Statistica Sinica **20** (2010), 1363-1393

INFERENCE FOR UNIT-ROOT MODELS WITH INFINITE VARIANCE GARCH ERRORS

Ngai Hang Chan and Rong-Mao Zhang

Chinese University of Hong Kong and Zhejiang University

Abstract: Random walk models driven by GARCH errors are widely applicable in diverse areas in finance and econometrics. For a first-order autoregressive model driven by GARCH errors, let $\widehat{\phi}_n$ be the least squares estimate of the autoregressive coefficient. The asymptotic distribution of $\widehat{\phi}_n$ is given in Ling and Li (2003) when the GARCH errors have finite variances. In this paper, the limit distribution of $\widehat{\phi}_n$ is established as functionals of a stable process when the GARCH errors are heavy-tailed with infinite variances. An estimate of the tail index of the limiting stable process is proposed and its asymptotic properties are derived. Furthermore, it is shown that the least absolute deviations procedure works well under the unit-root and heavy-tailed GARCH setting. This research provides a relatively broad treatment of unit-root GARCH models that includes the commonly entertained unit-root IGARCH scenario.

Key words and phrases: Autoregressive process, GARCH, heavy-tailed, IGARCH, stable processes and unit-root.

1. Introduction

Consider the models

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

$$Y_t = \mu + \phi Y_{t-1} + \varepsilon_t, \tag{1.2}$$

where $Y_0 = 0$ and the ε_t follow a first-order generalized autoregressive conditional heteroscedasticity model (GARCH(1, 1))

$$\varepsilon_t = \sigma_t \eta_t, \quad \sigma_t \geq 0\,, \tag{1.3}$$

$$\sigma_t^2 = \omega + a\sigma_{t-1}^2 + b\varepsilon_{t-1}^2,\ \omega > 0,\ a \geq 0,\ b \geq 0\,, \tag{1.4}$$

where $\{\eta_t\}$ are i.i.d. symmetric random variables.

There is an extensive literature on unit-root estimation and testing for the case $a = b = 0$, i.e., $\{\varepsilon_t\}$ are i.i.d. random variables. For a concise review on the recent developments on this topic, see Chan (2009) and the references therein.

The unit-root problem for the case of non i.i.d. errors ($a \neq 0$ or $b \neq 0$) has also been receiving considerable attention in the literature. One of the main reasons for this is that when Y_t in model (1.1) represents the log price of an underlying asset, then it is often found that the return process ε_t follows a GARCH$(1,1)$ model as prescribed by (1.3) and (1.4), see for example Bernard et al. (2008). Under these circumstances, the original testing for unit-root in (1.1) is tantamount to testing for unit-root with GARCH$(1,1)$ errors. Motivated by this consideration, extensive research have been conducted. For example, Weiss (1986) studied the distribution of quasi-maximum likelihood estimation (QMLE) when $\mathrm{E}\varepsilon_t^4 < \infty$ for ARCH models. 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) and Seo (1999) 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) and Ling, Li and McAleer (2003) generalized the results to the case $\mathrm{E}\varepsilon_t^2 < \infty$ but $\mathrm{E}\eta_t^4 < \infty$, and obtained the limit distribution of the estimated unit-root as a functional of the Brownian motion. Chan and Peng (2005) studied least absolute deviations estimation for the AR(1) process with $a = 0$ and $\mathrm{E}\eta_t^2 < \infty$. Recently, Wang (2006) studied the asymptotic distribution of the Dickey-Fuller test under $\mathrm{E}\varepsilon_t^2 < \infty$, i.e., $a + b\mathrm{E}\eta_t^2 < 1$.

Although the GARCH error model enjoys tremendous popularity in modeling stock returns, one critical issue remains. In fitting the log returns to (1.4), it is often reported in the data that the estimates of the parameter $a + b\mathrm{E}\eta_t^2$ are very close to unity, see Mikosch and Stărică (2000). In this case, the model exhibits the so-called IGARCH effect where $\mathrm{E}\varepsilon_t^2 = \infty$. As a result, many of the previously developed theories are not applicable. One of the main purposes of this paper is to study the asymptotic distribution of unit-root estimators when $\{\varepsilon_t\}$ is a GARCH$(1,1)$ process with infinite variance. In particular, it is shown that when $a + b\mathrm{E}\eta_t^2 > 1$, the asymptotic distribution of the LSE converges to a functional of a stable process, and when $a + b\mathrm{E}\eta_t^2 = 1$, the asymptotic distribution of the LSE converges to a functional of the Brownian motion, similar to the case when $a + b\mathrm{E}\eta_t^2 < 1$.

Another objective of this paper is to study the asymptotic properties of the least absolute deviations estimators of the parameters a and b in (1.4) when $\mathrm{E}\varepsilon_t^2 = \infty$, extending results of Peng and Yao (2003). By completing the asymptotic theory for both finite and infinite variance scenarios, we offer a relatively broad coverage of the unit-root problem for the AR model driven by GARCH errors, including the commonly entertained IGARCH case.

The paper is organized as follows. Section 2 gives the main results. As the limit process depends on an unknown parameter of the tail index, estimation of parameters and the corresponding limit distributions are given in Section 3. Properties of the least absolute deviations estimators of the GARCH parameters are given in Section 4. Simulations are reported in Section 5. Section 6 presents the proofs of the main theorems. Preliminary lemmas and a critical result of the weak convergence of a stable process for strongly mixing sequence are relegated to Appendix. In the sequel, we use the symbol C to denote an unspecified positive and finite constant that may take a different value at each appearance.

2. Asymptotic Distributions

Given $Y_0 = 0$ and observations $Y_1, \dots Y_n$, to test $\phi = 1$ against $\phi < 1$, the Dickey-Fuller (DF) test $\widehat{\rho}_n$ based on least squares (LS) regression of Y_t on Y_{t-1} for model (1.1) is

$$\widehat{\rho}_n = n(\widehat{\phi}_n - 1) = \Big(\frac{1}{n}\sum_{i=1}^{n} Y_{i-1}^2\Big)^{-1}\Big(\sum_{i=1}^{n} Y_{i-1}\varepsilon_i\Big). \tag{2.1}$$

Similarly, the unit-root statistic $\widehat{\rho}_{\mu n}$ for model (1.2) when there exists a drift in the autoregressive model is

$$\begin{aligned}\widehat{\rho}_{\mu n} = n(\widehat{\phi}_{\mu n} - 1) &= \Big(\frac{1}{n}\sum_{i=1}^{n}(Y_{i-1} - \overline{Y})^2\Big)^{-1}\Big(\sum_{i=1}^{n}(Y_{i-1} - \overline{Y})\varepsilon_i\Big)\\ &= \Big(\frac{1}{n}\sum_{i=1}^{n} Y_{i-1}^2 - (\overline{Y})^2\Big)^{-1}\Big(\sum_{i=1}^{n} Y_{i-1}\varepsilon_i - \overline{Y}\sum_{i=1}^{n}\varepsilon_i\Big),\end{aligned} \tag{2.2}$$

where $\overline{Y} = \sum_{i=1}^{n} Y_{i-1}/n$. We impose the following assumptions.

H1. $\mathrm{E}\log(a + b\eta_t^2) < 0$.

H2. 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 $\log^+(x) = \max\{0, \log(x)\}$.

H3. The distribution F of η_1 is a mixture of an absolutely continuous component with respect to the Lebesgue measure λ on R and Dirac masses at some points $\mu_i \in R, i = 1, \dots, N$, that is,

$$dF = \sum_{i=1}^{N} p_i d\,\delta_{\mu_i} + (1-p) f\, d\lambda,\ \ p_i \geq 0,\ \ \sum_{i=1}^{N} p_i = p \in [0, 1),$$

where f is a density of continuous component and satisfying

$$(x_-^0-\delta, x_-^0)\cup(x_+^0, x_+^0+\delta)\subset\{f>0\} \tag{2.3}$$

for some $\delta>0$ and $x_-^0=\sup\{x|x<0, f(x)>0\}$, $x_+^0=\inf\{x|x>0, f(x)>0\}$.

The first result gives weak convergence for the partial sum process of the sequence $\{\varepsilon_i\}$. Herein, the symbol $\xrightarrow{f.d.d.}$ denotes the convergence of the finite-dimension distribution in $D[0,1]$. Moreover, the symbol $\Rightarrow$ denotes the weak convergence in $D[0,1]$ (the space of functions on $[0,1]$ which are right-continuous and have left-hand limits, see Billingsley (1999)).

Theorem 2.1. *Under the conditions* H1, H2 *and* H3*, the following hold.*

(a) *There exists a unique* $\alpha\in(0,k_0]$ *such that* $\mathrm{E}(a+b\eta_1^2)^\alpha=1$.

(b) *If* $\alpha\in(0,1)$*, then*

$$\Big(\frac{1}{(c_1n)^{1/(2\alpha)}}\sum_{i=1}^{[nt]}\varepsilon_i, \frac{1}{(c_1n)^{1/\alpha}}\sum_{i=1}^{[nt]}\varepsilon_i^2\Big)\xrightarrow{f.d.d.}\Big(Z_{2\alpha}(t), Z_\alpha(t)\Big),$$

where $Z_{2\alpha}(t)$ *is a stable process with index* 2α *and* $c_1=c_0\mathrm{E}|\eta_1|^{2\alpha}$*, where* c_0 *is a constant to be defined in Lemma* A.1.

(c) *If* $\alpha=1$*, then*

$$\frac{1}{\sqrt{c_1n\log n}}\sum_{i=1}^{[nt]}\varepsilon_i\Rightarrow W(t),$$

where $c_1=c_0\mathrm{E}\eta_1^2=: c_0\sigma^2$ *and* $\{W(t), 0\le t\le 1\}$ *is a standard Brownian motion.*

Applying Theorem 2.1 yields the asymptotic distributions of $\widehat{\rho}_n$ and $\widehat{\rho}_{\mu n}$.

Theorem 2.2. *Suppose that* $\phi=1$ *in model* (1.1) *or* $(\phi,\mu)=(1,0)$ *in model* (1.2). *Under the conditions of Theorem* 2.1*, the following hold.*

(a) *For* $\alpha\in(0,1)$,

$$\widehat{\rho}_n = n(\widehat{\phi}_n-1)\xrightarrow{d}\frac{\int_0^1 Z_{2\alpha}^-(t)\,dZ_{2\alpha}(t)}{\int_0^1 Z_{2\alpha}^2(t)\,dt}, \tag{2.4}$$

$$\begin{aligned}\widehat{\rho}_{\mu n} &= n(\widehat{\phi}_{\mu n}-1)\\ &\xrightarrow{d}\frac{\int_0^1 Z_{2\alpha}^-(t)\,dZ_{2\alpha}(t)-Z_{2\alpha}(1)\int_0^1 Z_{2\alpha}(t)\,dt}{\int_0^1 Z_{2\alpha}^2(t)\,dt-(\int_0^1 Z_{2\alpha}(t)\,dt)^2},\end{aligned} \tag{2.5}$$

where $Z_{2\alpha}^-(t)$ *denotes the left-hand limit of* $Z_{2\alpha}(t)$ *and* $\xrightarrow{d}$ *denotes convergence in distribution.*

(b) *For* $\alpha \geq 1$,

$$\widehat{\rho}_n = n(\widehat{\phi}_n - 1) \xrightarrow{d} \frac{\int_0^1 W(t)\, dW(t)}{\int_0^1 W^2(t)\, dt} \tag{2.6}$$

$$\rho_{\mu n} = n(\phi_{\mu n} - 1) \xrightarrow{d} \frac{\int_0^1 W(t)\, dW(t) - W(1)\int_0^1 W(t)\, dt}{\int_0^1 W^2(t)\, dt - (\int_0^1 W(t)\, dt)^2}. \tag{2.7}$$

Remark 2.1. Condition H1 is a necessary and sufficient condition for the existence of stationary solution of σ_t^2 (see Nelson (1990)). If condition (H2) holds, then condition (H1) is equivalent to $\mathrm{E}(a + b\eta_1^2)^\mu < 1$ for some $\mu > 0$ (see Remark 2.9 of Basrak, Davis and Mikosch (2002)).

Remark 2.2. If there exists $h_0 > 0$ such that $\mathrm{E}|\eta_t|^{h_0} = \infty$ and $\mathrm{E}|\eta_t|^h < \infty$ for all $h < h_0$, then conditions H1 and H2 are satisfied.

Remark 2.3. If $f(x)$ is positive in a neighborhood of zero, then (2.3) is true with $x_-^0 = x_+^0 = 0$. Condition H3 is the weakest condition among the existing results on mixing conditions for GARCH(1, 1) process. For more information, refer to Francq and Zakoïan (2006).

Remark 2.4. If $\alpha > 1$, by condition H2, $\mathrm{E}|\eta_1|^{2\alpha} < \infty$ and as a result, $\mathrm{E}\eta_1^2 < \infty$. By virtue of Hölder's inequality, $\mathrm{E}(a + b\eta_1^2) \leq [\mathrm{E}(a + b\eta_1^2)^\alpha]^{1/\alpha} = 1$. Since η_1^2 is non-degenerate, we have $\mathrm{E}(a + b\eta_1^2) < 1$. When $\mathrm{E}\eta_1^2 = 1$ and $a + b < 1$, the asymptotic distribution of Theorem 2.2 is given in Wang (2006). Therefore, for the proof of Theorem 2.2, it is enough to show the case that $\alpha = 1$.

Remark 2.5. The limit distributions of Theorem 2.2 are the same as those in Chan and Tran (1989), in which the errors $\{\varepsilon_t\}$ are assumed to be i.i.d. with infinite variance.

Remark 2.6. To apply Theorem 2.1, one needs to estimate c_0. By its definition in Lemma A.1, it can be estimated by

$$\frac{n^{-1}\sum_{t=1}^n \Big([\widehat{\omega} + (\widehat{a} + \widehat{b}\widehat{\eta}_t^2)\widehat{\sigma}_t^2]^{\widehat{\alpha}} - [(\widehat{a} + \widehat{b}\widehat{\eta}_t^2)\widehat{\sigma}_t^2]^{\widehat{\alpha}}\Big)}{\widehat{\alpha} n^{-1}\sum_{t=1}^n \Big((\widehat{a} + \widehat{b}\widehat{\eta}_t^2)^{\widehat{\alpha}} \log^+(\widehat{a} + \widehat{b}\widehat{\eta}_t^2)\Big)},$$

where $\widehat{\omega}$, $\widehat{a}$, $\widehat{b}$, $\widehat{\alpha}$ are consistent estimators of ω, a, b, α, and $\widehat{\sigma}_t^2 = \widehat{\omega} + (\widehat{a} + \widehat{b}\widehat{\eta}_{t-1}^2)\widehat{\sigma}_{t-1}^2$.

3. Hill Estimators

Note that the limit partial sum process of ε is a stable process with index 2α, which is unknown a priori. To apply Theorem 2.1, we need to estimate α. In this section, we construct a Hill estimator for α and study its asymptotic property based on the empirical residuals $\widehat{\varepsilon}_i = Y_i - \widehat{\phi} Y_{i-1}, i = 1, \ldots, n$, where $\widehat{\phi}$ represents the LSE $\widehat{\phi}_n$ or $\widehat{\phi}_{\mu n}$.

Let $\Lambda = 1/(2\alpha)$ be the tail index. Let $\max_{1\le i\le n} \widehat{\varepsilon}_i = \widehat{\varepsilon}_{n:n} \ge \widehat{\varepsilon}_{n-1:n} \ge \cdots \ge \widehat{\varepsilon}_{n-k_n:n}$ be the $k_n + 1$ largest order statistics. Estimate Λ by

$$\widehat{\Lambda} = \frac{1}{k_n} \sum_{i=1}^{k_n} \log \frac{\widehat{\varepsilon}_{n-i-1:n}}{\widehat{\varepsilon}_{n-k_n:n}}.$$

Theorem 3.1. *Suppose the conditions of Theorem* 2.1 *hold and that there exists some* $\beta > 0$ *such that*

$$P(\varepsilon_t > x) = c_0 x^{-2\alpha} (\mathrm{E}\eta_t^{2\alpha})(1 + O(x^{-\beta})), \quad x \to \infty. \tag{3.1}$$

For $\alpha > 1/2$, $\log^2 n \log^4(\log n) = o(k_n)$ *and* $k_n = o(\min\{n^{\beta/(\beta+\alpha)}, n^{2\alpha/(1+\alpha)}\})$, *we have*

$$\sqrt{k_n}(\widehat{\Lambda} - \Lambda) \xrightarrow{d} \mathrm{N}(0, \Sigma^2),$$

where $\Sigma^2 = \Lambda^2 \mathrm{Var}[\sum_{i=1}^n (I(\varepsilon_i > F^{-1}(1 - k_n/n)) - k_n/n)]/k_n$, *and* F *is the distribution of* ε_1.

Remark 3.1. For the asymptotic normality of the Hill estimator, the second order regularly varying condition is necessary. This condition is satisfied for a GARCH model when (i) $\mathrm{E}(a+b\eta_1^2)^{\alpha+\delta} < \infty$, and (ii) $g(\mu) = \mathrm{E}(a+b\eta_1^2)^{\alpha+i\mu}, \mu \in \mathbb{C}$ is analytic in a neighborhood of $\mu = 0$ and $g(\mu) \neq 1$ for $-\delta < Im(\mu) < 0$ (η_1 is normal, for example). In fact, apart from a constant, Goldie (1991) shows that under conditions (i) and (ii), σ satisfies the second order regularly varying condition (3.1). Since $\varepsilon_t = \sigma_t \eta_t$, it follows that under the same condition, ε_t also has the second order regularly varying property.

Remark 3.2. To apply Theorem 3.1, it is necessary to give an estimate for the variance Σ^2. To this end, we first estimate F by the empirical distribution $F_n(x) = \sum_{i=1}^b I(\widehat{\varepsilon}_i \le x)/n$ based on the residuals $\{\widehat{\varepsilon}_i\}$. We suggest running m replications $\{\widehat{\varepsilon}_i^{(j)}, 1 \le i \le n, 1 \le j \le m\}$ for the residuals $\{\widehat{\varepsilon}_i, 1 \le i \le n\}$, and estimating Σ^2 by

$$\widehat{\Sigma}^2 = \widehat{\Lambda}^2 m^{-1} \sum_{j=1}^m \left[\sum_{i=1}^n \left[I(\widehat{\varepsilon}_i^{(j)} > F_n^{-1}(1 - \frac{k_n}{n})) - \frac{k_n}{n} \right] (k_n)^{-1} \right]^2.$$

Remark 3.3. Another related problem to the Hill estimator is selecting the binwidth k_n. A natural choice would be the k_n that minimizes the mean squared error of the estimate $\widehat{\Lambda}$, $k_n = \text{argmin}_k \text{E}[\widehat{\Lambda} - \Lambda]^2$. Since Λ is unknown, we suggest using the bootstrap procedure of Danielsson et al. (2001) to select k_n. Specifically, let $n_1 < n$ and $\mathfrak{X}^*_{n_1} = \{X^*_1, \ldots, X^*_{n_1}\}$ be resamples drawn from $\mathfrak{X}_n = \{X_1, \ldots, X_n\}$ with replacement. Let $X^*_{1,n_1} \leq X^*_{2,n_1} \leq X^*_{n_1,n_1}$ denote the order statistics of $\mathfrak{X}^*_{n_1}$, and define

$$\widehat{\Lambda}^*_{n_1}(k) := \frac{1}{k}\sum_{i=1}^{k} \frac{\log X^*_{n_1-i+1,n_1}}{X^*_{n_1-k,n_1}},\ M^*_{n_1}(k) = \frac{1}{k}\sum_{i=1}^{k}\Big(\frac{\log X^*_{n_1-i-1,n_1}}{X^*_{n_1-k,n_1}}\Big)^2.$$

Then let $\widehat{k}_n = \text{argmin}_k \text{E}((M^*_{n_1}(k) - 2\widehat{\Lambda}^*_{n_1}(k)^2)^2 | \mathfrak{X}_n)$. It is shown in Danielsson et al. (2001) that such a $\widehat{k}_n$ is consistent and asymptotically optimal in terms of attaining a minimum mean squared error.

4. Estimation of $\boldsymbol{\nu_0 = (\omega_0, a_0, b_0)}$

To fit the model, we need to estimate the parameters $\nu_0 = (\omega_0, a_0, b_0)$ prescribed in (1.4) from the data. Although there is an extensive literature on this topic, most of it focuses on cases with stationary data. When models (1.3) and (1.4) are entertained, we are dealing with unit-root nonstationarity compounded with heavy-tailedness. It is therefore interesting to investigate to what extent the unit-root and/or the heavy-tailedness affects the asymptotic properties of the estimates of a and b.

To estimate the parameters of a GARCH model, QMLE is usually applied and this requires the existence of the fourth moment of the error η_t and the second moment of ε_t in model (1.4). But when the unit-root is present and IGARCH is entertained, these moment assumptions fail to hold. Other estimation procedures should then be used. One commonly used alternative is the least absolute deviations estimator. LAD is studied in Peng and Yao (2003) when $\text{E}\varepsilon_t^2 < \infty$ and $\text{E}\varepsilon_t^4 = \infty$. No result for the LAD, however, seems to be available for the case of an IGARCH model when $\text{E}\varepsilon_t^2 = \infty$.

To this end, let $\varepsilon_i(z) = Y_i - zY_{i-1}$, $h_i(\nu, z) = \sigma_i^2(\omega, a, b, z) = \omega + a\sigma_{i-1}^2(\omega, a, b, z) + b\varepsilon_{i-1}(z)^2$, and set $\widehat{\varepsilon}_i = \varepsilon_i(\widehat{\phi})$, where $\widehat{\phi}$ is the LSE $\widehat{\phi}_n$ in model (1.1) and $\widehat{\phi}_{\mu n}$ in model (1.2). Let Θ be the parameter space of (ω, a, b) satisfying $a < 1$ and H_1. We estimate ν_0 by the LAD estimator

$$\widehat{\nu} = \arg\min_{\nu\in\Theta} \sum_{i=1}^{n} \Big(\Big|\frac{\widehat{\varepsilon}_i}{\sqrt{h_i(\nu, \widehat{\phi})}}\Big| + \frac{1}{2}\log h_i(\nu, \widehat{\phi})\Big), \tag{4.1}$$

and establish its limit distributions when $\text{E}\varepsilon_t^2 = \infty$, which is encompassed by the case $\text{E}(a + b\eta_t^2) \geq 1$.

Theorem 4.1. *Suppose that* (i) $\phi = 1$ *in model* (1.1) *and* $(\phi, \mu) = (1, 0)$ *in model* (1.2); (ii) $\mathrm{E}|\eta_t| = 1$ *and* $\mathrm{E}\eta_t^2 < \infty$; (iii) Θ *is a compact set containing* ν_0. *Then we have the following.*

(1) *If* $\mathrm{E}(a_0 + b_0\eta_t^2)^\alpha = 1$ *for some* $\alpha > 1/2$, *then* $\widehat{\nu} \xrightarrow{p} \nu_0$.

(2) *If* $\mathrm{E}(a_0 + b_0\eta_t^2)^\alpha = 1$ *for some* $\alpha > 1$, *then*

$$\sqrt{n}(\widehat{\nu} - \nu_0) \xrightarrow{d} 2(\mathrm{N}(0, A) - F\zeta)/\mathrm{E}[h^{-2}(\nu_0)[h'(\nu_0)]^T h'(\nu_0)], \qquad (4.2)$$

where $A = \mathrm{E}(|\eta_i| - 1)^2\mathrm{E}[h_i^{-2}(\nu_0)[h_i'(\nu_0)]^T h_i'(\nu_0)]$, $h_i'(\nu) = \partial h_i(\nu)/\partial\nu$, $h_i(\nu) := h_i(\nu, 1)$, $F = \mathrm{E}\{(\sum_{j=1}^\infty b_0 a_0^{j-1}\varepsilon_{i-j})\ [h_i'(\nu_0)]^T h_i^{-2}(\nu_0)\}$, *and*

$$\zeta = \begin{cases} \frac{\int_0^1 W(t)dW(t) \int_0^1 W(t)dt}{\int_0 W^2(t)dt}, & \text{for model (1.1)}, \\ \frac{(\int_0^1 W(t)dW(t) - W(1)\int_0^1 W(t)dt)\int_0^1 W(t)dt}{\int_0^1 W^2(t)dt - (\int_0^1 W(t)dt)^2}, & \text{for model (1.2)}. \end{cases}$$

Remark 4.1. When $\alpha \le 1$, to establish the asymptotic distribution of $\widehat{\nu}$ defined in (4.1), the convergence rate of $\widehat{\phi}$ has to be $n^{-1/2-1/(2\alpha)}$ for $\alpha < 1$ and $(n\log n)^{-1}$ for $\alpha = 1$. But according to Theorem 2.2, the convergence rate of the $\widehat{\phi}$ is only n^{-1}. This difficulty can be resolved if we estimate (ν, ϕ) simultaneously as

$$(\widehat{\nu}, \widehat{\phi}) = \arg\min_{(\nu,\phi)} \sum_{i=1}^n \Big(\Big|\frac{\varepsilon_i(\phi)}{\sqrt{h_i(\nu, \phi)}}\Big| + \frac{1}{2}\log h_i(\nu, \phi)\Big).$$

5. Simulations

Numerical simulations were conducted to demonstrate the effectiveness of the asymptotic results. Tables 1 lists the empirical percentiles of the limit distribution of $\widehat{\rho}_n$ and $\widehat{\rho}_{\mu n}$ in Theorem 2.2, where the data was simulated from model (1.1) and (1.2) with $n = 1,000$, $\omega = 0.5, b = 0.7$, $a = 0.3$, 0.5 and 0.1, and $\{\eta_t\}$ were i.i.d. standard normal. Five thousand independent samples were generated for each simulated series. Since $\{\eta_t\}$ was normal, bigger a resulted in a smaller tail index α. Table 1 shows that the smaller is the tail index α, the heavier is the tail of the density of $\widehat{\rho}_n$ and $\widehat{\rho}_{\mu n}$. Further, when $\alpha < 1$ ($a = 0.5$), the thickness of the tail was more pronounced. When $\alpha \ge 1$ ($a = 0.1$ and 0.3), not only were the tails of the limit distribution thinner, but the changes of the tails were also becoming less pronounced. These phenomena can be explained by virtue of Theorem 2.2. As $\alpha < 1$, the limit distribution of $\widehat{\rho}_n$ or $\widehat{\rho}_{\mu n}$ is a functional of an integral of stable processes, but when $\alpha \ge 1$, the limit distribution is a functional

Table 1. Empirical percentiles of $\widehat{\rho}_n$, $\widehat{\rho}_{\mu n}$ with $\omega = 0.5$, $b = 0.7$.

	a	Probability of a smaller value									
		0.01	0.025	0.05	0.10	0.25	0.75	0.90	0.95	0.975	0.99
$\widehat{\rho}_n$	0.3	-23.25	-15.320	-11.28	-7.06	-3.09	0.30	0.93	1.34	1.700	2.26
	0.5	-102.11	-59.840	-36.25	-20.08	-6.46	0.30	1.06	2.07	4.430	12.72
	0.1	-15.75	-11.080	-8.33	-5.88	-2.93	0.23	0.93	1.27	1.590	2.08
$\widehat{\rho}_{\mu n}$	0.3	-37.28	-27.410	-19.80	-14.68	-8.72	-2.23	-0.83	-0.06	0.610	1.71
	0.5	-112.85	-71.780	-49.92	-29.69	-12.67	-1.81	-0.12	1.42	5.030	13.19
	0.1	-25.47	-19.570	-15.79	-12.32	-7.78	-2.18	-0.72	-0.01	0.560	1.21

Table 2. Bias, variance, and MSE of $\widehat{\Lambda}$ for different k_n when $\omega = 0.5, a = 0.3, b = 0.7$.

$\widehat{\lambda}$	Number (k_n) of the order statistics										
	50	100	150	200	250	300	350	400	450	500	550
Bias	0.028	0.007	0.030	0.053	0.078	0.104	0.136	0.171	0.212	0.259	0.316
Var	0.027	0.017	0.012	0.009	0.008	0.007	0.006	0.005	0.004	0.004	0.004
MSE	0.028	0.017	0.013	0.012	0.014	0.018	0.024	0.034	0.049	0.071	0.103

of an integral of Brownian motions only. These results are similar to those of Chan and Tran (1989).

To gain a further understanding of these phenomena, the probability density functions of $\widehat{\rho}_n$ and $\widehat{\rho}_{\mu n}$ of the three cases reported in Table 1 are plotted in Figure 1.

To shed some light on how the number (k_n) affects the Hill estimator ($\widehat{\Lambda}$), we simulated model (1.1) with $n = 1,000$ and $2,000$ repetitions with normal $\{\eta_t\}$. Table 2 gives the bias, the variance and the mean squared error (MSE) of $\widehat{\Lambda}$ for various k_n when $\omega = 0.5, a = 0.3$, and $b = 0.7$. Table 2 shows that $\widehat{\Lambda}$ is very sensitive to k_n. To obtain a more robust $\widehat{\Lambda}$, k_n cannot be too small or too large. When k_n is too small, it results in a small bias but a big variance; when k_n is too big, it leads to a small variance but a big bias. Therefore, adequately choosing k_n is important in using the Hill estimator.

To assess the effect of k_n, the graphs of the bias (filled square), the variance (filled triangle) and the MSE (filled circle) of $\widehat{\Lambda}(k_n)$ for different values of k_n are plotted in the left-hand panels in Figures 2–4. The marked dot on the MSE curve corresponds to the selected k_n when minimum MSE was achieved. For such a selected k_n, the estimated probability density function (solid line) and the asymptotic normal limit (dashed line) are plotted in the right-hand panels of the same figures. Figure 2 plots the case $\omega = 0.5, a = 0.3$, and $b = 0.7$ ($\alpha = 1$), Figure 3 is for the case $\omega = 0.5, a = 0.4$, and $b = 0.7$ ($\alpha < 1$), and Figure 4 is for the case $\omega = 0.5, a = 0.1$, and $b = 0.7$ ($\alpha > 1$). These graphs show the bias-variance trade off in k_n. Comparing the three cases, we see that if the tail

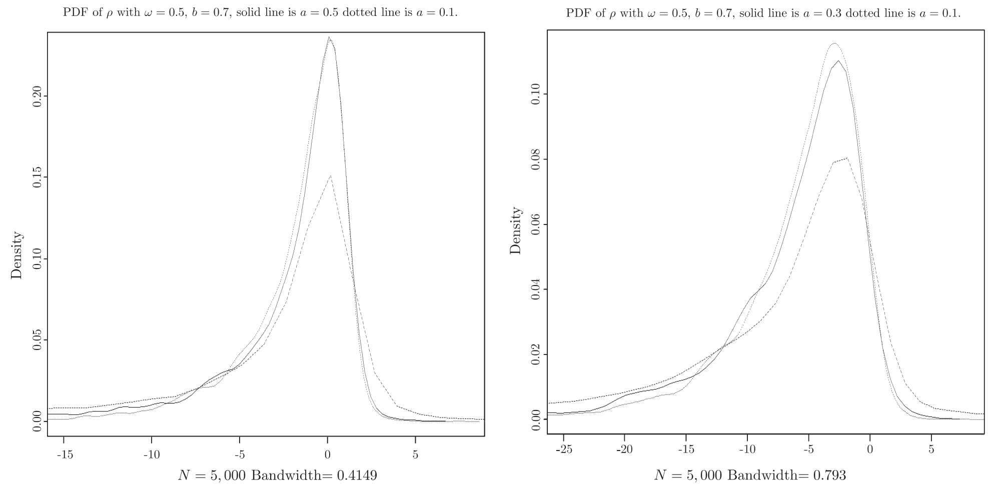

Figure 1. Probability density function of $\widehat{\rho}_n$ and $\widehat{\rho}_{\mu n}$.

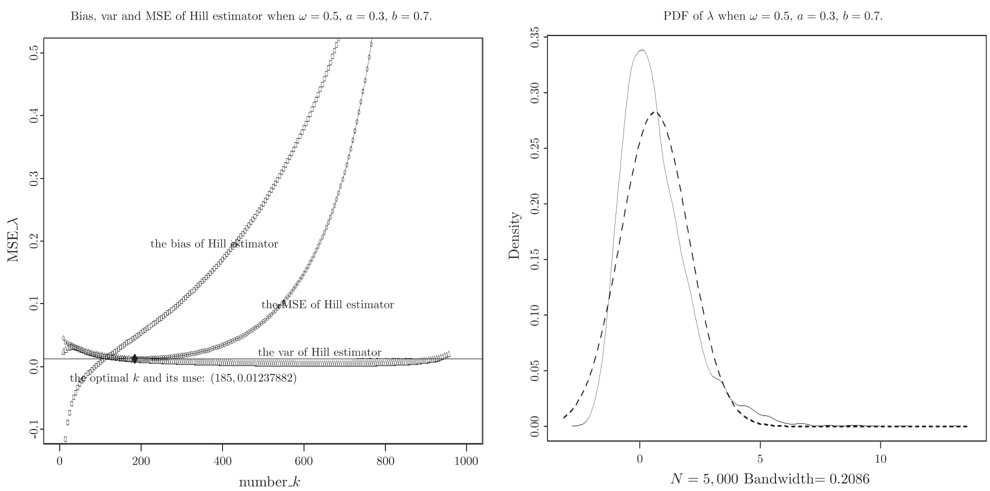

Figure 2. Hill estimators $\widehat{\Lambda}$ and their densities for $\omega = 0.5$, $a = 0.3$ and $b = 0.7$.

index α is small, then a large k_n should be chosen to minimize the mean squared error. This observation concurs with the findings reported in Danielsson et al. (2001). An explanation for this phenomenon is that such a k_n is determined by the quantity $n^{\beta/(\beta+\alpha)}$ defined in Theorem 3.1. As a result, a small α leads to a big k_n. From Figures 2−4, we see that when the tail index $\alpha > 1$, the pdf of the limit distribution approximates the normal limit reasonably well; when the tail index $\alpha < 1$, however, the approximation becomes less satisfactory.

Finally, small-scale simulations for the LAD estimator were also conducted. We simulated Y_t, $t = 1, \ldots, 400$ from model (1.1) and estimated ϕ by the LSE

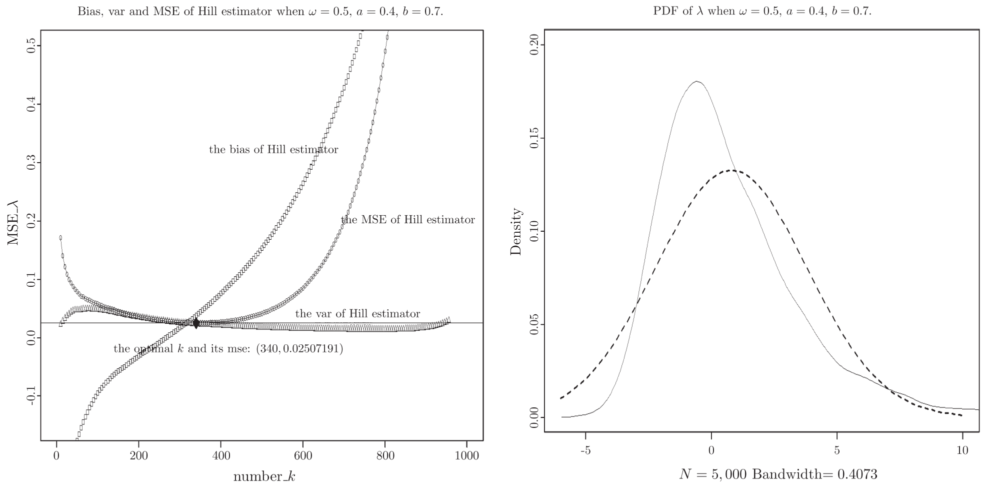

Figure 3. Hill estimators $\widehat{\Lambda}$ and their densities for $\omega = 0.5$, $a = 0.4$ and $b = 0.7$.

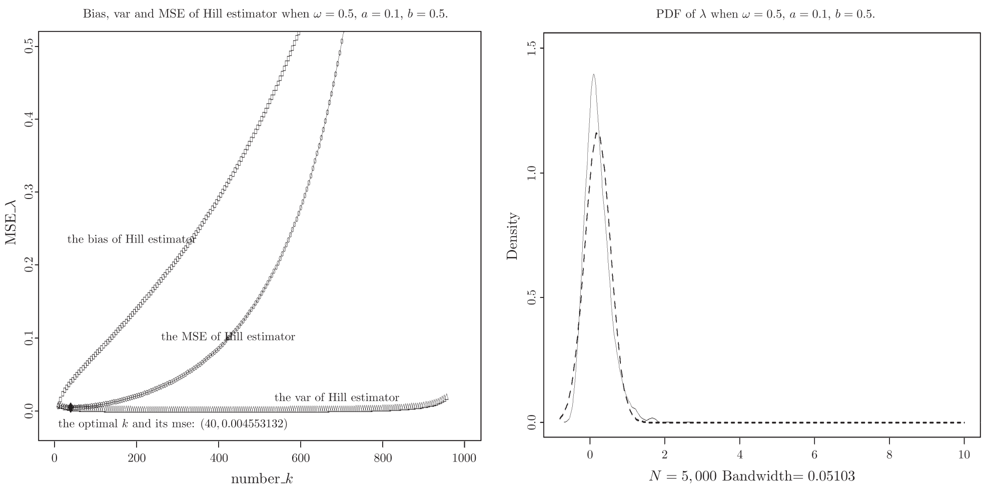

Figure 4. Hill estimators $\widehat{\Lambda}$ and their densities for $\omega = 0.5$, $a = 0.1$ and $b = 0.5$.

$\widehat{\phi}_n$. Using residuals $\{\widehat{\varepsilon}_t = Y_t - \widehat{\phi}Y_{t-1}\}$, we computed the LAD estimate and the QMLE for (ω, a, b). Three hundred independent samples were drawn for three cases: Case 1, $(\omega, a, b) = (0.5, 0.6, 0.3)$ and $\{\eta_t\}$ are standard normal; Case 2, $(\omega, a, b) = (0.5, 0.6, 0.1)$ and $\{\eta_t\}$ are t_3; Case 3, $(\omega, a, b) = (0.5, 0.6, 0.18)$ and $\{\eta_t\}$ are t_5. The bias and MSE of LAD and the QMLE are presented in Table 3. From Table 3, we see that the LAD estimate approximates the true (ω, a, b) better than the QMLE when the errors $\{\varepsilon_t\}$ are heavy-tailed. Even when ε_t has a light tail (Case 1), the LAD estimate performs comparably well with the QMLE. From these simulations, it seems reasonable to argue that the LAD estimate performs

Table 3. Bias and MSE of estimator $\widehat{\nu}$.

	Parameter	Case 1: N(0, 1)			Case 2: t_3			Case 3: t_5		
	(ω, a, b)	0.500	0.600	0.300	0.500	0.600	0.100	0.500	0.600	0.180
LAD	Bias	−0.018	−0.093	0.046	−0.019	−0.073	0.067	−0.007	−0.065	0.058
	MSE	0.031	0.039	0.045	0.026	0.032	0.029	0.026	0.029	0.039
QMLE	Bias	−0.023	−0.072	0.058	0.015	−0.052	0.081	−0.005	−0.065	0.086
	MSE	0.023	0.027	0.043	0.029	0.032	0.043	0.026	0.032	0.049

reasonably well when the errors are heavy-tailed, and is a viable alternative to the QMLE for the light-tailed case.

6. Proofs

6.1. Proofs of Theorems 2.1 and 2.2

Proof of Theorem 2.1. Conclusion (a) follows from Lemma A.1, leaving parts (b) and (c). For $\alpha \in (0,1)$, let $S_{jn}(t) = \sum_{i=1}^{[nt]} \varepsilon_i^j/(c_1 n)^{j/(2\alpha)}$, $j = 1,2$, and let $S_n(t) = (S_{1n}(t), S_{2n}(t))$. To show the finite-dimensional distributions of $S_n(t)$ converge to $Z(t) = (Z_{2\alpha}(t), Z_\alpha(t))$, it is enough to show that for any $t_1, \ldots, t_k \in (0,1]$,

$$(S_n(t_1), \ldots, S_n(t_k)) \xrightarrow{d} (Z(t_1), \ldots, Z(t_k)).$$

We only give $k = 2$ in detail, other cases can be shown similarly. It follows from Lemma A.6 that for any given $t \in (0,1)$, $S_n(t) \xrightarrow{d} Z(t)$. With the Cramér-Wold device, to have $(S_n(t_1), S_n(t_2)) \xrightarrow{d} (Z(t_1), Z(t_2))$, it is sufficient to show that for any $c = (c_1, c_2), d = (d_1, d_2) \in \mathbf{R}^2$ and $t_1 < t_2$, $c \cdot S_n(t_1) + d \cdot S_n(t_2) \xrightarrow{d} c \cdot Z(t_1) + d \cdot Z(t_2)$.

Let $a_n = (c_1 n)^{1/(2\alpha)}$, then for any $\delta > 0$, $P(|\varepsilon_i| > \delta a_n) \le \delta^{-2\alpha}/n$. It follows that for any $m = o(n)$, $\sum_{i=[nt_1]+1}^{[nt_1]+m} (d_1\varepsilon_i/a_n + d_2\varepsilon_i^2/a_n^2) \xrightarrow{p} 0$. Thus, when $m = o(n)$,

$$\begin{aligned} & c \cdot S_n(t_1) + d \cdot S_n(t_2) \\ & \quad = (c_1 + d_1) S_{1n}(t_1) + (c_2 + d_2) S_{2n}(t_1) + \sum_{i=[nt_1]+1+m}^{[nt_2]} \Big(\frac{d_1 \varepsilon_i}{a_n} + \frac{d_2 \varepsilon_i^2}{a_n^2} \Big) + o_p(1). \end{aligned}$$

By Lemma A.4 we have, for m large enough, the right side has the same distribution as

$$(c_1 + d_1) S_{1n}(t_1) + (c_2 + d_2) S_{2n}(t_1) + \sum_{i=[nt_1]+1+m}^{[nt_2]} \Big(\frac{d_1 \varepsilon_i'}{a_n} + \frac{d_2 \varepsilon_i'^2}{a_n^2} \Big),$$

where $\{\varepsilon_i'\}$ is an independent copy of $\{\varepsilon_i\}$. Thus

$$\begin{aligned}
& c\cdot S_n(t_1)+d\cdot S_n(t_2)\\
& \xrightarrow{d} (c_1+d_1)Z_{2\alpha}(t_1)+(c_2+d_2)Z_\alpha(t_1)+d_1(Z_{2\alpha}(t_2)-Z_{2\alpha}(t_1))+d_2(Z_\alpha(t_2)-Z_\alpha(t_1))\\
& = \ c\cdot Z(t_1)+d\cdot Z(t_2). \qquad (6.1)
\end{aligned}$$

This completes the proof of (b).

In the following, we apply a central limit theorem for a triangular array of martingale difference sequences (see Theorem 18.2 of Billingsley (1999)) to deal with weak convergence for the case $\alpha = 1$. Let $a_n = \sqrt{c_1 n\log n}$, $b_n = \sqrt{c_0 n\log\log n}$ and $\zeta_{ni} = \varepsilon_i I(\sigma_i \le b_n)/\sqrt{c_1 n\log n}$. Then

$$\frac{1}{a_n}\sum_{i=1}^{[nt]}\varepsilon_i = \frac{1}{a_n}\sum_{i=1}^{[nt]}\varepsilon_i I(\sigma_i\le b_n)+\frac{1}{a_n}\sum_{i=1}^{[nt]}\varepsilon_i I(\sigma_i > b_n) =: S_{n1}(t)+S_{n2}(t).$$

It is easy to see that for any $\delta > 0$,

$$P(\sup_{0\le t\le 1}|S_{n2}(t)|>\delta)\le P(\sup_{1\le i\le n}\sigma_i>b_n)\le nP(|\sigma_i|>b_n)\le \frac{1}{\log\log n}\to 0. \qquad (6.2)$$

So, it is enough to show that

$$S_{n1}(t)=\sum_{i=1}^{[nt]}\zeta_{ni}\Rightarrow W(t)\quad \text{in } D[0,1]. \qquad (6.3)$$

If $\mathcal{F}_t$ is the $\sigma-$field generated by $\{\varepsilon_i, i\le t\}$, then $\mathrm{E}(\zeta_{ni}|\mathcal{F}_{i-1})=0$. Since $\mathrm{E}\eta_1^2<\infty$ by virtue of H_2, it follows that for any $\delta>0$,

$$\begin{aligned}
\sum_{i=1}^{n}\mathrm{E}[\zeta_{ni}^2 I(|\zeta_{ni}|\ge\delta)] &= (c_1\log n)^{-1}\mathrm{E}[(\varepsilon_1 I(\sigma_1\le b_n))^2 I(|(\varepsilon_1 I(\sigma_1\le b_n)|\ge \delta a_n)]\\
&\le (c_1\log n)^{-1}\mathrm{E}[\sigma_1^2 I(\sigma_1\le b_n)]\mathrm{E}[\eta_1^2 I(|\eta_1|\ge \frac{\delta a_n}{b_n}]\\
&= (c_1\log n)^{-1}\int_0^{b_n} 2x[P(|\sigma_1|>x)-P(\sigma_1>b_n)]\,dx\\
&\quad\ \mathrm{E}\Big[\eta_1^2 I(|\eta_1|\ge\frac{\delta a_n}{b_n}\Big]\\
&\le (c_1\log n)^{-1}\Big(1+\int_1^{b_n}2y^{-1}\,dy\Big)\mathrm{E}\Big[\eta_1^2 I(|\eta_1|\ge\frac{\delta a_n}{b_n}\Big]\\
&\to 0. \qquad (6.4)
\end{aligned}$$

By Theorem 18.2 of Billingsley (1999), for the proof of (6.3) it is enough to show that for any $t \in [0,1]$,

$$\sum_{i=1}^{[nt]} \mathrm{E}(\zeta_{ni}^2|\mathcal{F}_{i-1}) = \mathrm{E}\eta_1^2 \sum_{i=1}^{[nt]} \frac{\sigma_i^2 I(\sigma_i \le b_n)}{a_n^2} = \frac{1}{c_0 n \log n} \sum_{i=1}^{[nt]} \sigma_i^2 I(\sigma_i \le b_n) \xrightarrow{p} t. \tag{6.5}$$

Note that

$$\begin{aligned}
\mathrm{E}\sigma_1^2 I(\sigma_1 \le b_n)) &= \int_0^{b_n} 2x[P(|\sigma_1| > x) - P(\sigma_1 > b_n)]\,dx \\
&\le 1 + \int_1^{b_n} 2c_0 y^{-1}\,dy \sim c_0 \log n, \\
\mathrm{E}\sigma_1^2 I(\sigma_1 \le b_n)) &\ge \int_1^{\sqrt{n}} 2x[P(|\sigma_1| > x) - P(\sigma_1 > b_n)]\,dx \\
&\ge \int_1^{\sqrt{n}} 2c_0 y^{-1}\,dy - 1 \sim c_0 \log n.
\end{aligned}$$

Thus, for n large enough, $\mathrm{E}\sigma_1^2 I(\sigma_1 \le b_n)) \sim c_0 \log n$, and (6.5) is equivalent to showing

$$\frac{1}{c_0 n \log n} \sum_{i=1}^{[nt]} [\sigma_i^2 I(\sigma_i \le b_n) - \mathrm{E}\sigma_i^2 I(\sigma_i \le b_n)] \xrightarrow{p} 0. \tag{6.6}$$

To this end, split $\sigma_i^2 I(\sigma_i \le b_n) - \mathrm{E}\sigma_i^2 I(\sigma_i \le b_n)$ into two parts: $\xi_{n,i} = \sigma_i^2 I(\sigma_i \le \sqrt{n}y) - \mathrm{E}\sigma_i^2 I(\sigma_i \le \sqrt{n}y)$ and $X_{n,i} = \sigma_i^2 I(\sqrt{n}y < \sigma_i \le b_n) - \mathrm{E}\sigma_i^2 I(\sqrt{n}y < \sigma_i \le b_n)$, where $y > 0$. Since σ_i^2 satisfies the β-mixing condition with exponential decay, it follows that for n large enough and for any $\delta > 0$,

$$\begin{aligned}
&\mathrm{E}\Big(\frac{1}{c_0 n \log n} \sum_{i=1}^{[nt]} \xi_{n,i}\Big)^2 \\
&\quad = \frac{1}{(c_0 n \log n)^2}\Big(\sum_{i=1}^{[nt]} \mathrm{E}\xi_{n,i}^2 + 2\sum_{i=1}^{[nt]} \sum_{j=i+1}^{[nt]} \mathrm{E}\xi_{n,i}\xi_{n,j}\Big) \\
&\quad \le 2tn^{-1}ny^2(c_0 \log n)^{-2}\mathrm{E}(\sigma_i^2 I(\sigma_i \le b_n)) + 2tn^{-1}(c_0 \log n)^{-2}\Big\{\sum_{i=1}^{m} \mathrm{E}\xi_{n,1}\xi_{n,i+1} \\
&\qquad + \sum_{i=m+1}^{[nt]} [(\rho^i)^{\delta/(2+\delta)}(\mathrm{E}|\xi_{n,1}|^{2+\delta})^{2/(2+\delta)} + \mathrm{E}\xi_{n,1}\mathrm{E}\xi_{n,i+1}]\Big\} \\
&\quad \le 4ty^2(c_0 \log n)^{-1} + 4tn^{-1}mny^2(c_0 \log n)^{-1}
\end{aligned}$$

$$+4tn^{-1}(c_0\log n)^{-2}\rho^{m\delta/(2+\delta)}[(\sqrt{n}y)^{2(1+\delta)}\mathrm{E}|\xi_{n,1}|]^{2/(2+\delta)}\sum_{i=1}^{[nt]}\rho^{i\delta/(2+\delta)}$$

$$\to 0, \tag{6.7}$$

for any $t \in [0,1]$ by taking $m = [-2\log n/\log\rho]$ and $y \to 0$. This implies that for large n and small y,

$$\frac{1}{c_0 n\log n}\sum_{i=1}^{[nt]}\xi_{ni} \xrightarrow{p} 0. \tag{6.8}$$

By Lemma A.1, it follows that for large x, $P(\sigma_1^2 > x) \sim c_0 x^{-1}$. As with Theorem 4.1 of Davis and Mikosch (1998), it can be seen that Conditions $1, 2, 3$ in Lemma A.5 are satisfied for the process $\{\sigma_t^2\}$. Therefore, there exists a process N (see Lemma A.5) such that

$$N_n = \sum_{i=1}^{n}\delta_{\sigma_i^2/n} \xrightarrow{d} N.$$

For any $y > 0$, let $T : \sum_{i=1}^{\infty}\delta_{x_i} \to \sum_{i=1}^{\infty}x_i I_{(y,\infty)}(|x_i|)$. Then T is continuous and by the Continuous Mapping Theorem (see (3.8) of Davis and Hsing (1995)), we have

$$\frac{1}{c_0 n}\sum_{i=1}^{n}\{\sigma_i^2 I(\sqrt{n}y < \sigma_i) - \mathrm{E}[\sigma_i^2 I(\sqrt{n}y < \sigma_i \le \sqrt{n})]\} \xrightarrow{d} T(N) + \log y.$$

This implies that for any $t > 0$,

$$\frac{1}{c_0 n\log n}\sum_{i=1}^{[nt]}\{\sigma_i^2 I(\sqrt{n}y < \sigma_i) - \mathrm{E}[\sigma_i^2 I(\sqrt{n}y < \sigma_i \le b_n)]\} \xrightarrow{p} 0.$$

On the other hand, from (6.2), it follows that for any $\delta > 0$,

$$P\Big\{\frac{1}{c_0 n\log n}\sum_{i=1}^{n}\sigma_i^2 I(\sigma_i > b_n) > \delta\Big\} \le P\{\max_{1\le i\le n}\sigma_i > b_n\} \to 0.$$

Thus, $(c_0 n\log n)^{-1}\sum_{i=1}^{[nt]}X_{n,i} \xrightarrow{p} 0$, which combined with (6.8) yields (6.6). This completes the proof of (c) and the proof of Theorem 2.1.

Proof of Theorem 2.2. Note that under either $\phi = 1$ in model (1.1) or $(\phi,\mu) = (1,0)$ in model (1.2), $Y_i = Y_{i-1} + \varepsilon_i = \sum_{j=1}^{i}\varepsilon_j$. This implies that $Y_{i-1}\varepsilon_i = Y_{i-1}(Y_i - Y_{i-1})$. Thus,

$$\widehat{\rho}_n = \Big(\frac{1}{n}\sum_{i=1}^{n}Y_{i-1}^2\Big)^{-1}\Big(\sum_{i=1}^{n}Y_{i-1}\varepsilon_i\Big) = \Big(\frac{1}{n}\sum_{i=1}^{n}Y_{i-1}^2\Big)^{-1}\Big(\sum_{i=1}^{n}Y_{i-1}(Y_i - Y_{i-1})\Big), \tag{6.9}$$

$$\widehat{\rho}_{\mu n} = \Big(\frac{1}{n}\sum_{i=1}^{n} Y_{i-1}^2\Big)\Big(\sum_{i=1}^{n} Y_{i-1}(Y_i - Y_{i-1}) - \overline{Y}\sum_{i=1}^{n}\varepsilon_i\Big). \tag{6.10}$$

When $\alpha = 1$, by (c) of Theorem 2.1 and the Continuous Mapping Theorem, we have

$$\begin{aligned}\widehat{\rho}_n &= \Big(\frac{1}{n}\sum_{i=1}^{n}[(c_1 n)^{-1/2\alpha}Y_{i-1}]^2\Big)^{-1}\Big(\sum_{i=1}^{n}[(c_1 n)^{-1/2\alpha}Y_{i-1}][(c_1 n)^{-1/2\alpha}(Y_i - Y_{i-1})]\Big)\\ &\xrightarrow{d} \int_0^1 W(t)\,dW(t)\Big/\int_0^1 W^2(t)\,dt. \end{aligned}\tag{6.11}$$

Similarly, by (6.10) and the Continuous Mapping Theorem, we have

$$\widehat{\rho}_{\mu n} \xrightarrow{d} \frac{\int_0^1 W(t)\,dW(t) - W(1)\int_0^1 W(t)\,dt}{\int_0^1 W^2(t)\,dt - (\int_0^1 W(t)\,dt)^2}.$$

This, combined with Remark 2.4, completes the proof of (b) in Theorem 2.2.

When $\alpha \in (0,1)$, rewrite $\widehat{\rho}_n = \Big(\frac{1}{n}\sum_{i=1}^{n} Y_{i-1}^2\Big)^{-1}\Big[\frac{1}{2}\Big(Y_n^2 - \sum_{i=1}^{n}\varepsilon_i^2\Big)\Big]$ and

$$\widehat{\rho}_{\mu n} = \Big(\frac{1}{n}\sum_{i=1}^{n} Y_{i-1}^2\Big)^{-1}\Big(\frac{Y_n^2}{2} - \frac{1}{2}\sum_{i=1}^{n}\varepsilon_i^2 - \overline{Y}\sum_{i=1}^{n}\varepsilon_i\Big).$$

Let $Y_{1n}(t) = \sum_{i=1}^{[nt]}\varepsilon_i I(|\varepsilon_i| \le a_n)/a_n$, $Y_{2n}(t) = \sum_{i=1}^{[nt]}\varepsilon_i I(|\varepsilon_i| > a_n)/a_n$, and $Y_n(t) = Y_{1n}(t) + Y_{2n}(t)$. By the symmetric assumption we have, for all n, $\{\varepsilon_i I(|\varepsilon_i| \le a_n)/a_n\}$ is a martingale difference sequence and $\sup_n \sup_{t\in[0,1]} \mathrm{E}|Y_{1n}(t)| < \infty$. By Doob's inequality, we have the following results.

(a) $\{\max_{0\le t\le 1}|Y_{1n}(t)|\}$ is stochastically bounded.

(b) For any $a < b, a, b \in \mathbf{R}$, if $N^{a,b}(Y_{1n})$ is the number of up-crossings of $[a,b]$ by the process Y_{1n}, then $\{N^{a,b}(Y_{1n})\}$ is stochastically bounded.

If $\beta = \min\{2\alpha - \tau, 1\}$, $\tau < 2\alpha$, and $N^{a,b}(Y_{2n})$ is the number of up-crossings of $[a,b]$ by the process Y_{2n}, then

$$\begin{aligned}\max_n \mathrm{E}(\max_{0\le t\le 1}|Y_{2n}(t)|)^\beta &\le \max_n a_n^{-\beta}\sum_{i=1}^{n}\mathrm{E}[|\varepsilon_i|^\beta I(|\varepsilon_i| > a_n)] < \infty,\\ \mathrm{E}\{N^{a,b}(Y_{2n})\} &\le \mathrm{E}\Big(\textstyle\sum_{i=1}^{n} I(|\varepsilon_i| > a_n)\Big) = nP(|\varepsilon_1| > a_n) < \infty.\end{aligned}$$

This implies that $\{\max_{0\le t\le 1}|Y_{2n}(t)|\}$ and $\{N^{a,b}(Y_{2n})\}$ are stochastically bounded. Since

$$\max_{0\le t\le 1}|Y_n(t)| \le \max_{0\le t\le 1}|Y_{1n}(t)| + \max_{0\le t\le 1}|Y_{2n}(t)|$$

and the number $N^{a,b}(Y_n)$ of up-crossings of $[a,b]$ by Y_n is no more than $N^{a,b}(Y_{1n})+N^{a,b}(Y_{2n})$, it follows that $\{\max_{0\le t\le 1}|Y_n(t)|\}$ and $\{N^{a,b}(Y_n)\}$ are stochastically bounded sequences. By Theorem 3.2 of Jakubowski (1997), it follows that $\{Y_n(\cdot)\}$ is uniformly S-tight in $D[0,1]$. Let $\tau_m=\{0=t_{m,0}<t_{m,1}<\cdots<t_{m,k_m}=1\}$ be such that $\max_k|t_{m,k}-t_{m,k-1}|\to 0$ and $X_n(t)=\sum_{i=1}^{[nt]}1/n$. Along the lines of Lemma 8 of Jakubowski (1996), we have for any $\delta>0$,

$$\lim_{m\to\infty}\sup_n P\Big[\Big|\Big(\int_0^1(Y_n(t-))^{\tau_m}\,dX_n(t),\int_0^1(Y_n^2(t-))^{\tau_m}\,dX_n(t)\Big)-\Big(\int_0^1 Y_n(t-)\,dX_n(t),\int_0^1 Y_n^2(t-)\,dX_n(t)\Big)\Big|>\delta\Big]=0, \quad (6.12)$$

where $(X_n(t))^{\tau_m}$ is given by $(X_n(1))^{\tau_m}=X_n(1)$ and $(X_n(t))^{\tau_m}=X_n(t_k)$ when $t_k\le t<t_{k+1}, k=1,\ldots,m-1$. By (a) of Theorem 2.1, we have

$$\Big(\int_0^1(Y_n(t-))^{\tau_m}\,dX_n(t),\int_0^1(Y_n^2(t-))^{\tau_m}\,dX_n(t),Y_n,\sum_{i=1}^n\frac{\varepsilon_i^2}{a_n^2}\Big)\xrightarrow{d}\Big(\int_0^1(Z_{2\alpha}(t-))^{\tau_m}\,dt,\int_0^1(Z_{2\alpha}^2(t-))^{\tau_m}\,dt,Z_{2\alpha}(1),Z_\alpha(1)\Big). \quad (6.13)$$

By (6.12) and (6.13), we have (a) of Theorem 2.2, as desired.

6.2. Proof of Theorem 3.1.

By Theorem 2.1 of Drees (2000) it can be shown that

$$\frac{1}{\sqrt{k_n}}\sum_{i=1}^n\Big(I(\varepsilon_i\ge F^{-1}(1-\frac{k_nu}{n}))-\frac{k_nu}{n}\Big)\Rightarrow B(u)\text{ in }D[0,1], \quad (6.14)$$

where $B(\cdot)$ is a standard Brownian motion. By standard arguments (see de Haan and Resnick (1998)), for the proof of Theorem 3.1 it is enough to show that

$$\sup_{0\le u\le 1}\frac{1}{\sqrt{k_n}}\sum_{i=1}^n[I(\widehat{\varepsilon}_i\ge F^{-1}(1-\frac{k_nu}{n}))-I(\varepsilon_i\ge F^{-1}(1-\frac{k_nu}{n}))]\xrightarrow{p}0. \quad (6.15)$$

Let

$$S_n(u,\theta)=\frac{1}{\sqrt{k_n}}\sum_{i=1}^n[I(\varepsilon_i\ge F^{-1}(1-\frac{k_nu}{n})+\frac{\theta Y_{i-1}}{n})-I(\varepsilon_i\ge F^{-1}(1-\frac{k_nu}{n}))]$$
$$=:\frac{1}{\sqrt{k_n}}\sum_{i=1}^n Z_i(u,\theta). \quad (6.16)$$

Let $\mathcal{F}_t$ be the σ-field generated by $\{\eta_j, j \le t\}$. Then

$$\begin{aligned}
&\mathrm{E}(Z_i(u,\theta)|\mathcal{F}_{i-1})\\
&\quad= \mathrm{E}\{I[F^{-1}(1-\frac{k_n u}{n}) + \frac{\theta Y_{i-1}}{n} < \varepsilon_i \le F^{-1}(1-\frac{k_n u}{n})]\\
&\qquad -I[F^{-1}(1-\frac{k_n u}{n}) < \varepsilon_i \le F^{-1}(1-\frac{k_n u}{n}) + \frac{\theta Y_{i-1}}{n}]|\mathcal{F}_{i-1}\}\\
&\quad= \mathrm{E}\Big[I\Big(\frac{\theta Y_{i-1}}{n} < 0\Big)\int_{(F^{-1}(1-k_n u/n)+\theta Y_{i-1}/n)/\sigma_i}^{(F^{-1}(1-k_n u/n))/\sigma_i} \sigma_i f(\sigma_i x)\, dx|\mathcal{F}_{i-1}\Big]\\
&\qquad -\mathrm{E}\Big[I\Big(\theta Y_{i-1}/n > 0\Big)\int_{(F^{-1}(1-k_n u/n))/\sigma_i}^{(F^{-1}(1-k_n u/n)+\theta Y_{i-1}/n)/\sigma_i} \sigma_i f(\sigma_i x)\, dx|\mathcal{F}_{i-1}\Big]\\
&\quad=: H_{i1} - H_{i2},
\end{aligned} \tag{6.17}$$

where $f(x)$ is the density of ε. Note that $\sup_{1\le i\le n}|Y_i|/n^{1/(2\alpha)} = O_p(1)$. It follows from $\alpha > 1/2$ that $\sup_{1\le i\le n} Y_i/n = o_p(1)$. By the continuity of $f(\cdot)$, we have

$$H_{i1} \overset{p}{=} -f(F^{-1}(1-\frac{k_n u}{n}))(\frac{\theta Y_{i-1}}{n})I(\theta Y_{i-1} < 0). \tag{6.18}$$

Similarly, it can be seen that

$$H_{i2} \overset{p}{=} f(F^{-1}(1-\frac{k_n u}{n}))(\frac{\theta Y_{i-1}}{n})I(\theta Y_{i-1} > 0). \tag{6.19}$$

Furthermore, for all $|\theta| \le M$ and $0 \le u \le 1$,

$$\begin{aligned}
\frac{1}{k_n}\sum_{i=1}^n \mathrm{E}(Z_i^2(u,\theta)|\mathcal{F}_{i-1}) &\le \frac{2}{k_n}\sum_{i=1}^n (H_{i1} + H_{i2})\\
&\le \frac{2}{k_n}\max_{1\le i\le n} M|Y_i| f(F^{-1}(1-\frac{k_n}{n}))\\
&= O_p\Big(\frac{k_n^{1/2\alpha}}{n}\Big) = o_p(1).
\end{aligned} \tag{6.20}$$

This implies that, for all $|\theta| \le M$ and $0 \le u \le 1$,

$$\frac{1}{\sqrt{k_n}}\sum_{i=1}^n Z_i(u,\theta) \overset{p}{=} \frac{1}{\sqrt{k_n}}\sum_{i=1}^n \mathrm{E}(Z_i(u,\theta)|\mathcal{F}_{i-1}) = \frac{1}{\sqrt{k_n}}\sum_{i=1}^n (H_{i1} - H_{i2}). \tag{6.21}$$

Similar to the argument as (6.20), the right side of (6.21) is $O_p(k_n^{1/2\alpha+1/2}n^{-1}) = o_p(1)$ by virtue of the condition that $k_n = o(n^{(2\alpha)/(\alpha+1)})$. Since $S_n(u,\theta)$ has convex sample paths in u and θ, the above convergence implies uniform convergence on compact sets (see Pollard (1991)). Thus, (6.15) follows from $n(\widehat{\phi}-\phi)=O_p(1)$ by Theorem 2.2.

6.3. Proof of Theorem 4.1

Set $l_n(\nu)=n^{-1}\sum_{i=1}^n(|\varepsilon_i|/\sqrt{h_i(\nu,1)}+\ln\sqrt{h_i(\nu,1)})$ and $l(\nu)=\mathrm{E}l_n(\nu)$. We show the consistency of $\widehat{\nu}$ via the following steps. First, along the lines of Theorem 1 of Lumsdaine (1996), we have ν_0 as the unique minimization of $l(\nu)$. Second, by the Ergodic Theorem,

$$\sup_{\nu\in\Theta}|l_n(\nu)-l(\nu)|=o_p(1), \tag{6.22}$$

and we have

$$\sup_{\nu\in\Theta}\Big|\frac{1}{n}\sum_{i=1}^n|\widehat{\varepsilon}_i|/\sqrt{h_i(\nu,\widehat{\phi})}+\ln\sqrt{h_i(\nu,\widehat{\phi})}-l_n(\nu)\Big|=o_p(1). \tag{6.23}$$

Note that the left side of (6.23) is no more than

$$\sup_{\nu\in\Theta}\frac{1}{n}\sum_{i=1}^n\Big[|\varepsilon_i|\Big|\frac{1}{\sqrt{h_i(\nu,\widehat{\phi})}}-\frac{1}{\sqrt{h_i(\nu,1)}}\Big|+\Big|\ln\frac{\sqrt{h_i(\nu,\widehat{\phi})}}{\sqrt{h_i(\nu,1)}}\Big|+\Big|\frac{(1-\widehat{\phi})Y_{i-1}}{\sqrt{h_i(\nu,\widehat{\phi})}}\Big|\Big]$$
$$\le\sup_{\nu\in\Theta}\frac{1}{n}\sum_{i=1}^n\Big[|\varepsilon_i|\frac{\sqrt{|h_i(\nu,\widehat{\phi})-h_i(\nu,1)|}}{\sqrt{\omega h_i(\nu,1)}}+\frac{1}{2}\frac{|h_i(\nu,\widehat{\phi})-h_i(\nu,1)|}{h_i(\nu,1)}+\frac{1}{\sqrt{\omega}}|(1-\widehat{\phi})Y_{i-1}|\Big].$$

Since $\alpha>1/2$, it follows that $Y_n^*:=\max_{1\le i\le n}|Y_i/n|=o_p(1)$. Thus,

$$\begin{aligned}
&\max_i\frac{|h_i(\nu,\widehat{\phi})-h_i(\nu,1)|}{h_i(\nu,1)}\\
&\quad=\max_i\frac{b\Big|\sum_{j=0}^{i-1}a^{i-1-j}[2(1-\widehat{\phi})\varepsilon_jY_{j-1}+(1-\widehat{\phi})^2Y_{j-1}^2]\Big|}{(\omega\sum_{j=0}^{i-1}a^j+b\sum_{j=0}^{i-1}a^{i-1-j}\varepsilon_j^2)}\\
&\quad\le C|n(1-\widehat{\phi})|\sum_{j=0}^{i-1}a^{(i-1-j)/2}Y_n^*+C[n(1-\widehat{\phi})]^2\sum_{j=0}^{i-1}a^{i-1-j}(Y_n^*)^2=o_p(1).
\end{aligned}$$

Thus, by $\sum_{i=1}^n|\varepsilon_i|/n=O_p(1)$, we have (6.23). By (6.22) and (6.23),

$$\begin{aligned}
&\Big|\frac{1}{n}\sum_{i=1}^n\Big(\frac{|\widehat{\varepsilon}_i|}{\sqrt{h_i(\nu,\widehat{\phi})}}+\ln\sqrt{h_i(\nu,\widehat{\phi})}\Big)-l_n(\nu_0)\Big|\\
&\quad=\Big|l(\nu)-l_n(\nu)+l_n(\nu)-\frac{1}{n}\sum_{i=1}^n\Big(|\widehat{\varepsilon}_i|/\sqrt{h_i(\nu,\widehat{\phi})}+\ln\sqrt{h_i(\nu,\widehat{\phi})}\Big)
\end{aligned}$$

$$
\begin{aligned}
&\quad +l_n(\nu_0) - l(\nu_0) + l(\nu_0) - l(\nu)\Big| \\
&= |l(\nu) - l(\nu_0)| + o_p(1) \qquad (6.24)
\end{aligned}
$$

holds uniformly for all $\nu \in \Theta$. Thus, with ν_0 as the unique minimization of $l(\nu)$,

$$
\inf_{\nu\in\Theta}\Big|\frac{1}{n}\sum_{i=1}^{n}\Big(\frac{|\widehat{\varepsilon}_i|}{\sqrt{h_i(\nu,\widehat{\phi})}}+\ln\sqrt{h_i(\nu,\widehat{\phi})}\Big)-l_n(\nu_0)\Big| = \inf_{\nu\in\Theta}|l(\nu)-l(\nu_0)|+o_p(1) = o_p(1).
$$

From the definition of $\widehat{\nu}$, it follows that

$$
|l(\widehat{\nu}) - l(\nu_0)| = \Big|\frac{1}{n}\sum_{i=1}^{n}\Big(\frac{|\widehat{\varepsilon}_i|}{\sqrt{h_i(\widehat{\nu},\widehat{\phi})}} + \ln\sqrt{h_i(\widehat{\nu},\widehat{\phi})}\Big) - l_n(\nu_0)\Big| + o_p(1) = o_p(1),
$$

which implies that $\widehat{\nu} \xrightarrow{p} \nu_0$. This completes the proof of the consistency of $\widehat{\nu}$.

Next, we show (4.2). Let $\widetilde{\theta} = \sqrt{n}(\widehat{\nu} - \nu_0)$, $x = (x_1, x_2, x_3) \in \mathbb{R}^3$, and

$$
\begin{aligned}
g(x,\widehat{\phi}) = \sum_{i=1}^{n}\Big[&\Big(\frac{|\widehat{\varepsilon}_i|}{\sqrt{h_i(\nu_0+\frac{x}{\sqrt{n}},\widehat{\phi})}} + \frac{1}{2}\log h_i(\nu_0+\frac{x}{\sqrt{n}},\widehat{\phi})\Big) \\
&-\Big(\frac{|\widehat{\varepsilon}_i|}{\sqrt{h_i(\nu_0,\widehat{\phi})}} + \frac{1}{2}\log h_i(\nu_0,\widehat{\phi})\Big)\Big].
\end{aligned}
$$

Then $\widetilde{\theta}$ minimizes $g(x,\widehat{\phi})$. With $\widehat{\eta}_i = \widehat{\varepsilon}_i/\sqrt{h_i(\nu_0,\widehat{\phi})}$, by a Taylor expansion and elementary computations, it can be seen that under the conditions of Theorem 4.1,

$$
\begin{aligned}
g(x,\widehat{\phi}) &= -\frac{1}{2\sqrt{n}}\sum_{i=1}^{n} x(|\widehat{\eta}_i|-1)h_i^{-1}(\nu_0,\widehat{\phi})[h_i'(\nu_0,\widehat{\phi})]^T \\
&\quad +\frac{3}{8n}\sum_{i=1}^{n} x(|\widehat{\eta}_i|-1)h_i^{-2}(\nu_0,\widehat{\phi})(h_i'(\nu_0,\widehat{\phi}))^T h_i'(\nu_0,\widehat{\phi})x^T \\
&\quad +\frac{1}{8n}\sum_{i=1}^{n} x h_i^{-2}(\nu_0,\widehat{\phi})(h_i'(\nu_0,\widehat{\phi}))^T h_i'(\nu_0,\widehat{\phi})x^T \\
&\quad -\frac{1}{4n}\sum_{i=1}^{n} x(|\widehat{\eta}_i|-1)h_i^{-1}(\nu_0,\widehat{\phi})h_i''(\nu_0,\widehat{\phi})x^T + o_p(1) \\
&=: J_1(x,\widehat{\phi}) + J_2(x,\widehat{\phi}) + J_3(x,\widehat{\phi}) - J_4(x,\widehat{\phi}) + o_p(1), \qquad (6.25)
\end{aligned}
$$

where $h''(\nu_0,\widehat{\phi}) = \partial^2 h(\nu,\widehat{\phi})/\partial\nu^2|_{\nu=\nu_0}$. Let $h'(\nu_0)$ and $h''(\nu_0)$ be the first and second derivative of $h(\nu,1)$ at ν_0, respectively. The conclusion of Theorem 4.1

follows once the following assertions are established. For any $M > 0$,

$$\sup_{|x|\le M} |J_1(x,\widehat{\phi})+\frac{x}{2\sqrt{n}}\sum_{i=1}^{n}[(|\eta_i|-1)]h_i^{-1}(\nu_0)[h_i'(\nu_0)]^T-\frac{xFn(\widehat{\phi}-1)}{2n\sqrt{n}}\sum_{i=1}^{n}Y_{i-1}|=o_p(1). \tag{6.26}$$

$$\sup_{|x|\le M} |J_2(x,\widehat{\phi}) - \frac{3}{8n}\sum_{i=1}^{n} x(|\eta_i|-1)h_i^{-2}(\nu_0)(h_i'(\nu_0)^T h_i'(\nu_0)x^T| = o_p(1), \tag{6.27}$$

$$\sup_{|x|\le M} |J_3(x,\widehat{\phi}) - \frac{1}{8n}\sum_{i=1}^{n} xh_i^{-2}(\nu_0)(h_i'(\nu_0))^T h_i'(\nu_0)x^T| = o_p(1), \tag{6.28}$$

$$\sup_{|x|\le M} |J_4(x,\widehat{\phi}) - \frac{1}{4n}\sum_{i=1}^{n} x(|\eta_i|-1)h_i^{-1}(\nu_0)h_i''(\nu_0)x^T| = o_p(1), \tag{6.29}$$

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(|\eta_i|-1)h_i^{-1}(\nu_0)[h_i'(\nu_0)]^T \xrightarrow{d} \mathrm{N}(0,\mathrm{E}(|\eta_i|-1)^2\mathrm{E}(h_i^{-2}(\nu_0)[h_i'(\nu_0)]^T h_i'(\nu_0))), \tag{6.30}$$

$$\frac{Fn(\widehat{\phi}-1)}{n\sqrt{n}}\sum_{i=1}^{n}Y_{i-1} \xrightarrow{d} F\zeta, \tag{6.31}$$

$$\frac{1}{n}\sum_{i=1}^{n} h_i^{-2}(\nu_0)(h_i'(\nu_0))^T h_i'(\nu_0) \xrightarrow{p} \mathrm{E}(h_i^{-2}(\nu_0)[h_i'(\nu_0)]^T h_i'(\nu_0)), \tag{6.32}$$

$$\frac{1}{n}\sum_{i=1}^{n} x(|\eta_i|-1)h_i^{-2}(\nu_0)(h_i'(\nu_0)^T h_i'(\nu_0)x^T \xrightarrow{p} 0, \tag{6.33}$$

$$\frac{1}{n}\sum_{i=1}^{n} x(|\eta_i|-1)h_i^{-1}(\nu_0)h_i''(\nu_0)x^T \xrightarrow{p} 0. \tag{6.34}$$

As in the proof of Theorem 2.2 in Francq and Zakoïan (2004), it can be shown that the second moment of $h_i^{-1}(\nu_0)[h_i'(\nu_0)]^T$ exists. Therefore, (6.30) follows from the Martingale Central Limit Theorem (see Theorem 3.2 of Hall and Heyde (1980)), (6.31) follows from Theorem 2.2 and the Martingale Central Limit Theorem for $\{Y_i\}$, and (6.32), (6.33), and (6.34) follow from the Ergodic Theorem. When $\alpha > 1$, we have $\widehat{\varepsilon}_i^2 = \varepsilon_i^2 + O_p(n^{-1}) + \varepsilon_i O_p(n^{-1/2})$ and $h_i(v_0,\widehat{\phi}) = h_i(v_0) + h_i^{1/2}(v_0)O_p(n^{-1/2})$. With these, (6.28), (6.27), and (6.29) can be derived similarly to the proof of Theorem C of Ling and Li (2003). We give only the proof of (6.26) in detail. Note that

$$J_1(x,\widehat{\phi}) = -\frac{x}{2\sqrt{n}}\sum_{i=1}^{n}(|\eta_i|-1)h_i^{-1}(\nu_0)[h_i'(\nu_0)]^T$$

$$
\begin{aligned}
&+\frac{x}{2\sqrt{n}}\sum_{i=1}^{n}(|\eta_i|-1)[h_i^{-1}(\nu_0)(h_i'(\nu_0))^T-h_i^{-1}(\nu_0,\widehat{\phi})(h_i'(\nu_0,\widehat{\phi}))^T]\\
&-\frac{x}{2\sqrt{n}}\sum_{i=1}^{n}(|\eta_i|-1)(\sqrt{h_i(\nu_0)}-\sqrt{h_i(\nu_0,\widehat{\phi})})\frac{[h_i'(\nu_0,\widehat{\phi})]^T}{h_i^{3/2}(\nu_0,\widehat{\phi})}\\
&-\frac{x}{2\sqrt{n}}\sum_{i=1}^{n}(|\varepsilon_i+(1-\widehat{\phi})Y_{i-1}|-|\varepsilon_i|)\frac{[h_i'(\nu_0,\widehat{\phi})]^T}{h_i^{3/2}((\nu_0,\widehat{\phi})}\\
&-\frac{x}{2\sqrt{n}}\sum_{i=1}^{n}(\sqrt{h_i(\nu_0)}-\sqrt{h_i(\nu_0,\widehat{\phi})})\frac{[h_i'(\nu_0,\widehat{\phi})]^T}{h_i^{3/2}(\nu_0,\widehat{\phi})}\\
&=: I_{1n}+I_{2n}+I_{3n}+I_{4n}+I_{5n}. \qquad (6.35)
\end{aligned}
$$

We show $I_{2n}=o_p(1)$ for all $x\in\mathbf{R^3}$. Let

$$
\begin{aligned}
I_{2n}(z) &= \frac{bx}{2\sqrt{n}}\sum_{i=1}^{n}(|\eta_i|-1)[h_i^{-1}(\nu_0)(h_i'(\nu_0))^T-h_i^{-1}(\nu_0,1-\frac{z}{n})(h_i'(\nu_0,1-\frac{z}{n}))^T]\\
&=: \frac{bx}{2\sqrt{n}}\sum_{i=1}^{n}(|\eta_i|-1)\xi_i(z).
\end{aligned}
$$

Note that for any $M>0$,

$$
\sup_{|x|,|z|\le M}|bx\sum_{i=1}^{n}(|\eta_i|-1)(\xi_i(z)|\le \sup_{|z|\le M} bM||\sum_{i=1}^{n}(|\eta_i|-1)\xi_i(z)||,
$$

and for any $\varepsilon>0$, $\delta>0$,

$$
\begin{aligned}
&P\{\frac{1}{\sqrt{n}}||\sum_{i=1}^{n}(|\eta_i|-1)\xi_i(z)||\ge\varepsilon\}\\
&\le \frac{4\mathrm{E}(|\eta_1|-1)^2}{n\varepsilon^2}\sum_{i=1}^{n}\mathrm{E}\Big\{||\xi_i(z)||^2 I\Big[\max_{1\le j\le i-1}|Y_j|\le n^{1/2+\delta}\Big]\Big\}\\
&\quad+\frac{4\mathrm{E}(|\eta_1|-1)^2}{n\varepsilon^2}\sum_{i=1}^{n}\mathrm{E}\Big\{||\xi_i(z)||^2 I\Big[\max_{1\le j\le i-1}|Y_j|> n^{1/2+\delta}\Big]\Big\}\\
&=: \Xi_1+\Xi_2. \qquad (6.36)
\end{aligned}
$$

If $\max_{1\le j\le i-1}|Y_j|\le n^{1/2+\delta}$, we have when n is large, for all $M>0$ and $|z|\le M$,

$$
\omega+b\varepsilon_j^2(1-z/n)=\omega+b[\varepsilon_j^2+2zY_{j-1}\varepsilon_j/n+(zY_{j-1}/n)^2]\ge(\omega+b\varepsilon_j^2)/2,\ j\in\mathbf{N},
$$

and

$$
||\xi_i(z)||\le Cn^{-1/2+\delta}(||h_i^{-1}(\nu_0)[h_i'(\nu_0)]^T||+1). \qquad (6.37)
$$

Thus,

$$\Xi_1 \le Cn^{-1+2\delta} E\{h_i^{-2}(\nu_0)[h_i'(\nu_0)]^T h_i'(\nu_0)\} =: C'n^{-1+2\delta}. \tag{6.38}$$

Since $\alpha > 1$, $Y_i = \sum_{j=1}^{i} \varepsilon_j$ is a martingale with zero-mean, and from the well-known martingale inequality (see Theorem 2.4 of Hall and Heyde (1980)) we have, as $x \to \infty$,

$$\begin{aligned} P\{\sup_{1\le i\le n} |Y_i| > 2n^{1/2+\delta}\} &\le \mathrm{E}[|Y_n| I(|Y_n| > n^{1/2+\delta})/n^{1/2+\delta}] \\ &\le \{\mathrm{E}|Y_n|^2/n^{1+2\delta}\}^{1/2}[P(|Y_n| > n^{1/2+\delta})]^{1/2} \le Cn^{-2\delta} \end{aligned} \tag{6.39}$$

Using $y/(1+y) < y^{\tau}$ for any $\tau > 0$ and $y > 0$ and some elementary computations, we can show $\mathrm{E}(||\xi_i(z)||^{2/\theta}) < \infty$ for any $\theta > 0$ and $|z| \le M$. This yields

$$\begin{aligned} &\mathrm{E}\{||\xi_i(z)||^2 I[\max_{1\le j\le i-1} |Y_j| > n^{1/2+\delta}]\} \\ &\quad \le \{\mathrm{E}(||\xi_i(z)||^{2/\theta})\}^{\theta}\{P[\max_{1\le j\le i-1} |Y_j| > n^{1/2+\delta}]\}^{1-\theta} \le Cn^{-2(1-\theta)\delta}. \end{aligned} \tag{6.40}$$

It follows that

$$\Xi_2 \le Cn^{-2(1-\theta)\delta}. \tag{6.41}$$

Divide $[-M, M]$ into $[2M/\Delta]+1$ subintervals $Q_i =: [z_{i-1}, z_i]$ with interval length Δ, such that $z_0 = -M$. By (6.38) and (6.41), it follows that

$$P\Big\{\sup_{z_r,\, 0\le r\le [2M/\Delta]+1} ||\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(|\eta_i| - 1)\xi_i(z_r)|| \ge \varepsilon\Big\} \le \frac{C(n^{-1+2\delta} + n^{-2(1-\theta)\delta})}{\Delta}. \tag{6.42}$$

Let $g_i(s,t) = \max\{||h_i^{-1}(\nu_0, s)h_i'(\nu_0, s)||, ||h_i^{-1}(\nu_0, t)h_i'(\nu_0, t)||\}$ and $\overline{Y}_i = \max_{1\le j\le i-1} |Y_j|$. Using $2|y| \le 1 + y^2$, it can be seen that

$$||\xi_i(s) - \xi_i(t)|| \le C|s - t|(1 + g_i(s,t))(\frac{\overline{Y}_i}{n} + \frac{C\overline{Y}_i^2}{n^2}).$$

Clearly, $(1/\sqrt{n})\sum_{i=1}^{n}(|\eta_i| - 1)|s - t|(1 + g_i(s,t))\overline{Y}_i^2/n^2 = o_p(|s - t|)$. Further, since $g_i(s,t)$ has all finite moments and $\mathrm{E}|\varepsilon_i|^{2\beta} < \infty$ for any $1 < \beta < \alpha$, we have $\mathrm{E}[g_i(s,t)\overline{Y}_n/n]^2 < \infty$. This implies

$$\begin{aligned} &P\Big\{\frac{1}{\sqrt{n}}\Big|\Big|\sum_{i=1}^{n}(|\eta_i| - 1)|s - t|(1 + g_i(s,t))\frac{\overline{Y}_i}{n}\Big|\Big| \ge \varepsilon\Big\} \\ &\quad \le C\varepsilon^{-2}n^{-1}|s - t|^2\sum_{i=1}^{n}\mathrm{E}\Big[g_i(s,t)\frac{\overline{Y}_n}{n}\Big]^2 = O(|s - t|^2). \end{aligned} \tag{6.43}$$

Thus, by taking $\Delta = n^{-\delta}$,

$$\frac{1}{\sqrt{n}}||\sum_{i=1}^{n}(|\eta_i|-1)(\xi_i(s)-\xi_i(t))|| = o_p(|s-t|). \tag{6.44}$$

By (6.42), and (6.44), we have $I_{2n}(z) = o_p(1)$ uniformly for all $z \in \mathbf{R}$. Because $n(1-\widehat{\phi}) = O_p(1)$, it follows that $I_{2n} = o_p(1)$. Similarly, we have $I_{3n} = o_p(1)$.

We now show that $I_{4n} = o_p(1)$. Let

$$I_{4n}(z) = -\frac{x}{2\sqrt{n}}\sum_{i=1}^{n}(|\varepsilon_i + \frac{zY_{i-1}}{n}| - |\varepsilon_i|)\frac{[h_i'(\nu_0, 1-z/n)]^T}{h_i^{3/2}}(\nu_0, 1-\frac{z}{n}).$$

From $|x-y| - |x| = -y\mathrm{sgn}(x) + 2\int_0^y [I(x \le t) - I(x \le 0)]\, dt$ for $x \ne 0$ and $\mathrm{sgn}(\varepsilon_i) = \mathrm{sgn}(\eta_i)$, we have

$$\begin{aligned} I_{4n}(z) &= -\frac{xz}{2n\sqrt{n}}\sum_{i=1}^{n} Y_{i-1}\mathrm{sgn}(\eta_i)\frac{[h_i'(\nu_0, 1-z/n)]^T}{h_i^{3/2}(\nu_0, 1-z/n)} \\ &\quad -\frac{x}{\sqrt{n}}\sum_{i=1}^{n}\int_0^{-zY_{i-1}/n}(I(\varepsilon_i \le t) - I(\varepsilon_i \le 0))\, dt \frac{[h_i'(\nu_0, 1-z/n)]^T}{h_i^{3/2}(\nu_0, 1-z/n)} \\ &=: I_{4n1}(z) + I_{4n2}(z). \end{aligned}$$

By a standard argument (see for example Li and Li (2009)), we have that

$$\begin{aligned} I_{4n2}(z) &= -\frac{xz^2 f(0)}{n^2\sqrt{n}}\sum_{i=1}^{n} Y_{i-1}^2 [h_i'(\nu_0, 1-\frac{z}{n})]^T h_i^{-3/2}(\nu_0, 1-\frac{z}{n}) h_i^{-1/2}(\nu_0) + o_p(1) \\ &= o_p(1) \end{aligned}$$

holds uniformly for all $z \le M, M > 0$. Similar to the argument for I_{2n}, we have $I_{4n1}(z) = o_p(1)$ holds uniformly for all $z \le M, M > 0$. Thus, $I_{4n} = o_p(1)$.

For I_{5n}, a Taylor expansion has

$$\Big(\sqrt{h_i(\nu_0, 1-\frac{z}{n})} - \sqrt{h_i(\nu_0)}\Big) = \frac{1}{2} z h_i^{-1/2}(\nu_0, 1-\frac{z^*}{n})(\frac{\partial h_i(\nu_0, 1-z^*/n)}{\partial z}),$$

where $\partial h_i(\nu_0, 1-z^*/n)/\partial z = (\partial h_i(\nu_0, 1-z/n)/\partial z)|z = z^*$, and z^* lies between 0 and z. Since $h_i(\nu_0, 1-z/n) = \sum_{j=1}^{i}(\omega_0 a_0^{j-1} + b_0 a_0^{j-1}(\varepsilon_{i-j} + zY_{i-j-1}/n)^2)$, it follows that

$$\frac{\partial h_i(\nu_0, 1-z/n)}{\partial z} = 2b_0 \sum_{j=1}^{i} a_0^{j-1}(\varepsilon_{i-j} + \frac{zY_{i-j-1}}{n})\frac{Y_{i-j-1}}{n}.$$

This gives that

$$\frac{x}{2\sqrt{n}}\sum_{i=1}^{n}(\sqrt{h_i(\nu_0,1-\frac{z}{n})}-\sqrt{h_i(\nu_0)})h_i'(\nu_0,1-\frac{z}{n})h_i^{-3/2}(\nu_0,1-\frac{z}{n})$$
$$=\frac{xz}{4\sqrt{n}}\sum_{i=1}^{n}h_i^{-2}(\nu_0)(\frac{\partial h_i(\nu_0,1-z^*n)}{\partial z})h_i'(\nu_0)+o_p(1)$$
$$=\frac{xz}{2\sqrt{n}}\sum_{i=1}^{n}h_i^{-2}(\nu_0)\Big(\sum_{j=1}^{i}b_0a_0^{j-1}\varepsilon_{i-j}\frac{Y_{i-j-1}}{n}\Big)h_i'(\nu_0)+o_p(1).$$

uniformly for all $z\in\mathbf{R}$. Thus,

$$I_{5n}=\frac{xn(\widehat{\phi}-1)}{2\sqrt{n}}\sum_{i=1}^{n}\Big(\sum_{j=1}^{i}b_0a_0^{j-1}\varepsilon_{i-j}\frac{Y_{i-j-1}}{n}\Big)[h_i'(\nu_0)]^Th_i^{-2}(\nu_0)+o_p(1).$$

Along the lines of Theorem 2.1 of Li and Li (2009), we have

$$\frac{1}{2\sqrt{n}}\sum_{i=1}^{n}\Big(\sum_{j=1}^{i}b_0a_0^{j-1}\varepsilon_{i-j}\frac{Y_{i-j-1}}{n}\Big)[h_i'(\nu_0)]^Th_i^{-2}(\nu_0)=\frac{1}{2n\sqrt{n}}\sum_{i=1}^{n}Y_{i-1}F.$$

This gives

$$I_{5n}=\frac{xFn(\widehat{\phi}-1)}{2n\sqrt{n}}\sum_{i=1}^{n}Y_{i-1}+o_p(1)$$

and completes the proof of (6.26).

Acknowledgement

We would like to thank the Co-Editor, an associate editor and two referees for many helpful comments and suggestions which lead to a substantial improvement of this paper. Research supported in part by grants from the HKSAR-RGC-General Research Fund: 400306 and 400408, the National Science Foundation of China (10801118), and the Ph.D. Programs Foundation of the Ministry of Education of China (200803351094).

Appendix: Auxiliary Lemmas

Preliminary lemmas needed to prove the main theorems are stated and discussed in this section. We first note that $\{\varepsilon_t\}$ of model (1.3) is a stationary sequence with regularly varying tails.

Lemma A.1. *Under conditions* H1, H2, *and* H3, *the following assertions hold.*

(a) *There exists a unique solution* $\alpha \in (0, k_0]$ *such that* $\mathrm{E}(a + b\eta_1^2)^\alpha = 1$.

(b) σ_t^2 *has a unique strictly stationary solution and* σ_t^2 *satisfies the regularly varying condition* $\lim_{x\to\infty} x^\alpha P\{\sigma_t^2 > x\} = c_0$, *where*

$$c_0 = \frac{\mathrm{E}\Big([\omega + (a + b\eta_t^2)\sigma_t^2]^\alpha - [(a + b\eta_t^2)\sigma_t^2]^\alpha\Big)}{\alpha \mathrm{E}\Big((a + b\eta_t^2)^\alpha \log^+(a + b\eta_t^2)\Big)}.$$

Proof. Observe that $\sigma_t^2 = \omega + a\sigma_{t-1}^2 + b\varepsilon_{t-1}^2 = \omega + (a + b\eta_{t-1}^2)\sigma_{t-1}^2$. By Theorem 4 of Kesten (1973) (see also Goldie (1991)), we conclude (a) and (b) except for the exact value of c_0. However, c_0 can be deduced via Theorem 4.1 of Goldie (1991).

Lemma A.2. *Let* X, Y *be independent random variables with* $\mathrm{E}|X|^\beta < \infty$ *for some* $\beta > 0$ *and* Y *satisfying the regularly varying condition* $\lim_{x\to\infty} x^\beta P(|Y| > x) = C$ *for* $C \geq 0$. *Then,* $\lim_{x\to\infty} x^\beta P(|XY| > x) = C\mathrm{E}|X|^\beta$.

Proof. This result can be easily proved by virtue of Proposition 3 of Breiman (1965). We thank a referee for pointing this out.

Lemma A.3. *Under the conditions of Theorem* 2.1*,*

$$P\{|\varepsilon_1| > x\} \sim c_0 \mathrm{E}|\eta_1|^{2\alpha} x^{-2\alpha} \quad \text{as} \quad x \to \infty.$$

Proof. By Lemma A.1, we have $P(\sigma_t > x) = P(\sigma_t^2 > x^2) \sim c_0 x^{-2\alpha}$ as $x \to \infty$. Clearly, $\mathrm{E}(a+b\eta_1^2)^{k_0} \log^+(a+b\eta_1^2) < \infty$, $a, b \geq 0$ and $\alpha \in (0, k_0]$ implies $\mathrm{E}|\eta_1|^{2\alpha} < \infty$. By virtue of Lemma A.2, we have the result.

Lemma A.4. *Suppose that* H1*,* H2*, and* H3 *are satisfied. Then* $\{\varepsilon_t\}$ *is a* β*-mixing stationary sequence with* $\beta_k = \mathrm{E}[\sup_{B\in\sigma(\varepsilon_t, t\geq k)} |P(B|\sigma(\varepsilon_s, s \leq 0)) - P(B)|] = O(\rho^k)$ *for some* $0 < \rho < 1$.

Proof. If $\mathrm{E}\log(a+b\eta_1^2) < 0$, we have that σ_t^2 has a stationary solution. Therefore, $\{\varepsilon_t\} = \{\sigma_t \eta_t\}$ is stationary. By Theorem 3 of Francq and Zakoïan (2006), we see that under conditions H1 and H3, $\{\sigma_t\}$ is a β-mixing process with an exponential decay. Since η_t are i.i.d. random variables, $\{\varepsilon_t\}$ satisfies the β-mixing condition with an exponential decay. That is, $\beta_k = E[\sup_{B\in\sigma(\varepsilon_t, t\geq k)} |P(B|\sigma(\varepsilon_s, s \leq 0)) - P(B)|] = O(\rho^k)$ for some $\rho < 1$. The desired conclusion thus follows.

Let $\{X_t\}$ be a strictly stationary sequence, $\{a_n\}$, $\{b_n\}$ be sequence of real numbers such that $\lim_{n\to\infty} nP\{|X_1| > a_n\} = 1$ and $b_n = a_n^2$. Let $S_{1n} = \sum_{i=1}^n X_i$ and $S_{2n} = \sum_{i=1}^n X_i^2$.

Lemma A.5. *Suppose the following hold*

1. *The finite-dimensional distributions of* $\{X_k\}$ *are regularly varying with index* $\alpha > 0$.
2. $\{X_t\}$ *satisfies the mixing condition* $\mathcal{A}(n)$: *there exists a sequence of positive integers* $\{r_n\}$ *such that* $r_n \uparrow \infty$, $k_n = [n/r_n] \uparrow \infty$ *as* $n \to \infty$, *and*

$$\mathrm{E}\exp\Big\{-\sum_{i=1}^{n} f(\frac{X_i}{a_n})\Big\} - \Big(\mathrm{E}\exp\Big\{-\sum_{i=1}^{r_n} f(\frac{X_i}{a_n})|Big\}\Big)^{k_n} \to 0$$

for any $f \in \mathcal{G}_b$, *where* $\mathcal{G}_b$ *is the collection of bounded non-negative step functions on* $[-\infty, 0) \cup (0, \infty]$.
3. $\lim_{k\to\infty} \limsup_{n\to\infty} P\Big(\sup_{k\le|i|\le r_n} |X_i| > a_n y \Big| |X_0| > a_n y\Big) = 0$, $y > 0$.

Then the following hold.

1. *The extremal index* γ *exists and*

$$\gamma = \lim_{n\to\infty} k_n P(\sup_{1\le i\le r_n} |X_i| > a_n) = \lim_{k\to\infty} \frac{\mathrm{E}\Big(|\theta_0|^\alpha - \sup_{1\le j\le k} |\theta_j|^\alpha\Big)_+}{\mathrm{E}|\theta_0|^\alpha},$$

where $\theta_j = X_j/(\max_{1\le 0\le i\le k} |X_i|)$, $j = 0, \ldots, k$, *are the* $(k+1)$*-dimensional random vectors with values in the unit sphere* $\mathcal{S}^k$.
2. *If* $\gamma > 0$, *the following hold.*
   (a) *If* δ_x *is a unit point measure at the point* x *and* $N_n = \sum_{i=1}^n \delta_{X_i/a_n}$, *then*

$$N_n \xrightarrow{d} N = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} \delta_{P_i Q_{ij}},$$

where $\sum_{i=1}^\infty \delta_{P_i}$ *is a Poisson process on* R_+ *with intensity measure* $v(dy) = \gamma\alpha y^{-\alpha-1}\,dy$. *Here* $\{\sum_{j=1}^\infty \delta_{Q_{ij}}\}$ *is a sequence of i.i.d. point processes with common distribution* Q *and independent of* $\{P_i\}$ *and*

$$Q(\cdot) = P\Big(\sum_{i=1}^{r_n} \delta_{X_i/(\sup_{1\le i\le r_n}|X_i|)} \in \cdot \Big| \sup_{1\le i\le r_n} |X_i| > a_n\Big).$$

   (b) *For* $\alpha \in (0, 1)$,

$$\Big(\frac{S_{1n}}{a_n}, \frac{S_{2n}}{b_n}\Big) \xrightarrow{d} \Big(\xi_\alpha, \xi_{\alpha/2}\Big) = \Big(\sum_{i=1}^{\infty}\sum_{j=1}^{\infty} P_i Q_{ij}, \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} P_i^2 Q_{ij}^2\Big).$$

(c) *For* $\alpha \in [1,2)$, *if for all* $\delta > 0$,

$$\lim_{y\to 0}\limsup_{n\to\infty} P\Big\{\Big|\sum_{i=1}^{n} X_i I(|X_i| \le a_n y) - \mathrm{E}X_i(|X_i| \le a_n y)\Big| > \delta a_n\Big\} = 0,$$

then

$$\Big(\frac{1}{a_n}\sum_{i=1}^{n}(X_i - \mathrm{E}X_i I(|X_i| \le a_n)), \frac{1}{b_n}\sum_{i=1}^{n}(X_i^2 - \mathrm{E}X_i^2 I(|X_i| \le a_n))\Big)$$
$$\xrightarrow{d} (\xi_\alpha, \xi_{\alpha/2}),$$

where ξ_α *(resp.,* $\xi_{\alpha/2}$*) is a stable variable given at* (b), *with index* α *(resp.,* $\alpha/2$*).*

Proof. (a) follows from Theorem 2.7 of Davis and Hsing (1995). Using (a), (b) and (c) can be shown similar to Theorem 3.1 of Davis and Hsing (1995), see also Mikosch and Stărică (2000).

Lemma A.6. *Let* $c_1 = c_0\mathrm{E}|\eta_1|^{2\alpha}$. *Under the conditions of Theorem* 2.1,

$$\Big(\frac{1}{(c_1 n)^{1/(2\alpha)}}\sum_{i=1}^{n}\varepsilon_i, \frac{1}{(c_1 n)^{1/\alpha}}\sum_{i=1}^{n}\varepsilon_i^2\Big) \xrightarrow{d} \Big(Z_{2\alpha}, Z_\alpha\Big), \alpha \in (0,1).$$

Proof. For the proof, it is enough to check that the conditions of Lemma A.5 are satisfied by $\{\varepsilon_t\}$. By Corollary 2.7 of Basrak, Davis and Mikosch (2002), for any finite integer k, $(\sigma_1^2, \ldots, \sigma_k^2)$ are regularly varying with index α. It follows that $(\sigma_1, \ldots, \sigma_k)$ are regularly varying with index 2α. Then using an induction method similar to that in the proof of (B) in Corollary 3.5 of Basrak, Davis and Mikosch (2002), we conclude that $(\varepsilon_1, \ldots, \varepsilon_k) = (\sigma_1\eta_1, \ldots, \sigma_k\eta_k)$ are regularly varying with index 2α. Condition 2 of Lemma A.5 (the mixing condition), follows from Lemma A.4. Let $a_n = (c_1 n)^{1/(2\alpha)}$ and $A_t = a + b\eta_{t-1}^2$. Since $\sigma_t^2 = \omega + (a + b\eta_{t-1}^2)\sigma_{t-1}^2$, it follows that $\sigma_t^2 = \omega + \prod_{j=1}^{t} A_j\sigma_0^2 + \sum_{j=1}^{t}\prod_{m=j+1}^{t}\omega A_m$. This implies that for any $0 < \delta < 1$ and $M > 0$,

$$P\Big(\sup_{k\le|i|\le r_n}\Big|\varepsilon_i| > a_n y\Big||\varepsilon_0| > a_n y\Big) = P\Big(\sup_{|i|}\varepsilon_i^2 > a_n^2 y^2||\varepsilon_0| > a_n y\Big)$$
$$\le P\Big(\sup_{|i|}\prod_{j=1}^{i} A_j\sigma_0^2\eta_i^2 > \frac{a_n^2 y^2}{4}\Big||\varepsilon_0| > a_n y\Big) + P\Big(\sup_{|i|}\sum_{j=1}^{i}\prod_{m=j+1}^{i}\omega A_m\eta_i^2 > \frac{a_n^2 y^2}{4}\Big)$$
$$\le P\Big(\sup_{|i|}\prod_{j=1}^{i} A_j\sigma_0^2\eta_i^2 > \frac{a_n^2 y^2}{4}, |\eta_0| \le M\Big||\sigma_0\eta_0| > a_n y\Big)$$

$$+P(\sup_{|i|}\prod_{j=1}^{i}A_j\sigma_0^2\eta_i^2>\frac{a_n^2y^2}{4},|\eta_0|>M\Big||\varepsilon_0|>a_ny)+Cr_n(a_ny)^{-2\delta}\sum_{j=1}^{\infty}(\mathrm{E}|A_1|^{\delta})^j$$

$$\le Cny^{2\alpha}\Big[\sum_{i=k}^{r_n}y^{-2\delta}\mathrm{E}(\prod_{j=1}^{i}(A_j\eta_i^2)^{\delta})\mathrm{E}(\frac{\sigma_0}{a_n})^{2\delta}I(\sigma_0>\frac{a_ny}{M})\Big]$$

$$+Cny^{2\alpha}[P(\sigma_0|\eta_0|>a_ny,|\eta_0|>M)]+Cr_na_n^{-2\delta}y^{-2\delta}$$

$$\le M^{2\alpha-2\delta}\sum_{i=k}^{r_n}(\mathrm{E}A_1^{\delta})^i+C\mathrm{E}[|\eta_0|I(|\eta_0|>M)]+Cr_na_n^{-2\delta}y^{-2\delta}\to 0\,,$$

by taking $r_n=o(n^{\delta/\alpha})$, $n\to\infty$, then letting $M\to\infty$ and finally $k\to\infty$, where $\sup_{|i|}$ is the abbreviation of $\sup_{k\le|i|\le r_n}$ and the last inequality follows by Karamata's Theorem and Lemma A.2.

Along the lines of Theorem 4.1 of Mikosch and Stărică (2000), we have

$$\gamma=\lim_{m\to\infty}\frac{\mathrm{E}\Big(|\eta_1|^{2\alpha}-\max_{j=2,\dots,m+1}|\eta_j|^{2\alpha}\prod_{i=1}^{j-1}(a+b\eta_i^2)^{\alpha}\Big)_+}{\mathrm{E}|\eta_1^{2\alpha}}>0.$$

Since ε_t is regularly varying with index 2α, by Lemma A.5 it suffices to show that, as $\alpha\in[1/2,1)$, for all $\delta>0$, $\lim_{y\to 0}\limsup_{n\to\infty}P\{|\sum_{i=1}^n\varepsilon_iI(|\varepsilon_i|\le a_ny)-\mathrm{E}\varepsilon_iI(|\varepsilon_i|\le a_ny)|>\delta a_n\}=0$. Let $\mathcal{F}_t=\sigma(\eta_i,i\le t)$. By the symmetric assumption on η_t, we have $\mathrm{E}[\sum_{i=1}^n\varepsilon_iI(|\varepsilon_i|\le a_ny)]=0$ and $X_{i1}:=\varepsilon_iI(|\varepsilon_i|\le a_ny)=\varepsilon_iI(|\varepsilon_i|\le a_ny)-\mathrm{E}[\varepsilon_iI(|\varepsilon_i|\le a_ny)|\mathcal{F}_{i-1}]$. Furthermore, $\{X_{i1}\}$ is a sequence of martingale differences. By Bahr-Esseen's inequality and Karamata's Theorem, it follows that as $n\to\infty$ and $y\to 0$, for any $\alpha\in[1/2,1)$, $P\{|\sum_{i=1}^nX_{i1}|>\delta a_n/2\}\le 4n(\delta a_n)^{-2}\mathrm{E}X_{11}^2\le Cy^{2-2\alpha}\to 0$. Lemma A.6 follows.

References

Basrak, B., Davis, R. A. and Mikosch, T. (2002). Regular variation of GARCH processes. *Stochastic Process. Appl.* **99**, 95-115.

Bernard, J., Khalaf, L., Kichian, M. and McMahon, S. (2008). Forecasting commodity prices: GARCH, jumps and mean reversion. *J. Forecasting* **27**, 279-291.

Billingsley, P. (1999). *Convergence of Probability Measures, 2nd Ed.* Wiley, New York.

Breiman, L. (1965). On some limit theorems similar to the arc-sin law. *Theo. Probab. Appl.* **10**, 323-331.

Chan, N. H. (2009). Time series with roots on or near the unit circle. In *Handbook of Financial Time Series*, 695-707. (Edited by T. G. Andersen, R. A. Davis, J. P. Kreiss and T. Mikosch). Springer-Verlag, New York.

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

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

Danielsson, J., de Hann, L., Peng, L. and de Vries, C. (2001). Using a bootstrap method to choose the optimal sample fraction in tail index estimation. *J. Multivariate Anal.* **76**, 226-248.

Davis, R. A. and Hsing, T. (1995). Point process and partial sum convergence for weakly dependent random variables with infinite variance. *Ann. Probab.* **23**, 879-917.

Davis, R. A. and Mikosch, T. (1998). The sample autocorrelations of heavy-tailed processes with applications to ARCH. *Ann. Statist.* **26**, 2049-2080.

Drees, H. (2000). Weighted approximations of tail processes for β mixing random variables. *Ann. Appl. Probab.* **10**, 1274-1301.

Francq, C. and Zakoïan, J. M., (2004). Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes. *Bernoulli* **10**, 605-37.

Francq, C. and Zakoïan, J. M. (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. *Ann. Appl. Probab.* **1**, 126-166.

de Haan, L. and Resnick, S. I. (1998). On asymptotic normality of Hill estimator. *Stochastic Models* **14**, 849-867.

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

Hall, P. and Yao, Q. W. (2003). Inference in ARCH and GARCH models. *Econometrica* **71**, 285-317.

Jakubowski, A. (1996). Convergence in various topologies for stochastic integrals driven by semimartingales. *Ann. Probab.* **24**, 2141-2153.

Jakubowski, A. (1997). A non-Skorohod topology on Skorohod space. *Eletronic J. Probab.* **2**, 1-21.

Kesten, H. (1973). Random difference equations and renewal theory for products of random matrices. *Acta Math.* **131**, 207-248.

Li, G. D. and Li, W. K. (2009). Least absolute deviation estimation for unit root processes with GARCH errors. *Econom. Theory* **25**, 1208-1227.

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

Ling, S. and Li, W. K. (2003). Asymptotic inference for unit-root with GARCH (1,1) errors. *Econom. Theory* **19**, 541-564

Ling, S., Li, W. K. and McAleer, M. (2003). Estimation and testing for a unit-root process with GARCH (1,1) errors. *Econom. Rev.* **22**, 179-202.

Lumsdaine, R. L. (1996). Consistence and asymptotic normality of the quasi-maximum likelihood estimator in IGARCH(1, 1) and covariance stationary GARCH(1, 1) models. *Econometrica* **64**, 575-596.

Mikosch, T. and Stărică, C. (2000). Limit theory for the sample autocorrelations and extremes of a GARCH (1,1) process. *Ann. Statist.* **28**, 1427-1451.

Nelson, D. B. (1990). Stationary and persistence in GARCH (1, 1) models. *Econom. Theory* **6**, 318-334.

Peng, L. and Yao, Q. W. (2003). Least absolute deviation estimation for ARC and GARCH models. *Biometrika* **90**, 967-975.

Pollard, D. (1991). Asymptotic for least absolute deviation regression estimators. *Econom. Theory*, **7**, 186-199.

Seo, B. (1999). Distribution theory for unit root tests with conditional heteroskedasticity. *J. Econom.* **91**, 113-144.

Wang, G. W. (2006). A note on unit-root tests with heavy-tailed GARCH errors. *Statist. Probab. Lett.* **76**, 1075-1079.

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

Department of Statistics, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong.

E-mail: nhchan@sta.cuhk.edu.hk

Department of Mathematics, Zhejiang University (Yuquan Campus), Hangzhou 310027, China.

E-mail: rmzhang@zju.edu.cn

(Received September 2008; accepted July 2009)