ESTIMATION AND TESTING FOR
UNIT ROOT PROCESSES
WITH GARCH(1,1) ERRORS:
THEORY AND MONTE CARLO EVIDENCE

Shiqing ling
W. K. Li
and
Michael McAleer

June 2001

Estimation and Testing for Unit Root Processes with GARCH (1, 1) Errors: Theory and Monte Carlo Evidence*

Shiqing Ling

Department of Mathematics

Hong Kong University of Science and Technology

W. K. Li

Department of Statistics and Actuarial Science

University of Hong Kong

Michael McAleer

Department of Economics

University of Western Australia

Revised: June 2001

Abstract

Least squares (LS) and maximum likelihood (ML) estimation are considered for unit root processes with GARCH (1, 1) errors. The asymptotic distributions of LS and ML estimators are derived under the condition $\alpha + \beta < 1$. The former has the usual unit root distribution and the latter is a functional of a bivariate Brownian motion, as in Ling and Li (1998). Several unit root tests based on LS estimators, ML estimators, and mixing LS and ML estimators, are constructed. Simulation results show that tests based on mixing LS and ML estimators perform better than Dickey-Fuller tests which are based on LS estimators, and that tests based on the ML estimators perform better than the mixed estimators.

Key Words and Phrases: Asymptotic distribution, Brownian motion, GARCH model, Least squares estimator, Maximum likelihood estimator, Unit root.

1 Introduction

Consider two unit root processes,

$$y_t = \phi y_{t-1} + \varepsilon_t \tag{1.1}$$

and

$$y_t = \mu + \phi y_{t-1} + \varepsilon_t, \tag{1.2}$$

where $\phi = 1$, $\mu = 0$ and ε_t follows the first-order generalized autoregressive conditional heteroscedasticity model, denoted as GARCH (1, 1),

$$\varepsilon_t = \eta_t \sqrt{h_t}, \quad h_t = \omega + \alpha \varepsilon_{t-1}^2 + \beta h_{t-1}, \tag{1.3}$$

where $\omega > 0$, $\alpha \geq 0$, $\beta \geq 0$, and the η_t are a sequence of independently and identically distributed (i.i.d.) random variables with zero mean and unit variance. For model (1.3), we make the following assumptions:

Assumption 1. $\alpha + \beta < 1$;

Assumption 2. The parameter vector $\delta = (\omega, \alpha, \beta)' \in \Theta$, where $\Theta = \{(\omega, \alpha, \beta)|$ $\omega > 0$, $\bar{\alpha} \leq \alpha \leq 1 - \bar{\alpha}$, $\bar{\beta} \leq \beta \leq 1 - \bar{\beta}$, for some $\bar{\alpha}, \bar{\beta} > 0\}$.

Assumption 3. η_t has a symmetric distribution and $E\eta_t^4 < \infty$.

The GARCH model was proposed by Bollerslev (1986) and has had many important applications in financial and econometric time series. Some recent reviews can be found in Bollerslev et al. (1992), Bollerslev et al. (1994) and Li et al. (1999). When $\alpha = \beta = 0$, the ε_t defined by model (1.3) reduce to i.i.d. white noise and, for this case, the unit root process has been investigated extensively. Motivated by practical applications, in recent decades many statisticians and econometricians have considered various unit root processes with non-i.i.d. errors. Some related results

on estimating and testing unit roots can be found in Phillips and Durlauf (1986), Phillips (1987), Chan and Wei (1988), Lucas (1995), and Herce (1996), and the references cited therein. When the error term follows a GARCH process, estimation and testing for a unit root involves intrinsic problems, an issue that was first raised by Pantula (1989). He derived the asymptotic distribution of least squares (LS) estimators for a unit root process with a first-order ARCH error (i.e. model (1.3) with $\beta = 0$), and showed that Dickey-Fuller tests could still be employed in this case. Pantula (1986, p.73) also stated without proof that Dickey-Fuller tests could be used for unit root processes with GARCH errors.

Peters and Veloce (1988) and Kim and Schmidt (1993) provided simulation results to show that Dickey-Fuller tests based on LS estimators are often sensitive and, when $\alpha + \beta$ is close to 1, the problem can be very serious. It seems that this phenomenon can be explained partly by the loss of efficiency of the LS estimator. Ling and Li (1998) derived the limiting distribution of the maximum likelihood (ML) estimator for a general nonstationary autoregressive moving average time series process with higher-order GARCH errors, and demonstrated that it is more efficient than the LS estimator. As for stationary time series with GARCH errors (see Weiss, 1986, and Ling and Li, 1997a), Ling and Li's (1998) results are obtained under the assumption that the fourth moment is finite. However, for the GARCH (1, 1) process, the condition for strict stationarity is $E(\ln(\alpha\eta_t^2 + \beta)) < 0$ (see Nelson, 1990), the condition for a finite variance is $\alpha + \beta < 1$, and the condition for a finite fourth moment under normality is $3\alpha^2 + 2\alpha\beta + \beta^2 < 1$. The fourth moment condition is clearly the most stringent.

For the pure GARCH (1, 1) model, Lee and Hansen (1994) and Lumsdaine (1996) proved that ML estimators are consistent and asymptotically normal under

the condition that $E(\ln(\alpha\eta_t^2+\beta)) < 0$. A challenging problem is whether the limiting distribution of ML estimators can be derived under weaker conditions for the unit root process with GARCH errors.

Ling and Li (1997b) obtained the asymptotic distributions of the ML estimators for models (1.1) and (1.2) under Assumptions 1-3. The limiting distribution of the estimated unit root is a functional of a bivariate Brownian motion and is the same as that obtained in Ling and Li (1998). Based on these asymptotic results, we can construct several new unit root tests. Simulation results reported in the paper show that tests based on mixing LS and ML estimators perform better than those based on LS estimators alone.

This paper proceeds as follows. Section 2 considers LS estimation and its asymptotic properties. Section 3 considers ML estimation and its asymptotic properties. Section 4 constructs some unit root tests. Section 5 reports some simulation results. The proofs of the theorems are given in section 6.

Throughout the paper, we use the following notation. U' denotes the transpose of the vector U; $o(1)$ $(o_p(1))$ denotes a series of numbers (random numbers) converging to zero (in probability); $O(1)$ $(O_p(1))$ denotes a series of numbers (random numbers) that are bounded (in probability); $\xrightarrow{p}$ and $\xrightarrow{\mathcal{L}}$ denote convergence in probability and in distribution, respectively. $D = D[0, 1]$ denotes the space of functions $f(s)$ on $[0, 1]$, which is defined and equipped with the Skorokhod topology (Billingsley, 1968). $||\cdot||$ denotes the Euclidean norm.

2   Preliminary Estimation

The observations $y_1, \cdots, y_n$, with initial value $y_0 = 0$, are generated by model (1.1) or (1.2). Denote $\hat{\phi}_{LS}$ as the LS estimator of ϕ in model (1.1) and $(\tilde{\mu}_{LS}, \tilde{\phi}_{LS})$ as the

LS estimator of (μ, ϕ) in model (1.2). Then

$$\hat{\phi}_{LS} = \left(\sum_{t=1}^{n} y_{t-1}^2 \right)^{-1} \left(\sum_{t=1}^{n} y_t y_{t-1} \right) \tag{2.1}$$

and

$$\begin{pmatrix} \tilde{\mu}_{LS} \\ \tilde{\phi}_{LS} \end{pmatrix} = \begin{pmatrix} n & \sum_{t=1}^{n} y_{t-1} \\ \sum_{t=1}^{n} y_{t-1} & \sum_{t=1}^{n} y_{t-1}^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum_{t=1}^{n} y_t \\ \sum_{t=1}^{n} y_t y_{t-1} \end{pmatrix}. \tag{2.2}$$

Theorem 2.1. Suppose that assumptions 1-3 hold. Then

(a) $\qquad n(\hat{\phi}_{LS} - 1) \xrightarrow{\mathcal{L}} \dfrac{\int_0^1 B_1(t) dB_1(t)}{\int_0^1 B_1^2(t) dt};$

(b) $\qquad N_n \begin{pmatrix} \tilde{\mu}_{LS} \\ \tilde{\phi}_{LS} - 1 \end{pmatrix} \xrightarrow{\mathcal{L}} \begin{pmatrix} 1 & \int_0^1 B_1(t) dt \\ \int_0^1 B_1(t) dt & \int_0^1 B_1^2(t) dt \end{pmatrix}^{-1} \begin{pmatrix} B_1(1) \\ \int_0^1 B_1(t) dB_1(t) \end{pmatrix},$

where $N_n = diag(\sqrt{n}, n)$ and $B_1(t)$ is a standard Brownian motion.

Let $\hat{\varepsilon}_t = y_t - \hat{\phi}_{LS} y_{t-1}$ or $\hat{\varepsilon}_t = y_t - \tilde{\mu}_{LS} - \tilde{\phi}_{LS} y_{t-1}$. Then $\{\hat{\varepsilon}_1, \hat{\varepsilon}_2, \cdots, \hat{\varepsilon}_n\}$ is a sequence of artificial observations of model (1.3). Denote

$$l_t(\varepsilon_t | \delta) = -\frac{1}{2} \ln h_t - \frac{\varepsilon_t^2}{2h_t} \tag{2.3}$$

and

$$l_t(\hat{\varepsilon}_t | \delta) = -\frac{1}{2} \ln \hat{h}_t - \frac{\hat{\varepsilon}_t^2}{2\hat{h}_t}, \tag{2.4}$$

where $\hat{h}_t = \omega + \alpha \hat{\varepsilon}_t^2 + \beta \hat{h}_{t-1}$ with $\hat{h}_0 = 1$, and $\delta = (\omega, \alpha, \beta)'$.

Lemma 2.1. Under assumptions 1-3

(a) $\qquad \sup\limits_{\delta \in \Theta} |\dfrac{1}{n} \sum\limits_{t=1}^{n} l_t(\hat{\varepsilon}_t | \delta) - \dfrac{1}{n} \sum\limits_{t=1}^{n} l_t(\varepsilon_t | \delta)| = o_p(1),$

(b) $\qquad \sup\limits_{\delta \in \Theta} |\dfrac{1}{\sqrt{n}} \sum\limits_{t=1}^{n} \dfrac{\partial l_t(\hat{\varepsilon}_t | \delta)}{\partial \delta} - \dfrac{1}{\sqrt{n}} \sum\limits_{t=1}^{n} \dfrac{\partial l_t(\varepsilon_t | \delta)}{\partial \delta}| = o_p(1),$

(c) $\qquad \sup\limits_{\delta \in \Theta} |\dfrac{1}{n} \sum\limits_{t=1}^{n} \dfrac{\partial^2 l_t(\hat{\varepsilon}_t | \delta)}{\partial \delta \partial \delta'} - \dfrac{1}{n} \sum\limits_{t=1}^{n} \dfrac{\partial^2 l_t(\varepsilon_t | \delta)}{\partial \delta \partial \delta'}| = o_p(1).$

Suppose that δ_0 is the true value of δ in Θ and let

$$\hat{\delta}_n = \arg\max_{\delta \in \Theta} \left[\frac{1}{n} \sum_{t=1}^{n} l_t(\hat{\varepsilon}_t | \delta) \right].$$

4

The following shows the asymptotic properties of $\hat{\delta}_n$.

Theorem 2.2. Under assumptions 1-3

$$(a) \qquad \hat{\delta}_n - \delta_0 \xrightarrow{p} 0,$$

$$(b) \qquad \sqrt{n}(\hat{\delta}_n - \delta_0) \xrightarrow{\mathcal{L}} N(O, V_0),$$

where $V_0 = B_0^{-1} A_0 B_0^{-1}$,

$$B_0 = -E\left[\frac{\partial^2 l_t(\varepsilon_t|\delta_0)}{\partial\delta\partial\delta'}\right] \quad \text{and} \quad A_0 = E\left[\frac{\partial l_t(\varepsilon_t|\delta_0)}{\partial\delta}\frac{\partial l_t(\varepsilon_t|\delta_0)}{\partial\delta'}\right].$$

From Theorems 2.1 and 2.2, we see that ϕ or (μ, ϕ) and δ can be estimated separately with $\hat{\phi}_{LS} - \phi = O(n^{-1})$ or $(\tilde{\mu}_{LS}, \tilde{\phi}_{LS}) - (\mu, \phi) = O(N_n^{-1})$ and $\hat{\delta} - \delta_0 = O(n^{-1/2})$. However, under GARCH errors the LS estimator of ϕ or (μ, ϕ) loses some efficiency (see the next section), which will result in a loss of efficiency for the 'ML' estimator of δ above in finite samples. A more efficient estimation procedure will be given in the next section, and all estimators in this section can be used as preliminary estimators.

3 ML Estimation

To simplify notation, in this section the true parameter δ_0 is denoted as δ. Let $\lambda = (\phi, \delta')'$, $\lambda_\mu = (\phi'_\mu, \delta')'$ and $\phi_\mu = (\mu, \phi)'$. The ML estimators of λ and λ_μ are the estimators denoted by $\hat{\lambda} = (\hat{\phi}_{ML}, \hat{\delta}'_{ML})'$ and $\tilde{\lambda}_\mu = (\tilde{\phi}'_{\mu,ML}, \tilde{\delta}'_{\mu,ML})'$, respectively, with $\tilde{\phi}_{\mu,ML} = (\tilde{\mu}_{ML}, \tilde{\phi}_{ML})'$, that maximize the log-likelihood

$$L = \frac{1}{n}\sum_{t=1}^{n} l_t, \tag{3.1}$$

where l_t is defined as in (2.3).

Since the likelihood equation $\partial l_t/\partial\lambda = 0$ and $\partial l_t/\partial\lambda_\mu = 0$ are nonlinear in λ and λ_μ, respectively, an iterative numerical procedure is required to obtain the solutions

to these equations. By Lemma 6.6, ϕ (or ϕ_μ) and δ can be estimated separately without loss of efficiency. Thus, we can define an algorithm scheme by the iterative approximate Newton-Raphson relation for $\hat{\phi}_{ML}$ and $\hat{\delta}_{ML}$:

$$
\begin{aligned}
\hat{\phi}^{(i+1)} &= \hat{\phi}^{(i)} + \left[\sum_{t=1}^{n} \frac{\partial^2 l_t}{\partial \phi^2} \right]^{-1}_{\hat{\lambda}^{(i)}} \left[\sum_{t=1}^{n} \frac{\partial l_t}{\partial \phi} \right]_{\hat{\lambda}^{(i)}} \\
\hat{\delta}^{(i+1)} &= \hat{\delta}^{(i)} + \left[\sum_{t=1}^{n} \frac{\partial^2 l_t}{\partial \delta \partial \delta'} \right]^{-1}_{\hat{\lambda}^{(i)}} \left[\sum_{t=1}^{n} \frac{\partial l_t}{\partial \delta} \right]_{\hat{\lambda}^{(i)}},
\end{aligned} \tag{3.2}
$$

where $\hat{\lambda}^{(i)} = (\hat{\phi}^{(i)}, \hat{\delta}'^{(i)})'$ is the estimated value at the i-th iteration. Similarly, define a scheme for $\tilde{\phi}_{\mu,ML}$ and $\tilde{\delta}_{\mu,ML}$. To derive our asymptotic results, we need the following theorem.

Theorem 3.1. Let $\Theta_n = \{\tilde{\lambda} : ||G_n(\tilde{\lambda} - \lambda)|| \leq M\}$ and $\Theta_{\mu,n} = \{\tilde{\lambda}_\mu : ||G_{\mu,n}(\tilde{\lambda}_\mu - \lambda_\mu)|| \leq M\}$, where M is an positive constant, $G_n^{-1} = \text{diag}\{n, \sqrt{n}, \sqrt{n}, \sqrt{n}\}$ and $G_{\mu,n}^{-1} = \text{diag}\{\sqrt{n}, n, \sqrt{n}, \sqrt{n}, \sqrt{n}\}$. Then

(a) $\quad G_n \sum_{t=1}^{n} \left(\frac{\partial^2 l_t}{\partial \lambda \partial \lambda'}\big|_{\lambda=\tilde{\lambda}} - \frac{\partial^2 l_t}{\partial \lambda \partial \lambda'} \right) G_n = o_p(1),$

(b) $\quad G_n \sum_{t=1}^{n} \left(\frac{\partial l_t}{\partial \lambda}\big|_{\lambda=\tilde{\lambda}} - \frac{\partial l_t}{\partial \lambda} \right) = G_n \sum_{t=1}^{n} \left(\frac{\partial^2 l_t}{\partial \lambda \partial \lambda'} \right) (\tilde{\lambda} - \lambda) + o_p(1),$

(c) $\quad G_{\mu,n} \sum_{t=1}^{n} \left(\frac{\partial^2 l_t}{\partial \lambda_\mu \partial \lambda'_\mu}\big|_{\lambda_\mu=\tilde{\lambda}_\mu} - \frac{\partial^2 l_t}{\partial \lambda_\mu \partial \lambda'_\mu} \right) G_{\mu,n} = o_p(1)$

(d) $\quad G_{\mu,n} \sum_{t=1}^{n} \left(\frac{\partial l_t}{\partial \lambda_\mu}\big|_{\lambda_\mu=\tilde{\lambda}_\mu} - \frac{\partial l_t}{\partial \lambda_\mu} \right) = G_{\mu,n} \sum_{t=1}^{n} \left(\frac{\partial^2 l_t}{\partial \lambda_\mu \partial \lambda'_\mu} \right) (\tilde{\lambda}_\mu - \lambda_\mu) + o_p(1),$

where $o_p(1)$ holds uniformly in Θ_n and $\Theta_{\mu,n}$.

By Theorems 2.1 and 2.2 in section 2, we can obtain the initial estimator of λ (or λ_μ) such that $\hat{\phi} - \phi = O_p(n^{-1})$ (or $\hat{\phi}_\mu - \phi_\mu = O_p(N_n^{-1})$) and $\hat{\delta} - \delta = O_p(n^{-1/2})$. With these consistent initial estimators, we can obtain the following asymptotic representations by Theorem 3.1:

$$
n(\hat{\phi}_{ML} - \phi) = \left[\frac{1}{n^2} \sum_{t=1}^{n} \frac{\partial^2 l_t}{\partial \phi^2} \right]^{-1} \left[\sum_{t=1}^{n} \frac{1}{n} \frac{\partial l_t}{\partial \phi} \right] + o_p(1), \tag{3.3}
$$

$$
\sqrt{n}(\hat{\delta}_{ML} - \delta) = \left[\frac{1}{n} \sum_{t=1}^{n} \frac{\partial^2 l_t}{\partial \delta \partial \delta'} \right]^{-1} \left[\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{\partial l_t}{\partial \delta} \right] + o_p(1). \tag{3.4}
$$

The asymptotic representations of $n(\tilde{\phi}_{\mu,ML} - \phi)$ and $\sqrt{n}(\tilde{\delta}_{\mu,ML} - \delta)$ can be obtained from (3.3)-(3.4) with $\hat{\phi}_{ML}$ and $\hat{\delta}_{ML}$ replaced by $\tilde{\phi}_{\mu,ML}$ and $\tilde{\delta}_{\mu,ML}$, respectively. The following theorem gives the asymptotic distributions of $(\hat{\phi}_{ML}, \hat{\delta}')'$ and $(\tilde{\phi}'_{\mu,ML}, \tilde{\delta}'_{ML})'$.

Theorem 3.2. Let $(\hat{\phi}_{ML}, \hat{\delta}'_{ML})'$ and $(\tilde{\phi}'_{\mu,ML}, \tilde{\delta}'_{\mu,ML})'$ be the estimators of $(\phi, \delta')'$ and $(\phi'_{\mu}, \delta')'$ obtained from (3.2) by using initial estimators $\hat{\phi}$ or $\hat{\phi}_{\mu}$ and $\hat{\delta}$ with $\hat{\phi} - \phi = O_p(n^{-1})$ or $\hat{\phi}_{\mu} - \phi_{\mu} = O_p(N_n^{-1})$ and $\hat{\delta} - \delta = O_p(n^{-1/2})$, respectively. Then under assumptions 1-3,

(a) $\quad n(\hat{\phi}_{ML} - 1) \xrightarrow{\mathcal{L}} \dfrac{\int_0^1 w_1(t)dw_2(t)}{F \int_0^1 w_1^2(t)dt};$

(b) $\quad N_n \begin{pmatrix} \tilde{\mu}_{ML} \\ \tilde{\phi}_{\mu,ML} - 1 \end{pmatrix} \xrightarrow{\mathcal{L}} F^{-1} \begin{pmatrix} 1 & \int_0^1 w_1(t)dt \\ \int_0^1 w_1(t)dt & \int_0^1 w_1^2(t)dt \end{pmatrix}^{-1} \begin{pmatrix} w_2(1) \\ \int_0^1 w_1(t)dw_2(t) \end{pmatrix},$

where N_n is defined in Theorem 2.1 and $(w_1(t), w_2(t))$ is a bivariate Brownian motion with covariance

$$t\Omega = t \begin{pmatrix} Eh_t & 1 \\ 1 & E(1/h_t) + \kappa\alpha^2 \sum_{k=1}^{\infty} \beta^{2(k-1)} E(\varepsilon_{t-k}^2/h_t^2) \end{pmatrix}, \qquad (3.5)$$

$\kappa = E\eta_t^4 - 1$ and, when η_t is normal, $\kappa = 2$. For $\sqrt{n}(\hat{\delta}_{ML} - \delta)$ and $\sqrt{n}(\tilde{\delta}_{\mu,ML} - \delta)$, their asymptotic distributions are the same as those given by Theorem 2.2(b).

Let

$$B_1(t) = \frac{1}{\sigma} w_1(t) \quad \text{and} \quad B_2(t) = -\frac{1}{\sigma^2}\sqrt{\frac{\sigma^2}{\sigma^2 K - 1}} w_1(t) + \sqrt{\frac{\sigma^2}{\sigma^2 K - 1}} w_2(t),$$

where $\sigma^2 = Eh_t$ and K is the (2,2)-th element of Ω in (3.5). Denote K by F when $\kappa = 2$. Then $B_1(t)$ and $B_2(t)$ are two independent standard Brownian motions. As shown in Ling and Li (2000),

$$n(\tilde{\phi}_{ML} - 1) \xrightarrow{\mathcal{L}} \frac{\int_0^1 B_1(t)dB_1(t)}{\sigma^2 F \int_0^1 B_1^2(t)dt} + \frac{\sqrt{\sigma^2 K - 1} \int_0^1 B_1(t)dB_2(t)}{\sigma^2 F \int_0^1 B_1^2(t)dt}. \qquad (3.6)$$

The second term of (3.6) can be simplified as $[\sqrt{\sigma^2 K - 1}/F\sigma^2] (\int_0^1 B_1^2(t) \, dt)^{-1/2}\xi$, where ξ is a standard normal random variable independent of $\int_0^1 B_1^2(t)dt$ (see Phillips,

1989). Thus,

$$n(\hat{\phi}_{ML} - 1) \xrightarrow{\mathcal{L}} \frac{\int_0^1 B_1(t)dB_1(t)}{\sigma^2 F \int_0^1 B_1^2(t)dt} + \frac{\sqrt{\sigma^2 K - 1}}{\sigma^2 F}(\int_0^1 B_1^2(t)dt)^{-1/2}\xi. \qquad (3.7)$$

Similarly, we can simplify the limiting distribution of $n(\tilde{\phi}_{\mu,ML} - 1)$ as

$$n(\tilde{\phi}_{\mu,ML} - 1) \xrightarrow{\mathcal{L}} \frac{\int_0^1 B_1(t)dB_1(t) - B_1(1)\int_0^1 B_1(t)dt}{\sigma^2 F[\int_0^1 B_1^2(t)dt - (\int_0^1 B_1(t)dt)^2]}$$
$$+ \frac{\sqrt{\sigma^2 K - 1}}{\sigma^2 F}[\int_0^1 B_1^2(t)dt - (\int_0^1 B_1(t)dt)^2]^{-1/2}\xi. \qquad (3.8)$$

From (3.7)-(3.8), we see that the asymptotic distributions of $\hat{\phi}_{ML}$ and $\tilde{\phi}_{\mu,ML}$ can be represented, respectively, as combinations of those of $\hat{\phi}_{LS}$ and $\tilde{\phi}_{\mu,LS}$ and a scale mixture of normals. These properties are similar to those of the least absolute deviation estimators of unit roots given by Herce (1996). The ML estimator of ϕ or ϕ_μ is more efficient than the LS estimator given in the last section (see the work by Ling and Li (1998)).

4   Unit Root Tests

4.1   Method A

Based on the asymptotic results in section 2, we can construct some unit root tests for the nonstationary model (1.1) with GARCH error (1.3) and model (1.2) with GARCH error (1.3). First we define the test statistics based on LS estimators:

$$
\begin{aligned}
L_\phi &= n(\hat{\phi}_{LS} - 1); \\
L_t &= \left(\frac{1}{n^2}\sum_{t=1}^n y_{t-1}^2\right)^{1/2} L_\phi; \\
L_{\mu,\phi} &= n(\hat{\phi}_{\mu,LS} - 1); \\
L_{\mu,t} &= \left[\frac{1}{n^2}\sum_{t=1}^n (y_{t-1} - \bar{y})^2\right]^{1/2} L_{\mu,\phi},
\end{aligned}
$$

where $\bar{y} = n^{-1}\sum_{t=1}^n y_{t-1}$. The following theorem comes from Theorem 2.1.

8

Theorem 4.1. Under assumptions 1-3,

$$(a) \qquad L_\phi \xrightarrow{\mathcal{L}} \frac{\int_0^1 B_1(t)dB_1(t)}{\int_0^1 B_1^2(t)};$$

$$(b) \qquad L_t \xrightarrow{\mathcal{L}} \frac{\int_0^1 B_1(t)dB_1(t)}{\sqrt{\int_0^1 B_1^2(t)}};$$

$$(c) \qquad L_{\mu,\phi} \xrightarrow{\mathcal{L}} \frac{\int_0^1 B_1(t)dB_1(t) - B_1(1)\int_0^1 B_1(t)dt}{\int_0^1 B_1^2(t) - (\int_0^1 B_1(t)dt)^2};$$

$$(d) \qquad L_{\mu,t} \xrightarrow{\mathcal{L}} \frac{\int_0^1 B_1(t)dB_1(t) - B_1(1)\int_0^1 B_1(t)dt}{[\int_0^1 B_1^2(t) - (\int_0^1 B_1(t)dt)^2]^{1/2}}.$$

These limiting distributions of the LS-based tests, i.e. L_ϕ, L_t, $L_{\mu,\phi}$ and $L_{\mu,t}$, are the same as those given by Dickey and Fuller (1979). The critical values of these distributions can be found in Tables 8.5.2 and 8.5.3 of Fuller (1976).

4.2 Method B

In order to apply ML estimators for unit root tests, we need to modify $n(\hat{\phi}_{ML} - 1)$ and $n(\tilde{\phi}_{\mu,ML} - 1)$ in Theorem 3.2 because these limiting distributions depend on nuisance parameters. These nuisance parameters should be replaced by their consistent estimators, an approach that was first used by Phillips (1987). Recently a similar approach was employed by Lucas (1995) and Herce (1996). By (3.7)-(3.8), we can define test statistics by mixing LS estimators and ML estimators as follows:

$$M_\phi = \frac{\hat{\sigma}^2\hat{F}}{\sqrt{\hat{\sigma}^2\hat{K} - 1}}\left\{n(\hat{\phi}_{ML} - 1) - (\hat{F}\hat{\sigma}^2)^{-1}[n(\hat{\phi}_{LS} - 1)]\right\};$$

$$M_t = \left(\frac{1}{n^2}\sum_{t=1}^n y_{t-1}^2\right)^{1/2}M_\phi;$$

$$M_\mu = \frac{\tilde{\sigma}^2\tilde{F}}{\sqrt{\tilde{\sigma}^2\tilde{K} - 1}}\left\{n(\tilde{\phi}_{\mu,ML} - 1) - (\tilde{F}\tilde{\sigma}^2)^{-1}[n(\tilde{\phi}_{\mu,LS} - 1)]\right\};$$

$$M_{\mu,t} = \left(\frac{1}{n^2}\sum_{t=1}^n (y_{t-1} - \bar{y})^2\right)^{1/2}M_\mu.$$

We call these ML-based tests. When $\eta_t \sim i.i.d.N(0,1)$, $F = K$ and hence $\hat{F} = \hat{K}$ and $\tilde{F} = \tilde{K}$. The following theorem comes from Theorems 2.1 and 3.2.

Theorem 4.2. Under assumptions 1-3,

$$(a) \quad M_\phi \xrightarrow{\mathcal{L}} \left[\int_0^1 B_1^2(t)dt \right]^{-1/2} \xi;$$

$$(b) \quad M_t \xrightarrow{\mathcal{L}} \xi;$$

$$(c) \quad M_\mu \xrightarrow{\mathcal{L}} \left[\int_0^1 B_1^2(t)dt - \left(\int_0^1 B_1(t) \right)^2 \right]^{-1/2} \xi;$$

$$(d) \quad M_{\mu,t} \xrightarrow{\mathcal{L}} \xi,$$

where ξ is a standard normal random variable independent of $\int_0^1 B_1(t)dt$.

The limiting distributions of the ML-based tests, M_ϕ, M_t, M_μ and $M_{\mu,t}$, are the same as those based on the least absolute deviations estimators by Herce (1996), but the test statistics are quite different. When $\alpha = 0$, h_t is a constant and hence $K\sigma^2 = 1$. In this case, assumption 2 is violated and obviously the above tests cannot be used. Therefore, it is necessary to check if the coefficient α is equal to zero before using the ML-based test statistics. This can be done easily by applying the diagnostic checking method in Li and Mak (1994) for the pure GARCH model (1.3) with the artificial observations, $\hat{\varepsilon}_t$, in Section 2.

Using the 20,000 simulated values, the $\alpha-$quantiles of the distributions of the ML-based tests are estimated. For $n = 200$, 300 and 5000, some of the empirical quantiles are summarized in Table 5 in Appendix A.

4.3 Method C

The unit root tests in the last method may not be very powerful since LS is employed. The asymptotic distributions (3.7)-(3.8) can be used to construct the unit root test

without using LS estimation. First, we can write (3.7) and (3.8) as

$$nc_1(\tilde{\phi}_{ML} - 1) \xrightarrow{\mathcal{L}} \frac{\rho \int_0^1 B_1(t)dB_1(t)}{\int_0^1 B_1^2(t)dt} + \sqrt{1-\rho^2}\frac{\int_0^1 B_1(t)dB_2(t)}{\int_0^1 B_1^2(t)dt}, \quad (4.1)$$

$$nc_1(\tilde{\phi}_{\mu,ML} - 1) \xrightarrow{\mathcal{L}} \frac{\rho \int_0^1 B_1(t)dB_1(t) - B_1(1)\int_0^1 B_1(t)dt}{[\int_0^1 B_1^2(t)dt - (\int_0^1 B_1(t)dt)^2]} \quad (4.2)$$

$$+\sqrt{1-\rho^2}[\int_0^1 B_1^2(t)dt - (\int_0^1 B_1(t)dt)^2]^{-1/2}\xi.$$

where $c_1 = \sigma F/K^{1/2}$ and $\rho^2 = 1/\sigma^2 K \in (0,1)$. Let $c_2 = c_1/\sigma$,

$$\tilde{M}_\phi = nc_1(\tilde{\phi}_{ML} - 1), \tilde{M}_t = nc_2(\frac{1}{n^2}\sum_{t=1}^n y_{t-1}^2)^{1/2}(\tilde{\phi}_{ML} - 1),$$

$$\tilde{M}_\mu = nc_1(\tilde{\phi}_{\mu,ML} - 1) \text{ and } \tilde{M}_{\mu,t} = nc_2(\frac{1}{n^2}\sum_{t=1}^n (y_{t-1} - \bar{y})^2)^{1/2}(\tilde{\phi}_{\mu,ML} - 1).$$

Then, we have the following theorem:

Theorem 4.3. Under assumptions 1-3, $\tilde{M}_t$ and $\tilde{M}_{\mu,t}$ have asymptotic distributions (4.1)-(4.2), and

$$\tilde{M}_t \xrightarrow{\mathcal{L}} \frac{\rho \int_0^1 B_1(t)dB_1(t)}{(\int_0^1 B_1^2(t)dt)^{1/2}} + \sqrt{1-\rho^2}\frac{\int_0^1 B_1(t)dB_2(t)}{(\int_0^1 B_1^2(t)dt)^{1/2}}, \quad (4.3)$$

$$\tilde{M}_{\mu,t} \xrightarrow{\mathcal{L}} \frac{\rho \int_0^1 B_1(t)dB_1(t) - B_1(1)\int_0^1 B_1(t)dt}{[\int_0^1 B_1^2(t)dt - (\int_0^1 B_1(t)dt)^2]^{1/2}} + \sqrt{1-\rho^2}\xi. \quad (4.4)$$

The asymptotic distribution of $\hat{\tau}_{AEn}$ depends on a nuisance parameter ρ. Its critical values can be obtained through the simulation method, with the estimated $\hat{\rho}$ as given in Hansen (1995), Seo (1999) and Shin and So (1999). Some critical values of $\tilde{M}_t$ and $\tilde{M}_{\mu,t}$ were given by Seo (1999). In Appendix B, Table 6, we give the critical values of $\tilde{M}_\phi$ and $\tilde{M}_\mu$. All these critical values are generated through 40,000 replications of an i.i.d. bivariate $N(0, I_2)$ process.

5 Simulation Study

5.1 Methods A versus B

In order to investigate the empirical sizes and powers of the test statistics in Theorems 4.1 and 4.2, we generate data sets from the following model:

$$y_t = \phi y_{t-1} + \varepsilon_t, \ \varepsilon_t = \eta_t \sqrt{h_t}, \ h_t = \omega + \alpha \varepsilon_{t-1}^2 + \beta h_{t-1}, \ \eta_t \sim i.i.d.N(0,1)$$

with $\phi = 0.9$, 0.95, 0.99, 1.0, $\omega = 1 - \alpha - \beta$ and $(\alpha, \beta) = (0.2, 0.7)$, $(0.3, 0.6)$ and $(0.4, 0.5)$. Each data set is estimated by model (1.1) with GARCH error (1.3) and model (1.2) with GARCH error (1.3). For model (1.1) with GARCH error (1.3), we first estimate ϕ by LS and then obtain a series of artificial observations of ε_t which are used to estimate (ω, α, β) by the IMSL subroutine DBCOAH. Using these estimators as the initial values, we obtain the ML estimator of $(\phi, \omega, \alpha, \beta)$ by the estimation procedure in Section 3. A similar estimation procedure is employed for model (1.2) with GARCH error (1.3). For each parameter vector (ϕ, α, β) or $(\mu, \phi, \alpha, \beta)$, we use 1000 independent replications. The empirical sizes and powers of eight test statistics, L_ϕ, L_t, $L_{\mu,\phi}$, $L_{\mu,t}$, M_ϕ, M_t, M_μ and $M_{\mu,t}$, are summarized for $n = 200$ and 300 in Tables 1 and 2, respectively, at the 5% significance level.

When $n = 200$, the empirical sizes of the LS-based tests, especially L_t and $L_{\mu,t}$, tend to overreject. For the ML-based tests, the sizes are closer to the nominal 5% level and powers are also acceptable as compared with those reported in other studies under i.i.d. errors (see, for example, Dickey and Fuller (1979)). When $n = 300$, all test statistics for the fitted model (1.1) with GARCH error (1.3) have similar sizes and powers. However, for the fitted model (1.2) with GARCH (1.3), the LS-based tests still tend to overreject, which is consistent with the findings in Kim and Schmidt (1993). In this case, the ML-based tests basically solve the overrejection problem. From Tables 1-2, we see that when α increases, the LS-based tests became more

sensitive, which results in rising sizes and decreasing powers. This phenomenon can be explained by the fact that when α increases, the unit root process, y_t, has more and more heavy-tailed innovations. Meanwhile, when α increases, the power of the ML-based tests improves. This is because, in this case, $\hat{\sigma}^2 \hat{K} - 1$ (or $\tilde{\sigma}^2 \tilde{K} - 1$ in the ML-based tests) can be evaluated more accurately. All these results clearly suggest that the ML-based tests are more robust and perform better than the LS-based tests.

TABLE 1

Powers and Sizes for Unit Root Processes with GARCH(1,1) error.
n=200, 1000 Replications, $\omega = 1 - \alpha - \beta$.

α	β	Test	0.90	0.95	0.99	1.0	Test	0.90	0.95	0.99	1.0
				ϕ					ϕ		
0.2	0.7	L_ϕ	.996	.848	.132	.064	$L_{\mu,\phi}$	.901	.466	.104	.064
		L_t	.994	.747	.129	.065	$L_{\mu,t}$	.830	.325	.073	.066
		M_ϕ	.799	.556	.133	.061	M_μ	.648	.335	.064	.037
		M_t	.460	.283	.132	.057	$M_{\mu,t}$	.356	.215	.070	.041
0.3	0.6	L_ϕ	.992	.739	.134	.074	$L_{\mu,\phi}$	.892	.470	.113	.077
		L_t	.987	.748	.135	.069	$L_{\mu,t}$	.819	.328	.083	.079
		M_ϕ	.890	.700	.168	.055	M_μ	.786	.483	.089	.058
		M_t	.626	.470	.170	.057	$M_{\mu,t}$	.543	.330	.097	.049
0.4	0.5	L_ϕ	.986	.737	.138	.075	$L_{\mu,\phi}$	.891	.487	.122	.083
		L_t	.981	.743	.144	.070	$L_{\mu,t}$	.810	.355	.093	.084
		M_ϕ	.934	.777	.211	.048	M_μ	.863	.578	.111	.063
		M_t	.741	.528	.212	.049	$M_{\mu,t}$	.653	.440	.119	.059

TABLE 2

Powers and Sizes for Unit Root Processes with GARCH(1,1) error.
n=300, 1000 Replications, $\omega = 1 - \alpha - \beta$.

α	β	Test	ϕ				Test	ϕ			
			0.90	0.95	0.99	1.0		0.90	0.95	0.99	1.0
0.2	0.7	L_ϕ	.999	.949	.191	.062	$L_{\mu,\phi}$	.998	.771	.142	.067
		L_t	.999	.945	.185	.063	$L_{\mu,t}$	.988	.604	.101	.073
		M_ϕ	.921	.739	.205	.052	M_μ	.843	.538	.094	.052
		M_t	.615	.381	.164	.057	$M_{\mu,t}$	.521	.306	.083	.043
0.3	0.6	L_ϕ	.999	.944	.192	.070	$L_{\mu,\phi}$	.993	.776	.161	.076
		L_t	.999	.937	.202	.070	$L_{\mu,t}$	.981	.627	.109	.073
		M_ϕ	.917	.715	.277	.042	M_μ	.947	.693	.126	.054
		M_t	.866	.472	.226	.056	$M_{\mu,t}$	.723	.465	.115	.050
0.4	0.5	L_ϕ	.999	.943	.223	.069	$L_{\mu,\phi}$	.988	.770	.171	.083
		L_t	.999	.933	.210	.068	$L_{\mu,t}$	.973	.624	.118	.079
		M_ϕ	.989	.925	.337	.043	M_μ	.975	.866	.156	.056
		M_t	.899	.673	.282	.058	$M_{\mu,t}$	.838	.582	.153	.064

5.2 Methods A versus C

We now examine the sizes and powers of the test statistics $\tilde{M}_\phi$ and $\tilde{M}_t$. The simulation experiments are similar to those in the previous section, and the results are summarized in Tables 3 and 4. The empirical sizes are slightly larger than those of method B, but comparable to those of method A. Overall, the empirical size improves quickly at $n = 300$. The powers are much better than those of method B for both $n = 200$ and 300, at the upper 10% and 5% levels, and also dominate those of method A uniformly. Such dominance is also rather substantial at the 5% and 1% levels when the sample size is only 200. This clearly suggests the usefulness of the proposed testing procedures in empirical applications. Based on these simulations, we would recommend method C if the sample size is 300 or larger and method B if the sample size is smaller than 300.

TABLE 3

Powers and sizes for Unit Root Processes with GARCH(1,1) error.

$n = 200, 1000$ Replications, $\omega = 1 - \alpha - \beta$.

			ϕ			
α	β	Test	0.90	0.95	0.99	1.00
0.2	0.7	L_ϕ	0.996	0.848	0.132	0.064
		L_t	0.994	0.747	0.129	0.065
		$\tilde{M}_\phi$	1.000	0.869	0.162	0.071
		$\tilde{M}_t$	0.999	0.865	0.200	0.080
0.3	0.6	L_ϕ	0.992	0.739	0.134	0.074
		L_t	0.987	0.748	0.135	0.069
		$\tilde{M}_\phi$	0.999	0.926	0.208	0.071
		$\tilde{M}_t$	0.998	0.910	0.262	0.080
0.4	0.5	L_ϕ	0.986	0.737	0.138	0.075
		L_t	0.981	0.743	0.144	0.070
		$\tilde{M}_\phi$	0.999	0.948	0.241	0.079
		$\tilde{M}_t$	0.999	0.937	0.291	0.078

TABLE 4

Powers and sizes for Unit Root Processes with GARCH(1,1) error.

$n = 300, 1000$ Replications, $\omega = 1 - \alpha - \beta$.

			ϕ			
α	β	Test	0.90	0.95	0.99	1.00
0.2	0.7	L_ϕ	0.999	0.949	0.191	0.062
		L_t	0.999	0.945	0.185	0.063
		$\tilde{M}_\phi$	1.000	1.000	0.236	0.065
		$\tilde{M}_t$	1.000	0.986	0.270	0.064
0.3	0.6	L_ϕ	0.999	0.944	0.192	0.070
		L_t	0.999	0.937	0.202	0.070
		$\tilde{M}_\phi$	1.000	1.000	0.291	0.068
		$\tilde{M}_t$	1.000	0.994	0.355	0.065
0.4	0.5	L_ϕ	0.999	0.943	0.223	0.069
		L_t	0.999	0.933	0.210	0.068
		$\tilde{M}_\phi$	1.000	1.000	0.363	0.063
		$\tilde{M}_t$	1.000	1.000	0.449	0.062

6 Technical Proofs

Lemma 6.1. Under assumption 1, the process h_t and ε_t defined by model (1.3) are strictly stationary and ergodic, and have the expansions

$$h_t = w \left[1 + \sum_{k=1}^{\infty} \prod_{i=1}^{k} (\alpha \eta_{t-i}^2 + \beta) \right] \text{ a.s.} \tag{6.1}$$

and

$$\varepsilon_t = w^{1/2} \eta_t \left[1 + \sum_{k=1}^{\infty} \prod_{i=1}^{k} (\alpha \eta_{t-i}^2 + \beta) \right]^{1/2} \text{ a.s..} \tag{6.2}$$

Denote $\mathcal{F}_t$ as the σ-field generated by $\{\eta_t, \eta_{t-1}, \cdots\}$. Then ε_{t-1} and h_t are measurable with respect to $\mathcal{F}_{t-1}$.

Proof of Lemma 6.1. It comes straightforwardly from Theorem 2 in Nelson (1990) (for another expansion, see also Ling and Li, 1997). □

Lemma 6.2. Suppose that the process ε_t is generated by model (1.3) and assumptions 1-3 are satisfied. Then

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{[n\tau]} \left(\varepsilon_t, \frac{\varepsilon_t}{h_t} - \frac{1}{h_t} \left(\frac{\varepsilon_t^2}{h_t} - 1 \right) \sum_{k=1}^{t-1} \beta^{k-1} \varepsilon_{t-k} \right) \xrightarrow{\mathcal{L}} (w_1(\tau), w_2(\tau)) \text{ in } D \times D,$$

where $(w_1(\tau), w_2(\tau))$ is a bivariate Brownian motion with mean zero and covariance $\tau\Omega$, and Ω is defined in Theorem 3.2.

Proof. See Lemma 3.2 in Ling and Li (2000). □

Proof of Theorem 2.1. By Lemma 6.1 and the continuous mapping theorem, it is easy to show

$$
\begin{aligned}
n(\hat{\phi}_{LS} - 1) &= \left(\frac{1}{n^2} \sum_{t=1}^{n} y_{t-1}^2 \right)^{-1} \left(\frac{1}{n} \sum_{t=1}^{n} y_{t-1} \varepsilon_t \right) \\
&\xrightarrow{\mathcal{L}} \frac{\int_0^1 w_1(t) dw_1(t)}{\int_0^1 w_1^2(t) dt},
\end{aligned}
$$

16

where $w_1(t)$ is defined by Lemma 6.2. Let $B_1(t) = w_1(t)/\sigma$, then

$$n(\hat{\phi}_{LS} - 1) \xrightarrow{\mathcal{L}} \frac{\int_0^1 B_1(t)dB_1(t)}{\int_0^1 B_1^2(t)dt}.$$

Similarly, we can show that part (b) of Theorem 2.1 holds. $\qquad\square$

Proof of Lemma 2.1. First we note that, by Theorem 2.1,

$$
\begin{aligned}
\hat{\varepsilon}_t &= y_t - \hat{\phi}_{LS} y_{t-1} \\
&= \varepsilon_t - [n(\hat{\phi}_{LS} - 1)]\left[\frac{1}{\sqrt{n}}y_{t-1}\right]\frac{1}{\sqrt{n}} = \varepsilon_t + O_p\left(\frac{1}{\sqrt{n}}\right). \\
\hat{h}_t &= \omega + \alpha\hat{\varepsilon}_{t-1} + \beta\hat{h}_{t-1} = \omega + \alpha\sum_{i=1}^t \beta^{i-1}\hat{\varepsilon}_{t-i}^2 + \beta^t\hat{h}_0 \\
&= \omega + \alpha\sum_{i=1}^t \beta^{i-1}\left[\varepsilon_{t-i} + O_p(\frac{1}{\sqrt{n}})\right]^2 + \beta^t\hat{h}_0 \\
&= h_t + O_p\left(\frac{1}{\sqrt{n}}\right) + \xi_0\beta^t,
\end{aligned}
$$

where $\xi_0 = 1 - h_0$ and $O_p(\cdot)$ holds uniformly in t. Thus,

$$
\begin{aligned}
&\sup_{\delta\in\Theta}|\frac{1}{n}\sum_{t=1}^n l_t(\hat{\varepsilon}_t|\delta) - \frac{1}{n}\sum_{t=1}^n l_t(\varepsilon_t|\delta)| \le \sup_{\delta\in\Theta}\frac{1}{n}\sum_{t=1}^n |l_t(\hat{\varepsilon}_t|\delta) - l_t(\varepsilon_t|\delta)| \\
&= \sup_{\delta\in\Theta}\frac{1}{n}\sum_{t=1}^n |-\frac{1}{2}\ln\frac{h_t + O_p(\frac{1}{\sqrt{n}}) + \xi_0\beta^t}{h_t} - \frac{1}{2}[\frac{\varepsilon_t^2}{h_t} - \frac{\varepsilon_t^2 + O_p(\frac{1}{\sqrt{n}})}{h_t + O_p(\frac{1}{\sqrt{n}}) + \xi_0\beta^t}]| \\
&\le \sup_{\delta\in\Theta}\frac{1}{2n}\sum_{t=1}^n \left|\ln\left[1 + O_p(\frac{1}{\sqrt{n}}) + \frac{\xi_0\beta^t}{h_t}\right]\right| \\
&\quad + \sup_{\delta\in\Theta}\frac{1}{2n}\sum_{t=1}^n \left|\frac{[O_p(\frac{1}{\sqrt{n}}) + \xi_0\beta^t]\varepsilon_t^2 + O_p(\frac{1}{\sqrt{n}})h_t}{h_t[h_t + O_p(\frac{1}{\sqrt{n}}) + \xi_0\beta^t]}\right| \\
&= O_p\left(\frac{1}{\sqrt{n}}\right) = o_p(1),
\end{aligned}
$$

where $o_p(\cdot)$ holds uniformly in t. Similarly, we can show that (b) and (c) hold. When $\hat{\varepsilon}_t = y_t - \tilde{\mu}_{LS} - \tilde{\phi}_{LS}y_{t-1}$, the proofs are similar and hence the details are omitted.

$\qquad\square$

Proof of Theorem 2.2. By Lemma 2.1, (a) and (b) come directly from Theorems 2 and 3 in Lee and Hansen (1994) (see also Lumsdaine, 1996). □

The following lemma is obvious and hence the detials are omitted (see also Ling and Li (2000)).

Lemma 6.3. Let $h_{t-k,l} = w[1+\sum_{j=1}^{l-k}\prod_{i=1}^{j}(\alpha\eta_{t-i-k}^2+\beta)]$ and $\varepsilon_{t-k,l} = \eta_{t-k}\sqrt{h_{t-k,l}}$. If assumption 1 holds, then (1) for any k,

$$(a) \quad h_t = \frac{\omega}{1-\beta} + \alpha\sum_{k=1}^{\infty}\beta^{k-1}\varepsilon_{t-k}^2 \text{ a.s.;}$$

$$(b) \quad \frac{|\varepsilon_{t-k}|}{\sqrt{h_t}} = O(\beta^{-(k-1)/2}) \, a.s.;$$

$$(c) \quad \frac{h_{t-k}}{h_t} \le [\prod_{i=1}^{k}(\alpha\eta_{t-i}^2+\beta)]^{-1} \, a.s.,$$

(2) for $k = 0, 1, \cdots, l-1$,

$$(a) \quad \frac{|\varepsilon_{t-k,l}|}{\sqrt{h_{t,l}}} = O(\beta^{-(k-1)/2});$$

$$(b) \quad \frac{h_{t-k,l}}{h_{t,l}} \le [\prod_{i=1}^{k}(\alpha\eta_{t-i}^2+\beta)]^{-1};$$

$$(c) \quad E|h_{t-k} - h_{t-k,l}| = O(\rho^{l-k+1}) \text{ with } 0 < \rho < 1.$$

Lemma 6.4 (Ling and Li, Theorem 2.1, 1998) Let $\{S_n(t), 0 \le t \le 1\}$ and $\{\xi_k, k = 1, 2, \cdots\}$ be two sequences of random processes such that

$$(a) \quad S_n(t) \xrightarrow{\mathcal{L}} S(t) \text{ in } D;$$

$$(b) \quad \frac{1}{\sqrt{n}}\sum_{k=1}^{[nt]}\xi_k \xrightarrow{\mathcal{L}} \xi(t) \text{ in } D;$$

$$(c) \quad \max_{1\le k\le n}|\xi_k|/\sqrt{n} \xrightarrow{p} 0;$$

$$(d) \quad \frac{1}{n}\sum_{t=1}^{n}|\xi_t| \text{ is bounded in probability,}$$

18

and almost all trajectories of $S(t)$ and $\xi(t)$ are continuous. Then

$$\sup_{0 \leq t \leq 1} |\frac{1}{n} \sum_{k=1}^{[nt]} S_n(\frac{k}{n})\xi_k| \xrightarrow{p} 0 \text{ as } n \to \infty.$$

Lemma 6.5. Under assumptions 1-3,

$$(a) \quad \frac{1}{n^{3/2}} \sum_{t=1}^{n} \frac{\partial^2 l_t}{\partial\phi\partial\delta} \xrightarrow{p} 0;$$

$$(b) \quad (n^{1/2} N_n)^{-1} \sum_{t=1}^{n} \frac{\partial^2 l_t}{\partial\phi_\mu\partial\delta'} \xrightarrow{p} 0;$$

$$(c) \quad \frac{1}{n} \sum_{t=1}^{n} \frac{\partial^2 l_t}{\partial\delta\partial\delta'} \xrightarrow{p} -E[\frac{\partial^2 l_t}{\partial\delta\partial\delta'}],$$

where $w_1(t)$ is a Brownian motion with covariance $t\sigma^2$, N_n is defined in Theorem 2.1, and F is defined in Theorem 3.2.

Proof of Lemma 6.5. Note that

$$\begin{aligned}
\frac{\partial^2 l_t}{\partial\phi\partial\delta} &= -\frac{y_{t-1}\varepsilon_t}{h_t^2}\frac{\partial h_t}{\partial\delta} + \frac{1}{2h_t}(\frac{\varepsilon_t^2}{h_t} - 1)\frac{\partial^2 h_t}{\partial\phi\partial\delta} \\
&\quad -\frac{1}{2h_t^2}(1 - \frac{2\varepsilon_t^2}{h_t})(\sum_{i=1}^{t-1}\beta^{i-1}y_{t-i-1}\varepsilon_{t-i})\frac{\partial h_t}{\partial\delta} \\
&= I_{1t} + I_{2t} - I_{3t}.
\end{aligned} \tag{6.3}$$

Let $\xi_t = [\varepsilon_t, \varepsilon_t(\partial h_t/\partial\delta)/h_t^2]'$. Then $S_{[n\tau]} = \sum_{t=1}^{[n\tau]} \xi_t$ is a martingale. By Lemma 6.3, it is not difficult to show that

$$E(\frac{1}{h_t^3}\frac{\partial h_t}{\partial\delta}\frac{\partial h_t}{\partial\delta'}) \leq \frac{1}{\omega}E(\frac{1}{h_t^2}\frac{\partial h_t}{\partial\delta}\frac{\partial h_t}{\partial\delta'}) < \infty$$

and hence we can show that $\Omega = E(\xi_t\xi_t') < \infty$. In a similar manner to the proof of Lemma 6.2 in Ling and Li (2000), we can show that

$$\frac{1}{\sqrt{n}}S_{[n\tau]} \xrightarrow{\mathcal{L}} W(t) \text{ in } D^2, \tag{6.4}$$

where $W(t)$ is a bivariate Brownian motion. Since ξ_t is a martingale difference, we can use Theorem 2.1 in Kurtz and Protter (1991) such that $n^{-1}\sum_{t=1}^{n} y_{t-1}(\varepsilon_t(\partial h_t/\partial\delta)/$

19

$h_t^2) = O_p(1)$. Thus,

$$\frac{1}{n^{3/2}} \sum_{t=1}^{n} I_{1t} = \frac{1}{n^{3/2}} \sum_{t=1}^{n} y_{t-1}[\varepsilon_t(\partial h_t/\partial \delta)/h_t^2] \xrightarrow{p} 0. \tag{6.5}$$

Now we consider the last term I_{3t}:

$$I_{3t} = \frac{1}{2h_t^2}(1 - \frac{2\varepsilon_t^2}{h_t})(\sum_{i=1}^{t-1} \beta^{i-1}\varepsilon_{t-i})\frac{\partial h_t}{\partial \delta}y_{t-1} - \frac{1}{2h_t^2}(1 - \frac{2\varepsilon_t^2}{h_t})(\sum_{i=1}^{t-1} \beta^{i-1}r_{t,i}\varepsilon_{t-i})\frac{\partial h_t}{\partial \delta}$$
$$= T_{1t} + T_{2t}. \tag{6.6}$$

Let $T_{1t} = \xi_t^* y_{t-1}$ and $S_{[n\tau]}^* = \sum_{t=1}^{[n\tau]} \xi_t^*$. In a similar manner to the proof of Lemma 3.4 in Ling and Li (2000), we can show that

$$\frac{1}{\sqrt{n}} S_{[n\tau]}^* \xrightarrow{\mathcal{L}} W^*(t) \text{ in } D, \tag{6.7}$$

where $W^*(t)$ is a Brownian motion. By Theorem 2.1 in Ling and Li (1998) and Lemma 6.2, we have $n^{-3/2} \sum_{t=1}^{n} T_{1t} = o_p(1)$. By Lemma 6.3.(1).(b),

$$\frac{1}{n^{3/2}} \sum_{t=1}^{n} E||T_{2t}||$$
$$\leq \frac{c}{n^{3/2}} \sum_{t=1}^{n} E[||\frac{1}{h_t}(\sum_{i=1}^{t-1} \beta^{i-1}r_{t,i}\varepsilon_{t-i})\frac{\partial h_t}{\partial \delta}||]$$
$$\leq \frac{c}{n^{3/2}} \sum_{t=1}^{n} [\sum_{i=1}^{t-1} \beta^{i-1} \sqrt{E(\frac{r_{t,i}^2\varepsilon_{t-i}^2}{h_t})}](E||\frac{\partial h_t}{\partial \delta}||^2)^{1/2}$$
$$= O(\frac{1}{n^{3/2}} \sum_{t=1}^{n} [\sum_{i=1}^{t-1} \beta^{(i-1)/2} \sqrt{Er_{t,i}^2}](E||\frac{\partial h_t}{\partial \delta}||^2)^{1/2}$$
$$= O(\frac{1}{\sqrt{n}}) = o(1), \tag{6.8}$$

where c is a constant and the last equation holds by $E||\partial h_t/\partial \delta||^2 < \infty$. Thus,

$$\frac{1}{n^{3/2}} \sum_{t=1}^{n} I_{3t} = o_p(1). \tag{6.9}$$

Similar to (6.10), we can show that

$$\frac{1}{n^{3/2}} \sum_{t=1}^{n} I_{2t} = o_p(1). \tag{6.10}$$

20

By (6.3), (6.6) and (6.9)-(6.10), we complete the proof of (a). The proof of (b) is similar to (a) and hence the details are omitted. Part (c) comes directly from Lemma 11 in Lee and Hansen (1992). This completes the proof. $\quad\square$

Proof of Theorem 3.2. (a) and (b) come directly from Ling and Li (2000). For $\sqrt{n}(\hat{\delta}_{ML} - \delta)$ and $\sqrt{n}(\hat{\delta}_{\mu,ML} - \delta)$, their asymptotic distributions come directly from Lemma 9 in Lee and Hansen (1996), Theorem 3.1(e) and (3.4). $\quad\square$

Proof of Theorem 3.1. We first show that

$$\sup_{\lambda \in \Theta_n} \Big| \frac{1}{n^2} \sum_{t=1}^{n} \Big(\frac{\partial^2 l_t}{\partial \phi^2} \Big|_{\lambda=\tilde{\lambda}} - \frac{\partial^2 l_t}{\partial \phi^2} \Big) \Big| = o_p(1). \tag{6.11}$$

By direct differentiation,

$$\begin{aligned}
\frac{\partial^2 l_t}{\partial \phi^2} &= -\frac{1}{h_t}\Big(\frac{\partial \varepsilon_t}{\partial \phi}\Big)^2 - \frac{\varepsilon_t^2}{2h_t^3}\Big(\frac{\partial h_t}{\partial \phi}\Big)^2 \\
&\quad + \Big[\Big(\frac{\varepsilon_t^2}{h_t} - 1\Big)\frac{\partial}{\partial \phi}\Big[\frac{1}{2h_t}\frac{\partial h_t}{\partial \phi}\Big] + \frac{2\varepsilon_t}{h_t^2}\frac{\partial \varepsilon_t}{\partial \phi}\frac{\partial h_t}{\partial \phi} - \frac{\varepsilon_t}{h_t}\frac{\partial^2 \varepsilon_t}{\partial \phi^2} \Big] \\
&= -I_{1t} - I_{2t} + I_{3t}. \tag{6.12}
\end{aligned}$$

In the following, all $O_p(1)$ and $o_p(1)$ hold uniformly in Θ_n.

As in the proof of Lemma 2.1, we can show that

$$\varepsilon_t(\tilde{\phi}) = \varepsilon_t + O_p(\frac{1}{\sqrt{n}}), \tag{6.13}$$

$$h_t(\tilde{\lambda}) = h_t + O_p(\frac{1}{\sqrt{n}})h_t + O_p(\frac{1}{\sqrt{n}}) + \xi_0\beta^t, \tag{6.14}$$

$$\frac{1}{h_t(\tilde{\lambda})} - \frac{1}{h_t} = O_p(\frac{1}{\sqrt{n}}) + \xi_0\beta^t, \tag{6.15}$$

$$\frac{1}{h_t(\tilde{\lambda})}\frac{\partial h_t(\lambda)}{\partial \phi} - \frac{1}{h_t}\frac{\partial h_t}{\partial \phi} = O_p(1)\Big(\frac{\partial h_t(\lambda)}{\partial \phi} - \frac{\partial h_t}{\partial \phi}\Big) + \Big[O_p(\frac{1}{\sqrt{n}}) + \xi_0\beta^t\Big]\frac{\partial h_t}{\partial \phi}. \tag{6.16}$$

By Lemma 6.3 (1)-(b),

$$\Big| \frac{1}{h_t(\tilde{\lambda})}\frac{\partial h_t(\lambda)}{\partial \phi} \Big| \le \sum_{i=1}^{t-1} \rho^i |y_{t-i-1}| \tag{6.17}$$

21

$$\left| \frac{\partial h_t(\lambda)}{\partial \phi} - \frac{\partial h_t}{\partial \phi} \right| \leq O(\frac{1}{\sqrt{n}}) \sum_{i=1}^{t-1} \rho^i |\varepsilon_{t-i} y_{t-i-1}| + O(\frac{1}{n}) \sum_{i=1}^{t-1} \rho^i y_{t-i-1}^2$$
$$= A_{1t} + A_{2t} \tag{6.18}$$

where $\rho \in (0,1)$ and ρ is independent of λ. Since $n^{-1/2} \max_{1 \leq t \leq n} |y_t| = O_p(1)$,

$$\frac{1}{n^2} \sum_{t=1}^{n} \eta_t^2 \sum_{i=1}^{t-1} \rho^i |y_{t-i-1}| A_{1t} = \frac{O_p(1)}{n\sqrt{n}} \sum_{t=1}^{n} (\eta_t^2 \sum_{i=1}^{t-1} \rho^i |\varepsilon_{t-i}|) = o_p(1) \tag{6.19}$$

$$\frac{1}{n^2} \sum_{t=1}^{n} \eta_t^2 \sum_{i=1}^{t-1} \rho^i |y_{t-i-1}| A_{2t} = \frac{O_p(1)}{n^{3/2}} \sum_{t=1}^{n} \eta_t^2 = o_p(1). \tag{6.20}$$

Thus, by (6.18)-(6.20),

$$\frac{1}{n^2} \sum_{t=1}^{n} \eta_t^2 \frac{\partial h_t(\lambda)}{\partial \phi} \left[\frac{\partial h_t(\lambda)}{\partial \phi} - \frac{\partial h_t}{\partial \phi} \right] = o_p(1) \tag{6.21}$$

$$\frac{1}{n^{5/2}} \sum_{t=1}^{n} \eta_t^2 \sum_{i=1}^{t-1} \rho^i |y_{t-i-1}| \sum_{i=1}^{t-1} \rho^i |\varepsilon_{t-i} y_{t-i-1}| = \frac{O_p(1)}{n^{3/2}} \sum_{t=1}^{n} \eta_t^2 \sum_{i=1}^{t-1} \rho^i |\varepsilon_{t-i}| = o_p(1) \tag{6.22}$$

$$\frac{1}{n^2} \sum_{t=1}^{n} \beta^t \eta_t^2 \sum_{i=1}^{t-1} \rho^i |y_{t-i-1}| \sum_{i=1}^{t-1} \rho^i |\varepsilon_{t-i} y_{t-i-1}| = \frac{O_p(1)}{n} \sum_{t=1}^{n} \beta^t \eta_t^2 \sum_{i=1}^{t-1} \rho^i |\varepsilon_{t-i}| = o_p(1) \tag{6.23}$$

because $E(\sum_{t=1}^{\infty} \beta^t \eta_t^2 \sum_{i=1}^{\infty} \rho^i |\varepsilon_{t-i}|)^2 < \infty$. By (6.22)-(6.23),

$$\frac{1}{n^2} \sum_{t=1}^{n} \eta_t^2 [O_p(\frac{1}{\sqrt{n}}) + \xi_0 \beta^t] \frac{\partial h_t(\lambda)}{\partial \phi} \frac{\partial h_t}{\partial \phi} = o_p(1). \tag{6.24}$$

By (6.21)-(6.24), we have

$$\frac{1}{n^2} \sum_{t=1}^{n} \frac{\varepsilon_t^2}{h_t^2} \left| \left(\frac{1}{h_t(\tilde{\lambda})} \frac{\partial h_t(\tilde{\lambda})}{\partial \phi} \right)^2 - \left(\frac{1}{h_t} \frac{\partial h_t}{\partial \phi} \right)^2 \right|$$
$$\leq \frac{1}{n^2} \sum_{t=1}^{n} \eta_t^2 (\sum_{i=1}^{t-1} \rho^i |y_{t-i-1}|) \left| \frac{1}{h_t(\tilde{\lambda})} \frac{\partial h_t(\lambda)}{\partial \phi} - \frac{1}{h_t} \frac{\partial h_t}{\partial \phi} \right| = o_p(1). \tag{6.25}$$

By (6.13)- (6.15), it is straightforward to show that

$$\frac{1}{n^2} \sum_{t=1}^{n} \left[\frac{\varepsilon_t^2(\tilde{\phi})}{h_t(\tilde{\lambda})} - \frac{\varepsilon_t^2}{h_t} \right] \left(\frac{1}{h_t(\tilde{\lambda})} \frac{\partial h_t(\lambda)}{\partial \phi} \right)^2$$
$$\leq O_p(1) \left[\frac{1}{n} \sum_{t=1}^{n} |\varepsilon_t^2(\tilde{\phi}) - \varepsilon_t^2| + \frac{1}{n} \sum_{t=1}^{n} \varepsilon_t^2 \left| \frac{1}{h_t(\tilde{\lambda})} - \frac{1}{h_t} \right| \right]$$
$$= o_p(1). \tag{6.26}$$

22

Thus, by (6.25)-(6.26), we have $n^{-2} \sum_{t=1}^{n} I_{2t} = o_p(1)$. Similarly, we can prove that $n^{-2} \sum_{t=1}^{n} I_{1t} = o_p(1)$ and $n^{-2} \sum_{t=1}^{n} I_{3t} = o_p(1)$. Thus, (6.11) holds. Using the same method, we can show

$$\sup_{\lambda \in \Theta_n} \left| \frac{1}{n^2} \sum_{t=1}^{n} \left(\frac{\partial^2 l_t}{\partial \phi \partial \delta} \Big|_{\lambda = \tilde{\lambda}} - \frac{\partial^2 l_t}{\partial \phi \partial \delta} \right) \right| = o_p(1) \tag{6.27}$$

and

$$\sup_{\lambda \in \Theta_n} \left| \frac{1}{n^2} \sum_{t=1}^{n} \left(\frac{\partial^2 l_t}{\partial \delta \partial \delta} \Big|_{\lambda = \tilde{\lambda}} - \frac{\partial^2 l_t}{\partial \delta \partial \delta} \right) \right| = o_p(1) \tag{6.28}$$

By (6.27) and (6.28), we complete the proof of (a). (b) comes directly from (a). The proofs of (c)-(d) are similar and hence the details are omitted. This completes the proof. $\qquad \square$

Appendix A: Critical Values of M_ϕ, M_t, M_μ and $M_{\mu,t}$

TABLE 5

Empirical Critical Values for M_ϕ, M_t, M_μ and $M_{\mu,t}$

Statistic	n	empirical quantiles							
		.010	.025	.050	.100	.900	.950	.975	.990
M_ϕ	200	-9.78	-6.71	-4.77	-3.09	2.42	3.43	4.40	5.59
	300	-8.96	-6.17	-4.36	-2.92	2.47	3.49	4.52	5.87
	5000	-6.53	-4.98	-3.68	-2.60	2.56	3.73	5.01	6.70
M_t	200	-2.37	-2.00	-1.66	-1.28	1.20	1.56	1.90	2.27
	300	-2.39	-1.98	-1.64	-1.28	1.24	1.60	1.91	2.28
	5000	-2.31	-1.96	-1.64	-1.28	1.24	1.60	1.92	2.31
M_μ	200	-16.41	-11.93	-9.03	-6.25	2.96	4.20	5.30	6.83
	300	-15.04	-11.08	-8.43	-5.73	3.12	4.36	5.56	7.09
	5000	-8.79	-6.90	-5.44	-3.95	3.86	5.31	6.69	8.74
$M_{\mu,t}$	200	-3.54	-2.88	-2.33	-1.81	1.05	1.42	1.75	2.18
	300	-3.38	-2.69	-2.22	-1.70	1.10	1.46	1.80	2.24
	5000	-2.31	-1.95	-1.67	-1.29	1.28	1.64	1.94	2.31

Note that these critical values are different from those in Herce (1996). When $n = 5000$, the critical values of M_ϕ and M_μ are almost the same as those in Herce (1996) and for M_t and $M_{\mu,t}$, the critical values are very close to those of the standard normal distribution.

Appendix B: Critical Values of $\tilde{M}_\phi$ and $\tilde{M}_\mu$

TABLE 6
Empirical Critical Values for $\tilde{M}_\phi$ and $\tilde{M}_\mu$

$\tilde{M}_\phi$	Significance	empirical quantiles				ρ				
	level	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
	0.01	-7.2048	-7.7703	-8.4536	-9.1138	-9.8088	-10.4791	-11.2213	-12.0237	-12.6903
$n = 200$	0.05	-4.1400	-4.5243	-4.9235	-5.3093	-5.7489	-6.1933	-6.6477	-7.0860	-7.5276
	0.10	-2.8947	-3.1637	-3.4493	-3.7311	-4.0221	-4.3391	-4.6455	-4.9472	-5.3329
	0.01	-7.1767	-7.8657	-8.5276	-9.2474	-9.9391	-10.6458	-11.4435	-12.1834	-12.7046
$n = 300$	0.05	-4.1873	-4.5845	-4.9897	-5.3982	-5.8291	-6.2552	-6.6822	-7.1470	-7.6011
	0.10	-2.9175	-3.1909	-3.4847	-3.7530	-4.0877	-4.3808	-4.7071	-5.0605	-5.3862
	0.01	-7.1858	-7.8878	-8.6966	-9.3821	-10.1568	-10.9151	-11.6742	-12.4343	-13.3237
$n = 500$	0.05	-4.1353	-4.4981	-4.9380	-5.3618	-5.7809	-6.2036	-6.6221	-7.0833	-7.5942
	0.10	-2.8689	-3.1488	-3.4230	-3.7228	-4.0261	-4.3353	-4.6831	-5.0267	-5.3624
	0.01	-7.2306	-7.8792	-8.5188	-9.1402	-9.9170	-10.6027	-11.2109	-11.9214	-12.6366
$n = 1000$	0.05	-4.1777	-4.5854	-4.9991	-5.4000	-5.8436	-6.3073	-6.7436	-7.1699	-7.5716
	0.10	-2.8998	-3.1848	-3.4805	-3.7769	-4.0834	-4.4002	-4.7277	-5.0657	-5.4198
	0.01	-7.1929	-7.8418	-8.6367	-9.2195	-9.9625	-10.6262	-11.2750	-12.0096	-12.7079
$n = 2000$	0.05	-4.1432	-4.5507	-4.9300	-5.3355	-5.7320	-6.1619	-6.6188	-7.0367	-7.5155
	0.10	-2.8486	-3.1393	-3.4129	-3.7170	-3.9937	-4.3191	-4.6449	-4.9544	-5.3103

$\tilde{M}_\mu$	Significance					ρ				
	level	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
	0.01	-11.5202	-11.3406	-11.3849	-11.7817	-12.5982	-13.6858	-15.0089	-16.6388	-18.2212
$n = 200$	0.05	-8.0208	-7.9747	-8.1686	-8.5588	-9.1045	-9.8264	-10.7116	-11.6695	-12.7012
	0.10	-6.4842	-6.4935	-6.6774	-7.0174	-7.5152	-8.0952	-8.7411	-9.4588	-10.2230
	0.01	-11.7107	-11.5681	-11.7253	-12.2209	-13.0345	-14.1097	-15.4398	-17.0025	-18.6410
$n = 300$	0.05	-8.1237	-8.0775	-8.2456	-8.6434	-9.1668	-9.8984	-10.7341	-11.7156	-12.7288
	0.10	-6.5140	-6.5484	-6.7104	-7.0366	-7.5086	-8.0909	-8.7602	-9.4833	-10.2556
	0.01	-11.9368	-11.8154	-11.9275	-12.3257	-13.1836	-14.4093	-15.8985	-17.4551	-18.9715
$n = 500$	0.05	-8.1980	-8.1506	-8.3142	-8.7487	-9.3397	-10.1524	-11.0065	-11.9671	-13.0069
	0.10	-6.5390	-6.5833	-6.7855	-7.1287	-7.6290	-8.2365	-8.8857	-9.6247	-10.3914
	0.01	-11.9275	-11.8197	-11.9128	-12.2675	-13.0854	-14.1518	-15.4855	-16.9825	-18.4980
$n = 1000$	0.05	-8.2414	-8.1764	-8.3328	-8.7073	-9.2618	-10.0293	-10.9329	-11.8729	-12.9321
	0.10	-6.5710	-6.5899	-6.7825	-7.1243	-7.6151	-8.1727	-8.8065	-9.5302	-10.3156
	0.01	-12.2837	-12.0579	-12.0785	-12.3478	-13.1710	-14.2054	-15.6072	-17.1130	-18.6337
$n = 2000$	0.05	-8.3099	-8.2566	-8.3617	-8.7794	-9.3596	-10.0190	-10.8896	-11.8102	-12.7995
	0.10	-6.6327	-6.6435	-6.8519	-7.1941	-7.6634	-8.2416	-8.9093	-9.6205	-10.3788

REFERENCES

Billingsley, P. (1968). *Convergence of Probability Measures.* New York: Wiley.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, **31**, 307-327.

Bollerslev, T., Chou, R.Y. and Kroner, K.F. (1992). ARCH modeling in finance. *Journal of Econometrics*, **52**, 5-59 .

Bollerslev, T., Engle, R.F. and Nelson, D.B. (1994). ARCH model, in *Handbook of Econometrics* **IV**, 2959-3038, R.F. Engle and D.C. McFadden (eds); Amsterdam: Elsevier Science.

Chan, N.H. and Wei, C.Z. (1988). Limiting distributions of least squares estimates of unstable autoregressive processes. *Ann. Statist.*, **16**, 367-401.

Dickey, D.A. and Fuller, W.A.(1979). Distribution of the estimators for autoregressive time series with a unit root. *J. Amer. Statist. Assoc.*, **74**, 427-431.

Fuller, W.A. (1976). *Introduction to Statistical Time Series.* New York: Wiley.

Hall, P. and Heyde, C. (1968). *Martingale Limit Theory and Its Applications.* Academic Press: San Diego.

Herce, M.A. (1996). Asymptotic theory of LAD estimation in a unit root process with finite variance errors. *Econometric Theory*, **12**, 129-53.

Kim, K. and Schmidt, P. (1993). Unit root tests with conditional heteroskedasticity. *Journal of Econometrics*, **59**, 287-300.

Kurtz, T.G. and Protter, P. (1991). Weak limit theorems to stochastic integrals and stochastic differential equations. *Ann. Probability*, **19**, 1035-1070.

Lee, S.-W. and Hansen, R.E. (1994). Asymptotic theory for GARCH(1,1) quasi-maximum likelihood estimator. *Econometric Theory*, **10**, 29-52.

Li, W.K., Ling, S. and McAleer, M. (1999). A survey of recent theoretical results for time series models with GARCH errors. To appear in *Journal of Economic Surveys.*

Li, W.K. and Mak, T.K. (1994). On the squared residual autocorrelations in nonlinear time series modeling. *Journal of Time Series Analysis*, **15**, 627-636.

Ling, S. and Li, W.K. (1997) On fractionally integrated autoregressive moving-average time series models with conditional heteroscedasticity. *J. Amer. Statist. Assoc.*, **92**, 1184-1194.

Ling, S. and Li, W.K. (1998). Limiting distributions of maximum likelihood estimators for unstable autoregressive moving-average time series with GARCH errors. *Ann. Statist.*, **26**, 84-125.

Ling, S. and Li, W.K. (2000). Asymptotic inference for unit root with GARCH (1, 1) errors. Technical report. Department of Mathematics, Hong Kong University of Science and Technology.

Lucas, A. (1995). Unit root tests based on M estimators. *Econometric Theory*, **11**, 331-346.

Lumsdaine, R. (1996). Consistency and asymptotic normality of the quasi-maximum likelihood estimator in IGARCH(1,1) and covariance stationary GARCH(1,1) Models. *Econometrica*, **3**, 575-596.

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

Pantula, S.G. (1986). Comment on modeling the persistence of conditional variance. *Econometric Reviews*, **5**, 71-74.

Pantula, S.G. (1989). Estimation of autoregressive models with ARCH errors. *Sankhya B*, **50**, 119-38.

Peters, T.A. and Veloce, W. (1988). Robustness of unit root tests in ARMA models with generalized ARCH errors. Unpublished paper, Brock University, St. Catherine, Canada.

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

Phillips, P.C.B. (1989). Partially identified econometric models. *Econometric Theory*, **5**, 181-240.

Phillips, P.C.B. and Durlauf, S.N. (1986). Multiple time series regression with integrated processes. *Review of Economic Studies*, **LIII**, 473-495.

Weiss. A.A. (1986). Asymptotic theory for ARCH models: estimation and testing. *Econometric Theory*, **2**, 107-131.

Yap, S.F. and Reinsel, G.C. (1995). Estimation and testing for unit roots in a partially nonstationary vector autoregressive moving average model. *J. Amer. Statist. Assoc.*, **90**, 253-267.