Supplementary material for "Linear double autoregression"

Qianqian Zhu, Yao Zheng and Guodong Li
Shanghai University of Finance and Economics
and University of Hong Kong

Abstract

This supplementary material provides four additional simulation experiments to evaluate the finite-sample performance of the proposed inference tools for the linear double autoregressive (AR) model, and also contains detailed proofs of all lemmas and theorems in the paper.

1 Additional simulation experiments

As a complement to the simulation studies in the paper, we present four additional experiments to evaluate the finite-sample performance of the proposed inference tools.

The first experiment aims to compare the performance of the self-weighted quantile regression estimator $\widetilde{\lambda}_{\tau n}$ for a specific quantile level τ under different weights $\{w_t\}$. We consider two choices of $\{w_t\}$: $\{1/(1+|y_{t-1}|)\}$ (denoted by W_1) and $\{1/(1+\widetilde{\beta}^{int}|y_{t-1}|)\}$ (denoted by W_2), where $\widetilde{\beta}^{int}$ is calculated from (3.4). The data are generated from

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

where $\{\varepsilon_t\}$ are *i.i.d.* random variables following the normal, Student's t_3 or Cauchy distribution 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. Table 1 presents the biases and the empirical standard deviations (ESDs) of $\widetilde{\lambda}_{\tau n}$ at quantile levels $\tau = 0.25$

and 0.35. Clearly, both the biases and the ESDs decrease as the sample size increases. It can also be seen that the weights W_2 slightly outperform W_1 when the sample size n is larger, which confirms our asymptotic result in Section 3.1 that the estimator is most efficient at $w_t = \sigma_t^{-1}$. On the other hand, the weights W_2 perform worse than W_1 when $n = 200$, mainly due to a less accurate $\widetilde{\beta}^{int}$ for small sample sizes. Moreover, $\widetilde{\phi}_{\tau n}$ has smaller ESDs for $\tau = 0.35$, which is as expected since there are more data points around the quantile level closer to the center. On the contrary, $\widetilde{\beta}_{\tau n}$ has smaller ESDs for $\tau = 0.25$, probably because its asymptotic variance as shown in Theorem 3 is smaller when b_τ is larger in magnitude. Finally, when the distribution of ε_t is more heavy-tailed, the performance of $\widetilde{\phi}_{\tau n}$ is improved in the sense that both its biases and ESDs decrease. However, the results for $\widetilde{\beta}_{\tau n}$ are mixed: they get better for $\tau = 0.35$ but worse for $\tau = 0.25$, when the tails become heavier. Based on the results, we recommend using W_2 when the sample size n is relatively large, say $n \geqslant 500$, and W_1 when n is relatively small.

The objective of the second experiment is to examine the performance of the optimal doubly weighted quantile regression estimator $\widehat{\lambda}_n^{opt}$ in Section 3.2. We preserve all settings from the first experiment and employ the quantile levels $\tau_k = k/10$ with $k = 1, \ldots, 9$. 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 2 lists the biases, empirical standard deviations (ESDs) and asymptotic standard deviations (ASDs) of $\widehat{\lambda}_n^{opt}$. 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. Finally, the results show that for the doubly weighted quantile regression estimator, the weights W_2 slightly outperform W_1 in terms of the biases, the ESDs and the ASDs, and hence we focus on W_2 in the following experiments.

In the third experiment, we examine the sample approximations for the asymptotic distributions of the residual ACFs $\widehat{\rho}_\ell$ and $\widehat{r}_\ell$ in Section 4. All settings are preserved from the first two experiments, and the maximum lag L is set at six. To transform the residuals from the fitted models, we consider the function G being the distribution function of the standard normal, Student's t_3 or standard Cauchy distribution, denoted by G_N, G_T and G_C respectively. The reference distributions are approximated based on $B = 10000$ generated multivariate normal random numbers. Table 3 provides the empirical coverage rates (ECRs) of $\widehat{\rho}_\ell$ and $\widehat{r}_\ell$ at lags $\ell = 2$, 4 and 6, at the 5% significance level. It shows that all the ECRs are close to their nominal values when n is as small as 200, and the results for the three transformations are quite similar.

The last experiment aims to evaluate the performance of the two goodness-of-fit tests, $Q_1^{BP}(L)$ and $Q_2^{BP}(L)$, proposed in Section 4. The data generating process is

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

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 second 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 the misspecification in the conditional mean, and the case of $c_2 > 0$ corresponds to the misspecification in the conditional scale. We consider three departure levels, 0.1, 0.2 and 0.3, and set the significance level at 5%. The transformations G_N, G_T and G_C are considered as in the previous experiment.

Tables 4 and 5 report the rejection rates of $Q_1^{BP}(6)$ and $Q_2^{BP}(6)$, respectively. Firstly, all sizes are close to the nominal rate even when the sample size n is as small as 200, and all powers increase as the sample size or the departure level increases. Secondly, $Q_1^{BP}(6)$ performs well in detecting the misspecification in the conditional mean (i.e., $c_1 \neq 0$ and $c_2 = 0$) and is especially powerful when the distribution of ε_t is more heavy-tailed, but it has little power for that in the scale structure (i.e., $c_1 = 0$ and $c_2 > 0$) regardless of the distribution of ε_t. On the other hand, $Q_2^{BP}(6)$ performs well in detecting the misspecification in the conditional scale, especially when the distribution of ε_t is light-tailed. Its power for the misspecification in the conditional mean may have opposite results for different distributions of ε_t: it is useless when the innovations follow the normal or Student's t_3 distribution, but is surprisingly powerful when they follow the Cauchy distribution. These findings seem consistent with the result in the first two experiments 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. Lastly, comparing the three transformations G_N, G_T and G_C for the residuals, it can be seen that their results are fairly similar for $Q_1^{BP}(6)$, whereas for $Q_2^{BP}(6)$, the transformation G_C outperforms the other two by a visible margin, probably because the Cauchy distribution function G_C is more spread out, and consequently, the serial dependence in the original sequence is better kept.

We summarize our findings from the four simulation experiments as follows:

(1) When the distribution of the innovation is more heavy-tailed, the performance of the proposed inference tools becomes better for ϕ_0 but worse for β_0.

(2) For the weights $\{w_t\}$ in Section 3.1, we recommend W_2 when the sample size is relatively large, say $n \geqslant 500$, and W_1 when it is relatively small.

(3) The transformation G_C for the residuals may be more favorable for constructing the goodness-of-fit test statistics.

(4) The test statistics $Q_1^{BP}(L)$ and $Q_2^{BP}(L)$ should be used in conjunction to check the adequacy of fitted linear double AR models.

2 Technical details

2.1 Proof of Theorem 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. We begin by proving that $\{Y_t\}$ and $\{Y_t^*\}$ are Markov chains with the same transition probability.

Let $\mathcal{B}^p$ be the class of Borel sets of $\mathbb{R}^p$ and ν_p be the Lebesgue measure on $(\mathbb{R}^p, \mathcal{B}^p)$. Let $m : \mathbb{R}^p \to \mathbb{R}$ be the projection map onto the first coordinate, i.e. $m(x) = x_1$ for $x = (x_1, \ldots, x_p)'$. Then, $\{Y_t\}$ is a homogeneous Markov chain on the state space $(\mathbb{R}^p, \mathcal{B}^p, \nu_p)$, with transition probability

$$P(x, A) = \int_{m(A)} \frac{1}{1 + x_{\mathrm{a}}'\beta} f\left(\frac{z - x'\phi}{1 + x_{\mathrm{a}}'\beta}\right) dz, \quad x \in \mathbb{R}^p \text{ and } A \in \mathcal{B}^p,$$

where $x_{\mathrm{a}} = (|x_1|, \ldots, |x_p|)'$, $\phi = (\phi_1, ..., \phi_p)'$, $\beta = (\beta_1, \cdots, \beta_p)'$, and $f(\cdot)$ is the density function of ε_t. Note that $\{Y_t^*\}$ can be rewritten in the vector form,

$$Y_t^* = A_t Y_{t-1}^* + e_t, \tag{S.2}$$

where $e_t = (\varepsilon_t, 0, \ldots, 0)'$, and $\{A_t\}$ are *i.i.d.* random matrices independent of $\{e_t\}$. Thus, $\{Y_t^*\}$ is also a homogeneous Markov chain on $(\mathbb{R}^p, \mathcal{B}^p, \nu_p)$. To verify that $\{Y_t\}$ and $\{Y_t^*\}$ have the same transition probability, it is sufficient to show that the conditional characteristic functions $E\{\exp(iu'Y_t)|Y_{t-1} = x\}$ and $E\{\exp(iu'Y_t^*)|Y_{t-1}^* = x\}$ are the same, where $u = (u_1, \ldots, u_p)'$. Since $E\{\exp(is\varepsilon_t)\} = \exp(-\sigma|s|)$, we have

$$\begin{aligned}
E\{\exp(iu'Y_t)|Y_{t-1} = x\} &= \exp\left(i\sum_{j=2}^{p} u_j x_{j-1}\right) E\{\exp(iu_1 y_t)|Y_{t-1} = x\} \\
&= \exp\left(i\sum_{j=2}^{p} u_j x_{j-1} + iu_1 x'\phi\right) E[\exp\{iu_1(1 + x_a'\beta)\varepsilon_t\}] \\
&= \exp\left\{i\sum_{j=2}^{p} u_j x_{j-1} + iu_1 x'\phi - \sigma|u_1|(1 + x_a'\beta)\right\}.
\end{aligned}$$

On the other hand, since ξ_{it} and ε_t are *i.i.d.*, we have

$$\begin{aligned}
E\{\exp(iu'Y_t^*)|Y_{t-1}^* = x\} &= \exp\left(i\sum_{j=2}^{p} u_j x_{j-1}\right) E\{\exp(iu_1 y_t^*)|Y_{t-1}^* = x\} \\
&= \exp\left(i\sum_{j=2}^{p} u_j x_{j-1} + iu_1 x'\phi\right) E\{\exp(iu_1 \sum_{i=1}^{p} \beta_i |x_i| \xi_{it} + iu_1 \varepsilon_t)\} \\
&= \exp\left\{i\sum_{j=2}^{p} u_j x_{j-1} + iu_1 x'\phi - \sigma|u_1|(1 + x_a'\beta)\right\}.
\end{aligned}$$

This proves that $\{Y_t\}$ and $\{Y_t^*\}$ have the same transition probability.

We can further verify that the p-step transition probability of $\{Y_t\}$ is

$$P^p(x, A) = \int_A \prod_{i=1}^{p} \frac{1}{1 + X_{\mathrm{a},i-1}'\beta} f\left(\frac{z_i - X_{i-1}'\phi}{1 + X_{\mathrm{a},i-1}'\beta}\right) dz_1 \ldots dz_p, \tag{S.3}$$

where $X_i = (z_i, \ldots, z_1, x_1, \ldots, x_{p-i})'$ and $X_{\mathrm{a},i} = (|z_i|, \ldots, |z_1|, |x_1|, \ldots, |x_{p-i}|)'$. Observe that, from Assumption 1, the transition density kernel in (S.3) is positive everywhere, and then $\{Y_t\}$ is ν_p-irreducible.

First suppose $\gamma < 0$. Then, there exists an integer s such that $E(\ln \|A_1 \cdots A_s\|) < 0$. Let $\widetilde{A}_t = \prod_{i=0}^{s-1} A_{t-i}$ and define $q(u) = E(\|\widetilde{A}_t\|^u)$. Due to the continuity of the density $f(\cdot)$, it can be shown that $q(u)$ is continuous and differentiable on $[0, 1)$, and its derivative function has the form of $\dot{q}(u) = E(\|\widetilde{A}_t\|^u \ln \|\widetilde{A}_t\|)$. For any given $\kappa^* \in (0,1)$, it can be verified that $E\{\sup_{u\in[0,\kappa^*]}(\|\widetilde{A}_t\|^u \ln \|\widetilde{A}_t\|)\} < \infty$, which, together with the dominated convergence theorem, implies that $\lim_{u\to 0} \dot{q}(u) = E(\ln \|\widetilde{A}_t\|) < 0$. As a result, there exists a constant $0 < \kappa < \kappa^*$ such that

$$E(\|\widetilde{A}_t\|^\kappa) < q(0) = 1.$$

We next prove that s-step Markov chain $\{Y_{ts}\}$ satisfies Tweedie's drift criterion (Tweedie, 1983, Theorem 4), i.e., there exists a small set G with $\nu_p(G) > 0$ and a non-negative continuous function $g(x)$ such that

$$E\left\{g(Y_{ts})|Y_{(t-1)s} = x\right\} \leqslant (1-\epsilon)g(x), \quad x \notin G, \tag{S.4}$$

$$E\left\{g(Y_{ts})|Y_{(t-1)s} = x\right\} \leqslant M, \quad x \in G, \tag{S.5}$$

for some constant $0 < \epsilon < 1$ and $0 < M < \infty$. By iterating the random coefficient AR model (S.2) s times, we have that

$$Y_{ts}^* = \widetilde{A}_{ts} Y_{(t-1)s}^* + \left(e_{ts} + \sum_{j=1}^{s-1}\prod_{r=0}^{j-1} A_{ts-r} e_{ts-j}\right). \tag{S.6}$$

Let $g(x) = 1 + \|x\|^\kappa$, and it can be verified that

$$\begin{aligned} E[g(Y_{ts}^*)|Y_{(t-1)s}^* = x] &\leqslant 1 + E(\|\widetilde{A}_{ts}\|^\kappa \cdot \|x\|^\kappa) + E\left(\left\|e_{ts} + \sum_{j=1}^{s-1}\prod_{r=0}^{j-1} A_{ts-r} e_{ts-j}\right\|^\kappa\right) \\ &= C + g(x)E(\|\widetilde{A}_{ts}\|^\kappa), \end{aligned}$$

where $C = 1 + E(\|e_{ts} + \sum_{j=1}^{s-1}\prod_{r=0}^{j-1} A_{ts-r}e_{ts-j}\|^\kappa) - E(\|\widetilde{A}_{ts}\|^\kappa) < \infty$. Note that $E(\|\widetilde{A}_{ts}\|^\kappa) = E(\|\widetilde{A}_t\|^\kappa) < 1$. Then there exists a $L > 0$ such that

$$E\{g(Y_{ts}^*)|Y_{(t-1)s}^* = x\} \leqslant (1-\epsilon)g(x), \quad \|x\| > L, \tag{S.7}$$

$$E\{g(Y_{ts}^*)|Y_{(t-1)s}^* = x\} \leqslant M < \infty, \quad \|x\| \leqslant L, \tag{S.8}$$

where $\epsilon = 0.5 - 0.5E(\|\widetilde{A}_t\|^\kappa)$, and $\nu_p(G) > 0$ with $G = \{x : \|x\| \leqslant L\}$. Note that $\{Y_t\}$ and $\{Y_t^*\}$ have the same transition probability, and then Claims (S.4) and (S.5) are implied by (S.7) and (S.8).

Moreover, $\{Y_{ts}\}$ is a Feller chain since, for each bounded continuous function $g^*(\cdot)$, $E\{g^*(Y_{ts})|Y_{(t-1)s} = x\}]$ is continuous with respect to x, and $\{Y_{ts}\}$ is also ν_p-irreducible.

This implies that G is a small set. As a result, from 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 $\pi(\cdot)$, and

$$\int \|Y_{ts}\|^{\kappa} d\pi = \int_{\mathbb{R}^p} g(x)\pi(dx) - 1 < \infty. \tag{S.9}$$

By Lemma 3.1 of Tjøstheim (1990), $\{Y_t\}$ is geometrically ergodic, and it is the unique strictly stationary solution to model (2.1). Moreover, it is implied by (S.9) that $E(|y_t|^{\kappa}) < \infty$.

Finally we prove the necessity. Suppose that there exists a strictly stationary solution $\{y_t\}$ to model (2.1), and then the Markov chain $\{Y_t\}$ has a stationary distribution $\pi(\cdot)$. Generate Y_0^* with the distribution of $\pi(\cdot)$ and, by iterating the random coefficient AR model in (S.2), it leads to $\{Y_t^* : t \in N\}$, which is a strictly stationary solution to model (S.2) since $\{Y_t\}$ and $\{Y_t^*\}$ have the same transition probability. Moreover, it is also nonanticipative.

By letting $s = p$ in (S.6), we can obtain a vector random coefficient AR model,

$$Y_{tp}^* = \widetilde{A}_{tp} Y_{(t-1)p}^* + B_{tp}, \tag{S.10}$$

where $\widetilde{A}_t = \prod_{i=0}^{p-1} A_{t-i}$, $B_{tp} = e_{tp} + \sum_{j=1}^{p-1}\prod_{r=0}^{j-1} A_{tp-r} e_{tp-j}$, and $\{(\widetilde{A}_{tp}, B_{tp}) : t \in N\}$ is an independent and identically distributed sequence. For a $\kappa^* \in (0,1)$, it holds that $\ln^+(x) \leqslant \max\{x^{\kappa^*}, C\}$ for $x > 0$ and a positive number C, where $\ln^+(x) = \max\{\ln(x), 0\}$. Moreover, $\{A_t\}$ are independent and identically distributed random matrices, and the κ^*th moment of Cauchy distributions is finite. As a result,

$$E(\ln^+ \|\widetilde{A}_{tp}\|) < \infty \quad \text{and} \quad E(\ln^+ \|B_{tp}\|) < \infty. \tag{S.11}$$

In addition, $\{Y_{tp}^* : t \in N\}$ is a nonanticipative and strictly stationary solution to (S.10).

From (S.3), it holds that

$$P(Y_{tp}^* \in A | Y_{(t-1)p}^* = x) = P(Y_p^* \in A | Y_0^* = x) = P^p(x, A) > 0$$

as $\nu_p(A) > 0$. Let H be any affine invariant subspace of $\mathbb{R}^p$ under model (S.10), i.e. $\{\widetilde{A}_{tp}x + B_{tp} : x \in H\} \subseteq H$ with probability one (Bougerol and Picard, 1992). If $\nu_p(\mathbb{R}^p - H) \neq 0$, then for any $x \in H$,

$$\begin{aligned} P(\widetilde{A}_{tp}x + B_{tp} \in H) &= P(Y_{tp}^* \in H | Y_{(t-1)p}^* = x) \\ &= P(Y_{tp}^* \in \mathbb{R}^p | Y_{(t-1)p}^* = x) - P(Y_{tp}^* \in \mathbb{R}^p - H | Y_{(t-1)p}^* = x) < 1. \end{aligned}$$

As a result, $\mathbb{R}^p$ is the unique affine invariant subspace, and hence model (S.10) is irreducible. Applying Theorem 2.5 of Bougerol and Picard (1992), we have that the corresponding top Lyapounov exponent is strictly negative, i.e

$$\widetilde{\gamma} = \inf\{\frac{1}{t}E(\ln \|\widetilde{A}_p\widetilde{A}_{2p}\cdots\widetilde{A}_{tp}\|), t \geqslant 1\} < 0, \tag{S.12}$$

which implies that $\gamma \leqslant \widetilde{\gamma}/p < 0$ since $\widetilde{A}_p\widetilde{A}_{2p}\cdots\widetilde{A}_{tp} = A_1A_2\cdots A_{tp}$. This completes the proof.

2.2 Proof of Theorem 2

We can define the Markov chain $\{Y_t\}$ and its state space as in the proof of Theorem 1. Note that, for the p-step transition probability, its density kernel is positive everywhere due to Assumption 1. As a result, $\{Y_t\}$ is ν_p-irreducible.

It can be verified that

$$\begin{aligned} E(|y_{t+1}|^\kappa \mid Y_t = x) &\leqslant \sum_{i=1}^p E(|\phi_i\text{sign}(x_i) + \beta_i\varepsilon_{t+1}|^\kappa)|x_i|^\kappa + E(|\varepsilon_{t+1}|^\kappa) \\ &\leqslant \sum_{i=1}^p a_i|x_i|^\kappa + E(|\varepsilon_{t+1}|^\kappa), \end{aligned}$$

where $x = (x_1, \ldots, x_p)'$ and $a_i = \max\{E(|\phi_i + \beta_i\varepsilon_t|^\kappa), E(|\phi_i - \beta_i\varepsilon_t|^\kappa)\}$ for $1 \leqslant i \leqslant p$. Note that $\sum_{i=1}^p a_i < 1$, and we can then find positive values $\{r_1, \ldots, r_{p-1}\}$ such that

$$a_p < r_{p-1} < 1 - \sum_{i=1}^{p-1} a_i \quad \text{and} \quad a_{i+1} + r_{i+1} < r_i < 1 - \sum_{k=1}^{i} a_k \text{ for } 1 \leqslant i \leqslant p-2. \tag{S.13}$$

Consider the test function $g(x) = 1 + |x_1|^\kappa + \sum_{i=1}^{p-1} r_i|x_{i+1}|^\kappa$, and we have that

$$\begin{aligned} &E\{g(Y_{t+1})|Y_t = x\} \\ &\qquad \leqslant 1 + \sum_{i=1}^p a_i|x_i|^\kappa + \sum_{i=1}^{p-1} r_i|x_i|^\kappa + E(|\varepsilon_{t+1}|^\kappa) \\ &\qquad = 1 + (a_1 + r_1)|x_1|^\kappa + \sum_{i=2}^{p-1} \frac{a_i + r_i}{r_{i-1}} r_{i-1}|x_i|^\kappa + \frac{a_p}{r_{p-1}} r_{p-1}|x_p|^\kappa + E(|\varepsilon_{t+1}|^\kappa) \\ &\qquad \leqslant \rho g(x) + 1 - \rho + E(|\varepsilon_{t+1}|^\kappa), \end{aligned}$$

where, from (S.13),

$$\rho = \max\left\{a_1 + r_1, \frac{a_2 + r_2}{r_1}, \cdots, \frac{a_{p-1} + r_{p-1}}{r_{p-2}}, \frac{a_p}{r_{p-1}}\right\} < 1.$$

Denote $\epsilon = 1 - \rho - \{1 - \rho + E(|\varepsilon_{t+1}|^\kappa)\}/g(x)$, and $G = \{x : \|x\| \leqslant L\}$, where L is a positive constant such that $g(x) > 1 + E(|\varepsilon_{t+1}|^\kappa)/(1-\rho)$ as $\|x\| > L$. We can verify that

$$E\{g(Y_{t+1})|Y_t = x\} \leqslant (1-\epsilon)g(x), \quad x \notin G,$$

and

$$E\{g(Y_{t+1})|Y_t = x\} \leqslant M < \infty, \quad x \in G,$$

i.e. Tweedie's drift criterion (Tweedie, 1983, Theorem 4) holds. Moreover, $\{Y_t\}$ is a Feller chain since, for each bounded continuous function $g^*(\cdot)$, $E\{g^*(Y_t)|Y_{t-1} = x\}$ is continuous with respect to x, and then G is a small set. As a result, from Theorem 4(ii) in Tweedie (1983) and Theorems 1 and 2 in Feigin and Tweedie (1985), $\{Y_t\}$ is geometrically ergodic with a unique stationary distribution $\pi(\cdot)$, and

$$\int_{\mathbb{R}^p} g(x)\pi(dx) = 1 + \left(1 + \sum_{i=1}^{p-1} r_i\right) E(|y_t|^\kappa) < \infty,$$

which implies that $E(|y_t|^\kappa) < \infty$. This accomplishes the proof.

2.3 Proofs of Lemma 1 and Theorem 3

Proof of Lemma 1. Denote by θ_τ the parameter vector corresponding to the true value $\theta_{\tau 0}$, and define the function $L_n(\theta_\tau) = \sum_{t=p+1}^n w_t \rho_\tau(y_t - x_t'\theta_\tau)$. Note that, for $u \neq 0$,

$$\rho_\tau(u - \nu) - \rho_\tau(u) = -\nu\psi_\tau(u) + \int_0^\nu \{I(u \leqslant s) - I(u \leqslant 0)\}ds, \tag{S.14}$$

where $\psi_\tau(u) = \tau - I(u < 0)$; see Knight (1998). For any fixed $u \in \mathbb{R}^{2p+1}$, applying (S.14) we have

$$\begin{aligned} &L_n(\theta_{\tau 0} + n^{-1/2}u) - L_n(\theta_{\tau 0}) \\ &\quad = \sum_{t=p+1}^n w_t[\rho_\tau\{(\varepsilon_t - b_\tau)\sigma_t - n^{-1/2}x_t'u\} - \rho_\tau\{(\varepsilon_t - b_\tau)\sigma_t\}] \\ &\quad = -\frac{u'}{\sqrt{n}} \sum_{t=p+1}^n \psi_\tau(\varepsilon_t - b_\tau)w_t x_t + \sum_{t=p+1}^n \xi_t(u), \end{aligned} \tag{S.15}$$

where $\sigma_t = 1 + Y_{a,t-1}'\beta_0$ and

$$\xi_t(u) = w_t \int_0^{n^{-1/2}x_t'u} \left\{I(\varepsilon_t \leqslant \sigma_t^{-1}s + b_\tau) - I(\varepsilon_t \leqslant b_\tau)\right\} ds.$$

We can further verify that, by Taylor expansion,

$$
\begin{aligned}
\sum_{t=p+1}^{n} \xi_t(u) &= \sum_{t=p+1}^{n} E\{\xi_t(u) \mid \mathcal{F}_{t-1}\} + R_{1n}(u) \\
&= \sum_{t=p+1}^{n} w_t \int_0^{n^{-1/2}x_t'u} \{F(\sigma_t^{-1}s + b_\tau) - F(b_\tau)\} ds + R_{1n}(u) \\
&= \frac{1}{2} f(b_\tau) u' \left(\frac{1}{n} \sum_{t=p+1}^{n} \sigma_t^{-1} w_t x_t x_t' \right) u + R_{2n}(u) + R_{1n}(u), \qquad \text{(S.16)}
\end{aligned}
$$

where $R_{1n}(u) = \sum_{t=p+1}^{n} [\xi_t(u) - E\{\xi_t(u) \mid \mathcal{F}_{t-1}\}]$ and

$$
R_{2n}(u) = \sum_{t=p+1}^{n} w_t \int_0^{n^{-1/2}x_t'u} \sigma_t^{-1} s \{f(\sigma_t^{-1}s^* + b_\tau) - f(b_\tau)\} ds
$$

with s^* between 0 and s.

By the compactness of the parameter space Λ, we have $\min\{\beta_1, \ldots, \beta_p\} \geqslant \underline{\omega}$ for some $\underline{\omega} > 0$, and then

$$
\sup_{\lambda \in \Lambda} \left\| \frac{Y_{t-1}}{1 + Y_{a,t-1}'\beta} \right\| \leqslant \sup_{\lambda \in \Lambda} \frac{\sum_{i=1}^{p} |y_{t-i}|}{1 + Y_{a,t-1}'\beta} \leqslant \frac{\sum_{i=1}^{p} |y_{t-i}|}{1 + \underline{\omega} \sum_{i=1}^{p} |y_{t-i}|} \leqslant \frac{1}{\underline{\omega}} \qquad \text{(S.17)}
$$

uniformly for all $1 \leqslant t \leqslant n$. It is then implied by Assumptions 2 and 3 that

$$
\begin{aligned}
|R_{2n}(u)| &\leqslant \sup_{0 \leqslant x \leqslant n^{-1/2}C} |f(x + b_\tau) - f(b_\tau)| \sum_{t=p+1}^{n} w_t \int_0^{n^{-1/2}|x_t'u|} \sigma_t^{-1} s ds \\
&= \frac{1}{2} \sup_{0 \leqslant x \leqslant n^{-1/2}C} |f(x + b_\tau) - f(b_\tau)| u' \left(\frac{1}{n} \sum_{t=p+1}^{n} \sigma_t^{-1} w_t x_t x_t' \right) u \\
&= o_p(1), \qquad \text{(S.18)}
\end{aligned}
$$

where $C = (1 + 2\underline{\omega}^{-1})\|u\|$ is a constant. For $R_{1n}(u)$, it can be similarly shown that

$$
\begin{aligned}
E[\xi_t^2(u)] &\leqslant \frac{1}{\sqrt{n}} E\left[w_t^2 |x_t'u| \int_0^{n^{-1/2}|x_t'u|} \{I(\varepsilon_t \leqslant \sigma_t^{-1}s + b_\tau) - I(\varepsilon_t \leqslant b_\tau)\} ds \right] \\
&= \frac{1}{\sqrt{n}} E\left[w_t^2 |x_t'u| \int_0^{n^{-1/2}|x_t'u|} \{F(\sigma_t^{-1}s + b_\tau) - F(b_\tau)\} ds \right] \\
&\leqslant \frac{\|u\|^3}{2n^{3/2}} \sup_{0 \leqslant x \leqslant n^{-1/2}C} f(x + b_\tau) E(\sigma_t^{-1} w_t^2 \|x_t\|^3) \\
&= o(n^{-1}),
\end{aligned}
$$

for all t, which implies that

$$
E\{R_{1n}^2(u)\} \leqslant \sum_{t=p+1}^{n} E\{\xi_t^2(u)\} = o(1). \qquad \text{(S.19)}
$$

By the central limit theorem and the ergodic theorem, together with (S.15), (S.16), (S.18) and (S.19), it can be verified that

$$L_n(\theta_{\tau 0} + n^{-1/2}u) - L_n(\theta_{\tau 0}) \to -u'T + \frac{1}{2}f(b_\tau)u'\Omega_0(w)u$$

in distribution as $n \to \infty$, where T is a normal random variable with mean zero and variance matrix $\tau(1-\tau)E(w_t^2 x_t x_t')$. Applying Corollary 2 in Knight (1998), together with the convexity of $L_n(\theta_\tau)$, we have

$$\sqrt{n}(\widetilde{\theta}_{\tau n} - \theta_{\tau 0}) \to N\left(0, \frac{\tau(1-\tau)}{f^2(b_\tau)}\Omega_1(w)\right)$$

in distribution as $n \to \infty$. □

Proof of Theorem 3. By the Delta method (van der Vaart, 1998, Chapter 3), we can show that

$$\sqrt{n}(\widetilde{\lambda}_{\tau n} - \lambda_0) = \begin{pmatrix} -b_\tau^{-1}\beta_0 & -b_\tau^{-1}I_p & 0 \\ 0 & 0 & I_p \end{pmatrix} \sqrt{n}(\widetilde{\theta}_{\tau n} - \theta_{\tau 0}) + o_p(1),$$

which, together with Lemma 1, implies the asymptotic normality result of $\sqrt{n}(\widetilde{\lambda}_{\tau n} - \lambda_0)$.

To prove the minimum of $\Omega_1(w)$, as in Xu (2017), we consider the *i.i.d.* samples $(x_1, z_1), \ldots, (x_n, z_n)$ from $z_t = \dfrac{x_t'}{\sigma_t}\gamma + e_t$, where $\{e_t\}$ are *i.i.d.* standard normal and are independent of $\{x_t\}$, and γ is the unknown parameter to be estimated from the data. Consider the weighted least-squares estimation of γ:

$$\widehat{\gamma}(\lambda) = \operatorname*{argmin}_r \sum_{t=1}^n \lambda_t \left(z_t - \frac{x_t'}{\sigma_t}r\right)^2, \quad \text{with weights } \lambda_t = \sigma_t w_t. \tag{S.20}$$

By the classical least-squares estimation theory and the central limit theorem, we have

$$\sqrt{n}[\widehat{\gamma}(\lambda) - \gamma] \to N(0, \Omega_1(w))$$

in distribution as $n \to \infty$. On the other hand, by letting $\lambda_t \equiv 1$ or equivalently $w_t = \sigma_t^{-1}$ in (S.20), the resulting ordinary least-squares estimator has asymptotic covariance matrix $\Omega_1(\sigma_t^{-1})$. Since e_t has a standard normal distribution, the ordinary least-squares estimator is exactly the maximum likelihood estimator, which is the most efficient and has the smallest variance. Thus, we conclude that $\Omega_1(w) \geqslant \Omega_1(\sigma_t^{-1})$. That is, $\Omega_1(w)$ is minimized at $w_t = \sigma_t^{-1}$. Finally, $\Omega_2(w) \geqslant \Omega_2(\sigma_t^{-1})$ follows from the fact that, for a symmetric positive semidefinite matrix A and a matrix B, the matrix BAB' is also positive semidefinite. □

2.4 Proof of Theorem 4

Let

$$\Sigma_2(\tau) = \Sigma_1^{-1}(\tau) \begin{pmatrix} -\beta_0 & I_p & 0 \\ 0 & 0 & I_p \end{pmatrix}.$$

From the proofs of Lemma 1 and Theorem 3, we have the Bahadur representation,

$$\sqrt{n}(\widetilde{\lambda}_{\tau n} - \lambda_0) = \Sigma_2(\tau)\Omega_0^{-1} \frac{1}{\sqrt{n}} \sum_{t=p+1}^{n} \psi_\tau(\varepsilon_t - b_\tau) \frac{x_t}{\sigma_t} + o_p(1), \tag{S.21}$$

where $\Omega_0^{-1} = \Omega_1$ since $w_t = \sigma_t^{-1}$. Let

$$H(\varepsilon_t) = H(\varepsilon_t; \Pi) = \sum_{k=1}^{K} \psi_{\tau_k}(\varepsilon_t - b_{\tau_k})\pi_k \Sigma_2(\tau_k).$$

As a result, by the central limit theorem, we have

$$\sqrt{n}(\widehat{\lambda}_n - \lambda_0) = \frac{1}{\sqrt{n}} \sum_{t=p+1}^{n} H(\varepsilon_t)\Omega_0^{-1} \frac{x_t}{\sigma_t} + o_p(1) \to N(0, \mathcal{V}(\Pi)) \tag{S.22}$$

in distribution as $n \to \infty$.

Let $\widetilde{\lambda}_n = (\widetilde{\lambda}'_{\tau_1 n}, ..., \widetilde{\lambda}'_{\tau_K n})'$. Note that $b_{\tau_k} \neq 0$ for $1 \leqslant k \leqslant K$ and, from (S.21), we have

$$\sqrt{n}[\widetilde{\lambda}_n - (1_K \otimes I_{2p})\lambda_0] \to N(0, \Sigma_1^{*-1}(\Gamma \otimes \Omega_2)\Sigma_1^{*-1})$$

in distribution as $n \to \infty$, where 1_K is a $K \times 1$ vector of ones, $\otimes$ is the Kronecker product, and $\Sigma_1^* = \text{diag}\{\Sigma_1(\tau_1), ..., \Sigma_1(\tau_K)\}$. 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. Let $\Pi = (\pi_1, ..., \pi_K) = \{(1_K \otimes I_{2p})' \Xi (1_K \otimes I_{2p})\}^{-1}(1_K \otimes I_{2p})'\Xi$ be a $2p \times 2pK$ matrix, and it can be verified that

$$\widehat{\lambda}_n^* = \Pi \widetilde{\lambda}_n = \sum_{k=1}^{K} \pi_k \widetilde{\lambda}_{\tau_k n}.$$

The minimum distance estimator will have a minimum variance when the matrix Ξ is proportional to $[\Sigma_1^{*-1}(\Gamma \otimes \Omega_2)\Sigma_1^{*-1}]^{-1} = \Sigma_1^*(\Gamma^{-1} \otimes \Omega_2^{-1})\Sigma_1^*$, and this corresponds to

$$\Pi^{opt} = [(1_K \otimes I_{2p})'\Sigma_1^*(\Gamma^{-1} \otimes \Omega_2^{-1})\Sigma_1^*(1_K \otimes I_{2p})]^{-1}(1_K \otimes I_{2p})'\Sigma_1^*(\Gamma^{-1} \otimes \Omega_2^{-1})\Sigma_1^*;$$

see also Chen et al. (2016). Hence the proof of this theorem is accomplished.

2.5 Proof of Theorem 5

Note that $\widetilde{\beta}^{int} - \beta_0 = O_p(n^{-1/2})$, $\widetilde{\sigma}_t = 1 + Y'_{a,t-1}\widetilde{\beta}^{int}$ and $w_t = (\widetilde{\sigma}_t + c\sum_{j=1}^{p_{\max}} |y_{t-j}|)^{-1}$. It is then asymptotically equivalent to use the weights $w_t = (\sigma_t + c\sum_{j=1}^{p_{\max}} |y_{t-j}|)^{-1} = (1+\sum_{j=1}^{p_{\max}}(c+\beta_{0j})|y_{t-j}|)^{-1}$, which satisfy Assumption 2 since $c+\beta_{0j} > 0$ for $1 \leqslant j \leqslant p_{\max}$. For simplicity, we here only provide the proof for the consistency of

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

In the following, we use θ^p, $\theta^p_{\tau 0}$, $\widetilde{\theta}^p_{\tau n}$, $\sigma^p_{\tau 0}$ and $\widetilde{\sigma}^p_{\tau n}$ to emphasize their dependence on the order p, and x_t refers to the corresponding regressor with a compatible dimension. Denote $\widetilde{\theta}^p_{\tau n} = \mathrm{argmin}_{\theta^p} \sum_{t=p_{\max}+1}^{n} w_t \rho_\tau(y_t - x'_t\theta^p)$,

$$\sigma^p_{\tau 0} = \min_{\theta^p} E[w_t \rho_\tau(y_t - x'_t\theta^p)] \quad \text{and} \quad \widetilde{\sigma}^p_{\tau n} = \frac{1}{n - p_{\max}} \sum_{t=p_{\max}+1}^{n} w_t \rho_\tau(y_t - x'_t\widetilde{\theta}^p_{\tau n}).$$

We can show that $\sigma^1_{\tau 0} > \cdots > \sigma^{p_0}_{\tau 0} = \cdots = \sigma^{p_{\max}}_{\tau 0}$, and $\widetilde{\sigma}^p_{\tau n} = \sigma^p_{\tau 0} + o_p(1)$ for all $1 \leqslant p \leqslant p_{\max}$.

We first consider the case with $p < p_0$, where we have

$$\begin{aligned}\mathrm{BIC}_\tau(p) - \mathrm{BIC}_\tau(p_0) &= 2(n - p_{\max})(\log \widetilde{\sigma}^p_{\tau n} - \log \widetilde{\sigma}^{p_0}_{\tau n}) + 2(p - p_0) \log(n - p_{\max}) \\ &= 2(n - p_{\max})\{(\log \sigma^p_{\tau 0} - \log \sigma^{p_0}_{\tau 0}) + o_p(1)\} + o(n), \qquad \text{(S.23)}\end{aligned}$$

which tends to $+\infty$ as $n \to \infty$.

We next consider the case with $p > p_0$. From the proof of Lemma 1, we have

$$\sum_{t=p_{\max}+1}^{n} \Big[w_t \rho_\tau(y_t - x'_t\widetilde{\theta}^p_{\tau n}) - w_t \rho_\tau(y_t - x'_t\theta^p_{\tau 0})\Big] = O_p(1),$$

which, together with the fact that $\sum_{t=p_{\max}+1}^{n} w_t \rho_\tau(y_t - x'_t\theta^p_{\tau 0})$ is a constant for all $p \geqslant p_0$, implies that

$$\widetilde{\sigma}^p_{\tau n} - \widetilde{\sigma}^{p_0}_{\tau n} = O_p(n^{-1}).$$

As a result, by the Taylor expansion,

$$|\log \widetilde{\sigma}^p_{\tau n} - \log \widetilde{\sigma}^{p_0}_{\tau n}| \leqslant \frac{1}{\widetilde{\sigma}^p_{\tau n}} |\widetilde{\sigma}^p_{\tau n} - \widetilde{\sigma}^{p_0}_{\tau n}| = O_p(n^{-1}),$$

and then

$$\begin{aligned}\mathrm{BIC}_\tau(p) - \mathrm{BIC}_\tau(p_0) &= 2(n - p_{\max})(\log \widetilde{\sigma}^p_{\tau n} - \log \widetilde{\sigma}^{p_0}_{\tau n}) + 2(p - p_0) \log(n - p_{\max}) \\ &= O_p(1) + 2(p - p_0) \log(n - p_{\max}), \qquad \text{(S.24)}\end{aligned}$$

which tends to $+\infty$ as $n \to \infty$. We can accomplish the proof by combining (S.23) and (S.24).

2.6 Proofs of Lemma 2 and Theorem 6

Proof of Lemma 2. Denote $Q_n(b) = \sum_{t=p+1}^{n} \rho_\tau(\widehat{\varepsilon}_t - b)$. For any fixed $v \in \mathbb{R}$, by (S.14), we have

$$
\begin{aligned}
& Q_n(b_\tau + n^{-1/2}v) - Q_n(b_\tau) \\
& \quad = -\frac{v}{\sqrt{n}} \sum_{t=p+1}^{n} \psi_\tau(\widehat{\varepsilon}_t - b_\tau) + \sum_{t=p+1}^{n} \int_0^{n^{-1/2}v} \{I(\widehat{\varepsilon}_t - b_\tau \leqslant s) - I(\widehat{\varepsilon}_t - b_\tau \leqslant 0)\} ds.
\end{aligned}
$$

From Theorem 4 and Knight (1998), together with the convexity of $Q_n(b)$, it is sufficient to show that

$$
\frac{1}{\sqrt{n}} \sum_{t=p+1}^{n} \psi_\tau(\widehat{\varepsilon}_t - b_\tau) = \frac{1}{\sqrt{n}} \sum_{t=p+1}^{n} \psi_\tau(\varepsilon_t - b_\tau) - d_0'(\tau)\sqrt{n}(\widehat{\lambda}_n^{opt} - \lambda_0) + o_p(1), \quad \text{(S.25)}
$$

and

$$
\sum_{t=p+1}^{n} \int_0^{n^{-1/2}v} \{I(\widehat{\varepsilon}_t - b_\tau \leqslant s) - I(\widehat{\varepsilon}_t - b_\tau \leqslant 0)\} ds = \frac{1}{2} f(b_\tau) v^2 + o_p(1), \quad \text{(S.26)}
$$

where $d_0(\tau) = f(b_\tau)(b_\tau E(\sigma_t^{-1} Y_{a,t-1}'), E(\sigma_t^{-1} Y_{t-1}'))'$. For any $u_1, u_2 \in \mathbb{R}^p$, let $u = (u_1', u_2')'$, and denote

$$
\sigma_t(u_1) = 1 + Y_{a,t-1}'(\beta_0 + n^{-1/2}u_1) \quad \text{and} \quad \varepsilon_t(u) = \frac{y_t - Y_{t-1}'(\phi_0 + n^{-1/2}u_2)}{\sigma_t(u_1)},
$$

where despite their dependence on n, it is suppressed in the notations without causing confusion. Since $\sqrt{n}(\widehat{\lambda}_n - \lambda_0) = O_p(1)$, to prove (S.25) and (S.26) it is sufficient to respectively establish that, for any fixed $M > 0$,

$$
\sup_{\|u\| \leqslant M} \left| \frac{1}{\sqrt{n}} \sum_{t=p+1}^{n} [\psi_\tau\{\varepsilon_t(u) - b_\tau\} - \psi_\tau(\varepsilon_t - b_\tau)] + d_0'(\tau) u \right| = o_p(1), \quad \text{(S.27)}
$$

and

$$
\sup_{\|u\| \leqslant M} \left| \sum_{t=p+1}^{n} \int_0^{n^{-1/2}v} [I\{\varepsilon_t(u) - b_\tau \leqslant s\} - I\{\varepsilon_t(u) - b_\tau \leqslant 0\}] ds - \frac{1}{2} f(b_\tau) v^2 \right| = o_p(1). \quad \text{(S.28)}
$$

We first prove (S.27). Let $b_t(u) = b_\tau \sigma_t^{-1} \sigma_t(u_1) + n^{-1/2} \sigma_t^{-1} Y_{t-1}' u_2$. We have that $b_t(0) = b_\tau$ and $b_t(u) \in \mathcal{F}_{t-1}$ for all t. Moreover, it is implied by (S.17) that, for any $u, u^* \in \mathbb{R}^{2p}$,

$$
|b_t(u^*) - b_t(u)| = n^{-1/2} \left| b_\tau \frac{Y_{a,t-1}'(u_1^* - u_1)}{\sigma_t} + \frac{Y_{t-1}'(u_2^* - u_2)}{\sigma_t} \right| \leqslant n^{-1/2} C \|u^* - u\|, \quad \text{(S.29)}
$$

where C is a constant independent of t.

Denote

$$\zeta_t(u) = \psi_\tau\{\varepsilon_t(u) - b_\tau\} - \psi_\tau(\varepsilon_t - b_\tau) \quad \text{and} \quad S_n(u) = \frac{1}{\sqrt{n}} \sum_{t=p+1}^{n} [\zeta_t(u) - E\{\zeta_t(u) \mid \mathcal{F}_{t-1}\}] .$$

It holds that

$$\zeta_t(u) = I(\varepsilon_t < b_\tau) - I\{\varepsilon_t < b_t(u)\} = I\{b_t(u) < \varepsilon_t < b_\tau\} - I\{b_t(u) > \varepsilon_t > b_\tau\},$$

and then, by the Taylor expansion, (S.29) and Assumption 3,

$$E\{\zeta_t^2(u)\} = E\{|F\{b_t(u)\} - F(b_\tau)|\} \leqslant \sup_x f(x) E\{|b_t(u) - b_\tau|\} \leqslant n^{-1/2} C \|u\|,$$

which implies that

$$E\{S_n^2(u)\} \leqslant \frac{1}{n} \sum_{t=p+1}^{n} E\{\zeta_t^2(u)\} = o(1). \tag{S.30}$$

Similarly, for any $u, u^* \in \mathbb{R}^{2p}$ and any $\delta > 0$,

$$\begin{aligned}
&\sup_{\|u^*-u\| \leqslant \delta} |\zeta_t(u^*) - \zeta_t(u)| \\
&\qquad = \sup_{\|u^*-u\| \leqslant \delta} |I\{b_t(u^*) < \varepsilon_t < b_t(u)\} - I\{b_t(u^*) > \varepsilon_t > b_t(u)\}| \\
&\qquad \leqslant I\left\{ |\varepsilon_t - b_t(u)| \leqslant \sup_{\|u^*-u\| \leqslant \delta} |b_t(u^*) - b_t(u)| \right\}.
\end{aligned}$$

This, together with (S.29) and Assumption 3, leads to

$$E \sup_{\|u^*-u\| \leqslant \delta} |\zeta_t(u^*) - \zeta_t(u)| \leqslant \mathrm{pr}\{|\varepsilon_t - b_t(u)| \leqslant n^{-1/2} \delta C\} \leqslant 2C \sup_x f(x) \cdot n^{-1/2} \delta,$$

which implies that

$$E \sup_{\|u^*-u\| \leqslant \delta} |S_n(u^*) - S_n(u)| \leqslant \frac{2}{\sqrt{n}} \sum_{t=p+1}^{n} E \sup_{\|u^*-u\| \leqslant \delta} |\zeta_t(u^*) - \zeta_t(u)| \leqslant \delta C. \tag{S.31}$$

Therefore, it follows from (S.30), (S.31) and the finite covering theorem that

$$\sup_{\|u\| \leqslant M} |S_n(u)| = o_p(1). \tag{S.32}$$

Observe that $E\{\zeta_t(u) \mid \mathcal{F}_{t-1}\} = F(b_\tau) - F\{b_t(u)\}$ and, by (S.29) and the law of large numbers, $n^{-1/2} \sum_{t=p+1}^{n} \{b_\tau - b_t(u)\} f(b_\tau) = -d_0'(\tau) u + o_p(1)$. As a result, by the Taylor expansion, (S.29) and Assumption 3,

$$\begin{aligned}
&\sup_{\|u\|\leqslant M}\left|\frac{1}{\sqrt{n}}\sum_{t=p+1}^{n}E\{\zeta_t(u)\mid\mathcal{F}_{t-1}\}+d_0'(\tau)u\right| \\
&\qquad=\sup_{\|u\|\leqslant M}\left|\frac{1}{\sqrt{n}}\sum_{t=p+1}^{n}\{f(b_t^*)-f(b_\tau)\}\{b_\tau-b_t(u)\}\right|+o_p(1) \\
&\qquad\leqslant CM\sup_{0\leqslant x\leqslant n^{-1/2}CM}|f(x+b_\tau)-f(b_\tau)|+o_p(1)=o_p(1),
\end{aligned}$$

where b_t^* is between $b_t(u)$ and b_τ. This together with (S.32), implies (S.27).

We next prove (S.28). By a method similar to that of (S.16), (S.18) and (S.19), it can be readily shown that

$$\sum_{t=p+1}^{n}\int_0^{n^{-1/2}v}\{I(\varepsilon_t-b_\tau\leqslant s)-I(\varepsilon_t-b_\tau\leqslant 0)\}ds=\frac{1}{2}f(b_\tau)v^2+o_p(1). \tag{S.33}$$

Denote

$$\eta_t(u)=\int_0^{n^{-1/2}v}\left[I\{\varepsilon_t(u)-b_\tau\leqslant s\}-I\{\varepsilon_t(u)-b_\tau\leqslant 0\}-I(\varepsilon_t-b_\tau\leqslant s)+I(\varepsilon_t-b_\tau\leqslant 0)\right]ds.$$

For any $u\in\mathbb{R}^{2p}$ and $s\in\mathbb{R}$, let $b_t(u,s)=\sigma_t^{-1}\sigma_t(u_1)s+b_t(u)$ and, by (S.17), we have that

$$\begin{aligned}
|b_t(u^*,s)-b_t(u,s)|&=n^{-1/2}\left|(b_\tau+s)\frac{Y_{a,t-1}'(u_1^*-u_1)}{\sigma_t}+\frac{Y_{t-1}'(u_2^*-u_2)}{\sigma_t}\right| \\
&\leqslant n^{-1/2}\|u^*-u\|C(1+|s|)
\end{aligned} \tag{S.34}$$

uniformly for all $1\leqslant t\leqslant n$. It can be further verified that

$$\eta_t(u)=\int_0^{n^{-1/2}v}[I\{\varepsilon_t\leqslant b_t(u,s)\}-I(\varepsilon_t\leqslant s+b_\tau)]-[I\{\varepsilon_t\leqslant b_t(u)\}-I(\varepsilon_t\leqslant b_\tau)]ds,$$

and for any $\delta>0$,

$$\begin{aligned}
&\sup_{\|u^*-u\|\leqslant\delta}|\eta_t(u^*)-\eta_t(u)| \\
&\qquad\leqslant\int_0^{n^{-1/2}|v|}I\left\{|\varepsilon_t-b_t(u,s)|\leqslant\sup_{\|u^*-u\|\leqslant\delta}|b_t(u^*,s)-b_t(u,s)|\right\}ds \\
&\qquad\quad+\int_0^{n^{-1/2}|v|}I\left\{|\varepsilon_t-b_t(u)|\leqslant\sup_{\|u^*-u\|\leqslant\delta}|b_t(u^*)-b_t(u)|\right\}ds.
\end{aligned}$$

By a method similar to the proof of (S.32), together with (S.29) and (S.34), we can show that

$$\sup_{\|u\|\leqslant M}\left|\sum_{t=p+1}^{n}[\eta_t(u)-E\{\eta_t(u)\mid\mathcal{F}_{t-1}\}]\right|=o_p(1). \tag{S.35}$$

Note that $b_t(0,s) = s + b_\tau$ and, by the Taylor expansion, we have

$$\begin{aligned}
&\sup_{\|u\|\leqslant M}\left|\sum_{t=p+1}^{n} E\{\eta_t(u) \mid \mathcal{F}_{t-1}\}\right| \\
&\quad = \sup_{\|u\|\leqslant M}\left|\sum_{t=p+1}^{n}\int_0^{n^{-1/2}v}[f(b_t^{**})\{b_t(u,s)-(s+b_\tau)\} - f(b_t^*)\{b_t(u)-b_\tau\}]ds\right| \\
&\quad \leqslant \sup_{\|u\|\leqslant M}\left|\sum_{t=p+1}^{n}\int_0^{n^{-1/2}v} f(b_t^*)[\{b_t(u,s)-(s+b_\tau)\} - \{b_t(u)-b_\tau\}]ds\right| \\
&\qquad + \sup_{\|u\|\leqslant M}\left|\sum_{t=p+1}^{n}\int_0^{n^{-1/2}v}\{f(b_t^{**}) - f(b_t^*)\}\{b_t(u,s)-(s+b_\tau)\}ds\right|, \qquad \text{(S.36)}
\end{aligned}$$

where b_t^* is between $b_t(u)$ and b_τ, and b_t^{**} is between $b_t(u,s)$ and $s+b_\tau$. Since it is implied by (S.17) that

$$|\{b_t(u,s)-(s+b_\tau)\} - \{b_t(u)-b_\tau\}| = n^{-1/2}|s|\frac{Y'_{a,t-1}u_1}{\sigma_t} \leqslant n^{-1/2}|s|\|u\|C$$

uniformly for all $1 \leqslant t \leqslant n$, by Assumption 3, we then have that

$$\begin{aligned}
&\sup_{\|u\|\leqslant M}\left|\sum_{t=p+1}^{n}\int_0^{n^{-1/2}v} f(b_t^*)[\{b_t(u,s)-(s+b_\tau)\} - \{b_t(u)-b_\tau\}]ds\right| \\
&\quad \leqslant \sqrt{n}CM\sup_x f(x)\int_0^{n^{-1/2}|v|}|s|ds = o_p(1). \qquad \text{(S.37)}
\end{aligned}$$

Moreover, by (S.29) and (S.34), we have

$$\sup_{\|u\|\leqslant M}\ \sup_{0\leqslant s\leqslant n^{-1/2}|v|}|b_t^{**} - b_t^*| \leqslant n^{-1/2}CM$$

uniformly for all $1 \leqslant t \leqslant n$. Then, it follows from Assumption 3 and (S.34) again that

$$\begin{aligned}
&\sup_{\|u\|\leqslant M}\left|\sum_{t=p+1}^{n}\int_0^{n^{-1/2}v}\{f(b_t^{**}) - f(b_t^*)\}\{b_t(u,s)-(s+b_\tau)\}ds\right| \\
&\quad \leqslant \sqrt{n}CM\int_0^{n^{-1/2}|v|}(1+|s|)ds\sup_{\|u\|\leqslant M}\ \sup_{0\leqslant s\leqslant n^{-1/2}|v|}|f(b_t^{**}) - f(b_t^*)| \\
&\quad = o_p(1). \qquad \text{(S.38)}
\end{aligned}$$

Hence, (S.28) follows from (S.33) and (S.35)-(S.38). This completes the derivation of the Bahadur representation of $\widehat{b}_{\tau n}$, and hence the proof of the lemma. □

Proof of Theorem 6. We first show that $\widehat{\mu}_{G,m} = \mu_{G,m} + o_p(1)$, $\widehat{\sigma}^2_{G,m} = \sigma^2_{G,m} + o_p(1)$ for

$m = 1$ and 2. By the Taylor expansion, (S.17) and Assumption 4(i) and (ii), we have

$$\begin{aligned}|G(\widehat{\varepsilon}_t) - G(\varepsilon_t)| &= |G\{\varepsilon_t(\widehat{\lambda}_n)\} - G\{\varepsilon_t(\lambda_0)\}| \\ &= \left| g\{\varepsilon_t(\lambda^*)\}\varepsilon_t(\lambda^*)\frac{Y'_{a,t-1}(\widehat{\beta}_n - \beta_0)}{\sigma_t(\lambda^*)} + g\{\varepsilon_t(\lambda^*)\}\frac{Y'_{t-1}(\widehat{\phi}_n - \phi_0)}{\sigma_t(\lambda^*)}\right| \\ &\leqslant C\|\widehat{\lambda}_n - \lambda_0\|\end{aligned}$$

uniformly for all $1 \leqslant t \leqslant n$, where λ^* is between $\widehat{\lambda}_n$ and λ_0. As a result, by the law of large numbers, the boundedness of $G(\cdot)$ and the fact that $\widehat{\lambda}_n - \lambda_0 = O_p(n^{-1/2})$,

$$\widehat{\mu}_{G,1} = \frac{1}{n}\sum_{t=p+1}^{n} G(\widehat{\varepsilon}_t) = \frac{1}{n}\sum_{t=p+1}^{n} G(\varepsilon_t) + \frac{1}{n}\sum_{t=p+1}^{n}\{G(\widehat{\varepsilon}_t) - G(\varepsilon_t)\} = \mu_{G,1} + o_p(1),$$

and

$$\begin{aligned}\widehat{\sigma}^2_{G,1} &= \frac{1}{n}\sum_{t=p+1}^{n}\{G(\widehat{\varepsilon}_t) - \widehat{\mu}_{G,1}\}^2 = \frac{1}{n}\sum_{t=p+1}^{n}\{G(\widehat{\varepsilon}_t)\}^2 - \mu^2_{G,1} + o_p(1) \\ &= \frac{1}{n}\sum_{t=p+1}^{n}\{G(\varepsilon_t)\}^2 + \frac{1}{n}\sum_{t=p+1}^{n}\{G(\widehat{\varepsilon}_t) - G(\varepsilon_t)\}\{G(\widehat{\varepsilon}_t) + G(\varepsilon_t)\} - \mu^2_{G,1} + o_p(1) \\ &= \sigma^2_{G,1} + o_p(1).\end{aligned}$$

Similarly, we can show that $\widehat{\mu}_{G,2} = \mu_{G,2} + o_p(1)$ and $\widehat{\sigma}^2_{G,2} = \sigma^2_{G,2} + o_p(1)$.

Let $\varepsilon^*_t = \varepsilon_t - b_\tau$ and $\widehat{\varepsilon}^*_t = \widehat{\varepsilon}_t - \widehat{b}_{\tau n}$ for simplicity. Since $|\sum_{t=p+1}^{n}\psi_\tau(\widehat{\varepsilon}^*_t)| < 1$, by an elementary calculation, we have

$$\begin{aligned}&\frac{1}{\sqrt{n}}\sum_{t=p+\ell+1}^{n}\psi_\tau(\widehat{\varepsilon}^*_t)\{G(\widehat{\varepsilon}_{t-\ell}) - \widehat{\mu}_{G,1}\} \\ =&\frac{1}{\sqrt{n}}\sum_{t=p+\ell+1}^{n}\psi_\tau(\widehat{\varepsilon}^*_t)G(\widehat{\varepsilon}_{t-\ell}) + O_p(n^{-1/2}) \\ =&\frac{1}{\sqrt{n}}\sum_{t=p+\ell+1}^{n}\psi_\tau(\varepsilon^*_t)G(\varepsilon_{t-\ell}) + A_{n1} + A_{n2} + A_{n3} + O_p(n^{-1/2}),\end{aligned} \tag{S.39}$$

where

$$\begin{aligned}A_{n1} &= \frac{1}{\sqrt{n}}\sum_{t=p+\ell+1}^{n}\{\psi_\tau(\widehat{\varepsilon}^*_t) - \psi_\tau(\varepsilon^*_t)\}G(\varepsilon_{t-\ell}), \\ A_{n2} &= \frac{1}{\sqrt{n}}\sum_{t=p+\ell+1}^{n}\psi_\tau(\varepsilon^*_t)\{G(\widehat{\varepsilon}_{t-\ell}) - G(\varepsilon_{t-\ell})\}, \\ A_{n3} &= \frac{1}{\sqrt{n}}\sum_{t=p+\ell+1}^{n}\{\psi_\tau(\widehat{\varepsilon}^*_t) - \psi_\tau(\varepsilon^*_t)\}\{G(\widehat{\varepsilon}_{t-\ell}) - G(\varepsilon_{t-\ell})\}.\end{aligned}$$

Let $d_\ell(\tau) = f(b_\tau)(b_\tau E\{G(\varepsilon_{t-\ell})Y'_{a,t-1}/\sigma_t\}, E\{G(\varepsilon_{t-\ell})Y'_{t-1}/\sigma_t\})'$ for $\ell \geqslant 1$. By a method similar to the proof of (S.25), we can show that

$$A_{n1} = -f(b_\tau)\mu_{G,1}\sqrt{n}(\widehat{b}_{\tau n} - b_\tau) - d'_\ell(\tau)\sqrt{n}(\widehat{\lambda}_n - \lambda_0) + o_p(1). \tag{S.40}$$

Denote $\sigma_t(\lambda) = 1 + Y'_{a,t-1}\beta$ and $\varepsilon_t(\lambda) = (y_t - Y'_{t-1}\phi)/\sigma_t(\lambda)$. Note that

$$\frac{\partial \varepsilon_t(\lambda)}{\partial \lambda} = \begin{pmatrix} -\varepsilon_t(\lambda)\dfrac{Y_{a,t-1}}{\sigma_t(\lambda)} \\ -\dfrac{Y_{t-1}}{\sigma_t(\lambda)} \end{pmatrix} \quad \text{and} \quad \frac{\partial^2 \varepsilon_t(\lambda)}{\partial \lambda \partial \lambda'} = \begin{pmatrix} 2\varepsilon_t(\lambda)\dfrac{Y_{a,t-1}Y'_{a,t-1}}{\sigma_t^2(\lambda)} & \dfrac{Y_{a,t-1}Y'_{t-1}}{\sigma_t^2(\lambda)} \\ \dfrac{Y_{t-1}Y'_{a,t-1}}{\sigma_t^2(\lambda)} & 0 \end{pmatrix}.$$

From (S.17) and Assumption 4, we can verify that

$$E\left(\left|\sup_{\lambda\in\Lambda} \frac{\partial^2 G\{\varepsilon_t(\lambda)\}}{\partial\lambda\partial\lambda'}\right|\right) = E\left(\left|\sup_{\lambda\in\Lambda}\left[g\{\varepsilon_t(\lambda)\}\frac{\partial^2\varepsilon_t(\lambda)}{\partial\lambda\partial\lambda'} + \dot{g}\{\varepsilon_t(\lambda)\}\frac{\partial\varepsilon_t(\lambda)}{\partial\lambda}\frac{\partial\varepsilon_t(\lambda)}{\partial\lambda'}\right]\right|\right) < \infty,$$

which, together with the Taylor expansion and the fact that $\sqrt{n}(\widehat{\lambda}_n - \lambda_0) = O_p(1)$, implies

$$A_{n2} = o_p(1). \tag{S.41}$$

Finally we consider A_{n3}. For any $v \in \mathbb{R}$ and $u \in \mathbb{R}^{2p}$, let $v = (v, u')'$, and denote $\varepsilon_t^*(v) = \varepsilon_t(u) - (b_\tau + n^{-1/2}v)$, where $\varepsilon_t(u)$ is defined as in the proof of Lemma 2. Let $\varsigma_t(v) = [\psi_\tau\{\varepsilon_t^*(v)\} - \psi_\tau(\varepsilon_t^*)][G\{\varepsilon_{t-\ell}(u)\} - G(\varepsilon_{t-\ell})]$. By a method similar to the proof of (S.27), we can show that, for any $M > 0$,

$$\sup_{\|v\|\leqslant M}\left|\frac{1}{\sqrt{n}}\sum_{t=p+\ell+1}^{n}[\varsigma_t(v) - E\{\varsigma_t(v) \mid \mathcal{F}_{t-1}\}]\right| = o_p(1)$$

and

$$\sup_{\|v\|\leqslant M}\left|\frac{1}{\sqrt{n}}\sum_{t=p+\ell+1}^{n}E\{\varsigma_t(v) \mid \mathcal{F}_{t-1}\}\right| = o_p(1).$$

As a result,

$$\sup_{\|v\|\leqslant M}\left|\frac{1}{\sqrt{n}}\sum_{t=p+\ell+1}^{n}[\psi_\tau\{\varepsilon_t^*(v)\} - \psi_\tau(\varepsilon_t^*)][G\{\varepsilon_{t-\ell}(u)\} - G(\varepsilon_{t-\ell})]\right| = o_p(1),$$

which, together with $\sqrt{n}(\widehat{b}_{\tau n} - b_\tau) = O_p(1)$ and $\sqrt{n}(\widehat{\lambda}_n - \lambda_0) = O_p(1)$, implies that

$$A_{n3} = o_p(1). \tag{S.42}$$

By (S.39)-(S.42), (S.22) and Lemma 2, we have

$$\widehat{\rho}_{\ell,\tau} = \frac{1}{\sqrt{\tau-\tau^2}\sigma_{G,1}}\frac{1}{n}\sum_{t=p+\ell+1}^{n}\left[\psi_\tau(\varepsilon_t^*)\{G(\varepsilon_{t-\ell}) - \mu_{G,1}\} - \widetilde{d}'_{\ell,1}(\tau)H(\varepsilon_t)\Omega_1\frac{x_t}{\sigma_t}\right] + o_p(n^{-1/2}),$$

where $\widetilde{d}_{\ell,1}(\tau) = f(b_\tau)\left(b_\tau E[\{G(\varepsilon_{t-\ell}) - \mu_{G,1}\}Y'_{a,t-1}/\sigma_t], E[\{G(\varepsilon_{t-\ell}) - \mu_{G,1}\}Y'_{t-1}/\sigma_t]\right)'$ and $H(\varepsilon_t) = \sum_{k=1}^K \psi_{\tau_k}(\varepsilon_t - b_{\tau_k})\pi_k^{opt}\Sigma_2(\tau_k)$. Let $\widehat{\rho}_\tau = (\widehat{\rho}_{1,\tau}, \ldots, \widehat{\rho}_{L,\tau})'$. Then we have

$$\widehat{\rho}_\tau = \frac{1}{\sqrt{\tau - \tau^2}\sigma_{G,1}} \frac{1}{n} \sum_{t=p+L+1}^{n} \left[\psi_\tau(\varepsilon_t^*)(G_1 - \mu_{G,1}1_L) - D_1'(\tau)H(\varepsilon_t)\Omega_1 \frac{x_t}{\sigma_t}\right] + o_p(n^{-1/2}),$$

where $G_1 = (G(\varepsilon_{t-1}), \ldots, G(\varepsilon_{t-L}))'$ and $D_1(\tau) = (\widetilde{d}_{1,1}(\tau), \ldots, \widetilde{d}_{L,1}(\tau))$.

Let $\widehat{r}_\tau = (\widehat{r}_{1,\tau}, \ldots, \widehat{r}_{L,\tau})'$, and then by a method similar to the proof for $\widehat{\rho}_\tau$ above,

$$\widehat{r}_\tau = \frac{1}{\sqrt{\tau - \tau^2}\sigma_{G,2}} \frac{1}{n} \sum_{t=p+L+1}^{n} \left[\psi_\tau(\varepsilon_t^*)(G_2 - \mu_{G,2}1_L) - D_2'(\tau)H(\varepsilon_t)\Omega_1 \frac{x_t}{\sigma_t}\right] + o_p(n^{-1/2}),$$

where $G_2 = (G(\varepsilon_{t-1}^2), \ldots, G(\varepsilon_{t-L}^2))'$ and $D_2(\tau) = (\widetilde{d}_{1,2}(\tau), \ldots, \widetilde{d}_{L,2}(\tau))$ with $\widetilde{d}_{\ell,2}(\tau) = f(b_\tau)\left(b_\tau E[\{G(\varepsilon_{t-\ell}^2) - \mu_{G,2}\}Y'_{a,t-1}/\sigma_t], E[\{G(\varepsilon_{t-\ell}^2) - \mu_{G,2}\}Y'_{t-1}/\sigma_t]\right)'$. Therefore, we complete the proof by the central limit theorem and the Cramér-Wold device. $\square$

References

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.

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.

Knight, K. (1998). Limiting distributions for L_1 regression estimators under general conditions. *The Annals of Statistics 26*, 755–770.

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.

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

Table 1: Biases and ESDs of $\widetilde{\lambda}_{\tau n}$ with the weights W_1 or W_2 at $\tau = 0.25$ or 0.35, when the innovations follow the normal, Student's t_3 or Cauchy distribution.

		$\tau = 0.25$				$\tau = 0.35$			
		W_1		W_2		W_1		W_2	
	n	Bias	ESD	Bias	ESD	Bias	ESD	Bias	ESD
				Normal distribution					
β	200	0.1093	0.6472	0.1240	0.7757	-0.0133	10.8221	-0.4226	12.5201
	500	0.0247	0.2426	0.0268	0.2487	0.1275	1.3010	0.1446	0.8197
	1000	0.0178	0.1660	0.0191	0.1639	0.0586	0.3211	0.0604	0.3289
ϕ	200	-0.0199	0.1300	-0.0215	0.1307	-0.0176	0.1190	-0.0204	0.1209
	500	-0.0076	0.0834	-0.0078	0.0828	-0.0080	0.0781	-0.0085	0.0775
	1000	-0.0021	0.0582	-0.0018	0.0577	-0.0003	0.0535	-0.0001	0.0530
				Student's t_3 distribution					
β	200	0.1396	0.6785	0.1691	0.9003	0.0788	6.6405	0.0056	6.5120
	500	0.0342	0.2690	0.0347	0.2827	0.1418	0.7987	0.1415	0.7684
	1000	0.0187	0.1691	0.0198	0.1679	0.0613	0.3157	0.0628	0.3156
ϕ	200	-0.0108	0.1258	-0.0104	0.1271	-0.0095	0.1077	-0.0115	0.1066
	500	-0.0063	0.0768	-0.0062	0.0758	-0.0081	0.0660	-0.0084	0.0659
	1000	-0.0049	0.0561	-0.0046	0.0551	-0.0036	0.0471	-0.0036	0.0465
				Cauchy distribution					
β	200	0.1413	0.8636	0.2298	1.6190	0.1010	2.6814	0.2891	2.9141
	500	0.0551	0.3855	0.0698	0.3433	0.0874	0.5839	0.1000	0.5827
	1000	0.0220	0.2264	0.0264	0.2243	0.0213	0.2854	0.0267	0.2836
ϕ	200	-0.0078	0.0799	-0.0095	0.0830	-0.0068	0.0556	-0.0071	0.0569
	500	-0.0036	0.0460	-0.0034	0.0456	-0.0037	0.0312	-0.0039	0.0314
	1000	-0.0022	0.0308	-0.0024	0.0306	-0.0009	0.0215	-0.0011	0.0213

Table 2: Biases, ESDs and ASDs of the doubly weighted estimator $\widehat{\lambda}_n^{opt}$ with the number of quantile levels $K = 9$ and the weights W_1 or W_2, when the innovations follow the normal, Student's t_3 or Cauchy distribution.

		W_1			W_2		
	n	Bias	ESD	ASD	Bias	ESD	ASD
				Normal distribution			
β	200	-0.0148	0.1703	0.1352	-0.0203	0.1658	0.1320
	500	-0.0035	0.0955	0.0879	-0.0063	0.0938	0.0863
	1000	0.0003	0.0632	0.0628	-0.0008	0.0618	0.0619
ϕ	200	-0.0106	0.1091	0.0896	-0.0106	0.1082	0.0889
	500	-0.0055	0.0631	0.0596	-0.0057	0.0630	0.0592
	1000	-0.0020	0.0449	0.0429	-0.0021	0.0448	0.0426
				Student's t_3 distribution			
β	200	0.0230	0.2176	0.1596	0.0204	0.2180	0.1569
	500	0.0134	0.1200	0.1025	0.0123	0.1180	0.1010
	1000	0.0082	0.0790	0.0728	0.0066	0.0780	0.0716
ϕ	200	-0.0076	0.1122	0.0863	-0.0082	0.1115	0.0856
	500	-0.0032	0.0616	0.0567	-0.0037	0.0607	0.0563
	1000	-0.0027	0.0423	0.0406	-0.0030	0.0421	0.0403
				Cauchy distribution			
β	200	0.1666	0.4452	0.2621	0.1674	0.4794	0.2619
	500	0.0777	0.2279	0.1564	0.0803	0.2111	0.1550
	1000	0.0326	0.1236	0.1085	0.0337	0.1224	0.1074
ϕ	200	-0.0072	0.0585	0.0435	-0.0081	0.0575	0.0430
	500	-0.0022	0.0272	0.0258	-0.0022	0.0272	0.0255
	1000	-0.0005	0.0175	0.0172	-0.0005	0.0173	0.0170

Table 3: Empirical coverage rates of $\widehat{\rho}_\ell$ and $\widehat{r}_\ell$ at lags $\ell = 2, 4, 6$ at the 5% significance level, when the transformation G is the normal (G_N), Student's t_3 (G_T) or Cauchy (G_C) distribution function, and the innovations follow the normal, Student's t_3 or Cauchy distribution.

		G_N		G_T		G_C	
n	lag	$\widehat{\rho}_\ell$	$\widehat{r}_\ell$	$\widehat{\rho}_\ell$	$\widehat{r}_\ell$	$\widehat{\rho}_\ell$	$\widehat{r}_\ell$
				Normal distribution			
200	2	0.949	0.945	0.949	0.948	0.948	0.945
	4	0.958	0.953	0.955	0.952	0.953	0.955
	6	0.960	0.940	0.957	0.942	0.959	0.947
500	2	0.941	0.941	0.942	0.941	0.942	0.945
	4	0.959	0.944	0.960	0.944	0.963	0.940
	6	0.951	0.957	0.949	0.959	0.951	0.960
1000	2	0.946	0.951	0.948	0.951	0.947	0.952
	4	0.948	0.953	0.950	0.949	0.947	0.958
	6	0.957	0.951	0.955	0.951	0.953	0.946
				t_3 distribution			
200	2	0.957	0.942	0.957	0.945	0.963	0.943
	4	0.952	0.953	0.956	0.953	0.953	0.955
	6	0.939	0.953	0.940	0.949	0.941	0.957
500	2	0.959	0.952	0.955	0.948	0.958	0.948
	4	0.957	0.955	0.955	0.962	0.952	0.956
	6	0.948	0.946	0.953	0.947	0.952	0.947
1000	2	0.963	0.946	0.963	0.952	0.960	0.954
	4	0.947	0.946	0.947	0.944	0.943	0.948
	6	0.948	0.944	0.948	0.948	0.946	0.947
				Cauchy distribution			
200	2	0.960	0.951	0.960	0.951	0.961	0.957
	4	0.952	0.949	0.950	0.953	0.947	0.949
	6	0.958	0.951	0.956	0.949	0.950	0.947
500	2	0.950	0.960	0.956	0.959	0.958	0.955
	4	0.958	0.956	0.955	0.949	0.958	0.962
	6	0.944	0.955	0.941	0.953	0.945	0.955
1000	2	0.942	0.941	0.945	0.942	0.948	0.947
	4	0.940	0.951	0.942	0.954	0.944	0.957
	6	0.948	0.953	0.949	0.954	0.951	0.954

Table 4: Rejection rates of the test $Q_1^{BP}(6)$ at the 5% significance level, when the transformation G is the normal (G_N), Student's t_3 (G_T) or Cauchy (G_C) distribution function and the innovations follow the normal, Student's t_3 or Cauchy distribution.

			Normal distribution			t_3 distribution			Cauchy distribution		
n	c_1	c_2	G_N	G_T	G_C	G_N	G_T	G_C	G_N	G_T	G_C
200	0.0	0.0	0.039	0.041	0.041	0.042	0.041	0.042	0.046	0.046	0.047
	0.0	0.1	0.042	0.040	0.042	0.049	0.048	0.049	0.056	0.057	0.055
	0.0	0.2	0.045	0.048	0.047	0.054	0.049	0.050	0.053	0.060	0.065
	0.0	0.3	0.051	0.055	0.054	0.054	0.052	0.054	0.080	0.080	0.084
	0.1	0.0	0.079	0.080	0.076	0.107	0.111	0.110	0.523	0.534	0.551
	0.2	0.0	0.264	0.267	0.270	0.415	0.425	0.426	0.946	0.950	0.953
	0.3	0.0	0.637	0.642	0.639	0.797	0.813	0.822	0.992	0.992	0.993
500	0.0	0.0	0.046	0.047	0.046	0.047	0.044	0.044	0.048	0.048	0.053
	0.0	0.1	0.033	0.032	0.035	0.048	0.050	0.050	0.053	0.054	0.049
	0.0	0.2	0.052	0.053	0.055	0.049	0.046	0.047	0.056	0.055	0.061
	0.0	0.3	0.049	0.051	0.048	0.048	0.050	0.050	0.072	0.076	0.081
	0.1	0.0	0.179	0.178	0.178	0.303	0.302	0.303	0.956	0.965	0.972
	0.2	0.0	0.714	0.721	0.724	0.879	0.894	0.896	1.000	1.000	1.000
	0.3	0.0	0.991	0.991	0.991	0.998	0.998	0.998	1.000	1.000	1.000
1000	0.0	0.0	0.051	0.051	0.052	0.048	0.047	0.050	0.049	0.052	0.051
	0.0	0.1	0.048	0.050	0.051	0.051	0.046	0.044	0.046	0.043	0.044
	0.0	0.2	0.064	0.062	0.062	0.049	0.046	0.048	0.060	0.061	0.065
	0.0	0.3	0.065	0.065	0.064	0.060	0.069	0.066	0.067	0.070	0.070
	0.1	0.0	0.380	0.380	0.386	0.567	0.576	0.586	1.000	1.000	1.000
	0.2	0.0	0.979	0.979	0.980	0.999	0.999	1.000	1.000	1.000	1.000
	0.3	0.0	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table 5: Rejection rates of the test $Q_2^{BP}(6)$ at the 5% significance level, when the transformation G is the normal (G_N), Student's t_3 (G_T) or Cauchy (G_C) distribution function and the innovations follow the normal, Student's t_3 or Cauchy distribution.

			Normal distribution			t_3 distribution			Cauchy distribution		
n	c_1	c_2	G_N	G_T	G_C	G_N	G_T	G_C	G_N	G_T	G_C
200	0.0	0.0	0.043	0.046	0.044	0.044	0.045	0.048	0.059	0.057	0.056
	0.0	0.1	0.063	0.072	0.073	0.052	0.057	0.061	0.069	0.070	0.085
	0.0	0.2	0.102	0.116	0.123	0.110	0.111	0.124	0.100	0.109	0.122
	0.0	0.3	0.179	0.202	0.252	0.178	0.197	0.228	0.185	0.194	0.213
	0.1	0.0	0.043	0.039	0.044	0.046	0.044	0.039	0.171	0.186	0.210
	0.2	0.0	0.045	0.051	0.052	0.046	0.047	0.052	0.430	0.450	0.485
	0.3	0.0	0.055	0.057	0.059	0.092	0.094	0.110	0.739	0.761	0.796
500	0.0	0.0	0.050	0.055	0.056	0.040	0.044	0.050	0.047	0.051	0.052
	0.0	0.1	0.089	0.097	0.107	0.102	0.109	0.117	0.104	0.109	0.123
	0.0	0.2	0.261	0.292	0.313	0.236	0.274	0.308	0.220	0.242	0.283
	0.0	0.3	0.660	0.700	0.763	0.506	0.556	0.628	0.397	0.408	0.468
	0.1	0.0	0.041	0.038	0.040	0.055	0.055	0.055	0.366	0.381	0.433
	0.2	0.0	0.059	0.062	0.062	0.067	0.072	0.088	0.835	0.868	0.905
	0.3	0.0	0.066	0.068	0.075	0.140	0.159	0.191	0.987	0.992	0.998
1000	0.0	0.0	0.050	0.049	0.049	0.054	0.053	0.051	0.050	0.048	0.047
	0.0	0.1	0.163	0.176	0.194	0.158	0.165	0.181	0.156	0.164	0.191
	0.0	0.2	0.600	0.646	0.699	0.530	0.579	0.641	0.444	0.463	0.507
	0.0	0.3	0.974	0.987	0.997	0.910	0.940	0.961	0.688	0.712	0.765
	0.1	0.0	0.054	0.057	0.061	0.068	0.063	0.064	0.641	0.677	0.735
	0.2	0.0	0.080	0.079	0.082	0.100	0.103	0.108	0.991	0.995	1.000
	0.3	0.0	0.126	0.139	0.146	0.261	0.282	0.339	1.000	1.000	1.000