Supplementary Materials to 'A Finite Mixture Model for Working Correlation Matrices in GEE'

Lili Xu[1], Nan Lin[2], Baoxue Zhang[1] and Ning-Zhong Shi[1]

[1]*Northeast Normal University and* [2] *Washington University in St. Louis*

1. A multivariate continuous model

In this section, we provide the following concrete example that satisfies all the assumptions of Theorems 3 and 5 in our manuscript.

Let y_{ik} be the response at time k for the i th subject from $N(\mu_{ik} = \beta_0 + x_{ik}\beta_1, 1)$, $k = 1, \ldots, T = 3$, $i = 1, \ldots, n = 50$. Let $\boldsymbol{y}_i = (y_{i1}, \cdots, y_{iT})^\tau$. The true value of (β_0, β_1) is (1, -1). Covariates x_{ik}'s are iid randomly sampled from $N(k, 1)$. We consider the case in which the true correlation of $\boldsymbol{y}_i$ is given by a two component mixture of AR(1) and CS. Let $R^{(1)}(\alpha_1) = R_{AR(1)}(\alpha_1)$ and $R^{(2)}(\alpha_2) = R_{CS}(\alpha_2)$. The true values of the correlation parameters (α_1, α_2) and the relative component proportions (π_1, π_2) are (0.6,0.6) and (0.3,0.7).

1.1 *Conditions of Theorem 3.*

In the following, we will present detailed verification of the conditions of Theorem 3. It is quite obvious that assumptions (i) and (ii) of Theorem 3 and the condition $E\{\frac{\partial \ell(\boldsymbol{\psi})}{\partial \boldsymbol{\psi}}\} = \frac{\partial E\{\ell(\boldsymbol{\psi})\}}{\partial \boldsymbol{\psi}}$ are satisfied.

For assumption(iii), we first have

$$|\log \sum_{l=1}^{L} \pi_l \phi_l(u, \alpha_l)| \leq |\log[\max \phi_l(u, \alpha_l)]| + |\log[\min \phi_l(u, \alpha_l)]| \leq \sum_{l=1}^{L} |\log \phi_l(u, \alpha_l)|$$
$$= \sum_{l=1}^{L} |1/\sqrt{2\pi}|^T + \sum_{l=1}^{L} |\frac{1}{\sqrt{|R^{(l)}(\alpha_l)|}}| + \frac{1}{2} \sum_{l=1}^{L} |trace[{R^{(l)}(\alpha_l)}^{-1}(uu^\tau)]|,$$

where

$$|R_{AR(1)}(\alpha_1)| = (1 - {\alpha_1}^2)^2, |R_{CS}(\alpha_2)| = (1 - \alpha_2)^2(1 + 2\alpha_2),$$

$$[R_{AR(1)}(\alpha_1)]^{-1} = \frac{1}{(1-{\alpha_1}^2)} \begin{bmatrix} 1 & -\alpha_1 & 0 \\ -\alpha_1 & 1+{\alpha_1}^2 & -\alpha_1 \\ 0 & -\alpha_1 & 1 \end{bmatrix},$$

$$\text{and} \quad [R_{CS}(\alpha_2)]^{-1} = \frac{1}{(1-\alpha_2)(1+2\alpha_2)} \begin{bmatrix} 1 & -\alpha_2 & -\alpha_2 \\ -\alpha_2 & 1 & -\alpha_2 \\ -\alpha_2 & -\alpha_2 & 1 \end{bmatrix}.$$

Second, the multivariate normal distribution G has a finite second-order moment. Therefore, $|\log \sum_{l=1}^{L} \pi_l \phi_l(u, \alpha_l)| \leq M(u)$ for all $\boldsymbol{\psi} = (\pi_l, \alpha_l)_{l=1}^{L} \in \Psi$ and all $u \in \Omega$, where M is an integrable function with respect to G. Hence assumption(iii) is satisfied.

Next, we verify the rest conditions required by Theorem 3. First, $\left[\frac{\partial^2 \ell(\boldsymbol{\psi})}{\partial \boldsymbol{\psi} \partial \boldsymbol{\psi}^\tau}\right]$ contains a finite number of elements. Second, we have

$$\left[\frac{\partial \ell(\boldsymbol{\psi})}{\partial \pi_1}\right] = \left[\frac{\partial \sum_{i=1}^{N} \left(\log \sum_{l=1}^{L} \pi_l \phi_l(\boldsymbol{\varepsilon_i}, \alpha_l)\right)}{\partial \pi_1}\right] = \sum_{i=1}^{N} \left[\frac{(\phi_1(\boldsymbol{\varepsilon_i}, \alpha_1) - \phi_2(\boldsymbol{\varepsilon_i}, \alpha_2))}{\left(\sum_{l=1}^{L} \pi_l \phi_l(\boldsymbol{\varepsilon_i}, \alpha_l)\right)}\right]$$

and

$$\left[\frac{\partial \ell(\boldsymbol{\psi})}{\partial \alpha_l}\right] = \left[\frac{\partial \sum_{i=1}^{N} \left(\log \sum_{l=1}^{L} \pi_l \phi_l(\boldsymbol{\varepsilon_i}, \alpha_l)\right)}{\partial \alpha_l}\right] = -\frac{1}{2} \sum_{i=1}^{N} \left[\frac{\pi_l \phi_l(\boldsymbol{\varepsilon_i}, \alpha_l)}{\left(\sum_{l=1}^{L} \pi_l \phi_l(\boldsymbol{\varepsilon_i}, \alpha_l)\right)} h_l(\boldsymbol{\varepsilon}_i, \alpha_l)\right],$$

where

$$\phi_l(\boldsymbol{\varepsilon}_i, \alpha_l) = (1/\sqrt{2\pi}) |R^{(l)}(\alpha_l)|^{-1/2} \exp\left\{-\frac{1}{2} {\boldsymbol{\varepsilon}_i}^\tau [R^{(l)}(\alpha_l)]^{-1} \boldsymbol{\varepsilon}_i\right\}$$

$$\text{and} \quad h_l(\boldsymbol{\varepsilon}_i, \alpha_l) = \{\frac{\partial |R^{(l)}(\alpha_l)|^{-1/2}}{\partial \alpha_l} |R^{(l)}(\alpha_l)|^{1/2} + \frac{\partial tr(R^{(l)}(\alpha_l)^{-1} \boldsymbol{\varepsilon_i} {\boldsymbol{\varepsilon_i}}^\tau)}{\partial \alpha_l}\}.$$

Therefore,

$$\left[\frac{\partial^2 \ell(\boldsymbol{\psi})}{\partial \pi_1 \partial \pi_1}\right] = -\sum_{i=1}^{N}\left[\frac{(\phi_1(\boldsymbol{\varepsilon_i},\alpha_1)-\phi_2(\boldsymbol{\varepsilon_i},\alpha_2))^2}{\left(\sum_{l=1}^{L}\pi_l\phi_l(\boldsymbol{\varepsilon_i},\alpha_l)\right)^2}\right],$$

$$\left[\frac{\partial^2 \ell(\boldsymbol{\psi})}{\partial \alpha_l \partial \alpha_l}\right] = \frac{1}{2}\sum_{i=1}^{N}\left[\frac{\pi_l\phi_l(\boldsymbol{\varepsilon_i},\alpha_l)}{\left(\sum_{l=1}^{L}\pi_l\phi_l(\boldsymbol{\varepsilon_i},\alpha_l)\right)}[h_l^2(\boldsymbol{\varepsilon}_i,\alpha_l)+\frac{\partial h_l(\boldsymbol{\varepsilon}_i,\alpha_l)}{\partial\alpha_l}]\right] - \frac{1}{2}\sum_{i=1}^{N}\left[\frac{\pi_l\phi_l^2(\boldsymbol{\varepsilon_i},\alpha_l)}{\left(\sum_{l=1}^{L}\pi_l\phi_l(\boldsymbol{\varepsilon_i},\alpha_l)\right)^2}h_l(\boldsymbol{\varepsilon}_i,\alpha_l)\right],$$

$$\text{and } \left[\frac{\partial^2 \ell(\boldsymbol{\psi})}{\partial \pi_1 \partial \alpha_1}\right] = \frac{1}{2}\sum_{i=1}^{N}\left[\frac{\phi_l(\boldsymbol{\varepsilon_i},\alpha_l)}{\left(\sum_{l=1}^{L}\pi_l\phi_l(\boldsymbol{\varepsilon_i},\alpha_l)\right)}h_l(\boldsymbol{\varepsilon}_i,\alpha_l)\right] - \frac{1}{2}\sum_{i=1}^{N}\left[\frac{\pi_l^2\phi_l^2(\boldsymbol{\varepsilon_i},\alpha_l)}{\left(\sum_{l=1}^{L}\pi_l\phi_l(\boldsymbol{\varepsilon_i},\alpha_l)\right)^2}h_l^2(\boldsymbol{\varepsilon}_i,\alpha_l)\right].$$

The elements of $\left[\frac{\partial^2 \ell(\boldsymbol{\psi})}{\partial \boldsymbol{\psi}\partial \boldsymbol{\psi}^\tau}\right]$ contain $\frac{\partial^2 \ell(\boldsymbol{\psi})}{\partial \pi_1 \partial \pi_1}$, $\frac{\partial^2 \ell(\boldsymbol{\psi})}{\partial \pi_1 \partial \pi_1}$ and $\frac{\partial^2 \ell(\boldsymbol{\psi})}{\partial \pi_1 \partial \alpha_1}$. Because $\frac{\pi_l\phi_l(\boldsymbol{\varepsilon_i},\alpha_l)}{\left(\sum_{l=1}^{L}\pi_l\phi_l(\boldsymbol{\varepsilon_i},\alpha_l)\right)}$ and $\frac{\pi_l^2\phi_l^2(\boldsymbol{\varepsilon_i},\alpha_l)}{\left(\sum_{l=1}^{L}\pi_l\phi_l(\boldsymbol{\varepsilon_i},\alpha_l)\right)^2}$ are bounded, and $h_l(\boldsymbol{\varepsilon}_i,\alpha_l)$ is a quadratic function of $\boldsymbol{\varepsilon}_i$, it follows that $E\left[\frac{\partial^2 \ell(\boldsymbol{\psi})}{\partial \boldsymbol{\psi}\partial \boldsymbol{\psi}^\tau}\right]$ is continuous in Ψ, and there is a neighborhood $\mathcal{N}$ of $\boldsymbol{\psi}$ such that $E\left[\sup_{\boldsymbol{\psi}\in\mathcal{N}}\left|\frac{\partial^2 \ell(\boldsymbol{\psi})}{\partial \boldsymbol{\psi}\partial \boldsymbol{\psi}^\tau}\right|\right] < \infty$.

The elements of $\mathrm{E}\left[\left(\frac{\partial \log\sum_{l=1}^{L}\pi_l\phi_l(\varepsilon_\mathbf{i},\alpha_l)}{\partial\boldsymbol{\psi}}\right)\left(\frac{\partial \log\sum_{l=1}^{L}\pi_l\phi_l(\varepsilon_\mathbf{i},\alpha_l)}{\partial\boldsymbol{\psi}}\right)^T\Big|_{\boldsymbol{\psi}=\boldsymbol{\psi}^*}\right]$ contain $E\left[\frac{\partial \ell(\boldsymbol{\psi})}{\partial \pi_1}\right]^2$, $E\left[\frac{\partial \ell(\boldsymbol{\psi})}{\partial \pi_1}\frac{\partial \ell(\boldsymbol{\psi})}{\partial \alpha_l}\right]$ and $E\left[\frac{\partial \ell(\boldsymbol{\psi})}{\partial \alpha_l}\right]^2$, where

$$E\left[\frac{\partial \ell(\boldsymbol{\psi})}{\partial \pi_1}\right]^2 = E\left[\sum_{i=1}^{N}\frac{(\phi_1(\boldsymbol{\varepsilon_i},\alpha_1)-\phi_2(\boldsymbol{\varepsilon_i},\alpha_2))}{\left(\sum_{l=1}^{L}\pi_l\phi_l(\boldsymbol{\varepsilon_i},\alpha_l)\right)}\right]^2 = E\sum_{i=1}^{N}\left[\frac{(\phi_1(\boldsymbol{\varepsilon_i},\alpha_1)-\phi_2(\boldsymbol{\varepsilon_i},\alpha_2))}{\left(\sum_{l=1}^{L}\pi_l\phi_l(\boldsymbol{\varepsilon_i},\alpha_l)\right)}\right]^2,$$

$$E\left[\frac{\partial \ell(\boldsymbol{\psi})}{\partial \pi_1}\frac{\partial \ell(\boldsymbol{\psi})}{\partial \alpha_l}\right] = \frac{1}{2}\sum_{i=1}^{N}E\left[\frac{(\phi_1(\boldsymbol{\varepsilon_i},\alpha_1)-\phi_2(\boldsymbol{\varepsilon_i},\alpha_2))}{\left(\sum_{l=1}^{L}\pi_l\phi_l(\boldsymbol{\varepsilon_i},\alpha_l)\right)}\frac{\pi_l\phi_l(\boldsymbol{\varepsilon_i},\alpha_l)}{\left(\sum_{l=1}^{L}\pi_l\phi_l(\boldsymbol{\varepsilon_i},\alpha_l)\right)}h_l(\boldsymbol{\varepsilon}_i,\alpha_l)\right],$$

$$E\left[\frac{\partial \ell(\boldsymbol{\psi})}{\partial \alpha_l}\right]^2 = \sum_{i=1}^{N}E\left[\frac{\pi_l\phi_l(\boldsymbol{\varepsilon_i},\alpha_l)}{\left(\sum_{l=1}^{L}\pi_l\phi_l(\boldsymbol{\varepsilon_i},\alpha_l)\right)}\right]^2 h_l^2(\boldsymbol{\varepsilon}_i,\alpha_l).$$

It is easy to see that $\frac{(\phi_1(\boldsymbol{\varepsilon_i},\alpha_1)-\phi_2(\boldsymbol{\varepsilon_i},\alpha_2))}{\left(\sum_{l=1}^{L}\pi_l\phi_l(\boldsymbol{\varepsilon_i},\alpha_l)\right)}$ is bounded because $\frac{(\phi_1(\boldsymbol{\varepsilon_i},\alpha_1)-\phi_2(\boldsymbol{\varepsilon_i},\alpha_2))}{\left(\sum_{l=1}^{L}\pi_l\phi_l(\boldsymbol{\varepsilon_i},\alpha_l)\right)} = \frac{1}{\pi_1}\frac{\pi_1\phi_1(\boldsymbol{\varepsilon_i},\alpha_1)}{\left(\sum_{l=1}^{L}\pi_l\phi_l(\boldsymbol{\varepsilon_i},\alpha_l)\right)} - \frac{1}{\pi_2}\frac{\pi_2\phi_2(\boldsymbol{\varepsilon_i},\alpha_1)}{\left(\sum_{l=1}^{L}\pi_l\phi_l(\boldsymbol{\varepsilon_i},\alpha_l)\right)}$ and $E[h_l(\boldsymbol{\varepsilon}_i,\alpha_l)] < \infty$. Therefore, $E\left[\frac{\partial \ell(\boldsymbol{\psi})}{\partial \pi_1}\right]^2 < \infty$, $E\left[\frac{\partial \ell(\boldsymbol{\psi})}{\partial \pi_1}\frac{\partial \ell(\boldsymbol{\psi})}{\partial \alpha_l}\right] < \infty$ and $E\left[\frac{\partial \ell(\boldsymbol{\psi})}{\partial \alpha_l}\right]^2 < \infty$, and then $E\left[\left(\frac{\partial \log\sum_{l=1}^{L}\pi_l\phi_l(\varepsilon_\mathbf{i},\alpha_l)}{\partial\boldsymbol{\psi}}\right)\left(\frac{\partial \log\sum_{l=1}^{L}\pi_l\phi_l(\varepsilon_\mathbf{i},\alpha_l)}{\partial\boldsymbol{\psi}}\right)^T\Big|_{\boldsymbol{\psi}=\boldsymbol{\psi}^*}\right] < \infty$.

In summary, our model satisfies all the conditions of Theorem 3. More

generally, even when G is a non-normal distribution, as long as it has a finite fourth moment, the assumptions of Theorem 3 are satisfied.

1.2. *Conditions of Theorem 5*

In this following, we verify the six conditions in Theorem 5 one by one.

Condition 1 The condition 'There exists $H(\mathbf{Y}, \boldsymbol{\beta}) = O_p(1)$ such that $|\partial\hat{\boldsymbol{\psi}}/\partial\boldsymbol{\beta}| \leq H(\mathbf{Y}, \boldsymbol{\beta})$ which is $O_p(1)$, where $\mathbf{Y} = \{\boldsymbol{y}_i, i = 1, 2, ..., N\}$' is a general condition and is similar to condition (iii) of Theorem 2 in Liang and Zeger (1986). In deriving the asymptotic property of our mix-GEE estimator, the estimating equation for regression parameters is Taylor expanded as a sum of independent variables plus a remainder term. This condition ensures that the remainder term is neglegible.

Condition 2 Because $\partial^2 U_i(\boldsymbol{\beta}, M(\boldsymbol{\psi}))/\partial\boldsymbol{\psi}\partial\boldsymbol{\psi}^\tau = X_i[\partial^2 M^{-1}(\boldsymbol{\psi})/\partial\boldsymbol{\psi}\partial\boldsymbol{\psi}^\tau](\boldsymbol{y}_i - \boldsymbol{X}_i^\tau\boldsymbol{\beta})$, it is continuous at $\boldsymbol{\psi}^*$ with probability one and there is a neighborhood $\mathcal{N}$ of $\boldsymbol{\psi}$ such that

$$\begin{aligned} E\left[\sup_{\boldsymbol{\psi}\in\mathcal{N}} \left|\partial^2 U_i(\boldsymbol{\beta}, M(\boldsymbol{\psi}))/\partial\boldsymbol{\psi}\partial\boldsymbol{\psi}^\tau\right|\right] &= \sup_{\boldsymbol{\psi}\in\mathcal{N}} E|X_i[\partial^2 M^{-1}(\boldsymbol{\psi})/\partial\boldsymbol{\psi}\partial\boldsymbol{\psi}^\tau](\boldsymbol{y}_i - \boldsymbol{X}_i^\tau\boldsymbol{\beta})| \\ &= \sup_{\boldsymbol{\psi}\in\mathcal{N}} |X_i[\partial^2 M^{-1}(\boldsymbol{\psi})/\partial\boldsymbol{\psi}\partial\boldsymbol{\psi}^\tau]|E|(\boldsymbol{y}_i - \boldsymbol{X}_i^\tau\boldsymbol{\beta})| \\ &= |X_i[\partial^2 M^{-1}(\boldsymbol{\psi})/\partial\boldsymbol{\psi}\partial\boldsymbol{\psi}^\tau]|_{\boldsymbol{\psi}=\boldsymbol{\psi}^{\mathcal{N}}} E|(\boldsymbol{y}_i - \boldsymbol{X}_i^\tau\boldsymbol{\beta})|, \end{aligned}$$

where $|X_i[\partial^2 M^{-1}(\boldsymbol{\psi})/\partial\boldsymbol{\psi}\partial\boldsymbol{\psi}^\tau|$ has a maximum at $\boldsymbol{\psi}^{\mathcal{N}}$ in $\mathcal{N}$. Therefore, as long as $E\,|(\boldsymbol{y}_i - \boldsymbol{X}_i^\tau\boldsymbol{\beta})| < \infty$, we have $E\left[\sup_{\boldsymbol{\psi}\in\mathcal{N}} |\partial^2 U_i(\boldsymbol{\beta}, M(\boldsymbol{\psi}))/\partial\boldsymbol{\psi}\partial\boldsymbol{\psi}^\tau|\right] < \infty$. Finally, note that

$$E\,|(\boldsymbol{y}_i - \boldsymbol{X}_i^\tau\boldsymbol{\beta})| = E\,|(\boldsymbol{y}_i - \boldsymbol{X}_i^\tau\boldsymbol{\beta}_0 + \boldsymbol{X}_i^\tau\boldsymbol{\beta}_0 - \boldsymbol{X}_i^\tau\boldsymbol{\beta})| = E\,|\boldsymbol{y}_i - \boldsymbol{X}_i^\tau\boldsymbol{\beta}_0| + |\boldsymbol{X}_i^\tau\boldsymbol{\beta}_0 - \boldsymbol{X}_i^\tau\boldsymbol{\beta}| < \infty.$$

So, condition 2 is satisfied.

Condition 3 First, it is easy to see that

$$\left[\frac{1}{N}\sum_{i=1}^{N} \frac{\partial U_i\{\boldsymbol{\beta}, M(\boldsymbol{\psi})\}}{\partial\boldsymbol{\beta}}\right]^{-1} = \left[\frac{1}{N}\sum_{i=1}^{N} X_i M^{-1}(\boldsymbol{\psi}) X_i\right]^{-1}.$$

Second, because covariates x_{ik}'s are iid randomly sampled from $N(k, 1)$,

we have

$$\left[\frac{1}{N}\sum_{i=1}^{N} X_i M^{-1}(\boldsymbol{\psi}) X_i\right] = \frac{1}{N}\sum_{i=1}^{N}\sum_{j,k=1}^{T} x_{ij}x_{ik}M_{jk} \to E[\sum_{j,k=1}^{T} x_{ij}x_{ik}M_{jk}] \quad a.e.,$$

where M_{jk} denotes the element at the jth row and kth column of $M^{-1}(\boldsymbol{\psi})$ and

$$E[\sum_{j,k=1}^{T} x_{ij}x_{ik}M_{jk}] = \sum_{k=1}^{T} E[x_{ik}]^2 M_{jk} = \sum_{k=1}^{T} M_{jk}(1+k^2).$$

Therefore, there is a neighborhood $\mathcal{N}$ of $\boldsymbol{\psi}^*$ such that $\left[\frac{1}{N}\sum_{i=1}^{N}\frac{\partial U_i\{\boldsymbol{\beta}, M(\boldsymbol{\psi})\}}{\partial \boldsymbol{\beta}}\right]^{-1}$ is bounded in $\mathcal{N}$ with probability one;

Condition 4 $E[U_i(\boldsymbol{\beta}, M(\boldsymbol{\psi}^*))] = 0$ has a unique solution $\boldsymbol{\beta}_0$, because $E[U_i(\boldsymbol{\beta}, M(\boldsymbol{\psi}^*))] = \boldsymbol{X}_i M^{-1}(\boldsymbol{\psi}^*)(\boldsymbol{X}_i^\tau\boldsymbol{\beta} - \boldsymbol{X}_i^\tau\boldsymbol{\beta}_0)$ and $\boldsymbol{X}_i M^{-1}(\boldsymbol{\psi}^*)\boldsymbol{X}_i^\tau$ is a positive definite matrix;

Condition 5 'The support of $\boldsymbol{\beta}$ is compact' is a general condition.

Condition 6 It is easy to see that

$$\left[\frac{\partial^2 U_i\{\boldsymbol{\beta}, M(\hat{\boldsymbol{\psi}}(\boldsymbol{\beta}))\}}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^\tau}\right] = -2X_i[\partial M^{-1}(\boldsymbol{\psi})/\partial\boldsymbol{\psi}]X_i^\tau + X_i[\partial^2 M^{-1}(\boldsymbol{\psi})/\partial\boldsymbol{\psi}\partial\boldsymbol{\psi}^\tau](\boldsymbol{y}_i - \boldsymbol{X}_i^\tau\boldsymbol{\beta}).$$

It follows that

$$\begin{aligned} E\left[\sup_{\boldsymbol{\beta}\in\mathcal{N}}\left|\frac{\partial^2 U_i\{\boldsymbol{\beta}, M(\hat{\boldsymbol{\psi}}(\boldsymbol{\beta}))\}}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^\tau}\right|\right] &= \sup_{\boldsymbol{\beta}\in\mathcal{N}}\left|2X_i[\partial M^{-1}(\boldsymbol{\psi})/\partial\boldsymbol{\psi}]X_i^\tau\right| + E\left[\sup_{\boldsymbol{\beta}\in\mathcal{N}}\left|X_i[\partial^2 M^{-1}(\boldsymbol{\psi})/\partial\boldsymbol{\psi}\partial\boldsymbol{\psi}^\tau](\boldsymbol{y}_i - \boldsymbol{X}_i^\tau\boldsymbol{\beta})\right|\right] \\ &= \left|2X_i[\partial M^{-1}(\boldsymbol{\psi})/\partial\boldsymbol{\psi}]X_i^\tau\right| + \left|X_i[\partial^2 M^{-1}(\boldsymbol{\psi})/\partial\boldsymbol{\psi}\partial\boldsymbol{\psi}^\tau]\right| E\left[\sup_{\boldsymbol{\beta}\in\mathcal{N}}\left|(\boldsymbol{y}_i - \boldsymbol{X}_i^\tau\boldsymbol{\beta})\right|\right]. \end{aligned}$$

Since

$$E\left[\sup_{\boldsymbol{\beta}\in\mathcal{N}}\left|(\boldsymbol{y}_i - \boldsymbol{X}_i^\tau\boldsymbol{\beta})\right|\right] = E\left[\sup_{\boldsymbol{\beta}\in\mathcal{N}}\left|(\boldsymbol{y}_i - \boldsymbol{X}_i^\tau\boldsymbol{\beta}_0 + \boldsymbol{X}_i^\tau\boldsymbol{\beta}_0 - \boldsymbol{X}_i^\tau\boldsymbol{\beta})\right|\right] = E\left|\boldsymbol{y}_i - \boldsymbol{X}_i^\tau\boldsymbol{\beta}_0\right| + \sup_{\boldsymbol{\beta}\in\mathcal{N}}\left|\boldsymbol{X}_i^\tau\boldsymbol{\beta}_0 - \boldsymbol{X}_i^\tau\boldsymbol{\beta}\right| < \infty,$$

we have $E\left[\sup_{\boldsymbol{\beta}\in\mathcal{N}}\left|\frac{\partial^2 U_i\{\boldsymbol{\beta}, M(\hat{\boldsymbol{\psi}}(\boldsymbol{\beta}))\}}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^\tau}\right|\right] < \infty$.

2. A multivariate binary model

Let $y_{it} = I[z_{it} \le \beta_1 x_{it}]$ be the binary response at time k for the ith

subject, $k = 1, \ldots, T = 3$, $i = 1, \ldots, n = 50$. The latent variable $Z_i = (z_{i1}, ...z_{iT})$ is form a two-component multivariate normal mixture, where the two components have both mean 0 and correlation structures as CS and AR(1), respectively. The true value of β_1 is 1. Covariates x_{ik}'s are randomly sampled from $U(-1, 1)$.

For the binary probit model, $\boldsymbol{\mu}_i = E\boldsymbol{y}_i = (\Phi(\boldsymbol{x}_{i1}^\tau\boldsymbol{\beta}), ...\Phi(\boldsymbol{x}_{iT}^\tau\boldsymbol{\beta}))^\tau$ and $E|(\boldsymbol{y}_i - \boldsymbol{\mu}_i)| = (|\boldsymbol{y}_{i1} - \boldsymbol{\mu}_{i1}|, ..., |\boldsymbol{y}_{iT} - \boldsymbol{\mu}_{iT}|)^\tau$, where Φ denotes normal distribution function. Then we have

$$\begin{aligned} E|(\boldsymbol{y}_{ij} - \boldsymbol{\mu}_{ij})| &\le \sqrt{E|(\boldsymbol{y}_{ij} - \boldsymbol{\mu}_{ij}(\boldsymbol{\beta}_0))|^2} + |\boldsymbol{\mu}_{ij}(\boldsymbol{\beta}_0) - \boldsymbol{\mu}_{ij}(\boldsymbol{\beta})| \\ &= \sqrt{\Phi(\boldsymbol{x}_{ij}^\tau\boldsymbol{\beta}_0)(1 - \Phi(\boldsymbol{x}_{ij}^\tau\boldsymbol{\beta}_0))} + |\boldsymbol{\mu}_{ij}(\boldsymbol{\beta}_0) - \boldsymbol{\mu}_{ij}(\boldsymbol{\beta})| \\ &\le \infty \qquad\qquad j = 1, 2..., T. \end{aligned}$$

Verification of the conditions in Theorem 3 for this binary model is similar to that for the continuous model in Section 1.1. In the following, we give some details for $E[h_l(\boldsymbol{\varepsilon}_i, \alpha_l)] < \infty$.

$$\begin{aligned} E[h_l(\boldsymbol{\varepsilon}_i, \alpha_l)] &= E\{\frac{\partial|R^{(l)}(\alpha_l)|^{-1/2}}{\partial\alpha_l}|R^{(l)}(\alpha_l)|^{1/2} + \frac{\partial tr(R^{(l)}(\alpha_l)^{-1}\boldsymbol{\varepsilon_i}\boldsymbol{\varepsilon_i}^\tau)}{\partial\alpha_l}\} \\ &= \{\frac{\partial|R^{(l)}(\alpha_l)|^{-1/2}}{\partial\alpha_l}|R^{(l)}(\alpha_l)|^{1/2} + \frac{\partial tr(R^{(l)}(\alpha_l)^{-1}E[\boldsymbol{\varepsilon_i}\boldsymbol{\varepsilon_i}^\tau])}{\partial\alpha_l}\}, \end{aligned}$$

where $E[\varepsilon_{\mathbf{ij}}\varepsilon_{\mathbf{ik}}] = E[(\boldsymbol{y}_{ij} - \boldsymbol{\mu}_{ij})(\boldsymbol{y}_{ik} - \boldsymbol{\mu}_{ik})] < \infty$.

Most of the verification of the conditions in Theorem 5 for this binary model is similar as in Section 1.2. In the following, we present these with slight difference.

Condition 2 $\partial^2 U_i(\boldsymbol{\beta}, M(\boldsymbol{\psi}))/\partial\boldsymbol{\psi}\partial\boldsymbol{\psi}^\tau = D_i^\tau A_i^{-\frac{1}{2}}[\partial^2 M^{-1}(\boldsymbol{\psi})/\partial\boldsymbol{\psi}\partial\boldsymbol{\psi}^\tau]A_i^{-\frac{1}{2}}(\boldsymbol{y}_i - \boldsymbol{\mu}_i)$ is continuous at $\boldsymbol{\psi}^*$ with probability one and there is a neighborhood $\mathcal{N}$ of $\boldsymbol{\psi}$ such that

$$\begin{aligned} E\left[\sup_{\boldsymbol{\psi}\in\mathcal{N}}|\partial^2 U_i(\boldsymbol{\beta}, M(\boldsymbol{\psi}))/\partial\boldsymbol{\psi}\partial\boldsymbol{\psi}^\tau|\right] &= \sup_{\boldsymbol{\psi}\in\mathcal{N}} E|D_i^\tau A_i^{-\frac{1}{2}}[\partial^2 M^{-1}(\boldsymbol{\psi})/\partial\boldsymbol{\psi}\partial\boldsymbol{\psi}^\tau]A_i^{-\frac{1}{2}}(\boldsymbol{y}_i - \boldsymbol{\mu}_i)| \\ &= |D_i^\tau A_i^{-\frac{1}{2}}[\partial^2 M^{-1}(\boldsymbol{\psi})/\partial\boldsymbol{\psi}\partial\boldsymbol{\psi}^\tau]A_i^{-\frac{1}{2}}|_{\boldsymbol{\psi}=\boldsymbol{\psi}^\mathcal{N}} E|(\boldsymbol{y}_i - \boldsymbol{\mu}_i)| < \infty, \end{aligned}$$

where $|D_i^\tau A_i^{-\frac{1}{2}}[\partial^2 M^{-1}(\boldsymbol{\psi})/\partial\boldsymbol{\psi}\partial\boldsymbol{\psi}^\tau]A_i^{-\frac{1}{2}}|$ has a maximum at $\boldsymbol{\psi}^\mathcal{N}$ in $\mathcal{N}$.

Condition 3 First, it is easy to see that

$$\left[\frac{1}{N}\sum_{i=1}^{N}\frac{\partial U_i\{\boldsymbol{\beta}, M(\boldsymbol{\psi})\}}{\partial \boldsymbol{\beta}}\right]^{-1} = \left[\frac{1}{N}\sum_{i=1}^{N} D_i^\tau A_i^{-\frac{1}{2}} M^{-1}(\boldsymbol{\psi}) A_i^{-\frac{1}{2}} D_i\right]^{-1}.$$

Second, because covariates x_{ik}'s are randomly sampled from $U(-1,1)$, we have

$$\left[\frac{1}{N}\sum_{i=1}^{N} D_i^\tau A_i^{-\frac{1}{2}} M^{-1}(\boldsymbol{\psi}) A_i^{-\frac{1}{2}} D_i\right] = \frac{1}{N}\sum_{i=1}^{N}\sum_{j,k=1}^{T} D_{ij} A_{ij}^{-\frac{1}{2}} A_{ik}^{-\frac{1}{2}} D_{ik} M_{jk} \to E[\sum_{j,k=1}^{T} D_{ij} A_{ij}^{-\frac{1}{2}} A_{ik}^{-\frac{1}{2}} D_{ik} M_{jk}] \ \ a.e.,$$

where M_{jk} denotes the element at the jth row and kth column of $M^{-1}(\boldsymbol{\psi})$ and

$$E[\sum_{j,k=1}^{T} D_{ij} A_{ij}^{-\frac{1}{2}} A_{ik}^{-\frac{1}{2}} D_{ik} M_{jk}] = \sum_{k=1}^{T} E[D_{ik}^2 A_{ik}^{-1}] M_{jk} = \sum_{k=1}^{T} E[x_{jk}^2 \Phi^4(x_{jk}^\tau\beta)(1-\Phi(x_{jk}^\tau\beta))^2] M_{jk}.$$

Because $\Phi^4(x_{jk}^\tau\beta)(1-\Phi(x_{jk}^\tau\beta))^2$ is bounded with probability one, and $E[x_{jk}^2] = \frac{1}{3}$. Therefore, there is a neighborhood $\mathcal{N}$ of $\boldsymbol{\psi}^*$ such that $\left[\frac{1}{N}\sum_{i=1}^{N}\frac{\partial U_i\{\boldsymbol{\beta}, M(\boldsymbol{\psi})\}}{\partial \boldsymbol{\beta}}\right]^{-1}$ is bounded in $\mathcal{N}$ with probability one;

Condition 4 $E[U_i(\boldsymbol{\beta}, M(\boldsymbol{\psi}^*))] = 0$ has a unique solution $\boldsymbol{\beta}_0$, because $E[U_i(\boldsymbol{\beta}, M(\boldsymbol{\psi}^*))] = D_i^\tau A_i^{-\frac{1}{2}} M^{-1}(\boldsymbol{\psi}^*)(\boldsymbol{\mu}_i(\boldsymbol{\beta}) - \boldsymbol{\mu}_i(\boldsymbol{\beta}_0))$ has a solution $\boldsymbol{\beta}_0$ and $\frac{\partial E[U_i(\boldsymbol{\beta}, M(\boldsymbol{\psi}^*))]}{\partial \beta} = D_i^\tau A_i^{-\frac{1}{2}} M^{-1}(\boldsymbol{\psi}^*) A_i^{-\frac{1}{2}} D_i$ is a positive definite matrix.

Condition 6 Because

$$\left[\frac{\partial^2 U_i\{\boldsymbol{\beta}, M(\hat{\boldsymbol{\psi}}(\boldsymbol{\beta}))\}}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}^\tau}\right] = -2D_i^\tau A_i^{-\frac{1}{2}}[\partial M^{-1}(\boldsymbol{\psi})/\partial\boldsymbol{\psi}] A_i^{-\frac{1}{2}} D_i + D_i^\tau A_i^{-\frac{1}{2}}[\partial^2 M^{-1}(\boldsymbol{\psi})/\partial\boldsymbol{\psi}\partial\boldsymbol{\psi}^\tau](\boldsymbol{y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta})),$$

we have

$$\begin{aligned} E\left[\sup_{\boldsymbol{\beta}\in\mathcal{N}}\left|\frac{\partial^2 U_i\{\boldsymbol{\beta}, M(\hat{\boldsymbol{\psi}}(\boldsymbol{\beta}))\}}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}^\tau}\right|\right] &= \sup_{\boldsymbol{\beta}\in\mathcal{N}}\left|2D_i^\tau A_i^{-\frac{1}{2}}[\partial M^{-1}(\boldsymbol{\psi})/\partial\boldsymbol{\psi}] A_i^{-\frac{1}{2}} D_i\right| + E\left[\sup_{\boldsymbol{\beta}\in\mathcal{N}}\left|D_i^\tau A_i^{-\frac{1}{2}}[\frac{\partial^2 M^{-1}(\boldsymbol{\psi})}{\partial\boldsymbol{\psi}\partial\boldsymbol{\psi}^\tau}](\boldsymbol{y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta}))\right|\right] \\ &= \left|2D_i^\tau A_i^{-\frac{1}{2}}[\partial M^{-1}(\boldsymbol{\psi})/\partial\boldsymbol{\psi}] A_i^{-\frac{1}{2}} D_i\right| + \left|D_i^\tau A_i^{-\frac{1}{2}}[\frac{\partial^2 M^{-1}(\boldsymbol{\psi})}{\partial\boldsymbol{\psi}\partial\boldsymbol{\psi}^\tau}]\right| E\left[\sup_{\boldsymbol{\beta}\in\mathcal{N}}|(\boldsymbol{y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta}))|\right]. \end{aligned}$$

Since

$$E\left[\sup_{\boldsymbol{\beta}\in\mathcal{N}}|(\boldsymbol{y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta}))|\right] = E\left[\sup_{\boldsymbol{\beta}\in\mathcal{N}}|(\boldsymbol{y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta}_0) + \boldsymbol{\mu}_i(\boldsymbol{\beta}_0) - \boldsymbol{\mu}_i(\boldsymbol{\beta}))|\right] = E|\boldsymbol{y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta}_0)| + \sup_{\boldsymbol{\beta}\in\mathcal{N}}|\boldsymbol{\mu}_i(\boldsymbol{\beta}_0) - \boldsymbol{\mu}_i(\boldsymbol{\beta})| < \infty,$$

we have $E\left[\sup_{\boldsymbol{\beta}\in\mathcal{N}}\left|\frac{\partial^2 U_i\{\boldsymbol{\beta},M(\hat{\boldsymbol{\psi}}(\boldsymbol{\beta}))\}}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^\tau}\right|\right]<\infty$.

3. Confidence Intervals for Simulation in Section 3.1.

In the following (see Tables 3.1 and 3.2), we present confidence intervals for estimates in Section 3.1 of our manuscript. In these tables, the 'Optimal' column refers to the results from the GEE estimator using the true correlation matrix. Numbers in parentheses are the average length of the 95% confidence intervals.

Table 3.1: Comparison for continuous data generated from three component mixture distributions.

	true parameter						average coverage probabilities of 95% confidence intervals (average length)							
							β_0				β_1			
n	α_1	α_2	α_3	π_1	π_2	π_3	Ind	Optimal	hybrid	mix	Ind	Optimal	hybrid	mix
10	0.7	0.7	0.4	0.3	0.3	0.4	0.890	0.935	0.645	0.893	0.904	0.948	0.610	0.888
							(1.1801)	(1.1207)	(0.7775)	(1.0334)	(0.1665)	(0.1480)	(0.1009)	(0.1349)
50	0.7	0.7	0.4	0.3	0.3	0.4	0.938	0.946	0.911	0.934	0.943	0.941	0.912	0.939
							(0.5009)	(0.4861)	(0.4574)	(0.4773)	(0.0653)	(0.0607)	(0.0571)	(0.0597)
100	0.7	0.7	0.4	0.3	0.3	0.4	0.956	0.960	0.949	0.962	0.934	0.950	0.926	0.945
							(0.3556)	(0.3439)	(0.3338)	(0.3400)	(0.0459)	(0.0429)	(0.0412)	(0.0421)
10	0.4	0.4	0.4	0.3	0.3	0.4	0.905	0.949	0.674	0.897	0.906	0.944	0.672	0.892
							(0.1501)	(0.1543)	(0.1036)	(0.1397)	(0.1501)	(0.1543)	(0.1036)	(0.1397)
50	0.4	0.4	0.4	0.3	0.3	0.4	0.934	0.937	0.914	0.931	0.934	0.945	0.921	0.937
							(0.4535)	(0.4487)	(0.4236)	(0.4420)	(0.0642)	(0.0627)	(0.0589)	(0.0616)
100	0.4	0.4	0.4	0.3	0.3	0.4	0.957	0.965	0.958	0.963	0.936	0.944	0.924	0.944
							(0.3213)	(0.3171)	(0.3069)	(0.3134)	(0.0453)	(0.0443)	(0.0425)	(0.0435)

From Tables 3.1 and 3.2, we see that the mix-GEE estimator outperforms the independent-GEE and hybrid-GEE estimators. Compared to independent-GEE, its average coverage probabilities are almost the same but its average length is smaller. In small sample situations, mix-GEE performs much better than hybrid-GEE in that its average coverage probabilities of confidence intervals are much closer to the nominal confidence level 95%.

Table 3.2: Comparison for continuous data generated from a single correlation structure.

			average coverage probabilities of 95% confidence intervals (average length)							
			$\boldsymbol{\beta}_0$				$\boldsymbol{\beta}_1$			
$R(\alpha)$	n	α	Ind	Optimal	hybrid	mix	Ind	Optimal	hybrid	mix
CS	10	0.4	0.898	0.953	0.672	0.905	0.907	0.957	0.681	0.899
			(0.9247)	(1.0066)	(0.6859)	(0.9184)	(0.0937)	(0.1005)	(0.0680)	(0.0909)
CS	10	0.7	0.902	0.946	0.640	0.902	0.892	0.942	0.630	0.906
			(1.0604)	(1.1292)	(0.7821)	(1.0492)	(0.0714)	(0.0712)	(0.0487)	(0.0653)
AR	10	0.4	0.905	0.949	0.674	0.897	0.906	0.944	0.672	0.892
			(0.1501)	(0.1543)	(0.1036)	(0.1397)	(0.1501)	(0.1543)	(0.1036)	(0.1397)
AR	10	0.7	0.890	0.935	0.645	0.893	0.904	0.948	0.610	0.888
			(1.1801)	(1.1207)	(0.7775)	(1.0334)	(0.1665)	(0.1480)	(0.1009)	(0.1349)
MA	10	0.4	0.917	0.965	0.690	0.903	0.916	0.957	0.687	0.903
			(0.9352)	(0.9570)	(0.6616)	(0.8771)	(0.1460)	(0.1472)	(0.1012)	(0.1350)
CS	50	0.4	0.937	0.945	0.915	0.941	0.953	0.950	0.929	0.948
			(0.4459)	(0.4498)	(0.4218)	(0.4416)	(0.0457)	(0.0449)	(0.0424)	(0.0443)
CS	50	0.7	0.942	0.957	0.910	0.945	0.949	0.952	0.909	0.943
			(0.4989)	(0.5048)	(0.4787)	(0.4929)	(0.0345)	(0.0317)	(0.0300)	(0.0312)
AR	50	0.4	0.932	0.954	0.912	0.942	0.926	0.939	0.908	0.924
			(0.4697)	(0.4552)	(0.4297)	(0.4488)	(0.0722)	(0.0690)	(0.0652)	(0.0680)
AR	50	0.7	0.939	0.964	0.925	0.947	0.931	0.939	0.892	0.931
			(0.5641)	(0.4998)	(0.4739)	(0.4934)	(0.0804)	(0.0658)	(0.0622)	(0.0650)
MA	50	0.4	0.951	0.960	0.917	0.952	0.946	0.950	0.923	0.941
			(0.4464)	(0.4269)	(0.4033)	(0.4219)	(0.0697)	(0.0656)	(0.0619)	(0.0647)
CS	100	0.4	0.951	0.961	0.941	0.954	0.942	0.943	0.923	0.938
			(0.3192)	(0.3181)	(0.3093)	(0.3162)	(0.0324)	(0.0317)	(0.0307)	(0.0314)
CS	100	0.7	0.950	0.951	0.930	0.950	0.946	0.948	0.928	0.946
			(0.3579)	(0.3569)	(0.3462)	(0.3533)	(0.0248)	(0.0224)	(0.0219)	(0.0224)
AR	100	0.4	0.932	0.944	0.927	0.944	0.935	0.938	0.931	0.935
			(0.3337)	(0.3219)	(0.3117)	(0.3182)	(0.0513)	(0.0488)	(0.0471)	(0.0481)
AR	100	0.7	0.940	0.952	0.935	0.939	0.944	0.952	0.938	0.946
			(0.4033)	(0.3534)	(0.3432)	(0.3506)	(0.0574)	(0.0465)	(0.0450)	(0.0461)
MA	100	0.4	0.946	0.947	0.927	0.944	0.943	0.947	0.927	0.944
			(0.3164)	(0.3020)	(0.2931)	(0.2992)	(0.0494)	(0.0464)	(0.0449)	(0.0459)

4. Indonesian Children's Infection Data

In this section, we analyze a respiratory infection data set from Diggle et al (2002), which was also analyzed by Leung et. al. (2009). To study the relationship between respiratory infection and Vitamin A deficiency, a

Table 4.3: Parameter estimates for Indonesian children's infection data.

parameter	ind	PL_{cs}	PL_{ar1}	PL_{ma1}	hybrid	mix
Intercept	-2.667	-2.652	-2.665	-2.667	-2.610	-2.653
Age	-0.022	-0.021	-0.022	-0.022	-0.022	-0.021
Xerophthalmia	0.823	0.718	0.800	0.802	0.929	0.726
Cos(season)	-0.782	-0.782	-0.781	-0.781	-0.733	-0.783
Sex	-0.101	-0.124	-0.103	-0.103	-0.153	-0.125
Height for age	-0.023	-0.025	-0.024	-0.024	-0.025	-0.025

group of 275 Indonesian children are chosen to take a medical examination at 3-month intervals for 18 months. The maximum number of visits equals 6. Due to missing data, we consider only a subset of $N = 122$ children with complete information. The response variable is given by a binary variable indicating whether whether a child suffered from a respiratory infection. We fit a logistic regression using five predictors listed in Table 4.3, including *sex*; *age* in months (centered at 36); *Xerophthalmia*, an ocular manifestation of vitamin A deficiency; *Height for age*, a percentage that indicates longer-term nutritional status; *Cos(season)*, a numerical variable that denotes the seasonality and has value -1, 0, 1, 0, -1 and 0 at 3, 6, 9, 12, 15, and 18 months, respectively. We apply the following methods to this data set, our mix-GEE estimator, the hybrid GEE estimator and the PL-GEE estimator. Here, we consider four different choices, including independent, CS, AR(1) and MA(1). For the hybrid-GEE and the mix-GEE method, three correlation structures, $R^1(\alpha_1) = AR(1)$, $R^2(\alpha_2) = CS$, and $R^3(\alpha_3) = MA(1)$, are considered. Results are shown in Table 4.3. Columns 2-5 are for the naive GEE estimators with different working correlations structures. It is interesting to see that other methods give very similar parameter estimates, but our mix-GEE estimates are quite different for a few predictors. For example, the mix-GEE estimates indicate a much smaller effect of *Xerophthalmia* but a stronger effect of *sex* and *height for age*. The component proportion estimates from our mix-GEE estimator are $\hat{\pi}_{AR} = 0.261$, $\hat{\pi}_{cs} = 0.259$, and $\hat{\pi}_{ma} = 0.480$.