Supplemental material

Appendix A - Derivations of MLE

The partial derivatives of the log-likelihood function are:

$$\frac{\partial l(\boldsymbol{\theta})}{\partial a_0} = -\sum_{d_{all}} \frac{\exp(a_0 + \mathbf{x}_i^T\boldsymbol{\beta})}{1+\exp(a_0 + \mathbf{x}_i^T\boldsymbol{\beta})} + n_1, \tag{1}$$

$$\begin{aligned}\frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}} =& -\sum_{d_{all}} \frac{\exp(a_0 + \mathbf{x}_i^T\boldsymbol{\beta})}{1+\exp(a_0 + \mathbf{x}_i^T\boldsymbol{\beta})}\mathbf{x}_i + \sum_{d_+}\mathbf{x}_i \\ &+ \frac{b_1}{\sigma_1^2(1-\rho^2)}\sum_{d_+}(\log(y_{i,1}) - a_1 - b_1\mathbf{x}_i^T\boldsymbol{\beta})\mathbf{x}_i \\ &+ \frac{b_2}{\sigma_2^2(1-\rho^2)}\sum_{d_+}(\log(y_{i,2}) - a_2 - b_2\mathbf{x}_i^T\boldsymbol{\beta})\mathbf{x}_i \\ &- \frac{\rho}{\sigma_1\sigma_2(1-\rho^2)}\Bigg[\sum_{d_+}(\log(y_{i,1}) - a_1 - b_1\mathbf{x}_i^T\boldsymbol{\beta})b_2\mathbf{x}_i \\ &+ \sum_{d_+}(\log(y_{i,2}) - a_2 - b_2\mathbf{x}_i^T\boldsymbol{\beta})b_1\mathbf{x}_i\Bigg],\end{aligned} \tag{2}$$

$$\frac{\partial l(\boldsymbol{\theta})}{\partial a_1} = \frac{1}{\sigma_1^2(1-\rho^2)}\sum_{d_+}(\log(y_{i,1}) - a_1 - b_1\mathbf{x}_i^T\boldsymbol{\beta}) - \frac{\rho}{\sigma_1\sigma_2(1-\rho^2)}\sum_{d_+}(\log(y_{i,2}) - a_2 - b_2\mathbf{x}_i^T\boldsymbol{\beta}), \tag{3}$$

$$\begin{aligned}\frac{\partial l(\boldsymbol{\theta})}{\partial b_1} =& \frac{1}{\sigma_1^2(1-\rho^2)}\sum_{d_+}(\log(y_{i,1}) - a_1 - b_1\mathbf{x}_i^T\boldsymbol{\beta})\mathbf{x}_i^T\boldsymbol{\beta} \\ &- \frac{\rho}{\sigma_1\sigma_2(1-\rho^2)}\sum_{d_+}(\log(y_{i,2}) - a_2 - b_2\mathbf{x}_i^T\boldsymbol{\beta})\mathbf{x}_i^T\boldsymbol{\beta},\end{aligned} \tag{4}$$

$$\frac{\partial l(\boldsymbol{\theta})}{\partial a_2} = \frac{1}{\sigma_2^2(1-\rho^2)}\sum_{d_+}(\log(y_{i,2}) - a_2 - b_2\mathbf{x}_i^T\boldsymbol{\beta}) - \frac{\rho}{\sigma_1\sigma_2(1-\rho^2)}\sum_{d_+}(\log(y_{i,1}) - a_1 - b_1\mathbf{x}_i^T\boldsymbol{\beta}), \tag{5}$$

$$\begin{aligned}\frac{\partial l(\boldsymbol{\theta})}{\partial b_2} =& \frac{1}{\sigma_2^2(1-\rho^2)}\sum_{d_+}(\log(y_{i,2}) - a_2 - b_2\mathbf{x}_i^T\boldsymbol{\beta})\mathbf{x}_i^T\boldsymbol{\beta} \\ &- \frac{\rho}{\sigma_1\sigma_2(1-\rho^2)}\sum_{d_+}(\log(y_{i,1}) - a_1 - b_1\mathbf{x}_i^T\boldsymbol{\beta})\mathbf{x}_i^T\boldsymbol{\beta},\end{aligned} \tag{6}$$

$$\begin{aligned}\frac{\partial l(\boldsymbol{\theta})}{\partial \sigma_1^2} =& -\frac{n_1}{2\sigma_1^2} + \frac{1}{2\sigma_1^4(1-\rho^2)}\sum_{d_+}(\log(y_{i,1}) - a_1 - b_1\mathbf{x}_i^T\boldsymbol{\beta})^2 \\ &- \frac{\rho}{2\sigma_1^3\sigma_2(1-\rho^2)}\sum_{d_+}(\log(y_{i,1}) - a_1 - b_1\mathbf{x}_i^T\boldsymbol{\beta})(\log(y_{i,2}) - a_2 - b_2\mathbf{x}_i^T\boldsymbol{\beta}),\end{aligned} \tag{7}$$

$$\begin{aligned}\frac{\partial l(\boldsymbol{\theta})}{\partial \sigma_2^2} =& -\frac{n_1}{2\sigma_2^2} + \frac{1}{2\sigma_2^4(1-\rho^2)}\sum_{d_+}(\log(y_{i,2}) - a_2 - b_2\mathbf{x}_i^T\boldsymbol{\beta})^2 \\ & - \frac{\rho}{2\sigma_1\sigma_2^3(1-\rho^2)}\sum_{d_+}(\log(y_{i,1}) - a_1 - b_1\mathbf{x}_i^T\boldsymbol{\beta})(\log(y_{i,2}) - a_2 - b_2\mathbf{x}_i^T\boldsymbol{\beta}), \end{aligned} \tag{8}$$

$$\begin{aligned}\frac{\partial l(\boldsymbol{\theta})}{\partial \rho} =& \frac{n_1\rho}{(1-\rho^2)} - \frac{\rho}{\sigma_1^2(1-\rho^2)^2}\sum_{d_+}(\log(y_{i,1}) - a_1 - b_1\mathbf{x}_i^T\boldsymbol{\beta})^2 \\ & - \frac{\rho}{\sigma_2^2(1-\rho^2)^2}\sum_{d_+}(\log(y_{i,2}) - a_2 - b_2\mathbf{x}_i^T\boldsymbol{\beta})^2 \\ & + \frac{1+\rho^2}{\sigma_1\sigma_2(1-\rho^2)^2}\sum_{d_+}(\log(y_{i,1}) - a_1 - b_1\mathbf{x}_i^T\boldsymbol{\beta})(\log(y_{i,2}) - a_2 - b_2\mathbf{x}_i^T\boldsymbol{\beta}). \end{aligned} \tag{9}$$

From the above equations (1) to (9), there is no analytic form for the MLE $\hat{\boldsymbol{\theta}}$, thus a numeric method is needed. Applying the Newton-Raphson method, we have the following iterative procedure:

$$\hat{\boldsymbol{\theta}}^{(t+1)} = \hat{\boldsymbol{\theta}}^{(t)} - \mathbf{H}^{-1}(\hat{\boldsymbol{\theta}}^{(t)})\mathbf{s}(\hat{\boldsymbol{\theta}}^{(t)}), \quad t = 0, 1, \cdots, \tag{10}$$

where $\mathbf{H}(\boldsymbol{\theta})$ denotes the Hessian matrix of second derivatives of the log-likelihood function,

$$\mathbf{H}(\boldsymbol{\theta}) = \frac{\partial^2 l(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T} = \frac{\partial \mathbf{s}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}. \tag{11}$$

When the convergence is attained, the observed Fisher information, i.e. $-\mathbf{H}(\hat{\boldsymbol{\theta}})$ becomes the inverse of the estimated covariance matrix. Note that for equations (7) to (9), we can write them concisely using the matrix form. If we define $\mathbf{a} = (a_1, a_2)^T$ and $\mathbf{b} = (b_1, b_2)^T$, we have

$$\frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\Sigma}} = -\frac{n_1}{2}\boldsymbol{\Sigma}^{-1} + \frac{1}{2}\sum_{d_+}\boldsymbol{\Sigma}^{-1}(\log(\mathbf{y}_i) - \mathbf{a} - \mathbf{b}\mathbf{x}_i^T\boldsymbol{\beta})(\log(\mathbf{y}_i) - \mathbf{a} - \mathbf{b}\mathbf{x}_i^T\boldsymbol{\beta})^T\boldsymbol{\Sigma}^{-1}. \tag{12}$$

From (10), the MLE of $\boldsymbol{\Sigma}$ is

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{n_1}\sum_{d_+}(\log(\mathbf{y}_i) - \hat{\mathbf{a}} - \hat{\mathbf{b}}\mathbf{x}_i^T\hat{\boldsymbol{\beta}})(\log(\mathbf{y}_i) - \hat{\mathbf{a}} - \hat{\mathbf{b}}\mathbf{x}_i^T\hat{\boldsymbol{\beta}})^T, \tag{13}$$

which is a consistent estimator of $\boldsymbol{\Sigma}$, and is asymptotically independent with $(\hat{a}_0, \hat{\boldsymbol{\beta}}^T, \hat{a}_1, \hat{b}_1, \hat{a}_2, \hat{b}_2)$. Since our primary interest is not on $\boldsymbol{\Sigma}$, we will use this result directly in the next part, more discussions can be found in Mardia and Marshall (1984) and Hamilton (1994) (Page 300).

Appendix B - Form of $\mathbf{I}_{BC2F}$

Let $\mathbf{P}, \mathbf{Q}_1, \mathbf{Q}_2, \mathbf{Q}_1^{\frac{1}{2}}, \mathbf{Q}_2^{\frac{1}{2}}$ and $\mathbf{Q}_0$ be $n \times n$ diagonal matrices, for $i = 1, \cdots, n$, the ith diagonal elements of these five matrices are given by

$$\frac{\exp(a_0 + \mathbf{x}_i^T \boldsymbol{\beta})}{(1 + \exp(a_0 + \mathbf{x}_i^T \boldsymbol{\beta}))^2}, \frac{\exp(a_0 + \mathbf{x}_i^T \boldsymbol{\beta})}{\sigma_1^2(1-\rho^2)(1 + \exp(a_0 + \mathbf{x}_i^T \boldsymbol{\beta}))}, \frac{\exp(a_0 + \mathbf{x}_i^T \boldsymbol{\beta})}{\sigma_1^2(1-\rho^2)(1 + \exp(a_0 + \mathbf{x}_i^T \boldsymbol{\beta}))},$$
$$\sqrt{\frac{\exp(a_0 + \mathbf{x}_i^T \boldsymbol{\beta})}{\sigma_1^2(1-\rho^2)(1 + \exp(a_0 + \mathbf{x}_i^T \boldsymbol{\beta}))}}, \sqrt{\frac{\exp(a_0 + \mathbf{x}_i^T \boldsymbol{\beta})}{\sigma_2^2(1-\rho^2)(1 + \exp(a_0 + \mathbf{x}_i^T \boldsymbol{\beta}))}},$$
and $\frac{\exp(a_0 + \mathbf{x}_i^T \boldsymbol{\beta})}{(1-\rho^2)(1 + \exp(a_0 + \mathbf{x}_i^T \boldsymbol{\beta}))}$, respectively.

For $\mathbf{I}_{\boldsymbol{\Sigma}}$, we have

$$\mathbf{I}_{\boldsymbol{\Sigma}} = \frac{\mathbf{J}^T \mathbf{Q}_0 \mathbf{J}}{2} \begin{pmatrix} \frac{2-\rho^2}{2\sigma_1^4} & \frac{-\rho^2}{2\sigma_1^2\sigma_2^2} & \frac{-\rho}{\sigma_1^2} \\ \frac{-\rho^2}{2\sigma_1^2\sigma_2^2} & \frac{2-\rho^2}{2\sigma_2^4} & \frac{-\rho}{\sigma_2^2} \\ \frac{-\rho}{\sigma_1^2} & \frac{-\rho}{\sigma_2^2} & \frac{2+2\rho^2}{1-\rho^2} \end{pmatrix}. \tag{14}$$

For $\mathbf{I}_{BC2}$, we have

$$\mathbf{I}_{BC2} = (\mathbf{a}_{rs})_{(p+5)\times(p+5)}, \tag{15}$$

where $\mathbf{a}_{rs}$ can be scalars, vectors or block matrices. Since $\mathbf{I}_{BC2}$ is symmetric, we have $\mathbf{a}_{rs} = \mathbf{a}_{sr}^T$.

Appendix C - Form of Fisher Information Matrix $\mathbf{I}_{BC2}$

The elements $\mathbf{a}_{rs}$ of $\mathbf{I}_{BC2}$ are given by (upper triangular parts):

$$\begin{aligned}
\mathbf{a}_{11} =& \mathbf{I}_{a_0 a_0} = \mathbf{J}^T \mathbf{P} \mathbf{J}, \mathbf{a}_{12} = \mathbf{I}_{a_0 \boldsymbol{\beta}} = \mathbf{J}^T \mathbf{P} \mathbf{X}, \\
\mathbf{a}_{13} =& \mathbf{I}_{a_0 a_1} = \mathbf{0}, \mathbf{a}_{14} = \mathbf{I}_{a_0 b_1} = \mathbf{0}, \mathbf{a}_{15} = \mathbf{I}_{a_0 a_2} = \mathbf{0}, \mathbf{a}_{16} = \mathbf{I}_{a_0 b_2} = \mathbf{0}, \\
\mathbf{a}_{22} =& \mathbf{I}_{\boldsymbol{\beta}\boldsymbol{\beta}} = \mathbf{X}^T \mathbf{P} \mathbf{X} + \mathbf{X}^T (b_1^2 \mathbf{Q}_1 + b_2^2 \mathbf{Q}_2 - 2\rho b_1 b_2 \mathbf{Q}_1^{\frac{1}{2}} \mathbf{Q}_2^{\frac{1}{2}}) \mathbf{X}, \\
\mathbf{a}_{23} =& \mathbf{I}_{\boldsymbol{\beta} a_1} = \mathbf{X}^T (b_1 \mathbf{Q}_1 - \rho b_2 \mathbf{Q}_1^{\frac{1}{2}} \mathbf{Q}_2^{\frac{1}{2}}) \mathbf{J}, \mathbf{a}_{24} = \mathbf{I}_{\boldsymbol{\beta} b_1} = \mathbf{X}^T (b_1 \mathbf{Q}_1 - \rho b_2 \mathbf{Q}_1^{\frac{1}{2}} \mathbf{Q}_2^{\frac{1}{2}}) \mathbf{X}\boldsymbol{\beta}, \\
\mathbf{a}_{25} =& \mathbf{I}_{\boldsymbol{\beta} a_2} = \mathbf{X}^T (b_2 \mathbf{Q}_2 - \rho b_1 \mathbf{Q}_1^{\frac{1}{2}} \mathbf{Q}_2^{\frac{1}{2}}) \mathbf{J}, \mathbf{a}_{26} = \mathbf{I}_{\boldsymbol{\beta} b_2} = \mathbf{X}^T (b_2 \mathbf{Q}_2 - \rho b_1 \mathbf{Q}_1^{\frac{1}{2}} \mathbf{Q}_2^{\frac{1}{2}}) \mathbf{X}\boldsymbol{\beta}, \\
\mathbf{a}_{33} =& \mathbf{I}_{a_1 a_1} = \mathbf{J}^T \mathbf{Q}_1 \mathbf{J}, \mathbf{a}_{34} = \mathbf{I}_{a_1 b_1} = \mathbf{J}^T \mathbf{Q}_1 \mathbf{X}\boldsymbol{\beta}, \mathbf{a}_{35} = \mathbf{I}_{a_1 a_2} = -\rho \mathbf{J}^T \mathbf{Q}_1^{\frac{1}{2}} \mathbf{Q}_2^{\frac{1}{2}} \mathbf{J}, \\
\mathbf{a}_{36} =& \mathbf{I}_{a_1 b_2} = -\rho \mathbf{J}^T \mathbf{Q}_1^{\frac{1}{2}} \mathbf{Q}_2^{\frac{1}{2}} \mathbf{X}\boldsymbol{\beta}, \mathbf{a}_{44} = \mathbf{I}_{b_1 b_1} = \boldsymbol{\beta}^T \mathbf{X}^T \mathbf{Q}_1 \mathbf{X}\boldsymbol{\beta}, \\
\mathbf{a}_{45} =& \mathbf{I}_{b_1 a_2} = -\rho \boldsymbol{\beta}^T \mathbf{X}^T \mathbf{Q}_1^{\frac{1}{2}} \mathbf{Q}_2^{\frac{1}{2}} \mathbf{J}, \mathbf{a}_{46} = \mathbf{I}_{b_1 b_2} = -\rho \boldsymbol{\beta}^T \mathbf{X}^T \mathbf{Q}_1^{\frac{1}{2}} \mathbf{Q}_2^{\frac{1}{2}} \mathbf{X}\boldsymbol{\beta}, \\
\mathbf{a}_{55} =& \mathbf{I}_{a_2 a_2} = \mathbf{J}^T \mathbf{Q}_2 \mathbf{J}, \mathbf{a}_{56} = \mathbf{I}_{a_2 b_2} = \mathbf{J}^T \mathbf{Q}_2 \mathbf{X}\boldsymbol{\beta}, \mathbf{a}_{66} = \mathbf{I}_{b_2 b_2} = \boldsymbol{\beta}^T \mathbf{X}^T \mathbf{Q}_2 \mathbf{X}\boldsymbol{\beta}.
\end{aligned}$$

Appendix D - Proof of Theorem 1

For the proof of Theorem 1, note that the MLE is a special case of the GEE estimator, we will apply the theorems from GEE theory. The proof of consistency is similar to Proposition 5.5 in Shao (2003). Note that the assumption of $h_i(X_i)$ and equicontinuity of the score function can be satisfied under the compactness assumption A1 and the differentiability of likelihood function, the identifiability assumption can be satisfied by assumption A2. To prove the $\sqrt{n}$ consistency and asymptotic normality, we use Theorem 1 in Ma and Kosorok (2005). Application of this theorem requires the following conditions to hold: (a) consistent sequence of GEE estimators, which is established before; (b) finite asymptotic variance, which is shown below; (c) equicontinuity of the gradient of the score function, which can be established using the assumption A1, the the consistency result and the differentiability of likelihood function.

Thus, we only need to establish the non-singularity of the information matrix, that is, to show $\mathbf{I}_{BC2F}$ is PD. To this end, we first show that $\mathbf{I}_{\boldsymbol{\Sigma}}$ in (14) is PD. Note that $\mathbf{J}^T\mathbf{Q}_0\mathbf{J}$ is PD, and we define the leading principal minors of the matrix

$$\begin{pmatrix} \frac{2-\rho^2}{2\sigma_1^4} & \frac{-\rho^2}{2\sigma_1^2\sigma_2^2} & \frac{-\rho}{\sigma_1^2} \\ \frac{-\rho^2}{2\sigma_1^2\sigma_2^2} & \frac{2-\rho^2}{2\sigma_2^4} & \frac{-\rho}{\sigma_2^2} \\ \frac{-\rho}{\sigma_1^2} & \frac{-\rho}{\sigma_2^2} & \frac{2+2\rho^2}{1-\rho^2} \end{pmatrix}$$

as $\mathbf{D}_1, \mathbf{D}_2$ and $\mathbf{D}_3$. Direct computation shows that

$$\mathbf{D}_1 = \frac{2-\rho^2}{2\sigma_1^4} > 0, \mathbf{D}_2 = \frac{1-1\rho^2}{\sigma_1^4\sigma_2^4} > 0, \mathbf{D}_3 = \frac{2}{\sigma_1^4\sigma_2^4} > 0.$$

Thus $\mathbf{I}_{\boldsymbol{\Sigma}}$ is PD.

Next we will show that $\mathbf{I}_{BC2}$ is also PD. The matrix given by (15) is complicated, however, we can rearrange it using simpler notations. If we let $\mathbf{a} = (a_1, a_2)^T$ and $\mathbf{b} = (b_1, b_2)^T$, it can be shown that the Fisher information matrix WRT $(a_0, \boldsymbol{\beta}, \mathbf{a}, \mathbf{b})$ is given by

$$\mathbf{I}^*_{BC2} = \begin{pmatrix} \mathbf{J}^T\mathbf{P}\mathbf{J} & \mathbf{J}^T\mathbf{P}\mathbf{X} & \mathbf{0} & \mathbf{0} \\ \mathbf{X}^T\mathbf{P}\mathbf{J} & \mathbf{X}^T\mathbf{P}\mathbf{X} + \mathbf{X}^T\mathbf{B}^T\mathbf{Q}^*\mathbf{B}\mathbf{X} & \mathbf{X}^T\mathbf{B}^T\mathbf{Q}^*\mathbf{J}^* & \mathbf{X}^T\mathbf{B}^T\mathbf{Q}^*\mathbf{T} \\ \mathbf{0} & \mathbf{J}^{*T}\mathbf{Q}^*\mathbf{B}\mathbf{X} & \mathbf{J}^{*T}\mathbf{Q}^*\mathbf{J}^* & \mathbf{J}^{*T}\mathbf{Q}^*\mathbf{T} \\ \mathbf{0} & \mathbf{T}^T\mathbf{Q}^*\mathbf{B}\mathbf{X} & \mathbf{T}^T\mathbf{Q}^*\mathbf{J}^* & \mathbf{T}^T\mathbf{Q}^*\mathbf{T} \end{pmatrix},$$

where $\mathbf{Q}^*$ is a block diagonal matrix with diagonal element

$$\frac{\exp(a_0 + \mathbf{x}_i^T\boldsymbol{\beta})}{1 + \exp(a_0 + \mathbf{x}_i^T\boldsymbol{\beta})}\boldsymbol{\Sigma}^{-1},$$

and

$$\mathbf{B} = \mathbf{I}_n \otimes \mathbf{b} = \begin{pmatrix} b_1 & 0 & \cdots & 0 \\ b_2 & 0 & \cdots & 0 \\ 0 & b_1 & \cdots & 0 \\ 0 & b_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & b_1 \\ 0 & 0 & \cdots & b_2 \end{pmatrix}, \mathbf{T} = \mathbf{X}\boldsymbol{\beta} \otimes \mathbf{I}_2 = \begin{pmatrix} \mathbf{x}_1^T\boldsymbol{\beta} & 0 \\ 0 & \mathbf{x}_1^T\boldsymbol{\beta} \\ \mathbf{x}_2^T\boldsymbol{\beta} & 0 \\ 0 & \mathbf{x}_2^T\boldsymbol{\beta} \\ \vdots & \vdots \\ \mathbf{x}_n^T\boldsymbol{\beta} & 0 \\ 0 & \mathbf{x}_n^T\boldsymbol{\beta} \end{pmatrix}, \mathbf{J}^* = \mathbf{J} \otimes \mathbf{I}_2 = \begin{pmatrix} \mathbf{I}_2 \\ \mathbf{I}_2 \\ \vdots \\ \mathbf{I}_2 \end{pmatrix},$$

where $\mathbf{I}_n$ and $\mathbf{I}_2$ are identity matrices of size n and 2, respectively, $\mathbf{J}, \mathbf{X}$ are defined as before. To show that the $\mathbf{I}^*_{BC2}$ is PD, one can see that $\mathbf{I}^*_{BC2} = \mathrm{I} + \mathrm{II}$, where

$$\mathrm{I} = \begin{pmatrix} \mathbf{J}^T\mathbf{P}\mathbf{J} & \mathbf{J}^T\mathbf{P}\mathbf{X} & \mathbf{0} & \mathbf{0} \\ \mathbf{X}^T\mathbf{P}\mathbf{J} & \mathbf{X}^T\mathbf{P}\mathbf{X} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix}, \mathrm{II} = \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{X}^T\mathbf{B}^T\mathbf{Q}^*\mathbf{B}\mathbf{X} & \mathbf{X}^T\mathbf{B}^T\mathbf{Q}^*\mathbf{J}^* & \mathbf{X}^T\mathbf{B}^T\mathbf{Q}^*\mathbf{T} \\ \mathbf{0} & \mathbf{J}^{*T}\mathbf{Q}^*\mathbf{B}\mathbf{X} & \mathbf{J}^{*T}\mathbf{Q}^*\mathbf{J}^* & \mathbf{J}^{*T}\mathbf{Q}^*\mathbf{T} \\ \mathbf{0} & \mathbf{T}^T\mathbf{Q}^*\mathbf{B}\mathbf{X} & \mathbf{T}^T\mathbf{Q}^*\mathbf{J}^* & \mathbf{T}^T\mathbf{Q}^*\mathbf{T} \end{pmatrix}.$$

Note that both I and II are PSD, if we assume there exists a vector $\boldsymbol{\alpha} = (\alpha_1, \boldsymbol{\alpha}_2, \boldsymbol{\alpha}_3, \boldsymbol{\alpha}_4)$, such that $\boldsymbol{\alpha}^T\mathbf{I}^*_{BC2}\boldsymbol{\alpha} = 0$, we must have $(\alpha_1, \boldsymbol{\alpha}_2) = (0, \mathbf{0})$ (since I is PSD), and $(\boldsymbol{\alpha}_3, \boldsymbol{\alpha}_4) = (\mathbf{0}, \mathbf{0})$ (since II is PSD). It yields that $\mathbf{I}^*_{BC2}$ is PD.

Appendix E - Proof of Theorem 2

To prove (i) and (ii) in Theorem 2, we need to use the fact given in Lemma 2.1 (Han and Kronmal 2006). We also introduce an intermediate model "$BC1$", where the proportional constraints only appear in the response $Y_{i,1}$. The model $BC1$ is given by,

$$\begin{aligned} \mathrm{logit}(p_{ij}) &= a_0 + \mathbf{x}_{ij}^T\boldsymbol{\beta}, \\ V_{ij,1} &= a_1 + b_1\mathbf{x}_{ij}^T\boldsymbol{\beta} + \epsilon_{ij,1} \quad \text{for } Y_{ij,1} > 0, \\ V_{ij,2} &= a_2 + \mathbf{x}_{ij}^T\boldsymbol{\xi} + \epsilon_{ij,2} \quad \text{for } Y_{ij,2} > 0. \end{aligned}$$

Our idea is to first show the efficiency of $BC1$ over $BC0$ and $UC1$, then show the efficiency of $BC2$ over $BC1$. To this end, it is suffice to show the following statements: In Loewner ordering,

(a) the asymptotic covariance matrix of the MLE of $(\hat{a}_0, \hat{\boldsymbol{\beta}})$, $\mathrm{Cov}(\hat{a}_0, \hat{\boldsymbol{\beta}})$ is no larger in $BC1$ than in $BC0$. The result also holds for $(\hat{a}_1, \hat{\boldsymbol{\delta}})$ and $(\hat{a}_2, \hat{\boldsymbol{\xi}})$;

(b) the asymptotic covariance matrix of the MLE of $(\hat{a}_0, \hat{\boldsymbol{\beta}})$, $\mathrm{Cov}(\hat{a}_0, \hat{\boldsymbol{\beta}})$ is no larger in $BC1$ than in $UC1$. The result also holds for $(\hat{a}_1, \hat{\boldsymbol{\delta}})$;

(c) the asymptotic covariance matrix of the MLE of $(\hat{a}_0, \hat{\boldsymbol{\beta}})$, $\mathrm{Cov}(\hat{a}_0, \hat{\boldsymbol{\beta}})$ is no larger in $BC2$ than in $BC1$. The result also holds for $(\hat{a}_1, \hat{\boldsymbol{\delta}})$ and $(\hat{a}_2, \hat{\boldsymbol{\xi}})$.

In Theorem 2, it is easy to see that (i) follows from (a) and (c), and (ii) follows from (b) and (c). Before showing (a), (b) and (c), we need the following facts, the computation details are similar to those for $BC2$ and are skipped here. Under $BC0$, the expected Fisher information matrix WRT $(a_0, \boldsymbol{\beta}, a_1, \boldsymbol{\delta}, a_2, \boldsymbol{\xi}, \sigma_1^2, \sigma_2^2, \rho)$ is given by $\begin{pmatrix} \mathbf{I}_{BC0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{\boldsymbol{\Sigma}} \end{pmatrix}$. Under $BC1$, the expected Fisher information matrix WRT $(a_0, \boldsymbol{\beta}, a_1, b_1, a_2, \boldsymbol{\xi}, \sigma_1^2, \sigma_2^2, \rho)$ is given by $\begin{pmatrix} \mathbf{I}_{BC1} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{\boldsymbol{\Sigma}} \end{pmatrix}$. $\mathbf{I}_{\boldsymbol{\Sigma}}$ is the same as that defined in (14), $\mathbf{I}_{BC0}$ is given by,

$$\begin{pmatrix}
\mathbf{J}^T\mathbf{P}\mathbf{J} & \mathbf{J}^T\mathbf{P}\mathbf{X} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
\mathbf{X}^T\mathbf{P}\mathbf{J} & \mathbf{X}^T\mathbf{P}\mathbf{X} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \mathbf{J}^T\mathbf{Q}_1\mathbf{J} & \mathbf{J}^T\mathbf{Q}_1\mathbf{X} & -\rho\mathbf{J}^T\mathbf{Q}_1^{\frac{1}{2}}\mathbf{Q}_2^{\frac{1}{2}}\mathbf{J} & -\rho\mathbf{J}^T\mathbf{Q}_1^{\frac{1}{2}}\mathbf{Q}_2^{\frac{1}{2}}\mathbf{X} \\
\mathbf{0} & \mathbf{0} & \mathbf{X}^T\mathbf{Q}_1\mathbf{J} & \mathbf{X}^T\mathbf{Q}_1\mathbf{X} & -\rho\mathbf{X}^T\mathbf{Q}_1^{\frac{1}{2}}\mathbf{Q}_2^{\frac{1}{2}}\mathbf{J} & -\rho\mathbf{X}^T\mathbf{Q}_1^{\frac{1}{2}}\mathbf{Q}_2^{\frac{1}{2}}\mathbf{X} \\
\mathbf{0} & \mathbf{0} & -\rho\mathbf{J}^T\mathbf{Q}_1^{\frac{1}{2}}\mathbf{Q}_2^{\frac{1}{2}}\mathbf{J} & -\rho\mathbf{J}^T\mathbf{Q}_1^{\frac{1}{2}}\mathbf{Q}_2^{\frac{1}{2}}\mathbf{X} & \mathbf{J}^T\mathbf{Q}_2\mathbf{J} & \mathbf{J}^T\mathbf{Q}_2\mathbf{X} \\
\mathbf{0} & \mathbf{0} & -\rho\mathbf{X}^T\mathbf{Q}_1^{\frac{1}{2}}\mathbf{Q}_2^{\frac{1}{2}}\mathbf{J} & -\rho\mathbf{X}^T\mathbf{Q}_1^{\frac{1}{2}}\mathbf{Q}_2^{\frac{1}{2}}\mathbf{X} & \mathbf{X}^T\mathbf{Q}_2\mathbf{J} & \mathbf{X}^T\mathbf{Q}_2\mathbf{X}
\end{pmatrix},$$

and

$$\mathbf{I}_{BC1} = (\mathbf{b}_{rs})_{(2p+4)\times(2p+4)},$$

where $\mathbf{b}_{rs}$ can be scalars, vectors or block matrices. Since $\mathbf{I}_{BC1}$ is symmetrical, we have $\mathbf{b}_{rs} = \mathbf{b}_{sr}^T$. The elements $\mathbf{b}_{rs}$ are given by (upper triangular parts):

$$\begin{aligned}
\mathbf{b}_{11} =& \mathbf{I}_{a_0a_0} = \mathbf{J}^T\mathbf{P}\mathbf{J}, \mathbf{b}_{12} = \mathbf{I}_{a_0\boldsymbol{\beta}} = \mathbf{J}^T\mathbf{P}\mathbf{X}, \mathbf{b}_{13} = \mathbf{I}_{a_0a_1} = \mathbf{0}, \mathbf{b}_{14} = \mathbf{I}_{a_0b_1} = \mathbf{0}, \\
\mathbf{b}_{15} =& \mathbf{I}_{a_0a_2} = \mathbf{0}, \mathbf{b}_{16} = \mathbf{I}_{a_0\boldsymbol{\xi}} = \mathbf{0}, \mathbf{b}_{22} = \mathbf{I}_{\boldsymbol{\beta}\boldsymbol{\beta}} = \mathbf{X}^T\mathbf{P}\mathbf{X} + b_1^2\mathbf{X}^T\mathbf{Q}_1\mathbf{X}, \\
\mathbf{b}_{23} =& \mathbf{I}_{\boldsymbol{\beta}a_1} = b_1\mathbf{X}^T\mathbf{Q}_1\mathbf{J}, \mathbf{b}_{24} = \mathbf{I}_{\boldsymbol{\beta}b_1} = b_1\mathbf{X}^T\mathbf{Q}_1\mathbf{X}\boldsymbol{\beta}, \mathbf{b}_{25} = \mathbf{I}_{\boldsymbol{\beta}a_2} = -\rho b_1\mathbf{X}^T\mathbf{Q}_1^{\frac{1}{2}}\mathbf{Q}_2^{\frac{1}{2}}\mathbf{J}, \\
\mathbf{b}_{26} =& \mathbf{I}_{\boldsymbol{\beta}\boldsymbol{\xi}} = -\rho b_1\mathbf{X}^T\mathbf{Q}_1^{\frac{1}{2}}\mathbf{Q}_2^{\frac{1}{2}}\mathbf{X}, \mathbf{b}_{33} = \mathbf{I}_{a_1a_1} = \mathbf{J}^T\mathbf{Q}_1\mathbf{J}, \mathbf{b}_{34} = \mathbf{I}_{a_1b_1} = \mathbf{J}^T\mathbf{Q}_1\mathbf{X}\boldsymbol{\beta}, \\
\mathbf{b}_{35} =& \mathbf{I}_{a_1a_2} = -\rho\mathbf{J}^T\mathbf{Q}_1^{\frac{1}{2}}\mathbf{Q}_2^{\frac{1}{2}}\mathbf{J}, \mathbf{b}_{36} = \mathbf{I}_{a_1\boldsymbol{\xi}} = -\rho\mathbf{J}^T\mathbf{Q}_1^{\frac{1}{2}}\mathbf{Q}_2^{\frac{1}{2}}\mathbf{X}, \mathbf{b}_{44} = \mathbf{I}_{b_1b_1} = \boldsymbol{\beta}^T\mathbf{X}^T\mathbf{Q}_1\mathbf{X}\boldsymbol{\beta}, \\
\mathbf{b}_{45} =& \mathbf{I}_{b_1a_2} = -\rho\boldsymbol{\beta}^T\mathbf{X}^T\mathbf{Q}_1^{\frac{1}{2}}\mathbf{Q}_2^{\frac{1}{2}}\mathbf{J}, \mathbf{b}_{46} = \mathbf{I}_{b_1\boldsymbol{\xi}} = -\rho\boldsymbol{\beta}^T\mathbf{X}^T\mathbf{Q}_1^{\frac{1}{2}}\mathbf{Q}_2^{\frac{1}{2}}\mathbf{X}, \\
\mathbf{b}_{55} =& \mathbf{I}_{a_2a_2} = \mathbf{J}^T\mathbf{Q}_2\mathbf{J}, \mathbf{b}_{56} = \mathbf{I}_{a_2\boldsymbol{\xi}} = \mathbf{J}^T\mathbf{Q}_2\mathbf{X}, \mathbf{b}_{66} = \mathbf{I}_{\boldsymbol{\xi}\boldsymbol{\xi}} = \mathbf{X}^T\mathbf{Q}_2\mathbf{X}.
\end{aligned}$$

Armed with these results, we can now prove the statements (a), (b) and (c). To show (a), it is suffice to show a more general result: the asymptotic covariance matrix of the MLE of $(\hat{a}_0, \hat{\boldsymbol{\beta}}^T, \hat{a}_1, \hat{a}_2, \hat{\boldsymbol{\xi}}^T)^T$ is no larger in $BC1$ than in $BC0$. Applying Lemma 2.1 to both $\mathbf{I}_{BC1}$ and $\mathbf{I}_{BC0}$, we have,

$$
\begin{aligned}
&\mathrm{Cov}^{-1}_{BC1}((\hat{a}_0, \hat{\boldsymbol{\beta}}^T, \hat{a}_1, \hat{a}_2, \hat{\boldsymbol{\xi}}^T)^T) - \mathrm{Cov}^{-1}_{BC0}((\hat{a}_0, \hat{\boldsymbol{\beta}}^T, \hat{a}_1, \hat{a}_2, \hat{\boldsymbol{\xi}}^T)^T) \\
&= \begin{pmatrix}
\mathbf{J}^T\mathbf{P}\mathbf{J} & \mathbf{J}^T\mathbf{P}\mathbf{X} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
\mathbf{X}^T\mathbf{P}\mathbf{J} & \mathbf{X}^T\mathbf{P}\mathbf{X} + b_1^2\mathbf{X}^T\mathbf{Q}_1\mathbf{X} & b_1\mathbf{X}^T\mathbf{Q}_1\mathbf{J} & -\rho b_1\mathbf{X}^T\mathbf{Q}_1^{\frac{1}{2}}\mathbf{Q}_2^{\frac{1}{2}}\mathbf{J} & -\rho b_1\mathbf{X}^T\mathbf{Q}_1^{\frac{1}{2}}\mathbf{Q}_2^{\frac{1}{2}}\mathbf{X} \\
\mathbf{0} & b_1\mathbf{J}^T\mathbf{Q}_1\mathbf{X} & \mathbf{J}^T\mathbf{Q}_1\mathbf{J} & -\rho\mathbf{J}^T\mathbf{Q}_1^{\frac{1}{2}}\mathbf{Q}_2^{\frac{1}{2}}\mathbf{J} & -\rho\mathbf{J}^T\mathbf{Q}_1^{\frac{1}{2}}\mathbf{Q}_2^{\frac{1}{2}}\mathbf{X} \\
\mathbf{0} & -\rho b_1\mathbf{J}^T\mathbf{Q}_1^{\frac{1}{2}}\mathbf{Q}_2^{\frac{1}{2}}\mathbf{X} & -\rho\mathbf{J}^T\mathbf{Q}_1^{\frac{1}{2}}\mathbf{Q}_2^{\frac{1}{2}}\mathbf{J} & \mathbf{J}^T\mathbf{Q}_2\mathbf{J} & \mathbf{J}^T\mathbf{Q}_2\mathbf{X} \\
\mathbf{0} & -\rho b_1\mathbf{X}^T\mathbf{Q}_1^{\frac{1}{2}}\mathbf{Q}_2^{\frac{1}{2}}\mathbf{X} & -\rho\mathbf{X}^T\mathbf{Q}_1^{\frac{1}{2}}\mathbf{Q}_2^{\frac{1}{2}}\mathbf{J} & \mathbf{X}^T\mathbf{Q}_2\mathbf{J} & \mathbf{X}^T\mathbf{Q}_2\mathbf{X}
\end{pmatrix} \\
&- \begin{pmatrix}
\mathbf{0} \\
b_1\mathbf{X}^T\mathbf{Q}_1\mathbf{X}\boldsymbol{\beta} \\
\mathbf{J}^T\mathbf{Q}_1\mathbf{X}\boldsymbol{\beta} \\
-\rho\mathbf{J}^T\mathbf{Q}_1^{\frac{1}{2}}\mathbf{Q}_2^{\frac{1}{2}}\mathbf{X}\boldsymbol{\beta} \\
-\rho\mathbf{X}^T\mathbf{Q}_1^{\frac{1}{2}}\mathbf{Q}_2^{\frac{1}{2}}\mathbf{X}\boldsymbol{\beta}
\end{pmatrix} (\boldsymbol{\beta}^T\mathbf{X}^T\mathbf{Q}_1\mathbf{X}\boldsymbol{\beta})^{-1} \\
&\times \begin{pmatrix} \mathbf{0} & b_1\boldsymbol{\beta}^T\mathbf{X}^T\mathbf{Q}_1\mathbf{X} & \boldsymbol{\beta}^T\mathbf{X}^T\mathbf{Q}_1\mathbf{J} & -\rho\boldsymbol{\beta}^T\mathbf{X}^T\mathbf{Q}_1^{\frac{1}{2}}\mathbf{Q}_2^{\frac{1}{2}}\mathbf{J} & -\rho\boldsymbol{\beta}^T\mathbf{X}^T\mathbf{Q}_1^{\frac{1}{2}}\mathbf{Q}_2^{\frac{1}{2}}\mathbf{X} \end{pmatrix} \\
&- \begin{pmatrix}
\mathbf{J}^T\mathbf{P}\mathbf{J} & \mathbf{J}^T\mathbf{P}\mathbf{X} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
\mathbf{X}^T\mathbf{P}\mathbf{J} & \mathbf{X}^T\mathbf{P}\mathbf{X} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \mathbf{J}^T\mathbf{Q}_1\mathbf{J} & -\rho\mathbf{J}^T\mathbf{Q}_1^{\frac{1}{2}}\mathbf{Q}_2^{\frac{1}{2}}\mathbf{J} & -\rho\mathbf{J}^T\mathbf{Q}_1^{\frac{1}{2}}\mathbf{Q}_2^{\frac{1}{2}}\mathbf{X} \\
\mathbf{0} & \mathbf{0} & -\rho\mathbf{J}^T\mathbf{Q}_1^{\frac{1}{2}}\mathbf{Q}_2^{\frac{1}{2}}\mathbf{J} & \mathbf{J}^T\mathbf{Q}_2\mathbf{J} & \mathbf{J}^T\mathbf{Q}_2\mathbf{X} \\
\mathbf{0} & \mathbf{0} & -\rho\mathbf{X}^T\mathbf{Q}_1^{\frac{1}{2}}\mathbf{Q}_2^{\frac{1}{2}}\mathbf{J} & \mathbf{X}^T\mathbf{Q}_2\mathbf{J} & \mathbf{X}^T\mathbf{Q}_2\mathbf{X}
\end{pmatrix} \\
&+ \begin{pmatrix}
\mathbf{0} \\
\mathbf{0} \\
\mathbf{J}^T\mathbf{Q}_1\mathbf{X} \\
-\rho\mathbf{J}^T\mathbf{Q}_1^{\frac{1}{2}}\mathbf{Q}_2^{\frac{1}{2}}\mathbf{X} \\
-\rho\mathbf{X}^T\mathbf{Q}_1^{\frac{1}{2}}\mathbf{Q}_2^{\frac{1}{2}}\mathbf{X}
\end{pmatrix} (\mathbf{X}^T\mathbf{Q}_1\mathbf{X})^{-1} \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{X}^T\mathbf{Q}_1\mathbf{J} & -\rho\mathbf{X}^T\mathbf{Q}_1^{\frac{1}{2}}\mathbf{Q}_2^{\frac{1}{2}}\mathbf{J} & -\rho\mathbf{X}^T\mathbf{Q}_1^{\frac{1}{2}}\mathbf{Q}_2^{\frac{1}{2}}\mathbf{X} \end{pmatrix} \\
&= \begin{pmatrix}
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & b_1^2\mathbf{X}^T\mathbf{Q}_1\mathbf{X} & b_1\mathbf{X}^T\mathbf{Q}_1\mathbf{J} & -\rho b_1\mathbf{X}^T\mathbf{Q}_1^{\frac{1}{2}}\mathbf{Q}_2^{\frac{1}{2}}\mathbf{J} & -\rho b_1\mathbf{X}^T\mathbf{Q}_1^{\frac{1}{2}}\mathbf{Q}_2^{\frac{1}{2}}\mathbf{X} \\
\mathbf{0} & b_1\mathbf{J}^T\mathbf{Q}_1\mathbf{X} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & -\rho b_1\mathbf{J}^T\mathbf{Q}_1^{\frac{1}{2}}\mathbf{Q}_2^{\frac{1}{2}}\mathbf{X} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & -\rho b_1\mathbf{X}^T\mathbf{Q}_1^{\frac{1}{2}}\mathbf{Q}_2^{\frac{1}{2}}\mathbf{X} & \mathbf{0} & \mathbf{0} & \mathbf{0}
\end{pmatrix}
\end{aligned}
$$

$$
\begin{aligned}
&-\begin{pmatrix} \mathbf{0} \\ b_1\mathbf{X}^T\mathbf{Q}_1\mathbf{X} \\ \mathbf{J}^T\mathbf{Q}_1\mathbf{X} \\ -\rho\mathbf{J}^T\mathbf{Q}_1^{\frac{1}{2}}\mathbf{Q}_2^{\frac{1}{2}}\mathbf{X} \\ -\rho\mathbf{X}^T\mathbf{Q}_1^{\frac{1}{2}}\mathbf{Q}_2^{\frac{1}{2}}\mathbf{X} \end{pmatrix} \boldsymbol{\beta}(\boldsymbol{\beta}^T\mathbf{X}^T\mathbf{Q}_1\mathbf{X}\boldsymbol{\beta})^{-1}\boldsymbol{\beta}^T \\
&\times \begin{pmatrix} \mathbf{0} & b_1\mathbf{X}^T\mathbf{Q}_1\mathbf{X} & \mathbf{X}^T\mathbf{Q}_1\mathbf{J} & -\rho\mathbf{X}^T\mathbf{Q}_1^{\frac{1}{2}}\mathbf{Q}_2^{\frac{1}{2}}\mathbf{J} & -\rho\mathbf{X}^T\mathbf{Q}_1^{\frac{1}{2}}\mathbf{Q}_2^{\frac{1}{2}}\mathbf{X} \end{pmatrix} \\
&+\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{J}^T\mathbf{Q}_1\mathbf{X} \\ -\rho\mathbf{J}^T\mathbf{Q}_1^{\frac{1}{2}}\mathbf{Q}_2^{\frac{1}{2}}\mathbf{X} \\ -\rho\mathbf{X}^T\mathbf{Q}_1^{\frac{1}{2}}\mathbf{Q}_2^{\frac{1}{2}}\mathbf{X} \end{pmatrix} (\mathbf{X}^T\mathbf{Q}_1\mathbf{X})^{-1} \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{X}^T\mathbf{Q}_1\mathbf{J} & -\rho\mathbf{X}^T\mathbf{Q}_1^{\frac{1}{2}}\mathbf{Q}_2^{\frac{1}{2}}\mathbf{J} & -\rho\mathbf{X}^T\mathbf{Q}_1^{\frac{1}{2}}\mathbf{Q}_2^{\frac{1}{2}}\mathbf{X} \end{pmatrix} \\
\geq &\begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & b_1^2\mathbf{X}^T\mathbf{Q}_1\mathbf{X} & b_1\mathbf{X}^T\mathbf{Q}_1\mathbf{J} & -\rho b_1\mathbf{X}^T\mathbf{Q}_1^{\frac{1}{2}}\mathbf{Q}_2^{\frac{1}{2}}\mathbf{J} & -\rho b_1\mathbf{X}^T\mathbf{Q}_1^{\frac{1}{2}}\mathbf{Q}_2^{\frac{1}{2}}\mathbf{X} \\ \mathbf{0} & b_1\mathbf{J}^T\mathbf{Q}_1\mathbf{X} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -\rho b_1\mathbf{J}^T\mathbf{Q}_1^{\frac{1}{2}}\mathbf{Q}_2^{\frac{1}{2}}\mathbf{X} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -\rho b_1\mathbf{X}^T\mathbf{Q}_1^{\frac{1}{2}}\mathbf{Q}_2^{\frac{1}{2}}\mathbf{X} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix} \\
&-\begin{pmatrix} \mathbf{0} \\ b_1\mathbf{X}^T\mathbf{Q}_1\mathbf{X} \\ \mathbf{J}^T\mathbf{Q}_1\mathbf{X} \\ -\rho\mathbf{J}^T\mathbf{Q}_1^{\frac{1}{2}}\mathbf{Q}_2^{\frac{1}{2}}\mathbf{X} \\ -\rho\mathbf{X}^T\mathbf{Q}_1^{\frac{1}{2}}\mathbf{Q}_2^{\frac{1}{2}}\mathbf{X} \end{pmatrix} (\mathbf{X}^T\mathbf{Q}_1\mathbf{X})^{-1} \\
&\times \begin{pmatrix} \mathbf{0} & b_1\mathbf{X}^T\mathbf{Q}_1\mathbf{X} & \mathbf{X}^T\mathbf{Q}_1\mathbf{J} & -\rho\mathbf{X}^T\mathbf{Q}_1^{\frac{1}{2}}\mathbf{Q}_2^{\frac{1}{2}}\mathbf{J} & -\rho\mathbf{X}^T\mathbf{Q}_1^{\frac{1}{2}}\mathbf{Q}_2^{\frac{1}{2}}\mathbf{X} \end{pmatrix} \\
&+\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{J}^T\mathbf{Q}_1\mathbf{X} \\ -\rho\mathbf{J}^T\mathbf{Q}_1^{\frac{1}{2}}\mathbf{Q}_2^{\frac{1}{2}}\mathbf{X} \\ -\rho\mathbf{X}^T\mathbf{Q}_1^{\frac{1}{2}}\mathbf{Q}_2^{\frac{1}{2}}\mathbf{X} \end{pmatrix} (\mathbf{X}^T\mathbf{Q}_1\mathbf{X})^{-1} \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{X}^T\mathbf{Q}_1\mathbf{J} & -\rho\mathbf{X}^T\mathbf{Q}_1^{\frac{1}{2}}\mathbf{Q}_2^{\frac{1}{2}}\mathbf{J} & -\rho\mathbf{X}^T\mathbf{Q}_1^{\frac{1}{2}}\mathbf{Q}_2^{\frac{1}{2}}\mathbf{X} \end{pmatrix} \\
= &\mathbf{0}.
\end{aligned}
$$

Thus we have shown the asymptotic covariance matrix of the MLE of $(\hat{a}_0, \hat{\boldsymbol{\beta}}^T, \hat{a}_1, \hat{a}_2, \hat{\boldsymbol{\xi}}^T)^T$ is no larger in $BC1$ than in $BC0$, by the delta method (Corollary 1.1 in Shao (2003)), the result of asymptotic efficiency must hold for $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\xi}}$. To show the result also holds for $\hat{\boldsymbol{\delta}}$, we can reparameterize $BC1$ by using $(a_0, c_1, a_1, \boldsymbol{\delta}, a_2, \boldsymbol{\xi})$ in place of $(a_0, \boldsymbol{\beta}, a_1, b_1, a_2, \boldsymbol{\xi})$, where $c_1 = 1/b_1$. For the reparameterized model, we can compute the expected Fisher information WRT $(a_0, c_1, a_1, \boldsymbol{\delta}, a_2, \boldsymbol{\xi})$, arguments similar to those used in the above proof.

Next we are going to show (b). Note that the parameters $(a_2, \boldsymbol{\xi})$ do not appear in the model $UC1$. Thus, it is suffice to show: the asymptotic covariance matrix of the MLE of $(\hat{a}_0, \hat{\boldsymbol{\beta}}^T, \hat{a}_1, \hat{b}_1)^T$ is no larger in $BC1$ than in $UC1$. (b) then follows from the delta method. Applying Lemma 2.1 to $\mathbf{I}_{BC1}$, together with the form of $\mathbf{I}_{UC1}$ from (8) (Han and Kronmal 2006), we have,

$$
\begin{aligned}
&\mathrm{Cov}_{BC1}^{-1}((\hat{a}_0, \hat{\boldsymbol{\beta}}^T, \hat{a}_1, \hat{b}_1)^T) - \mathrm{Cov}_{UC1}^{-1}((\hat{a}_0, \hat{\boldsymbol{\beta}}^T, \hat{a}_1, \hat{b}_1)^T) \\
&= \begin{pmatrix} \mathbf{J}^T\mathbf{P}\mathbf{J} & \mathbf{J}^T\mathbf{P}\mathbf{X} & \mathbf{0} & \mathbf{0} \\ \mathbf{X}^T\mathbf{P}\mathbf{J} & \mathbf{X}^T\mathbf{P}\mathbf{X} + b_1^2\mathbf{X}^T\mathbf{Q}_1\mathbf{X} & b_1\mathbf{X}^T\mathbf{Q}_1\mathbf{J} & b_1\mathbf{X}^T\mathbf{Q}_1\mathbf{X}\boldsymbol{\beta} \\ \mathbf{0} & b_1\mathbf{J}^T\mathbf{Q}_1\mathbf{X} & \mathbf{J}^T\mathbf{Q}_1\mathbf{J} & \mathbf{J}^T\mathbf{Q}_1\mathbf{X}\boldsymbol{\beta} \\ \mathbf{0} & b_1\boldsymbol{\beta}^T\mathbf{X}^T\mathbf{Q}_1\mathbf{X} & \boldsymbol{\beta}^T\mathbf{X}^T\mathbf{Q}_1\mathbf{J} & \boldsymbol{\beta}^T\mathbf{X}^T\mathbf{Q}_1\mathbf{X}\boldsymbol{\beta} \end{pmatrix} \\
&\quad - \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ -\rho b_1\mathbf{X}^T\mathbf{Q}_1^{\frac{1}{2}}\mathbf{Q}_2^{\frac{1}{2}}\mathbf{J} & -\rho b_1\mathbf{X}^T\mathbf{Q}_1^{\frac{1}{2}}\mathbf{Q}_2^{\frac{1}{2}}\mathbf{X} \\ -\rho\mathbf{J}^T\mathbf{Q}_1^{\frac{1}{2}}\mathbf{Q}_2^{\frac{1}{2}}\mathbf{J} & -\rho\mathbf{J}^T\mathbf{Q}_1^{\frac{1}{2}}\mathbf{Q}_2^{\frac{1}{2}}\mathbf{X} \\ -\rho\boldsymbol{\beta}^T\mathbf{X}^T\mathbf{Q}_1^{\frac{1}{2}}\mathbf{Q}_2^{\frac{1}{2}}\mathbf{J} & -\rho\boldsymbol{\beta}^T\mathbf{X}^T\mathbf{Q}_1^{\frac{1}{2}}\mathbf{Q}_2^{\frac{1}{2}}\mathbf{X} \end{pmatrix} \begin{pmatrix} \mathbf{J}^T\mathbf{Q}_2\mathbf{J} & \mathbf{J}^T\mathbf{Q}_2\mathbf{X} \\ \mathbf{X}^T\mathbf{Q}_2\mathbf{J} & \mathbf{X}^T\mathbf{Q}_2\mathbf{X} \end{pmatrix}^{-1} \\
&\quad \times \begin{pmatrix} \mathbf{0} & -\rho b_1\mathbf{J}^T\mathbf{Q}_1^{\frac{1}{2}}\mathbf{Q}_2^{\frac{1}{2}}\mathbf{X} & -\rho\mathbf{J}^T\mathbf{Q}_1^{\frac{1}{2}}\mathbf{Q}_2^{\frac{1}{2}}\mathbf{J} & -\rho\mathbf{J}^T\mathbf{Q}_1^{\frac{1}{2}}\mathbf{Q}_2^{\frac{1}{2}}\mathbf{X}\boldsymbol{\beta} \\ \mathbf{0} & -\rho b_1\mathbf{X}^T\mathbf{Q}_1^{\frac{1}{2}}\mathbf{Q}_2^{\frac{1}{2}}\mathbf{X} & -\rho\mathbf{X}^T\mathbf{Q}_1^{\frac{1}{2}}\mathbf{Q}_2^{\frac{1}{2}}\mathbf{J} & -\rho\mathbf{X}^T\mathbf{Q}_1^{\frac{1}{2}}\mathbf{Q}_2^{\frac{1}{2}}\mathbf{X}\boldsymbol{\beta} \end{pmatrix} \\
&\quad - \begin{pmatrix} \mathbf{J}^T\mathbf{P}\mathbf{J} & \mathbf{J}^T\mathbf{P}\mathbf{X} & \mathbf{0} & \mathbf{0} \\ \mathbf{X}^T\mathbf{P}\mathbf{J} & \mathbf{X}^T\mathbf{P}\mathbf{X} + b_1^2(1-\rho^2)\mathbf{X}^T\mathbf{Q}_1\mathbf{X} & b_1(1-\rho^2)\mathbf{X}^T\mathbf{Q}_1\mathbf{J} & b_1(1-\rho^2)\mathbf{X}^T\mathbf{Q}_1\mathbf{X}\boldsymbol{\beta} \\ \mathbf{0} & b_1(1-\rho^2)\mathbf{J}^T\mathbf{Q}_1\mathbf{X} & (1-\rho^2)\mathbf{J}^T\mathbf{Q}_1\mathbf{J} & (1-\rho^2)\mathbf{J}^T\mathbf{Q}_1\mathbf{X}\boldsymbol{\beta} \\ \mathbf{0} & b_1(1-\rho^2)\boldsymbol{\beta}^T\mathbf{X}^T\mathbf{Q}_1\mathbf{X} & (1-\rho^2)\boldsymbol{\beta}^T\mathbf{X}^T\mathbf{Q}_1\mathbf{J} & (1-\rho^2)\boldsymbol{\beta}^T\mathbf{X}^T\mathbf{Q}_1\mathbf{X}\boldsymbol{\beta} \end{pmatrix} \\
&= \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \rho^2\mathbf{V} \end{pmatrix},
\end{aligned}
$$

where

$$
\begin{aligned}
\mathbf{V} = &\begin{pmatrix} b_1^2\mathbf{X}^T\mathbf{Q}_1\mathbf{X} & b_1\mathbf{X}^T\mathbf{Q}_1\mathbf{J} & b_1\mathbf{X}^T\mathbf{Q}_1\mathbf{X}\boldsymbol{\beta} \\ b_1\mathbf{J}^T\mathbf{Q}_1\mathbf{X} & \mathbf{J}^T\mathbf{Q}_1\mathbf{J} & \mathbf{J}^T\mathbf{Q}_1\mathbf{X}\boldsymbol{\beta} \\ b_1\boldsymbol{\beta}^T\mathbf{X}^T\mathbf{Q}_1\mathbf{X} & \boldsymbol{\beta}^T\mathbf{X}^T\mathbf{Q}_1\mathbf{J} & \boldsymbol{\beta}^T\mathbf{X}^T\mathbf{Q}_1\mathbf{X}\boldsymbol{\beta} \end{pmatrix} \\
&- \begin{pmatrix} b_1\mathbf{X}^T\mathbf{Q}_1^{\frac{1}{2}}\mathbf{Q}_2^{\frac{1}{2}}\mathbf{J} & b_1\mathbf{X}^T\mathbf{Q}_1^{\frac{1}{2}}\mathbf{Q}_2^{\frac{1}{2}}\mathbf{X} \\ \mathbf{J}^T\mathbf{Q}_1^{\frac{1}{2}}\mathbf{Q}_2^{\frac{1}{2}}\mathbf{J} & \mathbf{J}^T\mathbf{Q}_1^{\frac{1}{2}}\mathbf{Q}_2^{\frac{1}{2}}\mathbf{X} \\ \boldsymbol{\beta}^T\mathbf{X}^T\mathbf{Q}_1^{\frac{1}{2}}\mathbf{Q}_2^{\frac{1}{2}}\mathbf{J} & \boldsymbol{\beta}^T\mathbf{X}^T\mathbf{Q}_1^{\frac{1}{2}}\mathbf{Q}_2^{\frac{1}{2}}\mathbf{X} \end{pmatrix} \begin{pmatrix} \mathbf{J}^T\mathbf{Q}_2\mathbf{J} & \mathbf{J}^T\mathbf{Q}_2\mathbf{X} \\ \mathbf{X}^T\mathbf{Q}_2\mathbf{J} & \mathbf{X}^T\mathbf{Q}_2\mathbf{X} \end{pmatrix}^{-1}
\end{aligned}
$$

$$
\begin{aligned}
&\times \begin{pmatrix} b_1\mathbf{J}^T\mathbf{Q}_1^{\frac{1}{2}}\mathbf{Q}_2^{\frac{1}{2}}\mathbf{X} & \mathbf{J}^T\mathbf{Q}_1^{\frac{1}{2}}\mathbf{Q}_2^{\frac{1}{2}}\mathbf{J} & \mathbf{J}^T\mathbf{Q}_1^{\frac{1}{2}}\mathbf{Q}_2^{\frac{1}{2}}\mathbf{X}\boldsymbol{\beta} \\ b_1\mathbf{X}^T\mathbf{Q}_1^{\frac{1}{2}}\mathbf{Q}_2^{\frac{1}{2}}\mathbf{X} & \mathbf{X}^T\mathbf{Q}_1^{\frac{1}{2}}\mathbf{Q}_2^{\frac{1}{2}}\mathbf{J} & \mathbf{X}^T\mathbf{Q}_1^{\frac{1}{2}}\mathbf{Q}_2^{\frac{1}{2}}\mathbf{X}\boldsymbol{\beta} \end{pmatrix} \\
=&\mathbf{A} - \mathbf{B}\mathbf{C}^{-1}\mathbf{B}^T,
\end{aligned}
$$

where $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ are defined accordingly. Hence, it is suffice to show that $\mathbf{V}$ is PSD. To this end, we need to show

$$
\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^T & \mathbf{C} \end{pmatrix}
$$

is PSD, $\mathbf{A}$ is PSD and $\mathbf{C}$ is PD. The latter two are obvious because $\mathbf{Q}_2$ is PD and $(\mathbf{J}\ \mathbf{X})$ has full column rank. We also have

$$
\begin{aligned}
&\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^T & \mathbf{C} \end{pmatrix} \\
=&\begin{pmatrix} b_1\mathbf{X}^T\mathbf{Q}_1^{\frac{1}{2}} \\ \mathbf{J}^T\mathbf{Q}_1^{\frac{1}{2}} \\ \boldsymbol{\beta}^T\mathbf{X}^T\mathbf{Q}_1^{\frac{1}{2}} \\ \mathbf{J}^T\mathbf{Q}_2^{\frac{1}{2}} \\ \mathbf{X}^T\mathbf{Q}_2^{\frac{1}{2}} \end{pmatrix} \begin{pmatrix} b_1\mathbf{Q}_1^{\frac{1}{2}}\mathbf{X} & \mathbf{Q}_1^{\frac{1}{2}}\mathbf{J} & \mathbf{Q}_1^{\frac{1}{2}}\mathbf{X}\boldsymbol{\beta} & \mathbf{Q}_2^{\frac{1}{2}}\mathbf{J} & \mathbf{Q}_2^{\frac{1}{2}}\mathbf{X} \end{pmatrix}
\end{aligned}
$$

is PSD.

Last we will show (c), again we will show a more general result: the asymptotic covariance matrix of the MLE of $(\hat{a}_0, \hat{\boldsymbol{\beta}}^T, \hat{a}_1, \hat{b}_1, \hat{a}_2)^T$ is no larger in $BC2$ than in $BC1$. Applying Lemma 2.1 to $\mathbf{I}_{BC2}$ and $\mathbf{I}_{BC1}$, also notice that from the form of $\mathbf{I}_{BC2}$ and $\mathbf{I}_{BC1}$, many elements can be cancelled, we have,

$$
\begin{aligned}
&\mathrm{Cov}_{BC2}^{-1}((\hat{a}_0, \hat{\boldsymbol{\beta}}^T, \hat{a}_1, \hat{b}_1, \hat{a}_2)^T) - \mathrm{Cov}_{BC1}^{-1}((\hat{a}_0, \hat{\boldsymbol{\beta}}^T, \hat{a}_1, \hat{b}_1, \hat{a}_2)^T) \\
=&\begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{X}^T(b_2^2\mathbf{Q}_2 - 2\rho b_1 b_2\mathbf{Q}_1^{\frac{1}{2}}\mathbf{Q}_2^{\frac{1}{2}})\mathbf{X} & -\rho b_2\mathbf{X}^T\mathbf{Q}_1^{\frac{1}{2}}\mathbf{Q}_2^{\frac{1}{2}}\mathbf{J} & -\rho b_2\mathbf{X}^T\mathbf{Q}_1^{\frac{1}{2}}\mathbf{Q}_2^{\frac{1}{2}}\mathbf{X}\boldsymbol{\beta} & b_2\mathbf{X}^T\mathbf{Q}_2\mathbf{J} \\ \mathbf{0} & -\rho b_2\mathbf{J}^T\mathbf{Q}_1^{\frac{1}{2}}\mathbf{Q}_2^{\frac{1}{2}}\mathbf{X} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -\rho b_2\boldsymbol{\beta}^T\mathbf{X}^T\mathbf{Q}_1^{\frac{1}{2}}\mathbf{Q}_2^{\frac{1}{2}}\mathbf{X} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & b_2\mathbf{J}^T\mathbf{Q}_2\mathbf{X} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix}
\end{aligned}
$$

$$
\begin{aligned}
&-\begin{pmatrix} \mathbf{0} \\ \mathbf{X}^T(b_2\mathbf{Q}_2 - \rho b_1 \mathbf{Q}_1^{\frac{1}{2}}\mathbf{Q}_2^{\frac{1}{2}})\mathbf{X} \\ -\rho \mathbf{J}^T\mathbf{Q}_1^{\frac{1}{2}}\mathbf{Q}_2^{\frac{1}{2}}\mathbf{X} \\ -\rho\boldsymbol{\beta}^T\mathbf{X}^T\mathbf{Q}_1^{\frac{1}{2}}\mathbf{Q}_2^{\frac{1}{2}}\mathbf{X} \\ \mathbf{J}^T\mathbf{Q}_2\mathbf{X} \end{pmatrix} \boldsymbol{\beta}(\boldsymbol{\beta}^T\mathbf{X}^T\mathbf{Q}_2\mathbf{X}\boldsymbol{\beta})^{-1}\boldsymbol{\beta}^T \\
&\times \begin{pmatrix} \mathbf{0} & \mathbf{X}^T(b_2\mathbf{Q}_2 - \rho b_1\mathbf{Q}_1^{\frac{1}{2}}\mathbf{Q}_2^{\frac{1}{2}})\mathbf{X} & -\rho\mathbf{X}^T\mathbf{Q}_1^{\frac{1}{2}}\mathbf{Q}_2^{\frac{1}{2}}\mathbf{J} & -\rho\mathbf{X}^T\mathbf{Q}_1^{\frac{1}{2}}\mathbf{Q}_2^{\frac{1}{2}}\mathbf{X}\boldsymbol{\beta} & \mathbf{X}^T\mathbf{Q}_2\mathbf{J} \end{pmatrix} \\
&+\begin{pmatrix} \mathbf{0} \\ -\rho b_1\mathbf{X}^T\mathbf{Q}_1^{\frac{1}{2}}\mathbf{Q}_2^{\frac{1}{2}}\mathbf{X} \\ -\rho \mathbf{J}^T\mathbf{Q}_1^{\frac{1}{2}}\mathbf{Q}_2^{\frac{1}{2}}\mathbf{X} \\ -\rho\boldsymbol{\beta}^T\mathbf{X}^T\mathbf{Q}_1^{\frac{1}{2}}\mathbf{Q}_2^{\frac{1}{2}}\mathbf{X} \\ \mathbf{J}^T\mathbf{Q}_2\mathbf{X} \end{pmatrix} (\mathbf{X}^T\mathbf{Q}_2\mathbf{X})^{-1} \\
&\times \begin{pmatrix} \mathbf{0} & -\rho b_1\mathbf{X}^T\mathbf{Q}_1^{\frac{1}{2}}\mathbf{Q}_2^{\frac{1}{2}}\mathbf{X} & -\rho\mathbf{X}^T\mathbf{Q}_1^{\frac{1}{2}}\mathbf{Q}_2^{\frac{1}{2}}\mathbf{J} & -\rho\mathbf{X}^T\mathbf{Q}_1^{\frac{1}{2}}\mathbf{Q}_2^{\frac{1}{2}}\mathbf{X}\boldsymbol{\beta} & \mathbf{X}^T\mathbf{Q}_2\mathbf{J} \end{pmatrix} \\
&\geq \mathbf{0},
\end{aligned}
$$

where the inequality comes from using the same technique as in proving (a). To show the result also holds for $\hat{\boldsymbol{\xi}}$, we reparameterize $BC2$ by using $(a_0, c_2, a_1, b_1, a_2, \boldsymbol{\xi})$ in place of $(a_0, \boldsymbol{\beta}, a_1, b_1, a_2, b_2)$, where $c_2 = 1/b_2$. For the reparameterized model, we can compute the expected Fisher information WRT $(a_0, c_2, a_1, b_1, a_2, \boldsymbol{\xi})$, arguments similar to those used in the above proof.

Appendix F - Simulation of ARE

We considered values of ρ between 0 and 0.9, and values of β_1 between 0 and 1.5; other parameters were set to be the same as in Table 1. Note that when $\rho = 0$, $BC2$ was not equivalent to $UC1$ since mean values of $\log(Y^*_{ij,1})$ and $\log(Y^*_{ij,2})$ were correlated. In $BC2$, the dependence need to be considered, while in $UC1$, the $\log(Y^*_{ij,d})$ part was ignored. Although we had explicit forms of the Fisher information matrices, in practice, we computed the estimated covariance matrices using the inverse of observed Fisher information, thus in the previous simulation, we increased the number n_i to 1000, and the sample size $n = 3000$. Table 1 shows the comparison between $BC2$ and $UC1$, and Table 2 shows the comparison between $BC2$ and $BC0$. From the results in the table, one can see that, when ρ changes, the ARE are not changing much, however, when β_1 (δ_1) increases, the ARE are uniformly going to 1. This is not surprising, because when β_1 (δ_1) is large, it becomes more significant in the model, and all the three models can detect it. However, when β_1 (δ_1) is not so significant, the SE of the estimates in models $UC1$ and $BC0$ become larger than those in the proposed model $BC2$. This empirical evidence suggests that, when the proportionality structure holds, the proposed model $BC2$ is more powerful than models $UC1$ and $BC0$ in determining the significance of the covariate effects.

Table 1. ARE of β_1 and $\delta_1 = b_1\beta_1$, Values Represent Variances under Model $BC2$ over Model $UC1$.

	β_1				$\delta_1 = b_1\beta_1$			
ρ	0	0.4	0.8	1.5	0	0.24	0.48	0.9
0.00	0.43	0.71	0.91	0.98	0.43	0.73	0.88	0.95
0.20	0.48	0.74	0.91	0.98	0.48	0.75	0.89	0.95
0.40	0.51	0.76	0.92	0.98	0.52	0.77	0.90	0.96
0.70	0.52	0.75	0.92	0.98	0.52	0.77	0.90	0.96
0.90	0.39	0.68	0.90	0.95	0.40	0.70	0.87	0.95

Table 2. ARE of β_1 and $\delta_1 = b_1\beta_1$, Values Represent Variances under Model $BC2$ over Model $BC0$.

	β_1				$\delta_1 = b_1\beta_1$			
ρ	0	0.4	0.8	1.5	0	0.24	0.48	0.9
0.00	0.27	0.52	0.76	0.94	0.17	0.43	0.71	0.91
0.20	0.29	0.54	0.77	0.94	0.19	0.45	0.71	0.92
0.40	0.31	0.56	0.77	0.94	0.20	0.46	0.72	0.92
0.70	0.31	0.55	0.77	0.94	0.21	0.45	0.72	0.92
0.90	0.24	0.50	0.76	0.94	0.16	0.41	0.70	0.91

Appendix G - Model Assumptions

Assumptions (A1)-(A4)

(A1) $\mathbf{X}$ is component-wise bounded. Denote the true value of $\boldsymbol{\theta}$ as $\boldsymbol{\theta}_T$, $\boldsymbol{\theta}_T$ is an interior point of an open set in the parameter space, $b_d \in \|\boldsymbol{\beta}\| \cdot \mathbf{R}$, $-1 < \rho < 1$.

(A2) The design matrix $(\mathbf{J}\ \mathbf{X})$ is assumed to have full column rank.

(A3) The model is correctly specified as in equation (6) of the main contents of the paper, and the error terms follow the distribution in equation (5) of the main contents of the paper.

(A4) The matrix $\mathbf{R}_{BC2F} = \lim_{n\to\infty} \frac{\mathbf{I}_{BC2F}}{n}$ exists.

Modified Assumptions (A3')-(A4') for mis-specified model

(A3') The independent vectors $\mathbf{Z}_i$ have a distribution with measurable (Radon-Nikodym) density $h(\cdot)$, and the parametric family of distribution functions all have densities $f(z, \boldsymbol{\theta})$, which is specified in equation (6) of the main contents of the paper. The Kullback-Leibler info $I(h : f, \boldsymbol{\theta}) = E(\ln \frac{h(\mathbf{Z})}{f(\mathbf{Z},\boldsymbol{\theta})})$ has a unique minimum at $\boldsymbol{\theta}_0$.

(A4') $E|(\ln h(\mathbf{Z}))| < \infty$; $|\frac{\partial^2 \ln f}{\partial\theta_i \partial\theta_j}|$ and $|\frac{\partial \ln f}{\partial\theta_i} \frac{\partial \ln f}{\partial\theta_i}|$ are dominated by functions integrable in z. $|B(\boldsymbol{\theta})| \neq 0$, rank of $A(\boldsymbol{\theta})$ is constant in a neighborhood of $\boldsymbol{\theta}_0$.

References

Hamilton JD (1994) *Time Series Analysis*, volume 2. Princeton University Press.

Han C and Kronmal R (2006) Two-part models for analysis of agatston scores with possible proportionality constraints. *Communications in Statistics Theory and Methods* 35: 99–111.

Ma S and Kosorok MR (2005) Robust semiparametric m-estimation and the weighted bootstrap. *Journal of Multivariate Analysis* 96: 190–217.

Mardia KV and Marshall R (1984) Maximum likelihood estimation of models for residual covariance in spatial regression. *Biometrika* 71: 135–146.

Shao J (2003) *Mathematical Statistics*. Springer Texts in Statistics. Springer.