One-Sided Testing for ARCH Effects Using Wavelets

Author(s): Yongmiao Hong and Jin Lee
Source: *Econometric Theory*, Vol. 17, No. 6 (Dec., 2001), pp. 1051-1081
Published by: Cambridge University Press
Stable URL: http://www.jstor.org/stable/3533244
Accessed: 22/11/2013 14:11

Econometric Theory, **17**, 2001, 1051–1081. Printed in the United States of America.

ONE-SIDED TESTING FOR ARCH EFFECTS USING WAVELETS

YONGMIAO HONG
Cornell University

JIN LEE
National University of Singapore

There has been increasing interest recently in hypothesis testing with inequality restrictions. An important example in time series econometrics is hypotheses on autoregressive conditional heteroskedasticity (ARCH). We propose a one-sided test for ARCH effects using a wavelet spectral density estimator at frequency zero of a squared regression residual series. The square of an ARCH process is positively correlated at all lags, resulting in a spectral mode at frequency zero. In particular, it has a spectral peak at frequency zero when ARCH effects are persistent or when ARCH effects are small at each individual lag but carry over a long distributional lag. As a joint time-frequency decomposition method, wavelets can effectively capture spectral peaks. We expect that wavelets are more powerful than kernels in small samples when ARCH effects are persistent or when ARCH effects have a long distributional lag. This is confirmed in a simulation study.

1. INTRODUCTION

Hypothesis testing with inequality restrictions is important in econometrics and statistics (e.g., Andrews 1998, 2001; Bera, Ra, and Sakar, 1998; Gourieroux, Holly, and Monfort, 1982; King and Wu, 1997; SenGupta and Vermeire, 1986; Silvapulle and Silvapulle, 1995; Wolak, 1989). An important example in time series econometrics is hypotheses on autoregressive conditional heteroskedasticity (ARCH). Here, parameters of interest are zero if there is no ARCH effect and are nonnegative if ARCH effects exist.

Detecting ARCH effects is important from both theoretical and practical points of view. Neglecting ARCH effects may lead to loss in asymptotic efficiency of parameter estimation (Engle, 1982); cause overrejection of conventional tests for serial correlation (Diebold, 1987; Milhøj, 1985); and result in overparameterization of ARMA models (Weiss, 1984). Although the one-sided nature of ARCH has been long known, most ARCH tests are two-sided. Among them are Bera and Higgins (1992), Engle (1982), Gregory (1989), Hong and Shehadeh (1999), McLeod and Li (1983), Robinson (1991), and Weiss (1984).

We thank two referees for insightful comments and suggestions. All remaining errors are solely ours. Address correspondence to: Yongmiao Hong, Department of Economics and Department of Statistical Science, Cornell University, 492 Uris Hall, Ithaca, NY 14853-7601, USA; e-mail: yh20@cornell.edu.

 0266-4666/01 $9.50

Exploiting the one-sided nature of ARCH is expected to increase power in small samples. Engle, Hendry, and Trumble (1985) suggest using the square root of a Lagrangian multiplier (LM) test statistic to test ARCH(1). This approach, however, could not be generalized to test ARCH(q) for $q > 1$. Lee and King (1993, 1994) are the first to develop one-sided tests for ARCH(q). They propose a locally most mean powerful score-based test for ARCH(q). Demos and Sentana (1998) consider a convenient one-sided LM test for ARCH(q) in a spirit similar to Kuhn–Tucker multiplier tests (Gourieroux et al., 1982). Andrews (2001), Lee and King (1993), and Demos and Sentana (1998) consider one-sided tests for GARCH(1,1). Simulation studies show that these tests outperform two-sided tests (e.g., Engle, 1982) in finite samples, indicating nontrivial gains of exploiting the one-sided nature of ARCH.

Hong (1997) recently proposed a one-sided ARCH test by observing that the spectral density of a squared regression residual series is uniform when there is no ARCH effect and is larger than the uniform density at frequency zero whenever ARCH effects exist. Hong (1997) uses a kernel method. The test is shown to perform reasonably well in comparison with some popular one-sided and two-sided ARCH tests.

It is well known that in finite samples kernels tend to underestimate the spectral density wherever there is a mode, even if a finite sample optimal bandwidth is used (Priestley, 1981). Kernels are not an ideal tool in capturing nonsmooth spectral features. In the present context, the one-sided nature of ARCH implies that the square of a linear ARCH process is positively correlated at all lags, resulting in a spectral mode at frequency zero. In particular, the squared series has a spectral peak at frequency zero when ARCH effects are persistent or when ARCH effects are small at each individual lag but carry over a long distributional lag. Examples are nearly integrated generalized autoregressive conditional heteroskedasticity (GARCH) and fractionally integrated GARCH processes (Baillie, Bollerslev, and Mikkelsen, 1996). In these situations, kernels cannot be expected to perform well in small and finite samples.

The recent development of wavelet analysis provides a new approach to constructing a potentially more powerful one-sided test for ARCH effects. Wavelets are a new mathematical tool developed over the last decade. As a joint time-frequency decomposition method, wavelets can effectively capture significantly spatially inhomogeneous features (e.g., Donoho and Johnstone, 1994, 1995a, 1995b; Donoho, Johnstone, Kerkyacharian, and Picard, 1996; Gao, 1993; Neumann, 1996; Wang, 1995). Here, we propose a one-sided test for ARCH effects using a wavelet spectral density estimator. Wavelets are expected to be more powerful than kernels in small samples when there exist persistent ARCH effects. Besides ARCH, spectral peaks may arise as a result of strong dependence, seasonality, and business cycles. Therefore, our approach might have potential applications to testing a broad range of other hypotheses in econometrics. This paper merely provides an example to illustrate how wavelets can be used to develop powerful econometric procedures.

In time series analysis, Gao (1993) uses a Meyer wavelet to estimate the spectral density of a stationary Gaussian series. Neumann (1996) considers wavelet estimation of the spectral density of a stationary non-Gaussian series. Priestley (1996) explores potential applications of wavelets to nonstationary time series (see also Subba Rao and Indukumar, 1996). Jensen (2000) uses wavelets to estimate a long memory model via maximum likelihood. There are also some applications of wavelet analysis to economic and financial time series (e.g., Goffe, 1994; Ramsey and Lampart, 1998a, 1998b; Ramsey, Usikov, and Zaslavsky, 1995).

We describe hypotheses of interest in Section 2. Section 3 introduces wavelet analysis and its application to spectral analysis. In Section 4, we propose our test statistic and derive its asymptotic distribution. Asymptotic local power is examined in Section 5. Section 6 presents a Monte Carlo study on the proposed test, three existing one-sided tests, and Engle's (1982) LM test. Section 7 concludes. All proofs are in the Appendix. Unless indicated, all limits are taken as the sample size $n \to \infty$; A^* denotes the complex conjugate of A; $\|A\| = [\mathrm{tr}(A'A)]^{1/2}$ the Euclidean norm of A; C a bounded constant that may differ from place to place; and $\mathbb{Z} = \{0, \pm 1, \dots\}$ the set of integers.

2. HYPOTHESES

Throughout, we consider the following data generating process.

Assumption A.1. $\{Y_t\}$ is a stochastic time series process

$$Y_t = g(X_t, b_0) + \varepsilon_t, \qquad \varepsilon_t = \xi_t h_t^{1/2}, \tag{2.1}$$

where $g(\cdot,\cdot)$ is a known functional form, b_0 is an unknown finite-dimensional parameter vector, X_t is a vector consisting of exogenous and lagged dependent variables, and h_t is a positive time-varying measurable function of $\mathcal{I}_{t-1}$, the information set available at period $t-1$. The innovation sequence $\{\xi_t\}$ is independent and identically distributed (i.i.d.) with $E(\xi_t) = 0$, $E(\xi_t^2) = 1$, and $E(\xi_t^8) < \infty$. Moreover, ξ_t is independent of X_s for all $s \leq t$.

This framework is often assumed in the ARCH literature (e.g., Bollerslev, Chou, and Kroner, 1992). We make no distributional assumption on innovation ξ_t except the existence of its eighth moment. Because $E(\varepsilon_t|\mathcal{I}_{t-1}) = 0$ almost surely, $\{\varepsilon_t, \mathcal{I}_{t-1}\}$ is an adapted martingale difference sequence with respect to $\mathcal{I}_{t-1}$. Its conditional variance, $E(\varepsilon_t^2|\mathcal{I}_{t-1}) = h_t$, is time-varying. Throughout, we consider a generalized linear ARCH process

$$h_t = \beta_0 + \sum_{l=1}^{\infty} \beta_l \varepsilon_{t-l}^2, \tag{2.2}$$

where $\beta_0 > 0$, $\sum_{l=1}^{\infty} \beta_l < \infty$, and $\beta_l \geq 0$ for all $l \geq 1$, which ensures $h_t > 0$ (cf. Drost and Nijman, 1993; Nelson and Cao, 1992). One example is Engle's (1982) ARCH(q) process:

$$h_t = \beta_0 + \sum_{l=1}^{q} \beta_l \varepsilon_{t-l}^2. \tag{2.3}$$

Another example is Bollerslev's (1986) GARCH(p,q) process:

$$h_t = \gamma_0 + \sum_{l=1}^{q} \alpha_l \varepsilon_{t-l}^2 + \sum_{l=1}^{p} \gamma_l h_{t-l}, \tag{2.4}$$

whose coefficient β_l, which is a function of $\{\alpha_l, \gamma_l\}$, decays to 0 exponentially as $l \to \infty$. The class (2.2) also includes fractionally integrated GARCH processes, whose coefficients $\beta_l \to 0$ as $l \to \infty$ at a slow hyperbolic rate (e.g., Baillie et al., 1996).

Under (2.2), the null hypothesis of no ARCH effect can be stated as

$$\mathbb{H}_0 : \beta_l = 0 \quad \text{for all } l = 1, 2, \ldots .$$

The alternative hypothesis that ARCH effects exist is

$$\mathbb{H}_A : \beta_l \geq 0 \quad \text{for all } l = 1, 2, \ldots, \text{ with at least one strict inequality.}$$

The alternative $\mathbb{H}_A$ is thus one sided. To test this hypothesis, we use a frequency domain approach. Let $f(\omega)$ be the standardized spectral density of $\{\varepsilon_t^2\}$; i.e.,

$$f(\omega) = (2\pi)^{-1} \sum_{l=-\infty}^{\infty} \rho(l) e^{-il\omega}, \qquad \omega \in [-\pi, \pi], \tag{2.5}$$

where $\rho(l)$ is the autocorrelation function of $\{\varepsilon_t^2\}$ and $i = \sqrt{-1}$. Note that (2.2) implies

$$\varepsilon_t^2 = \beta_0 + \sum_{l=1}^{\infty} \beta_l \varepsilon_{t-l}^2 + v_t, \tag{2.6}$$

where $E(v_t | \mathcal{I}_{t-1}) = 0$ almost surely. Under $\mathbb{H}_0$, $\{\varepsilon_t^2\} = \beta_0 + v_t$ is a white noise (in fact $\{\varepsilon_t^2\}$ is i.i.d. given Assumption A.1), so we have $f(0) = (2\pi)^{-1}$. Under $\mathbb{H}_A$, we have $\rho(l) \geq 0$ for all $l \in \mathbb{Z}$ and there exists at least one $l \neq 0$ such that $\rho(l) > 0$. Thus, $f(0) > (2\pi)^{-1}$ under $\mathbb{H}_A$. This forms a basis for constructing a consistent one-sided test for $\mathbb{H}_0$ vs. $\mathbb{H}_A$. We can use a consistent estimator $\hat{f}(0)$ for $f(0)$ and check if $\hat{f}(0) > (2\pi)^{-1}$ significantly. In Hong (1997), a kernel method is used. In the subsequent discussion, we use a wavelet method. We note that this spectral approach can be extended to check whether a nonstationary time series is a unit root process or a trend-stationary process; see Section 7 for more discussion.

3. WAVELET METHOD

Throughout, we use multiresolution analysis (MRA), introduced in Mallat (1989). MRA is a mathematical method to decompose a square-integrable function or signal $g(\cdot)$ at different scales. The key of MRA is a mother wavelet function $\psi(\cdot)$.

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

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

$$\psi_{jk}(x) = 2^{j/2}\psi(2^j x - k), \qquad j, k \in \mathbb{Z}. \tag{3.1}$$

This sequence is obtained from a single mother wavelet $\psi(\cdot)$ by dilations and translations. The integers j and k are called the dilation and translation parameters. Intuitively, j localizes analysis in frequency and k localizes analysis in time (or space). This joint time-frequency decomposition of information is the key feature of wavelet analysis, explaining why wavelets are attractive for approximating nonsmooth functions. We note that the dilation factor could differ from 2, but "2" ensures the L_2-invariance that $\int_{-\infty}^{\infty} \psi_{jk}^2(x)dx = \int_{-\infty}^{\infty} \psi^2(x)dx$. Often $\psi(\cdot)$ is well localized (i.e., $\psi(x) \to 0$ sufficiently fast as $x \to \infty$), so $\psi_{jk}(\cdot)$ is effectively nonzero only around an interval of width 2^{-j} centered at $k/2^j$.

The mother wavelet $\psi(\cdot)$ can have bounded support. An example is the Haar wavelet:

$$\psi(x) = \begin{cases} 1 & \text{if } 0 < x \leq \dfrac{1}{2}, \\ -1 & \text{if } -\dfrac{1}{2} \leq x < 0, \\ 0 & \text{otherwise.} \end{cases} \tag{3.2}$$

Compact support ensures that $\psi(\cdot)$ is well localized in time domain. Daubechies (1992) shows that for any nonnegative integer D, there exists an orthonormal compact supported wavelet whose first D moments vanish. The mother wavelet $\psi(\cdot)$ can also have infinite support, but it must decay to 0 sufficiently fast at ∞. An example is the Littlewood–Paley (or Shannon) wavelet $\psi(\cdot)$, which is defined via its Fourier transform

$$\hat{\psi}(z) \equiv (2\pi)^{-1/2} \int \psi(x)e^{-izx}dx = (2\pi)^{-1/2}\mathbf{1}(|z| \leq 2\pi), \qquad z \in \mathbb{R}, \tag{3.3}$$

where $\mathbf{1}(\cdot)$ is the indicator function. Other examples include the Franklin wavelet, Lemarie–Meyer wavelets, and spline wavelets. See, e.g., Hernandez and Weiss (1996).

For any $g(\cdot) \in L_2(\mathbb{R})$, there exists a wavelet representation

$$g(x) = \sum_{j\in\mathbb{Z}} \sum_{k\in\mathbb{Z}} \alpha_{jk} \psi_{jk}(x), \qquad x \in \mathbb{R}, \tag{3.4}$$

where the wavelet coefficient

$$\alpha_{jk} = \int_{-\infty}^{\infty} g(x)\psi_{jk}(x)dx \tag{3.5}$$

(cf. Mallat, 1989; Daubechies, 1992). Given the localization property of $\psi(\cdot)$, α_{jk} basically depends on the local property of $g(\cdot)$ on an interval of width 2^{-j} centered at $k/2^j$. This differs from the Fourier representation, where each Fourier coefficient depends on the global property of $g(\cdot)$. An essential feature of wavelet analysis is that wavelets, in an "automatic manner," evaluate high frequency components of $g(\cdot)$ on small intervals and low frequency components of $g(\cdot)$ on large intervals. Consequently, they can effectively represent nonsmooth functions with a relatively small number of wavelet coefficients.

To represent the standardized spectral density $f(\cdot)$ of $\{\varepsilon_t^2\}$, which is 2π-periodic and thus is not square-integrable on $\mathbb{R}$, we need to periodize the wavelet basis $\{\psi_{jk}(\cdot)\}$ via

$$\boldsymbol{\Psi}_{jk}(\omega) = (2\pi)^{-1/2} \sum_{m=-\infty}^{\infty} \psi_{jk}\left(\frac{\omega}{2\pi} + m\right), \qquad \omega \in \mathbb{R}, \tag{3.6}$$

which is 2π-periodic. With these periodic orthonormal bases for $L_2[-\pi,\pi]$, we can write

$$f(\omega) = (2\pi)^{-1} + \sum_{j=0}^{\infty} \sum_{k=1}^{2^j} \alpha_{jk} \boldsymbol{\Psi}_{jk}(\omega), \qquad \omega \in [-\pi,\pi], \tag{3.7}$$

where the wavelet coefficient

$$\alpha_{jk} = \int_{-\pi}^{\pi} f(\omega)\boldsymbol{\Psi}_{jk}(\omega)d\omega. \tag{3.8}$$

See Lee and Hong (2001) and Hong (2001) for more discussions.

Now, we denote the Fourier transform of $\psi(\cdot)$ by

$$\hat{\psi}(z) = (2\pi)^{-1/2} \int_{-\infty}^{\infty} \psi(x)e^{-izx}dx, \qquad z \in \mathbb{R}. \tag{3.9}$$

Assumption A.2 ensures that $\hat{\psi}(\cdot)$ exists and is continuous almost everywhere in $\mathbb{R}$, with $|\hat{\psi}(z)| \le C$, $\hat{\psi}(-z) = \hat{\psi}^*(z), \hat{\psi}(0) = 0$, and $\int_{-\infty}^{\infty} |\hat{\psi}(z)|^2 dz = 1$. By Parseval's identity, we can express the wavelet coefficient of (3.8) as

$$\alpha_{jk} = (2\pi)^{-1/2} \sum_{l=-\infty}^{\infty} \rho(l)\hat{\Psi}_{jk}^{*}(l), \tag{3.10}$$

where $\hat{\Psi}_{jk}(\cdot)$ is the Fourier transform of $\Psi_{jk}(\cdot)$; i.e.,

$$\hat{\Psi}_{jk}(l) = (2\pi)^{-1/2}\int_{-\pi}^{\pi} \Psi_{jk}(\omega)e^{-il\omega}d\omega = e^{-2\pi lk/2^j}(2\pi/2^j)\hat{\psi}(2\pi l/2^j). \tag{3.11}$$

In (3.11) the second equality follows from (3.6) and a change of variable. Note that the translation parameter k is converted into a "modulation," i.e., the multiplication of an exponential. This is a natural consequence of the Fourier transform of convolution.

We impose an additional assumption on $\psi(\cdot)$.

Assumption A.3. $|\hat{\psi}(z)| \le C\min\{|z|^{\kappa},(1+|z|)^{-\tau}\}$ for some $\kappa > 0$ and $\tau > 1$.

This requires that $\hat{\psi}(\cdot)$ have some regularity (i.e., smoothness) at 0 and sufficiently fast decay at ∞. The condition $|\hat{\psi}(z)| \le C|z|^{\kappa}$ is effective as $z \to 0$, where κ governs the degree of smoothness of $\hat{\psi}(\cdot)$ at 0. If $\int_{-\infty}^{\infty}(1+|x|^{\nu})|\psi(x)|dx < \infty$ for some $\nu > 0$, then $|\hat{\psi}(z)| \le C|z|^{\kappa}$ for $\kappa = \min(\nu,1)$ (cf. Priestley, 1996). When $\psi(\cdot)$ has first D vanishing moments (i.e., $\int_{-\infty}^{\infty}x^r\psi(x)dx = 0$ for $r = 0,\dots,D-1$), we have $|\hat{\psi}(z)| \le C|z|^D$ as $z \to 0$. On the other hand, the condition $|\hat{\psi}(z)| \le C(1+|z|)^{-\tau}$ is effective as $z \to \infty$. This holds trivially for the so-called band-limited wavelets, whose $\hat{\psi}(\cdot)$'s have compact supports.

Most commonly used wavelets satisfy Assumption A.3. Examples are Daubechies's (1992) compactly supported wavelets of positive order, the Franklin wavelet, Lemarie–Meyer wavelets, Littlewood–Paley wavelets, and spline wavelets. Assumption A.3 rules out the Haar wavelet, however, because its $\hat{\psi}(z) = -ie^{iz/2}\sin^2(z/4)/(z/4) \to 0$ at a rate of $|z|^{-1}$ only.

To obtain a feasible wavelet estimator of $f(0)$, we use the estimated regression residual

$$\hat{\varepsilon}_t = Y_t - g(X_t,\hat{b}), \tag{3.12}$$

where $\hat{b}$ is an estimator of b_0. We impose the following conditions on $g(\cdot,\cdot)$ and $\hat{b}$.

Assumption A.4.

(i) For each $b \in \mathbb{B}$, where $\mathbb{B}$ is a finite-dimensional subset, $g(\cdot,b)$ is a measurable function of $X_t \in \mathcal{I}_{t-1}$.

(ii) $g(X_t,\cdot)$ is twice continuously differentiable with respect to b in an open convex neighborhood $\mathbb{B}_0$ of b_0 almost surely, with $\lim_{n\to\infty}\{n^{-1}\sum_{t=1}^{n}E \times \sup_{b\in\mathbb{B}_0}\|(\partial/\partial b)g(X_t,b)\|^4\} < \infty$ and $\lim_{n\to\infty}\{n^{-1}\sum_{t=1}^{n}E\sup_{b\in\mathbb{B}_0}\|(\partial^2/\partial b\partial b') \times g(X_t,b)\|^2\} < \infty$.

Assumption A.5. $n^{1/2}(\hat{b} - b_0) = O_P(1)$.

We permit but do not require that $\hat{b}$ be the ordinary least square (OLS) or quasi-maximum likelihood estimators (e.g., Lee and Hansen, 1994; Lumsdaine, 1996).

Now, define the sample autocorrelation function of the squared residual series $\{\hat{\varepsilon}_t^2\}$

$$\hat{\rho}(l) = \hat{R}(l)/\hat{R}(0), \qquad l = 0, \pm 1, \dots, \pm(n-1), \tag{3.13}$$

where the sample autocovariance of $\{\hat{\varepsilon}_t^2\}$

$$\hat{R}(l) = n^{-1} \sum_{t=|l|+1}^{n} (\hat{\varepsilon}_t^2/\hat{\sigma}^2 - 1)(\hat{\varepsilon}_{t-|l|}^2/\hat{\sigma}^2 - 1), \tag{3.14}$$

with $\hat{\sigma}^2 = n^{-1} \sum_{t=1}^{n} \hat{\varepsilon}_t^2$. A wavelet spectral estimator for $f(0)$ can be given as

$$\hat{f}(0) = (2\pi)^{-1} + \sum_{j=0}^{J} \sum_{k=1}^{2^j} \hat{\alpha}_{jk} \Psi_{jk}(0), \tag{3.15}$$

where the empirical wavelet coefficient

$$\hat{\alpha}_{jk} = \int_{-\pi}^{\pi} \hat{I}(\omega) \Psi_{jk}(\omega) d\omega = (2\pi)^{-1/2} \sum_{l=1-n}^{n-1} \hat{\rho}(l) \hat{\Psi}_{jk}^*(l), \tag{3.16}$$

with $\hat{I}(\omega) = (2\pi n)^{-1} |\sum_{t=1}^{n} \hat{\varepsilon}_t^2 e^{i\omega t}|^2$ the periodogram of $\{\hat{\varepsilon}_t^2\}$. For compactly supported wavelets $\psi(\cdot)$, $\Psi_{jk}(\cdot)$ in (3.6) is a sum of finite terms, so the first expression of $\hat{\alpha}_{jk}$ in (3.16) is efficient to compute. Alternatively, for band-limited wavelets (whose $\hat{\psi}(\cdot)$'s have compact supports), the second expression of $\hat{\alpha}_{jk}$ is efficient to compute.

The integer J is called the finest scale parameter. It corresponds to the highest resolution level used in the wavelet approximation. Given each J, there are in total $2^{J+1} - 1$ empirical wavelet coefficients used in $\hat{f}(0)$. To reduce the bias of $\hat{f}(0)$, we must let $J \equiv J_n \to \infty$ as $n \to \infty$. To ensure that the variance of $\hat{f}(0)$ vanishes, however, $2^{J+1} - 1$ must grow more slowly than n. Thus, we need to choose J properly to balance the bias and variance. Although J is a smoothing parameter, it cannot be viewed as a lag truncation parameter, because even if $J = 0$ the empirical wavelet coefficient $\hat{\alpha}_{Jk}$ is still a weighted sum of all $n - 1$ sample autocorrelations $\{\hat{\rho}(l)\}_{l=1}^{n-1}$ provided $\hat{\psi}(\cdot)$ has unbounded support. We will provide proper conditions on J to ensure that the proposed test is well behaved.

The estimator $\hat{f}(0)$ is essentially a consistent long-run variance estimator for $\{\varepsilon_t^2\}$. Long-run variance-covariance estimation is important in time series econometrics. The existing popular long-run variance-covariance estimators are kernel-based (cf. Andrews, 1991; Newey and West, 1987). Our approach provides an alternative. It is expected to perform better in finite samples when data exhibit strong autocorrelation, which generates a spectral peak at frequency zero. See Hong (2001) for more discussion.

4. TEST STATISTIC AND ITS DISTRIBUTION

To introduce our test statistic for $\mathbb{H}_0$, we define the function

$$\lambda(z) = 2\pi\hat{\psi}^*(z) \sum_{m=-\infty}^{\infty} \hat{\psi}(z + 2\pi m), \qquad z \in \mathbb{R}. \tag{4.1}$$

Assumptions A.2 and A.3 imply that $\lambda(\cdot)$ is continuous almost everywhere in $\mathbb{R}, \lambda(0) = 0$, and $|\lambda(z)| \leq C$. Note that the tail behavior of $\lambda(\cdot)$ is governed by $\hat{\psi}(\cdot)$, because $\sum_{m=-\infty}^{\infty} \hat{\psi}(z + 2\pi m)$ is 2π-periodic. We impose a condition on $\lambda(\cdot)$.

Assumption A.6. $\lambda : \mathbb{R} \to \mathbb{R}$ is square-integrable.

Most commonly used wavelets satisfy this assumption. Because $\hat{\psi}^*(z) = \hat{\psi}(-z)$ given Assumption A.2, the condition that $\lambda(\cdot)$ is real-valued implies $\lambda(-z) = \lambda(z)$.

Our test statistic for $\mathbb{H}_0$ vs. $\mathbb{H}_A$ is defined as

$$S_n(J) = V_n^{-1/2}(J) n^{1/2} \pi[\hat{f}(0) - (2\pi)^{-1}], \qquad J \in \mathbb{Z}, \tag{4.2}$$

where the asymptotic variance estimator

$$V_n(J) = \sum_{l=1}^{n-1} (1 - l/n) \left[\sum_{j=0}^{J} \lambda(2\pi l/2^j) \right]^2. \tag{4.3}$$

The factor $1 - l/n$ is a finite sample correction; it could be replaced by unity.

The statistic $S_n(J)$ is applicable for both small J (i.e., J is fixed) and large J (i.e., $J \equiv J_n \to \infty$ as $n \to \infty$). For (and only for) large J, we could also use the statistic

$$\tilde{S}_n(J) = (n/2^J)^{1/2} V_0^{-1/2} \pi[\hat{f}(0) - (2\pi)^{-1}], \tag{4.4}$$

where

$$V_0 = \int_0^{2\pi} |\Gamma(z)|^2 dz \tag{4.5}$$

and

$$\Gamma(z) = \sum_{m=-\infty}^{\infty} \hat{\psi}(z + 2m\pi). \tag{4.6}$$

This statistic has the same null asymptotic distribution as $S_n(J)$ when J is large, because $2^{-J} V_n(J) \to V_0$ as $J \to \infty$ (see Lemma A.2 in the Appendix). It is simpler to compute than $S_n(J)$ but may have less desirable sizes in finite samples, especially when J is small.

THEOREM 1. *Suppose that Assumptions A.1–A.6 hold and $2^J/n \to 0$. Then under $\mathbb{H}_0$*

$$S_n(J) \to^d N(0,1).$$

Both small and large (i.e., fixed and increasing) finest scales J are allowed here. Thus, the choice of J has no impact on the null limit distribution of $S_n(J)$, as long as 2^J grows more slowly than the sample size n. Of course, it may have impact on the finite sample distribution of $S_n(J)$, and it is expected to significantly affect the power of $S_n(J)$ under $\mathbb{H}_A$. We will examine the impact of the choice of J in simulation. We note that because $S_n(J)$ is a one-sided test, it is appropriate to use upper-tailed $N(0,1)$ critical values. The critical value at the 5% level, e.g., is 1.645.

5. ASYMPTOTIC LOCAL POWER

We now consider a class of generalized linear local alternatives

$$\mathbb{H}_n(a_n): h_t = \sigma_0^2\left[1 + a_n \sum_{l=1}^{\infty} \beta_l(\varepsilon_{t-l}^2 - 1)\right], \tag{5.1}$$

where $\beta_l \geq 0, \sum_{l=1}^{\infty} \beta_l < \infty$, and $a_n \to 0$ governs the rate at which the local alternatives converge to $\mathbb{H}_0$. Without loss of generality we assume $a_n \sum_{l=1}^{\infty} \beta_l < 1$ for all n to ensure $h_t > 0$. The class $\mathbb{H}_n(a_n)$ describes all linear local ARCH alternatives, including ARCH, GARCH, and fractionally integrated GARCH processes of known or unknown orders.

THEOREM 2. *Suppose that Assumptions A.1–A.6 hold.*

(i) Let $J \in \mathbb{Z}$ be fixed. Define $\mu(J) = V_0(J)^{-1/2} \sum_{l=1}^{\infty} d_J(l)\beta_l$, where $V_0(J) = \sum_{l=1}^{\infty} d_J^2(l)$ and $d_J(l) = \sum_{j=0}^{J} \lambda(2\pi l/2^j)$. Then under $\mathbb{H}_A(n^{-1/2})$,

$$S_n(J) \to^d N[\mu(J),1].$$

(ii) Let $J \to \infty, 2^{2J}/n \to 0$. Define $\mu = V_0^{-1/2} \sum_{l=1}^{\infty} \beta_l$, where V_0 is as in (4.5). Then under $\mathbb{H}_n(2^{J/2}/n^{1/2})$,

$$S_n(J) \to^d N(\mu,1).$$

Theorem 2(i) implies that with fixed finest scale J, $S_n(J)$ has nontrivial power against $\mathbb{H}_n(a_n)$ with parametric rate $a_n = n^{-1/2}$, provided $\sum_{l=1}^{\infty} d_J(l)\beta_l > 0$. It has no power when $\sum_{l=1}^{\infty} d_J(l)\beta_l = 0$, which may occur for a fixed J, because $d_J(l)$ is a local average, depending on J and wavelet $\psi(\cdot)$. On the other hand, Theorem 2(ii) implies that with $J \equiv J_n \to \infty$, $S_n(J)$ has nontrivial power against all linear local ARCH processes in $\mathbb{H}_n(2^{J/2}/n^{1/2})$. This follows because the noncentrality parameter $\mu > 0$ whenever linear local ARCH effects exist (i.e., there exists at least one parameter $\beta_l > 0$ for some $l > 0$ in (5.1) with $a_n = 2^{J/2}/n^{1/2}$). Hong's (1997) one-sided kernel test also has nontrivial power against $\mathbb{H}_n(a_n)$ with an analogous rate. The one-sided tests of Lee and King (1993) and Demos and Sentana (1998) may not have nontrivial power against $\mathbb{H}_n(a_n)$ for any rate, because they are interested in testing a parametric ARCH(q) for fixed q or GARCH(1,1). The one-sided tests for GARCH(1,1) of Lee and King (1993) and Demos and Sentana (1998) numerically coincide with their tests for ARCH(1), respectively. The extension to testing GARCH(p,q) for $p,q > 1$ is more difficult, because some parameters do not lie on the boundary of the parameter space (cf. Lee and King, 1993; Demos and Sentana, 1998). In addition to the nontrivial power against $\mathbb{H}_n(2^{J/2}/n^{1/2})$, the $S_n(J)$ test is also convenient to use because it does not requires formulating an alternative model. For example, there is no need to specify the order of an ARCH, a GARCH, or a fractionally integrated GARCH model.

The ability of $S_n(J)$ to detect $\mathbb{H}_n(2^{J/2}/n^{1/2})$ is desirable when no prior information about the alternative is known. This is, however, achieved at the price that $S_n(J)$ can detect $\mathbb{H}_n(a_n)$ with a rate of $a_n = 2^{J/2}/n^{1/2}$, which is slower than the parametric rate $n^{-1/2}$ because $J \to \infty$. Nevertheless, this may not be taken too literally in practice. For example, if $2^J \propto (\ln n)^2$, we have $a_n \propto n^{-1/2}\ln(n)$, which is only slightly slower than $n^{-1/2}$. We note that in other contexts, some consistent nonparametric tests allow using a fixed smoothing parameter and are thus able to detect local alternatives with parametric rate $n^{-1/2}$. An example is the class of consistent kernel-based specification tests considered by Fan and Li (2000), who show that Bierens-type consistent integrated conditional moment tests for model specification can be viewed as a kernel-based test with a fixed bandwidth or vice versa. See Fan and Li (2000) for more discussion.

It should be emphasized that the ability of $S_n(J)$ to detect all linear local ARCH processes in $\mathbb{H}_n(2^{J/2}/n^{1/2})$ does not mean that rejecting $\mathbb{H}_0$ will accept a linear ARCH alternative. In fact, with $J \to \infty$ as $n \to \infty$, $S_n(J)$ will have power against all linear and nonlinear ARCH processes whenever $f(0) > (2\pi)^{-1}$. Thus, it may be better to interpret $S_n(J)$ as a test for conditional homoskedasticity ($E(\varepsilon_t^2|\mathcal{I}_{t-1}) = \sigma_0^2$ for some σ_0^2 almost surely) versus general conditional heteroskedasticity ($P[E(\varepsilon_t^2|\mathcal{I}_{t-1}) = \sigma^2] < 1$ for all σ^2). Of course, it may be noted that $S_n(J)$ has no power against nonlinear ARCH processes whose $f(0) = (2\pi)^{-1/2}$. This may occur, e.g., if h_t follows a tent map process.

Although the choice of J has no impact on the null limit distribution of $S_n(J)$, it may significantly affect the power of $S_n(J)$ in finite samples. It is not easy, however, to choose an optimal J that maximizes the power, especially in light of the facts that J is not a lag order and that usually no prior information on the alternative is available. Therefore, it is desirable to choose J via data-driven methods, which are more objective than any arbitrary choice of J or any simple "rule of thumb." To allow for this possibility, we consider using a data-dependent finest scale $\hat{J}$ (say). It can be shown (the proofs are available from the authors) that under the conditions that Assumptions A.1–A.6 hold and $\hat{J} - J = o_P(2^{-J/2})$, where J is nonstochastic and $2^J/n \to 0$, we have $S_n(\hat{J}) - S_n(J) \to^p 0$ and $S_n(\hat{J}) \to^d N(0,1)$ under $\mathbb{H}_0$. Thus, the randomness of data-driven $\hat{J}$ has no impact on the null limit distribution of $S_n(\hat{J})$, as long as $\hat{J} - J \to^p 0$ at a rate faster than $2^{-J/2}$. Note that when J is fixed ($J = 0$, say), as may occur under $\mathbb{H}_0$ for sensible data-driven methods, $\hat{J} - J = o_P(2^{-J/2})$ becomes $\hat{J} - J \to^p 0$; no rate condition on $\hat{J}$ is required.

So far very few data-driven methods to choose J are available in the literature. To our knowledge, only Walter (1994) proposes a data-driven $\hat{J}$, using an integrated mean square error (IMSE) criterion. It is based on the fact that the change in IMSE of $\hat{f}(\cdot)$ 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 a scale J. One starts from $J = 0$ and checks how IMSE changes from $J = 0$ to $J = 1$. The grid search is iterated until one gets the scale $\hat{J}$ at that IMSE increases most rapidly. Then, one obtains the finest scale $\hat{J}$. This $\hat{J}$ gives an IMSE that we cannot improve practically by further increasing J. Such an increase would increase the variance without a corresponding reduction in bias. This method is more suitable for estimation of $f(\cdot)$ on $[-\pi, \pi]$ rather than at 0, but it has a great appeal—simplicity. Particularly, there is no need to use any preliminary estimators for $f(0)$ and its derivative. We will use it in our simulation that follows. It should be noted, however, that although we conjecture that Walter's algorithm might satisfy the condition $\hat{J} - J = o_P(2^{-J/2})$ under $\mathbb{H}_0$ where the optimal finest scale is $J = 0$, there is no formal result on the rate of Walter's (1994) $\hat{J}$ in the literature.

6. MONTE CARLO EVIDENCE

We now study the finite sample performance of $S_n(J)$. We use the Franklin wavelet and the second-order spline wavelets. The Franklin wavelet is defined via its Fourier transform,

$$\hat{\psi}(z) = (2\pi)^{-1/2} e^{iz/2} \frac{\sin^4(z/4)}{(z/4)^2} \left\{ \frac{1 - (2/3)\cos^2(z/4)}{[1 - (2/3)\sin^2(z/2)][1 - (2/3)\sin^2(z/4)]} \right\}^{1/2}. \tag{6.1}$$

For the second-order spline wavelet, its Fourier transform

$$\hat{\psi}(z) = -(2\pi)^{-1/2} i e^{iz/2} \frac{\sin^6(z/4)}{(z/4)^3} \left[\frac{P(z/4 + \pi/2)}{P(z/2)P(z/4)} \right]^{1/2}, \tag{6.2}$$

where $P(z) = \frac{1}{30}\cos^2(2z) + \frac{13}{30}\cos(2z) + \frac{8}{15}$. The tests are denoted as S_1 and S_2.

To examine the impact of the choice of finest scale J, we consider $J = 0,1,2,3$ for each sample size n, which corresponds to using $2^{J+1} - 1 = 1,3,7,15$ empirical wavelet coefficients. We also use a data-driven $\hat{J}$ via Walter's (1994) algorithm. We choose the finest scale $\hat{J}$ for which the change in IMSE from $\hat{J}$ to $\hat{J}+1$ exceeds 100%.

We compare S_1 and S_2 with three one-sided ARCH tests—Hong's (1997) kernel test (denoted K), Lee and King's (1993, 1994) locally most mean powerful test (denoted LK), and Demos and Sentana's (1998) one-sided LM test (denoted DS). We also include Engle's (1982) two-sided LM test (LM). For the kernel test K, we use the quadratic-spectral kernel and a data-driven bandwidth via Andrews's (1991) plug-in method with an ARCH(1) approximating model. For the LK test, we use a statistic robust to nonnormality (Lee and King (13), 1993). The test statistics S_1, S_2, K, and LK are all asymptotically one-sided $N(0,1)$ under $\mathbb{H}_0$. The DS statistic is the sum of the squared t-statistics of positive estimated coefficients in the regression of $\hat{\varepsilon}_t^2$ on a constant and the first q lags of $\hat{\varepsilon}_t^2$. It has a nonstandard limit mixed χ^2 distribution; critical values are in Demos and Sentana (1998, Table 1). The LM test statistic has a null limit χ_q^2 distribution and is computed as $(n-q)R^2$, where R^2 is the squared correlation coefficient in the regression of $\hat{\varepsilon}_t^2$ on a constant and the first q lags of $\hat{\varepsilon}_t^2$. For LK, DS, and LM, the lag order q has to be chosen a priori. These tests will attain their maximal powers when using the optimal lag order, which depends on the true alternative. If the alternative is unknown, as often occurs in practice, these tests may suffer from power losses when using a suboptimal lag order. To examine the effect of the choice of q for these tests, we use $q = 1,12$. The corresponding tests are denoted as $LK(1), DS(1), LM(1), LK(12)$, $DS(12)$, and $LM(12)$.

We consider the data generating process

$$Y_t = X_t' b_0 + \varepsilon_t, \qquad \varepsilon_t = \xi_t h_t^{1/2}, \qquad t = 1,\dots,n, \tag{6.3}$$

where $X_t = (1, m_t)'$, $m_t = 0.8m_{t-1} + v_t, v_t \sim$ i.i.d. $N(0,4)$, $\xi_t \sim$ i.i.d. $N(0,1)$, and $b_0 = (1,1)'$. Both $\{\xi_t\}$ and $\{v_t\}$ are mutually independent. We estimate b_0 by OLS. As in Engle et al. (1985), the exogenous variable m_t is generated for each experiment and held fixed from iteration to iteration. We consider $n =$ 100, 200, 500, 1,000 for the size of the tests and $n =$ 100, 200 for the power. The initial values for ε_t, $t \leq 0$, are set to be 0, and h_t, $t \leq 0$, is set to be 1. To reduce the effects of the initial values, we generate $n +$ 1,000 observations and then discard the first 1,000 ones. For each experiment, we generate 1,000 iterations using GAUSS Windows/NT version random number generator.

We study the size by setting $h_t = 1$. Table 1 reports the size at the 10% and 5% levels using asymptotic critical values. We first look at the wavelet tests. When $J = 0,1$, both S_1 and S_2 perform reasonably well for all sample sizes, with S_2 slightly better than S_1 in most cases. When $J = 2,3$ (particularly $J = 3$), S_1 and S_2 are undersized when $n =$ 100, 200, but they become reasonable when $n =$ 500, 1,000. Walter's algorithm gives reasonable sizes for both S_1 and S_2. (In the notes to Tables 1–4, which follow, we report the mean and standard deviation of Walter's $\hat{J}$.) For the other tests, the sizes of the tests $K, LK(1)$, and

Table 1. Size at 10% and 5% levels

	$n = 100$		$n = 200$		$n = 500$		$n = 1{,}000$	
	10%	5%	10%	5%	10%	5%	10%	5%
$S_1(Walter)$	7.3	4.5	8.2	4.8	10.4	6.0	8.4	4.4
$S_1(0)$	9.6	5.0	8.6	4.2	9.1	5.1	9.4	5.1
$S_1(1)$	7.7	5.2	6.7	4.1	9.2	6.1	8.3	3.7
$S_1(2)$	6.3	3.9	8.1	4.6	8.7	4.4	7.4	4.0
$S_1(3)$	4.2	2.6	6.5	4.2	8.9	4.6	8.7	3.9
$S_2(Walter)$	9.0	5.4	8.2	4.2	11.3	6.2	9.1	5.1
$S_2(0)$	9.9	5.1	8.2	4.0	9.0	5.1	9.3	4.7
$S_2(1)$	9.0	5.3	8.1	4.2	10.3	4.8	8.3	4.7
$S_2(2)$	7.2	4.2	8.6	4.8	9.9	5.0	7.2	3.3
$S_2(3)$	5.5	2.9	7.7	4.2	9.1	5.1	8.2	4.7
K	8.4	4.0	7.5	3.5	8.0	4.2	8.5	4.2
$LK(1)$	9.8	5.0	8.5	4.0	8.9	5.0	9.3	4.8
$DS(1)$	9.9	5.3	8.3	4.1	9.0	5.1	9.4	4.7
$LM(1)$	7.8	4.0	8.0	4.3	9.5	4.3	9.1	4.6
$LK(12)$	6.3	2.8	6.8	3.0	8.0	4.2	7.3	3.7
$DS(12)$	11.6	7.3	11.5	6.7	12.6	7.5	10.8	6.0
$LM(12)$	6.1	1.9	7.1	3.2	10.1	5.1	8.8	4.7

Notes: Model: $Y_t = 1 + m_t + \varepsilon_t$, $m_t = 0.8m_{t-1} + v_t$, $v_t \sim NID(0,4)$, $\varepsilon_t \sim NID(0,1)$. Number of iterations = 1,000. The mean and standard deviation (in parentheses) of Walter's (1994) $\hat{J}$ are shown. The S_1 test: 1.87(1.01), 2.12(1.07), 2.57(1.20), 2.72(1.28) for n = 100, 200, 500, 1,000, respectively. The S_2 test: 1.89(1.02), 2.15(1.06), 2.56(1.19), 2.77(1.26) for n = 100, 200, 500, 1,000, respectively.

$DS(1)$ are reasonable for all sample sizes. The tests $LK(12)$ and $LM(12)$ show some underrejections, whereas $DS(12)$ tends to overreject slightly except for $n = 1{,}000$.

Next, we investigate the power under the following ARCH alternatives:

$$\text{ARCH}(1): \qquad h_t = 1 + \beta\varepsilon_{t-1}^2,$$

$$\text{ARCH}(12a): \qquad h_t = 1 + \beta\sum_{l=1}^{12}\varepsilon_{t-l}^2,$$

$$\text{ARCH}(12b): \qquad h_t = 1 + \beta\sum_{l=1}^{12}(1 - l/13)\varepsilon_{t-l}^2,$$

$$\text{GARCH}(1,1): \qquad h_t = 1 + \alpha\varepsilon_{t-1}^2 + \beta h_{t-1}.$$

For ARCH(1), we set β = 0.3, 0.95. ARCH(1) has no sharp peak for $f(\cdot)$ at any frequency. In contrast, ARCH(12a) and ARCH(12b) have a relatively long distributional lag, which generates a peak for $f(\cdot)$ at 0. Linearly decaying weights in ARCH(12b) are often considered in the literature (e.g., Engle, 1982). We set $\beta = 0.95/12$ for ARCH(12a) and $\beta = 0.95/\sum_{l=1}^{12}(1 - l/13)$ for ARCH(12b). GARCH(1,1) is a workhorse in modeling economic and financial time series. When $\alpha + \beta < 1$, GARCH(1,1) can be expressed as ARCH(∞) with exponentially decaying coefficients. We set $(\alpha, \beta) = (0.3, 0.3), (0.3, 0.65)$. The latter displays relatively persistent ARCH effects, yielding a peak for $f(\cdot)$ at 0. We consider the size-corrected power under these alternatives, using the empirical critical values obtained from 1,000 replications under $\mathbb{H}_0$.

Table 2 reports the power against ARCH(1). We first look at the wavelet tests. Both S_1 and S_2 have similar power. The choice of $J = 0$ and in some cases the choice of $J = 1$ give the best power, whereas the choice of $J = 3$ gives the smallest power. Walter's algorithm yields power similar to $J = 2$ in most cases. Next, we compare S_1 and S_2 using Walter's algorithm with the other tests. As expected, $LK(1)$ and $DS(1)$ have similar powers and are most power-

TABLE 2. Size-adjusted power against ARCH(1) at 10% and 5% levels

	$\beta = 0.3$				$\beta = 0.95$			
	$n = 100$		$n = 200$		$n = 100$		$n = 200$	
	10%	5%	10%	5%	10%	5%	10%	5%
$S_1(Walter)$	51.1	39.5	58.0	47.3	82.3	75.3	90.8	87.7
$S_1(0)$	71.4	60.9	90.8	84.9	97.2	94.9	100	100
$S_1(1)$	64.8	49.3	82.5	74.8	95.1	89.8	99.7	99.3
$S_1(2)$	47.5	34.0	57.6	46.7	85.5	75.8	96.2	93.6
$S_1(3)$	32.7	22.7	40.5	31.6	67.5	57.2	84.0	76.1
$S_2(Walter)$	55.8	43.5	69.3	61.2	87.9	82.2	95.6	93.7
$S_2(0)$	73.1	61.4	90.8	85.6	97.3	95.2	100	99.9
$S_2(1)$	70.2	58.6	94.0	93.0	97.5	93.3	100	100
$S_2(2)$	54.4	42.7	70.3	60.3	91.0	84.8	98.9	97.4
$S_2(3)$	40.7	29.5	52.9	39.4	79.2	70.2	92.9	87.6
K	70.8	60.4	88.8	82.7	97.5	94.8	100	99.7
$LK(1)$	72.8	62.7	90.8	86.0	97.3	95.4	100	100
$DS(1)$	73.1	61.7	90.9	85.7	97.4	95.7	100	99.9
$LM(1)$	64.7	56.0	85.8	81.2	95.9	93.4	100	99.2
$LK(12)$	24.1	16.0	36.8	27.6	46.4	33.6	73.3	63.1
$DS(12)$	40.2	30.4	62.7	54.1	77.5	70.0	94.5	92.1
$LM(12)$	35.4	23.8	60.7	50.8	72.5	60.0	92.3	89.4

Notes: Model: $Y_t = 1 + m_t + \varepsilon_t$, $m_t = 0.8m_{t-1} + v_t$, $v_t \sim NID(0,4)$, $\varepsilon_t = \xi_t h_t^{1/2}$, $\xi_t \sim NID(0,1)$, $h_t = 1 + \beta\varepsilon_{t-1}^2$. Number of iterations = 1,000. The mean and standard deviation (in parentheses) of Walter's (1994) $\hat{J}$ are shown. (a) $\beta = 0.3$: The S_1 test: 1.80(0.97), 2.10(1.08) for $n = 100$, 200, respectively. The S_2 test: 1.81(0.97), 2.08(1.08) for $n = 100$, 200, respectively. (b) $\beta = 0.95$: The S_1 test: 1.82(1.03), 2.28(1.10) for $n = 100$, 200, respectively. The S_2 test: 1.90(1.06), 2.37(1.23) for $n = 100$, 200, respectively.

ful. The kernel test K has power very close to that of $LK(1)$ and $DS(1)$. These three tests have better power than $LM(1)$, which has better power than S_1 and S_2. Both S_1 and S_2 suffer from nontrivial power loss when there is no sharp spectral peak. The fact that $LK(1)$ and $DS(1)$ are most powerful against ARCH(1) is hardly surprising, because they use the optimal lag $q = 1$. The powers of $LK(12)$, $DS(12)$, and $LM(12)$ are substantially smaller. These tests are less powerful than S_1 and S_2.

Table 3 reports the power under ARCH(12a) and ARCH(12b). Again, S_1 and S_2 have similar power. Now, in contrast to ARCH(1), the choice of $J = 3$ gives the best power for S_1 and S_2, whereas $J = 0$ gives the smallest power. Walter's algorithm gives power comparable to the choice of $J = 2$ in most cases. Next, we compare S_1 and S_2 using Walter's algorithm to the other tests. Under ARCH(12a), $LK(12)$ has the best power, and it dominates $DS(12)$. (These

TABLE 3. Size-adjusted power against ARCH(12) at 10% and 5% levels

	ARCH 12(a)				ARCH 12(b)			
	$n = 100$		$n = 200$		$n = 100$		$n = 200$	
	10%	5%	10%	5%	10%	5%	10%	5%
$S_1(Walter)$	64.0	53.7	84.9	80.5	78.7	69.2	86.6	83.6
$S_1(0)$	40.5	31.8	70.3	62.2	59.7	49.4	86.6	80.2
$S_1(1)$	52.3	40.3	83.9	77.6	72.3	58.8	93.8	90.8
$S_1(2)$	65.5	55.1	90.6	86.4	81.6	72.7	97.0	95.4
$S_1(3)$	78.4	68.6	96.0	93.5	87.6	83.2	98.1	97.6
$S_2(Walter)$	60.3	51.4	87.2	84.4	77.3	66.7	91.1	88.9
$S_2(0)$	36.8	27.7	65.0	53.5	54.1	43.0	80.7	72.6
$S_2(1)$	47.1	35.1	86.4	84.4	65.9	53.7	95.5	94.5
$S_2(2)$	58.6	49.9	86.9	81.8	76.7	68.9	95.9	92.8
$S_2(3)$	73.8	63.3	94.0	91.1	87.4	81.7	98.2	97.1
K	39.6	32.9	65.1	59.3	57.3	49.7	81.4	76.8
$LK(1)$	36.8	29.2	64.6	53.9	53.5	44.6	80.7	72.4
$DS(1)$	36.9	28.3	65.1	53.7	53.9	43.4	80.8	72.4
$LM(1)$	31.3	25.4	54.1	46.7	46.7	39.0	72.2	65.8
$LK(12)$	65.8	59.3	93.0	89.8	72.8	65.7	94.1	92.0
$DS(12)$	57.1	46.5	89.8	84.1	67.1	55.6	93.2	89.0
$LM(12)$	49.7	41.2	87.0	81.1	60.0	50.6	91.6	88.1

Notes: Model: $Y_t = 1 + m_t + \varepsilon_t$, $m_t = 0.8m_{t-1} + v_t$, $v_t \sim NID(0,4)$, $\varepsilon_t = \xi_t h_t^{1/2}$, $\xi_t \sim NID(0,1)$. ARCH(12a): $h_t = 1 + \beta\sum_{l=1}^{12}\varepsilon_{t-l}^2$, $\beta = 0.95/12$. ARCH(12b): $h_t = 1 + \beta\sum_{l=1}^{12}(1 - l/13)\varepsilon_{t-j}^2$, $\beta = 0.95/\sum_{l=1}^{12}(1 - l/13)$. Number of iterations = 1,000. The mean and standard deviation (in parentheses) of Walter's (1994) $\hat{J}$ are shown. (a) ARCH(12a): The S_1 test: 2.43(1.16), 3.88(1.36) for $n = 100$, 200, respectively. The S_2 test: 2.50(1.16), 3.93(1.37) for $n = 100$, 200, respectively. (b) ARCH(12b): The S_1 test: 2.62(1.16), 3.95(1.28) for $n = 100$, 200, respectively. The S_2 test: 2.70(1.18), 4.07(1.29) for $n = 100$, 200, respectively.

two tests have used the optimal lag order of 12.) The wavelet tests S_1 and S_2 have power close to that of $LK(12)$. These three tests have better power than the kernel test K. The tests S_1 and S_2 are slightly better than $DS(12)$ and are substantially better than $LM(12)$ for $n = 100$, although the latter uses the optimal lag of 12. This indicates that wavelets work well when ARCH effects have a relatively long distributional lag. Under ARCH(12b), S_1, S_2, and $LK(12)$ have comparable power, and they are more powerful than $DS(12)$, $LM(12)$, K, $LK(1)$, and $DS(1)$.

Table 4 reports the power against GARCH(1,1). When $(\alpha,\beta) = (0,3, 0.3)$, ARCH effects are relatively weak. For S_1 and S_2, the choice of $J = 1$ gives the best power whereas $J = 3$ gives the smallest power. This is similar to the case of ARCH(1). When $(\alpha,\beta) = (0.3, 0.65)$, ARCH effects are relatively persistent. The choice of $J = 2$ gives the best power for S_1 and S_2, whereas $J = 0$ gives the smallest power. This is analogous to the case with ARCH(12a) and

TABLE 4. Size-adjusted power against GARCH(1,1) at 10% and 5% levels

	$(\alpha,\beta) = (0.3,0.3)$				$(\alpha,\beta) = (0.3,0.65)$			
	$n = 100$		$n = 200$		$n = 100$		$n = 200$	
	10%	5%	10%	5%	10%	5%	10%	5%
$S_1(Walter)$	66.6	55.9	76.6	68.2	85.7	77.9	88.1	86.2
$S_1(0)$	73.6	65.0	91.6	87.1	79.6	71.0	97.0	94.0
$S_1(1)$	76.4	63.6	93.0	88.6	88.3	78.9	98.5	97.4
$S_1(2)$	67.1	54.3	81.8	74.2	90.9	85.4	98.6	98.2
$S_1(3)$	48.2	38.2	61.5	53.1	88.6	81.7	96.9	96.1
$S_2(Walter)$	68.1	58.3	83.2	76.6	85.2	77.3	93.6	92.3
$S_2(0)$	73.3	63.4	91.6	86.2	75.8	66.1	95.3	90.2
$S_2(1)$	77.4	66.5	97.2	96.1	86.0	77.3	99.0	98.7
$S_2(2)$	71.3	62.8	88.9	82.1	90.0	85.0	99.0	97.9
$S_2(3)$	58.1	47.5	76.3	64.9	91.4	86.1	98.2	97.2
K	75.2	66.5	91.6	88.6	78.8	72.8	95.4	93.9
$LK(1)$	73.7	64.4	91.0	86.6	76.1	66.7	95.1	90.4
$DS(1)$	73.3	63.8	91.6	86.2	76.2	66.3	95.2	90.4
$LM(1)$	66.3	57.8	86.2	82.6	68.3	62.5	90.6	86.3
$LK(12)$	35.8	24.3	54.3	43.5	70.0	63.1	92.0	88.9
$DS(12)$	46.2	34.0	70.5	61.5	70.4	58.6	94.2	89.3
$LM(12)$	41.1	30.6	67.8	59.5	65.3	56.3	93.0	88.7

Notes: Model: $Y_t = 1 + m_t + \varepsilon_t$, $m_t = 0.8m_{t-1} + v_t$, $v_t \sim NID(0,4)$, $\varepsilon_t = \xi_t h_t^{1/2}$, $\xi_t \sim NID(0,1)$, $h_t = 1 + \alpha\varepsilon_{t-1}^2 + \beta h_{t-1}$. Number of iterations = 1,000. The mean and standard deviation (in parentheses) of Walter's (1994) $\hat{J}$ are shown. (a) $(\alpha,\beta) = (0.3,0.3)$: The S_1 test: 1.88(0.99), 2.29(1.08) for $n = 100$, 200, respectively. The S_2 test: 1.94(1.01), 2.33(1.11) for $n = 100$, 200, respectively. (b) $(\alpha,\beta) = (0.3,0.65)$: The S_1 test: 2.42(1.11), 3.45(1.27) for $n = 100$, 200, respectively. The S_2 test: 2.52(1.14), 3.58(1.25) for $n = 100$, 200, respectively.

ARCH(12b). Walter's algorithm gives power similar to the choice of $J = 1$. We now compare S_1 and S_2 using Walter's algorithm with the other tests. When $(\alpha, \beta) = (0.3, 0.3)$, the kernel test K has the best power, followed closely by $LK(1)$ and $DS(1)$, then by S_1 and S_2, and finally by $LM(1)$. Nevertheless, the power difference among these tests is marginal. The tests $DS(12), LM(12)$, and $LK(12)$ suffer from severe power losses. When $(\alpha, \beta) = (0.3, 0.65)$, S_1 and S_2 perform the best among all the tests. They outperform K, which is more powerful than $LK(1)$, $DS(1)$, and $LM(1)$. The powers of $LK(12)$, $DS(12)$, and $LM(12)$ are smaller than those of $LK(1)$, $DS(1)$, and $LM(1)$, respectively, but the differences are rather small. It appears that using a long lag order may not suffer from severe power loss when ARCH effects are persistent.

In summary, we make to following observations. (1) The wavelet tests, S_1 and S_2, have similar size and power in almost all the cases. The choice of mother wavelet $\psi(\cdot)$ is not important for both the size and the power. The choice of finest scale J is not important for the size (unless J is large and n is small), but it significantly affects the power. Walter's algorithm yields an objective finest scale $\hat{J}$ that yields reasonable power. (2) The powers of the one-sided kernel and wavelet tests depend on the alternative. When ARCH effects are less persistent, the kernel test is more powerful than the wavelet tests. When ARCH effects are relatively persistent, the wavelet tests outperform the kernel test. (3) The tests of LK, DS, LM attain their own maximal powers when using the optimal lag order, but they may suffer from severe power loss when using a suboptimal lag. Under each alternative, the two-sided LM test is always dominated by some one-sided tests using the same lag order. (4) None of the one-sided tests dominates all the others in power for all the alternatives under study. When ARCH effects are less persistent, the kernel test has power comparable to that of $LK(1)$ and $DS(1)$, which use the correct lag order and are most powerful. When ARCH effects are relatively persistent, the wavelet tests using Walter's algorithm have power close to or even better than that of LK and DS using the optimal lag orders. (5) The kernel and wavelet tests using data-driven methods do not require the knowledge of the alternative.

The fact that the kernel test K has good power against less persistent ARCH effects whereas the wavelet tests S_1 and S_2 have good power against relatively persistent ARCH effects suggests that a suitable Bonferroni procedure that combines the wavelet and kernel tests may have good power against both persistent and less persistent ARCH effects. We consider two simple Bonferroni procedures, BF_1, which combines S_1 and K, and BF_2, which combines S_2 and K, where both S_1 and S_2 use Walter's algorithm. The simple BF_1 procedure works as follows. Let P_1 and P_2 be the smaller and larger asymptotic p-values of test statistics $\{S_1, K\}$. Then one rejects $\mathbb{H}_0$ at level α if $P_1 < \alpha/2$. The same procedure applies to BF_2. Table 5 reports the size and power of BF_1 and BF_2 at the 10% and 5% levels. Both BF_1 and BF_2 are undersized, as is expected. In spite of this, however, they do have all-round good power against all the four alternatives. In particular, they have better power than the wavelet tests S_1 and S_2

TABLE 5. Size and power of Bonferroni procedures at 10% and 5% levels

	$n = 100$				$n = 200$			
	$BF1$		$BF2$		$BF1$		$BF2$	
	10%	5%	10%	5%	10%	5%	10%	5%
Size	6.6	3.8	6.8	4.3	6.8	3.5	5.6	3.3
			Power					
ARCH(1): $\beta = 0.3$	59.4	49.9	60.1	50.9	80.6	72.3	81.4	72.9
ARCH(1): $\beta = 0.95$	93.4	90.0	93.3	90.3	99.5	98.7	99.7	98.9
ARCH 12(a)	55.0	49.5	54.5	47.9	86.4	83.2	87.3	84.7
ARCH 12(b)	72.3	65.8	70.5	65.8	94.0	92.4	94.9	93.4
GARCH(1,1): (0.3,0.3)	69.6	62.5	69.9	63.2	89.2	84.8	89.5	85.9
GARCH(1,1): (0.3,0.65)	84.3	79.1	84.1	78.8	98.0	97.1	98.2	97.7
			Size-adjusted power					
ARCH(1): $\beta = 0.3$	66.7	53.7	67.2	55.0	85.6	76.3	86.4	79.5
ARCH(1): $\beta = 0.95$	96.3	91.9	96.4	92.0	99.7	99.2	99.9	99.6
ARCH 12(a)	60.3	51.5	59.0	50.4	89.1	84.3	89.1	86.7
ARCH 12(b)	77.4	68.2	76.1	67.7	95.6	92.9	95.8	94.7
GARCH(1,1): (0.3,0.3)	75.1	66.0	75.8	66.0	91.9	87.0	93.2	88.9
GARCH(1,1): (0.3,0.65)	87.0	81.6	87.5	80.8	98.0	97.3	99.0	98.0

Notes: BF_1, Bonferroni procedure combining S_1 and K; BF_2, Bonferroni procedure consisting of S_2 and K. The size-adjusted power of BF_1 and BF_2 is based on their empirical p-values under H_0. Number of iterations = 1,000.

when ARCH effects are less persistent, and they have better power than the kernel test K when ARCH effects are persistent. It appears that BF_1 and BF_2 do capture the advantages of both wavelets and kernels. We note that the bootstrap could be used to obtain an accurate size for the Bonferroni procedures. The method in Horowitz and Spokoiny (2001) may be useful here.

7. CONCLUSION

We have proposed a one-sided test for ARCH effects using a wavelet spectral density estimator at frequency zero of a squared regression residual series. An essential feature of ARCH is that the squared series is positively correlated at all lags, resulting in a spectral mode at frequency zero. In particular, a spectral peak of the squared series arises when ARCH effects are persistent, or when ARCH effects are small at each individual lag but carry over a long distributional lag. Because the kernel method tends to underestimate peaks, it may not be a powerful tool in small samples when ARCH effects are persistent. In this case, wavelets are expected to perform better because they can effectively capture spectral peaks. This is confirmed in a simulation study. A Bonferroni procedure captures the advantages of both kernel and wavelet approaches.

Our approach can be extended to check adequacy of various conditional variance models. This can be done by using a wavelet spectral density at frequency zero of the squared residual series standardized by the conditional variance estimator. The underlying spectral density is uniform when the conditional variance model is adequate, but it will be generally larger than the uniform density at frequency zero when the conditional variance model is misspecified. We conjecture that parameter estimation in conditional variance models will have no impact on the limit distribution of the proposed test for large J.

Our approach can also be extended to test hypotheses in contexts other than ARCH. In nonstationary time series analysis, e.g., it is important to determine whether a nonstationary time series is a unit root process or a trend-stationary process. Because the spectral density at frequency zero of the first differenced series is zero if the time series is a trend-stationary process and is strictly positive if it is a unit root process, one can distinguish these two processes by testing whether a wavelet spectral density estimator at frequency zero of the first differenced series is significantly positive. Another potential application is detection of conditional duration effects and evaluation of autoregressive conditional duration models for irregularly spaced transaction data (cf. Engle and Russell, 1998) where the spectral density is also larger than the uniform density at frequency zero under the alternative hypothesis. These are left for future study.

REFERENCES

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

Andrews, D.W.K. (1998) Hypothesis testing with a restricted parameter space. *Journal of Econometrics* 84, 155–199.

Andrews, D.W.K. (2001) Testing When a Parameter Is on the Boundary of the Maintained Hypothesis. *Econometrica* 69, 683–734.

Baillie, R., T. Bollerslev, & H. Mikkelsen (1996) Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 74, 3–30.

Bera, A.K. & M.L. Higgins (1992) A test for conditional heteroskedasticity in time series models. *Journal of Time Series Analysis* 13, 501–519.

Bera, A.K., S. Ra, & N. Sakar (1998) Hypothesis testing for some nonregular cases in econometrics. In S. Chakravorty, D. Coondoo, & R. Mookherjee (eds.), *Quantitative Economics: Theory and Practice*, pp. 324–356. New Delhi: Applied Publishers.

Bollerslev, T. (1986) Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31, 307–327.

Bollerslev, T., R. Chou, & K.F. Kroner (1992) ARCH modelling in finance. *Journal of Econometrics* 52, 5–59.

Daubechies, I. (1992) *Ten Lectures on Wavelets*. CBS-NSF Regional Conferences in Applied Mathematics 61, Society for Industrial and Applied Mathematics.

Demos, A. & E. Sentana (1998) Testing for GARCH effects: A one-sided approach. *Journal of Econometrics* 86, 97–127.

Diebold, F.X. (1987) Testing for serial correlation in the presence of ARCH. In *Proceedings from the American Statistical Association, Business and Economic Statistics Section*, pp. 323–328.

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

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

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

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

Drost, F.C. & T.E. Nijman (1993) Temporal aggregation of GARCH processes. *Econometrica* 61, 909–927.

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

Engle, R.F., D.R. Hendry, & D. Trumble (1985) Small-sample properties of ARCH estimators and tests. *Canadian Journal of Economics* 18, 67–93.

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

Fan, Y. & Q. Li (2000) Consistent model specification tests: Kernel-based tests versus Bierens' ICM tests. *Econometric Theory* 16, 1016–1041.

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

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.

Gourieroux, C., A. Holly, & A. Monfort (1982) Likelihood ratio test, Wald test, and Kuhn–Tucker test in linear models with inequality constraints on the regression parameters. *Econometrica* 50, 63–80.

Gregory, A.W. (1989) A nonparametric test for autoregressive conditional heteroskedasticity: A Markov-chain approach. *Journal of Business and Economic Statistics* 7, 107–115.

Hall, P. & C.C. Heyde (1980) *Martingale Limit Theory and Its Application.* New York: Academic Press.

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

Hernandez, E. & G. Weiss (1996) *A First Course on Wavelets.* New York: CRC Press.

Hong, Y. (1997) One-sided ARCH testing in time series models. *Journal of Time Series Analysis* 18, 253–277.

Hong, Y. (2001) Wavelet-Based Estimation for Heteroskedasticity and Autocorrelation Consistent Variance-Covariance Matrices, Manuscript, Department of Economics and Department of Statistical Science, Cornell University.

Hong, Y. & R. Shehadeh (1999) A new test for ARCH effects and its finite sample performance. *Journal of Business and Economic Statistics* 17, 91–108.

Horowitz, J. & V.G. Spokoiny (2001) An adaptive, rate-optimal test of a parametric model against a nonparametric alternative. *Econometrica* 69, 599–631.

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.

King, M. & P. Wu (1997) Locally optimal one-sided tests for multiparameter hypotheses. *Econometric Reviews* 16, 131–156.9.

Lee, J. & Y. Hong (2001) Testing for serial correlation of unknown form using wavelet methods. *Econometric Theory* 17, 386–423.

Lee, J.H.H. & M.L. King (1993) A locally most mean powerful based score test for ARCH and GARCH regression disturbances. *Journal of Business and Economic Statistics* 11, 17–27.

Lee, J.H.H. & M.L. King (1994) Correction to a locally most mean powerful based score test for ARCH and GARCH regression disturbances. *Journal of Business and Economic Statistics* 12, 139.

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

Lumsdaine, R. (1996) Consistency and asymptotic normality of the quasi-maximum likelihood estimator in IGARCH(1,1) and covariance stationary GARCH(1,1) models. *Econometrica* 64, 575–596.

Mallat, S.G. (1989) A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 11, 674–693.

McLeod, A.I. & W.K. Li (1983) Diagnostic checking ARMA time series models using squared residual autocorrelations. *Journal of Time Series Analysis* 4, 269–273.

Milhøj, A. (1985) The moment structure of ARCH processes. *Scandinavian Journal of Statistics* 12, 281–292.

Nelson, D. & C.Q. Cao (1992) Inequality constraints in the univariate GARCH model. *Journal of Business and Economic Statistics* 10, 229–235.

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.

Newey, W. & K. West (1987) A simple positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55, 703–708.

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

Priestley, M.B. (1996) Wavelets and time-dependent spectral analysis. *Journal of Time Series Analysis* 17, 85–103.

Ramsey, J., D. Usikov, & G. Zaslavsky (1995) An analysis of U.S. stock price behavior using wavelets. *Fractals* 3, 377–389.

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 in Nonlinear Dynamics and Econometrics* 3, 23–42.

Robinson, P.M. (1991) Testing for strong serial correlation and dynamic conditional heteroskedasticity in multiple regression. *Journal of Econometrics* 47, 67–84.

SenGupta, A. & L. Vermeire (1986) Locally optimal tests for multiparameter hypotheses. *Journal of the American Statistical Association* 81, 819–825.

Silvapulle, M. & P. Silvapulle (1995) A score test against one-sided alternatives. *Journal of the American Statistical Association* 90, 342–349.

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: CRC Press.

Wang, Y. (1995) Jump and sharp cusp detection by wavelets. *Biometrika* 82, 385–397.

Weiss, A. (1984) ARMA models with ARCH errors. *Journal of Time Series Analysis* 5, 129–143.

Wolak, F. (1989) Testing inequality constraints in linear econometric models. *Journal of Econometrics* 41, 205–235.

APPENDIX

To prove Theorems 1 and 2, we first state some useful lemmas.

LEMMA A.1. *Define* $d_J(l) \equiv \sum_{j=0}^{J} \lambda(2\pi l/2^j)$, $l, J \in \mathbb{Z}$, *where* $\lambda(z)$ *is as in (4.1). Then*

(i) $d_J(0) = 0$ *and* $d_J(-l) = d_J(l)$ *for all* $l, J \in \mathbb{Z}, J > 0$;

(ii) $|d_J(l)| \leq C < \infty$ *uniformly in* $l, J \in \mathbb{Z}, J > 0$;

(iii) *For any given* $l \in \mathbb{Z}, l \neq 0$, $d_J(l) \to 1$ *as* $J \to \infty$;
(iv) *For any given* $r \geq 1$, $\sum_{l=1}^{n-1} |d_J(l)|^r = O(2^J)$ *as* $J, n \to \infty$.

Proof of Lemma A.1. See Hong and Lee (2000, proof of Lemma A.1).

LEMMA A.2. *Let* $V_n(J)$ *and* V_0 *be defined as in Theorem 1. Suppose that* $J \to \infty$, $2^J/n \to 0$. *Then* $2^{-J}V_n(J) \to V_0$.

Proof of Lemma A.2. Given the definition of $d_J(l)$, we put

$$\widetilde{V}_n(J) \equiv \sum_{l=1}^{n-1} d_J^2(l) = \sum_{p=-J}^{J} \sum_{j=|p|}^{J} \sum_{l=1}^{n-1} \lambda(2\pi l/2^j)\lambda(2^{|p|}2\pi l/2^j),$$

where the last equality follows by reindexing. We shall show $2^{-J}\widetilde{V}_n(J) \to V_0$, which implies $2^{-J}V_n(J) \to V_0$ by dominated convergence. Put $I \equiv I_n \to \infty$, $I/J \to 0$. We write

$$\widetilde{V}_n(J) = \widetilde{V}_n(I) + Q_{1n} + Q_{2n}, \tag{A.1}$$

where

$$Q_{1n} \equiv \sum_{p=-I}^{I} \sum_{j=I+1}^{J} \sum_{l=1}^{n-1} \lambda(2\pi l/2^j)\lambda(2^{|p|}2\pi l/2^j),$$

$$Q_{2n} \equiv \sum_{|p|=I+1}^{J} \sum_{j=|p|}^{J} \sum_{l=1}^{n-1} \lambda(2\pi l/2^j)\lambda(2^{|p|}2\pi l/2^j).$$

For the second term Q_{1n} in (A.1), we have as $n \to \infty$,

$$\begin{aligned} Q_{1n} &= 2^J \sum_{p=-I}^{I} \sum_{j=I+1}^{J} 2^{-(J-j)} \frac{1}{2\pi} \left[(2\pi/2^j) \sum_{l=1}^{n-1} \lambda(2\pi l/2^j)\lambda(2^{|p|}2\pi l/2^j) \right] \\ &= 2^{J+1} \sum_{p=-\infty}^{\infty} \frac{1}{2\pi} \int_0^\infty \lambda(z)\lambda(2^{|p|}z)dz[1+o(1)] \end{aligned} \tag{A.2}$$

by dominated convergence, $(2\pi/2^j)\sum_{l=1}^{n-1} \lambda(2\pi l/2^j)\lambda(2^{|p|}2\pi l/2^j) \to \int_0^\infty \lambda(z) \times \lambda(2^{|p|}z)dz$ for any given p as $j \to \infty$, and $\sum_{j=I+1}^{J} 2^{-(J-j)} \to 2$ as $I \to \infty, J/I \to \infty$.

Similarly, for the last term in (A.1), we have

$$Q_{2n} = o(2^J). \tag{A.3}$$

Next, for the first term $\widetilde{V}_n(j)$ in (A.1), by the Cauchy–Schwarz inequality, we have

$$\begin{aligned} \widetilde{V}_n(I) &\leq \sum_{p=-I}^{I} 2^{-|p|/2} \sum_{j=|p|}^{I} 2^j \left[2^{-j} \sum_{l=1}^{n-1} \lambda^2(2\pi l/2^j) \right]^{1/2} \left[2^{-(j-|p|)} \sum_{l=1}^{n-1} \lambda^2(2\pi l/2^{j-|p|}) \right]^{1/2} \\ &\leq C^2 \sum_{p=-\infty}^{\infty} 2^{-|p|/2} \sum_{j=0}^{I} 2^j \leq 8C^2 2^I, \end{aligned} \tag{A.4}$$

where we used the fact that for any $l > 0, j > 0$,

$$2^{-j}\sum_{l=1}^{n-1}\lambda^2(2\pi l/2^j) \le 2^{-j}\sum_{l=1}^{2^j} C(2\pi l/2^j)^{2\kappa} + 2^{-j}\sum_{l=2^j+1}^{n-1} C(2\pi l/2^j)^{-2\tau}$$
$$\le C + C\left[2^{-j}\sum_{l=1}^{n}(1+2\pi l/2^j)^{-2\tau}\right]$$
$$\le C\left[1+\frac{1}{2\pi}\int_0^\infty (1+x)^{-2\tau}dx\right],$$

where the first inequality follows by Assumption A.3 and the last one follows from the fact that $(1+x)^{-2\tau}$ is decreasing in $x > 0$. Note that $\int_0^\infty (1+x)^{-2\tau}dx < \infty$ given $\tau > 1$. Collecting (A.1)–(A.4) and $I/J \to 0$, we obtain $2^{-J}\widetilde{V}_n(J) \to 2\sum_{p=-\infty}^{\infty}(2\pi)^{-1} \times \int_0^\infty \lambda(z)\lambda(2^{|p|}z)dz$. Thus, it remains to show $V_0 = 2\sum_{p=-\infty}^{\infty}(2\pi)^{-1}\int_0^\infty \lambda(z)\lambda(2^{|p|}z)dz$, where V_0 is as in (4.5).

From the definitions of $\lambda(\cdot)$ in (4.1) and $\Gamma(z)$ in (4.6), we have $\lambda(z) = 2\pi\hat{\psi}^*(z)\Gamma(z), \lambda(\cdot)$ is real-valued and symmetric about 0, and $\Gamma(\cdot)$ is 2π-periodic. It follows that

$$2(2\pi)^{-1}\int_0^\infty \lambda(z)\lambda(2^{|p|}z)dz$$
$$= (2\pi)^{-1}\int_{-\infty}^\infty \lambda(z)\lambda^*(2^{|p|}z)dz$$
$$= (2\pi)^{-1}\sum_{l\in\mathbb{Z}}\int_0^{2\pi}\lambda(z+2l\pi)\lambda^*[2^{|p|}(z+2l\pi)]dz$$
$$= 2\pi\int_0^{2\pi}\left\{\sum_{l\in\mathbb{Z}}\hat{\psi}^*(z+2l\pi)\hat{\psi}[2^{|p|}(z+2l\pi)]\right\}\Gamma(z)\Gamma^*(2^{|p|}z)dz$$
$$= \delta_{p,0}\int_0^{2\pi}|\Gamma(z)|^2dz = V_0\delta_{p,0},$$

where $\delta_{p,0} = 1$ if $p = 0$ and $\delta_{p,0} = 0$ otherwise, and the fourth equality follows from the orthogonality condition that $2\pi\sum_{l\in\mathbb{Z}}\hat{\psi}^*(z+2l\pi)\hat{\psi}[2^{|p|}(z+2l\pi)] = \delta_{p,0}$ for $z \in \mathbb{R}$ almost everywhere (cf. Hernandez and Weiss, 1996, (1.4) and (1.5), p. 332; note that the Fourier transform $\hat{\psi}(\cdot)$ there differs from our $\hat{\psi}(\cdot)$ by a factor of 2π). Hence, we have $2\sum_{p=-\infty}^{\infty}(2\pi)^{-1}\int_0^\infty \lambda(z)\lambda(2^{|p|}z)dz = V_0$. This completes the proof. ■

LEMMA A.3. *Let $\beta(l)$ be a sequence of autocovariances with $\sum_{l=1}^{\infty}|\beta(l)| < \infty$ and let $d_J(l)$ be defined as in Lemma A.1. Then $\sum_{l=1}^{n-1} d_J(l)\beta(l) \to \sum_{l=1}^{\infty}\beta(l)$ as $J, n \to \infty$.*

Proof of Lemma A.3. We write

$$\sum_{l=1}^{n-1} d_J(l)\beta(l) - \sum_{l=1}^{\infty}\beta(l) = \sum_{l=1}^{n-1}[d_J(l)-1]\beta(l) - \sum_{l=n}^{\infty}\beta(l). \quad \textbf{(A.5)}$$

For the second term in (A.5), we have $|\sum_{l=n}^{\infty}\beta(l)| \le \sum_{l=n}^{\infty}|\beta(l)| \to 0$ given $\sum_{l=1}^{\infty}|\beta(l)| < \infty$. For the first term in (A.5), we have $\sum_{l=1}^{n-1}[d_J(l) - 1]\beta(l) \to 0$ as $J, n \to \infty$ by dominated convergence, $d_J(l) - 1 \to 0$ for any $l \in \mathbb{Z}$ as $J \to \infty$, and $|d_J(l) - 1| \le C$ from Lemma A.1. The desired result follows immediately.

Proof of Theorem 1. Put $u_t \equiv \xi_t^2 - 1$ and $R(l) \equiv \text{Cov}(u_t, u_{t-l})$. Define

$$\tilde{f}(0) \equiv \frac{1}{2\pi} + \sum_{j=0}^{J}\sum_{k=0}^{2^j} \tilde{\alpha}_{jk}\Psi_{jk}(0), \tag{A.6}$$

where $\tilde{\alpha}_{jk} \equiv (2\pi)^{-1/2}\sum_{l=1-n}^{n-1}\tilde{\rho}(j)\hat{\Psi}_{jk}^{*}(l), \tilde{\rho}(j) \equiv \tilde{R}(j)/R(0), \tilde{R}(l) \equiv n^{-1} \times \sum_{t=|l|+1}^{n} u_t u_{t-|l|}$, and $\hat{\Psi}_{jk}(l) \equiv (2\pi)^{-1/2}\int_{-\pi}^{\pi}\Psi_{jk}(\omega)e^{-il\omega}d\omega$ as in (3.11). Note that we have abused the notation $\tilde{\rho}(j)$ by defining $\tilde{\rho}(j) \equiv \tilde{R}(j)/R(0)$ rather than $\tilde{\rho}(j) \equiv \tilde{R}(j)/\tilde{R}(0)$.

Writing $\hat{f}(0) - (2\pi)^{-1} = [\hat{f}(0) - \tilde{f}(0)] + [\tilde{f}(0) - (2\pi)^{-1}]$, we shall prove Theorem 1 by showing Theorems A.1 and A.2, which follow.

THEOREM A.1. $V_n^{-1/2}(J)n^{1/2}[\hat{f}(0) - \tilde{f}(0)] \to^{p} 0$.

THEOREM A.2. $V_n^{-1/2}(J)n^{1/2}\pi[\tilde{f}(0) - (2\pi)^{-1}] \to^{d} N(0,1)$.

Proof of Theorem A.1. Because $\hat{\Psi}_{jk}(\cdot)$ is the Fourier transform of $\Psi_{jk}(\cdot)$, we have

$$\Psi_{jk}(0) = (2\pi)^{-1/2}\sum_{h=-\infty}^{\infty}\hat{\Psi}_{jk}(h) = (2\pi)^{-1/2}\sum_{h=-\infty}^{\infty} e^{-i2\pi hk/2^j}(2\pi/2^j)^{1/2}\hat{\psi}(2\pi h/2^j) \tag{A.7}$$

given (3.11). Moreover, by (3.11) and (3.16), we have

$$\hat{\alpha}_{jk} = (2\pi)^{-1/2}\sum_{l=1-n}^{n-1}\hat{\rho}(l)e^{i2\pi lk/2^j}(2\pi/2^j)^{1/2}\hat{\psi}^{*}(2\pi l/2^j). \tag{A.8}$$

Collecting (3.15), (A.7) and (A.8), and Lemma A.1 yields

$$\begin{aligned}\hat{f}(0) &= \frac{1}{2\pi} + \frac{1}{2\pi} \\ &\quad \times \sum_{l=1-n}^{n-1}\left[\sum_{j=0}^{J}\sum_{h=-\infty}^{\infty}\sum_{k=1}^{2^j} e^{i2\pi(l-h)k/2^j}(2\pi/2^j)\hat{\psi}^{*}(2\pi l/2^j)\hat{\psi}(2\pi h/2^j)\right]\hat{\rho}(l) \\ &= \frac{1}{2\pi} + \frac{1}{2\pi}\sum_{l=1-n}^{n-1} d_J(l)\hat{\rho}(l) = \frac{1}{2\pi} + \frac{1}{\pi}\sum_{l=1}^{n-1} d_J(l)\hat{\rho}(l),\end{aligned} \tag{A.9}$$

where the second equality follows because by the change of variable $l = h + m$, we have

$$\begin{aligned}&\sum_{j=0}^{J}\sum_{h=-\infty}^{\infty}\sum_{k=1}^{2^j} e^{i2\pi(l-h)k/2^j}(2\pi/2^j)\hat{\psi}^{*}(2\pi l/2^j)\hat{\psi}(2\pi h/2^j) \\ &\quad = \sum_{j=0}^{J} 2\pi \sum_{m=-\infty}^{\infty}\left(2^{-j}\sum_{k=1}^{2^j} e^{i2\pi mk/2^j}\right)\hat{\psi}^{*}(2\pi l/2^j)\hat{\psi}[2\pi(l+m)/2^j] \\ &\quad = \sum_{j=0}^{J}\lambda(2\pi l/2^j) = d_J(l),\end{aligned}$$

where we used the fact that $\sum_{k=1}^{2^j} e^{i2\pi mk/2^j} = 2^j$ if $m = 2^j q, q \in \mathbb{Z}$ and $\sum_{k=1}^{2^j} e^{i2\pi mk/2^j} = 0$ otherwise (e.g., Priestley, 1981, (6.19), p. 392). Moreover, the last equality in (A.9) follows from $\hat{\rho}(-l) = \hat{\rho}(l)$, $d_J(0) = 0$, and $d_J(-l) = d_J(l)$ by Lemma A.1.

Similarly, we have

$$\tilde{f}(0) = (2\pi)^{-1} + \pi^{-1} \sum_{l=1}^{n-1} d_J(l)\tilde{\rho}(j). \tag{A.10}$$

Combining (A.9) and (A.10), we obtain

$$\pi[\hat{f}(0) - \tilde{f}(0)] = \sum_{l=1}^{n-1} d_J(l)[\hat{\rho}(l) - \tilde{\rho}(l)]. \tag{A.11}$$

Because $\hat{R}(0) - \tilde{R}(0) = O_P(n^{-1/2})$ given Assumptions A.4 and A.5, it suffices to show

$$V_n^{-1/2}(J) n^{1/2} \sum_{l=1}^{n-1} d_J(l)[\hat{R}(l) - \tilde{R}(l)] \to^p 0. \tag{A.12}$$

We shall show (A.12) for the case $J \to \infty$, where $2^{-J}V_n(J) \to V_0$ by Lemma A.2. The proof for fixed J is similar, with $V_n(J) \to V_0(J)$, where $V_0(J)$ is as in Theorem 1(i).

Put $\hat{\xi}_t \equiv \hat{\varepsilon}_t/\hat{\sigma}$ and recall $u_t \equiv \xi_t^2 - 1$. Straightforward algebra yields

$$\hat{R}(l) - \tilde{R}(l) = \hat{A}_1(l) + \hat{A}_2(l) + \hat{A}_3(l), \tag{A.13}$$

where $\hat{A}_1(l) \equiv n^{-1}\sum_{t=l+1}^n u_t(\hat{\xi}_{t-l}^2 - \xi_{t-l}^2), \hat{A}_2(l) \equiv n^{-1}\sum_{t=l+1}^n (\hat{\xi}_t^2 - \xi_t^2)u_{t-l}$, and $\hat{A}_3(l) \equiv n^{-1}\sum_{t=l+1}^n (\hat{\xi}_t^2 - \xi_t^2)(\hat{\xi}_{t-l}^2 - \hat{\xi}_{t-l}^2)$. We first consider $\hat{A}_1(l)$ in (A.13). Noting $\xi_t = \varepsilon_t/\sigma_0$ under $\mathbb{H}_0$, where $\sigma_0^2 \equiv E(\varepsilon_t^2)$, we have

$$\begin{aligned}\hat{A}_1(l) &= \hat{\sigma}^{-2} n^{-1} \sum_{t=l+1}^n u_t(\hat{\varepsilon}_{t-l}^2 - \varepsilon_{t-l}^2) + (\hat{\sigma}^{-2} - \sigma_0^{-2}) n^{-1} \sum_{t=l+1}^n u_t \varepsilon_{t-l}^2 \\ &= \hat{\sigma}^{-2}\hat{A}_{11}(l) + 2\hat{\sigma}^{-2}\hat{A}_{12}(l) + (\hat{\sigma}^{-2} - \sigma_0^{-2})\hat{A}_{13}(l),\end{aligned}$$

where $\hat{A}_{11}(l) \equiv n^{-1}\sum_{t=l+1}^n u_t(\hat{\varepsilon}_{t-l} - \varepsilon_{t-l})^2, \hat{A}_{12}(l) \equiv n^{-1}\sum_{t=l+1}^n u_t\varepsilon_{t-l}(\hat{\varepsilon}_{t-l} - \varepsilon_{t-l})$, and $\hat{A}_{13}(l) \equiv n^{-1}\sum_{t=l+1}^n u_t\varepsilon_{t-l}^2$. By the Cauchy–Schwarz inequality, the mean value theorem, Assumptions A.1, A.4, and A.5, and Lemma A.1(iv), we have

$$\begin{aligned}\left|\sum_{l=1}^{n-1} d_J(l)\hat{A}_{11}(l)\right| &\le \|\hat{b} - b_0\|^2 \left[\sum_{l=1}^{n-1} |d_J(l)|\right] \left(n^{-1}\sum_{t=1}^n u_t^2\right)^{1/2} \\ &\quad \times \left[n^{-1}\sum_{t=1}^n \sup_{b\in\mathbb{B}_0} \left\|\frac{\partial}{\partial b} g(X_t, b)\right\|^4\right]^{1/2} \\ &= O_P(2^J/n).\end{aligned} \tag{A.14}$$

Next, by a second-order Taylor expansion and Assumptions A.1, A.4, and A.5, we have

$$\begin{aligned}\left|\sum_{l=1}^{n-1} d_J(l)\hat{A}_{12}(l)\right| &\leq \|\hat{b}-b_0\| \sum_{l=1}^{n-1} |d_J(l)| \left\| n^{-1}\sum_{t=1}^{n} u_t \varepsilon_{t-l} \frac{\partial}{\partial b} g(X_t, b_0)\right\| \\ &\quad + \frac{1}{2}\|\hat{b}-b_0\|^2 \sum_{l=1}^{n-1} |d_J(l)| \\ &\quad \times \left[n^{-1}\sum_{t=1}^{n} |u_t \varepsilon_{t-l}| \sup_{b\in\mathbb{B}_0} \left\| \frac{\partial^2}{\partial b \partial b'} g(X_t, b)\right\| \right] \\ &= O_P(2^J/n) \end{aligned} \qquad \textbf{(A.15)}$$

by Markov's inequality and Lemma A.1(iv), where we used $E\|n^{-1}\sum_{t=1}^{n} u_t \varepsilon_{t-l}(\partial/\partial b) g(X_t, b_0)\| = O(n^{-1})$ given $E(u_t|\mathcal{I}_{t-1}) = 0$ a.s. and $X_t \in \mathcal{I}_{t-1}$ as in Assumption A.4. Finally, we have

$$\left|\sum_{l=1}^{n-1} d_J(l)\hat{A}_{13}(l)\right| = O(2^J/n^{1/2}) \qquad \textbf{(A.16)}$$

by Markov's inequality and $\sup_{0<l<n} E\hat{A}_{13}^2(l) = O(n^{-1})$, which follows from $E(u_t|\mathcal{I}_{t-1}) = 0$ a.s. and Assumption A.1. Combining (A.14)–(A.16) and $\hat{\sigma}^2 - \sigma_0^2 = O_P(n^{-1/2})$, we obtain

$$\sum_{l=1}^{n-1} d_J(l)\hat{A}_1(l) = O_P(2^J/n). \qquad \textbf{(A.17)}$$

Similarly, for the second term $\hat{A}_2(l)$ in (A.13), we have

$$\sum_{l=1}^{n-1} d_J(l)\hat{A}_2(l) = O_P(2^J/n). \qquad \textbf{(A.18)}$$

Finally, we consider the last term $\hat{A}_3(l)$ in (A.13). As shown in Hong (1997, p. 272), $\sup_{0<l<n}|\hat{A}_3(l)| \leq n^{-1}\sum_{t=1}^{n}(\hat{\xi}_t^2 - \xi_t^2)^2 = O_P(n^{-1})$. This and Lemma A.1(iv) imply

$$\left|\sum_{l=1}^{n-1} d_J(l)\hat{A}_3(l)\right| \leq \sup_{0<l<n} |\hat{A}_3(l)| \left[\sum_{l=1}^{n-1} |d_J(l)|\right] = O_P(2^J/n). \qquad \textbf{(A.19)}$$

Combining (A.17)–(A.19) and $2^{-J}V_n(J) \to V_0$ yields $V_n^{-1/2} n^{1/2} \sum_{l=1}^{n-1} d_J(l)[\hat{R}(l) - \tilde{R}(l)] \to^p 0$ given $2^J/n \to 0$. This completes the proof for (A.12) and thus for Theorem A.1.

Proof of Theorem A.2. Put $\hat{W} \equiv \sum_{l=1}^{n-1} d_J(l)\tilde{R}(l)/R(0)$. By (A.10), we have

$$\pi[\tilde{f}(0) - (2\pi)^{-1}] = \hat{W} + [R(0)/\tilde{R}(0) - 1]\hat{W} = \hat{W} + o_P(\hat{W}) \qquad \textbf{(A.20)}$$

given $\tilde{R}(0) - R(0) = O_P(n^{-1/2})$ by Assumption A.1 and $\mathbb{H}_0$. Write

$$\hat{W} = n^{-1}\sum_{t=2}^{n} W_t, \qquad \textbf{(A.21)}$$

where $W_t \equiv R^{-1}(0)u_t \sum_{l=1}^{t-1} d_J(l)u_{t-l}$. Observe that $\{W_t, \mathcal{F}_{t-1}\}$ is an adapted martingale difference sequence, where $\mathcal{F}_t$ is the sigma field consisting of all $u_s, s \leq t$. Thus, from (A.21), we obtain

$$\mathrm{Var}(n^{1/2}\hat{W}) = n^{-1}\sum_{t=2}^{n} EW_t^2 = \sum_{t=2}^{n}\sum_{l=1}^{t-1} d_J^2(l) = \sum_{l=1}^{n}(1-l/n)d_J^2(l) = V_n(J). \qquad \textbf{(A.22)}$$

By Hall and Heyde's (1980, pp. 10–11) martingale theorem, $V_n^{-1/2}(J)n^{1/2}\hat{W} \to^d N(0,1)$ if

$$V_n^{-1}(J)n^{-1}\sum_{t=2}^{n} E\{W_t^2 \mathbf{1}[|W_t| > \eta n^{1/2}V_n^{1/2}(J)]\} \to 0 \quad \text{for any } \eta > 0, \qquad \textbf{(A.23)}$$

$$V_n^{-1}(J)n^{-1}\sum_{t=2}^{n} E(W_t^2|\mathcal{F}_{t-1}) \to^p 1. \qquad \textbf{(A.24)}$$

For space, we shall show the central limit theorem for $\hat{W}$ for large J (i.e., $J \to \infty$). The proof for fixed J is similar and simpler because $d_J(l)$ is finite and summable.

Given (A.22) and Lemma A.2, we shall verify condition (A.23) by showing $2^{-2J}n^{-2}\sum_{t=1}^{n} EW_t^4 \to 0$. Put $\mu_4 \equiv E(u_t^4)$. By Assumption A.1, we have

$$\begin{aligned} EW_t^4 &= \mu_4 R^{-4}(0)E\left[\sum_{l=1}^{t-1} d_J(l)u_{t-l}\right]^4 \\ &= \mu_4^2 R^{-4}(0)\sum_{l=1}^{t-1} d_J^4(l) + 6\mu_4 R^{-2}(0)\sum_{l=2}^{t-1}\sum_{h=1}^{l-1} d_J^2(l)d_J^2(h) \\ &\leq 3\mu_4^2 R^{-4}(0)\left[\sum_{l=1}^{n-1} d_J^2(l)\right]^2. \end{aligned}$$

It follows from Lemma A.2 that $2^{-2J}n^{-2}\sum_{t=1}^{n} EW_t^4 = O(n^{-1})$, ensuring (A.23).

Next, given Lemma A.2, it suffices for (A.24) if $2^{-2J}\mathrm{Var}[n^{-1}\sum_{t=2}^{n} E(W_t^2|\mathcal{F}_{t-1})] \to 0$. By the definition of W_t, we have

$$\begin{aligned} E(W_t^2|\mathcal{F}_{t-1}) &= R^{-1}(0)\left[\sum_{l=1}^{t-1} d_J(l)u_{t-l}\right]^2 \\ &= EW_t^2 + R^{-1}(0)\sum_{l=1}^{t-1} d_J(l)[u_{t-l}^2 - R(0)] \\ &\quad + 2R^{-1}(0)\sum_{l=2}^{t-1}\sum_{h=1}^{l-1} d_J(l)d_J(h)u_{t-l}u_{t-h} \\ &\equiv EW_t^2 + R^{-1}(0)A_t + 2R^{-1}(0)B_t, \text{ say.} \end{aligned}$$

It follows that

$$\begin{aligned} n^{-1}\sum_{t=2}^{n}[E(W_t^2|\mathcal{F}_{t-1}) - EW_t^2] &= R^{-1}(0)n^{-1}\sum_{t=2}^{n} A_t + 2R^{-1}(0)n^{-1}\sum_{t=2}^{n} B_t \\ &\equiv R^{-1}(0)\hat{A} + 2R^{-1}(0)\hat{B}, \text{ say.} \end{aligned} \qquad \textbf{(A.25)}$$

Whence, it suffices to show $2^{-2J}[\mathrm{Var}(\hat{A}) + \mathrm{Var}(\hat{B})] \to 0$. First, noting that A_t is a weighted sum of independent zero-mean variables $\{u_{t-j}^2 - R(0)\}$, we have $EA_t^2 = [\mu_4 - R^2(0)]\sum_{l=1}^{t-1} d_J^4(l)$. It follows by Minkowski's inequality and Lemma A.1(iv) that

$$E\hat{A}^2 \le \left[n^{-1}\sum_{t=2}^{n}(EA_t^2)^{1/2}\right]^2 \le [\mu_4 - R^2(0)]\left[\sum_{l=1}^{n-1} d_J^4(l)\right] = O(2^J). \quad \textbf{(A.26)}$$

Next, we consider $\mathrm{Var}(\hat{B})$. For all $t \ge s$, we have

$$EB_tB_s = R^2(0)\sum_{l_2=2}^{t-1}\sum_{h_2=1}^{l_2-1}\sum_{l_1=2}^{s-1}\sum_{h_1=1}^{l_1-1} d_J(l_1)d_2(h_1)d_J(l_2)d_J(h_2)\delta_{t-h_1,s-h_2}\delta_{t-l_1,s-l_2}$$

$$= R^2(0)\sum_{l=2}^{t-1}\sum_{h=1}^{l-1} d_J(t-s+l)d_J(t-s+h)d_J(l)d_J(h),$$

where, as before, $\delta_{j,h} = 1$ if $h = j$ and $\delta_{j,h} = 0$ otherwise. It follows that

$$E\hat{B}^2 \le 2n^{-2}\sum_{t=3}^{n}\sum_{s=2}^{t} EB_tB_s \le 2R^2(0)n^{-1}\sum_{\tau=0}^{n-1}\sum_{l=2}^{n-1}\sum_{h=1}^{l-1}|d_J(\tau+l)d_J(\tau+h)d_J(l)d_J(h)|$$

$$\le 2R^2(0)n^{-1}\left[\sum_{\tau=0}^{n-1} d_J^2(\tau)\right]\left[\sum_{l=1}^{n-1}|d_J(l)|\right]^2 = O(2^{3J}/n) \quad \textbf{(A.27)}$$

by Lemma A.1(iv). Combining (A.25)–(A.27) yields $2^{-2J}[\mathrm{Var}(\hat{A}) + \mathrm{Var}(\hat{B})] = O(2^{-J} + 2^J/n) \to 0$ given $J \to \infty$, $2^J/n \to 0$. Thus, condition (A.24) holds. By Hall and Heyde (1980, pp. 10–11), $V_n^{-1/2}(J)n^{1/2}\hat{W} \to^d N(0,1)$. This completes the proof.

Proof of Theorem 2. Put $\dot{R}(l) \equiv \sum_{t=l+1}^{n}(\varepsilon_t^2/\sigma_0^2 - 1)(\varepsilon_{t-l}^2/\sigma_0^2 - 1)$, where $\sigma_0^2 = E(\varepsilon_t^2)$ under $\mathbb{H}_n(a_n)$. Because $\varepsilon_t/\sigma_0 \ne \xi_t$ under $\mathbb{H}_n(a_n)$, we have $\dot{R}(l) \ne \tilde{R}(l)$, where, as before, $\tilde{R}(l) \equiv n^{-1}\sum_{t=|l|+1}^{n} u_tu_{t-|l|}$ and $u_t \equiv \xi_t^2 - 1$ as before. Under $\mathbb{H}_n(a_n)$, we have

$$\varepsilon_t^2/\sigma_0^2 = \xi_t^2\left[1 + a_n\sum_{l=1}^{\infty}\beta_l(\xi_{t-l}^2 - 1)\right]. \quad \textbf{(A.28)}$$

We now define

$$\dot{f}(0) \equiv \sum_{j=0}^{2^j}\sum_{k=1}^{2^j}\dot{\alpha}_{jk}\Psi_{jk}(0),$$

where $\dot{\alpha}_{jk} \equiv (2\pi)^{-1/2}\sum_{l=1-n}^{n-1}\dot{\rho}(l)\hat{\Psi}_{jk}^*(l)$ and $\dot{\rho}(l) \equiv \dot{R}(l)/R(0)$. Write $\hat{f}(0) - (2\pi)^{-1} = \hat{f}(0) - \dot{f}(0) + \dot{f}(0) - (2\pi)^{-1}$. The proof of Theorem 2 consists of Theorems A.3 and A.4, which follow.

THEOREM A.3. $V_n^{-1/2}(J)n^{1/2}[\hat{\dot f}(0) - \dot f(0)] \to^p 0$.

THEOREM A.4. $V_n^{-1/2}(J)n^{1/2}\pi[\dot{f}(0) - (2\pi)^{-1}] \to^d N(\bar\mu,1)$, *where* $\bar\mu \equiv \mu(J)$ *under Theorem 2(i) and* $\bar\mu \equiv \mu$ *under Theorem 2(ii).*

Proof of Theorem A.3. The proof is analogous to that of Theorem A.1 with the more restrictive condition $J \to \infty$, $2^{2J}/n \to 0$. We omit it here for the sake of space.

Proof of Theorem A.4. We shall only show results for the case $J \to \infty$. Because $\pi[\dot f(0) - (2\pi)^{-1}] = \sum_{l=1}^{n-1} d_J(l)\dot R(l)$, it suffices to show

$$V_n^{-1/2}(J)n^{1/2}\sum_{l=1}^{n-1} d_J(l)\dot R(l)/R(0) \to^d N(\mu,1). \tag{A.29}$$

Recall that $u_t \equiv \xi_t^2 - 1$ and put $V_t \equiv \xi_t^2\sum_{j=1}^\infty \beta_j u_{t-j}$. By (A.28) and (A.29) and $\mathbb{H}_n(a_n)$, we have

$$\begin{aligned}\dot R(l) &= n^{-1}\sum_{t=l+1}^n (\xi_t^2 h_t/\sigma_0^2 - 1)(\xi_{t-l}^2 h_{t-l}/\sigma_0^2 - 1)\\ &= n^{-1}\sum_{t=l+1}^n (u_t + a_n V_t)(u_{t-l} + a_n V_{t-l})\\ &= n^{-1}\sum_{t=l+1}^n u_t u_{t-l} + a_n n^{-1}\sum_{t=l+1}^n V_t u_{t-l} + a_n n^{-1}\sum_{t=l+1}^n u_t V_{t-l} + a_n^2 n^{-1}\sum_{t=l+1}^n V_t V_{t-l}\\ &\equiv \tilde R(l) + a_n\hat A_4(l) + a_n \hat A_5(l) + a_n^2\hat A_6(l), \text{ say,}\end{aligned} \tag{A.30}$$

where $\tilde R(l) \equiv n^{-1}\sum_{t=|l|+1} u_t u_{t-|l|}$ as before. Now, put $V_t(l) \equiv \xi_t^2\sum_{j=1,j\neq l}^\infty \beta_j u_{t-j}$. For the second term $\hat A_4(l)$ in (A.30), we have

$$\begin{aligned}\sum_{l=1}^{n-1} d_J(l)\hat A_4(l) &= \sum_{l=1}^{n-1} d_J(l)\left(n^{-1}\sum_{t=l+1}^n \xi_t^2\sum_{j=1}^\infty \beta_j u_{t-j}u_{t-l}\right)\\ &= R(0)\sum_{l=1}^{n-1} d_J(l)(1 - l/n)\beta_l + \sum_{l=1}^{n-1} d_J(l)\beta_l\left\{n^{-1}\sum_{t=l+1}^n \xi_t^2[u_{t-l}^2 - R(0)]\right\}\\ &\quad + \sum_{l=1}^{n-1} d_J(l)\beta_l\left[n^{-1}\sum_{t=l+1}^n V_t(l)u_{t-l}\right]\\ &= R(0)\sum_{l=1}^\infty \beta_l + O_P(2^{J/2}/n^{1/2}),\end{aligned} \tag{A.31}$$

where $\sum_{l=1}^{n-1} d_J(l)(1 - l/n)\beta_l \to \sum_{l=1}^\infty \beta_l < \infty$ as $J \to \infty$ by Lemma A.3 and dominated convergence, and $|\sum_{l=1}^n d_J(l)\beta_l\{n^{-1}\sum_{t=l+1}^n \xi_t^2[u_{t-l}^2 - R(0)]\}| = O_P(2^{J/2}/n^{1/2})$ by the Cauchy–Schwarz inequality, Lemma A.1(iv), $\sum_{l=1}^\infty \beta_l^2 < \infty$, and $E|n^{-1}\sum_{t=l+1}^n \xi_t^2[u_{t-l}^2 - R(0)]|^2 \le Cn^{-1}$ given Assumption A.1. In addition, we also used the fact that $\sum_{l=1}^n d_J(l)\beta_l[n^{-1}\sum_{t=l+1}^n V_t(l)u_{t-l}] = O_P(2^{J/2}/n^{1/2})$ given independence between $V_t(l)$ and u_{t-l}.

Similarly, for the third term $\hat{A}_5(l)$ in (A.30), we have

$$\sum_{l=1}^{n-1} d_J(l)\hat{A}_5(l) = O_P(2^{J/2}/n^{1/2}) \tag{A.32}$$

by the Cauchy–Schwarz inequality, Lemma A.1(iv), and $E\hat{A}_5^2(l) \leq Cn^{-1}$ given independence between u_t and V_{t-l} for $l > 0$.

Finally, for the last term $\hat{A}_6(l)$ in (A.30), we put $R_V(l) \equiv \text{Cov}(V_t, V_{t-l})$. Then

$$\sum_{l=1}^{n-1} d_J(l)\hat{A}_6(l) = \sum_{l=1}^{n-1} d_J(l)R_V(l) + \sum_{l=1}^{n-1} d_J(l)[\hat{A}_6(l) - R_V(l)].$$

Because $V_t = \sum_{j=1}^{\infty} \beta_j(\xi_t^2 u_{t-j})$ is a linear process with $\sum_{j=1}^{\infty}|\beta_j| < \infty$ and $E(\xi_t^2 u_{t-j})^4 < \infty$, we have $\sum_{j=-\infty}^{\infty}\sum_{m=-\infty}^{\infty}\sum_{l=-\infty}^{\infty}|k(0,j,m,l)| < \infty$, where $k(0,j,m,l)$ is the fourth order cumulant of $\{V_t, V_{t+j}, V_{t+m}, V_{t+l}\}$ (see, e.g., Hannan 1970, p. 211). It follows that $\sup_{0<l<n}\text{Var}[\hat{A}_6(l)] \leq Cn^{-1}$ by Hannan (1970, (5.1)). Consequently, we obtain $|\sum_{l=1}^{n-1} d_J(l)[\hat{A}_6(l) - R_V(l)]| \leq \sum_{l=1}^{n-1}|d_J(l)||\hat{A}_6(l) - R_V(l)| = O_P(2^J/n^{1/2})$ by Markov's inequality and Lemma A.1(iv). On the other hand, because $R_V(l)$ is absolutely summable (i.e., $\sum_{l=-\infty}^{\infty}|R_V(l)| < \infty$), it follows from Lemma A.3 that $\sum_{l=1}^{n-1} d_J(l)R_V(l) \to \sum_{l=1}^{\infty} R_V(l) < \infty$ as $J \to \infty$. Therefore, we have

$$\sum_{l=1}^{n-1} d_J(l)\hat{A}_6 = O_P(1) \tag{A.33}$$

given $2^{2J}/n \to 0$. Combining (A.30)–(A.33), $a_n = 2^{J/2}/n^{1/2}$, and $2^{2J}/n \to 0$ yields

$$\sum_{l=1}^{n-1} d_J(l)\dot{R}(l)/R(0) = \sum_{l=1}^{n-1} d_J(l)\tilde{R}(l)/R(0) + (2^{J/2}/n^{1/2})\sum_{l=1}^{\infty} \beta_l + o_P(2^{J/2}/n^{1/2}).$$

Consequently, we have (A.29) by Theorem A.2. This completes the proof. ■