Appl. Math. J. Chinese Univ.
2022, 37(2): 246-257

Nearly nonstationary processes under infinite variance GARCH noises

ZHANG Rong-mao[1,2] LIU Qi-meng[2,*] SHI Jian-hua[1]

Abstract. Let Y_t be an autoregressive process with order one, i.e., $Y_t = \mu + \phi_n Y_{t-1} + \varepsilon_t$, where $\{\varepsilon_t\}$ is a heavy tailed general GARCH noise with tail index α. Let $\hat{\phi}_n$ be the least squares estimator (LSE) of ϕ_n. For $\mu = 0$ and $\alpha < 2$, it is shown by Zhang and Ling (2015) that $\hat{\phi}_n$ is inconsistent when Y_t is stationary (i.e., $\phi_n \equiv \phi < 1$), however, Chan and Zhang (2010) showed that $\hat{\phi}_n$ is still consistent with convergence rate n when Y_t is a unit-root process (i.e., $\phi_n = 1$) and $\{\varepsilon_t\}$ is a GARCH(1, 1) noise. There is a gap between the stationary and nonstationary cases. In this paper, two important issues will be considered: (1) what about the nearly unit root case? (2) When can ϕ be estimated consistently by the LSE? We show that when $\phi_n = 1 - c/n$, then $\widehat{\phi}_n$ converges to a functional of stable process with convergence rate n. Further, we show that if $\lim_{n\to\infty} k_n(1-\phi_n) = c$ for a positive constant c, then $k_n(\hat{\phi}_n - \phi)$ converges to a functional of two stable variables with tail index $\alpha/2$, which means that ϕ_n can be estimated consistently only when $k_n \to \infty$.

§1 Introduction

Consider the following models

$$Y_t = \mu + \phi_n Y_{t-1} + \varepsilon_t, \tag{1.1}$$

where $Y_0 = 0$ and ε_t follows a general first-order generalized autoregressive conditional heteroscedasticity model (GGARCH(1, 1)) defined by

$$\varepsilon_t = \sigma_t \eta_t, \quad \sigma_t \geq 0\,, \tag{1.2}$$

$$\sigma_t^\delta = g(\eta_{t-1}) + c(\eta_{t-1})\sigma_{t-1}^\delta, \tag{1.3}$$

where $\delta > 0$, $Pr\{\sigma_t^\delta > 0\} = 1$, $c(0) < 1$, $c(\cdot)$ and $g(\cdot)$ are non-negative functions, and $\{\eta_t\}$ is a sequence of i.i.d. symmetric white noises with unit variance. The general model (1.3) was

Received: 2021-04-30. Revised: 2021-07-31.
MR Subject Classification: 62M10, 62E20, 60F17.
Keywords: GARCH noises, heavy-tailed, stable processes and unit-root.
Digital Object Identifier(DOI): https://doi.org/10.1007/s11766-022-4442-5.
Zhang is supported by the National Natural Science Foundation of China(11771390, 12171427), ZPNSFC(LZ21A010002), Fundamental Research Funds for the Central Universities (2021XZZX002) and Shi is supported by Natural Science Foundation of Fujian Province(2020J01794), Fujian Key Laboratory of Granular Computing and Applications(Minnan Normal University).
*Corresponding author.

defined by He and Terasvirta (1999). It includes many models as special cases, for example, the GARCH(1,1) model of Bollerslev (1986), the absolute value GARCH (1,1) model of Taylor (1986) and Schwert (1989), the nonlinear GARCH(1,1) model of Engle (1990), the volatility switching GARCH (1,1) model of Fornari and Mele (1997), the threshold GARCH (1,1) model of Zakoian (1994), and the generalized quadratic ARCH (1,1) model of Sentana (1995).

There is an extensive literature on unit-root estimation and testing for the case $c(\cdot) \equiv 0$ and $g(\cdot) \equiv \sigma^\delta$, i.e., $\{\varepsilon_t\}$ are i.i.d. random variables. For a concise review on the recent developments on this topic, see Chan (2008) and the references therein.

On the other hand, the unit-root problem for the case of non i.i.d errors has also been receiving considerable attention in the literature. Under these circumstances, the original testing for unit-root in (1.1) is tantamount to testing for unit-root with GGARCH(1,1) errors. Motivated by this consideration, extensive research have been conducted. For example, Hall and Yao (2003) considered QMLE and Peng and Yao (2003) studied the least absolute deviations estimation (LAD) when $\mathrm{E}\varepsilon_t^2 < \infty$ and $\mathrm{E}\eta_t^4 = \infty$. Ling and Li (1998) considered the distribution of the maximum likelihood estimation for non-stationary autoregressive moving average time series with GARCH errors for the case $\mathrm{E}\varepsilon_t^4 < \infty$. Ling and Li (2003), Ling, Li and McAleer (2003) and Li and Li (2009) generalized the results to the case $\mathrm{E}\varepsilon_t^2 < \infty$ and obtained that the limit distribution of the estimated unit-root as a functional of the Brownian motion. Chan and Peng (2005) studied the least absolute deviations estimation for stationary AR(1) process with heavy-tailed ARCH(1) noise, see also Zhu and Ling (2015). Chan and Zhang (2010) studied the asymptotic distribution of Dickey-Fuller test for $\phi_n = 1$ under an infinite-variance GARCH(1, 1) noise with tail index α, they showed that with convergence rate n, the asymptotic distribution of the LSE converges to a functional of a stable process when $\alpha < 2$ and a functional of the Brownian motion when $\alpha = 2$. On the other hand, Zhang and Ling (2015) showed that the LSE of a stationary AR(p) model is inconsistent when $\alpha < 2$, see also Zhang and Chan (2021). This means a big gap exists between the stationary and nonstationary cases when the noise is an infinite variance GARCH noise.

To shed some intuitive insight into these phenomena, consider the following simple simulation exercise. Let Y_t be a unit-root model, i.e., $Y_t = Y_{t-1} + \varepsilon_t$ with $\varepsilon_t = \sqrt{\omega + \beta\varepsilon_{t-1}^2}\eta_t, \quad t = 1, 2, \ldots, n$, where $\{\eta_t\}$ is a sequence of i.i.d. standard normal noise. Note that the tail index of ε_t is given by the solution of

$$\mathrm{E}(\beta\eta_t^2)^{\alpha/2} = 1,$$

see Kesten (1973). Thus, if $\beta = 1$, then the tail index $\alpha = 2$; if $\beta = \pi/2$, then the tail index $\alpha = 1$. We simulate $Y_t = \phi Y_{t-1} + \varepsilon_t$ with various ϕ and ARCH(1) noise with $\omega = 0.4$ and $\beta = 0.5$ for the finite variance case, $\beta = 1$, 1.5 for the finite mean but infinite variance case (i.e., $1 < \alpha < 2$) and $\beta = 2$ for the infinite mean case (i.e., $\alpha < 1$). For each setting, we replicate the exercise 500 times and take $n = 500$, 1000, 1500. The empirical sampling bias (Bias, $\hat{\phi}_n - \phi$) and standard deviation (SD) for the corresponding estimates (Est) $\widehat{\phi}$ based on the 500 repetitions are reported in Table 1. It can be seen from this table that for all ϕ, as β increases, i.e., the tail index α decreases, the bias and SD for the autoregressive parameter ϕ tend to increase. When $\beta > 1$, i.e., the noise has infinite variance, the autoregressive parameter ϕ cannot be estimated well if $\phi < 1$, but it can still be estimated consistently when $\phi = 1$, and the nearer the ϕ closes to 1, the smaller the bias and SD are. One natural question is when

$\phi = 1 - \gamma/n$ for some constant γ, does an Ornstein-Uhlenbeck (O-U) limit distribution still hold? We will show that the limit distribution of $n(\hat{\phi}_n - \phi)$ converges to functional of fractional Ornstein-Uhlenbeck (O-U) stable processes.

The second question is when ϕ_n can be estimated consistently by $\hat{\phi}_n$. We will show that if $\phi_n = 1 - c/k_n$ for a positive constant c, then $\hat{\phi}_n - \phi_n = O(1/k_n)$, which implies ϕ_n can be estimated consistently only when $k_n \to \infty$. This also gives a smoothing transition from stationary process to nonstationary process similar to Phillips and Magdalinos (2007), who showed that the convergence rate is $\sqrt{nk_n}$ when the noise ε_t is a sequence of i.i.d. variables with finite variance.

Table 1. Bias and standard deviation for the LSE of ϕ.

	$\beta = 0.5$		$\beta = 1$		$\beta = 1.5$		$\beta = 2$	
	Bias	SD	Bias	SD	Bias	SD	Bias	SD
				$n = 500$				
-0.8	0.0062	0.0386	0.0213	0.0849	0.0600	0.1612	0.0730	0.1868
-0.6	0.0065	0.0527	0.0312	0.1280	0.0561	0.1952	0.0734	0.2413
0.6	-0.0114	0.0539	-0.0296	0.1276	-0.0624	0.2039	-0.0782	0.2429
0.8	-0.0101	0.0390	-0.0276	0.0921	-0.0643	0.1641	-0.0849	0.1962
0.95	-0.0093	0.0182	-0.0187	0.0409	-0.0400	0.0814	-0.0617	0.1195
0.99	-0.0091	0.0109	-0.0131	0.0261	-0.0213	0.0459	-0.0348	0.0837
1	-0.0110	0.0086	-0.0130	0.0211	-0.0200	0.0333	-0.0300	0.0714
				$n = 1000$				
-0.8	0.0021	0.0267	0.0134	0.0724	0.0638	0.1494	0.0761	0.1785
-0.6	0.0031	0.0441	0.0258	0.1179	0.0512	0.1898	0.0781	0.2462
0.6	-0.0027	0.0406	-0.0377	0.1176	-0.0526	0.1933	-0.0832	0.2270
0.8	-0.0052	0.0281	-0.0248	0.0867	-0.0569	0.1776	-0.0854	0.1816
0.95	-0.0051	0.0127	-0.0138	0.0352	-0.0381	0.0914	-0.0560	0.1317
0.99	-0.0045	0.0067	-0.0069	0.0114	-0.0147	0.0377	-0.0287	0.0760
1	-0.0055	0.0046	-0.0069	0.0143	-0.0158	0.0458	-0.0251	0.0801
				$n = 1500$				
-0.8	0.0022	0.0235	0.0274	0.0834	0.0582	0.1363	0.0860	0.1885
-0.6	0.0038	0.0358	0.0163	0.1060	0.0555	0.1828	0.0733	0.2234
0.6	-0.0033	0.0364	-0.0197	0.1085	-0.0636	0.1967	-0.0817	0.2425
0.8	-0.0028	0.0237	-0.0319	0.0918	-0.0530	0.1383	-0.0908	0.1916
0.95	-0.0034	0.0102	-0.0115	0.0335	-0.0437	0.0992	-0.0518	0.1223
0.99	-0.0028	0.0050	-0.0054	0.0145	-0.0125	0.0298	-0.0252	0.0600
1	-0.0034	0.0031	-0.0044	0.0064	-0.0081	0.0201	-0.0175	0.0606

Throughout the paper, $o(1)$ ($o_P(1)$) 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); when two sequences a_n and b_n are of the same order, we denote $a_n \sim b_n$; $\xrightarrow{P}$ and $\xrightarrow{\mathcal{L}}$ denote convergence in probability and in distribution, respectively. And C denotes a positive bounded constant taking different values at different places. The rest of the paper is organized as follows. The Dickey-Fuller test and asymptotic theory are developed in Section 2. Section 3 concludes. All the technical proofs are relegated to Section 4.

§2 Tests and Asymptotic Distribution

2.1 Dickey-Fuller Test

Given Y_0 and observations $Y_1, \ldots Y_n$, the least squares estimator (LSE) of ϕ for model (1.1) is given by

$$\hat{\phi}_n = \Big(\sum_{i=1}^{n}(Y_{i-1} - \overline{Y})^2\Big)^{-1}\Big(\sum_{i=1}^{n}(Y_{i-1} - \overline{Y})Y_i\Big), \tag{2.1}$$

where $\overline{Y} = \sum_{i=1}^{n} Y_{i-1}/n$. When $\phi_n = 1 - \gamma/n$ for some constant γ, the Dickey-Fuller (DF) test $\widehat{\phi}_n$ for model (1.1) is defined by

$$n(\widehat{\phi}_n - \phi_n) = \Big(\frac{1}{n}\sum_{i=1}^{n}(Y_{i-1} - \overline{Y})^2\Big)^{-1}\Big(\sum_{i=1}^{n}(Y_{i-1} - \overline{Y})\varepsilon_i\Big). \tag{2.2}$$

Throughout the paper, we impose the following conditions.

Condition 1.

(i) $\mathrm{E}\log(c(\eta_t)) < 0$.

(ii) There exists a $k_0 > 0$ such that $\mathrm{E}(c(\eta_t))^{k_0} \geq 1$, $\mathrm{E}[(c(\eta_t))^{k_0}\log^+(c(\eta_t))] < \infty$ and $\mathrm{E}(g(\eta_t) + |\eta_t|^{\delta})^{k_0} < \infty$, where $\log^+(x) = \max\{0, \log(x)\}$.

(iii) The density of η_1 is positive in a neighborhood of zero.

Condition 1(i) is a necessary and sufficient condition for the existence of a stationary solution of σ_t^2 (see Nelson (1990)). If Condition 1(ii) holds, then Condition 1(i) is equivalent to is equivalent to $\mathrm{E}(c(\eta_1))^{\mu} < 1$ for some $\mu > 0$ (see Remark 2.9 of Basrak, Davis and Mikosch (2002)). Conditions 1(i) and (iii) also imply that h_t is not a constant and hence exclude the i.i.d. case. By Lemma 2.1 of Zhang and Ling (2015), it follows that there exists a unique $\alpha \in (0, k_0]$ such that

$$\mathrm{E}(c(\eta_t))^{\alpha/\delta} = 1. \tag{2.3}$$

Further, if $E|\eta_1|^{\alpha} < \infty$, then as $x \to \infty$,

$$P(|\varepsilon_1| > x) \sim c_0^{(\alpha)}\mathrm{E}|\eta_1|^{\alpha}x^{-\alpha},$$

where

$$c_0^{(\alpha)} = \frac{E\left(\left[g(\eta_1) + c(\eta_1)\sigma_1^{\delta}\right]^{\alpha/\delta} - \left[c(\eta_1)\sigma_1^{\delta}\right]^{\alpha/\delta}\right)}{\alpha E\left(c(\eta_1)^{\alpha/\delta}\log^+(c(\eta_1))\right)}.$$

Condition 1(iii) can be weakened as the distribution of F of η_1 is a mixture of an absolutely continuous component with respect to the Lebesgue measure λ on $\mathbb{R}$ and Dirac masses at some points $\mu_i \in \mathbb{R}, i = 1, \ldots, N$. See Francq and Zakoïan (2006).

2.2 Asymptotic Distributions

We now derive the limit distributions of the LSE in (2.1). Our first result is about whether the DF test given in (2.2) has power when ε_t is a heavy tailed GARCH noise with index $\alpha < 2$.

Theorem 2.1. *Let α be the tail index defined in (2.3). Suppose that $\alpha < 2$ and Condition 1 holds.*

(i) When $\phi_n \equiv \phi \in (-1, 1)$,

$$\hat{\phi}_n - \phi \xrightarrow{\mathcal{L}} (S_{\alpha/2})^{-1} Z_{\alpha/2}, \tag{2.4}$$

where both $S_{\alpha/2}$ and $Z_{\alpha/2}$ are stable variables with index $\alpha/2$.

(ii) When $\lim_{n\to\infty} n(1-\phi_n) = \gamma$ for some constant γ and $\mu = 0$,

$$n(\hat{\phi}_n - \phi) \xrightarrow{\mathcal{L}} \frac{\int_0^1 Z_{\alpha,\gamma}(t)\, dZ_\alpha(t) - Z_\alpha(1) \int_0^1 Z_{\alpha,\gamma}(t)\, dt}{\int_0^1 Z^2_{\alpha,\gamma}(t)\, dt - (\int_0^1 Z_{\alpha,\gamma}(t)\, dt)^2}, \tag{2.5}$$

where $Z_\alpha(t)$ is a stable process with index α and $Z_{\alpha,\gamma}(t)$ is an O-U stable process given by

$$Z_{\alpha,\gamma}(t) = Z_\alpha(t) - \gamma \int_0^t e^{-\gamma(t-s)} Z_\alpha(s)\, ds, \quad Z_\gamma(0) = 0, \quad t \in [0, 1].$$

Remark 2.1. *It can be seen from the proof that the results also work for more general higher order GARCH(p, q) cases.*

Remark 2.2. *$\phi_n \equiv 1$ is a special case of Theorem 2.1(ii). When $\phi_n \equiv 1$, $\gamma = 0$, hence $Z_{\alpha,0}(t) = Z_\alpha(t)$.*

Remark 2.3. *Theorem 2.1 shows that the asymptotic behaviors for a stationary AR model derived by heavy-tailed GARCH noise are completely different from those derived by i.i.d. noise. In fact, when $\{\varepsilon_t\}$ is a sequence of i.i.d. random variables belonging to the attraction domain of a stable law with index $\alpha \in (0, 2)$ and $\phi_n \equiv \phi \in (-1, 1)$, then*

$$a_n(\widehat{\phi}_n - \phi) \xrightarrow{\mathcal{L}} (S_{\alpha/2})^{-1} Z_\alpha,$$

where $S_{\alpha/2}$ is a stable variables with index $\alpha/2$, and Z_α is a stable variables with index α. However, when $\lim_{n\to\infty} n(1-\phi_n) = \gamma$ for some constant γ and $\mu = 0$, the asymptotic distribution is the same as Theorem 2.1(ii).

From (2.4) and (2.5), we see that the asymptotic behavior of $\hat{\phi}_n$ is totally different between a stationary and a nearly nonstationary case. Note that $\hat{\phi}_n$ is not consistent when $\phi_n < 1$ and does not depend on n (i.e., Y_t is a stationary process), while super consistent with convergence rate n when $\phi_n = 1 - \gamma/n$ for a certain constant γ. An interesting question is when ϕ can be consistently estimated? Does there exist a smoothing transition from a stationary to a nonstationary case? To address this issue, we consider a moderate deviation from unity model as in Phillips and Magdalinos (2007), i.e., $Y_t = \mu + \phi_n Y_{t-1} + \varepsilon_t$, with $\phi_n = 1 - c/k_n$, $c > 0$. The next theorem is about the limit distribution of $\hat{\phi}_n$ under such setting.

Theorem 2.2. *Suppose that $\mu = 0$, $\alpha < 2$, and Condition 1 holds. If there exists a $k_n = o(n)$ such that $\lim_{n\to\infty} k_n(1-\phi_n) = c > 0$, then*

$$k_n(\hat{\phi}_n - \phi) \xrightarrow{\mathcal{L}} (S^*_{\alpha/2})^{-1} Z^*_{\alpha/2}, \tag{2.6}$$

where both $S^*_{\alpha/2}$ and $Z^*_{\alpha/2}$ are stable variables with index $\alpha/2$.

Remark 2.4. *From Theorem 2.2, we have that when* $\phi_n = 1 - c/n$, *i.e.,* $k_n = n$, *then the convergence rate is n; when* $\phi_n = 1 - c/k_n \equiv \phi \in (-1, 1)$, *i.e.,* $k_n \equiv d$ *for some* $d > c/2$, *then* Y_t *is stationary and the convergence rate reduces to* $O(1)$, *which is consistent with Therem 2.1. Thus, Theorem 2.2 also gives a smoothing transition from stationary to nonstationary cases and show that* $\hat{\phi}_n$ *estimates* ϕ_n *consistently only when* $k_n \to \infty$.

§3 Conclusions

In this paper, we discuss the limit behaviors of the Dickey-Fuller statistic for a unit-root model with noises driven by heavy-tailed GARCH innovations. It is shown that when the tail index $\alpha < 2$ of the GARCH innovations, the autoregressive parameter ϕ cannot be consistently estimated by the LSE. However, for such GARCH noise, when $\lim_{n\to\infty} n(1 - \phi_n) = \gamma \in \mathcal{R}$, the LSE $\hat{\phi}_n$ is still a super consistent estimator and converges to a functional of O-U stable processes. Further, we also develop an asymptotic theory of the LSE for an AR(1) process with coefficient $\phi = 1 - c/k_n, c > 0$, which gives a smoothing transition from stationary to nonstationary cases, explains why their convergence rates are so different, and shows that the LSE is consistent only when $\phi_n \to 1$, i.e., $k_n \to \infty$. The results of this paper can be easily extended to higher order heavy-tailed GARCH-type processes, like GARCH(p, q). Further, using the same argument as in Chan and Zhang (2009), it is east to extend the results to the case with nonzero μ. This paper also opens several interesting questions. First, if a robust procedure instead of LSE is used, could one detect the unit-root more efficiently? Note that Knight (1989) (see also Phillips (1991)) showed that L_1 estimation has significant gains in this framework for the infinite variance case. In view of this fact, one possible way to handle the inconsistency and efficient testing issue is to adopt the L_1 estimate. Second, since the limit distribution of the DF test is complicated, their critical values are difficult to derive, how to construct a new test to avoid deriving the critical values? These issues will be explored in a future work.

§4 Technical Proofs

In this section, we prove the main results. For any given integers l and H, we define a $(H + 1)$-dimensional random vector:

$$\mathbf{Z}_{t,l,H} = (\varepsilon_{t-l}^2, \varepsilon_{t-l}\varepsilon_{t-l-1}, \cdots, \varepsilon_{t-l}\varepsilon_{t-l-H}). \tag{4.1}$$

And denote $a_n = \left(c_0^{(\alpha)} E|\eta_1|^\alpha n\right)^{1/\alpha}$.

Lemma 4.1. *Suppose that Condition 1 holds and* $\alpha < 2$. *Then, for any positive integer* K *and* H, *as* $n \to \infty$,

$$\sum_{t=1}^{[ns]} \varepsilon_t / a_n \xrightarrow{\mathcal{S}} Z_\alpha(s), \tag{4.2}$$

$$\frac{1}{a_n^2}\Big(\sum_{t=1}^{n} \mathbf{Z}_{t,0,H}, \sum_{t=1}^{n} \boldsymbol{Z}_{t,1,H}, \ldots, \sum_{t=1}^{n} \mathbf{Z}_{t,K,H}\Big) \xrightarrow{\mathcal{L}} (\boldsymbol{S}_{\alpha/2}, \boldsymbol{S}_{\alpha/2}, \ldots, \boldsymbol{S}_{\alpha/2})_{1\times(K+1)}, \tag{4.3}$$

where $\xrightarrow{\mathcal{S}}$ *denotes weak convergence under S-topology in* $D[0, 1]$, $Z_{\alpha/2}(s)$ *is a stable process with index* α, *and* $\boldsymbol{S}_{\alpha/2}$ *is a* $H + 1$*-dimensional stable random vector with index* $\alpha/2$.

Conclusion (4.3) can be found in Lemma 3.1 of Zhang and Ling (2015) and (4.2) can be shown similarly to Theorem 2.2 of Chan and Zhang (2010), here we omit the details.

Proof of Theorem 2.1. Note that when $\phi < 1$,

$$Y_i = \frac{\mu}{1-\phi} + \sum_{j=0}^{\infty} \phi^j \varepsilon_{i-j} =: \frac{\mu}{1-\phi} + \xi_i.$$

This implies that

$$\begin{aligned} \hat{\phi}_n - \phi &= \Big(\sum_{i=1}^{n} (Y_{i-1} - \overline{Y})^2\Big)^{-1} \Big(\sum_{i=1}^{n} (Y_{i-1} - \overline{Y})\varepsilon_i\Big) \\ &= \Big(\sum_{i=1}^{n} \xi_{i-1}^2 - n\overline{\xi}^2\Big)^{-1} \Big(\sum_{i=1}^{n} \xi_{i-1}\varepsilon_i - \overline{\xi} \sum_{i=1}^{n} \varepsilon_i\Big) \end{aligned} \tag{4.4}$$

By (4.3), Zhang and Ling (2015) showed that there exist two stable variables $S_{\alpha/2}$ and $Z_{\alpha/2}$ with tail index $\alpha/2$ such that

$$\frac{1}{a_n^2}\Big(\sum_{i=1}^{n} \xi_{i-1}^2, \sum_{i=1}^{n} \xi_{i-1}\varepsilon_i\Big) \xrightarrow{\mathcal{L}} (S_{\alpha/2}, Z_{\alpha/2}). \tag{4.5}$$

On the other hand, by (4.2) and a similar argument of Zhang, Sin and Ling (2015), we have

$$\frac{1}{a_n} \sum_{i=1}^{n} \xi_i \xrightarrow{\mathcal{L}} Z_\alpha(1)/(1-\phi),$$

which implies that

$$\frac{n\overline{\xi}^2}{a_n^2} = O_p(1/n) \quad \text{and} \quad \frac{\overline{\xi}}{a_n^2} \sum_{i=1}^{n} \varepsilon_i = O_p(1/n).$$

Thus, by (4.4) and (4.5), it follows that

$$\hat{\phi}_n - \phi = \Big(\sum_{i=1}^{n} \xi_{i-1}^2\Big)^{-1} \sum_{i=1}^{n} \xi_{i-1}\varepsilon_i + o_p(1) \xrightarrow{\mathcal{L}} S_{\alpha/2}^{-1} Z_{\alpha/2}, \tag{4.6}$$

this gives (2.4) as desired.

Next, we show (2.5). Let $S_{[nt]} = \sum_{i=1}^{[nt]} \varepsilon_i$. Then $S_0 = Y_0 = 0$ and

$$\begin{aligned} Y_{[nt]} &= \sum_{i=1}^{[nt]} \phi_n^{[nt]-i}(S_i - S_{i-1}) \\ &= S_{[nt]} - \sum_{i=1}^{[nt]} (\phi_n^{[nt]-i} - \phi_n^{[nt]-(i-1)}) S_{i-1} \\ &= S_{[nt]} - \frac{1}{n} \sum_{i=1}^{[nt]} \phi_n^{[nt]-i}(1-\phi_n) n S_{i-1} \\ &= S_{[nt]} - \frac{\gamma}{n} \sum_{i=1}^{[nt]} e^{-\gamma(t-i/n)} S_{i-1} + \frac{1}{n} \sum_{i=1}^{[nt]} (\gamma e^{-\gamma(t-i/n)} - \phi_n^{[nt]-i}(1-\phi_n)n) S_{i-1} \\ &= S_{[nt]} - \gamma \int_0^t e^{-\gamma(t-s)} S_{[ns]} d\, s + R_n(t). \end{aligned} \tag{4.7}$$

Since $\gamma e^{-\gamma(t-i/n)} - \phi_n^{[nt]-i}(1-\phi_n)n = o(1)$, it follows from (4.7), Lemma 4.1 and the continuous mapping that

$$Y_{[nt]}/a_n \xrightarrow{\mathcal{S}} Z_{\alpha,\gamma}(t). \tag{4.8}$$

Thus, by (2.2), Lemma 4.1 and the continuous mapping, we have (2.5) and complete the proof of Theorem 2.1. □

Lemma 4.2. *Under conditions of Theorem 2.2, we have*

$$\frac{c}{k_n a_n}\sum_{i=1}^{n} Y_i \xrightarrow{\mathcal{S}} Z_\alpha(1) \tag{4.9}$$

and

$$\frac{2c}{k_n a_n^2}\sum_{i=1}^{n} Y_i^2 \xrightarrow{\mathcal{S}} \boldsymbol{S}_{\alpha/2,0} + 2\sum_{i=1}^{\infty}\boldsymbol{S}_{\alpha/2,i}, \tag{4.10}$$

where $Z_\alpha(1)$ is a stable variable with tail index α given in (4.2), and $\boldsymbol{S}_{\alpha/2,i-1}$ denotes the i-th components of $\boldsymbol{S}_{\alpha/2}$ defined in (4.3).

Proof. Since $Y_t = \phi_n Y_{t-1} + \varepsilon_t$, it follows that

$$Y_i = \phi_n^i Y_0 + \sum_{j=1}^{i}\phi_n^{i-j}\varepsilon_j. \tag{4.11}$$

Thus,

$$\begin{aligned}
\sum_{i=1}^{n} Y_i &= \sum_{i=1}^{n}\phi_n^i Y_0 + \sum_{i=1}^{n}\sum_{j=1}^{i}\phi_n^{i-j}\varepsilon_j \\
&= \frac{\phi_n(1-\phi_n^n)}{1-\phi_n}Y_0 + \sum_{j=1}^{n}\sum_{i=j}^{n}\phi_n^{i-j}\varepsilon_j \\
&= c^{-1}k_n(1+o(1))Y_0 + \frac{1}{1-\phi_n}\sum_{j=1}^{n}(1-\phi_n^{n-j})\varepsilon_j \\
&= c^{-1}k_n(1+o(1))Y_0 + \frac{1}{1-\phi_n}\sum_{j=1}^{n-[v_n]}(1-\phi_n^{n-j})\varepsilon_j \\
&\quad + \frac{1}{1-\phi_n}\sum_{j=n-[v_n]+1}^{n}(1-\phi_n^{n-j})\varepsilon_j \\
&=: S_{n1} + S_{n2} + S_{n3},
\end{aligned}$$

where v_n is a constant sequence satisfying $v_n/k_n \to \infty$ and $v_n/n \to 0$. Since Y_0 is a given random variable, it follows that

$$\frac{c}{k_n a_n}S_{n1} = \frac{(1+o(1))Y_0}{a_n} = o_p(1). \tag{4.12}$$

By Lemma 4.1 and $(1-c/k_n)^j \to 0$ for all $j > v_n$, we have

$$\frac{c}{k_n a_n}S_{n2} = \frac{(1+o(1))}{a_n}\sum_{j=1}^{n-[v_n]}\varepsilon_j \xrightarrow{\mathcal{L}} Z_\alpha(1). \tag{4.13}$$

For S_{n3}, we write $\varepsilon_j = \varepsilon_j I(|\sigma_j| > a_n) + \varepsilon_j I(|\sigma_j| \le a_n) =: \varepsilon_{j,1} + \varepsilon_{j,2}$. Note that $\{\varepsilon_{j,2}/a_n\}$ is a martingale difference sequence and $\mathrm{E}\eta_t^2 = 1$. By Karamata's theorem, it follows that

$$\mathrm{E}\Big(\frac{1}{a_n}\sum_{j=n-[v_n]+1}^{n}(1-\phi_n^{n-j})\varepsilon_{j,2}\Big)^2 = \frac{v_n}{a_n^2}(1-\phi_n^{n-j})^2\mathrm{E}(\varepsilon_{j,2}^2) = O(v_n/n) = o(1).$$

This implies that

$$\frac{1}{a_n}\sum_{j=n-[v_n]+1}^{n}(1-\phi_n^{n-j})\varepsilon_{j,2} = o_p(1).$$

As a result, we have

$$\frac{c}{k_n a_n}\Big(\frac{1}{1-\phi_n}\Big)\sum_{j=n-[v_n]+1}^{n}(1-\phi_n^{n-j})\varepsilon_{j,2} = o_p(1).$$

By Karamata's theorem again, we have for any $0 < p < \min\{1, \alpha\}$,

$$\mathrm{E}\Big|\frac{1}{a_n}\sum_{j=n-[v_n]+1}^{n}(1-\phi_n^{n-j})\varepsilon_{j,1}\Big|^p = \frac{v_n}{a_n^p}(1-\phi_n^{n-j})^p\mathrm{E}(|\varepsilon_{j,1}|^p) = O(v_n/n) = o(1),$$

which implies that

$$\frac{c}{k_n a_n}\Big(\frac{1}{1-\phi_n}\Big)\sum_{j=n-[v_n]+1}^{n}(1-\phi_n^{n-j})\varepsilon_{j,1} = o_p(1).$$

Thus,

$$\frac{c}{k_n a_n}S_{n3} = o_p(1).$$

By (4.12), (4.13), we have (4.9).

Next, we show (4.10). Without loss of generality, we set $Y_0 = 0$. By (4.11), we have

$$\begin{aligned}\sum_{i=1}^{n}Y_i^2 &= \sum_{i=1}^{n}\Big(\sum_{j=1}^{i}\phi_n^{i-j}\varepsilon_j\Big)^2\\ &= \sum_{i=1}^{n}\sum_{j=1}^{i}\phi_n^{2(i-j)}\varepsilon_j^2 + 2\sum_{i=1}^{n}\sum_{j=1}^{i}\sum_{k=1}^{i-j}\phi_n^{2i-k-2j}\varepsilon_j\varepsilon_{j+k}\\ &= \sum_{j=1}^{n}\sum_{i=j}^{n}\phi_n^{2(i-j)}\varepsilon_j^2 + 2\sum_{j=1}^{n}\sum_{k=1}^{n-j}\sum_{i=j+k}^{n}\phi_n^{2i-k-2j}\varepsilon_j\varepsilon_{j+k}\\ &= \frac{1}{1-\phi_n^2}\sum_{j=1}^{n}(1-\phi_n^{2(n-j+1)})\varepsilon_j^2 + \frac{2}{1-\phi_n^2}\sum_{j=1}^{n}\sum_{k=1}^{n-j}\phi_n^k(1-\phi_n^{2(n-j-k+1)})\varepsilon_j\varepsilon_{j+k}\\ &=: L_{n1} + L_{n2}.\end{aligned}\tag{4.14}$$

Denote $X_{j1} = \varepsilon_j^2 I(|\varepsilon_j^2| > a_n)$ and $X_{j2} = \varepsilon_j^2 I(|\varepsilon_j^2| \le a_n)$. Then, by Karamata's theorem, we have that for any $\nu < 2/\alpha < 1$,

$$\begin{aligned}\mathrm{E}\Big|\frac{1}{a_n^2}\sum_{j=1}^{n}\phi_n^{2(n-j+1)}X_{j1}\Big|^{\nu} &\le \frac{1}{a_n^{2\nu}}\sum_{j=1}^{n}\phi_n^{2\nu(n-j+1)}\mathrm{E}|X_{j1}|^{\nu}\\ &\le \frac{C}{n}\sum_{j=1}^{n}\phi_n^{2\nu(n-j+1)} = o(1).\end{aligned}\tag{4.15}$$

Similarly, by Karamata's theorem again, we have

$$\mathrm{E}\Big|\frac{1}{a_n^2}\sum_{j=1}^n \phi_n^{2(n-j+1)}X_{j2}\Big| \le \frac{1}{a_n^2}\sum_{j=1}^n \phi_n^{2(n-j+1)}\mathrm{E}|X_{j2}| \le \frac{C}{n}\sum_{j=1}^n \phi_n^{2(n-j+1)} = o(1). \tag{4.16}$$

Thus, by (4.15), (4.16), and Lemma 4.1,

$$\begin{aligned}\frac{2cL_{n1}}{k_n a_n^2} &= \frac{2c}{k_n(1-\phi_n^2)a_n^2}\sum_{j=1}^n \varepsilon_j^2 - \frac{2c}{k_n(1-\phi_n^2)a_n^2}\Big(\sum_{j=1}^n \phi_n^{2(n-j+1)}X_{j1} + \sum_{j=1}^n \phi_n^{2(n-j+1)}X_{j2}\Big)\\ &= \frac{2c}{k_n(1-\phi_n^2)}\Big(\frac{1}{a_n^2}\sum_{j=1}^n \varepsilon_j^2\Big) + o_p(1) \xrightarrow{\mathcal{L}} \boldsymbol{S}_{\alpha/2,0}.\end{aligned} \tag{4.17}$$

One the other hand, by Lemma 3.1 of Zhang and Ling (2015), there exist a $0<\rho<1$ such that for all j,k,

$$P(|\varepsilon_j\varepsilon_{j+k}|>x) \le C(x^{-\alpha} + \rho^k x^{-\alpha/2}). \tag{4.18}$$

By (4.18) and the same arguments as in proving L_{n1}, we have that

$$\begin{aligned}\frac{2cL_{n2}}{k_n a_n^2} &= \frac{4c}{k_n(1-\phi_n^2)a_n^2}\sum_{j=1}^n\sum_{k=1}^H \varepsilon_j\varepsilon_{j+k} + o_p(1)\\ &\xrightarrow{\mathcal{L}} 2\sum_{k=1}^H \boldsymbol{S}_{\alpha/2,k} = 2\sum_{k=1}^\infty \boldsymbol{S}_{\alpha/2,k} + o_p(1),\end{aligned} \tag{4.19}$$

by letting $H\to\infty$.

By (4.17), (4.19), and their jointly convergence (by Lemma 4.1), we have (4.10) and complete the proof of Lemma 4.2. □

Proof of Theorem 2.2. Note that when $\mu=0$, $Y_t=\phi_n Y_{t-1}+\varepsilon_t$, implying that

$$\begin{aligned}\sum_{t=1}^n Y_{t-1}\varepsilon_t &= \frac{1}{2\phi_n}\sum_{t=1}^n (Y_t^2 - \phi_n^2 Y_{t-1}^2 - \varepsilon_t^2)\\ &= \frac{1}{2\phi_n}\Big[(1-\phi_n^2)\sum_{t=1}^n Y_{t-1}^2 + Y_n^2 - Y_0^2 - \sum_{t=1}^n \varepsilon_t^2\Big].\end{aligned} \tag{4.20}$$

Further, by (4.11), we have

$$\begin{aligned}\frac{Y_n^2}{a_n^2} &= \frac{1}{a_n^2}\Big(\phi_n^n Y_0 + \sum_{j=1}^n \phi_n^{n-j}\varepsilon_j\Big)^2\\ &\le \frac{2\phi_n^{2n}Y_0^2}{a_n^2} + 2\Big(\frac{1}{a_n}\sum_{j=1}^n \phi_n^{n-j}\varepsilon_j\Big)^2 = o_p(1).\end{aligned}$$

Thus, by (4.17), (4.20), and Lemma 4.2,

$$\begin{aligned}& k_n(\hat{\phi}_n - \phi_n)\\ =\ & \Big(\frac{1}{k_n a_n^2}\sum_{i=1}^n (Y_{i-1}-\overline{Y})^2\Big)^{-1}\Big(\frac{1}{a_n^2}\sum_{i=1}^n (Y_{i-1}-\overline{Y})\varepsilon_i\Big)\\ =\ & \Big[\frac{1}{k_n a_n^2}\Big(\sum_{i=1}^n Y_{i-1}^2 - n\overline{Y}^2\Big)\Big]^{-1}\Big(\frac{1}{a_n^2}\sum_{i=1}^n Y_{i-1}\varepsilon_i - \frac{1}{a_n^2}\overline{Y}\sum_{i=1}^n \varepsilon_i\Big)\end{aligned}$$

$$\begin{aligned}
&= \Big(\frac{1}{k_n a_n^2}\sum_{i=1}^n Y_{i-1}^2\Big)^{-1}\Big(\frac{1}{a_n^2}\sum_{i=1}^n Y_{i-1}\varepsilon_i\Big)+o_p(1)\\
&= \Big(\frac{2}{k_n a_n^2}\sum_{i=1}^n Y_{i-1}^2\Big)^{-1}\Big[\frac{(1-\phi_n^2)}{a_n^2}\sum_{t=1}^n Y_{t-1}^2+\frac{Y_n^2}{a_n^2}-\frac{Y_0^2}{a_n^2}-\frac{1}{a_n^2}\sum_{t=1}^n \varepsilon_t^2\Big]+o_p(1)\\
&= c\Big(\frac{2c}{k_n a_n^2}\sum_{i=1}^n Y_{i-1}^2\Big)^{-1}\Big[\frac{2c}{k_n a_n^2}\sum_{t=1}^n Y_{t-1}^2-\frac{1}{a_n^2}\sum_{t=1}^n \varepsilon_t^2\Big]+o_p(1)\\
&\xrightarrow{\mathcal{L}} \Big(\boldsymbol{S}_{\alpha/2,0}+2\sum_{i=1}^\infty \boldsymbol{S}_{\alpha/2,i}\Big)^{-1}\Big(2\sum_{i=1}^\infty \boldsymbol{S}_{\alpha/2,i}\Big)=:(S_{\alpha/2}^*)^{-1}Z_{\alpha/2}^*
\end{aligned}$$

i.e., we complete the proof of Theorem 2.2. □

Acknowledgement

We would like to thank the Editor and the anonymous referees for their critical comments and thoughtful suggestions, which led to a much improved version of this paper.

References

[1] B Basrak, R A Davis, T Mikosch. *Regular variation of GARCH processes*, Stoch Proc Appl, 2002, 99(1): 95-115.

[2] T Bollerslev. *Generalized autoregressive conditional heteroskedasticity*, J Econometrics, 1986, 31(3): 307-327.

[3] N H Chan. *Time series with roots on or near the unit circle*, In: Mikosch T, Kreiß JP, Davis R, Andersen T, (eds), Handbook of Financial Time Series, Springer-Verlag, New York, 2009.

[4] N H Chan, L Peng. *Weighted least absolute deviations estimation for AR(1) process with ARCH(1) errors*, Biometrika, 2005, 92(2): 477-484.

[5] N H Chan, R M Zhang. *Inference for nearly nonstationary processes under strong dependence with infinite variance*, Stat Sinica, 2009, 19: 925-947.

[6] N H Chan, R M Zhang. *Inference for unit-root models with infinite variance GARCH errors*, Stat Sinica, 2010, 20: 1363-1393.

[7] R F Engle. *Discussion: stock market volatility and the crash of 1987*, Rev Financ Stud, 1990, 3(1): 103-106.

[8] F Fornari, A Mele. *Sign- and volatility-switching ARCH models: theory and applications to international stock markets*, J Appl Econom, 1997, 12: 49-65.

[9] C Francq, J M Zakoïan. *Mixing properties of a general class of GARCH(1, 1) models without moment assumptions on the observed process*, Econometric Theory, 2006, 22(5): 815-834.

[10] P Hall, Q W Yao. *Inference in ARCH and GARCH models with heavy-tailed errors*, Econometrica, 2003, 71(1): 285-317.

[11] C He, T Terasvirta. *Properties of moments of a family of GARCH processes*, J Econometrics, 1999, 92(1): 173-192.

[12] H Kesten. *Random difference equations and renewal theory for products of random matrices*, Acta Math, 1973, 131: 207-248.

[13] K Knight. *Limit theory for autoregressive-parameter estimates in an infinite-variance random walk*, Can J Stat, 1989, 17(3): 261-278.

[14] G D Li, W K Li. *Least Absolute Deviations Estimation for unit root processes with GARCH errors*, Ecomomet Theor, 2009, 25(5): 1208-1227.

[15] S Ling, W K Li. *Limiting distributions of maximum likelihood estimators for unstable autoregressive moving-average time series with general autoregressive heteroscedastic errors*, Ann Stat, 1998, 26(1): 84-125.

[16] S Ling, W K Li. *Asymptotic inference for unit-root with GARCH (1,1) errors*, Economet Theor, 2003, 19(4): 541-564.

[17] S Ling, W K Li. *Estimation and testing for a unit-root process with GARCH (1,1) errors*, Economet Rev, 2003, 22(2): 179-202.

[18] D B Nelson. *Stationary and persistence in GARCH (1, 1) models*, Economet Theor, 1990, 6(3): 318-334.

[19] L Peng, Q W Yao. *Least absolute deviations estimation for ARC and GARCH models*, Biometrika, 2003, 90(4): 967-975.

[20] P C B Phillips. *A shortcut to LAD asymptotics*, Economet Theor, 1991, 7: 450-463.

[21] P C B Phillips, T Magdalinos. *Limit theory for moderate deviations from a unit root*, J Econometrics, 2007, 136(1): 115-130.

[22] G W Schwert. *Why does stock market volatility change over time*? J Financ, 1989, 44(5): 1115-1153.

[23] E Sentana. *Quadratic ARCH models*, Rev Econ Stud, 1995, 62(4): 639- 661.

[24] S Taylor. *Modelling Financial Time Series*, Wiley, New York, 1986.

[25] J M Zakoian. *Threshold heteroskedastic models*, J Econ Dyn Control, 1994, 18(5): 931-955.

[26] R M Zhang, S Ling. *Asymptotic inference for AR models with heavy-tailed G-GARCH noises*, Economet Theor, 2015, 31(4): 880-890.

[27] R M Zhang, C Y Sin, S Ling. *On functional limits of short- and long-memory linear processes with GARCH(1,1) noises*, Stoch Proc Appl, 2015, 125(2): 482-512.

[28] R M Zhang, N H Chan. *Nonstationary Linear Processes with Infinite Variance GARCH Errors*, Economet Theor, 2021, 37(5): 892-925.

[29] K Zhu, S Ling. *LADE-based inference for ARMA models with unspecified and heavy-tailed heteroscedastic noises*, J Am Stat Assoc, 2015, 110(510): 784-794.

[1]School of Mathematics and Statistics, Minnan Normal University, Zhangzhou 363000, China.

[2]School of Mathematical Science, Zhejiang University, Hangzhou 310027, China.

Email: rmzhang@zju.edu.cn