Diagnostic Checking for Adequacy of Linear and Nonlinear Time Series Models

Yongmiao Hong
Department of Economics &
Department of Statistical Science
Cornell University
Ithaca, NY 14853-7601
yh20@cornell.edu

Tae-Hwy Lee
Department of Economics
University of California
Riverside, CA 92521-0427
taelee@mail.ucr.edu

May 2000

ABSTRACT

This paper proposes a new diagnostic test for linear and nonlinear time series models, using a generalized spectral density approach. Under a wide class of time series models that includes ARCH models, the proposed test enjoys the appealing "nuisance parameter-free" property that model parameter estimation has no impact on the limit distribution of the test statistic. In addition, it is consistent against all pairwise serial dependencies and allows the choice of a proper lag order via data-driven methods. It is asymptotically more efficient than the test of Brock *et al.* (1991, 1996), the well-known BDS test, in detecting a class of plausible local alternatives (although not for ARCH). A simulation study compares the finite sample performance of the new test and the tests of BDS, Box and Pierce (1971), Ljung and Box (1978), McLeod and Li (1983), and Li and Mak (1994). The new procedure has good power against a wide variety of stochastic and chaotic alternatives to the null models and can play a valuable role in evaluating adequacy of linear and nonlinear time series models.

Key words: Diagnostic test, Generalized spectrum, Model adequacy, Nonlinear time series model, Standardized residual.

1 Introduction

The development of nonlinear time series analysis has been advancing rapidly (cf. Subba Rao and Gabr 1984, Priestley 1988, Tong 1990, Brock *et al.* 1991, Granger and Teräsvirta 1993, Teräsvirta *et al.* 1994, Terdik 1999). Many time series in practice display non-Gaussian and nonlinear features. In modern time series analysis, one often considers the time series process

$$Y_t = g_0(I_{t-1}) + h_0(I_{t-1})\varepsilon_t, \qquad t = 0, \pm 1, ..., \tag{1.1}$$

where I_t is the information set available at period t, $g_0(I_{t-1}) = E(Y_t|I_{t-1})$ $a.s.$, $h_0^2(I_{t-1}) = \mathrm{var}(Y_t|I_{t-1})$ $a.s.$, and $\{\varepsilon_t\}$ is i.i.d. $(0,1)$. For various linear and nonlinear time series that belong to class (1.1), see (e.g.) Tong (1990) and Granger and Teräsvirta (1993).

Various models for $g_0(\cdot)$ and $h_0(\cdot)$ have been proposed. Consider a parametric model

$$Y_t = g(I_{t-1}, \theta) + h(I_{t-1}, \theta)e_t(\theta), \tag{1.2}$$

where $g(\cdot,\theta)$ and $h(\cdot,\theta)$ are some parametric specifications for $g_0(\cdot)$ and $h_0(\cdot)$, θ is an unknown finite dimensional parameter vector, and $\{e_t(\theta)\}$ is an unobservable series with mean 0 and variance 1. Specification (1.2) covers most commonly used linear and nonlinear time series models. Examples include ARMA, ARCH, bilinear, nonlinear moving average, Markov-switching, smooth transition, exponential and threshold autoregressive models. When $g(\cdot,\theta)$ and $h(\cdot,\theta)$ are correctly specified for $g_0(\cdot)$ and $h_0(\cdot)$; that is, when there exists some θ_0 such that $g(\cdot,\theta_0) = g_0(\cdot)$ $a.s.$ and $h(\cdot,\theta_0) = h_0(\cdot)$ $a.s.$, the series $\{e_t(\theta_0)\}$ coincides with $\{\varepsilon_t\}$, and therefore, is i.i.d. In contrast, if $g(\cdot,\theta)$ is inadequate for $g_0(\cdot)$ and/or $h(\cdot,\theta)$ is inadequate for $h_0(\cdot)$; that is, if there exists no θ such that $g(\cdot,\theta) = g_0(\cdot)$ $a.s.$ and/or $h(\cdot,\theta) = h_0(\cdot)$ $a.s.$, $\{e_t(\theta)\}$ will be serially dependent for all θ. Consequently, to test adequacy of model (1.2), one can check whether there exists some θ_0 such that $\{e_t(\theta_0)\}$ is i.i.d. Because $\{e_t(\theta_0)\}$ is unobservable, one has to consider the standardized estimated residual

$$\hat{e}_t = \left[Y_t - g(\hat{I}_{t-1}, \hat{\theta})\right] / h(\hat{I}_{t-1}, \hat{\theta}), \qquad t = 1, ..., n, \tag{1.3}$$

where $\hat{\theta}$ is an estimator of θ_0 and $\hat{I}_t$ is the *observed* information set available at period t that may use certain initial values. In constructing an asymptotically valid test procedure, it is important to examine whether and how the use of $\{\hat{e}_t\}_{t=1}^n$ rather than $\{e_t \equiv e_t(\mathrm{plim}\ \hat{\theta})\}_{t=1}^n$ affects the limit distribution of a test statistic, besides its power property.

When $g_0(I_{t-1})$ is linear in I_{t-1}, Y_t is called linear in conditional mean on I_{t-1} (cf. Lee *et al.* 1993). If in addition $h_0^2(I_{t-1}) = \sigma^2$ $a.s.$, $\{Y_t\}$ is called completely linear in I_{t-1} (cf. Granger 1998, Granger and Lee 1999). Assuming $h_0^2(I_{t-1}) = \sigma^2$ *a.s.*, Box and Pierce (1971) and Ljung and Box (1978) propose a diagnostic test for an ARMA(p_0, q_0) model:

$$\text{BPL}(p) = n(n+2)\sum_{j=1}^{p}(n-j)^{-1}\hat{\rho}^2(j) \xrightarrow{d} \chi^2_{p-(p_0+q_0)}, \qquad p > p_0 + q_0, \tag{1.4}$$

where $\hat{\rho}(j)$ is the sample autocorrelation function of $\{\hat{e}_t\}_{t=1}^n$. The degrees of freedom of the test depends on p_0+q_0, the number of the estimated parameters. Hong (1996) and Paparoditis (2000a,b) propose spectrum-based diagnostic tests which generalize BPL(p). The null limit distributions of these spectral tests do not depend on estimated parameters.

It is well-known that Box-Pierce-Ljung's test has no power against nonlinear dependencies with zero autocorrelations, such as some bilinear and nonlinear moving-average processes (cf. Granger and Anderson 1978, Granger 1983). Using the sample autocorrelation of squared residuals, McLeod and Li (1983) suggest a test for linearity against unspecified nonlinearity. It has high power against departures from linearity that have apparent ARCH structures; ARCH itself can be detected as well. The null limit distribution of the test is a χ^2 distribution; the degrees of freedom need not be adjusted when only an ARMA model is estimated, but have to be adjusted when conditional variance estimation is involved (cf. Li and Mak 1994). For an ARMA(p_0, q_0)-ARCH(r) model, a version of McLeod-Li-Mak test statistic can be written as

$$\text{MLM}(p) = n(n+2)\sum_{j=r+1}^{p}(n-j)^{-1}\hat{\rho}_2^2(j) \xrightarrow{d} \chi^2_{p-r}, \qquad p > r, \tag{1.5}$$

where $\hat{\rho}_2(j)$ is the sample autocorrelation function of $\{\hat{e}_t^2\}_{t=1}^n$. The null limit distribution depends on r, the order of the ARCH model. The test statistic itself has to be modified if other conditional variance model is estimated.

Brock *et al.* (1991, 1996) propose a diagnostic test for model (1.2), using chaos theory:

$$\text{BDS}(m, d) = n^{\frac{1}{2}}\left[\hat{C}_m(d) - \hat{C}_1(d)^m\right]/\hat{V}_m^{1/2}, \tag{1.6}$$

where the sample correlation integral

$$\begin{aligned}\hat{C}_m(d) &= \frac{2}{n(n-1)}\sum_{t=m+1}^{n}\sum_{s=m}^{t-1}\prod_{j=0}^{m-1}\mathbf{1}(|\hat{e}_{t-j} - \hat{e}_{s-j}| < d) \\ &\xrightarrow{p} P\left[\prod_{j=0}^{m-1}\mathbf{1}(|e_{t-j} - e_{s-j}| < d)\right] \equiv C_m(d),\end{aligned} \tag{1.7}$$

$\mathbf{1}(\cdot)$ is an indicator function, the integer m is called the embedding dimension, d is a distance parameter, and $\hat{V}_m$ is an asymptotic variance estimator. The BDS test was originally proposed to test whether $\{e_t\}$ is stochastic or chaotic. The statistic $\hat{C}_m(d)$ measures the fraction of pairs of histories $\{\hat{e}_{t-j}, \hat{e}_{s-j}\}_{j=0}^{m-1}$ that are within distance d of each other. If $\hat{e}_t$ and $\hat{e}_s$ are close in value, so will be subsequent pairs for a chaotic process, but not for a stochastic one. Thus, BDS(m,d) is expected to have good power against chaos. In addition, it also has power against a wide range of stochastic dependent alternatives to the i.i.d. hypothesis. To see this, observe that when $\{e_t\}$ is i.i.d.,

$$C_m(d) = C_1(d)^m \tag{1.8}$$

for *all* positive integers m and *all* distances $d > 0$. In other words, the correlation integral $C_m(d)$ behaves like the characteristic function of a serial string in the sense that the correlation integral of a serial string is the product of correlation integrals of component substrings. If $C_m(d) \neq C_1(d)^m$, there is evidence against the i.i.d. hypothesis and BDS will gain power.

As shown in Brock *et al.* (1991, Ch. 2 & Appendix D), BDS(m,d) has the appealing "nuisance parameter-free" property that any $\sqrt{n}$-consistent parameter estimator $\hat{\theta}$ has no impact on its null limit distribution under a class of linear and nonlinear conditional mean models $g(\cdot,\theta)$. This, together with good power against a wide range of dependent alternatives, makes BDS(m,d) a convenient and powerful diagnostic tool for time series models. It has been recommended by Brock *et al.* (1991) as a portmanteau lack of fit test in the nonlinear time series context much in the same spirit as Box and Jenkins (1976, p. 29) recommend Box-Pierce-Ljung's test in the linear time series context.

On the other hand, BDS(m,d) also has certain features one may consider undesirable. First, the "nuisance parameter-free" property holds only under conditional mean models but not under ARCH models (cf. Brock *et al.* 1991, Appendix D). When conditional variance estimation is involved, the limit distribution of BDS(m,d) depends on the nature of $\hat{\theta}$ and how to modify the test statistic remains unknown. Second, although serial independence implies (1.8), the converse is not true (Brock *et al.* 1991, p. 47). There are examples in which $\{e_t\}$ is not i.i.d. but (1.8) holds. For such alternatives, BDS(m,d) may have no power. Also, BDS(m,d) involves the choice of two parameters — m and d. Both m and d are arbitrary and fixed. No theory for their choice is available. Since $m-1$ is the largest lag order used, BDS(m,d) has no power against alternatives for which e_t and e_{t-j} are not independent only when $j \geq m$. Ideally, a proper choice of m should depend on the alternative, which, however, is unknown when serial dependence of $\{e_t\}$ is of unknown form. Similarly, some choice of d may render BDS(m,d) inconsistent against certain alternatives. In fact, as shown in Section 5 below, BDS(m,d) also has suboptimal power against some plausible local alternatives. For example, it can detect ARCH(1) with parametric rate $n^{-\frac{1}{2}}$ but MA(1) with rate $n^{-\frac{1}{4}}$ only.

Here, we propose a new diagnostic test for linear and nonlinear time series models, using a generalized spectral density function recently proposed in Hong (1999). The proposed test enjoys the "nuisance parameter-free" property of the BDS test under a *wider* class of time series models, which include but are not restricted to ARCH models. It is consistent against any type of pairwise serial dependence, a property not attainable by the BDS test. It can detect a class of local alternatives with a rate slightly slower than the parametric rate $n^{-\frac{1}{2}}$ but much faster than $n^{-\frac{1}{4}}$. This class includes both MA and ARCH. Finally, generalized spectral smoothing allows one to choose a lag order via some data-driven methods, which are more objective than an arbitrary choice or a rule-of-thumb and give more robust power. A simulation study compares the proposed test and the tests of BDS, Box-Pierce-Ljung, and McLeod-Li-Mak in finite samples. The new test has reasonable power against a variety of stochastic and chaotic alternatives to the null models. It is a useful addition to the existing diagnostic toolkit for time series models (see Barnett *et al.* 1997).

It should be pointed out that there are a variety of tests for serial dependence in the literature. These include the tests of Chan and Tran (1992), Cameron and Trivedi (1993), Delgado (1996), Hong (1998, 2000), Pinkse (1998), Skaug and Tjøstheim (1993a,b, 1996), and Robinson (1991). All of these tests are based on *observed* raw data rather than estimated standardized residuals. Whether and how the limit distributions of these tests will change when applied to estimated standardized residuals has not been investigated. For space we do not consider how to adapt these tests to estimated standardized residuals $\{\hat{e}_t\}$.

2 A New Diagnostic Test

A generalized spectral density function is recently proposed in Hong (1999) as an analytic tool for linear and nonlinear time series analysis. Suppose that $\{e_t\}$ is strictly stationary. The basic idea of the generalized spectrum is to consider the spectrum of the transformed series $\{e^{iue_t}\}$, where $u \in \mathbb{R}$. Define

$$\sigma_j(u,v) = \mathrm{cov}(e^{iue_t}, e^{ive_{t-j}}), \qquad j = 0, \pm 1, ..., \tag{2.1}$$

the covariance between e^{iue_t} and $e^{ive_{t-j}}$. Straightforward algebra yields

$$\sigma_j(u,v) = \varphi_j(u,v) - \varphi(u)\varphi(v), \tag{2.2}$$

where $\varphi_j(u,v) = E[e^{i(ue_t+ve_{t-j})}]$ and $\varphi(u) = E(e^{iue_t})$ are the joint and marginal characteristic functions of (e_t, e_{t-j}). Therefore, $\sigma_j(u,v) = 0$ for all $(u,v) \in \mathbb{R}^2$ if and only if e_t and e_{t-j} are independent. Suppose that $\sup_{(u,v)\in\mathbb{R}^2} \sum_{j=-\infty}^{\infty} |\sigma_j(u,v)| < \infty$, which holds when, for example, $\{e_t\}$ is a stationary α-mixing process with the mixing coefficients $\{\alpha(j)\}$ satisfying $\sum_{j=0}^{\infty} \alpha(j)^{(\nu-1)/\nu} < \infty$ for some $\nu > 1$. Then the Fourier transform of $\sigma_j(u,v)$

$$f(\omega, u, v) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \sigma_j(u,v) e^{-ij\omega}, \qquad \omega \in [-\pi, \pi], \tag{2.3}$$

exists. No moment condition on e_t is required. When $\text{var}(e_t)$ exists, however, the negative partial derivative of $f(\omega, u, v)$ with respect to (u,v) at $(0,0)$ delivers the conventional spectral density:

$$-\frac{\partial^2 f(\omega, u, v)}{\partial u \partial v}|_{(0,0)} = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} R(j) e^{-ij\omega},$$

where $R(j) = \text{cov}(e_t, e_{t-j})$. For this reason we call $f(\omega, u, v)$ a "generalized spectral density" of $\{e_t\}$. The introduction of auxiliary parameters (u,v) offers much flexibility in capturing serial dependence. The generalized spectrum $f(\omega, u, v)$ can capture all pairwise dependencies, including those with zero autocorrelations. Searching over the domain of (u,v), one can find the "maximal dependence" at each frequency ω, as given by

$$s(\omega) = \sup_{(u,v)\in\mathbb{R}^2} \left| \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \sigma_j(u,v) e^{-ij\omega} \right|, \qquad \omega \in [-\pi, \pi],$$

where $|\cdot|$ is the usual Euclidean norm. This maximal spectral dependence may be contributed from linear or nonlinear serial dependence in $\{e_t\}$. Therefore, a generalized spectral peak at some frequency may indicate a cycle or periodicity due to nonlinear dependence.

The generalized spectrum $f(\omega, u, v)$ differs from higher order spectra, which are the Fourier transforms of higher order cumulants and have been popular in nonlinear time series analysis (cf. Brillinger 1965, Brillinger and Rosenblatt 1967a,b, Subba and Gabr 1980, 1984, Hinich 1982, Terdik 1999). The generalized spectrum $f(\omega, u, v)$ does not require any moment condition on e_t. This may be appealing because, for example, many high frequency economic and financial time series have infinite variances (e.g., Fama and Roll 1968, Pagan and Schwert 1990). It can effectively capture any pairwise serially dependent processes, including ARCH with zero third cumulants. For such ARCH processes, the bispectrum, which is the Fourier transform of third order cumulants, will miss them. However, $f(\omega, u, v)$ cannot capture dependent processes that are pairwise serially independent, which may be captured by the bispectrum. It would be interesting to compare the generalized spectral density and the bispectrum thoroughly, but this is beyond the scope of this paper and will be pursued in subsequent study.

When $\{e_t\}$ is i.i.d., $f(\omega, u, v)$ becomes

$$f_0(\omega, u, v) = \frac{1}{2\pi}\sigma_0(u, v), \qquad \omega \in [-\pi, \pi], \tag{2.4}$$

a constant function of frequency ω. Any deviation of $f(\omega, u, v)$ from $f_0(\omega, u, v)$ is evidence of serial dependence of $\{e_t\}$. To test the i.i.d. hypothesis for $\{e_t\}$, one can compare two consistent estimators of $f(\omega, u, v)$ and $f_0(\omega, u, v)$ via an L_2-norm. Define

$$\hat{\sigma}_j(u, v) = \hat{\varphi}_j(u, v) - \hat{\varphi}_j(u, 0)\hat{\varphi}_j(0, v), \qquad j = 0, \pm 1, ..., \pm(n-1), \tag{2.5}$$

where

$$\hat{\varphi}_j(u, v) = \begin{cases} (n-j)^{-1}\sum_{t=1+j}^{n} e^{i(u\hat{e}_t + v\hat{e}_{t-j})} & \text{if } j \geq 0, \\ (n+j)^{-1}\sum_{t=1-j}^{n} e^{i(u\hat{e}_{t+j} + v\hat{e}_t)} & \text{if } j < 0. \end{cases} \tag{2.6}$$

Note that $\hat{\varphi}_j(u, v) = \hat{\varphi}_{-j}(v, u)$. A kernel estimator for $f(\omega, u, v)$ can be defined as

$$\hat{f}_n(\omega, u, v) = \frac{1}{2\pi}\sum_{j=1-n}^{n-1}(1 - |j|/n)^{\frac{1}{2}} k(j/p)\hat{\sigma}_j(u, v)e^{-ij\omega}, \tag{2.7}$$

where $k : \mathbb{R} \to [-1, 1]$ is a symmetric kernel and $p \equiv p_n$ is a bandwidth (or lag order) such that $p \to \infty$, $p/n \to 0$ as $n \to \infty$. Examples of $k(\cdot)$ include Bartlett, Daniell, Quadratic-Spectral, and truncated kernels (see, e.g., Priestley 1981, p. 441). The factor $(1 - |j|/n)^{\frac{1}{2}}$ is a finite sample correction which delivers better approximation to the finite sample distribution. We also have a consistent estimator for $f_0(\omega, u, v)$; namely,

$$\hat{f}_0(\omega, u, v) = \frac{1}{2\pi}\hat{\sigma}_0(u, v), \qquad \omega \in [-\pi, \pi]. \tag{2.8}$$

Let $W : \mathbb{R} \to \mathbb{R}^+$ be a nondecreasing function that weights sets about 0 equally. Without loss of generality, we assume that it has unit total variation. Examples of $W(\cdot)$ are the cumulative distribution functions of $N(0, 1)$, double exponential, and uniform distributions. Then a test for the i.i.d. hypothesis of $\{e_t\}$ can be based on a standardized L_2-norm:

$$\begin{aligned} &\hat{M}(p) \\ &= \left[n\pi \int\int_{-\pi}^{\pi} |\hat{f}_n(\omega, u, v) - \hat{f}_0(\omega, u, v)|^2 d\omega dW(u)dW(v) - \hat{C}_0\sum_{j=1}^{n-1}k^2(j/p)\right] \Big/ \left[2\hat{D}_0\sum_{j=1}^{n-1}k^4(j/p)\right]^{\frac{1}{2}} \\ &= \left[\int\sum_{j=1}^{n-1}k^2(j/p)(n-j)\left|\hat{\sigma}_j(u, v)\right|^2 dW(u)dW(v) - \hat{C}_0\sum_{j=1}^{n-1}k^2(j/p)\right] \Big/ \left[2\hat{D}_0\sum_{j=1}^{n-1}k^4(j/p)\right]^{\frac{1}{2}}, \end{aligned} \tag{2.9}$$

where the second equality follows from Parseval's identity, $\hat{\sigma}_j(u,v) = \hat{\sigma}_{-j}(v,u)$, and symmetry of $k(\cdot)$ and $W(\cdot)$. Moreover,

$$\hat{C}_0 = \left[\int \hat{\sigma}_0(u,-u)dW(u)\right]^2, \tag{2.10}$$

$$\hat{D}_0 = \left[\int |\hat{\sigma}_0(u,v)|^2 dW(u)dW(v)\right]^2. \tag{2.11}$$

Throughout this paper, unspecified integrals are taken over the entire Euclidean space of proper dimension. The test statistic $\hat{M}(p)$ involves 1- and 2-dimensional numerical integrations with respect to (u,v), which can be implemented using, for example, Gauss-Legendre quadratures. Note that $\hat{M}(p)$ involves no numerical integration over frequency ω, which has been integrated out due to the use of the L_2-norm. Divergence measures rather than the L_2-norm could be used, but they would generally involve numerical integrations over ω as well as over (u,v), and the distribution theory might be different as well. A GAUSS code for computing $\hat{M}(p)$ with p chosen via a data-driven method is available from the authors.

3 Asymptotic Distribution

We first derive the limit distribution of $\hat{M}(p)$ and establish its "nuisance parameter-free" property under a wide class of time series models. Below are regularity conditions.

ASSUMPTION A.1: $\{Y_t\}$ is a strictly stationary α-mixing process with $\sum_{j=0}^{\infty} \alpha(j)^{\frac{\nu-1}{\nu}} < \infty$ for some $\nu > 1$.

ASSUMPTION A.2: $n^{\frac{1}{2}}(\hat{\theta} - \theta_0) = O_P(1)$, where $\theta_0 = \text{plim}(\hat{\theta})$.

ASSUMPTION A.3: Let I_t be the pseudo information set from period t to the infinite past, and let Θ_0 be a small convex neighborhood of θ_0. The functions $g(I_t,\cdot)$ and $h(I_t,\cdot)$ are twice continuously differentiable with respect to $\theta \in \Theta_0$ *a.s.*, with $E\sup_{\theta\in\Theta_0} \left\|h^{-1}(I_t,\theta)\frac{\partial}{\partial\theta}g(I_t,\theta)\right\|^4 \le C$, $E\sup_{\theta\in\Theta_0} \left\|h^{-1}(I_t,\theta)\frac{\partial}{\partial\theta}h(I_t,\theta)\right\|^4 \le C$, $E\sup_{\theta\in\Theta_0} \left\|h^{-1}(I_t,\theta)\frac{\partial^2}{\partial\theta^2}g(I_t,\theta)\right\|^2 \le C$, $E\sup_{\theta\in\Theta_0} \left\|h^{-1}(I_t,\theta)\frac{\partial^2}{\partial\theta^2}h(I_t,\theta)\right\|^2 \le C$, and $E\left[\sup_{\theta\in\Theta_0} e_t^4(\theta)|I_{t-1}\right] \le C$, where $e_t(\theta) = [Y_t - g(I_{t-1},\theta)]/h(I_{t-1},\theta)$ and $C \in (0,\infty)$ is a bounded constant.

ASSUMPTION A.4: Let $\hat{I}_t$ be the observed information set available at period t that may use certain initial values. Then

$$\lim_{n\to\infty}\sum_{t=1}^{n} E\sup_{\theta\in\Theta_0}\left|\frac{h(I_t,\theta)-h(\hat{I}_t,\theta)}{h(\hat{I}_t,\theta)}\right| < \infty \text{ and } \lim_{n\to\infty}\sum_{t=1}^{n} E\sup_{\theta\in\Theta_0}\left|\frac{g(I_t,\theta)-g(\hat{I}_t,\theta)}{g(\hat{I}_t,\theta)}\right| < \infty.$$

ASSUMPTION A.5: $k : \mathbb{R} \to [-1, 1]$ is symmetric about 0, and is continuous at 0 and all points except a finite number of points, with $k(0) = 1, \int_0^\infty k^2(z)dz < \infty, |k(z)| \leq C|z|^{-b}$ as $z \to \infty$ for some $b > \frac{1}{2}$ and $C \in (0, \infty)$.

ASSUMPTION A.6: $W : \mathbb{R} \to \mathbb{R}^+$ is nondecreasing and weighs sets about 0 equally, with $\int_{-\infty}^\infty dW(u) = 1$ and $\int_{-\infty}^\infty u^4 dW(u) < \infty$.

ASSUMPTION A.7: $D_0 \equiv [\int |\sigma_0(u, v)|^2 dW(u)dW(v)]^2 > 0$.

These are conditions on the data generating process $\{Y_t\}$, parameter estimator $\hat{\theta}$, initial value conditions, models $g(\cdot, \theta)$ and $h(\cdot, \theta)$, and weight functions $k(\cdot)$ and $W(\cdot)$. In Assumption A.1, we permit but do not require $\mathrm{var}(Y_t) < \infty$. In Assumption A.2, we permit but do not require $\hat{\theta}$ to be a quasi-maximum likelihood estimator (e.g., Lee and Hansen 1994, Lumsdaine 1996). Any $\sqrt{n}$-consistent estimator $\hat{\theta}$ suffices. Assumption A.4 is a start-up value condition. It ensures that the impact of initial values (if any) assumed in $\hat{I}_t$ is asymptotically negligible. In Assumption A.5, the constant b governs the rate at which the kernel $k(z) \to 0$ as $z \to \infty$. For kernels with bounded support (e.g., Bartlett, Parzen, Tuckey and truncated kernels), $b = \infty$. For Daniell kernel and Quadratic-Spectral kernel, $b = 1$ and 2 respectively. Assumption A.7 ensures that the choice of $W(\cdot)$ does not lead to a degenerate test statistic.

Theorem 1: *Suppose that Assumptions A.1-A.7 hold, and* $p = cn^\lambda$ for $\lambda \in (0, 1)$ and $c \in (0, \infty)$. *Then if* $\{e_t\}$ *is i.i.d.,* $\hat{M}(p) \to^d N(0, 1)$.

Throughout, all the proofs are collected in the mathematical appendix. In Theorem 1, the use of any $\sqrt{n}$-consistent estimator $\hat{\theta}$ rather than θ_0 has no impact on the limit distribution of $\hat{M}(p)$. Thus, $\hat{M}(p)$ enjoys the same "nuisance parameter-free" property as the BDS test, but under a *wider* class of time series models—the "nuisance parameter-free" property holds under ARCH models for $\hat{M}(p)$ but not for BDS(m, d).

The asymptotic normality and "nuisance parameter-free" property of $\hat{M}(p)$ provide a convenient inference procedure. In finite samples, however, it may not be accurate. Such resampling methods as bootstrap are ideally suited to the present context, and are expected to give accurate sizes (cf. Skaug and Tjøstheim 1993b, 1996). The bootstrap procedure can be described as follows: (i) Estimate model (1.2) to obtain the standardized residuals $\{\hat{e}_t\}_{t=1}^n$ via (1.3). Compute the test statistic $\hat{M}(p)$ in (2.9) using the residuals $\{\hat{e}_t\}_{t=1}^n$. (ii) Draw a bootstrap sample $\{e_t^*\}_{t=1}^n$, with replacement, from $\{\hat{e}_t\}_{t=1}^n$. (iii) Obtain a bootstrap sample $\{Y_t^*\}_{t=1}^n$, where $Y_t^* = g(\hat{I}_{t-1}^*, \hat{\theta}) + h(\hat{I}_{t-1}^*, \hat{\theta})e_t^*$, and $\hat{I}_t^*$ is the observed information set at period t that is based on the bootstrap sample $\{Y_s^*\}_{s=1}^t$ and certain initial values. (iv) Estimate model (1.2) using the bootstrap sample $\{Y_t^*\}_{t=1}^n$ and obtain the standardized residuals $\{\hat{e}_t^*\}_{t=1}^n$. (v) Use $\{\hat{e}_t^*\}_{t=1}^n$ to compute the test statistic $\hat{M}^*(p)$. (vi) Repeat steps (i)–(v) B times, and obtain B bootstrap test statistics $\{\hat{M}_l^*(p)\}_{l=1}^B$. (vii) Compute the bootstrap P-value $p_B^* = B^{-1}\sum_{l=1}^B \mathbf{1}(\hat{M}_l^*(p) > \hat{M}(p))$.

4 Consistency

We now establish the consistency of $\hat{M}(p)$ under the alternative to the i.i.d. hypothesis.

Theorem 2: *Suppose that Assumptions A.1-A.7 hold, and $p = cn^\lambda$ for $\lambda \in (0,1)$ and $c \in (0,\infty)$. Then*

$$\frac{p^{1/2}}{n}\hat{M}(p) \xrightarrow{p} \left[2D_0\int_0^\infty k^4(z)dz\right]^{-\frac{1}{2}} \pi\int\int_{-\pi}^{\pi} |f(\omega,u,v) - f_0(\omega,u,v)|^2\, d\omega dW(u)dW(v)$$

$$= \left[2D_0\int_0^\infty k^4(z)dz\right]^{-\frac{1}{2}} \sum_{j=1}^{\infty}\int |\sigma_j(u,v)|^2\, dW(u)dW(v). \tag{4.1}$$

Suppose that e_t and e_{t-j} are not independent for some $j > 0$; then $\int|\sigma_j(u,v)|^2 dW(u)dW(v) > 0$ for any weight function $W(\cdot)$ that is positive, monotonically increasing and continuous with unbounded support on $\mathbb{R}$. Therefore, $\hat{M}(p)$ is consistent against any type of pairwise dependence for *any* $W(\cdot)$ satisfying the aforementioned conditions. The examples of $W(\cdot)$ include the cumulative distribution functions of $N(0,1)$, Double Exponential, and Student's t_ν distribution with degrees of freedom $\nu \geq 5$. Thus, we expect that $\hat{M}(p)$ may have relatively omnibus power against a wide variety of alternatives. Since the L_2-norm in (4.1) is positive whenever pairwise serial dependence exists, $\hat{M}(p)$ is an asymptotically one-sided $N(0,1)$ test. Upper-tailed asymptotic critical values should be used.

The choice of $W(\cdot)$ for $\hat{M}(p)$ may not be as important as the choice of distance parameter d for BDS(m,d), because the latter may render BDS(m,d) inconsistent against some alternatives whereas any $W(\cdot)$ that is positive, monotonically increasing and continuous with unbounded support on $\mathbb{R}$ always ensures consistency of $\hat{M}(p)$ against all pairwise dependencies. Nevertheless, the choice of $W(\cdot)$ might have impact on the power of $\hat{M}(p)$ in finite samples. We investigate this in our simulation below. Our results show that a variety of choices of $W(\cdot)$ have little impact on the size and power of $\hat{M}(p)$, while the choice of d has significant impact on the size and power of BDS(m,d).

5 Asymptotic Local Power

Local power analysis is insightful for the power property of a test. As noted by Tjøstheim (1996), however, it is rather difficult to do asymptotic local power analysis in the context of nonparametric testing for serial dependence. For simplicity, we consider a class of local alternatives for which there only exists first order serial dependence in $\{e_t\}$, and the joint probability density of (e_t, e_{t-1}) is

$$\mathbb{H}_n(a_n) : f_{1n}(x,y) = f_0(x)f_0(y)\left[1 + a_n g(x,y) + r_n(x,y)\right], \tag{5.1}$$

where $f_0(\cdot)$ is a marginal probability density, $r_n(\cdot,\cdot)$ is a remainder that may rise from the asymptotic expansion of $f_{1n}(\cdot,\cdot)$, and $a_n \to 0$ as $n \to \infty$ is the rate at which $\mathbb{H}_n(a_n)$ converges to the i.i.d. hypothesis. To ensure that $f_{1n}(\cdot,\cdot)$ is a valid joint probability density, we assume the following.

ASSUMPTION A.8: (i) $1 + a_n g(x,y) + r_n(x,y) \geqslant 0$ for all $(x,y) \in \mathbb{R}^2$ and all $n \geq 1$; (ii) $\int g(x,y)f_0(x)f_0(y)dxdy = 0$ and $\int r_n(x,y)f_0(x)f_0(y)dxdy = 0$ for all $n \geq 1$; and (iii) $\int g^4(x,y)f_0(x)f_0(y)dxdy < \infty$ and $\int |r_n(x,y)|^4 f_0(x)f_0(y)dxdy = o(a_n^4)$.

The condition on $r_n(\cdot,\cdot)$ ensures that the remainder $r_n(\cdot,\cdot)$ has no impact on the limit distribution of $\hat{M}(p)$. Two examples of $\mathbb{H}_n(a_n)$ are an MA(1) process

$$e_t = a_n \varepsilon_{t-1} + \varepsilon_t, \tag{5.2}$$

and an ARCH(1) process

$$e_t = \varepsilon_t \sqrt{1 + a_n \varepsilon_{t-1}^2}, \tag{5.3}$$

where ε_t is i.i.d. $N(0,1)$. Here $g(x,y) = xy$ for (5.2) and $g(x,y) = (x^2-1)(y^2-1)$ for (5.3).

Theorem 3: *Suppose that Assumptions A.1-A.8 hold, and* $p = cn^{\lambda}$ for $\lambda \in (0, \frac{1}{2})$ and $c \in (0, \infty)$. *Then* $\hat{M}(p) \to^d N(\mu, 1)$ *under* $\mathbb{H}_n(p^{\frac{1}{4}}/n^{\frac{1}{2}})$, *where the noncentrality*

$$\mu = \left[2D_0 \int_0^{\infty} k^4(z)dz\right]^{-\frac{1}{2}} \int \left| e^{i(ux+vy)} g(x,y) f_0(x) f_0(y) dxdy \right|^2 dW(u)dW(v).$$

When $g(x,y) \neq 0$, we have $\mu > 0$ provided $W(u)$ is positive, monotonically increasing and continuous with unbounded support on $\mathbb{R}$. Consequently, $\hat{M}(p)$ has nontrivial power against $\mathbb{H}_n(p^{\frac{1}{4}}/n^{\frac{1}{2}})$. The rate $p^{\frac{1}{4}}/n^{\frac{1}{2}}$ is slower than $n^{-\frac{1}{2}}$, because $p \to \infty$ as $n \to \infty$. This is the price one has to pay in order to achieve consistency against all pairwise serial dependencies. However, it is faster than $n^{-\frac{1}{4}}$ given $p/n \to 0$. If $p \propto \log(n)$, then $p^{\frac{1}{4}}/n^{\frac{1}{2}} \propto n^{-\frac{1}{2}} \log^{\frac{1}{4}}(n)$, which is nearly the same as $n^{-\frac{1}{2}}$. If $p \propto n^{\frac{1}{5}}$, as is the case with the data-driven method described below for some kernels, $p^{\frac{1}{4}}/n^{\frac{1}{2}} \propto n^{-\frac{1}{2}+\frac{1}{20}}$, which is only slightly slower than $n^{-\frac{1}{2}}$. We note that the use of $\{\hat{e}_t\}_{t=1}^n$ rather than $\{e_t\}_{t=1}^n$ has no impact on the asymptotic local power of $\hat{M}(p)$, so the conclusion of Theorem 3 applies to the tests considered in Hong (1999), where no local power analysis was provided.

The statistic $\hat{M}(p)$ involves the choice of kernel $k(\cdot)$. Given two kernels $k_1(\cdot)$ and $k_2(\cdot)$, Theorem 3 implies that Pitman's asymptotic relative efficiency of $k_2(\cdot)$ to $k_1(\cdot)$ under $\mathbb{H}_n(p^{\frac{1}{4}}/n^{\frac{1}{2}})$ in (5.1) is

$$\text{PARE}(k_2 : k_1) = \left[\int_0^\infty k_1^4(z)dz \Big/ \int_0^\infty k_2^4(z)dz\right]^{\frac{1}{2-\lambda}}.$$

For each kernel $k(\cdot)$, let $q \in \mathbb{R}$ be the largest positive number such that

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

exists, is nonzero and finite. We consider the class of kernels

$$\mathbb{K}(\tau) = \left\{k : \mathbb{R} \to [-1,1] \,|\, k(\cdot) \text{ satisfies Assumption A.5, } k^{(2)} = \frac{\tau^2}{2}, K(\eta) > 0\right\},$$

where $K(\eta) = (2\pi)^{-1}\int_{-\infty}^{\infty} k(z)e^{i\eta z}dz$ is the Fourier transform of $k(z)$. The condition $k^{(2)} = \tau^2/2$ provides a normalization for the kernels in $\mathbb{K}(\tau)$. This class of kernels is often considered in deriving the optimal kernel for spectral estimation (cf. Priestley 1981). It has been shown (cf. Hong 1999, the proof of Theorem A.6) that Daniell kernel $k(z) = \sin(\sqrt{3}z)/\sqrt{3}z$ minimizes $\int_0^\infty k^4(z)dz$ over $\mathbb{K}(\tau)$. Therefore, Daniell kernel is most efficient over $\mathbb{K}(\tau)$ in terms of Pitman's efficiency criterion. However, numerical calculations show that some commonly used kernels have rather similar asymptotic efficiencies. Pitman's relative efficiency of Daniell kernel to Parzen and Quadratic-Spectral kernels, for example, are $(1.0961)^{1/(2-\lambda)}$ and $(1.0079)^{1/(2-\lambda)}$ respectively. Thus, the choice of $k(\cdot)$ may be of secondary importance.

It is of interest to compare the asymptotic local power of $\hat{M}(p)$ and BDS(m,d). For simplicity, we consider BDS$(2,d)$ under a subclass of $\mathbb{H}_n(a_n)$ where $g(x,y) = g_1(x)g_2(y)$ for some functions $g_l : \mathbb{R} \to \mathbb{R}$, $l = 1,2$. BDS$(2,d)$ has nontrivial power under $\mathbb{H}_n(a_n)$ if the limit noncentrality

$$\lim_{n\to\infty} \sqrt{n}\left[C_2(d) - C_1(d)^2\right] \neq 0. \tag{5.4}$$

Straightforward algebra shows that under $\mathbb{H}_n(a_n)$ with $g(x,y) = g_1(x)g_2(y)$,

$$C_2(d) - C_1(d)^2$$

$$= \int\int\int\int \mathbf{1}(|x-x'|<d)\mathbf{1}(|y-y'|<d)f_{1n}(x,y)f_{1n}(x',y')dxdx'dydy'$$

$$
\begin{aligned}
& - \left[\int\int \mathbf{1}(|x-y|<d) f_{0n}(x) f_{0n}(y) dx dy \right]^2 \\
= \quad & 2a_n \int\int \mathbf{1}(|x-y|<d) g_1(x) f_0(x) f_0(y) dx dy \int\int \mathbf{1}(|x-y|<d) g_2(y) f_0(x) f_0(y) dx dy \\
& + a_n^2 \left[\int\int \mathbf{1}(|x-y|<d) g_1(x) g_2(y) f_0(x) f_0(y) dx dy \right]^2 + o(a_n^2), \qquad (5.5)
\end{aligned}
$$

where $f_{0n}(\cdot)$ is the marginal density of e_t under $\mathbb{H}_n(a_n)$, which may not be the same as $f_0(\cdot)$, the marginal density of e_t when $\{e_t\}$ is i.i.d. If the first term in (5.5) is identically 0 for all n, the asymptotic local power of BDS$(2,d)$ will depend on the second term, which renders BDS$(2,d)$ only able to detect $\mathbb{H}_n(n^{-\frac{1}{4}})$. This occurs when the marginal density $f_0(\cdot)$ is uniform. Alternatively, suppose that $f_0(x) = (2\pi)^{-\frac{1}{2}} \exp(-x^2/2)$, and

$$
g_j(-x) = -g_j(x) \text{ for all } x \in \mathbb{R}, \text{ and for } j = 1 \text{ or } 2. \qquad (5.6)
$$

Then the first term in (5.5) is identically 0 for all n, because the integral

$$
\begin{aligned}
& \int\int \mathbf{1}(|x-y|<d) f_0(y) dy g_j(x) f_0(x) dx \\
= \quad & \int \int_{x_1-d}^{x_1+d} f_0(y) dy g_l(x) f_0(x) dx \\
= \quad & \int \int_{x_1-d}^{x_1+d} (2\pi)^{-\frac{1}{2}} e^{-\frac{y^2}{2}} dy g_l(x) (2\pi)^{-\frac{1}{2}} e^{-\frac{x^2}{2}} dx \\
= \quad & \pi^{-1} \int \int_{-d/2}^{d/2} g_l(y-z) e^{-(y^2+z^2)} dy dz \\
= \quad & 0 \qquad\qquad \text{for } l = 1 \text{ or } 2, \qquad (5.7)
\end{aligned}
$$

where the third equality follows from changes of variable, and the last one follows from (5.6). Note that the MA(1) process in (5.2), where $g_1(x) = x$ and $g_2(y) = y$, satisfies condition (5.6), and the ARCH(1) process in (5.3), where $g_1(x) = x^2 - 1$ and $g_2(y) = y^2 - 1$, does not satisfy condition (5.6). Thus, BDS$(2,d)$ can detect ARCH(1) with rate $n^{-\frac{1}{2}}$ but MA(1) with rate $n^{-\frac{1}{4}}$ only. This explains why it is often found in practice that BDS(m,d) has excellent power against ARCH (e.g. Brock *et al.* 1991).

6 Choice of Data-Driven Bandwidth

Both BDS(m,d) and $\hat{M}(p)$ involve the choice of m or p. Brock *et al.* (1991) recommend some simple "rule-of-thumb" that m be small for finite sample sizes. In contrast, the generalized spectral smoothing approach provides a data-driven method to choose p, which, to some extent, lets data themselves speak for a proper p for $\hat{M}(p)$. Before discussing specific data-driven methods, we first justify the use of a data-driven lag order $\hat{p}$. We impose a Lipschitz continuity condition on $k(\cdot)$, which rules out the truncated kernel $k(z) = \mathbf{1}(|z| \leq 1)$, but it includes most commonly used kernels.

ASSUMPTION A.9: *For any $x, y \in \mathbb{R}$, $|k(x) - k(y)| \leq C|x - y|$ for some constant $C \in (0, \infty)$.*

Theorem 4: *Suppose that Assumptions A.1-A.7 and A.9 hold, and $\hat{p}$ is a data-driven bandwidth such that $\hat{p}/p = 1 + O_P(p^{-(\frac{3}{2}\beta - 1)})$ for some $\beta > (2b - \frac{1}{2})/(2b - 1)$, where b is as in Assumption A.5, $p = cn^{\lambda}$ with $\lambda \in (0, 1)$* and *$c \in (0, \infty)$. Then if $\{e_t\}$ is i.i.d., $\hat{M}(\hat{p}) - \hat{M}(p) \to^p 0$ and $\hat{M}(\hat{p}) \to^d N(0, 1)$.*

Thus, as long as $\hat{p}$ converges to p sufficiently fast, the use of $\hat{p}$ rather than p has no impact on the limit distribution of $\hat{M}(\hat{p})$, an additional "nuisance parameter-free" property. This extends Hong's (1999) results to the standardized residuals of model (1.2).

Theorem 4 allows for a wide range of admissible rates for $\hat{p}$. One plausible choice of $\hat{p}$ is the plug-in method considered in Hong (1999), which asymptotically minimizes an integrated mean square error criterion for $\hat{f}_n(\cdot, \cdot, \cdot)$. This method is described as follows. Consider the "pilot" estimators based on a preliminary bandwidth $\bar{p}$:

$$\bar{f}_n(\omega, u, v) = \frac{1}{2\pi} \sum_{j=1-n}^{n-1} (1 - |j|/n)^{\frac{1}{2}} \bar{k}(j/\bar{p}) \hat{\sigma}_j(u, v) e^{-ij\omega}, \tag{6.1}$$

$$\bar{f}_n^{(q,0,0)}(\omega, u, v) = \frac{1}{2\pi} \sum_{j=1-n}^{n-1} (1 - |j|/n)^{\frac{1}{2}} \bar{k}(j/\bar{p}) \hat{\sigma}_j(u, v) |j|^q e^{-ij\omega}, \tag{6.2}$$

where $\bar{k} : \mathbb{R} \to [-1, 1]$ is a kernel not necessarily the same as the kernel $k(\cdot)$ used in (2.7). For example, $\bar{k}(\cdot)$ can be Bartlett kernel while $k(\cdot)$ is Daniell kernel. Note that $\bar{f}_n(\cdot, \cdot, \cdot)$ is an estimator for $f(\cdot, \cdot, \cdot)$ and $\bar{f}_n^{(q,0,0)}(\cdot, \cdot, \cdot)$ is an estimator for the generalized spectral derivative

$$f^{(q,0,0)}(\omega, u, v) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \sigma_j(u, v) |j|^q e^{-ij\omega}. \tag{6.3}$$

The plug-in bandwidth is then defined as

$$\hat{p}_0 = \hat{c}_0 n^{\frac{1}{2q+1}}, \tag{6.4}$$

where $\hat{c}_0$ is the tuning parameter estimator given by

$$\hat{c}_0 = \left[\frac{2q(k^{(q)})^2}{\int_{-\infty}^{\infty} k^2(z) dz} \frac{\int \int_{-\pi}^{\pi} |\bar{f}^{(q,0,0)}(\omega, u, v)|^2 d\omega dW(u) dW(v)}{\int_{-\pi}^{\pi} [\int \bar{f}_n(\omega, u, -u) dW(u)]^2 d\omega} \right]^{\frac{1}{2q+1}}$$

$$= \left[\frac{2q(k^{(q)})^2}{\int_{-\infty}^{\infty} k^2(z)dz} \frac{\sum_{j=1-n}^{n-1}(n-|j|)\bar{k}^2(j/\bar{p})|j|^{2q} \int |\hat{\sigma}_j(u,v)|^2 dW(u)dW(v)}{\sum_{j=1-n}^{n-1}(n-|j|)\bar{k}^2(j/\bar{p}) \left[\int \hat{\sigma}_j(u,-u)dW(u)\right]^2} \right]^{\frac{1}{2q+1}}. \quad (6.5)$$

The second equality in (6.5) follows from Parseval's identity.

The data-driven $\hat{p}_0$ still involves the choice of a preliminary bandwidth $\bar{p}$, which can be fixed or grow with the sample size n. If $\bar{p}$ is fixed, $\hat{p}_0$ still grows at rate $n^{\frac{1}{2q+1}}$ in general, but $\hat{c}_0$ does not converge to the optimal tuning constant. Hong (1999) shows that when $\bar{p}$ grows with n properly, the data-driven bandwidth $\hat{p}_0$ in (6.4) asymptotically minimizes an integrated mean square error of $\hat{f}_n(\cdot,\cdot,\cdot)$. Note that $\hat{p}_0$ is real-valued. One can take its integer part, and the impact of integer-clipping is expected to be negligible. The choice of $\bar{p}$ is somewhat arbitrary, but we expect that the choice of $\bar{p}$ is of secondary importance and may have no significant impact on $\hat{M}(\hat{p}_0)$. This is confirmed in our simulation below.

7 Monte Carlo Evidence

We now compare the size and power of $\hat{M}(\hat{p}_0)$, BDS(m,d), Box-Pierce-Ljung's test (BPL(p)), and McLeod-Li-Mak's test (MLM(p)), in finite samples. We check adequacy of two basic time series models—AR(1) and ARCH(1). With the null AR(1) model, we examine the size of the tests, their power against a variety of neglected dynamics and neglected nonlinearities in conditional mean, and their power to distinguish AR(1) from a chaotic alternative that has the same autocorrelation structure as AR(1). With the null ARCH(1) model, we examine the size of the tests, their power against misspecification in conditional variance, their power to distinguish ARCH(1) from nonlinearities in mean that result in apparent ARCH structures, and their power to distinguish ARCH(1) from a chaotic process that behaves like a white noise but has similar autocorrelations in squares to ARCH(1).

7.1 Testing Conditional Mean Model

We first examine the adequacy of an AR(1) model:

$$\text{Model A}: \qquad Y_t = a + bY_{t-1} + e_t, \qquad t = 1, ..., n,$$

under the following data generating processes (DGP).

DGP A.0 (AR(1))

$$Y_t = 0.6Y_{t-1} + \varepsilon_t.$$

DGP A.1 (AR(2))

$$Y_t = 0.6Y_{t-1} - 0.5Y_{t-2} + \varepsilon_t.$$

DGP A.2 (ARMA(1,1))

$$Y_t = 0.6Y_{t-1} + 0.5\varepsilon_{t-1} + \varepsilon_t.$$

DGP A.3 (Bilinear)

$$Y_t = 0.6Y_{t-1} + 0.7Y_{t-2}\varepsilon_{t-1} + \varepsilon_t.$$

DGP A.4 (Nonlinear MA)

$$Y_t = 0.6Y_{t-1} + 0.7\varepsilon_{t-1}\varepsilon_{t-2} + \varepsilon_t.$$

DGP A.5 (Threshold AR, TAR)

$$Y_t = \begin{cases} 0.6Y_{t-1} + \varepsilon_t, & \text{if } Y_{t-1} < 1, \\ -0.5Y_{t-1} + \varepsilon_t, & \text{if } Y_{t-1} \geq 1. \end{cases}$$

DGP A.6 (Markov Switching)

$$Y_t = \begin{cases} 0.6Y_{t-1} + \varepsilon_t, & \text{if } S_t = 0, \\ -0.5Y_{t-1} + \varepsilon_t, & \text{if } S_t = 1, \end{cases}$$

where S_t is an unobservable 2-state Markov chain with transition probabilities $P(S_t = 1|S_{t-1} = 0) = P(S_t = 0|S_{t-1} = 1) = 0.3$.

DGP A.7 (Sign Autoregressive, SIGN)

$$Y_t = \text{sign}(Y_{t-1}) + \sigma\varepsilon_t, \qquad \sigma = 0.43,$$

where $\text{sign}(x) = \mathbf{1}(x > 0) - \mathbf{1}(x < 0)$.

DGP A.8 (Tent Map)

$$Y_t = \begin{cases} \alpha^{-1}Y_{t-1}, & \text{if } 0 \leq Y_{t-1} < \alpha, \\ (1-\alpha)^{-1}(1 - Y_{t-1}), & \text{if } \alpha \leq Y_{t-1} \leq 1, \end{cases}$$

where $\alpha = 0.49999$ and Y_0 is generated from the uniform distribution on [0 1].

DGPs A.1–A.2 are used to check the power of tests against neglected dynamics in mean, and DGPs A.3–A.6 are used to check against various neglected nonlinearities in mean. DGP A.7, the SIGN autoregressive model, examined in Granger and Teräsvirta (1999), is a first-order nonlinear autoregressive process but has the same autocorrelation function as an AR(1) process: $\rho(j) = (1 - 2q)^{|j|}$, where $q = P(\varepsilon_t < -1) = P(\varepsilon_t > -1)$, when ε_t is symmetric. Following Granger and Teräsvirta (1999), we choose $\sigma = 0.43$ so that $q = 0.01$ if ε_t is $N(0, 1)$. Granger and Teräsvirta

(1999) find that such a nonlinear process has a misleading linear property in the sense that its estimated autocorrelations decline slower than the exponential rate and their pattern resembles a stationary fractionally integrated process. DGP A.8, the tent map, is a deterministic chaotic process, but it resembles in autocorrelations an AR(1) process with the AR coefficient $(2\alpha - 1)$ and the j-th autocorrelation $\rho(j) = (2\alpha - 1)^{|j|}$; see Sakai and Tokumaru (1980). DGPs A.7 and A.8 allow us to examine how a test can distinguish an AR(1) model from nonlinear stochastic and chaotic processes that behave like linear models in terms of autocorrelations.

7.2 Testing Conditional Variance Model

Next, we examine the adequacy of an ARCH(1) model:

$$\text{Model B}: \qquad Y_t = h_t e_t,\ h_t^2 = a + bY_{t-1}^2,\ \{e_t\} \sim \text{i.i.d.}(0,1), \qquad t = 1, ..., n,$$

when Y_t is generated from the following generating processes:

DGP B.0 (ARCH(1))

$$Y_t = h_t \varepsilon_t,\ h_t^2 = 0.9 + 0.1Y_{t-1}^2.$$

DGP B.1 (ARCH(2))

$$Y_t = h_t \varepsilon_t,\ h_t^2 = 0.1 + 0.1Y_{t-1}^2 + 0.8Y_{t-2}^2.$$

DGP B.2 (GARCH(1,1))

$$Y_t = h_t \varepsilon_t,\ h_t^2 = 0.1 + 0.1Y_{t-1}^2 + 0.8h_{t-1}.$$

DGP B.3 (EGARCH(1,1))

$$Y_t = h_t \varepsilon_t,\ \ln h_t^2 = 0.01 + 0.9 \ln h_{t-1}^2 + 0.3(|\varepsilon_{t-1}| - (2/\pi)^{1/2}) - 0.8\varepsilon_{t-1}.$$

DGP B.4 (Stochastic Volatility)

$$Y_t = h_t \varepsilon_t,\ h_t^2 = 0.1Y_{t-1}^2 + \exp(\varphi \ln h_{t-1}^2 + v_t),\ \{v_t\} \sim \text{i.i.d. N}(0,1),\ \varphi = 0.98.$$

DGP B.5 (Bilinear)

$$Y_t = 0.8Y_{t-1}\varepsilon_{t-1} + \varepsilon_t.$$

DGP B.6 (TAR)

$$Y_t = \begin{cases} -0.5Y_{t-1} + \varepsilon_t, & Y_{t-1} \geq 1, \\ \ \ 0.8Y_{t-1} + \varepsilon_t, & Y_{t-1} < 1. \end{cases}$$

DGP B.7 (Nonlinear MA)

$$Y_t = 0.8\varepsilon_{t-1}^2 + \varepsilon_t.$$

DGP B.8 (Logistic Map)

$$Y_t = 4Y_{t-1}(1 - Y_{t-1}),$$

where Y_0 is generated from the uniform distribution on [0, 1].

DGPs B.1–B.4 are used to examine the power of the tests against misspecification in conditional variance. In DGP B.4, parameter $\varphi = 0.98$ is an empirically relevant value; Harvey *et al.* (1994) estimated it in range of 0.9575–0.9948 for four different daily foreign exchange rates. DGPs B.5–B.7 allow us to examine the power to distinguish ARCH from a variety of nonlinearities in mean that result in apparent ARCH structures. Such distinction has important implications in practice (Weiss 1986, Bera and Higgins 1997, Diebold 1986). DGP B.8, the logistic map, behaves like a white noise but has similar autocorrelations in squares to an ARCH(1) process. It is used to examine the power of a test to distinguish ARCH from a chaotic process with similar autocorrelations in squares; see Granger and Teräsvirta (1993, p. 34).

For all the DGPs except the chaotic processes A.8 and B.8, we use the GAUSS random number generator to generate the i.i.d. innovation $\{\varepsilon_t\}$ from four distributions: (i) $N(0,1)$; (ii) Exponential; (iii) Mixed Normal, $P[\varepsilon_t \sim N(-3,1)] = P[\varepsilon_t \sim N(3,1)] = 0.5$; and (iv) Student's t_5. All the ε_t have been rescaled to have mean 0 and variance 1. We generate $n + 1000$ observations for $\{\varepsilon_t\}$ under each distribution, and then discard the first 1000 ones to alleviate the impact of using some initial values. We report the sizes for all four error distributions, but for space we report the power only for the normal error.

7.3 Monte Carlo Evidence

To compute $\hat{M}(\hat{p}_0)$, BDS(m,d), BPL(p) and MLM(p), we use the residual series $\hat{e}_t = Y_t - \hat{a} - \hat{b}Y_{t-1}$ from Model A, estimated by ordinary least squares method, and the standardized residual series $\hat{e}_t = Y_t/\hat{h}_t$ where $\hat{h}_t^2 = \hat{a} + \hat{b}Y_{t-1}^2$ from Model B, estimated by quasi-maximum likelihood method with a Gaussian likelihood function.

For $\hat{M}(\hat{p}_0)$, we use Daniell kernel $k(z) = \sin(\pi z)/\pi z$, which enjoys the optimal power property over a class of kernels (cf. Section 5). To examine the impact of the choice of preliminary bandwidths $\bar{p}$ on the size and power of $\hat{M}(\hat{p}_0)$, we consider $\bar{p} = 1$ to 10. To investigate the impact of the choice of weight function $W(\cdot)$ on the size and power of $\hat{M}(\hat{p}_0)$, we consider the three distribution functions:

(i) $N(0,1)$, (ii) Double Exponential, (iii) and t_5-distribution. They are all scaled to have mean 0 and variance 1.

For BDS(m,d), Brock *et al.* (1991) recommend using d in range $0.5\sigma - 1.5\sigma$, and m in range $2-5$, for $n = 500$ to 1000, where $\sigma^2 = \text{var}(Y_t)$. To examine the impact of the choice of embedding dimension m on the size and power of BDS(m,d), we use $m = 2$ to 11, which is equivalent to the choice of a lag order p from 1 to 10. As some data generating processes may have no finite variance, we consider three choices of distance parameter: $d = 0.5, 0.25, 0.125$, in the unit of data range. For normal random samples, these choices roughly correspond to $2\sigma, \sigma$ and 0.5σ, respectively.

For Box-Pierce-Ljung test, BPL(p), we use $p = 2$ to 10 for Model A ($p = 1$ cannot be chosen due to the adjustment of the degree of freedoms for its asymptotic distribution), and $p = 1$ to 10 for Model B.

For MLM(p), we use $p = 1$ to 10 for Model A, and $p = 2$ to 10 for Model B (similarly to BPL(p), $p = 1$ cannot be chosen here).

To examine the sizes of the tests under the null AR(1) model and under the null ARCH(1) model, we estimate Model A and Model B under DGP A.0 and DGP B.0 respectively. We consider the size at the 10%, 5% and 1% levels for $n = 100$ and 200, using asymptotic critical values and 1000 monte carlo iterations. To conserve space, Figures 1 and 2 only report the sizes of the tests at the 5% level for $n = 100$. The results at 10%, 1% levels and the results for $n = 200$ are available from the authors. The sizes with $n = 200$ are very similar to those with $n = 100$.

To examine the powers of the tests against various misspecifications of the AR(1) model, we report in Figure 3 the powers of the tests under DGPs A.1–A.8, each of which is fitted by an AR(1) model. To examine the powers of the tests against various misspecifications of the ARCH(1) model, we report in Figure 4 the powers of the tests under DGPs B.1–B.8, each of which is fitted by an ARCH(1) model. The power is size-adjusted by using empirical critical values obtained under DGPs A.0 and B.0, which provide a fair comparison among the tests under study. We only report the power at the 5% level, for $n = 100$ and normal errors $\{\varepsilon_t\}$, using 1000 replications.

In Figures 1–4, the sizes or powers of $\hat{M}(\hat{p}_0)$, BDS(m,d), BPL(p), and MLM(p) are plotted as functions of $\bar{p}$, $(m-1)$, p, and p in the horizontal axis, respectively. In each graph, there are three plots for $\hat{M}(\hat{p}_0)$ in solid lines, denoted as M_l $(l = 1, 2, 3)$, that correspond to three weight functions $W(\cdot)$ — the distribution functions of $N(0,1)$, Double Exponential, and t_5. There are also three plots for BDS(m,d) in dashed lines, denoted as BDS_l $(l = 1, 2, 3)$, that correspond to three distance parameter values — $d = (0.5)^l$. MLM(p) is plotted in dotted lines and BPL(p) is plotted with more closely spaced dots.

We first examine the sizes in Figures 1 and 2. We observe the following patterns:

1. Overall, the sizes of the tests $\hat{M}(\hat{p}_0)$, BPL(p) and MLM(p) under the null AR(1) model and the null ARCH(1) model are relatively reasonable while the size of BDS(m, d) appears not very satisfactory. The unsatisfactory size performance of BDS(m, d) under the null ARCH(1) model may be due to its violation of the "nuisance parameter free" property under the ARCH(1) model.

2. The size of $\hat{M}(\hat{p}_0)$ is robust to the choice of weight function $W(\cdot)$ and preliminary bandwidth $\bar{p}$. The sizes of BPL(p) and MLM(p) are excellent and robust to the choice of lag order p. On the other hand, the size of BDS(m, d) is sensitive to the choices of distance parameter d and embedding dimension m. The fact that BDS(m, d) is sensitive to m while $\hat{M}(\hat{p}_0)$ is not sensitive to $\bar{p}$ indicates the practical value of the data-driven choice of lag order $\hat{p}_0$ for $\hat{M}(\hat{p}_0)$.

3. $\hat{M}(\hat{p}_0)$ is slightly under-sized under the null AR(1) with normal errors or is slightly over-sized under the ARCH(1) model with exponential and mixed normal errors, but not excessive. The size distortion of BDS(m, d) is quite large, especially under the mixed normal errors, which is consistent with the findings of Brock *et al.* (1991, p. 50).

4. The size patterns of each test under the null AR(1) model and the null ARCH(1) model are more or less similar.

In Figures 1 and 2 we consider the sizes of the tests under an AR coefficient 0.6 for DGP A.0 and the sizes of the tests with ARCH coefficient 0.1 for DGP B.0. We have also experimented (not reported) with the a variety of coefficient values: 0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, and 0.9 in both cases. In most cases, the sizes of $\hat{M}(\hat{p}_0)$, BDS(m, d), BPL(p), and MLM(p) are generally robust to the values of the AR(1) and ARCH(1) coefficients in DGP A.0 and DGP B.0. On exception is BPL(p), which tends to overreject under the null AR(1) model when the AR coefficient is close to 1 (say, 0.9) in DGP A.0 and lag order p is small, but its size becomes reasonable for larger lag order p, say $p > 5$.

We now turn to examine the powers of the tests. For fair comparison of power and to take into account the BDS's violation of the nuisance parameter free property under the ARCH(1) model, we use empirical critical values. We first examine the powers of the tests against various misspecification of an AR(1) model, as reported in Figure 3. We observe the following patterns:

1. The power of $\hat{M}(\hat{p}_0)$ is generally not sensitive to the choice of the preliminary bandwidth (or lag order) $\bar{p}$. The power of BDS(m, d) is sensitive to the choice of the embedding dimension m, which is equivalent to the choice of a lag order. The tests BPL(p) and MLM(p) are also sensitive to the choice of lag order p in some cases.

2. The power of $\hat{M}(\hat{p}_0)$ is robust to the choices of weight function $W(\cdot)$, whereas the power of BDS(m, d) is sensitive to the choice of distance parameter d.

3. (a) The autocorrelation test BPL(p) has an excellent power against AR(2), ARMA(1,1) and SIGN alternatives to the AR(1) model. Nevertheless, as expected, BPL(p) cannot detect the nonlinear alternatives — bilinear, nonlinear MA, TAR, Markov switching. It also can not distinguish AR(1) from the tent map, which resembles an AR(1) process in autocorrelations but is completely deterministic.

   (b) The correlation in squares test MLM(p) has good power against bilinear, nonlinear MA, Markov switching, and tent map alternatives to the AR(1) model. However, it has low power against AR(2), ARMA(1,1), TAR and SIGN alternatives.

   (c) BDS(m, d) has good power against bilinear, nonlinear MA, and the tent map alternatives to the AR(1) model. However, it has low power against ARMA(1,1), TAR, and SIGN alternatives.

   (d) The generalized spectral test $\hat{M}(\hat{p}_0)$ has excellent power against AR(2), ARMA(1,1), bilinear, TAR, SIGN, and tent map alternatives to the AR(1) model.

4. Overall speaking, $\hat{M}(\hat{p}_0)$ has relatively omnibus power against all linear and nonlinear dependent alternatives except for nonlinear MA. Moreover, it is more powerful than the other tests in many cases.

Next, we turn to examine the powers of the tests against various misspecifications of the ARCH(1) model, as reported in Figure 4. We observe the following patterns.

1. Like in testing misspecifications in conditional mean, the power of $\hat{M}(\hat{p}_0)$ is generally robust to the choice of the preliminary bandwidth (or lag order) $\bar{p}$ in most cases. The power of BDS(m, d) is sensitive to the choice of the embedding dimension m. The tests BPL(p) and MLM(p) are also sensitive to the choice of lag order p in some cases.

2. The power of $\hat{M}(\hat{p}_0)$ is robust to the choices of weight function $W(\cdot)$, whereas the power of BDS(m, d) is sensitive to the choice of distance parameter d.

3. (a) The correlation-based test BPL(p) has low power against ARCH(2), GARCH(1,1), EGARCH(1,1), stochastic volatility, bilinear, nonlinear MA, and logistic map alternatives to the ARCH(1) models. These alternatives are either martingale difference sequences or serially uncorrelated processes. However, BPL(p) has good power to distinguish TAR from ARCH(1).

   (b) The correlation in squares test MLM(p) is most powerful against ARCH(2), for which it has the optimal power by its design. Nevertheless, MLM(p) has low power against other forms of conditional heteroskedastic alternatives to ARCH(1), such as GARCH(1,1), EGARCH(1,1), and stochastic volatility models. Moreover, it cannot distinguish ARCH(1) from bilinear, TAR, nonlinear MA and logistic map processes. Many of these nonlinear conditional mean models have similar moment structures to ARCH(1). In particular, the logistic map behaves like a white noise but has similar autocorrelations in squares to ARCH(1) (cf. Granger and Teräsvirta, 1993, p. 34).

   (c) BDS(m, d) has poor power against bilinear, TAR, and nonlinear MA alternatives to the null ARCH(1) model. This finding is consistent with the findings of Brooks and Heravi (1999), who documents that BDS(m, d) is a fairly poor discriminator of bilinear and TAR processes from ARCH processes. Such distinctions have important implications in terms of predictivity (cf. Bera and Higgins 1997, Weiss 1986).

   (d) $\hat{M}(\hat{p}_0)$ has high power against EGARCH(1,1) and stochastic volatility alternatives to the ARCH(1) alternatives. It has high power to distinguish ARCH(1) from bilinear, TAR, nonlinear MA and logistic map processes. It may be interesting to note that $\hat{M}(\hat{p}_0)$ has better power than BDS(m, d) against the chaotic logistic map alternative.

4. Overall speaking, the generalized spectral test $\hat{M}(\hat{p}_0)$ has omnibus power against all alternatives except for GARCH(1,1). For the GARCH(1,1) alternative to ARCH(1), all the tests have low or little power. In most cases, $\hat{M}(\hat{p}_0)$ is the most powerful.

8 Conclusions

In addition to the popular tests of McLeod and Li (1983) and Li and Mak (1994), the correlation integral-based test by Brock *et al.* (1991, 1996) has been recently proposed as a portmanteau test for adequacy of nonlinear time series models. The test has the nice "nuisance parameter-free" property in the sense that parameter estimation of conditional mean models has no impact on its limit distribution. It has been documented to have high power against a wide variety of linear and nonlinear alternatives of practical importance.

In this paper, we have proposed a new diagnostic test for adequacy of linear and nonlinear time series models, using a new generalized spectral density approach. The test has the "nuisance parameter-free" property under a wider class of time series models than the BDS test, that includes but is not restricted to ARCH models. It is consistent against all pairwise serial dependencies, and has better asymptotic local power than the BDS test against local alternatives in testing many conditional mean models (but not in testing ARCH models). The generalized spectral smoothing allows the choice of a lag order via data-driven methods, which let data themselves speak for a proper lag order and give more robust power. A simulation experiment examines the finite sample performance of the proposed test and the tests of BDS, Box-Pierce-Ljung, and McLeod-Li-Mak. The generalized spectral density test has good power against a variety of stochastic and chaotic alternatives to the null models of conditional mean and conditional variance. It is a useful addition to the existing diagnostic toolkit for time series models, and can play a valuable role in evaluating adequacy of linear and nonlinear time series models.

REFERENCES

Barnett, W., A. R. Gallant, M.J. Hinich, J.A. Jungeilges, D. T. Kaplan, and M.J. Jensen (1997), "A Single-Blind Controlled Competition Among Tests for Nonlinearity and Chaos", *Journal of Econometrics*, 82, 157-192.

Bera, A.K. and M.L. Higgins (1997), "ARCH and Bilinearity as Competing Models for Nonlinear Dependence", *Journal of Business and Economic Statistics* 15, 43-50.

Box, G. and G. Jenkins (1970), *Time Series Analysis, Forecasting and Control.* Holden Day: San Francisco.

Box, G. and D. Pierce (1970), "Distribution of Residual Autocorrelations in Autoregressive Integrated Moving Average Time Series Models", *Journal of American Statistical Association* 65, 1509-1527.

Brillinger, D.R. and M. Rosenblatt (1967a,b), "Asymptotic Theory of Estimates of the kth Order Spectra" and "Computation and Interpretation of the kth Order Spectra", in *Spectral Analysis of Time Series*, B. Harris (ed.), pp. 153-188, 189-232. Wiley.

Brock, W., D. Hsieh and B. LeBaron (1991), *Nonlinear Dynamics, Chaos, and Instability: Statistical Theory and Economic Evidence.* MIT Press: Cambridge, MA.

Brock, W., D. Dechert, J. Scheinkman and B. LeBaron (1996), "A Test for Independence Based on the Correlation Dimension", *Econometric Reviews* 15, 197-235.

Brooks, C. and S.M. Heravi (1999), "The Effect of (Mis-Specified) GARCH Filters on the Finite Sample Distribution of the BDS Test", *Computational Economics* 13, 147-162.

Cameron, A.C. and P. K. Trivedi (1993), "Tests of Independence in Parametric Models with Applications and Illustrations", *Journal of Business and Economic Statistics,* 11, 29-43.

Chan, N. H. and L. T. Tran (1992), "Nonparametric Tests for Serial Dependence", *Journal of Time Series Analysis* 13, 102-113.

Delgado, M. (1996), "Testing Serial Independence Using the Sample Distribution Function", *Journal of Time Series Analysis* 17, 271-285.

Diebold, F. X. (1986), "Modeling the Persistence of Conditional Variances: A Comment", *Econometric Reviews* 5, 51-56.

Fama, E. F. and R. Roll (1968), "Some properties of Symmetric Stable Distributions", *Journal of American Statistical Associations* 63, 817-836.

Granger, C.W.J. (1983), "Forecasting White Noise", in *Applied Time Series Analysis of Eco-*

nomic Data, Proceedings of the Conference on Applied Time Series Analysis of Economic Data (October 1981), edited by A. Zellner, U.S. Government Printing Office.

Granger, C.W.J. (1998), "Overview of Nonlinear Time Series Specification in Economics", Discussion Paper, Department of Economics, University of California, San Diego.

Granger, C.W.J. and A.P. Andersen (1978), *An Introduction to Bilinear Time Series Models*, Gottingen: Vandenhoech und Ruprecht.

Granger, C.W.J. and T.-H. Lee (1999), "The Effect of Aggregation on Nonlinearity", *Econometric Reviews*, 18(3), 259-269.

Granger, C.W.J. and T. Teräsvirta (1993), *Modelling Nonlinear Economic Relationships.* Oxford University Press: New York.

Granger, C.W.J. and T. Teräsvirta (1999), "A Simple Nonlinear Time Series Model with Misleading Linear Properties", *Economics Letters*, 62, 161-165.

Harvey, A., E. Ruiz, and N. Shephard (1994), "Multivariate Stochastic Variance Models", *Review of Economic Studies*, 61, 247-264.

Hinich, M. (1982), "Testing for Gaussianity and Linearity of Stationary Time Series", *Journal of Time Series Analysis,* 3, 169-176.

Hong, Y. (1996), "Consistent Testing for Serial Correlation of Unknown Form", *Econometrica,* 64(4), 837-864.

Hong, Y. (1998), "Testing for Pairwise Independence via the Empirical Distribution Function", *Journal of Royal Statistical Society Series B,* 60, 429-453.

Hong, Y. (1999), "Hypothesis Testing in Time Series via the Empirical Characteristic Function: A Generalized Spectral Density Approach", *Journal of American Statistical Association*, 84, 1201-1220.

Hong, Y. (2000), "Generalized Spectral Tests for Serial Dependence", Forthcoming, *Journal of Royal Statistical Society, Series B,* 62, Part 3.

Lee, S. and B. Hansen (1994), "Asymptotic Theory for the GARCH(1,1) Quasi-Maximum Likelihood Estimator", *Econometric Theory* 10, 29-52.

Lee, T.-H., H. White and C.W.J. Granger (1993), "Testing for Neglected Nonlinearity in Time Series Models: A Comparison of Neural Network Methods and Alternative Tests", *Journal of Econometrics*, 56, 269-290.

Li, W.K. and T.K. Mak (1994), "On the Squared Residual Autocorrelations in Nonlinear Time Series with Conditional Heteroskedasticity", *Journal of Time Series Analysis*, 15(6), 627-636.

Ljung, G.M. and G.E.P. Box (1978), "A Measure of Lack of Fit in Time Series Models", *Biometrika* 65, 297-303.

Lumsdaine, R. (1996), "Consistency and Asymptotic Normality of the Quasi-Maximum Likelihood Estimator in IGARCH(1,1) and Covariance Stationary GARCH(1,1) Models", *Econometrica* 64, 575-596.

McLeod, A.I. and W.K. Li (1983), "Diagnostic Checking ARMA Time Series Models Using Squared-residual Autocorrelations", *Journal of Time Series Analysis*, 4, 269-273.

Pagan, A. and W. Schwert (1990), "Alternative Models for Conditional Stock Volatility", *Journal of Econometrics* 84, 205-231.

Paparoditis, E. (2000a), "Spectral Density Based Goodness-of-Fit Tests for Time Series Models", *Scandinavian Journal of Statistics,* 27(1), 143-176.

Paparoditis, E. (2000b), "On Some Power Properties of Goodness-of-Fit Tests in Time Series Analysis", University of Cyprus.

Pinkse, J. (1998), "Consistent Nonparametric Testing for Serial Independence", *Journal of Econometrics,* 84, 205-231.

Priestley, M.B. (1981), *Spectral Analysis and Time Series.* Academic press: London.

Priestley, M.B. (1988), *Non-Linear and Non-Stationary Time Series Analysis.* Academic Press: London.

Robinson, P. M. (1991), "Consistent Nonparametric Entropy-Based Testing", *Review of Economic Studies,* 58, 437-453.

Sakai, H. and Tokumaru, H. (1980), "Autocorrelations of a Certain Chaos", *IEEE Transactions on Acoustics, Speech, and Signal Processing, ASSP-28,* 588-590.

Skaug, H. J. and D. Tjøstheim. (1993a), "Nonparametric Test of Serial Independence based on the Empirical Distribution Function", *Biometrika,* 80, 591-602.

Skaug, H. J. and D. Tjøstheim. (1993b), "Nonparametric Tests of Serial Independence", In *Developments in Time Series Analysis.* The Priestley Birthday Volume, T. Subba Rao (ed.), pp. 207-229, Chapman and Hall: London.

Skaug, H. J. and D. Tjøstheim. (1996), "Measures of Distance Between Densities with Application to Testing for Serial Independence", In *Time Series Analysis in Memory of E. J. Hannan*, P. Robinson and M. Rosenblatt (eds.), pp. 363-377, Springer: New York.

Subba Rao, T. and M. Gabr (1980), "A Test for Linearity of Stationary Time Series", *Journal of Time Series Analysis,* 1, 145-158.

Subba Rao, T. and M. Gabr (1984), *An Introduction to Bispectral Analysis and Bilinear Time Series Models.* Lecture Notes in Statistics 24. Springer: New York.

Teräsvirta, T., D. Tjøstheim and C.W.J. Granger (1994), "Aspects of Modelling Nonlinear Time Series", *Handbook of Econometrics,* Vol. IV, Chapter 48, pp. 2917-2957, R.F. Engle and D.L. McFadden (ed.), North Holland, Elsevier Science.

Terdik. G. (1999), *Bilinear Stochastic Models and Related Problems of Nonlinear Time Series Analysis: A Frequency Domain Approach*, Lecture Notes in Statistics, No. 142, Springer-verlag: New York.

Tjøstheim, D. (1996), "Measures and Tests of Independence: A Survey", *Statistics*, 28, 249-284.

Tong, H. (1990), *Nonlinear Time Series: A Dynamic System Approach.* Clarendon Press: Oxford.

Weiss, A.A. (1986), "ARCH and Bilinear Time Series Models: Comparison and Combination", *Journal of Business and Economic Statistics,* 4, 59-70.

MATHEMATICAL APPENDIX

Throughout the appendix, we let $M(p)$ be defined in the same way as $\hat{M}(p)$ in (2.9), with the unobservable sample $\{e_t\}_{t=1}^n$ replacing the standardized residual sample $\{\hat{e}_t\}_{t=1}^n$. Also, C denotes a generic bounded constant that may differ from place to place.

Proof of Theorem 1: It suffices to show Theorems A.1-A.2 below. Theorem A.1 implies that the use of $\{\hat{e}_t\}_{t=1}^n$ rather than $\{e_t\}_{t=1}^n$ has no impact on the limit distribution of $\hat{M}(p)$.

Theorem A.1: *Under the conditions of Theorem 1,* $\hat{M}(p) - M(p) \xrightarrow{p} 0$.

Theorem A.2: *Under the conditions of Theorem 1,* $M(p) \xrightarrow{d} N(0,1)$.

Proof of Theorem A.1: Noting that $e_t(\theta) \equiv [Y_t - g(I_{t-1},\theta)]/h(I_{t-1},\theta)$, where I_t is the unobservable information set from period t to the infinite past, we write

$$\begin{aligned}
\hat{e}_t &\equiv \frac{Y_t - g(\hat{I}_{t-1},\hat{\theta})}{h(\hat{I}_{t-1},\hat{\theta})} \\
&= \frac{Y_t - g(I_{t-1},\hat{\theta})}{h(\hat{I}_{t-1},\hat{\theta})} + \frac{g(I_{t-1},\hat{\theta}) - g(\hat{I}_{t-1},\hat{\theta})}{h(\hat{I}_{t-1},\hat{\theta})} \\
&= e_t(\hat{\theta}) + e_t(\hat{\theta})\frac{h(I_{t-1},\hat{\theta}) - h(\hat{I}_{t-1},\hat{\theta})}{h(\hat{I}_{t-1},\hat{\theta})} + \frac{g(I_{t-1},\hat{\theta}) - g(\hat{I}_{t-1},\hat{\theta})}{h(\hat{I}_{t-1},\hat{\theta})}. \qquad \text{(A1)}
\end{aligned}$$

By the mean value theorem,

$$e_t(\hat{\theta}) = \frac{Y_t - g(I_{t-1},\theta_0)}{h(I_{t-1},\theta_0)} + \xi_t(\bar{\theta})'(\hat{\theta} - \theta_0) = e_t(\theta_0) + \xi_t(\bar{\theta})'(\hat{\theta} - \theta_0) \qquad \text{(A2)}$$

for some $\bar{\theta}$ between $\hat{\theta}$ and θ_0, where

$$\xi_t(\theta) \equiv \frac{\partial}{\partial\theta}e_t(\theta) = e_t(\theta)h^{-1}(I_{t-1},\theta)\frac{\partial}{\partial\theta}h(I_{t-1},\theta) - h^{-1}(I_{t-1},\theta)\frac{\partial}{\partial\theta}g(I_{t-1},\theta).$$

It follows from (A1)-(A2) and the Cauchy-Schwarz inequality that

$$\begin{aligned}
\sum_{t=1}^n (\hat{e}_t - \tilde{e}_t)^2 &\leq 2\sum_{t=1}^n e_t^2(\hat{\theta})\left|\frac{h(I_{t-1},\hat{\theta}) - h(\hat{I}_{t-1},\hat{\theta})}{h(\hat{I}_{t-1},\hat{\theta})}\right|^2 + 2\sum_{t=1}^n\left|\frac{g(I_{t-1},\hat{\theta}) - g(\hat{I}_{t-1},\hat{\theta})}{h(\hat{I}_{t-1},\hat{\theta})}\right|^2 \\
&\leq 4\sum_{t=1}^n e_t^2(\theta_0)\sup_{\theta\in\Theta_0}\left[\frac{h(I_{t-1},\theta) - h(\hat{I}_{t-1},\theta)}{h(\hat{I}_{t-1},\theta)}\right]^2 \\
&\quad +4\left\|\hat{\theta} - \theta_0\right\|^2\left[\sum_{t=1}^n\sup_{\theta\in\Theta_0}\|\xi_t(\theta)\|^4\right]^{\frac{1}{2}}\left\{\sum_{t=1}^n\sup_{\theta\in\Theta_0}\left[\frac{h(I_{t-1},\theta) - h(\hat{I}_{t-1},\theta)}{h(\hat{I}_{t-1},\theta)}\right]^4\right\}^{\frac{1}{2}} \\
&\quad +2\sum_{t=1}^n\sup_{\theta\in\Theta_0}\left[\frac{g(I_{t-1},\theta) - g(\hat{I}_{t-1},\theta)}{h(\hat{I}_{t-1},\theta)}\right]^2 \\
&= O_P(1), \qquad \text{(A3)}
\end{aligned}$$

given Assumptions A.1-A.4, where we made use of $E\sup_{\theta\in\Theta_0}\|\xi_t(\theta)\|^4 \le C$ given Assumption A.3. Here, the first term in the second inequality is $O_P(1)$ by Markov's inequality, independence between $e_t(\theta_0)=\varepsilon_t$ and $(I_{t-1},\hat{I}_{t-1})$, and Assumption A.4. On the other hand, by (A2) and Assumptions A.2-A.4, we have

$$\sum_{t=1}^{n}\left[e_t(\hat{\theta})-e_t(\theta_0)\right]^2 \le n\left\|\hat{\theta}-\theta_0\right\|^2\left[n^{-1}\sum_{t=1}^{n}\sup_{\theta\in\Theta_0}\|\xi_t(\theta)\|^2\right]=O_P(1). \tag{A4}$$

Both (A3) and (A4) imply

$$\sum_{t=1}^{n}\left[\hat{e}_t-e_t(\theta_0)\right]^2=O_P(1). \tag{A5}$$

Put $n_j=n-|j|$. Observe that $p\to\infty, p/n\to 0, p^{-1}\sum_{j=1}^{n-1}k^r(j/p)\to\int_0^\infty k^r(z)dz$ for $r=2,4$ given Assumption A.3. To show $\hat{M}(p)-M(p)\to^p 0$, it suffices to show that

$$p^{-\frac{1}{2}}\int\sum_{j=1}^{n-1}k^2(j/p)n_j\left[|\hat{\sigma}_j(u,v)|^2-|\tilde{\sigma}_j(u,v)|^2\right]dW(u)dW(v)\xrightarrow{p}0, \tag{A6}$$

$\hat{C}_0-\tilde{C}_0=O_P(n^{-\frac{1}{2}})$ and $\hat{D}_0-\tilde{D}_0\xrightarrow{p}0$, where $\tilde{C}_0$ and $\tilde{D}_0$ are defined in the same way as $\hat{C}_0$ and $\hat{D}_0$ in (2.10) and (2.11), with $\{e_t\}_{t=1}^n$ replacing $\{\hat{e}_t\}_{t=1}^n$. For space, we focus on the proof of (A6); the proofs for $\hat{C}_0-\tilde{C}_0=O_P(n^{-\frac{1}{2}})$ and $\hat{D}_0-\tilde{D}_0\to^p 0$ are straightforward. We note that here it is necessary to obtain the convergence rate for $\hat{C}_0-\tilde{C}_0$ in order to ensure that replacing $\hat{C}_0$ with $\tilde{C}_0$ has asymptotically negligible impact given $p/n\to 0$.

To show (A6), we first decompose

$$\int\sum_{j=1}^{n-1}k^2(j/p)n_j\left[|\hat{\sigma}_j(u,v)|^2-|\tilde{\sigma}_j(u,v)|^2\right]dW(u)dW(v)=\hat{A}_1+2\operatorname{Re}(\hat{A}_2), \tag{A7}$$

where

$$\begin{aligned}\hat{A}_1 &= \int\sum_{j=1}^{n-1}k^2(j/p)n_j\left|\hat{\sigma}_j(u,v)-\tilde{\sigma}_j(u,v)\right|^2dW(u)dW(v),\\ \hat{A}_2 &= \int\sum_{j=1}^{n-1}k^2(j/p)n_j\left[\hat{\sigma}_j(u,v)-\tilde{\sigma}_j(u,v)\right]\hat{\sigma}_j(u,v)^*dW(u)dW(v),\end{aligned}$$

where $\operatorname{Re}(\hat{A}_2)$ is the real part of $\hat{A}_2$ and $\hat{\sigma}_j(u,v)^*$ is the complex conjugate of $\hat{\sigma}_j(u,v)$. Then, (A6) follows from Propositions A.1-A.2 below, and $p\to\infty$ as $n\to\infty$.

Proposition A.1: *Under the conditions of Theorem 1,* $\hat{A}_1=O_P(1)$.

Proposition A.2: *Under the conditions of Theorem 1,* $p^{-\frac{1}{2}}\hat{A}_2\to^p 0$.

Proof of Proposition A.1: Put $\hat{\delta}_t(u) = e^{iu\hat{e}_t} - e^{iue_t}$ and $\varphi_t(u) = e^{iue_t} - \varphi(u)$, where, as before, $\varphi(u) = E(e^{iue_t})$. Let $\tilde{\sigma}_j(u,v)$ be defined in the same way as $\hat{\sigma}_j(u,v)$ in (2.5), with $\{e_t\}_{t=1}^n$ replacing $\{\hat{e}_t\}_{t=1}^n$. Then straightforward algebra yields

$$\begin{aligned}
&\hat{\sigma}_j(u,v) - \tilde{\sigma}_j(u,v)\\
&= n_j^{-1}\sum_{t=j+1}^n \hat{\delta}_t(u)\hat{\delta}_{t-j}(v) - \left[n_j^{-1}\sum_{t=j+1}^n \hat{\delta}_t(u)\right]\left[n_j^{-1}\sum_{t=j+1}^n \hat{\delta}_{t-j}(v)\right]\\
&\quad + n_j^{-1}\sum_{t=j+1}^n \varphi_t(u)\hat{\delta}_{t-j}(v) - \left[n_j^{-1}\sum_{t=j+1}^n \varphi_t(u)\right]\left[n_j^{-1}\sum_{t=j+1}^n \hat{\delta}_{t-j}(v)\right]\\
&\quad + n_j^{-1}\sum_{t=j+1}^n \hat{\delta}_t(u)\varphi_{t-j}(v) - \left[n_j^{-1}\sum_{t=j+1}^n \hat{\delta}_t(u)\right]\left[n_j^{-1}\sum_{t=j+1}^n \varphi_{t-j}(v)\right]\\
&= \hat{B}_{1j}(u,v) - \hat{B}_{2j}(u,v) + \hat{B}_{3j}(u,v) - \hat{B}_{4j}(u,v) + \hat{B}_{5j}(u,v) - \hat{B}_{6j}(u,v), \text{ say.}
\end{aligned}$$

It follows from (A10) and the C_r-inequality that

$$\hat{A}_1 \le 2^5 \sum_{a=1}^6 \int \sum_{j=1}^{n-1} k^2(j/p) n_j |\hat{B}_{aj}(u,v)|^2 dW(u)dW(v).$$

Proposition A.1 follows from Lemmas A.1-A.6 below, and $p/n \to 0$. We shall show these lemmas under the conditions of Theorem 1.

Lemma A.1: $\int \sum_{j=1}^{n-1} k^2(j/p) n_j |\hat{B}_{1j}(u,v)|^2 dW(u)dW(v) = O_P(p/n)$.

Lemma A.2: $\int \sum_{j=1}^{n-1} k^2(j/p) n_j |\hat{B}_{2j}(u,v)|^2 dW(u)dW(v) = O_P(p/n)$.

Lemma A.3: $\int \sum_{j=1}^{n-1} k^2(j/p) n_j |\hat{B}_{3j}(u,v)|^2 dW(u)dW(v) = O_P(p/n)$.

Lemma A.4: $\int \sum_{j=1}^{n-1} k^2(j/p) n_j |\hat{B}_{4j}(u,v)|^2 dW(u)dW(v) = O_P(p/n)$.

Lemma A.5: $\int \sum_{j=1}^{n-1} k^2(j/p) n_j |\hat{B}_{5j}(u,v)|^2 dW(u)dW(v) = O_P(1)$.

Lemma A.6: $\int \sum_{j=1}^{n-1} k^2(j/p) n_j |\hat{B}_{6j}(u,v)|^2 dW(u)dW(v) = O_P(p/n)$.

Proof of Lemma A.1: By the Cauchy-Schwarz inequality and inequality $|e^{iz_1} - e^{iz_2}| \le |z_1 - z_2|$ for any complex-valued variables z_1 and z_2, we have

$$|\hat{B}_{1j}(u,v)|^2 \le \left[n_j^{-1}\sum_{t=1}^n |\hat{\delta}_t(u)|^2\right]\left[n_j^{-1}\sum_{t=1}^n |\hat{\delta}_t(v)|^2\right] \le (uv)^2 \left[n_j^{-1}\sum_{t=1}^n (\hat{e}_t - e_t)^2\right]^2. \qquad \text{(A11)}$$

It follows from (A11), (A5), and Assumptions A.5-A.6 that

$$\begin{aligned}\int \sum_{j=1}^{n-1} k^2(j/p) n_j |\hat{B}_{1j}(u,v)|^2 dW &\leq \left[\sum_{j=1}^{n-1} k^2(j/p) n_j^{-1}\right] \left[\sum_{t=1}^{n} (\hat{e}_t - e_t)^2\right]^2 \left[\int u^2 dW(u)\right]^2 \\ &= O_P(p/n).\end{aligned}$$

where we made use of the fact that

$$\sum_{j=1}^{n-1} k^2(j/p) n_j^{-1} = O(p/n) \tag{A12}$$

given Assumption A.3 and $p = cn^\lambda$ for $\lambda \in (0,1)$, as shown in Hong (1999, (A.15), p.1213). ■

Proof of Lemma A.2: Similar to the proof of Lemma A.1. ■

Proof of Lemma A.3: Using inequality $|e^{iz} - 1 - iz| \leq |z|^2$ for any complex-valued variable z, we have

$$|\hat{\delta}_t(u) - iu(\hat{e}_t - e_t)e^{iue_t}| \leq u^2(\hat{e}_t - e_t)^2. \tag{A13}$$

Also, a second order Taylor expansion yields

$$e_t(\hat{\theta}) = e_t(\theta_0) + \xi_t(\theta_0)'(\hat{\theta} - \theta_0) + \frac{1}{2}(\hat{\theta} - \theta_0)' \frac{\partial}{\partial \theta}\xi_t(\bar{\theta})(\hat{\theta} - \theta_0) \tag{A14}$$

for some $\bar{\theta}$ between $\hat{\theta}$ and θ_0, where $\xi_t(\theta)$ is as in (A2). Both (A13) and (A14) imply

$$|\hat{\delta}_t(u) - iu\tilde{\xi}_t(u)(\hat{\theta} - \theta_0)| \leq u^2 [\hat{e}_t - e_t(\theta_0)]^2 + |u| \left|\hat{e}_t - e_t(\hat{\theta})\right| + |u| \left\|\hat{\theta} - \theta_0\right\|^2 \sup_{\theta \in \Theta_0} \left\|\frac{\partial}{\partial \theta}\xi_t(\theta)\right\|, \tag{A15}$$

where $\tilde{\xi}_t(u) = \xi_t(\theta_0)e^{iue_t}$. Therefore, from the definition of $\hat{B}_{3j}(u,v)$ and $|\varphi_t(u)| \leq C$, we obtain

$$\begin{aligned}n_j \left|\hat{B}_{3j}(u,v)\right| \leq\ & |v| \left\|\hat{\theta} - \theta_0\right\| \left|\sum_{t=j+1}^{n} \varphi_t(u)\tilde{\xi}_{t-j}(v)\right| + v^2 \sum_{t=1}^{n} (\hat{e}_t - e_t)^2 \\ & + |v| \sum_{t=1}^{n} |\hat{e}_t - \tilde{e}_t| + |v| \left\|\hat{\theta} - \theta_0\right\|^2 \sum_{t=1}^{n} \sup_{\theta \in \Theta_0} \left\|\frac{\partial}{\partial \theta}\xi_t(\theta)\right\|. \end{aligned} \tag{A16}$$

It follows from (A16), the C_r-inequality, (A3), (A5), (A12), and Assumptions A.1-A.6 that

$$\int \sum_{j=1}^{n-1} k^2(j/p) n_j |\hat{B}_{3j}(u,v)|^2 dW$$

$$\leq 8 \left\|\hat{\theta} - \theta_0\right\|^2 \sum_{j=1}^{n-1} k^2(j/p) n_j^{-1} \int \left|\sum_{t=j+1}^{n} \varphi_t(u)\tilde{\xi}_{t-j}(v)\right|^2 v^2 dW(u) dW(v)$$

$$
\begin{aligned}
&+8\left[\sum_{t=1}^{n}(\hat{e}_t - e_t(\theta_0))^2\right]^2\left[\sum_{j=1}^{n-1}k^2(j/p)n_j^{-1}\right]\int v^4 dW(u)dW(v) \\
&+8\left(\sum_{t=1}^{n}|\hat{e}_t - e_t(\hat{\theta})|\right)^2\left[\sum_{j=1}^{n-1}k^2(j/p)n_j^{-1}\right]\int v^2 dW(u)dW(v) \\
&+8\left\|\sqrt{n}(\hat{\theta}-\theta_0)\right\|^4\left[n^{-1}\sum_{t=1}^{n}\sup_{\theta\in\Theta_0}\left\|\frac{\partial}{\partial\theta}\xi_t(\theta)\right\|\right]^2\left[\sum_{j=1}^{n-1}k^2(j/p)n_j^{-1}\right]\int v^2 dW(u)dW(v) \\
= \;& O_P(p/n)
\end{aligned}
$$

where we made use of the fact that $E|\sum_{t=j+1}^{n}\varphi_t(u)\tilde{\xi}_{t-j}(v)|^2 \leq Cn_j$ since $\varphi_t(u)$ is independent of $\tilde{\xi}_{t-j}(v)$ for $j>0$ under the i.i.d. hypothesis of $\{e_t\}_{t=1}^{n}$. We also made use of the fact that from (A.1), the Cauchy-Schwarz inequality and Assumption A.3

$$
\begin{aligned}
\sum_{t=1}^{n}|\hat{e}_t - e_t(\hat{\theta})| &\leq \sum_{t=1}^{n}|e_t(\hat{\theta})|\left|\frac{h(I_{t-1},\hat{\theta})-h(\hat{I}_{t-1},\hat{\theta})}{h(\hat{I}_{t-1},\hat{\theta})}\right| + \sum_{t=1}^{n}\left|\frac{g(I_{t-1},\hat{\theta})-g(\hat{I}_{t-1},\hat{\theta})}{h(\hat{I}_{t-1},\hat{\theta})}\right| \\
&\leq \sum_{t=1}^{n}|e_t(\theta_0)|\sup_{\theta\in\Theta_0}\left|\frac{h(I_{t-1},\theta)-h(\hat{I}_{t-1},\theta)}{h(\hat{I}_{t-1},\theta)}\right| \\
&\quad+\left\|\hat{\theta}-\theta_0\right\|\left[\sum_{t=1}^{n}\sup_{\theta\in\Theta_0}\|\xi_t(\theta)\|^2\right]^{\frac{1}{2}}\left\{\sum_{t=1}^{n}\sup_{\theta\in\Theta_0}\left[\frac{g(I_{t-1},\theta)-g(\hat{I}_{t-1},\theta)}{h(\hat{I}_{t-1},\theta)}\right]^2\right\}^{\frac{1}{2}} \\
&\quad+\sum_{t=1}^{n}\sup_{\theta\in\Theta_0}\left|\frac{g(I_{t-1},\theta)-g(\hat{I}_{t-1},\theta)}{h(\hat{I}_{t-1},\theta)}\right| \\
&= O_P(1). \qquad\qquad (A17)
\end{aligned}
$$

Here, the first term in the second inequality is $O_P(1)$ by Markov's inequality, independence between $e_t(\theta_0)=\varepsilon_t$ and $(I_{t-1},\hat{I}_{t-1})$, and Assumption A.3. ■

Proof of Lemma A.4: By the Cauchy-Schwarz inequality, we have

$$
|\hat{B}_{4j}(u,v)|^2 \leq \left|n_j^{-1}\sum_{t=j+1}^{n}\varphi_t(u)\right|^2\left[n_j^{-1}\sum_{t=1}^{n}|\hat{\delta}_t(v)|^2\right]. \qquad (A18)
$$

It follows from (A18), the Cauchy-Schwarz inequality, and $|\hat{\delta}_t(v)| \leq |v\hat{e}_t - ve_t|$ that

$$
\begin{aligned}
\int\sum_{j=1}^{n-1}k^2(j/p)n_j|\hat{B}_{4j}(u,v)|^2 dW &\leq \sum_{j=1}^{n-1}k^2(j/p)\int\left|n_j^{-1}\sum_{t=j+1}^{n}\varphi_t(u)\right|^2 v^2 dW(u)dW(v) \\
&\quad\times\sum_{t=1}^{n}(\hat{e}_t-e_t)^2 \\
&= O_P(p/n)
\end{aligned}
$$

given (A5), (A12), and $E\left|\sum_{t=j+1}^{n}\varphi_{t-j}(u)\right|^2 \leq Cn_j$ under the i.i.d. hypothesis of $\{e_t\}$. ■

Proof of Lemma A.5: Using (A15) and the C_r-inequality, we have

$$\begin{aligned}
&\int \sum_{j=1}^{n-1} k^2(j/p)n_j|\hat{B}_{5j}(u,v)|^2 dW(u)dW(v) \\
\leq\ & 8\left\|\hat{\theta}-\theta_0\right\|^2 \sum_{j=1}^{n-1} k^2(j/p)n_j^{-1}\int\left|\sum_{t=j+1}^{n}\tilde{\xi}_t(u)\varphi_{t-j}(v)\right|^2 u^2 dW(u)dW(v) \\
&+8\left[\sum_{t=1}^{n}(\hat{e}_t-e_t(\theta_0))^2\right]^2\left[\sum_{j=1}^{n-1}k^2(j/p)n_j^{-1}\right]\int u^4 dW(u)dW(v) \\
&+8\left(\sum_{t=1}^{n}|\hat{e}_t-e_t(\hat{\theta})|\right)^2\left[\sum_{j=1}^{n-1}k^2(j/p)n_j^{-1}\right]\int u^2 dW(u)dW(v) \\
&+8\left\|\sqrt{n}(\hat{\theta}-\theta_0)\right\|^4\left[n^{-1}\sum_{t=1}^{n}\sup_{\theta\in\Theta_0}\left\|\frac{\partial}{\partial\theta}\xi_t(\theta)\right\|\right]^2\left[\sum_{j=1}^{n-1}k^2(j/p)n_j^{-1}\right]\int u^2 dW(u)dW(v) \\
=\ & O_P(1), \qquad \text{(A19)}
\end{aligned}$$

where the last three terms are $O_P(p/n)$ given (A5), (A17), (A12), and Assumptions A.1-A.6; and the first term is $O_P(1)$, as is shown below:

Put $\eta_j(u,v)=E[\tilde{\xi}_t(u)\varphi_{t-j}(v)]$. Note that $\tilde{\xi}_t(u)$ is a function of I_{t-1} and thus is not independent of $\varphi_{t-j}(v)$. By the standard α-mixing inequality, we have

$$|\eta_j(u,v)| \leq \left[E|\tilde{\xi}_t(u)|^{2\nu}\right]^{\frac{1}{2\nu}}\left[E|\varphi_{t-j}(v)|^{2\nu}\right]^{\frac{1}{2\nu}}\alpha(j)^{(\nu-1)/\nu} \leq C\alpha(j)^{(\nu-1)/\nu}. \qquad \text{(A20)}$$

Moreover, given Assumptions A.1-A.2, we have

$$E\left|n_j^{-1}\sum_{t=j+1}^{n}\left[\tilde{\xi}_t(u)\varphi_{t-j}(v)-\eta_j(u,v)\right]\right|^2 \leq Cn_j^{-1}, \qquad \text{(A21)}$$

using a reasoning analogous to (A.7)-(A.10) in the proof of Theorem 1 of Hong (1999, pp.1212-1213). Consequently, from (A20) and (A21), we have

$$\begin{aligned}
\sum_{j=1}^{n-1}k^2(j/p)E\int\left|n_j^{-1}\sum_{t=j+1}^{n}\tilde{\xi}_t(u)\varphi_{t-j}(v)\right|^2 u^2 dW(u)dW(v) &\leq C\sum_{j=1}^{n-1}\int|\eta_j(u,v)|^2 v^2 dW(u)dW(v) \\
&\quad +C\sum_{j=1}^{n-1}k^2(j/p)n_j^{-1} \\
&= O(1)+O(p/n) \\
&= O(1).
\end{aligned}$$

It follows that the first term in (A19) is $O_P(1)$ by Markov's inequality. ■

Proof of Lemma A.6: The proof is analogous to that of Lemma A.4. ■

Proof of Proposition A.2: Given (A10), we have

$$\left|[\hat{\sigma}_j(u,v) - \tilde{\sigma}_j(u,v)]\, \tilde{\sigma}_j(u,v)^*\right| \leq \sum_{a=1}^{6} \left|\hat{B}_{aj}(u,v)\right| \left|\tilde{\sigma}_j(u,v)\right|, \tag{A22}$$

where the $\hat{B}_{aj}(u,v)$ are defined in (A10). For $a = 1,2,3,4$ and 6, we have, by the Cauchy-Schwarz inequality,

$$\begin{aligned}
&\sum_{j=1}^{n-1} k^2(j/p) n_j \int \left|\hat{B}_{aj}(u,v)\right| \left|\tilde{\sigma}_j(u,v)\right| dW(u)dW(v) \\
\leq\ & \left[\sum_{j=1}^{n-1} k^2(j/p) n_j \int |\hat{B}_{aj}(u,v)|^2 dW(u)dW(v)\right]^{\frac{1}{2}} \left[\sum_{j=1}^{n-1} k^2(j/p) n_j \int \left|\tilde{\sigma}_j(u,v)\right|^2 dW(u)dW(v)\right]^{\frac{1}{2}} \\
=\ & O_P(p^{\frac{1}{2}}/n^{\frac{1}{2}}) O_P(p^{\frac{1}{2}}) \\
=\ & o_P(p^{\frac{1}{2}})
\end{aligned}$$

given Lemmas A.1-A.4 and A.6, and $p/n \to 0$, where

$$p^{-1} \sum_{j=1}^{n-1} k^2(j/p) n_j \int \left|\tilde{\sigma}_j(u,v)\right|^2 dW = O_P(1)$$

as follows from Markov's inequality, the i.i.d. hypothesis of $\{e_t\}$, and (A12).

It remains to consider $a = 5$. Using (A15), the triangular inequality, (A5) and (A17), we have

$$\begin{aligned}
&\sum_{j=1}^{n-1} k^2(j/p) n_j \left|\hat{B}_{5j}(u,v)\right| \left|\tilde{\sigma}_j(u,v)\right| \\
\leq\ & \left\|\hat{\theta} - \theta_0\right\| \sum_{j=1}^{n-1} k^2(j/p) n_j \int \left| n_j^{-1} \sum_{t=j+1}^{n} \tilde{\xi}_t(u) \varphi_{t-j}(v) \right| \left|\tilde{\sigma}_j(u,v)\right| |u| dW(u)dW(v) \\
&+ \left\{\sum_{t=1}^{n} [\hat{e}_t - e_t(\theta_0)]^2\right\} \sum_{j=1}^{n-1} k^2(j/p) n_j \int \left|\tilde{\sigma}_j(u,v)\right| u^2 dW(u)dW(v) \\
&+ \left[\sum_{t=1}^{n} \left|\hat{e}_t - e_t(\hat{\theta})\right|\right] \sum_{j=1}^{n-1} k^2(j/p) n_j \int \left|\tilde{\sigma}_j(u,v)\right| |u| dW(u)dW(v) \\
&+ \left\|\hat{\theta} - \theta_0\right\|^2 \left[\sum_{t=1}^{n} \sup_{\theta \in \Theta_0} \left\|\frac{\partial}{\partial \theta} \xi_t(\theta)\right\|\right] \sum_{j=1}^{n-1} k^2(j/p) n_j \int \left|\tilde{\sigma}_j(u,v)\right| |u| dW(u)dW(v) \\
=\ & O_P(1 + p/n^{\frac{1}{2}}) + O_P(p/n^{\frac{1}{2}}) + O_P(p/n^{\frac{1}{2}}) + O_P(p/n^{\frac{1}{2}}) \\
=\ & o_P(p^{\frac{1}{2}})
\end{aligned}$$

given $p \to \infty$, $p/n \to 0$, Assumptions A.1-A.6, where we have made use of the fact that $n_j E|\tilde{\sigma}_j(u,v)|^2 \le C$ under the i.i.d. hypothesis of $\{e_t\}$. Note that here for the first term in the first inequality, we have made use of the fact that

$$\begin{aligned} E\left[\left|n_j^{-1}\sum_{t=j+1}^{n}\tilde{\xi}_t(u)\varphi_{t-j}(v)\right||\tilde{\sigma}_j(u,v)|\right] &\le \left[E\left|n_j^{-1}\sum_{t=j+1}^{n}\tilde{\xi}_t(u)\varphi_{t-j}(v)\right|^2\right]^{\frac{1}{2}}\left[E|\tilde{\sigma}_j(u,v)|^2\right]^{\frac{1}{2}} \\ &\le C\left[\alpha(j)^{\frac{\nu-1}{\nu}}+n_j^{-\frac{1}{2}}\right]n_j^{-\frac{1}{2}} \end{aligned}$$

given (A20) and (A21), and consequently,

$$\begin{aligned} & n^{-\frac{1}{2}}\sum_{j=1}^{n-1}k^2(j/p)n_j E\int\left[\left|n_j^{-1}\sum_{t=j+1}^{n}\tilde{\xi}_t(u)\varphi_{t-j}(v)\right||\tilde{\sigma}_j(u,v)|\right]|u|dW(u)dW(v) \\ &\le C\sum_{j=1}^{n-1}\alpha(j)^{\frac{\nu-1}{\nu}}+Cn^{-\frac{1}{2}}\sum_{j=1}^{n-1}k^2(j/p) \\ &= O(1)+O(p/n^{\frac{1}{2}}) \end{aligned}$$

given $\sum_{j=1}^{\infty}\alpha(j)^{\frac{\nu-1}{\nu}} < \infty$, $|k(\cdot)| \le 1$ and (A12). ■

Proof of Theorem A.2: See Hong (1999, Proof of Theorem 3, for the case $(m,l) = (0,0)$). ■

Proof of Theorem 2: The proof of Theorem 2 consists of the proofs of Theorems A.3 and A.4 below.

Theorem A.3: *Under the conditions of Theorem 1,* $(p^{\frac{1}{2}}/n)[\hat{M}(p) - M(p)] \to^p 0$.

Theorem A.4: *Under the conditions of Theorem 1,*

$$(p^{\frac{1}{2}}/n)M(p) \xrightarrow{p} (\pi/2)\int\int_{-\pi}^{\pi}|f(\omega,u,v)-f_0(\omega,u,v)|^2 d\omega dW(u)dW(v).$$

Proof of Theorem A.3: Given that $p \to \infty, p/n \to 0$ and $p^{-1}\sum_{j=1}^{n-1}k^r(j/p) \to \int_0^\infty k^r(z)dz$, it suffices to show that

$$n^{-1}\int\sum_{j=1}^{n-1}k^2(j/p)n_j\left[|\hat{\sigma}_j(u,v)|^2-|\tilde{\sigma}_j(u,v)|^2\right]dW(u)dW(v) \xrightarrow{p} 0, \qquad \text{(A23)}$$

$\hat{C}_0 - \tilde{C}_0 = O_P(1)$, and $\hat{D}_0 - \tilde{D}_0 \to^p 0$, where $\tilde{C}_0$ and $\tilde{D}_0$ are defined in the same way as $\hat{C}_0$ and $\hat{D}_0$ in (2.10) and (2.11), with $\{e_t\}_{t=1}^n$ replacing $\{\hat{e}_t\}_{t=1}^n$. Since the proofs for $\hat{C}_0 - \tilde{C}_0 = O_P(1)$ and

$\hat{D}_0 - \tilde{D}_0 \to^p 0$ are straightforward, we focus on the proof of (A23). From (A7), the Cauchy-Schwarz inequality, and the fact that

$$n^{-1} \int \sum_{j=1}^{n-1} k^2(j/p) n_j |\tilde{\sigma}_j(u,v)|^2 dW(u) dW(v) = O_P(1)$$

as is implied by Theorem A.4 (the proof of Theorem A.4 does not depend on Theorem A.3), it suffices to show that

$$n^{-1} \hat{A}_1 \xrightarrow{p} 0, \tag{A24}$$

where $\hat{A}_1$ is defined in (A8). Given (A10), we shall show that

$$n^{-1} \int \sum_{j=1}^{n-1} k^2(j/p) n_j |\hat{B}_{aj}(u,v)|^2 dW(u) dW(v) \xrightarrow{p} 0, \qquad a = 1, 2, ..., 6.$$

We first consider $a = 1$. By the Cauchy-Schwarz inequality and inequality $|\hat{\delta}_t(u)| \leq |u\hat{e}_t - ue_t|$, we have

$$|\hat{B}_{1j}(u,v)|^2 \leq \left[n_j^{-1} \sum_{t=j+1}^{n} |\hat{\delta}_t(u)|^2 \right] \left[n_j^{-1} \sum_{t=j+1}^{n} |\hat{\delta}_t(v)|^2 \right] \leq n_j^{-2} (uv)^2 \left[\sum_{t=1}^{n} (\hat{e}_t - e_t)^2 \right]^2, \tag{A25}$$

where

$$n^{-\frac{1}{2}} \sum_{t=1}^{n} (\hat{e}_t - e_t)^2 = O_P(1) \tag{A26}$$

as can be shown in a way similar to that for (A5), given the condition that $E(e_t^4) \leq C$. Note that compared to (A5), a factor of $n^{-\frac{1}{2}}$ arises here because we no longer have independence between e_t and $\{I_{t-1}, \hat{I}_{t-1}\}$, and thus have to use the Cauchy-Schwarz inequality. It follows from (A25)-(A26), (A5) and (A12) that

$$\begin{aligned} n^{-1} \int \sum_{j=1}^{n-1} k^2(j/p) n_j |\hat{B}_{1j}(u,v)|^2 dW(u) dW(v) &\leq \left[n^{-\frac{1}{2}} \sum_{t=1}^{n} (\hat{e}_t - e_t)^2 \right]^2 \sum_{j=1}^{n-1} k^2(j/p) n_j^{-1} \left[\int u^2 dW(u) \right]^2 \\ &= O_P(p/n). \end{aligned}$$

The proof for $a = 2$ is similar to that for $a = 1$, by noting that $|n_j^{-1} \sum_{t=j+1}^{n} \hat{\delta}_t(u)|^2 \leq n_j^{-1} \sum_{t=j+1}^{n} |\hat{\delta}_t(u)|^2$.

Next, we consider $a = 3$. By the Cauchy-Schwarz inequality and $|\varphi_t(u)| \leq C$, we have

$$\begin{aligned} |B_{3j}(u,v)|^2 &\leq \left[n_j^{-1} \sum_{t=j+1}^{n} |\varphi_t(u)|^2 \right] \left[n_j^{-1} \sum_{t=j+1}^{n} |\hat{\delta}_{t-j}(v)|^2 \right] \\ &\leq v^2 n_j^{-1} \sum_{t=1}^{n} (\hat{e}_t - e_t)^2. \end{aligned} \tag{A27}$$

It follows that

$$n^{-1}\int\sum_{j=1}^{n-1}k^2(j/p)n_j|\hat{B}_{3j}(u,v)|^2dW(u)dW(v) \le n^{-1}\sum_{t=1}^{n}(\hat{e}_t-e_t)^2\sum_{j=1}^{n-1}k^2(j/p)\int v^2dW(u)dW(v) = O_P(p/n^{\frac{1}{2}}).$$

The proof for $a = 4,5,6$ is similar to that for $a = 3$, by noting that $|n_j^{-1}\sum_{t=j+1}^{n}\hat{\delta}_t(u)|^2 \le n_j^{-1}\sum_{t=j+1}^{n}|\hat{\delta}_t(u)|^2$. This completes the proof for Theorem A.3. ■

Proof of Theorem A.4: See Hong (1999, Proof of Theorem 5, for $(m,l)=(0,0)$). ■

Proof of Theorem 3: The proof of Theorem 3 consists of Theorems A.5 and A.6 below.

Theorem A.5: *Under the conditions of Theorem 6.1,* $\hat{M}(p)-M(p)\to^p 0$.

Theorem A.6: *Under the conditions of Theorem 6.1,* $M(p)\to^d N(\mu,1)$.

Proof of Theorem A.5: The proof is more tedious than but similar to that of Theorem A.1. We omit it here. ■

Proof of Theorem A.6: Define

$$\bar{\sigma}_j(u,v)=n_j^{-1}\sum_{t=j+1}^{n}\psi_t(u)\psi_{t-j}(v), \qquad j=0,1,...,n-1,$$

where $\psi_t(u)=e^{iue_t}-\varphi(u)$. Recall that $M(p)$ is defined in the same way as $\hat{M}(p)$ in (2.9) with $\{e_t\}_{t=1}^n$ replacing $\{\hat{e}_t\}_{t=1}^n$. We let $\bar{M}(p)$ be defined in the same way as $\hat{M}(p)$ with $\{\bar{\sigma}_j(u,v)\}_{j=0}^{n-1}$ replacing $\{\hat{\sigma}_j(u,v)\}_{j=1}^{n-1}$. Then it suffices to show Propositions A.3 and A.4 below.

Proposition A.3: *Under the conditions of Theorem 3,* $M(p)-\bar{M}(p)\to^p 0$.

Proposition A.4: *Under the conditions of Theorem 3,* $\bar{M}(p)\to^d N(\mu,0)$.

Proof of Proposition A.3: Given that $p\to\infty, p/n\to 0, p^{-1}\sum_{j=1}^{n-1}k^r(j/p)\to\int_0^\infty k^r(z)dz$ for $r=2,4$, it suffices to show

$$p^{-\frac{1}{2}}\int\sum_{j=1}^{n-1}k^2(j/p)n_j\left[|\tilde{\sigma}_j(u,v)|^2-|\bar{\sigma}_j(u,v)|^2\right]dW(u)dW(v)\xrightarrow{p}0, \tag{A28}$$

$\tilde{C}_0-\bar{C}_0=O_P(1)$, and $\tilde{D}_0-\bar{D}_0\to^p 0$, where $\bar{C}_0$ and $\bar{D}_0$ are defined in the same way as $\hat{C}_0$ and $\hat{D}_0$ in (2.10) and (2.11) with $\bar{\sigma}_0(u,v)$ replacing $\hat{\sigma}_0(u,v)$. We focus on the proof of (A28) only. By straightforward algebra, we have

$$\int\sum_{j=1}^{n-1}k^2(j/p)n_j\left[|\tilde{\sigma}_j(u,v)|^2-|\bar{\sigma}_j(u,v)|^2\right]dW(u)dW(v)=\hat{B}_1+2\,\mathrm{Re}(\hat{B}_2), \tag{A29}$$

where

$$\hat{A}_1 = \int \sum_{j=1}^{n-1} k^2(j/p) n_j |\tilde{\sigma}_j(u,v) - \bar{\sigma}_j(u,v)|^2 dW(u)dW(v),$$

$$\hat{B}_2 = \int \sum_{j=1}^{n-1} k^2(j/p) n_j \left[\tilde{\sigma}_j(u,v) - \bar{\sigma}_j(u,v)\right] \bar{\sigma}_j(u,v)^* dW(u)dW(v).$$

Because

$$\tilde{\sigma}_j(u,v) = \bar{\sigma}_j(u,v) - \left[n_j^{-1} \sum_{t=j+1}^{n} \psi_t(u)\right] \left[n_j^{-1} \sum_{t=j+1}^{n} \psi_{t-j}(v)\right],$$

we have

$$\begin{aligned} E|\tilde{\sigma}_j(u,v) - \bar{\sigma}_j(u,v)|^2 &\leq \left[E\left|n_j^{-1} \sum_{t=j+1}^{n} \psi_t(u)\right|^4\right]^{\frac{1}{2}} \left[E\left|n_j^{-1} \sum_{t=j+1}^{n} \psi_{t-j}(v)\right|^4\right]^{\frac{1}{2}} \\ &\leq Cn_j^{-2}, \end{aligned} \tag{A30}$$

where $E|\sum_{t=j+1}^{n} \psi_t(u)|^4 \leq Cn_j^2$ since $\{\psi_t(u)\}$ is a bounded 1-dependent random sequence with mean 0. It follows from Markov's inequality, (A30), (A12) and $p/n \to 0$ that

$$\hat{B}_1 = O_P(p/n) = o_P(1). \tag{A31}$$

Next we consider $\hat{B}_2$ in (A29). Observe that

$$|\bar{\sigma}_j(u,v)|^2 \leq 2|\sigma_j(u,v)|^2 + 2|\bar{\sigma}_j(u,v) - \sigma_j(u,v)|^2. \tag{A32}$$

Also, since $\{\psi_t(u)\}$ is 1-dependent, we have

$$\sigma_j(u,v) = \begin{cases} E\left[\psi_t(u)\psi_{t-1}(v)\right], & j = 1 \\ 0, & j > 1 \end{cases} \tag{A33}$$

and

$$E\left|\bar{\sigma}_j(u,v) - \sigma_j(u,v)\right|^2 = \left|n_j^{-1} \sum_{t=j+1}^{n} \left[\psi_t(u)\psi_{t-j}(v) - \sigma_j(u,v)\right]\right|^2 \leq Cn_j^{-1} \tag{A34}$$

given that $\psi_t(u)\psi_{t-j}(v)$ and $\psi_s(u)\psi_{s-j}(v)$ are mutually independent unless $t = s, s \pm 1, s \pm j$ and $s + 1 \pm j, s - 1 \pm j$. It follows from (A32)-(A34), $|k(z)| \leq 1$ and Markov's inequality that

$$\int \sum_{j=1}^{n-1} k^2(j/p) n_j |\bar{\sigma}_j(u,v)|^2 dW(u)dW(v)$$

$$
\begin{aligned}
&\le 2\int \sum_{j=1}^{n-1} k^2(j/p) n_j |\sigma_j(u,v)|^2 dW(u)dW(v) \\
&\quad +2\int \sum_{j=1}^{n-1} k^2(j/p) n_j |\bar{\sigma}_j(u,v)-\sigma_j(u,v)|^2 dW(u)dW(v) \\
&\le 2n\int |\sigma_1(u,v)|^2 dW(u)dW(v) \\
&\quad +2\int \sum_{j=1}^{n-1} k^2(j/p) n_j |\bar{\sigma}_j(u,v)-\sigma_j(u,v)|^2 dW(u)dW(v) \\
&= O(p^{\frac{1}{2}}) + O_P(p), \qquad \text{(A35)}
\end{aligned}
$$

where we used the fact that $n\int |\sigma_1(u,v)|^2 dW(u)dW(v) = O(p^{\frac{1}{2}})$ under $\mathbb{H}_n(p^{\frac{1}{4}}/n^{\frac{1}{2}})$. Combining (A31), A(35), $p/n \to 0$, and Cauchy–Schwarz inequality, we obtain

$$
p^{-\frac{1}{2}}|\hat{B}_2| \le |\hat{A}_1|^{\frac{1}{2}} \left[p^{-1}\int \sum_{j=1}^{n-1} k^2(j/p) n_j |\bar{\sigma}_j(u,v)|^2 \right]^{\frac{1}{2}} = O_P(p^{\frac{1}{2}}/n^{\frac{1}{2}}) = o_P(1).
$$

This completes the proof for Proposition A.3. ■

Proof of Proposition A.4: We write

$$
\int \sum_{j=1}^{n-1} k^2(j/p) n_j |\bar{\sigma}_j(u,v)|^2 dW(u)dW(v) - \bar{C}_0 \sum_{j=1}^{n-1} k^2(j/p)
$$

$$
\begin{aligned}
&= k^2(1/p)\left[n_1\int |\bar{\sigma}_1(u,v)|^2 dW(u)dW(v) - \bar{C}_0\right] + k^2(2/p)\left[n_2\int |\bar{\sigma}_2(u,v)|^2 dW(u)dW(v) - \bar{C}_0\right] \\
&\quad + \sum_{j=3}^{n-1} k^2(j/p)\left[n_j\int |\bar{\sigma}_j(u,v)|^2 dW(u)dW(v) - \bar{C}_0\right].
\end{aligned}
$$

Given that $p\to\infty, p/n\to 0, p^{-1}\sum_{j=1}^{n-1} k^r(j/p) \to \int_0^\infty k^r(z)dz$ for $r=2,4, \bar{C}_0 - C_0 = O_P(n^{-\frac{1}{2}})$ and $\bar{D}_0 - D_0 \to^p 0$, where C_0 and D_0 are defined in the same way as $\hat{C}_0$ and $\hat{D}_0$ in (2.10) and (2.11) with $\sigma_0(u,v)$ replacing $\hat{\sigma}_0(u,v)$, it suffices to show Lemmas A.7-A.11 below. These lemmas hold under $\mathbb{H}_n(p^{\frac{1}{4}}/n^{\frac{1}{2}})$.

Lemma A.7: $p^{-\frac{1}{2}}k^2(1/p)[n_1\int |\bar{\sigma}_1(u,v)|^2 dW(u)dW(v) - C_0] \to^p [2D_0\int_0^\infty k^4(z)dz]^{\frac{1}{2}}\mu$.

Lemma A.8: $p^{-\frac{1}{2}}k^2(2/p)[n_2\int |\bar{\sigma}_2(u,v)|^2 dW(u)dW(v) - C_0] \to^p 0$.

Lemma A.9: Define

$$
\hat{V} = \sum_{j=3}^{n-2} k^2(j/p) n_j^{-1} \sum_{t=j+3}^{n} \sum_{s=j+1}^{t-2} \int V_{tsj}(u,v)dW(u)dW(v),
$$

where $V_{tsj}(u,v) = C_{tsj}(u,v) + C_{stj}(u,v)^*$ and $C_{tsj}(u,v) = \psi_t(u)\psi_s(u)^*\psi_{t-j}(v)\psi_{s-j}(v)^*$. Then

$$p^{-\frac{1}{2}} \sum_{j=3}^{n-1} k^2(j/p) \left[n_j \int |\bar{\sigma}_j(u,v)| dW(u)dW(v) - C_0 \right] = p^{-\frac{1}{2}} \hat{V} + o_P(1).$$

Lemma A.10: $p^{-\frac{1}{2}}\hat{V} = p^{-\frac{1}{2}}\hat{V}_g + o_P(1)$, where

$$\hat{V}_g = \sum_{j=3}^{g} \sum_{t=g+2}^{n} \sum_{s=1}^{t-g-1} k^2(j/p) n_j^{-1} \int V_{tsj}(u,v) dW(u)dW(v),$$

and $g/p \to \infty, g/n \to 0$ as $n \to \infty$.

Lemma A.11: $[2pD_0 \int_0^\infty k^4(z)dz]^{-\frac{1}{2}} \hat{V}_g \to^d N(0,1)$.

Proof of Lemma A.7: Since $p^{-\frac{1}{2}}k^2(1/p)C_0 \to 0$ and $k(1/p) \to 1$, it suffices to show

$$p^{-\frac{1}{2}} n_1 \int |\bar{\sigma}_1(u,v)|^2 dW(u)dW(v) \xrightarrow{p} \left[2D_0 \int_0^\infty k^4(z)dz \right]^{\frac{1}{2}} \mu.$$

Because

$$|\bar{\sigma}_1(u,v)|^2 - |\sigma_1(u,v)|^2 = |\bar{\sigma}_1(u,v) - \sigma_1(u,v)|^2 + 2\,\mathrm{Re}\left\{ [\bar{\sigma}_1(u,v) - \sigma_1(u,v)]\sigma_1(u,v)^* \right\},$$

we have

$$\begin{aligned}
&\left| p^{-\frac{1}{2}} n_1 \int |\bar{\sigma}_1(u,v)|^2 dW(u)dW(v) - p^{-\frac{1}{2}} n_1 \int |\sigma_1(u,v)|^2 dW(u)dW(v) \right| \\
\le\; & p^{-\frac{1}{2}} n_1 \int |\bar{\sigma}_1(u,v) - \sigma_1(u,v)|^2 dW(u)dW(v) \\
& +2 \left[p^{-\frac{1}{2}} n_1 \int |\bar{\sigma}_1(u,v) - \sigma_1(u,v)|^2 dW(u)dW(v) \right]^{\frac{1}{2}} \left[p^{-\frac{1}{2}} n_1 \int |\sigma_1(u,v)|^2 dW(u)dW(v) \right]^{\frac{1}{2}} \\
=\; & O_P(p^{-\frac{1}{2}}) + O_P(p^{-\frac{1}{4}})
\end{aligned} \tag{A36}$$

where the equality follows from (A34) and

$$p^{-\frac{1}{2}} n_1 \int |\sigma_1(u,v)|^2 dW(u)dW(v) \to \left[2D_0 \int_0^\infty k^4(z)dz \right]^{\frac{1}{2}} \mu \tag{A37}$$

under $\mathbb{H}_n(p^{\frac{1}{4}}/n^{\frac{1}{2}})$. Combining (A36), (A37) and $p \to \infty$ yields the desired result. ■

Proof of Lemma A.8: The proof is similar to and simpler than that of Lemma A.7 because $\sigma_2(u,v) = 0$ given that $\{\psi_t(u)\}$ is 1-dependent. ■

Proof of Lemma A.9: Given the definitions of $C_{tsj}(u,v)$ and $V_{tsj}(u,v)$, we can decompose

$$\begin{aligned}\sum_{j=3}^{n-1} k^2(j/p)n_j \int |\bar{\sigma}_j(u,v)|^2 dW(u)dW(v) &= \sum_{j=3}^{n-1} k^2(j/p)n_j^{-1} \sum_{t=j+1}^{n} \int C_{ttj}(u,v)dW(u)dW(v) \\ &\quad + \sum_{j=3}^{n-2} k^2(j/p)n_j^{-1} \sum_{t=j+2}^{n} \sum_{s=j+1}^{t-1} \int V_{tsj}(u,v)dW(u)dW(v) \\ &= \hat{C} + \hat{V}, \qquad \text{(A38)}\end{aligned}$$

where $\hat{V}$ is defined in Lemma A.9, and

$$\hat{C} = \sum_{j=3}^{n-1} k^2(j/p)n_j^{-1} \sum_{t=j+1}^{n} \int C_{ttj}(u,v)dW(u)dW(v).$$

We shall show $p^{-\frac{1}{2}}[\hat{C} - C_0 \sum_{j=3}^{n-1} k^2(j/p)] \to^p 0$. Because $\{\psi_t(u)\}$ is 1-dependent, we have

$$\int E[C_{ttj}(u,v)]dW(u)dW(v) = \int E|\psi_t(u)|^2 dW(u) \int E|\psi_{t-j}(v)|^2 dW(v) = C_0. \qquad \text{(A39)}$$

Also, since $\int C_{ttj}(u,v)dW(u)dW(v)$ and $\int C_{ssj}(u,v)|\psi_s(u)|^2 dW(u)dW(v)$ are independent unless $t = s, s \pm 1, s+1 \pm j, s-1 \pm j$, we have

$$E\left\{\sum_{t=j+1}^{n} \int \left[C_{ttj}(u,v) - EC_{ttj}(u,v)\right] dW(u)dW(v)\right\}^2 \leq Cn_j. \qquad \text{(A40)}$$

It follows from (A39), (A40), (A12) and Cauchy-Schwarz inequality that

$$\begin{aligned}p^{-\frac{1}{2}}\left[\hat{C} - C_0 \sum_{j=3}^{n-1} k^2(j/p)\right] &= p^{-\frac{1}{2}} \sum_{j=3}^{n-1} k^2(j/p)n_j^{-1} \sum_{t=j+1}^{n} \int \left[C_{ttj}(u,v) - EC_{ttj}(u,v)\right] dW(u)dW(v) \\ &= O_P(p^{\frac{1}{2}}/n^{\frac{1}{2}}). \qquad \text{(A41)}\end{aligned}$$

The desired result follows from (A38), (A41) and $p/n \to 0$. ■

Proof of Lemma A.10: Following the partition technique of Hong (1999, Proof of Theorem A.3, p.1215), we first decompose $\hat{V}$ into the sums with $j \leq g$ and $j > g$ respectively:

$$\begin{aligned}\hat{V} &= \left(\sum_{j=3}^{g} \sum_{t=j+2}^{n} \sum_{s=j+1}^{t-1} + \sum_{j=g+1}^{n-2} \sum_{t=j+2}^{n} \sum_{s=j+1}^{t-1}\right) k^2(j/p)n_j^{-1} \int V_{tsj}(u,v)dW(u)dW(v) \\ &= \hat{U} + \hat{R}_1, \text{ say.} \qquad \text{(A42)}\end{aligned}$$

Next, using the fact that the sum over (t,s), where $1 \leq s < t \leq n$, can be partitioned into a sum over (t,s), where $j < s < t \leq n$ and a sum over (t,s), where $1 \leq s \leq j$ and $s < t \leq n$, we can decompose

$$\begin{aligned}\hat{U} &= \left(\sum_{j=3}^{g}\sum_{t=2}^{n}\sum_{s=1}^{t-1} - \sum_{j=3}^{g}\sum_{s=1}^{j}\sum_{t=s+1}^{n}\right) k^2(j/p)n_j^{-1}\int V_{tsj}(u,v)dW(u)dW(v) \\ &= \hat{W} - \hat{R}_2, \text{ say.} \qquad \text{(A43)}\end{aligned}$$

Moreover, $\hat{W}$ can be decomposed into the sums over $t-s > g$ and $t-s \le g$ respectively:

$$\begin{aligned}\hat{W} &= \left(\sum_{j=3}^{g}\sum_{t=g+2}^{n}\sum_{s=1}^{t-g-1} + \sum_{j=3}^{g}\sum_{t=2}^{n}\sum_{s=\max(1,t-g)}^{t-1}\right) k^2(j/p)n_j^{-1}\int V_{tsj}(u,v)dW(u)dW(v) \\ &= \hat{V}_g + \hat{R}_3, \text{ say,} \qquad \text{(A44)}\end{aligned}$$

where $\hat{V}_g$ is defined in Lemma A.10. Combining (A42)-(A44), we obtain

$$\hat{V} = \hat{V}_g + \hat{R}_1 - \hat{R}_2 + \hat{R}_3,$$

Thus, it suffices to show $p^{-\frac{1}{2}}\hat{R}_a \to^p 0$ for $a = 1, 2, 3$.

We shall compute the order of magnitude for $\hat{R}_1$ in detail; the computation of the other reminder terms is similar. We first write $\hat{R}_1$ as the sums with $s = t-1$ and $s < t-1$:

$$\begin{aligned}\hat{R}_1 &= \sum_{j=g+1}^{n-2}\sum_{t=j+2}^{n} k^2(j/p)n_j^{-1}\int V_{t(t-1)j}(u,v)dW(u)dW(v) \\ &\quad + \sum_{j=g+1}^{n-2}\sum_{t=j+2}^{n}\sum_{s=j+1}^{t-2} k^2(j/p)n_j^{-1}\int V_{tsj}(u,v)dW(u)dW(v) \\ &= \hat{R}_{11} + \hat{R}_{12}, \text{ say.} \qquad \text{(A45)}\end{aligned}$$

Now consider the first term in (A45). Because $\{\psi_t(u)\}$ is 1-dependent, $V_{t(t-1)j}(u,v)$ and $V_{s(s-1)j}(u,v)$ are independent unless $t = s, s\pm 1, s-2, s\pm j, s+1\pm, s-1\pm j, s-2\pm j$. Hence, we have

$$E\left\{\sum_{t=j+2}^{n}\int \left[V_{t(t-1)j}(u,v) - EV_{t(t-1)j}(u,v)\right] dW(u)dW(v)\right\}^2 \le Cn_j. \qquad \text{(A46)}$$

Also, since $\psi_t(u)\psi_{t-1}(u)^*$ is independent of $\psi_{t-j}(v)\psi_{t-1-j}(v)^*$ for $j > 2$, we have

$$\begin{aligned}EV_{t(t-1)j} &= 2\,\mathrm{Re}\int EC_{t(t-1)j}(u,v)dW(u)dW(v) \\ &= 2\left|\int E\left[\psi_t(u)\psi_{t-1}(u)^*\right]dW(u)\right|^2 \\ &\le C(p^{\frac{1}{2}}/n) \qquad \text{(A47)}\end{aligned}$$

under $\mathbb{H}_n(p^{\frac{1}{4}}/n^{\frac{1}{2}})$. It follows from (A46), (A47), (A12) and Cauchy-Schwarz inequality that

$$\begin{aligned}
\hat{R}_{11} &= \sum_{j=g+1}^{n-2} k^2(j/p) n_j^{-1} \sum_{t=j+2}^{n} \int EV_{t(t-1)j}(u,v) dW(u,v) \\
&\quad + \sum_{j=g+1}^{n-2} k^2(j/p) n_j^{-1} \sum_{t=j+2}^{n} \int \left[V_{t(t-1)j}(u,v) - EV_{t(t-1)j}(u,v)\right] dW(u,v) \\
&= o(p^{\frac{3}{2}}/n) + o_P(p/n^{\frac{1}{2}}), \qquad \text{(A48)}
\end{aligned}$$

where we made use of the fact that $p^{-1}\sum_{j=g+1}^{n-2} k^2(j/p) n_j^{-1} \to 0$ given (A12) and $g/p \to \infty$.

Next, we consider $\hat{R}_{12}$. Given the definition of $V_{tsj}(u,v)$, we have

$$\begin{aligned}
E\hat{R}_{12}^2 &\leq 4E\left|\sum_{j=g+1}^{n-2}\sum_{t=j+2}^{n}\sum_{s=j+1}^{t-2} k^2(j/p) n_j^{-1} \int \psi_t(u)\psi_s(u)^*\psi_{t-j}(v)\psi_{s-j}(v)^* dW(u)dW(v)\right|^2 \\
&= 4\sum_{t=g+4}^{n} E\left|\int \psi_t(u) \sum_{s=g+2}^{t-2}\sum_{j=g+1}^{s-1} k^2(j/p) n_j^{-1}\psi_{t-j}(v)\psi_s(u)\psi_{s-j}(v) dW(u)dW(v)\right|^2 \\
&\leq 4\sum_{t=g+4}^{n} \int E|\psi_t(u)|^2 E\left|\sum_{s=g+2}^{t-2}\sum_{j=g+1}^{s-1} k^2(j/p) n_j^{-1}\psi_{t-j}(v)\psi_s(u)\psi_{s-j}(v)\right|^2 dW(u)dW(v) \qquad \text{(A49)}
\end{aligned}$$

where the equality and last inequality follow because $\psi_t(u)$ is independent of $\psi_s(u)^*, \psi_{t-j}(v)$ and $\psi_{s-j}(v)^*$ for $t > s-1$ and $j > 2$.

For the second expectation in (A49), we have

$$E\left|\sum_{s=g+2}^{t-2}\sum_{j=g+1}^{s-1} k^2(j/p) n_j^{-1}\psi_{t-j}(v)\psi_s(u)\psi_{s-j}(v)\right|^2$$

$$\begin{aligned}
&= 4E\left|\sum_{s=g+2}^{t-2}\sum_{j=1}^{s-1} k^2(j/p) n_j^{-1}\psi_{t-j}(v)\psi_s(u)\psi_{s-j}(v)\mathbf{1}(t-j > s+1)\right|^2 \\
&\quad +4E\left|\sum_{s=g+2}^{t-2}\sum_{j=g+1}^{s-1} k^2(j/p) n_j^{-1}\psi_{t-j}(v)\psi_s(u)\psi_{s-j}(v)\mathbf{1}(t-j < s-1)\right|^2 \\
&\quad +4E\left|\sum_{s=g+2}^{t-2}\sum_{j=g+1}^{s-1} k^2(j/p) n_j^{-1}\psi_{t-j}(v)\psi_s(u)\psi_{s-j}(v)\mathbf{1}(t-j = s, s\pm 1)\right|^2 \\
&\leq C\sum_{s=g+2}^{t-2}\sum_{j=g+1}^{s-1} k^4(j/p) n_j^{-2}\mathbf{1}(t-j > s+1) + C\sum_{s=g+2}^{t-2}\sum_{j=g+1}^{s-1} k^4(j/p) n_j^{-2}\mathbf{1}(t-j < s-1)
\end{aligned}$$

$$+C\left[\sum_{j=g+1}^{t-2} k^2(j/p)n_j^{-1}\right]^2. \tag{A50}$$

Combining (A49), (A50) and (A12), we obtain

$$E\hat{R}_{12}^2 \leq C\sum_{j=g+1}^{n-2} k^4(j/p) + Cn\left[\sum_{j=g+1}^{t-2} k^2(j/p)n_j^{-1}\right]^2 = o(p+p^2/n)$$

given $g/p \to \infty$. It follows that $p^{-\frac{1}{2}}\hat{R}_{12} \to^p 0$ by Chebyshev's inequality. Similarly, we can obtain $E\hat{R}_2^2 = O(pg/n)$ and $E\hat{R}_3^2 = O(pg/n)$. Therefore, $p^{-\frac{1}{2}}\hat{R}_2 \to^p 0$ and $p^{-\frac{1}{2}}\hat{R}_3 \to^p 0$ given $g/n \to 0$. This completes the proof. ■

Proof of Lemma A.11: The proof is exactly the same as the proof for Theorem A.4 of Hong (1999, pp.1215-1217), which applies Brown's (1971) martingale central limit theorem. The fact that $\{e_t\}$ is 1-dependent rather than i.i.d. does not alter any change of the proof there because $\psi_t(u), \psi_{t-j}(v), \psi_s(u)$ and $\psi_{s-j}(v)$ are mutually independent for $t-s>g\to\infty$ and $2<j<g$. ■

Proof of Theorem 4: We shall show Theorems A.7 and A.8 below.

Theorem A.7: *Under the conditions of Theorem 4,* $\hat{M}(\hat{p}) - M(\hat{p}) \xrightarrow{p} 0$.

Theorem A.8: *Under the conditions of Theorem 4,* $M(\hat{p}) - M(p) \xrightarrow{p} 0$.

Proof of Theorem A.7: Given that $p\to\infty, p/n\to 0, p^{-1}\sum_{j=1}^{n-1}k^r(j/p) \to \int_0^\infty k^r(z)dz$ for $r=2,4$, it suffices to show

$$\hat{B} = p^{-\frac{1}{2}}\sum_{j=1}^{n-1} k^2(j/\hat{p})n_j\left[|\hat{\sigma}_j(u,v)|^2 - |\tilde{\sigma}_j(u,v)|^2\right] \xrightarrow{p} 0. \tag{1}$$

Given the conditions on $k(\cdot)$, there exists a symmetric monotonic decreasing function $k_0(z)$ of $z>0$ such that $|k(z)| \leq k_0(z)$ for all $z>0$, and $k_0(\cdot)$ satisfies Assumption A.2. It follows that for any constants $\epsilon, \eta > 0$,

$$P\left(|\hat{B}|>\epsilon\right) \leq P\left(|\hat{B}|>\epsilon, |\hat{p}/p-1|\leq\eta\right) + P\left(|\hat{p}/p-1|>\eta\right), \tag{2}$$

where the second term vanishes for all $\eta>0$ given $\hat{p}/p-1 \to^p 0$. Thus it remains to show that the first term also vanishes as $n\to\infty$.

Because $|\hat{p}/p-1|\leq\eta$ implies $\hat{p}\leq(1+\eta)p$, we have that for $|\hat{p}/p-1|\leq\eta$,

$$|\hat{B}| \leq (1+\eta)^{\frac{1}{2}}[(1+\eta)p]^{-\frac{1}{2}}\sum_{j=1}^{n-1} k_0^2[j/(1+\eta)p]n_j\left[|\hat{\sigma}_j(u,v)|^2 - |\tilde{\sigma}_j(u,v)|^2\right] \xrightarrow{p} 0$$

for any $\eta > 0$ given (A6), where the inequality follows from the fact that $|k(z)| \leq |k_0(z)|$. This completes the proof of Theorem A.7. ■

Proof of Theorem A.8: See Hong (1999, Proof of Theorem 4, for $(m, l) = (0, 0)$). ■

Figure 1. Size of Testing Conditional Mean at 5% Level

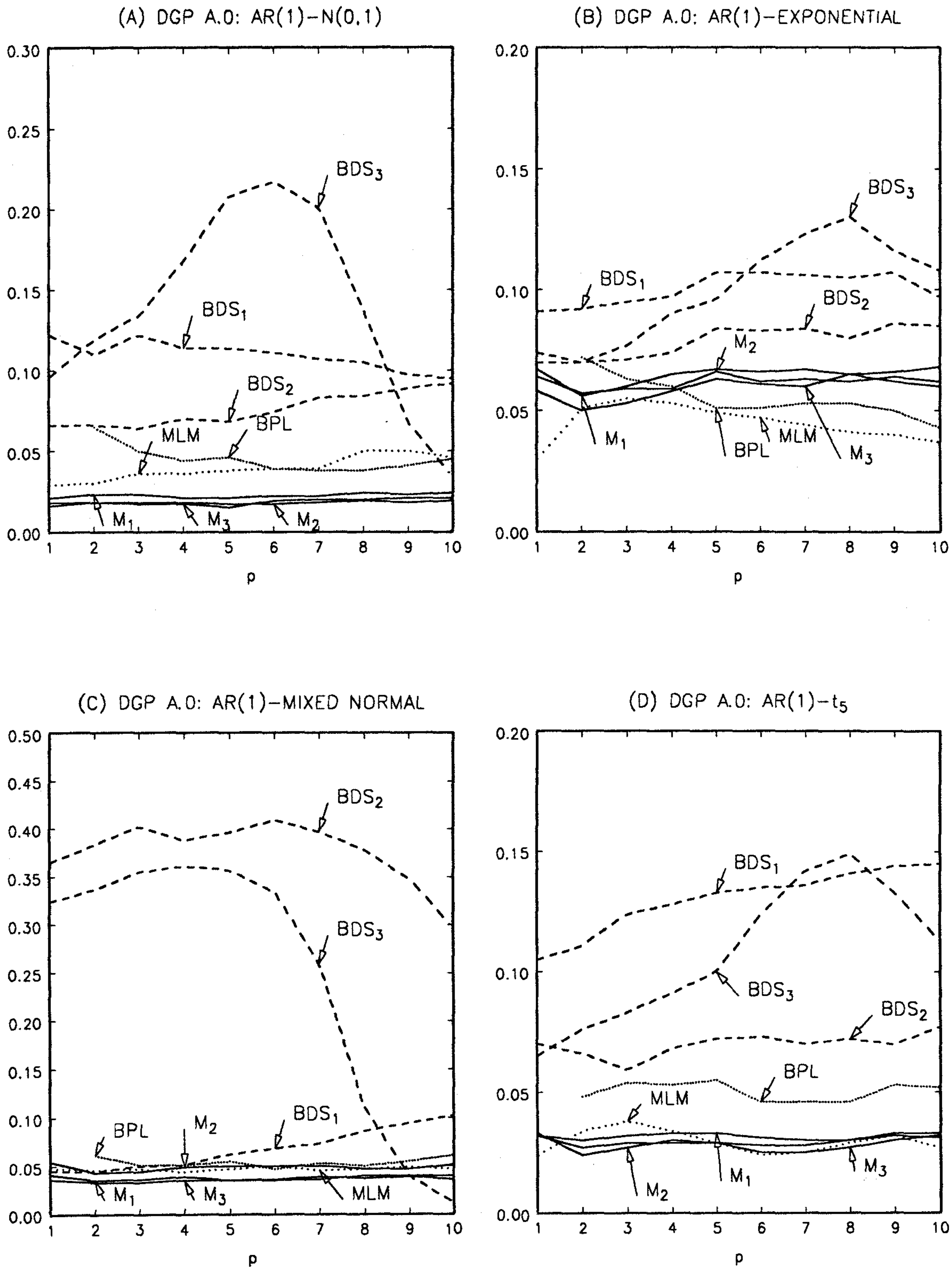

Figure 2. Size of Testing Conditional Variance at 5% Level

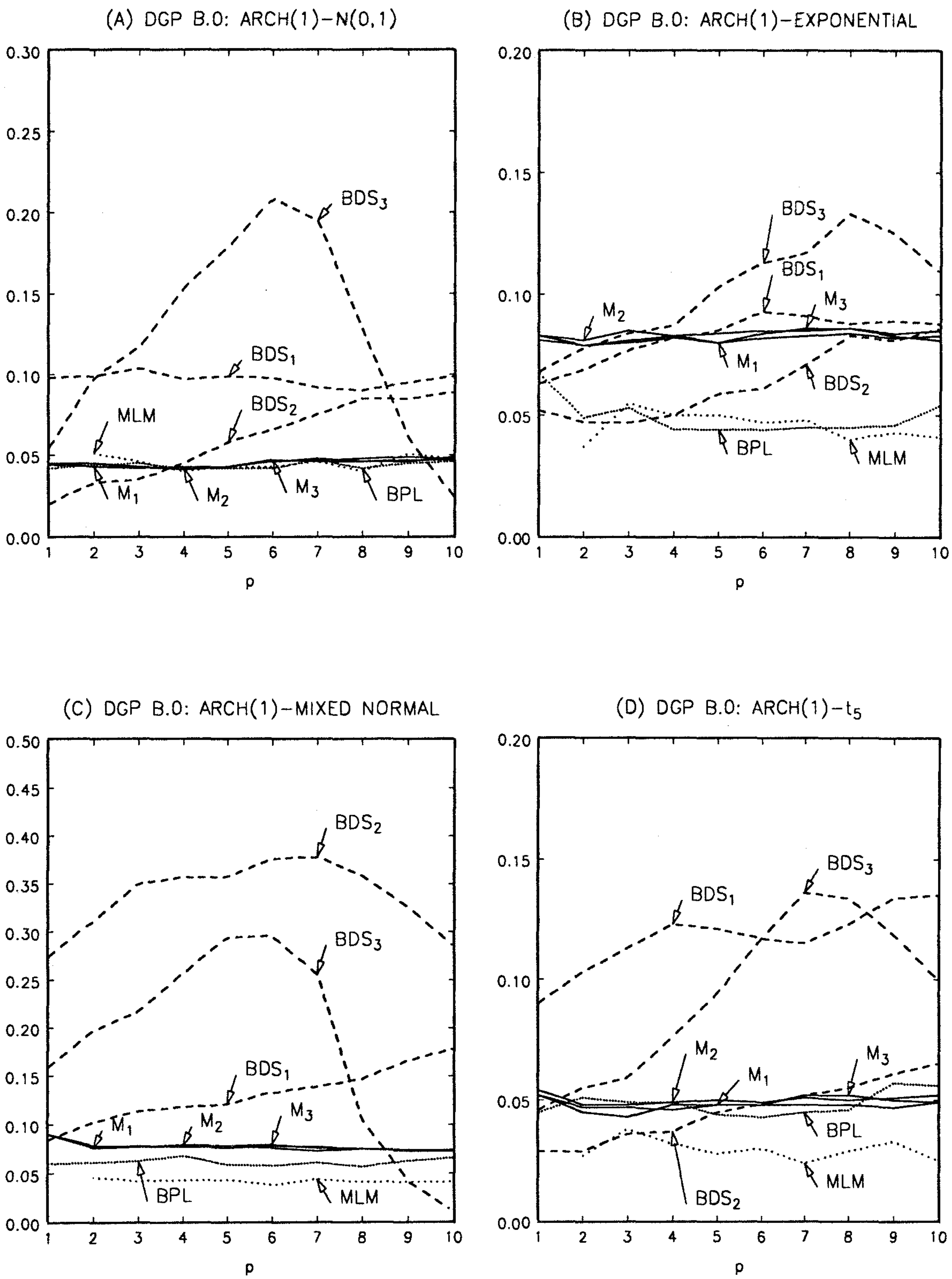

Figure 3. Size-Corrected Power of Testing Conditional Mean at 5% Level

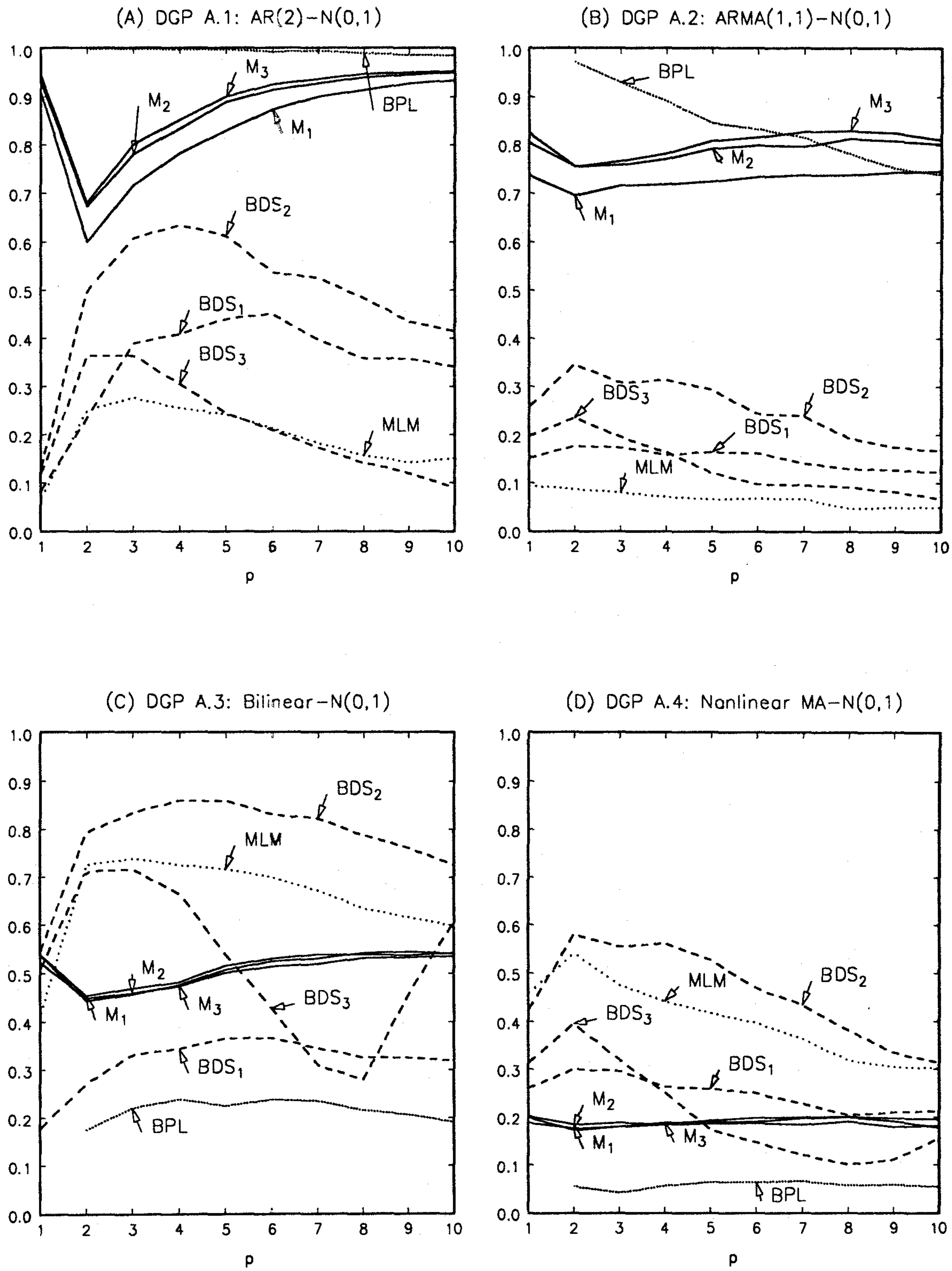

Figure 3. Continued

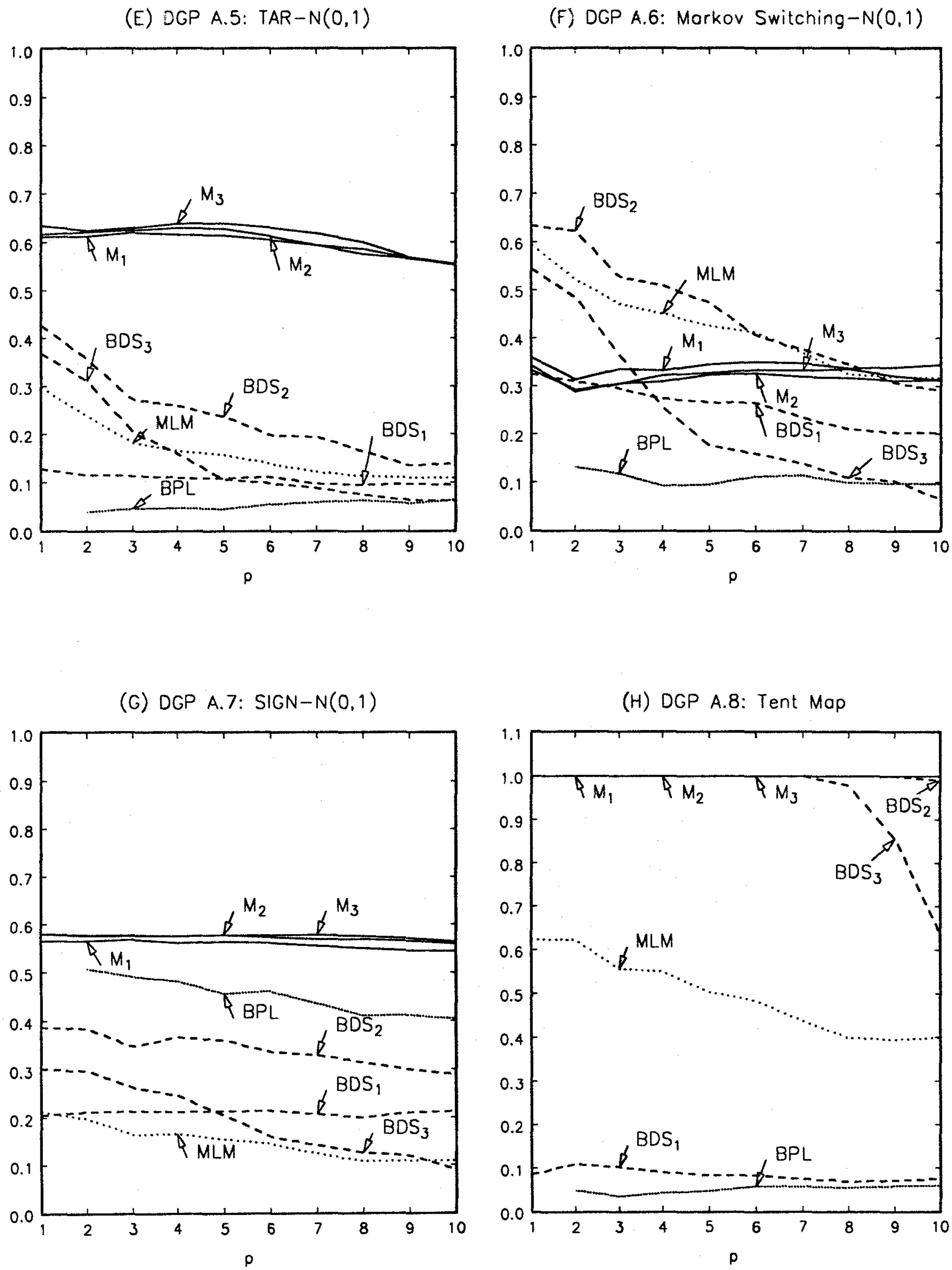

Figure 4. Size-Corrected Power of Testing Conditional Variance at 5% Level

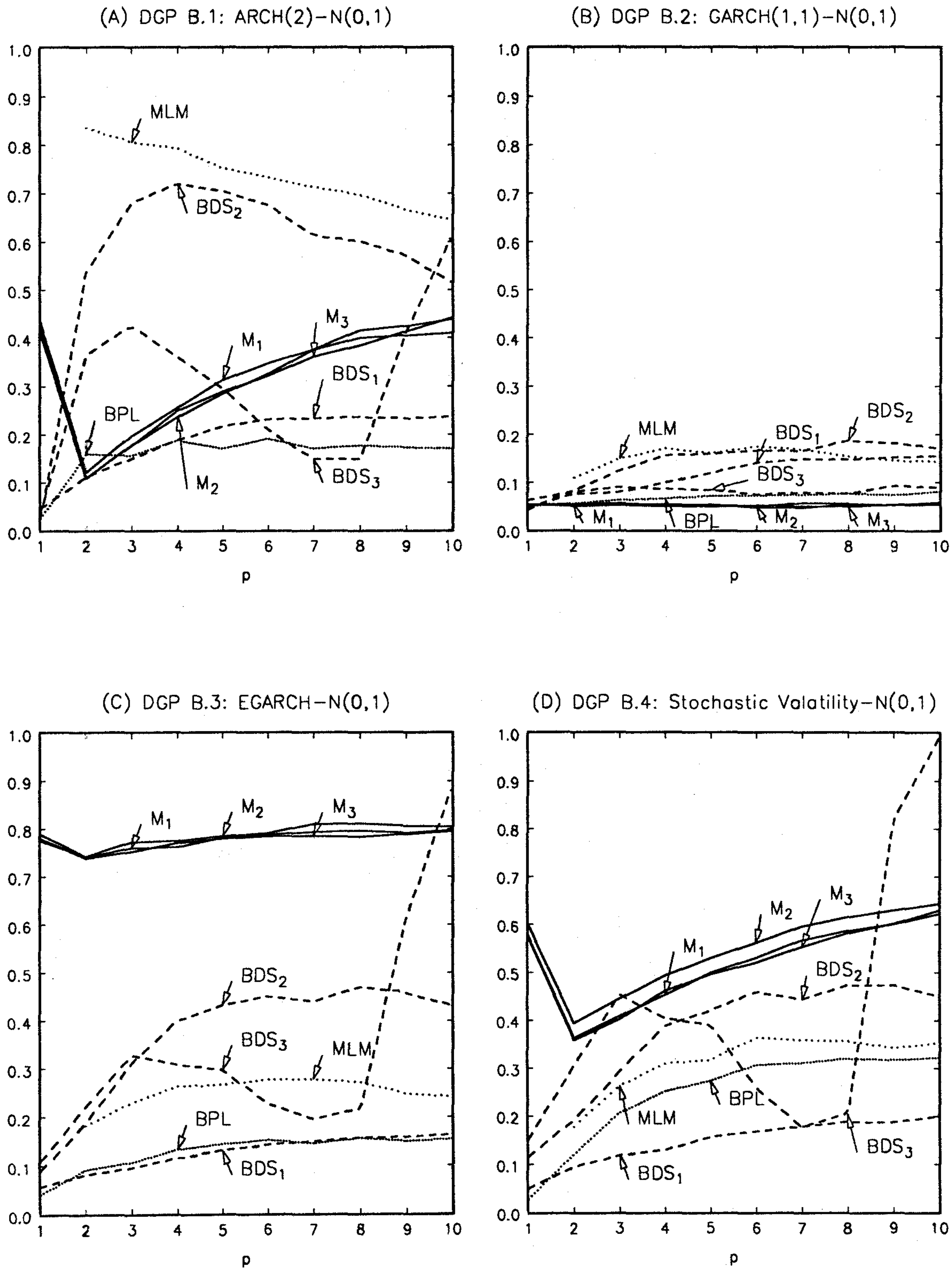

Figure 4. Continued