Econometric Theory, **37**, 2021, 892–925.
doi:10.1017/S0266466620000377

NONSTATIONARY LINEAR PROCESSES WITH INFINITE VARIANCE GARCH ERRORS*

RONGMAO ZHANG
Zhejiang University

NGAI HANG CHAN
Southwestern University of Finance
The Chinese University of Hong Kong

Recently, Cavaliere, Georgiev, and Taylor (2018, *Econometric Theory* 34, 302–348) (CGT) considered the augmented Dickey–Fuller (ADF) test for a unit-root model with linear noise driven by i.i.d. infinite variance innovations and showed that ordinary least squares (OLS)-based ADF statistics have the same distribution as in Chan and Tran (1989, *Econometric Theory* 5, 354–362) for i.i.d. infinite variance noise. They also proposed an interesting question to extend their results to the case with infinite variance GARCH innovations as considered in Zhang, Sin, and Ling (2015, *Stochastic Processes and their Applications* 125, 482–512). This paper addresses this question. In particular, the limit distributions of the ADF for random walk models with short-memory linear noise driven by infinite variance GARCH innovations are studied. We show that when the tail index $\alpha < 2$, the limit distributions are completely different from that of CGT and the estimator of the parameters of the lag terms used in the ADF regression is not consistent. This paper provides a broad treatment of unit-root models with linear GARCH noises, which encompasses the commonly entertained unit-root IGARCH model as a special case.

1. INTRODUCTION

The asymptotic theory of autoregressive time series with roots on or near the unit circle has been actively pursued by statisticians and econometricians alike. As of today, a relatively complete theory has been established for inference for time series with unit roots or near unit roots when the variance is finite, see, for example, the survey article of Chan (2008) and the references therein.

*We would like to thank the Editor, Professor Peter C.B. Phillips and the Co-Editor, Professor G. Cavaliere, and two anonymous referees for their critical comments and thoughtful suggestions, which led to a much improved version of this paper. Research supported in part by grants from NSFC (No. 11771390/11371318), the USyd-ZJU Partnership Collaboration Awards and the Fundamental Research Funds for the Central Universities (Zhang) and HKSAR-RGC-GRF Nos 14308218 and 14325216 and HKSR-RGC-TRF No. T32-101/15-R (Chan). Address correspondence to Ngai Hang Chan, Department of Statistics, The Chinese University of Hong Kong, Hong Kong; e-mail: nhchan@sta.cuhk.edu.hk.

© The Author(s), 2020. Published by Cambridge University Press.

https://doi.org/10.1017/S0266466620000377 Published online by Cambridge University Press

However, a large number of empirical studies, ranging from signal processing and network traffic to finance, indicate that time series with heavy-tails would be more appropriate. For background information on heavy-tailed time series and their applications, see Finkenstädt and Rootzén (2004) for a survey of important theories and applications of extreme values. In financial econometrics, there has also been increasing interest in modeling financial phenomena by time series driven by heavy-tailed innovation, see, for example, Rachev and Mittnik (2000), Lux and Marchesi (2000), and Bayraktar, Horst, and Sircar (2006).

Although relatively complete theories for the inference for finite variance, nonstationary autoregressive moving average (ARMA) models are available, see, for example, the recent monograph of Andersen et al. (2009), Part IV, less can be said for the infinite variance counterpart. For a unit-root model with i.i.d. heavy-tailed noise, Knight (1989) considered the asymptotic distribution of M-estimate and a least absolute deviation estimate, see also Samarakoon and Knight (2009); Chan and Tran (1989) considered the Dickey–Fuller test; Chan, Peng, and Qi (2006) studied a quantile inference for near-integrated autoregressive time series. Similar results have been extended to unit-root models with linear noises driven by i.i.d. infinite variance innovations, see, for example, Knight (1991) for an M-estimate, Hasan (2001) for a rank test, and Chan and Zhang (2009) for least squares estimate (LSE). Further, to address the dependence, Phillips and Perron (1988) proposed the Phillips–Perron test (PP test) and developed its limit distribution and Cavaliere, Georgiev, and Taylor, 2018 (CGT) proposed to use the augmented Dickey–Fuller (ADF) test, and showed that OLS-based ADF statistics have the same distribution as in Chan and Tran (1989). In that paper, they also proposed the interesting question of extending their results to the case with infinite variance GARCH innovations. This paper addresses this issue.

To shed some intuitive insight into this question, consider the following simple simulation exercise. Let Y_t be a unit-root model, that is, $Y_t = Y_{t-1} + \mu_t$ and the errors $\{\mu_t\}$ be generated from

$$\mu_t = a_1\mu_{t-1} + \varepsilon_t, \quad s\varepsilon_t = \sqrt{\omega + \beta\varepsilon_{t-1}^2}\eta_t, \quad t = 1, 2, \ldots, n, \tag{1.1}$$

where $\{\eta_t\}$ is a sequence of i.i.d. standard normal noise. Note that the tail index of ε_t in model (1.1) 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 model (1.1) with $\omega = 1$, $a_1 = 0.6$, 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 1,000 times and take $n = 300, 500, 1{,}000$. For the generated Y_t, represent it as: $\Delta Y_t = \rho Y_{t-1} + a_1\Delta Y_{t-1} + \varepsilon_t$, $\rho = 0$. Now use the ADF procedure to estimate the parameters ρ and a_1. The empirical sampling bias (Bias) and standard deviation (SD) for the corresponding estimates (Est) $\widehat{\rho}$ and $\widehat{a}_1$ based on the 1,000 repetitions are reported in Table 1, where the lag order in ADF

https://doi.org/10.1017/S0266466620000377 Published online by Cambridge University Press

TABLE 1. Bias and standard deviation for the estimates of ρ and a_1

n	Est	$\beta = 0.5$		$\beta = 1$		$\beta = 1.5$		$\beta = 2$	
		Bias	*SD*	Bias	*SD*	Bias	*SD*	Bias	*SD*
$n = 300$	$\widehat{\rho}$	−0.0085	0.0084	−0.0122	0.0203	−0.0163	0.0326	−0.0287	0.1414
	$\widehat{a}_1$	−0.0437	0.1131	−0.1374	0.2168	−0.2321	0.2889	−0.3073	0.4933
$n = 500$	$\widehat{\rho}$	−0.0052	0.0047	−0.0069	0.0104	−0.0097	0.0224	−0.0164	0.0763
	$\widehat{a}_1$	−0.0373	0.1001	−0.1359	0.2112	−0.2097	0.2911	−0.2789	0.3350
$n = 1{,}000$	$\widehat{\rho}$	−0.0023	0.0021	−0.0038	0.0069	−0.0059	0.0271	−0.0083	0.0730
	$\widehat{a}_1$	−0.0253	0.0872	−0.1394	0.2002	−0.2248	0.2892	−0.3066	0.3388

regression is taken as given in the adf.test in R. It can be seen from this table that for all cases, ρ is estimated satisfactorily. As β increases, that is, the tail index α decreases, the bias and SD for the autoregressive parameter a_1 increase, however. When $\beta > 1$, that is, the noise has infinite variance, the autoregressive parameter a_1 cannot be estimated consistently. This is similar to Zhang and Ling (2015) who showed that the LSE of a stationary AR(p) model is consistent only when $\alpha \geq 2$. Results given in Table 1 are in contrast with the conclusion of CGT for i.i.d. infinite variance innovations, where it is shown that the autoregressive parameters in ADF regression are consistent. Table 1 indicates that the finite sample behavior of the ADF is different from the i.i.d. infinite variance innovations. One main objective of this paper is to investigate this discrepancy.

It should also be pointed out that Knight (1989) (see also Phillips, 1991) showed that L_1 estimation has significant gains in this context for the infinite variance case because it takes advantage of the strong signal in the regressor while controlling the effect of large noises. While the analysis of this latter approach is beyond the scope of the current paper, it certainly offers an interesting and challenging issue to be pursued in future work.

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 proposed methodology is presented in Section 2. Asymptotic theory is developed in Section 3. Section 4 concludes. All the technical proofs are relegated to Section 5.

2. THE MODELS AND TESTS

2.1. Models

Consider the following AR(1) models with linear GARCH noise:

$$Y_t = \phi Y_{t-1} + \mu_t, \tag{2.1}$$

https://doi.org/10.1017/S0266466620000377 Published online by Cambridge University Press

where $Y_0 = 0$ and the μ_t follows a linear process driven by first-order generalized autoregressive conditional heteroscedasticity innovations (GARCH(1, 1))

$$\mu_t = \sum_{i=0}^{\infty} c_i \varepsilon_{t-i},\ c_0 = 1, \tag{2.2}$$

$$\varepsilon_t = \sigma_t \eta_t,\ \sigma_t = \sqrt{\omega + a\sigma_{t-1}^2 + b\varepsilon_{t-1}^2}, \tag{2.3}$$

where $\omega > 0$, $a, b \geq 0$, and $\{\eta_t\}$ are i.i.d. symmetric random variables and satisfy the following conditions.

Condition 1.

(i) $\mathrm{E}\log(a + b\eta_1^2) < 0$.
(ii) There exists a $k_0 > 0$ such that $\mathrm{E}(a + b\eta_t^2)^{k_0} \geq 1$ and $\mathrm{E}(a + b\eta_t^2)^{k_0} \log^+(a + b\eta_t^2) < \infty$, where $\log^+(x) = \max\{0, \log(x)\}$.
(iii) The density of η_1 is positive in a neighborhood of zero.[1]

Under Condition 1(ii), Condition 1(i) is a necessary and sufficient condition for the existence of a stationary solution of σ_t^2 (see Nelson, 1990), further Condition 1(i) is equivalent to $\mathrm{E}(a + b\eta_1^2)^{\mu} < 1$ for some $\mu > 0$, see Rem. 2.9 of Basrak, Davis, and Mikosch (2002). If there exists an $h_0 > 0$ such that $\mathrm{E}|\eta_t|^{h_0} = \infty$ and $\mathrm{E}|\eta_t|^h < \infty$ for all $h < h_0$, then (ii) of Condition 1 holds. By Lem. 7.1 of Chan and Zhan (2010) (see also Kesten, 1973 or Goldie, 1991), it follows that there exists a unique $\alpha \in (0, k_0]$ such that

$$\mathrm{E}(a + b\eta_1^2)^{\alpha/2} = 1. \tag{2.4}$$

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

$$P\left(\sigma_1^2 > x\right) = c_{\alpha} x^{-\alpha/2}\{1 + o(1)\}, \text{ and} \tag{2.5}$$

$$P\left(|\varepsilon_1| > x\right) = P\left(|\sigma_1 \eta_1| > x\right) = \left(E|\eta_1|^{\alpha}\right) c_{\alpha} x^{-\alpha}\{1 + o(1)\}, \tag{2.6}$$

where

$$c_{\alpha} = \frac{2\mathrm{E}\left(\left[\omega + \left(a + b\eta_1^2\right)\sigma_1^2\right]^{\alpha/2} - \left[\left(a + b\eta_1^2\right)\sigma_1^2\right]^{\alpha/2}\right)}{\alpha \mathrm{E}\left((a + b\eta_1^2)^{\alpha/2} \log^+(a + b\eta_1^2)\right)}.$$

Although the unit-root testing problem

$$H_0 : \phi = 1 \quad \leftrightarrow \quad H_1 : |\phi| < 1, \tag{2.7}$$

has been extensively studied, little is known for the case with linear GARCH noise. This paper presents a comprehensive treatment of the limit distribution of the ADF test for models (2.1)–(2.3) under various tail index α.

[1] 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).

https://doi.org/10.1017/S0266466620000377 Published online by Cambridge University Press

2.2. ADF Tests

Throughout, we assume that the coefficients c_is of μ_t in models (2.2) satisfy:

Condition 2. $\sum_{i=0}^{\infty} i|c_i|^{\gamma} < \infty$ for some $\gamma \leq 1$, $\gamma < \alpha$ and the power series $\varphi(z) = \sum_{i=0}^{\infty} c_i z^i$ has no roots on the unit circle.

Under Conditions 1 and 2, μ_t is stationary and invertible. By Thm. 3.8.2 of Brillinger (1981), there exists a summable constant sequence $\{v_i\}$ with $d_0 = 1$ such that

$$\varepsilon_t = \sum_{i=0}^{\infty} v_i \mu_{t-i}.$$

Thus, under H_0,

$$\Delta Y_t = \sum_{i=1}^{\infty} (-v_i)\mu_{t-i} + \varepsilon_t =: \sum_{l=1}^{\infty} \beta_l \Delta Y_{t-l} + \varepsilon_t.$$

To eliminate the effect of the dependence in the noise and improve the unit-root test power, we adopt the ADF regression as Cavaliere et al. (2018). That is, we represent model (2.1)–(2.3) as

$$\Delta Y_t = \rho Y_{t-1} + \sum_{i=1}^{k} \beta_i \Delta Y_{t-i} + e_{t,k}, \tag{2.8}$$

where $\rho = \phi - 1$ and under H_0,

$$e_{t,k} = \sum_{i=k+1}^{\infty} \beta_i \Delta Y_{t-i} + \varepsilon_t.$$

Let $\boldsymbol{\beta}_k = (\beta_1, \ldots, \beta_k)$. Given $Y_0 = 0$ and observations $\{Y_1, \ldots Y_n\}$, we estimate $\rho, \boldsymbol{\beta}$ in model (2.8) by least squares procedure:

$$\begin{aligned}(\widehat{\rho}, \widehat{\boldsymbol{\beta}}_k)^T &= \operatorname{argmin}_{\rho, \boldsymbol{\beta}_k} \sum_{t=k+1}^{n} (\Delta Y_t - \rho Y_{t-1} - \sum_{i=1}^{k} \beta_i \Delta Y_{t-i})^2 \\ &= \begin{pmatrix} \sum_{t=k+1}^{n} Y_{t-1}^2 & \sum_{t=k+1}^{n} \mathbf{X}_{t-1}^T Y_{t-1} \\ \sum_{t=k+1}^{n} \mathbf{X}_{t-1} Y_{t-1} & \sum_{t=k+1}^{n} \mathbf{X}_{t-1} \mathbf{X}_{t-1}^T \end{pmatrix}^{-1} \begin{pmatrix} \sum_{t=k+1}^{n} Y_{t-1} \Delta Y_t \\ \sum_{t=k+1}^{n} \mathbf{X}_{t-1} \Delta Y_t \end{pmatrix}, \end{aligned} \tag{2.9}$$

where $\mathbf{X}_t = (\Delta Y_t, \ldots, \Delta Y_{t-k+1})^T$ and x^T denotes the transpose of x.

3. ASYMPTOTIC DISTRIBUTIONS

We now derive the limit distributions of the LSE in (2.9). From (2.9), it follows that

$$\widehat{\rho} = \left[\sum_{t=k+1}^{n} Y_{t-1}^2 - \left(\sum_{t=k+1}^{n} \mathbf{X}_{t-1}^T Y_{t-1} \right) \left(\sum_{t=k+1}^{n} \mathbf{X}_{t-1} \mathbf{X}_{t-1}^T \right)^{-1} \left(\sum_{t=k+1}^{n} \mathbf{X}_{t-1} Y_{t-1} \right) \right]^{-1}$$

https://doi.org/10.1017/S0266466620000377 Published online by Cambridge University Press

$$\times\left[\sum_{t=k+1}^{n} Y_{t-1}\Delta Y_t-\left(\sum_{t=k+1}^{n}\mathbf{X}_{t-1}^T Y_{t-1}\right)\left(\sum_{t=k+1}^{n}\mathbf{X}_{t-1}\mathbf{X}_{t-1}^T\right)^{-1}\sum_{t=k+1}^{n}\mathbf{X}_{t-1}\Delta Y_t\right] \tag{3.1}$$

and

$$\begin{aligned}&\widehat{\boldsymbol{\beta}}_k^T-\boldsymbol{\beta}_k^T\\&=\left[\sum_{t=k+1}^{n}\mathbf{X}_{t-1}\mathbf{X}_{t-1}^T-\left(\sum_{t=k+1}^{n}\mathbf{X}_{t-1}Y_{t-1}\right)\left(\sum_{t=k+1}^{n}Y_{t-1}^2\right)^{-1}\sum_{t=k+1}^{n}\mathbf{X}_{t-1}^T Y_{t-1}\right]^{-1}\\&\times\left[\sum_{t=k+1}^{n}\mathbf{X}_{t-1}^T e_{tk}-\left(\sum_{t=k+1}^{n}\mathbf{X}_{t-1}Y_{t-1}\right)\left(\sum_{t=k+1}^{n}Y_{t-1}^2\right)^{-1}\sum_{t=k+1}^{n}Y_{t-1}e_{tk}\right].\end{aligned} \tag{3.2}$$

Our first result is about the consistency of $\widehat{\rho}$ and $\widehat{\boldsymbol{\beta}}_k$.

THEOREM 3.1. *Let α be the tail index defined in (2.4). Suppose that Conditions 1, 2 and $k^3\log n/n\to 0$ hold, then under $H_0:\phi=1$, we have*

(a) *For all $\alpha>0$,*

$$\widehat{\rho}=o_p(1). \tag{3.3}$$

(b) *For $\alpha\in[2,\infty)$,*

$$\|\widehat{\boldsymbol{\beta}}_k-\boldsymbol{\beta}_k\|=o_p(1),\ \text{as } k\to\infty. \tag{3.4}$$

(c) *For $\alpha\in(0,2)$, then*

(i) *$\widehat{\boldsymbol{\beta}}_k$ is not a consistent estimate of $\boldsymbol{\beta}_k$.*

(ii) *for any fixed k,*

$$\widehat{\boldsymbol{\beta}}_k-\boldsymbol{\beta}_k\xrightarrow{\mathcal{L}}\boldsymbol{\Sigma}_{\alpha/2}^{-1}(k)\mathbf{Z}_{\alpha/2}(k), \tag{3.5}$$

where $\mathbf{Z}_{\alpha/2}(k)$ is a k-dimensional stable vector with index $\alpha/2$, $\boldsymbol{\Sigma}_{\alpha/2}(k)$ is a $k\times k$ matrix given in (5.55) whose elements are composed of stable variables with index $\alpha/2$.

Remark 3.1. In Theorem 3.1, we only require the dimension k to satisfy $k^3\log n/n\to 0$, that is, fixed k is also allowed. From conclusion (c), we see that $\boldsymbol{\beta}_k$ cannot be consistently estimated by $\widehat{\boldsymbol{\beta}}_k$ when the noise μ_t is driven by heavy-tailed GARCH innovations with tail index $\alpha<2$. This is different from the i.i.d. case when the innovations belong to the domain of attraction of a stable law with index α, see Thm. 1 of CGT, which asserts that for $1<\alpha<2$, if $1/k+k^2/n$ and $k^{1/2}n^{1/\alpha-1}l(n)\sum_{i=k+1}^{\infty}|\beta_i|\to 0$, then the LSE $\widehat{\boldsymbol{\beta}}_k$ is a consistent estimate of $\boldsymbol{\beta}_k$.

https://doi.org/10.1017/S0266466620000377 Published online by Cambridge University Press

Remark 3.2. For simplicity, this paper focuses on GARCH(1,1) noise. It can be seen from the proof that the results also work for GGARCH noise considered in Zhang and Ling (2015).

Now consider the limit distribution of the LSE based ADF

$$R_n^* = (n-k)\widehat{\rho}$$

and the ADF t-statistic

$$Q_n = \widehat{\rho}/s(\widehat{\rho}),$$

where $s(\widehat{\rho}) = \upsilon^{-1}s_e$, s_e is the standard deviation given by

$$s_e^2 = \frac{1}{n-k}\sum_{t=k+1}^{n}(\Delta Y_t - \widehat{\rho}Y_{t-1} - \widehat{\boldsymbol{\beta}}^T\boldsymbol{X}_{t-1})^2, \text{ and}$$

$$\upsilon = \left[\sum_{t=k+1}^{n} Y_{t-1}^2 - \left(\sum_{t=k+1}^{n}\mathbf{X}_{t-1}^T Y_{t-1}\right)\left(\sum_{t=k+1}^{n}\mathbf{X}_{t-1}\mathbf{X}_{t-1}^T\right)^{-1}\left(\sum_{t=k+1}^{n}\mathbf{X}_{t-1}Y_{t-1}\right)\right]^{1/2}.$$

Remark 3.3. The definition of Q_n is the same as that of CGT but R_n^* is a little different from that of R_n in CGT. In fact, $R_n = R_n^*/(1-\sum_{i=1}^{k}\widehat{\beta}_i)$. From Theorem 3.1, we have that when $\alpha < 2$, $\boldsymbol{\beta}$ cannot be consistently estimated by the LSE. Thus, it would be impractical to consider the term $1-\sum_{i=1}^{k}\widehat{\beta}_i$.

THEOREM 3.2. *Suppose that Conditions 1 and 2 hold, then under* $H_0 : \phi = 1$, *we have*

(i) *For* $\alpha \geq 2$, *if* $\sqrt{k}\omega_n + k^3/n + 1/k = o(1)$,

$$R_n \xrightarrow{\mathcal{L}} \left(\int_0^1 W^2(s)\,ds\right)^{-1}\int_0^1 W(s)\,dW(s), \tag{3.6}$$

where $\omega_n = \log n^{-1}I(\alpha = 2) + n^{2/\alpha-1}I(2 < \alpha < 4)$, $I(x)$ *is an indicator function and* $W(s)$ *is a standard Brownian motion.*

(ii) *For* $0 < \alpha < 2$, *if* k *is fixed, then*

$$R_n^* \xrightarrow{\mathcal{L}} \left(\int_0^1 \xi_\alpha^2(s)\,ds\right)^{-1}\left[\int_0^1 \xi_\alpha(s)\,d\xi_\alpha(s) - \mathbf{Z}_\alpha^{*T}(k)\boldsymbol{\Sigma}_{\alpha/2}^{-1}(k)\mathbf{Z}_{\alpha/2}^*(k)\right)\Bigg]$$
$$=: S_{\alpha,k}, \tag{3.7}$$

where $\xi(s)$ *is defined in Lemma 5.1,* $\mathbf{Z}_\alpha^*(k)$ *is a* k*-dimensional stable vector defined in (5.53) and* $\mathbf{Z}_{\alpha/2}^*(k)$, $\boldsymbol{\Sigma}_{\alpha/2}(k)$ *are defined in (5.54) and (5.55). If* $k \to \infty$, *then*

$$R_n^* \xrightarrow{\mathcal{L}} \lim_{k\to\infty} S_{\alpha,k}. \tag{3.8}$$

https://doi.org/10.1017/S0266466620000377 Published online by Cambridge University Press

The next theorem establishes the limit distribution of Q_n.

THEOREM 3.3. *Suppose that Conditions 1 and 2 hold, then under* $H_0 : \phi = 1$, *we have*

(i) *For* $\alpha \geq 2$, *if* $\sqrt{k}\omega_n + k^3/n + 1/k = o(1)$,

$$Q_n \xrightarrow{\mathcal{L}} \left(\int_0^1 W^2(s)\, ds \right)^{-1/2} \int_0^1 W(s)\, dW(s). \tag{3.9}$$

(ii) *For* $0 < \alpha < 2$, *if* k *is fixed, then*

$$Q_n \xrightarrow{\mathcal{L}} \left(\gamma_\alpha(k) \int_0^1 \xi_\alpha^2(t)\, dt \right)^{-1/2} \left[\int_0^1 \xi_\alpha(t)\, d\xi_\alpha(t) - \mathbf{Z}_\alpha^{*T}(k) \boldsymbol{\Sigma}_{\alpha/2}^{-1}(k) \mathbf{Z}_{\alpha/2}^*(k) \right) \Bigg]$$
$$=: \xi_{\alpha,k}, \tag{3.10}$$

where $\gamma_\alpha(k)$ *is a functional of stable processes with index* $\alpha/2$. *If* $k \to \infty$, *then*

$$Q_n \xrightarrow{\mathcal{L}} \lim_{k\to\infty} \xi_{\alpha,k}. \tag{3.11}$$

Remark 3.4. To derive the limit distribution of the LSE, CGT require the coefficients to satisfy $\sum_{i=0}^{\infty} i|c_i|^{\delta/2} < \infty$ for some $\delta \in (0, \alpha) \cap [0, 1]$, we require that $\sum_{i=0}^{\infty} i|c_i| < \infty$, which is much weaker. From the definition of ω_n, when $\alpha > 12/5$, the condition $\sqrt{k}\omega_n + k^3/n = o(1)$ in Theorems 3.2 and 3.3 reduces to $k/n^3 = o(1)$, which is a regular condition and required even for the finite variance case with i.i.d. innovations. When $2 \leq \alpha < 12/5$, due to the cross-product terms $\sum_{t=k}^{n-1} \mu_{t-i}\mu_{t-j}$ in the sample covariance matrix, which have also the same rate as the sample variance (i.e., $i = j$) and are completely different from the i.i.d. innovations case, stricter condition on k is required. We assume that k satisfies $kn^{4/\alpha-2} = o(1)$ for $\alpha > 2$ and $k = o(\log n)$ when $\alpha = 2$.

Remark 3.5. The limit distributions of R_n^* and Q_n are different from that of CGT for $\alpha < 2$. This is because when the innovations ε_ts are i.i.d. variables belonging to the domain of attraction of a stable law, the effect of

$$\left(\sum_{t=k+1}^{n} \mathbf{X}_{t-1}^T Y_{t-1} \right) \left(\sum_{t=k+1}^{n} \mathbf{X}_{t-1}\mathbf{X}_{t-1}^T \right)^{-1} \sum_{t=k+1}^{n} \mathbf{X}_{t-1} e_{tk}$$

in expressing $\widehat{\rho}$ is negligible, see the second term in the first equation of (5.33). When the innovations are heavy-tailed GARCH noise, this effect cannot be neglected, however.

4. CONCLUSION

In this paper, we develop the limit distribution of the augmented Dickey–Fuller (ADF) statistic for a unit-root model with linear noises driven by heavy-tailed

https://doi.org/10.1017/S0266466620000377 Published online by Cambridge University Press

GARCH innovations. It is shown that when the tail index $\alpha \geq 2$ of the GARCH innovations, all parameters in ADF regression can be consistently estimated. However, when $\alpha < 2$, the LSE of the coefficients in the lag terms of the ADF regression is not consistent. The limit distributions of the ADF statistics become technically complex and are different from that driven by i.i.d. heavy-tailed innovations, see Cavaliere et al. (2018). These limit distributions and the numerical results show that ADF may not always be a good device for unit-root testing. Caution should be taken for cases with heavy-tailed GARCH noise. This paper also opens several interesting questions. First, when does the Dickey–Fuller test perform better than the ADF test? Second, is there any efficient testing procedure based on ADF regression? 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 means to handle the inconsistency and efficient testing issue is to adopt the L_1 estimate. Another possibility is to apply the rank test as considered in Hasan (2001). These issues will be explored in a future work. Finally, this paper only deals with the unit-root case without any deterministic trend. It would be straightforward to extend the results to the case with a trend and to the nearly nonstationary case.

5. TECHNICAL PROOFS

In this section, we prove the main results. From the expression of $\widehat{\rho}$ and $\widehat{\boldsymbol{\beta}}^T - \boldsymbol{\beta}^T$, we see that their limit distribution will follow by the asymptotic behaviors of $\sum_{t=k+1}^{n} Y_{t-1}\mu_t$ and $\sum_{t=k}^{n-1} \mu_{t-i}\mu_{t-j}$ for $0 \leq i,j \leq k$. Since $Y_t = \sum_{s=1}^{t} \mu_s + Y_0$ under H_0 and $\mu_t = \sum_{l=0}^{\infty} c_l \varepsilon_{t-l}$ with $\sum_{l=0}^{\infty} l|c_l|^{\gamma} < \infty$, the limit distribution of $\sum_{t=k+1}^{n} Y_{t-1}\mu_t$ can be deduced by the weak convergence of $\sum_{s=1}^{[nt]} \varepsilon_s$ and the continuous mapping theorem and $\sum_{t=k}^{n-1} \mu_{t-i}\mu_{t-j}$ can be approximated by limit distribution of $\left(\sum_{t=k}^{n-1} \varepsilon_{t-h}\varepsilon_t, 0 \leq h \leq M\right)$ as $M \to \infty$ (see Lemma 5.4 below) Thus, the key for the limit distribution of the LSE is to derive the weak convergence of

$$\left(\sum_{t=1}^{[ns]} \varepsilon_t, \sum_{t=1}^{n} \varepsilon_t\varepsilon_{t-h}, h = 1,2,\ldots,M\right),$$

which will be established in Lemma 5.1 based on a point process convergence technique.

Let $a_n^{(\alpha)} = (c_\alpha E|\eta_1|^\alpha n)^{1/\alpha}$. For simplicity, we write $a_n = a_n^{(\alpha)}$ and assume $E\eta_1^2 = 1$ when $\alpha \geq 2$. For any given integers l and M, we define two $(M+1)$-dimensional random vectors:

$$\boldsymbol{X}_{t,l,M} = (\varepsilon_{t-l},\ldots,\varepsilon_{t-l-M}) \text{ and } \boldsymbol{Z}_{t,l,M} = (\varepsilon_{t-l}^2 - w_n, \varepsilon_{t-l}\varepsilon_{t-l-1},\ldots,\varepsilon_{t-l}\varepsilon_{t-l-M}), \tag{5.1}$$

where $w_n = 0$ as $0 < \alpha < 2$, $w_n = E\varepsilon_1^2 I_{(|\varepsilon_1| \leq \sqrt{c_\alpha n})} = c_\alpha \log n$ as $\alpha = 2$, and $w_n = E\varepsilon_1^2$ as $\alpha > 2$. Under Condition 1, $\{\boldsymbol{X}_{t,l,H}, t \geq 1\}$ is a regularly varying random

https://doi.org/10.1017/S0266466620000377 Published online by Cambridge University Press

vector sequence with index α and satisfies the conditions of Thm. 2.8 of Davis and Mikosch (1998), which can be checked easily by the results of Francq and Zakoïan (2006) and Pedersen (2016). Thus, by Thm. 2.8 of Davis and Mikosch (1998) (see also Thm. 3.1 of Mikosch and Stărică, 2000), there exists a Poisson process $\sum_{i=1}^{\infty}\delta_{P_i}$ defined on $\mathbb{R}_+$ with intensity measure $\nu(dy)=\Upsilon\alpha y^{-\alpha-1}\,dy$ and a sequence of i.i.d. point processes $\{\sum_{j=1}^{\infty}\delta_{\boldsymbol{Q}_{ij,H}}\}$ with distribution Q_H, which depends on α, such that

$$\sum_{t=1}^{n}\delta_{\boldsymbol{X}_{t,l,M}/a_n}\xrightarrow{\mathcal{L}}\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}\delta_{P_i\boldsymbol{Q}_{ij,H}}, \tag{5.2}$$

where $\boldsymbol{Q}_{ij,M}=\left(Q_{ij}^0,Q_{ij}^1,\ldots,Q_{ij}^M\right)$, $\left\{\sum_{j=1}^{\infty}\delta_{\boldsymbol{Q}_{ij,H}}\right\}$ is independent of the process $\{P_i\}$, and Υ and Q_M are similarly defined as in Davis and Mikosch (1998). Thus, by Prop. 3.3 of Davis and Mikosch (1998) and the continuous mapping theorem, we have for all l,M,

$$\begin{aligned}\frac{1}{a_n^2}\sum_{t=1}^{n}\boldsymbol{Z}_{t,l,M}&\xrightarrow{\mathcal{L}}\left(\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}P_i^2(Q_{ij}^0)^2,\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}P_i^2Q_{ij}^0Q_{ij}^1,\ldots,\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}P_i^2Q_{ij}^0Q_{ij}^M\right)\\&=:(S^0(\alpha),S^1(\alpha),\ldots,S^M(\alpha))=\boldsymbol{S}_M(\alpha),\end{aligned} \tag{5.3}$$

if $0<\alpha<2$ and

$$\begin{aligned}&\frac{1}{a_n^2}\sum_{t=1}^{n}\boldsymbol{Z}_{t,l,M}\\&\xrightarrow{\mathcal{L}}\left(\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}P_i^2Q_{ij}^0Q_{ij}^hI(|P_i^2Q_{ij}^0Q_{ij}^h|>0)-\int_{(\boldsymbol{x}:|x_0x_h|>0)}x_0x_h\,d\mu(\boldsymbol{x})\right)_{h=0,\cdots,M}\\&=:(S^0(\alpha),S^1(\alpha),\ldots,S^M(\alpha))=\boldsymbol{S}_M(\alpha),\end{aligned} \tag{5.4}$$

if $2\le\alpha<4$, where $\mu(\cdot)=\lim_{n\to\infty}nP(\boldsymbol{Z}_{t,j,M}/n^{\frac{2}{\alpha}}\in\cdot)$ is a measure on $\mathbb{R}^{M+1}$ and $\boldsymbol{x}=(x_0,x_1,\ldots,x_M)$. $\boldsymbol{S}_M(\alpha)$ is a stable random vector with index $\alpha/2$.

Let $S_n(s)=\sum_{t=1}^{[ns]}\varepsilon_t$ and for any positive integer K, define

$$\boldsymbol{S}_n(K,M)=\left(\sum_{t=1}^{n}\boldsymbol{Z}_{t,0,M},\sum_{t=1}^{n}\boldsymbol{Z}_{t,1,M},\ldots,\sum_{t=1}^{n}\boldsymbol{Z}_{t,K,M}\right).$$

And let $\xRightarrow{A}$ denote the weak convergence under A topology, where $A=J_1,M_1,S$; details about these topologies can be found in Avram and Taqqu (1992) and Jakubowski (1997).

LEMMA 5.1. *Under Condition 1, for any positive integer l and M, we have*

(1) *if* $0<\alpha<2$, *then as* $n\to\infty$,

$$\left(S_n(s)/a_n,\boldsymbol{S}_n(K,M)/a_n^2\right)\xrightarrow{\mathcal{S}}(\xi_\alpha(s),\boldsymbol{S}_M(\alpha),\boldsymbol{S}_M(\alpha),\ldots,\boldsymbol{S}_M(\alpha)), \tag{5.5}$$

https://doi.org/10.1017/S0266466620000377 Published online by Cambridge University Press

where $\xi_\alpha(s)$ *is an* α*-stable process with* $\xi_\alpha = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} P_i(\alpha)Q_{ij}^0(\alpha)$ *and* $\boldsymbol{S}_M(\alpha)$ *is defined in (5.3).*

(2) *if* $2 \le \alpha < 4$*, then as* $n \to \infty$*,*

$$\left(S_n(s)/a_n^*, \boldsymbol{S}_n(K,M)/a_n^2\right) \xrightarrow{\mathcal{S}} (W(s), \boldsymbol{S}_M(\alpha), \boldsymbol{S}_M(\alpha), \ldots, \boldsymbol{S}_M(\alpha)), \tag{5.6}$$

where $a_n^* = \sqrt{c_\alpha n \log n} I(\alpha = 2) + \sqrt{n(\mathrm{E}\sigma_1^2)} I(\alpha > 2)$*,* $W(s)$ *is a standard Brownian motion and* $\boldsymbol{S}_M(\alpha)$ *is defined in (5.4).*

(3) *if* $\alpha \ge 4$*, then as* $n \to \infty$*,*

$$\left(S_n(s)/a_n^*, \boldsymbol{S}_n(K,M)/\kappa_n\right) \xrightarrow{J_1} (W(s), \boldsymbol{G}_M(\alpha), \boldsymbol{G}_M(\alpha), \ldots, \boldsymbol{G}_M(\alpha)), \tag{5.7}$$

where $\kappa_n = a_n^2 I(\alpha < 4) + \sqrt{n \log n} I(\alpha = 4) + \sqrt{n} I(\alpha > 4)$ *and* $\boldsymbol{G}_M$ *is an* M*-dimensional normal vector.*

Proof. The weak convergence of $S_n(s)$ can be found in Lem. 4.1 of Zhang, Sin, and Ling (2015) and the convergence of $\boldsymbol{S}_n(K,M)$ can refer to the proof of Thm. 3.3 of Zhang et al. (2015). Thus, it is enough to show their joint convergence.

When $0 < \alpha < 2$, the joint convergence for $0 < \alpha < 2$ follows from (5.2) and the continuous mapping theorem.

When $2 \le \alpha < 4$, since $S_n(s)/a_n^*$ converges weakly to a Brownian motion and $\boldsymbol{S}_n(K,M)/a_n^2$ converges to a α-stable vector, the joint convergence can be proved by Thm. 3 of Resnick and Greenwood (1979).

When $\alpha \ge 4$, the joint convergence can be easily shown by Cramér–Wold device, we omit the details here. □

Let $A_t = a + b\eta_{t-1}^2$ and for all t, h, write

$$\sigma_t^2 = \prod_{k=t-h+1}^{t} A_k \sigma_{t-h}^2 + \omega\left[1 + \sum_{k=t-h+1}^{t-1}\prod_{m=k+1}^{t} A_m\right] =: \Delta_1(t,h) + \Delta_2(t,h).$$

For all i, j, define

$$I_2(n,H,i,j) = \sum_{t=k}^{n-1}\sum_{l=0}^{\infty}\sum_{h=H+1}^{\infty} c_{i-j+l+h} c_l \varepsilon_{t-i-l-h} \Delta_2^{1/2}(t-i-l,h)\eta_{t-i-l},$$

$$I_1(n,H,i,j) = \sum_{t=k}^{n-1}\sum_{l=0}^{\infty}\sum_{h=H+1}^{\infty} c_{i-j+l+h} c_l \varepsilon_{t-i-l-h} (\sigma_{t-i-l} - \Delta_2^{1/2}(t-i-l,h))\eta_{t-i-l}.$$

The following lemma is about the bounds of $I_l(n,H,i,j)$, $l = 1, 2$, which are used to control the upper bound of the largest eigenvalue of $\sum_{t=k}^{n-1} \boldsymbol{X}_t \boldsymbol{X}_t^T$.

https://doi.org/10.1017/S0266466620000377 Published online by Cambridge University Press

LEMMA 5.2. *Under the conditions of Theorem 3.1, there exists a* $0 < \varrho < 1$ *such that for all* $\delta < \min\{\alpha/2, 1\}$ *and* $1 \le i,j \le k$,

$$\mathrm{E}\left|\frac{1}{\kappa_n} I_1(n,H,i,j)\right|^{\delta} = O(\varrho^{\delta H}), \tag{5.8}$$

and for all $\gamma < \theta < \alpha$ *and* $\theta \le 2$,

$$\mathrm{E}\left|\frac{1}{\kappa_n} I_2(n,H,i,j)\right|^{\theta} = O(n\kappa_n^{-\theta}). \tag{5.9}$$

Proof. We first show (5.8). Let $\varsigma = \alpha I(\alpha \le 2)/4 + I(\alpha > 2)$. Since $\varsigma < \alpha/2$, by Hölder inequality, we have

$$\mathrm{E}(a + b\eta_t^2)^{\varsigma} \le [\mathrm{E}(a + b\eta_t^2)^{\alpha/2}]^{2\varsigma/\alpha} = 1.$$

Since $\alpha/2$ is the unique solution of $\mathrm{E}(a+b\eta_t^2)^x = 1$, it follows that $\mathrm{E}(a+b\eta_t^2)^{\varsigma} < 1$. Define $\varrho = [\mathrm{E}(a+b\eta_t^2)^{\varsigma}]^{1/(2\varsigma)}$, then $\varrho < 1$.

Let $\varepsilon_{t,1} = \varepsilon_t I(\sigma_t^2 \le \kappa_n)$, $\varepsilon_{t,2} = \varepsilon_t I(\sigma_t^2 > \kappa_n)$ and

$$I_{11}(n,H,i,j) = \sum_{t=k}^{n-1} \sum_{l=0}^{\infty} \sum_{h=H+1}^{\infty} c_{i-j+l+h} c_l \varepsilon_{t-i-l-h,1} (\sigma_{t-i-l} - \Delta_2^{1/2}(t-i-l,h)) \eta_{t-i-l},$$

$$I_{12}(n,H,i,j) = \sum_{t=k}^{n-1} \sum_{l=0}^{\infty} \sum_{h=H+1}^{\infty} c_{i-j+l+h} c_l \varepsilon_{t-i-l-h,2} (\sigma_{t-i-l} - \Delta_2^{1/2}(t-i-l,h)) \eta_{t-i-l}.$$

Then $I_1(n,H,i,j) = I_{11}(n,H,i,j) + I_{12}(n,H,i,j)$. By (2.5) and Karamata's theorem, we have that for all $\gamma/2 < \delta < \min\{\alpha/2, 1\}$,

$$\begin{aligned}
&\mathrm{E}|\kappa_n^{-1} I_{12}(n,H,i,j)|^{\delta} \\
&\quad \le C\kappa_n^{-\delta} \sum_{t=k}^{n-1} \sum_{l=0}^{\infty} \sum_{h=H+1}^{\infty} |c_{i-j+l+h} c_l|^{\delta} \mathrm{E}|\sigma_{t-i-l-h} \varepsilon_{t-i-l-h,2}|^{\delta} (\mathrm{E}A_1^{\delta/2})^h \\
&\quad = C\kappa_n^{-\delta} n \sum_{h=H+1}^{\infty} \kappa_n^{\delta-\alpha/2} (\mathrm{E}A_1^{\varsigma})^{h\delta/(2\varsigma)} \\
&\quad = C \sum_{h=H+1}^{\infty} \varrho^{\delta h} = O(\varrho^{\delta H}).
\end{aligned} \tag{5.10}$$

Further, by Doob's inequality and Karamata's theorem, we have that for $\alpha/2 < \theta < \alpha$ if $\alpha < 4$ and $\theta = 2$ if $\alpha \ge 4$,

$$\begin{aligned}
&E\left|\kappa_n^{-1} I_{11}(\lfloor n\tau \rfloor, H)\right|^{\theta} \\
&\le C\kappa_n^{-\theta} \sum_{s=-\infty}^{n-i} \mathrm{E}\left| \sum_{h=H+1}^{\infty} \sum_{t=k\vee(i+s)\vee(j+s-h)}^{n-1} c_{t-j-s+h} c_{t-i-s} \varepsilon_{s-h,1} (\sigma_s - \Delta_2^{1/2}(s,h)) \right|^{\theta} (E|\eta_s|^{\theta})
\end{aligned}$$

https://doi.org/10.1017/S0266466620000377 Published online by Cambridge University Press

$$\le C'\kappa_n^{-\theta}\sum_{s=-\infty}^{n-i}\sum_{h=H+1}^{\infty}\left(\sum_{t=k\vee(i+s)\vee(j+s-h)}^{n-1} c_{t-j-s+h}c_{t-i-s}\right)^{\theta}\mathrm{E}\left|\varepsilon_{s-h,1}(\sigma_s-\Delta_2^{1/2}(s,h))\right|^{\theta}$$

$$\le C''\sum_{h=H+1}^{\infty} n\kappa_n^{-\theta}E\left[\sigma_{t-l-h}^{2\theta}I\left(\sigma_{t-l-h}^2\le\kappa_n\right)\right][EA_1^{\theta/2}]^h$$

$$= O\left(\sum_{h=H+1}^{\infty}[EA_1^{\varsigma}]^{\theta h/(2\varsigma)}\right) = O(\varrho^{\theta H}). \tag{5.11}$$

Combining (5.10), (5.11), and Hölder inequality gives (5.8) as desired.

Next, we show (5.9). When $\alpha \le 1$, (5.9) can be shown as (A.9) of Zhang et al. (2015). When $\alpha > 1$, we take $1 < \theta < \alpha$. Note that

$$\left\{\sum_{h=H+1}^{\infty}\sum_{t=k\vee(i+s)\vee(j+s-h)}^{n-1} c_{t-j-s+h}c_{t-i-s}\varepsilon_{s-h}\Delta_2^{1/2}(s,h)\eta_s\right\}$$

and $\{\Delta_2^{1/2}(s,h)\varepsilon_{s-h}\}$ are two sequences of martingale differences with respect to s and h, respectively, using Burkholder's inequality twice, we have

$$\mathrm{E}|I_2(n,H,i,j)|^{\theta}$$

$$\le C\sum_{s=-\infty}^{n-i}\mathrm{E}\left|\sum_{h=H+1}^{\infty}\sum_{t=k\vee(i+s)\vee(j+s-h)}^{n-1} c_{t-j-s+h}c_{t-i-s}\varepsilon_{s-h}\Delta_2^{1/2}(s,h)\right|^{\theta}(E|\eta_s|^{\theta})$$

$$\le C'\sum_{s=-\infty}^{n-i}\sum_{h=H+1}^{\infty}\left(\sum_{t=k\vee(i+s)\vee(j+s-h)}^{n-1} c_{t-j-s+h}c_{t-i-s}\right)^{\theta}\mathrm{E}\left|\varepsilon_{s-h}\Delta_2^{1/2}(s,h)\right|^{\theta}$$

$$= O(n), \tag{5.12}$$

where the last equation follows by $\sum_{i=0}^{\infty}|c_i|<\infty$, see Condition 2. Thus, by (5.12), we have (5.9) and complete the proof of Lemma 5.2. □

Let $d(s,h)=\sum_{l=0\vee(-s-h)}^{\infty} c_{s+l+h}c_l$,

$$S_n(i,j)=\sum_{t=k}^{n-1}\sum_{l=0\vee(i-j)}^{\infty} c_{j-i+l}c_l\varepsilon_{t-j-l}^2+\sum_{t=k}^{n-1}\sum_{l=0}^{\infty}\sum_{h=(j-i-l)\vee1}^{H} c_{i-j+l+h}c_l\varepsilon_{t-i-l}\varepsilon_{t-i-l-h}$$

$$+\sum_{t=k}^{n-1}\sum_{l=0}^{\infty}\sum_{h=(i-j-l)\vee1}^{H} c_{j-i+l+h}c_l\varepsilon_{t-j-l}\varepsilon_{t-j-l-h}=:\sum_{l=1}^{3}S_{nl}(i,j)$$

and A_n be a $k\times k$ matrix with elements in the ith row and jth column as follows:

$$b_n^{-1}\left(S_n(i,j)-d(j-i,0)\sum_{t=k}^{n-1}\varepsilon_t^2-\sum_{h=1}^{H}(d(i-j,h)+d(j-i,h))\sum_{t=k}^{n-1}\varepsilon_t\varepsilon_{t-h}\right),$$

https://doi.org/10.1017/S0266466620000377 Published online by Cambridge University Press

where

$$b_n = a_n^2 I(\alpha < 2) + c_\alpha n \log n I(\alpha = 2) + n(\mathrm{E}\sigma_1^2) I(\alpha > 2).$$

LEMMA 5.3. *Let* $H = -2\log n/(\delta \log \varrho)$. *Suppose that Conditions 1 and 2 and* $k^3 \log n = o(n)$ *hold, then*

$$\|A_n\|_F = o_p(1), \tag{5.13}$$

where $\|\cdot\|_F$ *denotes the Frobenius norm of a matrix.*

LEMMA 5.4. *Under conditions of Lemma 5.3, we have*

$$b_n^{-1} \left\| \sum_{t=k}^{n-1} \boldsymbol{X}_t \boldsymbol{X}_t^T - \sum_{h=0}^{M} D_h \left(\sum_{t=k}^{n-1} \varepsilon_t \varepsilon_{t-h} \right) \right\| = o_p(1) \tag{5.14}$$

by letting $n \to \infty$ *and then* $M \to \infty$, *where* $D_0 = (d(j-i,0))_{ij}$ *and* $D_h = (d(i-j,h) + d(j-i,h))_{ij}$.

The proof of Lemmas 5.3 and 5.4 are put in the appendix.

Remark 5.1. The term $\log n$ in the condition $k^3 \log n = o(n)$ is used only in proving (A.3) and (A.4) in Lemma 5.3. However, from the proof of these equations, we see that when $\alpha \geq 2$, this term is not necessary, that is, we only need $k^3 = o(n)$.

LEMMA 5.5. *Under conditions of Lemma 5.3, there exist* $C_1, C_2 > 0$ *such that*

$$0 < C_1 \leq \lambda_{min} \left(\frac{1}{b_n} \sum_{t=k}^{n-1} \boldsymbol{X}_t \boldsymbol{X}_t^T \right) \leq \lambda_{max} \left(\frac{1}{b_n} \sum_{t=k}^{n-1} \boldsymbol{X}_t \boldsymbol{X}_t^T \right) \leq C_2 \tag{5.15}$$

holds in probability as $n \to 0$, *where* $\lambda_{min}(A)$, $\lambda_{max}(A)$ *denote the smallest and largest eigenvalues of matrix A.*

Proof. By Lemma 5.1, it follows that for any integer $M > 1$ and $h \geq 0$,

$$\begin{aligned} &\max_{0 \leq t \leq M} \left| b_n^{-1} \sum_{t=k}^{n-1} \varepsilon_t \varepsilon_{t-h} \right| \\ &\quad = \max_{0 \leq t \leq M} \left| b_n^{-1} \sum_{t=k}^{n-1} (\varepsilon_t \varepsilon_{t-h} - w_n I(h=0)) + b_n^{-1}(n-k) w_n I(h=0) \right| = O_p(1). \end{aligned}$$

Thus, by Lemma 5.4 and (A.8), we have

$$\lambda_{max} \left(\frac{1}{b_n} \sum_{t=k}^{n-1} \boldsymbol{X}_t \boldsymbol{X}_t^T \right) \leq \left(\sup_h \|D_h\| \right) \left(\max_{0 \leq t \leq M} \left| b_n^{-1} \sum_{t=k}^{n-1} \varepsilon_t \varepsilon_{t-h} \right| \right) = O_p(1), \tag{5.16}$$

which gives the upper bound as desired.

https://doi.org/10.1017/S0266466620000377 Published online by Cambridge University Press

Next, we show the lower bound for the smallest eigenvalue. Since for any $\boldsymbol{x} = (x_1, \ldots, x_k)^T$ with $\|\boldsymbol{x}\| = \sqrt{\sum_{t=1}^{k} x_t^2} = 1$,

$$\boldsymbol{x}\left(\frac{1}{b_n}\sum_{t=k}^{n-1}\boldsymbol{X}_t\boldsymbol{X}_t^T\right)\boldsymbol{x}^T = \frac{1}{b_n}\sum_{t=k}^{n-1}\left(\sum_{l=1}^{k} x_l\mu_{t-1}\right)^2 \geq 0,$$

it is enough to show that

$$\left|\min_{\boldsymbol{x}}\boldsymbol{x}\left(\frac{1}{b_n}\sum_{t=k}^{n-1}\boldsymbol{X}_t\boldsymbol{X}_t^T\right)\boldsymbol{x}^T\right| > C_1 \tag{5.17}$$

holds in probability as $n \to \infty$. By Lemma 5.4, for any $\epsilon > 0$, there exists a large M such that for large n,

$$\left|\boldsymbol{x}\left(\frac{1}{b_n}\sum_{t=k}^{n-1}\boldsymbol{X}_t\boldsymbol{X}_t^T\right)\boldsymbol{x}^T - \boldsymbol{x}\left[\frac{1}{b_n}\sum_{h=0}^{M}D_h\left(\sum_{t=k}^{n-1}\varepsilon_t\varepsilon_{t-h}\right)\right]\boldsymbol{x}^T\right| \leq \epsilon \tag{5.18}$$

holds in probability for all $\boldsymbol{x}$. On the other hand, by Lemma 5.1, it follows that

$$\begin{aligned}\boldsymbol{x}\left[\frac{1}{b_n}\sum_{h=0}^{M}D_h\left(\sum_{t=k}^{n-1}\varepsilon_t\varepsilon_{t-h}\right)\right]\boldsymbol{x}^T &= \sum_{h=0}^{M}(\boldsymbol{x}D_h\boldsymbol{x}^T)\left(\frac{1}{b_n}\sum_{t=k}^{n-1}\varepsilon_t\varepsilon_{t-h}\right) \\ &\xrightarrow{\mathcal{L}} \begin{cases}\sum_{h=0}^{M}(\boldsymbol{x}D_h\boldsymbol{x}^T)Z_{\alpha/2}^h, & \text{if} \quad \alpha < 2, \\ \boldsymbol{x}D_0\boldsymbol{x}^T, & \text{if} \quad \alpha \geq 2.\end{cases}\end{aligned} \tag{5.19}$$

Since D_0 is a symmetric Toeplitz matrix, there exists a constant $C > 0$ such that $\min_{\boldsymbol{x}}\boldsymbol{x}D_0\boldsymbol{x}^T > C$. By (5.19) and (5.18), we have (5.17) for $\alpha \geq 2$.

For $\alpha < 2$, since $\min_{\boldsymbol{x}}\boldsymbol{x}D_0\boldsymbol{x}^T > C$, by the independence of $\{\eta_t\}$, it can be shown that $\sum_{h=0}^{M}(\boldsymbol{x}D_h\boldsymbol{x}^T)Z_{\alpha/2}^h$ is a nondegenerate stable variable with index $\alpha/2$ for all $\boldsymbol{x}$. Since a stable variable is a continuous random variable, it follows that for all $\boldsymbol{x}$,

$$P\left(\sum_{h=0}^{M}(\boldsymbol{x}D_h\boldsymbol{x}^T)Z_{\alpha/2}^h = 0\right) = 0, \tag{5.20}$$

which combining with (5.19) implies that there exists a constant $C_0 > 0$ such that

$$\left|\boldsymbol{x}\left[\frac{1}{b_n}\sum_{h=0}^{M}D_h\left(\sum_{t=k}^{n-1}\varepsilon_t\varepsilon_{t-h}\right)\right]\boldsymbol{x}^T\right| \geq C_0 \tag{5.21}$$

holds in probability as $n \to \infty$. Thus, by (5.18), we have (5.17) for $\alpha < 2$ by taking $C_1 = C_0 - \epsilon > 0$ and complete the proof of Lemma 5.5. □

https://doi.org/10.1017/S0266466620000377 Published online by Cambridge University Press

LEMMA 5.6. *Under Conditions 1, 2 and* $k^3 \log n/n \to 0$, *then under* H_0, *for any* $1 \le i \le k$,

$$\sup_{1\le i\le k}\left|\frac{1}{b_n}\sum_{t=k+1}^{n}\mu_{t-i}Y_{t-1}\right| = O_p(1). \tag{5.22}$$

Proof. Under H_0, for any $1 \le i \le k$, we have

$$\sum_{t=k+1}^{n}\mu_{t-i}Y_{t-1} = \frac{1}{2}\left(Y_{n-i}^2 - Y_{k-i}^2 - \sum_{t=k+1}^{n}\mu_{t-i}^2 + 2\sum_{j=1}^{i}\sum_{t=k+1}^{n}\mu_{t-j}\mu_{t-i}\right). \tag{5.23}$$

By the weak convergence of $S_n(s)$ in Lemma 5.1, Zhang et al. (2015) (see their Thm. 3.2) showed that under the same topology, the weak convergence results also hold for $\sum_{t=1}^{[ns]}\mu_t$, that is,

$$\frac{1}{\sqrt{b_n}}\sum_{t=1}^{[ns]}\mu_t \xrightarrow{\mathcal{S}} \left(\sum_{j=0}^{\infty}c_j\right)(\xi_\alpha(s)I(\alpha<2) + W(s)I(\alpha\ge 2)) =: Z_\alpha(s), \tag{5.24}$$

where the S-topology can be replaced by the J_1-topology when $\alpha \ge 2$. This implies that

$$\sup_{1\le t\le n}|Y_t/\sqrt{b_n}| = O_p(1). \tag{5.25}$$

Further, by Thm. 3.4 of Zhang et al. (2015), it follows that

$$\begin{aligned}&\sup_{0\le s\le 1} b_n^{-1}\sum_{t=1}^{[ns]}\mu_t^2\\ &\quad= \sup_{0\le s\le 1} b_n^{-1}\sum_{t=1}^{[ns]}\left[\mu_t^2 - \left(\sum_{l=0}c_l^2\right)w_n\right] + \left(\sum_{l=0}^{\infty}c_l^2\right)b_n^{-1}nw_n = O_p(1).\end{aligned} \tag{5.26}$$

Combining (5.25) with (5.26) gives

$$\sup_{1\le i\le k} b_n^{-1}\left|Y_{n-i}^2 - Y_{k-i}^2 - \sum_{t=k+1}^{n}\mu_{t-i}^2\right| = O_p(1). \tag{5.27}$$

On the other hand, since for each $1 \le i \le k$, $\sum_{j=1}^{i}\sum_{t=k+1}^{n}\mu_{t-j}\mu_{t-i}$ can be seen as the sum of the last column of the last $i \times i$ submatrix of $\sum_{t=k}^{n-1}\mathbf{X}_t^T\mathbf{X}_t$. By the proof of Lemma 5.4 (see (A.10)), we have that for $H = -2\log n/(\delta \log \varrho)$,

$$\sup_{1\le i\le k} b_n^{-1}\left|\sum_{j=1}^{i}\sum_{t=k+1}^{n}\mu_{t-j}\mu_{t-i} - \sum_{j=1}^{i}\sum_{h=0}^{H}(d(i-j,h)+d(j-i,h))\sum_{t=k}^{n-1}\varepsilon_t\varepsilon_{t-h}\right| = o_p(1).$$

https://doi.org/10.1017/S0266466620000377 Published online by Cambridge University Press

On the other hand, since

$$\max_h \sum_{j=1}^{\infty} |d(i-j,h)+d(j-i,h)| < \infty,$$

it follows from Lemma 5.2 that

$$\frac{1}{b_n}\left|\sum_{j=1}^{i}\sum_{h=0}^{H}(d(i-j,h)+d(j-i,h))\sum_{t=k}^{n-1}\varepsilon_t\varepsilon_{t-h}\right| \le C\sum_{h=0}^{H}\left|\frac{1}{b_n}\sum_{t=k}^{n-1}\varepsilon_t\varepsilon_{t-h}\right| = O_p(1).$$

Thus, by (5.27), we have (5.22) and complete the proof. □

Let $\boldsymbol{d}_0 = (d(1,0), d(2,0), \ldots, d(k,0))^T$ and

$$\boldsymbol{d}_h = (d(-1,h)+d(1,h), d(-2,h)+d(2,h), \ldots, d(-k,h)+d(k,h))^T,\ h \ge 1.$$

LEMMA 5.7. *Under Conditions 1, 2 and $k^3 \log n/n \to 0$, we have that under H_0,*

$$\frac{1}{b_n}\left\|\sum_{t=k+1}^{n}\boldsymbol{X}_{t-1}^T\mu_t - \sum_{h=0}^{M}\boldsymbol{d}_h\sum_{t=k+1}^{n}\varepsilon_t\varepsilon_{t-h}\right\| = o_p(1), \tag{5.28}$$

by letting $n \to \infty$ and then $M \to \infty$.

Proof. Let $H = -2\log n/(\delta \log \varrho)$. By (A.10) in the proof of Lemma 5.4, it follows that when $k^3 = o(n)$,

$$\frac{1}{b_n}\left\|\sum_{t=k+1}^{n}\boldsymbol{X}_{t-1}^T\mu_t - \sum_{h=0}^{H}\boldsymbol{d}_h\sum_{t=k+1}^{n}\varepsilon_t\varepsilon_{t-h}\right\| = o_p(1). \tag{5.29}$$

Further, for a large M, we have

$$\begin{aligned}\frac{1}{b_n}\left\|\sum_{h=M}^{H}\boldsymbol{d}_h\sum_{t=k+1}^{n}\varepsilon_t\varepsilon_{t-h}\right\| &\le \left(\max_{M\le h\le H}\|\boldsymbol{d}_h\|\right)\left(\sum_{h=M}^{H}\left|\frac{1}{b_n}\sum_{t=k+1}^{n}\varepsilon_t\varepsilon_{t-h}\right|\right)\\ &\le C\sum_{h=M}^{H}\left|\frac{1}{b_n}\sum_{t=k+1}^{n}\varepsilon_t\varepsilon_{t-h}\right| = o_p(1)\end{aligned}$$

by Lemma 5.2 and letting $M \to \infty$. This combining with (5.29) gives the conclusion of Lemma 5.7 as desired. □

Now, we are ready to show the theorems.

The proof of Theorem 3.1. By Lemma 5.5, it follows that

$$\left\|\left(\frac{1}{b_n}\sum_{t=k+1}^{n}\mathbf{X}_{t-1}\mathbf{X}_{t-1}^T\right)^{-1}\right\| = O_p(1).$$

https://doi.org/10.1017/S0266466620000377 Published online by Cambridge University Press

Thus, by Lemma 5.6, we have

$$\left(\frac{1}{b_n}\sum_{t=k+1}^{n}\mathbf{X}_{t-1}^{T}Y_{t-1}\right)\left(\frac{1}{b_n}\sum_{t=k+1}^{n}\mathbf{X}_{t-1}\mathbf{X}_{t-1}^{T}\right)^{-1}\left(\frac{1}{b_n}\sum_{t=k+1}^{n}\mathbf{X}_{t-1}Y_{t-1}\right)$$
$$= O_p\left(\left\|\frac{1}{b_n}\sum_{t=k+1}^{n}\mathbf{X}_{t-1}Y_{t-1}\right\|^2\right) = O_p(k). \tag{5.30}$$

Further, since the l_1 norm of each $\boldsymbol{d}_h$ is bounded, by Lemma 5.6 and Lemma 5.7,

$$\left(\frac{1}{b_n}\sum_{t=k+1}^{n}\mathbf{X}_{t-1}^{T}\mu_t\right)\left(\frac{1}{b_n}\sum_{t=k+1}^{n}\mathbf{X}_{t-1}\mathbf{X}_{t-1}^{T}\right)^{-1}\left(\frac{1}{b_n}\sum_{t=k+1}^{n}\mathbf{X}_{t-1}\mu_t\right)$$
$$= O_p\left(\sqrt{k}\left\|\frac{1}{b_n}\sum_{t=k+1}^{n}\mathbf{X}_{t-1}\mu_t\right\|\right) = O_p(\sqrt{k}). \tag{5.31}$$

However, by (5.24),

$$\frac{1}{nb_n}\sum_{t=k+1}^{n}Y_{t-1}^2 \xrightarrow{\mathcal{L}} \int_0^1 Z_\alpha^2(s)\,ds. \tag{5.32}$$

By (3.1), (5.30), and (5.32), we have that under $H_0 : \rho = 0$,

$$\widehat{\rho} = \left(\sum_{t=k+1}^{n}Y_{t-1}^2\right)^{-1}\left[\sum_{t=k+1}^{n}Y_{t-1}\Delta Y_t\right.$$
$$\left.-\left(\sum_{t=k+1}^{n}\mathbf{X}_{t-1}^{T}Y_{t-1}\right)\left(\sum_{t=k+1}^{n}\mathbf{X}_{t-1}\mathbf{X}_{t-1}^{T}\right)^{-1}\sum_{t=k+1}^{n}\mathbf{X}_{t-1}\Delta Y_t\right](1+o_p(1))$$
$$= O(\sqrt{k}/n) = o_p(1),$$

which completes the proof of (3.3).

By Lemma 5.6 and (5.32), it follows that

$$\left\|\frac{1}{b_n}\left(\sum_{t=k+1}^{n}\mathbf{X}_{t-1}Y_{t-1}\right)\left(\sum_{t=k+1}^{n}Y_{t-1}^2\right)^{-1}\sum_{t=k+1}^{n}\mathbf{X}_{t-1}^{T}Y_{t-1}\right\| = O_p(k/n) = o_p(1).$$

This gives

$$\widehat{\boldsymbol{\beta}}_k^T - \boldsymbol{\beta}_k^T = \left(\sum_{t=k+1}^{n}\mathbf{X}_{t-1}\mathbf{X}_{t-1}^{T}\right)^{-1}\left[\sum_{t=k+1}^{n}\mathbf{X}_{t-1}^{T}e_{tk}\right.$$
$$\left.-\left(\sum_{t=k+1}^{n}\mathbf{X}_{t-1}Y_{t-1}\right)\left(\sum_{t=k+1}^{n}Y_{t-1}^2\right)^{-1}\sum_{t=k+1}^{n}Y_{t-1}e_{tk}\right](1+o_p(1)). \tag{5.33}$$

https://doi.org/10.1017/S0266466620000377 Published online by Cambridge University Press

Next, we show (3.4). By (5.24), it follows that

$$\frac{1}{b_n}\sum_{t=k+1}^{n} Y_{t-1}\mu_t \xrightarrow{\mathcal{L}} \int_0^1 Z_\alpha(s)\,dZ_\alpha(s). \tag{5.34}$$

By Lemma 5.6, we have

$$\begin{aligned}\frac{1}{b_n}\left|\sum_{t=k+1}^{n} \boldsymbol{\beta}_k^T \boldsymbol{X}_{t-1}Y_{t-1}\right| &\le \left(\sum_{i=1}^{k}|\beta_i|\right)\left(\max_{1\le i\le k}\frac{1}{b_n}\left|\sum_{t=k+1}^{n}\mu_{t-i}Y_{t-1}\right|\right)\\ &= O_p(1).\end{aligned} \tag{5.35}$$

Thus, by (5.34) and (5.35),

$$\frac{1}{b_n}\sum_{t=k+1}^{n} Y_{t-1}e_{tk} = \frac{1}{b_n}\sum_{t=k+1}^{n} Y_{t-1}(\mu_t - \boldsymbol{\beta}_k^T\boldsymbol{X}_{t-1}) = O_p(1),$$

which combining with (5.32) and Lemma 5.6 yields

$$\frac{1}{b_n}\left\|\left(\sum_{t=k+1}^{n}\mathbf{X}_{t-1}Y_{t-1}\right)\left(\sum_{t=k+1}^{n}Y_{t-1}^2\right)^{-1}\sum_{t=k+1}^{n}Y_{t-1}e_{tk}\right\| = O_p(\sqrt{k}/n). \tag{5.36}$$

Next, we give a bound for $b_n^{-1}\sum_{t=k+1}^{n}\boldsymbol{X}_{t-1}^T e_{tk}$. Let θ be given as in Lemma 5.2, then along the same argument as in Lemma 5.2, it can be shown that

$$\begin{aligned}&\|b_n^{-1}\sum_{t=k+1}^{n}\boldsymbol{X}_{t-1}^T e_{tk}\|\\ &\le \|b_n^{-1}\sum_{t=k+1}^{n}\boldsymbol{X}_{t-1}^T\varepsilon_t\| + \|b_n^{-1}\sum_{j=k+1}^{\infty}\beta_j\sum_{t=k+1}^{n}\boldsymbol{X}_{t-1}^T\mu_{t-j}\|\\ &\le C\kappa_n b_n^{-1}\sum_{i=1}^{k}\left|\frac{1}{\kappa_n}\sum_{t=k+1}^{n}\sum_{j=0}^{\infty}c_j\varepsilon_{t-i-j}\varepsilon_t\right| + \sum_{j=k+1}^{\infty}|\beta_j|b_n^{-1}\sum_{i=1}^{k}\left|\sum_{t=k+1}^{n}\mu_{t-i}\mu_{t-j}\right|\\ &\le C\kappa_n b_n^{-1}\sum_{i=1}^{k}(\varrho^{i+j}+n^{1/\theta}\kappa_n^{-1})(1+o_p(1))\\ &\quad + b_n^{-1}\sum_{j=k+1}^{\infty}|\beta_j|\sum_{t=k+1}^{n}\sum_{s=0}^{\infty}\sum_{i=1}^{k}|c_{j+s-i}c_s|\left[(\varepsilon_{t-j-s}^2-w_n)+w_n\right]\\ &\quad + \kappa_n b_n^{-1}\sum_{j=k+1}^{\infty}\sum_{i=1}^{k}|\beta_j|\left(\left|\kappa_n^{-1}\sum_{t=k+1}^{n}\sum_{l\ne j+s-i}c_l c_s\varepsilon_{t-i-l}\varepsilon_{t-j-s}\right|\right)\\ &= O_p\left(\kappa_n b_n^{-1}+kn^{1/\theta}b_n^{-1}+\sum_{j=k+1}^{\infty}|\beta_j|\right) = o_p(1), \quad \text{as } k\to\infty,\end{aligned} \tag{5.37}$$

https://doi.org/10.1017/S0266466620000377 Published online by Cambridge University Press

where the last inequality follows by taking θ near enough to α for $\alpha \le 2$ and $\theta = 2$ for $\alpha > 2$ and noting $\kappa_n/b_n + k^3/n \to 0$ and $\sum_{i=1}^{\infty} i|\beta_i| < \infty$ if $\alpha \ge 2$. Thus, by (5.33), (5.36), (5.37), and Lemma 5.4, we have that when $\alpha \ge 2$,

$$\widehat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k = o_p(1), \tag{5.38}$$

this completes the proof of (3.4).

Finally, we show conclusion (c) of Theorem 3.1. By (5.33) and (5.36), it follows that under H_0, for $\alpha < 2$,

$$\widehat{\boldsymbol{\beta}}_k^T - \boldsymbol{\beta}_k^T = \left(\sum_{t=k+1}^{n} \mathbf{X}_{t-1}\mathbf{X}_{t-1}^T \right)^{-1} \left(\sum_{t=k+1}^{n} \mathbf{X}_{t-1}^T e_{tk} \right) (1 + o_p(1)). \tag{5.39}$$

Based on (5.39), the limit distribution of (3.5) can be shown similarly to Thm. 2.1 of Zhang and Ling (2015), details are omitted.

By (3.5), we know it suffices to show that conclusion c(i) holds for $k \to \infty$. By Lemma 5.4, it follows that $\lambda_{\min}\left(b_n^{-1}\sum_{t=k+1}^{n} \mathbf{X}_{t-1}\mathbf{X}_{t-1}^T\right)^{-1} > 0$ in probability. This implies that if there exists a positive constant C such that

$$P\left\{ \frac{1}{b_n} \left\| \sum_{t=k+1}^{n} \mathbf{X}_{t-1}^T e_{tk} \right\| > C \right\} > 0, \tag{5.40}$$

as $n \to \infty$, then by (5.39), there must exist a constant $C' > 0$ such that

$$\|\widehat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k\| > C'$$

with a positive probability. Since $\sum_{l=0}^{\infty} l|\beta_l| < \infty$, using the same argument as that in (5.37), we can show that as $k \to \infty$,

$$\frac{1}{b_n} \left\| \sum_{j=k+1}^{\infty} \beta_j \sum_{t=k+1}^{n} \boldsymbol{X}_{t-1}^T \mu_{t-j} \right\| = O_p\left(\sum_{j=k+1}^{\infty} k|\beta_j| \right) = O_p\left(\sum_{j=k+1}^{\infty} j|\beta_j| \right) = o_p(1).$$

It follows that for any given i,

$$\frac{1}{b_n} \left\| \sum_{t=k+1}^{n} \boldsymbol{X}_{t-1}^T e_{tk} - \sum_{t=k+1}^{n} \boldsymbol{X}_{t-1}^T \varepsilon_t \right\| = o_p(1). \tag{5.41}$$

By Lemma 5.1, it follows that for any i,

$$\frac{1}{b_n} \sum_{t=k+1}^{n} \mu_{t-i}\varepsilon_t \xrightarrow{\mathcal{L}} \varphi(1) S_M(\alpha), \text{ where } \varphi(1) = \sum_{i=0}^{\infty} c_i.$$

This combining with (5.41) implies that for any i,

$$\frac{1}{b_n} \sum_{t=k+1}^{n} \mu_{t-i} e_{tk} \xrightarrow{\mathcal{L}} \varphi(1) S_M(\alpha). \tag{5.42}$$

Thus, (5.40) holds and $\widehat{\boldsymbol{\beta}}_k$ cannot be a consistent estimator of $\boldsymbol{\beta}_k$ when $\alpha < 2$. □

https://doi.org/10.1017/S0266466620000377 Published online by Cambridge University Press

The proof of Theorem 3.2. We first consider the case with $\alpha \geq 2$. By Lemma 5.1 and the continuous mapping theorem, it follows that for any given M,

$$\sup_{0\leq h\leq M} \frac{1}{\kappa_n} \sum_{t=k}^{n-1} (\varepsilon_t \varepsilon_{t-h} - w_n I(h=0)) \xrightarrow{\mathcal{L}} \sup_{0\leq h\leq M} S^h(\alpha), \tag{5.43}$$

where $S^h(\alpha)$s are defined in (5.4). Since $\|D_h\| < C$, by (5.43) and Lemma 5.4, we have when $\alpha \geq 2$,

$$\left\| b_n^{-1} \sum_{t=k}^{n-1} \boldsymbol{X}_t \boldsymbol{X}_t^T - D_0 \right\| = o_p(1) \tag{5.44}$$

by taking $n \to \infty$ and then $M \to \infty$. Note that under $H_0 : \rho = 0$, we can rewrite $\widehat{\rho}$ in (3.1) with ΔY_t being replaced by e_{tk}. By Lemmas 5.6 and (5.37) and $\sqrt{k}\kappa_n/b_n = o(1)$ for $\alpha \geq 2$,

$$\begin{aligned} &\left(\frac{1}{b_n} \sum_{t=k+1}^{n} \mathbf{X}_{t-1}^T Y_{t-1} \right) \left(\frac{1}{b_n} \sum_{t=k+1}^{n} \mathbf{X}_{t-1} \mathbf{X}_{t-1}^T \right)^{-1} \left(\frac{1}{b_n} \sum_{t=k+1}^{n} \mathbf{X}_{t-1} e_{tk} \right) \\ &\quad = O_p \left(\sqrt{k}\kappa_n b_n^{-1} + \sqrt{k} \sum_{j=k+1}^{\infty} |\beta_j| \right) = o_p(1). \end{aligned} \tag{5.45}$$

Thus, by (3.1) and (5.30), we have

$$\begin{aligned} \widehat{\rho} &= \left(\sum_{t=k+1}^{n} Y_{t-1}^2 \right)^{-1} \left(\sum_{t=k+1}^{n} Y_{t-1} e_{tk} \right) (1 + o_p(1)) \\ &= \left(\sum_{t=k+1}^{n} Y_{t-1}^2 \right)^{-1} \left(\sum_{t=k+1}^{n} Y_{t-1} \left(\varepsilon_t + \sum_{i=k+1}^{\infty} \beta_i \mu_{t-i} \right) \right) (1 + o_p(1)). \end{aligned}$$

Under Condition 2, by Beveridge–Nelson representation, we can write μ_t as

$$\mu_t = \varphi(1)\varepsilon_t + (v_{t-1} - v_t), \tag{5.46}$$

where $v_t = \sum_{i=0}^{\infty} \pi_i \varepsilon_{t-i}$ and $\pi_i = \sum_{k=i+1}^{\infty} c_i$ satisfying $\sum_{i=0}^{\infty} |\pi_i| < \infty$. By (5.46) and the same argument as in (5.37), it follows that

$$\begin{aligned} \frac{1}{b_n} \sum_{t=k+1}^{n} Y_{t-1}\varepsilon_t &= \varphi(1) \frac{1}{b_n} \sum_{t=k+1}^{n} \sum_{i=1}^{t-1} \varepsilon_i \varepsilon_t + \frac{1}{b_n} \sum_{t=k+1}^{n} (v_0 - v_{t-1})\varepsilon_t \\ &= \varphi(1) \frac{1}{b_n} \sum_{t=k+1}^{n} \sum_{i=1}^{t-1} \varepsilon_i \varepsilon_t + O_p(\kappa_n/b_n) \\ &\xrightarrow{\mathcal{L}} \varphi(1) \int_0^1 W(s)\, dW(s). \end{aligned} \tag{5.47}$$

https://doi.org/10.1017/S0266466620000377 Published online by Cambridge University Press

On the other hand, by (5.24) and the continuous mapping theorem, it follows that

$$\frac{1}{nb_n}\sum_{t=k+1}^{n} Y_{t-1}^2 \xrightarrow{\mathcal{L}} \varphi^2(1)\int_0^1 W^2(s)\,ds. \tag{5.48}$$

Thus, by (5.47), (5.48), (3.4), and $\varphi(1) = 1/(1-\sum_{i=1}^{\infty}\beta_i)$, for the proof of Theorem 3.2(i), it is enough to show

$$\frac{1}{b_n}\sum_{t=k+1}^{n} Y_{t-1}\sum_{i=k+1}^{\infty}\beta_i\mu_{t-i} = \frac{\varphi(1)}{b_n}\sum_{t=k+1}^{n}\sum_{i=k+1}^{\infty}\beta_i Y_{t-1}\varepsilon_{t-i} = o_p(1). \tag{5.49}$$

By (5.46) and a similar argument as in (5.37), it follows that

$$\begin{aligned}
&\frac{1}{b_n}\sum_{t=k+1}^{n}\sum_{i=k+1}^{\infty}\beta_i Y_{t-1}\varepsilon_{t-i}\\
&= \frac{\varphi(1)}{b_n}\sum_{t=k+1}^{n}\sum_{i=k+1}^{\infty}\beta_i\sum_{j=1}^{t-1}\varepsilon_{t-j}\varepsilon_{t-i} - \frac{1}{b_n}\sum_{t=k+1}^{n}\sum_{i=k+1}^{\infty}\beta_i v_{t-1}\varepsilon_{t-i} + \frac{v_0}{b_n}\sum_{t=k+1}^{n}\sum_{i=k+1}^{\infty}\beta_i\varepsilon_{t-i}\\
&= \frac{\varphi(1)}{b_n}\sum_{t=k+1}^{n}\sum_{i=k+1,\,i\neq j}^{\infty}\sum_{j=1}^{t-1}\beta_i\varepsilon_{t-j}\varepsilon_{t-i} + \frac{\varphi(1)}{b_n}\sum_{i=k+1}^{n-1}\beta_i\sum_{t=k+1}^{n}(\varepsilon_{t-i}^2 - w_n)\\
&+ \frac{w_n\varphi(1)}{b_n}\sum_{i=k+1}^{n-1}(n-k-1)\beta_i + o_p(1)\\
&= \frac{\varphi(1)}{b_n}\sum_{i=k+1}^{\infty}\sum_{t=k+1}^{n}\sum_{j=1,j\neq i}^{t-1}\beta_i\varepsilon_{t-j}\varepsilon_{t-i} + O_p\left(\sum_{i=k+1}^{\infty}|\beta_i|\right) + o_p(1) \quad \text{(by (5.26))}\\
&= \frac{\varphi(1)}{b_n}\sum_{i=k+1}^{\infty}\sum_{t=k+1}^{n}\sum_{j=i+1}^{t-1}\beta_i\varepsilon_{t-j}\varepsilon_{t-i} + \frac{\varphi(1)}{b_n}\sum_{i=k+1}^{\infty}\sum_{t=k+1}^{n}\sum_{j=1}^{i-1}\beta_i\varepsilon_{t-j}\varepsilon_{t-i} + o_p(1)\\
&= \frac{\varphi(1)}{b_n}\sum_{i=k+1}^{\infty}\sum_{t=k+1}^{n}\sum_{j=i+1}^{t-1}\beta_i\varepsilon_{t-j}\varepsilon_{t-i} + O_p\left(\sum_{i=k+1}^{\infty} i|\beta_i|\right)\\
&= \frac{\varphi(1)}{b_n}\sum_{l=1}^{2}\sum_{i=k+1}^{\infty}\sum_{t=k+1}^{n}\sum_{j=i+1}^{t-1}\beta_i\widetilde{\varepsilon}_{t-j,l}\Delta_2^{1/2}(t-i,j-i)\eta_{t-i}\\
&+ \frac{\varphi(1)}{b_n}\sum_{i=k+1}^{\infty}\sum_{t=k+1}^{n}\sum_{j=i+1}^{t-1}\beta_i\varepsilon_{t-j}(\sigma_{t-i} - \Delta_2^{1/2}(t-i,j-i))\eta_{t-i} + o_p(1)\\
&=: \varphi(1)(\Gamma_1+\Gamma_2+\Gamma_3),
\end{aligned}$$

https://doi.org/10.1017/S0266466620000377 Published online by Cambridge University Press

where $\widetilde{\varepsilon}_{t,1} = \varepsilon_t I(\sigma_t \le n), \widetilde{\varepsilon}_{t,2} = \varepsilon_t I(\sigma_t > n)$. Using the same argument as in the proof of Lemma 5.2, it can be shown that

$$\mathrm{E}|\Gamma_1| \le \sum_{i=k+1}^{\infty} \frac{|\beta_i|}{b_n} \left[\mathrm{E} \left(\sum_{t=k+1}^{n} \sum_{j=i+1}^{t-1} \widetilde{\varepsilon}_{t-j,1} \Delta_2^{1/2}(t-i,j-i) \eta_{t-i} \right)^2 \right]^{1/2} = \left(\sum_{i=k+1}^{\infty} |\beta_i| \right), \tag{5.50}$$

for any $\theta < \alpha$ if $\alpha = 2$ and $\theta = 2$ if $\alpha > 2$,

$$\mathrm{E}|\Gamma_2|^{\theta} \le \sum_{i=k+1}^{\infty} \frac{|\beta_i|}{b_n} \left[\mathrm{E} \left(\sum_{t=k+1}^{n} \sum_{j=i+1}^{t-1} \widetilde{\varepsilon}_{t-j,2} \Delta_2^{1/2}(t-i,j-i) \eta_{t-i} \right)^{\theta} \right]^{1/\theta} = \left(\sum_{i=k+1}^{\infty} |\beta_i| \right), \tag{5.51}$$

and for $\delta < \min\{\alpha/2, 1\}$,

$$\begin{aligned} \mathrm{E}|\Gamma_3|^{\delta} &\le \frac{1}{b_n} \sum_{i=k+1}^{\infty} |\beta_i| \sum_{j=i+1}^{n-1} \mathrm{E} \Big| \sum_{t=k\vee j+1}^{n} \varepsilon_{t-j} (\sigma_{t-i} - \Delta_2^{1/2}(t-i,j-i)) \eta_{t-i} \Big|^{\delta} \\ &= O\left(\sum_{i=k+1}^{\infty} |\beta_i| \sum_{j=i+1}^{n-1} \varrho^{\delta(j-i)} \right) = O\left(\sum_{i=k+1}^{\infty} |\beta_i| \right). \end{aligned} \tag{5.52}$$

Thus, by (5.50), (5.51), and (5.52),

$$b_n^{-1} \sum_{t=k+1}^{n} \sum_{i=k+1}^{\infty} \beta_i Y_{t-1} \varepsilon_{t-i} = o_p(1)$$

and complete the proof of Theorem 3.2(i).

Next, we show conclusion (ii), that is, the case with $\alpha < 2$. When k is fixed, by Lemma 5.1, using the same argument as in Lem. 3.2 of Zhang and Ling (2015), we can show that

$$\begin{aligned} &\frac{1}{b_n} \sum_{t=k+1}^{n} \boldsymbol{X}_{t-1}^T Y_{t-1} \\ &= \frac{1}{b_n} \left(\sum_{t=k+1}^{n} \mu_{t-1} Y_{t-1}, \ldots, \sum_{t=k+1}^{n} \mu_{t-k} Y_{t-1} \right) \\ &= \frac{1}{2b_n} \left(Y_{n-1}^2 - Y_{k-1}^2 + \sum_{t=k+1}^{n} \mu_{t-1}^2, \ldots, Y_{n-k}^2 + 2 \sum_{j=1}^{k-1} \sum_{t=k+1}^{n} \mu_{t-j} \mu_{t-k} + \sum_{t=k+1}^{n} \mu_{t-k}^2 \right) \\ &= \frac{1}{2b_n} \left(Y_{n-1}^2 + \sum_{t=k+1}^{n} \mu_t^2, \ldots, Y_{n-k}^2 + \sum_{t=k+1}^{n} \mu_t^2 + 2 \sum_{j=1}^{k-1} \sum_{t=k+1}^{n} \mu_{t-j} \mu_{t-k} \right) + o_p(1) \end{aligned}$$

https://doi.org/10.1017/S0266466620000377 Published online by Cambridge University Press

$$\xrightarrow{\mathcal{L}} \frac{(Z_\alpha^2(1)+Z_{\alpha/2}^0(\mu))\times \mathbf{1}}{2}+\left(0, Z_{\alpha/2}^1(\mu), \ldots, \sum_{j=1}^{k-1} Z_{\alpha/2}^j(\mu)\right) =: Z_\alpha^{*T}(k), \tag{5.53}$$

and

$$\frac{1}{b_n}\sum_{t=k+1}^{n} \boldsymbol{X}_t \Delta Y_t \xrightarrow{\mathcal{L}} (Z_{\alpha/2}^1(\mu), Z_{\alpha/2}^2(\mu), \ldots, Z_{\alpha/2}^k(\mu))^T =: \boldsymbol{Z}_{\alpha/2}^*(k), \tag{5.54}$$

$$\frac{1}{b_n}\sum_{t=k+1} \boldsymbol{X}_t \boldsymbol{X}_t^T \xrightarrow{\mathcal{L}} (Z_{\alpha/2}^{|i-j|}(\mu)) =: \boldsymbol{\Sigma}_{\alpha/2}(k), \tag{5.55}$$

where $Z_\alpha(s)$ is defined in (5.24), $\mathbf{1} = (1, 1, \ldots, 1)$ and $Z_{\alpha/2}^j(\mu), j = 0, 1, \ldots, k$ are defined as $Z_{\alpha/2}^j(Y)$ in Zhang and Ling (2015). Further, the convergences in (5.53)–(5.55) hold jointly. Thus, by the continuous mapping theorem, we have Theorem 3.2(ii) and complete the proof. □

Proof of Theorem 3.3. Let $\boldsymbol{Q}_t = (Y_t, \boldsymbol{X}_t^T)^T$, then

$$s_e^2 = \frac{1}{n-k}\left[\sum_{t=k+1}^{n} e_{tk}^2 - \left(\sum_{t=k+1}^{n} \boldsymbol{Q}_t^T e_{tk}\right)\left(\sum_{t=k+1}^{n} \boldsymbol{Q}_t \boldsymbol{Q}_t^T\right)^{-1}\left(\sum_{t=k+1}^{n} \boldsymbol{Q}_t e_{tk}\right)\right]. \tag{5.56}$$

By this expression, Theorem 3.3 can be shown similarly to Theorem 3.2, we omit the details here. □

REFERENCES

Andersen, T.G., R.A. Davis, J. Kreiss, & T. Mikosch (2009) *Springer Handbook of Financial Time Series*. Springer.

Avram, F. & M.S. Taqqu (1992) Weak convergence of sums of moving averages in the α-stable domain of attraction. *Annals of Probability* 20, 483–503.

Basrak, B., R.A. Davis, & T. Mikosch (2002) Regular variation of GARCH processes. *Stochastic Processes and their Applications* 99, 95–115.

Bayraktar, E., U. Horst, & K.R. Sircar (2006) A limiting theorem for financial markets with inert investors. *Mathematics of Operations Research* 31, 789–810.

Brillinger, D. (1981) *Time Series: Data Analysis and Theory*. Holden Day.

Cavaliere, G., I. Georgiev, & A.M.R. Taylor (2018) Unit root inference for non-stationary linear processes driven by infinite variance innovations. *Econometric Theory* 34, 302–348.

Chan, N.H. (2008) Time series with roots on or near the unit circle. In T.G. Andersen, R.A. Davis, J. Kreiss and T. Mikosch (eds.), *Springer Handbook of Financial Time Series*, Springer.

Chan, N.H., L. Peng, & Y.C. Qi (2006) Quantile inference for near-integrated autoregressive time series with infinite variance. *Statistica Sinica* 16, 15–28.

Chan, N.H. & L.T. Tran (1989) On the first order autoregressive process with infinite variance. *Econometric Theory* 5, 354–362.

Chan, N.H. & R.M. Zhan (2010) Inference for unit-root models with infinite variance GARCH errors. *Statistica Sinica* 20, 1363–1393.

https://doi.org/10.1017/S0266466620000377 Published online by Cambridge University Press

Chan, N.H. & R.M. Zhang (2009) Inference for nearly nonstationary processes under strong dependence with infinite variance. *Statistica Sinica* 19, 925–947.
Davis, R.A. & T. Mikosch (1998) The sample autocorrelations of heavy-tailed processes with applications to ARCH. *Annals of Statistics* 26, 2049–2080.
Finkenstädt, B. & H. Rootzén (2004) *Extreme Values in Finance, Telecommunications, and the Environment*. Chapman and Hall/CRC Press.
Francq, C. & J.M. Zakoïan (2006) Mixing properties of a general class of GARCH(1, 1) models without moment assumptions on the observed process. *Econometric Theory* 22, 815–834.
Goldie, C.M. (1991) Implicit renewal theory and tails of solutions of random equations. *Annals of Applied Probability* 1, 126–166.
Hasan, M.N. (2001) Rank tests of unit root hypothesis with infinite variance errors. *Journal of Econometrics* 104, 49–65.
Jakubowski, A. (1997) A non-Skorohod topology on Okhotsk space. *Electronic Journal of Probability* 2, 1–21.
Kesten, H. (1973) Random difference equations and renewal theory for products of random matrices. *Acta Mathematica* 131, 207–248.
Knight, K. (1989) Limit theory for autoregressive-parameter estimates in an infinite-variance random walk. *Canadian Journal of Statistics* 17, 261–278.
Knight, K. (1991) Limit theory for M-estimates in an integrated infinite variance processes. *Econometric Theory* 7, 200–212.
Lux, T. & M. Marchesi (2000) Volatility clustering in financial markets: a microsimulation of interesting agents. *International Journal of Theoretical and Applied Finance* 3, 675–702.
Mikosch, T. & C. Stărică (2000) Limit theory for the sample autocorrelations and extremes of a GARCH (1,1) process. *Annals of Statistics* 28, 1427–1451.
Nelson, D.B. (1990) Stationary and persistence in GARCH (1, 1) models. *Econometric Theory* 6, 318–334.
Pedersen, R. (2016) Targeting estimation of CCC-GARCH models with infinite fourth moments. *Econometric Theory* 32, 498–531.
Phillips, P.C.B. (1991) A shortcut to the LAD estimator asymptotics. *Econometric Theory* 7, 450–463.
Phillips, P.C.B. & P. Perron (1988) Testing for a unit root in time series regression. *Biometrika* 75, 335–346.
Rachev, S.T. & S. Mittnik (2000) *Stable Paretian Models in Finance*. Wiley.
Resnick, S. & P. Greenwood (1979) A bivariate stable characterization and domains of attraction. *Journal of Multivariate Analysis* 9, 206–221.
Samarakoon, D.M.M. & K. Knight (2009) A note on unit root tests with infinite variance noise. *Econometric Review* 28, 314–334.
Zhang, R.M. & S. Ling (2015) Asymptotic inference for AR models with heavy-tailed G-GARCH noises. *Econometric Theory* 31, 880–890.
Zhang, R.M., C.Y. Sin, & S. Ling (2015) On functional limits of short- and long-memory linear processes with GARCH(1,1) noises. *Stochastic Processes and their Applications* 125, 482–512.

APPENDIX

A.1. Simulations

In this subsection, we present some simulations for the test $R_n^* = (n-k)\widehat{\rho}$. We consider a unit-root process $\{Y_t\}$, that is, $Y_t = Y_{t-1} + \mu_t$, where the noise $\{\mu_t\}$ is generated from

$$\mu_t = a_1\mu_{t-1} + \varepsilon_t - b_1\varepsilon_{t-1}, \quad \varepsilon_t = \sqrt{\omega + \beta\varepsilon_{t-1}^2}\,\eta_t, \quad t = 1, 2, \ldots, n, \tag{A.1}$$

https://doi.org/10.1017/S0266466620000377 Published online by Cambridge University Press

and $\{\eta_t\}$ is a sequence of i.i.d. standard normal innovations. In our simulations, β is taken as 0.5, 1, 1.5, and 2 for the cases with tail index $\alpha > 2, \alpha = 2$ and $1 < \alpha < 2$ and $\alpha < 1$, respectively. For each setting, we replicate the simulation 10,000 times with sample size 5,000 and the lag order k is taken as the same as in the adf.test function in R. The empirical critical values for different β with $(a_1, b_1) = (0.8, 0)$ (and $(0.6, 0.5)$) over the 10,000 replications are listed in Table A1 (and Table A2, respectively) with label "IR."

To illustrate whether the wild bootstrap works for our case, we also simulate the empirical values based on the wild bootstrap technique developed by CGT with sample size $n = 2,000$ and $m = 5,000$ repetitions, which is included in Tables A1 and A2 with label "WB."

Tables A1 and A2 show that when $\beta = 0.5$ and 1, that is, the tail index $\alpha \geq 2$, then the empirical critical values based on independent repetitions and wild bootstrap are close. This is because that when $\alpha \geq 2$, the effect of

$$\left(\sum_{t=k+1}^{n} \mathbf{X}_{t-1}^T Y_{t-1}\right)\left(\sum_{t=k+1}^{n} \mathbf{X}_{t-1}\mathbf{X}_{t-1}^T\right)^{-1} \sum_{t=k+1}^{n} \mathbf{X}_{t-1}\Delta Y_t \tag{A.2}$$

in expressing $\widehat{\rho}$ is negligible, consequently, this term does not appear in the limit distribution of $R_n^* = (n-k)\widehat{\rho}$, see Theorem 3.2(i), which is consistent with the limit distribution based on wild bootstrap. On the other hand, Tables A1 and A2 show that when $\beta = 1.5$ and 2, that is, the tail index $\alpha < 2$, the empirical critical values based on independent repetitions and wild bootstrap are very different. This is because that when the innovations are heavy-tailed GARCH noise with index $\alpha < 2$, the (A.2) effect cannot be neglected. Consequently, the limit distributions of the ADF test $R_n^* = (n-k)\widehat{\rho}$ depend on not only the tail index, but also the lag order used in the ADF regression, see Theorem 3.2(ii). Because conditional on the observations, the cross-product terms in wild bootstrap samples are independent, the term (A.2) in the bootstrap sample becomes negligible, which leads to a different limit distribution from this paper. As a result, their critical values are different.

Tables A1 and A2 indicate that the wild bootstrap technique developed by CGT works only for the case with $\alpha \geq 2$ but not for $\alpha < 2$.

Based on the empirical values in Tables A1 and A2, we also consider the size for the test of R_n^*. We simulate the unit-root process $\{Y_t\}$ with noise μ_t in model (A.1) as above, where (a_1, b_1) is taken as $(0.8, 0)$ and $(0.6, 0.5)$ and the sample size n is taken as 300, 500, and 1,000. In each setting, we repeat 1,000 simulations. The size results over the 1000 repetitions are reported in Table A3, which shows that both sizes based on the critical values of independent repetitions and wild bootstrap perform reasonably when $\beta = 0.5$ or $\beta = 1$, that is, the tail index $\alpha \geq 2$. However, when $\beta = 1.5$ or $\beta = 2$, that is, the tail index $\alpha > 2$, then the size based on the critical values of wild bootstrap performs very poorly. This also confirms our results that when $\alpha < 2$, the limit distribution is affected by the cross-product term, that is, the (A.2) term and the wild bootstrap technique cannot be applied directly in this case.

Finally, we consider the test R_n^* at the frontier, that is, the cases with little GARCH effect and/or when the finite variance assumption on $\mathrm{E}\eta_t^2$ is violated. To this end, we simulate the unit-root process with noise derived by model (A.1) and η_t is taken as a sequence of i.i.d. stable innovations with tail index $\alpha = 0.5, 1,$ and 1.5. We take $(a_1, b_1) = (0.6, 0)$ and sample size $n = 5,000$. In each setting, we replicate the simulations 10,000 times. Just as stated in analyzing Table 1, the generated process $\{Y_t\}$ can be represented as

$$\Delta Y_t = \rho Y_{t-1} + a_1 \Delta Y_{t-1} + \varepsilon_t, \rho = 0.$$

https://doi.org/10.1017/S0266466620000377 Published online by Cambridge University Press

https://doi.org/10.1017/S0266466620000377 Published online by Cambridge University Press

TABLE A1. Empirical critical values based on the wild bootstrap and independent repetitions for $\omega = 1, a_1 = 0.8,$ and $b_1 = 0$ in model (A.1)

β	Method	0.01	0.025	0.05	0.10	0.20	0.80	0.95	0.975	0.99
0.5	IR	−7.6060	−5.7004	−4.5299	−3.4755	−2.4088	−0.4376	−0.0186	0.1514	0.3876
	WB	−7.8043	−6.2790	−5.2746	−4.1425	−2.9356	−0.5477	−0.0558	0.1422	0.3483
1.0	IR	−35.1308	−19.1908	−11.5868	−7.1048	−4.2879	−0.5970	−0.0836	0.2212	0.7646
	WB	−10.7281	−8.1392	−6.4956	−5.1114	−3.5641	−0.5840	−0.0657	0.1414	0.4090
1.5	IR	−109.3002	−45.9664	−23.8466	−12.6629	−6.6720	−0.7046	−0.1236	0.4112	1.7195
	WB	−74.4056	−47.7300	−31.8655	−19.6662	−11.8565	−3.8242	−2.6779	−2.2846	−1.6962
2.0	IR	−233.4973	−89.9985	−44.8004	−20.2960	−8.5925	−0.7582	−0.0875	0.6209	3.3710
	WB	−50.7721	−39.5083	−30.3800	−23.5211	−17.6834	−6.9507	−3.7050	−2.4184	−1.1239

https://doi.org/10.1017/S0266466620000377 Published online by Cambridge University Press

TABLE A2. Empirical critical values based on the wild bootstrap and independent repetitions for $\omega = 1$, $a_1 = 0.6$, and $b_1 = 0.5$ in model (A.1)

β	Method	0.01	0.025	0.05	0.10	0.20	0.80	0.95	0.975	0.99
0.5	IR	−16.8492	−13.6800	−11.4395	−9.1815	−6.8084	−1.4716	−0.1268	0.3359	0.9455
	WB	−19.3680	−16.0606	−13.1001	−10.5251	−7.7374	−1.6686	−0.1465	0.3872	0.8033
1.0	IR	−19.4450	−15.3862	−12.5082	−9.6011	−6.9531	−1.3734	−0.1025	0.4337	1.0772
	WB	−18.9817	−15.5351	−12.9337	−10.2693	−7.3456	−1.4106	0.1259	0.6428	1.1011
1.5	IR	−46.9624	−27.7281	−18.6230	−12.7934	−8.1840	−1.1803	−0.0568	0.5489	1.6628
	WB	−135.0948	−78.1896	−53.6238	−31.4109	−14.8175	−0.0547	3.3009	6.3047	10.4250
2.0	IR	−143.1385	−65.2896	−34.5130	−18.8654	−9.7865	−1.1402	−0.1775	0.5342	2.4360
	WB	−134.3028	−84.8120	−53.4753	−27.1830	−13.3954	2.1294	10.1357	15.3513	22.9815

TABLE A3. The size for the test R_n^* with significant level 0.05

(a_1, b_1)	n	$\beta = 0.5$		$\beta = 1$		$\beta = 1.5$		$\beta = 2$	
		IR	*WB*	IR	*WB*	IR	*WB*	IR	*WB*
	300	0.068	0.064	0.062	0.132	0.055	0.541	0.075	0.551
(0.8, 0)	500	0.060	0.048	0.051	0.130	0.044	0.518	0.054	0.554
	1,000	0.064	0.057	0.057	0.136	0.048	0.526	0.049	0.517
	300	0.049	0.037	0.052	0.047	0.047	0.002	0.042	0.009
(0.6, 0.5)	500	0.071	0.047	0.073	0.064	0.044	0.010	0.047	0.022
	1,000	0.056	0.039	0.063	0.060	0.067	0.008	0.045	0.012

Based on the above regression, we can get the ADF estimates for ρ and a_1. The bias and standard deviation of these estimates over the 10,000 repetitions are reported in Table A4. It is shown in Table A4 that when little GARCH effect exists in the noise, both ρ and a_1 can be estimated well, which is consistent with CGT, where they showed that when the noise is derived by i.i.d. heavy-tailed innovations, then ρ and the autoregressive parameters in ADF regression can be estimated consistently. From Table A4, we also observe that if the tail index $\alpha < 2$, then as the GARCH effect increases, performance of the estimates worsens, especially the estimate of a_1, which is no longer consistent. This conforms with the results derived in our paper.

To gain some insight of Q_n, we plot the densities and distribution functions of Q_n for different tail indexes αs. We simulate model (A.1) with $(a_1, b_1) = (0.6, 0.5)$ and β is taken to be 0.7322959, 1, 1.222668, $\pi/2$, and 2.188439 when $\alpha = 3$, 2, 1.5, 1, and 0.5, respectively. We replicate the simulation 10,000 times with a sample size 5,000 and the lag order k is taken as the same as in the adf.test function in R. Based on 10,000 repetitions with a sample size $n = 5{,}000$, the densities and distribution functions of Q_n are plotted in Figure 1, which shows that as the tail index α gets smaller, the tail of Q_n becomes heavier.

A.2. Proofs of Lemmas

The proof of Lemma 5.3. Note that

$$\sum_{l=0\vee(j-i-h)}^{\infty} c_{i-j+l+h}c_l \left(\sum_{t=k}^{n-1} \varepsilon_{t-i-l}\varepsilon_{t-i-l-h} - \sum_{t=k}^{n-1} \varepsilon_t\varepsilon_{t-h} \right)$$
$$= \sum_{l=0\vee(j-i-h)}^{\infty} c_{i-j+l+h}c_l \left(\sum_{t=k-i-l}^{k-1} \varepsilon_t\varepsilon_{t-h} - \sum_{t=n-i-l}^{n-1} \varepsilon_t\varepsilon_{t-h} \right).$$

For the proof of Lemma 5.3, it is enough to show that

$$\left\| \sum_{i,j=1}^{k} \left[b_n^{-1} \sum_{h=1}^{H} \sum_{l=0\vee(j-i-h)}^{\infty} c_{i-j+l+h}c_l \left(\sum_{t=k-i-l}^{k-1} \varepsilon_t\varepsilon_{t-h} - \sum_{t=n-i-l}^{n-1} \varepsilon_t\varepsilon_{t-h} \right) \right]^2 \right\|_F = o_p(1), \tag{A.3}$$

https://doi.org/10.1017/S0266466620000377 Published online by Cambridge University Press

https://doi.org/10.1017/S0266466620000377 Published online by Cambridge University Press

TABLE A4. Bias and standard deviation for the estimates of ρ and a_1 in model (A.1) with i.i.d. stable innovations η_t

α	Est	$\beta = 0$		$\beta = 0.1$		$\beta = 0.2$		$\beta = 0.3$		$\beta = 0.4$		$\beta = 0.5$	
		Bias	*SD*	Bias	*SD*	Bias	*SD*	Bias	*SD*	Bias	*SD*	Bias	*SD*
1	$\widehat{\rho}$	−0.0004	0.0009	−0.0014	0.0131	−0.0018	0.0115	0.0708	6.8801	−0.0053	0.8261	0.0536	20.3655
	$\widehat{a}_1$	−0.0005	0.0118	−0.0840	0.3178	−0.1429	0.3542	−0.2601	6.8885	−0.2136	0.9150	−0.2959	20.3726
1.5	$\widehat{\rho}$	−0.0004	0.0004	−0.0005	0.0012	−0.0007	0.0067	−0.0009	0.0035	−0.0014	0.0312	−0.0012	0.0400
	$\widehat{a}_1$	−0.0004	0.0115	−0.0332	0.2091	−0.0783	0.2609	−0.1237	0.2923	−0.1649	0.3161	−0.2041	0.3305
2	$\widehat{\rho}$	−0.0004	0.0003	−0.0004	0.0003	−0.0004	0.0003	−0.0005	0.0004	−0.0006	0.0008	−0.0006	0.0014
	$\widehat{a}_1$	−0.0006	0.0142	−0.0010	0.0188	−0.0049	0.0434	−0.0277	0.0930	−0.0711	0.1524	−0.1163	0.2071

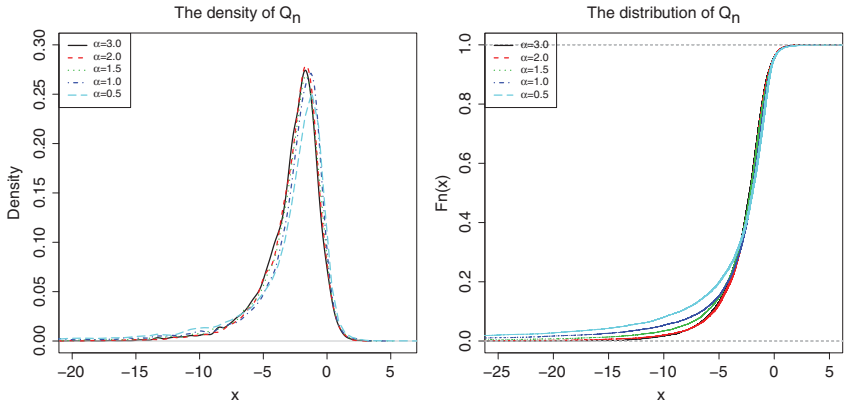


FIGURE A1. Probability densities and distribution functions of Q_n and $\widehat{\rho}_{\mu n}$ for various tail index values α.

and

$$\left\| \sum_{i,j=1}^{k} \left[b_n^{-1} \sum_{h=0}^{H} \sum_{l=0\vee(i-j-h)}^{\infty} c_{j-i+l+h} c_l \left(\sum_{t=k-j-l}^{k-1} \varepsilon_t \varepsilon_{t-h} - \sum_{t=n-j-l}^{n-1} \varepsilon_t \varepsilon_{t-h} \right) \right]^2 \right\|_F = o_p(1). \tag{A.4}$$

Since the proofs for (A.3) and (A.4) are similar, we only give the proof of (A.3) in detail. Let $\zeta_{t,h}^1 = \varepsilon_t \varepsilon_{t-h} I(|\sigma_t \varepsilon_{t-h}| \le b_n)$ and $\zeta_{t,h}^2 = \varepsilon_t \varepsilon_{t-h} I(|\sigma_t \varepsilon_{t-h}| > b_n)$. By the triangular inequality and Karamata's theorem, it follows that

$$\begin{aligned}
&\mathrm{E} \left| \sum_{i,j=1}^{k} \left[b_n^{-1} \sum_{h=1}^{H} \sum_{l=0\vee(j-i-h)}^{\infty} c_{i-j+l+h} c_l \sum_{t=k-i-l}^{k-1} \zeta_{t,h}^1 \right]^2 \right| \\
&\quad \le \sum_{i,j=1}^{k} \mathrm{E} \left[b_n^{-1} \sum_{h=1}^{H} \sum_{l=0\vee(j-i-h)}^{\infty} c_{i-j+l+h} c_l \sum_{t=k-i-l}^{k-1} \zeta_{t,h}^1 \right]^2 \\
&\quad \le C n^{-1} \sum_{i,j=1}^{k} \sum_{h=1}^{H} \left(\sum_{l=0\vee(j-i-h)}^{\infty} c_{i-j+l+h} c_l \right)^2 (i+l) \mathrm{E}(\zeta_{t,h}^1)^2 \\
&\quad = O(k^3 n^{-1}) = O(k^3 n^{-1}).
\end{aligned} \tag{A.5}$$

Further, by taking θ in Lemma 5.2 near enough to α as $\alpha \le 2$ and $\theta = 2$ for $\alpha > 2$, then by the triangular inequality and Karamata's theorem again,

$$\mathrm{E} \left| \sum_{i,j=1}^{k} \left[b_n^{-1} \sum_{h=1}^{H} \sum_{l=0\vee(j-i-h)}^{\infty} c_{i-j+l+h} c_l \sum_{t=k-i-l}^{k-1} \zeta_{t,h}^2 \right]^2 \right|^{\theta/4}$$

https://doi.org/10.1017/S0266466620000377 Published online by Cambridge University Press

$$\le \sum_{i,j=1}^{k} \mathrm{E}\left[b_n^{-1} \sum_{h=1}^{H} \sum_{l=0\vee(j-i-h)}^{\infty} c_{i-j+l+h} c_l \sum_{t=k-i-l}^{k-1} \zeta_{t,h}^2 \right]^{\theta/2}$$

$$\le \sum_{i,j=1}^{k} \sum_{h=1}^{H} \sum_{l=0\vee(j-i-h)}^{\infty} b_n^{-\theta/2} |c_{i-j+l+h} c_l|^{\theta/2} (i+l) \mathrm{E}|\zeta_{t,h}^2|^{\theta/2}$$

$$= O(k^3 H n^{-1}) = O(k^3 n^{-1} \log n). \tag{A.6}$$

By (A.5) and (A.6), we have as $k^3 n^{-1} \log n = o(1)$, then

$$\left\| \sum_{i,j=1}^{k} \left[b_n^{-1} \sum_{h=1}^{H} \sum_{l=0\vee(j-i-h)}^{\infty} c_{i-j+l+h} c_l \sum_{t=k-i-l}^{k-1} \varepsilon_t \varepsilon_{t-h} \right]^2 \right\|_F = o_p(1). \tag{A.7}$$

Similarly, we can show that

$$\left\| \sum_{i,j=1}^{k} \left[b_n^{-1} \sum_{h=1}^{H} \sum_{l=0\vee(j-i-h)}^{\infty} c_{i-j+l+h} c_l \sum_{t=n-i-l}^{n-1} \varepsilon_t \varepsilon_{t-h} \right]^2 \right\|_F = o_p(1),$$

which combining with (A.7) yields (A.3) and completes the proof of Lemma 5.3. □

The proof of Lemma 5.4. As the proof for $\alpha > 2$ is simpler than the other two cases and the proof of $\alpha = 2$ is similar to $\alpha < 2$, we only give the proof for $\alpha < 2$. Note that

$$\sum_{t=k}^{n-1} \boldsymbol{X}_t \boldsymbol{X}_t^T = \left(\sum_{t=k}^{n-1} \mu_{t-i} \mu_{t-j} \right)_{k\times k}$$

and for any $1 \le i,j \le k$,

$$\begin{aligned}
\sum_{t=k}^{n-1} \mu_{t-i} \mu_{t-j} &= \sum_{t=k}^{n-1} \sum_{l,s=0}^{\infty} c_s c_l \varepsilon_{t-i-s} \varepsilon_{t-j-l} \\
&= \sum_{t=k}^{n-1} \sum_{l=0\vee(i-j)}^{\infty} c_{j-i+l} c_l \varepsilon_{t-j-l}^2 \\
&\quad + \sum_{t=k}^{n-1} \sum_{l=0}^{\infty} \sum_{h=(j-i-l)\vee 1}^{\infty} c_{i-j+l+h} c_l \varepsilon_{t-i-l} \varepsilon_{t-i-l-h} \\
&\quad + \sum_{t=k}^{n-1} \sum_{l=0}^{\infty} \sum_{h=(i-j-l)\vee 1}^{\infty} c_{j-i+l+h} c_l \varepsilon_{t-j-l} \varepsilon_{t-j-l-h} \\
&= \sum_{t=k}^{n-1} \sum_{l=0\vee(i-j)}^{\infty} c_{j-i+l} c_l \varepsilon_{t-j-l}^2 \\
&\quad + \sum_{t=k}^{n-1} \sum_{l=0}^{\infty} \sum_{h=(j-i-l)\vee 1}^{H} c_{i-j+l+h} c_l \varepsilon_{t-i-l} \varepsilon_{t-i-l-h}
\end{aligned}$$

https://doi.org/10.1017/S0266466620000377 Published online by Cambridge University Press

$$
\begin{aligned}
&+\sum_{t=k}^{n-1}\sum_{l=0}^{\infty}\sum_{h=(i-j-l)\vee 1}^{H} c_{j-i+l+h}c_l\varepsilon_{t-j-l}\varepsilon_{t-j-l-h}\\
&+\sum_{t=k}^{n-1}\sum_{l=0}^{\infty}\sum_{h=(j-i-l)\vee (H+1)}^{\infty} c_{i-j+l+h}c_l\varepsilon_{t-i-l}\varepsilon_{t-i-l-h}\\
&+\sum_{t=k}^{n-1}\sum_{l=0}^{\infty}\sum_{h=(i-j-l)\vee (H+1)}^{\infty} c_{j-i+l+h}c_l\varepsilon_{t-j-l}\varepsilon_{t-j-l-h}\\
&=:\sum_{l=1}^{5} S_{nl}(i,j).
\end{aligned}
$$

By the definition of $I_l(n,H,i,j)$, $l=1,2$, we have

$$S_{n4}(i,j)=I_1(n,H,i,j)+I_2(n,H,i,j).$$

Since $k^3=o(n)$, it follows from Lemma 5.2 that when $H=-2\log n/(\delta\log\varrho)$,

$$
\begin{aligned}
\mathrm{E}\left(\sum_{i=1}^{k}\sum_{j=1}^{k}|b_n^{-1}I_1(n,H,i,j))|^2\right)^{\delta/2} &\le \sum_{i=1}^{k}\sum_{j=1}^{k}\mathrm{E}|b_n^{-1}I_1(n,H,i,j))|^{\delta}\\
&= O(k^2\varrho^{\delta H})=O(k^2n^{-2})=O(n^{-1}). \qquad \textbf{(A.8)}
\end{aligned}
$$

Further, by taking θ as in showing (A.6),

$$
\begin{aligned}
\mathrm{E}\left(\sum_{i=1}^{k}\sum_{j=1}^{k}|b_n^{-1}I_2(n,H,i,j))|^2\right)^{\theta/2} &\le \sum_{i=1}^{k}\sum_{j=1}^{k}\mathrm{E}|b_n^{-1}I_2(n,H,i,j))|^{\theta}\\
&\le Ck^2nb_n^{-\theta}=o(1), \qquad \textbf{(A.9)}
\end{aligned}
$$

since $k^3\log n=o(n)$. Let A_{n4} be the matrix with elements with $b_n^{-1}(S_{n4}(i,j)$, that is,

$$A_{n4}=\left(b_n^{-1}S_{n4}(i,j)\right)_{k\times k}.$$

Thus, combining (A.8) and (A.9), we have if $k=O(n^a)$ for some $a<1/2$, then

$$\|A_{n4}\|_F=o_p(1).$$

Similarly, for the matrix with elements with $\{b_n^{-1}S_{n5}(i,j)\}_{ij}$, say A_{n5}, we also have

$$\|A_{n5}\|_F=o_p(1).$$

Thus, by Lemma 5.3,

$$b_n^{-1}\left\|\sum_{t=k}^{n-1}\boldsymbol{X}_t\boldsymbol{X}_t^T-\sum_{h=0}^{H}D_h\left(\sum_{t=k}^{n-1}\varepsilon_t\varepsilon_{t-h}\right)\right\|_F=o_p(1) \qquad \textbf{(A.10)}$$

https://doi.org/10.1017/S0266466620000377 Published online by Cambridge University Press

holds for $H = -2\log n/(\delta \log \varrho)$. It remains to prove that for any $\nu > 0$,

$$\lim_{M\to\infty}\lim_{n\to\infty} P\left\{ b_n^{-1}\left\| \sum_{h=M+1}^{H} D_h \left(\sum_{t=k}^{n-1} \varepsilon_t \varepsilon_{t-h} \right) \right\| > \nu \right\} = 0. \tag{A.11}$$

Since $\{D_h\}$ is a sequence of symmetric Toeplitz matrices with $\sup_h \sum_{s=0}^{\infty}(|d(s,h)| + |d(-s,h)|) < \infty$, it follows that

$\sup_h \|D_h\| < \infty$.

Hence,

$$\begin{aligned} b_n^{-1}\left\| \sum_{h=M+1}^{H} D_h \left(\sum_{t=k}^{n-1} \varepsilon_t \varepsilon_{t-h} \right) \right\| &\le \left(\sup_h \|D_h\|\right) \sum_{h=M+1}^{H} \left| b_n^{-1} \sum_{t=k}^{n-1} \varepsilon_t \varepsilon_{t-h} \right| \\ &\le C \sum_{h=M+1}^{H} \left| b_n^{-1} \sum_{t=k}^{n-1} \varepsilon_t \varepsilon_{t-h} \right|. \end{aligned} \tag{A.12}$$

By Lemma 5.2, we have

$$\lim_{M\to\infty}\lim_{n\to\infty} P\left\{ \sum_{h=M+1}^{H} \left| b_n^{-1} \sum_{t=k}^{n-1} \varepsilon_t \varepsilon_{t-h} \right| > \nu \right\} = 0.$$

Thus, by (A.12), we have (A.11) and complete the proof. □

The Proof of (5.4). Let $z_{t,l,h}(y) = \varepsilon_{t-l-h}\varepsilon_{t-l} I(|\varepsilon_{t-l-h}\varepsilon_{t-l}| \le a_n^2 y)$. To apply Prop. 3.3 of Davis and Mikosch (1998) for the case of $2 \le \alpha < 4$, we need to show that for any given integer l and $h = 0, 1, \ldots, M$, and for all $\delta > 0$,

$$\lim_{y\to 0} \limsup_{n\to\infty} P\left\{ \left| \sum_{t=1}^{n} (z_{t,l,h}(y) - \mathrm{E} z_{t,l,h}(y)) \right| > \delta a_n^2 \right\} = 0. \tag{A.13}$$

Since $\{\eta_t\}$ are symmetric random variables, then

$$\begin{aligned} \mathrm{E}\left(a_n^{-2} \sum_{t=1}^{n} (z_{t,l,h}(y) - \mathrm{E} z_{t,l,h}(y)) \right)^2 &= a_n^{-4} \sum_{i=1}^{n} \mathrm{E} z_{i,l,h}^2(y) \\ &\le C a_n^{-4} n (a_n^2 y)^{2-\alpha/2} \\ &= O(y^{2-\alpha/2}) \to 0, \end{aligned}$$

as $y \to 0$, which implies (A.13) as desired. □

https://doi.org/10.1017/S0266466620000377 Published online by Cambridge University Press