Supplementary Material for "A Naive Least Squares Method for Spatial Autoregression with Covariates"

Yingying Ma[1], Rui Pan[2], Tao Zou[3] and Hansheng Wang[4]

[1]*Beihang University,* [2]*Central University of Finance and Economics*
[3]*Australian National University, and* [4]*Peking University*

Abstract

This document includes two sections. Section S1 is about the robustness check and the corresponding simulation results are presented in Tables S1 – S4. Sections S2 provides some useful lemmas and the theoretical proofs.

Section S1: Robustness Check

It is clear that technical condition (C3) requires that $\rho \to 0$ as the network size $n \to \infty$. We thus perform a robustness check for this condition. More specifically, we are interested in examining the finite sample performance of the proposed estimator (NLSE), when condition (C3) is violated (e.g., a fixed ρ). To that end, we conduct a number of robustness studies. We replicate all of the 4 simulation examples in the main article, the only difference being that we fix $\rho = 0.5$ instead of allowing $\rho = 1/\log(n) \to 0$. The detailed results are given in Tables S1–S4 in this supplement material. We are glad to report that the empirical results are qualitatively similar to those of the previous simulation studies. The performance of the proposed NLSE remains fairly outstanding.

Take Table S1 for illustration purpose. As we can see from this table, for the NLSE, the RMSE (%), SE (%), and $\widehat{\text{SE}}$ (%) values all drop towards 0, as the network size n increases. Besides, the BIAS (%) values are much smaller than the RMSE (%) values and thus, can be ignorable. Moreover, the estimated SE (%) (i.e., $\widehat{\text{SE}}$ (%)) approximates the true SE (%) quite well. As a consequence, the reported coverage probabilities are fairly close to their nominal level 95%. These results corroborate the asymptotic theory given in Theorem 1 quite well and suggest that the proposed NLSE is indeed consistent.

Comparatively speaking, as the network size increases, the coverage probability of our estimator (NLSE) is stable at the nominal level 95%. However, it becomes much harder for the MLE to get reliable $\widehat{\rho}$. Especially, when $n = 5,000$, the reported coverage probability for the MLE equals to 46.20%, which is far from its nominal level 95%. Moreover, the NLSE is computationally much more efficient than the MLE. The CPU time consumed by the MLE is substantially larger than that of the NLSE; see the last two columns in Table S1. For regression coefficients, the NLSE and MLE have better performance than OLSE, and therefore omitted the description for it here. Qualitatively similar patterns can be observed for Tables S2-S4 and therefore omitted here.

Section S2: Proof of Theorem 1

To prove Theorem 1, we first need to establish the following conclusions for $\lambda_{\max}(\cdot)$, which are to be used extensively in the proof. Let A and B be two arbitrary matrices with compatible dimensions, we should have

$$\lambda_{\max}(A+B) \leq \lambda_{\max}(A) + \lambda_{\max}(B), \tag{A.1}$$
$$\lambda_{\max}(A^\top B) \leq \lambda_{\max}(A)\lambda_{\max}(B), \tag{A.2}$$
$$\lambda_{\max}(A^k) \leq \lambda_{\max}^k(A)., \tag{A.3}$$

for any positive integer $k > 0$.

Proof of (A.1): Let u be a vector with unit length and a compatible dimension. Then, by definition, we should have $\lambda_{\max}^2(A+B) = \sup_{\|u\|=1} u^\top (A+B)^\top (A+B)u$, which can be further bounded by

$$\begin{aligned}
&\leq \sup_{\|u\|=1} u^\top A^\top A u + \sup_{\|u\|=1} u^\top B^\top B u + 2 \sup_{\|u\|=1} u^\top A^\top B u \\
&= \lambda_{\max}^2(A) + \lambda_{\max}^2(B) + 2 \sup_{\|u\|=1} u^\top A^\top B u, \\
&\leq \lambda_{\max}^2(A) + \lambda_{\max}^2(B) + 2\{\sup_{\|u\|=1} u^\top A^\top A u\}^{1/2}\{\sup_{\|u\|=1} u^\top B^\top B u\}^{1/2} \\
&= \Big\{\lambda_{\max}(A) + \lambda_{\max}(B)\Big\}^2,
\end{aligned}$$

where the first and the last equality is implied by the definition of $\lambda_{\max}(\cdot)$ and the last inequality is implied by Cauchy's inequality. Thus, (A.1) holds.

Proof of (A.2): Similarly with the proof of (A.1), we have

$$\begin{aligned}\lambda_{\max}^2(A^\top B) &= \sup_{\|u\|=1} u^\top B^\top A A^\top B u \le \lambda_{\max}(A^\top A) \sup_{\|u\|=1} u^\top B^\top B u \\ &\le \lambda_{\max}(A^\top A)\lambda_{\max}(B^\top B) = \lambda_{\max}^2(A)\lambda_{\max}^2(B),\end{aligned}$$

which completes the proof of (A.2).

Proof of (A.3): Note that

$$\begin{aligned}\lambda_{\max}(A^k) &= \Big\{\sup_{\|u\|=1} u^\top (A^k)^\top (A^k) u\Big\}^{1/2} = \Big\{\sup_{\|u\|=1} u^\top (A^{k-1})^\top A^\top A (A^{k-1}) u\Big\}^{1/2} \\ &\le \lambda_{\max}(A)\Big\{\sup_{\|u\|=1} u^\top (A^{k-1})^\top (A^{k-1}) u\Big\}^{1/2} = \lambda_{\max}(A)\lambda_{\max}(A^{k-1}).\end{aligned}$$

If $k > 2$, we can further obtain that $\lambda_{\max}(A^{k-1}) \le \lambda_{\max}(A)\lambda_{\max}(A^{k-2})$. Finally, we have $\lambda_{\max}(A^k) \le \lambda_{\max}^k(A)$, which completes the proof of (A.3).

Proof of Theorem 1: Since $\mathbf{Y}$ can be expressed as $\mathbf{Y} = (I_n - \rho W)^{-1}(\mathbf{X}\beta + \mathcal{E})$. We then have $\widehat{\theta} = (\widehat{\rho}, \widehat{\beta}^\top)^\top$. Specifically, $\widehat{\theta}$ can be expressed as

$$\widehat{\theta} = \begin{pmatrix} \mathbf{Y}^\top W^\top W \mathbf{Y}/n & \mathbf{Y}^\top W^\top \mathbf{X}/n \\ \mathbf{X}^\top W \mathbf{Y}/n & \mathbf{X}^\top \mathbf{X}/n \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{Y}^\top W^\top \mathbf{Y}/n \\ \mathbf{X}^\top \mathbf{Y}/n \end{pmatrix}$$

$$= \begin{pmatrix} \mathbf{Y}^\top W^\top W \mathbf{Y} & \mathbf{Y}^\top W^\top \mathbf{X} \\ \mathbf{X}^\top W \mathbf{Y} & \mathbf{X}^\top \mathbf{X} \end{pmatrix}^{-1} \begin{pmatrix} \rho \mathbf{Y}^\top W^\top W \mathbf{Y} + \mathbf{Y}^\top W^\top \mathbf{X}\beta + \mathbf{Y}^\top W^\top \mathcal{E} \\ \rho \mathbf{X}^\top W \mathbf{Y} + \mathbf{X}^\top \mathbf{X}\beta + \mathbf{X}^\top \mathcal{E} \end{pmatrix} \tag{A.4}$$

Let $\widehat{\Sigma} = (\widehat{\Sigma}_{11}, \widehat{\Sigma}_{12}; \widehat{\Sigma}_{21}, \widehat{\Sigma}_{22})$, where $\widehat{\Sigma}_{11} = \mathbf{Y}^\top W^\top W \mathbf{Y}/n$, $\widehat{\Sigma}_{22} = \mathbf{X}^\top \mathbf{X}/n$, $\widehat{\Sigma}_{12} = \mathbf{Y}^\top W^\top \mathbf{X}/n$, and $\widehat{\Sigma}_{21} = \widehat{\Sigma}_{12}^\top$. Then (A.4) can be further expressed as

$$= \widehat{\Sigma}^{-1}\left[\widehat{\Sigma}\begin{pmatrix} \rho \\ \beta \end{pmatrix} + \begin{pmatrix} \mathbf{Y}^\top W^\top \mathcal{E}/n \\ \mathbf{X}^\top \mathcal{E}/n \end{pmatrix}\right] = \theta + \widehat{\Sigma}^{-1}\begin{pmatrix} \mathbf{Y}^\top W^\top \mathcal{E}/n \\ \mathbf{X}^\top \mathcal{E}/n \end{pmatrix}.$$

Then, $\sqrt{n}(\widehat{\theta} - \theta)$ can be further expressed as $\widehat{\Sigma}^{-1}(\mathbf{Y}^\top W^\top \mathcal{E}/\sqrt{n}, \mathcal{E}^\top \mathbf{X}/\sqrt{n})^\top$. To establish the conclusion of Theorem 1, we consider the following three steps. In the first step, we demonstrate that $\xi = (\mathbf{Y}^\top W^\top \mathcal{E}/\sqrt{n}, \mathcal{E}^\top \mathbf{X}/\sqrt{n})^\top$ is asymptotically normal. In the second step, we are going to show that $\widehat{\Sigma} \to \Sigma$, where Σ is a positive definite matrix. Once Step 2 is established, we are then ready to prove the asymptotic normality of $\widehat{\theta}$, which is given in the third step.

Step 1. We are going to show that $\xi = (\mathbf{Y}^\top W^\top \mathcal{E}/\sqrt{n}, \mathcal{E}^\top \mathbf{X}/\sqrt{n})^\top$ is asymptotically distributed as normal distribution. Firstly, we start with $\mathbf{Y}^\top W^\top \mathcal{E}/\sqrt{n}$, which is the first component of ξ. Note that

$$\frac{1}{\sqrt{n}}\mathbf{Y}^\top W^\top \mathcal{E} = \Big[\frac{1}{\sqrt{n}}(I_n - \rho W)^{-1}(\mathbf{X}\beta + \mathcal{E})\Big]^\top W^\top \mathcal{E}, \tag{A.5}$$

which can be further expressed as

$$\Big\{\frac{1}{\sqrt{n}}\big[(I_n - \rho W)^{-1} - I_n\big](\mathbf{X}\beta + \mathcal{E})\Big\}^\top W^\top \mathcal{E} + \frac{1}{\sqrt{n}}(\mathbf{X}\beta + \mathcal{E})^\top W^\top \mathcal{E}. \tag{A.6}$$

Since $G = (I_n - \rho W)^{-1}$, the first part of equation (A.6) can be further expressed as $[(G - I_n)(\mathbf{X}\beta + \mathcal{E})]^\top W^\top \mathcal{E}/\sqrt{n}$. By Cauchy's inequality, it can be bounded by

$$\begin{aligned}
&\Big[(\mathbf{X}\beta + \mathcal{E})^\top \big\{(G^\top - I_n)(G - I_n)\big\}(\mathbf{X}\beta + \mathcal{E})/n\Big]^{1/2} \|W^\top \mathcal{E}\| \\
&\le \lambda_{\max}^{1/2}\{(G^\top - I_n)(G - I_n)\}\|\mathbf{X}\beta + \mathcal{E}\|\|W^\top \mathcal{E}\|/\sqrt{n} \\
&\le \lambda_{\max}^{1/2}\{(G^\top - I_n)(G - I_n)\}\lambda_{\max}^{1/2}(W^\top W)\|\mathbf{X}\beta + \mathcal{E}\|\|\mathcal{E}\|
\end{aligned}$$

Note that $G - I_n = \sum_{k=1}^\infty \rho^k W^k$. By the definition of $\lambda_{\max}(\cdot)$, we further have

$$\begin{aligned}
\lambda_{\max}\Big\{\sum_{k_1=1}^\infty \rho^{k_1}(W^{k_1})^\top \sum_{k_2=1}^\infty \rho^{k_2} W^{k_2}\Big\} &= \lambda_{\max}\Big\{\sum_{k_1,k_2=1}^\infty \rho^{k_1+k_2}(W^{k_1})^\top W^{k_2}\Big\} \\
&\le \sum_{k_1,k_2=1}^\infty \rho^{k_1+k_2}\lambda_{\max}\Big\{(W^{k_1})^\top W^{k_2}\Big\},
\end{aligned} \tag{A.7}$$

where the last inequality is due to (A.1). Next, by (A.2), we have $\lambda_{\max}\{(W^{k_1})^\top W^{k_2}\} \le \lambda_{\max}(W^{k_1})\lambda_{\max}(W^{k_2})$. Then, by (A.3), the right hand side of (A.7) can be further bounded by $\sum_{k_1,k_2=1}^{\infty}\rho^{k_1+k_2}\lambda_{\max}^{k_1}(W)\lambda_{\max}^{k_2}(W)$. We finally have

$$\begin{aligned}\lambda_{\max}\{(G-I_n)^\top(G-I_n)\} &\leq \sum_{k_1,k_2=1}^{\infty}\rho^{k_1+k_2}\lambda_{\max}^{k_1}(W)\lambda_{\max}^{k_2}(W) \leq \sum_{k_1,k_2=1}^{\infty}\rho^{k_1+k_2}c_{\max}^{(k_1+k_2)/2} \\ &= \sum_{k=2}^{\infty}\rho^k c_{\max}^{k/2} = \frac{\rho^2 c_{\max}}{1-\rho c_{\max}^{1/2}} = o(1),\end{aligned}$$

where the last inequality is implied by condition (C2) and the last equality is implied by condition (C3). On the other hand, implied by condition (C2), we have $|(\mathbf{X}\beta+\mathcal{E})^\top W^\top\mathcal{E}| \leq \lambda_{\max}^{1/2}(W^\top W)\|\mathbf{X}\beta+\mathcal{E}\|\|\mathcal{E}\|$ and $|(\mathbf{X}\beta+\mathcal{E})^\top W^\top\mathcal{E}| \geq \lambda_{\min}^{1/2}(W^\top W)\|\mathbf{X}\beta+\mathcal{E}\|\|\mathcal{E}\|$. Thus, $(\mathbf{X}\beta+\mathcal{E})^\top W^\top\mathcal{E}$ and $\|\mathbf{X}\beta+\mathcal{E}\|\|\mathcal{E}\|$ are of the same order. The above conclusions together with (A.7), we further have (A.6) can be approximated by

$$\frac{1}{\sqrt{n}}(\mathbf{X}\beta+\mathcal{E})^\top W^\top\mathcal{E}\Big\{1+o_p(1)\Big\} = \frac{1}{\sqrt{n}}(\beta^\top\mathbf{X}^\top W^\top\mathcal{E}+\mathcal{E}^\top W^\top\mathcal{E})\Big\{1+o_p(1)\Big\}. \quad \text{(A.8)}$$

Here, $\mathcal{E}^\top W^\top\mathcal{E} = [\mathcal{E}^\top(W^\top+W)\mathcal{E}]/2 = \mathcal{E}^{*\top}\Lambda_B\mathcal{E}^*/2$, where $W^\top+W = P^\top\Lambda_B P$ and $\mathcal{E}^* = P^\top\mathcal{E} \to_d N(0,\sigma^2 I_n)$. Thus, $\mathcal{E}^{*\top}\Lambda_B\mathcal{E}^*/(2\sqrt{n})$ can be further expressed as $\sum_{i=1}^n\lambda_i\varepsilon_i^{*2}/(2\sqrt{n})$, where $\Lambda_B = (\lambda_1,\cdots,\lambda_n)^\top$ with its i-th component represents the i-th largest eigenvalue of $\widehat{W}$ and $\mathcal{E}^* = (\varepsilon_1^*,\cdots,\varepsilon_n^*)^\top \in \mathbb{R}^{n\times 1}$.

Moreover, $\beta^\top\mathbf{X}^\top W^\top\mathcal{E}/\sqrt{n} = \beta^\top\mathbf{X}^\top W^\top P\mathcal{E}^*/\sqrt{n}$. Let $\alpha = P^\top W\mathbf{X}\beta$ and $\alpha = (\alpha_1,\cdots,\alpha_n)^\top \in \mathbb{R}^{n\times 1}$. Then, $\beta^\top\mathbf{X}^\top W^\top\mathcal{E}/\sqrt{n} = \sum_{i=1}^n\alpha_i\varepsilon_i^*/\sqrt{n}$. Thus, (A.8) can be further expressed as $\sum_{i=1}^n(\lambda_i\varepsilon_i^{*2}/2+\alpha_i\varepsilon_i^*)/\sqrt{n}$. On the other hand, $\mathcal{E}^\top\mathbf{X}/\sqrt{n} = \sum_{i=1}^n\gamma_i^\top\varepsilon_i^*/\sqrt{n}$, where γ_i is a $p\times 1$ vector, which represents the i-th column of $\mathbf{X}^\top P$. Thus, $\xi = (\beta^\top\mathbf{X}^\top W^\top\mathcal{E}+\mathcal{E}^\top W^\top\mathcal{E},\mathcal{E}^\top\mathbf{X})^\top/\sqrt{n}\{1+o_p(1)\} =: \tilde{\xi}\{1+o_p(1)\}$, say. Here, $\tilde{\xi} = \sum_{i=1}^n(\lambda_i/2,\alpha_i;\kappa_0,\gamma_i)\widetilde{\varepsilon}_i^*/\sqrt{n}$, $\widetilde{\varepsilon}_i^* = (\varepsilon_i^{*2},\varepsilon_i^*)^\top$ and $\kappa_0 \in \mathbb{R}^{p\times 1}$ with each element equals to 0. By the Lindeberg-Feller central limit theorem, we have $\tilde{\xi}|\mathbf{X}$ is asymptotic normal. Firstly, by simple calculation, we can obtain that $E(\tilde{\xi}) = 0$. Secondly, since

$$\tilde{\xi} = \frac{1}{2\sqrt{n}}\begin{pmatrix}\text{vec}^\top(W+W^\top)\text{vec}(\mathcal{E}\mathcal{E}^\top-\sigma^2 I_n)\\ 0\end{pmatrix} + \begin{pmatrix}\beta^\top\mathbf{X}^\top W^\top\\ \mathbf{X}^\top\end{pmatrix}\mathcal{E},$$

by Theorem 1 of Kelejian and Prucha (2001), the asymptotic covariance matrix of $\tilde{\xi}$

is the limitation of

$$\frac{1}{n}\left[\frac{1}{2}\sigma^4\begin{pmatrix} tr\left\{(W+W^\top)^2\right\} & 0 \\ 0 & 0\end{pmatrix} + \sigma^2\begin{pmatrix}\beta^\top\mathbf{X}^\top W^\top \\ \mathbf{X}^\top\end{pmatrix}\begin{pmatrix}\beta^\top\mathbf{X}^\top W^\top \\ \mathbf{X}^\top\end{pmatrix}^\top\right] =: \widehat{\Sigma}^*, \text{ say,}$$

where $\widehat{\Sigma}^* = (\widehat{\Sigma}^*_{11}, \widehat{\Sigma}^*_{12}; \widehat{\Sigma}^*_{21}, \widehat{\Sigma}^*_{22})$ with

$$\widehat{\Sigma}^*_{11} = \frac{1}{2n}\sigma^4 tr\left\{(W+W^\top)^2\right\} + \frac{1}{n}\sigma^2\beta^\top\mathbf{X}^\top W^\top W\mathbf{X}\beta,$$

$$\widehat{\Sigma}^*_{12} = \frac{1}{n}\sigma^2\beta^\top\mathbb{X}^\top W^\top\mathbb{X}, \widehat{\Sigma}^*_{21} = \widehat{\Sigma}^{*\top}_{12}, \text{ and } \widehat{\Sigma}^*_{22} = \frac{1}{n}\sigma^2\mathbb{X}^\top\mathbb{X}.$$

Moreover, $n^{-1}(W\mathbf{X}\beta)^\top(W\mathbf{X}\beta) = n^{-1}(\mathbf{X}\beta)^\top\widetilde{W}(\mathbf{X}\beta) = n^{-1}\sum_{i,j=1}^n \widetilde{w}_{ij}(X_i^\top\beta)(X_j^\top\beta)$. By condition (C1), we know that X_i and X_j are independent with $i \neq j$ and $E(X_i) = E(X_j) = 0$. We then have

$$n^{-1}\sum_{i,j=1}^n \widetilde{w}_{ij}E(X_i{}^\top\beta)(X_j{}^\top\beta) = n^{-1}\sum_{i=1}^n \widetilde{w}_{ii}E(X_i{}^\top\beta)^2 = n^{-1}tr(\widetilde{W})\beta^\top\Sigma_X\beta. \quad \text{(A.9)}$$

By condition (C2), we know that (A.9) tends to $C_1\beta^\top\Sigma_X\beta$, as $n \to \infty$. Next, we want to compute var$\{n^{-1}(W\mathbf{X}\beta)^\top(W\mathbf{X}\beta)\}$. To this end, we start with

$$E\{n^{-1}(W\mathbf{X}\beta)^\top(W\mathbf{X}\beta)\}^2 = n^{-2}E\Big\{\sum_{i_1,j_1,i_2,j_2=1}^n \widetilde{w}_{i_1j_1}\widetilde{w}_{i_2j_2}(X_{i_1}^\top\beta)(X_{j_1}^\top\beta)(X_{i_2}^\top\beta)(X_{j_2}^\top\beta)\Big\}.$$

Then, according to the index relationship between (i_1, j_1) and (i_2, j_2), we can decompose the above quantity into the following three parts: A_1, A_2, and A_3 respectively.

Specifically,

$$
\begin{aligned}
A_1 &= n^{-2} E \sum_{i_1 \neq i_2} \widetilde{w}_{i_1 i_1} \widetilde{w}_{i_2 i_2} (X_{i_1}^\top \beta)^2 (X_{i_2}^\top \beta)^2 \\
A_2 &= 2n^{-2} E \sum_{i_1 \neq j_1} \widetilde{w}_{i_1 j_1}^2 (X_{i_1}^\top \beta)^2 (X_{j_1}^\top \beta)^2 \\
A_3 &= n^{-2} E \sum_{i=1}^{n} \widetilde{w}_{ii}^2 (X_i^\top \beta)^4.
\end{aligned}
$$

We then evaluate the the above three parts separately. Firstly,

$$
\begin{aligned}
A_1 &= n^{-2} \sum_{i_1 \neq i_2} \widetilde{w}_{i_1 i_1} \widetilde{w}_{i_2 i_2} \Big\{ E(X_{i_1}^\top \beta)^2 \Big\}^2 = 4n^{-2} \sum_{i_1 \neq i_2} \widetilde{w}_{i_1 i_1} \widetilde{w}_{i_2 i_2} (\beta^\top \Sigma_X \beta)^2 \\
&= n^{-2} (\beta^\top \Sigma_X \beta)^2 \Big\{ tr^2(\widetilde{W}) - \sum_{i=1}^{n} \widetilde{w}_{ii}^2 \Big\}. \\
&= (\beta^\top \Sigma_X \beta)^2 \Big\{ \{n^{-1} tr(\widetilde{W})\}^2 - n^{-2} \sum_{i=1}^{n} \widetilde{w}_{ii}^2 \Big\}.
\end{aligned}
$$

Because, $n^{-2} \sum_{i=1}^n \widetilde{w}_{ii}^2 \le n^{-2} tr(\widetilde{W^2}) \le n^{-2} \lambda_{\max}(\widetilde{W}) tr(\widetilde{W}) \le n^{-2} c_{\max} tr(\widetilde{W}) = o_p(1)$. Thus, we have $A_1 = (\beta^\top \Sigma_X \beta)^2 \{n^{-1} tr(\widetilde{W})\}^2 + o_p(1)$. In addition,

$$
\begin{aligned}
A_2 &= 2n^{-2} (\beta^\top \Sigma_X \beta)^2 \sum_{i_1 \neq j_1} \widetilde{w}_{i_1 j_1}^2 = 2n^{-2} (\beta^\top \Sigma_X \beta)^2 \Big\{ tr(\widetilde{W^2}) - \sum_{i=1}^{n} \widetilde{w}_{ii}^2 \Big\}. \\
&= 2n^{-2} (\beta^\top \Sigma_X \beta)^2 tr(\widetilde{W^2}) + o_p(1).
\end{aligned}
$$

We next examine A_3. Note that $A_3 = n^{-2} \sum_{i=1}^n \widetilde{w}_{ii}^2 E(X_i^\top \beta)^4$. By condition (C1), we know that $E(X_i^\top \beta)^4$ exists. Combine with the above results, we immediately have $A_3 = o(1)$. Thus, A_3 is ignorable. Combining $A_1 + A_2 + A_3$ together, we have

$$
E\{n^{-1} (W\mathbf{X}\beta)^\top (W\mathbf{X}\beta)\}^2 = (\beta^\top \Sigma_X \beta)^2 \Big\{ n^{-2} tr^2(\widetilde{W}) + 2n^{-2} tr(\widetilde{W^2}) \Big\} + o(1).
$$

Implied by (A.9), we have $E^2\{n^{-1}(W\mathbf{X}\beta)^\top(W\mathbf{X}\beta)\} = n^{-2}(\beta^\top\Sigma_X\beta)^2 tr^2(\widetilde{W})$. Thus, $\mathrm{var}\{n^{-1}(W\mathbf{X}\beta)^\top(W\mathbf{X}\beta)\} = 2(\beta^\top\Sigma_X\beta)^2 n^{-2} tr(\widetilde{W^2}) + o(1) = o(1)$, where the last equality is implied by condition (C2). Therefore, $n^{-1}\sigma^2(W\mathbf{X}\beta)^\top(W\mathbf{X}\beta) \to_p C_1\sigma^2\beta^\top\Sigma_X\beta$. Then, implied by condition (C2), we have $\widehat{\Sigma}^*_{11} \to_p C_1\sigma^2\beta^\top\Sigma_X\beta + C_2\sigma^4/2$.

On the other hand, $\widehat{\Sigma}^*_{22} = \sigma^2 n^{-1}\mathbf{X}^\top\mathbf{X} \to_p \sigma^2\Sigma_X$. Note that $\widehat{\Sigma}^*_{12} = n^{-1}\sigma^2\beta^\top\sum_{i,j=1}^n X_i \times w_{ji}X_j^\top$. By similar technique, we can obtain that $E(\widehat{\Sigma}^*_{12}) = 0$ and $\mathrm{var}(\widehat{\Sigma}^*_{12}) = o(1)$. Thus, $\widehat{\Sigma}^* \to_p \Sigma^*$, where $\Sigma^* = (C_1\sigma^2\beta^\top\Sigma_X\beta + C_2\sigma^4/2, 0; 0, \sigma^2\Sigma_X)$. Therefore, $\xi \to_d N(0, \Sigma^*)$, which completes the proof of Step 1.

Step 2. Note that $\widehat{\Sigma}_{11} = \mathbf{Y}^\top W^\top W\mathbf{Y}/n$, which can be expressed into $(S_1 + S_2 + S_3)/n$. Here,

$$\begin{aligned}
S_1 &= (\mathbf{X}\beta + \mathcal{E})^\top(G-I)^\top W^\top W(G-I)(\mathbf{X}\beta + \mathcal{E}),\\
S_2 &= (\mathbf{X}\beta + \mathcal{E})^\top W^\top W(G-I)(\mathbf{X}\beta + \mathcal{E}) + (\mathbf{X}\beta + \mathcal{E})^\top(G-I)^\top W^\top W(\mathbf{X}\beta + \mathcal{E}),\\
S_3 &= (\mathbf{X}\beta + \mathcal{E})^\top W^\top W(\mathbf{X}\beta + \mathcal{E}).
\end{aligned}$$

We can verify that $S_1 = o(S_3)$ and $S_2 = o(S_3)$. We start with S_2 first and S_1 is a smaller order of S_2. By Cauchy's inequality, we have

$$\begin{aligned}
S_2 &\le 2\sqrt{(\mathbf{X}\beta + \mathcal{E})^\top W^\top W(\mathbf{X}\beta + \mathcal{E})}\sqrt{(\mathbf{X}\beta + \mathcal{E})^\top(G-I)^\top W^\top W(G-I)(\mathbf{X}\beta + \mathcal{E})}\\
&\le 2\lambda_{\max}(W^\top W)\lambda_{\max}^{1/2}\{(G-I)^\top(G-I)\}\|\mathbf{X}\beta + \mathcal{E}\|^2
\end{aligned}$$

Since $\lambda_{\max}\{(G-I)^\top(G-I)\} = o(1)$, we further have $S_2 = o_p(\|\mathbf{X}\beta+\mathcal{E}\|^2)$. Similarly, we can obtain that $S_1 = o_p(\|\mathbf{X}\beta+\mathcal{E}\|^2)$. On the hand, we have $S_3 \le \lambda_{\max}(W^\top W)\|\mathbf{X}\beta + \mathcal{E}\|^2$ and $S_3 \ge \lambda_{\min}(W^\top W)\|\mathbf{X}\beta + \mathcal{E}\|^2$. Then, by condition (C2), we can find that S_3 has the same order with $\|\mathbf{X}\beta + \mathcal{E}\|^2$. Therefore, $S_1 = o_p(S_3)$ and $S_2 = o_p(S_3)$, which implies that $\widehat{\Sigma}_{11} = S_3\{1 + o_p(1)\}/n$. Similar technique with Step 1, we have $\widehat{\Sigma}_{11} \to_p C_1(\beta^\top\Sigma_X\beta + \sigma^2)$, $\widehat{\Sigma}_{22} \to_p \Sigma_X$, $\widehat{\Sigma}_{21} \to_p 0$, and $\widehat{\Sigma}_{12} \to_p 0$. Let $\Sigma = (\Sigma_{11}, \Sigma_{12}; \Sigma_{21}, \Sigma_{22})$. Specifically, $\Sigma_{11} = C_1(\beta^\top\Sigma_X\beta + \sigma^2)$, $\Sigma_{12} = 0$, $\Sigma_{21} = 0$, and $\Sigma_{22} = \Sigma_X$. We then have $\widehat{\Sigma} \to_p \Sigma$, which completes the proof of Step 2.

Step 3. Since $\sqrt{n}(\widehat{\theta} - \theta_0) = \widehat{\Sigma}^{-1}\xi$, by the results of Steps 1 and 2, we know that $\sqrt{n}(\widehat{\theta} - \theta) \to_d N(0, \Sigma_1^{-1})$. Specifically, $\Sigma_1^{-1} = \Sigma^{-1}\Sigma^*\Sigma^{-1}$ and

$$\Sigma_1 = \Sigma\Sigma^{*-1}\Sigma = \begin{pmatrix} \Sigma_{11}{\Sigma^*_{11}}^{-1}\Sigma_{11} & 0 \\ 0 & \Sigma_{22}{\Sigma^*_{22}}^{-1}\Sigma_{22} \end{pmatrix} =: (\sigma_{11}, 0; 0, \sigma_{22}), \text{ say,}$$

where $\sigma_{11} = C_1^2(\beta^\top \Sigma_X \beta + \sigma^2)^2 / \{\sigma^2(C_1 \beta^\top \Sigma_X \beta + C_2 \sigma^2/2)\}$ and $\sigma_{22} = \sigma^{-2}\Sigma_X$. This completes the proof of Step 3 and thus the whole proof is completed.

REFERENCES

LeSage, J. and Pace, R. K. (2009), *Introduction to Spatial Econometrics*, New York: Chapman & Hall.

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

Table S1: Detailed simulation results for ER networks with $\rho = 0.5$ and $p = 7$. In this table, BIAS (%) = bias, RMSE (%) = root mean square error, SE(%) = estimated true standard error, $\widehat{\text{SE}}$ (%) = average standard error estimate, and CP (%) = coverage probability. Computational time (TIME) in seconds is also reported.

		BIAS(%)			RMSE(%)			SE (%)			$\widehat{\text{SE}}$(%)			CP (%)			TIME	
n	Parameter	NLSE	OLSE	MLE	NLSE	OLSE	MLE	NLSE	OLSE	MLE	NLSE	OLSE	MLE	NLSE	OLSE	MLE	NLSE	MLE
500	ρ	0.20	-	0.03	4.28	-	4.25	4.28	-	4.25	4.41	-	4.14	95.56	-	94.28	0.009	0.209
	β_1	0.12	0.39	0.12	5.17	5.86	5.17	5.17	5.85	5.17	5.19	5.93	5.16	95.04	95.06	94.94		
	β_2	-0.06	0.09	-0.06	5.92	6.76	5.92	5.92	6.76	5.92	5.81	6.63	5.78	94.98	94.64	94.92		
	β_3	0.11	0.13	0.11	5.81	6.59	5.81	5.81	6.59	5.81	5.81	6.63	5.77	95.06	95.00	94.90		
	β_4	-0.08	-0.02	-0.08	5.81	6.63	5.81	5.81	6.63	5.81	5.81	6.63	5.78	94.32	94.64	94.22		
	β_5	0.07	0.24	0.07	5.85	6.66	5.85	5.85	6.66	5.85	5.81	6.63	5.78	94.54	95.06	94.28		
	β_6	0.03	-0.01	0.03	5.76	6.58	5.76	5.76	6.58	5.76	5.81	6.63	5.78	95.00	95.38	94.90		
	β_7	-0.08	-0.07	-0.08	5.27	5.97	5.27	5.27	5.97	5.27	5.19	5.93	5.16	94.88	94.90	94.84		
2,000	ρ	0.10	-	0.02	3.06	-	3.05	3.06	-	3.05	3.14	-	6.44	95.76	-	97.82	0.225	2.855
	β_1	0.01	0.09	0.01	2.59	2.76	2.59	2.59	2.76	2.59	2.59	2.76	2.58	94.78	95.10	94.74		
	β_2	0.01	0.05	0.01	2.90	3.08	2.90	2.90	3.08	2.90	2.89	3.08	2.89	94.90	94.92	94.90		
	β_3	0.03	0.03	0.03	2.90	3.12	2.90	2.90	3.12	2.90	2.89	3.08	2.89	94.88	94.80	94.86		
	β_4	-0.02	-0.01	-0.02	2.84	3.04	2.84	2.84	3.04	2.84	2.89	3.08	2.89	95.80	95.34	95.80		
	β_5	-0.06	-0.02	-0.06	2.88	3.09	2.88	2.88	3.09	2.88	2.89	3.08	2.89	94.98	95.06	95.00		
	β_6	0.02	0.01	0.02	2.85	3.05	2.85	2.85	3.05	2.85	2.89	3.08	2.89	95.32	95.28	95.28		
	β_7	0.05	0.06	0.05	2.59	2.77	2.59	2.59	2.77	2.59	2.59	2.76	2.58	94.70	94.80	94.70		
5,000	ρ	0.01	-	-2.03	2.42	-	3.11	2.42	-	2.35	2.50	-	1.18	95.68	-	46.20	2.083	27.791
	β_1	-0.01	0.02	0.00	1.62	1.68	1.62	1.62	1.68	1.62	1.63	1.70	1.63	95.28	95.42	95.34		
	β_2	-0.02	0.00	-0.02	1.79	1.87	1.79	1.79	1.87	1.79	1.83	1.90	1.82	95.40	95.16	95.42		
	β_3	0.00	0.00	0.00	1.81	1.89	1.81	1.81	1.89	1.81	1.83	1.90	1.83	95.30	95.14	95.40		
	β_4	-0.01	0.00	-0.01	1.83	1.91	1.83	1.83	1.91	1.83	1.83	1.90	1.83	95.24	95.22	95.24		
	β_5	0.01	0.02	0.01	1.83	1.90	1.83	1.83	1.90	1.83	1.83	1.90	1.83	94.52	94.70	94.50		
	β_6	0.00	0.00	0.00	1.79	1.86	1.79	1.79	1.86	1.79	1.83	1.90	1.83	95.34	95.60	95.38		
	β_7	0.01	0.01	0.01	1.64	1.70	1.64	1.64	1.70	1.64	1.63	1.70	1.63	94.96	95.42	94.96		

Table S2: Detailed simulation results for Dyad Independence networks with $\rho = 0.5$ and $p = 3$. In this table, BIAS (%) = bias, RMSE (%) = root mean square error, SE(%) = estimated true standard error, $\widehat{\mathrm{SE}}$ (%) = average standard error estimate, and CP (%) = coverage probability. Computational time (TIME) in seconds is also reported.

		BIAS(%)			RMSE(%)			SE (%)			$\widehat{\mathrm{SE}}$(%)			CP (%)			TIME	
n	Parameter	NLSE	OLSE	MLE	NLSE	OLSE	MLE	NLSE	OLSE	MLE	NLSE	OLSE	MLE	NLSE	OLSE	MLE	NLSE	MLE
1,000	ρ	0.22	-	0.04	1.18	-	1.16	1.16	-	1.16	1.23	-	1.40	95.98	-	95.88	0.051	0.779
	β_1	-0.08	3.58	-0.06	3.96	7.66	3.96	3.96	6.77	3.96	3.88	6.54	3.87	94.50	90.78	94.50		
	β_2	0.04	3.77	0.06	3.92	7.65	3.92	3.92	6.65	3.92	3.88	6.54	3.87	94.82	91.16	94.78		
	β_3	0.03	3.76	0.04	3.84	7.69	3.84	3.84	6.71	3.84	3.88	6.54	3.87	95.14	90.72	95.08		
1,500	ρ	0.19	-	0.01	1.07	-	1.05	1.05	-	1.05	1.09	-	1.22	95.42	-	95.78	0.107	1.510
	β_1	-0.07	2.91	-0.06	3.17	5.90	3.17	3.17	5.13	3.17	3.16	5.03	3.16	94.96	90.34	94.94		
	β_2	0.05	3.33	0.06	3.16	6.08	3.16	3.16	5.09	3.16	3.16	5.03	3.16	95.04	89.34	95.00		
	β_3	-0.01	3.05	0.00	3.18	5.96	3.18	3.18	5.12	3.18	3.17	5.03	3.16	95.16	90.62	95.12		
2,000	ρ	0.18	-	-0.02	0.99	-	0.96	0.97	-	0.96	1.00	-	1.11	95.12	-	96.20	0.231	2.909
	β_1	0.06	2.80	0.08	2.74	5.05	2.74	2.74	4.20	2.74	2.74	4.20	2.74	94.84	90.00	94.80		
	β_2	-0.02	2.67	-0.01	2.73	5.09	2.73	2.73	4.33	2.73	2.74	4.20	2.74	95.08	89.90	95.08		
	β_3	-0.03	2.70	-0.02	2.67	4.99	2.67	2.67	4.20	2.67	2.74	4.20	2.74	95.22	90.50	95.22		

Table S3: Detailed simulation results for Stochastic Block networks with $\rho = 0.5$ and $p = 7$. In this table, BIAS (%) = bias, RMSE (%) = root mean square error, SE(%) = estimated true standard error, $\widehat{\text{SE}}$ (%) = average standard error estimate, and CP (%) = coverage probability. Computational time (TIME) in seconds is also reported.

		BIAS(%)			RMSE(%)			SE (%)			$\widehat{\text{SE}}$(%)			CP (%)			TIME	
n	Parameter	NLSE	OLSE	MLE	NLSE	OLSE	MLE	NLSE	OLSE	MLE	NLSE	OLSE	MLE	NLSE	OLSE	MLE	NLSE	MLE
1,000	ρ	0.19	-	-0.16	5.82	-	5.75	5.82	-	5.75	6.15	-	4.54	96.36	-	88.94	0.040	0.711
	β_1	0.04	0.07	0.04	3.71	3.86	3.71	3.71	3.86	3.71	3.66	3.80	3.65	94.98	94.56	94.94		
	β_2	-0.14	-0.10	-0.14	4.07	4.20	4.07	4.07	4.20	4.07	4.09	4.25	4.08	94.80	94.96	94.72		
	β_3	0.06	0.10	0.06	4.12	4.27	4.12	4.12	4.27	4.12	4.09	4.25	4.08	95.14	94.86	94.96		
	β_4	-0.07	0.00	-0.07	4.04	4.20	4.04	4.04	4.20	4.04	4.09	4.25	4.08	95.60	95.50	95.54		
	β_5	0.00	0.00	0.00	4.05	4.22	4.05	4.05	4.22	4.05	4.09	4.25	4.08	95.26	95.28	95.18		
	β_6	-0.05	0.02	-0.05	4.16	4.26	4.16	4.16	4.26	4.16	4.09	4.25	4.08	94.82	94.82	94.64		
	β_7	0.06	0.11	0.06	3.67	3.82	3.67	3.67	3.82	3.67	3.66	3.80	3.65	95.16	95.28	95.06		
2,000	ρ	0.14	-	-0.11	5.14	-	5.09	5.14	-	5.09	5.36	-	4.51	95.64	-	92.36	0.253	3.496
	β_1	0.01	0.04	0.01	2.59	2.65	2.59	2.59	2.65	2.59	2.59	2.65	2.58	94.74	95.04	94.72		
	β_2	0.01	0.04	0.01	2.91	2.97	2.91	2.91	2.97	2.91	2.89	2.96	2.89	94.90	94.70	94.88		
	β_3	0.03	0.06	0.03	2.90	2.99	2.90	2.90	2.99	2.90	2.89	2.96	2.89	94.94	95.00	94.86		
	β_4	-0.02	-0.01	-0.02	2.84	2.92	2.84	2.84	2.92	2.84	2.89	2.96	2.89	95.88	95.54	95.84		
	β_5	-0.06	-0.03	-0.06	2.88	2.95	2.88	2.88	2.95	2.88	2.89	2.96	2.89	94.96	95.08	94.94		
	β_6	0.02	0.06	0.02	2.85	2.93	2.85	2.85	2.93	2.85	2.89	2.96	2.89	95.20	95.40	95.18		
	β_7	0.05	0.07	0.05	2.59	2.66	2.59	2.59	2.65	2.59	2.59	2.65	2.58	94.68	94.68	94.68		
5,000	ρ	0.12	-	-3.15	4.41	-	5.22	4.41	-	4.16	4.46	-	2.46	95.28	-	59.46	2.484	28.398
	β_1	-0.01	0.00	-0.01	1.62	1.64	1.62	1.62	1.64	1.62	1.63	1.66	1.63	95.28	95.70	95.34		
	β_2	-0.02	-0.01	-0.02	1.79	1.82	1.79	1.79	1.82	1.79	1.83	1.85	1.82	95.36	95.42	95.38		
	β_3	0.00	0.00	0.00	1.81	1.83	1.81	1.81	1.83	1.81	1.83	1.85	1.83	95.34	95.22	95.34		
	β_4	0.00	0.01	0.00	1.83	1.85	1.83	1.83	1.85	1.83	1.83	1.85	1.83	95.22	95.18	95.18		
	β_5	0.01	0.02	0.01	1.83	1.86	1.83	1.83	1.86	1.83	1.83	1.85	1.83	94.54	94.44	94.46		
	β_6	0.00	0.00	0.00	1.79	1.82	1.79	1.79	1.82	1.79	1.83	1.85	1.83	95.38	95.32	95.34		
	β_7	0.01	0.02	0.01	1.64	1.67	1.64	1.64	1.67	1.64	1.63	1.66	1.63	94.96	95.24	95.00		

Table S4: Detailed simulation results for Power-Law type networks with $\rho = 0.5$ and $p = 4$. In this table, BIAS (%) = bias, RMSE (%) = root mean square error, SE(%) = estimated true standard error, $\widehat{\text{SE}}$ (%) = average standard error estimate, and CP (%) = coverage probability. Computational time (TIME) in seconds is also reported.

		BIAS(%)			RMSE(%)			SE (%)			$\widehat{\text{SE}}$(%)			CP (%)			TIME	
n	Parameter	NLSE	OLSE	MLE	NLSE	OLSE	MLE	NLSE	OLSE	MLE	NLSE	OLSE	MLE	NLSE	OLSE	MLE	NLSE	MLE
100	ρ	0.02	-	-11.73	1.89	-	29.42	1.89	-	26.98	2.07	-	2.63	96.86	-	79.04	0.001	0.093
	β_1	0.36	2.51	1.08	14.59	42.36	28.03	14.58	42.29	28.01	14.36	42.32	21.43	94.42	95.38	94.50		
	β_2	-0.10	2.48	0.28	14.60	42.62	27.54	14.60	42.55	27.54	14.35	42.29	21.42	94.14	94.54	94.18		
	β_3	-0.64	2.01	0.27	14.58	42.39	27.84	14.57	42.35	27.84	14.39	42.40	21.47	94.62	95.02	94.54		
	β_4	0.41	-4.81	-0.89	21.07	60.35	39.33	21.07	60.16	39.32	20.32	59.76	30.32	94.30	94.42	94.04		
500	ρ	0.01	-	0.01	0.84	-	0.84	0.84	-	0.84	0.90	-	0.82	96.34	-	94.52	0.009	0.210
	β_1	0.04	0.69	0.04	6.39	18.34	6.39	6.39	18.33	6.39	6.33	18.46	6.32	94.92	95.24	94.82		
	β_2	0.05	1.03	0.05	6.28	18.58	6.28	6.28	18.56	6.28	6.34	18.46	6.32	95.18	94.88	95.16		
	β_3	-0.01	0.84	-0.01	6.29	18.64	6.29	6.30	18.62	6.30	6.34	18.48	6.32	94.74	94.84	94.70		
	β_4	-0.14	-1.89	-0.14	8.96	26.17	8.96	8.96	26.10	8.96	8.96	26.11	8.94	94.64	95.10	94.60		
1,000	ρ	0.01	-	0.01	0.59	-	0.59	0.59	-	0.59	0.64	-	0.58	96.40	-	94.78	0.044	0.645
	β_1	0.08	0.25	0.08	4.50	12.96	4.50	4.50	12.95	4.50	4.47	13.02	4.47	94.96	94.82	94.94		
	β_2	0.00	0.41	0.00	4.54	13.19	4.54	4.54	13.19	4.54	4.48	13.02	4.47	94.70	94.78	94.64		
	β_3	0.02	0.21	0.02	4.48	13.06	4.48	4.48	13.06	4.48	4.48	13.02	4.47	94.96	94.98	94.88		
	β_4	-0.03	-0.41	-0.03	6.41	18.35	6.41	6.41	18.35	6.41	6.33	18.41	6.32	94.88	95.12	94.84		