Joint Mean-Angle Model for Spatial Binary Data

Cheng Peng[1], Renwen Luo[2], Yang Han[1], and Jianxin Pan[3,2]

[1]University of Manchester, UK

[2]Guangdong Provincial Key Laboratory of Interdisciplinary Research and Application for Data Science, BNU-HKBU United International College, China

[3]Research Center for Mathematics, Beijing Normal University, China

Supplementary Material

The supplementary materials include all technical proofs and further simulation studies, which are not shown in the article due to the limitation of its length.

S1 Technical proofs

S1.1 The upper and lower bounds of covariance element σ_{ij}

The following inequalities can be easily derived from the properties of probability,

$$\Pr(Y_i = 1) + \Pr(Y_j = 1) - 1 < \Pr(Y_i = 1, Y_j = 1) < \min\{\Pr(Y_i = 1), \Pr(Y_j = 1)\}.$$

Because $\sigma_{ij} = \mathrm{E}(Y_iY_j) - \mathrm{E}(Y_i)\mathrm{E}(Y_j) = \Pr(Y_i = 1, Y_j = 1) - \mu_i\mu_j$, $\Pr(Y_i = 1) = \mu_i$ and $\Pr(Y_j = 1) = \mu_j$, we rewrite the inequalities as follows,

$$\max\{0, \mu_i + \mu_j - 1\} < \Pr(Y_i = 1, Y_j = 1) < \min\{\mu_i, \mu_j\},$$

and then we have

$$\max\{-\mu_i\mu_j, -1\} \leq \max\{0, \mu_i + \mu_j - 1\} - \mu_i\mu_j < \sigma_{ij} < \min\{\mu_i, \mu_j\} - \mu_i\mu_j.$$

Since $\sigma_{ii} = \mu_i(1 - \mu_i)$ and $\sigma_{jj} = \mu_j(1 - \mu_j)$, the correlation coefficient $\varrho_{ij} = \sigma_{ij}/(\sigma_{ii}\sigma_{jj})^{1/2}$ satisfies

$$\max\left\{-\left[\mu_i\mu_j/((1-\mu_i)(1-\mu_j))\right]^{1/2}, -\left[(1-\mu_i)(1-\mu_j)/(\mu_i\mu_j)\right]^{1/2}\right\} < \varrho_{ij}$$
$$< \min\left\{\left[\mu_i(1-\mu_j)/(\mu_j(1-\mu_i))\right]^{1/2}, \left[\mu_j(1-\mu_i)/(\mu_i(1-\mu_j))\right]^{1/2}\right\},$$

S1.2 HPC method is order-dependent

Demonstration: suppose we have a p-dimensional correlation matrix $\mathbf{R} = (R_{jk})_{1\leq j,k\leq p}$, where $p \geq 4$ and the elements of the corresponding angle matrix $\omega_{31} \in (0, \pi/2)$, $\omega_{21} \in (0, \pi/2)$, $\omega_{32} \in (\pi/2, \pi)$, $\omega_{j2} \in (\pi/4, \pi/2)$ and $\omega_{j3} \in (\pi/4, \pi/2)$ for some $j \geq 4$. Here we exchange the labels j and 2 in the data. Denote $\Omega(\mathbf{R}) = \{\omega_{mk} | 1 \leq k < m \leq p\}$ by the set of all angles involved in the HPC decomposition of $\mathbf{R}$. Furthermore, for fixed l, $\Omega_l(\mathbf{R}) = \{\omega_{ml} | m > l\}$.

We first consider the set $\Omega_1(\mathbf{R})$ under exchange between labels j and

2. We denote the original correlation matrix and correlation matrix after change by $\mathbf{R}$ and $\mathbf{R}'$, respectively. Note that we have

$$\cos(\omega'_{21}) = R'_{21} = R_{j1} = \cos(\omega_{j1}),$$

$$\cos(\omega'_{j1}) = R'_{j1} = R_{21} = \cos(\omega_{21}),$$

$$\cos(\omega'_{k1}) = R'_{k1} = R_{k1} = \cos(\omega_{k1}),$$

for $j > 2$, $k > 2$ and $k \neq j$. Because cosine function is monotonic in $(0, \pi)$, we conclude that $\Omega_1(\mathbf{R}) = \Omega_1(\mathbf{R}')$. That is, $\Omega_1(\mathbf{R})$ is invariant under exchange between labels j and 2.

We next consider $\Omega_2(\mathbf{R})$. Regarding R_{j2} and R'_{j2}, it is easy to see that $R'_{j2} = R_{2j} = R_{j2}$, since $\mathbf{R}$ is a symmetric matrix. And by the equation (6) in Zhang et al. (2015), we have

$$R_{j2} = \sin(\omega_{j1})\sin(\omega_{21})[\cos(\omega_{j2}) + \cos(\omega_{j1})\cos(\omega_{21})],$$

$$R'_{j2} = \sin(\omega'_{j1})\sin(\omega'_{21})[\cos(\omega'_{j2}) + \cos(\omega'_{j1})\cos(\omega'_{21})].$$

Thus we have $\cos(\omega_{j2}) = \cos(\omega'_{j2})$. Since the angle $\omega \in (0, \pi)$ and cosine function is monotonic in $(0, \pi)$, then we can obtain that

$$\omega_{j2} = \omega'_{j2}.$$

In terms of the correlation coefficient R_{32}, we have $R_{32} = R'_{j3}$ under exchange between labels j and 2. In addition,

$$R_{32} = \sin(\omega_{31})\sin(\omega_{21})[\cos(\omega_{32}) + \cos(\omega_{31})\cos(\omega_{21})],$$
$$R'_{j3} = \cos(\omega'_{j3})\prod_{l=1}^{2}\sin(\omega'_{jl})\sin(\omega'_{3l}) + \sum_{l=1}^{2}\left[\cos(\omega'_{jl})\cos(\omega'_{3l})\prod_{t=1}^{l-1}\sin(\omega'_{jt})\sin(\omega'_{3t})\right],$$

where $\prod_{l=1}^{0} = 1$.

If HPC decomposition is order independent, we then have

$$\omega'_{j3} = \omega_{32}, \quad \omega'_{32} = \omega_{j3}.$$

Putting all the above results together, after simplification, we can obtain that

$$\begin{aligned} &\sin(\omega_{31})\sin(\omega_{21})[\cos(\omega_{j3})\cos(\omega_{j2}) + \cos(\omega_{32})\sin(\omega_{j2})\sin(\omega_{j3}) - \cos(\omega_{32})] \\ &= \cos(\omega_{31})\cos(\omega_{21})[\sin(\omega_{31})\sin(\omega_{21}) - 1]. \end{aligned} \tag{8}$$

Since we assume that all the angles are in $(0, \pi)$, we have

$$\mathrm{sgn}[\sin(\omega_{31})\sin(\omega_{21}) - 1] = -1, \quad \mathrm{sgn}[\sin(\omega_{31})\sin(\omega_{21})] = 1,$$

Note

$$\cos(\omega_{j3})\cos(\omega_{j2}) + \cos(\omega_{32})\sin(\omega_{j2})\sin(\omega_{j3}) - \cos(\omega_{32}) > 0,$$

if and only if

$$\tan(\omega_{j2})\tan(\omega_{j3}) > \frac{1}{1 - \cos(\omega_{32})},$$

which is implied by $\omega_{32} \in (\pi/2, \pi)$, $\omega_{j2} \in (\pi/4, \pi/2)$ and $\omega_{j3} \in (\pi/4, \pi/2)$. Since $\omega_{31} \in (0, \pi/2)$ and $\omega_{21} \in (0, \pi/2)$, we can conclude that

$$\sin(\omega_{31})\sin(\omega_{21})[\cos(\omega_{j3})\cos(\omega_{j2}) + \cos(\omega_{32})\sin(\omega_{j2})\sin(\omega_{j3}) - \cos(\omega_{32})] > 0,$$

$$\cos(\omega_{31})\cos(\omega_{21})[\sin(\omega_{31})\sin(\omega_{21}) - 1] < 0.$$

This is a contradiction with the equation (8). Therefore, the HPC method is order-dependent.

S1.3 Proof of Theorem 1

We first show that $\hat{\beta}$ is a $\sqrt{n}$-consistent estimator for parameter β. According to McCullagh (1983), we have

$$\hat{\beta} - \beta = \left\{ \frac{1}{n} \left(\frac{\partial \boldsymbol{\mu}^\top}{\partial \beta} \right) \boldsymbol{\Sigma}^{-1} \left(\frac{\partial \boldsymbol{\mu}^\top}{\partial \beta} \right)^\top \right\}^{-1} \left\{ \frac{1}{n} \left(\frac{\partial \boldsymbol{\mu}^\top}{\partial \beta} \right) \boldsymbol{\Sigma}^{-1} (\boldsymbol{y} - \boldsymbol{\mu}) \right\} + o_p(n^{-1/2}).$$

The expectation of $S_1(\beta) = (\partial \boldsymbol{\mu}^\top / \partial \beta) \boldsymbol{\Sigma}^{-1} (\boldsymbol{y} - \boldsymbol{\mu})$ at the true value β is $\mathrm{E}(S_1(\beta)) = 0$ and the matrix

$$\mathrm{E}\left(\frac{\partial S_1(\beta)}{\partial \beta} \right) = \left\{ \left(\frac{\partial \boldsymbol{\mu}^\top}{\partial \beta} \right) \boldsymbol{\Sigma}^{-1} \left(\frac{\partial \boldsymbol{\mu}^\top}{\partial \beta} \right)^\top \right\} = (v(\boldsymbol{X}\beta)\boldsymbol{X})^\top \boldsymbol{\Sigma}^{-1} (v(\boldsymbol{X}\beta)\boldsymbol{X}),$$

where

$$v(\boldsymbol{X}\beta) = \mathrm{diag}\left(\frac{\exp(X_1^\top \beta)}{(1 + \exp(X_1^\top \beta))^2}, \ldots, \frac{\exp(X_n^\top \beta)}{(1 + \exp(X_n^\top \beta))^2} \right)$$

is an $n \times n$ diagonal matrix. Since by Cholesky decomposition, $\boldsymbol{\Sigma}^{-1} = T^\top D^{-1} T$, where $D^{-1} = \mathrm{diag}(\sigma_{11}^{-1}, \ldots, \sigma_{nn}^{-1})$, the matrix $\mathrm{E}(\partial S_1(\beta)/\partial \beta) = (Tv(\boldsymbol{X}\beta)\boldsymbol{X})^\top D^{-1} (Tv(\boldsymbol{X}\beta)\boldsymbol{X})$. By condition (A3), there exists a constant

λ_β such that $\mathrm{E}(\partial S_1(\beta)/\partial\beta) \le \lambda_\beta \mathbf{1}_{p_\beta\times p_\beta}$, where $\mathbf{1}_{p_\beta\times p_\beta}$ is the p_β-dimensional matrix whose elements are all 1. Since $S_1(\beta)$ involves a sum of n zero-mean random variables, by the law of large numbers,

$$\frac{1}{n}S_1(\beta) = \left\{\frac{1}{n}\left(\frac{\partial\boldsymbol{\mu}^\top}{\partial\beta}\right)\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}-\boldsymbol{\mu})\right\} \to 0$$

with probability tending to 1 as $n\to\infty$ and $(1/n)S_1(\beta) = O_p(n^{-1/2})$. Similarly, we can prove that

$$\frac{1}{n}\mathrm{E}\left(\frac{\partial S_1(\beta)}{\partial\beta}\right) = \left\{\frac{1}{n}\left(\frac{\partial\boldsymbol{\mu}^\top}{\partial\beta}\right)\boldsymbol{\Sigma}^{-1}\left(\frac{\partial\boldsymbol{\mu}^\top}{\partial\beta}\right)^\top\right\}$$

is a bounded matrix and also $(1/n)\mathrm{E}(\partial S_1(\beta)/\partial\beta) = O(1)$. Finally, combining all these results, we have $\hat\beta\to\beta$ as $n\to\infty$ and

$$||\hat\beta-\beta||_2 = O_p(n^{-1/2}). \quad \square$$

And then, the consistency of parameter estimator $\hat\gamma$ can be obtained similarly. According to the generalized estimating equation $S_2(\gamma)$, we have

$$\begin{aligned}\hat\gamma-\gamma &= \left\{\frac{1}{N}\left(\frac{\partial\boldsymbol{\eta}^\top}{\partial\gamma}\right)\mathbf{M}^{-1}\left(\frac{\partial\boldsymbol{\eta}^\top}{\partial\gamma}\right)^\top\right\}^{-1}\left\{\frac{1}{N}\left(\frac{\partial\boldsymbol{\eta}^\top}{\partial\gamma}\right)\mathbf{M}^{-1}\mathrm{vech}^*(\hat{\mathbf{r}}\hat{\mathbf{r}}^\top - F(\mathbf{R}(\gamma);\mathbf{c}(\hat\beta)))\right\} + o_p(N^{-1/2})\\ &= D_1^{-1}B_1 + o_p(N^{-1/2}),\end{aligned}$$

where

$$\begin{aligned}D_1 &= \frac{1}{N}\left(\frac{\partial\boldsymbol{\eta}^\top}{\partial\gamma}\right)\mathbf{M}^{-1}\left(\frac{\partial\boldsymbol{\eta}^\top}{\partial\gamma}\right)^\top,\\ B_1 &= \frac{1}{N}\left(\frac{\partial\boldsymbol{\eta}^\top}{\partial\gamma}\right)\mathbf{M}^{-1}\mathrm{vech}^*(\hat{\mathbf{r}}\hat{\mathbf{r}}^\top - F(\mathbf{R}(\gamma);\mathbf{c}(\hat\beta)))\end{aligned}$$

Next, we decompose B_1 as follows,

$$B_1 = \frac{1}{N}\left(\frac{\partial \boldsymbol{\eta}^\top}{\partial \gamma}\right)\mathbf{M}^{-1}\mathrm{vech}^*(\hat{\mathbf{r}}\hat{\mathbf{r}}^\top - F(\mathbf{R}(\gamma);\mathbf{c}(\hat{\beta}))) = J_1 + J_2 + J_3,$$

where $\hat{\mathbf{r}} = \boldsymbol{y} - \hat{\boldsymbol{\mu}}$,

$$\begin{aligned}
J_1 &= \frac{1}{N}\left(\frac{\partial \boldsymbol{\eta}^\top}{\partial \gamma}\right)\mathbf{M}^{-1}\mathrm{vech}^*(\hat{\mathbf{r}}\hat{\mathbf{r}}^\top - \mathbf{r}\mathbf{r}^\top),\\
J_2 &= \frac{1}{N}\left(\frac{\partial \boldsymbol{\eta}^\top}{\partial \gamma}\right)\mathbf{M}^{-1}\mathrm{vech}^*(\mathbf{r}\mathbf{r}^\top - F(\mathbf{R}(\gamma);\mathbf{c}(\beta))),\\
J_3 &= \frac{1}{N}\left(\frac{\partial \boldsymbol{\eta}^\top}{\partial \gamma}\right)\mathbf{M}^{-1}\mathrm{vech}^*(F(\mathbf{R}(\gamma);\mathbf{c}(\beta)) - F(\mathbf{R}(\gamma);\mathbf{c}(\hat{\beta}))).
\end{aligned}$$

Similar to the proof of $\hat{\beta}$, it is easy to derive that $J_2 = O_p(N^{-1/2})$ since $\mathrm{E}(\mathbf{r}\mathbf{r}^\top) = F(\mathbf{R}(\gamma);\mathbf{c}(\beta))$. Then it suffices to show that

$$\begin{aligned}
J_1 &= \frac{1}{N}\left(\frac{\partial \boldsymbol{\eta}^\top}{\partial \gamma}\right)\mathbf{M}^{-1}\mathrm{vech}^*(\hat{\mathbf{r}}\hat{\mathbf{r}}^\top - \mathbf{r}\mathbf{r}^\top)\\
&= \frac{1}{N}\left(\frac{\partial \boldsymbol{\eta}^\top}{\partial \gamma}\right)\mathbf{M}^{-1}\mathrm{vech}^*(\hat{\boldsymbol{\mu}}\hat{\boldsymbol{\mu}}^\top - \boldsymbol{\mu}\boldsymbol{\mu} - \boldsymbol{y}(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})^\top - (\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})\boldsymbol{y}^\top)\\
&= \frac{1}{N}\left(\frac{\partial \boldsymbol{\eta}^\top}{\partial \gamma}\right)\mathbf{M}^{-1}\mathrm{vech}^*[(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})(\hat{\boldsymbol{\mu}} - \boldsymbol{y})^\top + (\boldsymbol{\mu} - \boldsymbol{y})(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})^\top]\\
&= \frac{1}{N}\left(\frac{\partial \boldsymbol{\eta}^\top}{\partial \gamma}\right)\mathbf{M}^{-1}\mathrm{vech}^*\left\{\left[(\hat{\beta} - \beta)^\top \frac{\partial \boldsymbol{\mu}^\top}{\partial \beta}\right](\hat{\boldsymbol{\mu}} - \boldsymbol{y})^\top + (\boldsymbol{\mu} - \boldsymbol{y})\left(\frac{\partial \boldsymbol{\mu}^\top}{\partial \beta}\right)^\top(\hat{\beta} - \beta)\right\}
\end{aligned}$$

Therefore, $J_1 = O_p(n^{-1/2})$. And also, following Taylor's expansion,

$$\begin{aligned}
J_3 &= \frac{1}{N}\left(\frac{\partial \boldsymbol{\eta}^\top}{\partial \gamma}\right)\mathbf{M}^{-1}\mathrm{vech}^*(F(\mathbf{R}(\gamma);\mathbf{c}(\beta)) - F(\mathbf{R}(\gamma);\mathbf{c}(\hat{\beta})))\\
&= \frac{1}{N}\left(\frac{\partial \boldsymbol{\eta}^\top}{\partial \gamma}\right)\mathbf{M}^{-1}\left(\frac{\partial \boldsymbol{\eta}^\top}{\partial \beta}\right)^\top(\hat{\beta} - \beta) = O_p(n^{-1/2}).
\end{aligned}$$

Since all the elements of matrix $\mathbf{M}^{-1}$ are bounded, we can obtain that $D_1 = O(1)$. Putting all the results together, we have

$$\|\hat{\gamma} - \gamma\|_2 = O_p(n^{-1/2}). \quad \square$$

S1.4 Proof of Theorem 2

Next, we give proof of the asymptotic normality of the generalized estimating equation estimators $\hat{\beta}$ and $\hat{\gamma}$. According to the proof of Theorem 1, we have

$$\sqrt{n}(\hat{\beta}-\beta) = \left\{ \frac{1}{n} \left(\frac{\partial \boldsymbol{\mu}^\top}{\partial \beta} \right) \boldsymbol{\Sigma}^{-1} \left(\frac{\partial \boldsymbol{\mu}^\top}{\partial \beta} \right)^\top \right\}^{-1} \left\{ \frac{1}{\sqrt{n}} \left(\frac{\partial \boldsymbol{\mu}^\top}{\partial \beta} \right) \boldsymbol{\Sigma}^{-1} (\boldsymbol{y}-\boldsymbol{\mu}) \right\} + o_p(1)$$

By Condition (A4) in Section 3, we can obtain that

$$\frac{1}{\sqrt{n}} \left(\frac{\partial \boldsymbol{\mu}^\top}{\partial \beta} \right) \boldsymbol{\Sigma}^{-1} (\boldsymbol{y}-\boldsymbol{\mu}) \xrightarrow{D} N_{p_\beta}(0, \boldsymbol{\Lambda}_\beta).$$

Using Condition (A7) in Ye and Pan (2006), we have, as $n \to \infty$,

$$\frac{1}{n} \left(\frac{\partial \boldsymbol{\mu}^\top}{\partial \beta} \right) \boldsymbol{\Sigma}^{-1} \left(\frac{\partial \boldsymbol{\mu}^\top}{\partial \beta} \right)^\top \xrightarrow{P} \mathbf{V}_\beta.$$

Therefore, $\sqrt{n}(\hat{\beta}-\beta) \xrightarrow{D} \mathcal{N}_{p_\beta}(\mathbf{0}, \mathbf{V}_\beta^{-1} \boldsymbol{\Lambda}_\beta \mathbf{V}_\beta^{-1})$. In a similar way, the asymptotic normality of parameter estimator $\hat{\gamma}$ can be easily derived. Based on the proof of the consistency of parameter estimator $\hat{\gamma}$, we have

$$\sqrt{N}(\hat{\gamma}-\gamma) = \left\{ \frac{1}{N} \left(\frac{\partial \boldsymbol{\eta}^\top}{\partial \gamma} \right) \mathbf{M}^{-1} \left(\frac{\partial \boldsymbol{\eta}^\top}{\partial \gamma} \right)^\top \right\}^{-1} \left\{ \frac{1}{\sqrt{N}} \left(\frac{\partial \boldsymbol{\eta}^\top}{\partial \gamma} \right) \mathbf{M}^{-1} \mathrm{vech}^*(\hat{\mathbf{r}}\hat{\mathbf{r}}^\top - F(\mathbf{R}(\gamma); \mathbf{c}(\hat{\beta}))) \right\} + o_p(1).$$

By Condition (A5) in Section 3, we have, as $N \to \infty$,

$$\begin{aligned}
\sqrt{N} J_2 &= \frac{1}{\sqrt{N}} \left(\frac{\partial \boldsymbol{\eta}^\top}{\partial \gamma} \right) \mathbf{M}^{-1} \mathrm{vech}^*(\mathbf{r}\mathbf{r}^\top - F(\mathbf{R}(\gamma); \mathbf{c}(\beta))) + o_p(1) \\
&= \frac{1}{\sqrt{N}} \left(\frac{\partial \boldsymbol{\eta}^\top}{\partial \gamma} \right) \mathbf{M}^{-1} \mathrm{vech}^*(\mathbf{r}\mathbf{r}^\top - \mathrm{E}(\mathbf{r}\mathbf{r}^\top)) + o_p(1) \xrightarrow{D} \mathcal{N}_{p_\gamma}(\mathbf{0}, \boldsymbol{\Lambda}_\gamma).
\end{aligned}$$

We expand $\sqrt{N}J_1$ and $\sqrt{N}J_3$ in a Taylor series as follows,

$$
\begin{aligned}
\sqrt{N}J_1 =& \frac{1}{\sqrt{N}}\left(\frac{\partial \boldsymbol{\eta}^\top}{\partial \gamma}\right)\mathbf{M}^{-1}\left(\frac{\partial \mathrm{vec}(\mathbf{r}\mathbf{r}^\top)^\top}{\partial \beta}\bigg|_{\tilde{\beta}^*}\right)^\top (\hat{\beta}-\beta),\\
\sqrt{N}J_3 =& \frac{1}{\sqrt{N}}\left(\frac{\partial \boldsymbol{\eta}^\top}{\partial \gamma}\right)\mathbf{M}^{-1}\left(\frac{\partial \mathrm{vech}^*(F(\mathbf{R}(\gamma);\mathbf{c}(\beta)))^\top}{\partial \beta}\bigg|_{\beta^*}\right)^\top (\hat{\beta}-\beta),
\end{aligned}
$$

where $\tilde{\beta}^*$ and β^* are two values in the neighborhood $\mathscr{N}(\hat{\beta},\beta)$, and for the (i,j)-th element of matrix $\mathbf{r}\mathbf{r}^\top$ and $F(\mathbf{R}(\gamma);\mathbf{c}(\beta))$, $1 \le i,j \le n$,

$$
\frac{\partial r_i r_j}{\partial \beta} = (\mu_i - y_i)\left(\frac{\partial g^{-1}(X_j^\top \beta)}{\partial \beta}\right) + (\mu_j - y_j)\left(\frac{\partial g^{-1}(X_i^\top \beta)}{\partial \beta}\right),
$$

$$
\begin{aligned}
\frac{\partial [F(R_{ij}(\gamma);c_i(\beta),c_j(\beta))]}{\partial \beta} =& \left\{\frac{\Phi\left(\frac{c_j(\beta)-R_{ij}(\gamma)c_i(\beta)}{\sqrt{1-(R_{ij}(\gamma))^2}}\right)\phi(c_i(\beta))}{\phi(1-\mu_i)} - (1-\mu_j)\right\}\left(\frac{\partial g^{-1}(X_i^\top \beta)}{\partial \beta}\right)\\
&+\left\{\frac{\Phi\left(\frac{c_i(\beta)-R_{ij}(\gamma)c_j(\beta)}{\sqrt{1-(R_{ij}(\gamma))^2}}\right)\phi(c_j(\beta))}{\phi(1-\mu_j)} - (1-\mu_i)\right\}\left(\frac{\partial g^{-1}(X_j^\top \beta)}{\partial \beta}\right),
\end{aligned}
$$

and

$$
\frac{\partial F(R_{ij}(\gamma);c_i(\beta),c_j(\beta))}{\partial \gamma} = \frac{\partial}{\partial \gamma}\int_{-\infty}^{c_i(\beta)} \Phi\left(\frac{c_j(\beta)-R_{ij}(\gamma)x}{\sqrt{1-(R_{ij}(\gamma))^2}}\right)\phi(x)dx.
$$

It is easy to obtain that $\sqrt{N}J_1 = O_p(1)$ and $\sqrt{N}J_3 = O_p(1)$ since $\|\hat{\beta}-\beta\|_2 = O_p(n^{-1/2})$ and $N = n(n-1)/2$. On the other hand, as $n \to \infty$,

$$
\frac{1}{N}\left(\frac{\partial \boldsymbol{\eta}^\top}{\partial \gamma}\right)\mathbf{M}^{-1}\left(\frac{\partial \boldsymbol{\eta}^\top}{\partial \gamma}\right)^\top \xrightarrow{P} \mathbf{V}_\gamma.
$$

Hence,

$$
\sqrt{N}(\hat{\gamma}-\gamma)|_{\hat{\beta}} \xrightarrow{D} \mathcal{N}_{p_\gamma}(\mathbf{V}_\gamma^{-1}(J_1^* + J_3^*), \mathbf{V}_\gamma^{-1}\boldsymbol{\Lambda}_\gamma \mathbf{V}_\gamma^{-1}),
$$

where $J_1^* = \lim_{N\to\infty} \sqrt{N} J_1$ and $J_3^* = \lim_{N\to\infty} \sqrt{N} J_3$. Let $J^* = J_1^* + J_3^*$, and then

$$\sqrt{N}(\hat{\gamma} - \gamma)|_{\hat{\beta}} \xrightarrow{D} \mathcal{N}_{p_\gamma}(\mathbf{V}_\gamma^{-1} J^*, \mathbf{V}_\gamma^{-1} \mathbf{\Lambda}_\gamma \mathbf{V}_\gamma^{-1}). \quad \square$$

S2 Further simulation studies

S2.1 Varying correlation strengths

In this subsection, we conduct numerical simulations to compare all the methods under varying correlation strengths and summarize the simulation results in Table 1. The settings of latent angle parameter γ are given as follows.

1. Weak latent correlation strength: $\gamma = (-0.2, 0.4, -0.2)^\top$ and $\zeta_{ij} = (1, d(s_i, s_j), d^2(s_i, s_j))^\top$. The ranges of latent correlation coefficients and correlations between binary responses are [6.123e-17,0.1561] and [-6.8239e-16,0.0997], respectively.

2. Moderate latent correlation strength: $\gamma = (-1, 0.4)^\top$ and $\zeta_{ij} = (1, d(s_i, s_j))^\top$. The ranges of latent correlation coefficients and correlations between pairs of binary responses are [0.3984,0.6909] and [0.2154,0.4855], respectively.

3. Strong latent correlation strength: $\gamma = (-2, 1)^\top$ and $\zeta_{ij} = (1, d(s_i, s_j))^\top$. The ranges of latent correlation coefficients and

correlations between pairs of binary responses are [0.5054,0.8838] and [0.2316,0.6900], respectively.

From Table 1, we can find that when the latent correlation strength is moderate or strong, the proposed joint modelling method is an improvement over the existing marginal model methodologies, in terms of the standard deviations (SDs) of mean structure parameters. In addition, the improvements are more obvious than that of the weak association case.

S2.2 Large sample size

In this subsection, we implement simulation experiments when the sample size n increases to 144, 225, and 400 (i.e., 12×12, 15×15, and 20×20 grids). Table 2 presents the estimate of each parameter and the associated standard deviation in parenthesis when the sample size $n \in \{100, 144, 225, 400\}$. It can be seen that the biases and standard deviations of parameter estimators improve as the sample size n increases.

S2.3 The convergence of algorithm

In Section 2.4, we introduce the quasi-Fisher scoring algorithm to solve the generalized estimating equations (2.5) & (2.6) and reach the numerical solutions for the parameters in our joint model. The number of iterations

Table 1: Simulation results on spatial binary data over 500 replications. The biases of parameter estimates are presented in the table (the standard deviations of parameter estimates are in parentheses). Various latent correlation strengths (weak, moderate, and strong) are considered. The sample size $n = 100$.

Latent correlation strength	Parameter	True	JMA	GLGM	GEE	ML
Weak	β_1	1	0.0527(0.4790)	0.1040(0.4962)	0.0580(0.4898)	0.0551(0.4920)
	β_2	−0.5	−0.0471(0.4661)	−0.0599(0.4810)	−0.0485(0.4758)	−0.0534(0.4797)
	β_3	0.3	0.0028(0.4517)	0.0085(0.4824)	0.0063(0.4500)	0.0093(0.4526)
	β_4	−0.7	−0.0793(0.8132)	−0.0915(0.8098)	−0.0751(0.8176)	−0.0662(0.8188)
	γ_1	−0.2	−0.0117(0.2280)			
	γ_2	0.4	0.0327(0.6780)			
	γ_3	−0.2	−0.0160(0.5194)			
Moderate	β_1	1	0.2876(0.6016)	1.1384(2.1508)	0.2964(0.6064)	0.3187(0.6262)
	β_2	−0.5	−0.1606(0.4768)	−0.6008(1.7934)	−0.1588(0.4734)	−0.1655(0.5286)
	β_3	0.3	0.0665(0.4116)	0.3153(1.7150)	0.0763(0.4230)	0.0847(0.4799)
	β_4	−0.7	−0.2358(0.7118)	−0.7959(3.0312)	−0.2159(0.7143)	−0.2159(0.8342)
	γ_1	−1	0.0225(0.9745)			
	γ_2	0.4	0.0832(0.8549)			
Strong	β_1	1	0.6955(0.8451)	2.9471(4.2224)	0.7549(0.8802)	0.7855(0.8929)
	β_2	−0.5	−0.3231(0.5198)	−1.3912(2.8638)	−0.3501(0.5529)	−0.3681(0.6174)
	β_3	0.3	0.2100(0.4677)	0.9769(2.7131)	0.2495(0.4976)	0.2462(0.5687)
	β_4	−0.7	−0.4407(0.8220)	−2.0427(4.2480)	−0.4609(0.8753)	−0.4751(0.9861)
	γ_1	−2	0.0931(1.9336)			
	γ_2	1	0.3085(1.5228)			

needed for convergence of the algorithm depends on the convergence criterion (stopping rule) and starting parameter value. we conduct further simulation studies to study the effect of iterations on the convergence. For different convergence criteria (the l_2-norm of parameter estimate change: $||\beta^{(k+1)} - \beta^{(k)}||_2 \leq \epsilon$ and $||\gamma^{(k+1)} - \gamma^{(k)}||_2 \leq \epsilon$ for any given $\epsilon > 0$), the numbers of iterations needed to meet them are recorded and then displayed in Figure 1 The sample means and standard errors of parameter estimates are summarized in Table 3. Based on the simulation results shown in Figure 1 and Table 3, we can conclude that a higher degree of convergence of the parameter estimates obtained by the proposed algorithm is achieved as the number of iterations grows.

Moreover, we also implement two simulation experiments to study the sensitivity of the parameter estimates concerning the starting values: (1) for each replication, the starting value of the angle parameter vector γ is independently sampled from a uniform distribution $\mathcal{U}_3(-1, 1)$; (2) for each replication, the starting values of parameter vectors β and γ are independently drawn from uniform distributions $\mathcal{U}_3(-1, 1)$ and $\mathcal{U}_4(-1, 1)$, respectively. Table 4 displays the sample means and standard errors of parameter estimates in the above settings. Compared with the results shown in Table 3, it is easy to find that the parameter estimates are not sensitive to the starting values, although the number of iterations needed for convergence is larger than that of the starting value setting proposed in section 2.4.

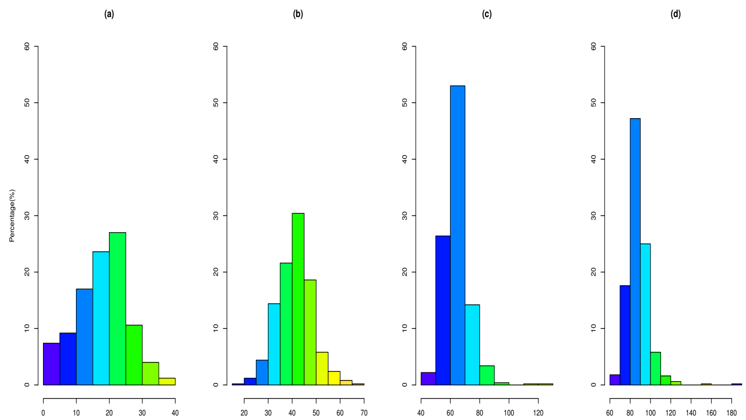

Figure 1: The number of iterations used to update the mean parameter β and angle parameter γ under different convergence criteria. The x-axis is the number of iterations in one replication, and the y-axis is the percentage of replications which require that number of iterations to reach the convergence criterion. The four plots correspond to four convergence criteria (a) $\epsilon = 10^{-2}$, (b) $\epsilon = 10^{-3}$, (c) $\epsilon = 10^{-4}$ and (d) $\epsilon = 10^{-5}$. For each convergence criterion, the number of replications is 500.

Table 2: Simulation results on spatial binary data over 500 replications. The biases of parameter estimates are presented in the table (the standard deviations of parameter estimates are in parentheses). The sample size $n \in \{100, 144, 225, 400\}$.

	True	JMA	GLGM	GEE	ML
$n = 100$					
β_1	1	0.0527(0.4790)	0.1040(0.4962)	0.0580(0.4898)	0.0551(0.4920)
β_2	−0.5	−0.0471(0.4661)	−0.0599(0.4810)	−0.0485(0.4758)	−0.0534(0.4797)
β_3	0.3	0.0028(0.4517)	0.0085(0.4824)	0.0063(0.4500)	0.0093(0.4526)
β_4	−0.7	−0.0793(0.8132)	−0.0915(0.8098)	−0.0751(0.8176)	−0.0662(0.8188)
γ_1	−0.2	−0.0117(0.2280)			
γ_2	0.4	0.0327(0.6780)			
γ_3	−0.2	−0.0160(0.5194)			
$n = 144$					
β_1	1	0.0411(0.3903)	0.1040(0.4962)	0.0400(0.3983)	0.0396(0.4021)
β_2	−0.5	−0.0132(0.3731)	−0.0599(0.4810)	−0.0163(0.3749)	−0.0170(0.3826)
β_3	0.3	0.0020(0.3688)	0.0085(0.4824)	0.0016(0.3756)	0.0032(0.3801)
β_4	−0.7	−0.1019(0.6437)	−0.0915(0.8098)	−0.0937(0.6464)	−0.0961(0.6450)
γ_1	−0.2	−0.0105(0.1988)			
γ_2	0.4	0.0323(0.5105)			
γ_3	−0.2	−0.0232(0.3797)			
$n = 225$					
β_1	1	0.0182(0.3045)	0.0691(0.3149)	0.0211(0.3074)	0.0204(0.3104)
β_2	−0.5	−0.0009(0.3024)	−0.0258(0.3165)	−0.0004(0.3049)	−0.0020(0.3097)
β_3	0.3	0.0217(0.2771)	0.0390(0.2846)	0.0215(0.2756)	0.0238(0.2787)
β_4	−0.7	−0.0015(0.4664)	−0.0333(0.4943)	−0.0012(0.4758)	−0.0071(0.4785)
γ_1	−0.2	−0.0062(0.1514)			
γ_2	0.4	0.0145(0.4138)			
γ_3	−0.2	−0.0112(0.3261)			
$n = 400$					
β_1	1	0.0245(0.2354)	0.0626(0.2313)	0.0294(0.2382)	0.0280(0.2442)
β_2	−0.5	−0.0187(0.2168)	−0.0399(0.2341)	−0.0193(0.2196)	−0.0175(0.2243)
β_3	0.3	−0.0050(0.2102)	0.0165(0.2159)	−0.0056(0.2145)	−0.0102(0.2204)
β_4	−0.7	−0.0070(0.3728)	−0.0337(0.3760)	−0.0061(0.3766)	−0.0053(0.3802)
γ_1	−0.2	0(0.1293)			
γ_2	0.4	−0.0168(0.3131)			
γ_3	−0.2	0.0148(0.2597)			

Table 3: Simulation results on balanced data over 500 replications. The sample means of all the parameter estimates with sample standard errors in parentheses are shown in the table. ANI is the averaged number of iterations used to update the estimates of mean and angle parameters, and the values of ϵ listed below are the convergence criteria (estimation precision).

	True	(a) $\epsilon = 10^{-2}$	(b) $\epsilon = 10^{-3}$	(c) $\epsilon = 10^{-4}$	(d) $\epsilon = 10^{-5}$
β_1	1	1.0536(0.4755)	1.0527(0.4790)	1.0527(0.4791)	1.0527(0.4791)
β_2	-0.5	-0.5464(0.4662)	-0.5471(0.4661)	-0.5472(0.4663)	-0.5472(0.4663)
β_3	0.3	0.3017(0.4488)	0.3028(0.4517)	0.3029(0.4519)	0.3029(0.4519)
β_4	-0.7	-0.7817(0.8124)	-0.7793(0.8132)	-0.7792(0.8132)	-0.7792(0.8132)
γ_1	-0.2	-0.2166(0.2091)	-0.2117(0.2280)	-0.2111(0.2298)	-0.2111(0.2299)
γ_2	0.4	0.4425(0.6337)	0.4327(0.6780)	0.4314(0.6821)	0.4312(0.6825)
γ_3	-0.2	-0.2213(0.4842)	-0.2160(0.5194)	-0.2152(0.5227)	-0.2151(0.5231)
ANI		18.32	41.66	64.948	88.398

Table 4: Simulation results on balanced data over 500 replications. The sample means of all the parameter estimates with sample standard errors in parentheses are shown in the table. ANI is the averaged number of iterations used to update the estimates of mean and angle parameters. (1) and (2) are two starting value settings of parameters.

	True	(1)	(2)
β_1	1	1.0525(0.4792)	1.0526(0.4788)
β_2	-0.5	$-0.5471(0.4660)$	$-0.5472(0.4664)$
β_3	0.3	0.3033(0.4526)	0.3030(0.4520)
β_4	-0.7	$-0.7794(0.8130)$	$-0.7791(0.8131)$
γ_1	-0.2	$-0.2112(0.2308)$	$-0.2103(0.2290)$
γ_2	0.4	0.4272(0.6808)	0.4295(0.6801)
γ_3	-0.2	$-0.2133(0.5211)$	$-0.2146(0.5208)$
ANI		51.696	51.858

Bibliography

McCullagh, P. (1983). Quasi-likelihood functions. *The Annals of Statistics 11*(1), 59–67.

Ye, H. and J. Pan (2006). Modelling covariance structures in generalized estimating equations for longitudinal data. *Biometrika 93*(4), 927–941.

Zhang, W., C. Leng, and C. Y. Tang (2015). A joint modelling approach for longitudinal studies. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 77*(1), 219–238.