Estimation and Tests for Power-Transformed and Threshold GARCH Models*

JIAZHU PAN[1,2†] HUI WANG[1] HOWELL TONG[2]

[1]*School of Mathematical Sciences, Peking University, Beijing 100871, China.*

[2]*Department of Statistics, London School of Economics, London, WC2A 2AE, UK*

Abstract. Consider a class of power transformed and threshold GARCH(p,q) (PTTGRACH(p,q)) model, which is a natural generalization of power-transformed and threshold GARCH(1,1) model in Hwang and Basawa (2004) and includes the standard GARCH model and many other models as special cases. We first establish the asymptotic normality for quasi-maximum likelihood estimators (QMLE) of the parameters under the condition that the error distribution has finite fourth moment. For the case of heavy-tailed errors, we propose a least absolute deviations estimation (LADE) for PTTGARCH(p,q) model, and prove that the LADE is asymptotically normally distributed under very weak moment conditions. This paves the way for a statistical inference based on asymptotic normality for heavy-tailed PTTGARCH(p,q) models. As a consequence, we can construct the Wald test for GARCH structure and discuss the order selection problem in heavy-tailed cases. Numerical results show that LADE is more accurate than QMLE for heavy tailed errors. Furthermore the theory is applied to the daily returns of the Hong Kong Hang Seng Index, which suggests that asymmetry and nonlinearity could be present in the financial time series and the PTTGARCH model is capable of capturing these characteristics. As for the probabilistic structure of PTTGARCH(p,q), we give in the appendix a necessary and sufficient condition for the existence of a strictly stationary solution of the model, the existence of the moments and the tail behavior of the strictly stationary solution.

JEL Classification: C4, G0.

Keywords: Threshold GARCH; power transformation; asymptotic normality; quasi-maximum likelihood estimator; least absolute deviations estimation; Wald test; order selection; PTTGARCH structure.

---

*Partially supported by an EPSRC research grant and the Natural Science Foundation of China (grant: 10471005)

†Address correspondence to Jiazhu Pan, Department of Statistics, London School of Economics, Houghton Street, London, WC2A 2AE, UK; e-mail: j.pan@lse.ac.uk.

1 Introduction

The autoregressive conditional heteroscedastic (ARCH) model proposed by Engle (1982) has led to considerable interest in models in which the conditional variance (volatility) of the current observation, σ_t^2, is a function of the past observations. Engle's ARCH model formulated the conditional variance of the process as "linear" in squared past values. Bollerslev (1986) generalized ARCH model to allow the conditional variance to depend additionally on its past realizations. Since then many empirical and theoretical aspects of the ARCH/GARCH model have been developed. Shephard (1996) and Rydberg (2000) gave excellent surveys of ARCH/GARCH modelling for financial data. Weiss (1986) and Berkes et al. (2003) established consistency and asymptotic normality of maximum likelihood estimators for ARCH and GARCH model respectively. The former assumes that the errors have finite fourth moment and the latter requires a moment of errors slightly higher than the fourth. Hall and Yao (2003) showed that when the error is heavy tailed (without finite fourth moment), quasi-maximum likelihood estimators (QMLE) are not asymptotically normal and suffer from slow convergence rate and complex asymptotic distribution, which do not facilitate, among others, statistical tests and interval estimation in the standard manner; see Hall and Yao (2003), and Mikosh and Straumann (2006). Peng and Yao (2003) pointed out that a kind of least absolute deviations estimator (LADE) has asymptotic normality if the error distribution has finite second moment.

Many extensions and generalizations of the ARCH model have appeared (see Engle and Bollerslev (1986), Higgins and Bera (1992), Li and Li (1996), Hwang and Kim (2004), and Hwang and Basawa (2004)). Among all the extensions, the functional form for σ_t^2 is of great importance (See Higgins and Bera (1992)). Even Engle (1982) has acknowledged that "it is likely that other formulations of the variance may be more appropriate for the particular applications". Hsieh (1989) found that the GARCH models cannot fit some exchange rates satisfactorily; Scheinkman and LeBaron (1989) found evidence that volatility in stock market data cannot be captured completely by linear ARCH models; Gouriéroux (1997, page 90) indicated that the heteroscedasticity varies depending on whether the error is positive or negative. This leads to asymmetric threshold ARCH modelling. The study of Li and Li (1996) has showed that threshold-asymmetric modelling provides better fitting compared with symmetric ARCH in the field of financial time series. Therefore, combining the above ideas, Hwang and Kim (2004) proposed a broad class of power transformed and threshold ARCH model:

$$X_t = \sigma_t \varepsilon_t \quad and \quad \sigma_t^{2\delta} = \alpha_0 + \sum_{i=1}^{p} \alpha_{1i} (X_{t-i}^{+})^{2\delta} + \sum_{i=1}^{p} \alpha_{2i} (X_{t-i}^{-})^{2\delta} \tag{1.1}$$

where $\delta > 0, \alpha_0 > 0, \alpha_{1i} \geq 0, \alpha_{2i} \geq 0$ are unknown parameters, $i = 1, \cdots, p$. Here, $\{\varepsilon_t\}$ is a sequence of independent and identically distributed random variables, and ε_t is independent of $\{X_{t-k}, k \geq 1\}$ for all t. They studied the geometric ergodicity and existence of moments of the model, and investigated a large sample test for ARCH structures based on the uniform local asymptotic normality approach. However, the σ_t in Hwang and Kim (2004)'s model is only a function of the past p observations. Hwang and Basawa (2004) introduced a Box-Cox transformed threshold GARCH(1,1) model by allowing σ_t to depend on σ_{t-1} and studied the stationarity and moment structure of the model. Liu (2006) investigated the tail behavior of the Box-Cox transformed threshold GARCH(1,1) model.

We consider a more general power transformed and threshold GARCH model, in which σ_t is a function of not only the past p observations but also the past q values of σ_t itself. A power-transformed and threshold GARCH(p,q) model (PTTGARCH(p,q)) is defined as

$$X_t = \sigma_t \varepsilon_t \quad and \quad \sigma_t^{2\delta} = \alpha_0 + \sum_{i=1}^{p} \alpha_{1i} (X_{t-i}^{+})^{2\delta} + \sum_{i=1}^{p} \alpha_{2i} (X_{t-i}^{-})^{2\delta} + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^{2\delta} \tag{1.2}$$

where $\delta, \alpha_0, \alpha_{1i}, \alpha_{2i}, \varepsilon_t$ are the same as those in model (1.1), and $\beta_j \geq 0, j = 1, \cdots, q$. As we can see, besides the standard GARCH model (Bollerslev (1986)) i.e. $\delta = 1$ and $\alpha_{1i} = \alpha_{2i}, i = 1, \cdots, p$, model (1.2) includes diverse nonlinear and asymmetric models as special cases. For example, it becomes a Box-Cox transformed ARCH model (Higgns and Bera (1992)) when $\delta = 2, q = 0$, a TARCH model (Li and Li (1996)) when $\delta = 1/2, q = 0$, a power-transformed and threshold ARCH model (Hwang and Kim (2004)) when $q = 0$, a Box-Cox transformed threshold GARCH(1,1) model (Hwang and Basawa (2004)) when $p = q = 1$.

The main goal of this paper is to study the estimation and tests for model (1.2). We differ from Hwang and Kim (2004) and Hwang and Basawa (2004) in the following ways.

(a) Our model is not a pure ARCH model or a simple GARCH(1,1). There are q GARCH terms in model (1.2).

(b) Instead of the uniform local asymptotic normality approach of maximum likelihood estimation (MLE), we consider Gaussian quasi-maximum likelihood estimation (QMLE) for PTTGARCH(p,q) model and obtain asymptotic normality of QMLE under the condition that the error distribution has finite fourth moment.

(c) Our LADE approach relaxes the moment condition for the error distribution to the minimum. Its asymptotic normality enables us to do statistical inference on PTTGARCH(p,q) model with heavy-tailed errors.

We also give a necessary and sufficient condition for the existence of a strictly stationary solution of model (1.2), and study the existence of the moments and the tail behavior of the model. Furthermore, an order selection method is established by using the Wald statistic based on the asymptotic normality of LADE for a heavy-tailed PTTGARCH(p,q) model. A simulation study indicates that the LADE is more accurate than the QMLE when the errors are heavy-tailed. We give a real data example to illustrate the practicality of our theory. Our results in this paper is relevant because much empirical evidence shows that financial data often have heavy tails (see Adler et al. (1997), Mittnik and Rachev (2000)).

The rest of this paper is organized as follows. In Section 2, asymptotic normality of QMLE and LADE is established. Section 3 investigates tests for GARCH structures and the order selection problem. Section 4 presents a simulation study and a real data example. All the proofs of the main results in Section 2 and Section 3 are presented in Section 5. The Appendix presents the stationarity and existence of moments for PTTGARCH(p,q) model.

In the sequel, $\stackrel{\mathcal{L}}{\to}$, $\stackrel{P}{\to}$ and $\stackrel{a.s}{\to}$ denote convergence in distribution, in probability and almost surely respectively. A' denotes the transpose of a vector or a matrix A, $\|\cdot\|$ denotes the Euclidean norm unless declared otherwise and C is a constant which may be different at different places.

2 Estimation

Suppose that the data generating process is model (1.2). To avoid pathological cases, we assume that α_{1p} or $\alpha_{2p} > 0$, and $\beta_q > 0$ if $q > 0$. Let

$$\phi = (\delta, \alpha_0, \alpha_{11}, \alpha_{21}, \cdots, \alpha_{1p}, \alpha_{2p}, \beta_1, \cdots, \beta_q)'$$

be the parametric vector with true value $\phi^0 = (\delta^0, \alpha_0^0, \alpha_{11}^0, \alpha_{21}^0, \cdots, \alpha_{1p}^0, \alpha_{2p}^0, \beta_1^0, \cdots, \beta_q^0)'$. Define

$$\sigma_t(\phi) = [\alpha_0 + \sum_{i=1}^{p} \alpha_{1i}(X_{t-i}^{+})^{2\delta} + \sum_{i=1}^{p} \alpha_{2i}(X_{t-i}^{-})^{2\delta} + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^{2\delta}(\phi)]^{1/2\delta}.$$

Our basic assumptions are as follows.

A1 ε_t is non-degenerate and symmetrically distributed. Furthermore, $E|\varepsilon_t|^{\Delta} < +\infty$ for some $\Delta > 0$, and

$$\lim_{t\to 0} t^{-\mu} P\{\varepsilon_t^2 \le t\} = 0, \quad \textit{for some } \mu > 0. \tag{2.1}$$

A2 Θ is a compact subset of R^d, ϕ_0 is in the interior of Θ, and the Lyapunov exponent $\gamma(\phi) < 0$ for all $\phi \in \Theta$ (see (A.4) in the Appendix).

Remark 1. Because of the compactness of Θ, there exist positive constants $\delta_1, \delta_2, \rho_0$ such that $0 < \delta_1 < \delta, \alpha_0 < \delta_2$, $\delta_1 \le \sum_{i=1}^q \beta_i \le \rho_0 < 1$ for any $\phi \in \Theta$.

Under Assumption A1-A2, it may be deduced that (1.2) implies that

$$\begin{aligned} \sigma_t(\phi)^{2\delta} &= \frac{\alpha_0}{1-\sum_{j=1}^q \beta_j} + \sum_{i=1}^{p} \alpha_{1i}(X_{t-i}^{+})^{2\delta} + \sum_{i=1}^{p} \alpha_{2i}(X_{t-i}^{-})^{2\delta} \\ &\quad + \sum_{i=1}^{p} \alpha_{1i} \sum_{k=1}^{\infty} \sum_{j_1=1}^{q} \cdots \sum_{j_k=1}^{q} \beta_{j_1} \cdots \beta_{j_k} (X_{t-i-j_1-\cdots-j_k}^{+})^{2\delta} \\ &\quad + \sum_{i=1}^{p} \alpha_{2i} \sum_{k=1}^{\infty} \sum_{j_1=1}^{q} \cdots \sum_{j_k=1}^{q} \beta_{j_1} \cdots \beta_{j_k} (X_{t-i-j_1-\cdots-j_k}^{-})^{2\delta}. \end{aligned} \tag{2.2}$$

The derivatives of $\sigma_t(\phi)^{2\delta}$, which are very useful in the sequel, may be deduced from (2.2) as follows.

$$\begin{aligned}
\frac{\partial\sigma_t(\phi)^{2\delta}}{\partial\delta} &= \sum_{i=1}^{p}\alpha_{1i}X_{t-i}^{+2\delta}\log(X_{t-i}^{+})^2 + \sum_{i=1}^{p}\alpha_{2i}(X_{t-i}^{-})^{2\delta}\log(X_{t-i}^{-})^2 \qquad (2.3)\\
&\quad + \sum_{i=1}^{p}\alpha_{1i}\sum_{k=1}^{\infty}\sum_{j_1=1}^{q}\cdots\sum_{j_k=1}^{q}\beta_{j_1}\cdots\beta_{j_k}(X_{t-i-j_1-\cdots-j_k}^{+})^{2\delta}\log(X_{t-i-j_1-\cdots-j_k}^{+})^2\\
&\quad + \sum_{i=1}^{p}\alpha_{2i}\sum_{k=1}^{\infty}\sum_{j_1=1}^{q}\cdots\sum_{j_k=1}^{q}\beta_{j_1}\cdots\beta_{j_k}(X_{t-i-j_1-\cdots-j_k}^{-})^{2\delta}\log(X_{t-i-j_1-\cdots-j_k}^{-})^2\\
&=: L_{1t}^{\phi_1}(\phi) + L_{2t}^{\phi_1}(\phi) + L_{3t}^{\phi_1}(\phi);
\end{aligned}$$

$$\frac{\partial\sigma_t(\phi)^{2\delta}}{\partial\alpha_0} = \frac{1}{1-\sum_{j=1}^{q}\beta_j}; \qquad (2.4)$$

$$\frac{\partial\sigma_t(\phi)^{2\delta}}{\partial\alpha_{1i}} = (X_{t-i}^{+})^{2\delta} + \sum_{k=1}^{\infty}\sum_{j_1=1}^{q}\cdots\sum_{j_k=1}^{q}\beta_{j_1}\cdots\beta_{j_k}(X_{t-i-j_1-\cdots-j_k}^{+})^{2\delta}; \qquad (2.5)$$

$$\frac{\partial\sigma_t(\phi)^{2\delta}}{\partial\alpha_{2i}} = (X_{t-i}^{-})^{2\delta} + \sum_{k=1}^{\infty}\sum_{j_1=1}^{q}\cdots\sum_{j_k=1}^{q}\beta_{j_1}\cdots\beta_{j_k}(X_{t-i-j_1-\cdots-j_k}^{-})^{2\delta}; \qquad (2.6)$$

$$\begin{aligned}
\frac{\partial\sigma_t(\phi)^{2\delta}}{\partial\beta_j} &= \frac{\alpha_0}{(1-\sum_{j=1}^{q}\beta_j)^2} + \sum_{i=1}^{p}\alpha_{1i}\sum_{k=0}^{\infty}(k+1)\sum_{j_1=1}^{q}\cdots\sum_{j_k=1}^{q}\beta_{j_1}\cdots\beta_{j_k}(X_{t-i-j-j_1-\cdots-j_k}^{+})^{2\delta}\\
&\quad + \sum_{i=1}^{p}\alpha_{2i}\sum_{k=0}^{\infty}(k+1)\sum_{j_1=1}^{q}\cdots\sum_{j_k=1}^{q}\beta_{j_1}\cdots\beta_{j_k}(X_{t-i-j-j_1-\cdots-j_k}^{-})^{2\delta} \qquad (2.7)\\
&=: L_{1t}^{\beta_j}(\phi) + L_{2t}^{\beta_j}(\phi) + L_{3t}^{\beta_j}(\phi).
\end{aligned}$$

In the above expressions, we set $(X_t^{+})^{2\delta}\log(X_t^{+})^2 = 0$ if $X_t \le 0$ and $(X_t^{-})^{2\delta}\log(X_t^{-})^2 = 0$ if $X_t \ge 0$. In practice, however, $\sigma_t(\phi)$ cannot be computed using equation (2.2), since X_t is only observed for $1 \le t \le n$. We have to use the following approximation for $\sigma_t(\phi)$ based on $\{X_1, \cdots, X_n\}$.

$$\begin{aligned}
\tilde{\sigma}_t(\phi)^{2\delta} &= \frac{\alpha_0}{1-\sum_{j=1}^{q}\beta_j} + \sum_{i=1}^{p}\alpha_{1i}(X_{t-i}^{+})^{2\delta} + \sum_{i=1}^{p}\alpha_{2i}(X_{t-i}^{-})^{2\delta}\\
&+ \sum_{i=1}^{p}\alpha_{1i}\sum_{k=1}^{\infty}\sum_{j_1=1}^{q}\cdots\sum_{j_k=1}^{q}\beta_{j_1}\cdots\beta_{j_k}(X_{t-i-j_1-\cdots-j_k}^{+})^{2\delta}I(t-i-j_1-\cdots-j_k \ge 1)\\
&+ \sum_{i=1}^{p}\alpha_{2i}\sum_{k=1}^{\infty}\sum_{j_1=1}^{q}\cdots\sum_{j_k=1}^{q}\beta_{j_1}\cdots\beta_{j_k}(X_{t-i-j_1-\cdots-j_k}^{-})^{2\delta}I(t-i-j_1-\cdots-j_k \ge 1). \quad (2.8)
\end{aligned}$$

2.1 Quasi Maximum Likelihood Estimation (QMLE)

In this subsection, we deal with QMLE of the parameters. The logarithm of the quasi-likelihood function (omitted some constant) is defined as

$$L_n(\phi) = \sum_{t=1}^{n} -\frac{1}{2}\{\log \sigma_t^2(\phi) + \frac{X_t^2}{\sigma_t^2(\phi)}\} = \sum_{t=1}^{n} l_t(\phi), \tag{2.9}$$

where $\sigma_t(\phi)$ is defined by (2.2). The QMLE of ϕ is $\bar{\phi}_n = arg\max_{\phi\in\Theta} L_n(\phi)$. Define

$$\Lambda(\phi) = E\{\frac{\partial^2 l_t(\phi)}{\partial\phi\partial\phi'}\} \quad and \quad \Omega(\phi) = E\{\frac{\partial l_t(\phi)}{\partial\phi}\frac{\partial l_t(\phi)}{\partial\phi'}\}.$$

In order to obtain the consistency and asymptotic normality of $\bar{\phi}_n$, we need an additional condition, namely

A3 $E\varepsilon_t^2 = 1$, and $E\varepsilon_t^4 < \infty$.

Remark 2. If ε_t has density at 0, then (2.1) is satisfied for any $\mu < 1/2$.

The following theorem shows that $\bar{\phi}_n$ is consistent and asymptotically normal.

Theorem 1. *Under assumptions A1-A3, it follows that*
(i) $\bar{\phi}_n \xrightarrow{a.s} \phi^0$,
(ii) $\sqrt{n}(\bar{\phi}_n - \phi^0) \xrightarrow{\mathcal{L}} N(0, \Lambda_0^{-1}\Omega_0\Lambda_0^{-1})$, *where* $\Lambda_0 = \Lambda(\phi^0)$ *and* $\Omega_0 = \Omega(\phi^0)$.

As mentioned earlier, we can only observe $X_1, \cdots, X_n$ in practice. So we replace $L_n(\phi)$ by

$$\tilde{L}_n(\phi) = \sum_{t=1}^{n} -\frac{1}{2}\{\log \tilde{\sigma}_t^2(\phi) + \frac{X_t^2}{\tilde{\sigma}_t^2(\phi)}\}.$$

Similarly, we define $\tilde{\phi}_n = arg\max_{\phi\in\Theta} \tilde{L}_n(\phi)$. Let

$$\tilde{\Lambda}(\phi) = \frac{1}{n}\sum_{t=1}^{n} \frac{\partial^2 \tilde{l}_t(\phi)}{\partial\phi\partial\phi'} \quad and \quad \tilde{\Omega}(\phi) = \frac{1}{n}\sum_{t=1}^{n} \frac{\partial \tilde{l}_t(\phi)}{\partial\phi}\frac{\partial \tilde{l}_t(\phi)}{\partial\phi'},$$

where $\tilde{l}_t(\phi)$ are defined similar to $l_t(\phi)$ by replacing $\sigma_t(\phi)$ by $\tilde{\sigma}_t(\phi)$. The next theorem shows that our results for $\tilde{\phi}_n$ are the same as those for $\bar{\phi}_n$, and $\tilde{\Lambda}(\tilde{\phi}_n)$ and $\tilde{\Omega}(\tilde{\phi}_n)$ are consistent estimators of Λ_0 and Ω_0 respectively.

Theorem 2. *Under assumptions A1-A3, it follows that*

(i) $\tilde{\phi}_n \xrightarrow{a.s} \phi^0$,

(ii) $\sqrt{n}(\tilde{\phi}_n - \phi^0) \xrightarrow{\mathcal{L}} N(0, \Lambda_0^{-1}\Omega_0\Lambda_0^{-1})$,

(iii) $\tilde{\Lambda}_n \equiv \tilde{\Lambda}(\tilde{\phi}_n) \xrightarrow{a.s} \Lambda_0$ *and* $\tilde{\Omega}_n \equiv \tilde{\Omega}(\tilde{\phi}_n) \xrightarrow{a.s} \Omega_0$

Based on Theorem 1, we can develop some statistical inference about model (1.2). For example, we can consider a general form of the linear null hypothesis

$$H_0 : \Gamma\phi^0 = \theta, \tag{2.10}$$

where Γ is a $s \times d$ constant matrix with rank s, and θ is $s \times 1$ constant vector. By Theorem 2 and Theorem 4 in the next section, the asymptotic distributions of the likelihood ratio (LR) test statistic, the Lagrange multiplier (LM) test statistic and the Wald test statistic are χ^2.

Remark 3. From the above discussion, it can be seen that if we apply QMLE, we need Asumption A3, which is quite restrictive on the parameter vector and excludes the heavy tailed cases.

2.2 Least Absolute Deviations Estimation (LADE)

We have seen from the above that the QMLE requires stringent moment conditions on ε_t and X_t. However, empirical evidence indicates that financial data may have heavy tails. In recent years, the problem of statistical inference about GARCH-type models with weak moment conditions on x_t and ε_t has attracted much attention (see Hall and Yao (2003)). We introduce LADE for PTTGARCH(p,q) model, which only requires conditions for strict stationarity and assumption A4. Define an objective function as in Peng and Yao (2003)

$$\tilde{S}_n(\phi) = \sum_{t=u}^{n} |\log|X_t| - \log\tilde{\sigma}_t(\phi)|,$$

where $u = u(n)$ is a positive number satisfying $u(n) \to \infty$ and $u(n)/n \to 0$ as $n \to \infty$. The LADE is a minimizer of the objective function on the parameter space

$$\hat{\phi}_n = arg\min_{\phi\in\Theta} \tilde{S}_n(\phi).$$

Denote

$$v = (v_1, \cdots, v_d)' = \sqrt{n}(\phi - \phi^0), \ \ \tilde{Q}_t(\phi) = \log|X_t| - \log\tilde{\sigma}_t(\phi), \ \ \textit{and} \ \ Q_t(\phi) = \log|X_t| - \log\sigma_t(\phi),$$

where $d = 2p + q + 2$. It is easy to see that $\hat{\phi}_n = \phi^0 + \hat{v}/\sqrt{n}$, where $\hat{v}$ is a minimizer of

$$\tilde{T}_n(v) = \sum_{t=u}^{n}(|\tilde{Q}_t(\phi^0 + n^{-1/2}v)| - |\tilde{Q}_t(\phi^0)|).$$

Define

$$D_t(\phi) = (D_{t,1}(\phi), \cdots, D_{t,d}(\phi))', \quad \Sigma = E(D_t(\phi^0)D_t'(\phi^0)),$$

where

$$D_{t,1}(\phi) = -\frac{\partial Q_t(\phi)}{\partial \phi_1} = -\frac{1}{2\delta^2}\log\sigma_t^{2\delta}(\phi) + \frac{1}{2\delta\sigma_t^{2\delta}(\phi)}\frac{\partial\sigma_t^{2\delta}(\phi)}{\partial\delta}, \tag{2.11}$$

$$D_{t,i}(\phi) = -\frac{\partial Q_t(\phi)}{\partial \phi_i} = \frac{1}{2\delta\sigma_t^{2\delta}(\phi)}\frac{\partial\sigma_t^{2\delta}(\phi)}{\partial\phi_i}, \quad i = 2, \cdots, d. \tag{2.12}$$

We need the following condition on the error distribution instead of Assumption A3.

A4 $\log|\varepsilon_t|$ has zero median and a differentiable positive density function $f(x)$ such that $\sup_{x\in R}|f(x)| < B_1 < \infty$ and $\sup_{x\in R}|f'(x)| < B_2 < \infty$.

Theorem 3. *Suppose that conditions A1, A2 and A4 hold. Then for any given positive random variable M with $P(0 < M < \infty) = 1$, there exists a local minimizer $\hat{\phi}$ of $S_n(\phi)$ which lies in the random region $\{\phi : \|\phi - \phi_0 - \xi/\sqrt{n}\| \leq M/\sqrt{n}\}$ for which*

$$\sqrt{n}(\hat{\phi}_n - \phi^0) \longrightarrow_{\mathcal{L}} N\big(0, \frac{1}{4f^2(0)}\Sigma^{-1}\big).$$

Here ξ is a normal random vector with mean 0 and covariance matrix $\frac{1}{4f^2(0)}\Sigma^{-1}$.

3 Order Selection and Tests for GARCH Structure

The likelihood ratio (LR) test, the Lagrange multiplier (LM) test and the Wald test are the three standard approaches to constructing test statistics for parametric hypotheses. However, the first two depend on the likelihood function and MLE, which are very sensitive to heavy tails (see Hall and Yao (2003)). Therefore, we use a Wald test statistic based on LADE for the heavy-tailed case. We consider a general form of linear null hypothesis (2.10). A Wald test statistic is defined as

$$W_n(s) \quad = \quad (\Gamma\hat{\phi}_n - \theta)'\big\{\Gamma\frac{1}{4n\hat{f}^2(0)}\hat{\Sigma}^{-1}\Gamma'\big\}^{-1}(\Gamma\hat{\phi}_n - \theta).$$

We reject H_0 for large values of $W_n(s)$. In the above expression,

$$\hat{\Sigma} = \frac{1}{n}\sum_{t=1}^{n}\tilde{D}_t(\hat{\phi}_n)\tilde{D}_t(\hat{\phi}_n)', \qquad \hat{f}(0) = \frac{1}{nb_n}\sum_{t=1}^{n}K\big(\frac{1}{b_n}\log\frac{|X_t|}{\tilde{\sigma}_t(\hat{\phi}_n)}\big), \tag{3.1}$$

where $\tilde{D}_t(\hat{\phi}_n)$ is defined similar to $D_t(\hat{\phi}_n)$ by replacing $\sigma_t(\phi)$ by $\tilde{\sigma}_t(\phi)$, $K(\cdot)$ is a density function on R, and $b_n > 0$ is a bandwidth. The following theorem gives the limiting distribution of $W_n(s)$ under H_0.

Theorem 4. *Suppose the conditions of Theorem 3 hold. If the kernel function K and the bandwidth b_n satisfy the following assumptions*

(i) K is Lipschitz continuous and of finite first moment;

(ii) $b_n \to 0$ and $nb_n^4 \to \infty$ as $n \to \infty$,

then $W_n(s) \to_{\mathcal{L}} \chi_s^2$ under H_0.

For testing an order (p, q) against a higher order (p, Q_0) or (P_0, q) with $P_0 > p$ and $Q_0 > q$, we can take a Γ in (2.10) such that

$$\Gamma\phi^0 = (\beta^0_{q+1}, \cdots, \beta^0_{Q_0})', \;\; or \;\; \Gamma\phi^0 = (\alpha^0_{1p+1}, \alpha^0_{2p+1} \cdots, \alpha^0_{1P_0}, \alpha^0_{2P_0})' \tag{3.2}$$

and

$$\theta = (0, \cdots, 0)'_{(Q_0-q)\times 1}, \;\; or \;\; \theta = (0, \cdots, 0)'_{2(P_0-p)\times 1}.$$

Notice that a GARCH(p,q) model cannot be tested directly against an $GARCH(P_0, Q_0)$ using the standard technique because of the identification problem already discussed by Bollerslev (1986), and the situation is the same for PTTGARCH (p,q) models. As pointed out by Ling (2005), the above test procedure is very useful in model building. In fact, we can use it to select the order. Suppose that the order of model (1.2) does not exceed (P_0, Q_0). For a given significant level η, we can take the above test in order for $p = P_0 - 1, \cdots, s_0$ and $q = Q_0 - 1, \cdots, r_0$ in (3.2) until $p = s_0$ and $q = r_0$ such that $W_n(s) > \chi_s^2(\eta)$. Then we can declare that the order of model (1.2) is (s_0, r_0).

Because model (1.2) is a very general framework, we can test whether some special case is true or not. In the following, we mainly discuss testing problems about GARCH structures for $\{X_t\}$.

(i) Bollerslev's standard GARCH: $\delta = 1$ and $\alpha_{1i} = \alpha_{2i},\ \ i = 1, \cdots, p$;

(ii) IGARCH: $\delta = 1$, $\alpha_{1i} = \alpha_{2i} = \alpha_i,\ \ i = 1, \cdots, p$ and $\sum_{i=1}^p \alpha_i + \sum_{j=1}^q \beta_j = 1$;

(iii) Symmetric GARCH: $\alpha_{1i} = \alpha_{2i},\ \ i = 1, \cdots, p$

(iv) No power transformation: $\delta = 1$.

Let

$$\Gamma_1 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & -1 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 0 & 0 & \cdots & 1 & -1 & 0 & \cdots & 0 \end{pmatrix}_{(p+1)\times d},$$

$$\Gamma_2 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & -1 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 0 & 0 & \cdots & 1 & -1 & 0 & \cdots & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \cdots & \frac{1}{2} & \frac{1}{2} & 1 & \cdots & 1 \end{pmatrix}_{(p+2)\times d},$$

$$\Gamma_3 = \begin{pmatrix} 0 & 0 & 1 & -1 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 0 & 0 & \cdots & 1 & -1 & 0 & \cdots & 0 \end{pmatrix}_{p\times d},$$

$$\Gamma_4 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 \end{pmatrix}_d',$$

$$\theta_1 = (1, \cdots, 0, 0)_{p+1}', \quad \theta_1 = (1, 0, \cdots, 0, 1)_{p+2}' \quad \theta_3 = (0, \cdots, 0)_p', \quad \theta_4 = 1.$$

Then the above four testing problems can be written in the form with

$$H_{i0} : \Gamma_i \phi^0 = \theta_i, \quad i = 1, \cdots, 4.$$

So the Wald-test provides a simple way to test the null hypothesis of a particular specification against wider nonlinear alternatives. For example, we can determine whether Bollerslev's standard model provides an adequate description of the data by testing H_{10}.

4 Simulations and empirical results

In this section, we perform a simulation study to demonstrate the accuracy of LADE in heavy-tailed case and apply the theory in Section 2 and Section 3 to the Hong Kong Hang Seng Index (HSI) series.

Firstly, we compare numerically LADE and QMLE for the PTTGARCH(1,1) model. The data are generated by the PTTGARCH(1,1) model

$$X_t = \sigma_t \varepsilon_t \quad and \quad \sigma_t^{2\delta} = \alpha_0 + \alpha_{11}(X_{t-1}^+)^{2\delta} + \alpha_{21}(X_{t-1}^-)^{2\delta} + \beta_1 \sigma_{t-1}^{2\delta},$$

with the true parameter $(\delta^0, \alpha_0^0, \alpha_{11}^0, \alpha_{21}^0, \beta_1^0)' = (0.8, 0.2, 0.2, 0.1, 0.4)'$. We take the errors ε_t to have either a standard normal distribution or a standardized Student's t-distribution with degrees of freedom $d = 2, 3, 4, 5$. The sample size is $n = 600$ and we draw 1000 independent samples. For LADE, u was set to be $u = 10$.

Figure 1 presents the boxplots of the average absolute error (AAE) $\big(|\hat{\delta} - 0.8| + |\hat{\alpha}_0 - 0.2| + |\hat{\alpha}_{11} - 0.2| + |\hat{\alpha}_{21} - 0.1| + |\hat{\beta} - 0.4|\big)/5$ for both LADE and QMLE. For samples with heavy-tailed errors, i.e., $t(2)$, $t(3)$ and $t(4)$, LADE performs better than QMLE especially for $t(2)$ and $t(3)$. As expected, QMLE is better when the errors are $t(5)$ and $N(0, 1)$.

Then we apply the PTTGARCH model to daily HSI from 2001 to 2003, which has a total of 738 observations. The return series X_t is defined as the percentage of the log difference of the index. Figure 2 and figure 3 are the time plots of the index and the return respectively. They display some drastic shocks, which are caused by the 11/9 terrorist attack on $11 - 09 - 2001$ and the severe acute respiratory syndrome (SARS) in China erupted in March 2003.

We first study whether or not $\{X_t\}$ is heavy-tailed. The Hill estimator and the QQ-plot are used for this. The Hill estimator is defined as

$$H_{n,m} = \Big[\frac{1}{m}\sum_{i=1}^{m} \log(\frac{X_{(i)}}{X_{(m+1)}})\Big]^{-1}, \quad n = 737,$$

where $X_{(1)} \geq X_{(2)} \geq \cdots \geq X_{(n)}$ are the order statistics of $X_1, \cdots, X_n$. We plot $\{(m, H_{n,m}), 1 \leq m \leq 600\}$ in Figure 4, which suggests that X_t has an infinite fourth moment or even probably infinite variance, since the estimator is less than 4 for $m \geq 20$ and less than 2 for $m \geq 200$. Figure 5 presents the QQ-plot for $\{X_t\}$, which suggests that $\{X_t\}$ is very heavy-tailed.

To test whether $\{X_t\}$ is white noise, we use the Wald test statistics based on the weighted least absolute estimators with the weight function

$$w_t = \begin{cases} 1, & \text{if } a_t = 0; \\ 0.5C^3/a_t^3, & \text{if } a_t \neq 0; \end{cases}$$

where $a_t = \sum_{i=1}^{p} |X_{t-i}| I(|X_{t-i}| \geq C)$ and C is the 90 percent quantile of the data $\{X_t\}$ (see Ling (2005) for details), since the Box-Pierce statistic is not applicable for the heavy-tailed case. Here and in the following we take the kernel function $K(x) = e^{-x}/(1 + e^{-x})$ and $b_n = 1.06n^{-1/5}$. We obtain that $W_n(8) = 2.31$ and $W_n(12) = 2.76$. Both are not significant at 0.05 level. However, for the squared series $\{X_t^2\}$, we obtain that $W_n^2(8) = 105.85$ and $W_n^2(12) = 108.55$, which are highly significant at the level 0.05. This suggests that the series $\{X_t\}$ has conditional heteroscedastic structure.

Now, we fit a PTTGARCH(1,1) model using QMLE to the data. The estimates are

$$(\tilde{\delta}, \tilde{\alpha}_0, \tilde{\alpha}_{11}, \tilde{\alpha}_{21}, \tilde{\beta}) = (1.6616, 0.3414, 0.0368, 0.1561, 0.7405)$$

with standard errors 0.0824, 0.1025, 0.0229, 0.1211, and 0.0476 respectively. For the standardized residuals, we obtain $W_n(8) = 4.88$ and $W_n(12) = 7.28$; for the squared standardized residuals, we obtained $W_n^2(8) = 15.21$ and $W_n^2(12) = 17.64$. Based on the 5% significance level of the $\chi^2(8)$ and $\chi^2(12)$ distribution, the PTTGARCH(1,1) model fits the data adequately according to both statistics $W_n^2(8)$,$W_n^2(12)$,$W_n^2(8)$, and $W_n^2(12)$. Figure 6 shows the Hill estimator of the standardized residuals, which indicates that the residuals may have infinite fourth moment since the Hill estimator is less than 3.5 when $m \geq 80$. The QQ-plot in Figure 7 also shows that the residuals are heavy-tailed. Thus, we fit a PTTGARCH model to the data with LADE.

For order selection, we assume that $p, q \leq 3$ for simplicity. Using the procedure for order selection in Section 3, we test $p = 3, q = 2$ v.s $p = 3, q = 3$; $p = 3, q = 1$ v.s $p = 3, q = 2$; and $p = 2, q = 1$ v.s $p = 1, q = 1$ in order, and all the Wald statistics are less than 1, namely not significant. Then we test $p = 1, q = 1$ v.s $p = 1, q = 0$, and the Wald statistic is 34.9, which rejects the null hypothesis and we take $p = 1, q = 1$. The LADEs are

$$(\hat{\delta}, \hat{\alpha}_0, \hat{\alpha}_{11}, \hat{\alpha}_{21}, \hat{\beta}) = (0.5384, 0.0495, 0.0189, 0.0899, 0.8642)$$

with standard errors 0.0524, 0.0123, 0.0296, 0.0571, and 0.0464 respectively. To check the adequacy of the estimated PTTGARCH(1,1) model, we conduct the white noise test for the residuals and the squared residuals using the same method as for X_t before. We have $W_n(8) = 2.60$ and $W_n(12) = 3.41$ for the residuals and $W_n^2(8) = 11.99$ and $W_n^2(12) = 17.46$ for the squared residuals, which are all not significant at 0.05 level. Hence, the estimated PTTGARCH(1,1) model is adequate for the data $\{X_t\}$. Notice that for both the residuals and the squared residuals, all the Wald statistics based on LADE are less than those based on QMLE, which suggests that the fitted model based on LADE is the more adequate. For the fitted model using LADE, we also test the hypothese $\delta = 1$ and $\alpha_{11} = \alpha_{21}$ respectively. The Wald statistic for the former is 14.24 and is highly significant. The Wald statistic for the latter is 2.41 and is not significant, which may be caused by the small values of α_{11} and α_{21}. In fact, as we can see from the estimators, $\hat{\alpha}_{21}$ is about five times $\hat{\alpha}_{11}$. This example illustrates that the data are asymmetric and nonlinear and the PTTGARCH model is capable of capturing these characteristics.

5 Theoretical Proofs

We use the same notation as in the Section 2. Before we prove Theorem 1—Theorem 3, we introduce some lemmas first.

Lemma 1. *Under assumptions A1—A2, there exist positive constants $0 < r < 1$ and $C > 0$ independent from ϕ such that*

$$\sup_{\phi \in \Theta} |\sigma_t^{2\delta}(\phi) - \tilde{\sigma}_t^{2\delta}(\phi)| \leq C \sum_{j=t}^{\infty} r^t (|X_{t-j}|^{2\delta_2} + 1) \tag{5.1}$$

and

$$\sup_{\phi \in \Theta} \| \frac{\partial \sigma_t^{2\delta}(\phi)}{\partial \phi} - \frac{\partial \tilde{\sigma}_t^{2\delta}(\phi)}{\partial \phi} \| \leq C \sum_{j=t}^{\infty} r^t (|X_{t-j}|^{2\delta_2} + 1), \tag{5.2}$$

where δ_2 is the positive constant in Remark 1.

Proof. Denote

$$\frac{1+\alpha_{11}x+\cdots+\alpha_{1p}x^p}{1-\beta_1 x-\cdots-\beta_q x^q}=\sum_{i=0}^{\infty} d_{1i}x^i, \quad |x|\le 1$$

and

$$\frac{1+\alpha_{21}x+\cdots+\alpha_{2p}x^p}{1-\beta_1 x-\cdots-\beta_q x^q}=\sum_{i=0}^{\infty} d_{2i}x^i, \quad |x|\le 1.$$

By Lemma 3.1 of Berkes et al. (2003), there exist some constants $C>0$ and $0<r<1$ such that

$$0<d_{it}\le Cr^t, \quad i=1,2; \quad t=0,1,\cdots,$$

where C and r are both independent of ϕ. Notice that

$$\sigma_t^{2\delta}(\phi)-\tilde{\sigma}_t^{2\delta}(\phi)=\sum_{i=t}^{\infty} d_{1i}(X_{t-i}^{+})^{2\delta}+\sum_{i=t}^{\infty} d_{2i}(X_{t-i}^{-})^{2\delta},$$

thus (5.1) holds. Similarly, we can show that (5.2) holds.

$\square$

Lemma 2. *Under assumption A1—A2, it follows that*

$$E\sup_{\phi\in\Theta}\left|\frac{\partial^m Q_t(\phi)}{\partial\phi_{j_1}\cdots\partial\phi_{j_m}}\right|^h<+\infty,$$

where m, h are any positive integers, $1\le j_1,\cdots,j_m\le d$ *and* $Q_t(\phi)=\log|X_t|-\log\sigma_t(\phi)$.

Proof. We only prove the case $m=1$. The proof of the case $m>1$ is similar. From (2.11) and (2.12), it is sufficient to prove that

$$E\sup_{\phi\in\Theta}|\log\sigma_t^{2\delta}(\phi)|^h<+\infty; \tag{5.3}$$

$$E\sup_{\phi\in\Theta}\left|\frac{1}{\sigma_t^{2\delta}(\phi)}\frac{\partial\sigma_t^{2\delta}(\phi)}{\partial\phi_i}\right|^h<+\infty, \quad i=2,\cdots,d; \tag{5.4}$$

$$E\sup_{\phi\in\Theta}\left|\frac{1}{\sigma_t^{2\delta}(\phi)}\frac{\partial\sigma_t^{2\delta}(\phi)}{\partial\delta}\right|^h<+\infty. \tag{5.5}$$

By the definition of $\sigma_t^{2\delta}$, we have $\sigma_t^{2\delta}>\alpha_0\ge\delta_1$, where δ_1 is the positive constant in Remark 1. Notice that for any given positive integers h, there exists $M_1>0$ such that $(\log x)^h\le x^{\tilde{h}}$

for $x > M_1$, where $\tilde{h} = \min\{1, \tau/(2\delta_2)\}$, τ and δ_2 are defined respectively in Theorem 6 and Remark 1. By Lemma 1, it follows that

$$\delta_1 \leq \sup_{\phi\in\Theta} \sigma_t^{2\delta}(\phi) \leq C \sum_{j=1}^{\infty} r^j (|X_{t-j}|^{2\delta_2} + 1).$$

Noticing that

$$\log(\delta_1) \leq \log \sigma_t^{2\delta}(\phi) \leq \log \big(\sup_{\phi\in\Theta} \sigma_t^{2\delta}(\phi) \big),$$

we have

$$\sup_{\phi\in\Theta} |\log \sigma_t^{2\delta}(\phi)| \leq |\log \delta_1| + \log[C \sum_{j=1}^{\infty} r^j (|X_{t-j}|^{2\delta_2} + 1)] \leq [C + \sum_{j=1}^{\infty} r^{j\tilde{h}} (|X_{t-j}|^{\tau} + 1)]^{1/h},$$

which implies (5.3) holds by Theorem 6 (i).

It is obvious that (5.4) holds for $i = 2, \cdots, 2p+2$ by (2.4)—(2.6). From Lemma 5.2 of Berkes et al. (2003), we can obtain (5.4) for $i = 2p+3, \cdots, d$.

Now we prove (5.5). Using the same argument as (A.5), we can obtain $E \sup_{\phi\in\Theta} |\frac{L_{1t}^{\phi_1}(\phi)}{\sigma_t^{2\delta}(\phi)}|^h < +\infty$, where $L_{1t}^{\phi_1}(\phi)$ is defined in (2.3). On the other hand,

$$\begin{aligned} L_{2t}^{\phi_1}(\phi) &\leq \rho_0 \sum_{j=1}^{q} \sum_{i=1}^{p} \alpha_{1i} \sum_{k=0}^{\infty} \sum_{j_1=1}^{q} \cdots \sum_{j_k=1}^{q} \beta_{j_1} \cdots \beta_{j_k} (X^+_{t-i-j-j_1-\cdots-j_k})^{2\delta} \log(X^+_{t-i-j-j_1-\cdots-j_k})^2 \\ &= \rho_0 \sum_{j=1}^{q} L_{t,j}(\phi), \end{aligned}$$

where ρ_0 is the same one as in Remark 1 and $L_{2t}^{\phi_1}(\phi)$ is defined in (2.3). From $H\ddot{o}lder$ inequality and (5.4), it follows that

$$\big(E \sup_{\phi\in\Theta} |\frac{L_{t,j}(\phi)}{\sigma_t^{2\delta}(\phi)}|^h\big)^2 \leq E \sup_{\phi\in\Theta} |\frac{L_{t,j}(\phi)}{L_{2t}^{\beta_j}(\phi)}|^{2h} E \sup_{\phi\in\Theta} |\frac{L_{2t}^{\beta_j}(\phi)}{\sigma_t^{2\delta}(\phi)}|^{2h} \leq CE \sup_{\phi\in\Theta} |\frac{L_{t,j}(\phi)}{L_{2t}^{\beta_j}(\phi)}|^{2h}.$$

It can be easily verified that

$$\frac{x_1 + x_2}{y_1 + y_2} \leq \max\{\frac{x_1}{y_1}, \frac{x_2}{y_2}\}, \;\; for \;\; any \;\; x_1, x_2, y_1, y_2 > 0.$$

Let $I_1 = I\{|\log X^+_{t-i-j-j_1-\cdots-j_k}| \le 1\}$ and $I_2 = 1 - I_1$. Noticing (2.1) holds, We obtain

$$\begin{aligned}
& E\sup_{\phi\in\Theta} |\frac{L_{t,j}(\phi)}{L_{2t}^{\beta_j}(\phi)}|^{2h} \\
= \;& E\sup_{\phi\in\Theta} \Big\{ \frac{\sum_{i=1}^{p}\alpha_{1i}\sum_{k=0}^{\infty}\sum_{j_1=1}^{q}\cdots\sum_{j_k=1}^{q}\beta_{j_1}\cdots\beta_{j_k}(X^+_{t-i-j-j_1-\cdots-j_k})^{2\delta}}{\tau L_{2t}^{\beta_j}(\phi)} \\
& \cdot \log(X^+_{t-i-j-j_1-\cdots-j_k})^{2\tau}(I_1+I_2)\Big\}^{2h} \\
\le \;& C + CE\{\max_{k\ge 0}\max_{j_1,\cdots,j_k}\frac{|\log(X^+_{t-i-j-j_1-\cdots-j_k})^{\tau}|I_2}{k+1}\}^{2h} \\
\le \;& C + C\int_1^{\infty} 2hy^{2h-1}P\{\max_{k\ge 0}\max_{j_1,\cdots,j_k}\frac{|\log(X^+_{t-i-j-j_1-\cdots-j_k})^{\tau}|I_2}{k+1} > y\}dy \\
\le \;& C + C\sum_{k=0}^{\infty}\int_1^{\infty} 2hy^{2h-1}P\{\frac{\max_{j_1,\cdots,j_k}|\log(X^+_{t-i-j-j_1-\cdots-j_k})^{\tau}|}{k+1} > y\}dy \\
= \;& C + C\sum_{k=0}^{\infty}\int_1^{\infty} 2hy^{2h-1}k[P\{(X_t^+)^{\tau} > e^{(k+1)y}\} + P\{(X_t^+)^{\tau} < e^{-(k+1)y}\}]dy \\
\le \;& C + C\sum_{k=0}^{\infty}\int_1^{\infty} ky^{2h-1}\cdot[e^{-(k+1)y}E(X_t^+)^{\tau} + P\{\varepsilon_t^+ < (\alpha_0)^{-1/(2\delta\tau)}e^{-(k+1)y/\tau}\}]dy \\
\le \;& C + C\sum_{k=0}^{\infty}\int_1^{\infty} ky^{2h-1}\cdot[e^{-(k+1)y}E(X_t^+)^{\tau} + e^{-\mu(k+1)y/\tau}]dy \\
< \;& +\infty.
\end{aligned}$$

Thus, $E\sup_{\phi\in\Theta}|\frac{L_{2t}^{\phi_1}(\phi)}{\sigma_t^{2\delta}(\phi)}|^h < +\infty$. Similarly, we can obtain that $E\sup_{\phi\in\Theta}|\frac{L_{3t}^{\phi_1}(\phi)}{\sigma_t^{2\delta}(\phi)}|^h < +\infty$. This completes the proof. □

Lemma 3. *Suppose that assumptions A1-A2 hold. If* $E|\varepsilon_t^2|^{1+2\eta} < \infty$ *for some* $\eta > 0$*, it follows that*

$$E\sup_{\phi\in\Theta}|\frac{\partial^m l_t(\phi)}{\partial\phi_{j_1}\cdots\partial\phi_{j_m}}| < +\infty,$$

where $1 \le j_1, \cdots, j_m \le d$.

Proof. We only prove the Lamma for $m = 1$, noting that similar arguments apply for $m > 1$. By the definition of $l_t(\phi)$, we have

$$\frac{\partial l_t(\phi)}{\partial\phi} = -\frac{1}{2\sigma_t^2(\phi)}\frac{\partial\sigma_t^2(\phi)}{\partial\phi} + \frac{X_t^2}{2\sigma_t^4(\phi)}\frac{\partial\sigma_t^2(\phi)}{\partial\phi}.$$

By similar argument to Lemma 5.1 of Berkes et al. (2003), we can obtain that

$$E|\sup_{\phi\in\Theta}\frac{\sigma_t^2(\phi^0)}{\sigma_t^2(\phi)}|^{1+\eta}<\infty. \tag{5.6}$$

From Hölder inequality, Lemma 2 and (5.6), we have

$$\begin{aligned} & E\sup_{\phi\in\Theta}\|\frac{X_t^2}{2\sigma_t^4(\phi)}\frac{\partial\sigma_t^2(\phi)}{\partial\phi}\| = E\sup_{\phi\in\Theta}\|\frac{\sigma_t^2(\phi^0)}{\sigma_t^2(\phi)}\frac{1}{2\sigma_t^2(\phi)}\frac{\partial\sigma_t^2(\phi)}{\partial\phi}\| \\ \leq\ & \big\{E|\sup_{\phi\in\Theta}\frac{\sigma_t^2(\phi^0)}{\sigma_t^2(\phi)}|^{1+\eta}\big\}^{\frac{1}{1+\eta}}\big\{E\sup_{\phi\in\Theta}\|\frac{1}{2\sigma_t^2(\phi)}\frac{\partial\sigma_t^2(\phi)}{\partial\phi}\|^{\frac{1+\eta}{\eta}}\big\}^{\frac{\eta}{1+\eta}}<\infty. \end{aligned}$$

Using Lemma 2 again, we obtain

$$E\sup_{\phi\in\Theta}\|\frac{1}{2\sigma_t^2(\phi)}\frac{\partial\sigma_t^2(\phi)}{\partial\phi}\|<\infty.$$

Thus, $E\sup_{\phi\in\Theta}\|\frac{\partial l_t(\phi)}{\partial\phi}\|<\infty$.

□

Lemma 4. *Suppose that assumptions A1-A3 hold, and denote* $L(\phi)=El_t(\phi)$ *for all* $\phi\in\Theta$*. Then* $L(\phi)$ *is well defined and* ϕ^0 *is the unique maximizer of* $L(\phi)$*.*

Proof. By Lemma 2 and Lemma 3, it is obvious that $L(\phi)$ is well defined.

Maximizing $L(\phi)$ is equivalent to minimizing $L(\phi^0)-L(\phi)$. But,

$$L(\phi^0)-L(\phi)=\frac{1}{2}E\big\{\frac{\sigma_t^2(\phi^0)}{\sigma_t^2(\phi)}-\log\frac{\sigma_t^2(\phi^0)}{\sigma_t^2(\phi)}\big\}-\frac{1}{2}.$$

Note that the function $x-\log x>0$ for any $x>0$ and reaches its unique minimum value at $x=1$. Since $\sigma_t^2(\phi)=\sigma_t^2(\phi^0)$ if and only if $\phi=\phi^0$, we obtain the result. □

Lemma 5. *If the conditions of Theorem 1 are satisfied, then as* $n\to\infty$*, we have*

$$\begin{aligned} &\sup_{\phi\in\Theta}|\frac{1}{n}L_n(\phi)-L(\phi)|\overset{a.s}{\to}0,\\ &\sup_{\phi\in\Theta}|\frac{1}{n}\frac{\partial L_n(\phi)}{\partial\phi}-\frac{\partial L(\phi)}{\partial\phi}|\overset{a.s}{\to}0\\ &\sup_{\phi\in\Theta}|\frac{1}{n}\frac{\partial^2 L_n(\phi)}{\partial\phi\partial\phi'}-\frac{\partial^2 L(\phi)}{\partial\phi\partial\phi'}|\overset{a.s}{\to}0. \end{aligned}$$

Proof. By the ergodic theorem,

$$\frac{1}{n} L_n(\phi) \xrightarrow{a.s} L(\phi), \tag{5.7}$$

for any $\phi \in \Theta$. Using the mean value theorem, we have

$$\begin{aligned}
& \sup_{\phi^1, \phi^2 \in \Theta} |L_n(\phi^1) - L_n(\phi^2)| \frac{1}{\|\phi^1 - \phi^2\|} \\
\leq \; & \frac{1}{2\|\phi^1 - \phi^2\|} \sum_{t=1}^{n} \sup_{\phi^1, \phi^2 \in \Theta} \{ |\log \sigma_t^2(\phi^1) - \log \sigma_t^2(\phi^2)| + |\frac{X_t^2}{\sigma_t^2(\phi^1)} - \frac{X_t^2}{\sigma_t^2(\phi^2)}| \} \\
\leq \; & \frac{C}{2} \sum_{t=1}^{n} \sup_{\phi^*, \phi^{**} \in \Theta} \{ \| \frac{1}{\sigma_t^2(\phi^*)} \frac{\partial \sigma_t^2(\phi^*)}{\partial \phi} \| + \| \frac{\varepsilon_t^2 \sigma_t^2(\phi^0)}{\sigma_t^4(\phi^{**})} \frac{\partial \sigma_t^2(\phi^{**})}{\partial \phi} \| \},
\end{aligned}$$

where ϕ^*, ϕ^{**} lie on the line form ϕ^1 to ϕ^2. We have

$$E\{ \sup_{\phi^*, \phi^{**} \in \Theta} [\| \frac{1}{\sigma_t^2(\phi^*)} \frac{\partial \sigma_t^2(\phi^*)}{\partial \phi} \| + \| \frac{\varepsilon_t^2 \sigma_t^2(\phi^0)}{\sigma_t^4(\phi^{**})} \frac{\partial \sigma_t^2(\phi^{**})}{\partial \phi} \|] \} < \infty$$

by Lemma 2 and Lemma 3. Then,

$$\sup_{\phi^1, \phi^2 \in \Theta} |\frac{1}{n} L_n(\phi^1) - \frac{1}{n} L_n(\phi^2)| \frac{1}{\|\phi^1 - \phi^2\|} = O(1), \quad a.s,$$

which shows that $\frac{L_n(\phi)}{n}$ is equicontinuous with probability one. Combining this fact, (5.7) and the compactness of Θ, the uniform convergence of $\frac{L_n(\phi)}{n}$ follows. By the same method, we can prove that the results hold for $\frac{1}{n} \frac{\partial L_n(\phi)}{\partial \phi}$ and $\frac{1}{n} \frac{\partial^2 L_n(\phi)}{\partial \phi \partial \phi'}$. □

Lemma 6. *If the conditions of Theorem 1 are satisfied, then it follows that*

$$\begin{aligned}
& \sup_{\phi \in \Theta} |\frac{1}{\sqrt{n}} L_n(\phi) - \frac{1}{\sqrt{n}} \tilde{L}_n(\phi)| \xrightarrow{a.s.} 0, \\
& \sup_{\phi \in \Theta} \| \frac{1}{\sqrt{n}} \frac{\partial L_n(\phi)}{\partial \phi} - \frac{1}{\sqrt{n}} \frac{\partial \tilde{L}_n(\phi)}{\partial \phi} \| \xrightarrow{a.s.} 0, \\
& \sup_{\phi \in \Theta} \| \frac{1}{\sqrt{n}} \frac{\partial^2 L_n(\phi)}{\partial \phi \partial \phi'} - \frac{1}{\sqrt{n}} \frac{\partial^2 \tilde{L}_n(\phi)}{\partial \phi \partial \phi'} \| \xrightarrow{a.s.} 0.
\end{aligned}$$

Proof. Let

$$U_1 = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \sup_{\phi \in \Theta} |\log \sigma_t^2(\phi) - \log \tilde{\sigma}_t^2(\phi)| \quad and \quad U_2 = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \sup_{\phi \in \Theta} |\frac{X_t^2}{\sigma_t^2(\phi)} - \frac{X_t^2}{\tilde{\sigma}_t^2(\phi)}|,$$

then

$$\sup_{\phi\in\Theta}|\frac{1}{\sqrt{n}}L_n(\phi)-\frac{1}{\sqrt{n}}\tilde{L}_n(\phi)| \quad \leq U_1+U_2.$$

By Lemma 1, we have

$$\begin{aligned}
U_1 &= \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\sup_{\phi\in\Theta}\frac{1}{\delta}\big|\log(1+\frac{\sigma_t^{2\delta}(\phi)-\tilde{\sigma}_t^{2\delta}(\phi)}{\tilde{\sigma}_t^{2\delta}(\phi)})\big| \\
&\leq \frac{C}{\sqrt{n}}\sum_{t=1}^{n}\sup_{\phi\in\Theta}\frac{1}{\delta}\frac{|\sigma_t^{2\delta}(\phi)-\tilde{\sigma}_t^{2\delta}(\phi)|}{\tilde{\sigma}_t^{2\delta}(\phi)} \\
&\leq \frac{C}{\sqrt{n}}\sum_{t=1}^{n}\sup_{\phi\in\Theta}|\sigma_t^{2\delta}(\phi)-\tilde{\sigma}_t^{2\delta}(\phi)| \\
&\leq \frac{C}{\sqrt{n}}\sum_{t=1}^{n}\sum_{j=t}^{\infty}r^t(1+|X_{t-j}|^{\delta_2}) \\
&\leq \frac{C}{\sqrt{n}}\sum_{t=1}^{n}r^t+\frac{C}{\sqrt{n}}\sum_{t=1}^{n}r^t\sum_{h=0}^{\infty}r^h|X_{-h}|^{\delta_2} \\
&\stackrel{a.s}{\to} 0.
\end{aligned}$$

By the mean value theorem, we have

$$\sigma_t^2(\phi)-\tilde{\sigma}_t^2(\phi)=\big(\sigma_t^{2\delta}(\phi)\big)^{\frac{1}{\delta}}-\big(\tilde{\sigma}_t^{2\delta}(\phi)\big)^{\frac{1}{\delta}}=\frac{1}{\delta}(\sigma_t^*)^{\frac{1}{\delta}-1}(\sigma_t^{2\delta}(\phi)-\tilde{\sigma}_t^{2\delta}(\phi)),$$

where σ_t^* lies between $\sigma_t^{2\delta}(\phi)$ and $\tilde{\sigma}_t^{2\delta}(\phi)$. Thus we have $U_2\stackrel{a.s}{\to}0$ by a similar way to the proof of $U_1\stackrel{a.s}{\to}0$ from Lemma 1. Therefore,

$$\sup_{\phi\in\Theta}|\frac{1}{\sqrt{n}}L_n(\phi)-\frac{1}{\sqrt{n}}\tilde{L}_n(\phi)| \xrightarrow{a.s.} 0. \tag{5.8}$$

Using the same method as the proof of (5.8), we get

$$\begin{aligned}
&\sup_{\phi\in\Theta}|\frac{1}{\sqrt{n}}\frac{\partial L_n(\phi)}{\partial\phi}-\frac{1}{\sqrt{n}}\frac{\partial\tilde{L}_n(\phi)}{\partial\phi}| \xrightarrow{a.s.} 0, \\
&\sup_{\phi\in\Theta}\|\frac{1}{\sqrt{n}}\frac{\partial^2 L_n(\phi)}{\partial\phi\partial\phi'}-\frac{1}{\sqrt{n}}\frac{\partial^2\tilde{L}_n(\phi)}{\partial\phi\partial\phi'}\| \xrightarrow{a.s.} 0.
\end{aligned}$$

This completes the proof. □

Proof of Theorem 1. (i) By Theorem 4.1.1 and the associated in Amemiya (1985), we have $\bar{\phi}_n\stackrel{a.s}{\to}\phi^0$ if the following conditions hold

(a) Θ is a compact parameter space;

(b) $L_n(\phi, X)$ is continuous in $\phi \in \Theta$ for all X and is a measurable function of X for all $\phi \in \Theta$;

(c) $\frac{1}{n}L_n(\phi) \xrightarrow{a.s} L(\phi)$ uniformly in $\phi \in \Theta$;

(d) $L(\phi)$ attains a unique global maximum at ϕ^0.

By Assumption A2, Lemma 4, and Lemma 5, we know conditions (a)-(d) are satisfied. Thus, $\bar{\phi}_n \xrightarrow{a.s} \phi^0$.

(ii) By the mean value theorem, we obtain that

$$\frac{\partial L_n(\bar{\phi}_n)}{\partial \phi} = \frac{\partial L_n(\phi^0)}{\partial \phi} + \frac{\partial^2 L_n(\xi)}{\partial \phi \partial \phi'}(\bar{\phi}_n - \phi^0),$$

where ξ lies between ϕ^0 and $\bar{\phi}_n$. But, $\frac{\partial L_n(\bar{\phi}_n)}{\partial \phi} = 0$. Then

$$\frac{\partial^2 L_n(\xi)}{\partial \phi \partial \phi'}(\bar{\phi}_n - \phi^0) = -\frac{\partial L_n(\phi^0)}{\partial \phi}.$$

Using Lemma 3 and the continuity of $\frac{\partial^2 L_n(\phi)}{\partial \phi \partial \phi'}$, we obtain

$$(\Lambda_0 + o(1))(\bar{\phi}_n - \phi^0) = -\frac{1}{n}\frac{\partial L_n(\phi^0)}{\partial \phi}.$$

It can be easily verified that

$$\frac{\partial l_t(\phi^0)}{\partial \phi} = \frac{1}{2}(1 - \varepsilon_t^2)\frac{1}{\sigma_t^2(\phi^0)}\frac{\partial \sigma_t^2(\phi^0)}{\partial \phi}$$

is a stationary sequence of martingale differences. Therefore, by applying a central limit theorem of martingale (Hall and Hegde (1980)), we obtain

$$\sqrt{n}(\bar{\phi}_n - \phi^0) \xrightarrow{\mathcal{L}} N(0, \Lambda_0^{-1}\Omega_0\Lambda_0^{-1}).$$

Proof of Theorem 2. (i) By Lemma 5 and Lemma 6, we have

$$\sup_{\phi \in \Theta} |\frac{1}{n}\tilde{L}_n(\phi) - L(\phi)| \leq \sup_{\phi \in \Theta} |\frac{1}{n}L_n(\phi) - \frac{1}{n}\tilde{L}_n(\phi)| + \sup_{\phi \in \Theta} |\frac{1}{n}L_n(\phi) - L(\phi)| \xrightarrow{a.s.} 0.$$

Imitating the proof of Theorem 3 (i), we obtain the result.

(ii) Notice that

$$\begin{aligned} & \frac{1}{\sqrt{n}}\frac{\partial L_n(\bar{\phi}_n)}{\partial \phi} - \frac{1}{\sqrt{n}}\frac{\partial L_n(\tilde{\phi}_n)}{\partial \phi} \\ = \quad & [\frac{1}{\sqrt{n}}\frac{\partial L_n(\bar{\phi}_n)}{\partial \phi} - \frac{1}{\sqrt{n}}\frac{\partial \tilde{L}_n(\tilde{\phi}_n)}{\partial \phi}] + [\frac{1}{\sqrt{n}}\frac{\partial \tilde{L}_n(\tilde{\phi}_n)}{\partial \phi} - \frac{1}{\sqrt{n}}\frac{\partial L_n(\tilde{\phi}_n)}{\partial \phi}]. \end{aligned}$$

From Lemma 5 and the mean value theorem, we have

$$\frac{1}{n}\frac{\partial^2 L_n(\xi^*)}{\partial \phi \partial \phi'}\sqrt{n}(\tilde{\phi}_n - \bar{\phi}_n) \xrightarrow{a.s.} 0.$$

Thus, $\sqrt{n}(\tilde{\phi}_n - \bar{\phi}_n) = o(1)$ by Lemma 3 and the continuity of $\frac{\partial^2 L_n(\phi)}{\partial \phi \partial \phi'}$. Then the result follows from Theorem 1.

(iii)By Lemma 3 and Lemma 6 and Theorem 2 (ii), we have

$$E \sup_{\phi \in \Theta} |\tilde{\Lambda}(\phi) - \Lambda(\phi)| \stackrel{a.s}{\to} 0, \ \ and \ \ [\Lambda(\tilde{\phi}_n) - \Lambda(\phi^0)] = O_p(1)(\tilde{\phi}_n - \phi^0) \to 0.$$

Therefore,

$$\tilde{\Lambda}(\tilde{\phi}_n) = [\tilde{\Lambda}(\tilde{\phi}_n) - \Lambda(\tilde{\phi}_n)] + [\Lambda(\tilde{\phi}_n) - \Lambda(\phi^0)] + \Lambda(\phi^0) \to \Lambda_0.$$

by applying an ergodic theorem to $\Lambda(\phi^0)$. This completes the proof of Theorem 2.

Proof of Theorem 3. Define

$$\begin{aligned} T_n(v) &= \sum_{t=u}^{n}(|Q_t(\phi^0 + n^{-1/2}v)| - |Q_t(\phi^0)|), \\ T_n^*(v) &= \sum_{t=u}^{n}(|Q_t(\phi^0) - n^{-1/2}v'D_t| - |Q_t(\phi^0)|), \end{aligned}$$

where $D_t \equiv D_t(\phi^0)$. By Lemma 1 and the same argument as Lemma 6, we obtain that

$$T_n(v) - \tilde{T}_n(v) \xrightarrow{\mathrm{P}} 0, \tag{5.9}$$

uniformly on compact sets. Using the equality

$$|z-y| - |z| = -ysgn(z) + 2(y-z)\{I(0<z<y) - I(y<z<0)\}, z \neq 0, \tag{5.10}$$

we have

$$
\begin{aligned}
T_n^*(v) &= n^{-1/2}\sum_{t=u}^{n} v'D_t sgn(Q_t(\phi^0)) \\
&\quad +2\sum_{t=u}^{n}(n^{-1/2}v'D_t - Q_t(\phi^0))[I(0<Q_t(\phi^0)<n^{-1/2}v'D_t) - I(n^{-1/2}v'D_t < Q_t(\phi^0) < 0)] \\
&=: A_n + B_n.
\end{aligned}
$$

Since $Q_t(\phi^0) = \log|\varepsilon_t|$, we know $\{v'D_t sgn(\log|\varepsilon_t|)\}$ is a stationary sequence of martingale differences by assumption A4 and Lemma 2. Therefore, applying a martingale central limit theorem (Hall and Heyde (1980)), we obtain $A_n \to_{\mathcal{L}} v'N$, where N denotes a $N(0,\Sigma)$ random vector.

Now turning to B_n, let

$$U_{nt} = (n^{-1/2}v'D_t - \log|\varepsilon_t|)I(0<\log|\varepsilon_t|<n^{-1/2}v'D_t).$$

Then

$$
\begin{aligned}
nEU_{nt}^2 &= nE(E(U_{nt}^2|\mathcal{F}_{t-1})) \\
&= nE\big(\int_0^{n^{-1/2}v'D_t}(n^{-1/2}v'D_t - x)^2 f(x)dx\big) \\
&= nE[\int_0^{n^{-1/2}v'D_t}(n^{-1/2}v'D_t - x)^2(f(x)-f(0))dx \\
&\quad +\int_0^{n^{-1/2}v'D_t}(n^{-1/2}v'D_t - x)^2 f(0)dx]) \\
&\le E\big(B_2 n^{-2}(v'D_t)^4 + B_1 n^{-3/2}(v'D_t)^3\big).
\end{aligned}
$$

By Lemma 2, we have $E(v'D_t)^4 < \infty$ and $E(v'D_t)^3 < \infty$. Therefore, we have proved that

$$\limsup_{n\to\infty} nEU_{nt}^2 = 0. \tag{5.11}$$

On the other hand, on the set $\{D_t'v > 0\}$, we may show that

$$\sum_{t=u}^{n} E(U_{nt}|\mathcal{F}_{t-1}) \to \frac{f(0)}{2}E[(v'D_t)^2 I(v'D_t>0)],$$

and

$$Var\big(\sum_{t=u}^{n}(U_{nt} - E(U_{nt}|\mathcal{F}_{t-1}))\big) \to 0.$$

Therefore,

$$\sum_{t=u}^{n} U_{nt} \to \frac{f(0)}{2} E[(v'D_t)^2 I(v'D_t > 0)].$$

Using the same argument for the second indicator in the summands of B_n, we obtain that

$$B_n \xrightarrow{P} f(0)v'\Sigma v. \tag{5.12}$$

Let $T = f(0)v'\Sigma v + v'N$, then the finite dimensional distributions of T_n^* converge to those of T. But, since T_n^* has convex sample paths, this implies that the convergence is in fact on $C(R^d)$ (see the proof of Proposition 1 in Davis and Dunsmuir (1997)).

Denote $H_t(\phi) = \frac{\partial^2 Q_t(\phi)}{\partial\phi\partial\phi'}$, then we have $E\|\sup_{\phi\in\Theta} H_t(\phi)\| < \infty$ from Lemma 2. By Taylor expansion and (5.10), it follows that

$$\begin{aligned}
T_n(v) - T_n^*(v) &= \sum_{t=u}^{n} \left[|Q_t(\phi^0) - n^{-1/2}v'D_t - n^{-1}v'H_t(\phi^*)v| - |Q_t(\phi^0) - n^{-1/2}v'D_t|\right] \\
= \quad & -\sum_{t=u}^{n} n^{-1}v'H_t(\phi^*)v sgn\big(\log|\varepsilon_t| - n^{-1/2}v'D_t\big) \\
& +2\sum_{t=u}^{n}(n^{-1}v'H_t(\phi^*)v - \log|\varepsilon_t| + n^{-1/2}v'D_t)I\big(0 < \log|\varepsilon_t| - n^{-1/2}v'D_t < n^{-1}v'H_t(\phi^*)v\big) \\
& +2\sum_{t=u}^{n}(n^{-1}v'H_t(\phi^*)v - \log|\varepsilon_t| + n^{-1/2}v'D_t)I\big(n^{-1}v'H_t(\phi^*)v < \log|\varepsilon_t| - n^{-1/2}v'D_t < 0\big).
\end{aligned}$$

By a similar argument for $S_n^{**}(v) - S_n^*(v) \to 0$ in Pan et al. (2005), we can obtain that $T_n(v) - T_n^*(v) \xrightarrow{P} 0$ uniformly on compact sets, which implies $\tilde{T}_n(v) \xrightarrow{\mathcal{L}} T$ on $C(R^d)$ from (5.9). By the proof of Theorem 1 in Pan et.al. (2005), we obtain the result.

Proof of Theorem 4. Based on Theorem 3, Theorem 4 follows immediately from the following two assertions

$$\hat{\Sigma} \xrightarrow{P} \Sigma \quad \text{and} \quad \hat{f}(0) \xrightarrow{P} f(0). \tag{5.13}$$

For the first assertion, defining

$$\Sigma_n(\phi) = \frac{1}{n}\sum_{t=1}^{n} D_t(\phi)D_t(\phi)' \quad and \quad \tilde{\Sigma}_n(\phi) = \frac{1}{n}\sum_{t=1}^{n} \tilde{D}_t(\phi)\tilde{D}_t(\phi)',$$

we have

$$\hat{\Sigma} = [\tilde{\Sigma}_n(\hat{\phi}_n) - \tilde{\Sigma}_n(\phi^0)] + [\tilde{\Sigma}_n(\phi^0) - \Sigma_n(\phi^0)] + \Sigma_n(\phi^0) =: \Sigma_{1n} + \Sigma_{2n} + \Sigma_{3n}.$$

Obviously, $\Sigma_{3n} \stackrel{a.s}{\to} \sigma$ by the ergodic theorem. But,

$$\begin{aligned}
\Sigma_{1n} &= \frac{1}{n}\sum_{t=1}^{n}[\tilde{D}_t(\hat{\phi}_n) - \tilde{D}_t(\phi^0)]\tilde{D}_t(\hat{\phi}_n)' + \frac{1}{n}\sum_{t=1}^{n}\tilde{D}_t(\phi^0)[\tilde{D}_t(\hat{\phi}_n) - \tilde{D}_t(\phi^0)]' \\
&\leq \|\hat{\phi}_n - \phi_0\|\frac{1}{n}\sum_{t=1}^{n}\{\|\frac{\partial \tilde{D}_t(\phi^*)}{\partial \phi'}\|\|\tilde{D}_t(\hat{\phi}_n)\| + \|\frac{\partial \tilde{D}_t(\phi^{**})}{\partial \phi'}\|\|\tilde{D}_t(\phi^0)\|\} \\
&\leq \|\hat{\phi}_n - \phi_0\|\frac{2}{n}\sum_{t=1}^{n}\sup_{\phi\in\Theta}\|\frac{\partial \tilde{D}_t(\phi)}{\partial \phi'}\|\sup_{\phi\in\Theta}\|\tilde{D}_t(\phi)\| \stackrel{a.s}{\to} 0,
\end{aligned}$$

by Lemma 2 and Theorem 3. In the above expression, ϕ^* and ϕ^{**} lie on the line from $\hat{\phi}_n$ to ϕ^0. Using Lemma 1, by a similar argument leading to Lemma 6, we can conclude that $\Sigma_{2n} \stackrel{a.s}{\to} 0$. The proof of the first assertion is completed.

For the second assertion, defining

$$f^*(0) = \frac{1}{nb_n}\sum_{t=1}^{n}K\big(\frac{1}{b_n}\log\frac{|X_t|}{\sigma_t(\hat{\phi}_n)}\big),$$

we have

$$\begin{aligned}
|\hat{f}(0) - f^*(0)| &\leq \frac{C}{2nb_n^2\hat{\delta}}\sum_{t=1}^{n}|\log\sigma_t^{2\hat{\delta}}(\hat{\phi}_n) - \log\tilde{\sigma}_t^{2\hat{\delta}}(\hat{\phi}_n)| \\
&= \frac{C}{2nb_n^2\hat{\delta}}\sum_{t=1}^{n}\log\big(1 + \frac{\sigma_t^{2\hat{\delta}}(\hat{\phi}_n) - \log\tilde{\sigma}_t^{2\hat{\delta}}(\hat{\phi}_n)}{\tilde{\sigma}_t^{2\hat{\delta}}(\hat{\phi}_n)}\big) \\
&\leq C\frac{1}{nb_n^2}\sum_{t=1}^{n}|\sigma_t^{2\hat{\delta}}(\hat{\phi}_n) - \tilde{\sigma}_t^{2\hat{\delta}}(\hat{\phi}_n)| \\
&\leq C\frac{1}{nb_n^2}\sum_{t=1}^{n}\sup_{\phi\in\Theta}|\sigma_t^{2\delta}(\phi) - \tilde{\sigma}_t^{2\delta}(\phi)| \stackrel{a.s}{\to} 0,
\end{aligned}$$

provided $nb_n^2 \to \infty$, by a proof similar to that of Lemma 6. Notice that

$$\begin{aligned}
|f^*(0) - f(0)| &\leq \frac{1}{nb_n}\sum_{t=1}^{n}|K\big(\frac{1}{b_n}\log\frac{|X_t|}{\sigma_t(\hat{\phi}_n)}\big) - K\big(\frac{\log|\varepsilon_t|}{b_n}\big)| + |\frac{1}{nb_n}\sum_{t=1}^{n}K\big(\frac{\log|\varepsilon_t|}{b_n}\big) - f(0)| \\
&= L_1 + L_2.
\end{aligned}$$

Due to Theorem 3 (i), it follows from Lemma 2 and Theorem 3 that

$$
\begin{aligned}
L_1 &\leq \frac{C}{nb_n^2}\sum_{t=1}^{n}|\log\sigma_t^2(\hat{\phi})-\log\sigma_t^2(\phi^0)| \\
&= \frac{C}{nb_n^2}\|\hat{\phi_n}-\phi^0\|\sum_{t=1}^{n}\|\frac{1}{\sigma_t^2(\phi^*)}\frac{\partial\sigma_t^2(\phi^*)}{\partial\phi}\| \\
&= \frac{C}{\sqrt{n}b_n^2}\sqrt{n}\|\hat{\phi_n}-\phi^0\|\cdot\frac{1}{n}\sum_{t=1}^{n}\|\frac{1}{\sigma_t^2(\phi^*)}\frac{\partial\sigma_t^2(\phi^*)}{\partial\phi}\| \\
&= O(1)\frac{1}{\sqrt{n}b_n^2}\rightarrow 0.
\end{aligned}
$$

On the other hand, since

$$
\begin{aligned}
E\{\frac{1}{b_n}K(\frac{\log|\varepsilon_t|}{b_n})\} &= \int_{-\infty}^{\infty}K(x)f(b_nx)dx \\
= \int_{-\infty}^{\infty}K(x)f(0)dx+\int_{-\infty}^{\infty}K(x)[f(b_nx)-f(0)]dx &= f(0)+o_p(1),
\end{aligned}
$$

it follows that

$$
\begin{aligned}
E(K_2)^2 &= \frac{1}{n^2}\sum_{t=1}^{n}E[\frac{1}{b_n}K(\frac{\log|\varepsilon_t|}{b_n})-f(0)]^2 \\
&\quad+\frac{2}{n^2}\sum_{1\leq i<j\leq n}E[(\frac{1}{b_n}K(\frac{\log|\varepsilon_i|}{b_n})-f(0))(\frac{1}{b_n}K(\frac{\log|\varepsilon_j|}{b_n})-f(0))] \\
&\leq \frac{2}{nb_n}\int_{-\infty}^{\infty}K^2(x)f(b_nx)dx+\frac{2}{n}f^2(0) \\
&\quad+\frac{2}{n^2}\sum_{1\leq i<j\leq n}E[\frac{1}{b_n}K(\frac{\log|\varepsilon_i|}{b_n})-f(0)]E[\frac{1}{b_n}K(\frac{\log|\varepsilon_j|}{b_n})-f(0)]\longrightarrow 0
\end{aligned}
$$

by Assumption A4. Then, $K_2\xrightarrow{P}0$. This completes the proof of the second assertion.

A Appendix. Probabilistic Properties of PTTGARCH Model

A.1 The Stochastic Recurence Form of The Proposed Model

Model (1.2) can be represented in a form of stochastic recurrence equation. We can always assume that $p\geq 2, q\geq 2$ because, otherwise, we can add some α_{1i},α_{2i} or β_i which is equal to 0. Denote

$$
\eta_t=\beta_1+\alpha_{11}(\varepsilon_t^+)^{2\delta}+\alpha_{21}(\varepsilon_t^-)^{2\delta},
$$

$$Y_t = (\sigma_{t+1}^{2\delta}, \cdots, \sigma_{t-q+2}^{2\delta}, (X_t^+)^{2\delta}, (X_t^-)^{2\delta}, \cdots, (X_{t-p+2}^+)^{2\delta}, (X_{t-p+2}^-)^{2\delta})' \in R^\kappa \tag{A.1}$$

$$B \equiv B(\phi) = (\alpha_0, 0, \cdots, 0)' \in R^\kappa$$

$$A_t \equiv A_t(\phi) = \begin{pmatrix} \eta_t & \beta_2 & \cdots & \beta_{q-1} & \beta_q & \alpha_{12} & \alpha_{22} & \cdots & \alpha_{1p-1} & \alpha_{2p-1} & \alpha_{1p} & \alpha_{2p} \\ 1 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ \varepsilon_t^{+2\delta} & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ \varepsilon_t^{-2\delta} & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 & 1 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 & 0 & 1 & \cdots & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 1 & 0 & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & 1 & 0 & 0 \end{pmatrix}_{\kappa \times \kappa} \tag{A.2}$$

with $\kappa = 2p + q - 2$ and $\phi = (\delta, \alpha_0, \alpha_{11}, \alpha_{21}, \cdots, \alpha_{1p}, \alpha_{2p}, \beta_1, \cdots, \beta_q)'$. Then, X_t is a solution of (1.2) if and only if Y_t is a solution of the following equation

$$Y_t = A_t Y_{t-1} + B. \tag{A.3}$$

A.2 Strict Stationarity of The Model

Choose a norm $\|\cdot\|$ on R^κ, say, $\|x\| = |x_1| + \cdots + |x_d|$, for any $x \in R^\kappa$. Then for any $\kappa \times \kappa$ matrix A, the corresponding operator norm is

$$\|A\| = \sup_{\|x\|=1} \|Ax\|.$$

The top Lyapunov exponent associated with the sequence $\{A_t\}$ given by (A.2) is defined as

$$\gamma(\phi) = \inf\{E(\frac{1}{t+1} \log \|A_0 A_{-1} \cdots A_{-t}\|), t \in N\},$$

provided $E \log^+ \|A_0\|$ is finite, which is satisfied under assumption A1.

Furthermore, under assumption A1, we have

$$\gamma(\phi) = \lim_{t\to+\infty} \frac{1}{t} \log \|A_0 A_{-1} \cdots A_{-t}\|. \tag{A.4}$$

This enables us to approximate the value of the top Lyapunov exponent numerically by simulation.

Theorem 5. *Under assumption A1, the following assertions hold:*

(i) There is a unique strictly stationary solution to model (1.2) if and only if the top Lyapunov exponent $\gamma(\phi)$ given by (A.4) is strictly negative. Moreover, this stationary solution is ergodic.

(ii) If there is a strictly stationary solution to model (1.2), then $\sum_{j=1}^{q} \beta_j < 1$.

Proof. (i) Our proof is similar to that of Bougerol and Picard (1992a). Firstly, we prove that $E \log^+ \|A_0\| < +\infty$, which ensures that the Lyapunov exponent $\gamma(\phi)$ is well defined. Due to the equivalence of norms of A_0, we have

$$\begin{aligned} \|A_0\| &\leq C\{\beta_1 + \alpha_{11}\varepsilon_t^{+2\delta} + \alpha_{21}\varepsilon_t^{-2\delta} + \sum_{j=2}^{q} \beta_j + \sum_{i=2}^{p} (\alpha_{1i} + \alpha_{2i}) + q - 1 + |\varepsilon_t|^{2\delta} + 2p - 4\} \\ &\leq C(1 + |\varepsilon_t|^{2\delta}). \end{aligned}$$

For $\tilde{\Delta} = \min\{1, \frac{\Delta}{2\delta}\}$, there exists an $\tilde{M} > 0$ such that $\log^+ x < x^{\tilde{\Delta}}$, for $x > \tilde{M}$. Then, by assumption A1,

$$E \log^+ \|A_0\| \leq \tilde{M} + C(1 + E|\varepsilon_t|^{\Delta}) < +\infty. \tag{A.5}$$

Necessity. Suppose $\{X_t, t \in Z\}$ is a strictly stationary solution of (1.2). Then $\{Y_t, t \in Z\}$ defined by (A.1) is a strictly stationary solution of (A.3). From (A.3), we have, for all $t > 0$

$$\begin{aligned} Y_0 &= A_0 Y_{-1} + B \\ &= A_0 A_{-1} Y_{-2} + B + A_0 B \\ &= \cdots \\ &= A_0 A_{-1} \cdots A_{-t} Y_{-t-1} + B + \sum_{k=0}^{t-1} A_0 \cdots A_{-k} B. \end{aligned}$$

Then for any $t > 0$, $\sum_{k=0}^{t-1} A_0 \cdots A_{-k} B \leq Y_0$ by the nonnegativity of all elements of A_t, Y_t and B. This indicates that $\sum_{k=0}^{t-1} A_0 \cdots A_{-t} B$ converges almost surely, as $t \to \infty$. Therefore,

$A_0 \cdots A_{-t} B \stackrel{a.s}{\to} 0$ as $t \to \infty$. By Lemma 2.1 of Bougerol and Picard (1992a), it is sufficient to prove $A_0 \cdots A_{-t} \stackrel{a.s}{\to} 0$. Let $\{e_i\}$ be the canonical basis of $\mathbb{R}^\kappa$, and we only need to prove that

$$\lim_{t \to \infty} A_0 \cdots A_{-t} e_i = 0, \tag{A.6}$$

$i = 1, \cdots, \kappa$. Since $B = \delta e_1$ and $\delta > 0$, (A.6) holds for $i = 1$. Notice that

$$\begin{aligned}
&A_{-t} e_q = \beta_q e_1, \\
&A_{-t} e_{j-1} = \beta_{j-1} e_1 + e_j, \quad 2 < j \le q \\
&A_{-t} e_\kappa = \alpha_{2p} e_1, \\
&A_{-t} e_{\kappa-1} = \alpha_{1p} e_1, \\
&A_{-t} e_{q+2j} = \alpha_{2,j+1} e_1 + e_{q+2j+1}, \quad 1 \le j \le p-2 \\
&A_{-t} e_{q+2j-1} = \alpha_{1,j+1} e_1 + e_{q+2j}, \quad 1 \le j \le p-2,
\end{aligned}$$

we obtain (A.6) holds for $1 \le i \le \kappa$.

Sufficiency. Assume $\gamma(\phi) < 0$. Then (A.4) implies that the series $\sum_{k=0}^{\infty} A_t \cdots A_{t-k} B$ converges almost surely for all t. Define $\{Y_t, t \in Z\}$ as follows

$$Y_t = B + \sum_{k=0}^{\infty} A_t \cdots A_{t-k} B. \tag{A.7}$$

It is easy to verify that $\{Y_t, t \in Z\}$ is a nonnegative solution of (A.3). Let $\sigma_t = (Y_{t-1}^1)^{\frac{1}{2\delta}}$, where Y_{t-1}^1 is the first component of Y_{t-1}. Then $X_t = \sigma_t \varepsilon_t$ is a solution of model (1.2). The strict stationarity and ergodicity of X_t can be derived by noticing that $\{(A_t, \varepsilon_t), t \in Z\}$ is strictly stationary and ergodic. In the following, we will prove Y_t is the unique solution of (A.3). Suppose $\tilde{Y}_t$ is another strictly stationary solution of (A.3), then we have

$$\|Y_0 - \tilde{Y}_0\| = \|A_0 \cdots A_{-t}(Y_{-t-1} - \tilde{Y}_{-t-1})\| \le \|A_0 \cdots A_{-t}\| \|Y_{-t-1} - \tilde{Y}_{-t-1}\| \stackrel{P}{\to} 0$$

So $Y_0 = \tilde{Y}_0$ a.s. This shows that (1.2) has a unique stationary solution and the proof of (i) is completed.

(ii) Let $\tilde{A}$ denote the matrix which is obtained by replacing $(\varepsilon_t^+)^{2\delta}$ and $(\varepsilon_t^-)^{2\delta}$ with 0 in the matrices A_t defined by (A.2). Using the notation in Bougerol and Picard (1992a), we have

$A_0A_1\cdots A_{-t} \geq \tilde{A}^{t+1}$. By (A.4), we have $\|\tilde{A}\|^{t+1} \leq \|A_0A_1\cdots A_{-t}\| \to 0$, which implies that the largest eigenvalue of $\tilde{A}$ is less than 1. On the other hand, it is easily seen that

$$Det|\lambda I_\kappa - \tilde{A}| = \lambda^\kappa(1 - \sum_{i=1}^{q} \beta_i \lambda^{-i})$$

Since the function $g(x) = 1 - \sum_{i=1}^{q} \beta_i x^i$ has no zero point on $x \in [0,1]$ and $g(0) = 1$, we obtain that $g(1) = 1 - \sum_{i=1}^{q} \beta_i > 0$.

The proof of Theorem 5 is completed. □

Remark A1. For the PTTGARCH(1,1) process, Hwang and Basawa (2004) pointed out that $E[\log(\beta_1 + \alpha_{11}(\varepsilon_t^+)^{2\delta} + \alpha_{21}(\varepsilon_t^-)^{2\delta})] < 0$ could ensure that the model has a unique strictly stationary solution. In fact, it is easily verified that $\gamma(\phi) = E[\log(\beta_1 + \alpha_{11}(\varepsilon_t^+)^{2\delta} + \alpha_{21}(\varepsilon_t^-)^{2\delta})]$ provided $p = q = 1$, which means by Theorem 5 (i) that $E[\log(\beta_1 + \alpha_{11}(\varepsilon_t^+)^{2\delta} + \alpha_{21}(\varepsilon_t^-)^{2\delta})] < 0$ is also a necessary condition for PTTGARCH(1,1) model to define a unique strictly stationary solution, see also Theorem 2.1 of Liu (2006).

A.3 The existence of Moments

Theorem 6. *(i) If $\{X_t\}$ is a strictly stationary solution of model (1.2) and assumption A1 holds, then there exists a constant $\tau > 0$ such that*

$$E|X_t|^\tau < +\infty.$$

(ii) Suppose $E|\varepsilon_t|^{2\delta} < +\infty$. Then, Model (1.2) has a stationary solution with $E|X_t|^{2\delta} < +\infty$ if and only if

$$\sum_{i=1}^{p}[\alpha_{1i}E(\varepsilon_t^+)^{2\delta} + \alpha_{2i}E(\varepsilon_t^-)^{2\delta}] + \sum_{j=1}^{q}\beta_j < 1. \tag{A.8}$$

(iii) For any $k \in \mathbb{N}$, if $E|\varepsilon_t|^{2k\delta} < +\infty$ and $E(\|A_t\|^k) < 1$, we have $E|X_t|^{2k\delta} < \infty$. Furthermore, if we assume in addition that

$$\sum_{i=1}^{p}[\alpha_{1i}(E(\varepsilon_t^+)^{2k\delta})^{\frac{1}{k}} + \alpha_{2i}(E(\varepsilon_t^-)^{2k\delta})^{\frac{1}{k}}] + \sum_{j=1}^{q}\beta_j < 1, \tag{A.9}$$

then

$$E|X_t|^{2k\delta} \le \alpha_0^k\{1 - \sum_{i=1}^{p}[\alpha_{1i}(E(\varepsilon_t^+)^{2k\delta})^{\frac{1}{k}} + \alpha_{2i}(E(\varepsilon_t^-)^{2k\delta})^{\frac{1}{k}}] - \sum_{j=1}^{q}\beta_j\}^{-k}E|\varepsilon_t|^{2k\delta}.$$

Proof. (i) By the definition of $\gamma(\phi)$, there exists an integer $m \ge 1$ such that $E\log\|A_0A_1\cdots A_{m-1}\| < 0$. From the proof of Theorem 5 (i), we know $E\|A_0\|^{\tilde{\Delta}} \le C < +\infty$. Therefore,

$$E\|A_0A_1\cdots A_{m-1}\|^{\tilde{\Delta}} \le (E\|A_0\|^{\tilde{\Delta}})^m < +\infty.$$

We introduce a function $h(x) = E\|A_0A_1\cdots A_{m-1}\|^x$, $0 < x \le \tilde{\Delta}$. Since $h'(0) = E\log\|A_0A_1\cdots A_{m-1}\| < 0$, $h(x)$ decreases in neighborhood of 0. Notice that $h(0) = 1$, so there exists a Δ^* such that $0 < \Delta^* < \min\{\tilde{\Delta}, 1\}$ such that

$$E\|A_0A_1\cdots A_{m-1}\|^{\Delta^*} < 1. \tag{A.10}$$

Using (A.7), we obtain that

$$\|Y_0\|^{\Delta^*} \le \|B\|^{\Delta^*} + \sum_{k=0}^{\infty}\|A_0A_1\cdots A_{-k}\|^{\Delta^*}\|B\|^{\Delta^*}.$$

By (A.10), it follows easily that there exist $0 < C_0 < \infty$ and $0 < \rho < 1$ such that $E\|A_0A_1\cdots A_{-k}\|^{\tau} \le C_0\rho^k$ for any k, which implies that $E\|Y_0\|^{\Delta^*} < \infty$. Let $\tau = (2\delta)\Delta^*$ and we complete the proof of (i).

(ii) *Necessity.* Suppose that $\{X_t, t \in Z\}$ is a strictly solution of (1.2) with $E|X_t|^{2\delta} < \infty$. Then $E\sigma_t^{2\delta} < +\infty$. Hence we have

$$E\sigma_t^{2\delta} = \alpha_0 + \sum_{i=1}^{p}(\alpha_{1i}E\sigma_{t-i}^{2\delta}E(\varepsilon_t^+)^{2\delta} + \alpha_{2i}E\sigma_{t-i}^{2\delta}E(\varepsilon_t^-)^{2\delta}) + \sum_{j=1}^{q}\beta_jE\sigma_{t-j}^{2\delta}.$$

Because $\{\sigma_t^{2\delta}\}$ is stationary, then

$$(1 - \sum_{i=1}^{p}(\alpha_{1i}E(\varepsilon_t^+)^{2\delta} + \alpha_{2i}E(\varepsilon_t^-)^{2\delta}) - \sum_{j=1}^{q}\beta_j)E\sigma_t^{2\delta} = \alpha_0.$$

Therefore,

$$1 - \sum_{i=1}^{p}(\alpha_{1i}E(\varepsilon_t^+)^{2\delta} + \alpha_{2i}E(\varepsilon_t^-)^{2\delta}) - \sum_{j=1}^{q}\beta_j > 0.$$

Sufficiency. Suppose (A.8) holds. Define

$$Y_t = B + \sum_{k=0}^{\infty} A_t \cdots A_{t-k} B, \quad \textit{and} \quad A = EA_0.$$

Notice that $EA_t \cdots A_{t-k} = A^{k+1}$, we only need to prove that $\rho(A) < 1$. In fact,

$$Det(\lambda I_\kappa - A) = \lambda^\kappa \{1 - \sum_{i=1}^{p} (\alpha_{1i} E(\varepsilon_t^+)^{2\delta} + \alpha_{2i} E(\varepsilon_t^-)^{2\delta}) \lambda^{-i} - \sum_{j=1}^{q} \beta_j \lambda^{-j}\}.$$

By (A.8), we have

$$|Det(\lambda I_\kappa - A)| \geq 1 - \sum_{i=1}^{p} (\alpha_{1i} E(\varepsilon_t^+)^{2\delta} + \alpha_{2i} E(\varepsilon_t^-)^{2\delta}) - \sum_{j=1}^{q} \beta_j > 0$$

provided $|\lambda| \geq 1$. Thus $\rho(A) < 1$ and $EY_t = B + \sum_{k=0}^{\infty} A^{k+1} B < \infty$, which implies that $E|X_t|^{2\delta} < \infty$.

(iii) Applying Minkowski's inequality to (A.7), we obtain

$$\begin{aligned}
\left(E\|Y_t\|^k\right)^{1/k} &\leq \|B\| + \sum_{i=0}^{\infty} \left[E\left(\|A_t \cdots A_{t-i} B\|^k\right)\right]^{1/k} \\
&\leq \|B\| + \|B\| \sum_{i=0}^{\infty} \left[E(\|A_t\|^k)\right]^{i/k} \\
&< +\infty,
\end{aligned}$$

since $E(\|A_t\|^k) < 1$. Therefore, we have $E|X_t|^{2k\delta} < \infty$. In the following, we assume (A.9) holds. It can be seen from (1.2) that $E|X_t|^{2k\delta} = E\sigma_t^{2k\delta} E|\varepsilon_t|^{2k\delta}$. But, by Minkowski inequality,

$$\begin{aligned}
(E\sigma_t^{2k\delta})^{\frac{1}{k}} &= \left(E\{\alpha_0 + \sum_{i=1}^{p} \alpha_{1i} (X_{t-i}^+)^{2\delta} + \sum_{i=1}^{p} \alpha_{2i} (X_{t-i}^-)^{2\delta} + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^{2\delta}\}^k\right)^{\frac{1}{k}} \\
&\leq \alpha_0 + [\sum_{i=1}^{p} \alpha_{1i} (E(\varepsilon_{t-i}^+)^{2k\delta})^{\frac{1}{k}} + \sum_{i=1}^{p} \alpha_{2i} (E(\varepsilon_{t-i}^-)^{2k\delta})^{\frac{1}{k}} + \sum_{j=1}^{q} \beta_j] (E\sigma_t^{2k\delta})^{\frac{1}{k}}.
\end{aligned}$$

Then,

$$E|X_t|^{2k\delta} \leq \alpha_0^k \{1 - \sum_{i=1}^{p} [\alpha_{1i} (E(\varepsilon_t^+)^{2k\delta})^{\frac{1}{k}} + \alpha_{2i} (E(\varepsilon_t^-)^{2k\delta})^{\frac{1}{k}}] - \sum_{j=1}^{q} \beta_j\}^{-k} E|\varepsilon_t|^{2k\delta}.$$

This completes the proof of (iii). □

Remark A2. Notice that $E(\|A_t\|^k) = E\big(\beta_1 + \alpha_{11}(\varepsilon_t^+)^{2\delta} + \alpha_{21}(\varepsilon_t^-)^{2\delta}\big)^k$ provided $p = q = 1$, the result of Hwang and Basawa (2004) Theorem 3 (i) is a special case of our result (iii).

A.4 The tail behavior of the model

The following theorem shows that under some regular conditions, the tail of PTTGATCH(p,q) model is Pareto-like, which indicates that light-tailed input may cause heavy-tailed output.

Theorem 7. *Assume that model (1.2) satisfies that $\alpha_0 > 0$, $\gamma(\phi) < 0$ and not all of the parameters α_{1i}, α_{2i}, and β_j vanish, $i = 1, \cdots, p; j = 1, \cdots, q$. If furthermore ε_t has a positive density on $\mathbb{R}$ such that $E|\varepsilon_t|^\xi < +\infty$ for some $\xi > 0$, then it follows that the limit*

$$\lim_{x \to +\infty} x^{2\kappa_0 \delta} P\big(X_1 > x\big)$$

exists and is positive, where $\kappa_0 = 2\kappa_1$ and κ_1 is the unique solution of

$$\lim_{n \to +\infty} \frac{1}{n} \log E\|A_n \cdots A_1\|^\kappa = 0.$$

Proof. See the proof of Corollary 3.5 of Basrak et.al (2002). □

Remark A3. Liu (2006) studies the tail behavior for PTTGARCH(1,1) model. In this simple case, the limit in Theorem 7 can be expressed explicitly, see Liu (2006).

References

[1] Adler, R. J., Feldaman, R. E. and Gallagher, C. (1997). In *A User's Guide to Heavy Tails: Statistical techniques for analyzing heavy tailed distribution and processes*, Ed. by Adler, R. J., Feldaman, R. E. and Taqqu, M, pp.133-158. Boston: Birkhauser.

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

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

[4] Billingsley, P. (1968). *Convergence of Probability Measures.* Wiley, New York.

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

[6] Bougerol, P. and Picard, N. (1992a). Stationarity of GARCH processes and of some nonnegative time series. *J. Econometrics.* 52, 115-127.

[7] Bougerol, P. and Picard, N. (1992b). Strict stationarity of generalized autoregressive processes. *Ann. Probab.* 20, 1714-1730.

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

[9] Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of variance of U.K. inflation. *Econometrica* 50, 987-1008.

[10] Engle, R. F.and Bollerslev, T. (1986). Modelling the persistence of conditional variance. *Econometric Reviews* 5, 1-50.

[11] Gouriéroux, C. (1997). *ARCH Models and Financial Applications.* Springer, New York.

[12] Hall, P. C. and Heyde, C. C. (1980). *Martingale Limit Theory and its Applications.* Academic Press, New York.

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

[14] Higgins, M. L. and Bera, A. K. (1992). A claas of nonliear ARCH models. *Internat. Econom. Rev.* 33, 137-158.

[15] Hsieh, D. A. (1989). Testing nonliear dependence in daily foreign exchange rates. *J. Business* 62, 339-368.

[16] Hwang, S. Y. and Basawa, I. V. (2004). Stationarity and moment structure for Box-Cox transformed threshold GARCH(1,1) processes. *Statistics & Probability Letters* 68, 209-220.

[17] Hwang, S. Y. and Kim, T. Y. (2004). Power transformation and threshold modeling for ARCH innovations with applications to tests for ARCH structure. *Stochastic Processes and Their Applications* 110, 295-314.

[18] Li, C. W. and Li, W. K. (1996). On a double-threshold auroressive heteroscedastic time series model. *J. Appl. Econometrics*, 11, 253-274.

[19] Ling, S. (2005). Self-weighted LAD estimation for infinite variance autoregressive models. *J. Roy. Statist. Soc.* B, 67, 381-393.

[20] Liu, J. (2006). On the tail behaviors of Box-Cox transformed threshold GARCH(1,1) processes. *Statistics & Probability Letters*, 76, 1323-1330.

[21] Mikosch, T. and Straumann, D. (2006).Stable limits of martingale transforms with application to the estimation of GARCH parameters. *The Annals of Statistics*, 34(1), 493-522.

[22] Mittnik, S. and Rachev, S. T. (2000). *Stable Paretian Models in Finance.* New York: Wiley.

[23] Pan, J. Z., Wang, H. and Yao, Q. (2005). Weighted least absolute deviations estimation for ARMA models with infinite variance. *Econometric Theory*, to appear.

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

[25] Rydberg, T. (2000). Realistic statistical modelling of financial data. *Int. Statist. Rev.* 68, 233-258. 25, 1805-1869.

[26] Scheinkman, J. A. and LeBaron, B. (1989). Nonliear dynamics and stock returns. *J. Business* 62, 311-337.

[27] Shepard, 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.* Chapman and Hall, London.

[28] Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis.* Chapman and Hall, London.

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

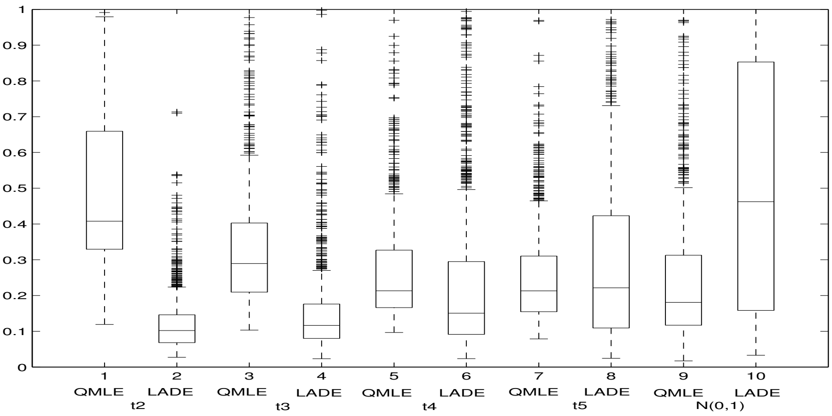

Figure 1: Boxplots of AAE for LADE and QMLE.

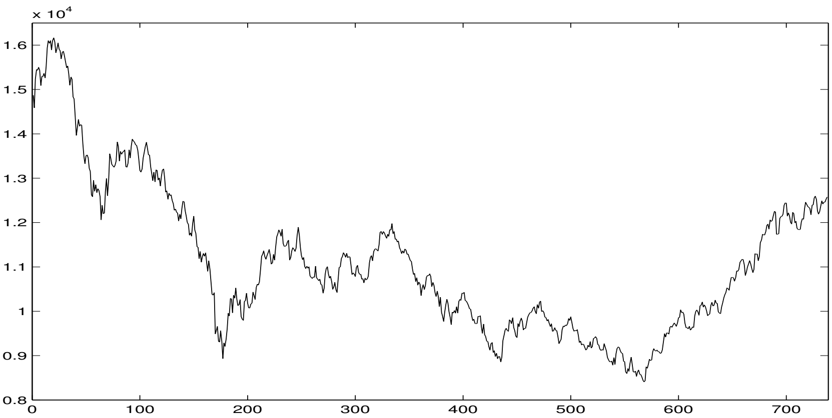

Figure 2: The time plot of the original HSI.

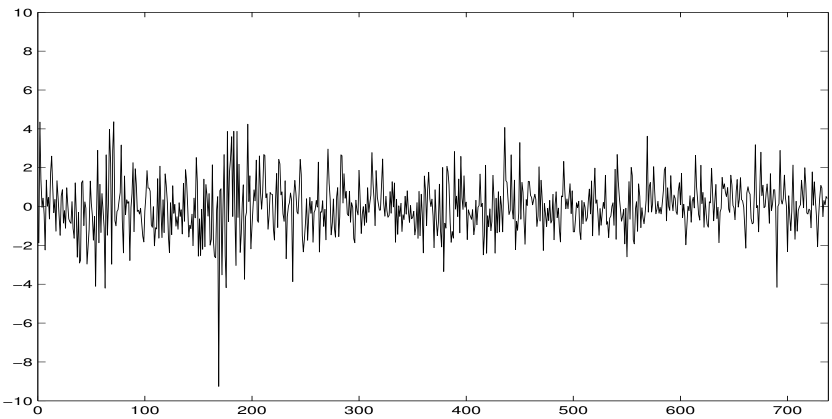

Figure 3: The time plot of the percentage of the log return of HSI $\{X_t\}$.

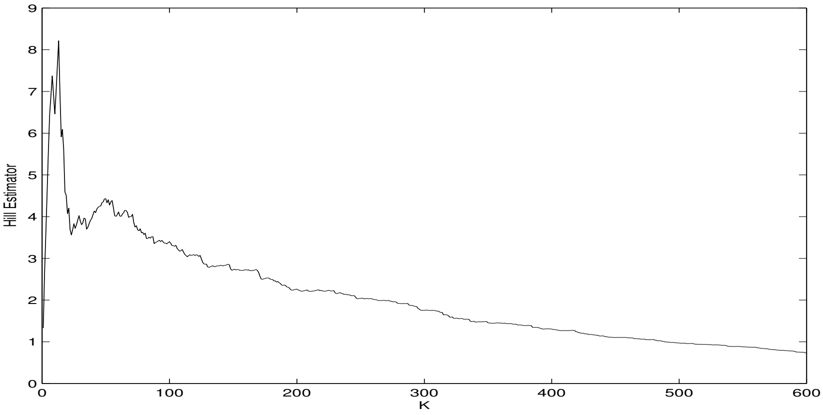

Figure 4: The Hill estimator of $\{X_t\}$.

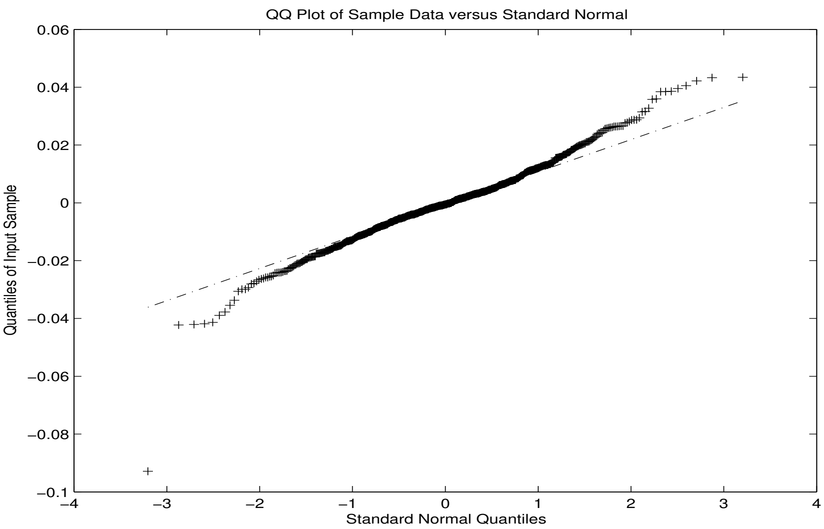

Figure 5: The QQ-plot of $\{X_t\}$.

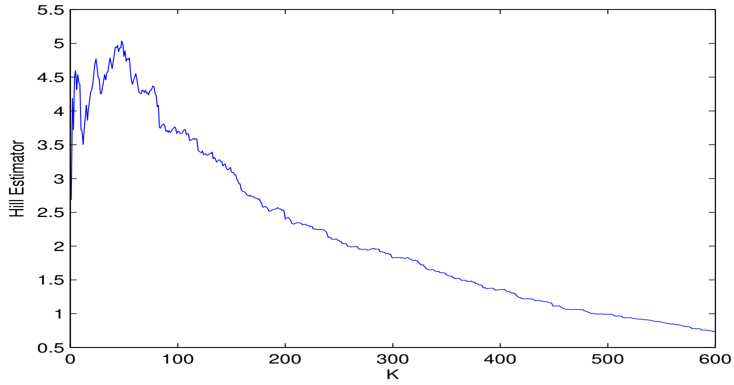

Figure 6: The Hill estimator of the standardized residuals

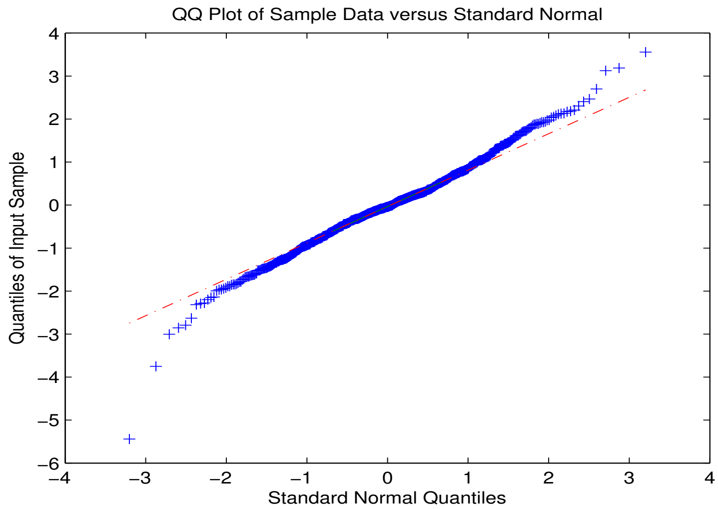

Figure 7: The QQ-plot of the standardized residuals