Statistica Sinica: Supplement

Mutual Influence Regression Model

Xinyan Fan, Wei Lan, Tao Zou and Chih-Ling Tsai

Renmin University of China, Southwestern University of Finance and Economics,

The Australian National University and University of California, Davis

Supplementary Material

This supplementary material consists of four sections. Section S1 introduces seven conditions and four useful Lemmas. Section S2 demonstrates the proofs of Lemmas 3 and 4, and Section S3 provides the proofs of Theorems 1–4. Section S4 presents additional simulation results.

S1 Technical Conditions and Useful Lemmas

To study the asymptotic properties of parameter estimators and test statistics, we introduce the following eleven technical conditions and four useful lemmas. For the sake of convenience, we let $\varrho_{\min}(A)$ and $\varrho_{\max}(A)$ denote the smallest and largest eigenvalues of any arbitrary matrix A. In addition, let $\|G\| = \{\varrho_{\max}(G^\top G)\}^{1/2}$ be the L_2 (i.e., spectral) norm, $\|G\|_F = \{\text{tr}(G^\top G)\}^{1/2}$ be the Frobenius norm, and $\|G\|_R = \max_i \sum_j |G_{ij}|$ be the maximum absolute row-sum norm for any generic matrix $G = (G_{ij})$. For any candidate model $\mathcal{S} \subseteq \{1, \cdots, d\}$, let $|\mathcal{S}|$ be the size of $\mathcal{S}$.

(C1) Assume that $\epsilon_t = (\epsilon_{1t}, \cdots, \epsilon_{nt})^\top \in \mathbb{R}^n$ is independent and identically generated with mean 0 and covariance matrix $\sigma_0^2 I_n$ for $t = 1, \cdots, T$, and assume that $\epsilon_{1t}, \cdots, \epsilon_{nt}$ are independent and identically random variables satisfying $\sup_{\psi \geq 1} \psi^{-1/2}\{E(|\epsilon_{it}|^\psi)\}^{1/\psi} < \infty$ for $i = 1, \cdots, n$ and $t = 1, \cdots, T$.

(C2) The parameter space of regression coefficients is $\Lambda = \big\{\lambda = (\lambda_1, \cdots, \lambda_d)^\top : \sum_{k=1}^d |\lambda_k| < 1 - \varsigma, \text{for some } \varsigma \in (0,1)\big\}$, and the true parameter is $\lambda_0 \in \Lambda$.

(C3) For similarity matrices in $\big\{W_k^{(t)} \in \mathbb{R}^{n \times n} : k = 1, \cdots, d\big\}$ and λ in a small neighborhood of λ_0, there exists $C_w > 0$ such that $\sup_{t \leq T, n \geq 1}(\|W_k^{(t)}\|_R + \|W_k^{(t)\top}\|_R + \|\Delta_t^{-1}(\lambda)\|_R + \|\{\Delta_t^\top(\lambda)\}^{-1}\|_R) \leq C_w < \infty$.

(C4) Assume that $\mathcal{I}_{nT}(\theta_0) \to \mathcal{I}(\theta_0)$ in Frobenius norm and $\mathcal{J}_{nT}(\theta_0) \to \mathcal{J}(\theta_0)$ in L_2 norm as $nT \to \infty$, where $\mathcal{I}_{nT}(\theta_0) = -(nT)^{-1}E(\frac{\partial^2 \ell(\theta_0)}{\partial\theta\partial\theta^\top})$, $\mathcal{J}_{nT}(\theta_0) = (nT)^{-1}\mathrm{Var}(\frac{\partial \ell(\theta_0)}{\partial\theta})$, and $\mathcal{I}(\theta_0)$ and $\mathcal{J}(\theta_0)$ are positive definite matrices. In addition, assume that, for θ in a small neighborhood of θ_0, there exist two finite positive constants $c_{\min,1}$ and $c_{\min,2}$ such that $0 < c_{\min,1} < \varrho_{\min}(\mathcal{I}(\theta)) \leq \varrho_{\max}(\mathcal{I}(\theta)) = O(d)$ and $0 < c_{\min,2} < \varrho_{\min}(\mathcal{J}(\theta)) \leq \varrho_{\max}(\mathcal{J}(\theta)) = O(d)$.

(C5) Assume that $d = o\{(nT)^{1/4}\}$ as $nT \to \infty$, and $|\mathcal{S}_T|$ is finite.

(C6) Assume that $\min_{k\in\mathcal{S}_T}|\lambda_{0k}|\sqrt{nT/\log(nT)}\to\infty$ as $nT\to\infty$.

(C7) There exist finite positive constants $c_{\min,3}$, $c_{\max,3}$, δ and q such that, for sufficiently large nT,

$$c_{\min,3}<\varrho_{\min}\left(\frac{1}{nT}\frac{\partial\ell(\tilde{\theta}_{\mathcal{S}})}{\partial\theta_{\mathcal{S}}\partial\theta_{\mathcal{S}}^{\top}}\right)\leq\varrho_{\max}\left(\frac{1}{nT}\frac{\partial\ell(\tilde{\theta}_{\mathcal{S}})}{\partial\theta_{\mathcal{S}}\partial\theta_{\mathcal{S}}^{\top}}\right)<c_{\max,3},$$

where $\mathcal{S}$ and $\tilde{\theta}_{\mathcal{S}}$ satisfy $|\mathcal{S}|\leq q$ for $q>|\mathcal{S}_T|$ and $\|\tilde{\theta}_{\mathcal{S}}-\theta_{0\mathcal{S}}\|\leq\delta$.

The above conditions are mild and sensible. Condition (C1) is a moment condition, which is much weaker than commonly used distribution assumptions; see, for example, the normal assumption in Zhou et al. (2017). Condition (C2) specifies the parameter space of regression coefficients. A similar condition can be found in Gupta and Robinson (2018). Condition (C3) has been carefully studied in Lee (2004) and Gupta and Robinson (2018). Condition (C4) is used for showing the asymptotic normality of QMLE and it is a sufficient condition for local identification. A similar condition can be found in Gupta and Robinson (2015). Under this condition, the quasi-loglikelihood function is strictly concave near θ_0 and a local maximizer exists. Condition (C5) addresses the order of d and the number of non-zero elements in λ_0. The order condition of d allows the number of weight matrices to diverge to infinity. Condition (C6) is a minimum signal assumption placed on the non-zero coefficients. Thus, if some of the nonzero coefficients converge to zero too fast, they cannot be consistently identified. Similar conditions have been commonly used in extant literature such as Fan and Li (2001) and Chen and Chen (2012). Condition (C7) is used for showing the asymptotic property of EBIC. A similar condition can be found in Chen and Chen (2012).

To prove the theorems, we next introduce the following four useful lemmas. Since Lemma 1 below is directly modified from Theorem 1 of Kelejian and Prucha (2001) and Lemma 2 is modified from Proposition A.1 of Chen, Zhang and Zhong (2010), we only present the proofs of Lemmas 3 and 4 in Section S2 of this supplementary material. It is worth mentioning that Lemma 4 is a general result for a combination of quadratic forms of random errors, and it plays a critical role for obtaining the asymptotic variance of T_{ql} (i.e., σ^2_{ql}) and proving the result of Theorem 4.

Lemma 1. *Let* $\mathcal{E} = (\varepsilon_1, \cdots, \varepsilon_m)^\top$, *where* $\varepsilon_1, \cdots, \varepsilon_m$ *are independent and identically distributed random variables with mean 0 and finite variance* σ^2. *Define*

$$Q_m = \mathcal{E}^\top B \mathcal{E} - \sigma^2 \mathrm{tr}(B),$$

where $B = (b_{ij})_{m\times m} \in \mathbb{R}^{m\times m}$. *Suppose the following assumptions are satisfied:*

(1) for $i, j = 1, \cdots, m$, $b_{ij} = b_{ji}$;

(2) $\|B\|_R < \infty$*;*

(3) there exists some $\eta > 0$ *such that* $\mathrm{E}|\varepsilon_i|^{4+\eta} < \infty$*.*

Then, we have $\mathrm{E}(Q_m) = 0$ *and*

$$\sigma^2_{Q_m} := \mathrm{Var}(Q_m) = 4\sigma^4 \sum_{i=1}^{m} \sum_{j=1}^{i-1} b^2_{ij} + \sum_{i=1}^{m} \left[\left\{ \mu^{(4)}\sigma^4 - \sigma^4 \right\} b^2_{ii} \right],$$

where $\mu^{(k)} = E(\epsilon_i/\sigma)^k$*. Furthermore, suppose*

(4) $m^{-1}\sigma^2_{Q_m} \geq c_q$ *for some finite* $c_q > 0$*.*

Then, we obtain

$$\sigma^{-1}_{Q_m} Q_m \xrightarrow{d} N(0,1).$$

Lemma 2. *Let* $V = (V_1, \cdots, V_m)^\top \in \mathbb{R}^m$ *be a random vector. Assume that* $V_1, \cdots, V_m$ *are independent and identically distributed with mean 0 and unit variance. Then, for any arbitrary symmetric matrices* A_1*,* A_2*,* A_3 *and* A_4 *of bounded eigenvalues, we have (i)* $E\{(V^\top A_1 V)^2\} = tr^2(A_1) + 2tr(A_1^2) + \bar{\Delta} tr(A_1^{\otimes 2})$ *with* $\bar{\Delta} = E(V_i^4) - 3$*, where* $A_1 = (a_{j_1 j_2})$ *and* $A_1^{\otimes 2} = (a^2_{j_1 j_2})$*; (ii)* $E\{(V^\top A_1 V)(V^\top A_2 V)\} = tr(A_1)tr(A_2) + 2tr(A_1 A_2) + \bar{\Delta} tr(A_1 \otimes A_2)$*; (iii) there exists finite positive constant* C^* *such that* $E\big[\{V^\top A_1 V - tr(A_1)\}\{V^\top A_2 V - tr(A_2)\}\{V^\top A_3 V - tr(A_3)\}\{V^\top A_4 V - tr(A_4)\}\big] \leq C^* m^2$*.*

Lemma 3. *Under Conditions (C1)-(C5) in Section S1, as* $nT \to \infty$*, we obtain the following results.*

(i)

$$(nTd)^{-1/2}D\frac{\partial \ell(\theta_0)}{\partial \theta} \xrightarrow{d} N\left(0, G(\theta_0)\right),$$

where D is a $M\times(d+1)$ matrix satisfying $\|D\| < \infty$ and $d^{-1}D\mathcal{J}(\theta_0)D^\top \to G(\theta_0)$, $M < \infty$, and $G(\theta_0)$ is a positive definite matrix.

(ii) For any $\nu > 0$,

$$P\left(\frac{1}{\sqrt{nT}}\left|\frac{\partial \ell(\theta_0)}{\partial \theta_k}\right| > \nu\right) \le 2\exp\left\{-\min\left(\frac{\tau_1\nu\sigma_0^2}{\|U_k\|/\sqrt{nT}}, \frac{\tau_2\sigma_0^4\nu^2}{\|U_k\|_F^2/nT}\right)\right\},$$

for $k = 1, \cdots, d+1$. τ_1 and τ_2 are two finite positive constants.

(iii)

$$\left\|(nT)^{-1}\frac{\partial^2 \ell(\theta_0)}{\partial\theta\partial\theta^\top} + \mathcal{I}(\theta_0)\right\|_F = o_p(1).$$

To state the following lemma, let $\epsilon_t = (\epsilon_{1t}, \cdots, \epsilon_{nt})^\top \in \mathbb{R}^n$ be independent and identically distributed random variables with mean 0 and covariance matrix I_n for $t = 1, \cdots, T$. In addition, let $\bar{\Delta} = E(\epsilon_{it}^4) - 3$ and $W = (w_{kl}) \in \mathbb{R}^{d^*\times d^*}$ be a positive definite matrix, and assume that U_{tk} and V_{tk} are symmetric matrices satisfying $\max_k \sup_t \|U_{tk}\| < \infty$ and $\max_k \sup_t \|V_{tk}\| < \infty$. Moreover, define $Z_{nT} = (nT)^{-1}\sum_t(\epsilon_t^\top\epsilon_t)^2 + n^{-1}T^{-2}\sum_{t_1,t_2}\sum_{k,l}^{d^*} w_{kl}\{\epsilon_{t_1}^\top U_{t_1k}\epsilon_{t_1} - tr(U_{t_1k})\}\epsilon_{t_2}^\top V_{t_2l}\epsilon_{t_2}$ and $T_{nT} = Z_{nT} - (n + 2 + \bar{\Delta})$.

Lemma 4. *Let $\sigma_{nT}^2 = (8+4\bar{\Delta})c + n^{-2}T^{-4}\sum_{t_1\neq t_2\neq t_3}\sum_{k_1,l_1}\sum_{k_2,l_2} w_{k_1l_1}w_{k_2l_2}$ $\{2tr(U_{t_1k_1}U_{t_1k_2}) + \bar{\Delta}tr(U_{t_1k_1}\otimes U_{t_1k_2})\}tr(V_{t_2l_1})tr(V_{t_3l_2}) + (8+4\bar{\Delta})n^{-1}T^{-3}\sum_{t_1\neq t_2}\sum_{k,l} w_{kl}$*

$tr(U_{t_1k})tr(V_{t_2l})$. *Assume that there is a finite positive constant* c_σ *such that* $\sigma_{nT} > c_\sigma$. *Under the null hypothesis of* H_0 *and Conditions (C1)-(C5) in Section S1, we then have that, as* $nT \to \infty$ *and* $n/T \to c$ *for some finite constant* $c > 0$,

$$T_{nT}/\sigma_{nT} \to_d N(0,1).$$

S2 Proofs of Lemmas 3–4

Proof of Lemma 3: To show part (i) in the above lemma, we define $\epsilon = (\epsilon_1^\top, \cdots, \epsilon_T^\top)^\top \in \mathbb{R}^{nT}$. After simple calculation, we obtain that

$$\frac{\partial \ell(\theta_0)}{\partial \theta} = \big(\frac{1}{\sigma_0^2}\epsilon^\top U_1 \epsilon - tr(U_1), \cdots, \frac{1}{\sigma_0^2}\epsilon^\top U_d \epsilon - tr(U_d), \frac{1}{2\sigma_0^4}\epsilon^\top \epsilon - \frac{nT}{2\sigma_0^2}\big),$$

where U_k for $k = 1, \cdots, d$ is defined below Theorem 1. Denote $U_{d+1} = \frac{1}{2\sigma_0^2} I_{nT}$. By Cramér's theorem, it suffices to show that $d^{-1/2}c^\top D\frac{\partial \ell(\theta_0)}{\partial \theta}$ is asymptotic normal for any finite vector $c = (c_1, \cdots, c_M)^\top \in \mathbb{R}^M$. Define $U_c = d^{-1/2}\sum_{k=1}^{d+1}(\sum_{m=1}^{M} c_m D_{mk})U_k$ with $D = (D_{mk})$, then we have $d^{-1/2}c^\top D\frac{\partial \ell(\theta_0)}{\partial \theta} = (\sigma_0^2)^{-1}\big\{\epsilon^\top U_c \epsilon - \sigma_0^2 tr(U_c)\big\}$. By Lemma 1, we only need to verify that U_c satisfies Assumptions (1)–(4) listed in Lemma 1. Using the fact that U_k is symmetric for any $k = 1, \cdots, d$, Assumption (1) is satisfied. By Conditions (C2) and (C3), we have that $\|U_k\|_R \le \sup_t \|W_k^{(t)}\Delta_t(\lambda_0)\|_R \le$

$\sup_t \|W_k^{(t)}\|_R \|\Delta_t(\lambda_0)\|_R < \infty$. This implies that

$$\|U_c\|_R \le \sum_{k=1}^{d+1} d^{-1/2} \sum_{m=1}^{M} |c_m D_{mk}| \|U_k\|_R \le c_{u,1} d^{-1/2} \sum_{k=1}^{d+1} \sum_{m=1}^{M} |D_{mk}| \le c_{u,2} \|D\| < \infty$$

for two finite positive constants $c_{u,1}$ and $c_{u,2}$. Thus, U_c satisfies Assumption (2). By Condition (C1), Assumption (3) holds. In addition, since $(nT)^{-1}\mathrm{var}(d^{-1/2}c^\top D \frac{\partial \ell(\theta_0)}{\partial \theta}) = d^{-1}c^\top D \mathcal{J}_{nT}(\theta_0) D^\top c > 2^{-1}c^\top G(\theta_0) c > c_g$ for some positive constant c_g with probability approaching 1, Assumption (4) holds, which completes the first part of the proof.

We next very part (ii). Note that $P\left((nT)^{-1/2}|\frac{\partial \ell(\theta_0)}{\partial \theta_k}| > \nu\right) = P\big\{(nT)^{-1/2} |\epsilon^\top U_k \epsilon/\sigma_0^2 - tr(U_k)| > \nu\big\}$. By Condition (C1) and the Hanson-Wright inequality (Hanson and Wright (1971); Wright (1973)), we have

$$P\Big\{(nT)^{-1/2}|\epsilon^\top U_k \epsilon/\sigma_0^2 - tr(U_k)| > \nu\Big\} \le 2\exp\left[-\min\left(\frac{\tau_1 \nu \sigma_0^2}{\|U_k\|/\sqrt{nT}}, \frac{\tau_2 \sigma_0^4 \nu^2}{\|U_k\|_F^2/nT}\right)\right],$$

which completes the second part of the proof.

We lastly prove (iii). By Condition (C4), it suffices to show that $\|(nT)^{-1}\frac{\partial^2 \ell(\theta_0)}{\partial \theta \partial \theta^\top} + \mathcal{I}_{nT}(\theta_0)\|_F = o_p(1)$. After simple calculation, we have that

$$-(nT)^{-1}\frac{\partial^2 \ell(\theta_0)}{\partial \theta \partial \theta^\top} = \begin{pmatrix} -(nT)^{-1}\frac{\partial^2 \ell(\theta_0)}{\partial \lambda \partial \lambda^\top} & -(nT)^{-1}\frac{\partial^2 \ell(\theta_0)}{\partial \lambda \partial \sigma^2} \\ -(nT)^{-1}\frac{\partial^2 \ell(\theta_0)}{\partial \sigma^2 \partial \lambda} & -(nT)^{-1}\frac{\partial^2 \ell(\theta_0)}{\partial^2 \sigma^2} \end{pmatrix}, \text{ where}$$

$-(nT)^{-1}\frac{\partial^2 \ell(\theta_0)}{\partial^2 \sigma^2} = \frac{1}{nT\sigma_0^6}\epsilon^\top \epsilon - \frac{1}{2\sigma_0^4}$, $-(nT)^{-1}\frac{\partial^2 \ell(\theta_0)}{\partial \lambda_k \partial \sigma^2} = \frac{1}{nT\sigma_0^4}\epsilon^\top U_k \epsilon$ and $-(nT)^{-1}\frac{\partial^2 \ell(\theta_0)}{\partial \lambda_k \partial \lambda_l} = \frac{1}{nT\sigma_0^2}\epsilon^\top U_k U_l \epsilon + (nT)^{-1} tr(U_k U_l)$ for $k = 1, \cdots, d$ and $l = 1, \cdots, d$. Note that $\mathrm{Var}\big(\frac{1}{nT\sigma_0^6}\epsilon^\top \epsilon\big) = O((nT)^{-1})$. Accordingly, we only need to show that

$\mathrm{Var}(\frac{1}{nT\sigma_0^4}\epsilon^\top U_k\epsilon) = O((nT)^{-1})$ and $\mathrm{Var}(\frac{1}{nT\sigma_0^2}\epsilon^\top U_k U_l\epsilon) = O((nT)^{-1})$ and they are given below.

By Lemma 2 (i) and the fact that $(nT)^{-1}E\big(\epsilon^\top U_k\epsilon\big) = (nT)^{-1}\sigma_0^2 tr(U_k)$, we have $(nT)^{-2}\mathrm{Var}\big(\epsilon^\top U_k\epsilon\big) = (nT)^{-2}\sigma_0^4\{2tr(U_k^2) + (\mu^{(4)} - 3)tr(U_k^{\otimes 2})\}$. By Condition (C3), we further have $tr(U_k^2) = O(nT)$ and $tr(U_k^{\otimes 2}) \le tr(U_k^2) = O(nT)$. This leads to $(nT\sigma_0^4)^{-2}\mathrm{Var}\big(\epsilon^\top U_k\epsilon\big) = O((nT)^{-1})$. Analogously, by Conditions (C3) and Lemma 2 (i), we can show that $\mathrm{Var}(\frac{1}{nT\sigma_0^2}\epsilon^\top U_k U_l\epsilon) = O((nT)^{-1})$.

Consequently, for any $\tau > 0$, by Condition (C5) and employ Chebyshev's inequality, we have

$$P\left(\left\|(nT)^{-1}\frac{\partial^2\ell(\theta_0)}{\partial\theta\partial\theta^\top} + \mathcal{I}_{nT}(\theta_0)\right\|_F > \tau/d\right) \le d^2/\tau^2\sum_{k=1}^{d+1}\sum_{l=1}^{d+1}\mathrm{Var}\Big\{(nT)^{-1}\frac{\partial^2\ell(\theta_0)}{\partial\theta_k\partial\theta_l}\Big\}$$
$$\le d^2/\tau^2\sum_{k=1}^{d+1}\sum_{l=1}^{d+1}O\big\{(nT)^{-1}\big\} = O\big\{d^4/(nT\tau^2)\big\} = o(1).$$

The above result, together with Condition (C4), leads to

$$\left\|(nT)^{-1}\frac{\partial^2\ell(\theta_0)}{\partial\theta\partial\theta^\top}+\mathcal{I}(\theta_0)\right\|_F \le \left\|(nT)^{-1}\frac{\partial^2\ell(\theta_0)}{\partial\theta\partial\theta^\top}+\mathcal{I}_{nT}(\theta_0)\right\|_F + \left\|\mathcal{I}(\theta_0)-\mathcal{I}_{nT}(\theta_0)\right\|_F = o_p(1),$$

which completes the proof of the lemma.

Proof of Lemma 4: Let $Z_{nT} = Z_{nT,1}+Z_{nT,2}$, where $Z_{nT,1} = (nT)^{-1}\sum_t(\epsilon_t^\top\epsilon_t)^2$ and $Z_{nT,2} = n^{-1}T^{-2}\sum_{t_1,t_2}\sum_{k,l}^{d^*} w_{kl}\{\epsilon_{t_1}^\top U_{t_1k}\epsilon_{t_1} - tr(U_{t_1k})\}\epsilon_{t_2}^\top V_{t_2l}\epsilon_{t_2}$. By Condition (C4), after tedious calculations, we obtain that $E(Z_{nT,1}) = n+2+\bar{\Delta}$,

$E(Z_{nT,2}) = o(1)$, $\text{Var}(Z_{nT,1}) = (8+4\bar{\Delta})c\{1+o(1)\}$,

$$\text{Var}(Z_{nT,2}) = n^{-2}T^{-4} \sum_{t_1 \neq t_2 \neq t_3} \sum_{k_1,l_1} \sum_{k_2,l_2} w_{k_1l_1} w_{k_2l_2}$$

$$\times\{2tr(U_{t_1k_1}U_{t_1k_2}) + \bar{\Delta}tr(U_{t_1k_1} \otimes U_{t_1k_2})\}tr(V_{t_2l_1})tr(V_{t_3l_2})\{1+o(1)\}$$

and $\text{cov}(Z_{nT,1}, Z_{nT,2}) = (4+2\bar{\Delta})n^{-1}T^{-3}\sum_{t_1 \neq t_2}\sum_{k,l} w_{kl} tr(U_{t_1k})tr(V_{t_2l})\{1+o(1)\}$. Then, one can straightforwardly demonstrate that $E(Z_{nT}) = n+2+\bar{\Delta}+o(1)$. In addition, by the lemma assumption that $\sigma_{nT} > c_\sigma$, we have $\text{Var}(Z_{nT}) > c_\sigma/2$.

We next define $\mathcal{F}_r = \{\epsilon_t, t \leq r\}$ to be the σ-field generated by $\{\epsilon_t\}$ for $t \leq r$. In addition, define $T_{nT,r} = (nT)^{-1}\sum_{t=1}^{r}(\epsilon_t^\top \epsilon_t)^2 + n^{-1}T^{-2}\sum_{t_1 \leq r, t_2 \leq r}\sum_{k,l}^{d^*} w_{kl}$ $\{\epsilon_{t_1}^\top U_{t_1k}\epsilon_{t_1} - tr(U_{t_1k})\}\epsilon_{t_2}^\top V_{t_2l}\epsilon_{t_2} - T^{-1}(n+2+\bar{\Delta})r$. Obviously, we have $T_{nT,r} \in \mathcal{F}_r$. Then, set $\Delta_{nT,r} = T_{nT,r} - T_{nT,r-1}$ with $\Delta_{nT,0} = 0$. One can easily verify that $E(\Delta_{nT,r}|\mathcal{F}_q) = 0$ and $E(T_{nT,r}|\mathcal{F}_q) = T_{nT,q}$ for any $q < r$. This implies that, for an arbitrary fixed T, $\{\Delta_{nT,r}, 0 \leq r \leq T\}$ is a martingale difference sequence with respect to $\{\mathcal{F}_r, 0 \leq r \leq T\}$ and $\mathcal{F}_0 = \emptyset$. Accordingly, by the Martingale Central Limit Theorem (Hall and Heyde (1980)), it suffices to show that

$$\text{Var}\Big(\sum_{r=1}^{T} \sigma^*_{nT,r}\Big) \to_p 0, \text{ and } \sum_{r=1}^{T} E(\Delta^4_{nT,r}) \to 0, \qquad \text{(S2.1)}$$

where $\sigma^*_{nT,r} = E(\Delta^2_{nT,r}|\mathcal{F}_{r-1})$. We next verify the above two parts separately via the following two steps. Without loss of generality, we assume that

$w_{kk} = 1$ and $w_{kl} = 0$ for any $k \neq l$ to ease the calculation.

STEP I. We begin by showing the first term of (S2.1). After simple calculation, we have $\Delta_{nT,r} = (nT)^{-1}(\epsilon_r^\top \epsilon_r)^2 - T^{-1}(n+2+\bar{\Delta}) + n^{-1}T^{-2}\sum_k \sum_{t<r}\{\epsilon_t^\top U_{tk}\epsilon_t - tr(U_{tk})\}\epsilon_r^\top V_{rk}\epsilon_r + n^{-1}T^{-2}\sum_k\sum_{t<r}\{\epsilon_r^\top U_{rk}\epsilon_r - tr(U_{rk})\}\epsilon_t^\top V_{tk}\epsilon_t$. Based on the above three components of $\Delta_{nT,r}$, we further obtain $\sigma^*_{nT,r} = E(\Delta^2_{nT,r}|\mathcal{F}_{r-1}) = \sum_{k=1}^6 \Pi_{nT,r(k)}$, where

$$\Pi_{nT,r(1)} = (nT)^{-2}E[\{(\epsilon_r^\top \epsilon_r)^2 - n(n+2+\bar{\Delta})\}^2],$$

$$\Pi_{nT,r(2)} = n^{-2}T^{-4}\sum_{k_1,k_2}^{d^*}\sum_{t_1<r}\sum_{t_2<r}\{tr(V_{rk_1})tr(V_{rk_2}) + 2tr(V_{rk_1}V_{rk_2}) + \bar{\Delta}tr(V_{rk_1}\otimes V_{rk_2})\}$$
$$\times\{\epsilon_{t_1}^\top U_{t_1k_1}\epsilon_{t_1} - tr(U_{t_1k_1})\}\{\epsilon_{t_2}^\top U_{t_2k_2}\epsilon_{t_2} - tr(U_{t_2k_2})\},$$

$$\Pi_{nT,r(3)} = n^{-2}T^{-4}\sum_{k_1,k_2}^{d^*}\sum_{t_1<r}\sum_{t_2<r}\{2tr(U_{rk_1}U_{rk_2}) + \bar{\Delta}tr(U_{rk_1}\otimes U_{rk_2})\}\epsilon_{t_1}^\top V_{t_1k_1}\epsilon_{t_1}\epsilon_{t_2}^\top V_{t_2k_2}\epsilon_{t_2},$$

$$\Pi_{nT,r(4)} = n^{-1}T^{-2}\sum_k\sum_{t<r}Q_{r(1)k}\{\epsilon_t^\top U_{tk}\epsilon_t - tr(U_{tk})\},$$

$$\Pi_{nT,r(5)} = n^{-1}T^{-2}\sum_k\sum_{t<r}Q_{r(2)k}\epsilon_t^\top V_{tk}\epsilon_t$$

$$\Pi_{nT,r(6)} = n^{-2}T^{-4}\sum_{k_1,k_2}\sum_{t_1<r}\sum_{t_2<r}\{2tr(U_{rk_1}V_{rk_2}) + \bar{\Delta}tr(U_{rk_1}\otimes V_{rk_2})\}\{\epsilon_{t_1}^\top U_{t_1k_1}\epsilon_{t_1} - tr(U_{t_1k_1})\}$$
$$\times\epsilon_{t_2}^\top V_{t_2k_2}\epsilon_{t_2},$$

where $Q_{r(1)k} = (nT)^{-1}E[\epsilon_r^\top V_{rk}\epsilon_r\{(\epsilon_r^\top\epsilon_r)^2 - n(n+2+\bar{\Delta})\}]$ and $Q_{r(2)k} = (nT)^{-1}E[\{\epsilon_r^\top U_{rk}\epsilon_r - tr(U_r)\}\{(\epsilon_r^\top\epsilon_r)^2 - n(n+2+\bar{\Delta})\}]$. Accordingly, to prove the first term of (S2.1), it suffices to show that $(d^*)^{-4}\text{Var}\{\sum_{r=1}^T \Pi_{nT,r(k)}\} \to$

0 for $k = 1, \cdots, 6$. It is worth noting that $\mathrm{Var}\{\sum_{r=1}^T \Pi_{nT,r(1)}\} = 0$. Hence, we only need to verify five of them.

We first consider $\Pi_{nT,r(2)}$, and decompose it into two terms $\Pi_{nT,r(2)} \triangleq \Pi_{nT,r(2)}^{(1)} + \Pi_{nT,r(2)}^{(2)}$, where

$$\Pi_{nT,r(2)}^{(1)} = n^{-2}T^{-4} \sum_{k_1,k_2} \sum_{t_1 \neq t_2 < r} \{tr(V_{rk_1})tr(V_{rk_2}) + 2tr(V_{rk_1}V_{rk_2}) + \bar{\Delta} tr(V_{rk_1} \otimes V_{rk_2})\}$$
$$\times \{\epsilon_{t_1}^\top U_{t_1k_1}\epsilon_{t_1} - tr(U_{t_1k_1})\}\{\epsilon_{t_2}^\top U_{t_2k_2}\epsilon_{t_2} - tr(U_{t_2k_2})\}$$

and $\Pi_{nT,r(2)}^{(2)} = n^{-2}T^{-4} \sum_{k_1,k_2} \sum_{t<r} \{tr(V_{rk_1})tr(V_{rk_2}) + 2tr(V_{rk_1}V_{rk_2}) + \bar{\Delta} tr(V_{rk_1} \otimes V_{rk_2})\}$

$$\times \{\epsilon_t^\top U_{tk_1}\epsilon_t - tr(U_{tk_1})\}\{\epsilon_t^\top U_{tk_2}\epsilon_t - tr(U_{tk_2})\}.$$

Using the fact that $E(\Pi_{nT,r(2)}^{(1)}) = 0$, we can have

$$\mathrm{Var}\big(\sum_{r=1}^T \Pi_{nT,r(2)}^{(1)}\big) = E\big(\sum_{r=1}^T \Pi_{nT,r(1)}^{(1)}\big)^2$$

$$= n^{-4}T^{-8} \sum_{k_1,k_2,k_3,k_4} \sum_{r_1=1}^T \sum_{r_2=1}^T \sum_{t_1 \neq t_2 < r_1} \sum_{t_3 \neq t_4 < r_2} \{tr(V_{r_1k_1})tr(V_{r_1k_2}) + 2tr(V_{r_1k_1}V_{r_1k_2})$$
$$+ \bar{\Delta} tr(V_{r_1k_1} \otimes V_{r_1k_2})\} \times \{tr(V_{r_2k_3})tr(V_{r_2k_4}) + 2tr(V_{r_2k_3}V_{r_2k_4}) + \bar{\Delta} tr(V_{r_2k_3} \otimes V_{r_2k_4})\}$$
$$\times E\big[\{\epsilon_{t_1}^\top U_{t_1k_1}\epsilon_{t_1} - tr(U_{t_1k_1})\}\{\epsilon_{t_2}^\top U_{t_2k_2}\epsilon_{t_2} - tr(U_{t_2k_2})\}\{\epsilon_{t_3}^\top U_{t_3k_3}\epsilon_{t_3} - tr(U_{t_3k_3})\}$$
$$\times \{\epsilon_{t_4}^\top U_{t_4k_4}\epsilon_{t_4} - tr(U_{t_4k_4})\}\big]$$
$$= n^{-4}T^{-8} \sum_{k_1,k_2,k_3,k_4} \sum_{r_1=1}^T \sum_{r_2=1}^T \sum_{t_1 \neq t_2 < \max\{r_1,r_2\}} \{tr(V_{r_1k_1})tr(V_{r_1k_2}) + 2tr(V_{r_1k_1}V_{r_1k_2})$$
$$+ \bar{\Delta} tr(V_{r_1k_1} \otimes V_{r_1k_2})\} \times \{tr(V_{r_2k_3})tr(V_{r_2k_4}) + 2tr(V_{r_2k_3}V_{r_2k_4}) + \bar{\Delta} tr(V_{r_2k_3} \otimes V_{r_2k_4})\}$$
$$\times E\big[\{\epsilon_{t_1}^\top U_{t_1k_1}\epsilon_{t_1} - tr(U_{t_1k_1})\}\{\epsilon_{t_1}^\top U_{t_1k_3}\epsilon_{t_1} - tr(U_{t_1k_3})\}\big]$$

$$\times E\big[\{\epsilon_{t_2}^\top U_{t_2k_2}\epsilon_{t_2} - tr(U_{t_2k_2})\}\{\epsilon_{t_2}^\top U_{t_2k_4}\epsilon_{t_2} - tr(U_{t_2k_4})\}\big]$$

$$+n^{-4}T^{-8}\sum_{k_1,k_2,k_3,k_4}\sum_{r_1=1}^{T}\sum_{r_2=1}^{T}\sum_{t_1\neq t_2<\max\{r_1,r_2\}}\{tr(V_{r_1k_1})tr(V_{r_1k_2})+2tr(V_{r_1k_1}V_{r_1k_2})$$

$$+\bar{\Delta}tr(V_{r_1k_1}\otimes V_{r_1k_2})\}\times\{tr(V_{r_2k_3})tr(V_{r_2k_4})+2tr(V_{r_2k_3}V_{r_2k_4})+\bar{\Delta}tr(V_{r_2k_3}\otimes V_{r_2k_4})\}$$

$$\times E\big[\{\epsilon_{t_1}^\top U_{t_1k_1}\epsilon_{t_1} - tr(U_{t_1k_1})\}\{\epsilon_{t_1}^\top U_{t_1k_4}\epsilon_{t_1} - tr(U_{t_1k_4})\}\big]$$

$$\times E\big[\{\epsilon_{t_2}^\top U_{t_2k_2}\epsilon_{t_2} - tr(U_{t_2k_2})\}\{\epsilon_{t_2}^\top U_{t_2k_3}\epsilon_{t_2} - tr(U_{t_2k_3})\}\big]$$

$$= n^{-4}T^{-8}\times O\{(d^*)^4\}\times T^4\times O(n^4)\times O(n^2) = O\{(d^*)^4n^2T^{-4}\} = o(1),$$

where the last equality follows from the lemma's assumption that $n/T\to c$ for some finite positive constant c and Condition (C5). In addition,

$$\mathrm{Var}\big(\sum_{r=1}^{T}\Pi_{nT,r(2)}^{(2)}\big) = n^{-4}T^{-8}\mathrm{Var}\big[\sum_{k_1,k_2}\sum_{r}\sum_{t<r}\{tr(V_{rk_1})tr(V_{rk_2})+2tr(V_{rk_1}V_{rk_2})$$

$$+\bar{\Delta}tr(V_{rk_1}\otimes V_{rk_2})\}\times\{\epsilon_t^\top U_{tk_1}\epsilon_t - tr(U_{tk_1})\}\{\epsilon_t^\top U_{tk_2}\epsilon_t - tr(U_{tk_2})\}\big]$$

$$\leq n^{-4}T^{-8}\times(d^*)^2T^2\sum_{k_1,k_2}\sum_{r}\sum_{t<r}\mathrm{Var}\big[\{tr(V_{rk_1})tr(V_{rk_2})+2tr(V_{rk_1}V_{rk_2})$$

$$+\bar{\Delta}tr(V_{rk_1}\otimes V_{rk_2})\}\times\{\epsilon_t^\top U_{tk_1}\epsilon_t - tr(U_{tk_1})\}\{\epsilon_t^\top U_{tk_2}\epsilon_t - tr(U_{tk_2})\}\big].$$

By Lemma 2 (iii), we have $\mathrm{Var}\big[\{\epsilon_t^\top U_{tk_1}\epsilon_t - tr(U_{tk_1})\}\{\epsilon_t^\top U_{tk_2}\epsilon_t - tr(U_{tk_2})\}\big] = O(n^2)$ uniformly for any k_1 and k_2. As a result, $\mathrm{Var}\big(\sum_{r=1}^{T}\Pi_{nT,r(1)}^{(2)}\big) = n^{-4}T^{-8}\times(d^*)^4\times T^2\times T^2\times O(n^4)O(n^2) = O\{(d^*)^4n^2T^{-4}\} = o(1)$. This, together with the above result, implies that $\mathrm{Var}\{\sum_{r=1}^{T}\Pi_{nT,r(2)}\}\to 0$.

We next consider $\Pi_{nT,r(3)}$, and analogously decompose $\Pi_{nT,r(3)}$ into two parts $\Pi_{nT,r(3)} = \Pi_{nT,r(3)}^{(1)}+\Pi_{nT,r(3)}^{(2)}$, where $\Pi_{nT,r(3)}^{(1)} = n^{-2}T^{-4}\sum_{k_1,k_2}\sum_{t_1\neq t_2<r}\{2tr(U_{rk_1}U_{rk_2})+$

$\bar{\Delta}tr(U_{rk_1}\otimes U_{rk_2})\}\epsilon_{t_1}^\top V_{t_1k_1}\epsilon_{t_1}\epsilon_{t_2}^\top V_{t_2k_2}\epsilon_{t_2}$ and $\Pi_{nT,r(3)}^{(2)} = n^{-2}T^{-4}\sum_{k_1,k_2}\{2tr(U_{rk_1}U_{rk_2})+$ $\bar{\Delta}tr(U_{rk_1}\otimes U_{rk_2})\}\epsilon_t^\top V_{tk_1}\epsilon_t\epsilon_t^\top V_{tk_2}\epsilon_t$. Note that

$$\mathrm{Var}\Big(\sum_{r=1}^T \Pi_{nT,r(3)}^{(1)}\Big) = n^{-4}T^{-8}\mathrm{Var}\Big[\sum_{k_1,k_2}\sum_{r}\sum_{t_1\neq t_2<r}\{2tr(U_{rk_1}U_{rk_2})+\bar{\Delta}tr(U_{rk_1}\otimes U_{rk_2})\}$$
$$\times\epsilon_{t_1}^\top V_{t_1k_1}\epsilon_{t_1}\epsilon_{t_2}^\top V_{t_2k_2}\epsilon_{t_2}\Big]$$
$$\leq n^{-4}T^{-8}\times(d^*)^2T^3\sum_{k_1,k_2}\sum_{r}\sum_{t_1\neq t_2<r}\{2tr(U_{rk_1}U_{rk_2})+\bar{\Delta}tr(U_{rk_1}\otimes U_{rk_2})\}^2\mathrm{Var}\big(\epsilon_{t_1}^\top V_{t_1k_1}\epsilon_{t_1}\epsilon_{t_2}^\top V_{t_2k_2}\epsilon_{t_2}\big).$$

By $\mathrm{Var}\big(\epsilon_{t_1}^\top V_{t_1k_1}\epsilon_{t_1}\epsilon_{t_2}^\top V_{t_2k_2}\epsilon_{t_2}\big) = \mathrm{Var}\big(\epsilon_{t_1}^\top V_{t_1k_1}\epsilon_{t_1}\big)\mathrm{Var}\big(\epsilon_{t_2}^\top V_{t_2k_2}\epsilon_{t_2}\big) = O(n)\times O(n) = O(n^2)$, we further have $\mathrm{Var}\big(\sum_{r=1}^T \Pi_{nT,r(3)}^{(1)}\big) \leq n^{-4}T^{-8}(d^*)^4\times T^3\times T^3\times O(n^2)O(n^2) = O\{(d^*)^4T^{-2}\} = o(1)$.

By Lemma 2 (iii), we obtain that $\mathrm{Var}(\epsilon_t^\top V_{tk_1}\epsilon_t\epsilon_t^\top V_{tk_2}\epsilon_t) \leq E\{(\epsilon_t^\top V_{tk_1}\epsilon_t)^2(\epsilon_t^\top V_{tk_2}\epsilon_t)^2\} = O(n^4)$. Then using Condition (C5), we can have

$$\mathrm{Var}\Big(\sum_{r=1}^T \Pi_{nT,r(3)}^{(2)}\Big) \leq n^{-4}T^{-8}\times(d^*)^2T^2\sum_{k_1,k_2}\sum_{r}\sum_{t<r}\{2tr(U_{rk_1}U_{rk_2})+\bar{\Delta}tr(U_{rk_1}\otimes U_{rk_2})\}^2$$
$$\times\mathrm{Var}(\epsilon_t^\top V_{tk_1}\epsilon_t\epsilon_t^\top V_{tk_2}\epsilon_t) = n^{-4}T^{-8}\times(d^*)^4\times T^2\times T^2\times O(n^2)O(n^4) = O\{(d^*)^4n^2T^{-4}\} = o(1)$$

This, in conjunction with the above result, leads to $\mathrm{Var}\{\sum_{r=1}^T \Pi_{nT,r(3)}\} \to 0$.

Applying techniques similar to those used in the above proofs, one can verify that $\mathrm{Var}\{\sum_{r=1}^T \Pi_{nT,r(6)}\} \to 0$. Finally, using the results of Bao and Ullah (2010), we can obtain $Q_{r(1)k} \leq (nT)^{-1}E\{(\epsilon_r^\top\epsilon_t)^3\} = O(nT^{-1}) = O(1)$ and $Q_{r(2)k} = O(1)$ uniformly for any k, one can also employ techniques similar to those used in the above proofs to demonstrate that $\mathrm{Var}\{\sum_{r=1}^T \Pi_{nT,r(4)}\} \to$

0 and $\mathrm{Var}\{\sum_{r=1}^{T}\Pi_{nT,r(5)}\}\to 0$. Consequently, we have completed the proof of the first term of (S2.1).

STEP II. To verify the second term of (S2.1), denote $\Delta_{nT,r}\triangleq\Delta_{nT,r(1)}+\Delta_{nT,r(2)}+\Delta_{nT,r(3)}$, where $\Delta_{nT,r(1)}=(nT)^{-1}(\epsilon_r^\top\epsilon_r)^2-T^{-1}(n+2+\bar{\Delta})$, $\Delta_{nT,r(2)}=n^{-1}T^{-2}\sum_k\sum_{t<r}\{\epsilon_t^\top U_{tk}\epsilon_t-tr(U_{tk})\}\epsilon_r^\top V_{rk}\epsilon_r$, and $\Delta_{nT,r(3)}=n^{-1}T^{-2}\sum_k\sum_{t<r}\{\epsilon_r^\top U_{rk}\epsilon_r-tr(U_{rk})\}\epsilon_t^\top V_{tk}\epsilon_t$. Accordingly, it suffices to show that, for $k=1,2$ and 3, $\sum_{r=1}^{T}E(\Delta^4_{nT,r(k)})\to 0$.

We first consider $k=1$. Note that $E\big\{(\epsilon_r^\top\epsilon_r)^2-n(n+2+\bar{\Delta})\big\}^4=O(n^6)$ by the results of Bao and Ullah (2010). This, together with the lemma's assumption that $n/T\to c$ for some finite positive constant c, implies that $\sum_{r=1}^{T}E(\Delta^4_{nT,r(1)})=(nT)^{-4}\sum_{r=1}^{T}E\big\{(\epsilon_r^\top\epsilon_r)^2-n(n+2+\bar{\Delta})\big\}^4=(nT)^{-4}T\times O(n^6)=o(1)$.

We next consider $k=2$. After algebraic simplification, we have that $\sum_{r=1}^{T}E(\Delta^4_{nT,r(2)})=n^{-4}T^{-8}\sum_{k_1,k_2,k_3,k_4}\sum_{r=1}^{T}\sum_{t_1<r}\sum_{t_2<r}\sum_{t_3<r}\sum_{t_4<r}E\big[\{\epsilon_{t_1}^\top U_{t_1k_1}\epsilon_{t_1}-tr(U_{t_1k_1})\}\{\epsilon_{t_2}^\top U_{t_2k_2}\epsilon_{t_2}-tr(U_{t_2k_2})\}\{\epsilon_{t_3}^\top U_{t_3k_3}\epsilon_{t_3}-tr(U_{t_3k_3})\}\{\epsilon_{t_4}^\top U_{t_4k_4}\epsilon_{t_4}-tr(U_{t_4k_4})\}\big]E\big[(\epsilon_r^\top V_{rk_1}\epsilon_r)(\epsilon_r^\top V_{rk_2}\epsilon_r)(\epsilon_r^\top V_{rk_3}\epsilon_r)(\epsilon_r^\top V_{rk_4}\epsilon_r)\big]$. Under the assumption that $\sup_t\|V_t\|<\infty$, we obtain $E\big[(\epsilon_r^\top V_{rk_1}\epsilon_r)(\epsilon_r^\top V_{rk_2}\epsilon_r)(\epsilon_r^\top V_{rk_3}\epsilon_r)(\epsilon_r^\top V_{rk_4}\epsilon_r)\big]=O(n^4)$, which immediately leads to

$$\sum_{r=1}^{T}E(\Delta^4_{nT,r(2)})=T^{-8}\sum_{k_1,k_2,k_3,k_4}\sum_{r=1}^{T}\sum_{t_1<r}\sum_{t_2<r}\sum_{t_3<r}\sum_{t_4<r}E\big[\{\epsilon_{t_1}^\top U_{t_1k_1}\epsilon_{t_1}-tr(U_{t_1k_1})\}$$
$$\times\{\epsilon_{t_2}^\top U_{t_2k_2}\epsilon_{t_2}-tr(U_{t_2k_2})\}\{\epsilon_{t_3}^\top U_{t_3k_3}\epsilon_{t_3}-tr(U_{t_3k_3})\}\{\epsilon_{t_4}^\top U_{t_4k_4}\epsilon_{t_4}-tr(U_{t_4k_4})\}\big]\times O(1).$$

Furthermore,

$$T^{-8}\sum_{k_1,k_2,k_3,k_4}\sum_{t=1}^{T}\sum_{t_1<r}\sum_{t_2<r}\sum_{t_3<r}\sum_{t_4<r}E\big[\{\epsilon_{t_1}^\top U_{t_1k_1}\epsilon_{t_1}-tr(U_{t_1k_1})\}\{\epsilon_{t_2}^\top U_{t_2k_2}\epsilon_{t_2}-tr(U_{t_2k_2})\}$$
$$\times\{\epsilon_{t_3}^\top U_{t_3k_3}\epsilon_{t_3}-tr(U_{t_3k_3})\}\{\epsilon_{t_4}^\top U_{t_4k_4}\epsilon_{t_4}-tr(U_{t_4k_4})\}\big]$$
$$=T^{-8}\sum_{k_1,k_2,k_3,k_4}\sum_{t=1}^{T}\sum_{t<r}E\big[\{\epsilon_{t}^\top U_{tk_1}\epsilon_{t}-tr(U_{tk_1})\}\{\epsilon_{t}^\top U_{tk_2}\epsilon_{t}-tr(U_{tk_2})\}$$
$$\times\{\epsilon_{t}^\top U_{tk_3}\epsilon_{t}-tr(U_{tk_3})\}\{\epsilon_{t}^\top U_{tk_4}\epsilon_{t}-tr(U_{tk_4})\}\big]$$
$$+T^{-8}\sum_{k_1,k_2,k_3,k_4}\sum_{t=1}^{T}\sum_{t_1<r,t_2<r,t_1\neq t_2}E\big[\{\epsilon_{t_1}^\top U_{t_1k_1}\epsilon_{t_1}-tr(U_{t_1k_1})\}\{\epsilon_{t_1}^\top U_{t_1k_2}\epsilon_{t_1}-tr(U_{t_1k_2})\}\big]$$
$$\times E\big[\{\epsilon_{t_2}^\top U_{t_2k_3}\epsilon_{t_2}-tr(U_{t_2k_3})\}\{\epsilon_{t_2}^\top U_{t_2k_4}\epsilon_{t_2}-tr(U_{t_2k_4})\}\big]$$
$$+T^{-8}\sum_{k_1,k_2,k_3,k_4}\sum_{t=1}^{T}\sum_{t_1<r,t_2<r,t_1\neq t_2}E\big[\{\epsilon_{t_1}^\top U_{t_1k_1}\epsilon_{t_1}-tr(U_{t_1k_1})\}\{\epsilon_{t_1}^\top U_{t_1k_2}\epsilon_{t_1}-tr(U_{t_1k_2})\}\big]$$
$$\times E\big[\{\epsilon_{t_2}^\top U_{t_2k_3}\epsilon_{t_2}-tr(U_{t_2k_3})\}\{\epsilon_{t_2}^\top U_{t_2k_4}\epsilon_{t_2}-tr(U_{t_2k_4})\}\big]$$
$$+T^{-8}\sum_{k_1,k_2,k_3,k_4}\sum_{t=1}^{T}\sum_{t_1<r,t_2<r,t_1\neq t_2}E\big[\{\epsilon_{t_1}^\top U_{t_1k_1}\epsilon_{t_1}-tr(U_{t_1k_1})\}\{\epsilon_{t_1}^\top U_{t_1k_2}\epsilon_{t_1}-tr(U_{t_1k_2})\}\big]$$
$$\times E\big[\{\epsilon_{t_2}^\top U_{t_2k_3}\epsilon_{t_2}-tr(U_{t_2k_3})\}\{\epsilon_{t_2}^\top U_{t_2k_4}\epsilon_{t_2}-tr(U_{t_2k_4})\}\big]\triangleq\Pi_1+\Pi_2+\Pi_3+\Pi_4.$$

By Lemma 2 (iii), we have $E\big[\{\epsilon_t^\top U_{tk_1}\epsilon_t-tr(U_{tk_1})\}\{\epsilon_t^\top U_{tk_2}\epsilon_t-tr(U_{tk_2})\}\{\epsilon_t^\top U_{tk_3}\epsilon_t-tr(U_{tk_3})\}\{\epsilon_t^\top U_{tk_4}\epsilon_t-tr(U_{tk_4})\}\big]=O(n^2)$. Thus, $\Pi_1=T^{-8}\times T(T-1)\times O(n^2)\times(d^*)^4=o(1)$. By Lemma 2 (ii), we have $E\big[\{\epsilon_{t_1}^\top U_{t_1k_1}\epsilon_{t_1}-tr(U_{t_1k_1})\}\{\epsilon_{t_1}^\top U_{t_1k_2}\epsilon_{t_1}-tr(U_{t_1k_2})\}\big]=O(n)$. As a result, $\Pi_2=T^{-8}\times T(T-1)^2\times O(n^2)\times(d^*)^2=o(1)$. This, in conjunction with the above results, implies that $\sum_{r=1}^T E(\Delta^4_{nT,r(2)})\to 0$. Analogously, we can demonstrate that $\sum_{r=1}^T E(\Delta^4_{nT,r(3)})\to 0$, which completes the entire proof.

S3 Proofs of Theorems 1–4

Proof of Theorem 1: To prove this theorem, we take the following two steps: (i) showing that $\hat{\theta}$ is $(nT/d)^{1/2}$-consistent; (ii) verifying that $\hat{\theta}$ is asymptotically normal.

Step I. To show the consistency, it suffices to follow the technique of Fan and Li (2001) to demonstrate that, for an arbitrarily small positive constant $\xi > 0$, there exists a constant $C_\xi > 0$ such that

$$\mathrm{P}\left\{\sup_{u\in\mathbb{R}^{d+1}:\|u\|=C_\xi} \ell\{\theta_0 + (nT/d)^{-1/2}u\} < \ell(\theta_0)\right\} \geq 1-\xi \qquad \text{(S3.2)}$$

for nT sufficiently large. To this end, we employ a Taylor series expansion and obtain that

$$\begin{aligned} &\sup_{u\in\mathbb{R}^{d+1}:\|u\|=C_\xi} \ell\left(\theta_0 + (nT/d)^{-1/2}u\right) - \ell(\theta_0) \\ = &\sup_{u\in\mathbb{R}^{d+1}:\|u\|=C_\xi}\left[\frac{1}{(nT/d)^{1/2}}u^\top\frac{\partial\ell(\theta_0)}{\partial\theta} - \frac{d}{2nT}u^\top\left\{-\frac{\partial^2\ell(\theta_0)}{\partial\theta\partial\theta^\top}\right\}u + R_n(u)\right], \end{aligned}$$

where $R_n(u)$ is a negligible term that satisfies $R_n(u) = o_p(d)$. According to Lemma 3 (i) and Condition (C4), we have $(nT/d)^{-1/2}u^\top\frac{\partial\ell(\theta_0)}{\partial\theta} = dC_\xi O_p(1)$ and

$$-\frac{d}{2nT}u^\top\Big\{-\frac{\partial^2\ell(\theta_0)}{\partial\theta\partial\theta^\top}\Big\}u = -\frac{1}{2}du^\top\mathcal{I}(\theta_0)u + o_p(d) \leq -\frac{1}{2}dc_{\min,1}C_\xi^2.$$

Note that $dC_\xi O_p(1) - \frac{1}{2}dc_{\min,1}C_\xi^2$ is a quadratic function of C_ξ. Hence, as long as C_ξ is sufficient large, we have

$$\sup_{u\in\mathbb{R}^{d+1}:\|u\|=C_\xi}\Big[\ell\big\{\theta_0+(nT/d)^{-1/2}u\big\}-\ell(\theta_0)\Big]<0, \tag{S3.3}$$

with probability tending to 1, which demonstrates (S3.2). Based on the result of (S3.3), there exists a local maximizer $\hat\theta$ such that $\|\hat\theta-\theta_0\|\le(nT/d)^{-1/2}C_\xi$ for nT sufficiently large. This, in conjunction with (S3.2), implies

$$P\Big(\|\hat\theta-\theta_0\|\le(nT/d)^{-1/2}C_\xi\Big)\ge P\Big\{\sup_{u\in\mathbb{R}^{d+1}:\|u\|=C_\xi}\ell\left(\theta_0+(nT/d)^{-1/2}u\right)<\ell(\theta_0)\Big\}\ge 1-\xi.$$

As a result, $(nT/d)^{1/2}\|\hat\theta-\theta_0\|=O_p(1)$, which completes the proof of Step I.

Step II. By the result of STEP I and a Taylor series expansion, we have that $0=\partial\ell(\hat\theta)/\partial\theta=\partial\ell(\theta_0)/\partial\theta+\{\partial^2\ell(\theta_0)/\partial\theta\partial\theta^\top\}(\hat\theta-\theta_0)\{1+o_p(1)\}$. Thus,

$$(nT/d)^{-1/2}\partial\ell(\theta_0)/\partial\theta=-(nT/d)^{1/2}\frac{1}{nT}\{\partial^2\ell(\theta_0)/\partial\theta\partial\theta^\top\}(\hat\theta-\theta_0)\{1+o_p(1)\}.$$

By Lemma 3 (iii), we obtain

$$\Big\|(nT/d)^{1/2}\Big(\frac{1}{nT}\frac{\partial^2\ell(\theta_0)}{\partial\theta\partial\theta^\top}+\mathcal{I}(\theta_0)\Big)(\hat\theta-\theta_0)\Big\|\le(nT/d)^{1/2}\Big\|\frac{1}{nT}\frac{\partial^2\ell(\theta_0)}{\partial\theta\partial\theta^\top}+\mathcal{I}(\theta_0)\Big\|\|\hat\theta-\theta_0\|=o_p(1).$$

Thus, we have

$$-(nT/d)^{1/2}\frac{1}{nT}\frac{\partial^2\ell(\theta_0)}{\partial\theta\partial\theta^\top}(\hat\theta-\theta_0)=(nT/d)^{1/2}\mathcal{I}(\theta_0)(\hat\theta-\theta_0)+o_p(1).$$

This, together with Lemma 3 (i), implies

$$\sqrt{nT/d}D\mathcal{I}(\theta_0)(\hat{\theta}-\theta_0) \xrightarrow{d} N\left(0, G(\theta_0)\right),$$

which completes the entire proof.

Proof of Theorem 2: By definition, $\hat{B}_t = \hat{\lambda}_1 W_1^{(t)} + \cdots + \hat{\lambda}_d W_d^{(t)}$. Thus, $\hat{B}_t - B_t = (\hat{\lambda}_1 - \lambda_{01})W_1^{(t)} + \cdots + (\hat{\lambda}_d - \lambda_{0d})W_d^{(t)}$. By Theorem 1, we have $\|\hat{\lambda} - \lambda_0\| = O_p\{(nT/d)^{-1/2}\}$. By the Triangle inequality and Condition (C3), we obtain

$$\|\hat{B}_t - B_t\| = \|(\hat{\lambda}_1 - \lambda_{01})W_1^{(t)} + \cdots + (\hat{\lambda}_d - \lambda_{0d})W_d^{(t)}\|$$

$$\leq |\hat{\lambda}_1 - \lambda_{01}| \times \|W_1^{(t)}\| + \cdots + |\hat{\lambda}_d - \lambda_{0d}| \times \|W_d^{(t)}\|$$

$$\leq \max_k \|W_k^{(t)}\| \times \sum_{k=1}^{d} |\hat{\lambda}_k - \lambda_{0k}| \leq C_w\sqrt{d}\|\hat{\lambda} - \lambda_0\| = O_p\{d(nT)^{-1/2}\},$$

which completes the proof.

Proof of Theorem 3: To prove the theorem, we consider the following two cases: (i) $\mathcal{S}$ is underfitted; (ii) $\mathcal{S}$ is overfitted.

CASE I: Underfitted models (i.e., $\mathcal{S} \in \mathcal{A}_1$). In this case, we need to prove $P\big\{\min_{\mathcal{S}\in\mathcal{A}_1} \text{EBIC}(\mathcal{S}) \leq \text{EBIC}(\mathcal{S}_T)\big\} \to 0$. Note that $|\mathcal{S}| - |\mathcal{S}_T| \geq$

$-|\mathcal{S}_T| \geq -q$ and $\ell(\hat{\theta}_{\mathcal{S}_T}) \geq \ell(\theta_{0\mathcal{S}_T})$. Then, we have

$$\begin{aligned}
&P\Big\{ \min_{\mathcal{S}\in\mathcal{A}_1} \mathrm{EBIC}(\mathcal{S}) \leq \mathrm{EBIC}(\mathcal{S}_T)\Big\} \\
=&P\Big\{ \min_{\mathcal{S}\in\mathcal{A}_1} \ell(\hat{\theta}_{\mathcal{S}}) - \ell(\hat{\theta}_{\mathcal{S}_T}) \geq \{\log(nT) + \gamma\log(d)\}(|\mathcal{S}| - |\mathcal{S}_T|)/2\Big\} \\
\leq&P\Big\{ \min_{\mathcal{S}\in\mathcal{A}_1} \ell(\hat{\theta}_{\mathcal{S}}) - \ell(\hat{\theta}_{\mathcal{S}_T}) \geq -q/2(\log(nT) + \gamma\log(d))\Big\} \\
\leq&P\Big\{ \min_{\mathcal{S}\in\mathcal{A}_1} \ell(\hat{\theta}_{\mathcal{S}}) - \ell(\theta_{0\mathcal{S}_T}) \geq -q/2(\log(nT) + \gamma\log(d))\Big\}.
\end{aligned}$$

Accordingly, it suffices to show $P\Big\{ \min_{\mathcal{S}\in\mathcal{A}_1} \ell(\hat{\theta}_{\mathcal{S}}) - \ell(\theta_{0\mathcal{S}_T}) \geq -q/2(\log(nT) + \gamma\log(d))\Big\} \to 0$.

For any $\mathcal{S} \in \mathcal{A}_1$, let $\bar{\mathcal{S}} = \mathcal{S} \cup \mathcal{S}_T$. We then consider those $\tilde{\theta}_{\bar{\mathcal{S}}}$ such that $\|\tilde{\theta}_{\bar{\mathcal{S}}} - \theta_{0\bar{\mathcal{S}}}\| = \rho_{nT}$ with $\rho_{nT} \to 0$ and $\rho_{nT}\sqrt{nT/\log(nT)} \to \infty$. By a Taylor series expansion, we have

$$\ell(\tilde{\theta}_{\bar{\mathcal{S}}}) - \ell(\theta_{0\bar{\mathcal{S}}}) = (\tilde{\theta}_{\bar{\mathcal{S}}} - \theta_{0\bar{\mathcal{S}}})^\top \frac{\partial \ell(\theta_{0\bar{\mathcal{S}}})}{\partial \theta_{\bar{\mathcal{S}}}} - \frac{1}{2}(\tilde{\theta}_{\bar{\mathcal{S}}} - \theta_{0\bar{\mathcal{S}}})^\top \Big\{ - \frac{\partial^2 \ell(\tilde{\theta}^*_{\bar{\mathcal{S}}})}{\partial \tilde{\theta}^*_{\bar{\mathcal{S}}} \partial \tilde{\theta}^{*\top}_{\bar{\mathcal{S}}}}\Big\}(\tilde{\theta}_{\bar{\mathcal{S}}} - \theta_{0\bar{\mathcal{S}}}),$$

where $\tilde{\theta}^*_{\bar{\mathcal{S}}}$ lies between $\tilde{\theta}_{\bar{\mathcal{S}}}$ and $\theta_{0\bar{\mathcal{S}}}$. By Condition (C7),

$$\ell(\tilde{\theta}_{\bar{\mathcal{S}}}) - \ell(\theta_{0\bar{\mathcal{S}}}) \leq (\tilde{\theta}_{\bar{\mathcal{S}}} - \theta_{0\bar{\mathcal{S}}})^\top \frac{\partial \ell(\theta_{0\bar{\mathcal{S}}})}{\partial \theta_{\bar{\mathcal{S}}}} - \frac{c_{\min,3}}{2} nT \|\tilde{\theta}_{\bar{\mathcal{S}}} - \theta_{0\bar{\mathcal{S}}}\|^2 \leq \rho_{nT} \Big\| \frac{\partial \ell(\theta_{0\bar{\mathcal{S}}})}{\partial \theta_{\bar{\mathcal{S}}}} \Big\| - \frac{c_{\min,3}}{2} nT \rho_{nT}^2.$$

We next prove that $\sup_{\bar{\mathcal{S}}} \|\partial \ell(\theta_{0\bar{\mathcal{S}}})/\partial \theta_{\bar{\mathcal{S}}}\| = O_p\big\{\sqrt{nT\log(nT)}\big\}$. According to Lemma 3 (ii), for sufficiently large constant c_l,

$$P\Big(\sup_{\bar{\mathcal{S}}}\|\frac{\partial\ell(\theta_{0\bar{\mathcal{S}}})}{\partial\theta_{\bar{\mathcal{S}}}}\| > c_l\sqrt{nT\log(nT)}\Big) \leq \sum_{\bar{\mathcal{S}}} P\Big(\|\frac{\partial\ell(\theta_{0\bar{\mathcal{S}}})}{\partial\theta_{\bar{\mathcal{S}}}}\| > c_l\sqrt{nT\log(nT)}\Big)$$

$$\leq \sum_{\bar{\mathcal{S}}}\sum_{k=1}^{|\bar{\mathcal{S}}|} P\Big(|\frac{\partial\ell(\theta_{0\bar{\mathcal{S}}})}{\partial\theta_{\bar{\mathcal{S}},k}}| > c_l\sqrt{nT\log(nT)/|\bar{\mathcal{S}}|}\Big)$$

$$\leq \sum_{\bar{\mathcal{S}}}\sum_{k=1}^{|\bar{\mathcal{S}}|} 2\exp\left[-\min\Big(\frac{\tau_1 c_l\sigma_0^2\sqrt{\log(nT)/|\bar{\mathcal{S}}|}}{\|U_k\|/\sqrt{nT}}, \frac{\tau_2 c_l^2\sigma_0^4\log(nT)/|\bar{\mathcal{S}}|}{\|U_k\|_F^2/nT}\Big)\right]. \tag{S3.4}$$

By Conditions (C2) and (C3), we have $\|U_k\| = \sup_t\|W_k^{(t)}\Delta_t^{-1}(\lambda_{0\bar{\mathcal{S}}})\| \leq \sup_t\|W_k^{(t)}\|\|\Delta_t^{-1}(\lambda_{0\bar{\mathcal{S}}})\| < \infty$ and $\|U_k\|_F^2 = O(nT)$ uniformly for any $k \leq |\bar{\mathcal{S}}|$. Hence,

$$\frac{\tau_1 c_l\sigma_0^2\sqrt{\log(nT)/|\bar{\mathcal{S}}|}}{\|U_k\|/\sqrt{nT}} = O\Big\{\sqrt{nT\log(nT)}\Big\} > \frac{\tau_2 c_l^2\sigma_0^4\log(nT)/|\bar{\mathcal{S}}|}{\|U_k\|_F^2/nT} = O\{\log(nT)\}.$$

Accordingly, for sufficiently large constants c_l and c_L, we have

$$P\left(\sup_{\bar{\mathcal{S}}}\|\frac{\partial\ell(\theta_{0\bar{\mathcal{S}}})}{\partial\theta_{\bar{\mathcal{S}}}}\| > c_l\sqrt{nT\log(nT)}\right) \leq 2q\exp\Big\{-c_L\log(nT)+q\log(d)\Big\} \to 0,$$

which immediately leads to $\sup_{\bar{\mathcal{S}}}\|\partial\ell(\theta_{0\bar{\mathcal{S}}})/\partial\theta_{\bar{\mathcal{S}}}\| = O_p\big\{\sqrt{nT\log(nT)}\big\}$.

Using the above result, we obtain that, for $\|\tilde{\theta}_{\bar{\mathcal{S}}} - \theta_{0\bar{\mathcal{S}}}\| = \rho_{nT}$,

$$\ell(\tilde{\theta}_{\bar{\mathcal{S}}}) - \ell(\theta_{0\bar{\mathcal{S}}}) \leq \rho_{nT}\Big\{c_l\sqrt{nT\log(nT)} - \frac{c_{\min,3}}{2}\rho_{nT}nT\Big\}$$

uniformly over $\bar{\mathcal{S}}$ and $\tilde{\theta}_{\bar{\mathcal{S}}}$ with probability tending to 1. Since $\rho_{nT}\sqrt{nT/\log(nT)} \to \infty$, we have $\sqrt{nT\log(nT)} = o(\rho_{nT}nT)$. Then $\ell(\tilde{\theta}_{\bar{\mathcal{S}}}) - \ell(\theta_{0\bar{\mathcal{S}}}) \leq -c_l'\rho_{nT}^2 nT$, where $c_l^{'}$ is a positive constant. In addition, the concavity of $\ell(\cdot)$ in a neighborhood of $\theta_{0\bar{\mathcal{S}}}$ implies that

$$\sup\Big\{\ell(\tilde{\theta}_{\bar{\mathcal{S}}})-\ell(\theta_{0\bar{\mathcal{S}}}) : \|\tilde{\theta}_{\bar{\mathcal{S}}}-\theta_{0\bar{\mathcal{S}}}\| \geq \rho_{nT}\Big\} \leq \sup\Big\{\ell(\tilde{\theta}_{\bar{\mathcal{S}}})-\ell(\theta_{0\bar{\mathcal{S}}}) : \|\tilde{\theta}_{\bar{\mathcal{S}}}-\theta_{0\bar{\mathcal{S}}}\| = \rho_{nT}\Big\} \leq -c_l^{'}\rho_{nT}^2 nT$$

with probability approaching 1. Let $\breve{\theta}_{\bar{\mathcal{S}}}$ be $\hat{\theta}_{\mathcal{S}}$ augmented with zeros corresponding to the elements in $\bar{\mathcal{S}}/\mathcal{S}$, and let $\rho_{nT} = \min_{k\in\mathcal{S}_T}|\lambda_{0k}|$. It can be shown that $\|\breve{\theta}_{\bar{\mathcal{S}}} - \hat{\theta}_{\mathcal{S}}\| \geq \rho_{nT}$. Therefore, $\ell(\hat{\theta}_{\mathcal{S}}) - \ell(\theta_{0\mathcal{S}_T}) \leq -c_l^{'}\rho_{nT}^2 nT$ uniformly over $\mathcal{S} \in \mathcal{A}_1$. Since $\rho_{nT}\sqrt{nT/\log(nT)} \to \infty$ and $-c_l^{'}\rho_{nT}^2 nT < -q/2\big\{\log(nT)+\gamma\log(d)\big\}$ for large nT. The above results imply $P\big\{\min_{\mathcal{S}\in\mathcal{A}_1}\ell(\hat{\theta}_{\mathcal{S}})-\ell(\theta_{0\mathcal{S}_T}) \geq -q/2(\log(nT)+\gamma\log(d))\big\} \to 0$, which completes the first part of the proof.

CASE II: Overfitted models (i.e., $\mathcal{S} \in \mathcal{A}_0, \mathcal{S} \neq \mathcal{S}_T$). For any $\mathcal{S} \in \mathcal{A}_0$, let $m = |\mathcal{S}| - |\mathcal{S}_T|$. By definition, EBIC($\mathcal{S}$) $\leq$ EBIC($\mathcal{S}_T$) if and only if $\ell(\hat{\theta}_{\mathcal{S}}) - \ell(\hat{\theta}_{\mathcal{S}_T}) \geq m/2\big\{\log(nT)+\gamma\log(d)\big\}$. For sufficiently large nT,

$$\ell(\hat{\theta}_{\mathcal{S}}) - \ell(\hat{\theta}_{\mathcal{S}_T}) \leq \ell(\hat{\theta}_{\mathcal{S}}) - \ell(\theta_{0\mathcal{S}}) = (\hat{\theta}_{\mathcal{S}} - \theta_{0\mathcal{S}})^\top\frac{\partial\ell(\theta_{0\mathcal{S}})}{\partial\theta_{\mathcal{S}}}$$
$$-\frac{1}{2}(\hat{\theta}_{\mathcal{S}} - \theta_{0\mathcal{S}})^\top\Big(-\frac{\partial^2\ell(\theta^*_{\mathcal{S}})}{\partial\theta^*_{\mathcal{S}}\partial\theta^{*\top}_{\mathcal{S}}}\Big)(\hat{\theta}_{\mathcal{S}} - \theta_{0\mathcal{S}}),$$

where $\theta^*_{\mathcal{S}}$ lies between $\hat{\theta}_{\mathcal{S}}$ and $\theta_{0\mathcal{S}}$. Define $H(\theta^*_{\mathcal{S}}) = -\frac{1}{nT}\frac{\partial^2\ell(\theta^*_{\mathcal{S}})}{\partial\theta^*_{\mathcal{S}}\partial\theta^{*\top}_{\mathcal{S}}}$. By Condition (C7), we have

$$\ell(\hat{\theta}_{\mathcal{S}}) - \ell(\theta_{0\mathcal{S}_T}) = \frac{1}{2nT}\frac{\partial\ell^\top(\theta_{0\mathcal{S}})}{\partial\theta_{\mathcal{S}}}(H(\theta^*_{\mathcal{S}}))^{-1}\frac{\partial\ell(\theta_{0\mathcal{S}})}{\partial\theta_{\mathcal{S}}} \leq \frac{1}{2nTc_{\min,3}}\|\frac{\partial\ell(\theta_{0\mathcal{S}})}{\partial\theta_{\mathcal{S}}}\|^2.$$

Then, employing similar techniques to those used for obtaining (A.4) in the proof of CASE I, we have

$$P\left\{\frac{1}{2nTc_{\min,3}}\sup_{\mathcal{S}\in\mathcal{A}_0,\mathcal{S}\neq\mathcal{S}_T}\|\frac{\partial\ell(\theta_{0\mathcal{S}})}{\partial\theta_{\mathcal{S}}}\|^2\geq m/2\big(\log(nT)+\gamma\log(d)\big)\right\}$$
$$\leq\sum_{\mathcal{S}}\sum_{k=1}^{|\mathcal{S}|}2\exp\left[-\frac{\tau_2c_{\min,3}m\sigma_0^4(\log(nT)+\gamma\log(d))/|\mathcal{S}|}{\|U_k\|_F^2/nT}\right].$$

Since $\|U_k\|_F^2/nT\leq\|U_k\|^2\leq\big(\sup_{t\leq T}\sup_{n\geq1}\|W_k^{(t)}\|\|\Delta^{-1}(\lambda_0)\|\big)^2\leq C_w^2$, we further have that

$$2\exp\left[-\tau_2c_{\min,3}m\sigma_0^4(\log(nT)+\gamma\log(d))/(|\mathcal{S}|\|U_k\|_F^2/nT)\right]$$
$$\leq 2\exp\left[-\tau_2c_{\min,3}m\sigma_0^4(\log(nT)+\gamma\log(d))/qC_w^2\right].$$

The above results imply that

$$P\Big\{\frac{1}{2nTc_{\min,3}}\sup_{\mathcal{S}\in\mathcal{A}_0,\mathcal{S}\neq\mathcal{S}_T}\|\frac{\partial\ell(\theta_{0\mathcal{S}})}{\partial\theta_{\mathcal{S}}}\|^2\geq m/2\big(\log(nT)+\gamma\log(d)\big)\Big\}$$
$$\leq 2q\exp\left[-\tau_2c_{\min,3}m\sigma_0^4(\log(nT)+\gamma\log(d))/qC_w^2+m\log(d)\right]$$
$$\leq 2q\exp\left[-\tau_2c_{\min,3}m\sigma_0^4/(qC_w^2)\{4+\gamma-qC_w^2/(\tau_2c_{\min,3}\sigma_0^4)\}\log(d)\right]\to0$$

for $\gamma>qC_w^2/(\tau_2c_{\min,3}\sigma_0^4)-4$. Accordingly, we have

$$P\Big\{\text{EBIC}(\mathcal{S})\leq\text{EBIC}(\mathcal{S}_T)\Big\}=P\Big\{\sup_{\mathcal{S}\in\mathcal{A}_0,\mathcal{S}\neq\mathcal{S}_T}\ell(\hat\theta_{\mathcal{S}})-\ell(\hat\theta_{\mathcal{S}_T})\geq m/2\big\{\log(nT)+\gamma\log(d)\big\}\Big\}$$
$$\leq P\Big\{\frac{1}{2nTc_{\min,3}}\sup_{\mathcal{S}\in\mathcal{A}_0,\mathcal{S}\neq\mathcal{S}_T}\|\frac{\partial\ell(\theta_{0\mathcal{S}})}{\partial\theta_{\mathcal{S}}}\|^2\geq m/2\big(\log(nT)+\gamma\log(d)\big)\Big\}\to0,$$

which completes the entire proof.

Proof of Theorem 4: To prove the theorem, we decompose T_{ql} into the following three parts,

$$T_{ql} = (nT)^{-1}\sum_{t=1}^{T} tr(Y_tY_t^\top\Sigma_t^{-1}-I_n)^2+2(nT)^{-1}\sum_{t=1}^{T} tr\big\{(Y_tY_t^\top\Sigma_t^{-1}-I_n)(Y_tY_t^\top\hat{\Sigma}_t^{-1}-Y_tY_t^\top\Sigma_t^{-1})\big\}$$

$$+(nT)^{-1}\sum_{t=1}^{T} tr\big\{Y_tY_t^\top\hat{\Sigma}_t^{-1}-Y_tY_t^\top\Sigma_t^{-1})\big\}^2 \triangleq T_{ql}^{(1)}+2T_{ql}^{(2)}+T_{ql}^{(3)}.$$

Then we consider the following three steps: (i) STEP I, showing that $T_{ql}^{(3)}$ is negligible; (ii) STEP II, showing that $T_{ql}^{(2)}$ is of order $O_p(d)$ and then employing a Taylor series expansion to derive the leading term of $T_{ql}^{(2)}$ as a combination of quadratic forms of the random error ϵ_t, which directly connects to the result of Lemma 4; (iii) STEP III, demonstrating the asymptotic normality of T_{ql} based on the results of Lemma 4.

STEP I: Note that $T_{ql}^{(3)} = (nT)^{-1}\sum_{t=1}^{T}\{Y_t^\top(\hat{\Sigma}_t^{-1}-\Sigma_t^{-1})Y_t\}^2 \le \max_t\|\hat{\Sigma}_t^{-1}-\Sigma_t^{-1}\|^2\times(nT)^{-1}\sum_{t=1}^{T}(Y_t^\top Y_t)^2$. By Condition (C3), we obtain $Y_t^\top Y_t = \epsilon_t^\top\{\Delta_t^{-1}(\lambda_0)\}^\top\Delta_t^{-1}(\lambda_0)\epsilon_t \le \max_t\|\{\Delta_t^{-1}(\lambda_0)\}^\top\Delta_t^{-1}(\lambda_0)\|\times\epsilon_t^\top\epsilon_t = O(1)\times\epsilon_t^\top\epsilon_t = O_p(n)$. Then $(nT)^{-1}\sum_{t=1}^{T}(Y_t^\top Y_t)^2 = O_p(n)$. In addition, $\hat{\Sigma}_t^{-1} = (\hat{\sigma}^2)^{-1}\Delta_t^\top(\hat{\lambda})\Delta_t(\hat{\lambda})$ and $\Sigma_t^{-1} = (\sigma_0^2)^{-1}\Delta_t^\top(\lambda_0)\Delta_t(\lambda_0)$. Thus, $\hat{\Sigma}_t^{-1}-\Sigma_t^{-1} = (\sigma_0^2)^{-1}\big\{\Delta_t^\top(\hat{\lambda})\Delta_t(\hat{\lambda})-\Delta_t^\top(\lambda_0)\Delta_t(\lambda_0)\big\}+\{(\hat{\sigma}^2)^{-1}-(\sigma_0^2)^{-1}\}\Delta_t^\top(\hat{\lambda})\Delta_t(\hat{\lambda}) \triangleq \Delta_{11t}+\Delta_{12t}$. Note that $\|\Delta_{11t}\| = \|(\sigma_0^2)^{-1}\big\{\Delta_t^\top(\hat{\lambda})\Delta_t(\hat{\lambda})-\Delta_t^\top(\hat{\lambda})\Delta_t(\lambda_0)+\Delta_t^\top(\hat{\lambda})\Delta_t(\lambda_0)-\Delta_t^\top(\lambda_0)\Delta_t(\lambda_0)\big\}\| \le (\sigma_0^2)^{-1}\|\Delta_t^\top(\hat{\lambda})\|\times\|\Delta_t(\hat{\lambda})-\Delta_t(\lambda_0)\|+(\sigma_0^2)^{-1}\|\Delta_t(\lambda_0)\|\times\|\Delta_t(\hat{\lambda})-\Delta_t(\lambda_0)\| \le (\sigma_0^2)^{-1}(\|\Delta_t^\top(\lambda)\|+\|\Delta_t(\hat{\lambda})-\Delta_t(\lambda_0)\|)\times$

$\|\Delta_t(\hat{\lambda}) - \Delta_t(\lambda_0)\| + (\sigma_0^2)^{-1}\|\Delta_t(\lambda_0)\| \times \|\Delta_t(\hat{\lambda}) - \Delta_t(\lambda_0)\|$. By Condition (C3), we have $\max_t \|\Delta_t^\top(\lambda)\| = O(1)$. Moreover, by the results of Theorem 2, $\max_t \|\Delta_t(\hat{\lambda}) - \Delta_t(\lambda_0)\| = O_p\{d(nT)^{-1/2}\}$, which leads to $\max_t \|\Delta_{11t}\| \leq (\sigma_0^2)^{-1}\max_t(\|\Delta_t^\top(\lambda)\| + \|\Delta_t(\hat{\lambda}) - \Delta_t(\lambda_0)\|) \times \|\Delta_t(\hat{\lambda}) - \Delta_t(\lambda_0)\| + (\sigma_0^2)^{-1}\max_t \|\Delta_t(\lambda_0)\| \times \|\Delta_t(\hat{\lambda}) - \Delta_t(\lambda_0)\| = O_p\big\{\|\Delta_t(\hat{\lambda}) - \Delta_t(\lambda_0)\|\big\} = O_p\{d(nT)^{-1/2}\}$. Analogously, we can show that $\max_t \Delta_{12t} = O_p\{d(nT)^{-1/2}\}$. Combining the above results, we have $\max_t \|\hat{\Sigma}_t - \Sigma_t\| = O_p\{d(nT)^{-1/2}\}$, which leads to $T_{ql}^{(3)} \leq \max_t \|\hat{\Sigma}_t^{-1} - \Sigma_t^{-1}\|^2 \times (nT)^{-1}\sum_{t=1}^T (Y_t^\top Y_t)^2 = O_p\{d(nT)^{-1}\} \times O_p(n) = o_p(1)$. The proof is complete.

STEP II: Note that $T_{ql}^{(2)} = (nT)^{-1}\sum_{t=1}^T Y_t^\top(\hat{\Sigma}_t^{-1} - \Sigma_t^{-1})Y_t\{Y_t^\top\Sigma_t^{-1}Y_t - 1\}$. In addition, $Y_t^\top\Sigma_t^{-1}Y_t = \epsilon_t^\top\epsilon_t = n\sigma_0^2\{1 + o_p(1)\} = O_p(n)$. By the theorem's assumption that $n/T \to c > 0$, we have $Y_t^\top(\hat{\Sigma}_t^{-1} - \Sigma_t^{-1})Y_t \leq \varrho_{\max}(\hat{\Sigma}_t^{-1} - \Sigma_t^{-1})Y_t^\top Y_t = O_p\{d(nT)^{-1/2}\} \times O_p(n) = O_p(dn^{1/2}T^{-1/2}) = O_p(d)$. Accordingly, $T_{ql}^{(2)} = O_p(d)$, and it is non-negligible. Hence, we need to evaluate the mean and variance of $T_{ql}^{(2)}$ for the bias-correction. However, $T_{ql}^{(2)}$ involves the estimator $\hat{\lambda}$, which does not have a closed form. As a result, it is not plausible to obtain explicit formulas for $E\{T_{ql}^{(2)}\}$ and $\mathrm{Var}\{T_{ql}^{(2)}\}$. Thus, we apply a Taylor series expansion with respect to $\hat{\theta} = (\hat{\lambda}^\top, \hat{\sigma}^2)^\top$ given below to obtain the leading term of $T_{ql}^{(2)}$.

Let $\widetilde{T}_{ql}^{(2)} = T^{-1}\sum_{t=1}^T Y_t^\top(\hat{\Sigma}_t^{-1} - \Sigma_t^{-1})Y_t$. We then obtain $T_{ql}^{(2)} = \widetilde{T}_{ql}^{(2)} +$

$o_p(1)$. Thus, we only need to consider $\widetilde{T}_{ql}^{(2)}$ in the rest of proof. Define $\Sigma_t^{-1} \triangleq \Sigma_t^{-1}(\theta_0) = (\sigma_0^2)^{-1}\Delta_t^\top(\lambda_0)\Delta_t(\lambda_0)$ and $\hat{\Sigma}_t^{-1} \triangleq \Sigma_t^{-1}(\hat{\theta}) = (\hat{\sigma}^2)^{-1}\Delta_t^\top(\hat{\lambda})\Delta_t(\hat{\lambda})$. By Theorem 1, we have $\hat{\theta} - \theta_0 = \mathcal{I}^{-1}(\theta_0)(nT)^{-1}\frac{\partial \ell(\theta_0)}{\partial \theta}\{1 + o_p(1)\}$. Furthermore, employing a Taylor series expansion, we have

$$\widetilde{T}_{ql}^{(2)} = T^{-1}\sum_{t=1}^T \text{vec}^\top\{\Sigma_t^{-1}(\hat{\theta}) - \Sigma_t^{-1}(\theta_0)\}\text{vec}(Y_tY_t^\top)$$

$$= T^{-1}\sum_{t=1}^T (\hat{\theta} - \theta_0)^\top \frac{\partial \text{vec}^\top\{\Sigma_t^{-1}(\theta_0)\}}{\partial \theta}\text{vec}(Y_tY_t^\top) + o_p(1)$$

$$= n^{-1}T^{-2}\sum_{t=1}^T \frac{\partial^\top \ell(\theta_0)}{\partial \theta}\mathcal{I}^{-1}(\theta_0)\Lambda_t^\top \text{vec}(Y_tY_t^\top) + o_p(1),$$

where $\Lambda_t = \partial\text{vec}\{\Sigma_t^{-1}(\theta_0)\}/\partial\theta = (\partial\text{vec}\{\Sigma_t^{-1}(\theta_0)\}/\partial\lambda_1, \cdots, \partial\text{vec}\{\Sigma_t^{-1}(\theta_0)\}/\partial\sigma^2) \in \mathbb{R}^{n^2\times(d+1)}$. It is worth noting that

$$\Lambda_t^\top \text{vec}(Y_tY_t^\top) = (Y_t^\top \widetilde{\Lambda}_{t1}Y_t, \cdots, Y_t^\top \widetilde{\Lambda}_{t(d+1)}Y_t)^\top$$

$$= (\epsilon_t^\top\{\Delta_t^{-1}(\lambda_0)\}^\top \widetilde{\Lambda}_{t1}\Delta_t^{-1}(\lambda_0)\epsilon_t, \cdots, \epsilon_t^\top\{\Delta_t^{-1}(\lambda_0)\}^\top \widetilde{\Lambda}_{t(d+1)}\Delta_t^{-1}(\lambda_0)\epsilon_t)^\top \in \mathbb{R}^{d+1},$$

where $\widetilde{\Lambda}_{tk}$ is the matrix form of $\partial\text{vec}\{\Sigma_t^{-1}(\theta_0)\}/\partial\theta_k$ for $k = 1, \cdots, d+1$. In addition, $\frac{\partial \ell(\theta_0)}{\partial \theta} = \sum_{t=1}^T(\epsilon_t^\top U_{t1}\epsilon_t/\sigma_0^2 - tr(U_{t1}), \cdots, \epsilon_t^\top U_{t(d+1)}\epsilon_t/\sigma_0^2 - tr(U_{t(d+1)}))^\top$, where $U_{tk} = s\{W_k^{(t)}\Delta_t^{-1}(\lambda_0)\}$, for $k = 1, \cdots, d$, and $U_{t(d+1)} = I_n/(2\sigma_0^2)$. Accordingly,

$$\widetilde{T}_{ql}^{(2)} = n^{-1}T^{-2}\sum_{t_1=1}^T\sum_{t_2=1}^T \Big(\epsilon_{t_1}^\top U_{t_11}\epsilon_{t_1}/\sigma_0^2 - tr(U_{t_11}), \cdots, \epsilon_{t_1}^\top U_{t_1(d+1)}\epsilon_{t_1}/\sigma_0^2 - tr(U_{t_1(d+1)})\Big)$$

$$\times \mathcal{I}^{-1}(\theta_0)(\epsilon_{t_2}^\top V_{t_21}\epsilon_{t_2}, \cdots, \epsilon_{t_2}^\top V_{t_2(d+1)}\epsilon_{t_2})^\top + o_p(1),$$

where $V_{tk} = \{\Delta_t^{-1}(\lambda_0)\}^\top \widetilde{\Lambda}_{tk} \Delta_t^{-1}(\lambda_0)$ for $k = 1, \cdots, d+1$. Consequently, the leading term of $T_{ql}^{(2)}$ is a combination of quadratic forms of the random errors ϵ_t, which allows us to directly apply the results from Lemma 4 and complete this part of the proof.

STEP III: Note that $T_{ql}^{(1)} = (nT)^{-1} \sum_{t=1}^T tr(\epsilon_t \epsilon_t^\top / \sigma_0^2 - I_n)^2 = (nT)^{-1} \sum_t (\epsilon_t^\top \epsilon_t / \sigma_0^2)^2 + (nT)^{-1} \sum_t (n - 2\epsilon_t^\top \epsilon_t / \sigma_0^2) \triangleq T_{ql}^{(11)} + T_{ql}^{(12)}$. Since $E\{T_{ql}^{(12)}\} = -1$ and $\mathrm{Var}\{T_{ql}^{(12)}\} = (nT)^{-2} \sum_t \mathrm{Var}(\epsilon_t^\top \epsilon_t / \sigma_0^2) = O\{(nT)^{-1}\} = o(1)$, we obtain that $T_{ql}^{(1)} = T_{ql}^{(11)} - 1 + o_p(1)$. This, together with the results from STEPs I–II, leads to

$$T_{ql} = T_{ql}^{(11)} + 2\widetilde{T}_{ql}^{(2)} - 1 + o_p(1) = (nT)^{-1} \sum_{t=1}^T (\epsilon_t^\top \epsilon_t / \sigma_0^2)^2 - 1$$

$$+2n^{-1}T^{-2} \sum_{t_1=1}^T \sum_{t_2=1}^T \Big(\epsilon_{t_1}^\top U_{t_1 1} \epsilon_{t_1} / \sigma_0^2 - tr(U_{t_1 1}), \cdots, \epsilon_{t_1}^\top U_{t_1 (d+1)} \epsilon_{t_1} / \sigma_0^2 - tr(U_{t_1 (d+1)}) \Big)$$

$$\times \mathcal{I}^{-1}(\theta_0) (\epsilon_{t_2}^\top V_{t_2 1} \epsilon_{t_2}, \cdots, \epsilon_{t_2}^\top V_{t_2 (d+1)} \epsilon_{t_2})^\top + o_p(1)$$

$$= (nT)^{-1} \sum_{t=1}^T (\epsilon_t^\top \epsilon_t / \sigma_0^2)^2 - 1 + 2n^{-1}T^{-2} \sum_{t_1=1}^T \sum_{t_2=1}^T \sum_{k,l}^{d+1} \mathcal{I}_{kl}^{-1}(\theta_0) \{\epsilon_{t_1}^\top U_{t_1 k} \epsilon_{t_1} / \sigma_0^2$$

$$- tr(U_{t_1 k})\} \epsilon_{t_2}^\top V_{t_2 l} \epsilon_{t_2} + o_p(1)$$

with $\mathcal{I}^{-1}(\theta_0) = (\mathcal{I}_{kl}^{-1}(\theta_0))$. One can further apply Condition (C3) to verify that $\max_t \max_k \|U_{tk}\| < \infty$ and $\max_t \max_k \|V_{tk}\| < \infty$. Moreover, $\mathcal{I}^{-1}(\theta_0)$ is a positive definite matrix with bounded upper eigenvalues by Condition (C4). Finally, we apply Lemma 4 and demonstrate that T_{ql} is asymptotic normal, which completes the entire proof.

S4 Additional Simulation Results

In this section, we consider five settings for additional simulation studies. Setting I is used to study a larger weight matrices with $d = 12$. Settings II and III introduce distributions with non-normal errors (i.e., exponential and mixture normal) to demonstrate the robustness of our proposed methods. Setting IV is designed to demonstrate the robustness of EBIC with different choices of γ. Section V presents simulation results with more dense weight matrices.

Setting I: Weight Matrices with $d = 12$. In addition to $d = 2$ and 6 presented in the paper, we consider weight matrices with $d = 12$. The Monte Carlo settings are the same as those in Section 4 of the manuscript, and the simulation results are summarized in Table S.9. We find that the results yield similar patterns to those in Table 1 of the manuscript, which indicates that our method is applicable for larger weight matrices.

Setting II: Exponential Errors. The simulation settings are the same as those in Section 4 of the manuscript, except that the random error terms are iid from an exponential distribution ($\exp(1) - 1$). The simulation results for assessing the finite sample performance of the QMLE estimates, the EBIC criterion and the influence matrix test are presented in Tables S.2-S.4, respectively. We find that the results yield similar patterns to those in Tables 1–3 of the manuscript, which demonstrate the robustness of the parameter estimate, weight matrix selection and the test statistic against non-normal errors.

Setting III: Mixture Normal Errors. The simulation settings are the same as those in Section 4 of the manuscript, except that the random error terms are iid from a mixture normal distribution $(0.9N(0,5/9)+0.1N(0,5))$. The simulation results for assessing the finite sample performance of the QMLE estimates, the EBIC criterion and the influence matrix test are given in Tables S.5-S.7, respectively. The results show similar patterns to those in Tables 1–3 of the manuscript, which further demonstrate the robustness of our proposed methods.

Setting IV: EBIC with Different γs. The simulation settings are the same as those in Section 4 of the manuscript. We consider three values of γ to implement EBIC, i.e., $\gamma = 0.5, 1$ and 2. All simulation results are summarized in Table S.8. This table shows that the model selection results for different values of γ are comparable, which indicates that EBIC is robust for different choices of γ.

Setting V: More Dense Weight Matrices. The simulation settings are the same as those in Section 4 of the manuscript, except that the density of $A_k^{(t)}$ (i.e., the proportion of nonzero elements) is defined as $20/n$ for any k

and t. The results are summarized in Table S.9, and they indicate that the performance gets worse (i.e., standard error increases) as the density of the weight matrices increases. This finding is expected, since Condition (C3) requires that the maximum absolute row-sum and column norms of weight matrices are bounded. Accordingly, (C3) is well satisfied if the weight matrices are extremely sparse and may be violated if the weight matrices are relatively dense.

REFERENCES

Bao, Y. and Ullah, A. (2010). *Expectation of quadratic forms in normal and nonnormal variables with applications. Journal of Statistical Planning and Inference* **140**, 1193–1205.

Chen, J. and Chen, Z. (2012). Extended BIC for small-n-large-p sparse GLM. *Statistica Sinica* **22**, 555–574.

Chen, S., Zhang, L. and Zhong, P. (2010). Tests for high dimensional covariance matrices. *Journal of the American Statistical Association* **105**, 810–819.

Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical*

Association **96**, 1348–1360.

Gupta, A. and Robinson, P. (2015). Inference on higher-order spatial autoregressive models with increasingly many parameters. *Journal of Econometrics* **186**, 19–31.

Gupta, A. and Robinson, P. (2018). Pseudo maximum likelihood estimation of spatial autoregressive models with increasing dimension. *Journal of Econometrics* **202**, 92–107.

Hall, P. and Heyde, C. (1980). *Martingale Limit Theory and Its Application*, New York: Academic Press.

Hanson, D. L. and Wright, E. T. (1971). A bound on tail probabilities for quadratic forms in independent random variables, *Annals of Mathematical Statistics* **42**,1079–1083.

Kelejian, H. and Prucha, I. (2001). On the asymptotic distribution of the Moran I test statistic with applications.*Journal of Econometrics* **104**, 219–257.

Lee, L. F. (2004). Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. *Econometrica* **72**, 1899–1925.

Wright, E. T. (1973). A bound on tail probabilities for quadratic forms in independent random variables whose distributions are not necessarily symmetric. *Annals of Probability* **1**, 1068–1070.

Zhou, J., Tu, Y., Chen, Y. and Wang, H. (2017). Estimating spatial autocorrelation with sampled network data. *Journal of Business & Economic Statistics* **35**, 130–138.

Table S.1: The bias and standard error of the parameter estimates when the true parameters are $\lambda_k = 0.1$ for $k = 1, \cdots, 12$, and the random errors are normally distributed. BIAS: the average bias; SE: the average of the estimated standard errors via Theorem 1; SE*: the standard error of parameter estimates calculated from 500 realizations.

n	T		λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	λ_7	λ_8	λ_9	λ_{10}	λ_{11}	λ_{12}
25	25	BIAS	-0.000	-0.006	0.001	-0.002	-0.005	-0.001	-0.001	0.004	-0.003	-0.001	-0.003	-0.004
		SE	0.052	0.052	0.052	0.052	0.052	0.052	0.052	0.052	0.052	0.052	0.052	0.052
		SE*	0.055	0.050	0.053	0.053	0.048	0.055	0.053	0.051	0.054	0.053	0.056	0.053
25	50	BIAS	-0.002	-0.004	-0.001	-0.000	-0.003	-0.003	0.002	-0.001	-0.001	-0.001	0.002	-0.002
		SE	0.037	0.037	0.037	0.037	0.037	0.037	0.037	0.037	0.037	0.037	0.037	0.037
		SE*	0.039	0.036	0.037	0.035	0.038	0.037	0.037	0.035	0.037	0.036	0.037	0.037
25	100	BIAS	0.000	0.001	-0.002	-0.002	-0.000	-0.001	-0.003	0.001	0.001	-0.001	0.001	-0.002
		SE	0.026	0.026	0.027	0.026	0.026	0.026	0.026	0.026	0.026	0.027	0.026	0.027
		SE*	0.027	0.026	0.026	0.026	0.027	0.027	0.025	0.028	0.027	0.026	0.026	0.027
50	25	BIAS	-0.001	-0.004	0.001	-0.004	-0.002	0.001	-0.004	-0.000	-0.004	0.002	-0.003	-0.003
		SE	0.032	0.032	0.032	0.032	0.032	0.032	0.032	0.032	0.032	0.032	0.032	0.032
		SE*	0.032	0.033	0.032	0.032	0.033	0.035	0.031	0.033	0.034	0.032	0.033	0.031
50	50	BIAS	0.000	-0.002	-0.001	-0.001	-0.001	-0.002	-0.000	-0.001	-0.000	-0.001	-0.000	-0.001
		SE	0.023	0.023	0.023	0.023	0.023	0.023	0.023	0.023	0.023	0.023	0.023	0.023
		SE*	0.024	0.023	0.024	0.023	0.023	0.021	0.022	0.022	0.022	0.022	0.023	0.022
50	100	BIAS	-0.001	0.001	-0.002	0.000	-0.001	-0.000	-0.001	-0.000	0.000	0.000	-0.002	0.000
		SE	0.016	0.016	0.016	0.016	0.016	0.016	0.016	0.016	0.016	0.016	0.016	0.016
		SE*	0.016	0.017	0.016	0.016	0.016	0.016	0.017	0.016	0.016	0.016	0.017	0.016
100	25	BIAS	0.000	-0.000	-0.001	-0.001	-0.003	-0.002	0.001	-0.001	-0.001	-0.001	-0.001	-0.000
		SE	0.021	0.021	0.021	0.021	0.021	0.021	0.021	0.021	0.021	0.021	0.021	0.021
		SE*	0.022	0.021	0.022	0.022	0.022	0.021	0.020	0.022	0.022	0.021	0.021	0.021
100	50	BIAS	-0.001	-0.000	-0.000	-0.001	-0.001	-0.001	-0.002	-0.000	-0.001	0.001	-0.001	-0.000
		SE	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015
		SE*	0.015	0.015	0.015	0.015	0.016	0.015	0.014	0.015	0.015	0.014	0.016	0.014
100	100	BIAS	0.000	-0.000	-0.000	-0.001	0.000	0.000	-0.000	0.000	-0.001	0.000	-0.000	0.000
		SE	0.011	0.011	0.011	0.011	0.011	0.011	0.011	0.011	0.011	0.011	0.011	0.011
		SE*	0.011	0.011	0.010	0.011	0.011	0.010	0.011	0.010	0.011	0.011	0.010	0.010

Table S.2: The bias and standard error of the parameter estimates when the true parameters are $\lambda_k = 0.1$ for $k = 1, \cdots, d$, and the random errors follow a standardized exponential distribution. BIAS: the average bias; SE: the average of the estimated standard errors via Theorem 1; SE*: the standard error of parameter estimates calculated from 500 realizations.

			$d=2$		$d=6$					
n	T		λ_1	λ_2	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6
25	25	BIAS	-0.001	-0.002	-0.004	-0.000	-0.011	0.001	0.003	-0.002
		SE	0.055	0.055	0.055	0.055	0.055	0.055	0.055	0.055
		SE*	0.059	0.058	0.054	0.053	0.063	0.061	0.053	0.053
25	50	BIAS	0.003	-0.000	-0.000	-0.001	-0.003	0.003	-0.004	-0.001
		SE	0.039	0.039	0.039	0.039	0.039	0.039	0.039	0.039
		SE*	0.035	0.037	0.040	0.035	0.041	0.038	0.037	0.035
25	100	BIAS	0.002	0.004	-0.002	0.001	0.001	-0.003	-0.003	0.001
		SE	0.027	0.027	0.028	0.028	0.027	0.028	0.028	0.028
		SE*	0.026	0.026	0.026	0.028	0.027	0.026	0.032	0.028
50	25	BIAS	0.000	0.002	-0.004	-0.006	0.003	0.006	-0.000	-0.005
		SE	0.038	0.038	0.037	0.037	0.037	0.037	0.037	0.037
		SE*	0.037	0.041	0.033	0.031	0.041	0.035	0.033	0.038
50	50	BIAS	-0.001	0.001	0.001	-0.004	0.004	-0.005	-0.007	0.006
		SE	0.027	0.027	0.026	0.026	0.026	0.026	0.026	0.026
		SE*	0.027	0.026	0.024	0.026	0.030	0.026	0.026	0.026
50	100	BIAS	0.001	0.001	-0.002	-0.002	-0.000	0.002	0.000	-0.000
		SE	0.019	0.019	0.019	0.018	0.019	0.018	0.018	0.018
		SE*	0.019	0.020	0.018	0.020	0.019	0.019	0.018	0.016
100	25	BIAS	0.003	-0.002	-0.001	-0.003	0.001	-0.004	-0.003	-0.004
		SE	0.027	0.027	0.026	0.026	0.026	0.026	0.026	0.026
		SE*	0.026	0.026	0.025	0.026	0.027	0.025	0.025	0.026
100	50	BIAS	0.000	-0.001	-0.003	-0.001	0.002	-0.002	0.000	0.000
		SE	0.019	0.019	0.018	0.018	0.018	0.018	0.018	0.018
		SE*	0.018	0.018	0.019	0.019	0.019	0.018	0.018	0.016
100	100	BIAS	0.000	-0.000	0.000	-0.001	-0.001	0.000	-0.001	-0.000
		SE	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013
		SE*	0.013	0.014	0.015	0.013	0.012	0.014	0.011	0.015

Table S.3: Model selection via EBIC when $d = 8$ and the random errors are distributed as standardized exponential. AS: the average size of the selected model; CT: the average percentage of the correct fit; TPR: the average true positive rate; FPR: the average false positive rate.

n	T	AS	CT	TPR	FPR
25	25	3.2	75.8	90.6	9.9
	50	3.2	80.4	96.3	7.6
	100	3.1	83.7	99.8	5.4
50	25	3.2	77.6	92.7	8.2
	50	3.1	82.9	98.2	6.8
	100	3.1	85.4	100.0	4.7
100	25	3.2	80.3	96.8	7.2
	50	3.1	85.0	100.0	4.8
	100	3.0	88.5	100.0	3.9

Table S.4: The empirical sizes and powers of the influence matrix test. The case of $\kappa = 0$ corresponds to the null model and $\kappa > 0$ represents alternative models. The random errors are distributed as standardized exponential, and the full model sizes are $d = 2$ and 6.

			d=2			d=6	
n	T	κ=0	κ=0.1	κ=0.2	κ=0	κ=0.1	κ=0.2
25	25	0.024	0.268	0.654	0.026	0.232	0.564
	50	0.028	0.482	0.780	0.030	0.376	0.724
	100	0.040	0.602	0.902	0.038	0.548	0.820
50	25	0.028	0.386	0.746	0.030	0.346	0.656
	50	0.032	0.562	0.872	0.038	0.480	0.776
	100	0.044	0.642	0.946	0.044	0.624	0.922
100	25	0.032	0.522	0.902	0.030	0.424	0.876
	50	0.040	0.704	0.986	0.036	0.568	0.928
	100	0.056	0.880	1.000	0.048	0.822	1.000

Table S.5: The bias and standard error of the parameter estimates when the true parameters are $\lambda_k = 0.1$ for $k = 1, \cdots, d$, and the random errors follow a mixture normal distribution. BIAS: the average bias; SE: the average of the estimated standard errors via Theorem 1; SE*: the standard error of parameter estimates calculated from 500 realizations.

			$d=2$		$d=6$					
n	T		λ_1	λ_2	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6
25	25	BIAS	-0.008	0.003	-0.007	-0.005	-0.000	-0.002	-0.002	0.001
		SE	0.055	0.054	0.055	0.055	0.055	0.055	0.055	0.055
		SE*	0.056	0.054	0.054	0.057	0.057	0.054	0.055	0.054
25	50	BIAS	0.002	-0.002	-0.004	0.001	0.005	-0.003	-0.000	-0.004
		SE	0.038	0.038	0.039	0.039	0.039	0.039	0.039	0.039
		SE*	0.041	0.038	0.039	0.041	0.039	0.039	0.042	0.039
25	100	BIAS	-0.002	0.002	-0.002	-0.001	-0.001	0.002	-0.002	-0.001
		SE	0.027	0.027	0.027	0.027	0.027	0.027	0.027	0.027
		SE*	0.027	0.027	0.027	0.025	0.026	0.024	0.027	0.027
50	25	BIAS	0.003	-0.006	-0.000	-0.001	-0.002	0.001	0.002	-0.002
		SE	0.038	0.038	0.037	0.037	0.037	0.037	0.037	0.037
		SE*	0.039	0.038	0.038	0.033	0.035	0.039	0.038	0.035
50	50	BIAS	0.001	-0.002	-0.001	-0.002	0.001	-0.002	-0.002	-0.001
		SE	0.027	0.027	0.026	0.026	0.026	0.026	0.026	0.026
		SE*	0.025	0.028	0.028	0.028	0.027	0.026	0.027	0.026
50	100	BIAS	0.000	-0.001	0.001	-0.000	0.002	-0.000	-0.001	-0.002
		SE	0.019	0.019	0.018	0.018	0.018	0.018	0.018	0.018
		SE*	0.019	0.019	0.018	0.020	0.018	0.016	0.018	0.019
100	25	BIAS	0.002	-0.004	-0.000	-0.000	-0.002	0.000	-0.001	-0.001
		SE	0.026	0.027	0.026	0.026	0.026	0.026	0.026	0.026
		SE*	0.027	0.026	0.024	0.024	0.028	0.027	0.025	0.027
100	50	BIAS	-0.001	-0.002	0.002	0.002	-0.001	-0.003	0.001	-0.001
		SE	0.019	0.019	0.018	0.018	0.018	0.018	0.018	0.018
		SE*	0.020	0.019	0.019	0.019	0.019	0.018	0.018	0.018
100	100	BIAS	-0.000	-0.001	-0.000	0.000	-0.000	0.000	-0.000	-0.002
		SE	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013
		SE*	0.013	0.014	0.012	0.013	0.013	0.012	0.012	0.014

Table S.6: Model selection via EBIC when $d = 8$ and the random errors are distributed as mixture normal. AS: the average size of the selected model; CT: the average percentage of the correct fit; TPR: the average true positive rate; FPR: the average false positive rate.

n	T	AS	CT	TPR	FPR
25	25	3.2	79.2	94.6	8.1
	50	3.1	83.4	98.7	6.3
	100	3.0	87.6	100.0	4.4
50	25	3.2	83.5	97.1	6.6
	50	3.1	86.1	100.0	5.1
	100	3.0	89.3	100.0	3.7
100	25	3.1	86.5	100.0	4.7
	50	3.0	88.7	100.0	4.0
	100	3.0	91.1	100.0	3.1

Table S.7: The empirical sizes and powers of the influence matrix test. The case of $\kappa = 0$ corresponds to the null model and $\kappa > 0$ represents alternative models. The random errors are distributed as mixture normal, and the full model sizes are $d = 2$ and 6.

		d=2			d=6		
n	T	κ=0	κ=0.1	κ=0.2	κ=0	κ=0.1	κ=0.2
25	25	0.026	0.266	0.640	0.024	0.242	0.574
	50	0.032	0.482	0.806	0.030	0.428	0.714
	100	0.042	0.592	0.902	0.040	0.570	0.804
50	25	0.028	0.374	0.748	0.030	0.378	0.620
	50	0.038	0.542	0.856	0.032	0.504	0.768
	100	0.044	0.662	0.986	0.046	0.644	0.952
100	25	0.030	0.448	0.918	0.034	0.408	0.882
	50	0.034	0.650	0.944	0.042	0.564	0.924
	100	0.048	0.842	1.000	0.040	0.802	1.000

Table S.8: Model selection via EBIC for different choices of γ when $d = 8$ and the random errors are normally distributed. AS: the average size of the selected model; CT: the average percentage of the correct fit; TPR: the average true positive rate; FPR: the average false positive rate.

		γ=2				γ=1				γ=0.5			
n	T	AS	CT	TPR	FPR	AS	CT	TPR	FPR	AS	CT	TPR	FPR
25	25	3.3	72.6	91.8	9.8	3.3	73.2	92.3	9.9	3.4	72.9	93.2	10.2
	50	3.2	77.1	95.7	8.5	3.2	76.9	96.7	8.7	3.3	77.2	97.0	9.0
	100	3.1	81.2	100.0	5.9	3.1	81.7	100.0	6.2	3.1	80.9	100.0	6.4
50	25	3.2	78.2	94.0	7.9	3.2	77.8	94.8	8.4	3.3	78.5	95.3	8.6
	50	3.1	80.8	97.2	5.8	3.1	81.1	97.7	6.3	3.2	80.5	98.2	6.8
	100	3.1	84.7	100.0	5.1	3.1	84.5	100.0	5.6	3.1	85.1	100.0	6.0
100	25	3.1	82.3	100.0	6.7	3.1	82.4	100.0	7.1	3.2	82.0	100.0	7.4
	50	3.1	83.8	100.0	5.1	3.1	83.2	100.0	5.6	3.1	84.0	100.0	5.9
	100	3.0	87.7	100.0	4.2	3.0	87.5	100.0	4.3	3.0	87.6	100.0	4.4

Table S.9: The bias and standard error of the parameter estimates when the true parameters are $\lambda_k = 0.1$ for $k = 1, \cdots, d$, the random errors follow a normal distribution and the densities of the weight matrices are set to be $20/n$. BIAS: the average bias; SE: the average of the estimated standard errors via Theorem 1; SE*: the standard error of parameter estimates calculated from 500 realizations.

			$d=2$		$d=6$					
n	T		λ_1	λ_2	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6
25	25	BIAS	-0.022	0.004	0.001	-0.003	-0.001	-0.008	0.001	-0.009
		SE	0.113	0.112	0.132	0.132	0.132	0.132	0.131	0.132
		SE*	0.113	0.115	0.135	0.123	0.137	0.155	0.134	0.143
25	50	BIAS	-0.009	-0.005	0.004	0.008	-0.008	-0.010	0.004	-0.005
		SE	0.079	0.080	0.094	0.093	0.094	0.094	0.094	0.094
		SE*	0.079	0.083	0.092	0.099	0.093	0.090	0.088	0.094
25	100	BIAS	-0.002	0.002	0.006	-0.003	-0.001	-0.008	0.009	-0.006
		SE	0.056	0.056	0.066	0.067	0.066	0.066	0.066	0.066
		SE*	0.057	0.057	0.069	0.062	0.074	0.071	0.071	0.069
50	25	BIAS	-0.005	-0.001	0.001	-0.001	0.010	-0.007	-0.007	0.001
		SE	0.073	0.073	0.078	0.078	0.077	0.078	0.078	0.078
		SE*	0.075	0.074	0.082	0.077	0.080	0.081	0.071	0.079
50	50	BIAS	-0.009	-0.004	-0.003	0.002	-0.002	-0.000	-0.001	-0.002
		SE	0.052	0.052	0.055	0.055	0.055	0.055	0.055	0.055
		SE*	0.054	0.052	0.059	0.057	0.056	0.054	0.059	0.053
50	100	BIAS	-0.000	-0.004	-0.001	0.002	-0.004	0.001	-0.001	0.001
		SE	0.036	0.036	0.039	0.039	0.039	0.039	0.039	0.039
		SE*	0.036	0.034	0.038	0.041	0.042	0.037	0.039	0.038
100	25	BIAS	-0.001	-0.002	-0.005	0.001	-0.002	0.001	-0.007	0.006
		SE	0.048	0.048	0.048	0.048	0.048	0.048	0.048	0.048
		SE*	0.048	0.046	0.045	0.050	0.046	0.045	0.047	0.050
100	50	BIAS	-0.002	-0.006	0.001	0.001	-0.002	-0.003	-0.000	0.002
		SE	0.034	0.034	0.034	0.034	0.034	0.034	0.034	0.034
		SE*	0.033	0.034	0.032	0.033	0.032	0.033	0.033	0.034
100	100	BIAS	-0.003	0.000	0.000	-0.002	0.001	0.002	0.002	-0.005
		SE	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024
		SE*	0.024	0.025	0.024	0.023	0.022	0.024	0.024	0.023