Web Appendix I

Gibbs sampler

We derive all full conditional distributions required for the Gibbs sampling algorithm below. Based on the likelihood (3.2) and the prior for $\boldsymbol{b}_i$ in (3.10), we obtain the distribution of $\boldsymbol{V}_{ij} = (\boldsymbol{Y}_{ij}', \boldsymbol{W}_{ij}')'$,

$$\begin{aligned} f(\boldsymbol{V}|\boldsymbol{\beta}, \boldsymbol{R}_1, \boldsymbol{B}, \boldsymbol{\Sigma}_2^*) &= \prod_{i=1}^{m}\prod_{j=1}^{n_i} f(\boldsymbol{V}_{ij}|\boldsymbol{\beta}, \boldsymbol{\Sigma}_i) \\ &\propto \exp\left\{-\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{n_i}(\boldsymbol{V}_{ij} - \boldsymbol{X}_{ij}\boldsymbol{\beta})'\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{Y}_{ij} - \boldsymbol{X}_{1ij}\boldsymbol{\beta})\right\} \end{aligned} \tag{7.1}$$

which is a multivariate normal with mean $\boldsymbol{X}_{ij}\boldsymbol{\beta}$ and covariance $\boldsymbol{\Sigma}_i$. Hence, the factors $f(\boldsymbol{W}_{mis}|\boldsymbol{\theta}, \boldsymbol{Y}_{obs}, \boldsymbol{Y}_{mis}, \boldsymbol{W}_{obs})$, $f(\boldsymbol{Y}_{mis}|\boldsymbol{\theta}, \boldsymbol{Y}_{obs}, \boldsymbol{W}_{obs})$ and $f(\boldsymbol{Y}_{obs}|\boldsymbol{\theta}, \boldsymbol{Q}_{obs}, \boldsymbol{W}_{obs})$ from (4.1) are multivariate normal, multivariate normal and truncated multivariate normal, respectively. The full conditional distribution of $\boldsymbol{\beta}$ is given as

$$\begin{aligned} \boldsymbol{\beta}|\boldsymbol{V}, \boldsymbol{\Sigma}_1, \cdots, \boldsymbol{\Sigma}_m \sim \mathrm{N}\Bigg(\Big(\sum_{i=1}^{m}\sum_{j=1}^{n_i}\boldsymbol{X}_{ij}'\boldsymbol{\Sigma}_i^{-1}\boldsymbol{X}_{ij}\Big)^{-1}\Big(\sum_{i=1}^{m}\sum_{j=1}^{n_i}\boldsymbol{X}_{ij}'\boldsymbol{\Sigma}_i^{-1}\boldsymbol{V}_{ij}\Big), \\ \Big(\sum_{i=1}^{m}\sum_{j=1}^{n_i}\boldsymbol{X}_{ij}'\boldsymbol{\Sigma}_i^{-1}\boldsymbol{X}_{ij}\Big)^{-1}\Bigg). \end{aligned} \tag{7.2}$$

The correlation matrix $\boldsymbol{R}_{i,11}$ can be generated from

$$\begin{aligned} &\pi(\boldsymbol{R}_{i,11}|\boldsymbol{Y}, \boldsymbol{\beta}_1) \\ \propto&|\boldsymbol{R}_{i,11}|^{-\frac{(n_i-T-1)+T+1}{2}} \exp\Big\{-\frac{1}{2}\mathrm{tr}\Big[\sum_{j=1}^{n_i}\big(\boldsymbol{Y}_{ij} - \boldsymbol{X}_{1ij}\boldsymbol{\beta}_1\big)\big(\boldsymbol{Y}_{ij} - \boldsymbol{X}_{1ij}\boldsymbol{\beta}_1\big)'\boldsymbol{R}_{i,11}^{-1}\Big]\Big\}\times \\ &\mathrm{I}\{r_{i,jk} : r_{i,jk} = 1(j = k), |r_{i,jk}| < 1\ (j \neq k) \text{ and } \boldsymbol{R}_{i,11} \text{ is positive definite}\}, \end{aligned} \tag{7.3}$$

where $r_{i,tl}$ $(t, l = 1, \cdots, T)$ is the element of the tth row and lth column in $\boldsymbol{R}_{i,11}$. This is a constrained inverse Wishart with degrees of freedom $n_i - T - 1$ and scale matrix $(1 + 1/n_i)\sum_{j=1}^{n_i}\big(\boldsymbol{Y}_{ij} - \boldsymbol{X}_{1ij}\boldsymbol{\beta}_1\big)\big(\boldsymbol{Y}_{ij} - \boldsymbol{X}_{1ij}\boldsymbol{\beta}_1\big)'$. How to sample $\boldsymbol{R}_{i,11}$ from this distribution was discussed in Section 4.2.2.

The full conditional distribution of $\boldsymbol{b}_{i\delta}$ is

$$\boldsymbol{b}_{i\delta}|\boldsymbol{Z}_i, \boldsymbol{C}_i, \boldsymbol{\delta}_i, \boldsymbol{\Sigma}^*_{22} \sim \mathrm{N}\Bigg(\left(\sum_{j=1}^{n_i} \boldsymbol{C}'_{ij\delta}\boldsymbol{\Sigma}^{*^{-1}}_{i,22}\boldsymbol{C}_{ij\delta} + \hat{\boldsymbol{\Sigma}}^{-1}_{b_{i\delta}}\right)^{-1} \left(\sum_{j=1}^{n_i} \boldsymbol{C}_{ij\delta}\boldsymbol{\Sigma}^{*^{-1}}_{i,22}\boldsymbol{Z}_{ij} + \hat{\boldsymbol{\Sigma}}^{-1}_{b_{i\delta}}\hat{\boldsymbol{\mu}}_{b_{i\delta}}\right), \left(\sum_{j=1}^{n_i} \boldsymbol{C}'_{ij\delta}\boldsymbol{\Sigma}^{*^{-1}}_{i,22}\boldsymbol{C}_{ij\delta} + \hat{\boldsymbol{\Sigma}}^{-1}_{b_{i\delta}}\right)^{-1}\Bigg), \tag{7.4}$$

where $\hat{\boldsymbol{\mu}}_{b_{i\delta}}$ and $\hat{\boldsymbol{\Sigma}}_{b_{i\delta}}$ were defined the same as in Section 3.2.2, $\boldsymbol{Z}_{ij} = \boldsymbol{W}_{ij} - \boldsymbol{X}_{2ij}\boldsymbol{\beta}_2$, and $\boldsymbol{b}_{i\delta}$ and $\boldsymbol{C}_{ij\delta}$ were defined in Section 3.2.1.

The full conditional distribution of each component $\delta_{i,tl}$ of $\boldsymbol{\delta}_i$ is Bernoulli with probability

$$\pi(\delta_{i,tl} = 1|\boldsymbol{b}_i, p_i, \delta_{(i,tl)}) = \frac{a_{i1}}{a_{i1} + a_{i0}}, \tag{7.5}$$

where $\delta_{(i,tl)}$ denotes a vector consisting of all elements of $\boldsymbol{\delta}_i$ except $\delta_{i,tl}$,

$$a_{i0} = \pi(\boldsymbol{b}_i|\delta_{i,tl} = 0, \delta_{(i,tl)})\pi(\delta_{i,tl} = 0|p_i)$$

and

$$a_{i1} = \pi(\boldsymbol{b}_i|\delta_{i,tl} = 1, \delta_{(i,tl)})\pi(\delta_{i,tl} = 1|p_i)$$

Next, the full conditional distribution of p_i is

$$\pi(p_i|\boldsymbol{Z}_i, \boldsymbol{C}_i, \boldsymbol{b}_i, \boldsymbol{\delta}_i, \boldsymbol{\Sigma}^*_{22}) = \pi(p_i|\boldsymbol{\delta}_i) \propto \prod_{t=1}^{T}\prod_{l=1}^{T}(p_i^{|t-l|^{1/a_0}+1})^{\delta_{i,tl}}(1-p_i^{|t-l|^{1/a_0}+1})^{1-\delta_{i,tl}} p_i^{r_i-1}(1-p_i)^{\lambda_i-1} \tag{7.6}$$

which can be sampled using the Metropolis-Hastings algorithm. Finally, $\boldsymbol{\Sigma}^*_{i,22}$ is sampled from

$$\boldsymbol{\Sigma}^*_{i,22}|\boldsymbol{Z}_i, \boldsymbol{C}_i, \boldsymbol{b}_i, \boldsymbol{\delta}_i \sim \mathrm{IW}(\nu_\Sigma, \boldsymbol{S}^{-1}_\Sigma) \tag{7.7}$$

with

$$\nu_\Sigma = n_i - T - 1 \quad \text{and} \quad \boldsymbol{S}_\Sigma = \sum_{j=1}^{n_i}\big(\boldsymbol{Z}_{ij} - \boldsymbol{C}_{ij}\boldsymbol{b}_i\big)\big(\boldsymbol{Z}_{ij} - \boldsymbol{C}_{ij}\boldsymbol{b}_i\big)'.$$

Web Appendix II

Proof of Proposition 1

Since $\boldsymbol{Y} \sim TN(\boldsymbol{\mu}, \boldsymbol{\Sigma})\mathrm{I}_{U_1}$, then by the property of a TMVN distribution, $\boldsymbol{Z} = \boldsymbol{P}^{-1}\boldsymbol{Y}$ is also distributed as a TMVND with mean $\boldsymbol{P}^{-1}\boldsymbol{\mu}$ and variance $\mathrm{I} = \boldsymbol{P}^{-1}\boldsymbol{\Sigma}\boldsymbol{P}^{-1}$, i.e. $\boldsymbol{Z} \sim TN(\boldsymbol{P}^{-1}\boldsymbol{\mu}, \boldsymbol{I})I_{U_2}$, where $\boldsymbol{P}$ is defined in the statement of Proposition 1 and H is the $\boldsymbol{P}$-transformed version of U_1. H is a convex set because it is transformed by non-singular matrix $\boldsymbol{P}$ from U_1 which is also a convex set. This implies that U_{2t}, the tth ordinate of U_2 $(t = 1, \cdots, T)$, has one of the following three forms: 1) $z \geq a$; 2) $z \leq b$; 3) $a \leq z \leq b$, where a and b are bounded constants. Hence generating $\boldsymbol{Y}$ by Gibbs sampling from $\pi(y_t^{(k)}|y_1^{(k)}, \cdots, y_{t-1}^{(k)}, y_{t+1}^{(k-1)}, \cdots, y_T^{(k-1)})$ with mean and variance in (4.2) is equivalent to first generating $\boldsymbol{Z}$ by Gibbs sampling from $\pi(z_t^{(k)}|z_1^{(k)}, \cdots, z_{t-1}^{(k)}, z_{t+1}^{(k-1)}, \cdots, z_T^{(k-1)})$ with mean ν_t and variance 1, where ν_t is the tth element of $\boldsymbol{P}^{-1}\boldsymbol{\mu}$ and then transforming back to $\boldsymbol{Y}$ via $\boldsymbol{PZ}$. $\square$

Proof of Proposition 2

Extending the idea in Liu (2007) and Liu and Daniels (2006), based on transformation (4.4), we derive the Jacobian. The following identities are needed

$$\frac{\partial(\boldsymbol{Y}_{i1}, \cdots, \boldsymbol{Y}_{in_i})'}{\partial(\psi_{i,21}, \cdots, \psi_{i,T(T-1)})} = 0, \quad \frac{\partial(\boldsymbol{Y}_{i1}, \cdots, \boldsymbol{Y}_{in_i})'}{\partial(b_{i,11}, \cdots, b_{i,TT})} = 0, \quad \frac{\partial(r_{i,21}, \cdots, r_{i,T(T-1)})'}{\partial(b_{i,11}, \cdots, b_{i,TT})} = 0,$$

$$\frac{\partial(\boldsymbol{Y}_{i1}, \cdots, \boldsymbol{Y}_{in_i})'}{\partial(\boldsymbol{Y}^*_{i1}, \cdots, \boldsymbol{Y}^*_{i(n_i-1)})} = \left(\boldsymbol{I}_{n_i-1} \bigotimes \boldsymbol{D}_i^{-1}\right)\boldsymbol{P}_{i1},$$

$$\frac{\partial(\boldsymbol{Y}_{i1}, \cdots, \boldsymbol{Y}_{in_i})'}{\partial(\psi^{-1}_{i,11}, \cdots, \psi^{-1}_{i,TT})} = \boldsymbol{P}_{i2},$$

$$\left|\frac{\partial(r_{i,21}, \cdots, r_{i,T(T-1)})'}{\partial(\psi_{i,21}, \cdots, \psi_{i,T(T-1)})}\right| = \prod_{j=1}^{T}(\psi^{-1}_{i,jj})^{T-1} = |\boldsymbol{D}_i^{-1}|^{T-1},$$

$$\left|\frac{\partial(b_{i,11}, \cdots, b_{i,TT})'}{\partial(b^*_{i,11}, \cdots, b^*_{i,TT})}\right| = |\boldsymbol{D}_i^{-1}|^{-T}$$

$$\left|\frac{\partial(\psi^{-1}_{i,11}, \cdots, \psi^{-1}_{i,TT})'}{\partial(\psi^2_{i,11}, \cdots, \psi^2_{i,TT})}\right| = (-\frac{1}{2})^T \prod_{j=1}^{T}(\psi^{-2}_{i,jj})^{\frac{3}{2}} \propto |\boldsymbol{D}_i^{-1}|^3$$

where $\boldsymbol{P}_{i1}$ and $\boldsymbol{P}_{i2}$ are matrices, not depending on $\boldsymbol{D}_i$. Then the Jacobian is

$$
\begin{aligned}
J =& \left| \frac{\partial(\boldsymbol{Y}_{i1},\cdots,\boldsymbol{Y}_{in_i},\boldsymbol{R}_i,\boldsymbol{B}_i)}{\partial(\boldsymbol{Y}^*_{i1},\cdots,\boldsymbol{Y}^*_{i(n_i-1)},\boldsymbol{\Sigma}_i,\boldsymbol{B}^*_i)} \right|_+ \\
=& \left| \frac{\partial(\boldsymbol{Y}_{i1},\cdots,\boldsymbol{Y}_{in_i},r_{i,21},\cdots,r_{i,T(T-1)},b_{i,11},\cdots,b_{i,TT})'}{\partial(\boldsymbol{Y}^*_{i1},\cdots,\boldsymbol{Y}^*_{i(n_i-1)},\psi^2_{i,11},\cdots,\psi^2_{i,TT},\psi_{i,21},\cdots,\psi_{i,T(T-1)},b^*_{i,11},\cdots,b^*_{i,TT})} \right|_+ \\
\propto& |\boldsymbol{D}_i^{-1}|^3 \left| \begin{matrix} (\boldsymbol{I}_{n_i-1} \bigotimes \boldsymbol{D}_i^{-1})\boldsymbol{P}_{i1} \\ \boldsymbol{P}_{i2} \end{matrix} \right|_+ |\boldsymbol{D}_i^{-1}|^{T-1} |\boldsymbol{D}_i^{-1}|^{-T} \\
=& |\boldsymbol{D}_i^{-1}|^2 \left| \begin{pmatrix} \boldsymbol{I}_{n_i-1} \bigotimes \boldsymbol{D}_i^{-1} & 0 \\ 0 & \boldsymbol{I} \end{pmatrix} \begin{pmatrix} \boldsymbol{P}_{i1} \\ \boldsymbol{P}_{i2} \end{pmatrix} \right|_+ \\
=& |\boldsymbol{D}_i^{-1}|^2 |\boldsymbol{D}_i^{-1}|^{n_i-1} = |\boldsymbol{D}_i^{-1}|^{n_i+1},
\end{aligned}
$$

where T is the number of time points, n_i is the number of subjects in group i, and $|\ |_+$ means the absolute value of the corresponding determinant. Using prior (4.5) with $a_i = 1$, the joint distribution of $\boldsymbol{Y}_i$, $\boldsymbol{R}_i$ and $\boldsymbol{B}_i$ given $\boldsymbol{\beta}$ is

$$
\begin{aligned}
p(\boldsymbol{Y}_i, \boldsymbol{R}_i, \boldsymbol{B}_i|\boldsymbol{\beta}) \propto& |\boldsymbol{R}_i|^{-\frac{n_i+1}{2}} \exp\left\{ -\frac{1}{2} \sum_{j=1}^{n_i} (\boldsymbol{Y}_{ij} - \boldsymbol{X}_{ij}\boldsymbol{\beta})' \boldsymbol{R}_i^{-1} (\boldsymbol{Y}_{ij} - \boldsymbol{X}_{ij}\boldsymbol{\beta}) \right\} \times \\
& \exp\left\{ -\frac{1}{2} \sum_{j=1}^{n_i} (\boldsymbol{Z}_{ij} - \boldsymbol{C}_{ij}\boldsymbol{b}_i)' \boldsymbol{\Sigma}^{*-1}_{i,22} (\boldsymbol{Z}_{ij} - \boldsymbol{C}_{ij}\boldsymbol{b}_i) \right\},
\end{aligned} \tag{A.1}
$$

where $\boldsymbol{Y}_i = (\boldsymbol{Y}_{i1}, \cdots, \boldsymbol{Y}_{in_i})'$ and $\boldsymbol{Z}_{ij}$, $\boldsymbol{C}_{ij}$, and $\boldsymbol{b}_i$ were defined previously. Through the one-to-one mapping $\mathbb{T}:\ (\boldsymbol{Y}_i, \boldsymbol{R}_i, \boldsymbol{B}_i) \rightarrow (\boldsymbol{Y}_i^*, \boldsymbol{\Sigma}_i, \boldsymbol{B}_i^*)$, we have

$$
\begin{aligned}
& p(\boldsymbol{Y}_i^*, \boldsymbol{\Psi}_i, \boldsymbol{B}_i^*|\boldsymbol{\beta}) \propto |\boldsymbol{R}_i|^{-\frac{n_i+1}{2}} \times J \times \\
& \exp\left\{ -\frac{1}{2} \sum_{j=1}^{n_i} \boldsymbol{Y}^{*'}_{ij} \boldsymbol{\Psi}_i^{-1} \boldsymbol{Y}^*_{ij} \right\} \cdot \exp\left\{ -\frac{1}{2} \sum_{j=1}^{n_i} \left(\boldsymbol{Z}_{ij} - \boldsymbol{B}_i^* \boldsymbol{Y}^*_{ij}\right)' \boldsymbol{\Sigma}^{*-1}_{i,22} \left(\boldsymbol{Z}_{ij} - \boldsymbol{B}_i^* \boldsymbol{Y}^*_{ij}\right) \right\}.
\end{aligned} \tag{A.2}
$$

So,

$$
\pi(\boldsymbol{\Psi}_i|\boldsymbol{Y}_i^*, \boldsymbol{B}_i, \boldsymbol{\beta}) \propto p(\boldsymbol{Y}^*, \boldsymbol{\Psi}_i, \boldsymbol{B}_i^*|\boldsymbol{\beta}) \propto |\boldsymbol{\Psi}_i|^{-\frac{\nu_i+T+1}{2}} \exp\left\{ -\frac{1}{2} \boldsymbol{S}_i \boldsymbol{\Psi}_i^{-1} \right\},
$$

where $\nu_i = n_i - T$, $\boldsymbol{S}_i = \sum_{j=1}^{n_i} \boldsymbol{Y}^*_{ij} \boldsymbol{Y}^{*'}_{ij}$, $\boldsymbol{Y}_i^* = (\boldsymbol{Y}^*_{i1}, \cdots, \boldsymbol{Y}^*_{in_i})$ and $\boldsymbol{Y}^*_{ij} = \boldsymbol{D}_i(\boldsymbol{Y}_{ij} - \boldsymbol{X}_{1ij}\boldsymbol{\beta})$. $\square$

Before we give the proof of Theorem 2, we introduce three important lemmas. The first two lemmas come from Marshall and Olkin (1979) and the third one is from Chen and Shao (1999).

Lemma 1. *Assume that* $\boldsymbol{A}_1$ *and* $\boldsymbol{A}_2$ *are two positive definite* $p \times p$ *matrices, then*

$$|\boldsymbol{A}_1 + \boldsymbol{A}_2| \geq |\boldsymbol{A}_1| \geq \lambda_p^p(\boldsymbol{A}_1),$$

where $\lambda_1(\boldsymbol{A}_1) \geq \lambda_2(\boldsymbol{A}_1) \geq \cdots \geq \lambda_p(\boldsymbol{A}_1)$ *are the ordered eigenvalues of* $\boldsymbol{A}_1$.

Lemma 2. *For two positive matrices* $\boldsymbol{A}_1$ *and* $\boldsymbol{A}_2$ *with* p *by* p *dimensions, the following inequality holds:*

$$tr(\boldsymbol{A}_1\boldsymbol{A}_2) \geq \sum_{k=1}^{p} \lambda_{p-k+1}(\boldsymbol{A}_1)\lambda_k(\boldsymbol{A}_2),$$

where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$ *are the ordered eigenvalues of the corresponding matrices.*

Lemma 3. *Let* $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \cdots, \boldsymbol{\theta}_k)$, $\boldsymbol{A}$ *be an* $n \times k$ *matrix. Assume that* $\boldsymbol{A}$ *is of full rank and that there exists a positive vector* $\boldsymbol{a}$ *such that*

$$\boldsymbol{a}'\boldsymbol{A} = 0.$$

Then there exists a constant c *depending only on* $\boldsymbol{A}$ *such that*

$$||\boldsymbol{\theta}|| \leq c||\boldsymbol{v}||$$

whenever

$$\boldsymbol{A\theta} \leq \boldsymbol{v},$$

where $||\cdot||$ *denote the Euclidean norm.*

Theorem for posterior propriety Based on model (3.2) and (3.3), the likelihood function is

$$L(\boldsymbol{\beta}, \boldsymbol{R}_1, \boldsymbol{b}, \boldsymbol{\Sigma}_2^*|\boldsymbol{Q}, \boldsymbol{W}) = \prod_{i=1}^{m}\prod_{j=1}^{n_i}\int_{[\boldsymbol{Y}_{ij}\in\mathbb{Y}_{ij}]} f(\boldsymbol{Y}_{ij}|\boldsymbol{\beta}, \boldsymbol{R}_{i,11}) f(\boldsymbol{W}_{ij}|\boldsymbol{Y}_{ij}, \boldsymbol{\beta}, \boldsymbol{b}_i, \boldsymbol{\Sigma}_{i,22}^*) d\boldsymbol{Y}_{ij},$$

where $\boldsymbol{R}_1$, $\boldsymbol{b}$ and $\boldsymbol{\Sigma}_2^*$ are defined the same as in (3.4), and $\mathbb{Y}_{ij}$ is a region of $\boldsymbol{Y}_{ij}$ such that $\mathbb{Y}_{ij} = \{\boldsymbol{Y}_{ij} : \cap_{t=1}^{T}\big[(Y_{ij,t} > 0)I\{Q_{ij,t} = 1\} + (Y_{ij,t} \leq 0)I\{Q_{ij,t} = 0\}\big]|\boldsymbol{Q}_{ij}\}$.

Considering the hierarchical priors (3.10), (3.8) and (3.9) for $\pi(\boldsymbol{b}_i, \boldsymbol{\delta}_i, p_i|\boldsymbol{\beta}, \boldsymbol{\Sigma}^*_{i,22})$, the joint posterior distribution is given as

$$\pi(\boldsymbol{\beta}, \boldsymbol{R}_1, \boldsymbol{b}, \boldsymbol{\delta}, \boldsymbol{p}, \boldsymbol{\Sigma}^*_2|\boldsymbol{Q}, \boldsymbol{W}) \propto L(\boldsymbol{\beta}, \boldsymbol{R}_1, \boldsymbol{b}, \boldsymbol{\Sigma}^*_2|\boldsymbol{Q}, \boldsymbol{W})\pi(\boldsymbol{\beta}, \boldsymbol{R}_1, \boldsymbol{b}, \boldsymbol{\delta}, \boldsymbol{p}, \boldsymbol{\Sigma}^*_2),$$

where $\boldsymbol{\delta} = (\boldsymbol{\delta}_1, \cdots, \boldsymbol{\delta}_m)'$, $\boldsymbol{p} = (p_1, \cdots, p_m)'$ and $\pi(\boldsymbol{\beta}, \boldsymbol{R}_1, \boldsymbol{b}, \boldsymbol{\delta}, \boldsymbol{p}, \boldsymbol{\Sigma}^*_2) = \prod_{i=1}^m \pi(\boldsymbol{\beta})\pi(\boldsymbol{R}_{i,11})\times$ $\pi(\boldsymbol{b}_i|\boldsymbol{\delta}_i, \boldsymbol{\beta}, \boldsymbol{\Sigma}^*_{i,22})\pi(\boldsymbol{\delta}_i|p_i)\pi(p_i)\pi(\boldsymbol{\Sigma}^*_{i,22})$. Then the posterior distribution is integrable if and only if

$$\Delta = \int L(\boldsymbol{\beta}, \boldsymbol{R}_1, \boldsymbol{b}, \boldsymbol{\Sigma}^*_2|\boldsymbol{Q}, \boldsymbol{W})\pi(\boldsymbol{\beta}, \boldsymbol{R}_1, \boldsymbol{b}, \boldsymbol{\delta}, \boldsymbol{p}, \boldsymbol{\Sigma}^*_2)d\boldsymbol{\beta}d\boldsymbol{R}_1 d\boldsymbol{b}d\boldsymbol{\delta}d\boldsymbol{p}d\boldsymbol{\Sigma}^*_2 < \infty. \quad (7.8)$$

The following theorem gives conditions for Δ to be finite.

Theorem 2. *Let $h_{ij,t} = 1$ if $y_{ij,t} = 0$ and $h_{ij,t} = -1$ if $y_{ij,t} = 1$, where $y_{ij,t}$ is the t-th element of $\boldsymbol{Y}_{ij}$ in (3.2). Write $\boldsymbol{X}^*_{1ij} = (\boldsymbol{x}^{*'}_{1ij}, \cdots, \boldsymbol{x}^{*'}_{1ij})'$ and $\boldsymbol{x}^*_{1ij} = h_{ij,t}\boldsymbol{x}_{1ij}$, where $\boldsymbol{x}'_{1ij}$ is the any row of $\boldsymbol{X}_{1ij}$. Let $I_{[S]}$ denote the indicator function such that $I_{[S]} = 1$ if S is true and 0 otherwise. Assume that the following conditions are satisfied for any i ($i = 1, \cdots, m$)*

- (T_1) *$n_i > T + 1$, where n_i is the number of subjects in group (treatment) i;*
- (T_2) *$\sum_{j=1}^{n_i} \boldsymbol{X}_{1ij}\boldsymbol{X}'_{1ij}$ and $\sum_{j=1}^{n_i} \boldsymbol{X}_{2ij}\boldsymbol{X}'_{2ij}$ are of full rank;*
- (T_3) *$\boldsymbol{X}^*_{1i}$ is of full rank, where $\boldsymbol{X}^*_{1i} = (\boldsymbol{X}^{*'}_{1i1}, \cdots, \boldsymbol{X}^{*'}_{1in_i})'$;*
- (T_4) *There exists a positive constant vector $\boldsymbol{l}$ such that $\boldsymbol{l}'\boldsymbol{X}^*_{1i} = 0$.*

Then the posterior distribution is proper, i.e., $\Delta < \infty$ from (7.8).

Conditions $(T_1) - (T_3)$ are easily verified for a given dataset. Condition (T_4) can be verified using a simple technique discussed in Roy and Hobert (2007). the proof follows.

Proof of Theorem 2

We first focus on the conditional distribution of $\boldsymbol{W}|\boldsymbol{Y}$ and the prior for the association matrix $\boldsymbol{B}$ between $\boldsymbol{W}$ and $\boldsymbol{Y}$. Let $\eta = \int f(\boldsymbol{W}|\boldsymbol{Y}, \boldsymbol{\beta}, \boldsymbol{b}, \boldsymbol{\Sigma}^*_2)\pi(\boldsymbol{b}|\boldsymbol{\delta}, \boldsymbol{\beta}, \boldsymbol{\Sigma}^*_2)d\boldsymbol{\beta}_2 d\boldsymbol{\Sigma}^*_2$ and

$\boldsymbol{Z}_{ij} = \boldsymbol{W}_{ij} - \boldsymbol{C}_{ij}\boldsymbol{b}_i$. Integrating out $\boldsymbol{\beta}_2$, we have

$$\begin{aligned}
\eta_1 =& \int f(\boldsymbol{W}|\boldsymbol{Y}, \boldsymbol{\beta}_2, \boldsymbol{b}, \boldsymbol{\Sigma}_2^*)\pi(\boldsymbol{b}|\boldsymbol{\delta}, \boldsymbol{\beta}, \boldsymbol{\Sigma}_2^*)d\boldsymbol{\beta}_2 \\
=&(2\pi)^{-\frac{\sum_{i=1}^{c} n_i T}{2}} \int \prod_{i=1}^{m} |\boldsymbol{\Sigma}_{i,22}^*|^{-\frac{n_i}{2}} \exp\Big\{ -\frac{1}{2}\sum_{i}^{m}\sum_{j=1}^{n_i} \Big(\boldsymbol{Z}_{ij} - \boldsymbol{X}_{2ij}\boldsymbol{\beta}_2\Big)'\boldsymbol{\Sigma}_{i,22}^{*^{-1}}\Big(\boldsymbol{Z}_{ij} - \boldsymbol{X}_{2ij}\boldsymbol{\beta}_2\Big)\Big\} d\boldsymbol{\beta}_2 \\
=&(2\pi)^{-\frac{\sum_{i=1}^{m} n_i T}{2}} \int \prod_{i=1}^{m} |\boldsymbol{\Sigma}_{i,22}^*|^{-\frac{n_i}{2}} \exp\Big\{ -\frac{1}{2}\sum_{i}^{m}\sum_{j=1}^{n_i} \boldsymbol{Z}_{ij}'\boldsymbol{\Sigma}_{i,22}^{*}{}^{-1}\boldsymbol{Z}_{ij}\Big\} \\
& \exp\Big\{ -\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{n_i} \Big(\boldsymbol{\beta}_2'\boldsymbol{X}_{2ij}'\boldsymbol{\Sigma}_{i,22}^{*^{-1}}\boldsymbol{X}_{2ij}\boldsymbol{\beta}_2 - 2\boldsymbol{\beta}_2'\boldsymbol{X}_{2ij}'\boldsymbol{\Sigma}_{i,22}^{*^{-1}}\boldsymbol{Z}_{ij}\Big)\Big\} d\boldsymbol{\beta}_2 \\
=&(2\pi)^{-\frac{\sum_{i=1}^{m} n_i T - p}{2}} \prod_{i=1}^{m} |\boldsymbol{\Sigma}_{i,22}^*|^{-\frac{n_i}{2}} \Big|\sum_{i=1}^{m}\sum_{j=1}^{n_i} \boldsymbol{X}_{2ij}'\boldsymbol{\Sigma}_{22}^{*^{-1}}\boldsymbol{X}_{2ij}\Big|^{-\frac{1}{2}} \exp\Big\{ -\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{n_i} \boldsymbol{Z}_{ij}'\boldsymbol{\Sigma}_{i,22}^{*^{-1}}\boldsymbol{Z}_{ij}\Big\} \\
& \exp\Big\{\frac{1}{2}\Big(\sum_{i=1}^{m}\sum_{j=1}^{n_i} \boldsymbol{X}_{2ij}'\boldsymbol{\Sigma}_{i,22}^{*^{-1}}\boldsymbol{Z}_{ij}\Big)'\Big(\sum_{i=1}^{c}\sum_{j=1}^{n_i} \boldsymbol{X}_{2ij}'\boldsymbol{\Sigma}_{i,22}^{*^{-1}}\boldsymbol{X}_{2ij}\Big)^{-1}\Big(\sum_{i=1}^{m}\sum_{j=1}^{n_i} \boldsymbol{X}_{2ij}'\boldsymbol{\Sigma}_{i,22}^{*^{-1}}\boldsymbol{Z}_{ij}\Big)\Big\} \\
=&(2\pi)^{-\frac{\sum_{i=1}^{m} n_i T - p}{2}} \prod_{i=1}^{m} |\boldsymbol{\Sigma}_{i,22}^*|^{-\frac{n_i}{2}} \Big|\sum_{i=1}^{m}\sum_{j=1}^{n_i} \boldsymbol{X}_{2ij}'\boldsymbol{\Sigma}_{22}^{*^{-1}}\boldsymbol{X}_{2ij}\Big|^{-\frac{1}{2}} \\
& \exp\Big\{ -\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{n_i} \Big(\boldsymbol{Z}_{ij} - \boldsymbol{X}_{2ij}\hat{\boldsymbol{\beta}}_2\Big)'\boldsymbol{\Sigma}_{i,22}^{*^{-1}}\Big(\boldsymbol{Z}_{ij} - \boldsymbol{X}_{2ij}\hat{\boldsymbol{\beta}}_2\Big)\Big\} \\
=&(2\pi)^{-\frac{\sum_{i=1}^{m} n_i T - p}{2}} \prod_{i=1}^{m} |\boldsymbol{\Sigma}_{i,22}^*|^{-\frac{n_i}{2}} \Big|\sum_{i=1}^{m}\sum_{j=1}^{n_i} \boldsymbol{X}_{2ij}'\boldsymbol{\Sigma}_{22}^{*^{-1}}\boldsymbol{X}_{2ij}\Big|^{-\frac{1}{2}} \\
& \exp\Big\{ -\frac{1}{2}\mathrm{tr}\Big[\sum_{i=1}^{m}\sum_{j=1}^{n_i} \Big(\boldsymbol{Z}_{ij} - \boldsymbol{X}_{2ij}\hat{\boldsymbol{\beta}}_2\Big)\Big(\boldsymbol{Z}_{ij} - \boldsymbol{X}_{2ij}\hat{\boldsymbol{\beta}}_2\Big)'\boldsymbol{\Sigma}_{i,22}^{*^{-1}}\Big]\Big\} \\
=&(2\pi)^{-\frac{\sum_{i=1}^{m} n_i T - p}{2}} \Big|\sum_{i=1}^{m}\sum_{j=1}^{n_i} \boldsymbol{X}_{2ij}'\boldsymbol{\Sigma}_{22}^{*^{-1}}\boldsymbol{X}_{2ij}\Big|^{-\frac{1}{2}} \prod_{i=1}^{m}\Big[|\boldsymbol{\Sigma}_{i,22}^*|^{-\frac{n_i}{2}} \exp\Big\{ -\frac{1}{2}\mathrm{tr}\Big[\boldsymbol{S}_i\boldsymbol{\Sigma}_{i,22}^{*^{-1}}\Big]\Big\}\Big],
\end{aligned} \tag{7.9}$$

where $\hat{\boldsymbol{\beta}}_2 = \Big(\sum_{i=1}^{m}\sum_{j=1}^{n_i} \boldsymbol{X}_{2ij}'\boldsymbol{\Sigma}_{22}^{*^{-1}}\boldsymbol{X}_{2ij}\Big)^{-1}\Big(\sum_{i=1}^{m}\sum_{j=1}^{n_i} \boldsymbol{X}_{2ij}'\boldsymbol{\Sigma}_{22}^{*^{-1}}\boldsymbol{Z}_{ij}\Big)$ and $\boldsymbol{S}_i = \sum_{j=1}^{n_i}\Big(\boldsymbol{Z}_{ij} - \boldsymbol{X}_{2ij}\hat{\boldsymbol{\beta}}_2\Big)\Big(\boldsymbol{Z}_{ij} - \boldsymbol{X}_{2ij}\hat{\boldsymbol{\beta}}_2\Big)'$. Note that here $\boldsymbol{S}_i$ is the function of $\boldsymbol{\Sigma}_{i,22}^*$, $\boldsymbol{\beta}_1$ and $\boldsymbol{B}_i$. By Lemma 1, we obtain

$$\Big|\sum_{i=1}^{m}\sum_{j=1}^{n_i} \boldsymbol{X}_{2ij}'\boldsymbol{\Sigma}_{22}^{*^{-1}}\boldsymbol{X}_{2ij}\Big|^{-\frac{1}{2}} \le \lambda_p^{-\frac{p}{2}} \le \tau_0^{-\frac{p}{2}},$$

where λ_p is the smallest eigenvalue of $\sum_{i=1}^{m}\sum_{j=1}^{n_i} \boldsymbol{X}_{2ij}'\boldsymbol{\Sigma}_{22}^{*^{-1}}\boldsymbol{X}_{2ij}$ and τ_0 is some positive value such that $\lambda_p \geq \tau_0 > 0$. By Lemma 2,

$$\mathrm{tr}\Big(\boldsymbol{S}_i\boldsymbol{\Sigma}_{i,22}^{*^{-1}}\Big) \geq \sum_{k=1}^{T}\lambda_{p-k+1}(\boldsymbol{S}_i)\lambda_k(\boldsymbol{\Sigma}_{i,22}^{*^{-1}}) \geq \lambda_{min}(\boldsymbol{S}_i)\sum_{k=1}^{T}\lambda_k(\boldsymbol{\Sigma}_{i,22}^{*^{-1}}) \geq \tau_i\sum_{k=1}^{T}\lambda_k(\boldsymbol{\Sigma}_{i,22}^{*^{-1}}),$$

where $\tau_i = \min_{\boldsymbol{\Sigma}_{i,22}^{*^{-1}},\boldsymbol{\beta}_1,\boldsymbol{B}_i}\lambda_{min}(\boldsymbol{S}_i)$. From (T_1) and (T_2), we have $n_i > T$ and $\boldsymbol{S}_i$ is positive definite with $0 < |\boldsymbol{S}_i| < \infty$. Hence under (T_1) and (T_2), it is obvious to see that $0 < \lambda_{min}(\boldsymbol{S}_i) < \infty$ and $0 < \tau_i < \infty$.

Then,

$$\begin{aligned}
&\exp\Big\{-\frac{1}{2}\mathrm{tr}\Big(\boldsymbol{S}_i\boldsymbol{\Sigma}_{i,22}^{*^{-1}}\Big)\Big\} \leq \exp\Big\{-\frac{1}{2}\tau_i\sum_{k=1}^{T}\lambda_k(\boldsymbol{\Sigma}_{i,22}^{*^{-1}})\Big\}\\
&= \exp\Big\{-\frac{1}{2}\tau_i\mathrm{tr}\Big(\boldsymbol{\Sigma}_{i,22}^{*^{-1}}\Big)\Big\} = \exp\Big\{-\frac{1}{2}\mathrm{tr}\Big(\tau_i\boldsymbol{I}_{T\times T}\boldsymbol{\Sigma}_{i,22}^{*^{-1}}\Big)\Big\} \qquad (7.10)\\
&= \exp\Big\{-\frac{1}{2}\mathrm{tr}\Big(\sum_{j=1}^{n_i}\boldsymbol{z}_{ij}^*\boldsymbol{z}_{ij}^{*'}\boldsymbol{\Sigma}_{i,22}^{*^{-1}}\Big)\Big\} = \exp\Big\{-\frac{1}{2}\sum_{j=1}^{n_i}\boldsymbol{z}_{ij}^{*'}\boldsymbol{\Sigma}_{i,22}^{*^{-1}}\boldsymbol{z}_{ij}^*\Big\},
\end{aligned}$$

where $\boldsymbol{z}_{ij}^*$ is an any constant vector satisfying $\sum_{j=1}^{n_i}\boldsymbol{z}_{ij}^*\boldsymbol{z}_{ij}^{*'} = \tau_i\boldsymbol{I}_{T\times T}$.

Substituting (7.10) into (7.9), we have:

$$\begin{aligned}
\eta_1 &= \int f(\boldsymbol{W}|\boldsymbol{Y},\boldsymbol{\beta},\boldsymbol{b},\boldsymbol{\Sigma}_2^*)d\boldsymbol{\beta}_2\\
&\leq (2\pi)^{-\frac{\sum_{i=1}^{c}n_iT}{2}}\tau_0^{-\frac{p}{2}}\prod_{i=1}^{m}\Big[|\boldsymbol{\Sigma}_{i,22}^*|^{-\frac{n_i}{2}}\exp\Big\{-\frac{1}{2}\sum_{j=1}^{n_i}\boldsymbol{z}_{ij}^{*'}\boldsymbol{\Sigma}_{i,22}^{*^{-1}}\boldsymbol{z}_{ij}^*\Big\}\Big]\\
&= (2\pi)^{-\frac{\sum_{i=1}^{c}n_iT}{2}}\tau_0^{-\frac{p}{2}}\prod_{i=1}^{m}\Big[|\boldsymbol{\Sigma}_{i,22}^*|^{-\frac{(n_i-T-1)+T+1}{2}}\exp\Big\{-\frac{1}{2}\mathrm{tr}\Big[\Big(\sum_{j=1}^{n_i}\boldsymbol{z}_{ij}^*\boldsymbol{z}_{ij}^{*'}\Big)\boldsymbol{\Sigma}_{i,22}^{*^{-1}}\Big]\Big\}\Big].
\end{aligned}$$

Then,

$$\begin{aligned}
&\eta = \int\eta_1 d\boldsymbol{\Sigma}_2^* \leq\\
&(2\pi)^{-\frac{\sum_{i=1}^{c}n_iT}{2}}\lambda_p^{-\frac{p}{2}}\int\prod_{i=1}^{m}\Big[|\boldsymbol{\Sigma}_{i,22}^*|^{-\frac{(n_i-T-1)+T+1}{2}}\exp\Big\{-\frac{1}{2}\mathrm{tr}\Big[\Big(\sum_{j=1}^{n_i}\boldsymbol{z}_{ij}^*\boldsymbol{z}_{ij}^{*'}\Big)\boldsymbol{\Sigma}_{i,22}^{*^{-1}}\Big]\Big\}\Big]d\boldsymbol{\Sigma}_2^*\\
&\leq M_0\prod_{i=1}^{m}\Big|\sum_{j=1}^{n_i}\boldsymbol{z}_{ij}^*\boldsymbol{z}_{ij}^{*'}\Big|^{-\frac{n_i-T-1}{2}} \qquad \Big(\text{here } \frac{n_i-T-1}{2} \geq 0\ (i=1,\cdots,m) \text{ based on } (T_1)\Big).
\end{aligned}$$

So η is bounded from above by some constant, say M_1, i.e.

$$\eta \leq M_1. \qquad (7.11)$$

By (7.8) and (7.11),

$$\Delta \le M_1 \int \Big[\int_{[\boldsymbol{Y}\in \boldsymbol{Y}_Q]} f(\boldsymbol{Y}|\boldsymbol{\beta}_1, \boldsymbol{R}_{11}) d\boldsymbol{\beta}_1 d\boldsymbol{R}_1 d\boldsymbol{Y} \Big] \pi(\boldsymbol{b}|\boldsymbol{\delta})\pi(\boldsymbol{\delta}|\boldsymbol{p})\pi(\boldsymbol{p}) d\boldsymbol{b} d\boldsymbol{\delta} d\boldsymbol{p}$$
$$= M_1 \int \int_{[\boldsymbol{Y}\in \boldsymbol{Y}_Q]} f(\boldsymbol{Y}|\boldsymbol{\beta}_1, \boldsymbol{R}_1) d\boldsymbol{\beta}_1 d\boldsymbol{R}_1 d\boldsymbol{Y} \int \pi(\boldsymbol{b}|\boldsymbol{\delta})\pi(\boldsymbol{\delta}|\boldsymbol{p})\pi(\boldsymbol{p}) d\boldsymbol{b} d\boldsymbol{\delta} d\boldsymbol{p}.$$

Let η^* and η^{**} denote $\int \pi(\boldsymbol{b}|\boldsymbol{\delta})\pi(\boldsymbol{\delta}|\boldsymbol{p})\pi(\boldsymbol{p}) d\boldsymbol{b} d\boldsymbol{\delta} d\boldsymbol{p}$ and $\int \int_{[\boldsymbol{Y}\in \boldsymbol{Y}_Q]} f(\boldsymbol{Y}|\boldsymbol{\beta}_1, \boldsymbol{R}_1) d\boldsymbol{\beta}_1 d\boldsymbol{R}_1 d\boldsymbol{Y}$, respectively. It is obvious that

$$\eta^* = \int \pi(\boldsymbol{b}|\boldsymbol{\delta})\pi(\boldsymbol{\delta}|\boldsymbol{p})\pi(\boldsymbol{p}) d\boldsymbol{b} d\boldsymbol{\delta} d\boldsymbol{p} < \infty.$$

Since $\pi(\boldsymbol{b}|\boldsymbol{\delta})$, $\pi(\boldsymbol{\delta}|p)$ and $\pi(p)$ are proper, so is $\pi(\boldsymbol{b}, \boldsymbol{\delta}, p)$. Now we just need to show that η^{**} is integrable, i.e. $\eta^{**} < \infty$.

Let $\boldsymbol{Y}_{ij}^* = (Y_{ij1}^*, \cdots, Y_{ijT}^*)$ be independent random variables such that

$$\boldsymbol{Y}_{ij}^* | \boldsymbol{R}_{i,11} \sim \mathrm{N}(\boldsymbol{0}, \boldsymbol{R}_{i,11});$$

that is, given $\boldsymbol{R}_{i,11}$, $\boldsymbol{Y}_{ij}^*$ is normally distributed with mean $\boldsymbol{0}$ and variance $\boldsymbol{R}_{i,11}$. Let $S = \prod_{i=1}^m \prod_{j=1}^{n_i} S_{ij}$ denote the set $[\boldsymbol{Y} \in \mathbf{Y}_Q]$, and put

$$S_{ij} = S_{ij1} \times S_{ij2} \times \cdots \times S_{ijT},$$

where $S_{ijt} = [Y_{ijt} \in Y_{Q_{ijt}}]$ $(t = 1, \cdots, T)$. Then

$$\begin{aligned} \eta^{**} &= \int \int_{[\boldsymbol{Y}\in \boldsymbol{Y}_Q]} f(\boldsymbol{Y}|\boldsymbol{\beta}_1, \boldsymbol{R}_1) d\boldsymbol{\beta}_1 d\boldsymbol{R}_1 d\boldsymbol{Y} \\ &= \int \Big[\prod_{i=1}^m \prod_{j=1}^{n_i} \int_{S_{ij1}} \cdots \int_{S_{ijT}} f(\boldsymbol{Y}|\boldsymbol{\beta}_1, \boldsymbol{R}_1) d\boldsymbol{Y}_i \Big] d\boldsymbol{\beta}_1 d\boldsymbol{R}_1 \qquad (7.12) \\ &= \int L(\boldsymbol{\beta}_1, \boldsymbol{R}_1 | \boldsymbol{Y}) d\boldsymbol{\beta}_1 d\boldsymbol{R}_1, \end{aligned}$$

where $L(\boldsymbol{\beta}_1, \boldsymbol{R}_1|\boldsymbol{Y})$ is the likelihood function associated with (3.2). We can rewrite it as

$$L(\boldsymbol{\beta}_1, \boldsymbol{R}_1|\boldsymbol{Y}) = \mathrm{E}\{I_{[\boldsymbol{Y}_{ij}^* + \boldsymbol{X}_{1ij}\boldsymbol{\beta}_1 \in S_{ij}, 1\le i\le m, 1\le j\le n_i]}\} = \mathrm{E}\{I_{[Y_{ijt}^* + \boldsymbol{x}_{1ijt}\boldsymbol{\beta}_1 \in S_{ijt}, 1\le t\le T, 1\le i\le m, 1\le j\le n_i]}\}. \qquad (7.13)$$

It is easy to see that

$$
\begin{aligned}
&[Y^*_{ijt} + \boldsymbol{x}_{1ijt}\boldsymbol{\beta}_1 \in S_{ijt}, 1 \le t \le T, 1 \le i \le m, 1 \le j \le n_i] \\
=&\bigcap_{1\le t\le T} \Big([Y^*_{ijt} + \boldsymbol{x}_{1ijt}\boldsymbol{\beta}_1 < 0, Y_{ijt} = 0, 1 \le i \le m, 1 \le j \le n_i] \cup \\
&\qquad [Y^*_{ijt} + \boldsymbol{x}_{1ijt}\boldsymbol{\beta}_1 \ge 0, Y_{ijt} = 1, 1 \le i \le m, 1 \le j \le n_i] \Big) \\
\subset& [h_{ijt}\boldsymbol{x}_{1ijt}\boldsymbol{\beta}_1 \le -h_{ijt}Y_{ijt}, 1 \le t \le T, 1 \le i \le m, 1 \le j \le n_i] = [\boldsymbol{X}^*_1\boldsymbol{\beta}_1 \le \boldsymbol{Y}^*],
\end{aligned}
\tag{7.14}
$$

where $\boldsymbol{X}^*_1 = (\boldsymbol{X}'_{1,11}, \cdots, \boldsymbol{X}'_{1,mn_m})'$, $\boldsymbol{Y}^* = (\boldsymbol{Y}^*_{11}, \cdots, \boldsymbol{Y}^*_{mn_m})$ and $\boldsymbol{Y}^*_{ij} = (Y^*_{ij1}, \cdots, Y^*_{ijT})$. Combining (7.12), (7.13), and (7.14), we have

$$
\eta^{**} = \int \mathrm{E}\{I_{[\boldsymbol{X}^*_1\boldsymbol{\beta}_1 \le \boldsymbol{Y}^*]}\} d\boldsymbol{\beta}_1 d\boldsymbol{R}_1.
$$

Therefore, by (T3), (T4) and Lemma 3

$$
\begin{aligned}
\eta^{**} &\le \int \mathrm{E}\{I_{[||\boldsymbol{\beta}_1|| \le c||\boldsymbol{Y}^*||]}\} d\boldsymbol{\beta}_1 d\boldsymbol{R}_1 \le c \int \mathrm{E}\Big\{ \max_{1\le i\le m, 1\le j\le n_i} ||\boldsymbol{Y}^*_{ij}|| \Big\}^p d\boldsymbol{R}_1 \\
&\le c \int \sum_{i=1}^{m} \sum_{j=1}^{n_i} \sum_{t=1}^{T} \mathrm{E}|Y^*_{ijt}|^p d\boldsymbol{R}_1 \le c \int \sum_{t=1}^{T} \mathrm{E}(1) d\boldsymbol{R}_1 \\
&< \infty,
\end{aligned}
$$

where $\boldsymbol{R}_1 = (\boldsymbol{R}'_{1,11}, \cdots, \boldsymbol{R}'_{m,11})'$. □

Proof of Theorem 1

By the candidate transformation $\mathbb{T}$ given in (4.4), and (A.1) and (A.2) in the proof of Proposition 2, drawing $\boldsymbol{Y}_i$, $\boldsymbol{R}_i$ and $\boldsymbol{B}_i$ given $\boldsymbol{\beta}$ is equivalent to drawing $\boldsymbol{Y}^*_i$, $\boldsymbol{\Psi}_i$ and $\boldsymbol{B}^*_i$ first, and then translating back to $\boldsymbol{Y}_i$, $\boldsymbol{R}_i$ and $\boldsymbol{B}_i$ through $\mathbb{T}$. $\boldsymbol{R}_i = \boldsymbol{D}_i^{-1}\boldsymbol{\Psi}_i\boldsymbol{D}_i^{-1}$ will be used as the candidate value. It is accepted in the Metropolis Hastings step with probability α_i. The α_i can be derived as follows.

Let π_1 denote the prior or full conditional distribution of $\boldsymbol{R}$ and π_2 denote the corresponding *candidate prior* or proposal density. Then

$$
\pi_1(\boldsymbol{R}_i|\boldsymbol{Y}_i, \boldsymbol{\beta}) \propto \pi_1(\boldsymbol{R}_i) f(\boldsymbol{Y}_i|\boldsymbol{\beta}, \boldsymbol{R}_i)
$$

and

$$\pi_2(\boldsymbol{R}_i|\boldsymbol{Y}_i,\boldsymbol{\beta}) \propto \pi_2(\boldsymbol{R}_i) f(\boldsymbol{Y}_i|\boldsymbol{\beta},\boldsymbol{R}_i),$$

where $\pi_1(\boldsymbol{R}_i)$, $\pi_2(\boldsymbol{R}_i)$ and $f(\boldsymbol{Y}_i|\boldsymbol{\beta},\boldsymbol{R}_i)$ are given by (3.7), (4.5) and (3.2), respectively. The probability of acceptance at iteration $k+1$ is

$$\begin{aligned}
\alpha_i &= \min\left\{1, \frac{\pi_1(\boldsymbol{R}_i|\boldsymbol{Y}_i,\boldsymbol{\beta})\pi_2(\boldsymbol{R}_i^{(k)}|\boldsymbol{Y}_i,\boldsymbol{\beta})}{\pi_1(\boldsymbol{R}_i^{(k)}|\boldsymbol{Y}_i,\boldsymbol{\beta})\pi_2(\boldsymbol{R}_i|\boldsymbol{Y}_i,\boldsymbol{\beta})}\right\} = \min\left\{1, \frac{\pi_1(\boldsymbol{R}_i)\pi_2(\boldsymbol{R}_i^{(k)})}{\pi_1(\boldsymbol{R}_i^{(k)})\pi_2(\boldsymbol{R}_i)}\right\} \\
&= \min\left\{1, \frac{\pi_2(\boldsymbol{R}_i^{(k)})}{\pi_2(\boldsymbol{R}_i)}\right\} = \min\left\{1, \frac{|\boldsymbol{R}_i^{(k)}|^{-\frac{a_i}{2}}}{|\boldsymbol{R}_i|^{-\frac{a_i}{2}}}\right\} \\
&= \min\left\{1, \exp\left(\frac{a_i}{2}(\log|\boldsymbol{R}_i| - \log|\boldsymbol{R}_i^{(k)}|)\right)\right\},
\end{aligned}$$

where $a_i = 1$. □

Web Appendix III

We conducted several simulations to examine the efficiency of our algorithm to sample from a truncated multivariate normal distribution. In the following, we assume that $\boldsymbol{Y}$ has a truncated normal distribution with mean $\mathbf{0}$ and covariance $\boldsymbol{\Sigma}$. For one of the simulations, we assumed the truncation region was $U_1 = \{\boldsymbol{y} \in \mathcal{C}^4 : y_1 > 0, y_2 < 0, y_3 < 0, y_4 > 0\}$ and $\boldsymbol{\Sigma} = \begin{pmatrix} 15 & 7 & 2 & 0 \\ 7 & 5 & 1.34 & 0.14 \\ 2 & 1.34 & 1 & 0.13 \\ 0 & 0.14 & 0.13 & 0.2 \end{pmatrix}$.

We ran two chains of 2,000 iterations using the LD-A and the R-A, respectively to sample from the distribution of $\boldsymbol{Y}$. Estimates of the lag-n autocorrelation are presented in Table 1. The results show that the decay of autocorrelation of each element of $\boldsymbol{Y}$ is much faster in the LD-A than in the R-A. Similar results were obtained using other choices for $\boldsymbol{\Sigma}$ and U_1.

Table 1: Lag-n autocorrelation estimates of Y_1, Y_2, Y_3 and Y_4 drawn from the TMVN distribution using the LD-A and the R-A

Lag	LD-A*				R-A*			
	Y_1	Y_2	Y_3	Y_4	Y_1	Y_2	Y_3	Y_4
1	0.10	0.07	-0.05	0.01	0.98	0.97	0.97	0.95
2	0.03	-0.04	0.02	-0.00	0.94	0.90	0.92	0.89
3	-0.01	0.01	-0.04	0.02	0.88	0.86	0.87	0.83
4	0.02	-0.03	0.00	-0.00	0.82	0.79	0.77	0.77
5	-0.00	-0.02	-0.02	0.03	0.77	0.76	0.75	0.73
6	-0.03	0.00	0.03	-0.00	0.72	0.70	0.70	0.69
7	0.04	-0.00	-0.00	0.02	0.69	0.68	0.67	0.67
10	-0.00	0.01	-0.02	-0.00	0.65	0.62	0.64	0.62
15	0.00	0.00	-0.00	0.01	0.50	0.47	0.47	0.45
20	0.01	-0.01	0.00	0.00	0.42	0.40	0.42	0.38

∗ LD-A: The algorithm in Section 5; R-A: The algorithm from Robert (1995)

Web Appendix IV

Here we provide details on computation of components of the DIC for the joint models. Note that the kernel of the DIC can be written as

$$\text{DIC} = -4\text{E}_{\boldsymbol{\theta}|\boldsymbol{Q}_{obs},\boldsymbol{W}_{obs}}(\log\, f(\boldsymbol{Q}_{obs},\boldsymbol{W}_{obs}|\boldsymbol{\theta})) + 2\log\, f(\boldsymbol{Q}_{obs},\boldsymbol{W}_{obs}|\overline{\boldsymbol{\theta}}). \tag{7.15}$$

Details on the computation of $f(\boldsymbol{Q}_{obs},\boldsymbol{W}_{obs}|\overline{\boldsymbol{\theta}})$ and $E_{\boldsymbol{\theta}|\boldsymbol{Q}_{obs},\boldsymbol{W}_{obs}}(\log\, f(\boldsymbol{Q}_{obs},\boldsymbol{W}_{obs}|\boldsymbol{\theta}))$ are given below.

To compute $f(\boldsymbol{Q}_{obs},\boldsymbol{W}_{obs}|\overline{\boldsymbol{\theta}})$, we need to evaluate the following integral,

$$f(\boldsymbol{Q}_{obs},\boldsymbol{W}_{obs}|\overline{\boldsymbol{\theta}}) = \int_{A_{\boldsymbol{Q}_{obs}}} \frac{1}{\sqrt{2\pi}|\boldsymbol{\Sigma}_{obs}|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(\boldsymbol{V}_{obs}-\boldsymbol{X}_{obs}\boldsymbol{\beta})'\boldsymbol{\Sigma}_{obs}^{-1}(\boldsymbol{V}_{obs}-\boldsymbol{X}_{obs}\boldsymbol{\beta})\right\} d\boldsymbol{Y}_{obs}, \tag{7.16}$$

where $A_{\boldsymbol{Q}_{obs}}$ is the truncation region of $\boldsymbol{Y}_{obs}$, $\boldsymbol{V}_{obs} = (\boldsymbol{Y}_{obs},\boldsymbol{W}_{obs})$, $\boldsymbol{\Sigma}_{obs}$ is the covariance matrix of $\boldsymbol{V}_{obs}$ and $\boldsymbol{X}_{obs}$ is the matrix composed of the rows in the design matrix in model (3.1). We can calculate this integral using Monte Carlo integration.

Next, we address the computation of the first term. Assume that $\boldsymbol{\theta}^{(g)}$ is a sampled value from the posterior distribution $\pi(\boldsymbol{\theta}|\boldsymbol{Q}_{obs}, \boldsymbol{W}_{obs})$. Then $\mathrm{E}_{\boldsymbol{\theta}|\boldsymbol{Q}_{obs},\boldsymbol{W}_{obs}}(\log\ f(\boldsymbol{Q}_{obs}, \boldsymbol{W}_{obs}|\boldsymbol{\theta}))$ can be estimated as

$$\hat{\mathrm{E}}_{\boldsymbol{\theta}|\boldsymbol{Q}_{obs},\boldsymbol{W}_{obs}}(\log\ f(\boldsymbol{Q}_{obs}, \boldsymbol{W}_{obs}|\boldsymbol{\theta})) = \frac{1}{G}\sum_{g=1}^{G} \log f(\boldsymbol{Q}_{obs}, \boldsymbol{W}_{obs}|\boldsymbol{\theta}^{(g)}).$$

We can use importance sampling techniques to compute each term in the sum as follows. We have

$$f(\boldsymbol{Q}_{obs}, \boldsymbol{W}_{obs}|\boldsymbol{\theta}^{(g)}) = \int_{A_{\boldsymbol{Q}_{obs}}} f(\boldsymbol{Y}_{obs}, \boldsymbol{W}_{obs}|\boldsymbol{\theta}^{(g)}) d\boldsymbol{Y}_{obs}$$

$$= \int_{A_{\boldsymbol{Q}_{obs}}} \frac{f(\boldsymbol{Y}_{obs}, \boldsymbol{W}_{obs}|\boldsymbol{\theta}^{(g)})}{f(\boldsymbol{Y}_{obs}, \boldsymbol{W}_{obs}|\overline{\boldsymbol{\theta}})} \cdot f(\boldsymbol{Y}_{obs}, \boldsymbol{W}_{obs}|\overline{\boldsymbol{\theta}}) dY_{obs} = \mathrm{E}\left[\frac{f(\boldsymbol{Y}_{obs}, \boldsymbol{W}_{obs}|\boldsymbol{\theta}^{(g)})}{f(\boldsymbol{Y}_{obs}, \boldsymbol{W}_{obs}|\overline{\boldsymbol{\theta}})} \cdot I_{A_{\boldsymbol{Q}_{obs}}}\right], \tag{7.17}$$

where $A_{\boldsymbol{Q}_{obs}}$ is the truncation region of $\boldsymbol{Y}_{obs}$, and $f(\boldsymbol{Y}_{obs}, \boldsymbol{W}_{obs}|\overline{\boldsymbol{\theta}})$ is the joint likelihood at $\boldsymbol{\theta} = \overline{\boldsymbol{\theta}}$. The expectation in (7.17) can be estimated as

$$\hat{\mathrm{E}}\left[\frac{f(\boldsymbol{Y}_{obs}, \boldsymbol{W}_{obs}|\boldsymbol{\theta}^{(g)})}{f(\boldsymbol{Y}_{obs}, \boldsymbol{W}_{obs}|\overline{\boldsymbol{\theta}})} \cdot I_{A_{\boldsymbol{Q}_{obs}}}\right] = \frac{1}{K}\sum_{k=1}^{K} \frac{f(\boldsymbol{Y}_{obs}^{(k)}, \boldsymbol{W}_{obs}|\boldsymbol{\theta}^{(g)})}{f(\boldsymbol{Y}_{obs}^{(k)}, \boldsymbol{W}_{obs}|\overline{\boldsymbol{\theta}})} \cdot I_{A_{\boldsymbol{Q}_{obs}}},$$

using the same Monte Carlo sample we used to compute the integral in (7.16).

Web appendix references

Roy, V., and Hobert, J. (2007). Convergence rates and asymptotic standard errors for MCMC algorithms for Bayesian probit regression. *Journal of the Royal Statistical Society, Series B*, 69, 607-623.