A Mixture Generalized Estimating Equations Approach for Complex Spatially-Dependent Data

Huichen Zhu[1]∗,Fangzheng Lin[2], Huixia Judy Wang[3], Zhongyi Zhu[4]

[1] *Department of Statistics, The Chinese University of Hong Kong, China*

[2] *Department of Statistics, Fudan University, China*

[*] *Corresponding author*

[3] *Department of Statistics, The George Washington University*

[4] *Department of Statistics, Fudan University, China*

Supplementary Material

S1 Technical conditions

THe following conditions are some regularity conditions needed for the theoretical results in the main paper.

(C1) *These exists a constant* $\Delta > 0$ *so that for all* $a \neq b \in \mathbb{N}^+$, $||\boldsymbol{s}_a - \boldsymbol{s}_b|| \geq \Delta$.

(C2) *For all fixed* $h \in \mathbb{R}^+$, $k = 1, \cdots, K$, $\rho_k(h, \boldsymbol{\alpha}_k)$ *is* $p_k + 1$ *times continuously differentiable w.r.t.* $\boldsymbol{\alpha}_k$. *For all* $i_1, \cdots, i_{p_k} \in \mathbb{N}$ *satisfying*

$i_1 + \cdots + i_{p_k} \leq p_k + 1$, there exists a positive constant $A < +\infty$, so that for all $h \in \mathbb{R}^+$ and $\boldsymbol{\alpha}_k \in \boldsymbol{\Psi}_{\boldsymbol{\alpha}_k}$,

$$|\frac{\partial^{i_1}}{\partial \alpha_{k,1}^{i_1}} \cdots \frac{\partial^{i_{p_k}}}{\partial \alpha_{k,p_k}^{i_{p_k}}} \rho_k(h, \boldsymbol{\alpha}_k)| \leq \frac{A}{1 + |h|^3}.$$

The Fourier transforms $\widetilde{\rho}_k(\omega, \boldsymbol{\alpha}_k)$ of $\rho_k(h, \boldsymbol{\alpha}_k)$ are jointly continuous and strictly positive in $\omega \in \mathbb{R}$ and $\boldsymbol{\alpha}_k \in \boldsymbol{\Psi}_{\alpha_k}$.

(C3) $\inf_n \lambda_{\min}(\boldsymbol{\Sigma}_0) > 0$, $\sup_n \lambda_{\max}(\boldsymbol{\Sigma}_0) < \infty$, where $\lambda_{\max}(\cdot)$ and $\lambda_{\min}(\cdot)$ represent the maximum and minimum eigenvalues of a matrix, respectively.

(C4) The first-order and second-order derivatives of the link function $\mu(\cdot)$ are finite in the parameter space, the variance function $V(\cdot)$ is positive, and the first-order derivative of $V(\cdot)$ is finite.

(C5) $\rho_{k,l}(m) = o(m^{-2})$ for $k+l \leq 4$, where the $\rho-$product mixing coefficients are defined by

$$\rho_{k,l}(m) = \sup\left[\left|\mathrm{Cov}\left\{\prod_{\boldsymbol{s}_i \in \Lambda_1} Y(\boldsymbol{s}_i), \prod_{\boldsymbol{s}_j \in \Lambda_2} Y(\boldsymbol{s}_j)\right\}\right| : \mathrm{E}(|Y(s)|^{2+\delta}) < \infty, \right.$$
$$\left. |\Lambda_1| \leq k, |\Lambda_2| \leq l, d(\Lambda_1, \Lambda_2) \geq m\right],$$

Λ denotes a subset of locations, and $|\Lambda|$ denotes the number of elements in Λ, $d(\Lambda_1, \Lambda_2) = \inf\{||\boldsymbol{s}_1 - \boldsymbol{s}_2|| : s_1 \in \Lambda_1, s_2 \in \Lambda_2\}$, and $\delta, m \in \mathbb{R}^+$, $k, l \in \mathbb{N}^+$.

(C6) The true correlation matrix of $\boldsymbol{Y}$ is $\sum_{k=1}^{K} \pi_k^0 \boldsymbol{R}^{(k)}(\boldsymbol{\alpha}_k^0)$, and $\mathrm{E}\{\ell_n(\boldsymbol{\psi}, \boldsymbol{\beta}_0; \gamma_2)\}$ attains its minimum over $\boldsymbol{\Psi}$ at $\boldsymbol{\psi}_0 = (\boldsymbol{\alpha}_k^{0\,T}, \pi_k^0)_{k=1}^K$.

S2 Technical proofs

Lemma 1. *Under Conditions (C1)-(C4) and the identifiability condition 1 in the main paper, suppose there exists an estimator of regression parameter, satisfying $\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 = O_p(n^{-1/2})$, then for the estimator of nuisance parameter through minimizing $\ell_n(\boldsymbol{\psi}, \hat{\boldsymbol{\beta}})$, denoted as $\hat{\boldsymbol{\psi}}$, we have for every $\zeta > 0$, as $n \to \infty$,*

$$P\left\{d(\hat{\boldsymbol{\psi}}, \boldsymbol{\Psi}_0) \geq \zeta\right\} \to 0.$$

Proof: First, we will prove $\sup_{\boldsymbol{\psi}\in\boldsymbol{\Psi}} |\ell_n(\boldsymbol{\psi}, \hat{\boldsymbol{\beta}}) - \bar{\ell}(\boldsymbol{\psi}, \boldsymbol{\beta}_0)| = o_p(1)$. Let $I_{1n}(\boldsymbol{\psi}) = \ell_n(\boldsymbol{\psi}, \hat{\boldsymbol{\beta}}) - \ell_n(\boldsymbol{\psi}, \boldsymbol{\beta}_0)$, $I_{2n}(\boldsymbol{\psi}) = \ell_n(\boldsymbol{\psi}, \boldsymbol{\beta}_0) - E\{\ell_n(\boldsymbol{\psi}, \boldsymbol{\beta}_0)\}$, and $I_{3n}(\boldsymbol{\psi}) = E\{\ell_n(\boldsymbol{\psi}, \boldsymbol{\beta}_0)\} - \bar{\ell}(\boldsymbol{\psi}, \boldsymbol{\beta}_0)$. Notice that

$$\ell_n(\boldsymbol{\psi}, \hat{\boldsymbol{\beta}}) - \bar{\ell}(\boldsymbol{\psi}, \boldsymbol{\beta}_0) = I_{1n}(\boldsymbol{\psi}) + I_{2n}(\boldsymbol{\psi}) + I_{3n}(\boldsymbol{\psi}), \tag{S2.1}$$

thus we only need to prove that $I_{1n}(\boldsymbol{\psi})$, $I_{2n}(\boldsymbol{\psi})$ and $I_{3n}(\boldsymbol{\psi})$ converge uniformly to zero over $\boldsymbol{\Psi}$.

For $I_{1n}(\boldsymbol{\psi})$, we have

$$
\begin{aligned}
I_{1n}(\boldsymbol{\psi}) &= \ell_n(\boldsymbol{\psi}, \hat{\boldsymbol{\beta}}) - \ell_n(\boldsymbol{\psi}, \boldsymbol{\beta}_0) \\
&= \frac{1}{n}\{\boldsymbol{Y} - \boldsymbol{\mu}(\hat{\boldsymbol{\beta}})\}^T \boldsymbol{A}(\hat{\boldsymbol{\beta}})^{-1/2} \boldsymbol{R}(\boldsymbol{\psi})^{-1} \boldsymbol{A}(\hat{\boldsymbol{\beta}})^{-1/2} \{\boldsymbol{Y} - \boldsymbol{\mu}(\hat{\boldsymbol{\beta}})\} \\
&\quad - \frac{1}{n}\{\boldsymbol{Y} - \boldsymbol{\mu}(\boldsymbol{\beta}_0)\}^T \boldsymbol{A}(\boldsymbol{\beta}_0)^{-1/2} \boldsymbol{R}(\boldsymbol{\psi})^{-1} \boldsymbol{A}(\boldsymbol{\beta}_0)^{-1/2} \{\boldsymbol{Y} - \boldsymbol{\mu}(\boldsymbol{\beta}_0)\} \\
&= \frac{1}{n}[\boldsymbol{Y} - \boldsymbol{\mu}(\boldsymbol{\beta}_0) + \boldsymbol{\nabla}\{\boldsymbol{\mu}(\boldsymbol{\beta}^*)\}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)]^T \boldsymbol{A}(\hat{\boldsymbol{\beta}})^{-1/2} \boldsymbol{R}(\boldsymbol{\psi})^{-1} \\
&\quad \boldsymbol{A}(\hat{\boldsymbol{\beta}})^{-1/2} [\boldsymbol{Y} - \boldsymbol{\mu}(\boldsymbol{\beta}_0) + \boldsymbol{\nabla}\{\boldsymbol{\mu}(\boldsymbol{\beta}^*)\}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)] \\
&\quad - \frac{1}{n}\{\boldsymbol{Y} - \boldsymbol{\mu}(\boldsymbol{\beta}_0)\}^T \boldsymbol{A}(\boldsymbol{\beta}_0)^{-1/2} \boldsymbol{R}(\boldsymbol{\psi})^{-1} \boldsymbol{A}(\boldsymbol{\beta}_0)^{-1/2} \{\boldsymbol{Y} - \boldsymbol{\mu}(\boldsymbol{\beta}_0)\} \\
&= \frac{2}{n}[\boldsymbol{\nabla}\{\boldsymbol{\mu}(\boldsymbol{\beta}^*)\}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)]^T \boldsymbol{A}(\hat{\boldsymbol{\beta}})^{-1/2} \boldsymbol{R}(\boldsymbol{\psi})^{-1} \boldsymbol{A}(\hat{\boldsymbol{\beta}})^{-1/2} \{\boldsymbol{Y} - \boldsymbol{\mu}(\boldsymbol{\beta}_0)\} \\
&\quad + \frac{1}{n}[\boldsymbol{\nabla}\{\boldsymbol{\mu}(\boldsymbol{\beta}^*)\}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)]^T \boldsymbol{A}(\hat{\boldsymbol{\beta}})^{-1/2} \boldsymbol{R}(\boldsymbol{\psi})^{-1} \boldsymbol{A}(\hat{\boldsymbol{\beta}})^{-1/2} [\boldsymbol{\nabla}\{\boldsymbol{\mu}(\boldsymbol{\beta}^*)\}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)] \\
&\quad + \frac{1}{n}\{\boldsymbol{Y} - \boldsymbol{\mu}(\boldsymbol{\beta}_0)\}^T \left\{\boldsymbol{A}(\hat{\boldsymbol{\beta}})^{-1/2} - \boldsymbol{A}(\boldsymbol{\beta}_0)^{-1/2}\right\} \boldsymbol{R}(\boldsymbol{\psi})^{-1} \boldsymbol{A}(\hat{\boldsymbol{\beta}})^{-1/2} \{\boldsymbol{Y} - \boldsymbol{\mu}(\boldsymbol{\beta}_0)\} \\
&\quad + \frac{1}{n}\{\boldsymbol{Y} - \boldsymbol{\mu}(\boldsymbol{\beta}_0)\}^T \left\{\boldsymbol{A}(\hat{\boldsymbol{\beta}})^{-1/2} - \boldsymbol{A}(\boldsymbol{\beta}_0)^{-1/2}\right\} \boldsymbol{R}(\boldsymbol{\psi})^{-1} \boldsymbol{A}(\boldsymbol{\beta}_0)^{-1/2} \{\boldsymbol{Y} - \boldsymbol{\mu}(\boldsymbol{\beta}_0)\},
\end{aligned}
\tag{S2.2}
$$

where $\boldsymbol{\nabla}$ represents the first-order partial derivative w.r.t. $\boldsymbol{\beta}$, and $\boldsymbol{\beta}^*$ satisfies $||\boldsymbol{\beta}^* - \boldsymbol{\beta}_0|| \leq ||\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0||$. Under Conditions (C1)-(C2), by Theorem 5 in Bachoc and Furrer (2016), we have $\inf_{\boldsymbol{\psi} \in \boldsymbol{\Psi}}{}_n \lambda_{\min}\{\boldsymbol{R}(\boldsymbol{\psi})\} > 0$, and by Lemma 6 in Furrer et al. (2016), we have $\sup_{\boldsymbol{\psi} \in \boldsymbol{\Psi}}{}_n \lambda_{\max}\{\boldsymbol{R}(\boldsymbol{\psi})\} < \infty$. Together with assumption (C4) and $\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 = O_p(n^{-1/2})$, we can obtain that

$$\sup_{\boldsymbol{\psi}\in\boldsymbol{\Psi}}\left|\frac{1}{n}[\boldsymbol{\nabla}\{\boldsymbol{\mu}(\boldsymbol{\beta}^*)\}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)]^T \boldsymbol{A}(\hat{\boldsymbol{\beta}})^{-1/2}\boldsymbol{R}(\boldsymbol{\psi})^{-1}\boldsymbol{A}(\hat{\boldsymbol{\beta}})^{-1/2}[\boldsymbol{\nabla}\{\boldsymbol{\mu}(\boldsymbol{\beta}^*)\}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)]\right| = o_p(1),$$
$$\sup_{\boldsymbol{\psi}\in\boldsymbol{\Psi}}\left|\frac{1}{n}[\boldsymbol{\nabla}\{\boldsymbol{\mu}(\boldsymbol{\beta}^*)\}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)]^T \boldsymbol{A}(\hat{\boldsymbol{\beta}})^{-1/2}\boldsymbol{R}(\boldsymbol{\psi})^{-1}\boldsymbol{A}(\hat{\boldsymbol{\beta}})^{-1/2}\{\boldsymbol{Y} - \boldsymbol{\mu}(\boldsymbol{\beta}_0)\}\right| = o_p(1),$$
$$\sup_{\boldsymbol{\psi}\in\boldsymbol{\Psi}}\left|\frac{1}{n}\{\boldsymbol{Y} - \boldsymbol{\mu}(\boldsymbol{\beta}_0)\}^T\left\{\boldsymbol{A}(\hat{\boldsymbol{\beta}})^{-1/2} - \boldsymbol{A}(\boldsymbol{\beta}_0)^{-1/2}\right\}\boldsymbol{R}(\boldsymbol{\psi})^{-1}\boldsymbol{A}(\hat{\boldsymbol{\beta}})^{-1/2}\{\boldsymbol{Y} - \boldsymbol{\mu}(\boldsymbol{\beta}_0)\}\right| = o_p(1),$$
$$\sup_{\boldsymbol{\psi}\in\boldsymbol{\Psi}}\left|\frac{1}{n}\{\boldsymbol{Y} - \boldsymbol{\mu}(\boldsymbol{\beta}_0)\}^T\left\{\boldsymbol{A}(\hat{\boldsymbol{\beta}})^{-1/2} - \boldsymbol{A}(\boldsymbol{\beta}_0)^{-1/2}\right\}\boldsymbol{R}(\boldsymbol{\psi})^{-1}\boldsymbol{A}(\boldsymbol{\beta}_0)^{-1/2}\{\boldsymbol{Y} - \boldsymbol{\mu}(\boldsymbol{\beta}_0)\}\right| = o_p(1). \tag{S2.3}$$

Thus, together with (S2.2), we have $\sup_{\boldsymbol{\psi}\in\boldsymbol{\Psi}}|I_{1n}(\boldsymbol{\psi})| = o_p(1)$.

Under Conditions (C1)-(C3), similar with the proof of Theorem 3.3 in Bachoc et al. (2018), we can obtain $\sup_{\boldsymbol{\psi}\in\boldsymbol{\Psi}}|I_{2n}(\boldsymbol{\psi})| = o_p(1)$. Directly from the definition of $\bar{\ell}(\boldsymbol{\psi}, \boldsymbol{\beta}_0)$, we have $\sup_{\boldsymbol{\psi}\in\boldsymbol{\Psi}}|I_{3n}(\boldsymbol{\psi})| = o_p(1)$. Through the above discussion, we have

$$\sup_{\boldsymbol{\psi}\in\boldsymbol{\Psi}}|\ell_n(\boldsymbol{\psi}, \hat{\boldsymbol{\beta}}) - \bar{\ell}(\boldsymbol{\psi}, \boldsymbol{\beta}_0)| = o_p(1). \tag{S2.4}$$

Combing (S2.4) with the identifiability condition 1 in the main paper, by Theorem 5.7 in Van der Vaart (2000), we have for every $\zeta > 0$, as $n \to \infty$, $P\left\{d(\hat{\boldsymbol{\psi}}, \boldsymbol{\Psi}_0) \geq \zeta\right\} \to 0$. □

Lemma 2. *Under the conditions of Lemma 1, assumption (C5) and the identifiability condition 2 in the main paper, suppose there exists an estimator of nuisance parameter, satisfying* $d(\hat{\boldsymbol{\psi}}, \boldsymbol{\Psi}_0) = o_p(1)$, *then for the* *estimator of regression parameter through solving equation (2.4) with $\boldsymbol{\psi}$ replaced by $\hat{\boldsymbol{\psi}}$, denoted as $\hat{\boldsymbol{\beta}}$, we have*

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \xrightarrow{D} N\left\{0, \Xi(\boldsymbol{\beta}_0, \boldsymbol{\psi}_0)\right\}.$$

Proof: We first prove the consistency of $\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 = o_p(1)$, then first-order Taylor expansion can be used to prove the asymptotic normality of $\hat{\boldsymbol{\beta}}$.

The proof of consistency is quite similar to the proof for Theorem 3 in Lin (2008). We only need to verify that the conditions for Theorem 3 of Lin (2008) are stilled satisfied under Conditions (C1)-(C5). The verifying process is quite straightforward, thus omitted here.

After obtaining the consistency of $\hat{\boldsymbol{\beta}}$, the proof for Theorem A.6 in the supplementary file of Oman et al. (2007) gives details for proving the asymptotic normality of the spatial GEE-type estimator $\hat{\boldsymbol{\beta}}$ using first-order Taylor expansion. As showed by Oman et al. (2007), the crucial step to bound the reminder terms in the Taylor expansion is to prove that, the inverse of estimated working correlation matrix converges in probability to the inverse of working correlation matrix with the unknown nuisance parameter replaced by the convergence value of the nuisance estimator. In our paper, it means to prove that

$$||\boldsymbol{R}(\hat{\boldsymbol{\psi}})^{-1} - \boldsymbol{R}(\boldsymbol{\psi}_0)^{-1}||_2 = o_p(1), \tag{S2.5}$$

where for a $n \times n$ matrix $\boldsymbol{A}$, the L_2-norm $||\boldsymbol{A}||_2 = \sup\{||\boldsymbol{A}\boldsymbol{a}||, ||a|| = 1\}$, and the L_2-norm of $\boldsymbol{A}$ is equal to the largest singular value of $\boldsymbol{A}$.

Under the identifiability condition 2 in the main paper, $\boldsymbol{\Psi}_0$ either contains a unique element, or the boundary, i.e., $\pi_k = 0$ for some $k \in \{1, \cdots, K\}$. For simplicity, we consider $\boldsymbol{\Psi}_0 = \{(\boldsymbol{\alpha}_k^T, \pi_k)_{k=1}^K : \pi_1 = 1, \boldsymbol{\alpha}_1 = \boldsymbol{\alpha}_1^*$, and $\pi_k = 0, \boldsymbol{\alpha}_k \in \boldsymbol{\Psi}_{\boldsymbol{\alpha}_k}$ for $k \geq 2\}$, the proofs for (S2.5) under other structures of $\boldsymbol{\Psi}_0$ are almost the same.

To prove (S2.5), noticing for $n \times n$ invertible matrices $\boldsymbol{A}$ and $\boldsymbol{B}$, we have $\boldsymbol{B}^{-1} - \boldsymbol{A}^{-1} = \boldsymbol{B}^{-1}(\boldsymbol{A} - \boldsymbol{B})\boldsymbol{A}^{-1}$, thus

$$\begin{aligned}
||\boldsymbol{R}(\hat{\boldsymbol{\psi}})^{-1} - \boldsymbol{R}(\boldsymbol{\psi}_0)^{-1}||_2 &= ||\boldsymbol{R}(\hat{\boldsymbol{\psi}})^{-1}\left\{\boldsymbol{R}(\boldsymbol{\psi}_0) - \boldsymbol{R}(\hat{\boldsymbol{\psi}})\right\}\boldsymbol{R}(\boldsymbol{\psi}_0)^{-1}||_2 \\
&\leq ||\boldsymbol{R}(\hat{\boldsymbol{\psi}})^{-1}||_2||\boldsymbol{R}(\boldsymbol{\psi}_0) - \boldsymbol{R}(\hat{\boldsymbol{\psi}})||_2||\boldsymbol{R}(\boldsymbol{\psi}_0)^{-1}||_2 \\
&\leq ||\boldsymbol{R}(\hat{\boldsymbol{\psi}})^{-1}||_2\Big\{|\hat{\pi}_1 - 1|\ ||\boldsymbol{R}^{(1)}(\hat{\boldsymbol{\alpha}}_1)||_2 + ||\boldsymbol{R}^{(1)}(\hat{\boldsymbol{\alpha}}_1) - \boldsymbol{R}^{(1)}(\boldsymbol{\alpha}_1^*)||_2 \\
&\quad + \sum_{k=2}^K |\hat{\pi}_k|\ ||\boldsymbol{R}^{(k)}(\hat{\boldsymbol{\alpha}}_k)||_2\Big\}||\boldsymbol{R}(\boldsymbol{\psi}_0)^{-1}||_2,
\end{aligned} \tag{S2.6}$$

Under Conditions (C1)-(C2), by Theorem 5 in Bachoc and Furrer (2016) and Lemma 6 in Furrer et al. (2016), we have $\inf_{n,\ \boldsymbol{\alpha}_k \in \boldsymbol{\Psi}_{\boldsymbol{\alpha}_k}} \lambda_{\min}\{\boldsymbol{R}^{(k)}(\boldsymbol{\alpha}_k)\} > 0$, $\sup_{n,\ \boldsymbol{\alpha}_k \in \boldsymbol{\Psi}_{\boldsymbol{\alpha}_k}} \lambda_{\max}\{\boldsymbol{R}^{(k)}(\boldsymbol{\alpha}_k)\} < \infty$. According to $d(\hat{\boldsymbol{\psi}}, \boldsymbol{\Psi}_0) = o_p(1)$, we have $\hat{\pi}_1 - 1 = o_p(1)$, $\hat{\boldsymbol{\alpha}}_1 - \boldsymbol{\alpha}_1^* = o_p(1)$, $\hat{\pi}_k = o_p(1)$ for $k \geq 2$. Thus, we can obtain $||\boldsymbol{R}(\hat{\boldsymbol{\psi}})^{-1}||_2 = O(1)$, $||\boldsymbol{R}(\boldsymbol{\psi}_0)^{-1}||_2 = O(1)$, $|\hat{\pi}_1 - 1|\ ||\boldsymbol{R}^{(1)}(\hat{\boldsymbol{\alpha}}_1)||_2 = o_p(1)$,

$\sum_{k=2}^{K} |\hat{\pi}_k| \; ||\boldsymbol{R}^{(k)}(\hat{\boldsymbol{\alpha}}_k)||_2 = o_p(1)$. Then, by first-order Taylor expansion, assumption (C2) and Lemma 6 in Furrer et al. (2016), we have $||\boldsymbol{R}^{(1)}(\hat{\boldsymbol{\alpha}}_1) - \boldsymbol{R}^{(1)}(\boldsymbol{\alpha}_1^*)||_2 = o_p(1)$. Combing above results, (S2.5) is proven. Then, following the same steps as the proof for Theorem A.6 in the supplementary file of Oman et al. (2007), we can obtain $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{D} N\left\{0, \Xi(\boldsymbol{\beta}_0, \boldsymbol{\psi}_0)\right\}$. $\square$

Proof of Theorem 1: For ease of illustration, let $p_k = 1$, $k = 1, \cdots, K$, and the proof for $p_k > 1$ is similar.

Taking the initial values $\pi_k^{(0)} > 0$ and $\alpha_k^{(0)} > 0$, $k = 1, \cdots, K$. Then, the logarithmic function in (2.6) of the main paper will force each update $\boldsymbol{\psi}^{(t)}$ to remain within the interior of the feasible region.

Now we prove $\ell_n(\boldsymbol{\psi}^{(t+1)}, \boldsymbol{\beta}) < \ell_n(\boldsymbol{\psi}^{(t)}, \boldsymbol{\beta})$. By (2.7) of the main paper, we have

$$\ell_n(\boldsymbol{\psi}^{(t+1)}, \boldsymbol{\beta}) = \mathcal{S}(\boldsymbol{\psi}^{(t+1)}|\boldsymbol{\psi}^{(t)}) + \mathcal{F}(\boldsymbol{\psi}^{(t+1)}|\boldsymbol{\psi}^{(t)}).$$

By the definition of $\boldsymbol{\psi}^{(t+1)}$, we have

$$\mathcal{S}(\boldsymbol{\psi}^{(t+1)}|\boldsymbol{\psi}^{(t)}) \leq \mathcal{S}(\boldsymbol{\psi}^{(t)}|\boldsymbol{\psi}^{(t)}), \tag{S2.7}$$

thus we only need to prove $\mathcal{F}(\boldsymbol{\psi}^{(t+1)}|\boldsymbol{\psi}^{(t)}) < \mathcal{F}(\boldsymbol{\psi}^{(t)}|\boldsymbol{\psi}^{(t)})$. Calculating the first-order and the second-order derivatives of $\mathcal{F}(\boldsymbol{\psi}|\boldsymbol{\psi}^{(t)})$ w.r.t. $(\pi_1, \cdots, \pi_{K-1}, \alpha_1, \cdots, \alpha_K)$, we have

$$\begin{aligned}
\frac{\partial \mathcal{F}(\boldsymbol{\psi}|\boldsymbol{\psi}^{(t)})}{\partial \pi_k} &= -\frac{1-\sum_{k=1}^{K-1}\pi_k^{(t)}}{1-\sum_{k=1}^{K-1}\pi_k} + \frac{\pi_k^{(t)}}{\pi_k}, \\
\frac{\partial \mathcal{F}(\boldsymbol{\psi}|\boldsymbol{\psi}^{(t)})}{\partial \alpha_k} &= \frac{\alpha_k^{(t)}}{\alpha_k} - 1, \\
\frac{\partial^2 \mathcal{F}(\boldsymbol{\psi}|\boldsymbol{\psi}^{(t)})}{\partial \pi_{k_1}\partial \pi_{k_2}} &= -\frac{1-\sum_{k=1}^{K-1}\pi_k^{(t)}}{1-\sum_{k=1}^{K-1}\pi_k}(\text{if } k_1 \neq k_2), \text{ or} \\
&= -\frac{1-\sum_{k=1}^{K-1}\pi_k^{(t)}}{1-\sum_{k=1}^{K-1}\pi_k} - \frac{\pi_{k_1}^{(t)}}{\pi_{k_1}^2}(\text{if } k_1 = k_2), \\
\frac{\partial^2 \mathcal{F}(\boldsymbol{\psi}|\boldsymbol{\psi}^{(t)})}{\partial \pi_{k_1}\partial \alpha_{k_2}} &= 0, \\
\frac{\partial^2 \mathcal{F}(\boldsymbol{\psi}|\boldsymbol{\psi}^{(t)})}{\partial \alpha_{k_1}\partial \alpha_{k_2}} &= 0(\text{if } k_1 \neq k_2), \text{ or } -\frac{\alpha_{k_1}^{(t)}}{\alpha_{k_1}^2}(\text{if } k_1 = k_2).
\end{aligned} \tag{S2.8}$$

From (S2.8), it is easy to show the Hessian matrix of $\mathcal{F}(\boldsymbol{\psi}|\boldsymbol{\psi}^{(t)})$ is negative definite for all values in the feasible region, which means $\mathcal{F}(\boldsymbol{\psi}|\boldsymbol{\psi}^{(t)})$ is strictly concave. Together with the first-order derivatives in (S2.8), $\mathcal{F}(\boldsymbol{\psi}|\boldsymbol{\psi}^{(t)})$ attains its unique maximum at $\boldsymbol{\psi}^{(t)}$. Through the above discussions, if $\boldsymbol{\psi}^{(t+1)} \neq \boldsymbol{\psi}^{(t)}$, we have $\mathcal{F}(\boldsymbol{\psi}^{(t+1)}|\boldsymbol{\psi}^{(t)}) < \mathcal{F}(\boldsymbol{\psi}^{(t)}|\boldsymbol{\psi}^{(t)})$. Proof is thus completed. $\square$

Proof of Theorem 2-3: Notice the initial estimator $\hat{\boldsymbol{\beta}}^0$ in the iterative algorithm is $\sqrt{n}-$consistent (Lin, 2008), then by Lemma 1 and Lemma 2, Theorem 2-3 can be directly obtained by deductive reasoning.

Now we prove that when the true correlation matrix of $\boldsymbol{Y}$ is $\sum_{k=1}^{K}\pi_k^0\boldsymbol{R}^{(k)}(\boldsymbol{\alpha}_k^0)$, if $\pi_k^0 > 0$ for all $k \in \{1, \cdots, K\}$, we have $\boldsymbol{\Psi}_0 = \{(\boldsymbol{\alpha}_k^{0T}, \pi_k^0)_{k=1}^K\}$, otherwise

$\boldsymbol{\Psi}_0 = \{(\boldsymbol{\alpha}_k{}^T, \pi_k^0)_{k=1}^K : \boldsymbol{\alpha}_k = \boldsymbol{\alpha}_k^0 \ for \ \pi_k^0 > 0, \ and \ \boldsymbol{\alpha}_k \in \boldsymbol{\Psi}_{\boldsymbol{\alpha}_k} \ for \ \pi_k^0 = 0\}$. First, notice $\mathrm{E}\{\ell_n(\boldsymbol{\psi}, \boldsymbol{\beta}_0)\} = \mathrm{E}_N\{\ell_n(\boldsymbol{\psi}, \boldsymbol{\beta}_0)\}$, where $\mathrm{E}_N\{\ell_n(\boldsymbol{\psi}, \boldsymbol{\beta}_0)\}$ represents that the expectation is taken pretending $\epsilon(\boldsymbol{\beta_0})$ is a standard Gaussian distribution with correlation matrix $\mathrm{E}\{\epsilon(\boldsymbol{\beta_0})\epsilon(\boldsymbol{\beta_0})^T\}$. Thus, when the true correlation matrix of $\boldsymbol{Y}$ is $\sum_{k=1}^K \pi_k^0 \boldsymbol{R}^{(k)}(\boldsymbol{\alpha}_k^0)$, i.e., $\mathrm{E}\{\epsilon(\boldsymbol{\beta_0})\epsilon(\boldsymbol{\beta_0})^T\} = \sum_{k=1}^K \pi_k^0 \boldsymbol{R}^{(k)}(\boldsymbol{\alpha}_k^0)$, according to Lemma 5.35 in Van der Vaart (2000), $\{(\boldsymbol{\alpha}_k^{0T}, \pi_k^0)_{k=1}^K\}$ is the minimum point of $\mathrm{E}_N\{\ell_n(\boldsymbol{\psi}, \boldsymbol{\beta}_0)\}$, so that it is also the minimum point of $\mathrm{E}\{\ell_n(\boldsymbol{\psi}, \boldsymbol{\beta}_0)\}$. Because $\mathrm{E}\{\ell_n(\boldsymbol{\psi}, \boldsymbol{\beta}_0)\}$ converges uniformly to $\bar{\ell}(\boldsymbol{\psi}, \boldsymbol{\beta}_0)$ over $\boldsymbol{\Psi}$, $\{(\boldsymbol{\alpha}_k^{0T}, \pi_k^0)_{k=1}^K\}$ is the minimum point of $\bar{\ell}(\boldsymbol{\psi}, \boldsymbol{\beta}_0)$. Then, by the identifiability condition 2 and the definition of $\boldsymbol{\Psi}_0$, we have, if $\pi_k^0 > 0$ for all $k \in \{1, \cdots, K\}$, we have $\boldsymbol{\Psi}_0 = \{(\boldsymbol{\alpha}_k^{0T}, \pi_k^0)_{k=1}^K\}$, otherwise $\boldsymbol{\Psi}_0 = \{(\boldsymbol{\alpha}_k{}^T, \pi_k^0)_{k=1}^K : \boldsymbol{\alpha}_k = \boldsymbol{\alpha}_k^0 \ for \ \pi_k^0 > 0, \ and \ \boldsymbol{\alpha}_k \in \boldsymbol{\Psi}_{\boldsymbol{\alpha}_k} \ for \ \pi_k^0 = 0\}$. Proof is completed. □

Proof of Theorem 4: The proof of Theorem 4 is similar to the proof of Theorem 2-3. We only need to obtain similar results with Lemma 1 and Lemma 2, then Theorem 4 is proven. We first present a useful fact for Schur product:

(F1) If $\boldsymbol{A}$ and $\boldsymbol{B}$ are $n \times n$ positive semi-definite matrices, then any eigenvalue $\lambda(\boldsymbol{A} \circ \boldsymbol{B})$ of $\boldsymbol{A} \circ \boldsymbol{B}$ satisfies $\min_{1 \le i \le n}(b_{ii})\lambda_{\min}(\boldsymbol{A}) \le \lambda(\boldsymbol{A} \circ \boldsymbol{B}) \le$

$\max_{1\le i\le n}(b_{ii})\lambda_{\max}(\boldsymbol{A})$, where b_{ii} is the diagonal element of $\boldsymbol{B}$.

For proving similar results of Lemma 1, we only need to show that (S2.3) still holds with $\boldsymbol{R}(\boldsymbol{\psi})^{-1}$ replaced by $\{\boldsymbol{R}(\boldsymbol{\psi})\circ\boldsymbol{T}(\gamma_2)\}^{-1}\circ\boldsymbol{T}(\gamma_2)$, and other steps for proof of Lemma 1 are similar. To show (S2.3) holds with $\{\boldsymbol{R}(\boldsymbol{\psi})\circ\boldsymbol{T}(\gamma_2)\}^{-1}\circ\boldsymbol{T}(\gamma_2)$, it is sufficient to prove that

$$\inf_{\boldsymbol{\psi}\in\boldsymbol{\Psi}}{}_n\,\lambda_{\min}[\{\boldsymbol{R}(\boldsymbol{\psi})\circ\boldsymbol{T}(\gamma_2)\}^{-1}\circ\boldsymbol{T}(\gamma_2)]>0,\quad \sup_{\boldsymbol{\psi}\in\boldsymbol{\Psi}}{}_n\,\lambda_{\max}[\{\boldsymbol{R}(\boldsymbol{\psi})\circ\boldsymbol{T}(\gamma_2)\}^{-1}\circ\boldsymbol{T}(\gamma_2)]<\infty. \tag{S2.9}$$

Noticing $\boldsymbol{T}(\gamma_2)$ is positive definite and its diagonal elements equal one, (S2.9) is immediately obtained by using (F1).

For proving similar results of Lemma 2, we first show (S2.5) holds with $\boldsymbol{R}(\boldsymbol{\psi})^{-1}$ replaced by $\{\boldsymbol{R}(\boldsymbol{\psi})\circ\boldsymbol{T}(\gamma_{1n})\}^{-1}$, and it can be directly obtained by using Lemma 9 in Furrer et al. (2016), which indicates

$$\sup_{\boldsymbol{\psi}\in\boldsymbol{\Psi}}{}_n\,||\{\boldsymbol{R}(\boldsymbol{\psi})\circ\boldsymbol{T}(\gamma_{1n})\}^{-1}-\boldsymbol{R}(\boldsymbol{\psi})^{-1}||_2=o(1). \tag{S2.10}$$

Then, by by similar steps in Lemma 2, we can obtain

$$\sqrt{n}\;\Xi(\boldsymbol{\beta}_0,\boldsymbol{\psi}_0;\gamma_{1n})^{-1}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})\xrightarrow{D}N\left\{0,\boldsymbol{I}\right\}.$$

where $\Xi(\boldsymbol{\beta}_0,\boldsymbol{\psi}_0;\gamma_{1n})=\boldsymbol{\Pi}(\boldsymbol{\beta}_{\boldsymbol{0}},\boldsymbol{\psi}_0;\gamma_{1n})\boldsymbol{\Sigma}_0\boldsymbol{\Pi}(\boldsymbol{\beta}_{\boldsymbol{0}},\boldsymbol{\psi}_0;\gamma_{1n})^T$, and

$$\begin{aligned}\boldsymbol{\Pi}(\boldsymbol{\beta},\boldsymbol{\psi};\gamma_{1n})=&\sqrt{n}\left[\boldsymbol{D}(\boldsymbol{\beta})^T\boldsymbol{A}(\boldsymbol{\beta})^{-1/2}\{\boldsymbol{R}(\boldsymbol{\psi})\circ\boldsymbol{T}(\gamma_{1n})\}^{-1}\boldsymbol{A}(\boldsymbol{\beta})^{-1/2}\boldsymbol{D}(\boldsymbol{\beta})\right]^{-1}\\&\boldsymbol{D}(\boldsymbol{\beta})^T\boldsymbol{A}(\boldsymbol{\beta})^{-1/2}\{\boldsymbol{R}(\boldsymbol{\psi})\circ\boldsymbol{T}(\gamma_{1n})\}^{-1}\boldsymbol{A}(\boldsymbol{\beta})^{-1/2}.\end{aligned}$$

Again by (S2.10), similar with (S2.6), we have $||\Xi(\boldsymbol{\beta}_0, \boldsymbol{\psi}_0; \gamma_{1n})^{-1} - \Xi(\boldsymbol{\beta}_0, \boldsymbol{\psi}_0)^{-1}||_2 = o(1)$. Thus, we can obtain

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{D} N\left\{0, \Xi(\boldsymbol{\beta}_0, \boldsymbol{\psi}_0)\right\}. \tag{S2.11}$$

Now we prove that when the true correlation matrix of $\boldsymbol{Y}$ is $\sum_{k=1}^{K} \pi_k^0 \boldsymbol{R}^{(k)}(\boldsymbol{\alpha}_k^0)$, we have $\widetilde{\boldsymbol{\Psi}}_0 = \boldsymbol{\Psi}_0$. By assumption (C6), when the true correlation matrix is $\sum_{k=1}^{K} \pi_k^0 \boldsymbol{R}^{(k)}(\boldsymbol{\alpha}_k^0)$, $\mathrm{E}\{\ell_n(\boldsymbol{\psi}, \boldsymbol{\beta}_0; \gamma_2)\}$ attains its minimum at $(\boldsymbol{\alpha}_k^{0T}, \pi_k^0)_{k=1}^K$. Because $E\{\ell_n(\boldsymbol{\psi}, \boldsymbol{\beta}_0; \gamma_2)\}$ converges uniformly to $\bar{\ell}(\boldsymbol{\psi}, \boldsymbol{\beta}_0; \gamma_2)$ over $\boldsymbol{\Psi}$, $(\boldsymbol{\alpha}_k^{0T}, \pi_k^0)_{k=1}^K$ is the minimum point of $\bar{\ell}(\boldsymbol{\psi}, \boldsymbol{\beta}_0; \gamma_2)$. Then, by the identifiability condition 2 and the definition of $\widetilde{\boldsymbol{\Psi}}_0$, if $\pi_k^0 > 0$ for all $k \in \{1, \cdots, K\}$, we have $\widetilde{\boldsymbol{\Psi}}_0 = \{(\boldsymbol{\alpha}_k^{0T}, \pi_k^0)_{k=1}^K\}$, otherwise $\widetilde{\boldsymbol{\Psi}}_0 = \{(\boldsymbol{\alpha}_k{}^T, \pi_k^0)_{k=1}^K : \boldsymbol{\alpha}_k = \boldsymbol{\alpha}_k^0 \ for \ \pi_k^0 > 0, \ and \ \boldsymbol{\alpha}_k \in \boldsymbol{\Psi}_{\boldsymbol{\alpha}_k} \ for \ \pi_k^0 = 0\}$. Proof is completed. $\square$

Bibliography

Bachoc, F. et al. (2018). Asymptotic analysis of covariance parameter estimation for gaussian processes in the misspecified case. *Bernoulli 24*(2), 1531–1575.

Bachoc, F. and R. Furrer (2016). On the smallest eigenvalues of covariance matrices of multivariate spatial processes. *Stat 5*(1), 102–107.

Furrer, R., F. Bachoc, and J. Du (2016). Asymptotic properties of multivariate tapering for estimation and prediction. *Journal of Multivariate Analysis 149*, 177–191.

Lin, P.-S. (2008). Estimating equations for spatially correlated data in multi-dimensional space. *Biometrika 95*(4), 847–858.

Oman, S. D., V. Landsman, Y. Carmel, and R. Kadmon (2007). Analyzing spatially distributed binary data using independent-block estimating equations. *Biometrics 63*(3), 892–900.

Van der Vaart, A. W. (2000). *Asymptotic statistics*, Volume 3. Cambridge university press.