Estimation for a non-stationary semi-strong GARCH(1,1) model with heavy-tailed errors

Oliver Linton
London School of Economics

Jiazhu Pan
University of Strathclyde and Peking University

Hui Wang
Central University of Finance and Economics, China

This paper studies the estimation of a semi-strong GARCH(1,1) model when it does not have a stationary solution, where semi-strong means that we do not require the errors to be independent over time. We establish necessary and sufficient conditions for a semi-strong GARCH(1,1) process to have a unique stationary solution. For the non-stationary semi-strong GARCH(1,1) model, we prove that a local minimizer of the least absolute deviations (LAD) criterion converges at the rate $\sqrt{n}$ to a normal distribution under very mild moment conditions for the errors. Furthermore, when the distributions of the errors are in the domain of attraction of a stable law with the exponent $\kappa \in (1,2)$, it is shown that the asymptotic distribution of the Gaussian quasi-maximum likelihood estimator (QMLE) is non-Gaussian but is some stable law with the exponent $\kappa \in (0,2)$. The asymptotic distribution is difficult to estimate using standard parametric methods. Therefore, we propose a percentile-t subsampling bootstrap method to do inference when the errors are independent and identically distributed, as in Hall and Yao (2003). Our result implies that the least absolute deviations estimator (LADE) is always asymptotically normal regardless of whether there exists a stationary solution or not even when the errors are heavy-tailed. So the LADE is more appealing when the errors are heavy-tailed. Numerical results lend further support to our theoretical results.

The authors thank the two referees and the co-editor for their valuable suggestions. Thanks to Professor Qiwei

Yao for helpful discussion on programming of bootstrap method. Oliver Linton was supported by the ESRC (UK), and Jiazhu Pan was partially supported by the starter grant from University of Strathclyde (UK) and the National Basic Research Program of China (grant No. 2007CB814902). Address correspondence to Jiazhu Pan, Department of Statistics and Modelling Science, University of Strathclyde, Richmond Street, Glasgow, G1 1XH, UK; e-mail:jiazhu@stams.strath.ac.uk. Or to Oliver Linton, Department of Economics, London School of Economics, Houghton Street, London, WC2A 2AE, UK; e-mail: o.linton@lse.ac.uk.

1 Introduction

Since the seminal work of Engle (1982), ARCH/GARCH models have been widely used in finance and economics, see Shephard (1996) and Rydberg (2000). The first order generalized autoregressive conditional heteroscedastic (GARCH (1,1)) model is given by

$$X_t = \sigma_t \varepsilon_t \quad and \quad \sigma_t^2 = \omega + \alpha X_{t-1}^2 + \beta \sigma_{t-1}^2,$$

where $\omega > 0, \alpha \geq 0, \beta \geq 0$ are unknown parameters, while $\{\varepsilon_t\}$ is a sequence of independent and identically distributed (i.i.d) random variables with mean 0 and variance 1, and ε_t is independent of $\{X_{t-k}, k \geq 1\}$ for all t, see Bollerslev (1986).

Nelson (1990) proved that there exists a unique strictly stationary and ergodic solution to GARCH(1,1) model if and only if

$$E \log(\alpha \varepsilon_t^2 + \beta) < 0.$$

Bougerol and Picard (1992) extended this result to the GARCH(p,q) case. Pan et al. (2008) establish this result for a more general class of models under an additional moment condition that $E|\varepsilon_t|^{\varrho} < +\infty$ for some $\varrho > 0$. Many authors have studied the asymptotic inference for stationary ARCH/GARCH models. When the errors have finite fourth moment, i.e., $E\varepsilon_t^4 < \infty$, the consistency and asymptotic normality of quasi-maximum likelihood estimators (QMLE) for ARCH/GARCH models have been established under different conditions, see Weiss (1986), Lee and Hansen (1994), Lumsdaine (1996), Berkes et.al. (2003) etc. Mikosch and Straumann (2002) adapted Whittle estimation to a heavy-tailed GARCH(1,1) model where X_t has a Pareto-like tail with tail index $\kappa > 4$. They showed that the Whittle estimator converges in distribution to an infinite series of a sequence of $\kappa/4$-stable random variables provided $\kappa < 8$ and $E\varepsilon_t^8 < +\infty$, and a normal random variable provided $\kappa > 8$. For a heavy-tailed ARCH (1) processes where X_t has a Pareto-like tail with tail index $0 < \kappa < 4$, Davis and Mikosch (1998) established the asymptotic theory for sample autocorrelation functions with the speed of convergence slower than $\sqrt{n}$, and Mikosch and Stărică (2000) extended the results to GARCH(1,1) model. In the case that $E\varepsilon_t^4 = \infty$, the asymptotic theory for QMLE becomes quite complicated and difficult. Hall and Yao (2003) studied the QMLE for heavy-tailed GARCH models with the errors in the domain of attraction of a stable law with exponent between 1 and 2. They showed that the asymptotic distribution may be non-Gaussian and the convergence rate is slower than $\sqrt{n}$. Straumann (2005) established similar results for a more general class of GARCH-type models. Peng and Yao (2003) show that, in contrast, the least absolute deviations estimator (LADE) is asymptotically Gaussian with convergence rate $\sqrt{n}$ provided $E\varepsilon_t^2 < +\infty$. In fact, their conditions on the error moments can be reduced to $E|\varepsilon_t|^{\varrho} < +\infty$ for some $\varrho > 0$, which is more appealing in dealing with heavy-tailed processes; see Pan et al.(2008).

Jensen and Rahbek (2004 a, 2004 b) were the first to consider the asymptotic theory of the QMLE for non-stationary ARCH/GARCH models. They showed that the likelihood-based estimator for the parameters in the first order ARCH/GARCH model is consistent and asymptotically Gaussian in the entire parameter region regardless of whether the process is strictly stationary or explosive, i.e. even for the case that $E\log(\alpha\varepsilon_t^2 + \beta) \geq 0$. But they assumed that the errors have finite fourth moment, i.e. $E\varepsilon_t^4 < \infty$. So the inferential theory for a non-stationary ARCH/GARCH model with errors with infinite fourth moments remains open.

Economic and financial time series often appear to be non-stationary and/or driven by heavy-tailed noises, see Mandelbrot (1963), Mittnik et al. (1998), Mittnik and Rachev (2000), Engle and Rangel (2005), and Polzehl and Spokoiny (2004). Furthermore, as Lee and Hansen (1994) have pointed out, there is no reason to assume that all of the conditional dependence is contained in the conditional variance. Thus, we assume that $\{\varepsilon_t\}$ are stationary and ergodic, and call a GARCH(1, 1) model with such errors a semi-strong GARCH(1, 1) model, following Drost and Nijman (1993). Lee and Hansen (1994) established the asymptotic normality of the QMLE for strictly stationary semi-strong GARCH(1, 1) model with errors such that their fourth moments conditional on the past are uniformly bounded. If we assume that the conditional second and fourth moments of the error ε_t equals its unconditional second and fourth moments a.s. respectively, the proof of Jensen and Rahbek (2004 a, 2004 b) still gets through for non-stationary semi-strong GARCH (1, 1) models with minor modification. Hence, it is meaningful to study the estimation problem for non-stationary semi-strong ARCH/GARCH models with errors with infinite fourth moment.

In this paper, we give necessary and sufficient conditions for a semi-strong GARCH(1, 1) model to have a unique stationary solution. We then study the estimation for the non-stationary semi-strong GARCH(1, 1) model in the case that $E\varepsilon_t^4 = \infty$. We show that the proposed LADE is asymptotically normal if the conditional expectation of $|\varepsilon_t|^{2+\delta}$ is uniformly bounded for some $\delta > 0$ and the conditional densities of $\log \varepsilon_t^2$ given the past satisfy some regular conditions. If the errors of a non-stationary GARCH(1, 1) model are i.i.d., the moment condition of ε_t for the LADE to have the asymptotic normality can be reduced to $E|\varepsilon_t|^{\varrho} < +\infty$ for some $\varrho > 0$. Based on the asymptotic normality of LADE, some inference on the model can be easily undertaken. For example, a Wald test of some interesting hypotheses can be built. We also investigate the properties of the (Gaussian) QMLE when some mixing condition holds and the distribution of the errors is in the domain of attraction of a stable law with exponent between 1 and 2 and the tails of the conditional distribution of $|\varepsilon_t^2 - 1|$ given the past are uniformly bounded by the tail of some distribution which is in the domain of attraction of a stable law with the same exponent as ε_t^2. The asymptotic distribution of the QMLE is non-Gaussian but some stable law with unknown index $\kappa \in (1, 2)$, which makes inference difficult, and we will use the percentile-t subsampling bootstrap method employed by Hall and Yao (2003) to do statistical inference. Thus, the proposed LADE seems more appealing for the non-stationary semi-strong GARCH(1, 1) model with heavy-tailed errors. Finally, the asymptotic results for QMLE and LADE hold independently of the choice of initial values and the scale parameter.

The rest of paper is organized as follows. Section 2 discusses when a semi-strong GARCH(1, 1) model defines a strictly stationary and ergodic solution and when it has no stationary version. Section 3 discusses estimation of a non-stationary semi-strong GARCH(1, 1) model. Subsection 3.1.1 gives the LADE and its asymptotic properties and Subsection 3.1.2 presents a Wald test based on the result of Subsection 3.1.1. The asymptotic results of QMLE for a non-stationary semi-strong GARCH(1, 1) model with κ-stable errors are presented in Subsection 3.2.1 and Subsection 3.2.2 provides subsampling bootstrap methods to construct confidence intervals. Section 4 reports some numerical results. Section 5 concludes. The appendix contains the proofs of all results.

We denote by $\xrightarrow{\mathrm{P}}$, $\xrightarrow{\mathrm{d}}$ and $\xrightarrow{\mathrm{L_p}}$ the convergence, respectively, in probability, in distribution and in L_p. Denote the Euclidean norm of a vector V by $\| V \|$. Let $A^\top$ denote the transpose of a matrix or a vector A, and let C be a generic constant which may be different at different places. $I(\cdot)$ stands for the indicator function through the whole paper.

2 The solution of the semi-strong GARCH(1, 1) model

Consider the first order semi-strong GARCH(1, 1) model given by

$$X_t = \sigma_t \varepsilon_t \quad and \quad \sigma_t^2 = \omega + \alpha X_{t-1}^2 + \beta \sigma_{t-1}^2, \tag{2.1}$$

where $\omega > 0, \alpha \geq 0, \beta \geq 0$ are unknown parameters, and $\{\varepsilon_t\}$ is a strictly stationary and ergodic sequence of random variables. Denote

$$\gamma = E \log(\alpha_0 \varepsilon_t^2 + \beta_0),$$

where $(\omega_0, \alpha_0, \beta_0)$ is the true value of the parameter of model (2.1). For the semi-strong GARCH(1, 1) model, we can not get the necessary and sufficient conditions for stationarity under the original assumptions on $\{\varepsilon_t\}$. However, imposing some mixing condition on $\{\varepsilon_t\}$ when $\gamma = 0$, we can get Theorem 1 which shows that model (2.1) has a unique strictly stationary and ergodic solution if and only if $\gamma < 0$. The assumptions needed are in the following:

A1. ε_t is strictly stationary and ergodic, ε_t^2 is non-degenerate (ε_t thus need to be different from scaled symmetric Bernoulli or degenerate random variables) and $E|\varepsilon_t|^\varrho < +\infty$ for some $\varrho > 0$.

A2. In the case of $\gamma = 0$, ε_t^2 is φ-mixing with $\sum_{n=1}^{+\infty} \varphi_n^{1/2} < +\infty$, where

$$\varphi_n = \sup_{A \in \mathcal{F}_{-\infty}^0, B \in \mathcal{F}_n^{+\infty}, Pr(A) > 0} \left| Pr(B) - Pr(B|A) \right|$$

and $\mathcal{F}_i^j = \sigma(\varepsilon_t, i \leq t \leq j)$.

Theorem 1. *Suppose that Assumptions A1-A2 hold and $\omega_0 > 0$. Then it follows that the semi-strong GARCH(1,1) model (2.1) defines a unique strictly stationary and ergodic solution if and only if $\gamma < 0$. Furthermore, $\sigma_t^2 \longrightarrow +\infty$ a.s provided $\gamma \geq 0$.*

Remark 1. When ε_t is i.i.d. with $E\varepsilon_t^2 = 1$, the condition for strict stationarity $\gamma < 0$ is weaker than the requirement for weak stationarity, $\beta_0 + \alpha_0 < 1$. What drives the surprising result is the well- known fact that the second moments of a stationary solution to model (2.1) are finite if and only if $\beta_0 + \alpha_0 < 1$. So that, while strict stationarity still holds if $\gamma < 0$ and $\beta_0 + \alpha_0 \geq 1$, weak stationarity fails since variances are infinite and autocovariances are not defined. In the Gaussian ARCH(1) case, one can have even $\alpha_0 < 0.5exp\big(-\Psi(0.5)\big) \approx 3.56$, where $\Psi(\cdot)$ is the Euler psi function. Thus the set of allowable parameter values for strict stationarity is larger than the set of values for weak stationarity. This situation is a bit more complicated when $E\varepsilon_t^2 = \infty$. In particular, Nelson (1990) shows that when ε_t is standard Cauchy, $\gamma = 2\ln({\alpha_0}^{1/2} + {\beta_0}^{1/2})$, so that the set of allowable parameter values for strict stationarity is smaller than the set $\alpha_0 + \beta_0 < 1$ (although in that case the set of parameter values implying weak stationarity is empty due to the infinite second moments).

3 Estimation for a non-stationary semi-strong GARCH(1,1) model

We assume the initial value of X_t is X_0 and that the unobserved σ_0^2 is parameterized by η_0, i.e. $\sigma_0^2 = \eta_0$. The parameter of the model (2.1) is then $\phi = (\alpha, \beta, \omega, \eta)^\top$ with true value $\phi_0 = (\alpha_0, \beta_0, \omega_0, \eta_0)^\top$. Denote $\theta = (\alpha, \beta)^\top$ and $\psi = (\omega, \eta)^\top$ with true value $\theta_0 = (\alpha_0, \beta_0)^\top$ and $\psi_0 = (\omega_0, \eta_0)^\top$ respectively. Let

$$\sigma_t^2(\phi) = \omega + \alpha X_{t-1}^2 + \beta \sigma_{t-1}^2(\phi), \tag{3.1}$$

with $\sigma_0^2(\phi) = \eta$ and $\sigma_t^2(\phi_0) = \sigma_t^2$.

The least absolute deviations estimator (LADE) for model (2.1) is defined as a minimizer of the following objective function

$$S_n(\phi) = \sum_{t=u+1}^{n} |\log X_t^2 - \log \sigma_t^2(\phi)|, \tag{3.2}$$

where $\sigma_t^2(\phi)$ is defined in (3.1), $u = u(n)$ is a nonnegative integer. The quasi-maximum likelihood estimator (QMLE) for model (2.1) is a minimizer of

$$l_n(\phi) = \frac{1}{n}\sum_{t=1}^{n} \big(\log \sigma_t^2(\phi) + \frac{X_t^2}{\sigma_t^2(\phi)}\big), \tag{3.3}$$

where $\sigma_t^2(\phi)$ is defined by (3.1). These objective functions can be computed quite cheaply and many algorithms are available for finding the minima.

We collect here some notation that will be useful in the sequel. Let $Z_t(\phi) = \log X_t^2 - \log \sigma_t^2(\phi)$ and denote $A_t(\phi) = (A_{1t}(\phi), A_{2t}(\phi))^\top$, where

$$A_{1t}(\phi) = \frac{\partial \sigma_t^2(\phi)}{\partial \alpha}\frac{1}{\sigma_t^2(\phi)} = \sum_{j=1}^{t} \beta^{j-1}\frac{X_{t-j}^2}{\sigma_t^2(\phi)}, \tag{3.4}$$

$$A_{2t}(\phi) = \frac{\partial \sigma_t^2(\phi)}{\partial \beta}\frac{1}{\sigma_t^2(\phi)} = \sum_{j=1}^{t} \beta^{j-1}\frac{\sigma_{t-j}^2(\phi)}{\sigma_t^2(\phi)}. \tag{3.5}$$

Then,

$$\frac{\partial Z_t(\phi)}{\partial \alpha} = -A_{1t}(\phi), \qquad \frac{\partial Z_t(\phi)}{\partial \beta} = -A_{2t}(\phi),$$

$$\frac{\partial l_n(\phi)}{\partial \alpha} = \frac{1}{n}\sum_{t=1}^{n}\big(A_{1t}(\phi) - \frac{X_t^2}{\sigma_t^2(\phi)}A_{1t}(\phi)\big), \qquad \frac{\partial l_n(\phi)}{\partial \beta} = \frac{1}{n}\sum_{t=1}^{n}\big(A_{2t}(\phi) - \frac{X_t^2}{\sigma_t^2(\phi)}A_{2t}(\phi)\big).$$

Let $A_t = (A_{1t}, A_{2t})^\top =: A_t(\phi_0)$. Define $D_t(a,b) = (D_{1t}(a,b), D_{2t}(a,b))^\top$, where $a > 0$, $b > 0$ and

$$D_{1t}(a,b) = \sum_{j=1}^{+\infty} a^{j-1}\varepsilon_{t-j}^2 \prod_{k=1}^{j}\frac{1}{\alpha_0\varepsilon_{t-k}^2 + b}, \quad D_{2t}(a,b) = \sum_{j=1}^{+\infty} a^{j-1} \prod_{k=1}^{j}\frac{1}{\alpha_0\varepsilon_{t-k}^2 + b}.$$

Denote $D_t = (D_{1t}, D_{2t})^\top$ with $D_{it} =: D_{it}(\beta_0, \beta_0)$, $i = 1, 2$.

Remark 2. By Lemma 3 in the appendix of this paper, we use two stationary ergodic processes D_{1t} and D_{2t} to approximate A_{1t} and A_{2t} respectively.

3.1 Least absolute deviation estimator

3.1.1 Asymptotic properties of the LADE

The QMLE can be viewed as an extended version of least squares estimation, which is known to be sensitive to heavy-tails, while the LADE would be more robust, see Peng and Yao (2003). In this subsection, we establish the properties of the LADE defined as a minimizer of (3.2) for the non-stationary semi-strong GARCH(1,1) model (2.1) with heavy-tailed errors in the sense that the errors have infinite fourth moment. Denote $\mathcal{F}_t = \sigma(\varepsilon_s, s \leq t)$. We need the following assumptions.

A3. For some $\delta > 0$, there exists a $G_\delta < \infty$ such that $E(|\varepsilon_t|^{2+\delta}|\mathcal{F}_{t-1}) \leq G_\delta < \infty$ a.s.

A4. Conditional on $\mathcal{F}_{t-1}$, $\log(\varepsilon_t^2)$ has zero median and a differentiable density function $f_t(x)$ satisfying $f_t(0) \equiv f(0) > 0$, and $\sup_{x\in R, t\geq 1} |f_t'(x)| < B_1 < \infty$.

A5. $u \to \infty$ and $u/n \to 0$, as $n \to \infty$.

Remark 3. The class of adapted sequences with bounded conditional moments is quite wide and includes, for instance, the classes of randomly stopped sequences and martingale transforms (e.g., Remark 3.3 in de la Pena et al.(2003)). Interestingly, moment inequalities for nonnegative adapted sequences and martingales with bounded conditional moments have the same form as under independence (see also the discussion in Sections B.3 and B.4 in Nze and Doukhan, (2004)). This is the essence of why the results of Jensen and Rahbek (2004 a, 2004 b) still hold for non-stationary semi-strong GARCH (1, 1) models when the conditional second and fourth moments of the error ε_t equal its unconditional second and fourth moments a.s. respectively.

Theorem 2. *Suppose that $\gamma \geq 0$ and Assumptions A1-A4 hold.*

(i) Denote $S_n(\phi)|_{\phi=(\theta^\top,\psi_0^\top)^\top}$ by $S_n(\theta)$ with $u = 0$. Then there exists a local minimizer $\hat{\theta} = (\hat{\alpha}, \hat{\beta})^\top$ of $S_n(\theta)$ such that

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{\text{d}} N\big(0, \frac{1}{4f^2(0)}\Omega^{-1}\big),$$

where $\Omega = E(D_t D_t^\top)$ *and* $S_n(\phi)$ *is defined in (3.2),*

(ii) Let ψ_* *be any fixed value of* ψ *and denote* $S_n(\phi)|_{\phi=(\theta^\top,\psi_*^\top)^\top}$ *by* $S_{n*}(\theta)$*. Assume, in addition,* $\gamma > 0$ *and Assumption A5 hold. Then there exists a local minimizer* $\hat{\theta}_* = (\hat{\alpha}_*, \hat{\beta}_*)^\top$ *of* $S_{n*}(\theta)$ *such that*

$$\sqrt{n}(\hat{\theta}_* - \theta_0) \xrightarrow{\mathrm{d}} N\big(0, \frac{1}{4f^2(0)}\Omega^{-1}\big),$$

where Ω *and* $S_n(\phi)$ *are the same as in (i) of this theorem.*

Remark 4. Since Assumption A4 assumes $\log(\varepsilon_t^2)$ has zero median conditional on $\mathcal{F}_{t-1}$, we have $P(\varepsilon_t^2 > 1|\mathcal{F}_{t-1}) = 1/2$ implying $P(\varepsilon_t^2 \leq 1/2|\mathcal{F}_{t-1}) < 1$, which and Assumption A3 together ensure the validity of Lemma 3. If $\{\varepsilon_t\}$ are i.i.d., Assumption A3 is redundant, and in this case, the moment condition for ε_t in Theorem 2 can be reduced to $E|\varepsilon_t|^\varrho < \infty$ for some $\varrho > 0$.

Remark 5. The result of (ii) implies that (α, β) can be estimated by taking any value of ψ. One may estimate ψ, but the asymptotic properties of the estimated ψ have not been obtained.

3.1.2 Wald test for linear hypotheses

In this subsection, we use the same notation as in Subsection 3.1.1. We can use the result of Theorem 2 to do some inference for a subset of the parameters of the model (2.1). For example, we may consider a general form of linear null hypothesis

$$H_0 : \Gamma\theta_0 = \Lambda,$$

where Γ is a $s \times 2$ constant matrix with rank $s \leq 2$ and Λ is a $s \times 1$ constant vector. A Wald test statistic may be defined as

$$P_n = 4\hat{f}_*^2(0)(\Gamma\hat{\theta}_* - \Lambda)^\top \big(\Gamma\hat{\Omega}_*^{-1}\Gamma^\top\big)^{-1}(\Gamma\hat{\theta}_* - \Lambda)$$

and we reject H_0 for large values of P_n. In the above expression

$$\hat{f}_*(0) = \frac{1}{nbw_n}\sum_{t=1}^{n} K\left(\frac{\log\hat{\varepsilon}_{t*}^2}{bw_n}\right), \quad \hat{\varepsilon}_{t*}^2 = \frac{X_t^2}{\sigma_t^2(\hat{\phi}_*)}, \quad \textit{and} \quad \hat{\Omega}_* = \frac{1}{n}\sum_{t=1}^{n}\big[A_t(\hat{\phi}_*)A_t^\top(\hat{\phi}_*)\big],$$

where $\hat{\phi}_* = (\hat{\theta}_*^\top, \psi_*^\top)^\top$, $\psi_* = (\omega_*, \eta_*)^\top$ is some fixed value of ψ, $K(\cdot)$ is a kernel function on R and $bw_n > 0$ is a bandwidth. By Theorem 2 and using the same method of Theorem 3 in

Pan et. al. (2007), we can obtain that $\hat{f}_*(0)$ and $\hat{\Omega}_*$ are consistent estimators for $f(0)$ and Ω respectively. Thus, we have the following theorem.

Theorem 3. *Suppose that the conditions of Theorem 2 hold. Moreover, we assume that the kernel function $K(\cdot)$ is bounded, Lipschitz continuous and of finite first moment. Let $b_n \to 0$ and $nb_n^4 \to \infty$, as $n \to \infty$. Then it follows that $P_n \xrightarrow{\mathrm{d}} \chi_s^2$.*

3.2 The Gaussian QMLE

3.2.1 Asymptotic properties of the QMLE

In this subsection, we give the asymptotic behavior of QMLE defined as a minimizer of (3.3) for a non-stationary semi-strong GARCH(1,1) model when the distributions of the errors are in the domain of attraction of a stable law with the exponent $\kappa \in (1,2)$. Jensen and Rahbek (2004b) have established the consistency and asymptotic normality of the QMLE for model (2.1) with i.i.d. errors under the conditions $\gamma \geq 0$ and $E\varepsilon_t^4 < \infty$. The asymptotic properties of the QMLE in Jensen and Rahbek (2004b) still hold for the non-stationary semi-strong GARCH(1,1) model if we assume in addition that $E(\varepsilon_t|\mathcal{F}_{t-1}) = 0$, $E(\varepsilon_t^2|\mathcal{F}_{t-1}) = 1$, and $E(\varepsilon_t^4|\mathcal{F}_{t-1}) = E\varepsilon_t^4$ a.s. However, we will show that the limiting distribution of QMLE for non-stationary semi-strong GARCH(1,1) model is non-Gaussian but some stable law if the following assumption holds.

A6. $E(\varepsilon_t|\mathcal{F}_{t-1}) = 0$ a.s., $E(\varepsilon_t^2|\mathcal{F}_{t-1}) = 1$ a.s., and the distribution of ε_t^2 is in the domain of attraction of a stable law with the exponent $\kappa \in (1,2)$. Moreover, there exists a positive random variable Y with distribution function F_Y such that

$$\sup_{t\geq 1} Pr\big(|\varepsilon_t^2 - 1| > x|\mathcal{F}_{t-1}\big) \leq 1 - F_Y(x), \quad a.s \tag{3.6}$$

for sufficiently large x, where $1 - F_Y(x) \sim x^{-\kappa}L_Y(x)$ as $x \to \infty$ and $L_Y(x)$ is a slowly varying function which means that $\lim_{x\to\infty} \frac{L_Y(\lambda x)}{L_Y(x)} \to 1$ for any $\lambda > 0$.

A7. $\liminf_{y\to+\infty} U_\varepsilon(y)/U_Y(y) = 2\lambda_0 > 0$, where $U_Y(y) = \big(\frac{1}{1-F_Y}\big)^{\leftarrow}(y) = \inf\{x : \frac{1}{1-F_Y(x)} \geq y\}$, $U_\varepsilon(y) = \big(\frac{1}{1-F_\varepsilon}\big)^{\leftarrow}(y) = \inf\{x : \frac{1}{1-F_\varepsilon(x)} \geq y\}$ and $F_\varepsilon(x)$ is the distribution function of $|\varepsilon_t^2 - 1|$.

A8. $|\varepsilon_t^2 - 1|$ is strongly mixing with geometric rate, namely,

$$\alpha(k) = \sup_{\substack{A\in\sigma\left((\varepsilon_t^2-1)D_t;t>k\right)\\ B\in\sigma\left((\varepsilon_t^2-1)D_t;t\le 0\right)}} |Pr(A\cap B) - Pr(A)Pr(B))| \le C\iota^k \to 0, \quad k\to\infty,$$

where $0 < \iota < 1$ and C are constants. Furthermore, we assume $(\varepsilon_t^2 - 1)D_t$ is strongly mixing with geometric rate too.

Remark 6. Denote $RV_\rho = \{H : \lim_{y\to\infty} H(yx)/H(y) = x^\rho, \quad \text{for any} \quad x > 0\}$. Under Assumption A6, $1-F_\varepsilon(x) \in RV_{-\kappa}$ and $1-F_Y(x) \in RV_{-\kappa}$, and then we know that $U_\varepsilon(y) \in RV_{1/\kappa}$ and $U_Y(y) \in RV_{1/\kappa}$ by the theory of regular variation, see Resnick (1987). Thus, $U_\varepsilon(y) = y^{1/\kappa}Q_\varepsilon(y)$ and $U_Y(y) = y^{1/\kappa}Q_Y(y)$, where $Q_\varepsilon(x)$ and $Q_Y(x)$ are both slowly varying functions. Here we give an example of a class of slowly varying functions ensuring Assumption A7 hold. Let $\aleph = \{Q(y) : Q(y) = a(1 + by^{-\xi}), \xi > 0, a > 0, b > 0\}$. It is easy to verify that for any $Q(y) \in \aleph$ and $\tilde{Q}(y) \in \aleph$, $Q(y)$ and $\tilde{Q}(y)$ are both slowly varying functions satisfying $\lim_{y\to+\infty} Q(y)/\tilde{Q}(y) = C > 0$.

Denote $l_n(\phi)|_{\phi=(\theta^\top,\psi_0^\top)^\top}$ by $l_n(\theta)$ and $l_n(\phi)|_{\phi=(\theta^\top,\psi_*^\top)^\top}$ by $l_{n*}(\theta)$, where $l_n(\phi)$ is defined in (3.3), and $\psi_* = (\omega_*, \eta_*)^\top$ is some fixed value of ψ.

Theorem 4. *Suppose $\gamma \ge 0$ and Assumptions A2 and A6-A8 hold. Assume that ε_t has a Lebesgue density $g(x)$ and the origin lies in the closure of the interior of $\{g > 0\}$. Then it follows that*

(i)There exists a fixed open neighborhood $U(\theta_0)$ of θ_0 such that $l_n(\theta)$ has a unique minimizer $\tilde{\theta}$ in $U(\theta_0)$ with probability tending to one as $n \to \infty$. Furthermore, $\tilde{\theta}$ is consistent and

$$na_n^{-1}(\tilde{\theta} - \theta_0) \xrightarrow{\mathrm{d}} W_\kappa,$$

where $a_n = \inf\{x : P(\varepsilon_t^2 > x) \le 1/n\}$ and W_κ is a non-degenerate κ-stable random vector.

(ii)If $\gamma > 0$ holds, the results in (i) hold for $l_{n*}(\theta)$.

Remark 7. First, Assumption A6 implies Assumption A3 and $P(\varepsilon_t^2 \le 1/2|\mathcal{F}_{t-1}) < 1$, thus the results of Lemma 3 hold. Second, in the case when $\{\varepsilon_t\}$ are i.i.d., it is obvious that

Assumption A7 and the latter part of Assumption A6 hold, and furthermore, we can prove that Assumption A8 holds. In fact, by the definition of D_t, we have

$$D_t = \frac{\beta_0}{\alpha_0\varepsilon_{t-1}^2 + \beta_0} D_{t-1} + \big(\frac{\varepsilon_{t-1}^2}{\alpha_0\varepsilon_{t-1}^2 + \beta_0}, \quad \frac{1}{\alpha_0\varepsilon_{t-1}^2 + \beta_0}\big)^\top. \tag{3.7}$$

Thus, it follows that

$$\sigma\big((\varepsilon_t^2 - 1)D_t; t > k\big) \subseteq \sigma\big(D_{t+1}; t > k-1\big) \quad and \quad \sigma\big((\varepsilon_t^2 - 1)D_t; t \le 0\big) \subseteq \sigma\big(D_{t+1}; t \le 0\big).$$

Therefore,

$$\sup_{\substack{A \in \sigma\left((\varepsilon_t^2-1)D_t; t>k\right) \\ B \in \sigma\left((\varepsilon_t^2-1)D_t; t\le 0\right)}} |Pr(A \cap B) - Pr(A)Pr(B))| \le \sup_{\substack{A \in \sigma\left(D_{t+1}; t>k-1\right) \\ B \in \sigma\left(D_{t+1}; t\le 0\right)}} |Pr(A \cap B) - Pr(A)Pr(B)|.$$

But, (3.7), Assumption A6 and the conditions that ε_t has a Lebesgue density $g(x)$ and the origin lies in the closure of the interior of $\{g > 0\}$ ensure that D_t satisfies the assumptions in Theorem 7.4.1 of Straumann (2005), which implies that D_t is strongly mixing with geometric rate.

3.2.2 Bootstrap methods

Note that from Theorem 4 the scale na_n^{-1} depends intimately on the particular law in whose domain of the distribution ε_t^2 lies. In fact, the scale depends on the unknown tail exponent κ. Since the law is unknown, it is awkward to determine the scale empirically. In the following, we use a similar method to that in Hall and Yao (2003) to demonstrate how to apply the result of Theorem 4 in practice. In this subsection, we use the same notation as in Section 3.2.1 and assume that the errors are i.i.d. Define

$$\hat{\tau}^2 = \frac{1}{n}\sum_{t=1}^{n} \varepsilon_t^4 - \big(\frac{1}{n}\sum_{t=1}^{n} \varepsilon_t^2\big)^2.$$

Using the same method of Theorem 3.1 in Hall and Yao (2003), we can obtain that

$$a_n^{-1}\big(n(\tilde{\theta}_* - \theta_0)^\top, n^{1/2}\hat{\tau}\big)^\top \xrightarrow{\text{d}} ((W_\kappa^{(1)})^\top, W_\kappa^{(2)})^\top, \tag{3.8}$$

where $((W_\kappa^{(1)})^\top, W_\kappa^{(2)})^\top$ is a κ-stable vector with dimension 3 and $\tilde{\theta}_* = (\tilde{\alpha}_*, \tilde{\beta}_*)^\top$ is a minimizer of $l_{n*}(\theta)$. Obviously, (3.8) means that

$$n^{1/2}\frac{\tilde{\theta}_* - \theta_0}{\hat{\tau}} \xrightarrow{\text{d}} \frac{W_\kappa^{(1)}}{W_\kappa^{(2)}} \tag{3.9}$$

Due to (3.9), we can use the subsample bootstrap to approximate the distribution of $\tilde{\theta}_* - \theta_0$, but we must take account of the fact that the errors $\{\varepsilon_t\}$ are unknown. Suppose we observe a sample $\mathcal{X} = \{X_1, \cdots, X_n\}$ from the model (2.1), a natural approach is to use the standardized residuals computed by $\tilde{\varepsilon}_t = X_t/\tilde{\sigma}_t$, where $\tilde{\sigma}_t = \sigma_t(\tilde{\theta}_*, \psi_*), 1 \le t \le n$. Define

$$\tilde{\tau}^2 = \frac{1}{n}\sum_{t=1}^{n} \tilde{\varepsilon}_t^4 - \big(\frac{1}{n}\sum_{t=1}^{n} \tilde{\varepsilon}_t^2\big)^2$$

Then, by the same way as in Hall and Yao (2003), we have

$$a_n^{-1}\big(n(\tilde{\theta}_* - \theta_0)^\top, n^{1/2}\tilde{\tau}\big)^\top \xrightarrow{\ \mathrm{d}\ } ((W_\kappa^{(1)})^\top, W_\kappa^{(2)})^\top, \tag{3.10}$$

where $W_\kappa^{(1)}$ and $W_\kappa^{(2)}$ are the same as in (3.8). Result (3.10) demonstrates that the replacement of $\tilde{\varepsilon}_t$ for ε_t comes at no cost.

Since we require that ε_t has mean 0 and variance 1, in practice we standardize $\tilde{\varepsilon}_t$ as follows

$$\hat{\varepsilon}_t = \frac{\tilde{\varepsilon}_t - n^{-1}\sum_{j=1}^n \tilde{\varepsilon}_j}{\big(n^{-1}\sum_{j=1}^n \tilde{\varepsilon}_j^2 - (n^{-1}\sum_{j=1}^n \tilde{\varepsilon}_j)^2\big)^{1/2}}.$$

Now we can construct confidence intervals using subsampling bootstrap. Suppose ε_t^*, for $0 < t < +\infty$ are drawn randomly from $\{\hat{\varepsilon}_t, t = 1, \ldots, n\}$. Consider the process (conditional on $\mathcal{X}$) defined by $X_t^* = \sigma_t^*\varepsilon_t^*$, where $(\sigma_0^*)^2 = \eta_*$ and

$$(\sigma_t^*)^2 = \omega^* + \tilde{\alpha}_*(X_{t-i}^*)^2 + \tilde{\beta}_*(\sigma_{t-j}^*)^2, \quad 0 < t < +\infty.$$

Since $\tilde{\theta}_*$ is a consistent estimator of θ_0, it follows that the probability, conditional on $\mathcal{X}$, of X_t^* being non-stationary converges to 1 as $n \to +\infty$. Let $m < n$, and compute the QMLE $\tilde{\theta}_*^*$ of θ_0 using the data set $\mathcal{X}^* = \{X_1^*, \cdots, X_m^*\}$, namely, $\tilde{\theta}_*^* = (\tilde{\alpha}_*^*, \tilde{\beta}_*^*)^\top$ is a maximizer of the quasi-maximum likelihood function based on $\mathcal{X}^*$. Define $\tilde{\varepsilon}_t^* = X_t^*/\tilde{\sigma}_t^*$, where $(\tilde{\sigma}_t^*)^2 = \omega_* + \tilde{\alpha}_*^*(X_{t-1}^*)^2 + \tilde{\beta}_*^*(\tilde{\sigma}_{t-1}^*)^2, 1 \le t \le m$. Let

$$(\tilde{\tau}^*)^2 = \frac{1}{m}\sum_{t=1}^{m}(\tilde{\varepsilon}_t^*)^4 - \big(\frac{1}{m}\sum_{t=1}^{m}(\tilde{\varepsilon}_t^*)^2\big)^2,$$

be the bootstrap versions of $\hat{\tau}^2$. If $m/n \to 0$, as in Hall and Yao (2003), it follows that

$$\begin{aligned} &Pr\Big\{a_m^{-1}\big[m(\tilde{\theta}_*^* - \tilde{\theta}_*)^\top, m^{1/2}\tilde{\tau}^*\big] \in V \times [y_1, y_2] | \mathcal{X}\Big\} \\ &\longrightarrow Pr\big\{\big((W_\kappa^{(1)})^\top, W_\kappa^{(2)}\big) \in V \times [y_1, y_2]\big\}, \end{aligned} \tag{3.11}$$

in probability for each cylindrical set V of R^2 and all continuity points $0 < y_1 < y_2 < \infty$ of $W_\kappa^{(2)}$, where $W_\kappa^{(1)}$ and $W_\kappa^{(2)}$ are the same as in (3.8). Therefore, multivariate confidence regions for θ_0 can be developed. However, as Hall and Yao (2003) has pointed out, such regions can be difficult to interpret. Notice that a two sided interval may be obtained by taking the intersection of the two one-sided intervals, thus we shall consider only one-sided confidence interval for individual parameter component. Given $\pi \in (0, 1)$, let

$$\hat{I}_\pi^1 = \inf\big\{u : Pr[m^{1/2}(\tilde{\tau}^*)^{-1}(\tilde{\alpha}_*^* - \tilde{\alpha}_*) \le u|\mathcal{X}] \ge \pi\big\}.$$

and

$$\hat{I}_\pi^2 = \inf\big\{u : Pr[m^{1/2}(\tilde{\tau}^*)^{-1}(\tilde{\beta}_*^* - \tilde{\beta}_*) \le u|\mathcal{X}] \ge \pi\big\}.$$

By (3.11), we know both $[\tilde{\alpha}_* - n^{-1/2}\tilde{\tau}\hat{I}_\pi^1, +\infty)$ and $[\tilde{\beta}_* - n^{-1/2}\tilde{\tau}\hat{I}_\pi^2, +\infty)$ have nominal coverage π in the sense that $Pr\{\alpha_0 \in [\tilde{\alpha}_* - n^{-1/2}\tilde{\tau}\hat{I}_\pi^1, +\infty)\} \to \pi$ and $Pr\{\beta_0 \in [\tilde{\beta}_* - n^{-1/2}\tilde{\tau}\hat{I}_\pi^2, +\infty)\} \to \pi$.

4 Numerical Properties

This section presents some numerical evidence on the performance of asymptotic results of the proposed LADE and QMLE in finite samples through a simulation study. The data are generated from the non-stationary GARCH(1,1) model (2.1) with the true parameter $\phi_0 = (0.1, 1, 0.1, 0.5)^\top$. In all experiments, we use the sample size $n = 600$ with 1000 replications.

We first give some numerical comparisons between LADE and QMLE. Here we take $u = 10$ and consider four error distributions, $t(2)$, $t(3)$, $t(4)$, and $N(0, 1)$, where $t(i)$ stands for Student's t-distribution with degree of freedom i, $i = 2, 3, 4$. Notice that the variances are infinite for GARCH processes (2.1) with the true parameter $\phi_0 = (0.1, 1, 0.1, 0.5)^\top$ driven by all the error distributions considered above, including the normal case $N(0, 1)$. Figure 1 gives the boxplots of the average absolute error (AAE) $\big(|\hat{\alpha} - 0.1| + |\hat{\beta} - 1|\big)/2$ for both LADE and QMLE when ω and η are fixed at their true values, namely, $\omega = 0.1$ and $\eta = 0.5$. For heavy tailed errors, i.e., $t(2)$, $t(3)$, and $t(4)$, LADE outperforms QMLE. This is natural since LADE converges faster than QMLE in this case by Theorem 1 and Theorem 2. As we expected, MLE is better when the errors are normal. The boxplots of AAE for LADE and QMLE when ω and η take different values are presented in Figure 2. Figure 2 indicates that there is almost no influence on the estimation error of α and β when the values of ω and η vary; see Remark 5.

Then we investigate numerically the construction of confidence intervals for model (2.1) using bootstrap methods. For the sake of simplicity we only consider the case of the one sided intervals $[\tilde{\alpha}_* - n^{-1/2}\tilde{\tau}\hat{I}^1_\pi, +\infty)$ and $[\tilde{\beta}_* - n^{-1/2}\tilde{\tau}\hat{I}^2_\pi, +\infty)$. In this experiment, we take $\pi = 0.9$ with 1000 replications for bootstrap sampling and take $(\omega, \eta) = (0.1, 0.5), (0.3, 0.4), (0.1, 0), (0.2, 0.5)$ respectively. Three error distributions, $t(3)$, $t(4)$, and $t(5)$ are considered. To investigate the impact of subsampling size m, we take m = 150, 200, 250, 300, 350, 400, 450, 500, 550, and 600 respectively. Figure 3 presents the difference of the nominal level and the real level of the confidence intervals. Figure 3 indicates that the difference is very close to zero, and the variation of (ω, η) has little impact on the results. Although the method is quite robust against the selection of m, it seems that $m = 400$ is a good selection for almost all cases.

5 Conclusion

The contribution of this paper is to extend the domain of coverage of existing asymptotic theory to cover non-stationary and heavy tailed GARCH processes. We found that the LADE estimator is asymptotically normal even under our extremely demanding conditions, while the Gaussian QMLE requires stronger moment conditions and even then may have non-normal limiting distributions and slower rates of convergence. We provided explicit methods for conducting inference for both estimation methods.

Our results have some practical significance. Ibragimov (2004) argues that a number of economic and financial series can have very heavy tails. Although the tails of standardized residuals from estimated GARCH models are typically lighter than the tails of the raw series itself the residual series still has 'heavy tails' and in some cases the tail thickness may approach the region where our theory is relevant.

REFERENCES

Berkes, I., L. Horváth & P.Kokoszka (2003) GARCH processes: structure and estimation. *Bernoulli.* 9(2), 201-227.

Billingsley, P. (1968) *Convergence of Probability Measures.* New York: John Wiley & Sons.

Bollerslev, T. (1986) Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics.* 31, 307-327.

Bougerol, P. & N. Picard (1992) Stationarity of GARCH process and of some nonnegative time series. *Journal of Econometrics.* 52, 115-127.

Breiman, L. (1965) On some limit theorems similar to the arc-sin law. *Teor. Verojatnost. i Primenen.* 10, 351-360.

Brown, B. M. (1971) Martingale central limit theorems. *The Annals of Mathematical Statistics.* 42, 59-66.

Davis, R. A. & W. T. M. Dunsmuir (1997) Least absolute deviation estimation for regression with ARMA errors. *J. Theor. Prob.* 10, 481-497.

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

de la Pena, V. H., R. Ibragimov & S. Sharakhmetov (2003) On extremeal distributions and sharp L_p-bounds for sums of multilinear forms. *Annals of Probability* 31, 630-675.

Drost, F.C. & T.E. Nijman (1993) Temporal Aggregation of GARCH Processes. *Econometrica* 61, 909-927.

Embrechts, P., C. Klüppelberg & T. Mikosch. (1997) *Modelling Extremal Events for Insurance and Finance*, first edition. Berlin: Springer.

Engle, R. F. (1982) Autoregressive conditional heteroscedasticity with estimates of the variance of U.K. inflation. *Econometrica* 50, 987-1007.

Engle, R. F. & J.G. Rangel (2005) The Spline GARCH Model for Unconditional Volatility and its Global Macroeconomic Causes. NYU Stern School working paper.

Feller, W. (1971) *An Introduction to Probability Theory and Its Applications.* Vol II, John Wiley & Sons.

Hall, P. & Q. Yao (2003) Inference in ARCH and GARCH models with heavy-tailed errors. *Econometrica.* 71, 285-317.

Ibragimov, R. (2004) On the robustness of economic models to heavy-tailedness assumptions. Manuscript, Yale University.

Jensen, S. T. & A. Rahbek (2004 a) Asymptotic normality of the QMLE estimator of ARCH in the nonstatinary case. *Econometrica* 72, 641-646.

Jensen, S. T. & A. Rahbek (2004 b) Asymptotic inference for nonstationary GARCH. *Econometric Theory* 20, 1203-1226.

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

Leadbetter, M. R., G. Lindgren & H. Rootzén (1983) *Extremes and Related Properties of Random Sequences and Processes.* Berlin: Springer.

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

Mandelbrot, B. (1963) The variation of certain speculative prices. Journal of Business 36, 394-419.

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

Mikosch, T. & D. Straumann (2002) Whittle estimation in a heavy-tailed GARCH(1,1) model. *Stochastic Processes and Their Applications* 100, 187-222.

Mittnik, S. & S. T. Rachev (2000) *Stable Paretian Models in Finance.* New York: Wiley.

Mittnik, S., S. T. Rachev & M. S. Paolella (1998) Stable Paretian modelling in finance: some empirical and theoretical aspects. *In a Practical Guide to Heavy Tails, Ed. by R. J. Adler, R. E. Feldman and M. S. Taqqu, pp.79-110.* Boston: Birkhäuser.

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

Nze, P. A. & P. Doukhan (2004) Weak dependence: models and applications to econometrics. *Econometric Theory* 20, 995-1045.

Pan, J., H. Wang & H. Tong (2008) Estimation and tests for power-transformed and threshold GARCH models. *Journal of Econometrics* 142, 352-378.

Pan, J., H. Wang & Q. Yao (2007) Weighted least absolute deviations estimation for ARMA models with infinite variance. *Econometric Theory* 23, 852-879.

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

Polzehl, J. & V. Spokoiny (2004) Varying coefificent GARCH versus local constant volatility modeling. Comparison of the predictive power. Weierstrass Institute no 977.

Rvačeva, E. L. (1962) On domains of attraction of multi-dimensional distributions. *Select. Transl. Math. Statist. Probab.* American Mathematical Society, Providence, R.I. 2, 183-205.

Rydberg, T. (2000) Realistic statistical modelling of financial data. *Int. Statist. Rev.* 68, 233-258.

Shephard, N. (1996) Statistical aspects of ARCH and stochastic volatility. *In Time Series Models in Econometrics, Finance and Other Fields, Ed. by D.R. Cox, D.V. Hinkley and O.E. Barndorff-Nielsen, pp.1-67.* London: Chapman and Hall.

Straumann, D. (2005) *Estimation in Conditionally Heteroscedastic Time Series Models.* Lecture Notes in Statistics 181. Berlin Heidelberg: Springer.

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

Resnick, S. I. (1987) *Extreme Values, Regular Variation, Point Processes.* Springer, New York.

A Appendix

A.1 Proof of Theorem 1

Denote

$$y_i = \log(\alpha_0 \varepsilon_i^2 + \beta_0) \quad and \quad S_t = \sum_{i=0}^{t-1} y_i.$$

First of all, we introduce a lemma.

Lemma 1. *Suppose $\gamma = 0$ and the conditions of Theorem 1 hold. Then it follows that*

$$\limsup_{t \to +\infty} S_t = +\infty, \quad a.s.$$

Proof. Since y_i is a measurable function of ε_i, Assumption A2 and the definition of φ-mixing (see page 166 of Billingsley (1968)) ensure that $\{y_i\}$ is also φ-mixing and with $\sum_{n=1}^{\infty} \tilde{\varphi}_n^{1/2} < \infty$, where $\tilde{\varphi}_n$ is the φ-mixing coefficients of $\{y_i\}$. Notice that for $\varrho > 0$, there exists some constant C such that

$$\log \beta_0 < \log(\alpha_0 \varepsilon_i^2 + \beta_0) < C + (\alpha_0 \varepsilon_i^2 + \beta_0)^{\varrho/4}.$$

By Assumption A1, we obtain that $Ey_i^2 < +\infty$. Note that $E(Y_i) = \gamma = 0$. Applying the functional central limit theorem (see Theorem 20.1 of Billingsley (1968)), we have

$$\frac{1}{\sqrt{t}} S_t \xrightarrow{\mathrm{d}} N(0, \sigma^2), \tag{A.1}$$

where $\sigma^2 = Ey_0^2 + 2\sum_{i=1}^{\infty} E(y_0 y_i)$. Since $\{y_i\}$ is ergodic and $A = \{\omega : \limsup_{t \to +\infty} S_t(\omega) = +\infty\}$ is an invariant set, we obtain that $Pr(A) = 0$ or $Pr(A) = 1$. Notice that

$$\begin{aligned}
Pr(A) &= Pr\big\{ \cap_{m=1}^{+\infty} \cup_{t=m}^{+\infty} \{S_t \geq m\} \big\} \\
&= \lim_{m \to +\infty} Pr\big\{ \cup_{t=m}^{+\infty} \{S_t \geq m\} \big\} \\
&\geq \limsup_{m \to +\infty} Pr\big\{ S_{m^2} \geq m \big\} \\
&= 1 - \Phi(1/\sigma) > 0,
\end{aligned}$$

where the last equality is from (A.1) and $\Phi(\cdot)$ stands for the cumulative probability function of a standard normal random variable. Thus, $Pr(A) = 1$, which means that $\limsup_{t \to +\infty} S_t = +\infty$ a.s. $\square$

Proof of Theorem 1.

Sufficiency. Suppose $\gamma < 0$. By the ergodic theorem it follows that

$$\frac{1}{n}\sum_{i=1}^{n}\log(\alpha_0\varepsilon_{t-i}^2+\beta_0)\longrightarrow\gamma<0 \quad a.s,$$

as $n\to\infty$. Using the same argument of Theorem 2 of Nelson (1990), we obtain that semi-strong GARCH(1,1) model (2.1) defines a unique strictly stationary and ergodic solution.

Necessity. Now we suppose the semi-strong GARCH(1,1) model (2.1) has a strictly stationary and ergodic solution $\{X_t\}$. If $\gamma > 0$, we have

$$\frac{1}{t}S_t\longrightarrow\gamma>0, \quad a.s,$$

as $t\to+\infty$ by the ergodic theorem, which implies that

$$S_t\longrightarrow+\infty, \quad a.s, \tag{A.2}$$

as $t\to+\infty$. On the other hand, by reduction we have

$$\begin{aligned}\sigma_t^2 &= \sigma_0^2\prod_{i=1}^{t}(\alpha_0\varepsilon_{t-i}^2+\beta_0)+\omega\big[1+\sum_{k=1}^{t-1}\prod_{i=1}^{k}(\alpha_0\varepsilon_{t-i}^2+\beta_0)\big]\\ &\geq \sigma_0^2\prod_{i=1}^{t}(\alpha_0\varepsilon_{t-i}^2+\beta_0)\\ &= \sigma_0^2\prod_{i=0}^{t-1}(\alpha_0\varepsilon_{i}^2+\beta_0).\end{aligned}$$

Thus, $\log\sigma_t^2\geq\log\omega_0+S_t\longrightarrow+\infty$ a.s as $t\to+\infty$ by (A.2). However, this contradicts the assumption that X_t a strictly stationary and ergodic solution. So, we have that $\gamma\leq 0$. But, Lemma 1 implies that $\gamma\neq 0$. This completes the proof.

A.2 Proof of Theorem 2

Let $Z_t(\phi)=\log X_t^2-\log\sigma_t^2(\phi)$. Put

$$Z_t(\theta)=Z_t(\phi)|_{\phi=(\theta^\top,\psi_0^\top)^\top}, \quad Z_{t*}(\theta)=Z_t(\phi)|_{\phi=(\theta^\top,\psi_*^\top)^\top},$$

where $\theta = \theta_0 + \frac{1}{\sqrt{n}}v$, $v = (v_1, v_2)^\top \in R^2$. It is easy to verify $\hat{\theta} = \theta_0 + \frac{1}{\sqrt{n}}\hat{v}$ and $\hat{\theta}_* = \theta_0 + \frac{1}{\sqrt{n}}\hat{v}_*$, where $\hat{v}$ and $\hat{v}_*$ are the minimizer of $T_n(v)$ and $T_{n*}(v)$, respectively. Here

$$\begin{aligned} T_n(v) &= \sum_{t=u+1}^{n} \big(|Z_t(\theta_0 + \frac{1}{\sqrt{n}}v)| - |Z_t(\theta_0)|\big), \\ T_{n*}(v) &= \sum_{t=u+1}^{n} \big(|Z_{t*}(\theta_0 + \frac{1}{\sqrt{n}}v)| - |Z_{t*}(\theta_0)|\big). \end{aligned}$$

The proof of Theorem 2 needs the following lemmas.

Lemma 2. *Suppose Assumptions A1 and A3 hold. Define* $q_p(a,b) = E\{[a/(\alpha_0\varepsilon_t^2 + b)]^p|\mathcal{F}_{t-1}\}$. *If* $Pr(\varepsilon_t^2 \le \frac{1}{2}|\mathcal{F}_{t-1}) < 1$, *then for any* $p \ge 1$, *there exists a constant* ρ *such that* $q_p(\beta_0, \beta_0) \le \rho < 1$, *a.s. Furthermore, for any* $p \ge 1$, *there exist some constants* β_L, β_U, ρ_L *and* ρ_U *such that* $\beta_L < \beta_0 < \beta_U$, $q_p(\beta_U, \beta_0) \le \rho_U < 1$ *and* $q_p(\beta_0, \beta_L) \le \rho_L < 1$.

Proof. By Lemma 4 (1) of Lee and Hansen (1994), we have that

$$Pr(\varepsilon_t^2 \le \frac{1}{2}|\mathcal{F}_{t-1}) \le r, \quad a.s. \tag{A.3}$$

with $r = 1 - 1/\big[2^{(2+\delta)/\delta}G_\delta^{2/\delta}\big] \in (0,1)$. Denote the conditional distribution function of ε_t given $\mathcal{F}_{t-1}$ by F_t, then by (A.3) it follows that

$$\begin{aligned} q_p(a,b) &= \int_{\{x^2 \le \frac{1}{2}\}} (\frac{a}{\alpha_0 x^2 + b})^p dF_t + \int_{\{x^2 > \frac{1}{2}\}} (\frac{a}{\alpha_0 x^2 + b})^p dF_t \\ &\le (\frac{a}{b})^p Pr(\varepsilon_t^2 \le \frac{1}{2}|\mathcal{F}_{t-1}) + (\frac{a}{b + \alpha_0/2})^p Pr(\varepsilon_t^2 > \frac{1}{2}|\mathcal{F}_{t-1}) \\ &= (\frac{a}{b + \alpha_0/2})^p + \big[(\frac{a}{b})^p - (\frac{a}{b + \alpha_0/2})^p\big] Pr(\varepsilon_t^2 \le \frac{1}{2}|\mathcal{F}_{t-1}) \\ &\le \frac{a^p}{b^p} \cdot \frac{b^p + (b + \alpha_0/2)^p r - b^p r}{(b + \alpha_0/2)^p}. \end{aligned}$$

Therefore,

$$q_p(\beta_0, \beta_0) \le \rho < 1, \quad with \quad \rho = \frac{\beta_0^p + (\beta_0 + \alpha_0/2)^p r - \beta_0^p r}{(\beta_0 + \alpha_0/2)^p}.$$

Notice that the function $h(a) = \frac{a^p}{\beta_0^p} \cdot \frac{\beta_0^p + (\beta_0 + \alpha_0/2)^p r - \beta_0^p r}{(\beta_0 + \alpha_0/2)^p}$ is continuous and increasing with $h(\beta_0) = \rho < 1$. Then, there exists some $\beta_U > \beta_0$ such that $h(\beta_U) = \rho_U < 1$, which means $q_p(\beta_U, \beta_0) \le \rho_U < 1$. Similarly, we can prove that there exists some $\beta_L < \beta_0$ such that $q_p(\beta_0, \beta_L) \le \rho_L < 1$. □

Lemma 3. *Suppose Assumptions A1 - A3 hold. Then, it follows that* $E\|D_t\|^k < \infty$ *for any integer* $k > 0$*,* $A_{it} < D_{it}$ *and*

$$E(D_{it} - A_{it})^p \to 0, \ \frac{1}{n}\sum_{t=1}^{n}(D_{it} - A_{it}) \xrightarrow{L_p} 0, \ \text{and} \ \frac{1}{n}\sum_{t=1}^{n}(D_{it} - A_{it})^2 \xrightarrow{L_p} 0, \quad i = 1, 2, \qquad \text{(A.4)}$$

for all $p \geq 1$*. Furthermore, for any* $p > 0$*, there exist some constants* $\beta_L < \beta_0 < \beta_U$ *such that* $E\|D_t(\beta_0, \beta_L)\|^p < \infty$ *and* $E\|D_t(\beta_U, \beta_0)\|^p < \infty$*.*

Proof. Since Lemmas 3 and 4 of Jensen and Rahbek (2004 b) deals with the case for A_{2t}, we will prove that the results hold for A_{1t} in the following and the case for A_{2t} is similar. By Lemma 2, it follows that

$$q_p = E\big\{[\beta_0/(\alpha_0\varepsilon_t^2 + \beta_0)]^p|\mathcal{F}_{t-1}\big\} \leq \rho < 1, \qquad a.s.$$

for any $p \geq 1$, where ρ is a constant independent with t. Using Minkowski's inequality, we have

$$\begin{aligned}
\big[E(D_{1t})^p\big]^{1/p} &\leq \sum_{j=1}^{\infty}\Big\{E\big[\beta_0^{(j-1)p}\varepsilon_{t-j}^{2p}\prod_{k=1}^{j}\frac{1}{(\alpha_0\varepsilon_{t-k}^2 + \beta_0)^p}\big]\Big\}^{1/p} \\
&\leq \frac{1}{\alpha_0}\sum_{j=1}^{\infty}\Big\{E\big[\prod_{k=1}^{j-1}\big(\frac{\beta_0}{\alpha_0\varepsilon_{t-k}^2 + \beta_0}\big)^p\big]\Big\}^{1/p} \\
&= \frac{1}{\alpha_0}\sum_{j=1}^{\infty}\Big\{E\big[\prod_{k=2}^{j-1}\big(\frac{\beta_0}{\alpha_0\varepsilon_{t-k}^2 + \beta_0}\big)^p E\big(\big(\frac{\beta_0}{\alpha_0\varepsilon_{t-1}^2 + \beta_0}\big)^p|\mathcal{F}_{t-2}\big)\big]\Big\}^{1/p} \\
&\leq \frac{\rho^{1/p}}{\alpha_0}\sum_{j=1}^{\infty}\Big\{E\big[\prod_{k=2}^{j-1}\big(\frac{\beta_0}{\alpha_0\varepsilon_{t-k}^2 + \beta_0}\big)^p\big]\Big\}^{1/p} \\
&\leq \frac{1}{\alpha_0}\sum_{j=1}^{\infty}\rho^{(j-1)/p} < \infty.
\end{aligned}$$

By Lemma 2, we can obtain with the same method as above that there exist some constants $\beta_L < \beta_0 < \beta_U$ such that $E\|D_t(\beta_0, \beta_L)\|^p < \infty$ and $E\|D_t(\beta_U, \beta_0)\|^p < \infty$ for any $p > 0$. Next, we will establish (A.4). From (3.4), we have $A_{1t} = \sum_{j=1}^{t}\beta^{j-1}\varepsilon_{t-j}^2\prod_{k=1}^{j}\frac{\sigma_{t-k}^2}{\sigma_{t-k+1}^2}$. Notice that

$$\frac{\sigma_{t-j}^2}{\sigma_{t-j+1}^2} = \frac{\sigma_{t-j}^2}{(\alpha_0\varepsilon_{t-j}^2 + \beta_0)\sigma_{t-j}^2 + \omega_0} \leq \frac{1}{\alpha_0\varepsilon_{t-j}^2 + \beta_0},$$

we obtain $A_{1t} \leq D_{1t}$ holds. By theorem 1, for any fixed j,

$$\frac{\beta_0}{\alpha_0\varepsilon_{t-j}^2 + \beta_0} - \frac{\beta_0\sigma_{t-j}^2}{(\alpha_0\varepsilon_{t-j}^2 + \beta_0)\sigma_{t-j}^2 + \omega_0} \to 0 \quad a.s.,$$

which implies that

$$\frac{1}{\alpha_0} \geq \varepsilon_{t-j}^2 \prod_{k=1}^{j} \frac{\beta_0}{\alpha_0 \varepsilon_{t-k}^2 + \beta_0} - \frac{\beta_0^j X_{t-j}^2}{\sigma_t^2} \to 0 \quad a.s. \tag{A.5}$$

Therefore, L_1 convergence holds by dominated convergence in (A.5). Now, using the same statement as Lemma 4 of Jensen and Rahbek (2004 b), we can obtain (A.4) holds. □

Lemma 4. *Suppose that Assumptions A1-A4 hold. Then*

$$n^{-1/2} \sum_{t=1}^{n} v^\top A_t sgn(\log \varepsilon_t^2) \xrightarrow{\text{d}} N(0, v^\top \Omega v),$$

for any $v \in R^2$, where $sgn(x)$ stands for sign of x.

Proof. By Assumption A4, we obtain that $E\big(v^\top A_t sgn(\log \varepsilon_t^2) | \mathcal{F}_{t-1}\big) = 0$. Thus $\{v^\top A_t sgn(\log \varepsilon_t^2)\}$ is a martingale difference with respect to $\mathcal{F}_{t-1}$. First,

$$\begin{aligned} \frac{1}{n} \sum_{t=1}^{n} E\big([v^\top A_t sgn(\log \varepsilon_t^2)]^2 | \mathcal{F}_{t-1}\big) &= \frac{1}{n} \sum_{t=1}^{n} (v^\top A_t)^2 = \frac{1}{n} \sum_{t=1}^{n} (v^\top D_t)^2 + \frac{1}{n} \sum_{t=1}^{n} [(v^\top A_t)^2 - (v^\top D_t)^2] \\ &\xrightarrow{\text{P}} E((v^\top D_t)^2) = v^\top \Omega v, \end{aligned}$$

by Lemma 3 and the ergodic theorem. Next, we can verify the Linderberg condition. Notice that $A_{it} \leq D_{it}$, $i = 1, 2$ (see Lemma 3). Then, by dominated convergence theorem, for any $\tilde{\delta} > 0$, we have

$$\begin{aligned} & \frac{1}{n} \sum_{t=1}^{n} E\big[(v^\top A_t)^2 I(|v^\top A_t| \geq \sqrt{n}\tilde{\delta})\big] \\ \leq \;& \frac{1}{n} \|v\|^2 \sum_{t=1}^{n} E\big[\|D_t\|^2 I(\|D_t\| \geq \sqrt{n}\tilde{\delta}/\|v\|)\big] \\ = \;& \|v\|^2 E\big[\|D_t\|^2 I(\|D_t\| \geq \sqrt{n}\tilde{\delta}/\|v\|)\big] \to 0, \end{aligned}$$

as $n \to \infty$, since $E\|D_t\|^2 < \infty$. Now we can obtain the result by applying the central limit theorem for martingale differences in Brown (1971). □

Lemma 5. *Suppose that the conditions of Theorem 2 (ii) hold. Then it follows that*

$$T_n(v) - T_{n*}(v) \xrightarrow{\text{P}} 0,$$

uniformly on compact sets.

Proof. Notice that Lemma 2 of this paper ensures that Lemma 12, and Lemma 14 of Jensen and Rahbek (2004b) still hold. By the mean value theorem, we have

$$\begin{aligned}
&\sup_{\|v\|\le M} |Z_t(\theta_0 + \frac{1}{\sqrt{n}}v) - Z_{t*}(\theta_0 + \frac{1}{\sqrt{n}}v)| \\
= &\sup_{\|v\|\le M} |\frac{\partial Z_t(\tilde{\phi})}{\partial \psi'}(\psi - \psi_*)| \le C D_{2t}(\beta_0, \beta_L) r_t,
\end{aligned}$$

where $\tilde{\phi} = \lambda(\theta_0, \psi_*)^\top + (1-\lambda)(\theta_0, \psi_0)^\top$ for some $\lambda \in [0, 1]$, and $\beta_L > \beta_0$ satisfying $E[D_{2t}(\beta_0, \beta_L)]^p < \infty$ by Lemma 3 and $E(r_t)^p = r^t$ for some $0 < p < 1$ and $0 < r < 1$ by Jensen and Rahbek (2004b). Therefore, by Hölder's inequality,

$$\begin{aligned}
&E\big(\sup_{\|v\|\le M} \sum_{t=u+1}^{n} |Z_t(\theta_0 + \frac{1}{\sqrt{n}}v) - Z_{t*}(\theta_0 + \frac{1}{\sqrt{n}}v)|\big)^{p/2} \\
\le\ & C^{p/2} \sum_{t=u+1}^{n} \big[E(D_{2t}(\beta_0, \beta_L))^p\big]^{1/2} r^{t/2} \to 0
\end{aligned}$$

as $n \to \infty$. But

$$\sup_{\|v\|\le M} |T_n(v) - T_{n*}(v)| \le 2 \sup_{\|v\|\le M} \sum_{t=u+1}^{n} |Z_t(\theta_0 + \frac{1}{\sqrt{n}}v) - Z_{t*}(\theta_0 + \frac{1}{\sqrt{n}}v)|.$$

This completes the proof of this lemma. □

Proof of Theorem 2.

(i) Define

$$T_n^+(v) = \sum_{t=u+1}^{n} (|Z_t(\theta_0) - n^{-1/2} v^\top A_t| - |Z_t(\theta_0)|).$$

It holds that for $z \neq 0$,

$$|z - y| - |z| = -y sgn(z) + 2(y - z)\{I(0 < z < y) - I(y < z < 0)\}.$$

Noticing that $Z_t(\varphi_0) = \log \varepsilon_t^2$, we have

$$\begin{aligned}
T_n^+(v) &= -n^{-1/2} \sum_{t=1}^{n} v^\top A_t sgn(\log \varepsilon_t^2) \\
&\quad + 2 \sum_{t=1}^{n} (n^{-1/2} v^\top A_t - \log \varepsilon_t^2)[I(0 < \log \varepsilon_t^2 < n^{-1/2} v^\top A_t) - I(n^{-1/2} v^\top A_t < \log \varepsilon_t^2 < 0)] \\
&=: J_{1n} + J_{2n}.
\end{aligned}$$

By Lemma 4, we have $J_{1n} \xrightarrow{\mathrm{d}} v^\top \xi$, where $\xi \sim N(0, \Omega)$. Now turning to J_{2n}, let

$$B_{nt} = (n^{-1/2} v^\top A_t - \log \varepsilon_t^2) I(0 < \log \varepsilon_t^2 < n^{-1/2} v^\top A_t),$$

and

$$C_{nt} = (n^{-1/2} v^\top A_t - \log \varepsilon_t^2) I(n^{-1/2} v^\top A_t < \log \varepsilon_t^2 < 0).$$

Then

$$\begin{aligned}
\sum_{t=1}^n E B_{nt}^2 &= \sum_{t=1}^n E\big(I(v^\top A_t > 0) \int_0^{n^{-1/2} v^\top A_t} (n^{-1/2} v^\top A_t - x)^2 f_t(x) dx\big) \\
&\le \sum_{t=1}^n E[\int_0^{n^{-1/2} v^\top A_t} (n^{-1/2} v^\top A_t - x)^2 (f_t(x) - f(0)) dx \\
&+ \int_0^{n^{-1/2} v^\top A_t} (n^{-1/2} v^\top A_t - x)^2 f(0) dx] \\
&\le \sum_{t=1}^n E\big(B_1 n^{-2} (v^\top A_t)^4 + f(0) n^{-3/2} (v^\top A_t)^3\big) \\
&\le \frac{C}{n^{1/2}} E\big(\|D_t\|^3 + \|D_t\|^4\big).
\end{aligned}$$

Hence

$$\lim_{n \to \infty} \sum_{t=1}^n E B_{nt}^2 = 0. \tag{A.6}$$

Similarly, we can prove that

$$\lim_{n \to \infty} \sum_{t=1}^n E C_{nt}^2 = 0. \tag{A.7}$$

Next, we will establish that

$$\sum_{t=1}^n E\big[(B_{nt} - C_{nt}) | \mathcal{F}_{t-1}\big] \xrightarrow{\mathrm{P}} \frac{f(0)}{2} E(v^\top D_t)^2.$$

Put

$$\begin{aligned}
B_{1n} &= \sum_{t=1}^n I(v^\top A_t > 0) \int_0^{n^{-1/2} v^\top A_t} (n^{-1/2} v^\top A_t - x) f(0) dx \\
B_{2n} &= \sum_{t=1}^n I(v^\top A_t > 0) \int_0^{n^{-1/2} v^\top A_t} (n^{-1/2} v^\top A_t - x)(f_t(x) - f(0)) dx \\
C_{1n} &= \sum_{t=1}^n I(v^\top A_t \le 0) \int_{n^{-1/2} v^\top A_t}^0 (n^{-1/2} v^\top A_t - x) f(0) dx \\
C_{2n} &= \sum_{t=1}^n I(v^\top A_t \le 0) \int_{n^{-1/2} v^\top A_t}^0 (n^{-1/2} v^\top A_t - x)(f_t(x) - f(0)) dx
\end{aligned}$$

Using the same method as (A.6), we can show that

$$B_{2n} \xrightarrow{\mathrm{P}} 0 \quad \text{and} \quad C_{2n} \xrightarrow{\mathrm{P}} 0 \quad \text{as } n \to \infty \tag{A.8}$$

By Lemma 3, we have

$$\frac{1}{n}\sum_{t=1}^{n}(v^\top A_t)^2 \xrightarrow{\mathrm{P}} E(v^\top D_t)^2, \tag{A.9}$$

as $n \to +\infty$. Notice that

$$B_{1n} + C_{1n} = \frac{f(0)}{2n}\sum_{t=1}^{n}(v^\top A_t)^2,$$

then we obtain

$$\sum_{t=1}^{n} E\big[B_{nt}|\mathcal{F}_{t-1}\big] = B_{1n} + C_{1n} + B_{2n} + C_{2n} \xrightarrow{\mathrm{P}} \frac{f(0)}{2}E(v^\top D_t)^2$$

by (A.9) and (A.8). But, from (A.6) and (A.7), it follows that

$$\begin{aligned} Var\Big(\sum_{t=1}^{n}\big(B_{nt} - C_{nt} - E((B_{nt} - C_{nt})|\mathcal{F}_{t-1})\big)\Big) &= \sum_{t=1}^{n} Var\big(B_{nt} - C_{nt} - E((B_{nt} - C_{nt})|\mathcal{F}_{t-1})\big) \\ &\le \sum_{t=1}^{n} 2E(B_{nt}^2 + C_{nt}^2) \to 0. \end{aligned}$$

Therefore,

$$\sum_{t=1}^{n}(B_{nt} - C_{nt}) \xrightarrow{\mathrm{P}} \frac{f(0)}{2}E(v^\top D_t)^2,$$

which implies that

$$J_{2n} \xrightarrow{\mathrm{P}} f(0)v^\top \Omega v.$$

Let $T(v) = f(0)v^\top\Omega v + v^\top\xi$. Then the finite dimensional distributions of $T_n^+(v)$ converge to those of $T(v)$. But, since $T_n^+(v)$ has convex sample paths, this implies that the convergence is in fact on $C(R^2)$ (see the proof of Proposition 1 in Davis and Dunsmuir (1997)). Let $H_t(\theta) = \frac{\partial^2 Z_t(\theta)}{\partial\theta\theta'}$, and we have $\sup_{\theta\in U(\theta_0)}\|H_t(\theta)\| \le \xi_t$, where $U(\theta_0)$ is some fixed neighborhood of θ_0 and ξ_t is strictly stationary and ergodic with $E\|\xi_t\| < \infty$, see Jensen and Rahbek (2004b). Hence, the result of (i) holds by a similar proof to that of Theorem 1 of Pan et. al (2005). (ii) By Lemma 5, we have

$$T_{n*}(v) \xrightarrow{\mathrm{d}} T(v), \quad on \quad C(R^2).$$

By the same argument as in (i), we obtain the result.

A.3 Proof of Theorem 4

We need the following lemmas to prove Theorem 4.

Lemma 6. *Suppose that Assumptions A6 and A7 hold. Then*

$$a_n^{-1}\sum_{t=1}^{n}(\varepsilon_t^2-1)D_t - a_n^{-1}\sum_{t=1}^{n}(\varepsilon_t^2-1)A_t \xrightarrow{P} 0.$$

Proof. Define

$$\begin{aligned}
&H_Y(b)=E\big[Y^2I(Y\le b)\big],\ \ b_n=U_\varepsilon(n),\ \ c_n=U_Y(n),\\
&\Upsilon_{nt}=I(|\varepsilon_t^2-1|\le b_n),\ \ J_{nt}=1-\Upsilon_{nt},\ \ \tau_{nt}=E\big[(\varepsilon_t^2-1)\Upsilon_{nt}|\mathcal{F}_{t-1}\big],\\
&L_1=\sum_{t=1}^{n}(\varepsilon_t^2-1)A_t,\ \ L_2=\sum_{t=1}^{n}(\varepsilon_t^2-1)J_{nt}A_t,\\
&L_3=\sum_{t=1}^{n}\big[(\varepsilon_t^2-1)\Upsilon_{nt}-\tau_{nt}\big]A_t,\ \ L_4=\sum_{t=1}^{n}\tau_{nt}A_t,
\end{aligned}$$

where $U_\varepsilon(x)$ and $U_Y(x)$ are defined in Assumption A7. Replacing A_t by D_t, we define $\tilde{L}_i$ in the same way as the definition of L_i, $i=1,\cdots,4$. Note that (ε_t^2-1) is still in the domain of attraction of a κ-stable law, and $a_n=b_n+1$ for sufficiently large n. Thus,

$$a_n\sim b_n,\quad as\quad n\to\infty. \tag{A.10}$$

By Theorem 2 of Feller (1971, P283), it follows that

$$\lim_{b\to+\infty}\frac{bE\big[YI(Y>b)\big]}{H_Y(b)}=\frac{2-\kappa}{\kappa-1}\quad and\quad \lim_{b\to+\infty}\frac{b^2Pr(Y>b)}{H_Y(b)}=\frac{2-\kappa}{\kappa}. \tag{A.11}$$

Hence

$$\lim_{b\to\infty}\frac{E\big[YI(Y>b)\big]}{bPr(Y>b)}=\frac{\kappa}{\kappa-1}. \tag{A.12}$$

From the definition of c_n and Assumption A6, we obtain that, for any fixed $\lambda>0$

$$\lim_{n\to+\infty}nPr(Y>c_n)=1,\quad and\quad \lim_{b\to+\infty}\frac{Pr(Y>\lambda b)}{Pr(Y>b)}=\lambda^{-\kappa} \tag{A.13}$$

By (A.11), (A.12) and (A.13), we have that for any fixed $\lambda>0$

$$\lim_{n\to\infty}\frac{n}{c_n}E\big[YI(Y>\lambda c_n)\big]=\lambda^{1-\kappa}\frac{\kappa}{\kappa-1}\quad and\quad \lim_{n\to\infty}\frac{nH_Y(\lambda c_n)}{c_n^2}=\lambda^{2-\kappa}\frac{\kappa}{2-\kappa}. \tag{A.14}$$

Since Assumptions A6-A7 imply that $\lambda_0 \le b_n/c_n \le 1$ for sufficiently large n, (3.6) ensures that

$$
\begin{aligned}
& E\big(|\varepsilon_t^2-1|J_{nt}|\mathcal{F}_{t-1}\big) \\
= & \int_0^{+\infty} Pr\big(|\varepsilon_t^2-1|J_{nt} > y|\mathcal{F}_{t-1}\big)dy \\
= & \int_0^{b_n} Pr\big(|\varepsilon_t^2-1| > b_n|\mathcal{F}_{t-1}\big)dy + \int_{b_n}^{+\infty} Pr\big(|\varepsilon_t^2-1| > y|\mathcal{F}_{t-1}\big)dy \\
\le & \int_0^{b_n} Pr\big(Y > b_n\big)dy + \int_{b_n}^{+\infty} Pr\big(Y > y\big)dy \\
= & E\big(YI(Y > b_n)\big) \le E\big[YI(Y > \lambda_0 c_n)\big]
\end{aligned}
$$

for sufficiently large n. Therefore, it follows from Lemma 3, (A.14) and (A.15) that

$$
\begin{aligned}
\frac{E|L_2-\tilde{L}_2|}{b_n} & \le \frac{1}{\lambda_0 c_n}\sum_{t=1}^{n} E\big[(D_t-A_t)E(|\varepsilon_t^2-1|J_{nt}|\mathcal{F}_{t-1})\big] \\
& \le \frac{n}{\lambda_0 c_n}E\big[YI(Y>\lambda_0 c_n)\big]\frac{1}{n}\sum_{t=1}^{n}E(D_t-A_t) \to 0.
\end{aligned}
$$

This implies that

$$
\frac{L_2-\tilde{L}_2}{b_n} \xrightarrow{P} 0. \tag{A.15}
$$

For L_3, we have, from (3.6), that

$$
\begin{aligned}
& E\Big\{\big[(\varepsilon_t^2-1)\Upsilon_{nt}-\tau_{nt}\big]^2|\mathcal{F}_{t-1})\Big\} \\
= & E\big[(\varepsilon_t^2-1)^2\Upsilon_{nt}-\tau_{nt}^2|\mathcal{F}_{t-1})\big] \\
\le & E\big[(\varepsilon_t^2-1)^2\Upsilon_{nt}|\mathcal{F}_{t-1})\big] \\
= & 2\int_0^{+\infty} yPr\big(|\varepsilon_t^2-1|\Upsilon_{nt} > y|\mathcal{F}_{t-1}\big)dy \\
= & 2\int_0^{b_n} yPr\big(y < |\varepsilon_t^2-1| \le b_n|\mathcal{F}_{t-1}\big)dy \\
= & 2\int_0^{b_n} yPr\big(|\varepsilon_t^2-1| > y|\mathcal{F}_{t-1}\big)dy - 2\int_0^{b_n} yPr\big(|\varepsilon_t^2-1| > b_n|\mathcal{F}_{t-1}\big)dy \\
\le & 2\int_0^{A} ydy + 2\int_A^{b_n} yPr\big(Y>y\big)dy \\
\le & A^2 + E\big[Y^2I(Y\le b_n)\big] + 2\int_0^{b_n} yPr\big(Y>b_n\big)dy \\
\le & A^2 + H(c_n) + c_n^2 Pr\big(Y>\lambda_0 c_n\big)
\end{aligned} \tag{A.16}
$$

for sufficiently large n. Notice that

$$E\Big\{\big[(\varepsilon_t^2-1)\Upsilon_{nt}-\tau_{nt}\big](A_{it}-D_{it})\big[(\varepsilon_s^2-1)\Upsilon_{ns}-\tau_{nt}\big](A_{is}-D_{is})\Big\}=0$$

for $t\neq s$, $i=1,2$. Then, it follows from Lemma 3, (A.13), (A.14) and (A.16) that

$$E\Big(\frac{L_3^{(i)}-\tilde{L}_3^{(i)}}{b_n}\Big)^2=\frac{1}{b_n^2}\sum_{t=1}^{n}E\Big\{(A_{it}-D_{it})^2E\big[[(\varepsilon_t^2-1)\Upsilon_{nt}-\tau_{nt}]^2|\mathcal{F}_{t-1}\big]\Big\}$$

$$\leq \frac{n}{\lambda_0^2c_n^2}[A^2+H(c_n)+c_n^2Pr\big(Y>\lambda_0c_n\big)]\frac{1}{n}\sum_{t=1}^{n}E(A_{it}-D_{it})^2\to 0,$$

where $L_3^{(i)}$ and $\tilde{L}_3^{(i)}$ denote the ith element of L_3 and $\tilde{L}_3$ respectively, $i=1,2$. Therefore,

$$\frac{L_3-\tilde{L}_3}{b_n}\xrightarrow{P}0. \tag{A.17}$$

Finally, we will show that

$$\frac{L_4-\tilde{L}_4}{b_n}\xrightarrow{P}0. \tag{A.18}$$

Notice that $E\big[(\varepsilon_t^2-1)|\mathcal{F}_{t-1}\big]=0$. Then $\tau_n=-E\big((\varepsilon_t^2-1)J_{nt}\big)$. Hence, using the same argument as for (A.15), we declare that (A.18) holds. It is easily verified that

$$L_1=L_2+L_3+L_4 \quad and \quad \tilde{L}_1=\tilde{L}_2+\tilde{L}_3+\tilde{L}_4.$$

Combining (A.10), (A.15), (A.17) and (A.18), we complete the proof of this lemma. □

Lemma 7. *Suppose that the conditions of Theorem 4 hold. Then* $(\varepsilon_t^2-1)D_t$ *has an extremal index* $\Delta>0$.

Proof. Let us recall the definition of the extremal index (see Leadbetter et al.(1983)) first. We say that a stationary process $\{\xi_n\}$ has extrema index Δ if for each $\vartheta>0$, there exists $a_n(\vartheta)$ such that $n\big(1-F(a_n(\vartheta))\big)\to\vartheta$ and $P(max_{i=1,\cdots,n}\xi_i\leq a_n(\vartheta))\to e^{-\Delta\vartheta}$ as $n\to\infty$. Suppose by contradiction that $\Delta=0$. Let $Y_n=\max_{t\in\{1,\cdots,n\}}\{|\varepsilon_t^2-1|D_t\}$ and $\tilde{Y}_n$ be the partial maxima of the corresponding iid sequence $\{Q_t\}$, where Q_1 has the same distribution as $|\varepsilon_1^2-1|D_1$. Since $|\varepsilon_t^2-1|D_t$ is regularly varying with index κ by Remark 7 and Breiman (1965), we have

$$\liminf_{n\to\infty}P(\tilde{Y}_n\leq a_nx)=\exp\{-(\varsigma x)^{-\kappa}\}>0, \quad for\ all\ x>0,$$

where $\varsigma = \left[E\|D_t\|^\kappa\right]^{1/\kappa}$ (see Chapter 3 of Embrechts et al.(1997)). By Theorem 3.7.2 of Leadbetter et al. (1983), we obtain that

$$\lim_{n\to\infty} P(Y_n \le a_n x) = 1 \tag{A.19}$$

provided $\Delta = 0$. However, it holds that for any $x_i > 0$, $y_i > 0$, $i = 1, 2$,

$$\frac{x_1 + x_2}{y_1 + y_2} \ge \min\{\frac{x_1}{y_1}, \frac{x_2}{y_2}\}.$$

Hence,

$$\begin{aligned}
\|D_t\| &\ge \|A_t\| = \|\frac{\partial\sigma_t^2(\theta_0)}{\partial\theta}\frac{1}{\sigma_t^2}\| \ge \frac{1}{\sqrt{2}}\frac{1}{\sigma_t^2}[\frac{\partial\sigma_t^2(\theta_0)}{\partial\alpha} + \frac{\partial\sigma_t^2(\theta_0)}{\partial\beta}] \\
&= \frac{1}{\sqrt{2}}\frac{\sum_{j=1}^t \beta_0^{j-1}(\varepsilon_{t-j}^2 + 1)\sigma_{t-j}^2}{\omega_0\sum_{j=1}^t \beta_0^{j-1} + \alpha_0\sum_{j=1}^t \beta_0^{j-1}\varepsilon_{t-j}^2\sigma_{t-j}^2 + \beta_0^t\sigma_0^2} \\
&= \frac{1}{\sqrt{2}}\frac{\sum_{j=1}^{t-1} \beta_0^{j-1}(\varepsilon_{t-j}^2 + 1)\sigma_{t-j}^2 + \beta_0^{t-1}(\varepsilon_0^2 + 1)\sigma_0^2}{\sum_{j=1}^{t-1} \beta_0^{j-1}(\omega_0 + \alpha_0\varepsilon_{t-j}^2\sigma_{t-j}^2) + \beta_0^{t-1}(\omega_0 + \alpha_0\varepsilon_0^2\sigma_0^2 + \beta_0\sigma_0^2)} \\
&\ge \frac{1}{\sqrt{2}}\min\{\frac{\beta_0^{j-1}(\varepsilon_{t-j}^2 + 1)\sigma_{t-j}^2}{\beta_0^{j-1}(\omega_0 + \alpha_0\varepsilon_{t-j}^2\sigma_{t-j}^2)}; \quad \frac{\beta_0^{t-1}(\varepsilon_0^2 + 1)\sigma_0^2}{\beta_0^{t-1}(\omega_0 + \alpha_0\varepsilon_0^2\sigma_0^2 + \beta_0\sigma_0^2)}; \quad 1 \le j \le t-1\} \\
&\ge \frac{1}{\sqrt{2}}\min\{\frac{\sigma_{t-j}^2}{\omega_0}, \frac{1}{\alpha_0}; \frac{\sigma_0^2/2}{\omega_0}, \frac{1}{\alpha_0}, \frac{\sigma_0^2/2}{\beta_0\sigma_0^2}; 1 \le j \le t-1\} \\
&\ge \frac{1}{\sqrt{2}}\min\{1, \frac{1}{\alpha_0}, \frac{1}{2}, \frac{1}{2\beta_0}\} \\
&= c_0 > 0.
\end{aligned}$$

Note that, as $n \to \infty$,

$$P(|\varepsilon_t^2 - 1| > a_n) \sim P(\varepsilon_t^2 > a_n) \sim n^{-1}. \tag{A.20}$$

Let $F(\cdot)$ be the marginal distribution of $|\varepsilon_t^2 - 1|$. For any $x > 0$, we have

$$\begin{aligned}
n\big(1 - F(q_n(y))\big) &= nP\big(|\varepsilon_t^2 - 1| > q_n(y)\big) \\
&= \frac{nP\big(|\varepsilon_t^2 - 1| > q_n(y)\big)}{nP(|\varepsilon_t^2 - 1| > a_n)} \cdot nP(|\varepsilon_t^2 - 1| > a_n) \\
&\to y,
\end{aligned} \tag{A.21}$$

where $y = (c_0^{-1}x)^{-\kappa}$ and $q_n(y) = a_n y^{-1/\kappa} = c_0^{-1}a_n x$. The convergence above holds because $|\varepsilon_t^2 - 1|$ is regularly varying with index κ and (A.20). By Assumption (A8) and (A.21), for any fixed y defined in (A.21) and integer $k > 1$,

$$
\begin{aligned}
& n \sum_{j=2}^{[n/k]} P\big(|\varepsilon_1^2 - 1| > q_n(y), |\varepsilon_j^2 - 1| > q_n(y)\big) \\
= \;& n \sum_{j=2}^{[n/k]} \Big[P\big(|\varepsilon_1^2 - 1| > q_n(y), |\varepsilon_j^2 - 1| > q_n(y)\big) - P\big(|\varepsilon_1^2 - 1| > q_n(y)\big)P\big(|\varepsilon_j^2 - 1| > q_n(y)\big) \\
& + P\big(|\varepsilon_1^2 - 1| > q_n(y)\big)P\big(|\varepsilon_j^2 - 1| > q_n(y)\big)\Big] \\
\leq \;& n \sum_{j=2}^{[n/k]} \alpha(j-1) + n[n/k]\big(1 - F(q_n(y))\big)^2 \\
\leq \;& o(1) + \frac{1}{k-1}\big(1 - F(q_n(y))\big)^2 \\
\to \;& \frac{y^2}{k-1}
\end{aligned}
$$

as $n \to \infty$, where [ ] denotes the integer part and $\alpha(j)$ is defined in Assumption (A8). Therefore,

$$
\lim_{k\to\infty} \limsup_{n\to\infty} \quad n \sum_{j=2}^{[n/k]} P\big(|\varepsilon_1^2 - 1| > q_n(y), |\varepsilon_j^2 - 1| > q_n(y)\big) = 0 \tag{A.22}
$$

By Assumption (A8), (A.21), (A.22) and Theorem 3.4.1 of Leadbetter et al. (1983), we obtain

$$
\lim_{n\to\infty} P\big(\max_{t\leq n} |\varepsilon_t^2 - 1| \leq c_0^{-1} a_n x\big) = \exp\{-(c_0^{-1}x)^{-\kappa}\}
$$

Thus,

$$
P(Y_n \leq a_n x) \leq P\big(\max_{t\leq n} |\varepsilon_t^2 - 1| \leq c_0^{-1} a_n x\big) \to \exp\{-(c_0^{-1}x)^{-\kappa}\} < 1, \quad n \to \infty
$$

which contradicts (A.19). Thus, $\Delta > 0$. $\square$

Proof of Theorem 4.

The proof for consistency in Theorem 1 and Theorem 2 of Jensen and Rahbek (2004 b) is still valid for the consistency of $\tilde{\theta}$. For the asymptotic normality, we follow the routine lines. According to Lemma 12 -14 of of Jensen and Rahbek (2004 b), it is sufficient to deal with $l_n(\theta) \equiv l_n(\theta, \psi_0)$. By Taylor expansion, we have

$$
\frac{\partial l_n(\theta_0)}{\partial \theta} = \frac{\partial l_n(\tilde{\theta})}{\partial \theta} + \frac{\partial^2 l_n(\tilde{\theta}^1)}{\partial \theta \partial \theta'}(\theta_0 - \tilde{\theta}),
$$

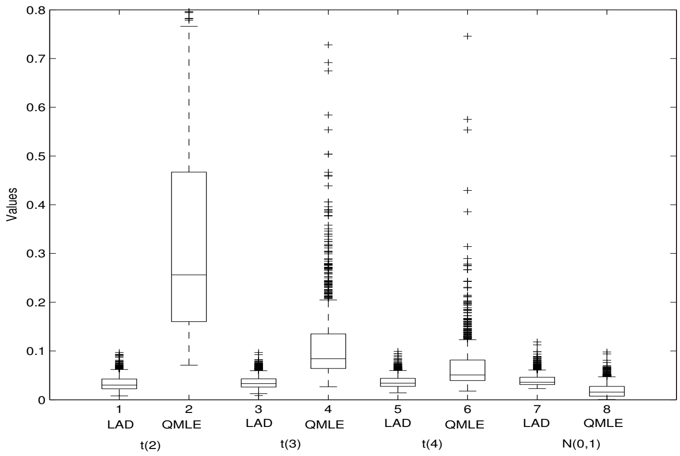

Figure 1: Boxplots of AAE for LADE and QMLE when ω and η are fixed at their true values for model (2.1).

where $\tilde{\theta}$ is the minimizer of $l_n(\theta)$ and θ^1 is on the line from $\tilde{\theta}$ to θ_0. Notice that $\frac{\partial l_n(\tilde{\theta})}{\partial \theta} = 0$ and $\frac{\partial^2 l_n(\tilde{\theta}^1)}{\partial \theta \partial \theta'} = \frac{\partial^2 l_n(\theta_0)}{\partial \theta \partial \theta'} + o_P(1)$ (see Jensen and Rahbek (2004 b)), and $\frac{\partial l_n(\theta_0)}{\partial \theta} = \frac{1}{n}\sum_{t=1}^n (\varepsilon_t^2 - 1)A_t$, we have

$$na_n^{-1}(\tilde{\theta} - \theta_0)(\frac{\partial^2 l_n(\theta_0)}{\partial \theta \partial \theta'} + o_P(1)) = a_n^{-1}\sum_{t=1}^n (\varepsilon_t^2 - 1)A_t.$$

By Lemma 6, it is enough to prove that

$$a_n^{-1}\sum_{t=1}^n (\varepsilon_t^2 - 1)D_t \xrightarrow{\text{d}} W_\kappa. \tag{A.23}$$

By Lemma 7 and the conditions of this theorem, the assumptions of Theorem 7.1.1 of Straumann (2005) hold. It follows that (A.23) holds. This completes the proof.

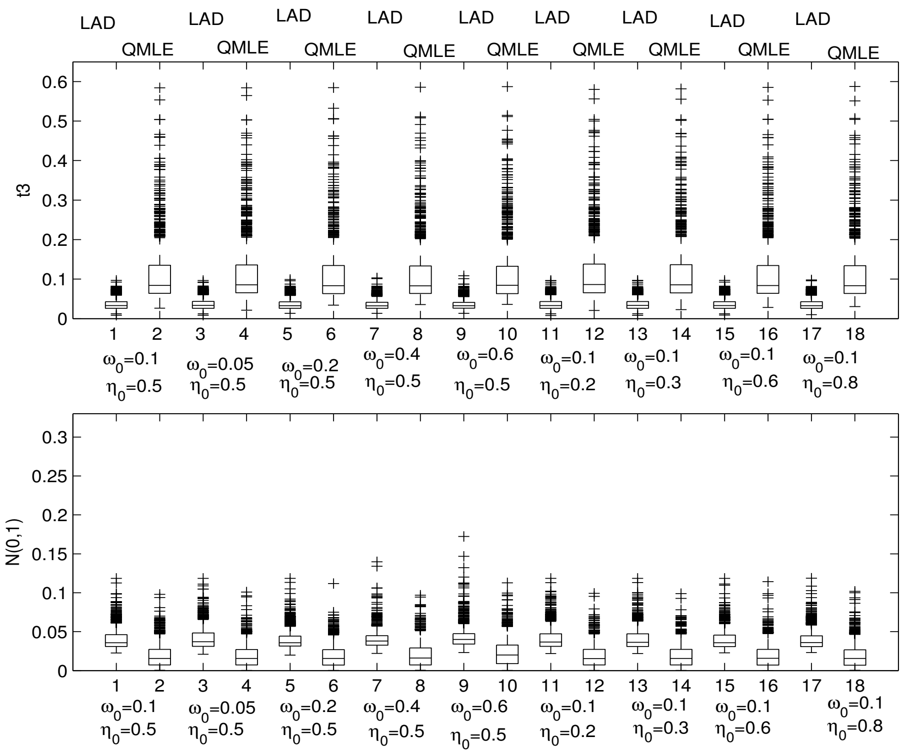

Figure 2: Boxplots of AAE for LADE and QMLE when ω and η take different values for model (2.1).

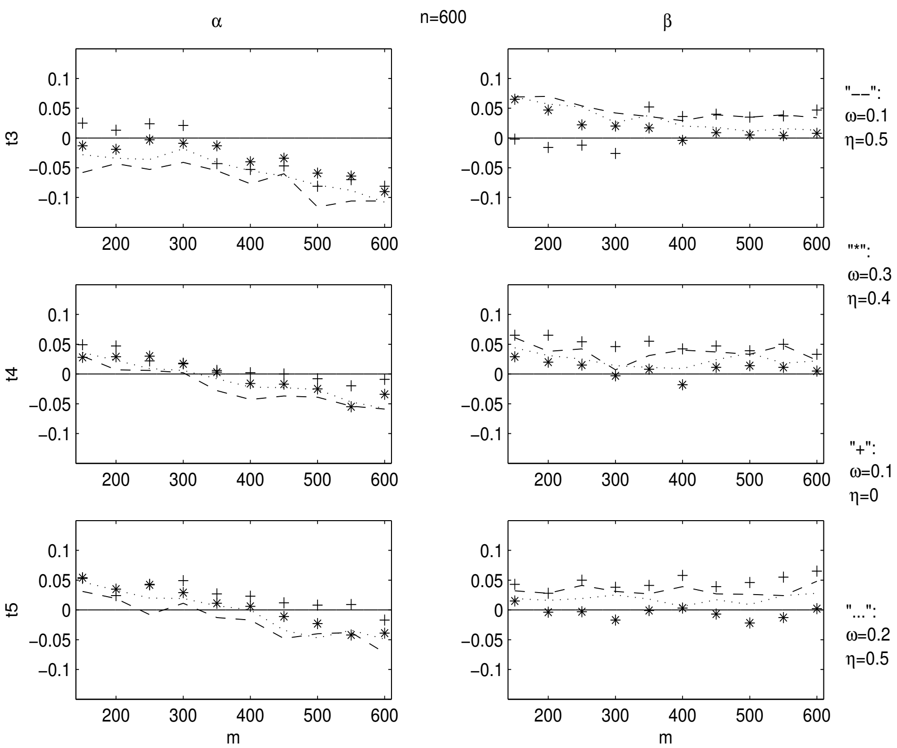

Figure 3: Differences between the nominal level and the real level of the confidence intervals when ω and η take different values for model (2.1).