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:

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

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 teletraffic 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 λ. Furthermore, Hannan 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 ε_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 $n^{1/\alpha}L_\nu(n)$, where $L_\nu(n)$ 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 $n^{1/\delta}$. The asymptotic theory of the LAE and M-estimator of λ 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 $a_n = \inf\{x{:}P(|\varepsilon_t| > x) \le n^{-1}\}$. Recently, Mikosch, 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 λ is asymptotically normal if the density function of ε_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 Muk -herjee (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 $\lambda(\tau) \equiv \lambda + (F^{-1}(\tau), 0, \cdots, 0)'$ as

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

where $\rho_\tau(u) = u[\tau - I(u < 0)]$, $u \in R$, $\tau \in (0,1)$, $X_t = (1, y_t, \cdots, y_{t-p+1})'$, 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

$$T_n(s,\tau) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{X_{t-1}}{w_t} [I(\varepsilon_t \leq F^{-1}(\tau) + s'X_{t-1}/\sqrt{n}) - \tau], \tag{2.1}$$

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 $\to_{\mathcal{L}}$ 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*

$$\sqrt{n}[\hat{\lambda}_n(\tau) - \lambda(\tau)] = -\frac{\Sigma^{-1}}{q(\tau)} T_n(0,\tau) + o_p(1) \longrightarrow_{\mathcal{L}} N\Big(0, \frac{\tau(1-\tau)}{q^2(\tau)} \Sigma^{-1}\Omega\Sigma^{-1}\Big),$$

where $q(\tau) = f(F^{-1}(\tau))$, $\Sigma = E(X_{t-1}X_{t-1}'/w_t)$ and $\Omega = E(X_{t-1}X_{t-1}'/w_t^2)$.

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 $1/w_t$ 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

$$\hat{\Sigma}_n = \frac{1}{n}\sum_{t=1}^{n}\frac{X_{t-1}X_{t-1}'}{w_t},$$
$$\hat{\Omega}_n = \frac{1}{n}\sum_{t=1}^{n}\frac{X_{t-1}X_{t-1}'}{w_t^2}, \tag{2.2}$$

respectively. Using the uniform kernel and the bandwidth $b_n = c/n^\nu$ with $\nu \in (0, 1/2)$ and constant $c > 0$, we can estimate $q(\tau)$ by

$$\hat{q}_n(\tau) = \frac{1}{2\hat{\sigma}_w n b_n}\sum_{t=1}^{n}\frac{1}{w_t}I\Big[-b_n + \hat{\lambda}_n'(\tau)X_{t-1} \le y_t \le \hat{\lambda}_n'(\tau)X_{t-1} + b_n\Big], \tag{2.3}$$

where $\hat{\sigma}_w = n^{-1}\sum_{t=1}^{n}(1/w_t)$. 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: $H_0 : R\lambda(\tau) = r$, in the usual notation, and give the corollary as follows.

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

$$W_n = \frac{n\hat{q}_n^2(\tau)}{\tau(1-\tau)}[R\hat{\lambda}_n(\tau) - r]'\Big[R\hat{\Sigma}_n^{-1}\hat{\Omega}_n\hat{\Sigma}_n^{-1}R'\Big]^{-1}[R\hat{\lambda}_n(\tau) - r] \longrightarrow_{\mathcal{L}} \chi^2_{p_1}.$$

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

$$H_{1n} : R\lambda_n(\tau) = r,$$

where $\lambda_n(\tau) = \lambda(\tau) + \nu/\sqrt{n}$ and $\nu \in R^{p+1}$ is a constant vector. To study the local power, the standard method is to show that the probability measures of $(y_1, \cdots, y_n)$ 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*) $$\sqrt{n}[\hat{\lambda}_n(\tau) - \lambda(\tau)] \longrightarrow_{\mathcal{L}} N\Big(\nu, \frac{\tau(1-\tau)}{q^2(\tau)}\Sigma^{-1}\Omega\Sigma^{-1}\Big),$$

(*ii*) $$W_n \longrightarrow_{\mathcal{L}} \chi^2_{p_1}(\mu),$$

where $\mu = q^2(\tau)\nu'R'(R\Sigma^{-1}\Omega\Sigma^{-1}R')^{-1}R\nu/[\tau(1-\tau)]$ 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,

$$y_t = \phi_0 + \phi y_{t-1} + \varepsilon_t.$$

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 $(\phi_0, \phi) = (0, -0.5)$, $(0, 0.5)$ 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 $\phi = 0.5$, $\eta_t \sim t_2$ and $n = 400$. Denote $N_{SWLn} = \sqrt{n}[\hat{\phi}_n(0.5) - 05]/\hat{\sigma}_{SWL}$, where $\hat{\sigma}_{SWL}$ is the SDs of the SWL of ϕ. Figure 1 shows the density curves of N_{SWLn} and

$N(0,1)$. The density curve of N_{SWLn} is approximated by $f(x_i) \approx \sum_{i=1}^{27000} I(x_{i-1} \le N_{SWLn} \le x_i)/(27000b)$ 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 signi- 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 $\mathcal{F}_t = \sigma\{\varepsilon_t, \varepsilon_{t-1}, \cdots\}$.

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

$$(i) \qquad \|T_n(\sqrt{n}[\hat{\lambda}_n(\tau) - \lambda(\tau)], \tau)\| = O_p(\frac{1}{\sqrt{n}}),$$

$$(ii) \qquad T_n(0,\tau) \longrightarrow_{\mathcal{L}} N(0, \tau(1-\tau)\Omega).$$

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

$$\|T_n(\sqrt{n}[\hat{\lambda}_n(\tau) - \lambda(\tau)], \tau)\| \le 2(p+1) \cdot \max_{1\le t\le n} \frac{\|X_{t-1}\|}{\sqrt{n} w_t} = O_p(\frac{1}{\sqrt{n}}),$$

i.e. (i) holds. Since $a_t \equiv (X_{t-1}/w_t)[I(\varepsilon_t \le F^{-1}(\tau)) - \tau]$ is strictly stationary and ergodic with $E(a_t|\mathcal{F}_{t-1}) = 0$ 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 $M \ge 0$,*

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

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

$$Z_t(s,u) = I[\varepsilon_t \le x + g_t(s,u)] - I(\varepsilon_t \le x) - F[x + g_t(s,u)] + F(x),$$

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

$$\begin{aligned} &|Z_t(s,u)| \\ &\le I(x - \frac{2M}{\sqrt{n}}\|X_{t-1}\| \le \varepsilon_t \le x + \frac{2M}{\sqrt{n}}\|X_{t-1}\|) \\ &+F(x + \frac{2M}{\sqrt{n}}\|X_{t-1}\|) - F(x - \frac{2M}{\sqrt{n}}\|X_{t-1}\|). \end{aligned}$$

Thus, we have

$$\begin{aligned} &E[Z_t^2(s,u)|\mathcal{F}_{t-1}] \\ &\le 4[F(x + \frac{2M}{\sqrt{n}}\|X_{t-1}\|) - F(x - \frac{2M}{\sqrt{n}}\|X_{t-1}\|)] \\ &\le \frac{16M\|X_{t-1}\|}{\sqrt{n}} f(x + n^{-1/2}\xi_{t-1}^*) \le \frac{C\|X_{t-1}\|}{\sqrt{n}}, \end{aligned}$$

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

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

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

$$\begin{aligned} &P(\Big|T_{in}^{\pm}(s,\tau,u)\Big| \ge \epsilon) \le \frac{1}{n\epsilon^2} \sum_{t=1}^n E[\xi_{it}^{\pm} Z_t(s,u)]^2 \\ (4.1) \quad &\le \frac{C^2}{n^2\epsilon^2} \sum_{t=1}^n E\Big(\frac{\|X_{t-1}\|^4}{w_t^2}\Big) \to 0, \end{aligned}$$

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

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 $\Delta_i(\delta)$, $i = 1, \cdots, m$, each with diameter δ, such that $\bigcup_{i=1}^m \Delta_i(\delta) \supset D_M$ and $\tilde{\Delta}_i \equiv \Delta_i(\delta) \cap D_M$ is not empty . Let s_r be any fixed point in $\tilde{\Delta}_r$. Then for any $u \in \tilde{\Delta}_r$, 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

$$T_{in}^{\pm}(s,\tau,0) \leq T_{in}^{\pm}(s_r,\tau,\delta) + \frac{1}{\sqrt{n}} \cdot \sum_{t=1}^n \xi_{it}^{\pm}[F(x+g_t(s_r,\delta)) - F(x+g_t(s,0))]$$

and a reverse inequality holds as δ is replaced by $-\delta$.

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

$$\Big|\frac{1}{\sqrt{n}}\sum_{t=1}^n \xi_{it}^{\pm}[F(x+g_t(s_r,\pm\delta)) - F(x+g_t(s,0))]\Big| \\ \leq \sup_x |f(x)| \frac{1}{\sqrt{n}}\sum_{t=1}^n \xi_{it}^{\pm}|g_t(s_r,\pm\delta) - g_t(s,0)| \\ \leq \frac{2\delta \sup_x |f(x)|}{n}\sum_{t=1}^n \frac{\|X_{t-1}\|^2}{w_t} = \delta O_p(1), \tag{4.2}$$

where $O_p(1)$ uniformly holds for all $s \in \tilde{\Delta}_r$ and all $r = 1, \cdots, m$. Given any small $\varepsilon > 0$ and $\eta > 0$, by (4.2), there exists a $\delta_\varepsilon > 0$ such that

$$P\{\frac{1}{\sqrt{n}}\sup_r \sup_{s\in\tilde{\Delta}_r}\Big|\sum_{t=1}^n [F(x+g_t(s_r,\pm\delta_\varepsilon)) - F(x+g_t(s,0))]\Big| \geq \frac{\varepsilon}{3}\} \leq \eta\,. \tag{4.3}$$

For the $\pm\delta_\varepsilon$, by (4.1), it follows that

$$P\{\max_r |T_{in}^{\pm}(s_r,\tau,\pm\delta_\varepsilon)| \geq \frac{\varepsilon}{3}\} \\ \leq r\max_r P\{|T_{in}(s_r,\tau,\pm\delta_\varepsilon)| \geq \frac{\varepsilon}{3}\} \leq \eta, \tag{4.4}$$

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

$$P\{\sup_{s\in D_M} |T_{in}^{\pm}(s,\tau,0)| \geq \varepsilon\} \\ \leq P\{\max_r |T_{in}^{\pm}(s_r,\tau,\delta_\varepsilon)| \geq \frac{\varepsilon}{3}\} \\ +P\{\max_r |T_{in}^{\pm}(s_r,\tau,-\delta_\varepsilon)| \geq \frac{\varepsilon}{3}\} \\ +P\{\frac{1}{\sqrt{n}}\sup_r \sup_{s\in\tilde{\Delta}_r} |\sum_{t=1}^n [F(x+g_t(s_r,\pm\delta_\varepsilon)) - F(x+g_t(s,0))]| \geq \frac{\varepsilon}{3}\} \\ \leq 3\eta. \tag{4.5}$$

By (4.5), we can show that

$$\sup_{\|s\|\leq M}\Big|\frac{1}{\sqrt{n}}\sum_{t=1}^n \frac{X_{t-1}}{w_t} Z_t(s,0)\Big| = o_p(1). \tag{4.6}$$

Furthermore, by Taylor's expansion and the assumption given, we have

$$\sup_{s\in D_M}\Big|\frac{1}{\sqrt{n}}\sum_{t=1}^n \frac{X_{t-1}}{w_t}\Big[F(x+\frac{1}{\sqrt{n}}s'X_{t-1}) - F(x) - \frac{f(x)}{\sqrt{n}}s'X_{t-1}\Big]\Big| \\ \leq \sup_{s\in D_M} f'(\xi_t^*)\frac{M^2}{n\sqrt{n}}\sum_{t=1}^n \frac{\|X_{t-1}\|^3}{w_t} \\ = O_p(\frac{1}{\sqrt{n}}). \tag{4.7}$$

By the ergodic theorem, $\sum_{t=1}^n (X_{t-1}X'_{t-1}/w_t)/n = \Sigma + o_p(1)$. Furthermore, by (4.6) and (4.7), we can claim that the conclusion holds. This completes the proof. □

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

$$P\Big\{\|T_n(\Upsilon_n(\tau),\tau)\| > \eta\Big\} < \varepsilon.$$

Thus, for a positive constant M, when $n > n_1$,

$$P\{\|\Upsilon_n(\tau)\| \geq M\} \\ \leq P\{\|\Upsilon_n(\tau)\| \geq M, \|T_n(\Upsilon_n(\tau),\tau)\| \leq \eta\} \\ +P\{\|T_n(\Upsilon_n(\tau),\tau)\| \geq \eta\} \\ \leq P\Big\{\inf_{\|s_1\|\geq M}\|T_n(s_1,\tau)\| \leq \eta\Big\} + \varepsilon. \tag{4.8}$$

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

$$\inf_{\|s\|=M}|s'T_n(s,\tau)| \leq \inf_{\|s\|=M,\nu\geq 1}|s'T_n(\nu s,\tau)| \\ \leq M\inf_{\|s_1\|\geq M}\|T_n(s_1,\tau)\|.$$

Thus, by (4.8),

$$P\{\|\Upsilon_n(\tau)\| \geq M\} \leq P\Big\{\inf_{\|s\|=M} |s' T_n(s,\tau)| \leq\leq \eta M\Big\} + \varepsilon. \tag{4.9}$$

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

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

by (4.9), it follows that

$$\begin{aligned} &P\Big\{\|\Upsilon_n(\tau)\| \geq M\Big\} \\ &\quad \leq P\Big\{R_n(\tau) \geq \inf_{\|s\|=M}[s'\Sigma s q(\tau)] \\ &- \sup_{\|s\|=M} |s' T_n(0,\tau)| - \eta M\Big\} + \varepsilon \\ &\quad \leq P\Big\{R_n(\tau) \geq - \sup_{\|s\|=M} |s' T_n(0,\tau)| \\ &-\eta M + c_0 M^2 q(\tau)]\Big\} + \varepsilon. \end{aligned} \tag{4.10}$$

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

$$\begin{aligned} &P(\sup_{\|s\|=M} |s' T_n(0,\tau)| > M M_1) \\ &\leq P(\|T_n(0,\tau)\| > M_1) < \varepsilon. \end{aligned} \tag{4.11}$$

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

$$\begin{aligned} &P\Big\{R_n(\tau) \geq - \sup_{\|s\|=M} |s' T_n(0,\tau)| \\ &-\eta M + c_0 M^2 q(\tau)\Big\} \\ &\quad \leq P\Big\{R_n(\tau) \geq - \sup_{\|s\|=M} |s' T_n(0,\tau)| \\ &-\eta M + c_0 M^2 q(\tau), \\ &\qquad \sup_{\|s\|=M} |s' T_n(0,\tau)| \leq M M_1\Big\} \\ &+P\Big(\sup_{\|s\|=M} |s' T_n(0,\tau)| > M M_1\Big) \\ &\quad \leq P\Big\{R_n(\tau) \geq c_0 M^2 q(\tau) - M M_1 - \\ &\eta M\Big\} + \varepsilon. \end{aligned} \tag{4.12}$$

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 $n > n_3$,

$$\begin{aligned} &P\Big\{R_n(\tau) \geq Mc\Big\} \\ &\leq P\Big\{\sup_{\|s\|=M} \|[T_n(s,\tau) - T_n(0,\tau)] \\ &-q(\tau)\Sigma s]\| \geq c\Big\} < \varepsilon. \end{aligned} \tag{4.13}$$

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. □

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

$$\begin{aligned} A_n &\equiv \frac{1}{2nb_n} \sum_{t=1}^{n} \frac{1}{w_t} \\ &\Big\{I\Big[- b_n + \hat{\theta}_n'(\tau) X_{t-1} + x \leq \varepsilon_t \\ &\leq x + \hat{\theta}_n'(\tau) X_{t-1} + b_n\Big] \\ &\quad -F[x + \hat{\theta}_n'(\tau) X_{t-1} + b_n\Big] \\ &+F[x + \hat{\theta}_n'(\tau) X_{t-1} - b_n]\Big\} = o_p(1), \end{aligned} \tag{4.14}$$

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

$$\begin{aligned} &P(|A_n| \geq \varepsilon) \leq \frac{1}{4n^2 b_n^2 \varepsilon^2} \sum_{t=1}^{n} E\{I[-b_n + \hat{\theta}_n'(\tau) X_{t-1} \\ &+x \leq \varepsilon_t \leq x + \hat{\theta}_n'(\tau) X_{t-1} + b_n] \\ &\qquad -F[x + \hat{\theta}_n'(\tau) X_{t-1} + b_n] \\ &+F[x + \hat{\theta}_n'(\tau) X_{t-1} - b_n]\}^2 \leq \frac{1}{4nb_n^2 \varepsilon^2} \to 0. \end{aligned}$$

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

$$\frac{1}{2nb_n} \sum_{t=1}^{n} \frac{1}{w_t} \Big| F[x + \hat{\theta}_n'(\tau) X_{t-1} \pm b_n] - F(x)$$

$$-f(x)[\hat{\theta}_n'(\tau)X_{t-1} \pm b_n]\Big|$$

$$\leq \frac{1}{nb_n}\sum_{t=1}^n \frac{1}{w_t}|f''(\xi_t^*)|[\hat{\theta}_n'(\tau)X_{t-1} \pm b_n]^2$$

$$\leq O(\frac{1}{nb_n})\sum_{t=1}^n \frac{1}{w_t}[\hat{\theta}_n'(\tau)X_{t-1} \pm b_n]^2$$

$$\leq O(\frac{\|\hat{\theta}_n(\tau)\|^2}{nb_n})\sum_{t=1}^n \frac{\|X_{t-1}\|^2}{w_t}$$

$$+O(\frac{b_n}{n})\sum_{t=1}^n \frac{1}{w_t} = O_p(\frac{1}{nb_n}) + O_p(b_n) = o_p(1),$$

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

$$\frac{1}{2nb_n}\sum_{t=1}^n \frac{1}{w_t}\Big\{F[x+\hat{\theta}_n'(\tau)X_{t-1}+b_n] - F[x+\hat{\theta}_n'(\tau)X_{t-1}-b_n]\Big\} = q(\tau)\hat{\sigma}_w + o_p(1).$$

Furthermore, by (4.14), we can readily show that $\hat{q}_n(\tau) = q(\tau) + o_p(1)$. Finally, by Theorem 2.1, the conclusion holds. This completes the proof. □

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, λ, ν and $\{\varepsilon_t\}$. When $\nu = 0$, $y_{nt} = y_t$. Here, y_t comes from model (1.1) under H_0. It is easy to see that $y_{nt} \to y_t$ a.s.. when $n \to \infty$. Similarly define w_{nt} and X_{nt}. Now, under H_{1n},

$$\hat{\lambda}_n(\tau) = \mathrm{argmin}_{\lambda \in R^{p+1}} \sum_{t=1}^n \frac{1}{w_{nt}}\rho_\tau(y_{nt} - X_{nt-1}'\lambda).$$

Define $\tilde{T}_n(s,\tau) = \sum_{t=1}^n X_{nt-1}/w_{nt}[I(\varepsilon_t \leq F^{-1}(\tau) + s'X_{nt-1}/\sqrt{n}) - \tau]/\sqrt{n}$, where $s \in R^{p+1}$. As for Lemma 4.1(i), we can show that

$$(4.15)\quad \|\tilde{T}_n(\sqrt{n}[\hat{\lambda}_n(\tau) - \lambda_n(\tau)], \tau)\| \leq 2(p+1)\max_{1\leq t\leq n}\frac{\|X_{nt-1}\|}{\sqrt{n}w_{nt}} = O_p(\frac{1}{\sqrt{n}}).$$

Let $a_{nt} = (X_{nt-1}/w_{nt})[I(\varepsilon_t \leq F^{-1}(\tau)) - \tau]$ and $a_t = (X_{t-1}/w_t)[I(\varepsilon_t \leq F^{-1}(\tau)) - \tau]$. By Markov's inequality, the dominated convergence theorem and the ergodic theorem, we can show that

$$(4.16)\quad \frac{1}{n}\sum_{t=1}^n a_{nt}a_{nt}' = \tau(1-\tau)\Omega$$

$$\frac{1}{n}\sum_{t=1}^n E(a_{nt}a_{nt}'|\mathcal{F}_{t-1}) = \tau(1-\tau)\Omega.$$

Since a_{nt} is a martingale difference in terms of $\mathcal{F}_t$, 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 $\sqrt{n}[\hat{\lambda}_n(\tau) - \lambda_n(\tau)] = \sqrt{n}[\hat{\lambda}_n(\tau) - \lambda(\tau)] + \nu$, it is straightforward to show that (ii) holds, see also the proof of Theorem 6 in Weiss (1991). This completes the proof. □

REFERENCES

An, H.Z. and Chen, Z.G. (1982). On convergence of LAD estimates in autoregression with infinite variance. *J. Multivariate Anal.* **12**, 335-345.

Bassett, G. W. and Koenker, R. W. (1982). An empirical quantiles function for linear models with i.i.d. errors. *J. Amer. Statist. Assoc.* **77**, 407-415.

Brockwell, P.J. and Davis, R.A. (1986). *Time Series: Theory and Methods.* Second edition. Springer Series in Statistics. Springer-Verlag, New York.

Castillo, E. (1988). *Extreme Value Theory in Engineering.* Academic Univ. Press, Cambridge.

Davis, R.A. and Resnick, S.I. (1985). Limit theory for moving averages of random variables with regularly varying tail probabilities. *Ann. Probab.* **13**, 179-195.

Davis, R.A. and Resnick, S.I. (1986). Limit theory for the sample covariance and correlation functions of moving averages. *Ann. Statist.* **14**, 533-558.

Davis, R.A. ; Knight, K. and Liu, J. (1992). *M*-estimation for autoregressions with infinite variance. *Stochastic Process. Appl.* **40**, 145-180.

Duffy, D.; Mcintosh, A.; Rosenstein, M. and Willinger, W. (1994). Statistical analysis of CCSN/SS7 traffic data from working

CCS subnetworks. *IEEE Journal on Selected Areas in Communications* **12**, 544-551.

Gross, S. and Steiger, W.L. (1979). Least absolute deviation estimates in autoregression with infinite variance. *J. Appl. Probab.* **16**, 104-116.

Kanter, M. and Steiger, W.L. (1974). Regression and autoregression with infinite variance. *Advances in Appl. Probability* **6**, 768-783.

Koedlj, K.; Schafgans, M. and De Vries, C. (1990). The tail index of exchange rate returns. *Journal of International Economics* **29**, 93-108.

Koenker, R. W. and Bassett, G. W. (1978). Regression quantiles. *Econometrica* **46**, 33-50.

Koenker, R. W. and Bassett, G. W. (1982). Robust tests for heteroscedasticity based on regression quantiles. *Econometrica* **50**, 43-61.

Koenker, R. W. and D'Orey, V. (1987). Algorithm AS 229: Computing regression quantiles. *J. Roy. Statist. Soc. Ser. C* **36**, 383-393.

Koenker, R. and Zhao, Q.S. (1996). Conditional quantile estimation and inference for ARCH models. *Econometric Theory* **12**, 793-813.

Kokoszka, P.S. and Taqqu, M.S. (1996). Parameter estimation for infinite variance fractional ARIMA. *Ann. Statist.* **24**, 1880-1913.

Koul, H. L. and Saleh, A. K. Md. E. (1995). Autoregression quantiles and related rank scores processes. *Ann. Statist.* **23**, 670-689.

Hannan, E.J. and Kanter, M. (1977). Autoregressive processes with infinite variance. *J. Appl. Probability* **14**, 411-415.

Jansen, D. and De Vries, C. (1991). On the frequency of large stock returns: putting booms and busts into perspective. *Review of Economics and Statistics.* **73**, 18-2A.

Ling, S. and McAleer, M. (2003). Regression quantiles for unstable autoregression model. To appear in *Journal of Multivariate Analysis.*

Ling, S. and McAleer, M. (2003). Adaptive estimation in nonstationry ARMA models with GARCH noises. *Ann. Statist.* **31**, 642-674.

Mukherjee, K. (1999). Asymptotics of quantiles and rank scores in nonlinear time series. *J. Time Ser. Anal.* **20**, 173-192.

Mikosch, T., Gadrich, T., Klppelberg, C. and Adler, R.J. (1995). Parameter estimation for ARMA models with infinite variance innovations. *Ann. Statist.* **23**, 305-326.

Portnoy, S. and Koenker, R. W. (1989). Adaptive $L-$estimation for linear models. *Ann. Statist.* **17**, 362-381.

Resnick, S.I. (1997). Heavy tail modeling and teletraffic data. With discussion and a rejoinder by the author. *Ann. Statist.* **25**, 1805-1869.

Ruppert, D. and Carroll, R. J. (1980). Trimmed least squares estimation in the linear model. *J. Amer. Statist. Assoc.* **75**, 828-838.

Silverman, B. W. (1986). *Density estimation for statistics and data analysis.* Chapman & Hall, London.

Weiss, A.A (1991). Estimating nonlinear dynamic models using least absolute error estimation. *Econometric Theory* **7**, 46-68.

SHIQING LING is Assistant Professor of Mathematics in the Department of Mathematics , Hong Kong University of Science and Technology. His research interests are in tme series econometrics, financial econometrics, and statistics.

[

TABLE 1

Means and Standard Deviations of SWL for AR Models with $\phi_0 = 0$(1000 replications)

ϕ		n=200 $\hat{\phi}_0$	n=200 $\hat{\phi}$	n=400 $\hat{\phi}_0$	n=400 $\hat{\phi}$	n=200 $\hat{\phi}_0$	n=200 $\hat{\phi}$	n=400 $\hat{\phi}_0$	n=400 $\hat{\phi}$
				$\varepsilon_t \sim$ Cauchy					$\varepsilon_t \sim t_2$
-0.5	Mean	.002	$-.505$	.001	$-.503$	.003	$-.495$	$-.002$	$-.494$
	SD	.134	.103	.098	.071	.130	.107	.093	.071
	AD	.139	.101	.098	.071	.134	.104	.095	.074
0.5	Mean	-.007	.491	-.008	.495	-.004	.487	-.003	.496
	SD	.136	.107	.094	.073	.135	.105	.092	.075
	AD	.139	.102	.098	.071	.134	.104	.094	.073
0.8	Mean	-.014	.787	-.014	.794	-.002	.779	-.001	.792
	SD	.155	.092	.116	.063	.164	.093	.110	.065
	AD	.165	.085	.117	.060	.163	.088	.115	.062

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)

ε_t	$\sim$	n=200 Cauchy	n=200 t_2	n=400 Cauchy	n=400 t_2
ϕ_0	ϕ				
-.1	.3	.453	.437	.742	.732
.0	.3	.416	.371	.698	.676
-.1	.4	.196	.189	.321	.328
.0	.4	.152	.152	.232	.212
.0	.5	.066	.066	.060	.057
.0	.6	.150	.156	.228	.209
.1	.6	.199	.214	.350	.339
.0	.7	.416	.407	.699	.661
.1	.7	.472	.465	.772	.742

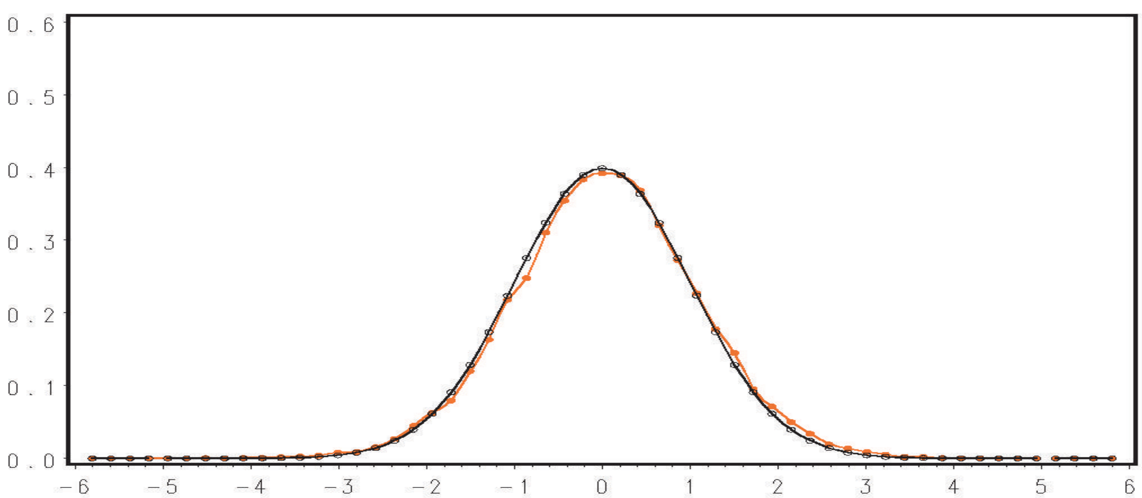

Figure 1: Density Curves of N_{SWLn} and $N(0,1)$: 'circles' and 'dots', respectively

]