Supplemental Material for "Autoregression Model with Spatial Dependence and Missing Data"

Jing Zhou[1], Jin Liu[2], Feifei Wang[1], and Hansheng Wang[2]

[1]*Renmin University of China, and* [2]*Peking University*

APPENDIX

Appendix A: Technical Lemmas

To establish the theoretical results in Section 3, the following technical lemmas are considered.

Lemma 1. *The OLS estimator of α_i is not consistent.*

PROOF: Define

$$A_{1i} = \frac{1}{T-1}\sum_{t=2}^{T} Z_{it} Z_{i(t-1)} Y_{i(t-1)}^2, \quad B_{1i} = \frac{1}{T-1}\sum_{t=2}^{T} Z_{it} Z_{i(t-1)} Y_{i(t-1)} \varepsilon_{it}.$$

Then, we have $\hat{\alpha}_i^{LSE} - \alpha_i = A_{1i}^{-1} B_{1i}$ and compute the following equation,

$$\begin{aligned}
E(B_{1i}) &= \frac{1}{T-1}\sum_{t=2}^{T} E\{Z_{it} Z_{i(t-1)} Y_{i(t-1)} \varepsilon_{it}\} \\
&= \frac{1}{T-1}\sum_{t=2}^{T} E\Big[Y_{i(t-1)} \varepsilon_{it} \Big\{ E\big(Z_{it}\big|\mathcal{F}\big) E\big(Z_{i(t-1)}\big|\mathcal{F}\big)\Big\}\Big] \\
&= \frac{1}{T-1}\sum_{t=2}^{T} E\Big\{Y_{i(t-1)} \varepsilon_{it} p_{it} p_{i(t-1)}\Big\} = E\Big\{Y_{i(t-1)} \varepsilon_{it} p_{it} p_{i(t-1)}\Big\} \neq 0 \\
E(A_{1i}) &= \frac{1}{T-1}\sum_{t=2}^{T} E\{Z_{it} Z_{i(t-1)} Y_{i(t-1)}^2\} = E\Big\{Y_{i(t-1)}^2 p_{it} p_{i(t-1)}\Big\} \to c > 0.
\end{aligned}$$

By the law of large numbers, we have $A_{1i} \to_p E(A_{1i}) \to c$ and $B_{1i} \to_p E(B_{1i}) \nrightarrow 0$. With Slutsky's Theorem, we can show that $\hat{\alpha}_i^{LSE}$ is inconsistent.

Lemma 2. *Assume M_1 and M_2 are $N \times N$ square matrices. Let $\mathbb{U} = (u_1, \cdots, u_N)^\top$ follows a N-dimensional standard normal distribution. Then, we have*

(i) $E(\mathbb{U}^\top M_1 \mathbb{U}) = tr(M_1)$;

(ii) $E(\mathbb{U}^\top M_1 \mathbb{U} \cdot \mathbb{U}^\top M_2 \mathbb{U}) = tr(M_1)tr(M_2) + 2tr(M_1 M_2)$;

(iii) $cov(\mathbb{U}^\top M_1 \mathbb{U}, \mathbb{U}^\top M_2 \mathbb{U}) = 2tr(M_1 M_2)$;

(iv) $var(\mathbb{U}^\top M_1 \mathbb{U}) = 2tr(M_1^2)$.

Proof: This Lemma follows Lemma A.11 in the supplement of Lee (2004).

Lemma 3. *Define $\mathcal{S} = (N(T-1))^{-1}(2\sigma^2)^{-1}\sum_{t=2}^T \mathcal{E}_t^\top \dot{A}_t(\rho)\mathcal{E}_t$, and $\mathcal{S}_1 = (N(T-1))^{-1}(2\sigma^2)^{-1}\sum_{t=2}^T \mathcal{E}_t^\top A_t(\rho)\mathcal{E}_t$. Assume the conditions in Theorem 3 hold, we have*

$$\sqrt{N(T-1)}\big\{\mathcal{S} - E(\mathcal{S})\big\} \to_d N\big(0, \Delta_{11}\big),$$
$$\sqrt{N(T-1)}\big\{\mathcal{S}_1 - E(\mathcal{S}_1)\big\} \to_d N\big(0, \sigma^4\Delta_{22}\big).$$

Proof: Denote $\mathbb{U}_t = \Sigma^{-1/2}\mathcal{E}_t = \sigma^{-1}\Omega(\rho)^{1/2}\mathcal{E}_t = (U_{1t}, \cdots, U_{Nt})^\top$, where $\mathbb{U}_t$ follows a N-dimensional standard normal distribution. Denote $B_t(\rho) = \Omega(\rho)^{-1/2}A_t(\rho)\Omega(\rho)^{-1/2}$, and $\dot{B}_t(\rho) = \Omega(\rho)^{-1/2}\dot{A}_t(\rho)\Omega(\rho)^{-1/2}$, i.e.,

$$\dot{B}_t(\rho) = \Omega(\rho)^{-1/2}\big[\mathcal{Z}_t\mathcal{P}_t^{-1}\{\dot{\Omega}(\rho) - \mathrm{diag}(\dot{\Omega}(\rho))\}\mathcal{P}_t^{-1}\mathcal{Z}_t + \mathrm{diag}(\dot{\Omega}(\rho))\mathcal{P}_t^{-1}\mathcal{Z}_t\big]\Omega(\rho)^{-1/2}.$$

Then, we have

$$\mathcal{S} = \frac{1}{2N(T-1)}\sum_{t=2}^T \mathbb{U}_t^\top \Omega(\rho)^{-1/2}\dot{A}_t(\rho)\Omega(\rho)^{-1/2}\mathbb{U}_t = \frac{1}{2N(T-1)}\sum_{t=2}^T \mathbb{U}_t^\top \dot{B}_t(\rho)\mathbb{U}_t.$$

Then, by Lemma 2, we have the expectation of $\mathcal{S}$ as,

$$E(\mathcal{S}) = \frac{1}{2N(T-1)}\sum_{t=2}^{T} E\Big[E\Big\{\mathbb{U}_t^\top \dot{B}_t(\rho)\mathbb{U}_t|\mathcal{F}\Big\}\Big] = \frac{1}{2N(T-1)}\sum_{t=2}^{T} E\Big[\mathrm{tr}\Big\{\dot{B}_t(\rho)\Big\}\Big]$$
$$= \frac{1}{2N(T-1)}\sum_{t=2}^{T} E\Big[\mathrm{tr}\Big\{\dot{A}_t(\rho)\Omega(\rho)^{-1}\Big\}\Big] = \frac{1}{2N}\mathrm{tr}\Big\{\dot{\Omega}(\rho)\Omega(\rho)^{-1}\Big\}.$$

The associated second moment is

$$E(\mathcal{S}^2) = E\Big\{E(\mathcal{S}^2|\mathcal{F})\Big\} = \frac{1}{4N^2(T-1)^2}E\Bigg(\sum_{t=2}^{T} E\Big[\big\{\mathbb{U}_t^\top \dot{B}_t(\rho)\mathbb{U}_t\big\}^2|\mathcal{F}\Big] +$$
$$\sum_{t_1\neq t_2} E\Big[\big\{\mathbb{U}_{t_1}^\top \dot{B}_{t_1}(\rho)\mathbb{U}_{t_1}\big\}\big\{\mathbb{U}_{t_2}^\top \dot{B}_{t_2}(\rho)\mathbb{U}_{t_2}\big\}|\mathcal{F}\Big]\Bigg)$$
$$= \frac{1}{4N^2(T-1)^2}\Bigg(\sum_{t=2}^{T} E\Big[2\mathrm{tr}\{\dot{B}_t^2(\rho)\}\Big] + \sum_{t_1=2}^{T}\sum_{t_2=2}^{T} E\Big[\mathrm{tr}\{\dot{B}_{t_1}(\rho)\}\mathrm{tr}\{\dot{B}_{t_2}(\rho)\}\Big]\Bigg)$$
$$= \frac{1}{4N^2(T-1)^2}\Bigg(\sum_{t=2}^{T} E\Big[2\mathrm{tr}\{\dot{B}_t^2(\rho)\} + \mathrm{tr}^2\{\dot{B}_t(\rho)\}\Big] + \sum_{t_1\neq t_2}\mathrm{tr}^2\Big\{\dot{\Omega}(\rho)\Omega(\rho)^{-1}\Big\}\Bigg).$$

Consequently, we have

$$(N(T-1))\mathrm{cov}(\mathcal{S}) = (N(T-1))\{E(\mathcal{S}^2) - E^2(\mathcal{S})\}$$
$$= \frac{1}{4N(T-1)}\sum_{t=2}^{T} E\Big[2\mathrm{tr}\{\dot{B}_t^2(\rho)\} + \mathrm{tr}^2\{\dot{B}_t(\rho)\}\Big] - \frac{1}{4N}\mathrm{tr}^2\Big\{\dot{\Omega}(\rho)\Omega(\rho)^{-1}\Big\}$$
$$= \frac{1}{4N(T-1)}\sum_{t=2}^{T} E\Bigg(2\mathrm{tr}\Big[\Big\{\dot{A}_t(\rho)\Omega(\rho)^{-1}\Big\}^2\Big] + \mathrm{tr}^2\Big\{\dot{A}_t(\rho)\Omega(\rho)^{-1}\Big\}\Bigg) - \frac{1}{4N}\mathrm{tr}^2\Big\{\dot{\Omega}(\rho)\Omega(\rho)^{-1}\Big\}$$
$$\to \Delta_{11}.$$

According to Lemma 2 of Sun and Wang (2019), we can obtain the asymptotic normality of $\mathcal{S}$. Similar results can be obtained for $\mathcal{S}_1$ following the same logic, where $E(\mathcal{S}_1) = 1/2$, and $(N(T-1))\mathrm{cov}(\mathcal{S}_1) = (4NT-4N)^{-1}\sum_{t=2}^{T} E\Bigg(2\mathrm{tr}\Big[\Big\{A_t(\rho)\Omega(\rho)^{-1}\Big\}^2\Big] +$

$\mathrm{tr}^2\{A_t(\rho)\Omega(\rho)^{-1}\}\Big] - N/4 \to \sigma^4\Delta_{22}$. In addition, we have

$$(N(T-1))\mathrm{cov}(\mathcal{S},\mathcal{S}_1) = (4NT-4N)^{-1}\sum_{t=2}^{T} E\Big[2\mathrm{tr}\Big\{\dot{A}_t(\rho)A_t(\rho)\Omega(\rho)^{-2}\Big\}$$

$$+\,\mathrm{tr}\Big\{\dot{A}_t(\rho)\Omega(\rho)^{-1}\Big\}\mathrm{tr}\Big\{A_t(\rho)\Omega(\rho)^{-1}\Big\}\Big] - \frac{1}{4}\mathrm{tr}\Big\{\dot{\Omega}(\rho)\Omega(\rho)^{-1}\Big\} \to -\sigma^2\Delta_{12}.$$

This completes the proof of Lemma 3.

Lemma 4 (Asymptotic normality of the ideal estimation equation)**.** *Under the conditions in Theorem 3, we have*

$$\frac{1}{\sqrt{NT}}\frac{d\ell_1(\theta)}{d\theta} \to_d N(0,\Delta),$$

where Δ is defined as in condition (C1).

Proof: Recall the likelihood function for θ,

$$\ell_1(\theta) = \ell_1(\rho,\sigma^2) = \frac{T-1}{2}\log\big|\Omega(\rho)\big| - \frac{N(T-1)}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{t=2}^{T}\mathcal{E}_t^\top A_t(\rho)\mathcal{E}_t.$$

The associated first derivative is

$$\frac{d\ell_1(\theta)}{d\rho} = \frac{T-1}{2}\mathrm{tr}\{\Omega^{-1}(\rho)\dot{\Omega}(\rho)\} - \frac{1}{2\sigma^2}\sum_{t=2}^{T}\mathcal{E}_t^\top \dot{A}_t(\rho)\mathcal{E}_t = -N(T-1)(\mathcal{S}-E\mathcal{S})$$

$$\frac{d\ell_1(\theta)}{d\sigma^2} = -\frac{N(T-1)}{2}\frac{1}{\sigma^2} + \frac{1}{2\sigma^4}\sum_{t=2}^{T}\mathcal{E}_t^\top A_t(\rho)\mathcal{E}_t = \frac{N(T-1)}{\sigma^2}(\mathcal{S}_1 - E\mathcal{S}_1).$$

Then, we have

$$\frac{1}{\sqrt{NT}}\frac{d\ell_1(\theta)}{d\theta} = \begin{pmatrix} \frac{1}{\sqrt{N(T-1)}}\frac{d\ell_1(\theta)}{d\rho} \\ \frac{1}{\sqrt{N(T-1)}}\frac{d\ell_1(\theta)}{d\sigma^2} \end{pmatrix} + o_p(1) = \begin{pmatrix} -\sqrt{N(T-1)}(\mathcal{S}-E\mathcal{S}) \\ \frac{1}{\sigma^2}\sqrt{N(T-1)}(\mathcal{S}_1 - E\mathcal{S}_1) \end{pmatrix} + o_p(1).$$

By Lemma 3 and condition (C1), we have $(N(T-1))^{-1/2}d\ell_1(\theta)/d\rho \rightarrow_d N(0,\Delta_{11})$, $(N(T-1))^{-1/2}d\ell_1(\theta)/d\sigma^2 \rightarrow_d N(0,\Delta_{22})$, and $(N(T-1))^{-1}\mathrm{cov}\{d\ell_1(\theta)/d\rho, d\ell_1(\theta)/d\sigma^2\}$ $\rightarrow \Delta_{12}$. Then, we can derive the asymptotic normal distribution of $(NT)^{-1/2}d\ell_1(\theta)/d\theta$ using the central limit theorem for linear-quadratic forms. Thus, the proof of Lemma 4 is completed.

Lemma 5. *Under the conditions in Theorem 3, the symmetric information matrix* $\Lambda_{1,NT} = -(NT)^{-1}\{d^2\ell_1(\theta)/d\theta d\theta^\top\}$, *we have* $\Lambda_{1,NT} \rightarrow_p \Lambda$.

PROOF: The second derivative for the likelihood of θ is specified as follows

$$
\begin{aligned}
-\frac{d^2\ell_1(\theta)}{d\rho^2} &= \frac{T-1}{2}\mathrm{tr}\Big[\{\Omega^{-1}(\rho)\dot{\Omega}(\rho)\}^2 - \Omega^{-1}(\rho)\ddot{\Omega}(\rho)\Big] + \frac{1}{2\sigma^2}\sum_{t=2}^{T}\mathcal{E}_t^\top \ddot{A}_t(\rho)\mathcal{E}_t \\
-\frac{d^2\ell_1(\theta)}{d\rho d\sigma^2} &= -\frac{d^2\ell_1(\theta)}{d\sigma^2 d\rho} = -\frac{1}{2\sigma^4}\sum_{t=2}^{T}\mathcal{E}_t^\top \dot{A}_t(\rho)\mathcal{E}_t \\
-\frac{d^2\ell_1(\theta)}{d\sigma^2 d\sigma^2} &= -\frac{N(T-1)}{2}\frac{1}{\sigma^4} + \frac{1}{\sigma^4}\frac{1}{\sigma^2}\sum_{t=2}^{T}\mathcal{E}_t^\top A_t(\rho)\mathcal{E}_t.
\end{aligned}
$$

Note that $E\{\sigma^{-2}\mathcal{E}_t^\top \ddot{A}_t(\rho)\mathcal{E}_t\} = \mathrm{tr}\{\Omega^{-1}(\rho)\ddot{\Omega}(\rho)\}$, then, we have

$$
\begin{aligned}
-\frac{1}{N(T-1)}E\Big\{\frac{d^2\ell_1(\theta)}{d\rho^2}\Big\} &= \frac{1}{2N}\mathrm{tr}\Big[\{\Omega^{-1}(\rho)\dot{\Omega}(\rho)\}^2\Big] \rightarrow \Lambda_{11} \\
-\frac{1}{N(T-1)}E\Big\{\frac{d^2\ell_1(\theta)}{d\rho d\sigma^2}\Big\} &= -\frac{1}{2N\sigma^2}\mathrm{tr}\{\Omega^{-1}(\rho)\dot{\Omega}(\rho)\} \rightarrow \Lambda_{12} \\
-\frac{1}{N(T-1)}E\Big\{\frac{d^2\ell_1(\theta)}{d\sigma^2 d\sigma^2}\Big\} &= \frac{1}{2\sigma^4} \rightarrow \Lambda_{22}.
\end{aligned}
$$

For $\Lambda_{1,NT}$, we have

$$
\Lambda_{1,NT} = \frac{1}{NT}\Big\{\frac{d^2\ell_1(\theta)}{d\theta d\theta^\top}\Big\} = \frac{1}{N(T-1)}E\Big\{\frac{d^2\ell_1(\theta)}{d\theta d\theta^\top}\Big\} + o_p(1) \rightarrow \Lambda.
$$

This follows by condition (C2). Thus, we have $\Lambda_{1,NT} \rightarrow_p \Lambda$. This completes the proof of Lemma 5.

Lemma 6 (Asymptotic normality of the feasible estimation equation)**.** *Assume conditions (C1)–(C3), we have*

$$\frac{1}{\sqrt{NT}}\Big\{\frac{d\ell_1(\theta)}{d\theta} - \frac{d\ell_2(\theta)}{d\theta}\Big\} = o_p(1).$$

Then, by the Lemma 4, we have $(NT)^{-1/2}d\ell_2(\theta)/d\theta \to_d N(0,\Delta)$.

Proof: Define $\widehat{\mathcal{P}}_t = \mathrm{diag}\{\hat{p}_{it}\hat{p}_{i(t-1)}\} \in \mathbb{R}^{N\times N}$. Recall the feasible weighted log-likelihood is specified as :

$$\ell_2(\theta) = \ell_2(\rho,\sigma^2) = \frac{T-1}{2}\log\big|\Omega(\rho)\big| - \frac{N(T-1)}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{t=2}^{T}\hat{\mathcal{E}}_t^\top \hat{A}_t(\rho)\hat{\mathcal{E}}_t,$$

where $\hat{A}_t(\rho) = \mathcal{Z}_t\hat{\mathcal{P}}_t^{-1}\{\Omega(\rho) - \mathrm{diag}(\Omega(\rho))\}\hat{\mathcal{P}}_t^{-1}\mathcal{Z}_t + \mathrm{diag}(\Omega(\rho))\hat{\mathcal{P}}_t^{-1}\mathcal{Z}_t$, $\hat{\mathcal{E}}_t = (\hat{\varepsilon}_{1t},\cdots,\hat{\varepsilon}_{Nt})^\top$, and $\hat{\varepsilon}_{it} = Y_{it} - \hat{\alpha}_i Y_{i(t-1)}$. The associated first derivative is

$$\begin{aligned}
\frac{d\ell_2(\theta)}{d\rho} &= \frac{T-1}{2}\mathrm{tr}\{\Omega^{-1}(\rho)\dot{\Omega}(\rho)\} - \frac{1}{2\sigma^2}\sum_{t=2}^{T}\hat{\mathcal{E}}_t^\top \dot{\hat{A}}_t(\rho)\hat{\mathcal{E}}_t \\
\frac{d\ell_2(\theta)}{d\sigma^2} &= -\frac{N(T-1)}{2}\frac{1}{\sigma^2} + \frac{1}{2\sigma^4}\sum_{t=2}^{T}\hat{\mathcal{E}}_t^\top \hat{A}_t(\rho)\hat{\mathcal{E}}_t.
\end{aligned}$$

Compared to $d\ell_1(\theta)/d\theta = (d\ell_1(\theta)/d\rho, d\ell_1(\theta)/d\sigma^2)^\top$, we have

$$-\Big\{\frac{d\ell_2(\theta)}{d\rho} - \frac{d\ell_1(\theta)}{d\rho}\Big\} = \frac{1}{2\sigma^2}\Big\{\sum_{t=2}^{T}\hat{\mathcal{E}}_t^\top \dot{\hat{A}}_t(\rho)\hat{\mathcal{E}}_t - \sum_{t=2}^{T}\mathcal{E}_t^\top \dot{A}_t(\rho)\mathcal{E}_t\Big\}. \tag{A.1}$$

To prove the Lemma, we can assume $\sigma^2 = 1$, and it suffices to show that

$$\frac{1}{\sqrt{N(T-1)}}\sum_{t=2}^{T}\Big\{\hat{\mathcal{E}}_t^\top \dot{\hat{A}}_t(\rho)\hat{\mathcal{E}}_t - \mathcal{E}_t^\top \dot{A}_t(\rho)\mathcal{E}_t\Big\} = \frac{1}{\sqrt{N(T-1)}}\sum_{t=2}^{T}L_{1,t} = o_p(1), \tag{A.2}$$

where $L_{1,t} = \hat{\mathcal{E}}_t^\top \hat{\dot{A}}_t(\rho)\hat{\mathcal{E}}_t - \mathcal{E}_t^\top \dot{A}_t(\rho)\mathcal{E}_t = \hat{\mathcal{E}}_t^\top \dot{A}_t(\rho)\hat{\mathcal{E}}_t - \mathcal{E}_t^\top \dot{A}_t(\rho)\mathcal{E}_t + \hat{\mathcal{E}}_t^\top\{\hat{\dot{A}}_t(\rho) - \dot{A}_t(\rho)\}\hat{\mathcal{E}}_t$. Note that $\mathcal{E}_t = \mathbb{Y}_t - \text{diag}(\mathbb{Y}_{t-1})\alpha$ and $\hat{\mathcal{E}}_t = \mathbb{Y}_t - \text{diag}(\mathbb{Y}_{t-1})\hat{\alpha} = \mathcal{E}_t - \text{diag}(\mathbb{Y}_{t-1})(\hat{\alpha} - \alpha)$, where $\mathbb{Y}_t = (Y_{1t}, \cdots, Y_{Nt})^\top$ and $\hat{\alpha} = (\hat{\alpha}_1, \cdots, \hat{\alpha}_N)^\top$. Then, we have $\hat{\mathcal{E}}_t^\top \dot{A}_t(\rho)\hat{\mathcal{E}}_t = \{\mathcal{E}_t - \text{diag}(\mathbb{Y}_{t-1})(\hat{\alpha} - \alpha)\}^\top \dot{A}_t(\rho)\{\mathcal{E}_t - \text{diag}(\mathbb{Y}_{t-1})(\hat{\alpha} - \alpha)\}$, and

$$\begin{aligned}\hat{\mathcal{E}}_t^\top \dot{A}_t(\rho)\hat{\mathcal{E}}_t - \mathcal{E}_t^\top \dot{A}_t(\rho)\mathcal{E}_t =& (\hat{\alpha} - \alpha)^\top \text{diag}(\mathbb{Y}_{t-1})\dot{A}_t(\rho)\text{diag}(\mathbb{Y}_{t-1})(\hat{\alpha} - \alpha) \\ & - 2\mathcal{E}_t^\top \dot{A}_t(\rho)\text{diag}(\mathbb{Y}_{t-1})(\hat{\alpha} - \alpha) = L_{1,t1} + L_{1,t2}.\end{aligned}$$

Then, we have

$$\sum_{t=2}^{T} L_{1,t2} = 2\sum_{t=2}^{T} \mathcal{E}_t^\top \dot{A}_t(\rho)\text{diag}(\mathbb{Y}_{t-1})(\alpha - \hat{\alpha}) \leq \sup_i(\alpha_i - \hat{\alpha}_i)\Big|2\sum_{t=2}^{T} \text{diag}(\mathbb{Y}_{t-1})\dot{A}_t(\rho)\mathcal{E}_t\Big|$$

Since $\sqrt{T}(\widetilde{\alpha}_i^{WLSE} - \alpha_i)$ is asymptotic normal and $(NT)^{-1/2}\sum_{t=2}^{T} \text{diag}(\mathbb{Y}_{t-1})\dot{A}_t(\rho)\mathcal{E}_t = O_p(1)$. We can show that $(NT)^{-1/2}\sum_{t=2}^{T} L_{1,t2} = o_p(1)$. Similarly, we can show that $(NT)^{-1/2}\sum_{t=2}^{T} L_{1,t1} = o_p(1)$. Next, considering

$$\begin{aligned}\hat{A}_t(\rho) - A_t(\rho) =& \mathcal{Z}_t\Big(\widehat{\mathcal{P}}_t^{-1} - \mathcal{P}_t^{-1}\Big)\Big[\Omega(\rho) - \text{diag}\{\Omega(\rho)\}\Big]\Big(\widehat{\mathcal{P}}_t^{-1} - \mathcal{P}_t^{-1}\Big)\mathcal{Z}_t \\ & + \mathcal{Z}_t\Big(\widehat{\mathcal{P}}_t^{-1} - \mathcal{P}_t^{-1}\Big)\Big[\Omega(\rho) - \text{diag}\{\Omega(\rho)\}\Big]\mathcal{P}_t^{-1}\mathcal{Z}_t \\ & + \mathcal{Z}_t\mathcal{P}_t^{-1}\Big[\Omega(\rho) - \text{diag}\{\Omega(\rho)\}\Big]\Big(\widehat{\mathcal{P}}_t^{-1} - \mathcal{P}_t^{-1}\Big)\mathcal{Z}_t + \text{diag}\{\Omega(\rho)\}\Big(\widehat{\mathcal{P}}_t^{-1} - \mathcal{P}_t^{-1}\Big)\mathcal{Z}_t.\end{aligned}$$

Then, we can show that $\hat{A}_t(\rho) - A_t(\rho) = o_p(1)$, and $\hat{\dot{A}}_t(\rho) - \dot{A}_t(\rho) = o_p(1)$. We have $L_{1,t3} = \hat{\mathcal{E}}_t^\top\{\hat{\dot{A}}_t(\rho) - \dot{A}_t(\rho)\}\hat{\mathcal{E}}_t = \mathcal{E}_t^\top\{\hat{\dot{A}}_t(\rho) - \dot{A}_t(\rho)\}\mathcal{E}_t + o_p(1) = o_p(1)$. Therefore, $(NT)^{-1/2}\sum_{t=2}^{T} L_{1,t} = o_p(1)$, i.e., (A.2) holds. In addition, $\frac{1}{\sqrt{NT}}\Big\{\frac{d\ell_2(\theta)}{d\sigma^2} - \frac{d\ell_1(\theta)}{d\sigma^2}\Big\} = \frac{1}{2\sigma^4}\sum_{t=2}^{T}\{\hat{\mathcal{E}}_t^\top \hat{A}_t(\rho)\hat{\mathcal{E}}_t - \mathcal{E}_t^\top A_t(\rho)\mathcal{E}_t\} = o_p(1)$. Thus, we have

$$\frac{1}{\sqrt{NT}}\Big\{\frac{d\ell_2(\theta)}{d\theta} - \frac{d\ell_1(\theta)}{d\theta}\Big\} = o_p(1).$$

Following Lemma 4, we have the asymptotic normality of $(NT)^{-1/2}d\ell_2(\theta)/d\theta$. This completes the proof of Lemma 6.

Lemma 7. *The symmetric information matrix* $\Lambda_{2,NT} = -(NT)^{-1}\{d^2\ell_2(\theta)/d\theta d\theta^\top\}$, *we have* $\Lambda_{2,NT} = \Lambda_{1,NT} + o_p(1) \to_p \Lambda$.

PROOF: The second derivative of $\ell_2(\theta)$ for the likelihood of θ is specified as follows

$$
\begin{aligned}
-\frac{d^2\ell_2(\theta)}{d\rho^2} &= \frac{T-1}{2}\mathrm{tr}\Big[\{\Omega^{-1}(\rho)\dot{\Omega}(\rho)\}^2 - \Omega^{-1}(\rho)\ddot{\Omega}(\rho)\Big] + \frac{1}{2\sigma^2}\sum_{t=2}^{T}\hat{\mathcal{E}}_t^\top \hat{\hat{A}}_t(\rho)\hat{\mathcal{E}}_t \\
-\frac{d^2\ell_2(\theta)}{d\rho d\sigma^2} &= -\frac{d^2\ell_2(\theta)}{d\sigma^2 d\rho} = -\frac{1}{2\sigma^4}\sum_{t=2}^{T}\hat{\mathcal{E}}_t^\top \hat{A}_t(\rho)\hat{\mathcal{E}}_t \\
-\frac{d^2\ell_2(\theta)}{d\sigma^2 d\sigma^2} &= -\frac{N(T-1)}{2}\frac{1}{\sigma^4} + \frac{1}{\sigma^4}\frac{1}{\sigma^2}\sum_{t=2}^{T}\hat{\mathcal{E}}_t^\top \hat{A}_t(\rho)\hat{\mathcal{E}}_t.
\end{aligned}
$$

Following the similar proof of (A.1), we can easily verify $\Lambda_{2,NT} - \Lambda_{1,NT} = o_p(1)$. By Lemma 5, $\Lambda_{1,NT} - \Lambda = o_p(1)$. Therefore, we have $\Lambda_{2,NT} - \Lambda = o_p(1)$. Thus, the proof of Lemma 7 is completed.

Appendix B: Proof of Theorem 1

PROOF: For the ideal WLS estimation of α_i, define

$$
A_{2i} = \frac{1}{T-1}\sum_{t=2}^{T}\frac{Z_{it}}{p_{it}}\frac{Z_{i(t-1)}}{p_{i(t-1)}}Y_{i(t-1)}^2, \quad B_{2i} = \sqrt{\frac{1}{T-1}}\sum_{t=2}^{T}\frac{Z_{it}}{p_{it}}\frac{Z_{i(t-1)}}{p_{i(t-1)}}Y_{i(t-1)}\varepsilon_{it}.
$$

Then, we have

$$
\sqrt{T-1}(\widetilde{\alpha}_i^{WLSE} - \alpha_i) = A_{2i}^{-1}B_{2i}. \tag{A.3}
$$

With Slutsky's Theorem, it is sufficient to prove that

$$A_{2i} \to_p \sigma^2_{Y_i}, \tag{A.4}$$

$$B_{2i} \to_d N(0, \sigma^2_{1i}). \tag{A.5}$$

For A_{2i}, we have

$$\begin{aligned} E(A_{2i}) &= (T-1)^{-1} \sum_{t=2}^{T} E\Big[Y^2_{i(t-1)} E\Big\{\frac{Z_{it}}{p_{it}} E\big(\frac{Z_{i(t-1)}}{p_{i(t-1)}} \big| \mathcal{F}\big) \Big| \mathcal{F}\Big\}\Big] \\ &= (T-1)^{-1} \sum_{t=2}^{T} E\big\{Y^2_{i(t-1)}\big\} = \sigma^2_{Y_i}. \end{aligned}$$

Then, by the law of large numbers, we have $A_{2i} \to_p \sigma^2_{Y_i}$. For B_{2i}, the expectation and variance are computed as follows

$$\begin{aligned} E(B_{2i}) =& \sqrt{\frac{1}{T-1}} \sum_{t=2}^{T} E\Big[Y_{i(t-1)}\varepsilon_{it} E\Big\{\frac{Z_{it}}{p_{it}} E\big(\frac{Z_{i(t-1)}}{p_{i(t-1)}} \big| \mathcal{F}\big) \Big| \mathcal{F}\Big\}\Big] = 0 \\ E(B^2_{2i}) =& \frac{1}{T-1} \sum_{t=2}^{T} E\Big[Y^2_{i(t-1)}\varepsilon^2_{it} E\Big\{\frac{Z^2_{it}}{p^2_{it}} E\big(\frac{Z^2_{i(t-1)}}{p^2_{i(t-1)}} \big| \mathcal{F}\big) \Big| \mathcal{F}\Big\}\Big] \\ &+ \frac{2}{T-1} \sum_{t_1 \neq t_2} \quad E\Big[Y_{i(t_1-1)} Y_{i(t_2-1)} \varepsilon_{it_1}\varepsilon_{it_2} E\Big\{\frac{Z_{it_2}}{p_{it_2}} \frac{Z_{i(t_2-1)}}{p_{i(t_2-1)}} E\big(\frac{Z_{it_1}}{p_{it_1}} \frac{Z_{i(t_1-1)}}{p_{i(t_1-1)}} \big| \mathcal{F}\big) \Big| \mathcal{F}\Big\}\Big] \\ =& \frac{1}{T-1} \sum_{t=2}^{T} E\Big\{Y^2_{i(t-1)}\varepsilon^2_{it} p^{-1}_{it} p^{-1}_{i(t-1)}\Big\} + \frac{2}{T-1} \sum_{t_1 \neq t_2} E\Big\{Y_{i(t_1-1)} Y_{i(t_2-1)} \varepsilon_{it_1}\varepsilon_{it_2}\Big\} \\ =& E\Big\{Y^2_{i(t-1)}\varepsilon^2_{it} p^{-1}_{it} p^{-1}_{i(t-1)}\Big\} = \sigma^2_{1i}. \end{aligned}$$

Then, we have $(T-1)^{-1/2} B_{2i} \to_p 0$. Thus, $\widetilde{\alpha}_i^{WLSE} - \alpha_i \to_p 0$. By a slight extension of Example 7.15 in Hamilton (1994), we can obtain the result of equation (A.5). Thus, the proof of Theorem 1 has been completed.

Appendix C: Proof of Theorem 2

By the theory of logistic regression, we can obtain $(\hat{\beta} - \beta) = O_p(1/\sqrt{NT})$. Then with Delta method, we have

$$\hat{p}_{it} - p_{it} = O_p(1/\sqrt{NT}), \quad \hat{p}_{it}^{-1} - p_{it}^{-1} = O_p(1/\sqrt{NT}),$$
$$\hat{p}_{it}\hat{p}_{i(t-1)} - p_{it}p_{i(t-1)} = O_p(1/\sqrt{NT}).$$

For the feasible WLSE, define

$$A_{3i} = \frac{1}{T-1}\sum_{t=2}^{T} \frac{Z_{it}}{\hat{p}_{it}} \frac{Z_{i(t-1)}}{\hat{p}_{i(t-1)}} Y_{i(t-1)}^2, \quad B_{3i} = \frac{1}{\sqrt{T-1}}\sum_{t=2}^{T} \frac{Z_{it}}{\hat{p}_{it}} \frac{Z_{i(t-1)}}{\hat{p}_{i(t-1)}} Y_{i(t-1)}\varepsilon_{it}.$$

Then, we have

$$\sqrt{T-1}(\widehat{\alpha}_i^{WLS} - \alpha_i) = A_{3i}^{-1} B_{3i}. \tag{A.6}$$

With Slutsky's Theorem, it is sufficient to prove that

$$A_{3i} \to_p \sigma_{Y_i}^2, \tag{A.7}$$
$$B_{3i} \to_d N(0, \sigma_{1i}^2). \tag{A.8}$$

For A_{3i}, we have

$$\begin{aligned} E(A_{3i}) &= (T-1)^{-1}\sum_{t=2}^{T} E\Big[Y_{i(t-1)}^2 E\Big\{\frac{Z_{it}}{\hat{p}_{it}} E\big(\frac{Z_{i(t-1)}}{\hat{p}_{i(t-1)}}\big|\mathcal{F}\big)\Big|\mathcal{F}\Big\}\Big] \\ &= (T-1)^{-1}\sum_{t=2}^{T} E\big\{Y_{i(t-1)}^2 \frac{p_{it}}{\hat{p}_{it}}\frac{p_{i(t-1)}}{\hat{p}_{i(t-1)}}\big\} = E(A_{2i}) + O_p(1/\sqrt{NT}) \to_p \sigma_{Y_i}^2. \end{aligned}$$

Then, by the law of large numbers, we have $A_{3i} \to_p \sigma_{Y_i}^2$. For B_3, the expectation is computed as

$$E(B_{3i}) = \sqrt{\frac{1}{T-1}}\sum_{t=2}^{T} E\Big[Y_{i(t-1)}\varepsilon_{it}E\Big\{\frac{Z_{it}}{\hat{p}_{it}}E\big(\frac{Z_{i(t-1)}}{\hat{p}_{i(t-1)}}\big|\mathcal{F}\big)\Big|\mathcal{F}\Big\}\Big] \to_p 0.$$

Its variance is computed as

$$\begin{aligned}
E(B_{3i}^2) = {} & \frac{1}{T-1}\sum_{t=2}^{T} E\Big[Y_{i(t-1)}^2\varepsilon_{it}^2E\Big\{\frac{Z_{it}^2}{\hat{p}_{it}^2}E\big(\frac{Z_{i(t-1)}^2}{\hat{p}_{i(t-1)}^2}\big|\mathcal{F}\big)\Big|\mathcal{F}\Big\}\Big] + \frac{2}{T-1}\sum_{t_1\neq t_2} \\
& E\Big[Y_{i(t_1-1)}Y_{i(t_2-1)}\varepsilon_{it_1}\varepsilon_{it_2}E\Big\{\frac{Z_{it_2}}{\hat{p}_{it_2}}\frac{Z_{i(t_2-1)}}{\hat{p}_{i(t_2-1)}}E\big(\frac{Z_{it_1}}{\hat{p}_{it_1}}\frac{Z_{i(t_1-1)}}{\hat{p}_{i(t_1-1)}}\big|\mathcal{F}\big)\Big|\mathcal{F}\Big\}\Big] \\
= {} & \frac{1}{T-1}\sum_{t=2}^{T} E\Big\{Y_{i(t-1)}^2\varepsilon_{it}^2\frac{p_{it}p_{i(t-1)}}{\hat{p}_{it}^2\hat{p}_{i(t-1)}^2}\Big\} + \frac{2}{T-1}\sum_{t_1\neq t_2} E\Big\{Y_{i(t_1-1)}Y_{i(t_2-1)}\varepsilon_{it_1}\varepsilon_{it_2}\frac{p_{it}p_{i(t-1)}}{\hat{p}_{it}\hat{p}_{i(t-1)}}\Big\} \\
= {} & \frac{1}{T-1}\sum_{t=2}^{T} E\Big\{Y_{i(t-1)}^2\varepsilon_{it}^2 p_{it}^{-1}p_{i(t-1)}^{-1}\Big\}\Big\{1+O_p\Big(\frac{1}{\sqrt{NT}}\Big)\Big\} \\
& + \frac{2}{T-1}\sum_{t_1\neq t_2} E\Big\{Y_{i(t_1-1)}Y_{i(t_2-1)}\varepsilon_{it_1}\varepsilon_{it_2}\Big\}\Big\{1+O_p\Big(\frac{1}{\sqrt{NT}}\Big)\Big\} \\
= {} & E(B_{2i}^2) + O_p(1/\sqrt{NT}) \to_p \sigma_{1i}^2.
\end{aligned}$$

Then we have $(T-1)^{-1/2}B_{3i} \to_p 0$. Thus, $\widehat{\alpha}_i^{WLS} - \alpha_i \to_p 0$. By a slight extension of Example 7.15 in Hamilton (1994), we can obtain the result of equation (A.8). Thus, the proof of Theorem 2 has been completed.

Appendix D: Proof of Theorem 3

The proof is similar to the proof of Theorem 1 in Sun and Wang (2019). The details are given in the following two steps.

Step 1: To demonstrate the consistency of $\widetilde{\theta}^{MLE}$, we first show that there exists some constant $C > 0$ so that

$$\lim_{N,T\to\infty} P\{\sup_{\|t\|=C} \ell_1(\theta + a(NT)^{-1/2}t) < \ell_1(\theta)\} = 1. \tag{A.9}$$

Applying Taylor's expansion to $\ell_1(\theta + a(NT)^{-1/2}t)$, we have

$$\begin{aligned} R_{NT}(\theta) &= \ell_1(\theta + (NT)^{-1/2}t) - \ell_1(\theta) \\ &= (NT)^{-1/2}t^\top \frac{d\ell_1(\theta)}{d\theta} + (2NT)^{-1}t^\top \frac{d^2\ell_1(\theta)}{d\theta d\theta^\top} t + o_p(1). \end{aligned} \tag{A.10}$$

From Lemma 2, we know that $(NT)^{-1/2} d\ell_1(\theta)/d\theta = O_p(1)$. In addition, we have $(NT)^{-1}d^2\ell_1(\theta)/d\theta d\theta^\top = -\Lambda_{1,NT} + o_p(1) \to -\Lambda_1$. By the similar arguments of Theorem 1 in Sun and Wang (2019), we can obtain the consistency of $\widetilde{\theta}^{MLE}$. Specifically, the second term of (A.10) is quadratic and negative, and the first term is linear. Then, for a sufficiently large C, the second term would dominate the first one. Thus, (A.9) holds. In addition, we maximize $l(\theta)$ at $\widetilde{\theta}^{MLE}$, which means $\widetilde{\theta}^{MLE}$ is controlled by $\{\theta + (NT)^{-1/2}t : \|t\| \le C\}$. Consequently, $\|\widetilde{\theta}^{MLE}\| = O_p(NT)^{-1/2}$.

Step 2: To demonstrate the asymptotic normality of $\widetilde{\theta}^{MLE}$, we take a Taylor expansion of $d\ell_1(\widetilde{\theta}^{MLE})/d\theta = 0$ at the true value of θ. This leads to

$$\sqrt{NT}(\widetilde{\theta}^{MLE} - \theta) = \Big\{ -\frac{1}{NT}\frac{d^2\ell_1(\breve{\theta})}{d\theta d\theta^\top}\Big\}^{-1} \frac{1}{\sqrt{NT}}\frac{d\ell_1(\theta)}{d\theta},$$

where $\breve{\theta}$ lies between $\widetilde{\theta}^{MLE}$ and θ. By the Slutsky's Theorem and Lemma 4, it suffices to show

$$\frac{1}{NT}\frac{d^2\ell_1(\breve{\theta})}{d\theta d\theta^\top} - \frac{1}{NT}\frac{d^2\ell_1(\theta)}{d\theta d\theta^\top} = o_p(1). \tag{A.11}$$

Consequently, we only to consider each block of the two related matrices, respectively.

First, we consider

$$\frac{1}{NT}\frac{d^2\ell_1(\breve{\theta})}{d\rho^2} - \frac{1}{NT}\frac{d^2\ell_1(\theta)}{d\rho^2} = L_{2,NT} = o_p(1). \tag{A.12}$$

The details are given below. Note that $\ddot{\Omega}(\breve{\rho}) \equiv \ddot{\Omega}(\rho)$, and

$$\Omega^{-1}(\breve{\rho})\ddot{\Omega}(\breve{\rho}) = \Omega^{-1}(\rho)\ddot{\Omega}(\rho) + \Omega^{-1}(\bar{\rho})\ddot{\Omega}(\bar{\rho})\Omega^{-1}(\bar{\rho})\ddot{\Omega}(\bar{\rho})(\breve{\rho}-\rho) \tag{A.13}$$

$$\{\Omega^{-1}(\breve{\rho})\dot{\Omega}(\breve{\rho})\}^2 = \{\Omega^{-1}(\rho)\dot{\Omega}(\rho)\}^2 + 2\Big[\Omega^{-1}(\bar{\rho})\big\{I - \dot{\Omega}(\bar{\rho})\Omega^{-1}(\bar{\rho})\big\}\ddot{\Omega}(\bar{\rho})\Big](\breve{\rho}-\rho) \tag{A.14}$$

with $\bar{\rho}$ lying between $\breve{\rho}$ and ρ. We know that

$$\begin{aligned} L_{2,NT} \doteq{}& \frac{1}{2N}\Big[\mathrm{tr}\{\Omega^{-1}(\breve{\rho})\dot{\Omega}(\breve{\rho})\}^2 - \mathrm{tr}\{\Omega^{-1}(\rho)\dot{\Omega}(\rho)\}^2\Big] \\ &- \frac{1}{2N}\Big[\mathrm{tr}\{\Omega^{-1}(\breve{\rho})\ddot{\Omega}(\breve{\rho}) - \mathrm{tr}\{\Omega^{-1}(\rho)\ddot{\Omega}(\rho)\}\Big] + \frac{1}{2}\Big(\breve{\sigma}^{-2} - \sigma^{-2}\Big)\sum_{t=2}^{T}\mathcal{E}_t^\top \ddot{A}_t(\breve{\rho})\mathcal{E}_t \\ &+ \frac{1}{2\sigma^2}\Big[\sum_{t=2}^{T}\mathcal{E}_t^\top \ddot{A}_t(\breve{\rho})\mathcal{E}_t - \sum_{t=2}^{T}\mathcal{E}_t^\top \ddot{A}_t(\rho)\mathcal{E}_t\Big] \\ ={}& L_{2,NT,1} + L_{2,NT,2} + L_{2,NT,3} + L_{2,NT,4}. \end{aligned}$$

With the consistency of $\widetilde{\rho}^{MLE}$ and (A.13), we have $L_{NT,1} = o_p(1)$. Similarly, by (A.14), we have $L_{NT,2} = o_p(1)$. By a similar calculation of (A.1) and the consistency of $\widetilde{\theta}^{MLE} = (\widetilde{\rho}^{MLE}, \widetilde{\sigma^2}^{MLE})^\top$, we also have $L_{NT,3} = o_p(1)$, and $L_{NT,4} = o_p(1)$. Then, the equation (A.12) holds, and other blocks of $\frac{1}{NT}\frac{d^2\ell_1(\breve{\theta})}{d\theta d\theta^\top} - \frac{1}{NT}\frac{d^2\ell_1(\theta)}{d\theta d\theta^\top}$ are also $o_p(1)$. Thus, (A.11) holds, and we have completed the proof of Theorem 3.

Appendix E: Proof of Theorem 4

The proof of Theorem 4 is similar to the proof of Theorem 3. We can prove the consistency by the same argument with Lemmas 6 and 7. To show the asymptotic normality of $\widehat{\theta}^{MLE}$, we apply Taylor expansion to $d\ell_2(\widehat{\theta})/d\theta = 0$ at the true value of θ. This leads to

$$\sqrt{NT}(\widehat{\theta}^{MLE} - \theta) = \Big\{ -\frac{1}{NT}\frac{d^2\ell_2(\breve{\theta})}{d\theta d\theta^\top}\Big\}^{-1}\frac{1}{\sqrt{NT}}\frac{d\ell_2(\theta)}{d\theta},$$

where $\breve{\theta}$ lies between $\widehat{\theta}$ and θ. By the Slutsky's Theorem and Lemmas 6 and 7, it suffices to show

$$\frac{1}{NT}\frac{d^2\ell_2(\breve{\theta})}{d\theta d\theta^\top} - \frac{1}{NT}\frac{d^2\ell_2(\theta)}{d\theta d\theta^\top} = o_p(1). \tag{A.15}$$

Consequently, similar to (A.12), we can consider each block of the two related matrices, respectively. First, we have

$$\begin{aligned} L_{3,NT} = \frac{1}{NT}\frac{d^2\ell_2(\breve{\theta})}{d\rho^2} - \frac{1}{NT}\frac{d^2\ell_2(\theta)}{d\rho^2} &\doteq \frac{1}{2N}\Big[\mathrm{tr}\{\Omega^{-1}(\breve{\rho})\dot{\Omega}(\breve{\rho})\}^2 - \mathrm{tr}\{\Omega^{-1}(\rho)\dot{\Omega}(\rho)\}^2\Big] \\ &- \frac{1}{2N}\Big[\mathrm{tr}\{\Omega^{-1}(\breve{\rho})\ddot{\Omega}(\breve{\rho}) - \mathrm{tr}\{\Omega^{-1}(\rho)\ddot{\Omega}(\rho)\}\Big] \\ &+ \frac{1}{2NT}(\breve{\sigma}^{-2} - \sigma^{-2})\sum_{t=2}^{T}\widehat{\mathcal{E}}_t^\top \ddot{A}_t(\breve{\rho})\widehat{\mathcal{E}}_t + \frac{1}{2NT\sigma^2}\sum_{t=2}^{T}\widehat{\mathcal{E}}_t^\top\{\ddot{A}_t(\breve{\rho}) - \ddot{A}_t(\rho)\}\widehat{\mathcal{E}}_t \\ = L_{3,NT,1} + L_{3,NT,2} + L_{3,NT,3} + L_{3,NT,4} = o_p(1). \end{aligned}$$

This can be established by a similar calculation for $\{L_{2,NT,k}\}_{k=1}^4$ in Theorem 3. Then, $L_{3,NT} = o_p(1)$, and other blocks of $\frac{1}{NT}\frac{d^2\ell_2(\breve{\theta})}{d\theta d\theta^\top} - \frac{1}{NT}\frac{d^2\ell_2(\theta)}{d\theta d\theta^\top}$ are also $o_p(1)$. Thus, (A.15) holds, and this completes the proof of Theorem 4.

References

Hamilton, J. (1994). *Time Series Analysis*. Princeton, NJ: Princeton University Press.

Lee, L.-F. (2004). Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. *Econometrica*, 72(6):1899–1925.

Sun, Z. and Wang, H. (2019). Network imputation for a spatial autoregression model with incomplete data. *Statistica Sinica, forthcoming*.