GLOBAL SELF-WEIGHTED AND LOCAL QUASI-MAXIMUM EXPONENTIAL LIKELIHOOD ESTIMATORS FOR ARMA–GARCH/IGARCH MODELS

BY KE ZHU AND SHIQING LING

Hong Kong University of Science and Technology

This paper investigates the asymptotic theory of the quasi-maximum exponential likelihood estimators (QMELE) for ARMA–GARCH models. Under only a fractional moment condition, the strong consistency and the asymptotic normality of the global self-weighted QMELE are obtained. Based on this self-weighted QMELE, the local QMELE is showed to be asymptotically normal for the ARMA model with GARCH (finite variance) and IGARCH errors. A formal comparison of two estimators is given for some cases. A simulation study is carried out to assess the performance of these estimators, and a real example on the world crude oil price is given.

1. Introduction. Assume that \( \{ y_t : t = 0, \pm 1, \pm 2, \ldots \} \) is generated by the ARMA–GARCH model

\[
\begin{align*}
y_t &= \mu + \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{i=1}^{q} \psi_i \epsilon_{t-i} + \epsilon_t, \\
\epsilon_t &= \eta_t \sqrt{h_t} \quad \text{and} \quad h_t = \alpha_0 + \sum_{i=1}^{r} \alpha_i \epsilon_{t-i}^2 + \sum_{i=1}^{s} \beta_i h_{t-i},
\end{align*}
\]

where \( \alpha_0 > 0, \alpha_i \geq 0 \ (i = 1, \ldots, r), \beta_j \geq 0 \ (j = 1, \ldots, s) \), and \( \eta_t \) is a sequence of i.i.d. random variables with \( E\eta_t = 0 \). As we all know, since Engle (1982) and Bollerslev (1986), model (1.1)–(1.2) has been widely used in economics and finance; see Bollerslev, Chou and Kroner (1992), Bera and Higgins (1993), Bollerslev, Engle and Nelson (1994) and Francq and Zakoïan (2010). The asymptotic theory of the quasi-maximum likelihood estimator (QMLE) was established by Ling and Li (1997) and by Francq and Zakoïan (2004) when \( E\epsilon_t^4 < \infty \). Under the strict stationarity condition, the consistency and the asymptotic normality of the QMLE were obtained by Lee and Hansen (1994) and Lumsdaine (1996) for the GARCH(1, 1) model, and by Berkes, Horváth and Kokoszka (2003) and Francq and Zakoïan (2004) for the GARCH(r, s) model. Hall and Yao (2003)
established the asymptotic theory of the QMLE for the GARCH model when \( E \varepsilon_t^2 < \infty \), including both cases in which \( E \eta_t^4 = \infty \) and \( E \eta_t^4 < \infty \). Under the geometric ergodicity condition, Lang, Rahbek and Jensen (2011) gave the asymptotic properties of the modified QMLE for the first order AR–ARCH model. Moreover, when \( E |\varepsilon_t|^{\xi} < \infty \) for some \( \xi > 0 \), the asymptotic theory of the global self-weighted QMLE and the local QMLE was established by Ling (2007) for model (1.1)–(1.2).

It is well known that the asymptotic normality of the QMLE requires \( E \eta_t^4 < \infty \) and this property is lost when \( E \eta_t^4 = \infty \); see Hall and Yao (2003). Usually, the least absolute deviation (LAD) approach can be used to reduce the moment condition of \( \eta_t \) and provide a robust estimator. The local LAD estimator was studied by Peng and Yao (2003) and Li and Li (2005) for the pure GARCH model, Chan and Peng (2005) for the double AR(1) model, and Li and Li (2008) for the ARFIMA–GARCH model. The global LAD estimator was studied by Horváth and Liese (2004) for the pure ARCH model and by Berkes and Horváth (2004) for the pure GARCH model, and by Zhu and Ling (2011a) for the double AR(\( p \)) model. Except for the AR models studied by Davis, Knight and Liu (1992) and Ling (2005) [see also Knight (1987, 1998)], the nondifferentiable and nonconvex objective function appears when one studies the LAD estimator for the ARMA model with i.i.d. errors. By assuming the existence of a \( \sqrt{n} \)-consistent estimator, the asymptotic normality of the LAD estimator is established for the ARMA model with i.i.d. errors by Davis and Dunsmuir (1997) for the finite variance case and by Pan, Wang and Yao (2007) for the infinite variance case; see also Wu and Davis (2010) for the noncausal or noninvertible ARMA model. Recently, Zhu and Ling (2011b) proved the asymptotic normality of the global LAD estimator for the finite/infinite variance ARMA model with i.i.d. errors.

In this paper, we investigate the self-weighted quasi-maximum exponential likelihood estimator (QMELE) for model (1.1)–(1.2). Under only a fractional moment condition of \( \varepsilon_t \) with \( E \eta_t^2 < \infty \), the strong consistency and the asymptotic normality of the global self-weighted QMELE are obtained by using the bracketing method in Pollard (1985). Based on this global self-weighted QMELE, the local QMELE is showed to be asymptotically normal for the ARMA–GARCH (finite variance) and –IGARCH models. A formal comparison of two estimators is given for some cases.

To motivate our estimation procedure, we revisit the GNP deflator example of Bollerslev (1986), in which the GARCH model was proposed for the first time. The model he specified is an AR(4)–GARCH(1, 1) model for the quarterly data from 1948.2 to 1983.4 with a total of 143 observations. We use this data set and his fitted model to obtain the residuals \( \{\hat{\eta}_t\} \). The tail index of \( \{\eta_t^2\} \) is estimated by Hill’s estimator \( \hat{\alpha}_\eta(k) \) with the largest \( k \) data of \( \{\hat{\eta}_t^2\} \), that is,

\[
\hat{\alpha}_\eta(k) = \frac{k}{\sum_{j=1}^{k} (\log \hat{\eta}_{143-j} - \log \hat{\eta}_{143-k})},
\]
where \( \tilde{\eta}_j \) is the \( j \)th order statistic of \( \{\hat{\eta}_i^2\} \). The plot of \( \{\hat{\alpha}_\eta(k)\}_{k=1}^{70} \) is given in Figure 1. From this figure, we can see that \( \hat{\alpha}_\eta(k) > 2 \) when \( k \leq 20 \), and \( \hat{\alpha}_\eta(k) < 2 \) when \( k > 20 \). Note that Hill’s estimator is not so reliable when \( k \) is too small. Thus, the tail of \( \{\eta_i^2\} \) is most likely less than 2, that is, \( E\eta_i^4 = \infty \). Thus, the setup that \( \eta_i \) has a finite forth moment may not be suitable, and hence the standard QMLE procedure may not be reliable in this case. The estimation procedure in this paper only requires \( E\eta_i^2 < \infty \). It may provide a more reliable alternative to practitioners. To further illustrate this advantage, a simulation study is carried out to compare the performance of our estimators and the self-weighted/local QMLE in Ling (2007), and a new real example on the world crude oil price is given in this paper.

This paper is organized as follows. Section 2 gives our results on the global self-weighted QMELE. Section 3 proposes a local QMELE estimator and gives its limiting distribution. The simulation results are reported in Section 4. A real example is given in Section 5. The proofs of two technical lemmas are provided in Section 6. Concluding remarks are offered in Section 7. The remaining proofs are given in the Appendix.

2. Global self-weighted QMELE. Let \( \theta = (\gamma', \delta')' \) be the unknown parameter of model (1.1)–(1.2) and its true value be \( \theta_0 \), where \( \gamma = (\mu, \phi_1, \ldots, \phi_p, \psi_1, \ldots, \psi_q)' \) and \( \delta = (\alpha_0, \ldots, \alpha_r, \beta_1, \ldots, \beta_s)' \). Given the observations \( \{y_n, \ldots, y_1\} \) and the initial values \( Y_0 \equiv \{y_0, y_{-1}, \ldots\} \), we can rewrite the parametric model (1.1)–(1.2)
as
\begin{equation}
\varepsilon_t(\gamma) = y_t - \mu - \sum_{i=1}^{p} \phi_i y_{t-i} - \sum_{i=1}^{q} \psi_i \varepsilon_{t-i}(\gamma),
\end{equation}
\begin{equation}
\eta_t(\theta) = \varepsilon_t(\gamma)/\sqrt{h_t(\theta)} \quad \text{and}
\end{equation}
\begin{equation}
h_t(\theta) = \alpha_0 + \sum_{i=1}^{r} \alpha_i \varepsilon_{t-i}^2(\gamma) + \sum_{i=1}^{s} \beta_i h_{t-i}(\theta).
\end{equation}

Here, \( \eta_t(\theta_0) = \eta_t, \ varepsilon_t(\gamma_0) = \varepsilon_t \) and \( h_t(\theta_0) = h_t \). The parameter space is \( \theta = \Theta_\gamma \times \Theta_\delta \), where \( \Theta_\gamma \subset R^{p+q+1}, \Theta_\delta \subset R^{r+s+1}, R = (-\infty, \infty) \) and \( R_0 = [0, \infty) \). Assume that \( \Theta_\gamma \) and \( \Theta_\delta \) are compact and \( \theta_0 \) is an interior point in \( \Theta \). Denote \( \alpha(z) = \sum_{i=1}^{r} \alpha_iz^i, \beta(z) = 1 - \sum_{i=1}^{s} \beta_iz^i, \phi(z) = 1 - \sum_{i=1}^{p} \phi_iz^i \) and \( \psi(z) = 1 + \sum_{i=1}^{q} \psi_iz^i \). We introduce the following assumptions:

**ASSUMPTION 2.1.** For each \( \theta \in \Theta, \phi(z) \neq 0 \) and \( \psi(z) \neq 0 \) when \( |z| \leq 1 \), and \( \phi(z) \) and \( \psi(z) \) have no common root with \( \phi_p \neq 0 \) or \( \psi_q \neq 0 \).

**ASSUMPTION 2.2.** For each \( \theta \in \Theta, \alpha(z) \) and \( \beta(z) \) have no common root, \( \alpha(1) \neq 1, \alpha_r + \beta_s \neq 0 \) and \( \sum_{i=1}^{s} \beta_i < 1 \).

**ASSUMPTION 2.3.** \( \eta_t^2 \) has a nondegenerate distribution with \( E\eta_t^2 < \infty \).

Assumption 2.1 implies the stationarity, invertibility and identifiability of model (1.1), and Assumption 2.2 is the identifiability condition for model (1.2). Assumption 2.3 is necessary to ensure that \( \eta_t^2 \) is not almost surely (a.s.) a constant. When \( \eta_t \) follows the standard double exponential distribution, the weighted log-likelihood function (ignoring a constant) can be written as follows:

\begin{equation}
L_{sn}(\theta) = \frac{1}{n} \sum_{t=1}^{n} w_t l_t(\theta) \quad \text{and} \quad l_t(\theta) = \log \sqrt{h_t(\theta)} + \frac{|\varepsilon_t(\gamma)|}{\sqrt{h_t(\theta)}},
\end{equation}

where \( w_t = w(y_{t-1}, y_{t-2}, \ldots) \) and \( w \) is a measurable, positive and bounded function on \( R^{Z_0} \) with \( Z_0 = \{0, 1, 2, \ldots\} \). We look for the minimizer, \( \hat{\theta}_{sn} = (\hat{\gamma}_{sn}', \hat{\delta}_{sn}') \), of \( L_{sn}(\theta) \) on \( \Theta \), that is,

\begin{equation}
\hat{\theta}_{sn} = \arg \min_{\Theta} L_{sn}(\theta).
\end{equation}

Since the weight \( w_t \) only depends on \( \{y_t\} \) itself and we do not assume that \( \eta_t \) follows the standard double exponential distribution, \( \hat{\theta}_{sn} \) is called the self-weighted quasi-maximum exponential likelihood estimator (QMELE) of \( \theta_0 \). When \( h_t \) is a constant, the self-weighted QMELE reduces to the weighted LAD estimator of the ARMA model in Pan, Wang and Yao (2007) and Zhu and Ling (2011b).

The weight \( w_t \) is to reduce the moment condition of \( \varepsilon_t \) [see more discussions in Ling (2007)], and it satisfies the following assumption:
ASSUMPTION 2.4. \( E[(w_t + w_t^2)\xi_{\rho t-1}^3] < \infty \) for any \( \rho \in (0, 1) \), where \( \xi_{\rho t} = 1 + \sum_{i=0}^{\infty} \rho^i |y_{t-i}|. \)

When \( w_t \equiv 1 \), the \( \hat{\theta}_{sn} \) is the global QMELE and it needs the moment condition \( E|\varepsilon_t|^3 < \infty \) for its asymptotic normality, which is weaker than the moment condition \( E\varepsilon_t^4 < \infty \) as for the QMLE of \( \theta_0 \) in Francq and Zakoïan (2004). It is well known that the higher is the moment condition of \( \varepsilon_t \), the smaller is the parameter space. Figure 2 gives the strict stationarity region and regions for \( E|\varepsilon_t|^{2i} < \infty \) of the GARCH(1, 1) model: \( \varepsilon_t = \eta_t \sqrt{h_t} \) and \( h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1} \), where \( \eta_t \sim \text{Laplace}(0, 1) \). From Figure 2, we can see that the region for \( E|\varepsilon_t|^{0.1} < \infty \) is

![Parameter region of GARCH(1,1) model](image)

**Fig. 2.** The regions bounded by the indicated curves are for the strict stationarity and for \( E|\varepsilon_t|^{2i} < \infty \) with \( i = 0.05, 0.5, 1, 1.5 \) and 2, respectively.
very close to the region for strict stationarity of $\varepsilon_t$, and is much bigger than the region for $E\varepsilon_t^4 < \infty$.

Under Assumption 2.4, we only need a fractional moment condition for the asymptotic property of $\hat{\theta}_{sn}$ as follows:

**Assumption 2.5.** $E|\varepsilon_t|^{2\iota} < \infty$ for some $\iota > 0$.

The sufficient and necessary condition of Assumption 2.5 is given in Theorem 2.1 of Ling (2007). In practice, we can use Hill’s estimator to estimate the tail index of $\{y_t\}$ and its estimator may provide some useful guidelines for the choice of $\iota$. For instance, the quantity $2\iota$ can be any value less than the tail index $\{y_t\}$. However, so far we do not know how to choose the optimal $\iota$. As in Ling (2007) and Pan, Wang and Yao (2007), we choose the weight function $w_t$ according to $\iota$. When $\iota = 1/2$ (i.e., $E|\varepsilon_t| < \infty$), we can choose the weight function as

$$w_t = \left( \max \left\{ 1, C^{-1} \sum_{k=1}^{\infty} \frac{1}{k^{9}} |y_{t-k}| I\{|y_{t-k}| > C\} \right\} \right)^{-4}, \quad (2.4)$$

where $C > 0$ is a constant. In practice, it works well when we select $C$ as the 90% quantile of data $\{y_1, \ldots, y_n\}$. When $q = s = 0$ (AR–ARCH model), for any $\iota > 0$, the weight can be selected as

$$w_t = \left( \max \left\{ 1, C^{-1} \sum_{k=1}^{p+r} \frac{1}{k^{9}} |y_{t-k}| I\{|y_{t-k}| > C\} \right\} \right)^{-4}.$$

When $\iota \in (0, 1/2)$ and $q > 0$ or $s > 0$, the weight function need to be modified as follows:

$$w_t = \left( \max \left\{ 1, C^{-1} \sum_{k=1}^{\infty} \frac{1}{k^{1+8/\iota}} |y_{t-k}| I\{|y_{t-k}| > C\} \right\} \right)^{-4}.$$

Obviously, these weight functions satisfy Assumptions 2.4 and 2.7. For more choices of $w_t$, we refer to Ling (2005) and Pan, Wang and Yao (2007). We first state the strong convergence of $\hat{\theta}_{sn}$ in the following theorem and its proof is given in the Appendix.

**Theorem 2.1.** Suppose $\eta_t$ has a median zero with $E|\eta_t| = 1$. If Assumptions 2.1–2.5 hold, then

$$\hat{\theta}_{sn} \to \theta_0 \quad \text{a.s., as } n \to \infty.$$

To study the rate of convergence of $\hat{\theta}_{sn}$, we reparameterize the weighted log-likelihood function (2.3) as follows:

$$L_n(u) \equiv nL_{sn}(\theta_0 + u) - nL_{sn}(\theta_0),$$

where $L_{sn}(\omega)$ is the weighted log-likelihood function.
where $u \in \Lambda \equiv \{(u_1', u_2') : u + \theta_0 \in \Theta\}$. Let $\hat{\theta}_n = \hat{\theta}_n - \theta_0$. Then, $\hat{u}_n$ is the minimizer of $L_n(u)$ on $\Lambda$. Furthermore, we have

$$L_n(u) = \sum_{t=1}^{n} w_t A_t(u) + \sum_{t=1}^{n} w_t B_t(u) + \sum_{t=1}^{n} w_t C_t(u),$$

where

$$A_t(u) = \frac{1}{\sqrt{h_t(\theta_0)}} |\varepsilon_t(\gamma_0 + u_1) - |\varepsilon_t(\gamma_0)||,$$

$$B_t(u) = \log \sqrt{h_t(\theta_0 + u)} - \log \sqrt{h_t(\theta_0)} + \frac{|\varepsilon_t(\gamma_0)|}{\sqrt{h_t(\theta_0 + u)}} - \frac{|\varepsilon_t(\gamma_0)|}{\sqrt{h_t(\theta_0)}},$$

$$C_t(u) = \left[ \frac{1}{\sqrt{h_t(\theta_0 + u)}} - \frac{1}{\sqrt{h_t(\theta_0)}} \right] [|\varepsilon_t(\gamma_0 + u_1)| - |\varepsilon_t(\gamma_0)|].$$

Let $I(\cdot)$ be the indicator function. Using the identity

$$|x - y| - |x| = -y[I(x > 0) - I(x < 0)]$$

$$+ 2 \int_0^y [I(x \leq s) - I(x \leq 0)] ds$$

for $x \neq 0$, we can show that

$$A_t(u) = q_t(u)[I(\eta_t > 0) - I(\eta_t < 0)] + 2 \int_0^{-q_t(u)} X_t(s) ds,$$

where $X_t(s) = I(\eta_t \leq s) - I(\eta_t \leq 0)$, $q_t(u) = q_{1t}(u) + q_{2t}(u)$ with

$$q_{1t}(u) = \frac{u'}{\sqrt{h_t(\theta_0)}} \frac{\partial \varepsilon_t(\gamma_0)}{\partial \theta} [I(\eta_t > 0) - I(\eta_t < 0)],$$

and $\xi^*$ lies between $\gamma_0$ and $\gamma_0 + u_1$. Moreover, let $\mathcal{F}_t = \sigma\{\eta_k : k \leq t\}$ and

$$\xi_t(u) = 2w_t \int_0^{-q_t(u)} X_t(s) ds.$$

Then, from (2.7), we have

$$\sum_{t=1}^{n} w_t A_t(u) = u' T_{1n} + \Pi_{1n}(u) + \Pi_{2n}(u) + \Pi_{3n}(u),$$

where

$$T_{1n} = \sum_{t=1}^{n} \frac{w_t}{\sqrt{h_t(\theta_0)}} \frac{\partial \varepsilon_t(\gamma_0)}{\partial \theta} [I(\eta_t > 0) - I(\eta_t < 0)],$$

$$\Pi_{1n}(u) = \sum_{t=1}^{n} \{\xi_t(u) - E[\xi_t(u)|\mathcal{F}_{t-1}]\}.$$
\[ \Pi_{2n}(u) = \sum_{t=1}^{n} E[\xi_t(u)|\mathcal{F}_{t-1}], \]
\[ \Pi_{3n}(u) = \sum_{t=1}^{n} w_t q_2(u) [I(\eta_t > 0) - I(\eta_t < 0)] \]
\[ + 2 \sum_{t=1}^{n} w_t \int_{-q_1(u)}^{q_1(u)} X_t(s) ds. \]

By Taylor’s expansion, we can see that
\[ \sum_{t=1}^{n} w_t B_t(u) = u^T T_{2n} + \Pi_{4n}(u) + \Pi_{5n}(u), \]
where
\[ T_{2n} = \sum_{t=1}^{n} \frac{w_t}{2h_t(\theta_0)} \frac{\partial h_t(\theta_0)}{\partial \theta} (1 - |\eta_t|), \]
\[ \Pi_{4n}(u) = u' \sum_{t=1}^{n} w_t \left( \frac{3}{8} \frac{\epsilon_t(\gamma_0)}{\sqrt{h_t(\xi^*)}} - \frac{1}{4} \frac{1}{h_t^2(\xi^*)} \frac{\partial h_t(\xi^*)}{\partial \theta} \frac{\partial h_t(\xi^*)}{\partial \theta'} u, \right) \]
\[ \Pi_{5n}(u) = u' \sum_{t=1}^{n} w_t \left( \frac{1}{4} - \frac{1}{4} \frac{1}{\sqrt{h_t(\xi^*)}} \right) \frac{1}{h_t(\xi^*)} \frac{\partial^2 h_t(\xi^*)}{\partial \theta \partial \theta'} u, \]
and \( \xi^* \) lies between \( \theta_0 \) and \( \theta_0 + u \).

We further need one assumption and three lemmas. The first lemma is directly from the central limit theorem for a martingale difference sequence. The second and third-lemmas give the expansions of \( \Pi_{in}(u) \) for \( i = 1, \ldots, 5 \) and \( \sum_{t=1}^{n} C_t(u) \). The key technical argument is for the second lemma for which we use the bracketing method in Pollard (1985).

**Assumption 2.6.** \( \eta_t \) has zero median with \( E|\eta_t| = 1 \) and a continuous density function \( g(x) \) satisfying \( g(0) > 0 \) and \( \sup_{x \in \mathbb{R}} g(x) < \infty \).

**Lemma 2.1.** Let \( T_n = T_{1n} + T_{2n} \). If Assumptions 2.1–2.6 hold, then
\[ \frac{1}{\sqrt{n}} T_n \rightarrow_d N(0, \Omega_0) \quad \text{as} \ n \rightarrow \infty, \]
where \( \rightarrow_d \) denotes the convergence in distribution and
\[ \Omega_0 = E \left( \frac{w_t^2}{h_t(\theta_0)} \frac{\partial^2 \epsilon_t(\gamma_0)}{\partial \theta \partial \theta'} + \frac{E h_t^2(\theta_0) - 1}{4} E \frac{w_t^2}{h_t^2(\theta_0)} \frac{\partial h_t(\theta_0)}{\partial \theta} \frac{\partial h_t(\theta_0)}{\partial \theta'} \right). \]
LEMMA 2.2. If Assumptions 2.1–2.6 hold, then for any sequence of random variables \( u_n \) such that \( u_n = o_p(1) \), it follows that
\[
\Pi_{1n}(u_n) = o_p(\sqrt{n} \| u_n \| + n \| u_n \|^2),
\]
where \( o_p(\cdot) \to 0 \) in probability as \( n \to \infty \).

LEMMA 2.3. If Assumptions 2.1–2.6 hold, then for any sequence of random variables \( u_n \) such that \( u_n = o_p(1) \), it follows that:

(i) \( \Pi_{2n}(u_n) = (\sqrt{n} u_n)^\prime \Sigma_1(\sqrt{n} u_n) + o_p(n \| u_n \|^2) \),

(ii) \( \Pi_{3n}(u_n) = o_p(n \| u_n \|^2) \),

(iii) \( \Pi_{4n}(u_n) = (\sqrt{n} u_n)^\prime \Sigma_2(\sqrt{n} u_n) + o_p(n \| u_n \|^2) \),

(iv) \( \Pi_{5n}(u_n) = o_p(n \| u_n \|^2) \),

(v) \( \sum_{t=1}^n C_t(u_n) = o_p(n \| u_n \|^2) \),

where
\[
\Sigma_1 = g(0) E\left( \frac{w_t}{h_t(\theta_0)} \frac{\partial \varepsilon_t(\gamma_0)}{\partial \theta} \frac{\partial \varepsilon_t(\gamma_0)}{\partial \theta^\prime} \right)
\]
and
\[
\Sigma_2 = \frac{1}{8} E\left( \frac{w_t}{h_t^2(\theta_0)} \frac{\partial h_t(\theta_0)}{\partial \theta} \frac{\partial h_t(\theta_0)}{\partial \theta^\prime} \right).
\]

The proofs of Lemmas 2.2 and 2.3 are given in Section 6. We now can state one main result as follows:

THEOREM 2.2. If Assumptions 2.1–2.6 hold, then:

(i) \( \sqrt{n}(\hat{\theta}_n - \theta_0) = O_p(1) \),

(ii) \( \sqrt{n}(\hat{\theta}_n - \theta_0) \to_d N(0, \Sigma_0^{-1} \Omega_0 \Sigma_0^{-1}) \) as \( n \to \infty \),

where \( \Sigma_0 = \Sigma_1 + \Sigma_2 \).

PROOF. (i) First, we have \( \hat{u}_n = o_p(1) \) by Theorem 2.1. Furthermore, by (2.5), (2.8) and (2.9) and Lemmas 2.2 and 2.3, we have
\[
L_n(\hat{u}_n) = \hat{u}_n^\prime T_n + (\sqrt{n} \hat{u}_n)^\prime \Sigma_0(\sqrt{n} \hat{u}_n) + o_p(\sqrt{n} \| \hat{u}_n \|^2 + n \| \hat{u}_n \|^2)
\]
Let \( \lambda_{\text{min}} > 0 \) be the minimum eigenvalue of \( \Sigma_0 \). Then
\[
L_n(\hat{u}_n) \geq -\| \sqrt{n} \hat{u}_n \| \left[ \frac{1}{\sqrt{n}} T_n + o_p(1) \right] + n \| \hat{u}_n \|^2 [\lambda_{\text{min}} + o_p(1)]
\]
Note that $L_n(\hat{u}_n) \leq 0$. By the previous inequality, it follows that

$$\sqrt{n}\|\hat{u}_n\| \leq [\lambda_{\min} + o_p(1)]^{-1} \left[ \frac{1}{\sqrt{n}} T_n + o_p(1) \right] = O_p(1), \tag{2.11}$$

where the last step holds by Lemma 2.1. Thus, (i) holds.

(ii) Let $u_n^* = -\Sigma_0^{-1} T_n / 2n$. Then, by Lemma 2.1, we have

$$\sqrt{n}u_n^* \rightarrow_d N(0, \frac{1}{4} \Sigma_0^{-1} \Omega_0 \Sigma_0^{-1}) \quad \text{as } n \rightarrow \infty.$$ 
Hence, it is sufficient to show that $\sqrt{n}\hat{u}_n - \sqrt{n}u_n^* = o_p(1)$. By (2.10) and (2.11), we have

$$L_n(\hat{u}_n) = \left( \sqrt{n}\hat{u}_n \right)' \frac{1}{\sqrt{n}} T_n + \left( \sqrt{n}\hat{u}_n \right)' \Sigma_0 (\sqrt{n}\hat{u}_n) + o_p(1)$$
$$= \left( \sqrt{n}\hat{u}_n \right)' \Sigma_0 (\sqrt{n}\hat{u}_n) - 2 \left( \sqrt{n}\hat{u}_n \right)' \Sigma_0 (\sqrt{n}u_n^*) + o_p(1).$$
Note that (2.10) still holds when $\hat{u}_n$ is replaced by $u_n^*$. Thus,

$$L_n(u_n^*) = \left( \sqrt{n}u_n^* \right)' \frac{1}{\sqrt{n}} T_n + \left( \sqrt{n}u_n^* \right)' \Sigma_0 (\sqrt{n}u_n^*) + o_p(1)$$
$$= -\left( \sqrt{n}u_n^* \right)' \Sigma_0 (\sqrt{n}u_n^*) + o_p(1).$$
By the previous two equations, it follows that

$$L_n(\hat{u}_n) - L_n(u_n^*) = \left( \sqrt{n}\hat{u}_n - \sqrt{n}u_n^* \right)' \Sigma_0 (\sqrt{n}\hat{u}_n - \sqrt{n}u_n^*) + o_p(1) \geq \lambda_{\min} \|\sqrt{n}\hat{u}_n - \sqrt{n}u_n^*\|^2 + o_p(1). \tag{2.12}$$
Since $L_n(\hat{u}_n) - L_n(u_n^*) = n[L_{sn}(\theta_0 + \hat{u}_n) - L_{sn}(\theta_0 + u_n^*)] \leq 0$ a.s., by (2.12), we have $\|\sqrt{n}\hat{u}_n - \sqrt{n}u_n^*\| = o_p(1)$. This completes the proof. □

Remark 2.1. When $w_t \equiv 1$, the limiting distribution in Theorem 2.2 is the same as that in Li and Li (2008). When $r = s = 0$ (ARMA model), it reduces to the case in Pan, Wang and Yao (2007) and Zhu and Ling (2011b). In general, it is not easy to compare the asymptotic efficiency of the self-weighted QMELE and the self-weight QMLE in Ling (2007). However, for the pure ARCH model, a formal comparison of these two estimators is given in Section 3. For the general ARMA–GARCH model, a comparison based on simulation is given in Section 4.

In practice, the initial values $Y_0$ are unknown, and have to be replaced by some constants. Let $\tilde{\epsilon}_t(\theta)$, $\tilde{h}_t(\theta)$ and $\tilde{w}_t$ be $\epsilon_t(\theta)$, $h_t(\theta)$ and $w_t$, respectively, when $Y_0$ are constants not depending on parameters. Usually, $Y_0$ are taken to be zeros. The objective function (2.3) is modified as

$$\tilde{L}_{sn}(\theta) = \frac{1}{n} \sum_{t=1}^{n} \tilde{w}_t \tilde{l}_t(\theta) \quad \text{and} \quad \tilde{l}_t(\theta) = \log \sqrt{\tilde{h}_t(\theta)} + \frac{\tilde{\epsilon}_t(\theta)}{\sqrt{\tilde{h}_t(\theta)}}.$$
To make the initial values $Y_0$ ignorable, we need the following assumption.
ASSUMPTION 2.7. \( E|w_t - \tilde{w}_t|^{\gamma_0/4} = O(t^{-2}) \), where \( t_0 = \min\{\tau, 1\} \).

Let \( \tilde{\theta}_{sn} \) be the minimizer of \( \tilde{L}_{sn}(\theta) \), that is,
\[
\tilde{\theta}_{sn} = \arg \min_{\Theta} \tilde{L}_{sn}(\theta).
\]

Theorem 2.3 below shows that \( \tilde{\theta}_{sn} \) and \( \hat{\theta}_{sn} \) have the same limiting property. Its proof is straightforward and can be found in Zhu (2011).

**THEOREM 2.3.** Suppose that Assumption 2.7 holds. Then, as \( n \to \infty \),

(i) \( \tilde{\theta}_{sn} \to \theta_0 \) a.s.,

(ii) \( \sqrt{n}(\tilde{\theta}_{sn} - \theta_0) \to d N(0, \frac{1}{4} \Sigma_0^{-1} \Omega_0 \Sigma_0^{-1}) \).

3. **Local QMELE.** The self-weighted QMELE in Section 2 reduces the moment condition of \( \varepsilon_t \), but it may not be efficient. In this section, we propose a local QMELE based on the self-weighted QMELE and derive its asymptotic property. For some special cases, a formal comparison of the local QMELE and the self-weighted QMELE is given.

Using \( \hat{\theta}_{sn} \) in Theorem 2.2 as an initial estimator of \( \theta_0 \), we obtain the local QMELE \( \hat{\theta}_n \) through the following one-step iteration:

\[
\hat{\theta}_n = \hat{\theta}_{sn} - 2\Sigma^*_n(\hat{\theta}_{sn})^{-1} T^*_n(\hat{\theta}_{sn}),
\]

where
\[
\Sigma^*_n(\theta) = \sum_{t=1}^{n} \left\{ g(0) \frac{\partial \varepsilon_t(\gamma)}{\partial \theta} \frac{\partial \varepsilon_t(\gamma)}{\partial \theta'} + \frac{1}{8h_t^2(\theta)} \frac{\partial h_t(\theta)}{\partial \theta} \frac{\partial h_t(\theta)}{\partial \theta'} \right\},
\]
\[
T^*_n(\theta) = \sum_{t=1}^{n} \left\{ \frac{1}{\sqrt{h_t(\theta)}} \frac{\partial \varepsilon_t(\gamma)}{\partial \theta} \left[ I(\eta_t(\theta) > 0) - I(\eta_t(\theta) < 0) \right] + \frac{1}{2h_t(\theta)} \frac{\partial h_t(\theta)}{\partial \theta} (1 - |\eta_t(\theta)|) \right\}.
\]

In order to get the asymptotic normality of \( \hat{\theta}_n \), we need one more assumption as follows:

**ASSUMPTION 3.1.** \( E\eta_t^2 \sum_{i=1}^{r} \alpha_{0i} + \sum_{i=1}^{s} \beta_{0i} < 1 \) or
\[
E\eta_t^2 \sum_{i=1}^{r} \alpha_{0i} + \sum_{i=1}^{s} \beta_{0i} = 1
\]
with \( \eta_t \) having a positive density on \( \mathbb{R} \) such that \( E|\eta_t|^{\tau} < \infty \) for all \( \tau < \tau_0 \) and \( E|\eta_t|^{\tau_0} = \infty \) for some \( \tau_0 \in (0, \infty) \).
Under Assumption 3.1, there exists a unique strictly stationary causal solution to GARCH model (1.2); see Bougerol and Picard (1992) and Basrak, Davis and Mikosch (2002). The condition $E\eta_t^2 \sum_{i=1}^r \alpha_{0i} + \sum_{i=1}^s \beta_{0i} < 1$ is necessary and sufficient for $E\varepsilon_t^2 < \infty$ under which model (1.2) has a finite variance. When $E\eta_t^2 \sum_{i=1}^r \alpha_{0i} + \sum_{i=1}^s \beta_{0i} = 1$, model (1.2) is called IGARCH model. The IGARCH model has an infinite variance, but $E|\varepsilon_t|^3 < \infty$ for all $t \in (0, 1)$ under Assumption 3.1; see Ling (2007). Assumption 3.1 is crucial for the ARMA–IGARCH model. From Figure 2 in Section 2, we can see that the parameter region specified in Assumption 3.1 is much bigger than that for $E|\varepsilon_t|^3 < \infty$ which is required for the asymptotic normality of the global QMELE. Now, we give one lemma as follows and its proof is straightforward and can be found in Zhu (2011).

**Lemma 3.1.** If Assumptions 2.1–2.3, 2.6 and 3.1 hold, then for any sequence of random variables $\theta_n$ such that $\sqrt{n}(\theta_n - \theta_0) = O_p(1)$, it follows that:

(i) $\frac{1}{n} [T_n^*(\theta_n) - T_n^*(\theta_0)] = [2\Sigma + o_p(1)](\theta_n - \theta_0) + o_p\left(\frac{1}{\sqrt{n}}\right)$,

(ii) $\frac{1}{n}\Sigma_n^*(\theta_n) = \Sigma + o_p(1)$,

(iii) $\frac{1}{\sqrt{n}}T_n^*(\theta_0) \rightarrow_d N(0, \Omega)$ as $n \rightarrow \infty$,

where

$\Omega = E\left(\frac{1}{h_t(\theta_0)} \frac{\partial \varepsilon_t(\gamma_0)}{\partial \theta} \frac{\partial \varepsilon_t(\gamma_0)}{\partial \theta^\prime} \right) + \frac{E\eta_t^2 - 1}{4} E\left(\frac{1}{h_t(\theta_0)} \frac{\partial h_t(\theta_0)}{\partial \theta} \frac{\partial h_t(\theta_0)}{\partial \theta^\prime} \right)$.

$\Sigma = g(0) E\left(\frac{1}{h_t(\theta_0)} \frac{\partial \varepsilon_t(\gamma_0)}{\partial \theta} \frac{\partial \varepsilon_t(\gamma_0)}{\partial \theta^\prime} \right) + \frac{1}{8} E\left(\frac{1}{h_t^2(\theta_0)} \frac{\partial h_t(\theta_0)}{\partial \theta} \frac{\partial h_t(\theta_0)}{\partial \theta^\prime} \right)$.

**Theorem 3.1.** If the conditions in Lemma 3.1 are satisfied, then

$\sqrt{n}(\hat{\theta}_n - \theta_0) \rightarrow_d N\left(0, \frac{1}{4} \Sigma^{-1} \Omega \Sigma^{-1}\right)$ as $n \rightarrow \infty$.

**Proof.** Note that $\sqrt{n}(\hat{\theta}_n - \theta_0) = O_p(1)$. By (3.1) and Lemma 3.1, we have that

$\hat{\theta}_n = \hat{\theta}_n - \left[\frac{2}{n}\Sigma_n^*(\hat{\theta}_n)\right]^{-1}\left[\frac{1}{n}T_n^*(\hat{\theta}_n)\right]$

$= \hat{\theta}_n - [2\Sigma + o_p(1)]^{-1}\left\{\frac{1}{n}T_n^*(\theta_0) + [2\Sigma + o_p(1)](\hat{\theta}_n - \theta_0) + o_p\left(\frac{1}{\sqrt{n}}\right)\right\}$

$= \theta_0 + \frac{\Sigma^{-1}T_n^*(\theta_0)}{2n} + o_p\left(\frac{1}{\sqrt{n}}\right)$. 

It follows that
\[ \sqrt{n}(\hat{\theta}_n - \theta_0) = \frac{\Sigma^{-1}T_n^*(\theta_0)}{2\sqrt{n}} + o_p(1). \]

By Lemma 3.1(iii), we can see that the conclusion holds. This completes the proof. \( \square \)

**Remark 3.1.** In practice, by using \( \tilde{\theta}_{sn} \) in Theorem 2.3 as an initial estimator of \( \theta_0 \), the local QMELE has to be modified as follows:
\[
\hat{\theta}_n = \tilde{\theta}_{sn} - \left[ 2\tilde{\Sigma}_n^*(\tilde{\theta}_{sn}) \right]^{-1} \tilde{T}_n^*(\tilde{\theta}_{sn}),
\]
where \( \tilde{\Sigma}_n^*(\theta) \) and \( \tilde{T}_n^*(\theta) \) are defined in the same way as \( \Sigma_n^*(\theta) \) and \( T_n^*(\theta) \), respectively, with \( \varepsilon_t(\theta) \) and \( h_t(\theta) \) being replaced by \( \tilde{\varepsilon}_t(\theta) \) and \( \tilde{h}_t(\theta) \). However, this does not affect the asymptotic property of \( \hat{\theta}_n \); see Theorem 4.3.2 in Zhu (2011).

We now compare the asymptotic efficiency of the local QMELE and the self-weighted QMELE. First, we consider the pure ARMA model, that is, model (1.1)–(1.2) with \( h_t \) being a constant. In this case,
\[ \Omega_0 = E(w_t^2 X_{1t} X'_{1t}), \quad \Sigma_0 = g(0) E(w_t X_{1t} X'_{1t}), \]
\[ \Omega = E(X_{1t} X'_{1t}) \quad \text{and} \quad \Sigma = g(0) \Omega, \]
where \( X_{1t} = h_t^{-1/2} \partial \varepsilon_t(\gamma_0)/\partial \theta \). Let \( b \) and \( c \) be two any \( m \)-dimensional constant vectors. Then,
\[ c' \Sigma_0 b b' \Sigma_0 c = \left[ E \left( \left( c' \sqrt{g(0) w_t X_{1t}} \right) \left( \sqrt{g(0) X'_{1t}} \right) \right) \right]^2 \leq E(c' \sqrt{g(0) w_t X_{1t}})^2 E(\sqrt{g(0) X'_{1t}})^2 \]
\[ = [c' g(0) \Omega_0 c] [b' \Sigma b] = c' [g(0) \Omega_0 b' \Sigma b] c. \]
Thus, \( g(0) \Omega_0 b' \Sigma b - \Sigma_0 b b' \Sigma_0 \geq 0 \) (a positive semi-definite matrix) and hence
\[ b' \Sigma_0 \Omega_0^{-1} \Sigma_0 b = \text{tr}(\Omega_0^{-1/2} \Sigma_0 b b' \Sigma_0 \Omega_0^{-1/2}) \leq \text{tr}(g(0) b' \Sigma b) = g(0) b' \Sigma b. \]
It follows that \( \Sigma_0^{-1} \Omega_0^{-1} \Sigma_0 \geq [g(0) \Sigma]^{-1} = \Sigma^{-1} \Omega \Sigma^{-1} \). Thus, the local QMELE is more efficient than the self-weighted QMELE. Similarly, we can show that the local QMELE is more efficient than the self-weighted QMELE for the pure GARCH model.

For the general model (1.1)–(1.2), it is not easy to compare the asymptotic efficiency of the self-weighted QMELE and the local QMELE. However, when \( \eta_t \sim \text{Laplace}(0, 1) \), we have
\[ \Sigma_0 = E \left( \frac{w_t}{2} X_{1t} X'_{1t} + \frac{w_t}{8} X_{2t} X'_{2t} \right), \]
\[ \Omega_0 = E \left( w_t^2 X_{1t} X'_{1t} + \frac{w_t^2}{4} X_{2t} X'_{2t} \right), \]
\[ \Sigma = E(\frac{1}{2} X_{1t} X'_{1t} + \frac{1}{8} X_{2t} X'_{2t}) \quad \text{and} \quad \Omega = 2 \Sigma, \]
where $X_{2t} = h_t^{-1} \partial h_t(\theta_0)/\partial \theta$. Then, it is easy to see that
\[ c^* \Sigma_0 \Sigma_0^{bb'} \Sigma_0 c = \{ E[(c' 2^{-1/4} w_t X_{1t})(2^{-3/4} X_{1t}^t b) + (c' 2^{-5/4} w_t X_{2t})(2^{-7/4} X_{2t}^t b)] \}^2 \]
\[ \leq \left\{ \sqrt{E(c' 2^{-1/4} w_t X_{1t})^2 E(2^{-3/4} X_{1t}^t b)^2} \right\}^2 \]
\[ + \left\{ \sqrt{E(c' 2^{-5/4} w_t X_{2t})^2 E(2^{-7/4} X_{2t}^t b)^2} \right\}^2 \]
\[ \leq [E(c' 2^{-1/4} w_t X_{1t})^2 + E(c' 2^{-5/4} w_t X_{2t})^2] \times [E(2^{-3/4} X_{1t}^t b)^2 + E(2^{-7/4} X_{2t}^t b)^2] \]
\[ = [c' 2^{-1/2} \Sigma_0 c][b' 2^{-1/2} \Sigma b] = c'[2^{-1} \Sigma_0 b' \Sigma b] c. \]

Thus, $2^{-1} \Sigma_0 b' \Sigma b - \Sigma_0 \Sigma_0^{-1} \Sigma_0 b = \text{tr}(\Sigma_0^{-1/2} \Sigma_0 b' \times \Sigma_0 \Sigma_0^{-1/2}) \leq \text{tr}(b' \Sigma b) = 2^{-1} b' \Sigma b$. It follows that $\Sigma_0^{-1} \Sigma_0^{-1} \geq 2 \Sigma^{-1} = \Sigma^{-1} \Sigma^{-1}$. Thus, the local QMLE is more efficient than the global self-weighted QMLE.

In the end, we compare the asymptotic efficiency of the self-weighted QMLE and the self-weighted QMLE in Ling (2007) for the pure ARCH model, when $E \eta_t^4 < \infty$. We reparametrize model (1.2) when $s = 0$ as follows:
\[ (3.2) \quad y_t = \eta_t^* \sqrt{h_t^*} \quad \text{and} \quad h_t^* = \alpha_0^* + \sum_{i=1}^{r} \alpha_i^* y_{t-i}^2, \]
where $\eta_t^* = \eta_t / \sqrt{E \eta_t^2}$, $h_t^* = (E \eta_t^2) h_t$ and $\theta^* = (\alpha_0^*, \alpha_1^*, \ldots, \alpha_r^*)' = (E \eta_t^2)^{1/2} \theta$. Let $\hat{\theta}_{sn}^*$ be the self-weighted QMLE of the true parameter, $\theta_0^*$, in model (3.2). Then, $\hat{\theta}_{sn}^* / E \eta_t^2$ is the self-weighted QMLE of $\theta_0$, and its asymptotic covariance is
\[ \Gamma_1 = \kappa_1 [(E(w_t X_{2t} X_{2t}^t))^{-1} E(w_t^2 X_{2t} X_{2t}^t)] (E(w_t X_{2t} X_{2t}^t))^{-1}, \]
where $\kappa_1 = E \eta_t^4 / (E \eta_t^2)^2 - 1$. By Theorem 2.2, the asymptotic variance of the self-weighted QMLE is
\[ \Gamma_2 = \kappa_2 [(E(w_t X_{2t} X_{2t}^t))^{-1} E(w_t^2 X_{2t} X_{2t}^t)] (E(w_t X_{2t} X_{2t}^t))^{-1}, \]
where $\kappa_2 = 4(E \eta_t^2 - 1)$. When $\eta_t \sim \text{Laplace}(0, 1)$, $\kappa_1 = 5$ and $\kappa_2 = 4$. Thus, $\Gamma_1 > \Gamma_2$, meaning that the self-weighted QMLE is more efficient than the self-weighted QMLE. When $\eta_t = \tilde{\eta}_t / E|\tilde{\eta}_t|$, with $\tilde{\eta}_t$ having the following mixing normal density:
\[ f(x) = (1 - \varepsilon) \phi(x) + \frac{\varepsilon}{\tau} \phi \left( \frac{x}{\tau} \right), \]
we have $E|\eta_t| = 1$.
\[ E \eta_t^2 = \frac{\pi (1 - \varepsilon + \varepsilon \tau^2)}{2(1 - \varepsilon + \varepsilon \tau)^2}, \]
and
\[ E\eta_i^4 = \frac{3\pi (1 - \varepsilon + \varepsilon \tau^4)}{2(1 - \varepsilon + \varepsilon \tau)^2(1 - \varepsilon + \varepsilon \tau^2)}, \]
where \( \phi(x) \) is the pdf of standard normal, \( 0 \leq \varepsilon \leq 1 \) and \( \tau > 0 \). The asymptotic efficiencies of the self-weighted QMELE and the self-weighted QMLE depend on \( \varepsilon \) and \( \tau \). For example, when \( \varepsilon = 1 \) and \( \tau = \sqrt{\pi/2} \), we have \( \kappa_1 = (6 - \pi)/\pi \) and \( \kappa_2 = 2\pi - 4 \), and hence the self-weighted QMLE is more efficient than the self-weighted QMELE since \( \Gamma_1 < \Gamma_2 \). When \( \varepsilon = 0.99 \) and \( \tau = 0.1 \), we have \( \kappa_1 = 28.1 \) and \( \kappa_2 = 6.5 \), and hence the self-weighted QMELE is more efficient than the self-weighted QMLE since \( \Gamma_1 > \Gamma_2 \).

4. Simulation. In this section, we compare the performance of the global self-weighted QMELE \( (\hat{\theta}_{sn}) \), the global self-weighted QMLE \( (\hat{\theta}_{sn}) \), the local QMELE \( (\hat{\theta}_n) \) and the local QMLE \( (\hat{\theta}_n) \). The following AR(1)–GARCH(1, 1) model is used to generate data samples:

\[ y_t = \mu + \phi_1 y_{t-1} + \varepsilon_t, \]
\[ \varepsilon_t = \eta_t \sqrt{h_t} \quad \text{and} \quad h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1}. \]

We set the sample size \( n = 1,000 \) and use 1,000 replications, and study the cases when \( \eta_t \) has Laplace \((0, 1), N(0, 1) \) and \( t_3 \) distribution. For the case with \( \varepsilon^2 = \beta \leq 1 \) (i.e., \( \varepsilon^2 = \alpha_0 + \beta_0 = 1 \)), we take \( \theta_0 = (0.0, 0.5, 0.1, 0.18, 0.4) \). For the IGARCH case (i.e., \( \varepsilon^2 = \alpha_0 + \beta_0 = 1 \)), we take \( \theta_0 = (0.0, 0.5, 0.1, 0.3, 0.4) \) when \( \eta_t \sim \text{Laplace}(0, 1) \), \( \theta_0 = (0.0, 0.5, 0.1, 0.6, 0.4) \) when \( \eta_t \sim N(0, 1) \) and \( \theta_0 = (0.0, 0.5, 0.1, 0.2, 0.4) \) and \( \eta_t \sim t_3 \). We standardize the distribution of \( \eta_t \) to ensure that \( E|\eta_t| = 1 \) for the QMELE. Tables 1–3 list the sample biases, the sample standard deviations (SD) and the asymptotic standard deviations (AD) of \( \hat{\theta}_{sn}, \hat{\theta}_{sn}, \hat{\theta}_n \) and \( \hat{\theta}_n \). We choose \( w_t \) as in (2.4) with \( C \) being 90% quantile of \( \{y_1, \ldots, y_n\} \) and \( y_i \equiv 0 \) for \( i \leq 0 \). The ADs in Theorems 2.2 and 3.1 are estimated by \( \hat{\chi}_{sn} = 1/4 \hat{\Sigma}_{sn}^{-1} \hat{\Omega}_{sn} \hat{\Sigma}_{sn}^{-1} \) and \( \hat{\chi}_n = 1/4 \hat{\Sigma}_n^{-1} \hat{\Omega}_n \hat{\Sigma}_n^{-1} \), respectively, where

\[ \hat{\Sigma}_{sn} = \frac{1}{n} \sum_{t=1}^{n} \left\{ \frac{g(0)w_t}{h_t(\hat{\theta}_{sn})} \frac{\partial \varepsilon_t(\hat{\theta}_{sn})}{\partial \theta} \frac{\partial \varepsilon_t(\hat{\theta}_{sn})}{\partial \theta'} + \frac{w_t}{8h^2_t(\hat{\theta}_{sn})} \frac{\partial h_t(\hat{\theta}_{sn})}{\partial \theta} \frac{\partial h_t(\hat{\theta}_{sn})}{\partial \theta'} \right\}, \]
\[ \hat{\Sigma}_n = \frac{1}{n} \sum_{t=1}^{n} \left\{ \frac{g(0)}{h_t(\hat{\theta}_n)} \frac{\partial \varepsilon_t(\hat{\theta}_n)}{\partial \theta} \frac{\partial \varepsilon_t(\hat{\theta}_n)}{\partial \theta'} + \frac{1}{8h^2_t(\hat{\theta}_n)} \frac{\partial h_t(\hat{\theta}_n)}{\partial \theta} \frac{\partial h_t(\hat{\theta}_n)}{\partial \theta'} \right\}, \]
\[ \hat{\Omega}_{sn} = \frac{1}{n} \sum_{t=1}^{n} \left\{ \frac{w_t^2}{h_t(\hat{\theta}_{sn})} \frac{\partial \varepsilon_t(\hat{\theta}_{sn})}{\partial \theta} \frac{\partial \varepsilon_t(\hat{\theta}_{sn})}{\partial \theta'} + \frac{E\eta_t^2 - 1}{4h^2_t(\hat{\theta}_{sn})} \frac{\partial h_t(\hat{\theta}_{sn})}{\partial \theta} \frac{\partial h_t(\hat{\theta}_{sn})}{\partial \theta'} \right\}, \]
\[ \hat{\Omega}_n = \frac{1}{n} \sum_{t=1}^{n} \left\{ \frac{1}{h_t(\hat{\theta}_n)} \frac{\partial \varepsilon_t(\hat{\theta}_n)}{\partial \theta} \frac{\partial \varepsilon_t(\hat{\theta}_n)}{\partial \theta'} + \frac{E\eta_t^2 - 1}{4h^2_t(\hat{\theta}_n)} \frac{\partial h_t(\hat{\theta}_n)}{\partial \theta} \frac{\partial h_t(\hat{\theta}_n)}{\partial \theta'} \right\}. \]
and the local QMLE. When QMELE has smaller AD and SD than those of both the self-weighted QMLE since their asymptotic variances are infinite, while the SD and AD of the self-weighted QMLE and the local QMLE are not close to each other.

From Table 1, when \( \eta_t \sim \text{Laplace}(0, 1) \), we can see that the self-weighted QMELE has smaller AD and SD than those of both the self-weighted QMLE and the local QMLE. When \( \eta_t \sim N(0, 1) \), in Table 2, we can see that the self-weighted QMLE has smaller AD and SD than those of both the self-weighted QMELE and the local QMLE. From Table 3, we note that the SD and AD of both the self-weighted QMLE and the local QMLE are not close to each other since their asymptotic variances are infinite, while the SD and AD of the self-weighted QMELE and the local QMELE are very close to each other. Except \( \tilde{\theta}_n \) in Table 3, we can see that all four estimators in Tables 1–3 have very small biases, and the local QMELE and local QMLE always have the smaller SD and AD than those of the self-weighted QMELE and self-weighted QMLE, respectively. This conclusion holds no matter with GARCH errors (finite variance) or IGARCH er-

\[
\theta_0 = (0.0, 0.5, 0.1, 0.18, 0.4)
\]

**Table 1**

*Estimators for model (4.1) when \( \eta_t \sim \text{Laplace}(0, 1) \)*

| \( \theta_0 = (0.0, 0.5, 0.1, 0.18, 0.4) \) | \( \theta_0 = (0.0, 0.5, 0.1, 0.3, 0.4) \) |
|---------------------------------------------|---------------------------------------------|
| \( \hat{\mu}_{sn} \) | \( \hat{\mu}_{sn} \) | \( \hat{\phi}_{1sn} \) | \( \hat{\phi}_{1sn} \) | \( \hat{\alpha}_{0sn} \) | \( \hat{\alpha}_{1sn} \) | \( \hat{\alpha}_{1sn} \) | \( \hat{\beta}_{1sn} \) | \( \hat{\beta}_{1sn} \) |
| Bias | 0.0004 | 0.0003 | 0.0023 | 0.0034 | 0.0078 | -0.0154 | 0.0003 | -0.0049 | 0.0031 | 0.0054 | -0.0068 |
| SD | 0.0172 | 0.0195 | 0.0317 | 0.0274 | 0.0548 | 0.1125 | 0.0192 | 0.0311 | 0.0218 | 0.0640 | 0.0673 |
| AD | 0.0166 | 0.0190 | 0.0304 | 0.0255 | 0.0540 | 0.1061 | 0.0192 | 0.0311 | 0.0218 | 0.0624 | 0.0664 |

\[
\theta_0 = (0.0, 0.5, 0.1, 0.3, 0.4)
\]

| \( \hat{\mu}_{sn} \) | \( \hat{\phi}_{1sn} \) | \( \hat{\alpha}_{0sn} \) | \( \hat{\alpha}_{1sn} \) | \( \hat{\alpha}_{1sn} \) | \( \hat{\beta}_{1sn} \) | \( \hat{\beta}_{1sn} \) |
| Bias | 0.0008 | 0.0010 | 0.0019 | 0.0027 | 0.0002 | -0.0094 | 0.0003 | -0.0044 | 0.0024 | -0.0008 | -0.0025 |
| SD | 0.0170 | 0.0192 | 0.0253 | 0.0249 | 0.0400 | 0.0989 | 0.0192 | 0.0261 | 0.0203 | 0.0502 | 0.0591 |
| AD | 0.0162 | 0.0190 | 0.0245 | 0.0234 | 0.0407 | 0.0920 | 0.0190 | 0.0258 | 0.0206 | 0.0499 | 0.0591 |

\[
\hat{\theta}_n
\]

| \( \hat{\mu}_{sn} \) | \( \hat{\phi}_{1sn} \) | \( \hat{\alpha}_{0sn} \) | \( \hat{\alpha}_{1sn} \) | \( \hat{\alpha}_{1sn} \) | \( \hat{\beta}_{1sn} \) | \( \hat{\beta}_{1sn} \) |
| Bias | -0.0003 | 0.0005 | -0.0016 | 0.0041 | 0.0114 | -0.0227 | 0.0005 | -0.0039 | 0.0031 | 0.0104 | -0.0127 |
| SD | 0.0243 | 0.0283 | 0.0451 | 0.0301 | 0.0624 | 0.1237 | 0.0283 | 0.0458 | 0.0242 | 0.0750 | 0.0755 |
| AD | 0.0240 | 0.0283 | 0.0443 | 0.0285 | 0.0607 | 0.1184 | 0.0283 | 0.0461 | 0.0243 | 0.0704 | 0.0741 |

\[
\tilde{\theta}_n
\]

| \( \hat{\mu}_{sn} \) | \( \hat{\phi}_{1sn} \) | \( \hat{\alpha}_{0sn} \) | \( \hat{\alpha}_{1sn} \) | \( \hat{\alpha}_{1sn} \) | \( \hat{\beta}_{1sn} \) | \( \hat{\beta}_{1sn} \) |
| Bias | 0.0007 | 0.0022 | -0.0034 | 0.0026 | 0.0037 | -0.1444 | 0.0022 | -0.0045 | 0.0020 | 0.0044 | -0.0081 |
| SD | 0.0243 | 0.0282 | 0.0368 | 0.0279 | 0.0461 | 0.1115 | 0.0282 | 0.0377 | 0.0227 | 0.0579 | 0.0674 |
| AD | 0.0236 | 0.0281 | 0.0361 | 0.0261 | 0.0459 | 0.1026 | 0.0281 | 0.0384 | 0.0230 | 0.0564 | 0.0659 |
heavy-tailed innovations. Local QMELE have a good performance in the finite sample, especially for the 710 observations; see Figure 3(a). Its 100 times log-return, denoted by \( \text{dollars per barrel} \) from January 3, 1997 to August 6, 2010, which has in total also Ling (2007) for a discussion.

5. A real example. In this section, we study the weekly world crude oil price (dollars per barrel) from January 3, 1997 to August 6, 2010, which has in total 710 observations; see Figure 3(a). Its 100 times log-return, denoted by \( \{y_t\}_{t=1}^{T_0} \), is plotted in Figure 3(b). The classic method based on the Akaike’s information
where the standard errors are in parentheses, and the estimated value of $\sigma^2$ is 16.83. Model (5.1) is stationary, and none of the first ten autocorrelations or partial autocorrelations of the residuals $\{\hat{\epsilon}_t\}$ are significant at the 5% level. However, looking at the autocorrelations of $\{\hat{\epsilon}_t^2\}$, it turns out that the 1st, 2nd and 8th all exceed two asymptotic standard errors; see Figure 4(a). Similar results hold for the partial autocorrelations of $\{\hat{\epsilon}_t^2\}$ in Figure 4(b). This shows that $\{\hat{\epsilon}_t^2\}$ may be highly correlated, and hence there may exist ARCH effects.

\begin{equation}
y_t = 0.2876\epsilon_{t-1} + 0.1524\epsilon_{t-3} + \epsilon_t,
\end{equation}

\begin{equation}
(0.0357) \quad (0.0357)
\end{equation}
Thus, we try to use a MA(3)–GARCH(1, 1) model to fit the data set \{y_t\}. To begin with our estimation, we first estimate the tail index of \{y_t^2\} by using Hill’s estimator \{\hat{\alpha}_y(k)\} with \(k = 1, \ldots, 180\), based on \{y_t^2\}_{t=1}^{709}. The plot of \{\hat{\alpha}_y(k)\}_{k=1}^{180} is given in Figure 5, from which we can see that the tail index of \{y_t^2\} is between 1 and 2, that is, \(E y_t^4 = \infty\). So, the standard QMLE procedure is not suitable. Therefore, we first use the self-weighted QMELE to estimate the MA(3)–GARCH(1, 1) model, and then use the one-step iteration as in Section 3 to obtain
its local QMELE. The fitted model is as follows:

\[ y_t = 0.3276 \varepsilon_{t-1} + 0.1217 \varepsilon_{t-3} + \varepsilon_t, \]

\[ h_t = 0.5147 + 0.0435 \varepsilon_{t-1}^2 + 0.8756 h_{t-1}, \]

\[ (5.2) \]

where the standard errors are in parentheses. Again model (5.2) is stationary, and none of the first ten autocorrelations or partial autocorrelations of the residuals \( \hat{\varepsilon}_t \triangleq \hat{\varepsilon}_t \hat{h}_t^{-1/2} \) are significant at the 5% level. Moreover, the first ten autocorrelations and partial autocorrelations of \( \{\hat{\eta}_t^2\} \) are also within two asymptotic standard errors; see Figure 6(a) and (b). All these results suggest that model (5.2) is adequate for the data set \( \{y_t\} \).

Finally, we estimate the tail index of \( \eta_t^2 \) in model (5.2) by using Hill’s estimator \( \hat{\alpha}_\eta(k) \) with \( k = 1, \ldots, 180 \), base on \( \{\hat{\eta}_t^2\} \). The plot of \( \{\hat{\alpha}_\eta(k)\}_{k=1}^{180} \) is given in Figure 7, from which we can see that \( E\eta_t^2 \) is most likely finite, but \( E\eta_t^4 \) is infinite. Furthermore, the estimator of \( E\eta_t^2 \) is \( \sum_{i=1}^n \hat{\eta}_t^2 / n = 1.6994 \), and it turns out that \( \hat{\alpha}_1 \{\sum_{i=1}^n \hat{\eta}_t^2 / n\} + \hat{\beta}_1 n = 0.9495 \). This means that \( E\varepsilon_t^2 < \infty \). Therefore, all the assumptions of Theorem 3.1 are most likely satisfied. In particular, the estimated tail indices of \( \{y_t^2\} \) and \( \{\hat{\eta}_t^2\} \) show the evidence that the self-weighted/local QMELE is necessary in modeling the crude oil price.

6. Proofs of Lemmas 2.2 and 2.3. In this section, we give the proofs of Lemmas 2.2 and 2.3. In the rest of this paper, we denote \( C \) as a universal constant, and \( G(x) \) be the distribution function of \( \eta_t \).
FIG. 6. (a) The autocorrelations for $\hat{\eta}^2_t$ and (b) the partial autocorrelations for $\hat{\eta}^2_t$.

PROOF OF LEMMA 2.2. A direct calculation gives

$$\xi_t(u) = -u' \frac{2w_t}{\sqrt{h_t}} \frac{\partial \varepsilon_t(\gamma_0)}{\partial \theta} M_t(u),$$

where $M_t(u) = \int_0^1 X_t(-q_{1t}(u)s) ds$. Thus, we have

$$|\Pi_{1n}(u)| \leq 2\|u\| \sum_{j=1}^m \frac{w_t}{\sqrt{h_t}} \frac{\partial \varepsilon_t(\gamma_0)}{\partial \theta_j} \sum_{t=1}^n \left[M_t(u) - E[M_t(u)|F_{t-1}]\right].$$

FIG. 7. The Hill estimators $\hat{\alpha}_n(k)$ for $\hat{\eta}^2_t$ of model (5.2).
It is sufficient to show that
\[
\left| \frac{w_t}{\sqrt{h_t}} \frac{\partial \epsilon_t(\theta_0)}{\partial \theta_j} \sum_{i=1}^{n} \{M_t(u_n) - E[M_t(u_n)|\mathcal{F}_{t-1}]\} \right| = o_p(\sqrt{n} + n\|u_n\|),
\]
for each \(1 \leq j \leq m\).

Let \(m_t = w_t h_t^{-1/2} \frac{\partial \epsilon_t(\theta_0)}{\partial \theta_j} \), \(f_t(u) = m_t M_t(u)\) and
\[
D_n(u) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \{f_t(u) - E[f_t(u)|\mathcal{F}_{t-1}]\}.
\]

Then, in order to prove (6.1), we only need to show that for any \(\eta > 0\),
\[
\sup_{\|u\| \leq \eta} |D_n(u)| = o_p(1).
\]

Note that \(m_t = \max\{m_t, 0\} - \max\{-m_t, 0\}\). To make it simple, we only prove the case when \(m_t \geq 0\).

We adopt the method in Lemma 4 of Pollard (1985). Let \(\mathcal{F} = \{f_t(u) : \|u\| \leq \eta\}\) be a collection of functions indexed by \(u\). We first verify that \(\mathcal{F}\) satisfies the bracketing condition in Pollard (1985), page 304. Denote \(\mathcal{B}\) to be an open neighborhood of \(\zeta\) with radius \(r > 0\). For any \(\varepsilon > 0\) and \(0 < \delta \leq \eta\), there is a sequence of small cubes \(\{B_{\varepsilon\delta/C_1}(u_i)\}_{i=1}^{K_\varepsilon}\) to cover \(B_\delta(0)\), where \(K_\varepsilon\) is an integer less than \(c_0 \varepsilon^{-m}\) and \(c_0\) is a constant not depending on \(\varepsilon\) and \(\delta\); see Huber (1967), page 227. Here, \(C_1\) is a constant to be selected later. Moreover, we can choose \(U_i(\delta) \subseteq B_{\varepsilon\delta/C_1}(u_i)\) such that \(\{U_i(\delta)\}_{i=1}^{K_\varepsilon}\) be a partition of \(B_\delta(0)\). For each \(u \in U_i(\delta)\), we define the bracketing functions as follows:
\[
f_t^{\pm}(u) = m_t \int_0^1 X_t(-q_{1t}(u)s \pm \frac{\varepsilon \delta}{C_1 \sqrt{h_t}} \frac{\partial \epsilon_t(\gamma_0)}{\partial \theta}) ds.
\]
Since the indicator function is nondecreasing and \(m_t \geq 0\), we can see that, for any \(u \in U_i(\delta)\),
\[
f_t^{-}(u_i) \leq f_t(u) \leq f_t^{+}(u_i).
\]

Note that \(\sup_{x \in R} g(x) < \infty\). It is straightforward to see that
\[
E[f_t^{+}(u_i) - f_t^{-}(u_i)|\mathcal{F}_{t-1}] \leq 2\varepsilon \delta \sup_{x \in R} g(x) \frac{w_t}{h_t} \left(\frac{\partial \epsilon_t(\gamma_0)}{\partial \theta}\right)^2 \leq \frac{\varepsilon \delta \Delta_t}{C_1}.
\]
Setting \(C_1 = E(\Delta_t)\), we have
\[
E[f_t^{+}(u_i) - f_t^{-}(u_i)] = E[E[f_t^{+}(u_i) - f_t^{-}(u_i)|\mathcal{F}_{t-1}] \leq \varepsilon \delta.
\]

Thus, the family \(\mathcal{F}\) satisfies the bracketing condition.

Put \(\delta_k = 2^{-k}\eta\). Define \(B(k) = B_{\delta_k}(0)\), and \(A(k)\) to be the annulus \(B(k)/B(k + 1)\). Fix \(\varepsilon > 0\), for each \(1 \leq i \leq K_\varepsilon\), by the bracketing condition, there exists a partition \(\{U_i(\delta_k)\}_{i=1}^{K_\varepsilon}\) of \(B(k)\).
We first consider the upper tail. For \( u \in U_i(\delta_k) \), by (6.3) with \( \delta = \delta_k \), we have

\[
D_n(u) \leq \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \{ f_t^+(u_i) - E[f_t^-(u_i)|\mathcal{F}_{t-1}] \}
\]

\[
= D_n^+(u_i) + \frac{1}{\sqrt{n}} \sum_{t=1}^{n} E[f_t^+(u_i) - f_t^-(u_i)|\mathcal{F}_{t-1}]
\]

\[
\leq D_n^+(u_i) + \sqrt{n} \varepsilon \delta_k \left[ \frac{1}{nC_1} \sum_{t=1}^{n} \Delta_t \right],
\]

where

\[
D_n^+(u_i) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \{ f_t^+(u_i) - E[f_t^+(u_i)|\mathcal{F}_{t-1}] \}.
\]

Denote the event

\[
E_n = \left\{ \omega : \frac{1}{nC_1} \sum_{t=1}^{n} \Delta_t(\omega) < 2 \right\}.
\]

On \( E_n \) with \( u \in U_i(\delta_k) \), it follows that

\[
D_n(u) \leq D_n^+(u_i) + 2\sqrt{n} \varepsilon \delta_k.
\]

On \( A(k) \), the divisor \( 1 + \sqrt{n} \|u\| > \sqrt{n} \delta_{k+1} = \sqrt{n} \delta_k / 2 \). Thus, by (6.4) and Chebyshev’s inequality, it follows that

\[
P\left( \sup_{u \in A(k)} \frac{D_n(u)}{1 + \sqrt{n} \|u\|} > 6 \varepsilon, E_n \right)
\]

\[
\leq P\left( \sup_{u \in A(k)} D_n(u) > 3\sqrt{n} \varepsilon \delta_k, E_n \right)
\]

\[
\leq P\left( \max_{1 \leq i \leq K_{\varepsilon}} \sup_{u \in U_i(\delta_k) \cap A(k)} D_n(u) > 3\sqrt{n} \varepsilon \delta_k, E_n \right)
\]

\[
(6.5)
\]

\[
\leq P\left( \max_{1 \leq i \leq K_{\varepsilon}} D_n^+(u_i) > \sqrt{n} \varepsilon \delta_k, E_n \right)
\]

\[
\leq K_{\varepsilon} \max_{1 \leq i \leq K_{\varepsilon}} P(D_n^+(u_i) > \sqrt{n} \varepsilon \delta_k)
\]

\[
\leq K_{\varepsilon} \max_{1 \leq i \leq K_{\varepsilon}} \frac{E[(D_n^+(u_i))^2]}{n \varepsilon^2 \delta_k^2}.
\]

Note that \( |q_{1t}(u_i)| \leq C \delta_k \xi_{\rho t-1} \) and \( m_t^2 \leq C \xi_{\rho t-1}^2 \) for some \( \rho \in (0, 1) \) by Lemma A.1(i), and \( \sup_{x \in \mathcal{R}} g(x) < \infty \) by Assumption 2.6. By Taylor’s expansion, we
have
\[ E[(f_i^+(u_i))^2] = E[E[(f_i^+(u_i))^2|\mathcal{F}_{t-1}]] \]
\[ \leq E \left[ \int_0^1 E \left[ X_t \left( -q_{1T}(u_i)s + \frac{\varepsilon \delta_k}{C_1\sqrt{t}} \left\| \frac{\partial \varepsilon_t(\gamma_0)}{\partial \theta} \right\| \right) \right] \left| \mathcal{F}_{t-1} \right| ds \right] \]
\[ \leq C E \left[ \sup_{|x| \leq \delta_k C_\varepsilon \rho_{t-1}} |G(x) - G(0)| w_{t}^2 \varepsilon_{\rho_{t-1}}^3 \right] \]
\[ \leq \delta_k C E(w_{t}^2 \varepsilon_{\rho_{t-1}}^3) \]
Since \( f_i^+(u_i) - E[f_i^+(u_i)|\mathcal{F}_{t-1}] \) is a martingale difference sequence, by the previous inequality, it follows that
\[ E[(D_n^+(u_i))^2] = \frac{1}{n} \sum_{t=1}^{n} E[f_i^+(u_i) - E[f_i^+(u_i)|\mathcal{F}_{t-1}]]^2 \]
\[ \leq \frac{1}{n} \sum_{t=1}^{n} E[(f_i^+(u_i))^2] \]
\[ \leq \frac{\delta_k}{n} \sum_{t=1}^{n} C E(w_{t}^2 \varepsilon_{\rho_{t-1}}^3) \]
\[ = \pi_n(\delta_k) \]
Thus, by (6.5) and (6.6), we have
\[ P \left( \sup_{u \in A(k)} \frac{D_n(u)}{1 + \sqrt{n} \|u\|} > 6\varepsilon, E_n \right) \leq K_{\varepsilon} \frac{\pi_n(\delta_k)}{n\varepsilon^2 \delta_k^2}. \]
By a similar argument, we can get the same bound for the lower tail. Thus, we can show that
\[ P \left( \sup_{u \in A(k)} \frac{|D_n(u)|}{1 + \sqrt{n} \|u\|} > 6\varepsilon, E_n \right) \leq 2K_{\varepsilon} \frac{\pi_n(\delta_k)}{n\varepsilon^2 \delta_k^2}. \]
Since \( \pi_n(\delta_k) \to 0 \) as \( k \to \infty \), we can choose \( k_{\varepsilon} \) so that
\[ 2\pi_n(\delta_k)K_{\varepsilon}/(\varepsilon \eta)^2 < \varepsilon \]
for \( k \geq k_{\varepsilon} \). Let \( k_n \) be an integer so that \( n^{-1/2} \leq 2^{-k_n} < 2n^{-1/2} \). Split \( \{ u : \|u\| \leq \eta \} \) into two sets \( B(k_n + 1) \) and \( B(k_n + 1)^c = \bigcup_{k=0}^{k_n} A(k) \). By (6.7), since \( \pi_n(\delta_k) \) is bounded, we have
\[ P \left( \sup_{u \in B(k_n+1)^c} \frac{|D_n(u)|}{1 + \sqrt{n} \|u\|} > 6\varepsilon \right) \]
\[ \leq \sum_{k=0}^{k_n} P \left( \sup_{u \in A(k)} \frac{|D_n(u)|}{1 + \sqrt{n} \|u\|} > 6\varepsilon, E_n \right) + P(E_n^c) \]
\[
\leq \frac{1}{n} \sum_{k=0}^{k_n-1} C K \varepsilon 2^k + \varepsilon \frac{1}{n} \sum_{k=k_\varepsilon}^{k_n} 2^k + P(E_n^c)
\]
\[
\leq O \left( \frac{1}{n} \right) + 4 \varepsilon \frac{2^k_n}{n} + P(E_n^c)
\]
\[
\leq O \left( \frac{1}{n} \right) + 4 \varepsilon + P(E_n^c).
\]

Since \(1 + \sqrt{n}\|u\| > 1\) and \(\sqrt{n}\delta_{k_n+1} < 1\), using a similar argument as for (6.5) together with (6.6), we have
\[
P \left( \sup_{u \in B(k_n+1)} \frac{D_n(u)}{1 + \sqrt{n}\|u\|} > 3\varepsilon, E_n \right) \leq P \left( \max_{1 \leq i \leq K_\varepsilon} D_n^+(u_i) > \varepsilon, E_n \right)
\]
\[
\leq \frac{K_\varepsilon \pi_n(\delta_{k_n+1})}{\varepsilon^2}.
\]

We can get the same bound for the lower tail. Thus, we have
\[
P \left( \sup_{u \in B(k_n+1)} \frac{|D_n(u)|}{1 + \sqrt{n}\|u\|} > 3\varepsilon \right)
\]
\[
= P \left( \sup_{u \in B(k_n+1)} \frac{|D_n(u)|}{1 + \sqrt{n}\|u\|} > 3\varepsilon, E_n \right) + P(E_n^c)
\]
\[
\leq \frac{2 K_\varepsilon \pi_n(\delta_{k_n+1})}{\varepsilon^2} + P(E_n^c).
\]

Note that \(\pi_n(\delta_{k_n+1}) \to 0\) as \(n \to \infty\). Furthermore, \(P(E_n) \to 1\) by the ergodic theorem. Hence,
\[
P(E_n^c) \to 0\quad \text{as } n \to \infty.
\]

Finally, (6.2) follows by (6.8) and (6.9). This completes the proof. \(\square\)

**Proof of Lemma 2.3.** (i). By a direct calculation, we have
\[
\Pi_{2n}(u) = 2 \sum_{t=1}^{n} w_t \int_0^{-q_{11}(u)} G(s) - G(0) \, ds
\]
\[
= 2 \sum_{t=1}^{n} w_t \int_0^{-q_{11}(u)} sg(\varsigma^*) \, ds
\]
\[
= (\sqrt{n}\|u\|) [K_1n + K_{2n}(u)](\sqrt{n}\|u\|),
\]
where \(\varsigma^*\) lies between 0 and \(s\), and
\[
K_{1n} = \frac{g(0)}{n} \sum_{t=1}^{n} \frac{w_t}{h_t(\theta_0)} \frac{\partial \varepsilon_t(\gamma_0)}{\partial \theta} \frac{\partial \varepsilon_t(\gamma_0)}{\partial \theta'},
\]
\[ K_{2n}(u) = \frac{2}{n\|u\|^2} \sum_{t=1}^{n} w_t \int_{-q_{1t}(u)}^{0} s[g(\zeta^*) - g(0)] ds. \]

By the ergodic theorem, it is easy to see that

\[ K_{1n} = \Sigma_1 + o_p(1). \]

Furthermore, since \(|q_{1t}(u)| \leq C \|u\| \xi \rho t - 1\) for some \(\rho \in (0, 1)\) by Lemma A.1(i), it is straightforward to see that for any \(\eta > 0\),

\[ \sup_{\|u\| \leq \eta} |K_{2n}(u)| \leq \frac{2}{n\|u\|^2} \sum_{t=1}^{n} w_t \int_{-|q_{1t}(u)|}^{\|q_{1t}(u)\|} s|g(\zeta^*) - g(0)| ds \]
\[ \leq \frac{1}{n} \sum_{t=1}^{n} \sup_{|s| \leq C \eta \rho t - 1} |g(s) - g(0)| w_t \xi^2_{\rho t - 1} \]

By Assumptions 2.4 and 2.6, \(E(w_t \xi^2_{\rho t - 1}) < \infty\) and \(\sup_{x \in R} g(x) < \infty\). Then, by the dominated convergence theorem, we have

\[ \lim_{\eta \to 0} E\left[ \sup_{|s| \leq C \eta \rho t - 1} |g(s) - g(0)| w_t \xi^2_{\rho t - 1} \right] = 0. \]

Thus, by the stationarity of \(\{y_t\}\) and Markov’s theorem, for \(\forall \varepsilon, \delta > 0\), \(\exists \eta_0(\varepsilon) > 0\), such that

\[ P\left( \sup_{\|u\| \leq \eta_0} |K_{2n}(u)| > \delta \right) < \frac{\varepsilon}{2} \]

for all \(n \geq 1\). On the other hand, since \(u_n = o_p(1)\), it follows that

\[ P(\|u_n\| > \eta_0) < \frac{\varepsilon}{2} \]

as \(n\) is large enough. By (6.12) and (6.13), for \(\forall \varepsilon, \delta > 0\), we have

\[ P\left( |K_{2n}(u_n)| > \delta \right) \leq P\left( |K_{2n}(u_n)| > \delta, \|u_n\| \leq \eta_0 \right) + P(\|u_n\| > \eta_0) \]
\[ < P\left( \sup_{\|u\| \leq \eta_0} |K_{2n}(u)| > \delta \right) + \frac{\varepsilon}{2} \]
\[ < \varepsilon \]

as \(n\) is large enough, that is, \(K_{2n}(u_n) = o_p(1)\). Furthermore, combining (6.10) and (6.11), we can see that (i) holds.

(ii) Let \(\Pi_{3n}(u) = (\sqrt{n}u)'K_{3n}(\xi^*)(\sqrt{n}u) + K_{4n}(u)\), where

\[ K_{3n}(\xi^*) = \frac{1}{n} \sum_{t=1}^{n} \frac{w_t}{\sqrt{n}_t} \frac{\partial^2 \epsilon_t(\xi^*)}{\partial \theta \partial \theta'} [I(\eta_t > 0) - I(\eta_t < 0)], \]

\[ K_{4n}(u) = \sum_{t=1}^{n} w_t \int_{-q_{1t}(u)}^{0} X_t(s) ds. \]
By Assumption 2.4 and Lemma A.1(i), there exists a constant \( \rho \in (0, 1) \) such that

\[
E \left( \sup_{\xi^* \in \Lambda} \frac{w_t}{\sqrt{h_t}} \left| \frac{\partial^2 \epsilon_t(\xi^*)}{\partial \theta \partial \theta'} [I(\eta_t > 0) - I(\eta_t < 0)] \right| \right) \leq C E(w_t \xi_{\rho t-1}^2) < \infty.
\]

Since \( \eta_t \) has median 0, the conditional expectation property gives

\[
E \left( \frac{w_t}{\sqrt{h_t}} \frac{\partial^2 \epsilon_t(\xi^*)}{\partial \theta \partial \theta'} [I(\eta_t > 0) - I(\eta_t < 0)] \right) = 0.
\]

Then, by Theorem 3.1 in Ling and McAleer (2003), we have

\[
\sup_{\xi^* \in \Lambda} |K_{3n}(\xi^*)| = o_p(1).
\]

On the other hand,

\[
\frac{K_{4n}(u)}{n\|u\|^2} = \frac{2}{n} \sum_{t=1}^{n} w_t \int_{0}^{-q_{2t}(u)/\|u\|^2} X_t(\|u\|^2 s - q_{1t}(u)) \, ds \equiv \frac{2}{n} \sum_{t=1}^{n} J_{1t}(u).
\]

By Lemma A.1, we have \( \|u\|^2 q_{2t}(u) \leq C \xi_{\rho t-1}^2 \) and \( |q_{1t}(u)| \leq C \|u\| \xi_{\rho t-1} \) for some \( \rho \in (0, 1) \). Then, for any \( \eta > 0 \), we have

\[
\sup_{\|u\| \leq \eta} |J_{1t}(u)| \leq w_t \int_{-C \xi_{\rho t-1}}^{C \xi_{\rho t-1}} \left\{ X_t(C \eta^2 \xi_{\rho t-1} - C \eta \xi_{\rho t-1}) - X_t(-C \eta^2 \xi_{\rho t-1} - C \eta \xi_{\rho t-1}) \right\} \, ds \\
\leq 2C w_t \xi_{\rho t-1} \int_{-C \xi_{\rho t-1}}^{C \xi_{\rho t-1}} \left\{ X_t(C \eta^2 \xi_{\rho t-1} - C \eta \xi_{\rho t-1}) - X_t(-C \eta^2 \xi_{\rho t-1} - C \eta \xi_{\rho t-1}) \right\} \, ds.
\]

By Assumptions 2.4 and 2.6 and the double expectation property, it follows that

\[
E \left[ \sup_{\|u\| \leq \eta} |J_{1t}(u)| \right] \leq 2CE[w_t \xi_{\rho t-1} \{G(C \eta^2 \xi_{\rho t-1} - C \eta \xi_{\rho t-1}) - G(-C \eta^2 \xi_{\rho t-1} - C \eta \xi_{\rho t-1})\}] \\
\leq C(\eta^2 + \eta) \sup_{x} g(x) E(w_t \xi_{\rho t-1}^2) \to 0
\]
as \( \eta \to 0 \). Thus, as for (6.12) and (6.13), we can show that \( K_{4n}(u_n) = o_p(n\|u_n\|^2) \).

This completes the proof of (ii).

(iii) Let \( \Pi_n(u) = (\sqrt{n} u) \left( n^{\frac{1}{2}} \sum_{t=1}^{n} J_{2t}(\xi^*) \right)(\sqrt{n} u) \), where

\[
J_{2t}(\xi^*) = w_t \left( \frac{3}{8} \frac{\epsilon_t(\gamma_0)}{\sqrt{h_t(\xi^*)}} - \frac{1}{4} \frac{1}{h_t(\xi^*)} \frac{\partial h_t(\xi^*)}{\partial \theta} \frac{\partial h_t(\xi^*)}{\partial \theta'} \right).
\]
By Assumption 2.4 and Lemma A.1(ii)–(iv), there exists a constant \( \rho \in (0, 1) \) and a neighborhood \( \Theta_0 \) of \( \theta_0 \) such that
\[
E\left[ \sup_{\xi^* \in \Theta_0} |J_{2l}(\xi^*)| \right] \leq C E[|w_l\xi_{\rho l-1}^2(\eta_l|\xi_{\rho l-1}| + 1)] < \infty.
\]
Then, by Theorem 3.1 of Ling and McAleer (2003), we have
\[
\sup_{\xi^* \in \Theta_0} \left| \frac{1}{n} \sum_{t=1}^{n} J_{2l}(\xi^*) - E[J_{2l}(\xi^*)] \right| = o_p(1).
\]
Moreover, since \( \xi_n^* \to \theta_0 \) a.s., by the dominated convergence theorem, we have
\[
\lim_{n \to \infty} E[J_{2l}(\xi_n^*)] = E[J_{2l}(\theta_0)] = \Sigma_2.
\]
Thus, (iii) follows from the previous two equations. This completes the proof of (iii).

(iv) Since \( E|\eta_l| = 1 \), a similar argument as for part (iii) shows that (iv) holds.

(v) By Taylor’s expansion, we have
\[
\frac{1}{\sqrt{h_t(\theta_0 + u)}} - \frac{1}{\sqrt{h_t(\theta_0)}} = \frac{-u'}{2(h_t(\xi^*))^{3/2}} \frac{\partial h_t(\xi^*)}{\partial \theta},
\]
where \( \xi^* \) lies between \( \theta_0 \) and \( \theta_0 + u \). By identity (2.6), it is easy to see that
\[
|\varepsilon_t(\gamma_0 + u_1)| - |\varepsilon_t(\gamma_0)| = u' \frac{\partial \varepsilon_t(\xi^*)}{\partial \theta} [I(\eta_l > 0) - I(\eta_l < 0)] + 2u' \frac{\partial \varepsilon_t(\xi^*)}{\partial \theta} \int_0^1 X_t \left( \frac{-u'}{\sqrt{h_t}} \frac{\partial \varepsilon_t(\xi^*)}{\partial \theta} s \right) ds,
\]
where \( \xi^* \) lies between \( \gamma_0 \) and \( \gamma_0 + u_1 \). By the previous two equations, it follows that
\[
\sum_{t=1}^{n} w_tC_t(u) = (\sqrt{n}u)'[K_{5n}(u) + K_{6n}(u)](\sqrt{n}u),
\]
where
\[
K_{5n}(u) = \frac{1}{n} \sum_{t=1}^{n} \frac{w_t}{2h_t^{3/2}(\xi^*)} \frac{\partial h_t(\xi^*)}{\partial \theta} \frac{\partial \varepsilon_t(\xi^*)}{\partial \theta} [I(\eta_l < 0) - I(\eta_l > 0)],
\]
\[
K_{6n}(u) = -\frac{1}{n} \sum_{t=1}^{n} \frac{w_t}{h_t^{3/2}(\xi^*)} \frac{\partial h_t(\xi^*)}{\partial \theta} \frac{\partial \varepsilon_t(\xi^*)}{\partial \theta} \int_0^1 X_t \left( \frac{-u'}{\sqrt{h_t}} \frac{\partial \varepsilon_t(\xi^*)}{\partial \theta} s \right) ds.
\]
By Lemma A.1(i), (iii), (iv) and a similar argument as for part (ii), it is easy to see that \( K_{5n}(u_n) = o_p(1) \) and \( K_{6n}(u_n) = o_p(1) \). Thus, it follows that (v) holds. This completes all of the proofs. \( \Box \)
7. Concluding remarks. In this paper, we first propose a self-weighted QMELE for the ARMA–GARCH model. The strong consistency and asymptotic normality of the global self-weighted QMELE are established under a fractional moment condition of $\varepsilon_t$ with $E\eta_t^2 < \infty$. Based on this estimator, the local QMELE is showed to be asymptotically normal for the ARMA–GARCH (finite variance) and –IGARCH models. The empirical study shows that the self-weighted/local QMELE has a better performance than the self-weighted/local QMLE when $\eta_t$ has a heavy-tailed distribution, while the local QMELE is more efficient than the self-weighted QMELE for the cases with a finite variance and –IGARCH errors. We also give a real example to illustrate that our new estimation procedure is necessary. According to our limit experience, the estimated tail index of most of data sets lies in $[2, 4)$ in economics and finance. Thus, the local QMELE may be the most suitable in practice if there is a further evidence to show that $E\eta_t^4 = \infty$.

APPENDIX

The Lemma A.1 below is from Ling (2007).

**Lemma A.1.** Let $\xi_{\rho t}$ be defined as in Assumption 2.4. If Assumptions 2.1 and 2.2 hold, then there exists a constant $\rho \in (0, 1)$ and a neighborhood $\Theta_0$ of $\theta_0$ such that:

\[
\sup_{\Theta} |\varepsilon_{t-1}(\gamma)| \leq C\xi_{\rho t-1},
\]

(i) \[
\sup_{\Theta} \left\| \frac{\partial \varepsilon_t(\gamma)}{\partial \gamma} \right\| \leq C\xi_{\rho t-1} \quad \text{and}
\]

\[
\sup_{\Theta} \left\| \frac{\partial^2 \varepsilon_t(\gamma)}{\partial \gamma \partial \gamma'} \right\| \leq C\xi_{\rho t-1},
\]

(ii) \[
\sup_{\Theta} h_t(\theta) \leq C\xi_{\rho t-1}^2,
\]

(iii) \[
\sup_{\Theta_0} \left\| \frac{1}{h_t(\theta)} \frac{\partial h_t(\theta)}{\partial \delta} \right\| \leq C\xi_{\rho t-1}^{t_1} \quad \text{for any } t_1 \in (0, 1),
\]

(iv) \[
\sup_{\Theta} \left\| \frac{1}{\sqrt{h_t(\theta)}} \frac{\partial h_t(\theta)}{\partial \gamma} \right\| \leq C\xi_{\rho t-1}.
\]

**Lemma A.2.** For any $\theta^* \in \Theta$, let $B_\eta(\theta^*) = \{\theta \in \Theta : \|\theta - \theta^*\| < \eta\}$ be an open neighborhood of $\theta^*$ with radius $\eta > 0$. If Assumptions 2.1–2.5 hold, then:

(i) $E\left[ \sup_{\theta \in \Theta} w_t l_t(\theta) \right] < \infty$,

(ii) $E[w_t l_t(\theta)]$ has a unique minimum at $\theta_0$.

(iii) $E\left[ \sup_{\theta \in B_\eta(\theta^*)} w_t l_t(\theta) - l_t(\theta^*) \right] \rightarrow 0$ as $\eta \rightarrow 0$. 

PROOF. First, by (A.13) and (A.14) in Ling (2007) and Assumptions 2.4 and 2.5, it follows that
\[
E\left[ \sup_{\theta \in \Theta} w_t |\varepsilon_t(\gamma)\right] \leq CE\left[ w_t \xi_{\rho t-1}(1 + |\eta_t|)\right] < \infty
\]
for some \( \rho \in (0, 1) \), and
\[
E\left[ \sup_{\theta \in \Theta} w_t \log \sqrt{h_t(\theta)} \right] < \infty;
\]
see Ling (2007), page 864. Thus, (i) holds.

Next, by a direct calculation, we have
\[
E[w_t l_t(\theta)] = E\left[ w_t \log \sqrt{h_t(\theta)} + \frac{w_t |\varepsilon_t(\gamma_0) + (\gamma - \gamma_0)(\partial \varepsilon_t(\xi^*)/\partial \theta)|}{\sqrt{h_t(\theta)}} \right]
\]
\[
= E\left[ w_t \log \sqrt{h_t(\theta)} + \frac{w_t}{\sqrt{h_t(\theta)}} E\left[ |\varepsilon_t(\gamma_0) + (\gamma - \gamma_0)(\partial \varepsilon_t(\xi^*)/\partial \theta)| \right] \right]
\]
\[
\geq E\left[ w_t \log \sqrt{h_t(\theta)} + \frac{w_t}{\sqrt{h_t(\theta)}} E(|\varepsilon_t|) \right]
\]
\[
= E\left[ w_t \left( \log \frac{h_t(\theta_0)}{h_t(\theta)} + \sqrt{h_t(\theta_0)/h_t(\theta)} \right) \right] + E[w_t \log \sqrt{h_t(\theta)}],
\]
where the last inequality holds since \( \eta_t \) has a unique median 0, and obtains the minimum if and only if \( \gamma = \gamma_0 \) a.s.; see Ling (2007). Here, \( \xi^* \) lies between \( \gamma \) and \( \gamma_0 \). Considering the function \( f(x) = \log x + a/x \) when \( a \geq 0 \), it reaches the minimum at \( x = a \). Thus, \( E[w_t l_t(\theta)] \) reaches the minimum if and only if \( \sqrt{h_t(\theta)} = \sqrt{h_t(\theta_0)} \) a.s., and hence \( \theta = \theta_0 \); see Ling (2007). Thus, we can claim that \( E[w_t l_t(\theta)] \) is uniformly minimized at \( \theta_0 \), that is, (ii) holds.

Third, let \( \theta^* = (\gamma^*, \delta^*)' \in \Theta \). For any \( \theta \in B_\eta(\theta^*) \), using Taylor’s expansion, we can see that
\[
\log \sqrt{h_t(\theta)} - \log \sqrt{h_t(\theta^*)} = \frac{(\theta - \theta^*)' \partial h_t(\theta^{**})}{2 h_t(\theta^{**})},
\]
where \( \theta^{**} \) lies between \( \theta \) and \( \theta^* \). By Lemma A.1(iii)–(iv) and Assumption 2.4, for some \( \rho \in (0, 1) \), we have
\[
E\left[ \sup_{\theta \in B_\eta(\theta^*)} w_t |\log \sqrt{h_t(\theta)} - \log \sqrt{h_t(\theta^*)}| \right] \leq C\eta E(w_t \xi_{\rho t-1}) \to 0
\]
as \( \eta \to 0 \). Similarly,
\[
E\left[ \sup_{\theta \in B_\eta(\theta^*)} \frac{w_t}{\sqrt{h_t(\theta)}} |\varepsilon_t(\gamma) - |\varepsilon_t(\gamma^*)|| \right] \to 0 \quad \text{as } \eta \to 0,
\]
\[
E\left[ \sup_{\theta \in B_\eta(\theta^*)} w_t |\varepsilon_t(\gamma^*)| \left( \frac{1}{\sqrt{h_t(\theta)}} - \frac{1}{\sqrt{h_t(\theta^*)}} \right) \right] \to 0 \quad \text{as } \eta \to 0.
\]
Then, it follows that (iii) holds. This completes all of the proofs of Lemma A.2. □

PROOF OF THEOREM 2.1. We use the method in Huber (1967). Let $V$ be any open neighborhood of $\theta_0 \in \Theta$. By Lemma A.2(iii), for any $\theta^* \in V^c = \Theta / V$ and $\varepsilon > 0$, there exists an $\eta_0 > 0$ such that

$$E \left[ \inf_{\theta \in B_{\eta_0}(\theta^*)} w_i l_t(\theta) \right] \geq E[w_i l_t(\theta^*)] - \varepsilon. \quad (A.1)$$

From Lemma A.2(i), by the ergodic theorem, it follows that

$$\frac{1}{n} \sum_{t=1}^{n} \inf_{\theta \in B_{\eta_0}(\theta^*)} w_i l_t(\theta) \geq E \left[ \inf_{\theta \in B_{\eta_0}(\theta^*)} w_i l_t(\theta) \right] - \varepsilon \quad (A.2)$$

as $n$ is large enough. Since $V^c$ is compact, we can choose $\{B_{\eta_0}(\theta_i) : \theta_i \in V^c, i = 1, 2, \ldots, k\}$ to be a finite covering of $V^c$. Thus, from (A.1) and (A.2), we have

$$\inf_{\theta \in V^c} L_{sn}(\theta) = \min_{1 \leq i \leq k} \inf_{\theta \in B_{\eta_0}(\theta_i)} L_{sn}(\theta)$$

$$\geq \min_{1 \leq i \leq k} \frac{1}{n} \sum_{t=1}^{n} \inf_{\theta \in B_{\eta_0}(\theta_i)} w_i l_t(\theta)$$

$$\geq \min_{1 \leq i \leq k} E \left[ \inf_{\theta \in B_{\eta_0}(\theta_i)} w_i l_t(\theta) \right] - \varepsilon \quad (A.3)$$

as $n$ is large enough. Note that the infimum on the compact set $V^c$ is attained. For each $\theta_i \in V^c$, from Lemma A.2(ii), there exists an $\varepsilon_0 > 0$ such that

$$E \left[ \inf_{\theta \in B_{\eta_0}(\theta_i)} w_i l_t(\theta) \right] \geq E[w_i l_t(\theta_0)] + 3\varepsilon_0. \quad (A.4)$$

Thus, from (A.3) and (A.4), taking $\varepsilon = \varepsilon_0$, it follows that

$$\inf_{\theta \in V^c} L_{sn}(\theta) \geq E[w_i l_t(\theta_0)] + 2\varepsilon_0. \quad (A.5)$$

On the other hand, by the ergodic theorem, it follows that

$$\inf_{\theta \in V} L_{sn}(\theta) \leq L_{sn}(\theta_0) = \frac{1}{n} \sum_{t=1}^{n} w_i l_t(\theta_0) \leq E[w_i l_t(\theta_0)] + \varepsilon_0. \quad (A.6)$$

Hence, combining (A.5) and (A.6), it gives us

$$\inf_{\theta \in V^c} L_{sn}(\theta) \geq E[w_i l_t(\theta_0)] + 2\varepsilon_0 > E[w_i l_t(\theta_0)] + \varepsilon_0 \geq \inf_{\theta \in V} L_{sn}(\theta),$$

which implies that

$$\hat{\theta}_{sn} \in V \quad \text{a.s. for } \forall V, \text{ as } n \text{ is large enough.}$$

By the arbitrariness of $V$, it yields $\hat{\theta}_{sn} \to \theta_0$ a.s. This completes the proof. □
Acknowledgments. The authors greatly appreciate the very helpful comments of two anonymous referees, the Associate Editor and the Editor T. Tony, Cai.

REFERENCES

BAŞRAK, B., DAVIS, R. A. and MIKOSCH, T. (2002). Regular variation of GARCH processes. Stochastic Process. Appl. 99 95–115. MR1894253

BERA, A. K. and HIGGINS, M. L. (1993). ARCH models: Properties, estimation and testing. Journal of Economic Surveys 7 305–366; reprinted in Surveys in Econometrics (L. Oxley et al., eds.) 215–272. Blackwell, Oxford 1995.

BERKES, I., HORVÁTH, L. and KOKOSZKA, P. (2003). GARCH processes: Structure and estimation. Bernoulli 9 201–227. MR1997027

BERKES, I. and HORVÁTH, L. (2004). The efficiency of the estimators of the parameters in GARCH processes. Ann. Statist. 32 633–655. MR2060172

BOLLERSLEV, T. (1986). Generalized autoregressive conditional heteroskedasticity. J. Econometrics 31 307–327. MR0853051

BOLLERSLEV, T., CHOU, R. Y. and KRONER, K. F. (1992). ARCH modeling in finance: A review of the theory and empirical evidence. J. Econometrics 52 5–59.

BOLLERSLEV, T., ENGEL, R. F. and NELSON, D. B. (1994). ARCH models. In Handbook of Econometrics, 4 (R. F. Engle and D. L. McFadden, eds.) 2961–3038. North-Holland, Amsterdam.

BOUGEROL, P. and PICARD, N. (1992). Stationarity of GARCH processes and of some nonnegative time series. J. Econometrics 10 285–317. MR1956860

CHAN, N. H. and PENG, L. (2005). Weighted least absolute deviations estimation for an AR(1) process with ARCH(1) errors. Biometrika 92 477–484. MR2201372

DAVIS, R. A. and DUNSMUIR, W. T. M. (1997). Least absolute deviation estimation for regression with ARMA errors. J. Theoret. Probab. 10 481–497. MR1455154

DAVIS, R. A., KNIGHT, K. and LIU, J. (1992). M-estimation for autoregressions with infinite variance. Stochastic Process. Appl. 40 145–180. MR1145464

ENGLE, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. Econometrica 50 987–1007. MR0666121

FRANÇQ, C. and ZAKOIAN, J.-M. (2004). Maximum likelihood estimation of pure GARCH and ARMA–GARCH processes. Bernoulli 10 605–637. MR2076065

FRANÇQ, C. and ZAKOIAN, J. M. (2010). GARCH Models: Structure, Statistical Inference and Financial Applications. Wiley, Chichester, UK.

HALL, P. and YAO, Q. (2003). Inference in ARCH and GARCH models with heavy-tailed errors. Econometrica 71 285–317. MR1956860

HORVÁTH, L. and LIESE, F. (2004). Lp-estimators in ARCH models. J. Statist. Plann. Inference 119 277–309. MR2019642

HUBER, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. In Proc. Fifth Berkeley Sympos. Math. Statist. and Probability (Berkeley, Calif., 1965/66), Vol. 1: Statistics 221–233. Univ. California Press, Berkeley, CA. MR0216620

KNIGHT, K. (1987). Rate of convergence of centred estimates of autoregressive parameters for infinite variance autoregressions. J. Time Series Anal. 8 51–60. MR0868017

KNIGHT, K. (1998). Limiting distributions for L1 regression estimators under general conditions. Ann. Statist. 26 755–770. MR1626024

LANG, W. T., RAHBEK, A. and JENSEN, S. T. (2011). Estimation and asymptotic inference in the first order AR–ARCH model. Econometric Rev. 30 129–153.

LEE, S.-W. and HANSEN, B. E. (1994). Asymptotic theory for the GARCH(1, 1) quasi-maximum likelihood estimator. Econometric Theory 10 29–52. MR1279689
LI, G. and LI, W. K. (2005). Diagnostic checking for time series models with conditional heteroscedasticity estimated by the least absolute deviation approach. *Biometrika* **92** 691–701. MR2202655

LI, G. and LI, W. K. (2008). Least absolute deviation estimation for fractionally integrated autoregressive moving average time series models with conditional heteroscedasticity. *Biometrika* **95** 399–414. MR2521590

LING, S. (2005). Self-weighted least absolute deviation estimation for infinite variance autoregressive models. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **67** 381–393. MR2155344

LING, S. (2007). Self-weighted and local quasi-maximum likelihood estimators for ARMA–GARCH/IGARCH models. *J. Econometrics* **140** 849–873. MR2408929

LING, S. and LI, W. K. (1997). On fractionally integrated autoregressive moving-average time series models with conditional heteroscedasticity. *J. Amer. Statist. Assoc.* **92** 1184–1194. MR1482150

LING, S. and McALEER, M. (2003). Asymptotic theory for a vector ARMA–GARCH model. *Econometric Theory* **19** 280–310. MR1966031

LUMSDAINE, R. L. (1996). Consistency and asymptotic normality of the quasi-maximum likelihood estimator in IGARCH(1, 1) and covariance stationary GARCH(1, 1) models. *Econometrica* **64** 575–596. MR1385558

PAN, J., WANG, H. and YAO, Q. (2007). Weighted least absolute deviations estimation for ARMA models with infinite variance. *Econometric Theory* **23** 852–879. MR2395837

PENG, L. and YAO, Q. (2003). Least absolute deviations estimation for ARCH and GARCH models. *Biometrika* **90** 967–975. MR2024770

POLLARD, D. (1985). New ways to prove central limit theorems. *Econometric Theory* **1** 295–314.

WU, R. and DAVIS, R. A. (2010). Least absolute deviation estimation for general autoregressive moving average time-series models. *J. Time Series Anal.* **31** 98–112. MR2677341

ZHU, K. (2011). On the LAD estimation and likelihood ratio test in time series. Thesis Dissertation, Hong Kong Univ. Science and Technology.

ZHU, K. and LING, S. (2011a). Quasi-maximum exponential likelihood estimators for a double AR(p) model. Manual Scripts.

ZHU, K. and LING, S. (2011b). The global weighted LAD estimators for finite/infinite variance ARMA(p, q) models. *Econometric Theory*. To appear.