Network Quantile Autoregression– Supplementary Material

There are 3 Appendixes in this supplementary material. The technical proofs related to model stationarity and estimation are presented in Appendix A and B. The extensive numerical studies are given in Appendix C.

APPENDIX A

In Appendix A, we are going to prove the stationarity results (Theorem 1) and Theorem 2 in Appendix A.1. The verification of (2.7) and (2.8) are given in Appendix A.2. The proofs of Theorem 2 and 3 are given in Appendix A.3 and A.4 respectively.

We first list necessary notations as follows. Let $|M|_a = (|m_{ij}|) \in \mathbb{R}^{m\times n}$ for any arbitrary matrix $M \in \mathbb{R}^{m\times n}$. For any two arbitrary matrices $M_1 = (m_{1,ij}) \in \mathbb{R}^{m\times n}$ and $M_2 = (m_{2,ij}) \in \mathbb{R}^{m\times n}$, define $M_1 \preceq M_2$ as $m_{1,ij} \leq m_{2,ij}$ for $1 \leq i \leq m$ and $1 \leq j \leq n$. In addition, for a p-dimensional vector $Y = (Y_1, \cdots, Y_p)^\top \in \mathbb{R}^p$, denote $|Y|_a^\delta = (|Y_1|^\delta, \cdots, |Y_p|^\delta)^\top \in \mathbb{R}^p$ for any finite real value $\delta > 0$.

Appendix A.1: Proof of Theorem 1

By iteration, we obtain the NQAR solution of model (2.1) as

$$\mathbb{Y}_t = \sum_{l=0}^{L-1} \Pi_l \Gamma + \Pi_L \mathbb{Y}_{t-L} + \sum_{l}^{L-1} \Pi_l V_{t-l} = \sum_{l=0}^{\infty} \Pi_l \Gamma + \sum_{l=0}^{\infty} \Pi_l V_{t-l}, \tag{A.1}$$

where $\Pi_l \stackrel{\text{def}}{=} G_t G_{t-1} \cdots G_{t-l+1}$ for $l > 0$ and $\Pi_0 = 1$. We now prove Theorem 1 in two steps. In the first step, we prove the covariance stationarity of the solution (A.1). Next, we prove the uniqueness of the stationary solution (A.1).

Step 1. (Proof of Covariance Stationarity) In this step, we show the covariance stationarity by calculating $\mathsf{E}(\mathbb{Y}_t)$ and $\mathsf{Cov}(\mathbb{Y}_t, \mathbb{Y}_{t-h})$ respectively. Denote $\lambda_i(M)$ as the ith eigenvalue of any arbitrary matrix $M \in \mathbb{R}^{N\times N}$ such that $|\lambda_1(M)| > |\lambda_2(M)| > \cdots > |\lambda_N(M)|$. Recall that $\mathsf{E}(G_t) = G = b_1 W + b_2 I$. We firstly verify that $\mathsf{E}(\mathbb{Y}_t) = \mu_Y$ for $1 \le t \le T$. To this end, note that $|\lambda_1(W)| = 1$ by Banerjee et al. (2014), we have

$$|\lambda_1(G)| \le |b_1||\lambda_1(W)| + |b_2| < 1 \tag{A.2}$$

due to the stationarity condition in Theorem 1 (note that $|b_1|$ is bounded by $\widetilde{b}_1$). Then it could be computed that $\mathsf{E}(\mathbb{Y}_t) = \sum_{l=0}^{\infty} G^l \Gamma = (I - G)^{-1}\Gamma$ due to the independence of G_ts over t, and $\mathsf{E}(V_{t-l}) = 0$ for $l \ge 0$. Recall that we have $\Gamma = c_0 \mathbf{1}_N$. Then it is straightforward to have $\mu_Y = c_1^{-1} c_0 \mathbf{1}_N$, where $c_1 = (1 - b_1 - b_2)^{-1}$ is defined in Theorem 1.

We next calculate the covariance of $\mathbb{Y}_t$. Specifically, it can be expressed as

$$\begin{aligned}\mathsf{Cov}(\mathbb{Y}_t) = &\mathsf{Cov}(\sum_{l=0}^{\infty} \Pi_l \Gamma) + \mathsf{Cov}(\sum_{l_1=0}^{\infty} \Pi_{l_1}\Gamma, \sum_{l_2=0}^{\infty} \Pi_{l_2} V_{t-l_2}) \\ &+ \mathsf{Cov}(\sum_{l_2=0}^{\infty} \Pi_{l_2} V_{t-l_2}, \sum_{l_1=0}^{\infty} \Pi_{l_1}\Gamma) + \mathsf{Cov}(\sum_{l=0}^{\infty} \Pi_l V_{t-l}).\end{aligned} \tag{A.3}$$

Recall that $G^* = \mathsf{E}(G_t \otimes G_t) = \mathsf{E}\{\mathbf{B}_1(U_t)\otimes \mathbf{B}_1(U_t)\}(W\otimes W) + \mathsf{E}\{\mathbf{B}_1(U_t)\otimes \mathbf{B}_2(U_t)\}(W \otimes I) + \mathsf{E}\{\mathbf{B}_2(U_t)\otimes \mathbf{B}_1(U_t)\}(I\otimes W) + \mathsf{E}\{\mathbf{B}_2(U_t)\otimes \mathbf{B}_2(U_t)\}(I\otimes I)$, $\widetilde{b}_1 = \{\mathsf{E}\beta_1^2(U_{it})\}^{1/2}$, and $\widetilde{b}_2 = \{\mathsf{E}\beta_2^2(U_{it})\}^{1/2}$. Then we have

$$|\lambda_1(G^*)| \le \widetilde{b}_1^2 |\lambda_1(W)|^2 + 2\widetilde{b}_1\widetilde{b}_2|\lambda_1(W)| + \widetilde{b}_2^2 \le (\widetilde{b}_1 + \widetilde{b}_2)^2 < 1, \tag{A.4}$$

by the fact that $|\lambda_1(W)| < 1$ and the stationarity condition in Theorem 1. Note the matrix G^* can be represented in the Jordan canonical form as $P\Lambda P^{-1}$, where Λ

is a matrix of the Jordan block diagonal form with diagonal elements being $\lambda_i(G^*)$ $(1 \le i \le N)$ and P is an invertible matrix. Then by (A.4), $(G^*)^l$ converges to zero at a geometric rate as $l \to \infty$ and therefore we have

$$\sum_{l=0}^{\infty}(G^*)^l = (I - G^*)^{-1}. \tag{A.5}$$

Similarly, by (A.4) we have $\sum_{l=0}^{\infty} G^l = (I - G)^{-1}$. We then calculate the terms of $\mathsf{Cov}(\mathbb{Y}_t)$ in (A.3) one by one.

For the first term it can be calculated $\mathsf{Cov}(\sum_{l=0}^{\infty}\Pi_l\Gamma) = \mathsf{E}\{(\sum_{l_1}^{\infty}\Pi_{l_1}\Gamma)(\sum_{l_2}^{\infty}\Gamma^\top\Pi_{l_2}^\top)\} - \mu_Y\mu_Y^\top = \sum_{l_1,l_2=0}^{\infty}\mathsf{E}(\Pi_{l_1}\Gamma\Gamma^\top\Pi_{l_2}^\top) - \mu_Y\mu_Y^\top$. Firstly we have $\mathrm{vec}(\Pi_{l_1}\Gamma\Gamma^\top\Pi_{l_2}^\top) = (\Pi_{l_2} \otimes \Pi_{l_1})\mathrm{vec}(\Gamma\Gamma^\top)$. Without loss of generality, we assume $l_1 \ge l_2$. Then it can be obtained that $\mathsf{E}(\Pi_{l_2} \otimes \Pi_{l_1}) = (G^*)^{l_2}(I_N \otimes G)^{l_1 - l_2}$ and

$$\mathsf{E}(\Pi_{l_2} \otimes \Pi_{l_1})\mathrm{vec}(\Gamma\Gamma^\top) = (G^*)^{l_2}(b_1 + b_2)^{l_1 - l_2}c_0^2\mathbf{1}_{N^2} \tag{A.6}$$

due to that $\mathrm{vec}(\Gamma\Gamma^\top) = c_0^2\mathbf{1}_{N^2}$, $(I_N \otimes G)\mathbf{1} = (b_1 + b_2)\mathbf{1}$. Therefore, by (A.4), (A.5), and (A.6) we have $\sum_{l_1,l_2=0}^{\infty}\mathsf{E}\{\mathrm{vec}(\Pi_{l_1}\Gamma\Gamma^\top\Pi_{l_2}^\top)\} = \{\sum_{l_2=0}^{\infty}(G^*)^{l_2}\sum_{l_1>l_2}(b_1 + b_2)^{l_1 - l_2} + \sum_{l_1=0}^{\infty}(G^*)^{l_1}\sum_{l_1 \le l_2}(b_1 + b_2)^{l_2 - l_1}\}c_0^2\mathbf{1} = M_1\mathbf{1}_{N^2}$, where $M_1 = c_1^{-1}c_0^2(1 + b_1 + b_2)(I - G^*)^{-1}$. As a result, we have $\mathrm{vec}\{\mathsf{Cov}(\sum_{l=0}^{\infty}\Pi_l\Gamma)\} = M_1\mathbf{1}_{N^2} - c_1^{-2}c_0^2\mathbf{1}_{N^2}$.

Next, for the second term, it can be derived that $\mathsf{Cov}(\sum_{l_1=0}^{\infty}\Pi_{l_1}\Gamma, \sum_{l_2=0}^{\infty}\Pi_{l_2}V_{t-l_2}) = \sum_{l_1,l_2=0}^{\infty}\mathsf{E}(\Pi_{l_1}\Gamma V_{t-l_2}^\top\Pi_{l_2}^\top)$, due to $\mathsf{E}(\sum_{l_2=0}^{\infty}\Pi_{l_2}V_{t-l_2}) = 0$. It is straightforward to verify that for $l_2 \le l_1$, we have $\mathsf{E}(\Pi_{l_1}V_{t-l_1}\Gamma^\top\Pi_{l_2}^\top) = 0$. Therefore, by (A.4) and (A.5), one could verify that $\sum_{l_1,l_2=0}^{\infty}\mathsf{E}\{\mathrm{vec}(\Pi_{l_1}V_{t-l_1}\Gamma^\top\Pi_{l_2}^\top)\} = \sum_{l_1=0}^{\infty}\sum_{l_2=l_1+1}^{\infty}(G^*)^{l_1}\mathsf{E}\{(G_{t-l_1} \otimes I)(G \otimes I)^{l_2 - l_1 - 1}(I \otimes V_{t-l_1})\}\Gamma = \sum_{l_1=0}^{\infty}\sum_{l_2=l_1+1}^{\infty}(G^*)^{l_1}(b_1 + b_2)^{l_2 - l_1 - 1}\mathrm{vec}(\Sigma_{bv}) = c_1^{-1}c_0(I - G^*)^{-1}\mathrm{vec}(\Sigma_{bv})$. Similarly, one could verify the third term $\mathsf{Cov}(\sum_{l_1=0}^{\infty}\Pi_{l_1}V_{t-l_1}, \sum_{l_2=0}^{\infty}\Pi_{l_2}\Gamma)$ is also equivalent to $c_1^{-1}c_0(I - G^*)^{-1}\mathrm{vec}(\Sigma_{bv})$.

For the last term, we have $\mathsf{Cov}(\sum_{l=0}^{\infty}\Pi_l V_{t-l}) = \sum_{l=0}^{\infty}\mathsf{E}(\Pi_l V_{t-l}V_{t-l}^\top \Pi_l^\top)$ due to that $\mathsf{E}(\Pi_{l_1}V_{t-l_1}V_{t-l_2}^\top \Pi_{l_2}^\top) = 0$ for any $l_1 \neq l_2$. Also note that $\mathsf{E}\{\mathrm{vec}(\Pi_l V_{t-l}V_{t-l}^\top \Pi_l^\top)\} = \mathsf{E}\{(\Pi_l \otimes \Pi_l)\mathrm{vec}(V_{t-l}V_{t-l}^\top)\} = (G^*)^l \mathrm{vec}(\Sigma_V)$. Then by (A.5) we have $\mathsf{Cov}(\sum_{l=0}^{\infty}\Pi_l V_{t-l}) = \sum_{l=0}^{\infty}(G^*)^l\mathrm{vec}(\Sigma_V) = \sum_{l=0}^{\infty} P\Lambda^l P^{-1}\mathrm{vec}(\Sigma_V) = (I - G^*)^{-1}\mathrm{vec}(\Sigma_V)$. Consequently, by (A.4) one obtains that $\mathrm{vec}\{\mathsf{Cov}(\sum_{l=0}^{\infty}\Pi_l V_{t-l})\} = (I - G^*)^{-1}\mathrm{vec}(\Sigma_V)$. As derived in the previous paragraph, $\mathrm{vec}(\Sigma_Y)$ takes the form $(I - G^*)^{-1}\mathrm{vec}(\Sigma_V) + 2c_1^{-1}c_0(1 - G^*)^{-1}\mathrm{vec}(\Sigma_{bv}) + (M_1 - c_1^{-2}c_0^2)\mathbf{1}_{N^2}$. To prove the covariance stationary, it suffices to prove that $\mathsf{Cov}(\mathbb{Y}_t, \mathbb{Y}_{t-h})$ only depends on h. As we can see that for $h \geq 1$, $\mathsf{Cov}(\mathbb{Y}_t, \mathbb{Y}_{t-h}) = \mathsf{E}(Y_t Y_{t-h}^\top) - \mathsf{E}(Y_t)\mathsf{E}(Y_{t-h})^\top = \mathsf{E}\{\mathsf{E}(Y_t|\mathcal{F}_{t-h})Y_{t-h}\} - \mu_y\mu_y^\top$. Furthermore one can obtain that $\mathsf{E}(Y_t|\mathcal{F}_{t-h}) = \sum_{l=0}^{h-1}G^l\Gamma + G^h Y_{t-h}$. So it is straightforward to conclude that $\mathsf{Cov}(\mathbb{Y}_t, \mathbb{Y}_{t-h}) = G^h\Sigma_Y$, which is only related to h. This completes the proof of Step 1.

Step 2. (Uniqueness of the Solution) Assume that $\mathbb{Y}_t^*$ is another covariance stationary solution to the NQAR model. Then we know that $\mathsf{E}|\mathbb{Y}_t^*|_a \preceq C_1\mathbf{1}_N$ for some constant C_1. Similarly we have $\mathbb{Y}_t^* = \sum_{l=0}^{m-1}\Pi_l(\Gamma + V_{t-l}) + \Pi_m\mathbb{Y}_{t-m}^*$. To calculate the difference between $\mathbb{Y}_t$ and $\mathbb{Y}_t^*$, we have $\mathsf{E}|\mathbb{Y}_t - \mathbb{Y}_t^*|_a = \mathsf{E}|(\sum_{l=m}^{\infty}\Pi_l\Gamma + \sum_{l=m}^{\infty}\Pi_l V_{t-l}) - \Pi_m\mathbb{Y}_{t-m}^*|_a \preceq C_2(\sum_{l=m}^{\infty}\mathsf{E}|\Pi_l|_a + \mathsf{E}|\Pi_m|_a)\mathbf{1}_N$, where $C_2 = \max\{C_1, c_0, \mathsf{E}|V_{it}|\}$. It can be verified that $\mathsf{E}|\Pi_l|_a\mathbf{1}_N = \mathsf{E}|\beta_1(U_{it})W + \beta_2(U_{it})I_N|_a^l\mathbf{1}_N \preceq (\widetilde{b}_1 + \widetilde{b}_2)^l\mathbf{1}_N$. Therefore we have $(\sum_{l=m}^{\infty}\mathsf{E}|\Pi_l|_a + \mathsf{E}|\Pi_m|_a)\mathbf{1}_N \preceq C_3(\widetilde{b}_1 + \widetilde{b}_2)^m\mathbf{1}_N$ for some positive constant C_3. As this holds for any m, we can then prove that $\mathbb{Y}_t = \mathbb{Y}_t^*$ with probability 1 due to the stationary condition that $\widetilde{b}_1 + \widetilde{b}_2 < 1$. This completes the proof.

Appendix A.2: Verification of (2.7) and (2.8)

Assume $\widetilde{b}_1 = |\int_0^1 \beta_1^2(u)^2 du|^{1/2} = o(1)$. Recall that $\widetilde{b}_{22} = \int_0^1 \beta_2^2(u)du$, $\widetilde{b}_{12} = \int_0^1 \beta_1(u)\beta_2(u)du$. By the stationary condition we have $\widetilde{b}_{22} < 1$, then by the Cauchy's inequality we have $|\widetilde{b}_{12}| \leq \widetilde{b}_1 |\widetilde{b}_{22}|^{1/2} = o(1)$. Recall that $\text{vec}(\Sigma_Y) = S_1 + S_2 + S_3$, where $S_1 = M_1 \mathbf{1}_{N^2} - c_1^{-2} c_0^2 \mathbf{1}_{N^2}$ $(M_1 = c_1^{-1} c_0^2 (1 + b_1 + b_2)(I - G^*)^{-1})$, $S_2 = 2c_1^{-1} c_0 (1 - G^*)^{-1} \text{vec}(\Sigma_{bv})$, and $S_3 = (I - G^*)^{-1} \text{vec}(\Sigma_V)$. Next, we approximate Σ_Y by neglecting higher order terms of $b_1, \widetilde{b}_{12}, \widetilde{b}_1$. To this end, we first approximate $(I - G^*)^{-1}$ and c_1^{-1} as follows

$$(I - G^*)^{-1} \approx (I - \widetilde{B}_{22})^{-1}(I + M_{12}), \tag{A.7}$$

$$c_1^{-1} \approx (1 - b_2)^{-1}\big\{1 + (1 - b_2)^{-1} b_1\big\}, \tag{A.8}$$

$$c_1^{-2} \approx (1 - b_2)^{-2}\big\{1 + 2(1 - b_2)^{-1} b_1\big\}, \tag{A.9}$$

where $M_{12} = (I - \widetilde{B}_{22})^{-1}\{\widetilde{B}_{12}(W \otimes I) + \widetilde{B}_{21}(I \otimes W)\}$, $\widetilde{B}_{22} = E\{B_2(U_t) \otimes B_2(U_t)\}$, $\widetilde{B}_{12} = E\{B_1(U_t) \otimes B_2(U_t)\}$, and $\widetilde{B}_{21} = E\{B_2(U_t) \otimes B_1(U_t)\}$.

Recall that $\widetilde{b}_{01} = \mathsf{E}\{\beta_1(U_{it})(\beta_0(U_{it}) - b_0)\}$ and $\widetilde{b}_{02} = \mathsf{E}\{\beta_2(U_{it})(\beta_0(U_{it}) - b_0)\}$. Then, by (A.7)-(A.9) one could verify that $S_1 \approx (I - \widetilde{B}_{22})^{-1}\{(1 + 2b_1 - b_2^2) I \otimes I + (1 - b_2^2)(I - \widetilde{B}_{22})^{-1}(\widetilde{B}_{12} + \widetilde{B}_{21})\}(1 - b_2)^{-2} c_0^2 \mathbf{1}_{N^2} - \{1 + 2(1 - b_2)^{-1} b_1\}(1 - b_2)^{-2} c_0^2 \mathbf{1}_{N^2}$, $S_2 \approx 2(1 - b_2)^{-1}(I - \widetilde{B}_{22})^{-1}\{\widetilde{b}_{02} I + (1 - b_2)^{-1} b_1 \widetilde{b}_{02} I + \widetilde{b}_{02} M_{12} + \widetilde{b}_{01} I\} c_0 \text{vec}(I_N)$, and $S_3 \approx (I - \widetilde{B}_{22})^{-1}(I + M_{12})\text{vec}(\Sigma_V)$. Let $S_j = \text{vec}(\Sigma_j)$ for $j = 1, 2, 3$ and $\Sigma_j = (\Sigma_{j,kl}) \in \mathbb{R}^{N \times N}$. Specifically, one can verify for the diagonal elements that $\Sigma_{1,ii} \approx [(1 - \widetilde{b}_{22})^{-1}\{1 - b_2^2 + 2b_1 + 2(1 - \widetilde{b}_{22})^{-1}(1 - b_2^2)\widetilde{b}_{12}\} - (1 - b_2)^{-1}(1 - b_2 + 2b_1)](1 - b_2)^{-2} c_0^2$, $\Sigma_{2,ii} \approx 2(1 - \widetilde{b}_{22})^{-1}(1 - b_2)^{-2}\{\sigma_{bv}(1 - b_2) + b_1 \widetilde{b}_{02}\} c_0$ $(\sigma_{bv} = \widetilde{b}_{01} + \widetilde{b}_{02})$, $\Sigma_{3,ii} \approx (1 - \widetilde{b}_{22})^{-1} \sigma_V^2$. Similarly, we have $\Sigma_{1,i_1 i_2} \approx \{(1 - b_2)^{-2}(1 - b_2^2)^{-2}(1 - b_2^2 + 2b_1 + 2b_1 b_2) - (1 - b_2)^{-3}(1 - b_2 + 2b_1)\} c_0^2$, $\Sigma_{2,i_1 i_2} \approx 2(1 - b_2^2)^{-2}(1 - b_2)^{-1} c_0 b_1 b_2 \widetilde{b}_{02}(w_{i_1 i_2} + w_{i_2 i_1})$, $\Sigma_{3,i_1 i_2} \approx (1 - b_2^2)^{-2} b_1 b_2 (w_{i_1 i_2} + w_{i_2 i_1}) \sigma_V^2$ for $i_1 \neq i_2$, where $w_{i_1 i_2} = n_{i_1}^{-1} a_{i_1 i_2}$ is the (i, j)th element of W. This leads to the desired results.

Appendix A.3: Proof of Theorem 2

In this subsection, we establish the asymptotic normality of $\mathbb{Y}_t$. Define $\widetilde{\mathbb{Y}}_t = \mathbb{Y}_t - \mu_Y = (\widetilde{Y}_{1t}, \cdots, \widetilde{Y}_{Nt})^\top \in \mathbb{R}^N$. We then adopt the dependent Lindeberg central limit theorem (theorem 2) in Bardet et al. (2008) on $(NT)^{-1/2}\widetilde{\mathbb{Y}}_t$. We verify the two conditions in the following two parts. Step 1 is concerning moments bounds, and Step 2 is regarding the time dependency.

Step 1. (Bounding Moments) First, it suffices to show that there exists $0 < \delta \le 1$ satisfying

$$S_T = (NT)^{-(2+\delta)/2}\mathbf{1}_N^\top \sum_{t=0}^T \mathsf{E}|\widetilde{\mathbb{Y}}_t|_a^{2+\delta} \to 0 \tag{A.10}$$

as $T \to \infty$. Then, one can verify that $\mathsf{E}|\widetilde{\mathbb{Y}}_t|_a^{2+\delta} = \mathsf{E}|\sum_{l=0}^\infty(\Pi_l\Gamma + \Pi_l V_{t-l}) - \mu_Y|_a^{2+\delta}$. Further by the Jensen's inequality $S_T = (NT)^{-(2+\delta)/2}\mathbf{1}_N^\top\sum_{t=0}^T \mathsf{E}|\widetilde{\mathbb{Y}}_t|_a^{2+\delta} \le (NT)^{-(2+\delta)/2}\mathbf{1}_N^\top$ $\sum_{t=0}^T\{\mathsf{E}|\sum_{l=0}^\infty \Pi_l V_{t-l}|_a^{2+\delta} + \mathsf{E}|\sum_l \Pi_l\Gamma|_a^{2+\delta} + \mathsf{E}|\mu_Y|_a^{2+\delta}\}(3^{2+\delta}/3)$. It is not hard to see that $(NT)^{-(2+\delta)/2}\mathbf{1}_N^\top\sum_{t=1}^T|\mu_Y|_a^{2+\delta} \to 0$. Let $\delta = 1$ and $S_{Tv} = N^{-3/2}T^{-1/2}\mathbf{1}_N^\top\mathsf{E}|\sum_{l=0}^\infty \Pi_l V_{t-l}|_a^3$. It suffices to show $S_{Tv} \to 0$. We then have $S_{Tv} = S_{Tv1} + S_{Tv2} + S_{Tv3} + S_{Tv4}$, where

$$\begin{aligned}
S_{Tv1} &= N^{-3/2}T^{-1/2}\mathbf{1}_N^\top\mathsf{E}\big\{\sum_l |\Pi_l V_{t-l}|_a^3\big\},\\
S_{Tv2} &= 3N^{-3/2}T^{-1/2}\sum_{l_1}\sum_{l_2>l_1}\mathsf{E}\big\{|\Pi_{l_1}V_{t-l_1}|_a^2 \circ |\Pi_{l_2}V_{t-l_2}|_a\big\},\\
S_{Tv3} &= 3N^{-3/2}T^{-1/2}\sum_{l_1}\sum_{l_2>l_1}\mathsf{E}\big\{|\Pi_{l_1}V_{t-l_1}|_a \circ |\Pi_{l_2}V_{t-l_2}|_a^2\big\},\\
S_{Tv4} &= 6N^{-3/2}T^{-1/2}\sum_{l_1}\sum_{l_2>l_1}\sum_{l_3>l_2>l_1}\mathsf{E}\big\{|\Pi_{l_1}V_{t-l_1}|_a \circ |\Pi_{l_2}V_{t-l_2}|_a \circ |\Pi_{l_3}V_{t-l_3}|_a\big\},
\end{aligned}$$

and $\circ$ means point wise product. We then verify the terms $S_{Tvj} \to 0$ for $j = 1, \cdots, 4$ as follows.

Firstly we have $S_{Tv1} \le N^{-3/2}T^{-1/2}\sum_{l=0}^\infty \mathsf{E}|\Pi_l(V_{t-1})|_a^3 \preceq N^{-3/2}T^{-1/2}\sum_{l=0}^\infty C_3\mathsf{E}|(|\Pi_l|_a$

$\mathbf{1}_N)|_a^3$, where $C_3 = \max_{it} \mathsf{E}|V_{it}|^3$ is finite by Theorem 2. Further, the above term is elementwisely bounded by $C_3 N^{-3/2} T^{-1/2} \sum_{l=0}^{\infty} C_b^l \mathbf{1}_N$, where $C_b = \mathsf{E}(|\beta_1(U_{it})| + |\beta_2(U_{it})|)^3 < 1$ by Theorem 2. Consequently we have $S_{Tv1} \to 0$. Next we look at the second term in S_{Tv}. It can be firstly verified that $\mathsf{E}(3\sum_{l_1}\sum_{l_2>l_1} |\Pi_{l_1} V_{t-l_1}|_a^2 \circ |\Pi_{l_2} V_{t-l_2}|_a) \preceq 3C_v \sum_{l_1}\sum_{l_2>l_1} \mathsf{E}(|\Pi_{l_1} V_{t-l_1}|_a^2 \circ |\Pi_{l_2} \mathbf{1}_N|_a) \preceq 3C_v(\widetilde{b}_1 + \widetilde{b}_2)^{l_2-l_1} \sum_{l_1}\sum_{l_2>l_1} \mathsf{E}(|\Pi_{l_1} V_{t-l_1}|_a^2 \circ |\Pi_{l_1-1}\mathbf{1}_N|_a)$ due to the independence of Π_{l_1-1} and $\prod_{k=1}^{l_2-l_1} G_{t-l_1-k}$, and the inequality $\mathsf{E}(\prod_{k=1}^{l_2-l_1} |G_{t-l_1-k}|_a \mathbf{1}_N) \preceq (\widetilde{b}_1 + \widetilde{b}_2)^{l_2-l_1}\mathbf{1}_N$, where $C_v = \max_i \mathsf{E}(|V_{it}|)$. Further it can be derived that $\mathbf{1}_N^\top \mathsf{E}(|\Pi_{l_1} V_{t-l_1}|_a^2 \circ |\Pi_{l_1-1}\mathbf{1}_N|_a) =$

$$\mathsf{E}(|\Pi_{l_1} V_{t-l_1}|_a^{2\top} |\Pi_{l_1-1}\mathbf{1}_N|_a) \le C_{bv}\mathsf{E}(|\Pi_{l_1}\mathbf{1}_N|_a^{2\top} |\Pi_{l_1}\mathbf{1}_N|_a) \le NC_{bv}C_b^{l_1}, \tag{A.11}$$

where $C_{bv} = (\mathsf{E}\{V_{it}^4\})^{1/2}[\mathsf{E}\{(|\beta_1(U_{it})| + |\beta_2(U_{it})|)^2\}]^{1/2}$. As a result, we have $S_{Tv2} \to 0$. Then, by iteratively applying (A.11), one could obtain $S_{Tv3} \to 0$ and $S_{Tv4} \to 0$, where the details are omitted here. As a result, (A.10) can be obtained.

Step 2. (Time Dependency) Next, we verify conditions imposed on the dynamic dependency structure of $\widetilde{\mathbb{Y}}_t$. To this end, we show the definition of λ dependency as in Bardet et al. (2008).

DEFINITION 1. *(λ **dependency**) A process X_t in $\mathbf{R}^d$ is said to be λ dependent if there exists a sequence $\{\lambda_r\}$ such that $\lambda_r \to 0$ when $r \to \infty$ satisfying*

$$\Big|\mathsf{Cov}\{f(X_{m_1}, X_{m_2}, \cdots, X_{m_v}), g(X_{s_1}, X_{s_2}, \cdots, X_{s_u})\}\Big| \le (uvL_fL_g + vL_f + uL_g)\lambda_r,$$

for all $v, u \in N^* \times N^*$ (N^* denotes the natural number space), L_f and L_g as constants, where v, u are two integers corresponding to support of f and g respectively.

Next we prove that $T^{-1/2}\widetilde{\mathbb{Y}}_t$ is λ dependent with a satisfactory rate. For this propose, we rewrite the NQAR model to be $\widetilde{\mathbb{Y}}_t = G_t\widetilde{\mathbb{Y}}_{t-1} + V_t'$, and $V_t' = V_t + (G_t - G)\mu_Y$. Then we have $\widetilde{\mathbb{Y}}_t = \sum_l^\infty \Pi_l V_{t-l}'$. For convenience, we define $\widetilde{\mathbb{Y}}_t^L = \sum_l^L \Pi_l V_{t-l}'$ as the truncated form of $\widetilde{\mathbb{Y}}_t$.

First of all define $\Im_v = \{\widetilde{\mathbb{Y}}_{m1}, \widetilde{\mathbb{Y}}_{m2}, \cdots, \widetilde{\mathbb{Y}}_{m_v}\}$ and $\Im_u = \{\widetilde{\mathbb{Y}}_{s_1}, \widetilde{\mathbb{Y}}_{s_2}, \cdots, \widetilde{\mathbb{Y}}_{s_u}\}$. Further denote $\Im_v^L = \{\widetilde{\mathbb{Y}}_{m_1}^L, \widetilde{\mathbb{Y}}_{m_2}^L, \cdots, \widetilde{\mathbb{Y}}_{m_v}^L\}$ and $\Im_u^L = \{\widetilde{\mathbb{Y}}_{s_1}^L, \widetilde{\mathbb{Y}}_{s_2}^L, \cdots, \widetilde{\mathbb{Y}}_{s_u}^L\}$. We then have $\mathsf{Cov}\left(f(\Im_v), g(\Im_u)\right) = \mathsf{Cov}\left(f(\Im_v) - f(\Im_v^L), g(\Im_u)\right) + \mathsf{Cov}\left(f(\Im_v^L), g(\Im_u) - g(\Im_u^L)\right) + \mathsf{Cov}\left(f(\Im_v^L), g(\Im_u^L)\right)$. Define $\tilde{f}(X) = f(X) - \mathsf{E}(f(X))$. Without loss of generality, we set $L = r - 1$. Then $\mathsf{Cov}(f(\Im_v^L), g(\Im_u^L)) = 0$, and $|\mathsf{Cov}(f(\Im_v), g(\Im_u))|$ can be bounded by $\|\widetilde{g}\|_\infty \mathsf{E}|\widetilde{f}(\Im_v) - \widetilde{f}(\Im_v^L)| + \|\widetilde{f}\|_\infty \mathsf{E}|\widetilde{g}(\Im_u) - \widetilde{g}(\Im_u^L)| \le c(vL_f + uL_g)\mathbf{1}_N^\top \mathsf{E}|\widetilde{\mathbb{Y}}_m - \widetilde{\mathbb{Y}}_m^L|_a$, where $m = m_1 \wedge m_2$, c is a constant and $\|\cdot\|_\infty$ is the uniform norm of a function, which takes the supremum of the absolute value of a function on its support. Then it can be verified that $\mathsf{E}|\widetilde{\mathbb{Y}}_m - \widetilde{\mathbb{Y}}_m^L|_a = \mathsf{E}|\sum_{l=L+1}^\infty \Pi_l V_{t-l}'|_a \le \sum_{l=L+1}^\infty C_{v'} \mathsf{E}|\Pi_l \mathbf{1}_N|_a \preceq \sum_{l=L+1}^\infty C_{v'}[\mathsf{E}\{|\beta_1(U_{it})| + |\beta_2(U_{it})|\}]^l \mathbf{1}_N \le C_{v'}(\widetilde{b}_1 + \widetilde{b}_2)^{L+1}(1 - \widetilde{b}_1 - \widetilde{b}_2)^{-1}$, where $C_{v'} = \mathsf{E}(|V_{it}|) + 2(\widetilde{b}_1 + \widetilde{b}_2)$. As a result, it can be concluded $\mathsf{Cov}\left(f(\Im_v), g(\Im_u)\right) \to 0$ as $r \to \infty$. This completes the proof.

Appendix A.4: Proof of Theorem 3

The proof follows the conclusion of Theorem 3.3 in Zhang and Cheng (2014), and one could find a more general result by Zhang and Wu (2015). To this end, first define a Gaussian counterpart of $\widetilde{\mathbb{Y}}_t$ by a sequence of Gaussian random vectors $\{\mathbb{Y}_t^*\}_{t=1}^T$, which is independent of $\{\widetilde{\mathbb{Y}}_t\}_{t=1}^T$. Particularly, $\{\mathbb{Y}_t^*\}_{t=1}^T$ preserves the autocovariance structure of $\{\widetilde{\mathbb{Y}}_t\}$ in the sense that $\mathsf{E}(\mathbb{Y}_t^*) = 0$ and $\mathsf{Cov}(\mathbb{Y}_t^*, \mathbb{Y}_{t-h}^*) = \Sigma(h)$. Let $Z^{(T)} = T^{-1/2}\sum_{t=1}^T \mathbb{Y}_t^*$. To accomplish our goal, we break the proof into two steps. Firstly we show that $\rho(T^{-1/2}\sum_{t=1}^T \widetilde{\mathbb{Y}}_t, Z^{(T)}) \to 0$ based on to verify the assumptions. Secondly, we show that $\rho(Z^{(T)}, Z) \to 0$.

Step 1 (Proof of $\rho(T^{-1/2}\sum_{t=1}^{T}\widetilde{\mathbb{Y}}_t, Z^{(T)}) \to 0$)

Before we go into the details, we first give some notations to facilitate the proofs. Let $\Sigma_Y^* = (\sigma_{ij}^*)$ and $\Sigma_Y^* = (\sigma_{ij}^{*(T)})$. In addition, define $\Sigma(h) = (\sigma_{ij}^{(h)})$ for $-\infty < h < \infty$. Following Zhang and Cheng (2014), $\widetilde{\mathbb{Y}}_t$ has a causal representation as $\widetilde{\mathbb{Y}}_t = \mathcal{G}(\cdots, \mathbf{U}_{t-1}, \mathbf{U}_t)$, where $\mathbf{U}_t = (U_{1t}, \cdots, U_{Nt})^\top$ is the *iid* noise vector at time t, and $\mathcal{G}(\cdot)$ is the representation function. Next, define an *iid* copy of $\mathbf{U}_0$ as $\mathbf{U}_0'$, and with $\mathbf{U}_0'$ we define $\widetilde{\mathbb{Y}}_t^* = \mathcal{G}(\cdots, \mathbf{U}_0', \mathbf{U}_{t-1}, \mathbf{U}_t)$. The new dependency measure is defined as,

$$\varrho_{t,i,m} = (\mathsf{E}|Y_{it} - Y_{it}^*|^m)^{1/m} \quad \text{and} \quad \rho_{t_0,i,m} = \sum_{t=t_0}^{\infty} \varrho_{t,i,m}.$$

We then aim to verify the following conditions.

(a) Both $\mathsf{E}(Y_{it}^4)$ and $\max_{1\le i\le N}\sigma_{ii}^*$ are bounded. Moreover, $\sigma_{ii}^{*(T)}$ is bounded from above and below for $1 \le i \le N$.

(b) There exists l_t such that $\max_{1\le i\le N} t\varrho_{t,i,2} \le \ell_t$, with $\sum_{t=0}^{\infty}\ell_t < \infty$.

(c) There exists a constant $\rho < 1$ such that $\max_{1\le i\le N}\rho_{t_0,i,2} = \mathcal{O}(\rho^{t_0})$.

We next verify the conditions one by one in the following.

Verification of (a): First, similar to the proof of (A.10) in Theorem 2, we have $\mathsf{E}(Y_{it}^4)$ to be bounded. Hence it is not hard to see $\sigma_{ii}^{(0)}$ is bounded. Let $\max_i \sigma_{ii}^{(0)} \le c_\sigma$. Suppose $l > 0$, we then have $|\sigma_{ii}^{(l)}| = |e_i^\top \Sigma(l) e_i| = |e_i^\top G^l \Sigma_Y e_i| \le cc_\sigma c_\beta^l e_i^\top \mathbf{1}\mathbf{1}^\top e_i = cc_\sigma c_\beta^l$, where e_i is a zero vector with only the ith element being 1, and c is a finite constant. Subsequently we have $|\sigma_{ii}^*| = |\sum_{l=-\infty}^{\infty}\sigma_{ii}^{(l)}|$ to be bounded by $2cc_\sigma/(1-c_\beta)$ for $1 \le i \le N$ since $c_\beta < 1$. Similarly, one could see $|\sigma_{ii}^{*(T)}| \le \sum_{l=-(T-1)}^{T-1}|\sigma_{ii}^{(l)}|$, which is obviously bounded by the previous argument. For the lower bound, it can be derived $\min_i \sigma_{ii}^{*(T)} = \min_i e_i^\top \Sigma_Y^{*(T)} e_i \ge \lambda_{\min}(\Sigma_Y^{*(T)}) \ge \tau$, where e_i is an N dimensional vector with all elements being 1 but only the ith element being 1, and the last inequality is implied by condition in Theorem 3.

VERIFICATION OF (b) and (c): It can be written that $\widetilde{\mathbb{Y}}_t = \sum_{l=0}^{\infty} \Pi_l V_{t-l} + \sum_{l=0}^{\infty} (\Pi_l - G^l)\Gamma$, where $\Pi_l \stackrel{\text{def}}{=} G_t G_{t-1} \cdots G_{t-l+1}$ for $l > 0$ and $\Pi_0 = 1$. Similarly we have $\widetilde{\mathbb{Y}}_t^* = \{\sum_{l=0}^{t-1} \Pi_l V_{t-l} + \sum_{l=0}^{t} (\Pi_l - G^l)\Gamma\} + \{\Pi_t V_0^* + \sum_{l=t+1}^{\infty} \Pi_l^* V_{t-l} + \sum_{l=t+1}^{\infty} (\Pi_l^* - G^l)\Gamma\}$, where Π_l^*, V_{t-l}^* are to replace $\mathbf{U}_0$ with $\mathbf{U}_0'$ in the representations. Therefore we have $|\widetilde{\mathbb{Y}}_t - \widetilde{\mathbb{Y}}_t^*|_a^2 \preccurlyeq$

$$2\Big|\Pi_t V_0 + \sum_{l=t+1}^{\infty} \Pi_l V_{t-l} + \sum_{l=t+1}^{\infty} (\Pi_l - G^l)\Gamma\Big|_a^2 + 2\Big|\Pi_t V_0^* + \sum_{l=t+1}^{\infty} \Pi_l^* V_{t-l} + \sum_{l=t+1}^{\infty} (\Pi_l^* - G^l)\Gamma\Big|_a^2.$$

The rest to prove the rate of each part follows the same idea as in the proof of (A.10). That leads to the result that $\mathsf{E}|\widetilde{\mathbb{Y}}_t - \widetilde{\mathbb{Y}}_t^*|_a^2 \preccurlyeq c^* c_\beta^{2t} \mathbf{1}$, where c^* is a finite constant. Let $\ell_t = c^* t c_\beta^{2t}$ and then we have $\sum_{t=0}^{\infty} \ell_t < \infty$. Moreover, we have $\max_{1 \le i \le N} \rho_{t_0,i,2} \le c^* c_\beta^{2t_0} / (1 - c_\beta^2)$. The result in (c) can be achieved by letting $\rho = c_\beta^2$.

STEP 2 (PROOF OF $\rho(Z^{(T)}, Z) \to 0$)

It can be easily verified that $\Sigma_Y^{*(T)} = T^{-1} \mathsf{Cov}(Z^{(T)})$. According to Theorem 2 of Chernozhukov et al. (2015), we have

$$\rho\Big(Z^{(T)}, Z\Big) \le c\Delta^{1/3} \Big[\max\big\{1, \log(N/\Delta)\big\}\Big]^{2/3}, \tag{A.12}$$

where $\Delta = \max_{1 \le j,k \le N} |\Sigma_{Y,jk}^{*(T)} - \Sigma_{Y,jk}^*|$ and c is a finite constant. We now prove that $\Delta (\log N)^2 = o(1)$. It can be calculated that

$$\Sigma_Y^{*(T)} = \Sigma_Y^* - T^{-1}(I - G)^{-2}(I - G^T) G \Sigma_Y - T^{-1} \Sigma_Y (G^\top)^2 (I - G^T)^\top (I - G^\top)^{-2}.$$

Further we have $|(I - G)^{-2}(I - G^T) G \Sigma_Y|_a \preccurlyeq c_\sigma |(I - \mathbb{G})^{-2}(I + \mathbb{G}^T) \mathbf{1}_N \mathbf{1}_N^\top| \preccurlyeq c_\sigma (1 - c_\beta)^{-2}(1 + c_\beta^T) c_\beta \mathbf{1}_N \mathbf{1}_N^\top$ and $|\Sigma_Y (G^\top)^2 (I - G^T)^\top (I - G^\top)^{-2}|_a \preccurlyeq c_\sigma (1 - c_\beta)^{-2} (1 + c_\beta^T) c_\beta^2 \mathbf{1}_N \mathbf{1}_N^\top$, where $\mathbb{G} = \|\beta_1\|_4 W + \|\beta_2\|_4 I$. Therefore we have $\Delta \le 2T^{-1} c_\sigma (1 - c_\beta)^{-2} (1 + c_\beta^T) c_\beta =$

$O(T^{-1})$. Recall that $\log N = \mathcal{O}(T^{\delta})$ where $0 \le \delta < 1/11$, hence it can be obtained $\Delta(\log N)^2 = o(1)$. We then have the right side of (A.12) tends to 0 as $N \to \infty$. This completes the proof.

APPENDIX B

In Appendix B, we give the proof of the asymptotic properties in the estimation part. Specifically, a lemma is first proved in Appendix B.1 as a useful tool. Next, Theorem 4 and Theorem 5 are proved in Appendix B.2 and Appendix B.3 respectively. Lastly, the misspecification of adjacency matrix A is discussed in Appendix B.4.

Appendix B.1: A Useful Lemma

In this section, we give the proof of a useful lemma, which is needed for a later proof of the asymptotic properties.

LEMMA 1. *Assume $c_\beta < 1$ and (C1)–(C2), where c_β is defined in (C1). Let $U = (U_1, \cdots, U_N)^\top \in \mathbb{R}^N$ and $V = (V_1, \cdots, V_N)^\top \in \mathbb{R}^N$, where U_i and V_i are iid distributed respectively for $1 \le i \le N$, and independent with Π_l. Assume $\mathsf{E}(U_i^4)^{1/4} \le \nu_u$, $\mathsf{E}(V_i^4)^{1/4} \le \nu_v$, $\mathsf{Cov}(U_i, V_i) \ne 0$, and $\mathsf{Cov}(U_i, U_j) = 0$ for $i \ne j$. Define $\mathbb{G} = \|\beta_1\|_4 W + \|\beta_2\|_4 I \in \mathbb{R}^{N\times N}$. Then the following results hold.*

(a) For any integer $l_1, l_2, l_3, l_4 > 0$ we have

$$\mathsf{E}(|\Pi_{l_1}^\top \Pi_{l_2} \Pi_{l_3}^\top \Pi_{l_4}|_a) \preccurlyeq |\mathbb{G}^{l_1\top} \mathbb{G}^{l_2} \mathbb{G}^{l_3\top} \mathbb{G}^{l_4}|_a. \tag{B.1}$$

(b) There exists a finite integer $K > 0$, such that for any $l > 0$, we have

$$\mathbb{G}^l \mathbb{G}^{l\top} \preccurlyeq l^K c_\beta^{2l} \mathcal{M}, \tag{B.2}$$

where $\mathcal{M} = MM^\top$ with $M = c_m \mathbf{1}\pi^\top + \sum_{j=1}^K W^j$, $c_m > 1$ is a constant, and π is defined in (C2.1). Denote $\mathcal{M}_{ij}$ as the (i,j)th element of $\mathcal{M}$. We then have

$$N^{-2}\mathbf{1}^\top \mathcal{M}\mathbf{1} \to 0, \tag{B.3}$$

$$N^{-2}\, tr(\mathcal{M}^2) \to 0, \tag{B.4}$$

as $N \to \infty$.

(c) For any integer $l_1 \geq l_2$, it holds that

$$\mathsf{Var}(U^\top \Pi_{l_1}^\top \Pi_{l_2} V) \leq 8\nu_u^2 \nu_v^2 c_\beta^{2(l_1+l_2)} l_1^{2K} \{\mathbf{1}^\top \mathcal{M}\mathbf{1} + tr(\mathcal{M}^2)\}. \tag{B.5}$$

(d) We have $\widehat{\Omega}_0 \to_p \Omega_0$ as $\min\{N, T\} \to \infty$.

PROOF OF (a). We first derive an inequality of $\mathsf{E}|\Pi_{l_1}^\top \Pi_{l_2} \Pi_{l_3}^\top \Pi_{l_4}|_a$ as

$$\mathsf{E}|\Pi_{l_1}^\top \Pi_{l_2} \Pi_{l_3}^\top \Pi_{l_4}|_a \preccurlyeq \mathsf{E}\Big\{\big(\prod_{l=0}^{l_1-1}|G_{t-l}|_a\big)^\top\big(\prod_{l=0}^{l_2-1}|G_{t-l}|_a\big)\big(\prod_{l=0}^{l_3-1}|G_{t-l}|_a\big)^\top\big(\prod_{l=0}^{l_4-1}|G_{t-l}|_a\big)\Big\}.$$

We first prove

$$\mathsf{E}(|G_t|_a^\top |G_t|_a M |G_t|_a^\top |G_t|_a) \preccurlyeq \mathbb{G}^\top \mathbb{G}\mathsf{E}(M)\mathbb{G}^\top \mathbb{G} \tag{B.6}$$

for an arbitrary elementwisely positive stochastic matrix M, where $M = (m_{ij}) \in \mathbb{R}^{N\times N}$ is assumed to be independent with G_t. Let $\mathbb{W}_{11} = W^\top W$, $\mathbb{W}_{10} = W^\top$, $\mathbb{W}_{01} = W$, $\mathbb{W}_{00} = I$. Then one could verify that the (i,j)th element of $\mathsf{E}(|G_t|_a^\top |G_t|_a M |G_t|_a^\top |G_t|_a)$ involves a sum of terms like $\mathsf{E}\{|\beta_1^{k_1}(U_{i_1 t})\beta_2^{k_2}(U_{i_2 t})\beta_1^{k_3}(U_{i_3 t})\beta_2^{k_4}(U_{i_4 t})|(\mathbb{W}_{q_1 q_2} M \mathbb{W}_{q_3 q_4})_{ij}\}$, where $q_1, q_2, q_3, q_4 \in \{0, 1\}$, $k_1 + k_2 + k_3 + k_4 = 4$, $0 \leq i_1, i_2, i_3, i_4 \leq N$, and k_1, k_2, k_3, k_4 are integers satisfying $0 \leq k_i \leq 4$. By Hölder's inequality, we have for all i_1, i_2, i_3, i_4, $\mathsf{E}|\beta_1^{k_1}(U_{i_1 t})\beta_2^{k_2}(U_{i_2 t})\beta_1^{k_3}(U_{i_3 t})\beta_2^{k_4}(U_{i_4 t})| \leq \|\beta_1\|_4^{k_1+k_3}\|\beta_2\|_4^{4-(k_1+k_3)}$. By applying the inequality one could obtain (B.6). Subsequently, (B.1) can be derived by recursively applying (B.6).

Proof of (b). Note we have $\|\beta_1\|_4+\|\beta_2\|_4 < 1$. Then (B.2) can be obtained by (5.1) of Lemma 2 (a) in the supplementary material of Zhu et al. (2017) by the condition (C2). For the completeness of the proof, we briefly state the main idea as below. Firstly, for any integer $l > 0$, we have $\mathbb{G}^l = (\|\beta_1\|_4 W + \|\beta_2\|_4 I)^l = \sum_{j=0}^{l} C_l^j \|\beta_1\|_4^j \|\beta_2\|_4^{l-j} W^j$, where $C_l^j = l!/\{j!(l-j)!\}$. Since W is an element-wise non-negative matrix, $|\mathbb{G}^l|_a \preccurlyeq \sum_{j=0}^{l} C_l^j \|\beta_1\|_4^j \|\beta_2\|_4^{l-j} W^j$. Then for $l > K$ we have $|\mathbb{G}^l|_a \preccurlyeq (\|\beta_1\|_4 + \|\beta_2\|_4)^l C\mathbf{1}\pi^\top + \sum_{j=0}^{K} C_l^j \|\beta_1\|_4^j \|\beta_2\|_4^{l-j} W^j$, where this fact is due to $W^l \preccurlyeq C\mathbf{1}\pi^\top$ by the Markov property in condition (C2.1). Further note that $\|\beta_1\|_4^j \|\beta_2\|_4^{n-j} < c_\beta^l$ $(0 \le j \le l)$, and $C_l^K \le l^K$. As a result, for $l > K$ we have,

$$|\mathbb{G}^l|_a \preccurlyeq l^K (\|\beta_1\|_4 + \|\beta_2\|_4)^l M, \tag{B.7}$$

where recall that $M = C\mathbf{1}\pi^\top + \sum_{j=0}^{K} W^j$. It is easy to verify that (B.7) also holds for $l = 1, \cdots, K-1$. Then we have $|\mathbb{G}^l(\mathbb{G}^\top)^l|_a \preccurlyeq l^{2K} c_\beta^{2l} MM^\top$ for any positive integer n. As a result, (B.2) can be proved. Next, (B.3) and (B.4) can be obtained by (5.11) and (5.12) of the supplementary material by Zhu et al. (2017) by condition (C2) respectively.

Proof of (c). We first prove that (B.5) holds for $l_1 = l_2 = l$, then extend the results to $l_1 > l_2$. Let $l_1 = l_2 = l$, we have

$$\mathsf{Var}(U^\top \Pi_l^\top \Pi_l V) = \mathsf{Var}\{\mathsf{E}(U^\top \Pi_l^\top \Pi_l V | \Pi_l)\} + \mathsf{E}\{\mathsf{Var}(U^\top \Pi_l^\top \Pi_l V | \Pi_l)\}. \tag{B.8}$$

We then derive the upper bound for $\mathsf{E}\{\mathsf{Var}(U^\top \Pi_l^\top \Pi_l V)|\Pi_l\}$ and $\mathsf{Var}\{\mathsf{E}(U^\top \Pi_l^\top \Pi_l V)|\Pi_l\}$ in the following respectively.

(c.1) Upper Bound for $\mathsf{E}\{\mathsf{Var}(U^\top \Pi_l^\top \Pi_l V)|\Pi_l\}$. One could first verify that $U^\top \Pi_l^\top \Pi_l V =$

$\text{vec}(\Pi_l)^\top(I\otimes\Pi_l)\text{vec}(VU^\top)$. Denote $\mathcal{V}=\text{vec}(VU^\top)\in\mathbb{R}^{N^2}$. As a consequence, we have

$$\mathsf{Var}(U^\top\Pi_l^\top\Pi_l V|\Pi_l)=\text{vec}(\Pi_l)^\top(I\otimes\Pi_l)\,\mathsf{Cov}(\mathcal{V})(I\otimes\Pi_l^\top)\text{vec}(\Pi_l).$$

Further denote $\mathsf{Cov}(\mathcal{V})=(\Sigma_{\mathcal{V},ij})\in\mathbb{R}^{N^2\times N^2}$, where $\Sigma_{\mathcal{V},ij}\in\mathbb{R}^{N\times N}$ is the (i,j)th block matrix of $\mathsf{Cov}(\mathcal{V})$. Further more, by the Cauchy's inequality, the following bound can be attained,

$$\Sigma_{\mathcal{V},ii}=\mathsf{Cov}(V_iU)\preceq 2\nu_u^2\nu_v^2\mathbf{1}\mathbf{1}^\top, \tag{B.9}$$

$$\Sigma_{\mathcal{V},ij}=\mathsf{Cov}(V_iU,V_jU)\preceq 2\nu_u^2\nu_v^2(I+e_j\mathbf{1}^\top+\mathbf{1}e_i^\top) \tag{B.10}$$

for $i\neq j$, where $e_i\in\mathbb{R}^N$ is a vector with all elements to be 0 but only the ith element being 1. Denote $\Pi_{l,\cdot i}$ as the ith column vector of Π_l. Then we have $|\text{vec}(\Pi_l)^\top(I\otimes\Pi_l)\,\mathsf{Cov}(\mathcal{V})(I\otimes\Pi_l^\top)\text{vec}(\Pi_l)|=|\sum_{i,j=1}^N\Pi_{l,\cdot i}^\top\Pi_l\Sigma_{\mathcal{V},ij}\Pi_l^\top\Pi_{l,\cdot j}|=|\sum_{i=1}^N\Pi_{l,\cdot i}^\top\Pi_l\Sigma_{\mathcal{V},ii}\Pi_l^\top\Pi_{l,\cdot i}+\sum_{i\neq j}\Pi_{l,\cdot i}^\top\Pi_l\Sigma_{\mathcal{V},ij}\Pi_l^\top\Pi_{l,\cdot j}|\ \le\ 2\nu_u^2\nu_v^2\{3\,\text{tr}(|\Pi_l^\top\Pi_l|_a\mathbf{1}\mathbf{1}^\top|\Pi_l^\top\Pi_l|_a)+\text{tr}(|\Pi_l^\top\Pi_l|_a|\Pi_l^\top\Pi_l|_a)\}\ \le\ 6\nu_u^2\nu_v^2\mathbf{1}^\top|\Pi_l^\top|_a|\Pi_l|_a|\Pi_l|_a^\top|\Pi_l|_a\mathbf{1}+2\nu_u^2\nu_v^2\,\text{tr}(|\Pi_l^\top|_a|\Pi_l|_a|\Pi_l|_a^\top|\Pi_l|_a)$. By taking expectation on the right side we have

$$\mathsf{E}\big\{\,\mathsf{Var}(U^\top|\Pi_l^\top\Pi_l|_aV||\Pi_l|_a)\big\}\le 6\nu_u^2\nu_v^2c_\beta^{4l}l^{2K}\mathbf{1}^\top\mathcal{M}\mathbf{1}+2\nu_u^2\nu_v^2c_\beta^{4l}l^K\,\text{tr}(\mathcal{M}^2). \tag{B.11}$$

(c.2) Upper Bound for $\mathsf{Var}\{\mathsf{E}(U^\top\Pi_l^\top\Pi_lV|\Pi_l)\}$. It can be calculated $\mathsf{E}(U^\top\Pi_l^\top\Pi_lV|\Pi_l)\le\nu_u\nu_v\,\text{tr}(|\Pi_l|_a^\top|\Pi_l|_a)$. Firstly we have $\mathsf{Var}\{\text{tr}(|\Pi_l|_a^\top|\Pi_l|_a)\}=$

$$\mathsf{E}[\mathsf{Var}\{\text{tr}(|\Pi_l|_a^\top|\Pi_l|_a)|\Pi_{l-1}\}]+\mathsf{Var}[\mathsf{E}\{\text{tr}(|\Pi_l|_a^\top|\Pi_l|_a)|\Pi_{l-1}\}]. \tag{B.12}$$

Write $\text{tr}(|\Pi_l|_a^\top|\Pi_l|_a)=\sum_i G_{t-l+1,\cdot i}^\top|\Pi_{l-1}|_a^\top|\Pi_{l-1}|_aG_{t-l+1,\cdot i}$. It can be derived $\mathsf{Var}\{\text{tr}(|\Pi_l|_a^\top|\Pi_l|_a)|\Pi_{l-1}\}=\sum_i\mathsf{Var}(G_{t-l+1,\cdot i}^\top|\Pi_{l-1}|_a^\top|\Pi_{l-1}|_aG_{t-l+1,\cdot i}|\Pi_{l-1})\le 2\sum_i(\mathbb{G}_{\cdot i}^\top|\Pi_{l-1}|_a^\top|\Pi_{l-1}|_a$

$\mathbb{G}_{\cdot i})^2 \leq 2c_\beta^2 \sum_i \mathbb{G}_{\cdot i}^\top |\Pi_{l-1}|_a^\top |\Pi_{l-1}|_a \mathbf{1}\mathbf{1}^\top |\Pi_{l-1}|_a^\top |\Pi_{l-1}|_a \mathbb{G}_{\cdot i} = 2c_\beta^2 \mathbf{1}^\top |\Pi_{l-1}|_a^\top |\Pi_{l-1}|_a \mathbb{G}\mathbb{G}^\top |\Pi_{l-1}|_a^\top |\Pi_{l-1}|_a \mathbf{1}$. By similar proofs of (B.1), we have $\mathsf{E}(\mathbf{1}^\top |\Pi_{l-1}|_a^\top |\Pi_{l-1}|_a \mathbb{G}\mathbb{G}^\top |\Pi_{l-1}|_a^\top |\Pi_{l-1}|_a \mathbf{1}) \leq l^K c_\beta^{4l-2} \mathbf{1}^\top \mathcal{M}\mathbf{1}$ by (B.1) and (B.3). Lastly, one could verify that $\mathsf{E}\{\mathrm{tr}(|\Pi_l|_a^\top |\Pi_l|_a)|\Pi_{l-1}\} \leq \sum_i \mathbb{G}_{\cdot i}^\top |\Pi_{l-1}|_a^\top |\Pi_{l-1}|_a \mathbb{G}_{\cdot i} = \mathrm{tr}(\mathbb{G}^\top |\Pi_{l-1}|_a^\top |\Pi_{l-1}|_a \mathbb{G})$. Consequently, by (B.12) we have $\mathsf{Var}\{\mathrm{tr}(|\Pi_l|_a^\top |\Pi_l|_a)\} \leq 2l^K c_\beta^{4l} \mathbf{1}^\top \mathcal{M}\mathbf{1} + \mathsf{Var}\{\mathrm{tr}(\mathbb{G}^\top |\Pi_{l-1}|_a^\top |\Pi_{l-1}|_a \mathbb{G})\}$. By applying similar technique of (B.12) to $\mathsf{Var}\{\mathrm{tr}(\mathbb{G}^\top |\Pi_{l-1}|_a^\top |\Pi_{l-1}|_a \mathbb{G})\}$, one could have $\mathsf{Var}\{\mathrm{tr}(\mathbb{G}^\top |\Pi_{l-1}|_a^\top |\Pi_{l-1}|_a \mathbb{G})\} \leq 2l^K c_\beta^{4l} \mathbf{1}^\top \mathcal{M}\mathbf{1} + \mathsf{Var}\{\mathrm{tr}(\mathbb{G}^{\top 2} |\Pi_{l-2}|_a^\top |\Pi_{l-2}|_a \mathbb{G}^2)\}$. As a result, by using the deduction recursively, one should have $\mathsf{Var}\{\mathsf{E}(U^\top |\Pi_l|_a^\top |\Pi_l|_a V|\Pi_l)\} \leq 2\nu_u^2 \nu_v^2 c_\beta^{4l} l^{K+1} \mathbf{1}^\top \mathcal{M}\mathbf{1}$. By combining the results of (B.11), we have

$$\mathsf{Var}(U^\top |\Pi_l|_a^\top |\Pi_l|_a V) \leq 2\nu_u^2 \nu_v^2 c_\beta^{4l} \Big\{3l^{2K} \mathbf{1}^\top \mathcal{M}\mathbf{1} + l^{K+1} \mathbf{1}^\top \mathcal{M}\mathbf{1} + l^K \,\mathrm{tr}(\mathcal{M}^2)\Big\}. \quad \text{(B.13)}$$

Consequently (B.5) holds.

(c.3) Extend to $l_1 > l_2$. For $l_1 > l_2$, it can be derived that $\mathsf{Var}(U^\top \Pi_{l_1}^\top \Pi_{l_2} V) = \mathsf{Var}\{\mathsf{E}(U^\top \Pi_{l_1}^\top \Pi_{l_2} V|\Pi_{l_1})\} + \mathsf{E}\{\mathsf{Var}(U^\top \Pi_{l_1}^\top \Pi_{l_2} V|\Pi_{l_1})\}$. For $\mathsf{E}\{\mathsf{Var}(U^\top \Pi_{l_1}^\top \Pi_{l_2} V|\Pi_{l_1})\}$, it can be directly calculated that $\mathsf{E}\{\mathsf{Var}(U^\top \Pi_{l_1}^\top \Pi_{l_2} V|\Pi_{l_1})\} \leq 2\nu_u^2 \mathsf{E}(V^\top |\Pi_{l_2}|_a^\top |\Pi_{l_1}|_a |\Pi_{l_1}|_a^\top |\Pi_{l_2}|_a V) \leq 2\nu_u^2 \nu_v^2 c_\beta^{2(l_1+l_2)} l_1^K \mathbf{1}^\top \mathcal{M}\mathbf{1}$, which can be achieved by (B.1) and (B.2) similar to the case of $l_1 = l_2$. For $\mathsf{Var}\{\mathsf{E}(U^\top \Pi_{l_1}^\top \Pi_{l_2} V|\Pi_{l_1})\}$, we have $\mathsf{Var}\{\mathsf{E}(U^\top \Pi_{l_1}^\top \Pi_{l_2} V|\Pi_{l_1})\} \leq 2\nu_u^2 \,\mathsf{Var}(\mathbf{1}^\top |\Pi_{l_1}|_a^\top |\Pi_{l_2}|_a V)$. So we have $\mathsf{Var}(U^\top |\Pi_{l_1}|_a^\top |\Pi_{l_2}|_a V) \leq 2\nu_u^2 \nu_v^2 c_\beta^{2(l_1+l_2)} l_1^K \mathbf{1}^\top \mathcal{M}\mathbf{1} + \nu_u^2 \,\mathsf{Var}(\mathbf{1}^\top |\Pi_{l_1}|_a^\top |\Pi_{l_2}|_a V)$. Note $\mathbf{1}^\top |\Pi_{l_1}|_a^\top |\Pi_{l_2}|_a V = \mathcal{B}_{t-l_1+1}^\top |\Pi_{l_1-1}|_a^\top |\Pi_{l_2}|_a V$. Then by letting $U = \mathcal{B}_{t-l_1+1}$ one could obtain the result that $\mathsf{Var}(U^\top |\Pi_{l_1}|_a^\top |\Pi_{l_2}|_a V) \leq$

$$2\nu_u^2 \nu_v^2 c_\beta^{2(l_1+l_2)} \{l_1^K + (l_1-1)^K\} \mathbf{1}^\top \mathcal{M}\mathbf{1} + \nu_u^2 c_\beta^2 \,\mathsf{Var}(\mathbf{1}^\top |\Pi_{l_1-1}|_a^\top |\Pi_{l_2}|_a V). \quad \text{(B.14)}$$

By applying the recursive formula (B.14), we have $\mathsf{Var}(U^\top |\Pi_{l_1}|_a^\top |\Pi_{l_2}|_a V) \leq 2\nu_u^2 \nu_v^2 c_\beta^{2(l_1+l_2)} \sum_{k=l_2+1}^{l_1} k^K \mathbf{1}^\top \mathcal{M}\mathbf{1} + \nu_u^2 c_\beta^{2(l_1-l_2-1)} \,\mathsf{Var}(\mathcal{B}_{t-l_2}^\top |\Pi_{l_2}|_a^\top |\Pi_{l_2}|_a V)$. The deduction of the upper bound of the second term $\mathsf{Var}(\mathcal{B}_{t-l_2}^\top|\Pi_{l_2}|_a^\top|\Pi_{l_2}|_a V)$ reduces to the case $l_1 = l_2$ previously. By combining the results from (B.13), we have $\mathsf{Var}(U^\top|\Pi_{l_1}|_a^\top|\Pi_{l_2}|_a V) \le 2\nu_u^2\nu_v^2 c_\beta^{2(l_1+l_2)}\{(l_1^{K+1}+3l_2^K)\mathbf{1}^\top\mathcal{M}\mathbf{1}+l_2^{2K}\,\mathrm{tr}(\mathcal{M}^2)\} \le 8\nu_u^2\nu_v^2 c_\beta^{2(l_1+l_2)} l_1^{2K}\{\mathbf{1}^\top\mathcal{M}\mathbf{1}+\mathrm{tr}(\mathcal{M}^2)\}$, which proves (B.5).

Proof of (d): Write $\widehat{\Omega}_0 =$

$$\frac{1}{NT}\sum_{t=1}^T\sum_{i=1}^N X_{it}^\top X_{it} = \begin{pmatrix} 1 & \mathcal{S}_{12} & \mathcal{S}_{13} & \mathcal{S}_{14} \\ & \mathcal{S}_{22} & \mathcal{S}_{23} & \mathcal{S}_{24} \\ & & \mathcal{S}_{33} & \mathcal{S}_{34} \\ & & & \mathcal{S}_{44} \end{pmatrix},$$

where

$$\mathcal{S}_{12} = \frac{1}{N}\sum_{i=1}^N Z_i^\top, \quad \mathcal{S}_{13} = \frac{1}{NT}\sum_{t=1}^T\sum_{i=1}^N w_i^\top \mathbb{Y}_t, \quad \mathcal{S}_{14} = \frac{1}{NT}\sum_{t=1}^T\sum_{i=1}^N Y_{it},$$

$$\mathcal{S}_{22} = N^{-1}\sum_{i=1}^N Z_i Z_i^\top, \quad \mathcal{S}_{23} = \frac{1}{NT}\sum_{t=1}^T\sum_{i=1}^N w_i^\top \mathbb{Y}_t Z_i, \quad \mathcal{S}_{24} = (NT)^{-1}\sum_{t=1}^T\sum_{i=1}^N Y_{it} Z_i,$$

$$\mathcal{S}_{33} = \frac{1}{NT}\sum_{t=1}^T\sum_{i=1}^N (w_i^\top \mathbb{Y}_t)^2, \quad \mathcal{S}_{34} = \frac{1}{NT}\sum_{t=1}^T\sum_{i=1}^N w_i^\top \mathbb{Y}_t Y_{it}, \quad \mathcal{S}_{44} = \frac{1}{NT}\sum_{t=1}^T\sum_{i=1}^N Y_{it}^2.$$

One can directly conclude that $\mathcal{S}_{12} \to_p \mathbf{0}_q^\top$ and $\mathcal{S}_{22} \to_p \Sigma_z$ by the law of large numbers. Recall that $\kappa_1 = \lim_{N\to\infty} N^{-1}\mathrm{tr}(\Sigma_Y)$, $\kappa_2 = \lim_{N\to\infty} N^{-1}\mathrm{tr}(W\Sigma_Y)$, $\kappa_3 = \lim_{N\to\infty} N^{-1}\mathrm{tr}(W\Sigma_Y W^\top)$, $\kappa_4 = \lim_{N\to\infty} N^{-1}\mathrm{tr}\{(I-G)^{-1}\}$, and $\kappa_5 = \lim_{N\to\infty} N^{-1}\mathrm{tr}\{W (I-G)^{-1}\}$. We first list the component-wise limit for each element in $\widehat{\Omega}_0$ in expectation, and then verify that the variances of these components converge to 0 in the next steps. Denote $Z = (Z_1, Z_2, \cdots, Z_n)^\top \in \mathbb{R}^{N\times q}$ and recall that $r = (\int \gamma_1(u)du, \int \gamma_2(u)du, \cdots,$

$\int \gamma_q(u)du)^\top$. It can be derived that $\mathsf{E}(\mathcal{S}_{22}) = \Sigma_z$, $\mathsf{E}(\mathcal{S}_{23}) = N^{-1}\mathsf{E}\{Z^\top W(I-G)^{-1}Zr\} \to \kappa_5\Sigma_z r$, $\mathsf{E}(\mathcal{S}_{24}) = N^{-1}\mathsf{E}(Z^\top(I-G)^{-1}Z)r \to \kappa_4\Sigma_z r$, $\mathsf{E}(\mathcal{S}_{33}) = N^{-1}\operatorname{tr}\{W^\top W\Sigma_Y\} + c_b^2 \to \kappa_3 + c_b^2$, $\mathsf{E}(\mathcal{S}_{34}) = N^{-1}\operatorname{tr}\{W^\top\Sigma_Y\} + c_b^2$, and $\mathsf{E}(\mathcal{S}_{44}) = N^{-1}\operatorname{tr}\{\Sigma_Y\} + c_b^2 \to \kappa_1 + c_b^2$ by condition (C2.3).

We next verify $\mathsf{Var}(\mathcal{S}_{ij}) \to 0$ for $1 \le i, j \le 4$. Since the proofs are quite similar, we show $\mathcal{S}_{44} \to_p \kappa_1 + c_b^2$ for simplicity. The proof contains two steps. In the first step, we prove that for any fixed t, $\mathsf{Var}(\mathcal{S}_{44}) \to 0$ as $N \to \infty$. Next, we deal with the dependence over time (i.e., $1 \le t \le T$). Specifically, the near epoch dependence of Y_{it} and its functional forms are presented and consequently the desired law of large numbers results are established.

Step 1. Proof of $\mathsf{Var}(\mathcal{S}_{44}) \to 0$. Recall that $\mathbb{Y}_t$ has the decomposition in the (A.1). Without loss of generality, assume $\Gamma = \mathbf{1}_N$. Then we have $\mathbb{Y}_t^\top\mathbb{Y}_t = \sum_{l_1,l_2=0}^\infty(\mathbf{1}^\top\Pi_{l_1}^\top\Pi_{l_2}\mathbf{1} + 2\mathbf{1}^\top\Pi_{l_1}^\top\Pi_{l_2}V_{t-l_2} + V_{t-l_1}^\top\Pi_{l_1}^\top\Pi_{l_2}V_{t-l_2})$. By the Cauchy's inequality, it suffices to show $N^{-2}\,\mathsf{Var}\left(\sum_{l_1,l_2=0}^\infty V_{t-l_1}^\top\Pi_{l_1}^\top\Pi_{l_2}V_{t-l_2}\right) \to 0$, $N^{-2}\,\mathsf{Var}\left(\sum_{l_1,l_2=0}^\infty \mathbf{1}^\top\Pi_{l_1}^\top\Pi_{l_2}V_{t-l_2}\right) \to 0$, and $N^{-2}\,\mathsf{Var}\left(\sum_{l_1,l_2=0}^\infty \mathbf{1}^\top\Pi_{l_1}^\top\Pi_{l_2}\mathbf{1}\right) \to 0$ as $N \to \infty$. Since their proofs are almost the same, we prove $N^{-2}\,\mathsf{Var}\left(\sum_{l_1,l_2=0}^\infty \mathbf{1}^\top\Pi_{l_1}^\top\Pi_{l_2}\mathbf{1}\right) \to 0$ in the following for simplicity. To this end, first it can be shown that $\mathsf{Var}(\sum_{l_1,l_2=0}^\infty \mathbf{1}^\top\Pi_{l_1}^\top\Pi_{l_2}\mathbf{1}) = \sum_{l_1\ne l_2}^\infty \mathsf{Var}\left(\mathbf{1}^\top\Pi_{l_1}^\top\Pi_{l_2}\mathbf{1}\right) + \sum_{l_1,l_2=0}^\infty \mathsf{Cov}\left(\mathbf{1}^\top\Pi_{l_1}^\top\Pi_{l_1}\mathbf{1}, \mathbf{1}^\top\Pi_{l_2}^\top\Pi_{l_2}\mathbf{1}\right)$. Then it suffices to show that

$$N^{-2}\sum_{l_1\ne l_2}^\infty \mathsf{Var}\left(\mathbf{1}^\top\Pi_{l_1}^\top\Pi_{l_2}\mathbf{1}\right) \to 0, \tag{B.15}$$

$$N^{-2}\sum_{l_1,l_2=0}^\infty \mathsf{Cov}\left(\mathbf{1}^\top\Pi_{l_1}^\top\Pi_{l_1}\mathbf{1}, \mathbf{1}^\top\Pi_{l_2}^\top\Pi_{l_2}\mathbf{1}\right) \to 0 \tag{B.16}$$

$N \to \infty$. We then prove (B.15) and (B.16) separately as follows. Write $\mathbf{1}^\top\Pi_{l_1}^\top\Pi_{l_2}\mathbf{1} = \mathcal{B}_{t-l_1+1}^\top\Pi_{l_1-1}^\top\Pi_{l_2-1}\mathcal{B}_{t-l_2+1}$, where $\mathcal{B}_{t-l_1+1} = \mathbf{B}_{1(t-l_1+1)}\mathbf{1}_N + \mathbf{B}_{2(t-l_1+1)}\mathbf{1}_N$. It can be calculated $\sum_{l_1\ne l_2}^\infty \mathsf{Var}\left(\mathcal{B}_{t-l_1+1}^\top\Pi_{l_1-1}^\top\Pi_{l_2-1}\mathcal{B}_{t-l_2+1}\right) = 2\sum_{l_2=0}^\infty\sum_{l_1>l_2}\mathsf{Var}\left(\mathcal{B}_{t-l_1+1}^\top\Pi_{l_1-1}^\top\Pi_{l_2-1}\right.$

$\mathcal{B}_{t-l_2+1}\big)$. By (B.5) we have $\mathsf{Var}\left(\mathcal{B}_{t-l_1+1}^\top \Pi_{l_1-1}^\top \Pi_{l_2-1}\mathcal{B}_{t-l_2+1}\right) \le 8c_\beta^{2(l_1+l_2-2)} l_1^{2K}\{\mathbf{1}^\top\mathcal{M}\mathbf{1} + \mathrm{tr}(\mathcal{M}^2)\}$. Then we have (B.15) due to $\sum_{l_2=0}^\infty \sum_{l_1>l_2} c_\beta^{2(l_1+l_2-2)} l_1^{2K} < \infty$ and $N^{-2}\{\mathbf{1}^\top\mathcal{M}\mathbf{1}+\mathrm{tr}(\mathcal{M}^2)\} \to 0$ by (B.3) and (B.4). For (B.16), it can be shown by Cauchy's inequality that $\mathsf{Cov}(\mathbf{1}^\top\Pi_{l_1}^\top\Pi_{l_1}\mathbf{1}, \mathbf{1}^\top\Pi_{l_2}^\top\Pi_{l_2}\mathbf{1}) \le \mathsf{Var}(\mathbf{1}^\top\Pi_{l_1}^\top\Pi_{l_1}\mathbf{1})^{1/2}\,\mathsf{Var}(\mathbf{1}^\top\Pi_{l_2}^\top\Pi_{l_2}\mathbf{1})^{1/2} \le 8c_\beta^{2(l_1+l_2)} l_1^K l_2^K$ $\{\mathbf{1}^\top\mathcal{M}\mathbf{1} + \mathrm{tr}(\mathcal{M}^2)\}$ by (B.5). Then (B.16) holds since $\sum_{l_1,l_2} c_\beta^{2(l_1+l_2)} l_1^K l_2^K < \infty$ and $N^{-2}\{\mathbf{1}^\top\mathcal{M}\mathbf{1} + \mathrm{tr}(\mathcal{M}^2)\} \to 0$ as $N \to \infty$. This completes the proof.

Step 2. L_1 Near Epoch Dependence. In this step, we further prove that $N^{-1}\sum_i Y_{it}^2$ satisfies near epoch dependence for $1 \le t \le T$. First we give the definition of L_1 near epoch dependence as below.

DEFINITION 2. *(L_1 **near epoch dependence**) A triangular array U_{it} in $\mathbb{R}^1$ is said to be L_1 near epoch dependent (NED) if there exists constants c_{it} and a sequence $\{v_J, J \ge 1\}$ such that $v_J \to 0$ when $J \to \infty$ satisfying*

$$\mathsf{E}\Big|(U_{it}) - \mathsf{E}(U_{it}|\mathcal{F}_{t-J}, \cdots, \mathcal{F}_t, \cdots, \mathcal{F}_{t+J})\Big| \le c_{it} v_J.$$

Given the definition, we firstly prove that Y_{it}s are L_1 NED by Andrews (1988). Next, according to Chapter 7 Lemma 1 of Gallant (2009), the smooth transformations of Y_{it}s (e.g., $N^{-1}\sum_i Y_{it}^2$) are also NED. Since Y_{it} has finite forth moment, then by Gallant (2009) we have $N^{-1}\sum_{i=1}^N Y_{it}^2$ is a uniformly integrable L_1 mixinggale. Consequently, according to Theorem 1 of Andrews (1988), we could have $(NT)^{-1}\sum_{t=1}^T\sum_{i=1}^N Y_{it}^2$ converge in probability as $N \to \infty$ and $T \to \infty$. We then prove that Y_{it} is NED in the following.

Denote $\mathcal{F}_{t-J}^{t+J} = \{\mathcal{F}_{t-J}, \cdots, \mathcal{F}_t, \cdots, \mathcal{F}_{t+J}\}$ and $\Pi_{t_1}^{t_2} = \prod_{t=t_1}^{t_2} G_t$. We then have the following inequality as

$$\mathsf{E}\Big\{e_i^\top\big|\mathbb{Y}_t - \mathsf{E}(\mathbb{Y}_t|\mathcal{F}_{t-J}^{t+J})\big|_a\Big\} \;\le\; \mathsf{E}\Big[e_i^\top\{\sum_{l=J+1}^{\infty}\Pi_l V_{t-l} + \sum_{l=J+1}^{\infty}\Pi_{J+1}(\Pi_{t-l-J}^{t-(J+1)} - G^{l-J-1})\Gamma\}\Big]$$
$$\le \sum_{l=J+1}^{\infty}(b_1^a + b_2^a)^l c_v + \sum_{l=J+1}^{\infty} 2(b_1^a + b_2^a)^l c_0,$$

where $c_v = \mathsf{E}|V_{it}|$. Let $v_J = (b_1^a + b_2^a)^{J+1}$ and $c_{it} = (1 - b_1^a - b_2^a)^{-1}(2c_0 + c_v)$. By condition (C1) we have $b_1^a + b_2^a < 1$, thus Y_{it}s are L_1 NED according to Definition 2. This completes the proof of Step 2.

Appendix B.2: Proof of Theorem 4

Recall that $V_{it\tau} = Y_{it} - X_{i(t-1)}^\top\theta(\tau)$ and define $\hat{v} = \sqrt{NT}\big(\hat{\theta}(\tau) - \theta(\tau)\big)$. Then we have $\rho_\tau\big(Y_{it} - X_{i(t-1)}^\top\hat{\theta}(\tau)\big) = \rho_\tau\big(V_{it\tau} - (NT)^{-1/2}X_{i(t-1)}^\top\hat{v}\big)$, where $V_{it\tau} = Y_{it} - X_{i(t-1)}^\top\theta(\tau)$. Then the minimization of (3.1) is equivalent to minimizing for a fixed τ,

$$Z_{NT}(v,\tau) = \sum_{i=1}^{N}\sum_{t=1}^{T}\Big\{\rho_\tau\big(V_{it\tau} - (NT)^{-1/2}X_{i(t-1)}^\top v\big) - \rho_\tau(V_{it\tau})\Big\},$$

One could verify that $\hat{v} = \arg\min_v Z_{NT}(v,\tau)$. The objective function $Z_{NT}(v,\tau)$ is a convex random function. Recall that $\psi_\tau(u) = \tau - I(u < 0)$. Let $\nu_{it} = (NT)^{-1/2}v^\top X_{it}$, and one could further write $Z_{NT}(v,\tau)$ as $Z_{NT}(v,\tau) =$

$$-\sum_{i,t}\Big[(NT)^{-1/2}v^\top X_{i(t-1)}\psi_\tau(V_{it\tau}) + \int_0^{\nu_{i(t-1)}}\big\{\mathbf{1}(V_{it\tau} \le s) - \mathbf{1}(V_{it\tau} < 0)\big\}ds\Big]$$

$\stackrel{\text{def}}{=} v^\top\xi_1 + \xi_2$. It is implied by (C3) the $\Sigma_\theta(\tau)$ is with uniformly bounded eigenvalues over $\tau \in B$. According to Kato (2009), in order to prove that $\hat{v}$ takes the representation in (3.2), it suffices to prove (a) $\xi_2 \to_p v^\top\Omega_1 v$ with Ω_1 defined in (C3) being a positive definite matrix with uniformly bounded eigenvalues on B. Also we would like to prove

(b) ξ_1 is tight for $\tau \in B \in (0,1)$, and ξ_1 converges weakly to a Brownian Bridge.

Note that (a) and (b) ensure the objective function $\sup_{\tau \in B} |Z_{NT}(v,\tau)|$ is convex in v for each τ and bounded in τ for each v. (a) would lead to $\sup_{\tau\in B} |\xi_2| = \mathcal{O}_p(1)$, and (b) would lead to $\sup_{\tau \in B} \|\xi_1\| = \mathcal{O}_p(1)$ by continuous mapping theorem. We then prove (a) in what follows and then prove (b) in the Appendix B.3.

Define $\xi_{2it} = \int_0^{\nu_{i(t-1)}} \big\{\mathbf{1}(V_{it\tau} \le s) - \mathbf{1}(V_{it\tau} < 0)\big\} ds$. To prove $\xi_2 = \sum_{i=1}^N \sum_{t=1}^T \xi_{2it} \to_p v^\top \Omega_1 v$, we decompose ξ_{2it} as $\xi_{2it} = \mathsf{E}(\xi_{2it}|\mathcal{F}_{t-1}) + \overline{\xi}_{2it}$, where $\overline{\xi}_{2it} = \xi_{2it} - \mathsf{E}(\xi_{2it}|\mathcal{F}_{t-1})$. We then prove $\sum_{i=1}^N \sum_{t=1}^T \mathsf{E}(\xi_{2it}|\mathcal{F}_{t-1}) \to_p 2^{-1} v^\top \Omega_1 v$ and $\sum_{i=1}^N \sum_{t=1}^T \overline{\xi}_{2it} \to_p 0$ respectively as follows.

We first evaluate $\sum_{i=1}^N \sum_{t=1}^T \mathsf{E}(\xi_{2it}|\mathcal{F}_{t-1})$. It can be expressed that $\sum_{i,t} \mathsf{E}(\xi_{2it}|\mathcal{F}_{t-1}) = \sum_{i=1}^N \sum_{t=1}^T \mathsf{E}[\int_0^{\nu_{i(t-1)}} \{\mathbf{1}(V_{it\tau} \le s) - \mathbf{1}(V_{it\tau} < 0)\} ds | \mathcal{F}_{t-1}] = \sum_{i=1}^N \sum_{t=1}^T \int_0^{\nu_{i(t-1)}} \{F_{i(t-1)}(s + F^{-1}_{i(t-1)}(\tau)) - F_{it-1}(F^{-1}_{i(t-1)}(\tau))\}/s \cdot s ds$. This yields that

$$
\begin{aligned}
\sum_{i,t} \mathsf{E}(\xi_{2it}|\mathcal{F}_{t-1}) &= \sum_{i,t} \int_0^{\nu_{i(t-1)}} f_{it-1}(F^{-1}_{it-1}(\tau)) s ds + o_p(1) \\
&= \sum_{i,t} (2NT)^{-1} f_{i(t-1)}\big(X_{i(t-1)}^\top \theta(\tau)\big) v^\top X_{i(t-1)} X_{i(t-1)}^\top v + o_p(1) \to_p 1/2 v^\top \Omega_1 v \qquad \text{(B.17)}
\end{aligned}
$$

according to condition (C3).

Next, we prove $\sum_{i,t} \overline{\xi}_{2it} \to_p 0$. It is not difficult to see that $\overline{\xi}_{2it}$ is a martingale difference sequence, which can be written as $\overline{\xi}_{2it} = \int_0^{\nu_{i(t-1)}} \delta_{it\tau}(s) - \delta_{it\tau}(0) ds$, where $\delta_{it\tau}(s) = \big\{\mathbf{1}(V_{it\tau} \le s) - F_{i(t-1)}(s + X_{i(t-1)}^\top \theta(\tau))\big\}$. It suffices to show $\mathsf{E}(|\sum_{i,t} \overline{\xi}_{2it}|)^2 = \sum_{i_1,i_2} \sum_{t_1,t_2} \mathsf{E}(\overline{\xi}_{2i_1t_1} \overline{\xi}_{2i_2t_2}) \to 0$. Importantly, recall that $V_{it\tau} = X_{i(t-1)}^\top (\theta(U_{it}) - \theta(\tau))$, therefore $V_{it\tau}$ and $V_{jt\tau}$ would be conditionally independent on $\mathcal{F}_{t-1}$. Thus it can be shown that $\mathsf{E}\{\int_0^{\nu_{i(t-1)}} \delta_{it\tau}(s) ds \int_0^{\nu_{j(t-1)}} \delta_{jt\tau}(s) ds\} = \mathsf{E}[\mathsf{E}\{\int_0^{\nu_{i(t-1)}} \delta_{it\tau}(s) ds \int_0^{\nu_{j(t-1)}} \delta_{jt\tau}(s) ds \,|\mathcal{F}_{t-1}\}] = 0$ due to the conditional independence of $\delta_{it\tau}(s)$ and $\delta_{jt\tau}(s)$ given $\mathcal{F}_{t-1}$. Similarly, for $t_1 > t_2$ we have $\mathsf{E}\{\int_0^{\nu_{i(t_1-1)}} \delta_{it_1\tau}(s)ds \int_0^{\nu_{i(t_2-1)}} \delta_{jt_2\tau}(s)ds\} = \mathsf{E}[\mathsf{E}\{\int_0^{\nu_{i(t_1-1)}} \delta_{it_1\tau}(s)ds \int_0^{\nu_{i(t_2-1)}} \delta_{jt_2\tau}(s)ds|\mathcal{F}_{t_1-1}\}] = 0$. Therefore, we have $\mathsf{E}\{\overline{\xi}_{2i_1t_1\tau}\overline{\xi}_{2i_2t_2\tau}\} = 0$ for $i_1 \neq i_2$ or $t_1 \neq t_2$. Then $\sum_{i_1,i_2}\sum_{t_1,t_2} \mathsf{E}(\overline{\xi}_{2i_1t_1}\overline{\xi}_{2i_2t_2}) = \sum_i \sum_t \mathsf{E}(\overline{\xi}^2_{2it})$. Next, write $\mathsf{E}(\overline{\xi}^2_{2it}) = \mathsf{E}(\xi^2_{2it}) - \mathsf{E}\{\mathsf{E}(\xi_{2it}|\mathcal{F}_{t-1})\}^2$. Further it can be derived that $\mathsf{E}(\xi^2_{2it}) = \mathsf{E}|\int_0^{\nu_{i(t-1)}}\{\mathbf{1}(V_{it\tau} \leq s) - \mathbf{1}(V_{it\tau} \leq 0)\}ds|^2 \leq |\nu_{i(t-1)}|\mathsf{E}\int_0^{|\nu_{i(t-1)}|}\{\mathbf{1}(V_{it\tau} \leq s) - \mathbf{1}(V_{it\tau} \leq 0)\}^2 ds$ by the Chebyshev's inequality. Further we have $|\nu_{i(t-1)}|\mathsf{E}[\int_0^{|\nu_{i(t-1)}|}\{\mathbf{1}(V_{it\tau} \leq s) - \mathbf{1}(V_{it\tau} \leq 0)\}ds] = |\nu_{i(t-1)}|\mathsf{E}[\int_0^{|\nu_{i(t-1)}|}\{F_{i(t-1)}(s + F^{-1}_{i(t-1)}(\tau)) - F_{i(t-1)}(F^{-1}_{i(t-1)}(\tau))\}/s \cdot sds]$. By similar technique with (B.17), one could obtain $\sum_{i,t}\mathsf{E}(\overline{\xi}^2_{2it}) \leq \mathsf{E}\{\sum_{i,t} 2^{-1}(NT)^{-3/2}|f_{i(t-1)}(X^\top_{i(t-1)}v)| |X^\top_{i(t-1)}v|^{3/2}\} + o(1)$. Since we have $f_{it}(\cdot)$ is bounded and $(NT)^{-3/2}\sum_{i,t}\mathsf{E}(v^\top X_{it}X^\top_{it}v)^2 = \mathcal{O}\big((NT)^{-1/2}\big) \to 0$, then it can be obtained that $\sum_{i,t}\mathsf{E}(\overline{\xi}^2_{2it}) \to 0$. Lastly, following similar argument of tightness as in Wagener et al. (2012), we can prove that $\sum_{i,t}\xi_{2it} \to_p 0$ uniformly over $\tau \in B$. This completes the proof of $\xi_2 \to_p v^\top\Omega_1 v$ for any $\tau \in (0,1)$.

Appendix B.3: Proof of Theorem 5

In this section, we are going to show that ξ_1 converges in distribution to a Brownian Bridge $\Omega_0^{1/2}B_{q+3}(\tau)$, where Ω_0 is defined in (3.3), and $B_{q+3}(\tau)$ is a $(q+3)$-dimensional Brownian bridge. To prove this conclusion, we adopt two steps:

(I) For an arbitrary k-dimensional vector $(\tau_1, \tau_2, \cdots, \tau_k)^\top \in \mathbb{R}^p$ and $\eta \in \mathbb{R}^{q+3}$, $\big(\xi_1(\tau_1), \xi_1(\tau_2), \cdots, \xi_1(\tau_k)\big)^\top\eta \in \mathbb{R}^k$ converge to a k-dimensional multivariate normal distribution.

(II) $\eta^\top\xi_1(\tau)$ for $\tau \in B \in (0,1)$ is tight, where B is a compact set in $(0,1)$.

Step I. Denote $\psi_t = \big(\psi(V_{1t\tau}), \cdots, \psi(V_{Nt\tau})\big)^\top \in \mathbb{R}^N$ for convenience. We then have $\mathsf{E}(\mathbb{X}^\top_{t-1}\psi_t|\mathcal{F}_{t-1}) = 0$. Therefore, $\mathbb{X}^\top_{t-1}\psi_t$ is a martingale difference sequence for $1 \leq t \leq T$. To prove (B.1), we define $\zeta_t = (NT_N)^{-1/2}\eta^\top\mathbb{X}^\top_{t-1}\psi_t$ and $\mathbb{S}_{Nt} = \sum_{s=1}^t \zeta_{\eta t}$. Then one can see that $\{\zeta_t, \mathcal{F}_{t-1}, -\infty < t < T_N, N \geq 1\}$ is a martingale array, where the number of observed time points T_N is assumed to depend on N with $T_N \to \infty$ as $N \to \infty$. As a result, the double sequence $\{\mathbb{S}_{Nt}, \mathcal{F}_t, -\infty < t \leq T_N, N \geq 1\}$ is a martingale array. As a consequence, the martingale difference central limit theorem can be applied (Hall and Heyde, 2014). Specifically, it requires two conditions as follows. First we have

$$\sum_{t=1}^{T_N} \mathsf{E}\{\zeta_t^2 \mathbf{1}|\zeta_t^2| > \delta | \mathcal{F}_{t-1}\} \leq \delta^{-2} \sum_{t=1}^{T_N} \mathsf{E}(|\zeta_t|^4 | \mathcal{F}_{t-1})$$
$$\leq \delta^{-2}\tau^2(1-\tau)^2(NT_N)^{-2} \sum_{t=1}^{T_N} (\eta^\top \mathbb{X}_{t-1}^\top \mathbb{X}_{t-1}\eta)^2 \to_p 0, \qquad \text{(B.18)}$$

where the last inequality is due to $\mathsf{E}\psi^4(V_{it\tau}) \leq \tau^2(1-\tau)^2$. Since by the proof of (d) of Lemma 1, we have $(NT_N)^{-2}\sum_{t=1}^{T_N} \mathsf{E}(\eta^\top \mathbb{X}_{t-1}^\top \mathbb{X}_{t-1}\eta)^2 \to 0$. Therefore (B.18) can be implied. Secondly, we also have the condition

$$\sum_{t=1}^{T_N} \mathsf{E}\{\zeta_t^2 | \mathcal{F}_{t-1}\} = \frac{\tau(1-\tau)}{NT} \sum_{t=1}^{T_N} \eta^\top \mathbb{X}_{t-1}^\top \mathbb{X}_{t-1}\eta \to_p \tau(1-\tau)\eta^\top \Omega_0 \eta, \qquad \text{(B.19)}$$

by (d) of Lemma 1 in Appendix B.1. Therefore, by the central limit theorem for martingale difference sequence in Hall and Heyde (2014), we have that $\xi_1(\tau)$ converge in distribution to Gaussian distribution $\mathsf{N}(0, \tau(1-\tau)\eta^\top\Omega_0\eta)$ for fixed τ. The conclusion also holds for any finite dimensional vector $(\tau_1, \tau_2, \cdots, \tau_k)^\top$, which proves (B.1).

Step II Then we prove that $\eta^\top \xi_1(\tau)$ for $\tau \in B \in (0,1)$ is tight. The definition of tightness is given as follows.

DEFINITION 3. *A process $W_{NT}(\tau)$ is said to be tight if and only if for any $\delta > 0$ there exists a compact set E such that $\mathbf{sup}_{\tau\in\mathbf{E}}\mathrm{P}(W_{NT}(\tau) \in E) > 1 - \delta$.*

Define $\psi_1(D) = -(NT)^{-1/2}\sum_{i,t} X_{i(t-1)}\{\psi_{\tau_2}(V_{it\tau_2}) - \psi_{\tau_1}(V_{it\tau_1})\}$ for any interval $D = (\tau_1, \tau_2]$. To show the tightness, we adopt Theorem 15.6 in Billingsley (1968) and prove a sufficient Chentsov-Billingsley type of inequality as follows.

LEMMA 2. *For any two intervals $D_1 = (\tau_1, \tau_2]$ and $D_2 = (\tau_2, \tau_3]$, we have*

$$\mathsf{E}\Big[\big\{\eta^\top \xi_1(D_1)\big\}^2\big\{\eta^\top \xi_1(D_2)\big\}^2\Big] \leq C(\tau_3 - \tau_1), \tag{B.20}$$

where C is a finite positive constant.

To prove Lemma 2, we have $\mathsf{E}[\{\eta^\top \xi_1(D_1)\}^2\{\eta^\top \xi_1(D_2)\}^2] = (NT)^{-2}\mathsf{E}[\{\sum_{i,t}\eta^\top X_{i(t-1)} \delta_{it}(\tau_1,\tau_2)\}^2\{\sum_{i,t}\eta^\top X_{i(t-1)}\delta_{it}(\tau_2,\tau_3)\}^2]$, where $\delta_{it}(\tau,\tau') = \psi_{\tau'}(V_{it\tau'}) - \psi_\tau(V_{it\tau})$. Next, by Cauchy's inequality, we have $\mathsf{E}[\{\sum_{i,t}\eta^\top X_{i(t-1)}\delta_{it}(\tau_1,\tau_2)\}^2\{\sum_{i,t}\eta^\top X_{i(t-1)}\delta_{it}(\tau_2,\tau_3)\}^2] \leq [\mathsf{E}\{\sum_{i,t}\eta^\top X_{i(t-1)}\delta_{it}(\tau_1,\tau_2)\}^4]^{1/2}[\mathsf{E}\{\sum_{i,t}\eta^\top X_{i(t-1)}\delta_{it}(\tau_2,\tau_3)\}^4]^{1/2}$. Since it can be derived $\mathsf{E}\{\delta_{it}(\tau,\tau')|\mathcal{F}_{t-1}\} = 0$, then $\mathsf{E}[\{\eta^\top X_{i_1(t_1-1)}\delta_{i_1t_1}(\tau,\tau')\}\{\eta^\top X_{i_2(t_2-1)}\delta_{i_2t_2}(\tau,\tau')\}\{\eta^\top X_{i_3(t_3-1)} \delta_{i_3t_3}(\tau,\tau')\}\{\eta^\top X_{i_4(t_4-1)}\delta_{i_4t_4}(\tau,\tau')\}]$ is non-zero only if (a) $i_1 = i_2, t_1 = t_2$ and $i_3 = i_4 \neq i_1, t_3 = t_4 \neq t_1$ or (b) $i_1 = i_2 = i_3 = i_4$ and $t_1 = t_2 = t_3 = t_4$. It is straightforward to verify $(NT)^{-2}\mathsf{E}\{\sum_{i,t}\eta^\top X_{i(t-1)}\delta_{it}(\tau_1,\tau_2)\}^4 =$

$$(NT)^{-2}\Big[\sum_{i,t}\mathsf{E}\big\{(\eta^\top X_{i(t-1)})^2\delta_{it}^2(\tau_1,\tau_2)\big\}\Big]^2 + (NT)^{-2}\sum_{i,t}\mathsf{E}\big\{(\eta^\top X_{i(t-1)})^4\delta_{it}^4(\tau_1,\tau_2)\big\}.$$

By the proof of (d) in Lemma 1, we know that $\mathsf{E}(\eta^\top X_{it})^2 = \mathcal{O}(1)$ and $\mathsf{E}(\eta^\top X_{it})^4 = \mathcal{O}(1)$. Moreover, it can be verified $\mathsf{E}\{\delta_{it}^2(\tau_1,\tau_2)\} \leq \tau_2 - \tau_1$ and $\mathsf{E}\{\delta_{it}^4(\tau_1,\tau_2)\} \leq \tau_2 - \tau_1$. By combining the results together, we have

$$\mathsf{E}\Big[\big\{\eta^\top \xi_1(D_1)\big\}^2\big\{\eta^\top \xi_1(D_2)\big\}^2\Big] \leq C(\tau_2 - \tau_1)(\tau_3 - \tau_2) \leq C|\tau_3 - \tau_1|,$$

for some positive constant C. This completes the proof of Lemma 2. We then conclude that the $\xi_1(\tau)$ converge weakly to a $(q+3)$-dimensional Brownian bridge. Consequently, the Theorem 5 can be proved.

Appendix B.4: Mis-specification of A

Suppose Y_{it} is generated by the true adjacency matrix A. From the theoretical results in Section 3, it is shown that $\widehat{\theta}(\tau)$ is $\sqrt{NT}$-consistent. However, the consistency result might not hold when the adjacency matrix A is mis-specified to be $A^* = (a^*_{ij})$. Accordingly, let $W^* = (w^*_{ij})$ be the row-normalized A^* and $X^*_{it} = (1, Z_i^\top, n_i^{-1}\sum_{j=1}^N a^*_{ij}Y_{jt}, Y_{it})^\top \in \mathbb{R}^{q+3}$. The estimator is then given by

$$\widehat{\theta}^*(\tau) = \arg\min_\theta \sum_{i=1}^N \sum_{t=1}^T \rho_\tau\Big\{Y_{it} - X^{*\top}_{i(t-1)}\theta(\tau)\Big\}.$$

Define $\widehat{\Omega}^*_0 = (NT)^{-1}\sum_{i=1}^N\sum_{t=0}^{T-1} X^*_{it}X^{*\top}_{it}$ and $\widehat{\Omega}^*_1(\tau) = (NT)^{-1}\sum_{i=1}^N\sum_{t=0}^{T-1} f_{it}\big\{X^{\top*}_{it}\theta(\tau)\big\}$ $X^*_{it}X^{*\top}_{it}$ for $\tau \in (0,1)$. To perform the asymptotic analysis, the conditions associated with the misspecified coefficients are listed below.

(C3*) (Eigenvalue-bound) Let $\widehat{\Omega}^*_1(\tau) \to_p \Omega^*_1(\tau)$ as $\min\{N,T\} \to \infty$ for any $\tau \in (0,1)$, where $\Omega^*_1(\tau) \in \mathbb{R}^{N\times N}$ is a positive definite matrix. Moreover, there exists positive constants $0 < c_1 < c_2 < \infty$ such that $c_1 \le \lambda_{\min}(\Omega^*_1(\tau)) \le \lambda_{\max}(\Omega^*_1(\tau)) \le c_2$ for any $\tau \in (0,1)$.

(C4*) (Monotonicity) It is assumed that $X^{\top*}_{it}\theta(\tau)$ $(1 \le i \le N, 1 \le t \le T)$ is a monotone increasing function with respect to $\tau \in (0,1)$.

Then we have the following result.

COROLLARY 1. *Assume (C1), (C2), (C3*), (C4*). Let* $V^*_{it\tau} = Y_{it} - X^{*\top}_{i(t-1)}\theta(\tau)$. *In addition, define* $\delta(W^*, W) = \sum_{i,j}|w^*_{ij} - w_{ij}|$ *to be the total magnitude of misspecification of* W. *We then have*

$$\widehat{\theta}^*(\tau) - \theta(\tau) = -(NT)^{-1}\{\Omega^*_1(\tau)\}^{-1}\sum_{i,t} X^*_{i(t-1)}\psi_\tau(V^*_{it\tau}) + r^*_{NT}(\tau) \tag{B.21}$$

with $\sup_\tau \|r^*_{NT}(\tau)\| = o_p((NT)^{-1/2})$*. In addition, further assume* $\delta(W^*, W) = o(\sqrt{N/T})$*, then we have* $\widehat{\theta}^*(\tau) - \theta(\tau) = o_p((NT)^{-1/2})$.

Consequently, it shows that when the misspecification magnitude is under control $\delta(W^*, W) = o(\sqrt{N/T})$, the resulting estimator is still $\sqrt{NT}$-consistent. We give a proof as follows.

PROOF. The proof is similar to the proof of Theorem 4 and therefore is only briefly shown here. In the following we break the proof into two parts. In the first part we show the representation of (B.21). Next, we show that the consistency result for $\widehat{\theta}^*(\tau)$.

Part I. (Proof of (B.21)) Define $\Delta^*(\tau) = \widehat{\theta}^*(\tau) - \theta(\tau)$. Then we have

$$\rho_\tau\big(Y_{it} - X^{*\top}_{i(t-1)}\widehat{\theta}^*(\tau)\big) = \rho_\tau\big(V^*_{it\tau} - X^{*\top}_{i(t-1)}\Delta^*(\tau)\big), \tag{B.22}$$

where recall $V^*_{it\tau} = Y_{it} - X^{*\top}_{i(t-1)}\theta(\tau)$. Then the minimization of (B.22) is equivalent to minimizing for a fixed τ,

$$Z^*_{NT}(\Delta, \tau) = (NT)^{-1}\sum_{i=1}^N\sum_{t=1}^T \Big\{\rho_\tau\big(V^*_{it\tau} - X^{*\top}_{i(t-1)}\Delta\big) - \rho_\tau(V^*_{it\tau})\Big\},$$

One could verify that $\Delta^*(\tau) = \arg\min_\Delta Z^*_{NT}(\Delta, \tau)$. The objective function $Z^*_{NT}(\Delta, \tau)$ is a convex random function. Recall that $\psi_\tau(u) = \tau - I(u < 0)$. Let $\nu_{it} = \Delta^\top X_{it}$, and one could further write $Z^*_{NT}(\Delta, \tau)$ as $Z^*_{NT}(\Delta, \tau) =$

$$-(NT)^{-1}\sum_{i,t}\Big[\Delta^\top X^*_{i(t-1)}\psi_\tau(V^*_{it\tau}) + \int_0^{\nu_{i(t-1)}}\big\{\mathbf{1}(V^*_{it\tau} \le s) - \mathbf{1}(V^*_{it\tau} < 0)\big\}ds\Big]$$

$\stackrel{\text{def}}{=} \Delta^\top\xi^*_1 + \xi^*_2$. By Theorem 1 of Kato (2009), we have $\widehat{\Delta}^*(\tau) = -\{\Omega^*_1(\tau)\}^{-1}\xi^*_1 + r^*_{NT}(\tau)$ with $\sup_\tau \|r^*_{NT}(\tau)\| = o_p((NT)^{-1/2})$ if we further assume

(a) $\xi_2^* \to_p \Delta^\top \Omega_1^* \Delta$ with Ω_1^* is a positive definite matrix with uniformly bounded eigenvalues on B;

(b) $\xi_1^*(\tau)$ is a sequence of bounded stochastic processes.

Part II. (Consistency) Note that the asymptotic bias of $\widehat{\theta}^*(\tau)$ is given by $-\{\Omega_1^*(\tau)\}^{-1}$ $\mathsf{E}(\xi_1^*)$. We next calculate the amount of $\mathsf{E}(\xi_1^*)$.

Denote $\mathsf{E}(\cdot|\mathcal{F}_t) = \mathsf{E}_t(\cdot)$. It can be derived $\mathsf{E}\{\xi_1^*(\tau)\} = \mathsf{E}\{\mathsf{E}_{t-1}(\xi_1^*)\}$. Further we have

$$\mathsf{E}_{t-1}(\xi_1^*) = -(NT)^{-1} \sum_{i,t} X_{i(t-1)}^* \Big\{\tau - F_{it}\big(X_{i(t-1)}^{*\top}\theta(\tau)\big)\Big\}. \tag{B.23}$$

Define $\delta_{i(t-1)}^* = \{X_{i(t-1)}^* - X_{i(t-1)}\}^\top \theta(\tau) = (w_i^* - w_i)^\top \mathbb{Y}_{t-1}\beta(\tau)$. It can be derived

$$\begin{aligned} &F_{it}\big(X_{i(t-1)}^{*\top}\theta(\tau)\big) = F_{it}\big(X_{i(t-1)}^{\top}\theta(\tau) + \delta_{it}^*\big) \\ &= F_{it}\big(X_{i(t-1)}^{\top}\theta(\tau)\big) + \int_0^1 f_{it}\big(X_{i(t-1)}^{\top}\theta(\tau) + t\delta_{i(t-1)}^*\big)\delta_{i(t-1)}^* dt. \end{aligned}$$

Note that $F_{it}\big(X_{i(t-1)}^{\top}\theta(\tau)\big) = \tau$. By substituting into (B.23) one could obtain that $\mathsf{E}_{t-1}(\xi_1^*) = (NT)^{-1}\sum_{i,t} X_{i(t-1)}^* \{\int_0^1 f_{it}\big(X_{i(t-1)}^{\top}\theta(\tau) + t\delta_{i(t-1)}^*\big)dt\}\delta_{i(t-1)}^*$. Since the density function $f_{it} * (\cdot)$ is bounded, then for any $\eta \in \mathbb{R}^{q+3}$ we have $|\mathsf{E}\{\mathsf{E}_{t-1}(\eta^\top \xi_1^*)\}|_a \preccurlyeq c_1(NT)^{-1}\sum_{i,t}\mathsf{E}(|\eta^\top X_{i(t-1)}^*||w_i^* - w_i|_a^\top |\mathbb{Y}_{t-1}|_a)$, where c_1 is a finite positive constant. In addition, one could note $\mathsf{E}(|\eta^\top X_{i(t-1)}^*||\mathbb{Y}_{i(t-1)}|) \le c_2$ for a finite constant c_2. Therefore we have $(NT)^{-1}\mathsf{E}(|\eta^\top X_{i(t-1)}^*||w_i^* - w_i|_a^\top |\mathbb{Y}_{t-1}|_a) \le c_2 N^{-1}\delta(W^*, W)$. Since it is assumed $\delta(W^*, W) = o(\sqrt{N/T})$, then it can be concluded that $\widehat{\theta}^*(\tau)$ is still $\sqrt{NT}$-consistent.

□

APPENDIX C

In Appendix C, we conduct a number of numerical studies. Appendix C.1 and

C.2 present the simulation models, performance measurements, and numerical results. Appendix C.3 gives a model diagnosis procedure and applies it the real data example. Appendix C.4 conducts the sub-sample analysis of the real data.

Appendix C.1: Simulation Models

We consider three simulation settings in this subsection to illustrate the finite sample performance of the proposed NQAR model. The main difference lies in the generating mechanism of the network structure (i.e, A).

Before we state the details of the network structure specification, we first give the forms of the coefficient functions. For convenience, we denote $\beta_{j,it} = \beta_j(U_{it})$ for $0 \le j \le 2$ and $\gamma_{j,it} = \gamma_j(U_{it})$ for $1 \le j \le 5$ in this section. Following Koenker and Xiao (2006), we generate the random coefficients as follows,

$$\begin{aligned}\beta_{0,it} =& u_{it},\ \beta_{1,it} = 0.1\Phi(u_{it}),\ \beta_{2,it} = 0.4\{1+\exp(u_{it})\}^{-1}\exp(u_{it}),\\ \gamma_{1,it} =& 0.5\Phi(u_{it}),\ \gamma_{2,it} = 0.3\mathbf{G}(u_{it},1,2),\ \gamma_{3,it} = 0.2\mathbf{G}(u_{it},2,2),\\ &\gamma_{4,it} = 0.25\mathbf{G}(u_{it},23,2), \gamma_{5,it} = 0.2\mathbf{G}(u_{it},2,1),\end{aligned}$$

where u_{it}s are iid random variables, $\Phi(\cdot)$ is the standard normal cumulated distribution function (cdf), $\mathbf{G}(\cdot,a,b)$ is the Gamma cdf with shape parameter a and scale parameter b. We generate u_{it} either from (a) the standard normal distribution (i.e., $N(0,1)$) or from (b) the t-distribution with 5 degrees of freedom. It can be noted the U_{it} in (2.1) can be transformed as $U_{it} = F(u_{it})$, where $F(\cdot)$ is cdf of u_{it}. By this way, U_{it} will be assured to follow a uniform distribution. Given the random coefficients, we futher generate observations from the NQAR model (2.1). Next generate the nodal covariates $Z_i = (Z_{i1},\cdots,Z_{i5})^\top \in \mathbb{R}^5$ from a multivariate normal distribution $N(\mathbf{0},\Sigma_z)$, where

$\Sigma_z = (\sigma_{j_1 j_2})$ and $\sigma_{j_1 j_2} = 0.5^{|\gamma_1 - \gamma_2|}$. Then, we generate $\mathbb{Y}_t$s according to (2.1), where $\mathbb{Y}_0 = \mathbf{0}$. To check the finite sample performance of the proposed method, we adopt three kinds of adjacency matrix structures that are well-known in the literature. The details are given in the following.

Example 1. (Dyad Independence Model) Holland and Leinhardt (1981) introduce a Dyad Independence Model with a Dyad defined as $D_{ij} = (a_{ij}, a_{ji})$ for $1 \leq i < j \leq N$. It is assumed that D_{ij}s are independent. Specifically, we set the probability of mutually connected dyads to be $P\{D_{ij} = (1,1)\} = 20N^{-1}$ to ensure the network sparsity. Besides, set $P\{D_{ij} = (1,0)\} = P\{D_{ij} = (0,1)\} = 0.5N^{-0.8}$, which implies that the expected degree for each node is $\mathcal{O}(N^{0.2})$. Accordingly, we have $P(D_{ij} = (0,0)) = 1 - 20N^{-1} - N^{-0.8}$, which is close to 1 as $N \to \infty$. The simulated dyad independence network is visualized in the left panel of Figure 1.

Example 2. (Stochastic Block Model) The Stochastic Block Model (Wang and Wong, 1987; Nowicki and Snijders, 2001) has important applications in community detection (Zhao et al., 2012). To generate the block network structure, we follow Nowicki and Snijders (2001) to randomly assign each node a block label indexed from 1 to K, where $K \in \{5, 10, 20\}$. We then set $P(a_{ij} = 1) = 0.3N^{-0.3}$ if i and j are in the same block, and $P(a_{ij} = 1) = 0.3N^{-1}$. This indicates that the nodes within the same block have higher probability to connect with each other than between blocks. Lastly, the simulated stochastic block network is displayed in the middle panel of Figure 1, where a clear cluster effect can be visualized.

Example 3. (Power-law Distribution Network) According to Barabási and Albert (1999), it is a common phenomenon that the majority nodes in the network have small links, while a small amount of nodes have large number of links. The degrees of nodes could then be characterized by the power-law distribution. To generate the network

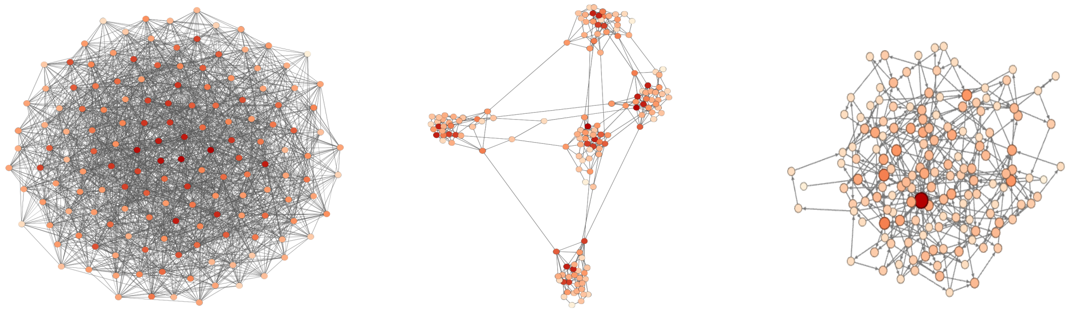

Figure 1: Left panel: dyad independence network; middle panel: stochastic block network; right panel: power-law distribution network. The larger and darker points imply higher in-degrees.

structure following this phenomenon, we simulate A as in Clauset et al. (2009). For each node, we generate the in-degree as $d_i = \sum_j a_{ji}$ according to the discrete power-law distribution as $P(d_i = k) = ck^{-\alpha}$, where c is a normalizing constant and the exponent parameter α is set to be $\alpha = 2.5$ by Clauset et al. (2009). Finally, for the ith node, we randomly select d_i nodes as its followers. The power-law distribution network structure is depicted in the right panel of Figure 1. It can be seen that only a limited number of nodes have high degrees.

Example 4. (Common Shareholder Network) To mimic the real data example, we consider the common shareholder network among stocks in the real data example in Section 6. The dataset contains $N = 2,442$ stocks traded in Shanghai Stock Exchange and the Shenzhen Stock Exchange. To construct the network structure, the top 10 shareholders' information is collected for each stock, which are referred to as major shareholders. Specifically, let $a_{ij} = 1$ if the two stock share at least one common major shareholder, otherwise $a_{ij} = 0$. The resulting network density is 3.9%. In the simulation study, we randomly sample a subset of stocks for the experiment, and the use their network relationships for the corresponding network structure.

Example 5. (Comparison With NAR) In this example, we compare the performance between the NAR model (Zhu et al., 2017) and the proposed QNAR model. Specifically, the data is generated by the NAR model with the true parameter fixed as $\beta = (0, 0.1, -0.2)^\top$ and $\gamma = (-0.5, 0.3, 0.8, 0, 0)^\top$. For each model, the innovation term ε_{it} is independently sampled from (a) a standard normal distribution $N(0, 1)$, and (b) t-distribution with 2 degrees. For the NQAR model, the median regression is fitted by fixing $\tau = 0.5$ to compare the estimation accuracy with the NAR model.

Example 6. (Mis-specification of A) We consider the misspecification of A in this example. Specifically, two possibly mis-specified patterns are evaluated, which is *partial misspecification* and *complete misspecification*. Let $n = \sum_{i,j} a_{ij}$ be the total number of edges. We first generate the true adjacency matrix A according to Example 1–3 respectively. We then construct the mis-specified $A^* = (a^*_{ij})$ in the following. For the partial misspecification, we first set $A^* = A$ and then randomly select $[0.1n]$ edges in the set $\{(i, j) : a_{ij} = 0\}$, and change them from 0 to 1. In this way, all the edges in the true adjacency matrix A are reserved but with few edges added. Second, for the complete misspecification, we randomly generate a new A^* according to the power-law distribution network. Therefore, the resulting mis-specified adjacency matrix A^* is non-related to the true adjacency matrix A.

Appendix C.2: Performance Measurements and Simulation Results

We consider different network sizes (i.e., $N = 100, 500, 1000$) and let $T = N/10$. For each case, the numerical performance is evaluated at $\tau = 0.1, 0.2, \cdots, 0.9$ respectively. The experiment is randomly replicated for $R = 1000$ times. Specifically, we use $\hat{\theta}^{(r)}(\tau) = \{\hat{\beta}_0^{(r)}(\tau), \hat{\beta}_1^{(r)}(\tau), \hat{\beta}_2^{(r)}(\tau), \hat{\gamma}^{(r)\top}(\tau)\}^\top$ to be the estimator from the rth replication. To evaluate the finite sample performance, the following measures are considered. Firstly the root mean square errors (RMSE) for $\beta_j(\tau)$s $(0 \le j \le 2)$ are calculated by

$\text{RMSE}_j(\tau) = \{R^{-1}\sum_{r=1}^{R}(\hat{\beta}_j^{(r)}(\tau) - \beta_j(\tau))^2\}^{1/2}$. Besides, for the nodal effect function vector γ, the RMSE is given by $\text{RMSE}_\gamma(\tau) = \{(5R)^{-1}\sum_r \|\hat{\gamma}^{(r)}(\tau) - \gamma(\tau)\|^2\}^{1/2}$. For Example 6, the estimation bias is further reported as $\text{Bias}_j(\tau) = R^{-1}\sum_{r=1}^{R}(\hat{\beta}_j^{(r)} - \beta_j(\tau))$ to compare the estimation accuracy. In addition, to compare the goodness-of-fit, we record $R^1(\tau)$ and $R^{1*}(\tau)$ by using A and A^* for model fitting. The average $\Delta R^1(\tau) = R^1(\tau) - R^{1*}(\tau)$ is calculate and reported. Secondly for each $\beta_j(\tau)$, a 95% confidence interval is constructed as $\text{CI}_j^{(r)}(\tau) = (\hat{\beta}_j^{(r)}(\tau) - z_{0.975}\widehat{\text{SE}}_j^{(r)}(\tau), \hat{\beta}_j^{(r)}(\tau) + z_{0.975}\widehat{\text{SE}}_j^{(r)}(\tau))$, where $\widehat{\text{SE}}_j^{(r)}(\tau)$ is the jth diagonal element of $(NT)^{-1}\tau(1-\tau)\widehat{\Sigma}_\theta(\tau)$, $\widehat{\Sigma}_\theta(\tau) = \widehat{\Omega}_1^{-1}\widehat{\Omega}_0\widehat{\Omega}_1^{-1}$, and z_α is the αth quantile of the standard normal distribution. Then, the coverage probability (CP) can be computed as $\text{CP}_j(\tau) = R^{-1}\sum_{m=1}^{R} I\{\beta_j(\tau) \in \text{CI}_j^{(r)}(\tau)\}$, where $I(\cdot)$ is the indicator function. Eventually the network density (ND) is given by $\{N(N-1)\}^{-1}\sum_{i_1,i_2} a_{i_1 i_2}$.

The detailed results of the simulation Example 1–4 are given from Table 1 to 4. It can be found that for a fixed τ the RMSE is decreased as N and T increased. For example, the RMSE of $\hat{\beta}_1(\tau)$ drops from 11.22×10^{-2} to 4.90×10^{-2} at $\tau = 0.1$ as N is increased from 100 to 500 in Example 1 for the t-distribution. It can also be noted that the RMSE for t-distribution of same network size N is slightly larger than standard normal distribution. Moreover, it can be concluded the computed coverage probabilities for $\beta_j(\tau)$s are stable at the nominal level 95%, which corroborates with the theoretical results. In addition, we plot the estimated $\hat{\beta}_j(\tau)$ with the 95% confidence interval against τ in Figure 2. A monotonic increasing pattern can be detected. Lastly, the network is becoming sparser as N increases (e.g. ND drops from 2.4% to 0.2% for the power-law distribution network from $N = 100$ to 1000).

The simulation result of Example 5–6 is given in Table 5, 6 and 7. First, it is found that compared to the maximum likelihood estimation of the NAR model, the estimator (3.1) of the QNAR model shows lower efficiency if the innovation is sampled from normal distribution. However, when the innovation term ε_{it} is sampled from a fat tail distribution (e.g., t-distribution), the QNAR model is able to produce more robust estimation results and overperforms NAR in terms of estimation efficiency. Next, for the model misspecification, it is observed that the bias of the complete misspecification is more severe than the partial misspecification (especially for the network effect $\beta_1(\tau)$). In addition, it is found that the difference of goodness-of-fit measure $\Delta R^1(\tau)$ is larger with the the complete misspecification. This suggests that the goodness-of-fit measure $R^1(\tau)$ is an informative tool to select different adjacency matrices.

Appendix C.3: Model Diagnosis

In this subsection, we conduct a model diagnosis for the NQAR model, and apply it to the real data example. Note that a direct residual analysis similar to the mean case of VAR is not feasible, as the error terms $V_{it\tau} = Y_{it} - X_{it}^\top \theta(\tau)$ are supposed to the correlated measuring by Pearson correlation. As an alternative, we follow Li et al. (2015) to use the QACF to measure the quantile correlation. The diagnosis procedures of the dynamic dependence and cross-sectional dependence are given as follows.

C.3.1. Dynamic Dependence. Assume the QNAR model is estimated with the fitted value for node i at time point t as $\widehat{Y}_{it\tau}$. Therefore, the residuals can be computed as $\widehat{V}_{it\tau} = Y_{it} - \widehat{Y}_{it\tau}$. Specifically, for each node i, we estimate the QACF according to Li et al. (2015) as follows

$$\hat{\rho}_{i\tau}^{(k)} = \frac{1}{\sqrt{(\tau - \tau^2)\widehat{\sigma}_v^2}} \cdot \frac{1}{T} \sum_{t=k}^{T} \psi_\tau(\widehat{V}_{it\tau})(\widehat{V}_{i(t-k)\tau} - \widehat{\mu}_v),$$

where $\widehat{\mu}_v$ and $\widehat{\sigma}_v^2$ are mean and variance estimates of $\widehat{V}_{it\tau}$. Note that the theoretical value $\rho_{i\tau}^{(k)} = 0$ as

$$\rho_{i\tau}^{(k)} = \mathsf{E}\{\psi_\tau(V_{it\tau})(V_{i(t-k)\tau} - \mu_v)\} - \mathsf{E}\{\psi_\tau(V_{it\tau})\}\mathsf{E}\{V_{i(t-k)\tau} - \mu_v\} = 0.$$

We plot all $\hat{\rho}_{i\tau}^{(k)}$s with $k = 1$ in a histogram in Figure 3. It can be visualized that the $\hat{\rho}_{i\tau}^{(k)}$s are around 0.

To test the significance of temporal dependence of residuals, we adopt a multiplier bootstrap procedure. We formulate the null hypothesis as $H_0 : \rho_{i\tau}^{(k)} = 0$ against the alternative $H_A : \rho_{i\tau}^{(k)} \neq 0$ Here we set $k = 1$ without of loss of generality. For each i, the multiplier bootstrap procedure is as follows.

***Bootstrap Steps.**

(1) Generate i.i.d standard normal random variables $\varepsilon_{it}^{[b]}$ for each bootstrap sample b.

(2) Compute $\widehat{\theta}^{[b]}(\tau)$ and $\widehat{V}_{it\tau}^{[b]}$ as

$$\widehat{\theta}^{[b]}(\tau) = \hat{\theta}(\tau) + (NT)^{-1}\widehat{\Omega}_1(\tau)^{-1}\sum_{i,t}\varepsilon_{it}^{[b]}X_{it}\psi_\tau(\widehat{V}_{it\tau})$$

$$\widehat{V}_{it\tau}^{[b]} = Y_{it} - X_{it}^\top\widehat{\theta}^{[b]}(\tau)$$

(3) Compute $\rho_{ik\tau}^{[b]}$ as

$$\widehat{\rho}_{i\tau}^{(k)[b]} = \frac{1}{\sqrt{(\tau - \tau^2)}} \cdot \frac{1}{T-k}\sum_{t=k}^{T}\left\{\psi_\tau(\widehat{V}_{it\tau}^{[b]})\frac{\widehat{V}_{i(t-k)\tau}^{[b]} - \widehat{\mu}_v^{[b]}}{\widehat{\sigma}_v^{[b]}} - \psi_\tau(\widehat{V}_{it\tau})\frac{\widehat{V}_{i(t-k)\tau} - \widehat{\mu}_v}{\widehat{\sigma}_v}\right\},$$

where $\widehat{\mu}_v^{[b]}$ and $(\widehat{\sigma}_v^{[b]})^2$ are mean and variance estimates of $\widehat{V}_{it\tau}^{[b]}$ for $t = k, \cdots, T$.

(4) For the significance level α, calculate the $\alpha/2$ and $1 - \alpha/2$ quantile of $\hat{\rho}_{i\tau}^{(k)[b]}$ to

produce the $(1-\alpha)$-confidence interval. Check whether the confidence interval contains 0.

The procedure is conducted for $i = 1, \cdots, N$. For the stock data with $N = 2442$, it is found that all the 95% confidence intervals cover 0. This illustrates a good fitness level of the NQAR model.

C.3.2. Cross-sectional Dependence. Similar to the test of the dynamic dependence of $V_{it\tau}$, we could define the cross sectional quantile correlation as

$$\hat{\rho}_{ij\tau} = \frac{1}{\sqrt{(\tau - \tau^2)\widehat{\sigma}_j^2}} \cdot \frac{1}{T}\sum_{t=1}^{T} \psi_\tau(\widehat{V}_{it\tau})(\widehat{V}_{jt\tau} - \widehat{\mu}_j),$$

where $\widehat{\mu}_j$ and $\widehat{\sigma}_j^2$ are mean and variance estimates of $\widehat{V}_{jt\tau}$. By similar bootstrap procedure, this leads to $(N^2 - N)/2$ test statistics. It is found that the 99.65% of all the 95% confidence intervals cover 0. This illustrates that the cross-sectional dependence is almost ignorable after fitting the NQAR model.

Lastly, to elaborate on the above test procedure, a multiple testing procedure can be developed and this is beyond the scope of the article. We leave it as a future research topic.

Appendix C.4: Sub-sample Analysis of the Real Data

In this section, we explore the empirical data performance when the NQAR model applied to sub-samples. Specifically, we have splitted the data in 2013 into the first half of the year and the second half of the year. Note that the returns are more volatile in the first half (as shown in the left panel of Figure 1). The results are shown in Table 8 and Table 9, which show different phenomenons. The network effect $\beta_1(\tau)$

is still significantly positive for the first half at $\tau = 0.95$; while for the second half the phenomenon does not exist. The empirical analysis suggests that one tends to see stronger asymmetric network effects when the market exhibits high turbulence.

Table 1: Simulation results for dyad independence network with 1000 replications. The random variable u_{it} is generated from standard normal distribution (i.e., Z) and t-distribution with 5 degrees of freedom (i.e., T). The RMSE ($\times 10^{-2}$) and the coverage probability (%) for the 95% confidence interval are reported for β_0 to β_1. The RMSE is also reported for γ. Lastly, the network density is computed and given.

N	Dist.	β_0	β_1	β_2	γ	ND
			$\tau = 0.1$			
100	Z	2.60(95.0)	10.10(95.8)	2.47(94.3)	3.09	22.7
	T	3.43(96.4)	11.22(95.2)	2.37(95.6)	4.17	
500	Z	1.08(96.2)	4.61(95.4)	1.04(96.0)	1.32	4.7
	T	1.51(95.4)	4.90(95.9)	1.03(96.1)	1.82	
1000	Z	0.77(95.8)	3.29(95.0)	0.80(94.0)	0.93	2.4
	T	1.06(95.8)	3.66(95.0)	0.75(95.0)	1.29	
			$\tau = 0.5$			
100	Z	1.90(95.5)	6.62(95.4)	1.65(96.7)	2.11	22.7
	T	1.99(95.7)	5.67(94.5)	1.32(93.3)	2.15	
500	Z	0.84(94.4)	2.99(95.5)	0.79(94.9)	0.87	4.7
	T	0.90(94.9)	2.43(96.2)	0.55(92.3)	0.91	
1000	Z	0.59(94.7)	2.17(95.0)	0.53(95.7)	0.63	2.4
	T	0.62(94.2)	1.77(95.0)	0.37(93.5)	0.66	
			$\tau = 0.9$			
100	Z	2.57(95.3)	9.96(95.1)	2.49(94.1)	2.92	22.7
	T	3.61(95.0)	10.61(95.4)	2.41(94.5)	3.98	
500	Z	1.08(96.3)	4.27(95.8)	1.10(94.0)	1.30	4.7
	T	1.53(95.6)	4.75(94.8)	1.11(93.9)	1.75	
1000	Z	0.78(95.5)	3.14(95.5)	0.76(95.0)	0.90	2.4
	T	1.09(95.9)	3.41(96.0)	0.84(93.5)	1.26	

Table 2: Simulation results for stochastic block network with 1000 replications. The random variable u_{it} is generated from standard normal distribution (i.e., Z) and t-distribution with 5 degrees of freedom (i.e., T). The RMSE ($\times 10^{-2}$) and the Coverage Probability (%) for the 95% confidence interval are reported for β_0 to β_1. The RMSE is also reported for γ. Lastly, the network density is computed and given.

N	Dist.	β_0	β_1	β_2	γ	ND
			$\tau = 0.1$			
100	Z	2.61(95.8)	3.29(94.9)	2.45(94.3)	3.03	2.6
	T	3.33(96.7)	3.37(96.0)	2.40(94.2)	4.29	
500	Z	1.14(94.3)	1.40(94.5)	1.08(94.9)	1.32	0.5
	T	1.57(94.0)	1.50(95.1)	1.04(95.6)	1.82	
1000	Z	0.79(94.6)	0.89(95.0)	0.74(95.9)	0.94	0.2
	T	1.09(95.4)	0.95(94.9)	0.78(94.5)	1.28	
			$\tau = 0.5$			
100	Z	1.88(94.5)	2.15(94.2)	1.74(95.2)	2.07	2.6
	T	2.03(94.0)	1.76(95.1)	1.28(93.4)	2.17	
500	Z	0.84(94.5)	0.92(94.5)	0.77(94.9)	0.90	0.5
	T	0.86(94.7)	0.75(94.5)	0.52(93.2)	0.90	
1000	Z	0.59(94.4)	0.59(95.9)	0.53(95.6)	0.63	0.2
	T	0.61(95.4)	0.47(95.6)	0.38(93.0)	0.64	
			$\tau = 0.9$			
100	Z	2.56(95.0)	2.91(96.0)	2.46(94.5)	2.94	2.6
	T	3.44(95.8)	3.28(94.3)	2.39(94.3)	4.07	
500	Z	1.08(95.4)	1.33(94.6)	1.07(95.3)	1.29	0.5
	T	1.52(95.9)	1.45(95.8)	1.12(94.0)	1.78	
1000	Z	0.80(95.2)	0.89(94.4)	0.75(96.0)	0.91	0.2
	T	1.03(96.4)	0.90(95.3)	0.82(93.4)	1.23	

Table 3: Simulation results for power-law distribution network with 1000 replications. The random variable u_{it} is generated from standard normal distribution (i.e., Z) and t-distribution with 5 degrees of freedom (i.e., T). The RMSE ($\times 10^{-2}$) and the coverage probability (%) for the 95% confidence interval are reported for β_0 to β_1. The RMSE is also reported for γ. Lastly, the network density is computed and given.

N	Dist.	β_0	β_1	β_2	γ	ND
			$\tau = 0.1$			
100	Z	2.44(95.9)	2.95(95.4)	2.32(96.2)	3.08	2.4
	T	3.45(96.3)	3.28(93.9)	2.36(95.1)	4.19	
500	Z	1.09(95.5)	1.24(96.3)	1.07(95.4)	1.35	0.5
	T	1.53(94.7)	1.42(94.8)	1.04(96.2)	1.79	
1000	Z	0.76(95.8)	0.91(95.6)	0.77(94.7)	0.94	0.2
	T	1.06(95.0)	0.99(95.5)	0.75(95.3)	1.28	
			$\tau = 0.5$			
100	Z	1.87(95.3)	1.96(95.7)	1.79(94.4)	2.07	2.4
	T	1.94(96.4)	1.55(96.2)	1.29(93.1)	2.15	
500	Z	0.82(95.7)	0.85(94.6)	0.77(95.8)	0.89	0.5
	T	0.90(95.5)	0.71(94.0)	0.54(93.3)	0.92	
1000	Z	0.58(95.1)	0.62(94.2)	0.54(96.0)	0.62	0.2
	T	0.63(94.4)	0.51(92.2)	0.37(94.6)	0.64	
			$\tau = 0.9$			
100	Z	2.55(95.8)	2.94(93.5)	2.43(94.3)	2.91	2.4
	T	3.53(95.3)	3.01(94.7)	2.43(94.1)	4.11	
500	Z	1.12(95.1)	1.20(96.3)	1.09(95.1)	1.29	0.5
	T	1.51(95.5)	1.33(95.1)	1.10(94.3)	1.80	
1000	Z	0.79(95.2)	0.87(95.7)	0.76(95.1)	0.90	0.2
	T	1.09(94.7)	0.98(95.3)	0.83(92.1)	1.26	

Table 4: Simulation results for using the real data network with 1000 replications. The random variable u_{it} is generated from standard normal distribution (i.e., Z) and t-distribution with 5 degrees of freedom (i.e., T). The RMSE ($\times 10^{-2}$) and the coverage probability (%) for the 95% confidence interval are reported for β_0 to β_1. The RMSE is also reported for γ. Lastly, the network density is computed and given.

N	Dist.	β_0	β_1	β_2	γ	ND
			$\tau = 0.1$			
100	Z	5.35(94.0)	10.92(89.0)	5.66(93.0)	7.04	4.6
	T	7.62(96.0)	11.48(89.0)	5.62(97.0)	9.20	
500	Z	1.18(93.0)	2.82(91.0)	1.12(95.0)	1.45	4.0
	T	1.44(96.0)	2.88(95.0)	1.05(98.0)	1.75	
1000	Z	0.54(97.0)	1.52(96.0)	0.49(97.0)	0.66	3.9
	T	0.73(92.0)	1.69(93.0)	0.49(96.0)	0.87	
			$\tau = 0.5$			
100	Z	3.76(97.0)	5.98(94.0)	4.19(96.0)	4.35	4.6
	T	4.02(98.0)	5.28(95.0)	3.31(92.0)	4.95	
500	Z	0.88(94.0)	1.64(96.0)	0.77(96.0)	0.85	4.0
	T	0.84(97.0)	1.47(94.0)	0.49(96.0)	0.88	
1000	Z	0.42(96.0)	0.91(93.0)	0.40(93.0)	0.45	3.9
	T	0.38(96.0)	0.85(93.0)	0.25(92.0)	0.47	
			$\tau = 0.9$			
100	Z	5.68(95.0)	8.75(96.0)	6.09(91.0)	6.83	4.6
	T	8.13(95.0)	8.86(95.0)	5.99(95.0)	9.39	
500	Z	1.15(97.0)	2.46(96.0)	1.00(97.0)	1.24	4.0
	T	1.27(98.0)	2.48(96.0)	1.02(94.0)	1.77	
1000	Z	0.57(93.0)	1.40(95.0)	0.58(93.0)	0.66	3.9
	T	0.67(98.0)	1.39(95.0)	0.63(90.0)	0.88	

Table 5: Simulation results with 500 replications for comparison between the NAR model and QNAR model. For QNAR model, the estimation is conducted at $\tau = 0.5$. The RMSEs ($\times 10^2$) of β and γ are reported.

N	Est.	β_0	β_1	β_2	γ_1	γ_2	γ_3	γ_4	γ_5
Case 1: Normal Distribution									
200	NAR	1.27	1.29	1.22	1.65	1.71	1.97	1.65	1.50
	QNAR	1.62	1.61	1.51	2.04	2.12	2.43	2.15	1.92
500	NAR	0.79	0.90	0.85	0.97	1.16	1.20	1.08	0.96
	QNAR	0.98	1.11	1.03	1.23	1.39	1.52	1.35	1.18
1000	NAR	0.62	0.61	0.55	0.70	0.74	0.80	0.76	0.68
	QNAR	0.76	0.74	0.69	0.86	0.94	0.99	0.89	0.83
Case 2: t-Distribution									
200	NAR	3.02	1.67	1.28	3.18	3.68	3.65	3.63	3.17
	QNAR	1.79	1.08	0.85	2.16	2.30	2.39	2.24	2.09
500	NAR	1.78	1.00	0.85	2.11	2.29	2.31	2.34	1.94
	QNAR	1.15	0.66	0.56	1.34	1.54	1.44	1.47	1.26
1000	NAR	1.26	0.73	0.58	1.53	1.59	1.63	1.55	1.46
	QNAR	0.80	0.46	0.40	0.93	0.99	1.08	1.00	0.93

Table 6: Simulation results for partially mis-specified A with power-law distribution network with 500 replications. The random variable u_{it} is generated from standard normal distribution (i.e., Z) and t-distribution with 5 degrees of freedom (i.e., T). The Bias ($\times 10^2$) and the RMSE ($\times 10^2$) are reported for β_0 to β_1. The RMSE is also reported for γ. Lastly, the average difference of the goodness-of-fit measures $\Delta R^1(\tau)$ is also reported.

N	Dist.	β_0	β_1	β_2	γ	$\Delta R^1(\tau)$
			$\tau = 0.1$			
100	Z	0.15(5.4)	0.12(7.4)	-0.62(5.8)	7.04	0.01
	T	-0.24(7.5)	-0.64(8.3)	-0.21(5.5)	9.38	-0.01
500	Z	0.02(1.1)	0.03(1.3)	-0.04(1.1)	1.30	0.00
	T	-0.08(1.5)	-0.01(1.5)	0.21(1.1)	1.81	0.00
1000	Z	0.01(0.5)	0.03(0.7)	-0.01(0.5)	0.66	0.00
	T	-0.03(0.8)	-0.01(0.8)	0.22(0.6)	0.88	0.00
			$\tau = 0.5$			
100	Z	-0.08(4.2)	0.15(5.1)	-0.54(4.1)	4.85	0.00
	T	0.17(4.4)	0.04(4.2)	-0.11(3.1)	4.90	0.01
500	Z	0.01(0.8)	0.06(0.9)	-0.03(0.8)	0.87	0.01
	T	-0.15(0.8)	0.01(0.7)	0.05(0.5)	0.93	0.01
1000	Z	-0.04(0.4)	0.00(0.5)	0.02(0.4)	0.46	0.01
	T	-0.07(0.4)	0.04(0.4)	0.03(0.3)	0.46	0.01
			$\tau = 0.9$			
100	Z	-0.04(5.7)	-0.66(7.0)	-0.86(6.0)	6.56	0.01
	T	0.19(7.6)	-1.01(7.6)	-1.59(5.5)	9.18	0.02
500	Z	0.05(1.1)	0.03(1.4)	0.00(1.1)	1.25	0.02
	T	0.00(1.5)	-0.04(1.4)	-0.38(1.2)	1.83	0.02
1000	Z	-0.01(0.5)	-0.04(0.7)	0.01(0.5)	0.64	0.02
	T	0.15(0.8)	-0.06(0.7)	-0.34(0.6)	0.87	0.02

Table 7: Simulation results for completely mis-specified A with power-law distribution network with 500 replications. The random variable u_{it} is generated from standard normal distribution (i.e., Z) and t-distribution with 5 degrees of freedom (i.e., T). The Bias ($\times 10^2$) and the RMSE ($\times 10^2$) are reported for β_0 to β_1. The RMSE is also reported for γ. Lastly, the average difference of the goodness-of-fit measures $\Delta R^1(\tau)$ is also reported.

N	Dist.	β_0	β_1	β_2	γ	$\Delta R^1(\tau)$
			$\tau = 0.1$			
100	Z	0.08(5.7)	-1.33(7.0)	-0.07(5.5)	7.14	0.01
	T	-0.25(7.6)	-1.01(7.5)	-0.07(5.8)	9.34	-0.01
500	Z	-0.02(1.1)	-1.01(1.6)	0.03(1.1)	1.32	0.00
	T	-0.02(1.5)	-0.69(1.6)	0.18(1.1)	1.80	0.00
1000	Z	-0.02(0.5)	-0.98(1.2)	-0.03(0.6)	0.66	0.00
	T	-0.07(0.7)	-0.76(1.0)	0.20(0.6)	0.90	0.00
			$\tau = 0.5$			
100	Z	-0.25(4.4)	-4.99(6.9)	-0.37(4.2)	4.65	0.09
	T	-0.27(4.8)	-5.05(6.4)	-0.16(3.3)	4.83	0.09
500	Z	-0.21(0.9)	-5.05(5.1)	0.01(0.8)	0.88	0.09
	T	-0.40(1.0)	-4.95(5.0)	0.11(0.6)	0.97	0.11
1000	Z	-0.17(0.5)	-4.99(5.0)	0.07(0.4)	0.48	0.09
	T	-0.32(0.6)	-4.98(5.0)	0.09(0.3)	0.51	0.11
			$\tau = 0.9$			
100	Z	0.18(6.1)	-8.99(11.1)	-0.69(5.8)	6.49	0.20
	T	1.12(8.0)	-9.24(11.8)	-1.31(5.9)	9.11	0.18
500	Z	0.27(1.3)	-9.01(9.1)	-0.02(1.1)	1.29	0.23
	T	0.36(1.7)	-9.23(9.3)	-0.33(1.2)	1.71	0.22
1000	Z	0.32(0.7)	-8.94(9.0)	0.06(0.5)	0.66	0.24
	T	0.55(1.1)	-9.23(9.3)	-0.28(0.6)	0.90	0.23

Table 8: The detailed NQAR and NAR analysis results of the first half year of 2013 for the Chinese Stock dataset (τ =0.05, 0.5, 0.95). The parameter estimates ($\times 10^2$) are reported for $\tau = 0.05, 0.5, 0.95$, where the standard error ($\times 10^2$) is given in parentheses. The sign of the estimates are reported instead if the absolute value of parameter estimate are less than 10^{-4}. The p-values are also reported.

	$\tau = 0.05$		$\tau = 0.5$		$\tau = 0.95$	
	Estimate	*p*-value	Estimate	*p*-value	Estimate	*p*-value
$\hat{\beta}_0$	0.05 (0.00)	< 0.01	0.96 (0.06)	< 0.01	2.75 (0.19)	< 0.01
$\hat{\beta}_1$	+ (0.04)	1.00	0.84 (1.07)	0.43	7.63 (2.69)	< 0.01
$\hat{\beta}_2$	4.28 (0.22)	< 0.01	37.10 (0.74)	< 0.01	69.54 (1.44)	< 0.01
SIZE	- (0.02)	0.97	-1.05 (0.13)	< 0.01	-3.27 (0.40)	< 0.01
BM	- (0.02)	0.99	-0.46 (0.05)	< 0.01	-0.93 (0.33)	< 0.01
PR	- (0.01)	0.97	-0.09 (0.17)	0.60	0.70 (0.54)	0.19
AR	-0.02 (0.03)	0.39	-0.69 (0.01)	< 0.01	-1.55 (0.04)	< 0.01
CASH	0.03 (0.04)	0.39	0.05 (0.09)	0.53	-0.07 (0.46)	0.87
LEV	- (0.03)	0.97	-1.03 (0.10)	< 0.01	-1.57 (0.55)	< 0.01

Table 9: The detailed NQAR and NAR analysis results of the second half year of 2013 for the Chinese Stock dataset (τ =0.05, 0.5, 0.95). The parameter estimates ($\times 10^{-2}$) are reported for $\tau = 0.05, 0.5, 0.95$, where the standard error ($\times 10^{-2}$) is given in parentheses. The sign of the estimates are reported instead if the absolute value of parameter estimate are less than 10^{-4}. The p-values are also reported.

	$\tau = 0.05$		$\tau = 0.5$		$\tau = 0.95$	
	Estimate	*p*-value	Estimate	*p*-value	Estimate	*p*-value
$\hat{\beta}_0$	0.06 (0.00)	< 0.01	1.09 (0.05)	< 0.01	3.12 (0.09)	< 0.01
$\hat{\beta}_1$	+ (0.03)	0.99	-3.77 (1.10)	< 0.01	-3.60 (3.79)	0.34
$\hat{\beta}_2$	3.80 (0.19)	< 0.01	33.44 (0.58)	< 0.01	63.71 (1.78)	< 0.01
SIZE	- (0.01)	0.97	-0.99 (0.12)	< 0.01	-5.17 (0.41)	< 0.01
BM	- (0.01)	0.99	-0.18 (0.07)	< 0.01	-0.34 (0.23)	0.14
PR	- (0.00)	0.99	-0.61 (0.14)	< 0.01	0.45 (0.16)	< 0.01
AR	-0.02 (0.10)	0.87	0.57 (0.47)	0.22	-0.01 (0.08)	0.89
CASH	-0.04 (0.04)	0.32	-0.18 (0.05)	< 0.01	-0.06 (0.32)	0.84
LEV	- (0.00)	0.82	-0.74 (0.02)	< 0.01	-2.81 (0.18)	< 0.01

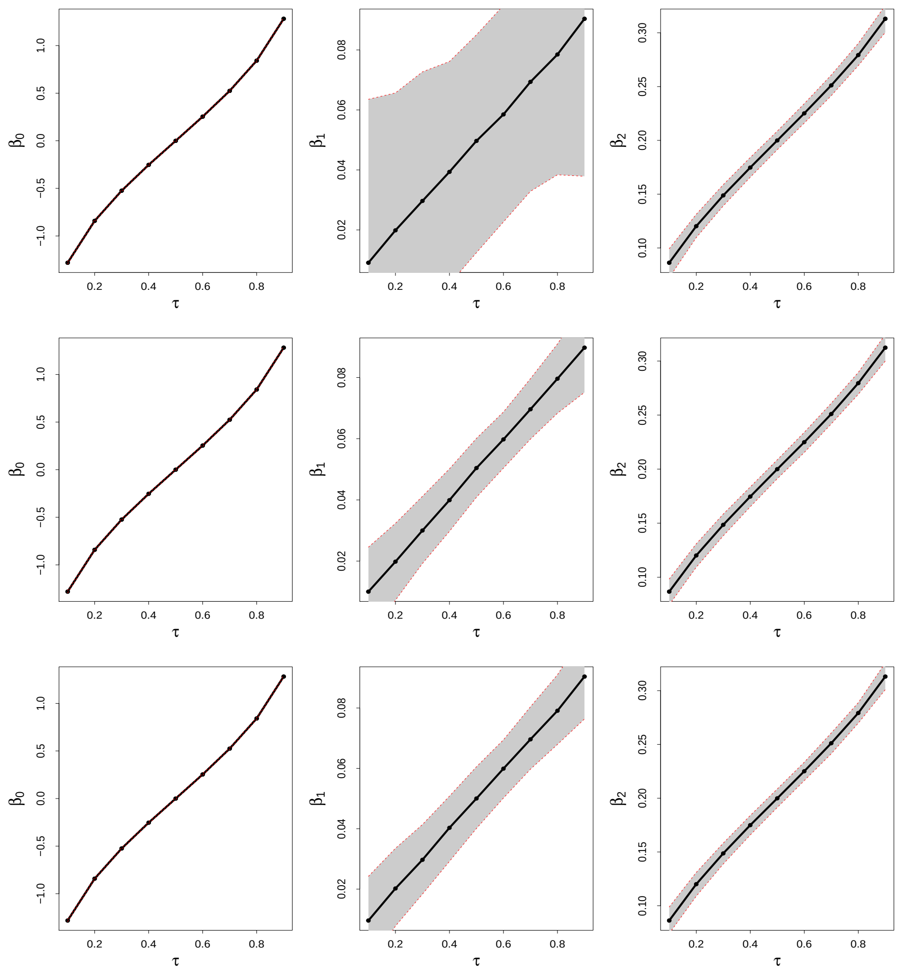

Figure 2: The estimated β_0 to β_2 against τ for three different network structures. The black line is the average estimated value over 1,000 replications, and the grey area is the empirical 95% confidence band. The top panel: dyad independence network; The middle panel: stochastic block network; the bottom panel: power-law distribution network.

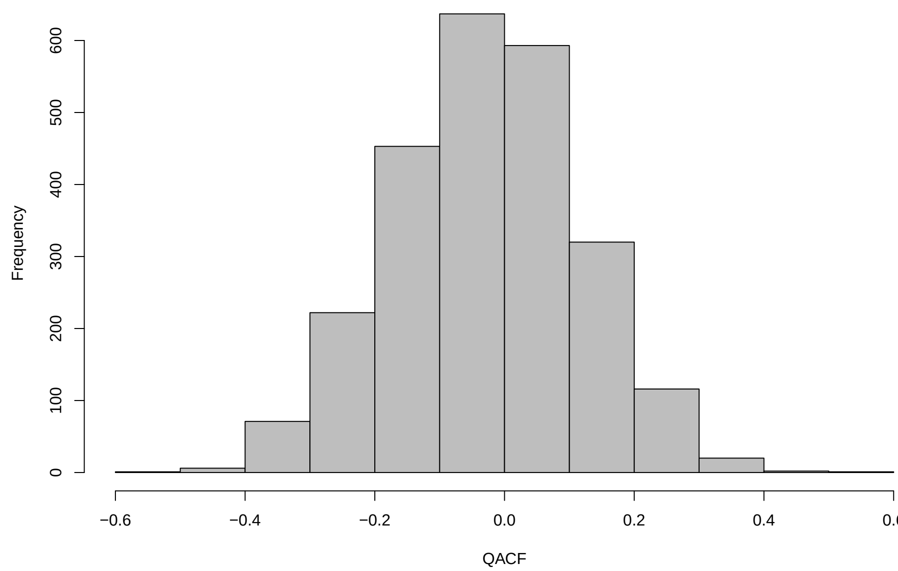

Figure 3: The histogram of QACF ($\hat{\rho}_{i\tau}$) of all nodes.

References

Andrews, D. W. (1988). Laws of large numbers for dependent non-identically distributed random variables. Econometric theory, 4(03):458–467.

Banerjee, S., Carlin, B. P., and Gelfand, A. E. (2014). Hierarchical modeling and analysis for spatial data. Crc Press.

Barabási, A.-L. and Albert, R. (1999). Emergence of scaling in random networks. Science, 286(5439):509–512.

Bardet, J.-M., Doukhan, P., Lang, G., and Ragache, N. (2008). Dependent lindeberg central limit theorem and some applications. ESAIM: Probability and Statistics, 12:154–172.

Billingsley, P. (1968). Convergence of Probability Measures. Wiley, New York.

Chernozhukov, V., Chetverikov, D., and Kato, K. (2015). Comparison and anticoncentration bounds for maxima of gaussian random vectors. Probability Theory and Related Fields, 162(1-2):47–70.

Clauset, A., Shalizi, C. R., and Newman, M. E. (2009). Power-law distributions in empirical data. SIAM review, 51(4):661–703.

Gallant, A. R. (2009). Nonlinear statistical models, volume 310. John Wiley & Sons.

Hall, P. and Heyde, C. C. (2014). Martingale limit theory and its application. Academic press.

Holland, P. W. and Leinhardt, S. (1981). An exponential family of probability distributions for directed graphs. Journal of the american Statistical association, 76(373):33–50.

Kato, K. (2009). Asymptotics for argmin processes: Convexity arguments. Journal of Multivariate Analysis, 100(8):1816–1829.

Koenker, R. and Xiao, Z. (2006). Quantile autoregression. Journal of the American Statistical Association, 101(475):980–990.

Li, G., Li, Y., and Tsai, C.-L. (2015). Quantile correlations and quantile autoregressive modeling. Journal of the American Statistical Association, 110(509):246–261.

Nowicki, K. and Snijders, T. A. B. (2001). Estimation and prediction for stochastic blockstructures. Journal of the American Statistical Association, 96(455):1077–1087.

Wagener, J., Volgushev, S., and Dette, H. (2012). The quantile process under random censoring. Mathematical Methods of Statistics, 21(2):127–141.

Wang, Y. J. and Wong, G. Y. (1987). Stochastic blockmodels for directed graphs. Journal of the American Statistical Association, 82(397):8–19.

Zhang, D. and Wu, W. B. (2015). Gaussian approximation for high dimensional time series. arXiv preprint arXiv:1508.07036.

Zhang, X. and Cheng, G. (2014). Bootstrapping high dimensional time series. arXiv preprint arXiv:1406.1037.

Zhao, Y., Levina, E., Zhu, J., et al. (2012). Consistency of community detection in networks under degree-corrected stochastic block models. The Annals of Statistics, 40(4):2266–2292.

Zhu, X., Pan, R., Li, G., Liu, Y., and Wang, H. (2017). Network vector autoregression. Annals of Statistics, 45(3):1096–1123.