Econometric Theory, **19**, 2003, 541–564. Printed in the United States of America.
DOI: 10.1017/S0266466603194029

ASYMPTOTIC INFERENCE FOR UNIT ROOT PROCESSES WITH GARCH(1,1) ERRORS

SHIQING LING
Hong Kong University of Science and Technology

W.K. LI
University of Hong Kong

This paper investigates the so-called one-step local quasi–maximum likelihood estimator for the unit root process with GARCH(1,1) errors. When the scaled conditional errors (the ratio of the disturbance to the conditional standard deviation) follow a symmetric distribution, the asymptotic distribution of the estimated unit root is derived only under the second-order moment condition. It is shown that this distribution is a functional of a bivariate Brownian motion as in Ling and Li (1998, *Annals of Statistics* 26, 84–125) and can be used to construct the unit root test.

1. INTRODUCTION

Consider the unit root process with the first-order general conditional heteroskedastic errors [GARCH(1,1)]:

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

$$\varepsilon_t = \eta_t \sqrt{h_t}, \qquad h_t = \alpha_0 + \alpha \varepsilon_{t-1}^2 + \beta h_{t-1}, \tag{1.2}$$

where $\phi = 1$ and $\alpha_0 > 0$, $\alpha \geq 0$, and $\beta \geq 0$, and η_t's are a sequence of independently and identically distributed (i.i.d.) random variables with zero mean and variance one.

The GARCH models were proposed by Bollerslev (1986) and have important applications in financial and econometric time series. Some reviews can be found in Bollerslev, Engle, and Nelson (1994). When $\alpha = \beta = 0$, ε_t's defined

The authors thank the co-editor, Bruce Hansen, and two referees for very helpful comments and suggestions. W.K. Li's research is partially supported by the Hong Kong Research Grants Council. Ling's research is supported by RGC Competitive Earmarked Research grant HKUST6113/02P. Address correspondence to: W.K. Li, Department of Statistics and Actuarial Science, The University of Hong Kong, Pokfulam Road, Hong Kong.

 0266-4666/03 $12.00

by model (1.2) reduce to i.i.d. white noises, and in this case the unit root process has been investigated for a long time. In recent decades, motivated by the practical applications in statistics and econometrics, 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 references therein. When the error terms follow a GARCH model, how to estimate and how to test the unit root are obviously important problems.

Ling and Li (1998) derive the limiting distribution of the local maximum likelihood estimator (MLE) for a general nonstationary autoregressive moving-average time series with general-order GARCH errors and demonstrate that the MLE is more efficient than the least squares estimator (LSE). Seo (1999) also independently has derived the limiting distribution of the local MLE unit root in the nonstationary AR(p) model. The simulation results in Seo (1999) and Ling, Li, and McAleer (2001) show that the unit root tests based on the MLE are not only more powerful than Dickey–Fuller tests based on the LSE but also have more stable sizes. However, the results in Ling and Li (1998) require that $E\varepsilon_t^4 < \infty$, whereas those in Seo (1999) require that $E\varepsilon_t^8 < \infty$.

Note that the condition for strict stationarity is $\ln(\alpha\eta_t^2 + \beta) < 0$ (see Nelson, 1990), the condition for $E\varepsilon_t^2 < \infty$ is $\alpha + \beta < 1$, and the condition for $E\varepsilon_t^4 < \infty$ is $3\alpha^2 + 2\alpha\beta + \beta^2 < 1$. The conditions for $E\varepsilon_t^4 < \infty$ or $E\varepsilon_t^8 < \infty$ are clearly much more stringent. For the pure GARCH(1,1) model, Lee and Hansen (1994) and Lumsdaine (1996) prove that MLE are consistent and asymptotically normal under $\ln(\alpha\eta_t^2 + \beta) < 0$. A challenging problem is whether or not we can derive the limiting distribution of the MLE under weaker conditions for the unit root process with GARCH errors. When η_t is symmetrically distributed, in this paper we obtain the asymptotic distribution of the so-called one-step local quasi-MLE of the unit root in model (1.1) under the assumption that $\alpha + \beta < 1$, that is, only the existence of the second moment of ε_t is required. In the literature on unit root with GARCH, this is the weakest condition for the unit root distribution to exist. This limiting distribution is a functional of a bivariate Brownian motion and is also the same as that obtained in Ling and Li (1998).

This paper proceeds as follows. Section 2 presents the one-step local MLE and main results. Section 3 extends the results in Section 2 to models with a constant intercept. The proof of main results is given in Section 4.

Throughout the paper, 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; and $D = D[0,1]$ denotes the space of functions $f(s)$ on $[0, 1]$, which is defined and equipped with the Skorokhod topology (Billingsley, 1968).

2. ONE-STEP LOCAL QMLE AND MAIN RESULTS

Given observations $y_1, \ldots, y_n$ with the initial value $y_0 = 0$, generated by model (1.1), the log-likelihood function (ignoring a constant) can be written as

$$L(\phi, \tilde{\delta}) = \frac{1}{n} \sum_{t=1}^{n} l_t(\phi, \tilde{\delta}) \quad \text{and} \quad l_t(\phi, \tilde{\delta}) = -\frac{1}{2} \ln h_t(\phi, \tilde{\delta}) - \frac{\varepsilon_t^2(\phi)}{2h_t(\phi, \tilde{\delta})}, \qquad \textbf{(2.1)}$$

where ϕ and $\tilde{\delta} = (\tilde{\alpha}_0, \tilde{\alpha}, \tilde{\beta})$ are unknown parameters, $\varepsilon_t(\phi) = y_t - \phi y_{t-1}$, and $h_t(\phi, \tilde{\delta}) = \tilde{\alpha}_0 + \tilde{\alpha}\varepsilon_t^2(\phi) + \tilde{\beta} h_{t-1}(\phi, \tilde{\delta})$ with $\varepsilon_0(\phi) = \varepsilon_0$ and $h_0(\phi, \tilde{\delta}) = h_0$. The true values of ϕ and $\tilde{\delta}$ are 1 and $\delta = (\alpha_0, \alpha, \beta)$, respectively. We make the following assumptions.

Assumption 1. $\Theta = \{(\tilde{\alpha}_0, \tilde{\alpha}, \tilde{\beta}) : 0 < \alpha_{0l} \le \tilde{\alpha}_0 \le \alpha_{0u}, 0 < \alpha_l \le \tilde{\alpha} \le \alpha_u, 0 < \beta_l \le \tilde{\beta} \le \beta_u, \tilde{\alpha} + \tilde{\beta} < 1\}$, $\delta \in \Theta$ and $\tilde{\delta} \in \Theta$.

Assumption 2. η_t has a symmetric distribution and $E\eta_t^4 < \infty$.

Because we do not assume that η_t is normal, the maximizer of $L(\phi, \tilde{\delta})$ on $R \times \Theta$ is called the quasi–maximum likelihood estimator (QMLE) of $\phi = 1$ and δ. In practice, ε_0 and h_0 are unavailable and can be replaced by some constants. These initial values do not affect our asymptotic results, which can be verified via some arguments similar to those in Lee and Hansen (1994).

Let $\hat{\phi}_{LS}$ be the LSE of the unit root $\phi = 1$ in model (1.1). Then $\hat{\phi}_{LS} = (\sum_{t=1}^{n} y_{t-1}^2)^{-1}(\sum_{t=1}^{n} y_t y_{t-1})$. The residual $\hat{\varepsilon}_t = y_t - \hat{\phi}_{LS} y_{t-1}$ can be used as the artificial observations of ε_t to estimate δ in model (1.2) through MLE as in Lee and Hansen (1994). Because $\hat{\phi}_{LS} - 1 = O_p(n^{-1})$[see (2.9), which follows], Theorem C in Appendix C shows that Hessian matrices based on ε_t and $\hat{\varepsilon}_t$ are asymptotically equivalent. The corresponding asymptotic equivalence on the log-likelihood functions and the score functions can be found in Lemma 2.1 in Ling et al. (2001). Thus, if Assumptions 1 and 2 hold then the MLE of δ based on $\hat{\varepsilon}_t$ is asymptotically equivalent in probability to that based on the true ε_t as in Lee and Hansen (1994). Hence, we assume that the estimator $\hat{\delta}_n$ of δ has been obtained and $\sqrt{n}(\hat{\delta}_n - \delta) = O_p(1)$.

Using $\hat{\delta}_n$ and an initial estimator $\tilde{\phi}_n$ with $n(\tilde{\phi}_n - 1) = O_p(1)$, the one-step local QMLE of $\phi = 1$, denoted by $\hat{\phi}_n$, is obtained by the one-step iteration

$$\hat{\phi}_n = \tilde{\phi}_n - \left[\sum_{t=1}^{n} \frac{\partial^2 l_t(\phi, \hat{\delta}_n)}{\partial \phi^2} \right]_{\phi = \tilde{\phi}_n}^{-1} \left[\sum_{t=1}^{n} \frac{\partial l_t(\phi, \hat{\delta}_n)}{\partial \phi} \right]_{\phi = \tilde{\phi}_n}. \qquad \textbf{(2.2)}$$

When $(\phi, \tilde{\delta}) = (1, \delta)$, we abbreviate $\partial l_t(\phi, \tilde{\delta})$ and $\partial^2 l_t(\phi, \tilde{\delta})$ to ∂l_t and $\partial^2 l_t$, respectively. If Assumptions 1 and 2 hold, then $n^{-3/2} \sum_{t=1}^{n} \partial^2 l_t / \partial\phi\partial\tilde{\delta} = o_p(1)$ by Lemma 6.5 in Ling et al. (2001), and hence, by Theorem 3.1 in the same paper, we have

$$\frac{1}{n^2}\sum_{t=1}^{n}\left[\frac{\partial^2 l_t(\phi,\hat{\delta}_n)}{\partial\phi^2}-\frac{\partial^2 l_t}{\partial\phi^2}\right]=o_p(1), \tag{2.3}$$

$$\frac{1}{n}\sum_{t=1}^{n}\left[\frac{\partial l_t(\phi,\hat{\delta}_n)}{\partial\phi}-\frac{\partial l_t}{\partial\phi}\right]=n(\phi-1)\left[\frac{1}{n^2}\sum_{t=1}^{n}\frac{\partial^2 l_t}{\partial\phi^2}\right]+o_p(1) \tag{2.4}$$

uniformly in the ball $\Theta_n=\{\phi:|n(\phi-1)|\le M\}$ for any fixed positive constant M. Thus, by (2.2)–(2.4), we have that

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

$$=-\left[\frac{1}{n^2}\sum_{t=1}^{n}\frac{\partial^2 l_t}{\partial\phi^2}\right]^{-1}\left[\sum_{t=1}^{n}\frac{1}{n}\frac{\partial l_t}{\partial\phi}\right]+o_p(1). \tag{2.5}$$

As pointed out by a referee and the co-editor Professor Bruce Hansen, $\hat{\phi}_n$ is not the local QMLE in the usual sense, but it has the same asymptotic distribution as the QMLE. We call $\hat{\phi}_n$ the one-step QMLE. In practice, we can use $\hat{\phi}_n$ as a new initial value to repeat the iterative procedure (2.2), and, for the estimated value from each iterative procedure, it has the same asymptotic representation as (2.5). The following is our main result.

THEOREM 2.1. *Let $\hat{\phi}_n$ be the estimator of the unit root $\phi=1$ such that (2.5) holds. If Assumptions 1 and 2 are satisfied, then*

$$n(\hat{\phi}_n-1)\xrightarrow{\mathcal{L}}\frac{\int_0^1 w_1(\tau)\,dw_2(\tau)}{F\int_0^1 w_1^2(\tau)\,d\tau},$$

where $(w_1(\tau),w_2(\tau))$ is a bivariate Brownian motion with covariance

$$\tau\Omega=\tau\begin{pmatrix} Eh_t & 1\\ 1 & E(1/h_t)+\kappa\alpha^2\sum_{k=1}^{\infty}\beta^{2(k-1)}E(\varepsilon_{t-k}^2/h_t^2)\end{pmatrix}, \tag{2.6}$$

$F=E(1/h_t)+2\alpha^2\sum_{k=1}^{\infty}\beta^{2(k-1)}E(\varepsilon_{t-k}^2/h_t^2)$ and $\kappa=E\eta_t^4-1$. In particular, when η_t is normal, $\kappa=2$.

Remark. Let

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

where $\sigma^2 = Eh_t$ and K is the (2,2)th element of Ω in (2.6). Then $B_1(\tau)$ and $B_2(\tau)$ are two independent standard Brownian motions. As shown in Ling and Li (1998),

$$n(\hat{\phi}_n - 1) \xrightarrow{\mathcal{L}} \frac{\int_0^1 B_1(\tau)\, dB_1(\tau)}{\sigma^2 F \int_0^1 B_1^2(\tau)\, d\tau} + \frac{\sqrt{\sigma^2 K - 1}}{\sigma^2 F} \frac{\int_0^1 B_1(\tau)\, dB_2(\tau)}{\int_0^1 B_1^2(\tau)\, d\tau}. \tag{2.7}$$

The second term of (2.7) can be simplified as $[\sqrt{\sigma^2 K - 1}/F\sigma^2] \times (\int_0^1 B_1^2(\tau)\, d\tau)^{-1/2}\xi$, where ξ is a standard normal random variable independent of $\int_0^1 B_1^2(\tau)\, d\tau$ (see Phillips, 1989). Let $c = \sigma F \big/ \sqrt{K}$ and $\rho^2 = 1/\sigma^2 K \in (0,1)$. We can obtain that

$$nc(\hat{\phi}_n - 1) \xrightarrow{\mathcal{L}} \frac{\rho \int_0^1 B_1(\tau)\, dB_1(\tau)}{\int_0^1 B_1^2(\tau)\, d\tau} + \sqrt{1-\rho^2} \left(\int_0^1 B_1^2(\tau)\, d\tau \right)^{-1/2} \xi. \tag{2.8}$$

Under Assumptions 1 and 2, Ling et al. (2001) show that

$$n(\hat{\phi}_{LS} - 1) \xrightarrow{\mathcal{L}} \frac{\int_0^1 B_1(\tau)\, dB_1(\tau)}{\int_0^1 B_1^2(\tau)\, d\tau}. \tag{2.9}$$

From (2.8) and (2.9), we see that the asymptotic distribution of $\hat{\phi}_n$ is a combination of that of $\hat{\phi}_{LS}$ and a scale mixture of normals. This property is similar to those of the least absolute deviation estimators of unit roots given by Herce (1996). Ling and Li (1998) show that $\hat{\phi}_n$ is more efficient than $\hat{\phi}_{LS}$, in the sense defined in Ling and McAleer (2003). Our result heavily relies on the symmetry assumption. When η_t is asymmetric, the MLEs of ϕ and δ are not asymptotically independent, and hence (2.4) and (2.5) do not hold. In this case, the limiting distribution of the local MLE of (ϕ, δ) can be obtained by using a similar method to that in Ling and McAleer (2003).

3. MODELS WITH A CONSTANT INTERCEPT

The co-editor, Bruce Hansen, pointed out that many economic data include an intercept. In this section, we consider the following model:

$$y_t = \mu + \phi y_{t-1} + \varepsilon_t, \tag{3.1}$$

$$\varepsilon_t = \eta_t \sqrt{h_t}, \qquad h_t = \alpha_0 + \alpha \varepsilon_{t-1}^2 + \beta h_{t-1}, \tag{3.2}$$

where $\delta = (\alpha_0, \alpha, \beta)$ is defined as in (2.1). Denote $\lambda = (\mu, \phi)'$ and assume that its true value is $\lambda_0 = (0,1)'$. Let $\hat{\lambda}_{LS} = (\hat{\mu}_{LS}, \hat{\phi}_{\mu LS})'$ be the LSE of λ_0. Then

$$\hat{\lambda}_{LS} = \begin{pmatrix} n & \sum_{t=1}^{n} y_{t-1} \\ \sum_{t=1}^{n} y_{t-1} & \sum_{t=1}^{n} y_{t-1}^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum_{t=1}^{n} y_t \\ \sum_{t=1}^{n} y_t y_{t-1} \end{pmatrix}. \tag{3.3}$$

Let the residual $\hat{\varepsilon}_t = y_t - \hat{\mu}_{LS} - \hat{\phi}_{LS} y_{t-1}$. Under Assumptions 1 and 2, Ling et al. (2001) prove that the MLE of δ based on $\hat{\varepsilon}_t$ is asymptotically equivalent in probability to that based on the true ε_t as in Lee and Hansen (1994). As in Section 2, we can assume that the estimator $\hat{\delta}_n$ of δ has been obtained and $\sqrt{n}(\hat{\delta}_n - \delta) = O_p(1)$.

The log-likelihood function for model (3.1) and (3.2) is similarly defined as (2.1) with ϕ replaced by λ. Let $N_n = \operatorname{diag}\{\sqrt{n}, n\}$. Using an initial value $\tilde{\lambda}_n$ with $N_n(\tilde{\lambda}_n - \lambda_0) = O_p(1)$, the one-step local QMLE $\hat{\lambda}_n$ of λ_0 is obtained by the one-step iterative procedure as (2.2) with ϕ replaced by λ. If Assumptions 1 and 2 hold, then $n^{-1/2} N_n^{-1} \sum_{t=1}^{n} \partial^2 l_t / \partial\lambda\partial\tilde{\delta} = o_p(1)$ by Lemma 6.5 in Ling et al. (2001), and hence, by Theorem 3.1 in the same paper, we can obtain the asymptotic representation

$$N_n(\hat{\lambda}_n - \lambda_0) = -\left[N_n^{-1} \sum_{t=1}^{n} \frac{\partial^2 l_t}{\partial\lambda\partial\lambda'} N_n^{-1} \right]^{-1} \left[N_n^{-1} \sum_{t=1}^{n} \frac{\partial l_t}{\partial\lambda} \right] + o_p(1). \tag{3.4}$$

In Appendix B, we show that

$$\frac{1}{n^{3/2}} \sum_{t=1}^{n} \frac{\partial^2 l_t}{\partial\phi\partial\mu} \xrightarrow{\mathcal{L}} F \int_0^1 w_1(\tau)\, d\tau. \tag{3.5}$$

Thus, by Theorem 2.1, Lemma 4.2 in the next section, (3.5), and the continuity mapping theorem, we have the following result.

THEOREM 3.1. *Let $\hat{\lambda}_n = (\hat{\mu}_n, \hat{\phi}_{\mu n})'$ be the estimator of $\lambda_0 = (0,1)'$ such that (3.4) holds. If Assumptions 1 and 2 hold, then*

$$N_n \begin{pmatrix} \hat{\mu}_n \\ \hat{\phi}_{\mu n} - 1 \end{pmatrix} \xrightarrow{\mathcal{L}} F^{-1} \begin{pmatrix} 1 & \int_0^1 w_1(\tau)\, d\tau \\ \int_0^1 w_1(\tau)\, d\tau & \int_0^1 w_1^2(\tau)\, d\tau \end{pmatrix}^{-1} \begin{pmatrix} w_2(1) \\ \int_0^1 w_1(\tau)\, dw_2(\tau) \end{pmatrix},$$

where F and $(w_1(\tau), w_2(\tau))$ are defined as in Theorem 2.1.

Remark. Similar to (2.8), by Theorem 3.1, we can show that

$$nc(\hat{\phi}_{\mu n}-1) \xrightarrow{\mathcal{L}} \frac{\rho\int_0^1 B_1(\tau)\,dB_1(\tau) - B_1(1)\int_0^1 B_1(\tau)\,d\tau}{\left[\int_0^1 B_1^2(\tau)\,d\tau - \left(\int_0^1 B_1(\tau)\,d\tau\right)^2\right]} + \sqrt{1-\rho^2}\left[\int_0^1 B_1^2(\tau)\,d\tau - \left(\int_0^1 B_1(\tau)\,d\tau\right)^2\right]^{-1/2}\xi, \tag{3.6}$$

where c and ρ are defined as in (2.8). Under Assumptions 1 and 2, Ling et al. (2001) show that

$$n(\hat{\phi}_{\mu LS}-1) \xrightarrow{\mathcal{L}} \frac{\int_0^1 B_1(\tau)\,dB_1(\tau) - B_1(1)\int_0^1 B_1(\tau)\,d\tau}{\left[\int_0^1 B_1^2(\tau)\,d\tau - \left(\int_0^1 B_1(\tau)\,d\tau\right)^2\right]}. \tag{3.7}$$

From (3.6) and (3.7), we see that the limiting distribution of $\hat{\phi}_{\mu n}$ is a combination of that of $\hat{\phi}_{\mu LS}$ and a scale mixture of normals. Some critical values of limiting distributions in (2.8) and (3.6) with different ρ are given in Ling et al. (2001), and those for the corresponding t-statistics are given in Seo (1999) and Ling et al. (2001).

4. PROOF OF MAIN RESULTS

LEMMA 4.1. *Under Assumption 1, the processes h_t and ε_t defined by model (1.2) are strictly stationary and ergodic and have the expansions*

$$h_t = \alpha_0\left[1+\sum_{k=1}^{\infty}\prod_{i=1}^{k}(\alpha\eta_{t-i}^2+\beta)\right] \quad a.s. \tag{4.1}$$

and

$$\varepsilon_t = \alpha_0^{1/2}\eta_t\left[1+\sum_{k=1}^{\infty}\prod_{i=1}^{k}(\alpha\eta_{t-i}^2+\beta)\right]^{1/2} \quad a.s. \tag{4.2}$$

Proof. This comes straightforwardly from Theorem 2 in Nelson (1990) (for another expansion, see also Ling and Li, 1997). ■

LEMMA 4.2. *Suppose that the process ε_t is generated by model (1.2) and Assumptions 1 and 2 are satisfied. Then*

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

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 2.1.

Proof. Let $\lambda = (\lambda_1, \lambda_2)'$ be a constant vector with $\lambda\lambda' \neq 0$. Denote

$$\xi_t = \lambda_1 \varepsilon_t + \lambda_2 \left[\frac{\varepsilon_t}{h_t} - \frac{\alpha}{h_t} \left(\frac{\varepsilon_t^2}{h_t} - 1 \right) \sum_{k=1}^{t-1} \beta^{k-1} \varepsilon_{t-k} \right],$$

$$\xi_t^* = \lambda_1 \varepsilon_t + \lambda_2 \left[\frac{\varepsilon_t}{h_t} - \frac{\alpha}{h_t} \left(\frac{\varepsilon_t^2}{h_t} - 1 \right) \sum_{k=1}^{\infty} \beta^{k-1} \varepsilon_{t-k} \right].$$

By Lemma 4.1, it is easy to show that both ξ_t and ξ_t^* are martingale differences with respect to $\mathcal{F}_t$, where $\mathcal{F}_t$ is the σ-field generated by $\{\eta_t, \eta_{t-1}, \ldots\}$. First we consider the asymptotic property of $S^*_{[n\tau]} = \sum_{t=1}^{[n\tau]} \xi_t^* / \sqrt{n}$. From Lemma 4.1, we see that h_t is a function in terms of $\{\eta^2_{t-1}, \eta^2_{t-2}, \ldots\}$. Note that $\varepsilon_t = \eta_t \sqrt{h_t}$ and η_t is symmetric. It is easy to see that $E(\varepsilon_{t-i}\varepsilon_{t-j}/h_t^2) = 0$ as $i \neq j$, and hence

$$\sigma^{*2} = ES_n^{*2} = \lambda_1^2 \sigma^2 + 2\lambda_1\lambda_2 + \lambda_2^2 \left[E\left(\frac{1}{h_t}\right) + \kappa\alpha^2 \sum_{k=1}^{\infty} \beta^{2(k-1)} E\left(\frac{\varepsilon_{t-k}^2}{h_t^2}\right) \right]$$

$$= \lambda'\Omega\lambda < \infty,$$

where $\sigma^2 = Eh_t$. By Assumption 2 and the Cauchy–Schwarz inequality, $\sigma^2 E(1/h_t) > 1$ and hence $\sigma^{*2} > \lambda_1\sigma^2 + 2\lambda_1\lambda_2 + \lambda_2^2 E(1/h_t) = (\sigma\lambda_1 + \lambda_2/\sigma)^2 + \lambda_2^2[\sigma^2 E(1/h_t) - 1]/\sigma^2 > 0$. Note that

$$E(\xi_t^{*2} | \mathcal{F}_{t-1}) = \lambda_1^2 h_t + 2\lambda_1\lambda_2 + \lambda_2^2 \left[\frac{1}{h_t} + \kappa \left(\frac{\alpha}{h_t} \sum_{k=1}^{\infty} \beta^{k-1} \varepsilon_{t-k} \right)^2 \right].$$

By Lemma 4.1, $\{E(\xi_t^{*2}|\mathcal{F}_{t-1})\}$ is a strictly stationary and ergodic time series. By the ergodic theorem and Assumption 2, it is easy to show that

$$[ES_n^{*2}]^{-1} \frac{1}{n} \sum_{t=1}^{n} E(\xi_t^{*2} | \mathcal{F}_{t-1}) \to 1 \quad \text{a.s.} \tag{4.3}$$

Furthermore, because $\{\xi_t^*\}$ is also a strictly stationary and ergodic time series with finite variance, it follows that, for any small $\epsilon > 0$,

$$\frac{1}{n} \sum_{t=1}^{n} E[\xi_t^{*2} I(\xi_t^* \geq \sqrt{n}\sigma^*\epsilon)]$$

$$= E[\xi_t^{*2} I(\xi_t^* \geq \sqrt{n}\sigma^*\varepsilon)] = \int_{x > \sqrt{n}\sigma^*\epsilon} x^2 \, dP(x) \to 0, \tag{4.4}$$

as $n \to \infty$, where $P(x)$ is the distribution function of ξ_t^*. By (4.3) and (4.4) and the invariance principle for martingale (Hall and Heyde, 1980, Theorem 4),

$$S_{[n\tau]}^* = \frac{1}{\sqrt{n}} \sum_{t=1}^{[n\tau]} \xi_t^* \xrightarrow{\mathcal{L}} W^*(\tau) \quad \text{in } D, \tag{4.5}$$

where $W^*(\tau)$ is a Brownian motion with variance $\sigma^{*2}\tau$.

Let $S_{[n\tau]} = \sum_{t=1}^{[n\tau]} \xi_t / \sqrt{n}$. Note that

$$\begin{aligned} E|S_{[n\tau]} - S_{[n\tau]}^*| &\le \frac{1}{\sqrt{n}} \sum_{t=1}^{[n\tau]} E\left| \frac{\lambda_2}{h_t}\left(\frac{\varepsilon_t^2}{h_t} - 1\right)\left(\alpha \sum_{k=t}^{\infty} \beta^{k-1}\varepsilon_{t-k}\right)\right| \\ &\le \frac{c}{\sqrt{n}} \sum_{t=1}^{[n\tau]} \left(\sum_{k=t}^{\infty} \beta^{k-1} E|\varepsilon_{t-k}|\right) = \frac{c}{\sqrt{n}} \sum_{t=1}^{[n\tau]} O(\beta^t) = o(1), \end{aligned} \tag{4.6}$$

where c is a constant and $o(1)$ holds uniformly in $\tau \in [0,1]$. By (4.5) and (4.6),

$$S_{[n\tau]} = \frac{1}{\sqrt{n}} \sum_{t=1}^{[n\tau]} \xi_t \xrightarrow{\mathcal{L}} W^*(\tau) \quad \text{in } D.$$

Furthermore, by Cramér's device, we complete the proof. ■

LEMMA 4.3. *Suppose that Assumption 1 holds. Then:*

(1) *for any k,*

(a) $h_t = \dfrac{\alpha_0}{1-\beta} + \alpha \sum_{k=1}^{\infty} \beta^{k-1}\varepsilon_{t-k}^2$ *a.s.*;

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

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

(2) *let* $h_{t-k,l} = \alpha_0[1 + \sum_{j=1}^{l-k}\prod_{i=1}^{j}(\alpha\eta_{t-i-k}^2 + \beta)]$ *and* $\varepsilon_{t-k,l} = \eta_{t-k}\sqrt{h_{t-k,l}}$; *it follows that, for* $k = 0,1,\dots,l-1$,

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

(b) $\dfrac{h_{t-k,l}}{h_{t,l}} \le \left[\prod_{i=1}^{k}(\alpha\eta_{t-i}^2 + \beta)\right]^{-1}$ *a.s.*;

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

Proof.

(1) Because $0 < \beta < 1$ and $E\varepsilon_t^2 < \infty$, (a) holds obviously. By (a), (b) holds obviously. Because $h_{t-i+1} > (\alpha\eta_{t-i}^2 + \beta)h_{t-i}$ a.s., we have that $h_{t-k}/h_t = \prod_{i=1}^k (h_{t-i}/h_{t-i+1}) < [\prod_{i=1}^k (\alpha\eta_{t-i}^2 + \beta)]^{-1}$ a.s., that is, (c) holds.

(2) It is easy to see that $h_{t,l} = \alpha_0[1 + \sum_{j=1}^{k-1}\prod_{i=1}^j (\alpha\eta_{t-i}^2 + \beta)] + \prod_{i=1}^k (\alpha\eta_{t-i}^2 + \beta)h_{t-k,l}$, and hence we can show that (a) holds. Similarly, we can show that (b) holds. By Lemma 4.1, $E|h_{t-k} - h_{t-k,l}| = \alpha_0 \sum_{j=l-k+1}^{\infty} E[\prod_{i=1}^j (\alpha\eta_{t-k-i}^2 + \beta)] = \alpha_0 \sum_{j=l-k+1}^{\infty} (\alpha + \beta)^j = O(\rho^{l-k+1})$, that is, (c) holds. This completes the proof. ■

LEMMA 4.4. *Suppose that Assumptions 1 and 2 hold and g_t is one of the following types of random variables:*

(a) $\dfrac{1}{h_t}$,

(b) $\dfrac{\varepsilon_t^2}{h_t^3}\left(\sum_{k=1}^{t-1}\beta^{k-1}\varepsilon_{t-k}\right)^2$,

(c) $\dfrac{1}{h_t^2}\left(1 - \dfrac{2\varepsilon_t^2}{h_t}\right)\left(\sum_{k=1}^{t-1}\beta^{k-1}\varepsilon_{t-k}\right)\left(\sum_{k=1}^{t-1}\beta^{k-1}h_{t-k}\right)$.

Then $E[\sum_{t=1}^n (g_t - Eg_t)]^2/n$ converges to a constant σ_0^2 and

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{[n\tau]}(g_t - Eg_t) \xrightarrow{\mathcal{L}} \sigma_0 w_0(\tau) \quad \text{in } D,$$

where $w_0(\tau)$ is a standard Brownian motion.

Remark. Davis, Mikosch, and Basrak (1999) prove that ε_t and h_t are strongly mixing with geometric rates. As a referee pointed out, it is possible to prove Lemma 4.4 under some mixing framework. However, one has not shown in the literature that $\sum_{k=1}^{\infty}\beta^{k-1}\varepsilon_{t-k}$ and $\sum_{k=1}^{\infty}\beta^{k-1}h_{t-k}$, which follow, are strongly mixing. Our proof uses Theorem 21.1 in Billingsley (1968) and heavily relies on Lemma 4.1.

Proof. Because the proofs of (a) and (b) are similar, we present only the proofs of (b) and (c). For (b), denote

$$g_t^* = \frac{\varepsilon_t^2}{h_t^3}\left(\sum_{k=1}^{\infty}\beta^{k-1}\varepsilon_{t-k}\right)^2 \quad \text{and} \quad g_{t,l}^* = \frac{\varepsilon_{t,l}^2}{h_{t,l}^3}\left(\sum_{k=1}^{l-1}\beta^{k-1}\varepsilon_{t-k,l}\right)^2,$$

where $\varepsilon_{t,l}$, $h_{t,l}$, and $\varepsilon_{t-k,l}$ are defined as in Lemma 4.3. By Lemma 4.1, g_t^* is a measurable function of $\eta_t, \eta_{t-1}, \ldots$. Meanwhile $g_{t,l}^*$ is a measurable function of $\{\eta_t, \ldots, \eta_{t-l}\}$.

$$\begin{aligned}
E|g_t^* - g_{t,l}^*|^2 &= E\left|\frac{1}{h_t^2}\left(\sum_{k=1}^{\infty}\beta^{k-1}\varepsilon_{t-k}\right)^2 - \frac{1}{h_{t,l}^2}\left(\sum_{k=1}^{l-1}\beta^{k-1}\varepsilon_{t-k,l}\right)^2\right|^2 \\
&\le 3E\left|\frac{1}{h_t^2}\left(\sum_{k=1}^{l-1}\beta^{k-1}\varepsilon_{t-k}\right)^2 - \frac{1}{h_t^2}\left(\sum_{k=1}^{l-1}\beta^{k-1}\varepsilon_{t-k,l}\right)^2\right|^2 \\
&\quad + 3E\left|\left(\frac{1}{h_t^2} - \frac{1}{h_{t,l}^2}\right)\left(\sum_{k=1}^{l-1}\beta^{k-1}\varepsilon_{t-k,l}\right)^2\right|^2 \\
&\quad + 3E\left|\frac{1}{h_t^2}\left(\sum_{k=1}^{\infty}\beta^{k-1}\varepsilon_{t-k}\right)^2 - \frac{1}{h_t^2}\left(\sum_{k=1}^{l-1}\beta^{k-1}\varepsilon_{t-k}\right)^2\right|^2 \\
&\equiv 3EI_1^2 + 3EI_2^2 + 3EI_3^2. \qquad \textbf{(4.7)}
\end{aligned}$$

In Appendix A, we prove that there exists a $\rho \in (0,1)$ such that

$$EI_1^2 = O(\rho^l), \qquad EI_2^2 = O(\rho^l), \qquad \text{and} \quad EI_3^2 = O(\rho^l). \qquad \textbf{(4.8)}$$

By (4.7) and (4.8), it is easy to show that

$$\sum_{l=1}^{\infty}[E|(g_t^* - Eg_t^*) - (g_{t,l}^* - Eg_{t,l}^*)|^2]^{1/2} < \infty. \qquad \textbf{(4.9)}$$

By Theorem 21.1 in Billingsley (1968), $E[\sum_{t=1}^{n}(g_t^* - Eg_t^*)]^2/n$ converges absolutely to a constant σ_0^2 and when $\sigma_0 > 0$,

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{[n\tau]}(g_t^* - Eg_t^*) \xrightarrow{\mathcal{L}} \sigma_0 w_0(\tau) \quad \text{in } D. \qquad \textbf{(4.10)}$$

When $\sigma_0 = 0$, it is not difficult to verify that the conditions C1–C4 of Theorem 3.2 in Ling and Li (1998) are satisfied and by the theorem,

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{[n\tau]}(g_t^* - Eg_t^*) \xrightarrow{\mathcal{L}} 0 \quad \text{in } D. \qquad \textbf{(4.11)}$$

Using a similar method as for (4.8), we can show that $E(g_t - g_t^*)^2 = O(\rho^t)$ and hence $\sum_{t=1}^{[n\tau]}E(g_t - g_t^*)^2/\sqrt{n} = o_p(1)$, where $o_p(1)$ holds uniformly in $\tau \in [0,1]$. Furthermore, by (4.10) and (4.11), we can claim that the conclusion with case (b) holds.

For case (c), let

$$g_t^* = \frac{1}{h_t^2}\left(1 - \frac{2\varepsilon_t^2}{h_t}\right)\left(\sum_{k=1}^{\infty}\beta^{k-1}\varepsilon_{t-k}\right)\left(\sum_{k=1}^{\infty}\beta^k h_{t-k}\right),$$

$$g_{t,l}^* = \frac{1}{h_{t,l}^2}\left(1 - \frac{2\varepsilon_{t,l}^2}{h_{t,l}}\right)\left(\sum_{k=1}^{l-1}\beta^{k-1}\varepsilon_{t-k,l}\right)\left(\sum_{k=1}^{l-1}\beta^k h_{t-k,l}\right).$$

Because $1 - 2\varepsilon_t^2/h_t = 1 - 2\varepsilon_{t,l}^2/h_{t,l} = 1 - 2\eta_t^2$ is independent of $\mathcal{F}_{t-1}$, it follows that

$$E|g_t^* - g_{t,l}^*|^2 \le 3cE\left[\left|\frac{1}{h_t^2}\sum_{k=1}^{\infty}\beta^k h_{t-k} - \frac{1}{h_t^2}\sum_{k=1}^{l-1}\beta^k h_{t-k,l}\right|^2 \left|\sum_{k=1}^{\infty}\beta^{k-1}\varepsilon_{t-k}\right|^2\right]$$
$$+ 3cE\left[\left|\frac{1}{h_t^2} - \frac{1}{h_{t,l}^2}\right|^2 \left(\sum_{k=1}^{l-1}\beta^{k-1}\varepsilon_{t-k,l}\right)^2 \left(\sum_{k=1}^{l-1}\beta^{k-1}h_{t-k,l}\right)^2\right]$$
$$+ 3cE\left[\left|\sum_{k=1}^{\infty}\beta^{k-1}\varepsilon_{t-k} - \sum_{k=1}^{l-1}\beta^{k-1}\varepsilon_{t-k,l}\right|^2 \left|\sum_{k=1}^{l-1}\beta^k h_{t-k,l}\right|^2\right]\Big/ h_t^4$$
$$\equiv 3cEL_{1t}^2 + 3cEL_{2t}^2 + 3cEL_{3t}^2, \tag{4.12}$$

where c is some constant. In Appendix A, we prove that there exists a constant $\rho \in (0,1)$ such that

$$EL_{1t}^2 = O(\rho^l), \qquad EL_{2t}^2 = O(\rho^l), \qquad \text{and} \quad EL_{3t}^2 = O(\rho^l). \tag{4.13}$$

By (4.12) and (4.13), we have that

$$\sum_{l=1}^{\infty} E|(g_t^* - Eg_t^*) - (g_{t,l}^* - Eg_{t,l}^*)|^2]^{1/2} < \infty.$$

Similar to (b), we can complete the remaining proof of case (c). This completes the proof. ∎

LEMMA 4.5. *(Ling and Li, 1998, Theorem 3.1). Let* $\{S_n(\tau), 0 \le \tau \le 1\}$ *and* $\{\xi_k, k = 1,2,\ldots\}$ *be two sequences of random processes such that*

(a) $S_n(\tau) \xrightarrow{\mathcal{L}} S(\tau)$ *in D;*

(b) $\dfrac{1}{\sqrt{n}}\sum_{k=1}^{[n\tau]}\xi_k \xrightarrow{\mathcal{L}} \xi(\tau)$ *in D;*

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

(d) $\dfrac{1}{n}\sum_{t=1}^{n}|\xi_t|$ *is bounded in probability*

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

$$\sup_{0\le\tau\le1}\left|\frac{1}{n}\sum_{k=1}^{[n\tau]}S_n\left(\frac{k}{n}\right)\xi_k\right| \xrightarrow{p} 0, \quad \text{as } n \to \infty.$$

LEMMA 4.6. *Let g_t be defined as in Lemma 4.4. Then under Assumptions 1 and 2,*

$$\frac{1}{n^2}\sum_{t=1}^{n}(g_t - Eg_t)y_{t-1}^2 = o_p(1).$$

Proof. It is not difficult to verify that g_t's satisfy the conditions (c) and (d) in Lemma 4.5 (see Ling and Li, 1998, the proof of Theorem 3.4). By Lemma 4.2 and the continuity mapping theorem, $y_{[n\tau]}^2/n \xrightarrow{\mathcal{L}} w_1^2(\tau)$. By Lemmas 4.4 and 4.5, $\sum_{t=1}^{n}(g_t - Eg_t)y_{t-1}^2/n^2 = n^{-1}\sum_{t=1}^{n}[(y_{t-1}^2/n)(g_t - Eg_t)] = o_p(1)$. This completes the proof. ∎

The following lemma gives the asymptotic properties of the information matrix.

LEMMA 4.7. *Under Assumptions 1 and 2,*

$$\frac{1}{n^2}\sum_{t=1}^{n}\frac{\partial^2 l_t}{\partial\phi^2} \xrightarrow{\mathcal{L}} F\int_0^1 w_1^2(\tau)\,d\tau,$$

where F and $w_1(\tau)$ are defined as in Theorem 2.1.

Proof.

$$\begin{aligned}\frac{\partial^2 l_t}{\partial\phi^2} &= -\frac{y_{t-1}^2}{h_t} - \frac{2\alpha^2\varepsilon_t^2}{h_t^3}\left(\sum_{i=1}^{t-1}\beta^{i-1}y_{t-i-1}\varepsilon_{t-i}\right)^2 \\ &\quad + \left(\frac{\varepsilon_t^2}{h_t} - 1\right)\frac{\partial}{\partial\phi}\left(\frac{1}{2h_t}\frac{\partial h_t}{\partial\phi}\right) + \frac{2\alpha\varepsilon_t y_{t-1}}{h_t^2}\left(\sum_{i=1}^{t-1}\beta^{i-1}y_{t-i-1}\varepsilon_{t-i}\right) \\ &\equiv I_{1t} - 2\alpha^2 I_{2t} + I_{3t} + 2\alpha I_{4t}. \end{aligned} \tag{4.14}$$

Note that

$$y_{t-i-1} = \sum_{k=1}^{t-i-1}\varepsilon_k = \sum_{k=1}^{t-1}\varepsilon_k - \sum_{k=t-i}^{t-1}\varepsilon_k = y_{t-1} - \sum_{r=1}^{i}\varepsilon_{t-r}. \tag{4.15}$$

Denote $r_{t,i} = \sum_{r=1}^{i}\varepsilon_{t-r}$. Then $Er_{t,i}^2 = i$. We first study I_{2t} in (4.14). By (4.15),

$$\begin{aligned} I_{2t} &= \frac{\varepsilon_t^2}{h_t^3}\left(\sum_{i=1}^{t-1}\beta^{i-1}\varepsilon_{t-i}\right)^2 y_{t-1}^2 + \frac{\varepsilon_t^2}{h_t^3}\left(\sum_{i=1}^{t-1}\beta^{i-1}r_{t,i}\,\varepsilon_{t-i}\right)^2 \\ &\quad - \frac{2\varepsilon_t^2}{h_t^3}\left(\sum_{i=1}^{t-1}\beta^{i-1}\varepsilon_{t-i}\right)\left(\sum_{i=1}^{t-1}\beta^{i-1}r_{t,i}\,\varepsilon_{t-i}\right)y_{t-1} \\ &\equiv T_{1t} + T_{2t} - 2T_{3t}. \end{aligned} \tag{4.16}$$

By Lemma 4.3(1)(b),

$$E|T_{3t}| \le \frac{1}{\alpha_0} E\left[\left(\sum_{i=1}^{t-1} \beta^{i-1} \frac{|\varepsilon_{t-i}|}{\sqrt{h_t}}\right)\left(\sum_{i=1}^{t-1} \beta^{i-1} \frac{|\varepsilon_{t-i}|}{\sqrt{h_t}} |r_{t,i}|\right)|y_{t-1}|\right]$$
$$= O\left(E\left[\left(\sum_{i=1}^{t-1} \beta^{(i-1)/2} |r_{t,i}|\right)|y_{t-1}|\right]\right)$$
$$= O\left(\sum_{i=1}^{t-1} \beta^{(i-1)/2} \sqrt{E|r_{t,i}|^2} \sqrt{Ey_{t-1}^2}\right) = O(\sqrt{n}) = o(n), \quad \textbf{(4.17)}$$

where $o(\cdot)$ holds uniformly in t. Similarly, we can show that the following equation holds uniformly in t:

$$E|T_{2t}| = o(n). \quad \textbf{(4.18)}$$

By (4.16) and (4.18), we know that

$$\frac{1}{n^2} \sum_{t=1}^{n} I_{2t} = \frac{1}{n^2} \sum_{t=1}^{n} \frac{\varepsilon_t^2}{h_t^3} \left(\sum_{i=1}^{t-1} \beta^{i-1} \varepsilon_{t-i}\right)^2 y_{t-1}^2 + o_p(1). \quad \textbf{(4.19)}$$

Next, we show that

$$\frac{1}{n^2} \sum_{t=1}^{n} I_{3t} = o_p(1) \quad \text{and} \quad \frac{1}{n^2} \sum_{t=1}^{n} I_{4t} = o_p(1). \quad \textbf{(4.20)}$$

Because their proofs are similar, we prove only the latter.

$$I_{4t} = \frac{\varepsilon_t y_{t-1}}{h_t^2} \left(\sum_{i=1}^{t-1} \beta^{i-1} y_{t-i-1} \varepsilon_{t-i}\right)$$
$$= \frac{\varepsilon_t}{h_t^2} \left(\sum_{i=1}^{t-1} \beta^{i-1} \varepsilon_{t-i}\right) y_{t-1}^2 - \frac{\varepsilon_t}{h_t^2} \left(\sum_{i=1}^{t-1} \beta^{i-1} r_{t,i} \varepsilon_{t-i}\right) y_{t-1}$$
$$\equiv P_{1t} + P_{2t}. \quad \textbf{(4.21)}$$

Similar to Lemma 4.2, we can show that

$$S_{[n\tau]} = \frac{1}{\sqrt{n}} \sum_{t=1}^{[n\tau]} \frac{\varepsilon_t}{h_t^2} \left(\sum_{i=1}^{t-1} \beta^{i-1} \varepsilon_{t-i}\right) \xrightarrow{\mathcal{L}} W(\tau) \quad \text{in } D, \quad \textbf{(4.22)}$$

where $W(\tau)$ is a Brownian motion. Similar to the proof of Lemma 4.6, by (4.22) and Lemma 4.5, it is easy to show that

$$\frac{1}{n^2} \sum_{t=1}^{n} P_{1t} = o_p(1). \quad \textbf{(4.23)}$$

Similar to (4.17), by Lemma 4.3(1)(b),

$$\frac{1}{n^2} \sum_{t=1}^{n} E|P_{2t}| \le \frac{1}{n^2} \sum_{t=1}^{n} E\left[\frac{|\varepsilon_t|}{h_t^{3/2}} \sum_{i=1}^{t-1} \beta^{i-1} \frac{|\varepsilon_{t-i}|}{\sqrt{h_t}} |r_{t,i}||y_{t-1}|\right] = o(1). \quad \textbf{(4.24)}$$

By (4.21), (4.23), and (4.24), it follows that

$$\frac{1}{n^2}\sum_{t=1}^{n} I_{4t} = o_p(1). \tag{4.25}$$

By (4.14), (4.19), and (4.20), we have

$$\begin{aligned}
\frac{1}{n^2}\sum_{t=1}^{n}\frac{\partial^2 l_t}{\partial\phi^2} &= -\frac{1}{n^2}\sum_{t=1}^{n}\left[\frac{y_{t-1}^2}{h_t} + \frac{2\alpha^2\varepsilon_t^2}{h_t^3}\left(\sum_{i=1}^{t-1}\beta^{i-1}\varepsilon_{t-i}\right)^2 y_{t-1}^2\right] + o_p(1)\\
&= -\frac{1}{n^2}\sum_{t=1}^{n} y_{t-1}^2\left[E\left(\frac{1}{h_t}\right) + 2\alpha^2\sum_{i=1}^{t-1}\beta^{2(i-1)}E\left(\frac{\varepsilon_{t-i}^2}{h_t^2}\right)\right]\\
&\quad - \frac{1}{n^2}\sum_{t=1}^{n} y_{t-1}^2\left[\frac{1}{h_t} - E\left(\frac{1}{h_t}\right)\right]\\
&\quad - \frac{2\alpha^2}{n^2}\sum_{t=1}^{n} y_{t-1}^2\left[\frac{\varepsilon_t^2}{h_t^3}\left(\sum_{i=1}^{t-1}\beta^{i-1}\varepsilon_{t-i}\right)^2\right.\\
&\qquad\qquad\left. - \sum_{i=1}^{t-1}\beta^{2(i-1)}E\left(\frac{\varepsilon_{t-i}^2}{h_t^2}\right)\right] + o_p(1),\\
&= -\frac{1}{n^2}\sum_{t=1}^{n} y_{t-1}^2\left[E\left(\frac{1}{h_t}\right) + 2\alpha^2\sum_{i=1}^{t-1}\beta^{2(i-1)}E\left(\frac{\varepsilon_{t-i}^2}{h_t^2}\right)\right] + o_p(1),
\end{aligned} \tag{4.26}$$

where the last equation holds by Lemma 4.6. Furthermore, note that

$$\begin{aligned}
&\frac{1}{n^2}\sum_{t=1}^{n} y_{t-1}^2\left[\sum_{i=1}^{t-1}\beta^{2(i-1)}E\left(\frac{\varepsilon_{t-i}^2}{h_t^2}\right)\right]\\
&\quad = \frac{1}{n^2}\sum_{t=1}^{n} y_{t-1}^2\left[\sum_{i=1}^{\infty}\beta^{2(i-1)}E\left(\frac{\varepsilon_{t-i}^2}{h_t^2}\right)\right] - \frac{1}{n^2}\sum_{t=1}^{n} y_{t-1}^2\left[\sum_{i=t}^{\infty}\beta^{2(i-1)}E\left(\frac{\varepsilon_{t-i}^2}{h_t^2}\right)\right]\\
&\quad = \left(\frac{1}{n^2}\sum_{t=1}^{n} y_{t-1}^2\right)\left[\sum_{i=1}^{\infty}\beta^{2(i-1)}E\left(\frac{\varepsilon_{t-i}^2}{h_t^2}\right)\right] + o_p(1),
\end{aligned}$$

where the last equation holds because $y_{t-1}^2/n = O_p(1)$ and $\varepsilon_{t-i}^2/h_t = O(\beta^i)$ a.s. Thus we have

$$\frac{1}{n^2}\sum_{t=1}^{n}\frac{\partial^2 l_t}{\partial\phi^2} = \left(\frac{1}{n^2}\sum_{t=1}^{n} y_{t-1}^2\right)F + o_p(1) \xrightarrow{\mathcal{L}} F\int_0^1 w_1^2(\tau)\,d\tau$$

by Lemma 4.2 and the continuity mapping theorem. This completes the proof. ■

LEMMA 4.8. *If Assumptions 1 and 2 hold, then*

$$\frac{1}{n}\sum_{t=1}^{n}\frac{\partial l_t}{\partial \phi} \xrightarrow{\mathcal{L}} \int_0^1 w_1(\tau)\,dw_2(\tau),$$

where $(w_1(\tau), w_2(\tau))$ *is defined as in Theorem 2.1.*

Proof.

$$\begin{aligned}\frac{\partial l_t}{\partial \phi} &= \frac{y_{t-1}\varepsilon_t}{h_t} - \frac{\alpha}{h_t}\left(\frac{\varepsilon_t^2}{h_t}-1\right)\left(\sum_{i=1}^{t-1}\beta^{i-1}y_{t-i-1}\varepsilon_{t-i}\right)\\ &= \left[\frac{\varepsilon_t}{h_t} - \frac{\alpha}{h_t}\left(\frac{\varepsilon_t^2}{h_t}-1\right)\left(\sum_{i=1}^{t-1}\beta^{i-1}\varepsilon_{t-i}\right)\right]y_{t-1}\\ &\quad + \frac{\alpha}{h_t}\left(\frac{\varepsilon_t^2}{h_t}-1\right)\left(\sum_{i=1}^{t-1}\beta^{i-1}r_{t,i}\,\varepsilon_{t-i}\right), \end{aligned} \tag{4.27}$$

where $r_{t,i} = \sum_{r=1}^{i}\varepsilon_{t-r}$. Denote the last term in (4.27) as R_t. Similar to (4.24), we can show that $E[n^{-1}\sum_{t=1}^{n}R_t]^2 = o_p(1)$ and thus $n^{-1}\sum_{t=1}^{n}R_t = o_p(1)$. Furthermore, by Lemma 4.2 and applying Theorem 2.2 in Kurtz and Protter (1991), we know that the conclusion holds. ■

Proof of Theorem 2.1. By Lemma 4.2 and the continuity mapping theorem, all limiting distributions in Lemmas 4.7 and 4.8 are jointly convergent. By (2.5) and Lemmas 4.7 and 4.8, we complete the proof. ■

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., R.F. Engle, & D.B. Nelson (1994) ARCH model. In R.F. Engle and D. C. McFadden (eds.), *Handbook of Econometrics*, vol. 4, pp. 2959–3038. Amsterdam: Elsevier Science.

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

Davis, R.A., T. Mikosch, & B. Basrak (1999) The Sample ACF of Multivariate Stochastic Recurrence Equations with Applications to GARCH. Technical Report, Laboratory of Actuarial Mathematics, University of Copenhagen.

Hall, P. & C.C. Heyde (1980) *Martingale Limit Theory and Its Applications*. San Diego: Academic Press.

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

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

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

Ling, S. & W.K. Li (1997) On fractionally integrated autoregressive moving-average time series models with conditional heteroscedasticity. *Journal of the American Statistical Association* 92, 1184–1194.

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

Ling, S., W.K. Li, & M. McAleer (2001) Estimation and testing for unit root processes with GARCH (1,1) errors: Theory and Monte Carlo study. *Econometric Reviews* (special issue), forthcoming.

Ling, S. and McAleer, M. (2003) On adaptive estimation in nonstationary ARMA models with GARCH errors. *Annals of Statistics* 31, forthcoming.

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* 64, 575–596.

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

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.

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

APPENDIX A: PROOFS OF (4.8) AND (4.13)

Proof of (4.8). By the definition of $\varepsilon_{t-k,l}$ in Lemma 4.3, $|\varepsilon_{t-k,l}| \le |\varepsilon_{t-k}|$ a.s. Furthermore, by Lemma 4.3(1)(b), we have

$$
\begin{aligned}
I_1 &= \left| \frac{1}{h_t^2}\left(\sum_{k=1}^{l-1}\beta^{k-1}\varepsilon_{t-k}\right)^2 - \frac{1}{h_t^2}\left(\sum_{k=1}^{l-1}\beta^{k-1}\varepsilon_{t-k,l}\right)^2 \right| \\
&= \frac{1}{h_t^2}\left|\sum_{k=1}^{l-1}\beta^{k-1}(\varepsilon_{t-k}-\varepsilon_{t-k,l})\right|\left|\sum_{k=1}^{l-1}\beta^{k-1}(\varepsilon_{t-k}+\varepsilon_{t-k,l})\right| \\
&\le \left|\sum_{k=1}^{l-1}\beta^{k-1}(\varepsilon_{t-k}-\varepsilon_{t-k,l})\right|^{1/2}\left[\sum_{k=1}^{l-1}\beta^{k-1}(|\varepsilon_{t-k}|+|\varepsilon_{t-k,l}|)\right]^{3/2}\Big/ h_t^2 \\
&\le \left|\sum_{k=1}^{l-1}\beta^{k-1}(\varepsilon_{t-k}-\varepsilon_{t-k,l})\right|^{1/2}\left[\sum_{k=1}^{l-1}\beta^{k-1}(|\varepsilon_{t-k}|\Big/\sqrt{h_t})\right]^{3/2}\cdot 2^{3/2}\alpha_0^{-1/2} \\
&= \left|\sum_{k=1}^{l-1}\beta^{k-1}(\varepsilon_{t-k}-\varepsilon_{t-k,l})\right|^{1/2} O\left(\left(\sum_{k=1}^{l-1}\beta^{k/2}\right)^{3/2}\right) \\
&= \left|\sum_{k=1}^{l-1}\beta^{k-1}(\varepsilon_{t-k}-\varepsilon_{t-k,l})\right|^{1/2}\cdot O(1).
\end{aligned}
$$

Thus, by Lemma 4.3(2)(c),

$$EI_1^2 = O\left(\sum_{k=1}^{l-1} \beta^{k-1} E|\varepsilon_{t-k} - \varepsilon_{t-k,l}|\right)$$

$$= O\left(\sum_{k=1}^{l-1} \beta^{k-1} E|\sqrt{h_t} - \sqrt{h_{t-k,l}}|\right) = O\left(\sum_{k=1}^{l-1} \beta^{k-1} E|h_t - h_{t-k,l}|\right)$$

$$= O\left(\sum_{k=1}^{l-1} \beta^{k-1} \rho^{l-k+1}\right) = O\left(\sum_{k=1}^{l-1} \rho_1^l\right) = O(l\rho_1^l) = O(\rho_2^l), \quad \textbf{(A.1)}$$

where $0 < \rho < 1$, $\rho_1 = \max\{\beta, \rho\}$, and $0 < \rho_2 < 1$ such that $l\rho_1^l < \rho_2^l$ for some large enough l.

By Lemma 4.3(2)(a),

$$I_2 = \left(\frac{1}{h_t^2} - \frac{1}{h_{t,l}^2}\right)\left(\sum_{k=1}^{l-1} \beta^{k-1} \varepsilon_{t-k,l}\right)^2 = \frac{|h_t - h_{t,l}||h_t + h_{t,l}|}{h_t^2 h_{t,l}^2}\left(\sum_{k=1}^{l-1} \beta^{k-1} \varepsilon_{t-k,l}\right)^2$$

$$\leq \frac{|h_t - h_{t,l}|^{1/2}|h_t + h_{t,l}|^{3/2}}{h_t^2 h_{t,l}}\left(\sum_{k=1}^{l-1} \beta^{k-1} \frac{|\varepsilon_{t-k,l}|}{\sqrt{h_{t,l}}}\right)^2$$

$$= O\left(|h_t - h_{t,l}|^{1/2}\left(\sum_{k=1}^{l-1} \beta^{k-1-(k-1)/2}\right)^2\right) = O(|h_t - h_{t,l}|^{1/2}).$$

Thus, by Lemma 4.3(2)(c),

$$EI_2^2 = O(E|h_t - h_{t,l}|) = O(\rho^l), \quad \textbf{(A.2)}$$

where $0 < \rho < 1$. By Lemma 4.3(1)(b),

$$I_3 = \frac{1}{h_t^2}\left|\sum_{k=l}^{\infty} \beta^{k-1} \varepsilon_{t-k}\right|\left|\sum_{k=1}^{\infty} \beta^{k-1} \varepsilon_{t-k} + \sum_{k=1}^{l-1} \beta^{k-1} \varepsilon_{t-k}\right|$$

$$\leq 2h_t^{-3/2}\left|\sum_{k=l}^{\infty} \beta^{k-1} \varepsilon_{t-k}\right|\left[\sum_{k=1}^{\infty} \beta^{k-1} \frac{|\varepsilon_{t-k}|}{\sqrt{h_t}}\right]$$

$$\leq 2w^{-3/2}\left|\sum_{k=l}^{\infty} \beta^{k-1} \varepsilon_{t-k}\right|\left[\sum_{k=1}^{\infty} O(\beta^{k-1-(k-1)/2})\right]$$

$$= \left|\sum_{k=l}^{\infty} \beta^{k-1} \varepsilon_{t-k}\right| O(1),$$

where $O(\cdot)$ holds uniformly in all t. Hence

$$EI_3^2 = O\left(\sum_{k=l}^{\infty} \beta^{2(k-1)} E\varepsilon_{t-k}^2\right) = O(\rho^l), \quad \textbf{(A.3)}$$

where $0 < \rho < 1$. By (A.1)–(A.3), we complete the proof of (4.8). ■

Proof of (4.13). Note that, by Lemma 4.3(1)(b), $\sum_{k=1}^{\infty}\beta^{k-1}|\varepsilon_{t-k}|/\sqrt{h_t}=O(1)$ a.s. Therefore

$$EL_{1t}^2 \le 2cE\left[\frac{1}{h_t^3}\left(\sum_{k=1}^{l-1}\beta^{k-1}|h_{t-k}-h_{t-k,l}|\right)^2\right]+2cE\left[\frac{1}{h_t^3}\left(\sum_{k=l}^{\infty}\beta^{k-1}h_{t-k}\right)^2\right]$$
$$\equiv 2cEI_1^2+2cEI_2^2, \tag{A.4}$$

where c is a constant. By Lemma 4.3(1)(c), $h_{t-k}/h_t<\beta^{-k}$ a.s., and hence

$$I_1^2 \le \frac{l}{h_t^3}\sum_{k=1}^{l-1}\beta^{2(k-1)}|h_{t-k}-h_{t-k,l}|^2$$
$$\le 2\alpha_0^{-2}l\sum_{k=1}^{l-1}\beta^{2(k-1)}|h_{t-k}-h_{t-k,l}|\,\frac{h_{t-k}}{h_t}\le 2\alpha_0^{-2}l\sum_{k=1}^{l-1}\beta^{k-2}|h_{t-k}-h_{t-k,l}|.$$

Thus, by Lemma 4.3(2)(c),

$$EI_1^2 \le 2\alpha_0^{-2}l\sum_{k=1}^{l-1}\beta^{k-2}\rho^{l-k}=O(l^2\rho_1^l)=O(\rho_2^l), \tag{A.5}$$

where $0\le\rho<1$, $\rho_1=\max\{\beta,\rho\}$, and $0<\rho_2<1$ such that $l^2\rho_1^l<\rho_2^l$ for some large enough l. By Lemma 4.3(2)(b) and 4.3(2)(c), the Cauchy–Schwarz inequality applying to h_{t-k}/h_t, and $E[\beta/(\alpha\eta_t^2+\beta)]^2<1$, we have

$$EI_2^2 \le cE\left[\frac{1}{h_t^2}\left(\sum_{k=l}^{\infty}\beta^{k-1}h_{t-k}\right)^2\right]$$
$$\le c\left\{\sum_{k=l}^{\infty}\beta^{k-1}\left[E\left(\frac{h_{t-k}}{h_t}\right)^2\right]^{1/2}\right\}^2\le c\left\{\sum_{k=l}^{\infty}\beta^{k-1}\left[E\left(\frac{1}{\alpha\eta_{t-1}^2+\beta}\right)^2\right]^{k/2}\right\}^2$$
$$\le c\left\{\beta^{-1}\sum_{k=l}^{\infty}\left[E\left(\frac{\beta}{\alpha\eta_{t-1}^2+\beta}\right)^2\right]^{k/2}\right\}^2=O(\rho^l), \tag{A.6}$$

where c is a constant. By (A.4)–(A.6), $EL_{1t}^2=O(\rho^l)$ with $0<\rho<1$. Now, we consider L_{2t} in (4.13).

$$L_{2t}^2=\left|\frac{1}{h_t^2}-\frac{1}{h_{t,l}^2}\right|^2\left(\sum_{k=1}^{l-1}\beta^{k-1}\varepsilon_{t-k,l}\right)^2\left(\sum_{k=1}^{l-1}\beta^{k-1}h_{t-k,l}\right)^2$$
$$=\frac{|h_t^2-h_{t,l}^2|^2}{h_t^4h_{t,l}^4}\left(\sum_{k=1}^{l-1}\beta^{k-1}\varepsilon_{t-k,l}\right)^2\left(\sum_{k=1}^{l-1}\beta^{k-1}h_{t-k,l}\right)^2$$
$$\le c|h_t-h_{t,l}|^{1/2}\left(\sum_{k=1}^{l-1}\beta^{k-1}|\varepsilon_{t-k,l}|/\sqrt{h_{t,l}}\right)^2\left(\sum_{k=1}^{l-1}\beta^{k-1}\frac{h_{t-k,l}}{h_{t,l}}\right)^2, \tag{A.7}$$

where c is a constant. By Lemma 4.3(2)(a) and 4.3(2)(b), the Cauchy–Schwarz inequality, and $E[\beta/(\alpha\eta_t^2+\beta)]^4 < 1$,

$$
\begin{aligned}
EL_{2t}^2 &\le cE|h_t - h_{t,l}|E\left(\sum_{k=1}^{l-1}\beta^{k-1}\frac{h_{t-k,l}}{h_{t,l}}\right)^4 \\
&\le cE|h_t - h_{t,l}|\left\{\sum_{k=1}^{l-1}\beta^{k-1}\left[E\left(\frac{h_{t-k,l}}{h_{t,l}}\right)^4\right]^{1/4}\right\}^4 \\
&\le O(\rho^l)\left\{\sum_{k=1}^{l-1}\beta^{k-1}\left[E\left(\frac{1}{\alpha\eta_{t-k}+\beta}\right)^4\right]^{(k-1)/4}\right\}^4 \\
&\le O(\rho^l)\left\{\sum_{k=1}^{l-1}\left[E\left(\frac{\beta}{\alpha\eta_{t-k}+\beta}\right)^4\right]^{(k-1)/4}\right\}^4 = O(\rho^l), \qquad \textbf{(A.8)}
\end{aligned}
$$

where the second inequality follows by the Minkowski inequality. Similarly we can show that $EL_{3t}^2 = O(\rho^l)$, where $0 < \rho < 1$. This completes the proof of (4.13). ■

APPENDIX B: PROOF OF (3.5)

Proof.

$$
\begin{aligned}
\frac{\partial^2 l_t}{\partial\phi\partial\mu} &= -\frac{y_{t-1}}{h_t} - \frac{2\alpha^2\varepsilon_t^2}{h_t^3}\left(\sum_{i=1}^{t-1}\beta^{i-1}\varepsilon_{t-i}\right)\left(\sum_{i=1}^{t-1}\beta^{i-1}y_{t-i-1}\varepsilon_{t-i}\right) \\
&\quad - \frac{2\alpha^2}{h_t^2}\left(\frac{\varepsilon_t^2}{h_t}-1\right)\left(\sum_{i=1}^{t-1}\beta^{i-1}\varepsilon_{t-i}\right)\left(\sum_{i=1}^{t-1}\beta^{i-1}y_{t-i-1}\varepsilon_{t-i}\right) \\
&\quad + \frac{1}{h_t}\left(\frac{\varepsilon_t^2}{h_t}-1\right)\left(\sum_{i=1}^{t-1}\beta^{i-1}y_{t-i-1}\right) + \frac{2\alpha\varepsilon_t}{h_t^2}\left(\sum_{i=1}^{t-1}\beta^{i-1}y_{t-i-1}\varepsilon_{t-i}\right) \\
&\equiv J_{1t} - 2\alpha^2 J_{2t} + J_{3t} + J_{4t} + 2\alpha J_{5t}. \qquad \textbf{(B.1)}
\end{aligned}
$$

By (4.15),

$$
\begin{aligned}
J_{2t} &= \frac{\varepsilon_t^2}{h_t^3}\left(\sum_{i=1}^{t-1}\beta^{i-1}\varepsilon_{t-i}\right)^2 y_{t-1} + \frac{\varepsilon_t^2}{h_t^3}\left(\sum_{i=1}^{t-1}\beta^{i-1}\varepsilon_{t-i}\right)\left(\sum_{i=1}^{t-1}\beta^{i-1}r_{t,i}\varepsilon_{t-i}\right) \\
&\equiv K_{1t} - K_{2t}. \qquad \textbf{(B.2)}
\end{aligned}
$$

By Lemma 4.3(1)(b),

$$
\begin{aligned}
E|K_{2t}| &\le \frac{1}{\alpha_0}E\left[\left(\sum_{i=1}^{t-1}\beta^{i-1}\frac{|\varepsilon_{t-i}|}{\sqrt{h_t}}\right)\left(\sum_{i=1}^{t-1}\beta^{i-1}\frac{|\varepsilon_{t-i}|}{\sqrt{h_t}}|r_{t,i}|\right)\right] \\
&= O\left(E\left[\left(\sum_{i=1}^{t-1}\beta^{(i-1)/2}|r_{t,i}|\right)\right]\right) = O\left(\sum_{i=1}^{t-1}\beta^{(i-1)/2}\sqrt{E|r_{t,i}|^2}\right) = O(1), \qquad \textbf{(B.3)}
\end{aligned}
$$

where $O(\cdot)$ holds uniformly in t.

$$\begin{aligned}\frac{1}{n^3}E\left(\sum_{t=1}^{n}J_{5t}\right)^2 &= \frac{1}{n^3}\sum_{t=1}^{n}E\left[\frac{1}{h_t^3}\left(\sum_{i=1}^{t-1}\beta^{i-1}y_{t-i-1}\varepsilon_{t-i}\right)^2\right]\\ &\le \frac{1}{n^3\alpha_0^2}\sum_{t=1}^{n}E\left[\left(\sum_{i=1}^{t-1}\beta^{i-1}|y_{t-i-1}|\frac{|\varepsilon_{t-i}|}{\sqrt{h_t}}\right)^2\right)\\ &\le \frac{1}{n^3\alpha_0^2}\sum_{t=1}^{n}\left(\sum_{i=1}^{t-1}\beta^{(i-1)/2}\sqrt{Ey_{t-i-1}^2}\right)^2 = \frac{1}{n^3}\sum_{t=1}^{n}O(n)=o(1).\end{aligned} \quad \textbf{(B.4)}$$

Similarly, we can show that

$$\frac{1}{n^3}E\left(\sum_{t=1}^{n}J_{3t}\right)^2 = o(1) \quad \text{and} \quad \frac{1}{n^3}E\left(\sum_{t=1}^{n}J_{4t}\right)^2 = o(1). \quad \textbf{(B.5)}$$

By (B.1)–(B.5), we have that

$$\begin{aligned}\frac{1}{n^{3/2}}\sum_{t=1}^{n}\frac{\partial^2 l_t}{\partial\phi\partial\mu} &= -\frac{1}{n^{3/2}}\sum_{t=1}^{n}\left[\frac{y_{t-1}}{h_t}+\frac{2\alpha^2\varepsilon_t^2}{h_t^3}\left(\sum_{i=1}^{t-1}\beta^{i-1}\varepsilon_{t-i}\right)^2 y_{t-1}\right]+o_p(1)\\ &= -\frac{1}{n^{3/2}}\sum_{t=1}^{n}y_{t-1}\left[E\left(\frac{1}{h_t}\right)+2\alpha^2\sum_{i=1}^{t-1}\beta^{2(i-1)}E\left(\frac{\varepsilon_{t-i}^2}{h_t^2}\right)\right]\\ &\quad -\frac{1}{n}\sum_{t=1}^{n}\frac{y_{t-1}}{\sqrt{n}}\left[\frac{1}{h_t}-E\left(\frac{1}{h_t}\right)\right]\\ &\quad -\frac{2\alpha^2}{n}\sum_{t=1}^{n}\frac{y_{t-1}}{\sqrt{n}}\left[\frac{\varepsilon_t^2}{h_t^3}\left(\sum_{i=1}^{t-1}\beta^{i-1}\varepsilon_{t-i}\right)^2-\sum_{i=1}^{t-1}\beta^{2(i-1)}E\left(\frac{\varepsilon_{t-i}^2}{h_t^2}\right)\right]+o_p(1),\\ &= -\frac{1}{n}\sum_{t=1}^{n}\frac{y_{t-1}}{\sqrt{n}}\left[E\left(\frac{1}{h_t}\right)+2\alpha^2\sum_{i=1}^{t-1}\beta^{2(i-1)}E\left(\frac{\varepsilon_{t-i}^2}{h_t^2}\right)\right]+o_p(1),\end{aligned} \quad \textbf{(B.6)}$$

where the last equation holds by Lemmas 4.2, 4.4, and 4.5. Furthermore, note that

$$\begin{aligned}&\frac{1}{n}\sum_{t=1}^{n}\frac{y_{t-1}}{\sqrt{n}}\left[\sum_{i=1}^{t-1}\beta^{2(i-1)}E\left(\frac{\varepsilon_{t-i}^2}{h_t^2}\right)\right]\\ &\quad = \frac{1}{n}\sum_{t=1}^{n}\frac{y_{t-1}}{\sqrt{n}}\left[\sum_{i=1}^{\infty}\beta^{2(i-1)}E\left(\frac{\varepsilon_{t-i}^2}{h_t^2}\right)\right]-\frac{1}{n}\sum_{t=1}^{n}\frac{y_{t-1}}{\sqrt{n}}\left[\sum_{i=t}^{\infty}\beta^{2(i-1)}E\left(\frac{\varepsilon_{t-i}^2}{h_t^2}\right)\right]\\ &\quad = \left(\frac{1}{n}\sum_{t=1}^{n}\frac{y_{t-1}}{\sqrt{n}}\right)\left[\sum_{i=1}^{\infty}\beta^{2(i-1)}E\left(\frac{\varepsilon_{t-i}^2}{h_t^2}\right)\right]+o_p(1),\end{aligned}$$

where the last equation holds because $y_{t-1}/\sqrt{n} = O_p(1)$ and $\varepsilon_{t-i}^2/h_t = O(\beta^i)$ a.s. Thus, we have that

$$\frac{1}{n^{3/2}}\sum_{t=1}^{n}\frac{\partial^2 l_t}{\partial\phi\partial\mu} = \left(\frac{1}{n^{3/2}}\sum_{t=1}^{n}y_{t-1}\right)F+o_p(1)\xrightarrow{\mathcal{L}}F\int_0^1 w_1(\tau)\,d\tau$$

by Lemma 4.2 and the continuity mapping theorem. This completes the proof. ∎

APPENDIX C: ASYMPTOTIC EQUIVALENCE OF HESSIAN MATRICES

The likelihood function based on ε_t can be written as

$$L_1(\tilde{\delta}) = \frac{1}{n}\sum_{t=1}^{n} l_t(\tilde{\delta}) \quad \text{and} \quad l_t(\tilde{\delta}) = -\frac{1}{2}\ln h_t(\tilde{\delta}) - \frac{\varepsilon_t^2}{2h_t(\tilde{\delta})},$$

where $h_t(\tilde{\delta}) = \tilde{\alpha}_0 + \tilde{\alpha}\varepsilon_{t-1}^2 + \tilde{\beta}h_{t-1}(\tilde{\delta}) = \sum_{i=0}^{t-1}\tilde{\beta}^i[\tilde{\alpha}_0 + \tilde{\alpha}\varepsilon_{t-1}^2]$ with $h_0(\tilde{\delta})$ = some constant. The following theorem shows that the Hessian matrices of $L(\phi,\tilde{\delta})$ and $L_1(\tilde{\delta})$ are asymptotically equivalent on the parameter space $\tilde{\Theta}_n \equiv \{(\phi,\tilde{\delta}) : n|\phi - 1| \le M$ and $\tilde{\delta} \in \Theta\}$ for any fixed $M > 0$.

THEOREM C. *Under Assumption 1, it follows that*

$$\sup_{(\phi,\tilde{\delta})\in\tilde{\Theta}_n} \frac{1}{n}\sum_{t=1}^{n}\left\| \frac{\partial^2 l_t(\phi,\tilde{\delta})}{\partial\tilde{\delta}\partial\tilde{\delta}'} - \frac{\partial^2 l_t(\tilde{\delta})}{\partial\tilde{\delta}\partial\tilde{\delta}'}\right\| = o_p(1).$$

Proof. For simplicity, we only consider the case with $h_0(\phi,\tilde{\delta}) = 0$ and $h_0(\tilde{\delta}) = 0$. It is easy to modify the proof for the case with $h_0(\phi,\tilde{\delta}) \neq 0$ or $h_0(\tilde{\delta}) \neq 0$.

$$\frac{\partial^2 l_t(\phi,\tilde{\delta})}{\partial\tilde{\delta}\partial\tilde{\delta}'} = -\frac{\varepsilon_t^2(\phi)}{2h_t^3(\phi,\tilde{\delta})}\frac{\partial h_t(\phi,\tilde{\delta})}{\partial\tilde{\delta}}\frac{\partial h_t(\phi,\tilde{\delta})}{\partial\tilde{\delta}'} + \frac{1}{2h_t(\phi,\tilde{\delta})}\left[\frac{\partial^2 h_t(\phi,\tilde{\delta})}{\partial\tilde{\delta}\partial\tilde{\delta}'} - \frac{1}{h_t(\phi,\tilde{\delta})}\frac{\partial h_t(\phi,\tilde{\delta})}{\partial\tilde{\delta}}\frac{\partial h_t(\phi,\tilde{\delta})}{\partial\tilde{\delta}'}\right]\left[1 - \frac{\varepsilon_t^2(\phi)}{h_t(\phi,\tilde{\delta})}\right]. \quad \textbf{(C.1)}$$

When $\phi = 1$, $\partial^2 l_t(\phi,\tilde{\delta})/\partial\tilde{\delta}\partial\tilde{\delta}' = \partial^2 l_t(\tilde{\delta})/\partial\tilde{\delta}\partial\tilde{\delta}'$. Because of the similarity, we only present the proof of the equation:

$$\sup_{(\phi,\tilde{\delta})\in\tilde{\Theta}_n} \frac{1}{n}\sum_{t=1}^{n}\left\| \frac{\varepsilon_t^2(\phi)}{h_t^3(\phi,\tilde{\delta})}\frac{\partial h_t(\phi,\tilde{\delta})}{\partial\tilde{\delta}}\frac{\partial h_t(\phi,\tilde{\delta})}{\partial\tilde{\delta}'} - \frac{\varepsilon_t^2}{h_t^3(\tilde{\delta})}\frac{\partial h_t(\tilde{\delta})}{\partial\tilde{\delta}}\frac{\partial h_t(\tilde{\delta})}{\partial\tilde{\delta}'}\right\| = o_p(1). \quad \textbf{(C.2)}$$

In the following discussion, all $O(\cdot)$, $O_p(\cdot)$, and $o_p(\cdot)$ hold uniformly in $(\phi,\tilde{\delta}) \in \tilde{\Theta}_n$ and $t = 1,\ldots,n$.

Denote $\varepsilon_t(\phi) = y_t - \phi y_{t-1}$. Because $\max_{1\le t\le n}|n^{-1/2}y_{t-1}| = O_p(1)$ under Assumption 1, it follows that $\varepsilon_t(\phi) = \varepsilon_t - [n(\phi-1)](n^{-1/2}y_{t-1})n^{-1/2} = \varepsilon_t + O_p(n^{-1/2})$. Thus,

$$\varepsilon_t^2(\phi) = \varepsilon_t^2 + O_p\left(\frac{1}{n}\right) + \varepsilon_t O_p\left(\frac{1}{\sqrt{n}}\right). \quad \textbf{(C.3)}$$

By (C.3), we can show that

$$h_t(\phi,\tilde{\delta}) = \sum_{i=0}^{t-1}\tilde{\beta}^i[\tilde{\alpha}_0 + \tilde{\alpha}\varepsilon_t^2(\phi)] = h_t(\tilde{\delta}) + h_t^{1/2}(\tilde{\delta})O_p\left(\frac{1}{\sqrt{n}}\right). \quad \textbf{(C.4)}$$

Note that $h_t(\phi,\tilde\delta) \ge \alpha_{0l}$ and $h_t(\tilde\delta) \ge \alpha_{0l}$. By (C.3) and (C.4), it follows that

$$\begin{aligned}\frac{\varepsilon_{t-i}^2(\phi)}{h_t(\phi,\tilde\delta)} - \frac{\varepsilon_{t-i}^2}{h_t(\tilde\delta)} &= \frac{\varepsilon_{t-i}^2(\phi)-\varepsilon_{t-i}^2}{h_t(\phi,\tilde\delta)} + \varepsilon_{t-i}^2\left[\frac{1}{h_t(\phi,\tilde\delta)} - \frac{1}{h_t(\tilde\delta)}\right] \\ &= \frac{O_p(n^{-1}) + O_p(n^{-1/2})\varepsilon_{t-i}}{h_t(\phi,\tilde\delta)} + \frac{\varepsilon_{t-i}^2 h_t^{1/2}(\tilde\delta) O_p(n^{-1/2})}{h_t(\phi,\tilde\delta)h_t(\tilde\delta)} \\ &= o_p(1) + o_p(1)\sqrt{\frac{\varepsilon_{t-i}^2}{h_t(\tilde\delta)}} = o_p(1) + o_p(1)\tilde\beta^{-i/2}, \end{aligned} \tag{C.5}$$

for $i = 0,1,\ldots$, where the next to last equation holds because $\max_{1\le t\le n}|\varepsilon_t|/\sqrt{n} = o_p(1)$. Again because $\max_{1\le t\le n}|\varepsilon_t|/\sqrt{n} = o_p(1)$, we can show that $\max_{1\le t\le n}\sup_{\tilde\delta\in\Theta} h_t^{1/2}(\tilde\delta)/\sqrt{n} = o_p(1)$. Using the fact that $h_t(\tilde\delta) \ge \tilde\alpha_0 + \tilde\beta^i h_{t-i}(\tilde\delta)$ and a similar method as for (C.5), we can obtain

$$\frac{h_{t-i}(\phi,\tilde\delta)}{h_t(\phi,\tilde\delta)} - \frac{h_{t-i}(\tilde\delta)}{h_t(\tilde\delta)} = o_p(1) + o_p(1)\sqrt{\frac{h_{t-i}(\tilde\delta)}{h_t(\tilde\delta)}} = o_p(1) + o_p(1)\tilde\beta^{-i/2}. \tag{C.6}$$

By (C.5), it is straightforward to show that

$$\frac{1}{h_t(\phi,\tilde\delta)}\frac{\partial h_t(\phi,\tilde\delta)}{\partial\tilde\alpha} - \frac{1}{h_t(\tilde\delta)}\frac{\partial h_t(\tilde\delta)}{\partial\tilde\alpha} = \sum_{i=1}^{t-1}\tilde\beta^i\left[\frac{\varepsilon_{t-i}^2(\phi)}{h_t(\phi,\tilde\delta)} - \frac{\varepsilon_{t-i}^2}{h_t(\tilde\delta)}\right] = o_p(1). \tag{C.7}$$

Similarly, by (C.6), we can obtain

$$\frac{1}{h_t(\phi,\tilde\delta)}\frac{\partial h_t(\phi,\tilde\delta)}{\partial\tilde\alpha_0} - \frac{1}{h_t(\tilde\delta)}\frac{\partial h_t(\tilde\delta)}{\partial\tilde\alpha_0} = o_p(1), \tag{C.8}$$

$$\frac{1}{h_t(\phi,\tilde\delta)}\frac{\partial h_t(\phi,\tilde\delta)}{\partial\tilde\beta} - \frac{1}{h_t(\tilde\delta)}\frac{\partial h_t(\tilde\delta)}{\partial\tilde\beta} = o_p(1). \tag{C.9}$$

By (C.7)–(C.9), it is easy to show that

$$\left\|\frac{1}{h_t(\phi,\tilde\delta)}\frac{\partial h_t(\phi,\tilde\delta)}{\partial\tilde\delta} - \frac{1}{h_t(\tilde\delta)}\frac{\partial h_t(\tilde\delta)}{\partial\tilde\delta}\right\| = o_p(1). \tag{C.10}$$

Because $\alpha_0 \ge \alpha_{0l}$ and $\tilde\beta^i\varepsilon_{t-i}^2(\phi)/[1+\tilde\beta^i\varepsilon_{t-i}^2(\phi)] \le 1$ uniformly in $\tilde\Theta_n$ and any t and i, we have

$$\begin{aligned}\frac{1}{h_t(\phi,\tilde\delta)}\frac{\partial h_t(\phi,\tilde\delta)}{\partial\tilde\alpha} &= \sum_{i=0}^{t-1}\frac{\tilde\beta^i\varepsilon_{t-i}^2(\phi)}{h_t(\phi,\tilde\delta)} \le O(1)\sum_{i=0}^{t-1}\frac{\tilde\beta^i\varepsilon_{t-i}^2(\phi)}{1+\tilde\beta^i\varepsilon_{t-i}^2(\phi)} \\ &\le O_p(1)\sum_{i=0}^{t-1}\tilde\beta^{i\nu}|\varepsilon_{t-i}|^{2\nu}(\phi) = o_p(1) + O_p(1)\sum_{i=0}^{t-1}\beta_u^{i\nu}|\varepsilon_{t-i}|^{2\nu}, \end{aligned} \tag{C.11}$$

where the last equation holds by (C.3). Similarly, it follows that

$$\frac{1}{h_t(\phi,\tilde\delta)}\frac{\partial h_t(\phi,\tilde\delta)}{\partial\tilde\beta} = o_p(1) + O_p(1)\sum_{i=0}^{t-1}\beta_u^{i\nu}|h_{t-i}(\tilde\delta)|^{\nu}. \tag{C.12}$$

Note that $h_t^{-1}(\phi,\tilde{\delta})[\partial h_t(\phi,\tilde{\delta})/\partial\tilde{\alpha}_0] = O_p(1)$. By (C.11) and (C.12), we have

$$\left\| \frac{1}{h_t(\phi,\tilde{\delta})} \frac{\partial h_t(\phi,\tilde{\delta})}{\partial\tilde{\delta}} \right\| = O_p(1) + O_p(1)\sum_{i=0}^{t-1}\beta_u^{i\nu}[h_{t-i}^{\nu}(\tilde{\delta}) + |\varepsilon_{t-i}|^{2\nu}]. \tag{C.13}$$

Taking ν_1 small enough such that $\beta_l^{\nu_1} > \beta$, we have

$$\begin{aligned}\frac{h_t}{h_t(\tilde{\delta})} &= O_p(1) + O_p(1)\sum_{i=1}^{t-1}\frac{\beta^i\varepsilon_{t-i}^2}{h_t(\tilde{\delta})} + O_p(1)\sum_{i=t}^{\infty}\beta^i\varepsilon_{t-i}^2 \\ &\le O_p(1) + O_p(1)\sum_{i=1}^{t-1}\frac{\beta^i\varepsilon_{t-i}^2}{(1+\beta_l^i\varepsilon_t^2)^{\nu_1}} + O_p(1)\rho^t\sum_{i=0}^{\infty}\beta^i\varepsilon_{-i}^2 \\ &\le O_p(1) + O_p(1)\sum_{i=1}^{t-1}\left(\frac{\beta}{\beta_l^{\nu_1}}\right)^i\varepsilon_{t-i}^{2-2\nu_1} + O_p(1)\rho^t\sum_{i=0}^{\infty}\beta^i\varepsilon_{-i}^2.\end{aligned} \tag{C.14}$$

Under Assumption 1, $E\varepsilon_t^2 < \infty$, and hence we can show that $E\max_{1\le t\le n}\sup_{\tilde{\delta}\in\Theta} \times h_t(\tilde{\delta}) < \infty$. Note that $\varepsilon_t^2 = \eta_t^2 h_t$. By (C.13) and (C.14), we can show that

$$\begin{aligned}&\frac{1}{n}\sum_{t=1}^{n}\frac{\varepsilon_t^2}{h_t(\tilde{\delta})}\left\| \frac{1}{h_t(\phi,\tilde{\delta})}\frac{\partial h_t(\phi,\tilde{\delta})}{\partial\tilde{\delta}} \right\| \\ &\quad = O_p(1) + O_p(1)\left(\sum_{i=0}^{\infty}\beta^i\varepsilon_{t-i}^2\right)\frac{1}{n}\sum_{t=1}^{n}\rho^t\left\{\sum_{i=0}^{t-1}\beta_u^{i\nu}[h_{t-i}^{\nu}(\tilde{\delta}) + |\varepsilon_{t-i}|^{2\nu}]\right\} \\ &\qquad + \frac{O_p(1)}{n}\sum_{t=1}^{n}\left\{\sum_{i=0}^{t-1}\beta_u^{i\nu}[h_{t-i}^{\nu}(\tilde{\delta}) + |\varepsilon_{t-i}|^{2\nu}]\sum_{i=1}^{t-1}\left(\frac{\beta}{\beta_l^{\nu_1}}\right)^i\varepsilon_{t-i}^{2-\nu_1}\right\} = O_p(1),\end{aligned} \tag{C.15}$$

where $2\nu < \nu_1$. By (C.10) and (C.15), we have

$$\frac{1}{n}\sum_{t=1}^{n}\frac{\varepsilon_t^2}{h_t(\tilde{\delta})}\left[\frac{1}{h_t(\phi,\tilde{\delta})}\frac{\partial h_t(\phi,\tilde{\delta})}{\partial\tilde{\delta}}\right]\left\| \frac{1}{h_t(\phi,\tilde{\delta})}\frac{\partial h_t(\phi,\tilde{\delta})}{\partial\tilde{\delta}} - \frac{1}{h_t(\tilde{\delta})}\frac{\partial h_t(\tilde{\delta})}{\partial\tilde{\delta}} \right\| = o_p(1), \tag{C.16}$$

$$\frac{1}{n}\sum_{t=1}^{n}\frac{\varepsilon_t^2}{h_t(\tilde{\delta})}\left[\frac{1}{h_t(\tilde{\delta})}\frac{\partial h_t(\tilde{\delta})}{\partial\tilde{\delta}}\right]\left\| \frac{1}{h_t(\phi,\tilde{\delta})}\frac{\partial h_t(\phi,\tilde{\delta})}{\partial\tilde{\delta}} - \frac{1}{h_t(\tilde{\delta})}\frac{\partial h_t(\tilde{\delta})}{\partial\tilde{\delta}} \right\| = o_p(1). \tag{C.17}$$

By (C.5) and (C.13), it follows that

$$\frac{1}{n}\sum_{t=1}^{n}\left|\frac{\varepsilon_t^2(\phi)}{h_t(\phi,\tilde{\delta})} - \frac{\varepsilon_t^2}{h_t(\tilde{\delta})}\right|\left\| \frac{1}{h_t(\phi,\tilde{\delta})}\frac{\partial h_t(\phi,\tilde{\delta})}{\partial\tilde{\delta}} \right\|^2 = o_p(1). \tag{C.18}$$

Finally, by (C.16)–(C.18), we can show that (C.2) holds. This completes the proof. ∎