Econometric Theory, **19**, 2003, 1065–1121. Printed in the United States of America.
DOI: 10.1017/S0266466603196089

DIAGNOSTIC CHECKING FOR THE ADEQUACY OF NONLINEAR TIME SERIES MODELS

YONGMIAO HONG
Cornell University

TAE-HWY LEE
University of California, Riverside

We propose a new diagnostic test for linear and nonlinear time series models, using a generalized spectral approach. Under a wide class of time series models that includes autoregressive conditional heteroskedasticity (ARCH) and autoregressive conditional duration (ACD) models, the proposed test enjoys the appealing "nuisance-parameter-free" property in the sense that model parameter estimation uncertainty has no impact on the limit distribution of the test statistic. It is consistent against any type of pairwise serial dependence in the model standardized residuals and allows the choice of a proper lag order via data-driven methods. Moreover, the new test is asymptotically more efficient than the correlation integral–based test of Brock, Hsieh, and LeBaron (1991, *Nonlinear Dynamics, Chaos, and Instability: Statistical Theory and Economic Evidence*) and Brock, Dechert, Scheinkman, and LeBaron (1996, *Econometric Reviews* 15, 197–235), the well-known BDS test, against a class of plausible local alternatives (not including ARCH). A simulation study compares the finite-sample performance of the proposed test and the tests of BDS, Box and Pierce (1970, *Journal of the American Statistical Association* 65, 1509–1527), Ljung and Box (1978, *Biometrika* 65, 297–303), McLeod and Li (1983, *Journal of Time Series Analysis* 4, 269–273), and Li and Mak (1994, *Journal of Time Series Analysis* 15, 627–636). The new test has good power against a wide variety of stochastic and chaotic alternatives to the null models for conditional mean and conditional variance. It can play a valuable role in evaluating adequacy of linear and nonlinear time series models. An empirical application to the daily S&P 500 price index highlights the merits of our approach.

We thank the co-editor (Don Andrews) and two referees for careful and constructive comments that have lead to significant improvement over an earlier version. We also thank C.W.J. Granger, D. Tjøstheim, and Z. Xiao for helpful comments. Hong's participation is supported by the National Science Foundation via NSF grant SES–0111769. Lee thanks the UCR Academic Senate for research support. Address correspondence to: Yongmiao Hong, Department of Economics and Department of Statistical Science, Cornell University, Ithaca, NY 14853-7601, USA; e-mail: yh20@cornell.edu; or to Tae-Hwy Lee, Department of Economics, University of California, Riverside, Riverside, CA 92521-0427, USA; e-mail: tae.lee@ucr.edu.

© 2003 Cambridge University Press 0266-4666/03 $12.00

https://doi.org/10.1017/S0266466603196089 Published online by Cambridge University Press

1. INTRODUCTION

The development of nonlinear time series analysis has been advancing rapidly (e.g., Subba Rao and Gabr, 1984; Priestley, 1988; Tong, 1990; Brock, Hsieh, and LeBaron, 1991; Granger and Teräsvirta, 1993; Tjøstheim, 1994; Teräsvirta, Tjøstheim, and Granger, 1994; Terdik, 1999). In modern time series analysis, one often considers the data generating process

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

where I_{t-1} is the information set available at time $t-1$ and $\{\varepsilon_t\}$ is a sequence of independently and identically distributed (i.i.d.) innovations. Often, $\{\varepsilon_t\}$ has mean 0 and variance 1 such that $g_0(I_{t-1}) = E(Y_t|I_{t-1})$ and $h_0^2(I_{t-1}) = \text{var}(Y_t|I_{t-1})$ almost surely (*a.s.*). For various examples that belong to class (1.1), see, e.g., Tong (1990) and Granger and Teräsvirta (1993). It is also possible that $\{\varepsilon_t\}$ is an i.i.d. nonnegative sequence with $E(\varepsilon_t) = 1$. An example is the autoregressive conditional duration (ACD) process introduced in Engle and Russell (1998) for irregularly spaced time series, where $g_0(I_{t-1}) = 0$ and $h_0(I_{t-1}) = E(Y_t|I_{t-1})$ is the conditional duration of the nonnegative time duration process $\{Y_t\}$.

Various models for $g_0(\cdot)$ and $h_0(\cdot)$ have been proposed in the literature. Consider a 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 known parametric specifications for $g_0(\cdot)$ and $h_0(\cdot)$, $\theta \in \Theta$ is an unknown finite-dimensional parameter vector, and $\{e_t(\theta)\}$ is an unobservable series. Specification (1.2) covers most commonly used linear and nonlinear time series models. Examples include ACD, autoregressive conditional heteroskedasticity (ARCH), autoregressive moving average (ARMA), bilinear, nonlinear moving average, Markov regime-switching, smooth transition, exponential, and threshold autoregressive models. When $g(\cdot, \theta)$ and $h(\cdot, \theta)$ are correctly specified for $g_0(\cdot)$ and $h_0(\cdot)$, i.e., when there exists some $\theta_0 \in \Theta$ such that $g(\cdot, \theta_0) = g_0(\cdot)$ and $h(\cdot, \theta_0) = h_0(\cdot)$ *a.s.*, the model standardized error series $\{e_t(\theta_0)\}$ coincides with the true innovation series $\{\varepsilon_t\}$ 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 $h_0(\cdot)$, i.e., if there exists no $\theta \in \Theta$ such that $g(\cdot, \theta) = g_0(\cdot)$ and/or $h(\cdot, \theta) = h_0(\cdot)$ *a.s.*, $\{e_t(\theta)\}$ will be serially dependent for all $\theta \in \Theta$. Consequently, to test adequacy of model (1.2), one can check whether there exists some $\theta_0 \in \Theta$ such that $\{e_t(\theta_0)\}$ is i.i.d. Tests for i.i.d. rather than for white noise are more suitable and useful in nonlinear time series analysis, in particular when higher order conditional moments or the conditional probability distribution are of interest (e.g., Brock et al., 1991; Brock, Dechert, Scheinkman, and LeBaron, 1996, Christoffersen, 1998; Diebold, Gunther, and Tay, 1998; Kim, Shephard, and Chib, 1998; Clement and Smith, 2000; Elerian, Chib, and Shephard, 2001). See also an empirical application in Section 8.

https://doi.org/10.1017/S0266466603196089 Published online by Cambridge University Press

Because $\{e_t(\theta_0)\}$ is unobservable, one has to use the standardized estimated residual

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

where $\hat{\theta}$ is a consistent estimator of θ_0 based on a random sample $\{Y_t\}_{t=1}^n$ of size n and $\hat{I}_t$ is the observed information set available at period t that may involve certain initial values. To construct an asymptotically valid test procedure, it is important to examine whether and how the use of $\{\hat{e}_t\}_{t=1}^n$ rather than $\{e_t \equiv e_t(\theta_0)\}_{t=1}^n$ affects the limit distribution of a test statistic, besides its power property (see Tjøstheim, 1996).

Often, the information set I_{t-1} consists of lagged variables $\{Y_{t-j}, j > 0\}$. When $g_0(I_{t-1})$ is linear in I_{t-1}, Y_t is called linear in conditional mean on I_{t-1} (Lee, White, and Granger, 1993). If in addition $h_0(I_{t-1}) = \sigma$ *a.s.*, $\{Y_t\}$ is called completely linear in I_{t-1} (Granger, 2001; Granger and Lee, 1999). Assuming $h_0(I_{t-1}) = \sigma$ *a.s.* in an ARMA framework, Box and Pierce (1970) and Ljung and Box (1978) propose a diagnostic test for an ARMA(p_0, q_0) model:

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

where $\hat{\rho}(j)$ is the sample autocorrelation function of $\{\hat{e}_t\}_{t=1}^n$, $\hat{e}_t = Y_t - g(\hat{I}_{t-1}, \hat{\theta})$, and $g(\hat{I}_{t-1}, \hat{\theta})$ is an estimated ARMA(p_0, q_0) model. Here, $\{\hat{e}_t\}$ is the usual estimated residuals because no conditional variance estimation is involved. The degrees of freedom of the Box–Pierce–Ljung (BPL) test depend on $p_0 + q_0$, the number of the estimated parameters. Hong (1996) and Paparoditis (2000a, 2000b) propose spectrum-based diagnostic tests that generalize the BPL(p) test.

It has been pointed out for a long time (e.g., Granger and Anderson, 1978; Granger, 1983) that the Box–Pierce–Ljung test has no power against nonlinear dependencies with zero autocorrelation, such as some bilinear and nonlinear moving-average (MA) processes. Using the sample autocorrelation function of squared residuals, McLeod and Li (1983) suggest a test for linearity against unspecified nonlinearity. For a null ARMA(p_0, q_0) model, the McLeod and Li (1983) test statistic is

$$\mathrm{ML}(p) = n(n+2)\sum_{j=1}^{p}(n-j)^{-1}\hat{\rho}_2^2(j) \xrightarrow{d} \chi^2_p, \tag{1.5}$$

where $\hat{\rho}_2(j)$ is the sample autocorrelation function of $\{\hat{e}_t^2\}_{t=1}^n$ and $\hat{e}_t = Y_t - g(\hat{I}_{t-1}, \hat{\theta})$. This test has good power against ARCH. It also has power against departures from linearity that have apparent ARCH structures. The null limit distribution of the test statistic is a χ^2_p distribution; the degrees of freedom need not be adjusted when only an ARMA model is estimated. As pointed out in

https://doi.org/10.1017/S0266466603196089 Published online by Cambridge University Press

Granger and Teräsvirta (1993), ML(p) is asymptotically equivalent to the Lagrange multiplier test for ARCH of Engle (1982).

When testing the adequacy of an ARCH/generalized autoregressive moving average (GARCH) model, many researchers have applied the McLeod–Li test to the squares of the estimated standardized residuals.[1] Li and Mak (1994) show that this procedure is misleading because its asymptotic distribution is not a χ^2 distribution if ML(p) is applied to the residuals standardized by estimated ARCH/GARCH models. Li and Mak (1994) propose corrected statistics. In fact, their test is asymptotically equivalent to the Lagrange multiplier test presented in Lundbergh and Teräsvirta (1998), which is a test of the standardized errors being i.i.d. against the alternative that they follow an ARCH model. Li and Mak (1994) also provide a simpler statistic when the fitted conditional variance model is ARCH(r). For a null ARMA(p_0, q_0)-ARCH(r) model, the simpler version of Li and Mak's test statistic can be written as

$$\mathrm{LM}(p,r) = n(n+2)\sum_{j=r+1}^{p}(n-j)^{-1}\hat{\rho}_2^2(j) \xrightarrow{d} \chi^2_{p-r}, \qquad p > r, \tag{1.6}$$

where $\hat{\rho}_2(j)$ is the sample autocorrelation function of $\{\hat{e}_t^2\}_{t=1}^n$, $\hat{e}_t = [Y_t - g(\hat{I}_{t-1}, \hat{\theta})]/h(\hat{I}_{t-1}, \hat{\theta})$, and $h(\hat{I}_{t-1}, \hat{\theta})$ is an estimated ARCH(r) model. The null limit distribution depends on r, the order of the ARCH model. When other conditional variance models are estimated, the test statistic itself has to be modified as suggested by Li and Mak (1994) or by Lundbergh and Teräsvirta (2002).

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

$$\mathrm{BDS}(m,d) = n^{1/2}[\hat{C}_m(d) - \hat{C}_1(d)^m]/\hat{V}_m^{1/2}, \tag{1.7}$$

where the sample correlation integral (cf. Grassberger and Procaccia, 1983)

$$\begin{aligned}\hat{C}_m(d) &= \frac{2}{n(n-1)}\sum_{t=m+1}^{n}\sum_{s=m}^{t-1}\prod_{j=0}^{m-1}\mathbf{1}(|\hat{e}_{t-j} - \hat{e}_{s-j}| < d) \\ &\xrightarrow{p} P\left[\prod_{j=0}^{m-1}\mathbf{1}(|e_{t-j} - e_{s-j}| < d)\right] \equiv C_m(d),\end{aligned} \tag{1.8}$$

$\mathbf{1}(\cdot)$ is an indicator function, m is the so-called embedding dimension, d is a distance parameter, and $\hat{V}_m$ is an asymptotic variance estimator. 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 an i.i.d. sequence. 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 processes. To see this, observe that when $\{e_t\}$ is i.i.d., we have

https://doi.org/10.1017/S0266466603196089 Published online by Cambridge University Press

$$C_m(d) = C_1(d)^m \tag{1.9}$$

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.[2]

As shown in Brock et al. (1991, Ch. 2 and Appendix D), BDS(m,d) has the appealing "nuisance-parameter-free" property that any $n^{1/2}$-consistent parameter estimator $\hat{\theta}$ has no impact on its null limit distribution under a class of conditional mean models $g(\cdot,\theta)$. This, together with good power against a wide range of dependent alternatives, has made BDS(m,d) a convenient and powerful diagnostic tool for nonlinear time series models. It has been recommended by Brock et al. (1991) as a portmanteau lack of fit test for nonlinear time series models in the same spirit as Box and Jenkins (1970, p. 29) recommend Box–Pierce–Ljung's test for linear time series models.

Nevertheless, BDS(m,d) has certain features one might 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). More generally, when conditional variance estimation is involved, the limit distribution of BDS(m,d) depends on the nature of estimator $\hat{\theta}$, and how to modify the test statistic is unknown.[3] This is troublesome in practice. 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 fixed but arbitrary. Because $m-1$ is actually the largest lag order used, BDS(m,d) has no power against alternatives for which serial dependence in $\{e_t\}$ occurs only at the lag orders equal to or larger than 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. There exists no rule guided by chaos theory for choosing parameters m and d, although Brock et al. (1991) have recommended a simple rule of thumb based on their simulation study. Moreover, as shown in Section 5, BDS(m,d) has suboptimal power against some local alternatives. For example, it can detect a local ARCH(1)-type alternative with parametric rate $n^{-1/2}$ but a local MA(1) alternative with rate $n^{-1/4}$ only.

In this paper, we propose a new diagnostic test for time series model (1.2), using a generalized spectrum proposed in Hong (1999). The 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 and ACD models. It is consistent against any type of pairwise serial dependence across various lags in the model standardized residuals, a property not attainable by the BDS

https://doi.org/10.1017/S0266466603196089 Published online by Cambridge University Press

test. It can detect a class of local alternatives with a rate slightly slower than the parametric rate $n^{-1/2}$ but much faster than $n^{-1/4}$. This class includes both MA and ARCH-type local alternatives. Finally, generalized spectral smoothing allows one to choose a lag order via data-driven methods, which are more objective than an arbitrary choice or a "rule of thumb" and thus give more robust power. A simulation study compares the proposed test and the tests of BDS, Box–Pierce–Ljung, and McLeod–Li/Li–Mak in finite samples. The new test has reasonable power against a wide variety of stochastic and chaotic alternatives to the null models. It is a useful addition to the existing diagnostic tool kit for time series models (see Barnett, Gallant, Hinich, Jungeilges, Kaplan, and Jensen, 1997). An empirical application to the daily S&P 500 index highlights the merits of the proposed test. We emphasize, however, that our procedure is best viewed as a complement rather than a substitute to the BDS test, which is motivated from an interesting chaotic theory.[4]

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

2. A NEW DIAGNOSTIC TEST

Hong (1999) proposes a generalized spectrum as an analytic tool for linear and nonlinear time series. Suppose the time series $\{e_t\}$ is strictly stationary. The basic idea of Hong (1999) is to consider the spectrum of the transformed series $\{e^{iue_t}\}$, where $u \in \mathbb{R} = (-\infty,\infty)$. Define

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

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}).[5] Thus, $\sigma_j(u,v) = 0$ for all $(u,v) \in \mathbb{R}^2$ if and only if e_t and e_{t-j} are independent. Suppose $\sup_{(u,v)\in\mathbb{R}^2} \sum_{j=-\infty}^{\infty} |\sigma_j(u,v)| < \infty$, 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 $\sigma_j(u,v)$ exists:

https://doi.org/10.1017/S0266466603196089 Published online by Cambridge University Press

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

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

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

where $R(j) = \mathrm{cov}(e_t, e_{t-j})$. For this reason $f(\omega,u,v)$ is called in Hong (1999) a "generalized spectral density" of $\{e_t\}$. The introduction of parameters (u,v) offers much flexibility in capturing serial dependence in $\{e_t\}$. The generalized spectrum $f(\omega,u,v)$ can capture any type of pairwise dependence across various lags in $\{e_t\}$, including those with zero autocorrelations. Searching over the domain of (u,v), for example, one can find the "maximal dependence" of $\{e_t\}$ at each frequency ω, as given by

$$s(\omega) = \sup_{(u,v)\in\mathbb{R}^2}|f(\omega,u,v)|, \qquad \omega\in[-\pi,\pi],$$

where $|\cdot|$ is the Euclidean norm. This maximal spectral dependence may be contributed from linear or nonlinear serial dependence in $\{e_t\}$. A generalized spectral peak at some frequency will indicate a cycle, seasonality, or periodicity due to nonlinear dependence (e.g., volatility clustering) when $\{e_t\}$ is a white noise.[6]

The generalized spectrum $f(\omega,u,v)$ differs from the well-known higher order spectra, which are the Fourier transforms of higher order cumulants (cf. Brillinger and Rosenblatt, 1967a, 1967b; Subba Rao and Gabr, 1980, 1984; Terdik, 1999). It does not require any moment condition on e_t. This is appealing because, for example, it has been argued that 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—the Fourier transform of third-order cumulants—will miss them. We note, however, that $f(\omega,u,v)$ cannot capture dependent processes that are pairwise serially independent (i.e., e_t and e_{t-j} are independent for any nonzero j but $\{e_t\}$ is serially dependent), which may or may not be captured by the bispectrum. It would be interesting to compare the generalized spectrum and the bispectrum thoroughly, but this is beyond the scope of this paper and should be pursued elsewhere.

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

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

https://doi.org/10.1017/S0266466603196089 Published online by Cambridge University Press

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 $\{e_t\}$, Hong (1999) suggests that one compare two consistent estimators of $f(\omega,u,v)$ and $f_0(\omega,u,v)$ via an L_2-norm. Define

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

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

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

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 the Bartlett, Daniell, quadratic-spectral, and truncated kernels (e.g., Priestley, 1981, p. 441). The factor $(1-|j|/n)^{1/2}$ is a finite-sample correction factor that delivers a better approximation to the finite-sample distribution. We also have a consistent estimator for $f_0(\omega,u,v)$:

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

Let $W:\mathbb{R}\to\mathbb{R}^+$ be a nondecreasing function such that $W'(u) = w(v)$ exists and is symmetric about 0, with $\int dW(u) = \int w(u)\,du < \infty$. Examples of $W(\cdot)$ are the cumulative distribution functions of $N(0,1)$, double exponential, and uniform distributions.[7] Then a test for the i.i.d. hypothesis of $\{e_t\}$ can be based on a properly standardized L_2-norm:

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

https://doi.org/10.1017/S0266466603196089 Published online by Cambridge University Press

where the second equality follows from Parseval's identity, $\hat{\sigma}_j(u,v) = \hat{\sigma}_{-j}(v,u)$, symmetry of weighting functions $k(\cdot)$ and $w(\cdot)$. Moreover, the centering and scaling factors have the following values:

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

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

Throughout, unspecified integrals are taken over the entire Euclidean space of proper dimension. The test statistic $\hat{M}(p)$ involves one- and two-dimensional numerical integrations with respect to (u,v), which can be implemented using, e.g., Gauss–Legendre quadratures. Note that $\hat{M}(p)$ involves no numerical integration over frequency ω, which has been integrated out as a result of 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 and also over (u,v), and the distribution theory might be different also. 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 now derive the null limit distribution of $\hat{M}(p)$ and establish its "nuisance-parameter-free" property under a wide class of time series models. Following are regularity conditions.

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

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

Assumption A.3. Let I_t be the pseudo information set from time 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 $E \sup_{\theta\in\Theta_0} \|h^{-1}(I_t,\theta)(\partial/\partial\theta)g(I_t,\theta)\|^4$, $E \sup_{\theta\in\Theta_0} \|h^{-1}(I_t,\theta)(\partial/\partial\theta)h(I_t,\theta)\|^4$, $E \sup_{\theta\in\Theta_0} \|h^{-1}(I_t,\theta)(\partial^2/\partial\theta^2)g(I_t,\theta)\|^2$, $E \sup_{\theta\in\Theta_0} \|h^{-1}(I_t,\theta)(\partial^2/\partial\theta^2)h(I_t,\theta)\|^2$, and $E \sup_{\theta\in\Theta_0} [e_t^4(\theta)]$ all bounded by some constant $C \in (0,\infty)$, where $e_t(\theta) = [Y_t - g(I_{t-1},\theta)]/h(I_{t-1},\theta)$.

Assumption A.4. Let $\hat{I}_t$ be the observed information set available at time t that may involve certain initial values. Then

https://doi.org/10.1017/S0266466603196089 Published online by Cambridge University Press

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

and

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

Assumption A.5. $k:\mathbb{R} \to [-1,1]$ is symmetric about 0 and is continuous at 0 and all points in $\mathbb{R}$ 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 such that the derivative $W'(u) = w(u)$ exists and is symmetric about 0, with $\int_{-\infty}^{\infty} dW(u) < \infty$ 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 (DGP) $\{Y_t\}$, model 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 $\mathrm{var}(Y_t) < \infty$. An example with $\mathrm{var}(Y_t) = \infty$ is the integrated GARCH(1,1) process (Engle and Bollerslev, 1986). In Assumption A.2, we permit but do not require $\hat{\theta}$ to be a quasi–maximum likelihood estimator (Lee and Hansen, 1994; Lumsdaine, 1996). Any $n^{1/2}$-consistent estimator $\hat{\theta}$ suffices. Assumption A.3 is a standard condition on the conditional mean and conditional variance models. We require that the fourth moment of the standardized error e_t exist.

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. This condition easily holds for many time series models. To illustrate this, we first consider an invertible MA(1) model $Y_t = \alpha u_{t-1} + u_t$, where $u_t = \sigma e_t, \{e_t\}$ is i.i.d.(0,1), and $|\alpha| < 1$. Here, we have $g(I_{t-1},\theta) = \alpha u_{t-1}$ and $h(I_{t-1},\theta) = \sigma$, where $\theta = (\alpha,\sigma)'$. Furthermore, we have $I_{t-1} = \{Y_{t-1},\ldots,Y_1,Y_0,\ldots\}$ and $\hat{I}_{t-1} = \{Y_{t-1},\ldots,Y_1,\hat{u}_0\}$, where $\hat{u}_0$ is some assumed value (e.g., $\hat{u}_0 = 0$) for u_0. The condition on $h(\cdot,\cdot)$ holds trivially, so we focus on the condition on $g(\cdot,\cdot)$. By recursive substitution, we obtain

$$g(I_{t-1},\theta) - g(\hat{I}_{t-1},\theta) = \left[\sum_{j=1}^{t-1} (-1)^{j-1} \alpha^j Y_{t-j} + \alpha^t u_0 - \sum_{j=1}^{t-1} (-1)^{j-1} \alpha^j Y_{t-j} - \alpha^t \hat{u}_0 \right].$$

It follows that

$$\sum_{t=1}^{n} E \sup_{\theta\in\Theta_0} \left| \frac{g(I_{t-1},\theta) - g(\hat{I}_{t-1},\theta)}{h(\hat{I}_{t-1},\theta)} \right| \leq \sum_{t=1}^{\infty} E \sup_{\theta\in\Theta_0} \left[\frac{|\alpha|^t(|u_0| + |\hat{u}_0|)}{\sigma} \right] \leq C$$

provided $|\alpha| < 1$, $0 < \sigma < \infty$, $E|e_t| < \infty$, and $E|\hat{u}_0| < \infty$.

https://doi.org/10.1017/S0266466603196089 Published online by Cambridge University Press

Next, we consider a GARCH(1,1) model: $Y_t = e_t h_t$, where $h_t^2 = \omega + \alpha h_{t-1} + \beta Y_{t-1}^2$. Here, we have $I_{t-1} = \{Y_{t-1}, \ldots, Y_1, Y_0, \ldots\}$ and $\hat{I}_{t-1} = \{Y_{t-1}, \ldots, Y_1, \hat{h}_0^2\}$, where $\hat{h}_0^2$ is an assumed value (e.g., $\hat{h}_0^2 = 1$) for initial variance h_0^2. In this case, $g(I_{t-1}, \theta) = 0$ so the condition on $g(\cdot,\cdot)$ holds trivially. By recursive substitution, we have

$$h^2(I_{t-1},\theta) - h^2(\hat{I}_{t-1},\theta) = \omega + \beta \sum_{j=0}^{t-2} \alpha^j Y_{t-1-j}^2 + \beta\alpha^{t-1} h^2(I_0,\theta) - \omega - \beta \sum_{j=0}^{t-2} \alpha^j Y_{t-1-j}^2 - \beta\alpha^{t-1}\hat{h}_0^2.$$

It follows that

$$\sum_{t=1}^{n} E \sup_{\theta\in\Theta_0} \left| \frac{h(I_{t-1},\theta) - h(\hat{I}_{t-1},\theta)}{h(\hat{I}_{t-1},\theta)} \right| \le \sum_{t=1}^{\infty} E \sup_{\theta\in\Theta_0} \left| \frac{\beta\alpha^{t-1}[h^2(I_0,\theta) + \hat{h}_0^2]}{\omega} \right| \le C$$

provided $\omega > 0$, $0 < \alpha, \beta < 1$, $\alpha + \beta < 1$, and $E(h_0^2) < \infty$.

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., the Bartlett, Parzen, Tukey, and truncated kernels), $b = \infty$. For the 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 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.,* $\hat{M}(p) \xrightarrow{d} N(0,1)$ *as* $n \to \infty$.

Throughout, all the proofs are collected in the Appendix. In the proof of Theorem 1, we find that the use of any $n^{1/2}$-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).

4. CONSISTENCY

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

THEOREM 2. *Suppose Assumptions A.1–A.7 hold and* $p = cn^{\lambda}$ *for* $\lambda \in (0,1)$ *and* $c \in (0,\infty)$. *Then as* $n \to \infty$,

$$(p^{1/2}/n)\hat{M}(p) \xrightarrow{p} \left[2D_0 \int_0^{\infty} k^4(z)\,dz\right]^{-1/2} \times \pi \iint_{-\pi}^{\pi} |f(\omega,u,v) - f_0(\omega,u,v)|^2\,d\omega\,dW(u)\,dW(v)$$
$$= \left[2D_0 \int_0^{\infty} k^4(z)\,dz\right]^{-1/2} \sum_{j=1}^{\infty} \int |\sigma_j(u,v)|^2\,dW(u)\,dW(v). \quad \textbf{(4.1)}$$

https://doi.org/10.1017/S0266466603196089 Published online by Cambridge University Press

Suppose e_t and e_{t-j} are not independent at some lag $j > 0$. Then $\int |\sigma_j(u,v)|^2 \, dW(u) \, dW(v) > 0$ for any weighting 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 $\nu \geq 5$. Thus, we expect that $\hat{M}(p)$ has relatively omnibus power against a wide variety of alternatives. Because the L_2-norm in (4.1) is positive whenever there exists pairwise serial dependence at any nonzero lag, $\hat{M}(p)$ is an asymptotically one-sided $N(0,1)$ test. Upper-tailed asymptotic critical values (e.g., 1.645 at the 5% level) 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 can render BDS(m,d) inconsistent against some alternatives. In contrast, any $W(\cdot)$ that is positive, monotonically increasing, and continuous with unbounded support on $\mathbb{R}$ always ensures consistency of $\hat{M}(p)$ against any type of pairwise dependence across various lags in $\{e_t\}$. 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 that follows. Our results show that a variety of choices of $W(\cdot)$ have little impact on the level and power of $\hat{M}(p)$, whereas the choice of d has significant impact on the level 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), 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 exists only first-order serial dependence in $\{e_t\}$ and the joint probability density of (e_t, e_{t-1}) is

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

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 density, we make the following assumption.

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

Note that $f_0(\cdot)$ is the marginal density of $\{e_t\}$ when $\{e_t\}$ is i.i.d. and $g(\cdot,\cdot)$ characterizes the type of serial dependence in $\{e_t\}$. The condition on $r_n(\cdot,\cdot)$

https://doi.org/10.1017/S0266466603196089 Published online by Cambridge University Press

ensures that the remainder $r_n(\cdot,\cdot)$ has no impact on the limit distribution of $\hat{M}(p)$ under $\mathbb{H}_n(a_n)$. 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 MA conditional heteroskedastic (MACH(1); Yang and Bewley, 1995) 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 $g(x,y) = (x^2 - 1) \times (y^2 - 1)$ for (5.3).

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

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

Whenever $g(x,y) \neq 0$, 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 $\mathbb{H}_n(p^{1/4}/n^{1/2})$. The rate $p^{1/4}/n^{1/2}$ is slower than $n^{-1/2}$, because $p \to \infty$ as $n \to \infty$. This is the price one has to pay to achieve consistency against any type of pairwise serial dependence in $\{e_t\}$. However, it is faster than $n^{-1/4}$ given $p/n \to 0$. If $p \propto \log(n)$, then $p^{1/4}/n^{1/2} \propto n^{-1/2} \log^{1/4}(n)$, which is nearly the same as $n^{-1/2}$. If $p \propto n^{1/5}$, as is the case with the data-driven method described subsequently for some commonly used kernels, $p^{1/4}/n^{1/2} \propto n^{-1/2+1/20}$, which is only slightly slower than $n^{-1/2}$. 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 $\hat{M}(p)$ (see Theorem A.3 in the Appendix), so the conclusion of Theorem 3 also applies to the tests considered in Hong (1999), where no local power analysis was given.

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 $g_l: \mathbb{R} \to \mathbb{R}$ such that $\int g_l(x) f_0(x)\, dx = 0$, $l = 1,2$. We find that BDS$(2,d)$ has nontrivial power under $\mathbb{H}_n(a_n)$ if the limit noncentrality

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

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

https://doi.org/10.1017/S0266466603196089 Published online by Cambridge University Press

$$
\begin{aligned}
C_2(d) - C_1(d)^2 = & \iiiint \mathbf{1}(|x - x'| < d)\mathbf{1}(|y - y'| < d) \\
& \times f_{1n}(x, y) f_{1n}(x', y')\, dx\, dx'\, dy\, dy' \\
& - \left[\iint \mathbf{1}(|x - y| < d) f_{0n}(x) f_{0n}(y)\, dx\, dy \right]^2 \\
= & \; 2a_n \iint \mathbf{1}(|x - y| < d) g_1(x) f_0(x) f_0(y)\, dx\, dy \\
& \times \iint \mathbf{1}(|x - y| < d) g_2(y) f_0(x) f_0(y)\, dx\, dy \\
& + a_n^2 \left[\iint \mathbf{1}(|x - y| < d) g_1(x) g_2(y) f_0(x) f_0(y)\, dx\, dy \right]^2 \\
& + o(a_n^2), \qquad \textbf{(5.5)}
\end{aligned}
$$

where $f_{0n}(\cdot)$ denotes 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 $\mathbb{H}_n(n^{-1/4})$. This occurs, e.g., when the marginal density $f_0(\cdot)$ is uniform.[8] Alternatively, suppose that $f_0(x) = (2\pi)^{-1/2} \exp(-\frac{1}{2}x^2)$, and $g_l(\cdot)$ is an odd function for $l = 1$ or 2; i.e.,

$$
g_l(-x) = -g_l(x) \quad \text{for all } x \in \mathbb{R}, \quad \text{and for } l = 1 \text{ or } 2. \qquad \textbf{(5.6)}
$$

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

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

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 MACH(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,

https://doi.org/10.1017/S0266466603196089 Published online by Cambridge University Press

BDS$(2,d)$ can detect MACH(1) with rate $n^{-1/2}$ but MA(1) with rate $n^{-1/4}$ 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). Note that a local MACH(1) is equivalent to a local ARCH(1).

6. CHOICE OF DATA-DRIVEN BANDWIDTH

Both BDS(m,d) and $\hat{M}(p)$ involve the choice of lag order $m-1$ or p. Brock et al. (1991), based on simulation experiments, recommend some simple rule of thumb that m be small for finite sample sizes. Our generalized spectral smoothing 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 Assumptions A.1–A.7 and A.9 hold and $\hat{p}$ is a data-driven bandwidth such that $\hat{p}/p=1+O_P(p^{-((3/2)\beta-1)})$ 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., $\hat{M}(\hat{p})-\hat{M}(p)\xrightarrow{p}0$ and $\hat{M}(\hat{p})\xrightarrow{d}N(0,1)$ as $n\to\infty$.*

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 the results of Hong (1999) to the estimated 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 minimizes an asymptotic integrated mean square error (IMSE) criterion for the estimator $\hat{f}_n(\cdot,\cdot,\cdot)$ in (2.7). This method is described as follows. Consider the "pilot" estimators based on a preliminary bandwidth $\bar{p}$:

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

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

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 the Bartlett kernel whereas $k(\cdot)$ is the Daniell kernel. Note that $\bar{f}_n(\cdot,\cdot,\cdot)$ is an estimator for $f(\cdot,\cdot,\cdot)$ and $\bar{f}_n^{(q)}(\cdot,\cdot,\cdot)$ is an estimator for the generalized spectral derivative

https://doi.org/10.1017/S0266466603196089 Published online by Cambridge University Press

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

The plug-in bandwidth is then defined as

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

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

$$\begin{aligned}\hat{c}_0 &= \left[\frac{2q(k^{(q)})^2}{\int_{-\infty}^{\infty}k^2(z)\,dz}\,\frac{\iint_{-\pi}^{\pi}|\bar{f}^{(q)}(\omega,u,v)|^2\,d\omega\,dW(u)\,dW(v)}{\int_{-\pi}^{\pi}\left[\int\bar{f}_n(\omega,u,-u)\,dW(u)\right]^2 d\omega}\right]^{1/(2q+1)} \\ &= \left[\frac{2q(k^{(q)})^2}{\int_{-\infty}^{\infty}k^2(z)\,dz}\,\frac{\sum_{j=1-n}^{n-1}(n-|j|)\bar{k}^2(j/\bar{p})|j|^{2q}\int|\hat{\sigma}_j(u,v)|^2\,dW(u)\,dW(v)}{\sum_{j=1-n}^{n-1}(n-|j|)\bar{k}^2(j/\bar{p})\left[\int\hat{\sigma}_j(u,-u)\,dW(u)\right]^2}\right]^{1/(2q+1)}.\end{aligned} \tag{6.5}$$

Note that 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 either can be fixed or can grow with the sample size n. If $\bar{p}$ is fixed, $\hat{p}_0$ still grows at rate $n^{1/(2q+1)}$ in general, but $\hat{c}_0$ does not converge to the optimal tuning constant. This is analogous in spirit to a parametric plug-in method. Hong (1999) shows that when $\bar{p}$ grows with n properly, the data-driven bandwidth $\hat{p}_0$ in (6.4) minimizes an asymptotic IMSE of $\hat{f}_n(\cdot,\cdot,\cdot)$.[9] 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 subsequent simulation.

7. MONTE CARLO EVIDENCE

We now compare the level and power of $\hat{M}(\hat{p}_0)$, BDS(m,d), the Box–Pierce–Ljung test BPL(p), the McLeod–Li test ML(p), and the Li–Mak test LM(p,r), in finite samples. We check adequacy of two basic time series models—AR(r) and ARCH(r)—for $r = 1,4$, respectively. With the null AR(1) model, we examine the level of the tests, their power against a variety of neglected dynamics and 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 level of the tests, their power against misspecification in conditional variance, their power to distinguish

https://doi.org/10.1017/S0266466603196089 Published online by Cambridge University Press

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). Finally, the null models AR(4) and ARCH(4) allow us to examine the impact of parameter estimation uncertainty on the level of the proposed tests when the parameter dimension increases. Such impact is asymptotically negligible but might be significant in finite samples.

7.1. Testing Conditional Mean Model

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

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

under each of the following 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 Regime-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 a latent state variable that follows a two-state Markov chain with transition probabilities $P(S_t = 1|S_{t-1} = 0) = P(S_t = 0|S_{t-1} = 1) = 0.3$.

https://doi.org/10.1017/S0266466603196089 Published online by Cambridge University Press

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)

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

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

In DGPs A.0–A.7, $\{\varepsilon_t\}$ is i.i.d.(0,1). DGPs A.1 and 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. DGPs A.3 and A.4 are not invertible but they are second-order stationary, as shown in Granger and Andersen (1978, pp. 90–91). The SIGN model examined in Granger and Teräsvirta (1999), DGP A.7, is a first-order nonlinear autoregressive process but has the same autocorrelation function as an AR(1) process: $\rho(j) = (1 - 2q)^{|j|}$, where $q = P(\sigma\varepsilon_t < -1) = P(\sigma\varepsilon_t > 1)$ when ε_t is symmetric. Following Granger and Teräsvirta (1999), we choose $\sigma = 0.43$ so that $q = 0.01$ if ε_t is $N(0,1)$. The tent map, DGP A.8, is a deterministic chaotic process, but it resembles in autocorrelation an AR(1) process with the AR coefficient $2\alpha - 1$ (see Sakai and Tokumaru, 1980). The 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 a linear process in terms of autocorrelation.

7.2. Testing Conditional Variance Model

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

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

when Y_t is generated from the following generating processes:

DGP B.0 (ARCH(1))

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

DGP B.1 (ARCH(2))

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

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

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

https://doi.org/10.1017/S0266466603196089 Published online by Cambridge University Press

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

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

DGP B.4 (Stochastic Volatility)

$$Y_t = h_t \varepsilon_t, \qquad h_t^2 = 0.1 Y_{t-1}^2 + \exp(\varphi \ln h_{t-1}^2 + v_t),$$

$$\{v_t\} \sim \text{i.i.d. } N(0,1), \qquad \varphi = 0.98.$$

DGP B.5 (Bilinear)

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

DGP B.6 (TAR)

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

DGP B.7 (Nonlinear MA)

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

DGP B.8 (Logistic Map)

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

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

In DGPs B.0–B.7, $\{\varepsilon_t\}$ is i.i.d.(0,1). The DGPs B.1–B.4 are used to examine the power of the tests against misspecification in conditional variance. In DGP B.4, parameter value $\varphi = 0.98$ is empirically relevant; Harvey, Ruiz, and Shephard (1994) obtain estimates of φ in range of 0.9575–0.9948 for four different daily foreign exchange rates. The 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 ARCH(1) (e.g., Granger and Teräsvirta, 1993, p. 34). It is used to examine the power of a test to distinguish ARCH from a chaotic process with similar autocorrelations in squares. From the results of He and Teräsvirta (1999), we note that $\{Y_t\}$ does not have finite fourth moments under DGP B.1 even when $\{\varepsilon_t\}$ is i.i.d.$N(0,1)$. However, $\{Y_t\}$ has all finite moments under DGP B.3 when $\{\varepsilon_t\}$ is i.i.d.$N(0,1)$, but $\text{var}(Y_t)$ is infinite when $\{\varepsilon_t\}$ is i.i.d.t_5 (see Nelson, 1991). We conjecture that $\text{var}(Y_t)$ is also infinite when $\{\varepsilon_t\}$ is an i.i.d. sequence of mixed normals. Note that strict stationarity holds under both DGPs B.1 and B.3. On the other hand, DGPs B.5 and B.7 are not invertible, but they are second-order stationary (see Granger and Andersen, 1978, pp. 90–91).

https://doi.org/10.1017/S0266466603196089 Published online by Cambridge University Press

7.3. Testing Higher Order Models

We now examine the impact of parameter estimation uncertainty on the level of the proposed tests. We examine how the level of the tests is affected by increasing the number of estimated parameters. Such impact is asymptotically negligible but might be significant in finite samples. We consider the following two higher order models under the null hypothesis.

First, we consider the adequacy of an AR(4) model for the conditional mean:

$$\text{Model C:} \qquad Y_t = a + \sum_{j=1}^{4} b_j Y_{t-j} + e_t, \qquad t = 1, \ldots, n,$$

under the following DGP:

DGP C.0 (AR(4))

$$Y_t = 0.9[0.4Y_{t-1} + 0.3Y_{t-2} + 0.2Y_{t-3} + 0.1Y_{t-4}] + \varepsilon_t, \qquad \{\varepsilon_t\} \sim \text{i.i.d.}(0,1).$$

Next, we consider the adequacy of an ARCH(4) model for the conditional variance:

$$\text{Model D:} \qquad Y_t = h_t e_t, \qquad h_t^2 = a + \sum_{i=1}^{4} b_i Y_{t-i}^2,$$
$$\{e_t\} \sim \text{i.i.d.}(0,1), \qquad t = 1, \ldots, n,$$

when Y_t is generated from the following DGP:

DGP D.0 (ARCH(4))

$$Y_t = h_t \varepsilon_t, \qquad h_t^2 = 0.1 + 0.9[0.4Y_{t-1}^2 + 0.3Y_{t-2}^2 + 0.2Y_{t-3}^2 + 0.1Y_{t-4}^2],$$
$$\{\varepsilon_t\} \sim \text{i.i.d.}(0,1).$$

For all the DGPs except the chaotic processes A.8 and B.8, we use the GAUSS Windows version random number generator to generate i.i.d. innovations $\{\varepsilon_t\}$ 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 $n + 1{,}000$ observations for $\{\varepsilon_t\}$ under each of the distributions (i)–(iv) and then discard the first 1,000 to alleviate the impact of using some initial values. We report the levels of the tests under all four error distributions, but for space we report the power under the normal error only.

7.4. Monte Carlo Evidence

To compute the test statistic $\hat{M}(\hat{p}_0)$, BDS(m,d), BPL(p), ML(p), and LM(p,r), we use the usual residual series $\hat{e}_t = Y_t - \hat{a} - \hat{b}Y_{t-1}$ from Model A and $\hat{e}_t = Y_t - \hat{a} - \sum_{j=1}^{4} \hat{b}_j Y_{t-j}$ from Model C, estimated by the ordinary least

https://doi.org/10.1017/S0266466603196089 Published online by Cambridge University Press

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 and $\hat{h}_t^2 = \hat{a} + \sum_{j=1}^4 \hat{b}_j Y_{t-j}^2$ from Model D, estimated by the quasi–maximum likelihood method with a Gaussian likelihood function.

For the generalized spectral test $\hat{M}(\hat{p}_0)$, we use Daniell kernel $k(z) = \sin(\pi z)/\pi z$, which enjoys the optimal power property over a class of kernels (Hong, 1999, Theorem 5).[10] To examine the impact of the choice of preliminary bandwidths $\bar{p}$ on the level and power of $\hat{M}(\hat{p}_0)$, we consider $\bar{p} = 1$–10. To investigate the impact of the choice of weight function $W(\cdot)$ on the level and power of $\hat{M}(\hat{p}_0)$, we consider the three distribution functions: (i) $N(0,1)$, (ii) double exponential, and (iii) 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 0.5σ–1.5σ and m in range 2–5, for $n = 500$–1,000, where $\sigma^2 = \text{var}(Y_t)$. To examine the impact of the choice of embedding dimension m on the level and power of BDS(m,d), we use $m = 2$–11, which is equivalent to the choice of a lag order p from 1 to 10. As some DGP 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 2σ, σ, and 0.5σ, respectively.

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

For the McLeod–Li test ML(p), which is suitable to test AR(r) models, we use $p = 1$–10 for Model A. For the Li–Mak test LM(p,r), which is suitable to test ARCH(r) models, $p = 2$–10 for Model B (similarly to BPL(p) for Model A, $p = 1$ cannot be chosen here for LM(p,r) when $r = 1$).

To examine the levels 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 empirical level at the 10%, 5%, and 1% significance levels for $n = 100$ and 200, using asymptotic critical values and 1,000 Monte Carlo iterations. To conserve space, Figures 1 and 2 only report the levels of the tests at the 5% level for $n = 100$. Similarly, to examine the levels of the tests under the null AR(4) model and under the null ARCH(4) model, we estimate Model C and Model D under DGP C.0 and DGP D.0, respectively, and report the results in Figures 1 and 2 also.[11]

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 level-adjusted by using the empirical critical values obtained under DGPs A.0 and B.0, respectively, 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 1,000 replications.

https://doi.org/10.1017/S0266466603196089 Published online by Cambridge University Press

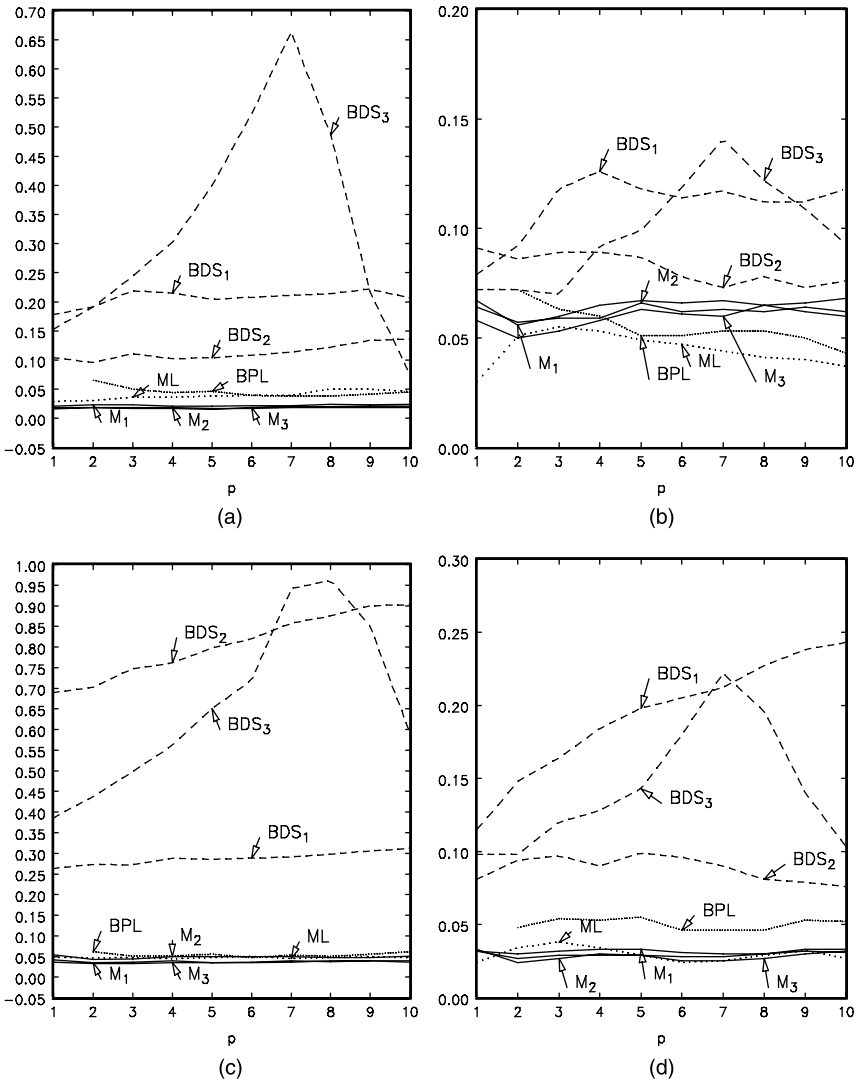


FIGURE 1. Size of testing conditional mean at 5% level—(a) DGP A.0: AR(1)–$N(0,1)$, (b) DGP A.0: AR(1)–exponential, (c) DGP A.0: AR(1)–mixed normal, (d) DGP A.0: AR(1)–t_5.

In Figures 1–4, the levels or powers of $\hat{M}(\hat{p}_0)$, BDS(m,d), BPL(p), ML(p), and LM(p,r) are plotted as functions of $\bar{p}$, $m-1$, p, 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

https://doi.org/10.1017/S0266466603196089 Published online by Cambridge University Press

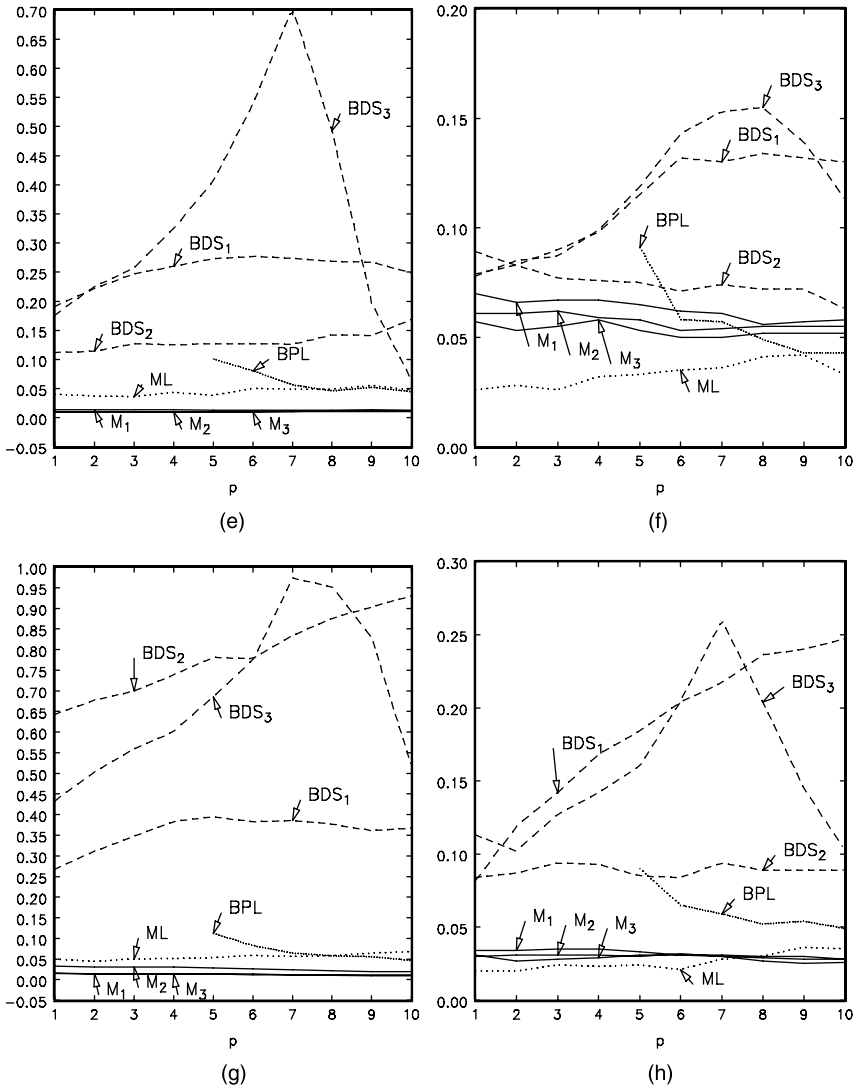


FIGURE 1. Size of testing conditional mean at 5% level—(e) DGP C.0: AR(4)–$N(0,1)$, (f) DGP C.0: AR(4)–exponential, (g) DGP C.0: AR(4)–mixed normal, (h) DGP C.0: AR(4)–t_5.

three plots for BDS(m,d) in dashed lines, denoted as BDS$_l$ ($l = 1,2,3$), that correspond to three distance parameter values—$d = 0.5^l$. Moreover, ML(p) or LM(p,r) is plotted in dotted lines, and BPL(p) is plotted with more closely spaced dots. The test ML(p) is reported in Figures 1 and 3, where the usual

https://doi.org/10.1017/S0266466603196089 Published online by Cambridge University Press

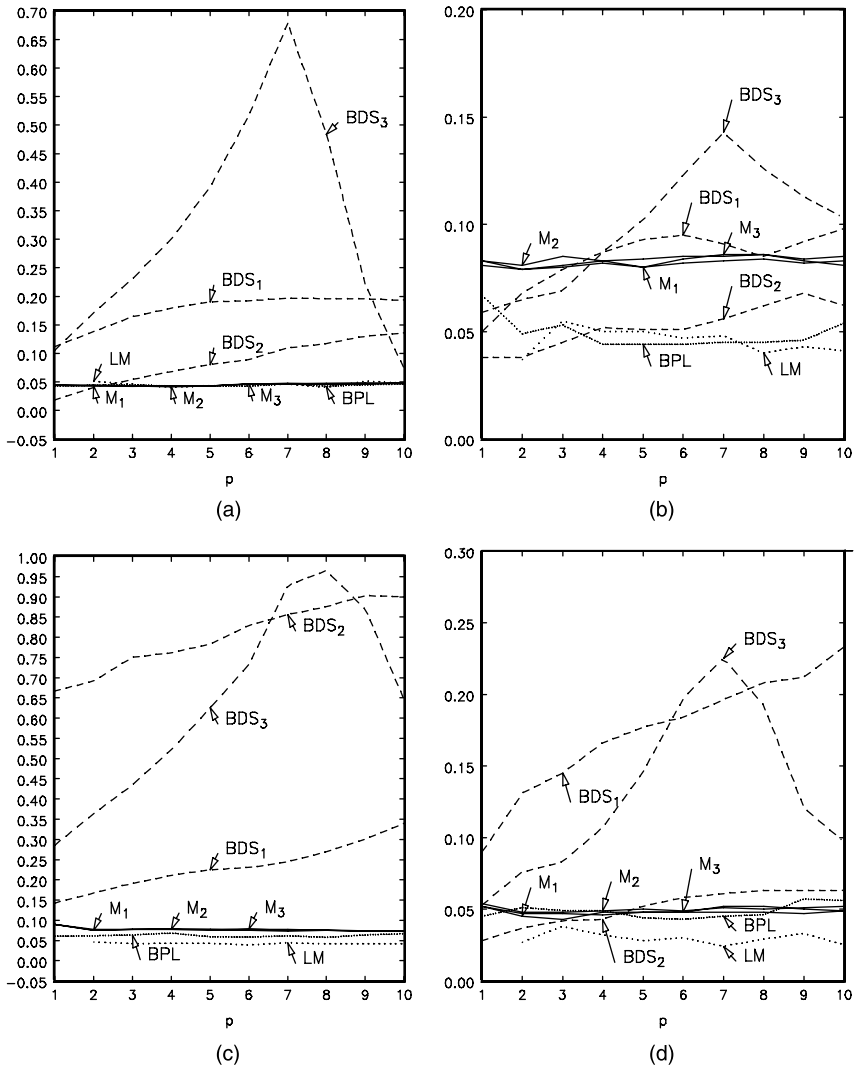


FIGURE 2. Size of testing conditional variance at 5% level—(a) DGP B.0: ARCH(1)–$N(0,1)$, (b) DGP B.0: ARCH(1)–exponential, (c) DGP B.0: ARCH(1)–mixed normal, (d) DGP B.0: ARCH(1)–t_5.

estimated residuals are used, and LM(p,r) is reported in Figures 2 and 4, where the standardized estimated residuals are used.

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

1. Overall, the levels of the tests $\hat{M}(\hat{p}_0)$, BPL(p), and ML(p) under the null AR(1) model and the levels of the tests $\hat{M}(\hat{p}_0)$, BPL(p), and LM(p,r) under the null

https://doi.org/10.1017/S0266466603196089 Published online by Cambridge University Press

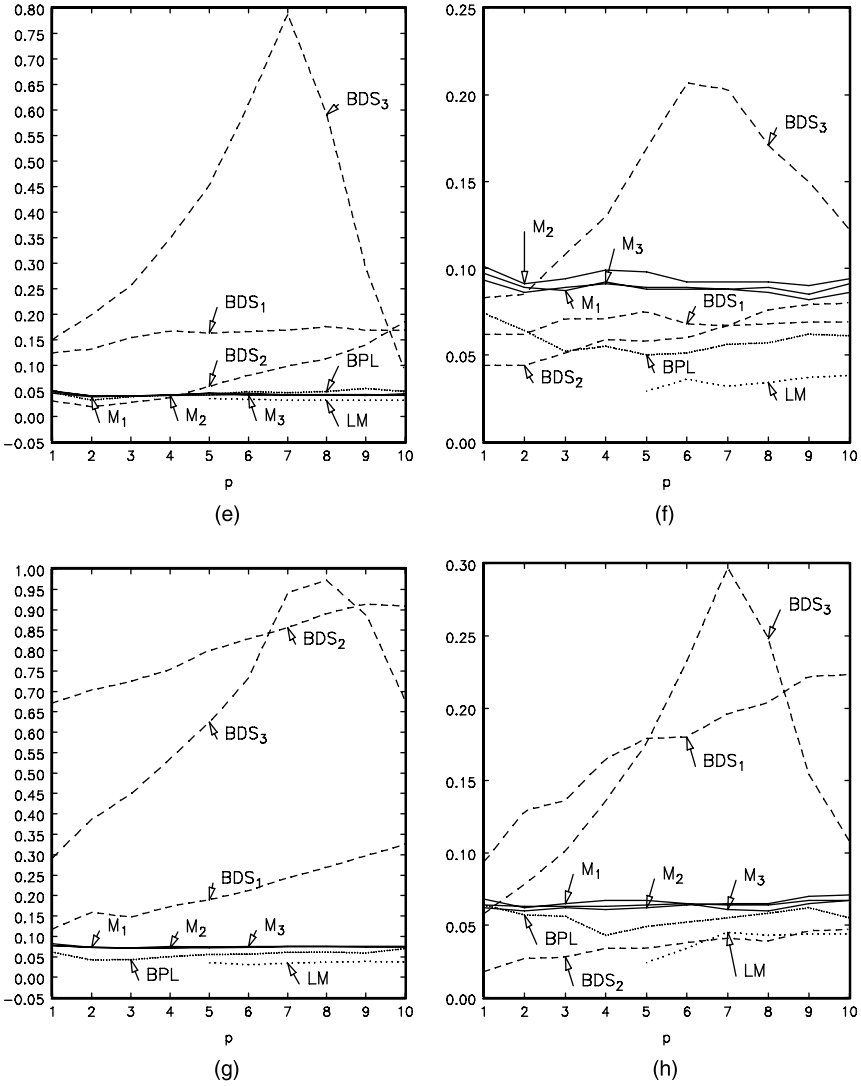


FIGURE 2. Size of testing conditional variance at 5% level—(e) DGP D.0: ARCH(4)–$N(0,1)$, (f) DGP D.0: ARCH(4)–exponential, (g) DGP D.0: ARCH(4)–mixed normal, (h) DGP D.0: ARCH(4)–t_5.

ARCH(1) model are more or less reasonable, whereas the level of BDS(m,d) appears not very satisfactory. The unsatisfactory level 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 level of $\hat{M}(\hat{p}_0)$ is robust to the choice of weight function $W(\cdot)$ and preliminary bandwidth $\bar{p}$. The levels of BPL(p) and ML(p)/LM(p,r) are excellent and

https://doi.org/10.1017/S0266466603196089 Published online by Cambridge University Press

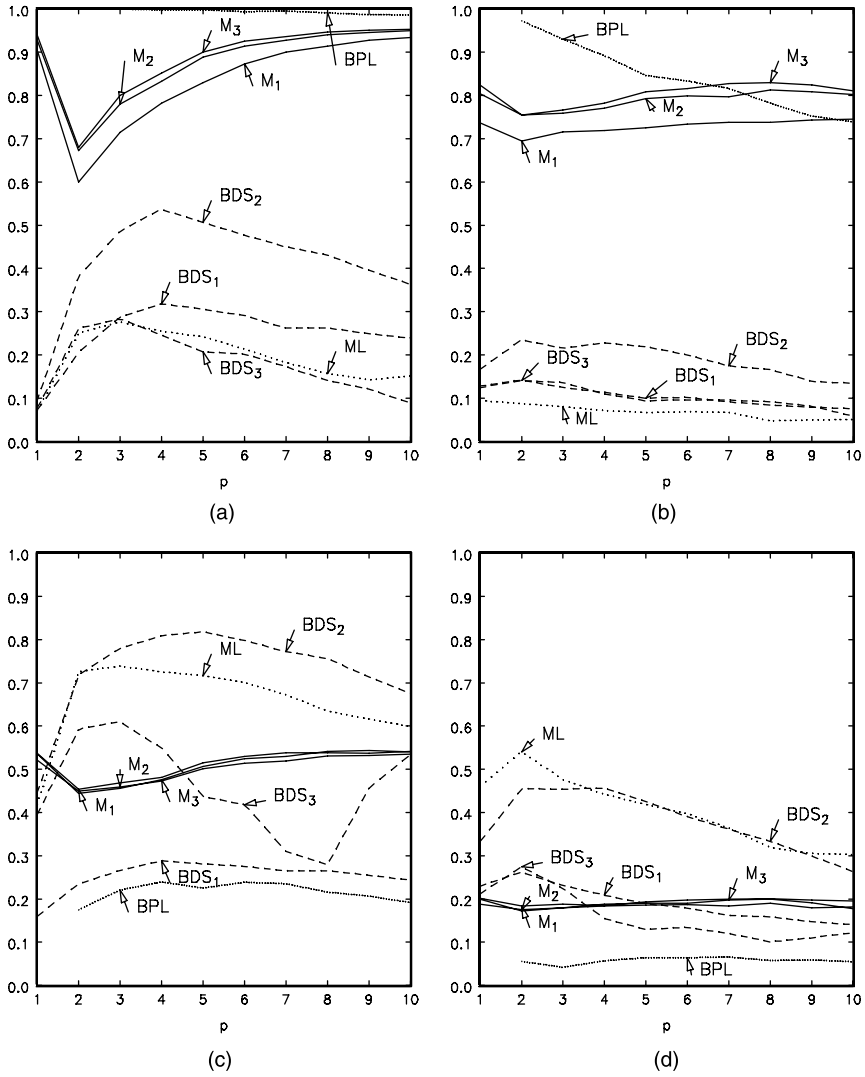


FIGURE 3. Size-corrected power of testing conditional mean at 5% level—(a) DGP A.1: AR(2)–$N(0,1)$, (b) DGP A.2: ARMA(1,1)–$N(0,1)$, (c) DGP A.3: bilinear–$N(0,1)$, (d) DGP A.4: nonlinear MA–$N(0,1)$.

robust to the choice of lag order p. On the other hand, the level 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 whereas $\hat{M}(\hat{p}_0)$ is not sensitive to $\bar{p}$ indicates the practical merit of the data-driven choice of lag order $\hat{p}_0$ for $\hat{M}(\hat{p}_0)$.

3. $\hat{M}(\hat{p}_0)$ displays some (not excessive) underrejection under the null AR(1) with normal errors or some (not excessive) overrejection under the ARCH(1) model

https://doi.org/10.1017/S0266466603196089 Published online by Cambridge University Press

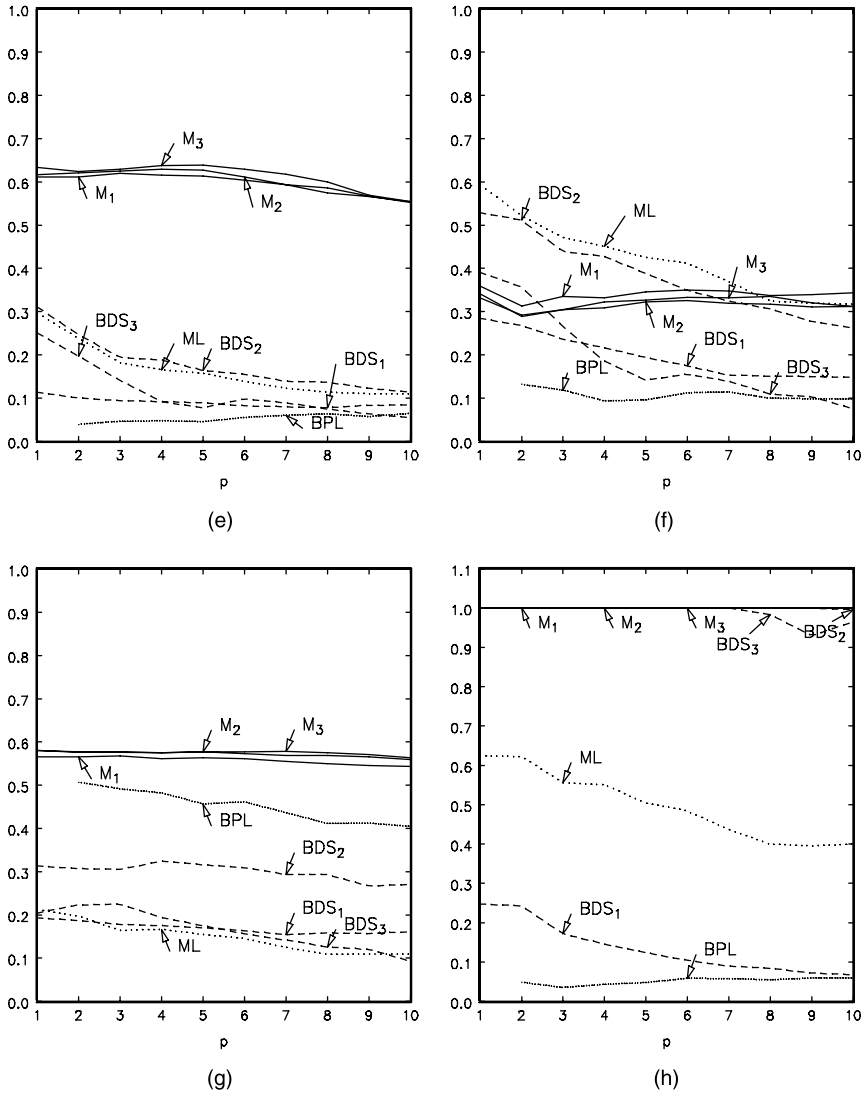


FIGURE 3. Size-corrected power of testing conditional mean at 5% level—(e) DGP A.5: TAR–$N(0,1)$, (f) DGP A.6: Markov switching–$N(0,1)$, (g) DGP A.7: SIGN–$N(0,1)$, (h) DGP A.8: tent map.

with exponential and mixed normal errors. On the other hand, the level 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 level patterns of each test under the null AR(1) model and the null ARCH(1) model are more or less similar.

https://doi.org/10.1017/S0266466603196089 Published online by Cambridge University Press

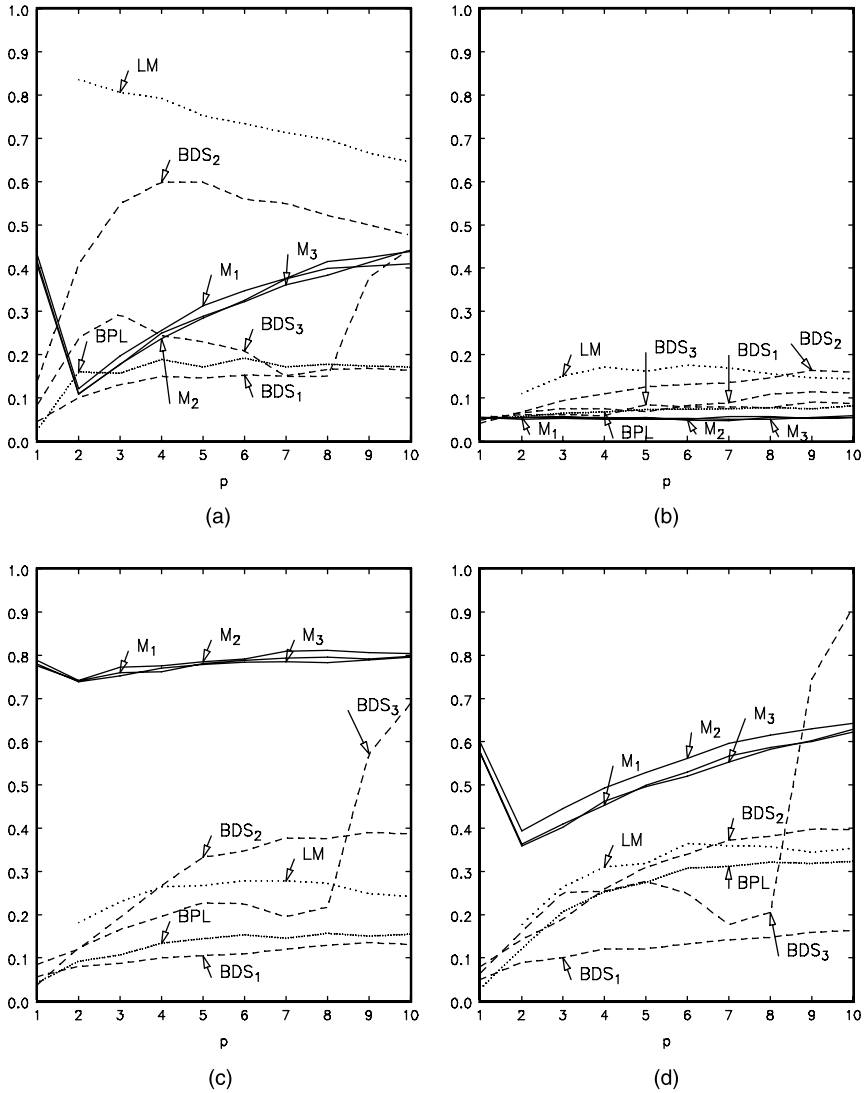


FIGURE 4. Size-corrected power of testing conditional variance at 5% level—(a) DGP B.1: ARCH(2)–$N(0,1)$, (b) DGP B.2: GARCH(1,1)–$N(0,1)$, (c) DGP B.3: EGARCH–$N(0,1)$, (d) DGP B.4: stochastic volatility–$N(0,1)$.

5. The preceding four observed level patterns for the null models of AR(1) and ARCH(1) carry over to the higher order null models of AR(4) and ARCH(4). This indicates that model parameter estimation uncertainty does not affect the level of the tests at least for the models and sample sizes considered.

https://doi.org/10.1017/S0266466603196089 Published online by Cambridge University Press

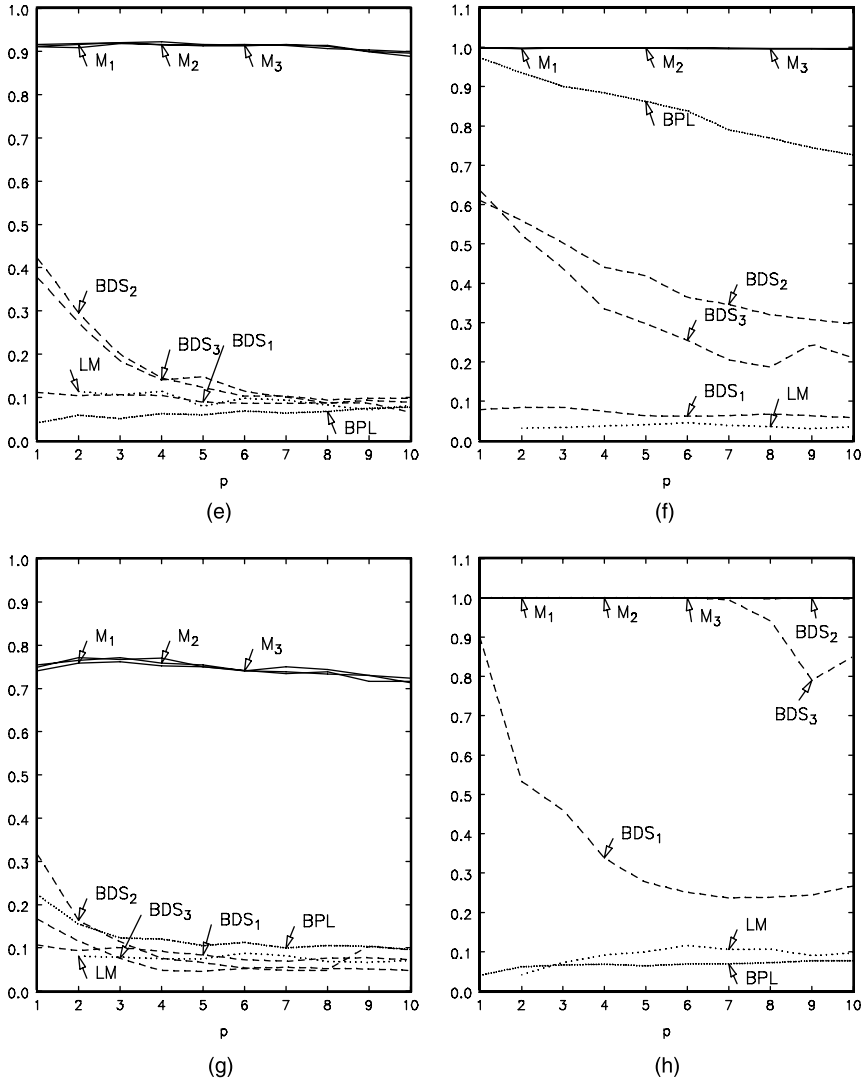


FIGURE 4. Size-corrected power of testing conditional variance at 5% level—(e) DGP B.5: bilinear–$N(0,1)$, (f) DGP B.6: TAR–$N(0,1)$, (g) DGP B.7: nonlinear MA–$N(0,1)$, (h) DGP B.8: logistic map.

In Figures 1 and 2 we consider the levels of the tests under an AR coefficient of 0.6 for DGP A.0 and the levels of the tests with an ARCH coefficient of 0.1 for DGP B.0. We have also experimented (not reported) with 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.

https://doi.org/10.1017/S0266466603196089 Published online by Cambridge University Press

We note that the fourth moment of Y_t does not exist when the ARCH(1) coefficient value is larger than or equal to $1/\sqrt{3} \simeq 0.577$ for DGP B.0. In most scenarios, the levels of $\hat{M}(\hat{p}_0)$, BDS(m,d), BPL(p), and ML(p)/LM(p,r) are generally robust to the values of the AR(1) and ARCH(1) coefficients in DGP A.0 and DGP B.0. One 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 level becomes reasonable for larger lag orders p, say, $p > 5$.

We now examine the powers of the tests. To be fair in comparison and to take into account the BDS(m,d)'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 misspecifications 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) seems sensitive to the choice of the embedding dimension m, which is equivalent to the choice of a lag order. The tests BPL(p) and ML(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 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, and Markov regime-switching. It also cannot distinguish AR(1) from the tent map, which resembles an AR(1) process in autocorrelation but is completely deterministic.
   (b) The correlation in squares test, ML(p), has good power against bilinear, nonlinear MA, Markov regime-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 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. And it has moderate power for Markov regime-switching and low power for nonlinear MA alternatives.
4. Overall, $\hat{M}(\hat{p}_0)$ has reasonable 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 examine the powers of the tests against various misspecifications of the ARCH(1) model, as reported in Figure 4. We observe the following patterns.

https://doi.org/10.1017/S0266466603196089 Published online by Cambridge University Press

1. As in testing misspecifications in conditional mean, the power of $\hat{M}(\hat{p}_0)$ is 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 LM(p,r) 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 (the Li–Mak test), LM(p,r), is most powerful against ARCH(2), for which it has good power by its design. Nevertheless, LM(p,r) 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 document 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 predictability in economics and finance (e.g., 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) model. It has high power to distinguish ARCH(1) from bilinear, TAR, nonlinear MA, and logistic map processes. Interestingly, $\hat{M}(\hat{p}_0)$ has better power than BDS(m,d) against the logistic map alternative.
4. Overall, 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. EMPIRICAL APPLICATION

To further highlight the merits of our generalized spectral test, we now apply it to evaluate an empirical financial time series model. As is well known, practical model-based financial decision making such as hedging, risk management, and option pricing will be satisfactory only if it builds on reasonable specification of the underlying asset price processes. In an important contribution, Andersen, Benzoni, and Lund (2002) use efficient method of moments (EMM) of Gallant and Tauchen (1996) to evaluate the adequacy of a variety of continuous-time parametric models for the daily S&P 500 equity index and the impact of different specifications on option pricing. The EMM method is based

https://doi.org/10.1017/S0266466603196089 Published online by Cambridge University Press

on the expectation of the score function of a discrete-time auxiliary semi-nonparametric model that adequately approximates the conditional distribution of the discretely sampled data in an asymptotic sense. An attractive feature of EMM is that it is generally applicable and achieves the same efficiency as the maximum likelihood estimator when the score function of the auxiliary model asymptotically spans the score function of the true conditional distribution. Moreover, the EMM criterion function can be used to construct an overall test of the overidentifying restrictions on the parametric model to be tested, and the fit of individual scores can be used to gauge how well the parametric model captures particular features of data.

In their EMM applications to the S&P 500 index, Anderson et al. (2002), based on some model selection criteria, choose the following auxiliary model for the daily S&P 500 price changes:

$$Y_t = \phi_0 + \phi_1 u_{t-1} + u_t, \qquad u_t = h_t e_t, \qquad \{e_t\} \sim \text{i.i.d. } (0,1), \qquad t = 1, \ldots, n,$$

$$\ln h_t^2 = \omega + \sum_{j=1}^{p} \beta_j \ln h_{t-j}^2 + \left(1 + \sum_{j=1}^{q} \alpha_j L^j\right)[\theta_1 e_{t-1} + \theta_2(|e_{t-1}| - E|e_{t-1}|)],$$

where the density of $\{e_t\}$ is approximated by Hermite polynomials. The i.i.d. property for $\{e_t\}$ is obtained as a consequence of using some model selection criteria. This model is referred to as MA(1)-EGARCH(p,q), as in the work of Andersen et al. (2002), who estimate the model in two steps: as the first-order autocorrelation in $\{Y_t\}$ is largely "artificial," induced by nonsynchronous trading effects, they filter out this effect before estimating the conditional variance.[12] Here, we adopt the same conditional mean model—MA(1)—but we estimate it jointly with the conditional variance model via the quasi–maximum likelihood method. We maximize the log likelihood for the generalized error distribution, described in Nelson (1991). Table 1 reports our quasi–maximum likelihood estimation of the MA-EGARCH models for both the whole sample (1953–1996) and the subsample (1980–1996). For comparison, we also include the estimation results in Anderson et al. (2002).

Anderson et al. (2002) find that an MA(1)-EGARCH(1,1) model can adequately capture the full serial dependence in the daily S&P 500 equity index changes from 01/02/1953 to 12/31/1996 (with $n = 11{,}076$), and an MA(1)-EGARCH(2,1) model can adequately capture the full serial dependence in daily changes of the S&P 500 index from 01/02/1980 to 12/31/1996 (with $n = 4{,}298$). In both cases, the density of the i.i.d. error e_t is approximated by a Hermite polynomial expansion.

The adequacy of the full dynamics of the auxiliary semi-nonparametric model for $\{Y_t\}$ is crucial for the efficiency of the EMM estimator and the validity of the related EMM diagnostic tests. Although a semi-nonparametric model is asymptotically free of model misspecification from a theoretical point of view, the use of some model selection criteria might lead to a misspecified model in practice. It is therefore important to check if the aforementioned MA(1)-

https://doi.org/10.1017/S0266466603196089 Published online by Cambridge University Press

https://doi.org/10.1017/S0266466603196089 Published online by Cambridge University Press

TABLE 1. Estimates of MA(1)-EGARCH models for S&P 500 index changes

	01/02/1953–12/31/1996 (n = 11,076) MA(1)-EGARCH(1,1)				01/03/1980–12/31/1996 (n = 4,298) MA(1)-EGARCH(2,1)			
	This paper		Andersen et al.		This paper		Andersen et al	
Parameter	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.
ϕ_0	0.0357	0.0063	0.0331	0.0142	0.0421	0.0109	−0.0398	0.0398
ϕ_1	0.1447	0.0116	—	—	0.0318	0.0151	—	—
ω	−0.0084	0.0019	4.3769	1.1249	−0.0090	0.0041	2.6724	0.4565
α_1	−0.4750	0.0666	−0.4391	0.0635	0.5537	0.2479	0.5712	0.1559
β_1	0.9887	0.0023	0.9893	0.0022	0.0508	0.0277	0.0388	0.0199
β_2	N.A.	N.A.	N.A.	N.A.	0.9268	0.0269	0.9213	0.0183
θ_1	−0.0997	0.0137	−0.1581	0.0195	−0.0472	0.0194	−0.0916	0.0234
θ_2	0.2259	0.0290	0.2973	0.0280	0.1268	0.0329	0.2011	0.0340
ν	1.3750	0.0446	N.A.	N.A.	1.2793	0.0569	N.A.	N.A.
Mean L	−1.0616				−1.1949			

Note: The results of Andersen et al. (2002) are taken from their Table II for the whole sample (1953–1996) and from their Table VII for the subsample (1980–1996). Summary statistics of the data are also provided in Andersen et al. (2002, Table I). They estimate the model in two steps with MA(1) estimated separately from the conditional variance model (instead of jointly) and do not report ϕ_1. In this paper we estimate the mean and variance models jointly. The log-likelihood for the generalized error distribution (GED), normalized to have zero mean and unit variance, and the expression of $E|e_{t-1}|$ can be found in Nelson (1991). Mean$L = n^{-1}\sum_{t=1}^{n} \log L_t$ is the mean log-likelihood, ν is a parameter for the GED distribution, and S.E. is the estimated robust standard error.

EGARCH models indeed completely capture the full serial dependence in the daily S&P 500 index data. Anderson et al. (2002) note that the standardized estimated residuals of these models pass Box–Pierce–Ljung type tests. We now check the adequacy of these MA(1)-EGARCH models by using our new procedure. For comparison, we also report the results for BPL(p), LM(p,r), and BDS(m,d) tests. Our results confirm that these models do pass the BLP(p), LM(p,r), and BDS(m,d) tests, but our new test finds very strong evidence of model misspecification missed by the BPL(p), LM(p,r), and BDS(m,d) tests.

Figure 5 reports the p-values of the proposed test $\hat{M}(\hat{p}_0)$, BPL(p), LM(p,r), and BDS(m,d) tests. We report $\hat{M}(\hat{p}_0)$ only with the normal cumulative distribution function (c.d.f.) as the weighting function, denoted as M_1 in Figure 5, because $\hat{M}(\hat{p}_0)$ is robust to the choice of the weight functions. However, we still report BDS(m,d) with three choices of distance parameters $d = 0.5$, 0.25, and 0.125 in unit of data range, denoted as BDS_1, BDS_2, and BDS_3, respectively. As LM(p,r) is computed for the standardized residuals with $h(\hat{I}_{t-1},\hat{\theta})$ an estimated EGARCH model (not ARCH(r)), its asymptotic distribution is not χ^2_{p-r} and is unknown to us. Thus, we report the bootstrap p-values in addition to the asymptotic p-values.[13] The bootstrap p-values are expected to provide more reliable inferences for all the tests.

We first consider the results for the whole sample from 1953 to 1996 (see Figures 5a and 5b). We report asymptotic and bootstrap p-values for M_1, BPL(p), and LM(p,r). We do not report BDS(m,d) for the whole sample, because we use the program, accompanied by the book by Brock et al. (1991), that can handle only up to 7,500 observations. The M_1 test has zero asymptotic and bootstrap p-values. In fact, the M_1 statistic values range from 13.3 to 14.5 when the preliminary lag order $\bar{p}$ changes from 21 to 50 and the Parzen kernel is used, and they range from 10.5 to 14.3 when $\bar{p}$ changes from 21 to 50 and the Bartlett kernel is used. The BPL(p) test has large asymptotic and bootstrap p-values, ranged approximately from 10% to 25%, well above the conventional 5% level. The LM(p,r) test statistic is clearly insignificant in terms of both asymptotic and bootstrap p-values. Because the asymptotic distribution of LM(p,r) is not χ^2_{p-r} when the conditional variance model is not ARCH(r), we rely on the bootstrap p-values in this case. The bootstrap p-values reported in Figure 5 are based on $r = 0$ for LM(p,r), but other values of r give similar bootstrap p-values (not reported). Using BPL(p) and LM(p,r), which have been commonly used in practice as diagnostic tools, we fail to detect any model inadequacy. In contrast, our proposed test indicates that the model does not adequately capture the full serial dependence in the S&P 500 index changes.

Next, we consider the results for the subsample (see Figures 5c and 5d). We now report BDS(m,d) also because the BDS program we use can handle this subsample ($n = 4{,}298$). Again, the M_1 test is the most powerful test: it has zero p-values in terms of both asymptotic and bootstrap p-values. The M_1 statistic values range from 5.9 to 6.6 when the preliminary lag order $\bar{p}$ changes from 21

https://doi.org/10.1017/S0266466603196089 Published online by Cambridge University Press

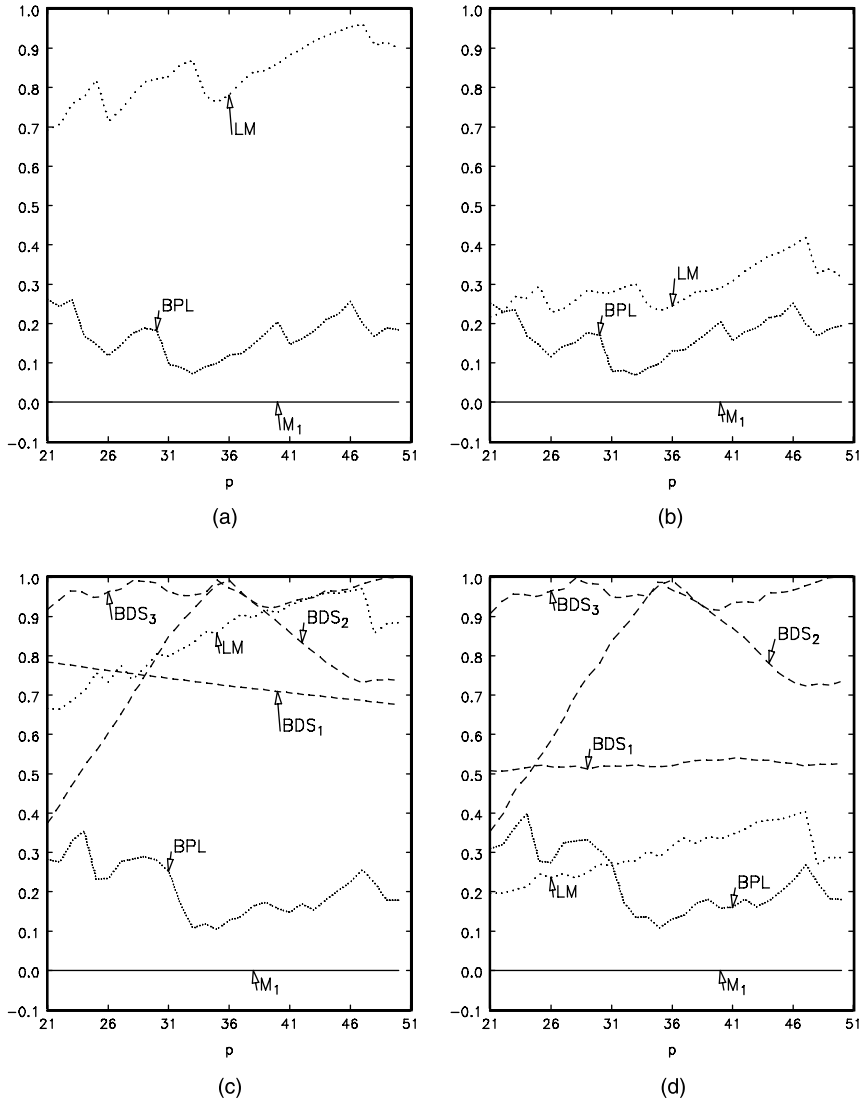


FIGURE 5. Empirical application—(a) Asymptotic *p*-values: 1953–1996, (b) bootstrap *p*-values: 1953–1996, (c) asymptotic *p*-values: 1980–1996, (d) bootstrap *p*-values: 1980–1996.

to 50 and the Parzen kernel is used, and it ranges from 5.2 to 6.5 when $\bar{p}$ changes from 21 to 50 and the Bartlett kernel is used. The BPL test has both asymptotic and bootstrap *p*-values well above 10%, consistent with the results of Anderson et al. (2002). The LM(p, r) test has bootstrap *p*-values above 20%. The

https://doi.org/10.1017/S0266466603196089 Published online by Cambridge University Press

BDS(m,d) tests are sensitive to the choice of distance parameter d. Nevertheless, all three BDS tests fail to reject the MA(1)-EGARCH(2,1) model. For $d = 0.125$, BDS_3 has large p-values close to one in terms of both asymptotic and bootstrap p-values. When $d = 0.25$, BDS_2 is very sensitive to m and has asymptotic and bootstrap p-values in the range 0.35–1.00. When $d = 0.5$, BDS_1 has bootstrap p-values about 0.50.

In summary, the proposed test $\hat{M}(\hat{p}_0)$ firmly rejects the adequacy of MA(1)-EGARCH models for daily S&P 500 index changes, whereas BDS(m,d), BPL(p), and LM(p,r) tests fail to detect any model inadequacy. This implies that care should be taken in interpreting the statistical significance of the EMM test statistics reported in Anderson et al. (2002), because an inadequate auxiliary model affects the efficiency of the EMM estimator and the validity of the EMM diagnostic tests that have used efficient scores to estimate the asymptotic variance-covariance matrices. But we emphasize that the consistency of the EMM estimator is not affected.

9. CONCLUSIONS

The correlation integral–based test by Brock et al. (1991, 1996) has been recently proposed as a portmanteau test for the adequacy of nonlinear time series models. The test has the nice "nuisance-parameter-free" property in the sense that parameter estimation uncertainty 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 the 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, which includes but is not restricted to ARCH and ACD models. It is consistent against any type of pairwise serial dependence across various lags in the model standardized residuals, and it has better asymptotic local power than the BDS test in testing many conditional mean models (but not 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, McLeod–Li, and Li–Mak. The generalized spectral test has omnibus 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 tool kit for time series models and can play a valuable role in evaluating the adequacy of linear and nonlinear time series models. An empirical application to the daily S&P 500 index highlights the merits of our approach.

https://doi.org/10.1017/S0266466603196089 Published online by Cambridge University Press

NOTES

1. A partial list of the papers using this procedure, as given in Tse and Zuo (1997, p. 238), includes Bollerslev (1987), Hsieh (1989), Hamao, Masulis, and Ng (1990), Lee and Tse (1991), Cheung and Ng (1992), Higgins and Bera (1992), and Karolyi (1995).

2. It should be emphasized, as correctly pointed out by one referee, that the BDS test is not a test for chaos against stochastic alternatives. This is the case because the test does not have chaos as the null hypothesis. Instead, the null hypothesis of the BDS test is i.i.d., and when it is rejected, chaos is just one of infinitely many possible reasons for the rejection.

3. The same problem is expected to occur when the BDS test is applied to a fitted ACD model.

4. Robinson (1991) and Pagan (1996) interpret the BDS test from a kernel-based density estimation perspective in which d is equivalent to a fixed bandwidth in a uniform kernel-based probability density estimation.

5. There has been increasing interest in using the characteristic function in time series analysis (e.g., Epps, 1987, 1988; Feuerverger, 1990; Jiang and Knight, 2002; Knight and Satchell, 1997; Knight and Yu, 2002; Pinkse, 1998; Singleton, 2001).

6. In a not unrelated context, Priestley (1988, p. 143) discusses the physical interpretation of frequency ω for nonlinear time series. In our context, the generalized spectrum is essentially the cross spectrum between the transformed variables e^{iue_t} and $e^{ive_{t-j}}$, which measures the correlation between ω-frequency components of e^{iue_t} and $e^{ive_{t-j}}$. Because the exponential function can be written as a sum of polynomials using a Taylor series expansion, it is expected that the generalized spectrum is able to effectively capture cycles and periodicities in higher moments (e.g., volatility clustering cycles) of $\{e_t\}$.

7. We examine via simulation in Section 7 the impact of choosing different weighting functions $W(\cdot)$ on the level and power of the proposed test.

8. Testing for i.i.d. when the marginal distribution of a generalized residual series is uniform has been of interest in the literature of evaluating probability density forecasts (e.g., Clement and Smith, 2000; Diebold et al., 1998; Elerian et al., 2001; Kim et al., 1998).

9. From a theoretical point of view, the choice of $\hat{p}$ based on the IMSE criterion may not maximize the power of the test. A more sensible alternative is to develop a data-driven $\hat{p}$ using a suitable power criterion or a criterion that trades off level distortion and power. This will necessitate higher order asymptotic analysis and is beyond the scope of this paper. We are content with the IMSE criterion here. Our simulation study shows that the power of the proposed test is relatively flat in the neighborhood of the optimal lag order that maximizes the power, and the data-driven $\hat{p}$ based on IMSE performs reasonably well.

10. We have also tried the Bartlett, Parzen and quadratic-spectral kernels; they give level and power very similar to that of the Daniell kernel.

11. The results at the 10% and 1% levels and the results for $n = 200$ are available from the authors. The results at the 10% and 1% levels have similar patterns, and the levels with $n = 200$ are very similar to those with $n = 100$.

12. The main reason that Anderson et al. (2002) first remove the first-order autocorrelation from the raw-observed returns is for convenient specification and estimation of continuous-time models. This is not the only way to go, but any reasonable approach will not affect the estimation results much. Moreover, it is a common practice to find that the conditional mean specification has little effect on the rest of the estimated model when using daily financial data.

13. Let T be one of the statistics $(\hat{M}(\hat{p}_0), \mathrm{BPL}(p), \mathrm{LM}(p,r), \mathrm{BDS}(m,d))$ based on the observed sample $\{Y_t\}_{t=1}^n$. The bootstrap p-value of T is computed as follows: (i) Estimate the MA(1)-EGARCH(p,q) model and get the estimated standardized residuals $\{\hat{e}_t\}_{t=1}^n$. (ii) Resample $\{\hat{e}_t\}_{t=1}^n$ with replacement to get $\{\hat{e}_t^b\}_{t=1}^n$. (iii) Generate the bootstrap samples $\{Y_t^b\}_{t=1}^n$ recursively from the estimated MA(1)-EGARCH(p,q) model using $\{\hat{e}_t^b\}_{t=1}^n$. (iv) Estimate the MA(1)-EGARCH(p,q) model using the bootstrapped data $\{Y_t^b\}_{t=1}^n$ to get the standardized residuals $\{\hat{e}_t^b\}_{t=1}^n$. (v) Compute

https://doi.org/10.1017/S0266466603196089 Published online by Cambridge University Press

the statistic T^b using $\{\hat{e}_t^b\}_{t=1}^n$. Repeat the steps (ii)–(v) B times (i.e., $b = 1,\ldots,B$), where B is the number of bootstrap replications. We use $B = 500$. The bootstrap p-value of the statistic T is then obtained from $B^{-1}\sum_{b=1}^{B}\mathbf{1}(T^b \geq T)$.

REFERENCES

Andersen, T.G., L. Benzoni, & J. Lund (2002) An empirical investigation of continuous-time equity return models. *Journal of Finance* 57, 1239–1284.

Barnett, W., A.R. Gallant, M.J. Hinich, J.A. Jungeilges, D.T. Kaplan, & M.J. Jensen (1997) A single-blind controlled competition among tests for nonlinearity and chaos. *Journal of Econometrics* 82, 157–192.

Bera, A.K. & M.L. Higgins (1997) ARCH and bilinearity as competing models for nonlinear dependence. *Journal of Business and Economic Statistics* 15, 43–50.

Bollerslev, T. (1987) A conditional heteroskedastic time series model for speculative prices and rates of return. *Review of Economics and Statistics* 69, 542–547.

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

Box, G. & D. Pierce (1970) Distribution of residual autocorrelations in autoregressive integrated moving average time series models. *Journal of the American Statistical Association* 65, 1509–1527.

Brillinger, D.R. & M. Rosenblatt (1967a) Asymptotic theory of estimates of the *k*th order spectra. In B. Harris (ed.), *Spectral Analysis of Time Series*, pp. 153–188. New York: Wiley.

Brillinger, D.R. & M. Rosenblatt (1967b) Computation and interpretation of the *k*th order spectra. In B. Harris (ed.), *Spectral Analysis of Time Series*, pp. 189–232. New York: Wiley.

Brock, W., D. Dechert, J. Scheinkman, & B. LeBaron (1996) A test for independence based on the correlation dimension. *Econometric Reviews* 15, 197–235.

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

Brooks, C. & 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.

Brown, B.M. (1971) Martingale limit theorems. *Annals of Mathematical Statistics* 42, 59–66.

Cameron, A.C. & 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. & L.T. Tran (1992) Nonparametric tests for serial dependence. *Journal of Time Series Analysis* 13, 102–113.

Cheung, Y.W. & L.K. Ng (1992) Interactions between the U.S. and Japan stock market indices. *Journal of International Financial Markets, Institutions, and Money* 2, 51–70.

Christoffersen, P. (1998) Evaluating interval forecasts. *International Economic Review* 39, 841–862.

Clement, M.P. & J. Smith (2000) Evaluating density forecasts of linear and nonlinear models: Applications to output growth and unemployment. *Journal of Forecasting* 19, 255–276.

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.

Diebold, F.X., T.A. Gunther, & A.S. Tay (1998) Evaluating density forecasts with applications to financial risk management. *International Economic Review* 39, 863–883.

Elerian, O., S. Chib, & N. Shephard (2001) Likelihood inference for discretely observed non-linear diffusions. *Econometrica* 69, 959–993.

Engle, R.F. (1982) Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* 50, 987–1007.

Engle, R.F. & T. Bollerslev (1986) Modelling the persistence of conditional variances. *Econometric Reviews* 5, 1–50.

Engle, R.F. & J.R. Russell (1998) Autoregressive conditional duration: A new model for irregularly spaced transaction data. *Econometrica* 66, 1127–1162.

https://doi.org/10.1017/S0266466603196089 Published online by Cambridge University Press

Epps, T.W. (1987) Testing that a stationary time series is Gaussian. *Annals of Statistics* 15, 1683–1698.

Epps, T.W. (1988) Testing that a Gaussian process is stationary. *Annals of Statistics* 16, 1667–1683.

Fama, E.F. & R. Roll (1968) Some properties of symmetric stable distributions. *Journal of the American Statistical Associations* 63, 817–836.

Feuerverger, A. (1990) An efficiency result for the empirical characteristic function in stationary time-series models. *Canadian Journal of Statistics* 18, 155–161.

Gallant, A.R. & G. Tauchen (1996) Which moment to match? *Econometric Theory* 12, 657–681.

Granger, C.W.J. (1983) Forecasting white noise. In A. Zellner (ed.), *Applied Time Series Analysis of Economic Data*, Proceedings of the Conference on Applied Time Series Analysis of Economic Data (October 1981). Washington, DC: U.S. Government Printing Office.

Granger, C.W.J. (2001) Overview of nonlinear macroeconometric empirical models. *Macroeconomic Dynamics* 5, 466–481.

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

Granger, C.W.J. & T.-H. Lee (1999) The effect of aggregation on nonlinearity. *Econometric Reviews* 18, 259–269.

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

Granger, C.W.J. & T. Teräsvirta (1999) A simple nonlinear time series model with misleading linear properties. *Economics Letters* 62, 161–165.

Grassberger, P. & I. Procaccia (1983) Measuring the strangeness of strange attractors. *Physica D* 9, 189–208.

Hamao, Y., R.W. Masulis, & V.K. Ng (1990) Correlations in price changes and volatility across international stock markets. *Review of Financial Studies* 3, 281–308.

Harvey, A., E. Ruiz, & N. Shephard (1994) Multivariate stochastic variance models. *Review of Economic Studies* 61, 247–264.

He, C. & T. Teräsvirta (1999) Fourth moment structure of the GARCH(p,q) process. *Econometric Theory* 15, 824–846.

Higgins, M. & A. Bera (1992) A class of nonlinear ARCH models. *International Economic Review* 33, 137–158.

Hong, Y. (1996) Consistent testing for serial correlation of unknown form. *Econometrica* 64, 837–864.

Hong, Y. (1998) Testing for pairwise independence via the empirical distribution function. *Journal of the 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 the American Statistical Association* 84, 1201–1220.

Hsieh, D. (1989) Modeling heteroskedasticity in daily-foreign exchange rates. *Journal of Business and Economic Statistics* 7, 307–317.

Jiang, G.J. & J.L. Knight (2002) Estimation of continuous time processes via the empirical characteristic function. *Journal of Business and Economic Statistics* 20, 198–212.

Karolyi, G.A. (1995) A multivariate GARCH model of international transmissions of stock returns and volatility: The case of the United States and Canada. *Journal of Business and Economic Statistics* 13, 11–25.

Kim, S., N. Shephard, & S. Chib (1998) Stochastic volatility: Likelihood inference and comparison with ARCH models. *Review of Economic Studies* 65, 361–393.

Knight, J.L. & S.E. Satchell (1997) The cumulant generating function estimation method. *Econometric Theory* 13, 170–184.

Knight, J.L. & J. Yu (2002) Empirical characteristic function in time series estimation. *Econometric Theory* 18, 691–721.

Lee, S.-W. & B. Hansen (1994) Asymptotic theory for the GARCH(1,1) quasi-maximum likelihood estimator. *Econometric Theory* 10, 29–52.

Lee, T.-H., H. White, & C.W.J. Granger (1993) Testing for neglected nonlinearity in time series

https://doi.org/10.1017/S0266466603196089 Published online by Cambridge University Press

models: A comparison of neural network methods and alternative tests. *Journal of Econometrics* 56, 269–290.
Lee, T.K.Y. & Y.K. Tse (1991) Term structure of interest rates in the Singapore Asian dollar market. *Journal of Applied Econometrics* 6, 143–152.
Li, W.K. & T.K. Mak (1994) On the squared residual autocorrelations in nonlinear time series with conditional heteroskedasticity. *Journal of Time Series Analysis* 15, 627–636.
Ljung, G.M. & 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.
Lundbergh, S. & T. Teräsvirta (2002) Evaluating GARCH models. *Journal of Econometrics* 110, 417–435.
McLeod, A.I. & W.K. Li (1983) Diagnostic checking ARMA time series models using squared-residual autocorrelations. *Journal of Time Series Analysis* 4, 269–273.
Nelson, D.B. (1991) Conditional heteroskedasticity in asset returns: A new approach. *Econometrica* 59, 347–370.
Pagan, A. (1996) The econometrics of financial markets. *Journal of Empirical Finance* 3, 15–102.
Pagan, A. & 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, 143–176.
Paparoditis, E. (2000b) On Some Power Properties of Goodness-of-Fit Tests in Time Series Analysis. Working paper, 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*. London: Academic Press.
Priestley, M.B. (1988) *Non-Linear and Non-Stationary Time Series Analysis*. London: Academic Press.
Robinson, P.M. (1991) Consistent nonparametric entropy-based testing. *Review of Economic Studies* 58, 437–453.
Sakai, H. & H. Tokumaru (1980) Autocorrelations of a certain chaos. *IEEE Transactions on Acoustics, Speech, and Signal Processing* 28, 588–590.
Singleton, K. (2001) Estimation of affine asset pricing models using the empirical characteristic function. *Journal of Econometrics* 102, 111–141.
Skaug, H.J. & D. Tjøstheim (1993a) Nonparametric test of serial independence based on the empirical distribution function. *Biometrika* 80, 591–602.
Skaug, H.J. & D. Tjøstheim (1993b) Nonparametric tests of serial independence. In T. Subba Rao (ed.), *Developments in Time Series Analysis*, the Priestley Birthday Volume, pp. 207–229. London: Chapman and Hall.
Skaug, H.J. & D. Tjøstheim (1996) Measures of distance between densities with application to testing for serial independence. In P. Robinson and M. Rosenblatt (eds.), *Time Series Analysis in Memory of E. J. Hannan*, pp. 363–377. New York: Springer.
Subba Rao, T. & M. Gabr (1980) A test for linearity of stationary time series. *Journal of Time Series Analysis* 1, 145–158.
Subba Rao, T. & M. Gabr (1984) *An Introduction to Bispectral Analysis and Bilinear Time Series Models.* Lecture Notes in Statistics 24. New York: Springer.
Teräsvirta, T., D. Tjøstheim, & C.W.J. Granger (1994) Aspects of modelling nonlinear time series. In R.F. Engle & D.L. McFadden (eds.), *Handbook of Econometrics*, vol. 4, pp. 2917–2957. Amsterdam: North-Holland.
Terdik, G. (1999) *Bilinear Stochastic Models and Related Problems of Nonlinear Time Series Analysis: A Frequency Domain Approach*. Lecture Notes in Statistics No. 142. New York: Springer-Verlag.

https://doi.org/10.1017/S0266466603196089 Published online by Cambridge University Press

Tjøstheim, D. (1994) Non-linear time series: A selective review. *Scandinavian Journal of Statistics* 21, 97–130.
Tjøstheim, D. (1996) Measures and tests of independence: A survey. *Statistics* 28, 249–284.
Tong, H. (1990) *Nonlinear Time Series: A Dynamic System Approach*. Oxford: Clarendon Press.
Tse, Y.K. & X.L. Zuo (1997) Testing for conditional heteroskedasticity: Some Monte Carlo results. *Journal of Statistical Computation and Simulation* 58, 237–253.
Weiss, A.A. (1986) ARCH and bilinear time series models: Comparison and combination. *Journal of Business and Economic Statistics* 4, 59–70.
Yang, M. & R. Bewley (1995) Moving average conditional heteroskedasticity processes. *Economics Letters* 49, 367–372.

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 $\{e_t = e_t(\theta_0)\}_{t=1}^n$ replacing the standardized residual sample $\{\hat{e}_t\}_{t=1}^n$, where $\theta_0 = p\lim \hat{\theta}$. 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 and A.2, which follow. 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

$$\hat{e}_t \equiv \frac{Y_t - g(\hat{I}_{t-1},\hat{\theta})}{h(\hat{I}_{t-1},\hat{\theta})} = \frac{Y_t - g(I_{t-1},\hat{\theta})}{h(\hat{I}_{t-1},\hat{\theta})} + \frac{g(I_{t-1},\hat{\theta}) - g(\hat{I}_{t-1},\hat{\theta})}{h(\hat{I}_{t-1},\hat{\theta})}$$

$$= e_t(\hat{\theta}) + e_t(\hat{\theta})\frac{h(I_{t-1},\hat{\theta}) - h(\hat{I}_{t-1},\hat{\theta})}{h(\hat{I}_{t-1},\hat{\theta})} + \frac{g(I_{t-1},\hat{\theta}) - g(\hat{I}_{t-1},\hat{\theta})}{h(\hat{I}_{t-1},\hat{\theta})}. \quad \textbf{(A.1)}$$

By the mean value theorem,

$$e_t(\hat{\theta}) = \frac{Y_t - g(I_{t-1},\theta_0)}{h(I_{t-1},\theta_0)} + \xi_t(\bar{\theta})'(\hat{\theta} - \theta_0) = e_t(\theta_0) + \xi_t(\bar{\theta})'(\hat{\theta} - \theta_0) \quad \textbf{(A.2)}$$

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

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

https://doi.org/10.1017/S0266466603196089 Published online by Cambridge University Press

It follows from (A.1), (A.2), and the Cauchy–Schwarz inequality that

$$\begin{aligned}
&\sum_{t=1}^{n}[\hat{e}_t - e_t(\hat{\theta})]^2 \\
&\quad \le 2\sum_{t=1}^{n} e_t^2(\hat{\theta})\left|\frac{h(I_{t-1},\hat{\theta}) - h(\hat{I}_{t-1},\hat{\theta})}{h(\hat{I}_{t-1},\hat{\theta})}\right|^2 + 2\sum_{t=1}^{n}\left|\frac{g(I_{t-1},\hat{\theta}) - g(\hat{I}_{t-1},\hat{\theta})}{h(\hat{I}_{t-1},\hat{\theta})}\right|^2 \\
&\quad \le 4\sum_{t=1}^{n} e_t^2(\theta_0)\sup_{\theta\in\Theta_0}\left[\frac{h(I_{t-1},\theta) - h(\hat{I}_{t-1},\theta)}{h(\hat{I}_{t-1},\theta)}\right]^2 \\
&\qquad + 4\|\hat{\theta} - \theta_0\|^2\left[\sum_{t=1}^{n}\sup_{\theta\in\Theta_0}\|\xi_t(\theta)\|^4\right]^{1/2}\left\{\sum_{t=1}^{n}\sup_{\theta\in\Theta_0}\left[\frac{h(I_{t-1},\theta) - h(\hat{I}_{t-1},\theta)}{h(\hat{I}_{t-1},\theta)}\right]^4\right\}^{1/2} \\
&\qquad + 2\sum_{t=1}^{n}\sup_{\theta\in\Theta_0}\left[\frac{g(I_{t-1},\theta) - g(\hat{I}_{t-1},\theta)}{h(\hat{I}_{t-1},\theta)}\right]^2 = O_P(1), \qquad \textbf{(A.3)}
\end{aligned}$$

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 (A.2) and Assumptions A.2–A.4, we have

$$\sum_{t=1}^{n}[e_t(\hat{\theta}) - e_t(\theta_0)]^2 \le n\|\hat{\theta} - \theta_0\|^2\left[n^{-1}\sum_{t=1}^{n}\sup_{\theta\in\Theta_0}\|\xi_t(\theta)\|^2\right] = O_P(1). \qquad \textbf{(A.4)}$$

Both (A.3) and (A.4) imply

$$\sum_{t=1}^{n}[\hat{e}_t - e_t(\theta_0)]^2 = O_P(1). \qquad \textbf{(A.5)}$$

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 $\hat{M}(p) - M(p) \xrightarrow{p} 0$, it suffices to show that

$$p^{-1/2}\int\sum_{j=1}^{n-1} k^2(j/p)n_j[|\hat{\sigma}_j(u,v)|^2 - |\tilde{\sigma}_j(u,v)|^2]\,dW(u)\,dW(v) \xrightarrow{p} 0, \qquad \textbf{(A.6)}$$

$\hat{C}_0 - \tilde{C}_0 = O_P(n^{-1/2})$, 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 (A.6); the proofs for $\hat{C}_0 - \tilde{C}_0 = O_P(n^{-1/2})$ and $\hat{D}_0 - \tilde{D}_0 \xrightarrow{p} 0$ are straightforward. We note that here it is necessary to obtain the convergence rate for $\hat{C}_0 - \tilde{C}_0$ to ensure that replacing $\hat{C}_0$ with $\tilde{C}_0$ has asymptotically negligible impact given $p/n \to 0$.

To show (A.6), we first decompose

$$\int\sum_{j=1}^{n-1} k^2(j/p)n_j[|\hat{\sigma}_j(u,v)|^2 - |\tilde{\sigma}_j(u,v)|^2]\,dW(u)\,dW(v) = \hat{A}_1 + 2\,\mathrm{Re}(\hat{A}_2), \qquad \textbf{(A.7)}$$

https://doi.org/10.1017/S0266466603196089 Published online by Cambridge University Press

where

$$\hat{A}_1 = \int \sum_{j=1}^{n-1} k^2(j/p) n_j |\hat{\sigma}_j(u,v) - \tilde{\sigma}_j(u,v)|^2 \, dW(u) \, dW(v), \tag{A.8}$$

$$\hat{A}_2 = \int \sum_{j=1}^{n-1} k^2(j/p) n_j [\hat{\sigma}_j(u,v) - \tilde{\sigma}_j(u,v)] \hat{\sigma}_j(u,v)^* \, dW(u) \, dW(v), \tag{A.9}$$

where $\mathrm{Re}(\hat{A}_2)$ 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, (A.6) follows from Propositions A.1–A.2, which appear subsequently, 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,* $p^{-1/2}\hat{A}_2 \xrightarrow{p} 0$.

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

$$\begin{aligned}
&\hat{\sigma}_j(u,v) - \tilde{\sigma}_j(u,v) \\
&\quad = n_j^{-1} \sum_{t=j+1}^{n} \hat{\delta}_t(u)\hat{\delta}_{t-j}(v) - \left[n_j^{-1} \sum_{t=j+1}^{n} \hat{\delta}_t(u) \right] \left[n_j^{-1} \sum_{t=j+1}^{n} \hat{\delta}_{t-j}(v) \right] \\
&\qquad + n_j^{-1} \sum_{t=j+1}^{n} \varphi_t(u)\hat{\delta}_{t-j}(v) - \left[n_j^{-1} \sum_{t=j+1}^{n} \varphi_t(u) \right] \left[n_j^{-1} \sum_{t=j+1}^{n} \hat{\delta}_{t-j}(v) \right] \\
&\qquad + n_j^{-1} \sum_{t=j+1}^{n} \hat{\delta}_t(u)\varphi_{t-j}(v) - \left[n_j^{-1} \sum_{t=j+1}^{n} \hat{\delta}_t(u) \right] \left[n_j^{-1} \sum_{t=j+1}^{n} \varphi_{t-j}(v) \right] \\
&\quad = \hat{B}_{1j}(u,v) - \hat{B}_{2j}(u,v) + \hat{B}_{3j}(u,v) - \hat{B}_{4j}(u,v) + \hat{B}_{5j}(u,v) - \hat{B}_{6j}(u,v), \quad \text{say.}
\end{aligned} \tag{A.10}$$

It follows from (A.10) 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, which appear subsequently, 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. $\int \sum_{j=1}^{n-1} k^2(j/p) n_j |\hat{B}_{6j}(u,v)|^2 \, dW(u) \, dW(v) = O_P(p/n)$.

https://doi.org/10.1017/S0266466603196089 Published online by Cambridge University Press

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

$$|\hat{B}_{1j}(u,v)|^2 \leq \left[n_j^{-1} \sum_{t=1}^{n} |\hat{\delta}_t(u)|^2 \right] \left[n_j^{-1} \sum_{t=1}^{n} |\hat{\delta}_t(v)|^2 \right]$$
$$\leq (uv)^2 \left[n_j^{-1} \sum_{t=1}^{n} (\hat{e}_t - e_t)^2 \right]^2. \tag{A.11}$$

It follows from (A.11), (A.5), and Assumptions A.5 and A.6 that

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

where we made use of the fact that

$$\sum_{j=1}^{n-1} k^2(j/p) n_j^{-1} = O(p/n) \tag{A.12}$$

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

$$|\hat{\delta}_t(u) - iu(\hat{e}_t - e_t) e^{iue_t}| \leq u^2 (\hat{e}_t - e_t)^2. \tag{A.13}$$

Also, a second-order Taylor expansion yields

$$e_t(\hat{\theta}) = e_t(\theta_0) + \xi_t(\theta_0)'(\hat{\theta} - \theta_0) + \frac{1}{2} (\hat{\theta} - \theta_0)' \frac{\partial}{\partial \theta} \xi_t(\bar{\theta})(\hat{\theta} - \theta_0) \tag{A.14}$$

for some $\bar{\theta}$ between $\hat{\theta}$ and θ_0, where $\xi_t(\theta)$ is as in (A.2). Both (A.13) and (A.14) imply

$$|\hat{\delta}_t(u) - iu\tilde{\xi}_t(u)(\hat{\theta} - \theta_0)| \leq u^2 [\hat{e}_t - e_t(\theta_0)]^2 + |u| |\hat{e}_t - e_t(\hat{\theta})|$$
$$+ |u| \|\hat{\theta} - \theta_0\|^2 \sup_{\theta \in \Theta_0} \left\| \frac{\partial}{\partial \theta} \xi_t(\theta) \right\|, \tag{A.15}$$

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

$$n_j |\hat{B}_{3j}(u,v)| \leq |v| \|\hat{\theta} - \theta_0\| \left| \sum_{t=j+1}^{n} \varphi_t(u) \tilde{\xi}_{t-j}(v) \right| + v^2 \sum_{t=1}^{n} (\hat{e}_t - e_t)^2$$
$$+ |v| \sum_{t=1}^{n} |\hat{e}_t - e_t(\hat{\theta})| + |v| \|\hat{\theta} - \theta_0\|^2 \sum_{t=1}^{n} \sup_{\theta \in \Theta_0} \left\| \frac{\partial}{\partial \theta} \xi_t(\theta) \right\|. \tag{A.16}$$

https://doi.org/10.1017/S0266466603196089 Published online by Cambridge University Press

It follows from (A.3), (A.5), (A.12), and (A.16), and Assumptions A.1–A.6 that

$$\begin{aligned}
&\int \sum_{j=1}^{n-1} k^2(j/p) n_j |\hat{B}_{3j}(u,v)|^2 \, dW \\
&\quad \le 8\|\hat{\theta} - \theta_0\|^2 \sum_{j=1}^{n-1} k^2(j/p) n_j^{-1} \int \left| \sum_{t=j+1}^{n} \varphi_t(u) \tilde{\xi}_{t-j}(v) \right|^2 v^2 \, dW(u) \, dW(v) \\
&\qquad + 8\left[\sum_{t=1}^{n} (\hat{e}_t - e_t(\theta_0))^2 \right]^2 \left[\sum_{j=1}^{n-1} k^2(j/p) n_j^{-1} \right] \int v^4 \, dW(u) \, dW(v) \\
&\qquad + 8\left(\sum_{t=1}^{n} |\hat{e}_t - e_t(\hat{\theta})| \right)^2 \left[\sum_{j=1}^{n-1} k^2(j/p) n_j^{-1} \right] \int v^2 \, dW(u) \, dW(v) \\
&\qquad + 8\|\sqrt{n}(\hat{\theta} - \theta_0)\|^4 \left[n^{-1} \sum_{t=1}^{n} \sup_{\theta \in \Theta_0} \left\| \frac{\partial}{\partial \theta} \xi_t(\theta) \right\| \right]^2 \left[\sum_{j=1}^{n-1} k^2(j/p) n_j^{-1} \right] \\
&\qquad \times \int v^2 \, dW(u) \, dW(v) = O_P(p/n),
\end{aligned}$$

where we made use of the fact that $E|\sum_{t=j+1}^{n} \varphi_t(u) \tilde{\xi}_{t-j}(v)|^2 \le C n_j$ because $\varphi_t(u)$ 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

$$\begin{aligned}
\sum_{t=1}^{n} |\hat{e}_t - e_t(\hat{\theta})| &\le \sum_{t=1}^{n} |e_t(\hat{\theta})| \left| \frac{h(I_{t-1}, \hat{\theta}) - h(\hat{I}_{t-1}, \hat{\theta})}{h(\hat{I}_{t-1}, \hat{\theta})} \right| + \sum_{t=1}^{n} \left| \frac{g(I_{t-1}, \hat{\theta}) - g(\hat{I}_{t-1}, \hat{\theta})}{h(\hat{I}_{t-1}, \hat{\theta})} \right| \\
&\le \sum_{t=1}^{n} |e_t(\theta_0)| \sup_{\theta \in \Theta_0} \left| \frac{h(I_{t-1}, \theta) - h(\hat{I}_{t-1}, \theta)}{h(\hat{I}_{t-1}, \theta)} \right| \\
&\quad + \|\hat{\theta} - \theta_0\| \left[\sum_{t=1}^{n} \sup_{\theta \in \Theta_0} \|\xi_t(\theta)\|^2 \right]^{1/2} \\
&\quad \times \left\{ \sum_{t=1}^{n} \sup_{\theta \in \Theta_0} \left[\frac{g(I_{t-1}, \theta) - g(\hat{I}_{t-1}, \theta)}{h(\hat{I}_{t-1}, \theta)} \right]^2 \right\}^{1/2} \\
&\quad + \sum_{t=1}^{n} \sup_{\theta \in \Theta_0} \left| \frac{g(I_{t-1}, \theta) - g(\hat{I}_{t-1}, \theta)}{h(\hat{I}_{t-1}, \theta)} \right| = O_P(1). \qquad \textbf{(A.17)}
\end{aligned}$$

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

$$|\hat{B}_{4j}(u,v)|^2 \le \left| n_j^{-1} \sum_{t=j+1}^{n} \varphi_t(u) \right|^2 \left[n_j^{-1} \sum_{t=1}^{n} |\hat{\delta}_t(v)|^2 \right]. \qquad \textbf{(A.18)}$$

https://doi.org/10.1017/S0266466603196089 Published online by Cambridge University Press

It follows from (A.18), the Cauchy–Schwarz inequality, and $|\hat{\delta}_t(v)| \leq |v\hat{e}_t - ve_t|$ that

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

given (A.5), (A.12), and $E|\sum_{t=j+1}^{n} \varphi_{t-j}(u)|^2 \leq Cn_j$ under the i.i.d. hypothesis of $\{e_t\}$. ■

Proof of Lemma A.5. Using (A.15), we have

$$\begin{aligned}
&\int \sum_{j=1}^{n-1} k^2(j/p) n_j |\hat{B}_{5j}(u,v)|^2 \, dW(u) \, dW(v) \\
&\quad \leq 8\|\hat{\theta} - \theta_0\|^2 \sum_{j=1}^{n-1} k^2(j/p) n_j^{-1} \int \left| \sum_{t=j+1}^{n} \tilde{\xi}_t(u) \varphi_{t-j}(v) \right|^2 u^2 \, dW(u) \, dW(v) \\
&\qquad + 8 \left[\sum_{t=1}^{n} (\hat{e}_t - e_t(\theta_0))^2 \right]^2 \left[\sum_{j=1}^{n-1} k^2(j/p) n_j^{-1} \right] \int u^4 \, dW(u) \, dW(v) \\
&\qquad + 8 \left(\sum_{t=1}^{n} |\hat{e}_t - e_t(\hat{\theta})| \right)^2 \left[\sum_{j=1}^{n-1} k^2(j/p) n_j^{-1} \right] \int u^2 \, dW(u) \, dW(v) \\
&\qquad + 8\|\sqrt{n}(\hat{\theta} - \theta_0)\|^4 \left[n^{-1} \sum_{t=1}^{n} \sup_{\theta \in \Theta_0} \left\| \frac{\partial}{\partial \theta} \xi_t(\theta) \right\| \right]^2 \\
&\qquad \times \left[\sum_{j=1}^{n-1} k^2(j/p) n_j^{-1} \right] \int u^2 \, dW(u) \, dW(v) = O_P(1),
\end{aligned} \tag{A.19}$$

where the last three terms are $O_P(p/n)$ given (A.5), (A.17), and (A.12), and Assumptions A.1–A.6 and the first term is $O_P(1)$, as is shown subsequently.

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

$$|\eta_j(u,v)| \leq [E|\tilde{\xi}_t(u)|^{2\nu}]^{1/2\nu} [E|\varphi_{t-j}(v)|^{2\nu}]^{1/2\nu} \alpha(j)^{(\nu-1)/\nu} \leq C\alpha(j)^{(\nu-1)/\nu}. \tag{A.20}$$

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

$$E \left| n_j^{-1} \sum_{t=j+1}^{n} [\tilde{\xi}_t(u)\varphi_{t-j}(v) - \eta_j(u,v)] \right|^2 \leq Cn_j^{-1}, \tag{A.21}$$

using reasoning analogous to (A.7)–(A.10) in the proof of Theorem 1 of Hong (1999, pp. 1212–1213). Consequently, from (A.20) and (A.21), we have

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

https://doi.org/10.1017/S0266466603196089 Published online by Cambridge University Press

It follows that the first term in (A.19) 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 (A.10), we have

$$|[\hat{\sigma}_j(u,v) - \tilde{\sigma}_j(u,v)]\tilde{\sigma}_j(u,v)^*| \leq \sum_{a=1}^{6} |\hat{B}_{aj}(u,v)||\tilde{\sigma}_j(u,v)|, \tag{A.22}$$

where the $\hat{B}_{aj}(u,v)$ are defined in (A.10). For $a = 1, 2, 3, 4$, and 6, we have, by the Cauchy–Schwarz inequality,

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

given Lemmas A.1–A.4 and A.6, and $p/n \to 0$, where $p^{-1}\sum_{j=1}^{n-1} k^2(j/p)n_j \times \int |\tilde{\sigma}_j(u,v)|^2\, dW = O_P(1)$ as follows from Markov's inequality, the i.i.d. hypothesis of $\{e_t\}$, and (A.12).

It remains to consider $a = 5$. Using (A.15), the triangular inequality, (A.5), and (A.17), we have

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

given $p \to \infty$, $p/n \to 0$, and Assumptions A.1–A.6, where we have made use of the fact that $n_j E|\tilde{\sigma}_j(u,v)|^2 \leq C$ 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

https://doi.org/10.1017/S0266466603196089 Published online by Cambridge University Press

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

given (A.20) and (A.21), and consequently,

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

given $\sum_{j=1}^{\infty}\alpha(j)^{(\nu-1)/\nu}<\infty$, $|k(\cdot)|\le 1$, and (A.12). ■

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, which follow.

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

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

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

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

$$n^{-1}\int\sum_{j=1}^{n-1}k^2(j/p)n_j[|\hat{\sigma}_j(u,v)|^2-|\tilde{\sigma}_j(u,v)|^2]\,dW(u)\,dW(v)\xrightarrow{p}0, \tag{A.23}$$

$\hat{C}_0-\tilde{C}_0=O_P(1)$, 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 $\{\hat{e}_t\}_{t=1}^n$. Because the proofs for $\hat{C}_0-\tilde{C}_0=O_P(1)$ and $\hat{D}_0-\tilde{D}_0\xrightarrow{p}0$ are straightforward, we focus on the proof of (A.23). From (A.7), 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

$$n^{-1}\hat{A}_1\xrightarrow{p}0, \tag{A.24}$$

where $\hat{A}_1$ is defined in (A.8). Given (A.10), we shall show that

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

https://doi.org/10.1017/S0266466603196089 Published online by Cambridge University Press

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

$$|\hat{B}_{1j}(u,v)|^2 \le \left[n_j^{-1}\sum_{t=j+1}^{n}|\hat{\delta}_t(u)|^2\right]\left[n_j^{-1}\sum_{t=j+1}^{n}|\hat{\delta}_t(v)|^2\right] \le n_j^{-2}(uv)^2\left[\sum_{t=1}^{n}(\hat{e}_t - e_t)^2\right]^2, \tag{A.25}$$

where

$$n^{-1/2}\sum_{t=1}^{n}(\hat{e}_t - e_t)^2 = O_P(1) \tag{A.26}$$

as can be shown in a way similar to that for (A.5), given the condition that $E(e_t^4) \le C$. Note that compared to (A.5), a factor of $n^{-1/2}$ arises here because we no longer have independence between e_t and $\{I_{t-1}, \hat{I}_{t-1}\}$, and we thus have to use the Cauchy–Schwarz inequality. It follows from (A.25), (A.26), (A.5), and (A.12) that

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

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

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

$$|B_{3j}(u,v)|^2 \le \left[n_j^{-1}\sum_{t=j+1}^{n}|\varphi_t(u)|^2\right]\left[n_j^{-1}\sum_{t=j+1}^{n}|\hat{\delta}_{t-j}(v)|^2\right] \le v^2 n_j^{-1}\sum_{t=1}^{n}(\hat{e}_t - e_t)^2. \tag{A.27}$$

It follows that

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

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, which follow.

THEOREM A.5. *Under the conditions of Theorem 3,* $\hat{M}(p) - M(p) \xrightarrow{p} 0$.

THEOREM A.6. *Under the conditions of Theorem 3,* $M(p) \xrightarrow{d} N(\mu,1)$.

https://doi.org/10.1017/S0266466603196089 Published online by Cambridge University Press

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

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

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 $\overline{M}(p)$ 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, which follow.

PROPOSITION A.3. *Under the conditions of Theorem 3,* $M(p) - \overline{M}(p) \xrightarrow{p} 0$.

PROPOSITION A.4. *Under the conditions of Theorem 3,* $\overline{M}(p) \xrightarrow{d} N(\mu,0)$.

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

$$p^{-1/2} \int \sum_{j=1}^{n-1} k^2(j/p) n_j [|\tilde{\sigma}_j(u,v)|^2 - |\bar{\sigma}_j(u,v)|^2]\, dW(u)\, dW(v) \xrightarrow{p} 0, \tag{A.28}$$

$\tilde{C}_0 - \bar{C}_0 = O_P(1)$, and $\tilde{D}_0 - \bar{D}_0 \xrightarrow{p} 0$, 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 (A.28) only. By straightforward algebra, we have

$$\int \sum_{j=1}^{n-1} k^2(j/p) n_j [|\tilde{\sigma}_j(u,v)|^2 - |\bar{\sigma}_j(u,v)|^2]\, dW(u)\, dW(v) = \hat{B}_1 + 2\,\mathrm{Re}(\hat{B}_2), \tag{A.29}$$

where

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

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

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

$$\begin{aligned} E|\tilde{\sigma}_j(u,v) - \bar{\sigma}_j(u,v)|^2 &\le \left[E\left|n_j^{-1}\sum_{t=j+1}^{n}\psi_t(u)\right|^4\right]^{1/2}\left[E\left|n_j^{-1}\sum_{t=j+1}^{n}\psi_{t-j}(v)\right|^4\right]^{1/2} \\ &\le Cn_j^{-2}, \end{aligned} \tag{A.30}$$

where $E|\sum_{t=j+1}^n \psi_t(u)|^4 \le Cn_j^2$ because $\{\psi_t(u)\}$ is a bounded one-dependent random sequence with mean 0. It follows from Markov's inequality, (A.30), (A.12), and $p/n \to 0$ that

$$\hat{B}_1 = O_P(p/n) = o_P(1). \tag{A.31}$$

https://doi.org/10.1017/S0266466603196089 Published online by Cambridge University Press

Next we consider $\hat{B}_2$ in (A.29). Observe that

$$|\bar{\sigma}_j(u,v)|^2 \le 2|\sigma_j(u,v)|^2 + 2|\bar{\sigma}_j(u,v) - \sigma_j(u,v)|^2. \tag{A.32}$$

Also, because $\{\psi_t(u)\}$ is one-dependent, we have

$$\sigma_j(u,v) = \begin{cases} E[\psi_t(u)\psi_{t-1}(v)], & j = 1 \\ 0, & j > 1 \end{cases} \tag{A.33}$$

and

$$E|\bar{\sigma}_j(u,v) - \sigma_j(u,v)|^2 = \left| n_j^{-1} \sum_{t=j+1}^{n} [\psi_t(u)\psi_{t-j}(v) - \sigma_j(u,v)] \right|^2 \le C n_j^{-1} \tag{A.34}$$

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 (A.32)–(A.34), $|k(z)| \le 1$, and Markov's inequality that

$$\begin{aligned} &\int \sum_{j=1}^{n-1} k^2(j/p) n_j |\bar{\sigma}_j(u,v)|^2 \, dW(u)\, dW(v) \\ &\quad \le 2 \int \sum_{j=1}^{n-1} k^2(j/p) n_j |\sigma_j(u,v)|^2 \, dW(u)\, dW(v) \\ &\qquad + 2 \int \sum_{j=1}^{n-1} k^2(j/p) n_j |\bar{\sigma}_j(u,v) - \sigma_j(u,v)|^2 \, dW(u)\, dW(v) \\ &\quad \le 2n \int |\sigma_1(u,v)|^2 \, dW(u)\, dW(v) \\ &\qquad + 2 \int \sum_{j=1}^{n-1} k^2(j/p) n_j |\bar{\sigma}_j(u,v) - \sigma_j(u,v)|^2 \, dW(u)\, dW(v) \\ &\quad = O(p^{1/2}) + O_P(p), \end{aligned} \tag{A.35}$$

where we used the fact that $n \int |\sigma_1(u,v)|^2 \, dW(u)\, dW(v) = O(p^{1/2})$ under $\mathbb{H}_n(p^{1/4}/n^{1/2})$. Combining (A.31), (A.35), $p/n \to 0$, and Cauchy–Schwarz inequality, we obtain

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

This completes the proof for Proposition A.3. ■

https://doi.org/10.1017/S0266466603196089 Published online by Cambridge University Press

Proof of Proposition A.4 We write

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

$$= k^2(1/p)\left[n_1 \int |\bar{\sigma}_1(u,v)|^2 \, dW(u)\, dW(v) - \bar{C}_0 \right]$$

$$+ k^2(2/p)\left[n_2 \int |\bar{\sigma}_2(u,v)|^2 \, dW(u)\, dW(v) - \bar{C}_0 \right]$$

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

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 $r = 2,4$, $\bar{C}_0 - C_0 = O_P(n^{-1/2})$, and $\bar{D}_0 - D_0 \xrightarrow{p} 0$, 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, which follow. These lemmas hold under $\mathbb{H}_n(p^{1/4}/n^{1/2})$.

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

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

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*

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

LEMMA A.10. $p^{-1/2}\hat{V} = p^{-1/2}\hat{V}_g + o_P(1)$, *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. $[2pD_0 \int_0^\infty k^4(z)\,dz]^{-1/2}\hat{V}_g \xrightarrow{d} N(0,1)$.

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

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

https://doi.org/10.1017/S0266466603196089 Published online by Cambridge University Press

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

$$\begin{aligned}
&\left| p^{-1/2} n_1 \int |\bar{\sigma}_1(u,v)|^2 \, dW(u)\, dW(v) - p^{-1/2} n_1 \int |\sigma_1(u,v)|^2 \, dW(u)\, dW(v) \right| \\
&\quad \le p^{-1/2} n_1 \int |\bar{\sigma}_1(u,v) - \sigma_1(u,v)|^2 \, dW(u)\, dW(v) \\
&\qquad + 2\left[p^{-1/2} n_1 \int |\bar{\sigma}_1(u,v) - \sigma_1(u,v)|^2 \, dW(u)\, dW(v) \right]^{1/2} \\
&\qquad \times \left[p^{-1/2} n_1 \int |\sigma_1(u,v)|^2 \, dW(u)\, dW(v) \right]^{1/2} \\
&\quad = O_P(p^{-1/2}) + O_P(p^{-1/4}),
\end{aligned} \tag{A.36}$$

where the equality follows from (A.34) and

$$p^{-1/2} n_1 \int |\sigma_1(u,v)|^2 \, dW(u)\, dW(v) \to \left[2D_0 \int_0^\infty k^4(z)\, dz \right]^{1/2} \mu \tag{A.37}$$

under $\mathbb{H}_n(p^{1/4}/n^{1/2})$. Combining (A.36), (A.37), 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 one-dependent. ■

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

$$\begin{aligned}
&\sum_{j=3}^{n-1} k^2(j/p) n_j \int |\bar{\sigma}_j(u,v)|^2 \, dW(u)\, dW(v) \\
&\quad = \sum_{j=3}^{n-1} k^2(j/p) n_j^{-1} \sum_{t=j+1}^{n} \int C_{ttj}(u,v)\, dW(u)\, dW(v) \\
&\qquad + \sum_{j=3}^{n-2} k^2(j/p) n_j^{-1} \sum_{t=j+2}^{n} \sum_{s=j+1}^{t-1} \int V_{tsj}(u,v)\, dW(u)\, dW(v) \\
&\quad = \hat{C} + \hat{V},
\end{aligned} \tag{A.38}$$

where $\hat{V}$ is defined in Lemma A.9 and $\hat{C} = \sum_{j=3}^{n-1} k^2(j/p) n_j^{-1} \sum_{t=j+1}^{n} \int C_{ttj}(u,v) \times dW(u)\, dW(v)$. We shall show $p^{-1/2}[\hat{C} - C_0 \sum_{j=3}^{n-1} k^2(j/p)] \xrightarrow{p} 0$. Because $\{\psi_t(u)\}$ is one-dependent, we have

$$\int E[C_{ttj}(u,v)]\, dW(u)\, dW(v) = \int E|\psi_t(u)|^2 \, dW(u) \int E|\psi_{t-j}(v)|^2 \, dW(v) = C_0. \tag{A.39}$$

https://doi.org/10.1017/S0266466603196089 Published online by Cambridge University Press

Also, because $\int C_{ttj}(u,v)\,dW(u)\,dW(v)$ 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

$$E\left\{\sum_{t=j+1}^{n}\int [C_{ttj}(u,v) - EC_{ttj}(u,v)]\,dW(u)\,dW(v)\right\}^2 \le Cn_j. \tag{A.40}$$

It follows from (A.39), (A.40), (A.12), and the Cauchy–Schwarz inequality that

$$\begin{aligned} p^{-1/2}&\left[\hat{C} - C_0\sum_{j=3}^{n-1}k^2(j/p)\right] \\ &= p^{-1/2}\sum_{j=3}^{n-1}k^2(j/p)n_j^{-1}\sum_{t=j+1}^{n}\int [C_{ttj}(u,v) - EC_{ttj}(u,v)]\,dW(u)\,dW(v) \\ &= O_P(p^{1/2}/n^{1/2}). \end{aligned} \tag{A.41}$$

The desired result follows from (A.38), (A.41), 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 $j \le g$ and $j > g$, respectively:

$$\begin{aligned} \hat{V} &= \left(\sum_{j=3}^{g}\sum_{t=j+2}^{n}\sum_{s=j+1}^{t-1} + \sum_{j=g+1}^{n-2}\sum_{t=j+2}^{n}\sum_{s=j+1}^{t-1}\right)k^2(j/p)n_j^{-1}\int V_{tsj}(u,v)\,dW(u)\,dW(v) \\ &= \hat{U} + \hat{R}_1, \quad \text{say.} \end{aligned} \tag{A.42}$$

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

$$\begin{aligned} \hat{U} &= \left(\sum_{j=3}^{g}\sum_{t=2}^{n}\sum_{s=1}^{t-1} - \sum_{j=3}^{g}\sum_{s=1}^{j}\sum_{t=s+1}^{n}\right)k^2(j/p)n_j^{-1}\int V_{tsj}(u,v)\,dW(u)\,dW(v) \\ &= \hat{W} - \hat{R}_2, \quad \text{say.} \end{aligned} \tag{A.43}$$

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

$$\begin{aligned} \hat{W} &= \left(\sum_{j=3}^{g}\sum_{t=g+2}^{n}\sum_{s=1}^{t-g-1} + \sum_{j=3}^{g}\sum_{t=2}^{n}\sum_{s=\max(1,\,t-g)}^{t-1}\right)k^2(j/p)n_j^{-1}\int V_{tsj}(u,v)\,dW(u)\,dW(v) \\ &= \hat{V}_g + \hat{R}_3, \quad \text{say,} \end{aligned} \tag{A.44}$$

where $\hat{V}_g$ is defined in Lemma A.10. Combining (A.42)–(A.44), we obtain $\hat{V} = \hat{V}_g + \hat{R}_1 - \hat{R}_2 + \hat{R}_3$. Thus, it suffices to show $p^{-1/2}\hat{R}_a \xrightarrow{p} 0$ for $a = 1,2,3$.

https://doi.org/10.1017/S0266466603196089 Published online by Cambridge University Press

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 $s < t-1$:

$$\hat{R}_1 = \sum_{j=g+1}^{n-2} \sum_{t=j+2}^{n} k^2(j/p) n_j^{-1} \int V_{t(t-1)j}(u,v)\, dW(u)\, dW(v)$$
$$+ \sum_{j=g+1}^{n-2} \sum_{t=j+2}^{n} \sum_{s=j+1}^{t-2} k^2(j/p) n_j^{-1} \int V_{tsj}(u,v)\, dW(u)\, dW(v) = \hat{R}_{11} + \hat{R}_{12}, \quad \text{say.} \tag{A.45}$$

Now consider the first term in (A.45). Because $\{\psi_t(u)\}$ is one-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

$$E\left\{ \sum_{t=j+2}^{n} \int [V_{t(t-1)j}(u,v) - EV_{t(t-1)j}(u,v)]\, dW(u)\, dW(v) \right\}^2 \le Cn_j. \tag{A.46}$$

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

$$EV_{t(t-1)j} = 2\,\mathrm{Re} \int EC_{t(t-1)j}(u,v)\, dW(u)\, dW(v)$$
$$= 2 \left| \int E[\psi_t(u)\psi_{t-1}(u)^*]\, dW(u) \right|^2 \le C(p^{1/2}/n) \tag{A.47}$$

under $\mathbb{H}_n(p^{1/4}/n^{1/2})$. It follows from (A.46), (A.47), (A.12), and Cauchy–Schwarz inequality that

$$\hat{R}_{11} = \sum_{j=g+1}^{n-2} k^2(j/p) n_j^{-1} \sum_{t=j+2}^{n} \int EV_{t(t-1)j}(u,v)\, dW(u,v)$$
$$+ \sum_{j=g+1}^{n-2} k^2(j/p) n_j^{-1} \sum_{t=j+2}^{n} \int [V_{t(t-1)j}(u,v) - EV_{t(t-1)j}(u,v)]\, dW(u,v)$$
$$= o(p^{3/2}/n) + o_P(p/n^{1/2}), \tag{A.48}$$

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 (A.12) and $g/p \to \infty$.

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

$$E\hat{R}_{12}^2 \le 4E \left| \sum_{j=g+1}^{n-2} \sum_{t=j+2}^{n} \sum_{s=j+1}^{t-2} k^2(j/p) n_j^{-1} \int \psi_t(u)\psi_s(u)^* \psi_{t-j}(v)\psi_{s-j}(v)^*\, dW(u)\, dW(v) \right|^2$$
$$= 4 \sum_{t=g+4}^{n} E \left| \int \psi_t(u) \sum_{s=g+2}^{t-2} \sum_{j=g+1}^{s-1} k^2(j/p) n_j^{-1} \psi_{t-j}(v)\psi_s(u)\psi_{s-j}(v)\, dW(u)\, dW(v) \right|^2$$
$$\le 4 \sum_{t=g+4}^{n} \int E|\psi_t(u)|^2 E \left| \sum_{s=g+2}^{t-2} \sum_{j=g+1}^{s-1} k^2(j/p) n_j^{-1} \psi_{t-j}(v)\psi_s(u)\psi_{s-j}(v) \right|^2 dW(u)\, dW(v), \tag{A.49}$$

https://doi.org/10.1017/S0266466603196089 Published online by Cambridge University Press

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 (A.49), we have

$$
\begin{aligned}
&E\left|\sum_{s=g+2}^{t-2}\sum_{j=g+1}^{s-1} k^2(j/p)n_j^{-1}\psi_{t-j}(v)\psi_s(u)\psi_{s-j}(v)\right|^2 \\
&\quad = 4E\left|\sum_{s=g+2}^{t-2}\sum_{j=1}^{s-1} k^2(j/p)n_j^{-1}\psi_{t-j}(v)\psi_s(u)\psi_{s-j}(v)\mathbf{1}(t-j>s+1)\right|^2 \\
&\qquad + 4E\left|\sum_{s=g+2}^{t-2}\sum_{j=g+1}^{s-1} k^2(j/p)n_j^{-1}\psi_{t-j}(v)\psi_s(u)\psi_{s-j}(v)\mathbf{1}(t-j<s-1)\right|^2 \\
&\qquad + 4E\left|\sum_{s=g+2}^{t-2}\sum_{j=g+1}^{s-1} k^2(j/p)n_j^{-1}\psi_{t-j}(v)\psi_s(u)\psi_{s-j}(v)\mathbf{1}(t-j=s,s\pm1)\right|^2 \\
&\quad \le C\sum_{s=g+2}^{t-2}\sum_{j=g+1}^{s-1} k^4(j/p)n_j^{-2}\mathbf{1}(t-j>s+1) \\
&\qquad + C\sum_{s=g+2}^{t-2}\sum_{j=g+1}^{s-1} k^4(j/p)n_j^{-2}\mathbf{1}(t-j<s-1) + C\left[\sum_{j=g+1}^{t-2} k^2(j/p)n_j^{-1}\right]^2. \qquad \textbf{(A.50)}
\end{aligned}
$$

Combining (A.49), (A.50), and (A.12), we obtain

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

given $g/p \to \infty$. It follows that $p^{-1/2}\hat{R}_{12} \xrightarrow{p} 0$ 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, $p^{-1/2}\hat{R}_2 \xrightarrow{p} 0$ and $p^{-1/2}\hat{R}_3 \xrightarrow{p} 0$ 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 the martingale central limit theorem of Brown (1971). The fact that $\{e_t\}$ is one-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 subsequently.

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

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

Given the conditions on $k(\cdot)$, there exists a symmetric monotonic decreasing function $k_0(z)$ of $z > 0$ such that $|k(z)| \le k_0(z)$ for all $z > 0$ and $k_0(\cdot)$ satisfies Assumption A.5. It follows that for any constants $\epsilon, \eta > 0$,

https://doi.org/10.1017/S0266466603196089 Published online by Cambridge University Press

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

where the second term vanishes for all $\eta > 0$ given $\hat{p}/p - 1 \xrightarrow{p} 0$. 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 $|\hat{p}/p - 1| \leq \eta$,

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

for any $\eta > 0$ given (A.6), 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 $(m, l) = (0,0)$). ■

https://doi.org/10.1017/S0266466603196089 Published online by Cambridge University Press