One-Sided Testing for ARCH E¤ect Using Wavelets

Yongmiao Hong and Jin Lee

Cornell University, Ithaca, U.S.A.

November 1999

Yongmiao Hong, Department of Economics and Department of Statistical Science, Cornell University, Uris Hall, Ithaca, NY 14853, U.S.A.; E-mail: yh20@cornell.edu; Jin Lee, Department of Economics, Cornell University, Uris Hall, Ithaca, NY 14853, U.S.A.; E-mail: jl79@cornell.edu.
This paper is based on part of Lee's dissertation.

ABSTRACT

There has been an increasing interest 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 using the wavelet method, a new analytic tool developed in the last decade or so. The test is based on a wavelet spectral density estimator at frequency zero of the square of estimated residuals from a regression model. 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 there exists persistent ARCH, or when ARCH effect is small at each lag but carries over a long distributional lag. Because wavelets can effectively capture spectral peaks, we expect that the wavelet test is more powerful than the kernel counterpart when there exists persistent ARCH or when ARCH effect has a long distributional lag. This is confirmed in a simulation study, which also compares a number of important one-sided and two-sided ARCH tests.

Key words: ARCH, one-sided hypothesis, time series, spectral analysis, wavelet

1. INTRODUCTION

Hypothesis testing with inequality restrictions has been important in econometrics and statistics (e.g., Andrews 1998, Bera et al. 1998, Gourieroux et al. 1982, King and Wu 1998, Self and Liang 1987, SenGupta and Vermeire 1986, Silvapulle and Silvapulle 1995, Wolak 1989, Wu and King 1994). An important example in time series econometrics is hypotheses on ARCH, where parameters of interest are zero if there is no ARCH and are nonnegative if ARCH exists.

Detection of ARCH is important from both theoretic and practical points of view. Neglecting ARCH may lead to arbitrarily large loss in asymptotic efficiency of parameter estimation (e.g., Engle, 1982); cause overrejection of conventional tests for serial correlation such as those of Box and Pierce (1970) and Ljung and Box (1978) (e.g., Taylor 1984, Milhøj 1985, Diebold 1987); and result in overparameterization of ARMA models (e.g., Weiss 1984). Although the one-sided nature of ARCH has been long well-known, most ARCH tests are two-sided. Among them are Engle (1982), McLeod and Li (1983), Bera and Higgins (1992), Gregory (1989), Hong and Shehadeh (1999), Lee (1991), and Weiss (1986). Brock et al.'s (1991,1996) chaotic correlation dimension test for serial dependence also has excellent power against ARCH.

Exploration of the one-sided nature of ARCH is expected to increase power in small samples. Engle et al. (1985) suggest using the square root of the Lagrangian Multiplier (LM) test, with proper sign, to test first order ARCH. This approach, however, could not be generalized to test higher order ARCH. Lee and King (1993,1994) are apparently the first to develop one-sided tests for ARCH of general order q. They propose a locally most mean powerful score-based test for ARCH(q), using SenGupta and Vermeire's (1986) approach for one-sided multiparameter hypotheses. Demos and Sentana (1998) consider a convenient one-sided LM test for ARCH(q) in spirit similar to Kuhn-Tucker Multiplier tests (cf. Gourieroux et al. 1982). Lee and King (1993) and Demos and Sentana (1998) also consider one-sided tests for GARCH(1,1), which are numerically identical to their tests for ARCH(1) respectively. Andrews (1999) also considers one-sided testing for GARCH(1,1). Simulation studies show that these tests outperform two-sided tests (e.g., Engle 1982), indicating nontrivial gains of exploring the one-sided nature of ARCH.

Hong (1997) recently proposed a one-sided ARCH test by observing that the spectral density of the square of an estimated residual from a regression model is uniform when there is no ARCH and is always larger than the uniform one at frequency zero whenever

ARCH exists. Hong (1997) uses Parzen's (1957) kernel estimator to construct the test. The test is shown to perform well in comparison with some popular one-sided and two-sided ARCH tests, and it requires no formulation of an alternative model (e.g., the orders of ARCH or GARCH processes).

It is well-known that in finite samples the kernel method tends to underestimate the spectral density at frequencies where there is a mode, no matter whether a finite sample optimal bandwidth is available (cf. Priestley 1981). The kernel method is not an ideal tool in capturing significantly inhomogeneous 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, always resulting in a spectral mode at frequency zero. In particular, the spectral density of the squared process exhibits a peak at frequency zero when there exists persistent ARCH, or when ARCH effect carries over a long distributional lag, although it may be small at each individual lag. Examples are nearly integrated GARCH processes, and fractionally integrated GARCH processes (cf. Baillie 1986). In such situations, the kernel method cannot be expected to perform well.

The recent development of wavelet analysis provides a tool to construct a potentially more powerful one-sided test for ARCH. Wavelet analysis is a new analytic tool developed over the last decade or so. It is a spatially adaptive analytic tool that can efficiently capture significantly inhomogeneous features (e.g., Donoho and Johnstone 1994,1995a,1995b, Donoho et al. 1996, Gao 1993, Neumann 1996, Wang 1995). In this paper, we propose a one-sided test for ARCH using a wavelet spectral density at frequency zero of the square of estimated residuals from a regression model. Because of the nature of ARCH, the wavelet method is expected to be more powerful than the kernel method where there exists persistent ARCH. Besides the ARCH context, spectral peaks may arise due to strong dependence, seasonality, and business cycles. Therefore, our approach might have potential applications to testing a broad range of one-sided hypotheses. The present paper merely provides an example to illustrate how wavelets can be used to develop powerful econometric procedures.

Wavelets have been applied to time series analysis in several directions. Gao (1993) uses the wavelet method to estimate the spectral density of a stationary Gaussian time series. Neumann (1996) considers wavelet estimation of the spectral density of a stationary non-Gaussian process. Priestley (1996) explores potential applications of wavelet analysis to nonstationary time series evolutionary spectral analysis. See also Subba Rao. In econometrics, Gilbert (1995) uses the wavelet method to estimate and test structural changes. Jensen (1996) uses wavelets to estimate a long memory model via maximum likelihood. There have been also some applications of wavelet methods to economic and financial time series (e.g., Goffe 1994, Ramsey 1998, Ramsey and Lampart 1998a,1998b, Ramsey and Zhang 1996,1997, Ramsey et al. 1995).

We first describe the basic framework and hypotheses of interest in Section 2. Section 3 is an introduction to wavelet analysis and especially its application to spectral analysis. In Section 4, we propose a test based on a wavelet spectral density estimator, and derive its asymptotic distribution. An asymptotic local power analysis is given in Section 5. In Section 6, we adapt the proposed test to data-dependent choice of finest scale parameter—the smoothing parameter in the wavelet estimation. Section 7 presents a Monte Carlo comparison between the proposed wavelet test, three existing one-sided ARCH tests, and Engle's (1982) popular two-sided LM test. Section 8 concludes the paper. All proofs are collected in the appendix. Unless indicated, all convergencies are taken as the sample size $n \to \infty$; A^* denotes the complex conjugate of A; $||A|| = \{tr(A'A)\}^{\frac{1}{2}}$ the Euclidean norm of A; C a generic bounded constant that may differ from place to place; and $\mathbb{Z} = \{0, \pm 1, \pm 2, \ldots\}$ the set of integers.

2. FRAMEWORK AND HYPOTHESES

Throughout, we consider the following data generating process:

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

$$Y_t = g(X_t; \beta_0) + \varepsilon_t; \quad \varepsilon_t = \xi_t h_t^{\frac{1}{2}}; \tag{2.1}$$

where X_t is a vector consisting of exogenous variables and lagged dependent variables, β_0 is a finite-dimensional parameter vector, and h_t is a positive time-varying measurable function with respect to the information set I_{t-1} 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 is a setup often seen in the ARCH literature (e.g., Bollerslev et al. 1992). We make no distributional assumption on innovation ξ_t except the existence of an eighth moment. The process $\{\varepsilon_t\}$ is an adapted martingale difference sequence with respect to I_{t-1}; namely $E(\varepsilon_t | I_{t-1}) = 0$ almost surely. Its conditional variance, $E(\varepsilon_t^2 | I_{t-1}) = h_t$; 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 $\beta_0 > 0$; $\sum_{l=1}^{\infty} < 1$; and $\beta_l \geq 0$ for all $l \geq 1$ to ensure positivity of h_t (cf. Nelson and Cao 1992, Drost and Nijman 1993). 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 (1987) GARCH(p, q)

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

whose coefficient β_l, which is a function of $\{\alpha_l, \gamma_l\}$, decays to zero exponentially as $l \to \infty$. The class (2.2) also includes Baillie et al.'s (1996) fractionally integrated GARCH process. For this process, β_j decays to zero slowly.

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

$$H_0 : \beta_j = 0 \qquad \text{for all } j = 1, 2, \ldots .$$

The alternative hypothesis that ARCH exists is

$$H_A : \beta_j \geq 0 \qquad \text{for all } j = 1, 2, \ldots, \text{ with at least one strict inequality.}$$

The alternative H_A is one-sided. To test such a hypothesis, we take a frequency domain approach. Let $f(\omega)$ be the standardized spectral density of ε_t^2, that is,

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

where $\rho(j)$ is the autocorrelation function of $\{\varepsilon_t^2\}$. Because (2.2) implies that $\{\varepsilon_t^2\}$ follows an AR(∞) process:

$$\varepsilon_t^2 = \beta_0 + \sum_{j=1}^{\infty} \beta_j \varepsilon_{t-j}^2 + w_t; \tag{2.6}$$

with $E(w_t | I_{t-1}) = 0$ almost surely. Under H_0, $\{\varepsilon_t^2\} = w_t$ is a white noise, we have $f(0) = (2\pi)^{-1}$. On the other hand, under H_A, we have $\rho(j) \geq 0$ for all $j \in \mathbb{Z}$ and there exists at least one j such that $\rho(j) > 0$. It follows that $f(0) > (2\pi)^{-1}$ under H_A.

This forms a basis for constructing a one-sided test for H_0 vs. H_A. We can compare a consistent estimator for $f(0)$ and $(2\pi)^{-1}$ and test if their difference is significantly larger than zero. Note that we do not specify any particular alternative model (e.g., the orders of GARCH(p, q)) under H_A; the proposed test will be consistent (i.e., has asymptotic unit power) against the class of general linear ARCH processes, which include ARCH, GARCH and fractionally integrated GARCH with known or unknown orders.

Hong (1997) proposes a consistent one-sided ARCH test using a Parzen's (1957) kernel estimator for $f(0)$. While the kernel estimator is consistent, it tends to underestimate $f(0)$ when there is a spectral mode at frequency zero (e.g., Priestley 1981). This is indeed the case under H_A, which implies that the autocorrelations of $\{\varepsilon_t^2\}$ are positive at all lags and consequently result in a spectral mode at frequency zero. In particular, when the β_j are small but decay to zero slowly, there is a spectral peak at frequency zero. This is the case with highly persistent volatility clustering. For such cases, the kernel method may not be expected to be most powerful.

3. WAVELET METHOD

The recent development of wavelet analysis provides a potentially useful tool to test ARCH. Wavelet analysis is a new mathematical tool. It can effectively estimate inhomogeneous spectral density functions (e.g., Gao 1993, Neumann 1996). We now propose a wavelet estimator for $f(0)$, the standardized spectral density at frequency zero of $\{\varepsilon_t^2\}$, and use it to construct a one-sided test for ARCH.

Throughout, we use multiresolution analysis (MRA), introduced by Mallat (1989). MRA is a mathematical method to describe a square-integrable function $g(\cdot) \in L_2(\mathbb{R})$ at different scales. The key of MRA is the introduction of the mother wavelet function ψ.

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

The orthonormality of ψ implies that the doubly infinite sequence $\{\psi_{jk}\}$ 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 ψ by dilations and translations. The integers j and k are called the dilation and translation parameters respectively. Intuitively, j localizes analysis in frequency and k localizes analysis in time (or space). This simultaneous time-frequency localization of information is the key feature of wavelet analysis, explaining why wavelets are attractive for function approximation. The dilation factor can differ from 2, but "2" ensures the L_2-invariance that $\int_{-\infty}^{\infty} \psi_{jk}(x)^2 dx = \int_{-\infty}^{\infty} \psi(x)^2 dx.$ Often $\psi(x)$ is well-localized (i.e., $\psi(x) \to 0$ sufficiently fast as $x \to \infty$), so $\psi_{jk}(x)$ is effectively nonzero only around an interval of width 2^{-j} centered at $k/2^j.$

The mother wavelet ψ can have bounded support. An example is Haar wavelet:

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

Compact support ensures that ψ 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 ψ can also have infinite support, but it must decay to zero sufficiently fast at infinity. An example is the Littlewood -Paley wavelet $\psi(\cdot)$, which is defined via its Fourier transform

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

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

For any $g(x) \in L_2(\mathbb{R}),$ we have the wavelet representation

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

where the wavelet coefficient

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

Cf. Mallat (1989) and Daubechies (1992). The localization property of ψ ensures that α_{jk} basically depends on the local property of g on an interval of width 2^{-j} centered at $k/2^j.$ This is fundamentally different from Fourier representation, where each Fourier coefficient depends on the global property of $g.$ An essential feature of wavelet analysis is that wavelets, in an "automatic manner", evaluate high frequency components of g on small intervals, and low frequency components of g on large intervals. Consequently, they can effectively represent significantly inhomogeneous functions with a relatively small number of wavelet coefficients. Wavelet coefficients are large where g exhibits significant inhomogeneity, and are small where g is smooth.

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

$$\Psi_{jk}(\omega) = (2\pi)^{-\frac{1}{2}} \sum_{m=-\infty}^{\infty} \psi_{jk}\left(\frac{\omega}{2\pi} + m\right), \tag{3.6}$$

which is 2π-periodic. With such periodic orthonormal bases for $L_2(I)$, where $I = [-\pi, \pi]$, we can represent $f(\omega)$ via wavelet bases:

$$f(\omega) = \frac{1}{2\pi} + \sum_{j=0}^{\infty} \sum_{k=1}^{2^j} \alpha_{jk} \Psi_{jk}(\omega), \tag{3.7}$$

where the wavelet coefficient

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

See Lee and Hong (1998) and Hong and Lee (1999). Denote the Fourier transform of $\psi(x)$ by

$$\hat{\psi}(z) = (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{\infty} \psi(x) e^{-izx} dx. \tag{3.9}$$

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

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

where $\hat{\Psi}_{jk}(l)$ is the Fourier transform of $\Psi_{jk}(\omega)$, that is,

$$\hat{\Psi}_{jk}(l) = (2\pi)^{-\frac{1}{2}} \int_{-\pi}^{\pi} \Psi_{jk}(\omega) e^{-il\omega} d\omega = e^{-i2\pi lk/2^j} (2\pi/2^j) \hat{\psi}(2\pi l/2^j). \tag{3.11}$$

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 ψ:

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

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

Most commonly used wavelets satisfy Assumptions A.2-A.3. Examples include Daubechies' (1992) compactly supported wavelets of positive order, Franklin wavelet, Lemarie-Meyer wavelets, Littlewood-Paley (or Shannon) wavelets, and spline wavelets. See (e.g.) Hernandez and Weiss (1996) for more discussions. Assumption A.3 rules out Haar wavelet, however, because its $\hat{\psi}(z) = -ie^{iz/2}\sin^2(z/4)/(z/4)$ decays to zero at a rate of $|z|^{-1}$ only.

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

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

where $\hat{\theta}$ is a consistent estimator of θ_0. We impose the following assumptions on the regression model $g(X_t, \theta)$ and parameter estimator $\hat{\theta}$:

ASSUMPTION A.4: (i) For each $\theta \in \Theta$, $g(\cdot, \theta)$ is a measurable function with respect to I_{t-1}; (ii) $g(X_t, \cdot)$ is twice continuously differentiable with respect to θ in an open convex neighborhood Θ_0 of θ_0 almost surely, with $\lim_{n\to\infty}\{n^{-1}\sum_{t=1}^{n} E\sup_{\theta\in\Theta_0}\|\frac{\partial}{\partial\theta}g(X_t,\theta)\|^4\} < \infty$ and $\lim_{n\to\infty}\{n^{-1}\sum_{t=1}^{n} E\sup_{\theta\in\Theta_0}\|\frac{\partial^2}{\partial\theta\partial\theta'}g(X_t,\theta)\|^2\} < \infty$.

ASSUMPTION A.5: $n^{\frac{1}{2}}(\hat{\theta} - \theta_0) = O_P(1)$.

We permit but do not require that $\hat{\theta}$ be the ordinary least square (OLS) or quasi-maximum likelihood estimators (e.g., Lee and Hansen 1994, Lumsdaine 1996). Any $\sqrt{n}$-consistent estimator of θ_0 suffices.

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

$$\hat{\rho}(l) = \hat{R}(l)/\hat{R}(0), \tag{3.13}$$

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

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

with $\hat{\sigma}^2 = n^{-1}\sum_{t=1}^{n}\hat{u}_t^2.$ 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

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

with $\hat{I}(\omega) = (2\pi n)^{-1}|\sum_{t=1}^{n}\hat{u}_t^2 e^{i\omega t}|^2$ the periodogram of $\{\hat{u}_t^2\}$. There are two ways to compute $\hat{\alpha}_{jk}.$ For compactly supported wavelets ψ, $\Psi_{jk}(\omega)$ in (3.6) is a sum of finite terms. The first expression of $\hat{\alpha}_{jk}$ in (3.16) is efficient to compute. For the band-limited wavelets (whose $\hat{\psi}$ has compact supports), the second expression of $\hat{\alpha}_{jk}$ in (3.16) is convenient to compute, as it is a sum of finite terms.

The integer J is called the finest scale parameter. Given n, a large J will lead to a smaller bias but a larger variance for $\hat{f}(0).$ We need to choose J properly to balance the bias and variance. In subsequent sections, we will provide proper conditions on J to ensure that the proposed test statistic have a well-defined limit distribution.

4. TEST STATISTIC AND ITS DISTRIBUTION

To introduce our test statistic, we define

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

Assumptions A.2-A.3 implies that $\lambda(z)$ is continuous almost everywhere, with $\lambda(0) = 0$ and $|\lambda(z)| \le C$. Note that the tail behavior of $\lambda(z)$ is governed by $\hat{\psi}(z)$, because $\sum_{m=-\infty}^{\infty}\hat{\psi}(z + 2\pi m)$ is 2π-periodic. For convenience, we impose a condition on $\lambda(z)$:

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 $\lambda(z)$ is real-valued implies $\lambda(-z) = \lambda(z).$

The test statistic for H_0 vs. H_A is defined as

$$S_n(J) \equiv V_n^{-\frac{1}{2}}(J)n^{\frac{1}{2}}\pi\left\{\hat{f}(0) - (2\pi)^{-1}\right\}, \tag{4.2}$$

where

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

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

The statistic S_n 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

$$\tilde{S}_n(J) \equiv \left[n/2^J \right]^{\frac{1}{2}} V_0^{-\frac{1}{2}} \pi \left\{ \hat{f}(0) - (2\pi)^{-1} \right\} , \qquad (4.4)$$

where

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

and

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

This statistic has the same null asymptotic distribution as $S_n(J)$ when J is large, because $V_n(J)/2^J \to V_0$ 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 $2^J/n \to 0$ as $n \to \infty$. Then under H_0

$$S_n(J) \to N(0,1) \text{ in distribution.}$$

Both small and large (i.e., fixed and increasing) finest scales J are allowed here. The choice of J may have important impact on the behavior of $S_n(J)$. We will use a data-driven method to choose J in the simulation study below.

5. ASYMPTOTIC LOCAL POWER

We now study the asymptotic power of $S_n(J)$ under the following class of generalized linear local alternatives

$$H_a(a_n) : h_t = \sigma_0^2 \left\{ 1 + a_n \sum_{j=1}^{\infty} \beta_j (\varepsilon_{t-j}^2 - 1) \right\} ,$$

where $\beta_j \geq 0$, $\sum_{j=1}^{\infty} \beta_j < \infty$ and $a_n \to 0$. Without loss of generality we further assume $a_n \sum_{j=1}^{\infty} \beta_j < 1$ for all n to ensure positivity of h_t. The class $H_a(a_n)$ describes all linear local ARCH alternatives, which include ARCH, GARCH and fractionally integrated GARCH of known or unknown orders.

Theorem 2: Suppose that Assumptions A.1-A.6 hold. (i) Let $J \in \mathbb{Z}$ be fixed. Define

$$\mu(J) = V_0(J)^{-\frac{1}{2}} \sum_{l=1}^{\infty} d_J(l)\beta_l;$$

where $V_0(J) = \sum_{l=1}^{\infty} d_J(l)^2$ and $d_J(l) = \sum_{j=0}^{J} \lambda(2\pi l/2^j)$. Then under $H_a(n^{-\frac{1}{2}})$,

$$S_n(J) \to N\{\mu(J), 1\} \text{ in distribution.}$$

(ii) Let $J \to \infty, 2^{2J}/n \to 0$. Define $\mu = V_0^{-\frac{1}{2}} \sum_{j=1}^{\infty} \beta_j$. Then under $H_a(2^{J/2}/n^{1/2})$,

$$S_n(J) \to N(\mu, 1) \text{ in distribution.}$$

Theorem 2(i) implies that with fixed finest scale J, $S_n(J)$ has nontrivial power against $H_a(a_n)$ with parametric rate $a_n = n^{-\frac{1}{2}}$, provided $\sum_{l=1}^{\infty} d_J(l)\beta_l \neq 0$. It has no power whenever $\sum_{l=1}^{\infty} d_J(l)\beta_l = 0$, which may occur for a fixed J, because $d_J(l)$ is a local average, depending on J and wavelet ψ. On the other hand, Theorem 2(ii) implies that with increasing finest scale J, $S_n(J)$ has nontrivial power against all linear local ARCH processes asymptotically. This follows because the noncentrality parameter $\mu > 0$ whenever ARCH exists (i.e., at least one parameter $\beta_j > 0$). Hong's (1997) kernel test is also consistent for all linear local ARCH processes. The tests of Lee and King (1993) and Demos and Sentana (1998) are not designed to test all linear local ARCH alternatives, since they are interested in testing a parametric ARCH(q) for fixed q. Lee and King (1993) and Demos and Sentana (1998) also consider one-sided tests for GARCH(1,1), which 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 of the parameters do not lie on the boundary of the parameter space (cf. Lee and King 1993, Demos and Sentana 1998). We note that Andrews (1999) recently also considered one-sided testing for GARCH(1,1) using a different approach.

The consistency of $S_n(J)$ against all possible linear local ARCH alternatives is desirable when no prior information about the alternative is known. This is, however, achieved at the price that $S_n(J)$ can detect $H_a(a_n)$ with $a_n = 2^{J/2}/n^{\frac{1}{2}}$ only. This rate is slower than the parametric rate $n^{-\frac{1}{2}}$, as is typical for nonparametric smoothed testing. However, it may not be taken too literally in practice. For example, if $2^J \propto (\ln n)^2$, the rate of the local alternatives is $n^{-\frac{1}{2}} \ln(n)$, only slightly slower than $n^{-\frac{1}{2}}$. Finally, we note that because of the one-sided nature of the tests, it is appropriate to use upper-tailed $N(0, 1)$ critical values for $S_n(J)$. For example, the upper-tailed $N(0, 1)$ critical value at the 5% level is 1.645.

6. ADAPTION TO DATA-DRIVEN FINEST SCALE

Theorem 2 shows that $S_n(J)$ is consistent for all linear locally ARCH processes as J increases. In practice, the choice of J may have important impact on the power. Because usually no prior information on the alternative is available, it may be desirable if J can be determined by suitable data-driven methods. To allow for such a possibility, we give the conditions on the data-dependent finest scale $\hat{J}$ (say) under which the randomness of $\hat{J}$ has asymptotically negligible effect on the limit distribution of $S_n(J)$.

ASSUMPTION A.7: $\hat{J}$ is a data-driven finest scale such that $\hat{J} - J = o_P(2^{-J/2})$, where J is a nonstochastic integer.

For fixed J, Assumption A.7 becomes $\hat{J} - J = o_P(1)$.

Theorem 3: Suppose that Assumptions A.1-A.7 hold, and $2^J/n \to 0$. Then under H_0, $S_n(\hat{J}) - S_n(J) \to 0$ in probability, and $S_n(\hat{J}) \to N(0,1)$ in distribution.

So far there are very few data-driven methods to choose J available in the literature. To our knowledge, only Walter (1995) proposes a data-driven J, using a mean square error criterion. We will use it in our simulation study below.

7. MONTE CARLO EVIDENCE

We now investigate the finite sample performance of the wavelet-based test $S_n(J)$. We use Franklin wavelet and the second order spline (S_1 and S_2, respectively). Franklin wavelet is defined via its Fourier transform,

$$\hat{\psi}(z) = (2\pi)^{-1/2} e^{iz/2} \frac{\sin^4(z/4)}{(z/4)^2} \left(\frac{1 - 2/3\cos^2(z/4)}{(1 - 2/3\sin^2(z/2))(1 - 2/3\sin^2(z/4))} \right)^{1/2}. \tag{7.1}$$

For the second order spline wavelet, its Fourier transform

$$\hat{\psi}(z) = -(2\pi)^{-1/2} i e^{iz/2} \frac{\sin^6(z/4)}{(z/4)^3} \left(\frac{P(z/4 + \pi/2)}{P(z/2)P(z/4)} \right)^{1/2} \tag{7.2}$$

where $P(z) = \frac{1}{30}\cos^2(2z) + \frac{13}{30}\cos(2z) + \frac{8}{15}$.

The choice of the finest scale parameter, J, may be important. We choose a data-driven J via Walter's (1994) algorithm, which makes use of the fact that 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 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 0 to 1. The grid search is iterated until we get the scale J at which the error increases most rapidly. Then, one obtains the finest scale. In our simulation, we choose the finest scale J for which the change in error between J and $J+1$ exceeds 100 %. We note that this method is more suitable for estimation of $f(\omega)$ on $[-\pi, \pi]$ rather than at frequency zero. Nevertheless, the simulation below shows that it works relatively well in the present context.

We compare S_1 and S_2 with three one-sided ARCH tests—Hong's (1997) kernel test (denoted K), Lee and King's (LK; 1993) locally most mean powerful test, and Demos and Sentana's (1998) one-sided LM test (DS). We also include Engle's (1982) two-sided LM test (LM); which is commonly used in practice. For the K test; we use Quadratic-Spectral kernel and select a data-driven bandwidth via Andrews' (1991) plug-in method based on an ARCH(1,1) approximating model. For the LK test, we use the version of the test statistic which is robust to non-normality (see Lee and King 1993, (13)). The tests of S_1; S_2; K; and LK are all asymptotically one-sided $N(0, 1)$ under H_0. The DS test is computed as the sum of the squared t-statistics of the positive coefficients in the regression of $\hat{\varepsilon}_t^2$ on a constant and the first q lags of $\hat{\varepsilon}_t^2$. This test has a nonstandard mixed chi-square distribution, whose critical values are given in Demos and Sentana's (1998, Table 1). The LM test has asymptotic χ_q^2 distribution under H_0 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 alternative. When the order of the alternative is unknown, as is often the case in practice, these tests may suffer from power losses when using a suboptimal lag order. To investigate the effect of using different choices of q for these tests; we consider $q = 1$ and 12 (denoted $LK(1)$; $LK(12)$; $DS(1)$; $DS(12)$; $LM(1)$; and $LM(12)$).

Consider the data generating process

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

where $X_t = (1, m_t)'$; $m_t = 0.8 m_{t-1} + \nu_t$ and $\nu_t \sim$ i.i.d. $N(0, 4)$; $\xi_t \sim$ i.i.d. $N(0, 1)$. Both $\{\xi_t\}$ and $\{\nu_t\}$ are mutually independent. We set $b_0 = (1, 1)'$ and estimate them 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. Two sample sizes, $n = 100; 200$; are considered. To reduce the possible effects of the initial condition, $n + 1000$ observations are generated and then the first 1000 ones are discarded. Also, the initial values for $\varepsilon_t;\ t \le 0$ are set to be zero, and $h_t;\ t \le 0$ is set to be 1. For each experiment, 1000 iterations are generated using the GAUSS random number generator on a personal computer.

We first study the size by setting $h_t = 1$. Table 1 reports the size at the 10 % and 5 % levels using asymptotic critical values. The tests S_1; S_2 and K attain reasonable sizes, though they tend to slightly underreject. The tests $LK(1)$ and $DS(1)$ have best sizes. The tests $LK(12)$ and $LM(12)$ show some underrejections, while $DS(12)$ tends to overreject slightly.

Next, we investigate the power under the following alternatives.

$$
\begin{array}{ll}
\text{ARCH(1):} & h_t = 1 + \beta\varepsilon_{t-1}^2; \\
\text{ARCH(12a):} & h_t = 1 + \beta\sum_{j=1}^{12}\varepsilon_{t-j}^2; \\
\text{ARCH(12b):} & h_t = 1 + \beta\sum_{j=1}^{12}(1 - j/13)\varepsilon_{t-j}^2; \\
\text{GARCH(1,1):} & h_t = 1 + \alpha\varepsilon_{t-1}^2 + \beta h_{t-1}.
\end{array}
$$

For these alternatives, we choose the values of parameters $(\alpha; \beta)$ to ensure strictly positive conditional variance and finite unconditional variance. For ARCH(1), we consider $\beta = 0.3; 0.95$. It does not have a sharp spectral peak at any frequency. In contrast, ARCH(12a) and ARCH(12b) are allowed to have a relatively long distributional lag, which generates a spectral peak at frequency zero. Linearly declining weights in ARCH(12b) were often considered in the literature (e.g., Engle 1982, Engle et al. 1987). We consider $\beta = 0.95/12$ for ARCH(12a) and $\beta = 0.95/\sum_{j=1}^{12}(1 - j/13)$ for ARCH(12b). GARCH(1,1) is a workhorse in modelling economic and financial time series (cf. Bollerslev 1986). When $\alpha + \beta < 1$, GARCH(1,1) can be expressed as ARCH(∞) with coefficients declining at exponential rate. We set $(\alpha; \beta) = (0.3; 0.3); (0.3; 0.65)$. The latter displays relatively persistent ARCH, which yields a spectral peak at frequency zero. Tables 2-4 report the size-corrected power under these alternatives. The empirical critical values are obtained from 1000 replications under H_0.

Table 2 reports the power against ARCH(1). For $\beta = 0.3$; $LK(1)$ and $DS(1)$ have similar powers and are the most powerful. The K test has power very close to that of $LK(1)$ and $DS(1)$. These three tests have better power than $LM(1)$; which in turn has better power than S_1 and S_2. Compared to the kernel test K; wavelets suffer from nontrivial power loss when there is no sharp spectral peak. The fact that $LK(1)$ and $DS(1)$ are most powerful here is not surprising, because they use the optimal lag $q = 1$.

The powers of $LK(12)$, $DS(12)$ and $LM(12)$ are substantially smaller, with $LK(12)$ the smallest. These tests are less powerful than S_1 and S_2. This suggests that power loss may be severe when one uses a suboptimal q for LM, DS and LM. Note that the power rankings remain largely the same when $\beta = 0.95$.

Table 3 reports the power under ARCH(12a) and ARCH(12b). Under ARCH(12a), $LK(12)$ has the best power, and dominates $DS(12)$. These two tests use the optimal lag order $q = 12$. Both S_1 and S_2 have power close to that of $LK(12)$. They have better power than K. The S_2 test is slightly better than $DS(12)$ and is substantially better than $LM(12)$ for $n = 100$, although the latter uses the optimal lag. This indicates that wavelets work pretty well when ARCH effect has a relatively long distributional lag. Under ARCH(12b), S_1, S_2 and $LK(12)$ have comparable power and are more powerful than $DS(12)$, $LM(12)$, K, $LK(1)$ and $DS(1)$.

Table 4 reports the power against GARCH(1,1). When $(\alpha, \beta) = (0.3, 0.3)$, there is relatively weak ARCH effect. In this case K attains the best power, followed very 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, especially for $LK(12)$. When $(\alpha, \beta) = (0.3, 0.65)$, there is relatively persistent ARCH, Here, S_1 and S_2 perform the best. They outperform K, which, in turn, 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. This suggests that the use of a long lag order may not suffer from severe power loss when there exists persistent ARCH. Finally, we note that while $DS(1)$ and $LK(1)$ have similar power when $(\alpha, \beta) = (0.3, 0.3)$ and $(0.3, 0.65)$, $DS(12)$ has better power than $LK(12)$ when $(\alpha, \beta) = (0.3, 0.3)$, and similar power when $(\alpha, \beta) = (0.3, 0.65)$.

In summary, we find:

1) The wavelet tests, S_1 and S_2, have similar size and power in almost all the cases. The choice of wavelet function is not important.

2) The relative power performance of the one-sided kernel and wavelet tests depends on the spectral shape of the squared residuals. When ARCH is of a short memory (as in ARCH(1), GARCH(1,1) with $(\alpha, \beta) = (0.3, 0.3)$), the one-sided kernel test is more powerful than the one-sided wavelet test. When there exists relatively persistent ARCH (i.e., GARCH(1,1) with $(\alpha, \beta) = (0.3, 0.65)$, or when ARCH effect has a long distributional

lag (i.e., ARCH(12)), there is a spectral peak at frequency for the squared time series process. In this case, the wavelet test outperforms the kernel test.

3) The tests LK, DS and LM attain their own maximal powers when the optimal lag order is used, but they may su¤er severe power loss when a suboptimal lag is used. Under each alternative, the two-sided LM test is always dominated by some one-sided tests using the same lag order. This suggests nontrivial power gain of exploiting the one-sided nature of the ARCH alternative.

4) None of the one-sided tests dominates the others in power for the alternatives under study. When ARCH e¤ect has short memory (ARCH(1) and GARCH(1,1) with (®;¯) = (0:3; 0:3)), the one-sided 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 e¤ect has relatively long memory (ARCH(12a,b) and GARCH(1,1) with (®;¯) = (0:3; 0:65)); the one-sided wavelet test has power close to or even better than that of LK and DS with the optimal lag orders. We note that both the kernel and wavelet tests do not require the knowledge of the optimal lag.

The fact that the kernel test K has good power when ARCH e¤ect is weak or of relatively short memory while the wavelet tests S_1 and S_2 have good power when ARCH e¤ect is persistent suggests that a suitable Bonferoni procedure that combines the kernel and wavelet tests may have good power against both weak and persistent ARCH. We consider two simple Bonferoni procedures, BF_1; which combines S_1 and K; and BF_2; which combines S_2 and K: 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 fS_1; Kg: Then one rejects H_0 at level ® if P_1 < ®=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 show some underrejections, which is consistent with the conservative nature of Bonferoni procedures. In spite of this underrejection in size, however, they do have all-round good power against all the alternatives under study. In particular, they have better power than the wavelet tests S_1 and S_2 when ARCH is less persistent, and have better power than the K test when ARCH is persistent. This suggests that BF_1 and BF_2 do combine the advantages of the wavelet and kernel approaches.

8. CONCLUSION

We consider a wavelet-based one-sided test for ARCH. The test statistic is based on a wavelet spectral density estimator at frequency zero of the square of estimated residuals from a regression model. An essential feature of ARCH is that the squared process is positively correlated at all lags, thus resulting in a spectral mode at frequency zero. In particular, a spectral peak arises when there exists persistent ARCH, or when ARCH effect carries over a long distributional lag, although its effect may be small at each lag. Because the kernel method tends to underestimate modes or peaks, it may not be a powerful tool when there exists persistent ARCH. In contrast, wavelets can efficiently capture such inhomogeneous features as spectral peaks, and are expected to perform well in these situations. This is confirmed in a simulation study. Since there exists unknown smoothness from the data, the wavelet-based test for ARCH is a useful complement to the existing one-sided tests for ARCH.

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. (1999), Testing When a Parameter is on the Boundary of the Maintained Hypothesis. Cowles Foundation, Yale University.

Baillie, R., Bollerslev, T and Mikkelsen, H. (1996), Fractionally Integrated Generalized Autoregressive Conditional Heteroskedasticity, Journal of Econometrics 74, 3-30.

Bera, A.K. and Higgins, M.L. (1992), A Test for Conditional Heteroskedasticity in Time Series Models, Journal of Time Series Analysis, 13, 501-519.

Bera, A.K., Ra, S., and Sakar, N. (1998), Hypothesis Testing for Some Nonregular Cases in Econometrics, in Quantitative Economics: Theory and Practice, eds. S. Chakravorty, D. Coondoo, and R. Mookherjee, New Delhi: Applied Publishers, pp.324-356.

Bollerslev, T. (1986), Generalized Autoregressive Conditional Heteroskedasticity, Journal of Econometrics, 31, 307-327.

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

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

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

Demos, A. and Sentana, E. (1998), Testing for GARCH E¤ects: A One-Sided Approach, Journal of Econometrics 86, 97-127.

Diebold, F. X. (1987), Testing for Serial Correlation in the Presence of ARCH, Proceedings from the American Statistical Association, Business and Economic Statistics Section, 323-328.

Diebold, F.X. and Pauly, P. (1989), Small Sample Properties of Asymptotically Equivalent Tests for Autoregressive Conditional Heteroskedasticity, Statistical Papers, 30, 105-131.

Domowitz, I. and Hakkio, C.S. (1987), Testing for Serial Correlation and Common Factor Dynamics in the Presence of Heteroskedasticity, Manuscript, Department of Economics, Northwestern University.

Donoho, D. and Johnstone, I. (1994), Ideal Spatial Adaptation by Wavelet Shrinkage, Biometrika 81, 425-455.

Donoho, D. and Johnstone, I. (1995), Wavelet Shrinkage: Asymptopia? (with discussion), Journal of the Royal Statistical Society Series B, 57, 301-369.

Donoho, D. and Johnstone, I. (1995), Adapting to Unknown Smoothness via Wavelet Shrinkage, Journal of the American Statistical Association 90, 1200-1224.

Donoho, D., Johnstone, I., Kerkyacharian, G., Picard, D. (1996), Density Estimation by Wavelet Thresholding, Annals of Statistics 24, 508-539.

Engle, R.F. (1982), Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom In‡ation, Econometrica, 50, 987-1007.

Engle, R.F., Hendry, D.F., Trumble, D. (1985), Small-sample Properties of ARCH Estimators and Tests, Canadian Journal of Economics, 18, 67-93.

Engle, R.F., Lilien, D.M., Robins, R.P. (1987), Estimating Time Varying Risk Premia in the Term Structure: The ARCH-M Model, Econometrica, 55, 391-407.

Gao, H. (1993), Wavelet Estimation of Spectral Densities in Time Series Analysis. Ph.D Dissertation, 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.

Go¤e, W.L.(1994) Wavelets in Macroeconomics: An Introduction, in Belsley, D. (ed.), Computational Techniques for Econometrics and Economic Analysis, The Netherlands: Kluwer Academic Publishers, pp.137-149.

Gourieroux, C., Holly, A., Monfort, A. (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. and Heyde, C.C. (1980), Martingale Limit Theory and Its Application, Academic-Press, New York.

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

Hong, Y. (1997), One-sided ARCH Testing in Time Series Models, Journal of Time Series Analysis 18, 253-277.

Hong, Y. and Lee, J. (1999), Wavelet-Based Estimation for Heteroskedasticity and Autocorrelation Consistent Variance-Covariance Matrices, manuscript, Department of Economics, Cornell University.

Hong, Y. and Shehadeh, R. (1999), A New Test for ARCH E¤ects and its Finite Sample Performance, Journal of Business and Economic Statistics, 17, 1-18.

Jensen, M.J.(1996), An Alternative Maximum Likelihood Estimator of Long Memory Processes Using Compactly Supported Wavelets, manuscript, Department of Economics, Southern Illinois University at Carbondale.

King, M. and Wu. P. (1997), Locally Optimal One-sided Tests for Multiparameter Hypotheses, Econometric Reviews 16, 131-156.

Lee, J.H.H. (1991), A Lagrange Multiplier Test for GARCH Models, Economics Letters, 37, 265-271.

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

Lee, J.H.H. and King, M.L. (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. and 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, J. and Hong (1998), Testing for Serial Correlation of Unknown Form Using Wavelet Methods, CAE Working Paper, Department of Economics, Cornell University.

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 Multiresultion Signal Decomposition: The Wavelet Representation, IEEE Transactions on Pattern Analysis and Machine Intelligence 11, 674-693.

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

MilhÁj, A. (1985), The Moment Structure of ARCH Processes, Scandinavian Journal of Statistics, 12, 281-292.

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

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 Estimates of the Spectrum of a Stationary Time Series, Annals of Mathematical Statistics, 28, 329-348.

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

Priestley, M.B. (1996), Wavelets and Time-dependent Spectral Analysis, Journal of Time Series Analysis 17, 85-103.

Ramsey, J.B.(1998) Regression over Time Scale Decomposition: A Sampling Analysis of Distributional Properties, manuscript, Department of Economics, New York University.

Ramsey, J.B. and Lampart, C.(1998a) The Decomposition of Economic Relationships by Time Scale Using Wavelets: Money and Income, forthcoming, Macroeconomic Dynamics.

———,(1998b) The Decomposition of Economic Relationships by Time Scale Using Wavelets: Expenditure and Income, forthcoming, Society for Nonlinear Dynamics and Econometrics.

Ramsey, J.B. and Zhang, Z.(1996) The Application of Wave Form Dictionaries to Stock Market Index Data, in Kadtke, J.B. and Kravtsov, Y.A. (eds.), Predictability of Complex Dynamical Systems, New York: Springer-Verlag, pp.189-205.

———(1997) The Analysis of Foreign Exchange Data Using Waveform Dictionaries, Journal of Empirical Finance, 4, 341-372.

Ramsey, J., Usikov, D. and Zaslavsky, G. (1995), An Analysis of U.S. Stock Price Behavior Using Wavelets, Fractals, 3(2), 377-389.

Robinson, P.M. (1991), Testing for Strong Serial Correlation and Dynamic Conditional Heteroskedasticity in Multiple Regression, Journal of Econometrics, 47, 67-84.

SenGupta, A. and L. Vermeire (1986), Locally Optimal Tests for Multiparameter Hypotheses, Journal of the American Statistical Association 81, 819-825.

Self, S., Liang, K. (1987), Asymptotic Properties of Maximum Likelihood Estimators and Likelihood Ratio Tests under Nonstandard Conditions, Journal of the American Statistical Association 82, 605-610.

Silvapulle, M. and P. Silvapulle (1995), A Score Test Against One-Sided Alternatives, Journal of the American Statistical Association 90, 342-349.

Subba Rao, T. and Indukumar, K.C.(1996) Spectral and Wavelet Methods for the Analysis of Nonlinear and Nonstationary Time Series, Journal of the Franklin Institute, 333(b), 3, 425-452.

Taylor, S. (1984), Modelling Financial Time Series. Wiley and Sons: New York.

Wang, Y. (1995), Jump and Sharp Cusp Detection by Wavelets, Biometrika, 82, 385-397.

Walter, G.(1994) Wavelets and Other Orthogonal Systems with Applications, CRC Press: Boca Raton.

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.

Wu, P. and M. King (1994), One-sided Hypothesis Testing in Econometrics: A Survey, Department of Econometrics Research Working Paper, Monash University.

MATHEMATICAL APPENDIX

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

Lemma A.1: Define

$$d_J(l) = \sum_{j=0}^{J} \lambda(2\pi l/2^j); \quad l, J \in \mathbb{Z};$$

where $\lambda(z)$ as in (4.1). Then

(i) $d_J(0) = 0$ and $d_J(-l) = d_J(l)$;

(ii) $|d_J(l)| \le C$ uniformly in J and l;

(iii) For any given $l \in \mathbb{Z}, l \neq 0$, $d_J(l) \to \infty$ as $J \to \infty$;

(iv) For $r \ge 1$, $\sum_{l=1}^{n-1} |d_J(l)|^r = O(2^J)$ as $J, n \to \infty$.

Proof of Lemma A.1: See Hong and Lee (1999, Proof of Lemma A.1).

Lemma A.2: Let $V_n(J)$ and V_0 be defined as in Theorem 1. Suppose $J \to \infty$, $2^J/n \to 0$. Then $V_n(J)/2^J \to V_0$ as $n \to \infty$.

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

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

where the second equality follows by reindexing. We shall show $\tilde{V}_n(J)/2^{J+1} \to V_0$, which, with dominated convergence, implies $V_n(J)/2^{J+1} \to V_0$. Let $I = I_n \to \infty$, $I/J \to 0$ as $n \to \infty$. Decompose

$$\tilde{V}_n(J) = \tilde{V}_n(I) + Q_{1n} + Q_{2n}; \qquad \text{(A1)}$$

where

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

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

For the second term Q_{1n} in (A1), we have that as $n \to \infty$,

$$\begin{aligned} Q_{1n} &= 2^J \sum_{p=-J}^{J} \sum_{j=I+1}^{J} 2^{-(J-j)} \frac{1}{2\pi} \left\{ (2\pi/2^j) \sum_{l=1}^{n-1} \lambda(2\pi l/2^j)\lambda(2^{|p|}2\pi l/2^j) \right\} \\ &= 2^{J+1} \sum_{p=-\infty}^{\infty} \frac{1}{2\pi} \int_0^{\infty} \lambda(z)\lambda(2^{|p|}z)dz \{1 + o(1)\} \\ &= 2^J V_0(1 + o(1)) \qquad \text{(A2)} \end{aligned}$$

by dominated convergence,

$$(2\pi/2^j)\sum_{l=1}^{n-1}\lambda(2\pi l/2^j)\lambda(2^{|p|}2\pi l/2^j) \to \int_0^\infty \lambda(z)\lambda(2^{|p|}z)dz \text{ for any given } p \text{ as } j\to\infty;$$

$$\sum_{j=l+1}^{J} 2^{-(J-j)} \to 2 \text{ as } l\to\infty;\ J/l\to\infty;$$

and symmetry of $\lambda(z)$ given Assumption A.5: Using a similar reasoning, for the last term in (A1), we have

$$Q_{2n} = o(2^{J+1}). \tag{A3}$$

Finally, for the first term in (A1), we can show

$$\begin{aligned}
\tilde{V}_n(l) &\le \left|\sum_{p=-l}^{l}\sum_{j=|p|}^{l}\sum_{l=1}^{n-1}\lambda(2\pi l/2^j)\lambda(2^{|p|}2\pi l/2^j)\right| \\
&\le \sum_{p=-l}^{l}2^{-|p|/2}\sum_{j=|p|}^{l}2^j\left\{2^{-j}\sum_{l=1}^{n-1}\lambda^2(2\pi l/2^j)\right\}^{\frac{1}{2}}\left\{2^{-(j-|p|)}\sum_{l=1}^{n-1}\lambda^2(2\pi l/2^{j-|p|})\right\}^{\frac{1}{2}} \\
&\le C^2\sum_{p=-\infty}^{\infty}2^{-|p|/2}\sum_{j=0}^{\infty}2^j \\
&\le 8C^2 2^l
\end{aligned} \tag{A4}$$

by Assumption A.5, where we used the fact that for any $l>0$; $j>0$;

$$\begin{aligned}
2^{-j}\sum_{l=1}^{n-1}\lambda^2(2\pi l/2^j) &= 2^{-j}\left(\sum_{l=1}^{2^j}+\sum_{l=2^j+1}^{n-1}\right)\lambda^2(2\pi l/2^j) \\
&\le 2^{-j}\sum_{l=1}^{2^j}C(2\pi l/2^j)^{2q} + 2^{-j}\sum_{l=2^j+1}^{n-1}C(2\pi l/2^j)^{-2\tau} \\
&\le C + C\left\{2^{-j}\sum_{l=1}^{\infty}(1+2\pi l/2^j)^{-2\tau}\right\} \\
&\le C\left\{1+\frac{1}{2\pi}\int_0^\infty(1+x)^{-2\tau}dx\right\};
\end{aligned}$$

where the first inequality follows by Assumption A.5 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 (A1)-(A4) and $l/J\to 0$ yields the desired result.

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

$$\sum_{l=1}^{n-1} d_J(l)\beta(l) - \sum_{l=1}^{\infty}\beta(l) = \sum_{l=1}^{n-1}\{d_J(l) - 1\}\beta(l) - \sum_{l=n+1}^{\infty}\beta(l). \quad \text{(A5)}$$

For the second term, we have

$$\left|\sum_{l=n+1}^{\infty}\beta(l)\right| \le \sum_{l=n+1}^{\infty}|\beta(l)| \to 0 \quad \text{(A6)}$$

as $n \to \infty$ given $\sum_{l=1}^{\infty}|\beta(l)| < \infty$. For the first term, we have

$$\sum_{l=1}^{n-1}\{d_J(l) - 1\}\beta(l) \to 0 \quad \text{(A7)}$$

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. Collecting (A6)-(A7) yields the desired result.

Lemma A.4: Let $V_n(J)$ be as defined in (4.3). Suppose $\hat{J} \in \mathbb{Z}$ is a data-driven integer such that $\hat{J} - J = o_P(1)$, where $J \in \mathbb{Z}$ is nonstochastic, then $V_n(\hat{J})/V_n(J) \to^p 1$.

Proof of Lemma A.4: By the definition of $V_n(J)$ in (4.3) and the Cauchy-Schwarz inequality, we have

$$\left|V_n(\hat{J}) - V_n(J)\right| \le \sum_{l=1}^{n-1}\{d_{\hat{J}}(l) - d_J(l)\}^2 + 2\left\{\sum_{l=1}^{n-1}[d_{\hat{J}}(l) - d_J(l)]^2\right\}^{\frac{1}{2}}\left\{\sum_{l=1}^{n-1}d_J(l)^2\right\}^{\frac{1}{2}}. \quad \text{(A8)}$$

Now, given Assumptions A.5 and A.7, we have $|d_{\hat{J}}(l) - d_J(l)| \le \sum_{j=\min(\hat{J},J)}^{\max(\hat{J},J)}|\lambda(2\pi l/2^j)|$, and so

$$\begin{aligned}
\sum_{l=1}^{n-1}\{d_{\hat{J}}(l) - d_J(l)\}^2 &\le C^2\sum_{l=1}^{m}\left\{\sum_{j=\min(J,\hat{J})}^{\max(J,\hat{J})}(2\pi l/2^j)^q\right\}^2 + C^2\sum_{l=m+1}^{n-1}\left\{\sum_{j=\min(J,\hat{J})}^{\max(J,\hat{J})}(2\pi l/2^j)^{-\tau}\right\}^2 \\
&\le C^2(\hat{J} - J)^2 2^{-2q\min(J,\hat{J})}m^{2q+1} + C^2(\hat{J} - J)^2 2^{2\tau\max(J,\hat{J})}m^{1-2\tau} \\
&= (\hat{J} - J)^2\left\{2^{-2qJ}2^{-2q\min(0,\hat{J}-J)}m^{2q+1} + 2^{2\tau J}2^{2\tau\max(0,\hat{J}-J)}m^{1-2\tau}\right\} \\
&= O_P\{(\hat{J} - J)^2 2^J\} \quad \text{(A9)}
\end{aligned}$$

by choosing $m = 2^J$ and noting $\hat J - J = o_P(1)$. This, together with $V_n(J) = O(2^J)$ from Lemma A.2, implies

$$|V_n(\hat J)/V_n(J) - 1| = V_n(J)^{-1}|V_n(\hat J) - V_n(J)| = O_P(\hat J - J) = o_P(1).$$

Proof of Theorem 1: Put $u_t = \xi_t^2 - 1$. Define

$$\tilde f(0) = \frac{1}{2\pi} + \sum_{j=0}^{J}\sum_{k=0}^{2^j} \tilde\alpha_{jk}\psi_{jk}(0), \tag{A10}$$

where $\tilde\alpha_{jk} = (2\pi)^{-\frac{1}{2}}\sum_{l=1-n}^{n-1}\tilde\rho(j)\hat\psi^*_{jk}(l)$; $\tilde\rho(j) = \tilde R(j)/\tilde R(0)$; $\tilde R(l) = n^{-1}\sum_{t=|l|+1}^{n} u_t u_{t-|l|}$; and $\hat\psi_{jk}(l) = (2\pi)^{-\frac{1}{2}}\int_{-\pi}^{\pi}\psi_{jk}(\omega)e^{-il\omega}d\omega$ is the Fourier transform of $\psi_{jk}(\omega)$.

Write $\hat f(0) - (2\pi)^{-1} = \{\hat f(0) - \tilde f(0)\} + \{\tilde f(0) - 1/2\pi\}$. We shall prove Theorem 1 by showing Theorems A.1-A.2 below.

Theorem A.1: $V_n^{-\frac{1}{2}} n^{\frac{1}{2}}\{\hat f(0) - \tilde f(0)\} \to^p 0$.

Theorem A.2: $V_n^{-\frac{1}{2}} n^{\frac{1}{2}}\pi\{\tilde f(0) - (2\pi)^{-1}\} \to^d N(0,1)$.

Proof of Theorem A.1: Recall that $\hat\psi_{jk}(h)$ is the Fourier transform of $\psi_{jk}(\omega)$; we have

$$\psi_{jk}(0) = (2\pi)^{-\frac{1}{2}}\sum_{h=-\infty}^{\infty}\hat\psi_{jk}(h) = (2\pi)^{-\frac{1}{2}}\sum_{h=-\infty}^{\infty} e^{-i2\pi hk/2^j}(2\pi/2^j)^{\frac{1}{2}}\hat\Psi(2\pi h/2^j) \tag{A11}$$

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

$$\hat\alpha_{jk} = (2\pi)^{-\frac{1}{2}}\sum_{l=1-n}^{n-1}\hat\rho(l)e^{i2\pi lk/2^j}(2\pi/2^j)^{\frac{1}{2}}\hat\Psi^*(2\pi l/2^j). \tag{A12}$$

Collecting (3.15) and (A11)-(A12) with Lemma A.1 yields

$$\begin{aligned}
\hat f(0) &= \frac{1}{2\pi} + \sum_{j=0}^{J}\sum_{k=1}^{2^j}\hat\alpha_{jk}\psi_{jk}(0)\\
&= \frac{1}{2\pi} + \frac{1}{2\pi}\sum_{l=1-n}^{n-1}\left\{\sum_{j=0}^{J}\sum_{h=-\infty}^{\infty}\sum_{k=1}^{2^j} e^{i2\pi(l-h)k/2^j}(2\pi/2^j)\hat\Psi^*(2\pi l/2^j)\hat\Psi(2\pi h/2^j)\right\}\hat\rho(l)\\
&= \frac{1}{2\pi} + \frac{1}{2\pi}\sum_{l=1-n}^{n-1} d_J(l)\hat\rho(l)\\
&= \frac{1}{2\pi} + \frac{1}{\pi}\sum_{l=1}^{n-1} d_J(l)\hat\rho(l)
\end{aligned} \tag{A13}$$

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

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

where we used the well-known 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 (A13) 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

$$\tilde f(0) = (2\pi)^{-1} + \pi^{-1}\sum_{l=1}^{n-1} d_J(l)\tilde\rho(j). \tag{A14}$$

Combining (A13)-(A14), we can write

$$\pi\left\{\hat f(0) - \tilde f(0)\right\} = \sum_{l=1}^{n-1} d_J(l)\{\hat\rho(l) - \tilde\rho(l)\}. \tag{A15}$$

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

$$V_n^{-\frac{1}{2}}(J) n^{\frac{1}{2}} \sum_{l=1}^{n-1} d_J(l)\{\hat R(l) - \tilde R(l)\} \to^p 0. \tag{A16}$$

We shall show (A16) for large J (i.e., $J \to \infty$ as $n \to \infty$), where $V_n(J) = 2^J V_0\{1 + o(1)\}$ by Lemma A.2. The proof for fixed J is similar, with $V_n(J) \to V_0(J) = O(1)$, where $V_0(J)$ is as in Theorem 1(i).

Put $\hat\xi_t = \hat\varepsilon_t/\hat\sigma$ and recall $u_t = \xi_t^2 - 1$. Straightforward algebra yields $\hat R(l) - \tilde R(l) = \hat A_1(l) + \hat A_2(l) + \hat A_3(l)$, where

$$\begin{aligned}
A_1(l) &= n^{-1}\sum_{t=l+1} u_t(\hat\xi^2_{t-l} - \xi^2_{t-l}),\\
\hat A_2(l) &= n^{-1}\sum_{t=l+1} (\hat\xi^2_t - \xi^2_t)u_{t-l},\\
\hat A_3(l) &= n^{-1}\sum_{t=l+1} (\hat\xi^2_t - \xi^2_t)(\hat\xi^2_{t-l} - \hat\xi^2_{t-l}).
\end{aligned}$$

Noting $\xi_t = \varepsilon_t/\sigma_0$ under H_0; where $\sigma_0 = E(\varepsilon_t^2)$; we have

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

where

$$\begin{aligned}\hat{\chi}_{11}(l) &= n^{-1} \sum_{t=l+1}^{n} u_t(\hat{\varepsilon}_{t-l} - \varepsilon_{t-l})^2; \\ \hat{\chi}_{12}(l) &= n^{-1} \sum_{t=l+1}^{n} u_t \varepsilon_{t-l}(\hat{\varepsilon}_{t-l} - \varepsilon_{t-l}); \\ \hat{\chi}_{13}(l) &= n^{-1} \sum_{t=l+1}^{n} u_t \varepsilon_{t-l}^2.\end{aligned}$$

By the Cauchy-Schwarz inequality, the mean value theorem and Assumptions A.1 and A.4–A.5, we have

$$\begin{aligned}\left|\sum_{l=1}^{n-1} d_J(l)\hat{\chi}_{11}(l)\right| &\leq \left\|\hat{b} - b_0\right\|^2 \left\{\sum_{l=1}^{n-1} |d_J(l)|\right\} \left\{n^{-1}\sum_{t=1}^{n} u_t^2\right\}^{\frac{1}{2}} \left\{n^{-1}\sum_{t=1}^{n} \sup_{b \in B_0} \left\|\frac{\partial}{\partial b} g(X_t; b)\right\|^4\right\}^{\frac{1}{2}} \\ &= O_P(2^J/n); \qquad \text{(A17)}\end{aligned}$$

where $\sum_{l=1}^{n-1} |d_J(l)| = O(2^J)$ by Lemma A.1(iv).

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

$$\begin{aligned}\left|\sum_{l=1}^{n-1} d_J(l)\hat{\chi}_{12}(l)\right| &\leq \left\|\hat{b} - b_0\right\| \sum_{l=1}^{n-1} |d_J(l)| \left\|n^{-1}\sum_{t=1}^{n} u_t \varepsilon_{t-l} \frac{\partial}{\partial b} g(X_t; b_0)\right\| \\ &\quad + \frac{1}{2}\left\|\hat{b} - b_0\right\|^2 \sum_{l=1}^{n-1} |d_J(l)| \left\{n^{-1}\sum_{t=1}^{n} |u_t \varepsilon_{t-l}| \sup_{b \in B_0} \left\|\frac{\partial^2}{\partial b \partial b'} g(X_t; b)\right\|\right\} \\ &= O_P(2^J/n) \qquad \text{(A18)}\end{aligned}$$

by Markov's inequality and Lemma A.1, where we have used

$$E\left\|n^{-1}\sum_{t=1}^{n} u_t \varepsilon_{t-l} \frac{\partial}{\partial b} g(X_t; b_0)\right\| = O(n^{-1})$$

given $E(u_t | I_{t-1}) = 0$ a.s. Finally, we also have

$$\left|\sum_{l=1}^{n-1} d_J(l)\hat{\chi}_{13}(l)\right| = O(2^J/n^{\frac{1}{2}}) \qquad \text{(A19)}$$

by Markov's' inequality and $\sup_{0<l<n} E\hat{A}_{13}^2(l) = O(n^{-1})$, which follows from $E(u_t|I_{t-1}) = 0$ a.s. and Assumption A.2. Combining (A17)-(A19) and $\hat\sigma^2 - \sigma_0^2 = O_P(n^{-\frac{1}{2}})$, we obtain

$$\sum_{l=1}^{n-1} d_J(l)\hat{A}_1(l) = O_P(2^J/n). \tag{A20}$$

Similarly, we have

$$\sum_{l=1}^{n-1} d_J(l)\hat{A}_2(l) = O_P(2^J/n). \tag{A21}$$

Next, we consider $\hat{A}_3$. As shown in Hong (1997, p.272),

$$\sup_{0<l<n} |\hat{A}_3(l)| \le n^{-1}\sum_{t=1} (\hat{\xi}_t^2 - \xi_t^2)^2 = O_P(n^{-1}).$$

This, together with $\sum_{l=1}^{n-1} |d_J(l)| = O(2^J)$ from Lemma A.1(iv), implies

$$\left|\sum_{l=1}^{n-1} d_J(l)\hat{A}_3(l)\right| \le \sup_{0<l<n} |\hat{A}_3(l)| \left\{\sum_{l=1}^{n-1} |d_J(l)|\right\} = O_P(2^J/n). \tag{A22}$$

Collecting (A20)-(A22) and $V_n(J) = 2^{J+1}V_0\{1 + o(1)\}$ by Lemma A.2, we have

$$V_n^{-\frac{1}{2}} n^{\frac{1}{2}} \sum_{l=1}^{n-1} d_J(l)\{\hat{R}(l) - \tilde{R}(l)\} = O_P(2^{J/2}/n^{\frac{1}{2}}) = o_P(1)$$

given $2^J/n \to 0$. This completes the proof for (A16), and thus for Theorem A.1.

Proof of Theorem A.2: Put $\hat{W} = \sum_{l=1}^{n-1} d_J(l)\tilde{R}(l)/\tilde{R}(0)$. By (A14), we have

$$\begin{aligned}\pi\{\tilde{f}(0) - (2\pi)^{-1}\} &= \hat{W} + \{\tilde{R}(0)/R(0) - 1\}\hat{W} \\ &= \hat{W} + o_P(\hat{W})\end{aligned} \tag{A23}$$

given $\tilde{R}(0) - R(0) = O_P(n^{-\frac{1}{2}})$ by Assumption A.1 and H_0.

Write

$$\hat{W} = n^{-1}\sum_{t=2} W_t, \tag{A24}$$

where $W_t = R^{-1}(0)u_t \sum_{l=1}^{t-1} d_J(l)u_{t-l}$. Observe that $\{W_t, I_t\}$ is an adapted martingale difference sequence, we shall prove the asymptotic normality of $\hat{W}$ by the martingale theorem (e.g. Hall and Heyde 1980, pp.10-11). First, from (A24), we have

$$\begin{aligned}Var(n^{\frac{1}{2}}\hat{W}) &= R^{-2}(0)n^{-1}\sum_{t=2} EW_t^2 = \sum_{t=2}\sum_{l=1}^{t-1} d_J^2(l) \\ &= \sum_{l=1} (1 - l/n)d_J^2(l) \\ &= V_n(J).\end{aligned} \tag{A25}$$

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

$$V_n^{-2}(J)n^{-1}\sum_{t=2}^{n} W_t^2 1\{|W_t| > \epsilon n^{\frac{1}{2}}V_n^{\frac{1}{2}}(J)\} \text{ for any } \epsilon > 0; \qquad \text{(A26)}$$

and

$$V_n^{-2}(J)n^{-1}\sum_{t=2}^{n}\{E(W_t^2|I_{t-1}) - EW_t^2\} \to 0. \qquad \text{(A27)}$$

For sake of 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 (A25) and Lemma A.2, we shall verify condition (A26) by showing $2^{-2J}n^{-2}\sum_{t=1}^{n} EW_t^4 \to 0$. Put $\mu_4 = E(u_t^4)$. By Assumption A.1, we have

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

It follows that $2^{-2J}n^{-2}\sum_{t=1}^{n} EW_t^4 = O(n^{-1})$, ensuring condition (A26).

Given Lemma A.2, it suffices for (A27) if $2^{-2J}\mathrm{Var}\{n^{-1}\sum_{t=2}^{n} E(W_t^2|I_{t-1})\} \to 0$, which we now focus on. By the definition of W_t, we have

$$\begin{aligned} E(W_t^2|I_{t-1}) &= R^{-1}(0)\left\{\sum_{l=1}^{t-1} d_J(l)u_{t-l}\right\}^2 \\ &= EW_t^2 + R^{-1}(0)\sum_{l=1}^{t-1} d_J(l)\{u_{t-l}^2 - R(0)\} \\ &\quad + 2R^{-1}(0)\sum_{l=2}^{t-1}\sum_{h=1}^{l-1} d_J(l)d_J(h)u_{t-l}u_{t-h} \\ &= EW_t^2 + R^{-1}(0)A_t + 2R^{-1}(0)B_t; \end{aligned}$$

it follows that

$$\begin{aligned} n^{-1}\sum_{t=2}^{n}\{E(W_t^2|I_{t-1}) - EW_t^2\} &= R^{-1}(0)n^{-1}\sum_{t=2}^{n} A_t + 2R^{-1}(0)n^{-1}\sum_{t=2}^{n} B_t \\ &= R^{-1}(0)\hat{A} + 2R^{-1}(0)\hat{B}, \text{ say.} \end{aligned} \qquad \text{(A28)}$$

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

$$E\hat{A}^2 \le 2^{-2J}\left\{n^{-1}\sum_{t=2}^{n}(EA_t^2)^{\frac{1}{2}}\right\}^2 \le \{\mu_4 - R^2(0)\}\left\{\sum_{l=1}^{n-1} d_J^4(l)\right\} = O(2^J) \qquad \text{(A29)}$$

where $\sum_{l=1}^{n-1} d_J^4(l) = O(2^J)$ by Lemma A.1(iv).

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

$$\begin{aligned} EB_tB_s &= R^2(0)\sum_{l_2=2}^{t-1}\sum_{h_2=1}^{l_2-1}\sum_{l_1=2}^{s-1}\sum_{h_1=1}^{l_1-1} d_J(l_1)d_2(h_1)d_J(l_2)d_J(h_2)\delta_{t-h_1,s-h_2}\delta_{t-l_1,s-l_2} \\ &= R^2(0)\sum_{l=2}^{t-1}\sum_{h=1}^{l-1} d_J(t-s+l)d_J(t-s+h)d_J(l)d_J(h); \end{aligned}$$

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

$$\begin{aligned} E\hat{B}^2 &\le 2n^{-2}\sum_{t=3}^{n}\sum_{s=2}^{t-1} EB_tB_s \le 2R^2(0)n^{-1}\sum_{\tau=0}^{n-1}\sum_{l=2}^{n-1}\sum_{h=1}^{l-1} |d_J(\tau+l)d_J(\tau+h)d_J(l)d_J(h)| \\ &\le 2R^2(0)n^{-1}\left\{\sum_{\tau=0}^{n-1} d_J^2(\tau)\right\}\left\{\sum_{l=1}^{n-1}|d_J(l)|\right\}^2 \\ &= O(2^{3J}/n) \end{aligned} \qquad \text{(A30)}$$

where $\sum_{l=1}^{n-1}|d_J(l)|^r = O(2^J)$ for $r > 1$, by Lemma A.1(iv). Collecting (A28)-(A30) yields $2^{-2J}\{Var(\hat{A}) + Var(\hat{B})\} = O(2^{-J} + 2^J/n) \to 0$ given $J \to \infty, 2^J/n \to 0$ as $n \to \infty$. Thus, condition (A27) holds. By Hall and Heyde (1980,pp.10-11), $V_n^{-\frac{1}{2}}(J)n^{\frac{1}{2}}\hat{W} \to^d N(0,1)$.

Proof of Theorem 2: Put

$$\underline{R}(l) = \sum_{t=l+1}^{n}(\varepsilon_t^2/\sigma_0^2 - 1)(\varepsilon_{t-l}^2/\sigma_0^2 - 1) \qquad \text{(A31)}$$

where $\sigma_0^2 = E(\varepsilon_t^2)$ under $H_a(a_n)$. Note that we have $\varepsilon_t/\sigma_0 \ne \xi_t$ under H_a. Instead, we have

$$\varepsilon_t/\sigma_0 = \xi_t\{1 + a_n\sum_{l=1}^{\infty}\beta_l(\xi_{t-l}^2 - 1)\}^2. \qquad \text{(A32)}$$

We now define

$$\underline{f}(0) = \sum_{j=0}^{J}\sum_{k=1}^{2^j}\underline{\alpha}_{jk}\Psi_{jk}(0)$$

where $\alpha_{jk} = (2\pi)^{-1/2}\sum_{l=1-n}^{n-1}\tilde\rho(l)\hat\Psi_{jk}^{*}(l)$ and $\tilde\rho(l) = \tilde R(l)/R(0)$. Write $\hat f(0) - 1/2\pi = \hat f(0) - \tilde f(0) + \tilde f(0) - 1/2\pi$. The proof of Theorem 2 consists of Theorems A.3-A.4 below.

Theorem A.3: $V_n^{-1/2}(J)n^{1/2}\{\hat f(0) - \tilde f(0)\} \to^{p} 0$.

Theorem A.4: $V_n^{-1/2}(J)n^{1/2}\pi\{\tilde f(0) - 1/2\pi\} \to^{d} N(\mu, 1)$.

Proof of Theorem A.3: The proof is analogous to that for 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 for the case where $J \to \infty$. Because $\pi\{\tilde f(0) - (2\pi)^{-1}\} = \sum_{l=1}^{n-1} d_J(l)\tilde R(j)$, it suffices to show

$$V_n^{-1/2}(J)n^{1/2}\sum_{l=1}^{n-1} d_J(l)\tilde R(l)/R(0) \to^{d} N(\mu, 1), \qquad \text{(A33)}$$

where $\mu = V_0^{-\frac{1}{2}}\sum_{j=1}^{\infty}\beta_j$. Recall $u_t = \xi_t^2 - 1$ and put $V_t = \xi_t^2\sum_{j=1}^{\infty}\beta_j u_{t-j}$. By (A32)-(A33) and $H_a(a_n)$, we have

$$\begin{aligned}
\tilde R(l) &= n^{-1}\sum_{t=l+1}(\xi_t^2 h_t/\sigma_0^2 - 1)(\xi_{t-l}^2 h_{t-l}/\sigma_0^2 - 1)\\
&= n^{-1}\sum_{t=l+1}\{u_t + a_n V_t\}\{u_{t-l} + a_n V_{t-l}\}\\
&= n^{-1}\sum_{t=l+1}u_t u_{t-l} + a_n n^{-1}\sum_{t=l+1}V_t u_{t-l} + a_n n^{-1}\sum_{t=l+1}u_t V_{t-l} + a_n^2 n^{-1}\sum_{t=l+1}V_t V_{t-l}\\
&= \bar R(l) + a_n\hat A_4(l) + a_n\hat A_5(l) + a_n^2\hat A_6(l), \qquad \text{(A34)}
\end{aligned}$$

where $\bar R(l) = n^{-1}\sum_{t=|l|+1}u_t u_{t-|l|}$ as before. Put $V_t(l) = \xi_t^2\sum_{j=1, j\neq l}^{\infty}\beta_j u_{t-j}$. For the second term in (A34), we have

$$\begin{aligned}
\sum_{l=1}^{n-1} d_J(l)\hat A_4(l) &= \sum_{l=1}^{n-1} d_J(l)\left\{n^{-1}\sum_{t=l+1}\xi_t^2\sum_{j=1}\beta_j u_{t-j}u_{t-l}\right\}\\
&= R(0)\sum_{l=1}^{n-1} d_J(l)(1 - l/n)\beta_l\\
&\quad + \sum_{l=1}^{n-1} d_J(l)\beta_l\left\{n^{-1}\sum_{t=l+1}\xi_t^2\{u_{t-l}^2 - R(0)\}\right\}\\
&\quad + \sum_{l=1}^{n-1} d_J(l)\beta_l\left\{n^{-1}\sum_{t=l+1}V_t(l)u_{t-l}\right\}
\end{aligned}$$

$$= R(0)\sum_{l=1}^{\infty}\beta_l + O_P(2^{J/2}/n^{1/2}) \quad \text{(A35)}$$

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

$$\left|\sum_{l=1} d_J(l)\beta_l \left\{ n^{-1}\sum_{t=l+1} \xi_t^2\{u_{t-l}^2 - R(0)\}\right\}\right| = O_P(2^{J/2}/n^{1/2})$$

by the Cauchy-Schwarz inequality, Lemma A.1(iv); $\sum_{j=1}^{\infty}\beta_j^2 < \infty$ and

$$E\left|n^{-1}\sum_{t=l+1} \xi_t^2\{u_{t-l}^2 - R(0)\}\right|^2 \le Cn^{-1}$$

given Assumption A.1. Similarly, for the last term in (A34), we have

$$\sum_{l=1} d_J(l)\beta_l \left\{ n^{-1}\sum_{t=l+1} V_t(l)u_{t-l}\right\} = O_P(2^{J/2}/n^{1/2})$$

given independence between $V_t(l)$ and u_{t-l}. Moreover, we have

$$\sum_{l=1}^{n-1} d_J(l)\hat{A}_5(l) = O_P(2^{J/2}/n^{1/2}) \quad \text{(A36)}$$

by the Cauchy-Schwarz inequality, Lemma A.1(iv) and $E\hat{A}_4^2(l) \le Cn^{-1}$ given independence between u_t and V_{t-l} for $l > 0$.

For the last term $\hat{A}_6(l)$ in (A34), we put $R_V(l) = \mathrm{Cov}(V_t, V_{t-l})$ and $\bar{R}_V(l) = n^{-1}\sum_{t=l+1}^{n} V_tV_{t-l}$. Then

$$\sum_{l=1}^{n-1} d_J(l)\bar{R}_V(l) = \sum_{l=1}^{n-1} d_J(l)R_V(l) + \sum_{l=1}^{n-1} d_J(l)\{\bar{R}_V(l) - R_V(l)\}.$$

Because $V_t = \sum_{j=1}^{\infty}\beta_j(\xi_t^2u_{t-j})$ is a linear process with $\sum_{j=1}^{\infty}|\beta_j| < \infty$ and $E(\xi_t^2u_{t-j})^4 < \infty$, the cumulant condition $\sum_{j=-\infty}^{\infty}\sum_{m=-\infty}^{\infty}\sum_{l=-\infty}^{\infty}|k(0,j,m,l)| < \infty$, where $k(0,j,m,l)$ is the fourth order cumulant of $V_tV_{t+j}V_{t+m}V_{t+l}$ (e.g., Hannan 1970, p.211). It follows that $\sup_{0<l<n}\mathrm{Var}\{\bar{R}_V(l)\} \le Cn^{-1}$ by Hannan (1970, (5.1)). Consequently, we have

$$\left|\sum_{l=1}^{n-1} d_J(l)\{\bar{R}_V(l) - R_V(l)\}\right| \le \sum_{l=1}^{n-1} |d_J(l)|\left|\bar{R}_V(l) - R_V(l)\right| = O_P(2^{J/2}/n^{1/2})$$

by Markov's inequality and Lemma A.1(iv). On the other hand, $R_V(l)$ 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, $\hat{A}_6 = O_P(1)$: This, with (A34)-(A36), $a_n = (2^J/n)^{\frac{1}{2}}$ and $2^{2J}/n \to 0$; yields

$$\sum_{l=1}^{n-1} d_J(l)\hat{R}(l)/R(0) = \sum_{l=1}^{n-1} d_J(l)\tilde{R}(l)/R(0) + (2^J/n)^{\frac{1}{2}}\sum_{j=1}^{J}\xi_j + o_P(2^{J/2}/n^{1/2}):$$

Consequently, we have (A33) by Theorem 1. It follows that $\tilde{S}_n \to^d N(0,1)$ given $\tilde{V}_n(J)\to 2^{J+1}V_0\{1+o(1)\}$ and Slutsky theorem. This completes the proof.

Proof of Theorem 3: (i) We shall show for large J only; the proof for fixed J is similar. Here we explicitly denote $\hat{f}_J(0)$ as the spectral estimator (3.15) with the finest scale J: Recall the definition of $S_n(J)$; we have

$$\begin{aligned} S_n(\hat{J}) - S_n(J) &= V_n(\hat{J})^{-\frac{1}{2}}n^{\frac{1}{2}}\pi\{\hat{f}_{\hat{J}}(0) - (2\pi)^{-1}\} - V_n(J)^{-\frac{1}{2}}n^{\frac{1}{2}}\pi\{\hat{f}_J(0) - (2\pi)^{-1}\} \\ &= \{V_n(\hat{J})/V_n(J)\}^{\frac{1}{2}}V_n(J)^{-\frac{1}{2}}n^{\frac{1}{2}}\pi\{\hat{f}_{\hat{J}}(0) - \hat{f}_J(0)\} \\ &\quad + \{[V_n(J)/V_n(\hat{J})]^{-\frac{1}{2}} - 1\}S_n(J) \end{aligned}$$

Because $S_n(J) = O_P(1)$ by Theorem 1 and $V_n(\hat{J})/V_n(J) \to^p 1$ by Lemma A.4, we have $S_n(\hat{J}) - S_n(J) \to^p 0$ provided $V_n^{-\frac{1}{2}}(J)n^{\frac{1}{2}}\pi\{\hat{f}_{\hat{J}}(0) - \hat{f}_J(0)\} \to^p 0$; which we shall show below. (The asymptotic normality of $S_n(\hat{J})$ follows from $S_n(\hat{J}) - S_n(J) \to^p 0$ and Theorem 1.)

Because $V_n(J) = O(2^J)$; it suffices to show $\hat{f}_{\hat{J}}(0) - \hat{f}_J(0) = o_P(2^{J/2}/n^{1/2})$: Write

$$\begin{aligned} \pi\{\hat{f}_{\hat{J}}(0) - \hat{f}_J(0)\} &= \hat{R}^{-1}(0)\sum_{l=1}^{n-1}\{d_{\hat{J}}(l) - d_J(l)\}\{\hat{R}(l) - R(l)\} \\ &\quad + \hat{R}^{-1}(0)\sum_{l=1}^{n-1}\{d_{\hat{J}}(l) - d_J(l)\}R(l): \end{aligned} \tag{A37}$$

Given $|d_{\hat{J}}(l) - d_J(l)| \le \sum_{j=\min(\hat{J},J)}^{\max(\hat{J},J)}|\lambda(2\pi l/2^j)|$; we have, by Assumption A.5,

$$\begin{aligned} \sum_{l=1}^{n-1}|d_{\hat{J}}(l) - d_J(l)|\,|R(l)| &\le C\sum_{j=\min(\hat{J},J)}^{\max(\hat{J},J)}\sum_{l=1}^{m}(2\pi l/2^j)^q|R(l)| + C\sum_{j=\min(\hat{J},J)}^{\max(\hat{J},J)}\sum_{l=m+1}^{n-1}(2\pi l/2^j)^{-\tau}|R(l)| \\ &\le 2C|\hat{J} - J|2^{-Jq}2^{q\min(0,\hat{J}-J)}\left\{\sum_{l=1}^{m} l^q|R(l)|\right\} \\ &\quad + 2C|\hat{J} - J|2^{-J\tau}2^{\tau\max(0,\hat{J}-J)}\left\{\sum_{l=m+1}^{n-1} l^{-\tau}|R(l)|\right\} \\ &= |\hat{J} - J|\left\{2^{-Jq}O_P(m^{q+1}/n^{\frac{1}{2}}) + 2^{-J\tau}O_P(m^{1-\tau}n^{-\frac{1}{2}})\right\} \\ &= o_P(2^{\frac{J}{2}}/n^{\frac{1}{2}}) \end{aligned} \tag{A38}$$

by choosing $m = 2^J$ and using $\hat{J} - J = o_P(2^{-J/2})$, where $\sum_{l=1}^{m} l^q |\bar{R}(l)| = O_P(m^{q+1}/n^{\frac{1}{2}})$ and $\sum_{l=m+1}^{n-1} l^{-\tau} |\bar{R}(l)| = O_P(m^{1-\tau}/n^{\frac{1}{2}})$ by Markov's inequality, $\tau > 1$ in Assumption A.3 and $E|\bar{R}(l)| = O(n^{-\frac{1}{2}})$.

Next, following reasoning analogous to that of (A16), we can obtain

$$\sum_{l=1}^{n-1} d_{\hat{J}}(l)\{\hat{R}(l) - \bar{R}(l)\} = o_P(2^J/n) \tag{A39}$$

given $\hat{J} - J = o_P(1)$, $\sum_{l=1}^{n-1} d_{\hat{J}}(l) = \{\sum_{l=1}^{n-1} d_J(l)\}\{1 + o_P(1)\}$ and $\sum_{l=1}^{n-1} d_J(l) = O(2^J)$ by Lemma A.1. Combining (A37)-(A39) and (A16), we obtain $\hat{f}_{\hat{J}}(0) - \hat{f}_J(0) = o_P(2^{J/2}/n^{1/2})$. This completes the proof.

Table 1: Size at the 10 % and 5 % Levels

	n = 100		n = 200	
	10%	5%	10%	5%
S_1	8.9	5.8	7.6	4.3
S_2	9.0	5.4	8.1	4.2
K	8.4	4.0	7.5	3.5
LK(1)	9.8	5.0	8.5	4.0
DS(1)	9.9	5.3	8.3	4.1
LM(1)	7.8	4.0	8.0	4.3
LK(12)	6.3	2.8	6.8	3.0
DS(12)	11.6	7.3	11.5	6.7
LM(12)	6.1	1.9	7.1	3.2

1) Model: $Y_t = 1 + m_t + "_t$; $m_t = 0{:}8m_{t\,¡\,1} + À_t$;
$À_t$ s NID(0; 4); $"_t = »_t h_t^{1=2}$; $»_t$ s NID(0; 1); $h_t = 1$:

2) 1000 iterations.

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	56.4	41.7	71.1	62.3	87.5	82.5	97.6	96.5
S_2	55.9	43.5	70.1	62.1	87.9	82.4	96.1	94.7
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

1) Model: $Y_t = 1 + m_t + \varepsilon_t$; $m_t = 0.8 m_{t-1} + \nu_t$; $\nu_t \sim NID(0;4)$; $\varepsilon_t = \xi_t h_t^{1/2}$; $\xi_t \sim NID(0;1)$; $h_t = 1 + \beta \varepsilon_{t-1}^2$.

2) 1000 iterations.

Table 3: Size-adjusted Power against ARCH(12a) and ARCH(12b) 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	59.1	43.3	88.4	82.9	76.6	62.1	95.8	92.9
S_2	60.1	51.2	89.3	86.4	77.2	66.6	93.0	90.8
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

1) Model: $Y_t = 1 + m_t + \varepsilon_t$; $m_t = 0.8 m_{t-1} + \psi_t$; $\psi_t \sim NID(0, 4)$; $\varepsilon_t = \xi_t h_t^{1/2}$; $\xi_t \sim NID(0, 1)$.

2) ARCH(12a): $h_t = 1 + \beta \sum_{j=1}^{12} \varepsilon_{t-j}^2$; $\beta = 0.95/12$.

3) ARCH(12b): $h_t = 1 + \beta \sum_{j=1}^{12} (1 - j/13) \varepsilon_{t-j}^2$; $\beta = 0.95 / \sum_{j=1}^{12} (1 - j/13)$.

4) 1000 iterations.

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	70.7	57.5	86.9	81.0	88.6	78.1	98.5	97.8
S_2	67.9	58.2	83.9	77.6	85.2	77.3	94.3	93.1
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

1) Model: $Y_t = 1 + m_t + \varepsilon_t$; $m_t = 0.8m_{t-1} + \upsilon_t$; $\upsilon_t \sim NID(0; 4)$; $\varepsilon_t = \xi_t h_t^{1/2}$; $\xi_t \sim NID(0; 1)$; $h_t = 1 + \alpha\varepsilon_{t-1}^2 + \beta h_{t-1}$.

2) 1000 iterations.

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	7.3	4.2	6.8	4.3	6.2	2.8	5.6	3.3
Power								
ARCH(1): ¯ = 0:3	60.3	50.3	60.1	50.9	81.5	73.4	81.4	72.9
ARCH(1): ¯ = 0:95	93.4	89.9	93.3	90.3	99.6	98.8	99.7	98.9
ARCH 12(a)	48.2	39.7	54.5	47.9	81.7	76.9	87.3	84.7
ARCH 12(b)	67.1	59.6	70.5	65.8	92.0	88.7	94.9	93.4
GARCH(1,1): (0:3; 0:3)	69.5	61.5	69.9	63.2	89.9	85.5	89.5	85.9
GARCH(1,1): (0:3; 0:65)	83.1	76.7	84.1	78.8	97.7	96.4	98.2	97.7
Size-Adjusted Power								
ARCH(1): ¯ = 0:3	66.0	54.0	67.2	55.0	87.7	79.8	86.4	79.5
ARCH(1): ¯ = 0:95	95.8	91.9	96.4	92.0	99.9	99.5	99.9	99.6
ARCH 12(a)	54.2	43.1	59.0	50.4	85.9	80.4	89.1	86.7
ARCH 12(b)	71.8	62.0	76.1	67.7	94.5	90.8	95.8	94.7
GARCH(1,1): (0:3; 0:3)	74.6	64.7	75.8	66.0	92.8	89.1	93.2	88.9
GARCH(1,1): (0:3; 0:65)	86.3	79.2	87.5	80.8	98.2	97.3	99.0	98.0

1) BF_1; Bonferoni procedure combining S_1 and K; BF_2; Bonferoni procedure consisting of S_2 and K:

2) The size-adjusted power of BF_1 and BF_2 is based on their empirical p-values under H_0:

3) 1000 iterations.