Published in final edited form as:

J Nonparametr Stat. 2012 September 1; 24(3): 715–731. doi:10.1080/10485252.2012.698280.

Weighted Quantile Regression for AR model with Infinite Variance Errors

Zhao Chen¹, Runze Li², and Yaohua Wu¹

¹Department of Statistics and Finance, University of Science and Technology of China, Hefei, Anhui 230026, China

²Department of Statistics, Pennsylvania State University, University Park, PA 16802-2111, USA

Abstract

Autoregressive (AR) models with finite variance errors have been well studied. This paper is concerned with AR models with heavy-tailed errors, which is useful in various scientific research areas. Statistical estimation for AR models with infinite variance errors is very different from those for AR models with finite variance errors. In this paper, we consider a weighted quantile regression for AR models to deal with infinite variance errors. We further propose an induced smoothing method to deal with computational challenges in weighted quantile regression. We show that the difference between weighted quantile regression estimate and its smoothed version is negligible. We further propose a test for linear hypothesis on the regression coefficients. We conduct Monte Carlo simulation study to assess the finite sample performance of the proposed procedures. We illustrate the proposed methodology by an empirical analysis of a real-life data set.

Keywords

Quantile regression; autoregressive model; infinite variance

1 Introduction

Suppose that $\{y_1, \dots, y_n\}$ is a sample from the following AR(p) model:

$$y_t = \varphi_0 + \varphi_1 y_{t-1} + \dots + \varphi_p y_{t-p} + \varepsilon_t, \quad (1)$$

where ε_t , $t = 1, \dots, n$, are a sequence of independent and identically distributed errors and $\boldsymbol{\varphi} = (\varphi_0, \varphi_1, \dots, \varphi_p)^T$ is the parameter vector. When $E(\varepsilon_t^2)$ is finite, this AR model has been well studied in the literature. AR models with heavy-tailed errors have been proved to be useful in a variety of research fields, including but not limited to economics and finance (see Koedijk *et al.*, 1990; Jansen and De Vries, 1991), teletraffic data (Duffy *et al.*, 1994, Willinger and Paxson, 1997) and hydrology (Castillo, 1988). This paper is concerned with infinite variance AR models (i.e. AR(p) with $E(\varepsilon_t^2) = \infty$). We refer it as IVAR for simplicity. Infinite variance imposes challenges in estimating $\boldsymbol{\varphi}$. It is of great interest to develop new statistical procedures for IVAR.

There have existed some theoretical results in the literature on IVAR models (see Davis and Resnick (1985), Davis *et al.* (1992), Davis and Dunsmuir (1997) and Ling (2005)). Kanter and Steiger (1974) proved weak consistency of the least squares estimator of $\boldsymbol{\varphi}$. Hannan and Kanter (1977) showed the strong consistency with a convergent rate $n^{1/\delta}$ (Knight, 1987). Davis and Resnick (1985, 1986) firstly gave the asymptotic result of the least squares estimator. Using point process techniques, they proved that least squares estimator

converges weakly to a ratio of least absolute deviation (LAD) estimator and showed its strong consistency. An and Chen (1982) proved that the convergence rate of the LAD estimator is of $n^{1/6}$ and the asymptotic theory of the LAD and M-estimators converge weakly to the minimum of a stochastic process with rate $a_n = \inf\{x: P(|e| > x) \sim n^{-1}\}$. Mikosch *et al.* (1995) studied the Whittle estimate of the parameters of IVAR moving average models. This result was extended to long memory autoregressive fractionally integrated moving average models by Kokoszka and Taqqu (1996). However, none of the limiting distribution in the mentioned has a closed form until Ling (2005), in which the author used a self-weighted least absolute deviation estimation for IVAR and showed that this estimator is asymptotically normal.

Introduced by Koenker and Bassett (1978, 1982), quantile regression has become a very useful alternative to the least squares approach. There is a growing literature devoted to applications of quantile regression to time-series models. Weiss (1991) studied least absolute deviation (LAD) estimation for a dynamic nonlinear model with neither independent nor identically distributed errors. Koul and Saleh (1995) extended the work by Weiss (1991) from the LAD estimation to quantile regression for AR models. Koul and Mukherjee (1994) showed asymptotic representations of the regression quantiles and the regression rank-scores processes in linear regression setting when the errors are a function of Gaussian random variables that are stationary and long range dependent. Koenker and Xiao (2006) proposed a quantile autoregression (QAR) model. Yang and Zhang (2008) proposed a self-weighted estimation procedure for the QAR with infinite variance. The QAR model may be regarded as a special case of model (1) since Koenker and Xiao (2006) demonstrated that the QAR has an equivalent form of model (1) under some conditions.

This paper considers weighted quantile regression for IVAR. To deal with computational issues related to quantile regression, we proposed a new smoothed quantile regression estimator. The proposed estimation procedure can be carried out by solving a differentiable convex optimization problem with Newton-Raphson algorithm or conjugate gradient algorithm effectively. The Newton-Raphson algorithm enables us to estimate the covariance matrix by directly using a sandwich formula, a conventional technique in the literature. Under mild conditions, we show that the difference between the smoothed estimator and original quantile regression estimator converges to zero in probability. This allows us to further establish the asymptotic normality of the smoothed estimator. Based on the asymptotic normality, we further develop statistical inference procedures for the regression coefficients.

The remainder of the paper is organized as follows. In Section 2, we propose a weighted quantile regression and its smoothed version for IVAR. Sampling properties of the proposed estimation procedures are studied. We also discuss the choice of the weights. An algorithm for the proposed procedure is given at the end of this section. Section 3 investigate the finite sample performance of the proposed procedures by Monte Carlo simulations and illustrate the proposed procedures by an empirical analysis of a real-world data set. Concluding remarks are given in section 4. All technical proofs are given in the Appendix.

2 Weighted quantile regression

Let us rewrite model (1) in the following form:

$$y_t = X_t^T \boldsymbol{\varphi} + \varepsilon_t, \quad (2)$$

where $\boldsymbol{\varphi} = (\varphi_0, \varphi_1, \dots, \varphi_p)^T$ and $X_t = (1, y_{t-1}, \dots, y_{t-p})^T$. For a given $\tau \in (0, 1)$, let $\rho(u) = \tau u - u\mathbb{I}(u < 0)$ be the quantile loss function. To ensure the identifiability of $\boldsymbol{\varphi}_0$, it is assumed

that $P(\varepsilon_t = 0) = \tau$. The weighted quantile regression is to minimize the following objective function:

$$L_n(\varphi) = n^{-1} \sum_{t=1}^n w_t \rho(y_t - X_t^T \varphi), \quad (3)$$

where w_t 's are weights depending on $\{y_{t-1}, \dots, y_{t-p}\}$. The purpose of using the weight w_t is to downweight the leverage points (see Davis *et al.* (1992)) in X_t such that the covariance matrices Ω and Σ in Proposition 1 below are finite. Conditions on the weights will be given in Assumption 2 below. We will further discuss the choice of the weights in Section 2.2.

Note that the first order derivative $\rho'(u) = \tau - I(u < 0)$ for any $u \neq 0$. The corresponding score function for φ is

$$U(\varphi) \triangleq -n^{-1} \sum_{t=1}^n w_t X_t \{\tau - I(y_t - X_t^T \varphi < 0)\}. \quad (4)$$

Minimizing $L_n(\varphi)$ yields a weighted quantile regression estimate of φ :

$$\hat{\varphi}_n = \operatorname{argmin}_{\varphi \in \Theta} \{L_n(\varphi)\}, \quad (5)$$

where Θ is the parameter space.

2.1 Asymptotic normality of $\hat{\varphi}_n$

We next study the asymptotic property of $\hat{\varphi}_n$. The strictly stationary and ergodic condition of model (1) and (2) is given below. Let $\{\varepsilon_t\}$ be an independent and identically distributed sequence of random variables with distribution function F satisfying that as $x \rightarrow \infty$,

$$\begin{cases} x^\alpha (1 - F(x)) = x^\alpha P(\varepsilon_t > x) \rightarrow pC, \\ x^\alpha F(x) = x^\alpha P(\varepsilon_t \leq -x) \rightarrow qC, \end{cases} \quad (6)$$

where $0 < \alpha < 2$, $0 < p = 1 - q < 1$, $C > 0$. Let $\eta(\cdot)$ be a polynomial such that $\eta(Z) \neq 0$, for $|Z| \geq 1$. Then $\eta(B)X_t = \varepsilon_t$, where B is a lag-operator, has the unique strictly stationary solution $X_t = \sum_{j=0}^{\infty} \phi_j \varepsilon_{t-j}$, $\{\phi_j\}$ is determined by the relation $\sum_{j=0}^{\infty} \phi_j Z^j = 1/\eta(Z)$. (See, for example, Proposition 13.3.2 in Brockwell and Davis (1986)).

The following assumptions are imposed throughout this paper. They may not be the weakest conditions, but they are used to facilitate the technical proofs.

- Assumption 1** The characteristic polynomial $1 - \varphi_1 Z - \dots - \varphi_p Z^p$ has all roots outside the unit circle and $\{\varepsilon_t\}$ are independent and identically distributed with $E|\varepsilon_t|^{\delta_1} < \infty$ for some $\delta_1 > 0$. Further assume that $E\|X_t\|^{\delta_2} < \infty$ for some $\delta_2 > 0$.
- Assumption 2** The weight $w_t = w(y_{t-1}, \dots, y_{t-p})$ and $w(x_1, \dots, x_p)$ is a real and positive function on R^p such that $E\{(w_t + w_t^2)(\|X_t\|^2 + \|X_t\|^3)\} < \infty$.
- Assumption 3** The distribution function F of the errors $\{\varepsilon_t\}$ is absolutely continuous, with continuous densities f uniformly bounded away from 0 and ∞ , and $f'(\cdot)$ exists and is uniformly bounded. Particularly, $\sup_{x \in \Xi} |f'(x)| < \infty$, where Ξ denotes some neighborhood of 0.

These assumptions shares the same spirit of the regularity conditions in general quantile regression (See Koenker, 2005, Ling, 2005 and Wang *et al.* 2009), but they are stated in the terminology of time series.

Proposition 1—Suppose that Assumptions 1–3 hold, then it follows that

$$n^{1/2}(\hat{\varphi}_n - \varphi_0) \xrightarrow{\mathcal{L}} N\left\{0, \frac{\tau(1-\tau)}{f^2(0)} \Sigma^{-1} \Omega \Sigma^{-1}\right\},$$

where φ_0 is the true value of parameter, $\Sigma = E(w_t X_t X_t^T)$ and $\Omega = E(w_t^2 X_t X_t^T)$. Here and hereafter, $\xrightarrow{\mathcal{L}}$ stands for convergence in distribution as $n \rightarrow \infty$.

There are two challenges in estimation of φ and statistical inference for φ . One is to minimize the weighted quantile loss function since $\rho(\cdot)$ is not differential at the origin. The other one is to estimate the covariance matrix of $\hat{\varphi}_n$ since the asymptotic covariance matrix involved unknown quantity. Under certain regularity conditions, it can be shown that the score function $U(\varphi_0)$ is asymptotically normal with mean zero and covariance matrix $\text{cov}\{U(\varphi_0)\}$ (See Koenker, 2005, Ling, 2005 and Wang *et al.* 2009). We can further show that the asymptotic covariance matrix of $\hat{\varphi}_n$ (see White, 1982)

$$\Lambda_n = A_n^{-1} B_n (A_n^{-1})^T,$$

where $A_n = E\{U(\varphi_0)\} / \varphi = f(0)\Sigma$ and $B_n = \text{cov}\{U(\varphi_0)\} = \tau(1-\tau)n^{-1}\Omega$. For a differentiable score function $U(\cdot)$, one may estimate A_n by $\hat{A}_n = U(\varphi) / \varphi$ evaluated at $\varphi = \hat{\varphi}_n$. For quantile regression, $U(\cdot)$ is not smooth, the derivative does not exist at the origin. This makes it difficult to obtain an estimator for A_n . In next section, we propose a smoothed version of weighted quantile regression, in which one may directly use Newton-Raphson algorithm to deal with the corresponding minimization problem. Furthermore, one may estimate the standard error of the resulting estimate by using the sandwich formula, a conventional technique in the literature of standard error estimation.

2.2 A smoothed weighted quantile regression

Let $\Lambda_n = (\tau(1-\tau)/n f^2(0))\Sigma^{-1}\Omega\Sigma^{-1}$, which is the asymptotic covariance matrix of $\hat{\varphi}_n$. From Proposition 1, $\hat{\varphi}_n \approx \varphi_0 + \Lambda_n^{1/2}Z$, where Z is a multivariate standard normal distribution $N(0, I)$. Instead of minimizing the nonsmooth objective function (3), we propose to construct a smoothed objective function, which is defined as $\tilde{L}(\varphi) = E_Z\{L(\varphi + \Lambda_n^{1/2}Z)\}$ where the expectation is taken over Z . However, this expectation cannot be evaluated as Λ_n is unknown. We nominate a known matrix Γ as an initial value of Λ_n , and then update Γ as an estimator for Λ_n . A simple interpretation is that $\Gamma^{1/2}Z$ can be regarded as a perturbation to φ .

Let $\sigma_t^2 = X_t^T \Gamma X_t$, $a_t = y_t - X_t^T \varphi$, $b_t = a_t / \sigma_t$. We can show that

$$\begin{aligned} \tilde{L}(\varphi) &= E_Z\{L(\varphi + \Lambda_n^{1/2}Z)\} = n^{-1} \sum_{t=1}^n E_Z w_t \rho\{y_t - X_t^T(\varphi + \Gamma^{1/2}Z)\} \\ &= n^{-1} \sum_{t=1}^n w_t \{\sigma_t \varphi(b_t) + a_t [\tau - 1 + \Phi(b_t)]\}, \end{aligned} \quad (7)$$

where $\Phi(\cdot)$ and $\varphi(\cdot)$ are the distribution function and density function of $N(0, 1)$, respectively. As we described before, $U(\boldsymbol{\varphi}) = L(\boldsymbol{\varphi})/\boldsymbol{\varphi}$. We also can show that the smoothed version of $U(\boldsymbol{\varphi})$ is given by

$$\tilde{U}(\boldsymbol{\varphi}) = \frac{\partial \tilde{L}(\boldsymbol{\varphi})}{\partial \boldsymbol{\varphi}} = E_Z\{U(\boldsymbol{\varphi} + \Gamma^{1/2}Z)\} = -n^{-1} \sum_{t=1}^n w_t [X_t(\Phi(b_t) - 1 + \tau)], \quad (8)$$

which is a smooth function of $\boldsymbol{\varphi}$. As a result of Theorem 1 below, we can easily use $\tilde{U}(\boldsymbol{\varphi})/\boldsymbol{\varphi}$ as a smoothing estimator of A_n . That is,

$$\tilde{A}_n = \frac{\partial \tilde{U}(\boldsymbol{\varphi})}{\partial \boldsymbol{\varphi}} = n^{-1} \sum_{t=1}^n w_t \frac{\varphi(b_t)}{\sigma_t} X_t X_t^T, \quad (9)$$

for given values of $\boldsymbol{\varphi}$ and Γ .

The new estimator $\tilde{\boldsymbol{\varphi}}_n$ for $\boldsymbol{\varphi}$ can be obtained from the smoothed score function $\tilde{U}(\boldsymbol{\varphi}) = 0$. Obviously, $\tilde{L}(\boldsymbol{\varphi})$ is a strictly convex function of $\boldsymbol{\varphi}$, then it exists a unique minimizer of $\boldsymbol{\varphi}$ for each given positive definite matrix Γ . However, how the difference between the unsmoothed estimator $\hat{\boldsymbol{\varphi}}_n$ and smoothed estimator $\tilde{\boldsymbol{\varphi}}_n$ is? It is a natural question. Theorem 1 below shows that the resulting estimator $\tilde{\boldsymbol{\varphi}}_n$ is not changed much by such a smoothing method.

Theorem 1—Suppose that Assumptions 1–3 hold. Assume $\Gamma = \Gamma_n$ are a sequence of positive definite matrices in n and $\Gamma_n = O(1/n)$.

A. It follows that

$$\sqrt{n}\{\tilde{U}(\boldsymbol{\varphi}_0) - U(\boldsymbol{\varphi}_0)\} \rightarrow 0, \quad \text{in probability} \quad (10)$$

B. We have the following asymptotic normality:

$$\sqrt{n}(\tilde{\boldsymbol{\varphi}}_n - \boldsymbol{\varphi}_0) \xrightarrow{\mathcal{L}} N(0, \frac{\tau(1-\tau)}{f^2(0)} \sum^{-1} \Omega \sum^{-1}).$$

The condition on Γ_n is to facilitate the proof of this theorem. We will empirically investigate the impact of different choices of Γ_n on the resulting estimate in our simulation study. From our simulation results, the resulting estimate is insensitive to the choice of Γ_n provided that Γ_n falls in a reasonable range. If we further assume that $\Gamma_n = O(n^3 \log n)^{-2}$, we can show in the earlier version of this paper that $\sqrt{n}\{\tilde{U}(\boldsymbol{\varphi}_0) - U(\boldsymbol{\varphi}_0)\} \rightarrow 0$ almost sure.

The covariance matrix of $\tilde{\boldsymbol{\varphi}}_n$ can be estimated by using the corresponding sandwich formula:

$$\tilde{\Lambda}_n = \tilde{A}_n^{-1} \text{cov}\{\tilde{U}(\boldsymbol{\varphi})\} \tilde{A}_n^{-1}. \quad (11)$$

This together with Theorem 1 enables us to make the statistical inference easily. The detailed algorithm will be given at the end of this section.

The choice of the weights—To implement the newly proposed smoothed weighted quantile regression, we need to determine an appropriate data-driven weight. Obviously,

there are many possible choices of the weights. Motivated by the influence function in Huber (1977), we consider the following weight function:

$$w_t = \begin{cases} 1, & \text{if } a_t = 0, \\ \max(|C_1 - m|, |C_2 - m|)^3 / a_t^3, & \text{if } a_t \neq 0, \end{cases}$$

where $a_t = \sum_{i=1}^p |y_{t-i} - m| (1 - I(C_1 \leq y_{t-i} \leq C_2))$ and m is the median of data $\{y_1, \dots, y_n\}$, C_1 and C_2 are the κ and $1 - \kappa$ quantiles of $\{y_1, \dots, y_n\}$, respectively. In practice, we set $\kappa = 0.05$ or 0.10 . This weight satisfies Assumption 2, and will be used in our numerical study. From our simulation comparison (cf. Example 1 below), we find the result with $\kappa = 0.05$ is almost the same as that for $\kappa = 0.10$. This implies that the choice of κ is not very sensitive provided that it falls in the reasonable range. The weight shrinks the potential outliers in both tails. In the numerical studies, we can find that this weight downweights the covariance matrices with outliers. Thus, the larger $C \triangleq \max(|C_1 - m|, |C_2 - m|)^3 / a_t^3$ is, the smaller the asymptotic variance is. However, when C is too large, the difference between distribution of estimate $\hat{\phi}_n$ and its limiting distribution becomes visible for small sample size n . As pointed out by Ling (2005) for LAD estimate, there is no theory to support the choice of constants C_1 and C_2 for weighted quantile regression. Further study on this issue is needed.

An algorithm—Let us summarize the newly proposed smoothed weighted quantile regression procedure as an algorithm as below.

- Step 1** Initialize the value of Γ . Set $\Gamma^{(0)} = \pi^{-1} I_{p+1}$.
- Step 2** In the j th iteration ($j = 1, 2, \dots$), we update $\phi^{(j)}$, which is the solution of function $\tilde{U}(\phi) = E_Z[U(\phi + \Gamma^{(j-1)1/2}Z)] = 0$ by using Newton-Raphson algorithm.
- Step 3** Use $\phi^{(j)}$ and $\Gamma^{(j-1)}$ to update $\tilde{A}_n$, and then obtain an updated

$$\Gamma^{(j)} = \tilde{A}_n^{-1} B_n \tilde{A}_n^{-1}.$$

- Step 4** Repeat the above iteration steps 2 and 3 until $\Gamma^{(j)}$ and $\Gamma^{(j+1)}$ is close enough, e.g., $\max |\Gamma^{(j)} - \Gamma^{(j+1)}| < \varepsilon$, a prespecified threshold. In our simulation, we set $\varepsilon = 10^{-6}$.

It is worth noting that B_n does not depend on Γ and ϕ . From our limited experience, the algorithm converges quickly. Suppose that the algorithm converges at J -step, the resulting estimates for ϕ and the covariance matrix Λ_n are $\phi^{(J)}$ and $\Gamma^{(J)}$, respectively. In this algorithm, we set an initial value of Γ to be $\pi^{-1} I_{p+1}$, and update it so that it can be an estimator of Λ_n .

By the updating Γ , we hope the working $\tilde{L}(\cdot)$ will be closer to $E_Z(\phi + \Lambda_n^{1/2}Z)$. Theoretical investigation on the optimal choice of Γ is out of scope of this paper.

2.3 Statistical inference for ϕ

Theorem 1 provides us a basis for statistical inference on ϕ for IVAR model. Consider linear hypothesis:

$$H_0: \Psi \phi_0 = \gamma,$$

where Ψ is a prespecified $p_1 \times (p+1)$ matrix with full rank p_1 , and γ is a prespecified p_1 -dimensional vector. Based on Theorem 1, we define the Wald's test statistic as follows

$$W_n(p_1) = (\Psi \tilde{\varphi}_n - \gamma)^T (\Psi \tilde{\Lambda}_n \Psi^T)^{-1} (\Psi \tilde{\varphi}_n - \gamma).$$

Under H_0 , we can show that under the assumptions of Theorem 1, $W_n(p_1) \rightarrow \chi_{p_1}^2$, the Chi-square distribution with p_1 degrees of freedom.

As a direct application of the linear hypothesis, we may formulate a test for independent white noise against an AR alternative as

$$H_0: \varphi_0 = \varphi_1 = \cdots = \varphi_p = 0.$$

Thus, the corresponding Wald's test is given by

$$W_{0n}(p+1) = \tilde{\varphi}_n^T \tilde{\Lambda}_n^{-1} \tilde{\varphi}_n,$$

which follows a χ_{p+1}^2 asymptotically, under H_0 .

Theorem 1 also enables us to investigate the local power of $W_n(p_1)$ by considering the local alternative hypothesis:

$$H_{1n}: \Psi \varphi_n = \gamma,$$

where $\varphi_n = \varphi_0 + (\nu / \sqrt{n})$ and $\nu \in R^{p+1}$ is a constant vector.

Using techniques in the proof of Theorem 1, we can show that under the assumptions of Theorem 1, and under hypothesis H_{1n} , it follows that

$$n^{1/2}(\tilde{\varphi}_n - \varphi_0) \xrightarrow{\mathcal{L}} N \left\{ \nu, \frac{\tau(1-\tau)}{f^2(0)} \Sigma^{-1} \Omega \Sigma^{-1} \right\}, \quad \text{and} \quad W_n(p_1) \xrightarrow{\mathcal{L}} \chi_{p_1}^2(\mu),$$

where $\mu = \frac{f^2(0)}{\tau(1-\tau)} \nu^T \Psi^T (\Psi \Sigma^{-1} \Omega \Sigma^{-1} \Psi^T)^{-1} \Psi \nu$ is a non-central parameter.

3 Simulation Studies

In this section, we examine the finite sample performance of the proposed smoothed weighted quantile regression, and illustrate the proposed methodology by a real data example. All numerical studies were conducted by using Matlab code, which is available from the authors upon request.

Example 1—In this example, random samples were generated from the following AR(1) model:

$$y_t = \varphi_0 + \varphi_1 y_{t-1} + \varepsilon_t,$$

where $\{\varepsilon_t\}$, $t = 1, \dots, n$ are a sequence of white noises. We consider $n = 200$ and 400 . In our simulation, we consider three error distributions: Cauchy distribution, t-distribution with 2 degrees of freedom (denoted by t_2 for short) and $N(0, 1)$. Note that both Cauchy distribution and t_2 have infinite variance. The normal distribution is included here as a benchmark. In our simulation, we consider $\tau = 0.1, 0.25, 0.5, 0.75$ and 0.9 . The sampled data is from a given distribution with a shift so that the condition $P(\varepsilon_t = 0) = \tau$ holds. The true values $\boldsymbol{\varphi}_0$ are set to be $(\varphi_0, \varphi_1) = (0, -.25), (0, -.5), (0, -.75), (0, -.8)$ and $(0, .25), (0, .5), (0, .75), (0, .8)$. For each case, we conducted 1000 simulations. We report results for $\boldsymbol{\varphi}_0 = (0, .5)$ only. Results for other true values are similar to those for $(0, .5)$, and thus are not reported here to save space.

We first investigate how the order of $I^{(0)}$ affects the resulting estimate. To this end, we set $I^{(0)} = \gamma_n I_{p+1}$, and take $\gamma_n = n^{-1}, n^{-1.5}, n^{-2}$ and $n^{-2.5}$. Table 1 summarizes the results for $\tau = 0.5$, $n = 200$ and $\kappa = 0.05$. Results for other cases are similar and not reported here to save space. In Tables 1 and other tables in this section, ‘Mean’ stands for the average of 1000 estimates over 1000 simulations, ‘SD’ for the standard deviation of the 1000 estimates, and ‘AD’ for the average of 1000 standard error estimates using the sandwich formula in (11) over 1000 simulations. Table 1 indicates that the resulting estimate is quite robust to the order of $I^{(0)}$. This is expected since the initial value of I should have little impact on the resulting estimate provided that the initial value falls in the reasonable range. Thus, we set $I^{(0)} = n^{-1} I_{p+1}$ in the rest of this section as in the algorithm given in Section 2.2.

We next investigate how the choice of κ in the weights affects the estimation procedure. To this end, we consider $\kappa = 0.05$ and 0.10 . Table 2 summarizes the simulation results. Table 2 reports results for $\tau = 0.5$ only. Results for other cases are similar. Labels in Table 2 are the same as those for Table 1. From Table 2, it can be seen that the results for $\kappa = 0.05$ and 0.10 are very close. This implies that the estimation procedure is insensitive to the choice of κ . Thus, we set $\kappa = 0.05$ in the rest of this section. Next we empirically compare the proposed method with that in Ling (2005), in which the author considered $\tau = 0.5$ only, and kernel density estimator was used to estimate $f(0)$. Note that our simulation setting is the same as that in Ling (2005). For purpose of comparison and to avoid duplicate efforts, we directly extracted the corresponding numerical results from Ling (2005) and present them in the bottom panel of Table 2. The newly proposed method with $\kappa = 0.05$ performs slightly better than Ling (2005)’s method for most cases.

Table 3 reports simulation results for $\tau = 0.1, 0.25, 0.75$ and 0.9 with $\kappa = 0.10$. Labels in Table 3 are the same as those for Table 1. From Tables 2 and 3, it can be seen that the variance for $\tau = 0.1$ and 0.9 are greater than that for $\tau = 0.25, 0.75$, and $\tilde{\boldsymbol{\varphi}}$ has smallest variance at $\tau = 0.5$. This is consistent with the theoretical results. Tables 2 and 3 also implies that the weighted quantile regression performs better for the lighter tail error distribution in terms of bias and variance. This is quite obvious by comparing the bias with Cauchy error and that with normal error. We also can see from Table 3 that the bias for Cauchy error distribution is quite large at $\tau = 0.1$ and 0.9 , but it reduces dramatically as the sample size increases from $n = 200$ to $n = 400$.

We further examine the accuracy and variation of the proposed covariance estimator by sandwich formula. Table 4 depicts the sample standard deviations of the 1000 estimated standard errors. From Table 4, we can see that the variation of the proposed variance estimator is quite small for $\tau = 0.5$. The variation become larger from $\tau = 0.5$ to $\tau = 0.25$ and 0.75 , and become even larger when $\tau = 0.1$ and 0.9 . This pattern is consistent with the pattern for SD and AD in Tables 2 and 3. It is expected that the larger asymptotic variance has, the large variation of its estimator will be. The differences between SD and AD in Tables 2 and 3 are the same magnitude as that of the corresponding standard deviation of the

1000 estimated standard errors. Indeed, most of the differences are less than twice of the corresponding standard deviations of the estimated standard errors. This implies that the proposed covariance estimator works reasonably well.

We next assess the finite sample performance of the proposed Wald's test. For simplicity, we fix $\varphi_0 = 0$ and set $H_0: \varphi_1 = 0.5$ and $H_1: \varphi_1 \neq 0.5$. Thus, the limiting distribution of the proposed Wald's test under H_0 is the χ^2 -distribution with 1 degree of freedom. We consider the confidence level $\alpha = 0.1, 0.05$ and 0.01 to investigate the size and power of the proposed test. The corresponding critical values are 2.7055, 3.8415 and 6.6349, respectively. We first examine Type I error rate. Under H_0 , we conduct 1000 simulations for $n = 200, 400$, and $\tau = 0.25, 0.5$ and 0.75 . The rejection rate under H_0 is reported in rows with $\varphi_1 = 0.5$ of Table 5.

For a given significance level α , the Monte Carlo error is about $2\sqrt{\alpha(1-\alpha)/1000}$ for 1000 Monte Carlo simulations. Thus, the Monte Carlo error is 0.0190, 0.0138 and 0.0063 for $\alpha = 0.1, 0.05$ and 0.01 , respectively. Table 5 indicates that the Wald's test keeps Type I error rate well for $\alpha = 0.10$ and 0.05 . For Cauchy error, the Wald's test also keeps Type I error rate well for $\alpha = 0.01$, while it does not keep Type I error rate well for normal error with $\alpha = 0.01$ and $n = 200$.

To investigate the power of the proposed Wald's test, we fix $\varphi_0 = 0$ and take $\varphi_1 = 0.3, 0.4, 0.6$ and 0.7 . We also conduct 1000 simulations. The rejection rate is depicted in Table 5, from which it can be seen that the Wald's test is quite powerful, particularly for Cauchy error.

Example 2—In this example, we illustrate the proposed methodology by an empirical analysis of daily electricity price collected in a district of New Jersey-Pennsylvania-Maryland area. This data set consists of 1000 daily electronic prices. After some preliminary analysis, we made a logarithm transformation on the electronic prices. That is, y_t equals the logarithm of electronic price at day $t, t = 1, \dots, n$. The sample autocorrelation function and sample partial auto correlation function are depicted in Figure 1, which implies that the process is approximately an AR(1) process.

We further estimate the tail index of $y_t, t = 1, \dots, n$ using Hill estimate (Resnick, 1997) with order 360. The tail index is displayed in Figure 2, from which we can see that both left- and right-tail indices are far from 2. This implies that the error may be heavy-tailed. This motivated us to consider weighted quantile regression for AR(1) model:

$$y_t = \varphi_0 + \varphi_1 y_{t-1} + \varepsilon_t. \quad (12)$$

Table 6 depicts the resulting estimates and their standard errors. From the estimates of φ_1 , there might exist mild heteroscedasticity. Figure 3 shows the scatter plot of y_t versus y_{t-1} , and quantile regression lines with $\tau = 0.1, \tau = 0.5, \tau = 0.9$. We also plot the linear regression lines along with its 80% pointwise prediction interval for the response variable. Note that the quantile regression lines with $\tau = 0.1$ and $\tau = 0.9$ consists of a 80% pointwise predictor interval of the response variable. From Figure 3, the coverage probability of prediction interval based on the quantile regression is much better than that based on the least squares methods.

4 Conclusion

We proposed a weighted quantile regression procedure for autoregression models with infinite variance errors in this paper. We proved the asymptotic properties of the estimator

and constructed a Wald statistic for testing the linear hypothesis of parameters. This method can be used to other time series models.

Acknowledgments

Zhao Chen's research was partially supported by National Institute on Drug Abuse (NIDA) grant R21-DA024260 as a graduate research assistant during his visit at the Pennsylvania State University. Runze Li is the corresponding author and his research was supported by NIDA grant P50-DA10075 and National Natural Science Foundation of China grants 11028103 and 10911120395. Yaohua Wu's research was supported by National Natural Science Foundation of China grant 10871188. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIDA or the NIH.

References

- An HZ, Chen ZG. On convergence of LAD estimates in autoregression with infinite variance. *J Multiv Anal.* 1982; 12:335–345.
- Brockwell, PJ.; Davis, RA. *Time Series: Theory and Methods*. New York: Springer; 1986.
- Castillo, E. *Extreme Value Theory in Engineering*. Cambridge: Cambridge University Press; 1988.
- Davis RA, Dunsmuir WTM. Least absolute deviation estimation for regression with ARMA errors. *J Theoret Probab.* 1997; 10:481–497.
- Davis RA, Knight K, Liu J. M-estimation for autoregressions with infinite variance. *Stoch Processes Appl.* 1992; 40:145–180.
- Davis RA, Resnick SI. Limit theory for moving averages of random variables with regularly varying tail probabilities. *Ann Probab.* 1985; 13:179–195.
- Davis RA, Resnick SI. Limit theory for the sample covariance and correlation functions of moving averages. *Ann Statist.* 1986; 14:533–558.
- Duffy D, McIntosh A, Rosenstein M, Willinger W. Statistical analysis of CCSN/SS7 traffic data from working CCS subnetworks. *IEEE J Selctd Areas Commun.* 1994; 12:544–551.
- Hannan EJ, Kanter M. Autoregressive processes with infinite variance. *J Appl Probab.* 1977; 14:411–415.
- Hjort, NL.; Pollard, D. *Asymptotics for minimizers of convex processes*. Department of Statistics, Yale University; 1993. Preprint
- Huber, PJ. *Robust Statistical Procedures*. Philadelphia: SIAM; 1977.
- Jansen D, De Vries C. On the frequency of large stock returns: putting booms and busts into perspective. *Rev Econ Statist.* 1991; 73:18–24.
- Kanter M, Steiger WL. Regression and autoregression with infinite variance. *Adv Appl Probab.* 1974; 6:768–783.
- Knight K. Rate of convergence of centred estimates of autoregressive parameters for infinite variance autoregressions. *J Time Ser Anal.* 1987; 8:51–60.
- Knight K. Limiting distributions for L1 regression estimators under general conditions. *Ann Statist.* 1998; 26:755–770.
- Koedijk C, Schafgans M, De Vries C. The tail index of exchange rate returns. *J Int Econ.* 1990; 29:93–108.
- Koenker, R. *Quantile Regression*. Cambridge: Cambridge University Press; 2005.
- Koenker RW, Bassett GW. Regression quantiles. *Econometrica.* 1978; 46:33–50.
- Koenker RW, Bassett GW. Tests of Linear Hypotheses and l_1 Estimation. *Econometrica.* 1982; 50:1577–83.
- Koenker R, Xiao ZJ. Quantile autoregression. *J Amer Statist Assoc.* 2006; 101:980–990.
- Kokoszka PS, Taqqu MS. Parameter estimation for infinite variance fractional ARIMA. *Ann Statist.* 1996; 24:1880–1913.
- Koul HL, Saleh AK, MdE. Autoregression quantiles and related rank scores processes. *Ann Statist.* 1995; 23:670–689.
- Koul H, Mukherjee R. Regression quantiles and related processes under long range dependent errors. *J Mult Anal.* 1994; 51:318–337.

- Ling S. Self-weighted least absolute deviation estimation for infinite variance autoregressive models. *J Roy Statist Soc Ser B*. 2005; 67:381–393.
- Mikosch T, Gadrich T, Klüppelberg C, Adler RJ. Parameter estimation for ARMA models with infinite variance innovations. *Ann Statist*. 1995; 23:305–326.
- Pollard D. Asymptotics for Least Absolute Deviation Regression Estimators. *Econometr Theory*. 1991; 7:186–199.
- Resnick SI. Heavy tail modeling and teletraffic data (with discussion). *Ann Statist*. 1997; 25:1805–1869.
- Willinger W, Paxson V. Discussion: Heavy Tail Modeling and Teletraffic Data. *Ann Statist*. 1997; 25:1856–1866.
- Wang YG, Shao QX, Zhu M. Quantile regression without the curse of unsmooth-ness. *Computational Statistics and Data Analysis*. 2009; 53:3696–3705.
- Weiss AA. Estimating nonlinear dynamic models using least absolute error estimation. *Econometr Theory*. 1991; 7:46–68.
- White H. Maximum likelihood estimation of misspecified models. *Econometrica*. 1982; 50:1–25.
- Yang XR, Zhang LX. A note on self-weighted quantile estimation for infinite variance quantile autoregression models. *Statistics and Probability Letters*. 2008; 78:2731–2738.

Appendix

Proof of Proposition 1

Using the same strategy in Ling (2005) for median regression and Yang and Zhang (2008) for QAR, we can prove this proposition. To save space, we will provide a sketch of the proof. The detailed proof is given in the dissertation of the first author, and is available upon request to the first author.

Denote $\hat{\theta}_n = n^{1/2}(\hat{\varphi}_n - \varphi_0)$ and

$$L_n^*(\mathbf{u}) = \sum_{t=1}^n w_t \{ \rho_\tau(\varepsilon_t - n^{-1/2} \mathbf{u}^T X_t) - \rho_\tau(\varepsilon_t) \} \quad (13)$$

where $\mathbf{u} \in R^{p+1}$. So it follows that $\hat{\theta}_n$ is the minimizer of $L_n^*(\mathbf{u})$ on R^{p+1} . Using the identity

$$\rho_\tau(u-v) - \rho_\tau(u) = -v(\tau - I(u < 0)) + \int_0^v (I(u \leq s) - I(u \leq 0)) ds$$

(Knight, 1998). Thus, it follows that

$$L_n^*(\mathbf{u}) = -\mathbf{u}^T U_n + \sum_{t=1}^n w_t \int_0^{\mathbf{u}^T X_t / \sqrt{n}} (I(\varepsilon_t \leq s) - I(\varepsilon \leq 0)) ds, \quad (14)$$

where $U_n = \sqrt{n}U(\varphi)$ and $U(\varphi)$ is defined in Section 2. Denote

$$\xi_t(\mathbf{u}) = w_t \int_0^{\mathbf{u}^T X_t / n^{1/2}} (I(\varepsilon_t \leq s) - I(\varepsilon \leq 0)) ds.$$

Recall that $F(x)$ is the distribution function of ε_t . It follows by the Taylor expansion and some calculations that

$$\sum_{t=1}^n E\{\xi_t(\mathbf{u})|\mathcal{F}_{t-1}\} = \mathbf{u}^T \left\{ \frac{f(0)}{2n} \sum_{t=1}^n w_t X_t X_t^T \right\} \mathbf{u} + \frac{1}{2} \sum_{t=1}^n w_t \int_0^{\mathbf{u}^T X_t / n^{1/2}} s^2 f'(s^*) ds, \quad (15)$$

where $s^* \in (0, s)$. Denote

$$R_n(\mathbf{u}) = \frac{1}{2} \sum_{t=1}^n w_t \int_0^{\mathbf{u}^T X_t / n^{1/2}} s^2 f'(s^*) ds + \sum_{t=1}^n \{\xi_t(\mathbf{u}) - E\{\xi_t(\mathbf{u})|\mathcal{F}_{t-1}\}\} \quad (16)$$

Thus, according to equations (14), (15) and (16), it follows that

$$L_n^*(\mathbf{u}) = -\mathbf{u}^T U_n + \mathbf{u}^T \left\{ \frac{f(0)}{2n} \sum_{t=1}^n w_t X_t X_t^T \right\} \mathbf{u} + R_n(\mathbf{u}). \quad (17)$$

We next deal with the residual term $R_n(\mathbf{u})$. By Assumptions 2 and 3, it can be shown that

$$\left| \sum_{t=1}^n w_t \int_0^{\mathbf{u}^T X_t / n^{1/2}} s^2 f'(s^*) ds \right| \leq \sup_x |f'(x)| \frac{\|\mathbf{u}\|^3}{n \cdot n^{1/2}} \sum_{t=1}^n w_t \|X_t\|^3 \rightarrow 0, \quad a.s.$$

Since

$$E\{\xi_t^2(\mathbf{u})\} = E\left\{w_t \int_0^{\mathbf{u}^T X_t / \sqrt{n}} \{I(\varepsilon_t \leq s) - I(\varepsilon_t \leq 0)\} ds\right\}^2, \quad (18)$$

For $s > 0$, $E\{\xi_t^2(\mathbf{u})\} \leq E w_t^2 \int_0^{|\mathbf{u}^T X_t| / \sqrt{n}} (I(\varepsilon_t \leq s) - I(\varepsilon_t \leq 0))^2 ds \cdot \int_0^{|\mathbf{u}^T X_t| / \sqrt{n}} 1 ds$ by the Cauchy-Schwarz inequality. Since

$$\begin{aligned} & E w_t^2 \int_0^{|\mathbf{u}^T X_t| / \sqrt{n}} (I(\varepsilon_t \leq s) - I(\varepsilon_t \leq 0))^2 ds \cdot \int_0^{|\mathbf{u}^T X_t| / \sqrt{n}} 1 ds \\ &= E \left\{ \frac{1}{\sqrt{n}} w_t^2 |\mathbf{u}^T X_t| \int_0^{|\mathbf{u}^T X_t| / \sqrt{n}} E\{|I(\varepsilon_t \leq s) - I(\varepsilon_t \leq 0)|^2 | \mathcal{F}_{t-1}\} ds \right\} \end{aligned}$$

which is not more than

$$E \left\{ \frac{1}{\sqrt{n}} w_t^2 |\mathbf{u}^T X_t| \int_0^{|\mathbf{u}^T X_t| / \sqrt{n}} E\{|I(\varepsilon_t \leq s) - I(\varepsilon_t \leq 0)|^2 | \mathcal{F}_{t-1}\}^{1/2} ds \right\}$$

which equals

$$E \left\{ \frac{1}{\sqrt{n}} w_t^2 |\mathbf{u}^T X_t| \int_0^{|\mathbf{u}^T X_t| / \sqrt{n}} (F(s) - F(0))^{1/2} ds \right\}$$

By the Taylor Expansion, it follows that

$$E\{\xi_t^2(\mathbf{u})\} \leq E\left\{\frac{1}{\sqrt{n}}w_t^2|\mathbf{u}^T X_t|\int_0^{|\mathbf{u}^T X_t|/\sqrt{n}}(f(0)s+o(s))^{1/2}ds\right\}$$

We can further show that

$$E\{\xi_t^2(\mathbf{u})\} \leq \frac{\|\mathbf{u}\|^{5/2}}{n^{5/4}} \max_x \{f(x)^{1/2}\} E\left\{w_t^2(\|X_t\|^2 + \|X_t\|^3)\right\}. \quad (19)$$

By Assumptions 2 and 3, for each $\mathbf{u}$, it follows that

$$E\left(\sum_{t=1}^n [\xi_t(\mathbf{u}) - E\{\xi_t(\mathbf{u})|\mathcal{F}_{t-1}\}]\right)^2 = \sum_{t=1}^n E[\xi_t(\mathbf{u}) - E\{\xi_t(\mathbf{u})|\mathcal{F}_{t-1}\}]^2 + 2 \sum_{t>s} E\{[\xi_t(\mathbf{u}) - E\{\xi_t(\mathbf{u})|\mathcal{F}_{t-1}\}] \cdot [\xi_s(\mathbf{u}) - E\{\xi_s(\mathbf{u})|\mathcal{F}_{s-1}\}]\} \quad (20)$$

Using the property of conditional expectation, it can be shown that the last term of equation (20),

$$E\{[\xi_t(\mathbf{u}) - E\{\xi_t(\mathbf{u})|\mathcal{F}_{t-1}\}] \cdot [\xi_s(\mathbf{u}) - E\{\xi_s(\mathbf{u})|\mathcal{F}_{s-1}\}]\} = 0$$

Thus, it follows by Assumption 2 that

$$\begin{aligned} E\left(\sum_{t=1}^n [\xi_t(\mathbf{u}) - E\{\xi_t(\mathbf{u})|\mathcal{F}_{t-1}\}]\right)^2 &= \sum_{t=1}^n E[\xi_t(\mathbf{u}) - E\{\xi_t(\mathbf{u})|\mathcal{F}_{t-1}\}]^2 \\ &\leq 2 \sum_{t=1}^n \{E\xi_t^2(\mathbf{u})\} \rightarrow 0, \quad \text{as } n \rightarrow \infty. \end{aligned}$$

This implies that $R_n(\mathbf{u}) = o_p(1)$ for each $\mathbf{u}$.

By Assumptions 1 and 2 and the Ergodic Theorem, we have

$$\begin{aligned} \frac{1}{n} \sum_{t=1}^n (w_t X_t X_t^T) &\rightarrow \Sigma, \quad \text{almost surely,} \\ \frac{1}{n} \sum_{t=1}^n (w_t^2 X_t X_t^T) &\rightarrow \Omega, \quad \text{almost surely, as } n \rightarrow \infty \end{aligned}$$

Thus, by the Central Limit Theorem, it follows that $U_n(\boldsymbol{\varphi}_0)$ weakly converges to $N(0, \boldsymbol{\tau}(1 - \boldsymbol{\tau})\Omega)$. Thus, for each $\mathbf{u}$, we have

$$L_n^*(\mathbf{u}) \xrightarrow{\mathcal{L}} L_0^*(\mathbf{u}) = -\mathbf{u}^T \Phi + f(0)\mathbf{u}^T \sum \mathbf{u}$$

where Φ follows $N(0, \tau(1-\tau)\Omega)$. The convexity of the limiting objective function, $L_0^*(u)$, assures the uniqueness of the minimizer. By the convexity lemma (see, e.g., Pollard, 1991; Hjort and Pollard, 1993; Knight, 1998), it follows that

$$\sqrt{n}(\widehat{\varphi}_n - \varphi_0) = \widehat{\theta}_n = \arg \min L_n^*(u) \rightarrow \widehat{u}_0 = \arg \min L_0^*(u)$$

Moreover, $\widehat{u}_0 = f(0)^{-1}\Sigma^{-1}\Phi$ almost surely. Using Corollary 2 in Knight(1998), we can show that

$$\widehat{\theta}_n = n^{1/2}(\widehat{\varphi}_n - \varphi_0) \xrightarrow{\mathcal{L}} N\left\{0, \tau(1-\tau)f(0)^{-2}\Sigma^{-1}\Omega\Sigma^{-1}\right\}.$$

This completes the proof of Proposition 1.

Proof of Theorem 1

We first show that $\sqrt{n}(\tilde{U}(\varphi_0) - U(\varphi_0))$ converges to zero in probability. By Markov's inequality, it is sufficient to show that $E\left(\sqrt{n}|\tilde{U}(\varphi_0) - U(\varphi_0)|\right) \rightarrow 0$. Let $\tilde{U}_t(\varphi)$ and $U_t(\varphi)$ be the contribution from the t -th observation to $\tilde{U}(\varphi)$ and $U(\varphi)$, respectively. Denote $d_t = n(\tilde{U}_t(\varphi_0) - U_t(\varphi_0))$, and d_{tj} to be the j -th element of d_t . Let $\mathbf{e}_j$ be $(p+1)$ -dimensional column vector with the j -th element being 1 and other elements being 0. Note that $E\left(\sqrt{n}|e_j^T\{\tilde{U}(\varphi_0) - U(\varphi_0)\}|\right) \leq E\left(\sum_{t=1}^n |d_{tj}|/\sqrt{n}\right)$ for $j = 0, \dots, p$. Thus, it suffices to prove that $(1/\sqrt{n})\sum_{t=1}^n E|d_{tj}| \rightarrow 0$. Recalling the definition of d_t it can be shown by some tedious calculation that

$$\begin{aligned} \sum_{t=1}^n \frac{E|d_{tj}|}{\sqrt{n}} &= \sum_{t=1}^n n^{1/2} E|w_t X_{tj} \Phi(-|b_t|) I(b_t < 0) - w_t X_{tj} \Phi(-|b_t|) I(b_t > 0)| \\ &= \sum_{t=1}^n n^{1/2} E\{|w_t X_{tj} \sigma_t| \cdot E[\Phi(-|b_t|) \cdot |2I(b_t < 0) - 1| | \mathcal{F}_{t-1}]\} \\ &= \sum_{t=1}^n n^{-1/2} E\{|w_t x_{tj} \sigma_t| \cdot \int \Phi(-|z|) |2I(z < 0) - 1| \{f(0) + O(z\sigma_t)\} dz\} \\ &= O(1) n^{-1/2} \sum_{t=1}^n E|w_t x_{tj} \sigma_t^2| \rightarrow 0 \end{aligned}$$

by using Assumptions 2 and 3 which ensure that $E|w_t x_{tj} \sigma_t^2| = O(1/n)$ since $\Gamma_n = O(1/n)$. This completes the proof of Theorem 1 (A).

We next establish the asymptotic normality of $\hat{\varphi}$. The proof strategy is similar to the proof of Proposition 1. Denote $\tilde{\delta}_n = n^{1/2}(\tilde{\varphi}_n - \varphi_0)$ and

$$\tilde{L}(\varphi) = E_z\{L(\varphi + \Lambda_n^{1/2} Z)\} = n^{-1} \sum_{t=1}^n E_z w_t \rho\{y_t - X_t^T(\varphi + \Gamma^{1/2} Z)\}. \text{ Define}$$

$$Z_n(\delta) = \tilde{L}_n(\varphi_0 + n^{-1/2}\delta) - \tilde{L}_n(\varphi_0)$$

where $\delta \in R^{p+1}$. So it follows that $\tilde{\delta}_n$ is the minimizer of $Z_n(\delta)$ on R^{p+1} . Using the Taylor expansion

$$\begin{aligned} Z_n(\delta) &= n^{1/2} \left[\frac{\partial \tilde{L}(\varphi_0)}{\partial \varphi} \right]^T \delta + \frac{1}{2} \delta^T \left[\frac{\partial^2 \tilde{L}(\varphi_0)}{\partial \varphi \partial \varphi^T} \right] \delta + o(\frac{1}{n} \delta^T \delta) \\ &= n^{1/2} \left[\frac{\partial \tilde{L}(\varphi_0)}{\partial \varphi} \right]^T \delta + \frac{1}{2n} \left(\sum_{t=1}^n w_t \delta^T X_t X_t^T \delta \frac{\varphi(b_t)}{\sigma_t} \right) + o(\frac{1}{n} \delta^T \delta) \\ &= n^{1/2} \tilde{U}(\varphi_0)^T \delta + \frac{1}{2n} \left(\sum_{t=1}^n w_t \delta^T X_t X_t^T \delta \frac{\varphi(b_t)}{\sigma_t} \right) + o(\frac{1}{n} \delta^T \delta), \end{aligned}$$

where $\sigma_t^2 = X_t^T \Gamma X_t$, $a_t = y_t - X_t^T \varphi$, $b_t = a_t / \sigma_t$. We denote the first and second terms of above last equation as Z_{1n} and Z_{2n} respectively. Thus,

$$Z_{1n} = n^{1/2} \tilde{U}(\varphi_0)^T \delta \quad (21)$$

$$Z_{2n} = \frac{1}{2n} \left(\sum_{t=1}^n w_t \delta^T X_t X_t^T \delta \frac{\varphi(b_t)}{\sigma_t} \right) \equiv \sum_{t=1}^n Z_{2nt} \quad (22)$$

As shown in the proof of Proposition 1, $n^{1/2} U(\varphi_0)$ follows an asymptotic normal distribution. By using Slutsky's theorem and (10), it follows that $n^{1/2} U(\tilde{\varphi}_0)$ and $n^{1/2} U(\varphi_0)$ have the same limiting distribution. Furthermore, $\lim_{n \rightarrow \infty} n^{1/2} E \tilde{U}(\varphi_0) = 0$. We consider

$$\begin{aligned} E(Z_{2n} | \mathcal{F}_{t-1}) &= \frac{1}{2n} \delta^T \left[\sum_{t=1}^n w_t X_t X_t^T E \left(\frac{\varphi(b_t)}{\sigma_t} \middle| \mathcal{F}_{t-1} \right) \right] \delta \\ &= \frac{1}{2n} \delta^T \left[\sum_{t=1}^n w_t X_t X_t^T \left(\frac{1}{\sqrt{2\pi}} \int \frac{1}{\sigma_t} \exp \left\{ -\frac{1}{2} \cdot \frac{u^2}{\sigma_t^2} \right\} f(u) du \right) \right] \delta \\ &= \frac{1}{2n} \delta^T \left[\sum_{t=1}^n w_t X_t X_t^T \left(\frac{1}{\sqrt{2\pi\sigma_t}} \int \exp \left\{ -\frac{u^2}{2\sigma_t^2} \right\} f(u) du \right) \right] \delta \end{aligned}$$

By the definition of σ_t^2 and the assumption $\Gamma = O(1/n)$, we can show by using the Markov inequality that under Assumption 1, $\sigma_t^2 = O_p(n^{-\delta_2})$. Thus, we can further show by employing techniques related to Gaussian kernel density estimate that

$$\begin{aligned} E(Z_{2n} | \mathcal{F}_{t-1}) &= \frac{1}{2n} \delta^T \left[\sum_{t=1}^n w_t X_t X_t^T \left(\frac{1}{\sqrt{2\pi\sigma_t}} \int \exp \left\{ -\frac{u^2}{2\sigma_t^2} \right\} (f(0) + f'(0)u + o(u)) du \right) \right] \delta \\ &= \frac{1}{2n} \delta^T \left[\sum_{t=1}^n w_t X_t X_t^T \left(\frac{1}{\sqrt{2\pi\sigma_t}} \int \exp \left\{ -\frac{u^2}{2\sigma_t^2} \right\} f(0) du \right) \right] \delta + \frac{1}{2n} \delta^T \left[\sum_{t=1}^n w_t X_t X_t^T \left(\frac{1}{\sqrt{2\pi\sigma_t}} \int \exp \left\{ -\frac{u^2}{2\sigma_t^2} \right\} f'(0)u du \right) \right] \delta (1 + o(1)) \\ &= \frac{1}{2n} \delta^T \left[\sum_{t=1}^n w_t X_t X_t^T f(0) \left(\frac{1}{\sqrt{2\pi\sigma_t}} \int \exp \left\{ -\frac{u^2}{2\sigma_t^2} \right\} du \right) \right] \delta + \frac{1}{2n} \delta^T \left[\sum_{t=1}^n w_t X_t X_t^T f'(0) \left(\frac{1}{\sqrt{2\pi\sigma_t}} \int \exp \left\{ -\frac{u^2}{2\sigma_t^2} \right\} u du \right) \right] \delta (1 + o(1)) \\ &= \frac{1}{2n} \delta^T \left[\sum_{t=1}^n w_t X_t X_t^T f(0) \right] \delta \rightarrow \frac{1}{2} \delta^T E \left[w_t X_t X_t^T f(0) \right] \delta \end{aligned}$$

We next show that $\text{Var}(Z_{2n}(\delta) - E(Z_{2n}(\delta) | \mathcal{F}_t))$ converges to 0.

$$\text{var}(Z_{2n}(\delta) - E(Z_{2n}(\delta)|\mathcal{F}_{t-1})) = E(Z_{2n}(\delta) - E(Z_{2n}(\delta)|\mathcal{F}_{t-1}))^2 = E\left(\sum_{t=1}^n (Z_{2nt}(\delta) - E(Z_{2nt}(\delta)|\mathcal{F}_{t-1}))^2\right)$$

Similar to the proof of Proposition 1, we can show that

$$\begin{aligned} E\left(\sum_{t=1}^n (Z_{2nt}(\delta) - E(Z_{2nt}(\delta)|\mathcal{F}_{t-1}))^2\right) &= \sum_{t=1}^n E(Z_{2nt}(\delta) - E(Z_{2nt}(\delta)|\mathcal{F}_{t-1}))^2 \\ &\leq 2 \sum_{t=1}^n E Z_{2nt}^2(\delta) = 2 \sum_{t=1}^n E\left(\delta^T \frac{1}{2n} w_t X_t X_t^T \frac{\varphi(b_t)}{\sigma_t} \delta\right)^2 \\ &= \frac{1}{2n^2} \sum_{t=1}^n E\left(\delta^T w_t X_t X_t^T \frac{\varphi(b_t)}{\sigma_t} \delta\right)^2 \\ &\leq \frac{1}{2n^2} \sum_{t=1}^n E\left(\delta^T w_t X_t X_t^T \frac{\varphi(b_t)}{\sigma_t} \delta\right) \cdot \sup_t g(t) \cdot E w_t \|X_t\|^2 \cdot \max_{\delta} \|\delta\|^2 \\ &= \frac{1}{2n} \left[\frac{1}{n} \sum_{t=1}^n E\left(\delta^T w_t X_t X_t^T \frac{\varphi(b_t)}{\sigma_t} \delta\right) \right] \cdot \sup_t g(t) \cdot E w_t \|X_t\|^2 \cdot \max_{\delta} \|\delta\|^2 \\ &= \frac{1}{2n} E[E(Z_{2n}(\delta)|\mathcal{F}_{t-1})] \cdot \sup_t g(t) \cdot E w_t \|X_t\|^2 \cdot \max_{\delta} \|\delta\|^2 \end{aligned}$$

where $g(u)$ is a density function of normal distribution with mean 0 and variance σ_t^2 . Thus, $g(u)$ is a bounded function. By Assumption 3, $E w_t \|X_t\|^2 < \infty$. It implies that $\text{Var}(Z_{2n}(\delta) - E(Z_{2n}(\delta)|\mathcal{F}_t))$ converges to 0. Thus, for any $\delta \in R^{p+1}$, $Z_{2n} \rightarrow \frac{1}{2} \delta^T E[w_t X_t X_t^T f(0)] \delta$. This combines with Theorem 1(A) and the Proposition 1, it follows that

$$\sqrt{n}(\tilde{\varphi}_n - \varphi) \xrightarrow{\mathcal{L}} N(0, \tau(1-\tau)f^{-2}(0) \sum^{-1} \Omega \sum^{-1}).$$

This completes the proof of Theorem 1(B).

Figure 1. Correlogram of log-price. The upper panel is the auto correlation and the lower one is partial auto correlation for log-price

Figure 2.
Hill estimators for left-hand tail(dashed curve) and right-hand tail(solid curve).

Figure 3.

The 'o' is the scatter of y_t versus y_{t-1} . The solid lines are quantile regression lines with $\tau = 0.9$, $\tau = 0.5$, $\tau = 0.1$ from top to bottom, respectively. The dashed lines are the linear regression line and its 80% pointwise prediction intervals for the response variable based on least squares method.

Table 1

Robustness to Choice of $J^{(0)}$ ($\tau = 0.5$, $n = 200$, $\kappa = 0.05$)

Cauchy			t_2			$N(0, 1)$		
γ_n	$\tilde{\gamma}_n$	$\tilde{\gamma}_1$	$\tilde{\gamma}_n$	$\tilde{\gamma}_1$	$\tilde{\gamma}_n$	$\tilde{\gamma}_n$	$\tilde{\gamma}_1$	$\tilde{\gamma}_1$
n^{-1}	Mean	0.003	0.498	0.002	0.492	0.004	0.491	
	SD	0.119	0.023	0.097	0.053	0.087	0.078	
	AD	0.122	0.025	0.105	0.056	0.090	0.083	
$n^{-1.5}$	Mean	0.006	0.498	0.004	0.493	-0.002	0.487	
	SD	0.116	0.022	0.100	0.052	0.089	0.079	
	AD	0.120	0.025	0.105	0.055	0.092	0.084	
n^{-2}	Mean	0.001	0.498	0.001	0.493	-0.003	0.484	
	SD	0.118	0.024	0.101	0.052	0.088	0.078	
	AD	0.120	0.025	0.105	0.056	0.090	0.083	
$n^{-2.5}$	Mean	0.005	0.498	-0.001	0.492	0.002	0.488	
	SD	0.113	0.022	0.106	0.053	0.092	0.079	
	AD	0.121	0.025	0.106	0.057	0.091	0.084	

Table 2

Robustness to Choice of κ ($\tau = 0.5$)

n	Cauchy			t_2			$N(0, 1)$		
	$\tilde{\varphi}_0$	$\tilde{\varphi}_1$	$\tilde{\varphi}_2$	$\tilde{\varphi}_0$	$\tilde{\varphi}_1$	$\tilde{\varphi}_2$	$\tilde{\varphi}_0$	$\tilde{\varphi}_1$	$\tilde{\varphi}_2$
$\kappa = 0.10$									
200	Mean	-0.000	0.496	0.001	0.493	0.001	-0.003	0.485	0.485
	SD	0.124	0.035	0.103	0.062	0.089	0.089	0.086	0.086
	AD	0.125	0.037	0.108	0.067	0.092	0.092	0.091	0.091
400	Mean	0.006	0.499	-0.003	0.496	-0.001	0.001	0.494	0.494
	SD	0.084	0.023	0.072	0.043	0.064	0.064	0.062	0.062
	AD	0.086	0.025	0.075	0.046	0.065	0.065	0.064	0.064
$\kappa = 0.05$									
200	Mean	0.001	0.498	0.001	0.493	-0.003	0.484	0.484	0.484
	SD	0.118	0.024	0.101	0.052	0.088	0.078	0.078	0.078
	AD	0.120	0.025	0.105	0.056	0.090	0.083	0.083	0.083
400	Mean	0.005	0.499	-0.003	0.497	-0.001	0.494	0.494	0.494
	SD	0.080	0.016	0.069	0.035	0.063	0.057	0.057	0.057
	AD	0.083	0.016	0.073	0.039	0.064	0.059	0.059	0.059
Ling (2005)'s method									
200	Mean	0.009	0.499	0.001	0.494	0.001	0.491	0.491	0.491
	SD	0.121	0.023	0.101	0.054	0.090	0.082	0.082	0.082
	AD	0.123	0.021	0.108	0.055	0.092	0.084	0.084	0.084
400	Mean	0.005	0.500	0.005	0.495	0.001	0.493	0.493	0.493
	SD	0.082	0.015	0.074	0.038	0.064	0.058	0.058	0.058
	AD	0.086	0.015	0.075	0.038	0.064	0.059	0.059	0.059

Table 3

Simulation results for $\tau = 0.1, 0.25, 0.75, 0.9$

n	τ	Cauchy			t_2	$N(0, 1)$		
		$\tilde{\varphi}_0$	$\tilde{\varphi}_1$	$\tilde{\varphi}_2$		$\tilde{\varphi}_0$	$\tilde{\varphi}_1$	$\tilde{\varphi}_2$
200	0.10	Mean	-0.297	0.499	-0.034	0.485	0.026	0.485
		SD	1.096	0.115	0.593	0.134	0.295	0.107
		AD	1.043	0.112	0.578	0.135	0.323	0.117
400	0.10	Mean	-0.104	0.489	-0.034	0.494	-0.010	0.500
		SD	0.744	0.078	0.408	0.095	0.197	0.071
		AD	0.678	0.072	0.413	0.097	0.220	0.079
200	0.25	Mean	-0.049	0.495	-0.006	0.490	0.012	0.485
		SD	0.248	0.058	0.192	0.080	0.155	0.093
		AD	0.243	0.059	0.201	0.087	0.167	0.099
400	0.25	Mean	-0.013	0.495	-0.011	0.496	-0.002	0.498
		SD	0.172	0.038	0.133	0.055	0.107	0.064
		AD	0.167	0.040	0.139	0.060	0.118	0.069
200	0.75	Mean	0.044	0.497	0.007	0.489	-0.006	0.495
		SD	0.244	0.055	0.199	0.080	0.159	0.090
		AD	0.245	0.057	0.199	0.087	0.167	0.099
400	0.75	Mean	0.026	0.496	0.007	0.497	-0.003	0.494
		SD	0.164	0.039	0.143	0.058	0.115	0.067
		AD	0.170	0.041	0.140	0.061	0.119	0.070
200	0.90	Mean	0.258	0.493	0.051	0.489	-0.029	0.485
		SD	1.065	0.108	0.573	0.131	0.301	0.108
		AD	1.026	0.107	0.593	0.138	0.314	0.114
400	0.90	Mean	0.141	0.497	0.018	0.491	0.000	0.498
		SD	0.741	0.075	0.400	0.095	0.199	0.070
		AD	0.685	0.074	0.412	0.096	0.223	0.080

Table 4

Sample Standard Deviation of the 1000 Estimated SEs

n	τ	Cauchy				t_2				$N(0, 1)$			
		$\tilde{\phi}_0$	$\tilde{\phi}_1$	$\tilde{\phi}_0$	$\tilde{\phi}_1$	$\tilde{\phi}_0$	$\tilde{\phi}_1$	$\tilde{\phi}_0$	$\tilde{\phi}_1$	$\tilde{\phi}_0$	$\tilde{\phi}_1$	$\tilde{\phi}_0$	$\tilde{\phi}_1$
200	0.10	1.3783	0.0156	0.2793	0.0107	0.0533	0.0059						
	0.25	0.0254	0.0011	0.0160	0.0027	0.0125	0.0035						
	0.50	0.0051	0.0005	0.0042	0.0014	0.0029	0.0026						
	0.75	0.0254	0.0012	0.0153	0.0025	0.0120	0.0034						
	0.90	1.1475	0.0140	0.2094	0.0100	0.0539	0.0067						
400	0.10	0.3309	0.0041	0.0881	0.0045	0.0223	0.0027						
	0.25	0.0087	0.0014	0.0062	0.0010	0.0050	0.0014						
	0.50	0.0022	0.0003	0.0018	0.0007	0.0014	0.0013						
	0.75	0.0086	0.0014	0.0064	0.0010	0.0049	0.0015						
	0.90	0.3046	0.0033	0.0879	0.0043	0.0236	0.0032						

Table 5

Sizes and powers of the Wald's test for $H_0: \phi_1 = 0.5$

n	τ	ϕ_0	ϕ_1	Cauchy				$t(2)$				$N(0, 1)$			
				$\alpha =$	0.10	0.05	0.01	0.10	0.05	0.01	0.05	0.10	0.05	0.01	0.01
200	0.25	0.0	0.3	0.969	0.950	0.887	0.804	0.701	0.489	0.682	0.564	0.335			
			0.4	0.761	0.702	0.516	0.400	0.283	0.130	0.311	0.222	0.089			
			0.0	0.5	0.111	0.060	0.012	0.084	0.039	0.009	0.101	0.057	0.018		
	0.5	0.0	0.6	0.912	0.863	0.740	0.491	0.370	0.194	0.315	0.240	0.109			
			0.7	0.997	0.994	0.989	0.931	0.893	0.806	0.798	0.704	0.489			
			0.0	0.3	0.999	0.998	0.989	0.946	0.901	0.756	0.771	0.654	0.410		
200	0.5	0.0	0.4	0.956	0.928	0.817	0.575	0.475	0.221	0.356	0.258	0.108			
			0.0	0.5	0.088	0.038	0.008	0.091	0.045	0.005	0.106	0.061	0.019		
			0.0	0.6	0.986	0.976	0.939	0.647	0.532	0.316	0.314	0.212	0.090		
	0.75	0.0	0.7	1.000	1.000	1.000	0.992	0.982	0.946	0.834	0.757	0.554			
			0.3	0.969	0.951	0.876	0.799	0.704	0.490	0.688	0.571	0.317			
			0.4	0.772	0.693	0.521	0.416	0.285	0.134	0.322	0.212	0.075			
200	0.75	0.0	0.5	0.106	0.061	0.015	0.109	0.059	0.016	0.093	0.052	0.017			
			0.6	0.904	0.864	0.747	0.457	0.351	0.181	0.298	0.201	0.079			
			0.7	0.998	0.997	0.994	0.945	0.913	0.810	0.785	0.713	0.527			
	0.0	0.3	0.999	0.998	0.991	0.974	0.945	0.831	0.923	0.859	0.672				
			0.4	0.938	0.902	0.776	0.639	0.520	0.280	0.483	0.358	0.163			
			0.5	0.100	0.053	0.012	0.101	0.059	0.020	0.098	0.051	0.015			
400	0.25	0.0	0.6	0.987	0.976	0.948	0.711	0.595	0.369	0.530	0.389	0.203			
			0.7	1.000	1.000	1.000	0.996	0.995	0.984	0.973	0.943	0.844			
			0.0	0.3	1.000	1.000	0.998	0.994	0.974	0.952	0.894	0.753			
	0.5	0.0	0.4	0.997	0.992	0.980	0.796	0.698	0.458	0.518	0.401	0.197			
			0.5	0.096	0.045	0.009	0.116	0.062	0.018	0.113	0.060	0.013			
			0.6	1.000	1.000	0.996	0.861	0.799	0.635	0.554	0.427	0.230			
400	0.75	0.0	0.7	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.981	0.961	0.887		
			0.3	0.999	0.998	0.991	0.974	0.945	0.831	0.923	0.859	0.672			
			0.4	0.938	0.902	0.776	0.639	0.520	0.280	0.483	0.358	0.163			
	0.0	0.5	0.100	0.053	0.012	0.101	0.059	0.020	0.098	0.051	0.015				
			0.6	0.987	0.976	0.948	0.711	0.595	0.369	0.530	0.389	0.203			
			0.7	1.000	1.000	1.000	0.996	0.995	0.984	0.973	0.943	0.844			

<i>n</i>	<i>z</i>	ϕ_0	ϕ_1	Cauchy			<i>t</i> (2)			<i>N</i> (0, 1)		
				0.10	0.05	0.01	0.10	0.05	0.01	0.10	0.05	0.01
			$\alpha =$									
400	0.75	0.0	0.3	0.998	0.995	0.985	0.969	0.936	0.833	0.916	0.852	0.635
		0.0	0.4	0.929	0.892	0.781	0.607	0.478	0.268	0.485	0.365	0.182
		0.0	0.5	0.094	0.051	0.008	0.097	0.054	0.013	0.093	0.052	0.010
		0.0	0.6	0.990	0.981	0.946	0.689	0.576	0.361	0.498	0.382	0.216
		0.0	0.7	1.000	1.000	1.000	0.998	0.992	0.978	0.965	0.939	0.834

Table 6

Estimated coefficients and their standard errors

	$\tilde{\varphi}_0$	$\tilde{\varphi}_1$
$\tau = 0.1$	1.939(0.11)	0.405(0.03)
$\tau = 0.2$	1.577(0.15)	0.535(0.04)
$\tau = 0.3$	1.640(0.13)	0.548(0.03)
$\tau = 0.4$	1.500(0.18)	0.610(0.04)
$\tau = 0.5$	1.553(0.17)	0.625(0.04)
$\tau = 0.6$	1.864(0.19)	0.575(0.04)
$\tau = 0.7$	2.078(0.17)	0.550(0.04)
$\tau = 0.8$	2.150(0.18)	0.566(0.04)
$\tau = 0.9$	2.355(0.13)	0.558(0.03)