Aust. N. Z. J. Stat. **50**(3), 2008, 273–292 doi: 10.1111/j.1467-842X.2008.00517.x

UNIT ROOT TESTING IN THE PRESENCE OF HEAVY-TAILED GARCH ERRORS

GAOWEN WANG[1]* AND WEI-LIN MAO[2]

Takming College and National Chengchi University

Summary

We derive the asymptotic distributions of the Dickey–Fuller (DF) and augmented DF (ADF) tests for unit root processes with Generalized Autoregressive Conditional Heteroscedastic (GARCH) errors under fairly mild conditions. We show that the asymptotic distributions of the DF tests and ADF t-type test are the same as those obtained in the independent and identically distributed Gaussian cases, regardless of whether the fourth moment of the underlying GARCH process is finite or not. Our results go beyond earlier ones by showing that the fourth moment condition on the scaled conditional errors is totally unnecessary. Some Monte Carlo simulations are provided to illustrate the finite-sample-size properties of the tests.

Key words: augmented Dickey–Fuller tests; Lindeberg condition; martingale invariance principle; self-normalized sums.

1. Introduction

Unit root tests with independent and identically distributed (i.i.d.) errors having zero mean and finite variance were proposed by Dickey & Fuller (1979, 1981) and are referred to as Dickey–Fuller (DF) tests. They have recently found widespread application in economic time series. Obviously, the i.i.d. assumption is very restrictive in practice. The tests have been extensively studied in the econometric literature, and various generalizations have been proposed for handling unit root processes with errors being dependent, heteroscedastic or heavy-tailed: see Phillips & Xiao (1998) and Stock (1994) for reviews. However, Autoregressive Conditional Heteroscedastic (ARCH) (Engle, 1982) and Generalized ARCH (GARCH) (Bollerslev, 1986) models are the most popular volatility models, and are widely used in empirical finance. Recently, non-stationary autoregressive (AR) models with GARCH-type errors have received increasing attention in the literature. Investigations by Pantula (1988), Ling & Li (1998, 2003), Seo (1999), and Ling, Li & McAleer (2003) have been at the forefront of research in this area; see Li, Ling & McAleer (2002) for a review of the theory.

In the present paper we are interested in the asymptotics of unit root processes with GARCH-type errors. Formally, consider the following AR(1)–GARCH(p, q)

*Author to whom correspondence should be addressed.

[1]Department of Banking and Finance, Takming College, No. 56, Sec. 1, Huanshan Rd., Neihu, Taipei 11451, Taiwan.
e-mail: GWWang@takming.edu.tw

[2]Department of Economics, National Chengchi University, No. 64, Sec. 2, Zhinan Rd., Wenshan, Taipei 11605, Taiwan.

Acknowledgements. The authors are grateful to the editor, an associate editor, and two reviewers for their helpful comments and suggestions. G.W.'s research was supported by NSC grant NSC-95-2415-H-147-001.

process:

$$y_t = \phi y_{t-1} + u_t, \quad \phi = 1, \tag{1}$$

$$u_t = z_t \sigma_t, \quad \sigma_t^2 = w + \alpha(L) u_t^2 + \beta(L) \sigma_t^2, \tag{2}$$

where $w > 0, \alpha(L) = \alpha_1 L + \cdots + \alpha_q L^q, \alpha_i \geq 0, i = 1, \ldots, q$, and $\beta(L) = \beta_1 L + \cdots + \beta_p L^p$, $\beta_j \geq 0$, $j = 1, \ldots, p$; L denotes the backshift operator and the *scaled conditional errors* z_t are i.i.d.(0,1) random variables. Furthermore, it is assumed in this paper that $\alpha(1) + \beta(1) < 1$ and that all the roots of the polynomials $[1 - \alpha(L) - \beta(L)]$ and $[1 - \beta(L)]$ lie outside the unit circle. Together, these assumptions ensure that the $\{u_t\}$ process is strictly stationary and ergodic with finite variance (Bougerol & Picard, 1992; Ling & McAleer 2002). Note that, when $\mathrm{E}(u_t^2) < \infty$ but $\mathrm{E}(u_t^4) = \infty$, the distribution of the GARCH(p, q) process is said to be heavy-tailed in the sense of Mikosch & Stărică (2000, p. 1429) and Li *et al.* (2002, p. 263).

When $\alpha_i = \beta_j = 0$ for all i, j, the $\{u_t\}$ process defined in (2) reduces to i.i.d. random variables. Given $y_0 = 0$ and observations $\{y_1, \ldots, y_n\}$, the asymptotic distributions of DF tests needed to test the null hypothesis of $\phi = 1$ against $\phi < 1$ can be found in White (1958), Dickey & Fuller (1979, 1981), Phillips (1987), and Chan & Wei (1988). In addition, it has been shown that augmented-DF (ADF) tests for an AR(M) process with i.i.d. errors, $M > 1$, developed by Dickey & Fuller (1979), have the same asymptotic distributions as DF tests. See chapter 17 of Hamilton (1994) for full details.

When the u_t are GARCH-type errors with $\mathrm{E}(u_t^4) < \infty$, the asymptotic distributions of DF tests are the same as those obtained in the i.i.d. case (Pantula 1988; Ling & Li, 1998, 2003; Ling *et al.*, 2003). It is reasonable to conjecture that ADF tests should have the same asymptotic distributions as DF tests (cf. Pantula, 1986, p. 73). However, when the u_t are GARCH-type errors with $\mathrm{E}(u_t^2) < \infty$ but $\mathrm{E}(z_t^4) = \infty$, there is presently no asymptotic theory for DF tests nor for ADF tests, see Ling (2004, p. 66) and Li et al. (2002).

With this in mind, we apply the self-normalized limit theorems for square-integrable martingale difference sequences in Hall (1979, theorem 2) or Hall & Heyde (1980, theorem 4.1) to derive the asymptotic distributions of DF and ADF tests. Our purpose is to eliminate or to weaken the fourth moment condition, i.e. $\mathrm{E}(z_t^4) < \infty$. Throughout the paper, we make the following assumptions.

Assumption 1. $\mathrm{E}(z_t^2) = 1$ and $\alpha(1) + \beta(1) < 1$.

We show in this paper that, under the above assumptions, the asymptotic distributions of DF tests for AR(1)–GARCH(p, q) models and of ADF t-type tests for AR(m)–GARCH(p, q) models, $m \geq M$, are the same as those given by Dickey & Fuller (1979). It means that the DF and ADF t-type tests are nuisance parameter-free and asymptotically robust to GARCH-type heteroscedasticity, even though the fourth moment condition is not satisfied. In other words, the fourth moment condition on the scaled conditional errors z_t has no effect on the validity of the DF and ADF t-type tests and is totally unnecessary. Moreover, the asymptotic distributions of ADF tests are obtained under the condition that the coefficients of a linear process are absolutely summable. By contrast, in the context of short-memory linear processes, the coefficients are typically assumed to be one-summable or (1/2)-summable. Because we impose less restrictive assumptions on the coefficients of the linear process, it may be regarded as an improvement, and further extension, of results in Phillips & Solo (1992), Stock (1994), Phillips & Xiao (1998) and Chang & Park (2002).

The paper contains some Monte Carlo simulations illustrating the finite-sample-size properties of the tests. In all the simulations, for simplicity, we let u_t follow a GARCH(1,1) process. Overall, the simulation results indicate that, for moderately large sample sizes (e.g. $n \geq 10^5$), the tests have reasonably good size performance even though the GARCH(1,1) process is near-integrated and its fourth moment is infinity. However, our results also reveal that the size of the tests deteriorates slightly when the volatility parameter in the GARCH(1,1) process is relatively large (cf., Kim & Schmidt, 1993).

The paper is organized as follows. Section 2 presents the self-normalized version of the DF tests. Their asymptotic distributions are given in the same section. Extensions to the ADF tests for AR(m)–GARCH(p, q) models, $m \geq M$, are made in Section 3. In these two sections we give some conclusions and provide comparisons with related works. Section 4 examines the finite-sample distortions of the DF and ADF tests by means of Monte Carlo simulations. The proofs are given in Section 5.

Throughout the paper, we use the following notation: $\to_{\text{a.s.}}$, $\to_{\text{p}}$, and $\Rightarrow$ denote convergence almost surely, convergence in probability, and weak convergence of probability measures on $D[0, 1]$ under the Skorokhod topology, respectively. $W(r)$ is a standard Brownian motion, $0 \leq r \leq 1$. $O_p(1)$ ($o_p(1)$) stands for a sequence of random variables that is bounded (converges to zero) in probability. The indicator of a set A is denoted by I (A). The symbol =: means equality by definition, and $\lfloor x \rfloor$ denotes the largest integer less than or equal to x.

2. DF tests with GARCH errors

Let

$$S_n = \sum_{t=1}^{n} u_t, \qquad V_n^2 = \sum_{t=1}^{n} u_t^2.$$

The quotient S_n/V_n is the so-called self-normalized sum. Given $y_0 = 0$ and observations $\{y_1, \ldots, y_n\}$, to test the null hypothesis $\phi = 1$ against the alternative $\phi < 1$, the DF $\hat{\rho}_n$ and $\hat{\tau}_n$ tests based on the least squares (LS) regression of y_t on y_{t-1} have the self-normalized representations as follows:

$$\begin{aligned}\hat{\rho}_n &= n(\hat{\phi}_n - 1) = \left(\frac{1}{n}\sum_{t=1}^{n} y_{t-1}^2\right)^{-1}\left(\sum_{t=1}^{n} y_{t-1}u_t\right) \\ &= \frac{\frac{1}{2}[(S_n/V_n)^2 - 1]}{\frac{1}{n}\sum_{t=1}^{n-1}(S_t/V_n)^2}, \end{aligned} \tag{3}$$

$$\hat{\tau}_n = (\hat{\phi}_n - 1)\left(\frac{\sum_{t=1}^{n} y_{t-1}^2}{\hat{\sigma}_u^2}\right)^{1/2} \tag{4}$$

$$= \frac{\frac{1}{2}[(S_n/V_n)^2 - 1]}{\sqrt{\frac{1}{n}\sum_{t=1}^{n-1}(S_t/V_n)^2}} + o_p(1), \tag{5}$$

where $\hat{\phi}_n = (\sum_{t=1}^n y_{t-1}^2)^{-1} \sum_{t=1}^n y_{t-1} y_t$ is the LS estimator of ϕ in (1), and $\hat{\sigma}_u^2 = (n-1)^{-1} \sum_{t=1}^n (y_t - \hat{\phi}_n y_{t-1})^2$. Similarly, under the null hypothesis, the DF $\hat{\rho}_{cn}$ and $\hat{\tau}_{cn}$ tests based on the LS regression of y_t on y_{t-1} with a constant c are as follows:

$$\hat{\rho}_{cn} = n(\hat{\phi}_{cn} - 1) = \left[\frac{1}{n}\sum_{t=1}^{n}(y_{t-1} - \bar{y}_{-1})^2\right]^{-1}\left[\sum_{t=1}^{n}(y_{t-1} - \bar{y}_{-1})u_t\right]$$

$$= \frac{\frac{1}{2}[(S_n/V_n)^2 - 1] - \frac{1}{n}(S_n/V_n)\sum_{t=1}^{n-1} S_t/V_n}{\frac{1}{n}\left[\sum_{t=1}^{n-1}(S_t/V_n)^2 - \frac{1}{n}\left(\sum_{t=1}^{n-1} S_t/V_n\right)^2\right]}, \tag{6}$$

$$\hat{\tau}_{cn} = (\hat{\phi}_{cn} - 1)\left[\frac{\sum_{t=1}^{n}(y_{t-1} - \bar{y}_{-1})^2}{\hat{\sigma}_{cu}^2}\right]^{1/2} \tag{7}$$

$$= \frac{\frac{1}{2}[(S_n/V_n)^2 - 1] - \frac{1}{n}(S_n/V_n)\sum_{t=1}^{n-1} S_t/V_n}{\sqrt{\frac{1}{n}\left[\sum_{t=1}^{n-1}(S_t/V_n)^2 - \frac{1}{n}\left(\sum_{t=1}^{n-1} S_t/V_n\right)^2\right]}} + o_p(1), \tag{8}$$

where $\hat{\sigma}_{cu}^2 = (n-2)^{-1} \sum_{t=1}^n (y_t - \hat{c}_n - \hat{\phi}_{cn} y_{t-1})^2$, $\bar{y}_{-1} = n^{-1} \sum_{t=1}^n y_{t-1}$, and $\hat{\phi}_{cn}$ and $\hat{c}_n$ are the LS estimators of ϕ and c, respectively. For more details on (3), (5), (6) and (8), see Wang (2006). Clearly, if the sequence $\{S_t/V_n, 1 \le t \le n\}$ has an asymptotic distribution, then so do the DF tests, $\hat{\rho}_n$, $\hat{\tau}_n$, $\hat{\rho}_{cn}$ and $\hat{\tau}_{cn}$.

We are now ready to state one of our main results.

Theorem 1. *Let $\{y_t\}$ and $\{u_t\}$ be generated according to (1)–(2). Suppose that Assumption 1 holds. Then, as $n \to \infty$,*

$$\frac{S_{\lfloor nr \rfloor}}{V_n} \Rightarrow W(r), \quad 0 \le r \le 1, \tag{9}$$

and

(a)

$$\hat{\rho}_n \Rightarrow \frac{1}{2}\frac{W^2(1) - 1}{\int_0^1 W^2(r)\,\mathrm{d}r};$$

(b)

$$\hat{\tau}_n \Rightarrow \frac{1}{2}\frac{W^2(1) - 1}{\left[\int_0^1 W^2(r)\,\mathrm{d}r\right]^{1/2}};$$

(c)

$$\hat{\rho}_{cn} \Rightarrow \frac{\frac{1}{2}\left[W^2(1) - 1\right] - W(1)\int_0^1 W(r)\,\mathrm{d}r}{\int_0^1 W^2(r)\,\mathrm{d}r - \left[\int_0^1 W(r)\,\mathrm{d}r\right]^2};$$

(d)

$$\hat{\tau}_{cn} \Rightarrow \frac{\frac{1}{2}\left[W^2(1)-1\right] - W(1)\int_0^1 W(r)\,\mathrm{d}r}{\left\{\int_0^1 W^2(r)\,\mathrm{d}r - \left[\int_0^1 W(r)\,\mathrm{d}r\right]^2\right\}^{1/2}}.$$

Remark 1. The study by Ling *et al.* (2003) contains a recent development in estimation and testing for unit root processes with GARCH(1,1) errors, especially the maximum likelihood estimation. However, there is an error in the cited paper on page 185. The error is that the DF t-tests should be defined by $\hat{\tau}_n$ and $\hat{\tau}_{cn}$ (see (4) and (7)), not by

$$L_t = \left(\sum_{t=1}^{n} y_{t-1}^2\right)^{1/2} (\hat{\phi}_n - 1) \text{ and } L_{\mu,t} = \left[\sum_{t=1}^{n} (y_{t-1} - \bar{y}_{-1})^2\right]^{1/2} (\hat{\phi}_{cn} - 1).$$

Remark 2. Pantula (1986, p. 73) posed a conjecture that DF tests are asymptotically valid for GARCH(p, q) errors except in the integrated variance case. Theorem 1 affirms the conjecture. Consequently, as shown by simulations in Kim & Schmidt (1993), the effects of GARCH errors in DF tests are a small-sample problem.

3. ADF tests with GARCH errors

In this section, we will establish asymptotic results for ADF tests for a unit root process with the errors being dependent and heavy-tailed. The ADF tests were derived by Dickey & Fuller (1979), and are often used to control for serial correlation by adding higher-order autoregressive terms in the regression. In the light of Theorem 1, it is natural to ask whether ADF tests could also hold under the same weaker assumptions. As this section will show, this is indeed the case.

Throughout this section, we assume that the time series y_t is generated by an AR(M) process,

$$(1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_M L^M) y_t = u_t, \tag{10}$$

where $\phi_1 + \cdots + \phi_M = 1$, $1 < M < \infty$, and the u_t are defined as in (2). As in Hamilton (1994, p. 518), under the null hypothesis that $\phi_1 + \cdots + \phi_M = 1$, the $\{y_t\}$ process can be written as

$$\Delta y_t = (1 - \xi_1 L - \cdots - \xi_{M-1} L^{M-1})^{-1} u_t$$

or

$$\Delta y_t = \varepsilon_t = \sum_{j=0}^{\infty} \theta_j u_{t-j}, \tag{11}$$

where $\xi_j = -(\phi_{j+1} + \cdots + \phi_M)$, $j = 1, 2, \ldots, M-1$. Let $\theta(1) = \sum_{j=0}^{\infty} \theta_j$. Assume that all roots of $(1 - \xi_1 z - \cdots - \xi_{M-1} z^{M-1}) = 0$ lie outside the unit circle and, without loss of generality, that $\theta(1) \neq 0$. Because the moving average coefficients θ_j in (11) decay geometrically, we say that the $\{\Delta y_t\}$ or $\{\varepsilon_t\}$ process is a short-memory linear process (cf. Hall, 1992, p. 118).

We now consider the following AR(m) regression model:

$$y_t = c + \phi_1 y_{t-1} + \cdots + \phi_m y_{t-m} + u_t,$$

which can equivalently be written as

$$\begin{aligned}\Delta y_t &= c + (\phi - 1)y_{t-1} + \xi_1 \Delta y_{t-1} + \cdots + \xi_{m-1}\Delta y_{t-(m-1)} + u_t \\ &= (\phi - 1)y_{t-1} + \boldsymbol{\Psi} \boldsymbol{Z}_t + u_t, \end{aligned} \tag{12}$$

where $M \le m < \infty$, $\phi = \phi_1 + \cdots + \phi_M + \cdots + \phi_m$, $\xi_j = -(\phi_{j+1} + \cdots + \phi_m)$, $j = 1, 2, \ldots, m-1$, $\boldsymbol{\Psi} = (c, \xi_1, \ldots, \xi_{m-1})$, and $\boldsymbol{Z}_t = (1, \Delta y_{t-1}, \ldots, \Delta y_{t-(m-1)})^\top$. Under the null hypothesis $\phi_1 + \cdots + \phi_M = 1$, we have $c = 0$, $\phi_{M+i} = 0$ for $i > 0$, $\phi = \phi_1 + \cdots + \phi_M + \cdots + \phi_m = \phi_1 + \cdots + \phi_M = 1$ and $\xi_i = 0$ for $i \ge M$.

Before applying the Frisch–Waugh–Lovell (FWL) theorem (see Davidson & MacKinnon, 1993, p. 19), we first introduce some notation. Let

$$R_{0t} = \Delta y_t - \left(\sum_{t=1}^{n} \Delta y_t \boldsymbol{Z}_t^\top\right)\left(\sum_{t=1}^{n} \boldsymbol{Z}_t \boldsymbol{Z}_t^\top\right)^{-1} \boldsymbol{Z}_t, \tag{13}$$

$$R_{1t} = y_{t-1} - \left(\sum_{t=1}^{n} y_{t-1} \boldsymbol{Z}_t^\top\right)\left(\sum_{t=1}^{n} \boldsymbol{Z}_t \boldsymbol{Z}_t^\top\right)^{-1} \boldsymbol{Z}_t \tag{14}$$

be the residuals obtained by regressing Δy_t and y_{t-1} on $\boldsymbol{Z}_t$, and let $\hat{\boldsymbol{\Psi}}$ be the LS estimator of $\boldsymbol{\Psi}$ in (12). Then regression (12) can be written as a regression equation in the residuals as follows:

$$R_{0t} = (\phi - 1)R_{1t} + \tilde{u}_t, \tag{15}$$

where $\tilde{u}_t = u_t + (\boldsymbol{\Psi} - \hat{\boldsymbol{\Psi}})\boldsymbol{Z}_t$ (cf. Johansen, 1995, p. 91). The FWL theorem states that the regressions (12) and (15) have the same residuals, and that the estimate of $(\phi - 1)$ is the same from the two regressions. As our main focus is on testing $\phi - 1 = 0$, in order to assist in deriving the asymptotic distributions of ADF tests we restrict our attention to (15).

Given $y_{-(m-1)} = \cdots = y_0 = 0$ and observations $\{y_1, \ldots, y_n\}$, when testing $\phi = 1$ against $\phi < 1$ the ADF ρ-type and t-type test statistics based on the LS regression of R_{0t} on R_{1t} are as follows:

$$ADF_\rho = n(\hat{\phi}_{A,n} - 1) = \frac{\sum_{t=1}^{n} R_{0t} R_{1t}}{n^{-1}\sum_{t=1}^{n} R_{1t}^2}, \tag{16}$$

$$ADF_\tau = (\hat{\phi}_{A,n} - 1)\left(\frac{\sum_{t=1}^{n} R_{1t}^2}{\hat{\sigma}_R^2}\right)^{1/2} = \frac{\sum_{t=1}^{n} R_{0t} R_{1t}}{\left(\sum_{t=1}^{n} R_{1t}^2 \hat{\sigma}_R^2\right)^{1/2}}, \tag{17}$$

where $(\hat{\phi}_{A,n} - 1) = (\sum_{t=1}^{n} R_{1t}^2)^{-1} \sum_{t=1}^{n} R_{0t} R_{1t}$ is the LS estimator of $(\phi - 1)$ in (15), and $\hat{\sigma}_R^2 = (n-1)^{-1}\sum_{t=1}^{n}[R_{0t} - (\hat{\phi}_{A,n} - 1)R_{1t}]^2$.

We are now in a position to present another main result of this paper.

Theorem 2. *Let $\{y_t\}$ and $\{u_t\}$ be generated according to (10) and (2), respectively. Suppose that Assumption 1 holds. If the regression (15) is estimated by ordinary LS, then, as $n \to \infty$,*

(a)

$$ADF_\rho \Rightarrow \frac{\frac{1}{2}\left[W^2(1)-1\right] - W(1)\int_0^1 W(r)\mathrm{d}r}{\theta(1)\left\{\int_0^1 W^2(r)\mathrm{d}r - \left[\int_0^1 W(r)\mathrm{d}r\right]^2\right\}};$$

(b)

$$ADF_\tau \Rightarrow \frac{\frac{1}{2}\left[W^2(1)-1\right] - W(1)\int_0^1 W(r)\mathrm{d}r}{\left\{\int_0^1 W^2(r)\mathrm{d}r - \left[\int_0^1 W(r)\mathrm{d}r\right]^2\right\}^{1/2}}.$$

Remark 3. We note again that, under Assumption 1, the GARCH errors given in (2) are allowed to have a heavy-tailed distribution. To the best of our knowledge, it has not yet been shown rigorously in the literature that the ADF t-type test for time series with weakly dependent, heteroscedastic and heavy-tailed errors has the same asymptotic distribution as the corresponding DF test for the data generating process given in (1)–(2). In addition, in the context of short-memory linear processes, the coefficients θ_j are typically restricted to be one-summable (i.e., $\sum_{j=0}^{\infty} j|\theta_j| < \infty$) or (1/2)-summable (i.e., $\sum_{j=0}^{\infty} j^{1/2}|\theta_j| < \infty$) for validating the functional central limit theorem (cf. Phillips & Solo, 1992; Phillips & Xiao, 1998; Chang & Park, 2002). These conditions are, however, weakened in Theorem 2.

Remark 4. A work by Hansen & Rahbek (1998), which was brought to our attention by a reviewer, contains material related to the present paper. These authors used an operational drift criterion from Markov chain theory to show that both the strong law of large numbers and the functional central limit theorem hold for a simple multivariate ARCH(1) process. Based on the obtained results, they showed that the procedure developed, for example in Johansen (1995) for cointegration analysis of vector autoregressive models, is robust to the errors. If their results are true, then, as DF and ADF tests are special cases of multivariate cointegration tests, the results given in Theorems 1 and 2 seem to be a consequence of the results given by Hansen & Rahbek (1998). However, the asymptotic distributions of DF tests in Theorem 1 are derived using the self-normalized representations of the tests given in (3), (5), (6) and (8) first and then by applying a self-normalized invariance principle found in Hall (1979, theorem 2) or Hall & Heyde (1980, theorem 4.1). The proofs in Section 5 below clearly show that the fourth moment condition on the scaled conditional errors of the GARCH process is totally unnecessary. We note in passing that the above-mentioned theorem 2 or theorem 4.1 is a self-normalized limit theorem for martingale differences. These theorems are a little different from the martingale functional central limit theorem in Brown (1971, theorem 3) or Hall & Heyde (1980, theorem 4.4). Based on the self-normalized representations and the limit theorem, our approach has the advantage that the results in Theorem 1 come very easily. On the other hand, the method adopted for obtaining Theorem 2 is based only on simple algebraic manipulations on the coefficients θ_j of the short-memory linear process ε_t in (11), together with the asymptotic distribution of self-normalized sums in (9). It seems to be quite simple at a technical level. Moreover, the role of the summable coefficients of the linear process is easily understood, and the coefficient condition required in Theorem 2 is weak, as noted in Remark 3. For these reasons, our approach can be quite useful for pedagogical purposes.

Remark 5. In the framework of Phillips (1987), the underlying errors in a unit root process are allowed to be serially dependent and heteroscedastic, as we permit in (10) or (11). Under the unit root hypothesis, we can see from (11) that

$$\frac{1}{n}\sum_{t=1}^{n} y_{t-1}\varepsilon_t = \frac{1}{2}\left[\frac{\left(\sum_{t=1}^{n}\varepsilon_t\right)^2}{n} - \frac{\sum_{t=1}^{n}\varepsilon_t^2}{n}\right]$$

(cf. Hamilton, 1994, p. 476). Phillips (1987, p. 296) applied an invariance principle established by Herrndorf (1984, corollary 1), together with the continuous mapping theorem, to derive the asymptotic distribution of $(\sum_{t=1}^{n}\varepsilon_t)^2/n$ under a sufficient condition that $\sup_t \mathrm{E}|\varepsilon_t|^{\gamma} < \infty$ for some $\gamma > 2$. He showed, by applying the strong law of McLeish (1975, Theorem 2.10), that $\sum_{t=1}^{n}\varepsilon_t^2/n \to_{a.s.} \lim_{n\to\infty} n^{-1}\sum_{t=1}^{n}\mathrm{E}(\varepsilon_t^2)$ under a sufficient condition that $\sup_t \mathrm{E}|\varepsilon_t|^{\gamma+\delta} < \infty$ for some $\gamma > 2$ and any $\delta > 0$. Since the moment conditions imposed in the present paper are slightly weaker than the above, and since mixing is a more restrictive assumption than ergodicity (see Stout, 1974, Theorem 3.5.4, p. 173), Theorem 2 can be considered as a slight extension of theorem 3.1 of Phillips (1987).

Remark 6. It is well known that the Lindeberg condition is a sufficient (as well as necessary) condition for the central limit theorem for independent random variables (see, for example, Chung, 2001, theorem 7.2.1). More precisely, let u_t be independent random variables with mean zero and finite variances, and set $s_n^2 = \mathrm{E}(\mathrm{S}_n^2) = \mathrm{E}(V_n^2)$. Then the Lindeberg condition is as follows:

$$\frac{1}{s_n^2}\sum_{t=1}^{n}\mathrm{E}\left[u_t^2\mathrm{I}(|u_t| > \epsilon s_n)\right] \to 0 \;\text{ for all }\; \epsilon > 0 \;\text{ and as }\; n \to \infty$$

(see also Lemma 1(b) in Section 5 below). As is usual, when the u_t are independent but not identically distributed, only the case where all variances are finite is not sufficient for the Lindeberg condition to hold. Thus, a sufficient condition for the Lindeberg condition would be Liapounov's condition, i.e.,

$$\frac{1}{s_n^{2+\delta}}\sum_{t=1}^{n}\mathrm{E}\,|u_t|^{2+\delta} \to 0 \;\text{ for some }\; \delta > 0 \;\text{ and as }\; n \to \infty$$

(see Chung, 2001, p. 219). It is stronger than the Lindeberg condition, but is easily checkable and frequently used in practice. Similarly, when $\{u_t\}$ is a zero-mean, square-integrable martingale difference sequence, the Lindeberg condition is sufficient for the martingale central limit theorem and the invariance principle; see Brown (1971, theorem 3), Hall (1979, theorem 2), Hall & Heyde (1980, theorems 4.1 and 4.4) and Billingsley (1995, theorem 35.12). However, as in the independent case, the assumption of finite variances is not sufficient to imply the Lindeberg condition (cf., Billingsley, 1995, pp. 362–363). If the assumption is sufficient, the Lindeberg condition required in the above theorems will be redundant. As a result, it seems that the statement in Hansen & Rahbek (1998, p. 10, lines 11–12) may not be rigorous. Moreover, the proof in Hansen & Rahbek (1998, p. 10, lines 17–19) also looks suspicious, because one can find a similar proof in Hamilton (1994, example 17.2, p. 482), where the finite fourth moment condition is imposed. A further paper related to ours is Pantula (1988). He applied the invariance principle in Hall & Heyde (1980, theorem 4.4) to derive the asymptotic distributions of unit root tests for AR–ARCH processes under the fourth moment condition on the ARCH errors. Similar to Liapounov's condition, a finite fourth moment is sufficient for the Lindeberg condition to hold. In contrast to Pantula (1988), our proof of the Lindeberg condition in Lemma 1(b) below is based on a generalization of Kolmogorov's inequality given in Chung (2001, corollary 1, p. 347). The inequality requires only the existence of the second moment.

4. Monte Carlo simulations

We have shown that the four DF-type tests and the ADF t-type test have standard asymptotic distributions in the preceding sections. Here we try to find out how large the sample size n should be so that the asymptotic results become approximately true. We examine the size (percentages of rejections under the null hypothesis) distortions of the tests in finite samples by means of Monte Carlo simulations. For the well-established empirical relevance, we let u_t follow the simple, but important, GARCH(1,1) process.

The data generating process for the DF tests was the following AR(1)–GARCH(1,1) process:

$$y_t = y_{t-1} + u_t, \qquad u_t = z_t\sigma_t, \qquad \sigma_t^2 = w + \alpha u_{t-1}^2 + \beta\sigma_{t-1}^2,$$

where $w > 0, \alpha + \beta < 1, z_t = T_{4,t}/\sqrt{2}$ and $T_{4,t}$ are i.i.d. Student-t random variables with four degrees of freedom. It is obvious that $\mathrm{E}(z_t) = 0$, $\mathrm{E}(z_t^2) = 1$ and $\mathrm{E}(z_t^4) = \infty$. The process also implies that $\mathrm{E}(u_t^2) = w/(1-\alpha-\beta) < \infty$ but $\mathrm{E}(u_t^4) = \infty$. Similarly to Kim & Schmidt (1993), without loss of generality we set $y_0 = 0$, $w = 0.05$ and $\sigma_0^2 = w/(1-\alpha-\beta)$. In this experiment, only three parameters, namely α, β and n, need to be specified. The DF $\hat{\rho}_n$, $\hat{\tau}_n$, $\hat{\rho}_{cn}$ and $\hat{\tau}_{cn}$ tests in (3), (5), (6) and (8) were then performed on the generated data. When the sample size n went to infinity (was equal to 100), the corresponding critical values of the tests (lower-tail tests) at the nominal 5% level were -8.1, -1.95, -14.1 and -2.86 (-7.9, -1.95, -13.7 and -2.89), respectively (see, for example, Banerjee et al., 1993, pp. 102–103).

In the Monte Carlo experiment for the ADF tests, for simplicity we chose only the ADF t-type test, i.e., ADF_τ, because its asymptotic distribution is nuisance parameter-free and the same as that of $\hat{\tau}_{cn}$; see Theorem 1(d) and Theorem 2(b). This means that the critical values of -2.86 and -2.89 could be applied directly. Specifically, the data generating process for ADF_τ was assumed to be the following AR(4)–GARCH(1,1) process:

$$y_t = 0.1y_{t-1} + 0.2y_{t-2} + 0.3y_{t-3} + 0.4y_{t-4} + u_t,$$

where the u_t were defined as in the above experiment. Similarly, we set $y_{-3} = \cdots = y_0 = 0$, $w = 0.05$ and $\sigma_0^2 = w/(1-\alpha-\beta)$. There are also only three parameters, namely α, β and n, that need to be specified in this experiment. Because the data were generated by setting $M = 4$, without loss of generality we set $m = 8$ for the regression model in (12). Based on regression (15), the ADF t-type test, ADF_τ, in (17) was performed on the generated data.

In the two experiments, we created $n + 50$ observations and discarded the first 50 observations to reduce the effect of the initial conditions. Samples of size $n = 10^2, 10^3, 10^4$ and 10^5 were used in the experiments, respectively. These large values of n can arise with high-frequency financial data, and, more importantly, they are provided to show the validity of the asymptotic distributions in Theorems 1 and 2.

Tables 1 and 2 give the proportion of rejections under the unit root null hypothesis for a 5% lower-tail test. The results reported in the tables were based on 10 000 replications for each simulation for various values of α, β and n. All simulations were performed using the econometric package EViews 3.1. (1999).

TABLE 1.

Finite-sample size properties of the DF and ADF tests for a 5% nominal level.

(α, β)	n	$\hat{\rho}_n$	$\hat{\tau}_n$	$\hat{\rho}_{cn}$	$\hat{\tau}_{cn}$	ADF_τ
(0.00, 0.00)	10^2	0.0497	0.0486	0.0453	0.0481	0.0573
	10^3	0.0485	0.0495	0.0473	0.0491	0.0489
	10^4	0.0519	0.0509	0.0502	0.0526	0.0533
	10^5	0.0502	0.0494	0.0520	0.0514	0.0512
(0.05, 0.90)	10^2	0.0562	0.0536	0.0526	0.0585	0.0667
	10^3	0.0517	0.0517	0.0550	0.0579	0.0572
	10^4	0.0481	0.0485	0.0509	0.0525	0.0545
	10^5	0.0512	0.0513	0.0527	0.0515	0.0505
(0.10, 0.85)	10^2	0.0602	0.0601	0.0619	0.0651	0.0737
	10^3	0.0568	0.0573	0.0622	0.0658	0.0652
	10^4	0.0486	0.0493	0.0540	0.0549	0.0573
	10^5	0.0504	0.0506	0.0524	0.0504	0.0511
(0.15, 0.80)	10^2	0.0638	0.0625	0.0692	0.0741	0.0766
	10^3	0.0605	0.0606	0.0695	0.0722	0.0689
	10^4	0.0531	0.0521	0.0588	0.0587	0.0586
	10^5	0.0500	0.0493	0.0530	0.0545	0.0534
(0.20, 0.75)	10^2	0.0684	0.0691	0.0746	0.0801	0.0793
	10^3	0.0616	0.0613	0.0766	0.0781	0.0697
	10^4	0.0540	0.0530	0.0622	0.0611	0.0595
	10^5	0.0512	0.0512	0.0564	0.0563	0.0541
(0.25, 0.70)	10^2	0.0739	0.0722	0.0796	0.0865	0.0797
	10^3	0.0658	0.0636	0.0804	0.0821	0.0711
	10^4	0.0558	0.0547	0.0665	0.0630	0.0623
	10^5	0.0517	0.0529	0.0630	0.0645	0.0574
(0.30, 0.65)	10^2	0.0763	0.0745	0.0861	0.0913	0.0805
	10^3	0.0676	0.0666	0.0853	0.0848	0.0715
	10^4	0.0577	0.0563	0.0680	0.0673	0.0645
	10^5	0.0554	0.0537	0.0632	0.0617	0.0582
(0.35, 0.60)	10^2	0.0778	0.0771	0.0900	0.0938	0.0803
	10^3	0.0679	0.0682	0.0871	0.0894	0.0736
	10^4	0.0607	0.0605	0.0703	0.0714	0.0650
	10^5	0.0573	0.0572	0.0640	0.0629	0.0581

Notes. The data generating process (DGP) for the DF tests in columns 3–6 is $y_t = y_{t-1} + u_t$, where $u_t = z_t\sigma_t = z_t(w + \alpha\, u_{t-1}^2 + \beta\, \sigma_{t-1}^2)^{1/2}$, $w = 0.05$, $\sigma_0^2 = w/(1-\alpha-\beta)$ and z_t are i.i.d. Student-t random variables with four degrees of freedom. The DGP for the ADF t-type test in column 7 is $y_t = 0.1\ y_{t-1} + 0.2\ y_{t-2} + 0.3\ y_{t-3} + 0.4\ y_{t-4} + u_t$, where u_t are defined as above. The ADF t-type test is based on an AR(8) regression model in Equation (12). The test statistics $\hat{\rho}_n$, $\hat{\tau}_n$, $\hat{\rho}_{cn}$, $\hat{\tau}_{cn}$ and ADF_τ are defined in Equations (3), (4), (6), (7) and (17), respectively.

TABLE 2.
Finite-sample size properties of the DF and ADF tests for a 5% nominal level.

(α, β)	$\hat{\rho}_n$	$\hat{\tau}_n$	$\hat{\rho}_{cn}$	$\hat{\tau}_{cn}$	ADF_τ
(0.05, 0.91)	0.0514	0.0508	0.0546	0.0515	0.0517
(0.05, 0.92)	0.0541	0.0521	0.0539	0.0530	0.0528
(0.05, 0.93)	0.0532	0.0529	0.0534	0.0525	0.0524
(0.05, 0.94)	0.0551	0.0535	0.0527	0.0551	0.0548
(0.10, 0.86)	0.0503	0.0502	0.0525	0.0524	0.0511
(0.10, 0.87)	0.0524	0.0513	0.0526	0.0529	0.0529
(0.10, 0.88)	0.0498	0.0497	0.0587	0.0571	0.0566
(0.10, 0.89)	0.0558	0.0553	0.0636	0.0646	0.0638
(0.15, 0.81)	0.0518	0.0508	0.0544	0.0562	0.0554
(0.15, 0.82)	0.0518	0.0510	0.0572	0.0578	0.0570
(0.15, 0.83)	0.0526	0.0517	0.0619	0.0634	0.0609
(0.15, 0.84)	0.0590	0.0573	0.0698	0.0717	0.0704

Note. $n = 10^5$. See notes to Table 1.

The first section in Table 1 reports results for i.i.d. samples (no conditional heteroscedasticity, i.e. $\alpha = 0$ and $\beta = 0$). It is proposed as a benchmark for comparison. The other results in Table 1 were obtained by varying n, α and β, for fixed $\alpha + \beta = 0.95$. As expected from Theorems 1 and 2, the proportion of rejections tends steadily to 0.05 as n increases. It turns out that the DF and ADF t-type tests are actually robust to general GARCH errors in larger sample sizes, even if the fourth moment of the errors goes to infinity. However, it can be seen from Table 1 that the size of the tests deteriorates slightly when the volatility parameter α is relatively large (cf. Kim & Schmidt, 1993). Note, however, that the degree of over-rejection is slight when the sample size n is greater than or equal to 10^5.

The results in Table 2 were obtained by varying α and β, for $0.95 < \alpha + \beta < 1$ and for fixed $n = 10^5$. Note that the volatility parameter α in the table was chosen to be in the range often observed in empirical studies; see Kim & Schmidt (1993, p. 288) and references therein. The main purpose of the table is to answer the following question: for such a large sample size, do the DF and ADF t-type tests suffer from serious size distortions when $\alpha + \beta$ approaches 1 (i.e. the GARCH(1,1) process is near-integrated)? It is readily seen from Table 2 that, when α is relatively small, the proportion of rejections is not too far away from 0.05 as $\alpha + \beta$ approaches 1. This means that, when the volatility parameter α is not too large and the sample size is large enough, the size of the tests deteriorates slightly even if the GARCH process is near-integrated and its fourth moment is infinity. This result is consistent with the finding of Kim & Schmidt (1993, p. 295) that near-integrated GARCH errors do not lead to a severe over-rejection problem for the tests.

In summary, the simulation results suggest that, if the GARCH process is near-integrated or the volatility parameter in the process is relatively large (or both), then smaller size distortions of the DF and ADF t-type tests require larger sample sizes, such as $n \geq 10^5$.

5. Proofs

In this section, we will prove Theorem 1 of Section 2 and Theorem 2 of Section 3. The proofs are based on the self-normalized limit theorem for martingale difference sequences in Hall (1979, theorem 2) or Hall & Heyde (1980, theorem 4.1, p. 99); see also Phillips & Solo

(1992, theorem 2.6). Before applying the theorem, we need the following lemma. Let u_t be as in (2) and set $s_n^2 = \mathrm{E}(\mathrm{S}_\mathrm{n}^2) = \mathrm{E}(V_n^2)$.

Lemma 1. *Let $\{u_t\}$ be a GARCH(p,q) process defined by (2). Then under Assumption 1, we have*

(a)

$$\frac{V_n^2}{s_n^2} \to_{a.s.} 1;$$

(b)

$$\frac{1}{s_n^2}\sum_{t=1}^{n} \mathrm{E}[u_t^2 \mathrm{I}(|u_t| > \epsilon s_n)] \to 0 \quad \text{for all } \epsilon > 0 \text{ and as } n \to \infty.$$

Proof of Lemma 1. For part (a), let $\eta_t = u_t^2 - \sigma_t^2 = \sigma_t^2(z_t^2 - 1)$. It follows from (2) that

$$[1 - \alpha(L) - \beta(L)]u_t^2 = w + [1 - \beta(L)]\eta_t. \tag{18}$$

Note that $\sigma_u^2 = \mathrm{E}(u_\mathrm{t}^2) = \mathrm{w}/[1\text{-}\alpha(1) - \beta(1)]$. Since all the roots of $[1 - \alpha(L) - \beta(L)]$ and $[1 - \beta(L)]$ lie outside the unit circle, we can rewrite (18) as

$$u_t^2 = \sigma_u^2 + \frac{1 - \beta(L)}{1 - \alpha(L) - \beta(L)}\eta_t =: \sigma_u^2 + \sum_{i=0}^{\infty} \pi_i \eta_{t-i}, \tag{19}$$

where the coefficients π_i are absolutely summable (cf. Brockwell & Davis, 1991, theorem 3.1.1). By (19), a simple calculation gives

$$V_n^2 = n\sigma_u^2 + \sum_{t=1}^{n}\sum_{i=0}^{\infty} \pi_i \eta_{t-i} = n\sigma_u^2 + \sum_{i=0}^{\infty} \pi_i \sum_{t=1}^{n} \eta_{t-i} =: n\sigma_u^2 + M_n. \tag{20}$$

Then, by (20) we have $s_n^2 = n\sigma_u^2$, since $\mathrm{E}(\eta_{\mathrm{t-i}}) = 0$ for $i \geq 0$ and the sequence $\{\pi_i\}$ is absolutely summable.

Now all we need to show is that $M_n/n \to_{\mathrm{a.s.}} 0$ as $n \to \infty$. It is well known that, when $w > 0$, $\alpha(1) + \beta(1) < 1$ and $\mathrm{E}(z_t^2) < \infty$, $\{u_t, \sigma_t^2\}$ is strictly stationary and ergodic (Bougerol & Picard, 1992). By theorem 3.5.8 in Stout (1974, p. 182), $\{\eta_t\}$ is stationary ergodic. On the other hand, it follows from Assumption 1 that $\mathrm{E}|z_t^2| < \infty$, and then from Chung (2001, p. 51, exercise 10) that $\mathrm{E}|z_t^2 - a| < \infty$ for every a. Since z_t is independent of σ_t^2, it is easy to see that

$$\mathrm{E}|\eta_t| = \mathrm{E}|\sigma_t^2(z_t^2 - 1)| = \mathrm{E}|\sigma_t^2|\, \mathrm{E}|z_t^2 - 1| < \infty,$$

provided that $\mathrm{E}|\sigma_t^2| = \mathrm{E}(\sigma_t^2) = \sigma_u^2 < \infty$. Now, applying theorem 3.5.7 in Stout (1974, p. 181), we obtain that, for all $i \geq 0$, $\sum_{t=1}^{n} \eta_{t-i}/n \to_{a.s.} \mathrm{E}(\eta_t) = 0$ as $n \to \infty$. This gives that $M_n/n \to_{a.s.} \mathrm{E}(\eta_t)\sum_{i=0}^{\infty} \pi_i = 0$, since, again, $\{\pi_i\}$ is absolutely summable. Hence, Lemma 1(a) holds.

For part (b), Hall (1979, p. 372) showed that, if $\{V_n^2/s_n^2\}$ is uniformly integrable, Lemma 1(b) is equivalent to

$$\max_{1\le t\le n} |u_t/s_n| \to_p 0 \tag{21}$$

(see also Phillips & Solo, 1992, theorem 2.6). So, with Lemma 1(a), we need only show that (21) holds. Since $s_n^2 \to \infty$ as $n \to \infty$ and $\{u_t\}$ is a strictly stationary and ergodic martingale difference sequence with $\mathrm{E}(u_t^2) < \infty$ for all t, by corollary 1 in Chung (2001, p. 347) we have, for all $\epsilon > 0$,

$$\lim_{n\to\infty} \Pr\left(\max_{1\le t\le n} |u_t| \ge \epsilon s_n\right) \le \lim_{n\to\infty} \frac{\mathrm{E}(u_t^2)}{\epsilon^2 s_n^2} = 0,$$

which implies condition (21). The proof of Lemma 1 is complete.

Proof of Theorem 1. By theorem 4.1 in Hall & Heyde (1980, p. 99), the invariance principle (9) follows directly from Lemma 1. The results of parts (a)–(d) follow from (3), (5), (6) and (8), together with the continuous mapping theorem (Billingsley, 1968). The proof of Theorem 1 is complete.

Remark 7. It is clear from the proof above that Theorem 1 remains true even if $\mathrm{E}(\mathrm{z}_\mathrm{t}^{2+\delta}) = \infty$ for all $\delta > 0$.

Next, to prove Theorem 2 of Section 3, we use the following notation and lemmas. Recall from (11) that $\Delta y_t = \varepsilon_t = \sum_{j=0}^{\infty} \theta_j u_{t-j}$ and $\theta(1) = \sum_{j=0}^{\infty} \theta_j \ne 0$. Assume without loss of generality that $\sum_{j=0}^{\infty} |\theta_j| < \infty$. Let $S_{\varepsilon,n} = \sum_{t=1}^{n} \varepsilon_t$, and let $S_{j,n} = \sum_{t=1}^{n} u_{t-j}$ and $V_{j,n}^2 = \sum_{t=1}^{n} u_{t-j}^2$, $j \ge 0$. In particular, put $S_{0,n} = S_n$ and $V_{0,n}^2 = V_n^2$.

Lemma 2. *Let $\{y_t\}$ and $\{u_t\}$ be generated according to (10) and (2), respectively. Suppose that Assumption 1 holds. Then, as $n \to \infty$,*

$$\frac{S_{\varepsilon,\lfloor nr\rfloor}}{V_n} \Rightarrow \theta(1)W(r), \quad 0 \le r \le 1.$$

Proof of Lemma 2. Some elementary algebra gives

$$\begin{aligned}\frac{S_{\varepsilon,\lfloor nr\rfloor}}{V_n} &= \frac{\sum_{t=1}^{\lfloor nr\rfloor}\sum_{j=0}^{\infty}\theta_j u_{t-j}}{V_n} = \frac{\sum_{t=1}^{\lfloor nr\rfloor}\sum_{j=0}^{\infty}\theta_j (u_t + u_{t-j} - u_t)}{V_n}\\ &= \sum_{j=0}^{\infty}\theta_j \frac{S_{\lfloor nr\rfloor}}{V_n} + \sum_{j=0}^{n} \frac{\theta_j (S_{j,\lfloor nr\rfloor} - S_{\lfloor nr\rfloor})}{V_n} + \sum_{j=n+1}^{\infty} \frac{\theta_j (S_{j,\lfloor nr\rfloor} - S_{\lfloor nr\rfloor})}{V_n}.\end{aligned} \tag{22}$$

It follows from Theorem 1 that

$$\frac{S_{j,\lfloor nr\rfloor} - S_{\lfloor nr\rfloor}}{V_n} = \frac{S_{j,\lfloor nr\rfloor}}{V_{j,n}}\frac{V_{j,n}}{V_n} - \frac{S_{\lfloor nr\rfloor}}{V_n} \tag{23}$$

is $O_p(1)$. Since the coefficients θ_j are absolutely summable (cf. Brockwell & Davis, 1991, theorem 3.1.1), it is necessary to have $\sum_{j=n+1}^{\infty} \theta_j \to 0$ as $n \to \infty$. By this and by (23), the third term on the right-hand side of (22) is obviously $O_p(1)$. Note that, for $1 \le j \le n$,

$$S_{j,\lfloor nr \rfloor} - S_{j-1,\lfloor nr \rfloor} = \sum_{t=1}^{\lfloor nr \rfloor} [u_{t-j} - u_{t-(j-1)}] = u_{1-j} - u_{\lfloor nr \rfloor-(j-1)}. \tag{24}$$

With this, the second term in (22) can be written as

$$\begin{aligned}
&\sum_{j=0}^{n} \frac{\theta_j \big(S_{j,\lfloor nr \rfloor} - S_{\lfloor nr \rfloor}\big)}{V_n} \\
&= \sum_{j=1}^{n} \frac{\theta_j}{V_n} \left\{ \big[S_{j,\lfloor nr \rfloor} - S_{j-1,\lfloor nr \rfloor}\big] + \big[S_{j-1,\lfloor nr \rfloor} - S_{j-2,\lfloor nr \rfloor}\big] + \cdots + \big[S_{1,\lfloor nr \rfloor} - S_{\lfloor nr \rfloor}\big] \right\} \\
&= \sum_{j=1}^{n} \frac{\theta_j}{V_n} \left\{ \big[u_{1-j} - u_{\lfloor nr \rfloor-(j-1)}\big] + \big[u_{1-(j-1)} - u_{\lfloor nr \rfloor-(j-2)}\big] + \cdots + \big[u_0 - u_{\lfloor nr \rfloor}\big] \right\} \\
&= \frac{1}{V_n} \left\{ (\theta_1 + \cdots + \theta_n)\big[u_0 - u_{\lfloor nr \rfloor}\big] + (\theta_2 + \cdots + \theta_n)\big[u_{-1} - u_{\lfloor nr \rfloor-1}\big] \right. \\
&\quad \left. + \cdots + \theta_n \big[u_{1-n} - u_{\lfloor nr \rfloor-(n-1)}\big] \right\}.
\end{aligned} \tag{25}$$

Let $\Theta_j = \sum_{k=j}^{n} \theta_k$ for $1 \le j \le n$. Then, (25) can be rewritten as

$$\sum_{j=0}^{n} \frac{\theta_j (S_{j,\lfloor nr \rfloor} - S_{\lfloor nr \rfloor})}{V_n} = \sum_{j=1}^{n} \frac{\Theta_j u_{1-j}}{V_n} - \sum_{j=1}^{n} \frac{\Theta_j u_{\lfloor nr \rfloor-(j-1)}}{V_n} =: Q_{1,n} - Q_{2,n}. \tag{26}$$

Note that $\{u_j\}$ is a strictly stationary ergodic martingale difference sequence. Since $V_n^2/n \to_{\text{a.s.}} \sigma_u^2$ (by Lemma 1(a)) and $\Theta_j \to 0$ as $j, n \to \infty$, it then follows from the Toeplitz lemma (Stout, 1974, p. 120) that

$$\mathrm{E}(Q_{1,n}^2) = \frac{1}{n} \sum_{j=1}^{n} \Theta_j^2 \to 0 \quad \text{as} \quad n \to \infty. \tag{27}$$

By (27), the weak law of large numbers (Petrov, 1995, p. 134) implies that $Q_{1,n} \to_p \mathrm{E}(Q_{1,n}) = 0$. The proof of $Q_{2,n} \to_p 0$ is similar to that of $Q_{1,n}$ and hence the details are omitted. Moreover, we also need to show that $\mathrm{E}(Q_{1,n} Q_{2,n}) = 0$. It follows from Hölder's inequality (see Chung, 2001, p. 50) that $|\ \mathrm{E}(Q_{1,n} Q_{2,n})| \le \mathrm{E}|Q_{1,n}\ \ Q_{2,n}\ | \le [\mathrm{E}(Q_{1,n}^2)\ \mathrm{E}(Q_{2,n}^2)]^{1/2} \to 0$. Putting these results together yields $\mathrm{E}(Q_{1,n} - Q_{2,n})^2 \to 0$, implying that $Q_{1,n} - Q_{2,n} \to_p \mathrm{E}(Q_{1,n} - Q_{2,n}) = 0$ by the weak law of large numbers. As for the first term of (22), it is readily seen from Theorem 1 that

$$\sum_{j=0}^{\infty} \theta_j \frac{S_{\lfloor nr \rfloor}}{V_n} \Rightarrow \theta(1) W(r), \quad 0 \le r \le 1.$$

This completes the proof of Lemma 2.

Lemma 3. *Let $\{y_t\}$ and $\{u_t\}$ be generated according to (10) and (2), respectively. Suppose that Assumption 1 holds. If the regression (15) is estimated by ordinary LS then, as $n \to \infty$,*

(a)

$$\frac{1}{V_n^2}\sum_{t=1}^{n} R_{0t}R_{1t} = \frac{\theta(1)}{2}\left[\left(\frac{S_n}{V_n}\right)^2 - 1\right] - \frac{S_n}{V_n}\frac{1}{n}\sum_{t=1}^{n-1}\frac{S_{\varepsilon,t}}{V_n} + o_p(1);$$

(b)

$$\frac{1}{nV_n^2}\sum_{t=1}^{n} R_{1t}^2 = \frac{1}{n}\sum_{t=1}^{n-1}\left(\frac{S_{\varepsilon,t}}{V_n}\right)^2 - \left(\frac{1}{n}\sum_{t=1}^{n-1}\frac{S_{\varepsilon,t}}{V_n}\right)^2 + o_p(1);$$

(c)

$$\hat\sigma_R^2 = \frac{V_n^2}{n} + o_p(1).$$

Proof of Lemma 3. Note from (11) that, under the unit root hypothesis, $\boldsymbol{Z}_t = (1, \varepsilon_{t-1}, \dots, \varepsilon_{t-(m-1)})^\top$. Since ε_t is a short-memory linear process whose innovations are stationary ergodic martingale differences with finite variance, and since, by Hölder's inequality, $\mathrm{E}|\varepsilon_t\ \varepsilon_{t-j}| \le [\mathrm{E}(\varepsilon_t^2)\ \mathrm{E}(\varepsilon_{t-j}^2)]^{1/2} < \infty$ for all j, it follows from Stout (1974, theorems 3.5.7 and 3.5.8) that for $j \ge 0$, $n^{-1}\sum_{t=1}^n \varepsilon_{t-j} \to_{a.s.} \mathrm{E}(\varepsilon_{t-j}) = 0$ and $n^{-1}\sum_{t=1}^n \varepsilon_t\varepsilon_{t-j} \to_{a.s.} \mathrm{E}(\varepsilon_t\varepsilon_{t-j}) = \sigma_u^2 c_j$ as $n\to\infty$, where $c_j = \sum_{i=0}^\infty \theta_i\theta_{i+j}$. Let $\boldsymbol{\Upsilon}_m = \mathrm{diag}[\sqrt{n}, V_n, \dots, V_n]$ be a diagonal matrix of order m with $\sqrt{n}$ at the (1,1)th entry and V_n elsewhere on the diagonal. Note from Lemma 1(a) that $V_n^2/(n\,\sigma_u^2) \to_{\text{a.s.}} 1$. Then, as $n\to\infty$,

$$\left(\boldsymbol{\Upsilon}_m^{-1}\sum_{t=1}^{n}\boldsymbol{Z}_t\boldsymbol{Z}_t^\top\boldsymbol{\Upsilon}_m^{-1}\right)^{-1} \to_{a.s.} \begin{bmatrix} 1 & 0 \\ 0 & \boldsymbol{C}^{-1}\end{bmatrix}, \tag{28}$$

where

$$\boldsymbol{C} = \begin{bmatrix} c_0 & c_1 & \dots & c_{m-2} \\ c_1 & c_0 & \dots & c_{m-3} \\ \vdots & \vdots & \ddots & \vdots \\ c_{m-2} & c_{m-3} & \dots & c_0 \end{bmatrix}. \tag{29}$$

On the other hand, it is easy to see that

$$\frac{1}{V_n}\sum_{t=1}^{n} u_t\boldsymbol{Z}_t^\top\boldsymbol{\Upsilon}_m^{-1} = \left(\frac{\sum_{t=1}^n u_t}{\sqrt{n}V_n}, \frac{\sum_{t=1}^n u_t\varepsilon_{t-1}}{V_n^2}, \dots, \frac{\sum_{t=1}^n u_t\varepsilon_{t-(m-1)}}{V_n^2}\right). \tag{30}$$

Write $y_{t-1} = \varepsilon_1 + \dots + \varepsilon_{t-j-1} + \varepsilon_{t-j} + \dots + \varepsilon_{t-1} = y_{t-j-1} + \varepsilon_{t-j} + \dots + \varepsilon_{t-1}$ for $1 \le j \le m-1$. Then it follows from Lemma 1(a) and the above that

$$\begin{aligned}\frac{\sum_{t=1}^n y_{t-1}\varepsilon_{t-j}}{V_n^2} &= \frac{\sum_{t=1}^n y_{t-j-1}\varepsilon_{t-j}}{V_n^2} + \frac{\sum_{t=1}^n \varepsilon_{t-j}^2}{V_n^2} + \dots + \frac{\sum_{t=1}^n \varepsilon_{t-1}\varepsilon_{t-j}}{V_n^2} \\ &= \frac{\sum_{t=1}^n y_{t-j-1}\varepsilon_{t-j}}{V_n^2} + c_0 + c_1 + \dots + c_{j-1} + o_p(1). \end{aligned} \tag{31}$$

Similarly to Hamilton (1994, p. 476), noting that $y_t = S_{\varepsilon,t}$, we have from Lemma 1(a) that, for $1 \le j \le m-1$,

$$\begin{aligned}\frac{1}{V_n^2}\sum_{t=1}^{n} y_{t-j-1}\varepsilon_{t-j} &= \frac{1}{2}\left[\frac{y_{n-j}^2}{V_n^2} - \frac{\sum_{t=1}^{n}\varepsilon_{t-j}^2}{V_n^2}\right] \\ &= \frac{1}{2}\left[\left(\frac{S_{\varepsilon,n-j}}{V_n}\right)^2 - c_0\right] + o_p(1). \end{aligned} \tag{32}$$

Then it follows from (31) and (32) that

$$\begin{aligned}\boldsymbol{\Upsilon}_m^{-1}\frac{\sum_{t=1}^{n}\mathbf{Z}_t y_{t-1}}{V_n} &= \left(\frac{\sum_{t=1}^{n} y_{t-1}}{\sqrt{n}V_n}, \frac{\sum_{t=1}^{n}\varepsilon_{t-1}y_{t-1}}{V_n^2}, \dots, \frac{\sum_{t=1}^{n}\varepsilon_{t-(m-1)}y_{t-1}}{V_n^2}\right)^\top \\ &= \left(\frac{\sum_{t=1}^{n} y_{t-1}}{\sqrt{n}V_n}, \frac{1}{2}\left[\left(\frac{S_{\varepsilon,n-1}}{V_n}\right)^2 - c_0\right] + c_0 + o_p(1), \dots, \right. \\ &\quad \left. \frac{1}{2}\left[\left(\frac{S_{\varepsilon,n-(m-1)}}{V_n}\right)^2 - c_0\right] + c_0 + \dots + c_{m-2} + o_p(1)\right)^\top. \end{aligned} \tag{33}$$

Now we prove part (a). Under the unit root hypothesis, the data generating process in (11), i.e. $\Delta y_t = \varepsilon_t$, can be written as

$$\Delta y_t = \boldsymbol{\Phi}\mathbf{Z}_t + u_t, \tag{34}$$

where $\boldsymbol{\Phi} = [c, \xi_1, \dots, \xi_{M-1}, \xi_M, \dots, \xi_{m-1}] = [0, \xi_1, \dots, \xi_{M-1}, 0, \dots, 0]$. By (13)–(14) and (34), it is easy to show that

$$\begin{aligned} R_{0t} &= \boldsymbol{\Phi}\mathbf{Z}_t + u_t - \left[\sum_{t=1}^{n}(\boldsymbol{\Phi}\mathbf{Z}_t + u_t)\mathbf{Z}_t^\top\right]\left(\sum_{t=1}^{n}\mathbf{Z}_t\mathbf{Z}_t^\top\right)^{-1}\mathbf{Z}_t \\ &= u_t - \left(\sum_{t=1}^{n}u_t\mathbf{Z}_t^\top\right)\left(\sum_{t=1}^{n}\mathbf{Z}_t\mathbf{Z}_t^\top\right)^{-1}\mathbf{Z}_t, \end{aligned} \tag{35}$$

and

$$\begin{aligned}\frac{1}{V_n^2}\sum_{t=1}^{n}R_{0t}R_{1t} &= \frac{\sum_{t=1}^{n}u_t y_{t-1}}{V_n^2} - \frac{\sum_{t=1}^{n}u_t\mathbf{Z}_t^\top}{V_n}\boldsymbol{\Upsilon}_m^{-1} \\ &\quad \times \left(\boldsymbol{\Upsilon}_m^{-1}\sum_{t=1}^{n}\mathbf{Z}_t\mathbf{Z}_t^\top\boldsymbol{\Upsilon}_m^{-1}\right)^{-1}\boldsymbol{\Upsilon}_m^{-1}\frac{\sum_{t=1}^{n}\mathbf{Z}_t y_{t-1}}{V_n} \\ &=: A_n - B_n. \end{aligned} \tag{36}$$

Recall from Lemma 1(a) that $V_n^2/n \to_{\text{a.s.}} \sigma_u^2$, and note from Lemma 2 that $S_{\varepsilon,n-j}/V_n = O_p(1)$ for $1 \le j \le m-1$. Since the θ_j are absolutely summable, it is sufficient to show that, for

$j \geq 1$, $\sum_{t=1}^{n} u_t \varepsilon_{t-j}/V_n^2 = \sum_{i=0}^{\infty} \theta_i \sum_{t=1}^{n} u_t u_{t-j-i}/V_n^2 = o_p(1)$ as $n \to \infty$. Then, by (28), (30), (33) and the above, it is evident that

$$B_n = \frac{\sum_{t=1}^{n} u_t}{\sqrt{n} V_n} \frac{\sum_{t=1}^{n} y_{t-1}}{\sqrt{n} V_n} + o_p(1) = \frac{S_n}{V_n} \frac{1}{n} \sum_{t=1}^{n-1} \frac{S_{\varepsilon,t}}{V_n} + o_p(1). \tag{37}$$

As for the term A_n in (36), noting that $y_{t-1} = \sum_{k=1}^{t-1} \varepsilon_k = \sum_{k=1}^{t-1} \sum_{j=0}^{\infty} \theta_j u_{k-j}$, we can write

$$\begin{aligned}
A_n =&\ \frac{1}{V_n^2} \sum_{t=1}^{n} u_t \sum_{k=1}^{t-1} \sum_{j=0}^{\infty} \theta_j u_{k-j} = \frac{1}{V_n^2} \sum_{j=0}^{\infty} \theta_j \sum_{t=1}^{n} u_t (S_{t-1} + S_{j,t-1} - S_{t-1}) \\
=&\ \frac{1}{V_n^2} \sum_{j=0}^{\infty} \theta_j \sum_{t=1}^{n} u_t S_{t-1} + \frac{1}{V_n^2} \sum_{j=0}^{n} \theta_j \sum_{t=1}^{n} u_t (S_{j,t-1} - S_{t-1}) \\
&+ \frac{1}{V_n^2} \sum_{j=n+1}^{\infty} \theta_j \sum_{t=1}^{n} u_t (S_{j,t-1} - S_{t-1}) \\
=:&\ A_{1,n} + A_{2,n} + A_{3,n}.
\end{aligned} \tag{38}$$

Similarly to (32), it follows that

$$A_{1,n} = \frac{1}{V_n^2} \sum_{j=0}^{\infty} \theta_j \frac{1}{2} \left(S_n^2 - \sum_{t=1}^{n} u_t^2 \right) = \frac{\theta(1)}{2} \left[\left(\frac{S_n}{V_n} \right)^2 - 1 \right]. \tag{39}$$

Similarly to (24)–(26), the second term in (38) can be written as

$$\begin{aligned}
A_{2,n} =&\ \frac{1}{V_n^2} \sum_{j=0}^{n} \theta_j \sum_{t=1}^{n} u_t (S_{j,t-1} - S_{t-1}) = \frac{1}{V_n^2} \sum_{t=1}^{n} u_t \sum_{j=0}^{n} \theta_j (S_{j,t-1} - S_{t-1}) \\
=&\ \frac{1}{V_n^2} \sum_{t=1}^{n} u_t \left(\sum_{j=1}^{n} \Theta_j u_{1-j} \right) - \frac{1}{V_n^2} \sum_{t=1}^{n} u_t \left(\sum_{j=1}^{n} \Theta_j u_{t-j} \right) \\
=:&\ A_{21,n} - A_{22,n}.
\end{aligned} \tag{40}$$

Recall from Lemma 1(a) that $V_n^2/n \to_{\text{a.s.}} \sigma_u^2$. Note that $\mathrm{E}(u_i u_j) = 0$, $\mathrm{E}(u_i u_j u_k u_l) = 0$, $\mathrm{E}(u_i^2 u_j u_{u\mathrm{k}}) = 0$ and $\mathrm{E}(u_i^2 u_j^2) = \mathrm{E}(u_i^2)\ \mathrm{E}(u_j^2) = \sigma_u^4$ for $i \neq j \neq k \neq l$. Then, similarly to (27), we have $\mathrm{E}(A_{21,n}) = 0$ and

$$\mathrm{E}(A_{21,n}^2) = \frac{1}{n} \sum_{j=1}^{n} \Theta_j^2 \to 0 \quad \text{as} \quad n \to \infty. \tag{41}$$

This implies that $A_{21,n} \to_{\mathrm{p}} 0$ by the weak law of large numbers. By arguments similar to those shown above and in (27), the proofs of $A_{22,n} \to_{\mathrm{p}} 0$ and then $A_{2,n} \to_{\mathrm{p}} 0$ are obvious and hence omitted. As for the third term in (38), note from the third term in (22) that $\sum_{j=n+1}^{\infty} \theta_j (S_{j,\lfloor nr \rfloor} - S_{\lfloor nr \rfloor})/V_n = o_p(1)$, and from (9) in Theorem 1 that $\sum_{t=1}^{n} u_t/V_n = S_n/V_n = O_p(1)$. These two results imply that

$$\begin{aligned} A_{3,n} &= \frac{1}{V_n^2} \sum_{j=n+1}^{\infty} \theta_j \sum_{t=1}^{n} u_t (S_{j,t-1} - S_{t-1}) = \frac{1}{V_n^2} \sum_{t=1}^{n} u_t \sum_{j=n+1}^{\infty} \theta_j (S_{j,t-1} - S_{t-1}) \\ &= \sum_{t=1}^{n} \frac{u_t}{V_n} \sum_{j=n+1}^{\infty} \frac{\theta_j (S_{j,t-1} - S_{t-1})}{V_n} \\ &\to_p 0 \quad \text{as} \quad n \to \infty. \end{aligned} \tag{42}$$

Considering the above results together, the proof of Lemma 3(a) follows.

For part (b), by (14) and by noting that $y_{t-1} = S_{\varepsilon,t-1}$, we have

$$\begin{aligned} \frac{1}{nV_n^2} \sum_{t=1}^{n} R_{1t}^2 &= \frac{\sum_{t=1}^{n} y_{t-1}^2}{nV_n^2} - \frac{1}{n} \left(\frac{\sum_{t=1}^{n} y_{t-1} \mathbf{Z}_t^\top}{V_n} \boldsymbol{\Upsilon}_m^{-1} \right) \left(\boldsymbol{\Upsilon}_m^{-1} \sum_{t=1}^{n} \mathbf{Z}_t \mathbf{Z}_t^\top \boldsymbol{\Upsilon}_m^{-1} \right)^{-1} \\ &\quad \times \boldsymbol{\Upsilon}_m^{-1} \frac{\sum_{t=1}^{n} \mathbf{Z}_t y_{t-1}}{V_n} \\ &=: \frac{1}{n} \sum_{t=1}^{n-1} \left(\frac{S_{\varepsilon,t}}{V_n} \right)^2 - H_n. \end{aligned} \tag{43}$$

It is readily seen from (28), (33) and Lemma 2 that

$$H_n = \left(\frac{1}{n} \sum_{t=1}^{n} \frac{y_{t-1}}{V_n} \right)^2 + o_p(1) = \left(\frac{1}{n} \sum_{t=1}^{n-1} \frac{S_{\varepsilon,t}}{V_n} \right)^2 + o_p(1). \tag{44}$$

Putting (43) and (44) together completes the proof of Lemma 3(b).

For part (c), recall that $\hat{\sigma}_R^2 = \sum_{t=1}^{n} [R_{0t} - (\hat{\phi}_{A,n} - 1) R_{1t}]^2/(n-1)$. Then a straightforward calculation shows that

$$\hat{\sigma}_R^2 = \frac{1}{n-1} \sum_{t=1}^{n} R_{0t}^2 - \left[n(\hat{\phi}_{A,n} - 1) \right]^2 \frac{\sum_{t=1}^{n} R_{1t}^2}{n^2(n-1)}. \tag{45}$$

Since $n(\hat{\phi}_{A,n} - 1) = O_p(1)$, as will be shown in Theorem 2(a) below, and since by Lemma 3(b) and Lemma 1(a), $\sum_{t=1}^{n} R_{1t}^2 = O_p(n^2)$, the last term in (45) is $o_p(1)$ as $n \to \infty$. This result, together with the fact that $n/(n-1) \to 1$ as $n \to \infty$, implies that $\hat{\sigma}_R^2 = n^{-1} \sum_{t=1}^{n} R_{0t}^2 + o_p(1)$. Under the null hypothesis, it is sufficient to show by Lemma 1(a), (30) and (35) that

$$\begin{aligned} \hat{\sigma}_R^2 &= \frac{1}{n} \sum_{t=1}^{n} R_{0t}^2 + o_p(1) \\ &= \frac{\sum_{t=1}^{n} u_t^2}{n} - \frac{\sum_{t=1}^{n} u_t \mathbf{Z}_t^\top}{\sqrt{n}} \boldsymbol{\Upsilon}_m^{-1} \left(\boldsymbol{\Upsilon}_m^{-1} \sum_{t=1}^{n} \mathbf{Z}_t \mathbf{Z}_t^\top \boldsymbol{\Upsilon}_m^{-1} \right)^{-1} \boldsymbol{\Upsilon}_m^{-1} \frac{\sum_{t=1}^{n} \mathbf{Z}_t u_t}{\sqrt{n}} + o_p(1) \\ &= \frac{V_n^2}{n} + o_p(1) \quad \text{as} \quad n \to \infty. \end{aligned} \tag{46}$$

This completes the proof of Lemma 3(c).

Proof of Theorem 2. It follows from (16)–(17) and Lemma 3 that

$$
\begin{aligned}
ADF_\rho &= \frac{\sum_{t=1}^{n} R_{0t} R_{1t}/V_n^2}{\sum_{t=1}^{n} R_{1t}^2/(nV_n^2)} \\
&= \frac{\frac{\theta(1)}{2}[(S_n/V_n)^2 - 1] - (S_n/V_n)n^{-1}\sum_{t=1}^{n-1} S_{\varepsilon,t}/V_n}{n^{-1}\sum_{t=1}^{n-1}(S_{\varepsilon,t}/V_n)^2 - (n^{-1}\sum_{t=1}^{n-1} S_{\varepsilon,t}/V_n)^2} + o_p(1), \\
ADF_\tau &= \frac{\sum_{t=1}^{n} R_{0t} R_{1t}/V_n^2}{\left\{(\sum_{t=1}^{n} R_{1t}^2/V_n^2)(\hat{\sigma}_R^2/V_n^2)\right\}^{1/2}} \\
&= \frac{\frac{\theta(1)}{2}[(S_n/V_n)^2 - 1] - (S_n/V_n)n^{-1}\sum_{t=1}^{n-1} S_{\varepsilon,t}/V_n}{\left\{n^{-1}\sum_{t=1}^{n-1}(S_{\varepsilon,t}/V_n)^2 - (n^{-1}\sum_{t=1}^{n-1} S_{\varepsilon,t}/V_n)^2\right\}^{1/2}} + o_p(1).
\end{aligned}
$$

Clearly, by the above two equations and by Lemma 2, the remainder of the proof is the same as for the proofs of Theorem 1(c) and Theorem 1(d), and thus we omit it here. This completes the proof.

References

BANERJEE, A., DOLADO, J.J., GALBRAITH, J.W. & HENDRY, D.F. (1993). *Co-integration, Error Correction, and the Econometric Analysis of Non-Stationary Data*. Oxford: Oxford University Press.

BILLINGSLEY, P. (1968). *Convergence of Probability Measure*. New York: Wiley.

BILLINGSLEY, P. (1995). *Probability and Measure*, 3rd edn. New York: John Wiley & Sons.

BOLLERSLEV, T. (1986). Generalized autoregressive conditional heteroskedasticity. *J. Econometrics* **31**, 307–327.

BOUGEROL, P. & PICARD, N. (1992). Stationarity of GARCH Processes and of some nonnegative time series. *J. Econometrics* **52**, 115–127.

BROCKWELL, P.J. & DAVIS, R.A. (1991). *Time Series: Theory and Methods*, 2nd edn. New York: Springer-Verlag.

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

CHAN, N.H. & WEI, C.Z. (1988). Limiting distributions of least squares estimates of unstable autoregressive processes. *Ann. Statist.* **16**, 367–401.

CHANG, Y. & PARK, J.Y. (2002). On the asymptotics of ADF tests for unit roots. *Econometric Rev.* **21**, 431–447.

CHUNG, K.L. (2001). *A Course in Probability Theory*, 3th edn. San Diego: Academic Press.

DAVIDSON, R. & MACKINNON, J.G. (1993). *Estimation and Inference in Econometrics*. Oxford: Oxford University Press.

DICKEY, D.A. & FULLER, W.A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *J. Amer. Statist. Assoc.* **74**, 427–431.

DICKEY, D.A. & FULLER, W.A. (1981). Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica* **49**, 1057–1072.

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

EVIEWS 3.1. (1999) QMS Quantitative Micro Software, 4521 Campus Drive Suite 336 Irvine, CA 92612-2699, USA. (www.eviews.com)

HALL, P. (1979) On the Skorokhod representation approach to martingale invariance principles. *Ann. Probab.* **7**, 371–376.

HALL, P. (1992). Convergence rates in the central limit theorem for means of autoregressive and moving average sequences. *Stochastic Process. Appl.* **43**, 115–131.

HALL, P. & HEYDE, C.C. (1980). *Martingale Limit Theory and its Application*. New York: Academic Press.

HAMILTON, J.D. (1994). *Time Series Analysis*. Princeton: Princeton University Press.

HANSEN, E. & RAHBEK, A. (1998). Stationarity and asymptotics of multivariate ARCH time series with an application to robustness of cointegration analysis. Preprint. Department of Theoretical Statistics, University of Copenhagen.

HERRNDORF, N. (1984). A functional central limit theorem for weakly dependent sequences of random variables. *Ann. Probab.* **12**, 141–153.

JOHANSEN, S. (1995). *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*. Oxford: Oxford University Press.

KIM, K. & SCHMIDT, P. (1993). Unit root tests with conditional heteroskedasticity. *J. Econometrics* **59**, 287–300.

LI, W.K., LING, S. & MCALEER, M. (2002). Recent theoretical results for time series models with GARCH errors. *J. Econom. Surv.* **16**, 245–269.

LING, S. (2004). Estimation and testing stationarity for double-autoregressive models. *J. Roy. Statist. Soc. Ser. B* **66**, 63–78.

LING, S. & LI, W.K. (1998). Limiting distributions of maximum likelihood estimators for unstable autoregressive moving-average time series with general autoregressive heteroscedastic errors. *Ann. Statist.* **26**, 84–125.

LING, S. & LI, W.K. (2003). Asymptotic inference for unit root processes with GARCH(1,1) errors. *Econometric Theory* **19**, 541–564.

LING, S. & MCALEER, M. (2002). Necessary and sufficient moment conditions for the GARCH(r,s) and asymmetric power GARCH(r,s) models. *Econometric Theory* **18**, 722–729.

LING, S., LI, W.K. & MCALEER, M. (2003). Estimation and testing for unit root processes with GARCH(1,1) errors: Theory and Monte Carlo evidence. *Econometric Rev.* **22**, 179–202.

MCLEISH, D.L. (1975). A maximal inequality and dependent strong laws. *Ann. Probab.* **3**, 829–839.

MIKOSCH, T. & STĂRICĂ, C. (2000). Limit theory for the sample autocorrelations and extremes of a GARCH(1,1) process. *Ann. Statist.* **28**, 1427–1451.

PANTULA, S.G. (1986). Comment on modelling the persistence of conditional variances. *Econometric Rev.* **5**, 71–74.

PANTULA, S.G. (1988). Estimation of autoregressive models with ARCH errors. *Sankhyā B* **50**, 119–138.

PETROV, V.V. (1995). *Limit Theorems of Probability Theory*. Oxford: Clarendon Press.

PHILLIPS, P.C.B. (1987). Time series regression with a unit root. *Econometrica* **55**, 277–301.

PHILLIPS, P.C.B. & SOLO, V. (1992). Asymptotics for linear processes. *Ann. Statist.* **20**, 971–1001.

PHILLIPS, P.C.B. & XIAO, Z. (1998). A primer on unit root testing. *J. Econom. Surv.* **12**, 423–469.

SEO, B. (1999). Distribution theory for unit root tests with conditional heteroskedasticity. *J. Econometrics* **91**, 113–144.

STOCK, J.H. (1994). Unit root, structural breaks and trends. In *Handbook of Econometrics*, Vol. IV, eds R.F. ENGLE & D.L. MCFADDEN, pp. 2739–2841. Amsterdam: Elsevier Science.

STOUT, W.F. (1974). *Almost Sure Convergence*. New York: Academic Press.

WANG, G. (2006). A note on unit root tests with heavy-tailed GARCH errors. *Statist. Probab. Lett.* **76**, 1075–1079.

WHITE, J.S. (1958). The limiting distribution of the serial correlation coefficient in the explosive case. *Ann. Math. Statist.* **29**, 1188–1197.