Compositional Mediation Model for Binary Outcomes: Application to Microbiome Samples

Supplementary Materials

By Michael B. Sohn, Jiarui Lu, and Hongzhe Li

A Estimation of Composition Parameters

To estimate the parameters in Model (1), we propose the following objective function, which minimizes the composition norm of the difference between observed and estimated compositions,

$$\begin{aligned}\hat{\boldsymbol{a}} &= \underset{\boldsymbol{a},\boldsymbol{h}_r,\boldsymbol{m}_0\in\mathbb{S}^{k-1}}{\text{argmin}} \sum_{i=1}^{n} \left\| M_i \ominus (\boldsymbol{m}_0 \oplus \boldsymbol{a}^{T_i} \oplus \boldsymbol{h}_1^{X_{i1}} \oplus \cdots \oplus \boldsymbol{h}_q^{X_{iq}}) \right\|^2 \qquad \text{(S1)}\\
&= \underset{\boldsymbol{a},\boldsymbol{h}_r,\boldsymbol{m}_0\in\mathbb{S}^{k-1}}{\text{argmin}} \sum_{i=1}^{n}\sum_{j=1}^{k-1} \left\{ (k-1)\left[\log\left(\frac{M_{ij} m_{0k} a_k^{T_i} \prod_{r=1}^{q} h_{rk}^{X_{ir}}}{M_{ik} m_{0j} a_j^{T_i} \prod_{r=1}^{q} h_{rj}^{X_{ir}}}\right)\right]^2 \right.\\
&\qquad \left. - \log\left(\frac{M_{ij} m_{0k} a_k^{T_i} \prod_{r=1}^{q} h_{rk}^{X_{ir}}}{M_{ik} m_{0j} a_j^{T_i} \prod_{r=1}^{q} h_{rj}^{X_{ir}}}\right) \sum_{\ell\neq j}^{k-1} \log\left(\frac{M_{i\ell} m_{0k} a_k^{T_i} \prod_{r=1}^{q} h_{rk}^{X_{ir}}}{M_{ik} m_{0\ell} a_\ell^{T_i} \prod_{r=1}^{q} h_{r\ell}^{X_{ir}}}\right) \right\}.\end{aligned}$$

The objective function (S1) is convex in terms of $\text{alt}(\boldsymbol{a})_j$, $\text{alt}(\boldsymbol{m}_0)_j$, and $\text{alt}(\boldsymbol{h}_r)_j$ for $j = 1, \ldots, k-1$; $r = 1, \ldots, q$. Thus, the optimal solution can be obtained by solving the following system of linear equations with constraints $\boldsymbol{m}_0, \boldsymbol{a}, \boldsymbol{h}_r \in \mathbb{S}^{k-1}$:

$$\begin{bmatrix} D(1) & D(T) & D(X_1) & \cdots & D(X_q) \\ D(T) & D(T^2) & D(TX_1) & \cdots & D(TX_q) \\ D(X_1) & D(TX_1) & D(X_1^2) & \cdots & D(X_1X_q) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ D(X_q) & D(TX_q) & D(X_1X_q) & \cdots & D(X_q) \end{bmatrix} \begin{bmatrix} \text{alt}(\boldsymbol{m}_0) \\ \text{alt}(\boldsymbol{a}) \\ \text{alt}(\boldsymbol{h}_1) \\ \vdots \\ \text{alt}(\boldsymbol{h}_q) \end{bmatrix} = \begin{bmatrix} \boldsymbol{\zeta}_0 \\ \boldsymbol{\zeta}_1 \\ \boldsymbol{\xi}_1 \\ \vdots \\ \boldsymbol{\xi}_q \end{bmatrix},$$

where $\zeta_{0j} = k\sum_{i=1}^{n} \log M_{ij} - \sum_{\ell=1}^{k}\sum_{i=1}^{n} \log M_{i\ell}$, $\zeta_{1j} = k\sum_{i=1}^{n} T_i \log M_{ij} - \sum_{\ell=1}^{k}\sum_{i=1}^{n} T_i \log M_{i\ell}$, $\xi_{rj} = k\sum_{i=1}^{n} X_{ir} \log M_{ij} - \sum_{\ell=1}^{k}\sum_{i=1}^{n} X_{ir} \log M_{i\ell}$, and for any ν, $D(\nu)$ is defined as

$$D(\nu) = \begin{bmatrix} (k-1)\sum_{i=1}^{n}\nu_i & -\sum_{i=1}^{n}\nu_i & \ldots & -\sum_{i=1}^{n}\nu_i \\ -\sum_{i=1}^{n}\nu_i & (k-1)\sum_{i=1}^{n}\nu_i & \ldots & -\sum_{i=1}^{n}\nu_i \\ \vdots & \vdots & \ddots & \vdots \\ -\sum_{i=1}^{n}\nu_i & -\sum_{i=1}^{n}\nu_i & \ldots & (k-1)\sum_{i=1}^{n}\nu_i \end{bmatrix}.$$

B Estimation of Regression Parameters

Let $\eta_i = 2y_i - 1$, $\boldsymbol{z}_i = (1, t_i, \log(\boldsymbol{m}_i)^\top, \boldsymbol{x}_i^\top)^\top$, $\boldsymbol{\beta} = (c_0, c, \boldsymbol{b}^\top, \boldsymbol{g}^\top)^\top$, and $q(\eta_i \boldsymbol{z}_i^\top \boldsymbol{\beta}) = -\log \Phi(\eta_i \boldsymbol{z}_i^\top \boldsymbol{\beta})$. Then, an L_1-penalized log-likelihood function for Model (2) is given by

$$\widehat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \left\{ \frac{1}{n} \sum_{i=1}^{n} q(\eta_i \boldsymbol{z}_i^\top \boldsymbol{\beta}) \right\}, \quad \text{subject to } \|\boldsymbol{\beta}\|_1 \leq t \text{ and } \mathbf{1}_k^\top \boldsymbol{b} = 0, \tag{S2}$$

where $t \geq 0$ is some constant. By the Taylor expansion, we have

$$\begin{aligned} q(\eta_i \boldsymbol{z}_i^\top \boldsymbol{\beta}) &= q(\eta_i \boldsymbol{z}_i^\top \boldsymbol{\beta}_0) + (\boldsymbol{\beta} - \boldsymbol{\beta}_0)^\top G(\eta_i \boldsymbol{z}_i^\top \boldsymbol{\beta}_0) + \frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)^\top H(\eta_i \boldsymbol{z}_i^\top \boldsymbol{\beta}^*)(\boldsymbol{\beta} - \boldsymbol{\beta}_0) \\ &= \frac{1}{2} \left\{ \boldsymbol{\beta}^\top H(\eta_i \boldsymbol{z}_i^\top \boldsymbol{\beta}^*) \boldsymbol{\beta} + 2 \left[G(\eta_i \boldsymbol{z}_i^\top \boldsymbol{\beta}_0) - H(\eta_i \boldsymbol{z}_i^\top \boldsymbol{\beta}^*) \boldsymbol{\beta}_0 \right]^\top \boldsymbol{\beta} + C \right\}, \end{aligned}$$

where $G(\eta_i \boldsymbol{z}_i^\top \boldsymbol{\beta}_0) = \nabla_\beta q(\eta_i \boldsymbol{z}_i^\top \boldsymbol{\beta}_0)$, $H(\eta_i \boldsymbol{z}_i^\top \boldsymbol{\beta}^*) = \nabla_\beta^2 q(\eta_i \boldsymbol{z}_i^\top \boldsymbol{\beta}^*)$, $\boldsymbol{\beta}^*$ a vector that lies between $\boldsymbol{\beta}_0$ and $\boldsymbol{\beta}$, and C is a constant with respect to $\boldsymbol{\beta}$. Since $\sum_{i=1}^n H(\eta_i \boldsymbol{z}_i^\top \boldsymbol{\beta}^*) \succeq 0$, finding a solution minimizing $\sum_{i=1}^n q(\eta_i \boldsymbol{z}_i^\top \boldsymbol{\beta})$ is equivalent to finding a solution of $\nabla_\beta q(\eta_i \boldsymbol{z}_i^\top \boldsymbol{\beta}) = 0$, that is,

$$\widehat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \left\{ \frac{1}{n} \sum_{i=1}^{n} q(\eta_i \boldsymbol{z}_i^\top \boldsymbol{\beta}) \right\} \quad \Leftrightarrow \quad \widehat{\boldsymbol{\beta}} = \boldsymbol{\beta}_0 - \left(\sum_{i=1}^{n} H(\eta_i \boldsymbol{z}_i^\top \boldsymbol{\beta}^*) \right)^{-} \left(\sum_{i=1}^{n} G(\eta_i \boldsymbol{z}_i^\top \boldsymbol{\beta}_0) \right),$$

where A^- is the Moore-Penrose inverse of a matrix A. Note that $\nabla_\beta q(\eta_i \boldsymbol{z}_i^\top \boldsymbol{\beta}) = -\xi_i(\eta_i \boldsymbol{z}_i^\top \boldsymbol{\beta}) \boldsymbol{z}_i$, where $\xi_i(\eta_i \boldsymbol{z}_i^\top \boldsymbol{\beta}) = \eta_i \phi(\eta_i \boldsymbol{z}_i^\top \boldsymbol{\beta}) / \Phi(\eta_i \boldsymbol{z}_i^\top \boldsymbol{\beta})$ and $\nabla_\beta^2 q(\eta_i \boldsymbol{z}_i^\top \boldsymbol{\beta}) = \xi_i(\eta_i \boldsymbol{z}_i^\top \boldsymbol{\beta})[\boldsymbol{z}_i^\top \boldsymbol{\beta} + \xi_i(\eta_i \boldsymbol{z}_i^\top \boldsymbol{\beta})] \boldsymbol{z}_i \boldsymbol{z}_i^\top$. Substituting these terms, we have

$$\widehat{\boldsymbol{\beta}} = \boldsymbol{\beta}_0 + (Z^\top \Xi Z)^- Z^\top \boldsymbol{\xi}(\boldsymbol{\beta}_0) = (Z^\top \Xi Z)^- Z^\top \Xi \boldsymbol{u},$$

where Ξ is an $n \times n$ diagonal matrix with the i^{th} diagonal term $\Xi_{ii} = \xi_i(\eta_i \boldsymbol{z}_i^\top \boldsymbol{\beta}^*)[\boldsymbol{z}_i^\top \boldsymbol{\beta}^* + \xi_i(\eta_i \boldsymbol{z}_i^\top \boldsymbol{\beta}^*)]$, $\boldsymbol{\xi}(\boldsymbol{\beta}_0) = (\xi_1(\eta_1 \boldsymbol{z}_1^\top \boldsymbol{\beta}_0), \ldots, \xi_1(\eta_n \boldsymbol{z}_n^\top \boldsymbol{\beta}_0))^\top$, and $\boldsymbol{u} = Z\boldsymbol{\beta}_0 + \Xi^- \boldsymbol{\xi}(\boldsymbol{\beta}_0)$. This is, given $\boldsymbol{\beta}^*$ and $\boldsymbol{\beta}_0$, the solution of a weighted least squares problem with a weight matrix Ξ, a dependent variable $\boldsymbol{u}$, and independent variables Z, that is,

$$\widehat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \|\Xi^{1/2}(\boldsymbol{u} - Z\boldsymbol{\beta})\|_2^2.$$

Therefore, optimization problem (S2) can be expressed as

$$\widehat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \left\{ \frac{1}{2n} \|\Xi^{1/2}(\boldsymbol{u} - \widetilde{Z}\boldsymbol{\beta})\|_2^2 + \lambda \|\boldsymbol{\beta}\|_1 \right\}, \quad \text{s.t. } \boldsymbol{\iota}^\top \boldsymbol{\beta} = 0, \tag{S3}$$

where $\widetilde{Z} = Z(\mathcal{I}_p - \boldsymbol{\iota}\boldsymbol{\iota}^\top / k)$ and $\boldsymbol{\iota}^\top = (0, 0, 1, \ldots, 1, 0, \ldots, 0)$. Note that $Z\boldsymbol{\beta} = \widetilde{Z}\boldsymbol{\beta}$ because $\boldsymbol{\iota}^\top \boldsymbol{\beta} = 0$. The objective function in this alternative optimization problem, particularly Ξ and $\boldsymbol{u}$, depend on unknown quantities, $\boldsymbol{\beta}^*$ and $\boldsymbol{\beta}_0$. Therefore, we propose a method that combines iteratively reweighted least squares and coordinate descent method of multipliers (IRLS-CDMM). To derive an algorithm for this constrained optimization problem, we first form the augmented Lagrangian,

$$L_\mu(\boldsymbol{\beta}, \varsigma) = \frac{1}{2n} \|\Xi^{1/2}(\boldsymbol{u} - \widetilde{Z}\boldsymbol{\beta})\|_2^2 + \lambda \|\boldsymbol{\beta}\|_1 + \varsigma \boldsymbol{\iota}^\top \boldsymbol{\beta} + \frac{\mu}{2}(\boldsymbol{\iota}^\top \boldsymbol{\beta})^2,$$

where ς is the Lagrange multiplier and $\mu > 0$ is a penalty parameter. Defining a scaled Lagrange multiplier $\alpha = \varsigma / \mu$, we obtain the solution of optimization problem (S3) given Ξ and $\boldsymbol{u}$ by iterating

$$\boldsymbol{\beta}^{(\ell+1)} \leftarrow \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \left\{ \frac{1}{2n} \|\Xi^{1/2}(\boldsymbol{u} - \widetilde{Z}\boldsymbol{\beta})\|_2^2 + \lambda \|\boldsymbol{\beta}\|_1 + \frac{\mu}{2} \left(\boldsymbol{\iota}^\top \boldsymbol{\beta} + \alpha^{(\ell)} \right)^2 \right\}; \tag{S4}$$

$$\alpha^{(\ell+1)} \leftarrow \alpha^{(\ell)} + \boldsymbol{\iota}^\top \boldsymbol{\beta}^{(\ell+1)}. \tag{S5}$$

Since the L_1 terms are now separable, optimization problem (S4), can be solved by the coordinate decent method,

$$\beta_j^{(\ell+1)} \leftarrow \frac{1}{\widetilde{w}_j} S_\lambda \left\{ \frac{1}{n} \widetilde{\boldsymbol{z}}_j^\top \Xi^{(\ell)} \left(\boldsymbol{u}^{(\ell)} - \sum_{i \neq j} \beta_i^{(\ell+1)} \widetilde{\boldsymbol{z}}_i \right) - \mu \left(\sum_{i \neq j} \beta_i^{(\ell+1)} \frac{\iota_i \iota_j}{k} + \alpha^{(\ell)} \frac{\iota_j}{\sqrt{k}} \right) \right\}, \tag{S6}$$

where $\widetilde{\boldsymbol{z}}_k$ is the k^{th} column vector of $\widetilde{Z}$, $\widetilde{w}_j = \|\widetilde{\boldsymbol{z}}_j\|_2^2 / n + \mu/k$, and $S_\lambda(t) = \text{sgn}(t)(|t| - \lambda)_+$. We repeat Iterations (S4)-(S5) with the updated $\Xi^{(\ell)}$ and $\boldsymbol{u}^{(\ell)}$, as in Algorithm 1.

Algorithm 1 IRLS-CDMM

Initialize $\boldsymbol{\beta}^{(0)}$, $\alpha^{(0)}$, $\Xi^{(0)}$, and $\boldsymbol{u}^{(0)}$
repeat
  for $j = 1$ to p **do**
    Update $\beta_j^{(\ell+1)}$ using (S6)
    Update $\alpha^{(\ell+1)}$ using (S5)
  end for
  Find $\boldsymbol{\beta}^{*(\ell+1)}$ by a line search that maximizes $\sum_{i=1}^n \log \Phi(\eta_i \boldsymbol{z}_i^\top \boldsymbol{\beta})$
  Update $\Xi^{(\ell+1)}$ and $\boldsymbol{u}^{(\ell+1)}$
  $\ell \leftarrow \ell + 1$
until convergence

C De-biasing Procedure

Let $\widehat{\Sigma} = \widetilde{Z}^\top \widehat{\Xi} \widetilde{Z}/n$, where $\widehat{\Xi}$ is an estimate of Ξ obtained from Algorithm 1, $\boldsymbol{e}_j \in \mathbb{R}^p$ be the vector with one at the j^{th} position and zero everywhere else, and γ be some constant. The matrix $\widetilde{\Theta}$ in Equation (6) can be obtained from Algorithm 2. To describe the logic behind Algorithm 2, define

Algorithm 2 Constructing a de-biased estimator

for $j = 1$ to p **do**
  $\widehat{\boldsymbol{\theta}}_j \leftarrow \min\limits_{\theta} \boldsymbol{\theta}^\top \widehat{\Sigma} \boldsymbol{\theta}$ subject to $\|\widehat{\Sigma}\boldsymbol{\theta} - (\mathcal{I}_p - \boldsymbol{\iota}\boldsymbol{\iota}^\top/k)\boldsymbol{e}_j\|_\infty \leq \gamma$
end for
$\widehat{\Theta} \leftarrow \left(\widehat{\boldsymbol{\theta}}_i, \ldots, \widehat{\boldsymbol{\theta}}_p\right)^\top$; $\widetilde{\Theta} \leftarrow (\mathcal{I}_p - \boldsymbol{\iota}\boldsymbol{\iota}^\top/k)\widehat{\Theta}$
$\widehat{\boldsymbol{\beta}}_{db} \leftarrow \widehat{\boldsymbol{\beta}} + \frac{1}{n} \widetilde{\Theta} \widetilde{Z}^\top \widehat{\Xi} (\widehat{\boldsymbol{u}} - \widetilde{Z}\widehat{\boldsymbol{\beta}})$

$\Sigma = \widetilde{Z}^\top \Xi \widetilde{Z}/n$ and suppose that $V\Lambda V^\top$ is the eigenvalue decomposition of Σ. Since $\left(V, \boldsymbol{\iota}/\sqrt{k}\right)$ is full rank and orthonormal, Σ can be expressed as

$$\Sigma = \left(V, \boldsymbol{\iota}/\sqrt{k}\right) \begin{pmatrix} \Lambda & 0 \\ 0 & 0 \end{pmatrix} \left(V, \boldsymbol{\iota}/\sqrt{k}\right)^\top,$$

where $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_{p-1})$. Defining

$$\Theta = \left(V, \boldsymbol{\iota}/\sqrt{k}\right) \begin{pmatrix} \Lambda^{-1} & 0 \\ 0 & 0 \end{pmatrix} \left(V, \boldsymbol{\iota}/\sqrt{k}\right)^\top,$$

we have $\Sigma\Theta = \mathcal{I}_p - \boldsymbol{\iota}\boldsymbol{\iota}^\top/k$, i.e., Θ is the inverse of Σ in the perpendicular space of $\boldsymbol{\iota}$.

D Identification and Asymptotic Properties

D.1 Notations

For an $n \times m$ matrix A, $\|A\|_p$ is the ℓ_p operator norm defined as

$$\|A\|_p = \sup_{\|\boldsymbol{x}\|_p=1} \|A\boldsymbol{x}\|_p,$$

where $\|\boldsymbol{x}\|_p$ is the standard ℓ_p-norm of a vector $\boldsymbol{x}$, and $|A|_p$ is the element-wise ℓ_p norm defined as

$$|A|_p = \big(\sum_{i,j}|A_{ij}|_p\big)^{1/p}.$$

In particular,

$$\|A\|_\infty = \max_{1\le i\le n}\sum_{j=1}^{m}|A_{ij}|; \qquad |A|_\infty = \max_{i,j}|A_{ij}|.$$

We denote by $\theta_{s_1,s_2}(A)$ the restricted orthogonal constant of s_1 and s_2, defined as

$$\theta_{s_1,s_2}(A) = \sup \frac{|\boldsymbol{r}_1^\top A^\top A\boldsymbol{r}_2|}{\|\boldsymbol{r}_1\|_2\|\boldsymbol{r}_2\|_2},$$

where $\boldsymbol{r}_1$ is a s_1-sparse vector, $\boldsymbol{r}_2$ is a s_2-sparse vector, and $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ have non-overlapping support. The upper and lower restricted isometry property constants of order l are denoted by $\varrho_l^+(A)$ and $\varrho_l^-(A)$, respectively, and defined as

$$\varrho_l^+(A) = \sup \frac{\|A\boldsymbol{r}\|_2^2}{\|\boldsymbol{r}\|_2^2}$$

and

$$\varrho_l^+(A) = \inf \frac{\|A\boldsymbol{r}\|_2^2}{\|\boldsymbol{r}\|_2^2},$$

where $\boldsymbol{r} \in \mathbb{R}^m$ is an l-sparse vector. For a random variable X, $\|X\|_{\psi_1}$ is the sub-exponential norm defined as

$$\|X\|_{\psi_1} = \sup_{q\ge 1} q^{-1}(\mathbb{E}|X|^q)^{1/q},$$

and $\|X\|_{\psi_2}$ is the sub-Gaussian norm defined as

$$\|X\|_{\psi_2} = \sup_{q\ge 1} q^{-1/2}(\mathbb{E}|X|^q)^{1/q}.$$

For a random vector $\boldsymbol{X} \in \mathbb{R}^n$, the sub-exponential norm is defined as

$$\|\boldsymbol{X}\|_{\psi_2} = \sup\big\{\|\boldsymbol{X}^\top\boldsymbol{\alpha}\|_{\psi_2} : \boldsymbol{\alpha} \in \mathbb{R}^n, \|\boldsymbol{\alpha}\|_2 = 1\big\}.$$

D.2 Regularity Conditions

Necessary regularity conditions for asymptotic properties of the de-biased estimator include:

C1. There exist uniform constants, $C_{\min}$ and $C_{\max}$, such that $0 < C_{\min} \le \sigma_{\min}(\Sigma) \le \sigma_{\max}(\Sigma) \le C_{\max} < \infty$, where $\sigma_{\max}(A)(\sigma_{\min}(A))$ is the largest (smallest) non-zero eigenvalue of matrix A.

C2. $\Xi(\boldsymbol{\beta})$ is Lipschitz continuous with a Lipschitz constant υ.

C3. $\big|\sum_{l=1}^{n}\Theta\widetilde{Z}_l\widetilde{Z}_l^\top/n\big|_\infty < \infty$, where $\widetilde{Z}_l$ is a column vector of the l^{th} row of $\widetilde{Z}$.

C4. There exists a uniform constant $\kappa \in (0,\infty)$ such that $\|\Xi_{ll}^{1/2}\Theta^{1/2}\widetilde{Z}_l\|_{\psi_2} \le \kappa$ for all $l = 1,\ldots,n$.

D.3 Model Assumptions

Combined with the stable unit treatment value assumption (SUTVA) (Imbens and Rubin, 2015) and the positivity assumption (i.e., $0 < P(T_i = t|\boldsymbol{X}_i = \boldsymbol{x})$ and $0 < P(\log \boldsymbol{M}_i(t) = \log \boldsymbol{m}|T_i = t, \boldsymbol{X}_i = \boldsymbol{x})$), the CMM requires the following assumptions:

$$\{Y_i(t', \log(\boldsymbol{m})), \log \boldsymbol{M}_i(t)\} \perp\!\!\!\perp T_i|\boldsymbol{X}_i = \boldsymbol{x} \tag{S7}$$

$$Y_i(t', \log(\boldsymbol{m})) \perp\!\!\!\perp \log \boldsymbol{M}_i(t)|T_i = t, \boldsymbol{X}_i = \boldsymbol{x} \tag{S8}$$

for $t, t' \in \mathcal{T}$, $\boldsymbol{m} \in \boldsymbol{\mathcal{M}}$, and $\boldsymbol{x} \in \boldsymbol{\mathcal{X}}$. Assumptions (S7)-(S8) basically state no unmeasured confounding effects after adjusting for $\boldsymbol{X}$.

D.4 Identification of Direct and Indirect Effects

Proof. With the causal assumptions in Section D.3, we have

$$\begin{aligned}
\delta(\tau) &= \mathbb{E}\left[Y_i(\tau, \log \boldsymbol{M}_i(t)) - Y_i(\tau, \log \boldsymbol{M}_i(t'))|\boldsymbol{X}_i = \boldsymbol{x}\right] \\
&= \int \cdots \int \mathbb{E}(Y_i|\text{alt}(\boldsymbol{M}_i) = \text{alt}(\boldsymbol{m}), T_i = \tau, \boldsymbol{X}_i = \boldsymbol{x}) \\
&\qquad \left[dF_{\text{alt}(\boldsymbol{M}_i)|T_i=t,\boldsymbol{X}_i=\boldsymbol{x}}(\text{alt}(\boldsymbol{m})) - dF_{\text{alt}(\boldsymbol{M}_i)|T_i=t',\boldsymbol{X}_i=\boldsymbol{x}}(\text{alt}(\boldsymbol{m}))\right] dF_{\boldsymbol{X}_i}(\boldsymbol{x}) \\
&= \int \cdots \int \Pr\{c_0 + c\,\tau + \boldsymbol{b}_{-k}^\top \text{alt}(\boldsymbol{M}_i) + \boldsymbol{g}^\top \boldsymbol{x} + U_{2i} > 0\} \\
&\qquad \left[dF_{\text{alt}(\boldsymbol{M}_i)|T_i=t,\boldsymbol{X}_i=\boldsymbol{x}}(\text{alt}(\boldsymbol{m})) - dF_{\text{alt}(\boldsymbol{M}_i)|T_i=t',\boldsymbol{X}_i=\boldsymbol{x}}(\text{alt}(\boldsymbol{m}))\right] dF_{\boldsymbol{X}_i}(\boldsymbol{x}) \\
&= \int \cdots \int \Pr\{f_\delta(\tau, \boldsymbol{x}) + \boldsymbol{b}_{-k}^\top \text{alt}(\boldsymbol{a})\, t + \boldsymbol{b}_{-k}^\top \text{alt}(\boldsymbol{U}_{1i}) + U_{2i} > 0\}\, dF(\text{alt}(\boldsymbol{U}_{1i})) \\
&\qquad - \int \cdots \int \Pr\{f_\delta(\tau, \boldsymbol{x}) + \boldsymbol{b}_{-k}^\top \text{alt}(\boldsymbol{a})\, t' + \boldsymbol{b}_{-k}^\top \text{alt}(\boldsymbol{U}_{1i}) + U_{2i} > 0\}\, dF(\text{alt}(\boldsymbol{U}_{1i})) \\
&= \int \cdots \int 1\{\boldsymbol{b}_{-k}^\top \text{alt}(\boldsymbol{U}_{1i}) + U_{2i} > -\boldsymbol{b}_{-k}^\top \text{alt}(\boldsymbol{a})\, t - f_\delta(\tau, \boldsymbol{x})\}\, dF(U_{2i})\, dF(\text{alt}(\boldsymbol{U}_{1i})) \\
&\qquad - \int \cdots \int 1\{\boldsymbol{b}_{-k}^\top \text{alt}(\boldsymbol{U}_{1i}) + U_{2i} > -\boldsymbol{b}_{-k}^\top \text{alt}(\boldsymbol{a})\, t' - f_\delta(\tau, \boldsymbol{x})\}\, dF(U_{2i})\, dF(\text{alt}(\boldsymbol{U}_{1i})) \\
&= \Pr\{\boldsymbol{\varepsilon}_i \leq -\boldsymbol{b}_{-k}^\top \text{alt}(\boldsymbol{a})\, t' - f_\delta(\tau, \boldsymbol{x})\} - \Pr\{\boldsymbol{\varepsilon}_i \leq -\boldsymbol{b}_{-k}^\top \text{alt}(\boldsymbol{a})\, t - f_\delta(\tau, \boldsymbol{x})\}
\end{aligned}$$

where $f_\delta(\tau, \boldsymbol{x}) = c_0 + c\tau + \boldsymbol{b}_{-k}^\top(\text{alt}(\boldsymbol{m}_0) + \sum_{r=1}^{n_x} x_r \text{alt}(\boldsymbol{h}_r)) + \boldsymbol{g}^\top \boldsymbol{x}$ and $\boldsymbol{\varepsilon}_i = \boldsymbol{b}_{-k}^\top \text{alt}(\boldsymbol{U}_{1i}) + U_{2i}$ The second equality is given in Sohn and Li (2019). The fourth equality follows from changing of variables and the independence between T_i and $\text{alt}(\boldsymbol{U}_{1i})$. The fifth equality is due to the independence between $\text{alt}(\boldsymbol{U}_{1i})$ and U_{2i}. Since we assume $U_{2i} \sim N(0, 1)$ and $\boldsymbol{U}_{1i} \sim LN(\boldsymbol{0}, \Sigma)$, we have $\boldsymbol{\varepsilon}_i \sim N(\boldsymbol{0}, \boldsymbol{b}_{-k}^\top \Sigma \boldsymbol{b}_{-k} + 1)$. Note that $\text{alt}(\boldsymbol{U}_{1i}) \sim N(\boldsymbol{0}, \Sigma)$ if and only if $\boldsymbol{U}_{1i} \sim LN(\boldsymbol{0}, \Sigma)$ (Aitchison, 1986). Thus, we have

$$\delta(\tau) = \mathbb{E}\left\{\Phi\left(\frac{(\log \boldsymbol{a})^\top \boldsymbol{b}\, t + f_\delta(\tau, \boldsymbol{X}_i)}{\sqrt{\boldsymbol{b}_{-k}^\top \Sigma \boldsymbol{b}_{-k} + 1}}\right) - \Phi\left(\frac{(\log \boldsymbol{a})^\top \boldsymbol{b}\, t' + f_\delta(\tau, \boldsymbol{X}_i)}{\sqrt{\boldsymbol{b}_{-k}^\top \Sigma \boldsymbol{b}_{-k} + 1}}\right)\right\}. \tag{S9}$$

Similarly, we have

$$\zeta(\tau) = \mathbb{E}\left\{\Phi\left(\frac{ct + f_\zeta(\tau, \boldsymbol{X}_i)}{\sqrt{\boldsymbol{b}_{-k}^\top \Sigma \boldsymbol{b}_{-k} + 1}}\right) - \Phi\left(\frac{ct' + f_\zeta(\tau, \boldsymbol{X}_i)}{\sqrt{\boldsymbol{b}_{-k}^\top \Sigma \boldsymbol{b}_{-k} + 1}}\right)\right\}. \tag{S10}$$

□

D.5 Asymptotic Properties of Debiased Estimators

To show asymptotic behaviors of debiased estimators, we will use the following Theorem S1 and Lemma S1.

Theorem S1. *Let $\boldsymbol{\beta}$ be s-sparse, $\widehat{\boldsymbol{\beta}}$ be the estimator for Objective function (5) given $\boldsymbol{u}$ and Ξ, and $\boldsymbol{\epsilon} = \boldsymbol{u} - \widetilde{Z}\boldsymbol{\beta}$ be sub-Gaussian. If $(3\tau - 1)\varrho^-_{2s}(\Xi^{1/2}\widetilde{Z}/\sqrt{n}) - (\tau + 1)\varrho^+_{2s}(\Xi^{1/2}\widetilde{Z}/\sqrt{n}) \geq 4\tau\phi_0$ for some constant $\phi_0 > 0$ and $\|\widetilde{Z}^\top\Xi\boldsymbol{\epsilon}\|_\infty \leq n\lambda/\tau$, then, with $\lambda = \tau\tilde{\omega}\sqrt{(\log p)/n}$ for some constant $\tilde{\omega} > 0$, the following holds true:*

$$\mathbb{P}\Big(\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\|_1 \geq s\lambda(2 + 1/\tau)/\phi_0\Big) \leq 2p^{-\omega'},$$

where $\omega' = \tilde{\omega}^2/(2K^2) - 1$ and $K^2 = \max_{1\leq j\leq p}\widehat{\Sigma}_{jj}$.

Proof. Theorem S1. Let $\boldsymbol{h} = \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}$, and S_h be the set of indices of the s largest absolute values of $\boldsymbol{h}$. Then, given Ξ and $\boldsymbol{u}$, we have the following inequality

$$\frac{1}{2n}\|\Xi^{1/2}(\boldsymbol{u} - \widetilde{Z}\widehat{\boldsymbol{\beta}})\|_2^2 + \lambda\|\widehat{\boldsymbol{\beta}}\|_1 \leq \frac{1}{2n}\|\Xi^{1/2}(\boldsymbol{u} - \widetilde{Z}\boldsymbol{\beta})\|_2^2 + \lambda\|\boldsymbol{\beta}\|_1$$

Thus, we have

$$\begin{aligned}
&\frac{1}{2n}(\|\Xi^{1/2}(\boldsymbol{\epsilon} - \widetilde{Z}\boldsymbol{h})\|_2^2 - \|\Xi^{1/2}\boldsymbol{\epsilon}\|_2^2) \leq \lambda(\|\boldsymbol{\beta}\|_1 - \|\widehat{\boldsymbol{\beta}}\|_1)\\
\Rightarrow\quad & -\frac{1}{2n}(\widetilde{Z}\boldsymbol{h})^\top\Xi(2\boldsymbol{\epsilon} - \widetilde{Z}\boldsymbol{h}) \leq \lambda(\|\boldsymbol{\beta}_{supp(\beta)}\|_1 - \|\widehat{\boldsymbol{\beta}}_{supp(\beta)}\|_1 - \|\widehat{\boldsymbol{\beta}}_{supp(\beta)^c}\|_1)\\
\Rightarrow\quad & -\frac{1}{n}\boldsymbol{h}^\top\widetilde{Z}^\top\Xi\boldsymbol{\epsilon} \leq \lambda(\|\boldsymbol{\beta}_{supp(\beta)} - \widehat{\boldsymbol{\beta}}_{supp(\beta)}\|_1 - \|\boldsymbol{h}_{supp(\beta)^c}\|_1)\\
\Rightarrow\quad & -\frac{1}{n}\|\widetilde{Z}^\top\Xi\boldsymbol{\epsilon}\|_\infty\|\boldsymbol{h}\|_1 \leq \lambda(\|\boldsymbol{h}_{supp(\beta)}\|_1 - \|\boldsymbol{h}_{supp(\beta)^c}\|_1)\\
\Rightarrow\quad & -\frac{1}{n}\|\widetilde{Z}^\top\Xi\boldsymbol{\epsilon}\|_\infty(\|\boldsymbol{h}_{S_h}\|_1 + \|\boldsymbol{h}_{S_h^c}\|_1) \leq \lambda(\|\boldsymbol{h}_{S_h}\|_1 - \|\boldsymbol{h}_{S_h^c}\|_1)\\
\Rightarrow\quad & -(\|\boldsymbol{h}_{S_h}\|_1 + \|\boldsymbol{h}_{S_h^c}\|_1) \leq \tau(\|\boldsymbol{h}_{S_h}\|_1 - \|\boldsymbol{h}_{S_h^c}\|_1)\ \text{ since } \|\widetilde{Z}^\top\Xi\boldsymbol{\epsilon}\|_\infty \leq n\lambda/\tau\\
\Rightarrow\quad & \|\boldsymbol{h}_{S_h^c}\|_1 \leq \frac{\tau + 1}{\tau - 1}\|\boldsymbol{h}_{S_h}\|_1 \qquad \text{(S11)}
\end{aligned}$$

From the KKT condition of Objective function (5), we have $\|\widetilde{Z}^\top\Xi(\boldsymbol{u} - \widetilde{Z}\widehat{\boldsymbol{\beta}}) - \eta\boldsymbol{\iota}\|_\infty \leq n\lambda$ for some $\eta \in \mathbb{R}$, which provides the following inequalities:

$$\begin{aligned}
\|\widetilde{Z}^\top\Xi(\boldsymbol{u} - \widetilde{Z}\widehat{\boldsymbol{\beta}})\|_\infty &= \|(\mathcal{I}_p - \boldsymbol{\iota}\boldsymbol{\iota}^\top/k)(\widetilde{Z}^\top\Xi(\boldsymbol{u} - \widetilde{Z}\widehat{\boldsymbol{\beta}}) - \eta\boldsymbol{\iota})\|_\infty\\
&= \|(\widetilde{Z}^\top\Xi(\boldsymbol{u} - \widetilde{Z}\widehat{\boldsymbol{\beta}}) - \eta\boldsymbol{\iota}) - \boldsymbol{\iota}\boldsymbol{\iota}^\top(\widetilde{Z}^\top\Xi(\boldsymbol{u} - \widetilde{Z}\widehat{\boldsymbol{\beta}}) - \eta\boldsymbol{\iota})/k\|_\infty\\
&\leq \|\widetilde{Z}^\top\Xi(\boldsymbol{u} - \widetilde{Z}\widehat{\boldsymbol{\beta}}) - \eta\boldsymbol{\iota}\|_\infty + \|\boldsymbol{\iota}\boldsymbol{\iota}^\top(\widetilde{Z}^\top\Xi(\boldsymbol{u} - \widetilde{Z}\widehat{\boldsymbol{\beta}}) - \eta\boldsymbol{\iota})/k\|_\infty\\
&= 2\|\widetilde{Z}^\top\Xi(\boldsymbol{u} - \widetilde{Z}\widehat{\boldsymbol{\beta}}) - \eta\boldsymbol{\iota}\|_\infty\\
&\leq 2n\lambda,
\end{aligned}$$

and

$$\begin{aligned}\|\widetilde{Z}^\top \Xi \widetilde{Z}\boldsymbol{h}\|_\infty &\le \|\widetilde{Z}^\top \Xi(\boldsymbol{u} - \widetilde{Z}\widehat{\boldsymbol{\beta}})\|_\infty + \|\widetilde{Z}^\top \Xi(\boldsymbol{u} - \widetilde{Z}\boldsymbol{\beta})\|_\infty \le 2n\lambda + \|\widetilde{Z}^\top \Xi \boldsymbol{\epsilon}\|_\infty \\ &\le n\lambda(2 + 1/\tau). \end{aligned} \tag{S12}$$

Thus, we have

$$\begin{aligned} n\lambda(2+1/\tau)\|\boldsymbol{h}_{S_h}\|_1 &\ge \|\widetilde{Z}^\top \Xi \widetilde{Z}\boldsymbol{h}\|_\infty \|\boldsymbol{h}_{S_h}\|_1 \ge \langle \widetilde{Z}^\top \Xi \widetilde{Z}\boldsymbol{h}, \boldsymbol{h}_{S_h}\rangle \\ &= \langle \Xi^{1/2}\widetilde{Z}\boldsymbol{h}_{S_h}, \Xi^{1/2}\widetilde{Z}\boldsymbol{h}_{S_h}\rangle + \langle \Xi^{1/2}\widetilde{Z}\boldsymbol{h}_{S_h}, \Xi^{1/2}\widetilde{Z}\boldsymbol{h}_{S_h^c}\rangle. \end{aligned}$$

By Lemma 5.1 in Cai and Zhang (2013), we have

$$\begin{aligned} |\langle \Xi^{1/2}\widetilde{Z}\boldsymbol{h}_{S_h}, \Xi^{1/2}\widetilde{Z}\boldsymbol{h}_{S_h^c}\rangle| &\le \sqrt{s}\theta_{s,s}(\Xi^{1/2}\widetilde{Z})\|\boldsymbol{h}_{S_h}\|_2 \cdot \max(\|\boldsymbol{h}_{S_h^c}\|_\infty, \|\boldsymbol{h}_{S_h^c}\|_1/s) \\ &\le \sqrt{s}\theta_{s,s}(\Xi^{1/2}\widetilde{Z})\|\boldsymbol{h}_{S_h}\|_2 \cdot \frac{\tau+1}{\tau-1}\|\boldsymbol{h}_{S_h}\|_1/s \\ &\le \frac{\tau+1}{\tau-1}\theta_{s,s}(\Xi^{1/2}\widetilde{Z})\|\boldsymbol{h}_{S_h}\|_2^2. \end{aligned}$$

Thus,

$$\begin{aligned} n\lambda(2+1/\tau)\|\boldsymbol{h}_{S_h}\|_1 &\ge \|\Xi^{1/2}\widetilde{Z}\boldsymbol{h}_{S_h}\|_2^2 - \frac{\tau+1}{\tau-1}\theta_{s,s}(\Xi^{1/2}\widetilde{Z})\|\boldsymbol{h}_{S_h}\|_2^2 \\ &\ge \left[\varrho_{2s}^-(\Xi^{1/2}\widetilde{Z}) - \frac{\tau+1}{\tau-1}\theta_{s,s}(\Xi^{1/2}\widetilde{Z})\right]\|\boldsymbol{h}_{S_h}\|_2^2 \\ &\ge \left[\frac{3\tau-1}{2\tau-2}\varrho_{2s}^-(\Xi^{1/2}\widetilde{Z}) - \frac{\tau+1}{2\tau-2}\varrho_{2s}^+(\Xi^{1/2}\widetilde{Z})\right]\|\boldsymbol{h}_{S_h}\|_1^2/s \end{aligned} \tag{S13}$$

The last inequality comes from Lemma 1 of Kang *et al.* (2016) that shows a relationship between $\theta_{s,s}$ and $\varrho_{2s}^\pm$, i.e., $\theta_{k_1.k_2}(A) \le [\varrho_{k_1+k_2}^+(A) - \varrho_{k_1+k_2}^-(A)]/2$ for any matrix A. By rearranging Inequality (S13), we have

$$\|\boldsymbol{h}_{S_h}\|_1 \le \frac{sn\lambda(2+1/\tau)}{n[(3\tau-1)\varrho_{2s}^-(\Xi^{1/2}\widetilde{Z}/\sqrt{n}) - (\tau+1)\varrho_{2s}^+(\Xi^{1/2}\widetilde{Z}/\sqrt{n})]/(2\tau-2)} \le \frac{s\lambda(2+1/\tau)}{2\tau\phi_0/(\tau-1)}.$$

Therefore,

$$\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\|_1 = \|\boldsymbol{h}_{S_h}\|_1 + \|\boldsymbol{h}_{S_h^c}\|_1 \le \frac{2\tau}{\tau-1}\|\boldsymbol{h}_{S_h}\|_1 \le s\lambda(2+1/\tau)/\phi_0,$$

so

$$\begin{aligned} \mathbb{P}(\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\|_1 \ge s\lambda(2+1/\tau)/\phi_0) &\le \mathbb{P}(\|\widetilde{Z}^\top \Xi\boldsymbol{\epsilon}\|_\infty > n\lambda/\tau) \le \sum_{j=1}^p \mathbb{P}(|(\widetilde{Z}^\top \Xi\boldsymbol{\epsilon})_j| > n\lambda/\tau) \\ &\le 2p\exp\left(-\frac{n\lambda^2}{2\tau^2K^2}\right) = 2p^{1-\tilde{\omega}^2/(2K^2)}. \end{aligned}$$

□

Lemma S1. *Suppose the regularity conditions hold, then for any constant $\omega > 0$, the following inequality holds:*

$$\mathbb{P}\left\{\left|\Theta\widehat{\Sigma} - (\mathcal{I}_p - \boldsymbol{\iota}\boldsymbol{\iota}^\top/k)\right|_\infty \ge \omega\sqrt{(\log p)/n}\right\} \le 2p^{-(\omega_1''+\omega_2'')},$$

where $\omega_1'' = \omega^2 C_{\min}/(24e^2\kappa^4 C_{\max}) - 2$ and $\omega_2'' = (\omega\phi_0)^2/[2(\upsilon' s(2\tau+1)K)^2] - 1$.

Proof. Lemma S1.

$$\left|\Theta\widehat{\Sigma} - (\mathcal{I}_p - \boldsymbol{\iota}\boldsymbol{\iota}^\top/k)\right|_\infty \le \left|\Theta\Sigma - (\mathcal{I}_p - \boldsymbol{\iota}\boldsymbol{\iota}^\top/k)\right|_\infty + \left|\Theta(\Sigma - \widehat{\Sigma})\right|_\infty$$

Since $\Sigma^{1/2}\Theta^{1/2}\widetilde{Z}_l = (\mathcal{I}_p - \boldsymbol{\iota}\boldsymbol{\iota}^\top/k)\widetilde{Z}_l = \widetilde{Z}_l$ for $l = 1, \ldots, n$, we have

$$\begin{aligned}
\Theta\Sigma - (\mathcal{I}_p - \boldsymbol{\iota}\boldsymbol{\iota}^\top/k) &= \frac{1}{n}\sum_{l=1}^{n}\Big\{\Xi_{ll}\Theta\widetilde{Z}_l\widetilde{Z}_l^\top - (\mathcal{I}_p - \boldsymbol{\iota}\boldsymbol{\iota}^\top/k)\Big\} \\
&= \frac{1}{n}\sum_{l=1}^{n}\Big\{\Xi_{ll}\Theta^{1/2}\Theta^{1/2}\widetilde{Z}_l\widetilde{Z}_l^\top\Theta^{1/2}\Sigma^{1/2} - (\mathcal{I}_p - \boldsymbol{\iota}\boldsymbol{\iota}^\top/k)\Big\}
\end{aligned}$$

For $i, j = 1, \ldots, p$, define $v_l^{(ij)} = \Xi_{ll}\Theta_{i\cdot}^{1/2}\Theta^{1/2}\widetilde{Z}_l\widetilde{Z}_l^\top\Theta^{1/2}\Sigma_{\cdot j}^{1/2} - (\mathcal{I}_p - \boldsymbol{\iota}\boldsymbol{\iota}^\top/k)_{ij}$, where $A_{k\cdot}$ is the k^{th} row vector of matrix A and $A_{\cdot k}$ is the k^{th} column vector of matrix A. Notice that $\mathbb{E}v_l^{(ij)} = 0$ since $\mathbb{E}(\Xi_{ll}\Theta\widetilde{Z}_l\widetilde{Z}_l^\top) = \Theta\Sigma = \mathcal{I}_p - \boldsymbol{\iota}\boldsymbol{\iota}^\top/k$. Following the proof of Lemma 23 in Javanmard and Montanari (2014), we have

$$\begin{aligned}
\|v_l^{(ij)}\|_{\psi_1} &\le 2\|\Xi_{ll}\Theta_{i\cdot}^{1/2}\Theta^{1/2}\widetilde{Z}_l\widetilde{Z}_l^\top\Theta^{1/2}\Sigma_{\cdot j}^{1/2}\|_{\psi_1} \\
&\le 2\|\Xi_{ll}^{1/2}\Theta_{i\cdot}^{1/2}\Theta^{1/2}\widetilde{Z}_l\|_{\psi_2}\|\Xi_{ll}^{1/2}\Sigma_{j\cdot}^{1/2}\Theta^{1/2}\widetilde{Z}_l\|_{\psi_2} \\
&\le 2\|\Theta_{i\cdot}^{1/2}\|_2\|\Sigma_{j\cdot}^{1/2}\|_2\|\Xi_{ll}^{1/2}\Theta^{1/2}\widetilde{Z}_l\|_{\psi_2}\|\Xi_{ll}^{1/2}\Theta^{1/2}\widetilde{Z}_l\|_{\psi_2} \\
&\le 2\kappa^2\sqrt{C_{\max}/C_{\min}}.
\end{aligned}$$

Let $\kappa' = 2\sqrt{C_{\max}/C_{\min}}\kappa^2$. Then, by the Bernstein-type inequality for centered sub-exponential random variable (Bühlmann and van de Geer, 2011), we have

$$\mathbb{P}\Big\{\frac{1}{n}\Big|\sum_{l=1}^{n}v_l^{(ij)}\Big| \ge \gamma\Big\} \le 2\exp\Big\{-\frac{n}{6}\min\Big[\Big(\frac{\gamma}{e\kappa'}\Big)^2, \frac{\gamma}{e\kappa'}\Big]\Big\}.$$

Choosing $\gamma = \omega\sqrt{(\log p)/n}$ with $\omega \le e\kappa'\sqrt{n/\log p}$, we have

$$\mathbb{P}\Big\{\frac{1}{n}\Big|\sum_{l=1}^{n}v_l^{(ij)}\Big| \ge \omega\sqrt{(\log p)/n}\Big\} \le 2p^{-\omega^2/(6e^2\kappa'^2)} = 2p^{-\omega^2 C_{\min}/(24e^2\kappa^4 C_{\max})}.$$

By union bounding over all pairs of (i, j), we have

$$\mathbb{P}\left\{\left|\Theta\Sigma - (\mathcal{I}_p - \boldsymbol{\iota}\boldsymbol{\iota}^\top/k)\right|_\infty \ge \omega\sqrt{(\log p)/n}\right\} \le 2p^{-\omega^2 C_{\min}/(24e^2\kappa^4 C_{\max})+2}. \tag{S14}$$

Define $\upsilon' = |\upsilon\sum_{l=1}^{n}\Theta\widetilde{Z}_l\widetilde{Z}_l^\top/n|_\infty$. Then,

$$\left|\Theta(\Sigma - \widehat{\Sigma})\right|_\infty = \frac{1}{n}\Big|\sum_{l=1}^{n}[(\Xi_{ll} - \widehat{\Xi}_{ll})\Theta\widetilde{Z}_l\widetilde{Z}_l^\top]\Big|_\infty \le \upsilon'\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\|_1.$$

Thus, by Theorem S1, we have

$$\begin{aligned}
\mathbb{P}\Big\{\left|\Theta(\Sigma - \widehat{\Sigma})\right|_\infty \ge \omega\sqrt{(\log p)/n}\Big\} &\le \mathbb{P}\Big\{\upsilon'\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\|_1 \ge \omega\sqrt{(\log p)/n}\Big\} \\
&\le 2p^{1-(\omega\phi_0)^2/[2(\upsilon' s(2\tau+1)K)^2]}.
\end{aligned} \tag{S15}$$

With Bounds (S14)-(S15), we have

$$\mathbb{P}\left\{\left|\Theta\widehat{\Sigma} - (\mathcal{I}_p - \boldsymbol{\iota}\boldsymbol{\iota}^\top/k)\right|_\infty \ge \omega\sqrt{(\log p)/n}\right\} \le 2p^{-\omega_1''} + 2p^{-\omega_2''}.$$

□

Proof. Asymptotic properties of debiased estimators. Since $\boldsymbol{\iota}^\top \boldsymbol{\beta} = \boldsymbol{\iota}^\top \widehat{\boldsymbol{\beta}} = 0$, we have

$$\begin{aligned}
\widehat{\boldsymbol{\beta}}_{db} - \boldsymbol{\beta} &= \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} + \frac{1}{n}\widetilde{\Theta}\widetilde{Z}^\top \widehat{\Xi}\,(\widehat{\boldsymbol{u}} - \widetilde{Z}\widehat{\boldsymbol{\beta}}) \\
&= \frac{1}{n}\widetilde{\Theta}\widetilde{Z}^\top \widehat{\Xi}\boldsymbol{\epsilon} + (\widetilde{\Theta}\widehat{\Sigma} - \mathcal{I}_p)(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}) \\
&= \frac{1}{n}\widetilde{\Theta}\widetilde{Z}^\top \widehat{\Xi}\boldsymbol{\epsilon} + [\widetilde{\Theta}\widehat{\Sigma} - (\mathcal{I}_p - \boldsymbol{\iota}\boldsymbol{\iota}^\top/k)](\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}).
\end{aligned}$$

Thus, $\sqrt{n}(\widehat{\boldsymbol{\beta}}_{db} - \boldsymbol{\beta}) = R + \Delta$, where $R = \widetilde{\Theta}\widetilde{Z}^\top \widehat{\Xi}\boldsymbol{\epsilon}/\sqrt{n}$ and $\Delta = \sqrt{n}[\widetilde{\Theta}\widehat{\Sigma} - (\mathcal{I}_p - \boldsymbol{\iota}\boldsymbol{\iota}^\top/k)](\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}})$. Since $\mathbb{E}(\boldsymbol{\epsilon}|\widehat{\Xi}, \widehat{\boldsymbol{u}}) = \mathbf{0}$, we have $\mathbb{E}(R|\widetilde{Z}, \widehat{\Xi}, \widehat{\boldsymbol{u}}) = \mathbf{0}$ and

$$\begin{aligned}
\|\Delta\|_\infty &\le \sqrt{n}\big|\widetilde{\Theta}\widehat{\Sigma} - (\mathcal{I}_p - \boldsymbol{\iota}\boldsymbol{\iota}^\top/k)\big|_\infty \|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\|_1 \\
&= \sqrt{n}\big|(\mathcal{I}_p - \boldsymbol{\iota}\boldsymbol{\iota}^\top/k)[\Theta\widehat{\Sigma} - (\mathcal{I}_p - \boldsymbol{\iota}\boldsymbol{\iota}^\top/k)]\big|_\infty \|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\|_1 \\
&\le \sqrt{n}\big\{\big|\Theta\widehat{\Sigma} - (\mathcal{I}_p - \boldsymbol{\iota}\boldsymbol{\iota}^\top/k)\big|_\infty + \big|\boldsymbol{\iota}\boldsymbol{\iota}^\top[\Theta\widehat{\Sigma} - (\mathcal{I}_p - \boldsymbol{\iota}\boldsymbol{\iota}^\top/k)]/k\big|_\infty\big\}\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\|_1 \\
&\le 2\sqrt{n}\big|\Theta\widehat{\Sigma} - (\mathcal{I}_p - \boldsymbol{\iota}\boldsymbol{\iota}^\top/k)\big|_\infty \|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\|_1.
\end{aligned}$$

Therefore, by Lemma S1 and Theorem S1, we have

$$\begin{aligned}
\mathbb{P}\Big(\|\Delta\|_\infty \ge \frac{2\omega\tilde{\omega}(2\tau+1)s\log p}{\phi_0\sqrt{n}}\Big) &\le \mathbb{P}\Big(\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\|_1 \ge s\lambda(2 + 1/\tau)/\phi_0\Big) \\
&\quad + \mathbb{P}\Big(\big|\Theta\widehat{\Sigma} - (\mathcal{I}_p - \boldsymbol{\iota}\boldsymbol{\iota}^\top/k)\big|_\infty \ge \omega\sqrt{(\log p)/n}\Big) \\
&\le 2p^{-\omega'} + 2p^{-\omega_1''} + 2p^{-\omega_2''}.
\end{aligned}$$

□

E Details of the Sensitivity Analysis

Because of the linear constraint, $\mathbf{1}_k^\top \boldsymbol{b} = 0$, we can write Model (2), excluding the covariates $\boldsymbol{X}_i$ for notation simplicity, as

$$\begin{aligned}
Y_i &= 1\{c_0 + c\,T_i + (\boldsymbol{b}_\rho)_{-k}^\top \text{alt}(\boldsymbol{M}_i) + U_{2i} > 0\} \\
&= 1\{c_0 + c\,T_i + (\boldsymbol{b}_\rho)_{-k}^\top (\text{alt}(\boldsymbol{m}_0) + \text{alt}(\boldsymbol{a})T_i + \text{alt}(\boldsymbol{U}_{1i})) + U_{2i} > 0\} \\
&= 1\{c_0^* + c^*\,T_i + U_{0i}^* > 0\},
\end{aligned}$$

where $(\boldsymbol{b}_\rho)_{-k} = ((b_1)_\rho, \ldots, (b_{k-1})_\rho)^\top$, $c_0^* = c_0 + (\boldsymbol{b}_\rho)_{-k}^\top \text{alt}(\boldsymbol{m}_0)$, $c^* = c + (\boldsymbol{b}_\rho)_{-k}^\top \text{alt}(\boldsymbol{a})$, and $U_{0i}^* = (\boldsymbol{b}_\rho)_{-k}^\top \text{alt}(\boldsymbol{U}_{1i}) + U_{2i}$. Thus, a probit regression model for the total effect of T on Y can be expressed as

$$Y_i = 1\{\tilde{c}_0 + \tilde{c}\,T_i + U_{0i} > 0\},$$

where $\tilde{c}_0 = c_0^*/\Psi(\boldsymbol{\rho}, \boldsymbol{b}_\rho, \Sigma)$, $\tilde{c} = c^*/\Psi(\boldsymbol{\rho}, \boldsymbol{b}_\rho, \Sigma)$, and $U_{0i} = U_{0i}^*/\Psi(\boldsymbol{\rho}, \boldsymbol{b}_\rho, \Sigma)$, where $\Psi(\boldsymbol{\rho}, \boldsymbol{b}_\rho, \Sigma) = \big[(\boldsymbol{b}_\rho)_{-k}^\top \Sigma (\boldsymbol{b}_\rho)_{-k} + 2\rho\,(\boldsymbol{b}_\rho)_{-k}^\top \text{diag}(\Sigma)^{1/2} + 1\big]^{1/2}$. For a given j, we have

$$\text{Cov}\,[U_{0i}, \text{alt}(\boldsymbol{U}_{1i})_j] = \big((\boldsymbol{b}_\rho)_{-k}^\top \Sigma_j + \rho\Sigma_{jj}^{1/2}\big)/\Psi(\boldsymbol{\rho}, \boldsymbol{b}_\rho, \Sigma), \tag{S16}$$

where Σ_j is the j^{th} column of Σ and Σ_{jj} is the variance of $\text{alt}(\boldsymbol{U}_{1i})_j$. Because of the constraint on U_{0i} and U_{2i}, i.e., they follow the standard gaussian distribution, we can estimate $\text{Cov}\,[U_{0i}, \text{alt}(\boldsymbol{U}_{1i})_j]$ with an estimated $\boldsymbol{b}_{-k}$ under the assumption of $\rho = 0$. That is, we have $k-1$ unknown parameters $(\boldsymbol{b}_\rho)_{-k}$ and $k-1$ equations for a given value of ρ.

References

Aitchison, J. (1986). *The Statistical Analysis of Compositional Data*. New York: Chapman & Hall.

Bühlmann, P. and van de Geer, S. (2011). *Statistics for High-Dimensional Data: Method, Theory and Applications*. Berlin: Springer.

Cai, T. T. and Zhang, A. (2013). Compressed sensing and affine rank minimization under restricted isometry. *Signal Processing, IEEE Transactions*. **61**, 3279-3290.

Imbens, G. and Rubin, D. (2015). *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge: Cambridge University Press.

Javanmard, A. and Montanari, A. (2014). Confidence intervals and hypothesis testing for high-dimensional regression. *J. Mach. Learn. Res.*, **15**(1), 2869-2909.

Kang, H., Zhang, A., Cai, T. T., and Small, D. S. (2016). Instrumental variables estimation with some invalid instruments and its application to mendelian randomization. *Journal of the American Statistical Association*. **111**(513),132-144.

Sohn, M. B. and Li, H. (2019). Compositional mediation analysis for microbiome studies. *Annals of Applied Statistics*, **13**(1), 661-681.