Linear double autoregression

Qianqian Zhu[a], Yao Zheng[b,*] and Guodong Li[b]

[a] *School of Statistics and Management, Shanghai University of Finance and Economics, Shanghai, China*

[b] *Department of Statistics and Actuarial Science, University of Hong Kong, Pokfulam Road, Hong Kong, China*

Abstract

This paper proposes the linear double autoregression, a conditional heteroscedastic model with a conditional mean structure but compatible with the quantile regression. The existence of a strictly stationary solution is discussed, for which a necessary and sufficient condition is established. A doubly weighted quantile regression estimation procedure is introduced, where the first set of weights ensures the asymptotic normality of the estimator and the second set improves its efficiency through balancing individual quantile regression estimators across multiple quantile levels. Bayesian information criteria are proposed for model selection, and two goodness-of-fit tests are constructed to check the adequacy of the fitted conditional mean and conditional scale structures. Simulation studies indicate that the proposed inference tools perform well in finite samples, and an empirical example illustrates the usefulness of the new model.

JEL classifications: C15; C22.

Key words: Conditional quantile estimation; Goodness-of-fit test; Heavy tail; Nonlinear time series model; Stationary solution.

*Correspondence to: Department of Statistics and Actuarial Science, University of Hong Kong, Pokfulam Road, Hong Kong. Email address: yaozheng@connect.hku.hk (Y. Zheng).

1 Introduction

Since the appearance of autoregressive conditional heteroscedastic (ARCH) and the generalized autoregressive conditional heteroscedastic (GARCH) models (Engle, 1982; Bollerslev, 1986), conditional heteroscedastic models have become extremely popular in volatility and financial risk modeling. In particular, they have been widely used for the prediction of quantile-based risk measures, e.g., the value at risk. Hence, it is natural to consider the quantile regression (Koenker and Bassett, 1978) for conditional heteroscedastic models; see, e.g., Engle and Manganelli (2004).

In the literature of quantile regression methods for conditional heteroscedastic models, for numerical feasibility, it is often assumed that the conditional standard deviation rather than the conditional variance of the model has a linear structure, which allows the linear programming (Koenker, 2005) to be used for efficient optimization; see, e.g., the linear ARCH model studied by Koenker and Zhao (1996), the linear GARCH model by Xiao and Koenker (2009) and the double-threshold ARCH model by Jiang et al. (2014). Moreover, this structure can result in more robust inference than the linear structure for the conditional variance (Xiao and Koenker, 2009). Nevertheless, when there is a conditional mean component, new challenges will arise. To see this, consider a simple autoregressive (AR) model with linear ARCH errors: $y_t = \phi y_{t-1} + e_t$, $e_t = \varepsilon_t \sigma_t$, $\sigma_t = \beta_0 + \beta_1 |e_{t-1}|$. The corresponding quantile regression can be defined as

$$\min_{\theta} \sum_{t=1}^{n} \rho_\tau (y_t - \phi y_{t-1} - b\beta_0 - b\beta_1 |y_{t-1} - \phi y_{t-2}|),$$

where $\theta = (b, \beta_0, \beta_1, \phi)'$, $\tau \in (0,1)$ is the quantile level, and $\rho_\tau(x) = x\{\tau - I(x < 0)\}$ is the check function. Because of the term $|y_{t-1} - \phi y_{t-2}|$, the above objective function is non-convex, causing difficulties for statistical inference and numerical optimization. This paper proposes a new conditional heteroscedastic model with a conditional mean structure but highly tractable for the quantile regression. The corresponding inference requires no moment restriction on the observed process or the innovations, and hence can realize the full potential of the quantile regression from a robustness perspective.

The proposed model is the linear double AR model, which adopts the basic form of the double autoregression introduced by Ling (2007) to make the conditional mean structure especially tractable for quantile inference and, at the same time, assumes a

linear structure for the conditional standard deviation. Recently the double AR model has attracted growing interest; see Ling and Li (2008), Zhu and Ling (2013) and the references therein. It has the form of

$$y_t = \sum_{i=1}^{p} \phi_i y_{t-i} + \varepsilon_t \sqrt{\beta_0 + \sum_{j=1}^{p} \beta_j y_{t-j}^2}, \tag{1.1}$$

where $\beta_0 > 0$, $\beta_j \geq 0$ for $1 \leq j \leq p$, and $\{\varepsilon_t\}$ are independent and identically distributed ($i.i.d.$) innovations with mean zero and variance one. Model (1.1) has two novel properties. First, it has a larger parameter space than conventional AR models. For example, model (1.1) with $p = 1$ may still be stationary even when $|\phi_1| \geq 1$ (Ling, 2004), which is impossible for AR-ARCH models. Second, it usually requires no moment condition on $\{y_t\}$ for the asymptotic normality of its parameter estimator (Ling, 2007). In contrast, for the ARMA-GARCH model, a finite fourth moment of the observed process is required for the asymptotic normality of the Gaussian quasi-maximum likelihood estimator (Francq and Zakoian, 2004), resulting in a much narrower parameter space (Li and Li, 2009). Similar to the double AR model, the proposed linear double AR model enjoys both novel properties. In particular, we establish a necessary and sufficient stationarity condition by borrowing the linearity of the random coefficient AR model.

Although the quantile regression is well known for its robustness against heavy tails, its efficiency at certain quantile levels can be arbitrarily low. The composite quantile regression (CQR) was proposed to improve the efficiency by combining multiple quantile levels (Koenker, 1984; Zou and Yuan, 2008). As argued in Jiang et al. (2012), by choosing the optimal weights, the weighted CQR estimator can be nearly as efficient as the maximum likelihood estimator (MLE); see also Jiang et al. (2014). However, the CQR for the proposed model is time-consuming due to the non-convexity of the objective function. Zhao and Xiao (2014) suggested using weighted averages of quantile regression estimators at different quantile levels and their simulation studies showed that the averaging estimator is more efficient than the CQR estimator. Chen et al. (2016) considered more general weights and the resulting estimator is hence supposed to be even more efficient. On the other hand, the consistency of the usual quantile regression estimator for the proposed model requires the observed process to have a finite first moment; see Section 3.1. To avoid such moment conditions, Ling (2005) proposed a self-weighted estimation

method for the infinite variance AR model; see also Zhu and Ling (2011). Motivated by Ling (2005) and Chen et al. (2016), we eliminate any moment condition on the observed process or the innovations by introducing a double weighting scheme, where the first set of weights guarantees the asymptotic normality, while the second set improves the efficiency through balancing the information across multiple quantile levels. As a result, the proposed model can handle more heavy-tailed data, as opposed to existing inference tools for conditional heteroscedastic models which all require the innovations to have at least a finite second moment. Moreover, the optimal doubly weighted estimator can approach the efficiency of the MLE under certain conditions.

To select the order of the proposed model in practice, Bayesian information criteria (BIC) are proposed in the quantile regression context. Furthermore, based on the quantile autocorrelation function (Li et al., 2015) for transformed residuals, two goodness-of-fit tests are constructed to detect misspecifications in the conditional mean and the conditional scale separately for the fitted model. Along the lines of robust inference, no further moment condition is required by the information criteria and goodness-of-fit tests. In this paper, for a matrix or column vector A, we define $\|A\| = \sqrt{\mathrm{tr}(AA')}$, where $\mathrm{tr}(\cdot)$ denotes the trace of a square matrix. For two matrices $A = (A_{ij})$ and $B = (B_{ij})$ with the same dimension, we define the element-wise product $A \circ B$ by $(A \circ B)_{i,j} = A_{ij}B_{ij}$, and define $A > B$ if $A - B$ is positive definite.

2 Linear double autoregression

Consider the linear double AR model,

$$y_t = \sum_{i=1}^{p} \phi_i y_{t-i} + \varepsilon_t \left(1 + \sum_{j=1}^{p} \beta_j |y_{t-j}| \right), \tag{2.1}$$

where the integer p is the order, $\beta_j \geq 0$ for $1 \leq j \leq p$, and $\{\varepsilon_t\}$ is a sequence of $i.i.d.$ innovations. When $E(\varepsilon_t^2) < \infty$, by further assuming that $E(\varepsilon_t) = 0$, the innovations can be standardized to have variance one, and model (2.1) can be rewritten as

$$y_t = \sum_{i=1}^{p} \phi_i y_{t-i} + \varepsilon_t^* \left(\sigma + \sum_{j=1}^{p} \varrho_j |y_{t-j}| \right), \tag{2.2}$$

where $\sigma = \{E(\varepsilon_t^2)\}^{1/2}$, $\varepsilon_t^* = \varepsilon_t/\sigma$, $\varrho_j = \sigma\beta_j$, $E(\varepsilon_t^*) = 0$ and $E(\varepsilon_t^{*2}) = 1$. The linear double AR model in the form of (2.2) is hence an extension of the double AR model in

(1.1) along the lines of the linear GARCH model. We aim to study model (2.1) without any moment or location restriction on ε_t.

We first consider the case where ε_t follows the Cauchy distribution with location zero and scale $\sigma > 0$, whose density function is $f(x) = \sigma/\{\pi(x^2 + \sigma^2)\}$ for $x \in \mathbb{R}$. Note that $E(|\varepsilon_t|) = \infty$ and $E(|\varepsilon_t|^\kappa) < \infty$ for any $0 < \kappa < 1$. Let $\{\xi_{it}, 1 \le i \le p, t \in \mathbb{Z}\}$ be a double array of independent random variables which have the same distribution as ε_t and are independent of $\{\varepsilon_t\}$. Consider the random coefficient AR model,

$$y_t^* = \sum_{i=1}^{p} (\phi_i + \beta_i \xi_{it}) y_{t-i}^* + \varepsilon_t, \tag{2.3}$$

where the ϕ_i's, β_i's and ε_t are from model (2.1). Let $Y_t = (y_t, \ldots, y_{t-p+1})'$ and $Y_t^* = (y_t^*, \ldots, y_{t-p+1}^*)'$, where $\{y_t\}$ and $\{y_t^*\}$ are generated by models (2.1) and (2.3), respectively. Noting that the characteristic function of ε_t is $E\{\exp(is\varepsilon_t)\} = \exp(-\sigma|s|)$, we can verify that $\{Y_t\}$ and $\{Y_t^*\}$ are Markov chains with the same transition probability. This observation enables us to derive a necessary and sufficient condition for the existence of a strictly stationary solution to model (2.1) by borrowing the linearity of model (2.3).

Let $\{A_t\}$ be a sequence of random matrices with

$$A_t = \begin{pmatrix} \phi_1 + \beta_1 \xi_{1t} & \cdots & \phi_{p-1} + \beta_{p-1}\xi_{p-1,t} & \phi_p + \beta_p \xi_{pt} \\ & I_{p-1} & & 0 \end{pmatrix},$$

where I_m is the $m \times m$ identity matrix, and 0 is a zero vector or matrix with compatible dimensions. We define the top Lyapounov exponent of $\{A_t\}$ as

$$\gamma = \inf \left\{ \frac{1}{n} E(\ln \|A_1 \cdots A_n\|), n \ge 1 \right\}.$$

It can be shown that $E(\ln^+ \|A_1\|) < \infty$, where $\ln^+(x) = \max\{\ln(x), 0\}$. Then, by the subadditive ergodic theorem (Kingman, 1973), $\gamma = \lim_{n\to\infty} n^{-1} \ln \|A_1 \cdots A_n\|$ with probability one. In particular, $\gamma = E(\ln|\phi_1 + \beta_1 \xi_{1t}|)$ when $p = 1$.

Theorem 1. *If ε_t follows the Cauchy distribution with location zero and scale $\sigma > 0$, then there exists a strictly stationary solution $\{y_t\}$ to model (2.1) if and only if $\gamma < 0$, and this solution is unique and geometrically ergodic with $E(|y_t|^\kappa) < \infty$ for some $0 < \kappa < 1$.*

For other distributions for ε_t, it is generally challenging to derive a necessary and sufficient condition for the strict stationarity, as model (2.1) is actually nonlinear. Alternatively, a sufficient condition is provided below.

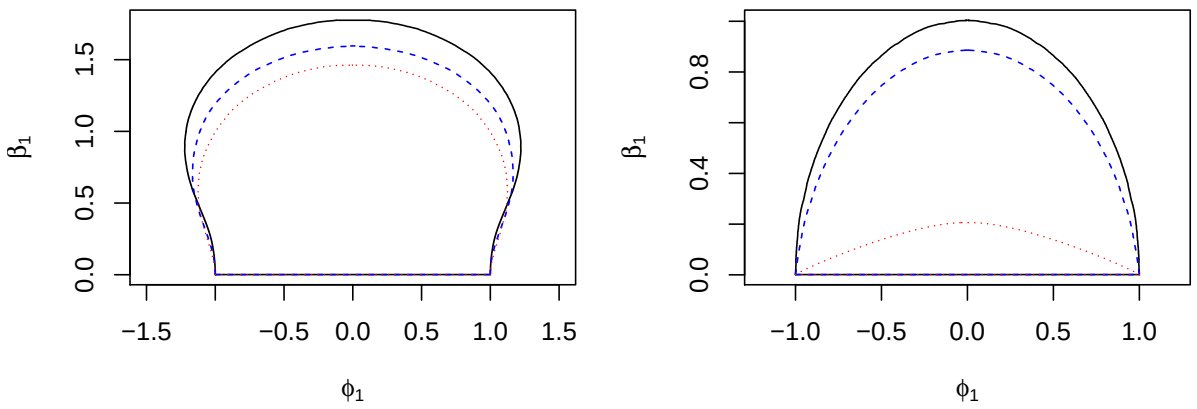

Figure 1: Stationarity regions of model (2.1) of order one. Left: $E(|\phi_1 + \beta_1\varepsilon_t|^\kappa) < 1$, where $\kappa = 0.1$ and ε_t follows the standard normal (solid line), Student's t_5 (dashed line) or Student's t_3 (dotted line) distribution. Right: $E(\ln|\phi_1 + \beta_1\varepsilon_t|) < 0$ (solid line), or $E(|\phi_1 + \beta_1\varepsilon_t|^\kappa) < 1$ with $\kappa = 0.1$ (dashed line) or 0.9 (dotted line), where ε_t follows the standard Cauchy distribution.

Assumption 1. *The density function of ε_t is continuous and positive everywhere on $\mathbb{R}$, and $E(|\varepsilon_t|^\kappa) < \infty$ for some $0 < \kappa \leq 1$.*

Theorem 2. *Under Assumption 1, if $\sum_{i=1}^p \max\{E(|\phi_i - \beta_i\varepsilon_t|^\kappa), E(|\phi_i + \beta_i\varepsilon_t|^\kappa)\} < 1$, then there exists a strictly stationary solution $\{y_t\}$ to model (2.1), and this solution is unique and geometrically ergodic with $E(|y_t|^\kappa) < \infty$.*

The stationarity region in Theorem 2 depends on the distribution of ε_t and implies a moment condition on y_t. In addition, when ε_t has a symmetric distribution, it simplifies to $\sum_{i=1}^p E(|\phi_i + \beta_i\varepsilon_t|^\kappa) < 1$. For illustration, the left panel of Figure 1 shows that model (2.1) with order $p = 1$ can be stationary even if $|\phi_1| \geq 1$, a feature inherited from the double AR model (Ling, 2004), and the right panel of the figure displays the different stationarity regions given by Theorems 1 and 2.

3 Doubly weighted quantile regression estimation

3.1 Self-weighted quantile regression estimation

Let $\lambda = (\beta', \phi')'$ be the parameter vector of model (2.1), where $\beta = (\beta_1, \ldots, \beta_p)'$ and $\phi = (\phi_1, \ldots, \phi_p)'$. We assume that the true parameter vector $\lambda_0 = (\beta_0', \phi_0')'$ is in the interior

of the parameter space Λ, where Λ is a compact subset of $\mathbb{R}_+^p \times \mathbb{R}^p$ with $\mathbb{R}_+ = (0, \infty)$.

Let $Y_t = (y_t, ..., y_{t-p+1})'$, $Y_{a,t} = (|y_t|, ..., |y_{t-p+1}|)'$ and $x_t = (1, Y_{a,t-1}', Y_{t-1}')'$. Denote the density and distribution functions of ε_t by $f(\cdot)$ and $F(\cdot)$, respectively. For any $\tau \in (0,1)$, let b_τ be the τth quantile of ε_t. Let $\mathcal{F}_t$ be the σ-field generated by $\{y_s, s \le t\}$. Then the τth conditional quantile of y_t can be written as

$$Q_\tau(y_t \mid \mathcal{F}_{t-1}) = b_\tau + b_\tau Y_{a,t-1}'\beta_0 + Y_{t-1}'\phi_0, \tag{3.1}$$

which motivates us to consider the self-weighted quantile regression estimator

$$(\widetilde{b}_{\tau n}, \widetilde{\lambda}_{\tau n}') = \underset{b,\lambda}{\operatorname{argmin}} \sum_{t=p+1}^{n} w_t \rho_\tau(y_t - b - bY_{a,t-1}'\beta - Y_{t-1}'\phi), \tag{3.2}$$

where $\widetilde{\lambda}_{\tau n} = (\widetilde{\beta}_{\tau n}', \widetilde{\phi}_{\tau n}')'$, $\rho_\tau(x) = x\{\tau - I(x < 0)\}$ is the check function and $\{w_t\}$ are random weights; see also Ling (2005). Numerically, we can first compute the weighted linear quantile regression estimator

$$\widetilde{\theta}_{\tau n} = \underset{\theta}{\operatorname{argmin}} \sum_{t=p+1}^{n} w_t \rho_\tau(y_t - x_t'\theta), \tag{3.3}$$

where $\widetilde{\theta}_{\tau n} = (\widetilde{b}_{\tau n}^*, \widetilde{\beta}_{\tau n}^{*\prime}, \widetilde{\phi}_{\tau n}^{*\prime})'$. Then it follows that $\widetilde{b}_{\tau n} = \widetilde{b}_{\tau n}^*$, $\widetilde{\beta}_{\tau n} = \widetilde{b}_{\tau n}^{*-1}\widetilde{\beta}_{\tau n}^*$ if $\widetilde{b}_{\tau n}^* \neq 0$, and $\widetilde{\phi}_{\tau n} = \widetilde{\phi}_{\tau n}^*$.

When $w_t = 1$ for all t, the weighted estimator becomes the common quantile regression estimator, and its consistency requires that $E(|\varepsilon_t|) < \infty$ and $E(|y_t|) < \infty$, since $y_t - Q_\tau(y_t \mid \mathcal{F}_{t-1}) = (\varepsilon_t - b_\tau)(1 + Y_{a,t-1}'\beta_0)$. If $E(y_t^2) = \infty$, the estimator will have a slower convergence rate than $\sqrt{n}$ and a more complicated asymptotic distribution than the normal distribution; see Gross and Steiger (1979), An and Chen (1982) and Davis et al. (1992) for the least absolute deviations estimation of infinite variance AR models.

Let $\sigma_t = 1 + Y_{a,t-1}'\beta_0$, and define the matrices $\Omega_0(w) = E(\sigma_t^{-1} w_t x_t x_t')$ and $\Omega_1(w) = \Omega_0^{-1}(w)[E(w_t^2 x_t x_t')]\Omega_0^{-1}(w)$, where w in $\Omega_i(\cdot)$ indicates dependence on the weights $\{w_t\}$.

Assumption 2. *The sequence of random weights $\{w_t\}$ is strictly stationary and ergodic, and w_t is nonnegative and measurable with respect to $\mathcal{F}_{t-1}$ for each t. Moreover, $\Omega_0(w)$ is a positive definite matrix and $E(\|w_t Y_{t-1}\|^2) < \infty$.*

Assumption 3. *The density function $f(\cdot)$ is bounded, positive and uniformly continuous on $\{x \in \mathbb{R} : 0 < F(x) < 1\}$.*

The matrix $\Omega_0(w)$ is degenerate if y_t is non-negative (or non-positive) with probability one, and hence its positive definiteness requires $0 < F(0) < 1$. For a fixed $\tau \in (0,1)$, restrictions on $f(\cdot)$ in a neighborhood of b_τ will be sufficient to derive asymptotic properties for the self-weighted estimator (Li et al., 2015). In fact, Assumption 3 is imposed mainly for the discussion in the next subsection.

Lemma 1. *Under Assumptions 2 and 3,* $\sqrt{n}(\widetilde{\theta}_{\tau n} - \theta_{\tau 0}) \to N\{0, \tau(1-\tau)[f(b_\tau)]^{-2}\Omega_1(w)\}$ *in distribution as* $n \to \infty$*, where* $\theta_{\tau 0} = (b_\tau, b_\tau \beta_0', \phi_0')'$.

When $b_\tau = 0$, since $Q_\tau(y_t \mid \mathcal{F}_{t-1}) = Y_{t-1}'\phi_0$, the parameter vector β_0 is not estimable, although $\widetilde{\phi}_{\tau n}$ is still asymptotically normal. By Lemma 1 and the Delta method (van der Vaart, 1998, Chapter 3), we have the following theorem.

Theorem 3. *Suppose that Assumptions 2 and 3 hold. If* $b_\tau \neq 0$*, then* $\sqrt{n}(\widetilde{\lambda}_{\tau n} - \lambda_0) \to N\left\{0, \tau(1-\tau)\Sigma_1^{-1}(\tau)\Omega_2(w)\Sigma_1^{-1}(\tau)\right\}$ *in distribution as* $n \to \infty$*, where*

$$\Sigma_1(\tau) = f(b_\tau)\begin{pmatrix} b_\tau I_p & 0 \\ 0 & I_p \end{pmatrix} \quad \text{and} \quad \Omega_2(w) = \begin{pmatrix} -\beta_0 & I_p & 0 \\ 0 & 0 & I_p \end{pmatrix}\Omega_1(w)\begin{pmatrix} -\beta_0 & I_p & 0 \\ 0 & 0 & I_p \end{pmatrix}'.$$

Moreover, the matrices $\Omega_1(w)$ *and* $\Omega_2(w)$ *are minimized if* $w_t = \sigma_t^{-1}$ *for all* t.

For the random weights w_t, one feasible choice is $w_t = 1/(1 + \sum_{i=1}^p |y_{t-i}|)$, which satisfies Assumption 2. However, from Theorem 3, $\widetilde{\lambda}_{\tau n}$ is asymptotically most efficient when $w_t = \sigma_t^{-1}$. Thus, in practice, when the sample size is relatively large, we may use the weights $\{\widetilde{\sigma}_t^{-1}\}$ with $\widetilde{\sigma}_t = 1 + Y_{a,t-1}'\widetilde{\beta}^{int}$, where $\widetilde{\beta}^{int}$ is an initial estimator with $\widetilde{\beta}^{int} - \beta_0 = O_p(n^{-1/2})$; e.g., we may use

$$\widetilde{\beta}^{int} = \frac{\sum_{k=1}^K |\widetilde{b}_{\tau_k n}||\widetilde{\beta}_{\tau_k n}|}{\sum_{k=1}^K |\widetilde{b}_{\tau_k n}|} = \frac{\sum_{k=1}^K |\widetilde{\beta}^*_{\tau_k n}|}{\sum_{k=1}^K |\widetilde{b}_{\tau_k n}|}, \tag{3.4}$$

where $\widetilde{b}_{\tau_k n}$ and $\widetilde{\beta}_{\tau_k n}$ for $1 \le k \le K$ are the self-weighted estimators computed based on the initial weights $w_t^{int} = 1/(1 + \sum_{i=1}^p |y_{t-i}|)$ for different quantile levels; see also Zhao and Xiao (2014). Although Assumption 2 does not hold for $\{\widetilde{\sigma}_t^{-1}\}$, we can show that the weights $\{\sigma_t^{-1}\}$ and $\{\widetilde{\sigma}_t^{-1}\}$ will lead to the same asymptotic distribution of $\widetilde{\lambda}_{\tau n}$. On the other hand, when the sample size is relatively small, the weights $\{w_t^{int}\}$ may be preferable to $\{\widetilde{\sigma}_t^{-1}\}$; see the simulation experiments in the supplementary materials. In the rest of the paper, we will use $\{\widetilde{\sigma}_t^{-1}\}$ unless otherwise specified, and Ω_i for $i = 0, 1, 2$ will refer to the matrices $\Omega_i(w)$ with $w_t = \sigma_t^{-1}$ for all t. Note that $\Omega_1 = \Omega_0^{-1}$.

Theorem 3 also implies that, when the value of b_τ is near zero, the variance of $\widetilde{\beta}_{\tau n}$ can be so large that $\widetilde{\beta}_{\tau n}$ may even be negative. This motivates us to consider a doubly weighted quantile regression estimator in what follows.

3.2 Doubly weighted quantile regression estimation

We next introduce a more efficient estimator of λ_0 by balancing the information across K quantile levels: $\tau_k = k/(K+1)$ for $1 \le k \le K$, where K is a fixed integer.

Specifically, we combine the self-weighted quantile regression estimators $\{\widetilde{\lambda}_{\tau_k n}, 1 \le k \le K\}$ linearly to define the doubly weighted quantile regression estimator

$$\widehat{\lambda}_n = (\widehat{\beta}_n', \widehat{\phi}_n')' = \sum_{k=1}^{K} \pi_k \widetilde{\lambda}_{\tau_k n},$$

where the π_k's are $2p \times 2p$ weighting matrices with possibly negative entries satisfying

$$\sum_{k=1}^{K} \pi_k = I_{2p}; \tag{3.5}$$

see also Chen et al. (2016). Define the $K \times K$ matrix $\Gamma = (\Gamma_{ij})_{1 \le i,j \le K}$, with $\Gamma_{ij} = \min(\tau_i, \tau_j) - \tau_i \tau_j$, and let Γ_{ij}^{inv} be the (i,j)th element of Γ^{-1}. Denote

$$\mathcal{V}(\Pi) = \sum_{i=1}^{K} \sum_{j=1}^{K} \Gamma_{ij} \pi_i \Sigma_1^{-1}(\tau_i) \Omega_2 \Sigma_1^{-1}(\tau_j) \pi_j',$$

where $\Pi = (\pi_1, ..., \pi_K)$ is a $2p \times 2pK$ matrix.

Theorem 4. *Suppose that Assumptions 2 and 3 hold. If $b_{\tau_k} \neq 0$ for $1 \le k \le K$, then $\sqrt{n}(\widehat{\lambda}_n - \lambda_0) \to N\{0, \mathcal{V}(\Pi)\}$ in distribution as $n \to \infty$. Moreover, denote*

$$\pi_k^{opt} = \left[\sum_{i=1}^{K} \sum_{j=1}^{K} \Gamma_{ij}^{inv} \Sigma_1(\tau_i) \Omega_2^{-1} \Sigma_1(\tau_j) \right]^{-1} \left[\sum_{i=1}^{K} \Gamma_{ik}^{inv} \Sigma_1(\tau_i) \Omega_2^{-1} \right] \Sigma_1(\tau_k), \quad 1 \le k \le K,$$

and let $\Pi^{opt} = (\pi_1^{opt}, ..., \pi_K^{opt})$ be a $2p \times 2pK$ weighting matrix. Then we have $\Pi^{opt} = \arg\min_\Pi \mathcal{V}(\Pi)$, and the asymptotic variance of the optimal doubly weighted quantile regression estimator is $\mathcal{V}(\Pi^{opt}) = \left[\sum_{i=1}^{K} \sum_{j=1}^{K} \Gamma_{ij}^{inv} \Sigma_1(\tau_i) \Omega_2^{-1} \Sigma_1(\tau_j) \right]^{-1}$.

For simplicity, denote $g(\tau) = f(b_\tau)$ and $h(\tau) = b_\tau f(b_\tau)$. Let $g_K = (g(\tau_1), \ldots, g(\tau_K))'$ and $h_K = (h(\tau_1), \ldots, h(\tau_K))'$. Suppose that $f(\cdot)$ is twice differentiable on $\{x \in \mathbb{R} : 0 < F(x) < 1\}$ and its derivative function is $\dot{f}(\cdot)$. Define the 2×2 matrices

$$\mathcal{I}_K = \begin{pmatrix} \mathcal{I}_{K,s} & \mathcal{I}_{K,ls} \\ \mathcal{I}_{K,ls} & \mathcal{I}_{K,l} \end{pmatrix} \quad \text{and} \quad \mathcal{I} = \begin{pmatrix} \mathcal{I}_s & \mathcal{I}_{ls} \\ \mathcal{I}_{ls} & \mathcal{I}_l \end{pmatrix},$$

where $\mathcal{I}_{K,l} = g_K'\Gamma^{-1}g_K$, $\mathcal{I}_{K,s} = h_K'\Gamma^{-1}h_K$, $\mathcal{I}_{K,ls} = g_K'\Gamma^{-1}h_K$, $\mathcal{I}_l = \int_{\mathbb{R}}[\dot{f}(u)]^2/f(u)du$, $\mathcal{I}_s = \int_{\mathbb{R}}[f(u)+u\dot{f}(u)]^2/f(u)du$, and $\mathcal{I}_{ls} = \int_{\mathbb{R}}\dot{f}(u)[f(u)+u\dot{f}(u)]/f(u)du$. Under Assumption 3, we have $\lim_{K\to\infty}\mathcal{I}_{K,s} = \mathcal{I}_s$ and $\lim_{K\to\infty}\mathcal{I}_{K,l} = \mathcal{I}_l$; see Theorems 6.1 and 6.2 in Zhao and Xiao (2014). Similarly, we can show that $\lim_{K\to\infty}\mathcal{I}_{K,ls} = \mathcal{I}_{ls}$, and hence

$$\lim_{K\to\infty}\mathcal{V}(\Pi^{opt}) = \lim_{K\to\infty}\{[\mathcal{I}_K\otimes\iota_{p\times p}]\circ\Omega_2^{-1}\}^{-1} = \{[\mathcal{I}\otimes\iota_{p\times p}]\circ\Omega_2^{-1}\}^{-1},$$

where $\otimes$ is the Kronecker product, and $\iota_{m\times n}$ is an $m\times n$ matrix with all elements being one. Denote by $(\widehat{\sigma}_n^{MLE}, \widehat{\lambda}_n^{MLE\prime})$ the maximum likelihood estimator (MLE) of model (2.2), where the parameter vector is (σ, λ') and the density of ε_t^* is assumed to be known. It can be shown that $\sqrt{n}(\widehat{\lambda}_n^{MLE} - \lambda_0) \to N(0, \mathcal{V}^{MLE})$ in distribution as $n\to\infty$, where

$$\mathcal{V}^{MLE} = \begin{pmatrix} -\beta_0 & I_p & 0 \\ 0 & 0 & I_p \end{pmatrix}\left[\begin{pmatrix} \mathcal{I}_s\iota_{(p+1)\times(p+1)} & \mathcal{I}_{ls}\iota_{(p+1)\times p} \\ \mathcal{I}_{ls}\iota_{p\times(p+1)} & \mathcal{I}_l\iota_{p\times p}\end{pmatrix}\circ\Omega_0\right]^{-1}\begin{pmatrix} -\beta_0 & I_p & 0 \\ 0 & 0 & I_p \end{pmatrix}'.$$

Moreover, $\lim_{K\to\infty}\mathcal{V}(\Pi^{opt}) = \mathcal{V}^{MLE}$ under the conditions that $E(\sigma_t^{-2}Y_{t-1}) = 0$ and $E(\sigma_t^{-2}Y_{a,t-1}'Y_{t-1}) = 0$, which is the case where parameters in the conditional mean and conditional scale can be separately estimated without loss of efficiency. In particular, when all ϕ_i's in model (2.1) are zero, these conditions are satisfied as long as the distribution of ε_t is symmetric about zero. Otherwise, $\mathcal{V}(\Pi^{opt})$ may not be able to attain the Crámer-Rao lower bound. This can probably be solved by using a nonlinear combination of the estimators, and we leave it for future research.

Theorem 4 requires that $b_{\tau_k} \neq 0$ for all $1\le k\le K$, which is not guaranteed in practice. Let $\pi_k = (\pi_{k1}, \pi_{k2})$, where π_{k1} and π_{k2} are $2p\times p$ matrices. To make practical the proposed doubly weighted estimator, in addition to (3.5), we further impose that

$$\pi_{k1} = 0 \quad \text{if} \quad b_{\tau_k} = 0, \quad \text{for all} \quad 1\le k\le K. \tag{3.6}$$

The optimal weighting matrix Π^{opt} actually satisfies both (3.5) and (3.6), which means that we can conduct the estimation procedure without worrying whether $b_{\tau_k} = 0$. But, by a method similar to that of the proof of Theorem 4, it can be shown that Π^{opt} is no longer optimal under both constraints (3.5) and (3.6), as the matrices π_k^{opt}'s in general are not diagonal or even block diagonal. This may be regarded as a necessary consequence of the lack of information about the zeroness of the quantiles b_{τ_k}'s.

To estimate the optimal weighting matrices π_k^{opt}, we can first obtain an estimator of Ω_0 using sample averages, i.e., $\widetilde{\Omega}_0 = n^{-1}\sum_{t=p+1}^n(1+Y_{a,t-1}'\widetilde{\beta}^{int})^{-2}x_tx_t'$, where the

method in (3.4) can be used to calculate $\widetilde{\beta}^{int}$, and similarly ϕ_0 can be approximated by $\widetilde{\phi}^{int} = K^{-1}\sum_{k=1}^{K}\widetilde{\phi}_{\tau_k n}$. The estimators of Ω_1 and Ω_2 can then be constructed, denoted by $\widetilde{\Omega}_1$ and $\widetilde{\Omega}_2$, respectively. Define the error function

$$\varepsilon_t(\lambda) = (y_t - Y_{t-1}'\phi)/(1 + Y_{a,t-1}'\beta), \tag{3.7}$$

and then the residuals $\{\widetilde{\varepsilon}_t\}$ can be calculated by $\widetilde{\varepsilon}_t = \varepsilon_t(\widetilde{\lambda}^{int})$ with $\widetilde{\lambda}^{int} = (\widetilde{\beta}^{int\prime}, \widetilde{\phi}^{int\prime})'$. As a result, the density function $f(\cdot)$ can be estimated by the kernel density estimator $\widetilde{f}(x) = (nh)^{-1}\sum_{t=p+1}^{n} K\{(x - \widetilde{\varepsilon}_t)/h\}$, where $K(\cdot)$ is the kernel function and h is the bandwidth. By Lemma 1, b_{τ_k} can be estimated by $\widetilde{b}_{\tau_k n}$, and hence an estimator of $\Sigma_1(\tau_k)$ can be obtained, denoted by $\widetilde{\Sigma}_1(\tau_k)$. Consequently, a consistent estimator $\widehat{\Pi}^{opt}$ of Π^{opt} can be obtained.

Now we are ready to compute the proposed optimal doubly weighted estimator

$$\widehat{\lambda}_n^{opt} = (\widehat{\beta}_n^{opt\prime}, \widehat{\phi}_n^{opt\prime})' = \sum_{k=1}^{K} \widehat{\pi}_k^{opt}\widetilde{\lambda}_{\tau_k n}.$$

It can be verified that $\sqrt{n}(\widehat{\lambda}_n^{opt} - \lambda_0) \to N(0, \mathcal{V}(\Pi^{opt}))$ in distribution as $n \to \infty$. Accordingly the residuals can be calculated by $\widehat{\varepsilon}_t = \varepsilon_t(\widehat{\lambda}_n^{opt})$, and an estimator of b_τ can be defined as the τth sample quantile of $\{\widehat{\varepsilon}_t\}$, i.e., $\widehat{b}_{\tau n} = \inf\{x : \widehat{F}_n(x) \geq \tau\}$ with $\widehat{F}_n(x) = (n-p)^{-1}\sum_{t=p+1}^{n} I(\widehat{\varepsilon}_t \leq x)$. To estimate $\mathcal{V}(\Pi^{opt})$, we can update the estimator of λ_0 by $\widehat{\lambda}_n^{opt}$ and the residuals by $\{\widehat{\varepsilon}_t\}$. By a method similar to that for calculating $\widehat{\Pi}^{opt}$, we can obtain a consistent estimator of $\mathcal{V}(\Pi^{opt})$, denoted by $\widehat{\mathcal{V}}(\widehat{\Pi}^{opt})$.

3.3 Model selection

This subsection considers the selection of the order p for model (2.1) in practice. We first discuss the case for a certain quantile level $\tau \in (0,1)$. Note that, from (3.1),

$$y_t = b_\tau + b_\tau Y_{a,t-1}'\beta_0 + Y_{t-1}'\phi_0 + e_t, \quad \text{with} \quad e_t = (\varepsilon_t - b_\tau)\sigma_t.$$

Suppose that $\{\sigma_t\}$ are observable and $\varepsilon_t - b_\tau$ follows the asymmetric Laplace distribution with location zero, unknown scale $\sigma > 0$ and the density function $f(x) = \tau(1-\tau)\sigma^{-1}\exp[-\rho_\tau(x/\sigma)]$ (Koenker and Machado, 1999). Then, the MLE of $(b_\tau, \lambda_0')'$ will have the same formula as the self-weighted quantile regression estimator in (3.2) with $w_t = \sigma_t^{-1}$. This motivates us to define the Bayesian information criterion (BIC):

$$\mathrm{BIC}_\tau(p) = 2(n - p_{\max})\log\widetilde{\sigma}_{\tau n} + (2p+1)\log(n - p_{\max}), \tag{3.8}$$

where p is searched over $\{1, \ldots, p_{\max}\}$, with $p_{\max}$ being a predetermined number, and $\widetilde{\sigma}_{\tau n} = (n - p_{\max})^{-1} \sum_{t=p_{\max}+1}^{n} w_t \rho_\tau(y_t - x_t' \widetilde{\theta}_{\tau n})$ is the MLE of the scale σ, with $\widetilde{\theta}_{\tau n}$ calculated by (3.3) and the weights defined as $w_t = (\widetilde{\sigma}_t + c \sum_{j=1}^{p_{\max}} |y_{t-j}|)^{-1}$ for a very small but fixed positive number c.

The proposed doubly weighted estimation, however, does not have a corresponding likelihood function since it consists of multiple quantile regressions. Nevertheless, the BIC in (3.8) yields consistent estimators of the true order p_0 for all $\tau \in (0, 1)$, and this motivates us to introduce an information criterion by combining the BIC across $\tau_1, \ldots, \tau_K$. Notice that the weights π_k's in Section 3.2 are matrices and thus cannot be directly applied to the BIC. In practice, we may use the simple average, $\mathrm{BIC}_1(p) = K^{-1} \sum_{k=1}^{K} \mathrm{BIC}_{\tau_k}(p)$. In addition, by replacing the self-weighted estimator $\widetilde{\theta}_{\tau n}$ in (3.8) with the doubly weighted estimator $(\widehat{b}_{\tau n}, \widehat{b}_{\tau n} \widehat{\beta}_n^{opt\prime}, \widehat{\phi}_n^{opt\prime})'$, we can define another BIC, denoted by $\mathrm{BIC}_2(p)$. Let $\widehat{p}_{1n} = \mathrm{argmin}_{1 \le p \le p_{\max}} \mathrm{BIC}_1(p)$ and $\widehat{p}_{2n} = \mathrm{argmin}_{1 \le p \le p_{\max}} \mathrm{BIC}_2(p)$.

Theorem 5. *Under Assumption 3, if $p_{\max} \ge p_0$, then $P(\widehat{p}_{1n} = p_0) \to 1$ and $P(\widehat{p}_{2n} = p_0) \to 1$ as $n \to \infty$, where p_0 is the true order.*

In the proposed estimation procedure, the key reason that we need no moment condition on y_t is that the condition $E(\|w_t Y_{t-1}\|^2) < \infty$ in Assumption 2 holds true for $w_t = \sigma_t^{-1}$ when the order p is correctly specified. But, since $\beta_{0j} = 0$ for $j > p_0$, this is not the case when $p > p_0$. To ensure that no additional moment condition is required by the proposed BIC, we add a small number $c > 0$ to all the β_{0j}'s, leading to the weights defined earlier in this subsection. In practice, the effect of c is ignorable; see the second simulation experiment in Section 5 for details.

4 Goodness-of-fit tests

To check the adequacy of fitted linear double AR models, we adopt the quantile autocorrelation function (QACF) in Li et al. (2015) to construct two goodness-of-fit tests to detect misspecifications in the conditional mean and conditional scale separately.

To make the QACF robust to arbitrarily heavy-tailed innovations, we consider the transformed innovations $\{G(\varepsilon_t)\}$, where $G : \mathbb{R} \to \mathbb{R}$ is a predetermined, bounded and strictly increasing function. Noticing that $\psi_\tau(\varepsilon_t - b_\tau) = \psi_\tau[G(\varepsilon_t) - G(b_\tau)]$, where $\psi_\tau(x) =$

$\tau - I(x < 0)$, the QACF of $\{G(\varepsilon_t)\}$ at lag ℓ can be defined as

$$\rho_{\ell,\tau} = \text{qcor}_\tau\{G(\varepsilon_t), G(\varepsilon_{t-\ell})\} = \frac{E\{\psi_\tau(\varepsilon_t - b_\tau)[G(\varepsilon_{t-\ell}) - \mu_{G,1}]\}}{\sqrt{\tau - \tau^2}\sigma_{G,1}}, \quad \ell = 1, 2, \ldots,$$

where $\mu_{G,1} = E[G(\varepsilon_t)]$ and $\sigma^2_{G,1} = \text{var}[G(\varepsilon_t)]$. By replacing $G(\varepsilon_{t-\ell})$ with $G(\varepsilon^2_{t-\ell})$, a variant of $\rho_{\ell,\tau}$ can be defined as

$$r_{\ell,\tau} = \text{qcor}_\tau\{G(\varepsilon_t), G(\varepsilon^2_{t-\ell})\} = \frac{E\{\psi_\tau(\varepsilon_t - b_\tau)[G(\varepsilon^2_{t-\ell}) - \mu_{G,2}]\}}{\sqrt{\tau - \tau^2}\sigma_{G,2}}, \quad \ell = 1, 2, \ldots,$$

where $\mu_{G,2} = E[G(\varepsilon^2_t)]$ and $\sigma^2_{G,2} = \text{var}[G(\varepsilon^2_t)]$. Notice that if model (2.1) is correctly specified, then $\rho_{\ell,\tau} = 0$ and $r_{\ell,\tau} = 0$ for all ℓ and all τ.

Accordingly the residual QACFs at lag ℓ can be defined as

$$\widehat{\rho}_{\ell,\tau} = \frac{1}{\sqrt{(\tau - \tau^2)}\widehat{\sigma}_{G,1}} \frac{1}{n-p} \sum_{t=p+\ell+1}^{n} \psi_\tau(\widehat{\varepsilon}_t - \widehat{b}_{\tau n})\{G(\widehat{\varepsilon}_{t-\ell}) - \widehat{\mu}_{G,1}\}$$

and

$$\widehat{r}_{\ell,\tau} = \frac{1}{\sqrt{(\tau - \tau^2)}\widehat{\sigma}_{G,2}} \frac{1}{n-p} \sum_{t=p+\ell+1}^{n} \psi_\tau(\widehat{\varepsilon}_t - \widehat{b}_{\tau n})\{G(\widehat{\varepsilon}^2_{t-\ell}) - \widehat{\mu}_{G,2}\},$$

where $\widehat{\mu}_{G,m} = (n-p)^{-1}\sum_{t=p+1}^{n} G(\widehat{\varepsilon}^m_t)$ and $\widehat{\sigma}^2_{G,m} = (n-p)^{-1}\sum_{t=p+1}^{n}\{G(\widehat{\varepsilon}^m_t) - \widehat{\mu}_{G,m}\}^2$ for $m = 1$ and 2. The two residual QACFs $\widehat{\rho}_{\ell,\tau}$ and $\widehat{r}_{\ell,\tau}$ will be used to construct goodness-of-fit tests for the conditional mean and conditional scale structures, respectively; see Li and Li (2008) for tests based on the conventional sample autocorrelation function.

To combine the information from multiple quantile levels, for any lag ℓ, we can define

$$\widehat{\rho}_\ell = \max_{1\le k\le K} |\widehat{\rho}_{\ell,\tau_k}| \quad \text{and} \quad \widehat{r}_\ell = \max_{1\le k\le K} |\widehat{r}_{\ell,\tau_k}|.$$

Let $\widehat{\rho} = (\widehat{\rho}_1, ..., \widehat{\rho}_L)'$ and $\widehat{r} = (\widehat{r}_1, ..., \widehat{r}_L)'$, where L is a predetermined positive integer.

Assumption 4. *$G : \mathbb{R} \to \mathbb{R}$ is a bounded, strictly increasing and twice-differentiable function, with its derivatives of first and second orders, g and $\dot{g}$, satisfying that (i) $\sup_{x\in\mathbb{R}} g(x) < \infty$; (ii) $\sup_{x\in\mathbb{R}} xg(x) < \infty$; (iii) $\sup_{x\in\mathbb{R}} \dot{g}(x) < \infty$; (iv) $\sup_{x\in\mathbb{R}} x\dot{g}(x) < \infty$; and (v) $\sup_{x\in\mathbb{R}} x^2\dot{g}(x) < \infty$.*

For $m = 1$ and 2, let $G_m = (G(\varepsilon^m_{t-1}), \ldots, G(\varepsilon^m_{t-L}))'$, $\Omega_{3,m} = E[\sigma_t^{-1}x_t(G_m - \mu_{G,m}1_L)']$ with 1_L being an $L \times 1$ vector of ones, and $D_m(\tau) = (\widetilde{d}_{1,m}(\tau), \ldots, \widetilde{d}_{L,m}(\tau))$ with

$$\widetilde{d}_{\ell,m}(\tau) = f(b_\tau)\left(b_\tau E\left\{[G(\varepsilon^m_{t-\ell}) - \mu_{G,m}]\frac{Y'_{a,t-1}}{\sigma_t}\right\}, E\left\{[G(\varepsilon^m_{t-\ell}) - \mu_{G,m}]\frac{Y'_{t-1}}{\sigma_t}\right\}\right)'.$$

In addition, for $m = 1$ and 2 and $1 \leq i, j \leq K$, let $\Psi_m(\tau_i, \tau_j)$ be

$$\frac{\Gamma_{ij}\sigma^2_{G,m} I_L - D'_m(\tau_i)\Sigma_3(\tau_j)\Omega_1\Omega_{3,m} - \Omega'_{3,m}\Omega_1\Sigma'_3(\tau_i)D_m(\tau_j) + D'_m(\tau_i)\mathcal{V}(\Pi^{opt})D_m(\tau_j)}{\sqrt{(\tau_i - \tau_i^2)(\tau_j - \tau_j^2)}\sigma_{G,m}},$$

where Ω_1 and Γ_{ij} are defined in Section 3, I_L is the $L \times L$ identity matrix, and

$$\Sigma_3(\tau) = \sum_{k=1}^{K} [\min(\tau, \tau_k) - \tau\tau_k]\pi_k^{opt}\Sigma_1^{-1}(\tau_k) \begin{pmatrix} -\beta_0 & I_p & 0 \\ 0 & 0 & I_p \end{pmatrix}.$$

Theorem 6. *Under Assumption 4 and the conditions of Theorem 4, we have* $\sqrt{n}\widehat{\rho} \to \max_{1\leq k\leq K} |B_1(\tau_k)|$ *and* $\sqrt{n}\widehat{r} \to \max_{1\leq k\leq K} |B_2(\tau_k)|$ *in distribution as* $n \to \infty$*, where* $|x| = (|x_1|, ..., |x_L|)'$ *for* $x = (x_1, ..., x_L)' \in \mathbb{R}^L$*, and* $B_m(\tau_k)$ *with* $1 \leq k \leq K$ *are multivariate normal random vectors such that* $cov(B_m(\tau_i), B_m(\tau_j)) = \Psi_m(\tau_i, \tau_j)$*, for* $m \in \{1, 2\}$*.*

We can construct consistent estimators of the covariance matrix $\Psi_m(\tau_i, \tau_j)$ by a method similar to that for $\widehat{\mathcal{V}}(\widehat{\Pi}^{opt})$ in Section 3.2. Then, by generating a sequence of, say $B = 10000$, multivariate normal random numbers, we can approximate the asymptotic distributions in Theorem 6 and then obtain confidence bounds for $\widehat{\rho}_\ell$ and $\widehat{r}_\ell$.

To check the first L lags jointly, we suggest the Box-Pierce type test statistics $Q_1^{BP}(L) = n\sum_{\ell=1}^{L} \widehat{\rho}_\ell^2$ and $Q_2^{BP}(L) = n\sum_{\ell=1}^{L} \widehat{r}_\ell^2$, which, as $n \to \infty$, converge in distribution to $\sum_{\ell=1}^{L} \max_{1\leq k\leq K} B_{1,\ell}^2(\tau_k)$ and $\sum_{\ell=1}^{L} \max_{1\leq k\leq K} B_{2,\ell}^2(\tau_k)$, respectively, where $B_m(\tau) = (B_{m,1}(\tau), ..., B_{m,L}(\tau))'$ for $m = 1$ and 2.

In practice, we may use the distribution function of the standard Cauchy random variable as the transformation $G(\cdot)$. Our simulation experiments in the supplementary material indicate that it performs slightly better than several other transformations in finite samples.

5 Simulation experiments

This section presents three simulation experiments to evaluate the finite-sample performance of the proposed doubly weighted quantile regression estimator, model selection method and goodness-of-fit tests. In all experiments, we employ the quantile levels $\tau_k = k/10$ with $k = 1, \ldots, 9$.

The first experiment aims to examine the finite-sample performance of the doubly

Table 1: Biases (×10), ESDs (×10) and ASDs (×10) of the doubly weighted estimator $\widehat{\lambda}_n^{opt}$ when the innovations follow the normal, Student's t_3 or Cauchy distribution.

		Normal			t_3			Cauchy		
	n	Bias	ESD	ASD	Bias	ESD	ASD	Bias	ESD	ASD
β	200	-0.203	1.658	1.320	0.204	2.180	1.569	1.674	4.794	2.619
	500	-0.063	0.938	0.863	0.123	1.180	1.010	0.803	2.111	1.550
	1000	-0.008	0.618	0.619	0.066	0.780	0.716	0.337	1.224	1.074
ϕ	200	-0.106	1.082	0.889	-0.082	1.115	0.856	-0.081	0.575	0.430
	500	-0.057	0.630	0.592	-0.037	0.607	0.563	-0.022	0.272	0.255
	1000	-0.021	0.448	0.426	-0.030	0.421	0.403	-0.005	0.173	0.170

weighted quantile regression estimator $\widehat{\lambda}_n^{opt}$, for which the data generating process is

$$y_t = 0.2y_{t-1} + \varepsilon_t(1 + 0.5|y_{t-1}|),$$

where $\{\varepsilon_t\}$ are *i.i.d.* normal, Student's t_3 or Cauchy random variables with location zero and $E(|\varepsilon_t|^\kappa) = 1$ for $\kappa = 0.9$. The sample size is set to $n = 200$, 500 or 1000, with 1000 replications for each sample size. The self weights $\{\sigma_t^{-1}\}$ are approximated by $\{1/(1 + \widetilde{\beta}^{int}|y_{t-1}|)\}$, where $\widetilde{\beta}^{int}$ is calculated by (3.4). The density function of ε_t is estimated by the kernel density method with the Gaussian kernel and its rule-of-thumb bandwidth, $h = 0.9n^{-1/5}\min\{s, \widehat{R}/1.34\}$, where s and $\widehat{R}$ are the sample standard deviation and interquartile of the residuals, respectively; see Silverman (1986). Table 1 lists the biases, empirical standard deviations (ESDs) and asymptotic standard deviations (ASDs) of $\widehat{\lambda}_n^{opt}$ for different innovation distributions and sample sizes. As the sample size increases, most of the biases, ESDs and ASDs become smaller, and the ESDs get closer to the corresponding ASDs. Moreover, when the distribution of ε_t has heavier tails, all these quantities of $\widehat{\phi}_n^{opt}$ decrease, whereas those of $\widehat{\beta}_n^{opt}$ increase.

In the second experiment, we evaluate the performance of the proposed model selection method in Section 3.3, and the data generating process is

$$y_t = 0.1y_{t-1} + 0.3y_{t-2} + \varepsilon_t(1 + 0.1|y_{t-1}| + 0.3|y_{t-2}|),$$

where the innovations $\{\varepsilon_t\}$ are defined as in the previous experiment. The two information criteria, BIC_1 and BIC_2, in Section 3.3 are employed with $c = 10^{-5}$ and $p_{\max} = 5$.

Table 2: Percentages of underfitted, correctly selected and overfitted models by BIC_1 and BIC_2 based on 1000 replications.

		Normal			t_3			Cauchy		
	n	Under	Exact	Over	Under	Exact	Over	Under	Exact	Over
BIC_1	200	7.9	91.6	0.5	7.8	92.1	0.1	16.4	83.6	0
	500	0	99.7	0.3	0	99.8	0.2	2.5	97.5	0
	1000	0	100	0	0	100	0	0.9	99.1	0
BIC_2	200	18.8	81.2	0	16.7	83.2	0.1	19.4	80.5	0.1
	500	0	99.8	0.2	0	100	0	2.4	97.6	0
	1000	0	100	0	0	100	0	0.9	99.1	0

Recall that BIC_1 is based on the self-weighted estimators, while BIC_2 is based on the doubly weighted estimator. For $i = 1$ or 2, the cases of underfitting, correct selection and overfitting by BIC_i correspond to $\widehat{p}_{i,n}$ being 1, 2 and greater than 2, respectively. Table 2 reports the percentages of underfitted, correctly selected and overfitted models by the two information criteria. It can be seen that both information criteria select the correct model in most of the replications when the sample size is as small as $n = 200$, while BIC_1 is slightly better. We have also conducted the experiment for BIC_1 with $c = 0$, and have found that the resulting percentages remain the same as those of BIC_1 in Table 2.

In the third experiment, we study the proposed goodness-of-fit tests, $Q_1^{BP}(L)$ and $Q_2^{BP}(L)$. The data are generated from

$$y_t = c_1 y_{t-2} + \varepsilon_t(1 + 0.2|y_{t-1}| + c_2|y_{t-2}|),$$

where the innovations $\{\varepsilon_t\}$ are defined as in the first experiment. We fit a linear double AR model with $p = 1$ using the same method as in the first experiment, so that the case of $c_1 = c_2 = 0$ corresponds to the size of the tests, the case of $c_1 \neq 0$ corresponds to misspecifications in the conditional mean, and the case of $c_2 > 0$ corresponds to misspecifications in the conditional scale. Two departure levels, 0.1 and 0.3, are considered for both c_1 and c_2, and the standard Cauchy distribution function is employed as the transformation $G(\cdot)$ for the residual sequence. Table 3 reports the rejection rates of $Q_1^{BP}(6)$ and $Q_2^{BP}(6)$ based on 1000 replications, for sample size $n = 200$, 500 or 1000.

Table 3: Rejection rates of the tests $Q_1^{BP}(6)$ and $Q_2^{BP}(6)$ at the 5% significance level when the innovations follow the normal, Student's t_3 or Cauchy distribution.

			Normal			t_3			Cauchy		
	c_1	c_2	200	500	1000	200	500	1000	200	500	1000
Q_1^{BP}	0.0	0.0	0.041	0.046	0.052	0.042	0.044	0.050	0.047	0.053	0.051
	0.0	0.1	0.042	0.035	0.051	0.049	0.050	0.044	0.055	0.049	0.044
	0.0	0.3	0.054	0.048	0.064	0.054	0.050	0.066	0.084	0.081	0.070
	0.1	0.0	0.076	0.178	0.386	0.110	0.303	0.586	0.551	0.972	1.000
	0.3	0.0	0.639	0.991	1.000	0.822	0.998	1.000	0.993	1.000	1.000
Q_2^{BP}	0.0	0.0	0.044	0.056	0.049	0.048	0.050	0.051	0.056	0.052	0.047
	0.0	0.1	0.073	0.107	0.194	0.061	0.117	0.181	0.085	0.123	0.191
	0.0	0.3	0.252	0.763	0.997	0.228	0.628	0.961	0.213	0.468	0.765
	0.1	0.0	0.044	0.040	0.061	0.039	0.055	0.064	0.210	0.433	0.735
	0.3	0.0	0.059	0.075	0.146	0.110	0.191	0.339	0.796	0.998	1.000

It can be observed that all sizes are close to the nominal rate when the sample size n is as small as 200, and all powers increase as n or the departure level increases. Moreover, $Q_1^{BP}(6)$ performs well in detecting the misspecification in the conditional mean (i.e., $c_1 \neq 0$ and $c_2 = 0$), especially when the innovation distribution is heavy-tailed, but has little power for the misspecification in the conditional scale (i.e., $c_1 = 0$ and $c_2 > 0$). In contrast, $Q_2^{BP}(6)$ performs well in detecting the misspecification in the conditional scale, especially when the innovation distribution is light-tailed. This indicates that $Q_1^{BP}(L)$ and $Q_2^{BP}(L)$ should be used in conjunction to check the adequacy of the fitted model. In addition, the findings seem consistent with the result in the first experiment that, as the innovation distribution becomes more heavy-tailed, the estimation performance for the location-type parameters ϕ_0 tends to improve, whereas that for the scale-type parameters β_0 tends to worsen. Furthermore, the performance of $Q_2^{BP}(6)$ for the misspecification in the conditional mean seems mixed: it is useless when the innovation distribution is relatively light-tailed, but is surprisingly powerful for the Cauchy distribution.

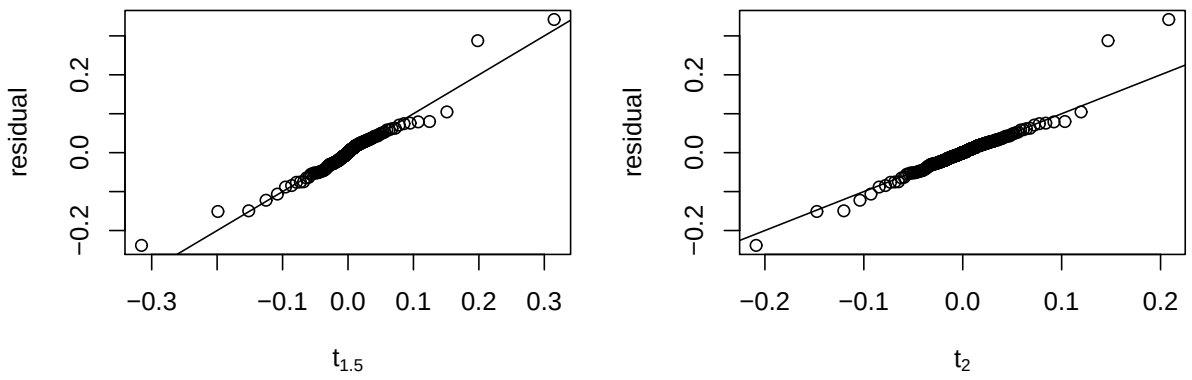

Figure 2: Q-Q plots of the residuals against the Student's $t_{1.5}$ (left panel) or t_2 (right panel) distribution.

6 An empirical example

We illustrate the proposed inference tools using the U.S. monthly interest rates (the effective federal funds rates) from January 1956 to December 2015. There are 720 observations in total, and we focus on their log returns, denoted by $\{y_t\}$.

Based on $\tau_k = k/10$ for $k = 1, \ldots, 9$ and $p_{\max} = 10$, the proposed BIC_1 and BIC_2 both select $p = 3$. By the doubly weighted estimation method in Section 3.2, the fitted model is

$$\begin{aligned} y_t =& 0.3659_{0.0385} y_{t-1} + 0.0938_{0.0295} y_{t-2} + 0.1234_{0.0320} y_{t-3} \\ &+ \varepsilon_t (1 + 27.8293_{5.9681} |y_{t-1}| + 6.9331_{3.4008} |y_{t-2}| + 14.9557_{4.2480} |y_{t-3}|), \end{aligned} \tag{6.1}$$

where the subscripts are the standard errors of the estimated coefficients, and all the estimated coefficients are significant at the 5% significance level. Figure 2 gives the Q-Q plots of the residuals from the fitted model against the Student's $t_{1.5}$ or t_2 distribution. It can be seen that the left tail of the residuals is heavier than t_2 yet lighter than $t_{1.5}$, while the right tail seems as heavy as $t_{1.5}$, which suggests that $E(\varepsilon_t^2) = \infty$ and $E(|\varepsilon_t|) < \infty$.

For the fitted model in (6.1), the p-values of the goodness-of-fit test $Q_1^{BP}(L)$ for $L = 6$, 12 and 18 are all greater than 0.5306, and those of the test $Q_2^{BP}(L)$ are all greater than 0.9597. This suggests that the fitted model is adequate. In addition, as shown in Figure 3, the residual QACFs $\widehat{\rho}_\ell$ and $\widehat{r}_\ell$ fall within the corresponding 95% confidence bounds at all the first 15 lags.

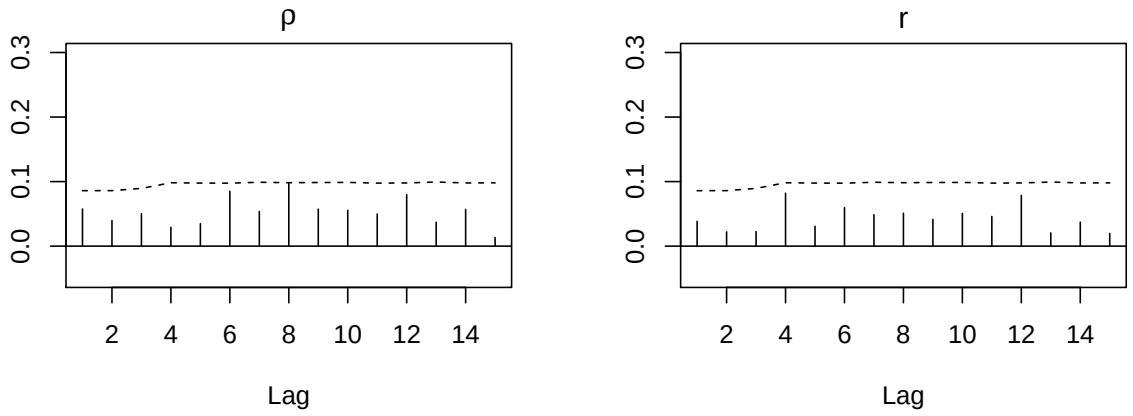

Figure 3: Residual QACF plots for $\widehat{\rho}_\ell$ (left panel) and $\widehat{r}_\ell$ (right panel), where the dashed lines are the corresponding 95% confidence bounds.

For comparison, the fitted double AR model is given by

$$\begin{aligned} y_t =& 0.4272_{0.0898} y_{t-1} + 0.0707_{0.0888} y_{t-2} + 0.1069_{0.0714} y_{t-3} \\ &+ \varepsilon_t \sqrt{0.0044_{0.0004} + 1.2561_{0.1895} y_{t-1}^2 + 0.9247_{0.1678} y_{t-2}^2 + 0.2599_{0.0896} y_{t-3}^2}, \end{aligned}$$

and the fitted AR-ARCH model is

$$\begin{aligned} y_t &= 0.3667_{0.0537} y_{t-1} + 0.1006_{0.0422} y_{t-2} + 0.1561_{0.0393} y_{t-3} + e_t, \quad e_t = \varepsilon_t \sqrt{h_t}, \\ h_t &= 0.0018_{0.0002} + 0.7699_{0.1240} e_{t-1}^2 + 0.4481_{0.0873} e_{t-2}^2 + 0.0642_{0.0452} e_{t-3}^2, \end{aligned} \tag{6.2}$$

where the innovations $\{\varepsilon_t\}$ are standardized to have mean zero and variance one. The three fitted models have similar conditional mean structures. In the ARCH component of (6.2), the coefficients of the e_{t-j}^2's add up to 1.2822, suggesting that $E(e_t^2) = \infty$. This, together with Figure 2, indicates that the double AR and AR-ARCH models and their inference tools may be misused here. Moreover, the significance of the conditional mean component implies that a linear ARCH model would not be suitable.

To examine the forecasting performance, we conduct one-step-ahead predictions using a rolling forecasting procedure. We start from the forecast origin $t = 200$ and fit the model using the data from the beginning to the forecast origin (exclusive). We compute the forecast of the τth conditional quantile of y_{t+1}, given by $\widehat{\mu}_{t+1} + \widehat{\sigma}_{t+1}\widehat{b}_\tau$, for $\tau = 1\%$, 5%, 10%, 90%, 95% and 99%, where $\widehat{\mu}_{t+1}$ and $\widehat{\sigma}_{t+1}$ are the predicted conditional mean and scale, respectively, and $\widehat{b}_\tau$ is the τth sample quantile of the residuals. Then we

Table 4: Empirical coverage rates (%) for two fitted models at different quantile levels. LDAR: linear double AR model. DAR: double AR model.

		1%	5%	10%	90%	95%	99%
$t \in [200, 399]$	LDAR	0.5	5.0	8.0	89.5	95.5	98.5
	DAR	0.5	5.5	7.5	91.0	96.5	97.5
$t \in [400, 599]$	LDAR	1.5	4.0	9.0	96.0	98.0	99.5
	DAR	1.5	3.5	8.5	96.5	99.0	99.5
$t \in [600, 719]$	LDAR	3.3	10.8	15.0	85.0	90.0	96.7
	DAR	3.3	10.8	18.3	85.8	91.7	98.3

advance the forecast origin by one and repeat the above procedure until all data are employed.

The forecasting subsample can be divided into three periods: $t \in [200, 399]$, $t \in [400, 599]$ and $t \in [600, 719]$, corresponding to the periods with moderate, low and high volatilities, respectively. Table 4 reports the empirical coverage rates (ECRs) of the one-step-ahead predictions by the fitted linear double AR model and the fitted double AR model, for the three periods and six quantile levels. Among the totally 18 cases, we find that the proposed model outperforms the double AR model in 10 cases, and is as good as the latter in 5 cases. In contrast, the double AR model is more favorable only at the three upper quantiles in the high volatility period. This is probably because the conditional scale $\widehat{\sigma}_{t+1}$ of the double AR model has a quadratic structure, which makes it more sensitive to sudden jumps in the magnitude, resulting in larger and more accurate ECRs than the proposed model in the high volatility period. Although not reported in the table, all the ECRs for the fitted AR-ARCH model deviate farther from the corresponding nominal rates than those for the other two models.

7 Conclusion

For conditional heteroscedastic time series models without a conditional mean component, the quantile regression is often made tractable by assuming a linear structure for the conditional standard deviation. However, when a conditional mean structure needs

to be incorporated, the objective function of the quantile regression is usually no longer convex and new challenges in the inference and optimization will arise.

This paper proposes the linear double AR model which is suitable for quantile inference even when there is a conditional mean component. It can be regarded as a modification of the double AR model along the lines of the linear GARCH model, but enjoys greater tractability for the quantile regression than both models. The proposed doubly weighted estimation achieves greater efficiency by optimally combining information across the quantiles. As with the estimation, the proposed information criteria and goodness-of-fit tests require no moment condition on the observed process or the innovations, whereas existing models and inference tools usually need stronger conditions. The necessity of such robustness is corroborated by the real data example in Section 6, where it is founded that the innovations may even have an infinite variance.

8 Supplementary material

The supplementary material contains additional simulation experiments and detailed proofs of all lemmas and theorems in the paper.

Acknowledgements

We are deeply grateful to the co-editor Jianqing Fan, the associate editor and two anonymous referees for their valuable comments that led to the substantial improvement of this paper. Special thanks go to one referee who kindly pointed out the best self weights in Theorem 3 and provided a detailed proof. Zhu and Zheng are co-first authors, and contributed to the paper equally. Li's research was partially supported by the Hong Kong Research Grant Council grant 17325416.

Appendix

This appendix presents an auxiliary lemma, which is crucial to the proof of Theorem 6, and gives the proof sketches of Theorems 1 and 3-5. Due to the space limit, detailed proofs of all lemmas and theorems are provided in the supplementary material.

Lemma 2. *Under Assumptions 2 and 3, we have the Bahadur representation for* $\widehat{b}_{\tau n}$*:*

$$\sqrt{n}(\widehat{b}_{\tau n} - b_\tau) = \frac{1}{f(b_\tau)}\left[\frac{1}{\sqrt{n}}\sum_{t=p+1}^{n}\psi_\tau(\varepsilon_t - b_\tau) - d_0'(\tau)\sqrt{n}(\widehat{\lambda}_n^{opt} - \lambda_0)\right] + o_p(1),$$

where $d_0(\tau) = f(b_\tau)\left(b_\tau E(\sigma_t^{-1}Y_{a,t-1}'), E(\sigma_t^{-1}Y_{t-1}')\right)'$.

Proof sketch of Theorem 1. Denote $Y_t = (y_t, y_{t-1}, \ldots, y_{t-p+1})'$. Let $\mathcal{B}^p$ be the class of Borel sets of $\mathbb{R}^p$ and ν_p be the Lebesgue measure on $(\mathbb{R}^p, \mathcal{B}^p)$. By Assumption 1, we can show that $\{Y_t\}$ is a homogeneous Markov chain on the state space $(\mathbb{R}^p, \mathcal{B}^p, \nu_p)$, has a p-step transition kernel that is positive everywhere, and hence is ν_p-irreducible.

To prove the sufficiency, suppose that $\gamma < 0$, i.e., there is an integer s such that $E(\ln\|A_1 \cdots A_s\|) < 0$. Let $\widetilde{A}_t = \prod_{i=0}^{s-1} A_{t-i}$. By the continuity of the density $f(\cdot)$ and the dominated convergence theorem, we can show that $\lim_{u\to 0}\dot{q}(u) = E(\ln\|\widetilde{A}_t\|) < 0$, where $\dot{q}(u)$ is the derivative of $q(u) = E(\|\widetilde{A}_t\|^u)$, and thus, there is a constant $\kappa \in (0,1)$ such that $E(\|\widetilde{A}_t\|^\kappa) < q(0) = 1$. Using this result and the test function $g(x) = 1 + \|x\|^\kappa$, we can verify Tweedie's drift criterion (Tweedie, 1983, Theorem 4) for the s-step Markov chain $\{Y_{ts}^*\}$ and hence that for $\{Y_{ts}\}$, since $\{Y_t\}$ and $\{Y_t^*\}$ have the same transition probability. We can further show that $\{Y_{ts}\}$ is a ν_p-irreducible Feller chain, and then by Theorem 4(ii) in Tweedie (1983) and Theorems 1 and 2 in Feigin and Tweedie (1985), $\{Y_{ts}\}$ is geometrically ergodic with a unique stationary distribution and $E(|y_t|^\kappa) < \infty$. By Lemma 3.1 of Tjøstheim (1990), we conclude that $\{Y_t\}$ is geometrically ergodic and is the unique strictly stationary solution to model (2.1).

To prove the necessity, suppose that there is a strictly stationary solution $\{y_t\}$ to model (2.1). Then we can generate iteratively a strictly stationary and nonanticipative solution $\{Y_t^* : t \in N\}$ for model (2.3) by letting Y_0^* follow the same distribution as Y_t. As a result, $\{Y_{tp}^* : t \in N\}$ is a nonanticipative and strictly stationary solution to $Y_{tp}^* = \widetilde{A}_{tp}Y_{(t-1)p}^* + B_{tp}$, where $\widetilde{A}_t = \prod_{i=0}^{p-1} A_{t-i}$ and $B_{tp} = e_{tp} + \sum_{j=1}^{p-1}\prod_{r=0}^{j-1} A_{tp-r}e_{tp-j}$ with $e_t = (\varepsilon_t, 0, \ldots, 0)'$. Moreover, it can be shown that $E(\ln^+\|\widetilde{A}_{tp}\|) < \infty$, $E(\ln^+\|B_{tp}\|) < \infty$, and $\{Y_{tp}^* : t \in N\}$ is irreducible. Finally, by Theorem 2.5 of Bougerol and Picard (1992), the top Lyapounov exponent $\widetilde{\gamma} = \inf\{t^{-1}E(\ln\|\widetilde{A}_p\widetilde{A}_{2p}\cdots\widetilde{A}_{tp}\|), t \geq 1\}$ is strictly negative, and it follows that $\gamma \leq \widetilde{\gamma}/p < 0$. □

Proof sketch of Theorem 3. The asymptotic normality of $\sqrt{n}(\widetilde{\lambda}_{\tau n} - \lambda_0)$ in Theorem 3 follows directly from Lemma 1 and the Delta method (van der Vaart, 1998, Chapter

3). To find the minimum of $\Omega_1(w)$, as in Xu (2017), we consider the regression model, $z_t = \sigma_t^{-1} x_t'\gamma + e_t$, where $\{e_t\}$ are *i.i.d.* standard normal, independent of $\{x_t\}$, and γ is the unknown parameter to be estimated. The weighted least-squares estimator $\widehat{\gamma}(\lambda) = \text{argmin}_r \sum_{t=1}^n \lambda_t \left(z_t - \sigma_t^{-1} x_t' r\right)^2$ with the weights $\lambda_t = \sigma_t w_t$ is asymptotically normal with mean zero and variance $\Omega_1(w)$. On the other hand, by the normality of e_t, the estimator is most efficient when $\lambda_t \equiv 1$. Thus, $\Omega_1(w)$ is minimized at $w_t = \sigma_t^{-1}$, and so is $\Omega_2(w)$. □

Proof sketch of Theorem 4. By Lemma 1 and the Delta method, we have the Bahadur representation $\sqrt{n}(\widetilde{\lambda}_{\tau n} - \lambda_0) = \Sigma_2(\tau)\Omega_0^{-1} n^{-1/2} \sum_{t=p+1}^n \psi_\tau(\varepsilon_t - b_\tau)\sigma_t^{-1} x_t + o_p(1)$, where $\Omega_0^{-1} = \Omega_1$ since $w_t = \sigma_t^{-1}$. It then follows from the central limit theorem that $\sqrt{n}(\widehat{\lambda}_n - \lambda_0) \to N(0, \mathcal{V}(\Pi))$ in distribution as $n \to \infty$. Consider a minimum distance estimator

$$\widehat{\lambda}_n^* = \underset{\lambda}{\text{argmin}} \{\widetilde{\lambda}_n - (1_K \otimes I_{2p})\lambda\}' \Xi \{\widetilde{\lambda}_n - (1_K \otimes I_{2p})\lambda\},$$

where Ξ is a $2pK \times 2pK$ matrix and $\widetilde{\lambda}_n = (\widetilde{\lambda}_{\tau_1 n}', ..., \widetilde{\lambda}_{\tau_K n}')'$. It can be verified that $\widehat{\lambda}_n^* = \Pi\widetilde{\lambda}_n = \sum_{k=1}^K \pi_k \widetilde{\lambda}_{\tau_k n}$, where $\Pi = (\pi_1, ..., \pi_K) = \{(1_K \otimes I_{2p})'\Xi(1_K \otimes I_{2p})\}^{-1}(1_K \otimes I_{2p})'\Xi$. As argued in Chen et al. (2016), the asymptotic variance of $\widehat{\lambda}_n^*$ is minimized when Ξ is proportional to the inverse of the asymptotic variance of $\sqrt{n}[\widetilde{\lambda}_n - (1_K \otimes I_{2p})\lambda_0]$. Thus, we can obtain Π^{opt} that corresponds to such a matrix Ξ and the results of the theorem. □

Proof sketch of Theorem 5. By $\widetilde{\beta}^{int} - \beta_0 = O_p(n^{-1/2})$, $\widetilde{\sigma}_t = 1 + Y_{a,t-1}'\widetilde{\beta}^{int}$ and $w_t = (\widetilde{\sigma}_t + c\sum_{j=1}^{p_{\max}} |y_{t-j}|)^{-1}$, it suffices to prove the theorem for the weights $w_t = (\sigma_t + c\sum_{j=1}^{p_{\max}} |y_{t-j}|)^{-1} = [1 + \sum_{j=1}^{p_{\max}} (c + \beta_{0j})|y_{t-j}|]^{-1}$, for which Assumption 2 holds since $c + \beta_{0j} > 0$ for $1 \le j \le p_{\max}$. By a standard argument, we can accomplish the proof. □

References

An, H. Z. and Z. G. Chen (1982). On convergence of LAD estimates in autoregresion with infinite variance. *Journal of Multivariate Analysis 12*, 335–345.

Bollerslev, T. (1986). Generalized autoregression conditional heteroscedasticity. *Journal of Econometrics 31*, 307–327.

Bougerol, P. and N. Picard (1992). Strict stationarity of generalized autoregressive processes. *The Annals of Probability 20*, 1714–1730.

Chen, X., D. Jacho-Chávez, and O. Linton (2016). Averaging of an increasing number of moment condition estimators. *Econometric Theory 32*, 30–70.

Davis, R. A., K. Knight, and J. Liu (1992). M-estimation for autoregressions with infinite variances. *Stochastic Processes and their Applications 40*, 145–180.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica 50*, 987–1007.

Engle, R. F. and S. Manganelli (2004). CAViaR: conditional autoregressive value at risk by regression quantiles. *Journal of Business and Economic Statistics 22*, 367–381.

Feigin, P. D. and R. L. Tweedie (1985). Random coefficient autoregressive processes: a markov chain analysis of stationarity and finiteness of moments. *Journal of Time Series Analysis 6*, 1–14.

Francq, C. and J. M. Zakoian (2004). Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes. *Bernoulli 10*, 605–637.

Gross, S. and W. L. Steiger (1979). Least absolute deviation estimates in autoregresion with infinite variance. *Journal of Applied Probability 16*, 104–116.

Jiang, J., X. Jiang, and X. Song (2014). Weighted composite quantile regression estimation of DTARCH models. *Econometrics Journal 17*, 1–23.

Jiang, X., J. Jiang, and X. Song (2012). Oracle model selection for nonlinear models based on weighted composite quantile regression. *Statistica Sinica 22*, 1479–1506.

Kingman, J. F. C. (1973). Subadditive ergodic theory. *The Annals of Probability 1*, 883–899.

Koenker, R. (1984). A note on L-estimates for linear models. *Statistics and Probability Letters 2*, 323–325.

Koenker, R. (2005). *Quantile regression*. Cambridge: Cambridge University Press.

Koenker, R. and G. Bassett (1978). Regression quantiles. *Econometrica 46*, 33–49.

Koenker, R. and J. A. F. Machado (1999). Goodness of fit and related inference processes for quantile regression. *Journal of the American Statistical Association 94*, 1296–1310.

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

Li, G. and W. K. Li (2008). Least absolute deviation estimation for fractionally integrated autoregressive moving average time series models with conditional heteroscedasticity. *Biometrika 95*, 399–414.

Li, G. and W. K. Li (2009). Least absolute deviation estimation for unit root processes with GARCH errors. *Econometric Theory 25*, 1208–1227.

Li, G., Y. Li, and C.-L. Tsai (2015). Quantile correlations and quantile autoregressive modeling. *Journal of the American Statistical Association 110*, 246–261.

Ling, S. (2004). Estimation and testing stationarity for double-autoregressive models. *Journal of the Royal Statistical Society, Series B 66*, 63–78.

Ling, S. (2005). Self-weighted least absolute deviation estimation for infinite variance autoregressive models. *Journal of the Royal Statistical Society: Series B 67*, 381–393.

Ling, S. (2007). A double AR(p) model: structure and estimation. *Statistica Sinica 17*, 161–175.

Ling, S. and D. Li (2008). Asymptotic inference for a nonstationary double AR(1) model. *Biometrika 95*, 257–263.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. London: Chapman and Hall.

Tjøstheim, D. (1990). Nonlinear time series and Markov chains. *Advances in Applied Probability 22*, 587–611.

Tweedie, R. L. (1983). Criteria for rates of convergence of Markov chains, with application to queueing and storage theory. In J. F. C. Kingman and G. E. H. Reuter (Eds.), *Probability, Statistics and Analysis*, pp. 260–276. Cambridge: Cambridge University Press.

van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge: Cambridge University Press.

Xiao, Z. and R. Koenker (2009). Conditional quantile estimation for generalized autoregressive conditional heteroscedasticity models. *Journal of the American Statistical Association 104*, 1696–1712.

Xu, Z. (2017). *Efficient parameter estimation methods using quantile regression in heteroscedastic models*. Ph. D. thesis, The Pennsylvania State University.

Zhao, Z. and Z. Xiao (2014). Efficient regressions via optimally combining quantile information. *Econometric Theory 30*, 1272–1314.

Zhu, K. and S. Ling (2011). Global self-weighted and local quasi-maximum exponential likelihood estimators for ARMA-GARCH/IGARCH models. *The Annals of Statistics 39*, 2131–2163.

Zhu, K. and S. Ling (2013). Quasi-maximum exponential likelihood estimators for a double AR(p) model. *Statistica Sinica 23*, 251–270.

Zou, H. and M. Yuan (2008). Composite quantile regression and the oracle model selection theory. *The Annals of Statistics 36*, 1108–1126.