Testing for Serial Correlation of Unknown Form Using Wavelet Methods
Author(s): Jin Lee and Yongmiao Hong
Source: *Econometric Theory*, Vol. 17, No. 2 (Apr., 2001), pp. 386-423
Published by: Cambridge University Press
Stable URL: http://www.jstor.org/stable/3533074
Accessed: 22/11/2013 14:12

*Econometric Theory*, **17**, 2001, 386–423. Printed in the United States of America.

# TESTING FOR SERIAL CORRELATION OF UNKNOWN FORM USING WAVELET METHODS

JIN LEE
*National University of Singapore*

YONGMIAO HONG
*Cornell University*

A wavelet-based consistent test for serial correlation of unknown form is proposed. As a spatially adaptive estimation method, wavelets can effectively detect local features such as peaks and spikes in a spectral density, which can arise as a result of strong autocorrelation or seasonal or business cycle periodicities in economic and financial time series. The proposed test statistic is constructed by comparing a wavelet-based spectral density estimator and the null spectral density. It is asymptotically one-sided $N(0,1)$ under the null hypothesis of no serial correlation and is consistent against serial correlation of unknown form. The test is expected to have better power than a kernel-based test (e.g., Hong, 1996, *Econometrica* 64, 837–864) when the true spectral density has significant spatial inhomogeneity. This is confirmed in a simulation study. Because the spectral densities of time series arising in practice usually have unknown smoothness, the wavelet-based test is a useful complement to the kernel-based test in practice.

## 1. INTRODUCTION

Wavelet analysis originated as a new analytic method alternative to Fourier analysis in signal analysis and has rapidly grown through interactions with mathematics over the last decade or so. As a spatially adaptive analytic method, wavelets provide a new useful tool for nonparametric function estimation. Because wavelets are spatially varying orthonormal bases with two parameters—scale and translation—they are fundamentally different from the Fourier basis or Gabor basis (i.e., the windowed Fourier basis) and have some appealing statistical advantages over traditional estimators such as kernel and spline methods in esti-

We are most grateful to a referee and the co-editor K. Tanaka for invaluable comments and encouragements that have led to significant improvements over the earlier versions. We also thank H. Gao, G. Jakubson, N. Kiefer, M. Neumann, J. Park, P. Phillips, A. Spanos, H. Wan, M. Wells, and seminar participants at the 1998 North American Summer Meeting of the Econometric Society in Montreal and the 1999 Far Eastern Econometric Society Meeting in Singapore for helpful comments. All the remaining errors are attributed solely to the authors. Address correspondence to: Jin Lee, Department of Economics, National University of Singapore, Singapore; e-mail: ecsleej@nus.edu.sg. Yongmiao Hong, Department of Economics and Statistical Science, Cornell University, Uris Hall, Ithaca, NY 14853-7601, USA; e-mail: yh20@cornell.edu.

  0266-4666/01 $9.50

mating a function with unknown smoothness (e.g., Donoho and Johnstone, 1994, 1995a, 1995b; Donoho et al., 1996). In particular, wavelets have good time-frequency localization properties. Both smooth and nonsmooth functions can be effectively reconstructed using wavelet bases.

Although the development of wavelet analysis has been rapidly growing, its application to time series analysis is relatively sparse. In time series spectral analysis, Gao (1993) uses wavelet shrinkage to estimate the spectral density of a stationary Gaussian process. Neumann (1996) derives the asymptotic normal distribution of empirical wavelet coefficients used to estimate the spectral density of a non-Gaussian process. He shows in simulations that wavelet estimators outperform kernel estimators in capturing such spatially inhomogeneous features as peaks and spikes in the spectral density. In another development, Priestley (1996) provides useful links between wavelet analysis and nonstationary evolutionary spectral analysis. See also Subba Rao and Indukumar (1996) for wavelet application to nonlinear and nonstationary time series.

There have been some applications of wavelet analysis to economics and econometrics. Goffe (1994) illustrates the application of the wavelet method to some nonstationary macroeconomic time series. Gilbert (1995) uses wavelets to estimate and test structural changes. Jensen (2000) proposes a wavelet-based algorithm to estimate a long memory model via the maximum likelihood estimation method. Wang (1995) applies the wavelet method to detect jumps and sharp cusps in stock market returns. Ramsey and his coauthors, in a series of papers (e.g., Ramsey, 1999; Ramsey and Lampart, 1998a, 1998b; Ramsey and Zhang, 1996, 1997; Ramsey, Usikov, and Zaslavsky, 1995), apply wavelets to various economic and financial time series and obtain some interesting results.

In this paper, we illustrate how wavelets can be used to effectively detect serial correlation of unknown form. Detection and inference of serial correlation have long been of interest in time series analysis (e.g., Anderson, 1993; Andrews and Ploberger, 1996; Box and Pierce, 1970; Durbin and Watson, 1950, 1951; Durlauf, 1991; Godfrey, 1978a, 1978b; Hong, 1996; Robinson, 1991; Whang, 1997). Among various existing tests for serial correlation, Hong (1996) proposes a consistent test for serial correlation using a Parzen (1957) kernel estimator for the spectral density of a stationary time series. The test is shown to have good power against both short and long memory processes. However, spatially nonadaptive estimation methods such as the kernel method cannot effectively detect spatially varying local features (e.g., kinks, peaks, or jumps). It is well known, for example, that kernel estimators tend to underestimate a mode in a spectral density (e.g., Priestley, 1981). Thus, it is expected that a kernel-based test may have relatively poor power when the spectral density has significant spatial inhomogeneity. Nonsmooth spectral densities with kinks, peaks, or spikes are not uncommon for time series arising in practice. They can arise as a result of strong autocorrelation or seasonal or business cycle periodicities in time series (e.g., Wen, 1998). It is therefore important to develop test procedures with good power against these alternatives. The wavelet method is particularly suitable for such a purpose. Here we propose a new consistent test for serial correlation using wavelets. The test is constructed by comparing a wavelet-based spectral density estimator and the null spectral density. We establish the asymptotic theory for the proposed test. The null distribution of the test statistic is asymptotically one-sided $N(0,1)$. No formulation of an alternative model is required, and the test is consistent against serial correlation of unknown form.

A simulation study compares the proposed wavelet-based test with the kernel-based test of Hong (1996). It confirms our conjecture that the wavelet method outperforms the kernel method in detecting spatially inhomogeneous spectral features. When the spectral density has distinctive peaks or spikes, the wavelet-based test has better power than the kernel-based test. On the other hand, when the spectral density is smooth, the kernel-based test performs better. Because the power of the tests depends on the shape of the unknown spectrum, the proposed test is a useful complement to the kernel-based test in practice.

In Section 2, we introduce the wavelet framework and construct the test statistic. The null asymptotic normality is derived in Section 3, and consistency is established in Section 4. Section 5 presents a simulation study on the finite sample performances of the proposed test and the kernel-based test of Hong (1996). All the proofs are given in the Appendixes. Throughout, $A^*$ denotes the complex conjugate of $A$; $\mathrm{Re}(A)$ the real part of $A$; $\mathbb{Z} = \{0,\pm 1,\dots\}$ and $\mathbb{Z}^+ = \{1,2,\dots\}$ the sets of integers, and positive integers, respectively; $C \in (0,\infty)$ a generic constant that may differ from place to place. All convergences are taken as the sample size $n \to \infty$.

## 2. BASIC FRAMEWORK AND TEST STATISTICS

### 2.1. Wavelet Analysis

Let $\psi: \mathbb{R} \to \mathbb{R}$ be an orthonormal wavelet such that the doubly infinite sequence $\{\psi_{jk}(x) = 2^{j/2}\psi(2^j x - k)\}$ forms a complete orthonormal basis of the $L_2(\mathbb{R})$-space of square integrable functions, where $j$ and $k \in \mathbb{Z}$ are integers corresponding to scale (dilation or compression) and translation (displacement), respectively. Any function $f(x) \in L_2(\mathbb{R})$ can be expressed as a sum of wavelets $\{\psi_{jk}(\cdot)\}$, which are generated from the single function $\psi(\cdot)$, the so-called mother wavelet. For excellent intuitive accounts of wavelet analysis, see, for example, Priestley (1996) and Ramsey (1998).

Throughout, we consider *multiresolution analysis*, which is the most commonly used analytic method in the wavelet literature (cf. Daubechies, 1992; Hernández and Weiss, 1996; Priestley, 1996; Strang and Nguyen, 1996).

DEFINITION. *A multiresolution analysis is a sequence of subspaces* $\{V_j, j \in \mathbb{Z}\}$ *of* $L_2(\mathbb{R})$ *satisfying the following requirements*:

(i) $V_j \subset V_{j+1}$ *and* $\cap V_j = \{\mathbf{0}\}$, $\overline{\cap V_j} = L_2$;
(ii) $f(x) \in V_j$ *if and only if* $f(2x) \in V_{j+1}$;

(iii) $f(x) \in V_j$ *if and only if* $f(x-k) \in V_j$, *for all* $k \in \mathbb{Z}$;
(iv) $V_0$ *has an orthonormal basis* $\{\phi(\cdot - k), k \in \mathbb{Z}\}$, *where* $\phi: \mathbb{R} \to \mathbb{R}$ *such that* $\int_{-\infty}^{\infty} \phi(x)dx = 1$.

From the nested structure of subspace $V_j$, the orthogonal complement $W_j$ of $V_j$ can be defined as $V_j \bigoplus W_j = V_{j+1}$, where $\bigoplus$ denotes the orthogonal sum. Let $V_0$ be the initial subspace; then

$$V_0 \bigoplus \sum_{j=0}^{\infty} W_j = L_2(\mathbb{R}).$$

Through dilation and translation of the scale function $\phi(\cdot)$ (also called father wavelet) and mother wavelet $\psi(\cdot)$, the sequences $\{\phi_{jk}(x) = 2^{j/2}\phi(2^j x - k)\}$ and $\{\psi_{jk}(x) = 2^{j/2}\psi(2^j x - k)\}$ constitute a complete orthonormal basis of $V_j$ and $W_j$, respectively. Each subspace $V_j$ encodes the information of the signal at resolution level $j$, which can be represented by scale functions $\{\phi_{jk}(\cdot),\ k \in \mathbb{Z}\}$. Each subspace $W_j$ orthogonal to $V_j$ encodes the details, namely, the difference of the information between the signals seen at two resolutions $V_j$ and $V_{j+1}$. Details at level $j$ can be represented by wavelets $\{\psi_{jk}(\cdot),\ k \in \mathbb{Z}\}$. Thus, signals at level $j$ combined with details at level $j$ provide signals at level $j+1$. Intuitively, a small $j$ or a low resolution level can capture smooth components of the signal, whereas a large $j$ or a high resolution level can capture variable components of the signal. Moreover, given a resolution level $j$, various values of translation parameter $k$ allow one to capture local features of the signal. In our application, the scale function $\phi(\cdot)$ will capture the smoothest component of the spectral density, whereas the wavelets $\{\psi_{jk}(\cdot)\}$ will capture the differences such as peaks and spikes. We note that requirement (i) implies that (a) the signal seen at a given resolution level contains all the information of the signal seen at coarser resolution levels and (b) any function in $L_2(\mathbb{R})$ can be approximated arbitrarily well by a sufficiently fine resolution (i.e., a sufficiently large $j$). Requirements (ii) and (iii) represent scale and shift invariance, respectively.

We first impose a standard condition on the mother wavelet $\psi(\cdot)$.

Assumption 1. $\psi: \mathbb{R} \to \mathbb{R}$ is an orthonormal wavelet such that $\int_{-\infty}^{\infty} \psi(x)dx = 0$, $\int_{-\infty}^{\infty} |\psi(x)|dx < \infty$, $\int_{-\infty}^{\infty} \psi(x)\psi(x-k)dx = 0$ for all $k \in \mathbb{Z}, k \neq 0$, and $\int_{-\infty}^{\infty} \psi^2(x)dx = 1$.

The orthonormality of $\psi(\cdot)$ implies that the doubly infinite sequence $\{\psi_{jk}(\cdot)\}$ constitutes an orthonormal basis for $L_2(\mathbb{R})$, that is,

$$\int_{-\infty}^{\infty} \psi_{jk}(x)\psi_{lm}(x)dx = \delta_{jl}\delta_{km}, \quad j,l,k,m \in \mathbb{Z},$$

where $\delta_{jl} = 1$ if $j = l$ and $\delta_{jl} = 0$ otherwise (cf. Daubechies, 1992). Assumption 1 ensures that the Fourier transform of $\psi(\cdot)$ defined by

$$\hat{\psi}(\omega) = (2\pi)^{-1/2} \int_{-\infty}^{\infty} \psi(x)e^{-i\omega x}dx, \quad i = \sqrt{-1}$$

exists and is continuous in $\omega$ almost everywhere. Note that $\hat{\psi}(0) = (2\pi)^{-1/2} \times \int_{-\infty}^{\infty} \psi(x)dx = 0$, which implies that the mother wavelet $\psi(\cdot)$ must have alternating signs. This is one of the characteristic properties of wavelets and a reason why wavelets are sensitive to changes or singularities. Note also that $\hat{\psi}^*(\omega) = \hat{\psi}(-\omega)$ for all $\omega \in \mathbb{R}$.

Most orthonormal wavelets $\psi(\cdot)$ are constructed from a father wavelet $\phi(\cdot)$. The mother wavelet $\psi(\cdot)$ can have bounded or unbounded support. A well-known compactly supported wavelet is the Haar wavelet,

$$\psi(x) = \begin{cases} 1, & 0 \le x < 1/2 \\ -1, & 1/2 \le x < 1 \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

It is generated from a linear combination of $\phi(x) = 1(0 \le x < 1)$, where $1(\cdot)$ is the indicator function. The sequence $\{\psi_{jk}(x) = 2^{j/2}\psi(2^j x - k), j, k \in \mathbb{Z}\}$ forms a complete orthonormal basis of $L_2(\mathbb{R})$. (See, e.g., Hernández and Weiss, 1996, pp. 59–61.)

An example of wavelets with unbounded support is the Shannon wavelet

$$\psi(x) = -2\,\frac{\sin(2\pi x) + \cos(\pi x)}{\pi(2x+1)}, \qquad x \in \mathbb{R}. \tag{2}$$

This is generated from the scaling function $\phi(x) = \sin(\pi x)/(\pi x)$. See Hernández and Weiss (1996, pp. 61–62) for more discussion.

We impose an additional condition for good localization of $\psi(\cdot)$ in the frequency domain.

Assumption 2.

(i) $|\hat{\psi}(\omega)| \le C(1 + |\omega|)^{-\alpha}$ for some $\alpha > \frac{3}{2}$ and some constant $C \in (0,\infty)$;
(ii) $\hat{\psi}(\omega) = e^{i\omega/2}b(\omega)$ or $\hat{\psi}(\omega) = -ie^{i\omega/2}b(\omega)$, where $b(\cdot)$ is a real-valued function.

Most commonly used wavelets satisfy these conditions (cf. Hernández and Weiss, 1996). For example, the Lemarie–Meyer family of wavelets is of the form $\hat{\psi}(\omega) = e^{i\omega/2}b(\omega)$, where $b(\cdot)$ is a real-valued and symmetric function on $\mathbb{R}$. An example of this family is the Meyer wavelet, defined via the Fourier transforms of $\phi(\cdot)$ and $\psi(\cdot)$:

$$\hat{\phi}(\omega) = \begin{cases} (2\pi)^{-1/2}, & |\omega| \le 2\pi/3 \\ (2\pi)^{-1/2}\cos\left[\dfrac{\pi}{2}\,v\left(\dfrac{3}{2\pi}\,|\omega| - 1\right)\right], & 2\pi/3 < |\omega| \le 4\pi/3, \\ 0, & \text{otherwise,} \end{cases} \tag{3}$$

$$\hat{\psi}(\omega) = \begin{cases} e^{i\omega/2}(2\pi)^{-1/2} \sin\left[\frac{\pi}{2} v\left(\frac{3}{2\pi}|\omega| - 1\right)\right], & 2\pi/3 \leq |\omega| \leq 4\pi/3 \\ e^{i\omega/2}(2\pi)^{-1/2} \cos\left[\frac{\pi}{2} v\left(\frac{3}{4\pi}|\omega| - 1\right)\right], & 4\pi/3 < |\omega| \leq 8\pi/3, \\ 0, & \text{otherwise} \end{cases} \tag{4}$$

where $v(x) = 0$ if $x \leq 0$ and $v(x) = 1$ if $x \geq 1$. For $x \in [0,1]$, $v(x)$ can be chosen in terms of the regularity (e.g., $v(x) = x^2(3 - 2x)$) with the restriction that $v(x) + v(1 - x) = 1$. The Meyer wavelet has compact support in the frequency domain and fast decay in the time domain.

Another family of wavelets that satisfy Assumption 2 is the spline wavelets of positive order $m \in \mathbb{Z}^+$. When $m$ is odd, this family is of the form $\hat{\psi}(\omega) = e^{i\omega/2}b(\omega)$, where $b(\cdot)$ is a real-valued and symmetric function; when $m$ is even, it is of the form $\hat{\psi}(\omega) = -ie^{i\omega/2}b(\omega)$, where $b(\cdot)$ is a real-valued and odd function (cf. Hernández and Weiss, 1996, (2.16), p. 161). One example of this family is the first order spline wavelet, often called the Franklin wavelet. It is given by

$$\hat{\phi}(\omega) = (2\pi)^{-1/2} \frac{\sin^2(\omega/2)}{(\omega/2)^2} (P_3(\omega/2))^{-1/2}, \tag{5}$$

$$\hat{\psi}(\omega) = e^{i\omega/2}(2\pi)^{-1/2} \frac{\sin^4(\omega/4)}{(\omega/4)^2} \left(\frac{P_3(\omega/4 + \pi/4)}{P_3(\omega/2)P_3(\omega/4)}\right)^{1/2}, \tag{6}$$

where $P_3(\omega) = \frac{2}{3} + \frac{1}{3}\cos(2\omega)$. Another example is the second order spline wavelet, given by

$$\hat{\phi}(\omega) = (2\pi)^{-1/2} \frac{\sin^3(\omega/2)}{(\omega/2)^3} (P_5(\omega/2))^{-1/2}, \tag{7}$$

$$\hat{\psi}(\omega) = -ie^{i\omega/2}(2\pi)^{-1/2} \frac{\sin^6(\omega/4)}{(\omega/4)^3} \left(\frac{P_5(\omega/4 + \pi/4)}{P_5(\omega/2)P_5(\omega/4)}\right)^{1/2}, \tag{8}$$

where $P_5(\omega) = \frac{1}{30}\cos^2(2\omega) + \frac{13}{30}\cos(2\omega) + \frac{8}{15}$. The Franklin wavelet and the second order spline wavelet are constructed from piecewise linear and quadratic functions, respectively. They have compact support in the time domain and have exponential decay in the frequency domain (cf. Hernández and Weiss, 1996, p. 149). In fact, the Harr wavelet (1) is the 0th order spline wavelet, but it does not satisfy Assumption 2 because its Fourier transform $\hat{\psi}(\omega) = -ie^{i\omega/2}(2\pi)^{-1/2}\sin^2(\omega/4)/(\omega/4)$ decays to 0 as $|\omega| \to \infty$ only at the rate of $|\omega|^{-1}$.

### 2.2. Wavelet Representation of the Spectral Density

The signal described previously can be a regression function or a probability density function in the time domain or a spectral density function in the frequency domain. We now consider a wavelet representation of the normalized spectral density function $f(\omega)$ of a covariance-stationary real-valued process $\{X_t, t \in \mathbb{Z}\}$. Because $f(\omega)$ is $2\pi$-periodic, it is not square integrable over $\mathbb{R}$. We need to construct a wavelet basis $\{\Phi_{jk}(\cdot), \Psi_{jk}(\cdot)\}$ for the $L_2(\Pi)$-space of $2\pi$-periodic functions, where $\Pi = [-\pi, \pi]$. Given an orthonormal wavelet basis $\{\phi_{jk}(\cdot), \psi_{jk}(\cdot)\}$ of $L_2(\mathbb{R})$, we can always construct an orthonormal wavelet basis $\{\Phi_{jk}(\cdot), \Psi_{jk}(\cdot)\}$ of $L_2(\Pi)$ by periodizing $\{\phi_{jk}(\cdot), \psi_{jk}(\cdot)\}$ via the formula

$$\Phi_{jk}(\omega) = (2\pi)^{-1/2} \sum_{m=-\infty}^{\infty} \phi_{jk}\left(\frac{\omega}{2\pi} + m\right), \tag{9}$$

$$\Psi_{jk}(\omega) = (2\pi)^{-1/2} \sum_{m=-\infty}^{\infty} \psi_{jk}\left(\frac{\omega}{2\pi} + m\right) \tag{10}$$

(cf. Daubechies, 1992, Ch. 9; Hernández and Weiss, 1996, Ch. 4). Both $\Phi_{jk}(\cdot)$ and $\Psi_{jk}(\cdot)$ are real valued.

Because (9) and (10) are an infinite sum, it is convenient to use compactly supported wavelets so that only a finite number of terms are nonzero. Alternatively, when $\psi(\cdot)$ has unbounded support, one can compute $\{\Phi_{jk}(\cdot), \Psi_{jk}(\cdot)\}$ from their Fourier transforms $\{\hat{\Phi}_{jk}(\cdot), \hat{\Psi}_{jk}(\cdot)\}$ via the formula

$$\Phi_{jk}(\omega) = (2\pi)^{-1/2} \sum_{h=-\infty}^{\infty} \hat{\Phi}_{jk}(h) e^{i\omega h}, \tag{11}$$

$$\Psi_{jk}(\omega) = (2\pi)^{-1/2} \sum_{h=-\infty}^{\infty} \hat{\Psi}_{jk}(h) e^{i\omega h}, \tag{12}$$

where

$$\hat{\Phi}_{jk}(h) = (2\pi)^{-1/2} \int_{-\pi}^{\pi} \Phi_{jk}(\omega) e^{-i\omega h} d\omega, \tag{13}$$

$$\hat{\Psi}_{jk}(h) = (2\pi)^{-1/2} \int_{-\pi}^{\pi} \Psi_{jk}(\omega) e^{-i\omega h} d\omega. \tag{14}$$

By the periodization techniques (9) and (10), and change of variables, we have

$$\hat{\Phi}_{jk}(h) = (2\pi)^{1/2} \hat{\phi}_{jk}(2\pi h) = e^{-i2\pi hk/2^j} (2\pi/2^j)^{1/2} \hat{\phi}(2\pi h/2^j), \tag{15}$$

$$\hat{\Psi}_{jk}(h) = (2\pi)^{1/2} \hat{\psi}_{jk}(2\pi h) = e^{-i2\pi hk/2^j} (2\pi/2^j)^{1/2} \hat{\psi}(2\pi h/2^j). \tag{16}$$

Note that the orthonormality of the periodized wavelet basis $\{\Phi_{jk}(\cdot), \Psi_{jk}(\cdot)\}$ implies

$$1 = \int_{-\pi}^{\pi} \Psi_{jk}^2(\omega)d\omega = \sum_{h=-\infty}^{\infty} |\hat{\Psi}_{jk}(h)|^2 = (2\pi/2^j) \sum_{h=-\infty}^{\infty} |\hat{\psi}(2\pi h/2^j)|^2, \tag{17}$$

where the second equality follows from Parseval's identity and the last one from (16). By utilizing the properties of wavelets and their Fourier transforms, one can see how dilation and translation in periodized wavelets play their roles. Dilation parameter $j$ varies dyadically, and translation parameter $k$ varies as the modulation.

Recall that $\{X_t, t \in \mathbb{Z}\}$ is a covariance-stationary real-valued time series with normalized spectral density $f(\omega)$, $\omega \in [-\pi, \pi]$. Fourier series are most often used in practice to represent $f(\omega)$, where the Fourier coefficients are autocorrelations at various lags; that is,

$$f(\omega) = (2\pi)^{-1} \sum_{h=-\infty}^{\infty} \rho(h)e^{-ih\omega}, \quad \omega \in [-\pi, \pi], \tag{18}$$

where $\rho(h) = R(h)/R(0)$, $R(h) = \mathrm{Cov}(X_t, X_{t-|h|})$. The wavelet basis is, however, more effective in capturing nonsmooth features of $f(\omega)$. Let $V_0$ be the initial subspace in $L_2(\Pi)$. Then, we can partition $L_2(\Pi) = V_0 \oplus \sum_{j=0}^{\infty} W_j$. Without loss of generality, we restrict $k \in [1, 2^j] \cap \mathbb{Z}$, because of the use of the periodized wavelet basis $\{\Phi_{jk}(\cdot), \Psi_{jk}(\cdot)\}$. Thus, with the orthonormal wavelet basis $\{\Phi_{jk}(\cdot), \Psi_{jk}(\cdot)\}$ in $L_2(\Pi)$, $f(\omega)$ can be expressed as

$$f(\omega) = \beta_{00}\Phi_{00}(\omega) + \sum_{j=0}^{\infty} \sum_{k=1}^{2^j} \alpha_{jk}\Psi_{jk}(\omega), \qquad \omega \in [-\pi, \pi], \tag{19}$$

where the wavelet coefficients

$$\beta_{00} = \int_{-\pi}^{\pi} f(\omega)\Phi_{00}(\omega)d\omega, \tag{20}$$

$$\alpha_{jk} = \int_{-\pi}^{\pi} f(\omega)\Psi_{jk}(\omega)d\omega. \tag{21}$$

The coefficients $\beta_{00}$ and $\{\alpha_{jk}\}$ are the orthogonal projections of $f(\omega)$ on wavelet bases. They are real valued. Without loss of generality, we can choose a scale function $\phi(\cdot)$ such that $\hat{\phi}(\omega) = 0$ for $|\omega| > \pi$ or $|\hat{\phi}(\omega)|$ is continuous. It follows that $(2\pi)^{1/2}\hat{\phi}(2k\pi) = 0$ for $k \in \mathbb{Z}, k \neq 0$, and $(2\pi)^{1/2}\hat{\phi}(0) = 1$ (cf. Hernández and Weiss, 1996, Proposition 2.17, p. 64). Thus, $\Phi_{00}(\omega) = (2\pi)^{-1/2}$ for $\omega \in [-\pi, \pi]$ and $\beta_{00} = (2\pi)^{-1/2}$. Consequently, we can write (19) as

$$f(\omega) = (2\pi)^{-1} + \sum_{j=0}^{\infty} \sum_{k=1}^{2^j} \alpha_{jk}\Psi_{jk}(\omega), \quad \omega \in [-\pi, \pi]. \tag{22}$$

By Parseval's identity and (16), we can also express $\alpha_{jk}$ in the time domain, namely,

$$\alpha_{jk} = (2\pi)^{-1/2} \sum_{h=-\infty}^{\infty} \rho(h)\hat{\Psi}_{jk}(h)$$
$$= \sum_{h=-\infty}^{\infty} \rho(h)\hat{\psi}_{jk}(2\pi h) = \sum_{h=-\infty}^{\infty} \rho(h)\hat{\psi}_{jk}^{*}(2\pi h), \tag{23}$$

where $\{\hat{\psi}_{jk}(\cdot)\}$ is given in (16) and the last equality follows given $\hat{\psi}^{*}(z) = \hat{\psi}(-z)$. Unlike the Fourier coefficients, $\{\alpha_{jk}\}$ do not represent autocorrelations at different lags. They are the weighted average of autocorrelations centered at varying locations. It is intuitively clear from the expression of $\{\alpha_{jk}\}$ why wavelets can capture the peaks of $f(\omega)$. Suppose, for example, that $f(\omega)$ has a peak at some frequency, say, $\omega = 0$, which can arise when $\{\rho(h)\}$ have the same, positive sign and decay to zero slowly. Such a pattern can be effectively captured by $\{\alpha_{jk}\}$ with sufficiently large $j$'s.

## 2.3. Wavelet Spectral Density Estimator and Test Statistic

Suppose that we observe a sample $\{X_t\}_{t=1}^{n}$. We define the sample autocorrelation function, $\hat{\rho}(h) = \hat{R}(h)/\hat{R}(0)$, where $\hat{R}(h) = n^{-1}\sum_{t=|h|+1}^{n}(X_t - \bar{X}_n) \times (X_{t-|h|} - \bar{X}_n)$, $\bar{X}_n = n^{-1}\sum_{t=1}^{n} X_t$. A natural choice of the estimator for $\alpha_{jk}$ is the empirical wavelet coefficient

$$\hat{\alpha}_{jk} = \sum_{h=1-n}^{n-1} \hat{\rho}(h)\hat{\psi}_{jk}(2\pi h) = \sum_{h=1-n}^{n-1} \hat{\rho}(h)\hat{\psi}_{jk}^{*}(2\pi h). \tag{24}$$

Then a wavelet estimator of the spectral density $f(\omega)$ can be given by

$$\hat{f}(\omega) = (2\pi)^{-1} + \sum_{j=0}^{J}\sum_{k=1}^{2^j} \hat{\alpha}_{jk}\Psi_{jk}(\omega), \qquad \omega \in [-\pi, \pi], \tag{25}$$

where $J \equiv J_n$ is the finest scale corresponding to the highest resolution level used in the approximation. The degree of approximation (or bias) depends on $J$. The larger $J$ is, the smaller the bias. On the other hand, $J$ also affects the sampling variation (i.e., variance) of $\hat{f}(\omega)$. The larger $J$ is, the larger the variance of $\hat{f}(\omega)$. Given each sample size $n$, a suitable $J$ should be chosen to balance the variance and the squared bias so that $\hat{f}(\omega)$ will be consistent for $f(\omega)$. There are a total of $\sum_{j=0}^{J} 2^j = 2^{J+1} - 1$ empirical wavelet coefficients in (25). The finest scale $J$ should be smaller than $\log_2 n$. Proper conditions on $J$ will be given to ensure that the proposed test statistic has a well-defined limit distribution. In our simulation, we will choose $J$ via an automatic data-driven method by Walter (1994), which, to a certain extent, lets the data themselves determine a proper $J$ given each $n$.

Now we construct a consistent test for serial correlation of unknown form, with expected good power against alternatives with nonsmooth spectrum. The hypotheses of interest are

$$H_0 : \rho(h) = 0 \text{ for all } h \in \mathbb{Z}, h \neq 0$$

versus

$$H_A : \rho(h) \neq 0 \text{ for some } h \in \mathbb{Z}, h \neq 0.$$

Note that $H_A$ includes all possible serially autocorrelated alternatives. Under the null hypothesis $H_0$, the spectral density $f(\omega)$ becomes

$$f_0(\omega) = (2\pi)^{-1} \text{ for all } \omega \in [-\pi, \pi]. \quad \textbf{(26)}$$

Hence, all the wavelet coefficients $\{\alpha_{jk}\}$ are zero under $H_0$. Under $H_A$, however, $f(\omega)$ is not a constant function of $\omega$. At least one wavelet coefficient is nonzero. Thus, testing for serial uncorrelatedness is the same as testing whether all the wavelet coefficients $\{\alpha_{jk}\}$ are jointly zero.

We now propose a test for $H_0$ versus $H_A$ using a quadratic form, defined as

$$Q(\hat{f};f_0) = \int_{-\pi}^{\pi} (\hat{f}(\omega) - f_0(\omega))^2 d\omega = \sum_{j=0}^{J} \sum_{k=1}^{2^j} \hat{\alpha}_{jk}^2, \quad \textbf{(27)}$$

where the second equality follows from the orthonormality that

$$\int_{-\pi}^{\pi} \Psi_{jk}(\omega)\Psi_{lm}(\omega)d\omega = \delta_{jl}\delta_{km}, \qquad j,k,l,m \in \mathbb{Z}. \quad \textbf{(28)}$$

Divergence measures other than the quadratic form could be used also, but (27) is convenient because it involves no numerical integration over frequency $\omega$. Our test statistic is constructed by properly standardizing (27); that is,

$$W_n = \left\{2\pi n \sum_{j=0}^{J} \sum_{k=1}^{2^j} \hat{\alpha}_{jk}^2 - (2^{J+1} - 1)\right\} \Big/ \{4(2^{J+1} - 1)\}^{1/2}, \quad \textbf{(29)}$$

where $2^{J+1} - 1$ and $4(2^{J+1} - 1)$ are approximately the mean and variance of $2\pi n Q(\hat{f};f_0)$. We could use $2^{J+1}$ to replace $2^{J+1} - 1$, but the latter is expected to give better finite sample performances when $J$ is small.

Hong (1996) considers a class of consistent tests for serial correlation of unknown form using the Parzen (1957) kernel estimator for the spectral density $f(\omega)$, which depends on Fourier analysis. Unlike the Fourier transform, (22) represents $f(\omega)$ via the wavelet transform. For the kernel-based test, kernels typically put more weight on low order autocorrelations and less weight on high order autocorrelations. In contrast, wavelets do not necessarily weigh down high order autocorrelations. Instead, they impose different weights via different scales and translations. In this sense, the wavelet-based test is expected to have better power than the kernel-based test when $f(\omega)$ has distinctive local features such as peaks and spikes. If $f(\omega)$ is smooth without singularities, however, the kernel-based test is expected to perform well. Our simulation study, which follows, confirms these claims.

## 3. ASYMPTOTIC NULL DISTRIBUTION

To derive the null limit distribution of $W_n$, we impose the following condition.

Assumption 3. $\{X_t\}_{t=-\infty}^{\infty}$ is independent and identically distributed with $E(X_t-\mu)^2=\sigma^2$ and $E(X_t-\mu)^4=\mu_4<\infty$, where $\mu=EX_t$. A random sample $\{X_t\}_{t=1}^{n}$ of size $n\in\mathbb{Z}^+$ is observed.

Assumption 3 allows non-Gaussian processes as are common for economic and financial time series. We now establish the asymptotic normality of the wavelet-based test.

THEOREM 1. *Suppose that Assumptions* 1–3 *hold and* $J\equiv J_n\to\infty$, $2^{2J}/n\to 0$. *Then* $W_n\to^d N(0,1)$.

The conditions on the finest scale $J$ ensure the asymptotic normality of $W_n$. They are analogous to the conditions on the bandwidth (or the lag order) for the kernel test of Hong (1996). The finest scale $J$ is restricted to increase at a slower rate than $\frac{1}{2}\log_2 n$. In our simulation, we use an automatic data-driven method proposed by Walter (1994) to choose $J$. Although both the finest scale $J$ in wavelet estimation and the bandwidth in kernel estimation are smoothing parameters, they are conceptually different. In particular, $J$ is not a lag order; it is the integer corresponding to the finest resolution in wavelet decomposition. At each level $j\le J$, all the sample autocorrelations $\{\hat\rho(h)\}_{h=1}^{n-1}$ are used to obtain the empirical wavelet coefficients $\{\hat\alpha_{jk}\}$ when $\hat\psi(\cdot)$ has unbounded support (see (24)).

## 4. CONSISTENCY

To establish consistency of the proposed test under the alternative hypothesis $H_A$, we impose a condition on the temporal dependence of $\{X_t\}$.

Assumption 4. $\{X_t\}_{t=-\infty}^{\infty}$ is fourth order stationary with $\sum_{h=-\infty}^{\infty}R^2(h)<\infty$ and $\sum_{j=-\infty}^{\infty}\sum_{k=-\infty}^{\infty}\sum_{l=-\infty}^{\infty}|\kappa(j,k,l)|<\infty$, where $\kappa(j,k,l)$ is the fourth order cumulant of the joint distribution of $\{X_t,X_{t+j},X_{t+k},X_{t+l}\}$, where $j,k,l\in\mathbb{Z}$.

The fourth order cumulant $\kappa(j,k,l)$ is defined as

$$\kappa(j,k,l)=E(X_tX_{t+j}X_{t+k}X_{t+l})-E(\widetilde X_t\widetilde X_{t+j}\widetilde X_{t+k}\widetilde X_{t+l}), \quad \textbf{(30)}$$

where $\{\widetilde X_t\}$ is a Gaussian sequence with the same mean and covariance function as $\{X_t\}$. Cumulant conditions are widely used in time series analysis (e.g., Anderson, 1971; Andrews, 1991; Hannan, 1970). The cumulant condition in Assumption 4 holds trivially for Gaussian processes. It is also satisfied if $\{X_t\}$ is a fourth order stationary linear process with absolutely summable coefficients and finite fourth moment (cf. Hannan, 1970). Andrews (1991, Lemma 1) provides a primitive mixing condition to ensure the cumulant condition. Neumann (1996) also uses a higher order cumulant condition.

THEOREM 2. *Suppose that Assumptions* 1, 2, *and* 4 *hold and* $J \equiv J_n \to \infty$, $2^{3J/2}/n \to 0$. *Let* $Q(f;f_0)$ *be defined as* $Q(\hat{f};f_0)$ *in* (27) *with* $\hat{f}(\cdot)$ *replaced by* $f(\cdot)$. *Then*

$$\frac{2(2^{J+1}-1)^{1/2}}{n} W_n \to^p 2\pi Q(f;f_0).$$

Theorem 2 implies that $W_n$ is consistent against $H_A$, because $Q(f;f_0) > 0$ if and only if $H_A$ holds. In particular, it is consistent against fractionally integrated processes, $I(d)$, for $d < \frac{1}{4}$. The conditions on the finest scale $J$ are weaker than those under Theorem 1.

Following reasoning analogous to that of Hong (1996), it can be shown that Bahadur's (1960) asymptotic slope of the wavelet test $W_n$ under $H_A$ does not depend on the choice of the mother wavelet $\psi(\cdot)$. In other words, the asymptotic power of $W_n$ does not depend on the choice of $\psi(\cdot)$. This is in contrast to the kernel-based test of Hong (1996, Sect. 5), for which the asymptotic power depends on the choice of a kernel function, with the Daniell kernel being optimal within a class of kernel functions.

## 5. FINITE SAMPLE PERFORMANCE

We now study the finite sample performances of the wavelet-based test $W_n$ in comparison with the kernel-based test of Hong (1996). To study the impact of the choice of mother wavelet $\psi(\cdot)$ on the size and power of $W_n$ in finite samples, we use three wavelets—Meyer, Franklin, and the second order spline wavelets. For the Meyer wavelet, we choose $v(x) = x$ for $x \in (0,1)$. We also consider more regular forms such as $v(x) = x^2(3-2x)$ for $x \in (0,1)$, but simulations show that the choice of $v(x)$ has little impact on the size and power of $W_n$.

Both the wavelet- and the kernel-based tests involve the choice of smoothing parameters—the finest scale and the lag order, which are not directly comparable because they are conceptually different. It is thus critical to choose the smoothing parameters by appropriate data-driven methods, which determine the smoothing parameters based on data information. For $W_n$, we employ the algorithm of Walter (1994) to choose $J$, where the change in the integrated mean squared error (IMSE) from one scale to the next finer scale is proportional to the sum of squared empirical wavelet coefficients. The IMSE at the scale $J$ is given by $e_J = E\int_{-\pi}^{\pi}[\hat{f}_J(\omega) - f(\omega)]^2 d\omega$, where $\hat{f}_J(\omega)$ is a wavelet spectral density estimator (25) using the finest scale $J$. The change in IMSE from $J-1$ to $J$ is proportional to $\sum_{k=1}^{2^J} \hat{\alpha}_{Jk}^2$, where $\hat{\alpha}_{Jk}$ is the empirical wavelet coefficient at the scale $J$. One starts from the initial scale $J = 0$ and checks how much the error changes from zero to one. The grid search is iterated until we get the scale $J$ at which the error increases most rapidly. Then, one obtains the finest scale. Here, we choose the finest scale $J$ for which the change in error between $J$ and $J+1$ exceeds 100%.

The kernel-based test of Hong (1996) is constructed by comparing a Parzen (1957) kernel-based spectral density estimator and the null spectral density $f_0(\omega)$. A Parzen (1957) kernel-based spectral density estimator is given by

$$\hat{f}_p(\omega) = (2\pi)^{-1} \sum_{h=1-n}^{n-1} k(h/p)\hat{\rho}(h)e^{-ih\omega}, \qquad \omega \in [-\pi, \pi],$$

where $k(\cdot)$ is a kernel function and $p \equiv p_n$ is the bandwidth such that $p \to \infty$, $p/n \to 0$.

From a standardized version of a quadratic form, the test statistic is

$$K_n = \left\{ n \sum_{h=1}^{n-1} k^2(h/p)\hat{\rho}^2(h) - C_n(k) \right\} \Big/ \{2D_n(k)\}^{1/2}, \tag{31}$$

where

$$C_n(k) = \sum_{h=1}^{n-1} (1 - h/n)k^2(h/p),$$

$$D_n(k) = \sum_{h=1}^{n-2} (1 - h/n)(1 - (h+1)/n)k^4(h/p).$$

For the choice of kernel $k(\cdot)$, we use the Daniell kernel, $k(z) = \sin(\pi z)/\pi z$, $z \in \mathbb{R}$, which maximizes the asymptotic power of $K_n$ over a class of kernel functions (cf. Hong, 1996). We choose a data-driven bandwidth $p$ using the cross-validation procedure of Beltrao and Bloomfield (1987). Here, the bandwidth is determined to maximize the cross-validated log likelihood in the frequency domain, which is asymptotically equivalent to minimizing a weighted IMSE. We use a grid search for the optimal integer-valued bandwidth $p$ over the range from 2 to 15. The algorithm is implemented by fast Fourier transform. See Beltrao and Bloomfield (1987) for more discussion.

Two sample sizes, $n = 64$ and $n = 128$, are considered. First, we study the size under normal and nonnormal processes using a GAUSS pseudo random number generator on a personal computer. Nonnormal cases include lognormal and uniform processes, scaled to have mean zero and variance one. Table 1 reports the percentage rejections of the tests $W_n$ and $K_n$ at the 10% and 5%

**TABLE 1.** Size at the 10% and 5% levels[a]

| | | $n = 64$ | | | $n = 128$ | | |
|---|---|---|---|---|---|---|---|
| | | Normal | Uniform | Lognormal | Normal | Uniform | Lognormal |
| $K_n$ | 10% | 0.115 | 0.138 | 0.097 | 0.124 | 0.138 | 0.098 |
| | 5% | 0.081 | 0.096 | 0.079 | 0.086 | 0.102 | 0.070 |
| $W_{1n}$ | 10% | 0.123 | 0.135 | 0.112 | 0.152 | 0.127 | 0.111 |
| | 5% | 0.091 | 0.096 | 0.089 | 0.096 | 0.088 | 0.072 |
| $W_{2n}$ | 10% | 0.138 | 0.136 | 0.114 | 0.145 | 0.128 | 0.114 |
| | 5% | 0.094 | 0.102 | 0.085 | 0.094 | 0.093 | 0.077 |
| $W_{3n}$ | 10% | 0.174 | 0.169 | 0.138 | 0.214 | 0.167 | 0.156 |
| | 5% | 0.121 | 0.131 | 0.103 | 0.150 | 0.126 | 0.096 |

[a]1,000 iterations.
*Abbreviations:* $K_n$, kernel-based test (Daniell kernel); $W_{1n}$, wavelet-based test (Meyer wavelet); $W_{2n}$, wavelet-based test (Franklin wavelet); $W_{3n}$, wavelet-based test (Spline wavelet of order two).

significance levels, based on 1,000 iterations. Both $K_n$ and $W_n$ have reasonable sizes at the 10% level, but they have some overrejections at the 5% level. In most cases, $W_n$ tends to overreject $H_0$ a little more than $K_n$. For all the three wavelets, $W_n$ has more accurate sizes under the lognormal process than under both normal and uniform processes at the 10% level.

Next we study the size-corrected power using empirical critical values, which ensure fair power comparison for the tests under study. The empirical critical values are obtained from 1,000 iterations for normal and nonnormal processes, respectively. Alternatives are chosen according to the shapes of their spectral densities. They are

$$ARMA(4,4): X_t + 0.1X_{t-1} + 0.3X_{t-4} = \varepsilon_t + \varepsilon_{t-4},$$

$$AR(4): X_t = 0.3X_{t-4} + \varepsilon_t,$$

$$ARIMA(0,0.2,0): (1 - B)^{0.2}X_t = \varepsilon_t, \qquad BX_t = X_{t-1},$$

$$AR(1): X_t = 0.3X_{t-1} + \varepsilon_t,$$

where $\varepsilon_t$ is i.i.d.(0,1). We consider normal, lognormal, and uniform distributions for $\varepsilon_t$, respectively. The first two alternatives generate spatially inhomogeneous features such as peaks or spikes in the spectral densities. Both ARMA(4,4) and AR(4) can arise from quarterly data. ARIMA(0,0.2,0) is a fractionally integrated process, a well-known long memory time series process. The spectral density of the long memory process is $f(\omega) = (1/2\pi)|1 - e^{-i\omega}|^{-2d}$, for $0 < d < 0.5$. It is infinite at $\omega = 0$ but is quite smooth with no jumps elsewhere. Figure 1 describes the autocorrelation function $\rho(h)$ and spectral density function $f(\omega)$ for each of the four alternatives. The autocorrelations are generated up to lag 20 using

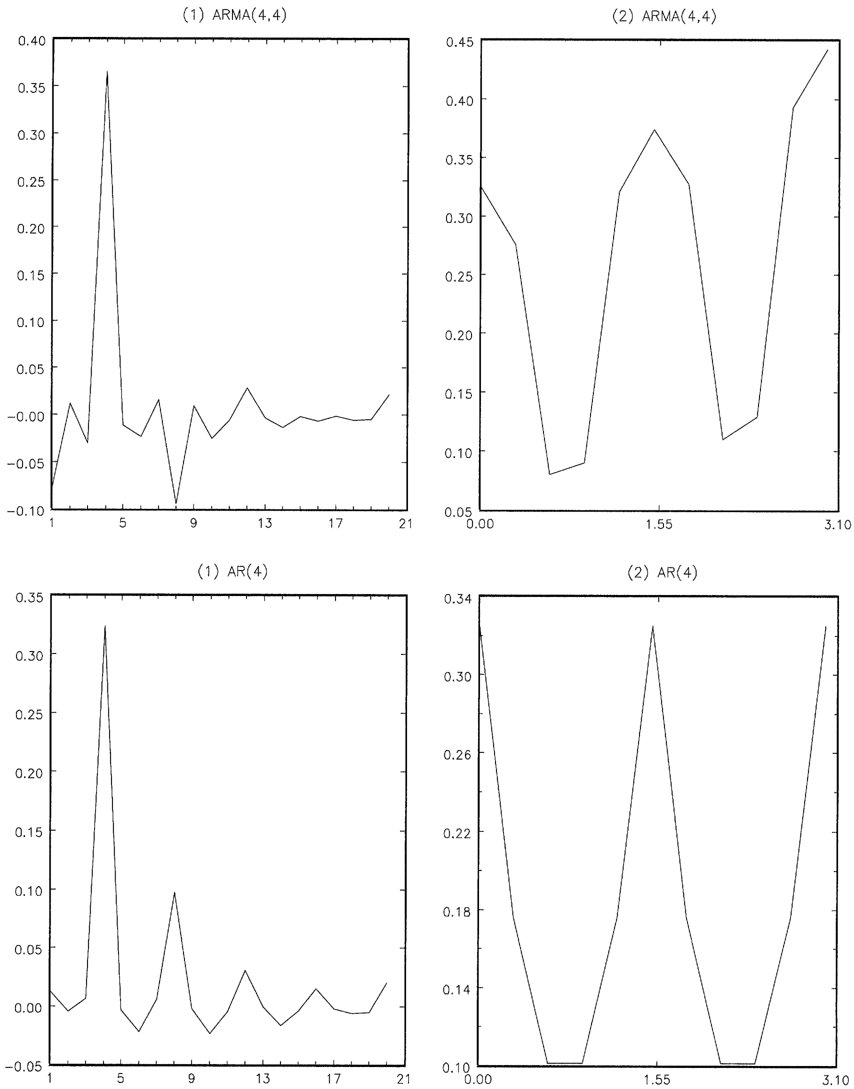


**FIGURE 1.** Sample autocorrelation (1) and true spectral density (2): ARMA(4,4) and AR(4).

a realization of the random sample $\{X_t\}_{t=1}^{10,000}$. The spectral densities are depicted for $\omega \in [0,\pi]$ (the spectral density of the long memory process is depicted for $\omega \in (0,\pi])$. There exist distinctive local features (kinks or spikes) for ARMA(4,4) and AR(4). On the other hand, the spectral density of AR(1) has no spike at any frequency.

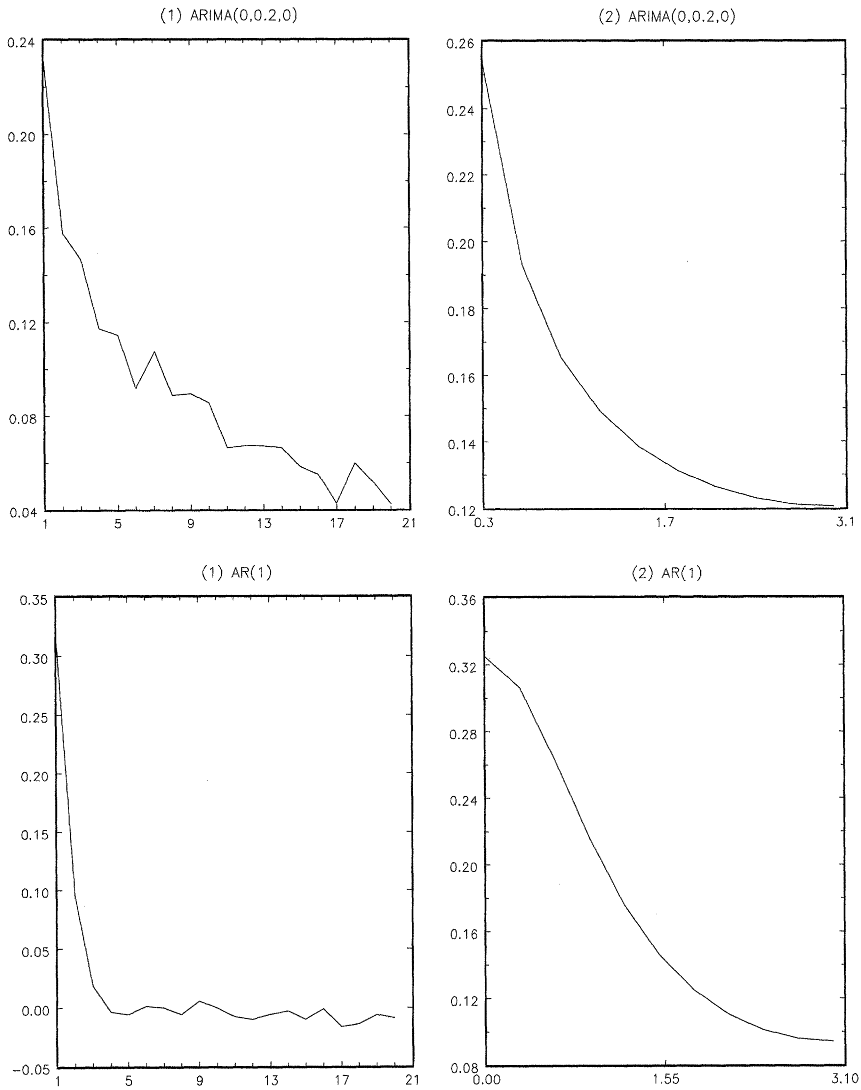


**FIGURE 1 (CONTINUED).** Sample autocorrelation (1) and true spectral density (2): ARFIMA$(0, d, 0)$ and AR(1).

Table 2 reports the power at the 10% and 5% levels when the innovations are normally distributed. For ARMA(4,4) and AR(4) processes, $W_n$ has better powers than $K_n$ regardless of the choice of the wavelet $\psi(\cdot)$ and the sample size $n$. Among the wavelets, the Meyer and Franklin wavelets have slightly better powers than the second order spline wavelet for the smaller sample size ($n = 64$).

**TABLE 2.** Size-corrected power under normal innovations[a]

| | $K_n$ | | $W_{1n}$ | | $W_{2n}$ | | $W_{3n}$ | |
|---|---|---|---|---|---|---|---|---|
| DGP | 10% | 5% | 10% | 5% | 10% | 5% | 10% | 5% |
| | | | | $n = 64$ | | | | |
| ARMA(4,4) | 0.507 | 0.392 | 0.719 | 0.534 | 0.707 | 0.533 | 0.674 | 0.516 |
| AR(4) | 0.310 | 0.219 | 0.476 | 0.340 | 0.482 | 0.347 | 0.447 | 0.332 |
| ARIMA(0,0.2,0) | 0.410 | 0.298 | 0.352 | 0.246 | 0.341 | 0.238 | 0.331 | 0.235 |
| AR(1) | 0.643 | 0.523 | 0.463 | 0.320 | 0.468 | 0.333 | 0.460 | 0.352 |
| | | | | $n = 128$ | | | | |
| ARMA(4,4) | 0.914 | 0.842 | 0.981 | 0.961 | 0.980 | 0.955 | 0.964 | 0.898 |
| AR(4) | 0.599 | 0.515 | 0.774 | 0.684 | 0.760 | 0.673 | 0.710 | 0.597 |
| ARIMA(0,0.2,0) | 0.710 | 0.650 | 0.618 | 0.521 | 0.601 | 0.522 | 0.544 | 0.435 |
| AR(1) | 0.912 | 0.869 | 0.731 | 0.652 | 0.723 | 0.636 | 0.683 | 0.585 |

[a] 1,000 iterations. Power is computed using empirical critical values obtained from 1,000 iterations under an i.i.d. $N(0,1)$ process.
*Abbreviations:* $K_n$, kernel-based test (Daniell kernel); $W_{1n}$, wavelet-based test (Meyer wavelet); $W_{2n}$, wavelet-based test (Franklin wavelet); $W_{3n}$, wavelet-based test (Spline wavelet of order two).

For the larger sample size ($n = 128$), all the three wavelets deliver similar powers, which is consistent with the asymptotic theory. For the long memory process, $W_n$ performs similarly to $K_n$. Although there is a peak at zero frequency, the spectral density of the long memory process is rather smooth elsewhere, and the kernel method performs well for this alternative. On the other hand, $K_n$ outperforms $W_n$ in detecting AR(1), which has a spatially homogeneous spectral density with no peaks.

Tables 3 and 4 report the power when the innovations are lognormally and uniformly distributed, respectively. Again, $W_n$ is superior to $K_n$ against ARMA(4,4) and AR(4). Each wavelet performs quite similarly and dominates $K_n$, particularly against AR(4). For the long memory process, $W_n$ has better power than $K_n$ for $n = 128$ when innovations are lognormally distributed, but when innovations are uniformly distributed, $K_n$ outperforms $W_n$. Also, $K_n$ obviously outperforms $W_n$ against AR(1) under both lognormal and uniform innovations. This evidence, though limited, confirms the theoretical claims that the wavelet method can more effectively detect spatial inhomogeneity of the spectrum than the kernel method. The power of the tests clearly depends on how spatially inhomogeneous the spectral density is.

In summary, we observe that (1) the choice of wavelets, in general, has little effect on the power of the wavelet-based test $W_n$, as expected from the asymptotic theory and (2) the wavelet-based test $W_n$ has better power than the kernel-based test $K_n$ when the spectral density exhibits distinct local features. When the spectral density is smooth, however, the kernel-based test $K_n$ performs better.

**TABLE 3.** Size-corrected power under lognormal innovations[a]

| | $K_n$ | | $W_{1n}$ | | $W_{2n}$ | | $W_{3n}$ | |
|---|---|---|---|---|---|---|---|---|
| DGP | 10% | 5% | 10% | 5% | 10% | 5% | 10% | 5% |
| | | | | $n = 64$ | | | | |
| ARMA(4,4) | 0.559 | 0.322 | 0.697 | 0.453 | 0.697 | 0.440 | 0.657 | 0.371 |
| AR(4) | 0.285 | 0.147 | 0.427 | 0.252 | 0.434 | 0.245 | 0.392 | 0.194 |
| ARIMA(0,0.2,0) | 0.676 | 0.538 | 0.689 | 0.552 | 0.688 | 0.531 | 0.707 | 0.557 |
| AR(1) | 0.658 | 0.440 | 0.371 | 0.222 | 0.381 | 0.215 | 0.385 | 0.204 |
| | | | | $n = 128$ | | | | |
| ARMA(4,4) | 0.938 | 0.842 | 0.982 | 0.937 | 0.980 | 0.942 | 0.972 | 0.920 |
| AR(4) | 0.630 | 0.467 | 0.808 | 0.672 | 0.787 | 0.688 | 0.773 | 0.640 |
| ARIMA(0,0.2,0) | 0.834 | 0.782 | 0.933 | 0.875 | 0.914 | 0.861 | 0.941 | 0.891 |
| AR(1) | 0.953 | 0.890 | 0.726 | 0.599 | 0.713 | 0.612 | 0.722 | 0.600 |

[a]1,000 iterations. Power is computed using empirical critical values obtained from 1,000 iterations under an i.i.d. lognormal process.
*Abbreviations:* $K_n$, kernel-based test (Daniell kernel); $W_{1n}$, wavelet-based test (Meyer wavelet); $W_{2n}$, wavelet-based test (Franklin wavelet); $W_{3n}$, wavelet-based test (Spline wavelet of order two).

## 6. CONCLUSION

We propose a wavelet-based consistent test for serial correlation of unknown form. The test is constructed by comparing a wavelet-based spectral density

**TABLE 4.** Size-corrected power under uniform innovations[a]

| | $K_n$ | | $W_{1n}$ | | $W_{2n}$ | | $W_{3n}$ | |
|---|---|---|---|---|---|---|---|---|
| DGP | 10% | 5% | 10% | 5% | 10% | 5% | 10% | 5% |
| | | | | $n = 64$ | | | | |
| ARMA(4,4) | 0.500 | 0.367 | 0.699 | 0.507 | 0.680 | 0.500 | 0.658 | 0.480 |
| AR(4) | 0.280 | 0.212 | 0.486 | 0.330 | 0.479 | 0.334 | 0.449 | 0.290 |
| ARIMA(0,0.2,0) | 0.373 | 0.288 | 0.287 | 0.191 | 0.277 | 0.185 | 0.263 | 0.177 |
| AR(1) | 0.629 | 0.520 | 0.449 | 0.301 | 0.435 | 0.289 | 0.435 | 0.291 |
| | | | | $n = 128$ | | | | |
| ARMA(4,4) | 0.875 | 0.746 | 0.979 | 0.943 | 0.979 | 0.950 | 0.952 | 0.942 |
| AR(4) | 0.570 | 0.441 | 0.785 | 0.692 | 0.769 | 0.681 | 0.746 | 0.716 |
| ARIMA(0,0.2,0) | 0.704 | 0.601 | 0.590 | 0.503 | 0.573 | 0.504 | 0.535 | 0.518 |
| AR(1) | 0.903 | 0.837 | 0.755 | 0.646 | 0.748 | 0.668 | 0.725 | 0.706 |

[a]1,000 iterations. Power is computed using empirical critical values obtained from 1,000 iterations under an i.i.d. uniform process.
*Abbreviations:* $K_n$, kernel-based test (Daniell kernel); $W_{1n}$, wavelet-based test (Meyer wavelet); $W_{2n}$, wavelet-based test (Franklin wavelet); $W_{3n}$, wavelet-based test (Spline wavelet of order two).

estimator and the null spectral density. The null asymptotic distribution of the proposed test is one-sided standard normal. Our simulation study shows that the wavelet-based test is more powerful than the kernel-based test of Hong (1996) when the data generating process has distinctive local spectral features, which confirms the theoretical claims that wavelets can effectively detect spatial inhomogeneous features. On the other hand, when the spectral density is smooth with no peaks or spikes, the kernel-based test outperforms the wavelet-based test. Because the spectral densities of time series arising in practice usually have unknown smoothness, the wavelet-based test is a useful complement to the kernel-based test in practice.

## REFERENCES

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

Anderson, T.W. (1993) Goodness of fit tests for spectral distributions. *Annals of Statistics* 21, 830–847.

Andrews, D.W.K. (1991) Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 59, 817–858.

Andrews, D.W.K. & W. Ploberger (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.

Beltrao, K. & P. Bloomfield (1987) Determining the bandwidth of a kernel spectrum estimate. *Journal of Time Series Analysis* 8, 21–38.

Box, G. & D. Pierce (1970) Distribution of residual autocorrelations in autoregressive integrated moving average time series models. *Journal of the American Statistical Association* 65,1509–1526.

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

Daubechies, I. (1992) *Ten Lectures on Wavelets*. Philadelphia: SIAM.

Donoho, D.L. & I.M. Johnstone (1994) Ideal spatial adaptation by wavelet shrinkage. *Biometrika* 81, 425–455.

Donoho, D.L. & I.M. Johnstone (1995a) Adapting to unknown smoothness via wavelet shrinkage. *Journal of the American Statistical Association* 90,1200–1224.

Donoho, D.L. & I.M. Johnstone (1995b) Wavelet shrinkage: Asymptopia? *Journal of the Royal Statistical Society, Series B* 57, 301–369.

Donoho, D.L., I.M. Johnstone, G. Kerkyacharian, & D. Picard (1996) Density estimation by wavelet thresholding. *Annals of Statistics* 24, 508–539.

Durbin, J. & G.S. Watson (1950) Testing for serial correlation in least squares regression: I. *Biometrika* 37, 409–428.

Durbin, J. & G.S. Watson (1951) Testing for serial correlation in least squares regression: II. *Biometrika* 38, 159–178.

Durlauf, S. (1991) Spectral based testing for the martingale hypothesis. *Journal of Econometrics* 50, 1–19.

Gao, H.Y. (1993) *Wavelet Estimation of Spectral Densities in Time Series Analysis*. Ph.D. thesis, Department of Statistics, University of California, Berkeley.

Gilbert, S. (1995) Structural Change: Estimation and Testing by Wavelet Regression. Manuscript, Department of Economics, University of California, San Diego.

Godfrey, L. (1978a) Testing against general autoregressive and moving average error models when the regressors include lagged dependent variables. *Econometrica* 46, 1293–1302.

Godfrey, L. (1978b) Testing for higher order serial correlation in regression equations when the regressors include lagged dependent variables. *Econometrica* 46, 1303–1310.

Goffe, W.L. (1994) Wavelets in macroeconomics: An introduction. In D. Belsley (ed.), *Computational Techniques for Econometrics and Economic Analysis*, pp. 137–149. The Netherlands: Kluwer Academic Publishers.

Hannan, E. (1970) *Multiple Time Series*. New York: Wiley.
Hernandez, E. & G. Weiss (1996) *A First Course on Wavelets*. Boca Raton, Florida: CRC Press.
Hong, Y. (1996) Consistent testing for serial correlation of unknown form. *Econometrica* 64, 837–864.
Jensen, M.J. (2000) An alternative maximum likelihood estimator of long memory processes using compactly supported wavelets. *Journal of Economic Dynamics and Control* 24, 361–387.
Neumann, M.H. (1996) Spectral density estimation via nonlinear wavelet methods for stationary non-Gaussian time series. *Journal of Time Series Analysis* 17, 601–633.
Parzen, E. (1957) On consistent estimate of the spectrum of a stationary time series. *Annals of Mathematical Statistics* 28, 329–348.
Priestley, M.B. (1981) *Spectral Analysis and Time Series*. New York: Academic Press.
Priestley, M.B. (1996) Wavelets and time-dependent spectral analysis. *Journal of Time Series Analysis* 17, 85–103.
Ramsey, J.B. (1999) Regression over time scale decomposition: A sampling analysis of distributional properties. *Economic Systems Research* 11, 163–183.
Ramsey, J.B. & C. Lampart (1998a) The decomposition of economic relationships by time scale using wavelets: Money and income. *Macroeconomic Dynamics* 2, 49–71.
Ramsey, J.B. & C. Lampart (1998b) The decomposition of economic relationships by time scale using wavelets: Expenditure and income. *Studies for Nonlinear Dynamics and Econometrics* 3, 23–42.
Ramsey, J.B., D. Usikov, & G.M. Zaslavsky (1995) An analysis of U.S. stock price behavior using wavelets. *Fractals* 3 (2), 377–389.
Ramsey, J.B. & Z. Zhang (1996) The application of wave form dictionaries to stock market index data. In J.B. Kadtke & Y.A. Kravtsov (eds.), *Predictability of Complex Dynamical Systems*, pp. 189–205. New York: Springer-Verlag.
Ramsey, J.B. & Z. Zhang (1997) The analysis of foreign exchange data using waveform dictionaries. *Journal of Empirical Finance* 4, 341–372.
Robinson, P.M. (1991) Testing for strong serial correlation and dynamic conditional heteroskedasticity in multiple regression. *Journal of Econometrics* 47, 67–84.
Strang, G. & T. Nguyen (1996) *Wavelets and Filter Banks*. Wellesley, Massachusetts: Wellesley-Cambridge Press.
Subba Rao, T. & K.C. Indukumar (1996) Spectral and wavelet methods for the analysis of nonlinear and nonstationary time series. *Journal of the Franklin Institute* 333(b) (3), 425–452.
Walter, G. (1994) *Wavelets and Other Orthogonal Systems with Applications*. Boca Raton, Florida: CRC Press.
Wang, Y. (1995) Jump and sharp cusp detection by wavelets. *Biometrika* 82, 385–397.
Wen, Y. (1998) Investment cycles. *Journal of Economic Dynamics and Control* 22, 1139–1165.
Whang, Y. (1997) A test of autocorrelation in the presence of heteroskedasticity of unknown form. *Econometric Theory* 14, 87–122.

# APPENDIX A

To prove Theorems 1 and 2, we first state a useful lemma, which is proved in Appendix B.

LEMMA A.1. *Suppose that Assumptions* 1 *and* 2 *hold and* $J \to \infty, 2^J/n \to 0$. *Define*

$$b_J(h,m) = a_J(h,m) + a_J(-h,-m) + a_J(h,-m) + a_J(-h,m),$$

*where* $a_J(h,m) = 2\pi \sum_{j=0}^{J} \sum_{k=1}^{2^j} \hat{\psi}_{jk}(2\pi h)\hat{\psi}_{jk}^*(2\pi m)$. *Then*

(i) $b_J(h,m)$ *is real valued*, $b_J(0,m) = b_J(h,0) = 0$, *and* $b_J(h,m) = b_J(m,h)$;
(ii) $\sum_{h=1}^{n-1}\sum_{m=1}^{n-1} h^\nu |b_J(h,m)| = O(2^{(1+\nu)J})$ *for* $0 \le \nu \le 1/2$;

(iii) $\sum_{h=1}^{n-1}\{\sum_{m=1}^{n-1}|b_J(h,m)|\}^2 = O(2^J)$;
(iv) $\sum_{h_1=1}^{n-1}\sum_{h_2=1}^{n-1}\{\sum_{m=1}^{n-1}|b_J(h_1,m)b_J(h_2,m)|\}^2 = O\{(J+1)2^J\}$;
(v) $\sum_{h=1}^{n-1} b_J(h,h) = (2^{J+1}-1)\{1 + O[(J+1)/2^J + (2^J/n)^{2\alpha-1}]\}$, *where* $\alpha$ *is in Assumption* 2;
(vi) $\sum_{h=1}^{n-1}\sum_{m=1}^{n-1} b_J^2(h,m) = 2(2^{J+1}-1)\{1+o(1)\}$.

**Proof of Theorem 1.** For simplicity and without loss of generality, we assume $E(X_t) = 0$ and consider $\hat{R}(h) = n^{-1}\sum_{t=|h|+1}^{n} X_t X_{t-|h|}$. It can be shown that subtracting the sample mean $\bar{X}_n = n^{-1}\sum_{t=1}^{n} X_t$ from $\{X_t\}$ has no impact on the limit distribution of the test statistic $W_n$. Using (24), we have

$$\begin{aligned}2\pi n \sum_{j=0}^{J}\sum_{k=1}^{2^j} \hat{\alpha}_{jk}^2 &= 2\pi n \sum_{j=0}^{J}\sum_{k=1}^{2^j}\left(\sum_{h=1-n}^{n-1} \hat{\rho}(h)\hat{\psi}_{jk}(2\pi h)\right)\left(\sum_{m=1-n}^{n-1} \hat{\rho}(m)\hat{\psi}_{jk}^*(2\pi m)\right)\\ &= n \sum_{h=1-n}^{n-1}\sum_{m=1-n}^{n-1} a_J(h,m)\hat{\rho}(h)\hat{\rho}(m)\\ &= n \sum_{h=1}^{n-1}\sum_{m=1}^{n-1} b_J(h,m)\hat{\rho}(h)\hat{\rho}(m), \qquad \textbf{(A.1)}\end{aligned}$$

where the last equality follows from reindexing and the definition of $b_J(h,m)$.

Using $\hat{\rho}(h) = \hat{R}(h)/\hat{R}(0)$ and $\hat{R}(0) - \sigma^2 = O_P(n^{-1/2})$ by Assumption 3, we have

$$\begin{aligned}n\sum_{h=1}^{n-1}\sum_{m=1}^{n-1} b_J(h,m)\hat{\rho}(h)\hat{\rho}(m) &= \sigma^{-4} n \sum_{h=1}^{n-1}\sum_{m=1}^{n-1} b_J(h,m)\hat{R}(h)\hat{R}(m)\\ &\quad + \{\hat{R}^{-2}(0) - \sigma^{-4}\} n \sum_{h=1}^{n-1}\sum_{m=1}^{n-1} b_J(h,m)\hat{R}(h)\hat{R}(m)\\ &= \sigma^{-4} n \sum_{h=1}^{n-1}\sum_{m=1}^{n-1} b_J(h,m)\hat{R}(h)\hat{R}(m) + O_P(2^J/n^{1/2}), \qquad \textbf{(A.2)}\end{aligned}$$

where the second term is of the indicated order of magnitude because

$$E\sum_{h=1}^{n-1}\sum_{m=1}^{n-1}|b_J(h,m)\hat{R}(h)\hat{R}(m)| \le Cn^{-1}\sum_{h=1}^{n-1}\sum_{m=1}^{n-1}|b_J(h,m)| = O(2^J/n)$$

given $E\hat{R}^2(h) \le Cn^{-1}$ and Lemma A.1(ii).

We now focus on the first term in (A.2). Given $\hat{R}(h) = n^{-1}\sum_{t=h+1}^{n} X_t X_{t-h}$,

$$\begin{aligned}n\sum_{h=1}^{n-1}\sum_{m=1}^{n-1} b_J(h,m)\hat{R}(h)\hat{R}(m) &= n^{-1}\sum_{h=1}^{n-1}\sum_{m=1}^{n-1} b_J(h,m)\sum_{t=h+1}^{n}\sum_{s=m+1}^{n} X_t X_{t-h} X_s X_{s-m}\\ &= n^{-1}\sum_{h=1}^{n-1}\sum_{m=1}^{n-1} b_J(h,m)\\ &\quad \times\left(\sum_{t=1}^{n}\sum_{s=1}^{n} - \sum_{t=1}^{h}\sum_{s=m+1}^{n} - \sum_{t=1}^{n}\sum_{s=1}^{m}\right) X_t X_{t-h} X_s X_{s-m}\\ &= \hat{A}_n + \hat{B}_{1n} - \hat{B}_{2n} - \hat{B}_{3n}, \qquad \textbf{(A.3)}\end{aligned}$$

where

$$\hat{A}_n = n^{-1} \sum_{h=1}^{n-1} \sum_{m=1}^{n-1} b_J(h,m) \left( \sum_{t=2}^{n} \sum_{s=1}^{t-1} + \sum_{s=2}^{n} \sum_{t=1}^{s-1} \right) X_t X_{t-h} X_s X_{s-m}$$

$$= 2n^{-1} \sum_{h=1}^{n-1} \sum_{m=1}^{n-1} b_J(h,m) \sum_{t=2}^{n} \sum_{s=1}^{t-1} X_t X_{t-h} X_s X_{s-m} \quad \text{given } b_J(h,m) = b_J(m,h),$$

$$\hat{B}_{1n} = n^{-1} \sum_{h=1}^{n-1} \sum_{m=1}^{n-1} b_J(h,m) \sum_{t=1}^{n} X_t^2 X_{t-h} X_{t-m},$$

$$\hat{B}_{2n} = n^{-1} \sum_{h=1}^{n-1} \sum_{m=1}^{n-1} b_J(h,m) \sum_{t=1}^{h} \sum_{s=m+1}^{n} X_t X_{t-h} X_s X_{s-m},$$

$$\hat{B}_{3n} = n^{-1} \sum_{h=1}^{n-1} \sum_{m=1}^{n-1} b_J(h,m) \sum_{t=1}^{n} \sum_{s=1}^{m} X_t X_{t-h} X_s X_{s-m}.$$

Proposition 1 shows that the $U$-statistic $\hat{A}_n$ is dominant.

PROPOSITION 1. *Suppose that Assumptions* 1–3 *hold and* $J \to \infty, 2^{2J}/n \to 0$. *Then* $2^{-J/2}\{2\pi n \sum_{j=0}^{J} \sum_{k=1}^{2^j} \hat{\alpha}_{jk}^2 - (2^{J+1} - 1)\} = 2^{-J/2}\sigma^{-4}\hat{A}_n + o_P(1)$.

We now decompose $\hat{A}_n$ into the terms with $t - s > q$ and $t - s \leq q$, for some $q \in \mathbb{Z}^+$:

$$\hat{A}_n = 2n^{-1} \sum_{h=1}^{n-1} \sum_{m=1}^{n-1} b_J(h,m) \left( \sum_{t=q+2}^{n} \sum_{s=1}^{t-q-1} + \sum_{t=2}^{n} \sum_{s=\max(t-q,1)}^{t-1} \right) X_t X_{t-h} X_s X_{s-m}$$

$$= \hat{B}_n + \hat{B}_{4n}, \quad \text{say.} \tag{A.4}$$

Furthermore, we decompose

$$\hat{B}_n = 2n^{-1} \left( \sum_{h=1}^{q} \sum_{m=1}^{q} + \sum_{h=1}^{q} \sum_{m=q+1}^{n-1} + \sum_{h=q+1}^{n-1} \sum_{m=1}^{n-1} \right) b_J(h,m) \sum_{t=q+2}^{n} \sum_{s=1}^{t-q-1} X_t X_{t-h} X_s X_{s-m}$$

$$= \hat{U}_n + \hat{B}_{5n} + \hat{B}_{6n}, \quad \text{say,} \tag{A.5}$$

where $B_{5n}$ and $\hat{B}_{6n}$ are the contributions from $m > q$ and $h > q$, respectively.

Proposition 2 shows that $\hat{A}_n$ can be approximated arbitrarily well by $\hat{U}_n$ under a proper condition on $q$.

PROPOSITION 2. *Suppose that Assumptions* 1–3 *hold*, $J \to \infty, 2^{2J}/n \to 0$, *and* $q \equiv q_n \to \infty, q/2^J \to \infty, q^2/n \to 0$. *Then* $2^{-J/2}\hat{A}_n = 2^{-J/2}\hat{U}_n + o_P(1)$.

It is much easier to show the asymptotic normality of $\hat{U}_n$ than of $\hat{A}_n$, because for $\hat{U}_n$, $\{X_t X_{t-h}\}$ and $\{X_s X_{s-m}\}$ are independent given $t - s > q$ and $0 < h, m \leq q$.

PROPOSITION 3. *Suppose that Assumptions* 1–3 *hold and* $J \to \infty, 2^{2J}/n \to 0$, $q/2^J \to \infty, q^2/n \to 0$. *Let* $\lambda_n^2 = E\hat{U}_n^2$. *Then* $4(2^{J+1} - 1)\sigma^8/\lambda_n^2 \to 1$, *and* $\lambda_n^{-1}\hat{U}_n \to^d N(0,1)$.

Propositions 1–3 and the Slutsky theorem imply $W_n \to^d N(0,1)$. The proof of Theorem 1 will be completed provided Propositions 1–3 are shown.

**Proof of Proposition 1.** By (A.1)–(A.3), we obtain

$$2\pi n \sum_{j=0}^{J} \sum_{k=1}^{2^j} \hat{\alpha}_{jk}^2 - (2^{J+1} - 1) = \sigma^{-4}\{\hat{A}_n + (\hat{B}_{1n} - \sigma^4(2^{J+1} - 1)) - \hat{B}_{2n} - \hat{B}_{3n}\} + O_P(2^J/n^{1/2}).$$

We shall show that (i) $2^{-J/2}\{\hat{B}_{1n} - \sigma^4(2^{J+1} - 1)\} \to^p 0$; (ii) $2^{-J/2}\hat{B}_{2n} \to^p 0$; (iii) $2^{-J/2}\hat{B}_{3n} \to^p 0$.

(i) Given Assumption 3 and $E(X_t^2 X_{t-h} X_{t-m}) = \sigma^4 \delta_{hm}$ for $h, m > 0$, we have

$$E\left[\sum_{t=1}^{n} (X_t^2 X_{t-h} X_{t-m} - \sigma^4 \delta_{hm})\right]^2 \le Cn \quad \text{for any } h, m > 0.$$

It follows by Minkowski's inequality and Lemma A.1(ii) that

$$\begin{aligned} E(\hat{B}_{1n} - E\hat{B}_{1n})^2 &\le n^{-2}\left\{\sum_{h=1}^{n-1} \sum_{m=1}^{n} |b_J(h,m)| \left[E\left(\sum_{t=1}^{n} (X_t^2 X_{t-h} X_{t-m} - \sigma^4 \delta_{hm})\right)^2\right]^{1/2}\right\}^2 \\ &\le Cn^{-1}\left\{\sum_{h=1}^{n-1} \sum_{m=1}^{n-1} |b_J(h,m)|\right\}^2 = O(2^{2J}/n). \end{aligned}$$

Hence, by Chebyshev's inequality, we have

$$\hat{B}_{1n} - E\hat{B}_{1n} = O_P(2^J/n^{1/2}). \tag{A.6}$$

Next, given $E(X_t^2 X_{t-h} X_{t-m}) = \sigma^4 \delta_{hm}$ for any $h, m > 0$, and Lemma A.1(v), we have

$$E\hat{B}_{1n} = \sigma^4 \sum_{h=1}^{n-1} b_J(h,h) = \sigma^4(2^{J+1} - 1)\{1 + O[(J+1)/2^J + (2^J/n)^{2\alpha - 1}]\}. \tag{A.7}$$

Combining (A.6), (A.7), $2^{2J}/n \to 0, J \to \infty$, and $\alpha > \frac{3}{2}$ then yields

$$\begin{aligned} 2^{-J/2}\{\hat{B}_{1n} - \sigma^4(2^{J+1} - 1)\} &= 2^{-J/2}\{\hat{B}_{1n} - E\hat{B}_{1n} + E\hat{B}_{1n} - \sigma^4(2^{J+1} - 1)\} \\ &= O_P(2^{J/2}/n^{1/2} + (J+1)/2^{J/2} + 2^{J(2\alpha - 1/2)}/n^{2\alpha - 1}) \\ &= o_P(1). \end{aligned}$$

(ii) Next, we consider $\hat{B}_{2n}$. Given Assumption 3, we have that for any $h, m > 0$,

$$E\left(\sum_{t=1}^{h} \sum_{s=m+1}^{n} X_t X_{t-h} X_s X_{s-m}\right)^2 \le \left[E\left(\sum_{t=1}^{h} X_t X_{t-h}\right)^4 E\left(\sum_{s=m+1}^{n} X_s X_{s-m}\right)^4\right]^{1/2} \le Cnh.$$

It follows by Minkowski's inequality and Lemma A.1(ii) that

$$\begin{aligned} E\hat{B}_{2n}^2 &\le 4n^{-2}\left\{\sum_{h=1}^{n-1} \sum_{m=1}^{n-1} |b_J(h,m)| \left[E\left(\sum_{t=1}^{h} \sum_{s=m+1}^{n} X_t X_{t-h} X_s X_{s-m}\right)^2\right]^{1/2}\right\}^2 \\ &\le 4Cn^{-1}\left\{\sum_{h=1}^{n-1} \sum_{m=1}^{n-1} h^{1/2} |b_J(h,m)|\right\}^2 = O(2^{3J}/n). \end{aligned}$$

Hence, $2^{-J/2}\hat{B}_{2n} = O_P(2^J/n^{1/2}) = o_P(1)$ by Chebyshev's inequality and $2^{2J}/n \to 0$.

(iii) By reasoning similar to (ii), we can obtain $2^{-J/2}\hat{B}_{3n} = O_P(2^J/n^{1/2}) = o_P(1)$. ■

**Proof of Proposition 2.** Given (A.4) and (A.5), $\hat{A}_n = \hat{U}_n + \hat{B}_{4n} + \hat{B}_{5n} + \hat{B}_{6n}$. It suffices to show $2^{-J/2}\hat{B}_{jn} \to^p 0$ for $j = 4,5,6$.

(i) We first consider $\hat{B}_{4n}$. Given Assumption 3,

$$E\left(\sum_{t=2}^{n}\sum_{s=\max(t-q,1)}^{t-1} X_t X_{t-h} X_s X_{s-m}\right)^2 \le Cnq \quad \text{for } h,m > 0.$$

It follows by Minkowski's inequality and Lemma A.1(ii) that

$$E\hat{B}_{4n}^2 \le 4n^{-2}\left\{\sum_{h=1}^{n-1}\sum_{m=1}^{n-1}|b_J(h,m)|\left[E\left(\sum_{t=2}^{n}\sum_{s=t-q}^{t-1} X_t X_{t-h} X_s X_{s-m}\right)^2\right]^{1/2}\right\}^2$$

$$\le 4Cqn^{-1}\left\{\sum_{h=1}^{n-1}\sum_{m=1}^{n-1}|b_J(h,m)|\right\}^2 = O(q2^{2J}/n).$$

This, along with Chebyshev's inequality and $q^2/n \to 0, 2^{2J}/n \to 0$, implies $2^{-J/2}\hat{B}_{4n} = O_P(q^{1/2}2^{J/2}/n^{1/2}) = o_P(1)$.

(ii) Next, we consider $\hat{B}_{5n}$. Define the partial sum

$$S_{t-q-1}(m) = \sum_{s=1}^{t-q-1} X_s X_{s-m}. \tag{A.8}$$

Noting that $X_t X_{t-h}$ is independent of $S_{t-q-1}(m)$ for $0 < h \le q$ and $m > 0$, we have

$$E\hat{B}_{5n}^2 = 4n^{-2}E\left\{\sum_{t=q+2}^{n}\sum_{h=1}^{q} X_t X_{t-h}\sum_{m=q+1}^{n-1} b_J(h,m)S_{t-q-1}(m)\right\}^2$$

$$= 4\sigma^8 n^{-2}\sum_{t=q+2}^{n}\sum_{h=1}^{q}\sum_{s=1}^{t-q-1}\sum_{m=q+1}^{n-1} b_J^2(h,m)$$

$$\le 2\sigma^8\sum_{h=1}^{n-1}\sum_{m=q+1}^{n-1} b_J^2(h,m) = o(2^J),$$

where the last inequality follows from $q \to \infty, q/2^J \to \infty$, and Lemma A.1(vi). Therefore, $2^{-J/2}\hat{B}_{5n} \to^p 0$ by Chebyshev's inequality.

(iii) Finally, we consider $\hat{B}_{6n}$. Because $X_t$ is independent of $X_{t-h}S_{t-q-1}(m)$ for $h$, $m > 0$,

$$E\hat{B}_{6n} = 4\sigma^2 n^{-2}\sum_{t=q+2}^{n} E\left[\sum_{h=q+1}^{n-1}\sum_{m=1}^{n-1} b_J(h,m)X_{t-h}S_{t-q-1}(m)\right]^2. \tag{A.9}$$

We now decompose the expectation in (A.9):

$$E\left[\sum_{h=q+1}^{n-1}\sum_{m=1}^{n-1} b_J(h,m)X_{t-h}S_{t-q-1}(m)\right]^2 = E\left[\sum_{h=q+1}^{n-1}\sum_{m=1}^{n-1}\sum_{s=1}^{t-q-1} b_J(h,m)X_{t-h}X_s X_{s-m}\right]^2$$

$$\le 4(C_{1nt} + C_{2nt} + C_{3nt}), \tag{A.10}$$

where

$$C_{1nt} = E\left[\sum_{h=q+1}^{n-1}\sum_{m=1}^{n-1}\sum_{s=1}^{t-q-1} b_J(h,m)X_{t-h}X_sX_{s-m}1(t-h>s)\right]^2,$$

$$C_{2nt} = E\left[\sum_{h=q+1}^{n-1}\sum_{m=1}^{n-1}\sum_{s=1}^{t-q-1} b_J(h,m)X_{t-h}X_sX_{s-m}1(t-h=s)\right]^2,$$

$$C_{3nt} = E\left[\sum_{h=q+1}^{n-1}\sum_{m=1}^{n-1}\sum_{s=1}^{t-q-1} b_J(h,m)X_{t-h}X_sX_{s-m}1(t-h<s)\right]^2.$$

For $C_{1nt}$, because $X_{t-h}$ is independent of $\{X_sX_{s-m}\}_{s=1}^{t-q-1}$ for $t-h>s$ and $m>0$, we have

$$C_{1nt} = \sigma^6\sum_{h=q+1}^{n-1}\sum_{s=1}^{t-q-1}\sum_{m=1}^{n-1} b_J^2(h,m)1(t-h>s) \le \sigma^6 t\sum_{h=q+1}^{n-1}\sum_{m=1}^{n-1} b_J^2(h,m). \tag{A.11}$$

For $C_{2nt}$, noting that $X_{t-h}X_sX_{s-m} = X_{t-h}^2X_{t-h-m}$ given $t-h=s$, we have

$$C_{2nt} = E\left(\sum_{h=q+1}^{n-1}\sum_{m=1}^{n-1} b_J(h,m)X_{t-h}^2X_{t-h-m}\right)^2 \le C\left(\sum_{h=q+1}^{n-1}\sum_{m=1}^{n-1}|b_J(h,m)|\right)^2. \tag{A.12}$$

For $C_{3nt}$, noting that $X_s$ is independent of $\{X_{t-h}X_{s-m}\}$ for $t-h<s$ and $m>0$, we have

$$C_{3nt} = \sigma^2\sum_{s=1}^{t-q-1} E\left(\sum_{h=q+1}^{n-1}\sum_{m=1}^{n-1} b_J(h,m)X_{t-h}X_{s-m}1(t-h<s)\right)^2,$$

where

$$\begin{aligned}
&E\left(\sum_{h=q+1}^{n-1}\sum_{m=1}^{n-1} b_J(h,m)X_{t-h}X_{s-m}1(t-h<s)\right)^2\\
&\quad\le 2E\left(\sum_{h=q+1}^{n-1}\sum_{m=1}^{n-1} b_J(h,m)X_{t-h}X_{s-m}1(t-h\ne s-m)\right)^2\\
&\qquad + 2E\left(\sum_{h=q+1}^{n-1}\sum_{m=1}^{n-1} b_J(h,m)X_{t-h}X_{s-m}1(t-h=s-m)\right)^2\\
&\quad\le C\sum_{h=q+1}^{n-1}\sum_{m=1}^{n-1} b_J^2(h,m) + C\left(\sum_{m=1}^{n-1}|b_J(m+t-s,m)|\right)^2.
\end{aligned}$$

It follows that

$$C_{3nt} \le Ct\sum_{h=q+1}^{n-1}\sum_{m=1}^{n-1} b_J^2(h,m) + C\left(\sum_{h=1}^{n-1}\sum_{m=1}^{n-1}|b_J(h,m)|\right)^2. \tag{A.13}$$

Combining (A.9)–(A.13), Lemma A.1(ii and vi), and $q \to \infty, q/2^J \to \infty$ yields

$$E\hat{B}_{6n}^2 \leq C \sum_{h=q+1}^{n-1} \sum_{m=1}^{n-1} b_J^2(h,m) + Cn^{-1}\left(\sum_{h=1}^{n-1}\sum_{m=1}^{n-1} |b_J(h,m)|\right)^2 = o(2^J) + O(2^{2J}/n).$$

Thus, $2^{-J/2}\hat{B}_{6n} \to^p 0$ by Chebyshev's inequality and $2^{2J}/n \to 0$, This completes the proof. ■

**Proof of Proposition 3.** We write $\hat{U}_n = n^{-1}\sum_{t=q+2}^n U_{nt}$, where

$$U_{nt} = 2X_t \sum_{h=1}^q X_{t-h}H_{t-q-1}(h), \tag{A.14}$$

where $H_{t-q-1}(h) = \sum_{m=1}^q b_J(h,m)S_{t-q-1}(m)$ and $S_{t-q-1}(m)$ is defined in (A.8). Let $\mathcal{F}_t$ be the sigma field consisting of $X_s, s \leq t$. Because $\{X_tX_{t-h}\}$ is independent of $H_{t-q-1}(h)$ for $0 < h \leq q$, $\{U_{nt}, \mathcal{F}_{t-1}\}$ is an adapted martingale difference sequence, with

$$\begin{aligned}
\lambda_n^2 &= n^{-2}\sum_{t=q+2}^n EU_{nt}^2 \\
&= 4\sigma^8 n^{-2}\sum_{t=q+2}^n (t-q-1)\sum_{h=1}^q\sum_{m=1}^q b_J^2(h,m) \\
&= 2\sigma^8(1-q/n)(1-(q+1)/n)\sum_{h=1}^q\sum_{m=1}^q b_J^2(h,m) \\
&= 4(2^{J+1}-1)\sigma^8\{1+o(1)\},
\end{aligned} \tag{A.15}$$

by Lemma A.1(vi) and $q \to \infty, q/2^J \to \infty, q^2/n \to 0$. It follows that $4(2^{J+1}-1)\sigma^8/\lambda_n^2 \to 1$.

We apply Brown's (1971) martingale limit theorem by verifying his two conditions: (i) $\lambda_n^{-2}n^{-2}\sum_{t=q+2}^n E[U_{nt}^2 1(|U_{nt}| \geq \epsilon n\lambda_n)] \to 0$ for all $\epsilon > 0$ and (ii) $\lambda_n^{-2}n^{-2} \times \sum_{t=q+2}^n E(U_{nt}^2|\mathcal{F}_{t-1}) \to^p 1$. Because

$$\begin{aligned}
\lambda_n^{-2}n^{-2}\sum_{t=q+2}^n E[U_{nt}^2 1(|U_{nt}| > \epsilon n\lambda_n)] &= \lambda_n^{-2}n^{-2}\sum_{t=q+2}^n \int_{|u|>\epsilon n\lambda_n} u^2 dF_{nt}(u) \\
&\leq \lambda_n^{-4}n^{-4}\epsilon^{-2}\sum_{t=q+2}^n EU_{nt}^4,
\end{aligned}$$

where $F_{nt}(u)$ is the cumulative distribution function (c.d.f.) of $U_{nt}$, it suffices for (i) if $\lambda_n^{-4}n^{-4}\sum_{tq+2}^n EU_{nt}^4 \to 0$. Given the independence between $X_{t-h}$ and $H_{t-q-1}(h)$ for $0 < h \leq q$, we can use the iterated expectation $EU_{nt}^4 = E[E(U_{nt}^4|\mathcal{F}_{t-q-1})]$ and obtain

$$\begin{aligned}
EU_{nt}^4 &= 16\mu_4 E\left[\sum_{h=1}^q X_{t-h}H_{t-q-1}(h)\right]^4 \\
&\leq 48\mu_4\left\{\sum_{h=1}^q (EX_{t-h}^4)^{1/2}(EH_{t-q-1}^4(h))^{1/2}\right\}^2 \\
&\leq Ct^2\left\{\sum_{h=1}^q\sum_{m=1}^q b_J^2(h,m)\right\}^2 = O(t^2 2^{2J})
\end{aligned}$$

by Lemma A.1(vi), where we used the fact that given Assumption 3,

$$EH_{t-q-1}^4(h) = E\left(\sum_{s=1}^{t-q-1} X_s \sum_{m=1}^{q} b_J(h,m)X_{s-m}\right)^4$$

$$\leq C\left(\sum_{s=1}^{t-q-1} (EX_s^4)^{1/2}\left[E\left(\sum_{m=1}^{q} b_J(h,m)X_{s-m}\right)^4\right]^{1/2}\right)^2$$

$$\leq Ct^2\left(\sum_{m=1}^{q} b_J^2(h,m)\right)^2. \tag{A.16}$$

It follows that condition (i) holds because $\lambda_n^{-4}n^{-4}\sum_{t=q+2}^{n} EU_{nt}^4 = O(n^{-1})$.

Next, we verify condition (ii), which holds if $\lambda_n^{-4}E(\tilde{U}_n^2 - \lambda_n^2) \to 0$, where

$$\tilde{U}_n^2 = n^{-2}\sum_{t=q+2}^{n} E(U_{nt}^2|\mathcal{F}_{t-1}).$$

Given $E(X_{t-h_1}X_{t-h_2}) = \sigma^2\delta_{h_1h_2}$ for $h_1, h_2 > 0$, we have

$$E(U_{nt}^2|\mathcal{F}_{t-1}) = 4\sigma^2\left(\sum_{h=1}^{q} X_{t-h}H_{t-q-1}(h)\right)^2$$

$$= 4\sigma^4\sum_{h=1}^{q} H_{t-q-1}^2(h) + 4\sigma^2\sum_{h_2=1}^{q}\sum_{h_1=1}^{q}(X_{t-h_1}X_{t-h_2} - \sigma^2\delta_{h_1h_2})$$

$$\times H_{t-q-1}(h_1)H_{t-q-1}(h_2)$$

$$= 4\sigma^4\sum_{h=1}^{q} H_{t-q-1}^2(h) + V_{1nt}, \quad \text{say.} \tag{A.17}$$

Next, using the definitions of $H_{t-q-1}(h)$ and $S_{t-q-1}(m)$ in (A.14) and (A.8), we decompose

$$4\sigma^4\sum_{h=1}^{q} H_{t-q-1}^2(h) = 4\sigma^4\sum_{h=1}^{q}\sum_{m_2=1}^{q}\sum_{m_1=1}^{q} b_J(h,m_1)b_J(h,m_2)S_{t-q-1}(m_1)S_{t-q-1}(m_2)$$

$$= 4\sigma^4\sum_{h=1}^{q}\sum_{m_2=1}^{q}\sum_{m_1=1}^{q} b_J(h,m_1)b_J(h,m_2)\sum_{s=1}^{t-q-1} X_s^2X_{s-m_1}X_{s-m_2}$$

$$+ 8\sigma^4\sum_{h=1}^{q}\sum_{m_2=1}^{q}\sum_{m_1=1}^{q} b_J(h,m_1)b_J(h,m_2)$$

$$\times\sum_{s_2=2}^{t-q-1}\sum_{s_1=1}^{s_2-1} X_{s_2}X_{s_2-m_2}X_{s_1}X_{s_1-m_1}$$

$$= W_{nt} + V_{2nt}, \quad \text{say.} \tag{A.18}$$

Finally, noting that $E(X_s X_{s-m_1} X_{s-m_2}) = \sigma^4 \delta_{m_1 m_2}$ for $m_1, m_2 > 0$, we obtain

$$
\begin{aligned}
W_{nt} &= 4(t-q-1)\sigma^8 \sum_{h=1}^{q} \sum_{m=1}^{q} b_J^2(h,m) \\
&\quad + 4\sigma^4 \sum_{h=1}^{q} \sum_{m_2=1}^{q} \sum_{m_1=1}^{q} b_J(h,m_1) b_J(h,m_2) \sum_{s=1}^{t-q-1} (X_s^2 X_{s-m_1} X_{s-m_2} - \sigma^4 \delta_{m_1 m_2}) \\
&= EU_{nt}^2 + V_{3nt}, \quad \text{say.}
\end{aligned} \tag{A.19}
$$

It follows from (A.17)–(A.19), $C_r$ inequality, Lemma A.2, which follows, and (A.15) that

$$
\begin{aligned}
\lambda_n^{-4} E(\tilde{U}_n^2 - \lambda_n^2)^2 &= \lambda_n^{-4} E\left( \sum_{j=1}^{3} n^{-2} \sum_{t=q+2}^{n} V_{jnt} \right)^2 \le 4\lambda_n^{-4} \sum_{j=1}^{3} E\left( n^{-2} \sum_{t=q+2}^{n} V_{jnt} \right)^2 \\
&= O\{q/n + (J+1)/2^J\} \to 0
\end{aligned}
$$

given $q^2/n \to 0, J \to \infty$. Hence, condition (ii) of Brown (1971) holds, and so $\lambda_n^{-1} \hat{U}_n \to^d N(0,1)$. The proof of Proposition 3 will be completed if Lemma A.2 is shown. ■

LEMMA A.2. *Suppose that the conditions of Proposition* 3 *hold. Then*

(i) $E(n^{-2} \sum_{t=q+2}^{n} V_{1nt})^2 = O(q2^{2J}/n)$;
(ii) $E(n^{-2} \sum_{t=q+2}^{n} V_{2nt})^2 = O\{q2^{2J}/n + (J+1)2^J\}$;
(iii) $E(n^{-2} \sum_{t=q+2}^{n} V_{3nt})^2 = O(2^{2J}/n)$.

**Proof of Lemma A.2.** (i) We first write

$$
\begin{aligned}
E\left( \sum_{t=q+2}^{n} V_{1nt} \right)^2 &\le 2 \sum_{t_2=q+2}^{n} \sum_{t_1=q+2}^{t_2} E(V_{1nt_2} V_{1nt_1}) \\
&= 2 \sum_{t_2=q+2}^{n} \sum_{t_1=\max(q+2, t_2-q)}^{t_2} E(V_{1nt_2} V_{1nt_1}) + 2 \sum_{t_2=2q+3}^{n} \sum_{t_1=q+2}^{t_2-q-1} E(V_{1nt_2} V_{1nt_1}).
\end{aligned} \tag{A.20}
$$

Because $\{X_{t_2-h}\}_{h=1}^{q}$ is independent of $\{H_{t_2-q-1}(h)\}_{h=1}^{q}$, we have $E(V_{1nt_2} | \mathcal{F}_{t_2-q-1}) = 0$. Moreover, $\{X_{t_2-h}\}_{h=1}^{q}$ is also independent of $V_{1nt_1}$ for $t_2 - t_1 > q$. Hence, we have $E(V_{1nt_2} V_{1nt_1}) = E[E(V_{1nt_2} | \mathcal{F}_{t_2-q-1}) V_{1nt_1}] = 0$ when $t_2 - t_1 > q$. Thus, the second term in (A.20) is zero.

We now compute the order of magnitude for the first term. Again, using the facts that $\{X_{t-h}\}_{h=1}^{q}$ is independent of $\{H_{t-q-1}(h)\}_{h=1}^{q}$ and $EV_{1nt}^2 = E[E(V_{1nt}^2 | \mathcal{F}_{t-q-1})]$, we obtain

$$
\begin{aligned}
EV_{1nt}^2 &= E\left[ 8\sigma^2 \sum_{h_2=1}^{q} \sum_{h_1=1}^{q} (X_{t-h_1} X_{t-h_2} - \sigma^2 \delta_{h_1 h_2}) H_{t-q-1}(h_1) H_{t-q-1}(h_2) \right]^2 \\
&\le C \sum_{h_2=1}^{q} \sum_{h_1=1}^{q} E\{H_{t-q-1}^2(h_1) H_{t-q-1}^2(h_2)\} \\
&\le Ct^2 \left( \sum_{h=1}^{q} \sum_{m=1}^{q} b_J^2(h,m) \right)^2 = O(t^2 2^{2J})
\end{aligned} \tag{A.21}
$$

given Lemma A.1(vi), where we made use of the fact that

$$E\{H_{t-q-1}^2(h_1)H_{t-q-1}^2(h_2)\} \le \{EH_{t-q-1}^4(h_1)EH_{t-q-1}^4(h_2)\}^{1/2}$$

$$\le Ct^2 \sum_{m=1}^{q} b_J^2(h_1,m) \sum_{m=1}^{q} b_J^2(h_2,m)$$

by Cauchy–Schwarz inequality and (A.16). Hence, we have $E(n^{-2}\sum_{t=q+2}^{n} V_{1nt})^2 = O(q2^{2J}/n)$ from (A.20) and (A.21) and Cauchy–Schwarz inequality.

(ii) We decompose $V_{2nt}$ into the sums with $s_2 - s_1 \le q$ and $s_2 - s_1 > q$:

$$V_{2nt} = 8\sigma^4 \sum_{h=1}^{q} \sum_{m_2=1}^{q} \sum_{m_1=1}^{q} b_J(h,m_1)b_J(h,m_2)\left(\sum_{s_2=2}^{t-q-1} \sum_{s_1=\max(1,s_2-q)}^{s_2-1} + \sum_{s_2=q+2}^{t-q-1} \sum_{s_1=1}^{s_2-q-1}\right)$$

$$\times X_{s_2}X_{s_2-m_2}X_{s_1}X_{s_1-m_1} = V_{21nt} + V_{22nt}, \quad \text{say.} \tag{A.22}$$

We first consider $V_{21nt}$. For any $m_1, m_2 > 0$, we have

$$E\left(\sum_{s_2=2}^{t-q-1} X_{s_2} \sum_{s_1=\max(1,s_2-q)}^{s_2-1} X_{s_2-m_2}X_{s_1}X_{s_1-m_1}\right)^2$$

$$= \sigma^2 \sum_{s_2=2}^{t-q-1} E\left(\sum_{s_1=\max(1,s_2-q)}^{s_2-1} X_{s_2-m_2}X_{s_1}X_{s_1-m_1}\right)^2$$

$$= \sigma^2 \sum_{s_2=2}^{t-q-1} \sum_{s_1=\max(1,s_2-q)}^{s_2-1} E(X_{s_2-m_2}X_{s_1}X_{s_1-m_1})^2 \le Ctq.$$

Hence, by Minkowski's inequality and Lemma A.1(iii),

$$E\left(n^{-2}\sum_{t=q+2}^{n} V_{21nt}\right)^2 \le Cqn^{-1}\left\{\sum_{h=1}^{q}\sum_{m_2=1}^{q}\sum_{m_1=1}^{q} |b_J(h,m_1)b_J(h,m_2)|\right\}^2 = O(q2^{2J}/n). \tag{A.23}$$

Next, we consider $V_{22nt}$. Put

$$Z_{s-1}(h) = \sum_{m=1}^{q} b_J(h,m)X_{s-m}.$$

Then we can write

$$V_{22nt} = 8\sigma^4 \sum_{h=1}^{q} \sum_{s_2=q+2}^{t-q-1} \sum_{s_1=1}^{s_2-q-1} X_{s_2}Z_{s_2-1}(h)X_{s_1}Z_{s_1-1}(h).$$

Because $X_s$ is independent of $Z_{s-1}(h)$ for $h > 0$, and because $\{X_{s_2}, Z_{s_2-1}(h)\}$ is independent of $\{X_{s_1}, Z_{s_1-1}(h)\}$ for $s_2 - s_1 > q$ and $0 < h \le q$, we have

$$\begin{aligned}
E(V_{22nt}^2) &= 64\sigma^{10} \sum_{s_2=q+2}^{t-q-1} E\left\{ \sum_{h=1}^{q} Z_{s_2-1}(h) \sum_{s_1=1}^{s_2-q-1} X_{s_1} Z_{s_1-1}(h) \right\}^2 \\
&= 64\sigma^{10} \sum_{s_2=q+2}^{t-q-1} \sum_{h_1=1}^{q} \sum_{h_2=1}^{q} E\{Z_{s_2-1}(h_1) Z_{s_2-1}(h_2)\} \\
&\quad \times E\left\{ \left( \sum_{s_1=1}^{s_2-q-1} X_{s_1} Z_{s_1-1}(h_1) \right) \left( \sum_{s_1'=1}^{s_2-q-1} X_{s_1'} Z_{s_1'-1}(h_2) \right) \right\} \\
&= 64\sigma^{12} \sum_{s_2=q+2}^{t-q-1} \sum_{s_1=1}^{s_2-q-1} \sum_{h_1=1}^{q} \sum_{h_2=1}^{q} E\{Z_{s_2-1}(h_1) Z_{s_2-1}(h_2)\} \\
&\quad \times E\{Z_{s_1-1}(h_1) Z_{s_1-1}(h_2)\} \\
&\le 32\sigma^{16} t^2 \sum_{h_1=1}^{q} \sum_{h_2=1}^{q} \left( \sum_{m=1}^{q} b_J(h_1, m) b_J(h_2, m) \right)^2,
\end{aligned}$$

where the last inequality follows from the fact that

$$E\{Z_{s-1}(h_1) Z_{s-1}(h_2)\} = \sigma^2 \sum_{m=1}^{q} b_J(h_1, m) b_J(h_2, m).$$

It follows from Minkowski's inequality and Lemma A.1(iv) that

$$E\left( n^{-2} \sum_{t=q+2}^{n} V_{22nt} \right)^2 \le \left( n^{-2} \sum_{t=q+2}^{n} (EV_{22nt}^2)^{1/2} \right)^2 = O\{(J+1)2^J\}. \tag{A.24}$$

Combining (A.22)–(A.24) yields $E(n^{-2} \sum_{t=q+2}^{n} V_{2nt})^2 = O\{q2^{2J}/n + (J+1)2^J\}$.

(iii) Write

$$\begin{aligned}
V_{3nt} &= 4\sigma^4 \sum_{h=1}^{q} \sum_{m_2=1}^{q} \sum_{m_1=1}^{q} b_J(h, m_1) b_J(h, m_2) \\
&\quad \times \left\{ \sum_{s=1}^{t-q-1} (X_s^2 - \sigma^2) X_{s-m_1} X_{s-m_2} \right. \\
&\quad \left. + \sum_{s=1}^{t-q-1} \sigma^2 (X_{s-m_1} X_{s-m_2} - \sigma^2 \delta_{m_1 m_2}) \right\} = V_{31nt} + V_{32nt}, \quad \text{say.}
\end{aligned} \tag{A.25}$$

By Minkowski's inequality, we have

$$\begin{aligned}
EV_{31nt}^2 &\le 16\sigma^8 \left\{ \sum_{h=1}^{q} \sum_{m_2=1}^{q} \sum_{m_1=1}^{q} |b_J(h, m_1) b_J(h, m_2)| \right. \\
&\quad \left. \times \left[ E\left( \sum_{s=1}^{t-q-1} (X_s^2 - \sigma^2) X_{s-m_1} X_{s-m_2} \right)^2 \right]^{1/2} \right\}^2 \\
&\le Ct \left\{ \sum_{h=1}^{q} \left( \sum_{m=1}^{q} |b_J(h, m)| \right)^2 \right\}^2 = O(t2^{2J})
\end{aligned} \tag{A.26}$$

by Lemma A.1(iii). Similarly, we have

$$EV_{32nt}^2 \le Ct\left\{\sum_{h=1}^{q}\left(\sum_{m=1}^{q}|b_J(h,m)|\right)^2\right\}^2 = O(t2^{2J}). \tag{A.27}$$

It follows from (A.25)–(A.27) and Minkowski's inequality that $E(n^{-2}\sum_{t=q+2}^{n} V_{3nt})^2 = O(2^{2J}/n)$. ■

**Proof of Theorem 2.** Given $2^{3J/2}/n \to 0$, we have $\{2(2^{J+1}-1)^{1/2}/n\}W_n = 2\pi Q(\hat{f};f_0) + o(1)$. To show $Q(\hat{f};f_o) \to^p Q(f;f_0)$, we decompose

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

It suffices to show $Q(\hat{f};f) \to^p 0$, because the last term is $o_P(1)$ by Cauchy–Schwarz inequality.

Now, define the pseudo estimator

$$\tilde{f}(\omega) = (2\pi)^{-1} + \sum_{j=0}^{J}\sum_{k=1}^{2^j}\tilde{\alpha}_{jk}\Psi_{jk}(\omega), \tag{A.28}$$

where $\tilde{\alpha}_{jk} = \sum_{h=1-n}^{n-1}\tilde{\rho}(h)\hat{\psi}_{jk}(2\pi h)$ and $\tilde{\rho}(h) = \hat{R}(h)/\sigma^2$. Also, put

$$f_J(\omega) = (2\pi)^{-1} + \sum_{j=0}^{J}\sum_{k=1}^{2^j}\alpha_{jk}\Psi_{jk}(\omega). \tag{A.29}$$

Noting that $\hat{f}(\omega) - f(\omega) = \hat{f}(\omega) - \tilde{f}(\omega) + \tilde{f}(\omega) - E\tilde{f}(\omega) + E\tilde{f}(\omega) - f_J(\omega) + f_J(\omega) - f(\omega)$, we have

$$Q(\hat{f};f) \le 8Q(\hat{f};\tilde{f}) + 8Q(\tilde{f};E\tilde{f}) + 8Q(E\tilde{f};f_J) + 8Q(f_J;f).$$

We shall show that (i) $Q(\hat{f};\tilde{f}) \to^p 0$; (ii) $Q(\tilde{f};E\tilde{f}) \to^p 0$; (iii) $Q(E\tilde{f};f_J) \to 0$; (iv) $Q(f_J;f) \to 0$.

(i) By the orthonormality (28) and noting that $\hat{\alpha}_{jk} = \{\sigma^2/\hat{R}(0)\}\tilde{\alpha}_{jk}$, we have

$$Q(\hat{f};\tilde{f}) = \sum_{j=0}^{J}\sum_{k=1}^{2^j}(\hat{\alpha}_{jk}-\tilde{\alpha}_{jk})^2 = \{\sigma^2/\hat{R}(0)-1\}^2\sum_{j=0}^{J}\sum_{k=1}^{2^j}\tilde{\alpha}_{jk}^2 \to^p 0$$

given $\hat{R}(0) - \sigma^2 \to^p 0$ and

$$\sum_{j=0}^{J}\sum_{k=1}^{2^j}\tilde{\alpha}_{jk}^2 \le 2Q(\tilde{f};f_J) + 2\sum_{j=0}^{\infty}\sum_{k=1}^{2^j}\alpha_{jk}^2 = O_P(1),$$

where $Q(\tilde{f};f_J) \le 4Q(\tilde{f};E\tilde{f}) + 4Q(E\tilde{f};f_J) \to^p 0$ by (ii) and (iii) as shown subsequently.

(ii) By the orthonormality (28) and noting that $\tilde{\alpha}_{jk}$ is real valued, we have

$$Q(\tilde{f};E\tilde{f}) = \sum_{j=0}^{J}\sum_{k=1}^{2^j}(\tilde{\alpha}_{jk} - E\tilde{\alpha}_{jk})^2 = \sigma^{-4}\sum_{j=0}^{J}\sum_{k=1}^{2^j}\left|\sum_{h=1-n}^{n-1}[\hat{R}(h) - E\hat{R}(h)]\hat{\psi}_{jk}(2\pi h)\right|^2.$$

Recalling the definitions of $a_J(h,m)$ and $b_J(h,m)$ in Lemma A.1, we obtain

$$0 \le EQ(\tilde{f};E\tilde{f}) = \sigma^{-4}\sum_{h=1-n}^{n-1}\sum_{m=1-n}^{n-1}|a_J(h,m)|\mathrm{Cov}\{\hat{R}(h),\hat{R}(m)\}$$

$$= \sigma^{-4}\sum_{h=1}^{n-1}\sum_{m=1}^{n-1}|b_J(h,m)|\mathrm{Cov}\{\hat{R}(h),\hat{R}(m)\}$$

$$\le \sup_{0<h<n}\mathrm{Var}\{\hat{R}(h)\}\sum_{h=1}^{n-1}\sum_{m=1}^{n-1}|b_J(h,m)| = O(2^J/n), \tag{A.30}$$

where the last inequality follows from Cauchy–Schwarz inequality and the last equality follows from Lemma A.1(ii) and $\sup_{0<h<n}\mathrm{Var}\{\hat{R}(h)\} = O(n^{-1})$, which follows from $\sum_{l=1}^{\infty}R^2(l) < \infty, \sum_{j=-\infty}^{\infty}\sum_{k=-\infty}^{\infty}\sum_{l=-\infty}^{\infty}|\kappa(j,k,l)| < \infty$, and

$$\mathrm{Var}\{\hat{R}(h)\} = n^{-1}\sum_{l=1-n}^{n-1}(1-|l|/n)\{R^2(l) + R(l-h)R(l+h) + \kappa(h,l,l+h)\}$$

(cf. Hannan, 1970, p. 209). Hence, $Q(\tilde{f};E\tilde{f}) \to^p 0$ by (A.30) and Markov's inequality.

(iii) We now show $Q(E\tilde{f};f_J) \to 0$. From the definition of $\tilde{\alpha}_{jk}$ in (A.28), we have

$$E\tilde{\alpha}_{jk} - \alpha_{jk} = \sigma^{-2}\sum_{h=1-n}^{n-1}(1-|h|/n)R(h)\hat{\psi}_{jk}(2\pi h) - \sigma^{-2}\sum_{h=-\infty}^{\infty}R(h)\hat{\psi}_{jk}(2\pi h)$$

$$= -\sigma^{-2}n^{-1}\sum_{h=1-n}^{n-1}|h|R(h)\hat{\psi}_{jk}(2\pi h) - \sigma^{-2}\sum_{|h|\ge n}R(h)\hat{\psi}_{jk}(2\pi h).$$

It follows that

$$Q(E\tilde{f};f_J) = \sum_{j=0}^{J}\sum_{k=1}^{2^j}(E\tilde{\alpha}_{jk} - \alpha_{jk})^2$$

$$\le 2\sigma^{-4}n^{-2}\sum_{j=0}^{J}\sum_{k=1}^{2^j}\left\{\sum_{h=1-n}^{n-1}|h|R(h)\hat{\psi}_{jk}(2\pi h)\right\}^2$$

$$+ 2\sigma^{-4}\sum_{j=0}^{J}\sum_{k=1}^{2^j}\left\{\sum_{|h|\ge n}R(h)\hat{\psi}_{jk}(2\pi h)\right\}^2$$

$$= 2\sigma^{-4}M_{1n} + 2\sigma^{-4}M_{2n}, \quad \text{say.} \tag{A.31}$$

For the first term $M_{1n}$, by Cauchy–Schwarz inequality and (16), we have

$$M_{1n} \le n^{-2}\left\{\sum_{h=1-n}^{n-1} R^2(h)\right\}\left\{\sum_{j=0}^{J}\sum_{k=1}^{2^j}\sum_{h=1-n}^{n-1} h^2|\hat{\psi}_{jk}(2\pi h)|^2\right\}$$

$$\le n^{-2}\left\{\sum_{h=1-n}^{n-1} R^2(h)\right\}\sum_{j=0}^{J} 2^{3j}\left\{(2\pi/2^j)\sum_{h=1-n}^{n-1}(2\pi h/2^j)^2|\hat{\psi}(2\pi h/2^j)|^2\right\}$$

$$= O(2^{3J}/n^2) \tag{A.32}$$

given $\sum_{h=-\infty}^{\infty} R^2(h) < \infty$ and the fact that given Assumption 2(i),

$$(2\pi/2^j)\sum_{h=1-n}^{n-1}(2\pi h/2^j)^2|\hat{\psi}(2\pi h/2^j)|^2 \le C\int_{-\infty}^{\infty} z^2/(1+|z|)^{2\alpha}dz < \infty, \quad \alpha > \frac{3}{2}.$$

For the second term $M_{2n}$ in (A.31), we have, by Cauchy–Schwarz inequality and (16),

$$\sum_{j=0}^{J}\sum_{k=1}^{2^j}\left\{\sum_{|h|\ge n} R(h)\hat{\psi}_{jk}(2\pi h)\right\}^2$$

$$\le \left\{\sum_{h\ge n} R^2(h)\right\}\sum_{j=0}^{J}\sum_{h\ge n}|\hat{\psi}(2\pi h/2^j)|^2 = o(2^{2\alpha J}/n^{2\alpha-1}), \tag{A.33}$$

where the last equality follows from $\sum_{h\ge n} R^2(h) \to 0$ and

$$\sum_{j=0}^{J}\sum_{|h|\ge n}|\hat{\psi}(2\pi h/2^j)|^2 \le C^2\sum_{j=0}^{J}\sum_{|h|\ge n}|2\pi h/2^j|^{-2\alpha} = O(2^{2\alpha J}/n^{2\alpha-1}) \tag{A.34}$$

by Assumption 2(i). Combining (A.31)–(A.34), $2^{3J/2}/n \to 0$, and $\alpha > \frac{3}{2}$ yields $Q(E\hat{f};f_J) \to 0$.

(iv) By the orthonormality (28) and $\sum_{j=0}^{\infty}\sum_{k=1}^{2^j}\alpha_{jk}^2 = \int_{-\pi}^{\pi} f^2(\omega)d\omega = \sum_{l=-\infty}^{\infty}\rho^2(l) < \infty$, we have $Q(f_J;f) = \sum_{j=J+1}^{\infty}\sum_{k=1}^{2^j}\alpha_{jk}^2 \to 0$ as $J \to \infty$. This completes the proof. ■

# APPENDIX B

**Proof of Lemma A.1.** (i) Given (16), $\hat{\psi}(0) = 0$, and $\hat{\psi}^*(-z) = \hat{\psi}(-z)$, we have $a_J(0,m) = a_J(h,0) = 0, a_J^*(h,m) = a_J(m,h) = a_J(-h,-m)$. Hence, $b_J(0,m) = b_J(h,0) = 0$, and

$$b_J(h,m) = a_J(h,m) + a_J^*(h,m) + a_J(h,-m) + a_J^*(h,-m)$$

$$= a_J(h,m) + a_J(m,h) + a_J(h,-m) + a_J(m,-h). \tag{B.1}$$

The first equality in (B.1) implies that $b_J(h,m)$ is real valued, and the second one implies that $b_J(h,m) = b_J(m,h)$.

(ii) Put $c_j(h,m) = 2^{-j}\sum_{k=1}^{2^j} e^{i2\pi(m-h)k/2^j}$. Then, using (16), we obtain

$$a_J(h,m) = 2\pi \sum_{j=0}^{J} c_j(h,m)\hat{\psi}(2\pi h/2^j)\hat{\psi}^*(2\pi m/2^j)$$
$$= \begin{cases} 2\pi \sum_{j=0}^{J} \hat{\psi}(2\pi h/2^j)\hat{\psi}^*(2\pi m/2^j), & \text{if } m-h = 2^j r \quad \text{for some } r \in \mathbb{Z}, \\ 0, & \text{otherwise,} \end{cases} \tag{B.2}$$

where we used the well-known identity that $c_j(h,m) = 1$ if $m-h = 2^j r$ for some $r \in \mathbb{Z}$ and $c_j(h,m) = 0$ otherwise (cf. Priestley, 1981, (6.19), p. 392).

Now, by the triangle inequality, reindexing, and (B.2), we obtain

$$\sum_{h=1}^{n-1}\sum_{m=1}^{n-1} h^\nu |b_J(h,m)| \le \sum_{h=-\infty}^{\infty}\sum_{m=-\infty}^{\infty} |h|^\nu |a_J(h,m)|$$
$$\le 2\pi \sum_{j=0}^{J}\sum_{h=-\infty}^{\infty}\sum_{r=-\infty}^{\infty} |h|^\nu |\hat{\psi}(2\pi h/2^j)\hat{\psi}^*(2\pi h/2^j + 2\pi r)|$$
$$\le \sum_{j=0}^{J} 2^{j(1+\nu)} \left( (2\pi/2^j) \sum_{h=-\infty}^{\infty} |2\pi h/2^j|^\nu |\hat{\psi}(2\pi h/2^j)| \right)$$
$$\times \left( \sum_{r=-\infty}^{\infty} |\hat{\psi}(2\pi h/2^j + 2\pi r)| \right) = O(2^{(1+\nu)J}),$$

where we used the facts that given Assumption 2(i) and $\nu \le \frac{1}{2}$,

$$(2\pi/2^j) \sum_{h=-\infty}^{\infty} |2\pi h/2^j|^\nu |\hat{\psi}(2\pi h/2^j)| \le C \int_{-\infty}^{\infty} |z|^\nu/(1+|z|)^\alpha dz < \infty, \tag{B.3}$$

$$\sup_{z\in\mathbb{R}} \sum_{r=-\infty}^{\infty} |\hat{\psi}(z + 2\pi r)| \le C. \tag{B.4}$$

(iii) By reindexing and (B.2), we have

$$\sum_{h=1}^{n-1}\left(\sum_{m=1}^{n-1} |b_J(h,m)|\right)^2 \le C \sum_{h=-\infty}^{\infty}\left(\sum_{m=-\infty}^{\infty} |a_J(h,m)|\right)^2$$
$$\le C \sum_{h=-\infty}^{\infty}\left(2\pi \sum_{j=0}^{J} |\hat{\psi}(2\pi h/2^j)| \sum_{r=-\infty}^{\infty} |\hat{\psi}(2\pi h/2^j + 2\pi r)|\right)^2$$
$$\le C^2 \sum_{h=-\infty}^{\infty}\left(2\pi \sum_{j=0}^{J} |\hat{\psi}(2\pi h/2^j)|\right)^2 \quad \text{given (B.4)}$$
$$\le 2\pi C^2 \left\{ \sum_{j=0}^{J} 2^{j/2} \left[ (2\pi/2^j) \sum_{h=-\infty}^{\infty} |\hat{\psi}(2\pi h/2^j)|^2 \right]^{1/2} \right\}^2$$
$$= O(2^J),$$

where we used the orthonormality (17) in obtaining the last equality.

(iv) We first show $\sup_{h\in\mathbb{Z}} \sum_{m=1}^{n-1} |b_J(h,m)|^2 \leq C(J+1)$. By reindexing and (B.2),

$$\sum_{m=-\infty}^{\infty} |a_J(h,m)|^2 = (2\pi)^2 \sum_{j=0}^{J} \sum_{m=-\infty}^{\infty} |c_j(h,m)|^2 |\hat{\psi}(2\pi h/2^j)|^2 |\hat{\psi}(2\pi m/2^j)|^2$$

$$+ 2(2\pi)^2 \operatorname{Re} \sum_{d=1}^{J} \sum_{j=d}^{J} \sum_{m=-\infty}^{\infty} c_j(h,m) c_{j-d}^*(h,m)$$

$$\times \hat{\psi}(2\pi h/2^j) \hat{\psi}^*(2^d 2\pi h/2^j) \hat{\psi}^*(2\pi m/2^j) \hat{\psi}(2^d 2\pi m/2^j)$$

$$= 2\pi \sum_{j=0}^{J} |\hat{\psi}(2\pi h/2^j)|^2 \left(2\pi \sum_{r=-\infty}^{\infty} |\hat{\psi}(2\pi h/2^j + 2\pi r)|^2\right)$$

$$+ 4\pi \operatorname{Re} \sum_{d=1}^{J} \sum_{j=d}^{J} \hat{\psi}(2\pi h/2^j) \hat{\psi}^*(2^d 2\pi h/2^j)$$

$$\times \left(2\pi \sum_{r=-\infty}^{\infty} \hat{\psi}^*(2\pi h/2^j + 2\pi r) \hat{\psi}(2^d(2\pi h/2^j + 2\pi r))\right)$$

$$= \sum_{j=0}^{J} |\hat{\psi}(2\pi h/2^j)|^2, \tag{B.5}$$

where the last equality follows from the orthonormality that for any $d \geq 0$,

$$2\pi \sum_{r=-\infty}^{\infty} \hat{\psi}(z + 2\pi r) \hat{\psi}^*(2^d(z + 2\pi r)) = \delta_{0d} \quad \text{a.e. } z \in \mathbb{R} \tag{B.6}$$

(cf. Hernández and Weiss, 1996, (1.2) and its proof, pp. 101–102). It follows from Cauchy–Schwarz inequality, $b_J(h,m) = b_J(m,h)$, Lemma A.1(iii), (B.5), and $|\hat{\psi}(z)| \leq C$ that

$$\sum_{h_1=1}^{n-1} \sum_{h_2=1}^{n-1} \left( \sum_{m=1}^{n-1} |b_J(h_1,m) b_J(h_2,m)| \right)^2$$

$$\leq \sup_{h\in\mathbb{Z}} \sum_{m=-\infty}^{\infty} |b_J(h,m)|^2$$

$$\times \sum_{h_1=1}^{n-1} \sum_{h_2=1}^{n-1} \left( \sum_{m=1}^{n-1} |b_J(h_1,m) b_J(h_2,m)| \right)$$

$$= \sup_{h\in\mathbb{Z}} \sum_{m=-\infty}^{\infty} |b_J(h,m)|^2 \left( \sum_{h=1}^{n-1} \left( \sum_{m=1}^{n-1} |b_J(h,m)| \right)^2 \right)$$

$$= O\{(J+1)2^J\}.$$

(v) By $a_J(0,h) = a_J(h,0) = 0$ and reindexing, we have

$$\sum_{h=1}^{n-1} b_J(h,h) = \sum_{h=1-n}^{n-1} a_J(h,h) + \sum_{h=1-n}^{n-1} a_J(h,-h). \tag{B.7}$$

Using (B.2) and the orthonormality (17), we have

$$\sum_{h=1-n}^{n-1} a_J(h,h) = \sum_{j=0}^{J} 2^j \left\{ (2\pi/2^j) \sum_{h=1-n}^{n-1} |\hat{\psi}(2\pi h/2^j)|^2 \right\}$$

$$= \sum_{j=0}^{J} 2^j \left\{ 1 - (2\pi/2^j) \sum_{|h|\geq n} |\hat{\psi}(2\pi h/2^j)|^2 \right\}$$

$$= (2^{J+1}-1)\{1 + O[(2^J/n)^{2\alpha-1}]\}, \tag{B.8}$$

where we used $\sum_{j=0}^{J} 2^j = 2^{J+1} - 1$ and (A.34). For the second term in (B.7), using (B.2) and Assumptions 1 and 2(i), we obtain

$$\left| \sum_{h=1-n}^{n-1} a_J(h,-h) \right| \leq \sum_{h=-\infty}^{\infty} |a_J(h,-h)| \leq \sum_{j=0}^{J} \sum_{r=-\infty}^{\infty} |\hat{\psi}(\pi r)|^2 \leq C(J+1). \tag{B.9}$$

Combining (B.7)–(B.9), $2^{3J/2}/n \to 0, J \to \infty$, and $\alpha > \frac{3}{2}$ then yields $\sum_{h=1}^{n-1} b_J(h,h) = (2^{J+1}-1)\{1 + O[(J+1)/2^J + (2^J/n)^{2\alpha-1}]\}$.

(vi) Using $a(h,0) = a(0,m) = 0$ and reindexing, we have

$$\sum_{h=1}^{n-1}\sum_{m=1}^{n-1} b_J^2(h,m) = \sum_{h=1}^{n-1}\sum_{m=1}^{n-1} [a_J(h,m) + a_J(-h,-m) + a_J(h,-m) + a_J(-h,m)]^2$$

$$= \sum_{h=1}^{n-1}\sum_{m=1}^{n-1} [a_J^2(h,m) + a_J^2(-h,-m) + a_J^2(h,-m) + a_J^2(-h,m)]$$

$$+ 2\sum_{h=1}^{n-1}\sum_{m=1}^{n-1} [a_J(h,m)a_J(-h,-m) + a_J(h,-m)a_J(-h,m)]$$

$$+ 2\sum_{h=1}^{n-1}\sum_{m=1}^{n-1} [a_J(h,m)a_J(h,-m) + a_J(-h,-m)a_J(-h,m)]$$

$$+ 2\sum_{h=1}^{n-1}\sum_{m=1}^{n-1} [a_J(h,m)a_J(-h,m) + a_J(-h,-m)a_J(h,-m)]$$

$$= \sum_{h=1-n}^{n-1}\sum_{m=1-n}^{n-1} a_J^2(h,m) + \sum_{h=1-n}^{n-1}\sum_{m=1-n}^{n-1} |a_J(h,m)|^2$$

$$+ \sum_{h=1-n}^{n-1}\sum_{m=1-n}^{n-1} a_J(h,m)a_J(h,-m) + \sum_{h=1-n}^{n-1}\sum_{m=1-n}^{n-1} a_J(h,m)a_J(-h,m)$$

$$= A_{1n} + A_{2n} + A_{3n} + A_{4n}, \quad \text{say}. \tag{B.10}$$

We first consider $A_{1n}$. Write

$$A_{1n} = \sum_{h=1-n}^{n-1}\sum_{m=-\infty}^{\infty} a_J^2(h,m) - \sum_{h=1-n}^{n-1}\sum_{|m|\geq n} a_J^2(h,m) = A_{11n} - A_{12n}, \quad \text{say}. \tag{B.11}$$

Following a reasoning analogous to that of (B.5), we have

$$A_{11n} = (2\pi)^2 \sum_{d=-J}^{J} \sum_{j=|d|}^{J} \sum_{h=1-n}^{n-1} \sum_{m=-\infty}^{\infty} c_j(h,m) c_{j-|d|}(h,m) \hat{\psi}(2\pi h/2^j) \hat{\psi}(2^{|d|} 2\pi h/2^j)$$

$$\times \hat{\psi}^*(2\pi m/2^j) \hat{\psi}^*(2^{|d|} 2\pi m/2^j)$$

$$= (2\pi)^2 \sum_{d=-J}^{J} \sum_{j=|d|}^{J} \sum_{h=1-n}^{n-1} \sum_{r=-\infty}^{\infty} \hat{\psi}(2\pi h/2^j) \hat{\psi}(2^{|d|} 2\pi h/2^j)$$

$$\times \hat{\psi}^*(2\pi h/2^j + 2\pi r) \hat{\psi}^*(2^{|d|}(2\pi h/2^j + 2\pi r)) \qquad \text{by (B.2)}$$

$$= 2\pi \sum_{d=-J}^{J} \sum_{j=|d|}^{J} \sum_{h=1-n}^{n-1} \hat{\psi}^*(2\pi h/2^j) \hat{\psi}(2^{|d|} 2\pi h/2^j)$$

$$\times 2\pi \sum_{r=-\infty}^{\infty} \hat{\psi}(2\pi h/2^j + 2\pi r) \hat{\psi}^*(2^{|d|}(2\pi h/2^j + 2\pi r))$$

by (B.13), which follows

$$= 2\pi \sum_{j=0}^{J} \sum_{h=1-n}^{n-1} |\hat{\psi}(2\pi h/2^j)|^2 \qquad \text{by (B.6)}$$

$$= (2^{J+1} - 1)\{1 + o(1)\}, \tag{B.12}$$

where the third equality follows because for any $z \in \mathbb{R}$, and any $d, r \in \mathbb{Z}$,

$$\hat{\psi}(z) \hat{\psi}(2^{|d|} z) \hat{\psi}^*(z + 2\pi r) \hat{\psi}^*(2^{|d|}(z + 2\pi r))$$

$$= \hat{\psi}^*(z) \hat{\psi}(2^{|d|} z) \hat{\psi}(z + 2\pi r) \hat{\psi}^*(2^{|d|}(z + 2\pi r)), \tag{B.13}$$

given Assumption 2(ii). Also, the last equality follows from (B.8) and $2^{3J/2}/n \to 0$.

Next, using $a_J(h,m) = a_J(m,h)^*$ and (B.5), we obtain

$$A_{12n} \le \sum_{h=-\infty}^{\infty} \sum_{|m| \ge n} |a_J(h,m)|^2 = \sum_{h=-\infty}^{\infty} \sum_{|m| \ge n} |a_J(m,h)|^2$$

$$= 2\pi \sum_{j=0}^{J} \sum_{|m| \ge n} |\hat{\psi}(2\pi m/2^j)|^2 = o(2^J), \tag{B.14}$$

where the last equality follows from (A.34) and $2^{3J/2}/n \to 0$. Consequently, from (B.11), (B.12), and (B.14), we obtain

$$A_{1n} = (2^{J+1} - 1)\{1 + o(1)\}. \tag{B.15}$$

By reasoning similar to that of $A_{1n}$, we can also show

$$A_{2n} = (2^{J+1} - 1)\{1 + o(1)\}. \tag{B.16}$$

Now, we consider $A_{3n}$. By reindexing, (B.2), and $\hat{\psi}^*(z) = \hat{\psi}(-z)$, we can write

$$A_{3n} = (2\pi)^2 \sum_{d=-J}^{J} \sum_{j=|d|}^{J} \sum_{h=1-n}^{n-1} \sum_{m=n-1}^{n-1} c_j(h,m)c_{j-|d|}(h,-m)\hat{\psi}(2\pi h/2^j)\hat{\psi}(2^{|d|}2\pi h/2^j)$$
$$\times \hat{\psi}^*(2\pi m/2^j)\hat{\psi}^*(-2^{|d|}2\pi m/2^j).$$

Given (B.2), we have $c_j(h,m)c_{j-|d|}(h,-m) = 1$ if $m-h = 2^j r$ and $m+h = 2^{j-|d|}r'$ for some $r, r' \in \mathbb{Z}$, and $c_j(h,m)c_{j-|d|}(h,-m) = 0$ otherwise. It follows that

$$|A_{3n}| \leq (2\pi)^2 \sum_{d=-J}^{J} \sum_{j=|d|}^{J} \sum_{r'=-\infty}^{\infty} \sum_{r=-\infty}^{\infty} |\hat{\psi}(\pi r'/2^{|d|} - \pi r)\hat{\psi}(\pi r' - 2^{|d|}\pi r)|$$
$$\times |\hat{\psi}(\pi r'/2^{|d|} + \pi r)\hat{\psi}(\pi r' + 2^{|d|}\pi r)|$$
$$\leq C \sum_{d=-J}^{J} \sum_{j=|d|}^{J} \left\{ \sum_{l=-\infty}^{\infty} |\hat{\psi}(\pi l)| \right\} \left\{ \sum_{r=-\infty}^{\infty} |\hat{\psi}(2\pi r)| \right\} = O\{(J+1)^2\}, \quad \textbf{(B.17)}$$

where the second inequality follows by change of variable $l = r' - 2^{|d|}r$ and $|\hat{\psi}(z)| \leq C$. Similarly, we have $A_{4n} = O\{(J+1)^2\}$. Combining this with (B.10) and (B.15)–(B.17), we obtain $\sum_{h=1}^{n-1}\sum_{m=1}^{n-1} b_J^2(h,m) = 2(2^{J+1}-1)\{1+o(1)\}$. This completes the proof. ■