ONE-SIDED TESTING FOR CONDITIONAL HETEROSKEDASTICITY IN TIME SERIES MODELS

By Yongmiao Hong

Cornell University

First version received February 1996

Abstract. Engle's autoregressive conditional heteroskedasticity (ARCH) model and its various generalizations have been widely used to model the volatility of economic and financial time series. Most existing ARCH tests fail to exploit the one-sided nature of the alternative hypothesis. Lee and King (A locally most mean powerful based score test for ARCH and GARCH regression disturbances. *J. Bus. Econ. Stat.* 11 (1993), 17–27) recently proposed a locally most mean powerful score-based one-sided test for ARCH effects. In this paper a new one-sided test for ARCH effects of the disturbance of a dynamic regression model is proposed. The test is based on a weighted sum of sample autocorrelations of squared regression residuals, with the weighting function typically giving more weight to lower orders of lags and less weight to higher orders of lags. Lee and King's (1993) test can be viewed as a special case of the present approach with the use of uniform weighting. Many non-uniform weighting schemes deliver better power than uniform weighting; the efficiency gain is substantial when a relatively long lag is used. A simulation experiment confirms the gains from exploiting the one-sided nature of the alternative hypothesis and from using non-uniform weighting.

Keywords. ARCH; frequency domain analysis; one-sided testing for multiparameter hypotheses; weighting.

1. INTRODUCTION

There has been considerable interest in estimating and testing dynamic conditional heteroskedasticity of the regression disturbance since Engle (1982) introduced the autoregressive conditional heteroskedasticity (ARCH) model. The ARCH model and its various generalizations (e.g. Bollerslev's (1986) generalized ARCH or GARCH, Higgins and Bera's (1992) nonlinear ARCH, Nelson's (1991) exponential GARCH, Sentana's (1995) quadratic ARCH) have been widely used to the model volatility of economic and financial time series. See Bera and Higgins (1993) and Bollerslev *et al*. (1992, 1994) for recent surveys.

In this paper, we propose a one-sided test for ARCH effects. Detection of ARCH effects is important. From the perspective of econometric inference, the neglect of ARCH effects may lead to arbitrarily large loss in asymptotic efficiency of parameter estimation (Engle, 1982); it also causes over-rejection of conventional tests for serial correlation in mean such as those of Box and Pierce (1970) and Ljung and Box (1978) (e.g. Taylor, 1984; Milhoj, 1985;

0143-9782/97/03 253–277 JOURNAL OF TIME SERIES ANALYSIS Vol. 18, No. 3

Diebold, 1987). Weiss (1984) pointed out that ignoring ARCH effects will result in overparameterization of an ARMA model. In practice, one is often concerned with testing for the existence of ARCH effects and then proceeding to investigate the model's dynamic nature. For example, as pointed out by Engle and Susmel (1993), it is necessary to test for the existence of ARCH effects before testing and modeling volatility spillover across different time series.

The most popular ARCH test is Engle's (1982) Lagrange multiplier (LM) test for ARCH(q). This test is simple to calculate and is asymptotically locally most powerful, a characterization it shares with the likelihood ratio and Wald tests. Lee (1991) showed that a modified LM test for GARCH(p, q) is the same as the LM test for ARCH(q). The portmanteau tests of Box and Pierce (1970) and Ljung and Box (1978) are also widely used (see also McLeod and Li, 1983). These tests are asymptotically equivalent to the LM test (e.g. Granger and Terasvirta, 1993, pp. 93–94). Other ARCH tests include those of Bera and Higgins (1992), Gregory (1989), Hong and Shehadeh (1997), Lee (1991), Lee and King (1993), Robinson (1991a) and Weiss (1986).

The one-sided nature of the ARCH alternative hypothesis has been known for a long time. However, most existing tests fail to exploit this. A one-sided test is expected to yield better power in small samples. Engle *et al*. (1985, p. 75) suggested a one-sided test for ARCH(1) by using the squared root of the LM test with an appropriate sign. This approach, however, cannot be extended to test higher-order ARCH(q) alternatives. Recognizing that the LM test fails to exploit the one-sided nature of ARCH alternatives, Lee and King (1993) proposed a one-sided locally most mean powerful score-based (LBS) test for ARCH(q) and GARCH(q), by applying SenGupta and Vermeire's (1986) approach for testing one-sided multiparameter problems. The test is based on the sum of the scores of the data evaluated under the null hypothesis. It is essentially based on the sum of the first q sample autocorrelations of squared ordinary least squares residuals. Lee and King (1993) showed in simulation that this test has better power than the LM test in small samples, illustrating the gain of exploiting the one-sided nature of the ARCH alternative. Sentana (1995, p. 644) noted that Demos and Sentana, in a 1994 working paper, also considered one-sided testing for GARCH effects by using an alternative approach.

Lee and King's LBS test gives equal weight for each of the q sample autocorrelations. Because economic agents normally discount past information, most ARCH processes have autocorrelations that decay to zero as the lag increases. Indeed, the key feature of volatility clustering is that large volatility changes tend to be followed by large volatility changes and periods of tranquility alternate with periods of high volatility. This implies that the recent past volatility has a larger impact on the current volatility than the distant past volatility. Therefore, it seems to be more efficient to give more weight to lower orders of lags and less weight to higher orders of lags. However, although Lee and King showed that the LBS test for GARCH(p, q) was identical to the LBS test for ARCH(q), simulation (Lee and King, 1993) showed that the power of the LBS test is hardly affected by changes in the coefficients associated with lagged conditional variances. In fact, the LBS test has better power for highly persistent GARCH(1, 1) if more lags (i.e. $q > 1$) are included in the test statistic (see Section 5 below). This suggests that it may be desirable to let the number of lags q grow with the sample size when testing persistent GARCH processes and strongly dependent processes whose autocorrelations decay to zero slowly (e.g. Robinson, 1991a; Baillie and Bollerslev, 1993; Baillie *et al.* 1993). The use of long lags is expected to have good power against persistent or strongly dependent alternatives.

In this paper, we propose an ARCH test that exploits the one-sided nature of the alternative hypothesis, permits flexible weights for sample autocorrelations, and allows for the number of lags to grow with the sample size. We use a frequency domain approach. It turns out that our test is based on a weighted sum of sample autocorrelations of squared residuals, with the weighting function typically giving more weight to lower orders of lags and less weight to higher orders of lags. Under appropriate conditions, our test has a one-sided asymptotic $N(0, 1)$ distribution under the null hypothesis of no ARCH and diverges to positive infinity as the sample size increases under the alternative. The test is expected to have good power against strong-dependence alternatives. Lee and King's (1993) LBS test can be viewed as a special case of our approach because it corresponds to the use of uniform weighting that gives equal weight to each of the q sample autocorrelations. Non-uniform weighting delivers better power than uniform weighting; the efficiency gain is substantial for relatively large q. In addition, our approach leads to a natural choice of q by such data-driven methods as those of Andrews (1991), Beltrão and Bloomfield (1987) and Robinson (1991b). Simulation shows that these data-driven methods deliver reasonable size and power.

Lee and King's (1993) approach is discussed in Section 2. In Section 3, we introduce our statistic and derive its asymptotic null distribution. In Section 4, we obtain the asymptotic local power of our test and discuss the relative efficiency of various weighting functions. Automatic choices of q via data-driven methods are also discussed. In Section 5, a simulation experiment is conducted to investigate the finite sample performances of our test in comparison with Lee and King's test (1993) and Engle's (1982) LM test. The last section concludes the paper. All mathematical proofs are collected in the Appendix.

2. THE LBS TEST

Suppose $\{Y_t\}$ is a stationary process such that

$$Y_t = X_t' b_0 + \varepsilon_t \qquad t = 1, \ldots, T \tag{1}$$

with an ARCH error

$$\varepsilon_t = \xi_t h_t^{1/2}.$$

Here, X_t is a vector consisting of exogenous variables and lagged dependent variables, b_0 is a finite-dimensional parameter, ξ_t is an identically and independently distributed (i.i.d.) sequence with $E(\xi_t) = 0$ and $E(\xi_t^2) = 1$, and h_t is a positive time-varying measurable function with respect to the information set I_{t-1}, available at period $t - 1$.

By construction, $\{\varepsilon_t, I_t\}$ is a martingale difference sequence and $h_t = E(\varepsilon_t^2 | I_{t-1})$ is the condition variance of ε_t. A well-known functional form for h_t is Engle's (1982) ARCH(q) process

$$h_t = \alpha_0 + \sum_{j=1}^{q} \alpha_j \varepsilon_{t-j}^2$$

where q is a fixed integer. To ensure that h_t is strictly positive for all realizations of ε_t, it is required that $\alpha_0 > 0$ and $\alpha_j \geqslant 0$ for $j = 1, \ldots, q$.

Another popular functional form for h_t is Bollerslev's (1986) GARCH(p, q) process

$$h_t = \alpha_0 + \sum_{j=1}^{q} \alpha_j \varepsilon_{t-j}^2 + \sum_{j=1}^{p} \beta_j h_{t-j}^2.$$

To ensure that h_t is strictly positive for all realizations of ε_t, one normally requires that $\alpha_0 > 0$, $\alpha_j \geqslant 0$, $j = 1, \ldots, q$, and $\beta_j \geqslant 0$, $j = 1, \ldots, p$. Nelson and Cao (1992) showed that these constraints may be weakened for processes of higher order than GARCH(1, 1). See also Drost and Nijman (1993).

To test for ARCH(q), the hypotheses of interest are

$$\mathrm{H}_0\colon \alpha_j = 0 \text{ for } j = 1, \ldots, q$$

versus

$$\mathrm{H}_\mathrm{A}\colon \alpha_j \geqslant 0,\ j = 1, \ldots, q, \text{ with at least one strict inequality.}$$

Engle's (1982) LM test for ARCH(q) can be obtained as TR^2, where R^2 is the squared multicorrelation coefficient of the regression

$$e_t^2 = \alpha_0 + \sum_{j=1}^{q} \alpha_j e_{t-j}^2 + v_t$$

where e_t^2 is the ordinary least squares residual from (1). This statistic is asymptotically χ_q^2 under H_0. However, it is a test for H_0 against a two-sided alternative.

Exploiting the one-sided nature of H_A is expected to yield better power in small samples. For $q = 1$, Engle *et al.* (1985) suggested that a one-sided test for ARCH(1) can be obtained by using the squared root of the LM test with an appropriate sign, but this approach cannot be extended to general cases for $q > 1$.

SenGupta and Vermeire (1986) introduced a class of locally most mean powerful unbiased tests for multiparameter hypotheses. These tests maximize the mean curvature of the power function in the neighborhood of the null hypothesis. Lee and King (1993) applied this approach to construct a one-sided test for H_0. More specifically, Lee and King's test is based on the sum of scores evaluated under H_0, namely,

$$\hat{s} = \sum_{j=1}^{q} \frac{\partial}{\partial \alpha_j} L_T(\alpha | e^2)|_{\alpha=0}$$

where $L_T(\alpha|e^2)$ is the log-likelihood function conditional on the sample $e^2 = (e_1^2, \ldots, e_T^2)'$, and $\alpha = (\alpha_1, \ldots, \alpha_q)'$. Lee and King's test statistic is a proper standardized version of $\hat{s}$:

$$\mathrm{LBS}(q) = \hat{s}/(l'\hat{J}_{\alpha\alpha} l)^{1/2}$$

where l is the $q \times 1$ vector of ones, and $\hat{J}_{\alpha\alpha}^{-1} = \nabla_\alpha^2 L_T(\alpha|e^2)|_{\alpha=0}$ is the sample information matrix evaluated under H_0. Under H_0, LBS(q) is asymptotically N(0, 1). For concreteness, Lee and King (1993) assumed that ξ_t is N(0, 1). This yields

$$\mathrm{LBS}(q) = \hat{V}^{-1/2} \sum_t \left(\frac{e_t^2}{\hat{\sigma}^2} - 1 \right) \sum_{j=1}^{q} e_{t-j}^2 \tag{2}$$

where $\hat{\sigma}^2 = T^{-1} \sum_{t=1}^{T} e_t^2$ and

$$\hat{V} = 2 \left\{ \sum_t \left(\sum_{i=1}^{q} e_{t-i}^2 \right)^2 - \left(\sum_t \sum_{i=1}^{q} e_{t-i}^2 \right)^2 \Big/ (T-q) \right\}.$$

This test is not robust to non-normality. Lee and King also proposed a modified version that is robust to non-normality, by using the variance estimator

$$\hat{V}_T = \left\{ (T-q)^{-1} \sum_t \left(\frac{e_t^2}{\hat{\sigma}^2} - 1 \right)^2 \right\} \left\{ \sum_t \left(\sum_{i=1}^{q} e_{t-i}^2 \right)^2 - \left(\sum_t \sum_{i=1}^{q} e_{t-i}^2 \right)^2 \Big/ (T-q) \right\}.$$

Furthermore, Lee and King (1993) showed that the LBS test against GARCH(p, q) is identical to the LBS against ARCH(q), because the sum of the scores evaluated under H_0 is precisely the same as that for the LBS test against ARCH(q). In their simulation, Lee and King (1993) showed that their test has better power than Engle's (1982) LM test. See also Section 5 below.

The LBS test (2) is essentially based on the sum of the first q sample autocorrelations of the squared residual, where q is a fixed integer. Obviously, the LBS test gives equal weight for each of the q sample autocorrelations. This may not be fully efficient against the alternatives whose autocorrelations decay to zero as the lag increases, which are often encountered in practice because economic agents usually discount past information. Indeed, a stylized fact for the volatility clustering of most financial time series is that high volatility 'today' tends to be followed by a similar volatility 'tomorrow' and vice versa. This implies that the recent past volatility has bigger impact on the current volatility than the distant past volatility. For such alternatives, a more efficient test may be obtained by giving more weight to lower orders of lags and less weight to higher orders of lags. On the other hand, when the LBS test is applied to test GARCH(p, q), only the first q sample autocorrelations are used, no matter how strong the persistence of GARCH effects is. We expect that this might not be optimal for testing highly persistent GARCH(p, q) and strongly dependent processes, whose autocorrelations decay to zero slowly. Here, it is better to include long lags. Below, we construct a one-sided test that uses non-uniform weights and allows for q to grow with the sample size.

3. A NEW ONE-SIDED ARCH TEST

We consider the following data-generating process.

ASSUMPTION A1. The stochastic process $\{Y_t\}$ is given by

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

where X_t is a vector consisting of exogenous variables and lagged dependent variables, b_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}. The innovation $\{\xi_t\}$ is an i.i.d. sequence with $E(\xi_t) = 0$, $E(\xi_t^2) = 1$ and $E(\xi_t^8) < \infty$. Also, ξ_t is independent of X_s for all $t \geqslant s$.

ASSUMPTION A2. (a) For each b, $g(\cdot, b)$ is a measurable function with respect to I_{t-1}; (b) $g(X_t, \cdot)$ is twice differentiable with respect to b in an open convex neighborhood $N(b_0)$ of b_0 almost surely, with

$$\lim_{T\to\infty} E \sup_{b\in N(b_0)} \left\{ T^{-1} \sum_{t=1}^{T} \|\nabla_b g(X_t, b)\|^4 \right\} < \infty$$

and

$$\lim_{T\to\infty} E \sup_{b\in N(b_0)} \left\{ T^{-1} \sum_{t=1}^{T} \|\nabla_b^2 g(X_t, b)\|^2 \right\} < \infty,$$

where ∇_b and ∇_b^2 are the gradient and Hessian operators respectively.

Note that we make no specific distribution (e.g. normality) assumptions on the innovation $\{\xi_t\}$ beyond the regularity moment conditions. Our extension to nonlinear regression models is straightforward, but it allows that the X_t are trending variables or nonstationary time series when $g(\cdot, \cdot)$ is an appropriate nonlinear functional form.

Suppose that h_t follows a general linear process

$$h_t = \alpha_0 + \sum_{j=1}^{\infty} \alpha_j \varepsilon_{t-j}^2 \qquad (3)$$

where, to ensure strict positiveness of h_t, $\alpha_0 > 0$ and $\alpha_j \geqslant 0$ for all $j = 1, \ldots, \infty$ (cf. Nelson and Cao, 1992). This process includes Engle's (1982) ARCH(q), Bollerslev's (1986) GARCH(p, q) and Engle and Bollerslev's (1986) integrated GARCH (IGARCH) and fractionally differenced GARCH processes. For ARCH(q), $\alpha_j = 0$ for $j > q$; for stationary GARCH(p, q), the α_j decay exponentially; for fractionally differenced GARCH, the α_j decay at a slow geometric rate. We note that Robinson (1991a) considered (3) in constructing a general class of LM tests that includes Engle's (1982) and Weiss's (1986) LM tests for ARCH(q) and many new ones designed to have good power against long memory processes. There, some parametric restrictions on the α_j were imposed because the LM principle was used.

Given (3), we consider the hypotheses of interest

$$\mathrm{H}_0\colon \alpha_j = 0 \text{ for all } j = 1, 2, \ldots$$

versus

$$\mathrm{H}_\mathrm{A}\colon \alpha_j \geqslant 0 \text{ for all } j = 1, 2, \ldots, \text{ with at least one strict inequality.}$$

Let $f(\omega)$ be the normalized spectral density of ε_t^2, where $\omega \in (-\pi, \pi)$. Then, for (3),

$$f(\omega) = \frac{1}{2\pi} |1 - \sum_{j=0}^{\infty} \alpha_j \exp(ij\omega)|^{-2} \qquad \omega \in (-\pi, \pi)$$

where $i = \sqrt{(-1)}$. We observe that $f(0) = 1/2\pi$ under H_0 and $f(0) > 1/2\pi$ under H_A. This observation forms the basis for our test, namely, a one-sided test can be obtained by comparing an estimator of $f(0)$ to $1/2\pi$. If an estimate of $f(0)$ is close to $1/2\pi$, then there is no ARCH; when ARCH effects exist, the estimate will be significantly larger than $1/2\pi$ asymptotically, thus delivering a one-sided test for H_0. In addition, a test based on an estimator of $f(0)$ is expected to be particularly powerful against strong-dependence alternatives, because their spectral densities are positive infinity at frequency zero (cf. Robinson, 1991a, 1994). Hence, our approach yields a test complementary to those of Robinson (1991a) in detecting strong-dependence alternatives. Granger (1969) pointed out that most economic time series typically have a spectral density that has a peak at frequency zero and then decays to zero as the frequency increases.

To obtain an estimate of $f(0)$, we define the regression residual

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

where $\hat{b}$ is a $\sqrt{T}$-consistent estimator for b_0.

Assumption A3. $\hat{b} - b_0 = \mathrm{O}_P(T^{-1/2})$.

One example of $\hat{b}$ is the nonlinear least squares estimator, namely,

$$\hat{b} = \arg\min_b T^{-1} \sum_{t=1}^{T} \{Y_t - g(X_t, b)\}^2.$$

Such an estimator ignores possible ARCH effects and thus may not be efficient, but it is still consistent for b_0 under both H_0 and H_A. Now define the sample autocorrelation function of $\hat{\varepsilon}_t^2$ as

$$\hat{\rho}(j) = \frac{\hat{R}(j)}{\hat{R}(0)} \qquad j = 0, 1, \ldots, T-1$$

where the autocovariance

$$\hat{R}(j) = T^{-1} \sum_{t=j+1}^{T} \left(\frac{\hat{\varepsilon}_t^2}{\hat{\sigma}^2} - 1\right)\left(\frac{\hat{\varepsilon}_{t-j}^2}{\hat{\sigma}^2} - 1\right)$$

with $\hat{\sigma}^2 = T^{-1} \sum_{t=1}^{T} \hat{\varepsilon}_t^2$. A kernel-based estimator for $f(0)$ is given by

$$\hat{f}(0) = \frac{1}{2\pi} + \frac{1}{\pi} \sum_{j=1}^{T-1} k\left(\frac{j}{q}\right) \hat{\rho}(j)$$

where q is a positive integer and k is a kernel function satisfying Assumption A4.

Assumption A4. The function k: $\mathrm{R} \rightarrow [-1, 1]$ is symmetric and continuous at zero and all but a finite number of points, with $k(0) = 1$ and $\int_0^\infty |k(z)|\, dz < \infty$. In addition, $0 < \sum_{j=1}^{\infty} k^2(j/q) < \infty$ for any positive finite integer q.

Most commonly used kernels typically give more weight to lower orders of lags and less weight to higher orders of lags. The exception is the truncated kernel $k(z) = 1$ for $|z| \leqslant 1$ and $k(z) = 0$ otherwise, which gives equal weight for each lag. As will be seen below, non-uniform kernels deliver better power than the truncated kernel.

Our test S (say) is based on comparison between $\hat{f}(0)$ and $1/2\pi$:

$$\begin{aligned} S &= V_T^{-1/2} T^{1/2} \pi \{\hat{f}(0) - 1/2\pi\} \\ &= V_T^{-1/2} T^{1/2} \sum_{j=1}^{T-1} k\left(\frac{j}{q}\right) \hat{\rho}(j) \end{aligned}$$

where

$$V_T = \sum_{j=1}^{T-1} \left(1 - \frac{j}{T}\right) k^2\left(\frac{j}{q}\right)$$

is approximately the variance of $T^{1/2}\pi\{\hat{f}(0) - 1/2\pi\}$. The factor $1 - j/T$ can be viewed as a finite sample correction. For large q, we can use $q \int_0^\infty k^2(z)\, dz$ to substitute for V_T.

THEOREM 1. *Suppose Assumptions A1–A4 hold, and let* $q/T \to 0$. *Then* $S \to \mathrm{N}(0, 1)$ *in distribution under* H_0.

The proof is given in the Appendix. We note that S is a one-sided test because S will diverge to $+\infty$ under H_A. Asymptotically, negative values of S occur only under H_0. Therefore, appropriate one-sided upper-tailed N(0, 1) critical values should be used.

The condition $q/T \to 0$ is very weak. It permits q to be fixed, or to grow with the sample size but at a slower rate. Automatic choices of q via data-driven methods are discussed in Section 4.

When the truncated kernel is used, our approach yields

$$S_{\mathrm{TRUN}} = \left(\frac{T}{q}\right)^{1/2} \sum_{j=1}^{q} \hat{\rho}(j) \tag{4}$$

where we have used q for V_T. This is essentially Lee and King's LBS test (2) for ARCH(q) and GARCH(p, q). The differences are that we permit (but do not require) q to grow with the sample size and use a different variance estimator. Thus, Lee and King's test can be interpreted as a special case of our approach with the use of the truncated kernel. As will be seen below, many non-uniform kernels have better power than the truncated kernel. Simulation also shows that the use of the different variance estimator in (4) gives better power than Lee and King's test.

4. ASYMPTOTIC LOCAL POWER

We now evaluate the asymptotic power of the test under the following sequence of local alternatives:

$$\mathrm{H}_\mathrm{a}\colon\ h_t = \sigma_0^2 \left\{ 1 + a_T \sum_{j=1}^{\infty} \beta_j (\xi_{t-j}^2 - 1) \right\}$$

where $\beta_j \geqslant 0$, $\sum_{j=1}^{\infty} \beta_j < \infty$, and $a_T \to 0$. To ensure positivity of h_t, we assume $a_T \sum_{j=1}^{\infty} \beta_j < 1$ for all $T \geqslant 1$.

THEOREM 2. *Suppose Assumptions A1–A4 hold, and let* $q^2/T \to 0$. *Then*

$$S \to N(\mu, 1) \quad \textit{in distribution}$$

under H_a *with* $a_T = (q/T)^{1/2}$, *where*

$$\mu = \sum_{j=1}^{\infty} k\left(\frac{j}{q}\right)\beta_j \Big/ \left\{ q^{-1} \sum_{j=1}^{\infty} k^2\left(\frac{j}{q}\right) \right\}^{1/2}. \tag{5}$$

If in addition $q \to \infty$, *then*

$$\mu = \sum_{j=1}^{\infty} \beta_j \Big/ \left\{ \int_0^{\infty} k^2(z)\, dz \right\}^{1/2}. \tag{6}$$

The proof is given in the Appendix. By Theorem 2, the asymptotic local power of S is $\lim_{T\to\infty} P(S > C_\alpha) = 1 - \Phi(C_\alpha - \mu)$, where C_α is the asymptotic upper-tailed N(0, 1) critical value at level $\alpha \in (0, 1)$, and Φ is the cumulative distribution function of N(0, 1).

We now compare the efficiencies of various weighting functions. First, we consider fixed q. In this case, we assume $\beta_j = 0$ for $j > q$. The optimal weight, to a proportionality, is $k(j/q) = \beta_j / \sum_{j=1}^{q} \beta_j$; this maximizes the noncentrality μ in (5), yielding the best asymptotic local power. Obviously, the optimal weight depends on the alternative $\{\beta_j\}$. If the β_j decreases as the lag j increases, the optimal weight should also decrease as j increases. When $\beta_j = \beta$ for $j = 1, \ldots, q$, the optimal weight is uniform weighting. Lee and King's (1993) test (2) or the truncated-kernel-based test (4) has the best power against this uniform alternative.

The conclusions on the relative efficiency of various weighting functions change if $q \to \infty$ as the sample size $T \to \infty$. In this case, the noncentrality μ is given by (6). The optimal weight minimizes $\int_0^\infty k^2(z)\, dz$ over an appropriate class of k, and this optimal kernel does not depend on the alternative. The uniform weighting function or the truncated kernel is no longer optimal. This is so even when $\beta_j = \beta$ for $j = 1, \ldots, q_0$ and $\beta_j = 0$ for $j > q_0$, where q_0 is a fixed integer. To see this, we compare, for example, the Bartlett kernel $k_{\mathrm{BAR}}(z) = (1 - |z|)1(|z| \leqslant 1)$ and the truncated kernel $k_{\mathrm{TRUN}}(z) = 1(|z| \leqslant 1)$, where $1(\cdot)$ denotes the indicator function. Let $q = cT^{v}$, for $0 < v < 1/2$ and $0 < c < \infty$. Following Pitman (1979), we obtain Pitman's relative efficiency of k_2 with respect to k_1

$$\mathrm{REF}(k_2\colon k_1) = \left\{ \int_0^{\infty} k_1^2(z)\, dz \Big/ \int_0^{\infty} k_2^2(z)\, dz \right\}^{1/(1-v)}.$$

We have $\mathrm{REF}(k_{\mathrm{BAR}}\colon k_{\mathrm{TRUN}}) = 3^{1/(1-v)} > 3$ for any $0 < v < 1/2$. Thus, the Bartlett kernel is three times as efficient as the truncated kernel for large q. In fact, most non-uniform kernels are more efficient than the truncated kernel. Therefore, we expect that for large q our test with non-uniform kernels will be more efficient than Lee and King's LBS test.

We now consider the optimal weighting function that maximizes the power of S over a suitable class of kernels. Let $r > 0$ be the largest positive integer such that

$$k^{(r)} = \lim_{z\to 0} \left\{ \frac{1 - k(z)}{|z|^r} \right\}$$

exists and is finite and non-zero. We consider the class of kernels with $r = 2$:

$$\mathscr{K}(\tau) = \left\{ k\colon k \text{ satisfies A4, } k^{(z)} = \tfrac{1}{2}\tau^2, \text{ and } \int_{-\infty}^{\infty} k(z) \exp(i\lambda z)\, dz \geqslant 0 \text{ for all } \lambda \right\}. \tag{7}$$

It is well known (e.g. Priestley, 1962; Andrews, 1991) that the quadratic-spectral (QS) kernel

$$k(z) = \left(\frac{3}{5z^2}\right)\left\{\frac{\sin(5^{1/2}z)}{5^{1/2}z} - \cos(5^{1/2}z)\right\} \qquad z \in (-\infty, \infty)$$

minimizes $\int_0^\infty k^2(z)\, dz$ over $\mathscr{K}(\tau)$. Therefore, the QS kernel maximizes the power of the S test over $\mathscr{K}(\tau)$. This kernel is also optimal for spectral density estimation in terms of the mean squared error (MSE) criterion. The class (7) includes the Daniell, Parzen and QS kernels, but rules out the truncated and Bartlett kernels. We emphasize that the optimal weighting function crucially depends on the form of the test statistic. For example, Hong (1996) showed that, for some frequency-domain-based tests in a different context, the Daniell kernel (rather than the QS kernel) is optimal within the class $\mathscr{K}(\tau)$.

For all ARCH tests, the choice of q is important. Because generally no prior information is available in practice, applied workers usually try several, perhaps many, different q. It is not uncommon that some of these test statistics are significant but some are not. Thus, it is a delicate business to determine the overall significance level, because these statistics are not independent. In the present context, our frequency domain approach leads to a natural choice of q via data-driven methods. For spectral density estimation, there are basically two types of data-driven methods: the cross-validation method (e.g. Wahba and Wold, 1975; Beltrão and Bloomfield, 1987; Hurvich, 1985; Robinson, 1991b) and the narrow-band method (e.g. Andrews, 1991; Newey and West, 1994). Most cross-validation methods choose q to minimize an appropriate MSE criterion. Beltrão and Bloomfield's (1987) method chooses q to maximize a Whittle approximation for the log-likelihood function. The chosen q, as shown in Robinson (1991b), asymptotically minimizes a weighted integrated MSE criterion for the spectral density estimator, with the weight depending on the true spectral density. Because cross-validation methods use information on an interval, typically $[-\pi, \pi]$, they are suitable for spectral density estimation over an interval. In contrast, the narrow-band method chooses q to minimize an appropriate MSE criterion for the spectral density estimator at a single frequency, typically frequency zero. Therefore, it is suitable for estimation of the spectral density at a single frequency. The method of Andrews (1991) is of the plug-in type and uses parametric estimates. This method is simple to use in practice. It may not deliver the optimal q in terms of the MSE criterion, but the optimal rate for q is still obtained. Newey and West's (1994) method is also of the plug-in type but uses nonparametric estimates. This method delivers the optimal q in terms of the MSE criterion. Because our test statistic is based on estimation of the spectral density of squared residuals at frequency zero, it seems more appropriate to use the narrow-band method here. We compare the narrow-band method and the cross-validation in simulation.

5. SIMULATION

We now investigate the finite sample performances of our tests in comparison with Lee and King's (1993) LBS test and Engle's (1982) LM test. Consider the data-generating process

$$Y_t = X'_t b_0 + \varepsilon_t \qquad \varepsilon_t = \xi_t h_t^{1/2} \qquad t = 1, \ldots, T$$

where $X_t = (1, m_t)'$, $m_t = \lambda m_{t-1} + v_t$ and $v_t \sim \mathrm{NID}(0, \sigma_v^2)$. We consider three processes for h_t: (a) $h_t = \omega$; (b) $h_t = \omega + \alpha\varepsilon_{t-1}^2$; and (c) $h_t = \omega + \alpha\varepsilon_{t-1}^2 + \beta h_{t-1}$.

There is no ARCH under (a). Alternative (b) is an ARCH(1) process, often examined in existing simulation studies (e.g. Engle *et al.*, 1985; Lee and King, 1993). Alternative (c) is a GARCH(1, 1) process. When $\alpha + \beta < 1$, the GARCH(1, 1) can be expressed as an ARCH(∞) with coefficients declining at an exponential rate. The GARCH(1, 1) model has been the workhorse in the literature since it was introduced by Bollerslev (1986). Empirical studies find that this model is adequate in modeling volatilities of most economic and financial time series.

We set $b_0 = (1, 1)'$ and $\omega = 1$. For the exogenous variable m_t, we set $\lambda = 0.8$ and $\sigma_v^2 = 4$. As in Engle *et al.* (1985), m_t is generated for each experiment and then held fixed from iteration to iteration. For ARCH(1), we use $\alpha = 0.3, 0.95$, and set the initial condition $\varepsilon_0^2 = 0$; for GARCH(1, 1), we use $(\alpha, \beta) = (0.3, 0.2)$, $(0.3, 0.65)$, and set the initial conditions $\varepsilon_0 = 0$ and $h_0 = 1$. Sample sizes of $T = 64$, 128, are considered. To reduce the possible effects of the initial conditions, we generate $T + 100$ observations and then discard the first 100. For each experiment, the replication number is 5000 for (a), and 1000 each for (b) and (c). The simulation is conducted using a GAUSS random number generator on IBM RISC System/6000.

To investigate the effect of the choice of kernel function k on size and power, we use five kernels: the Bartlett, Daniell, Parzen, QS and truncated kernels:

$$\text{Bartlett} \qquad k(z) = \begin{cases} 1 - |z| & \text{for } |z| \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

$$\text{Daniell} \qquad k(z) = \frac{\sin(\pi z)}{\pi z} \qquad z \in (-\infty, \infty)$$

$$\text{Parzen} \qquad k(z) = \begin{cases} 1 - 6|\pi z/6|^2 & \text{for } |z| \leq 3/\pi \\ 2(1 - |\pi z/6|)^3 & \text{for } 3/\pi \leq |z| \leq 6/\pi \\ 0 & \text{otherwise;} \end{cases}$$

QS $$k(z) = \left\{\frac{9}{5(\pi z)^2}\right\}\left[\frac{\sin\{(5/3)^{1/2}\pi z\}}{(5/3)^{1/2}\pi z} - \cos\{(5/3)^{1/2}\pi z\}\right] \quad z \in (-\infty, \infty)$$

Truncated $$k(z) = \begin{cases} 1 & \text{for } |z| \leqslant 1 \\ 0 & \text{otherwise.} \end{cases}$$

The first four kernels are non-uniform and the truncated kernel is uniform. The Daniell, Parzen and QS kernels belong to the class $\mathscr{K}(\pi/3^{1/2})$ in (7). To investigate the effect of the choice of q, we use five q for each sample size. The first three are deterministic: $q = 1, 4, 8$. The last two are data-driven methods: Beltrão and Bloomfield's (1987) cross-validation log-likelihood method and Andrews' (1991) narrow-band plug-in method. These data-driven methods are used for their simplicity. For the cross-validation, we use a fast Fourier transform algorithm and choose an integer q by a grid search over the range $q = 1$ to $q = 20$, with the grid interval equal to 1; for Andrews' plug-in method, we use an ARCH(1) approximating process and round the resulting real-valued q to the nearest integer. We also restrict q to the range 1–20. To ensure that the first q sample autocorrelations are used in the test statistic, we use the truncation lag number $q + 1$ for non-uniform kernels, because the weight for the lag $j = q + 1$ is zero.

In addition to the S test, we also consider Lee and King's (1993) LBS test that is robust to non-normality, as well as Engle's (1982) LM test TR^2. For these two tests against alternatives (b) and (c), the theoretical optimal lag number is $q = 1$ (cf. Lee, 1991; Lee and King, 1993), which requires knowledge of the alternatives. For comparison we also use $q = 4, 8$.

Table I reports size performances under (a). For S tests, non-uniform weights generally yield better sizes than uniform weighting. Both $S_{\mathrm{TRUN}}(q)$ and LBS(q) have similar sizes in all cases. All the tests have best sizes for $q = 1$ and poorer sizes for larger q. Both data-driven methods give reasonable sizes for the S test.

Table II reports power against ARCH(1) using the 10% and 5% empirical critical values obtained from the replications under (a). The use of empirical critical values provides an equal basis to compare powers. We first consider $q = 1$. The tests $S_{\mathrm{BAR}}(1)$, $S_{\mathrm{PAR}}(1)$, $S_{\mathrm{QS}}(1)$, $S_{\mathrm{TRUN}}(1)$ and LBS(1) have the same power and are more powerful than LM(1), suggesting the gain of exploiting the one-sided nature of the alternative. For the S test, the Daniell kernel has slightly lower power than the Bartlett, Parzen and QS kernels for $T = 64$. This may be due to the fact that the Daniell kernel gives negative, although small, weights for some lags j. Next, we consider $q > 1$. As expected, all the tests have less power than when $q = 1$. For our S test, non-uniform kernels have substantially better power than the truncated kernel, suggesting the gain of using non-uniform weights. The QS kernel has slightly better power than the Daniell and Parzen kernels in many cases. The S tests have better power than the LBS test. The efficiency gain of the S test over the LBS test is substantial

TABLE I

SIZE AT THE 10% AND 5% SIGNIFICANCE LEVELS

		$q = 1$		$q = 4$		$q = 8$		CV		PI	
		10%	5%	10%	5%	10%	5%	10%	5%	10%	5%
$T = 64$											
S	BAR	8.1	4.2	6.0	3.6	4.9	3.0	7.3	3.9	7.3	3.1
	DAN	8.5	4.4	7.0	4.2	6.0	3.6	8.7	5.2	7.5	3.7
	PAR	7.4	4.3	5.8	3.4	4.3	2.7	7.5	4.1	5.7	3.8
	QS	8.1	4.4	6.3	3.8	5.1	3.2	8.2	4.9	6.5	3.1
	TRUN	8.0	4.2	5.2	3.2	4.1	2.3				
LBS		8.2	4.4	5.3	2.7	5.3	2.5				
LM		7.3	3.5	6.8	3.4	5.1	2.4				
$T = 128$											
S	BAR	8.7	5.2	7.4	4.5	6.9	4.3	10.4	6.0	7.9	3.5
	DAN	9.0	4.8	8.0	4.6	8.0	4.7	11.0	6.5	8.5	4.4
	PAR	8.2	5.0	7.4	4.4	6.5	3.7	10.2	6.2	6.9	3.8
	QS	8.8	5.0	7.9	4.5	7.1	4.4	10.8	6.7	7.5	3.7
	TRUN	8.6	5.1	7.2	4.3	6.0	3.4				
LBS		8.8	5.1	6.5	3.2	5.6	2.8				
LM		8.6	4.1	8.3	4.2	7.8	3.9				

Notes: Model: $Y_t = 1 + m_t + \varepsilon_t$, $m_t = 0.8m_{t-1} + v_t$, $v_t \sim \mathrm{NID}(0, 4)$, $\varepsilon_t = \xi_t h_t^{0.5}$, $\xi_t \sim \mathrm{NID}(0, 1)$, $h_t = 1$.
5000 replications.
CV, cross-validation; PI, narrow-band plug-in method.
BAR, Bartlett kernel; DAN, Daniell kernel; PAR, Parzen kernel; QS, quadratic-spectral kernel; TRUN, truncated kernel.

when q is large, confirming our asymptotic analysis. Interestingly, although $S_{\mathrm{TRUN}}(q)$ is asymptotically equivalent to LBS(q), it has better power than LBS(q). This is perhaps due to use of a different asymptotic variance estimator. Finally, the data-driven methods give reasonable power for the S test, although slightly lower than when $q = 1$ is used. The narrow-band method gives better power than the cross-validation.

Table III reports power against GARCH(1, 1). First, we consider $q = 1$. For $(\alpha, \beta) = (0.3, 0.2)$, $S_{\mathrm{BAR}}(1)$, $S_{\mathrm{PAR}}(1)$, $S_{\mathrm{QS}}(1)$, $S_{\mathrm{TRUN}}(1)$ and LBS(1) have similar powers and are more powerful than LM(1). Again, for the S test, the Daniell kernel gives slightly lower power than the Bartlett, Parzen and QS kernels. Next, we examine $q > 1$. For $(\alpha, \beta) = (0.3, 0.2)$, all the tests become less powerful as q increases. For $(\alpha, \beta) = (0.3, 0.65)$, however, better powers are obtained for larger q. This clearly suggests that it is better to include more lags in detecting highly persistent alternatives. Indeed, when $q = 1$ is used, the powers of all tests are hardly affected by the change of β. For $q > 1$, the S tests with non-uniform kernels have better power than the truncated-kernel-based test and the LBS test. This is true even for $(\alpha, \beta) = (0.3, 0.65)$, the highly persistent alternative. The QS kernel has slightly better power than the Daniell kernel. Interestingly, for $(\alpha, \beta) = (0.3, 0.65)$, the Parzen kernel has slightly better power than the QS kernel for $q = 4$. Again, our truncated-kernel-

TABLE II

NUMBER OF REJECTIONS AGAINST ARCH(1) AT THE 10% AND 5% EMPIRICAL CRITICAL VALUES

		$T = 64$										$T = 128$									
		$q = 1$		$q = 4$		$q = 8$		CV		PI		$q = 1$		$q = 4$		$q = 8$		CV		PI	
		10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%
$\alpha = 0.3$																					
S	BAR	573	438	494	371	412	294	493	357	568	412	785	702	678	577	587	440	677	567	778	666
	DAN	528	404	481	345	369	261	479	343	522	415	769	664	669	550	524	400	660	529	730	611
	PAR	556	423	461	332	362	251	498	362	533	394	777	675	656	532	508	386	710	596	733	621
	QS	557	434	474	358	382	270	492	357	560	442	784	701	676	568	550	411	710	582	770	649
	TRUN	574	438	394	269	289	191					787	699	554	436	396	291				
LBS		563	441	351	224	216	139					797	695	516	405	349	241				
LM		454	364	344	260	285	199					725	632	585	490	478	383				
$\alpha = 0.95$																					
S	BAR	901	822	850	778	780	674	815	690	898	816	988	968	973	945	942	895	955	909	986	971
	DAN	866	778	834	745	728	622	847	731	859	778	974	956	965	944	921	855	954	910	969	952
	PAR	906	829	827	735	725	599	847	735	900	811	989	969	961	934	836	750	956	919	986	962
	QS	898	815	842	754	749	629	845	740	895	828	989	969	970	942	928	875	960	926	986	970
	TRUN	901	822	759	647	595	469					988	968	931	890	814	729				
LBS		896	825	681	553	432	314					989	970	911	854	728	603				
LM		839	778	705	636	605	537					972	952	920	889	857	811				

Notes: Model: $Y_t = 1 + m_t + \varepsilon_t$, $m_t = 0.8m_{t-1} + v_t$, $v_t \sim \text{NID}(0, 4)$, $\varepsilon_t = \xi_t h_t^{0.5}$, $\xi_t \sim \text{NID}(0, 1)$, $h_t = 1 + \alpha\varepsilon_{t-1}^2$.
1000 replications.
CV, cross-validation; PI, narrow-band plug-in method.
BAR, Bartlett kernel; DAN, Daniell kernel; PAR, Parzen kernel; QS, quadratic-spectral kernel; TRUN, truncated kernel.

TABLE III

Number of Rejections Against GARCH(1, 1) at the 10% and 5% Empirical Critical Values

		T = 64										T = 128									
		$q = 1$		$q = 4$		$q = 8$		CV		PI		$q = 1$		$q = 4$		$q = 8$		CV		PI	
		10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%
$(\alpha, \beta) = (0.3, 0.2)$																					
S	BAR	576	456	573	445	499	386	554	413	577	441	801	714	772	676	683	576	746	661	798	705
	DAN	507	399	559	428	462	337	519	417	522	419	761	649	754	663	634	531	733	620	746	663
	PAR	593	469	539	422	461	330	561	426	587	463	816	718	741	645	630	515	769	657	803	701
	QS	571	458	553	436	478	350	553	413	578	469	803	721	763	669	655	552	767	663	801	708
	TRUN	576	455	501	356	362	250					804	713	680	569	511	399				
LBS		571	451	445	313	270	180					804	711	642	537	453	322				
LM		468	380	369	292	303	233					734	644	629	536	516	438				
$(\alpha, \beta) = (0.3, 0.65)$																					
S	BAR	558	455	742	635	763	681	686	610	560	481	844	739	953	896	961	923	927	894	852	766
	DAN	465	336	701	589	748	648	688	601	487	418	728	616	926	863	955	905	933	886	770	708
	PAR	640	536	760	666	759	674	688	606	651	560	905	838	961	924	956	916	933	889	912	862
	QS	572	450	738	638	763	670	691	602	584	493	849	761	954	891	958	925	929	884	855	788
	TRUN	559	454	738	637	741	615					846	737	946	905	934	890				
LBS		556	452	687	586	590	484					851	732	924	882	879	813				
LM		472	391	537	439	480	400					759	700	837	776	793	715				

Notes: Model: $Y_t = 1 + m_t + \varepsilon_t$, $m_t = 0.8m_{t-1} + v_t$, $v_t \sim$ NID(0, 4), $\varepsilon_t = \xi_t h_t^{0.5}$, $\xi_t \sim$ NID(0, 1), $h_t = 1 + \alpha\varepsilon_{t-1}^2 + \beta h_{t-1}$.
1000 replications.
CV, cross-validation; PI, narrow-band plug-in method.
BAR, Bartlett kernel; DAN, Daniell kernel; PAR, Parzen kernel; QS, quadratic-spectral kernel; TRUN, truncated kernel.

based test has better power than LBS(q), perhaps because of the use of a different asymptotic variance. Also, the efficiency gain of non-uniform weighting over uniform weighting and LBS(q) is substantial for large q. Finally, for data-driven methods, when $(\alpha, \beta) = (0.3, 0.2)$, the narrow-band method gives better power for the S test. For $(\alpha, \beta) = (0.3, 0.65)$, however, the cross-validation gives better power.

We summarize our findings.

(i) For $q = 1$, the new S test with use of the Bartlett, Parzen, QS and truncated kernels and the LBS test have the same power against ARCH(1) and GARCH(1, 1) respectively. They all have better power than the LM test.

(ii) For $q > 1$, the S tests with non-uniform weights have better power than the truncated-kernel-based test, which in turn is better than the LBS test. The efficiency gain from using non-uniform weighting is substantial for large q. The S tests with non-uniform weights have better power than the LM test. The LBS test has better power than the LM test for ARCH(1) and less persistent GARCH(1, 1), but not for persistent GARCH(1, 1).

(iii) For ARCH(1) and less persistent GARCH(1, 1), all the tests with $q = 1$ have best power. For persistent GARCH(1, 1), however, none has best power when $q = 1$, although $q = 1$ is the theoretical optimal order for both the LM and the LBS tests. Instead, all the tests have better power when more lags are used.

(iv) The data-driven methods (the cross-validation and narrow-band methods) yield reasonable size and power for the S test. To some extent, these methods reveal information on the true alternative.

6. CONCLUSION

Most existing ARCH tests fail to exploit the one-sided nature of the alternative hypothesis. Lee and King (1993) recently proposed a one-sided locally most mean powerful score-based test for ARCH effects, by using SenGupta and Vermeire's (1986) approach for testing multiparameter hypotheses. Using a frequency domain approach, in this paper a new one-sided N(0, 1) test for ARCH effects of the disturbance of a dynamic regression model is proposed. The test is based on a weighted sum of sample autocorrelations of squared regression residuals, with the weighting function typically giving more weight to lower orders of lags and less weight to higher orders of lags. Lee and King's (1993) test can be viewed as a special case of the present approach with the use of uniform weighting. Many non-uniform weighting schemes deliver better power than uniform weighting; the efficiency gain is substantial when a relatively large q is used. A simulation experiment confirms the gains from exploiting the one-sided nature of the alternative hypothesis and from using non-uniform weighting. Some data-driven methods for choosing the number of lags deliver reasonable size and power in finite samples.

MATHEMATICAL APPENDIX

PROOF OF THEOREM 1. Throughout this appendix, '$\rightarrow^{p}$' and '$\rightarrow^{d}$' denote convergence in probability and in distribution respectively. For $j \geqslant 0$, put

$$\bar{R}(j) = T^{-1} \sum_{t=j+1}^{T} (\xi_t^2 - 1)(\xi_{t-j}^2 - 1)$$

and $R(j) = E\bar{R}(j)$. Then

$$\sum_{j=1}^{T-1} k\left(\frac{j}{q}\right)\hat{R}(j) = \sum_{j=1}^{T-1} k\left(\frac{j}{q}\right)\bar{R}(j) + \sum_{j=1}^{T-1} k\left(\frac{j}{q}\right)\{\hat{R}(j) - \bar{R}(j)\}.$$

By Lemmas A1 and A2 below, we obtain $V_T^{-1/2} T^{1/2} \sum_{j=1}^{T-1} k(j/q)\hat{R}(j)/R(0) \rightarrow^{d} \mathrm{N}(0, 1)$. The normality for S follows by Slutsky's theorem and $\hat{R}(0) - R(0) \rightarrow^{p} 0$.

LEMMA A1. $V_T^{-1/2} T^{1/2} \sum_{j=1}^{T-1} k(j/q)\{\hat{R}(j) - \bar{R}(j)\} = \mathrm{o}_P(1)$.

LEMMA A2. $V_T^{-1/2} T^{1/2} \sum_{j=1}^{T-1} k(j/q)\bar{R}(j)/R(0) \rightarrow^{d} \mathrm{N}(0, 1)$.

PROOF OF LEMMA A1. We consider $q \rightarrow \infty$ as $T \rightarrow \infty$; the proof for fixed q is similar. Put $\hat{\xi}_t = \hat{\varepsilon}_t/\hat{\sigma}$. Then we have

$$\begin{aligned}
\hat{R}(j) - \tilde{R}(j) &= T^{-1} \sum_{t=j+1}^{T} \{(\hat{\xi}_t^2 - 1)(\hat{\xi}_{t-j}^2 - 1) - (\xi_t^2 - 1)(\xi_{t-j}^2 - 1)\} \\
&= T^{-1} \sum_{t=j+1}^{T} (\xi_t^2 - 1)(\hat{\xi}_{t-j}^2 - \xi_{t-j}^2) + T^{-1} \sum_{t=j+1}^{T} (\hat{\xi}_t^2 - \xi_t^2)(\xi_{t-j}^2 - 1) \\
&\quad + T^{-1} \sum_{t=j+1}^{T} (\hat{\xi}_t^2 - \xi_t^2)(\hat{\xi}_{t-j}^2 - \xi_{t-j}^2) \\
&= \hat{A}_1(j) + \hat{A}_2(j) + \hat{A}_3(j), \text{ say.}
\end{aligned} \tag{A1}$$

Putting $u_t = \xi_t^2 - 1$ and noting $\xi_t = \varepsilon_t/\sigma_0$ under H_0, where $\sigma_0^2 = E(\varepsilon_t^2)$, we have

$$\begin{aligned}
\hat{A}_1(j) &= T^{-1} \sum_{t=j+1}^{T} (\xi_t^2 - 1)(\hat{\xi}_{t-j}^2 - \xi_{t-j}^2) \\
&= \hat{\sigma}^{-2} T^{-1} \sum_{t=j+1}^{T} u_t(\hat{\varepsilon}_{t-j}^2 - \varepsilon_{t-j}^2) + (\hat{\sigma}^{-2} - \sigma_0^{-2}) T^{-1} \sum_{t=j+1}^{T} u_t \varepsilon_{t-j}^2 \\
&= \hat{\sigma}^{-2} T^{-1} \sum_{t=j+1}^{T} u_t(\hat{\varepsilon}_{t-j} - \varepsilon_{t-j})^2 + 2\hat{\sigma}^{-2} T^{-1} \sum_{t=j+1}^{T} u_t \varepsilon_{t-j}(\hat{\varepsilon}_{t-j} - \varepsilon_{t-j}) \\
&\quad + (\hat{\sigma}^{-2} - \sigma_0^{-2}) T^{-1} \sum_{t=j+1}^{T} u_t \varepsilon_{t-j}^2 \\
&= \hat{\sigma}^{-2} \hat{B}_{11}(j) + 2\hat{\sigma}^{-2} \hat{B}_{12}(j) + (\hat{\sigma}^{-2} - \sigma_0^{-2}) \hat{B}_{13}(j), \text{ say.}
\end{aligned} \tag{A2}$$

Noting that $\hat{\varepsilon}_t - \varepsilon_t = (b_0 - \hat{b})'\nabla_b g(X_t, \bar{b})$, where $\|\bar{b} - b_0\| \leqslant \|\hat{b} - b_0\|$, we have

$$\begin{aligned}
\left|\sum_{j=1}^{T-1} k\left(\frac{j}{q}\right)\hat{B}_{11}(j)\right| &\leqslant \sum_{j=1}^{T-1}\left|k\left(\frac{j}{q}\right)\left\{T^{-1}\sum_{t=j+1}^{T} u_t(\hat{\varepsilon}_{t-j}-\varepsilon_{t-j})^2\right\}\right| \\
&\leqslant \|b_0-\hat{b}\|^2\left\{\sum_{j=1}^{T-1}\left|k\left(\frac{j}{q}\right)\right|\right\}\left(T^{-1}\sum_{t=1}^{T}u_t^2\right)^{1/2} \\
&\quad\times\left\{T^{-1}\sum_{t=1}^{T}\|\nabla_b g(X_t,\bar{b})\|^4\right\}^{1/2} \\
&= \mathrm{O}_P\left(\frac{q}{T}\right) \qquad \text{(A3)}
\end{aligned}$$

by the Cauchy–Schwarz inequality and Assumptions A1–A4, where

$$q^{-1}\sum_{j=1}^{T-1}|k(j/q)| \to \int_0^\infty |k(z)|\,dz.$$

Next, using $\hat{\varepsilon}_t-\varepsilon_t=(b_0-\hat{b})'\nabla_b g(X_t,b_0)+\frac{1}{2}(b_0-\hat{b})'\nabla_b^2 g(X_t,\bar{b})(b_0-\hat{b})$, we have

$$\begin{aligned}
\left|\sum_{j=1}^{T-1} k\left(\frac{j}{q}\right)\hat{B}_{12}(j)\right| &\leqslant \sum_{j=1}^{T-1}\left|k\left(\frac{j}{q}\right)\left\{T^{-1}\sum_{t=j+1}^{T} u_t\varepsilon_{t-j}(\hat{\varepsilon}_{t-j}-\varepsilon_{t-j})\right\}\right| \\
&\leqslant \|b_0-\hat{b}\|\left\{\sum_{j=1}^{T-1}\left|k\left(\frac{j}{q}\right)\right|\left\|T^{-1}\sum_{t=j+1}^{T}u_t\varepsilon_{t-j}\nabla_b g(X_{t-j},b_0)\right\|\right\} \\
&\quad+\frac{1}{2}\|b_0-\hat{b}\|^2\left\{\sum_{j=1}^{T-1}\left|k\left(\frac{j}{q}\right)\right|\left\|T^{-1}\sum_{t=j+1}^{T}u_t\varepsilon_{t-j}\nabla_b^2 g(X_{t-j},\bar{b})\right\|\right\} \\
&= \mathrm{O}_P\left(\frac{q}{T}\right) \qquad \text{(A4)}
\end{aligned}$$

by Markov's inequality and $E\|T^{-1}\sum_{t=j+1}^{T}u_t\varepsilon_{t-j}\nabla_b g(X_{t-j},b_0)\|^2=\mathrm{O}(T^{-1})$ given $E(u_t|I_{t-1})=0$. We also have

$$\sum_{j=1}^{T-1} k\left(\frac{j}{q}\right)\hat{B}_{13}(j)=\mathrm{O}_P\left(\frac{q}{T^{1/2}}\right) \qquad \text{(A5)}$$

by Markov's inequality and $E(T^{-1}\sum_{t=j+1}^{T}u_t\varepsilon_{t-j}^2)^2=\mathrm{O}(T^{-1})$ given Assumption A1.

Collecting (A2)–(A5) and noting $\hat{\sigma}^2-\sigma_0^2=\mathrm{O}_P(T^{-1/2})$, we obtain

$$\sum_{j=1}^{T-1} k\left(\frac{j}{q}\right)\hat{A}_1(j)=\mathrm{O}_P\left(\frac{q}{T}\right). \qquad \text{(A6)}$$

Similarly, we have

$$\sum_{j=1}^{T-1} k\left(\frac{j}{q}\right)\hat{A}_2(j)=\mathrm{O}_P\left(\frac{q}{T}\right). \qquad \text{(A7)}$$

Next, we turn to $\hat{A}_3(j)$. By the Cauchy–Schwarz inequality, we have

$$\sup_{1\leqslant j\leqslant T-1}|\hat{A}_3(j)| \leqslant T^{-1}\sum_{t=1}^{T}(\hat{\xi}_t^2-\xi_t^2)^2$$

$$\leqslant 2\hat{\sigma}^{-4}T^{-1}\sum_{t=1}^{T}(\hat{\varepsilon}_t^2-\varepsilon_t^2)^2+2(\hat{\sigma}^{-2}-\sigma_0^{-2})^2T^{-1}\sum_{t=1}^{T}\varepsilon_t^4.$$

Recalling that $\hat{\varepsilon}_t^2-\varepsilon_t^2=(\hat{\varepsilon}_t-\varepsilon_t)^2+2\varepsilon_t(\hat{\varepsilon}_t-\varepsilon_t)$, we have

$$\begin{aligned}T^{-1}\sum_{t=1}^{T}(\hat{\varepsilon}_t^2-\varepsilon_t^2)^2 &\leqslant 2T^{-1}\sum_{t=1}^{T}(\hat{\varepsilon}_t-\varepsilon_t)^4+8T^{-1}\sum_{t=1}^{T}\varepsilon_t^2(\hat{\varepsilon}_t-\varepsilon_t)^2\\ &\leqslant 2\|b_0-\hat{b}\|^4T^{-1}\sum_{t=1}^{T}\|\nabla_b g(X_t,\bar{b})\|^4\\ &\quad+8\|b_0-\hat{b}\|^2T^{-1}\sum_{t=1}^{T}\varepsilon_t^2\|\nabla_b g(X_t,\bar{b})\|^2\\ &=\mathrm{O}_P(T^{-1}).\end{aligned}$$

This, together with $\hat{\sigma}^2-\sigma_0^2=\mathrm{O}_P(T^{-1/2})$, yields

$$\left|\sum_{j=1}^{T-1}k\left(\frac{j}{q}\right)\hat{A}_3(j)\right| \leqslant \sum_{j=1}^{T-1}\left|k\left(\frac{j}{q}\right)\right|\left\{T^{-1}\sum_{t=1}^{T}(\hat{\xi}_t^2-\xi_t^2)^2\right\}=\mathrm{O}_P\left(\frac{q}{T}\right). \tag{A8}$$

Hence, from (A1), (A6)–(A8) and $V_T=\mathrm{O}(q)$, we have

$$V_T^{-1/2}T^{1/2}\sum_{j=1}^{T-1}k(j/q)\{\hat{R}(j)-\bar{R}(j)\}=\mathrm{O}_P(q^{1/2}/T^{1/2})=\mathrm{o}_P(1)$$

given $q/T\to 0$. The proof for fixed q is similar and is omitted here. This completes the proof.

Proof of Lemma A2. Put $\bar{W}_T=T^{1/2}\sum_{j=1}^{T-1}k(j/q)\bar{R}(j)/R(0)=T^{-1/2}\sum_{j=1}^{T-1}W_t$, where

$$W_t=R^{-1}(0)u_t\left\{\sum_{j=1}^{t-1}k\left(\frac{j}{q}\right)u_{t-j}\right\}. \tag{A9}$$

Because $\{W_t, F_t\}$ is a martingale difference sequence, where F_t is a sigma field consisting of u_s, $s\leqslant t$, we apply Brown's (1971) martingale limit theorem, which implies that $\{\mathrm{var}(\bar{W}_T)\}^{-1/2}\bar{W}_T\to^{\mathrm{d}}\mathrm{N}(0,1)$ if

$$\{\mathrm{var}(\bar{W}_T)\}^{-1}T^{-1}\sum_{t=2}^{T}E[W_t^2 1\{|W_t|>\delta T^{1/2}\,\mathrm{var}^{1/2}(\bar{W}_T)\}]\to 0 \tag{A10}$$

for every $\delta>0$, and

$$\{\mathrm{var}(\bar{W}_T)\}^{-1}T^{-1}\sum_{t=3}^{T}E(W_t^2|F_{t-1})\to^{\mathrm{p}}0. \tag{A11}$$

Given Assumption A1, $\mathrm{var}(\bar{W}_T)=\sum_{t=2}^{T}\sum_{j=1}^{t-1}k^2(j/q)=V_T$.

We first consider $q \to \infty$ as $T \to \infty$. Because $q^{-1} V_T \to \int_0^\infty k^2(z)\,dz$, we have $\mathrm{var}(\bar{W}_T) = \mathrm{O}(q)$. Thus, we can verify (A10) by showing that $q^{-2} T^{-2} \sum_{t=2}^{T} E(W_t^4) = \mathrm{o}(1)$. Put $\mu_4 = E(u_t^4)$. By Assumption A1 we have

$$
\begin{aligned}
E(W_t^4) &= \left\{\frac{\mu_4}{R^4(0)}\right\} E\left\{\sum_{j=1}^{t-1} k\left(\frac{j}{q}\right) u_{t-j}\right\}^4 \\
&= \left\{\frac{\mu_4^2}{R^4(0)}\right\} \sum_{j=1}^{t-1} k^4\left(\frac{j}{q}\right) + 6\left\{\frac{\mu_4}{R^2(0)}\right\} \sum_{j=2}^{t-1}\sum_{i=1}^{j-1} k^2\left(\frac{j}{q}\right) k^2\left(\frac{i}{q}\right) \\
&\leqslant 3\left\{\frac{\mu_4^2}{R^4(0)}\right\}\left\{\sum_{j=1}^{t-1} k^2\left(\frac{j}{q}\right)\right\}^2 .
\end{aligned}
$$

It follows that $q^{-2} T^{-2} \sum_{t=2}^{T} E(W_t^4) \leqslant 3\{\mu_4^2/R^4(0)\} T^{-1} \{q^{-1} \sum_{j=1}^{T-1} k^2(j/q)\}^2 = \mathrm{O}(T^{-1})$. Thus, (A10) holds.

Next, we verify (A11) by showing that $q^{-2}\,\mathrm{var}\{T^{-1} \sum_{t=2}^{T} E(W_t^2 | F_{t-1})\} \to 0$. By (A9),

$$
\begin{aligned}
E(W_t^2 | F_{t-1}) &= R^{-1}(0) \left\{\sum_{j=1}^{t-1} k\left(\frac{j}{q}\right) u_{t-j}\right\}^2 \\
&= E(W_t^2) + R^{-1}(0) \sum_{j=1}^{t-1} k^2\left(\frac{j}{q}\right)\{u_{t-j}^2 - R(0)\} \\
&\quad + 2R^{-1}(0) \sum_{j=2}^{t-1}\sum_{i=1}^{j-1} k\left(\frac{j}{q}\right) k\left(\frac{i}{q}\right) u_{t-i} u_{t-j} \\
&= E(W_t^2) + R^{-1}(0) A_t + 2R^{-1}(0) B_t, \text{ say.}
\end{aligned}
$$

It follows that

$$
T^{-1} \sum_{t=2}^{T} \{E(W_t^2 | F_{t-1}) - E(W_t^2)\} = R^{-1}(0) T^{-1} \sum_{t=2}^{T} A_t + 2R^{-1}(0) T^{-1} \sum_{t=2}^{T} B_t .
$$

Thus, it suffices to show that $q^{-2}\,\mathrm{var}(T^{-1} \sum_{t=2}^{T} A_t) \to 0$ and $q^{-2}\,\mathrm{var}\,(T^{-1} \sum_{t=2}^{T} B_t) \to 0$. Noting that A_t is a weighted sum of independent variables $u_{t-j}^2 - R(0)$, we have $E(A_t^2) = \{\mu_4 - R^2(0)\} \sum_{j=1}^{t-1} k^4(j/q)$. It follows by Minkowski's inequality that

$$
\begin{aligned}
q^{-2} E\left(T^{-1} \sum_{t=2}^{T} A_t\right)^2 &\leqslant q^{-2}\left\{T^{-1} \sum_{t=2}^{T} (EA_t^2)^{1/2}\right\}^2 \\
&\leqslant q^{-1}\{\mu_4 - R^2(0)\}\left\{q^{-1} \sum_{j=1}^{t-1} k^4\left(\frac{j}{q}\right)\right\} \\
&= \mathrm{O}(q^{-1}).
\end{aligned}
$$

On the other hand, for $t \geqslant s$, we have

$$E(B_t B_s) = R^2(0) \sum_{j_1=2}^{t-1} \sum_{i_1=1}^{j_1-1} \sum_{j_2=2}^{s-1} \sum_{i_2=1}^{j_2-1} k\left(\frac{j_1}{q}\right) k\left(\frac{i_1}{q}\right) k\left(\frac{j_2}{q}\right) k\left(\frac{i_2}{q}\right) \delta_{t-i_1, s-i_2} \delta_{t-j_1, s-j_2}$$

$$= R^2(0) \sum_{j_2=2}^{s-1} \sum_{i_2=1}^{j_2-1} k\left(\frac{t-s+j_2}{q}\right) k\left(\frac{t-s+i_2}{q}\right) k\left(\frac{j_2}{q}\right) k\left(\frac{i_2}{q}\right)$$

where $\delta_{ij} = 1$ for $i = j$ and $\delta_{ij} = 0$ otherwise. It follows that

$$q^{-2} E\left(T^{-1} \sum_{t=2}^{T} B_t\right)^2 = q^{-2} T^{-2} \sum_{s=2}^{T} E(B_t^2) + 2q^{-2} T^{-2} \sum_{t=3}^{T} \sum_{s=2}^{t-2} E(B_t B_s)$$

$$= \mathrm{O}(T^{-1}).$$

Therefore, (A11) holds. By Brown's theorem, we have $V_T^{-1/2} \bar{W}_T \to^{\mathrm{d}} \mathrm{N}(0, 1)$.

Next, we consider fixed q. We first decompose

$$\sum_{j=1}^{T-1} k\left(\frac{j}{q}\right) \bar{R}(j) = \sum_{j=1}^{l} k\left(\frac{j}{q}\right) \bar{R}(j) + \sum_{j=l+1}^{T-1} k\left(\frac{j}{q}\right) \bar{R}(j)$$

for some $l \to \infty$, $l/T \to 0$. Because $TE|\sum_{j=l+1}^{T-1} k(j/q)\bar{R}(j)|^2 \leqslant R^2(0) \sum_{j=l}^{\infty} k^2(j/q) \to 0$ as $l \to \infty$, we have $T^{1/2} \sum_{j=l+1}^{T-1} k(j/q)\bar{R}(j) = \mathrm{o}_P(1)$. By Brown's theorem, we can show

$$\left\{\sum_{j=1}^{l} \left(1 - \frac{j}{T}\right) k^2\left(\frac{j}{q}\right)\right\}^{-1/2} T^{1/2} \sum_{j=1}^{l} k\left(\frac{j}{q}\right) \frac{\bar{R}(j)}{R(0)} \to^{\mathrm{d}} \mathrm{N}(0, 1)$$

where $\sum_{j=1}^{l} (1 - j/T) k^2(j/q)$ is bounded from below and from above. Also, $V_T - \sum_{j=1}^{l} (1 - j/T) k^2(j/q) \to^{\mathrm{p}} 0$. It follows that $V_T^{-1/2} \bar{W}_T \to^{\mathrm{d}} \mathrm{N}(0, 1)$. This completes the proof.

Proof of Theorem 2. We consider only $q \to \infty$ as $T \to \infty$; the proof for fixed q is similar. Put $\tilde{R}(j) = T^{-1} \sum_{t=j+1}^{T} (\varepsilon_t^2/\sigma_0^2 - 1)(\varepsilon_{t-j}^2/\sigma_0^2 - 1)$, where $\sigma_0^2 = E(\varepsilon_t^2)$. Then following the analogous but more tedious reasoning of Lemma A1, we obtain

$$V_T^{-1/2} T^{1/2} \sum_{j=1}^{T-1} k\left(\frac{j}{q}\right) \{\hat{R}(j) - \tilde{R}(j)\} = \mathrm{O}_P\left(\frac{q}{T^{1/2}}\right)$$

$$= \mathrm{o}_P(1)$$

given that $q^2/T \to 0$. Therefore, it remains to show that

$$V_T^{-1/2} T^{1/2} \sum_{j=1}^{T-1} k\left(\frac{j}{q}\right) \frac{\tilde{R}(j)}{R(0)} \to^{\mathrm{d}} \mathrm{N}(\mu, 1)$$

where $\mu = \sum_{j=1}^{\infty} \beta_j / \{\int_0^{\infty} k^2(z)\, dz\}^{1/2}$. Putting $V_t = \sum_{i=1}^{\infty} \beta_i u_{t-i}$, where $u_t = \xi_t^2 - 1$, we have

$$\tilde{R}(j) = T^{-1} \sum_{t=j+1}^{T} (u_t + a_T V_t)(u_{t-j} + a_T V_{t-j})$$

$$= \bar{R}(j) + a_T T^{-1} \sum_{j=1}^{T} V_t u_{t-j} + a_T T^{-1} \sum_{t=j+1}^{T} u_t V_{t-j} + a_T^2 T^{-1} \sum_{t=j+1}^{T} V_t V_{t-j}$$

$$= \bar{R}(j) + a_T \hat{A}_4(j) + a_T \hat{A}_5(j) + a_T^2 \hat{A}_6, \text{ say.} \tag{A12}$$

where, as before, $\bar{R}(j) = n^{-1} \sum_{t=j+1}^{T} u_t u_{t-j}$. We first consider $\hat{A}_4(j)$. Write

$$\sum_{j=1}^{T-1} k\left(\frac{j}{q}\right) \hat{A}_4(j) = \sum_{j=1}^{T-1} k\left(\frac{j}{q}\right) T^{-1} \sum_{t=j+1}^{T} \left(\sum_{i=1}^{\infty} \beta_i u_{t-i}\right) u_{t-j}$$

$$= R(0) \sum_{j=1}^{T-1} \left(1 - \frac{j}{T}\right) \beta_j k\left(\frac{j}{q}\right)$$

$$+ \sum_{j=1}^{T-1} \beta_j k\left(\frac{j}{q}\right) T^{-1} \sum_{t=j+1}^{T} \{u_{t-j}^2 - R(0)\}$$

$$+ \sum_{j=1}^{T-1} \beta_j k\left(\frac{j}{q}\right) T^{-1} \sum_{t=j+1}^{T} V_t(j) u_{t-j}$$

$$= R(0) \sum_{t=j+1}^{T-1} \left(1 - \frac{j}{T}\right) \beta_j k\left(\frac{j}{q}\right) + \hat{B}_{41} + \hat{B}_{42}, \text{ say,}$$

where $V_t(j) = \sum_{i=1, i \neq j}^{\infty} \beta_i u_{t-i}$. Because

$$E|\hat{B}_{41}| \leq \sum_{j=1}^{T-1} \beta_j \left| k\left(\frac{j}{q}\right) \right| \left(E\left[T^{-1} \sum_{t=j+1}^{T} \{u_{t-j}^2 - R(0)\} \right]^2 \right)^{1/2}$$

$$\leq \{\mu_4 - R^2(0)\}^{1/2} T^{-1/2} \sum_{j=1}^{T-1} \beta_j \left| k\left(\frac{j}{q}\right) \right|$$

$$= \mathrm{O}(T^{-1/2})$$

we have $\hat{B}_{41} = \mathrm{O}_P(T^{-1/2})$ by Markov's inequality. Similarly, $\hat{B}_{42} = \mathrm{O}_P(T^{-1/2})$ by noting that $V_t(j)$ and u_{t-j} are mutually independent with finite variances. It follows that

$$\sum_{j=1}^{T-1} k\left(\frac{j}{q}\right) \hat{A}_4(j) = R(0) \sum_{j=1}^{T-1} \left(1 - \frac{j}{T}\right) \beta_j k\left(\frac{j}{q}\right) + \mathrm{O}_P(T^{-1/2}). \tag{A13}$$

Next, we consider $\hat{A}_5(j)$ in (A12). Because u_t is independent of V_{t-j} for $j \geq 1$, we have $E\{A_5^2(j)\} \leq T^{-1} E(u_1^2) E(V_1^2) = R^2(0) T^{-1} \sum_{j=1}^{\infty} \beta_j^2$. It follows that

$$E\left| \sum_{j=1}^{T-1} k\left(\frac{j}{q}\right) \hat{A}_5(j) \right| \leq \sum_{j=1}^{T-1} \left| k\left(\frac{j}{q}\right) \right| [E\{\hat{A}_5^2(j)\}]^{1/2}$$

$$\leq R(0) \left(\frac{q}{T^{1/2}}\right) \left(\sum_{j=1}^{\infty} \beta_j^2\right)^{1/2} \left\{ q^{-1} \sum_{j=1}^{T-1} \left| k\left(\frac{j}{q}\right) \right| \right\}.$$

Therefore, by Markov's inequality, we have

$$\sum_{j=1}^{T-1} k\left(\frac{j}{q}\right)\hat{A}_5(j) = \mathrm{O}_P\left(\frac{q}{T^{1/2}}\right). \tag{A14}$$

Finally, we consider the last term $\hat{A}_6(j)$ in (A12). Because $V_t = \sum_{i=1}^{\infty} \beta_i u_{t-i}$ is a linear process with $\sum_{j=1}^{\infty} \beta_j < \infty$ and $E(u_t^4) < \infty$, the condition $\sum_{l=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} |\kappa(0, m, n, l)| < \infty$ is satisfied, where $\kappa(0, m, n, l)$ is the fourth-order cumulant of $V_t V_{t+m} V_{t+n} V_{t+l}$ (e.g. Hannan, 1970, p. 211). It follows (e.g. Hannan, 1970) that

$$\sum_{j=1}^{T-1} k\left(\frac{j}{q}\right)\hat{A}_6(j) \rightarrow^{\mathrm{p}} f_V(0) < \infty \tag{A15}$$

where $f_V(0)$ is the spectral density of V_t at zero frequency. Collecting (A12)–(A15) with $a_T = (q/T)^{1/2}$ and $q^2/T \rightarrow 0$ yields

$$\sum_{j=1}^{T-1} k\left(\frac{j}{q}\right)\frac{\tilde{R}(j)}{R(0)} = \sum_{j=1}^{T-1} k\left(\frac{j}{q}\right)\frac{\bar{R}(j)}{R(0)} + \left(\frac{q}{T}\right)^{1/2} \sum_{j=1}^{T-1}\left(1-\frac{j}{T}\right)\beta_j k\left(\frac{j}{q}\right) + \mathrm{o}_P\left\{\left(\frac{q}{T}\right)^{1/2}\right\}.$$

It follows that $S \rightarrow^{\mathrm{d}} \mathrm{N}(\mu, 1)$ by Lemma A2, $\sum_{j=1}^{T-1}(1-j/T)\beta_j k(j/q) \rightarrow \sum_{j=1}^{\infty} \beta_j$ and $q^{-1} V_T \rightarrow \int_0^{\infty} k^2(z)\,dz$ as $q \rightarrow \infty$. This completes the proof.

REFERENCES

Andrews, D. W. K. (1991) Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 59, 817–58.

Baillie, R. and Bollerslev, T. (1994) The long memory of the forward premium. Manuscript, Department of Finance, J. L. Kellogg Graduate School of Management, Northwestern University.

——, —— and Mikkelsen, H. (1993) Fractionally integrated generalized autoregressive conditional heteroskedasticity. Manuscript, Department of Finance, J. L. Kellogg Graduate School of Management, Northwestern University.

Beltrão, K. and Bloomfield, P. (1987) Determining the bandwidth of a kernel spectrum estimate. *J. Time Ser. Anal.* 8, 21–38.

Bera, A. K. and Higgins, M. L. (1992) A test for conditional heterskedasticity in time series models. *J. Time Ser. Anal.* 13, 501–19.

—— and —— (1993) ARCH models: properties, estimation and testing. *J. Econ. Surv.* 7, 305–66.

Bollerslev, T. (1986) A generalized autoregressive conditional heteroskedasticity. *J. Econometrics* 31, 307–27.

——, Chou, R. Y. and Kroner, K. F. (1992) ARCH modeling in finance. *J. Econometrics* 52, 5–59.

——, Engle, R. F. and Nelson, D. B. (1994) ARCH models. In *Handbook of Econometrics*, Vol. IV (eds R. F. Engle and D. L. McFadden). Amsterdam: Elsevier Science.

Box, G. E. P. and Pierce, D. A. (1970) Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *J. Am. Stat. Assoc.* 65, 1509–26.

Brown, B. M. (1971) Martingale central limit theorems. *Ann. Math. Stat.* 42, 59–66.

Diebold, F. X. (1987) Testing for serial correlation in the presence of ARCH. *Proc. Am. Stat. Assoc. Bus. Econ. Stat. Sect.* 323–28.

Drost, F. C. and Nijman, T. E. (1993) Temporal aggregation of GARCH processes. *Econometrica* 61, 909–27.

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

—— and Bollerslev, T. (1986) Modeling the persistence of conditional variances. *Econometric Rev.* 5, 1–50, 81–87.

—— and Susmel, R. (1993) Common volatility in international equity markets. *J. Bus. Econ. Stat.* 11, 167–76.

——, Hendry, D. F. and Trumble, D. (1985) Small-sample properties of ARCH estimators and tests.

Can. J. Econ. 18, 67–93.

GRANGER, C. W. J. (1969) Investigating causal relations by econometric models and cross spectral methods. *Econometrica* 37, 424–38.

—— and TERASVIRTA, T. (1993) *Modelling Nonlinear Economic Relationships*. New York: Oxford University Press.

GREGORY, A. W. (1989) A nonparametric test for autoregressive conditional heteroskedasticity: a Markov-chain approach. *J. Bus. Econ. Stat.* 7, 107–15.

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

HIGGINS, M. L. and BERA, A. K. (1992) A class of nonlinear ARCH models. *Int. Econ. Rev.* 33, 137–58.

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

—— and Shehadeh, R. D. (1997) A new test for ARCH effects and its finite sample performance. Manuscript. Department of Economics, Cornell University

HURVICH, M. C. (1985) Data-driven choice of a spectrum estimate: extending the applicability of cross-validation methods. *J. Am. Stat. Assoc.* 80, 933–40.

LEE, J. H. H. (1991) A Lagrange multiplier test for GARCH models. *Econ. Lett.* 37, 265–71.

—— and KING, M. L. (1993) A locally most mean powerful based score test for ARCH and GARCH regression disturbances. *J. Bus. Econ. Stat.* 11, 17–27.

LJUNG, G. M. and BOX, G. E. P. (1978) On a measure of lack of fit in time series models. *Biometrika* 65, 297–30.

MCLEOD, A. I. and LI, W. K. (1983) Diagnostic checking ARMA time series models using squared residual autocorrelations. *J. Time Ser. Anal.* 4, 269–73.

MILHOJ, A. (1985) The moment structure of ARCH processes. *Scand. J. Stat.* 12, 281–92.

NELSON, D. (1991) Conditional heteroskedasticity in asset returns: a new approach. *Econometrica* 59, 347–70.

—— and CAO, C. Q. (1992) Inequality constraints in the univariate GARCH model. *J. Bus. Econ. Stat.* 10, 229–35.

NEWEY, W. and WEST, K. (1994) Automatic lag selection in covariance matrix estimation. *Rev. Econ. Stud.* 61, 631–53.

PITMAN, E. (1979) *Some Basic Theory for Statistical Inference*. London: Chapman and Hall.

PRIESTLEY, M. B. (1962) Basic considerations in the estimation of spectra. *Technometrics* 4, 551–64.

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

—— (1991b) Automatic frequency domain inference on semiparametric and nonparametric models. *Econometrica* 59, 1329–64.

—— (1994) Time series with strong dependence. In *Advances in Econometrics*, Sixth World Congress, Vol. 1 (ed. C. Sims). Cambridge: Cambridge University Press, pp. 47–95.

SENGUPTA, A. and VERMEIRE, L. (1986) Locally optimal tests for multiparameter hypotheses. *J. Am. Stat. Assoc.* 81, 819–25.

SENTANA, E. (1995) Quadratic ARCH models. *Rev. Econ. Stud.* 62, 639–61.

TAYLOR, S. (1984) *Modelling Financial Time Series*. New York: Wiley.

WAHBA, G. and WOLD, S. (1975) Periodic splines for spectral density estimation: the use of cross-validation for determining the degree of smoothing. *Commun. Stat.* 4, 125–41.

WEISS, A. (1984) ARMA models with ARCH errors. *J. Time Ser. Anal.* 5, 129–43.

—— (1986) Asymptotic theory for ARCH models: estimation and testing. *Econometric Theory* 20, 107–31.