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

Ke Zhu and Shiqing Ling

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 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.

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

1. Introduction

Consider the autoregressive (AR) model with conditional heteroscedasticity:

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

where \( w > 0, \alpha_i > 0 \) (\( i = 1, \ldots, p \)), \( \{\eta_t\} \) 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 \( \sigma \)-field generated by \( \{\eta_t, \ldots, \eta_1, y_0, \ldots, y_{t-p}\} \). The conditional variance of \( y_t \) is \( \text{var}(y_t|\mathcal{F}_{t-1}) = E\eta_t^2 \left(w + \sum_{i=1}^{p} \alpha_i y_{t-i}^2\right) \) 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

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 $\eta_t$ 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), 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 $\eta_t$ 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 $\gamma = (\phi_1, \ldots, \phi_p)'$ and $\delta = (w, \alpha_1, \ldots, \alpha_p)'$. 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 $\{y_1, \ldots, y_n\}$ are generated by model (1.1). When $\eta_t$ follows the standard double exponential distribution, the log-likelihood function (ignoring a constant) can be written as

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

(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 \( \eta_t \) follows the standard double exponential distribution, \( \hat{\theta}_n \) is called the global quasi-maximum exponential likelihood estimator (QMELE) of \( \theta_0 \).

**Assumption 1.** \( \Theta \) is compact with \( \underline{w} \leq w \leq \bar{w} \) and \( \underline{\alpha}_i \leq \alpha_i \leq \bar{\alpha}_i \) \((i = 1, \ldots, p)\), where \( \underline{w}, \bar{w}, \underline{\alpha}_i \) and \( \bar{\alpha}_i \) \((i = 1, \ldots, p)\) are some positive constants, and \( \theta_0 \) is an interior point in \( \Theta \).

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

**Assumption 3.** \( \eta_t \) 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 \((\phi, \alpha)\) for the DAR(1) models when \( \eta_t \sim N(0, \pi/2) \) 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 \( \eta_t \) has zero median with \( E|\eta_t| = 1 \). If Assumptions 1–2 hold, then \( \hat{\theta}_n \to \theta_0 \) almost surely \( (a.s.) \) as \( n \to \infty \).

**Theorem 2.** If Assumptions 1–3 hold with \( E|\eta_t| = 1 \), then
\[ \sqrt{n}(\hat{\theta}_n - \theta_0) \to_d N \left( 0, \frac{1}{4} \Sigma_0^{-1} \Omega_0 \Sigma_0^{-1} \right) \quad \text{as} \quad n \to \infty, \]
where \( \to_d \) denotes the convergence in distribution and
\[ \Sigma_0 = \text{diag} \left\{ f(0)E \left( Y_{1t}Y_{1t}' \right), \frac{1}{8} E \left( Y_{2t}Y_{2t}' \right) \right\}, \]
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

\[Y_{1t} = \frac{1}{\sqrt{h_t(\delta_0)}}(y_{t-1}, \ldots, y_{t-p})' \text{ and } Y_{2t} = \frac{1}{h_t(\delta_0)}(1, y_{2t-1}, \ldots, y_{2t-p})'.\]

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

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

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 matrices \( \Sigma_0 \) and \( \Omega_0 \) in Theorem 2 can be estimated via

\[
\hat{\Sigma}_0 = \text{diag} \left\{ \hat{f}(0) \hat{\Delta}_{1n}, \frac{1}{8} \hat{\Delta}_{2n} \right\} \quad \text{and} \quad \hat{\Omega}_0 = \left( \begin{array}{ccc} \hat{\Delta}_{1n} & \frac{1}{2} \hat{\Upsilon}_{1n} \hat{\Delta}_{3n} \\ \frac{1}{2} \hat{\Upsilon}_{1n} \hat{\Delta}_{3n} & \frac{1}{12} \hat{\Upsilon}_{1n} \hat{\Delta}_{3n} \end{array} \right) \left( \hat{\Upsilon}_{2n} - 1 \right) \hat{\Delta}_{2n},
\]

(2.2)

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

\[
\hat{f}(0) = \frac{1}{nb_n} \sum_{t=p+1}^{n} K \left( \frac{\hat{\eta}_t}{b_n} \right), \quad \hat{\Upsilon}_{1n} = \frac{1}{n} \sum_{t=p+1}^{n} \hat{\eta}_t, \quad \hat{\Upsilon}_{2n} = \frac{1}{n} \sum_{t=p+1}^{n} \hat{\eta}_t^2,
\]

\[
\hat{\Delta}_{1n} = \frac{1}{n} \sum_{t=p+1}^{n} \frac{1}{h_t(\delta_n)}(y_{t-1}, \ldots, y_{t-p})'(y_{t-1}, \ldots, y_{t-p}),
\]

\[
\hat{\Delta}_{2n} = \frac{1}{n} \sum_{t=p+1}^{n} \frac{1}{h_t^2(\delta_n)}(1, y_{t-1}^2, \ldots, y_{t-p}^2)'(1, y_{t-1}^2, \ldots, y_{t-p}^2),
\]

\[
\hat{\Delta}_{3n} = \frac{1}{n} \sum_{t=p+1}^{n} \frac{1}{h_t^{3/2}(\delta_n)}(y_{t-1}, \ldots, y_{t-p})'(1, y_{t-1}^2, \ldots, y_{t-p}^2).
\]

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 \( b_n \to 0 \), \( nb_n^4 \to \infty \) as \( n \to \infty \), then

\[
\hat{\Sigma}_0 \to \Sigma_0 \quad \text{and} \quad \hat{\Omega}_0 \to \Omega_0, \quad \text{as} \quad n \to \infty.
\]

**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

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

\( b_n^* = 1.06sn^{-1/5} \) 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 \( E\eta^3_t = 0 \) and \( E\eta^4_t < \infty \), and reparametrize (1.1) as

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

where \( \eta_t^* = \eta_t / \sqrt{E\eta_t^2} \) and \( \delta^* = (E\eta_t^2) \delta \). Let \( \tilde{\theta}_n^* = (\tilde{\gamma}_n^*, \tilde{\delta}_n^*)' \) be the QMLE of the true parameter \( \theta_0^* = (\gamma_0^*, \delta_0^*)' \) in model (2.3), Then, \( \hat{\theta}_n = (\hat{\gamma}_n, \hat{\delta}_n)' \) is the QMLE of \( \theta_0 = (\gamma_0, \delta_0)' \), where \( \hat{\delta}_n = \delta_n^* / E\eta_t^2 \). Thus, by Theorem 3.1 in Ling (2007), the asymptotic covariance of \( \hat{\theta}_n \) is \( \Gamma_1 \), where

\[
\Gamma_1 = \left( \begin{array}{cc}
E\eta_t^2 J_1^{-1} & 0 \\
0 & \kappa J_2^{-1}
\end{array} \right),
\]

with \( J_1 = E(Y_1Y_1') \), \( J_2 = E(Y_2Y_2') \), and \( \kappa = E\eta_t^4 / (E\eta_t^2)^2 - 1 \).

We now compare \( \Gamma_0 \) with \( \Gamma_1 \) 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 \( \varepsilon = 1 \), \( \tau = \sqrt{\pi/2} \),

\[
\Gamma_0 = \left( \begin{array}{cc}
\frac{\pi^2}{4} J_1^{-1} & 0 \\
0 & (2\pi - 4) J_2^{-1}
\end{array} \right) \quad \text{and} \quad \Gamma_1 = \left( \begin{array}{cc}
\frac{\pi^2}{4} J_1^{-1} & 0 \\
0 & 2 J_2^{-1}
\end{array} \right).
\]

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

\[
\Gamma_0 = \left( \begin{array}{cc}
1.64 J_1^{-1} & 0 \\
0 & 3.44 J_2^{-1}
\end{array} \right) \quad \text{and} \quad \Gamma_1 = \left( \begin{array}{cc}
1.86 J_1^{-1} & 0 \\
0 & 3.60 J_2^{-1}
\end{array} \right).
\]

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

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

\[
\Gamma_0 = \left( \begin{array}{cc}
J_1^{-1} & 0 \\
0 & 4 J_2^{-1}
\end{array} \right) \quad \text{and} \quad \Gamma_1 = \left( \begin{array}{cc}
2 J_1^{-1} & 0 \\
0 & 5 J_2^{-1}
\end{array} \right).
\]

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 \( \eta_t \sim \text{Laplace}(0, 1) \). Finally, we compare the asymptotic efficiency of the WLAD 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 WLDAE is
\[
\begin{pmatrix}
\frac{E_{\eta}^2}{4m} \left( E\frac{y_2^2}{\bar{s}^2} + E\frac{y_3^2}{\bar{s}^2} \right) J_1^{-2} & 0 \\
0 & \frac{1}{2m} \left( \frac{1}{J_2} J_1^{-1} \right)
\end{pmatrix},
\]
where \( m = \text{median}\{\eta^2_1\} \) and \( S = 1 + y^2_{t-1} \). 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 WLDAE for parameters \((\phi, w, \alpha)\). For the parameter \( \phi \), it is hard to compare them with each other in theory, since the asymptotic covariance of WLDAE 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 WLDAE in finite samples. The DAR(1) model used to generate data samples was
\[
y_t = \phi y_{t-1} + \eta_t \sqrt{w + \alpha y^2_{t-1}}. \tag{3.1}
\]
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 \( \eta_t \) as \( \text{Laplace}(0, 1) \), \( \text{N}(0, 1) \), 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 WLDAE \((\tilde{\theta}_n^*)\) are estimators of \((\phi_0, rw_0, r\alpha_0)\) with \( r = (E|\eta_t|)^2, E_{\eta_t}, \) and \( \text{median}\{\eta^2_t\} \), respectively. In order to make our comparison feasible, we let
\[
\hat{\theta}_n = \left( \hat{\phi}_n^*, \hat{w}_n^*, \hat{\alpha}_n^* \right), \quad \bar{\theta}_n = \left( \bar{\phi}_n^*, \bar{w}_n^*, \bar{\alpha}_n^* \right),
\]
and
\[
\tilde{\theta}_n = \left( \tilde{\phi}_n^*, \tilde{w}_n^*, \tilde{\alpha}_n^* \right)
\]
be the QMELE, QMLE, and WLDAE of \((\phi_0, w_0, \alpha_0)\), 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 \( E|\eta_t|, E_{\eta_t}, \) and \( \text{median}\{\eta^2_t\} \).

Tables 1–3 list the sample biases, the sample standard deviations (SD), and the average estimated asymptotic standard deviations (AD) of \( \hat{\theta}_n, \theta_n, \) 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 WLDAE, while the QMLE is better than the WLDAE. When
Table 1. $\eta_t \sim \text{Laplace}(0,1)$.

<table>
<thead>
<tr>
<th>$\phi_0$</th>
<th>$w_0$</th>
<th>$\alpha_0$</th>
<th>QMLE($\hat{\theta}_n$)</th>
<th>QMLE($\bar{\theta}_n$)</th>
<th>WLAE($\tilde{\theta}_n$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0</td>
<td>1.0</td>
<td>0.5</td>
<td>Bias -0.0040</td>
<td>-0.0045</td>
<td>-0.0043</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SD 0.0437</td>
<td>0.0576</td>
<td>0.0818</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>AD 0.0475</td>
<td>0.0569</td>
<td>0.0801</td>
</tr>
<tr>
<td>0.5</td>
<td>1.0</td>
<td>0.5</td>
<td>Bias -0.0040</td>
<td>-0.0033</td>
<td>-0.0029</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SD 0.0547</td>
<td>0.0729</td>
<td>0.1063</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>AD 0.0606</td>
<td>0.0729</td>
<td>0.1046</td>
</tr>
<tr>
<td>0.0</td>
<td>1.0</td>
<td>1.0</td>
<td>Bias -0.0042</td>
<td>-0.0029</td>
<td>0.0005</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SD 0.0689</td>
<td>0.0924</td>
<td>0.1348</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>AD 0.0765</td>
<td>0.0926</td>
<td>0.1311</td>
</tr>
<tr>
<td>0.0</td>
<td>1.0</td>
<td>0.5</td>
<td>Bias -0.0040</td>
<td>-0.0011</td>
<td>-0.0016</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SD 0.0591</td>
<td>0.0816</td>
<td>0.1199</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>AD 0.0660</td>
<td>0.0793</td>
<td>0.1145</td>
</tr>
</tbody>
</table>

$\eta_t \sim N(0, 1)$, in Table 2 we can see that the QMLE has smaller AD and SD than those of both the QMELE and the WLAE, and the SD and AD of the WLAE of $\phi_0$ is slightly smaller than those of the QMELE, while the SD and AD of the WLAE of $\omega_0$ and $\alpha_0$ 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 $\phi_0$ are smaller than those of the WLAE, while the SD and AD of the WLAE of $\omega_0$ and $\alpha_0$ are smaller than those of the QMLE. We also note that the SD and AD of the QMLE of $\omega_0$ and $\alpha_0$ 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 $\{y_t^2\}$,

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

where $\hat{y}_j$ is the $j$-th order statistic of $y_t^2$. The plot of $\{\hat{H}_y(k)\}_{k=1}^{180}$ 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. $\eta_t \sim N(0,1)$.

<table>
<thead>
<tr>
<th>$\phi_0$</th>
<th>$\psi_0$</th>
<th>$\alpha_0$</th>
<th>$\hat{\phi}_n$</th>
<th>$\hat{\psi}_n$</th>
<th>$\hat{\alpha}_n$</th>
<th>$\tilde{\phi}_n$</th>
<th>$\tilde{\psi}_n$</th>
<th>$\tilde{\alpha}_n$</th>
<th>$\bar{\phi}_n$</th>
<th>$\bar{\psi}_n$</th>
<th>$\bar{\alpha}_n$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0</td>
<td>1.0</td>
<td>0.5</td>
<td>Bias -0.0029</td>
<td>0.0094</td>
<td>-0.0030</td>
<td>-0.0009</td>
<td>0.0055</td>
<td>-0.0033</td>
<td>0.0003</td>
<td>0.0183</td>
<td>0.0040</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SD 0.0574</td>
<td>0.1568</td>
<td>0.0583</td>
<td>0.0448</td>
<td>0.1451</td>
<td>0.0533</td>
<td>0.0530</td>
<td>0.2470</td>
<td>0.0932</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>AD 0.0599</td>
<td>0.1553</td>
<td>0.0575</td>
<td>0.0455</td>
<td>0.1435</td>
<td>0.0532</td>
<td>0.0554</td>
<td>0.2522</td>
<td>0.0930</td>
</tr>
<tr>
<td>0.5</td>
<td>1.0</td>
<td>0.5</td>
<td>Bias -0.0009</td>
<td>0.0115</td>
<td>-0.0124</td>
<td>-0.0036</td>
<td>0.0115</td>
<td>0.0883</td>
<td>0.0723</td>
<td>0.1950</td>
<td>0.1458</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SD 0.0762</td>
<td>0.1258</td>
<td>0.0940</td>
<td>0.0605</td>
<td>0.1175</td>
<td>0.0883</td>
<td>0.0723</td>
<td>0.1950</td>
<td>0.1458</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>AD 0.0772</td>
<td>0.1228</td>
<td>0.0917</td>
<td>0.0586</td>
<td>0.1134</td>
<td>0.0852</td>
<td>0.0718</td>
<td>0.1993</td>
<td>0.1498</td>
</tr>
<tr>
<td>0.0</td>
<td>1.0</td>
<td>1.0</td>
<td>Bias -0.0015</td>
<td>0.0187</td>
<td>-0.0260</td>
<td>-0.0027</td>
<td>0.0142</td>
<td>-0.0227</td>
<td>0.0034</td>
<td>0.0224</td>
<td>-0.0131</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SD 0.0893</td>
<td>0.1457</td>
<td>0.1383</td>
<td>0.0693</td>
<td>0.1351</td>
<td>0.1298</td>
<td>0.0811</td>
<td>0.2249</td>
<td>0.2105</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>AD 0.0921</td>
<td>0.1427</td>
<td>0.1372</td>
<td>0.0701</td>
<td>0.1318</td>
<td>0.1276</td>
<td>0.0844</td>
<td>0.2278</td>
<td>0.2188</td>
</tr>
<tr>
<td>0.0</td>
<td>1.0</td>
<td>0.5</td>
<td>Bias -0.0026</td>
<td>0.0101</td>
<td>-0.0101</td>
<td>-0.0002</td>
<td>0.0088</td>
<td>-0.0087</td>
<td>0.0014</td>
<td>0.0069</td>
<td>0.0084</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SD 0.0769</td>
<td>0.1219</td>
<td>0.1108</td>
<td>0.0605</td>
<td>0.1134</td>
<td>0.1040</td>
<td>0.0733</td>
<td>0.1892</td>
<td>0.1679</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>AD 0.0823</td>
<td>0.1197</td>
<td>0.1051</td>
<td>0.0627</td>
<td>0.1109</td>
<td>0.0978</td>
<td>0.0773</td>
<td>0.1929</td>
<td>0.1723</td>
</tr>
</tbody>
</table>

Table 3. $\eta_t \sim t_3$.

<table>
<thead>
<tr>
<th>$\phi_0$</th>
<th>$\psi_0$</th>
<th>$\alpha_0$</th>
<th>$\hat{\phi}_n$</th>
<th>$\hat{\psi}_n$</th>
<th>$\hat{\alpha}_n$</th>
<th>$\tilde{\phi}_n$</th>
<th>$\tilde{\psi}_n$</th>
<th>$\tilde{\alpha}_n$</th>
<th>$\bar{\phi}_n$</th>
<th>$\bar{\psi}_n$</th>
<th>$\bar{\alpha}_n$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0</td>
<td>1.0</td>
<td>0.5</td>
<td>Bias -0.0021</td>
<td>0.0040</td>
<td>-0.0042</td>
<td>-0.0019</td>
<td>-0.0022</td>
<td>-0.0053</td>
<td>-0.0002</td>
<td>0.0760</td>
<td>0.0114</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SD 0.0550</td>
<td>0.3516</td>
<td>0.0751</td>
<td>0.0668</td>
<td>0.5706</td>
<td>0.2172</td>
<td>0.0860</td>
<td>0.3786</td>
<td>0.0849</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>AD 0.0552</td>
<td>0.3510</td>
<td>0.0727</td>
<td>0.0677</td>
<td>0.6113</td>
<td>0.1759</td>
<td>0.0911</td>
<td>0.3977</td>
<td>0.0804</td>
</tr>
<tr>
<td>0.5</td>
<td>1.0</td>
<td>0.5</td>
<td>Bias -0.0032</td>
<td>0.0218</td>
<td>-0.0175</td>
<td>-0.0024</td>
<td>-0.0089</td>
<td>-0.0746</td>
<td>-0.0023</td>
<td>0.0332</td>
<td>0.0061</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SD 0.0684</td>
<td>0.2155</td>
<td>0.1005</td>
<td>0.0846</td>
<td>0.4201</td>
<td>0.2108</td>
<td>0.1116</td>
<td>0.2310</td>
<td>0.1107</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>AD 0.0691</td>
<td>0.2114</td>
<td>0.1026</td>
<td>0.0831</td>
<td>0.4434</td>
<td>0.2055</td>
<td>0.1077</td>
<td>0.2313</td>
<td>0.1149</td>
</tr>
<tr>
<td>0.0</td>
<td>1.0</td>
<td>1.0</td>
<td>Bias -0.0039</td>
<td>0.0230</td>
<td>-0.0128</td>
<td>0.0052</td>
<td>-0.0096</td>
<td>-0.1276</td>
<td>0.0067</td>
<td>0.0424</td>
<td>0.0241</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SD 0.0833</td>
<td>0.2657</td>
<td>0.1778</td>
<td>0.1031</td>
<td>0.4808</td>
<td>0.3412</td>
<td>0.1365</td>
<td>0.2895</td>
<td>0.1982</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>AD 0.0857</td>
<td>0.2594</td>
<td>0.1688</td>
<td>0.1040</td>
<td>0.5572</td>
<td>0.4155</td>
<td>0.1317</td>
<td>0.2804</td>
<td>0.1848</td>
</tr>
<tr>
<td>0.0</td>
<td>1.0</td>
<td>0.5</td>
<td>Bias -0.0016</td>
<td>0.0223</td>
<td>-0.0089</td>
<td>-0.0010</td>
<td>-0.0056</td>
<td>-0.0542</td>
<td>-0.0037</td>
<td>0.0297</td>
<td>0.0153</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SD 0.0714</td>
<td>0.2095</td>
<td>0.1213</td>
<td>0.0924</td>
<td>0.4151</td>
<td>0.2835</td>
<td>0.1224</td>
<td>0.2144</td>
<td>0.1302</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>AD 0.0730</td>
<td>0.2009</td>
<td>0.1178</td>
<td>0.0899</td>
<td>0.4010</td>
<td>0.2850</td>
<td>0.1168</td>
<td>0.2145</td>
<td>0.1297</td>
</tr>
</tbody>
</table>

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

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

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

$$y_t = -0.3511y_{t-1} - 0.2583y_{t-2} + \eta_t \sqrt{0.1207 + 0.1562y_{t-1}^2 + 0.1180y_{t-2}^2} + \epsilon_t$$

$$\begin{align*}
(0.0424) & & (0.0443) & & (0.1456) & & (0.0339) & & (0.0396) \\
\end{align*}
$$

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 \( y^2_t \); (b) the kernel density of \( \eta_t \) (---) 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 \( \{\hat{\eta}_2^t\} \) and (d) the partial autocorrelations of \( \{\hat{\eta}_2^t\} \).

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 \( \theta^* \) with radius \( \eta > 0 \). If Assumptions 1–2 hold, then

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

(ii) \( E \left[ l_t(\theta) \right] \) has a unique minimum at \( \theta_0 \),

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

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

\[
E[l_t(\theta)] = E\left[ \log \sqrt{h_t(\delta)} + \frac{\varepsilon_t(\gamma_0) + \sqrt{h_t(\delta_0)}(\gamma_0 - \gamma)Y_{1t}}{\sqrt{h_t(\delta)}} \right]
\]

\[
= E\left[ \log \sqrt{h_t(\delta)} + \frac{h_t(\delta_0)}{h_t(\delta)}E\left[ |\eta_t(\theta_0) + (\gamma_0 - \gamma)'Y_{1t}| \right] \right]
\]

\[
\geq E\left[ \log \sqrt{h_t(\delta)} + \frac{h_t(\delta_0)}{h_t(\delta)}E\left[ |\eta_t| \right] \right]
\]

\[
= E\left[ \log \sqrt{h_t(\delta)} + \frac{h_t(\delta_0)}{h_t(\delta)} \right],
\]

where the last inequality holds since \( \eta_t \) has zero median, and the minimum is attained if and only if \( \gamma = \gamma_0 \) a.s.; see [Ling (2007)]. Here, \( \xi^* \) lies between \( \gamma \) and \( \gamma_0 \). 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 \( \theta_0 \), 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

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

where \( \zeta^* \) lies between \( \delta \) and \( \delta^* \). Then, by Assumption 2, we have

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

as \( \eta \to 0 \). Similarly,

\[
E\left[ \sup_{\theta \in B_\eta(\theta^*)} \frac{1}{\sqrt{h_t(\delta)}} |\varepsilon_t(\gamma) - |\varepsilon_t(\gamma^*)|| \right] \to 0 \text{ as } \eta \to 0,
\]

\[
E\left[ \sup_{\theta \in B_\eta(\theta^*)} \left| \frac{1}{\sqrt{h_t(\delta)}} - \frac{1}{\sqrt{h_t(\delta^*)}} \right| \right] \to 0 \text{ as } \eta \to 0.
\]

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 \( \theta^* \in \Theta \cap V, \) and \( \varepsilon > 0, \)
there exists an $\eta_0 > 0$ such that
\[
E \left[ \inf_{\theta \in B_{\eta_0}(\theta^*)} l_t(\theta) \right] \geq E[l_t(\theta^*)] - \varepsilon. \tag{A.1}
\]

From Lemma A.1 (i), by the Ergodic Theorem, it follows that
\[
\frac{1}{n} \sum_{t=p+1}^{n} \inf_{\theta \in B_{\eta_0}(\theta^*)} l_t(\theta) \geq E \left[ \inf_{\theta \in B_{\eta_0}(\theta^*)} l_t(\theta) \right] - \varepsilon \tag{A.2}
\]
if $n$ is large enough. Since $V^c$ is compact, we can choose \( \{B_{\eta_0}(\theta_i) : \theta_i \in V^c, i = 1, 2, \ldots, k\} \) to be a finite covering of $V^c$. Thus, from (A.1)−(A.2), we have
\[
\inf_{\theta \in V^c} L_n(\theta) = \min_{1 \leq i \leq k} \inf_{\theta \in B_{\eta_0}(\theta_i)} L_n(\theta) \\
\geq \min_{1 \leq i \leq k} \frac{1}{n} \sum_{t=p+1}^{n} \inf_{\theta \in B_{\eta_0}(\theta_i)} l_t(\theta) \\
\geq \min_{1 \leq i \leq k} E \left[ \inf_{\theta \in B_{\eta_0}(\theta_i)} l_t(\theta) \right] - \varepsilon \tag{A.3}
\]
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
\[
E \left[ \inf_{\theta \in B_{\eta_0}(\theta_i)} l_t(\theta) \right] \geq E[l_t(\theta_i)] + 3\varepsilon_0. \tag{A.4}
\]

Thus, from (A.3)−(A.4), taking $\varepsilon = \varepsilon_0$, it follows that
\[
\inf_{\theta \in V^c} L_n(\theta) \geq E[l_t(\theta_0)] + 2\varepsilon_0 \tag{A.5}
\]
if $n$ is large enough. On the other hand, by the Ergodic Theorem, it follows that
\[
\inf_{\theta \in V} L_n(\theta) \leq L_n(\theta_0) = \frac{1}{n} \sum_{t=p+1}^{n} l_t(\theta_0) \leq E[l_t(\theta_0)] + \varepsilon_0 \tag{A.6}
\]
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 $\hat{\theta}_n \to \theta_0$ 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 \( u \in \Lambda \equiv \{u = (u_1', \ldots, u_p') : u + \theta_0 \in \Theta \} \). Let \( \hat{u}_n = \hat{\theta}_n - \theta_0 \). Then, \( \hat{u}_n \) is the minimizer of \( H_n(u) \) in \( \Lambda \). 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

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

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

\[
A_t(u) = u'K_t[I(\eta_t < 0) - I(\eta_t > 0)] + 2 \int_0^{u'K_t} X_t(s)ds, \tag{A.8}
\]

where \( X_t(s) = I(\eta_t \leq s) - I(\eta_t \leq 0) \) and \( K_t = (Y_{1t}', 0_{1 \times (p+1)})' \). 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

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

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
\[ T_{2n} = \frac{1}{2} \sum_{t=p+1}^{n} K_{2t}(1 - |\eta_t|), \]
\[ \Pi_{3n}(u) = u' \sum_{t=p+1}^{n} \left( \frac{3}{8} \frac{\varepsilon_t(\gamma_0)}{\sqrt{h_t(\xi^*)}} \left( - \frac{1}{4} \right) \right) \frac{1}{h_t^2(\xi^*)} \frac{\partial h_t(\xi^*)}{\partial \theta} \frac{\partial h_t(\xi^*)}{\partial \theta'} u, \]
and \( \xi^* \) lies between \( \delta_0 \) 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
\[ \frac{1}{\sqrt{n}} T_n \to_d N(0, \Omega_0) \quad \text{as} \quad n \to \infty, \]
where
\[ \Omega_0 = \begin{pmatrix} E(Y_{1t}Y_{1t}') & E_n E(Y_{1t}Y_{2t}') \\ E_n E(Y_{2t}Y_{1t}') & E_n E(Y_{2t}Y_{2t}') \end{pmatrix}. \]

**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
\[
\begin{align*}
(i) \quad & \Pi_{1n}(u_n) = o_p(\sqrt{n}\|u_n\| + n\|u_n\|^2), \\
(ii) \quad & \Pi_{2n}(u_n) = (\sqrt{n}u_n)'\Sigma_1(\sqrt{n}u_n) + o_p(n\|u_n\|^2), \\
(iii) \quad & \Pi_{3n}(u_n) = (\sqrt{n}u_n)'\Sigma_2(\sqrt{n}u_n) + o_p(n\|u_n\|^2), \\
(iv) \quad & \sum_{t=1}^{n} C_t(u_n) = o_p(n\|u_n\|^2),
\end{align*}
\]
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
\[
H_n(\hat{u}_n) = \hat{u}_n' T_n + (\sqrt{n}\hat{u}_n)' \Sigma_0(\sqrt{n}\hat{u}_n) + o_p(\sqrt{n}\|\hat{u}_n\| + n\|\hat{u}_n\|^2), \tag{A.11}
\]
where \( \Sigma_0 = \Sigma_1 + \Sigma_2 \). Let \( \lambda_{\min} > 0 \) be the minimum eigenvalue of \( \Sigma_0 \). 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

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

(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

$$
H_n(\hat{u}_n) = (\sqrt{n}\hat{u}_n)' \frac{1}{\sqrt{n}} T_n + (\sqrt{n}\hat{u}_n)' \Sigma_0 (\sqrt{n}\hat{u}_n) + o_p(1)
$$

$$
= (\sqrt{n}\hat{u}_n)' \Sigma_0 (\sqrt{n}\hat{u}_n) - 2 (\sqrt{n}\hat{u}_n)' \Sigma_0 (\sqrt{n}u^*_n) + o_p(1).
$$

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

$$
H_n(u^*_n) = (\sqrt{n}u^*_n)' \frac{1}{\sqrt{n}} T_n + (\sqrt{n}u^*_n)' \Sigma_0 (\sqrt{n}u^*_n) + o_p(1)
$$

$$
= -(\sqrt{n}u^*_n)' \Sigma_0 (\sqrt{n}u^*_n) + o_p(1).
$$

By the previous two equations, it follows that

$$
H_n(\hat{u}_n) - H_n(u^*_n) = (\sqrt{n}\hat{u}_n - \sqrt{n}u^*_n)' \Sigma_0 (\sqrt{n}\hat{u}_n - \sqrt{n}u^*_n) + o_p(1)
$$

$$
\geq \lambda_{\text{min}} \|\sqrt{n}\hat{u}_n - \sqrt{n}u^*_n\|^2 + o_p(1). 
$$

(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

$$
\left| \hat{f}(0) - \frac{1}{nb_n} \sum_{t=p+1}^{n} K\left(\frac{\eta_t}{b_n}\right) \right| \leq \frac{L}{nb_n^2} \sum_{t=p+1}^{n} |\hat{\eta}_t - \eta_t|
$$

$$
= \frac{L}{nb_n^2} \sum_{t=p+1}^{n} \left| \frac{\partial \eta_t(\xi_n)}{\partial \theta} \right|, 
$$

(A.14)

where $\xi_n$ lies between $\theta_0$ and $\hat{\theta}_n$. Note that

$$
\frac{\partial \eta_t(\theta)}{\partial \theta} = \frac{\partial \varepsilon_t(\gamma)}{\partial \theta} \frac{1}{\sqrt{\hat{h}_t(\delta)}} - \frac{\varepsilon_t(\gamma)}{2h_t^{3/2}(\delta)} \frac{\partial h_t(\delta)}{\partial \theta}
$$
and \( \varepsilon_t(\gamma) = \varepsilon_t + \sum_{i=1}^{p} (\phi_0 - \phi_i) y_{t-i} \). By Assumption 1, we can show that

\[
E \left[ \sup_{\theta} \left\| \frac{\partial \eta_t(\theta)}{\partial \theta} \right\| \right] \leq E \left[ \frac{\sqrt{\sum_{i=1}^{p} y_{t-i}^2}}{w + \sum_{i=1}^{p} \sigma_i y_{t-i}^2} \right] + E \left[ \sup_{\theta} \frac{\varepsilon_t(\gamma)}{\sqrt{1 + \sum_{i=1}^{p} \sigma_i y_{t-i}^4}} \right] \\
\leq O(1) + O(1) E \left[ \frac{\sqrt{\sum_{i=1}^{p} \alpha_i y_{t-i}^2}}{w + \sum_{i=1}^{p} \sigma_i y_{t-i}^2} \right]
\]

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

\[
\frac{1}{n} \sum_{t=p+1}^{n} \left| \frac{\partial \eta_t(\xi_n)}{\partial \theta} \right| = E \left[ \frac{\partial \eta_t(\xi_0)}{\partial \theta} \right] + o_p(1) = E \left[ \frac{\partial \eta_t(\theta_0)}{\partial \theta} \right] + o_p(1) = O_p(1). \quad (A.15)
\]

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

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

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

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

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

\[
\left| E \left[ \frac{1}{b_n} K \left( \frac{\eta_t}{b_n} \right) \right] - f(0) \right| = \left| \int_{-\infty}^{\infty} K(x) [f(b_n x) - f(0)] dx \right| \\
\leq b_n \sup_x |f'(x)| \int_{-\infty}^{\infty} |x| K(x) dx \to 0 \quad (A.18)
\]
as \( n \to \infty \). 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

\[
\Delta_{1n} = E(Y_{1t}Y'_{1t}) + o_p(1),
\Delta_{2n} = E(Y_{2t}Y'_{2t}) + o_p(1),
\Delta_{3n} = E(Y_{1t}Y'_{2t}) + o_p(1),
\Upsilon_{1n} = E\eta_t + o_p(1),
\]

and

\[
\Upsilon_{2n} = E\eta^2 + o_p(1).
\]

This completes the proof.

References


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

E-mail: zkxxaa@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)