Econometric Theory, **25**, 2009, 1208–1227. Printed in the United States of America.
doi:10.1017/S0266466608090488

LEAST ABSOLUTE DEVIATION ESTIMATION FOR UNIT ROOT PROCESSES WITH GARCH ERRORS

GUODONG LI
Nanyang Technological University
and
University of Hong Kong

WAI KEUNG LI
University of Hong Kong

This paper considers a local least absolute deviation estimation for unit root processes with generalized autoregressive conditional heteroskedastic (GARCH) errors and derives its asymptotic properties under only finite second-order moment for both errors and innovations. When the innovations are symmetrically distributed, the asymptotic distribution of the estimated unit root is shown to be a functional of a bivariate Brownian motion, and then two unit root tests are derived. The simulation results demonstrate that the tests outperform those based on the Gaussian quasi maximum likelihood estimators with heavy-tailed innovations and those based on the simple least absolute deviation estimators.

1. INTRODUCTION

In this paper we consider autoregressive (AR) unit root processes with generalized autoregressive conditional heteroskedastic (GARCH) errors,

$$\Delta y_t = \phi y_{t-1} + \mu + \sum_{i=1}^{k-1} \psi_i \Delta y_{t-i} + e_t, \tag{1}$$

$$e_t = \varepsilon_t \sqrt{h_t}, \qquad h_t = \omega + \sum_{i=1}^{p} \alpha_i e_{t-i}^2 + \sum_{j=1}^{q} \beta_j h_{t-j}, \tag{2}$$

where k, p, q are known nonnegative integers, $\Delta y_t = y_t - y_{t-1}$, $\omega > 0$, $\alpha_i \geq 0, i = 1, \ldots, p$, $\beta_j \geq 0, j = 1, \ldots, q$, $k \geq 1$, $k = 1$ refers to the pure unit root case, and innovations $\{\varepsilon_t\}$ are independent and identically distributed (i.i.d.) with mean zero and finite variance. The GARCH models were proposed by Bollerslev (1986) and are usually considered in modeling the phenomena of time-varying conditional

We thank Bruce Hansen and two anonymous referees for valuable comments and the Hong Kong Research Grant Council, grant HKU 7031/04P, for partial support. Address correspondence to Wai Keung Li, Department of Statistics and Actuarial Sciences, University of Hong Kong, Pokfulam Road, Hong Kong; e-mail: hrntlwk@hku.hk.

 © 2009 Cambridge University Press 0266-4666/09 $15.00

https://doi.org/10.1017/S0266466608090488 Published online by Cambridge University Press

variance in financial and economic time series. If the parameters $\alpha_1, \ldots, \alpha_p$ and $\beta_1, \ldots, \beta_q$ are assumed to be zero, then $\{e_t\}$ will be an i.i.d. white noise sequence. For this case, the problem of unit root testing has been extensively and well discussed. For both empirical and theoretical considerations, more and more statisticians and econometricians recently focused on various unit root processes with non-i.i.d. errors; see Phillips and Durlauf (1986), Phillips (1987), Chan and Wei (1988), Herce (1996), etc. Obviously, in this literature, how to estimate and test hypotheses on the unit root parameter ϕ in models satisfying (1) and (2) are important problems.

For model (1) with conditional heteroskedasticity, the least squares (LS) estimation and the augmented Dickey–Fuller tests may encounter some intrinsic problems; see Pantula (1988). The simulation experiments in Kim and Schmidt (1993) also showed that these tests are often overrejecting, especially for the near integrated cases in which, e.g., $\alpha_1 + \beta_1 \approx 1$ for GARCH(1,1). This phenomenon can be explained partly by loss of efficiency because the score function of LS estimation did not involve the structure of conditional variance. Ling and Li (1998) and Seo (1999) investigated the local Gaussian quasi maximum likelihood estimation (QMLE) for models satisfying (1) and (2) under conditions $\mathrm{E}e_t^4 < \infty$ and $\mathrm{E}e_t^8 < \infty$, respectively. The tests based on the local Gaussian QMLE are shown not only to be more powerful but also to have more stable sizes than (augmented) Dickey–Fuller tests based on LS estimation; see Seo (1999) and Ling, Li, and McAleer (2003). However, the parameter space of GARCH models is much more limited for higher order moments, and the condition $\mathrm{E}e_t^4 < \infty$ may be too stringent for some real data; see Ling (2007). For AR(1) unit root processes with GARCH(1,1) errors, Ling and Li (2003) sharpened the results in Ling and Li (1998) and proposed a one-step local Gaussian QMLE for the unit root parameter ϕ with conditions $\mathrm{E}e_t^2 < \infty$ and $\mathrm{E}\varepsilon_t^4 < \infty$ and with ε_t symmetrically distributed. The unit root tests based on the one-step local Gaussian QMLE were shown to outperform Dickey–Fuller tests based on LS estimation; see Ling et al. (2003). However, how to perform estimation and test for unit roots in the general setting of (1) and (2) with only finite second-order moment for e_t are important open problems.

The innovations ε_t in (2) are generally assumed to be normally distributed, and it is well known that a stationary GARCH process can exhibit more heavy-tailed marginal distribution than the normal distribution. However, Mikosch and Stărică (2000) showed that the tails of fitted GARCH models with normal innovations seem to be much thinner than the tails appearing in the real data, and their empirical experiments demonstrated that the innovations may be so heavy-tailed that the fourth moment is infinite; see also Mittnik, Rachev, and Paolella (1998), Politis (2004), and Zhang, Li, and Yuen (2006). For pure GARCH processes with $\mathrm{E}\varepsilon_t^4 = \infty$, Hall and Yao (2003) showed that the Gaussian QMLE may not be asymptotically normal and the range of possible limit distributions is extraordinarily large; see also Mikosch and Straumann (2006). Hence, for model (1) with GARCH(1,1) errors under condition $\mathrm{E}\varepsilon_t^4 = \infty$, the asymptotic distribution for the

https://doi.org/10.1017/S0266466608090488 Published online by Cambridge University Press

estimated unit root in Ling and Li (2003) may no longer be obtained, and the unit root tests in Ling et al. (2003) may fail to correctly identify the unit root in the data. Herce (1996) considered a simple least absolute deviation (LAD) estimation for the unit root processes with finite variance errors, which may be correlated, and developed several unit root tests that can still be used for the case of GARCH errors with $\mathrm{E}\varepsilon_t^4 = \infty$. Because of the loss of efficiency, these tests in Herce (1996) may have lower powers, and this is supported by the simulation results in Section 4.

In this paper, we propose a LAD estimation methodology for unit root processes with GARCH errors, taking into account both efficiency and heavy tails. The asymptotic distribution of a local LAD estimator is derived under only finite second moments for both errors and innovations. When the innovations are symmetrically distributed, the asymptotic distribution of the estimated unit roots is shown to be a functional of a bivariate Brownian motion, and then two unit root tests are derived. These tests are found to have asymptotically either conditional or unconditional normal distributions, and the simulation results in Section 4 demonstrate that they outperform those based on the one-step local Gaussian QMLE in Ling and Li (2003) for the case with heavy-tailed innovations and outperform those based on the simple LAD estimators in Herce (1996) for all cases.

This paper is arranged as follows. Section 2 discusses the asymptotic properties of a local LAD estimator for models satisfying (1) and (2). Section 3 introduces two derived unit root tests, and the simulation results are presented in Section 4. The proof of Theorem 2.1 is given in the Appendix. Throughout this paper, $\|\cdot\|$ is the euclidean norm, $o_p(1)$ denotes a series of random numbers converging to zero in probability, $O_p(1)$ denotes a series of random numbers that are bounded in probability, $\rightarrow_d$ denotes convergence in distribution, and $D = D[0,1]$ denotes the space of functions on $[0,1]$ that is defined and equipped with the Skorokhod topology (Billingsley, 1968).

2. ASYMPTOTIC PROPERTIES OF LAD ESTIMATION

For models satisfying (1) and (2), denote their parameter space by Θ, a compact set in $R^{k+p+q+2}$. Let $\alpha = (\alpha_1, \ldots, \alpha_p)'$, $\beta = (\beta_1, \ldots, \beta_q)'$, $\psi = (\mu, \psi_1, \ldots, \psi_{k-1})'$, $\gamma = (\omega, \alpha', \beta')'$, $\delta = (\psi', \gamma')'$, and $\theta = (\phi, \delta')'$, where θ is the parameter vector. The true parameter vector $\theta_0 = (0, \delta_0')'$ is supposed to be an interior point of Θ. Note that the true value of the parameter μ is also equal to zero.

Assumption 2.1. The median of ε_t is equal to zero, $\mathrm{E}|\varepsilon_t| = 1$, $\mathrm{E}\varepsilon_t^2 = \sigma^2 < \infty$, and the density function $f(x)$ of ε_t is continuous at the origin.

Assumption 2.2. $\alpha_i > 0, i = 1, \ldots, p$, $\beta_j > 0, j = 1, \ldots, q$, $\mathrm{E}e_t^2 < \infty$, and the polynomials $\sum_{i=1}^{p} \alpha_i z^i$ and $1 - \sum_{j=1}^{q} \beta_j z^j$ have no common root.

Assumption 2.3. All roots of the polynomial $\psi(x) = 1 - \psi_1 x - \cdots - \psi_{k-1} x^{k-1}$ are outside the unit circle.

https://doi.org/10.1017/S0266466608090488 Published online by Cambridge University Press

Note that Assumption 2.1 unitizes the quantity $\mathrm{E}|\varepsilon_t|$ instead of $\mathrm{E}\varepsilon_t^2$, so that $\mathrm{E}\varepsilon_t^2 = \mathrm{var}(|\varepsilon_t|) + 1 > 1$. Hence, we need to rescale the parameters in the GARCH process (2), and the necessary and sufficient condition for $\mathrm{E}e_t^2 < \infty$ is $\sigma^2 \sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j < 1$; see Li and Li (2005). For simplicity, we assume that these parameters in (2) have already been rescaled to satisfy Assumption 2.1.

Given observations $y_1, \ldots, y_n$ from a model satisfying (1) and (2) with the initial value $y_0 = 0$, we consider the LS estimator $\widehat{\phi}_{\mathrm{LS}}$ and the simple LAD estimator $\widehat{\phi}_{\mathrm{SLAD}}$ as follows:

$$\left(\widehat{\phi}_{\mathrm{LS}}, \widehat{\psi}'_{\mathrm{LS}}\right)' = \operatorname{argmin} \sum_{t=1}^{n} [e_t(\theta)]^2 \quad \text{and} \quad \left(\widehat{\phi}_{\mathrm{SLAD}}, \widehat{\psi}'_{\mathrm{SLAD}}\right)' = \operatorname{argmin} \sum_{t=1}^{n} |e_t(\theta)|,$$

where $e_t(\theta) = e_t(\phi, \psi) = \Delta y_t - \phi y_{t-1} - \psi' z_{t-1}$ and $z_t = (1, \Delta y_t, \ldots, \Delta y_{t-k+2})$. The estimator $\widehat{\phi}_{\mathrm{LS}}$ can be used to construct augmented Dickey–Fuller tests and is well known to be sensitive to outliers. The estimator $\widehat{\phi}_{\mathrm{SLAD}}$ is a direct improvement of $\widehat{\phi}_{\mathrm{LS}}$, and the tests based on $\widehat{\phi}_{\mathrm{SLAD}}$ should be more powerful under heavy-tailed errors. Note that the structure of conditional variance is ignored by the score functions of these two estimators and hence they are expected to be inefficient.

To take into account the structure of conditional variance, we first define the function $h_t(\theta)$ satisfying the iterative equation $h_t(\theta) = \omega + \sum_{i=1}^{p} \alpha_i e_{t-i}^2(\theta) + \sum_{j=1}^{q} \beta_j h_{t-j}(\theta)$. In real applications, initial values are needed to calculate the functions $h_t(\theta)$ and are set to be zero for simplicity. By the proof in the Appendix, the initial values have no effect on the asymptotic results. Note that the two quantities, $h_t(\theta_0)$ and h_t, are quite different, where θ_0 is the true parameter vector and h_t is the exact conditional variance. On the other hand, h_t can be considered as a function of δ_0, and we can define its first- and second-order derivative functions, denoted, respectively, by $\partial h_t/\partial\delta$ and $\partial^2 h_t/\partial\delta\partial\delta'$. Furthermore, the behaviors of $\partial h_t(\theta)/\partial\phi$ are different from those of $\partial h_t(\theta)/\partial\delta$, and we cannot even define the quantity $\partial h_t/\partial\phi$.

We now introduce the LAD estimation for models satisfying (1) and (2), taking into account efficiency and heavy tails. As we know for linear models, the LAD estimation can be considered as the maximum likelihood estimation (MLE) under assumption of double exponential innovations; see Bassett and Koenker (1978), Davis and Dunsmuir (1997), etc. In this paper, we just use this approach to define the LAD estimation for a model satisfying (1) and (2). Suppose that ε_t, instead of e_t, in (2) follows a double exponential distribution with density $g(x) = 0.5e^{-|x|}$ and then, for time series $\{y_t\}$ generated by (1) and (2), the log transformation of the conditional likelihood function is equal to $-L_n(\theta) - n\log 2$, where

$$L_n(\theta) = \sum_{t=1}^{n} l_t(\theta) \quad \text{and} \quad l_t(\theta) = \frac{|e_t(\phi, \psi)|}{\sqrt{h_t(\theta)}} + \frac{1}{2}\log h_t(\theta).$$

Hence, the LAD estimator is defined as follows:

$$\widehat{\theta}_n = \operatorname*{argmin}_{\theta \in \Theta} L_n(\theta).$$

https://doi.org/10.1017/S0266466608090488 Published online by Cambridge University Press

To present the asymptotic results of the preceding LAD estimation, we first state some notation as follows:

$$F_1 = \mathrm{E}[\psi^{-1}(B)e_t]^2, \qquad F_2 = \mathbf{I}_1 + 0.5\kappa_1 \mathrm{E}\left[\frac{1}{\sqrt{h_t}}\frac{\partial h_t}{\partial \delta}\right], \qquad \Sigma_1 = \mathrm{E}\left[\frac{1}{h_t^2}\frac{\partial h_t}{\partial \delta}\frac{\partial h_t}{\partial \delta'}\right],$$

$$\Sigma_2 = \begin{pmatrix} \mathrm{E}[h_t^{-1} z_{t-1} z_{t-1}'] & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}, \qquad \Omega = \begin{pmatrix} F_1 & F_2' \\ F_2 & \Sigma_2 + 0.25\kappa_2 \Sigma_1 \end{pmatrix},$$

$$F = f(0)\mathrm{E}\left[\frac{1}{h_t}\right] + \frac{1}{8}\mathrm{E}\left[\frac{1}{h_t}\frac{\partial h_t}{\partial \mu}\right]^2,$$

$$F_3 = \left(f(0)\mathrm{E}\left[\frac{z_{t-1}'}{h_t}\right], 0, \ldots, 0\right)' + \frac{1}{8}\mathrm{E}\left[\frac{1}{h_t^2}\frac{\partial h_t}{\partial \mu}\frac{\partial h_t}{\partial \delta}\right],$$

$$\Sigma = f(0)\Sigma_2 + \frac{1}{8}\Sigma_1, \qquad \Gamma = \begin{pmatrix} F \int_0^1 B_1^2(\tau)\,d\tau & F_3' \int_0^1 B_1(\tau)\,d\tau \\ F_3 \int_0^1 B_1(\tau)\,d\tau & \Sigma \end{pmatrix},$$

where $z_t = (1, \Delta y_t, \ldots, \Delta y_{t-k+2})$, $\psi^{-1}(B)e_t = \sum_{j=0}^{\infty} \bar{\psi}_j e_{t-j}$ with $\psi(z) \cdot \sum_{j=0}^{\infty} \bar{\psi}_j z^j = 1$, $\mathbf{I}_1 = (1, 0, \ldots, 0)'$, $\kappa_1 = \mathrm{E}(|\varepsilon_t|\varepsilon_t)$, $\kappa_2 = \mathrm{E}(|\varepsilon_t| - 1)^2$, 0 is a scalar, and $\mathbf{0}$ is a zero vector or matrix with corresponding dimensions. Based on the preceding notation, we state the asymptotic distribution of the LAD estimator $\widehat{\theta}_n$ as follows.

THEOREM 2.1. *If Assumptions 2.1–2.3 are satisfied, then there exists a local minimizer* $\widehat{\theta}_n = (\widehat{\phi}_n, \widehat{\psi}_n', \widehat{\gamma}_n')'$ *of* $L_n(\theta)$ *such that*

$$\begin{bmatrix} n\widehat{\phi}_n \\ \sqrt{n}(\widehat{\psi}_n - \psi_0) \\ \sqrt{n}(\widehat{\gamma}_n - \gamma_0) \end{bmatrix} \longrightarrow_d 0.5\Gamma^{-1} \begin{bmatrix} \int_0^1 B_1(\tau)\,dB_2^{(1)}(\tau) \\ B_2(1) \\ B_3(1) \end{bmatrix},$$

where $\boldsymbol{B}(\tau) = [B_1(\tau), B_2'(\tau), B_3'(\tau)]'$ *is a* $(k + p + q + 2)$*-dimensional Brownian motion with covariance* $\tau\Omega$ *and* $B_2^{(1)}(\tau)$ *is the first component of* $B_2(\tau)$.

The proof of the preceding theorem is delayed to the Appendix. Note that the matrix Γ generally is not diagonal and it makes the asymptotic distribution of the estimated unit roots $\widehat{\phi}_n$ complicated. By some algebra, we can show that

$$n\widehat{\phi}_n \longrightarrow_d \frac{\int_0^1 B_1(\tau)\,dB_2^{(1)}(\tau) - F_3'\Sigma^{-1}[B_2'(1), B_3'(1)]' \int_0^1 B_1(\tau)\,d\tau}{2[F\int_0^1 B_1^2(\tau)\,d\tau - F_3'\Sigma^{-1}F_3(\int_0^1 B_1(\tau)\,d\tau)^2]}. \tag{3}$$

Denote by Ω_1 a lower triangular matrix such that $\Omega_1\Omega_1' = \Omega$ and it has the form

$$\Omega_1 = \left(\begin{array}{cc|c} \sqrt{F_1} & 0 & \mathbf{0} \\ \hline X_1 & X_2 & * \end{array}\right),$$

https://doi.org/10.1017/S0266466608090488 Published online by Cambridge University Press

where $X_i = (x_{i1}, \ldots, x_{i(k+p+q+1)})'$ with $i = 1, 2$. Note that $\omega(\tau) = \Omega_1^{-1}\mathbf{B}(\tau)$ is a $(k + p + q + 2)$-dimensional standard Brownian motion with independent components. Then, following Ling and Li (2003), we can rewrite the asymptotic distribution in (3) as

$$\frac{\begin{array}{c} x_{11}\int_0^1 \omega_1 d\omega_1 - F_3'\Sigma^{-1}X_1\omega_1(1)\int_0^1 \omega_1 d\tau + x_{21}\int_0^1 \omega_1 d\omega_2 \\ -(F_3'\Sigma^{-1}X_2\omega_2(1) + \eta^*)\int_0^1 \omega_1 d\tau \end{array}}{2\sqrt{F_1}[F\int_0^1 \omega_1^2 d\tau - F_3'\Sigma^{-1}F_3(\int_0^1 \omega_1 d\tau)^2]}, \tag{4}$$

where $\omega_1(\tau)$ and $\omega_2(\tau)$ are the first two components of $\omega(\tau)$ and η^* is a normal distribution with variance $F_3'\Sigma^{-1}(\Sigma_2 + 0.25\kappa_2\Sigma_1 - X_1X_1' - X_2X_2')\Sigma^{-1}F_3$ and independent of $w_1(\tau)$ and $w_2(\tau)$.

It is a challenging problem to directly determine the LAD estimators $\widehat{\theta}_n$ for models satisfying (1) and (2) because the log-likelihood function $L_n(\theta)$ has no derivatives with respect to ϕ and ψ. We here introduce the local quadratic approximation (Fan and Li, 2001) to overcome this problem; i.e., we can iteratively minimize

$$\sum_{t=1}^{n}\left\{\frac{e_t^2(\phi,\psi)}{|e_t(\phi^{(m)},\psi^{(m)})|\sqrt{h_t(\phi,\psi,\gamma)}} + \frac{1}{2}\log h_t(\phi,\psi,\gamma)\right\},$$

where $\phi^{(m)}$ and $\psi^{(m)}$ are the results of the mth iteration. For the initial values, we can take $\phi^{(1)}$ and $\psi^{(1)}$ to be the simple LAD estimators. Then some classical algorithms such as the Newton–Raphson method can be used to obtain estimates for the current iteration because the new objective function is smooth.

For the LS estimator $\widehat{\phi}_{\mathrm{LS}}$, by an argument similar to that of $\widehat{\phi}_n$, we can show that

$$n\widehat{\phi}_{\mathrm{LS}} \to_d \bar{\gamma}_1\frac{\int_0^1 \omega_1 d\omega_1 - \omega_1(1)\int_0^1 \omega_1 d\tau}{\int_0^1 \omega_1^2 d\tau - \left(\int_0^1 \omega_1 d\tau\right)^2} + \bar{\gamma}_2\left[\int_0^1 \omega_1^2 d\tau - \left(\int_0^1 \omega_1 d\tau\right)^2\right]^{-1/2}\eta_{\mathrm{LS}}, \tag{5}$$

where $\bar{\gamma}_1 = \mathrm{E}e_t^2/F_1$, $\bar{\gamma}_2 = \sqrt{\mathrm{E}e_t^2/F_1 - (\mathrm{E}e_t^2/F_1)^2}$, $\omega_1(\tau)$ is defined as in (4), and η_{LS} is a standard normal random variable independent of $\int_0^1\omega_1^2(\tau)\,d\tau - \left(\int_0^1 \omega_1 d\tau\right)^2$; see Phillips (1989). The asymptotic distribution of the simple LAD estimator $\widehat{\phi}_{\mathrm{SLAD}}$ has an even more complicated form than that of $\widehat{\phi}_n$, and we omit it here.

3. UNIT ROOT TESTS BASED ON LAD ESTIMATES

In the previous section, the asymptotic distribution of the local LAD estimator $\widehat{\phi}_n$ is derived under some very general conditions; however, it is rather complicated to be useful in practice. In this section, some simplifying conditions are imposed to make it more applicable for unit root tests.

https://doi.org/10.1017/S0266466608090488 Published online by Cambridge University Press

When ε_t is symmetrically distributed, it holds that

$$\mathrm{E}\left[h_t^{-1}\Delta y_{t-i}\right]=0 \quad \text{and} \quad \mathrm{E}\left[h_t^{-2}(\partial h_t/\partial\mu)(\partial h_t/\partial\delta_j)\right]=0, \tag{6}$$

where $i=1,\ldots,k-1$, $j=2,\ldots,k+p+q+1$, and δ_j is the jth element of the vector δ. The matrix Γ is then diagonal, and we can simplify the asymptotic distributions in (4) now. Furthermore, we can also obtain that $\kappa_1=0$, $F_2=\mathbf{I}_1$, and $F_3=F\mathbf{I}_1$. Hence, the asymptotic distribution in (4) can be simplified as

$$n\widehat{\phi}_n \to_d \gamma_1\frac{\int_0^1\omega_1 d\omega_1-\omega_1(1)\int_0^1\omega_1 d\tau}{\int_0^1\omega_1^2 d\tau-\left(\int_0^1\omega_1 d\tau\right)^2}+\gamma_2\left[\int_0^1\omega_1^2 d\tau-\left(\int_0^1\omega_1 d\tau\right)^2\right]^{-1/2}\eta_{\mathrm{LAD}}, \tag{7}$$

where $F_4=\mathrm{E}\left[h_t^{-1}\right]+0.25\kappa_2\mathrm{E}\left[h_t^{-1}(\partial h_t/\partial\mu)\right]^2$, $\gamma_1=1/(2FF_1)$, $\gamma_2=\sqrt{F_1F_4-1}/(2FF_1)$, $\omega_1(\tau)$ is defined as in (4), and η_{LAD} is a standard normal random variable independent of $\int_0^1\omega_1^2(\tau)\,d\tau-\left(\int_0^1\omega_1 d\tau\right)^2$.

Note that there are two nuisance parameters, γ_1 and γ_2, in (7). As in Phillips (1987) and Phillips and Perron (1988), we may modify the quantity $n\widehat{\phi}_n$ such that these two nuisance parameters can be removed from the asymptotic distributions of the resulting test statistics; see also Lucas (1995), Herce (1996), and Ling and Li (2003). The random variables, η_{LS} and η_{LAD}, in (5) and (7) generally depend on each other in a complicated way, and we need some extra quantities to simplify the asymptotic distributions. We here consider the LS estimator for an auxiliary regression,

$$(\widehat{\phi}_{\mathrm{SR}},\widehat{\mu}_{\mathrm{SR}})=\operatorname{argmin}\sum_{t=1}^{n}[\Delta y_t-\phi y_{t-1}-\mu]^2.$$

We can readily show the following lemma.

LEMMA 3.1. *If Assumptions 2.1–2.3 are satisfied, then*

$$n\widehat{\phi}_{\mathrm{SR}} \to_d \frac{\int_0^1\omega_1 d\omega_1-\omega_1(1)\int_0^1\omega_1 d\tau}{\int_0^1\omega_1^2 d\tau-\left(\int_0^1\omega_1 d\tau\right)^2},$$

where $\omega_1(\tau)$ is defined as in (4).

The asymptotic distribution in Lemma 3.1 is just the first term of that in (7), up to a constant γ_1. Using this result and (7), we can construct a unit root test,

$$L_\phi=\frac{n\widehat{\phi}_n-\widehat{\gamma}_1 n\widehat{\phi}_{\mathrm{SR}}}{\widehat{\gamma}_2},$$

where $\widehat{\gamma}_1$ (or $\widehat{\gamma}_2$) is a consistent estimator of γ_1 (or γ_2). Furthermore, we can also consider the corresponding t-ratio test as follows:

$$L_t=\left[\frac{1}{\widehat{F}_1 n^2}\sum_{t=2}^{n}(y_{t-1}-\bar{y})^2\right]^{1/2}L_\phi,$$

https://doi.org/10.1017/S0266466608090488 Published online by Cambridge University Press

where $\bar{y} = (1/n)\sum_{t=1}^{n} y_t$. For the quantities $\widehat{F}_1$, $\widehat{\gamma}_1$, and $\widehat{\gamma}_2$ in L_ϕ and L_t, we suggest the following method. Let $\widehat{e}_t = e_t(\widehat{\theta}_n)$, $\widehat{h}_t = h_t(\widehat{\theta}_n)$, and $\widehat{\varepsilon}_t = \widehat{e}_t/\sqrt{\widehat{h}_t}$. It can be shown that the quantities $(1/n)\sum_{t=2}^{n}(\Delta y_t)^2$, $(1/n)\sum_{t=1}^{n}\widehat{h}_t^{-1}$, $(1/n)\sum_{t=1}^{n}\left[\widehat{h}_t^{-1}(\partial h_t(\widehat{\theta}_n)\partial\mu)\right]^2$, and $(1/n)\sum_{t=1}^{n}(|\widehat{\varepsilon}_t| - 1)^2$ are consistent estimators of the items F_1, $\mathrm{E}h_t^{-1}$, $\mathrm{E}\left[h_t^{-1}(\partial h_t \partial\mu)\right]^2$, and κ_2, respectively. For $f(0)$, we can use a kernel method to estimate the density function of the sequence $\{\widehat{\varepsilon}_t\}$, which is assumed to be i.i.d., and it can be shown that $\widehat{f}(0)$ is a consistent estimator of $f(0)$, where $\widehat{f}(x)$ is the fitted density; see Silverman (1986) and Roussas (1988). By using the preceding estimated values to replace the corresponding terms in γ_1 (or γ_2), we can obtain a consistent estimator $\widehat{\gamma}_1$ (or $\widehat{\gamma}_2$).

Based on (7), Lemma 3.1, and Theorem 2.1, we can show that the preceding test statistics are asymptotically distributed with conditional or unconditional normality under the null hypothesis of unit root.

THEOREM 3.1. *Under Assumptions 2.1–2.3, if ε_t is symmetrically distributed, then*

(i) $L_\phi \to_d \left[\int_0^1 \omega_1^2 d\tau - \left(\int_0^1 \omega_1 d\tau\right)^2\right]^{-1/2}\eta$,

(ii) $L_t \to_d \eta$,

where $\omega_1(\tau)$ is a standard Brownian motion and η is a standard normal random variable independent of $\int_0^1 \omega_1^2 d\tau - \left(\int_0^1 \omega_1 d\tau\right)^2$.

The distribution of ε_t is assumed to be symmetrical in the preceding theorem, and, as pointed out by the referees, this restriction may be violated in real applications. However, to construct practical unit root tests, this condition, or condition (6) at least, is necessary to simplify the asymptotic distribution in (4). By the simulation results in Section 4, we note that the distortion of the sizes of L_ϕ and L_t is not serious when the condition of symmetry is violated. Note also that in practice ε_t is often assumed to have either a normal or Student's t-distribution.

The asymptotic distributions in Theorem 3.1 have the classical forms, and some of the critical values for the test L_ϕ can be referred to Table 1 of Herce (1996) or Table 5 of Ling et al. (2003). Furthermore, the test L_ϕ has normalization n; however, the test L_t is normalized by $\left[\sum_{t=2}^{n}(y_{t-1} - \bar{y})^2\right]^{1/2}$, which is $O_p(n^{1/2})$ under the stationary alternatives. This suggests that the test L_ϕ is more powerful than its t-ratio statistic when the sample size is large.

For the simple LAD estimator $\widehat{\phi}_{\mathrm{SLAD}}$, if ε_t is symmetrically distributed, we can show that

$$n\widehat{\phi}_{\mathrm{SLAD}} \to_d \tilde{\gamma}_1 \frac{\int_0^1 \omega_1 d\omega_1 - \omega_1(1)\int_0^1 \omega_1 d\tau}{\int_0^1 \omega_1^2 d\tau - \left(\int_0^1 \omega_1 d\tau\right)^2} + \tilde{\gamma}_2 \left[\int_0^1 \omega_1^2 d\tau - \left(\int_0^1 \omega_1 d\tau\right)^2\right]^{-1/2} \eta_{\mathrm{SLAD}}, \quad \textbf{(8)}$$

https://doi.org/10.1017/S0266466608090488 Published online by Cambridge University Press

where $\tilde{\gamma}_1 = \mathrm{E}|e_t|/\left[2F_1 f(0)\mathrm{E}\left(h_t^{-1/2}\right)\right]$, $\tilde{\gamma}_2 = \sqrt{F_1 - (\mathrm{E}|e_t|)^2}/\left[2F_1 f(0)\mathrm{E}\left(h_t^{-1/2}\right)\right]$, $\omega_1(\tau)$ is defined as in (4), and η_{SLAD} is a standard normal random variable independent of $\int_0^1 \omega_1^2 d\tau - \left(\int_0^1 \omega_1 d\tau\right)^2$. Note that, for AR(1) unit root processes, the condition of symmetry can be removed in deriving the asymptotic distribution in (8). Similarly, by Lemma 3.1 and (8), the unit root tests based on the simple LAD estimator can be constructed as follows:

$$\widetilde{L}_\phi = \frac{n\widehat{\phi}_{\mathrm{SLAD}} - \widehat{\tilde{\gamma}}_1 n\widehat{\phi}_{\mathrm{SR}}}{\widehat{\tilde{\gamma}}_2} \quad \text{and} \quad \widetilde{L}_t = \left[\frac{1}{\widehat{F}_1 n^2}\sum_{t=2}^{n}(y_{t-1}-\bar{y})^2\right]^{1/2}\widetilde{L}_\phi,$$

where $\widehat{\tilde{\gamma}}_1$ and $\widehat{\tilde{\gamma}}_2$ are, respectively, consistent estimators of $\tilde{\gamma}_1$ and $\tilde{\gamma}_2$. We call $\widetilde{L}_\phi$ and $\widetilde{L}_t$ Herce tests because they are the direct extension of those in Herce (1996). It can be shown that the asymptotic distributions of the preceding two tests are the same as those of the statistics L_ϕ and L_t, respectively.

Based on the LS estimator $\widehat{\phi}_{\mathrm{LS}}$, by Lemma 3.1 and (5), we can construct the augmented Dickey–Fuller tests for AR(k) unit root process with $k > 1$,

$$\widetilde{Z}_\phi = \frac{n\widehat{\phi}_{\mathrm{LS}} - \widehat{\bar{\gamma}}_1 n\widehat{\phi}_{\mathrm{SR}}}{\widehat{\bar{\gamma}}_2} \quad \text{and} \quad \widetilde{Z}_t = \left[\frac{1}{\widehat{F}_1 n^2}\sum_{t=2}^{n}(y_{t-1}-\bar{y})^2\right]^{1/2}\widetilde{Z}_\phi,$$

where $\widehat{\bar{\gamma}}_1$ and $\widehat{\bar{\gamma}}_2$ are, respectively, consistent estimators of $\bar{\gamma}_1$ and $\bar{\gamma}_2$. It can be shown that the preceding augmented Dickey–Fuller tests have the same asymptotic distributions as those of the statistics L_ϕ and L_t, respectively. However, when the true values of the parameters $\psi_1, \ldots, \psi_{k-1}$ are all equal to zero, the second item of the asymptotic distribution in (5) vanishes, i.e., $\bar{\gamma}_1 = 1$ and $\bar{\gamma}_2 = 0$, and the tests $\widetilde{Z}_\phi$ and $\widetilde{Z}_t$ cannot be used again. Let $\widetilde{Z}_\phi = n\widehat{\phi}_{\mathrm{LS}}$ and $\widetilde{Z}_t$ be its corrected t-ratio test. Then they are just the Dickey–Fuller tests, and some of their critical values can be found in Tables 10.A.1 and 10.A.2 of Fuller (1996). Note that the symmetry of ε_t is not used in deriving the augmented Dickey–Fuller tests.

4. SIMULATION RESULTS

In this section, we present the results of some simulation experiments that we performed to compare the test statistics designed in Section 3 with (augmented) Dickey–Fuller tests, Herce tests, and the unit root tests in Ling et al. (2003) (henceforth LLM tests).

Recall that, for the AR(1) unit root processes with GARCH(1, 1) errors, LLM tests have the forms of

$$Z_\phi = \frac{\mathrm{E}(e_t^2)F^* n\widehat{\phi}_{\mathrm{ML}} - n\widehat{\phi}_{\mathrm{SR}}}{\sqrt{\mathrm{E}(e_t^2)K^* - 1}} \quad \text{and} \quad Z_t = \left[\frac{1}{\widehat{F}_1 n^2}\sum_{t=2}^{n}(y_{t-1}-\bar{y})^2\right]^{1/2} Z_\phi,$$

where $\widehat{\phi}_{\mathrm{ML}}$ is the one-step local Gaussian QMLE and F^* (or K^*) is just the quantity F (or K) in Ling et al. (2003). It has been shown that these tests have the

https://doi.org/10.1017/S0266466608090488 Published online by Cambridge University Press

same asymptotic distributions as the Herce tests and our tests. In the following simulation experiments, we used the same critical values for these three types of tests. Specifically, the significance level was set to 0.05, and the acceptance region was $(-6.79, 6.71)$ for L_ϕ, $\widetilde{L}_\phi$, and Z_ϕ and $(-1.96, 1.96)$ for the other three; see Table 1 of Herce (1996).

For the sake of comparison, we employed the same generating processes as in Ling et al. (2003),

$$\Delta y_t = -\phi y_{t-1} + e_t, \qquad e_t = \varepsilon_t \sqrt{h_t}, \qquad h_t = \omega + \alpha_1 e_{t-1}^2 + \beta_1 h_{t-1},$$

where $\omega = 0.1$ and $(\alpha_1, \beta_1) = (0.2, 0.7), (0.3, 0.6)$, and $(0.4, 0.5)$. Four distributions, the standard normal, the double exponential, and t-distributions with 5 and 3 degrees of freedom ($t(5)$ and $t(3)$), were considered to check the performance of these tests with heavy-tailed innovations, and the pseudo random numbers were first standardized to be mean zero and variance one. The AR parameter ϕ was set to zero to check for the sizes and 0.05, 0.025, or 0.01 to check for the powers. We considered the sample size $n = 300$, and there were 1,000 replications for each combination of parameter vector $(\phi, \omega, \alpha_1, \beta_1)$ and distribution of innovations.

We applied all tests to each replication. Note that for replications driven by $t(3)$ innovations, the fourth moment of ε_t is infinite and LLM tests are not available. Hence, only the other three types of tests were considered for this case. The simple LAD estimator $\widehat{\phi}_{\mathrm{SLAD}}$ was calculated by the DRLAV subroutine, the double precision version of the least absolute value regression subroutine in the International Mathematical and Statistical Library (IMSL), as in Herce (1996). For the LAD estimator $\widehat{\phi}_n$ and the one-step local Gaussian QMLE $\widehat{\phi}_{\mathrm{ML}}$, we used, respectively, the algorithms in Section 2 and Ling and Li (2003), and the Newton–Raphson method was employed to optimize the corresponding objective functions. The nuisance parameters were estimated by the methods in Section 3 with a chosen bandwidth. It would be ideal, in practice, to use a data-driven bandwidth as in Jones, Marron, and Sheather (1996); however, such an approach may involve a huge amount of computation in the simulation experiment. For simplicity, the bandwidth was set to be 0.3. Our experiments and simulation results, which are available on request, demonstrate that the estimated values $\widehat{f}(0)$ are all close to the true values. The significance level was set to 0.05, and the critical values for Dickey–Fuller tests can be obtained from Tables 10.A.1 and 10.A.2 of Fuller (1996). The empirical powers and sizes are listed in Tables 1 and 2.

The sizes of the Dickey–Fuller tests in Tables 1 and 2 are all greater than the nominal value 0.05, which is consistent with the finding in Kim and Schmidt (1993). Note that the GARCH errors exhibit more heavy-tailed feature when the value of α_1 is larger and the innovations are more heavy-tailed. As in Ling et al. (2003), the Dickey–Fuller tests are also more sensitive when the errors are more heavy-tailed. Furthermore, in terms of power, the Dickey–Fuller tests also perform the worst for time series with a near to unit root situation, i.e., $\phi = 0.01$ or 0.025.

https://doi.org/10.1017/S0266466608090488 Published online by Cambridge University Press

TABLE 1. The empirical powers and sizes for the innovations with normal and double exponential distributions

			ϕ							
			Normal				Double exponential			
α_1	β_1	Test	0.05	0.025	0.01	0.0	0.05	0.025	0.01	0.0
0.2	0.7	$\widetilde{Z}_\phi$	0.592	0.221	0.073	0.065	0.599	0.221	0.096	0.081
		$\widetilde{Z}_t$	0.426	0.145	0.053	0.066	0.462	0.149	0.065	0.088
		Z_ϕ	0.458	0.402	0.194	0.082	0.645	0.387	0.200	0.107
		Z_t	0.229	0.228	0.137	0.088	0.393	0.254	0.174	0.111
		$\widetilde{L}_\phi$	0.363	0.242	0.144	0.066	0.712	0.378	0.134	0.057
		$\widetilde{L}_t$	0.130	0.105	0.088	0.069	0.398	0.226	0.100	0.041
		L_ϕ	0.458	0.262	0.148	0.067	0.875	0.598	0.265	0.066
		L_t	0.229	0.140	0.103	0.066	0.673	0.427	0.201	0.049
0.3	0.6	$\widetilde{Z}_\phi$	0.570	0.241	0.086	0.073	0.615	0.210	0.116	0.068
		$\widetilde{Z}_t$	0.449	0.164	0.056	0.072	0.488	0.142	0.091	0.087
		Z_ϕ	0.781	0.493	0.200	0.080	0.731	0.490	0.256	0.100
		Z_t	0.530	0.344	0.166	0.072	0.513	0.336	0.207	0.108
		$\widetilde{L}_\phi$	0.426	0.269	0.152	0.063	0.730	0.460	0.136	0.061
		$\widetilde{L}_t$	0.181	0.121	0.083	0.072	0.450	0.294	0.096	0.048
		L_ϕ	0.584	0.364	0.167	0.069	0.921	0.691	0.281	0.063
		L_t	0.344	0.204	0.130	0.069	0.774	0.529	0.238	0.054
0.4	0.5	$\widetilde{Z}_\phi$	0.617	0.245	0.095	0.086	0.612	0.263	0.116	0.082
		$\widetilde{Z}_t$	0.492	0.172	0.063	0.085	0.490	0.199	0.086	0.094
		Z_ϕ	0.861	0.605	0.247	0.082	0.768	0.570	0.267	0.120
		Z_t	0.648	0.425	0.210	0.085	0.547	0.393	0.216	0.100
		$\widetilde{L}_\phi$	0.462	0.283	0.154	0.080	0.752	0.452	0.123	0.059
		$\widetilde{L}_t$	0.232	0.152	0.115	0.070	0.496	0.275	0.102	0.043
		L_ϕ	0.656	0.432	0.215	0.073	0.951	0.743	0.310	0.070
		L_t	0.411	0.281	0.152	0.072	0.813	0.576	0.268	0.058

Note: Dickey–Fuller tests: $\widetilde{Z}_\phi$ and $\widetilde{Z}_t$; LLM tests: Z_ϕ and Z_t; Herce tests: $\widetilde{L}_\phi$ and $\widetilde{L}_t$; our tests: L_ϕ and L_t.

In contrast, the tests designed in Section 3 have acceptable sizes, which are all close to their nominal value of 0.05. The sizes of L_t are comparable to those of $\widetilde{L}_t$; however, the sizes of L_ϕ are slightly more sensitive than those of $\widetilde{L}_\phi$. Second, all powers of our tests in Tables 1 and 2 are significantly greater than those of the Herce tests. This may be due to loss of efficiency of the simple LAD estimators. Compared with the LLM tests, our tests also have larger powers for cases with heavy-tailed innovations, i.e., double exponential or $t(5)$. It is within expectation that LLM tests perform best for the case with normal innovations because the Gaussian QMLE is just the MLE and would be most efficient.

https://doi.org/10.1017/S0266466608090488 Published online by Cambridge University Press

TABLE 2. The empirical powers and sizes for the innovations with $t(5)$ and $t(3)$ distributions

			ϕ							
			$t(5)$				$t(3)$			
α_1	β_1	Test	0.05	0.025	0.01	0.0	0.05	0.025	0.01	0.0
0.2	0.7	$\widetilde{Z}_\phi$	0.590	0.231	0.096	0.070	0.618	0.225	0.117	0.068
		$\widetilde{Z}_t$	0.457	0.163	0.068	0.074	0.487	0.158	0.082	0.083
		Z_ϕ	0.642	0.398	0.210	0.110				
		Z_t	0.376	0.247	0.158	0.133				
		$\widetilde{L}_\phi$	0.501	0.308	0.146	0.053	0.692	0.399	0.164	0.056
		$\widetilde{L}_t$	0.244	0.170	0.092	0.053	0.428	0.246	0.118	0.058
		L_ϕ	0.667	0.401	0.175	0.061	0.825	0.557	0.255	0.069
		L_t	0.405	0.254	0.125	0.063	0.606	0.397	0.195	0.066
0.3	0.6	$\widetilde{Z}_\phi$	0.602	0.241	0.103	0.071	0.583	0.258	0.112	0.094
		$\widetilde{Z}_t$	0.460	0.179	0.069	0.077	0.459	0.177	0.082	0.094
		Z_ϕ	0.722	0.471	0.211	0.099				
		Z_t	0.488	0.318	0.175	0.097				
		$\widetilde{L}_\phi$	0.590	0.333	0.165	0.070	0.695	0.442	0.152	0.051
		$\widetilde{L}_t$	0.331	0.202	0.106	0.065	0.456	0.291	0.117	0.049
		L_ϕ	0.744	0.477	0.197	0.064	0.862	0.630	0.267	0.073
		L_t	0.528	0.343	0.160	0.062	0.668	0.469	0.216	0.066
0.4	0.5	$\widetilde{Z}_\phi$	0.612	0.252	0.105	0.080	0.618	0.286	0.122	0.097
		$\widetilde{Z}_t$	0.483	0.186	0.075	0.085	0.491	0.215	0.083	0.101
		Z_ϕ	0.757	0.556	0.230	0.107				
		Z_t	0.518	0.412	0.188	0.104				
		$\widetilde{L}_\phi$	0.652	0.399	0.177	0.069	0.711	0.424	0.163	0.049
		$\widetilde{L}_t$	0.403	0.253	0.120	0.056	0.447	0.264	0.125	0.037
		L_ϕ	0.816	0.565	0.244	0.073	0.875	0.628	0.279	0.073
		L_t	0.626	0.423	0.192	0.064	0.688	0.448	0.244	0.059

Note: Dickey–Fuller tests: $\widetilde{Z}_\phi$ and $\widetilde{Z}_t$; LLM tests: Z_ϕ and Z_t; Herce tests: $\widetilde{L}_\phi$ and $\widetilde{L}_t$; our tests: L_ϕ and L_t.

As suggested by the co-editor, we also conducted an experiment to check the sizes of our tests in Section 3 when the innovations were asymmetrically distributed. The same generating process as that in the previous experiments was employed here, and the innovation sequence $\{\varepsilon_t\}$ came from a mixed t-distribution, taking values of $a_1|t_{f_1}|$ and $-a_2|t_{f_2}|$, respectively, with probability 0.5, where the two constants a_1 and a_2 made sure that $\mathrm{E}\varepsilon_t = 0$ and $\mathrm{E}\varepsilon_t^2 = 1$. We considered the following AR(2) unit root process to fit each sample:

$$\Delta y_t = -\phi y_{t-1} + \mu + \psi \Delta y_{t-1} + e_t, \qquad e_t = \varepsilon_t \sqrt{h_t},$$
$$h_t = \omega + \alpha_1 e_{t-1}^2 + \beta_1 h_{t-1},$$

https://doi.org/10.1017/S0266466608090488 Published online by Cambridge University Press

TABLE 3. The empirical sizes in percentages for the mixed t-distributed innovations with 5 and t_{f_2} degrees of freedom

α_1	β_1	t_{f_2}	$\widetilde{Z}_\phi$	$\widetilde{Z}_t$	$\widetilde{L}_\phi$	$\widetilde{L}_t$	L_ϕ	L_t
0.2	0.7	7	6.4	6.7	7.2	6.5	6.6	6.5
		11	7.4	7.8	6.5	6.1	6.3	5.7
0.3	0.6	7	7.8	8.5	7.4	6.1	6.8	6.0
		11	7.8	8.4	7.7	6.5	7.3	6.7
0.4	0.5	7	8.5	8.0	7.8	6.7	7.5	6.1
		11	7.4	9.0	8.1	7.2	7.9	7.0

Note: Augmented Dickey–Fuller tests: $\widetilde{Z}_\phi$ and $\widetilde{Z}_t$; Herce tests: $\widetilde{L}_\phi$ and $\widetilde{L}_t$; our tests: L_ϕ and L_t.

and two different distributions of ε_t were considered in this experiment with (f_1, f_2), respectively, equal to $(5, 7)$ and $(5, 11)$. Note that LLM tests cannot be used here. The empirical sizes of the other three types of tests are listed in Table 3. Furthermore, from the derivation in Section 3, the sizes of the augmented Dickey–Fuller tests should be close to the nominal value, 5%, and those of the Herce tests and our tests are supposed to be distorted. Nevertheless, Table 3 shows that the sizes of all the tests are affected to some extent. Although the empirical sizes of L_ϕ and L_t are slightly distorted, the situation seems better than for the other tests.

Finally, we conclude that our tests can be usefully applied when time series have time-varying conditional variance with heavy-tailed innovations, even when the fourth moment of innovations is finite.

REFERENCES

Bassett, G.J.R. & R. Koenker (1978) Asymptotic theory of least absolute error regression. *Journal of the American Statistical Association* 73, 618–622.

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

Bollerslev, T. (1986) Generalized autoregression conditional heteroscedasticity. *Journal of Econometrics* 31, 307–327.

Chan, N.H. & C.Z. Wei (1988) Limiting distributions of least squares estimates of unstable autoregressive processes with general autoregressive heteroscedastic errors. *Annals of Statistics* 16, 367–401.

Davis, R.A. & W.T.M. Dunsmuir (1997) Least absolute deviation estimation for regression with ARMA errors. *Journal of Theoretical Probability* 10, 481–497.

Davis, R.A., K. Knight, & J. Liu (1992) M-estimation for autoregressions with infinite variances. *Stochastic Processes and Their Applications* 40, 145–180.

Fan, J. & R. Li (2001) Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association* 96, 1348–1360.

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

Hall, P. & C.C. Heyde (1980) *Martingale Limit Theory and Its Application*. Academic Press.

Hall, P. & Q. Yao (2003) Inference in ARCH and GARCH models with heavy-tailed errors. *Econometrica* 71, 285–317.

Herce, M.A. (1996) Asymptotic theory of LAD estimation in a unit root process with finite variance errors. *Econometric Theory* 12, 129–153.

https://doi.org/10.1017/S0266466608090488 Published online by Cambridge University Press

Jones, M.C., J.S. Marron, & S.J. Sheather (1996) A brief survey of bandwidth selection for density estimation. *Journal of the American Statistical Association* 91, 401–407.
Kim, K. & P. Schmidt (1993) Unit root tests with conditional heteroskedasticity. *Journal of Econometrics* 59, 287–300.
Knight, K. (1998) Limiting distributions for l_1 regression estimators under general conditions. *Annals of Statistics* 26, 755–770.
Kurtz, T.G. & P. Protter (1991) Weak limit theorems to stochastic integrals and stochastic differential equations. *Annals of Probability* 19, 1035–1070.
Li, G. & W.K. Li (2005) Diagnostic checking for time series models with conditional heteroskedasticity estimated by the least absolute deviation approach. *Biometrika* 92, 691–701.
Li, G. & W.K. Li (2008) Testing for threshold moving average with conditional heteroscedasticity. *Statistica Sinica* 18, 647–665.
Ling, S. (2007) A double Ar(p) model: Structure and estimation. *Statistica Sinica* 17, 161–175.
Ling, S. & W.K. Li (1998) Limiting distributions of maximum likelihood estimators for unstable autoregressive moving-average time series with general autoregressive heteroskedastic errors. *Annals of Statistics* 26, 84–125.
Ling, S. & W.K. Li (2003) Asymptotic inference for unit root processes with GARCH(1,1) errors. *Econometric Theory* 19, 541–564.
Ling, S., W.K. Li, & M. McAleer (2003) Estimation and testing for unit root processes with GARCH(1,1) errors: Theory and Monte Carlo evidence. *Econometric Reviews* 22, 179–202.
Lucas, A. (1995) Unit root tests based on M estimators. *Econometric Theory* 11, 331–346.
Mikosch, T. & C. Stărică (2000) Limit theory for the sample autocorrelations and extremes of a GARCH(1,1) process. *Annals of Statistics* 28, 1427–1451.
Mikosch, T. & D. Straumann (2006) Stable limits of martingale transforms with application to the estimation of GARCH parameters. *Annals of Statistics* 34, 493–522.
Mittnik, S., S.T. Rachev, & M.S. Paolella (1998) Stable Paretian models in finance: Some empirical and theoretical aspects. In R.J. Adler, R.E. Feldman, & M.S. Taqqu (eds.), *A Practical Guide to Heavy Tails*, pp. 79–110. Birkhauser.
Pantula, S.G. (1988) Estimation of autoregressive models with ARCH errors. *Sankhya B* 50, 119–138.
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. & S.N. Durlauf (1986) Multiple time series regression with integrated processes. *Review of Economic Studies* 53, 473–495.
Phillips, P.C.B. & P. Perron (1988) Testing for a unit root in time series regression. *Biometrika* 75, 335–346.
Politis, D.N. (2004) A heavy-tailed distribution for ARCH residuals with application to volatility prediction. *Annals of Economics and Finance* 5, 283–298.
Roussas, G.G. (1988) Nonparametric estimation in mixing sequences of random variable. *Journal of Statistical Planning and Inference* 18, 135–149.
Seo, B. (1999) Distribution theory for unit root tests with conditional heteroskedasticity. *Journal of Econometrics* 91, 113–144.
Silverman, B.W. (1986) *Density Estimation for Statistics and Data Analysis*. Chapman and Hall.
Zhang, Z., W.K. Li, & K.C. Yuen (2006) On a mixture GARCH time-series model. *Journal of Time Series Analysis* 27, 577–597.

APPENDIX: Proof of Theorem 2.1

In this Appendix, we first state two lemmas, Lemmas A.1 and A.2, and then present the proof of Theorem 2.1. To make the proof easy to understand, we also give the proofs of (A.7) and some important inequalities at the end of this Appendix.

https://doi.org/10.1017/S0266466608090488 Published online by Cambridge University Press

LEMMA A.1. *If Assumption 2.2 is satisfied, then it holds that, for* $0 < \rho < 1$,

$$\mathrm{E}|h_t - h_t(\theta_0)| = O(\rho^t), \qquad \mathrm{E}\left\| \frac{\partial h_t}{\partial \delta} - \frac{\partial h_t(\theta_0)}{\partial \delta} \right\| = O(\rho^t)$$

and

$$\mathrm{E}\left\| \frac{\partial^2 h_t}{\partial \delta \partial \delta'} - \frac{\partial^2 h_t(\theta_0)}{\partial \delta \partial \delta'} \right\| = O(\rho^t).$$

Proof. Let

$$u_q = (1, 0, \ldots, 0)'_{1\times q} \quad \text{and} \quad B = \begin{pmatrix} \beta_1 & \cdots & \beta_q \\ & I_{q-1} & \mathbf{0}_{(q-1)\times 1} \end{pmatrix}_{q\times q},$$

where I_k is a $k \times k$ identity matrix and $\mathbf{0}_{k\times l}$ is a $k \times l$ zero matrix. Then the function $h_t(\theta)$ can be rewritten as

$$h_t(\theta) = \sum_{j=0}^{t-2} \left[\omega + \sum_{i=1}^{\min\{p, t-j-1\}} \alpha_i e^2_{t-j-i}(\theta) \right] u'_q B^j u_q$$

in the almost sure sense. Note that, by Assumption 2.2 and Lemma A.1 of Li and Li (2008), it holds that $|u'_q B^j u_q| < C_1 \rho^j$ for all $j \geq 0$, where $0 < \rho < 1$ and C_1 is a constant. By some straightforward calculation, we finish the proof of this lemma. ■

LEMMA A.2. *Under Assumptions 2.1–2.3, the following conditions are jointly satisfied:*

(i) $\frac{1}{n^2} \sum_{t=1}^{n} y^2_{t-1} \longrightarrow_d \int_0^1 B_1^2(\tau)\, d\tau,$

(ii) $\frac{1}{n\sqrt{n}} \sum_{t=1}^{n} y_{t-1} \longrightarrow_d \int_0^1 B_1(\tau)\, d\tau,$

(iii) $\frac{1}{n} \sum_{t=1}^{n} y_{t-1} \left[\frac{\mathrm{sgn}(e_t)}{\sqrt{h_t}} + \frac{(|\varepsilon_t| - 1)}{2h_t} \frac{\partial h_t}{\partial \mu} \right] \longrightarrow_d \int_0^1 B_1(\tau)\, dB_2^{(1)}(\tau),$

(iv) $\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left[\mathrm{sgn}(e_t) \frac{z_{t-1}}{\sqrt{h_t}} + \frac{(|\varepsilon_t| - 1)}{2h_t} \frac{\partial h_t}{\partial \psi} \right] \longrightarrow_d B_2(1),$

(v) $\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left[\frac{(|\varepsilon_t| - 1)}{2h_t} \frac{\partial h_t}{\partial \gamma} \right] \longrightarrow_d B_3(1),$

where $\boldsymbol{B}(\tau) = [B_1(\tau), B'_2(\tau), B'_3(\tau)]'$ *is a* $(k + p + q + 2)$*-dimensional Brownian motion defined as in Theorem 2.1 and* $B_2^{(1)}(\tau)$ *is the first component of* $B_2(\tau)$.

Proof. Let $\lambda = (\lambda_1, \lambda'_2, \lambda'_3)'$ be a $(k + p + q + 2)$-dimensional constant vector with $\lambda'\lambda \neq 0$, where λ_2 and λ_3, are, respectively, k- and $(p + q + 1)$-dimensional vector. Let $\mathcal{F}_t = \sigma(\varepsilon_t, \varepsilon_{t-1}, \ldots)$ be a σ-field generated by $\{\varepsilon_t, \varepsilon_{t-1}, \ldots\}$ and denote

$$\varrho_t = \lambda_1 \psi^{-1}(B) e_t + \lambda'_2 \left[\mathrm{sgn}(e_t) \frac{z_{t-1}}{\sqrt{h_t}} + \frac{(|\varepsilon_t| - 1)}{2h_t} \frac{\partial h_t}{\partial \psi} \right] + \lambda'_3 \frac{(|\varepsilon_t| - 1)}{2h_t} \frac{\partial h_t}{\partial \gamma}.$$

Note that $\{\varrho_t, t \in Z\}$ is a martingale difference with respect to $\{\mathcal{F}_t, t \in Z\}$.

https://doi.org/10.1017/S0266466608090488 Published online by Cambridge University Press

Denote $S_m = n^{-1/2}\sum_{t=1}^{m} \varrho_t$. It is readily shown that $\mathrm{E}S_n^2 = \lambda'\Omega\lambda$ and $0 < \lambda'\Omega\lambda < \infty$, where Ω is defined as in Theorem 2.1. Note that the sequences $\{\varrho_t\}$ and $\{\mathrm{E}(\varrho_t{}^2|\mathcal{F}_{t-1})\}$ are both strictly stationary and ergodic. Hence, it is easy to show that

$$\frac{1}{n}\sum_{t=1}^{n}\frac{\mathrm{E}(\varrho_t{}^2|\mathcal{F}_{t-1})}{\mathrm{E}S_n{}^2} \longrightarrow 1 \tag{A.1}$$

in the almost sure sense and, for any $\epsilon > 0$,

$$\frac{1}{n}\sum_{t=1}^{n}\mathrm{E}\left[\varrho_t{}^2 I\left(\varrho_t \geq \sqrt{n\mathrm{var}(\varrho_t)}\epsilon\right)\right] = o(1). \tag{A.2}$$

By the invariance principle for martingales (Hall and Heyde, 1980), (A.1), and (A.2), we have that

$$S_{[n\tau]} = \frac{1}{\sqrt{n}}\sum_{t=1}^{[n\tau]}\varrho_t \longrightarrow_d W(\tau) \qquad \text{in } D,$$

where $W(\tau)$ is a Brownian motion with variance $\tau\lambda'\Omega\lambda$. Then, by Cramér's device,

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{[n\tau]}\left[\psi^{-1}(B)e_t, \mathrm{sgn}(e_t)\frac{z'_{t-1}}{\sqrt{h_t}} + \frac{(|\varepsilon_t|-1)}{2h_t}\frac{\partial h_t}{\partial \psi'}, \frac{(|\varepsilon_t|-1)}{2h_t}\frac{\partial h_t}{\partial \gamma'}\right]' \longrightarrow_d \mathbf{B}(\tau) \quad \text{in } D\times D.$$

Hence, following Theorem 2.2 in Kurtz and Protter (1991), (i)–(v) hold jointly. ■

Proof of Theorem 2.1. For any $\upsilon = (\upsilon_1, \upsilon_2', \upsilon_3')'$, where $\upsilon_1 \in R$, $\upsilon_2 \in R^k$, and $\upsilon_3 \in R^{p+q+1}$, let $\phi = \upsilon_1/n$, $\psi = \psi_0 + \upsilon_2/\sqrt{n}$, $\gamma = \gamma_0 + \upsilon_3/\sqrt{n}$, $\delta = (\psi', \gamma')'$, and $\theta = (\phi, \delta')'$. Then,

$$\begin{aligned} L_n(\theta) - L_n(\theta_0) &= \sum_{t=1}^{n}\left\{\frac{|e_t(\phi,\psi)|}{\sqrt{h_t(\theta)}} + \frac{1}{2}\log h_t(\theta) - \frac{|e_t|}{\sqrt{h_t(\theta_0)}} - \frac{1}{2}\log h_t(\theta_0)\right\} \\ &= \sum_{t=1}^{n}\left\{\frac{|e_t|}{\sqrt{h_t(\theta)}} + \frac{1}{2}\log h_t(\theta) - \frac{|e_t|}{\sqrt{h_t(\theta_0)}} - \frac{1}{2}\log h_t(\theta_0)\right\} \\ &\quad + \sum_{t=1}^{n}\frac{1}{\sqrt{h_t(\theta)}}[|e_t(\phi,\psi)| - |e_t|] \\ &:= s_n^{(1)}(\upsilon) + s_n^{(2)}(\upsilon), \end{aligned}$$

where $\theta_0 = (0, \delta_0')' = (0, \psi_0', \gamma_0')'$.

Note that $s_n^{(1)}(\upsilon)$ is a smooth function with respect to υ. Then, by Taylor expansion, Lemma A.1, and the inequalities at the end of this Appendix, it holds that

$$\begin{aligned} s_n^{(1)}(\upsilon) = &-\frac{\upsilon_1}{2n}\sum_{t=1}^{n}\frac{(|\varepsilon_t|-1)}{h_t}\frac{\partial h_t(\theta_0)}{\partial\phi} - \frac{(\upsilon_2', \upsilon_3')}{2\sqrt{n}}\sum_{t=1}^{n}\frac{(|\varepsilon_t|-1)}{h_t}\frac{\partial h_t}{\partial\delta} \\ &+ \frac{\upsilon_1^2}{8n^2}\sum_{t=1}^{n}\frac{1}{h_t^2}\left(\frac{\partial h_t(\theta_0)}{\partial\phi}\right)^2 + \frac{\upsilon_1(\upsilon_2', \upsilon_3')}{4n\sqrt{n}}\sum_{t=1}^{n}\frac{1}{h_t^2}\frac{\partial h_t(\theta_0)}{\partial\phi}\frac{\partial h_t}{\partial\delta} \\ &+ \frac{1}{8}(\upsilon_2', \upsilon_3')\mathrm{E}\left[\frac{1}{h_t^2}\frac{\partial h_t}{\partial\delta}\frac{\partial h_t}{\partial\delta'}\right](\upsilon_2', \upsilon_3')' + o_p(1). \end{aligned} \tag{A.3}$$

https://doi.org/10.1017/S0266466608090488 Published online by Cambridge University Press

We now consider the item $\partial h_t(\theta_0)/\partial\phi$ in the preceding approximation. Denote $r_{t,i} = \sum_{r=1}^{i} \psi^{-1}(B)e_{t-r}$. It holds that $y_{t-i-1} = y_{t-1} - r_{t,i}$ and $\mathrm{E}r_{t,i}^2 \le C_2 i$, where C_2 is a constant. Then,

$$\frac{\partial h_t(\theta_0)}{\partial\phi} = -2\sum_{j=0}^{t-2}\sum_{i=1}^{\min(p,t-j-1)} \alpha_{i0}e_{t-j-i}y_{t-j-i-1}u_q' B_0^j u_q = y_{t-1}\frac{\partial h_t(\theta_0)}{\partial\mu} + 2R_t^*,$$

where B_0 is just the matrix B in the proof of Lemma A.1 evaluated at the true parameter vector and

$$\mathrm{E}\left[\frac{R_t^*}{h_t}\right]^2 = \mathrm{E}\left[\frac{1}{h_t}\sum_{j=0}^{t-2}\sum_{i=1}^{\min(p,t-j-1)} \alpha_{i0}e_{t-j-i}r_{t,i+j}u_q' B^j u_q\right]^2$$

$$\le \mathrm{E}\left\{\sum_{j=0}^{t-2}\sum_{i=1}^{\min(p,t-j-1)} \sqrt{\alpha_{i0}u_q' B^j u_q}\, r_{t,i+j}\right\}^2 \le C_3$$

with a finite constant C_3. For the first item in (A.3), by Lemma A.1, we can show that

$$\frac{1}{n}\sum_{t=1}^{n}\frac{(|\varepsilon_t|-1)}{h_t}\frac{\partial h_t(\theta_0)}{\partial\phi} = \frac{1}{n}\sum_{t=1}^{n}\frac{y_{t-1}(|\varepsilon_t|-1)}{h_t}\frac{\partial h_t(\theta_0)}{\partial\mu} + \frac{2}{n}\sum_{t=1}^{n}\frac{(|\varepsilon_t|-1)}{h_t}R_t^*$$

$$= \frac{1}{n}\sum_{t=1}^{n}\frac{y_{t-1}(|\varepsilon_t|-1)}{h_t}\frac{\partial h_t}{\partial\mu} + o_p(1) \tag{A.4}$$

because

$$\mathrm{E}\left[\frac{1}{n}\sum_{t=1}^{n}\frac{(|\varepsilon_t|-1)}{h_t}R_t^*\right]^2 = \frac{\kappa}{n^2}\sum_{t=1}^{n}\mathrm{E}\left(\frac{R_t^*}{h_t}\right)^2 = o(1).$$

Similarly, by Theorem 3.1 in Ling and Li (1998), we can show that

$$\frac{1}{n^2}\sum_{t=1}^{n}\frac{1}{h_t^2}\left(\frac{\partial h_t(\theta_0)}{\partial\phi}\right)^2 = \frac{1}{n^2}\sum_{t=1}^{n}\frac{y_{t-1}^2}{h_t^2}\left(\frac{\partial h_t}{\partial\mu}\right)^2 + o_p(1)$$

$$= \mathrm{E}\left[\frac{1}{h_t}\frac{\partial h_t}{\partial\mu}\right]^2\frac{1}{n^2}\sum_{t=1}^{n}y_{t-1}^2 + o_p(1) \tag{A.5}$$

and

$$\frac{1}{n\sqrt{n}}\sum_{t=1}^{n}\frac{1}{h_t^2}\frac{\partial h_t(\theta_0)}{\partial\phi}\frac{\partial h_t}{\partial\delta} = \frac{1}{n\sqrt{n}}\sum_{t=1}^{n}\frac{y_{t-1}}{h_t^2}\frac{\partial h_t}{\partial\mu}\frac{\partial h_t}{\partial\delta} + o_p(1)$$

$$= \mathrm{E}\left[\frac{1}{h_t^2}\frac{\partial h_t}{\partial\mu}\frac{\partial h_t}{\partial\delta}\right]\frac{1}{n\sqrt{n}}\sum_{t=1}^{n}y_{t-1} + o_p(1); \tag{A.6}$$

see also the proofs of Lemmas 4.4 and 4.6 in Ling and Li (2003).

https://doi.org/10.1017/S0266466608090488 Published online by Cambridge University Press

For the item $s_n^{(2)}(\upsilon)$, we can show that

$$s_n^{(2)}(\upsilon) = -\frac{\upsilon_1}{n}\sum_{t=1}^{n}\frac{\operatorname{sgn}(e_t)}{\sqrt{h_t}}y_{t-1} - \frac{\upsilon_2'}{n}\sum_{t=1}^{n}\operatorname{sgn}(e_t)\frac{z_{t-1}}{\sqrt{h_t}}$$
$$+ f(0)\mathrm{E}\left[\frac{1}{h_t}\right]\frac{\upsilon_1^2}{n^2}\sum_{t=1}^{n}y_{t-1}^2 + 2f(0)\upsilon_2'\mathrm{E}\left[\frac{z_{t-1}}{h_t}\right]\frac{\upsilon_1}{n\sqrt{n}}\sum_{t=1}^{n}y_{t-1}$$
$$+ f(0)\upsilon_2'\mathrm{E}\left[\frac{z_{t-1}z_{t-1}'}{h_t}\right]\upsilon_2 + o_p(1). \tag{A.7}$$

Note that $s_n^{(2)}(\upsilon)$ is not a smooth function with respect to υ. We will finish the proof of this theorem first and then the proof of equation (A.7) because it is complicated.

By (A.3)–(A.7) and Lemma A.2, we have shown that

$$L_n(\theta) - L_n(\theta_0) \longrightarrow_d s(\upsilon) := -\upsilon_1\int_0^1 B_1(\tau)\,dB_2^{(1)}(\tau) - \upsilon_2'B_2(1) - \upsilon_3'B_3(1) + \upsilon'\Gamma\upsilon,$$

where Γ is defined in Theorem 2.1. Note that the random function $s(\upsilon)$ has a unique minimizer. Following Lemma 2.2 and Remark 1 of Davis, Knight, and Liu (1992), we complete the proof of Theorem 2.1. ■

Proof of Equation (A.7). It holds that, for $x, y \in R$ and $x \neq 0$,

$$|x-y| - |x| = -y\operatorname{sgn}(x) + 2\int_0^y I(x \le s) - I(x \le 0)\,ds,$$

where $\operatorname{sgn}(x)$ is equal to 1 for $x > 0$ and -1 for $x < 0$; see Knight (1998). Hence, we can rewrite the item $s_n^{(2)}(\upsilon)$ as follows:

$$s_n^{(2)}(\upsilon) = \sum_{t=1}^{n}\frac{1}{\sqrt{h_t(\theta)}}\left[\left|e_t - \frac{\upsilon_1}{n}y_{t-1} - \frac{\upsilon_2'}{\sqrt{n}}z_{t-1}\right| - |e_t|\right]$$
$$= \sum_{t=1}^{n}\frac{\operatorname{sgn}(e_t)}{\sqrt{h_t(\theta)}}\left[-\frac{\upsilon_1}{n}y_{t-1} - \frac{\upsilon_2'}{\sqrt{n}}z_{t-1}\right]$$
$$+ 2\sum_{t=1}^{n}\frac{1}{\sqrt{h_t(\theta)}}\int_0^{\upsilon_1 y_{t-1}/n + \upsilon_2' z_{t-1}/\sqrt{n}} I(e_t \le s) - I(e_t \le 0)\,ds, \tag{A.8}$$

where $z_t = (1, \Delta y_t, \ldots, \Delta y_{t-k+2})'$. By the inequalities given at the end of this Appendix, we know that

$$\sum_{t=1}^{n}\frac{\operatorname{sgn}(e_t)}{\sqrt{h_t(\theta)}}\left[-\frac{\upsilon_1}{n}y_{t-1} - \frac{\upsilon_2'}{\sqrt{n}}z_{t-1}\right] = -\frac{\upsilon_1}{n}\sum_{t=1}^{n}y_{t-1}\frac{\operatorname{sgn}(e_t)}{\sqrt{h_t}} - \frac{\upsilon_2'}{\sqrt{n}}\sum_{t=1}^{n}\operatorname{sgn}(e_t)\frac{z_{t-1}}{\sqrt{h_t}} + o_p(1).$$

For the second summation in (A.8), let

$$\zeta_t(\upsilon) = \frac{2}{\sqrt{h(\theta)}}\int_0^{\upsilon_1 y_{t-1}/n + \upsilon_2' z_{t-1}/\sqrt{n}} I(e_t \le s) - I(e_t \le 0)\,ds.$$

https://doi.org/10.1017/S0266466608090488 Published online by Cambridge University Press

Note that $\mathrm{E}(\zeta_t(\upsilon)|\mathcal{F}_{t-1}) = 2h^{-1/2}(\theta)\int_0^{\upsilon_1 y_{t-1}/n+\upsilon_2' z_{t-1}/\sqrt{n}} F\left(sh_t^{-1/2}\right) - F(0)\,ds$, where $F(\cdot)$ is the distribution function of ε_t. Denote

$$R_t = \mathrm{E}(\zeta_t(\upsilon)|\mathcal{F}_{t-1}) - \frac{f(0)}{\sqrt{h_t(\theta)h_t}}\left[\frac{\upsilon_1}{n}y_{t-1} + \frac{\upsilon_2'}{\sqrt{n}}z_{t-1}\right]^2$$
$$= \frac{2}{\sqrt{h_t(\theta)}}\int_0^{\upsilon_1 y_{t-1}/n+\upsilon_2' z_{t-1}/\sqrt{n}} F\left(sh_t^{-1/2}\right) - F(0) - sf(0)h_t^{-1/2}\,ds. \tag{A.9}$$

By Assumption 2.1, there exists a constant, say, π, such that the density function $f(x)$ is continuous on the set $\{x, |x| \le \pi\}$. For $\pi_1 \in (0, \pi)$, it holds that

$$|R_t| \le \frac{1}{\sqrt{\underline{\omega}}}\left|2\int_0^{\upsilon_1 y_{t-1}/n+\upsilon_2' z_{t-1}/\sqrt{n}} F\left(sh_t^{-1/2}\right) - F(0) - sf(0)h_t^{-1/2}ds\right|$$
$$\le \frac{1}{\sqrt{\underline{\omega}}}\sup_{|x|\le\pi_1}|f(x) - f(0)| \cdot \int_0^{|\upsilon_1 y_{t-1}/n+\upsilon_2' z_{t-1}/\sqrt{n}|} 2h_t^{-1/2}s\,ds$$
$$\le \frac{2}{n\sqrt{\underline{\omega}}}\sup_{|x|\le\pi_1}|f(x) - f(0)| \cdot \frac{\upsilon_1^2\xi_n^2 + |\upsilon_2' z_{t-1}|^2}{\sqrt{h_t}}$$

as $|\upsilon_1 y_{t-1}/n + \upsilon_2' z_{t-1}/\sqrt{n}| \le \pi_1$ and

$$|R_t| \le \frac{1}{\sqrt{\underline{\omega}}}\left|2\int_0^{\upsilon_1 y_{t-1}/n} F\left(sh_t^{-1/2}\right) - F(0)\,ds - \frac{\upsilon_1^2 f(0) y_{t-1}^2}{n^2\sqrt{h_t}}\right|$$
$$\le \frac{1}{\sqrt{\underline{\omega}}}\left(\frac{2}{\pi_1^2} + \frac{f(0)}{\pi_1}\right)\frac{1}{h_t}\left[\frac{|\upsilon_1 y_{t-1}|}{n} + \frac{|\upsilon_2' z_{t-1}|}{\sqrt{n}}\right]^3$$
$$\le \frac{4}{n\sqrt{\underline{\omega}}}\left(\frac{2}{\pi_1^2} + \frac{f(0)}{\pi_1}\right)\left[\frac{\upsilon_1^3\xi_n^3}{\sqrt{n}h_t} + \frac{|\upsilon_2' z_{t-1}|^2}{h_t}\cdot\max_{1\le t\le n} n^{-1/2}|\upsilon_2' z_t|\right]$$

as $|\upsilon_1 y_{t-1}/n + \upsilon_2' z_{t-1}/\sqrt{n}| > \pi_1$, where $\underline{\omega} = \inf_{\omega\in\Theta}\omega > 0$. Note that the item $\max_{1\le t\le n} n^{-1/2}|\upsilon_2' z_t| = o_p(1)$ and $\sup_{|x|\le\pi_1}|f(x) - f(0)|$ tends to zero as $\pi_1 \to 0$. Hence, by letting $n \to \infty$ and then $\pi_1 \to 0$, we can show that

$$\sum_{t=1}^n R_t = o_p(1). \tag{A.10}$$

Similarly, it can be shown that $n\mathrm{E}\zeta_t^2(\upsilon) = o(1)$, and then

$$\mathrm{E}\left[\sum_{t=1}^n \zeta_t(\upsilon) - \sum_{t=1}^n \mathrm{E}(\zeta_t(\upsilon)|\mathcal{F}_{t-1})\right]^2 \le n\mathrm{E}\zeta_t^2(\upsilon) = o(1),$$

which implies that

$$\sum_{t=1}^n \zeta_t(\upsilon) = \sum_{t=1}^n \mathrm{E}(\zeta_t(\upsilon)|\mathcal{F}_{t-1}) + o_p(1). \tag{A.11}$$

https://doi.org/10.1017/S0266466608090488 Published online by Cambridge University Press

By (A.9)–(A.11), Theorem 3.1 in Ling and Li (1998), and the inequalities at the end of this Appendix, we can show that

$$\begin{aligned}\sum_{t=1}^{n}\zeta_t(v) &= \sum_{t=1}^{n}\frac{f(0)}{h_t}\left[\frac{v_1}{n}y_{t-1}+\frac{v_2'}{\sqrt{n}}z_{t-1}\right]^2+o_p(1)\\ &= f(0)\mathrm{E}\left[\frac{1}{h_t}\right]\frac{v_1^2}{n^2}\sum_{t=1}^{n}y_{t-1}^2+2f(0)v_2'\mathrm{E}\left[\frac{z_{t-1}}{h_t}\right]\frac{v_1}{n\sqrt{n}}\sum_{t=1}^{n}y_{t-1}\\ &\quad+f(0)v_2'\mathrm{E}\left[\frac{z_{t-1}z_{t-1}'}{h_t}\right]v_2+o_p(1).\end{aligned}$$

Hence, equation (A.7) holds. ■

Some important inequalities. The Taylor expansion in (A.3) is readily obtained if higher order moments of e_t are assumed. However, only the second-order moments of e_t and ε_t are assumed in this paper, and hence the following inequalities are necessary for the proof of (A.3).

Denote $\xi_n = n^{-1/2}\max_{1\le i\le n}|y_i|$. By Doob's inequality (Hall and Heyde, 1980, p. 15), we know that $\mathrm{E}\xi_n^2 < \infty$ and $\xi_n = O_p(1)$. Let $G_n = (1/n, 1/\sqrt{n}, \ldots, 1/\sqrt{n})'$ be a $(k+p+q+2)$-dimensional vector, and denote, $\Theta_n = \{\theta : \|G_n'(\theta-\theta_0)\| \le M\}$, where M is a positive constant. Note that, for $i = 1, \ldots, q$ and $j \ge 0$, the inequality $h_t(\theta) > \beta_i h_{t-i}(\theta) > \beta_i u_q' B^j u_q h_{t-i-j}(\theta)$ is satisfied for all $\theta \in \Theta_n$, and then it is not difficult to show that

$$\sup_{\theta\in\Theta_n}\left|\frac{1}{\sqrt{h_t(\theta)}}\frac{\partial h_t(\theta)}{\partial\phi}\right| \le C\sqrt{n}\xi_n, \qquad \mathrm{E}\left\{\sup_{\theta\in\Theta_n}\left\|\frac{1}{\sqrt{h_t(\theta)}}\frac{\partial h_t(\theta)}{\partial\delta}\right\|\right\}^2 \le \infty,$$

$$\sup_{\theta\in\Theta_n}\left|\frac{1}{\sqrt{h_t(\theta)}}\frac{\partial^2 h_t(\theta)}{\partial\phi^2}\right| \le Cn\xi_n^2, \qquad \sup_{\theta\in\Theta_n}\left\|\frac{1}{\sqrt{h_t(\theta)}}\frac{\partial^2 h_t(\theta)}{\partial\phi\partial\delta}\right\| \le C\sqrt{n}\xi_n,$$

$$\mathrm{E}\left\{\sup_{\theta\in\Theta_n}\left\|\frac{1}{\sqrt{h_t(\theta)}}\frac{\partial^2 h_t(\theta)}{\partial\delta\partial\delta'}\right\|\right\}^2 \le \infty$$

and

$$\sup_{\theta\in\Theta_n}\left|\frac{1}{\sqrt{h_t(\theta)}}\frac{\partial^3 h_t(\theta)}{\partial\phi^2\partial\delta}\right| \le Cn\xi_n^2, \quad \sup_{\theta\in\Theta_n}\left\|\frac{1}{\sqrt{h_t(\theta)}}\frac{\partial^3 h_t(\theta)}{\partial\phi\partial\delta\partial\delta'}\right\| \le C\sqrt{n}\xi_n,$$

$$\mathrm{E}\left\{\sup_{\theta\in\Theta_n}\left\|\frac{1}{\sqrt{h_t(\theta)}}\frac{\partial^3 h_t(\theta)}{\partial\delta\partial\delta'\partial\delta}\right\|\right\}^2 \le \infty,$$

where C is a constant. ■

https://doi.org/10.1017/S0266466608090488 Published online by Cambridge University Press