# CONDITIONAL QUANTILE ESTIMATION FOR HYSTERETIC AUTOREGRESSIVE MODELS

Degao Li, Ruochen Zeng, Liwen Zhang, Wai Keung Li and Guodong Li

*Jiaxing University, University of Hong Kong*

*and Shanghai University of Finance and Economics*

**Supplementary Material**

This supplementary material gives the technical proofs of Theorems 1-4.

## S1 Proof of Theorem 1

For simplicity, we drop $\tau$ in all notations for model parameters in the subsequent section. For example, we denote $\theta_\tau = \theta, R_{L,\tau} = R_L, R_{U,\tau} = R_U$. As in the standard arguments in Huber (1967), it is sufficient to verify the following three claims:

(a) $\sup_{\lambda\in\Lambda} n^{-1}\left|\widetilde{L}_n(\lambda) - L_n(\lambda)\right| \to 0$ in the almost surely sense, where the parameter space $\Lambda$ is previously defined.

(b) $E\{\rho_\tau[y_t - M_t(\lambda)]\} \geq E\{\rho_\tau[y_t - M_t(\lambda^0)]\}$ for any $\lambda \in \Lambda$. Additionally, the equality holds if and only if $\lambda = \lambda^0$.

(c)

$$E\left\{\sup_{\widetilde{\lambda}\in\delta_\Lambda(\eta)}\left|\rho_\tau[y_t - M_t(\widetilde{\lambda})] - \rho_\tau[y_t - M_t(\lambda)]\right|\right\} \to 0, \quad \text{as} \;\; \eta\to 0,$$

where $\delta_\Lambda(\eta) = \{\widetilde{\lambda}\in\Lambda : \|\widetilde{\lambda}-\lambda\| < \eta\}$, $0<\eta<1$ and $\lambda\in\Lambda$. Therefore, this shows that $E\{\rho_\tau[M_t(\lambda)]\}$ is a continuous function of $\lambda$.

We first prove Claim (a). Let $\jmath_n = \min\{t : y_{t-d}\in(r_L, r_U]\}$. By the settings for the initial values, it holds that $\widetilde{R}_t = 1$ for $1\le t\le \jmath_n$ and $R_{\jmath_n} = R_{\jmath_n-1} = \ldots = R_1$. We have

$$\begin{aligned}
&\frac{1}{n}\left|\widetilde{L}_n(\lambda) - L_n(\lambda)\right| \\
&= (1-R_1)\prod_{t=1}^{\jmath_n} I\{r_L < y_{t-d}\le r_U\}\cdot\frac{1}{n}\left|\sum_{t=1}^{\jmath_n}[\rho_\tau(y_t - x_t^T\theta_1) - \rho_\tau(y_t - x_t^T\theta_2)]\right| \\
&\le \prod_{t=1}^{\jmath_n} I\{a\le y_{t-d}\le b\}\cdot\frac{1}{n}\sum_{t=1}^{\jmath_n}\sup_{\lambda\in\Lambda}\left[|x_t^T(\theta_2-\theta_1)\tau| + |y_t - x_t^T\theta_2| + |y_t - x_t^T\theta_1|\right],
\end{aligned}$$

which implies that Claim (a) holds if $\jmath_n$ is finite. On the other hand, applying the ergodic theorem, we obtain

$$\begin{aligned}
&\frac{1}{\jmath_n}\sum_{t=1}^{\jmath_n}\sup_{\lambda\in\Lambda}\left[|x_t^T(\theta_2-\theta_1)\tau| + |y_t - x_t^T\theta_2| + |y_t - x_t^T\theta_1|\right] \\
&\to E\left\{\sup_{\lambda\in\Lambda}\left[|x_t^T(\theta_2-\theta_1)\tau| + |y_t - x_t^T\theta_2| + |y_t - x_t^T\theta_1|\right]\right\} < \infty, \qquad \text{(A.1)}
\end{aligned}$$

when $\jmath_n \to \infty$ as $n \to \infty$, and $E(|y_t|^{2+\delta}) < \infty$. As in Li et al. (2015a), it is easy to show that

$$\prod_{t=1}^{\jmath_n} I\{a \leq y_{t-d} \leq b\} \to 0, \quad \text{as} \quad \jmath_n \to \infty. \tag{A.2}$$

Thus, (A.1) and (A.2) imply the validity of Claim (a).

Next, we prove Claim (b). Denote $\psi_\tau(w) = \tau - I\{w < 0\}$, and it holds that, for $u \neq 0$,

$$\begin{aligned}\rho_\tau(u-v) - \rho_\tau(u) &= -v\psi_\tau(u) + \int_0^v [I(u \leq s) - I(u < 0)]ds \\ &= -v\psi_\tau(u) + (u-v)[I(0 > u > v) - I(0 < u < v)];\end{aligned}$$

see Knight (1998). Moreover, $E\{\psi_\tau(y_t - x_t^T\theta_1^0)R_t^0|\mathcal{F}_{t-1}\} = 0$ and $E\{\psi_\tau(y_t - x_t^T\theta_2^0)(1-R_t^0)|\mathcal{F}_{t-1}\} = 0$. As a result,

$$\begin{aligned}E\{\rho_\tau[y_t - M_t(\lambda)]\} &= E\{\rho_\tau[y_t - M_t(\lambda^0)]\} + E\{[\rho_\tau(y_t - x_t^T\theta_1) - \rho_\tau(y_t - x_t^T\theta_1^0)]R_tR_t^0\} \\ &+ E\{[\rho_\tau(y_t - x_t^T\theta_1) - \rho_\tau(y_t - x_t^T\theta_2^0)]R_t(1-R_t^0)\} \\ &+ E\{[\rho_\tau(y_t - x_t^T\theta_2) - \rho_\tau(y_t - x_t^T\theta_1^0)](1-R_t)R_t^0\} \\ &+ E\{[\rho_\tau(y_t - x_t^T\theta_2) - \rho_\tau(y_t - x_t^T\theta_2^0)](1-R_t)(1-R_t^0)\} \\ &= I_1 + I_2 + I_3 + I_4 + I_5.\end{aligned}$$

where

$$I_1 = E\{\rho_\tau[y_t - M_t(\lambda^0)]\},$$

$$I_2 = E\{(y_t - x_t^T\theta_1)(I\{x_t^T\theta_1^0 > y_t > x_t^T\theta_1\} - I\{x_t^T\theta_1^0 < y_t < x_t^T\theta_1\})R_tR_t^0\},$$

$$I_3 = E\{(y_t - x_t^T\theta_1)(I\{x_t^T\theta_2^0 > y_t > x_t^T\theta_1\} - I\{x_t^T\theta_2^0 < y_t < x_t^T\theta_1\})R_t(1 - R_t^0)\},$$

$$I_4 = E\{(y_t - x_t^T\theta_2)(I\{x_t^T\theta_1^0 > y_t > x_t^T\theta_2\} - I\{x_t^T\theta_1^0 < y_t < x_t^T\theta_2\})R_t^0(1 - R_t)\},$$

$$I_5 = E\{(y_t - x_t^T\theta_2)(I\{x_t^T\theta_2^0 > y_t > x_t^T\theta_2\} - I\{x_t^T\theta_2^0 < y_t < x_t^T\theta_2\})(1 - R_t^0)(1 - R_t)\}.$$

It can be shown that $I_1, I_2, I_3, I_4$ and $I_5$ are all nonnegative, and thus $E\{\rho_\tau[y_t - M_t(\lambda)]\} \geq E\{\rho_\tau[y_t - M_t(\lambda^0)]\}$. Furthermore, the above equality holds only when these nonnegative terms are all equal to zero.

From the equality

$$E\{(y_t - x_t^T\theta_1)(I\{x_t^T\theta_1^0 > y_t > x_t^T\theta_1\} - I\{x_t^T\theta_1^0 < y_t < x_t^T\theta_1\})R_tR_t^0\} = 0,$$

we have

$$E\{(y_t - x_t^T\theta_1)I\{x_t^T\theta_1^0 > y_t > x_t^T\theta_1\}R_tR_t^0\} = 0, \tag{A.3}$$

and

$$E\{(y_t - x_t^T\theta_1)I\{x_t^T\theta_1^0 < y_t < x_t^T\theta_1\}R_tR_t^0\} = 0. \tag{A.4}$$

Thus (A.3) implies that

$$P\{I\{x_t^T\theta_1^0 > y_t > x_t^T\theta_1\} = 0\} \geq P\{R_tR_t^0 = 1\} \geq P\{y_{t-d} \leq r_L, y_{t-d^0} \leq r_L^0\} > 0.$$

We then have

$$P\{0 > y_t - x_t^T\theta_1^0 > x_t^T(\theta_1 - \theta_1^0)\} = P\{x_t^T\theta_1^0 > y_t > x_t^T\theta_1\} = 0.$$

By Assumption 4, we can obtain that $x_t^T(\theta_1 - \theta_1^0) \geq 0$ almost surely. On the other hand, by (A.4), we can obtain that $x_t^T(\theta_1 - \theta_1^0) \leq 0$ almost surely. Thus we have $\theta_1 = \theta_1^0$.

Similarly, we can obtain that $\theta_2 = \theta_2^0$ from the equality

$$E\{(y_t - x_t^T\theta_2)(I\{x_t^T\theta_2^0 > y_t > x_t^T\theta_2\} - I\{x_t^T\theta_2^0 < y_t < x_t^T\theta_2\})(1 - R_t^0)(1 - R_t)\}$$
$$= 0.$$

Based on the following two inequalities

$$E\{(y_t - x_t^T\theta_1)(I\{x_t^T\theta_2^0 > y_t > x_t^T\theta_1\} - I\{x_t^T\theta_2^0 < y_t < x_t^T\theta_1\})R_t(1 - R_t^0)\} = 0,$$

and

$$E\{(y_t - x_t^T\theta_2)(I\{x_t^T\theta_1^0 > y_t > x_t^T\theta_2\} - I\{x_t^T\theta_1^0 < y_t < x_t^T\theta_2\})R_t^0(1 - R_t)\} = 0.$$

As $\theta_1^0 \neq \theta_2^0$, we have $E\{(1 - R_t)R_t^0\} = 0$ and $E\{(1 - R_t^0)R_t\} = 0$, respectively. As in Li et al. (2015a), we have $r_L = r_L^0, r_U = r_U^0$, and $d = d^0$. Thus, Claim (b) holds.

To complete the proof, it is sufficient to verify Claim (c). We denote $\widetilde{\lambda} = (\widetilde{\theta}_1^T, \widetilde{\theta}_2^T, \widetilde{r}_L, \widetilde{r}_U, d)^T \in \delta_\Lambda(\eta)$ with $0 < \eta < 1$, then

$$\begin{aligned}
&\rho_\tau[y_t - M_t(\widetilde{\lambda})] - \rho_\tau[y_t - M_t(\lambda)] \\
&= [\rho_\tau(y_t - x_t^T\widetilde{\theta}_1) - \rho_\tau(y_t - x_t^T\theta_1)]R_t(\widetilde{r}_L, \widetilde{r}_U, d)R_t(r_L, r_U, d) \\
&\quad + [\rho_\tau(y_t - x_t^T\widetilde{\theta}_2) - \rho_\tau(y_t - x_t^T\theta_1)][1 - R_t(\widetilde{r}_L, \widetilde{r}_U, d)]R_t(r_L, r_U, d) \\
&\quad + [\rho_\tau(y_t - x_t^T\widetilde{\theta}_1) - \rho_\tau(y_t - x_t^T\theta_2)]R_t(\widetilde{r}_L, \widetilde{r}_U, d)[1 - R_t(r_L, r_U, d)] \\
&\quad + [\rho_\tau(y_t - x_t^T\widetilde{\theta}_2) - \rho_\tau(y_t - x_t^T\theta_2)][1 - R_t(\widetilde{r}_L, \widetilde{r}_U, d)][1 - R_t(r_L, r_U, d)].
\end{aligned}$$

Note that

$$\begin{aligned}
&[\rho_\tau(y_t - x_t^T\widetilde{\theta}_1) - \rho_\tau(y_t - x_t^T\theta_1)]R_t(\widetilde{r}_L, \widetilde{r}_U, d)R_t(r_L, r_U, d) \\
&= \Big\{ - x_t^T(\widetilde{\theta}_1 - \theta_1)\psi_\tau(y_t - x_t^T\theta_1) + \int_0^{x_t^T(\widetilde{\theta}_1 - \theta_1)} [I\{y_t - x_t^T\theta_1 \le s\} \\
&\quad - I\{y_t - x_t^T\theta_1 \le 0\}]ds \Big\} \times R_t(\widetilde{r}_L, \widetilde{r}_U, d)R_t(r_L, r_U, d) \\
&\le 3\eta\|x_t\|.
\end{aligned}$$

In the same way,

$$[\rho_\tau(y_t - x_t^T\widetilde{\theta}_2) - \rho_\tau(y_t - x_t^T\theta_2)][1 - R_t(\widetilde{r}_L, \widetilde{r}_U, d)][1 - R_t(r_L, r_U, d)] \le 3\eta\|x_t\|.$$

On the other hand, it is easy to obtain that

$$|[1 - R_t(\widetilde{r}_L, \widetilde{r}_U, d)]R_t(r_L, r_U, d)| \le |R_t(\widetilde{r}_L, \widetilde{r}_U, d) - R_t(r_L, r_U, d)|,$$

and

$$|[1 - R_t(r_L, r_U, d)]R_t(\widetilde{r}_L, \widetilde{r}_U, d)| \le |R_t(\widetilde{r}_L, \widetilde{r}_U, d) - R_t(r_L, r_U, d)|.$$

Moreover,

$$[\rho_\tau(y_t - x_t^T\widetilde{\theta}_1) - \rho_\tau(y_t - x_t^T\theta_2)]R_t(\widetilde{r}_L, \widetilde{r}_U, d)[1 - R_t(r_L, r_U, d)]$$

$$= [\rho_\tau(y_t - x_t^T\widetilde{\theta}_1) - \rho_\tau(y_t - x_t^T\theta_1)]R_t(\widetilde{r}_L, \widetilde{r}_U, d)[1 - R_t(r_L, r_U, d)]$$

$$+ [\rho_\tau(y_t - x_t^T\theta_1) - \rho_\tau(y_t - x_t^T\theta_2)]R_t(\widetilde{r}_L, \widetilde{r}_U, d)[1 - R_t(r_L, r_U, d)]$$

$$\leq 3\eta\|x_t\| + [\rho_\tau(y_t - x_t^T\theta_1) + \rho_\tau(y_t - x_t^T\theta_2)]|R_t(\widetilde{r}_L, \widetilde{r}_U, d) - R_t(r_L, r_U, d)|.$$

In a similar way, we obtain that

$$[\rho_\tau(y_t - x_t^T\widetilde{\theta}_1) - \rho_\tau(y_t - x_t^T\theta_2)]R_t(\widetilde{r}_L, \widetilde{r}_U, d)[1 - R_t(r_L, r_U, d)]$$

$$\leq 3\eta\|x_t\| + [\rho_\tau(y_t - x_t^T\theta_1) + \rho_\tau(y_t - x_t^T\theta_2)]|R_t(\widetilde{r}_L, \widetilde{r}_U, d) - R_t(r_L, r_U, d)|.$$

Denote $C_1 = 12\|x_t\|$ and $C_2 = 2E\{\rho_\tau(y_t - x_t^T\theta_1) + \rho_\tau(y_t - x_t^T\theta_2)\}^{1+\varsigma/2}$, respectively. It is easy to show both $C_1$ and $C_2$ are finite when $E\{|y_t|^{2+\varsigma}\} < \infty$. Applying Hölder's inequality, we have

$$E\left\{\sup_{\widetilde{\lambda}\in\delta_\Lambda(\eta)}\left|\rho_\tau[y_t - M_t(\widetilde{\lambda})] - \rho_\tau[y_t - M_t(\lambda)]\right|\right\}$$
$$\leq C_1\eta + C_2\left(E\left\{\sup_{\widetilde{\lambda}\in\delta_\Lambda(\eta)}|R_t(r_L, r_U, d) - R_t(\widetilde{r}_L, \widetilde{r}_U, d)|\right\}\right)^{\varsigma/(2+\varsigma)}.$$

Similar to Li et al. (2015a), we have

$$E\left\{\sup_{\widetilde{\lambda}\in\delta_\Lambda(\eta)}|R_t(r_L, r_U, d) - R_t(\widetilde{r}_L, \widetilde{r}_U, d)|\right\} \to 0, \quad \text{as} \quad \eta \to 0.$$

We thus finish the proof of Claim (c).

Making use of the standard argument for strong consistency in Huber (1967), based on the above three claims, it can be shown that $\widehat{\lambda}_n \to \lambda_0$ almost surely. We then complete the proof of Theorem 1.

## S2 Proof of Theorem 2

As Theorem 1 indicates that $\widehat{\theta}_n$ is strongly consistent, without loss of generality, we restrict the parameter space to a neighborhood of $\theta^0$, say,

$$\xi(\Delta) = \{\theta \in \Theta, a < r_L < r_U < b : |\theta - \theta^0| < \Delta, |r_L - r_L^0| < \Delta, |r_U - r_U^0| < \Delta\},$$

for $0 < \Delta < \min\{1, (r_U^0 - r_L^0)/2\}$. First, we assume $p = d = 1$. And for simplicity assume $z_L < 0, z_U > 0$.

Similar to Chan (1991), we need to verify that $\forall \varepsilon > 0, \exists K > 0$,

$$P\{\widetilde{L}_n(\theta, r_L^0 + z_L, r_U^0 + z_U) - \widetilde{L}_n(\theta, r_L^0, r_U^0) > 0\} > 1 - \varepsilon, \tag{A.5}$$

where $\theta \in \xi(\Delta), |z_L| > K/n$, and $|z_U| > K/n$.

Denote

$$\mathcal{A}_0 = \{r_U^0 < y_{t-1} \leq r_U^0 + z_U, R_{t-1} = 1\}, \quad \mathcal{B}_0 = \{r_L^0 + z_L < y_{t-1} \leq r_L^0, R_{t-1} = 0\},$$

$$\mathcal{A}_{jt} = \{y_{t-1}, \ldots, y_{t-j} \in (r_L^0, r_U^0), r_U^0 < y_{t-j-1} \leq r_U^0 + z_U, R_{t-j-1} = 1\} \text{ for } j \geq 1,$$

and

$$\mathcal{B}_{jt} = \{y_{t-1}, \ldots, y_{t-j} \in (r_L^0, r_U^0), r_L^0 + z_L < y_{t-j-1} \leq r_L^0, R_{t-j-1} = 0\} \text{ for } j \geq 1,$$

where $R_t = R_t(r_L^0 + z_L, r_U^0 + z_U)$. As Li et al. (2015a), we have

$$R_t(r_L^0 + z_L, r_U^0 + z_U) - R_t(r_L^0, r_U^0) = I\{\mathcal{A}_t(z_L, z_U)\} - I\{\mathcal{B}_t(z_L, z_U)\},$$

where $\mathcal{A}_t(z_L, z_U) = \bigcup_{j=1}^{\infty} \mathcal{A}_{jt}$, and $\mathcal{B}_t(z_L, z_U) = \bigcup_{j=1}^{\infty} \mathcal{B}_{jt}$. Moreover, it can be shown that $\mathcal{B}_t(z_L, z_U) \subset \{\mathcal{B}_t(r_L^0, r_U^0) = 1\}$ and $\mathcal{A}_t(z_L, z_U) \subset \{\mathcal{A}_t(r_L^0, r_U^0) =$

0}. Then,

$$
\begin{aligned}
&\widetilde{L}_n(\theta, r_L^0 + z_L, r_U^0 + z_U) - \widetilde{L}_n(\theta, r_L^0, r_U^0) \\
&= \sum_{t=1}^{n}[\rho(y_t - x^T\theta_1) - \rho(y_t - x^T\theta_2)][R_t(r_L^0 + z_L, r_U^0 + z_U) - R_t(r_L^0, r_U^0)] \\
&= \sum_{t=1}^{n}[\rho(y_t - x^T\theta_1) - \rho(y_t - x^T\theta_2)][I\{\mathcal{A}_t(z_L, z_U)\} - I\{\mathcal{B}_t(z_L, z_U)\}] \\
&= \widetilde{L}_n^1(z_L, z_U) + \widetilde{L}_n^2(z_L, z_U),
\end{aligned}
$$

where

$$
\widetilde{L}_n^1(z_L, z_U) = \sum_{t=1}^{n}[\rho(y_t - x^T\theta_1) - \rho(y_t - x^T\theta_2)]I\{\mathcal{A}_t(z_L, z_U)\},
$$

and

$$
\widetilde{L}_n^2(z_L, z_U) = \sum_{t=1}^{n}[\rho(y_t - x^T\theta_2) - \rho(y_t - x^T\theta_1)]I\{\mathcal{B}_t(z_L, z_U)\}.
$$

Next, we show that $\forall \varepsilon > 0$, $\exists K > 0$, such that

$$
P\{\widetilde{L}_n^1(z_L, z_U) > 0\} > 1 - \varepsilon,
$$

when $\theta \in \xi(\Delta)$, $-\Delta < z_L < 0$, and $K/n < z_U < \Delta$. Then

$$
\begin{aligned}
\widetilde{L}_n^1(z_L, z_U) &= \sum_{t=1}^{n} \{\rho(y_t - x^T\theta_1) - \rho(y_t - x^T\theta_2)\}\, I\{\mathcal{A}_t(z_L, z_U)\} \\
&= \sum_{t=1}^{n} \{-x_t^T(\theta_1 - \theta_2)\psi_\tau(y_t - x_t^T\theta_2)\}\, I\{\mathcal{A}_t(z_L, z_U)\} \\
&\quad + \sum_{t=1}^{n}\{(y_t - x_t^T\theta_1)(I\{x_t^T\theta_2 > y_t > x_t^T\theta_1\} \\
&\quad - I\{x_t^T\theta_2 < y_t < x_t^T\theta_1\})\}I\{\mathcal{A}_t(z_L, z_U)\}.
\end{aligned}
$$

By Assumption 4, when $\Delta$ is sufficiently small, for some $\upsilon_1 > 0$,

$$\sum_{t=1}^{n}\{(y_t - x_t^T\theta_1)(I\{x_t^T\theta_2 > y_t > x_t^T\theta_1\} - I\{x_t^T\theta_2 < y_t < x_t^T\theta_1\})\}I\{\mathcal{A}_t(z_L, z_U)\}$$
$$\geq \upsilon_1 \sum_{t=1}^{n} I\{\mathcal{A}_t(z_L, z_U)\}.$$

As $P\{y_t - x_t^T\theta_2^0 < 0\} = \tau$, we have for some constant $\upsilon_2 > 0$,

$$\begin{aligned} I\{y_t - x_t^T\theta_2 < 0\} &= I\{y_t - x_t^T\theta_2^0 + x_t^T(\theta_2^0 - \theta_2) < 0\} \\ &= I\{y_t - x_t^T\theta_2^0 < 0\} + I\{0 < y_t - x_t^T\theta_2^0 < -x_t^T(\theta_2^0 - \theta_2)\} \\ &= I\{y_t - x_t^T\theta_2^0 < 0\} + I\{x_t^T\theta_2^0 < y_t < x_t^T\theta_2\}. \end{aligned}$$

Furthermore,

$$\begin{aligned} &\sum_{t=1}^{n}\{-x_t^T(\theta_1 - \theta_2)\psi_\tau(y_t - x_t^T\theta_2)\}\, I\{\mathcal{A}_t(z_L, z_U)\} \\ &\leq \upsilon_2\left\{\left|\sum_{t=1}^{n}\psi_\tau(y_t - x_t^T\theta_2^0)I\{\mathcal{A}_t(z_L, z_U)\}\right| + \left|\sum_{t=1}^{n} y_{t-1}\psi_\tau(y_t - x_t^T\theta_2^0)I\{\mathcal{A}_t(z_L, z_U)\}\right|\right\} \\ &\quad + \upsilon_2\left\{\left|\sum_{t=1}^{n} I\{x_t^T\theta_2^0 < y_t < x_t^T\theta_2\}I\{\mathcal{A}_t(z_L, z_U)\}\right|\right. \\ &\qquad \left. + \left|\sum_{t=1}^{n} y_{t-1}I\{x_t^T\theta_2^0 < y_t < x_t^T\theta_2\}I\{\mathcal{A}_t(z_L, z_U)\}\right|\right\}. \end{aligned}$$

Denote

$$T_{z_L}(z_U) = E\{I\{\mathcal{A}_t(z_L, z_U)\}\}, \quad T_{n,z_L}(z_U) = \frac{1}{n}\sum_{t=1}^{n} I\{\mathcal{A}_t(z_L, z_U)\},$$

$$V_{z_L}(z_U) = E\{y_{t-1}\psi_\tau(y_t - x_t^T\theta_2^0)I\{\mathcal{A}_t(z_L, z_U)\}\},$$

$$V_{n,z_L}(z_U) = \frac{1}{n}\sum_{t=1}^{n} y_{t-1}\psi_\tau(y_t - x_t^T\theta_2^0)I\{\mathcal{A}_t(z_L, z_U)\},$$

$$\widetilde{V}_{n,z_L}(z_1, z_2) = \frac{1}{n}\sum_{t=1}^{n} |y_{t-1}\psi_\tau(y_t - x_t^T\theta_2^0)|I\{A_t(z_L, z_2) \cap \mathcal{A}_t^c(z_L, z_1)\},$$

$$\widetilde{V}_{z_L}(z_1, z_2) = E\{|y_{t-1}\psi_\tau(y_t - x_t^T\theta_2^0)|I\{\mathcal{A}_t(z_L, z_2) \cap \mathcal{A}_t^c(z_L, z_1)\}\},$$

where $\theta \in \xi(\Delta), -\Delta < z_L < 0$, $K/n < z_U < \Delta$, and $z_1 < z_2$, $\mathcal{A}_t^c(z_L, z_1)$ is the complement set of $\mathcal{A}_t(z_L, z_1)$. There thus exists $0 < m < M$ and $H > 0$ such that

$$mz_U \le T_{z_L}(z_U) \le Mz_U, \quad Var\{I\{\mathcal{A}_t(z_L, z_U)\}\} \le HT_{z_L}(z_U),$$

$$E\{|y_{t-1}\psi_\tau(y_t - x_t^T\theta_2^0)|I\{\mathcal{A}_t(z_L, z_2) \cap \mathcal{A}_t^c(z_L, z_1)\}\} \le H\{T_{z_L}(z_2) - T_{z_L}(z_1)\},$$

and

$$Var\{|y_{t-1}\psi_\tau(y_t - x_t^T\theta_2^0)|I\{\mathcal{A}_t(z_L, z_2) \cap \mathcal{A}_t^c(z_L, z_1)\}\} \le H\{T_{z_L}(z_2) - T_{z_L}(z_1)\},$$

$$\widetilde{V}_{z_L}(z_1, z_2) \le H\{T_{z_L}(z_2) - T_{z_L}(z_1)\}.$$

By Assumption 3, we have

$$Var\{nT_{n,z_L}(z_U)\} \le nHT_{z_L}(z_U), \quad \text{and} \quad Var\{nV_{n,z_L}(z_U)\} \le nHT_{z_L}(z_U),$$

and

$$Var\{n\widetilde{V}_{n,z_L}(z_1, z_2)\} \le nH\{T_{z_L}(z_2) - T_{z_L}(z_1)\}.$$

Moreover, we can obtain that, for any $\varepsilon > 0$, any $\eta > 0$ and all $n$,

$$P\left\{\sup_{K/n<z_U<\Delta,-\Delta<z_L\leq 0}\left|\frac{1}{nT_{z_L}(z_U)}\sum_{t=1}^{n} I\{\mathcal{A}_t(z_L,z_U)\}-1\right|<\eta\right\}>1-\varepsilon,$$

$$P\left\{\sup_{K/n<z_U<\Delta,-\Delta<z_L\leq 0}\left|\frac{1}{nT_{z_L}(z_U)}\sum_{t=1}^{n} \psi_\tau(y_t-x_t^T\theta_2^0)I\{\mathcal{A}_t(z_L,z_U)\}\right|<\eta\right\}>1-\varepsilon.$$

and

$$P\left\{\sup_{K/n<z_U<\Delta,-\Delta<z_L\leq 0}\left|\frac{1}{nT_{z_L}(z_U)}\sum_{t=1}^{n} y_{t-1}\psi_\tau(y_t-x_t^T\theta_2^0)I\{\mathcal{A}_t(z_L,z_U)\}\right|<\eta\right\}>1-\varepsilon.$$

By Assumption 4, we also have

$$P\left\{\sup_{K/n<z_U<\Delta,-\Delta<z_L\leq 0}\left|\frac{1}{nT_{z_L}(z_U)}\sum_{t=1}^{n} I\{x_t^T\theta_2^0<y_t<x_t^T\theta_2\}I\{\mathcal{A}_t(z_L,z_U)\}\right|<\eta\right\}>1-\varepsilon,$$

and

$$P\left\{\sup_{K/n<z_U<\Delta,-\Delta<z_L\leq 0}\left|\frac{1}{nT_{z_L}(z_U)}\sum_{t=1}^{n} y_{t-1}I\{x_t^T\theta_2^0<y_t<x_t^T\theta_2\}I\{A_t(z_L,z_U)\}\right|<\eta\right\}>1-\varepsilon.$$

Thus we show that for any $\varepsilon > 0$,

$$P\{\widetilde{L}_n^1(z_L,z_U)>0\}>1-\varepsilon.$$

Similarly, we obtain the result for $\widetilde{L}_n^2(z_L,z_U)$, and then that for $\widetilde{L}_n(\theta,r_L^0+z_L,r_U^0+z_U)-\widetilde{L}_n(\theta,r_L^0,r_U^0)$. Together with Claim (a) in the proof of Theorem 1, we finish the proof of (A.5) for $p=d=1$, $z_L<0$ and $z_U>0$. Moreover, the proof for the other cases is similar, and is hence ignored. The proof of (a) is then completed.

As the proof of Claim (b) is routine and similar to Qian (1998), we omit the the details.

Next, we prove Claim (c). Denote $X_t = (x_t^T \widetilde{R}_t, x_t^T(1-\widetilde{R}_t))^T$, $\widehat{V}_n = \sqrt{n}(\widehat{\theta} - \theta^0)$, and $u_{t\tau} = y_t - X_t^T\theta^0$. We have

$$\rho_\tau[(y_t - x_t^T\widehat{\theta}_1)\widetilde{R}_t + (y_t - x_t^T\widehat{\theta}_2)(1-\widetilde{R}_t)] = \rho_\tau[u_{t\tau} - X_t^T(n^{-1/2}\widehat{V})].$$

Thus we denote

$$Z_{n,\tau}(V) = \sum_{t=1}^{n}\{\rho_\tau[u_{t\tau} - X_t^T(n^{-1/2}V)] - \rho_\tau(u_{t\tau})\}, \tag{A.6}$$

and can obtain that if $\widehat{V}_n$ is a minimizer of $Z_{n,\tau}(V)$, then $\widehat{V}_n = \sqrt{n}(\widehat{\theta}_n - \theta^0)$. By Knight's identity (Knight, 1998),

$$\rho_\tau(u-v) - \rho_\tau(u) = -v\psi_\tau(u) + \int_0^v \{I\{u \le s\} - I\{u < 0\}\}ds,$$

we rewrite (A.6) as

$$\begin{aligned} Z_{n,\tau}(V) &= \sum_{t=1}^{n}\{\rho_\tau[u_{t\tau} - X_t^T(n^{-1/2}V)] - \rho_\tau(u_{t\tau})\} \\ &= Z_{n,\tau}^{(1)}(V) + Z_{n,\tau}^{(2)}(V), \end{aligned}$$

where

$$Z_{n,\tau}^{(1)}(V) = -\frac{1}{\sqrt{n}}\sum_{t=1}^{n} X_t^T V \psi_\tau(u_{t\tau}),$$

and

$$Z_{n,\tau}^{(2)}(V) = \sum_{t=1}^{n}\int_0^{\frac{1}{\sqrt{n}}X_t^T V}\{I\{u_{t\tau} \le s\} - I\{u_{t\tau} < 0\}\}.$$

Note that

$$\begin{aligned}\psi_\tau(u_{t\tau}) &= \tau - I\{(y_t - x_t^T\theta_1^0)\widetilde{R}_t + (y_t - x_t^T\theta_2^0)(1-\widetilde{R}_t) < 0\}\\ &= \tau - I\{[y_t - F_t^{-1}(\tau)]\widetilde{R}_t + [y_t - F_t^{-1}(\tau)](1-\widetilde{R}_t) < 0\}\\ &= \tau - I\{y_t - F_t^{-1}(\tau) < 0\}\\ &= \psi_\tau[y_t - F_t^{-1}(\tau)],\end{aligned}$$

and

$$\begin{aligned}X_tX_t^T &= \begin{pmatrix} x_t\widetilde{R}_t \\ x_t(1-\widetilde{R}_t)\end{pmatrix}\begin{pmatrix} x_t^T\widetilde{R}_t & x_t^T(1-\widetilde{R}_t)\end{pmatrix}\\ &= \begin{pmatrix} x_tx_t^T\widetilde{R}_t^2 & x_tx_t^T\widetilde{R}_t(1-\widetilde{R}_t)\\ x_tx_t^T\widetilde{R}_t(1-\widetilde{R}_t) & x_tx_t^T(1-\widetilde{R}_t)^2\end{pmatrix}\\ &= \operatorname{diag}[x_tx_t^T\widetilde{R}_t, \quad x_tx_t^T(1-\widetilde{R}_t)].\end{aligned}$$

Then using the martingale central limit theorem, we have

$$\frac{1}{\sqrt{n}}\sum_{t=1}^n X_t^T\psi_\tau[y_t - F_t^{-1}(\tau)] \xrightarrow{d} W,$$

where $W$ is a $2(p+1)$-dimensional vector normal variate with covariance matrix $\tau(1-\tau)\Omega_0$. On the other hand,

$$\begin{aligned}E\{Z_{n,\tau}^{(2)}(V)|\mathcal{F}_t\} &= \sum_{t=1}^n \int_0^{\frac{1}{\sqrt{n}}X_t^TV} \frac{F_t[s+F_t^{-1}(\tau)] - F_t[F_t^{-1}(\tau)]}{s} s\,ds\\ &= \sum_{t=1}^n \int_0^{\frac{1}{\sqrt{n}}X_t^TV} f_t[F_t^{-1}(\tau)]s\,ds + o_p(1)\\ &= \frac{1}{2n}\sum_{t=1}^n f_t[F_t^{-1}(\tau)]V^TX_t^TX_tV + o_p(1),\end{aligned}$$

and thus

$$E\{Z^{(2)}_{n,\tau}(V)|\mathcal{F}_{t-1}\} \xrightarrow{d} \frac{1}{2}V^T\Omega_1 V.$$

We have

$$\begin{aligned} Z_{n,\tau}(V) &= \sum_{t=1}^{n}\{\rho_\tau(u_{t\tau} - X_t^T(n^{-1/2}V)) - \rho_\tau(u_{t\tau})\} \\ &= Z^{(1)}_{n,\tau}(V) + Z^{(2)}_{n,\tau}(V) \\ &\Rightarrow \frac{1}{2}V^T\Omega_1 V - V^T W = Z_\tau(V), \end{aligned}$$

where " $\Rightarrow$ " denotes the weak convergence. Knight (1998) and Pollard (1991) have shown that if the finite dimensional distributions of $Z_n(\cdot)$ converge weakly to $Z(\cdot)$ and $Z(\cdot)$ has a unique minimum, then the convexity of $Z_n(\cdot)$ implies that $\widehat{V}_n$ converges in distribution to the minimizer of $Z(\cdot)$. Thus by the lemma A of Knight (1989), we have

$$\sqrt{n}(\widehat{\theta}_n - \theta^0) \xrightarrow{d} \tau(1-\tau)\Omega_1^{-1}\Omega_0\Omega_1^{-1}.$$

This finishes the proof of Theorem 2.

# S3 Proof of Theorem 3

Note that $u_{t\tau} = y_t - \theta^T X_t$, arguing as in the proof of Theorem 5 of Li et al. (2015b), for any $v \in R^{2p+2}$, we have

$$\begin{aligned}
L^*(v) &= \sum_{t=1}^{n} \omega_t \rho_\tau(y_t - (\theta + n^{-1/2}v)^T X_t) - \sum_{t=1}^{n} \omega_t \rho_\tau(y_t - \theta^T X_t) \\
&= -v^T \cdot \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \omega_t \psi_\tau(u_{t\tau}) X_t + \sum_{t=1}^{n} \omega_t \int_0^{n^{-1/2}v^T X_t} I(u_{t\tau} \le s) - I(e_{t,\tau} < 0) ds \\
&= -v^T \cdot \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \omega_t \psi_\tau(e_{t,\tau}) X_t + \frac{1}{2} v^T \Omega_1^* v + o_p^*(1),
\end{aligned}$$

where $\Omega_1^* = \frac{1}{n} \sum_{t=1}^{n} \omega_t f_t[F_t^{-1}(\tau)] X_t^T X_t = \Omega_1 + o_p^*(1)$, and the notation $o_p^*(1)$ is referred to the bootstrapped probability space. Moreover, $L^*(v)$ is a convex function with respect to $v$, thus we have the following Bahadur representation

$$\sqrt{n}(\hat{\theta}^* - \theta) = \Omega_1^{-1} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \omega_t \psi_\tau(u_{t\tau}) X_t + o_p^*(1). \tag{A.7}$$

On the other hand, by the proof of Claim (c), we have

$$\sqrt{n}(\hat{\theta} - \theta) = \Omega_1^{-1} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \psi_\tau(u_{t\tau}) X_t + o_p(1). \tag{A.8}$$

(A.7) and (A.8) imply that

$$\sqrt{n}(\hat{\theta}^* - \hat{\theta}) = \Omega_1^{-1} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} (\omega_t - 1) \psi_\tau(u_{t\tau}) X_t + o_p^*(1). \tag{A.9}$$

Similar to the proof of Theorem 5 of Li et al. (2015b), we have the left-hand-side of the (A.9) is tight. Subsequently, we complete the proof of Theorem

3 by the central limit theorem and Cramer-Wold device.

## S4 Proof of Theorem 4

By Theorems 1 and 2, $\widehat{d}$ is consistent with integer value, and the estimators of $\widehat{r}_L$ and $\widehat{r}_U$ are super-consistent. Therefore, we can assume that the true values of $(r_L^0, r_U^0, d^0)$ are known in advance. As the regime indicator $R_t^0$ only depends on $(r_L^0, r_U^0, d^0)$, $R_t^0$ is known as well. For each $0 \le p \le p_{\max}$, let

$$\theta^0_{1,(p)} = \operatorname{argmin} E\{\rho_\tau[(y_t - x_{t,p}^T\theta_{1,(p)})R_t^0]\},$$

and

$$\theta^0_{2,(p)} = \operatorname{argmin} E\{\rho_\tau[(y_t - x_{t,p}^T\theta_{2,(p)})(1 - R_t^0)]\},$$

where notation $_{(p)}$ indicates the dependence on $p$. Furthermore, let

$$\widehat{\sigma}_{1,(p)} = \frac{1}{n_1}\sum_{t=1}^n \rho_\tau\{(y_t - x_{t,p}^T\widehat{\theta}_{1,(p)})R_t^0\} \text{ and } \widehat{\sigma}_{2,(p)} = \frac{1}{n_2}\sum_{t=1}^n \rho_\tau\{(y_t - x_{t,p}^T\widehat{\theta}_{2,(p)})(1 - R_t^0)\}$$

with

$$\widehat{\theta}_{1,(p)} = \operatorname{argmin} \sum_{t=1}^n \rho_\tau\{(y_t - x_{t,p}^T\theta_{1,(p)})R_t^0\},$$

and

$$\widehat{\theta}_{2,(p)} = \operatorname{argmin} \sum_{t=1}^n \rho_\tau\{(y_t - x_{t,p}^T\theta_{2,(p)})(1 - R_t^0)\}.$$

Define $\sigma^0_{1,(p)} = E\{\rho_\tau\{(y_t - x_{t,p}^T\theta_{1,(p)})R_t^0\}\}$ and $\sigma^0_{2,(p)} = E\{\rho_\tau\{(y_t - x_{t,p}^T\theta_{2,(p)})(1 - R_t^0)\}\}$. By the fact that $\widehat{\theta}_{k,(p)} - \widehat{\theta}^0_{k,(p)} = O_p(\frac{1}{\sqrt{n}})$, we obtain that, for $k = 1, 2$, and for any $p$,

$$\widehat{\sigma}_{1,(p)} = \sigma^0_{1,(p)} + o_p(1) \quad \text{and} \quad \widehat{\sigma}_{2,(p)} = \sigma^0_{2,(p)} + o_p(1).$$

Furthermore, we have

$$\sigma^0_{1,(0)} \geq \sigma^0_{1,(1)} \geq \ldots \geq \sigma^0_{1,(p_0)} = \sigma^0_{1,(p_0+1)} = \ldots = \sigma^0_{1,(p_{\max})}$$

and

$$\sigma^0_{2,(0)} \geq \sigma^0_{2,(1)} \geq \ldots \geq \sigma^0_{2,(p_0)} = \sigma^0_{2,(p_0+1)} = \ldots = \sigma^0_{2,(p_{\max})}.$$

When $p < p_0$, we have

$$\text{BIC}(p) - \text{BIC}(p_0) = 2n_1 \ln(\widehat{\sigma}_{1,(p)}/\widehat{\sigma}_{1,(p_0)}) + 2n_2 \ln(\widehat{\sigma}_{2,(p)}/\widehat{\sigma}_{2,(p_0)}) + (p-p_0)\ln(n_1 n_2).$$

Therefore, we have $\sigma^0_{1,(p_0-1)} > \sigma^0_{1,(p_0)}$ when $|\theta^0_{1,p_0}| \neq 0$, and $\sigma^0_{2,(p_0-1)} > \sigma^0_{2,(p_0)}$ when $|\theta^0_{2,p_0}| \neq 0$. Denote $C = \ln[\sigma^0_{1,(p)}/\sigma^0_{1,(p_0)}]P(R^0_t = 1) + \ln[\sigma^0_{2,(p)}/\sigma^0_{2,(p_0)}]P(R^0_t = 0) > 0$. Then

$$\text{BIC}(p) - \text{BIC}(p_0) = Cn + o_p(n). \tag{A.10}$$

When $p > p_0$, it is easy to obtain that $n_1 \widehat{\sigma}_{1,(p)}/\widehat{\sigma}_{1,(p_0)} = O_p(1)$. Thus $n_1 \ln(\widehat{\sigma}_{1,(p)}/\widehat{\sigma}_{1,(p_0)}) = O_p(1)$. In the same way, we have $n_2 \ln(\widehat{\sigma}_{2,(p)}/\widehat{\sigma}_{2,(p_0)}) = O_p(1)$. Then by $n_1 = nP\{R^0_t = 1\}$ and $n_2 = nP\{R^0_t = 0\}$, with $p > p_0$ we have

$$\begin{aligned}
&\text{BIC}(p) - \text{BIC}(p_0)\\
&\quad = 2n_1 \ln(\widehat{\sigma}_{1,(p)}/\widehat{\sigma}_{1,(p_0)}) + 2n_2 \ln[\widehat{\sigma}_{2,(p)}/\widehat{\sigma}_{2,(p_0)}] + (p - p_0)\ln(n_1 n_2)\\
&\quad = (p - p_0)\ln(n_1 n_2) + O_p(1)\\
&\quad = 2(p - p_0)\ln(n) + (p - p_0)\ln[P(R_t^0 = 1)P(R_t^0 = 0)] + O_p(1).
\end{aligned}$$

Together with (A.10), we have $\text{BIC}(p) - \text{BIC}(p_0) > 0$ when $p \neq p_0$. We complete the proof of Theorem 4.

## References


Chan, K. S. (1991). Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model. *The Annals of Statistics*, 21:520–533.

Huber, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, volume 1, pages 221–233.

Knight, K. (1989). Limit theory for autoregressive-parameter estimates in an infinite-variance random walk. *Canadian Journal of Statistics*, 17:261–278.

Knight, K. (1998). Limiting distributions for $l_1$ regression estimators under general conditions. *The Annals of Statistics*, 26:755–770.

Li, G., Guan, B., Li, W. K., and Yu, P. L. H. (2015a). Hysteretic autoregressive time series models. *Biometrika*, 102:717–723.

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

Pollard, D. (1991). Asymptotics for least absolute deviation regression estimators. *Econometric Theory*, 7:186–199.

Qian, L. (1998). On maximum likelihood estimators for a threshold autoregression. *Journal of Statistical Planning and Inference*, 75:21–46.