Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Estimation and tests for power-transformed and threshold GARCH models

Jiazhu Pan\textsuperscript{a,c,*}, Hui Wang\textsuperscript{b}, Howell Tong\textsuperscript{c}

\textsuperscript{a}School of Mathematical Sciences, Peking University, Beijing 100871, China
\textsuperscript{b}School of Finance, Central University of Finance and Economics, China
\textsuperscript{c}Department of Statistics, London School of Economics, Houghton Street, London WC2A 2AE, UK

Received 11 May 2006; received in revised form 3 March 2007; accepted 20 June 2007
Available online 18 July 2007

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. Stationarity and moment structure for Box–Cox transformed threshold GARCH\((1,1)\) processes. Statistics & Probability Letters 68, 209–220.] 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\((p,q)\) model, 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.

© 2007 Elsevier B.V. All rights reserved.

\textit{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

\textsuperscript{* Partially supported by an EPSRC research grant and the Natural Science Foundation of China (Grant 10471005).}
\textsuperscript{*Corresponding author. Department of Statistics, London School of Economics, Houghton Street, London, WC2A 2AE, UK. Tel.: +44 20 7955 7953.
E-mail address: j.pan@lse.ac.uk (J. Pan).}
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, \( \sigma_t^2 \), 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 Mikosch and Straumann (2006).

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; Hwang and Basawa, 2004). Among all the extensions, the functional form for \( \sigma_t^2 \) 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 data satisfactorily; Scheinkman and LeBaron (1989) found evidence that volatility in stock market data cannot be modeled using a GARCH model alone. He showed that the volatility is likely to depend on its past realizations. As a result, a threshold GARCH model was proposed by Engle and Bollerslev (1986) to model the volatility of financial time series data. This model is a generalization of the GARCH model, allowing the conditional variance to depend on the signs of the past shocks. The threshold GARCH model is only a function of the past \( X_{t-k}, k \geq 1 \) 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 by allowing \( \sigma_t \) to depend on \( \sigma_{t-1} \) 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 \( \sigma_t \) is a function of not only the past \( p \) observations but also the past \( q \) values of \( \sigma_t \) itself. A power-transformed and threshold GARCH\((p, q)\) model (PTTGGARCH\((p, q)\)) is defined as

\[
X_t = \sigma_t \varepsilon_t \quad \text{and} \quad \sigma_t^{2q} = \sigma_0 + \sum_{i=1}^{p} \alpha_i (X_{t-i}^+)^{2q} + \sum_{i=1}^{q} \alpha_i (X_{t-i}^-)^{2q},
\]

where \( \delta > 0, \sigma_0 > 0, \alpha_1 \geq 0, \alpha_2 \geq 0 \) are unknown parameters, \( i = 1, \ldots, p \). Here, \( \{\varepsilon_t\} \) is a sequence of independent and identically distributed random variables, and \( \varepsilon_t \) is independent of \( \{X_{t-k}, k \geq 1\} \) for all \( t \). 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 \( \sigma_t \) in Hwang and Kim’s (2004) 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 \( \sigma_t \) to depend on \( \sigma_{t-1} \) 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 \( \sigma_t \) is a function of not only the past \( p \) observations but also the past \( q \) values of \( \sigma_t \) itself. A power-transformed and threshold GARCH\((p, q)\) model (PTTGGARCH\((p, q)\)) is defined as

\[
X_t = \sigma_t \varepsilon_t \quad \text{and} \quad \sigma_t^{2q} = \sigma_0 + \sum_{i=1}^{p} \alpha_i (X_{t-i}^+)^{2q} + \sum_{i=1}^{q} \alpha_i (X_{t-i}^-)^{2q} + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^{2q},
\]
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 \( \text{PTTGARCH}(p, q) \) 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 \( \text{PTTGARCH}(p, q) \) 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 \( \text{PTTGARCH}(p, q) \) 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 Sections 2 and 3 are presented in Section 5. The Appendix presents the stationarity and existence of moments for \( \text{PTTGARCH}(p, q) \) model.

In the sequel, $\rightarrow_F$, $\overset{a.s.}{\rightarrow}$ and $\overset{d}{\rightarrow}$ 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 $a_{1p}$ or $a_{2p} > 0$, and $\beta_q > 0$ if $q > 0$. Let

$$\phi = (\delta, a_0, a_{11}, \ldots, a_{1p}, a_{2}, \beta_1, \ldots, \beta_q)'$$

be the parametric vector with true value $\phi^0 = (\delta^0, a_{1p}^0, a_{11}^0, \ldots, a_{12}^0, a_{2}^0, \beta_1^0, \ldots, \beta_q^0)'$. Define

$$\sigma_i(\phi) = \left[ x_0 + \sum_{i=1}^p a_i(X_{i-1}^+) 2^{\delta} + \sum_{i=1}^p a_{2i}(X_{i-1}^-) 2^{\delta} + \sum_{j=1}^q \beta_j \sigma_{i-j}^{2\delta}(\phi) \right]^{1/2 \delta}.$$

Our basic assumptions are as follows.

(A1) $\varepsilon_i$ is non-degenerate and symmetrically distributed. Furthermore, $E|\varepsilon_i|^4 < +\infty$ for some $A > 0$, and

$$\lim_{\tau \to 0} \tau^{-\mu} P[\varepsilon_i^2 \leq \tau] = 0 \quad \text{for some } \mu > 0. \quad (2.1)$$

(A2) $\Theta$ is a compact subset of $\mathbb{R}^d$, $d = 2p + q + 2 \phi_0$ is in the interior of $\Theta$, 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 $\Theta$, there exist positive constants $\delta_1, \delta_2, \rho_0$ such that $0 < \delta_1 < \delta, \delta_0 < \delta_2, \delta_1 \leq \sum_{i=1}^q \beta_i \leq \rho_0 < 1$ for any $\phi \in \Theta$. 

Under assumptions (A1) and (A2), it may be deduced that (1.2) implies that

$$
\sigma_t(\phi)^{2\delta} = \frac{\gamma_0}{1 - \sum_{j=1}^{q} \beta_j} + \sum_{i=1}^{p} \gamma_{1i}(X_{t-i}^+)^{2\delta} + \sum_{i=1}^{p} \gamma_{2i}(X_{t-i}^-)^{2\delta} \\
+ \sum_{i=1}^{p} \gamma_{1i} \sum_{k=1}^{\infty} \sum_{j_1=1}^{q} \cdots \sum_{j_k=1}^{q} \beta_{j_1} \cdots \beta_{j_k} (X_{t-i-j_1-\cdots-j_k}^+)^{2\delta} \\
+ \sum_{i=1}^{p} \gamma_{2i} \sum_{k=1}^{\infty} \sum_{j_1=1}^{q} \cdots \sum_{j_k=1}^{q} \beta_{j_1} \cdots \beta_{j_k} (X_{t-i-j_1-\cdots-j_k}^-)^{2\delta}.
$$

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

$$
\frac{\partial \sigma_t(\phi)^{2\delta}}{\partial \phi} = \sum_{i=1}^{p} \gamma_{1i} X_{t-i}^+ \log(X_{t-i}^+) + \sum_{i=1}^{p} \gamma_{2i} X_{t-i}^- \log(X_{t-i}^-) \\
+ \sum_{i=1}^{p} \gamma_{1i} \sum_{k=1}^{\infty} \sum_{j_1=1}^{q} \cdots \sum_{j_k=1}^{q} \beta_{j_1} \cdots \beta_{j_k} (X_{t-i-j_1-\cdots-j_k}^+) \log(X_{t-i-j_1-\cdots-j_k}^+) \\
+ \sum_{i=1}^{p} \gamma_{2i} \sum_{k=1}^{\infty} \sum_{j_1=1}^{q} \cdots \sum_{j_k=1}^{q} \beta_{j_1} \cdots \beta_{j_k} (X_{t-i-j_1-\cdots-j_k}^-) \log(X_{t-i-j_1-\cdots-j_k}^-) =: L_{1t}^\phi(\phi) + L_{2t}^\phi(\phi) + L_{3t}^\phi(\phi),
$$

$$
\frac{\partial \sigma_t(\phi)^{2\delta}}{\partial \gamma_{0}} = \frac{1}{1 - \sum_{j=1}^{q} \beta_j},
$$

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

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

$$
\frac{\partial \sigma_t(\phi)^{2\delta}}{\partial \beta_j} = \frac{\gamma_0}{(1 - \sum_{j=1}^{q} \beta_j)^2} + \sum_{i=1}^{p} \gamma_{1i} \sum_{k=0}^{\infty} (k+1) \sum_{j_1=1}^{q} \cdots \sum_{j_k=1}^{q} \beta_{j_1} \cdots \beta_{j_k} (X_{t-i-j_1-\cdots-j_k}^+)^{2\delta} \\
+ \sum_{i=1}^{p} \gamma_{2i} \sum_{k=0}^{\infty} (k+1) \sum_{j_1=1}^{q} \cdots \sum_{j_k=1}^{q} \beta_{j_1} \cdots \beta_{j_k} (X_{t-i-j_1-\cdots-j_k}^-)^{2\delta} =: L_{1t}^{\beta_j}(\phi) + L_{2t}^{\beta_j}(\phi) + L_{3t}^{\beta_j}(\phi).
$$

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

$$
\tilde{\sigma}(\phi)^{2\delta} = \frac{x_0}{1 - \sum_{j=1}^{q} \beta_j} + \sum_{i=1}^{p} x_1i(X_{i}^{+})^{2\delta} + \sum_{i=1}^{p} x_2i(X_{i}^{-})^{2\delta}
+ \sum_{i=1}^{p} x_{1j} \sum_{k=1}^{q} \sum_{j_1=1}^{q} \cdots \sum_{j_k=1}^{q} \beta_j_1 \cdots \beta_j_{k}(X_{i-j_1-j_2-\cdots-j_k}^{+})^{2\delta} I(t - i - j_1 - \cdots - j_k \geq 1)
+ \sum_{i=1}^{p} x_{2j} \sum_{k=1}^{q} \sum_{j_1=1}^{q} \cdots \sum_{j_k=1}^{q} \beta_j_1 \cdots \beta_j_{k}(X_{i-j_1-j_2-\cdots-j_k}^{-})^{2\delta} I(t - i - j_1 - \cdots - j_k \geq 1).
$$

(2.8)

### 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

$$
L_n(\phi) = \sum_{i=1}^{n} -\frac{1}{2} \left\{ \log \sigma_i^2(\phi) + \frac{X_i^2}{\sigma_i^2(\phi)} \right\} = \sum_{i=1}^{n} l_i(\phi),
$$

(2.9)

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

$$
A(\phi) = E \left\{ \frac{\partial^2 L_i(\phi)}{\partial \phi \partial \phi^t} \right\} \quad \text{and} \quad \Omega(\phi) = E \left\{ \frac{\partial L_i(\phi)}{\partial \phi} \frac{\partial L_i(\phi)}{\partial \phi^t} \right\}.
$$

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

(A3) $E v_i^2 = 1$, and $E v_i^4 < \infty$.

**Remark 2.** If $v_i$ has density at 0, then (2.1) is satisfied for any $\mu < \frac{1}{2}$.

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

**Theorem 1.** Under assumptions (A1)–(A3), it follows that

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

(ii) $\sqrt{n} (\tilde{\phi}_n - \phi^0) \xrightarrow{d} N(0, A_0^{-1} \Omega_0 A_0^{-1})$, where $A_0 = A(\phi^0)$ and $\Omega_0 = \Omega(\phi^0)$.

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

$$
\tilde{L}_n(\phi) = \sum_{i=1}^{n} -\frac{1}{2} \left\{ \log \tilde{\sigma}_i^2(\phi) + \frac{X_i^2}{\tilde{\sigma}_i^2(\phi)} \right\}.
$$

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

$$
\tilde{A}(\phi) = \frac{1}{n} \sum_{i=1}^{n} \frac{\partial^2 \tilde{L}_i(\phi)}{\partial \phi \partial \phi^t} \quad \text{and} \quad \tilde{\Omega}(\phi) = \frac{1}{n} \sum_{i=1}^{n} \frac{\partial \tilde{L}_i(\phi)}{\partial \phi} \frac{\partial \tilde{L}_i(\phi)}{\partial \phi^t},
$$

where $\tilde{L}_i(\phi)$ are defined similar to $l_i(\phi)$ by replacing $\sigma_i(\phi)$ by $\tilde{\sigma}_i(\phi)$. The next theorem shows that our results for $\tilde{\phi}_n$ are the same as those for $\tilde{\phi}_n$, and $\tilde{A}(\tilde{\phi}_n)$ and $\tilde{\Omega}(\tilde{\phi}_n)$ are consistent estimators of $A_0$ and $\Omega_0$, respectively.

**Theorem 2.** Under assumptions (A1)–(A3), it follows that

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

(ii) $\sqrt{n} (\tilde{\phi}_n - \phi^0) \xrightarrow{d} N(0, A_0^{-1} \Omega_0 A_0^{-1})$. 
(iii) \( \tilde{L}_n \equiv \tilde{L}(\phi_n) \xrightarrow{a.s.} A_0 \) and \( \tilde{\Omega}_n \equiv \tilde{\Omega}(\phi_n) \xrightarrow{a.s.} \Omega_0 \).

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

\[
H_0 : \Gamma \phi^0 = \theta,
\]

where \( \Gamma \) is a \( s \times d \) constant matrix with rank \( s \), and \( \theta \) is \( s \times 1 \) constant vector. By Theorems 2 and 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 \( \chi^2 \).

**Remark 3.** From the above discussion, it can be seen that if we apply QMLE, we need Assumption (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 \( \varepsilon_i \) 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 \( \varepsilon_i \) 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)

\[
\tilde{S}_n(\phi) = \sum_{t=1}^n |\log |X_t|| - \log \tilde{\sigma}_t(\phi),
\]

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

\[
v = (v_1, \ldots, v_d)^t = \sqrt{n}(\phi - \phi^0), \quad \tilde{Q}_t(\phi) = \log |X_t| - \log \tilde{\sigma}_t(\phi) \quad \text{and} \quad Q_t(\phi) = \log |X_t| - \log \sigma_t(\phi),
\]

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

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

Define

\[
D_t(\phi) = (D_{1,1}(\phi), \ldots, D_{1,d}(\phi))^t, \quad \Sigma = E(D_t(\phi^0)D_t(\phi^0)),
\]

where

\[
D_{1,1}(\phi) = -\frac{\partial Q_t(\phi)}{\partial \phi_1} = -\frac{1}{2\sigma_t^2(\phi)} \log \sigma_t^{2\delta}(\phi) + \frac{1}{2\sigma_t^{2\delta}(\phi)} \frac{\partial \sigma_t^{2\delta}(\phi)}{\partial \delta}, \quad (2.11)
\]

\[
D_{i,1}(\phi) = -\frac{\partial Q_t(\phi)}{\partial \phi_i} = \frac{1}{2\sigma_t^{2\delta}(\phi)} \frac{\partial \sigma_t^{2\delta}(\phi)}{\partial \phi_i}, \quad i = 2, \ldots, d. \quad (2.12)
\]

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

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

**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 \( \tilde{S}_n(\phi) \) which lies in the random region.
(ϕ : ∥ϕ − ϕ0∥ − ξ/√n ≤ M/√n} for which
\[
\sqrt{n}(\hat{ϕ}_n − ϕ^0) \rightarrow \mathcal{N}\left(0, \frac{1}{4f^2(0)} \Sigma^{-1}\right).
\]

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

3. Order selection and tests for GARCH structure

The LR test, the 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
\[
W_n(s) = (I \hat{ϕ}_n − 0)\left\{\Gamma \frac{1}{4nf^2(0)} \hat{\Sigma}^{-1} \Gamma'\right\}^{-1} (I \hat{ϕ}_n − 0).
\]

We reject \(H_0\) for large values of \(W_n(s)\). In the above expression,
\[
\hat{\Sigma} = \frac{1}{n} \sum_{i=1}^{n} \tilde{D}_i(\hat{ϕ}_n) \tilde{D}_i(\hat{ϕ}_n)' \quad \hat{f}(0) = \frac{1}{nb_n} \sum_{i=1}^{n} K\left(\frac{1}{b_n} \log \frac{|X_i|}{\hat{a}_i(\hat{ϕ}_n)}\right),
\]
where \(\tilde{D}_i(\hat{ϕ}_n)\) is defined similar to \(D_i(\hat{ϕ}_n)\) by replacing \(a_i(ϕ)\) by \(\hat{a}_i(ϕ)\), \(K(\cdot)\) 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) \(b_n \rightarrow 0\) and \(nb_n^4 \rightarrow \infty\) as \(n \rightarrow \infty\), then \(W_n(s) \xrightarrow{D} \chi^2_1\) 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 \(\Gamma\) in (2.10) such that
\[
\Gamma \phi^0 = (φ^0_{q+1}, \ldots, φ^0_{Q_0})' \quad \Gamma \phi^0 = (φ^0_{p+1}, φ^0_{2p+1}, \ldots, φ^0_{1p_0}, φ^0_{2p_0})'
\]
and
\[
θ = (0, \ldots, 0)_{(Q_0-q)×1} \quad θ = (0, \ldots, 0)_{2(P_0-p)×1}.
\]

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 \(p = P_0 − 1, \ldots, s_0\) and \(q = Q_0 − 1, \ldots, r_0\) in (3.2) until \(p = s_0\) and \(q = r_0\) such that \(W_n(s) > \chi^2_1(η)\). 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: \(δ = 1\) and \(φ_{1i} = x_{2i}, i = 1, \ldots, p\).

(ii) IGARCH: \(δ = 1, \ x_{1i} = x_{2i} = x_i, i = 1, \ldots, p\) and \(\sum_{i=1}^{p} φ_i + \sum_{j=1}^{q} β_j = 1\).

(iii) Symmetric GARCH: \(φ_{1i} = x_{2i}, i = 1, \ldots, p\).

(iv) No power transformation: \(δ = 1\).
Let

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

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

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

$$
\Gamma_4 = \begin{pmatrix}
1 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 \\
\end{pmatrix}_{d}.
$$

$$
\theta_1 = (1, \ldots, 0, 0)_{p+1}^\prime, \quad \theta_2 = (1, 0, \ldots, 0, 1)_{p+2}^\prime, \quad \theta_3 = (0, \ldots, 0)_p^\prime, \quad \theta_4 = 1.
$$

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

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

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 Sections 2 and 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

$$
X_t = \sigma_t \epsilon_t \quad \text{and} \quad \sigma_t^{2\delta} = \sigma_0 + \sigma_{11}(X_{t-1}^{\delta})^{2\delta} + \sigma_{21}(X_{t-1}^{\delta})^{2\delta} + \beta_1 \sigma_{1_{t-1}}^{2\delta},
$$

with the true parameter $(\delta^0, \sigma_0^0, \sigma_{11}^0, \sigma_{21}^0, \beta_1^0)^\prime = (0.8, 0.2, 0.2, 0.1, 0.4)^\prime$. We take the errors $\epsilon_t$ 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 $u = 10$.

Fig. 1 presents the boxplots of the average absolute error (AAE) $(|\hat{\delta} - 0.8| + |\hat{\sigma}_0 - 0.2| + |\hat{\sigma}_{11} - 0.2| + |\hat{\sigma}_{21} - 0.1| + |\beta_1 - 0.4|)/5$ 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. Figs. 2 and 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 \( f_{X_t} \) is heavy-tailed. The Hill estimator and the QQ-plot are used for this. The Hill estimator is defined as

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

where \( X_{(1)} \geq X_{(2)} \geq \cdots \geq X_{(n)} \) are the order statistics of \( X_1, \ldots, X_n \). We plot \( (m, H_{n,m}), 1 \leq m \leq 600 \) in Fig. 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 \( m \geq 200 \). Fig. 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 \( a_t = \sum_{t=1}^{C_{00}} |X_{t-1}|I(|X_{t-1}| \geq C) \) 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 \( K(x) = e^{-x}/(1 + e^{-x})^2 \) and \( b_n = 1.06n^{-1/5} \) (Silverman, 1986, p. 40). 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 
\[ (\tilde{\alpha}, \tilde{\beta}, \tilde{\alpha}_{11}, \tilde{\alpha}_{21}, \tilde{\beta}) = (1.6616, 0.3414, 0.0368, 0.1561, 0.7405) \]
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 \( W_n^2(12) \). Fig. 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 Fig. 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 \( p = 3, q = 2 \) vs \( p = 3, q = 3; p = 3, q = 1 \) vs \( p = 3, q = 2; \) and \( p = 2, q = 1 \) vs \( p = 1, q = 1 \) in order, and all the Wald statistics are less than 1, namely not significant. Then we test \( p = 1, q = 1 \) vs \( p = 1, q = 0, \)
and the Wald statistic is 34.9, which rejects the null hypothesis and we take $p = 1$, $q = 1$. The LADEs are

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

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 $W_n(8) = 11.99$ and $W_n(12) = 17.46$ 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 hypotheses $\delta = 1$ and $z_{11} = z_{21}$, 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 $z_{11}$ and $z_{21}$. In fact, as we can see from the estimators, $\hat{z}_{21}$ is about five times $\hat{z}_{11}$. 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 Section 2 and Section 3. Before we prove Theorems 1–4, we introduce some lemmas first.

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

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

and

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

where $\delta_2$ is the positive constant in Remark 1.
Proof. Denote
\[
\frac{1 + x_1 + \cdots + x_p}{1 - \beta_1 x - \cdots - \beta_q x^q} = \sum_{i=0}^{\infty} d_i x^i, \quad |x| \leq 1
\]
and
\[
\frac{1 + x_1 + \cdots + x_p}{1 - \beta_1 x - \cdots - \beta_q x^q} = \sum_{i=0}^{\infty} d_i x^i, \quad |x| \leq 1.
\]
By Lemma 3.1 of Berkes et al. (2003), there exist some constants \( C > 0 \) and \( 0 < r < 1 \) such that
\[0 < d_i \leq C r^i, \quad i = 1, 2, \quad t = 0, 1, \ldots, \]
where \( C \) and \( r \) are both independent of \( \phi \). Notice that
\[
\sigma_i^2(\phi) - \bar{\sigma}_i^2(\phi) = \sum_{i=1}^{\infty} d_1(X_{t-i}^+)^{2^i} + \sum_{i=1}^{\infty} d_2(X_{t-i}^-)^{2^i},
\]
thus (5.1) holds. Similarly, we can show that (5.2) holds. \( \square \)

Lemma 2. Under assumption (A1) and (A2), it follows that
\[
E \sup_{\phi \in \Theta} \left| \frac{\partial^m Q_i(\phi)}{\partial \phi_{j_1} \cdots \partial \phi_{j_m}} \right|^h < + \infty,
\]
where \( m, h \) are any positive integers, \( 1 \leq j_1, \ldots, j_m \leq d \) and \( Q_i(\phi) = \log |X_i - \log \sigma_i(\phi)| \).

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
\[
E \sup_{\phi \in \Theta} |\log \sigma_i^2(\phi)|^h < + \infty, \quad (5.3)
\]
\[
E \sup_{\phi \in \Theta} \left| \frac{\partial \sigma_i^2(\phi)}{\partial \phi\phi_i} \right|^h < + \infty, \quad i = 2, \ldots, d, \quad (5.4)
\]
\[
E \sup_{\phi \in \Theta} \left| \frac{\partial \sigma_i^2(\phi)}{\partial \phi_{j_1} \cdots \partial \phi_{j_m}} \right|^h < + \infty. \quad (5.5)
\]
By the definition of \( \sigma_i^2 \), we have \( \sigma_i^2 > x_0 > \delta_1 \), where \( \delta_1 \) is the positive constant in Remark 1. Notice that for any given positive integers \( h \), there exists \( M_1 > 0 \) such that \( (\log x)^h \leq x^h \) for \( x > M_1 \), where \( \tilde{h} = \min\{1, \tau/(2\delta_2)\} \), \( \tau \) and \( \delta_2 \) are defined, respectively, in Theorem 6 and Remark 1. By Lemma 1, it follows that
\[
\delta_1 \leq \sup_{\phi \in \Theta} \sigma_i^2(\phi) \leq C \sum_{j=1}^{\infty} r^j(|X_{t-j}|^{2^j} + 1).
\]
Noticing that
\[
\log(\delta_1) \leq \log \sigma_i^2(\phi) \leq \log \left( \sup_{\phi \in \Theta} \sigma_i^2(\phi) \right),
\]
we have
\[
\sup_{\phi \in \Theta} |\log \sigma_i^2(\phi)| \leq |\log \delta_1| + \log \left[ C \sum_{j=1}^{\infty} r^j(|X_{t-j}|^{2^j} + 1) \right] \leq \left[ C + \sum_{j=1}^{\infty} r^j(|X_{t-j}|^{2^j} + 1) \right]^{1/h},
\]
which implies (5.3) holds by Theorem 6(i).
It is obvious that (5.4) holds for \( i = 2, \ldots, 2p + 2 \) by (2.4)--(2.6). From Lemma 5.2 of Berkes et al. (2003), we can obtain (5.4) for \( i = 2p + 3, \ldots, d \).

Now we prove (5.5). Using the same argument as (A.5), we can obtain \( \mathbb{E} \sup_{\phi \in \Theta} |L_{2i}^2(\phi)/\sigma_2^2(\phi)|^b < +\infty \), where \( L_{2i}^2(\phi) \) is defined in (2.3). On the other hand,

\[
L_{2i}^2(\phi) \leq \rho_0 \sum_{j=1}^{q} \sum_{j=1}^{q} \beta_j \cdots \beta_j (X^+_{t-i-j-1} \cdots - j_k)^{2\delta} \log(X^+_{t-i-j-1} \cdots - j_k)^2
\]

where \( \rho_0 \) is the same one as in Remark 1 and \( L_{2i}^2(\phi) \) is defined in (2.3). From Hölder inequality and (5.4), it follows that

\[
\left( \mathbb{E} \sup_{\phi \in \Theta} |L_{i,j}^1(\phi)/\sigma_1^2(\phi)|^b \right)^2 \leq \mathbb{E} \sup_{\phi \in \Theta} \left| L_{i,j}^1(\phi)/\sigma_1^2(\phi) \right|^{2h} \cdot \mathbb{E} \sup_{\phi \in \Theta} \left| L_{2i}^2(\phi)/\sigma_1^2(\phi) \right|^{2h} \leq C \mathbb{E} \sup_{\phi \in \Theta} \left| L_{i,j}^1(\phi)/\sigma_1^2(\phi) \right|^{2h}
\]

It can be easily verified that

\[
\frac{x_1 + x_2}{y_1 + y_2} \leq \max \left\{ \frac{x_1}{y_1}, \frac{x_2}{y_2} \right\}, \quad \text{for any } x_1, x_2, y_1, y_2 > 0.
\]

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

\[
\mathbb{E} \sup_{\phi \in \Theta} \left| L_{2i}^2(\phi)/\sigma_1^2(\phi) \right|^{2h} = \mathbb{E} \sup_{\phi \in \Theta} \left\{ \sum_{i=1}^{p} \sum_{j=1}^{q} \beta_j \cdots \beta_j (X^+_{t-i-j-1} \cdots - j_k)^{2\delta} \log(X^+_{t-i-j-1} \cdots - j_k)^2 \right\}^{2h}
\]

\[
\leq C + CE \left\{ \max_{k \geq 0} \frac{\log(X^+_{t-i-j-1} \cdots - j_k)^2}{k + 1} \right\}^{2h}
\]

\[
\leq C + C \int_1^{\infty} 2h y^{2h-1} P \left\{ \max_{k \geq 0} \frac{\log(X^+_{t-i-j-1} \cdots - j_k)^2}{k + 1} > y \right\} dy
\]

\[
\leq C + C \sum_{k=0}^{\infty} \int_1^{\infty} 2h y^{2h-1} P \left\{ \max_{k \geq 0} \frac{\log(X^+_{t-i-j-1} \cdots - j_k)^2}{k + 1} > y \right\} dy
\]

\[
= C + C \sum_{k=0}^{\infty} \int_1^{\infty} 2h y^{2h-1} [ P \{ (X^+_{t})^2 > e^{(k+1)y} \} + P \{ (X^+_{t})^2 < e^{-(k+1)y} \} ] dy
\]

\[
\leq C + C \sum_{k=0}^{\infty} \int_1^{\infty} k y^{2h-1} \left[ e^{-(k+1)y} E(X^+_{t})^2 + P \{ \varepsilon^2_t < (\varepsilon_0)^{-1/(2\delta)} e^{-(k+1)y} \} \right] dy
\]

\[
\leq C + C \sum_{k=0}^{\infty} \int_1^{\infty} k y^{2h-1} \left[ e^{-(k+1)y} E(X^+_{t})^2 + e^{-\mu(k+1)y/\varepsilon} \right] dy
\]

\[
< +\infty.
\]

Thus, \( \mathbb{E} \sup_{\phi \in \Theta} |L_{2i}^2(\phi)/\sigma_1^2(\phi)|^b < +\infty \). Similarly, we can obtain that \( \mathbb{E} \sup_{\phi \in \Theta} |L_{3i}^2(\phi)/\sigma_1^2(\phi)|^b < +\infty \). This completes the proof. \( \square \)
Lemma 3. Suppose that assumptions (A1) and (A2) hold. If \( E|\epsilon_i|^1+\eta < \infty \) for some \( \eta > 0 \), it follows that
\[
E \sup_{\phi \in \Theta} \left| \frac{\partial^{m} l_{1}(\phi)}{\partial \phi_1 \cdots \partial \phi_m} \right| < + \infty,
\]
where \( 1 \leq j_1, \ldots, j_m \leq d \).

Proof. We only prove the Lemma for \( m = 1 \), noting that similar arguments apply for \( m > 1 \). By the definition of \( l_{1}(\phi) \), we have
\[
\frac{\partial l_{1}(\phi)}{\partial \phi} = -\frac{1}{2\sigma_1^2(\phi)} \frac{\partial \sigma_1^2(\phi)}{\partial \phi} + \frac{X_1^2}{2\sigma_1^4(\phi)} \frac{\partial \sigma_1^2(\phi)}{\partial \phi}.
\]
By similar argument to Lemma 5.1 of Berkes et al. (2003), we can obtain that
\[
E \sup_{\phi \in \Theta} \left| \frac{\sigma_1^2(\phi^0)}{\sigma_1^2(\phi)} \right|^{1+\eta} < \infty.
\]
From Hölder inequality, Lemma 2 and (5.6), we have
\[
E \sup_{\phi \in \Theta} \left| \frac{X_1^2}{2\sigma_1^4(\phi)} \frac{\partial \sigma_1^2(\phi)}{\partial \phi} \right| = E \sup_{\phi \in \Theta} \left| \frac{\sigma_1^2(\phi^0)}{\sigma_1^2(\phi)} \frac{1}{2\sigma_1^2(\phi)} \frac{\partial \sigma_1^2(\phi)}{\partial \phi} \right| \leq \left\{ E \sup_{\phi \in \Theta} \left( \frac{\sigma_1^2(\phi^0)}{\sigma_1^2(\phi)} \right)^{1+\eta} \right\}^{\frac{1}{1+\eta}} \left\{ E \sup_{\phi \in \Theta} \left| \frac{1}{2\sigma_1^2(\phi)} \frac{\partial \sigma_1^2(\phi)}{\partial \phi} \right| \right\}^{\frac{1}{1+\eta}} < \infty.
\]
Using Lemma 2 again, we obtain
\[
E \sup_{\phi \in \Theta} \left| \frac{1}{2\sigma_1^2(\phi)} \frac{\partial \sigma_1^2(\phi)}{\partial \phi} \right| < \infty.
\]
Thus, \( E \sup_{\phi \in \Theta} \left| \frac{\partial l_{1}(\phi)}{\partial \phi} \right| < \infty. \)

Lemma 4. Suppose that assumptions (A1)–(A3) hold, and denote \( L(\phi) = E l_{1}(\phi) \) for all \( \phi \in \Theta \). Then \( L(\phi) \) is well defined and \( \phi^0 \) is the unique maximizer of \( L(\phi) \).

Proof. By Lemmas 2 and 3, it is obvious that \( L(\phi) \) is well defined. Maximizing \( L(\phi) \) is equivalent to minimizing \( L(\phi^0) - L(\phi) \). But,
\[
L(\phi^0) - L(\phi) = \frac{1}{2} E \left\{ \frac{\sigma_1^2(\phi^0)}{\sigma_1^2(\phi)} - \log \frac{\sigma_1^2(\phi^0)}{\sigma_1^2(\phi)} \right\} - \frac{1}{2}.
\]
Note that the function \( x - \log x > 0 \) for any \( x > 0 \) and reaches its unique minimum value at \( x = 1 \). Since \( \sigma_1^2(\phi) = \sigma_1^2(\phi^0) \) if and only if \( \phi = \phi^0 \), we obtain the result. \( \square \)

Lemma 5. If the conditions of Theorem 1 are satisfied, then as \( n \to \infty \), we have
\[
\sup_{\phi \in \Theta} \left| \frac{1}{n} L_{n}(\phi) - L(\phi) \right| \xrightarrow{a.s.} 0,
\]
\[
\sup_{\phi \in \Theta} \left| \frac{1}{n} \frac{\partial L_{n}(\phi)}{\partial \phi} - \frac{\partial L(\phi)}{\partial \phi} \right| \xrightarrow{a.s.} 0,
\]
\[
\sup_{\phi \in \Theta} \left| \frac{1}{n} \frac{\partial^2 L_{n}(\phi)}{\partial \phi \partial \phi} - \frac{\partial^2 L(\phi)}{\partial \phi \partial \phi} \right| \xrightarrow{a.s.} 0.
\]
Proof. By the ergodic theorem,
\[ \frac{1}{n} L_n(\phi) \overset{a.s.}{\to} L(\phi), \] (5.7)
for any \( \phi \in \Theta \). Using the mean value theorem, we have
\[ \sup_{\phi^1, \phi^2 \in \Theta} |L_n(\phi^1) - L_n(\phi^2)| \leq \frac{1}{2 \| \phi^1 - \phi^2 \|} \sum_{i=1}^{n} \sup_{\phi^1, \phi^2 \in \Theta} \left\{ \left[ \log \sigma_i^2(\phi^1) - \log \sigma_i^2(\phi^2) \right] + \left| \frac{X_i^2}{\sigma_i^2(\phi^1)} - \frac{X_i^2}{\sigma_i^2(\phi^2)} \right| \right\} \]
\[ \leq C \sum_{i=1}^{n} \sup_{\phi^1, \phi^2 \in \Theta} \left\{ \left| \frac{\partial \sigma_i^2(\phi^*)}{\partial \phi} \right| + \left| \frac{\varepsilon_i^2 \sigma_i^2(\phi^0) \partial \sigma_i^2(\phi^{**})}{\sigma_i^2(\phi^{**})} \right| \right\}, \]
where \( \phi^*, \phi^{**} \) lie on the line form \( \phi^1 \) to \( \phi^2 \). We have
\[ \mathbb{E} \left\{ \sup_{\phi^1, \phi^2 \in \Theta} \left[ \left| \frac{1}{\sigma_i^2(\phi^*)} \frac{\partial \sigma_i^2(\phi^*)}{\partial \phi} \right| + \left| \frac{\varepsilon_i^2 \sigma_i^2(\phi^0) \partial \sigma_i^2(\phi^{**})}{\sigma_i^2(\phi^{**})} \right| \right] \right\} < \infty \]
by Lemmas 2 and 3. Then,
\[ \sup_{\phi^1, \phi^2 \in \Theta} \left| \frac{1}{n} L_n(\phi^1) - \frac{1}{n} L_n(\phi^2) \right| \overset{a.s.}{\to} 0. \]
which shows that \( L_n(\phi)/n \) is equicontinuous with probability one. Combining this fact, (5.7) and the compactness of \( \Theta \), the uniform convergence of \( L_n(\phi)/n \) follows. By the same method, we can prove that the results hold for \((1/n)(\partial L_n(\phi)/\partial \phi)\) and \((1/n)(\partial^2 L_n(\phi)/\partial \phi \partial \phi')\). □

Lemma 6. If the conditions of Theorem 1 are satisfied, then it follows that:
\[ \sup_{\phi \in \Theta} \left| \frac{1}{\sqrt{n}} L_n(\phi) - \frac{1}{\sqrt{n}} \bar{L}_n(\phi) \right| \overset{a.s.}{\to} 0, \]
\[ \sup_{\phi \in \Theta} \left| \frac{1}{\sqrt{n}} \partial L_n(\phi) - \frac{1}{\sqrt{n}} \partial \bar{L}_n(\phi) \right| \overset{a.s.}{\to} 0, \]
\[ \sup_{\phi \in \Theta} \left| \frac{1}{\sqrt{n}} \partial^2 L_n(\phi) - \frac{1}{\sqrt{n}} \partial^2 \bar{L}_n(\phi) \right| \overset{a.s.}{\to} 0. \]

Proof. Let
\[ U_1 = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sup_{\phi \in \Theta} \left| \log \sigma_i^2(\phi) - \log \bar{\sigma}_i^2(\phi) \right| \quad \text{and} \quad U_2 = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sup_{\phi \in \Theta} \left| \frac{X_i^2}{\sigma_i^2(\phi)} - \frac{X_i^2}{\bar{\sigma}_i^2(\phi)} \right|, \]
then
\[ \sup_{\phi \in \Theta} \left| \frac{1}{\sqrt{n}} L_n(\phi) - \frac{1}{\sqrt{n}} \bar{L}_n(\phi) \right| \leq U_1 + U_2. \]
By Lemma 1, we have

\[
U_1 = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sup_{\phi \in \Theta} \frac{1}{\delta} \left| \log \left( 1 + \frac{\sigma_i^2(\phi) - \tilde{\sigma}_i^2(\phi)}{\tilde{\sigma}_i^2(\phi)} \right) \right| \\
\leq \frac{C}{\sqrt{n}} \sum_{i=1}^{n} \sup_{\phi \in \Theta} \frac{1}{\delta} \left| \sigma_i(\phi) - \tilde{\sigma}_i(\phi) \right| \\
\leq \frac{C}{\sqrt{n}} \sum_{i=1}^{n} \sum_{t=1}^{\infty} r'(1 + |X_{t-}|^{\delta_2}) \\
\leq \frac{C}{\sqrt{n}} \sum_{i=1}^{n} r' + \frac{C}{\sqrt{n}} \sum_{i=1}^{n} r' \sum_{k=0}^{\infty} r^k |X_{-h}|^{\delta_2} \\
\xrightarrow{a.s.} 0.
\]

By the mean value theorem, we have

\[
\sigma_i^2(\phi) - \tilde{\sigma}_i^2(\phi) = (\sigma_i^2(\phi))^{1/\delta} - (\tilde{\sigma}_i^2(\phi))^{1/\delta} = \frac{1}{\delta} (\sigma_i^2(\phi))^{1/\delta - 1} (\sigma_i^{2\delta}(\phi) - \tilde{\sigma}_i^{2\delta}(\phi)),
\]

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

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

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

\[
\sup_{\phi \in \Theta} \left| \frac{1}{\sqrt{n}} \frac{\partial L_n(\phi)}{\partial \phi} - \frac{1}{\sqrt{n}} \frac{\partial \tilde{L}_n(\phi)}{\partial \phi} \right| \xrightarrow{a.s.} 0,
\]

\[
\sup_{\phi \in \Theta} \left| \frac{1}{\sqrt{n}} \frac{\partial^2 L_n(\phi)}{\partial \phi \partial \phi'} - \frac{1}{\sqrt{n}} \frac{\partial^2 \tilde{L}_n(\phi)}{\partial \phi \partial \phi'} \right| \xrightarrow{a.s.} 0.
\]

This completes the proof. \( \square \)

**Proof of Theorem 1.** (i) By Theorem 4.1.1 and the associated in Amemiya (1985), we have \(\phi_n \xrightarrow{a.s.} \phi^0\) if the following conditions hold:

(a) \(\Theta\) 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) \((1/n)L_n(\phi) \xrightarrow{a.s.} L(\phi)\) uniformly in \(\phi \in \Theta\);
(d) \(L(\phi)\) attains a unique global maximum at \(\phi^0\).

By assumption (A.2), Lemmas 4 and 5, we know conditions (a)–(d) are satisfied. Thus, \(\phi_n \xrightarrow{a.s.} \phi^0\).

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

\[
\frac{\partial L_n(\phi_n)}{\partial \phi} = \frac{\partial L_n(\phi_0)}{\partial \phi} + \frac{\partial^2 L_n(\xi)}{\partial \phi \partial \phi'} (\phi_n - \phi_0),
\]

where \(\xi\) lies between \(\phi^0\) and \(\phi_n\). But, \(\partial L_n(\phi_n)/\partial \phi = 0\). Then

\[
\frac{\partial^2 L_n(\xi)}{\partial \phi \partial \phi'} (\phi_n - \phi_0) = -\frac{\partial L_n(\phi_0)}{\partial \phi}.
\]
Using Lemma 3 and the continuity of $\partial^2 L_n(\phi)/\partial \phi \partial \phi'$, we obtain

$$(A_0 + o(1))(\hat{\phi}_n - \phi^0) = -\frac{1}{n} \frac{\partial L_n(\phi^0)}{\partial \phi}.$$ 

It can be easily verified that

$$\frac{\partial L_n(\phi^0)}{\partial \phi} = \frac{1}{2} (1 - \epsilon^2) \frac{1}{\sigma^2(\phi^0)} \frac{\partial \sigma^2(\phi^0)}{\partial \phi}$$

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

$$\sqrt{n}(\hat{\phi}_n - \phi^0) \xrightarrow{d} N(0, A_0^{-1} \Omega_0 A_0^{-1}).$$

**Proof of Theorem 2.**

(i) By Lemmas 5, and 6, we have

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

Imitating the proof of Theorem 1 (i), we obtain the result.

(ii) Notice that

$$\frac{1}{\sqrt{n}} \frac{\partial L_n(\phi^0)}{\partial \phi} - \frac{1}{\sqrt{n}} \frac{\partial L_n(\phi^0)}{\partial \phi} = \left[ \frac{1}{\sqrt{n}} \frac{\partial L_n(\phi^0)}{\partial \phi} - \frac{1}{\sqrt{n}} \frac{\partial L_n(\phi^0)}{\partial \phi} \right] + \left[ \frac{1}{\sqrt{n}} \frac{\partial \tilde{L}_n(\phi^0)}{\partial \phi} - \frac{1}{\sqrt{n}} \frac{\partial \tilde{L}_n(\phi^0)}{\partial \phi} \right].$$

From Lemma 5 and the mean value theorem, we have

$$\frac{1}{n} \frac{\partial^2 L_n(\phi^0)}{\partial \phi \partial \phi'} \sqrt{n}(\hat{\phi}_n - \phi^0) \overset{a.s.}{\rightarrow} 0.$$ 

Thus, $\sqrt{n}(\hat{\phi}_n - \phi^0) = o(1)$ by Lemma 3 and the continuity of $\partial^2 L_n(\phi)/\partial \phi \partial \phi'$. Then the result follows from Theorem 1.

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

$$\mathbb{E} \sup_{\phi \in \Theta} |\hat{A}(\phi) - A(\phi)| \overset{a.s.}{\rightarrow} 0,$$ 

and $[\hat{A}(\phi^0) - A(\phi^0)] = O_p(1)(\hat{\phi}_n - \phi^0) \rightarrow 0$.

Therefore,

$$\hat{A}(\phi^0) = [\hat{A}(\phi^0) - A(\phi^0)] + [A(\phi^0) - A(\phi^0)] + A(\phi^0) \rightarrow A_0,$$

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

**Proof of Theorem 3.**

Define

$$T_n(v) = \sum_{i=1}^{n} \left( |Q_i(\phi^0) + n^{-1/2}v| - |Q_i(\phi^0)| \right),$$

$$T_n^*(v) = \sum_{i=1}^{n} \left( |Q_i(\phi^0) - n^{-1/2}vD_i| - |Q_i(\phi^0)| \right),$$

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

$$T_n(v) - \hat{T}_n(v) \overset{P}{\rightarrow} 0,$$

uniformly on compact sets. Using the equality

$$\left| |z - y| - |z| \right| = -y \text{sgn}(z) + 2(y - z)(1(y < z) - I(y < z < 0)), \quad z \neq 0,$$

(5.10)
we have

\[
T_n^*(v) = n^{-1/2} \sum_{i=u}^n v'D_i \text{sgn}(Q_i(\phi^0)) + 2 \sum_{i=u}^n (n^{-1/2}v'D_i - Q_i(\phi^0))(I(0 < Q_i(\phi^0) < n^{-1/2}v'D_i) - I(n^{-1/2}v'D_i < Q_i(\phi^0) < 0))
\]

\[\equiv A_n + B_n.\]

Since \(Q_i(\phi^0) = \log|\varepsilon_i|\), we know \(\{v'D_i \text{sgn}(\log|\varepsilon_i|)\}\) 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 \(A_n \xrightarrow{p} v'N\), 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

\[
nE U_{nt}^2 = nE(E(U_{nt}^2|\mathcal{F}_{t-1}))
\]

\[
= nE \left( \int_0^{n^{-1/2}v'D_t} (n^{-1/2}v'D_t - x)^2 f(x) \, dx \right)
\]

\[
= nE \left[ \int_0^{n^{-1/2}v'D_t} (n^{-1/2}v'D_t - x)^2 f(x) \, dx \right] + \int_0^{n^{-1/2}v'D_t} (n^{-1/2}v'D_t - x)^2 f(0) \, dx
\]

\[
\leq E(B_2n^{-2}(v'D_t)^4 + B_1n^{-3/2}(v'D_t)^3).
\]

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

\[
\lim_{n \to \infty} nE U_{nt}^2 = 0. \quad (5.11)
\]

On the other hand, on the set \(\{D'_v > 0\}\), we may show that

\[
\sum_{i=u}^n E(U_{nt}|\mathcal{F}_{i-1}) \to \frac{f(0)}{2} E[(v'D_t)^2 I(v'D_t > 0)]
\]

and

\[
\text{Var} \left( \sum_{i=u}^n (U_{nt} - E(U_{nt}|\mathcal{F}_{i-1})) \right) \to 0.
\]

Therefore,

\[
\sum_{i=u}^n U_{nt} \to \frac{f(0)}{2} E[(v'D_t)^2 I(v'D_t > 0)].
\]

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

\[
B_n \xrightarrow{p} f(0)v'\Sigma v. \quad (5.12)
\]

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

$$
T_n(v) - T_n^*(v) = \sum_{i=u}^{n} [Q_i(\phi^0) - n^{-1/2}v'D_i - n^{-1}v'H_i(\phi^*)v] - [Q_i(\phi^0) - n^{-1/2}v'D_i]
$$

$$
= -n^{-1}v'H_i(\phi^*)v \text{ sgn}(\log |\varepsilon_i| - n^{-1/2}v'D_i) + 2n^{-1}v'H_i(\phi^*)v - \log |\varepsilon_i| + n^{-1/2}v'D_i
$$

$$
I(0 < \log |\varepsilon_i| - n^{-1/2}v'D_i < n^{-1}v'H_i(\phi^*)v)
$$

$$
+ 2n^{-1}v'H_i(\phi^*)v - \log |\varepsilon_i| + n^{-1/2}v'D_i
$$

$$
I(n^{-1}v'H_i(\phi^*)v < \log |\varepsilon_i| - n^{-1/2}v'D_i < 0).
$$

By a similar argument for $S_n^*(v) - S_n^*_0(v) \to 0$ in Pan et al. (2007), we can obtain that $T_n(v) - T_n^*(v) \to 0$ uniformly on compact sets, which implies $\hat{T}_n(v) \to T$ on $C(R^d)$ from (5.9). By the proof of Theorem 1 in Pan et al. (2007), we obtain the result. \hfill \Box

**Proof of Theorem 4.** Based on Theorems 3 and 4 follows immediately from the following two assertions:

$$
\hat{\Sigma} \to \Sigma \quad \text{and} \quad \hat{f}(0) \to f(0).
$$

(5.13)

For the first assertion, defining

$$
\Sigma_n(\phi) = \frac{1}{n} \sum_{i=1}^{n} D_i(\phi)D_i(\phi)' \quad \text{and} \quad \hat{\Sigma}_n(\phi) = \frac{1}{n} \sum_{i=1}^{n} \hat{D}_i(\phi)\hat{D}_i(\phi)',
$$

we have

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

Obviously, $\Sigma_{3n} \xrightarrow{a.s.} \Sigma$ by the ergodic theorem. But,

$$
\Sigma_{1n} = \frac{1}{n} \sum_{i=1}^{n} [\hat{D}_i(\hat{\phi}_n) - \hat{D}_i(\phi^0)]\hat{D}_i(\hat{\phi}_n)' + \frac{1}{n} \sum_{i=1}^{n} \hat{D}_i(\phi^0)[\hat{D}_i(\hat{\phi}_n) - \hat{D}_i(\phi^0)]'
$$

$$
\leq \|\hat{\phi}_n - \phi_0\| \frac{1}{n} \sum_{i=1}^{n} \left\{ \left\| \frac{\partial \hat{D}_i(\phi^0)}{\partial \phi} \right\| \left\| \hat{D}_i(\phi^0) \right\| + \left\| \frac{\partial \hat{D}_i(\phi^0)}{\partial \phi} \right\| \left\| \hat{D}_i(\phi^0) \right\| \right\}
$$

$$
\leq \|\hat{\phi}_n - \phi_0\| \frac{2}{n} \sup_{\phi \in \Theta} \sup_{\phi \in \Theta} \left\| \frac{\partial \hat{D}_i(\phi)}{\partial \phi} \right\| \sup_{\phi \in \Theta} \| \hat{D}_i(\phi) \| \xrightarrow{a.s.} 0
$$

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

For the second assertion, defining

$$
f^*(0) = \frac{1}{nb_n} \sum_{i=1}^{n} K \left( \frac{1}{b_n} \log \frac{|X_i|}{\sigma_i(\hat{\phi}_n)} \right),
$$
we have

\[
|\hat{f}(0) - f^*(0)| \leq \frac{C}{2nb^2_n\delta} \sum_{t=1}^n \left| \log \sigma_t^2(\hat{\phi}_n) - \log \sigma_t^2(\phi_n) \right|
\]

\[
= \frac{C}{2nb^2_n\delta} \sum_{t=1}^n \log \left( 1 + \frac{\sigma_t^2(\hat{\phi}_n) - \log \sigma_t^2(\phi_n)}{\sigma_t^2(\phi_n)} \right)
\]

\[
\leq \frac{C}{nb^2_n} \sum_{t=1}^n \left| \sigma_t^2(\hat{\phi}_n) - \sigma_t^2(\phi_n) \right|
\]

\[
\leq \frac{C}{nb^2_n} \sum_{t=1}^n \sup_{\phi \in \Theta} \left| \sigma_t^2(\phi) - \sigma_t^2(\phi_n) \right| \overset{a.s.}{\to} 0,
\]

provided \(nb^2_n \to \infty\), by a proof similar to that of Lemma 6. Notice that

\[
|f^*(0) - f(0)| \leq \frac{1}{nb^2_n} \sum_{t=1}^n \left| K \left( \frac{1}{b_n} \log \left| \frac{X_t}{\sigma_t(\phi_n)} \right| \right) - K \left( \frac{\log |\varepsilon_t|}{b_n} \right) \right| + \frac{1}{nb^2_n} \sum_{t=1}^n K \left( \frac{\log |\varepsilon_t|}{b_n} \right) - f(0)
\]

\[
= L_1 + L_2.
\]

It follows from Lemma 2 and Theorem 3 that

\[
L_1 \leq \frac{C}{nb^2_n} \sum_{t=1}^n \left| \log \sigma_t^2(\hat{\phi}_n) - \log \sigma_t^2(\phi_n) \right|
\]

\[
= \frac{C}{nb^2_n} \left\| \hat{\phi}_n - \phi_n \right\| \sum_{t=1}^n \left| \frac{1}{\sigma_t^2(\phi_n)} \frac{\partial \sigma_t^2(\phi)}{\partial \phi} \right|
\]

\[
= \frac{C}{\sqrt{nb^2_n}} \left\| \hat{\phi}_n - \phi_n \right\| \cdot \frac{1}{n} \sum_{t=1}^n \left| \frac{1}{\sigma_t^2(\phi_n)} \frac{\partial \sigma_t^2(\phi)}{\partial \phi} \right|
\]

\[
= O(1) \cdot \frac{1}{\sqrt{nb^2_n}} \to 0.
\]

On the other hand, since

\[
E \left\{ \frac{1}{b_n} K \left( \frac{\log |\varepsilon_t|}{b_n} \right) \right\} = \int_{-\infty}^{\infty} K(x) f(b_n x) \, dx
\]

\[
= \int_{-\infty}^{\infty} K(x) f(0) \, dx + \int_{-\infty}^{\infty} K(x)[f(b_n x) - f(0)] \, dx = f(0) + o_p(1),
\]

it follows that

\[
E(K_2)^2 = \frac{1}{n^2} \sum_{t=1}^n E \left\{ \frac{1}{b_n} K \left( \frac{\log |\varepsilon_t|}{b_n} \right) - f(0) \right\}^2
\]

\[
+ \frac{2}{n^2} \sum_{1 \leq i < j \leq n} E \left\{ \left[ \frac{1}{b_n} K \left( \frac{\log |\varepsilon_t|}{b_n} \right) - f(0) \right] \left( \frac{1}{b_n} K \left( \frac{\log |\varepsilon_t|}{b_n} \right) - f(0) \right) \right\}
\]

\[
\leq \frac{2}{nb_n} \int_{-\infty}^{\infty} K^2(x) f(b_n x) \, dx + \frac{2}{n} f^2(0)
\]

\[
+ \frac{2}{n^2} \sum_{1 \leq i < j \leq n} E \left\{ \left[ \frac{1}{b_n} K \left( \frac{\log |\varepsilon_t|}{b_n} \right) - f(0) \right] \left[ \frac{1}{b_n} K \left( \frac{\log |\varepsilon_t|}{b_n} \right) - f(0) \right] \right\} \to 0
\]

by assumption (A4). Then, \(K_2 \overset{p}{\to} 0\). This completes the proof of the second assertion. \(\square\)
Appendix A. Probabilistic properties of PTTGARCH model

A.1. The stochastic Recurrence 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 $x_{1i}, x_{2i}$ or $\beta_i$ which is equal to 0. Denote

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

$$
Y_t = (\sigma_{t+1}^d, \ldots, \sigma_{t+q+2}^d, (X_t^+)^{2\delta}, (X_t^-)^{2\delta}, \ldots, (X_{t+p+2}^+)^{2\delta}, (X_{t+p+2}^-)^{2\delta})' \in \mathbb{R}^k,
$$

$$
B \equiv B(\phi) = (x_0, 0, \ldots, 0)' \in \mathbb{R}^k,
$$

$$
A_t \equiv A_t(\phi) = 
\begin{pmatrix}
\eta_t & \beta_2 & \ldots & \beta_{q-1} & \beta_q & x_{12} & x_{22} & \ldots & x_{1p-1} & x_{2p-1} & x_{1p} & x_{2p} \\
1 & 0 & \ldots & 0 & 0 & 0 & 0 & \ldots & 0 & 0 & 0 & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\
0 & 0 & \ldots & 1 & 0 & 0 & 0 & \ldots & 0 & 0 & 0 & 0 \\
0 & 0 & \ldots & 0 & 0 & 0 & 1 & \ldots & 0 & 0 & 0 & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\
0 & 0 & \ldots & 0 & 0 & 0 & 0 & \ldots & 1 & 0 & 0 & 0 \\
0 & 0 & \ldots & 0 & 0 & 0 & 0 & \ldots & 0 & 1 & 0 & 0 \\
\end{pmatrix}_{k \times k}
$$

(A.1)

with $k = 2p + q - 2$ and $\phi = (\delta, x_0, x_{11}, x_{21}, \ldots, x_{1p}, x_{2p}, \beta_1, \ldots, \beta_q)'$. Then, $Y_t$ is a solution of (1.2) if and only if $X_t$ is a solution of the following equation:

$$
Y_t = A_t Y_{t-1} + B.
$$

(A.3)

A.2. Strict stationarity of the model

Choose a norm $\| \cdot \|$ on $\mathbb{R}^k$, say, $\|x\| = |x_1| + \cdots + |x_d|$, for any $x \in \mathbb{R}^k$. Then for any $k \times k$ 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

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

provided $\mathbb{E} \log^+ \|A_0\|$ is finite, which is satisfied under assumption (A1).

Furthermore, under assumption (A1), we have

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

(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 (1992). Firstly, we prove that \( \log^+ ||A_0|| < +\infty \), which ensures that the Lyapunov exponent \( \gamma(\phi) \) is well defined. Due to the equivalence of norms of \( A_0 \), we have

\[
\|A_0\| \leq C(\beta_1 + \varepsilon_1 e_1^2 + \varepsilon_2 e_2^2 + \sum_{j=2}^{q} \beta_j + \sum_{t=2}^{p} (\varepsilon_{1t} + \varepsilon_{2t}) + q - 1 + |\varepsilon_t|^2 + 2p - 4) \\
\leq C(1 + |\varepsilon_t|^2). 
\]

For \( \hat{d} = \min\{1, \Delta/2\delta\} \), there exists an \( \hat{M} > 0 \) such that \( \log^+ x < x^3 \), for \( x > \hat{M} \). Then, by assumption (A1),

\[
\log^+ \|A_0\| \leq M + C(1 + E|\varepsilon_t|^4) < +\infty. 
\]  

(A.5)

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

\[
Y_0 = A_0 Y_{-1} + B \\
= A_0 A_{-1} Y_{-2} + B + A_0 B \\
= \ldots \\
= A_0 A_{-1} \cdots A_{-t} Y_{-t-1} + B + \sum_{k=0}^{t-1} A_0 \cdots A_{-k} B. 
\]

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

\[
\lim_{t \to \infty} A_0 \cdots A_{-t} e_i = 0, \quad i = 1, \ldots, \kappa. 
\]  

(A.6)

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

\[
A_{-j} e_q = \beta_q e_1, \\
A_{-j} e_{j-1} = \beta_{j-1} e_1 + e_j, \quad 2 < j \leq q, \\
A_{-j} e_k = \varepsilon_{2p} e_1, \\
A_{-j} e_{k+1} = \varepsilon_{1p} e_1, \\
A_{-j} e_{q+2j} = \varepsilon_{2j+1} e_1 + e_{q+2j+1}, \quad 1 \leq j \leq p - 2, \\
A_{-j} e_{q+2j-1} = \varepsilon_{1j+1} e_1 + e_{q+2j}, \quad 1 \leq j \leq p - 2, 
\]

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

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 \( \{Y_t, t \in \mathbb{Z}\} \) as follows:

\[
Y_t = B + \sum_{k=0}^{\infty} A_t \cdots A_{t-k} B. 
\]  

(A.7)

It is easy to verify that \( \{Y_t, t \in \mathbb{Z}\} \) is a nonnegative solution of (A.3). Let \( \sigma_t = (Y_{t-1}^2)^{1/2} \), where \( Y_{t-1} \) is the first component of \( Y_{t-1} \). Then \( X_t = \sigma_t \tilde{e}_i \) is a solution of model (1.2). The strict stationarity and ergodicity of \( X_t \) can be derived by noticing that \( \{(A_t, \tilde{e}_i), t \in \mathbb{Z}\} \) 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
\[ \|Y_0 - \tilde{Y}_0\| = \|A_0 \cdots A_{-t}(Y_{-t-1} - \tilde{Y}_{-t-1})\| = \|A_0 \cdots A_{-t}\| Y_{-t-1} - \tilde{Y}_{-t-1} \| \overset{p}{\to} 0. \]
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 $A$ denote the matrix which is obtained by replacing $(e_t^2)^{2\delta}$ and $(e_t^-)^{2\delta}$ with 0 in the matrices $A_t$ defined by (A.2). Using the notation in Bougerol and Picard (1992), we have $A_0 A_1 \cdots A_{-t} \geq A^{t+1}$. By (A.4), we have $\|A\|^{t+1} \leq \|A_0 A_1 \cdots A_{-t}\| \to 0$, which implies that the largest eigenvalue of $A$ is less than 1. On the other hand, it is easily seen that
\[ \text{Det}(\lambda I - A) = \lambda^k \left( 1 - \sum_{i=1}^{q} \beta_i \lambda^{-1} \right). \]
Since the function $g(x) = 1 - \sum_{i=1}^{q} \beta_i x$ 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. \hfill \Box

**Remark A1.** For the PTTGARCH(1,1) process, Hwang and Basawa (2004) pointed out that $\text{E} \log(\beta_1 + z_1(e_t^2)^{2\delta} + z_2(e_t^-)^{2\delta}) < 0$ could ensure that the model has a unique strictly stationary solution. In fact, it is easily verified that $\gamma(\phi) = \text{E} \log(\beta_1 + z_1(e_t^2)^{2\delta} + z_2(e_t^-)^{2\delta})$ provided $p = q = 1$, which means by Theorem 5 (i) that $\text{E} \log(\beta_1 + z_1(e_t^2)^{2\delta} + z_2(e_t^-)^{2\delta}) < 0$ 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
\[ \text{E} |X_t|^\tau < + \infty. \]

(ii) Suppose $\text{E} |e_t|^2 \delta < + \infty$. Then, model (1.2) has a stationary solution with $\text{E} |X_t|^\delta < + \infty$ if and only if
\[ \sum_{i=1}^{p} \left[ z_{1i} \text{E}(e_t^2)^{2\delta} + z_{2i} \text{E}(e_t^-)^{2\delta} \right] + \sum_{j=1}^{q} \beta_j < 1. \quad (A.8) \]

(iii) For any $k \in \mathbb{N}$, if $\text{E} |e_t|^{2k \delta} < + \infty$ and $\text{E}(\|A_t\|^k) < 1$, we have $\text{E} |X_t|^{2k \delta} < \infty$. Furthermore, if we assume in addition that
\[ \sum_{i=1}^{p} \left[ z_{1i} \left( \text{E}(e_t^2)^{2k \delta} \right)^{1/k} + z_{2i} \left( \text{E}(e_t^-)^{2k \delta} \right)^{1/k} \right] + \sum_{j=1}^{q} \beta_j < 1, \quad (A.9) \]
then
\[ \text{E} |X_t|^{2k \delta} \leq |x_0|^{\delta} \left( 1 - \sum_{i=1}^{p} \left[ z_{1i} \left( \text{E}(e_t^2)^{2k \delta} \right)^{1/k} + z_{2i} \left( \text{E}(e_t^-)^{2k \delta} \right)^{1/k} \right] - \sum_{j=1}^{q} \beta_j \right)^{-k} \text{E} |e_t|^{2k \delta}. \]

**Proof.** (i) By the definition of $\gamma(\phi)$, there exists an integer $m \geq 1$ such that $\text{E} \log \|A_0 A_1 \cdots A_{m-1}\| < 0$. From the proof of Theorem 5 (i), we know $\text{E} \|A_0\|^\delta \leq C < + \infty$. Therefore,
\[ \text{E} \|A_0 A_1 \cdots A_{m-1}\|^\delta \leq \text{E} \|A_0\|^\delta m < + \infty. \]
We introduce a function $h(x) = \text{E} \|A_0 A_1 \cdots A_{m-1}\|^x$, $0 < x < \delta$. Since $h'(0) = \text{E} \log \|A_0 A_1 \cdots A_{m-1}\| < 0$, $h(x)$ decreases in neighborhood of 0. Notice that $h(0) = 1$, so there exists a $\delta'$ such that $0 < \delta' < \min(\delta, 1)$ such that
\[ \text{E} \|A_0 A_1 \cdots A_{m-1}\|^\delta' < 1. \quad (A.10) \]
Using (A.7), we obtain that
\[
\|Y_0\|^{A^*} \leq \|B\|^{A^*} + \sum_{k=0}^{\infty} \|A_0A_1 \cdots A_{-k}\|^{A^*} \|B\|^{A^*}.
\]

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}\|^i \leq C_0\rho^k for any k, which implies that E\|Y_0\|^{A^*} < \infty. Let \( \tau = (2\delta)A^* \) and we complete the proof of (i).

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

Hence we have
\[
E\sigma_t^{2\delta} = \alpha_0 + \sum_{i=1}^{p} (x_{i1}E(e_{t-i}^{2\delta} + x_{22}E(e_{t-i}^{2\delta}) - \sum_{j=1}^{q} \beta_j E\sigma_{t-j}^{2\delta}) = \alpha_0.
\]

Because \( \{\sigma_t^{2\delta}\} \) is stationary, then
\[
\left( 1 - \sum_{i=1}^{p} (x_{i1}E(e_{t-i}^{2\delta} + x_{22}E(e_{t-i}^{2\delta})) - \sum_{j=1}^{q} \beta_j \right) E\sigma_t^{2\delta} = \alpha_0.
\]

Therefore,
\[
1 - \sum_{i=1}^{p} (x_{i1}E(e_{t-i}^{2\delta} + x_{22}E(e_{t-i}^{2\delta})) - \sum_{j=1}^{q} \beta_j > 0.
\]

Sufficiency: Suppose (A.8) holds. Define
\[
Y_t = B + \sum_{k=0}^{\infty} A_1 \cdots A_{-k} B, \quad \text{and} \quad A = E A_0.
\]

Notice that E\( A_1 \cdots A_{-k} = A^{k+1} \), we only need to prove that \( \rho(A) < 1 \). In fact,
\[
\text{Det}(\lambda I_k - A) = \lambda^k \left\{ 1 - \sum_{i=1}^{p} (x_{i1}E(e_{t-i}^{2\delta} + x_{22}E(e_{t-i}^{2\delta})) - \sum_{j=1}^{q} \beta_j \right\} \lambda^{-i} - \sum_{j=1}^{q} \beta_j \lambda^{-j}.
\]

By (A.8), we have
\[
|\text{Det}(\lambda I_k - A)| \geq 1 - \sum_{i=1}^{p} (x_{i1}E(e_{t-i}^{2\delta} + x_{22}E(e_{t-i}^{2\delta})) - \sum_{j=1}^{q} \beta_j > 0
\]

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

(iii) Applying Minkowski’s inequality to (A.7), we obtain
\[
(E\|Y_t\|^k)^{1/k} \leq \|B\| + \sum_{i=0}^{\infty} \left[ E(\|A_1 \cdots A_{i-1} B\|^k) \right]^{1/k}
\]
\[
\leq \|B\| + \|B\| \sum_{i=0}^{\infty} \left[ E(\|A_i\|^k) \right]^{1/k}
\]
\[
< +\infty,
\]

since E\( \|A_i\|^k \) < 1. Therefore, we have E\( \text{Y}_t \text{Y}_t^{2\delta} < \infty \). In the following, we assume (A.9) holds. It can be seen from (1.2) that E\( \text{Y}_t \text{Y}_t^{2k\delta} = E\sigma_t^{2k\delta} E|\epsilon_t|^{2k\delta} \). But, by Minkowski inequality,
\[
(E\sigma_t^{2k\delta})^{1/k} \leq \left( E \left\{ \alpha_0 + \sum_{i=1}^{p} (x_{i1}(X_{t-i})^{2\delta} + \sum_{i=1}^{p} (x_{22}(X_{t-i})^{2\delta} + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^{2\delta}) \right\} \right)^{1/k}
\]
\[
\leq \alpha_0 + \left[ \sum_{i=1}^{p} (x_{i1}(E(e_{t-i}^{2\delta}))^{1/k} + \sum_{i=1}^{p} (x_{22}(E(e_{t-i}^{2\delta}))^{1/k} + \sum_{j=1}^{q} \beta_j \right] (E\sigma_t^{2k\delta})^{1/k}.
\]
Then,
\[
E|X|^{2\delta} \leq \alpha_0 \left\{ 1 - \sum_{i=1}^{p} \left[ \alpha_i \left( E(e_i^+)^{2\delta} \right)^{1/k} + \alpha_2 \left( E(e_i^-)^{2\delta} \right)^{1/k} \right] - \sum_{j=1}^{q} \beta_j \right\} ^{-k} E|\alpha_1|^{2\delta}.
\]

This completes the proof of (iii). □

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

### 4. The tail behavior of the model

The following theorem shows that under some regular conditions, the tail of PTTGARCH\((p, q)\) 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 \(\alpha_{1i}, \alpha_{2i}\), and \(\beta_j\) vanish, \(i = 1, \ldots, p; j = 1, \ldots, q\). If furthermore \(e_i\) has a positive density on \(\mathbb{R}\) such that \(E|e_i|^\zeta < +\infty\) for some \(\zeta > 0\), then it follows that the limit
\[
\lim_{x \to +\infty} x^{2\alpha_0 \delta} P(X_1 > x)
\]
exists and is positive, where \(\kappa_0 = 2\kappa_1\) and \(\kappa_1\) 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

Adler, R.J., Feldman, R.E., Gallagher, C., 1997. In: Adler, R.J., Feldman, R.E., Taqqu, M. (Eds.), A user’s Guide to Heavy Tails: Statistical Techniques for Analyzing Heavy Tailed Distribution and Processes. Birkhauser, Boston, pp. 133–158.

Amemiya, T., 1985. Advanced Econometrics. Harvard University Press.

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

Bekes, I., Horváth, L., Kokoszka, P., 2003. GARCH processes: structure and estimation. Bernoulli 9 (2), 201–227.

Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics 31, 307–327.

Bougerol, P., Picard, N., 1992. Stationarity of GARCH processes and of some nonnegative time series. Journal of Econometrics 52, 115–127.

Davis, R.A., Dunsmuir, W.T.M., 1997. Least absolute deviation estimation for regression with ARMA errors. Journal of Theoretical Probability 10, 481–497.

Engle, R.F., 1982. Autoregressive conditional heteroskedasticity with estimates of variance of U.K. inflation. Econometrica 50, 987–1008.

Engle, R.F., Bollerslev, T., 1986. Modelling the persistence of conditional variance. Econometric Reviews 5, 1–50.

Gourieroux, C., 1997. ARCH Models and Financial Applications. Springer, New York.

Hall, P.C., Heyde, C.C., 1980. Martingale Limit Theory and its Applications. Academic Press, New York.

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

Higgins, M.L., Bera, A.K., 1992. A class of nonlinear ARCH models. International Economic Review 33, 137–158.

Hsieh, D.A., 1989. Testing nonlinear dependence in daily foreign exchange rates. Journal of Business 62, 339–368.

Hwang, S.Y., Basawa, I.V., 2004. Stationarity and moment structure for Box–Cox transformed threshold GARCH(1,1) processes.

Hwang, S.Y., 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.

Li, C.W., Li, W.K., 1996. On a double-threshold autoregressive heteroscedastic Time series model. Journal of Applied Econometrics 11, 253–274.

Ling, S., 2005. Self-weighted least absolute deviation estimation for infinite variance autoregressive models. Journal of the Royal Statistical Society Series B 67, 381–393.
Liu, J., 2006. On the tail behaviors of Box–Cox transformed threshold GARCH(1,1) processes. Statistics & Probability Letters 76, 1323–1330.
Mikosch, T., Straumann, D., 2006. Stable limits of martingale transforms with application to the estimation of GARCH parameters. The Annals of Statistics 34 (1), 493–522.
Mittnik, S., Rachev, S.T., 2000. Stable Paretian Models in Finance. Wiley, New York.
Pan, J.Z., Wang, H., Yao, Q., 2007. Weighted least absolute deviations estimation for ARMA models with infinite variance. Econometric Theory 23, 852–879.
Peng, L., Yao, Q., 2003. Least absolute deviation estimation for ARCH and GARCH models. Biometrika 90, 967–975.
Rydberg, T., 2000. Realistic statistical modelling of financial data. International Statististical Review 68, 233–258.
Scheinkman, J.A., LeBaron, B., 1989. Nonlinear dynamics and stock returns. Journal of Business 62, 311–337.
Shepard, N., 1996. Statistical aspects of ARCH and stochastic volatility. In: Cox, D.R., Hinkley, D.V., Barndorff-Nielsen, O.E. (Eds.), Time Series Models in Econometrics Finance and Other Fields. Chapman & Hall, London, pp. 1–67.
Silverman, B.W., 1986. Density Estimation for Statistics and Data Analysis. Chapman & Hall, London.
Weiss, A., 1986. Asymptotic theory for ARCH models: estimation and testing. Econometric Theory 2, 107–131.