Approximating Volatilities by Asymmetric Power GARCH Functions

Jeremy Penzer\textsuperscript{1} Mingjin Wang\textsuperscript{1,2} Qiwei Yao\textsuperscript{1,2}

\textsuperscript{1}Department of Statistics, London School of Economics, London WC2A 2AE, UK
\textsuperscript{2}Guanghua School of Management, Peking University, Beijing 100871, China

Abstract

ARCH/GARCH representations of financial series usually attempt to model the serial correlation structure of squared returns. While it is undoubtedly true that squared returns are correlated, there is increasing empirical evidence of stronger correlation in the absolute returns than in squared returns (Granger, Spear and Ding 2000). Rather than assuming an explicit form for volatility, we adopt an approximation approach; we approximate the $\gamma$-th power of volatility by an asymmetric GARCH function with the power index $\gamma$ chosen so that the approximation is optimum. Asymptotic normality is established for both the quasi-maximum likelihood estimator (qMLE) and the least absolute deviations estimator (LADE) estimators in our approximation setting. A consequence of our approach is a relaxation of the usual stationarity condition for GARCH models. In an application to real financial data sets, the estimated values for $\gamma$ are found to be close to one, consistent with the stylised fact that strongest autocorrelation is found in the absolute returns. A simulation study illustrates that the qMLE is inefficient for models with heavy-tailed errors, while the LADE estimation is more robust.

Key words: autoregressive conditional heteroscedasticity, financial returns, least absolute deviation estimation, leverage effects, quasi-maximum likelihood estimation, Taylor effect.
1 Introduction

Let \( \{X_t\} \) be a strictly stationary time series defined by a volatility model

\[
X_t = \sigma_t \varepsilon_t,
\]

where \( \{\varepsilon_t\} \) is a sequence of independent and identically distributed random variables with mean 0 and variance 1, \( \sigma_t \geq 0 \) is \( \mathcal{F}_{t-1} \)-measurable, and \( \mathcal{F}_{t-1} \) is the \( \sigma \)-algebra generated by \( \{X_{t-k}, k \geq 1\} \). Furthermore, we assume that \( \varepsilon_t \) is independent of \( \mathcal{F}_{t-1} \), and both \( X_t \) and \( \varepsilon_t \) have probability density functions. In financial time series, \( \{X_t\} \) is typically the (log) returns of an observed price; our aim is to explain and forecast the volatility of the returns. A GARCH model assumes the conditional second moments follow the recursive equation

\[
\sigma_t^2 = E(X_t^2 | X_{t-1}, X_{t-2}, \cdots) = \text{Var}(X_t | \mathcal{F}_{t-1}) = c + \sum_{i=1}^{p} b_i X_{t-i}^2 + \sum_{j=1}^{q} a_j \sigma_{t-j}^2,
\]

where \( c > 0 \) and \( b_i, a_j \) are non-negative; see Engle (1982), Bollerslev (1986) and Taylor (1986, chapter 3). Under the condition \( \sum_i b_i + \sum_j a_j < 1 \), (1.2) admits the representation

\[
\sigma_t^2 = E(X_t^2 | X_{t-1}, X_{t-2}, \cdots) = d_0 + \sum_{j=1}^{\infty} d_j X_{t-j}^2,
\]

where \( d_i \geq 0 \) are some constants; see, for example, (4.35) of Fan and Yao (2003). Thus, a GARCH model effectively assumes a linear autoregressive structure for the squared returns \( X_t^2 \). Therefore the stronger the autocorrelation of \( X_t^2 \) is, the better \( \sigma_t^2 \) would be explained by \( X_{t-1}^2, X_{t-2}^2, \cdots \) for a correctly specified GARCH model. While most financial squared returns are significantly auto-correlated, such an autocorrelation is typically weak. On the other hand, there is growing empirical evidence that stronger autocorrelation exists in other functions of returns; see Granger, Spear, and Ding (2000) and references therein. In fact, absolute returns \( |X_t| \) often exhibit stronger autocorrelation than squared returns. Furthermore, we may question whether a linear autoregressive structure for \( X_t^2 \) is realistic.

This paper puts forward approximations to the volatility function that exploit the stronger autocorrelation in the \( \gamma \)-th power of absolute returns, for some \( \gamma \in \)
We do not impose any explicit form on \( \sigma_t \). Instead we seek the index \( \gamma \) for which a GARCH-like model provides the *best* approximation for \( \sigma_t^\gamma \). More specifically, we approximate \( \sigma_t^\gamma \) by an asymmetric GARCH function

\[
\xi_{t,\gamma} = c + \sum_{i=1}^{p} b_i \{ |X_{t-i}| - d_i X_{t-i} \}^\gamma + \sum_{j=1}^{q} a_j \xi_{t-j,\gamma}
\]

(1.4)

where the parameters \( c, b_i, a_j \) are non-negative, and \( d_i \in (-1, 1) \). We then choose \( \gamma \in (0, 2] \) so that the approximation is *optimum* in a certain sense; see section 2.2.

The restriction \( \gamma \leq 2 \) is not essential and is imposed to avoid higher order moment conditions on \( X_t \). The presence of asymmetric parameters \( d_i \) is to reflect the so-called *leverage effect* in financial returns; see Ding, Engle, and Granger (1993). Proposition 1 in Appendix A indicates that equation (1.4) admits a unique strictly stationary solution

\[
\xi_{t,\gamma} = \frac{c}{1 - \sum_{j=1}^{q} a_j} + \sum_{i=1}^{p} b_i \{ |X_{t-i}| - d_i X_{t-i} \}^\gamma
\]

(1.5)

\[
+ \sum_{i=1}^{p} b_i \sum_{k=1}^{\infty} \sum_{j_1=1}^{q} \cdots \sum_{j_k=1}^{q} a_{j_1} \cdots a_{j_k} \{ |X_{t-i-j_1-\cdots-j_k}| - d_i X_{t-i-j_1-\cdots-j_k} \}^\gamma
\]

with \( E(\xi_{t,\gamma}) < \infty \), provided that \( \{X_t\} \) is strictly stationary with \( E|X_t|^\gamma < \infty \), and \( \theta \equiv (c, b_1, \cdots, b_p, a_1, \cdots, a_q, d_1, \cdots, d_p)^\tau \in \Theta \), where

\[
\Theta = \left\{ (c, b_1, \cdots, b_p, a_1, \cdots, a_q, d_1, \cdots, d_p) \mid c, b_i, a_j > 0, \ d_i \in [-1+\delta_0, 1-\delta_0], \ \sum_{j=1}^{q} a_j < 1 \right\},
\]

(1.6)

and \( \delta_0 \in [0, 1) \) is a small constant. We restrict \( d_i \) to be in a closed interval contained in \((-1, 1)\) to avoid some technical difficulties; see (C.5) in appendix C.

Attempts to make use of the stronger autocorrelation of power functions of returns for modelling volatility may be traced back to Ding *et al* (1993). In fact, the asymmetric power GARCH (APGARCH) model proposed by Ding *et al* (1993) is (1.4) with \( \xi_{t-j,\gamma} \) replaced by \( \sigma_t^\gamma \) for \( j = 0, 1, \cdots, q \); see also (B.1) and proposition 2 in appendix B. Hence an APGARCH model assumes that the \( \gamma \)-th power of the
volatility function $\sigma_t^\gamma$ is of the form of the right hand side of (1.5). We argue that the approximation paradigm adopted in this paper has at least two advantages over the assumption of an exact APGARCH model. First, it brings the relevant statistical theory one step closer to reality since any statistical model is merely an approximation under most circumstances. Second, the condition for the stationarity has been relaxed from $\sum_{i=1}^p b_i + \sum_{j=1}^q a_j < 1$ (proposition 2 in appendix B) for APGARCH models to the condition $\sum_{1 \leq j \leq q} a_j < 1$ (proposition 1 in appendix A) for APGARCH approximation. This relaxation is of practical relevance since the estimated $b_i$ and $a_i$ for financial data often sum up to 1 or beyond. By accepting that our model is only an approximation, we admit explicitly the possibility that parameters beyond the admissible bound may result from inadequacy of the model, in addition to the possibility of a non-stationary data generating process. The relaxation of the stationarity condition is due to the fact that the approximation process $\xi_t, \gamma$ is defined and caused by $X_t$ but not vise verse.

Statistical inference for the GARCH model and its variants is predominantly quasi-maximum likelihood estimation (qMLE), facilitated by treating $\varepsilon_t$ as a normal random variable. It is well documented that when $\varepsilon_t$ is heavy-tailed such as $E(\varepsilon_t^4) = \infty$, the qMLE for GARCH models suffers from slow convergence rates and complicated asymptotic distributions (Hall and Yao 2003), (Mikosch and Straumann 2003), (Straumann and Mikosch 2003), (Straumann 2005). On the other hand, least absolute deviations estimation (LADE) based on a log-transformation enjoys standard root-$n$ convergence rate regardless of whether $\varepsilon_t$ is heavy-tailed or not (Peng and Yao 2003); see also Horvath and Liese (2004). We consider both qMLE and LADE for parameters $c, b_i, d_i, a_j$ and $\gamma$ in (1.4) in section 2. In addition, a new estimator for $\gamma$ is proposed, based on minimizing serial dependence in the residuals from the fitted volatility function. The asymptotic properties of the estimators for $c, b_i, d_i, a_j$ are presented in section 3. In considering the asymptotic properties of qMLE and LADE for GARCH models, existing work assumes an exact model for the volatility function; see for example, Hall and Yao (2003), Peng and Yao (2003) and Mikosch and Straumann (2003). The asymptotic theory in section 3 is new; we consider estimators of the parameters of an optimal approximation to the volatility function rather than estimators of the parameters of the volatility function itself.

Application of our method to four financial return series in section 4 indicates
that a better approximation to the volatility function is obtained by using the \( \gamma \)-th power in place of the squared returns. The fact that estimates of \( \gamma \) are always close to 1 coincides with empirical evidence indicating that the strongest autocorrelation is found in absolute returns. Note that a \( \gamma \)-th power GARCH model implies a linear autoregressive structure for \(|X_t|^\gamma\); cf. (1.3). A larger autocorrelation of \(|X_t|^\gamma\) leads to a better (linear) autoregressive fitting. Ding, Engle, and Granger (1993) observe that qMLE for \( \gamma \) is inefficient when \( \varepsilon_t \) is heavy-tailed. A simulation study in section 5 confirms this observation and indicates that LADE is robust to the distribution of errors.

An APGARCH model may be viewed as a member of the so-called augmented GARCH class of Duan (1997). Theoretical properties such as stationarity, mixing properties, and higher order moment properties for APGARCH models are studied by, among others, He and Teräsvirta (1999), Carrasco and Chen (2002), and Ling and McAleer (2002). Applications of APGARCH models are reported in McKenzie and Mitchell (2002), Conrad and Karanasos (2002) and Brooks, Faff, McKenzie, and Mitchell (2003). Hagerud (1997) considers a statistical test for asymmetry under APGARCH models. We derive a simple condition for stationary APGARCH processes in appendix B, which includes a result of Ling and McAleer (2002) as a special case.

2 Methodology

2.1 Estimation of \( c, b_i, a_j, d_i \) for a given \( \gamma \)

2.1.1 Least absolute deviations estimator

To facilitate the LADE, we adopt a different parametrization. Namely we drop the assumption \( E\varepsilon_t^2 = 1 \) in (1.1). Instead we assume that the median of \(|\varepsilon_t|\) is equal to 1. Hence the median of \(|\varepsilon_t|^\gamma\) is equal to 1 for any \( \gamma > 0 \). Note that \( \sigma_t \) defined in (1.1) differs under the two parameterisations by a constant independent of \( t \). This affects the parameters in \( \xi_{t,\gamma} \) as follows; \( c \) and all \( b_i \) differ by a common multiplicative constant under the two parameterisation while \( d_i \) and \( a_j \) remain unchanged.

Let \( X_1, \cdots, X_n \) be observations. First we assume \( \gamma \) is known. Then an estimator
for $\theta$ is obtained by the least absolute deviations method as follows:

$$\tilde{\theta} \equiv \tilde{\theta}^{(\gamma)} = \arg \min_{\theta} \sum_{t=1}^{n} \left| X_t^\gamma - c - \sum_{i=1}^{p} b_i |X_{t-i}| - d_i X_{t-i}^\gamma - \sum_{j=1}^{q} a_j \xi_{t-j,\gamma}(\theta) \right|, \quad (2.1)$$

where $\nu \equiv \nu_n > 1$ is a large integer, and $\xi_{t,\gamma}(\theta) \equiv \xi_{t,\gamma}$ defined in (1.5). In practice, we let $X_k \equiv 0$ for any $k \leq 0$ in (1.5). The sum in the above expression is taken from $t = \nu$ to alleviate the effect of this truncation. See condition (A5) below.

We see from (1.1) that $|X_t^\gamma| = \sigma_t^{\gamma} + \tilde{\epsilon}_t^\gamma (|\epsilon_t|^{\gamma-1}) \equiv \sigma_t^{\gamma} + \epsilon_t$, and the conditional median of $\epsilon_t$ is 0 under the specified parametrisation. Hence $\sigma_t^{\gamma} = \arg \min_{a} E\{ |X_t^\gamma - a| \mid F_{t-1} \}$. Furthermore when $\sigma_t^{\gamma} = \xi_{t,\gamma}(\theta^0)$, it holds almost surely that

$$\theta^0 = \arg \min_{\theta} E\{ |X_t^\gamma - \xi_{t,\gamma}(\theta)| \mid F_{t-1} \} = \arg \min_{\theta} E\{ |X_t^\gamma - \xi_{t,\gamma}(\theta)| \}.$$  

This motivates the estimator (2.1). Note that $\{\epsilon_t\}$ is not a sequence of independent random variables and its (conditional) heteroscedasticity may compromise the performance of $\tilde{\theta}$. However, if we define $\epsilon_t^\dagger = \log(|\epsilon_t|) = \log(|X_t|) - \gamma^{-1} \log(\sigma_t^{\gamma})$ then $\epsilon_t^\dagger$ has median 0 and $\{\epsilon_t^\dagger\}$ is an i.i.d. sequence. Therefore, it holds that

$$\sigma_t^{\gamma} = \arg \min_{a>0} E\{ \log |X_t| - \frac{1}{\gamma} \log a \mid F_{t-1} \}.$$  

This leads to the estimator

$$\hat{\theta}_1 \equiv \hat{\theta}_1^{(\gamma)} = \arg \min_{\theta} \sum_{t=1}^{n} \left| \log |X_t| - \frac{1}{\gamma} \log \left\{ c + \sum_{i=1}^{p} b_i |X_{t-i}| - d_i X_{t-i}^\gamma + \sum_{j=1}^{q} a_j \xi_{t-j,\gamma}(\theta) \right\} \right|,$$

$$= \arg \min_{\theta} \sum_{t=1}^{n} \left| \log |X_t| - \frac{1}{\gamma} \log \left\{ \xi_{t,\gamma}(\theta) \right\} \right|, \quad (2.2)$$

where $\xi_{t,\gamma}$ is given in (1.5). Peng and Yao (2003) showed that in the context of estimation for GARCH models, the estimators of the type $\hat{\theta}_1$ enjoy better asymptotic properties than those of type $\tilde{\theta}$ in the sense that $\hat{\theta}_1$ is asymptotically unbiased while $\tilde{\theta}$ is typically a biased estimator. See also Theorem 1 below.

### 2.1.2 Quasi-maximum likelihood estimation

An approximate qMLE may also be entertained based on an additional assumption that $\epsilon_t$ in (1.1) are independent $N(0,1)$ random variables, therefore is constructed under the standard parametrisation implied by $E \epsilon_t^2 = 1$. The resulting estimator is

$$\hat{\theta}_2 \equiv \hat{\theta}_2^{(\gamma)} = \arg \min_{\theta} \sum_{t=1}^{n} \left[ X_t^2 / \{ \xi_{t,\gamma}(\theta) \}^{2/\gamma} + 2 \gamma^{-1} \log \{ \xi_{t,\gamma}(\theta) \} \right]. \quad (2.3)$$
2.2 Estimation of $\gamma$

The estimators $\hat{\theta}_1^{(\gamma)}$ and $\hat{\theta}_2^{(\gamma)}$ naturally facilitate estimation of $\gamma$. For example, with the least absolute deviations estimator $\hat{\theta}_1^{(\gamma)}$, we may choose $\gamma \in (0, 2]$ which minimises

$$
\sum_{t=\nu}^{n} D_t(\hat{\theta}_1^{(\gamma)}, \gamma),
$$

where

$$
D_t(\theta, \gamma) = \left| \log |X_t| - \frac{1}{\gamma} \log \{\xi_{t, \gamma}(\theta)\} \right|.
$$

With the MLE $\hat{\theta}_2^{(\gamma)}$, we may treat $\gamma$ as an additional parameter and estimate it by maximising the profile likelihood function derived from (2.3).

Our goal is to estimate the volatility function $\sigma_t$; a good estimate should ensure the residuals $\hat{\varepsilon}_t = X_t/\hat{\sigma}_t$ contain little information about $F_{t-1}$, where $\hat{\sigma}_t$ denotes an estimator for $\sigma_t$. We construct an alternative method for estimating $\gamma$ based on this idea. Let $\hat{\theta}^{(\gamma)}$ be an estimator for the parameters $\theta$ of $\xi_{t, \gamma}$, which may be either $\hat{\theta}_1^{(\gamma)}$ or $\hat{\theta}_2^{(\gamma)}$. Define residuals

$$
\hat{\varepsilon}_t^{(\gamma)} = X_t/\{\xi_{t, \gamma}(\hat{\theta}^{(\gamma)})\}^{1/\gamma}, \quad t = \nu, \ldots, n,
$$

(2.4)

If $\hat{\varepsilon}_t^{(\gamma)}$ is a good estimator for $\varepsilon_t$, $E\{|\hat{\varepsilon}_t^{(\gamma)}| I(A)\} \approx E|\varepsilon_t^{(\gamma)}| P(A)$ for any $A \in F_{t-1}$. This suggests a choice of $\hat{\gamma} \in (0, 2]$ which minimises

$$
R(\gamma) \equiv \sup_{B \in \mathcal{B}} \frac{1}{n - \nu + 1} \left| \frac{1}{n} \sum_{t=\nu}^{n} \{\hat{\varepsilon}_t^{(\gamma)} - \bar{\varepsilon}^{(\gamma)}\} I(X_t \in B) \right|,
$$

(2.5)

where $\bar{\varepsilon}^{(\gamma)}$ is the sample mean of $\{|\hat{\varepsilon}_t^{(\gamma)}|\}$, $X_t = (X_{t-1}, \ldots, X_{t-k})^T$ for some prescribed integer $k \geq 1$, and $\mathcal{B}$ consists of some subsets in $\mathbb{R}^k$. Statistics of this type have been used for model checking by, for example, Stute (1997), Chen and An (1997), Koul and Stute (1999), and Polonik and Yao (2005). In practice, we may use either the LADE $\hat{\theta}_1^{(\gamma)}$ or the qMLE $\hat{\theta}_2^{(\gamma)}$ as $\hat{\theta}^{(\gamma)}$ in (2.4). We may choose $\mathcal{B}$ consisting of the sets with balls centered at the origin as their images under the mapping $x \to S^{-1/2}(x - \bar{X})$, where $\bar{X}$ and $S$ denote, respectively, the sample mean and the sample covariance matrix of $\{X_t\}$. When the distribution of $X_t$ is symmetric, those sets are approximately the minimum-volume sets (Polonik and Yao 2005).
Without assuming a true model, the so-called true value of $\gamma$ needs to be clarified. From (2.5), the value to be estimated by $\hat{\gamma}$ is

$$\gamma^0 = \arg \min_{\gamma \in (0, 2]} \left( \sup_{B \in B} \left| E[\{\|\varepsilon_t\| - E[\varepsilon_t]\}I(X_t \in B)] \right| \right),$$

which is assumed to be unique. When $X_t$ is indeed an APGARCH process, $\gamma^0$ is the true value of the power index.

3 Theoretical properties

We always assume in this section that $\gamma \in (0, 2]$ is known. The asymptotic properties of the estimator $\hat{\gamma}$ is more complicated and will be investigated in a follow-up paper.

3.1 Asymptotic normality of LADEs

We introduce some notation first. Let $U_t(\theta)$ be the derivative of $\xi_t(\gamma)(\theta)$ with respect to $\theta$. Then it holds for $\theta \in \Theta$ that

$$E\{|U_t(\theta)/\xi_t(\gamma)(\theta)|^k\} < \infty, \quad \text{for any } k > 0, \quad 1 \leq \ell \leq 2p + q + 1, \quad (3.1)$$

see the first paragraph in appendix C below. In the expression above, $U_t$ denotes the $\ell$-th component of $U_t$. Define $Z_t(\theta) = \log |X_t| - \gamma^{-1} \log \{\xi_t(\gamma)(\theta)\}$. Then the derivative of $Z_t$ with respect to $\theta$ is $\dot{Z}_t(\theta) = -U_t(\theta)/\{\gamma \xi_t(\gamma)(\theta)\}$. Put

$$\Sigma = \sum_{k=-\infty}^{\infty} E\left\{\dot{Z}_t(\theta^0)\dot{Z}_{t+k}(\theta^0)^\tau \text{sgn} \{Z_t(\theta^0)Z_{t+k}(\theta^0)\} \right\}, \quad \Sigma_0 = E\{\dot{Z}_t(\theta^0)\dot{Z}_t(\theta^0)^\tau | Z_t(\theta^0) = 0\},$$

where $\theta^0$ is specified in condition (A2) below.

Some regularity conditions are now in order.

(A1) The process $\{X_t\}$ is strictly stationary and $\alpha$-mixing with the mixing coefficients satisfying condition $\lim_{n \to \infty} n^{8+\epsilon_0} \alpha(n) = 0$ for some $\epsilon_0 > 0$. Furthermore, $E|X_t|^\gamma < \infty$.

(A2) There exists a unique $\theta^0 \equiv \theta_{\gamma} \in \Theta$ for which

$$\theta^0 = \arg \min_{\theta} E\left[ |\log |X_t| - \frac{1}{\gamma} \log \{\xi_t(\gamma)(\theta)\}| \right]. \quad (3.2)$$
(A3) The matrix $\Sigma_0$ is nonsingular.

(A4) The density function $f$ of $\gamma Z_t(\theta^0)$ is positive and continuous at 0. The conditional density function $g(z|u)$ of $Z_t(\theta^0)$ given $\hat{Z}_t(\theta^0) = u$ is uniformly Lipschitz continuous at $z = 0$ in the sense that

$$|g(z|u) - g(0|u)| \leq C|z|, \text{ for all } |z| < \delta_1,$$

where $C, \delta_1 > 0$ are constants, and $C$ does not depend on $u$. Further, $\sup_u g(0|u) < \infty$.

(A5) As $n \to \infty$, $\nu/n \to 0$ and $\nu/\log n \to \infty$.

The mixing condition in (A1) is required to establish asymptotic normality. Together with (3.1), it also ensures that $\Sigma$ is well-defined; see Theorem 2.20(i) of Fan and Yao (2003). When $\sigma_t^\gamma \equiv \xi_{t,\gamma}(\theta^0)$, we may drop this mixing assumption, since the asymptotic normality is entailed by the resulting martingale differences structure (Davis and Dunsmuir 1997; Peng and Yao 2003). On the other hand, the condition for an APGARCH($p,q$) process to be strictly stationary is given in proposition 2 in appendix B. Proposition 5 of Carrasco and Chen (2002) characterises the condition for $\beta$-mixing APGARCH($p,q$) processes with exponential decaying coefficients, which implies the $\alpha$-mixing. The assumption of positive parameters in (A2) ensures the property (3.1); see also (1.6). Similar conditions are employed by, for example, Hall and Yao (2003) and Peng and Yao (2003). Note that $Z_t(\theta^0) = \log |\varepsilon_t|$ in the case where $\sigma_t^\gamma \equiv \xi_{t,\gamma}(\theta^0)$; (A4) can then be replaced by the condition that the density function of $\log |\varepsilon_t|$ is continuous at zero. (A5) requires $\nu \to \infty$ at appropriate speeds as $n \to \infty$, which ensures that the truncation $X_t \equiv 0$ for all $t \leq 0$ does not alter the asymptotic property of the estimator.

**Theorem 1.** Let conditions (A1) – (A5) holds and $\delta_0 \in (0,1)$ in (1.6). For any positive random variable $M > 0$, there exists a local minimiser $\hat{\theta}_1$ defined by (2.2) but with the minimization taken over $||\theta - \theta^0 - \eta/\sqrt{n}|| \leq M/\sqrt{n}$ only and $X_k \equiv 0$ for all $k \leq 0$, for which

$$n^{1/2}(\hat{\theta}_1 - \theta^0) \to N(0, \Sigma_0^{-1} \Sigma \Sigma_0^{-1}/(2\gamma f(0))^2) \quad (3.3)$$

in distribution, where $\eta \sim N(0, \Sigma_0^{-1} \Sigma \Sigma_0^{-1}/(2\gamma f(0))^2)$ is a random vector.
Remark 1. In the case that $\sigma^\gamma_t \equiv \xi_{t,\gamma}(\theta^0)$, condition (A1) may be removed while the condition $\sum_j a_j < 1$ implied implicitly in (A2) should be replaced by the condition $\sum_i b_i E(|\varepsilon_t| - d_i \varepsilon_t)^\gamma + \sum_j a_j < 1$; see proposition 2 in appendix B. The latter ensures that the equations (1.1) and (1.4), with $\xi_{t,\gamma}(\theta^0)$ replaced by $\sigma^\gamma_t$, defines a unique stationary solution $\{X_t\}$ with $E(|X_t|) < \infty$. Now $[\dot{Z}_t(\theta^0) \text{sgn}\{Z_t(\theta^0)\}]$ is a martingale difference, and $\Sigma_0 = \Sigma = E[\dot{Z}_t(\theta^0)\dot{Z}_t(\theta^0)^\gamma]$.

Remark 2. Kernel-based estimation of covariance matrix such as $\Sigma$ above has been discussed by Newey and West (1987), Newey and West (1994), Andrews (1991) and Andrews and Monahan (1992); see also Wooldridge (1994). For instance, a simple Newey-West’s Bartlett kernel estimator has the form

$$\hat{\Sigma} = \hat{\Gamma}_0 + \sum_{j=1}^{L_T} (1 - \frac{j}{L_T + 1})(\hat{\Gamma}_j + \hat{\Gamma}_j')$$

(3.4)

where $\hat{\Gamma}_j = 1/T \sum_{t=1}^T \dot{Z}_t(\theta^0)\dot{Z}_{t+j}(\theta^0)^\gamma \text{sgn}\{Z_t(\theta^0)Z_{t+j}(\theta^0)\}$, $j = 0, 1, 2, \cdots$ are the sample covariance matrices. $L_T$ is called the bandwidth of the kernel (Newey and West, 1987). In practice, $\Sigma_0$ may be estimated through some non-parametric regression methods, such as Nadaraya-Watson estimator. Moreover, $f(0)$ can be given straightway by the kernel density estimation of $\gamma Z_t(\hat{\theta}_1)$ at 0.

The proof of theorem 1 is given in appendix C.

3.2 Asymptotic normality of qMLEs

Although we continue to use the notation defined above, the parameters are now defined under a different parametrisation entailed by the condition $E(\varepsilon_t^2) = 1$; see the discussion in section 2.1.2.

Write $\dot{U}_t(\theta) = \partial U_t(\theta)/\partial \theta^\tau$. Put

$$\Sigma_1 = \sum_{k=-\infty}^{\infty} E\left[\frac{\dot{U}_t(\theta^0)\dot{U}_{t+k}(\theta^0)^\tau}{\xi_{t,\gamma}(\theta^0)\xi_{t+k,\gamma}(\theta^0)}\left\{\frac{X_t^2}{\xi_{t,\gamma}(\theta^0)^2/\gamma} - 1\right\}\left\{\frac{X_{t+k}^2}{\xi_{t+k,\gamma}(\theta^0)^2/\gamma} - 1\right\}\right],$$

$$\Sigma_2 = E\left[\left\{(1 + \frac{2}{\gamma})\frac{X_t^2}{\xi_{t,\gamma}(\theta^0)^2/\gamma} - 1\right\}\frac{\dot{U}_t(\theta^0)\dot{U}_t(\theta^0)^\tau}{\xi_{t,\gamma}(\theta^0)^2} + \left\{1 - \frac{X_t^2}{\xi_{t,\gamma}(\theta^0)^2/\gamma}\right\}\dot{U}_t(\theta^0)\right],$$

where $\theta^0$ is given in (B2) below. We list some regularity conditions now.
(B1) The process \( \{X_t\} \) is strictly stationary and \( \alpha \)-mixing with the mixing coefficients satisfying condition \( \sum_{j \geq 1} j^{2+\epsilon_0} \alpha(j)^{1-2/\delta} < \infty \) for some \( \epsilon_0 > 0 \). Furthermore, \( E|X_t|^{2\delta} < \infty \), where \( \delta > 2 \) is a constant.

(B2) Condition (A2) holds with (3.2) replaced by
\[
\theta^0 = \arg \min_{\theta} E\left[ X_t^2 / \{\xi_{t,\gamma}(\theta)\}^{2/\gamma} + 2\gamma^{-1} \log \{\xi_{t,\gamma}(\theta)\} \right].
\]

(B3) The matrix \( \Sigma_2 \) is nonsingular.

**Theorem 2.** Under conditions (B1) – (B3) and (A5), there exists a local minimiser \( \hat{\theta}_2 \) within radius \( r \) of \( \theta^0 \) for which
\[
n^{1/2}(\hat{\theta}_2 - \theta^0) \rightarrow N(0, \Sigma_2^{-1}\Sigma_1\Sigma_2^{-1})
\]
in distribution, as \( n \rightarrow \infty \), where \( r > 0 \) is a sufficiently small but fixed constant.

**Remark 3.** In case that \( \sigma^2_t \equiv \xi_{t,\gamma}(\theta^0) \), condition (B1) may be replaced by condition \( E(\varepsilon_t^4) < \infty \) while the condition \( \sum_j a_j < 1 \) in (1.6) be replaced by the condition \( \sum_i b_i E(|\varepsilon_t| - d_i \varepsilon_t)^\gamma + \sum_j a_j < 1 \); see also remark 1. Now note
\[
\Sigma_1 = \{E(\varepsilon_t^4) - 1\} E\left[ \frac{U_t(\theta^0)U_t(\theta^0)^\tau}{\xi_{t,\gamma}(\theta^0)^2} \right], \quad \Sigma_2 = \frac{2}{\gamma} E\left[ \frac{U_t(\theta^0)U_t(\theta^0)^\tau}{\xi_{t,\gamma}(\theta^0)^2} \right].
\]

Especially when \( \gamma = 2 \), the above theorem reproduces Theorem 2.1(a) of Hall and Yao (2003). See also Berkes, Horváth, and Kokoszka (2003) and Straumann and Mikosch (2003).

**Remark 4.** Comparing theorems 1 and 2, we can see that the asymptotic normality for the qMLE requires higher order moment conditions than that for the LADE. In fact, the condition that \( E(|\varepsilon_t|^{4-\epsilon}) < \infty \) for any \( \epsilon > 0 \) is necessary for the asymptotic normality of \( \hat{\theta}_2 \) (Hall and Yao 2003), and is not required for that of \( \hat{\theta}_1 \).

We omit the proof of theorem 2, since it is technically less involved than that of theorem 1, and is similar to the proof of theorem 5.1(a) of Hall and Yao (2003).

## 4 Real data examples

This section applies the volatility approximating procedures of section 2 to the returns of two real financial data sets; namely the daily closing prices of S&P500 index
in 3 January 1928 — 30 August 1991 analyzed extensively by Ding et al. (1993), and the daily closing prices of the IBM stock in 3 January 1962 – 30 December 1997 analyzed in Tsay (2001). Returns are defined as \(R_t = \log(p_t) - \log(p_{t-1})\), where \(p_t\) is the price or the index at time \(t\). See Figure 1 (a) and (b) for the plots of these two time series.

Ding et al (1993) compare the auto-correlation functions of \(|R_t|^\gamma\) with different \(\gamma\)-values and found that absolute returns (i.e. with \(\gamma = 1\)) are the most autocorrelated series. Figure 1 (c) and (e) show the sample autocorrelations of the squared return and absolute return of S&P500 data, respectively. The later obviously has a much stronger autocorrelation structure than the former. Similar phenomena has been observed in Figure 1 (d) and (f) for the returns of IBM stock. For further empirical evidence of the stronger autocorrelation of absolute returns, see Granger and Ding (1995) in which this phenomenon is called the Taylor effect after Taylor (1986). To explore this effect in modelling volatilities, Ding et al (1993) fitted an APGARCH model to the S&P500 data using qMLE method and obtained 1.43 as an estimate for \(\gamma\).

We apply the method proposed in section 2 to approximate the conditional volatility of the mean-adjusted returns \(X_t = R_t - \bar{R}\), where \(\bar{R}\) is the sample mean. We take \(X_t = \sigma_t \epsilon_t\) and approximate \(\sigma_t^\gamma\) by an asymmetric power GARCH(1,1) function,

\[\xi_{t,\gamma} = c + b_1 \{ |X_{t-1}| - d_1 X_{t-1} \}^\gamma + a_1 \xi_{t-1,\gamma}.\]

We set \(\nu = 21\). For each of \(\gamma = 0.1, 0.2, \cdots, 1.9, 2.0\), we estimate \(c, a_1, b_1, d_1\) by LAD and calculate the \(R(\gamma)\). Plots of \(R(\gamma)\) with \(k = 2\) for these two data sets are given in Figure 2. For the S&P500 data, \(R(\gamma)\) achieves its minimum value at \(\gamma = 0.9\), while for the IBM data, the minimum point of \(R(\gamma)\) is at \(\gamma = 1.2\). Results for other \(k\)-values are similar and are not reported here to save the space. The LAD estimates and their standard errors are listed in table 1. Note that, for the S&P data, our \(\gamma\) estimate is substantially smaller than that obtained by Ding et al (1993).

5 Simulation study

The results of section 4 suggest that qML may overestimate the power parameter \(\gamma\). In this section we perform a simulation study to verify this observation. We
choose $\gamma = 1$ and set the other parameter values to be close to those fitted to the S&P500 data by full-LAD, that is, $a_1 = 0.9$, $b_1 = 0.05$, $c = 10^{-4}$ and $d_1 = 0.5$.

We simulate 500 instances of an APGARCH(1,1) process with 1000 observations and $t_3$ distributed errors. Here $t_3$ denotes a $t$-distribution on 3 degrees of freedom. All parameters are estimated for both qML and full-LAD objective functions. We take $\nu = 21$ as in section 4 and, in order to ensure fair comparison, optimisation is performed by golden section search in both cases. The experiment is repeated with $t_4$ and standard normal errors.

Figure 3 shows boxplots of the estimates of the power parameter $\gamma$ across three error distributions for both estimation methods. For error distributions with heavy tails, that is, $t_3$ and $t_4$ it is clear that LAD out-performs qML. There is a marked worsening of qML performance going from $\varepsilon_t \sim t_4$ to $\varepsilon_t \sim t_3$, that is, a marked worsening as the weight in the tails of the error distribution increases. Figure 3 also provides evidence of slight bias in qML estimates for $\gamma$ when the error distribution is non-Gaussian. In both $\varepsilon_t \sim t_4$ and $\varepsilon_t \sim t_3$ cases, over 55% of the mass of the empirical distribution for qML estimator is above the true value, $\gamma = 1$. Similar behaviour is seen across estimates for ARCH, GARCH and asymmetry parameters. The performance of LAD is robust to the distribution of the errors while qML is inefficient for heavy tailed distributions.

A Stationary APGARCH approximation

**Proposition 1.** Let $\{X_t\}$ be a strictly stationary process with $E|X_t|^\gamma < \infty$, and $\{\varepsilon_t\}$ be a sequence of independent and identically distributed random variables. Let $\theta \in \Theta$ given in (1.6) with $\delta_0 \in [0, 1)$. Then $\xi_{t,\gamma}$ defined in (1.5) is the unique strictly stationary solution of equation (1.4) with $E|\xi_{t,\gamma}| < \infty$.

**Proof.** For $d_i \in [-1, 1]$, $E|X_{t-i-j_1-...-j_k}|^\gamma \{1 - d_i \text{sgn}(\varepsilon_{t-i-j_1-...-j_k})\}^\gamma \leq 2^\gamma E|X_t|^\gamma$. Hence the expectation of the multiple sum on the RHS of (1.5) is bounded from above by

$$2^\gamma E|X_t|^\gamma \sum_{i=1}^p b_i \sum_{j=1}^q a_j / (1 - \sum_{j=1}^q a_j).$$

Since all the terms are non-negative, the infinite sum on the RHS of (1.5) converges almost surely to a random variable with finite expectation. Hence $\xi_{t,\gamma}$ defined by
(1.5) is a well-defined strictly stationary process with \( E(\xi_{t,\gamma}) < \infty \). Now substituting \( \xi_{t-j,\gamma} \) on the RHS of (1.4) by (1.5) leads to the RHS of (1.5). Therefore \( \xi_{t,\gamma} \) defined in (1.5) is a solution of (1.4).

To prove the uniqueness, let \( \{\xi'_{t,\gamma}\} \) be a strictly stationary solution of (1.4) with \( E|\xi'_{t,\gamma}| < \infty \). For any integer \( \ell \geq 1 \), we iterate (1.4) \( \ell \) times and it leads to

\[
\xi'_{t,\gamma} = c \sum_{k=0}^{\ell} \left( \sum_{j=1}^{q} a_j \right)^k + \sum_{i=1}^{p} b_i |X_{t-i}|^\gamma \{1 - d_i \text{sgn}(\varepsilon_{t-i})\}^\gamma \\
+ \sum_{i=1}^{p} b_i \sum_{k=1}^{\ell} \sum_{j_1=1}^{q} \cdots \sum_{j_k=1}^{q} a_{j_1} \cdots a_{j_k} |X_{t-i-j_1-\cdots-j_k}|^\gamma \{1 - d_i \text{sgn}(\varepsilon_{t-i-j_1-\cdots-j_k})\}^\gamma \\
+ \sum_{j_1=1}^{q} \cdots \sum_{j_k=1}^{q} a_{j_1} \cdots a_{j_k} \xi'_{t-j_1-\cdots-j_k,\gamma}.
\]

Hence

\[
E|\xi_{t,\gamma} - \xi'_{t,\gamma}| \leq \left( \sum_{j=1}^{q} a_j \right) \left\{ \frac{c}{1 - \sum_{j=1}^{q} a_j} + 2^\gamma E(\xi_{t,\gamma}) \sum_{i=1}^{p} b_i + E|\xi'_{t,\gamma}| \right\}.
\]

Let \( A_\ell = \{|\xi_{t,\gamma} - \xi'_{t,\gamma}| > 1/\ell\} \). It holds that

\[
P(A_\ell) \leq \ell E|\xi_{t,\gamma} - \xi'_{t,\gamma}| \leq \ell \left( \sum_{j=1}^{q} a_j \right) \left\{ \frac{c}{1 - \sum_{j=1}^{q} a_j} + 2^\gamma E(\xi_{t,\gamma}) \sum_{i=1}^{p} b_i + E|\xi'_{t,\gamma}| \right\}.
\]

Thus \( \sum_{\ell \geq 1} P(A_\ell) < \infty \). It follows from the Borel-Cantelli lemma (see, for example, Theorem 3.2.1 in Chow and Teicher 1997) that \( P(A_\ell, i.o.) = 0 \). Since \( A_\ell \subset A_{\ell+1} \), it holds that \( P(A_\ell) = 0 \) for any \( \ell \geq 1 \). Hence \( \xi_{t,\gamma} = \xi'_{t,\gamma} \) a.s.. This completes the proof.

B Stationarity of APGARCH\((p, q)\) processes

Ding et al (1993) introduce an asymmetric power GARCH\((p, q)\) model

\[
X_t = \sigma_t \varepsilon_t, \quad \sigma_t^\gamma = c + \sum_{i=1}^{p} b_i |X_{t-i}|^\gamma \{1 - d_i \text{sgn}(\varepsilon_{t-i})\}^\gamma + \sum_{j=1}^{q} a_j \sigma_{t-j}^\gamma, \quad (B.1)
\]

where \( \{\varepsilon_t\} \) is a sequence of independent and identically distributed random variables with mean 0 and \( 0 < E|\varepsilon_t|^\gamma < \infty, \gamma \in (0, 2] \), \( c > 0 \), \( b_i, a_j \geq 0 \) and \( d_i \in (-1, 1) \) are
parameters. The stationarity condition for APGARCH\((p,q)\) models are stated in proposition 2 below. It is implied by proposition 3 which deals with a more general form of volatility models. Proposition 2 resembles the stationarity condition for the standard GARCH models in Chen and An (1998). Note that we require the strictly stationary solution of the finite moment \(E|X_t|^\gamma\), which simplifies the condition for the existence of such a solution substantially. Proposition 2 was established by Ling and McAleer (2002) for the special case \(d_1 = \cdots = d_p\).

**Proposition 2.** The necessary and sufficient condition for (B.1) defining a unique strictly stationary process \(\{X_t, t = 0, \pm 1, \pm 2, \cdots\}\) with \(E|X_t|^\gamma < \infty\) is

\[
\sum_{i=1}^{p} b_i E\{(|\varepsilon_t| - d_i \varepsilon_t)^\gamma\} + \sum_{j=1}^{q} a_j < 1. \quad (B.2)
\]

We consider now a general form of volatility model

\[
Y_t = \rho_t \psi(\varepsilon_t), \quad \rho_t = \varphi_0 + \sum_{i=1}^{\infty} \varphi_i(\varepsilon_{t-i}) \rho_{t-i}, \quad (B.3)
\]

where \(\{\varepsilon_t\}\) is a sequence of independent and identically distributed random variables, \(\varphi_0 > 0\) is a constant, \(\psi(\cdot)\) and \(\varphi_i(\cdot)\) are non-negative, and \(E\{\psi(\varepsilon_t)\} < \infty\). The form of model (B.3) is general. It contains, for example, (B.1) as a special case with \(Y_t = |X_t|^\gamma\), \(\rho_t = \sigma_t^\gamma\), \(\psi(x) = |x|^\gamma\), \(\varphi_i(x) = b_i(|x| - d_i x)^\gamma + a_i\). (We assume that \(b_{p+j} = a_{q+j} = 0\) for any \(j \geq 1\).) Although the form (B.3) is different from ARCH(\(\infty\)) model introduced by Robinson (1991), its stationarity may be established in the similar manner. In fact the proof of proposition 3 below adopted the idea of Giraitis, Kokoszka, and Leipus (2000); see also section 2.7.1 of Fan and Yao (2003).

**Proposition 3.** Equation (B.3) admits a unique strictly stationary solution

\[
Y_t \equiv \varphi_0 \psi(\varepsilon_t) \left\{ 1 + \sum_{\ell=1}^{\infty} \sum_{1 \leq i_1, \cdots, i_\ell < \infty} \varphi_{i_1}(\varepsilon_{t-i_1}) \cdots \varphi_{i_\ell}(\varepsilon_{t-i_1-\cdots-i_\ell}) \right\}, \quad t = 0, \pm 1, \pm 2, \cdots \quad (B.4)
\]

with \(|EY_t| < \infty\) if and only if

\[
\sum_{i=1}^{\infty} E\{\varphi_i(\varepsilon_t)\} < 1.
\]
In fact, $EY_t = \varphi_0 E\{\psi(\varepsilon_t)\}/[1-\sum_{i \geq 1} E\{\varphi_i(\varepsilon_t)\}]$, and $\rho_t$ is a function of $\{\varepsilon_{t-1}, \varepsilon_{t-2}, \cdots\}$ only.

**Proof.** The necessity follows directly from taking expectation at the both sides of (B.4), and the fact $|EY_t| < \infty$. We show the sufficiency below.

It follows from (B.3) that, for any integer $k \geq 1$,

$$Y_t = \varphi_0 \psi(\varepsilon_t) + \psi(\varepsilon_t) \sum_{i=1}^{\infty} \varphi_i(\varepsilon_{t-i}) \rho_{t-i}$$

$$= \varphi_0 \psi(\varepsilon_t) \left\{1 + \sum_{i=1}^{\infty} \varphi_i(\varepsilon_{t-i}) \right\} + \psi(\varepsilon_t) \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \varphi_i(\varepsilon_{t-i}) \varphi_j(\varepsilon_{t-i-j}) \rho_{t-i-j}$$

$$= \varphi_0 \psi(\varepsilon_t) \left\{1 + \sum_{\ell=1}^{k} \sum_{1 \leq i_1, \cdots, i_{\ell} < \infty} \varphi_{i_1}(\varepsilon_{t-i_1}) \cdots \varphi_{i_{\ell}}(\varepsilon_{t-i_1-\cdots-i_{\ell}}) \right\}$$

$$+ \psi(\varepsilon_t) \sum_{1 \leq i_1, \cdots, i_{k+1} < \infty} \varphi_{i_1}(\varepsilon_{t-i_1}) \cdots \varphi_{i_{k+1}}(\varepsilon_{t-i_1-\cdots-i_{k+1}}) \rho_{t-i_1-\cdots-i_{k+1}}. \quad \text{(B.5)}$$

Let $Y'_t$ be the random variable defined on the right-hand-side of (B.4). Then $Y'_t \geq 0$ a.s. Note that for any $\ell \geq 1$,

$$E\left\{ \sum_{1 \leq i_1, \cdots, i_{\ell} < \infty} \varphi_{i_1}(\varepsilon_{t-i_1}) \cdots \varphi_{i_\ell}(\varepsilon_{t-i_1-\cdots-i_{\ell}}) \right\} = \sum_{1 \leq i_1, \cdots, i_{\ell} < \infty} \prod_{j=1}^{\ell} E\{\varphi_{i_j}(\varepsilon_1)\} = \sum_{i=1}^{\infty} E\varphi_i(\varepsilon_1)^{\ell}.$$ 

Thus $0 \leq Y'_t < \infty$ a.s., $E(Y'_t) = \varphi_0 E\{\psi_1\}/[1-\sum_{i \geq 1} E\varphi_i(\varepsilon_1)]$, and $\{Y'_t\}$ is strictly stationary. It is easy to verify that $Y'_t$ fulfils (B.3).

To show the uniqueness, let $\{Y_t\}$ be a strictly stationary solution of (B.3) with $|EY_t| < \infty$. We will show now that $Y_t = Y'_t$ a.s. for any fixed $t$. By (B.5) it holds for any $k \geq 1$,

$$|Y_t - Y'_t| \leq \psi(\varepsilon_t) \sum_{1 \leq i_1, \cdots, i_{k+1} < \infty} \varphi_{i_1}(\varepsilon_{t-i_1}) \cdots \varphi_{i_{k+1}}(\varepsilon_{t-i_1-\cdots-i_{k+1}}) |\rho_{t-i_1-\cdots-i_{k+1}}|$$

$$+ \varphi_0 \psi(\varepsilon_t) \sum_{\ell=k+1}^{\infty} \sum_{1 \leq i_1, \cdots, i_{\ell} < \infty} \varphi_{i_1}(\varepsilon_{t-i_1}) \cdots \varphi_{i_{\ell}}(\varepsilon_{t-i_1-\cdots-i_{\ell}}).$$

Hence,

$$E|Y_t - Y'_t| \leq E|Y_t| + \frac{\varphi_0 E\psi(\varepsilon_1)}{1-\sum_{i=1}^{\infty} E\varphi_i(\varepsilon_1)} \left\{ \sum_{i=1}^{\infty} E\varphi_i(\varepsilon_1) \right\}^{k+1}.$$ 

Now using the same argument as showing $\xi_t, \gamma = \xi'_t, \gamma$ a.s. in the proof of proposition 1 above, we may show that $Y_t = Y'_t$ a.s. This completes the proof.
C Proof of Theorem 1

The basic idea of the proof is similar to Davis and Dunsmuir (1997), although technically it is more involved under current context; see also Pan, Wang, and Yao (2005). We use the same notation as in section 3.1. Furthermore for \( \boldsymbol{u} \in \mathbb{R}^{2p+q+1} \), put

\[
S_n(\boldsymbol{u}) = \sum_{t=1}^{n} \{|Z_t^0 + n^{-1/2} \boldsymbol{u}| - |Z_t^0|\}, \quad S_n^*(\boldsymbol{u}) = \sum_{t=1}^{n} \{|Z_t^0 + n^{-1/2} \boldsymbol{u} \hat{Z}_t(\theta_0)| - |Z_t(\theta_0)|\},
\]

and

\[
S(\boldsymbol{u}) = \gamma f(0) \mathbf{u}^\tau \Sigma_0 \mathbf{u} + \mathbf{u}^\tau \mathcal{N}, \quad (C.1)
\]

where \( \mathcal{N} \sim N(0, \Sigma) \). We also write \( Y_{t,i} = |X_t|^{\gamma} \{1 - d_i \text{sgn}(\varepsilon_t)\}^{\gamma} \). Recall \( U_t(\theta) \) is the derivative of \( \xi_t,\gamma(\theta) \) with respect to \( \theta \). Then the \( 2p + q + 1 \) components of \( U_t(\theta) \) can be expressed as follows.

\[
U_{t,1}(\theta) = \left\{1 - \sum_{\ell=1}^{q} a_{\ell}\right\}^{-1},
\]

\[
U_{t,1+i}(\theta) = Y_{t-i,i} + \sum_{k=1}^{p} \sum_{j_1=1}^{q} \cdots \sum_{j_k=1}^{q} a_{j_1} \cdots a_{j_k} Y_{t-i-j_1-\cdots-j_k,i} \quad (C.2)
\]

\[
U_{t,1+p+j}(\theta) = \frac{c}{(1 - \sum_{\ell=1}^{p} a_{\ell})^2} + \sum_{\ell=1}^{p} b_{\ell} Y_{t-\ell-j,\ell} + \sum_{\ell=1}^{p} b_{\ell} \sum_{k=1}^{\infty} (k + 1) \sum_{j_1=1}^{q} \cdots \sum_{j_k=1}^{q} a_{j_1} \cdots a_{j_k} Y_{t-\ell-j-j_1-\cdots-j_k,\ell} \quad (C.3)
\]

\[
U_{t,1+p+q+i}(\theta) = -\gamma b_{i} Y_{t-i,i} \frac{\text{sgn}(\varepsilon_{t-i})}{1 - d_i \text{sgn}(\varepsilon_{t-i})} \quad (C.4)
\]

where \( i = 1, \cdots, p, \ j = 1, \cdots, q \). Note that all \( c, b_i, a_j \) are positive and \( d_i \in [-1 + \delta_0, 1 - \delta_0] \), and all the terms occurred on the RHS of (C.2) – (C.5) are contained (with a different but positive coefficients) on the RHS of (1.5). Using the same argument as in section 2.5 of Hall and Yao (2003), we may show that (3.1) holds.
For integer $s \geq 1$, let $C(\mathbb{R}^s)$ be the space of the real-valued continuous functions on $\mathbb{R}^s$, topologized by the separating family of seminorms

$$p_m(f) = \sup\{|f(x)| : x \in K_m\}$$

where $\{K_m \neq \emptyset, m \geq 1\}$ is an increasing sequence of compact sets such that $K_m$ lies in the interior of $K_{m+1}$ and $R^s = \bigcup_{m=1}^{\infty} K_m$. Define a metric on $C(\mathbb{R}^s)$ as follows

$$d(f, g) = \max_{1 \leq m < \infty} \frac{2^{-m} p_m(f - g)}{1 + p_m(f - g)}.$$ 

Then $\{C(\mathbb{R}^s), d\}$ is a complete and separable metric space Rudin (1991, p. 33). For probability measures $P_n, P$ on $C(\mathbb{R}^s)$, we say that $P_n$ converges weakly to $P$ in $C(\mathbb{R}^s)$ if $\int f dP_n \to \int f dP$ for any bounded and continuous function $f$ defined on $C(\mathbb{R}^s)$. For random functions $S_n, S$ defined on $C(\mathbb{R}^s)$, we say that $S_n$ converges in distribution to $S$ if the distribution of $S_n$ converges weakly to that of $S$ in $C(\mathbb{R}^s)$ (Billingsley 1999). We denote by $||v||$ the Euclidean norm for a vector $v$.

We always assume that conditions (A1) – (A5) hold and $\delta_0 \in (0, 1)$ in (1.6). We first prove Theorem 1 under the assumption that we also observed $X_k$ for all $k \leq 0$, which splits into three lemmas below. Finally we show that the same asymptotic result holds with the truncation $X_k \equiv 0$ for all $k \leq 0$.

**Lemma 1.** Let $\hat{u}^*$ be the minimizer of $S_n^*(u)$. Then $\hat{u}^* \to N(0, \Sigma_0^{-1}\Sigma_0^{-1}/\{2\gamma f(0)\}^2)$ in distribution. In fact $S_n^*(u)$ converges in distribution to $S(u)$ in $C(\mathbb{R}^{2p+q+1})$.

**Proof.** We will show that for any $u \in \mathbb{R}^{2p+q+1}$,

$$S_n^*(u) = u^T N_n + \gamma f(0) u^T \Sigma_0 u + o_p(1), \quad (C.6)$$

where $N_n \to N$ in distribution, where $N$ is defined as in (C.1). Note that the quadratic function $S(u)$ has the minimizer $-\{\gamma f(0)\}^{-1}\Sigma_0^{-1} N$, and $S_n^*(u)$ is a convex function. Now the asymptotic normality of $\hat{u}^*$ follows from the Basic Corollary of Hjort and Pollard (1993). By the convexity lemma (see, for example, Lemma 1 of Hjort and Pollard 1993), the term $o_p(1)$ in (C.6) is uniform in $u$ over any compact sets in $\mathbb{R}^{2p+q+1}$. This implies that the probability measures of $S_n^*(u)$, for $u \in \mathbb{R}^{2p+q+1}$, are tight. By Theorem 7.1 of Billingsley (1999) that $S_n^*(u)$ converges in distribution to $S(u)$ in $C(\mathbb{R}^{2p+q+1})$. 

17
Now we prove (C.6). By the identity
\[ |z - y| - |z| = -y \text{sgn}(z) + 2(y - z)\{I(0 < z < y) - I(y < z < 0)\}, \quad z \neq 0 \]
(see Davis and Dunsmuir 1997), we have
\[
S_n^*(u) = n^{-1/2} \sum_{t=1}^{n} u^\tau \hat{Z}_t(\theta^0) \text{sgn}\{Z_t(\theta^0)\} \\
+ 2 \sum_{t=1}^{n} \{-n^{-1/2}u^\tau \hat{Z}_t(\theta^0) - Z_t(\theta^0)\} I\{0 < Z_t(\theta^0) < -n^{-1/2}u^\tau \hat{Z}_t(\theta^0)\} \\
+ 2 \sum_{t=1}^{n} \{n^{-1/2}u^\tau \hat{Z}_t(\theta^0) + Z_t(\theta^0)\} I\{0 > Z_t(\theta^0) > -n^{-1/2}u^\tau \hat{Z}_t(\theta^0)\}.
\]

Write the three terms on the right-hand side of the above expression as, respectively, \(I_1, I_2\) and \(I_3\). Then (C.6) follows immediately from the following three assertions:

(i) \(I_2 \to \gamma f(0)u^\tau E[\hat{Z}_t(\theta^0)\hat{Z}_t(\theta^0)^\tau I\{u^\tau \hat{Z}_t(\theta^0) < 0\}\{Z_t(\theta^0) = 0\}]u\) in probability,

(ii) \(I_3 \to \gamma f(0)u^\tau E[\hat{Z}_t(\theta^0)\hat{Z}_t(\theta^0)^\tau I\{u^\tau \hat{Z}_t(\theta^0) > 0\}\{Z_t(\theta^0) = 0\}]u\) in probability, and

(iii) \(I_1 \equiv u^\tau N_n \to N(0, u^\tau \Sigma u)\) in distribution.

To simplify notion, we write \(Z_t = Z_t(\theta^0)\) and \(\hat{Z}_t = \hat{Z}_t(\theta^0)\). The proofs for (i) and (ii) are similar. We only show (i). To this end, let \(\psi(w, z)\) be the joint density function of \((u^\tau \hat{Z}_t, Z_t)\), and \(\psi(z|w)\) and \(\psi(w)\) be the corresponding conditional and marginal densities. A simple Taylor expansion of \(\psi(z|w)\) around \(z = 0\) leads to
\[
EI_2 = 2(n - \nu + 1) \int_{0 < z < -w/\sqrt{n}} (-w/\sqrt{n} - z) \psi(w, z) dwdz \\
= 2(n - \nu + 1) \int_{-\infty}^{0} \psi(w) dw \int_{0}^{-w/\sqrt{n}} (-w/\sqrt{n} - z) \psi(0|w) dz + R_n \\
= \int_{-\infty}^{0} w^2 \psi(0, w) dw + o(1) + R_n \\
= \gamma f(0)E\{(u^\tau \hat{Z}_t)^2 I(u^\tau \hat{Z}_t < 0)|Z_t = 0\} + o(1) + R_n, \tag{C.7}
\]
where \(R_n\), due to condition (A4), may be bounded as follows:
\[
|R_n| \leq Cn \int_{-\infty}^{0} \varphi(w) dw \int_{0}^{-w/\sqrt{n}} (w/\sqrt{n} + z) dz = C_1 E\{|u^\tau \hat{Z}_t|^3|\sqrt{n} = O(1/\sqrt{n}),
\]

18
see (3.1). In the above expression, $C$ and $C_1$ are some positive constants. This, together with (C.7), implies

$$EI_2 \to \gamma f(0)E\{(u^\top \dot{Z}_t)^2 I(u^\top \dot{Z}_t < 0)|Z_t = 0\}. \tag{C.8}$$

Similarly to (C.8), we may show that for any $k \geq 2$,

$$E\{|(n^{-1/2}u^\top \dot{Z}_t + Z_t)(0 < Z_t < -n^{-1/2}u^\top \dot{Z}_t)|^k\} = O(n^{-(k+1)/2}). \tag{C.9}$$

To show $\text{Var}(I_2) \to 0$, we employ the small-block and large-block arguments as follows. We partition \{\nu, \nu+1, \ldots, n\} into $2k_n + 1$ subsets with large blocks of size $l_n$, small blocks of size $s_n$, and the last remaining set of size $n - \nu + 1 - k_n(l_n + s_n)$, where $k_n = [(n - \nu + 1)/(l_n + s_n)]$. We write accordingly

$$I_2 = \sum_{j=1}^{k_n} A_j + \sum_{j=1}^{k_n} B_j + R, \tag{C.10}$$

where

$$A_j = \sum_{t=(j-1)(l_n + s_n) + \nu}^{j(l_n + s_n) + \nu} (n^{-1/2}u^\top \dot{Z}_t + Z_t) I(0 < Z_t < -n^{-1/2}u^\top \dot{Z}_t),$$

$$B_j = \sum_{t=j(l_n + s_n) + \nu}^{j(l_n + s_n) + \nu} (n^{-1/2}u^\top \dot{Z}_t + Z_t) I(0 < Z_t < -n^{-1/2}u^\top \dot{Z}_t).$$

Put

$$l_n = \lfloor \sqrt{n/\log n} \rfloor, \quad s_n = \lfloor n^{1/4}/\log n \rfloor. \tag{C.11}$$

Then $k_n = O(\sqrt{n \log n})$. Now it follows from (C.9) that

$$E\left(\sum_{j=1}^{k_n} B_j\right)^2 \leq C\frac{k_n^2 s_n^2}{n^{3/2}} \to 0,$$

and $E(R^2) \leq C\ell_n^2/n^{3/2} \to 0$. On the other hand, it follows from proposition 2.5(i) of Fan and Yao (2003) that

$$\text{Var}\left(\sum_{j=1}^{k_n} A_j\right)^2 \leq k_n E(A_1^2) + 2 \sum_{i=1}^{k_n-1} (k_n - i) |\text{Cov}(A_1, A_{i+1})| \tag{C.12}$$

$$\leq C\frac{k_n l_n^2}{n^{3/2}} + 16 k_n \sum_{i=1}^{k_n-1} \alpha(i s_n)^{1/2} (EA_1)^{1/2} \leq C\frac{k_n l_n^2}{n^{3/2}} + C\frac{k_n l_n^2}{n^{5/4}} \sum_{i=1}^{k_n-1} \alpha(i s_n)^{1/2}$$

$$\leq C\frac{k_n l_n^2}{n^{3/2}} + C\frac{k_n l_n^2}{n^{5/4}} \alpha(s_n)^{1/2} \to 0. \tag{C.13}$$

19
The limit on the right hand side of the above expression is ensured by condition (A1). Therefore we conclude that \( \text{Var}(I_2) \to 0 \), which, together with (C.8), imply (i).

To show (iii), we note the fact that for any given \( u \in \mathbb{R}^{2p+q+1} \), the inequality

\[
E(I_1) + E(I_2 + I_3) \geq 0
\]

holds for all large values of \( n \); see the definition of \( S^*(u) \) and condition (A2). Note that \( E(I_2 + I_3) \to \gamma f(0)u^\top\Sigma_0 u \geq 0 \) (see (C.8)), and \( E(I_1) = u^\top [n^{1/2} E(\tilde{Z}_t \text{sgn}(Z_t))] \{1 + o(1)\} \). Hence \( n^{1/2} E(\tilde{Z}_t \text{sgn}(Z_t)) \to 0 \), in order that (C.14) holds for all large values of \( n \) with any given \( u \). Now we have proved that \( E(I_1) \to 0 \).

We apply the decomposition (C.10) for \( I_1 \), that is,

\[
I_1 = \sum_{j=1}^{k_n} (A'_j + B'_j) + R',
\]

with

\[
A'_j = \frac{u^\top}{n^{1/2}} \sum_{t=(j-1)l_n+s_n+\nu}^{j l_n+(j-1)s_n+\nu} \tilde{Z}_t \text{sgn}(Z_t), \quad B'_j = \frac{u^\top}{n^{1/2}} \sum_{t=(j-1)l_n+s_n+\nu}^{j l_n+(s_n)+\nu} \tilde{Z}_t \text{sgn}(Z_t),
\]

and where \( l_n \) and \( s_n \) are specified in (C.11). Recall \( \tilde{Z}_t = -U_t(\theta^0)/\{\gamma_l(\theta^0)\} \).

Based on (3.1), we may show in the same manner as for (C.12) that

\[
\text{Var} \left( \sum_{j=1}^{k_n} B'_j \right) = O \left\{ \frac{k_n s_n^2}{n} + \frac{k_n s_n^2}{n} \sum_{j=1}^{k_n-1} \alpha(j l_n)^{1/2} \right\} = O \left\{ \frac{k_n s_n^2}{n} + \frac{k_n^2 s_n^2}{n} \alpha(l_n)^{1/2} \right\} \to 0.
\]

It is easy to see that \( \text{Var}(R') = O(l_n^2/n) \to 0 \). Hence

\[
I_1 = \sum_{j=1}^{k_n} A'_j + o_p(1) \equiv Q_n + o_p(1). \tag{C.15}
\]

Now

\[
\text{Var}(Q_n) = k_n \text{Var}(A'_1) + 2 \sum_{j=1}^{k_n-1} (k_n - j) \text{Cov}(A'_1, A'_{1+j}).
\]

Note that

\[
k_n \text{Var}(A'_1) = \frac{k_n l_n}{n} u^\top (\tilde{Z}_1 \tilde{Z}_1^\top) u + \frac{2k_n l_n}{n} u^\top \sum_{j=1}^{l_n-1} (1 - j/l_n) E(\tilde{Z}_1 \tilde{Z}_{1+j}^\top \text{sgn}(Z_1 Z_{1+j})) u
\]

\[
\to u^\top (\tilde{Z}_1 \tilde{Z}_1^\top) u + 2u^\top \sum_{j=1}^{\infty} E(\tilde{Z}_1 \tilde{Z}_{1+j} \text{sgn}(Z_1 Z_{1+j})) u = u^\top \Sigma u.
\]
See, for example, Theorem 2.20(i) of Fan and Yao (2003). On the other hand, it follows from proposition 2.5(i) of Fan and Yao (2003) and condition (A1) that
\[ \sum_{j=1}^{k_n-1} (k_n - j)|\text{Cov}(A'_1, A'_{1+j})| \leq C \frac{k_n^2 l_n^2}{n} \alpha(s_n)^{1/2} \rightarrow 0. \]
Hence we have proved that
\[ \text{Var}(Q_n) \rightarrow u^\top \Sigma u. \] (C.16)

Now we employ a truncation argument to establish the asymptotic normality for \( Q_n \). Write
\[ \dot{Z}_t^L = \dot{Z}_t I(||\dot{Z}_t|| \leq L), \quad \dot{Z}_t^R = \dot{Z}_t I(||\dot{Z}_t|| > L). \]

Let \( Q_n^L \) and \( Q_n^R \) be defined in the same manner as \( Q_n \) with \( \dot{Z}_t \) replaced by, respectively, \( \dot{Z}_t^L \) and \( \dot{Z}_t^R \). Similar to the arguments leading to (C.16), we may show that
\[ \text{Var}(Q_n^L) \rightarrow u^\top \Sigma^L u, \quad \text{Var}(Q_n^R) \rightarrow u^\top \Sigma^R u, \]
where \( \Sigma^L \) and \( \Sigma^R \) are defined in the same manner as \( \Sigma \) with \( \dot{Z}_t \) replaced by, respectively, \( \dot{Z}_t^L \) and \( \dot{Z}_t^R \). It is easy to see that as \( L \rightarrow \infty, \Sigma^L \rightarrow \Sigma \), and therefore \( \Sigma^R \rightarrow 0 \). Put
\[ M_n = |E \exp(itQ_n) - \exp(-t^2 u^\top \Sigma u/2)|, \]
where \( i = \sqrt{-1} \). It is easy to see that
\begin{align*}
M_n &\leq E|\exp(itQ_n^L)\{\exp(itQ_n^R) - 1\}| + E|\exp(itQ_n^L) - \prod_{j=1}^{k_n} E\exp(itA_j^L)| \\
&\quad + |\prod_{j=1}^{k_n} E\exp(itA_j^L) - \exp(-t^2 u^\top \Sigma^L u/2)| \\
&\quad + |\exp(-t^2 u^\top \Sigma^L u/2) - \exp(-t^2 u^\top \Sigma^L u/2)|, \quad \text{(C.17)}
\end{align*}
where \( A_j^L \) is defined in the same manner as \( A'_j \) with \( Z_t \) replaced by \( Z_t^L \). For any given \( \epsilon > 0 \), the first term on the right-hand side of (C.17) is bounded by
\[ E|\exp(itQ_n^R) - 1| = O\{\text{Var}(Q_n^R)\} \quad \text{(as } n \rightarrow \infty\text{),} \]
which may be smaller than \( \epsilon/2 \) for all sufficiently large \( n \) as long as we choose \( L = L(\epsilon) \) large enough; see, for example, section 2.7.7 of Fan and Yao (2003), and
Masry and Fan (1997). The last term is also smaller than $\epsilon/2$ by choosing $L$ large. By proposition 2.6 of Fan and Yao (2003), the second term on the right hand side of (C.17) is bounded by $16k_n\alpha(s_n-\nu)$, which converges to 0 due to condition (A.1). To prove that the third term on the right hand side of (C.17) converges to 0, we may prove an equivalent limit:

$$\sum_{j=1}^{k_n} A_{ij} \rightarrow N(0, u^\tau \Sigma^L u/2)$$

in distribution while treating $\{A_{ij}\}$ as a sequence of independent random variables. The latter is implied by the Lindeberg condition

$$\sum_{j=1}^{k_n} E\{(A_{ij})^2 I(|A_{ij}| > \omega u^\tau \Sigma^L u)\} \rightarrow 0,$$

for any $\omega > 0$; see, for example, Chow and Teicher (1997, p.315). Note $|A_{ij}| \leq (l_n/n^{1/2})(||u||^2 + L^2) \leq 2(||u||^2 + L^2)/\log n \rightarrow 0$. Hence $\{A_{ij} > \omega u^\tau \Sigma^L u\}$ is an empty set for all large $n$. Therefore the limit above holds. We have shown that $Q_n \rightarrow N(0, u^\tau \Sigma u)$. Now assertion (iii) follows from (C.15). This completes the proof of Lemma 1.

**Lemma 2.** For any compact set $K \subset \mathbb{R}^{2p+q+1}$, $\sup_{u \in K} |S_n(u) - S_n^*(u)| \rightarrow 0$ in probability.

**Proof.** Let $S_n^*(u) = \sum_{\nu \leq t \leq n} \{ |Z_t(\theta^0) + n^{-1/2} u^\tau \hat{Z}_t(\theta^0) + \frac{1}{2n} u^\tau \hat{\bar{Z}}_t(\theta^0) u| - |Z_t(\theta^0)| \}$, where the Hessian matrix

$$\hat{Z}_t(\theta) = \frac{1}{\gamma} \left\{ \frac{U_t(\theta)U_t(\theta)^\tau}{\xi_t,\gamma(\theta)^2} - \frac{\hat{U}_t(\theta)}{\xi_t,\gamma(\theta)} \right\},$$

and $\hat{U}_t(\theta) = \frac{\partial U_t(\theta)}{\partial \theta}$. It follows from (3.1) that for $\theta \in \Theta$ all the elements of $U_t(\theta)U_t(\theta)^\tau/\xi_t,\gamma(\theta)^2$ have finite moments. In the same vein, we may show that all the elements of $\hat{U}_t(\theta)/\xi_t,\gamma(\theta)$ also have finite moments. Note that

$$|S_n(u) - S_n^*(u)| = \sum_{t=\nu}^{n} \left| |Z_t(\theta^0) + n^{-1/2} u^\tau \hat{Z}_t(\theta^0) + \frac{1}{2n} u^\tau \hat{\bar{Z}}_t(\theta^0) u| - |Z_t(\theta^0) + n^{-1/2} u| \right|$$

$$\leq \left| u^\tau \left\{ \frac{1}{2n} \sum_{t=\nu}^{n} \{ \hat{Z}_t(\theta^0) - \hat{\bar{Z}}_t(\theta^0) \} \right\} u \right|,$$
where $\mathbf{\theta}^*$ is between $\mathbf{\theta}^0$ and $\mathbf{\theta}^0 + n^{-1/2}\mathbf{u}$. Hence $S_n(\mathbf{u}) - S_n^*(\mathbf{u}) \to 0$ in probability uniformly on compact sets. Similar to Lemma 1, we may show that $S_n^*(\mathbf{u}) - S_n^*(\mathbf{u}) \to 0$ in probability uniformly on compact sets. Hence Lemma 2 holds.

**Lemma 3.** $S_n(\mathbf{u}) \to S(\mathbf{u})$ in distribution in $C(\mathbb{R}^{2p+q+1})$.

**Proof.** For any small $\epsilon > 0$, let $m_0 = -\log(\epsilon/2)$. Then $2^{-m} < \epsilon/2$ for any $m \geq m_0$. Lemma 2 implies that for any $\epsilon_0 > 0$, it holds $P\{p_{m_0}(S_n - S_n^*) \geq \epsilon/2\} < \epsilon_0$ for all sufficiently large values of $n$. Note that

$$
d(S_n, S_n^*) \leq \max_{1 \leq m \leq m_0} \frac{2^{-m}p_m(S_n - S_n^*)}{1 + p_m(S_n - S_n^*)} + \max_{m > m_0} \frac{2^{-m}p_m(S_n - S_n^*)}{1 + p_m(S_n - S_n^*)}
$$

$$
\leq \max_{1 \leq m \leq m_0} p_m(S_n - S_n^*) + \frac{\epsilon}{2} \leq p_{m_0}(S_n - S_n^*) + \frac{\epsilon}{2}.
$$

Hence it holds that for all sufficiently large $n$,

$$
P\{d(S_n, S_n^*) > \epsilon\} \leq P\{p_{m_0}(S_n - S_n^*) > \epsilon/2\} < \epsilon_0.
$$

Therefore $d(S_n, S_n^*) \to 0$ in probability. This together with Lemma 1 imply that $S_n(\mathbf{u}) \to S(\mathbf{u})$ in distribution in $C(\mathbb{R}^{2p+q+1})$.

**Proof of Theorem 1.** It follows from Lemma 3 and Skorokhod’s representation theorem (Pollard 1984, p.71-73) that there exist random functions $T_n$ and $T$ in $C(\mathbb{R}^{2p+q+1})$ for which $d(T_n, T) \to 0$ almost surely, while $T_n$ has the same distribution of $S_n$, and $T$ has the same distribution of $S$. Hence there exists a set $\Omega$ with $P(\Omega) = 1$, and for any $\omega \in \Omega$,

$$
\sup_{\mathbf{u} \in K} |T_n(\mathbf{u}, \omega) - T(\mathbf{u}, \omega)| \to 0 \quad \text{(C.18)}
$$

for any compact set $K$. Note $S(\mathbf{u}, \omega)$ is convex in $\mathbf{u}$ and it has unique minimizer $\eta = -\{\gamma f(0)^{-1}\Sigma_0^{-1}N$. Denote by $\eta^*$ the minimizer of $T(\mathbf{u})$. Then $\eta_n^*$ and $\eta$ have the same distribution. For any given positive random variable $M$, let

$$
\eta_n^* = \arg \min_{\|\mathbf{u} - \eta\| \leq M} T_n^*(\mathbf{u}).
$$

We now show that $\eta_n^*(\omega) \to \eta^*(\omega)$ for any $\omega \in \Omega$. Suppose it does not hold. Then there exists a subsequence $\{n'\}$ such that $\eta_{n'}^*(\omega) \to \eta'(\omega) \neq \eta^*(\omega)$. Note that

$$
0 \leq T_{n'}\{\eta^*(\omega), \omega\} - T_{n'}\{\eta_{n'}^*(\omega), \omega\} = T_{n'}\{\eta^*(\omega), \omega\} - T\{\eta^*(\omega), \omega\} + T\{\eta_{n'}^*(\omega), \omega\} - T\{\eta_{n'}^*(\omega), \omega\} - T\{\eta_{n'}^*(\omega), \omega\} - T\{\eta_{n'}^*(\omega), \omega\} - T\{\eta^*(\omega), \omega\} - T\{\eta'(\omega), \omega\} < 0.
$$
This contradiction shows that $\eta_n^*(\omega) \to \eta^*(\omega)$ for any $\omega \in \Omega$. Note that the two inequalities in the above expression follow from the definitions of $\eta_n^*$ and $\eta^*$, the limits are guaranteed by (C.18). Define

$$\eta_n = \arg \min_{\|u - \eta\| \leq M} S_n(u).$$

Then $\eta_n \to \eta$ in distribution. Therefore the required CLT holds.

Note in all the proofs so far, we assume that we observe $X_t$ for all $t \leq 0$. Below we show that the same conclusion holds even with the truncation $X_t \equiv 0$ for all $t \leq 0$. To this end, it suffices to show that

$$\sup_{\theta \in \Theta_0} \sum_{t=\nu}^{n} \sum_{i=1}^{p} b_i \sum_{k=1}^{\infty} \sum_{1 \leq j_1, \ldots, j_k \leq q \atop j_1 + \cdots + j_k \geq t-i} a_{j_1} \cdots a_{j_k} |X_{t-i-j_1-\cdots-j_k}|^{\gamma} \{1 - d_i \text{sgn}(\varepsilon_{t-i-j_1-\cdots-j_k})\}^{\gamma} = o_p(1),$$

where $\Theta_0 \subset \Theta$ is a ball with a small but fixed radius and centred at $\theta^0$, and $\xi_{t,\gamma}(\theta)$ is defined as the same as $\xi_{t,\gamma}(\theta)$ but with $X_t$ replaced by 0 for all $t \leq 0$. Hence we only need to show that

$$\sup_{\theta \in \Theta_0} \sum_{t=\nu}^{n} \sum_{i=1}^{p} b_i \sum_{k=1}^{\infty} \sum_{1 \leq j_1, \ldots, j_k \leq q \atop j_1 + \cdots + j_k \geq t-i} a_{j_1} \cdots a_{j_k} |X_{t-i-j_1-\cdots-j_k}|^{\gamma} \{1 - d_i \text{sgn}(\varepsilon_{t-i-j_1-\cdots-j_k})\}^{\gamma} = o_p(1).$$

This is true because of $E|X_t|^{\gamma} < \infty$ and the fact that for any $\delta > 0$ and $1 \leq i \leq p$,

$$P\{ \sup_{\theta \in \Theta_0} \sum_{t=\nu}^{n} \sum_{k=1}^{\infty} \sum_{1 \leq j_1, \ldots, j_k \leq q \atop j_1 + \cdots + j_k \geq t-i} a_{j_1} \cdots a_{j_k} |X_{t-i-j_1-\cdots-j_k}|^{\gamma} \{1 - d_i \text{sgn}(\varepsilon_{t-i-j_1-\cdots-j_k})\}^{\gamma} > \delta \} \leq C n \sup_{\theta \in \Theta_0} \sum_{k=1}^{\infty} \sum_{j_1 + \cdots + j_k \geq t-i} a_{j_1} \cdots a_{j_k} \leq C n \sup_{\theta \in \Theta_0} \sum_{k \geq (\nu-p)/q} (\sum_{j=1}^{k} a_j)^k \to 0.$$

The limit above is guaranteed by (A5). This completes the proof for theorem 1.

**Acknowledgements**

This project was partially supported by an EPSRC research grant. Mingjin Wang is partly supported by the Natural Science Foundation of China (grant 70201007). The comments and suggestions from Professor Xiaohong Chen are gratefully acknowledged.
References


Ling, S. and M. McAleer (2002). Necessary and sufficient moment conditions for the GARCH($r,s$) and asymmetric power GARCH($r,s$) models. *Econometric Theory* 18, 722–729.


### Table 1: LAD Estimation Results of the Volatility Functions

<table>
<thead>
<tr>
<th></th>
<th>$\gamma$</th>
<th>$\hat{c} \times 10^4$</th>
<th>$\hat{a}_1$</th>
<th>$\hat{b}_1$</th>
<th>$\hat{d}_1$</th>
<th>$R(\gamma)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>S&amp;P</td>
<td>0.9</td>
<td>0.9146</td>
<td>0.9345</td>
<td>0.0464</td>
<td>0.4961</td>
<td>0.0138</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.2417)</td>
<td>(0.0086)</td>
<td>(0.0062)</td>
<td>(0.0967)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.0</td>
<td>0.0032</td>
<td>0.9104</td>
<td>0.0265</td>
<td>0.2442</td>
<td>0.0159</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0008)</td>
<td>(0.0113)</td>
<td>(0.0041)</td>
<td>(0.0687)</td>
<td></td>
</tr>
<tr>
<td>IBM</td>
<td>1.2</td>
<td>0.7239</td>
<td>0.9211</td>
<td>0.0398</td>
<td>0.2558</td>
<td>0.0077</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.2693)</td>
<td>(0.0174)</td>
<td>(0.0087)</td>
<td>(0.1179)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.0</td>
<td>0.0097</td>
<td>0.9376</td>
<td>0.0178</td>
<td>0.1811</td>
<td>0.0094</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0040)</td>
<td>(0.0133)</td>
<td>(0.0041)</td>
<td>(0.0936)</td>
<td></td>
</tr>
</tbody>
</table>

**Note:** Standard errors in parentheses were calculated as suggested in Remark 2 in Section 3.1. Newey-West’s (1987) Bartlett kernel method was used to estimate $\Sigma$ with bandwidth $L_T = T^{1/3}$. The matrix $\Sigma_0$ was estimated by the Nadaraya-Watson kernel regression with Gaussian kernel and bandwidth $h = 0.05 \times \text{Range}(Z_t(\hat{\theta}^0))$. The value $f(0)$ was estimated using kernel density with Gaussian kernel and the simple reference bandwidth (see, for example, (5.9) of Fan and Yao 2003).
Figure 1: Time series plots of (a) S&P500 and (b) IBM stock daily return. (c) and (d) are the auto-correlations of their squared returns, and (e) and (f) are auto-correlation of their absolute returns.

Figure 2: Plots of $R(\gamma)$ functions of (a) S&P500 data and (b) IBM data.
Figure 3: Estimated values of power parameter $\gamma$ using Gaussian qML and full-LAD for $t_3$, $t_4$ and normal errors when true value is 1