

QUASI-MAXIMUM EXPONENTIAL LIKELIHOOD ESTIMATORS FOR A DOUBLE AR(p) MODEL

Ke Zhu and Shiqing Ling

*Chinese Academy of Sciences and
Hong Kong University of Science and Technology*

Abstract: The paper studies the quasi-maximum exponential likelihood estimator (QMELE) for the double AR(p) (DAR(p)) model:

$$y_t = \sum_{i=1}^p \phi_i y_{t-i} + \eta_t \sqrt{w + \sum_{i=1}^p \alpha_i y_{t-i}^2},$$

where $\{\eta_t\}$ is a white noise sequence. Under a fractional moment of y_t with $E\eta_t^2 < \infty$, strong consistency and asymptotic normality of the global QMELE are established. A formal comparison is given with the QMLE in Ling (2007) and WLAD in Chan and Peng (2005). A simulation study is carried out to compare the performance of these estimators in finite samples. An example on the exchange rate is given.

Key words and phrases: Asymptotic normality, double AR(p) model, QMELE and strong consistency.

1. Introduction

Consider the autoregressive (AR) model with conditional heteroscedasticity:

$$y_t = \sum_{i=1}^p \phi_i y_{t-i} + \eta_t \sqrt{w + \sum_{i=1}^p \alpha_i y_{t-i}^2}, \quad (1.1)$$

where $w > 0, \alpha_i > 0$ ($i = 1, \dots, p$), $\{\eta_t\}$ are independent and identically distributed random variables with $E|\eta_t| = 1$, and y_s is independent of $\{\eta_t : t \geq 1\}$ for $s \leq 0$. Let \mathcal{F}_t be the σ -field generated by $\{\eta_t, \dots, \eta_1, y_0, \dots, y_{1-p}\}$. The conditional variance of y_t is $\text{var}(y_t | \mathcal{F}_{t-1}) = E\eta_t^2 (w + \sum_{i=1}^p \alpha_i y_{t-i}^2)$ when $E\eta_t^2 < \infty$, which is changing over time. We call (1.1) the p-th order double AR(p) (DAR(p)) model. It is a special case of the ARMA-ARCH models in Weiss (1986) and an example of the weak ARMA models in Francq and Zakoian (1998, 2000). Model (1.1) reduces to Engle's (1982) ARCH(p) model when $\phi_i \equiv 0$, but they are different when $\phi_i \neq 0$. For some important results on Engle's ARCH models, we refer

to Bollerslev, Chou, and Kroner (1992), Bollerslev, Engel, Nelson (1994), Berkes, Horváth, and Kokoszka (2003), Hall and Yao (2003), Francq and Zakoian (2004, 2010), Jensen and Rahbek (2004), and Lang, Rahbek, and Jensen (2011).

The quasi-maximum likelihood estimator (QMLE) was studied by Ling (2004) and Ling and Li (2008) for the DAR(1) models, and by Ling (2007) for the DAR(p) models. That $E\eta_t^4 < \infty$ is necessary for its asymptotic normality; in practice, this may fail; and the standard QMLE procedure may not be reliable. The least absolute deviation (LAD) approach can be used to reduce the moment condition of η_t and provide a robust estimator; see Knight (1987, 1998), Davis and Dunsmuir (1997), Ling (2005), Pan, Wang, and Yao (2007), and Zhu and Ling (2012) for the ARMA models with i.i.d. errors, and Horváth and Liese (2004), Peng and Yao (2003), Berkes and Horváth (2004), Li and Li (2005, 2008), and Zhu and Ling (2011) for the ARMA-GARCH/GARCH models. Chan and Peng (2005) proposed a local weighted LAD estimator (WLADE) for the DAR(1) models, and established its asymptotic theory. Unlike QMLE, the WLADE only requires $E\eta_t^2 < \infty$ and shares a property of robust estimators. However, contrary to the LAD estimators for the regression or AR models, the WLADE is not an efficient estimator when η_t follows a double exponential distribution.

In this paper, we investigate the global quasi-maximum exponential likelihood estimator (QMELE) for model (1.1), which is a LAD-type estimator. If $E|y_t|^\iota < \infty$ for some $\iota > 0$, with $E\eta_t^2 < \infty$, strong consistency and asymptotic normality of the QMELE are obtained. A comparison is given with the QMLE in Ling (2007) and WLADE in Chan and Peng (2005). A simulation study is carried out to compare the performance of these estimators in finite samples. An example on the exchange rate is given to illustrate the advantage of our QMELE procedure.

This paper is organized as follows. Section 2 gives our main results. Simulation results are reported in Section 3. An example is given in Section 4. All of the proofs are in the Appendix.

2. Main Results

Let $\theta = (\gamma', \delta)'$ be the unknown parameter of model (1.1) with true value $\theta_0 = (\gamma_0', \delta_0)'$, where $\gamma = (\phi_1, \dots, \phi_p)'$ and $\delta = (w, \alpha_1, \dots, \alpha_p)'$. Denote the parameter space by $\Theta = \Theta_\gamma \times \Theta_\delta$, where $\Theta_\gamma \subset R^p$, $\Theta_\delta \subset R_0^{p+1}$, $R = (-\infty, \infty)$, and $R_0 = (0, \infty)$. Assume that $\{y_1, \dots, y_n\}$ are generated by model (1.1). When η_t follows the standard double exponential distribution, the log-likelihood function (ignoring a constant) can be written as

$$L_n(\theta) = \frac{1}{n} \sum_{t=p+1}^n l_t(\theta) \quad \text{and} \quad l_t(\theta) = \log \sqrt{h_t(\delta)} + \frac{|\varepsilon_t(\gamma)|}{\sqrt{h_t(\delta)}}, \quad (2.1)$$

where

$$\varepsilon_t(\gamma) = y_t - \sum_{i=1}^p \phi_i y_{t-i} \quad \text{and} \quad h_t(\delta) = w + \sum_{i=1}^p \alpha_i y_{t-i}^2.$$

Let $\hat{\theta}_n = \arg \min_{\Theta} L_n(\theta)$. Since we do not assume that η_t follows the standard double exponential distribution, $\hat{\theta}_n$ is called the global quasi-maximum exponential likelihood estimator (QMELE) of θ_0 .

Assumption 1. Θ is compact with $\underline{w} \leq w \leq \bar{w}$ and $\underline{\alpha}_i \leq \alpha_i \leq \bar{\alpha}_i$ ($i = 1, \dots, p$), where $\underline{w}, \bar{w}, \underline{\alpha}_i$ and $\bar{\alpha}_i$ ($i = 1, \dots, p$) are some positive constants, and θ_0 is an interior point in Θ .

Assumption 2. $\{y_t : t = 1 - p, \dots, 0, 1, 2, \dots\}$ is strictly stationary and ergodic with $E|y_t|^\iota < \infty$ for some $\iota > 0$.

Assumption 3. η_t has zero median with $E\eta_t^2 < \infty$ and a continuous density $f(x)$ in R satisfying $f(0) > 0$ and $\sup_x f(x) < \infty$.

When $\eta_t \sim N(0, 1)$, a necessary and sufficient condition for Assumption 2 is given in Ling (2007). When $p = 1$, Borkovec and Klüppelberg (2001) obtained a strict stationarity condition for the DAR(1) models, $E(\ln |\phi + \eta_t \sqrt{\alpha}|) < 0$. This condition implies that $E|y_t|^\iota < \infty$ for some $\iota > 0$; see Ling (2005). Figure 1 shows the stationary region of (ϕ, α) for the DAR(1) models when η_t is $N(0, \pi/2)$ and $(\pi/(2\sqrt{3}))t_3$, respectively, with t_3 the student's t distribution with three degrees of freedom. From the figure, we can see that $|\phi|$ can be 1 or slightly larger than 1, and the stationary region when $\eta_t \sim (\pi/(2\sqrt{3}))t_3$ is larger than that when $\eta_t \sim N(0, \pi/2)$; this is quite different from the stationarity condition, $|\phi| < 1$, of the AR(1) models with i.i.d. errors.

Our basic results on strong convergence and asymptotic normality are as follows:

Theorem 1. Suppose that η_t has zero median with $E|\eta_t| = 1$. If Assumptions 1–2 hold, then $\hat{\theta}_n \rightarrow \theta_0$ almost surely (a.s.) as $n \rightarrow \infty$.

Theorem 2. If Assumptions 1–3 hold with $E|\eta_t| = 1$, then

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \rightarrow_d N\left(0, \frac{1}{4}\Sigma_0^{-1}\Omega_0\Sigma_0^{-1}\right) \text{ as } n \rightarrow \infty,$$

where \rightarrow_d denotes the convergence in distribution and

$$\Sigma_0 = \text{diag}\left\{f(0)E(Y_{1t}Y'_{1t}), \frac{1}{8}E(Y_{2t}Y'_{2t})\right\},$$

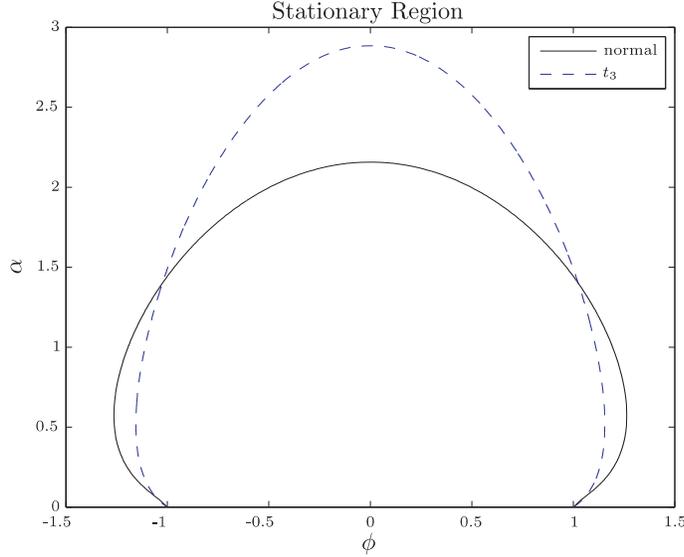


Figure 1. The stationary regions bounded by solid line and dashed line are $\{(\phi, \alpha) : E \ln |\phi + \eta_t \sqrt{\alpha}| < 0\}$ for $\eta_t \sim N(0, \pi/2)$ and $(\pi/(2\sqrt{3}))t_3$, respectively.

$$\Omega_0 = \begin{pmatrix} E(Y_{1t}Y'_{1t}) & \frac{E\eta_t}{2} E(Y_{1t}Y'_{2t}) \\ \frac{E\eta_t}{2} E(Y_{2t}Y'_{1t}) & \frac{E\eta_t^2 - 1}{4} E(Y_{2t}Y'_{2t}) \end{pmatrix},$$

with

$$Y_{1t} = \frac{1}{\sqrt{h_t(\delta_0)}}(y_{t-1}, \dots, y_{t-p})' \text{ and } Y_{2t} = \frac{1}{h_t(\delta_0)}(1, y_{t-1}^2, \dots, y_{t-p}^2)'$$

Remark 1. When $E\eta_t = 0$, the asymptotic variance in Theorem 2 reduces to the block diagonal matrix:

$$\Gamma_0 = \text{diag} \left\{ 4f(0)^2 E(Y_{1t}Y'_{1t}), \frac{1}{4(E\eta_t^2 - 1)} E(Y_{2t}Y'_{2t}) \right\}^{-1}.$$

The asymptotic normality of our QMELE only needs a fractional moment of y_t . It is well known that the asymptotic normality of the classical LAD estimator requires $Ey_t^2 < \infty$ for the AR models with i.i.d. errors or GARCH errors; see Knight (1987), Davis, Knight, and Liu (1992), Davis (1996), Davis and Dunsmuir (1997), and Li and Li (2008). Recently, the weighted LAD estimator was investigated for the AR models with i.i.d. errors or GARCH errors and shown to be asymptotically normal under a fractional moment of y_t ; see Ling (2005), Pan, Wang, and Yao (2007), and Zhu and Ling (2011, 2012). However, the weighted

LAD estimator may not be efficient in general. Since the conditional variance $h_t(\delta)$ can control the log-likelihood function at (2.1), the weight is not needed for the DAR(p) models. This advantage motivates us to consider the QMELE procedure in practice, especially when y_t is heavy-tailed with $Ey_t^2 = \infty$.

Given the data set $\{y_t\}$, the matrixes Σ_0 and Ω_0 in Theorem 2 can be estimated via

$$\hat{\Sigma}_{0n} = \text{diag} \left\{ \hat{f}(0)\hat{\Delta}_{1n}, \frac{1}{8}\hat{\Delta}_{2n} \right\} \text{ and } \hat{\Omega}_{0n} = \begin{pmatrix} \hat{\Delta}_{1n} & \frac{1}{2}\hat{\Upsilon}_{1n}\hat{\Delta}_{3n} \\ \frac{1}{2}\hat{\Upsilon}_{1n}\hat{\Delta}'_{3n} & \frac{1}{4}(\hat{\Upsilon}_{2n} - 1)\hat{\Delta}_{2n} \end{pmatrix}, \tag{2.2}$$

respectively, where $\hat{\eta}_t = \varepsilon_t(\hat{\gamma}_n)/\sqrt{h_t(\hat{\delta}_n)}$ is the residual, and

$$\begin{aligned} \hat{f}(0) &= \frac{1}{nb_n} \sum_{t=p+1}^n K\left(\frac{\hat{\eta}_t}{b_n}\right), \quad \hat{\Upsilon}_{1n} = \frac{1}{n} \sum_{t=p+1}^n \hat{\eta}_t, \quad \hat{\Upsilon}_{2n} = \frac{1}{n} \sum_{t=p+1}^n \hat{\eta}_t^2, \\ \hat{\Delta}_{1n} &= \frac{1}{n} \sum_{t=p+1}^n \frac{1}{h_t(\hat{\delta}_n)} (y_{t-1}, \dots, y_{t-p})' (y_{t-1}, \dots, y_{t-p}), \\ \hat{\Delta}_{2n} &= \frac{1}{n} \sum_{t=p+1}^n \frac{1}{h_t^2(\hat{\delta}_n)} (1, y_{t-1}^2, \dots, y_{t-p}^2)' (1, y_{t-1}^2, \dots, y_{t-p}^2), \\ \hat{\Delta}_{3n} &= \frac{1}{n} \sum_{t=p+1}^n \frac{1}{h_t^{3/2}(\hat{\delta}_n)} (y_{t-1}, \dots, y_{t-p})' (1, y_{t-1}^2, \dots, y_{t-p}^2). \end{aligned}$$

Here $K(x)$, with $\int_{-\infty}^{\infty} K(x)dx = 1$ and $\int_{-\infty}^{\infty} |x|K(x)dx < \infty$, is a kernel function and $b_n > 0$ is the bandwidth.

Corollary 1. *Suppose that the conditions in Theorem 2 hold and $\sup_x |f'(x)| < \infty$. If there exists a positive number $L > 0$ such that $|K(x) - K(y)| \leq L|x - y|$ for any x, y , and $b_n \rightarrow 0, nb_n^4 \rightarrow \infty$ as $n \rightarrow \infty$, then*

$$\hat{\Sigma}_{0n} \rightarrow \Sigma_0 \quad \text{and} \quad \hat{\Omega}_{0n} \rightarrow \Omega_0, \quad \text{as } n \rightarrow \infty.$$

Remark 2. The kernel function $K(x)$ in Corollary 1 could be normal, Epanechnikov, triangular, or one of many others. We use the normal kernel function and the bandwidth $b_{opt,n}$ for the numerical studies in Sections 3-4, where

$$b_{opt,n} = b_n^* \left(1 + \frac{35}{48}\hat{\gamma}_4 + \frac{35}{32}\hat{\gamma}_3^2 + \frac{385}{1024}\hat{\gamma}_4^2 \right)^{-1/5},$$

$b_n^* = 1.06sn^{-1/5}$ is the reference bandwidth selector, and $s, \hat{\gamma}_3$, and $\hat{\gamma}_4$ are the sample standard deviation, skewness, and kurtosis of the residuals $\{\hat{\eta}_t\}$, respectively. See Fan and Yao (2003, p.201)

To compare the asymptotic efficiency of our QMELE and the QMLE in Ling (2007), we assume that $E\eta_t^3 = 0$ and $E\eta_t^4 < \infty$, and reparametrize (1.1) as

$$y_t = \sum_{i=1}^p \phi_i y_{t-i} + \eta_t^* \sqrt{w^* + \sum_{i=1}^p \alpha_i^* y_{t-i}^2}, \tag{2.3}$$

where $\eta_t^* = \eta_t / \sqrt{E\eta_t^2}$ and $\delta^* = (E\eta_t^2)\delta$. Let $\bar{\theta}_n^* = (\bar{\gamma}'_n, \bar{\delta}'_n)^T$ be the QMLE of the true parameter $\theta_0^* = (\gamma'_0, \delta'_0)^T$ in model (2.3), Then, $\bar{\theta}_n = (\bar{\gamma}'_n, \bar{\delta}'_n)^T$ is the QMLE of $\theta_0 = (\gamma'_0, \delta'_0)^T$, where $\bar{\delta}_n = \bar{\delta}_n^* / E\eta_t^2$. Thus, by Theorem 3.1 in Ling (2007), the asymptotic covariance of $\bar{\theta}_n$ is Γ_1 , where

$$\Gamma_1 = \begin{pmatrix} E\eta_t^2 J_1^{-1} & 0 \\ 0 & \kappa J_2^{-1} \end{pmatrix},$$

with $J_1 = E(Y_{1t}Y'_{1t})$, $J_2 = E(Y_{2t}Y'_{2t})$, and $\kappa = E\eta_t^4 / (E\eta_t^2)^2 - 1$.

We now compare Γ_0 with Γ_1 for some specific cases. First, consider the case when $\eta_t = \tilde{\eta}_t / E|\tilde{\eta}_t|$, with $\tilde{\eta}_t$ having the mixing normal density

$$f(x) = (1 - \varepsilon)\phi(x) + \frac{\varepsilon}{\tau}\phi\left(\frac{x}{\tau}\right),$$

for $0 \leq \varepsilon \leq 1$ and $\tau > 0$. Here, $\phi(x)$ is the pdf of standard normal. After a simple calculation, we can show that, when $\varepsilon = 1$, $\tau = \sqrt{\pi/2}$,

$$\Gamma_0 = \begin{pmatrix} \frac{\pi^2}{4} J_1^{-1} & 0 \\ 0 & (2\pi - 4)J_2^{-1} \end{pmatrix} \quad \text{and} \quad \Gamma_1 = \begin{pmatrix} \frac{\pi}{2} J_1^{-1} & 0 \\ 0 & 2J_2^{-1} \end{pmatrix}.$$

Thus, $\Gamma_0 > \Gamma_1$, and the QMLE is asymptotically more efficient than the QMELE. When $\varepsilon = 1/2, \tau = 5/2$, we have

$$\Gamma_0 = \begin{pmatrix} 1.64J_1^{-1} & 0 \\ 0 & 3.44J_2^{-1} \end{pmatrix} \quad \text{and} \quad \Gamma_1 = \begin{pmatrix} 1.86J_1^{-1} & 0 \\ 0 & 3.60J_2^{-1} \end{pmatrix}.$$

Here, $\Gamma_1 > \Gamma_0$, and hence the QMELE is asymptotically more efficient than the QMLE.

We next consider the case when $\eta_t \sim Laplace(0, 1)$. In this case,

$$\Gamma_0 = \begin{pmatrix} J_1^{-1} & 0 \\ 0 & 4J_2^{-1} \end{pmatrix} \quad \text{and} \quad \Gamma_1 = \begin{pmatrix} 2J_1^{-1} & 0 \\ 0 & 5J_2^{-1} \end{pmatrix}.$$

Thus, $\Gamma_1 > \Gamma_0$, and the QMELE is asymptotically more efficient than the QMLE. It is not surprising because QMELE is the MLE when $\eta_t \sim Laplace(0, 1)$.

Finally, we compare the asymptotic efficiency of the WLADE in Chan and Peng (2005) and the QMELE for the DAR(1) models. To make it simple, we only

consider $\theta_0 = (0, 1, 1)'$, in which case the asymptotic covariance of the WLAD E is

$$\begin{pmatrix} \frac{E\eta_t^2}{4mf^2(\sqrt{m})} \left(E\frac{y_{t-1}^2}{S^2} + E\frac{y_{t-1}^4}{S^2} \right) J_1^{-2} & 0 \\ 0 & \frac{1}{f^2(\sqrt{m})m} J_2^{-1} \end{pmatrix},$$

where $m = \text{median}\{\eta_t^2\}$ and $S = 1 + y_{t-1}^2$. When $\eta_t \sim \text{Laplace}(0, 1)$, $m = 0.48$ and $f(\sqrt{m}) = 0.25$. We can see that the QMELE is asymptotically more efficient than the WLAD E for parameters (w, α) . For the parameter ϕ , it is hard to compare them with each other in theory, since the asymptotic covariance of WLAD E is quite complicated; simulation comparison is given in the next section.

3. Simulation

In this section, we compare the performance of the QMELE with those of the QMLE and the WLAD E in finite samples. The DAR(1) model used to generate data samples was

$$y_t = \phi y_{t-1} + \eta_t \sqrt{w + \alpha y_{t-1}^2}. \tag{3.1}$$

We set the sample size $n = 400$ and use 1000 replications. The true parameters were $(\phi_0, w_0, \alpha_0) = (1.0, 1.0, 0.5), (0.5, 1.0, 0.5), (0.0, 1.0, 1.0)$, and $(0.0, 1.0, 0.5)$. We took η_t as *Laplace*(0, 1), *N*(0, 1), and t_3 distribution, respectively. Since these three estimation methods require different conditions for model (3.1), the QMELE ($\hat{\theta}_n^*$), QMLE ($\bar{\theta}_n^*$), and WLAD E ($\tilde{\theta}_n^*$) are estimators of $(\phi_0, rw_0, r\alpha_0)$ with $r = (E|\eta_t|)^2, E\eta_t^2$, and $\text{median}\{\eta_t^2\}$, respectively. In order to make our comparison feasible, we let

$$\hat{\theta}_n = \left(\hat{\phi}_n^*, \frac{\hat{w}_n^*}{(E|\eta_t|)^2}, \frac{\hat{\alpha}_n^*}{(E|\eta_t|)^2} \right), \quad \bar{\theta}_n = \left(\bar{\phi}_n^*, \frac{\bar{w}_n^*}{E\eta_t^2}, \frac{\bar{\alpha}_n^*}{E\eta_t^2} \right),$$

and

$$\tilde{\theta}_n = \left(\tilde{\phi}_n^*, \frac{\tilde{w}_n^*}{\text{median}\{\eta_t^2\}}, \frac{\tilde{\alpha}_n^*}{\text{median}\{\eta_t^2\}} \right)$$

be the QMELE, QMLE, and WLAD E of (ϕ_0, w_0, α_0) , respectively. The estimated asymptotic standard deviations of $\hat{\theta}_n, \bar{\theta}_n$, and $\tilde{\theta}_n$ were derived in a similar way. In all calculations, we chose the kernel function and the bandwidth as in Remark 2, and used the true values of $E|\eta_t|, E\eta_t^2$, and $\text{median}\{\eta_t^2\}$.

Tables 1–3 list the sample biases, the sample standard deviations (SD), and the average estimated asymptotic standard deviations (AD) of $\hat{\theta}_n, \bar{\theta}_n$, and $\tilde{\theta}_n$. From them, we can see that all three estimators have very small biases. When $\eta_t \sim \text{Laplace}(0, 1)$, the QMELE has smaller AD and SD than those of both the QMLE and the WLAD E, while the QMLE is better than the WLAD E. When

Table 1. $\eta_t \sim \text{Laplace}(0,1)$.

ϕ_0	w_0	α_0		QMELE($\hat{\theta}_n$)			QMLE($\bar{\theta}_n$)			WLADE($\tilde{\theta}_n$)		
				$\hat{\phi}_n$	\hat{w}_n	$\hat{\alpha}_n$	$\bar{\phi}_n$	\bar{w}_n	$\bar{\alpha}_n$	$\tilde{\phi}_n$	\tilde{w}_n	$\tilde{\alpha}_n$
1.0	1.0	0.5	Bias	-0.0040	0.0518	-0.0034	-0.0045	0.0391	-0.0061	-0.0043	0.1082	0.0075
			SD	0.0437	0.3051	0.0622	0.0576	0.3370	0.0678	0.0818	0.4376	0.0928
			AD	0.0475	0.3038	0.0621	0.0569	0.3254	0.0671	0.0801	0.4571	0.0894
0.5	1.0	0.5	Bias	-0.0040	0.0195	-0.0098	-0.0033	0.0099	-0.0102	-0.0029	0.0337	0.0132
			SD	0.0547	0.1701	0.0941	0.0729	0.1857	0.1040	0.1063	0.2545	0.1406
			AD	0.0606	0.1734	0.0930	0.0729	0.1873	0.1013	0.1046	0.2524	0.1366
0.0	1.0	1.0	Bias	-0.0042	0.0253	-0.0151	-0.0029	0.0169	-0.0129	0.0005	0.0457	0.0172
			SD	0.0689	0.1984	0.1529	0.0924	0.2125	0.1700	0.1348	0.2953	0.2261
			AD	0.0765	0.1968	0.1542	0.0926	0.2134	0.1692	0.1311	0.2855	0.2227
0.0	1.0	0.5	Bias	0.0004	0.0124	-0.0083	-0.0011	0.0044	-0.0099	-0.0016	0.0303	0.0231
			SD	0.0591	0.1590	0.1052	0.0816	0.1764	0.1181	0.1199	0.2365	0.1598
			AD	0.0660	0.1579	0.1048	0.0793	0.1710	0.1145	0.1145	0.2310	0.1566

$\eta_t \sim N(0, 1)$, in Table 2 we can see that the QMLE has smaller AD and SD than those of both the QMELE and the WLADE, and the SD and AD of the WLADE of ϕ_0 is slightly smaller than those of the QMELE, while the SD and AD of the WLADE of ω_0 and α_0 are larger than those of the QMELE. From Table 3, we can see that the QMELE has the smallest SD and AD, and the SD and AD of the QMLE of ϕ_0 are smaller than those of the WLADE, while the SD and AD of the WLADE of ω_0 and α_0 are smaller than those of the QMLE. We also note that the SD and AD of the QMLE of ω_0 and α_0 are not close to each other since the asymptotic variance of the QMLE is infinite in this case. These results are consistent with our theory in Section 2. The simulation results indicate that the QMELE has a good performance in finite samples.

4. A Data Example

In this section, we consider the daily exchange rate of United States Dollars (USD) to New Taiwan Dollars (TWD) (Interbank rate) from January 1, 2010 to January 1, 2011, which has in total 366 observations; see Figure 2 (a). Here, 100 times log return (after mean adjustment), denoted by $\{y_t\}_{t=1}^{365}$, is plotted in Figure 2 (b). To begin with, we first estimate the tail index of $\{y_t\}$ by Hill's estimator $\hat{H}_y(k)$ with the largest k data of $\{y_t^2\}$,

$$\hat{H}_y(k) = \frac{k}{\sum_{j=1}^k (\log \tilde{y}_{365-j} - \log \tilde{y}_{365-k})},$$

where \tilde{y}_j is the j -th order statistic of y_t^2 . The plot of $\{\hat{H}_y(k)\}_{k=10}^{180}$ is given in Figure 3 (a), from which we can see that the tail of y_t^2 is most likely less than

Table 2. $\eta_t \sim N(0,1)$.

ϕ_0	w_0	α_0	QMELE($\hat{\theta}_n$)			QMLE($\bar{\theta}_n$)			WLADE($\tilde{\theta}_n$)			
			$\hat{\phi}_n$	\hat{w}_n	$\hat{\alpha}_n$	$\bar{\phi}_n$	\bar{w}_n	$\bar{\alpha}_n$	$\tilde{\phi}_n$	\tilde{w}_n	$\tilde{\alpha}_n$	
1.0	1.0	0.5	Bias	-0.0029	0.0094	-0.0030	-0.0009	0.0055	-0.0033	0.0003	0.0183	0.0040
			SD	0.0574	0.1568	0.0583	0.0448	0.1451	0.0533	0.0530	0.2470	0.0932
			AD	0.0599	0.1553	0.0575	0.0455	0.1435	0.0532	0.0554	0.2522	0.0930
0.5	1.0	0.5	Bias	-0.0009	0.0115	-0.0124	-0.0036	0.0067	-0.0109	-0.0044	0.0185	0.0002
			SD	0.0762	0.1258	0.0940	0.0605	0.1175	0.0883	0.0723	0.1950	0.1458
			AD	0.0772	0.1228	0.0917	0.0586	0.1134	0.0852	0.0718	0.1993	0.1498
0.0	1.0	1.0	Bias	0.0015	0.0187	-0.0260	0.0027	0.0142	-0.0227	0.0034	0.0224	-0.0131
			SD	0.0893	0.1457	0.1383	0.0693	0.1351	0.1298	0.0811	0.2249	0.2105
			AD	0.0921	0.1427	0.1372	0.0701	0.1318	0.1276	0.0844	0.2278	0.2188
0.0	1.0	0.5	Bias	-0.0026	0.0101	-0.0101	-0.0002	0.0088	-0.0087	0.0014	0.0069	0.0084
			SD	0.0769	0.1219	0.1108	0.0605	0.1134	0.1040	0.0733	0.1892	0.1679
			AD	0.0823	0.1197	0.1051	0.0627	0.1109	0.0978	0.0773	0.1929	0.1723

Table 3. $\eta_t \sim t_3$.

ϕ_0	w_0	α_0	QMELE($\hat{\theta}_n$)			QMLE($\bar{\theta}_n$)			WLADE($\tilde{\theta}_n$)			
			$\hat{\phi}_n$	\hat{w}_n	$\hat{\alpha}_n$	$\bar{\phi}_n$	\bar{w}_n	$\bar{\alpha}_n$	$\tilde{\phi}_n$	\tilde{w}_n	$\tilde{\alpha}_n$	
1.0	1.0	0.5	Bias	-0.0021	0.0400	-0.0042	-0.0019	-0.0922	-0.0537	-0.0002	0.0760	0.0114
			SD	0.0550	0.3516	0.0751	0.0668	0.5706	0.2172	0.0860	0.3786	0.0849
			AD	0.0552	0.3510	0.0727	0.0677	0.6113	0.1759	0.0911	0.3977	0.0804
0.5	1.0	0.5	Bias	-0.0032	0.0218	-0.0175	-0.0024	-0.0989	-0.0746	-0.0023	0.0332	0.0061
			SD	0.0684	0.2155	0.1005	0.0846	0.4201	0.2108	0.1116	0.2310	0.1107
			AD	0.0691	0.2114	0.1026	0.0831	0.4434	0.2055	0.1077	0.2313	0.1149
0.0	1.0	1.0	Bias	0.0039	0.0230	-0.0128	0.0052	-0.0996	-0.1276	0.0067	0.0424	0.0241
			SD	0.0833	0.2657	0.1778	0.1031	0.4808	0.3412	0.1365	0.2895	0.1982
			AD	0.0857	0.2594	0.1688	0.1040	0.5572	0.4155	0.1317	0.2804	0.1848
0.0	1.0	0.5	Bias	0.0016	0.0223	-0.0089	-0.0010	-0.0956	-0.0542	-0.0037	0.0297	0.0153
			SD	0.0714	0.2095	0.1213	0.0924	0.4151	0.2835	0.1224	0.2144	0.1302
			AD	0.0730	0.2009	0.1178	0.0899	0.4010	0.2850	0.1168	0.2145	0.1297

1, i.e., $Ey_t^2 = \infty$. Moreover, the estimated kurtosis of $\{y_t\}$ is 13.9, indicating that the tail distribution of $\{y_t\}$ is much heavier than for the normal distribution. Thus, the classical LAD or QMLE procedure is not reliable for the ARMA models with i.i.d. errors or GARCH errors. Therefore, we used a DAR(p) ($p \leq 4$) model to fit the data set $\{y_t\}$. According to Akaike's information criterion (AIC), the fitted model is

$$y_t = -0.2358y_{t-1} - 0.1457y_{t-2} + \eta_t \sqrt{0.0437 + 0.1232y_{t-1}^2 + 0.0712y_{t-2}^2}. \quad (4.1)$$

(0.0293)
(0.0313)
(0.1361)
(0.0087)
(0.0109)

Model (4.1) is estimated using the QMELE procedure. The standard errors, reported in parentheses, are calculated via (2.2), with the kernel function and

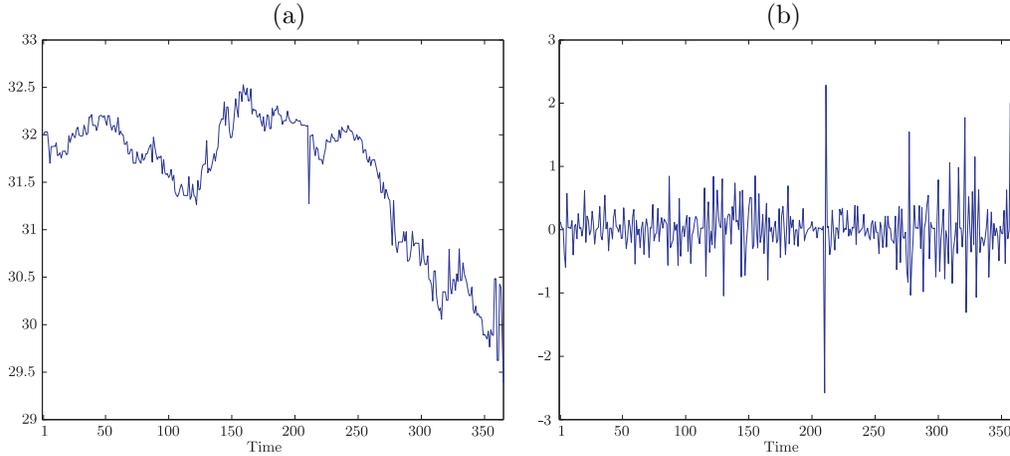


Figure 2. (a) the daily exchange rate of USD/TWD from January 1, 2010 to January 1, 2011 and (b) its 100 times log return (after mean adjustment).

bandwidth being chosen as in Remark 2. The value of the log-likelihood is -120.46, and the estimated value of $E|\eta_t|$ is 0.9998, close to 1. The first 25 autocorrelations or partial autocorrelations of the residuals $\{\hat{\eta}_t\}$ are not significant at the 5% level; see Figure 4 (a)–(b). Similar results hold for the autocorrelations or partial autocorrelations of $\{\hat{\eta}_t^2\}$; see Figure 4 (c)–(d). With bandwidth $b_n = 0.3022$, the normal kernel density estimator of η_t , based on $\{\hat{\eta}_t\}$, is plotted in Figure 3 (b). Apart from a small neighborhood of origin, this kernel density is very close to the density of the standard double exponential distribution. These results suggest that model (4.1) is adequate for the data set $\{y_t\}$.

Next, by using the QMLE procedure in Ling (2007), we get an alternative fitted model for the data set $\{y_t\}$ as

$$y_t = -0.3511y_{t-1} - 0.2583y_{t-2} + \eta_t \sqrt{0.1207 + 0.1562y_{t-1}^2 + 0.1180y_{t-2}^2}. \quad (4.2)$$

(0.0424) (0.0443) (0.1456) (0.0339) (0.0396)

Again, model (4.2) is selected by AIC, with the standard errors in parentheses. However, the log-likelihood value is -178.89 for model (4.2). From this, model (4.1) is clearly superior to model (4.2). A DAR(2) model, based on the QMELE procedure, seems be more reasonable and suitable choice to fit the data set $\{y_t\}$.

Acknowledgements

The authors greatly appreciate the helpful comments of two anonymous referees, and the editor Peter Hall. The first author's research is supported in part by NSFC (No.11201459) and the National Center for Mathematics and Inter-

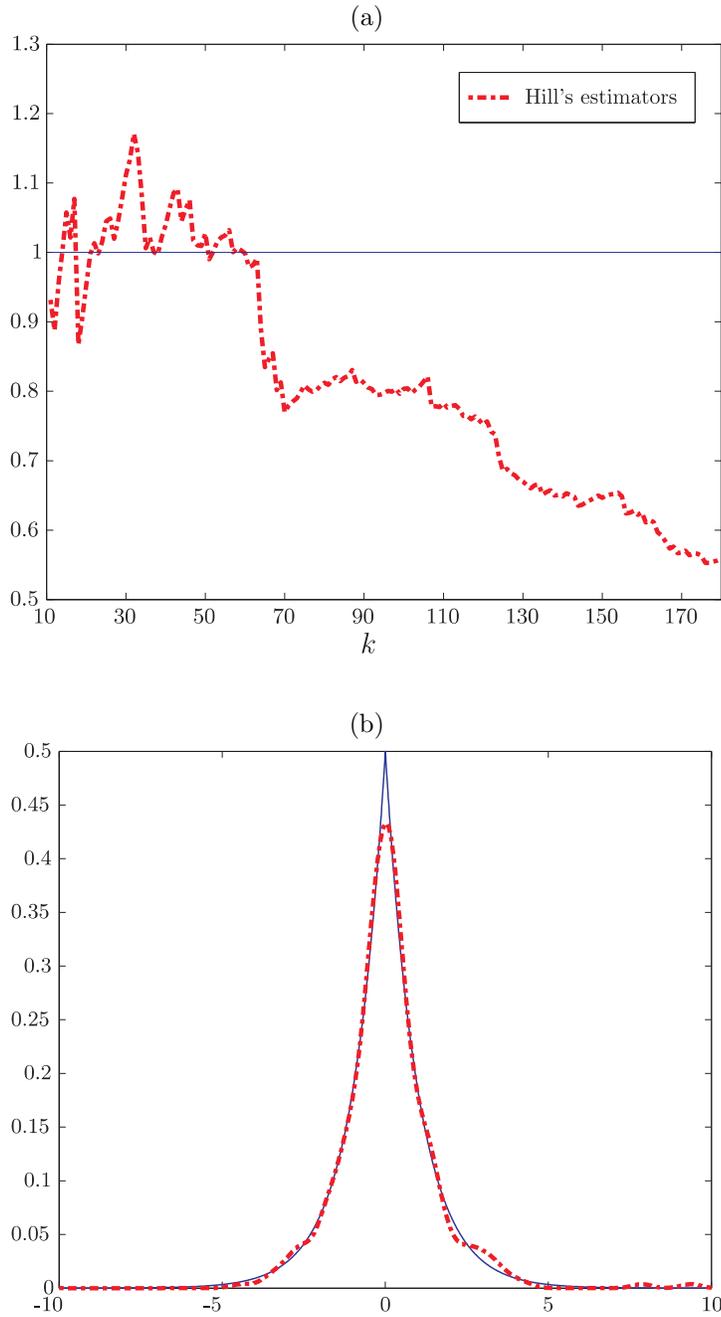


Figure 3. (a) Hill's estimators $\{\hat{H}_y(k)\}$ for y_t^2 ; (b) the kernel density of η_t (---) in model (4.1), together with the density of the standard double exponential distribution (—).

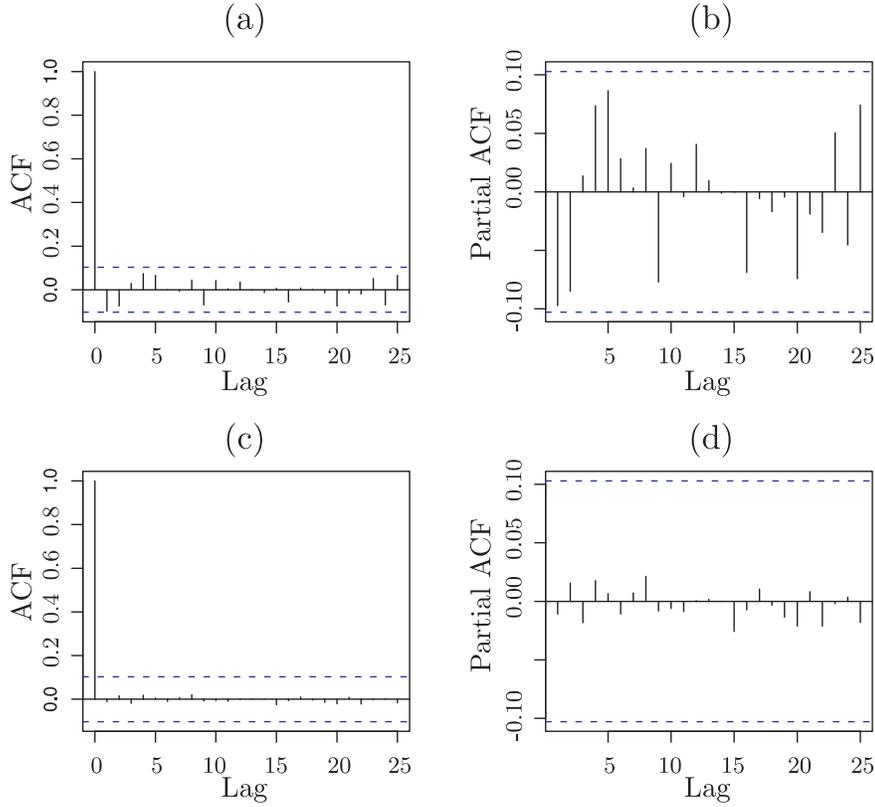


Figure 4. (a) the autocorrelations of $\{\hat{\eta}_t\}$, (b) the partial autocorrelations of $\{\hat{\eta}_t\}$, (c) the autocorrelations of $\{\hat{\eta}_t^2\}$ and (d) the partial autocorrelations of $\{\hat{\eta}_t^2\}$.

disciplinary Sciences, CAS. The second author's research is supported by Hong Kong Research Commission Grant HKUST641912.

Appendix

Lemma A.1. For any $\theta^* \in \Theta$, let $B_\eta(\theta^*) = \{\theta \in \Theta : \|\theta - \theta^*\| < \eta\}$ be an open neighborhood of θ^* with radius $\eta > 0$. If Assumptions 1–2 hold, then

- (i) $E \left[\sup_{\theta \in \Theta} l_t(\theta) \right] < \infty$,
- (ii) $E [l_t(\theta)]$ has a unique minimum at θ_0 ,
- (iii) $E \left[\sup_{\theta \in B_\eta(\theta^*)} |l_t(\theta) - l_t(\theta^*)| \right] \rightarrow 0$ as $\eta \rightarrow 0$.

Proof. First, by Assumptions 1–2 and Lemma B.2 in Ling (2007), it is straight-

forward to see that (i) holds. Next, by a direct calculation, we have

$$\begin{aligned} E[l_t(\theta)] &= E \left[\log \sqrt{h_t(\delta)} + \frac{|\varepsilon_t(\gamma_0) + \sqrt{h_t(\delta_0)}(\gamma_0 - \gamma)'Y_{1t}|}{\sqrt{h_t(\delta)}} \right] \\ &= E \left[\log \sqrt{h_t(\delta)} + \sqrt{\frac{h_t(\delta_0)}{h_t(\delta)}} E \left\{ |\eta_t(\theta_0) + (\gamma_0 - \gamma)'Y_{1t}| \middle| \mathcal{F}_{t-1} \right\} \right] \\ &\geq E \left[\log \sqrt{h_t(\delta)} + \sqrt{\frac{h_t(\delta_0)}{h_t(\delta)}} E(|\eta_t| | \mathcal{F}_{t-1}) \right] \\ &= E \left[\log \sqrt{h_t(\delta)} + \sqrt{\frac{h_t(\delta_0)}{h_t(\delta)}} \right], \end{aligned}$$

where the last inequality holds since η_t has zero median, and the minimum is attained if and only if $\gamma = \gamma_0$ a.s.; see Ling (2007). Here, ξ^* lies between γ and γ_0 . The function $f(x) = \log x + a/x$, $a \geq 0$, reaches its minimum at $x = a$. Thus, $E[l_t(\theta)]$ reaches its minimum if and only if $\sqrt{h_t(\delta)} = \sqrt{h_t(\delta_0)}$ a.s., and hence $\theta = \theta_0$. Thus, we can claim that $E[l_t(\theta)]$ is uniformly minimized at θ_0 , i.e., (ii) holds.

Let $\theta^* = (\gamma^*, \delta^*)' \in \Theta$. For any $\theta \in B_\eta(\theta^*)$, by using Taylor's expansion, we can see that

$$\log \sqrt{h_t(\delta)} - \log \sqrt{h_t(\delta^*)} = \frac{(\delta - \delta^*)'}{2h_t(\zeta^*)} (1, y_{t-1}^2, \dots, y_{t-p}^2)',$$

where ζ^* lies between δ and δ^* . Then, by Assumption 2, we have

$$E \left[\sup_{\theta \in B_\eta(\theta^*)} \left| \log \sqrt{h_t(\delta)} - \log \sqrt{h_t(\delta^*)} \right| \right] \leq \frac{\eta}{2} E \left[\frac{1 + \sum_{i=1}^p y_{t-i}^2}{\underline{w} + \sum_{i=1}^p \underline{\alpha}_i y_{t-i}^2} \right] \leq O(1)\eta \rightarrow 0$$

as $\eta \rightarrow 0$. Similarly,

$$\begin{aligned} E \left[\sup_{\theta \in B_\eta(\theta^*)} \frac{1}{\sqrt{h_t(\delta)}} \left| |\varepsilon_t(\gamma)| - |\varepsilon_t(\gamma^*)| \right| \right] &\rightarrow 0 \text{ as } \eta \rightarrow 0, \\ E \left[\sup_{\theta \in B_\eta(\theta^*)} |\varepsilon_t(\gamma^*)| \left| \frac{1}{\sqrt{h_t(\delta)}} - \frac{1}{\sqrt{h_t(\delta^*)}} \right| \right] &\rightarrow 0 \text{ as } \eta \rightarrow 0. \end{aligned}$$

Thus it follows that (iii) holds.

Proof of Theorem 1. We use the method in Huber (1967). Let V be any open neighborhood of $\theta_0 \in \Theta$. By Lemma A.1 (iii), for any $\theta^* \in V^c = \Theta/V$ and $\varepsilon > 0$,

there exists an $\eta_0 > 0$ such that

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

From Lemma A.1 (i), by the Ergodic Theorem, it follows that

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

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

$$\begin{aligned} \inf_{\theta \in V^c} L_n(\theta) &= \min_{1 \leq i \leq k} \inf_{\theta \in B_{\eta_0}(\theta_i)} L_n(\theta) \\ &\geq \min_{1 \leq i \leq k} \frac{1}{n} \sum_{t=p+1}^n \inf_{\theta \in B_{\eta_0}(\theta_i)} l_t(\theta) \\ &\geq \min_{1 \leq i \leq k} E \left[\inf_{\theta \in B_{\eta_0}(\theta_i)} l_t(\theta) \right] - \varepsilon \end{aligned} \quad (\text{A.3})$$

if 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.1 (ii), there exists an $\varepsilon_0 > 0$ such that

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

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

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

if n is large enough. On the other hand, by the Ergodic Theorem, it follows that

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

if n is large enough. Hence, combining (A.5) and (A.6),

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

which implies that $\hat{\theta}_n \in V$, a.s. for $\forall V$, if n is large enough. By the arbitrariness of V , it yields $\hat{\theta}_n \rightarrow \theta_0$ a.s. This completes the proof.

To prove Theorem 2, we use the technique in Zhu and Ling (2011). We first re-parameterize the objective function (2.1) as

$$H_n(u) = nL_n(\theta_0 + u) - nL_n(\theta_0),$$

where $u \in \Lambda \equiv \{u = (u'_1, u'_2)' : u + \theta_0 \in \Theta\}$. Let $\hat{u}_n = \hat{\theta}_n - \theta_0$. Then, \hat{u}_n is the minimizer of $H_n(u)$ in Λ . Furthermore, we have

$$H_n(u) = \sum_{t=p+1}^n A_t(u) + \sum_{t=p+1}^n B_t(u) + \sum_{t=p+1}^n C_t(u), \tag{A.7}$$

where

$$\begin{aligned} A_t(u) &= \frac{1}{\sqrt{h_t(\delta_0)}} [|\varepsilon_t(\gamma_0 + u_1)| - |\varepsilon_t(\gamma_0)|], \\ B_t(u) &= \log \sqrt{h_t(\delta_0 + u_2)} - \log \sqrt{h_t(\delta_0)} + \frac{|\varepsilon_t(\gamma_0)|}{\sqrt{h_t(\delta_0 + u_2)}} - \frac{|\varepsilon_t(\gamma_0)|}{\sqrt{h_t(\delta_0)}}, \\ C_t(u) &= \left[\frac{1}{\sqrt{h_t(\delta_0 + u_2)}} - \frac{1}{\sqrt{h_t(\delta_0)}} \right] [|\varepsilon_t(\gamma_0 + u_1)| - |\varepsilon_t(\gamma_0)|]. \end{aligned}$$

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

$$A_t(u) = u' K_{1t} [I(\eta_t < 0) - I(\eta_t > 0)] + 2 \int_0^{u' K_{1t}} X_t(s) ds, \tag{A.8}$$

where $X_t(s) = I(\eta_t \leq s) - I(\eta_t \leq 0)$ and $K_{1t} = (Y'_{1t}, 0_{1 \times (p+1)})'$. Then, from (A.8), we have

$$\sum_{t=p+1}^n A_t(u) = u' T_{1n} + \Pi_{1n}(u) + \Pi_{2n}(u), \tag{A.9}$$

where

$$\begin{aligned} T_{1n} &= \sum_{t=p+1}^n K_{1t} [I(\eta_t < 0) - I(\eta_t > 0)], \\ \Pi_{1n}(u) &= 2 \sum_{t=p+1}^n \int_0^{u' K_{1t}} \{X_t(s) - E[X_t(s) | \mathcal{F}_{t-1}]\} ds, \\ \Pi_{2n}(u) &= 2 \sum_{t=p+1}^n \int_0^{u' K_{1t}} E[X_t(s) | \mathcal{F}_{t-1}] ds. \end{aligned}$$

Let $K_{2t} = (0_{1 \times p}, Y'_{2t})'$. By Taylor's expansion, we can see that

$$\sum_{t=p+1}^n B_t(u) = u' T_{2n} + \Pi_{3n}(u), \tag{A.10}$$

where

$$T_{2n} = \frac{1}{2} \sum_{t=p+1}^n K_{2t}(1 - |\eta_t|),$$

$$\Pi_{3n}(u) = u' \sum_{t=p+1}^n \left(\frac{3}{8} \left| \frac{\varepsilon_t(\gamma_0)}{\sqrt{h_t(\xi^*)}} \right| - \frac{1}{4} \right) \frac{1}{h_t^2(\xi^*)} \frac{\partial h_t(\xi^*)}{\partial \theta} \frac{\partial h_t(\xi^*)}{\partial \theta'} u,$$

and ξ^* lies between δ_0 and $\delta_0 + u_2$.

We need two lemmas. The first is directly from the Central Limit Theorem; the second gives the expansions of $\Pi_{in}(u)$ for $i = 1, 2, 3$ and $\sum_{t=1}^n C_t(u)$, and its proof is analogous to those of Lemmas 2.2 and 2.3 in Zhu and Ling (2011).

Lemma A.2. Let $T_n = T_{1n} + T_{2n}$. If Assumptions 1–3 hold, then

$$\frac{1}{\sqrt{n}} T_n \rightarrow_d N(0, \Omega_0) \quad \text{as } n \rightarrow \infty,$$

where

$$\Omega_0 = \begin{pmatrix} E(Y_{1t}Y'_{1t}) & \frac{E\eta_t}{2} E(Y_{1t}Y'_{2t}) \\ \frac{E\eta_t}{2} E(Y_{2t}Y'_{1t}) & \frac{E\eta_t^2 - 1}{4} E(Y_{2t}Y'_{2t}) \end{pmatrix}.$$

Lemma A.3. If Assumptions 1–3 hold, then for any sequence of random variables u_n such that $u_n = o_p(1)$, it follows that

- (i) $\Pi_{1n}(u_n) = o_p(\sqrt{n}\|u_n\| + n\|u_n\|^2)$,
- (ii) $\Pi_{2n}(u_n) = (\sqrt{n}u_n)' \Sigma_1 (\sqrt{n}u_n) + o_p(n\|u_n\|^2)$,
- (iii) $\Pi_{3n}(u_n) = (\sqrt{n}u_n)' \Sigma_2 (\sqrt{n}u_n) + o_p(n\|u_n\|^2)$,
- (iv) $\sum_{t=1}^n C_t(u_n) = o_p(n\|u_n\|^2)$,

where

$$\Sigma_1 = \text{diag} \left\{ f(0)E(Y_{1t}Y'_{1t}), 0_{(p+1) \times (p+1)} \right\} \quad \text{and} \quad \Sigma_2 = \text{diag} \left\{ 0_{p \times p}, \frac{1}{8} E(Y_{2t}Y'_{2t}) \right\}.$$

Proof of Theorem 2. We have $\hat{u}_n = o_p(1)$ by Theorem 1. Furthermore, by (A.7), (A.9)–(A.10) and Lemma A.3, we have

$$H_n(\hat{u}_n) = \hat{u}'_n T_n + (\sqrt{n}\hat{u}_n)' \Sigma_0 (\sqrt{n}\hat{u}_n) + o_p(\sqrt{n}\|\hat{u}_n\| + n\|\hat{u}_n\|^2), \quad (\text{A.11})$$

where $\Sigma_0 = \Sigma_1 + \Sigma_2$. Let $\lambda_{\min} > 0$ be the minimum eigenvalue of Σ_0 . Then

$$H_n(\hat{u}_n) \geq -\|\sqrt{n}\hat{u}_n\| \left[\left\| \frac{1}{\sqrt{n}} T_n \right\| + o_p(1) \right] + n\|\hat{u}_n\|^2 [\lambda_{\min} + o_p(1)].$$

Note that $H_n(\hat{u}_n) \leq 0$. By the previous inequality, it follows that

$$\sqrt{n}\|\hat{u}_n\| \leq [\lambda_{\min} + o_p(1)]^{-1} \left[\left\| \frac{1}{\sqrt{n}}T_n \right\| + o_p(1) \right] = O_p(1), \quad (\text{A.12})$$

where the last step holds by Lemma A.2. Next, let $u_n^* = -\Sigma_0^{-1}T_n/2n$. Then, by Lemma A.2, we have

$$\sqrt{n}u_n^* \rightarrow_d N \left(0, \frac{1}{4}\Sigma_0^{-1}\Omega_0\Sigma_0^{-1} \right) \text{ as } n \rightarrow \infty.$$

Hence, it is sufficient to show that $\sqrt{n}\hat{u}_n - \sqrt{n}u_n^* = o_p(1)$. By (A.11) and (A.12), we have

$$\begin{aligned} H_n(\hat{u}_n) &= (\sqrt{n}\hat{u}_n)' \frac{1}{\sqrt{n}}T_n + (\sqrt{n}\hat{u}_n)' \Sigma_0(\sqrt{n}\hat{u}_n) + o_p(1) \\ &= (\sqrt{n}\hat{u}_n)' \Sigma_0(\sqrt{n}\hat{u}_n) - 2(\sqrt{n}\hat{u}_n)' \Sigma_0(\sqrt{n}u_n^*) + o_p(1). \end{aligned}$$

Note that (A.11) still holds when \hat{u}_n is replaced by u_n^* . Thus,

$$\begin{aligned} H_n(u_n^*) &= (\sqrt{n}u_n^*)' \frac{1}{\sqrt{n}}T_n + (\sqrt{n}u_n^*)' \Sigma_0(\sqrt{n}u_n^*) + o_p(1) \\ &= -(\sqrt{n}u_n^*)' \Sigma_0(\sqrt{n}u_n^*) + o_p(1). \end{aligned}$$

By the previous two equations, it follows that

$$\begin{aligned} H_n(\hat{u}_n) - H_n(u_n^*) &= (\sqrt{n}\hat{u}_n - \sqrt{n}u_n^*)' \Sigma_0(\sqrt{n}\hat{u}_n - \sqrt{n}u_n^*) + o_p(1) \\ &\geq \lambda_{\min} \|\sqrt{n}\hat{u}_n - \sqrt{n}u_n^*\|^2 + o_p(1). \end{aligned} \quad (\text{A.13})$$

Since $H_n(\hat{u}_n) - H_n(u_n^*) = n [L_n(\theta_0 + \hat{u}_n) - L_n(\theta_0 + u_n^*)] \leq 0$ a.s., by (A.13) we have $\|\sqrt{n}\hat{u}_n - \sqrt{n}u_n^*\| = o_p(1)$.

Proof of Corollary 1. First, since $|K(x) - K(y)| \leq L|x - y|$ for some $L > 0$, by Taylor's expansion we have

$$\begin{aligned} \left| \hat{f}(0) - \frac{1}{nb_n} \sum_{t=p+1}^n K \left(\frac{\eta_t}{b_n} \right) \right| &\leq \frac{L}{nb_n^2} \sum_{t=p+1}^n |\hat{\eta}_t - \eta_t| \\ &= \frac{L |\hat{\theta}_n - \theta_0|}{nb_n^2} \sum_{t=p+1}^n \left| \frac{\partial \eta_t(\xi_n)}{\partial \theta} \right|, \end{aligned} \quad (\text{A.14})$$

where ξ_n lies between θ_0 and $\hat{\theta}_n$. Note that

$$\frac{\partial \eta_t(\theta)}{\partial \theta} = \frac{\partial \varepsilon_t(\gamma)}{\partial \theta} \frac{1}{\sqrt{h_t(\delta)}} - \frac{\varepsilon_t(\gamma)}{2h_t^{3/2}(\delta)} \frac{\partial h_t(\delta)}{\partial \theta}$$

and $\varepsilon_t(\gamma) = \varepsilon_t + \sum_{i=1}^p (\phi_{i0} - \phi_i) y_{t-i}$. By Assumption 1, we can show that

$$\begin{aligned} E \left[\sup_{\theta} \left\| \frac{\partial \eta_t(\theta)}{\partial \theta} \right\| \right] &\leq E \left[\frac{\sqrt{\sum_{i=1}^p y_{t-i}^2}}{\sqrt{w + \sum_{i=1}^p \alpha_i y_{t-i}^2}} \right] + E \left[\sup_{\theta} \frac{|\varepsilon_t(\gamma)| \sqrt{1 + \sum_{i=1}^p y_{t-i}^4}}{2 (w + \sum_{i=1}^p \alpha_i y_{t-i}^2)^{3/2}} \right] \\ &\leq O(1) + O(1) E \left[\frac{\sqrt{w_0 + \sum_{i=1}^p \alpha_{i0} y_{t-i}^2}}{\sqrt{w + \sum_{i=1}^p \alpha_i y_{t-i}^2}} \right] \\ &\quad + O(1) E \left[\frac{\sum_{i=1}^p |y_{t-i}|}{\sqrt{w + \sum_{i=1}^p \alpha_i y_{t-i}^2}} \right] < \infty. \end{aligned}$$

Thus, by Theorem 3.1 in Ling and McAleer (2003) and the Dominated Convergence Theorem, we have

$$\frac{1}{n} \sum_{t=p+1}^n \left| \frac{\partial \eta_t(\xi_n)}{\partial \theta} \right| = E \left| \frac{\partial \eta_t(\xi_n)}{\partial \theta} \right| + o_p(1) = E \left| \frac{\partial \eta_t(\theta_0)}{\partial \theta} \right| + o_p(1) = O_p(1). \tag{A.15}$$

Since $\sqrt{n}(\hat{\theta}_n - \theta_0) = O_p(1)$ and $nb_n^4 \rightarrow \infty$ as $n \rightarrow \infty$, by (A.14) and (A.15), it follows that

$$\left| \hat{f}(0) - \frac{1}{nb_n} \sum_{t=p+1}^n K \left(\frac{\eta_t}{b_n} \right) \right| \leq O_p \left(\frac{1}{\sqrt{nb_n^2}} \right) = o_p(1). \tag{A.16}$$

Next, by a direct calculation we have

$$E \left[\frac{1}{b_n} K \left(\frac{\eta_t}{b_n} \right) \right] = \int_{-\infty}^{\infty} K(x) f(b_n x) dx < \infty,$$

where the last inequality holds since $\sup_x f(x) < \infty$ by Assumption 2.2 and $\int_{-\infty}^{\infty} K(x) dx = 1$. Then, by Theorem 3.1 in Ling and McAleer (2003), it follows that

$$\frac{1}{nb_n} \sum_{t=p+1}^n K \left(\frac{\eta_t}{b_n} \right) = E \left[\frac{1}{b_n} K \left(\frac{\eta_t}{b_n} \right) \right] + o_p(1). \tag{A.17}$$

Furthermore, since $\int_{-\infty}^{\infty} |x|K(x)dx < \infty$ and $b_n \rightarrow 0$ as $n \rightarrow \infty$, we have

$$\begin{aligned} \left| E \left[\frac{1}{b_n} K \left(\frac{\eta_t}{b_n} \right) \right] - f(0) \right| &= \left| \int_{-\infty}^{\infty} K(x) [f(b_n x) - f(0)] dx \right| \\ &\leq b_n \sup_x |f'(x)| \int_{-\infty}^{\infty} |x|K(x)dx \rightarrow 0 \end{aligned} \tag{A.18}$$

as $n \rightarrow \infty$. By (A.16)–(A.18), we know that $\hat{f}(0) = f(0) + o_p(1)$. Finally, by a similar argument as for (A.15), we can show that $\hat{\Delta}_{1n} = E(Y_{1t}Y'_{1t}) + o_p(1)$, $\hat{\Delta}_{2n} = E(Y_{2t}Y'_{2t}) + o_p(1)$, $\hat{\Delta}_{3n} = E(Y_{1t}Y'_{2t}) + o_p(1)$, $\hat{\Upsilon}_{1n} = E\eta_t + o_p(1)$, and $\hat{\Upsilon}_{2n} = E\eta_t^2 + o_p(1)$. This completes the proof.

References

- Berkes, I. and Horváth, L. (2004). The efficiency of the estimators of the parameters in GARCH processes. *Ann. Statist.* **32**, 633-655.
- Berkes, I., Horváth, L. and Kokoszka, P. (2003). GARCH processes: structure and estimation. *Bernoulli* **9**, 201-207.
- 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** (Edited by R. F. Engle and D. L. McFadden), 2961-3038, North-Holland, Amsterdam.
- Borkovec, M. and Klüppelberg, C. (2001). The tail of the stationary distribution of an autoregressive process with ARCH(1) errors. *Ann. Appl. Probab.* **11**, 1220-1241.
- Chan, N. H. and Peng, L. (2005). Weighted least absolute deviation estimation for an AR(1) process with ARCH(1) errors. *Biometrika* **92**, 477-484.
- Davis, R. A. (1996). Gauss-Newton and M-estimation for ARMA processes with infinite variance. *Stochastic Process. Appl.* **63**, 75-95.
- Davis, R. A. and Dunsmuir, W. T. M. (1997). Least absolute deviation estimation for regression with ARMA errors. *J. Theoret. Probab.* **10**, 481-497.
- Davis, R. A., Knight, K. and Liu, J. (1992). M-estimation for autoregressions with infinite variance. *Stochastic Process. Appl.* **40**, 145-180.
- Engle, R. F. (1982). Autoregressive conditional heteroskedasticity with estimates of variance of U.K. inflation. *Econometrica* **50**, 987-1008.
- Fan, J. and Yao, Q. (2003). *Nonlinear Time Series: Nonparametric and Parametric Methods*. Springer, New York.
- Francq, C. and Zakoïan, J. M. (1998). Estimating linear representations of nonlinear processes. *J. Statist. Plann. Inference* **68**, 145-165.
- Francq, C. and Zakoïan, J. M. (2000). Covariance matrix estimation for estimators of mixing weak ARMA models. *J. Statist. Plann. Inference* **83**, 369-394.
- Francq, C. and Zakoïan, J. M. (2004). Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes. *Bernoulli* **10**, 605-637.
- Francq, C. and Zakoïan, J. M. (2010). *GARCH Models: Structure, Statistical Inference and Financial Applications*. John Wiley.
- Hall, P. and Yao, Q. (2003). Inference in ARCH and GARCH models. *Econometrica* **71**, 285-317.
- Horváth, L. and Liese, F. (2004). L_p -estimators in ARCH models. *J. Statist. Plann. Inference* **119**, 277-309.
- Huber, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, Vol.1, 151-163, University of California Press, Berkeley, LA.

- Jensen, S. T. and Rahbek, A. (2004). Asymptotic normality of the QMLE estimator of ARCH in the nonstationary case. *Econometrica* **72**, 641-646.
- Knight, K. (1987). Rate of convergence of centred estimates of autoregressive parameters for infinite variance autoregressions. *J. Time Ser. Anal.* **8**, 51-60.
- Knight, K. (1998). Limiting distributions for L_1 regression estimators under general conditions. *Ann. Statist.* **26**, 755-770.
- Lang, W. T., Rahbek, A. and Jensen, S. T. (2011). Estimation and asymptotic inference in the first order AR-ARCH model. *Econometric Reviews* **30**, 129-153.
- 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.
- 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.
- Ling, S. (2004). Estimation and testing stationarity for double autoregressive models. *J. Roy. Statist. Soc. Ser. B* **66**, 63-78.
- Ling, S. (2005). Self-weighted least absolute deviation estimation for infinite variance autoregressive models. *J. Roy. Statist. Soc. Ser. B* **67**, 381-393.
- Ling, S. (2007). A double AR(p) model: structure and estimation. *Statistica Sinica* **17**, 161-175.
- Ling, S. and Li, D. (2008). Asymptotic inference for a nonstationary double AR(1) model. *Biometrika* **95**, 257-263.
- Ling, S. and McAleer, M. (2003). Asymptotic theory for a new vector ARMA-GARCH model. *Econometric Theory* **19**, 280-310.
- Pan, J., Wang, H. and Yao, Q. (2007). Weighted least absolute deviations estimation for ARMA models with infinite variance. *Econometric Theory* **23**, 852-879.
- Peng, L. and Yao, Q. (2003). Least absolute deviations estimation for ARCH and GARCH models. *Biometrika* **90**, 967-975.
- Weiss, A. A. (1986). Asymptotic theory for ARCH models: estimation and testing. *Econometric Theory* **2**, 107-131.
- Zhu, K. and Ling, S. (2011). Global self-weighted and local quasi-maximum exponential likelihood estimators for ARMA-GARCH/IGARCH models. *Ann. Statist.* **39**, 2131-2163.
- Zhu, K. and Ling, S. (2012). The global weighted LAD estimators for finite/infinite variance ARMA(p, q) models. *Econometric Theory*. **28**, 1065-1086.

Chinese Academy of Sciences, NCMIS, AMSS, Haidian District, Zhongguancun, Bei Jing, China.

E-mail: zkxaa@ust.hk

Department of Mathematics, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong.

E-mail: maling@ust.hk

(Received April 2011; accepted February 2012)