Hypothesis Testing in Time Series via the Empirical Characteristic Function: A Generalized Spectral Density Approach

Yongmiao HONG

The standardized spectral density completely describes serial dependence of a Gaussian process. For non-Gaussian processes, however, it may become an inappropriate analytic tool, because it misses the nonlinear processes with zero autocorrelation. By generalizing the concept of the standardized spectral density, I propose a new spectral tool suitable for both linear and nonlinear time series analysis. The generalized spectral density is indexed by frequency and a pair of auxiliary parameters. It is well defined for both continuous and discrete random variables, and requires no moment condition. Introduction of the auxiliary parameters renders the spectrum able to capture all pairwise dependencies, including those with zero autocorrelation. The standardized spectral density can be derived by properly differentiating the generalized spectral density with respect to the auxiliary parameters at the origin. The consistency of a class of Parzen's kernel-type estimators for the generalized spectral density is established, and their optimal convergence rates are derived using the integrated mean squared error criterion. A data-dependent asymptotically optimal bandwidth (or lag order) is introduced. The kernel estimators and their derivatives are applied to construct a class of asymptotically one-sided N(0, 1) tests for generic serial dependence and hypotheses on various specific aspects of serial dependence. The latter include serial uncorrelatedness, martingale, conditional homoscedasticity, conditional symmetry, and conditional homokurtosis. All of the proposed tests, which include Hong's spectral density test for serial correlation, can be derived from a unified framework. An empirical application to Deutschemark exchange rates highlights the approach.

KEY WORDS: Asymptotic normality; Bandwidth; Empirical characteristic function; Generalized spectral density; Hypothesis testing; Kernel function; Nonlinear dependence; Spectral density; Time series.

1. INTRODUCTION

Measuring and detecting serial dependence has long been of interest in time series analysis. Serial dependence is often characterized by the autocorrelation function, or equivalently by the standardized spectral density function. Frequency domain analysis is convenient and enlightening (Priestley 1981). When a stochastic process is serially independent, for example, its standardized spectral density is uniform. Any deviation of the spectral density from uniformity is evidence of serial dependence. This fact can be used to form consistent tests against serial correlation of unknown form (see, e.g., Anderson 1993; Hong 1996).

The standardized spectral density captures all serial dependencies of a Gaussian process. For non-Gaussian processes, however, it may become an inappropriate analytic tool, because it misses the nonlinear processes with zero autocorrelation. Well-known examples are autoregressive conditional heteroscedastic (ARCH), bilinear, nonlinear moving average, and threshold autoregressive processes (see, e.g., Tong 1990). It has been widely recognized that many time series arising in practice display non-Gaussian and nonlinear features (see, e.g., Brock, Hseih, and Le Baron 1991; Granger and Anderson 1978; Granger and Teräsvirta 1993; Priestley 1988; Subba Rao and Gabr 1984; Tong 1990).

The development of nonlinear time series analysis has been advancing rapidly, but there are relatively few useful analytic tools (Granger and Teräsvirta 1993). Higher-order spectra, which are the Fourier transforms of higher-order cumulants, have been proposed to capture nonlinear dependencies (Brillinger 1965; Brillinger and Rosenblatt 1967a, b; Hinich 1982; Hinich and Patterson 1992; Subba Rao and Gabr 1980, 1984). Although they can effectively capture many nonlinear dependencies, they may miss some important ones. The bispectrum, for example, may fail to capture ARCH processes because their third-order cumulants can be identically 0. Moreover, higher-order spectra need restrictive moment conditions (e.g., the existence of a sixth moment). These features make higher-order spectra less appealing in practice (Granger and Teräsvirta 1993, p. 22).

By generalizing the concept of the standardized spectral density, I propose a new spectral tool suitable for both linear and nonlinear time series analysis. The generalized spectral density can capture all pairwise dependencies while maintaining the nice features of the conventional spectrum. It is indexed by frequency and a pair of auxiliary parameters. The conventional spectral density can be derived as a special case by differentiating the generalized spectral density with respect to the auxiliary parameters at the origin. One of the central contributions of this article is to show that the introduction of the auxiliary parameters provides much flexibility for spectral analysis, rendering the spectrum able to capture all pairwise dependencies. The generalized spectrum can be estimated consistently by a class of Parzen's (1957) kernel-type estimators. I derive their optimal convergence rates using the integrated mean squared error (IMSE) crite-

Yongmiao Hong is Associate Professor, Department of Economics and Department of Statistical Science, Cornell University, Ithaca, NY 14853.
The author thanks the editor, an associate editor, and two referees for valuable comments and suggestions that have led to significant improvements in the manuscript. He also thanks W. Barnett, A. Bera, H. Bierens, M. Carrasco, N. Kiefer, T. Lee, J. McCulloch, J. Pinkse, J. Ramsey, P. Robinson, M. Stinchcombe, and M. Wells for helpful comments and discussions on the earlier versions.

Journal of the American Statistical Association
December 1999, Vol. 94, No. 448, Theory and Methods

rion, and introduce a data-dependent asymptotically optimal bandwidth that provides the choice of a lag order for any given sample size.

Unlike the standardized spectral density, the generalized spectral density needs no moment condition. It applies to time series generated from either discrete or continuous distribution with possibly infinite moments, as is often encountered in high-frequency economic and financial data (see, e.g., Fama and Roll 1968). Moreover, as a consequence of analyticity of the complex-valued exponential function, one can use the derivatives of the generalized spectrum to capture various specific aspects of serial dependence. I illustrate how the generalized spectrum and its derivatives can be used to test various hypotheses. These include tests of serial independence, serial uncorrelatedness, martingale, conditional homoscedasticity, conditional symmetry, and conditional homokurtosis. All of the proposed tests are derived from a unified framework and have a convenient null asymptotic one-sided N(0, 1) distribution. Hong's (1996) spectral density test is obtained as a special case by differentiating the generalized spectral density with respect to the auxiliary parameters at the origin.

My approach essentially uses empirical characteristic functions (ECFs) and their derivatives in a time series framework. The ECF has been widely used to test various hypotheses in the context of independent and identically distributed (iid) samples (Epps 1993). These include goodness of fit, symmetry, and sample heterogeneity (see, e.g., Baringhaus and Henze 1988; Csörgõ 1984; Epps and Pulley 1983; Epps and Singleton 1986; Feigin and Heathcote 1976; Feuerverger and Mureika 1977; Hall and Welsh 1983; Heathcote 1972; Henze and Wagner 1997; Koutrouvelis 1980; Koutrouvelis and Kellermeier 1981). Csörgõ (1985) and Feuerverger (1987, 1993) considered tests of independence among the components of an iid random vector. Epps (1987, 1988) was the first to use the ECF in a time series context to test whether a stationary time series is Gaussian or vice versa. Feuerverger (1987) noted the possibility of using the ECF to test serial independence. Pinkse (1998) considered a chi-squared test for first-order serial dependence via a characteristic function principle, although he did not use the ECF. This article differs from the aforementioned works in several ways, particularly in that I use spectral analysis and derivatives.

In Section 2, I introduce the generalized spectral density and a class of Parzen's (1957) kernel-type estimators for the generalized spectrum. I establish consistency of the kernel estimators and derive their optimal convergence rates using the IMSE criterion. A data-dependent asymptotically optimal bandwidth is introduced. In Sections 3 and 4 I use the kernel estimators and their derivatives to develop tests for generic serial dependence and various hypotheses of interest. I derive asymptotic normality of the proposed tests in Section 5 and study their asymptotic power in Section 6. In Section 7 I give an empirical application to Deutschmark exchange rates. Conclusions are provided in Section 8, and all the proofs are given in the Appendix. Throughout, $C \in (0, \infty)$ denotes a generic constant that may differ from place to place; $|A|, A^*, \mathrm{Re}(A)$, and $\mathrm{Im}(A)$ denote the usual Euclidean norm, complex conjugate, and real and imaginary parts of A. Unless indicated, all convergencies are taken as the sample size $n \to \infty$, and all unspecified integrals are taken over the entire Euclidean space with proper dimension.

2. GENERALIZED SPECTRAL DENSITY

Consider a stationary time series $\{X_t \in \mathbb{R}, t \in \mathbb{N}\}$ with marginal characteristic function $\varphi(u) = Ee^{iuX_0}$ and pairwise characteristic function $\varphi_j(u, v) = Ee^{i(uX_0 + vX_{-j})}$, where $\mathbb{N}$ is the set of integers, $i = \sqrt{-1}, j \in \mathbb{N}$, and $(u, v) \in \mathbb{R}^2$. Often, serial dependence of $\{X_t\}$ is described by its autocorrelation function, $\rho(j)$, or by its standardized spectral density,

$$h(\omega) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \rho(j) e^{-ij\omega}, \qquad \omega \in [-\pi, \pi].$$

Both $h(\omega)$ and $\rho(j)$ are the Fourier transforms of each other, containing the same information of serial dependence of $\{X_t\}$. It has been long known that nonlinear time series may have zero autocorrelation but display strong nonlinear dependence, and both $h(\omega)$ and $\rho(j)$ fail to capture such processes. Higher-order spectra have been proposed to capture nonlinear dependencies (Brillinger 1965). They can detect nonlinear dependencies but may still miss such important ones as ARCH. Moreover, they require restrictive moment conditions and relatively large sample sizes for reasonable estimation.

To avoid the undesired features of higher-order spectra and to enable capture of a wider range of nonlinear dependencies, I propose a new spectral tool. I first transform the original time series $\{X_t\}$ by $\{e^{iuX_t}\}$, and then consider $\sigma_j(u, v) \equiv \mathrm{cov}(e^{iuX_t}, e^{ivX_{t-|j|}})$, the covariance function between e^{iuX_t} and $e^{ivX_{t-|j|}}$. Straightforward algebra yields

$$\sigma_j(u, v) = \varphi_{|j|}(u, v) - \varphi(u)\varphi(v), \tag{1}$$

the difference between the joint characteristic function of $(X_t, X_{t-|j|})$ and the product of their marginals. Hence $\sigma_j(u, v) = 0$ for all $(u, v) \in \mathbb{R}^2$ if and only if X_t and $X_{t-|j|}$ are independent (Lukacs 1970). Consequently, $\sigma_j(u, v)$ is able to capture all pairwise dependencies, including those with zero autocorrelation.

Suppose that $\sup_{(u,v)\in\mathbb{R}^2} \sum_{j=-\infty}^{\infty} |\sigma_j(u, v)| < \infty$, which holds when, for example, $\{X_t\}$ is α mixing with $\sum_{j=-\infty}^{\infty} \alpha(j)^{(\nu-1)/\nu} < \infty$ for some $\nu > 1$. Then the Fourier transform of $\sigma_j(u, v)$ exists:

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

Both (1) and (2) contain the same information of serial dependence of $\{X_t\}$. In contrast to the standardized spectral density $h(\omega)$, $f(\omega, u, v)$ can capture all pairwise dependencies. It requires no moment condition and is well defined for both discrete and continuous random variables. When $\mathrm{var}(X_t) = \sigma^2$ exists, $h(\omega)$ can be obtained by differentiating

$f(\omega, u, v)$ with respect to (u, v) at (0, 0); that is,

$$h(\omega) = -\frac{1}{\sigma^2} \frac{\partial^2}{\partial u \partial v} f(\omega, u, v)|_{(u,v)=(0,0)}.$$

For this reason, I call $f(\omega, u, v)$ a "generalized spectral density function" of $\{X_t\}$, although it does not have the mathematical properties of a probability density function.

To estimate $f(\omega, u, v)$, I define the empirical measure

$$\hat{\sigma}_j(u, v) = \hat{\varphi}_j(u, v) - \hat{\varphi}_j(u, 0)\hat{\varphi}_j(0, v), \qquad j = 0, \pm 1, \ldots, \pm(n-1), \quad (3)$$

where $\hat{\varphi}_j(u, v) \equiv (n - |j|)^{-1} \sum_{t=|j|+1}^{n} e^{i(uX_t + vX_{t-|j|})}$ is an unbiased estimator for $\varphi_j(u, v)$. I introduce a class of Parzen's (1957) smoothed kernel estimators,

$$\hat{f}_n(\omega, u, v) = \frac{1}{2\pi} \sum_{j=1-n}^{n-1} (1 - |j|/n)^{1/2} k(j/p) \hat{\sigma}_j(u, v) e^{-ij\omega}, \quad (4)$$

where $k(\cdot)$ is a kernel function or "lag window" and p is a bandwidth or lag order. The factor $(1 - j/n)^{1/2}$ could be replaced by unity, but it gives better finite-sample performance.

I study the consistency property of (4) using the IMSE criterion,

$$\text{IMSE}(\hat{f}_n, f) = E \int \int_{-\pi}^{\pi} |\hat{f}_n(\omega, u, v) - f(\omega, u, v)|^2 \, d\omega \, dW(u, v), \quad (5)$$

where $W(u, v)$ is a weighting function. The following regularity conditions are imposed.

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

Assumption A.2. $W : \mathbb{R}^2 \to \mathbb{R}^+$ is nondecreasing with bounded total variation.

Assumption A.3. $k : \mathbb{R} \to [-1, 1]$ is symmetric and is continuous at 0 and all but a finite number of points, with $k(0) = 1, k_2 \equiv \int_{-\infty}^{\infty} k^2(z) \, dz < \infty$ and $|k(z)| \le C|z|^{-b}$ for large z, where $b > \frac{1}{2}$.

Assumption A.4. There exists $q \in (0, \infty)$ that is the largest real number such that $k^{(q)} \equiv \lim_{|z|\to 0}\{1 - k(z)\}/|z|^q \in (0, \infty)$.

In Assumption A.1(r), I allow both discrete and continuous random variables and require no moment condition. The mixing condition governs smoothness of $f(\omega, u, v)$. It ensures the existence of the generalized partial rth derivative of $f(\omega, u, v)$ with respect to ω,

$$f^{(r,0,0)}(\omega, u, v) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} |j|^r \sigma_j(u, v) e^{-ij\omega}. \quad (6)$$

Note that (6) is not the same as the ordinary rth partial derivative of $f(\omega, u, v)$ with respect to ω. If r is even, then $f^{(r,0,0)}(\omega, u, v) = (-1)^{r/2} \partial^r f(\omega, u, v)/\partial \omega^r$, but if r is odd, then there is no simple relation between the two.

Assumption A.2 ensures the existence of the integral (5). Assumption A.3 is standard for $k(\cdot)$. It allows $k(\cdot)$ with bounded or unbounded support. Examples include the truncated, Bartlett, Daniell, Parzen, quadratic-spectral (QS), and Tukey kernels (Priestley 1981). Assumption A.4 is a smoothness condition on $k(\cdot)$ at 0; q is called the "characteristic exponent" of $k(\cdot)$. For Bartlett kernel, $q = 1$; for Daniell, Parzen, QS, and Tukey kernels, $q = 2$. Zurbenko (1986) considered kernels with $q > 2$.

Theorem 1. (a) Suppose that Assumptions A.1(2) and A.2–A.3 hold, and $p = cn^\lambda$ for $c \in (0, \infty)$ and $\lambda \in (0, 1)$. Then $\text{IMSE}(\hat{f}_n, f) \to 0$.

(b) If, in addition, Assumptions A.1(q) and A.4 hold and $p = cn^{1/(2q+1)}$, then

$$\lim_{n\to\infty} n^{2q/(2q+1)} \text{IMSE}(\hat{f}_n, f) = ck_2 \, \text{Re} \int \int_{-\pi}^{\pi} f(\omega, u, -u) f(\omega, v, -v) \, d\omega \, dW(u, v) + (k^{(q)}/c^q)^2 \int \int_{-\pi}^{\pi} |f^{(q,0,0)}(\omega, u, v)|^2 \, d\omega \, dW(u, v). \quad (7)$$

The right side of (7) consists of the variance and bias-squared components of $\hat{f}_n(\omega, u, v)$. The rate $p = cn^{1/(2q+1)}$ balances the variance and squared bias, so that the optimal convergence rate $n^{-[2q/(2q+1)]}$ is attained for $\text{IMSE}(\hat{f}_n, f)$. Such an optimal rate depends on smoothness of $f(\omega, u, v)$ and smoothness of $k(\cdot)$ at 0. For the Bartlett kernel, this rate is $n^{-2/3}$; for the Daniell, Parzen, QS, and Tukey kernels, it is $n^{-4/5}$.

In practice, the choice of tuning constant c is crucial, as it governs smoothness of $\hat{f}_n(\omega, u, v)$. The IMSE criterion provides a basis for choosing an optimal c in (7). It can be obtained by setting to 0 the derivative of (7) with respect to c:

$$c_0 = \left\{ \frac{2q(k^{(q)})^2}{k_2} \frac{\int\int_{-\pi}^{\pi} |f^{(q,0,0)}(\omega, u, v)|^2 \times d\omega \, dW(u, v)}{\text{Re} \int\int_{-\pi}^{\pi} f(\omega, u, -u) \times f(\omega, v, -v) \, d\omega \, dW(u, v)} \right\}^{1/(2q+1)} \quad (8)$$

This delivers the theoretically optimal bandwidth $p_0 = c_0 n^{1/(2q+1)}$ that minimizes (5) asymptotically. Because $f(\omega, u, v)$ and $f^{(q,0,0)}(\omega, u, v)$ are unknown, p_0 is infeasible. One can, however, plug in (8) the following "pilot" estimators of $f(\omega, u, v)$ and $f^{(q,0,0)}(\omega, u, v)$ based on a preliminary bandwidth $\bar{p}$:

$$\bar{f}_n(\omega, u, v) = \frac{1}{2\pi} \sum_{j=1-n}^{n-1} (1 - |j|/n)^{1/2} \bar{k}(j/\bar{p}) \hat{\sigma}_j(u, v) e^{-ij\omega} \quad (9)$$

and

$$\bar{f}_n^{(q,0,0)}(\omega,u,v) = \frac{1}{2\pi}\sum_{j=1-n}^{n-1}(1-|j|/n)^{1/2}\bar{k}(j/\bar{p})|j|^q\hat{\sigma}_j(u,v)e^{-ij\omega}, \quad (10)$$

where $\bar{k}(\cdot)$ is a kernel not necessarily the same as that used in (4) and satisfying the following.

Assumption A.5. $\bar{k}:\mathbb{R}\to[-1,1]$ is symmetric and is continuous at 0 and all but a finite number of points, with $\bar{k}(0)=1, \int_0^\infty |z|^{2q}\bar{k}^2(z)\,dz<\infty$ and $|\bar{k}(z)|\le C|z|^{-\bar{b}}$ for large z, where $\bar{b}>q+\frac{1}{2}$ and q is as in Assumption A.4.

This yields a consistent estimator for c_0:

$$\hat{c}_0 = \left\{\frac{2q(k^{(q)})^2}{k_2}\,\frac{\int\int_{-\pi}^{\pi}|\bar{f}_n^{(q,0,0)}(\omega,u,v)|^2\times d\omega\,dW(u,v)}{\mathrm{Re}\int\int_{-\pi}^{\pi}\bar{f}_n(\omega,u,-u)\times\bar{f}_n(\omega,v,-v)\,d\omega\,dW(u,v)}\right\}^{1/(2q+1)}$$

$$= \left\{\frac{2q(k^{(q)})^2}{k_2}\,\frac{\sum_{j=1-n}^{n-1}(n-|j|)\bar{k}^2(j/\bar{p})|j|^{2q}\times\int|\hat{\sigma}_j(u,v)|^2\,dW(u,v)}{\sum_{j=1-n}^{n-1}(n-|j|)\bar{k}^2(j/\bar{p})\times\mathrm{Re}\int\hat{\sigma}_j(u,-u)\hat{\sigma}_j(v,-v)\times dW(u,v)}\right\}^{1/(2q+1)}, \quad (11)$$

where the last equality follows by Parseval's identity.

Theorem 2. Suppose that Assumptions A.1$(\max(2,q))$, A.2, and A.5 hold, and $\bar{p}\equiv\bar{p}(n)\to\infty, \bar{p}^{2q+1}/n\to 0$. Then $\hat{c}_0\to c_0$ and $\hat{p}_0/p_0\to 1$ in probability, where $\hat{p}_0=\hat{c}_0 n^{1/(2q+1)}$.

The estimator $\hat{c}_0$ delivers a data-dependent bandwidth $\hat{p}_0$, which provides the choice of a lag order for a given sample size n and is asymptotically optimal in terms of IMSE despite the fact that $f(\omega,u,v)$ is unknown. Note that $\hat{p}_0$ is real valued, but the effect of integer clipping of $\hat{p}_0$ is likely to be negligible. Nonparametric plug-in methods for selecting bandwidths are not uncommon in smoothed nonparametric estimation. Such methods do not completely avoid arbitrariness in choosing a preliminary bandwidth. There is, however, some evidence in both probability and spectral density estimation (see, e.g., Newey and West 1994; Silverman 1986, p. 58) that the final choice of a bandwidth is not very sensitive to the choice of a preliminary bandwidth. In an empirical application in Section 7, I study the effect of choosing different $\bar{p}$ on both $\hat{p}_0$ and $\hat{f}_n(\omega,u,v)$. It is found that $\hat{p}_0$ is not very sensitive to the choice of $\bar{p}$ for some commonly used kernels, and that related test statistics are robust to the choice of $\bar{p}$ over a relatively wide range of $\bar{p}$.

3. HYPOTHESIS TESTING IN TIME SERIES

To illustrate the scope and merits of the generalized spectrum. I apply $\hat{f}_n(\omega,u,v)$ and its derivatives to test generic serial dependence and hypotheses on various specific aspects of serial dependence. Detecting various forms of serial dependence has been a long-standing problem. A feature of the present approach is that all of the proposed tests for various hypotheses of interest are derived from a unified framework.

Under serial independence of $\{X_t\}, f(\omega,u,v)$ becomes a constant function of frequency ω,

$$f_0(\omega,u,v)=\frac{1}{2\pi}\,\sigma_0(u,v), \quad \forall\,\omega\in[-\pi,\pi], \quad (12)$$

where $\sigma_0(u,v)=\varphi(u+v)-\varphi(u)\varphi(v)$. This can be estimated consistently by

$$\hat{f}_0(\omega,u,v)=\frac{1}{2\pi}\,\hat{\sigma}_0(u,v), \quad \forall\,\omega\in[-\pi,\pi]. \quad (13)$$

To test serial independence, one can compare $\hat{f}_n(\omega,u,v)$ and $\hat{f}_0(\omega,u,v)$ via, for example, an L_2 norm. Such tests can detect all pairwise dependencies. They are useful, for example, in identifying nonlinear time series models, in testing the random walk hypothesis, and in situations where serial dependence is of unknown form. In addition to generic serial dependence, one may also be interested in testing hypotheses on various specific aspects of serial dependence, such as serial uncorrelatedness, martingale, conditional homoscedasticity, conditional symmetry, and conditional homokurtosis. Such hypotheses have their own rights. For example, market efficiency hypothesis is often equivalent to a martingale hypothesis. The analyticity of the complex-valued exponential function allows one to test these hypotheses by comparing the derivative estimators,

$$\hat{f}_n^{(0,m,l)}(\omega,u,v) = \frac{1}{2\pi}\sum_{j=1-n}^{n-1}(1-|j|/n)^{1/2}k(j/n)\hat{\sigma}_j^{(m,l)}(u,v)e^{-ij\omega} \quad (14)$$

and

$$\hat{f}_0^{(0,m,l)}(\omega,u,v)=\frac{1}{2\pi}\,\hat{\sigma}_0^{(m,l)}(u,v), \quad (15)$$

where $\hat{\sigma}_j^{(m,l)}(u,v)=\partial^{m+l}\hat{\sigma}_j(u,v)/\partial^m u\partial^l v$. As is illustrated in Section 4, (14) can only capture certain aspects of serial dependence, depending on the order (m,l). The order to choose is dictated by the aspect of serial dependence in which one is interested. When no prior information about possible alternatives is available, for example, one may be interested in generic serial dependence and set $(m,l)=(0,0)$. Alternatively, one can set $(m,l)=(1,0)$ to test the martingale hypothesis, or set $(m,l)=(1,1)$ to test serial correlation. Note that for $m,l>0$, proper moment conditions on $\{X_t\}$ are needed to ensure the existence of derivative $f^{(0,m,l)}(\omega,u,v)=\partial^{m+l}f(\omega,u,v)/\partial u^m\partial v^l$; see Assumption A.6 later.

Consider a squared L_2 norm between (14) and (15); namely,

$$L_2^2(\hat{f}_n^{(0,m,l)},\hat{f}_0^{(0,m,l)}) = \int\int_{-\pi}^{\pi}|\hat{f}_n^{(0,m,l)}(\omega,u,v)-\hat{f}_0^{(0,m,l)}(\omega,u,v)|^2\,d\omega\,dW(u,v)$$

$$= \left(\frac{\pi}{2}\right)^{-1} \int \left[\sum_{j=1}^{n-1} k^2(j/p)(1-j/n) |\hat{\sigma}_j^{(m,l)}(u,v)|^2 \right] \times dW(u,v), \quad (16)$$

where the second equality follows by Parseval's identity. I integrate out (ω, u, v) over $[-\pi, \pi] \times \mathbb{R}^2$ rather than use finitely many gridpoints. This ensures a reasonable omnibus test. Note that numerical integration over frequency ω is not needed in (16). One could also use divergence measures other than the L_2 norm, but they would generally involve numerical integrations over ω as well as (u, v).

My test statistic is a properly standardized version of (16):

$$M(m,l,p) = \left\{ \int \left[\sum_{j=1}^{n-1} k^2(j/p)(n-j) |\hat{\sigma}_j^{(m,l)}(u,v)|^2 \right] \times dW(u,v) - \hat{C}_0^{(m,l)} \sum_{j=1}^{n-1} k^2(j/p) \right\} \div \left[\hat{D}_0^{(m,l)} \sum_{j=1}^{n-2} k^4(j/p) \right]^{1/2}, \quad (17)$$

where

$$\hat{C}_0^{(m,l)} = \int \hat{\sigma}_0^{(m,m)}(u,-u) \hat{\sigma}_0^{(l,l)}(v,-v)\, dW(u,v)$$

and

$$\hat{D}_0^{(m,l)} = \int \{ |\hat{\sigma}_0^{(m,m)}(u,u') \hat{\sigma}_0^{(l,l)}(v,v')|^2 + |\hat{\sigma}_0^{(m,m)}(u,-u') \times \hat{\sigma}_0^{(l,l)}(v,-v')|^2 \}\, dW(u,v)\, dW(u',v').$$

Note that $\hat{D}_0^{(m,l)}$ involves a four-dimensional integration. For simplicity, one can use

$$W(u,v) = W_1(u) W_2(v), \quad (18)$$

where the $W_j : \mathbb{R} \to \mathbb{R}^+$ weight sets symmetric about 0 equally. This is analogous to a multiplicative symmetric kernel function commonly used in multivariate smoothed nonparametric estimation (Silverman 1986). With the use of (18), (17) simplifies to

$$M(m,l,p) = \left\{ \int \left[\sum_{j=1}^{n-1} k^2(j/p)(n-j) |\hat{\sigma}_j^{(m,l)}(u,v)|^2 \right] \times dW_1(u)\, dW_2(v) - \hat{C}_0^{(m,l)} \sum_{j=1}^{n-1} k^2(j/p) \right\} \div \left[\hat{D}_0^{(m,l)} \sum_{j=1}^{n-2} k^4(j/p) \right]^{1/2}, \quad (19)$$

where

$$\hat{C}_0^{(m,l)} = \int \hat{\sigma}_0^{(m,m)}(u,-u)\, dW_1(u) \int \hat{\sigma}_0^{(l,l)}(v,-v)\, dW_2(v)$$

and

$$\hat{D}_0^{(m,l)} = 2 \int |\hat{\sigma}_0^{(m,m)}(u,u')|^2\, dW_1(u)\, dW_1(u') \times \int |\hat{\sigma}_0^{(l,l)}(v,v')|^2\, dW_2(v)\, dW_2(v').$$

Now $\hat{D}_0^{(m,l)}$ involves only two-dimensional integrations.

4. TEST STATISTICS

The choices of (m,l) and $\{W_1(u), W_2(v)\}$ provide much flexibility in capturing various aspects of serial dependence. I now illustrate how different choices of (m,l) and $\{W_1(u), W_2(v)\}$ deliver tests for various hypotheses of interest. Throughout, I assume that $W_0 : \mathbb{R} \to \mathbb{R}^+$ is nondecreasing and weights set symmetric about 0 equally.

4.1 Testing Generic Serial Dependence

Put $(m,l) = (0,0), W_1(\cdot) = W_2(\cdot) = W_0(\cdot)$. Then

$$M(0,0,p) = \left\{ \int \left[\sum_{j=1}^{n-1} k^2(j/p)(n-j) |\hat{\sigma}_j(u,v)|^2 \right] \times dW_0(u)\, dW_0(v) - \hat{C}_0^{(0,0)} \sum_{j=1}^{n-1} k^2(j/p) \right\} \div \left[\hat{D}_0^{(0,0)} \sum_{j=1}^{n-2} k^4(j/p) \right]^{1/2}, \quad (20)$$

where

$$\hat{C}_0^{(0,0)} = \left[\int \hat{\sigma}_0(u,-u)\, dW_0(u) \right]^2$$

and

$$\hat{D}_0^{(0,0)} = 2 \left[\int |\hat{\sigma}_0(u,u')|^2\, dW_0(u)\, dW_0(u') \right]^2.$$

This tests generic serial dependence, which is useful when no prior information about the alternative is available. For any continuous and increasing $W_0(\cdot)$ with unbounded support, (20) is consistent against all pairwise dependencies. When $\varphi_j(u,v)$ is analytic, which requires a finite moment-generating function of (X_t, X_{t-j}), (20) is consistent against all pairwise serial dependencies even if $W_0(\cdot)$ has bounded support. This follows because two distributions are identical if and only if their analytic characteristic functions coincide for all (u,v) in a neighborhood of the origin. For a nonanalytic characteristic function, bounded support may lead to inconsistent tests. This follows because two distinct distributions can have the same characteristic functions on a finite interval (Chung 1974, pp. 184–185).

There have been a number of tests for generic serial dependence using various nonparametric time domain approaches, including those by Brock et al. (1991), Chan and Tran (1992), Delgado (1996), Hjellvik and Tjøstheim (1996), Hong (1998), Robinson (1991), and Skaug and Tjøstheim (1993a, b, 1996). These tests are based on correlation dimension in the chaotic theory, smoothed density estimation, and EDFs. Nearly all of these tests deal with serial dependence of a finite order and thus are not consistent against all pairwise dependencies. Pinkse (1998) considered a test for first-order serial dependence by using an upper bound of $\int |\hat{\sigma}_1(u,v)|^2\, dW(u,v)$ rather than the ECF directly.

Like the EDF, the ECF involves no smoothed nonparametric estimation at each lag j. It is well known (Feigin and Heathcote 1976) that ECF- and EDF-based tests for goodness of fit have different powers in the iid context: EDF-based tests have good power against shifts in mean but low power against changes in scale, whereas ECF-based tests have omnibus power against the two. A time series analog is documented for ECF- and EDF-based tests of serial independence. Hong (1998) and Skaug and Tjøstheim (1993a) found that EDF-based tests have good power against linear time series but low power against ARCH. A simulation study of an earlier version of this article found that the ECF-based test $M(0,0,p)$ has omnibus power against both time series processes. Moreover, EDF-based tests may have no power against some types of serial dependence of any order when X_t is a discrete random variable (see Skaug and Tjøstheim 1993a for an example).

4.2 Testing Serial Correlation of Unknown Form

Put $(m,l) = (1,1)$, and suppose that $W_1(\cdot)$ and $W_2(\cdot)$ have a density function $\delta: \mathbb{R} \to \mathbb{R}^+$, where $\delta(\cdot)$ is the Dirac delta function; that is, $\delta(u) = 0$ if and only if $u \neq 0$ and $\int \delta(u)\, du = 1$. Straightforward algebra yields

$$\int |\hat{\sigma}_j^{(1,1)}(u,v)|^2\, \delta(u)\, \delta(v)\, du\, dv = |\hat{\sigma}_j^{(1,1)}(0,0)|^2 = \hat{R}_1^2(j),$$

$$\hat{C}_0^{(1,1)} = \left[\int \hat{\sigma}^{(1,1)}(u,-u)\, \delta(u)\, du\right]^2 = \hat{R}_1^2(0),$$

and

$$\hat{D}_0^{(1,1)} = 2\left[\int |\hat{\sigma}_0^{(1,1)}(u,u')|^2\, \delta(u)\, \delta(u')\, du\, du'\right]^2 = 2\hat{R}_1^4(0),$$

where

$$\hat{R}_1(j) = (n-j)^{-1} \sum_{t=j+1}^{n} [X_t - \bar{X}_1(j)][X_{t-j} - \bar{X}_2(j)],$$

$$\bar{X}_1(j) = (n-j)^{-1} \sum_{t=j+1}^{n} X_t$$

and

$$\bar{X}_2(j) = (n-j)^{-1} \sum_{t=j+1}^{n} X_{t-j}.$$

It follows that

$$M(1,1,p) = \left[\sum_{j=1}^{n-1} k^2(j/p)(n-j)\hat{\rho}_1^2(j) - \sum_{j=1}^{n-1} k^2(j/p)\right] \div \left[2\sum_{j=1}^{n-2} k^4(j/p)\right]^{1/2}, \quad (21)$$

where $\hat{\rho}_1(j) = \hat{R}_1(j)/\hat{R}_1(0)$ is the sample autocorrelation function of $\{X_t\}_{t=1}^n$. This is a slightly modified version of Hong's (1996) spectral density test, which is consistent against serial correlation of unknown form. When uniform weighting (i.e., $k(z) = 1$ if $|z| \le 1$ and 0 otherwise) is used, then (21) delivers generalized portmanteau tests of Box and Pierce (1970). As shown by Hong (1996), however, uniform weighting is not optimal when p is large. Note that Anderson (1993), Andrews and Ploberger (1996), and Durlauf (1991) also considered consistent tests for serial correlation of unknown form.

4.3 Testing for the Martingale Hypothesis

To test the martingale difference sequence hypothesis that $E[(X_t - \mu)|X_{t-j}, j > 0] = 0$ a.s., where $\mu = E(X_t)$, one can put $(m,l) = (1,0)$. Suppose that $W_1(\cdot)$ has a Dirac $\delta(\cdot)$ density and $W_2(\cdot) = W_0(\cdot)$. Then

$$M(1,0,p) = \left\{\int \left[\sum_{j=1}^{n-1} k^2(j/p)(n-j)|\hat{\sigma}_j^{(1,0)}(0,v)|^2\right] \times dW_0(v) - \hat{C}_0^{(1,0)} \sum_{j=1}^{n-1} k^2(j/p)\right\} \div \left[\hat{D}_0^{(1,0)} \sum_{j=1}^{n-2} k^4(j/p)\right]^{1/2}, \quad (22)$$

where

$$\hat{C}_0^{(1,0)} = \hat{R}_1(0) \int \hat{\sigma}_0(v,-v)\, dW_0(v),$$

$$D_0^{(1,0)} = 2\hat{R}_1^2(0) \int |\hat{\sigma}_0(v,v')|^2\, dW_0(v)\, dW_0(v'),$$

and

$$\hat{\sigma}_j^{(1,0)}(0,v) = (n-j)^{-1} \sum_{t=j+1}^{n} X_t[e^{ivX_{t-j}} - \hat{\varphi}_j(0,v)].$$

Note that $\hat{\sigma}_j^{(1,0)}(0,v)$ is consistent for $\sigma_j^{(1,0)}(0,v) = E[(X_t - \mu)e^{ivX_{t-j}}]$. This delivers a test for the martingale hypothesis in spirit similar to Bierens's (1982) and Bierens and Ploberger's (1997) integrated conditional moment tests for model specification. Unlike the Bierens and Ploberger tests, whose null limit distributions are a sum of weighted chi-squared variables with weights depending on the unknown data-generating process and thus cannot be tabulated, (22) has a convenient one-sided N(0, 1) limit distribution. It has power against alternatives that have zero autocorrelation but a nonzero mean conditional on the X_{t-j}, such as some bilinear and nonlinear moving average processes. This is illustrated in an empirical application to Deutschemark exchange rates in Section 7, where tests based on the conventional spectrum fail. Therefore, (22) may be more useful than conventional spectral tests, which have been widely applied to test the martingale hypothesis in practice (see, e.g., Durlauf 1991). Hinich and Patterson (1992) also proposed a test for the martingale hypothesis using the bispectrum.

4.4 Testing for Linear ARCH

Put $(m,l)=(2,2)$ and suppose that both $W_1(\cdot)$ and $W_2(\cdot)$ have a Dirac $\delta(\cdot)$ density function. Then

$$M(2,2,p)=\left[\sum_{j=1}^{n-1}k^2(j/p)(n-j)\hat{\rho}_2^2(j)-\sum_{j=1}^{n-1}k^2(j/p)\right] \div\left[2\sum_{j=1}^{n-2}k^4(j/p)\right]^{1/2}, \quad (23)$$

where $\hat{\rho}_2(j)$ is the sample autocorrelation function of $\{X_t^2\}$ defined analogously as $\hat{\rho}_1(j)$. This is consistent against all linear ARCH processes; that is, all autocorrelations in X_t^2. When uniform weighting for $k(\cdot)$ is used, (23) delivers a generalized version of McLeod and Li's (1983) test. The tests of Engle (1982) and Granger and Anderson (1978) are similar in spirit to McLeod and Li's test. They are asymptotically equivalent under proper conditions (Granger and Teräsvirta 1993, pp. 93–94). Again, nonuniform weighting for $k(\cdot)$ is asymptotically more powerful than uniform weighting for large p.

4.5 Testing for Nonlinear ARCH

Put $(m,l)=(2,0)$. Suppose that $W_1(\cdot)$ has a $\delta(\cdot)$ density and $W_2(\cdot)=W_0(\cdot)$. Then

$$M(2,0,p)=\left\{\int\left[\sum_{j=1}^{n-1}k^2(j/p)(n-j)|\hat{\sigma}_j^{(2,0)}(0,v)|^2\right]\times dW_0(v)-\hat{C}_0^{(2,0)}\sum_{j=1}^{n-1}k^2(j/p)\right\} \div\left[\hat{D}_0^{(2,0)}\sum_{j=1}^{n-2}k^4(j/p)\right]^{1/2}, \quad (24)$$

where

$$\hat{C}_0^{(2,0)}=\hat{R}_2(0)\int\hat{\sigma}_0(v,-v)\,dW_0(v),$$

$$\hat{D}_0^{(2,0)}=2\hat{R}_2^2(0)\int|\hat{\sigma}_0(v,v')|^2\,dW_0(v)\,dW_0(v'),$$

and

$$\hat{\sigma}_j^{(2,0)}(0,v)=(n-j)^{-1}\sum_{t=j+1}^{n}X_t^2[e^{ivX_{t-j}}-\hat{\varphi}_j(0,v)].$$

Note that $\hat{\sigma}_j^{(2,0)}(0,v)$ is consistent for $\sigma_j^{(2,0)}(0,v)=E[(X_t^2-EX_t^2)e^{ivX_{t-j}}]$. Besides linear ARCH, (24) can also detect the nonlinear ARCH processes for which $\{X_t^2\}$ has zero autocorrelation but a nonconstant mean conditional on the $X_{t-j}, j>0$. An example is an ARCH process whose conditional variance follows a chaotic tent map process (see, e.g., Brock et al., 1991, pp. 11–12).

4.6 Testing for Conditional Symmetry

Put $(m,l)=(3,0)$. Suppose that $W_1(\cdot)$ has a $\delta(\cdot)$ density and $W_2(\cdot)=W_0(\cdot)$. Then

$$M(3,0,p)=\left\{\int\left[\sum_{j=1}^{n-1}k^2(j/p)(n-j)|\hat{\sigma}_j^{(3,0)}(0,v)|^2\right]\times dW_0(v)-\hat{C}_0^{(3,0)}\sum_{j=1}^{n-1}k^2(j/p)\right\} \div\left[\hat{D}_0^{(3,0)}\sum_{j=1}^{n-2}k^4(j/p)\right]^{1/2}, \quad (25)$$

where

$$\hat{C}_0^{(3,0)}=\hat{R}_3(0)\int\hat{\sigma}_0(v,-v)\,dW_0(v),$$

$$\hat{D}_0^{(3,0)}=2\hat{R}_3^2(0)\int|\hat{\sigma}_0(v,v')|^2\,dW_0(v)\,dW_0(v'),$$

and

$$\hat{\sigma}_j^{(3,0)}(0,v)=(n-j)^{-1}\sum_{t=j+1}^{n}X_t^3[e^{ivX_{t-j}}-\hat{\varphi}_j(0,v)],$$

where $\hat{R}_3(j)$ is the sample autocovariance function of $\{X_t^3\}_{t=1}^n$ defined analogously as $\hat{R}_1(j)$. This checks whether the conditional third moment of $\{X_t\}$ is time-varying, the so-called conditional heteroclicticity. It can be used to test conditional symmetry. This is important in finance, for example, because conditional skewness has important implications for portfolio selections. It can also be used to test whether the innovation distribution of an ARCH process is symmetric, which is crucial for efficiency of quasi-maximum likelihood estimators (Newey and Steigerward 1996).

4.7 Testing for Conditional Heterokurtosis

Put $(m,l)=(4,0)$. Suppose that $W_1(\cdot)$ has a $\delta(\cdot)$ density and $W_2(\cdot)=W_0(\cdot)$. Then

$$M(4,0,p)=\left\{\int\left[\sum_{j=1}^{n-1}k^2(j/p)(n-j)|\hat{\sigma}_j^{(4,0)}(0,v)|^2\right]\right.$$

$$\times\, dW_0(v) - \hat{C}_0^{(4,0)} \sum_{j=1}^{n-1} k^2(j/p) \Bigg\}$$

$$\div \left[\hat{D}_0^{(4,0)} \sum_{j=1}^{n-2} k^4(j/p) \right]^{1/2}, \tag{26}$$

where

$$\hat{C}_0^{(4,0)} = \hat{R}_4(0) \int \hat{\sigma}_0(v, -v)\, dW_0(v),$$

$$\hat{D}_0^{(4,0)} = 2\hat{R}_4^2(0) \int |\hat{\sigma}_0(v, v')|^2\, dW_0(v)\, dW_0(v'),$$

and

$$\hat{\sigma}_j^{(4,0)}(0, v) = (n-j)^{-1} \sum_{t=j+1}^{n} X_t^4 [e^{ivX_{t-j}} - \hat{\varphi}_j(0, v)],$$

where $\hat{R}_4(j)$ is the sample autocovariance function of $\{X_t^4\}_{t=1}^n$ defined analogously as $\hat{R}_1(j)$. This checks whether the conditional fourth-order moment of $\{X_t\}$ is time varying.

Tests for other moments can be derived analogously. For example, the choice of $(m, l) = (2, 1)$ with both $W_1(\cdot)$ and $W_2(\cdot)$ having a Dirac $\delta(\cdot)$ density will test whether the volatility of X_t depends on the levels of its own past history, the so-called "leverage effect." The choice of $(m, l) = (1, 2)$ with both $W_1(\cdot)$ and $W_2(\cdot)$ having a Dirac $\delta(\cdot)$ density will test whether X_t depends on its volatility, the so-called "ARCH-in-mean effect."

5. ASYMPTOTIC NULL DISTRIBUTION

To derive the asymptotic distribution of $M(m, l, p)$, I impose the following conditions.

Assumption A.6. $E|X_1|^{4d} < \infty$, where $d = \max(m, l)$.

Assumption A.7. $D_0^{(m,l)} > 0$, where $D_0^{(m,l)}$ is as $\hat{D}_0^{(m,l)}$ in (17) with $\hat{\sigma}_0(u, v)$ replaced by $\sigma_0(u, v)$.

Assumption A.6 ensures that the marginal characteristic function $\varphi(\cdot)$ has bounded continuous derivatives up to order $4d$ (Chung 1974, thm. 6.4.1). Note that no moment of $\{X_t\}$ is needed to test generic serial dependence $(m, l = 0)$. To test serial correlation $(m, l = 1)$ or the martingale hypothesis $(m = 1, l = 0), EX_1^4 < \infty$ is needed. To test ARCH $(m = 2, l = 0, 2), EX_1^8 < \infty$ is needed. Assumption A.7 rules out the possibility that $\sigma_0^{(m,m)}(u, v) = \sigma_0^{(l,l)}(u, v) = 0$ for all (u, v) on the support of $W(u, v)$, as can arise, for example, when the distribution of X_t is degenerate. This ensures that $M(m, l, p)$ is well behaved.

Theorem 3. Suppose that Assumptions A.2–A.3 and A.6–A.7 hold, and $p = cn^\lambda$ for $c \in (0, \infty)$ and $\lambda \in (0, 1)$. If $\{X_t\}$ is iid, then $M(m, l, p) \to \mathrm{N}(0, 1)$ in distribution.

Theorem 3 allows a wide range of admissible rates for nonstochastic bandwidth p. In practice, one may like to choose p via such data-driven methods as $\hat{p}_0$ in Theorem 2, which lets data themselves determine an appropriate lag order. To justify the use of a data-dependent bandwidth $\hat{p}$, I impose a Lipschitz condition on $k(\cdot)$. This includes most commonly used nonuniform kernels, but rules out the truncated kernel.

Assumption A.8. $|k(z_1) - k(z_2)| \le C|z_1 - z_2|$ for any $z_1, z_2 \in \mathbb{R}$.

Theorem 4. Suppose that Assumptions A.2–A.3 and A.6–A.8 hold, and $\hat{p}$ is data-dependent such that $\hat{p}/p = 1 + O_P(p^{-[(3\beta/2)-1]})$ for some $\beta > (2b - \frac{1}{2})/(2b - 1)$, where b is as in Assumption A.3, and nonstochastic bandwidth $p = cn^\lambda$ for $c \in (0, \infty)$ and $\lambda \in (0, (2b-1)/(2b - \frac{1}{2}))$. If $\{X_t\}$ is iid, then $M(m, l, \hat{p}) - M(m, l, p) \to 0$ in probability, and $M(m, l, \hat{p}) \to \mathrm{N}(0, 1)$ in distribution.

The use of $\hat{p}$ has an asymptotically negligible impact on the limit distribution of $M(m, l, \hat{p})$ provided that $\hat{p}/p \to 1$ in probability at a proper rate. For kernels with bounded support (e.g., the Bartlett, Parzen, and Tukey kernels), $\beta > 1$ suffices, because $b = \infty$. For the QS kernel $(b = 2), \beta > \frac{7}{6}$ suffices. For the Daniell kernel $(b = 1), \beta > \frac{3}{2}$ suffices. These conditions are mild. With some additional conditions, Theorem 2 could be extended to obtain a convergence rate for $\hat{p}_0$ to satisfy Theorem 4, but I do not do so for space considerations.

The asymptotic normality of Theorems 3 and 4 gives quick and convenient inference. For small and finite samples, however, asymptotic approximation may not be reasonable. In this case, bootstrap or permutation can be used. These methods are ideally suited to the present situation and can be expected to yield reasonably accurate sizes (Skaug and Tjøstheim 1996). I use the bootstrap in an empirical application in Section 7.

6. ASYMPTOTIC POWER

To state the consistency theorems for $M(m, l, p)$ and $M(m, l, \hat{p})$, I impose the following condition.

Assumption A.9. $E|X_1|^{4d\nu} < \infty$, where $d = \max(m, l)$ and $\nu > 1$ is as in Assumption A.1.

This is slightly stronger than Assumption A.6 if m or $l \ge 1$. For $(m, l) = (0, 0)$, no moment condition is required. Following a reasoning analogous to Chung (1974, thm. 6.4.1), it is easy to show that Assumption A.9 ensures that $\varphi_j^{(m,l)}(u, v)$ exists and is bounded and continuous in $\mathbb{R}^2$. This, along with Assumption A.1, ensures the existence of $f^{(0,m,l)}(\omega, u, v)$.

Theorem 5. Suppose that Assumptions A.1(2), A.2–A.3, A.7, and A.9 hold, and $p = cn^\lambda$ for $c \in (0, \infty)$ and $\lambda \in (0, 1)$. Then, in probability,

$$\frac{p^{1/2}}{n} M(m, l, p) \to$$

$$\frac{\pi}{2} \iint_{-\pi}^{\pi} |f^{(0,m,l)}(\omega, u, v) - f_0^{(0,m,l)}(\omega, u, v)|^2\, d\omega\, dW(u, v)$$

$$\div \left[D_0^{(m,l)} \int_0^\infty k^4(z)\, dz \right]^{1/2}.$$

Thus $M(m, l, p)$ has asymptotic power 1 whenever the L_2 norm between $f^{(0,m,l)}(\omega, u, v)$ and $f_0^{(0,m,l)}(\omega, u, v)$ is positive. For concreteness and succinctness, I focus the discussion here on $(m, l) = (0, 0)$ only. In this case $M(0, 0, p)$ has asymptotic power 1 whenever

$$\frac{\pi}{2} \int\int_{-\pi}^{\pi} |f(\omega, u, v) - f_0(\omega, u, v)|^2 \, d\omega \, dW(u, v) = \sum_{j=1}^{\infty} \int |\sigma_j(u, v)|^2 \, dW(u, v) \quad (27)$$

is positive. Suppose that $W(u, v)$ is continuous and increasing with unbounded support, $\int |\sigma_j(u, v)|^2 \, dW(u, v) = 0$ if and only if X_t and X_{t-j} are independent. It follows that (27) is 0 if and only if X_t and X_{t-j} are independent for all $j \neq 0$. Therefore, $M(0, 0, p)$ is consistent against all pairwise dependencies. A price for this is that $M(0, 0, p)$ can only detect a class of local alternatives converging to the null hypothesis at a rate slightly slower than $n^{-1/2}$. Nearly all of the existing nonparametric tests for serial dependence deal with a finite order and thus are not consistent for all pairwise dependencies. Note that Theorem 5 applies to both discrete and continuous random variables. This differs from some EDF-based tests (see, e.g., Delgado 1996; Hong 1998; Skaug and Tjøstheim 1993a), which may have no power against some types of serial dependence of any order when X_t is a discrete random variable.

The test $M(m, l, p)$ involves the choice of $k(\cdot)$. An interesting question is whether there exists an optimal $k(\cdot)$ that maximizes asymptotic power. Intuitively, because $\sigma_j^{(m,l)}(u, v) \to 0$ as $j \to \infty$, it seems more efficient to give more weights to lower-order lags. This is confirmed by the following asymptotic power analysis. Consider the class of kernels

$$\mathbb{K}(\tau) = \left\{ k(\cdot)\colon\ k(\cdot) \text{ satisfies Assumption A.3}, \; k^{(2)} = \frac{\tau^2}{2} > 0, K(\xi) \geq 0 \, \forall \, \xi \in \mathbb{R} \right\}, \quad (28)$$

where $K(\xi) = (2\pi)^{-1} \int k(z) e^{i\xi z} \, dz$ is the Fourier transform of $k(\cdot)$. This class of kernels is often considered in deriving the optimal kernel for spectral estimation (see, e.g., Andrews 1991; Priestley 1981). It includes the Daniell, Parzen, and QS kernels but rules out the truncated, Bartlett, and Tukey kernels. Theorem 6 shows that the Daniell kernel maximizes the asymptotic power of $M(m, l, p)$ over $\mathbb{K}(\tau)$ in terms of Bahadur's (1960) asymptotic slope criterion, which is suitable for large sample sizes.

Theorem 6. Under the conditions of Theorem 5, the Daniell kernel $k_D(z) = \sin(\sqrt{3}\tau z)/(\sqrt{3}\tau z), z \in \mathbb{R}$, maximizes the asymptotic power of $M(m, l, p)$ over $\mathbb{K}(\tau)$ in terms of Bahadur's asymptotic slope criterion.

Intuitively, Bahadur's (1960) asymptotic slope is the rate of minus twice the logarithm of the asymptotic significance level of a test statistic that goes to infinity under the fixed alternative. The larger the asymptotic slope, the more powerful the test. (For more discussion, see Bahadur 1960 and Geweke 1981.) It is well known that there is no good reason to use the Daniell kernel for spectral estimation (Hannan 1970, pp. 279–280). Theorem 6, however, implies that for hypothesis testing, the Daniell kernel is more powerful than the QS kernel. The latter can be shown, using the methods of Andrews (1991) and Priestley (1981), to be optimal over the same class $\mathbb{K}(\tau)$ for the estimation of $\hat{f}_n(\omega, u, v)$ in terms of the IMSE criterion. Note that the optimality of the Daniell kernel holds given any choices of function $W(u, v)$ and derivative orders (m, l), as well as any data-generating process satisfying Assumptions A.1 and A.9. This generalizes the results of Hong (1996), who considered the special case $(m, l) = (1, 1)$ where a Parzen (1957) estimator for the conventional spectrum is used.

Although the optimality of the Daniell kernel is of theoretical interest, some kernels in $\mathbb{K}(\tau)$ have rather similar asymptotic efficiency to the Daniell kernel. For example, the Bahadur asymptotic relative efficiencies of the Daniell kernel to the Parzen and QS kernels are $(1.0961)^{1/(2-\lambda)}$ and $(1.0079)^{1/(2-\lambda)}$. Therefore, the choices of $k(\cdot)$ in $\mathbb{K}(\tau)$ may have little impact on power, as is confirmed in an empirical application in Section 7.

Finally, I justify the use of a data-dependent $\hat{p}$ under the alternative hypothesis.

Theorem 7. Suppose that Assumptions A.1(2), A.2–A.3, and A.7–A.9 hold, and $\hat{p}/p = 1 + O_P(p^{-\beta})$ for some $\beta > 0$, where the nonstochastic bandwidth $p = cn^{\lambda}$ for $c \in (0, 1)$ and $\lambda \in (0, 1)$. Then in probability, $(p^{1/2}/n)\{M(m, l, \hat{p}) - M(m, l, p)\} \to 0$ and

$$\frac{p^{1/2}}{n} M(m, l, \hat{p}) \to \frac{\pi}{2} \int\int_{-\pi}^{\pi} |f^{(0,m,l)}(\omega, u, v) - f_0^{(0,m,l)}(\omega, u, v)|^2 \, d\omega \, dW(u, v) \div \left[D_0^{(m,l)} \int_0^{\infty} k^4(z) \, dz \right]^{1/2}.$$

7. EMPIRICAL APPLICATION

I now apply the proposed tools to Deutschemark exchange rates, one of the most actively traded currencies in the foreign exchange market. The data are the weekly spot rates measured in units of U.S. dollars from the first week of 1976:1 to the last week of 1995:11, with a total of 1,039 observations. These are interbank closing spot rates S_t on Wednesdays and are obtained from the Bloomberg L.P. Using the Wednesday data avoids the so-called weekend effect. Also, very few holidays occur on Wednesday; for these holidays, data on the following Thursdays are used. I use the logarithmic difference $X_t = 100 \ln(S_t/S_{t-1})$.

The interest here is in testing the random-walk hypothesis and the martingale hypothesis as well as exploring a possible nonlinear dependence structure of X_t. It has been hypothesized (e.g., Meese and Rogoff 1983) that exchange rates approximately follow a martingale process, so that the future changes are essentially unpredictable on the basis of public information. Such a hypothesis has been verified using correlation tests and various data (see, e.g., Bollerslev 1990; Brock et al., 1991; Engle, Ito, and Lin 1990; Meese and Rogoff 1983).

I first estimate $f(\omega, u, v)$ by $\hat{f}_n(\omega, u, v)$ via the plug-in method of Theorem 2 with $\bar{k}(z) = (1-|z|)1(|z| \leq 1)$, where $1(\cdot)$ is the indicator function. To examine the effect of the choice of preliminary bandwidth $\bar{p}$, I consider $\bar{p}$ in the range of 6–15. I also consider four kernels: Bartlett, Daniell, Parzen, and QS, the latter three belonging to $\mathbb{K}(\tau)$ with $\tau = \pi/\sqrt{3}$. The data-dependent bandwidth $\hat{p}_0$ is invariant to τ. Figure 1 reports $\hat{p}_0$ and statistic $M(0,0,\hat{p}_0)$ using various preliminary bandwidths $\bar{p}$. I use $W(u,v) = \Phi(u)\Phi(v)$, where $\Phi(\cdot)$ is the CDF N(0, 1). The numerical integrations involved in $\hat{p}_0$ and $M(0,0,\hat{p}_0)$ are calculated using the Gauss–Legendre quadratures INTQUAD1 and INTQUAD2 in GAUSS software. I set the order of integration equal to 24; it is found that numerical integrations are almost identical for larger integration orders. It takes about 9.5 minutes to compute $M(0,0,\hat{p}_0)$ for all four kernels $k(\cdot)$ and 10 preliminary bandwidths $\bar{p}$ together on a Pentium 500 personal computer. Figure 1 shows that given each $\bar{p}, \hat{p}_0$ is essentially the same for the Daniell, Parzen, and QS kernels but is significantly larger for the Bartlett kernel. This accords with the faster rate of the optimal bandwidth $p_0 \propto n^{1/3}$ for the Bartlett kernel. Given each $k(\cdot), \hat{p}_0$ depends on $\bar{p}$ in an increasing way, but by a small marginal rate. This is true especially for the Daniell, Parzen, and QS kernels. Thus the choice of $\hat{p}_0$ is not very sensitive to the choice of $\bar{p}$. The statistic $M(0,0,\hat{p}_0)$ is relatively robust to the choice of $\bar{p}$. The Daniell, Parzen, and QS kernels deliver rather similar values for $M(0,0,\hat{p}_0)$ for each $\bar{p}$. The Bartlett kernel gives a slightly larger value for $M(0,0,\hat{p}_0)$, but this is not inconsistent with Theorem 6, because the Bartlett kernel is outside $\mathbb{K}(\tau)$.

Figure 2 reports statistics $M(m,l,\hat{p}_0)$ and their p values for various (m,l). I consider the preliminary bandwidth $\bar{p}$ in the range of 6–15 again, but use the Daniell kernel only. Note that the data-dependent $\hat{p}_0$ depends on (m,l), because the plug-in derivative estimators $\bar{f}_n^{(0,m,l)}(\omega,u,v)$ and $\bar{f}_n^{(q,m,l)}(\omega,u,v)$ depend on (m,l). In addition to asymptotic p values, I also compute bootstrap p values for $M(m,l,\hat{p}_0)$. To compute each bootstrap statistic, I first use a GAUSS-386 random number generator to randomly draw a bootstrap sample $\{X_t^o\}_{t=1}^n$, with replacement, from the observed sample $\{X_t\}_{t=1}^n$. Based on $\{X_t^o\}_{t=1}^n$, I then compute $\hat{p}_0$ and $M(m,l,\hat{p}_0)$ for each (m,l). The bootstrap p-values are based on 500 iterations. Figure 2 shows that both asymptotic and bootstrap p-values of $M(0,0,\hat{p}_0)$ are essentially 0 for any given $\bar{p}$, strongly rejecting the random-walk hypothesis. In contrast, although asymptotic and bootstrap p-values of the conventional spectrum-based test $M(1,1,\hat{p}_0)$ differ from each other, both of them are well above .2 for all choices of $\bar{p}$, suggesting that $\{X_t\}$ is serially uncorrelated. This, however, does not necessarily imply that $\{X_t\}$ is a martingale difference sequence (as most existing studies conclude), because $\{X_t\}$ may have zero autocorrelation but a nonzero conditional mean. Indeed, the martingale test $M(1,0,\hat{p}_0)$ strongly rejects the martingale hypothesis in terms of both asymptotic and bootstrap p-values. Therefore, the change of Deutschemark exchange rates, though serially uncorrelated, has a nonzero mean conditional on its past history. This evidence might be useful in modeling exchange rates. The sharp differences of

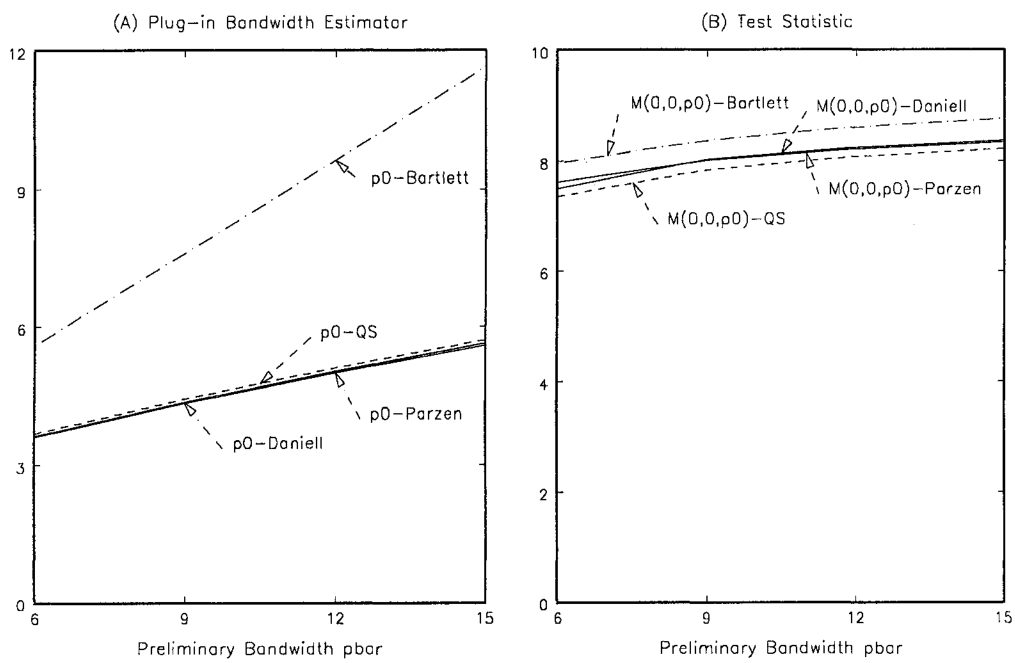

Figure 1. (a) The Data-Driven Bandwidth $\hat{p}_0$ as a Function of the Preliminary Bandwidth $\bar{p}$. The figure displays how sensitive the value of $\hat{p}_0$ is to the choice of $\bar{p}$. Four kernels—Bartlett, Daniell, Parzen, and Quadratic-Spectral kernels are considered. (b) The test statistic M(0, 0, $\hat{p}_0$) as a function of the preliminary bandwidth $\bar{p}$. The figure displays how sensitive the value of M(0, 0, $\hat{p}_0$) is to the choice of $\bar{p}$, which affects M(0, 0, $\hat{p}_0$) via the data-driven bandwidth $\hat{p}_0$. Four kernels—Bartlett, Daniell, Parzen, and Quadratic-Spectral kernels are considered.

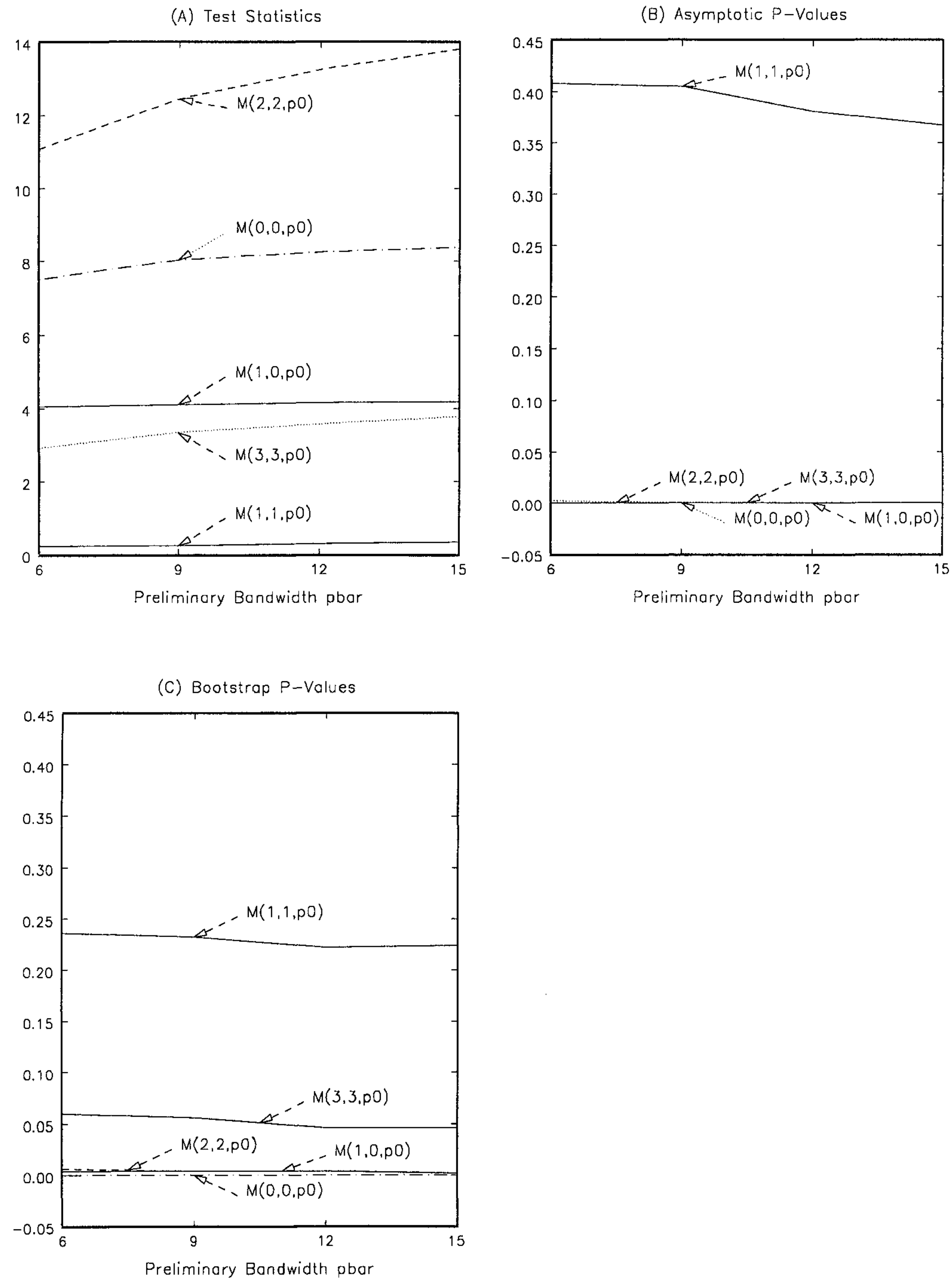

Figure 2. (a) The Test Statistic $M(m, l, \hat{p}_0)$ as a Function of the Preliminary Bandwidth $\bar{p}$. The figure shows how sensitive the value of $M(m, l, \hat{p}_0)$ is to the choice of $\bar{p}$. Only the Daniell kernel is considered. (b) The asymptotic p-value of the test statistic $M(m, l, \hat{p}_0)$ as a function of the preliminary bandwidth $\bar{p}$. The figure shows how sensitive the asymptotic p-value of $M(m, l, \hat{p}_0)$ is to the choice of $\bar{p}$. Only the Daniell kernel is considered. (c) The bootstrap p-value of the test statistic $M(m, l, \hat{p}_0)$ as a function of the preliminary bandwidth $\bar{p}$. The figure displays how sensitive the bootstrap p-value of $M(m, l, \hat{p}_0)$ is to the choice of $\bar{p}$. The bootstrap p-values are computed using 500 bootstrap samples. Only the Daniell kernel is used.

$M(0,0,\hat{p}_0)$ and $M(1,0,\hat{p}_0)$ from $M(1,1,\hat{p}_0)$ clearly illustrate that the generalized spectrum can detect the nonlinear dependence structure that would be missed by the conventional spectrum. Finally, both asymptotic and bootstrap p-values of the ARCH test $M(2,2,\hat{p}_0)$ are essentially 0, suggesting a strong ARCH effect. On the other hand, although asymptotic p-values of the conditional symmetry test $M(3,3,\hat{p}_0)$ are 0, its bootstrap p-values are only around .05, suggesting marginally significant conditional skewness.

8. CONCLUSIONS

Using the empirical characteristic function, I have proposed a generalized spectral tool suitable for both linear and nonlinear time series analysis. The generalized spectral density can capture all pairwise dependencies, requires no moment condition, and is well defined for both discrete and continuous random variables. The conventional spectral density can be derived as a special case by properly differentiating the generalized spectral density. I established consistency of a class of Parzen's (1957) kernel-type estimators for the generalized spectrum and derived their optimal convergence rates using the IMSE criterion. A data-dependent asymptotically optimal bandwidth was introduced, which provides the choice of a lag order for a given sample sizes. The kernel estimators and their derivatives were applied to construct tests for generic serial dependence and hypotheses on various specific aspects of serial dependence. The latter include serial uncorrelatedness, martingale, conditional homoscedasticity, conditional symmetry, and conditional homokurtosis. All of the proposed tests are derived from a unified framework and have a convenient null asymptotic one-sided N(0, 1) distribution. An empirical application to Deutschemark exchange rates highlighted the approach.

APPENDIX: PROOFS

Proof of Theorem 1

I show (b) only; the proof of (a) is similar to that of Theorem 2. Define the pseudoestimator

$$\tilde{f}_n(\omega,u,v) \equiv (2\pi)^{-1}\sum_{|j|<n}(1-|j|/n)^{1/2}k(j/p)\tilde{\sigma}_j(u,v)e^{-ij\omega}, \quad (A.1)$$

where $\tilde{\sigma}_j(u,v) \equiv (n-|j|)^{-1}\sum_{t=|j|+1}^{n}\psi_t(u)\psi_{t-|j|}(v)$ and $\psi_t(u) \equiv e^{iuX_t}-\varphi(u)$. Then $\mathrm{IMSE}(\hat{f}_n,f) = \mathrm{IMSE}(\tilde{f}_n,f) + \mathrm{IMSE}(\hat{f}_n,\tilde{f}_n) + 2\mathrm{Re}(R_n)$, where $R_n \equiv E\int\int_{-\pi}^{\pi}[\hat{f}_n(\omega,u,v) - \tilde{f}_n(\omega,u,v)][\tilde{f}_n^*(\omega,u,v)-f^*(\omega,u,v)]\,d\omega\,dW$ and $dW \equiv dW(u,v)$. It suffices to show Theorems A.1 and A.2, which imply $R_n = o\{\mathrm{IMSE}(\tilde{f}_n,f)\}$ by the Cauchy–Schwarz inequality.

Theorem A.1

Suppose that the conditions of Theorem 1 hold; then

$$n^{2q/(2q+1)}\mathrm{IMSE}(\tilde{f}_n,f) \to ck_2\mathrm{Re}\int\int_{-\pi}^{\pi}f(\omega,u,-u)f(\omega,v,-v)\,d\omega\,dW + \left(\frac{k^{(q)}}{c^q}\right)^2\int\int_{-\pi}^{\pi}|f^{(q,0,0)}(\omega,u,v)|^2\,d\omega\,dW.$$

Theorem A.2

Suppose that the conditions of Theorem 1 hold; then $\mathrm{IMSE}(\hat{f}_n,\tilde{f}_n) = o\{\mathrm{IMSE}(\tilde{f}_n,f)\}$.

Proof of Theorem A.1

Put $A_n(\omega,u,v) \equiv \tilde{f}_n(\omega,u,v) - E\tilde{f}_n(\omega,u,v)$ and $B_n(\omega,u,v) \equiv E\tilde{f}_n(\omega,u,v) - f(\omega,u,v)$. Then

$$\mathrm{IMSE}(\tilde{f}_n,f) = \int\int_{-\pi}^{\pi}E|A_n(\omega,u,v)|^2\,d\omega\,dW + \int\int_{-\pi}^{\pi}|B_n(\omega,u,v)|^2\,d\omega\,dW. \quad (A.2)$$

Given (A.1) and $E\tilde{\sigma}_j(u,v) = \sigma_j(u,v), 2\pi B_n(\omega,u,v) = B_{1n}(\omega,u,v) + B_{2n}(\omega,u,v) - B_{3n}(\omega,u,v)$, where

$$B_{1n}(\omega,u,v) \equiv \sum_{|j|<n}[k(j/p)-1]\sigma_j(u,v)e^{-ij\omega},$$

$$B_{2n}(\omega,u,v) \equiv \sum_{|j|<n}[(1-|j|/n)^{1/2}-1]k(j/p)\sigma_j(u,v)e^{-ij\omega},$$

and

$$B_{3n}(\omega,u,v) \equiv \sum_{|j|\geq n}\sigma_j(u,v)e^{-ij\omega}.$$

Using the mixing inequality $|\sigma_j(u,v)| \leq C\alpha(j)^{(\nu-1)/\nu}$ and Assumption A.1(q),

$$\sum_{j=-\infty}^{\infty}|j|^q\sup_{(u,v)\in\mathbb{R}^2}|\sigma_j(u,v)| \leq C\sum_{j=-\infty}^{\infty}|j|^q\alpha(j)^{(\nu-1)/\nu}<\infty. \quad (A.3)$$

By (6), the triangle inequality, Assumption A.4, (A.3), and dominated convergence,

$$|p^qB_{1n}(\omega,u,v)+2\pi k^{(q)}f^{(q,0,0)}(\omega,u,v)| \leq \sum_{|j|<n}\left|\frac{1-k(j/p)}{|j/p|^q}-k^{(q)}\right||j|^q|\sigma_j(u,v)| + k^{(q)}\sum_{|j|\geq n}|j|^q|\sigma_j(u,v)| \to 0 \quad (A.4)$$

uniformly in $(\omega,u,v)\in[-\pi,\pi]\times\mathbb{R}^2$ as $p\to\infty$. Using analogous reasoning,

$$\sup_{(\omega,u,v)}|nB_{2n}(\omega,u,v)| = O(1) \text{ and}$$

$$\sup_{(\omega,u,v)}|n^qB_{3n}(\omega,u,v)| \leq \sum_{|j|\geq n}|j|^q\alpha(j)^{(\nu-1)/\nu}\to 0, \quad (A.5)$$

where the supremum is taken over $[-\pi,\pi]\times\mathbb{R}^2$. Combining (A.4)–(A.5) and $p = cn^{1/(2q+1)}$ yields $|p^qB_n(\omega,u,v) + k^{(q)}f^{(q,0,0)}(\omega,u,v)| \to 0$ uniformly in $(\omega,u,v)\in[-\pi,\pi]\times\mathbb{R}^2$. Hence

$$p^{2q}\int\int_{-\pi}^{\pi}|B_n(\omega,u,v)|^2\,d\omega\,dW \to (k^{(q)})^2\int\int_{-\pi}^{\pi}|f^{(q,0,0)}(\omega,u,v)|^2\,d\omega\,dW. \quad (A.6)$$

Next, I consider the variance term in (A.2). By Parseval's identity,

$$\int\int_{-\pi}^{\pi}E|A_n(\omega,u,v)|^2\,d\omega\,dW = (2\pi)^{-1}\sum_{|j|<n}(1-|j|/n)k^2(j/p) \times\int E|\tilde{\sigma}_j(u,v)-\sigma_j(u,v)|^2\,dW.$$

Let $\kappa_{uvu^*0^*}(0,j,\tau,s)$ denote the fourth-order cumulant of the complex-valued stationary process $\{\psi_t(u),\psi_{t-j}(v),\psi_{t-\tau}^*(u),\psi_{t-s}^*(v)\}$. Using the definitions of $\tilde{\sigma}_j(u,v)$ and $\sigma_j(u,v)$ and expressing the moments by cumulants by well-known formulas (Hannan 1970, 5.1, p. 23, for real-valued processes), straightforward algebra yields

$$\begin{aligned}(n-|j|)E|\tilde{\sigma}_j(u,v)-\sigma_j(u,v)|^2 &= \mathrm{Re}\sum_{|\tau|<n-|j|}\{1-|\tau|/(n-|j|)\}\{\sigma_\tau(u,-u)\sigma_\tau(v,-v)\\ &+\sigma_{|j|+|\tau|}(u,-v)\sigma_{|j|-|\tau|}(-u,v)1(|\tau|\leq|j|)\\ &+\sigma_{|j|+|\tau|}(u,-v)\sigma_{|\tau|-|j|}(v,-u)1(|\tau|>|j|)\\ &+\kappa_{uvu^*v^*}(0,|j|,|\tau|,|j|+|\tau|)\}.\end{aligned} \quad (A.7)$$

Hence $2\pi n \int\int_{-\pi}^{\pi} E|A_n(\omega,u,v)|^2\,d\omega\,dW = A_{1n} + A_{2n} + A_{3n}$, where

$$A_{1n} \equiv \sum_{|j|<n} k^2(j/p) \sum_{|\tau|<n-|j|} \{1-|\tau|/(n-|j|)\} \times \operatorname{Re}\int \sigma_\tau(u,-u)\sigma_\tau(v,-v)\,dW,$$

$$A_{2n} \equiv \sum_{|j|<n} k^2(j/p) \sum_{|\tau|<n-|j|} \{1-|\tau|/(n-|j|)\} \times \operatorname{Re}\int \sigma_{|\tau|+|j|}(u,-v)\{\sigma_{|j|-|\tau|}(-u,v)1(|\tau|\le|j|) + \sigma_{|\tau|-|j|}(v,-u)1(|\tau|>|j|)\}\,dW,$$

and

$$A_{3n} \equiv \sum_{|j|<n} k^2(j/p) \sum_{|\tau|<n-|j|} \{1-|\tau|/(n-|j|)\} \times \operatorname{Re}\int \kappa_{uvu^*v^*}(0,|j|,|\tau|,|j|+|\tau|)\,dW.$$

Because $p^{-1}\sum_{|j|<n} k^2(j/p) \to k_2$ and $0 \le \operatorname{Re}\sum_{\tau=-\infty}^{\infty}\sigma_\tau(u,-u)\sigma_\tau(v,-v) < \infty$,

$$p^{-1}A_{1n} \to k_2 \operatorname{Re}\sum_{\tau=-\infty}^{\infty}\int \sigma_\tau(u,-u)\sigma_\tau(v,-v)\,dW = 2\pi k_2 \operatorname{Re}\int\int_{-\pi}^{\pi} f(\omega,u,-u)f(\omega,v,-v)\,d\omega\,dW \quad (A.8)$$

by dominated convergence. Also, given $|k(\cdot)|\le 1$, (A.3), and Assumption A.1(2),

$$|A_{2n}| \le \left\{\sum_{j=-\infty}^{\infty}\sup_{(u,v)\in\mathbb{R}^2}|\sigma_j(u,v)|\right\}^2 \le C \quad (A.9)$$

and

$$|A_{3n}| \le \sum_{j=-\infty}^{\infty}\sum_{s=-\infty}^{\infty}\sum_{t=-\infty}^{\infty}\sup_{(u,v)\in\mathbb{R}^2}|\kappa_{uvu^*v^*}(0,j,s,t)| \le C, \quad (A.10)$$

where the fourth-order cumulants are absolutely summable uniformly in (u,v) given Assumption A.1(2) (Andrews 1991, lem. 1). Note that uniformity in (u,v) follows from the fact that $\{\psi_t(\cdot)\}$ is a bounded random variable. Combining (A.7)–(A.10) and $p\to\infty$ yields

$$(n/p)\int\int_{-\pi}^{\pi} E|A_n(\omega,u,v)|^2\,d\omega\,dW \to k_2 \operatorname{Re}\int\int_{-\pi}^{\pi} f(\omega,u,-u)f(\omega,v,-v)\,d\omega\,dW. \quad (A.11)$$

The desired result follows from (A.2), (A.6), (A.11), and $p = cn^{1/(2q+1)}$.

Proof of Theorem A.2

By straightforward algebra,

$$(n-|j|)^2\{\hat\sigma_j(u,v)-\tilde\sigma_j(u,v)\} = -\left\{\sum_{t=|j|+1}^{n}\psi_t(u)\right\}\left\{\sum_{t=|j|+1}^{n}\psi_{t-|j|}(v)\right\}. \quad (A.12)$$

It follows from the Cauchy–Schwarz inequality that

$$(n-|j|)^4 E|\hat\sigma_j(u,v)-\tilde\sigma_j(u,v)|^2 \le \left\{E\left|\sum_{t=|j|+1}^{n}\psi_t(u)\right|^4 E\left|\sum_{t=|j|+1}^{n}\psi_{t-|j|}(u)\right|^4\right\}^{1/2} \le C(n-|j|)^2, \quad (A.13)$$

where $E|\sum_{t=|j|+1}^{n}\psi_t(u)|^4 \le C(n-|j|)^2$ given Assumption A.1(2).

Now, by Parseval's identity, (A.13), and $\int dW < \infty$ from Assumption A.2,

$$\operatorname{IMSE}(\hat f_n,\tilde f_n) = (2\pi)^{-1}\sum_{|j|<n}(1-|j|/n)k^2(j/p) \times E\int|\hat\sigma_j(u,v)-\tilde\sigma_j(u,v)|^2\,dW \le Cn^{-1}\sum_{|j|<n}a_{nj} = O(p/n^2) = o\{\operatorname{IMSE}(\tilde f_n,f)\}, \quad (A.14)$$

where and hereafter $a_{nj}\equiv(n-|j|)^{-1}k^2(j/p)$ and

$$\sum_{|j|<n}a_{nj} \le \{p/(n-l)\}p^{-1}\sum_{|j|\le l}k^2(j/p) + C^2n^{-1}p^{2b} \times \sum_{l<|j|<n}\{(n-|j|)^{-1}+|j|^{-1}\}|j|^{-2b+1} = O(p/n) \quad (A.15)$$

given $|k(z)|\le C|z|^{-b}$ for large $z, p = cn^\lambda$ for $\lambda\in(0,1)$, and choosing $l = p(\ln n)^{1/(2b-1)}$.

Proof of Theorem 2

The convergence $\hat c_0 \to^p c_0$ follows if $\operatorname{IMSE}(\bar f_n^{(q,0,0)}, f^{(q,0,0)}) \to 0$ and $\operatorname{IMSE}(\bar f_n, f)\to 0$, the latter implying $\int\int_{-\pi}^{\pi}|\bar f_n(\omega,u,-u) - f(\omega,u,-u)|^2\,d\omega\,dW \to^p 0$ and $\int\int_{-\pi}^{\pi}|\bar f_n(\omega,v,-v)-f(\omega,v,-v)|^2\,d\omega\,dW \to^p 0$. Thus it suffices to show that $\operatorname{IMSE}(\bar f_n^{(r,0,0)}, f^{(r,0,0)}) \to 0$ for $r\in[0,q]$. Define the pseudoestimator

$$\dot f_n^{(r,0,0)}(\omega,u,v) \equiv (2\pi)^{-1}\sum_{|j|<n}(1-|j|/n)^{1/2}\bar k(j/\bar p)\tilde\sigma_j(u,v)|j|^r e^{-ij\omega}.$$

By the C_r inequality,

$$\frac{1}{4}\operatorname{IMSE}(\bar f_n^{(r,0,0)}, f^{(r,0,0)}) \le \operatorname{IMSE}(\dot f_n^{(r,0,0)}, E\dot f_n^{(r,0,0)}) + \operatorname{IMSE}(E\dot f_n^{(r,0,0)}, f^{(r,0,0)}) + \operatorname{IMSE}(\bar f_n^{(r,0,0)}, \dot f_n^{(r,0,0)}). \quad (A.16)$$

Now, by Parseval's identity, $(n-|j|)E|\tilde\sigma_j(u,v)-\sigma_j(u,v)|^2 \le C$ from (A.7)–(A.10), and Assumption A.5, for the first term in (A.16),

$$\operatorname{IMSE}(\dot f_n^{(r,0,0)}, E\dot f_n^{(r,0,0)}) \le C(\bar p^{2r+1}/n)\bar p^{-1}\sum_{|j|<n}\bar k^2(j/\bar p)|j/\bar p|^{2r} = O(\bar p^{2r+1}/n). \quad (A.17)$$

Similarly, by (6), Parseval's identity, (A.3), Assumption A.5, and dominated convergence,

$$\begin{aligned} &\mathrm{IMSE}(E\dot{f}_n^{(r,0,0)}, f^{(r,0,0)}) \\ &= \sum_{|j|<n} \{(1-|j|/n)^{1/2}\bar{k}(j/\bar{p})-1\}^2 \\ &\quad \times \int |j|^{2r}|\sigma_j(u,v)|^2\, dW \\ &\quad + \sum_{|j|>n} \int |j|^{2r}|\sigma_j(u,v)|^2\, dW \to 0 \end{aligned} \quad (A.18)$$

as $\bar{p} \to \infty$. Finally, for the last term of (A.16), by Parseval's identity and (A.13),

$$\begin{aligned} &\mathrm{IMSE}(\bar{f}_n^{(r,0,0)}, \dot{f}_n^{(r,0,0)}) \\ &\le \frac{C\bar{p}^{2r+1}}{n^2}\left\{\bar{p}^{-1}\sum_{|j|<n}(1-|j|/n)^{-1}\bar{k}^2(j/\bar{p})|j/\bar{p}|^{2r}\right\} \\ &= O(\bar{p}^{2r+1}/n^2), \end{aligned} \quad (A.19)$$

where the quantity inside $\{\cdot\}$ is $O(1)$ given Assumption A.5 by reasoning similar to (A.15). The desired result follows from combining (A.16)–(A.19) and $\bar{p}^{2r+1}/n \to 0$.

Proof of Theorem 3

I first show that replacing $\hat{\sigma}_j(u,v)$ with $\tilde{\sigma}_j(u,v)$ in $M(m,l,p)$ is a higher-order effect. Write

$$\begin{aligned} &|\hat{\sigma}_j^{(m,l)}(u,v)|^2 - |\tilde{\sigma}_j^{(m,l)}(u,v)|^2 \\ &= |\hat{\sigma}_j^{(m,l)}(u,v) - \tilde{\sigma}_j^{(m,l)}(u,v)|^2 \\ &\quad + 2\mathrm{Re}\{[\hat{\sigma}_j^{(m,l)}(u,v) - \tilde{\sigma}_j^{(m,l)}(u,v)]\tilde{\sigma}_j^{(m,l)}(u,v)^*\}. \end{aligned}$$

Under serial independence of $\{X_t\}$ and Assumption A.6, $(n-j)^2E|\hat{\sigma}_j^{(m,l)}(u,v) - \tilde{\sigma}_j^{(m,l)}(u,v)|^2 \le C$ and $(n-j)E|\tilde{\sigma}_j^{(m,l)}(u,v)|^2 \le C$ uniformly in $(u,v) \in \mathbb{R}^2$ can be obtained. It follows from Markov's inequality, the Cauchy–Schwarz inequality, $\int dW < \infty$, and (A.15), that

$$\begin{aligned} &\int \left\{\sum_{j=1}^{n-1} k^2(j/p)(n-j)[|\hat{\sigma}_j^{(m,l)}(u,v)|^2 - |\tilde{\sigma}_j^{(m,l)}(u,v)|^2]\right\} dW \\ &\qquad = O_P(p/n^{1/2}) = o_P(p^{1/2}) \end{aligned} \quad (A.20)$$

given $p/n \to 0$. Now define $V_{tsj}^{(m,l)}(u,v) \equiv C_{tsj}^{(m,l)}(u,v) + C_{stj}^{(m,l)}(u,v)^*$, where

$$C_{tsj}^{(m,l)}(u,v) \equiv \psi_t^{(m)}(u)\psi_s^{(m)}(u)^*\psi_{t-j}^{(l)}(v)\psi_{s-j}^{(l)}(v)^*.$$

Because $C_{tsj}^{(m,l)}(u,v) = C_{stj}^{(m,l)}(u,v)^*$, $V_{tsj}^{(m,l)}(u,v)$ is real-valued and is symmetric in t and s; namely, $V_{tsj}^{(m,l)}(u,v) = V_{stj}^{(m,l)}(u,v)$. This leads to

$$\begin{aligned} &\sum_{j=1}^{n-1} k^2(j/p)(n-j)|\tilde{\sigma}_j^{(m,l)}(u,v)|^2 \\ &= \sum_{j=1}^{n-1} a_{nj}\left[\sum_{t=j+1}^{n} C_{ttj}^{(m,l)}(u,v) + \sum_{t=j+2}^{n}\sum_{s=j+1}^{t-1} V_{tsj}^{(m,l)}(u,v)\right] \\ &\equiv \hat{C}^{(m,l)}(u,v) + \hat{V}^{(m,l)}(u,v), \quad \text{say.} \end{aligned} \quad (A.21)$$

I now consider the first term in (A.21). Observe that $\int C_{ttj}^{(m,l)}(u,v)\,dW$ and $\int C_{ssj}^{(m,l)}(u,v)\,dW$ are independent unless $t = s$ or $s \pm j$, and that $E\int C_{ttj}^{(m,l)}(u,v)\,dW = C_0^{(m,l)} \equiv \int \sigma_0^{(m,m)}(u,-u)\sigma_0^{(l,l)}(v,-v)\,dW$. It follows that $E\{\sum_{t=j+1}^n [\int C_{ttj}^{(m,l)}(u,v)\,dW - C_0^{(m,l)}]\}^2 \le C(n-j)$. Hence, by Markov's inequality, the Cauchy–Schwarz inequality, and (A.15),

$$\begin{aligned} &\int \hat{C}^{(m,l)}(u,v)\,dW - C_0^{(m,l)}\sum_{j=1}^{n-1}k^2(j/p) \\ &= \sum_{j=1}^{n-1} a_{nj}\left[\sum_{t=j+1}^{n}\int C_{ttj}^{(m,l)}(u,v)\,dW - C_0^{(m,l)}\right] \\ &= O_P(p/n^{1/2}). \end{aligned} \quad (A.22)$$

Combining (A.20)–(A.22) and putting $\hat{V}_n^{(m,l)} \equiv \int \hat{V}^{(m,l)}(u,v)\,dW$ yields

$$\begin{aligned} &\int \left[\sum_{j=1}^{n-1} k^2(j/p)(n-j)|\hat{\sigma}_j^{(m,l)}(u,v)|^2\right] dW \\ &\qquad = C_0^{(m,l)}\sum_{j=1}^{n-1}k^2(j/p) + \hat{V}_n^{(m,l)} + O_P(p/n^{1/2}). \end{aligned} \quad (A.23)$$

Now, by rearrangement of indices (t,s,j), one can write $\hat{V}_n^{(m,l)} = \sum_{t=3}^n V_{nt}^{(m,l)}$, where $V_{nt}^{(m,l)} \equiv \sum_{s=2}^{t-1}\sum_{j=1}^{s-1} a_{nj}\int V_{tsj}^{(m,l)}(u,v)\,dW$. Because $E[V_{nt}^{(m,l)}|\mathcal{F}_{t-1}] = 0$ under serial independence of $\{X_t\}$, where $\{\mathcal{F}_t\}$ is the sequence of σ fields consisting of $\{X_s, s \le t\}$, $\{V_{nt}^{(m,l)}, \mathcal{F}_t\}$ is an adapted martingale difference sequence. Because $\hat{V}_n^{(m,l)}$ is a triple sum of nonlinear functions of $(X_t, X_{t-j}, X_s, X_{s-j})$, it is very difficult to apply a martingale limit theorem to $\hat{V}_n^{(m,l)}$ directly. I first use a truncation argument to partition $\hat{V}_n^{(m,l)}$ into a sum over (t,s,j), where $t - s > g \in \mathbb{N}$ and $1 \le j \le g$ plus a remainder term. The remainder term, as shown in Theorem A.3, is of a smaller order of magnitude if $g \equiv g(n) \to \infty$ such that $g/p \to \infty$ and $g/n \to 0$:

$$\begin{aligned} \hat{V}_n^{(m,l)} &= \sum_{t=g+2}^{n}\sum_{s=1}^{t-g-1}\sum_{j=1}^{g} a_{nj}\int V_{tsj}^{(m,l)}(u,v)\,dW + o_P(p^{1/2}) \\ &\equiv \hat{V}_{ng}^{(m,l)} + o_P(p^{1/2}), \end{aligned} \quad (A.24)$$

where $\hat{V}_{ng}^{(m,l)}$ is a triple sum over (t,s,j), where $t - s > g$. Note that (X_t, X_{t-j}) and (X_s, X_{s-j}) are independent if $t - s > g$. This greatly simplifies verification of Brown's (1971) conditions. In fact, $\hat{V}_n^{(m,l)}$ is a sum of degenerate dependent U statistics of processes (X_t, X_{t-j}), where lag order j can be as large as $n-1$. Hjellvik and Tjøstheim (1996) and Skaug and Tjøstheim (1996) considered a degenerate U statistic of (X_t, X_{t-j}) for fixed and finite j.

With application of Brown's theorem, one can obtain (as shown in Theorem A.4)

$$\left[pD_0^{(m,l)}\int_0^\infty k^4(z)\,dz\right]^{-1/2}\hat{V}_{ng}^{(m,l)} \to^d \mathrm{N}(0,1), \quad (A.25)$$

where $D_0^{(m,l)}$ is as in Assumption A.7. Combining (A.23)–(A.25),

$$\left[pD_0^{(m,l)}\int_0^\infty k^4(z)\,dz\right]^{-1/2} \times \left\{\int\left[\sum_{j=1}^{n-1}k^2(j/p)(n-j)|\hat{\sigma}_j^{(m,l)}(u,v)|^2\right]dW\right.$$

$$- C_0^{(m,l)} \sum_{j=1}^{n-1} k^2(j/p) \Bigg\} \to^d \mathrm{N}(0,1). \quad \text{(A.26)}$$

Because $\hat{C}_0^{(m,l)} - C_0^{(m,l)} = O_P(n^{-1/2})$, as can be easily shown, and $\sum_{j=1}^{n-1} k^2(j/p) = O(p)$, replacing $C_0^{(m,l)}$ with $\hat{C}_0^{(m,l)}$ in (A.26) has asymptotically negligible impact given $p/n \to 0$. By Slutsky's theorem, one can also replace $pD_0^{(m,l)} \int_0^\infty k^4(z)\,dz$ with $\hat{D}_0^{(m,l)} \sum_{j=1}^{n-2} k^4(j/p)$, where $\hat{D}_0^{(m,l)} \to^p D_0^{(m,l)}$ and $p^{-1}\sum_{j=1}^{n-2} k^4(j/p) \to \int_0^\infty k^4(z)\,dz$. It follows that $M(m,l,p) \to^d$ N(0, 1). The proof will be completed provided that Theorems A.3 and A.4 are proven.

Theorem A.3

Let $\hat{V}_n^{(m,l)}$ and $\hat{V}_{ng}^{(m,l)}$ be defined as in (A.23)–(A.24), where $g \equiv g(n)$ is such that $g/p \to \infty, g/n \to 0$. Then $\hat{V}_n^{(m,l)} = \hat{V}_{ng}^{(m,l)} + o_P(p^{1/2})$.

Theorem A.4

Let $\hat{V}_{ng}^{(m,l)}$ be defined as in (A.24), and $g \equiv g(n)$ be such that $g/p \to \infty, g/n \to 0$. Then $[pD_0^{(m,l)} \int_0^\infty k^4(z)\,dz]^{-1/2}\hat{V}_{ng}^{(m,l)} \to^d$ N(0, 1).

Proof of Theorem A.3

I first partition $\hat{V}_n^{(m,l)}$ into sums over $1 \le j \le g$ and $g < j \le n-2$:

$$\hat{V}_n^{(m,l)} = \left(\sum_{j=1}^{g}\sum_{t=j+2}^{n}\sum_{s=j+1}^{t-1} + \sum_{j=g+1}^{n-2}\sum_{t=j+2}^{n}\sum_{s=j+1}^{t-1}\right) \times a_{nj}\int V_{tsj}^{(m,l)}(u,v)\,dW \equiv \hat{U}_n^{(m,l)} + \hat{R}_{1n}^{(m,l)}, \quad \text{(A.27)}$$

where $\hat{R}_{1n}^{(m,l)}$ is the contribution from the tail of $k(j/p)$. 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 $s < t \le n$,

$$\hat{U}_n^{(m,l)} = \left(\sum_{j=1}^{g}\sum_{t=2}^{n}\sum_{s=1}^{t-1} - \sum_{j=1}^{g}\sum_{s=1}^{j}\sum_{t=s+1}^{n}\right) \times a_{nj}\int V_{tsj}^{(m,l)}(u,v)\,dW \equiv \hat{W}_n^{(m,l)} - \hat{R}_{2n}^{(m,l)}, \quad \text{say.} \quad \text{(A.28)}$$

Furthermore, $\hat{W}_n^{(m,l)}$ can be partitioned into sums over $t-s > g$ and $t-s \le g$:

$$\hat{W}_n^{(m,l)} = \left(\sum_{j=1}^{g}\sum_{t=g+2}^{n}\sum_{s=1}^{t-g-1} + \sum_{j=1}^{g}\sum_{t=2}^{g+1}\sum_{s=1}^{t-1} + \sum_{j=1}^{g}\sum_{t=g+2}^{n}\sum_{s=t-g}^{t-1}\right) \times a_{nj}\int V_{tsj}^{(m,l)}(u,v)\,dW \equiv \hat{V}_{ng}^{(m,l)} + \hat{R}_{3n}^{(m,l)} + \hat{R}_{4n}^{(m,l)}, \quad \text{say,} \quad \text{(A.29)}$$

where $\hat{R}_{3n}^{(m,l)}$ and $\hat{R}_{4n}^{(m,l)}$ are sums over (t,s,j) with $t-s \le g$.

Combining (A.27)–(A.29), yields $\hat{V}_n^{(m,l)} = \hat{V}_{ng}^{(m,l)} + \hat{R}_{1n}^{(m,l)} - \hat{R}_{2n}^{(m,l)} + \hat{R}_{3n}^{(m,l)} + \hat{R}_{4n}^{(m,l)}$. It remains to show that $E[\hat{R}_{hn}^{(m,l)}]^2 = o(p)$ for $1 \le h \le 4$. First, by rearranging the indices (t,s,j) in $\hat{R}_{1n}^{(m,l)}$,

$$\hat{R}_{1n}^{(m,l)} = \sum_{t=g+3}^{n}\left[\sum_{s=g+2}^{t-1}\sum_{j=g+1}^{s-1} a_{nj}\int V_{tsj}^{(m,l)}(u,v)\,dW\right] \equiv \sum_{t=g+3}^{n} R_{1nt}^{(m,l)}, \quad \text{say,}$$

where $\{R_{1nt}^{(m,l)}, \mathcal{F}_t\}$ is a martingale difference sequence because $E(R_{1nt}^{(m,l)}|\mathcal{F}_{t-1}) = 0$. Also, by straightforward but tedious algebra, it can be shown that for $t > s_1 > j_1$ and $t > s_2 > j_2$,

$$E[V_{ts_1j_1}^{(m,l)}(u,v)V_{ts_2j_2}^{(m,l)}(u',v')] = \begin{cases} E[V_{ts_1j_1}^{(m,l)}(u,v)V_{ts_2j_2}^{(m,l)}(u',v')]\delta_{s_1,s_2} & \text{if } j_1 = j_2 \\ E[V_{ts_1j_1}^{(m,l)}(j_1)V_{ts_2j_2}^{(m,l)}(u,v')]\delta_{s_1,t-j_2}\delta_{s_2,t-j_1} & \text{if } j_1 \ne j_2, \end{cases}$$

where $\delta_{jl} = 1$ for $j = 1$ and $\delta_{jl} = 0$ for $j \ne l$. It follows that

$$\begin{aligned} E(\hat{R}_{1n}^{(m,l)})^2 &= \sum_{t=g+3}^{n} E\left[\sum_{s=g+2}^{t-1}\sum_{j=g+1}^{s-1} a_{nj}\int V_{tsj}^{(m,l)}(u,v)\,dW\right]^2 \\ &= \sum_{t=g+3}^{n}\sum_{s=g+2}^{t-1}\sum_{j=g+1}^{s-1} a_{nj}^2 \int E[V_{tsj}^{(m,l)}(u,v)V_{tsj}^{(m,l)}(u',v')]\,dW\,dW' \\ &\quad + 2\sum_{t=g+3}^{n}\sum_{j_2=g+2}^{t-1}\sum_{j_1=g+1}^{j_2-1} a_{nj_1}a_{nj_2} \\ &\quad \times \int E[V_{t,t-j_2,j_1}^{(m,l)}(u,v)V_{t,t-j_1,j_2}^{(m,l)}(u',v')]\,dW\,dW' \\ &\le C\sum_{j=g+1}^{n-1} k^4(j/p) + C(n-g)\left(\sum_{j=g+1}^{n-1} a_{nj}\right)^2 \\ &= o(p) + O(p^2/n), \end{aligned}$$

where $\sum_{j=g+1}^{n-1} k^4(j/p) = o(p)$ by Assumption A.3 and $g/p \to \infty$. Similarly, $E(\hat{R}_{2n}^{(m,l)})^2 = O(pg/n + p^2g/n^2), E(\hat{R}_{3n}^{(m,l)})^2 = O(pg^2/n^2 + p^2g/n^2)$, and $E(\hat{R}_{4n}^{(m,l)})^2 = O(pg/n^2 + p^2/n)$ can be obtained. All of these are $o(p)$ given $g/p \to \infty, g/n \to 0$. This finishes the proof.

Proof of Theorem A.4

Put $S_{nt}^{(m,l)} \equiv S_{1nt}^{(m,l)} + (S_{1nt}^{(m,l)})^*$, where

$$S_{1nt}^{(m,l)} \equiv \int \psi_t^{(m)}(u)\left[\sum_{j=1}^{g} a_{nj}\psi_{t-j}^{(l)}(v)G_{t-g-1,j}^{(m,l)}(u,v)\right]dW, \quad \text{(A.30)}$$

and $G_{t-g-1,j}^{(m,l)}(u,v) \equiv \sum_{s=1}^{t-g-1}\psi_s^{(m)}(u)^*\psi_{s-j}^{(l)}(v)^*$. Then write $\hat{V}_{ng}^{(m,l)} = \sum_{t=g+2}^{n} S_{nt}^{(m,l)}$. Because $\{S_{nt}^{(m,l)}, \mathcal{F}_t\}$ is an adapted martingale difference sequence, one can apply Brown's (1971) theorem. First, compute $\lambda_n^2(m,l) \equiv \mathrm{var}(\hat{V}_{ng}^{(m,l)})$. Observing that $\psi_t^{(m)}(\cdot)\psi_{t-j}^{(l)}(\cdot)$ and $\psi_s^{(m)}(\cdot)\psi_{s-j}^{(l)}(\cdot)$ are independent for $t-s > g$ and $1 \le j \le g$, one can obtain

$$\begin{aligned} \lambda_n^2(m,l) &= D_0^{(m,l)}(n-g)(n-g-1)\sum_{j=1}^{g} a_{nj}^2 \\ &= pD_0^{(m,l)}\int_0^\infty k^4(z)\,dz\{1+o(1)\} \quad \text{(A.31)} \end{aligned}$$

given $g/p \to \infty$, and $g/n \to 0$, where $p^{-1}\sum_{j=1}^{g} n^2 a_{nj}^2 \to \int_0^\infty k^4(z)\,dz$.

By Brown's theorem, $\hat{V}_{ng}^{(m,l)}/\lambda_n(m,l) \to^d \mathrm{N}(0,1)$ if

$$\lambda_n^{-2}(m,l)\sum_{t=g+2}^{n} E\{(S_{nt}^{(m,l)})^2 1[|S_{nt}^{(m,l)}| > \epsilon\lambda_n(m,l)]\} \to 0 \quad \forall\epsilon > 0 \quad \text{(A.32)}$$

and

$$\lambda_n^{-2}(m,l)\sum_{t=g+2}^{n} E[(S_{nt}^{(m,l)})^2|\mathcal{F}_{t-1}] \to^p 1. \quad \text{(A.33)}$$

Given (A.31), it suffices for (A.32) if $\sum_{t=g+2}^{n} E|S_{1nt}^{(m,l)}|^4 = o(p^2)$. From (A.30),

$$|S_{1nt}^{(m,l)}|^4 \le \left\{\int\left[E|\psi_t^{(m)}(u)|^4 E\left|\sum_{j=1}^{g} a_{nj}\psi_{t-j}^{(l)}(v)G_{t-g-1,j}^{(m,l)}(u,v)\right|^4\right]^{1/4} dW\right\}^4 \quad \text{(A.34)}$$

by Minkowski's inequality. Now compute the second expectation inside (A.34). Put

$$Q_{ntj}^{(m,l)}(u,v;u',v') \equiv \psi_{t-j}^{(l)}(v)G_{t-g-1,j}^{(m,l)}(u,v)\sum_{r=j+1}^{g} a_{nr}\psi_{t-r}^{(l)}(v')^* \times G_{t-g-1,r}^{(m,l)}(u',v')^*, \quad \text{(A.35)}$$

where $1 \le j \le g$. Then

$$\left|\sum_{j=1}^{g} a_{nj}\psi_{t-j}^{(l)}(v)G_{t-g-1,j}^{(m,l)}(u,v)\right|^4 \le 2\left|\sum_{j=1}^{g} a_{nj}^2|\psi_{t-j}^{(l)}(v)|^2|G_{t-g-1,j}^{(m,l)}(u,v)|^2\right|^2 + 8\left|\sum_{j=1}^{g-1} a_{nj}Q_{ntj}^{(m,l)}(u,v;u,v)\right|^2. \quad \text{(A.36)}$$

For the first term of (A.36), by Minkowski's inequality and independence between $\psi_{t-j}^{(l)}(\cdot)$ and $G_{t-g-1,r}^{(m,l)}(\cdot,\cdot), 1 \le j, r \le g$,

$$E\left|\sum_{j=1}^{g} a_n^2(j)|\psi_{t-j}^{(l)}(v)|^2|G_{t-g-1,j}^{(m,l)}(u,v)|^2\right|^2 \le Ct^2p^2\left(p^{-1}\sum_{j=1}^{g} a_{nj}^2\right)^2 \quad \text{(A.37)}$$

uniformly in $(u,v) \in \mathbb{R}^2$, where I have used $E|G_{t-g-1,j}^{(m,l)}(u,v)|^4 \le Ct^2$ because $G_{t-g-1,j}^{(m,l)}(u,v), 1 \le j \le g$, is the sum of a martingale difference sequence with bounded fourth moment.

I now consider the second term in (A.36). Given each $t, E[Q_{ntj}^{(m,l)}(u,v;u',v')|\mathcal{F}_{t-j-1}] = 0$ for all $1 \le j \le g$, and so $\{Q_{ntj}^{(m,l)}(u,v;u',v'), \mathcal{F}_{t-j}\}$ is a martingale difference sequence. Moreover, conditional on $\psi_{t-j}^{(l)}(v)$ and $\{G_{t-g-1,r}^{(m,l)}(\cdot,\cdot)\}_{r=1}^{g}, Q_{ntj}^{(m,l)}(u,v;u',v')$ is a sum of independent random variables. It follows that the second term of (A.36),

$$E\left|\sum_{j=1}^{g-1} a_{nj}Q_{ntj}^{(m,l)}(u,v;u',v')\right|^2 = \sum_{j=1}^{g-1} a_{nj}^2 E|Q_{ntj}^{(m,l)}(u,v;u',v')|^2$$

$$= \sum_{j=1}^{g-1} a_{nj}^2 \sum_{r=j+1}^{g} a_{nr}^2 E|\psi_{t-j}^{(l)}(v)\psi_{t-r}^{(l)}(v')^* G_{t-g-1,j}^{(m,l)}(u,v) \times G_{t-g-1,r}^{(m,l)}(u',v')^*|^2$$

$$\le Ct^2p^2\left(p^{-1}\sum_{j=1}^{g-1} a_{nj}^2\right)^2 \quad \text{(A.38)}$$

uniformly in $(u,v,u',v') \in \mathbb{R}^4$, where I have used $E|G_{t-g-1,j}^{(m,l)}(u,v)|^4 \le Ct^2$. Combining (A.34) and (A.36)–(A.38), we have $E|S_{1nt}^{(m,l)}|^4 \le Ct^2p^2(p^{-1}\sum_{j=1}^{g} a_{nj}^2)^2$ and so $\sum_{t=g+2}^{n} E|S_{1nt}^{(m,l)}|^4 = O(p^2/n) = o(p^2)$. It follows that (A.32) holds.

I now turn to verify (A.33), for which it suffices to show that

$$E\left|\sum_{t=g+2}^{n} E[(S_{1nt}^{(m,l)})^2|\mathcal{F}_{t-1}] - E(S_{1nt}^{(m,l)})^2\right|^2 + E\left[\sum_{t=g+2}^{n} E[|S_{1nt}^{(m,l)}|^2|\mathcal{F}_{t-1}] - E|S_{1nt}^{(m,l)}|^2\right]^2 = o(p^2).$$

I show the first condition only; the proof for the second one is similar. Given (A.30), (A.35) and $E\psi_1^{(r)}(u)\psi_1^{(r)}(u) = \sigma_0^{(r,r)}(u,u)$,

$$E[(S_{1nt}^{(m,l)})^2|\mathcal{F}_{t-1}] = \int \sigma_0^{(m,m)}(u,u')\sum_{j=1}^{g} a_{nj}^2\psi_{t-j}^{(l)}(v)\psi_{t-j}^{(l)}(v')G_{t-g-1,j}^{(m,l)}(u,v) \times G_{t-g-1,j}^{(m,l)}(u',v')$$

$$+ 2\int \sigma_0^{(m,m)}(u,u')\sum_{j=1}^{g-1} a_{nj}\mathrm{Re}[Q_{ntj}^{(m,l)}(u,v;u',v')]$$

$$\equiv Z_{nt}^{(m,l)} + C_{1nt}^{(m,l)}, \quad \text{say}. \quad \text{(A.39)}$$

Putting $\tilde{\psi}_t^{(r,r)}(v,v') \equiv \psi_t^{(r)}(v)\psi_t^{(r)}(v') - \sigma_0^{(r,r)}(v,v')$, one can decompose

$$Z_{nt}^{(m,l)} = \int \sigma_0^{(m,m)}(u,u')\sigma_0^{(l,l)}(v,v')\sum_{j=1}^{g} a_{nj}^2 G_{t-g-1,j}^{(m,l)}(u,v) \times G_{t-g-1,j}^{(m,l)}(u',v')\,dW\,dW'$$

$$+ \sum_{j=1}^{g} a_{nj}^2 \int \sigma_0^{(m,m)}(u,u')\tilde{\psi}_{t-j}^{(l,l)}(v,v')G_{t-g-1,j}^{(m,l)}(u,v) \times G_{t-g-1,j}^{(m,l)}(u',v')\,dW\,dW'$$

$$\equiv U_{nt}^{(m,l)} + C_{2nt}^{(m,l)}, \quad \text{say}. \quad \text{(A.40)}$$

Now, using the definition of $G_{t-g-1,j}^{(m,l)}(\cdot,\cdot)$ given in (A.30),

$$\begin{aligned} U_{nt}^{(m,l)} &= \sum_{j=1}^{g} a_{nj}^2 \int \sigma_0^{(m,m)}(u,u')\sigma_0^{(l,l)}(v,v') \\ &\quad \times \sum_{s=1}^{t-g-1} \psi_s^{(m)}(u)^* \psi_s^{(m)}(u')^* \psi_{s-j}^{(l)}(v)^* \psi_{s-j}^{(l)}(v')^* \\ &\quad + 2\int \sigma_0^{(m,m)}(u,u')\sigma_0^{(l,l)}(v,v') \sum_{j=1}^{g} a_{nj}^2 \\ &\quad \times \sum_{s=1}^{t-g-1}\sum_{r=1}^{s-1} \psi_s^{(m)}(u)^*\psi_r^{(m)}(u')^*\psi_{s-j}^{(l)}(v)^*\psi_{r-j}^{(l)}(v')^* \\ &\equiv V_{nt}^{(m,l)} + C_{3nt}^{(m,l)} \end{aligned} \tag{A.41}$$

and

$$\begin{aligned} V_{nt}^{(m,l)} &= (t-g-1)\sum_{j=1}^{g} a_{nj}^2 \int |\sigma_0^{(m,m)}(u,u')|^2 |\sigma_0^{(l,l)}(v,v')|^2 \, dW\, dW' \\ &\quad + \sum_{j=1}^{g} a_{nj}^2 \int |\sigma_0^{(m,m)}(u,u')|^2 \sigma_0^{(l,l)}(v,v') \\ &\quad \times \sum_{s=1}^{t-g-1} \tilde{\psi}_{s-j}^{(m,m)}(u,u')^* \, dW\, dW' \\ &\quad + \sum_{j=1}^{g} a_{nj}^2 \int \sigma_0^{(m,m)}(u,u')\sigma_0^{(l,l)}(v,v') \\ &\quad \times \sum_{s=1}^{t-g-1} \tilde{\psi}_s^{(m,m)}(u,u')^* \psi_{s-j}^{(l)}(v)^* \psi_{s-j}^{(l)}(v')^* \, dW\, dW' \\ &\equiv E(S_{1nt}^{(m,l)})^2 + C_{4nt}^{(m,l)} + C_{5nt}^{(m,l)}, \quad \text{say.} \end{aligned} \tag{A.42}$$

Combining (A.39)–(A.42) and using the C_r inequality and Lemma A.1,

$$E\left|\sum_{t=g+2}^{n} E[(S_{1nt}^{(m,l)})^2|\mathcal{F}_{t-1}] - E(S_{1nt}^{(m,l)})^2\right|^2 = E\left|\sum_{h=1}^{5}\sum_{t=g+2}^{n} C_{hnt}^{(m,l)}\right|^2 = o(p^2).$$

Thus (A.33) holds, and $[D_0^{(m,l)} p\int_0^\infty k^4(z)\,dz]^{-1/2}\hat{V}_{ng}^{(m,l)} \to^d$ N(0, 1) by Brown's theorem.

Lemma A.1

Let $C_{hnt}^{(m,l)}, 1 \le h \le 5$, be as in (A.39)–(A.42), and let $g \equiv g(n)$ be such that $g/p \to \infty, g/n \to 0$. Then $E|\sum_{t=g+2}^{n} C_{hnt}^{(m,l)}|^2 = o(p^2)$ for $1 \le h \le 5$.

Proof of Lemma A.1

I only show the proof for $C_{1nt}^{(m,l)}$. By Minkowski's inequality and (A.38),

$$\begin{aligned} &E|C_{1nt}^{(m,l)}|^2 \\ &\le \left\{2\int |\sigma_0^{(m,m)}(u,u')| \times \left[E\left|\sum_{j=1}^{g-1} a_{nj} Q_{ntj}^{(m,l)}(u,v;u',v')\right|^2\right]^{1/2} dW\, dW'\right\}^2 \\ &\le Ct^2\left[\sum_{j=1}^{g} a_{nj}^2\right]^2. \end{aligned}$$

Also, because $E[Q_{ntj}^{(m,l)}(u,v;u',v')|\mathcal{F}_{t-g-1}] = 0$ for $1 \le j \le g$, $E[C_{1nt}^{(m,l)}(C_{1ns}^{(m,l)})^*] = 0$ for $t-s > g$. It follows that

$$\begin{aligned} E\left|\sum_{t=g+2}^{n} C_{1nt}^{(m,l)}\right|^2 &\le \sum_{|t-s|\le g} [E|C_{1nt}^{(m,l)}|^2 E|C_{1ns}^{(m,l)}|^2]^{1/2} \\ &\le C(n-g)^3 g\left(\sum_{j=1}^{g} a_{nj}^2\right)^2 = O(p^2 g/n). \end{aligned}$$

Moreover, similar results can be obtained for $2 \le h \le 5$. This completes the proof.

To show the proofs of Theorem 4 and Theorem 7, I first state a lemma.

Lemma A.2

Suppose that Assumptions A.3 and A.8 hold. (a) If $\hat{p}/p = 1 + O_P(p^{-(3/2\beta-1)})$ for some $\beta > (2b-\frac{1}{2})/(2b-1)$, where b is as in Assumption A.3 and $p = cn^\lambda$ for $c \in (0,\infty)$ and $\lambda \in (0, (2b-1)/(2b-\frac{1}{2}))$, then $p^{-1}\sum_{j=1}^{n-1}\{k^2(j/\hat{p}) - k^2(j/p)\} = o_P(p^{-1/2})$; (b) If $\hat{p}/p = 1 + O_P(p^{-\beta})$ for some $\beta > 0$, then $p^{-1}\sum_{j=1}^{n-2}\{k^r(j/\hat{p}) - k^r(j/p)\} \to^p 0$, where $r = 2, 4$.

Proof of Lemma A.2

I show (a) only; the proof of (b) is similar. First, I write

$$\begin{aligned} &p^{-1}\sum_{j=1}^{n-1}\{k^2(j/\hat{p}) - k^2(j/p)\} \\ &= p^{-1}\left(\sum_{j=1}^{d} + \sum_{j=d+1}^{n-1}\right)\{k^2(j/\hat{p}) - k^2(j/p)\} \\ &= \hat{A}_{1n} + \hat{A}_{2n}, \end{aligned} \tag{A.43}$$

where $d \equiv [p^{2b-1/2}\ln n]^{1/(2b-1)}$. Note that $d/p \to \infty$ and $d/n \to 0$. Further,

$$\begin{aligned} \hat{A}_{1n} &= p^{-1}\sum_{j=1}^{d}\{k(j/\hat{p}) - k(j/p)\}^2 \\ &\quad + 2p^{-1}\sum_{j=1}^{d}\{k(j/\hat{p}) - k(j/p)\}k(j/p) \\ &= \hat{A}_{11n} + 2\hat{A}_{12n}. \end{aligned} \tag{A.44}$$

Given the Lipschitz condition of $k(\cdot)$ in Assumption A.8, $\hat{p}/p - 1 = O_P(p^{-(3/2\beta-1)}), \beta > (2b-1/2)/(2b-1)$, and the Cauchy–Schwarz inequality,

$$|\hat{A}_{11n}| \le C^2(\hat{p}/p - 1)^2(p/\hat{p})(d/p)^3 = o_P(p^{-1}) \tag{A.45}$$

and

$$|\hat{A}_{12n}| \leq |\hat{A}_{11n}|^{1/2} \left\{ p^{-1} \sum_{j=1}^{d} k^2(j/p) \right\}^{1/2} = o_P(p^{-1/2}). \quad \text{(A.46)}$$

Next, consider the second term of (A.43). Noting $|k(z)| \leq C|z|^{-b}$ for large z,

$$|\hat{A}_{2n}| \leq C\{(\hat{p}/p)^{2b} + 1\}(p/d)^{2b-1} \left\{ d^{-1} \sum_{j=d+1}^{n-1} (j/d)^{-2b} \right\}$$
$$= O_P((p/d)^{2b-1}) = o_P(p^{-1/2}). \quad \text{(A.47)}$$

The desired result follows by combining (A.43)–(A.47).

Proof of Theorem 4

Put $\hat{B}_n \equiv \sum_{j=1}^{n-1} \{k^2(j/\hat{p}) - k^2(j/p)\}(n-j) \int |\hat{\sigma}_j^{(m,l)}(u,v)|^2 \, dW$. Then from (17),

$$M(m,l,\hat{p}) - M(m,l,p)$$
$$= \left\{ \hat{B}_n - \hat{C}_0^{(m,l)} \sum_{j=1}^{n-1} [k^2(j/\hat{p}) - k^2(j/p)] \right\} \div \left\{ \hat{D}_0^{(m,l)} \sum_{j=1}^{n-2} k^4(j/\hat{p}) \right\}^{1/2}$$
$$+ M(m,l,p) \left\{ \left(\sum_{j=1}^{n-1} k^4(j/\hat{p}) \Big/ \sum_{j=1}^{n-2} k^4(j/p) \right)^{1/2} - 1 \right\}$$
$$= \hat{B}_n \Big/ \left\{ \hat{D}_0^{(m,l)} \sum_{j=1}^{n-2} k^4(j/\hat{p}) \right\}^{1/2} + o_P(1),$$

where the second equality follows by Lemma A.2, $M(m,l,p) = O_P(1)$ from Theorem 3, $\sum_{j=1}^{n-2} k^4(j/p) = p \int_0^\infty k^4(z)\, dz\{1 + o(1)\}, \hat{C}_0^{(m,l)} \to^p C_0^{(m,l)}$, and $\hat{D}_0^{(m,l)} \to^p D_0^{(m,l)} > 0$.

It remains to show $p^{-1/2}\hat{B}_n = o_P(1)$. The proof is analogous to Lemma A.2. Write

$$\hat{B}_n = \left(\sum_{j=1}^{d} + \sum_{j=d+1}^{n-1} \right) [k^2(j/\hat{p}) - k^2(j/p)](n-j) \times \int |\hat{\sigma}_j^{(m,l)}(u,v)|^2 \, dW$$
$$= \hat{B}_{1n} + \hat{B}_{2n}, \quad \text{(A.48)}$$

where $d \equiv [p^{2b-1/2} \ln(n)]^{1/(2b-1)}$. For the second term, given $|k(z)| \leq C|z|^{-b}$ for large z,

$$|\hat{B}_{2n}| \leq C[(\hat{p}/p)^{2b} + 1]p^{2b} d^{1-2b} \times \left\{ d^{-1} \sum_{j=d+1}^{n-1} |j/d|^{-2b}(n-j) \int |\hat{\sigma}_j^{(m,l)}(u,v)|^2 \, dW \right\}$$
$$= O_P(p^{2b} d^{1-2b}) = o_P(p^{-1/2}), \quad \text{(A.49)}$$

where I have made use of $(n-j)E \int |\hat{\sigma}_j^{(m,l)}(u,v)|^2 \, dW(u,v) \leq C$ under independence.

For the first term of (A.48), write $\hat{B}_{1n} = \hat{B}_{11n} + 2\hat{B}_{12n}$, where

$$\hat{B}_{11n} \equiv \sum_{j=1}^{d} [k(j/\hat{p}) - k(j/p)]^2 (n-j) \int |\hat{\sigma}_j^{(m,l)}(u,v)|^2 \, dW$$

and

$$\hat{B}_{12n} \equiv \sum_{j=1}^{d} \{k(j/\hat{p}) - k(j/p)\} k(j/p)(n-j) \int |\hat{\sigma}_j^{(m,l)}(u,v)|^2 \, dW.$$

Given the Lipschitz condition of $k(\cdot)$, $(n-j)\int E|\hat{\sigma}_j^{(m,l)}(u,v)|^2 dW \leq C, \hat{p}/p = 1 + O_P(p^{-(3/2\beta-1)})$ with $\beta > (2b - \frac{1}{2})/(2b-1)$, it can be shown that

$$|\hat{B}_{11n}| \leq C^2(\hat{p}/p - 1)^2(p/\hat{p})^2(d^3/p^2) = o_P(1) \quad \text{(A.50)}$$

and

$$|\hat{B}_{12n}| \leq |\hat{B}_{11n}|^{1/2} \left\{ \sum_{j=1}^{d} k^2(j/p)(n-j) \int |\hat{\sigma}_j^{(m,l)}(u,v)|^2 \, dW \right\}^{1/2}$$
$$= o_P(p^{1/2}). \quad \text{(A.51)}$$

Combining (A.48)–(A.51) yields $p^{-1/2}\hat{B}_n \to^p 0$. This completes the proof.

Proof of Theorem 5

The consistency result follows from (a) $p^{-1}\sum_{j=1}^{n} k^4(j/p) \to \int_0^\infty k^4(j/p)$, (b) $\text{IMSE}(\hat{f}_n^{(0,m,l)}, f^{(0,m,l)}) \to 0$, (c) $\hat{C}_0^{(m,l)} = O_P(1)$, and (d) $\hat{D}_0^{(m,l)} \to^p D_0^{(m,l)}$. Part (a) follows from Assumption A.3 and $p \to \infty, p/n \to 0$. The proof of (b) is the same as that of $\text{IMSE}(\bar{f}_n, f) \to 0$ in the proof of Theorem 2 and thus is omitted here. The proof of (c)–(d) is straightforward by Markov's inequality and Assumptions A.1(2) and A.9.

Proof of Theorem 6

By Theorem 3, $M(m,l,p) \to^d \text{N}(0, 1)$ under serial independence. Hence the asymptotic p-value of $M(m,l,p)$ is $1 - \Phi[M(m,l,p)]$, where $\Phi(\cdot)$ is the CDF N(0, 1). Put $B_n(k) = -2\ln\{1 - \Phi[M(m,l,p)]\}$. Using Theorem 5 and $\ln[1 - \Phi(z)] = -\frac{1}{2}z^2\{1 + o(1)\}$ as $z \to +\infty$,

$$(p/n^2)B_n(k) = \left(\frac{\pi}{2}\right)^2 L_2^2(f^{(0,m,l)}, f_0^{(0,m,l)}) \div \left[D_0^{(m,l)} \int_0^\infty k^4(z)\, dz \right] + o_P(1). \quad \text{(A.52)}$$

Following Bahadur (1960), I call $(\pi/2)^2 L_2^2(f^{(0,m,l)}, f_0^{(0,m,l)}) / [D_0^{(m,l)} \int_0^\infty k^4(z)\, dz]$ the "asymptotic slope" of the test $M(m,l,p)$. Now consider two tests based on $M(m,l,p)$ using two different kernels $k_1(\cdot)$ and $k_2(\cdot)$ in $\mathbb{K}(\tau)$ under the fixed alternative. Bahadur's asymptotic relative efficiency $\text{REF}(k_2 : k_1)$ of $k_2(\cdot)$ to $k_1(\cdot)$ is the limit ratio of the sample sizes for both tests to attain the same asymptotic significance level under the same fixed alternative. Given (A.52) and $p = cn^\lambda$ for $\lambda \in (0,1)$,

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

Thus $k_2(\cdot)$ is more efficient than $k_1(\cdot)$ if $\int_0^\infty k_2^4(z)\,dz < \int_0^\infty k_1^4(z)\,dz$ in terms of Bahadur's criterion. Hong (1996), in the proof of his theorem 5, showed that the Daniell kernel $k_D(z) = \sin(\sqrt{3}\tau z)/\sqrt{3}\tau z$ minimizes $\int_0^\infty k^4(z)\,dz$ over the class of functions

$$\Theta(\tau) = \Bigg\{ k(\cdot) | k : \mathbb{R} \to [-1,1] \text{ is symmetric and continuous at 0 and all except a finite number of points, with } k(0) = 1, \int_0^\infty k^2(z)\,dz < \infty, k^{(2)} = \frac{\tau^2}{2} > 0, K(\xi) \geq 0 \, \forall \xi \in \mathbb{R} \Bigg\}.$$

Because $\mathbb{K}(\tau)$ is a subset of $\Theta(\tau)$ and $k_D(z) \in \mathbb{K}(\tau)$, it follows that $k_D(z)$ maximizes the asymptotic slope of $M(m,l,p)$ over $\mathbb{K}(\tau)$.

Proof of Theorem 7

Theorem 7 follows from Lemma A.2 (b), Theorem 5, $\hat{C}_0^{(m,l)} = O_P(1), \hat{D}_0^{(m,l)} \to^p D_0^{(m,l)} > 0, p = cn^\lambda$ for $\lambda \in (0,1)$, and

$$\hat{H}_n \equiv n^{-1} \sum_{j=1}^{n-1} [k^2(j/\hat{p}) - k^2(j/p)](n-j) \times \int |\hat{\sigma}_j^{(m,l)}(u,v)|^2 \, dW \to^p 0. \quad \text{(A.53)}$$

I now show (A.53). First, observe that

$$\begin{aligned} E|\hat{\sigma}_j^{(m,l)}(u,v)|^2 &\leq 2E|\hat{\sigma}_j^{(m,l)}(u,v) - \sigma_j^{(m,l)}(u,v)|^2 + 2|\sigma_j^{(m,l)}(u,v)|^2 \\ &\leq C(n-j)^{-1} + C\alpha(j)^{2(\nu-1)/\nu}, \end{aligned} \quad \text{(A.54)}$$

where the first term is bounded by $C(n-j)^{-1}$ given (A.13) and $E|\bar{\sigma}_j^{(m,l)}(u,v) - \sigma_j^{(m,l)}(u,v)|^2 \leq C(n-j)^{-1}$ from (A.7)–(A.10); the second term follows from $|\sigma_j^{(m,l)}(u,v)| \leq C\alpha(j)^{(\nu-1)/\nu}$ given Assumptions A.1(2) and A.9. To show (A.53), first write

$$\hat{H}_n = n^{-1} \left(\sum_{j=1}^{p} + \sum_{j=p+1}^{n-1} \right) [k^2(j/\hat{p}) - k^2(j/p)](n-j) \times \int |\hat{\sigma}_j^{(m,l)}(u,v)|^2 \, dW \equiv \hat{H}_{1n} + \hat{H}_{2n}. \quad \text{(A.55)}$$

Given the Lipschitz condition of $k(\cdot)$,

$$\begin{aligned} \hat{H}_{1n} &\leq 2C|\hat{p}^{-1} - p^{-1}| n^{-1} \sum_{j=1}^{p} j(n-j) \int |\hat{\sigma}_j^{(m,l)}(u,v)|^2 \, dW \\ &= 2C|\hat{p}/p - 1|(p/\hat{p})p^{-1} O_P \left(p^2/n + p \sum_{j=1}^{p} \alpha(j)^{2(\nu-1)/\nu} \right) \\ &= o_P(1), \end{aligned} \quad \text{(A.56)}$$

where the first equality follows from (A.54) and Chebysev's inequality, and the last equality follows from $\hat{p}/p - 1 = O_P(p^{-\beta})$ for $\beta > 0, p = cn^\lambda$ for $\lambda \in (0,1)$, and Assumption A.1(2).

For the second term of (A.55), given $|k(z)| \leq C|z|^{-b}$ for large z,

$$\begin{aligned} |\hat{H}_{2n}| &\leq C^2(\hat{p}^{2b} + p^{2b}) n^{-1} \sum_{j=p+1}^{n-1} j^{-2b}(n-j) \times \int |\hat{\sigma}_j(u,v)|^2 \, dW(u,v) \\ &= C^2[(\hat{p}/p)^{2b} + 1] \times O_P\left(n^{-1} \sum_{j=p+1}^{n-1} (j/p)^{-2b} + \sum_{j=p+1}^{n-1} \alpha(j)^{2(\nu-1)/\nu} \right) \to^p 0, \end{aligned} \quad \text{(A.57)}$$

where the first equality follows from (A.54) and the second follows from $p = cn^\lambda$ for $\lambda \in (0,1)$ and Assumption A.1(2). Combining (A.55)–(A.57) yields (A.53).

[Received July 1997. Revised April 1999.]

REFERENCES

Anderson, T. W. (1993), "Goodness-of-Fit Tests for Spectral Distributions," *The Annals of Statistics*, 21, 830–847.

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

Andrews, D. W. K., and Ploberger, W. (1996), "Testing for Serial Correlation Against an ARMA(1,1) Process," *Journal of the American Statistical Association*, 91, 1331–1342.

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

Baringhaus, L., and Henze, N. (1988), "A Consistent Test for Multivariate Normality Based on the Empirical Characteristic Function," *Metrika*, 35, 339–348.

Bierens, H. (1982), "Consistent Model Specification Tests," *Journal of Econometrics*, 20, 105–134.

Bierens, H., and Ploberger, W. (1997), "Asymptotic Theory of Integrated Conditional Moments Tests," *Econometrica*, 58, 1129–1151.

Bollerslev, T. (1990), "Modelling the Coherence in Short-Run Nominal Exchange Rates: A Multivariate Generalized ARCH Model," *Review of Economic Statistics*, 72, 498–505.

Box, G., and Pierce, D. (1970), "Distribution of Residual Autocorrelations in Autoregressive-Integrated Moving Average Time Series Models," *Journal of the American Statistical Association*, 65, 1509–1526.

Brillinger, D. R. (1965), "An Introduction to Polyspectra," *Annals of Mathematical Statistics*, 36, 1351–1374.

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

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

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

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

Chung, K. (1974), *A Course in Probability Theory* (2nd ed.), New York: Academic Press.

Csörgõ, S. (1984), "Testing by the Empirical Characteristic Function: A Survey," in *Asymptotic Statistics 2: Proceedings of the 3rd Prague Symposium on Asymptotic Statistics*, eds. P. Mandl and M. Muskova, Amsterdam: Elsevier, pp. 45–56.

———, (1985), "Testing for Independence by the Empirical Characteristic Function," *Journal of Multivariate Analysis*, 16, 290–299.

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

Durlauf, S. (1991), "Spectral-Based Tests for the Martingale Hypothesis," *Journal of Econometrics*, 50, 1–19.

Engle, R. (1982), "Autoregressive Conditional Heteroskedasticity With Estimates of the Variance of United Kingdom Inflation," *Econometrica*, 50, 987–1007.

Engle, R., Ito, T., and Lin, W. (1990), "Meteor Showers or Heat Waves? Heteroskedastic Intra-Daily Volatility in the Foreign Exchange Market," *Econometrica*, 58, 525–542.

Epps, T. W. (1987), "Testing That a Stationary Time Series is Gaussian," *The Annals of Statistics*, 15, 1683–1698.

———, (1988), "Testing That a Gaussian Process is Stationary," *The Annals of Statistics*, 16, 1667–1683.

———, (1993), "Characteristic Functions and Their Empirical Counterparts: Geometrical Interpretations and Applications to Statistical Inference," *The American Statistician*, 47, 33–48.

Epps, T. W., and Pulley, L. B. (1983), "A Test for Normality Based on the Empirical Characteristic Function," *Biometrika*, 70, 723–726.

Epps, T. W., and Singleton, K. J. (1986), "An Omnibus Test for the Two-Sample Problem Using the Empirical Characteristic Function," *Journal of Statistical Computation and Simulation*, 26, 177–203.

Fama, E. F., and Roll, R. (1968), "Some Properties of Symmetric Stable Distributions," *Journal of the American Statistical Association*, 63, 817–836.

Feigin, P. D., and Heathcote, C. R. (1976), "The Empirical Characteristic Function and the Cramer–von Mises Statistic," *Sankya*, Ser. A, 38, 309–325.

Feuerverger, A. (1987), "On Some ECF Procedures for Testing Independence," in *Time Series and Econometric Modelling*, eds. I. B. MacNeill and G. J. Umphrey, Boston: Reidel, pp. 189–206.

———, (1993), "A Consistent Test for Bivariate Dependence," *International Statistical Review*, 61, 419–433.

Feuerverger, A., and Mureika, R. (1977), "The Empirical Characteristic Function and Its Applications," *The Annals of Statistics*, 5, 88–97.

Geweke, J. (1981), "A Comparison of Tests of the Independence of Two Covariance Stationary Time Series," *Journal of the American Statistical Association*, 76, 363–373.

Granger, C., and Anderson, A. P. (1978), *An Introduction to Bilinear Time Series Models*, Gottingen: Vandenhoek and Ruprecht.

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

Hall, P., and Welsh, A. H. (1983), "A Test for Normality Based on the Empirical Characteristic Function," *Biometrika*, 70, 485–489.

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

Heathcote, C. R. (1972), "A Test of Goodness of Fit for Symmetric Random Variables," *Australian Journal of Statistics*, 14, 172–181.

Henze, N., and Wagner, T. (1997), "A New Approach to the BHEP Tests for Multivariate Normality," *Journal of Multivariate Analysis*, 62, 1–23.

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

Hinich, M., and Patterson, D. (1992), "A New Diagnostic Test of Model Inadequacy Which Uses the Martingale Difference Criterion," *Journal of Time Series Analysis*, 13, 233–252.

Hjellvik, V., and Tjøstheim, D. (1996), "Nonparametric Statistics for Testing of Linearity and Serial Independence," *Journal of Nonparametric Statistics*, 6, 223–251.

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

———, (1998), "Testing for Pairwise Independence via the Empirical Distribution Function," *Journal of the Royal Statistical Society*, Ser. B, 60, 429–453.

Koutrouvelis, I. A. (1980), "A Goodness-of-Fit Test of Simple Hypotheses Based on the Empirical Characteristic Function," *Biometrika*, 67, 238–240.

Koutrouvelis, I. A., and Kellermeier, J. (1981), "A Goodness-of-Fit Test Based on the Empirical Characteristic Function When Parameters Must be Estimated," *Journal of the Royal Statistical Society*, Ser. B, 43, 173–176.

Lukacs, E. (1970), *Characteristic Functions* (2nd ed.), London: Charles Griffin.

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

Meese, R., and Rogoff, K. (1983), "Empirical Exchange Rate Models of the Seventies: Do They Fit Out of Sample?," *Journal of International Economics*, 14, 2–24.

Newey, W. and Steigerward, D. (1996), "Asymptotic Bias for Quasi-Maximum Likelihood Estimators in Conditional Heteroskedastic Models," *Econometrica*, 65, 587–599.

Newey, W., and West, K. (1994), "Automatic Lag Selection in Covariance Matrix Estimation," *Review of Economic Studies*, 61, 631–653.

Parzen, E. (1957), "On Consistent Estimates of the Spectrum of a Stationary Time Series," *Annals of Mathematical Statistics*, 28, 329–348.

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.

———, (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.

Silverman, B. W. (1986), *Density Estimation for Statistics and Data Analysis*, London: Chapman Hall.

Skaug, H. J., and Tjøstheim, D. (1993a), "A Nonparametric Test of Serial Independence Based on the Empirical Distribution Function," *Biometrka*, 80, 591–602.

———, (1993b), "Nonparametric Tests of Serial Independence," in *Developments in Time Series Analysis*, The Priestley Birthday Volume, ed. T. Subba Rao, London: Chapman and Hall, pp. 207–229.

———, (1996), "Measures of Distance Between Densities With Application to Testing for Serial Independence," in *Time Series Analysis in Memory of E. J. Hannan*, eds. P. Robinson and M. Rosenblatt, New York: Springer, pp. 363–377.

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

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

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

Zurbenko, I. (1986), *The Spectral Analysis of Time Series*, Amsterdam: North-Holland.