Inference for Tail Index of GARCH(1,1) Model and AR(1) Model with ARCH(1) Errors

Rongmao Zhang

Department of Mathematics, Zhejiang University

Chenxue Li

Department of Mathematics and Statistics, Georgia State University

and Liang Peng

Department of Risk Management and Insurance, Georgia State University

Abstract

For a GARCH(1,1) sequence or an AR(1) model with ARCH(1) errors, it is known that the observations have a heavy tail and the tail index is determined by an estimating equation. Therefore, one can estimate the tail index by solving the estimating equation with unknown parameters replaced by quasi maximum likelihood estimation (QMLE), and profile empirical likelihood method can be employed to effectively construct a confidence interval for the tail index . However, this requires that the errors of such a model have at least finite fourth moment to ensure asymptotic normality with $\sqrt{n}$ rate of convergence and Wilks theorem. In this paper, we show that the finite fourth moment can be relaxed by employing a least absolute deviations estimate (LADE) instead of QMLE for the unknown parameters by noting that the estimating equation for determining the tail index is invariant to a scale transformation of the underlying model. The proposed tail index estimators have a normal limit with $\sqrt{n}$ rate of convergence and Wilks theorem holds for the proposed profile empirical likelihood methods. Hence a confidence interval for the tail index can be obtained without estimating any additional quantities such as asymptotic variance.

Keywords: AR model, empirical likelihood, GARCH sequence, tail index.

1 Introduction

A large number of empirical studies show that many financial data series, such as exchange rate returns and stock indices, often exhibit skewness and heavy tails (see Taylor (2005)). The heaviness of tails determines some unusual asymptotic behavior of sample covariance functions, sample correlation functions and extremes of the underlying sequence; see Davis and Resnick (1985, 1986) for ARMA processes, Mikosch and Stărică (2000) and Basrak, Davis and Mikosch (2002) for GARCH sequences, Davis and Mikosch (1998) and De Haan, Resnick, Rootzén and de Vries (1989) for ARCH models, Borkovec (2000, 2001) for an AR(1) process with ARCH(1) errors, and Davis and Resnick (1996) and Resnick and Van den Berg (2000) for bilinear time series. When the sequence follows from a time series model, heavy tailed errors play an important role in deriving the asymptotic limit of parameters estimation; see Hall and Yao (2003) for the study of quasi maximum likelihood estimation (QMLE) for a GARCH process, Lange (2011) and Zhang and Ling (2015) for the study of least squares estimation for AR-GARCH models. Some robust inference procedures for heavy-tailed GARCH models can be found in Hill (2015) and Hill and Prokhorov (2016). The tail index also plays an important role in testing structural changes in stock prices (see Quinton, Fan and Phillips (2001)) and calculating financial risk measures such as Value-at-Risk and expected shortfall (see Wagner and Marsh (2005)). Therefore inference for the tail index is useful in understanding and modeling time series data.

For a GARCH(p,q) sequence, i.e.,

$$Y_t = \sigma_t \varepsilon_t, \quad \sigma_t^2 = w + \sum_{i=1}^{p} a_i \sigma_{t-i}^2 + \sum_{j=1}^{q} b_j Y_{t-j}^2,$$

where $w > 0, a_i \geq 0, b_i \geq 0$ are unknown parameters, and $\varepsilon_t' s$ are independent and identically distributed random variables with zero mean and variance one, Basrak, Davis and Mikosch (2002) showed that, under some conditions, there is $\alpha > 0$ such that $\lim_{x\to\infty} x^\alpha P(|Y_t| > x) \in (0, \infty)$ by using results in Kesten (1973) for random difference equations. For estimating the tail index α, one could simply employ the Hill's estimator (see Hill (1975)) defined as

$$\tilde{\alpha}(k) = \{\frac{1}{k}\sum_{i=1}^{k} \log \frac{Y_{n,n-i+1}}{Y_{n,n-k}}\}^{-1}, \tag{1.1}$$

where $Y_{n,1} < \cdots < Y_{n,n}$ denote the order statistics of $Y_1, \cdots, Y_n$, and $k = k(n) \to \infty$ and $k/n \to 0$ as $n \to \infty$. Although the Hill's estimator has been studied extensively for independent data, existing research on dependent data such as m-dependence or β-mixing can be found in Hsing (1991), Resnick and Stărică (1998) and Drees (2000). However the choice of the sample fraction k for dependent data is much more complicated than that for independent data. Indeed, as far as we are aware, there is no data-driven method for choosing k for dependent data although several methods are available for independent data. For the case of an ARMA sequence, one can apply the Hill's estimator to either the sequence itself or the estimated errors since heavy tailed errors imply that the observations have a heavy tail with the same tail index (see Resnick and Stărică (1997) and Ling and Peng (2004)).

When $p = q = 1$, i.e., a GARCH(1,1) sequence, the tail index α is determined by an estimating equation (see Section 2 for details). In this case, the tail index can be estimated by using all observations rather than a small fraction of upper order statistics as Hill's estimator, and so the resulted estimator has a faster rate of convergence than the Hill's estimator and does not need to choose the sample fraction k. Asymptotic limit was first derived in Mikosch and Stărică (2000), and later its asymptotic variance was corrected by Berkes, Horváth and Kokoszka (2003). Since the asymptotic variance is very complicated, interval estimation relies on bootstrap method, which is computationally extensive due to the fact that one has to resample from estimated errors and refit the GARCH(1,1) model. Moreover, it is known that the performance of bootstrap method for nonpivotal statistics is not good in general. Therefore Chan, Peng and Zhang (2012) proposed a profile empirical likelihood method to construct a confidence interval for the index α by using score equations derived from QMLE, which requires finite fourth moment for ϵ_t. For an AR(1) model with ARCH(1) errors, which is sometimes called a double AR process in the literature, the tail index is determined by an estimating equation too. Therefore Chan, Li, Peng and Zhang (2013) derived the asymptotic limit of an estimator for the index based on an estimation equation with QMLE, and proposed a profile empirical likelihood method to construct a confidence interval without estimating the asymptotic variance explicitly. These results require the errors to have a finite fourth moment as well.

Motivated by the analysis of the exchange rates between Hong Kong dollar and US dollar in Zhu and Ling (2015), the purpose of this paper is to propose a robust method to estimate the tail index, which allows the errors to have an infinite fourth moment. More specifically, by noting that the estimating equation for determining the tail index is invariant to a scale transformation of the studied models, we propose to first estimate the unknown parameters by a least absolute deviations estimate (LADE) and then to estimate the tail index by the estimating equation. This leads to a tail index estimator with the $\sqrt{n}$ rate of convergence and asymptotic normality without requiring a finite fourth moment of errors. Since the asymptotic variance of the proposed tail index estimator is too complicated, we further propose to employ the profile empirical likelihood method to construct a confidence interval, which does not require to estimate the asymptotic variance explicitly. Unlike existing methods in Berkes, Horváth and Kokoszka (2003), Chan, Peng and Zhang (2012) and Chan, Li, Peng and Zhang (2013), the proposed methods in this paper not only relax the moment conditions of errors (see

Section 2), but also perform well because LADE is more robust than QMLE (see the empirical study in Section 3).

The paper is organized as follows. The proposed methodologies and their asymptotic results are presented in Section 2. Section 3 presents a simulation study and a data analysis. Some conclusions are given in Section 4. All technical proofs are put in Section 5.

2 Models, Methodologies and Theoretical Results

2.1 Heavy Tailed GARCH(1, 1) Model

Being a benchmark of GARCH family, GARCH(1, 1) model is simply used to capture the heteroscedastic and heavy-tailed phenomena in financial returns, which is defined as

$$Y_t = \sigma_t^* \varepsilon_t^*,\ (\sigma_t^*)^2 = \omega^* + a^*(\sigma_{t-1}^*)^2 + b^* Y_{t-1}^2, \tag{2.1}$$

where $\omega^* > 0$, $a^* \geq 0$, $b^* \geq 0$ and $\{\varepsilon_t^*\}$ is a sequence of independent and identically distributed random variables with zero mean and unit variance. For some general studies and applications of GARCH models in financial econometrics, we refer to Tayor (2005) and Francq and Zakoïan (2010). For model (**??**), it is known that, under some conditions, Y_t has a heavy tail with index $\alpha > 0$. More specifically, it follows from Basrak, Davis and Mikosch (2002) that

$$P(|Y_t| > x) = cx^{-\alpha}\{1 + o(1)\} \text{ for some } c > 0 \text{ as } x \to \infty, \tag{2.2}$$

and the tail index α is determined by

$$\mathrm{E}\{a^* + b^*(\varepsilon_t^*)^2\}^{\alpha/2} = 1. \tag{2.3}$$

Note that equations (**??**) and (**??**) can be derived from Kesten (1973) and Goldie (1991) too. When $\mathrm{E}|\varepsilon_t^*|^\delta < \infty$ for some $\delta > \max\{4, 2\alpha\}$, one can estimate the nuisance parameters $\theta^* = (\omega^*, a^*, b^*)^T$ by the QMLE (say $\hat{\theta}^* = (\hat{\omega}^*, \hat{a}^*, \hat{b}^*)^T$) and then estimate the tail index α by solving the following estimating equation:

$$\frac{1}{n}\sum_{t=1}^{n}\{\hat{a}^* + \hat{b}^*(\hat{\varepsilon}_t^*)^2\}^{\alpha/2} = 1, \tag{2.4}$$

where $\hat{\varepsilon}_t^* = Y_t/\hat{\sigma}_t^*$ and $\hat{\sigma}_t^*$ is an estimator of σ_t^* with θ^* being replaced by $\hat{\theta}^*$, see Berkes, Horváth and Kokoszka (2003) for the asymptotic distribution of the above estimator and Chan, Peng and Zhang (2012) for a profile empirical likelihood inference based on the above estimation procedure.

Note that $\delta > 4$ ensures that the QMLE $\hat{\theta}^*$ has a normal limit, and $\delta > 2\alpha$ ensures the asymptotic normality for estimating α via solving $n^{-1}\sum_{t=1}^{n}\{a^* + b^*(\epsilon_t^*)^2\}^{\alpha/2} = 1$. Therefore the condition of $\delta > 2\alpha$ can not be relaxed. However, we may be able to allow $\mathrm{E}|\varepsilon_t^*|^4 = \infty$ by using some different estimate for parameters θ^* such as least absolute deviations estimate, which generally requires to reparameterize model (**??**). Issues on reparameterization for GARCH sequences are discussed in Fan, Qi and Xiu (2014).

Assume the unknown median of $(\varepsilon_t^*)^2$ is $d > 0$ and put $\varepsilon_t = \varepsilon_t^*/\sqrt{d}$. Then the median of $\log \varepsilon_t^2$ becomes $\log\{\mathrm{median}((\varepsilon_t^*)^2/d)\} = 0$. Furthermore, model (**??**) and equation (**??**) can be written as

$$Y_t = \sigma_t \varepsilon_t, \quad \sigma_t^2 = \omega + a\sigma_{t-1}^2 + bY_{t-1}^2 \tag{2.5}$$

and

$$E\{a + b\epsilon_t^2\}^{\alpha/2} = 1, \tag{2.6}$$

where $\sigma_t = \sqrt{d}\sigma_t^*$, $\omega = d\omega^*$, $a = a^*$ and $b = db^*$. It is clear that the estimating equation for the tail index α remains unchanged. Therefore we propose to first apply the least absolute deviations estimate in Peng and Yao (2003) to (**??**) and then to estimate the tail index α via (**??**) so as to relax the moment condition on ε_t^* or equivalently on ε_t.

More specifically, for any $\theta = (\omega, a, b)^T$, by the recursion of (**??**), the conditional variance $\sigma_t^2 = \sigma_t^2(\theta)$ can be represented as follows:

$$\sigma_t^2(\theta) \quad = \quad \omega + a\sigma_{t-1}^2(\theta) + bY_{t-1}^2 = \frac{\omega(1-a^t)}{1-a} + \sum_{k=0}^{t-1} ba^k Y_{t-1-k}^2 + a^t\sigma_0^2(\theta). \tag{2.7}$$

Thus, given the observations $\{Y_1, Y_2, \cdots, Y_n\}$ and the initial value Y_0, we can estimate θ by the following LADE:

$$\hat{\theta}_{inital} = \arg\min_{\theta} \sum_{t=1}^{n} |\log Y_t^2 - \log \sigma_t^2(\theta)|. \tag{2.8}$$

However, since $\sigma_0^2(\theta)$ depends on the unobserved sample path $Y_{-1}, Y_{-2}, \ldots$, one cannot use the above expression of $\sigma_t^2(\theta)$ in practice. Instead, we consider the LADE based on a truncated version of $\sigma_t(\theta)$, which is

$$\hat{\theta} = \arg\min_{\theta} \sum_{t=1}^{n} |\log Y_t^2 - \log \bar{\sigma}_t^2(\theta)|, \tag{2.9}$$

where $\bar{\sigma}_t^2(\theta) = \omega(1-a^t)/(1-a) + b\sum_{0\le k\le t-1} a^k Y_{t-k-1}^2$. Using this LADE $\hat{\theta}$, we estimate α by solving

$$\frac{1}{n}\sum_{t=1}^{n}(\hat{a} + \hat{b}\bar{\varepsilon}_t^2(\hat{\theta}))^{\alpha/2} = 1,$$

where $\bar{\varepsilon}_t^2(\hat{\theta}) = Y_t^2/\bar{\sigma}_t^2(\hat{\theta})$. Denote this estimator by $\hat{\alpha}$. For deriving the asymptotic limit of $\hat{\alpha}$, we need some regularity conditions:

Condition 1. $\mathrm{E}\log(a_0^* + b_0^*(\epsilon_t^*)^2) < 0$ (i.e., $\mathrm{E}\log(a_0 + b_0\epsilon_t^2) < 0$) and $\mathrm{E}|\varepsilon_t^*|^{\delta_0} < \infty$ (i.e., $\mathrm{E}|\varepsilon_t|^{\delta_0} < \infty$) for some $\delta_0 > \max\{2, 2\alpha_0\}$, where $\theta_0 = (\omega_0, a_0, b_0)^T$, $\theta_0^* = (\omega_0^*, a_0^*, b_0^*)^T$ and α_0 denote the true values of θ, θ^* and α respectively.

Condition 2. $(\varepsilon_t^*)^2$ has an unknown median $d > 0$ and a continuous density at d, i.e., $\log\{\varepsilon_t^2\}$ has median zero and its density $f(x)$ is continuous at zero.

Remark 2.1. $\mathrm{E}\log(a_0 + b_0\epsilon_t^2) < 0$ *in Condition 1 is a sufficient and necessary condition for the existence of a stationary solution of* σ_t^2 *(see Nelson (1990)). Further, Condition 1 and (**??**) imply that* b_0 *can not be zero, as a result, we have* $a_0 < 1$. *Condition 2 is a standard condition for a LADE, which is the same as that in Peng and Yao (2003).*

Remark 2.2. *When* $a_0 + b_0 < 1$, *it is known that* Y_t *has a finite variance (see Fan and Yao (2003)), i.e.,* $\alpha_0 > 2$. *Therefore Condition 1 implies* ϵ_t^* *has a finite fourth moment* *in case of $a_0+b_0<1$. In order to consider the case of infinite fourth moment for errors, one has to study the case of $a+b\geq 1$.*

Theorem 2.1. *Assume Conditions 1 and 2 hold for model (??). Then, as $n\to\infty$*

$$\sqrt{n}(\hat{\alpha}-\alpha_0)\xrightarrow{d}N(0,\gamma_{\alpha_0}^2), \tag{2.10}$$

where

$$\begin{aligned}\gamma_{\alpha_0}^2 &= \{4A_0^2f^2(0)\}^{-1}(\mu_1,\mu_2,\mu_3)\Omega^{-1}(\mu_1,\mu_2,\mu_3)^T+4\{A_0^2\}^{-1}\mathrm{E}[(a_0+b_0\varepsilon_1^2)^{\frac{\alpha_0}{2}}-1]^2\\ &\quad +2\{A_0^2f(0)\}^{-1}(\mu_1,\mu_2,\mu_3)\Omega^{-1}\mathrm{E}\{A(1)[(a_0+b_0\varepsilon_1^2)^{\frac{\alpha_0}{2}}-1]\}\end{aligned}$$

with

$$\begin{cases} A_0=\mathrm{E}[(a_0+b_0\varepsilon_1^2)^{\frac{\alpha_0}{2}}\log(a_0+b_0\varepsilon_1^2)],\\ e_0=\alpha_0\mathrm{E}[(a_0+b_0\varepsilon_1^2)^{\alpha_0/2-1}\varepsilon_1^2],\\ \mu_1=-\frac{b_0e_0}{2}\mathrm{E}\frac{\partial\log\sigma_1^2(\theta_0)}{\partial w},\\ \mu_2=\alpha_0\mathrm{E}[(a_0+b_0\varepsilon_1^2)^{\frac{\alpha_0}{2}-1}]-\frac{b_0e_0}{2}\mathrm{E}\frac{\partial\log\sigma_1^2(\theta_0)}{\partial a},\\ \mu_3=e_0-\frac{b_0e_0}{2}\mathrm{E}\frac{\partial\log\sigma_1^2(\theta_0)}{\partial b},\\ \Omega=\mathrm{E}[A(1)A^T(1)],\\ A(t)=\left(\frac{\partial(\log\sigma_t^2(\theta))}{\partial\omega},\frac{\partial(\log\sigma_t^2(\theta))}{\partial a},\frac{\partial(\log\sigma_t^2(\theta))}{\partial b}\right)^T sgn\{\log(\varepsilon_t^2)\},\end{cases}$$

and sgn denotes the sign function.

Remark 2.3. *Although the moment condition on errors depends on the unknown parameter α_0, this can be checked when the error distribution has heavy tails. More specifically one can simply compute the Hill's estimator based on estimated errors via quasi maximum likelihood estimators for parameters in the GARCH(1,1) model and then compare it with $\max(2,2\hat{\alpha})$; see the data analysis in Section 3.*

To construct a confidence interval for the tail index α, an obvious approach is to estimate the asymptotic variance $\gamma_{\alpha_0}^2$. Due to the complexity of this asymptotic variance, one can simply employ a naive bootstrap method. However bootstrapping nonpivotal statistics is inefficient in general. Bootstrap method for a time series model is computationally intensive since one has to resample from the estimated errors and refit the time series model. Alternatively, we seek an empirical likelihood method to bypass estimating the asymptotic variance. A direct application of the empirical likelihood method to equation (**??**) with θ replaced by $\hat{\theta}$ cannot capture the variance of the plug-in estimator $\hat{\theta}$ since the asymptotic variance $\gamma_{\alpha_0}^2$ of the tail index estimator $\hat{\alpha}$ really depends on the asymptotic variances of $\hat{\theta}$. Hence, Wilks theorem fails for such a direct application of an empirical likelihood method. Instead we propose the following profile empirical likelihood method.

Note that the proposed LADE is a solution to the score equations

$$\sum_{t=1}^{n} \bar{Z}_{t,j}(\theta) = 0 \quad \text{for} \quad j = 2, 3, 4,$$

where

$$\begin{cases} \bar{Z}_{t,2}(\theta) = \big(\partial(\log \bar{\sigma}_t^2(\theta))/\partial\omega\big)\, \mathrm{sgn}\{\log(Y_t^2/\bar{\sigma}_t^2(\theta))\}, \\ \bar{Z}_{t,3}(\theta) = \big(\partial(\log \bar{\sigma}_t^2(\theta))/\partial a\big)\, \mathrm{sgn}\{\log(Y_t^2/\bar{\sigma}_t^2(\theta))\}, \\ \bar{Z}_{t,4}(\theta) = \big(\partial(\log \bar{\sigma}_t^2(\theta))/\partial b\big)\, \mathrm{sgn}\{\log(Y_t^2/\bar{\sigma}_t^2(\theta))\}. \end{cases}$$

It follows from (**??**) that θ and α can be estimated simultaneously by solving the following equations

$$\sum_{t=1}^{n} \bar{Z}_{t,j} = 0 \quad \text{for} \quad j = 1, 2, 3, 4,$$

where $\bar{Z}_{t,1} := \bar{Z}_{t,1}(\theta, \alpha) = \{a + bY_t^2/\bar{\sigma}_t^2(\theta)\}^{\alpha/2} - 1$. This simultaneous estimation procedure motivates us to apply the empirical likelihood method to the above four equations and then profile the nuisance parameters θ. This is the so-called profile empirical likelihood method based on estimating equations proposed by Qin and Lawless (1994).

Put $\bar{Z}_t(\theta, \alpha) = (\bar{Z}_{t,1}^T(\theta, \alpha), \bar{Z}_{t,2}(\theta), \bar{Z}_{t,3}(\theta), \bar{Z}_{t,4}(\theta))^T$ for $t = 1, \ldots, n$, and define the empirical likelihood function of (θ, α) as

$$L(\theta, \alpha) = \sup\{\prod_{t=1}^{n}(np_t) : p_1 \geq 0, \ldots, p_n \geq 0, \sum_{t=1}^{n} p_t = 1, \sum_{t=1}^{n} p_t \bar{Z}_t(\theta, \alpha) = 0\}.$$

By virtue of the Lagrange multipliers, it is clear that $p_t = n^{-1}\{1 + \lambda^T \bar{Z}_t(\theta, \alpha)\}^{-1}$ for $t = 1, \ldots, n$ and

$$l(\theta, \alpha) := -2 \log L(\theta, \alpha) = 2 \sum_{t=1}^{n} \log\{1 + \lambda^T \bar{Z}_t(\theta, \alpha)\},$$

where $\lambda = \lambda(\theta, \alpha)$ satisfies

$$\sum_{t=1}^{n} \frac{\bar{Z}_t(\theta, \alpha)}{1 + \lambda^T \bar{Z}_t(\theta, \alpha)} = 0. \tag{2.11}$$

Since we are interested in the tail index α, we consider the profile empirical likelihood ratio $l_p(\alpha) = l(\tilde{\theta}(\alpha), \alpha)$, where $\tilde{\theta}(\alpha) = \arg\min_\theta l(\theta, \alpha)$. Next theorem shows that Wilks theorem holds for the proposed profile empirical likelihood method.

Theorem 2.2. *Under conditions of Theorem* **??**, *the random variable* $l_p(\alpha_0)$ *converges in distribution to* $\chi^2(1)$ *as* $n \to \infty$.

Corollary 2.1. *For any* $0 < \xi < 1$, *let* $\chi^2_{1,\xi}$ *denote the* ξ*-th quantile of a* $\chi^2(1)$ *random variable and define the empirical likelihood confidence interval with level* ξ *as* $I_\xi = \{\alpha | \, l_p(\alpha) \leq \chi^2_{1,\xi}\}$. *Then, under conditions of Theorem* **??**, $P(\alpha_0 \in I_\xi) \longrightarrow \xi$ *as* $n \to \infty$.

Remark 2.4. *It is possible to develop similar estimation procedure and empirical likelihood method as above by replacing the LADE by other estimators such as those in Berkes and Horváth (2004).*

2.2 AR(1) with heavy tailed ARCH(1) noise

In this subsection, we study another time series model called the first-order autoregressive model (AR(1)) with autoregressive conditional heteroskedastic errors of order one (ARCH(1)), which is defined as

$$Y_t = a^* Y_{t-1} + \sqrt{\omega^* + b^* Y_{t-1}^2}\, \varepsilon_t^*, \tag{2.12}$$

where $\{\varepsilon_t^*\}$ is a sequence of independent and identically distributed random variables with zero mean and unit variance, $a^* \in \mathbb{R}$, $\omega^* > 0$ and $b^* > 0$. This model is also called a double AR model in the literature. Throughout this subsection, we assume model (**??**) satisfies the following regularity conditions:

Condition A. $\mathrm{E}\log(|a^* + \sqrt{b^*}\varepsilon_1^*|) < 0$;

Condition B. ε_t^* has a symmetric, positive and continuous Lebesgue density on $\mathbb{R}$.

Under Conditions A and B, it is known that Y_t has a heavy tail with index $\alpha > 0$, which is determined by

$$\mathrm{E}(|a^* + \sqrt{b^*}\varepsilon_t^*|^\alpha) = 1, \tag{2.13}$$

see Borkovec and Klüppelberg (2001) for details. Therefore, one can estimate α by solving

$$\frac{1}{n}\sum_{t=1}^{n}|\hat{a}^* + \sqrt{\hat{b}^*}\hat{\varepsilon}_t^*|^\alpha = 1,$$

where $\hat{a}^*, \hat{b}^*, \hat{\varepsilon}_t^*$ are some estimators for $a^*, b^*, \varepsilon_t^*$, respectively. Indeed Chan, Li, Peng and Zhang (2013) proposed to first employ the QMLE in Ling (2004) to estimate α and then to apply a profile empirical likelihood method for interval estimation, where finite fourth moment of ε_t^* is required to ensure a normal limit. Here we propose to relax this moment condition by using the weighted least absolute deviations estimate in Chan and Peng (2005) as follows by observing that equation (**??**) is invariant to a scale transformation of the model.

Assume the unknown median of $(\varepsilon_t^*)^2$ is $d > 0$. Put $\varepsilon_t = \varepsilon_t^*/\sqrt{d}$. Then the median of ε_t^2 becomes one, and model (**??**) and equation (**??**) can be written as

$$Y_t = aY_{t-1} + \sqrt{\omega + bY_{t-1}^2}\varepsilon_t \tag{2.14}$$

and

$$E\{|a + \sqrt{b}\epsilon_t|^\alpha\} = 1, \tag{2.15}$$

where $a = a^*, \omega = d\omega^*$ and $b = db^*$. Therefore, as before we first propose to estimate $\theta = (\omega, a, b)^T$ by the following weighted least absolute deviations estimate

$$\hat{\theta} = (\hat{\omega}, \hat{a}, \hat{b})^T = \arg\min_{\theta} \sum_{t=1}^{n} \frac{1}{1+Y_{t-1}^2} |(Y_t - aY_{t-1})^2 - (\omega + bY_{t-1}^2)|. \tag{2.16}$$

Put $\hat{\varepsilon}_t = (Y_t - \hat{a}Y_{t-1})/\sqrt{\hat{\omega} + \hat{b}Y_{t-1}^2}$. Then, α can be estimated by solving the following equation:

$$\frac{1}{n} \sum_{t=1}^{n} |\hat{a} + \sqrt{\hat{b}}\hat{\varepsilon}_t|^{\alpha} = 1. \tag{2.17}$$

Denote this estimator by $\hat{\alpha}$, and let α_0 denote the true value of α. Put $\Delta = (1+Y_1^2)(\omega_0 + b_0 Y_1^2)$, $S = 1 + Y_1^2$,

$$\Gamma_1 = \begin{pmatrix} \mathrm{E}\frac{a_0^2 Y_1^4}{\Delta} + \mathrm{E}\frac{Y_1^2}{S} & \mathrm{E}\frac{a_0 Y_1^2}{\Delta} & -\mathrm{E}\frac{a_0 Y_1^4}{\Delta} \\ \mathrm{E}\frac{a_0 Y_1^2}{\Delta} & \mathrm{E}\frac{1}{\Delta} & -\mathrm{E}\frac{Y_1^2}{\Delta} \\ -\mathrm{E}\frac{a_0 Y_1^4}{\Delta} & -\mathrm{E}\frac{Y_1^2}{\Delta} & \mathrm{E}\frac{Y_1^4}{\Delta} \end{pmatrix}, \ \Gamma_2 = \begin{pmatrix} \frac{1}{2} & 0 & 0 \\ 0 & 1 & 0 \\ \alpha_0 & 0 & 1 \end{pmatrix}.$$

Let $\bar{A}(t) = (Y_t Y_{t-1}, 1, -Y_{t-1}^2)^T \mathrm{sgn}(\varepsilon_t^2 - 1)/(1 + Y_{t-1}^2)$, $f(x)$ denote the density of ε_1,

$$\begin{aligned} \bar{\gamma}_{\alpha_0}^2 &= \{f(1)\}^{-2}(c_1, c_2, c_3)\Gamma_2\Gamma_1^{-1}\mathrm{Cov}\{\bar{A}(1)\}\Gamma_1^{-1}\Gamma_2(c_1, c_2, c_3)^T + \kappa_0^{-2}\mathrm{Var}(|a_0 + \sqrt{b_0}\varepsilon_1|^{\alpha_0}) \\ &\quad -2\{f(1)\}^{-1}\kappa_0^{-1}(c_1, c_2, c_3)\Gamma_2\Gamma_1^{-1}\mathrm{E}\{\bar{A}(1)(|a_0 + \sqrt{b_0}\varepsilon_1|^{\alpha_0} - \mathrm{E}|a_0 + \sqrt{b_0}\varepsilon_1|^{\alpha_0})\}, \end{aligned}$$

where

$$\begin{cases} \kappa_0 = \mathrm{E}\{|a_0 + \sqrt{b_0}\varepsilon_1|^{\alpha_0} \log |a_0 + \sqrt{b_0}\varepsilon_1|\}, \\ c_1 = \kappa_0^{-1}\mathrm{E}\frac{\sqrt{b_0}(\alpha_0|a_0+\sqrt{b_0}\varepsilon_2|^{\alpha_0-1}\mathrm{sgn}(a_0+\sqrt{b_0}\varepsilon_2))\epsilon_2}{2(w_0+b_0Y_1^2)}, \\ c_2 = \kappa_0^{-1}\mathrm{E}\{(\alpha_0|a_0 + \sqrt{b_0}\varepsilon_2|^{\alpha_0-1}\mathrm{sgn}(a_0 + \sqrt{b_0}\varepsilon_2))(\frac{\sqrt{b_0}Y_1}{\sqrt{w_0+b_0Y_1^2}} - 1)\}, \\ c_3 = \kappa_0^{-1}\mathrm{E}\{(\alpha_0|a_0 + \sqrt{b_0}\varepsilon_2|^{\alpha_0-1}\mathrm{sgn}(a_0 + \sqrt{b_0}\varepsilon_2))(\frac{\sqrt{b_0}\varepsilon_2 Y_1^2}{2(w_0+b_0Y_1^2)} - \frac{\varepsilon_2}{2\sqrt{b_0}})\}. \end{cases}$$

Then the following theorem holds.

Theorem 2.3. *In addition to Conditions A and B for model (**??**), we further assume that $\alpha_0 > 1$ and $\mathrm{E}|\varepsilon_t|^{\delta_0} < \infty$ for some $\delta_0 > 2\alpha_0$. Then as $n \to \infty$*

$$\sqrt{n}(\hat{\alpha} - \alpha_0) \xrightarrow{d} \mathrm{N}(0, \bar{\gamma}^2_{\alpha_0}).$$

Again, to avoid estimating $\bar{\gamma}^2_{\gamma_0}$, we develop a profile empirical likelihood method for constructing a confidence interval for α_0. Put $\varepsilon_t^2(\omega, a, b) = [(Y_t - aY_{t-1})^2 - (\omega + bY_{t-1}^2)]/(1 + Y_{t-1}^2)$, define

$$X_{t,1}(\theta, \alpha) = \left| a + \sqrt{b}(Y_t - aY_{t-1})/\sqrt{\omega + bY_{t-1}^2} \right|^{\alpha} - 1,$$

$$X_{t,2}(\theta) = (\partial(\varepsilon_t^2(\omega, a, b))/\partial\omega)\mathrm{sgn}\{\varepsilon_t^2(\omega, a, b)\},$$

$$X_{t,3}(\theta) = (\partial(\varepsilon_t^2(\omega, a, b))/\partial a)\mathrm{sgn}\{\varepsilon_t^2(\omega, a, b)\},$$

$$X_{t,4}(\theta) = (\partial(\varepsilon_t^2(\omega, a, b))/\partial b)\mathrm{sgn}\{\varepsilon_t^2(\omega, a, b)\},$$

and write $X_t(\theta, \alpha) = (X_{t,1}(\theta, \alpha), X_{t,2}(\theta), X_{t,3}(\theta), X_{t,4}(\theta))^T$. Based on the estimating equations $\sum_{t=1}^n X_t(\theta, \alpha) = 0$, we define the empirical likelihood function of (θ, α) as

$$L(\theta, \alpha) = \sup\{\prod_{t=1}^n (np_t) : p_1 \geq 0, \ldots, p_n \geq 0, \sum_{t=1}^n p_t = 1, \sum_{t=1}^n p_t X_t(\theta, \alpha) = 0\}.$$

Put $l(\theta, \alpha) = -2 \log L(\theta, \alpha)$. Since we are interested in α, we consider the profile empirical likelihood ratio $l_p(\alpha) = l(\tilde{\theta}(\alpha), \alpha)$, where $\tilde{\theta} = \tilde{\theta}(\alpha) := \arg\min_\theta l(\theta, \alpha)$. Next theorem shows that Wilks theorem holds for the proposed profile empirical likelihood method.

Theorem 2.4. *Under conditions of Theorem **??**, $l_p(\alpha_0)$ converges in distribution to $\chi^2(1)$ as $n \longrightarrow \infty$.*

Remark 2.5. *Based on Theorem (**??**), one can construct a confidence interval for the tail index α_0 under model (**??**) as in Corollary **??**.*

3 Data Analysis and Simulation Study

3.1 Data Analysis

We revisit the analysis of the daily HKD/USD exchange rate from January 21, 1998 to June 6, 2000 in Zhu and Ling (2015), where LADE-based inference is proposed to replace QMLE due to the lack of moments. Therefore it is useful to accurately estimate the tail index of this data set.

As in Zhu and Ling (2015), we consider the log-returns ($\times 100$) of this data sample denoted by $\{X_t\}_{t=1}^{600}$. First we fit an ARMA(10,10) model to the data and use the function 'auto.arima' in the R package 'forecast' with AIC to obtain the following best model:

$$
\begin{array}{rclllll}
X_t & = & 0.0012+ & 0.2374X_{t-1}+ & 0.0127X_{t-2}- & 0.1536X_{t-3}- & 0.1516X_{t-4} \\
 & & (0.0004) & (0.4867) & (0.7874) & (0.2930) & (0.0689) \\
 & + & 0.0283X_{t-5}- & 1.5400e_{t-1}+ & 0.2160e_{t-2}+ & 0.3375e_{t-3}+ & e_t. \\
 & & (0.0620) & (0.4868) & (0.9383) & (0.4560) &
\end{array}
\tag{3.1}
$$

Denote the resulted residuals by $Y_t's$. Next we use the function 'garchFit' in the R package 'fGarch' to fit a GARCH(1,1) to $Y_t's$ and obtain

$$
\begin{array}{llll}
Y_t = \sigma_t \varepsilon_t, \quad \sigma_t^2 = & 1.098 \times 10^{-5}+ & 0.7220\sigma_{t-1}^2+ & 0.3773Y_{t-1}^2. \\
 & (9.794 \times 10^{-6}) & (0.04657) & (0.14012)
\end{array}
\tag{3.2}
$$

Numbers in brackets mean standard deviations. After this fitting, we plot $\{X_t\}$, the autocorrelation functions of $\{Y_t\}$, $\{Y_t^2\}$ and estimated $\{\varepsilon_t\}$ in Figure 1, which indicate the fitting is good. However, as showed in Zhang and Ling (2015), the estimators in (**??**) would be inconsistent theoretically if $EY_t^2 = \infty$, and the standard deviations in (**??**) may be theoretically incorrect when $EY_t^4 = \infty$ since this case implies that the joint asymptotic limit of estimators in (**??**) is nonnormal.

Here we study the tail index of Y_t by applying the profile empirical likelihood methods based on both the LADE in this paper and the QMLE in Chan, Peng and Zhang (2012) to $\{Y_t\}$ without taking into account of the randomness in obtaining $\{Y_t\}$. Since the parameters in (**??**) satisfy $w > 0, a \in (0,1), b > 0$, we rewrite $w = \exp(\tilde{w})$, $a = \exp(\tilde{a})/\{1 + \exp(\tilde{a})\}$, $b = \exp(\tilde{b})$ in computing the profile empirical likelihood ratio based on QMLE. Similar transformation for θ^* in (**??**) is applied to computing the profile empirical likelihood ratio based on LADE.

First we use the R function 'garchFit' to obtain the QMLE for $\tilde{\theta}^*$, and then get an estimator for $(\epsilon_t^*)^2$, which results in an estimator for d. Hence we have initial values for both $\tilde{\theta}$ and $\tilde{\theta}^*$, which are the transformed θ and θ^*. Denote them by $\tilde{\theta}_{ini}$ and $\tilde{\theta}^*_{ini}$. Using the obtained initial value $\tilde{\theta}_{ini}$, we minimize $\Delta(\theta) = \sum_{j=1}^4 (\frac{1}{n}\sum_{t=1}^n \bar{Z}_{t,j}(\theta))^2$ to obtain $\bar{\theta}_{ini}$. Next we employ the R package 'emplik' and the R function 'optim' to compute the profile empirical likelihood ratio based on LADE for $\alpha = 0.7, 0.75, 0.8, ..., 4$ by using either $\tilde{\theta}_{ini}$ or $\bar{\theta}_{ini}$, depending on which one gives a smaller value of $\Delta(\theta)$, as an initial value. The same approach is applied to calculating the profile empirical likelihood ratio based on QMLE by using $\tilde{\theta}^*_{ini}$ instead of $\tilde{\theta}_{ini}$. We also compute the profile empirical likelihood ratios by restricting $|\tilde{\theta} - \tilde{\theta}_{ini}| \le \delta$ and $|\tilde{\theta}^* - \tilde{\theta}^*_{ini}| \le \delta$ with $\delta = 0.5$. Hence we use $\delta = \infty$ to mean no such a restriction in our calculation. The profile empirical likelihood ratio based on LADE in Figure 2 has its minimum around $\alpha = 1.8$ and indicates $\alpha_0 \in (1,3)$ at both level 90% and level 95%. The profile empirical likelihood ratio based on QMLE in Figure 2 gives a very large value when $\alpha = 2$, which rejects $H_0 : \alpha_0 = 2$. This is in line with the fact that the estimated value of $a + b$ in (**??**) is larger than one, i.e., $EY_t^2 = \infty$. However, the empirical likelihood ratio based on QMLE fails to reject other considered α's in $(0.7, 4)$ at levels 90% and 95%, which may indicate the method is not applicable to this data set. After plotting the Hill's estimator in (**??**) for both $\{Y_t\}$ and estimated $\{\varepsilon_t\}$ in Figure 3, we conclude that the method in Chan, Peng and Zhang (2012) is problematic since $E\varepsilon_t^4$ seems infinite, and the standard deviations in (**??**) are inaccurate since $EY_t^4 = \infty$. Note that the 95% confidence intervals in Figure 3 are based on $\sqrt{k}(\tilde{\alpha}(k)/\alpha_0 - 1) \xrightarrow{d} N(0,1)$ for independent data.

3.2 Simulation Study

In this section we examine the finite sample behavior of the proposed profile empirical likelihood for a GARCH(1,1) sequence and compare it with the method in Chan, Peng and Zhang (2012), where the errors are required to have a finite fourth moment.

Consider model (**??**) with $\varepsilon_t^* \sim t(\nu)/\sqrt{\nu/(\nu-2)}$ with $\nu = 3.2$, or 4, or 8, or 12, and $\theta_0^* = (1, 0.72, 0.38)^T$, or $(1, 0.65, 0.38)^T$, or $(1, 0.65, 0.25)^T$, or $(1, 0.6, 0.25)^T$. By drawing $5,000$ random samples with sample size $n = 500$, $n = 1,000$ and $n = 2,000$, we follow the procedure in the data analysis to compute the profile empirical likelihood ratios and calculate the coverage probabilities for the proposed profile empirical likelihood confidence interval in this paper and that in Chan, Peng and Zhang (2012) with levels $\xi = 0.9$ and $\xi = 0.95$, which are denoted by I_ξ^{LADE} and I_ξ^{QMLE}.

Coverage probabilities for these two methods are reported in Table 1, which shows that the proposed profile empirical likelihood method works well and even performs better than the method based on the QMLE in Chan, Peng and Zhang (2012). Results for $\nu = 3$ and 4 well indicate that the method in Chan, Peng and Zhang (2012) does not work since the errors have an infinite fourth moment.

4 Conclusions

It is known that the tail index of a GARCH(1,1) sequence or an AR(1) model with ARCH(1) errors is determined by an estimating equation, which can be employed to estimate the tail index at the rate of $\sqrt{n}$, where n is the sample size. That is, the resulted tail index estimator has a faster rate of convergence than an estimator based on extreme value theory. However, this estimation procedure requires that the plug-in estimators for the unknown parameters in the model should have a joint normal limit, which generally needs a finite fourth moment for the errors. By noting that the estimating equation for determining the tail index is invariant to a scale transformation of the underlying model, we propose to estimate the tail index by employing some least absolute deviations estimate so as to relax the moment condition on errors. Although the resulted tail index estimator has a $\sqrt{n}$ rate of convergence and a normal limit, the asymptotic variance is quite complicated. To effectively construct a confidence interval for the tail index, we further propose a profile empirical likelihood method, which does not need to estimate any additional quantities such as asymptotic variance. A simulation study confirms that the proposed new methods have good finite sample behavior.

5 Proofs

5.1 Proofs for GARCH(1, 1) Case

In this subsection, we define $\Theta = \{\theta : ||\theta - \theta_0|| \le n^{-\frac{1}{2\gamma}}\}$ for some $1 < \gamma < \min\{\delta_0/2, 2\}$, where $||\cdot||$ denotes the L_2 norm.

Lemma 5.1. *Under conditions of Theorem* **??**,

$$\sup_{1\le t\le n}\sup_{\theta\in\Theta}||\bar{Z}_t(\theta,\alpha_0)|| = o_p(n^{\frac{1}{2\gamma}}). \tag{5.1}$$

Proof. Put $g(t,\theta) = (\bar{Z}_{t,2}(\theta), \bar{Z}_{t,3}(\theta), \bar{Z}_{t,4}(\theta))^T =: (g_1(t,\theta), g_2(t,\theta), g_3(t,\theta))^T$ and

$$\begin{aligned} h(t,\theta) &= \Big(\frac{1-a^t}{1-a}, \frac{\omega\{(1-a^t) - ta^{t-1}(1-a)\}}{(1-a)^2} + b\sum_{k=0}^{t-1} ka^{k-1}Y_{t-k-1}^2, \sum_{k=0}^{t-1} a^k Y_{t-k-1}^2\Big)^T \\ &=: (h_1(t,\theta), h_2(t,\theta), h_3(t,\theta))^T. \end{aligned}$$

Then we have

$$g(t,\theta) = \mathrm{sgn}\{\log(Y_t^2/\bar{\sigma}_t^2(\theta))\}h(t,\theta)\bar{\sigma}_t^{-2}(\theta). \tag{5.2}$$

Write $\sigma_0 = \sigma_0(\theta_0)$. It follows from (**??**) that

$$\begin{aligned} \bar{\sigma}_t^2(\theta) - \sigma_t^2(\theta_0) &= (\omega - \omega_0)\frac{1-a_0^t}{1-a_0} + \omega\left\{\frac{a_0^t - a^t}{1-a} + \frac{(1-a_0^t)(a-a_0)}{(1-a)(1-a_0)}\right\} \\ &\quad + b\sum_{i=0}^{t-1}(a^i - a_0^i)Y_{t-1-i}^2 + (b-b_0)\sum_{i=0}^{t-1}a_0^iY_{t-1-i}^2 - a_0^t\sigma_0^2. \end{aligned} \tag{5.3}$$

Thus, there exists $C_1 > 0$ such that

$$\left|\bar{\sigma}_t^2(\theta) - \sigma_t^2(\theta_0)\right| / \sigma_t^2(\theta_0) \leq [C_1 n^{-\frac{1}{2\gamma}}\{1 + \sum_{i=0}^{t-1} i(\max(a, a_0))^{i-1} Y_{t-1-i}^2\} + a_0^t \sigma_0^2] / \sigma_t^2(\theta_0) \quad (5.4)$$

uniformly for $t \geq 1$ and $\theta \in \Theta$. By the inequality

$$x/(1+x) \leq x^\tau \quad \text{for} \quad x > 0 \quad \text{and} \quad 0 < \tau < 1,$$

it can be shown that there exist constants C_2 and $\rho \in (0,1)$ such that for any $\tau \in (0,1)$

$$\begin{aligned}
\sup_{\theta \in \Theta} \{1 + \sum_{i=0}^{t-1} i(\max(a, a_0))^{i-1} Y_{t-1-i}^2\} / \sigma_t^2(\theta_0) &\leq C_2 \Big(1 + \sum_{i=0}^{\infty} \rho^i |Y_{t-1-i}|\Big)^\tau, & (5.5) \\
a_0^t \sigma_0^2 / \sigma_t^2(\theta_0) &\leq C_2 \sum_{i=t}^{\infty} a_0^{i\tau} |Y_{t-1-i}|^\tau \text{ and} & (5.6) \\
\sup_{\theta \in \Theta} ||h(t,\theta)|| / \bar{\sigma}_t^2(\theta) &\leq C_2 \Big(1 + \sum_{i=0}^{\infty} \rho^i |Y_{t-1-i}|\Big)^\tau & (5.7)
\end{aligned}$$

uniformly for $t \geq 1$. By (**??**), (**??**) and $a_0 < 1$ (see Remark **??**), for any $0 < \delta < 1/2$ there exists a t_δ such that

$$\sup_{t \geq t_\delta} \sup_{\theta \in \Theta} \left|\bar{\sigma}_t^2(\theta) - \sigma_t^2(\theta_0)\right| / \sigma_t^2(\theta_0) < \delta \ \text{ in probability .} \quad (5.8)$$

Thus, by inequality $|\log(1+x)| \leq 2|x|$ for all $x > -1/2$, we have for all $t \geq t_\delta$,

$$\begin{aligned}
\left|\log(\bar{\sigma}_t^2(\theta)/\sigma_t^2(\theta_0))\right| &= \left|\log\{1 + (\bar{\sigma}_t^2(\theta) - \sigma_t^2(\theta_0))/\sigma_t^2(\theta_0)\}\right| \\
&\leq 2\left|\bar{\sigma}_t^2(\theta) - \sigma_t^2(\theta_0)\right| / \sigma_t^2(\theta_0) \\
&\leq 2C_2 \left[C_1 n^{-\frac{1}{2\gamma}} \Big(1 + \sum_{i=0}^{\infty} \rho^i |Y_{t-1-i}|\Big)^\tau + \Big(\sum_{i=t}^{\infty} a_0^i |Y_{t-1-i}|\Big)^\tau\right] \\
&=: d(n,t). \qquad (5.9)
\end{aligned}$$

It follows that

$$\begin{aligned}
& \left|\text{sgn}\{\log(Y_t^2/\bar{\sigma}_t^2(\theta))\} - \text{sgn}\{\log \varepsilon_t^2\}\right| \\
= & \left|\text{sgn}\{\log(Y_t^2/\sigma_t^2(\theta_0)) - \log(\bar{\sigma}_t^2(\theta)/\sigma_t^2(\theta_0))\} - \text{sgn}\{\log(Y_t^2/\sigma_t^2(\theta_0))\}\right| \\
\leq & 2|I\{\log(\bar{\sigma}_t^2(\theta)/\sigma_t^2(\theta_0)) < \log \varepsilon_t^2 \leq 0\} - I\{0 \leq \log \varepsilon_t^2 < \log(\bar{\sigma}_t^2(\theta)/\sigma_t^2(\theta_0))\}| \\
\leq & 2I\{|\log \varepsilon_t^2| \leq d(n,t)\}. \qquad (5.10)
\end{aligned}$$

This, combining with (??) and (??), yields that for any $0 < \tau < 1$,

$$\begin{aligned} ||g(t,\theta)|| &\leq C_2\Big(1+\sum_{i=0}^{\infty}\rho^i|Y_{t-1-i}|\Big)^{\tau} I\{|\log\varepsilon_t^2| \leq d(n,t)\} \\ &\quad +C_2\Big(1+\sum_{i=0}^{\infty}\rho^i|Y_{t-1-i}|\Big)^{\tau} \\ &=: I_1(t)+I_2(t) \end{aligned} \tag{5.11}$$

uniformly for $t \geq t_\delta$. By the corollary on p.322 of Nelson (1990), we have $\mathrm{E}\sigma_0^p < \infty$ for any $0 < p < \alpha_0$. Since $\log\varepsilon_t^2$ is independent of $\mathcal{F}_{t-1} = \sigma(Y_{t-1},\ldots,Y_{-\infty})$ and its density is continuous at zero, by taking τ small enough such that $\mathrm{E}|Y_t|^{8\gamma\tau} < \infty$ and δ small enough such that $\sup_{|x|\leq\delta} f(x) \leq 2f(0)$, there exists $C_3 > 0$ such that for all $t > t_\delta$,

$$\begin{aligned} &\mathrm{E}\left|\Big(1+\sum_{i=0}^{\infty}\rho^i|Y_{t-1-i}|\Big)^{\tau} I\{|\log\varepsilon_t^2| \leq d(n,t)\}\right|^{4\gamma} \\ = \ &\mathrm{E}\left[\Big(1+\sum_{i=0}^{\infty}\rho^i|Y_{t-1-i}|\Big)^{\tau} \mathrm{E}\left(I\{|\log\varepsilon_t^2| \leq d(n,t)\}|\mathcal{F}_{t-1}\right)\right]^{4\gamma} \\ \leq \ &C_3 f^{4\gamma}(0)\mathrm{E}\Big(1+\sum_{i=0}^{\infty}\tilde{\rho}^i|Y_{t-1-i}|\Big)^{8\gamma\tau}\left[n^{-2}+a_0^{4\gamma t\tau}\right], \end{aligned} \tag{5.12}$$

where $\tilde{\rho} = \max\{\rho, a_0\}$. Therefore, for any $\zeta > 0$,

$$\begin{aligned} &P\{\sup_{1\leq t\leq n} I_1(t) > \zeta n^{\frac{1}{2\gamma}}\} \\ \leq \ &\sum_{t=1}^{t_\delta} P\left\{C_2\Big(1+\sum_{i=0}^{\infty}\rho^i|Y_{t-1-i}|\Big)^{\tau} > \zeta n^{\frac{1}{2\gamma}}\right\} + \sum_{t=t_\delta+1}^{n}\zeta^{-4\gamma}n^{-2}\mathrm{E}|I_1(t)|^{4\gamma} \\ \to \ &0. \end{aligned} \tag{5.13}$$

Similarly, for any $\zeta > 0$,

$$P\{\sup_{1\leq t\leq n} I_2(t) > \zeta n^{\frac{1}{2\gamma}}\} \leq \sum_{t=1}^{n}\zeta^{-4\gamma}n^{-2}\mathrm{E}|I_2(t)|^{4\gamma} \leq C_3\zeta^{-4\gamma}n^{-1} \to 0. \tag{5.14}$$

So it follows from (??) and (??) that

$$\sup_{1\leq t\leq n}\sup_{\theta\in\Theta}||g(t,\theta)|| = o_p(n^{\frac{1}{2\gamma}}). \tag{5.15}$$

On the other hand, similar to (**??**), we have

$$
\begin{aligned}
& (a + bY_t^2/\bar{\sigma}_t^2(\theta))^{\alpha_0/2} \\
= \;& \{a_0 + b_0 Y_t^2/\bar{\sigma}_t^2(\theta) + (a - a_0) + (b - b_0) Y_t^2/\bar{\sigma}_t^2(\theta)\}^{\alpha_0/2} \qquad (5.16) \\
\leq \;& \{a_0 + b_0\varepsilon_t^2 + Cn^{-\frac{1}{2\gamma}}(1 + \varepsilon_t^2(1 + \textstyle\sum_{i=0}^\infty \rho^i |Y_{t-1-i}|)^\tau + \omega^{-1} a_0^t \varepsilon_t^2 \sigma_0^2)\}^{\alpha_0/2}
\end{aligned}
$$

uniformly for $t \geq 1$. Thus, by the inequality: $(1+x)^p \leq 1 + 2px$ for $p > 0$ and small $x > 0$, we have

$$\sup_{1\leq t\leq n} \sup_{\theta\in\Theta} |\{a + bY_t^2/\bar{\sigma}_t^2(\theta)\}^{\alpha_0/2} - (a_0 + b_0\varepsilon_t^2)^{\alpha_0/2}| = o_p(n^{\frac{1}{2\gamma}}). \qquad (5.17)$$

Since $\{\varepsilon_t^2\}$ is a sequence of independent and identically distributed random variables with $\mathrm{E}|\varepsilon_t|^{\delta_0} < \infty$ and $\mathrm{E}(a_0 + b_0\varepsilon_t^2)^{\alpha_0/2} = 1$, we have

$$\sup_{1\leq t\leq n} |(a_0 + b_0\varepsilon_t^2)^{\alpha_0/2} - 1| = o_p(n^{\frac{1}{2\gamma}}). \qquad (5.18)$$

Thus, the lemma follows from (**??**), (**??**) and (**??**). □

Lemma 5.2. *Under Conditions of Theorem* **??**,

$$(i) \qquad \sup_{\theta\in\Theta} ||\frac{1}{n}\sum_{t=1}^n \bar{Z}_t(\theta,\alpha_0)\bar{Z}_t^T(\theta,\alpha_0) - \mathrm{E}\{Z_1(\theta_0,\alpha_0)Z_1^T(\theta_0,\alpha_0)\}|| = o_p(1), \textit{ and}$$

$$(ii) \qquad \frac{1}{\sqrt{n}}\sum_{t=1}^n \bar{Z}_t(\theta_0,\alpha_0) \xrightarrow{d} \mathrm{N}\Big(0, \mathrm{E}\{Z_1(\theta_0,\alpha_0)Z_1^T(\theta_0,\alpha_0)\}\Big),$$

where $\bar{Z}_{t,i}(\theta,\alpha_0) = \bar{Z}_{t,i}(\theta)$ *when* $i = 2,3,4$, *and* $Z_t(\theta,\alpha_0)$ *is defined as* $\bar{Z}_t(\theta,\alpha_0)$ *with* $\sigma_t^2(\theta)$ *replaced by* $\bar{\sigma}_t^2(\theta)$.

Proof. For the proof of (i), it is sufficient to show that

$$\sup_{\theta\in\Theta}\Big|\Big|\frac{1}{n}\sum_{t=1}^n \{\bar{Z}_{t,i}(\theta,\alpha_0)\bar{Z}_{t,j}(\theta,\alpha_0)\} - \mathrm{E}\{Z_{t,i}(\theta_0,\alpha_0)Z_{t,j}(\theta_0,\alpha_0)\}\Big|\Big| = o_p(1) \qquad (5.19)$$

for $i,j = 1,2,3,4$. Here we only show the case of $i = 3$ and $j = 4$, since the other cases can be proved similarly. Define $\widetilde{h}_2(t,\theta) = \omega/(1-a)^2 + b\sum_{k=0}^{\infty} ka^{k-1}Y_{t-k-1}^2$. By (**??**),

$$\begin{aligned}
\bar{Z}_{t,3}(\theta,\alpha_0) &= \operatorname{sgn}\{\log(Y_t^2/\bar{\sigma}_t^2(\theta))\}\Big(\frac{\omega}{(1-a)^2} + b\sum_{k=0}^{\infty} ka^{k-1}Y_{t-k-1}^2\Big)\bar{\sigma}_t^{-2}(\theta) \\
&\quad -\operatorname{sgn}\{\log(Y_t^2/\bar{\sigma}_t^2(\theta))\}\Big(\frac{\omega a^t + ta^{t-1}(1-a)}{(1-a)^2} + b\sum_{k=t}^{\infty} ka^{k-1}Y_{t-k-1}^2\Big)\bar{\sigma}_t^{-2}(\theta) \\
&= \operatorname{sgn}(\log\varepsilon_t^2)\left[\widetilde{h}_2(t,\theta_0)/\sigma_t^2(\theta_0)\right] \\
&\quad +\operatorname{sgn}(\log\varepsilon_t^2)\left[(\widetilde{h}_2(t,\theta) - \widetilde{h}_2(t,\theta_0))/\sigma_t^2(\theta_0)\right] \\
&\quad +\operatorname{sgn}(\log\varepsilon_t^2)\left[(\sigma_t^2(\theta_0) - \bar{\sigma}_t^2(\theta))/\sigma_t^2(\theta_0)\right]\left[\widetilde{h}_2(t,\theta)/\bar{\sigma}_t^2(\theta)\right] \\
&\quad -\operatorname{sgn}\{\log(Y_t^2/\bar{\sigma}_t^2(\theta))\}a^{t-1}\left[\Big(\frac{\omega a + t(1-a)}{(1-a)^2} + b\sum_{k=0}^{\infty}(t+k)a^kY_{-k-1}^2\Big)/\bar{\sigma}_t^2(\theta)\right] \\
&\quad +\left[\operatorname{sgn}\{\log(Y_t^2/\bar{\sigma}_t^2(\theta))\} - \operatorname{sgn}(\log\varepsilon_t^2)\right]\left[\widetilde{h}_2(t,\theta)/\bar{\sigma}_t^2(\theta)\right] \\
&=: L_1(t) + L_2(t) + L_3(t) + L_4(t) + L_5(t)
\end{aligned}$$

and

$$\begin{aligned}
\bar{Z}_{t,4}(\theta,\alpha_0) &= \operatorname{sgn}\{\log(Y_t^2/\bar{\sigma}_t^2(\theta))\}\sum_{k=0}^{t-1} a^kY_{t-k-1}^2/\bar{\sigma}_t^2(\theta) \\
&= \operatorname{sgn}(\log\varepsilon_t^2)\sum_{k=0}^{\infty} a_0^kY_{t-k-1}^2/\sigma_t^2(\theta_0) \\
&\quad +\operatorname{sgn}\{\log(Y_t^2/\bar{\sigma}_t^2(\theta))\}\sum_{k=0}^{\infty}[a^kY_{t-k-1}^2/\sigma_t^2(\theta_0)][(\sigma_t^2(\theta_0) - \bar{\sigma}_t^2(\theta))/\bar{\sigma}_t^2(\theta)] \\
&\quad +\operatorname{sgn}(\log\varepsilon_t^2)\sum_{k=0}^{\infty}(a^k - a_0^k)Y_{t-k-1}^2/\sigma_t^2(\theta_0) \\
&\quad -\operatorname{sgn}\{\log(Y_t^2/\bar{\sigma}_t^2(\theta))\}\sum_{k=t}^{\infty} a^kY_{t-k-1}^2/\bar{\sigma}_t^2(\theta) \\
&\quad +\left[\operatorname{sgn}\{\log(Y_t^2/\bar{\sigma}_t^2(\theta))\} - \operatorname{sgn}(\log\varepsilon_t^2)\right]\sum_{k=0}^{\infty} a^kY_{t-k-1}^2/\sigma_t^2(\theta_0) \\
&=: M_1(t) + M_2(t) + M_3(t) + M_4(t) + M_5(t).
\end{aligned}$$

Similar to (??) and (??), there exist $C_4 > 0$ and $\rho \in (0,1)$ such that for any $0 < \tau < 1$,

$$
\begin{aligned}
& \sup_{\theta\in\Theta} |L_2(t) + L_3(t) + L_4(t)| \\
\leq\ & \sup_{\theta\in\Theta} C_4 \Big\{ \Big[n^{-\frac{1}{2\gamma}} \Big(1 + \sum_{i=0}^{\infty} \rho^i |Y_{t-i-1}|\Big)^{\tau} + a_0^t \sigma_0^2 \Big] \Big(1 + \sum_{i=0}^{\infty} \rho^i |Y_{t-i-1}|\Big)^{\tau} \Big\} \\
& + \omega^{-1} a^{t-1} \left[\Big(\frac{\omega a + t(1-a)}{(1-a)^2} + b \sum_{k=0}^{\infty} (t+k) a^k Y_{-k-1}^2 \Big) \right] \\
\leq\ & C_4 \Big[n^{-\frac{1}{2\gamma}} \Big(1 + \sum_{i=0}^{\infty} \rho^i |Y_{t-i-1}|\Big)^{\tau} + a_0^t \sigma_0^2 \Big] \Big(1 + \sum_{i=0}^{\infty} \rho^i |Y_{t-i-i}|\Big)^{\tau} \\
& + 2\omega_0^{-1} \left(\frac{a_0+1}{2} \right)^{t-1} \left\{ \frac{4\omega_0(1+a_0) + 6t(1-a_0)}{(1-a_0)^2} + 2b_0 \sum_{k=0}^{\infty} (t+k) \left(\frac{a_0+1}{2} \right)^k Y_{-k-1}^2 \right\},
\end{aligned} \tag{5.20}
$$

where the inequalities follow by taking n sufficiently large such that $C_4 n^{-\frac{1}{2\gamma}} \leq \min\{(1-a_0)/2, \omega_0/2, b_0/2\}$, and

$$
\begin{aligned}
& \sup_{\theta\in\Theta} |M_2(t) + M_3(t) + M_4(t)| \\
\leq\ & C_4 \Big\{ n^{-\frac{1}{2\gamma}} \Big(1 + \sum_{i=0}^{\infty} \rho^i |Y_{t-i-1}|\Big)^{\tau} + a_0^t \sigma_0^2 \Big\} \Big(1 + \sum_{i=0}^{\infty} \rho^i |Y_{t-i-1}|\Big)^{\tau} \\
& + 2\omega_0^{-1} \left(\frac{a_0+1}{2\gamma} \right)^{t} \sum_{k=0}^{\infty} \left(\frac{a_0+1}{2} \right)^k Y_{-k-1}^2.
\end{aligned} \tag{5.21}
$$

Thus, by (??) and (??), we can show that

$$
\frac{1}{n} \sum_{t=1}^{n} \left\{ \sup_{\theta\in\Theta} [|M_1(t)|(|L_2(t) + L_3(t) + L_4(t)|)] \right\} \xrightarrow{p} 0, \tag{5.22}
$$

$$
\frac{1}{n} \sum_{t=1}^{n} \left\{ \sup_{\theta\in\Theta} [|L_1(t)|(|M_2(t) + M_3(t) + M_4(t)|)] \right\} \xrightarrow{p} 0, \text{ and} \tag{5.23}
$$

$$
\frac{1}{n} \sum_{t=1}^{n} \left\{ \sup_{\theta\in\Theta} [(|L_2(t) + L_3(t) + L_4(t)|)(|M_2(t) + M_3(t) + M_4(t)|)] \right\} \xrightarrow{p} 0. \tag{5.24}
$$

Further, using the same arguments as in proving Lemma 5.1, we have

$$
\frac{1}{n} \sum_{t=1}^{n} \left\{ \sup_{\theta\in\Theta} [|M_5(t)|(|L_1(t) + L_2(t) + L_3(t) + L_4(t) + L_5(t)|)] \right\} \xrightarrow{p} 0, \text{ and} \tag{5.25}
$$

$$
\frac{1}{n} \sum_{t=1}^{n} \left\{ \sup_{\theta\in\Theta} [|L_5(t)|(|M_1(t) + M_2(t) + M_3(t) + M_4(t)|)] \right\} \xrightarrow{p} 0. \tag{5.26}
$$

Therefore,

$$\sup_{\theta\in\Theta}\Big|\frac{1}{n}\sum_{t=1}^{n}\{\bar{Z}_{t,3}(\theta,\alpha_0)\bar{Z}_{t,4}(\theta,\alpha_0)\} - \mathrm{E}\{Z_{t,3}(\theta_0,\alpha_0)Z_{t,4}(\theta_0,\alpha_0)\}\Big|$$
$$= \sup_{\theta\in\Theta}\Big|\frac{1}{n}\sum_{t=1}^{n}\{L_1(t)M_1(t)\} - \mathrm{E}\{Z_{t,3}(\theta_0,\alpha_0)Z_{t,4}(\theta_0,\alpha_0)\}\Big| + o_p(1) = o_p(1), \quad (5.27)$$

i.e., (**??**) holds for the case of $i = 3$ and $j = 4$.

Next, we prove (ii). Note that

$$\begin{aligned} g(t,\theta_0) &= \mathrm{sgn}\{\log\varepsilon_t^2\}h(t,\theta_0)/\sigma_t^2(\theta_0) \\ &\quad + \left(\mathrm{sgn}\{\log(Y_t^2/\bar{\sigma}_t^2(\theta_0))\} - \mathrm{sgn}\{\log\varepsilon_t^2\}\right) h(t,\theta_0)/\sigma_t^2(\theta_0) \\ &\quad + \mathrm{sgn}\{\log(Y_t^2/\bar{\sigma}_t^2(\theta_0))\}h(t,\theta_0)[1/\bar{\sigma}_t^2(\theta_0) - 1/\sigma_t^2(\theta_0)] \\ &=: H_1(t) + H_2(t) + H_3(t). \end{aligned} \quad (5.28)$$

Using $\sigma_t^2(\theta_0) - \bar{\sigma}_t^2(\theta_0) = a_0^t\sigma_0^2$ and the same arguments as in deriving (**??**)–(**??**), we have

$$\mathrm{E}\{\frac{1}{\sqrt{n}}\sum_{t=1}^{n}(||H_2(t)|| + ||H_3(t)||)\}^p = O\left\{n^{-p/2}\sum_{t=1}^{n}\mathrm{E}||a_0^t\sigma_0^2 h(t,\theta_0)||^p\right\} = O(n^{-p/2}) \quad (5.29)$$

for any $0 < p < \min\{1, \alpha_0/8\}$, which implies that

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}g(t,\theta_0) - \frac{1}{\sqrt{n}}\sum_{t=1}^{n}H_1(t) \xrightarrow{p} 0. \quad (5.30)$$

Similarly, we can show that

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\{(a_0 + b_0Y_t^2/\bar{\sigma}_t^2(\theta_0))^{\alpha_0/2} - 1\} - \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\{(a_0 + b_0\varepsilon_t^2)^{\alpha_0/2} - 1\} \xrightarrow{p} 0. \quad (5.31)$$

Note that $H_1(t) = (Z_{t,2}(\theta_0), Z_{t,3}(\theta_0), Z_{t,4}(\theta_0))^T$. By (**??**) and (**??**), we have

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\bar{Z}_t(\theta_0,\alpha_0) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}Z_t(\theta_0,\alpha_0) + o_p(1). \quad (5.32)$$

Since $Z_t(\theta_0,\alpha_0)$ is a martingale difference sequence, (ii) follows from (**??**) and the central limit theorem (CLT) for martingales (see Hall and Heyde (1980)). This completes the proof of Lemma 5.2. □

Lemma 5.3. *Under conditions of Theorem* **??**,

$$\sup_{\theta\in\Theta}\Big|\Big|\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\Big\{\Big[sgn\Big(\log\frac{Y_t^2}{\bar\sigma_t^2(\theta)}\Big)-sgn\Big(\log\frac{Y_t^2}{\bar\sigma_t^2(\theta_0)}\Big)\Big]\frac{h(t,\theta)}{\bar\sigma_t^2(\theta)}$$
$$-2f(0)\Big(\frac{\sigma_t^2(\theta_0)-\bar\sigma_t^2(\theta)}{\sigma_t^2(\theta_0)}\Big)\frac{h(t,\theta_0)}{\bar\sigma_t^2(\theta_0)}\Big\}\Big|\Big|=o_p(1),$$

where $h(t,\theta)$ *is defined in the proof of Lemma 5.1.*

Proof. By Taylor expansion, similar to Lemma 5.2, it can be shown that

$$\sup_{\theta\in\Theta}\Big|\Big|\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\Big\{\Big[\mathrm{sgn}\Big(\log\frac{Y_t^2}{\bar\sigma_t^2(\theta)}\Big)-\mathrm{sgn}\Big(\log\frac{Y_t^2}{\bar\sigma_t^2(\theta_0)}\Big)\Big]\Big(\frac{h(t,\theta)}{\bar\sigma_t^2(\theta)}-\frac{h(t,\theta_0)}{\bar\sigma_t^2(\theta_0)}\Big)\Big\}\Big|\Big|=o_p(1).$$

Further, similar to (**??**), we have

$$\sup_{\theta\in\Theta}\Big|\Big|\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left[\mathrm{sgn}\left(\log\frac{Y_t^2}{\bar\sigma_t^2(\theta_0)}\right)-\mathrm{sgn}(\log\varepsilon_t^2)\right]\frac{h(t,\theta_0)}{\bar\sigma_t^2(\theta_0)}\Big|\Big|=o_p(1).$$

Thus, for proving Lemma 5.3, it suffices to show that

$$\sup_{\theta\in\Theta}\Big|\Big|\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\Big[\mathrm{sgn}\Big(\log\frac{Y_t^2}{\bar\sigma_t^2(\theta)}\Big)-\mathrm{sgn}(\log\varepsilon_t^2)-2f(0)\Big(\frac{\sigma_t^2(\theta_0)-\bar\sigma_t^2(\theta)}{\sigma_t^2(\theta_0)}\Big)\Big]\frac{h(t,\theta_0)}{\bar\sigma_t^2(\theta_0)}\Big|\Big|=o_p(1).$$

Put $\xi_{1t}(\theta)=[2I(\log(\frac{\bar\sigma_t^2(\theta)}{\sigma_t^2(\theta_0)})<\log\varepsilon_t^2<0)+I(\log(\frac{\bar\sigma_t^2(\theta)}{\sigma_t^2(\theta_0)})<\log\varepsilon_t^2=0)]h(t,\theta_0)/\bar\sigma_t^2(\theta_0)$, $\xi_{2t}(\theta)=[2I(0<\log\varepsilon_t^2\le\log(\frac{\bar\sigma_t^2(\theta)}{\sigma_t^2(\theta_0)}))+I(0=\log\varepsilon_t^2\le\log(\frac{\bar\sigma_t^2(\theta)}{\sigma_t^2(\theta_0)}))]h(t,\theta_0)/\bar\sigma_t^2(\theta_0)$ and $\mathcal{F}_t=\sigma(\varepsilon_s,\,s\le t)$. Then $\Big[\mathrm{sgn}\Big(\log\frac{Y_t^2}{\bar\sigma_t^2(\theta)}\Big)-\mathrm{sgn}(\log\varepsilon_t^2)\Big]\frac{h^T(t,\theta_0)}{\bar\sigma_t^2(\theta_0)}=\xi_{1t}(\theta)-\xi_{2t}(\theta)$ and

$$\begin{aligned}\mathrm{E}[(\xi_{1t}(\theta)-\xi_{2t}(\theta))|\mathcal{F}_{t-1}] &= -2f(0)\log\left(\bar\sigma_t^2(\theta)/\sigma_t^2(\theta_0)\right)\left(h(t,\theta_0)/\bar\sigma_t^2(\theta_0)\right)(1+o_p(1))\\ &= 2f(0)\left(\frac{\sigma_t^2(\theta_0)-\bar\sigma_t^2(\theta)}{\sigma_t^2(\theta_0)}\right)\frac{h(t,\theta_0)}{\bar\sigma_t^2(\theta_0)}(1+o_p(1))\end{aligned}$$

holds uniformly in $\theta\in\Theta$. Hence, we only need to show that

$$\sup_{\theta\in\Theta}\Big\|\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\{\xi_{it}(\theta)-\mathrm{E}[\xi_{it}(\theta)|\mathcal{F}_{t-1}]\}\Big\|=o_p(1)\text{ for }i=1,2. \tag{5.33}$$

It follows from (**??**) and (**??**) that for $i=1$, 2

$$\frac{1}{n}\sum_{t=1}^{n}\mathrm{E}\|\xi_{it}(\theta)\|^2=o(1)\text{ and }\frac{1}{n}\sum_{t=1}^{n}\mathrm{E}\|\xi_{it}(\theta_1)-\xi_{it}(\theta_2)\|^2\le C||\theta_1-\theta_2||. \tag{5.34}$$

Note that for any given θ, $\{\xi_{it}(\theta)-\mathrm{E}[\xi_{it}(\theta)|\mathcal{F}_{t-1}]\}$ is a martingale difference sequence. By (**??**) and a chaining technique (see pp. 356-358 in Hansen (1996) or pp. 330–331 in Koul and Surgailis (2001)), (**??**) can be derived. □

Proof of Theorem 2.1. It follows from Theorem 1 of Peng and Yao (2003), Conditions 1 and 2 that

$$\sqrt{n}(\hat{\theta}-\theta_0)=\frac{\Omega^{-1}}{2f(0)\sqrt{n}}\sum_{t=1}^{n}A(t)+o_p(1)\xrightarrow{d}N(0,\Omega^{-1}/\{4f^2(0)\}), \tag{5.35}$$

where $A(t)$ is given in Theorem 2.1. Thus, by (**??**) with α instead of α_0 and some similar arguments as in proving (**??**), we have for any $0\le\alpha\le\delta_0$,

$$\frac{1}{n}\sum_{t=1}^{n}\{(\hat{a}+\hat{b}\bar{\varepsilon}_t^2(\hat{\theta}))^{\alpha/2}-(a_0+b_0\varepsilon_t^2)^{\alpha/2}\}\xrightarrow{p}0. \tag{5.36}$$

It follows from the weak law of large numbers that

$$\frac{1}{n}\sum_{t=1}^{n}\{(a_0+b_0\varepsilon_t^2)^{\alpha/2}-\mathrm{E}(a_0+b_0\varepsilon_t^2)^{\alpha/2}\}\xrightarrow{p}0.$$

Since the convergence of a monotone function to its limit is uniform over any closed interval, we have

$$\sup_{0\le\alpha\le\delta_0}\left|\frac{1}{n}\sum_{t=1}^{n}\{(\hat{a}+\hat{b}\bar{\varepsilon}_t^2(\hat{\theta}))^{\alpha/2}-\mathrm{E}(a_0+b_0\varepsilon_t^2)^{\alpha/2}\}\right|\xrightarrow{p}0. \tag{5.37}$$

Since α_0 is the unique positive solution to $\mathrm{E}(a_0+b_0\varepsilon_t^2)^{\alpha/2}=1$, we have $\hat{\alpha}\xrightarrow{p}\alpha_0$. Thus, by Taylor expansion, when $|\hat{\alpha}-\alpha_0|\le\nu$ with $\nu>0$ small enough,

$$\begin{aligned}
0 &= \frac{1}{n}\sum_{t=1}^{n}(\hat{a}+\hat{b}\bar{\varepsilon}_t^2(\hat{\theta}))^{\hat{\alpha}/2}-1\\
&= \frac{1}{n}\sum_{t=1}^{n}[(\hat{a}+\hat{b}\bar{\varepsilon}_t^2(\hat{\theta}))^{\hat{\alpha}/2}-(a_0+b_0\varepsilon_t^2)^{\alpha_0/2}]+\frac{1}{n}\sum_{t=1}^{n}[(a_0+b_0\varepsilon_t^2)^{\alpha_0/2}-1]\\
&= \frac{\hat{\alpha}}{2n}\sum_{t=1}^{n}[\tilde{a}_0+\tilde{b}_0\tilde{\varepsilon}_t^2]^{\hat{\alpha}/2-1}[(\hat{a}-a_0)+(\hat{b}\bar{\varepsilon}_t^2(\hat{\theta})-b_0\varepsilon_t^2)]\\
&\quad+\frac{1}{n}\sum_{t=1}^{n}[(a_0+b_0\varepsilon_t^2)^{\alpha_0/2}-1]+\frac{1}{2n}\sum_{t=1}^{n}(a_0+b_0\varepsilon_t^2)^{\alpha_0/2}\log(a_0+b_0\varepsilon_t^2)(\hat{\alpha}-\alpha_0)\\
&\quad+\frac{1}{4n}\sum_{t=1}^{n}(a_0+b_0\varepsilon_t^2)^{\tilde{\alpha}/2}\{\log(a_0+b_0\varepsilon_t^2)\}^2(\hat{\alpha}-\alpha_0)^2,
\end{aligned} \tag{5.38}$$

where $(\tilde a,\tilde b,\tilde\varepsilon_t^2,\tilde\alpha)$ lies between $(\hat a,\hat b,\bar\varepsilon_t^2(\hat\theta),\hat\alpha)$ and $(a_0,b_0,\varepsilon_t^2,\alpha_0)$. Thus, like the proof of Lemma 5.2, the right hand side of (**??**) is equal to

$$\begin{aligned}
&\Big\{\frac{\alpha_0}{2n}\sum_{t=1}^n[a_0+b_0\varepsilon_t^2]^{\alpha_0/2-1}[(\hat a-a_0)+(\hat b-b_0)\varepsilon_t^2-(\hat\theta-\theta_0)^T(b_0\varepsilon_t^2h(t,\theta_0)/\bar\sigma_t^2(\theta_0))]\\
&+\frac{1}{n}\sum_{t=1}^n[(a_0+b_0\varepsilon_t^2)^{\frac{\alpha_0}{2}}-1]+\frac{1}{2n}\sum_{t=1}^n(a_0+b_0\varepsilon_t^2)^{\frac{\alpha_0}{2}}\log(a_0+b_0\varepsilon_t^2)(\hat\alpha-\alpha_0)\Big\}(1+o_p(1))\\
=\;&\frac{\alpha_0}{2}(\hat a-a_0)\mathrm{E}[a_0+b_0\varepsilon_t^2]^{\frac{\alpha_0}{2}-1}+\frac{e_0}{2}(\hat b-b_0)-\frac{b_0e_0}{2}(\hat\theta-\theta_0)^T\lim_{t\to\infty}\mathrm{E}(h(t,\theta_0)/\bar\sigma_t^2(\theta_0))\\
&+\frac{1}{n}\sum_{t=1}^n[(a_0+b_0\varepsilon_t^2)^{\frac{\alpha_0}{2}}-1]+\frac{1}{2}(\hat\alpha-\alpha_0)\mathrm{E}[(a_0+b_0\varepsilon_t^2)^{\frac{\alpha_0}{2}}\log(a_0+b_0\varepsilon_t^2)]+o_p(n^{-\frac12}),
\end{aligned}$$

where e_0 is given in Theorem 2.1. Hence

$$\begin{aligned}
&\sqrt{n}(\hat\alpha-\alpha_0)\\
=\;&A_0^{-1}\sqrt{n}\left\{\alpha_0(\hat a-\hat a_0)\mathrm{E}[a_0+b_0\varepsilon_t^2]^{\frac{\alpha_0}{2}-1}+e_0(\hat b-\hat b_0)-\frac{b_0e_0}{2}(\hat\theta-\theta_0)^T\lim_{t\to\infty}\mathrm{E}\left(\frac{h(t,\theta_0)}{\bar\sigma_t^2(\theta_0)}\right)\right\}\\
&+\frac{2A_0^{-1}}{\sqrt{n}}\sum_{t=1}^n[(a_0+b_0\varepsilon_t^2)^{\frac{\alpha_0}{2}}-1]+o_p(1).
\end{aligned}\tag{5.39}$$

Thus, by (**??**) and the CLT for martingales (see Hall and Heyde (1980)), the right-hand side of (**??**) converges in distribution to a Gaussian distribution with asymptotic variance $\gamma_{\alpha_0}^2$. □

Proof of Theorem 2.2. Put

$$\theta=\theta_0+n^{-1/2}\nu,\quad \nu=(\nu_1,\nu_2,\nu_3)^T\quad\text{and}\quad S_{11}=\mathrm{E}\{Z_1(\theta_0,\alpha_0)Z_1^T(\theta_0,\alpha_0)\}.$$

Then by Lemmas 5.1, 5.2 and some similar arguments as in the proof of Theorem 1 of Owen (1990), we have

$$\begin{aligned}
&l(\theta,\alpha_0)\\
=\;&\left(\frac{1}{\sqrt n}\sum_{t=1}^n\bar Z_t(\theta,\alpha_0)\right)^T\left(\frac1n\sum_{t=1}^n\bar Z_t(\theta,\alpha_0)\bar Z_t^T(\theta,\alpha_0)\right)\left(\frac{1}{\sqrt n}\sum_{t=1}^n\bar Z_t(\theta,\alpha_0)\right)(1+o_p(1))\\
=\;&\left(\frac{1}{\sqrt n}\sum_{t=1}^n\bar Z_t(\theta,\alpha_0)\right)^TS_{11}^{-1}\left(\frac{1}{\sqrt n}\sum_{t=1}^n\bar Z_t(\theta,\alpha_0)\right)(1+o_p(1)),
\end{aligned}\tag{5.40}$$

holds uniformly for all $\theta \in \Theta$. Especially,

$$l(\theta_0, \alpha_0) = \left(\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \bar{Z}_t(\theta_0, \alpha_0)\right)^T S_{11}^{-1} \left(\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \bar{Z}_t(\theta_0, \alpha_0)\right) (1 + o_p(1)). \quad (5.41)$$

Put $\Delta_n(\theta) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \bar{Z}_t(\theta, \alpha_0)$. Then

$$\begin{aligned} l(\theta, \alpha_0) - l(\theta_0, \alpha_0) &= (\Delta_n(\theta) - \Delta_n(\theta_0))^T S_{11}^{-1} \Delta_n(\theta_0) + \Delta_n^T(\theta_0) S_{11}^{-1} (\Delta_n(\theta) - \Delta_n(\theta_0)) \\ &\quad + (\Delta_n(\theta) - \Delta_n(\theta_0))^T S_{11}^{-1} (\Delta_n(\theta) - \Delta_n(\theta_0)) + o_p(1) \quad \text{and} \\ \Delta_n(\theta) - \Delta_n(\theta_0) &= \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \Big\{ [(a + bY_t^2/\bar{\sigma}_t^2(\theta))^{\alpha_0/2} - (a_0 + b_0 Y_t^2/\bar{\sigma}_t^2(\theta_0))^{\alpha_0/2}], \\ &\quad \{\text{sgn}[\log(Y_t^2/\bar{\sigma}_t^2(\theta))] - \text{sgn}[\log(Y_t^2/\bar{\sigma}_t^2(\theta_0))]\} h^T(t, \theta)/\bar{\sigma}_t^2(\theta) \\ &\quad + \text{sgn}[\log(Y_t^2/\bar{\sigma}_t^2(\theta_0))][h^T(t, \theta)/\bar{\sigma}_t^2(\theta) - h^T(t, \theta_0)/\bar{\sigma}_t^2(\theta_0)] \Big\}^T \\ &=: \frac{1}{\sqrt{n}} \sum_{t=1}^{n} (Z_{t1}(\nu), g_A(t, \nu) + g_B(t, \nu))^T. \end{aligned}$$

By Taylor expansion (see also the expression for (**??**)), it follows that

$$\begin{aligned} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} Z_{t1}(\nu) &= \frac{\alpha_0}{2n} \sum_{t=1}^{n} [a_0 + b_0 \varepsilon_t^2]^{\frac{\alpha_0}{2} - 1} \left[\nu_2 + \nu_3 \varepsilon_t^2 - \nu^T b_0 \varepsilon_t^2 \frac{h(t, \theta_0)}{\bar{\sigma}_t^2(\theta_0)}\right] + o_p(1) \\ &= \frac{\nu^T \alpha_0}{2n} \sum_{t=1}^{n} [a_0 + b_0 \varepsilon_t^2]^{\frac{\alpha_0}{2} - 1} \Big[-\frac{b_0 \varepsilon_t^2 h_1(t, \theta_0)}{\bar{\sigma}_t^2(\theta_0)}, 1 - \frac{b_0 \varepsilon_t^2 h_2(t, \theta_0)}{\bar{\sigma}_t^2(\theta_0)}, \\ &\quad \varepsilon_t^2 - \frac{b_0 \varepsilon_t^2 h_3(t, \theta_0)}{\bar{\sigma}_t^2(\theta_0)} \Big]^T (1 + o_p(1)) \qquad (5.42) \end{aligned}$$

holds uniformly for all $\theta \in \Theta$. By Lemma 5.3 and Taylor expansion, we have

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} g_A(t, \nu) = \frac{-2\nu^T f(0)}{n} \sum_{t=1}^{n} \left(\frac{h(t, \theta_0)}{\sigma_t^2(\theta_0)}\right) \left(\frac{h(t, \theta_0)}{\sigma_t^2(\theta_0)}\right)^T (1 + o_p(1)) \text{ and} \qquad (5.43)$$

$$\begin{aligned} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} g_B(t, \nu) &= \frac{\nu^T}{n} \sum_{t=1}^{n} \text{sgn}(\log \varepsilon_t^2) \Big[\sigma_t^{-2}(\theta_0) \Big(\frac{\partial h_1(t, \theta_0)}{\partial \theta}, \frac{\partial h_2(t, \theta_0)}{\partial \theta}, \frac{\partial h_3(t, \theta_0)}{\partial \theta}\Big) \\ &\quad - \Big(h(t, \theta_0)/\sigma_t^2(\theta_0)\Big) \Big(h(t, \theta_0)/\sigma_t^2(\theta_0)\Big)^T \Big] (1 + o_p(1)) \qquad (5.44) \end{aligned}$$

holds uniformly for all $\theta \in \Theta$, where $\frac{\partial h_i(t,\theta_0)}{\partial \theta} = (\frac{\partial h_i(t,\theta_0)}{\partial \theta_1}, \frac{\partial h_i(t,\theta_0)}{\partial \theta_2}, \frac{\partial h_i(t,\theta_0)}{\partial \theta_3})^T$ for $i = 1, 2, 3$. Since the median of $\text{sgn}(\log \varepsilon_t^2)$ is zero, it follows from the weak law of large numbers for a martingale that the right-hand side of (**??**) converges to zero in probability. Put $d_0 = b_0 \mathrm{E}[(a_0 + b_0\varepsilon_t^2)^{\alpha_0/2-1}\varepsilon_t^2]$ and define

$$\begin{aligned} A_1 &= \lim_{t\to\infty} \frac{\alpha_0 d_0}{2}\left[-\mathrm{E}\left(\frac{h_1(t,\theta_0)}{\bar\sigma_t^2(\theta_0)}\right), \frac{\mathrm{E}[a_0+b_0\varepsilon_t^2]^{\frac{\alpha_0}{2}-1}}{d_0} - \mathrm{E}\left(\frac{h_2(t,\theta_0)}{\bar\sigma_t^2(\theta_0)}\right), \mathrm{E}\left(\frac{1}{b_0} - \frac{h_3(t,\theta_0)}{\bar\sigma_t^2(\theta_0)}\right)\right]^T, \\ A_2 &= \lim_{t\to\infty} \mathrm{E}\left[-2f(0)\left(h(t,\theta_0)/\sigma_t^2(\theta_0)\right)\left(h(t,\theta_0)/\sigma_t^2(\theta_0)\right)^T\right] \end{aligned}$$

and $A = (A_1, A_2)$. It follows from (**??**)–(**??**) that

$$l(\theta,\alpha_0) - l(\theta_0,\alpha_0) = \left(\nu^T A S_{11}^{-1}\Delta_n(\theta_0) + \Delta_n^T(\theta_0) S_{11}^{-1} A^T\nu + \nu^T A S_{11}^{-1} A^T \nu\right)(1+o_p(1)) \quad (5.45)$$

holds uniformly for all $\theta \in \Theta$. Like the proof of Lemma 1 of Qin and Jin (1994), we know that the minimizer $\hat\theta = \theta_0 + n^{-1/2}\nu$ of (**??**) must lie in Θ. Thus, by minimizing (**??**) with respect to ν, it follows that

$$\hat\nu = -(AS_{11}^{-1}A^T)^{-1} A S_{11}^{-1}\Delta_n(\theta_0) + o_p(1).$$

Substitute this into (**??**), we have

$$\begin{aligned} & l(\hat\theta,\alpha_0) \\ = \; & [S_{11}^{-1/2}\Delta_n(\theta_0)]^T [I - S_{11}^{-1/2} A^T (AS_{11}^{-1}A^T)^{-1} A S_{11}^{-1/2}][S_{11}^{-1/2}\Delta_n(\theta_0)](1+o_p(1)). \end{aligned} \quad (5.46)$$

By Lemma 5.2, $S_{11}^{-1/2}\Delta_n(\theta_0)$ converges in distribution to a multivariate standard normal distribution. Thus, by (**??**) and noting that the trace of $I - S_{11}^{-1/2}A^T(AS_{11}^{-1}A^T)^{-1}AS_{11}^{-1/2}$ is 1, we have $l(\hat\theta,\alpha_0) \xrightarrow{d} \chi^2(1)$, i.e., Theorem **??** follows. □

5.2 Proofs for AR(1)-ARCH(1) Case

In this subsection, we define $\Theta = \{\theta : ||\theta - \theta_0|| \le n^{-\frac{1}{2\gamma}}\}$ for some $\gamma \in (1, \alpha_0)$.

Lemma 5.4. *Under conditions of Theorem* **??***,*

$$\sup_{1\le t\le n}\sup_{\theta\in\Theta} ||X_t(\theta,\alpha_0)|| = o_p(n^{\frac{1}{2\gamma}}). \quad (5.47)$$

Proof. Define $G(t,\theta) = (X_{t2}(\theta), X_{t3}(\theta), X_{t4}(\theta))^T$ and

$$H(t,\theta) = (1, 2(Y_t - aY_{t-1})Y_{t-1}, Y_{t-1}^2)^T/(1+Y_{t-1}^2).$$

Then $G(t,\theta) = -\text{sgn}(\varepsilon_t^2(\omega,a,b))H(t,\theta)$ and

$$\sup_{\theta\in\Theta} ||G(t,\theta)|| \leq \frac{1}{(1+Y_{t-1}^2)} \left|\left|(1, 2|\varepsilon_t Y_{t-1}| + 2n^{-\frac{1}{2\gamma}} Y_{t-1}^2, Y_{t-1}^2)^T\right|\right| \leq C\left(1 + \frac{|\varepsilon_t Y_{t-1}|}{1+Y_{t-1}^2}\right).$$

Since $\text{E}[|\varepsilon_t Y_{t-1}|/(1+Y_{t-1}^2)]^{2\alpha_0} < \infty$, we can show that

$$\sup_{1\leq t\leq n}\sup_{\theta\in\Theta} ||G(t,\theta)|| = o_p(n^{\frac{1}{2\gamma}}).$$

Further, by Lemma 4.1 of Chan, Li, Peng and Zhang (2013), we have

$$\sup_{1\leq t\leq n}\sup_{\theta\in\Theta} |X_{t1}(\theta,\alpha_0)| = o_p(n^{\frac{1}{2\gamma}}).$$

Hence Lemma 5.4 follows from the above two equations. □

Proof of Theorem **??**. It follows from Theorem 1 of Chan and Peng (2005) that

$$\sqrt{n}\{\hat{\theta} - \theta_0\} = \frac{f^{-1}(1)\Gamma_2\Gamma_1^{-1}}{\sqrt{n}} \sum_{t=1}^n \bar{A}(t) + o_p(1) \xrightarrow{d} N(0,\Omega_1) \tag{5.48}$$

where $\Omega_1 = f^{-2}(1)\Gamma_2\Gamma_1^{-1}\text{Cov}\{\bar{A}(t)\}\Gamma_1^{-1}\Gamma_2$.

Like the proofs in Chan, Li, Peng and Zhang (2013), we can show that

$$\begin{aligned}
&\sqrt{n}(\hat{\alpha} - \alpha_0) \\
= \; & \sqrt{n}(\hat{\omega} - \omega_0)\kappa_0^{-1}\text{E}\Big[\frac{\sqrt{b_0}(\alpha_0|a_0 + \sqrt{b_0}\varepsilon_2|^{\alpha_0-1}\text{sgn}(a_0 + \sqrt{b_0}\varepsilon_2))\varepsilon_2}{2(\omega_0 + b_0Y_1^2)}\Big] \\
& + \sqrt{n}(\hat{a} - a_0)\kappa_0^{-1}\text{E}\Big[(\alpha_0|a_0 + \sqrt{b_0}\varepsilon_2|^{\alpha_0-1}\text{sgn}(a_0 + \sqrt{b_0}\varepsilon_2))\Big(\frac{\sqrt{b_0}Y_1}{\sqrt{\omega_0 + b_0Y_1^2}} - 1\Big)\Big] \\
& + \sqrt{n}(\hat{b} - b_0)\kappa_0^{-1}\text{E}\Big[(\alpha_0|a_0 + \sqrt{b_0}\varepsilon_2|^{\alpha_0-1}\text{sgn}(a_0 + \sqrt{b_0}\varepsilon_2))[\frac{\sqrt{b_0}\varepsilon_2Y_1^2}{2(\omega_0 + b_0Y_1^2)} - \frac{\varepsilon_2}{2\sqrt{b_0}}]\Big] \\
& - \frac{1}{\kappa_0\sqrt{n}} \sum_{t=1}^n (|a_0 + \sqrt{b_0}\varepsilon_t|^{\alpha_0} - \text{E}|a_0 + \sqrt{b_0}\varepsilon_1|^{\alpha_0}) + o_p(1).
\end{aligned} \tag{5.49}$$

Thus, Theorem **??** follows from (**??**), (**??**) and the CLT for martingales. □

Lemma 5.5. *Under conditions of Theorem 2.3, we have, as $n \to \infty$*

$$(a) \sup_{\theta\in\Theta} ||\frac{1}{n}\sum_{t=1}^{n} X_t(\theta,\alpha_0)X_t^T(\theta,\alpha_0) - \mathrm{E}\{X_1(\theta_0,\alpha_0)X_1^T(\theta_0,\alpha_0)\}|| = o_p(1);$$

$$(b) \frac{1}{\sqrt{n}}\sum_{t=1}^{n} X_t(\theta_0,\alpha_0) \xrightarrow{d} \mathrm{N}\Big(0, \mathrm{E}\{X_1(\theta_0,\alpha_0)X_1^T(\theta_0,\alpha_0)\}\Big).$$

Proof. Conclusion (a) can be proved in a way similar to the proof of Lemma 5.3.

Conclusion (b) follows from the CLT for martingales by noting that

$$\begin{aligned} X_t(\theta_0,\alpha_0) &= \Big(|a_0 + \sqrt{b_0}\varepsilon_t|^{\alpha_0} - 1, -\tfrac{\mathrm{sgn}(\varepsilon_t^2-1)}{1+Y_{t-1}^2}, \\ &\quad -\frac{2(\omega_0 + b_0 Y_{t-1}^2)^{\frac{1}{2}} Y_{t-1}\mathrm{sgn}(\varepsilon_t^2-1)\varepsilon_t}{1+Y_{t-1}^2}, \frac{Y_{t-1}^2\mathrm{sgn}(\varepsilon_t^2-1)}{1+Y_{t-1}^2}\Big)^T \end{aligned}$$

is a martingale difference sequence. □

Lemma 5.6. *Under conditions of Theorem* **??***,*

$$\begin{aligned} \sup_{\theta\in\Theta}\Big\|\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\Big\{ &\left[sgn\left(\varepsilon_t^2(\omega,a,b)\right) - sgn\left(\varepsilon_t^2 - 1\right)\right] H(t,\theta) \\ &+2(\theta-\theta_0)^T f_{\varepsilon_1^2}(1) H(t,\theta_0) H^T(t,\theta_0)\Big\}\Big\| = o_p(1), \end{aligned}$$

where $f_{\varepsilon_1^2}(\cdot)$ denotes the density of ε_1^2 and $H(t,\theta)$ is defined in the beginning of proof of Lemma 5.4.

Proof. This lemma can be proved in a way similar to the proof of Lemma 5.3, hence we omit the details. □

Proof of Theorem 2.4. Theorem 2.4 can be shown similar to the proof of Theorem 2.2 by using Lemmas 5.5 and 5.6. □

Acknowledgments

We appreciate the editor, an associate editor and two reviewers for helpful comments, which significantly improve the presentation of this paper. Zhang's research was partly supported by NSFC grant 11371318 and Natural Science Foundation of Zhejiang province

(R16A010001). Peng's research was partly supported by Simons Foundation and the Georgia State University CAE grant from the Society of Actuaries.

References

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

Berkes, I. and Horváth, L. (2004). The efficiency of the estimators of the parameters in GARCH processes. *Ann. Stat.* **32**, 633–655.

Berkes, I., Horváth, L. and Kokoszka, P. (2003). Estimation of the maximal moment exponent of a GARCH (1,1) sequence. *Econometric Theory* **19**, 565–586.

Borkovec, M. (2000). Extremal behavior of the autoregressive process with ARCH(1) errors. *Stochastic Processes and Their Applications* **85**, 189–207.

Borkovec, M. (2001). Asymptotic behavior of the sample autocovariance and autocorrelation function of the AR(1) process with ARCH errors. *Bernoulli* **7**, 847–872.

Borkovec, M. and Klüppelberg, C. (2001). The tail of the stationary distribution of an autoregressive process with ARCH(1) errors. *Ann. Appl. Probab.* **11**, 1220–1241.

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

Chan, N. H., Peng, L. and Zhang, R. M. (2012). Interval estimation for GARCH(1,1) tail index. *Test* **21**, 546–565.

Chan, N. H., Li, D. Y., Peng, L. and Zhang, R. M. (2013). Interval estimation of the tail index of an AR(1) with ARCH(1) errors. *Econometric Theory* **29**, 920–940.

Cont, R. and Tankov, P. (2004). *Financial Modeling with Jump Processes*. Chapman and Hall, New York.

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

Davis, R. A. and Resnick, S. I. (1985). Limit theory for moving averages of random variables with regularly varying tail probabilities. *Ann. Probab.* **13**, 179–195.

Davis, R. A. and Resnick, S. I. (1986). Limit theory for the sample covariance and correlation functions of moving averages. *Ann. Statist.* **14**, 533–558.

Davis, R. A. and Resnick, S. I. (1996). Limit theory for bilinear processes with heavy tailed noise. *Ann. Appl. Probab.* **6**, 1191–1210.

De Haan, L., Resnick, S. I., Rootzén, H. and de Vries, C. G. (1989). Extremal behavior of solutions to a stochastic difference equation with applications to ARCH processes. *Stochastic Processes and Their Applications* **32**, 213–224.

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

Fan, J. and Yao, Q. (2003). Nonlinear Time Series: Nonparametric and Parametric Methods. Springer, New York.

Fan, J., Qi, L., and Xiu, D. (2014). Quasi maximum likelihood estimation of GARCH models with heavy-tailed likelihoods. *Journal of Business and Economics Statistics* **32**, 178–205.

Francq, C. and Zakoïan, J. (2010). GARCH Models: Structure, Statistical Inference and Financial Applications. *John Wiley & Sons.*

Goldie, C. M. (1991). Implicit renewal theory and tails of solutions of random equations. *Ann. Appl. Probab.* **1**, 126–166.

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

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

Hansen, B. E. (1996). Stochastic equicontinuity for unbounded dependent heterogeneous arrays. *Econometric Theory* **12**, 347–359.

Hill, B. M. (1975). A simple general approach to inference about the tail of a distribution. *Ann. Statist.* **3**, 1163–1174.

Hill, J.B. (2015). Robust estimation and inference for heavy tailed GARCH. *Bernoulli* **21**, 1629–1669.

Hill, J.B. and Prokhorov, A. (2016). GEL estimation for heavy-tailed GARCH models with robust empirical likelihood inference. *J. Econometrics* **190**, 18–45.

Hsing, T. (1991). On tail index estimation using dependent data. *Ann. Statist.* **19**, 1547–1569.

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

Koul, H. L. and Surgailis, D. (2001). Asymptotics of empirical processes of long memory moving averages with infinite variance. *Stoch. Process. Appl.* **91**, 309–336.

Lange, T. (2011) Tail behavior and OLS estimation in AR-GARCH models. *Statistica Sinica* **21**, 1191–1200.

Ling, S. (2004). Estimation and testing stationarity for double autoregressive models. *J. R. Statist. Soc. B* **66**, 63–78.

Ling, S. and Peng, L. (2004). Hill's estimator for the tail index of an ARMA model. *Journal of Statistical Planning and Inference* **123**, 279–293.

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. *Econometric Theory* **6**, 318–334.

Owen, A. B. (1990). Empirical likelihood confidence regions. *Ann. Statist.* **18**, 90–120.

Peng, L. and Yao, Q. (2003). Least absolute deviations estimation for ARCH and GARCH models. *Biometrika* **90**, 967–975.

Qin, J. and Lawless, J. (1994). Empirical likelihood and general estimating equations. *Ann. Statist.* **22**, 300–325.

Quintos, C., Fan, Z. and Phillips, P. C. B. (2001). Structural change tests in tail behaviour and the Asian crisis. *Review of Economic Studies* **68**, 633–663.

Resnick, S. I. and Starica, C. (1997). Asymptotic behavior of Hill's estimator for autoregressive data. *Stochastic Models* **13**, 703–723.

Resnick, S. I. and Starica, C. (1998). Tail index estimation for dependent data. *Ann. Appl. Probab.* **8**, 1156–1183.

Resnick, S. I. and Van den Berg, E. (2000). Sample correlation behavior for the heavy tailed general bilinear process. *Stochastic Models* **16**, 233–258.

Taylor, S. J. (2005). *Asset Price Dynamics, Volatility, and Prediction.* Princeton University Press, New Jersey.

Wagner, N. and Marsh, T. A. (2005). Measuring tail thickness under GARCH and an application to extreme exchange rate changes. *Journal of Empirical Finance* **12** 165–185.

Zhang, R. M. and Ling, S. (2015). Asymptotic inference for AR models with heavy-tailed G-GARCH noises. *Econometric Theory* 31, 880–890.

Zhu, K. and Ling, S. (2015). LADE-based inference for ARMA models with unspecified and heavy-tailed heteroscedastic noises. *Journal of American Statistical Association* **110**, 784–794.

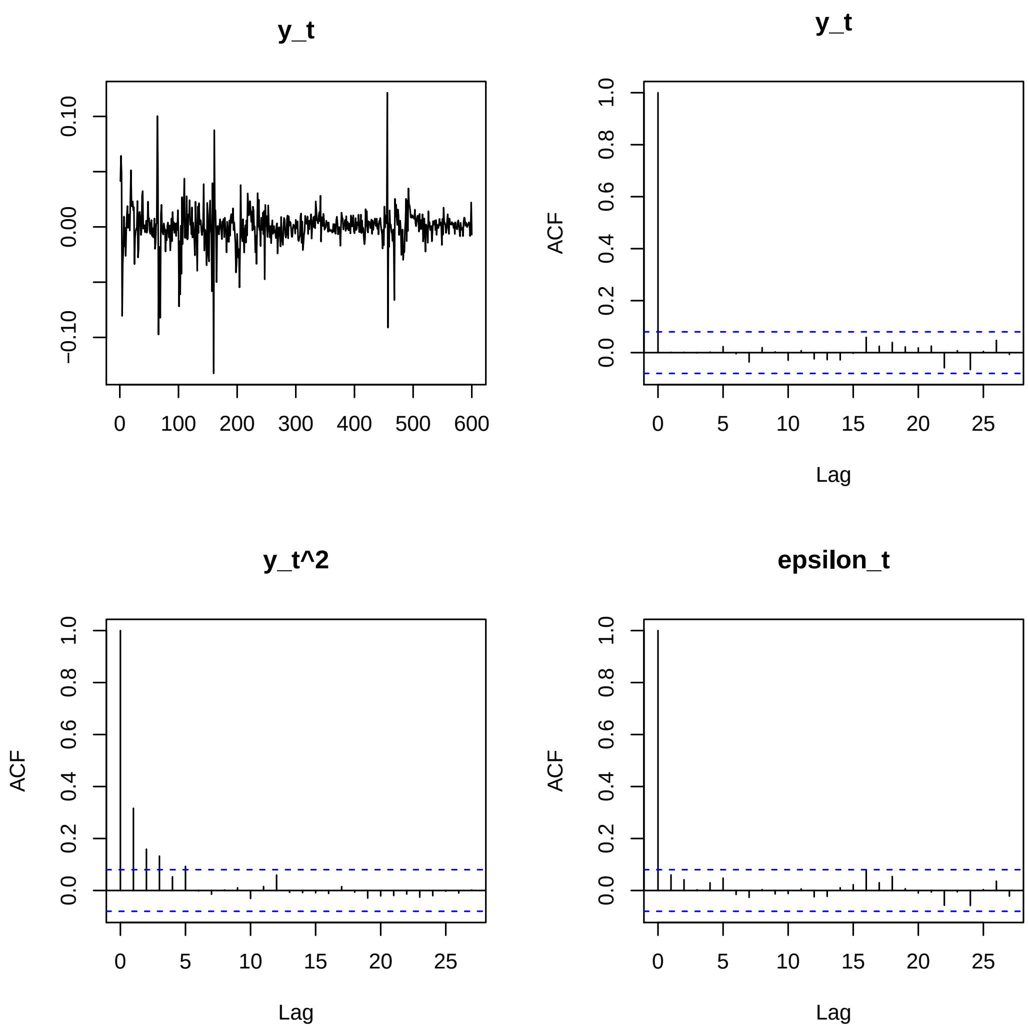

Figure 1: We plot $\{X_t\}$, the autocorrelation functions of $\{Y_t\}$, $\{Y_t^2\}$ and estimated $\{\varepsilon_t\}$ respectively.

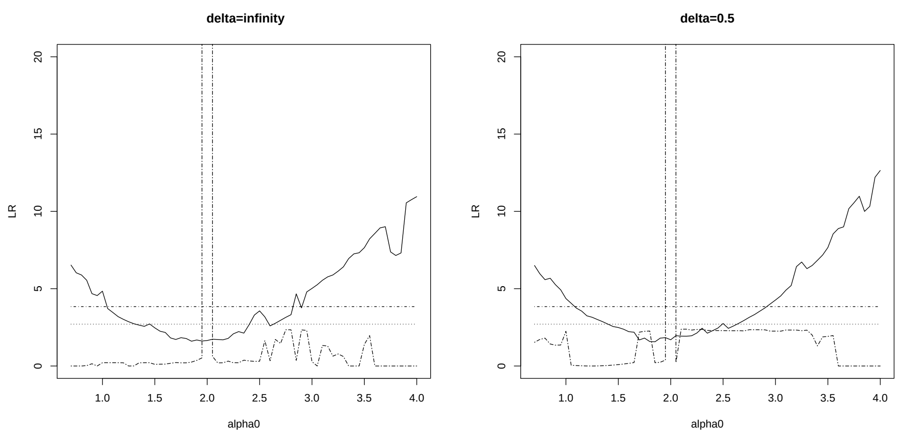

Figure 2: The profile empirical likelihood ratios based on both LADE and QMLE are plotted against $\alpha = 0.7, 0.75, \cdots, 4$ in solid line and dotted line, respectively. Two straight lines represent the 90th and 95th quantile of $\chi^2(1)$ respectively.

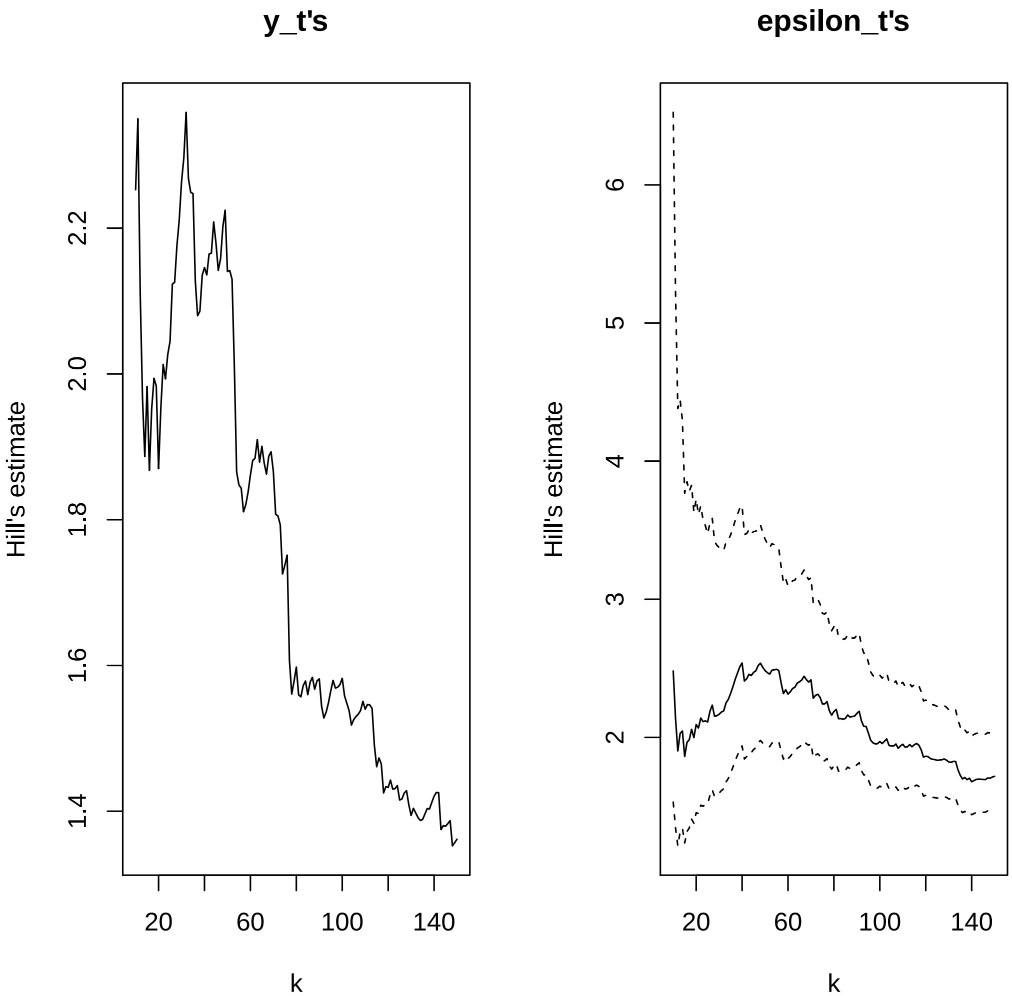

Figure 3: Hill's estimator in (**??**) is plotted against $k = 10, 11, \cdots, 150$ for $\{Y_t\}$ and estimated $\{\varepsilon_t\}$.

Table 1: Coverage probabilities based on the method in Section 2.1 (I_{ξ}^{LADE}) and the method in Chan, Peng and Zhang (2012) (I_{ξ}^{QMLE}) are calculated for $w_0 = 1$ and $\varepsilon_t^* \sim t(\nu)/\sqrt{\nu/(\nu-2)}$. Here $\nu = \infty$ means $\varepsilon_t^* \sim N(0,1)$.

(a_0, b_0, ν, n)	$I_{0.90}^{LADE}$	$I_{0.95}^{LADE}$	$I_{0.90}^{QMLE}$	$I_{0.95}^{QMLE}$	α_0
$(0.72, 0.38, 3.2, 500)$	0.8892	0.9340	0.9700	0.9882	1.1678
$(0.65, 0.38, 4, 500)$	0.8988	0.9528	0.9374	0.9682	1.7367
$(0.65, 0.25, 8, 500)$	0.9044	0.9530	0.9536	0.9734	3.8212
$(0.6, 0.25, 12, 500)$	0.9000	0.9526	0.9174	0.9546	4.7681
$(0.72, 0.38, 3.2, 1000)$	0.9052	0.9484	0.9842	0.9930	1.1678
$(0.65, 0.38, 4, 1000)$	0.9034	0.9538	0.9552	0.9848	1.7367
$(0.65, 0.25, 8, 1000)$	0.8940	0.9468	0.9504	0.9752	3.8212
$(0.6, 0.25, 12, 1000)$	0.8988	0.9514	0.9102	0.9586	4.7681
$(0.72, 0.38, 3.2, 2000)$	0.9064	0.9524	0.9854	0.9948	1.1678
$(0.65, 0.38, 4, 2000)$	0.9130	0.9616	0.9694	0.9872	1.7367
$(0.65, 0.25, 8, 2000)$	0.8978	0.9492	0.9222	0.9724	3.8212
$(0.6, 0.25, 12, 2000)$	0.9020	0.9476	0.8768	0.9416	4.7681