

THE VARIANCE RATIO STATISTIC AT LARGE HORIZONS

WILLA W. CHEN
Texas A&M University

ROHIT S. DEO
New York University

We make three contributions to using the variance ratio statistic at large horizons. Allowing for general heteroskedasticity in the data, we obtain the asymptotic distribution of the statistic when the horizon k is increasing with the sample size n but at a slower rate so that $k/n \rightarrow 0$. The test is shown to be consistent against a variety of relevant mean reverting alternatives when $k/n \rightarrow 0$. This is in contrast to the case when $k/n \rightarrow \delta > 0$, where the statistic has been recently shown to be inconsistent against such alternatives. Second, we provide and justify a simple power transformation of the statistic that yields almost perfectly normally distributed statistics in finite samples, solving the well-known right skewness problem. Third, we provide a more powerful way of pooling information from different horizons to test for mean reverting alternatives. Monte Carlo simulations illustrate the theoretical improvements provided.

1. INTRODUCTION

The variance ratio (VR) statistic is one of the popular tests that has been employed in the literature to test the random walk hypothesis for financial and economic data. The statistic is obtained as the sample variance of k -period differences, $x_t - x_{t-k}$, of the time series x_t , divided by k times the sample variance of the first difference, $x_t - x_{t-1}$, for some integer k . The VR statistic has been found by several authors (see, e.g., Faust, 1992) to be particularly powerful when testing against mean reverting alternatives to the random walk model, particularly when k is large. However, the practical use of the statistic has been impeded by the fact that the asymptotic theory provides a poor approximation to the small-sample distribution of the VR statistic. More specifically, rather than being normally distributed as the theory states, the statistics are severely biased and right skewed for large k (see Lo and MacKinlay, 1989), which makes application of the statistic problematic. To circumvent this problem, Richard-

The authors thank Bruce Hansen and the referees for useful suggestions and comments that greatly improved the paper. The first author's research was supported by NSF grant DMS-0306726. Address correspondence to Rohit S. Deo, Stern School of Business, New York University, 8-57 KMEC, 44 W. 4th Street, New York, NY 10012, USA; e-mail: rdeo@stern.nyu.edu.

son and Stock (1989) derived the asymptotic distribution of the *VR* statistic under the random walk null, assuming that both k and n increase to infinity but in such a way that k/n converges to a positive constant δ that is strictly less than 1. They showed that the *VR* statistic, without any normalization, converges to a functional of Brownian motion. Through Monte Carlo simulations, they demonstrated that this new distribution provides a far more robust approximation to the small-sample distribution of the *VR* statistic. However, Deo and Richardson (2003) have recently shown that the *VR* statistic is inconsistent against an important class of mean reverting alternatives under this framework. Thus, though the *VR* statistic would have vastly improved size properties under the null hypothesis of a random walk if k were chosen to be a fraction of the sample size n , it would fail to detect such alternatives with probability approaching 1 as the sample size increased. Currently there is no proposal in the literature that provides a way of using the *VR* statistic without compromising either its finite-sample size properties or its large-sample power properties.

With this backdrop, we provide several contributions to the literature. First, it is intuitively appealing to maintain the assumption that the multiperiod horizon k is large, not least because longer horizons have a better chance of capturing mean reversion in the series. Thus, under general conditions that allow for conditional heteroskedasticity in the innovations, we study the limiting behavior of the *VR* statistic for large k but now under the restriction that $k/n \rightarrow 0$. Specifically, we show that when $k \rightarrow \infty$, $n \rightarrow \infty$, but $k/n \rightarrow 0$, then under the null of a random walk, the *VR* statistic is asymptotically normal with a mean of 1. The requirement that k is large is important because as stated before, previous authors have shown that large values of k are to be preferred when testing for mean reversion. Furthermore, we prove that under this alternative distribution theory, the test is consistent, in that the probability of it detecting a wide variety of mean reversion alternatives approaches 1 as the sample size n increases.

Unfortunately, this new distribution does not solve the well-documented skewness problem of the *VR* statistic's sampling distribution. The second contribution of this paper is to propose a method that is shown to improve the asymptotic normal approximation to the distribution of the statistic by an order of magnitude in finite samples, via a simple power transformation of the *VR* statistic. Monte Carlo simulations confirm the theoretical assertion of the vast improvement of the normal approximation afforded by the power transformation. Our Monte Carlo simulations also show that this improvement in the normal approximation leads to significant gains in power against mean reverting alternatives. Our simulations also show that the performance of the test based on using the Richardson and Stock asymptotics is sensitive to both sample size and conditional heteroskedasticity. Furthermore, using the Richardson and Stock asymptotics also results in uniformly lower power when compared to the new asymptotic approach that we present. Thus, our new approach uniformly dominates the Richardson and Stock asymptotic approach.

The third contribution of this paper is to implement a new joint test that uses *VR* statistics computed at different differencing periods to test the random walk null hypothesis. The joint test statistic that has been studied so far in the literature is the Wald type chi-square test statistic that jointly tests whether a sequence of population variance ratios at several differencing periods all equal 1. However, this test is blind to the inherent one-sided nature of a mean reverting alternative hypothesis, because under such an alternative all the population variance ratios should be less than 1. See Lo and MacKinlay (1989). In this paper, we adapt a test procedure proposed by Follmann (1996) for testing against one-sided alternatives for the mean vector of a multivariate normal distribution. Our Monte Carlo simulations show that this adapted test in combination with the power transformation results in significant power gains over the usual chi-square test when testing for mean reverting alternatives, while retaining the appropriate size.

The paper is organized as follows. In Section 2, we define the *VR* statistic and provide its asymptotic distribution under conditional heteroskedasticity for large k such that $k^{-1} + k/n \rightarrow 0$. We also demonstrate in that section that in this framework the *VR* statistic is consistent against a wide range of alternatives. In Section 3, we provide an alternative equivalent representation of the *VR* statistic that motivates the power transformation that provides a better approximation to the normal distribution. A new joint test that combines information from several differencing periods and is useful against one-sided alternatives is also introduced. Section 4 presents Monte Carlo results for the various statistics that we have proposed under two different null hypotheses and three alternative hypotheses. All technical proofs are relegated to the Appendix.

2. ASYMPTOTIC THEORY FOR THE VARIANCE RATIO STATISTIC

Given $n + 1$ observations x_0, x_1, \dots, x_n of a time series, the *VR* statistic with a positive integer $k (< n)$ as differencing period is defined as

$$VR(k) = \hat{\sigma}_b^2(k) / \hat{\sigma}_a^2, \tag{1}$$

where

$$\hat{\sigma}_b^2(k) = \frac{n}{k(n - k + 1)(n - k)} \sum_{t=k}^n (x_t - x_{t-k} - k\hat{\mu})^2,$$

$$\hat{\sigma}_a^2 = \frac{1}{n - 1} \sum_{t=1}^n (x_t - x_{t-1} - \hat{\mu})^2,$$

and

$$\hat{\mu} = n^{-1} \sum_{t=1}^n (x_t - x_{t-1}).$$

In the usual fixed k asymptotic treatment, under the null hypothesis that the $\{x_t\}$ follow a random walk with possible drift, given by

$$x_t = \mu + x_{t-1} + \varepsilon_t, \tag{2}$$

where μ is a real number and $\{\varepsilon_t\}$ is a sequence of zero mean independent random variables, it is possible to show (see, e.g., Lo and MacKinlay, 1988) that

$$\sqrt{n}(VR(k) - 1) \xrightarrow{D} N(0, \sigma_k^2),$$

where σ_k^2 is some simple function of k . This result extends to the case where the $\{\varepsilon_t\}$ are a martingale difference series with conditional heteroskedasticity (see, e.g., Campbell, Lo, and MacKinlay, 1997), though the variance σ_k^2 has to be adjusted to account for the conditional heteroskedasticity. However, the asymptotic behavior of the VR statistic for large values of k , such that $k^{-1} + k/n \rightarrow 0$, is not known when the innovations ε_t are conditionally heteroskedastic. In this section, we provide precisely this asymptotic distribution, in obtaining which the following assumptions on the series of innovations $\{\varepsilon_t\}$ are made.

(A1) $\{\varepsilon_t\}$ is ergodic and $E(\varepsilon_t | \mathcal{F}_{t-1}) = 0$ for all t , where \mathcal{F}_t is a sigma field, ε_t is \mathcal{F}_t measurable, and $\mathcal{F}_{t-1} \subset \mathcal{F}_t$ for all t .

(A2) $E(\varepsilon_t^2) = \sigma^2 < \infty$.

(A3) For any integer q , $2 \leq q \leq 8$, and for q nonnegative integers s_i , $E(\prod_{i=1}^q \varepsilon_{t_i}^{s_i}) = 0$ when at least one s_i is exactly one and $\sum_{i=1}^q s_i \leq 8$.

(A4) For any integer r , $2 \leq r \leq 4$, and for r nonnegative integers s_i , $E(\prod_{i=1}^r \varepsilon_{t_i}^{s_i} | \mathcal{F}_{t_i}) = 0$ when at least one s_i is exactly one and $\sum_{i=1}^r s_i \leq 4$, for all $t < t_i$, $i = 1, 2, 3, 4$.

(A5) $\lim_{n \rightarrow \infty} \text{Var}[E(\varepsilon_{t+n}^2 \varepsilon_{t+n+j}^2 | \mathcal{F}_t)] = 0$ uniformly in j for every $j > 0$.

(A6) $\lim_{n \rightarrow \infty} E(\varepsilon_t^2 \varepsilon_{t-n}^2) = \sigma^4$.

Conditions (A1)–(A6) allow the innovations ε_t to be a martingale difference sequence with conditional heteroskedasticity. As a matter of fact, Lemmas 1 and 2, which follow, show that the stochastic volatility model (see Shephard, 1996) and the GARCH model (Bollerslev, 1986), which are two of the most popular models in the literature for conditional heteroskedastic martingale differences, satisfy conditions (A1)–(A6). Conditions (A3)–(A4) state that the series $\{\varepsilon_t\}$ shows product moment behavior similar to that of an independent white noise process. Conditions (A5)–(A6) state that ε_t and ε_{t-n} are roughly independent for large lags n .

The following two lemmas assert that two major models of conditionally heteroskedastic martingale differences, namely, the stochastic volatility model

and the generalized autoregressive conditionally heteroskedastic (GARCH) model, satisfy the assumptions (A1)–(A6). The proofs of the lemmas are in the Appendix.

LEMMA 1. *Let the series $\{\varepsilon_t\}$ be generated by the stochastic volatility model*

$$\varepsilon_t = v_t \exp(h_t), \tag{3}$$

where $\{v_t\}$ is an independent $(0, \sigma_v^2)$ stationary series, $\{h_t\}$ is a stationary zero mean Gaussian series, and $\{v_t\}$ and $\{h_t\}$ are independent. Assume that $E(v_t^8) < \infty$. Then $\{\varepsilon_t\}$ satisfies the assumptions (A1)–(A6).

See Shephard (1996) for a discussion of model (3) and its applications.

Our next lemma asserts that under some conditions the GARCH(1,1) family of models also satisfies conditions (A1)–(A6). We have restricted attention to the GARCH(1,1) case for simplicity of exposition. We conjecture that conditions (A1)–(A6) will continue to hold for a general GARCH(p, q) model, the proof following along similar lines by referring to the work of Bougerol and Picard (1992).

LEMMA 2. *Let the series $\{\varepsilon_t\}$ be a GARCH(1,1) process given by*

$$\varepsilon_t = \sigma_t v_t, \tag{4}$$

where $\sigma_t^2 = \omega + \beta\sigma_{t-1}^2 + \alpha\varepsilon_{t-1}^2$ and $\{v_t\}$ is a sequence of independent standard normal variables. Let $\omega > 0$, $\beta \geq 0$, and $\alpha > 0$. Furthermore, let α and β be such that $E\{\log_e(\beta + \alpha v_t^2)\} < 0$ and $E\{(\beta + \alpha v_t^2)^4\} < 1$. Then $\{\varepsilon_t\}$ satisfies the assumptions (A1)–(A6).

The condition $E\{\log_e(\beta + \alpha v_t^2)\} < 0$ in Lemma 2 is satisfied by any pair (α, β) in the set $S = \{(\alpha, \beta) : \alpha + \beta < 1\}$ (see Nelson, 1990), whereas the condition $E\{(\beta + \alpha v_t^2)^4\} < 1$ will be satisfied by a nonempty subset of S (see Bollerslev, 1986).

We now state our result on the limiting distribution of the VR statistic in the following theorem.

THEOREM 3. *Let the series $\{x_t\}$ satisfy equation (2) and assume that conditions (A1)–(A6) hold. For a fixed positive integer s , let $k_1 < k_2 < \dots < k_s < n$ be positive integers such that $k_1 \rightarrow \infty$, $k_s n^{-1} \rightarrow 0$, and $k_i k_j^{-1} \rightarrow a_{ij}$ for $1 \leq i \leq j \leq s$. Let \mathbf{D}_n be an $s \times s$ diagonal matrix with diagonal elements $d_{ii} = \sqrt{n/k_i}$ for $i = 1, 2, \dots, s$. Then*

$$\mathbf{V}_n \stackrel{D}{\approx} N(\mathbf{1}, \mathbf{D}_n^{-1} \mathbf{\Sigma} \mathbf{D}_n^{-1}),$$

where $\mathbf{V}_n = (VR(k_1), VR(k_2), \dots, VR(k_s))'$, $\mathbf{1}$ is an $s \times 1$ vector of ones and $\mathbf{\Sigma} = (\sigma_{ij})$ is an $s \times s$ matrix such that $\sigma_{ij} = 4a_{ij}^{1/2}(3 - a_{ij})/6$.

Note that the limiting distribution of the VR statistic is free of nuisance parameters and is identical to that obtained when the ε_t are assumed to be independent. See Theorem 9.4.1 of Anderson (1994). Furthermore, the VR statistics computed at different differencing periods k_i are asymptotically independent when $k_i k_j^{-1} \rightarrow 0$ for $i < j$. Both of these results are in contrast to those obtained when the differencing periods are fixed and not allowed to increase to infinity with the sample size. See Lo and MacKinlay (1989). It is interesting to note that the limiting distribution of the VR statistic is free of nuisance parameters depending on higher moments that might arise as a result of conditional heteroskedasticity. This is quite different from the behavior of other tests of the random walk hypothesis in the presence of conditional heteroskedasticity. See Deo (2000).

We have established the asymptotic distribution of the VR statistic under the null hypothesis of a random walk with conditional heteroskedasticity when $k \rightarrow \infty$, $n \rightarrow \infty$, and $k/n \rightarrow 0$. The next theorem states that under this framework, the VR statistic also provides a consistent test against a large class of mean reverting alternatives.

THEOREM 4. *Let $\{e_t\}$ and $\{u_t\}$ be two series of zero mean independent processes with finite fourth moments and which are independent of each other. Define the processes $\{y_t\}$ and $\{z_t\}$ by $y_t = \sum_{j=0}^{\infty} a_j u_{t-j}$ and $z_t = \sum_{j=0}^{\infty} b_j e_{t-j}$, where $|a_j| \leq C\lambda^j$ and $|b_j| \leq C\lambda^j$ for some constant C and $0 < \lambda < 1$. Let $r_t = \mu + r_{t-1} + z_t$ and $x_t = r_t + y_t$. If $k \rightarrow \infty$, $n \rightarrow \infty$, and $k/n \rightarrow 0$, then*

$$VR(k) \xrightarrow{P} \frac{\sigma_z^2 + 2 \sum_{j=1}^{\infty} \gamma_z(j)}{\sigma_z^2 + 2\sigma_y^2 - 2\gamma_y(1)},$$

where σ_z^2 and σ_y^2 are the variances of z_t and y_t , respectively, whereas $\gamma_z(j)$ and $\gamma_y(j)$ are the respective autocovariances at lag j .

Theorem 4 shows that the power properties of the VR statistic under the $k/n \rightarrow 0$ framework are markedly different from those when $k/n \rightarrow \delta > 0$, in which case Deo and Richardson (2003) have shown the VR statistic to be inconsistent against the alternatives considered in Theorem 4.

Though the VR statistic has an asymptotic normal distribution when $k/n \rightarrow 0$, it is obvious that in finite samples the normal distribution may not provide a good approximation because the statistic is a quadratic form and hence must be right skewed. A common method that has a long history in statistics to reduce skewness and induce normality in such random variables is to consider power transformations. The obvious question, naturally, is which power one should use, and we address this question for the VR statistic in the next section.

3. POWER TRANSFORMATIONS OF THE VARIANCE RATIO STATISTIC

In attempting to address the skewness of the finite-sample distribution of the *VR* statistic, it helps to express the *VR* statistic in an alternative form, which lends more insight into how the normal distribution approximation can be improved. Inspection of the proof of Theorem 3 in the Appendix shows that

$$VR(k) = \hat{\sigma}^{-2} \sum_{|j| \leq k} (1 - |j|/k) \hat{\gamma}_j + o_p(\sqrt{k/n}), \tag{5}$$

where $\hat{\gamma}_j = \hat{\gamma}_{-j} = n^{-1} \sum_{t=j+1}^n \varepsilon_t \varepsilon_{t-j}$ for $j \geq 0$ and

$$\hat{\sigma}^2 = (n - 1)^{-1} \sum_{t=1}^n (\varepsilon_t - \bar{\varepsilon})^2 = (n - 1)^{-1} \sum_{t=1}^n (x_t - x_{t-1} - \hat{\mu})^2.$$

Now, using the fact that

$$\hat{\gamma}_j = \int_0^{2\pi} I(\lambda) \exp(-ij\lambda) d\lambda,$$

where $I(\lambda) = (2\pi n)^{-1} |\sum_{t=1}^n \varepsilon_t \exp(-it\lambda)|^2$ is the periodogram, we get from (5)

$$VR(k) = \hat{\sigma}^{-2} \int_0^{2\pi} W_k(\lambda) I(\lambda) d\lambda + o_p(\sqrt{k/n}), \tag{6}$$

where

$$W_k(\lambda) = \sum_{|j| \leq k} (1 - |j|/k) \exp(-ij\lambda) = k^{-1} \left\{ \frac{\sin(k\lambda/2)}{\sin(\lambda/2)} \right\}^2.$$

As shown in part (i) of Lemma 7 in the Appendix, the integral in (6) can be approximated by a discrete sum over the Fourier frequencies $\lambda_j = 2\pi j/n$ with error $o_p(\sqrt{k/n})$, and hence we get

$$VR(k) = \frac{4\pi}{n\hat{\sigma}^2} \sum_{j=1}^{[(n-1)/2]} W_k(\lambda_j) I(\lambda_j) + o_p(\sqrt{k/n}). \tag{7}$$

The behavior of $VR(k)$ is thus dictated by the behavior of the periodogram values $I(\lambda_j)$ at the Fourier frequencies. If the ε_t series is Gaussian, then it is well known (Brockwell and Davis, 1996) that the variables $2\pi I(\lambda_j)/\sigma^2$ are exactly independent and identically distributed (i.i.d.) standard exponential random variables for all sample sizes. This behavior of the variables $2\pi I(\lambda_j)/\sigma^2$ can be shown to continue to hold asymptotically if the ε_t are a martingale difference sequence with finite fourth moment, by applying the central limit theorem for martingale differences to $n^{-1/2} \sum_{t=1}^n \varepsilon_t \exp(-i\lambda_j t)$. These observations in con-

junction with (7) and the fact that $\hat{\sigma}^2/\sigma^2 = 1 + O_p(n^{-1/2})$ imply that, in general, we may think of the VR statistic as being of the form

$$VR(k) = \frac{2}{n} \sum_{j=1}^{[(n-1)/2]} W_k(\lambda_j) V_j + o_p(\sqrt{k/n}), \tag{8}$$

where the V_j are independent standard exponential random variables. As we next show, this approximate expression for the VR statistic as a weighted linear combination of independent standard exponential random variables helps us both to understand why the normal distribution provides a bad approximation for large k and also to obtain an appropriate power transformation that improves the normal approximation.

It is known (see, e.g., Anderson, 1994, p. 509) that $W_k(\lambda)$ has a peak at the origin and then damps down to zero for values of λ further from the origin. Furthermore, the larger k is, the more quickly $W_k(\lambda)$ damps down to zero, which can be seen in Figure 1, where we plot $W_k(\lambda)$ for $n = 128$ and $k = 8$ and 16. Thus, for large values of k , we see from (8) that $VR(k)$ will essentially be a sum of too few independent standard exponential random variables for the central limit theorem to properly take effect, resulting in right skewed distributions. However, Chen and Deo (2004) have recently shown that power transformations may be gainfully applied to random variables that have approximate linear representations of the form in (8), yielding much better normal approximations. Using their results (see Chen and Deo, 2004, eqn. (9)), it follows that if one sets

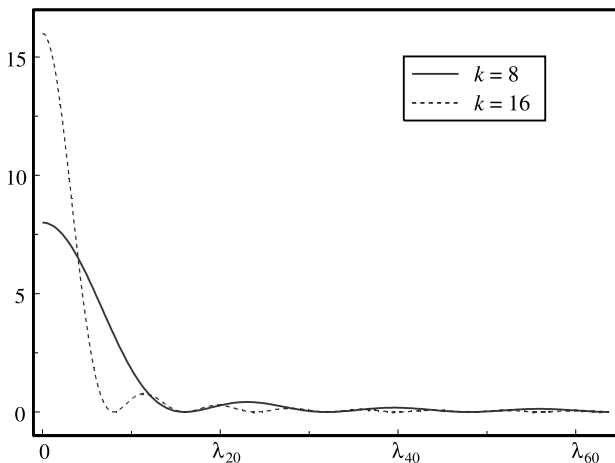


FIGURE 1. $W_k(\lambda)$ for $n = 128$ and $k = 8$ and 16.

$$\beta = 1 - \frac{2}{3} \frac{\left(\sum_{j=1}^{[(n-1)/2]} W_k(\lambda_j) \right) \left(\sum_{j=1}^{[(n-1)/2]} W_k^3(\lambda_j) \right)}{\left(\sum_{j=1}^{[(n-1)/2]} W_k^2(\lambda_j) \right)^2}, \tag{9}$$

then the Gaussian distribution provides a better approximation to the distribution of $VR^\beta(k)$ than to that of $VR(k)$. Indeed, from the results of Chen and Deo (2004), the Gaussian distribution approximation to the distribution of $VR^\beta(k)$ is an entire order of magnitude better than the Gaussian approximation to the distribution of $VR(k)$. A dramatic visual display of this improvement is shown in Figure 2. The plot on the left is a QQ plot of 20,000 replications of the $VR(k)$ statistic, based on a sample size of $n = 128$ and $k = 16$ where the ε_t are i.i.d. standard normal. The extreme curvature is indicative of the right skewness of the distribution of $VR(k)$. The plot on the right is a QQ plot of $VR^\beta(k)$, where β was computed using (9). The plot now shows a straight line as would be expected for observations from a normal distribution. The power transformation thus provides a very simple method of getting almost near perfect normality for the finite-sample distribution of the VR statistic. A standard Taylor series argument applied to the result of Theorem 3 yields the asymptotic distribution of $VR^\beta(k)$, which can then be used for inference. However, we feel that because the power transformation is motivated by the representation (7), it might be preferable to redefine the VR statistic and also its power transformation directly in terms of the leading term of that expression, thus avoiding any effects of the remainder term on its finite-sample distribution. Toward that end, we

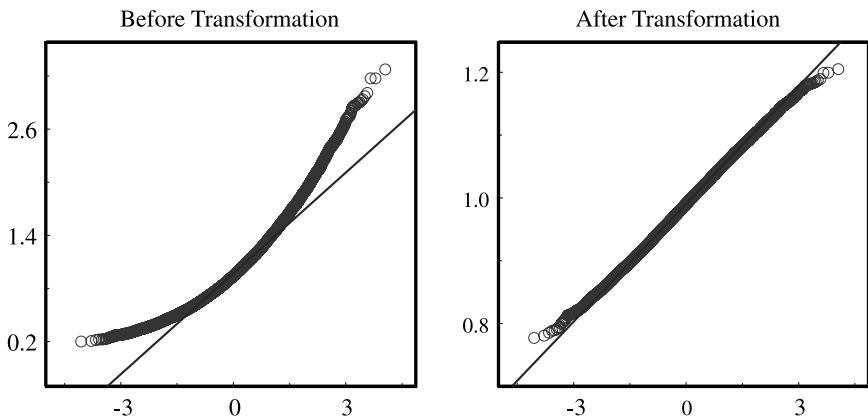


FIGURE 2. QQ plots of $VR(k)$ and $VR^\beta(k)$ on 20,000 replications with $n = 128$, $k = 16$ and $\varepsilon_t \sim N(0,1)$.

now define the *VR* statistic based on the periodogram, for differencing period k , as

$$VR_p(k) = \frac{1}{(1 - k/n)} \frac{4\pi}{n\hat{\sigma}^2} \sum_{j=1}^{[(n-1)/2]} W_k(\lambda_j) I_{\Delta X}(\lambda_j), \tag{10}$$

where $I_{\Delta X}(\lambda_j) = (2\pi n)^{-1} |\sum_{t=1}^n (x_t - x_{t-1} - \hat{\mu}) \exp(-i\lambda_j t)|^2$. Because the periodogram is shift invariant at nonzero Fourier frequencies, we have $I_{\Delta X}(\lambda_j) = I(\lambda_j)$, and hence the $VR_p(k)$ statistic as defined in (10) based on the observed data $x_t - x_{t-1} - \hat{\mu}$ is identical to the first term in (7), which is based on the unobserved ε_t . It should be noted that this expression for the *VR* statistic, apart from the normalization of $(1 - k/n)^{-1}$, which is just a finite-sample correction ensuring a unit mean, is precisely the normalized discrete periodogram average estimate of the spectral density of a stationary process at the origin and has a long tradition in time series analysis. See Brockwell and Davis (1996). From (7) it follows that $VR_p(k)$ will have the same asymptotic distribution as that of $VR(k)$ given in Theorem 3 and hence, by the usual Taylor series argument, the asymptotic distribution of $VR_p^\beta(k)$ may be obtained. It is however preferable to have an expression for the variance of $VR_p(k)$, and thus for that of $VR_p^\beta(k)$, that is accurate in finite samples and accounts for the finite-sample effects of conditional heteroskedasticity. Toward this end, we first define the quantities $C_{n,k} = n(n - k)^{-1}$ and

$$\hat{\tau}_j = \hat{\sigma}^{-4} (n - j - 4)^{-1} \sum_{t=j+1}^n (x_t - x_{t-1} - \hat{\mu})^2 (x_{t-j} - x_{t-j-1} - \hat{\mu})^2,$$

where $\hat{\tau}_j$ is an estimator of $\sigma^{-4} E(\varepsilon_t^2 \varepsilon_{t-j}^2)$. In part (ii) of Lemma 7, we show that the finite-sample variance covariance matrix of $\mathbf{V}_p = (VR_p(k_1), VR_p(k_2), \dots, VR_p(k_s))'$ with remainder terms of order $o(k_s^2/n^2)$ is consistently estimated by

$$\hat{\Sigma} = \mathbf{L}' \begin{bmatrix} \hat{\mathbf{A}} & \hat{\mathbf{b}} \\ \hat{\mathbf{b}}' & \hat{d} \end{bmatrix} \mathbf{L}, \tag{11}$$

where $\mathbf{L} = (l_{k_1}, \dots, l_{k_s})$,

$$l'_{k_i} = \left(2C_{n,k_i}(1 - 1/k_i), \dots, 2C_{n,k_i}(1 - (k_i - 1)/k_i), \underbrace{0, \dots, 0}_{k_s - k_i \text{ terms}} - (k_i C_{n,k_i} - n/(n - 1)) \right), \tag{12}$$

$$\hat{\mathbf{A}} = \text{diag} \left(\frac{n-j}{n^2} \hat{\tau}_j + \frac{j}{n^2} \right) \quad j = 1, \dots, k_s,$$

$\hat{\mathbf{b}}$ is a $k_s \times 1$ vector such that its j th element is given by $(2(n - j)n^{-3}\hat{\tau}_j + 2jn^{-3})$, and $\hat{d} = 2n^{-2}$. We are now in a position to state the following theorem.

THEOREM 5. *Let the series $\{x_i\}$ satisfy equation (2) and assume that conditions (A1)–(A6) hold. For a fixed positive integer s , let $k_1 < k_2 < \dots < k_s < n$ be positive integers such that $k_1 \rightarrow \infty$, $k_s n^{-1} \rightarrow 0$, and $k_i k_j^{-1} \rightarrow a_{ij}$ for $1 \leq i \leq j \leq s$. For each k_i , let β_i be given by (9) and define $\mathbf{V}_{p,\beta} = (VR_p^{\beta_1}(k_1), VR_p^{\beta_2}(k_2), \dots, VR_p^{\beta_s}(k_s))'$. Then*

$$\mathbf{V}_{p,\beta} \stackrel{D}{\approx} N(\boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta),$$

where the (i, j) th element of $\boldsymbol{\Sigma}_\beta$ is

$$\beta_i \beta_j \hat{\sigma}_{i,j}$$

and the i th element of $\boldsymbol{\mu}_\beta$ is

$$1 + 0.5\beta_i(\beta_i - 1)\hat{\sigma}_{i,i},$$

where $\hat{\sigma}_{i,j}$ is the (i, j) th entry of $\hat{\boldsymbol{\Sigma}}$ given in (11).

It is trivially seen that both $VR_p \xrightarrow{P} 1$ and $VR_p^\beta \xrightarrow{P} 1$ under conditions (A1)–(A6). Our next theorem shows that both VR_p and VR_p^β also retain the consistency of the VR statistic with regard to detecting the alternative hypotheses assumed in Theorem 4.

THEOREM 6. *Let the assumptions of Theorem 4 hold. Then*

$$VR_p(k) \xrightarrow{P} \frac{\sigma_z^2 + 2 \sum_{j=1}^{\infty} \gamma_z(j)}{\sigma_z^2 + 2\sigma_y^2 - 2\gamma_y(1)}$$

and

$$VR_p^\beta(k) \xrightarrow{P} \left(\frac{\sigma_z^2 + 2 \sum_{j=1}^{\infty} \gamma_z(j)}{\sigma_z^2 + 2\sigma_y^2 - 2\gamma_y(1)} \right)^\beta,$$

where σ_z^2 and σ_y^2 are the variances of z_t and y_t , respectively, whereas $\gamma_z(j)$ and $\gamma_y(j)$ are the respective autocovariances at lag j .

We have, so far, obtained the joint distribution of the VR_p statistic computed at various differencing periods. These VR statistics can be combined into a single statistic by computing the quadratic form

$$Q_n = (\mathbf{V}_p - \mathbf{E}(\mathbf{V}_p))' \mathbf{Var}(\mathbf{V}_p)^{-1} (\mathbf{V}_p - \mathbf{E}(\mathbf{V}_p)), \tag{13}$$

where $\mathbf{V}_p = (VR_p(k_1), \dots, VR_p(k_s))'$. Because of the asymptotic normality of \mathbf{V}_p , this quadratic form will have an asymptotic chi-squared distribution with s degrees of freedom under the null hypothesis of a random walk. The test statistic Q_n can then be used to test whether the sequence of population variance ratios all equal one for $i = 1, 2, \dots, s$. Because the quadratic form Q_n is always positive, rejection of the null hypothesis of a random walk occurs only in the upper tail of the distribution of Q_n . However, under the important alternative of mean reverting processes of the kind imposed in finance applications, the population variance ratios, given by $VRP(k) \equiv \text{Var}(\sum_{i=1}^k \varepsilon_i) / (k \text{Var}(\varepsilon_1))$, are generally expected to be less than 1 for large k . For example, it can be easily shown that for the alternative models that are the sum of permanent and transitory components (see Poterba and Summers, 1988; Fama and French, 1988), $VRP(k)$ is less than 1 for all values of k . Hence, under such mean reverting processes, the alternative hypothesis actually has the one-sided form $H_a: VRP(k) < 1$ for $i = 1, \dots, s$. In such circumstances, ignoring the one-sided nature of the alternative can lead to a loss of power of the test. However, Follmann (1996) has proposed a test for the null hypothesis that the mean vector of a multivariate normal random variable is zero, which has good power for alternatives where all the elements of the mean vector are negative. Thus, Follmann's procedure would be directly applicable in the setting where the alternative of interest is a mean reverting process. We now adapt Follmann's procedure to test for mean reverting alternatives using VR_p statistics as follows. In testing the null hypothesis of a random walk

$$H_0: VRP(k_1) = \dots = VRP(k_s) = 1 \quad i = 1, 2, \dots, s$$

versus the one-sided alternative

$$H_a: VRP(k_1) < 1, \dots, VRP(k_s) < 1 \quad i = 1, 2, \dots, s$$

at the α level of significance, reject the null hypothesis if

$$\sum_{i=1}^s [VR_p(k_i) - 1] < 0 \quad \text{and} \quad Q_n > \chi_{s, 2\alpha}^2 \tag{14}$$

where $\chi_{s, 2\alpha}^2$ is the upper 2α critical value of a chi-square distribution with s degrees of freedom. From the asymptotic normality of VR_p and Theorem 2.1 of Follmann (1996), it follows that the procedure given previously has an asymptotic level of significance equal to α . An analogous procedure can be developed using the power transformation as follows. Reject the null hypothesis if

$$\sum_{i=1}^s [VR_p^{\beta_i}(k_i) - 1] < 0 \quad \text{and} \quad QP_n > \chi_{s, 2\alpha}^2 \tag{15}$$

where

$$QP_n = (\mathbf{V}_{p,\beta} - \boldsymbol{\mu}_\beta)' \boldsymbol{\Sigma}_\beta^{-1} (\mathbf{V}_{p,\beta} - \boldsymbol{\mu}_\beta) \quad (16)$$

and $\boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta$ are as in Theorem 5. The test procedure based on the power transformation would be expected to have better size and power properties compared to the one based on the original VR_p statistics because the quadratic form QP_n should be expected to have a distribution closer to the expected chi-square distribution. In the next section, we report the results from a Monte Carlo study, which evaluates the effectiveness of the new proposals we have made.

4. SIMULATION RESULTS

We carried out Monte Carlo simulations to evaluate the finite-sample performance of tests based on our modified VR statistic. The size properties under the null hypothesis were evaluated using the following two models: (i) $x_t = x_{t-1} + \varepsilon_t$, where $\varepsilon_t \sim i.i.d. N(0,1)$, (ii) $x_t = x_{t-1} + \varepsilon_t$, where $\varepsilon_t = \sigma_t v_t$, $v_t \sim i.i.d. N(0,1)$, and $\sigma_t^2 = 0.0001 + 0.8575\sigma_{t-1}^2 + 0.1171\varepsilon_{t-1}^2$. The parameter values for the GARCH(1,1) model in (ii) were chosen to reflect values obtained when fitting such models to real data. The sample sizes we considered were $n = 128$ and 512 , and the number of replications was $20,000$. For $n = 128$, we used $k_1 = 8$ and $k_2 = 16$, whereas for $n = 512$ we used $k_1 = 16$ and $k_2 = 32$. The first part of Table 1 reports the Monte Carlo sizes of the test statistics under the Gaussian white noise case, whereas the second part of the table is for the GARCH(1,1) model. In both parts the nominal level of significance is 5% and the test is two tailed. The sizes are reported for the statistics VR_p and VR_p^β for each combination of sample size and k , where β was computed for each case using (9). The sizes are reported for both the left and right tail to demonstrate the skewness and the effect of the power transformation on it. We also report the sizes of the quadratic tests (13), denoted in the table by Q_n , based upon both the untransformed and transformed VR statistics. Sizes for the modified intersection tests given in (14) and (15), denoted in the table by IQ_n , are also shown.

It is also of interest to study the finite-sample performance of the VR statistic under the $k/n \rightarrow \delta > 0$ asymptotics as proposed by Richardson and Stock (1989). We therefore also present empirical sizes and power of the VR statistics for our configuration of (k, n) values based on asymptotic critical values of the Richardson–Stock distribution that were computed as follows. For each combination of (k, n) , we generated $20,000$ replications of the VR statistic based on Gaussian noise with $n_\infty = 12,000$ and $k_\infty = (k/n)n_\infty = (k/n)12,000$ and the percentiles of these $20,000$ values were used to obtain the asymptotic critical values. The empirical sizes and powers based on these critical values are presented in Tables 1–3 in the row labeled RS.

TABLE 1. Sizes in percentage of the null of random walk: $x_t = \mu + x_{t-1} + \varepsilon_t$

<i>n</i>	128				512			
	8		16		16		32	
<i>k</i>	Lower	Upper	Lower	Upper	Lower	Upper	Lower	Upper
Gaussian white noise errors, $\varepsilon_t \sim N(0,1)$								
VR_p	0.6	3.8	0.1	4.5	1.0	3.6	0.5	4.2
VR_p^β	2.4	2.3	2.4	2.4	2.4	2.4	2.4	2.5
RS	1.3	1.5	1.7	1.8	1.8	2.0	2.0	2.4
	Q_n	Q_n^β	IQ_n	IQ_n^β	Q_n	Q_n^β	IQ_n	IQ_n^β
	5.8	5.1	1.5	5.7	5.2	4.8	2.2	5.4
GARCH(1,1) white noise errors, $\varepsilon_t \sim N(0, \sigma_t^2)$, $\sigma_t^2 = 0.0001 + 0.8575\sigma_{t-1}^2 + 0.1171\varepsilon_{t-1}^2$								
VR_p	0.3	4.1	0.0	4.6	0.6	3.9	0.2	4.4
VR_p^β	2.0	2.4	2.0	2.6	2.0	2.5	1.7	2.4
RS	2.2	2.5	2.6	2.6	4.4	4.4	3.9	4.1
	Q_n	Q_n^β	IQ_n	IQ_n^β	Q_n	Q_n^β	IQ_n	IQ_n^β
	6.2	4.6	1.3	4.9	5.7	4.3	1.5	4.5

It is immediately apparent from Table 1 that whereas the distribution of VR_p is very right skewed, as is well known, the power transformation is able to correct it and provide near perfect normality with sizes in each tail that are very close to nominal. One can also see that the power transformed statistic VR_p^β is able to retain the size close to the nominal even in the presence of GARCH innovations. On the other hand, it is seen from these tables that the finite-sample performance of the VR statistic when compared to the critical values of the Richardson–Stock distribution is not as good. In the case of Gaussian noise, the test is undersized, particularly for $n = 128$, whereas in the case of GARCH innovations, the test is oversized for $n = 512$. Note that for our configuration of (k, n) values, the ratio k/n takes values 0.03125, 0.0625, and 0.125. This clearly indicates that the $k/n \rightarrow \delta > 0$ asymptotic distribution cannot approximate the finite-sample distribution of the VR statistic when k/n is small and is sensitive to the presence of conditional heteroskedasticity.

Table 1 also demonstrates that the quadratic and the modified intersection tests based on the transformed VR statistics have much better size properties than those using their untransformed counterparts.

To evaluate the power properties of our tests, we generated data from the mean reverting process given by $x_t = r_t + y_t$, where $r_t = r_{t-1} + w_t$, $y_t = 0.9y_{t-1} +$

TABLE 2. Power in percentage against the alternative of random walk + AR(1)

<i>n</i>	128				512			
<i>k</i>	8	16			16	32		
$X_t = r_t + y_t, r_t = r_t - 1 + w_t, w_t \sim N(0,0.1)$								
VR_p	4.0	0.5			70.0	74.9		
VR_p^β	12.2	15.3			82.6	94.1		
RS	7.5	12.0			78.9	92.8		
	Q_n	Q_n^β	IQ_n	IQ_n^β	Q_n	Q_n^β	IQ_n	IQ_n^β
	2.2	13.4	4.7	22.2	39.1	86.5	70.9	93.8
$X_t = r_t + y_t, r_t = r_t - 1 + w_t, w_t \sim N(0,0.25)$								
VR_p	3.4	0.4			51.5	47.3		
VR_p^β	10.1	11.6			67.1	78.2		
RS	6.1	9.0			62.1	74.9		
	Q_n	Q_n^β	IQ_n	IQ_n^β	Q_n	Q_n^β	IQ_n	IQ_n^β
	2.2	10.8	4.0	18.0	21.7	65.5	47.6	79.3
$X_t = r_t + y_t, r_t = r_t - 1 + w_t, w_t \sim N(0,0.5)$								
VR_p	2.8	0.5			30.9	22.4		
VR_p^β	8.0	8.5			45.9	51.4		
RS	4.7	6.5			40.7	46.5		
	Q_n	Q_n^β	IQ_n	IQ_n^β	Q_n	Q_n^β	IQ_n	IQ_n^β
	2.3	8.7	3.4	14.4	10.5	40.4	26.7	56.0

$y_t = 0.9y_{t-1} + u_t, u_t \sim N(0,1).$

u_t , and $u_t \sim i.i.d. N(0,1)$ and also independent of $\{w_t\}$. The errors w_t were assumed to be $i.i.d. N(0, \sigma_w^2)$ where $\sigma_w^2 = 0.1, 0.25, \text{ and } 0.5$. This model with similar parameter configurations was considered in Lo and MacKinlay (1989) and Richardson and Smith (1991). Table 2 reports the Monte Carlo power values at 5% level of significance for this alternative model for the three different values of σ_w^2 . As the value of σ_w^2 increases, the permanent component dominates the process and the power of all tests decreases, as is to be expected. However, similar behavior of the tests is seen across the table. It is clear that the individual tests based on the transformed VR statistics provide power that is significantly superior to that of the untransformed ones, in some cases increasing the power by as much as 10%. Furthermore, the test based on the transformed VR statistic provides power that is uniformly higher than the power of the VR statistics under the Richardson–Stock asymptotic distribution.

TABLE 3. Power in percentage against the alternative of AR(1)

<i>n</i>	128				512			
	8		16		16		32	
<i>VR_p</i>	3.4	0.4	59.8	72.0				
<i>VR_p^β</i>	10.1	12.8	73.8	92.3				
RS	6.1	9.7	69.4	90.7				
	<i>Q_n</i>	<i>Q_n^β</i>	<i>IQ_n</i>	<i>IQ_n^β</i>	<i>Q_n</i>	<i>Q_n^β</i>	<i>IQ_n</i>	<i>IQ_n^β</i>
	2.3	11.7	3.9	19.2	33.4	84.2	66.0	92.2

$$x_t = 0.92x_{t-1} + u_t, u_t \sim N(0,1).$$

The quadratic test based on the transformed statistics also provides significant power gain over that based on the untransformed statistics. Furthermore, it is seen that the modified intersection test, which is specially geared to take into account the unidirectional nature of mean reverting alternatives, is able to provide a significant advantage over the quadratic test, when based on the transformed *VR* statistics.

We also generated data from the alternative mean reverting process given by $x_t = 0.92x_{t-1} + u_t$ where $u_t \sim i.i.d. N(0,1)$. This process is also considered in Lo and MacKinlay (1989). The simulation results are presented in Table 3. It is seen that the test based on the transformed *VR* statistics once again provides significantly higher power than that based on the untransformed statistics as well as that based on the Richardson–Stock distribution.

The simulations we present here are for the modified variance ratio statistic, *VR_p*, which is defined in the frequency domain as given in equation (10). It is of interest to see how good an approximation this statistic is to the variance ratio statistic, *VR*, defined in the time domain in equation (1). In Table 4, we present the empirical size and power of the transformed statistics *VR^β* and *VR_p^β*. It is seen that the size and power are very similar, indicating that the approximation is good, though as the theory suggests, this approximation will worsen as *k* gets larger relative to *n*.

It should also be noted that according to our theory the normal approximation to the transformed statistics will be good only when *k* is not too large relative to *n*. In the simulations we present here, the largest value of the ratio *k/n* we consider is 0.125, and the normal approximation works well in this case. We also did a simulation study, not presented here, in which *k/n* was set to be 0.25. In this case, the normal approximation to even the transformed ratio statistic was poor. This is not surprising because *k* is now very large relative to *n* and thus violates the assumption. Furthermore, in practice, one should not be using such large values of *k* because, as Deo and Richardson (2003) have shown, the test would then be inconsistent against a wide class of alternatives.

TABLE 4. Comparison of size and power of VR_p^β and VR^β

n	128				512			
	8		16		16		32	
k	VR_p^β	VR^β	VR_p^β	VR^β	VR_p^β	VR^β	VR_p^β	VR^β
Size								
I. Gaussian white noise								
Lower	2.4	2.6	2.4	2.8	2.4	2.6	2.4	2.7
Upper	2.3	2.5	2.4	2.8	2.4	2.5	2.5	2.8
II. GARCH(1,1)								
Lower	2.0	2.3	2.0	2.5	2.0	2.1	1.7	2.0
Upper	2.4	2.6	2.6	2.9	2.5	2.6	2.4	2.7
III. AR(1) + noise								
Power	12.2	13.2	15.3	17.3	82.6	83.5	94.1	95.1

I. $x_t = \mu + x_{t-1} + \varepsilon_t$, $\varepsilon_t \sim N(0,1)$.
 II. $x_t = \mu + x_{t-1} + \varepsilon_t$, $\varepsilon_t \sim N(0, \sigma_t^2)$, $\sigma_t^2 = 0.0001 + 0.8575\sigma_{t-1}^2 + 0.1171\varepsilon_{t-1}^2$.
 III. $x_t = r_t + y_t$, $r_t = r_{t-1} + w_t$, $w_t \sim N(0,0.1)$, $y_t = 0.9y_{t-1} + u_t$, $u_t \sim N(0,1)$.

5. CONCLUSION

From Deo and Richardson (2003), it is clear that large values of k should not be used when testing for the mean revision using the VR statistic. From our theoretical results and Monte Carlo study, we conclude that when k is not too large, the transformed VR statistic proposed in the paper is able to solve the problem of skewness and is thus well approximated by the normal distribution in finite samples. This provides good size properties in addition to significant power gains. Furthermore, the distribution of the transformed VR statistic is shown, both theoretically and through simulations, to be robust to conditional heteroskedasticity.

Our simulation study also shows that the $k/n \rightarrow \delta > 0$ asymptotic distribution cannot approximate the finite-sample distribution of the VR statistic when k/n is small and is sensitive to conditional heteroskedasticity. Furthermore, our transformed VR statistic provides power that is uniformly higher than that of the VR statistic based on the $k/n \rightarrow \delta > 0$ asymptotic distribution.

Finally, the modified intersection test is also able to incorporate information from various differencing periods and yet maintain good power.

REFERENCES

Anderson, T.W. (1994) *The Statistical Analysis of Time Series*, Wiley Classics edition. Wiley.
 Bollerslev, T. (1986) Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics* 31, 307–327.

Bougerol, P. & N. Picard (1992) Stationarity of GARCH processes and of some non-negative time series. *Journal of Econometrics* 52, 115–127.

Brockwell, P. & R. Davis (1996) *Time Series: Theory and Methods*, 2nd ed. Springer.

Campbell, J., A. Lo, & A. MacKinlay (1997) *The Econometrics of Financial Markets*. Princeton University Press.

Chen, W. & R. Deo (2004) Power transformations to induce normality and their applications. *Journal of the Royal Statistical Society, Series B* 66, 117–130.

Deo, R. (2000) Spectral tests of the martingale hypothesis under conditional heteroscedasticity. *Journal of Econometrics* 99, 291–315.

Deo, R. & M. Richardson (2003) On the asymptotic power of the variance ratio test. *Econometric Theory* 19, 231–239.

Fama, E. & K. French (1988) Permanent and temporary components of stock prices. *Journal of Political Economy* 96, 246–273.

Faust, J. (1992) When are variance ratio tests for serial dependence optimal? *Econometrica* 60, 1215–1226.

Follmann, D. (1996) A simple multivariate test for one-sided alternatives. *Journal of the American Statistical Association* 91, 854–861.

Fuller, W. (1996) *Introduction to Time Series Analysis*, 2nd ed. Wiley.

Lo, A. & A.C. MacKinlay (1988) Stock market prices do not follow random walks: Evidence from a simple specification test. *Review of Financial Studies* 1, 41–66.

Lo, A. & A.C. MacKinlay (1989) The size and power of the variance ratio test in finite samples: A Monte Carlo investigation. *Journal of Econometrics* 40, 203–238.

Nelson, D. (1990) Stationarity and persistence in the GARCH(1,1) model. *Econometric Theory* 6, 318–334.

Poterba, J. & L. Summers (1988) Mean reversion in stock returns: Evidence and implications. *Journal of Financial Economics* 22, 27–60.

Richardson, M. & T. Smith (1991) Tests of financial models with the presence of overlapping observations. *Review of Financial Studies* 4, 227–254.

Richardson, M. & J. Stock (1989) Drawing inferences from statistics based on multi-year asset returns. *Journal of Financial Economics* 25, 323–348.

Shephard, N. (1996) Statistical aspects of ARCH and stochastic volatility. In D.R. Cox, D.V. Hinkley, & O.E. Barndorff-Nielsen (eds.), *Time Series Models in Econometrics, Finance and Other Fields*. Chapman and Hall.

Stout, W. (1974) *Almost Sure Convergence*. Academic Press.

APPENDIX

Proof of Lemma 1. Because $\{h_t\}$ is a Gaussian stationary series with zero mean, it can be expressed as $h_t = \sum_{j=0}^{\infty} \alpha_j u_{t-j}$, where $\sum \alpha_j^2 < \infty$ and $\{u_t\}$ is a sequence of independent standard normal variables. Furthermore, $\{u_t\}$ and $\{v_t\}$ will also be independent. Let $\tilde{\delta}_t = \sigma(u_t, u_{t-1}, u_{t-2}, \dots, v_t, v_{t-1}, v_{t-2}, \dots)$. By Lemma 3.5.8 and Theorem 3.5.8 of Stout (1974), $\{\varepsilon_t\}$ is an ergodic sequence. Furthermore, Lemma 1 in Deo (2000) shows that ε_t satisfies (A1)–(A3). Because $\{v_t\}$ is an independent zero mean sequence, (A4) is trivially true. Also,

$$\begin{aligned}
 & E(\varepsilon_{t+n}^2 \varepsilon_{t+n+j}^2 | \tilde{\delta}_t) \\
 &= E^2(v_t^2) \exp \left\{ 2 \sum_{p=0}^{\infty} (\alpha_{p+n} + \alpha_{p+n+j}) u_{t-p} \right\} \exp \left(2 \sum_{s=0}^{j-1} \alpha_s^2 + 2 \sum_{p=0}^{n-1} (\alpha_p + \alpha_{p+j})^2 \right).
 \end{aligned}$$

Because $\sum \alpha_j^2 < \infty$, to prove (A5) it suffices to show that

$$\lim_{n \rightarrow \infty} \text{Var} \left(\exp \left\{ 2 \sum_{p=0}^{\infty} (\alpha_{p+n} + \alpha_{p+n+j}) u_{t-p} \right\} \right) = 0 \tag{A.1}$$

uniformly in j . But

$$\begin{aligned} &\text{Var} \left(\exp \left\{ 2 \sum_{p=0}^{\infty} (\alpha_{p+n} + \alpha_{p+n+j}) u_{t-p} \right\} \right) \\ &= \exp \left\{ 8 \sum_{p=0}^{\infty} (\alpha_{p+n} + \alpha_{p+n+j})^2 \right\} - \exp \left\{ 4 \sum_{p=0}^{\infty} (\alpha_{p+n} + \alpha_{p+n+j})^2 \right\} \\ &= \exp \left\{ 4 \sum_{p=0}^{\infty} (\alpha_{p+n} + \alpha_{p+n+j})^2 \right\} \left(\exp \left\{ 4 \sum_{p=0}^{\infty} (\alpha_{p+n} + \alpha_{p+n+j})^2 \right\} - 1 \right). \end{aligned}$$

Because $\sum_{p=0}^{\infty} (\alpha_{p+n} + \alpha_{p+n+j})^2$ converges to 0 uniformly in j , (A.1) is established. The proof of (A6) follows along similar lines. ■

Proof of Lemma 2. Lemma 2 in Deo (2000) proves (A1)–(A3). An argument similar to the one provided on page 309 in the proof of Lemma 2 of Deo (2000) also establishes (A4). We now turn to proving (A5). Iterating the expression for ε_t , we have

$$\begin{aligned} \varepsilon_{t+n}^2 &= v_{t+n}^2 \omega + v_{t+n}^2 \omega \sum_{k=1}^{n-2} \prod_{i=1}^k (\alpha v_{t+n-i}^2 + \beta) + v_{t+n}^2 \sigma_{t+1}^2 \prod_{i=1}^{n-1} (\alpha v_{t+n-i}^2 + \beta) \tag{A.2} \\ &\equiv T_{11} + T_{12} + T_{13} \end{aligned}$$

and

$$\begin{aligned} \varepsilon_{t+n+j}^2 &= v_{t+n+j}^2 \omega + v_{t+n+j}^2 \omega \sum_{k=1}^{n+j-2} \prod_{i=1}^k (\alpha v_{t+n+j-i}^2 + \beta) \\ &\quad + v_{t+n+j}^2 \sigma_{t+1}^2 \prod_{i=1}^{n-1} (\alpha v_{t+n+j-i}^2 + \beta) \\ &\equiv T_{21} + T_{22} + T_{23}. \end{aligned}$$

Thus,

$$E(\varepsilon_{t+n}^2 \varepsilon_{t+n+j}^2 | \mathcal{F}_t) = \sum_{p,q=1}^3 E(T_{1p} T_{2q} | \mathcal{F}_t). \tag{A.3}$$

Consider the term $T_{12} T_{23}$. Then we can easily see that we can express $T_{12} T_{23}$ as the product $T_{12} T_{23} = AB$, where

$$A = \omega \sigma_{t+1}^2 v_{t+n+j}^2 v_{t+n}^2 (\alpha v_{t+n}^2 + \beta) \prod_{i=1}^{j-1} (\alpha v_{t+n+i}^2 + \beta)$$

and

$$B = \prod_{i=1}^{n-1} (\alpha v_{i+n-i}^2 + \beta) \sum_{k=1}^{n-2} \prod_{i=1}^k (\alpha v_{i+n-i}^2 + \beta).$$

Letting $\theta_1 = E(\alpha v_t^2 + \beta)$ and $\theta_2 = E(\alpha v_t^2 + \beta)^2$ and noting that $Ev_{t+n}^4 = 3$, we get

$$E(T_{12}T_{23}|\mathcal{F}_t) \leq \omega \sigma_{t+1}^2 (3\alpha + \beta) \theta_1^{j-1} \sum_{k=1}^{n-2} \theta_2^k \theta_1^{n-1-k}.$$

Because $\gamma = \max(\theta_1, \theta_2) < 1$, it follows that for all $j \geq 1$ there exists some finite constant C such that

$$E(T_{12}T_{23}|\mathcal{F}_t) \leq C \sigma_{t+1}^2 (n - 2) \gamma^{n-1},$$

and hence

$$\text{Var}(E(T_{12}T_{23}|\mathcal{F}_t)) \leq E(E(T_{12}T_{23}|\mathcal{F}_t))^2 \leq C^2 E(\sigma_{t+1}^4) n^2 \gamma^{2(n-1)}$$

uniformly in j . Thus,

$$\lim_{n \rightarrow \infty} \text{Var}(E(T_{12}T_{23}|\mathcal{F}_t)) = 0$$

uniformly in j . Similar arguments yield

$$\lim_{n \rightarrow \infty} \text{Var}(E(T_{1p}T_{2q}|\mathcal{F}_t)) = 0 \quad 1 \leq p, \quad q \leq 3 \tag{A.4}$$

uniformly in j . Thus, (A5) follows from (A.3), (A.4), and the Cauchy–Schwarz inequality. To prove (A6), we first note that using (A.2),

$$E(\varepsilon_t^2|\mathcal{F}_{t-n}) = \omega \left(\frac{1 - \theta_1^n}{1 - \theta_1} \right) + \theta_1^n \sigma_{t-n}^2.$$

Thus, $E(\varepsilon_t^2 \varepsilon_{t-n}^2) = E(\varepsilon_{t-n}^2 E(\varepsilon_t^2|\mathcal{F}_{t-n})) = \omega((1 - \theta_1^n)/(1 - \theta_1)) \sigma^2 + \theta_1^n E(\sigma_{t-n}^4)$, and so

$$\lim_{n \rightarrow \infty} E(\varepsilon_t^2 \varepsilon_{t-n}^2) = \omega \left(\frac{1 - \theta_1^n}{1 - \theta_1} \right) \sigma^2 = \sigma^4. \quad \blacksquare$$

Proof of Theorem 3. By simple but tedious algebraic manipulation, it can be shown that

$$\begin{aligned} [VR(k_i) - 1] &= \frac{2n^2}{\hat{\sigma}^2(n - k_i + 1)(n - k_i)} \sum_{j=1}^{k_i-1} \left(1 - \frac{j}{k_i}\right) \hat{\gamma}_j - \frac{n(A_i + B_i)}{\hat{\sigma}^2 k_i (n - k_i + 1)(n - k_i)} \\ &\quad + o_p \left(\sqrt{\frac{k_i}{n}} \right), \end{aligned}$$

where

$$\hat{\gamma}_j = n^{-1} \sum_{t=j+1}^n \varepsilon_t \varepsilon_{t-j},$$

$$A_i = -2 \sum_{v=0}^{k_i-2} \sum_{p=1}^{k_i-1-v} \sum_{s=p+1}^{k_i-v-1} \varepsilon_s \varepsilon_{s-p} - 2 \sum_{v=0}^{k_i-2} \sum_{p=1}^{k_i-1-v} \sum_{s=n-v+1}^n \varepsilon_s \varepsilon_{s-p}$$

$$= A_{i1} + A_{i2},$$

and

$$B_i = \sum_{v=0}^{k_i-1} \sum_{q=1}^{k_i-v-1} \varepsilon_q^2 + \sum_{v=0}^{k_i-1} \sum_{q=n-v+1}^n \varepsilon_q^2.$$

Because $E(B_i) = O(k_i^2)$ trivially, it follows that $[k_i(n - k_i + 1)(n - k_i)]^{-1}nB_i = o_p([n^{-1}k_i]^{1/2})$. By condition (A1), we have $E(A_i) = 0$. Furthermore, by using condition (A3), it can be easily seen that $E(A_{i1}^2) = E(A_{i2}^2) = O(k_i^4)$. By the Cauchy–Schwarz inequality, it follows that $\text{Var}(A_i) = O(k_i^4)$ and hence $[k_i(n - k_i + 1)(n - k_i)]^{-1}nA_i = o_p([n^{-1}k_i]^{1/2})$. Because $\hat{\sigma}^2 \xrightarrow{P} \sigma^2$, we have

$$\sqrt{\frac{n}{k_i}} [VR(k_i) - 1] = \frac{n^2}{\hat{\sigma}^2(n - k_i + 1)(n - k_i)} \sqrt{\frac{n}{k_i}} 2 \sum_{j=1}^{k_i-1} \left(1 - \frac{j}{k_i}\right) \hat{\gamma}_j + o_p(1). \tag{A.5}$$

Now consider

$$\sqrt{\frac{n}{k_i}} 2 \sum_{j=1}^{k_i-1} \left(1 - \frac{j}{k_i}\right) \hat{\gamma}_j = \frac{2}{\sqrt{nk_i}} \sum_{j=1}^{k_i-1} \left(1 - \frac{j}{k_i}\right) \sum_{q=1}^{n-j} \varepsilon_q \varepsilon_{q+j}$$

$$= \frac{2}{\sqrt{nk_i}} \sum_{q=1}^{n-k_i} \sum_{j=1}^{k_i} \left(1 - \frac{j}{k_i}\right) \varepsilon_q \varepsilon_{q+j}$$

$$+ \frac{2}{\sqrt{nk_i}} \sum_{q=n-k_i+1}^n \sum_{j=1}^{n-q} \left(1 - \frac{j}{k_i}\right) \varepsilon_q \varepsilon_{q+j}$$

$$= R_{i1} + R_{i2}.$$

By conditions (A1) and (A3), respectively, it follows that $E(R_{i2}) = 0$ and $E(R_{i2}^2) = o(1)$, and hence

$$\sqrt{\frac{n}{k_i}} 2 \sum_{j=1}^{k_i-1} \left(1 - \frac{j}{k_i}\right) \hat{\gamma}_j = R_{i1} + o_p(1). \tag{A.6}$$

Now define $N = [\sqrt{nk_s}]$ and $M = [N^{-1}n]$. Then, $M \rightarrow \infty, N \rightarrow \infty, n^{-1}N \rightarrow 0$, and $N^{-1}k_i \rightarrow 0$ for $i = 1, 2, \dots, s$. Also, define

$$W_{i,q} = \frac{1}{\sqrt{k_i}} \sum_{j=1}^{k_i} \left(1 - \frac{j}{k_i}\right) \varepsilon_q \varepsilon_{q+j} \quad q = 1, 2, \dots, n - k_i,$$

$$Z_{i,p} = \frac{1}{\sqrt{N}} \{W_{i,(p-1)N+1} + \dots + W_{i,pN-k_i}\} \quad p = 1, 2, \dots, M,$$

and

$$V_{i,l} = W_{i,lN-k_i+1} + \dots + W_{i,lN} \quad l = 1, 2, \dots, M - 1.$$

Then we can decompose R_{i1} as

$$R_{i1} = \frac{2}{\sqrt{M}} \sum_{p=1}^M Z_{i,p} + \frac{2}{\sqrt{n}} \sum_{l=1}^{M-1} V_{i,l} \tag{A.7}$$

$$\equiv U_{i1} + U_{i2}.$$

By condition (A3), it follows that $E(W_{i,a}W_{i,b}) = 0$ for $a < b$ and hence $E(V_{i,a}V_{i,b}) = 0$ for $a < b$. Thus,

$$E(U_{i2}^2) = \frac{4}{n} \sum_{l=1}^{M-1} E(V_{i,l}^2) = \frac{4}{n} \sum_{l=1}^{M-1} k_i E(W_{i,1}^2)$$

$$= O\left(\frac{k_i(M-1)}{n}\right) = o(1) \quad i = 1, 2, \dots, s. \tag{A.8}$$

From equations (A.6)–(A.8) it follows that

$$\sqrt{\frac{n}{k_i}} 2 \sum_{j=1}^{k_i-1} \left(1 - \frac{j}{k_i}\right) \hat{\gamma}_j = U_{i1} + o_p(1),$$

and hence, from equation (A.5),

$$\sqrt{\frac{n}{k_i}} [VR(k_i) - 1] = \frac{n^2}{\hat{\sigma}^2(n - k_i + 1)(n - k_i)} U_{i1} + o_p(1).$$

Because $\hat{\sigma}^2 \xrightarrow{P} \sigma^2$ and $[(n - k_i + 1)(n - k_i)]^{-1}n^2 \rightarrow 1$, the theorem will be proved if we show that the vector $(U_{11}, U_{21}, \dots, U_{s1})'$ converges in distribution to a multivariate normal distribution with mean zero and variance covariance matrix $\sigma^4 \Sigma$. To do this, it is sufficient to show that for any set of s real numbers c_i ,

$$\sum_{i=1}^s c_i U_{i1} = 2M^{-1/2} \sum_{p=1}^M \sum_{i=1}^s c_i Z_{i,p} \xrightarrow{D} N\left(0, \sigma^4 \sum_{i,j} c_i c_j \sigma_{ij}\right), \tag{A.9}$$

which we now proceed to demonstrate.

Let $\mathbb{G}_{p,n} = \sigma\{\varepsilon_{pN}, \varepsilon_{pN-1}, \varepsilon_{pN-2}, \dots\}$ be the sigma algebra generated by $\{\varepsilon_{pN}, \varepsilon_{pN-1}, \varepsilon_{pN-2}, \dots\}$. Then, for any set of s real numbers c_i , the sequence $\{\sum_{i=1}^s c_i Z_{i,p}\}$ forms a martingale difference with respect to $\mathbb{G}_{p,n}$. To show (A.9), we first need to establish that

$$\left(\sum_{p=1}^M E \left(\sum_{i=1}^s c_i Z_{i,p} \right)^2 \right)^{-1} \sum_{p=1}^M E \left[\left(\sum_{i=1}^s c_i Z_{i,p} \right)^2 \mid \mathbb{G}_{p-1,n} \right] \xrightarrow{P} 1. \tag{A.10}$$

Now, by condition (A3)

$$\begin{aligned} E \left(\left[\sum_{i=1}^s c_i Z_{i,p} \right]^2 \right) &= \sum_{i=1}^s c_i^2 E(Z_{i,p}^2) + 2 \sum_{i < u} c_i c_u E(Z_{i,p} Z_{u,p}) \\ &= \sum_{i=1}^s c_i^2 \frac{N - k_i}{N} E(W_{i,1}^2) + 2 \sum_{i < u} c_i c_u \frac{N - k_i}{N} E(W_{i,1} W_{u,1}). \end{aligned}$$

By conditions (A3) and (A6),

$$E(W_{i,1}^2) = k_i^{-1} \sum_{j=1}^{k_i} \left(1 - \frac{j}{k_i} \right)^2 E(\varepsilon_1^2 \varepsilon_{1+j}^2) \rightarrow 4^{-1} \sigma^4 \sigma_{ii}$$

and

$$E(W_{i,1} W_{u,1}) = (k_u k_i)^{-1/2} \sum_{j=1}^{k_i} \left(1 - \frac{j}{k_i} \right) \left(1 - \frac{j}{k_u} \right) E(\varepsilon_1^2 \varepsilon_{1+j}^2) \rightarrow 4^{-1} \sigma^4 \sigma_{iu}$$

for $i < u$. Hence, we have

$$\lim_{M \rightarrow \infty} \frac{1}{M} \sum_{p=1}^M E \left(\left[\sum_{i=1}^s c_i Z_{i,p} \right]^2 \right) = 4^{-1} \sigma^4 \sum_{i,j} c_i c_j \sigma_{ij}. \tag{A.11}$$

We now show that

$$M^{-1} \sum_{p=1}^M E \left[\left(\sum_{i=1}^s c_i Z_{i,p} \right)^2 \mid \mathbb{G}_{p-1,n} \right] \xrightarrow{P} 4^{-1} \sigma^4 \sum_{i,j} c_i c_j \sigma_{ij}, \tag{A.12}$$

which along with (A.11) will prove (A.10). We have

$$\begin{aligned} E \left[\left(\sum_{i=1}^s c_i Z_{i,p} \right)^2 \mid \mathbb{G}_{p-1,n} \right] &= \sum_{i=1}^s c_i^2 E(Z_{i,p}^2 \mid \mathbb{G}_{p-1,n}) \\ &\quad + 2 \sum_{i < u} c_i c_u E(Z_{i,p} Z_{u,p} \mid \mathbb{G}_{p-1,n}). \end{aligned}$$

Letting $f(x) = (1 - x)$, $Y_{i,u,p} = E(Z_{i,p} Z_{u,p} \mid \mathbb{G}_{p-1,n})$ and using condition (A4), we get for $i \leq u$,

$$Y_{i,u,p} = N^{-1} (k_i k_u)^{-1/2} \sum_{a=1}^{N-k_u} \sum_{b=1}^{k_i} f\left(\frac{b}{k_i}\right) f\left(\frac{b}{k_u}\right) E(\varepsilon_{pN-N+a}^2 \varepsilon_{pN-N+a+b}^2 \mid \mathbb{G}_{p-1,n}).$$

By condition (A6), there exists $C < \infty$ such that

$$E|E(\varepsilon_{pN-N+a}^2 \varepsilon_{pN-N+a+b}^2 | \mathbb{G}_{p-1,n}) - E(\varepsilon_t^2 \varepsilon_{t+b}^2)| < C \tag{A.13}$$

for all $p, a,$ and $b.$ Furthermore, given any $\delta > 0,$ by condition (A5) and Jensen’s inequality there exists an integer N_0 such that

$$\sup_{b > 0} E|E(\varepsilon_{pN-N+a}^2 \varepsilon_{pN-N+a+b}^2 | \mathbb{G}_{p-1,n}) - E(\varepsilon_t^2 \varepsilon_{t+b}^2)| < \delta \tag{A.14}$$

for all $a > N_0.$ Hence, letting $H_{p,a,b} = E(\varepsilon_{pN-N+a}^2 \varepsilon_{pN-N+a+b}^2 | \mathbb{G}_{p-1,n}) - E(\varepsilon_t^2 \varepsilon_{t+b}^2),$ we have for any $\varepsilon > 0$

$$\begin{aligned} &P \left\{ \left| M^{-1} \sum_{p=1}^M E(Z_{i,p}^2 | \mathbb{G}_{p-1,n}) - M^{-1} \sum_{p=1}^M N^{-1} k_i^{-1} \sum_{a=1}^{N-k_i} \sum_{b=1}^{k_i} f^2 \left(\frac{b}{k_i} \right) E(\varepsilon_t^2 \varepsilon_{t+b}^2) \right| > \varepsilon \right\} \\ &\leq P \left\{ N^{-1} k_i^{-1} \sum_{a=1}^{N-k_i} \sum_{b=1}^{k_i} f^2 \left(\frac{b}{k_i} \right) M^{-1} \sum_{p=1}^M |H_{p,a,b}| > \varepsilon \right\} \\ &\leq P \left\{ N^{-1} k_i^{-1} \sum_{a=1}^{N_0} \sum_{b=1}^{k_i} f^2 \left(\frac{b}{k_i} \right) M^{-1} \sum_{p=1}^M |H_{p,a,b}| > 2^{-1} \varepsilon \right\} \\ &\quad + P \left\{ N^{-1} k_i^{-1} \sum_{a=N_0+1}^{N-k_i} \sum_{b=1}^{k_i} f^2 \left(\frac{b}{k_i} \right) M^{-1} \sum_{p=1}^M |H_{p,a,b}| > 2^{-1} \varepsilon \right\} \\ &\leq 2\varepsilon^{-1} N^{-1} k_i^{-1} \sum_{a=1}^{N_0} \sum_{b=1}^{k_i} f^2 \left(\frac{b}{k_i} \right) M^{-1} \sum_{p=1}^M E|H_{p,a,b}| \\ &\quad + 2\varepsilon^{-1} N^{-1} k_i^{-1} \sum_{a=N_0+1}^{N-k_i} \sum_{b=1}^{k_i} f^2 \left(\frac{b}{k_i} \right) M^{-1} \sum_{p=1}^M E|H_{p,a,b}| \\ &\leq 2\varepsilon^{-1} N^{-1} N_0 k_i^{-1} \sum_{b=1}^{k_i} f^2 \left(\frac{b}{k_i} \right) C + 2\varepsilon^{-1} N^{-1} (N - k_i) k_i^{-1} \sum_{b=1}^{k_i} f^2 \left(\frac{b}{k_i} \right) \delta, \tag{A.15} \end{aligned}$$

where the last inequality follows from equations (A.13) and (A.14). Because δ can be chosen to be arbitrarily small and N large enough that $N^{-1}N_0 \rightarrow 0,$ it follows from equation (A.15) that

$$M^{-1} \sum_{p=1}^M E(Z_{i,p}^2 | \mathbb{G}_{p-1,n}) - M^{-1} \sum_{p=1}^M N^{-1} k_i^{-1} \sum_{a=1}^{N-k_i} \sum_{b=1}^{k_i} f^2 \left(\frac{b}{k_i} \right) E(\varepsilon_t^2 \varepsilon_{t+b}^2) \xrightarrow{P} 0.$$

Because, by condition (A6) we also have

$$M^{-1} \sum_{p=1}^M N^{-1} k_i^{-1} \sum_{a=1}^{N-k_i} \sum_{b=1}^{k_i} f^2 \left(\frac{b}{k_i} \right) E(\varepsilon_t^2 \varepsilon_{t+b}^2) \rightarrow \sigma^4 4^{-1} \sigma_{ii},$$

we obtain

$$M^{-1} \sum_{p=1}^M E(Z_{i,p}^2 | \mathcal{G}_{p-1,n}) \xrightarrow{P} \sigma^4 4^{-1} \sigma_{ii}.$$

A similar argument as before in conjunction with the fact that $k_u^{-1} k_i \rightarrow a_{iu}$ for $i < u$ yields

$$M^{-1} \sum_{p=1}^M E(Z_{i,p} Z_{u,p} | \mathcal{G}_{p-1,n}) \xrightarrow{P} \sigma^4 4^{-1} \sigma_{iu}.$$

Thus, (A.12) is established giving equation (A.10).

By using condition (A3), one can employ the same argument given on page 539 of Anderson (1994) to show that $E(Z_{i,p}^4)$ is uniformly bounded in n for $i = 1, 2, \dots, s$. This implies that $E(\sum_{i=1}^s c_i Z_{i,p})^4$ is also uniformly bounded in n , from whence we get

$$M^{-1} \sum_{p=1}^M E \left[\left(\sum_{i=1}^s c_i Z_{i,p} \right)^2 I \left(\left| \sum_{i=1}^s c_i Z_{i,p} \right| > \varepsilon \sqrt{M} \right) \right] \rightarrow 0 \tag{A.16}$$

for every $\varepsilon > 0$. By Chebyshev’s inequality, equation (A.16) implies that

$$M^{-1} \sum_{p=1}^M E \left[\left(\sum_{i=1}^s c_i Z_{i,p} \right)^2 I \left(\left| \sum_{i=1}^s c_i Z_{i,p} \right| > \varepsilon \sqrt{M} \right) | \mathcal{G}_{p-1,n} \right] \xrightarrow{P} 0. \tag{A.17}$$

Hence, equation (A.9) follows from equations (A.10) and (A.17) and Theorem 5.3.4 of Fuller (1996). ■

Proof of Theorem 4. We first note that by the weak law of large numbers, $\hat{\sigma}_a^2 \xrightarrow{P} \text{Var}(z_t) + \text{Var}(y_t - y_{t-1})$. Now, letting $V_{n,k} \equiv n(\hat{\sigma}_a^2 k(n - k + 1)(n - k))^{-1}$, we get

$$\begin{aligned} VR(k) &= V_{n,k} \sum_{t=k}^n (x_t - x_{t-k} - k\hat{\mu})^2 \\ &= V_{n,k} \sum_{t=k}^n \left[\left\{ \sum_{j=t-k+1}^t z_j - k\bar{z} \right\} + \sum_{t=k}^n \left(y_t - y_{t-k} - \frac{k}{n} \{y_n - y_0\} \right) \right]^2. \end{aligned} \tag{A.18}$$

It is trivial to show that

$$\sum_{t=k}^n \left(y_t - y_{t-k} - \frac{k}{n} \{y_n - y_0\} \right)^2 = o_p(nk). \tag{A.19}$$

Now

$$\begin{aligned} \sum_{t=k}^n \left\{ \sum_{j=t-k+1}^t z_j - k\bar{z} \right\}^2 &= \sum_{t=k}^n \left\{ \sum_{j=t-k+1}^t z_j \right\}^2 + (n - k)k^2 \bar{z}^2 - 2k\bar{z} \sum_{t=k}^n \sum_{j=t-k+1}^t z_j \\ &= \sum_{t=k}^n \left\{ \sum_{j=t-k+1}^t z_j \right\}^2 + (n - k)k^2 O_p(n^{-1}) \\ &\quad - 2k\bar{z} \left(\sum_{j=0}^k jz_j + \sum_{j=k}^{n-k} kz_j + \sum_{j=n-k}^n (n - j)z_j \right) \\ &= \sum_{t=k}^n \left\{ \sum_{j=t-k+1}^t z_j \right\}^2 + O_p(k^2) + O_p(k^2). \end{aligned} \tag{A.20}$$

From (A.20), we get

$$V_{n,k} \sum_{t=k}^n \left\{ \sum_{j=t-k+1}^t z_j - k\bar{z} \right\}^2 = V_{n,k} \sum_{t=k}^n \left\{ \sum_{j=t-k+1}^t z_j \right\}^2 + o_p(1). \tag{A.21}$$

Letting $\hat{\gamma}_j = \hat{\gamma}_{-j} = n^{-1} \sum_{t=j+1}^n z_t z_{t-j}$, some tedious algebra yields

$$\sum_{t=k}^n \left\{ \sum_{j=t-k+1}^t z_j \right\}^2 = nk \sum_{j=-(k-1)}^{k-1} (1 - |j|/k) \hat{\gamma}_j - A - B, \tag{A.22}$$

where

$$\begin{aligned} A &= -2 \sum_{v=0}^{k-2} \sum_{p=1}^{k-1-v} \sum_{s=p+1}^{k-v-1} z_s z_{s-p} - 2 \sum_{v=0}^{k-2} \sum_{p=1}^{k-1-v} \sum_{s=n-v+1}^n z_s z_{s-p} \\ &= A_1 + A_2 \end{aligned}$$

and

$$B = \sum_{v=0}^{k-1} \sum_{q=1}^{k-v-1} z_q^2 + \sum_{v=0}^{k-1} \sum_{q=n-v+1}^n z_q^2.$$

Now

$$E(A_1^2) = 4 \sum_{p=1}^{k-1} \sum_{v=1}^{k-1-p} \sum_{s=1}^{k-1} \sum_{j=1}^{k-1-s} (k-v-p)(k-j-s) E(z_v z_{v+p} z_j z_{j+s}).$$

From equation (6.2.5) on page 315 of Fuller (1996), we have $|E(z_v z_{v+p} z_j z_{j+s})| = O(\lambda^{|v|+|p|+|j|+|s|})$, and hence

$$E(A_1^2) = O(k^4).$$

A similar argument shows that $E(A_2^2) = O(k^4)$, and hence, by the Cauchy–Schwarz and Chebyshev inequalities, we get

$$A = O_p(k^2). \tag{A.23}$$

Because $E(B) = O(k^2)$ trivially, it follows from (A.23), (A.22), and (A.21) that

$$V_{n,k} \sum_{t=k}^n \left\{ \sum_{j=t-k+1}^t z_j - k\bar{z} \right\}^2 = V_{n,k} nk \sum_{j=-(k-1)}^{k-1} (1 - |j|/k) \hat{\gamma}_j + o_p(1).$$

From Theorem 9.3.3 and Theorem 9.4.1 of Anderson (1994), it follows that

$$\sum_{j=-(k-1)}^{k-1} (1 - |j|/k) \hat{\gamma}_j \xrightarrow{P} \sum_{|j|<\infty} \gamma_z(j),$$

and hence

$$V_{n,k} \sum_{t=k}^n \left\{ \sum_{j=t-k+1}^t z_j - k\bar{z} \right\}^2 \xrightarrow{P} (\text{Var}(z_t) + \text{Var}(y_t - y_{t-1}))^{-1} \sum_{|j|<\infty} \gamma_z(j). \tag{A.24}$$

From (A.18), (A.19), (A.24), and the Cauchy–Schwarz inequality, we get

$$VR(k) \xrightarrow{P} (\text{Var}(z_t) + \text{Var}(y_t - y_{t-1}))^{-1} \sum_{|j|<\infty} \gamma_z(j).$$

LEMMA 7.

- (i) $\int_0^{2\pi} W_k(\lambda)I(\lambda) d\lambda = (4\pi/n) \sum_{j=1}^{[(n-1)/2]} W_k(\lambda_j)I(\lambda_j) + o_p(\sqrt{k/n})$.
- (ii) *The finite-sample variance covariance matrix of $\mathbf{V}_p = (VR_p(k_1), VR_p(k_2), \dots, VR_p(k_s))'$ with remainder terms of order $o(k_s^2/n^2)$ is estimated consistently by the matrix $\hat{\Sigma}$ in (11).*

Proof of (i). Using the fact that $I(\lambda) = (2\pi)^{-1} \sum_{|s|<n} \hat{\gamma}_s \exp(-is\lambda)$ and that

$$\begin{aligned} \sum_{j=0}^{n-1} \exp(-i(s-p)\lambda_j) &= n \quad \text{if } s-p=0, \quad \pm n \\ &= 0 \quad \text{otherwise,} \end{aligned}$$

we get

$$\begin{aligned} \frac{2\pi}{n} \sum_{j=1}^{n-1} W_k(\lambda_j)I(\lambda_j) &= \frac{2\pi}{n} \sum_{j=0}^{n-1} W_k(\lambda_j)I(\lambda_j) - \frac{2\pi k}{n} I(0) \\ &= \frac{2\pi}{n} \sum_{j=0}^{n-1} \sum_{|p|<k} (1 - |p|/k) \exp(ip\lambda_j) (2\pi)^{-1} \\ &\quad \times \sum_{|s|<n} \hat{\gamma}_s \exp(-is\lambda_j) - \frac{2\pi k}{n} I(0) \\ &= \frac{1}{n} \sum_{|p|<k} \sum_{|s|<n} (1 - |p|/k) \hat{\gamma}_s \sum_{j=0}^{n-1} \exp(-i(s-p)\lambda_j) - \frac{2\pi k}{n} I(0) \\ &= \sum_{|p|<k} (1 - |p|/k) \hat{\gamma}_p + 2 \sum_{p=1}^k (1 - p/k) \hat{\gamma}_{n-p} - \frac{2\pi k}{n} I(0) \\ &= \int_0^{2\pi} W_k(\lambda)I(\lambda) d\lambda + 2 \sum_{p=1}^k (1 - p/k) \hat{\gamma}_{n-p} - \frac{2\pi k}{n} I(0), \tag{A.25} \end{aligned}$$

where the last step follows from the identity $\hat{\gamma}_j = \int_0^{2\pi} I(\lambda) \exp(-ij\lambda) d\lambda$. We now note that because $I(\lambda) = (2\pi n)^{-1} |\sum_{t=1}^n \varepsilon_t \exp(it\lambda)|^2$, it follows that $(2\pi k/n)I(0) = k\bar{\varepsilon}^2 = O_p(kn^{-1}) = O_p(\sqrt{k/n})$. Furthermore, $\text{Var}(\hat{\gamma}_{n-p}) = O(\frac{pn}{n^2})$, whereas $\text{Cov}(\hat{\gamma}_{n-p}, \hat{\gamma}_{n-s}) = 0$, which implies that $2 \sum_{p=1}^k (1 - p/k) \hat{\gamma}_{n-p} = O_p(\sqrt{k/n})$. Part (i) of the lemma now follows by noting that

$$(2\pi/n) \sum_{j=1}^{n-1} W_k(\lambda_j)I(\lambda_j) = (4\pi/n) \sum_{j=1}^{[(n-1)/2]} W_k(\lambda_j)I(\lambda_j) + \frac{2\pi}{nk} I(\lambda_{n/2})\delta_{\{n \text{ even}, k \text{ odd}\}},$$

where δ is the indicator function due to the periodicity of the sine and cosine functions on $[0, 2\pi]$. ■

Proof of (ii). Using a Taylor series expansion and equation (A.25) in the proof of part (i), we get

$$\begin{aligned} VR_p(k) &= 1 + C_{n,k}(4\pi/n) \sum_{j=1}^{[(n-1)/2]} W_k(\lambda_j)I(\lambda_j) - \hat{\sigma}^2 + O_p(k^{1/2}/n) \\ &= 1 + 2C_{n,k} \sum_{j=1}^{k-1} (1 - j/k)(\hat{\gamma}_j + \hat{\gamma}_{n-j}) - (kC_{n,k} - n/(n-1))\bar{\varepsilon}^2 \\ &\quad + O(k/n)\hat{\gamma}_0 + O_p(k^{1/2}/n). \end{aligned}$$

Now define the random vector $\mathbf{U} = (\hat{\gamma}_1 + \hat{\gamma}_{n-1}, \hat{\gamma}_2 + \hat{\gamma}_{n-2}, \dots, \hat{\gamma}_s + \hat{\gamma}_{n-s}, \bar{\varepsilon}^2)$. Because $\text{Var}(\hat{\gamma}_0) = O(n^{-1})$, it is seen that

$$\text{Var}(VR_p(k)) = l'_k \text{Var}(\mathbf{U})l_k + o(k^2/n^2), \tag{A.26}$$

where l_k is as defined in (12). Letting $\tau_j = \sigma^{-4}E(\varepsilon_t^2 \varepsilon_{t-j}^2)$, tedious but elementary calculation shows that

$$\text{Var}(\mathbf{U}) = \begin{bmatrix} \mathbf{A}_0 & \mathbf{b}_0 \\ \mathbf{b}'_0 & d_0 \end{bmatrix}, \tag{A.27}$$

where $\mathbf{A}_0 = \text{diag}(((n-j)/n^2)\tau_j + (j/n^2)\tau_{n-j})$ for $j = 1, \dots, k_s$, \mathbf{b}_0 is a $k_s \times 1$ vector such that its j th element is given by $(2(n-j)n^{-3}\tau_j + 2jn^{-3}\tau_{n-j})$, and $d_0 = n^{-3}\tau_0 + 6n^{-4} \sum_{u=1}^{n-1} (n-u)\tau_u - n^{-2}$. Using the fact that by assumption (A6) $\tau_j \rightarrow 1$ as $j \rightarrow \infty$, it is easily seen that $6n^{-2} \sum_{u=1}^{n-1} (n-u)\tau_u = 3 + o(1)$ and using these facts in conjunction with substituting (A.27) in (A.26), we get

$$\text{Var}(VR_p(k)) = l'_k \begin{bmatrix} \mathbf{A} & \mathbf{b} \\ \mathbf{b}' & d \end{bmatrix} l_k + o(k^2/n^2),$$

where $\mathbf{A} = \text{diag}(((n-j)/n^2)\tau_j + (j/n^2))$ for $j = 1, \dots, k_s$, \mathbf{b} is a $k_s \times 1$ vector such that its j th element is given by $(2(n-j)n^{-3}\tau_j + 2jn^{-3})$, and $d = 2n^{-2}$. The estimated variance covariance matrix is now obtained by replacing τ_j in the entries of \mathbf{A} and \mathbf{b} by $\hat{\tau}_j$, and standard arguments from smoothing theory establish consistency of the resulting estimated covariance matrix. ■

Proof of Theorem 6. In the proof of Lemma 7, we noted that

$$(4\pi/n) \sum_{j=1}^{[(n-1)/2]} W_k(\lambda_j)I(\lambda_j) = \sum_{|p|<k} (1 - |p|/k)\hat{\gamma}_p + 2 \sum_{p=1}^k (1 - p/k)\hat{\gamma}_{n-p} - k\bar{\varepsilon}^2.$$

It is trivially true that under the assumptions of Theorem 6, $\bar{\varepsilon}^2 = O_p(n^{-1})$. The result for $VR_p(k)$ now follows by noting that $\sum_{p=1}^k (1 - p/k) \hat{\gamma}_{n-p} = o_p(1)$, that $\hat{\sigma}^2 \xrightarrow{p} (\text{Var}(z_t) + \text{Var}(y_t - y_{t-1}))$, and that by Theorem 9.3.3 and Theorem 9.4.1 of Anderson (1994),

$$\sum_{j=-(k-1)}^{k-1} (1 - |j|/k) \hat{\gamma}_j \xrightarrow{p} \sum_{|j| < \infty} \gamma_z(j).$$

The result for $VR_p^\beta(k)$ follows by continuity. ■