A New Test for ARCH Effects and Its Finite-Sample Performance

Yongmiao HONG and Ramsey D. SHEHADEH
Department of Economics, Cornell University, Ithaca, NY 14853-7601 (yh20@cornell.edu)
(ramsey.shehadeh@nera.com)

We propose a test for autoregressive conditional heteroscedasticity based on a weighted sum of the squared sample autocorrelations of squared residuals from a regression, typically with greater weight given to lower-order lags. The tests of Engle, Box and Pierce, and Ljung and Box are equivalent to the test with equal weighting. Our test does not require formulation of an alternative and permits choice of the lag number via data-driven methods. Simulation studies show that the new test performs reasonably well in finite samples especially with greater weight on lower-order lags. We apply the test in two empirical examples.

KEY WORDS: Cross-validation; Efficiency; Frequency domain; Monte Carlo; Spectral density; Weighting.

Since Engle (1982) introduced the autoregressive conditional heteroscedasticity (ARCH) model, there has been considerable interest in estimation of and testing for dynamic conditional heteroscedasticity of the regression disturbance. The ARCH model and its various generalizations [e.g., Bollerslev's (1986) generalized ARCH (GARCH)] have proved quite useful in modeling the disturbance behavior of regression models of economic and financial time series. See Bera and Higgins (1993), Bollerslev, Chou, and Kroner (1992), and Bollerslev, Engle, and Nelson (1994) for recent surveys of this literature.

From the perspective of econometric inference, neglecting ARCH effects may lead to arbitrarily large loss in asymptotic efficiency (Engle 1982) and cause overrejection of standard tests for serial correlation in conditional mean (Taylor 1984; Milhøj 1985; Diebold 1987; Domowitz and Hakkio 1987). In the context of autoregressive moving average (ARMA) modeling, Weiss (1984) pointed out that ignoring the ARCH effect will result in overparameterization of an ARMA model.

In practice, the most popular test for ARCH is Engle's (1982) Lagrange multiplier (LM) test for ARCH(q) under a two-sided alternative formulation. When the null hypothesis of no ARCH is true, this statistic is asymptotically distributed as a chi-squared random variable with q degrees of freedom. It is simple to calculate and is asymptotically locally most powerful if the true alternative is ARCH(q), a characteristic it shares with the likelihood ratio and Wald tests (Engle 1982) under a two-sided alternative formulation. [In the ARCH context, it is also possible to construct a one-sided test. With the one-sided alternative, the LM test has a standard chi-squared distribution. On the other hand, the likelihood ratio and Wald tests may have an asymptotic mixture of chi-squared distributions. See Bera, Ra, and Sarkar (1998).] An alternative approach is to subject the squared residuals to such standard tests for serial correlation as the portmanteau tests of Box and Pierce (1970) and Ljung and Box (1978), which are based on the sum of the first q-squared sample autocorrelations of the squared residuals. These tests, as shown by McLeod and Li (1983), are also asymptotically distributed as chi-squared random variables with q df. Granger and Teräsvirta (1993, pp. 93–94) showed that they are asymptotically equivalent to Engle's (1982) LM test.

Detection of ARCH effects has attracted significant recent attention from researchers. Contributions include those of Bera and Higgins (1992), Brock, Dechert, and Scheinkman (BDS, 1987), Gregory (1989), Lee (1991), Lee and King (1993), and Robinson (1991a). In this article, we propose a new test for the presence of ARCH in the residuals from a possibly nonlinear regression model. The test is based on an extension of Hong's (1996) spectral density approach and results for testing serial correlation of unknown form in conditional mean. Specifically, the null hypothesis of no ARCH implies that the normalized spectral density of the squared residuals from a regression model is the uniform density on the frequency interval $[-\pi, \pi]$. Alternatively, if ARCH exists, the normalized spectral density of the squared residuals is nonuniform in general. Therefore, we can construct a test for ARCH by comparing a kernel-based spectral density estimator of the squared residuals to the uniform density via an appropriate divergence measure. When a quadratic norm is used, the new test turns out to be a properly standardized version of the weighted sum of squares of all $n-1$ sample autocorrelations of the squared residuals, with weighting depending on the kernel function. Most commonly used kernels typically give greater weight to lower-order lags. The exception is the truncated kernel, which gives equal weight to each lag.

Interestingly, when using the truncated kernel (i.e., uniform weighting), our approach delivers a generalized Box and Pierce (1970) type of test, which also is asymptotically equivalent to a generalized version of Engle's (1982) LM test for ARCH. (See Theorem 3 following.) Because economic agents normally discount past information, the older

Journal of Business & Economic Statistics
January 1999, Vol. 17, No. 1

the information, the less effect it has on current volatility. Therefore, when a relatively large q is used, it seems that uniform weighting tests are likely not fully efficient; a better test should put greater weight on recent information. Indeed, it can be shown that, when a large q is used, many nonuniform kernels deliver more powerful tests than the truncated kernel. Therefore, our approach might be more powerful than the Box–Pierce–Ljung and LM tests in many practical situations. In fact, a linearly declining weighting scheme with a relatively long lag often has been used in testing and estimating ARCH models (e.g., Engle 1982, 1983; Engle and Kraft 1983; Engle, Lilien, and Robins 1987; Bera and Higgins 1993). Among other things, such a scheme increases power, although it has been criticized by Bollerslev (1986) and others as an ad hoc approach. In comparison, our kernel-based weighting scheme arises endogenously from the frequency-domain approach. For persistent ARCH effects, frequency-domain inference suggests that it is better to employ a long lag. Therefore, it is desirable to allow growth of the number of lags with the sample size for such alternatives.

In addition to its relative efficiency gain, our test is also relatively convenient to implement. It is based on a weighted sum of squared autocorrelations of the squared residuals and has a null, asymptotically one-sided normal distribution. Furthermore, we do not require formulation of the alternative model. Our asymptotic theory allows the lag number to be chosen via such data-driven methods as cross-validation. Such methods may reveal some information about the true alternative. With a data-driven q, one can obtain a nonparametric spectral density estimate for the squared residuals. This estimate usually enjoys some optimality properties and contains information on autocorrelations in squared residuals—that is, volatility clustering. This may be appealing when no prior information about the true alternative is available, as is common in practice. In applications of the existing ARCH tests, one usually must calculate statistics for several, possibly many, values of q. It is common that some statistics are significant and some insignificant. Consequently, drawing conclusions from such a sequence can be a delicate business because these statistics are not independent.

In Section 1, we introduce a class of new ARCH tests. In Section 2, we discuss choice of the lag number for our test via data-dependent methods, particularly cross-validation. In Section 3, we discuss the relationships between our tests and some important existing tests. In Section 4, we conduct a simulation study to compare finite-sample performances of the new tests and those of Engle (1982), Box and Pierce (1970), Ljung and Box (1978), Lee and King (1993), and BDS. We also present two empirical examples in Section 5. In Section 6, we conclude the article. All proofs are collected in the Appendix.

1. A NEW TEST FOR ARCH

Consider the data-generating process (DGP)

$$Y_t = g(X_t, b_o) + \varepsilon_t, \qquad t = 1, \ldots, n, \tag{1}$$

with ARCH error

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

where Y_t is the dependent variable; X_t is a $d \times 1$ vector containing exogenous variables and lagged dependent variables; b_o is an $l \times 1$ unknown true parameter vector in $\mathbb{R}^l$; $g(X_t, b)$ is a given, possibly nonlinear, function such that, for each $b, g(\cdot, b)$ is measurable with respect to Ψ_{t-1}, the information set available at period $t-1$; and $g(X_t, \cdot)$ is twice differentiable with respect to b in an open neighborhood $N(b_o)$ of b_o almost surely, with $\lim_{n\to\infty} n^{-1} \sum_{t=1}^{n} E \sup_{b \in N(b_o)} \|\nabla_b g(X_t, b)\|^4 < \infty$ and $\lim_{n\to\infty} n^{-1} \sum_{t=1}^{n} E \sup_{b\in N(b_o)} \|\nabla_b^2 g(X_t, b)\|^2 < \infty$, where $\|\cdot\|$ is the Euclidean norm in $\mathbb{R}^l$. [For linear models—i.e., $g(X_t, b) = X_t' b$—these dominance conditions reduce to $lim_{n\to\infty} n^{-1} \sum_{t=1}^{n} E\|X_t\|^4 < \infty$.]

The function h_t is a positive, time-varying, and measurable function with respect to Ψ_{t-1}. We assume that ξ_t is iid with $E(\xi_t) = 0, E(\xi_t^2) = 1$, and $E(\xi_t^8) < \infty$. In particular, we do not require that ξ_t be N(0, 1), which may be too restrictive for many high-frequency financial data. In addition, we assume that ξ_t and X_s are mutually independent for $t \geq s$. By definition, ε_t is serially uncorrelated with $E(\varepsilon_t) = 0$. But, its conditional variance, $E(\varepsilon_t^2 | \Psi_{t-1}) = h_t$, may change over time. For example, if h_t follows an ARCH(q_o) process, we have $h_t = \alpha_0 + \sum_{i=1}^{q_o} \alpha_i \varepsilon_{t-i}^2$, where $\alpha_0 > 0$ and $\alpha_i \geq 0$ to ensure positivity of h_t. If h_t follows a GARCH(p_o, q_o) process, $h_t = \alpha_0 + \sum_{i=1}^{q_o} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p_o} \beta_j h_{t-j}$, where $\alpha_0 > 0$.

The null hypothesis of no ARCH effect is H_o: $h_t = \sigma_o^2$ almost surely for some $0 < \sigma_o^2 < \infty$ and all $t = 1, 2, \ldots$. This is equivalent to conditional homoscedasticity. Define $u_t = \varepsilon_t^2/\sigma_o^2 - 1$, where $\sigma_o^2 = E(\varepsilon_t^2)$. Then, under $H_o, u_t = \xi_t^2 - 1$, a white-noise process. The white-noise hypothesis implies a uniform normalized spectral density or distribution function. When ARCH exists, the spectral density or distribution function will not be uniform in general. Therefore, a test for H_o can be based on the shape of the spectral density or distribution function. Researchers have employed the shape of the spectral distribution function to test various hypotheses on serial dependence (e.g., white noise). These include Anderson (1993), Bartlett (1955), Durbin (1967), Durlauf (1991), and Grenander and Rosenblatt (1953, 1957). See Priestley (1981) and Anderson (1993) for surveys. Extending these works to the ARCH context should be possible, though we have not seen such results in the literature.

Relatively few tests are based on the spectral density function. Although the periodogram is not consistent for the spectral density, one can use Parzen's (1957) smoothed-kernel spectral density estimator to test the white-noise hypothesis, as did Hong (1996). This approach involves the choice of a smoothing parameter, a sometimes delicate business. But, maximal power is often attainable if this parameter is chosen via an appropriate data-driven method. In this article, we extend Hong's (1996) spectral density approach to the ARCH context. Our asymptotic theory here permits a data-dependent smoothing parameter.

Let $f(\omega)$ be the normalized spectral density function of u_t. Consequently, $f(\omega) = f_o(\omega) \equiv 1/2\pi$ for all frequencies $\omega \in [-\pi, \pi]$ under H_o. In contrast, $f(\omega) \neq f_o(\omega)$ in general when ARCH exists. It follows that a test for H_o can be based on the L^2 norm

$$L_2(\hat{f}; f_o) = \left[2\pi \int_{-\pi}^{\pi} (\hat{f}(\omega) - f_o(\omega))^2 \, d\omega\right]^{1/2}, \quad (3)$$

where $\hat{f}$ is a consistent spectral density estimator for f. As will be seen, tests based on (3) can be rather simple to compute; no integration is required.

We now construct an estimator for f. Let $\hat{b}$ be an $n^{1/2}$-consistent estimator for b_o. For example, $\hat{b}$ can be the nonlinear least squares estimator

$$\hat{b} = \underset{b}{\operatorname{argmin}}\, n^{-1} \sum_{t=1}^{n} (Y - g(X_t, b))^2.$$

Such an estimator does not take into account the possible ARCH effects but consistently estimates b_o under both the null and alternative. The estimated residual of (1) is $\hat{\varepsilon}_t = Y_t - g(X_t, \hat{b})$. Put $\hat{u}_t = \hat{\xi}_t^2 - 1$, where $\hat{\xi}_t = \hat{\varepsilon}_t / \hat{\sigma}_n$ and $\hat{\sigma}_n^2 = n^{-1} \sum_{t=1}^{n} \hat{\varepsilon}_t^2$. Then the sample autocorrelation function of $\{u_t\}$ is $\hat{\rho}(j) = \hat{r}(j)/\hat{r}(0), j = 0, \pm 1, \ldots, \pm(n-1)$, where $\hat{r}(j) = n^{-1} \sum_{t=|j|+1}^{n} \hat{u}_t \hat{u}_{t-|j|}$. A kernel estimator for f is given by

$$\hat{f}(\omega) = (2\pi)^{-1} \sum_{j=1-n}^{n-1} k(j/q) \hat{\rho}(j) \cos(j\omega)$$

for $\omega \in [-\pi, \pi]$. Here, the bandwidth $q \equiv q(n) \to \infty$ and $q/n \to 0$. The kernel function k: $\mathbb{R} \to [-1, 1]$ is symmetric, continuous at 0, and all but a finite number of points, with $k(0) = 1$ and $\int_{-\infty}^{\infty} k^2(z)\, dz < \infty$. This implies that, for j small relative to n, the weight given to $\hat{\rho}(j)$ is close to unity, the maximum weight. The square integrability of k implies that $k(z) \to 0$ as $|z| \to \infty$. Thus, eventually, less weight is given to $\hat{\rho}(j)$ as j increases.

Examples of k include the Bartlett, Daniell, general Tukey, Parzen, quadratic-spectral (QS), and truncated kernels. (See, e.g., Priestley 1981, p. 441.) Of these, the Bartlett, general Tukey, Parzen, and truncated kernels are of compact support—$k(z) = 0$ for $|z| > 1$. For these kernels, q is the lag truncation number because lags of order $j > q$ receive zero weight. In contrast, the Daniell and QS kernels are of unbounded support. Here, q is not a truncation point but determines the degree of smoothing for $\hat{f}(\omega)$. When the kernel is of unbounded support, all $n-1$ sample autocorrelations of the squared residual are used.

Following Hong (1996), we define our test statistic

$$Q(q) = \left(\frac{n}{2} L_2^2(\hat{f}; f_o) - C_n(k)\right) / (2D_n(k))^{1/2}$$
$$= \left(n \sum_{j=1}^{n-1} k^2(j/q) \hat{\rho}^2(j) - C_n(k)\right) \Bigg/ (2D_n(k))^{1/2},$$

where the second equality follows by Parseval's identity,

$$C_n(k) = \sum_{j=1}^{n-1} (1 - j/n) k^2(j/q)$$

and

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

We note that $C_n(k)$ and $D_n(k)$ are approximately the mean and variance of $n \sum_{j=1}^{n-1} k^2(j/q) \hat{\rho}^2(j)$. The factors $(1 - j/n)$ and $(1 - j/n)(1 - (j+1)/n)$ can be viewed as finite-sample corrections and are asymptotically negligible. Although we motivate our test using (3), $Q(q)$ involves neither numerical integration nor estimation of $\hat{f}$. This frequency-domain motivation leads to an appropriate choice of q via data-driven methods, as will be seen.

Our first result shows that $Q(q)$ is asymptotically standard normal under H_o.

Theorem 1. Let $q \to \infty, q/n \to 0$. Then under the stated conditions and H_o, $Q(q) \to$ N(0, 1) in distribution.

The test is one-sided because in general $Q(q)$ diverges to positive infinity when ARCH exists. Consequently, appropriate upper-tailed critical values of N(0, 1) must be used. For example, the critical value at the 5% significance level is 1.645.

For large $q, q^{-1} D_n(k) \to D(k) \equiv \int_0^{\infty} k^4(z)\, dz$. Thus, one can replace $D_n(k)$ by $qD(k)$ without affecting the asymptotic distribution of $Q(q)$. Under some additional conditions on k and/or $q, q^{-1} C_n(k) = C(k) + o(q^{-1/2})$, where $C(k) \equiv \int_0^{\infty} k^2(z)\, dz$; we also can replace $C_n(k)$ by $qC(k)$. Thus, a more compact, asymptotically equivalent test statistic is

$$Q^*(q) = \left(n \sum_{j=1}^{n-1} k^2(j/q) \hat{\rho}^2(j) - qC(k)\right) \Bigg/ (2qD(k))^{1/2}.$$

Because $Q(q)$ is based on the whole-sample autocorrelation function $\hat{\rho}(j)$ of the squared residual $\hat{u}_t$, it can detect the whole class of linear dependencies of h_t—that is, autocorrelation in the squared residuals. Hence, it may be powerful against ARCH and GARCH alternatives, especially strongly persistent ones when a long lag is used. The test cannot, however, detect all deviations from conditional homoscedasticity. For example, it has no power if h_t follows a tent map, a typical nonlinearity from chaos theory. Of course, like the LM test, the test has power against many types of nonlinearity.

When the regression model (1) is misspecified, the $Q(q)$ test may falsely reject the null hypothesis H_o. This is not peculiar to $Q(q)$. The same is true of all existing ARCH tests; all, explicitly or implicitly, assume correct specification of the regression model. In the present context, Equations (1) and (2) with existence of a consistent estimator $\hat{b}$ for b_o (cf. Assumptions A.1–A.5 in the Appendix) imply correct specification of regression model (1). Such an assumption makes sense because one is usually interested in ARCH only when the regression model is correctly specified.

The $Q(q)$ test is based on the squared deviation of $\hat{f}(\omega)$ from $f_o(\omega)$, weighted equally for all frequencies ω over $[-\pi, \pi]$. This is appropriate when no prior information is available. One can direct power toward specific directions of interest by giving different weights to different frequencies. For example, if one is interested in detecting strongly persistent ARCH effects, greater weight can be given to frequencies near 0. When nonuniform weighting for frequencies is used, an L_2-norm-based test will involve numerical integration. (We emphasize that weighting for frequencies is different from weighting for lags.)

Alternatively, tests can be based on other appropriate global divergence measures, such as the supremum over $[0, \pi]$ of the absolute value of the deviation $\hat{f}(\omega) - f_o(\omega)$:

$$\Omega(q) = \left(\frac{n}{2}\right)^{1/2} \sup_{\omega \in [0,\pi]} |\hat{f}(\omega) - f_o(\omega)|$$

$$= \eta^{1/2} \sup_{\omega \in [0,\pi]} \left| \sum_{j=1}^{n-1} k(j/q)\hat{\rho}(j)\sqrt{2}\cos(j\omega) \right|.$$

This test is computationally more complicated than $Q(q)$. Following reasoning analogous to, but more complicated than, that of Anderson (1993) and Durlauf (1991), one can expect that $\Omega(q)$ has a nonstandard null asymptotic distribution that is the supremum of a Gaussian process with mean 0 and covariance depending on both ω and k. For analytic simplicity, we focus on $Q(q)$ here, but we study the power of $\Omega(q)$ in our simulations.

2. CROSS-VALIDATION

An important practical issue with $Q(q)$ or $Q^*(q)$ is the choice of q, which may have significant impact on power in finite samples. Clearly, the optimal choice of q depends on the unknown alternative. It is, therefore, desirable to choose q via data-driven methods. Although we need not compute $\hat{f}$ to compute $Q(q)$, our approach suggests that it is appropriate to choose q using an optimality criterion for the estimation of $\hat{f}$. Before we discuss some data-driven methods in detail, we extend Theorem 1, which assumes a nonstochastic q, to allow for a data-dependent bandwidth $\hat{q}$. For this purpose, we further restrict the class of kernels k such that $|k(z_1) - k(z_2)| \le \Delta|z_1 - z_2|$ for any z_1, z_2 and some $0 < \Delta < \infty$, and $|k(z)| \le \Delta|z|^{-\tau}$ for all $z \in \mathbb{R}$ and for some $\tau > \frac{1}{2}$. The Lipschitz condition on k rules out the truncated kernel but allows for most commonly used nonuniform kernels.

Theorem 2. Suppose the data-dependent bandwidth $\hat{q}$ satisfies $\hat{q}/q - 1 = o_P(q^{-((3/2)\nu-1)})$, where $\nu > (2\tau - \frac{1}{2})/(2\tau - 1)$ and q is a nonstochastic bandwidth such that $q \to \infty, q^\nu/n \to 0$. Then, under the stated conditions and $H_o, Q(\hat{q}) - Q(q) = o_P(1)$ and $Q(\hat{q}) \to$ N(0, 1) in distribution.

The range of admissible $\hat{q}$ depends on ν, which in turn depends on τ or k. The smaller is ν, the larger is the range of admissible rates for $\hat{q}$. For kernels with bounded support such as the Bartlett and Parzen kernels, any $\nu > 1$ is allowed because $\tau = \infty$. Consequently, the condition $\hat{q}/q - 1 = o_P(q^{-((3/2)\nu-1)})$ is rather weak. For the Daniell kernel, any $\nu > 3/2$ is allowed because $\tau = 1$. In this case, the rate condition $\hat{q}/q - 1 = o_P(q^{-((3/2)\nu-1)})$ is also easy to satisfy. For such methods as Andrews's (1991) parametric "plug-in" bandwidth, $\hat{q}/q - 1$ vanishes in probability at the parametric rate $n^{-1/2}$ with $q \propto n^{1/5}$ for most kernels, but for such methods as Newey and West's (1994) nonparametric "plug-in" bandwidth, $\hat{q}/q - 1$ vanishes at a rate slower than the parametric rate with $q \propto n^{1/3}$ for the Bartlett kernel. In general, our condition on $\hat{q}$ allows for a variety of data-dependent methods.

One appropriate data-driven procedure is the cross-validation proposed by Beltrão and Bloomfield (1987). This delivers an automatic $\hat{q}$ by minimizing a cross-validated frequency-domain likelihood function. Robinson (1991b) showed that, asymptotically, such chosen $\hat{q}$ minimizes a weighted integrated mean squared error of $\hat{f}$ with suitable weights depending on the true spectral density $f(\omega)$. This global procedure is more appropriate in the present context than the narrow-band methods used in estimating an autocorrelation-consistent covariance matrix (e.g., Andrews 1991; Newey and West 1994) because $Q(q)$ is based on all frequencies over $[-\pi, \pi]$ rather than on frequency 0 only.

The Beltrão–Bloomfield procedure can be described as follows. Define

$$\hat{I}(\lambda) = n^{-1} \left| \sum_{t=0}^{n-1} \hat{u}_t \exp(-i\lambda t) \right|^2, \qquad \lambda \in (-\infty, \infty).$$

This is the periodogram of $\hat{u}_t = \hat{\varepsilon}_t^2/\hat{\sigma}_n^2 - 1$. Note that $\hat{I}$ is periodic in λ with periodicity 2π. Put the Fourier frequencies $\lambda_j = 2\pi j/n$ for $j = 0, 1, \ldots, n-1$. Then define the cross-validated spectral-density-function estimator at the Fourier frequencies as

$$\hat{f}_j(\lambda_j; q) = \sigma_j(q)^{-1} \sum_{l \notin N(n,j)} W(q\lambda_l)\hat{I}(\lambda_j - \lambda_l),$$

where $\sigma_j(q) = \sum_{l \notin N(n,j)} W(q\lambda_l), W(\lambda) = (2\pi)^{-1} \int_{-\infty}^{\infty} k(z) \exp(i\lambda z)\, dz$ is the Fourier transform of a kernel function k, and $N(n, j) = \{0, \pm n, \ldots\} \cup \{2j, 2j \pm n, \ldots\}$ is the set of indices j for which $\hat{I}(\lambda_j - \lambda_l) = \hat{I}(\lambda_j)$. Note that when W is of compact support such that $W(\lambda) = 0$ for $|\lambda| > \pi$, the effective summation is only over l for $|l| \le n/2q$.

The cross-validated $\hat{q}_c$ solves

$$\hat{q}_c = \underset{q \in [a_n, b_n]}{\operatorname{argmin}} \sum_{j=1}^{[n/2-1]} \{\ln \hat{f}_j(\lambda_j; q) + \hat{I}(\lambda_j)/\hat{f}_j(\lambda_j; q)\}, \quad (4)$$

where $[n/2 - 1]$ is the integer part of $n/2 - 1$ and the interval $[a_n, b_n]$ is predetermined. The objective function $\ln \hat{f}_j(\lambda_j; q) + \hat{I}(\lambda_j)/\hat{f}_j(\lambda_j; q)$ is the well-known Whittle approximation for the negative log-likelihood function of $\{\hat{u}_t\}$. Therefore, $\hat{q}_c$ approximately maximizes the log-likelihood of $\{\hat{u}_t\}$. The parameter $\hat{q}_c$ can be real-valued, but an integer-valued $\hat{q}_c$ is more convenient here. Then the problem (4) can be solved using grid search. By letting $b_n \to \infty$ as $n \to \infty, \hat{q}_c$ always will tend to infinity under the alternatives with nonuniform $f(\omega)$. Under the null hypothesis of no ARCH, $\hat{q}_c$ will converge to the lower bound a_n as $n \to \infty$ because the spectrum is flat. Consequently, we also let $a_n \to \infty$ to satisfy the conditions of Theorem 2; namely, $\hat{q}_c \to \infty$. As long as a_n diverges at a slow rate and b_n diverges at a fast rate, the theoretically optimal q_c will fall in this interval, and $\hat{q}_c$ will converge to it (cf. Robinson 1991b). For example, with the Daniell kernel, $q_c = Cn^{1/5}$ for the alternatives for which $f''(\omega)$ exists, where C depends on $f(\omega)$. Then $\hat{q}_c$ will converge to q_c under the alternative as $n \to \infty$ if $a_n/n^{1/5} \to 0$ and $b_n/n^{1/5} \to \infty$. We note that the choices of a_n and b_n are to some degree arbitrary, but these choices are of secondary importance because $\hat{q}_c$ will converge to q_c for large n. This is similar in spirit to Newey and West's (1994) method. Our simulation studies, like those of Beltrão and Bloomfield (1987) and Robinson (1991b), reveal that $\hat{q}_c$ is variable in finite samples. It gives better power, however, than simple "rule-of-thumb" choices of q, suggesting gains from its use. The procedure can be implemented efficiently using the fast Fourier transform (FFT).

We now summarize the procedures to implement the test:

1. Obtain any $n^{1/2}$-consistent estimator $\hat{b}$ and save the residuals $\hat{\varepsilon}_t = Y_t - g(X_t, \hat{b})$.
2. Construct the sample autocorrelation function $\hat{\rho}(j)$ of $\hat{u}_t = \hat{\varepsilon}_t^2/\hat{\sigma}^2 - 1$.
3. Choose a kernel k. For example, we can choose the Daniell kernel $k(z) = \sin(\pi z)/\pi z, z \in (-\infty, \infty)$. Its Fourier transform $W(\lambda) = \pi^{-1}1[|\lambda| \leq \pi]$.
4. Choose $\hat{q}_c$ by the Beltrão–Bloomfield (1987) cross-validation procedure.
5. Compute the statistic $Q(\hat{q}_c)$ and compare it to the upper-tailed critical value C_η of N(0, 1) at significance level η. If $Q(\hat{q}_c) > C_\eta$, then one rejects H_o at level η.

Step 4 is not always necessary. If some q is preferred a priori, one can omit this step. A GAUSS program implementing these steps is available from us.

Although the cross-validated $\hat{q}_c$ is asymptotically optimal for the estimation of $\hat{f}(\omega)$ in terms of an integrated weighted mean squared error criterion, it may not be optimal for power of $Q(q)$. For hypothesis testing, a sensible criterion is to choose q so as to maximize power and/or minimize size distortion. It can be expected that the optimal q chosen this way will generally be different from $\hat{q}_c$, although they may be of the same order of magnitude. This is a theoretically interesting but technically complicated issue that deserves further investigation; we defer it to future work. Nevertheless, our simulation and application suggest that $\hat{q}_c$ often delivers maximal or reasonably good power for $Q(q)$ in finite samples.

3. RELATIONSHIP TO SOME EXISTING TESTS

We now discuss the relationship of our test to some important existing tests for ARCH. We will show that, when q is large, a generalized version of Engle's (1982) LM test, as well as those of Box and Pierce (1970) and Ljung and Box (1978), can be viewed as a special case of our approach with the use of the truncated kernel—that is, uniform weighting. Because many nonuniform kernels deliver better power than the truncated kernel when q is large, we expect that our tests may have greater power than these tests for large q.

Suppose that k is the truncated kernel, $k(z) = 1$ for $|z| \leq 1$ and 0 for $|z| > 1$; $Q(q)$ becomes

$$Q_{\mathrm{TRUN}}(q) = (\mathrm{BP}(q) - q)/(2q)^{1/2},$$

given $q^{3/2}/n \to 0$, where

$$\mathrm{BP}(q) = n \sum_{j=1}^{q} \hat{\rho}^2(j)$$

is a Box and Pierce (1970) type of statistic for the squared residuals. The condition $q^{3/2}/n \to 0$ is stronger than $q/n \to 0$. This ensures $q^{-1}C_n(k) = 1 + o(q^{-1/2})$ and $q^{-1}D_n(k) = 2 + o(1)$ for the truncated kernel. The test BP(q), as shown by McLeod and Li (1983), is asymptotically χ_q^2 under H_o. Intuitively, when q is large, we can transform BP(q) into an N(0, 1) by first subtracting the mean q and dividing by the standard deviation $(2q)^{1/2}$. Therefore $Q_{\mathrm{TRUN}}(q)$ can be viewed as a generalized BP(q) test for large q. In practice, a modified but asymptotically equivalent statistic, originally proposed by Ljung and Box (1978),

$$\mathrm{LB}(q) = n^2 \sum_{j=1}^{q} \hat{\rho}^2(j)/(n-j),$$

often is used for testing ARCH. The weights $n/(n-j)$ are introduced to improve size performance and do not affect asymptotic power. These weights are fundamentally different from those of our $Q(q)$ test, which affects the asymptotic power. The test LB(q) is asymptotically equivalent to BP(q). It follows that $Q_{\mathrm{TRUN}}(q)$ is also asymptotically equivalent to a generalized version of LB(q) under H_o; namely,

$$Q_{\mathrm{TRUN}}(q) - (\mathrm{LB}(q) - q)/(2q)^{1/2} = o_P(1).$$

We now relate a generalized version of Engle's (1982) LM test to $Q_{\mathrm{TRUN}}(q)$. Put $\hat{U} = (\hat{u}_1, \ldots, \hat{u}_n)'$ and $\hat{Z} = (\hat{Z}_1', \ldots, \hat{Z}_n')'$, where $\hat{Z}_j = (1, \hat{\varepsilon}_{t-1}^2, \ldots, \hat{\varepsilon}_{t-q}^2)'$. Then the LM test

$$\begin{aligned}\mathrm{LM}(q) &= n\hat{U}'\hat{Z}(\hat{Z}'\hat{Z})^{-1}\hat{Z}'\hat{U}/\hat{U}'\hat{U} \\ &= nR^2,\end{aligned}$$

where R^2 is the uncentered squared multicorrelation coefficient from the ARCH regression

$$\hat{\varepsilon}_t^2 = \alpha_0 + \alpha_1\hat{\varepsilon}_{t-1}^2 + \alpha_2\hat{\varepsilon}_{t-2}^2 + \cdots + \alpha_q\hat{\varepsilon}_{t-q}^2 + v_{tq}, \qquad t = 1, \ldots, n,$$

with the initial values $\hat{\varepsilon}_t^2 = 0$ for $t = -q+1, \ldots, 0$. This test is asymptotically χ_q^2 under H_o and is asymptotically locally most powerful if the true alternative is ARCH(q) with q fixed (cf. Engle 1982, 1984). Lee (1991) showed that a modified LM(q) test for GARCH(p, q) is the same as the LM(q) test for ARCH(q). Granger and Teräsvirta (1993) showed that the LM(q) test is asymptotically equivalent to BP(q) and LB(q) for fixed q.

For large q it is natural to consider the generalized version of Engle's test,

$$Q_{\mathrm{REG}}(q) = (\mathrm{LM}(q) - q)/(2q)^{1/2} = (nR^2 - q)/(2q)^{1/2}.$$

The following theorem shows that $Q_{\mathrm{TRUN}}(q)$ is asymptotically equivalent to $Q_{\mathrm{REG}}(q)$.

Theorem 3. Let $q \to \infty, q^5/n \to 0$. Then, under the stated conditions and $H_o, Q_{\mathrm{TRUN}}(q) - Q_{\mathrm{REG}}(q) = o_P(1), Q_{\mathrm{REG}}(q) \to \mathrm{N}(0, 1)$ in distribution.

Like the LM test, $Q_{\mathrm{REG}}(q)$ can be viewed as a test for the hypothesis that all q coefficients of an ARCH(q) model are jointly equal to 0, where q grows with the sample size n. Intuitively, if the data are homoscedastic, then the variance cannot be predicted by any past squared residuals. If a large variance for $\hat{\varepsilon}_t$ can be predicted by large values of the past squared residuals, however, then ARCH effects are present. Any stationary invertible process in $\hat{\varepsilon}_t^2$ with continuous $f(\omega)$ can be approximated well by a truncated ARCH model of sufficiently high order. Therefore, $Q_{\mathrm{REG}}(q)$ eventually will capture a wide range of alternatives as more lags of $\hat{\varepsilon}_t^2$ are included when n increases. This test is convenient to implement, but it cannot be expected to be fully efficient, as we show next.

The preceding discussions show that the tests of $Q_{\mathrm{TRUN}}(q), Q_{\mathrm{REG}}(q)$, BP($q$), LB($q$), and LM($q$) put uniform, or roughly uniform, weights on all q sample autocorrelations. Intuitively, this might not be the optimal weighting scheme. For most stationary ARCH processes, the autocorrelation decays to 0 as the lag increases. Intuitively, a better test should put greater weight on lower-order lags. We thus expect that tests based on nonuniform weighting may deliver better power than BP(q), LB(q), LM(q), $Q_{\mathrm{TRUN}}(q)$, and $Q_{\mathrm{REG}}(q)$. Using Pitman's relative efficiency criterion, it can be shown that this is indeed the case. Consider the following class of local ARCH alternatives:

$$H_{an}: h_t = \sigma_o^2 \left(1 + (q^{1/4}/n^{1/2}) \sum_{j=1}^{\infty} \alpha_j (\xi_{t-j}^2 - 1)\right),$$

where $\alpha_j \geq 0$ and $\sum_{j=1}^{\infty} \alpha_j < \infty$, which implies $\alpha_j \to 0$ as $j \to \infty$. Without loss of generality, we assume $(q^{1/4}/n^{1/2}) \sum_{j=1}^{\infty} \alpha_j < 1$ for all n to ensure positivity of h_t. Following reasoning analogous to (but more tedious than) that of Hong (1996), it can be shown that, if $q^2/n \to 0$ and $q \to \infty, Q(q) \to \mathrm{N}(\mu, 1)$ in distribution under H_{an} with noncentrality parameter $\mu = (\sum_{j=1}^{\infty} \alpha_j)^2/\{D(k)\}^{1/2}$. Note that, whenever ARCH exists—that is, at least some $\alpha_j > 0$ exist—$Q(q)$ always has nontrivial power because $\mu > 0$. With $q = Cn^{\gamma}$, where $0 < \gamma < 1$ and $0 < C < \infty$, the relative efficiency of the $Q(q)$ test using kernel k_2 with respect to using kernel k_1 under H_{an} is given by

$$\mathrm{EFF}(k_2 : k_1) = (D(k_2)/D(k_1))^{1/(2-\gamma)}.$$

Therefore, the Bartlett kernel $k_B(z) = (1 - |z|)1[|z| \leq 1]$ is about 120% more efficient than the truncated kernel $k_T(z) = 1[|z| \leq 1]$ because $\mathrm{EFF}(k_B : k_T) \geq 5^{1/2}$. As done by Hong (1996), within a suitable class of kernel functions, the Daniell kernel maximizes the power of $Q(q)$ under H_{an}.

The relative efficiency over the LM test does not contradict the well-known result that the LM test is locally most powerful when the true alternative is ARCH(q) with fixed q; here we consider a different regime; namely, $q \to \infty$. Many earlier applications (e.g., Engle 1982, 1983; Engle and Kraft 1983; Engle et al. 1987; Bera and Higgins 1993) used linearly declining weighting schemes with relatively long lags. Such weighting schemes also serve to discount past information. But, as pointed out by Bollerslev (1986), these weighting schemes are somewhat ad hoc. In comparison, our weighting depends on the kernel function k, which arises endogenously from the kernel spectral estimation.

Recently, Lee and King (1993) proposed a locally most mean powerful scored-based (LBS) test for ARCH that exploits the one-sided nature of the ARCH alternative. Their test statistic for ARCH, which is robust to nonnormality, is given by

$$\mathrm{LBS}(q) = \frac{(n-q)\sum_t (\hat{\varepsilon}_t^2/\sigma_n^2 - 1)\sum_{i=1}^q \hat{\varepsilon}_{t-i}^2}{[\sum_t (\hat{\varepsilon}_t^2/\hat{\sigma}_n^2 - 1)^2]\{(n-q)\sum_t (\sum_{i=1}^q \hat{\varepsilon}_{t-i}^2)^2 - (\sum_t \sum_{i=1}^q \hat{\varepsilon}_{t-i}^2)^2\}^{1/2}}. \quad (5)$$

Under H_o, LBS(q) is asymptotically one-sided N(0, 1). Obviously, LBS(q) also puts uniform weight on all q sample autocorrelations of the squared residuals. Consequently, it may not be fully efficient in detecting the ARCH alternatives whose autocorrelation $\rho(j)$ decays to 0 as the lag j increases. Therefore, we expect that $Q(q)$ may be competitive with LBS(q) in some cases, as illustrated in our following simulations and empirical applications. Comparing $\hat{f}(\omega)$ and $f_o(\omega)$ at frequency 0, Hong (1997) constructed a test that exploits the one-sided nature of the ARCH alternative and uses a flexible weighting scheme. Although $Q(q)$ does not exploit this, it is asymptotically more efficient than Hong's (1997) test because $Q(q)$ can detect a class of local alternatives of $O(q^{1/4}/n^{1/2})$. Hong's (1997) test only can detect a class of local alternatives of $O(q^{1/2}/n^{1/2})$. The weighting schemes of $Q(q)$ and Hong's (1997) test also differ.

4. MONTE CARLO EVIDENCE

We now study the finite-sample performances of our tests in comparison to a variety of existing ARCH tests. Consider the following DGP: $Y_t = X_t'b_o + \varepsilon_t, \varepsilon_t = \xi_t h_t^{1/2}$, where $\xi_t \sim \mathrm{NID}(0, 1)$ and $X_t = (1, m_t)'$ with $m_t = \lambda m_{t-1} + v_t$ and $v_t \sim \mathrm{NID}(0, \sigma_v^2)$. This model was first used by Engle, Hendry, and Trumble (1985). We consider four processes for h_t—(1) $h_t = \omega$, (2) $h_t = \omega + \alpha\varepsilon_{t-1}^2$, (3) $h_t = \omega + \alpha(\sum_{i=1}^8 1/i^2)^{-1}\sum_{i=1}^8 \varepsilon_{t-i}^2/i^2$, and (4) $h_t = \omega + \alpha\varepsilon_{t-1}^2 + \beta h_{t-1}$.

Under (1), ARCH is not present. This permits us to examine the sizes. Alternative (2) is an ARCH(1) process often examined in existing simulation studies (e.g., Engle et al. 1985; Diebold and Pauly 1989; Luukkonen, Saikkonen, and Terasvirta 1988; Bollerslev and Wooldridge 1992; Lee and King 1993). Alternative (3) is an ARCH(8) process with weights declining at a geometric rate as the lag increases. Finally, alternative (4) is a GARCH(1, 1) process. The GARCH model has been the workhorse in the literature.

We set $b_o = (1,1)'$ and $\omega = 1$. For the exogenous variable m_t, we set $\lambda = .8$ and $\sigma_v^2 = 4$. As done by Engle et al. (1985), m_t is generated for each experiment and then held fixed from iteration to iteration. For alternatives (2) and (3), we consider two values of α, .3 and .95. For alternative (4), we choose three combinations of (α, β)—(.3, .2), (.5, .2), and (.3, .65). We consider sample sizes of $n = 2^m$ for $m = 6$, 7, 8, and 9. These sample sizes are chosen for convenience because we employ the Cooley–Tukey FFT algorithm in our cross-validation procedure. (Our procedure is applicable when n is not a power of 2, though at some sacrifice of computational efficiency. This sacrifice is costly in simulation.) We set $\varepsilon_t^2 = 0$ for $t \leq 0$ and $h_0 = 1$. To reduce the possible effects of these initial conditions, we generate $n + 100$ observations and then discard the first 100. The iteration number is 10,000 for (1), and 1,000 for (2)–(4). A GAUSS-386 random-number generator is used.

We choose the Daniell kernel for our tests, $Q_{\text{DAN}}(q)$ and $\Omega_{\text{DAN}}(q)$, and also consider their cross-validated versions, $Q_{\text{CV}} = Q_{\text{DAN}}(\hat{q}_c)$ and $\Omega_{\text{CV}} = \Omega_{\text{DAN}}(\hat{q}_c)$. We only study the powers of Ω_{CV} and $\Omega_{\text{DAN}}(q)$ because their null limit distributions are unknown. We compare our tests to Engle's LM(q), Box and Pierce's BP(q), Ljung and Box's LB(q), Lee and King's LBS(q), the truncated kernel-based test $Q_{\text{TRUN}}(q)$, the generalized LM test $Q_{\text{REG}}(q)$, and BDS$_\gamma(q)$ for $q = 1, \ldots, 20$. We include BDS$_\gamma(q)$ because it has been popular in the literature partly for its capacity to detect ARCH alternatives. This test involves choice of a distance parameter $\gamma \in (0, 1)$ in terms of the data spread. To study the impact of γ on size and power, we consider two values, $\gamma = 1/4$ and 1/2. (We use the BDS code of D. Dechert.)

For economy, we report results for $n = 128$ in detail and briefly mention results for other sample sizes. Figure 1 shows the sizes at the 5% level. The $Q_{\text{DAN}}(q)$ test performs well for all q and overrejects only slightly. For large q, BP(q), $Q_{\text{TRUN}}(q)$, LM(q), and $Q_{\text{REG}}(q)$ all tend to underreject, most seriously for LM(q). The LB(q) test underrejects for small q and performs well for medium and large q. The LBS(q) test has reasonable size for small q, underrejects for medium q, and overrejects for large q. The BDS$_\gamma(q)$ test significantly overrejects for the two choices of γ. Finally, the cross-validated Q_{CV} gives a rejection rate of 7.79%, illustrating some overrejection at the 5% level. (To facilitate comparison, the result for Q_{CV} is reported in each figure as a horizontal line, the same value for each q. Since $\hat{q}_c$ is chosen endogenously it may vary for each iteration of the simulation.) Comparison of $Q_{\text{DAN}}(q)$ and Q_{CV} indicates that this overrejection arises from the additional noise of cross-validation. (Here, the mean of $\hat{q}_c$ is 3.8 with a standard deviation of 3.7; the empirical distribution of $\hat{q}_c$ is left-skewed.) As will be seen, the gain in power from using $\hat{q}_c$ may compensate for this moderate overrejection. At the 10% and 1% levels, the size of LBS(q) exhibits a U-shaped

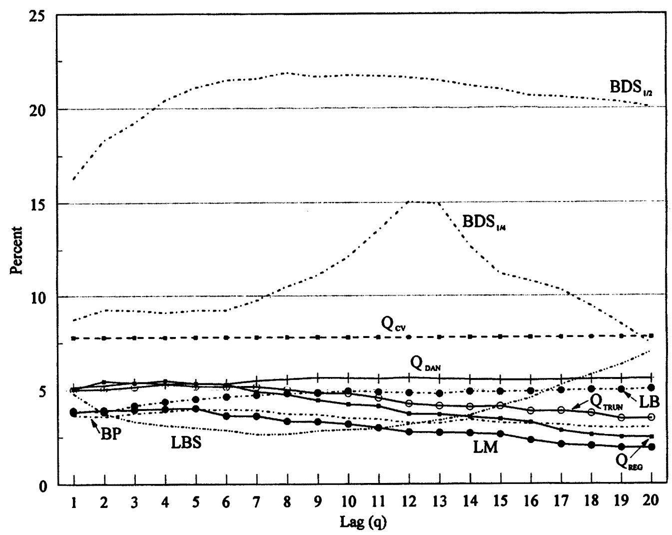

Figure 1. Empirical Size: Rejection Rates of Tests Under the Null at the 5% Level. The simulated model is $y_t = 1 + m_t + \varepsilon_t$, where $\varepsilon_t = \xi_t h_t^{1/2}$, $h_t = 1$, $\xi_t \sim N(0, 1)$, $m_t = \lambda m_{t-1} + v_t$, $\lambda = .8$, and $\nu_t \sim N(0, 4)$. Calculations are based on a sample size of 128 and 10,000 replications. The result for Q_{CV} (for which $\hat{q}_c$ is chosen via cross-validation and, therefore, may change for each iteration) is included for comparison as a horizontal line.

curve, performing best for small q. At the 10% level, the finite-sample correction again improves the sizes of LB(q) versus BP(q), while it induces some overrejection for LB(q) in the 1% level. Cross-validation for Q_{CV} delivers a rejection rate of 11.6% at the 10% level, and a strong rejection rate of 4.2% at the 1% level. The LM(q) test underrejects at both levels. The sizes of the tests improve, to varying degrees, in larger samples. For $n = 512$ and the 5% level, the Q tests have sizes between 5% and 6%; the other tests have sizes between 3% and 5%. The cross-validated Q_{CV} test still exhibits some overrejection, 7% at the 5% level. At the 10% level, the sizes of all of the tests are similar and range between 7% and 11%. At the 1% level, LBS(q), LB(q), and BP(q) perform well, close to 1%. The Q tests show some overrejection but never exceed 2.5%; Q_{CV} remains 3.5%. The LM(q) test exhibits some underrejection.

As suggested by asymptotic theory, the sizes of Q_{CV} improve as the sample size n increases, but very slowly. In a finite sample, one can use bootstrap to obtain more accurate sizes for Q_{CV}. Given a sample of size n, we first obtain the ordinary least squares (OLS) residuals $\hat{u}_t = Y_t - X_t'\hat{b}$, where $\hat{b}$ is the OLS estimator for b_o. We then draw a sample of residuals $\{\hat{u}_t^*\}$ of size n, with replacement, from the empirical distribution function of $\{\hat{u}_t\}$. Next, we obtain a sample $\{Y_t^*\}$ of size n, where $Y_t^* = X_t'\hat{b} + \hat{u}_t^*$. On each sample $\{Y_t^*\}$, we perform the regression described previously and submit the regression residuals to each test. We repeat this for 1,000 iterations and obtain 1,000 bootstrap test statistics, which are used to form the bootstrap empirical distribution function for each test. We then compare the actual test statistics to the bootstrap empirical distribution functions to find the bootstrap p values. The bootstrap sizes of Q_{CV} at the 10%, 5%, and 1% levels are 12.2%, 5.3%, and 1.5%, respectively. These are significantly better than the sizes using asymptotic critical values, especially at the 1% level.

Figure 2 shows powers against ARCH(1) using the 5% empirical critical values. The empirical critical values are obtained from the iterations under (1). First, LBS(1) is slightly more powerful than LM(1), LB(1), and $Q_{\mathrm{DAN}}(1)$, suggesting some gain from exploiting the one-sided nature of the alternative. The tests LM(1), LB(1), $Q_{\mathrm{DAN}}(1), Q_{\mathrm{TRUN}}(1)$, and $Q_{\mathrm{REG}}(1)$ have the same power, slightly better than $\Omega_{\mathrm{DAN}}(1)$ and $\mathrm{BDS}_{1/4}(1)$. Next, the powers of all the tests except $Q_{\mathrm{DAN}}(q)$ decrease rapidly as q increases, most dramatically for LBS(q) and $\mathrm{BDS}_{1/4}(q)$.

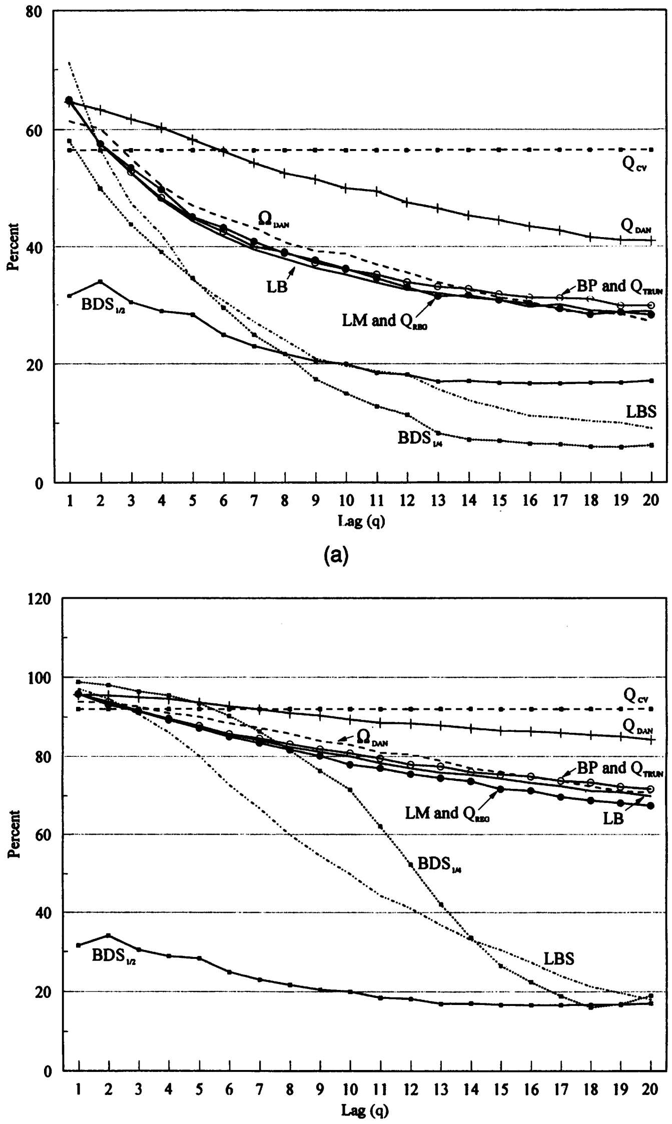

Figure 2. Empirical Power: Rejection Rates of Tests Under the ARCH(1) Alternative at the 5% Level. The simulated model is $y_t = 1 + m_t + \varepsilon_t$, where $\varepsilon_t = \xi_t h_t^{1/2}$, $h_t = 1 + \alpha\varepsilon_{t-1}^2$, $\xi_t \sim N(0, 1)$, $m_t = \lambda m_{t-1} + \nu_t$, $\lambda = .8$, and $\nu_t \sim N(0, 4)$. Calculations are based on a sample size of 128 and 1,000 replications. Rejection rates are based on the size-corrected empirical critical values for the 5% level. The result for Q_{CV} (for which $\hat{q}_c$ is chosen via cross-validation and, therefore, may change for each iteration) is included for comparison as a horizontal line. (a) $\alpha = .3$; (b) $\alpha = .95$.

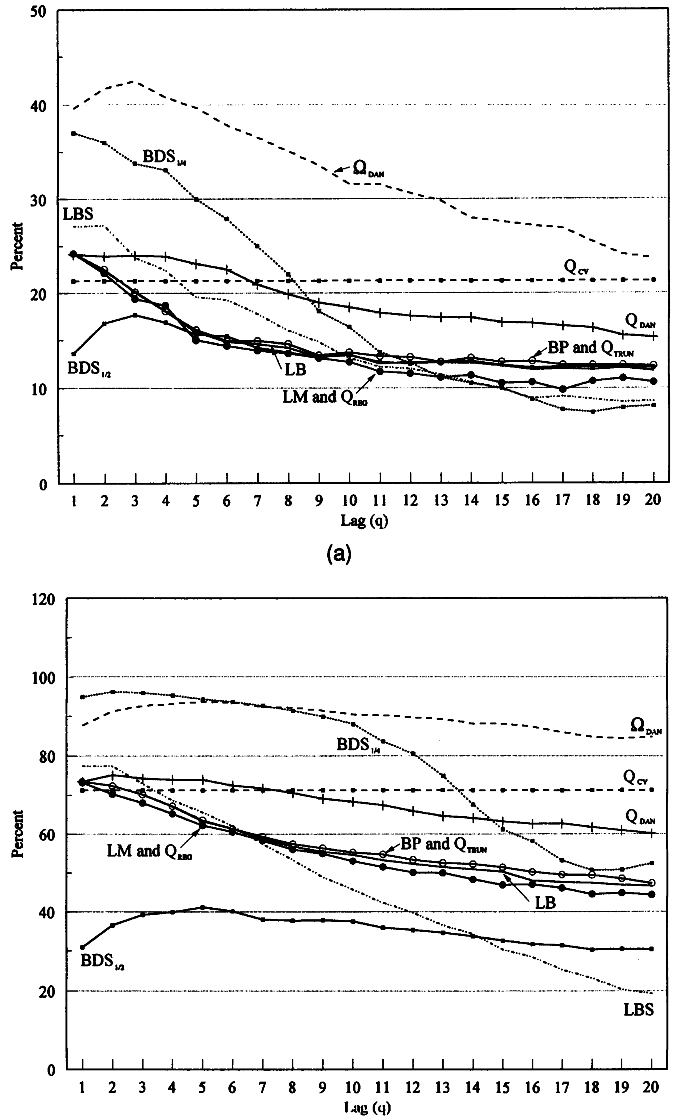

Figure 3. Empirical Power: Rejection Rates of Tests Under the ARCH(8) Alternative at the 5% Level. The simulated model is $y_t = 1 + m_t + \varepsilon_t$, where $\varepsilon_t = \xi_t h_t^{1/2}$, $h_t = 1 + \alpha(\sum_{i=1}^{8} i^{-2}\varepsilon_{t-i}^2)/(\sum_{i=1}^{8} i^{-2})$; $\xi_t \sim N(0, 1)$, $m_t = \lambda m_{t-1} + \nu_t$, $\lambda = .8$, and $\nu_t \sim N(0, 4)$. Calculations are based on a sample size of 128 and 1,000 replications. Rejection rates are based on the size-corrected empirical critical values for the 5% level. The result for Q_{CV} (for which $\hat{q}_c$ is chosen via cross-validation and, therefore, may change for each iteration) is included for comparison as a horizontal line. (a) $\alpha = .3$; (b) $\alpha = .95$.

The tests LM(q), LB(q), BP(q), $Q_{\mathrm{TRUN}}(q)$, $Q_{\mathrm{REG}}(q)$, and $\Omega_{\mathrm{DAN}}(q)$ perform similarly. For $q > 1$, $Q_{\mathrm{DAN}}(q)$ is the most powerful test. Its power loss as q increases is smaller than the losses of the other tests because $Q_{\mathrm{DAN}}(q)$ discounts higher-order lags. The power of $\mathrm{BDS}_{1/4}(q)$ is comparable to that of LBS(q) except for small q: $\mathrm{BDS}_{1/2}(q)$ has low power for small q but performs better than $\mathrm{BDS}_{1/4}(q)$ for $q > 8$. (We report $\gamma = 1/4$ because this choice appears to maximize the power of BDS for $n = 128$ in most cases. We know of no theoretical basis for choosing γ.) For large α, as in Figure 2(b), the power of $Q_{\mathrm{DAN}}(q)$ increases slightly and then decreases as q increases. This is not surprising; the spectral density of an ARCH(1) process includes contributions from $q > 1$ that are substantial for large α. Finally, we note that the cross-validation delivers reasonable power for Q_{CV}, although a little lower than those of LM(1) and LB(1), which use the correct information that the true ARCH order $q_o = 1$. The power of Ω_{CV} is low, 31% and 80%, respectively. (We do not include Ω_{CV} in the figures.)

Next, Figure 3 reports powers against ARCH(8). In Figure 3(a), $\Omega_{\text{DAN}}(q)$ has the highest power for all q, including $q = 8$. The tests $Q_{\text{DAN}}(8)$ and $\text{BDS}_{1/4}(8)$ perform similarly. The Q_{CV} test has roughly the same or better power than $Q_{\text{DAN}}(8)$; Ω_{CV} only has power 27%. The power of $Q_{\text{DAN}}(q)$ increases slightly and then decreases as q increases. Although LBS(8), LM(8), BP(8), and LB(8) are asymptotically locally most powerful against ARCH(8), none of them achieves its maximum power at $q = 8$. The powers of LBS(q), LM(q), LB(q), $Q_{\text{TRUN}}(q)$, and $Q_{\text{REG}}(q)$ attain their maxima at $q = 1$ and then decrease as q increases. Again, LBS(1) has better power than the related tests at $q = 1$. The tests LM(1), LB(1) and $Q_{\text{DAN}}(1)$ have similar power. As q increases, the power of LBS(q) drops dramatically. The tests LM(q), BP(q), LB(q), $Q_{\text{TRUN}}(q)$, and $Q_{\text{REG}}(q)$ perform similarly. $\text{BDS}_{1/4}(q)$ performs quite well for small q but is dominated by $\text{BDS}_{1/2}(q)$ for large q. Again, the power of $Q_{\text{DAN}}(q)$ declines slowly. In Figure 3(b), $\text{BDS}_{1/4}(q)$ dominates $\Omega_{\text{DAN}}(q)$ for $q < 6$.

Figure 4 reports powers against GARCH(1, 1). First, we examine Figure 4, (a) and (b), in which α changes but β is fixed. For $(\alpha, \beta) = (.3, .2)$, the power patterns and ranking of the tests are similar to those under ARCH(1). In particular, all the tests have maximum power at $q = 1$, except $Q_{\text{DAN}}(q)$ and $\Omega_{\text{DAN}}(q)$, which have maximum power at $q = 2$ and 3, respectively. LBS(q) is the most powerful at $q = 1$, and the other tests have similar power. [As pointed out by Lee (1991) and Lee and King (1993), LM(1) and LBS(1) are the appropriate LM and LBS tests for both ARCH(1) and GARCH(1, 1).] For $q > 2$, $Q_{\text{DAN}}(q)$ has the highest power. The power of Q_{CV} is close to $Q_{\text{DAN}}(1)$. For $(\alpha, \beta) = (.5, .2)$, we note two primary differences from the preceding. First, the power curves of $\text{BDS}_\gamma(q)$ and $\Omega_{\text{DAN}}(q)$ are concave and each achieves its maximum for $q > 5$. Second, both the advantage of LBS(1) over the other tests and the cost of using a cross-validated $\hat{q}_c$ are smaller. Last, Figure 4(c) contains results for $(\alpha, \beta) = (.3, .65)$, where $\alpha + \beta$ is close to 1, a characteristic common in empirical research. All of the power curves now become concave; none of the tests achieves its maximum power at $q = 1$. Indeed, it is reasonable that LM(1) and LBS(1) cannot be expected to be optimal here. For persistent GARCH, it is better to include more lags in the tests. Here, the power of $Q_{\text{DAN}}(q)$ decreases very slightly even for large q and achieves its maximum power at q as large as 7. As β increases, the relative power ranking of the tests remains much the same, though the concavity is more pronounced. Finally, we note that for most GARCH(1, 1) processes, cross-validation delivers Q_{CV} with reasonable power, performing better than LM(1), LB(1), and BP(1) in every case, while Ω_{CV} has low power.

[We also examined power against several other parameter values and variance processes including ARCH(4) and

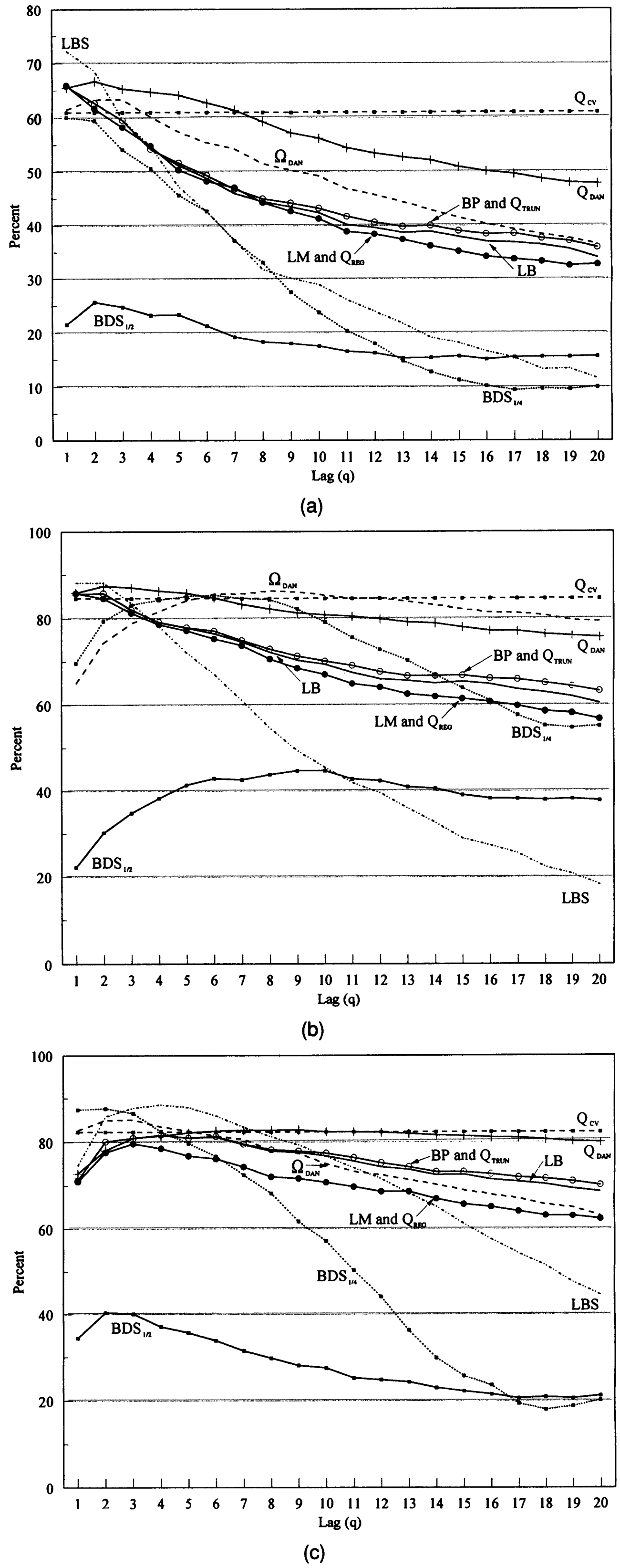

Figure 4. Empirical Power: Rejection Rates of Tests Under the GARCH(1, 1) Alternative at the 5% Level. The simulated model is $y_t = 1 + m_t + \varepsilon_t$, where $\varepsilon_t = \xi_t h_t^{1/2}$, $h_t = 1 + \alpha\varepsilon_{t-1}^2 + \beta h_{t-1}$, $\xi_t \sim N(0, 1)$, $m_t = \lambda m_{t-1} + v_t$, $\lambda = .8$, and $v_t \sim N(0, 4)$. Calculations are based on a sample size of 128 and 1,000 replications. Rejection rates are based on the size-corrected empirical critical values for the 5% level. The result for Q_{CV} (for which $\hat{q}_c$ is chosen via cross-validation and, therefore, may change for each iteration) is included for comparison as a horizontal line. (a) $(\alpha, \beta) = (.3, .2)$; (b) $(\alpha, \beta) = (.5, .2)$; (c) $(\alpha, \beta) = (.3, .65)$

ARCH(12), as well as the Student's t and gamma distributions for ξ_t. The results are largely similar to those presented previously. For some of these additional results, see Hong and Shehadeh (1996).]

In all of our simulations, the mean of $\hat{q}_c$ increases with the sample size and ARCH persistence. For Figure 2(a) ($\alpha = .3$), the mean of $\hat{q}_c$ is 5.8 with a standard deviation of 3.9. For $n = 512$, the mean and standard deviation are 7.7 and 4.7. For Figure 4(c) [$(\alpha, \beta) = (.3, .65)$], the means and standard deviations of $\hat{q}_c$ are 11.1 and 5.0, respectively, for $n = 128$, and 17.2 and 3.6 for $n = 512$. The distributions of $\hat{q}_c$ are skewed left for weak persistence and become increasingly symmetric under stronger persistence.

Summing up, we can draw the following conclusions from our simulation:

1. Both $Q_{\mathrm{DAN}}(q)$ and $\mathrm{LB}(q)$ have reasonable size and perform better than the uniform-weight tests for most q. The $Q_{\mathrm{REG}}(q)$ test has better size performance than $\mathrm{LM}(q)$ for all q; $Q_{\mathrm{TRUN}}(q)$ has better size performance than $\mathrm{BP}(q)$. $\mathrm{BDS}_\gamma(q)$ has poor size performance, overrejecting the null hypothesis in most cases.
2. For each q, our test with nonuniform weights, $Q_{\mathrm{DAN}}(q)$, has power as good or better than the uniform-weighting test, $Q_{\mathrm{TRUN}}(q)$, against all the alternatives under consideration. $\mathrm{LM}(q)$, $\mathrm{BP}(q)$, $\mathrm{LB}(q)$, and $Q_{\mathrm{REG}}(q)$ have powers similar to $Q_{\mathrm{TRUN}}(q)$.
3. Lee and King's $\mathrm{LBS}(q)$ test often has the best power for $q = 1$. For large q, however, $\mathrm{LBS}(q)$ is dominated by all of the other tests, independent of the true alternative. The supremum-norm test $\Omega_{\mathrm{DAN}}(q)$ and $\mathrm{BDS}_{1/4}(q)$ have the best power for persistent ARCH alternatives such as ARCH(8) but perform less well against GARCH alternatives. For large q, $Q_{\mathrm{DAN}}(q)$ is often the most powerful.
4. For persistent GARCH(1, 1) processes, the LM(1) and LBS(1) tests do not have the best power. Instead, these tests have better power when more lags are employed. On the other hand, the power of $Q_{\mathrm{DAN}}(q)$ is only slightly affected by the truncation lag number q over rather wide ranges of q and often is the greatest of all the tests.
5. For our test Q_{CV}, cross-validation exhibits some size overrejection but always delivers good power. To some extent, cross-validation reveals information about the true alternative. The costs associated with misspecification of an unknown true alternative appear high for all of the other tests, though lower for $Q_{\mathrm{DAN}}(q)$. As opposed to $Q_{\mathrm{CV}}, \Omega_{\mathrm{CV}}$ performs poorly relative to $\Omega_{\mathrm{DAN}}(q)$ in each experiment.

5. EMPIRICAL APPLICATIONS

We now apply our tests to two datasets, the U.S. implicit price deflator for gross national product (GNP) and the daily nominal U.S. dollar/Deutschemark exchange rate.

5.1 U.S. GNP Deflator

We consider inflation as measured by the log change in the quarterly implicit price deflator of U.S. GNP, $\pi_t = \ln(\mathrm{GD}_t/\mathrm{GD}_{t-1})$. (These deflator data are reported by the U.S. Bureau of Economic Analysis.) Among others, Engle and Kraft (1983) and Bollerslev (1986) examined this series in detail. These researchers reported that the conditional mean $E(\pi_t|\Psi_{t-1})$ appears well specified by an autoregressive (AR) model. Standard univariate time series methods lead to identification of the following AR(4) model for π_t:

$$\pi_t = \underset{(.076)}{.148} + \underset{(.090)}{.355}\ \pi_{t-1} + \underset{(.094)}{.265}\ \pi_{t-2} + \underset{(.093)}{.199}\ \pi_{t-3} + \underset{(.088)}{.059}\ \pi_{t-4} + \varepsilon_t.$$

The model is estimated on quarterly data from the second quarter of 1952 to the first quarter of 1984, a total of 128 observations. OLS standard errors appear in parentheses and the $R^2 = .648$. The p values of the LM, BP, and LB tests for serial correlation in the mean are .80 or greater for lags between 1 and 20, suggesting an absence of autocorrelation in the residuals. [In the presence of ARCH effects, these standard tests for serial correlation in mean will tend to overreject the null hypothesis of no serial correlation. This fact does not change the conclusion here. But, for an earlier period the BP, LB, and LM tests for serial correlation in mean are all significant at the 10% level for a lag of 5. As Bollerslev (1986, p. 323) noted, "[f]rom the late 1940s until the mid-1950s the inflation rate was ... hard to predict." Although the heteroscedasticity in this period is striking, the mean is difficult to specify. We therefore omit this period and begin our analysis at the earliest date at which serial correlation in mean is not evident with these tests.] We note that the dynamic regression model here is different from the static regression models with exogenous time series variables used in the simulation experiment. Nevertheless, one can expect that ARCH tests will perform similarly under both the regression models with the same ARCH alternative because ARCH tests, based on the squares of residuals, are not very sensitive to the estimation effect even in moderately small samples. The estimation effect has an asymptotically negligible impact on both size and power of ARCH tests, whether the regression model is static or dynamic.

We apply our tests, as well as those of Engle (1982), LB, BP, BDS, and Lee and King (1993) to the estimated squared residuals from the preceding regression. In Table 1, we report the p values for the tests at different q. With use of the Daniell kernel, cross-validation delivers $\hat{q}_c = 6$. Our test $Q_{\mathrm{CV}} = Q_{\mathrm{DAN}}(\hat{q}_c) = 1.901$, giving a p value of .028 using the asymptotic critical value. Thus, our test suggests that ARCH effects exist at the 5% significance level. Engle's LM test statistic $\mathrm{LM}(\hat{q}_c) = 11.00$. This gives a p value of .09; the LM statistic is just significant at the 10% level. Lee and King's test $\mathrm{LBS}(\hat{q}_c) = -2.75$. This has a p value equal to .997, delivering an opposite conclusion. This results from the fact that the sample autocorrelations of the squared residuals are negative at some lags, although these negative autocorrelations are not significant for any given lag at reasonable levels. Likewise, $\mathrm{BDS}_{1/4}(\hat{q}_c) = -.12$ which fails to reject the null. As can be seen in Table 1, the choices of $q = 4, 8$, and 12 deliver significant p val-

Table 1. p Values From Tests for ARCH Effects in the Residuals of the AR(4) Regression of the U.S. GNP Deflator

q	BP	LB	LM	LBS	Q_{TRUN}	Q_{DAN}	$BDS_{1/4}$
1	.442	.437	.511	.253	.614	.547	.792
	(.439)	(.439)	(.515)	(.225)	(.439)	(.362)	(.832)
4	.046	.040	.076	.912	.022	.052	.516
	(.039)	(.039)	(.055)	(.863)	(.039)	(.048)	(.587)
6	.104	.091	.088	.997	.096	.028	.905
	(.077)	(.077)	(.064)	(.989)	(.077)	(.037)	(.937)
8	.172	.150	.154	.999	.187	.032	.724
	(.101)	(.106)	(.104)	(.983)	(.101)	(.041)	(.768)
12	.394	.351	.367	1.000	.446	.070	.422
	(.263)	(.281)	(.282)	(.987)	(.263)	(.062)	(.619)
16	.544	.479	.340	1.000	.589	.138	.007
	(.361)	(.395)	(.234)	(.945)	(.361)	(.094)	(.075)

NOTE: The bootstrap p values appear in parentheses. Cross-validation generated $\hat{q}_c = 6$.

ues (smaller than 10%) for $Q_{\mathrm{DAN}}(q)$, but the choices of $q = 1$ and 16 deliver insignificant p values (larger than 10%). This suggests the importance of a proper choice of q. We see that the cross-validated $\hat{q}_c$ gives the smallest p value here.

We also perform a bootstrap procedure to better compare the tests in this empirical setting. The bootstrap results are largely similar to those using the asymptotic critical values and are reported in parentheses in Table 1. In particular, the relative ranking of the tests is unchanged.

[To confirm the source of the heteroscedasticity, we estimated an ARCH(3) model, in which only the intercept and coefficient on $q = 3$ were significant. We applied the ARCH tests to the new standardized squared residuals and failed to reject the null of homoscedasticity at reasonable levels for $q \leq 20$. To check robustness of the obtained results, we applied the tests to some later samples of the same size. The conclusions remain largely unchanged.]

5.2 Exchange-Rate Data

We now consider a high-frequency series, the daily nominal Deutschemark/U.S. dollar exchange rate. The data, obtained from Chicago Mercantile Exchange, cover the period June 8, 1983, to July 24, 1991—2,048 observations. As noted by Bollerslev et al. (1994), most researchers have reached a consensus that nominal exchange rates are modeled best as integrated processes. We therefore model $s_t \equiv 100 \cdot \ln(\mathrm{DM}_t/\mathrm{DM}_{t-1})$ as a first-order moving average [MA(1)], where DM_t is the spot exchange rate on day t. Quasi-maximum likelihood estimation of this process yields

$$s_t = \underset{(.0149)}{.0190} - \underset{(.0210)}{.0149} \quad \varepsilon_{t-1} + \varepsilon_t.$$

We note that neither the constant nor the MA(1) parameter is significant, consistent with the integration hypothesis.

We test for ARCH effects as in the previous example using the estimated squared residuals. A summary of our results appears in Table 2. All of the tests in Table 2 are significant at any reasonable level for $q = 10, 20, \ldots, 50$ using the asymptotic critical values. Therefore, all the tests under consideration deliver the same conclusion. [Rejection here is not surprising given the sample size. We estimated a GARCH(3, 3) and applied the tests to these new standardized squared residuals. The GARCH parameters were significant at any reasonable level, and all of the tests failed to reject the null of homoscedasticity at any conventional level.] Beyond this, we see that $Q_{\mathrm{DAN}}(q)$ is much larger than Lee and King's $\mathrm{LBS}(q)$ and $\mathrm{BDS}_{1/4}(q)$. Moreover, our $Q_{\mathrm{DAN}}(q)$ is always larger than the truncated kernel-based test $Q_{\mathrm{TRUN}}(q)$ for all q, which can be viewed as a normal approximation of Box and Pierce's $\mathrm{BP}(q)$ test and has approximately the same power as $\mathrm{BP}(q)$ and $\mathrm{LB}(q)$.

Table 2. Test Statistics for ARCH Effects in the Residuals of the MA(1) Regression for the DM/U.S. Dollar Exchange Rate

q	BP	LB	LM	LBS	Q_{TRUN}	Q_{DAN}	$BDS_{1/4}$
10	155.50	155.92	111.45	8.11	32.53	40.44	7.44
20	188.66	189.33	120.30	7.43	26.67	37.04	9.43
29	191.88	192.60	125.61	6.29	21.39	34.00	9.72
30	193.69	194.44	126.30	6.32	21.13	33.62	9.74
40	205.09	206.05	140.78	5.38	18.46	30.49	9.76
50	208.58	209.63	144.11	4.55	15.86	27.91	9.87

NOTE: Cross-validation generated $\hat{q}_c = 29$. Under H_o, BP, LB, and LM are χ^2_q; $\mathrm{BDS}_{1/4}$, LBS, Q_{TRUN}, and Q_{DAN} are N(0, 1).

Of special interest here is the long lag, $\hat{q}_c = 29$, chosen by cross-validation. This result is consistent with other evidence of strong persistence in the conditional volatility of high-frequency data. We have $Q_{\mathrm{CV}} = Q_{\mathrm{DAN}}(\hat{q}_c) = 34.00$, significant at any level. In particular, as we allow the sample size to grow, $\hat{q}_c$ grows as well. For sample sizes of $n = 2^m$ with $m = 6, 7, \ldots, 11$ and a common start date, cross-validation yielded $\hat{q}_c = 1, 12, 8, 11, 19$, and 29, respectively. This trend is consistent with our requirements for the theoretical results.

Again, we performed a bootstrap of each test. The conclusion using our test is the same based on the bootstrap critical values. In three other cases, the conclusion is changed. For lags greater than 20, the $\mathrm{LM}(q)$ and $Q_{\mathrm{REG}}(q)$ tests are now significant only at the 5% level. With the bootstrap critical values, $\mathrm{LBS}(q)$ is significant at the 1% level for $q < 16$, at the 5% level for $q < 20$, and at the 10% level for $q < 25$ and $q > 35$. This is consistent with the U-shaped size curve we observed for $\mathrm{LBS}(q)$ in the Monte Carlo section.

6. CONCLUSIONS

We propose a new ARCH test based on the sum of weighted squares of the sample autocorrelations, with the weights depending on a kernel function. Typically, the kernel function gives greater weight to lower-order lags. When a relatively large lag is employed, Engle's (1982) LM test, as well as the tests of Box and Pierce (1970) and Ljung and Box (1978), is equivalent to the truncated kernel-based test that imposes a uniform weighting scheme and is less efficient than nonuniform kernel-based tests. Our method permits choice of an appropriate lag via data-driven methods—for example, Beltrão and Bloomfield's (1987) cross-validation. Simulation studies show that the new test performs reasonably well in finite samples. The test also is applied in two empirical examples. These empirical examples illustrate some merits of the present test.

Some directions for further research may be pursued. As is well known, it is important to check the adequacy of an estimated GARCH model. For example, misspecification of a GARCH-M model will lead to inconsistent quasi maximum likelihood estimation. Our approach can be extended to check adequacy of a GARCH model. If the model is correctly specified, then the standardized squared residual is a white-noise process; otherwise, there is evidence of misspecification for GARCH. We conjecture that the proposed test also will be asymptotically N(0, 1) under correct specification of the GARCH, but the technicality involved appears nontrivial. These considerations are left for subsequent work.

ACKNOWLEDGMENTS

We thank a coeditor, an associate editor, three referees, and seminar participants at Cornell University for helpful comments and suggestions. Any remaining errors are solely attributable to the authors. We also thank Jau-Lian Jeng for providing us with data.

APPENDIX: MATHEMATICAL PROOFS

For rigor and completeness, we state explicitly the conditions for the theorems.

Assumption A.1. (a) $\{\xi_t\}$ is iid, with $E(\xi_t)=0, E(\xi_t^2)=1$, and $E(\xi_t^8)<\infty$. (b) ξ_t is independent of X_s for $s \leq t$.

Assumption A.2. (a) For each $b \in \mathbb{B} \subseteq \mathbb{R}^l$, where $l \in \mathbb{N}, g(\cdot, b)$ is a measurable function, and (b) $g(X_t,\cdot)$ is twice differentiable with respect to b in an open convex neighborhood $N(b_o)$ of $b_o \in \mathbb{B}$, with $\lim_{n\to\infty} n^{-1}\sum_{t=1}^n E\sup_{b\in N(b_o)}\|\nabla_b g(X_t,b)\|^4<\infty$ and $\lim_{n\to\infty} n^{-1}\sum_{t=1}^n E\sup_{b\in N(b_o)}\|\nabla_b^2 g(X_t,\, b)\|^2<\infty$, where ∇_b and ∇_b^2 are the gradient and Hessian operators, respectively.

Assumption A.3. $n^{1/2}(\hat{b}-b_o)=O_P(1)$.

Assumption A.4. k: $\mathbb{R}\to[-1,1]$ is a symmetric function continuous at 0 and all but a finite number of points, with $k(0)=1$ and $\int_0^\infty k^2(u)\,du<\infty$.

Assumption A.5. For any $z_1, z_2 \in \mathbb{R}, |k(z_1)-k(z_2)| \leq \Delta|z_1-z_2|$ for some $0<\Delta<\infty$ and $|k(z)|\leq \Delta|z|^{-\tau}$ for all $z\in\mathbb{R}$ and some $\tau>\frac{1}{2}$.

Theorem 1. Suppose Assumptions A.1–A.4 hold. Let $q\to\infty, q/n\to 0$. Then under H_o,

$$Q(q)\to^d \text{ N}(0,1).$$

To prove Theorem 1, we first state a lemma.

Lemma A.1. For $j\geq 0$, define $\tilde{r}(j)=n^{-1}\sum_{t=j+1}^n(\xi_t^2-1)(\xi_{t-j}^2-1)$. Then

$$\sum_{j=1}^{n-1} k^2(j/q)\{\hat{r}^2(j)-\tilde{r}^2(j)\}=o_P(q^{1/2}/n).$$

Proof of Theorem 1. Put $r(j)=E\tilde{r}(j)$. We can write

$$\begin{aligned}\sum_{j=1}^{n-1} k^2(j/q)\hat{\rho}^2(j) &= \sum_{j=1}^{n-1} k^2(j/q)\tilde{r}^2(j)/r^2(0)\\ &\quad + \{\hat{r}^{-2}(0)-r^{-2}(0)\}\sum_{j=1}^{n-1}k^2(j/q)\tilde{r}^2(j)\\ &\quad + \hat{r}^{-2}(0)\sum_{j=1}^{n-1}k^2(j/q)\{\hat{r}^2(j)-\tilde{r}^2(j)\}. \qquad (A.1)\end{aligned}$$

By Markov's inequality and $\sum_{j=1}^{n-1}k^2(j/q)E\tilde{r}^2(j)\leq(q/n)r^2(0)\{q^{-1}\sum_{j=1}^{n-1}k^2(j/q)\}=O(q/n)$, where $q^{-1}\sum_{j=1}^{n-1}k^2(j/q)\to\int_0^\infty k^2(z)\,dz$, we have $\sum_{j=1}^{n-1}k^2(j/q)\tilde{r}^2(j)=O_P(q/n)$. Moreover, $\hat{r}(0)-r(0)=O_P(n^{-1/2})$. It follows that the second term in (A.1) is $O_P(q/n^{3/2})=o_P(q^{1/2}/n)$ given $q/n\to 0$. The last term is also $o_P(q^{1/2}/n)$ by Lemma A.1. Therefore, we have

$$\sum_{j=1}^{n-1}k^2(j/q)\hat{\rho}^2(j) = \sum_{j=1}^{n-1}k^2(j/q)\tilde{r}^2(j)/r^2(0)+o_P(q^{1/2}/n). \qquad (A.2)$$

By Hong (1996, theorem A.2), Assumptions A.1 and A.4, and $q\to\infty, q/n\to 0$, we have

$$\left(n\sum_{j=1}^{n-1}k^2(j/q)\tilde{r}^2(j)/r^2(0)-C_n(k)\right) \div (2D_n(k))^{1/2}\to^d \text{N}(0,1). \qquad (A.3)$$

Both (A.2) and (A.3), together with $q^{-1}D_n(k)\to\int_0^\infty k^4(z)\,dz$, imply that $Q(q)\to^d$ N(0, 1). The proof will be completed provided Lemma A.1 is proven.

Proof of Lemma A.1. Noting $\hat{r}^2(j)-\tilde{r}^2(j)=(\hat{r}(j)-\tilde{r}(j))^2+2\tilde{r}(j)(\hat{r}(j)-\tilde{r}(j))$, we write

$$\begin{aligned}\sum_{j=1}^{n-1}k^2(j/q)\{\hat{r}^2(j)-\tilde{r}^2(j)\} &= \sum_{j=1}^{n-1}k^2(j/q)(\hat{r}(j)-\tilde{r}(j))^2\\ &\quad + 2\sum_{j=1}^{n-1}k^2(j/q)\tilde{r}(j)(\hat{r}(j)-\tilde{r}(j)). \qquad (A.4)\end{aligned}$$

We now consider the first term of (A.4). By straightforward algebra, we have

$$\hat{r}(j)-\tilde{r}(j) = n^{-1}\sum_{t=j+1}^n((\hat{\xi}_t^2-1)(\hat{\xi}_{t-j}^2-1)-(\xi_t^2-1)(\xi_{t-j}^2-1))$$

$$= n^{-1}\sum_{t=j+1}^{n}(\xi_t^2-1)(\hat{\xi}_{t-j}-\xi_{t-j}^2)$$
$$+\; n^{-1}\sum_{t=j+1}^{n}(\hat{\xi}_t^2-\xi_t^2)(\xi_{t-j}^2-1)$$
$$+\; n^{-1}\sum_{t=j+1}^{n}(\hat{\xi}_t^2-\xi_t^2)(\hat{\xi}_{t-j}^2-\xi_{t-j}^2)$$
$$= \hat{A}_1(j)+\hat{A}_2(j)+\hat{A}_3(j),\ \text{say}. \tag{A.5}$$

Recall that $\hat{\xi}_t=\hat{\varepsilon}_t/\hat{\sigma}_n, \xi_t=\varepsilon_t/\sigma_o$, and $u_t=\xi_t^2-1$. We have

$$\hat{A}_1(j) = n^{-1}\sum_{t=j+1}^{n}(\xi_t^2-1)(\hat{\xi}_{t-j}-\xi_{t-j}^2)$$
$$= \hat{\sigma}_n^{-2}n^{-1}\sum_{t=j+1}^{n}u_t(\hat{\varepsilon}_{t-j}^2-\varepsilon_{t-j}^2)$$
$$+(\hat{\sigma}_n^{-2}-\sigma_o^{-2})n^{-1}\sum_{t=j+1}^{n}u_t\varepsilon_{t-j}^2$$
$$= \hat{\sigma}_n^{-2}n^{-1}\sum_{t=j+1}^{n}u_t(\hat{\varepsilon}_{t-j}-\varepsilon_{t-j})^2$$
$$+2\hat{\sigma}_n^{-2}n^{-1}\sum_{t=j+1}^{n}u_t\varepsilon_{t-j}(\hat{\varepsilon}_{t-j}-\varepsilon_{t-j})$$
$$+(\hat{\sigma}_n^{-2}-\sigma_o^{-2})n^{-1}\sum_{t=j+1}^{n}u_t\varepsilon_{t-j}^2$$
$$= \hat{\sigma}_n^{-2}\hat{B}_{11}(j)+2\hat{\sigma}_n^{-2}\hat{B}_{12}(j)$$
$$+(\hat{\sigma}_n^{-2}-\sigma_o^{-2})\hat{B}_{13}(j),\ \text{say}. \tag{A.6}$$

By the Cauchy–Schwarz inequality and noting $\hat{\varepsilon}_t-\varepsilon_t=(b_o-\hat{b})'\nabla_b g(X_t,\bar{b})$ for some $\bar{b}$, where $\|\bar{b}-b_o\|\le\|\hat{b}-b_o\|$, we have

$$\sum_{j=1}^{n-1}k^2(j/q)\hat{B}_{11}^2(j)$$
$$=\sum_{j=1}^{n-1}k^2(j/q)\left(n^{-1}\sum_{t=j+1}^{n}u_t(\hat{\varepsilon}_{t-j}-\varepsilon_{t-j})^2\right)^2$$
$$\le\|b_o-\hat{b}\|^4\left(\sum_{j=1}^{n-1}k^2(j/q)\right)\left(n^{-1}\sum_{t=1}^{n}u_t^2\right)$$
$$\times\left(n^{-1}\sum_{t=1}^{n}\|\nabla_b g(X_t,\bar{b})\|^4\right)$$
$$=O_P(q/n^2), \tag{A.7}$$

given A.1–A.4, where $q^{-1}\sum_{j=1}^{n-1}k^2(j/q)\to\int_0^\infty k^2(z)\,dz$ and $n^{-1}\sum_{t=1}^{n}\|\nabla_b g(X_t,\bar{b})\|^4=O_P(1)$.

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

$$\sum_{j=1}^{n-1}k^2(j/q)\hat{B}_{12}^2(j)$$
$$=\sum_{j=1}^{n-1}k^2(j/q)\left(n^{-1}\sum_{t=j+1}^{n}u_t\varepsilon_{t-j}(\hat{\varepsilon}_{t-j}-\varepsilon_{t-j})\right)^2$$
$$\le 4\|b_o-\hat{b}\|^2\sum_{j=1}^{n-1}k^2(j/q)$$
$$\times\left\|n^{-1}\sum_{t=j+1}^{n}u_t\varepsilon_{t-j}\nabla_b g(X_{t-j},b_o)\right\|^2$$
$$+\;2\,\|b_o-\hat{b}\|^4\sum_{j=1}^{n-1}k^2(j/q)$$
$$\times\left\|n^{-1}\sum_{t=j+1}^{n}u_t\varepsilon_{t-j}\nabla_b^2 g(X_{t-j},\bar{b})\right\|^2$$
$$=O_P(q/n^2) \tag{A.8}$$

by Markov's inequality and $E\|n^{-1}\sum_{t=j+1}^{n}u_t\varepsilon_{t-j}\nabla_b g(X_{t-j},b_o)\|^2=O(n^{-1})$ given $E(u_t|\Psi_{t-1})=0$, where Ψ_{t-1} is the information set available at period $t-1$. We also have

$$\sum_{j=1}^{n-1}k^2(j/q)\hat{B}_{13}^2(j)=\sum_{j=1}^{n-1}k^2(j/q)\left(n^{-1}\sum_{t=j+1}^{n}u_t\varepsilon_{t-j}^2\right)^2$$
$$=O_P(q/n) \tag{A.9}$$

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

Combining (A.6)–(A.9) with $\hat{\sigma}_n^2-\sigma_o^2=O_P(n^{-1/2})$, we obtain

$$\sum_{j=1}^{n-1}k^2(j/q)\hat{A}_1^2(j)=O_P(q/n^2). \tag{A.10}$$

Similarly, we have

$$\sum_{j=1}^{n-1}k^2(j/q)\hat{A}_2^2(j)=O_P(q/n^2). \tag{A.11}$$

We now turn to $\sum_{j=1-n}^{n-1}k^2(j/q)\hat{A}_3^2(j)$. By the Cauchy–Schwarz inequality and noting $\hat{\xi}_t^2-\xi_t^2=\hat{\sigma}_n^{-2}(\hat{\varepsilon}_t^2-\varepsilon_t^2)+(\hat{\sigma}_n^{-2}-\sigma_o^{-2})\varepsilon_t^2$, we have

$$\sup_{1\le j\le n-1}|\hat{A}_3(j)|^2\le n^{-1}\sum_{t=1}^{n}(\hat{\xi}_t^2-\xi_t^2)^2$$
$$=2\hat{\sigma}_n^{-4}n^{-1}\sum_{t=1}^{n}(\hat{\varepsilon}_t^2-\varepsilon_t^2)^2$$

$$+ 2(\hat{\sigma}_n^{-2} - \sigma_o^{-2})^2 n^{-1} \sum_{t=1}^{n} \varepsilon_t^4$$
$$= O_P(n^{-2})$$

given Assumptions A.1–A.3. It follows that

$$\sum_{j=1}^{n-1} k^2(j/q)\hat{A}_3^2(j) = O_P(q/n^2). \tag{A.12}$$

Hence, from (A.5) and (A.10)–(A.12), we have

$$\sum_{j=1}^{n-1} k^2(j/q)(\hat{r}(j) - \tilde{r}(j))^2 = O_P(q/n^2). \tag{A.13}$$

On the other hand, because $\sum_{j=1}^{n-1} k^2(j/q)\tilde{r}^2(j) = O_P(q/n)$, we have

$$\sum_{j=1}^{n-1} k^2(j/q)(\hat{r}(j) - \tilde{r}(j))\tilde{r}(j) = o_P(q^{1/2}/n) \tag{A.14}$$

by the Cauchy–Schwarz inequality, (A.13), and $q/n \to 0$. The lemma then follows from (A.4), (A.13)–(A.14), and $q/n \to 0$. This completes the proof.

Theorem 2. Suppose Assumptions A.1–A.5 hold. Let $\hat{q}$ be a data-dependent bandwidth such that $\hat{q}/q - 1 = o_P(q^{-((3/2)\nu-1)})$ for some $\nu > (2\tau - \frac{1}{2})/(2\tau - 1)$, where $q \to \infty, q^\nu/n \to 0$. Then, under $H_o, Q(\hat{q}) - Q(q) = o_P(1)$, and

$$Q(\hat{q}) \to^d \mathrm{N}(0,1).$$

To show Theorem 2, we first state two lemmas.

Lemma A.2. Let $\hat{C}_n(k) = \sum_{j=1}^{n-1}(1 - j/n)k^2(j/\hat{q})$ and $\hat{D}_n(k) = \sum_{j=1}^{n-1}(1 - j/n)(1 - (j+1)/n)k^4(j/\hat{q})$. Then $q^{-1}\hat{C}_n(k) = q^{-1}C_n(k) + o_P(q^{-1/2})$, and $q^{-1}\hat{D}_n(k) = q^{-1}D_n(k) + o_P(1)$.

Lemma A.3. $\sum_{j=1}^{n-1}\{k^2(j/\hat{q}) - k^2(j/q)\}\hat{r}^2(j) = o_P(q^{1/2}/n)$.

Proof of Theorem 2. From the definition of $Q(q)$, we can write

$$\begin{aligned} Q(\hat{q}) - Q(q) &= \{\hat{D}_n(k)\}^{-1/2}\{\hat{C}_n(k) - C_n(k)\} \\ &\quad + Q(q)\{(D_n(k)/\hat{D}_n(k))^{1/2} - 1\} \\ &\quad + \{\hat{D}_n(k)\}^{-1/2}\left\{n\sum_{j=1}^{n-1}(k^2(j/\hat{q}) - k^2(j/q))\hat{\rho}^2(j)\right\}. \end{aligned}$$

Here, the first term vanishes in probability by Lemma A.2 and $\{\hat{D}_n(k)\}^{-1/2} = O_P(q^{-1/2})$, the latter follows from Lemma A.2, and $q^{-1}D_n(k) \to \int_0^\infty k^4(z)\,dz$. Next, the second term vanishes in probability because $\hat{D}_n(k)/D_n(k) \to^p 1$ by Lemma A.2 and $Q(q) = O_P(1)$ by Theorem 1. Finally, the last term vanishes in probability by Lemma A.3 and $\{\hat{D}_n(k)\}^{-1/2} = O_P(q^{-1/2})$. Consequently, $Q(\hat{q}) - Q(q) = o_P(1)$, so $Q(\hat{q}) \to^d$ N(0, 1) given $Q(q) \to^d$ N(0, 1) by Theorem 1. The proof is completed provided Lemmas A.2–A.3 are proven.

Proof of Lemma A.2. Let integer $d = [q^\nu]$, where $[x]$ denotes the integer part of x. Then $d/n \to 0, d/q \to \infty$. Write

$$\begin{aligned} \hat{C}_n(k) &= C_n(k) + \sum_{j=1}^{d}(1 - j/n)\{k^2(j/\hat{q}) - k^2(j/q)\} \\ &\quad + \sum_{j=d+1}^{n-1}(1 - j/n)k^2(j/\hat{q}) \\ &\quad - \sum_{j=d+1}^{n-1}(1 - j/n)k^2(j/q) \\ &= C_n(k) + \hat{C}_{1n} + \hat{C}_{2n} - \hat{C}_{3n}, \text{ say.} \end{aligned} \tag{A.15}$$

We show that the last three terms are $o_P(q^{1/2})$. First, we consider $\hat{C}_{2n}$. Given $|k(z)| \le \Delta|z|^{-\tau}$ in Assumption A.5 and $\hat{q}/q \to^p 0$, we have

$$\begin{aligned} |\hat{C}_{2n}| &\le \Delta^2 q^{2\tau}(\hat{q}/q)^{2\tau}\sum_{j=d+1}^{n-1} j^{-2\tau} \\ &= O_P(q^{2\tau}/d^{2\tau-1}) = o_P(q^{1/2}), \end{aligned} \tag{A.16}$$

where the last equality follows from $d = [q^\nu]$ for $\nu > (2\tau - \frac{1}{2})/(2\tau - 1)$. Similarly,

$$\hat{C}_{3n} = o(q^{1/2}). \tag{A.17}$$

For the second term in (A.15), we decompose

$$\hat{C}_{1n} \le \sum_{j=1}^{d}\{k(j/\hat{q}) - k(j/q)\}^2 + 2\sum_{j=1}^{d}|\{k(j/\hat{q}) - k(j/q)\}k(j/q)|.$$

Given the Lipschitz condition on k (in Assumption A.5), $\hat{q}/q - 1 = o_P(q^{-((3/2)\nu-1)})$, and $d = [q^\nu]$, we have

$$\sum_{j=1}^{d}\{k(j/\hat{q}) - k(j/q)\}^2 \le \Delta^2 q^{-2}(\hat{q}/q)^2(\hat{q}/q - 1)^2\sum_{j=1}^{d} j^2 = o_P(1).$$

This, together with $\sum_{j=1}^{d} k^2(j/q) = O(q)$ and the Cauchy–Schwarz inequality, implies

$$\sum_{j=1}^{d}|\{k(j/\hat{q}) - k(j/q)\}k(j/q)| = o_P(q^{1/2}).$$

It follows that

$$\hat{C}_{1n} = o_P(q^{1/2}). \tag{A.18}$$

Collecting (A.15)–(A.18) yields $q^{-1}\hat{C}_n(k) = q^{-1}C_n(k) + o_P(q^{-1/2})$. The result $q^{-1}\hat{D}_n(k) = q^{-1}D_n(k) + o_P(1)$ can be shown analogously.

Proof of Lemma A.3. Recalling the definition of $\tilde{r}(j)$ in Lemma A.1, we can write

$$\sum_{j=1}^{n-1}\{k^2(j/\hat{q})-k^2(j/q)\}\hat{\rho}^2(j)$$
$$= \hat{r}^{-2}(0)\sum_{j=1}^{n-1}\{k^2(j/\hat{q})-k^2(j/q)\}\tilde{r}^2(j)$$
$$+ \hat{r}^{-2}(0)\sum_{j=1}^{n-1}\{k^2(j/\hat{q})-k^2(j/q)\}$$
$$\times \{\hat{r}^2(j)-\tilde{r}^2(j)\}. \quad \text{(A.19)}$$

For the first term of (A.19), we have

$$\sum_{j=1}^{n-1}\{k^2(j/\hat{q})-k^2(j/q)\}\tilde{r}^2(j)$$
$$= \sum_{j=1}^{d}\{k^2(j/\hat{q})-k^2(j/q)\}\tilde{r}^2(j) + \sum_{j=d+1}^{n-1}k^2(j/\hat{q})\tilde{r}^2(j)$$
$$- \sum_{j=d+1}^{n-1}k^2(j/q)\tilde{r}^2(j)$$
$$= \hat{D}_{1n}+\hat{D}_{2n}-\hat{D}_{3n}, \text{ say.} \quad \text{(A.20)}$$

We first consider $\hat{D}_{2n}$. Given $|k(z)|\le \Delta|z|^{-\tau}$, we have

$$\hat{D}_{2n} \le \Delta^2 q^{2\tau}(\hat{q}/q)^{2\tau}\left\{\sum_{j=d+1}^{n-1} j^{-2\tau}\tilde{r}^2(j)\right\}$$
$$= o_P(q^{1/2}/n), \quad \text{(A.21)}$$

given $\hat{q}/q \to^p 1$ and $d=[q^\nu]$ for $\nu > (2\tau-\frac{1}{2})/(2\tau-1)$, where $\sum_{j=d+1}^{n-1} j^{-2\tau}\tilde{r}^2(j) = O_P(d^{1-2\tau}/n)$ by Markov's inequality and $E\tilde{r}^2(j)=O(n^{-1})$. Similarly, we have

$$\hat{D}_{3n} = o_P(q^{1/2}/n). \quad \text{(A.22)}$$

Next, we consider the first term in (A.20). We write

$$\hat{D}_{1n} = \sum_{j=1}^{d}\{k(j/\hat{q})-k(j/q)\}^2\tilde{r}^2(j)$$
$$+ 2\sum_{j=1}^{d}\{k(j/\hat{q})-k(j/q)\}k(j/q)\tilde{r}^2(j)$$
$$= \hat{D}_{11n}+2\hat{D}_{12n}, \text{ say.} \quad \text{(A.23)}$$

Given the Lipschitz condition on k (Assumption A.5), we have

$$\hat{D}_{11n} \le \Delta^2 q^{-2}(q/\hat{q})^2(\hat{q}/q-1)^2\sum_{j=1}^{d} j^2\tilde{r}^2(j)$$
$$= \Delta^2 q^{-2} o_P(q^{-(3\nu-2)})O_P(d^3/n)$$
$$= o_P(n^{-1}), \quad \text{(A.24)}$$

where $\hat{q}/q-1 = o_P(q^{-((3/2)\nu-1)})$ and $\sum_{j=1}^{d} j^2\tilde{r}^2(j) = O_P(d^3/n)$ by Markov's inequality. Next, by the Cauchy–Schwarz inequality and $\sum_{j=1}^{d} k^2(j/q)\tilde{r}^2(j) = O_P(q)$, we have

$$|\hat{D}_{12n}| \le (\hat{D}_{11n})^{1/2}\left\{\sum_{j=1}^{d} k^2(j/q)\tilde{r}^2(j)\right\}^{1/2}$$
$$= o_P(q^{1/2}/n). \quad \text{(A.25)}$$

It follows from (A.23)–(A.25) that

$$\hat{D}_{1n} = o_P(q^{1/2}/n). \quad \text{(A.26)}$$

Combining (A.20)–(A.22) and (A.26), we obtain that, for the first term of (A.19),

$$\sum_{j=1}^{n-1}\{k^2(j/\hat{q})-k^2(j/q)\}\tilde{r}^2(j) = o_P(q^{1/2}/n). \quad \text{(A.27)}$$

Next, we consider the second term of (A.19). Because $\sum_{j=1}^{n-1} k^2(j/q)\{\hat{r}^2(j)-\tilde{r}^2(j)\} = o_P(q^{1/2}/n)$ by Lemma A.1, it suffices to show $\sum_{j=1}^{n-1} k^2(j/\hat{q})\{\hat{r}^2(j)-\tilde{r}^2(j)\} = o_P(q^{1/2}/n)$. Put

$$\hat{W}_n = (n/q^{1/2})\sum_{j=1}^{n-1} k^2(j/\hat{q})|\hat{r}^2(j)-\tilde{r}^2(j)|.$$

We shall show $\hat{W}_n = o_P(1)$. For any given $\eta>0, \delta>0$,

$$P(\hat{W}_n>\eta) \le P(\hat{W}_n>\eta, |\hat{q}/q-1|\le\delta) + P(|\hat{q}/q-1|>\delta),$$

where the last probability converges to 0 as $n\to\infty$, given $\hat{q}/q-1 = o_P(1)$. Thus, it remains to show that the first probability also vanishes. Put

$$\bar{k}(z) = \begin{cases} 1 & \text{if } |z|\le z_0 \\ \Delta|z|^{-\tau} & \text{if } |z|>z_0, \end{cases}$$

where $0<z_0<\infty$ and Δ, τ are given in Assumption A.5. Then $|k(z)|\le \bar{k}(z)$, and $\bar{k}(z)$ is monotonically decreasing on $\mathbb{R}^+ = [0,\infty]$; that is, $\bar{k}(z_1)\ge\bar{k}(z_2)$ if $0\le z_1\le z_2$. Furthermore, we have $\int_0^\infty \bar{k}^2(z)\,dz<\infty$, given $\tau>\frac{1}{2}$. It follows that $|k(j/\hat{q})|\le\bar{k}(j/\hat{q})\le\bar{k}\{j/(1-\delta)q\}$ if $\hat{q}\ge(1-\delta)q$. Whence, the first probability will vanish as $n\to\infty$ if

$$(n/(1-\delta)q^{1/2})\sum_{j=1}^{n-1}\bar{k}^2\{j/(1-\delta)q\}|\hat{r}^2(j)-\tilde{r}^2(j)| = o_P(1)$$

for any given small $\delta>0$, which follows by a reasoning analogous to the proof of Lemma A.1 and the fact that $q^{-1}\sum_{j=1}^{n-1}\bar{k}^2(j/q)\to\int_0^\infty\bar{k}^2(z)\,dz<\infty$ as $q\to\infty$. Therefore, we have $\hat{W}_n = o_P(1)$, or equivalently,

$$\sum_{j=1}^{n-1} k^2(j/\hat{q})|\hat{r}^2(j)-\tilde{r}^2(j)| = o_P(q^{1/2}/n). \quad \text{(A.28)}$$

Combining (A.19), (A.27)–(A.28), and Lemma A.1 then yields the desired result.

Theorem 3. Suppose Assumptions A.1–A.3 hold. Let $q \to \infty, q^5/n \to 0$. Then under H_o, $Q_{\text{TRUN}}(q) - Q_{\text{REG}}(q) = o_P(1)$ and

$$Q_{\text{REG}}(q) \to^d \text{N}(0,1).$$

Proof of Theorem 3. Put $\hat{Z} = (\hat{Z}_1, \ldots, \hat{Z}_n)', \hat{Z}_t = (1, \hat{\varepsilon}_{t-1}^2, \ldots, \hat{\varepsilon}_{t-q}^2)', \hat{U} = (\hat{\varepsilon}_1^2/\hat{\sigma}_n^2 - 1, \ldots, \hat{\varepsilon}_n^2/\hat{\sigma}_n^2 - 1)$, and $\hat{R}_q = (\hat{R}(1), \ldots, \hat{R}(q))'$, where $\hat{R}(j) = n^{-1}\sum_{t=j+1}^n (\hat{\varepsilon}_t^2 - \hat{\sigma}_n^2)(\hat{\varepsilon}_{t-j}^2 - \hat{\sigma}_n^2)$. Then by definition (e.g., see Engle 1982), we have

$$R^2 = \hat{U}'\hat{Z}(\hat{Z}'\hat{Z})^{-1}\hat{Z}'\hat{U}/(\hat{U}'\hat{U}).$$

By straightforward algebra, we have $R^2 = (n\hat{R}^2(0))^2 \hat{R}_q'\hat{A}^{-1}\hat{R}_q$, where $\hat{A}$ is a $q \times q$ symmetric matrix with the (i,j) element

$$\hat{A}_{ij} = n^{-1} \sum_{t=\max(i,j)+1}^{n} (\hat{\varepsilon}_{t-i}^2 - \hat{\sigma}_i^2)(\hat{\varepsilon}_{t-j}^2 - \hat{\sigma}_j^2)$$

and $\hat{\sigma}_j^2 = n^{-1}\sum_{t=j+1}^n \hat{\varepsilon}_{t-j}^2$. Because $n\sum_{j=1}^q \hat{\rho}^2(j) = n\sum_{j=1}^q \hat{R}^2(j)/\hat{R}^2(0)$, it follows that

$$\begin{aligned} nR^2 - n\sum_{j=1}^{q} \hat{\rho}^2(j) &= n\hat{r}^{-2}(0)(\hat{R}(0)\hat{R}_q'\hat{A}^{-1}\hat{R}_q - \hat{R}_q'\hat{R}_q) \\ &= nR^2\hat{R}^{-2}(0)\hat{\theta}_q'(\hat{R}(0)I_q - \hat{A})\hat{\theta}_q, \end{aligned} \tag{A.29}$$

where $\hat{\theta}_q = \hat{A}^{-1/2}\hat{R}_q/\sqrt{R^2}$ is a $q \times 1$ unit vector in $\Theta_q = \{\theta \in \mathbb{R}^q\colon \theta'\theta = 1\}$.

Put $S_n \equiv \sup_{\theta \in \Theta_q} \theta'(\hat{r}(0)I_q - \hat{A})\theta$. We first show $S_n = O_P(q^2/n^{1/2})$. For this, we write

$$\begin{aligned} |S_n| &= \sup_{\theta\in\Theta_q}\left|\sum_{i=1}^q\sum_{j=1}^q n^{-1}\left(\sum_{t=1}^n(\hat{\varepsilon}_t^2 - \hat{\sigma}_n^2)\delta_{ij} \right.\right. \\ &\quad \left.\left. - \sum_{t=\max(i,j)+1}^{n}(\hat{\varepsilon}_{t-i}^2 - \hat{\sigma}_i^2)(\hat{\varepsilon}_{t-j}^2 - \hat{\sigma}_j^2)\right)\theta_i\theta_j\right| \\ &\le \sup_{\theta\in\Theta_q}\left|\sum_{j=1}^q n^{-1}\left(\sum_{t=1}^n(\hat{\varepsilon}_t^2 - \hat{\sigma}_n^2)^2 \right.\right. \\ &\quad \left.\left. - \sum_{t=j+1}^n(\hat{\varepsilon}_{t-j}^2 - \hat{\sigma}_j^2)^2\right)\theta_j^2\right| \\ &\quad + 2\sup_{\theta\in\Theta_q}\left|\sum_{j=2}^q\sum_{i=1}^{j-1} n^{-1}\sum_{t=j+1}^n \right. \\ &\quad \left. \times (\hat{\varepsilon}_{t-i}^2 - \hat{\sigma}_i^2)(\hat{\varepsilon}_{t-j}^2 - \hat{\sigma}_j^2)\theta_i\theta_j\right| \\ &= B_{1n} + 2B_{2n}, \text{ say.} \end{aligned} \tag{A.30}$$

For the first term of (A.30), we have

$$\begin{aligned} B_{1n} &= \sup_{\theta\in\Theta_q}\left|\sum_{j=1}^q \theta_j^2\left(n^{-1}\sum_{t=1}^n(\hat{\varepsilon}_t^4 - \hat{\sigma}_n^4) \right.\right. \\ &\quad \left.\left. - n^{-1}\sum_{t=j+1}^n(\hat{\varepsilon}_{t-j}^4 - \hat{\sigma}_j^4)\right)\right| \\ &= \sup_{\theta\in\Theta_q}\left|\sum_{j=1}^q\theta_j^2\left(n^{-1}\sum_{t=n-j+1}^n\hat{\varepsilon}_t^4 \right.\right. \\ &\quad \left.\left. - (\hat{\sigma}_n^2 + \hat{\sigma}_j^2)(\hat{\sigma}_n^2 - \hat{\sigma}_j^2)\right)\right| \\ &\le \sup_{\theta\in\Theta_q}\left|\sum_{j=1}^q\theta_j^2\left(n^{-1}\sum_{t=n-j+1}^n\hat{\varepsilon}_t^4\right)\right| \\ &\quad + \sup_{\theta\in\Theta_q}\left|\sum_{j=1}^q\theta_j^2(\hat{\sigma}_n^2 + \hat{\sigma}_j^2)\left(n^{-1}\sum_{t=j+1}^n\hat{\varepsilon}_{t-j}^2\right)\right| \\ &\le n^{-1}\sum_{t=n-q+1}^n\hat{\varepsilon}_t^4 + 2\hat{\sigma}_n^2 n^{-1}\sum_{t=n-q+1}^n\hat{\varepsilon}_t^2 \\ &= O_P(q/n). \end{aligned} \tag{A.31}$$

For the second term of (A.30), we have

$$\begin{aligned} B_{2n} &\le \sum_{j=2}^q\sum_{i=1}^{j-1}\left|n^{-1}\sum_{t=j+1}^n(\hat{\varepsilon}_{t-i}^2 - \hat{\sigma}_i^2)(\hat{\varepsilon}_{t-j}^2 - \hat{\sigma}_j^2)\right| \\ &= \sum_{j=2}^q\sum_{i=1}^{j-1}\left|n^{-1}\sum_{t=j+1}^n\hat{\varepsilon}_{t-i}^2(\hat{\varepsilon}_{t-j}^2 - \hat{\sigma}_j^2)\right| \\ &\le \sum_{j=2}^q\sum_{i=1}^{j-1}\left|n^{-1}\sum_{t=j+1}^n\hat{\varepsilon}_{t-i}^2(\hat{\varepsilon}_{t-j}^2 - \varepsilon_{t-j}^2)\right| \\ &\quad + \sum_{j=2}^q\sum_{i=1}^{j-1}\left|n^{-1}\sum_{t=j+1}^n(\hat{\varepsilon}_{t-i}^2 - \varepsilon_{t-i}^2)(\varepsilon_{t-j}^2 - \hat{\sigma}_j^2)\right| \\ &\quad + \sum_{j=2}^q\sum_{i=1}^{j-1}\left|n^{-1}\sum_{t=j+1}^n\varepsilon_{t-i}^2(\hat{\sigma}_j^2 - \sigma_o^2)\right| \\ &\quad + \sum_{j=2}^q\sum_{i=1}^{j-1}\left|n^{-1}\sum_{t=j+1}^n\varepsilon_{t-i}^2(\varepsilon_{t-j}^2 - \sigma_o^2)\right| \\ &= C_{1n} + C_{2n} + C_{3n} + C_{4n} = O_P(q^2/n^{1/2}). \end{aligned} \tag{A.32}$$

Here,

$$C_{1n} \leq \sum_{j=2}^{q}\sum_{i=1}^{j-1}\left(n^{-1}\sum_{t=1}^{n}\hat{\varepsilon}_t^4\right)^{1/2}\left(n^{-1}\sum_{t=1}^{n}(\hat{\varepsilon}_t^2-\varepsilon_t^2)^2\right)^{1/2}$$
$$= o_P(q^2/n^{1/2}) \quad \text{(A.33)}$$

given $n^{-1}\sum_{t=1}^{n}(\hat{\varepsilon}_t^2-\varepsilon_t^2)^2 = O_P(n^{-1})$. Similarly,

$$C_{2n} = O_P(q^2/n^{1/2}). \quad \text{(A.34)}$$

We also have

$$C_{3n} \leq q\left(n^{-1}\sum_{t=1}^{n}\varepsilon_t^2\right)\sum_{j=2}^{q}|\hat{\sigma}_j^2-\sigma_o^2|$$
$$\leq q^2\left(n^{-1}\sum_{t=1}^{n}\varepsilon_t^2\right)|\hat{\sigma}_n^2-\sigma_o^2|$$
$$+ q\left(n^{-1}\sum_{t=1}^{n}\varepsilon_t^2\right)\sum_{j=2}^{q}|\hat{\sigma}_n^2-\hat{\sigma}_j^2|$$
$$\leq q^2\left(n^{-1}\sum_{t=1}^{n}\varepsilon_t^2\right)|\hat{\sigma}_n^2-\sigma_o^2|$$
$$+ q^2\left(n^{-1}\sum_{t=1}^{n}\varepsilon_t^2\right)\left(n^{-1}\sum_{t=n-q+1}^{n}\hat{\varepsilon}_t^2\right)$$
$$= O_P(q^2/n^{1/2}). \quad \text{(A.35)}$$

Finally, by Markov's inequality and the fact that under H_o

$$E\left|n^{-1}\sum_{t=j+1}^{n}\varepsilon_{t-i}^2(\varepsilon_{t-j}^2-\sigma_o^2)\right|$$
$$\leq \left(E\left(n^{-1}\sum_{t=j+1}^{n}\varepsilon_{t-i}^2(\varepsilon_{t-j}^2-\sigma_o^2)\right)^2\right)^{1/2} = O(n^{-1/2}),$$

we have

$$C_{4n} = O_P(q^2/n^{1/2}). \quad \text{(A.36)}$$

Collecting (A.32)–(A.36) yields $B_{2n} = O_P(q^2/n^{1/2})$. This, together with (A.30)–(A.31) and $q^5/n \to 0$, implies $S_n = O_P(q^2/n^{1/2}) = o_P(1)$. It follows from (A.29) that nR^2 is of the same order of magnitude as $n\sum_{j=1}^{q}\hat{\rho}^2(j)$; that is, $nR^2 = O_P(q)$. Therefore, from (A.29) and $q^5/n \to 0$ again, we obtain $nR^2 - n\sum_{j=1}^{q}\hat{\rho}^2(j) = O_P(q^3/n^{1/2}) = o_P(q^{1/2})$. Consequently, $Q_{\mathrm{REG}}(q) - Q_{\mathrm{TRUN}}(q) = o_P(1)$ and $Q_{\mathrm{REG}}(q) \to^d$ N(0, 1) because $Q_{\mathrm{TRUN}}(q) \to^d$ N(0, 1) by Theorem 1 and the facts that $C_n(k) = q + o(q^{1/2})$ and $D_n(k) = 2q(1+o(1))$ as $q^5/n \to 0, q \to \infty$. This completes the proof.

[Received January 1996. Revised May 1998.]

REFERENCES

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

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

Bartlett, M. S. (1955), *An Introduction to Stochastic Processes*, Cambridge, U.K.: Cambridge University Press.

Beltrão, K., and Bloomfield, P. (1987), "Determining the Bandwidth of a Kernel Spectrum Estimate," *Journal of Time Series Analysis*, 8, 21–38.

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.

——— (1993), "ARCH Models: Properties, Estimation and Testing," *Journal of Economic Surveys*, 7, 305–366.

Bera, A. K., Ra, S., and Sarkar, 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: Allied Publishers, pp. 324–356.

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

Bollerslev, T., Chou, R. Y., and Kroner, K. F. (1992), "ARCH Modeling in Finance," *Journal of Econometrics*, 52, 5–59.

Bollerslev, T., 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, pp. 2959–3038.

Bollerslev, T., and Wooldridge, J. (1992), "Quasi-Maximum Likelihood Estimation and Inference in Dynamic Models With Time Varying Covariances," *Econometric Reviews*, 11, 143–172.

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

Brock, W., Dechert, D., and Scheinkman, J. (1987), "A Test for Independence Based on the Correlation Dimension," unpublished manuscript, University of Wisconsin, Dept. of Economics.

Diebold, F. X. (1987), "Testing for Serial Correlation in the Presence of ARCH," in *Proceedings of the Business and Economic Statistics Section, American Statistical Association*, pp. 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," unpublished manuscript, Northwestern University, Dept. of Economics.

Durbin, J. (1967), "Tests of Serial Independence Based on the Cumulated Periodogram," *Bulletin of the International Statistical Institute*, 42, 1040–1048.

Durlauf, S. (1991), "Spectral Based Testing of the Martingale Hypothesis," *Journal of Econometrics*, 50, 355–376.

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

——— (1983), "Estimates of the Variance of U.S. Inflation Based on the ARCH Model," *Journal of Money, Credit and Banking*, 15, 286–301.

——— (1984), "Wald, Likelihood Ratio, and Lagrange Multiplier Tests in Econometrics," in *Handbook of Econometrics* (vol. II), eds. Z. Griliches and M. Intriligator, Amsterdam: Elsevier Science, pp. 776–826.

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

Engle, R. F., and Kraft, D. (1983), "Multiperiod Forecast Error Variances of Inflation Estimated From ARCH Models," in *Applied Time Series Analysis of Economic Data*, ed. A. Zellner, Washington, DC: U.S. Department of Commerce, Bureau of the Census, pp. 293–302.

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

Granger, C. W. J., and Teräsvirta, 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," *Journal of Business & Economic Statistics*, 7, 107–115.

Grenander, U., and Rosenblatt, M. (1953), "Statistical Spectral Analysis

Arising From Stationary Stochastic Process," *The Annals of Mathematical Statistics*, 24, 537–558.
——— (1957), *Statistical Analysis of Stationary Time Series*, New York: Chelsea.
Hong, Y. (1996), "Consistent Testing for Serial Correlation of Unknown Form," *Econometrica*, 64, 837–864.
——— (1997), "One-sided ARCH Testing in Time Series Models," *Journal of Time Series Analysis*, 18, 253–277.
Hong, Y., and Shehadeh, R. (1996), "A New Test for ARCH Effects and its Finite Sample Performance," CAE Working Paper 96-01, Cornell University, Dept. of Economics.
Lee, J. H. H. (1991), "A Lagrange Multiplier Test for GARCH Models," *Economics Letters*, 37, 265–271.
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 & Economic Statistics*, 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, 297–303.
Luukkonen, R., Saikkonen, P., and Teräsvirta, T. (1988), "Testing Linearity in Univariate Time Series Models," *Scandinavian Journal of Statistics*, 15, 161–175.
McLeod, A. I., and Li, W. K. (1983), "Diagnostic Checking ARMA Time Series Models Using Squared Residual Autocorrelations," *Journal of Time Series Analysis*, 4, 269–273.
Milhøj, A. (1985), "The Moment Structure of ARCH Processes," *Scandinavian Journal of Statistics*, 12, 281–292.
Newey, W., and West, K. (1994), "Automatic Lag Selection in Covariance Matrix Estimation," *Review of Economic Studies*, 61, 631–653.
Parzen, E. (1957), "On Consistent Estimates of the Spectrum of a Stationary Time Series," *The Annals of Mathematical Statistics*, 28, 329–348.
Priestley, M. B. (1981), *Spectral Analysis and Time Series* (vol. 1), London: Academic Press.
Robinson, P. M. (1991a), "Testing for Strong Serial Correlation and Dynamic Conditional Heteroskedasticity in Multiple Regression," *Journal of Econometrics*, 47, 67–84.
——— (1991b), "Automatic Frequency Domain Inference on Semiparametric and Nonparametric Models," *Econometrica*, 59, 1329–1364.
Taylor, S. (1984), *Modelling Financial Time Series*, New York: Wiley.
Weiss, A. (1984), "ARMA Models With ARCH Errors," *Journal of Time Series Analysis*, 5, 129–143.