Self-weighted Quantile Estimation for Infinite Variance Autoregressive Models

Shiqing Ling

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

Abstract

It has been a long-term standing open problem to do inferences for infinite variance AR models. The difficulty is that the estimated parameters based on the existing methods in the literature asymptotically follow some unknown distributions. This paper proposes a self-weighted quantile estimation for this kind of models. It is shown that the estimated parameters are asymptotically normal if the density function of the errors and its derivative are uniformly bounded. The Wald test statistic is constructed for linear restrictions on the parameters and it is shown that the test has non-trivial local powers. Our results basically solve the problem as above and provide a new insight for future research on heavy tailed time series. Simulation studies are carried out to access the performance of the method and theory in finite samples.

Key words and phrases: AR model, Infinite variance, LAD and robust.

1 Introduction

Consider autoregressive (AR) time series process \( \{y_t\} \) generated by the equation:

\[
(1.1) y_t = \phi_0 + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \varepsilon_t,
\]

where \( \{\varepsilon_t\} \) is a sequence of independent and identically distributed (i.i.d.) errors with a common distribution \( F \) and \( 1 - \phi_1 z - \cdots - \phi_p z^p = 0 \) has all roots outside the unit circle. When \( E\varepsilon_t^2 \) is finite, it is well known that all kinds of the estimators of the parameter \( \lambda \equiv (\phi_0, \phi_1, \cdots, \phi_p)' \) are asymptotically normal and various methods are available to do inferences for the model. When \( E\varepsilon_t^2 \) is infinite, model (1.1) is called the infinite variance AR (IVAR) model. This kind of models displaying the features of heavy tails are encountered in several fields, such as tele-traffic engineering in Duffy, et al. (1994), hydrology in Castilo (1988), and economics and finance in Koedijk, et al. (1990) and Janson and de Vries (1991). A comprehensive review and more references can be found in Resnick (1997). The statistical theory of the IVAR model is essentially different from that of AR models with finite variances.

Kanter and Steiger (1974) showed the weak consistency of the least squares estimator (LSE) of \( \lambda \). Furthermore, Haman and Kanter (1977) proved its strong consistency with a convergent rate \( n^{1/\delta} \), where \( n \) is the sample size, \( \delta > \alpha \) and \( \alpha \in (0, 2) \) is the stable index of \( \varepsilon_t \). The limiting distribution of the LSE had not available until Davis and Resnick (1985, 1986). Based on the point processes, they showed that the LSE converges weakly to the ratio of two stable random variables with the rate \( n^{1/\alpha} L_\delta(n) \), where \( L_\delta(n) \) is a slowly varying function. The least absolute deviation estimator (LAE) was considered by Gross and Steiger (1979) and its strong consistency was proved. An and Chen (1982) showed that a convergent rate of the LAE is \( n^{1/\alpha} \). The asymptotic theory of the LAE and M-estimator of \( \lambda \) was completely established by Davis, et al. (1992). They showed that these estimators converge weakly to the minimum of a stochastic process with the rate \( n^{-1} \). Recently, Mikosh, et al. (1995) studied the Whittle estimator for the infinite variance ARMA model and showed that the estimated parameters converge to a function of a sequence of stable random variables. This result was extended by Kokoszka and Taqqu (1996) for the long memory ARFIMA model. All the limiting distributions in these works do not have a close form and hence they cannot be used to do
statistical inference in practice.

How to do statistical inferences for the IVAR model has been a long-term standing open problem. This paper proposes a self-weighted quantile estimation for this model. It is shown that the estimated $\lambda$ is asymptotically normal if the density function of $\varepsilon_t$ and its derivative are uniformly bounded. The Wald test statistic is constructed for linear restrictions on the parameters and it is shown to have non-trivial local powers. Our method and theory basically solve the problem as above. Our results can be extended for a lot of infinite variance time series models, such as ARMA, long memory fractional ARIMA and threshold AR models, and provide a new insight for future research on heavy tailed time series. This paper is organized as follows. Section 2 presents the estimation method and main results and Section 3 reports some simulation results. All the proofs are given in Section 4.

2 Self-weighted Estimation and Main Results

The quantile estimation was first proposed by Koenker and Bassett (1978). It includes the LAE as a special case and has been extensively investigated in the literature, see for examples, Ruppert and Carroll (1980), Bassett and Koenker (1982), Koenker and Bassett (1982), Koenker and D’Orey (1987), and Portnoy and Koenker (1989). In the regression setup, one of advantages of this estimation is that it does not require any moment condition on the errors to obtain the asymptotic normality of the estimated parameters. However, when we used this method to the time series setup, such as in Koul and Saleh (1995), Koenker and Zhao (1996) and Mukherjee (1999), and Ling and McAleer (2003), this advantage is disappeared. This is because the information-type matrix is required to have the finite expectation for using the central limit theorem. In regression models, this matrix is independent of errors. But in time series models such as model (1.1), the finite expectation of the information-type matrix requires the errors to have at least finite variances.

The key point is in the information-type matrix. This motivates us to define the self-weighted quantile estimator (SWE) of $\lambda(\tau) \equiv \lambda(F^{-1}(\tau), 0, \ldots, 0)'$ as

$$
\hat{\lambda}(\tau) = \arg\min_{\lambda \in R^{p+1}} \sum_{t=1}^{n} \frac{1}{w_t} \rho_{\tau}(y_t - X'_{t-1} \lambda),
$$

where $\rho_{\tau}(u) = u[\tau - I(u < 0)]$, $u \in R$, $\tau \in (0, 1)$, $X_t = (1, y_t, \ldots, y_{t-p+1})'$, and $w_t = (1 + \sum_{i=1}^{p} y_{t-i}^2)^{3/2}$. An important special case is when $\tau = 1/2$. We call $\hat{\lambda}_n(0.5)$ the self-weighted least absolute deviation estimator (SWL) of $\lambda(0.5)$. Define

$$
T_n(s, \tau) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{X_{t-1}}{w_t} \left[ I(\varepsilon_t \leq F^{-1}(\tau)) + s' X_{t-1} / \sqrt{n} - \tau \right],
$$

where $s \in R^{p+1}$. $T_n(s, \tau)$ serves as the score function in the maximum likelihood estimation. We can see that the corresponding information-type matrix is bounded.

Our assumption is as follows, which ensures that model (1.1) is strictly stationary and ergodic, see Proposition 13.3.2 in Brockwell and Davis (1996).

**Assumption 2.1** The characteristic polynomial $1 - \phi_1 z - \cdots - \phi_p z^p$ has all roots outside the unit circle and $E[\varepsilon_t]^{\alpha} < \infty$ for some $\alpha > 0$.

Here and in the sequel, $o_p(1)$ denotes a random sequence converging to zero in probability and $\rightarrow_{p}$ denotes convergence in distribution as $n \to \infty$. We now can state our main result as follows.

**Theorem 2.1** If Assumption 2.1 is satisfied and $F(x)$ has a positive density $f(x)$ on $\{ x : 0 < F(x) < 1 \}$ with $\sup_{x \in R} f(x) < \infty$ and $\sup_{x \in R} F'(x) < \infty$, then

$$
\sqrt{n}[\hat{\lambda}_n(\tau) - \lambda(\tau)] = \frac{1}{q(\tau)} T_n(0, \tau) + o_p(1)
$$

$\rightarrow_{p} N\left( 0, \frac{\tau(1 - \tau)}{q^2(\tau)} \Sigma^{-1} \Omega \Sigma^{-1} \right),$

where $q(\tau) = f(F^{-1}(\tau))$, $\Sigma = E(X_{t-1} X'_{t-1} / w_t)$ and $\Omega = E(X_{t-1} X'_{t-1} / w_t)$.

This result is surprising and novel when $E \varepsilon_t^2 = \infty$, compared with those discussed in Section 1. We note that the self-weighted principal can be used for other models and other estimation.
methods. It gives a new way to handle with the heavy tailed time series and will have a large applicable area. From the proof in Section 4, we can see that the weight 1/w_i is not unique. It remains an interesting topic to select an optimal weight such that the asymptotic covariance matrix is minimal.

The covariance matrix Σ and Ω can be estimated by

$$\hat{\Sigma}_n = \frac{1}{n} \sum_{t=1}^{n} \frac{X_{t-1}X'_{t-1}}{w_t},$$

$$\hat{\Omega}_n = \frac{1}{n} \sum_{t=1}^{n} \frac{X_{t-1}X'_{t-1}}{w_t^2},$$

(2.2)

respectively. Using the uniform kernel and the bandwidth b_n = c/n^ν with ν ∈ (0, 1/2) and constant c > 0, we can estimate q(τ) by

$$\hat{q}_n(τ) = \frac{1}{2\sigma_w b_n} \sum_{t=1}^{n} \frac{1}{w_t} I[-b_n + \hat{\lambda}'(τ)X_{t-1}],$$

(2.3)

where σ_w = n^{-1} \sum_{t=1}^{n} (1/w_t). Now, we can do statistical inferences for the IVAR model, such as testing linearity and the goodness-of-fit test. Here, we only consider the Wald test statistic, denoted by W_n, for the p_1 linear hypothesis of the form: H_0 : Rλ(τ) = r, in the usual notation, and give the corollary as follows.

**Corollary 2.1** If Assumptions of Theorem 2.1 holds and b_n = O(1/n^ν) with ν ∈ (0, 1/2), then under H_0, it follows that

$$W_n = \frac{n\hat{q}_n^2(τ)}{\tau(1-\tau)} [R\hat{\lambda}_n(τ) - r]' [R\hat{\Sigma}_{n-1}^{-1} \hat{\Omega}_n \hat{\Sigma}_{n-1}R'] [R\hat{\lambda}_n(τ) - r] \rightarrow_d \chi_{p_1}^2.$$

A natural question is whether or not W_n has local powers. For this, we consider the local alternative hypothesis:

$$H_{1n} : R\lambda_n(τ) = r,$$

where λ_n(τ) = λ(τ) + ν/\sqrt{n} and ν ∈ R^{p+1} is a constant vector. To study the local power, the standard method is to show that the probability measures of (y_1, ..., y_n) under H_0 and H_{1n} are contiguous and then to use Le Cam’s third lemma. We are not sure whether or not the contiguity holds for the IVAR model. Even if yes, it is difficult to prove that in the usual method as in Ling and McAleer (2003). In Section 4, we prove the following result by a direct method. This result implies that W_n has non-trivial local powers.

**Theorem 2.2** If the assumptions of Theorem 2.1 holds, then under H_{1n},

(i) \(\sqrt{n}[\hat{\lambda}_n(τ) - λ(τ)] \rightarrow_d N(ν, τ(1-τ)\Sigma^{-1}ΩΣ^{-1})\),

(ii) \(W_n \rightarrow_d \chi_{p_1}^2(μ)\),

where μ = (q^2(τ)v' (R\Sigma^{-1}Ω)^{-1}Rν/|τ(1-τ)| is a noncentral parameter.

### 3 Simulation Studies

This section examines the performance of the asymptotic results in finite samples through Monte Carlo experiments. Data are generated through the AR(1) model,

$$y_t = φ_0 + φy_{t-1} + ε_t.$$

In all the experiments, we use the optimal bandwidth b_n given in Silverman (1986, p.40) which is automatically searched from the data.

We first study the means and standard deviations of the SWL (a special SWE). The true parameters are taken to be (φ_0, φ) = (0, -0.5), (0, 0.5) and (0, 0.8). Two density functions, Cauchy and t_2, are considered. The sample sizes are n = 200 and n = 400. One thousand replications are used. Table 1 summarizes the empirical means, empirical standard deviations (SD) and asymptotic standard deviations (AD) of the SWLs of (φ_0, φ). The ADs are calculated using the estimated covariances in (2.2). Table 1 shows that all the biases are very small and all the SDs and ADs are very close, particularly, when n = 400. As n is increased from 200 to 400, all the SDs and ADs become smaller.

To give an overall view on the approximation of the limiting distribution to the finite sample distribution, we simulate 27000 replications for the case with φ = 0.5, η_t ~ t_2 and n = 400. Denote \(N_{SW,Ln} = \sqrt{n}[\hat{φ}_n(0.5) - 0.5]/\hat{σ}_{SWL}\), where \(σ_{SWL}\) is the SDs of the SWL of φ. Figure 1 shows the density curves of \(N_{SW,Ln}\) and
$N(0, 1)$. The density curve of $N_{SWLn}$ is approximated by $f(x_i) \approx \sum_{i=1}^{27000} I(x_{i-1} \leq N_{SWLn} \leq x_i)/(27000b)$ with $x_0 = -6.235$, $x_i = x_{i-1} + b$ and $b = 0.215$. From this figure, we can see that the density curve of $N_{SWLn}$ is very close to that of $N(0, 1)$. This is consistent with our theoretical results. These simulation results indicate that the SWL performs very well in the finite samples.

We now investigate the size and power of the statistic $W_n$. Again, the sample sizes are $n = 200$ and $400$ and the number of replications is 1000. Cauchy and $t_2$ distributions are used. The null hypothesis is $H_0$: $((\phi_0, \phi) = (0, 0.5)$ and the significance level is 5%.

Table 2 summarizes the sizes and powers of $W_n$. From this table, we can see that the sizes are a little large, but they are still acceptable. In particular, when $n = 400$, the sizes are getting close to the nominal significance level. The powers are increased when $n$ becomes large or when the distance between the alternative and the null $H_0$ becomes large. These simulation results indicate that the Wald test works well in the finite samples and should be useful in practice.

## 4 Proofs

In what follows, we denote Euclidean norm by $\| \cdot \|$ and a bounded random sequence in probability by $O_p(1)$, and let $F_i = \sigma(\xi_i, \xi_{i-1}, \ldots)$.

**Lemma 4.1** If the assumptions of Theorem 2.1 hold, then it follows that

(i) $\|T_n(\sqrt{n} [\lambda_n(\tau) - \lambda(\tau)], \tau)\| = O_p\left(\frac{1}{\sqrt{n}}\right)$,

(ii) $T_n(0, \tau) \rightarrow_L N(0, \tau(1 - \tau)\Omega)$.

**Proof.** Since $F$ is continuous, for each $t$, there exists no constant vector $c$ with $c \neq 0$ such that $c'X_t = 0$ almost surely (a.s.). Furthermore, note that $\max_{1 \leq t \leq n} \|X_{t-1}\|/w_t \leq 1$ a.s. Exactly following the arguments as for Lemma 4.2 in Ruppert and Carroll (1980), we can show that

$$\|T_n(\sqrt{n} [\lambda_n(\tau) - \lambda(\tau)], \tau)\| \leq 2(p + 1) \max_{1 \leq t \leq n} \frac{\|X_{t-1}\|}{\sqrt{n}w_t} = O_p\left(\frac{1}{\sqrt{n}}\right),$$

i.e. (i) holds. Since $a_t \equiv (X_{t-1}/w_t)[I(\xi_t \leq F^{-1}(\tau)) - \tau]$ is strictly stationary and ergodic with $E(a_t|F_{t-1}) = 0$ and $E(a_t'a_t') = \tau(1 - \tau)\Omega$.

(ii) holds by the central limit theorem. This completes the proof.

**Lemma 4.2** Under the assumptions of Theorem 2.1, for any constant $M \geq 0$,

$$\sup_{\|s\| \leq M} \|T_n(s, \tau) - T_n(0, \tau) - q(\tau)\Sigma s\| = o_p(1).$$

**Proof.** Let $g_t(s, u) = (s'X_{t-1} + u|s'X_{t-1}|)/\sqrt{n}$ with $u \in [0, M]$. We define

$$Z_t(s, u) = I[\xi_t \leq x + g_t(s, u)]$$

$$-I(\xi_t \leq x) - F[x + g_t(s, u)] + F(x),$$

where $x = F^{-1}(\tau)$. By the monotonicity of $F$ and indicator function, it follows that

$$\|Z_t(s, u)\| \leq I(x - \frac{2M}{\sqrt{n}}\|X_{t-1}\| \leq \xi_t \leq x + \frac{2M}{\sqrt{n}}\|X_{t-1}\|)$$

$$+ F(x + \frac{2M}{\sqrt{n}}\|X_{t-1}\|) - F(x - \frac{2M}{\sqrt{n}}\|X_{t-1}\|).$$

Thus, we have

$$E[Z_t^2(s, u)|F_{t-1}] \leq 4\|F(x + \frac{2M}{\sqrt{n}}\|X_{t-1}\|) - F(x - \frac{2M}{\sqrt{n}}\|X_{t-1}\|)\|$$

$$\leq \frac{16M\|X_{t-1}\|}{\sqrt{n}} f(x + n^{-1/2}\xi_{t-1}^*) \leq C\|X_{t-1}\| \sqrt{n},$$

where $-2M\|X_{t-1}\|/\sqrt{n} \leq \xi_{t-1}^* \leq 2M\|X_{t-1}\|/\sqrt{n}$ and $C$ is a constant. Let $\xi_{it}^* = \max\{y_{i-1}, 0\}/w_t$ and $\xi_{it}^- = \max\{-y_{i-1}, 0\}/w_t$. Denote

$$T_{in}^\pm(s, \tau, u) = \frac{1}{\sqrt{n}} \sum_{i=1}^{\infty} \xi_{it}^\pm Z_t(s, u),$$

where $i = 1, \ldots, p + 1$. For any $\epsilon > 0$, since $\xi_{it}^\pm Z_t(s, u)$ is a martingale difference in terms of $F_t$, by Markov’s inequality, we have

$$P(\|T_{in}^\pm(s, \tau, u)\| \geq \epsilon) \leq \frac{1}{n\epsilon^2} \sum_{i=1}^{\infty} E(\xi_{it}^\pm Z_t(s, u))^2$$

$$\leq \frac{C^2}{n^2\epsilon^2} \sum_{i=1}^{\infty} E\left(\frac{\|X_{t-1}\|}{w_t^2}\right) \to 0,$$

as $n \to \infty$, for each $s \in R^p$ and $u \in R$, where $i = 1, \ldots, p + 1$.

Denote $D_M = [-M, M]^{p+1}$. Since $D_M$ is a bounded and closed region of $R^{p+1}$, for every
\( \delta > 0 \), there is a finite number of open subsets \( \Delta_i(\delta) \), \( i = 1, \ldots, m \), each with diameter \( \delta \), such that \( \bigcup_{i=1}^{m} \Delta_i(\delta) \supset D_M \) and \( \Delta_i \equiv \Delta_i(\delta) \cap D_M \) is not empty. Let \( s_r \) be any fixed point in \( \Delta_r \). Then for any \( u \in \Delta_r \), we know that

\[
|g_t(s, u) - g_t(s_r, u)| \\
\leq \|s - s_r\| \cdot \|X_{t-1}\| / \sqrt{n} \leq \delta \|X_{t-1}\| / \sqrt{n},
\]

that is, \( g_t(s_r, u - \delta) \leq g_t(s, u) \leq g_t(s_r, u + \delta) \). By the monotonicity of the indicator function, we obtain that

\[
T^+_n(s, \tau, 0) \leq T^+_n(s_r, \tau, \delta) + \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \varepsilon_t^+ [F(x + g_t(s_r, \delta)) - F(x + g_t(s_r, 0))]
\]

and a reverse inequality holds as \( \delta \) is replaced by \( -\delta \).

By the assumption given and the mean value theorem, it follows that

\[
\left| \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \varepsilon_t^+ [F(x + g_t(s_r, \pm \delta))] - F(x + g_t(s, 0)) \right| \\
\leq \sup_x |f(x)| \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \varepsilon_t^+ |g_t(s_r, \pm \delta) - g_t(s_r, 0)|
\]

(4.2) \( \leq 2\delta \sup_x |f(x)| \sqrt{n} \sum_{t=1}^{n} \frac{\|X_{t-1}\|^2}{w_t} = \delta O_p(1), \)

where \( O_p(1) \) uniformly holds for all \( s \in \Delta_r \) and all \( r = 1, \ldots, m \). Given any small \( \varepsilon > 0 \) and \( \eta > 0 \), by (4.2), there exists a \( \delta_x > 0 \) such that

\[
P\left\{ \left| \frac{1}{\sqrt{n}} \sup_{s \in \Delta_r} \sum_{t=1}^{n} [F(x + g_t(s_r, \pm \delta)) - F(x + g_t(s, 0))] \right| \geq \frac{\varepsilon}{3} \right\} \leq \eta.
\]

For the \( \pm \delta_x \), by (4.1), it follows that

\[
P\left\{ \max_r \left| T^+_n(s_r, \tau, \pm \delta_x) \right| \geq \frac{\varepsilon}{3} \right\} \\
\leq \max_r P\left\{ \left| T^+_n(s_r, \tau, \pm \delta_x) \right| \geq \frac{\varepsilon}{3} \right\} \leq \eta,
\]

as \( n \) is large enough. By (4.3)-(4.4), we know that

\[
P\left\{ \sup_{s \in D_M} \left| T^+_n(s, \tau, 0) \right| \geq \varepsilon \right\} \\
\leq P\left\{ \max_r \left| T^+_n(s_r, \tau, \delta_x) \right| \geq \frac{\varepsilon}{3} \right\}.
\]

Proof of Theorem 2.1. Denote \( \Upsilon_n(\tau) = \sqrt{n} \lambda_n(\tau) - \lambda(\tau) \). For any \( \varepsilon, \eta > 0 \), by Lemma 4.1 (i), there exists an integer \( n_1 > 0 \) such that, when \( n > n_1 \),

\[
P\left\{ \| \Upsilon_n(\tau) \| > \eta \right\} < \varepsilon.
\]

Thus, for a positive constant \( M \), when \( n > n_1 \),

\[
P\{ \| \Upsilon_n(\tau) \| > M \} \\
\leq P \{ \| \Upsilon_n(\tau) \| > M, \| \Upsilon_n(\tau) \| \leq \eta \} \\
+ P \{ \| \Upsilon_n(\tau) \| > \eta \}
\]

(4.8) \( \leq P \left\{ \inf_{\|s\| \geq M} \| T_n(s, \tau, \tau) \| \leq \eta \right\} + \varepsilon.
\]

Note that \( s_1^1 T_n(\nu s_1, \tau) \) is a non-decreasing function of \( \nu \) for any \( \tau \in (0, 1) \) and \( s_1 \in \mathbb{R}^{d+1} \). Writing \( s_1 \) as \( s_1 = \nu s \) with \( \nu \geq 1 \) and \( \|s\| = M \) for any \( \|s\| \geq M \), by the Cauchy-Schwarz inequality, we have

\[
\inf_{\|s\| = M} \| s_1 T_n(s, \tau) \| \leq \inf_{\|s\| = M, \nu \geq 1} \| s_1 T_n(\nu s, \tau) \| \leq M \inf_{\|s\| \geq M} \| T_n(s_1, \tau) \|.
\]
Thus, by (4.8),

\[
(4.9) \quad P\left\{ \| \Upsilon_n(\tau) \| \geq M \right\} \leq P\left\{ \inf_{\| s \| = M} |s^T T_n(s, \tau)| \right. \\
\left. \quad \leq \eta M \right\} + \varepsilon.
\]

Denote \( R_n(\tau) = \sup_{\| s \| = M} |s^T T_n(s, \tau) - T_n(0, \tau)| - s^T \Sigma s q(\tau) \) and let \( c_0 \) be the minimum eigenvalue of \( \Sigma \). Since

\[
|s^T T_n(s, \tau)| \geq \inf_{\| s \| = M} |s^T \Sigma s q(\tau) - R_n(\tau) | \\
\leq - \sup_{\| s \| = M} |s^T T_n(0, \tau)| - \eta M,
\]

by (4.9), it follows that

\[
(4.10) \quad P\left\{ \| \Upsilon_n(\tau) \| \geq M \right\} \\
\leq P\left\{ R_n(\tau) \geq - \sup_{\| s \| = M} \left| s^T T_n(0, \tau) \right| - \eta M + c_0 M^2 q(\tau) \right\} + \varepsilon.
\]

By Lemma 4.1 (ii), there exists a large constant \( M_1 \) and an integer \( n_2 \) such that, when \( n > n_2 \),

\[
(4.11) \quad P\left( \sup_{\| s \| = M} |s^T T_n(0, \tau)| > N M M_1 \right) \\
\leq P\left( \| T_n(0, \tau) \| > M_1 \right) < \varepsilon.
\]

Thus, by (4.11), when \( n > \max\{n_2, n_3\} \),

\[
(4.12) \quad P\left\{ R_n(\tau) \geq - \sup_{\| s \| = M} |s^T T_n(0, \tau)| \right. \\
\left. \quad - \eta M + c_0 M^2 q(\tau), \sup_{\| s \| = M} |s^T T_n(0, \tau)| \leq M M_1 \right\} \\
+ P\left( \sup_{\| s \| = M} |s^T T_n(0, \tau)| > M M_1 \right) \\
\leq P\left\{ R_n(\tau) \geq c_0 M^2 q(\tau) - M M_1 - \eta M \right\} + \varepsilon.
\]

We may choose \( M \) large enough such that \( c = c_0 M q(\tau) - M_1 - \eta > 0 \). For the constant \( c \), by Lemma 4.2, there exists an integer \( n_3 \) such that, when \( n > n_3 \),

\[
(4.13) \quad P\left\{ R_n(\tau) \geq M \right\} \\
\leq P\left\{ \sup_{\| s \| = M} \left| T_n(s, \tau) - T_n(0, \tau) \right| - q(\tau) \Sigma s \right\} \geq c \right\} < \varepsilon.
\]

Thus, by (4.10) and (4.12)-(4.13), when \( n > \max\{n_1, n_2, n_3\} \), \( P\{ \| \Upsilon_n(\tau) \| \geq M \} < 4 \varepsilon \). Finally, by Lemma 4.1(i) and 4.2, we can show that

\[
\sqrt{n} \hat{\Lambda}_n(\tau) - \lambda(\tau) = -\frac{\Sigma^{-1}}{q(\tau)} T_n(0, \tau) + o_p(1).
\]

Furthermore, by Lemma 4.1(ii) and the equation above, the conclusion holds. This completes the proof. \( \square \)

**Proof of Corollary 2.1.** From the proofs of Lemmas 4.1 and 4.2, we know that \( \hat{\Sigma}_n = \Sigma + o_p(1) \) and \( \hat{\Omega}_n = \Omega + o_p(1) \). Let \( \hat{\theta}_n(\tau) = \hat{\lambda}_n(\tau) - \lambda(\tau) \). Then

\[
A_n = \frac{1}{2 nb_n} \sum_{t=1}^{n} \frac{1}{w_t} \\
\left\{ I \left[ -b_n + \hat{\theta}_n(\tau) X_{t-1} + x \leq \varepsilon_t \right] \right. \\
\left. \leq x + \hat{\theta}_n(\tau) X_{t-1} + b_n \right\} \\
(4.14) \quad -F[x + \hat{\theta}_n(\tau) X_{t-1} + b_n] \\
+ F[x + \hat{\theta}_n(\tau) X_{t-1} - b_n]\right\} = o_p(1),
\]

where \( x = F^{-1}(\tau) \). In fact, since each term in the summation in (4.14) is a martingale difference in terms of \( \mathcal{F}_t \), for any \( \varepsilon > 0 \), by Markov’s inequality, we have

\[
P(|A_n| \geq \varepsilon) \leq \frac{1}{4 n^2 b_n \varepsilon^2} \sum_{t=1}^{n} E \left[ I \left[ -b_n + \hat{\theta}_n(\tau) X_{t-1} \right] \right. \\
\left. + x \leq \varepsilon_t \leq x + \hat{\theta}_n(\tau) X_{t-1} + b_n \right] \\
-F[x + \hat{\theta}_n(\tau) X_{t-1} + b_n] \\
+ F[x + \hat{\theta}_n(\tau) X_{t-1} - b_n]\right]^2 \leq \frac{1}{4 nb_n^2 \varepsilon^2} \to 0.
\]

Since \( \sqrt{n} \hat{\theta}_n(\tau) = O_p(1) \), by Taylor’s expansion, it follows that

\[
\frac{1}{2 nb_n} \sum_{t=1}^{n} \frac{1}{w_t} \left| F[x + \hat{\theta}_n(\tau) X_{t-1} \pm b_n] - F(x) \right|
\]
\[-f(x)[\hat{\theta}_n'(\tau)X_{t-1} \pm b_n]\]
\[\leq \frac{1}{nb_n} \sum_{i=1}^{n} \frac{1}{w_t}[f''(\xi^*_t)[\hat{\theta}_n'(\tau)X_{t-1} \pm b_n]^2]
\[\leq O(\frac{1}{nb_n}) \sum_{i=1}^{n} \frac{1}{w_t}[\hat{\theta}_n'(\tau)X_{t-1} \pm b_n]^2\]
\[\leq O(\frac{\hat{\theta}_n'(\tau)}{nb_n})^2 \sum_{i=1}^{n} \frac{\|X_{t-1}\|^2}{w_t}\]
\[+ O(\frac{b_n}{n}) \sum_{i=1}^{n} \frac{1}{w_t} = O_p(\frac{1}{nb_n}) + O_p(b_n) = O_p(1),\]

where \(\xi^*_t\) lies between \(x\) and \(x + \hat{\theta}_n'(\tau)X_{t-1} \pm b_n\).

Thus,
\[
\frac{1}{2nb_n} \sum_{i=1}^{n} \frac{1}{w_t} \left\{ F[x + \hat{\theta}_n'(\tau)X_{t-1} + b_n]ight.
\[\left.- F[x + \hat{\theta}_n'(\tau)X_{t-1} - b_n]\right\} = q(\tau)\sigma_w + O_p(1).
\]

Furthermore, by (4.14), we can readily show that \(q_n(\tau) = q(\tau) + O_p(1)\). Finally, by Theorem 2.1, the conclusion holds. This completes the proof. \(\square\)

**Proof of Theorem 2.2.** First, we note that \(y_t\) under \(H_{1n}\) depends on \(n\). To emphasize this, we denote \(y_t\) by \(y_{nt}\) under \(H_{1n}\), \(y_{nt}\) is a function of \(n, \lambda, \nu\) and \(\{\xi_t\}\). When \(\nu = 0\), \(y_{nt} = y_t\). Here, \(y_t\) comes from model (1.1) under \(H_0\). It is easy to see that \(y_{nt} \rightarrow y_t\) a.s. when \(n \rightarrow \infty\).

Similarly define \(w_{nt}\) and \(X_{nt}\). Now, under \(H_{1n}\),
\[
\lambda_n(\tau) = \arg\min_{\lambda \in \mathbb{R}^{p+1}} \sum_{i=1}^{n} \frac{1}{w_{nt}} \rho_\tau(y_{nt} - X_{nt-1} - \lambda).
\]

Define \(\tilde{T}_n(s, \tau) = \sum_{i=1}^{n} X_{nt-1}I(e_t \leq F^{-1}(\tau) + s'X_{nt-1}/\sqrt{n} - \tau)/\sqrt{n}\), where \(s \in \mathbb{R}^{p+1}\). As for Lemma 4.1(i), we can show that
\[
(4.15) \left\| \tilde{T}_n(\sqrt{n}[\hat{\lambda}_n(\tau) - \lambda_n(\tau)], \tau) \right\|
\[\leq 2(p + 1) \max_{1 \leq t \leq n} \frac{\|X_{nt-1}\|}{\sqrt{w_{nt}}} = O_p\left(\frac{1}{\sqrt{n}}\right).
\]

Let \(a_{nt} = (X_{nt-1}/w_{nt})I(e_t \leq F^{-1}(\tau) - \tau)\) and \(a_t = (X_{t-1}/w_{t})I(e_t \leq F^{-1}(\tau) - \tau)\). By Markov’s inequality, the dominated convergence theorem and the ergodic theorem, we can show that
\[
(4.16) \frac{1}{n} \sum_{i=1}^{n} a_{nt}a'_{nt} = \tau(1 - \tau)\Omega
\]
\[
\frac{1}{n} \sum_{i=1}^{n} E(a_{nt}a'_{nt}|F_{t-1}) = \tau(1 - \tau)\Omega.
\]

Since \(a_{nt}\) is a martingale difference in terms of \(F_t\), by the central limit theorem for martingale differences and (4.16), it follows that
\[
(4.17) \tilde{T}_n(0, \tau) \longrightarrow_{\mathcal{L}} N(0, \tau(1 - \tau)\Omega).
\]

Using (4.15)-(4.17) and a similar method as for Theorem 2.1, we can show that (i) holds. Furthermore, we can show that \(\hat{\Sigma}_n = \Sigma + o_p(1), \hat{\Omega}_n = \Omega + o_p(1)\) and \(\hat{q}_n(\tau) = q(\tau) + o_p(1)\) under \(H_{1n}\). By (i) of this theorem and note that
\[
\sqrt{n}[\hat{\lambda}_n(\tau) - \lambda_n(\tau)] = \sqrt{n}[\lambda_n(\tau) - \lambda(\tau)] + \nu,
\]

it is straightforward to show that (ii) holds, see also the proof of Theorem 6 in Weiss (1991). This completes the proof.

**REFERENCES**


**SHIQING LING** is Assistant Professor of Mathematics in the Department of Mathematics, Hong Kong University of Science and Technology. His research interests are in time series econometrics, financial econometrics, and statistics.
### TABLE 1
Means and Standard Deviations of SWL for AR Models with $\phi_0 = 0$ (1000 replications)

<table>
<thead>
<tr>
<th>$\phi$</th>
<th>$n=200$</th>
<th>$n=400$</th>
<th>$n=200$</th>
<th>$n=400$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\phi_0$</td>
<td>$\phi_0$</td>
<td>$\phi_0$</td>
<td>$\phi_0$</td>
</tr>
<tr>
<td>$\varepsilon_t \sim \text{Cauchy}$</td>
<td>$\varepsilon_t \sim t_2$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-0.5</td>
<td>Mean</td>
<td>-.505</td>
<td>.001</td>
<td>-.503</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>.134</td>
<td>.103</td>
<td>.098</td>
</tr>
<tr>
<td></td>
<td>AD</td>
<td>.139</td>
<td>.101</td>
<td>.098</td>
</tr>
<tr>
<td>0.5</td>
<td>Mean</td>
<td>-.007</td>
<td>.491</td>
<td>-.008</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>.136</td>
<td>.107</td>
<td>.094</td>
</tr>
<tr>
<td></td>
<td>AD</td>
<td>.139</td>
<td>.102</td>
<td>.098</td>
</tr>
<tr>
<td>0.8</td>
<td>Mean</td>
<td>-.014</td>
<td>.787</td>
<td>-.014</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>.155</td>
<td>.092</td>
<td>.116</td>
</tr>
<tr>
<td></td>
<td>AD</td>
<td>.165</td>
<td>.085</td>
<td>.117</td>
</tr>
</tbody>
</table>

### TABLE 2
Sizes and Powers of Wald-test for Null Hypothesis $H_0: (\phi_0, \phi) = (0, 0.5)$ at Significance Level 5% in AR Models (1000 replications)

<table>
<thead>
<tr>
<th>$\varepsilon_t \sim$</th>
<th>$n=200$</th>
<th></th>
<th>$n=400$</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cauchy</td>
<td>$t_2$</td>
<td>Cauchy</td>
<td>$t_2$</td>
</tr>
<tr>
<td>$\phi_0$ $\phi$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-.1 .3</td>
<td>.453</td>
<td>.437</td>
<td>.742</td>
<td>.732</td>
</tr>
<tr>
<td>.0 .3</td>
<td>.416</td>
<td>.371</td>
<td>.698</td>
<td>.676</td>
</tr>
<tr>
<td>-.1 .4</td>
<td>.196</td>
<td>.189</td>
<td>.321</td>
<td>.328</td>
</tr>
<tr>
<td>.0 .4</td>
<td>.152</td>
<td>.152</td>
<td>.232</td>
<td>.212</td>
</tr>
<tr>
<td>.0 .5</td>
<td>.066</td>
<td>.066</td>
<td>.060</td>
<td>.057</td>
</tr>
<tr>
<td>.0 .6</td>
<td>.150</td>
<td>.156</td>
<td>.228</td>
<td>.209</td>
</tr>
<tr>
<td>.1 .6</td>
<td>.199</td>
<td>.214</td>
<td>.350</td>
<td>.339</td>
</tr>
<tr>
<td>.0 .7</td>
<td>.416</td>
<td>.407</td>
<td>.699</td>
<td>.661</td>
</tr>
<tr>
<td>.1 .7</td>
<td>.472</td>
<td>.465</td>
<td>.772</td>
<td>.742</td>
</tr>
</tbody>
</table>

Figure 1: Density Curves of $N_{SWL_{LR}}$ and $N(0, 1)$: ‘circles’ and ‘dots’, respectively