Supplemental Material for "Envelope-based Partial Partial Least Squares with Application to Cytokine-based Biomarker Analysis for COVID-19"

Yeonhee Park[1,*], Zhihua Su[2], and Dongjun Chung[3]
[1]Department of Biostatistics and Medical Informatics, University of Wisconsin-Madison
[2]Department of Statistics, University of Florida
[3]Department of Biomedical Informatics, The Ohio State University
*email: ypark56@wisc.edu

A. DETAILS OF ESTIMATION

For a fixed dimension d ($1 \leq d \leq p_1$) of the envelope $\mathcal{E}_{\boldsymbol{\Sigma}_{1|2}}(\boldsymbol{\beta}_1)$, the normal likelihood is given by

$$\begin{aligned} l = & \frac{n(r+p_1)}{2}\log(2\pi) - \frac{n}{2}\log|\boldsymbol{\Omega}| - \frac{n}{2}\log|\boldsymbol{\Omega}_0| - \frac{1}{2}\operatorname{tr}[\{\mathbb{X}_1 - \mathbf{1}_n\boldsymbol{\mu}_1^T - (\mathbb{X}_2 - \mathbf{1}_n\bar{\mathbf{X}}_2^T)\boldsymbol{\gamma}\} \\ & (\boldsymbol{\Gamma}\boldsymbol{\Omega}\boldsymbol{\Gamma}^T + \boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^T)^{-1}\{\mathbb{X}_1 - \mathbf{1}_n\boldsymbol{\mu}_1^T - (\mathbb{X}_2 - \mathbf{1}_n\bar{\mathbf{X}}_2^T)\boldsymbol{\gamma}\}^T] - \frac{n}{2}\log|\boldsymbol{\Sigma}_{\mathbf{Y}|\mathbf{X}}| \\ & - \frac{1}{2}\operatorname{tr}[\{\mathbb{Y} - \mathbf{1}_n\boldsymbol{\mu}_{\mathbf{Y}}^T - (\mathbb{X}_1 - \mathbf{1}_n\boldsymbol{\mu}_1^T)\boldsymbol{\Gamma}\boldsymbol{\eta} - (\mathbb{X}_2 - \mathbf{1}_n\bar{\mathbf{X}}_2^T)\boldsymbol{\beta}_2\}\boldsymbol{\Sigma}_{\mathbf{Y}|\mathbf{X}}^{-1}\{\mathbb{Y} - \mathbf{1}_n\boldsymbol{\mu}_{\mathbf{Y}}^T \\ & - (\mathbb{X}_1 - \mathbf{1}_n\boldsymbol{\mu}_1^T)\boldsymbol{\Gamma}\boldsymbol{\eta} - (\mathbb{X}_2 - \mathbf{1}_n\bar{\mathbf{X}}_2^T)\boldsymbol{\beta}_2\}^T], \end{aligned} \tag{1}$$

where $\bar{\mathbf{X}}_2 = (1/n)\sum_{i=1}^n \mathbf{X}_{2i}$ denotes the sample mean of $\mathbf{X}_2$, and $\mathbf{1}_n \in \mathbb{R}^n$ denotes an n-dimensional vector of 1's. We hold $\boldsymbol{\Gamma}$ fixed and derive the estimator of parameters $\boldsymbol{\mu}_{\mathbf{Y}}, \boldsymbol{\mu}_1, \boldsymbol{\gamma}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{\Sigma}_{1|2}, \boldsymbol{\eta}, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2$, and $\boldsymbol{\Sigma}_{\mathbf{Y}|\mathbf{X}}$ as a function of $\boldsymbol{\Gamma}$. We first differentiate the log likelihood function l in (1) with respect to $\boldsymbol{\mu}_{\mathbf{Y}}$ and set the derivative to be $\mathbf{0}$, i.e,

$$\frac{\partial l}{\partial \boldsymbol{\mu}_{\mathbf{Y}}} = \boldsymbol{\Sigma}_{\mathbf{Y}|\mathbf{X}}^{-1}\{\mathbb{Y} - \mathbf{1}_n\boldsymbol{\mu}_{\mathbf{Y}}^T - (\mathbb{X}_1 - \mathbf{1}_n\boldsymbol{\mu}_1^T)\boldsymbol{\Gamma}\boldsymbol{\eta} - (\mathbb{X}_2 - \mathbf{1}_n\bar{\mathbf{X}}_2^T)\boldsymbol{\beta}_2\}^T\mathbf{1}_n = \mathbf{0}. \tag{2}$$

The derivative of l with respect to $\boldsymbol{\mu}_1$ is

$$\begin{aligned} \frac{\partial l}{\partial \boldsymbol{\mu}_1} = & -\boldsymbol{\Sigma}_{1|2}^{-1}\{\mathbb{X}_1 - \mathbf{1}_n\boldsymbol{\mu}_1^T - (\mathbb{X}_2 - \mathbf{1}_n\bar{\mathbf{X}}_2^T)\boldsymbol{\gamma}\}^T\mathbf{1}_n \\ & -\boldsymbol{\Gamma}\boldsymbol{\eta}\boldsymbol{\Sigma}_{\mathbf{Y}|\mathbf{X}}^{-1}\{\mathbb{Y} - \mathbf{1}_n\boldsymbol{\mu}_{\mathbf{Y}}^T - (\mathbb{X}_1 - \mathbf{1}_n\boldsymbol{\mu}_1^T)\boldsymbol{\Gamma}\boldsymbol{\eta} - (\mathbb{X}_2 - \mathbf{1}_n\bar{\mathbf{X}}_2^T)\boldsymbol{\beta}_2\}^T\mathbf{1}_n. \end{aligned} \tag{3}$$

Setting the derivative in (3) to $\mathbf{0}$ and substituting (2) to (3), we have $\{\mathbb{X}_1 - \mathbf{1}_n\boldsymbol{\mu}_1^T - (\mathbb{X}_2 - \mathbf{1}_n\bar{\mathbf{X}}_2^T)\boldsymbol{\gamma}\}^T\mathbf{1}_n = \mathbf{0}$. Because of $(\mathbb{X}_2 - \mathbf{1}_n\bar{\mathbf{X}}_2^T)^T\mathbf{1}_n = \mathbf{0}$, we have $\hat{\boldsymbol{\mu}}_1 = \bar{\mathbf{X}}_1$. Substitute $\hat{\boldsymbol{\mu}}_1 = \bar{\mathbf{X}}_1$ back to (2), and we obtain $\hat{\boldsymbol{\mu}}_{\mathbf{Y}} = \bar{\mathbf{Y}}$. We now substitute $\hat{\boldsymbol{\mu}}_1$ to the log likelihood function l in (1) and consider the derivative of l with respect to $\boldsymbol{\gamma}$

$$\frac{\partial l}{\partial \boldsymbol{\gamma}} = -\{\mathbb{X}_{2c}^T\mathbb{X}_{2c}\boldsymbol{\gamma} - \mathbb{X}_{2c}^T\mathbb{X}_{1c}\}\boldsymbol{\Sigma}_{1|2}^{-1}. \tag{4}$$

We set the derivative in (4) to be $\mathbf{0}$ and obtain $\hat{\boldsymbol{\gamma}} = (\mathbb{X}_{2c}^T\mathbb{X}_{2c})^{-1}\mathbb{X}_{2c}^T\mathbb{X}_{1c}$. Substitute $\hat{\boldsymbol{\mu}}_{\mathbf{Y}}$, $\hat{\boldsymbol{\mu}}_1$ and $\hat{\boldsymbol{\gamma}}$ to the log likelihood function l in (1), we have

$$\begin{aligned} l = & \frac{n(p_1+r)}{2}\log(2\pi) - \frac{1}{2}\operatorname{tr}\{\mathbf{Q}_{\mathbb{X}_{2c}}\mathbb{X}_{1c}(\boldsymbol{\Gamma}\boldsymbol{\Omega}\boldsymbol{\Gamma}^T + \boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^T)^{-1}\mathbb{X}_{1c}^T\mathbf{Q}_{\mathbb{X}_{2c}}\} \\ & -\frac{n}{2}\log|\boldsymbol{\Omega}| - \frac{n}{2}\log|\boldsymbol{\Omega}_0| - \frac{n}{2}\log|\boldsymbol{\Sigma}_{\mathbf{Y}|\mathbf{X}}| \\ & -\frac{1}{2}\operatorname{tr}[\{\mathbb{Y}_c - \mathbb{X}_{1c}\boldsymbol{\Gamma}\boldsymbol{\eta} - \mathbb{X}_{2c}\boldsymbol{\beta}_2\}\boldsymbol{\Sigma}_{\mathbf{Y}|\mathbf{X}}^{-1}\{\mathbb{Y}_c - \mathbb{X}_{1c}\boldsymbol{\Gamma}\boldsymbol{\eta} - \mathbb{X}_{2c}\boldsymbol{\beta}_2\}^T]. \end{aligned} \tag{5}$$

Taking the derivatives of l in (5) with respect to $\boldsymbol{\Omega}$ and $\boldsymbol{\Omega}_0$, we have

$$\frac{\partial l}{\partial \boldsymbol{\Omega}} = -\frac{1}{2}(n\boldsymbol{\Omega}^{-1} - \boldsymbol{\Omega}^{-1}\boldsymbol{\Gamma}^T\mathbb{X}_{1c}^T\mathbf{Q}_{\mathbb{X}_{2c}}\mathbb{X}_{1c}\boldsymbol{\Gamma}\boldsymbol{\Omega}^{-1}), \tag{6}$$

and

$$\frac{\partial l}{\partial \boldsymbol{\Omega}_0} = -\frac{1}{2}(n\boldsymbol{\Omega}_0^{-1} - \boldsymbol{\Omega}_0^{-1}\boldsymbol{\Gamma}_0^T\mathbb{X}_{1c}^T\mathbf{Q}_{\mathbb{X}_{2c}}\mathbb{X}_{1c}\boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0^{-1}). \tag{7}$$

We obtain the estimator of $\mathbf{\Omega}$ and $\mathbf{\Omega}_0$ by setting the derivatives in (6) and (7) to be $\mathbf{0}$, and the estimators are $\widehat{\mathbf{\Omega}} = (1/n)\mathbf{\Gamma}^T\mathbf{R}_{1|2}^T\mathbf{R}_{1|2}\mathbf{\Gamma}$ and $\widehat{\mathbf{\Omega}}_0 = (1/n)\mathbf{\Gamma}_0^T\mathbf{R}_{1|2}^T\mathbf{R}_{1|2}\mathbf{\Gamma}_0$. Thus, $\widehat{\mathbf{\Sigma}}_{1|2} = \mathbf{\Gamma}\widehat{\mathbf{\Omega}}\mathbf{\Gamma}^T + \mathbf{\Gamma}_0\widehat{\mathbf{\Omega}}_0\mathbf{\Gamma}_0^T$. Now we take the derivative of l in (5) with respect to $\boldsymbol{\beta}_2$ and set it to $\mathbf{0}$, i.e.

$$\frac{\partial l}{\partial \boldsymbol{\beta}_2} = -\{\mathbb{X}_{2c}^T\mathbb{X}_{2c}\boldsymbol{\beta}_2 - \mathbb{X}_{2c}^T(\mathbb{Y}_c - \mathbb{X}_{1c}\mathbf{\Gamma}\boldsymbol{\eta})\}\mathbf{\Sigma}_{\mathbf{Y}|\mathbf{X}}^{-1} = \mathbf{0}. \tag{8}$$

Substituting the equality (8) to (5), simplifying the expression of l, and then taking the derivative with respect to $\boldsymbol{\eta}$, we have

$$\frac{\partial l}{\partial \boldsymbol{\eta}} = -\mathbf{\Gamma}^T\mathbf{R}_{1|2}^T\mathbf{R}_{1|2}\mathbf{\Gamma}\boldsymbol{\eta}\mathbf{\Sigma}_{\mathbf{Y}|\mathbf{X}}^{-1} + \mathbf{\Gamma}^T\mathbf{R}_{1|2}^T\mathbf{R}_{\mathbf{Y}|2}\mathbf{\Sigma}_{\mathbf{Y}|\mathbf{X}}^{-1}. \tag{9}$$

We set the derivative in (9) to be $\mathbf{0}$ and obtain $\hat{\boldsymbol{\eta}} = (\mathbf{\Gamma}^T\mathbf{R}_{1|2}^T\mathbf{R}_{1|2}\mathbf{\Gamma})^{-1}\mathbf{\Gamma}^T\mathbf{R}_{1|2}^T\mathbf{R}_{\mathbf{Y}|2}$, then $\widehat{\boldsymbol{\beta}}_1 = \mathbf{\Gamma}\hat{\boldsymbol{\eta}}$. Substituting $\widehat{\boldsymbol{\beta}}_1$ to (8), we have $\widehat{\boldsymbol{\beta}}_2 = (\mathbb{X}_{2c}^T\mathbb{X}_{2c})^{-1}\mathbb{X}_{2c}^T(\mathbb{Y}_c - \mathbb{X}_{1c}\mathbf{\Gamma}\hat{\boldsymbol{\eta}})$. We now substitute $\hat{\boldsymbol{\mu}}_{\mathbf{Y}}$, $\hat{\boldsymbol{\mu}}_1$, $\hat{\boldsymbol{\gamma}}$, $\widehat{\mathbf{\Omega}}$, $\widehat{\mathbf{\Omega}}_0$, $\hat{\boldsymbol{\eta}}$, and $\widehat{\boldsymbol{\beta}}_2$ to the log likelihood function in (5), arrange the terms, and take the derivative of l with respect to $\mathbf{\Sigma}_{\mathbf{Y}|\mathbf{X}}$, then

$$\frac{\partial l}{\partial \mathbf{\Sigma}_{\mathbf{Y}|\mathbf{X}}} = -\frac{n}{2}\mathbf{\Sigma}_{\mathbf{Y}|\mathbf{X}}^{-1} + \frac{1}{2}\mathbf{\Sigma}_{\mathbf{Y}|\mathbf{X}}^{-1}\mathbf{R}_{\mathbf{Y}|2}^T\mathbf{Q}_{\mathbf{R}_{1|2}\mathbf{\Gamma}}\mathbf{R}_{\mathbf{Y}|2}\mathbf{\Sigma}_{\mathbf{Y}|\mathbf{X}}^{-1}. \tag{10}$$

Setting the derivative in (10) to be $\mathbf{0}$, we obtain $\widehat{\mathbf{\Sigma}}_{\mathbf{Y}|\mathbf{X}} = (1/n)\mathbf{R}_{\mathbf{Y}|2}^T\mathbf{Q}_{\mathbf{R}_{1|2}\mathbf{\Gamma}}\mathbf{R}_{\mathbf{Y}|2}$, which can be expressed as $\widehat{\mathbf{\Sigma}}_{\mathbf{Y}|\mathbf{X}} = \mathbf{S}_{\mathbf{Y}|2} - \mathbf{S}_{(\mathbf{Y},1)|2}\mathbf{\Gamma}(\mathbf{\Gamma}^T\mathbf{S}_{1|2}\mathbf{\Gamma})^{-1}\mathbf{\Gamma}^T\mathbf{S}_{(\mathbf{Y},1)|2}^T$.

Substitution of $\hat{\boldsymbol{\mu}}_{\mathbf{Y}}$, $\hat{\boldsymbol{\mu}}_1$, $\hat{\boldsymbol{\gamma}}$, $\widehat{\mathbf{\Omega}}$, $\widehat{\mathbf{\Omega}}_0$, $\hat{\boldsymbol{\eta}}$, and $\widehat{\boldsymbol{\beta}}_2$ into the log likelihood function l in (5) gives

$$\begin{aligned} l = & -\frac{n(p_1+r)}{2}(\log(2\pi)+1) - \frac{n}{2}\log|\mathbf{\Gamma}^T\mathbf{S}_{1|2}\mathbf{\Gamma}| \\ & -\frac{n}{2}\log|\mathbf{\Gamma}_0^T\mathbf{S}_{1|2}\mathbf{\Gamma}_0| - \frac{n}{2}\log|\mathbf{S}_{\mathbf{Y}|2} - \mathbf{S}_{(\mathbf{Y},1)|2}\mathbf{\Gamma}(\mathbf{\Gamma}^T\mathbf{S}_{1|2}\mathbf{\Gamma})^{-1}\mathbf{\Gamma}^T\mathbf{S}_{(\mathbf{Y},1)|2}^T|. \end{aligned}$$

By Lemma 6.2. of Cook et al. [1],

$$\begin{aligned} l = & -\frac{n(p_1+r)}{2}(\log(2\pi)+1) - \frac{n}{2}\log|\mathbf{S}_{1|2}| - \frac{n}{2}\log|\mathbf{\Gamma}^T\mathbf{S}_{1|2}\mathbf{\Gamma}| \\ & -\frac{n}{2}\log|\mathbf{\Gamma}^T\mathbf{S}_{1|2}^{-1}\mathbf{\Gamma}| - \frac{n}{2}\log|\mathbf{S}_{\mathbf{Y}|2} - \mathbf{S}_{(\mathbf{Y},1)|2}\mathbf{\Gamma}(\mathbf{\Gamma}^T\mathbf{S}_{1|2}\mathbf{\Gamma})^{-1}\mathbf{\Gamma}^T\mathbf{S}_{(\mathbf{Y},1)|2}^T|. \end{aligned} \tag{11}$$

Then, $\widehat{\mathbf{\Gamma}}$ can be obtained by minimizing

$$\log|\mathbf{\Gamma}^T\mathbf{S}_{1|2}\mathbf{\Gamma}| + \log|\mathbf{\Gamma}^T\mathbf{S}_{1|2}^{-1}\mathbf{\Gamma}| + \log|\mathbf{S}_{\mathbf{Y}|2} - \mathbf{S}_{(\mathbf{Y},1)|2}\mathbf{\Gamma}(\mathbf{\Gamma}^T\mathbf{S}_{1|2}\mathbf{\Gamma})^{-1}\mathbf{\Gamma}^T\mathbf{S}_{(\mathbf{Y},1)|2}^T|. \tag{12}$$

To obtain more insight on the above objective function, we notice that

$$\begin{aligned} & \log|\mathbf{\Gamma}^T\mathbf{S}_{1|2}\mathbf{\Gamma}| + \log|\mathbf{\Gamma}^T\mathbf{S}_{1|2}^{-1}\mathbf{\Gamma}| + \log|\mathbf{S}_{\mathbf{Y}|2} - \mathbf{S}_{(\mathbf{Y},1)|2}\mathbf{\Gamma}(\mathbf{\Gamma}^T\mathbf{S}_{1|2}\mathbf{\Gamma})^{-1}\mathbf{\Gamma}^T\mathbf{S}_{(\mathbf{Y},1)|2}^T| \\ = & \log|\mathbf{\Gamma}^T\mathbf{S}_{1|2}^{-1}\mathbf{\Gamma}| + \log|\mathbf{S}_{\mathbf{Y}|2}| + \log|\mathbf{\Gamma}^T\mathbf{S}_{1|2}\mathbf{\Gamma} - \mathbf{\Gamma}^T\mathbf{S}_{(\mathbf{Y},1)|2}^T\mathbf{S}_{\mathbf{Y}|2}^{-1}\mathbf{S}_{(\mathbf{Y},1)|2}\mathbf{\Gamma}| \\ = & \log|\mathbf{S}_{\mathbf{Y}|2}| + \log|\mathbf{\Gamma}^T\mathbf{S}_{1|2}^{-1}\mathbf{\Gamma}| + \log|\mathbf{\Gamma}^T\mathbf{S}_{1|(\mathbf{Y},2)}\mathbf{\Gamma}|, \end{aligned}$$

where the first equality is because of Sylvester's determinant theorem, and $\mathbf{S}_{1|(\mathbf{Y},2)} = \mathbf{S}_{1|2} - \mathbf{S}_{(\mathbf{Y},1)|2}^T\mathbf{S}_{\mathbf{Y}|2}^{-1}\mathbf{S}_{(\mathbf{Y},1)|2}$. Then the optimization (12) is equivalent to the optimization

$$\widehat{\mathcal{E}}_{\mathbf{\Sigma}_{1|2}}(\boldsymbol{\beta}_1) = \underset{\mathrm{span}(\mathbf{\Gamma})\in\mathcal{L}_{p_1\times d}}{\arg\min}\{\log|\mathbf{S}_{\mathbf{Y}|2}| + \log|\mathbf{\Gamma}^T\mathbf{S}_{1|2}^{-1}\mathbf{\Gamma}| + \log|\mathbf{\Gamma}^T\mathbf{S}_{1|(\mathbf{Y},2)}\mathbf{\Gamma}|\}. \tag{13}$$

B. PROOFS OF THEORETICAL RESULTS

Proof of Proposition 1

Let $\mathbf{E}_m \in \mathbb{R}^{m^2\times m(m+1)/2}$ and $\mathbf{C}_m \in \mathbb{R}^{m(m+1)/2\times m^2}$ denote the expansion and contraction operators that connect vec and vech operator, i.e. $\mathrm{vec}(\mathbf{A}) = \mathbf{E}_m\mathrm{vech}(\mathbf{A})$ and $\mathrm{vech}(\mathbf{A}) = \mathbf{C}_m\mathrm{vec}(\mathbf{A})$. Since $\mathbf{\Gamma}$ in the EPPLS (9) is over-parametrized, we use Proposition

4.1 in[3] to derive the asymptotic distribution of the EPPLS estimator. Note that the gradient matrix $\boldsymbol{\Delta} = \partial \mathbf{h}(\boldsymbol{\phi})/\partial \boldsymbol{\phi}^T$ is

$$\boldsymbol{\Delta} = \begin{pmatrix} \mathbf{I}_r & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{p_1} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{g}_1 & \mathbf{g}_2 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}_{p_2 r} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}_{p_1 p_2} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{g}_3 & \mathbf{0} & \mathbf{0} & \mathbf{g}_4 & \mathbf{g}_5 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}_{r(r+1)/2} \end{pmatrix}$$

where $\mathbf{g}_1 = \mathbf{I}_r \otimes \boldsymbol{\Gamma}$, $\mathbf{g}_2 = \boldsymbol{\eta}^T \otimes \mathbf{I}_{p_1}$, $\mathbf{g}_3 = 2\mathbf{C}_{p_1}(\boldsymbol{\Gamma}\boldsymbol{\Omega} \otimes \mathbf{I}_{p_1} - \boldsymbol{\Gamma} \otimes \boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^T)$, $\mathbf{g}_4 = \mathbf{C}_{p_1}(\boldsymbol{\Gamma} \otimes \boldsymbol{\Gamma})\mathbf{E}_d$, $\mathbf{g}_5 = \mathbf{C}_{p_1}(\boldsymbol{\Gamma}_0 \otimes \boldsymbol{\Gamma}_0)\mathbf{E}_{p_1-d}$. We define Shapiro's discrepancy function to be $l_{\max} - l$, where l is the objective function defined in (1) and $l_{\max}$ is the maximum value of l evaluated at the standard estimator of $\mathbf{h}$. Since l is derived from normal likelihood, it is easy to check that $l_{\max} - l$ satisfies the four conditions in Section 3 of Shapiro[3]. Since $\mathbf{V}$ is a full rank matrix, $\text{rank}(\boldsymbol{\Delta}^T\mathbf{V}\boldsymbol{\Delta}) = \text{rank}(\boldsymbol{\Delta}^T\boldsymbol{\Delta}) = \text{rank}(\boldsymbol{\Delta})$. Therefore, all the assumptions of Proposition 4.1 in Shapiro[3] are satisfied, using Proposition 4.1, we have

$$\sqrt{n}(\widehat{\mathbf{h}} - \mathbf{h}) \xrightarrow{d} N(\mathbf{0}, \mathbf{U}), \qquad \mathbf{U} = \boldsymbol{\Delta}(\boldsymbol{\Delta}^T\mathbf{V}\boldsymbol{\Delta})^{\dagger}\boldsymbol{\Delta}.$$

Now since

$$\mathbf{V}^{-1} - \mathbf{U} = \mathbf{V}^{-1} - \mathbf{V}^{-1/2}[\mathbf{V}^{1/2}\boldsymbol{\Delta}(\boldsymbol{\Delta}^T\mathbf{V}\boldsymbol{\Delta})^{\dagger}\boldsymbol{\Delta}\mathbf{V}^{1/2}]\mathbf{V}^{-1/2} = \mathbf{V}^{-1/2}(\mathbf{I} - \mathbf{P}_{\mathbf{V}^{1/2}\boldsymbol{\Delta}})\mathbf{V}^{-1/2},$$

and $\mathbf{I} - \mathbf{P}_{\mathbf{V}^{1/2}\boldsymbol{\Delta}}$ is a positive semi-definite matrix, then $\mathbf{V}^{-1} - \mathbf{U}$ is a positive semi-definite matrix.

Proof of Proposition 2

When population residuals are normally distributed, the Fisher information matrix $\mathbf{V}$ has a closed form. We first derive the closed form as follows. Given the observed data, we have the following log likelihood function

$$\begin{aligned} l = & -\frac{n(r+p_1)}{2}\log(2\pi) - \frac{n}{2}\log|\boldsymbol{\Sigma}_{\mathbf{Y}|\mathbf{X}}| - \frac{1}{2}\text{tr}\left[\{\mathbb{Y} - \mathbf{1}_n\boldsymbol{\mu}_{\mathbf{Y}}^T - (\mathbb{X}_1 - \mathbf{1}_n\boldsymbol{\mu}_1^T)\boldsymbol{\beta}_1 - (\mathbb{X}_2 - \mathbf{1}_n\bar{\mathbf{X}}_2^T)\boldsymbol{\beta}_2\}\right. \\ & \left.\times\boldsymbol{\Sigma}_{\mathbf{Y}|\mathbf{X}}^{-1}\{\mathbb{Y} - \mathbf{1}_n\boldsymbol{\mu}_{\mathbf{Y}}^T - (\mathbb{X}_1 - \mathbf{1}_n\boldsymbol{\mu}_1^T)\boldsymbol{\beta}_1 - (\mathbb{X}_2 - \mathbf{1}_n\bar{\mathbf{X}}_2^T)\boldsymbol{\beta}_2\}^T\right] - \frac{n}{2}\log|\boldsymbol{\Sigma}_{1|2}| \\ & -\frac{1}{2}\text{tr}\left[\{\mathbb{X}_1 - \mathbf{1}_n\boldsymbol{\mu}_1^T - (\mathbb{X}_2 - \mathbf{1}_n\bar{\mathbf{X}}_2^T)\boldsymbol{\gamma}\}\boldsymbol{\Sigma}_{1|2}^{-1}\{\mathbb{X}_1 - \mathbf{1}_n\boldsymbol{\mu}_1^T - (\mathbb{X}_2 - \mathbf{1}_n\bar{\mathbf{X}}_2^T)\boldsymbol{\gamma}\}^T\right]. \end{aligned}$$

We first take first and second derivatives of log likelihood function l with respect to the parameters. We have

$$\begin{aligned} \frac{\partial l}{\partial \boldsymbol{\mu}_{\mathbf{Y}}^T} &= n\{\bar{\mathbf{Y}}^T - \boldsymbol{\mu}_{\mathbf{Y}}^T - (\bar{\mathbf{X}}_1 - \boldsymbol{\mu}_1)^T\boldsymbol{\beta}_1\}\boldsymbol{\Sigma}_{\mathbf{Y}|\mathbf{X}}^{-1}, \\ \frac{\partial^2 l}{\partial \boldsymbol{\mu}_{\mathbf{Y}}\partial \boldsymbol{\mu}_{\mathbf{Y}}^T} &= -n\boldsymbol{\Sigma}_{\mathbf{Y}|\mathbf{X}}^{-1}, \\ \frac{\partial^2 l}{\partial \boldsymbol{\mu}_1\partial \boldsymbol{\mu}_{\mathbf{Y}}^T} &= n\boldsymbol{\beta}_1\boldsymbol{\Sigma}_{\mathbf{Y}|\mathbf{X}}^{-1}, \\ \frac{\partial^2 l}{\partial \text{vec}(\boldsymbol{\beta}_1)\partial \boldsymbol{\mu}_{\mathbf{Y}}^T} &= -n\boldsymbol{\Sigma}_{\mathbf{Y}|\mathbf{X}}^{-1} \otimes (\bar{\mathbf{X}}_1 - \boldsymbol{\mu}_1), \\ \frac{\partial^2 l}{\partial \text{vec}(\boldsymbol{\beta}_2)\partial \boldsymbol{\mu}_{\mathbf{Y}}^T} &= \mathbf{0}, \quad \frac{\partial^2 l}{\partial \text{vec}(\boldsymbol{\gamma})\partial \boldsymbol{\mu}_{\mathbf{Y}}^T} = \mathbf{0}, \quad \frac{\partial^2 l}{\partial \text{vech}(\boldsymbol{\Sigma}_{1|2})\partial \boldsymbol{\mu}_{\mathbf{Y}}^T} = \mathbf{0}, \\ \frac{\partial^2 l}{\partial \text{vech}(\boldsymbol{\Sigma}_{\mathbf{Y}|\mathbf{X}})\partial \boldsymbol{\mu}_{\mathbf{Y}}^T} &= -n\mathbf{E}_r^T(\boldsymbol{\Sigma}_{\mathbf{Y}|\mathbf{X}}^{-1} \otimes \boldsymbol{\Sigma}_{\mathbf{Y}|\mathbf{X}}^{-1})[\mathbf{I}_r \otimes \{\bar{\mathbf{Y}} - \boldsymbol{\mu}_{\mathbf{Y}} - \boldsymbol{\beta}_1^T(\bar{\mathbf{X}} - \boldsymbol{\mu}_1)\}], \end{aligned}$$

$$
\begin{aligned}
\frac{\partial l}{\partial \boldsymbol{\mu}_1^T} &= -n\{\bar{\mathbf{Y}}^T - \boldsymbol{\mu}_{\mathbf{Y}}^T - (\bar{\mathbf{X}}_1 - \boldsymbol{\mu}_1)^T \boldsymbol{\beta}_1\}\boldsymbol{\Sigma}_{\mathbf{Y}|\mathbf{X}}^{-1}\boldsymbol{\beta}_1^T + n(\bar{\mathbf{X}}_1 - \boldsymbol{\mu}_1)^T\boldsymbol{\Sigma}_{1|2}^{-1},\\
\frac{\partial^2 l}{\partial \boldsymbol{\mu}_1 \partial \boldsymbol{\mu}_1^T} &= -n\boldsymbol{\beta}_1\boldsymbol{\Sigma}_{\mathbf{Y}|\mathbf{X}}^{-1}\boldsymbol{\beta}_1^T - n\boldsymbol{\Sigma}_{1|2}^{-1},\\
\frac{\partial^2 l}{\partial \mathrm{vec}(\boldsymbol{\beta}_1)\partial \boldsymbol{\mu}_1^T} &= n[\boldsymbol{\Sigma}_{\mathbf{Y}|\mathbf{X}}^{-1}\boldsymbol{\beta}_1^T \otimes (\bar{\mathbf{X}}_1 - \boldsymbol{\mu}_1) - \boldsymbol{\Sigma}_{\mathbf{Y}|\mathbf{X}}^{-1}\{\bar{\mathbf{Y}} - \boldsymbol{\mu}_{\mathbf{Y}} - \boldsymbol{\beta}_1^T(\bar{\mathbf{X}}_1 - \boldsymbol{\mu}_1)\} \otimes \mathbf{I}_{p_1}],\\
\frac{\partial^2 l}{\partial \mathrm{vec}(\boldsymbol{\beta}_2)\partial \boldsymbol{\mu}_1^T} &= \mathbf{0}, \quad \frac{\partial^2 l}{\partial \mathrm{vec}(\boldsymbol{\gamma})\partial \boldsymbol{\mu}_1^T} = \mathbf{0},\\
\frac{\partial^2 l}{\partial \mathrm{vech}(\boldsymbol{\Sigma}_{1|2})\partial \boldsymbol{\mu}_1^T} &= -n\mathbf{E}_{p_1}^T(\boldsymbol{\Sigma}_{1|2}^{-1} \otimes \boldsymbol{\Sigma}_{1|2}^{-1})\{\mathbf{I}_{p_1} \otimes (\bar{\mathbf{X}}_1 - \boldsymbol{\mu}_1)\},\\
\frac{\partial^2 l}{\partial \mathrm{vech}(\boldsymbol{\Sigma}_{\mathbf{Y}|\mathbf{X}})\partial \boldsymbol{\mu}_1^T} &= n\mathbf{E}_r^T(\boldsymbol{\Sigma}_{\mathbf{Y}|\mathbf{X}}^{-1} \otimes \boldsymbol{\Sigma}_{\mathbf{Y}|\mathbf{X}}^{-1})[\boldsymbol{\beta}_1^T \otimes \{\bar{\mathbf{Y}} - \boldsymbol{\mu}_{\mathbf{Y}} - \boldsymbol{\beta}_1^T(\bar{\mathbf{X}}_1 - \boldsymbol{\mu}_1)\}].
\end{aligned}
$$

Let $\tilde{\boldsymbol{S}}_{1\mathbf{Y}} = (\mathbb{X}_1 - \mathbf{1}_n\boldsymbol{\mu}_1^T)^T(\mathbb{Y} - \mathbf{1}_n\boldsymbol{\mu}_{\mathbf{Y}}^T)/n$, $\tilde{\boldsymbol{S}}_1 = (\mathbb{X}_1 - \mathbf{1}_n\boldsymbol{\mu}_1^T)^T(\mathbb{X}_1 - \mathbf{1}_n\boldsymbol{\mu}_1^T)/n$, and $\tilde{\boldsymbol{S}}_{12} = (\mathbb{X}_1 - \mathbf{1}_n\boldsymbol{\mu}_1^T)^T(\mathbb{X}_2 - \mathbf{1}_n\bar{\mathbf{X}}_2^T)/n$. Then

$$
\begin{aligned}
\frac{\partial l}{\partial \boldsymbol{\beta}_1} &= n\big(\tilde{\boldsymbol{S}}_{1\mathbf{Y}} - \tilde{\boldsymbol{S}}_1\boldsymbol{\beta}_1 - \tilde{\boldsymbol{S}}_{12}\boldsymbol{\beta}_2\big)\boldsymbol{\Sigma}_{\mathbf{Y}|\mathbf{X}}^{-1},\\
\frac{\partial l}{\partial \mathrm{vec}(\boldsymbol{\beta}_1)} &= n\mathrm{vec}\big\{\big(\tilde{\boldsymbol{S}}_{1\mathbf{Y}} - \tilde{\boldsymbol{S}}_1\boldsymbol{\beta}_1 - \tilde{\boldsymbol{S}}_{12}\boldsymbol{\beta}_2\big)\boldsymbol{\Sigma}_{\mathbf{Y}|\mathbf{X}}^{-1}\big\},\\
\frac{\partial^2 l}{\partial \mathrm{vec}^T(\boldsymbol{\beta}_1)\partial \mathrm{vec}(\boldsymbol{\beta}_1)} &= -n(\boldsymbol{\Sigma}_{\mathbf{Y}|\mathbf{X}}^{-1} \otimes \tilde{\boldsymbol{S}}_1),\\
\frac{\partial^2 l}{\partial \mathrm{vec}^T(\boldsymbol{\beta}_2)\partial \mathrm{vec}(\boldsymbol{\beta}_1)} &= -n(\boldsymbol{\Sigma}_{\mathbf{Y}|\mathbf{X}}^{-1} \otimes \tilde{\boldsymbol{S}}_{12}),\\
\frac{\partial^2 l}{\partial \mathrm{vec}^T(\boldsymbol{\gamma})\partial \mathrm{vec}(\boldsymbol{\beta}_1)} &= \mathbf{0}, \quad \frac{\partial^2 l}{\partial \mathrm{vech}^T(\boldsymbol{\Sigma}_{1|2})\partial \mathrm{vec}(\boldsymbol{\beta}_1)} = \mathbf{0},\\
\frac{\partial^2 l}{\partial \mathrm{vech}^T(\boldsymbol{\Sigma}_{\mathbf{Y}|\mathbf{X}})\partial \mathrm{vec}(\boldsymbol{\beta}_1)} &= -n\big\{\mathbf{I}_r \otimes \big(\tilde{\boldsymbol{S}}_{1\mathbf{Y}} - \tilde{\boldsymbol{S}}_1\boldsymbol{\beta}_1 - \tilde{\boldsymbol{S}}_{12}\boldsymbol{\beta}_2\big)\big\}(\boldsymbol{\Sigma}_{\mathbf{Y}|\mathbf{X}}^{-1} \otimes \boldsymbol{\Sigma}_{\mathbf{Y}|\mathbf{X}}^{-1})\mathbf{E}_r.
\end{aligned}
$$

Let $\tilde{\boldsymbol{S}}_{2\mathbf{Y}} = (\mathbb{X}_2 - \mathbf{1}_n\bar{\mathbf{X}}_2^T)^T(\mathbb{Y} - \mathbf{1}_n\boldsymbol{\mu}_{\mathbf{Y}}^T)/n$, $\tilde{\boldsymbol{S}}_2 = (\mathbb{X}_2 - \mathbf{1}_n\bar{\mathbf{X}}_2^T)^T(\mathbb{X}_2 - \mathbf{1}_n\bar{\mathbf{X}}_2^T)/n$, and $\tilde{\boldsymbol{S}}_{21} = (\mathbb{X}_2 - \mathbf{1}_n\bar{\mathbf{X}}_2^T)^T(\mathbb{X}_1 - \mathbf{1}_n\boldsymbol{\mu}_1^T)/n$. Then

$$
\begin{aligned}
\frac{\partial l}{\partial \boldsymbol{\beta}_2} &= n\big(\tilde{\boldsymbol{S}}_{2\mathbf{Y}} - \tilde{\boldsymbol{S}}_{21}\boldsymbol{\beta}_1 - \tilde{\boldsymbol{S}}_2\boldsymbol{\beta}_2\big)\boldsymbol{\Sigma}_{\mathbf{Y}|\mathbf{X}}^{-1},\\
\frac{\partial l}{\partial \mathrm{vec}(\boldsymbol{\beta}_2)} &= n\mathrm{vec}\big\{\big(\tilde{\boldsymbol{S}}_{2\mathbf{Y}} - \tilde{\boldsymbol{S}}_{21}\boldsymbol{\beta}_1 - \tilde{\boldsymbol{S}}_2\boldsymbol{\beta}_2\big)\boldsymbol{\Sigma}_{\mathbf{Y}|\mathbf{X}}^{-1}\big\},\\
\frac{\partial^2 l}{\partial \mathrm{vec}^T(\boldsymbol{\beta}_2)\partial \mathrm{vec}(\boldsymbol{\beta}_2)} &= -n(\boldsymbol{\Sigma}_{\mathbf{Y}|\mathbf{X}}^{-1} \otimes \tilde{\boldsymbol{S}}_2),\\
\frac{\partial^2 l}{\partial \mathrm{vec}^T(\boldsymbol{\gamma})\partial \mathrm{vec}(\boldsymbol{\beta}_2)} &= \mathbf{0}, \quad \frac{\partial^2 l}{\partial \mathrm{vech}^T(\boldsymbol{\Sigma}_{1|2})\partial \mathrm{vec}(\boldsymbol{\beta}_2)} = \mathbf{0},\\
\frac{\partial^2 l}{\partial \mathrm{vech}^T(\boldsymbol{\Sigma}_{\mathbf{Y}|\mathbf{X}})\partial \mathrm{vec}(\boldsymbol{\beta}_2)} &= -n\big\{\mathbf{I}_r \otimes \big(\tilde{\boldsymbol{S}}_{2\mathbf{Y}} - \tilde{\boldsymbol{S}}_{21}\boldsymbol{\beta}_1 - \tilde{\boldsymbol{S}}_2\boldsymbol{\beta}_2\big)\boldsymbol{\Sigma}_{\mathbf{Y}|\mathbf{X}}^{-1}\big\}(\boldsymbol{\Sigma}_{\mathbf{Y}|\mathbf{X}}^{-1} \otimes \boldsymbol{\Sigma}_{\mathbf{Y}|\mathbf{X}}^{-1})\mathbf{E}_r,
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial l}{\partial \boldsymbol{\gamma}} &= n\big(\tilde{\boldsymbol{S}}_{21} - \tilde{\boldsymbol{S}}_2\boldsymbol{\gamma}\big)\boldsymbol{\Sigma}_{1|2}^{-1},\\
\frac{\partial l}{\partial \mathrm{vec}(\boldsymbol{\gamma})} &= n\mathrm{vec}\big\{\big(\tilde{\boldsymbol{S}}_{21} - \tilde{\boldsymbol{S}}_2\boldsymbol{\gamma}\big)\boldsymbol{\Sigma}_{1|2}^{-1}\big\},\\
\frac{\partial^2 l}{\partial \mathrm{vec}^T(\boldsymbol{\gamma})\partial \mathrm{vec}(\boldsymbol{\gamma})} &= -n(\boldsymbol{\Sigma}_{1|2}^{-1} \otimes \tilde{\boldsymbol{S}}_2),\\
\frac{\partial^2 l}{\partial \mathrm{vech}^T(\boldsymbol{\Sigma}_{1|2})\partial \mathrm{vec}(\boldsymbol{\gamma})} &= -n\big\{\mathbf{I}_{p_1} \otimes (\tilde{\boldsymbol{S}}_{21} - \tilde{\boldsymbol{S}}_2\boldsymbol{\gamma})\big\}(\boldsymbol{\Sigma}_{1|2}^{-1} \otimes \boldsymbol{\Sigma}_{1|2}^{-1})\mathbf{E}_{p_1},\\
\frac{\partial^2 l}{\partial \mathrm{vech}^T(\boldsymbol{\Sigma}_{\mathbf{Y}|\mathbf{X}})\partial \mathrm{vec}(\boldsymbol{\gamma})} &= \mathbf{0}.
\end{aligned}
$$

Let $\mathbf{R}_{1|2} = \mathbb{X}_1 - \mathbf{1}_n\boldsymbol{\mu}_1^T - (\mathbb{X}_2 - \mathbf{1}_n\bar{\mathbf{X}}_2^T)\boldsymbol{\gamma}$. Then

$$\begin{aligned}
\frac{\partial l}{\partial \mathrm{vech}(\boldsymbol{\Sigma}_{1|2})} &= -\frac{1}{2}\mathbf{E}_{p_1}^T \mathrm{vec}\left(n\boldsymbol{\Sigma}_{1|2}^{-1} - \boldsymbol{\Sigma}_{1|2}^{-1}\mathbf{R}_{1|2}^T\mathbf{R}_{1|2}\boldsymbol{\Sigma}_{1|2}^{-1}\right),\\
\frac{\partial^2 l}{\partial \mathrm{vech}^T(\boldsymbol{\Sigma}_{1|2})\partial \mathrm{vech}(\boldsymbol{\Sigma}_{1|2})} &= -\frac{1}{2}\mathbf{E}_{p_1}^T\left\{(\mathbf{I}_{p_1} \otimes \boldsymbol{\Sigma}_{1|2}^{-1}\mathbf{R}_{1|2}^T\mathbf{R}_{1|2}) + (\boldsymbol{\Sigma}_{1|2}^{-1}\mathbf{R}_{1|2}^T\mathbf{R}_{1|2} \otimes \mathbf{I}_{p_1})\right.\\
&\quad \left. - n\mathbf{I}_{p_1} \otimes \mathbf{I}_{p_1}\right\}(\boldsymbol{\Sigma}_{1|2}^{-1} \otimes \boldsymbol{\Sigma}_{1|2}^{-1})\mathbf{E}_{p_1},\\
\frac{\partial^2 l}{\partial \mathrm{vech}^T(\boldsymbol{\Sigma}_{\mathbf{Y}|\mathbf{X}})\partial \mathrm{vech}(\boldsymbol{\Sigma}_{1|2})} &= \mathbf{0}.
\end{aligned}$$

Let $\mathbf{R}_{\mathbf{Y}|\mathbf{X}} = \mathbb{Y} - \mathbf{1}_n\boldsymbol{\mu}_{\mathbf{Y}}^T - (\mathbb{X}_1 - \mathbf{1}_n\boldsymbol{\mu}_1^T)\boldsymbol{\beta}_1 - (\mathbb{X}_2 - \mathbf{1}_n\bar{\mathbf{X}}_2^T)\boldsymbol{\beta}_2$. Then,

$$\begin{aligned}
\frac{\partial l}{\partial \mathrm{vech}(\boldsymbol{\Sigma}_{\mathbf{Y}|\mathbf{X}})} &= -\frac{1}{2}\mathbf{E}_r^T \mathrm{vec}\left(n\boldsymbol{\Sigma}_{\mathbf{Y}|\mathbf{X}}^{-1} - \boldsymbol{\Sigma}_{\mathbf{Y}|\mathbf{X}}^{-1}\mathbf{R}_{\mathbf{Y}|\mathbf{X}}^T\mathbf{R}_{\mathbf{Y}|\mathbf{X}}\boldsymbol{\Sigma}_{\mathbf{Y}|\mathbf{X}}^{-1}\right),\\
\frac{\partial^2 l}{\partial \mathrm{vech}^T(\boldsymbol{\Sigma}_{\mathbf{Y}|\mathbf{X}})\partial \mathrm{vech}(\boldsymbol{\Sigma}_{\mathbf{Y}|\mathbf{X}})} &= -\frac{1}{2}\mathbf{E}_r^T\left\{(\mathbf{I}_r \otimes \boldsymbol{\Sigma}_{\mathbf{Y}|\mathbf{X}}^{-1}\mathbf{R}_{\mathbf{Y}|\mathbf{X}}^T\mathbf{R}_{\mathbf{Y}|\mathbf{X}}) + (\boldsymbol{\Sigma}_{\mathbf{Y}|\mathbf{X}}^{-1}\mathbf{R}_{\mathbf{Y}|\mathbf{X}}^T\mathbf{R}_{\mathbf{Y}|\mathbf{X}} \otimes \mathbf{I}_r)\right.\\
&\quad \left. - n\mathbf{I}_r \otimes \mathbf{I}_r\right\}(\boldsymbol{\Sigma}_{\mathbf{Y}|\mathbf{X}}^{-1} \otimes \boldsymbol{\Sigma}_{\mathbf{Y}|\mathbf{X}}^{-1})\mathbf{E}_r.
\end{aligned}$$

By taking expectation of the second derivatives, we obtain

$$\mathbf{V} = \begin{pmatrix}
\boldsymbol{\Sigma}_{\mathbf{Y}|\mathbf{X}}^{-1} & -\boldsymbol{\Sigma}_{\mathbf{Y}|\mathbf{X}}^{-1}\boldsymbol{\beta}_1^T & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0}\\
-\boldsymbol{\beta}_1\boldsymbol{\Sigma}_{\mathbf{Y}|\mathbf{X}}^{-1} & \boldsymbol{\beta}_1\boldsymbol{\Sigma}_{\mathbf{Y}|\mathbf{X}}^{-1}\boldsymbol{\beta}_1^T + \boldsymbol{\Sigma}_{1|2}^{-1} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0}\\
\mathbf{0} & \mathbf{0} & \boldsymbol{\Sigma}_{\mathbf{Y}|\mathbf{X}}^{-1} \otimes \boldsymbol{\Sigma}_1 & \boldsymbol{\Sigma}_{\mathbf{Y}|\mathbf{X}}^{-1} \otimes \boldsymbol{\Sigma}_{12} & \mathbf{0} & \mathbf{0} & \mathbf{0}\\
\mathbf{0} & \mathbf{0} & \boldsymbol{\Sigma}_{\mathbf{Y}|\mathbf{X}}^{-1} \otimes \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{\mathbf{Y}|\mathbf{X}}^{-1} \otimes \boldsymbol{\Sigma}_2 & \mathbf{0} & \mathbf{0} & \mathbf{0}\\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{\Sigma}_{1|2}^{-1} \otimes \boldsymbol{\Sigma}_2 & \mathbf{0} & \mathbf{0}\\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \frac{1}{2}\mathbf{E}_{p_1}^T(\boldsymbol{\Sigma}_{1|2}^{-1} \otimes \boldsymbol{\Sigma}_{1|2}^{-1})\mathbf{E}_{p_1} & \mathbf{0}\\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \frac{1}{2}\mathbf{E}_r^T(\boldsymbol{\Sigma}_{\mathbf{Y}|\mathbf{X}}^{-1} \otimes \boldsymbol{\Sigma}_{\mathbf{Y}|\mathbf{X}}^{-1})\mathbf{E}_r
\end{pmatrix}$$

We consider the following partition of the gradient matrix $\boldsymbol{\Delta}$ (displayed in the proof of Proposition 1), and the Fisher information matrix $\mathbf{V}$ into submatrices

$$\boldsymbol{\Delta} = \begin{pmatrix} \boldsymbol{\Delta}_{\mathbf{I}} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{\Delta}^* \end{pmatrix} \quad \text{and} \quad \mathbf{V} = \begin{pmatrix} \mathbf{V}_b & \mathbf{0}\\ \mathbf{0} & \mathbf{V}^* \end{pmatrix},$$

where

$$\boldsymbol{\Delta}_{\mathbf{I}} = \begin{pmatrix}
\mathbf{I}_r & \mathbf{0} & \mathbf{0} & \mathbf{0}\\
\mathbf{0} & \mathbf{I}_{p_1} & \mathbf{0} & \mathbf{0}\\
\mathbf{0} & \mathbf{0} & \mathbf{I}_{p_1 p_2} & \mathbf{0}\\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}_{r(r+1)/2}
\end{pmatrix},$$

$$\boldsymbol{\Delta}^* = \begin{pmatrix}
\mathbf{I}_r \otimes \boldsymbol{\Gamma} & \boldsymbol{\eta}^T \otimes \mathbf{I}_{p_1} & \mathbf{0} & \mathbf{0} & \mathbf{0}\\
\mathbf{0} & \mathbf{0} & \mathbf{I}_{p_2 r} & \mathbf{0} & \mathbf{0}\\
\mathbf{0} & 2\mathbf{C}_{p_1}(\boldsymbol{\Gamma}\boldsymbol{\Omega} \otimes \mathbf{I}_{p_1} - \boldsymbol{\Gamma} \otimes \boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^T) & \mathbf{0} & \mathbf{C}_{p_1}(\boldsymbol{\Gamma} \otimes \boldsymbol{\Gamma})\mathbf{E}_d & \mathbf{C}_{p_1}(\boldsymbol{\Gamma}_0 \otimes \boldsymbol{\Gamma}_0)\mathbf{E}_{p_1 - d}
\end{pmatrix}$$

$$\mathbf{V}_b = \begin{pmatrix}
\boldsymbol{\Sigma}_{\mathbf{Y}|\mathbf{X}}^{-1} & -\boldsymbol{\Sigma}_{\mathbf{Y}|\mathbf{X}}^{-1}\boldsymbol{\beta}_1^T & \mathbf{0} & \mathbf{0}\\
-\boldsymbol{\beta}_1\boldsymbol{\Sigma}_{\mathbf{Y}|\mathbf{X}}^{-1} & \boldsymbol{\beta}_1\boldsymbol{\Sigma}_{\mathbf{Y}|\mathbf{X}}^{-1}\boldsymbol{\beta}_1^T + \boldsymbol{\Sigma}_{1|2}^{-1} & \mathbf{0} & \mathbf{0}\\
\mathbf{0} & \mathbf{0} & \boldsymbol{\Sigma}_{1|2}^{-1} \otimes \boldsymbol{\Sigma}_2 & \mathbf{0}\\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \frac{1}{2}\mathbf{E}_r^T(\boldsymbol{\Sigma}_{\mathbf{Y}|\mathbf{X}}^{-1} \otimes \boldsymbol{\Sigma}_{\mathbf{Y}|\mathbf{X}}^{-1})\mathbf{E}_r
\end{pmatrix}$$

and

$$\mathbf{V}^* = \begin{pmatrix}
\boldsymbol{\Sigma}_{\mathbf{Y}|\mathbf{X}}^{-1} \otimes \boldsymbol{\Sigma}_1 & \boldsymbol{\Sigma}_{\mathbf{Y}|\mathbf{X}}^{-1} \otimes \boldsymbol{\Sigma}_{12} & \mathbf{0}\\
\boldsymbol{\Sigma}_{\mathbf{Y}|\mathbf{X}}^{-1} \otimes \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{\mathbf{Y}|\mathbf{X}}^{-1} \otimes \boldsymbol{\Sigma}_2 & \mathbf{0}\\
\mathbf{0} & \mathbf{0} & \frac{1}{2}\mathbf{E}_{p_1}^T(\boldsymbol{\Sigma}_{1|2}^{-1} \otimes \boldsymbol{\Sigma}_{1|2}^{-1})\mathbf{E}_{p_1}
\end{pmatrix}.$$

The asymptotic covariance matrix of $(\mathrm{vec}^T(\boldsymbol{\beta}_1), \mathrm{vec}^T(\boldsymbol{\beta}_2), \mathrm{vech}^T(\boldsymbol{\Sigma}_{1|2}))^T$ is then

$$\mathbf{W} \equiv \boldsymbol{\Delta}^*\{(\boldsymbol{\Delta}^*)^T\mathbf{V}^*\boldsymbol{\Delta}^*\}^\dagger(\boldsymbol{\Delta}^*)^T.$$

We notice that $\mathbf{\Delta}^*$ does not have full rank, and write $\mathbf{\Delta}^* = \mathbf{G}_1\mathbf{G}_2$, where

$$\mathbf{G}_1 = \begin{pmatrix} \mathbf{I}_r \otimes \mathbf{\Gamma} & \boldsymbol{\eta}^T \otimes \mathbf{\Gamma}_0 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_{p_2 r} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & 2\mathbf{C}_{p_1}(\mathbf{\Gamma\Omega} \otimes \mathbf{\Gamma}_0 - \mathbf{\Gamma} \otimes \mathbf{\Gamma}_0\mathbf{\Omega}_0) & \mathbf{0} & \mathbf{C}_{p_1}(\mathbf{\Gamma} \otimes \mathbf{\Gamma})\mathbf{E}_d & \mathbf{C}_{p_1}(\mathbf{\Gamma}_0 \otimes \mathbf{\Gamma}_0)\mathbf{E}_{p_1-d} \end{pmatrix}$$

and

$$\mathbf{G}_2 = \begin{pmatrix} \mathbf{I}_{rd} & \boldsymbol{\eta}^T \otimes \mathbf{\Gamma}^T & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_d \otimes \mathbf{\Gamma}_0^T & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_{p_2 r} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & 2\mathbf{C}_d(\mathbf{\Omega} \otimes \mathbf{\Gamma}^T) & \mathbf{0} & \mathbf{I}_{d(d+1)/2} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}_{(p_1-d)(p_1-d+1)/2} \end{pmatrix}$$

Note that $\mathbf{G}_1$ has full column rank and $\mathbf{G}_2$ has full row rank. We have rank$\{(\mathbf{\Delta}^*)^T\mathbf{V}^*\mathbf{\Delta}^*\}$ =rank $(\mathbf{G}_1^T\mathbf{V}^*\mathbf{G}_1) = \text{rank}(\mathbf{G}_1)$ and $\mathbf{W} = \mathbf{G}_1(\mathbf{G}_1^T\mathbf{V}^*\mathbf{G}_1)^{-1}\mathbf{G}_1^T$. After some straightforward calculations, we have asymptotic variance of vec$(\widehat{\boldsymbol{\beta}}_1)$ and asymptotic variance of vec$(\widehat{\boldsymbol{\beta}}_2)$, i.e. $\sqrt{n}\{\text{vec}(\widehat{\boldsymbol{\beta}}_1) - \text{vec}(\boldsymbol{\beta}_1)\} \xrightarrow{d} \text{N}(\mathbf{0}, \mathbf{V}_1)$, where

$$\mathbf{V}_1 = \mathbf{\Sigma}_{\mathbf{Y}|\mathbf{X}} \otimes \mathbf{\Gamma\Omega}^{-1}\mathbf{\Gamma}^T + (\boldsymbol{\eta}^T \otimes \mathbf{\Gamma}_0)\{\boldsymbol{\eta}\mathbf{\Sigma}_{\mathbf{Y}|\mathbf{X}}^{-1}\boldsymbol{\eta}^T \otimes \mathbf{\Omega}_0 + \mathbf{\Omega} \otimes \mathbf{\Omega}_0^{-1} + \mathbf{\Omega}^{-1} \otimes \mathbf{\Omega}_0 - 2\mathbf{I}_d \otimes \mathbf{I}_{p_1-d}\}^{-1}(\boldsymbol{\eta} \otimes \mathbf{\Gamma}_0^T)$$

and $\sqrt{n}\{\text{vec}(\widehat{\boldsymbol{\beta}}_2) - \text{vec}(\boldsymbol{\beta}_2)\} \xrightarrow{d} \text{N}(\mathbf{0}, \mathbf{V}_2)$, where

$$\begin{aligned}\mathbf{V}_2 =\ & \mathbf{\Sigma}_{\mathbf{Y}|\mathbf{X}} \otimes \mathbf{\Sigma}_2^{-1} + \mathbf{\Sigma}_{\mathbf{Y}|\mathbf{X}} \otimes \boldsymbol{\gamma}^T\mathbf{\Gamma\Omega}^{-1}\mathbf{\Gamma}^T\boldsymbol{\gamma} \\ & +(\boldsymbol{\eta}^T \otimes \boldsymbol{\gamma}^T\mathbf{\Gamma}_0)\{\boldsymbol{\eta}\mathbf{\Sigma}_{\mathbf{Y}|\mathbf{X}}^{-1}\boldsymbol{\eta}^T \otimes \mathbf{\Omega}_0 + \mathbf{\Omega} \otimes \mathbf{\Omega}_0^{-1} + \mathbf{\Omega}^{-1} \otimes \mathbf{\Omega}_0 - 2\mathbf{I}_d \otimes \mathbf{I}_{p_1-d}\}^{-1}(\boldsymbol{\eta} \otimes \mathbf{\Gamma}_0^T\boldsymbol{\gamma}).\end{aligned}$$

Proof of Lemma 1

To prove $\text{cov}(\mathbf{Y}|\mathbf{X}) = \text{cov}(\mathbf{r}_{\mathbf{Y}|2}|\mathbf{r}_{1|2})$, we define $\mathbf{\Sigma}_{\mathbf{Y}} = \text{cov}(\mathbf{Y})$, $\mathbf{\Sigma}_{\mathbf{X}} = \text{cov}(\mathbf{X})$, $\mathbf{\Sigma}_{\mathbf{YX}} = \text{cov}(\mathbf{Y}, \mathbf{X})$, $\mathbf{\Sigma}_{\mathbf{Y}1} = \text{cov}(\mathbf{Y}, \mathbf{X}_1)$, $\mathbf{\Sigma}_{\mathbf{Y}2} = \text{cov}(\mathbf{Y}, \mathbf{X}_2)$, and $\mathbf{\Sigma}_{(\mathbf{Y},1)|2} = \text{cov}(\mathbf{Y}, \mathbf{X}_1|\mathbf{X}_2)$. We notice that $\text{cov}(\mathbf{r}_{\mathbf{Y}|2}|\mathbf{r}_{1|2}) = \mathbf{\Sigma}_{\mathbf{Y}|2} - \mathbf{\Sigma}_{(\mathbf{Y},1)|2}\mathbf{\Sigma}_{1|2}^{-1}\mathbf{\Sigma}_{(1,\mathbf{Y})|2}$. Then we have

$$\begin{aligned}
&\text{cov}(\mathbf{Y}|\mathbf{X}) = \mathbf{\Sigma}_{\mathbf{Y}} - \mathbf{\Sigma}_{\mathbf{YX}}\mathbf{\Sigma}_{\mathbf{X}}^{-1}\mathbf{\Sigma}_{\mathbf{XY}} \\
&= \mathbf{\Sigma}_{\mathbf{Y}} - \begin{pmatrix}\mathbf{\Sigma}_{\mathbf{Y}1} & \mathbf{\Sigma}_{\mathbf{Y}2}\end{pmatrix} \begin{pmatrix}\mathbf{\Sigma}_1 & \mathbf{\Sigma}_{12} \\ \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_2\end{pmatrix}^{-1} \begin{pmatrix}\mathbf{\Sigma}_{1\mathbf{Y}} \\ \mathbf{\Sigma}_{2\mathbf{Y}}\end{pmatrix} \\
&= \mathbf{\Sigma}_{\mathbf{Y}} - \begin{pmatrix}\mathbf{\Sigma}_{\mathbf{Y}1} & \mathbf{\Sigma}_{\mathbf{Y}2}\end{pmatrix} \begin{pmatrix}\mathbf{\Sigma}_{1|2}^{-1} & -\mathbf{\Sigma}_{1|2}^{-1}\mathbf{\Sigma}_{12}\mathbf{\Sigma}_2^{-1} \\ -\mathbf{\Sigma}_2^{-1}\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{1|2}^{-1} & \mathbf{\Sigma}_2^{-1} + \mathbf{\Sigma}_2^{-1}\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{1|2}^{-1}\mathbf{\Sigma}_{12}\mathbf{\Sigma}_2^{-1}\end{pmatrix} \begin{pmatrix}\mathbf{\Sigma}_{1\mathbf{Y}} \\ \mathbf{\Sigma}_{2\mathbf{Y}}\end{pmatrix} \\
&= \mathbf{\Sigma}_{\mathbf{Y}} - \mathbf{\Sigma}_{\mathbf{Y}2}\mathbf{\Sigma}_2^{-1}\mathbf{\Sigma}_{2\mathbf{Y}} - \begin{pmatrix}\mathbf{\Sigma}_{\mathbf{Y}1} & \mathbf{\Sigma}_{\mathbf{Y}2}\end{pmatrix} \begin{pmatrix}\mathbf{\Sigma}_{1|2}^{-1} & -\mathbf{\Sigma}_{1|2}^{-1}\mathbf{\Sigma}_{12}\mathbf{\Sigma}_2^{-1} \\ -\mathbf{\Sigma}_2^{-1}\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{1|2}^{-1} & \mathbf{\Sigma}_2^{-1}\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{1|2}^{-1}\mathbf{\Sigma}_{12}\mathbf{\Sigma}_2^{-1}\end{pmatrix} \begin{pmatrix}\mathbf{\Sigma}_{1\mathbf{Y}} \\ \mathbf{\Sigma}_{2\mathbf{Y}}\end{pmatrix} \\
&= \mathbf{\Sigma}_{\mathbf{Y}|2} - \begin{pmatrix}\mathbf{\Sigma}_{\mathbf{Y}1} & \mathbf{\Sigma}_{\mathbf{Y}2}\end{pmatrix} \begin{pmatrix}\mathbf{I}_{p_1} \\ -\mathbf{\Sigma}_2^{-1}\mathbf{\Sigma}_{21}\end{pmatrix} \mathbf{\Sigma}_{1|2}^{-1} \begin{pmatrix}\mathbf{I}_{p_1} & -\mathbf{\Sigma}_{12}\mathbf{\Sigma}_2^{-1}\end{pmatrix} \begin{pmatrix}\mathbf{\Sigma}_{1\mathbf{Y}} \\ \mathbf{\Sigma}_{2\mathbf{Y}}\end{pmatrix} \\
&= \mathbf{\Sigma}_{\mathbf{Y}|2} - (\mathbf{\Sigma}_{\mathbf{Y}1} - \mathbf{\Sigma}_{\mathbf{Y}2}\mathbf{\Sigma}_2^{-1}\mathbf{\Sigma}_{21})\mathbf{\Sigma}_{1|2}^{-1}(\mathbf{\Sigma}_{1\mathbf{Y}} - \mathbf{\Sigma}_{12}\mathbf{\Sigma}_2^{-1}\mathbf{\Sigma}_{2\mathbf{Y}}) \\
&= \mathbf{\Sigma}_{\mathbf{Y}|2} - \mathbf{\Sigma}_{(\mathbf{Y},1)|2}\mathbf{\Sigma}_{1|2}^{-1}\mathbf{\Sigma}_{(1,\mathbf{Y})|2} \\
&= \text{cov}(\mathbf{r}_{\mathbf{Y}|2}|\mathbf{r}_{1|2}).
\end{aligned}$$

Proof of Proposition 3

We first investigate the objective function l in (1). Since all the other parameters depend on $\mathbf{\Gamma}$, fixing a subspace span$(\mathbf{\Gamma}) \equiv \mathcal{S}$, the maximized value of l in (1) is in (11). Thus we write l as a function of $\mathcal{S}$, i.e. $l(\mathcal{S})$. Define $l_0(\mathcal{S})$ as the population version of $l(\mathcal{S})$, where all the $\mathbf{S}$ matrices in l are replaced by its population counterparts. Let $d_{\mathcal{S}}$ denote the dimension of $\mathcal{S}$ and $d_{\mathcal{E}_{1|2}}$ denote the dimension of $\mathcal{E}_{\mathbf{\Sigma}_{1|2}}(\boldsymbol{\beta}_1)$. We consider the following two scenarios.

Scenario 1: $\mathcal{E}_{\mathbf{\Sigma}_{1|2}}(\boldsymbol{\beta}_1) \subseteq \mathcal{S}$, and $d_{\mathcal{E}_{1|2}} < d_{\mathcal{S}}$. Then

$$l_0(\mathcal{S}) = l_0(\mathcal{E}_{\mathbf{\Sigma}_{1|2}}(\boldsymbol{\beta}_1)).$$

By the asymptotic distribution of Wilks statistic, see Li and Babu [2], Chapter 11, we have

$$l(\mathcal{S}) - l(\mathcal{E}_{\mathbf{\Sigma}_{1|2}}(\boldsymbol{\beta}_1)) = O_p(1).$$

Let $BIC(\mathcal{S})$ denote the BIC value when span($\boldsymbol{\Gamma}$) = $\mathcal{S}$. Then

$$\begin{aligned}
&BIC(\mathcal{S}) - BIC(\mathcal{E}_{\boldsymbol{\Sigma}_{1|2}}(\boldsymbol{\beta}_1)) \\
&= -2[l(\mathcal{S}) - l(\mathcal{E}_{\boldsymbol{\Sigma}_{1|2}}(\boldsymbol{\beta}_1))] + \log(n)[N(d_{\mathcal{S}}) - N(d_{\mathcal{E}_{1|2}})] \\
&= O_p(1) + \log(n)[N(d_{\mathcal{S}}) - N(d_{\mathcal{E}_{1|2}})].
\end{aligned}$$

According to the parameter counts $N(d) = r + p_1 + p_2 + p_1p_2 + p_1(p_1+1)/2 + dr + p_2r + r(r+1)/2$, when $d_{\mathcal{E}_{1|2}} < d_{\mathcal{S}}$, $N(d_{\mathcal{E}_{1|2}}) < N(d_{\mathcal{S}})$. So we have $BIC(\mathcal{S}) > BIC(\mathcal{E}_{\boldsymbol{\Sigma}_{1|2}}(\boldsymbol{\beta}_1))$ with probability tending to 1.

Scenario 2: $\mathcal{E}_{\boldsymbol{\Sigma}_{1|2}}(\boldsymbol{\beta}_1) \nsubseteq \mathcal{S}$. Then there exists some constant $c > 0$ such that

$$\frac{1}{n} l_0(\mathcal{S}) < \frac{1}{n} l_0(\mathcal{E}_{\boldsymbol{\Sigma}_{1|2}}(\boldsymbol{\beta}_1)) - c.$$

We then have

$$\begin{aligned}
&BIC(\mathcal{S}) - BIC(\mathcal{E}_{\boldsymbol{\Sigma}_{1|2}}(\boldsymbol{\beta}_1)) \\
&= 2n\left[\frac{1}{n} l(\mathcal{E}_{\boldsymbol{\Sigma}_{1|2}}(\boldsymbol{\beta}_1)) - \frac{1}{n} l(\mathcal{S})\right] + \log(n)[N(d_{\mathcal{S}}) - N(d_{\mathcal{E}_{1|2}})] \\
&> nc + \log(n)[N(d_{\mathcal{S}}) - N(d_{\mathcal{E}_{1|2}})].
\end{aligned}$$

When $n \to \infty$, the term nc dominates the term $\log(n)[N(d_{\mathcal{S}}) - N(d_{\mathcal{E}_{1|2}})]$. So we have $BIC(\mathcal{S}) > BIC(\mathcal{E}_{\boldsymbol{\Sigma}_{1|2}}(\boldsymbol{\beta}_1))$ with probability tending to 1.

Combining the two scenarios, $\mathcal{E}_{\boldsymbol{\Sigma}_{1|2}}(\boldsymbol{\beta}_1)$ minimizes BIC value with probability tending to 1. So BIC will choose the dimension of $\mathcal{E}_{\boldsymbol{\Sigma}_{1|2}}(\boldsymbol{\beta}_1)$, i.e. d, with probability tending to 1.

C. SENSITIVITY ANALYSIS

We considered a situation where the immaterial part of $\mathbf{X}_1$ has larger variation than the material part. We kept the generation of all the parameters the same as Table 2 in Simulation study and only changed $\boldsymbol{\Omega}$ and $\boldsymbol{\Omega}_0$. In this case, $\mathbf{A}$ was a diagonal matrix with diagonal elements being independent normal $(4, 2^2)$ variates, $\mathbf{B}$ was a diagonal matrix with diagonal elements being independent normal $(2.2, 0.3^2)$ variates, $\boldsymbol{\Omega} = \mathbf{A}\mathbf{A}^T$ and $\boldsymbol{\Omega}_0 = \mathbf{B}\mathbf{B}^T$. Then we have $\|\boldsymbol{\Omega}\| = 2.56$ and $\|\boldsymbol{\Omega}_0\| = 6.38$. The results of $\|\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1\|_F$ are provided in Web Table 1. In this scenario, EPPLS is still more efficient than OLS, especially when we have a larger number of response variables. PPLS is not as stable in this case because its algorithm seeks for the linear combination of $\mathbf{X}_1$ that has the largest covariance with the response, which may not be in the EPPLS subspace. EPLS also performs well, especially with the multivariate response. It loses some efficiency compared to EPPLS, but it still achieves efficiency gains by eliminating some immaterial information from the data. PLS, PCR, PRINCALS, and CA have a large bias and have worse performance compared to OLS. In both scenarios in Table 2 and Web Table 1, EPPLS is the most efficient, and its performance is quite stable. When the variation of the material part is large, PPLS is a fast and easy-to-compute method that also performs well.

We also considered another situation where the linearity assumption between $\mathbf{X}_1$ and $\mathbf{X}_2$ is relaxed. The simulation settings that produced Table 2 of the manuscript were considered, i.e., there are three categorical predictor variables $\mathbf{X}_2 = 10(W_{21}, W_{22}, W_{23})$, where W_{21}, W_{22} and W_{23} were independent Bernoulli variates that take value 1 with probability 0.4, 0.5 and 0.8, respectively. However, we assumed that the relationship between $\mathbf{X}_1$ and $\mathbf{X}_2$ is not linear. Specifically, we generated $\mathbf{X}_1$ such that $\mathbf{X}_1 = \boldsymbol{\mu}_1 + \boldsymbol{\gamma}^T(\mathbf{W} - \mathrm{E}(\mathbf{W})) + \mathbf{e}$, where $\mathbf{W} = (W_{21}W_{22}, W_{21}W_{23}, W_{22}W_{23})$. We took $\boldsymbol{\gamma} = (0.5\mathbf{1}_{p_1}, -0.5\mathbf{1}_{p_1}, 1.2\mathbf{1}_{p_1},)^T$ and $\boldsymbol{\beta}_2 = (1.5\mathbf{1}_r, 1.2\mathbf{1}_r, 2\mathbf{1}_r)^T$. The results are provided in Web Table 2. The performance of EPPLS is quite stable and still performs the best among all the methods. However, the performance of PPLS deteriorates a lot, that PPLS even underperforms OLS most of the time.

D. ADDITIONAL DETAILS FOR DATA APPLICATION

In Web Figures 1 and 2, we added the clustering structure of the responses ($\mathbf{Y}$) and predictor variables ($\mathbf{X}_1$ and $\mathbf{X}_2$) to the heatmaps in Figures 2 and 3 of the manuscript. In addition, Web Figures 3 and 4 show the heatmaps using the uniform color scale for all methods.

Web Table 1. Results of average (standard deviation / (number of replications)$^{1/2}$) of $\|\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1\|_F$ based on 100 replications.

r	Methods	$n = 100$		$n = 300$		$n = 1000$	
1	EPPLS	0.81	(0.032)	0.61	(0.023)	0.35	(0.013)
	EPLS	0.88	(0.022)	0.85	(0.018)	0.54	(0.010)
	PPLS	2.57	(0.047)	2.43	(0.046)	2.38	(0.039)
	PLS	0.88	(0.002)	0.94	(0.001)	3.04	(0.001)
	PCR	0.87	$(3.1 * 10^{-4})$	0.96	$(3.0 * 10^{-4})$	3.08	$(2.8 * 10^{-4})$
	PRINCALS	0.92	$(2.7 * 10^{-}3)$	1.04	$(1.1 * 10^{-3})$	3.60	$(9.7 * 10^{-4})$
	CA	0.87	$(4.2 * 10^{-4})$	0.94	$(4.6 * 10^{-4})$	3.64	$(3.3 * 10^{-4})$
	OLS	0.85	(0.031)	0.71	(0.023)	0.36	(0.013)
10	EPPLS	2.27	(0.042)	1.27	(0.024)	0.70	(0.014)
	EPLS	2.81	(0.079)	1.97	(0.049)	1.21	(0.029)
	PPLS	15.20	(0.855)	13.23	(0.321)	9.34	(0.291)
	PLS	5.71	(0.224)	6.15	(0.179)	10.39	(0.148)
	PCR	8.90	(0.003)	9.65	(0.001)	12.72	(0.001)
	PRINCALS	9.62	(0.005)	10.30	(0.006)	12.47	(0.003)
	CA	8.73	(0.002)	9.52	(0.001)	12.54	(0.001)
	OLS	4.10	(0.046)	2.29	(0.022)	1.24	(0.013)
30	EPPLS	3.67	(0.045)	2.06	(0.027)	1.38	(0.014)
	EPLS	4.96	(0.207)	2.84	(0.047)	1.79	(0.035)
	PPLS	25.78	(1.872)	24.28	(0.985)	16.14	(0.610)
	PLS	12.62	(0.578)	12.22	(0.345)	12.21	(0.316)
	PCR	20.10	(0.006)	19.21	(0.003)	16.10	(0.001)
	PRINCALS	21.68	(0.016)	20.51	(0.010)	15.79	(0.004)
	CA	19.75	(0.005)	18.91	(0.003)	15.88	(0.001)
	OLS	7.08	(0.054)	3.99	(0.024)	2.16	(0.011)

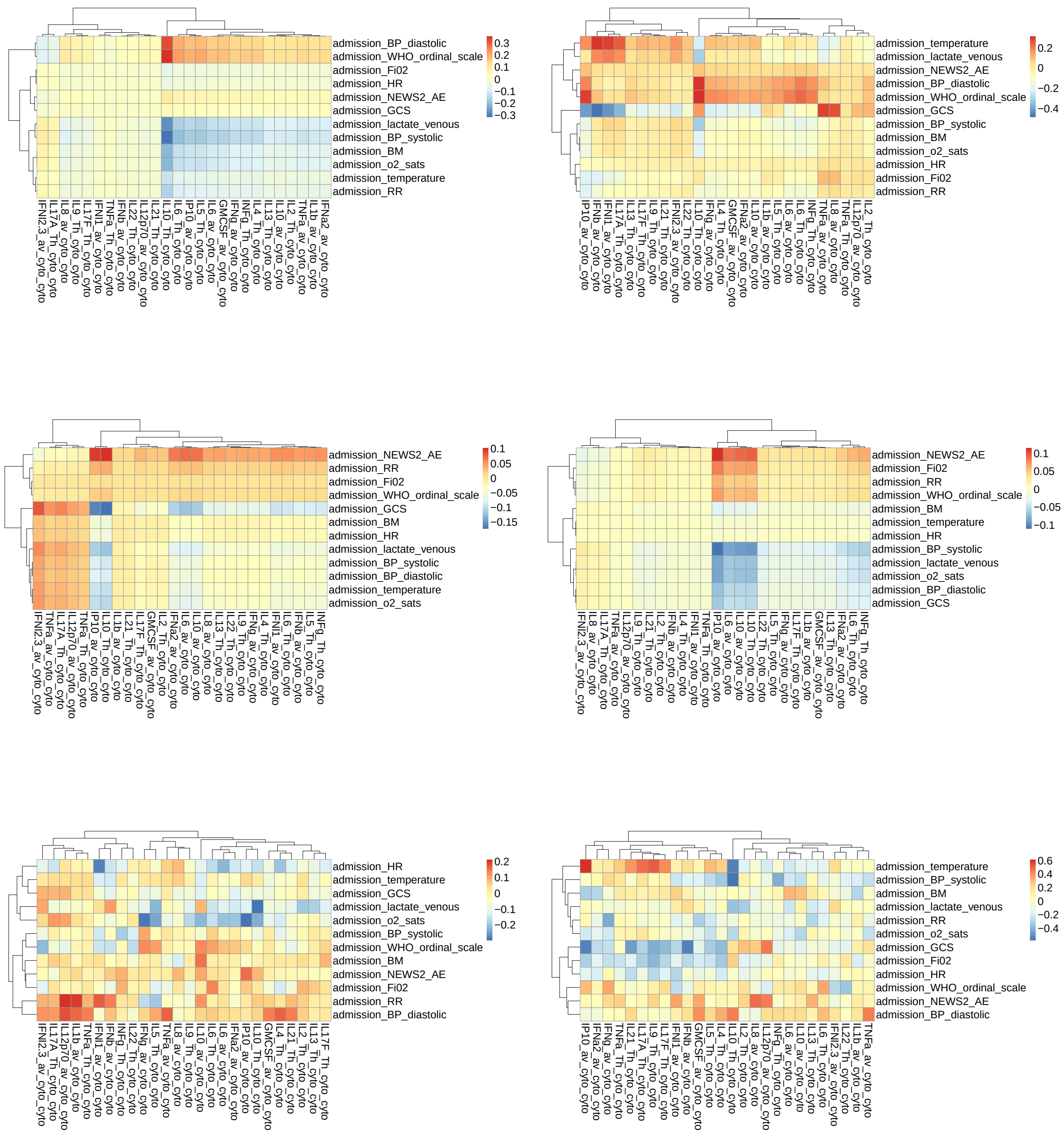
References

1. Cook, R. D., Helland, I. S. and Su, Z. (2013). Envelopes and partial least squares regression. *Journal of the Royal Statistical Society: Series B* **75** 851–877.
2. Li, B. and Babu, G. J. (2019). *A Graduate Course on Statistical Inference*. Springer.
3. Shapiro, A. (1986). Asymptotic theory of overparameterized structural models. *Journal of the American Statistical Association* **81** 142–149.

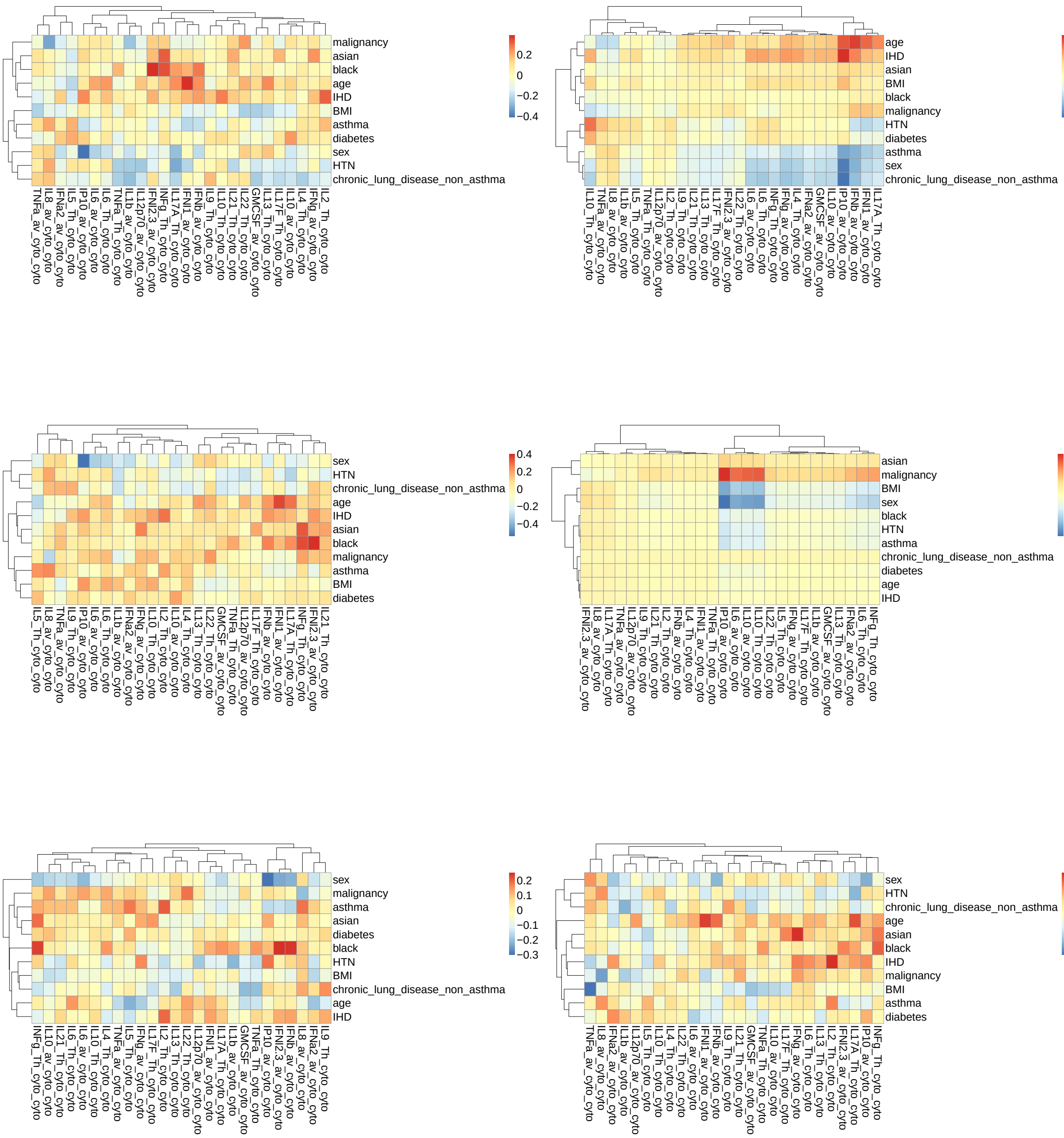
□

Web Table 2. Results of average (standard deviation / (number of replications)$^{1/2}$) of $\|\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1\|_F$ based on 100 replications.

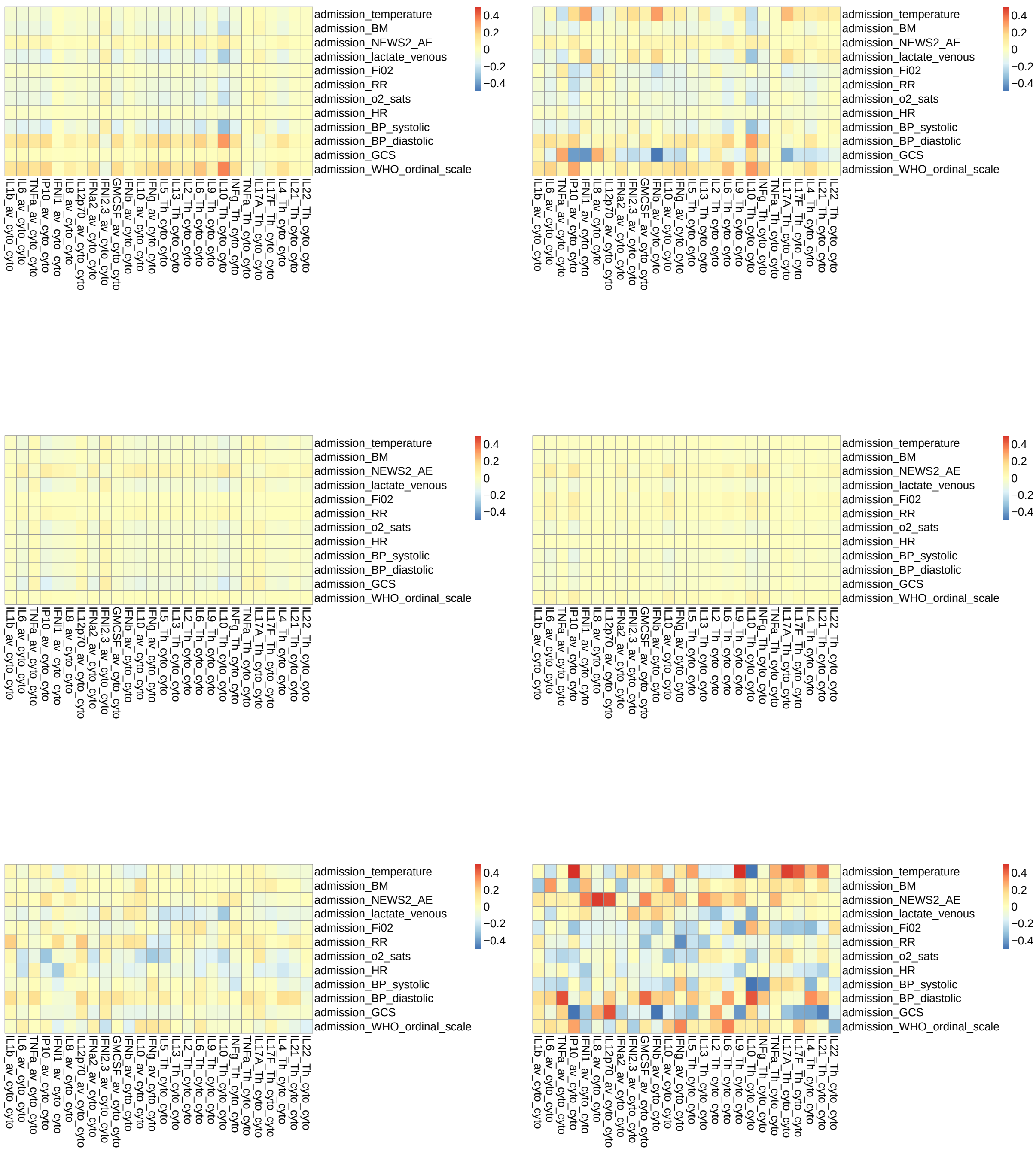
r	Methods	$n = 100$		$n = 300$		$n = 1000$	
1	EPPLS	0.68	(0.078)	0.39	(0.035)	0.12	(0.023)
	EPLS	1.43	(0.151)	1.25	(0.090)	0.22	(0.023)
	PPLS	3.87	(0.181)	7.34	(0.191)	0.60	(0.008)
	PLS	10.47	(0.008)	15.43	(0.006)	1.13	(0.007)
	PCR	8.67	(0.269)	4.72	(0.585)	0.38	(0.006)
	PRINCALS	3.77	(0.067)	4.89	(0.036)	0.44	(0.010)
	CA	10.32	(0.070)	16.48	(0.157)	1.08	(0.004)
	OLS	2.56	(0.141)	1.51	(0.078)	0.79	(0.042)
10	EPPLS	0.99	(0.026)	0.61	(0.018)	0.26	(0.006)
	EPLS	2.02	(0.201)	2.17	(0.196)	1.25	(0.114)
	PPLS	14.75	(0.702)	24.35	(0.631)	11.97	(0.122)
	PLS	40.12	(0.029)	51.11	(0.019)	20.72	(0.005)
	PCR	33.18	(1.018)	15.65	(1.932)	2.19	(0.189)
	PRINCALS	14.37	(0.194)	16.03	(0.114)	8.13	(0.033)
	CA	39.53	(0.266)	54.59	(0.517)	22.25	(0.054)
	OLS	9.59	(0.160)	5.36	(0.096)	2.98	(0.060)
30	EPPLS	1.43	(0.033)	0.86	(0.021)	0.43	(0.010)
	EPLS	1.93	(0.044)	1.27	(0.028)	1.42	(0.113)
	PPLS	22.04	(1.046)	35.09	(0.910)	31.26	(0.317)
	PLS	59.85	(0.043)	73.69	(0.027)	54.06	(0.014)
	PCR	49.52	(1.516)	22.68	(2.778)	5.15	(0.498)
	PRINCALS	21.47	(0.301)	23.15	(0.162)	21.11	(0.092)
	CA	58.97	(0.397)	78.71	(0.747)	58.01	(0.139)
	OLS	16.86	(0.189)	9.47	(0.109)	5.52	(0.061)



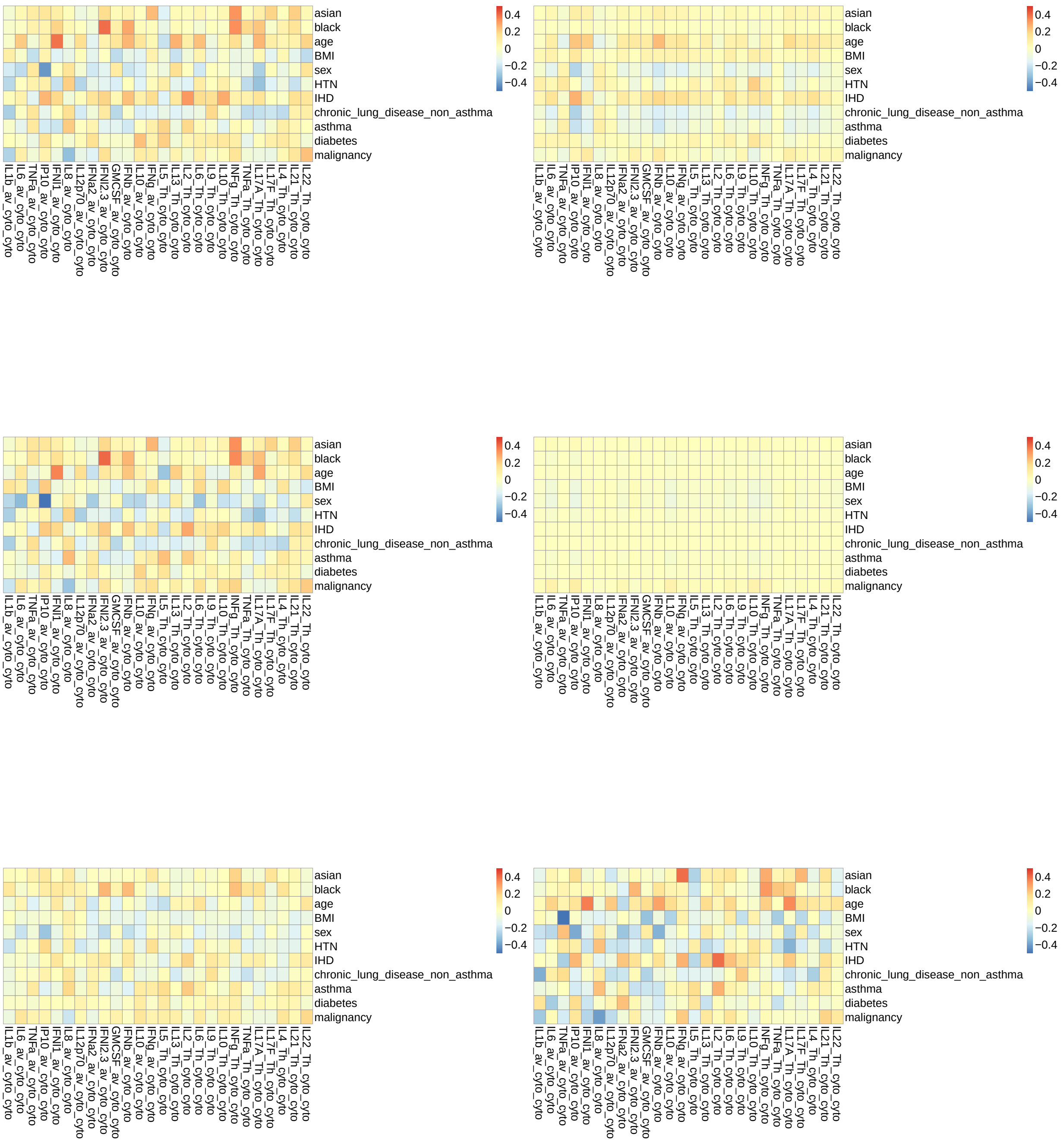
Web Figure 1. Heatmaps of the regression coefficients of $\widehat{\boldsymbol{\beta}}_1$ under EPPLS (left of 1st row), EPLS (right of 1st row), PPLS (left of 2nd row), PLS (right of 2nd row), PCR (left of 3rd row), and OLS (right of 3rd row).



Web Figure 2. Heatmaps of the regression coefficients of $\widehat{\beta}_2$ under EPPLS (left of 1st row), EPLS (right of 1st row), PPLS (left of 2nd row), PLS (right of 2nd row), PCR (left of 3rd row), and OLS (right of 3rd row).



Web Figure 3. Heatmaps of the regression coefficients of $\widehat{\boldsymbol{\beta}}_1$ under EPPLS (left of 1st row), EPLS (right of 1st row), PPLS (left of 2nd row), PLS (right of 2nd row), PCR (left of 3rd row), and OLS (right of 3rd row).



Web Figure 4. Heatmaps of the regression coefficients of $\widehat{\boldsymbol{\beta}}_2$ under EPPLS (left of 1st row), EPLS (right of 1st row), PPLS (left of 2nd row), PLS (right of 2nd row), PCR (left of 3rd row), and OLS (right of 3rd row).