Transfer Learning for High-Dimensional Regression with Compositional Covariates: Application to Microbiome Studies

Qinqin Hu *[1], Xiaojing Luo[1], Chencheng Ma[2] and Wang Zhou[3]

[1] *Shandong University at Weihai,* [2] *Shandong University and* [3] *National University of Singapore.*

Supplementary Material

This supplementary material provides additional technical details and numerical results that complement the main paper. Specifically, Section S1 presents the details of Algorithm 1, Section S2 and S3 contain the proofs of Theorems 1 and 2, respectively. Section S4 offers further comparison with other constrained methods, while Section S5 includes additional numerical and real data analysis results.

S1 The details of Algorithm 1

S1.1 $\boldsymbol{w}^{\mathcal{A}}$'s Expression

This part provides the detailed derivation of the expression for $\boldsymbol{w}^{A}$ used in the main paper. Since $\boldsymbol{w}^{\mathcal{A}}$ is the solution of the following constrained optimization

$$\min_{w} \sum_{k\in\{0\}\bigcup\mathcal{A}} \alpha_k \mathbb{E}\left[(\boldsymbol{y}^{(k)}-\mathbf{Z}^{(k)}\boldsymbol{w})^{\top}(\boldsymbol{y}^{(k)}-\mathbf{Z}^{(k)}\boldsymbol{w})\right], \quad \text{subject to } \mathbf{C}^{\top}\boldsymbol{w}=0,$$

First, the Lagrange of optimization problem is formed as,

$$L(\boldsymbol{w},\lambda) = -2\boldsymbol{w}^{\top}\mathbb{E}(\sum_{k\in\{0\}\bigcup\mathcal{A}} \alpha_k \mathbf{Z}^{(k)\top}\boldsymbol{y}^{(k)}) + \boldsymbol{w}^{\top}\sum_{k\in\{0\}\bigcup\mathcal{A}} \alpha_k \mathbb{E}(\mathbf{Z}^{(k)\top}\mathbf{Z}^{(k)})\boldsymbol{w} + \lambda\mathbf{C}^{\top}\boldsymbol{w},$$

where $\lambda \in \mathbb{R}$ is the Lagrange multiplier. Then we have that the solution $\boldsymbol{w}^{\mathcal{A}}$ satisfies the following equations.

$$\begin{aligned} 0 = \frac{\partial L}{\partial \boldsymbol{w}} &= -2\mathbb{E}(\sum_{k\in\{0\}\bigcup\mathcal{A}} \alpha_k \mathbf{Z}^{(k)\top}\boldsymbol{y}^{(k)}) + 2\sum_{k\in\{0\}\bigcup\mathcal{A}} \alpha_k \mathbb{E}(\mathbf{Z}^{(k)\top}\mathbf{Z}^{(k)})\boldsymbol{w} + \lambda\mathbf{C} \\ &= -2\sum_{k\in\{0\}\bigcup\mathcal{A}} \alpha_k \boldsymbol{\Sigma}^{(k)}\boldsymbol{w}^{*(k)} + 2\boldsymbol{\Sigma}\boldsymbol{w} + \lambda\mathbf{C}, \end{aligned}$$

where $\boldsymbol{\Sigma}^{(k)} = \mathbb{E}\mathbf{Z}^{(k)\top}\mathbf{Z}^{(k)}$ and $\boldsymbol{\Sigma} = \sum_{k\in\{0\}\bigcup\mathcal{A}} \alpha_k \boldsymbol{\Sigma}^{(k)}$.

Assuming that the composition data $\mathbf{X}$ are independent of the non-composition $\mathbf{N}$, then

$$\boldsymbol{\Sigma}^{(k)} = \begin{bmatrix} \boldsymbol{\Sigma}_X^{(k)} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_N^{(k)} \end{bmatrix},$$

where $\boldsymbol{\Sigma}_X^{(k)} = \mathbb{E}\tilde{\tilde{\mathbf{Z}}}^{(k)\top}\tilde{\tilde{\mathbf{Z}}}^{(k)}$ and $\boldsymbol{\Sigma}_N^{(k)} = \mathbb{E}\mathbf{N}^{(k)\top}\mathbf{N}^{(k)}$. Due to the zero-sum constraint on subgroup of $\tilde{\tilde{\mathbf{Z}}}^{(k)}$, $\boldsymbol{\Sigma}_X^{(k)}$ has an eigenvalue 0 with multiplicity G, which is equal to the number of groups, the matrix formed by its corresponding eigenvectors as rows is

$$\tilde{\mathbf{C}}_s^{\top} = \begin{bmatrix} \mathbf{1}_{p_1}^{\top}/\sqrt{p_1} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{1}_{p_2}^{\top}/\sqrt{p_2} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{1}_{p_G}^{\top}/\sqrt{p_G} \end{bmatrix}_{G\times q}.$$

Notice that for $k = 0, \dots, K$, $\alpha_k > 0$ and $\boldsymbol{\Sigma}_X^{(k)}$ has $q - G$ positive eigenvalues, and the zero eigenvalue has algebraic multiplicity G. The matrix $\tilde{\mathbf{C}}_s^{\top}$ is formed by taking the eigenvectors corresponding to the zero eigenvalue as its rows, hence $\boldsymbol{\Sigma}_X = \sum_{k\in\{0\}\bigcup\mathcal{A}} \alpha_k \boldsymbol{\Sigma}_X^{(k)}$ also has the same properties. Then let $\boldsymbol{\Sigma}_X = \mathbf{V}_X\mathbf{D}_X\mathbf{V}_X^{\top}$ be the spectral decomposition of $\boldsymbol{\Sigma}_X$, where $\mathbf{V}_X$

is a $q\times(q-G)$ matrix of eigenvectors and $\mathbf{D}_X = diag(d_1, \ldots, d_{q-G})$ with $d_1 \geq \ldots \geq d_{q-G} > 0$.

For any constant $\rho > 0$, $\mathbf{\Sigma}_X + \rho\tilde{\mathbf{C}}_s\tilde{\mathbf{C}}_s^\top$ is non-singular and has the spectral decomposition

$$\mathbf{\Sigma}_X + \rho\tilde{\mathbf{C}}_s\tilde{\mathbf{C}}_s^\top = (\mathbf{V}, \tilde{\mathbf{C}}_s)\begin{pmatrix} \mathbf{D} & \mathbf{0} \\ \mathbf{0} & \rho\mathbf{I}_G \end{pmatrix}(\mathbf{V}, \tilde{\mathbf{C}}_s)^\top,$$

whose inverse is given by

$$(\mathbf{\Sigma}_X + \rho\tilde{\mathbf{C}}_s\tilde{\mathbf{C}}_s^\top)^{-1} = (\mathbf{V}, \tilde{\mathbf{C}}_s)\begin{pmatrix} \mathbf{D}^{-1} & \mathbf{0} \\ \mathbf{0} & \rho^{-1}\mathbf{I}_G \end{pmatrix}(\mathbf{V}, \tilde{\mathbf{C}}_s)^\top,$$

and the Moore-Penrose inverse of $\mathbf{\Sigma}_X$ can be expressed as

$$\mathbf{\Omega}_{Xc} = (\mathbf{\Sigma}_X + \rho\tilde{\mathbf{C}}_s\tilde{\mathbf{C}}_s^\top)^{-1} - \frac{1}{\rho}\tilde{\mathbf{C}}_s\tilde{\mathbf{C}}_s^\top.$$

According to Zhang et al. (2025), the matrix $\mathbf{\Omega}_{Xc}$ satisfies the following properties

$$\mathbf{\Sigma}_X\mathbf{\Omega}_{Xc} = \mathbf{G}_X, \quad \mathbf{G}_X\mathbf{\Omega}_{Xc} = \mathbf{\Omega}_{Xc}, \quad \text{and} \quad \mathbf{\Omega}_{Xc}\mathbf{C}_s = \mathbf{\Omega}_{Xc}\tilde{\mathbf{C}}_s = \mathbf{0},$$

where $\mathbf{G}_X = \boldsymbol{I}_q - \tilde{\mathbf{C}}_s\tilde{\mathbf{C}}_s^\top$.

We can define

$$\mathbf{\Omega}_c = \begin{pmatrix} \mathbf{\Omega}_{Xc} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma}_N^{-1} \end{pmatrix}$$

as the Moore-Penrose inverse of $\mathbf{\Sigma}$, where

$$\mathbf{\Sigma} = \begin{pmatrix} \mathbf{\Sigma}_X & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma}_N \end{pmatrix}, \quad \text{and} \quad \mathbf{\Sigma}_N = \sum_{k\in\{0\}\bigcup\mathcal{A}} \alpha_k\mathbf{\Sigma}_N^{(k)}.$$

Then we can get

$$\boldsymbol{\Sigma\Omega}_c = \mathbf{G}, \quad \mathbf{G\Omega}_c = \boldsymbol{\Omega}_c, \quad \text{and} \quad \boldsymbol{\Omega}_c\mathbf{C} = \boldsymbol{\Omega}_c\tilde{\mathbf{C}} = \mathbf{0},$$

where $\mathbf{G} = \boldsymbol{I}_p - P_{\tilde{C}} = \boldsymbol{I}_p - \tilde{\mathbf{C}}\tilde{\mathbf{C}}^\top$ and $\tilde{\mathbf{C}}^\top = (\tilde{\mathbf{C}}_s^\top \mathbf{0}_{G\times q'})^\top$.

Therefore, we have

$$\begin{aligned} 0 &= -2\boldsymbol{\Omega}_c \sum_{k\in\{0\}\bigcup\mathcal{A}} \alpha_k \boldsymbol{\Sigma}^{(k)} \boldsymbol{w}^{*(k)} + 2\boldsymbol{\Omega}_c \boldsymbol{\Sigma} \boldsymbol{w}^{\mathcal{A}} + \lambda \boldsymbol{\Omega}_c \mathbf{C} \\ &= -2\boldsymbol{\Omega}_c \sum_{k\in\{0\}\bigcup\mathcal{A}} \alpha_k \boldsymbol{\Sigma}^{(k)} \boldsymbol{w}^{*(k)} + 2(\boldsymbol{I}_p - P_{\tilde{C}}) \boldsymbol{w}^{\mathcal{A}} \end{aligned}$$

That means

$$\boldsymbol{\Omega}_c \sum_{k\in\{0\}\bigcup\mathcal{A}} \alpha_k \boldsymbol{\Sigma}^{(k)} \boldsymbol{w}^{*(k)} = (\boldsymbol{I}_p - P_{\tilde{C}}) \boldsymbol{w}^{\mathcal{A}},$$

then

$$G\boldsymbol{\Omega}_c \sum_{k\in\{0\}\bigcup\mathcal{A}} \alpha_k \boldsymbol{\Sigma}^{(k)} \boldsymbol{w}^{*(k)} = \boldsymbol{\Sigma\Omega}_c\boldsymbol{\Omega}_c \sum_{k\in\{0\}\bigcup\mathcal{A}} \alpha_k \boldsymbol{\Sigma}^{(k)} \boldsymbol{w}^{*(k)} = \boldsymbol{\Sigma\Omega}_c(\boldsymbol{I}_p - P_{\tilde{C}})\boldsymbol{w}^{\mathcal{A}} = \boldsymbol{\Sigma\Omega}_c\boldsymbol{w}^{\mathcal{A}} = G\boldsymbol{w}^{\mathcal{A}},$$

the last second equality due to $\boldsymbol{\Omega}_c\mathbf{C} = \mathbf{0}$.

Denote $\tilde{\boldsymbol{w}} = \boldsymbol{\Omega}_c \sum_{k\in\{0\}\bigcup\mathcal{A}} \alpha_k \boldsymbol{\Sigma}^{(k)} \boldsymbol{w}^{*(k)}$, and hence $\boldsymbol{w}^{\mathcal{A}} = \tilde{\boldsymbol{w}} + \tilde{\mathbf{C}}\boldsymbol{t}$. Since $\tilde{C}^\top \boldsymbol{w}^{\mathcal{A}} = 0$ and $\tilde{\mathbf{C}}^\top \tilde{\boldsymbol{w}} = 0$, we can have

$$\mathbf{0} = \tilde{\mathbf{C}}^\top \boldsymbol{w}^{\mathcal{A}} = \tilde{\mathbf{C}}^\top(\tilde{\boldsymbol{w}} + \tilde{\mathbf{C}}\boldsymbol{t}) = \mathbf{0} + \tilde{\mathbf{C}}^\top\tilde{\mathbf{C}}\boldsymbol{t} = \tilde{\mathbf{C}}^\top\tilde{\mathbf{C}}\boldsymbol{t} = \boldsymbol{t},$$

hence we can get $\boldsymbol{t} = \mathbf{0}$. Therefore, we can get the solution

$$\boldsymbol{w}^{\mathcal{A}} = \tilde{\boldsymbol{w}} = \boldsymbol{\Omega}_c \sum_{k\in\{0\}\bigcup\mathcal{A}} \alpha_k \boldsymbol{\Sigma}^{(k)} \boldsymbol{w}^{*(k)}.$$

S1.2 Implementation of Algorithm 1

In order to solve the two constrained optimization problems in Algorithm 1 more conveniently, we define artificial data sets $\mathbb{Z} \in \mathbb{R}^{(n_{\mathcal{A}}+n_0)\times p}$ and $\mathbf{y} \in \mathbb{R}^{n_{\mathcal{A}}+n_0}$ as follows:

$$\mathbb{Z} = \begin{bmatrix} \mathbf{Z}^{(0)} \\ \mathbf{Z}^{(1)} \\ \cdots \\ \mathbf{Z}^{(K)} \end{bmatrix}, \mathbf{y} = \begin{bmatrix} \boldsymbol{y}^{(0)} \\ \boldsymbol{y}^{(1)} \\ \cdots \\ \boldsymbol{y}^{(K)} \end{bmatrix} \quad \text{and} \quad \widetilde{\boldsymbol{y}^{(0)}} = \boldsymbol{y}^{(0)} - \mathbf{Z}^{(0)}\hat{\boldsymbol{w}}^{\mathcal{A}}.$$

Hence, the criteria to be minimized in Algorithm 1 reduces to minimize the following problems:

$$\hat{\boldsymbol{w}}^{\mathcal{A}} = \arg\min_{\boldsymbol{w}\in\mathbb{R}^p} \left\{ \frac{1}{2(n_{\mathcal{A}}+n_0)} \|\mathbf{y} - \mathbb{Z}\boldsymbol{w}\|_2^2 + \lambda_w \|\boldsymbol{w}\|_1 \right\}, \text{s.t. } \mathbf{C}^\top \boldsymbol{w} = \mathbf{0}, \tag{S1.1}$$

$$\hat{\boldsymbol{\delta}}^{\mathcal{A}} = \arg\min_{\boldsymbol{\delta}\in\mathbb{R}^p} \left\{ \frac{1}{2n_0} \|\widetilde{\boldsymbol{y}^{(0)}} - \mathbf{Z}^{(0)}\boldsymbol{\delta}\|_2^2 + \lambda_\delta \|\boldsymbol{\delta}\|_1 \right\}, \text{s.t. } \mathbf{C}^\top \boldsymbol{\delta} = \mathbf{0}. \tag{S1.2}$$

Similarly to Lin et al. (2014) and Shi et al. (2016), a coordinate descent method of multipliers can be used to implement the constrained optimization problems (S1.1) and (S1.2). First, the augmented Lagrange of optimization problem (S1.1) (Bertsekas, 1996) is formed as

$$L_\mu(\boldsymbol{w}, \eta) = \frac{1}{2(n_{\mathcal{A}}+n_0)} \|\mathbf{y} - \mathbb{Z}\boldsymbol{w}\|_2^2 + \lambda_w \|\boldsymbol{w}\|_1 + \eta^\top \mathbf{C}^\top \boldsymbol{w} + \frac{\mu}{2} \|\mathbf{C}^\top \boldsymbol{w}\|_2^2,$$

where $\eta \in \mathbb{R}^G$ is the Lagrange multiplier, and $\mu > 0$ is a penalty parameter. The method of

Algorithm S1 : Step 1 of the Oracle Trans-sub-Coda-Lasso Algorithm

step 1.1: Initialize w^0 with 0 or a warm start, $\nu^0 = 0$, $\mu > 0$ and $t = 0$.
step 1.2: For $j = 1, \ldots, p, 1, \ldots, p, \ldots$, update w_j^{t+1} by (S1.5) until convergence.
step 1.3: Update ν^{t+1} by (S1.4).
step 1.4: $t \leftarrow t + 1$ and repeat the two steps above until convergence.

multipliers for problem (S1.1) consists of the iterations

$$\boldsymbol{w}^{t+1} \leftarrow \arg\min_{\boldsymbol{w}} L_\mu(\boldsymbol{w}, \eta^t), \qquad \eta^{t+1} \leftarrow \eta^t + \mu \mathbf{C}^\top \boldsymbol{w}^{t+1}.$$

Defining $\nu^t = \eta^t/\mu$, the iterations become

$$\boldsymbol{w}^{t+1} \leftarrow \arg\min_{\boldsymbol{w}} \left\{ \frac{1}{2(n_{\mathcal{A}} + n_0)} \|\mathbf{y} - \mathbb{Z}\boldsymbol{w}\|_2^2 + \lambda_w \|\boldsymbol{w}\|_1 + \frac{\mu}{2} \|\mathbf{C}^\top \boldsymbol{w} + \nu^t\|_2^2 \right\}, \tag{S1.3}$$

$$\nu^{t+1} \leftarrow \nu^t + \mathbf{C}^\top \boldsymbol{w}^{t+1}. \tag{S1.4}$$

The iteration of $\boldsymbol{w}$ can be further detailed as

$$\boldsymbol{w}_j^{t+1} \leftarrow \frac{1}{\frac{\|\boldsymbol{z}_j\|_2^2}{n_{\mathcal{A}}+n_0} + \mu \|C_j\|_2^2} S_{\lambda_w} \left\{ \frac{1}{n_{\mathcal{A}} + n_0} \boldsymbol{z}_j^\top \left(\mathbf{y} - \sum_{i \neq j} \boldsymbol{w}_i^{(t+1)} \boldsymbol{z}_i \right) - \mu \left(\sum_{i \neq j} \boldsymbol{w}_i^{(t+1)} C_j^\top C_i + C_j^\top \nu^t \right) \right\}, \tag{S1.5}$$

where $C_j, j = 1, \ldots, p$ are the columns of $\mathbf{C}^\top$, $\boldsymbol{z}_j, j = 1, \ldots, p$ are columns of $\mathbb{Z}$, and $S_{\lambda_w}(x) = sgn(x)(|x| - \lambda_w)_+$ is the soft thresholding operator. Combining (S1.3)-(S1.5) yields the following algorithm to solve problem (S1.1), that is, Step 1 in Algorithm 1 for Oracle Trans-sub-Coda-Lasso.

We can also use the similar coordinate descent method of multipliers to solve the problem (S1.2).

$$\boldsymbol{\delta}^{t+1} \leftarrow \arg\min_{\boldsymbol{\delta}} \left\{ \frac{1}{2n_0} \|\widetilde{\boldsymbol{y}^{(0)}} - \mathbf{Z}^{(0)} \boldsymbol{\delta}\|_2^2 + \lambda_\delta \|\boldsymbol{\delta}\|_1 + \frac{\mu}{2} \|\mathbf{C}^\top \boldsymbol{\delta} + \nu^t\|_2^2 \right\}, \tag{S1.6}$$

Algorithm S2 : Step 2 of the Oracle Trans-sub-Coda-Lasso Algorithm

step 2.1: Initialize δ^0 with 0 or a warm start, $\nu^0 = 0$, $\mu > 0$ and $t = 0$.
step 2.2: For $j = 1, \ldots, p, 1, \ldots, p, \ldots$, update δ_j^{t+1} by (S1.8) until convergence.
step 2.3: Update ν^{t+1} by (S1.7).
step 2.4: $t \leftarrow t + 1$ and repeat the two steps above until convergence.

$$\nu^{t+1} \leftarrow \nu^t + \mathbf{C}^\top \boldsymbol{\delta}^{t+1}. \tag{S1.7}$$

The iteration of $\boldsymbol{\delta}$ can also be further detailed as

$$\delta_j^{(t+1)} \leftarrow \frac{1}{\frac{\|\boldsymbol{Z}_j^{(0)}\|_2^2}{n_0} + \mu \|C_j\|_2^2} S_{\lambda_\delta} \left\{ \frac{1}{n_0} \boldsymbol{Z}_j^{(0)\top} \left(\widetilde{\boldsymbol{y}^{(0)}} - \sum_{i \neq j} \delta_i^{(t+1)} \boldsymbol{Z}_i^{(0)} \right) - \mu \left(\sum_{i \neq j} \delta_i^{(t+1)} C_j^\top C_i + C_j^T \nu^t \right) \right\}, \tag{S1.8}$$

where $\boldsymbol{Z}_j^{(0)}, j = 1, \ldots, p$ are columns of $\mathbf{Z}^{(0)}$, and $S_{\lambda_\delta}(x) = sgn(x)(|x| - \lambda_\delta)_+$. Combining (S1.6)-(S1.8) yields the following algorithm to solve problem (S1.2), that is, Step 2 in Algorithm 1 for Oracle Trans-sub-Coda-Lasso.

S2 Proof of Theorem 1

The proof of Theorem 1 following similar structure as those of Theorem 1 and Theorem 4 in Li et al. (2022), and the main challenge is to handle the zero-sum constraint. For simplicity, we do not use different notations for the constants appear in rates or probability expressions. For example, we always use c to describe any constants in probability expressions like $1 - \exp\{-c \log p\}$. We first list a few useful lemmas first, and these lemmas will be used for the proof of Theorem 1.

Lemma 1. *If Assumption 2 holds, then for each $k \in \mathcal{A} \cup \{0\}$, each row of $\mathbf{Z}^{(k)}$ is i.i.d. Sub-Gaussian distributed with mean zero and covariance matrix $\boldsymbol{\Sigma}^{(k)}$.*

Proof. By Assumption 2 and Lemma S1 in Zhang et al. (2025), the grouped clr-transformed compositional covariates $\tilde{\tilde{\mathbf{Z}}}_i$ are sub-Gaussian, the detailed proof is omitted here. In addition, assume that the non-compositional covariate vector $\mathbf{N}_i$ is also sub-Gaussian and independent of $\tilde{\tilde{\mathbf{Z}}}_i$. Then the concatenated covariate vector $\mathbf{Z}_i = (\tilde{\tilde{\mathbf{Z}}}_i^\top, \mathbf{N}_i^\top)$ is sub-Gaussian. □

Lemma 2. *Denote* $\hat{\mathbf{\Sigma}} = \sum_{k\in\{0\}\bigcup\mathcal{A}} \alpha_k \hat{\mathbf{\Sigma}}^{(k)}$, *where* $\hat{\mathbf{\Sigma}}^{(k)} = (\mathbf{Z}^{(k)})^\top \mathbf{Z}^{(k)}/n_k$. *If Assumption 2-3 hold, then for any* $\boldsymbol{u} \in \mathbb{R}^p$, *any constant* $\rho > 0$, *there exist strictly positive constants* ϕ_1 *and* ϕ_2, *such that*

$$\boldsymbol{u}^\top \hat{\mathbf{\Sigma}}^{(0)} \boldsymbol{u} \geq \phi_1 \|\boldsymbol{u}\|_2^2 - \rho \|\mathbf{C}^\top \boldsymbol{u}\|_2^2 - c_1 \sqrt{\frac{\log p}{n_0}} \|\boldsymbol{u}\|_1^2.$$

and

$$\boldsymbol{u}^\top \hat{\mathbf{\Sigma}} \boldsymbol{u} \geq \phi_2 \|\boldsymbol{u}\|_2^2 - \rho \|\mathbf{C}^\top \boldsymbol{u}\|_2^2 + c_1 \sqrt{\frac{\log p}{n_{\mathcal{A}} + n_0}} \|\boldsymbol{u}\|_1^2$$

with probability at least $1 - \exp(-c \log p)$.

Proof. First, since each row of $\mathbf{Z}^{(k)}$ is i.i.d. Sub-Gaussian distributed with mean zero and covariance matrix $\mathbf{\Sigma}^{(k)}$, we can get the convergence rate of $\hat{\mathbf{\Sigma}}^{(k)}$ as following,

$$Pr\left(\|\hat{\mathbf{\Sigma}}^{(k)} - \mathbf{\Sigma}^{(k)}\|_{\max} \geq c_1 \sqrt{\frac{\log p}{n_k}} \right) \leq \exp(-c_2 \log p).$$

Since

$$\begin{aligned} \boldsymbol{u}^\top \mathbf{\Sigma}^{(0)} \boldsymbol{u} - \boldsymbol{u}^\top \hat{\mathbf{\Sigma}}^{(0)} \boldsymbol{u} &= \boldsymbol{u}^\top (\mathbf{\Sigma}^{(0)} - \hat{\mathbf{\Sigma}}^{(0)}) u \\ &\leq \|\hat{\mathbf{\Sigma}}^{(0)} - \mathbf{\Sigma}^{(0)}\|_{\max} \|\boldsymbol{u}\|_1^2 \\ &\leq c_1 \sqrt{\frac{\log p}{n_0}} \|\boldsymbol{u}\|_1^2 \end{aligned}$$

with probability at least $1-\exp(-c_2\log p)$. Therefore, for any constant $\rho>0$,we have

$$\begin{aligned}\boldsymbol{u}^\top\hat{\boldsymbol{\Sigma}}^{(0)}\boldsymbol{u} &\geq \boldsymbol{u}^\top\boldsymbol{\Sigma}^{(0)}\boldsymbol{u} - c_1\sqrt{\frac{\log p}{n_0}}\|\boldsymbol{u}\|_1^2 \\ &= \boldsymbol{u}^\top(\boldsymbol{\Sigma}^{(0)}+\rho\tilde{\mathbf{C}}\tilde{\mathbf{C}}^\top)\boldsymbol{u} - \rho\boldsymbol{u}^\top\tilde{\mathbf{C}}\tilde{\mathbf{C}}^\top\boldsymbol{u} - c_1\sqrt{\frac{\log p}{n_0}}\|\boldsymbol{u}\|_1^2 \\ &= \boldsymbol{u}^\top(\boldsymbol{\Sigma}^{(0)}+\rho\tilde{\mathbf{C}}\tilde{\mathbf{C}}^\top)\boldsymbol{u} - \rho\|\mathbf{C}^\top\boldsymbol{u}\|_2^2 - c_1\sqrt{\frac{\log p}{n_0}}\|\boldsymbol{u}\|_1^2 \\ &\geq \lambda_{\min}(\boldsymbol{\Sigma}^{(0)}+\rho\tilde{\mathbf{C}}\tilde{\mathbf{C}}^\top)\|\boldsymbol{u}\|_2^2 - \rho\|\mathbf{C}^\top\boldsymbol{u}\|_2^2 - c_1\sqrt{\frac{\log p}{n_0}}\|\boldsymbol{u}\|_1^2.\end{aligned}$$

Since for any constant $\rho>0$, $\boldsymbol{\Sigma}^{(0)}+\rho\tilde{\mathbf{C}}\tilde{\mathbf{C}}^\top$ is nonsigular (Zhang et al., 2025), then the last inequality holds and we can get the result with $\phi_1=\lambda_{\min}(\boldsymbol{\Sigma}^{(0)}+\rho\tilde{\mathbf{C}}\tilde{\mathbf{C}}^\top)=\min(\rho,\lambda^+_{min}(\boldsymbol{\Sigma}^{(0)})$, where $\lambda^+_{min}(\boldsymbol{\Sigma}^{(0)})$ denotes the smallest positive eigenvalue of $\boldsymbol{\Sigma}^{(0)}$.

Then using reasoning similar to that used in the proof of Lemma 1.2 in He et al. (2022), we can obtain

$$Pr\left(\|\hat{\boldsymbol{\Sigma}}-\boldsymbol{\Sigma}\|_{\max}\geq c_1\sqrt{\frac{\log p}{n_{\mathcal{A}}+n_0}}\right)\leq\exp(-c_2\log p),$$

hence

$$\boldsymbol{u}^\top\boldsymbol{\Sigma}\boldsymbol{u}-\boldsymbol{u}^\top\hat{\boldsymbol{\Sigma}}\boldsymbol{u}\leq c_1\sqrt{\frac{\log p}{n_{\mathcal{A}}+n_0}}\|\boldsymbol{u}\|_1^2$$

with probability at least $1-\exp(-c_2\log p)$. Therefore, using the similar proof step as above, then we can get for any constant $\rho>0$,

$$\begin{aligned}\boldsymbol{u}^\top\hat{\boldsymbol{\Sigma}}\boldsymbol{u} &\geq \boldsymbol{u}^\top(\boldsymbol{\Sigma}+\rho\tilde{\mathbf{C}}\tilde{\mathbf{C}}^\top)\boldsymbol{u} - \rho\|\mathbf{C}^\top\boldsymbol{u}\|_2^2 - c_1\sqrt{\frac{\log p}{n_{\mathcal{A}}+n_0}}\|\boldsymbol{u}\|_1^2 \\ &\geq \lambda_{\min}(\boldsymbol{\Sigma}+\rho\tilde{\mathbf{C}}\tilde{\mathbf{C}}^\top)\|\boldsymbol{u}\|_2^2 - \rho\|\mathbf{C}^\top\boldsymbol{u}\|_2^2 - c_1\sqrt{\frac{\log p}{n_{\mathcal{A}}+n_0}}\|\boldsymbol{u}\|_1^2\end{aligned}$$

Notice that $\boldsymbol{\Sigma}+\rho\tilde{\mathbf{C}}\tilde{\mathbf{C}}^\top=\sum_{k\in\{0\}\bigcup\mathcal{A}}\alpha_k(\boldsymbol{\Sigma}^{(k)}+\rho\tilde{\mathbf{C}}\tilde{\mathbf{C}}^\top)$ and $\alpha_k>0$ for $k=0,\ldots,K$,

then the last Inequality holds. Then we can get the result with $\phi_2 = \lambda_{\min}(\mathbf{\Sigma} + \rho\tilde{\mathbf{C}}\tilde{\mathbf{C}}^\top) = \min(\rho, \lambda^+_{min}(\mathbf{\Sigma}))$.

□

In order to better elaborate the proof, we divide the whole process into three parts. Define $\hat{\boldsymbol{u}}^{\mathcal{A}} = \hat{\boldsymbol{w}}^{\mathcal{A}} - \boldsymbol{w}^{\mathcal{A}}$, $\hat{\boldsymbol{v}}^{\mathcal{A}} = \hat{\boldsymbol{\delta}}^{\mathcal{A}} - \boldsymbol{\delta}^{\mathcal{A}}$, $\tilde{\boldsymbol{r}} = \tilde{\mathbf{C}}^\top \sum_{k\in\{0\}\bigcup\mathcal{A}} \alpha_k \mathbf{\Sigma}^{(k)}\boldsymbol{\delta}^{(k)}$,

$$E_1 = \left\{ \left\| \frac{1}{n_{\mathcal{A}} + n_0} \sum_{k\in\{0\}\bigcup\mathcal{A}} (\mathbf{Z}^{(k)})^\top (\boldsymbol{y}^{(k)} - \mathbf{Z}^{(k)}\boldsymbol{w}^{\mathcal{A}}) + \tilde{\mathbf{C}}\tilde{\boldsymbol{r}} \right\|_\infty \leq \frac{\lambda_w}{2} \right\},$$

and

$$E_2 = \left\{ \frac{1}{n_0}\|(\mathbf{Z}^{(0)})^\top \boldsymbol{\varepsilon}^{(0)}\|_\infty \leq \frac{\lambda_\delta}{2}, \frac{1}{n_0}\|\mathbf{Z}^{(0)}\hat{\boldsymbol{u}}^{\mathcal{A}}\|_2^2 \leq 2\lambda_{max}(\mathbf{\Sigma}^{(0)})\|\hat{\boldsymbol{u}}^{\mathcal{A}}\|_2^2 \right\}.$$

In the first part, we will show that on the event of E_1 , the following inequality

$$(\hat{\boldsymbol{u}}^{\mathcal{A}})^\top \hat{\mathbf{\Sigma}}\hat{\boldsymbol{u}}^{\mathcal{A}} \vee \|\hat{\boldsymbol{u}}^{\mathcal{A}}\|_2^2 \lesssim \lambda_w(s\lambda_w + C_\Sigma^{\mathcal{A}} h) \quad \text{and} \quad \|\hat{\boldsymbol{u}}^{\mathcal{A}}\|_1 \lesssim s\lambda_w + C_\Sigma^{\mathcal{A}} h$$

with probability at least $1-\exp(-c\log p)$,where $s\log p/(n_{\mathcal{A}}+n_0) = o(1)$ and $C_\Sigma^{\mathcal{A}} h\sqrt{\log p/(n_{\mathcal{A}} + n_0)} = o(1)$. In the second part , we will show on the event of $E_1 \bigcap E_2$, the following inequality

$$\begin{aligned}
&(\hat{\boldsymbol{u}}^{\mathcal{A}})^\top \hat{\mathbf{\Sigma}}^{(0)}\hat{\boldsymbol{u}}^{\mathcal{A}} \vee \|\hat{\boldsymbol{v}}^{\mathcal{A}}\|_2^2 \lesssim \lambda_w(s\lambda_w + C_\Sigma^{\mathcal{A}} h),\\
&(\hat{\boldsymbol{v}}^{\mathcal{A}})^\top \hat{\mathbf{\Sigma}}^{(0)}\hat{\boldsymbol{v}}^{\mathcal{A}} \lesssim \lambda_\delta C_\Sigma^{\mathcal{A}} h,\\
&\|\hat{\boldsymbol{v}}^{\mathcal{A}}\|^2 \lesssim \lambda_\delta C_\Sigma^{\mathcal{A}} h \wedge (C_\Sigma^{\mathcal{A}} h)^2,
\end{aligned}$$

with probability at least $1-\exp(-c\log p)$. In the third part, we will show that $Pr(E_1) \geq 1-\exp(-c\log p)$ and $Pr(E_2) \geq 1-\exp(-c\log p)$ for some constant $c > 0$. Hence $Pr(E_1 \bigcap E_2) \to 1$ as $p \to \infty$.

Finally, we can get the desired result by utilizing the above result and the fact that

$$\begin{aligned}\|\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\|_2 &= \|(\hat{\boldsymbol{w}}^{\mathcal{A}}+\hat{\boldsymbol{\delta}}^{\mathcal{A}})-(\boldsymbol{w}^{\mathcal{A}}+\boldsymbol{\delta}^{\mathcal{A}})\|_2\\ &\leq \|\hat{\boldsymbol{u}}^{\mathcal{A}}\|_2+\|\hat{\boldsymbol{v}}^{\mathcal{A}}\|_2.\end{aligned}$$

Part 1. For $\hat{\boldsymbol{u}}^{\mathcal{A}}=\hat{\boldsymbol{w}}^{\mathcal{A}}-\boldsymbol{w}^{\mathcal{A}}$, we have

$$\frac{1}{2}(\hat{\boldsymbol{u}}^{\mathcal{A}})^\top\hat{\boldsymbol{\Sigma}}\hat{\boldsymbol{u}}^{\mathcal{A}}\leq\lambda_w\|\boldsymbol{w}^{\mathcal{A}}\|_1-\lambda_w\|\hat{\boldsymbol{w}}^{\mathcal{A}}\|_1+\frac{1}{n_{\mathcal{A}}+n_0}|(\hat{\boldsymbol{u}}^{\mathcal{A}})^\top\sum_{k\in\{0\}\bigcup\mathcal{A}}(\mathbf{Z}^{(k)})^\top(\boldsymbol{y}^{(k)}-\mathbf{Z}^{(k)}\boldsymbol{w}^{\mathcal{A}})|,$$

Note that $\boldsymbol{w}^{\mathcal{A}}=\boldsymbol{\Omega}_c\sum_{k\in\{0\}\bigcup\mathcal{A}}\alpha_k\boldsymbol{\Sigma}^{(k)}\boldsymbol{w}^{*(k)}$ satisfies that the linear constraint $\tilde{\mathbf{C}}^\top\boldsymbol{w}^{\mathcal{A}}=0$, then $\tilde{\mathbf{C}}^\top(\hat{\boldsymbol{w}}^{\mathcal{A}}-\boldsymbol{w}^{\mathcal{A}})=\tilde{\mathbf{C}}^\top\hat{\boldsymbol{u}}^{\mathcal{A}}=\mathbf{0}$. And $\tilde{\boldsymbol{r}}=\tilde{\mathbf{C}}^\top\sum_{k\in\{0\}\bigcup\mathcal{A}}\alpha_k\boldsymbol{\Sigma}^{(k)}\boldsymbol{\delta}^{(k)}$, hence

$$\begin{aligned}\frac{1}{2}(\hat{\boldsymbol{u}}^{\mathcal{A}})^\top\hat{\boldsymbol{\Sigma}}\hat{\boldsymbol{u}}^{\mathcal{A}} &\leq \lambda_w\|\boldsymbol{w}^{\mathcal{A}}\|_1-\lambda_w\|\hat{\boldsymbol{w}}^{\mathcal{A}}\|_1\\ &\quad+\frac{1}{n_{\mathcal{A}}+n_0}|(\hat{\boldsymbol{u}}^{\mathcal{A}})^\top\sum_{k\in\{0\}\bigcup\mathcal{A}}(\mathbf{Z}^{(k)})^\top(\boldsymbol{y}^{(k)}-\mathbf{Z}^{(k)}\boldsymbol{w}^{\mathcal{A}})|+\tilde{\boldsymbol{r}}^\top\tilde{\mathbf{C}}^\top\hat{\boldsymbol{u}}^{\mathcal{A}}\\ &\leq \lambda_w\|\boldsymbol{w}^{\mathcal{A}}\|_1-\lambda_w\|\hat{\boldsymbol{w}}^{\mathcal{A}}\|_1+\frac{\lambda_w}{2}\|\hat{\boldsymbol{u}}^{\mathcal{A}}\|_1\\ &\leq \frac{3}{2}\lambda_w\|\hat{\boldsymbol{u}}^{\mathcal{A}}_S\|_1+\lambda_w\|\boldsymbol{w}^{\mathcal{A}}_{S^c}\|_1-\lambda_w\|\hat{\boldsymbol{w}}^{\mathcal{A}}_{S^c}\|_1+\frac{1}{2}\lambda_w\|\hat{\boldsymbol{u}}^{\mathcal{A}}_{S^c}\|_1\end{aligned}$$

where last two steps are due to the event E_1 and the inequality $\|\boldsymbol{x}\|_1-\|\boldsymbol{y}\|_1\leq\|\boldsymbol{x}-\boldsymbol{y}\|_1$ for vectors $\boldsymbol{x},\boldsymbol{y}$, respectively. Using the fact that $\|\hat{\boldsymbol{u}}^{\mathcal{A}}_{S^c}\|_1\leq\|\boldsymbol{w}^{\mathcal{A}}_{S^c}\|_1+\|\hat{\boldsymbol{w}}^{\mathcal{A}}_{S^c}\|_1$, we have the oracle inequality

$$\frac{1}{2}(\hat{\boldsymbol{u}}^{\mathcal{A}})^\top\hat{\boldsymbol{\Sigma}}\hat{\boldsymbol{u}}^{\mathcal{A}}\leq\frac{3}{2}\lambda_w\|\hat{\boldsymbol{u}}^{\mathcal{A}}_S\|_1+2\lambda_w\|\boldsymbol{w}^{\mathcal{A}}_{S^c}\|_1-\frac{1}{2}\lambda_w\|\hat{\boldsymbol{u}}^{\mathcal{A}}_{S^c}\|_1. \tag{S2.1}$$

Let's consider the following two cases.

(i) If $\|\hat{\boldsymbol{u}}_S^{\mathcal{A}}\|_1 \geq \|\boldsymbol{w}_{S^c}^{\mathcal{A}}\|_1$, then (S2.1) can be further relaxed in two the following forms.

$$\frac{1}{2}(\hat{\boldsymbol{u}}^{\mathcal{A}})^\top \hat{\boldsymbol{\Sigma}} \hat{\boldsymbol{u}}^{\mathcal{A}} \leq \frac{7}{2}\lambda_w \|\hat{\boldsymbol{u}}_S^{\mathcal{A}}\|_1 - \frac{1}{2}\lambda_w \|\hat{\boldsymbol{u}}_{S^c}^{\mathcal{A}}\|_1. \tag{S2.2}$$

$$\frac{1}{2}(\hat{\boldsymbol{u}}^{\mathcal{A}})^\top \hat{\boldsymbol{\Sigma}} \hat{\boldsymbol{u}}^{\mathcal{A}} \leq \frac{7}{2}\lambda_w \|\hat{\boldsymbol{u}}_S^{\mathcal{A}}\|_1. \tag{S2.3}$$

Under the constraint of convex cone, $\|\hat{\boldsymbol{u}}^{\mathcal{A}}\|_1 \leq \|\hat{\boldsymbol{u}}_S^{\mathcal{A}}\|_1 + \|\hat{\boldsymbol{u}}_{S^c}^{\mathcal{A}}\|_1 \leq 8\|\hat{\boldsymbol{u}}_S^{\mathcal{A}}\|_1 \leq 8\sqrt{s}\|\hat{\boldsymbol{u}}^{\mathcal{A}}\|_2$. According Lemma 2, for some sufficiently large constant c, combining (S2.3), with probability at least $1 - \exp(-c \log p)$, we can further obtain

$$\phi_2 \|\hat{\boldsymbol{u}}^{\mathcal{A}}\|_2^2 - \rho \|\mathbf{C}^\top \hat{\boldsymbol{u}}^{\mathcal{A}}\|_2^2 - c_1 \sqrt{\frac{\log p}{n_{\mathcal{A}} + n_0}} \|\hat{\boldsymbol{u}}^{\mathcal{A}}\|_1^2 \lesssim \frac{1}{2}(\hat{\boldsymbol{u}}^{\mathcal{A}})^\top \hat{\boldsymbol{\Sigma}} \hat{\boldsymbol{u}}^{\mathcal{A}} \leq \frac{7}{2}\lambda_w \|\hat{\boldsymbol{u}}_S^{\mathcal{A}}\|_1 \leq \frac{7}{2}\sqrt{s}\lambda_w \|\hat{\boldsymbol{u}}_S^{\mathcal{A}}\|_2 \leq \frac{7}{2}\sqrt{s}\lambda_w \|\hat{\boldsymbol{u}}^{\mathcal{A}}\|_2.$$

Since $\mathbf{C}^\top(\hat{\boldsymbol{w}}^{\mathcal{A}} - \boldsymbol{w}^{\mathcal{A}}) = \mathbf{C}^\top \hat{\boldsymbol{u}}^{\mathcal{A}} = \mathbf{0}$ and for $\sqrt{\log p/(n_{\mathcal{A}} + n_0)} = o(1)$ derived from Assumption 3, we have

$$\|\hat{\boldsymbol{u}}^{\mathcal{A}}\|_2^2 \lesssim s\lambda_w^2, \quad \|\hat{\boldsymbol{u}}^{\mathcal{A}}\|_1 \lesssim s\lambda_w.$$

(ii) If $\|\hat{\boldsymbol{u}}_S^{\mathcal{A}}\|_1 \leq \|\boldsymbol{w}_{S^c}^{\mathcal{A}}\|_1$, then (S2.1) also can be further relaxed into the following two forms.

$$\frac{1}{2}(\hat{\boldsymbol{u}}^{\mathcal{A}})^\top \hat{\boldsymbol{\Sigma}} \hat{\boldsymbol{u}}^{\mathcal{A}} \leq \frac{7}{2}\lambda_w \|\boldsymbol{w}_{S^c}^{\mathcal{A}}\|_1 - \frac{1}{2}\lambda_w \|\hat{\boldsymbol{u}}_{S^c}^{\mathcal{A}}\|_1. \tag{S2.4}$$

$$\frac{1}{2}(\hat{\boldsymbol{u}}^{\mathcal{A}})^\top \hat{\boldsymbol{\Sigma}} \hat{\boldsymbol{u}}^{\mathcal{A}} \leq \frac{7}{2}\lambda_w \|\boldsymbol{w}_{S^c}^{\mathcal{A}}\|_1. \tag{S2.5}$$

Since

$$\|\hat{\boldsymbol{u}}^{\mathcal{A}}\|_1 = \|\hat{\boldsymbol{u}}_S^{\mathcal{A}}\|_1 + \|\hat{\boldsymbol{u}}_{S^c}^{\mathcal{A}}\|_1 \leq 8\|\boldsymbol{w}_{S^c}^{\mathcal{A}}\|_1 \leq 8\|\boldsymbol{\delta}_{S^c}^{\mathcal{A}}\|_1 \leq 8C_\Sigma h.$$

Hence, a direct upper bound on $\|\hat{u}^{\mathcal{A}}\|_2$ is

$$\|\hat{\boldsymbol{u}}^{\mathcal{A}}\|_2 \leq \|\hat{\boldsymbol{u}}^{\mathcal{A}}\|_1 \leq 8C_\Sigma h.$$

By the proof of Lemma 2, with probability at least $1-\exp(-c\log p)$, we have

$$\|\hat{\boldsymbol{u}}^{\mathcal{A}}\|_2^2-\rho\|\mathbf{C}^\top\hat{\boldsymbol{u}}^{\mathcal{A}}\|_2^2-c_1\sqrt{\frac{\log p}{n_{\mathcal{A}}+n_0}}\|\hat{\boldsymbol{u}}^{\mathcal{A}}\|_1^2\lesssim\frac{1}{2}(\hat{\boldsymbol{u}}^{\mathcal{A}})^\top\hat{\boldsymbol{\Sigma}}\hat{\boldsymbol{u}}^{\mathcal{A}}\leq\frac{7}{2}\lambda_w\|\boldsymbol{w}^{\mathcal{A}}_{S^c}\|_1\leq\frac{7}{2}\lambda_w\|\boldsymbol{\delta}^{\mathcal{A}}_{S^c}\|_1\leq\frac{7}{2}\lambda_w C_\Sigma h.$$

Noticing that $\mathbf{C}^\top\hat{\boldsymbol{u}}^{\mathcal{A}}=\mathbf{0}$ and $C_\Sigma h(\rho/p+\sqrt{\log p/(n_{\mathcal{A}}+n_0)})=o(1)$ as derived form Assumption 3, we can get

$$\|\hat{u}^{\mathcal{A}}\|_2^2\lesssim\lambda_w C_\Sigma h.$$

To summarize, in event E_1, we get the result

$$(\hat{\boldsymbol{u}}^{\mathcal{A}})^\top\hat{\boldsymbol{\Sigma}}\hat{\boldsymbol{u}}^{\mathcal{A}}\vee\|\hat{\boldsymbol{u}}^{\mathcal{A}}\|_2^2\lesssim\lambda_w(s\lambda_w+C^{\mathcal{A}}_\Sigma h)\quad\text{and}\quad\|\hat{\boldsymbol{u}}^{\mathcal{A}}\|_1\lesssim(s\lambda_w+C^{\mathcal{A}}_\Sigma h)$$

with probability at least $1-\exp(-c\log p)$.

Part 2. Denote that $\hat{\boldsymbol{v}}^{\mathcal{A}}=\hat{\boldsymbol{\delta}}^{\mathcal{A}}-\boldsymbol{\delta}^{\mathcal{A}}$, under $E_1\bigcap E_2$, we have the following oracle inequality, the same as the one form the proof of Theorem 1 and 4 in Li et al. (2022) and Theorem 4.1 inJin et al. (2024)

$$\begin{aligned}\frac{1}{2}(\hat{\boldsymbol{v}}^{\mathcal{A}})^\top\hat{\boldsymbol{\Sigma}}^{(0)}\hat{\boldsymbol{v}}^{\mathcal{A}}&\leq\lambda_\delta\|\boldsymbol{\delta}^{\mathcal{A}}\|_1-\lambda_\delta\|\hat{\boldsymbol{\delta}}^{\mathcal{A}}\|+\frac{1}{n_0}|<\mathbf{Z}^{(0)}\hat{\boldsymbol{v}}^{\mathcal{A}},\boldsymbol{\varepsilon}^{(0)}-\mathbf{Z}^{(0)}\hat{\boldsymbol{u}}^{\mathcal{A}}>|\\&\leq\lambda_\delta\|\boldsymbol{\delta}^{\mathcal{A}}\|_1-\lambda_\delta\|\hat{\boldsymbol{\delta}}^{\mathcal{A}}\|+\frac{\lambda_\delta}{2}\|\hat{\boldsymbol{v}}^{\mathcal{A}}\|_1+(\hat{\boldsymbol{u}}^{\mathcal{A}})^\top\hat{\boldsymbol{\Sigma}}^{(0)}\hat{\boldsymbol{u}}^{\mathcal{A}}+\frac{1}{4}(\hat{\boldsymbol{v}}^{\mathcal{A}})^\top\hat{\boldsymbol{\Sigma}}^{(0)}\hat{\boldsymbol{v}}^{\mathcal{A}},\end{aligned}$$

where the last step is due to the event $E_1\bigcap E_2$ and the inequality $|ab|\leq a^2+b^2/4$. Therefore, by noticing $\|\hat{\boldsymbol{v}}^{\mathcal{A}}\|_1=\|\hat{\boldsymbol{\delta}}^{\mathcal{A}}-\boldsymbol{\delta}^{\mathcal{A}}\|_1\geq\|\hat{\boldsymbol{\delta}}^{\mathcal{A}}\|_1-\|\boldsymbol{\delta}^{\mathcal{A}}\|_1$, we have the following oracle inequality

$$\frac{1}{4}(\hat{\boldsymbol{v}}^{\mathcal{A}})^\top\hat{\boldsymbol{\Sigma}}^{(0)}\hat{\boldsymbol{v}}^{\mathcal{A}}\leq2\lambda_\delta\|\boldsymbol{\delta}^{\mathcal{A}}\|_1-\frac{1}{2}\lambda_\delta\|\hat{\boldsymbol{v}}^{\mathcal{A}}\|_1+(\hat{\boldsymbol{u}}^{\mathcal{A}})^\top\hat{\boldsymbol{\Sigma}}^{(0)}\hat{\boldsymbol{u}}^{\mathcal{A}}.$$

If $(\hat{\boldsymbol{u}}^{\mathcal{A}})^\top \hat{\boldsymbol{\Sigma}}^{(0)} \hat{\boldsymbol{u}}^{\mathcal{A}} \leq \lambda_\delta \|\boldsymbol{\delta}^{\mathcal{A}}\|_1$, then

$$\begin{aligned} 0 \leq \frac{1}{4}(\hat{\boldsymbol{v}}^{\mathcal{A}})^\top \hat{\boldsymbol{\Sigma}}^{(0)} \hat{\boldsymbol{v}}^{\mathcal{A}} &\leq 3\lambda_\delta \|\boldsymbol{\delta}^{\mathcal{A}}\|_1 - \frac{\lambda_\delta}{2}\|\hat{\boldsymbol{v}}^{\mathcal{A}}\|_1 \\ &\leq 3\lambda_\delta \|\boldsymbol{\delta}^{\mathcal{A}}\|_1 \\ &\leq 3\lambda_\delta C_\Sigma^{\mathcal{A}} h, \end{aligned}$$

and $\|\hat{\boldsymbol{v}}^{\mathcal{A}}\|_1 \leq 6\|\boldsymbol{\delta}^{\mathcal{A}}\|_1 \leq 6C_\Sigma^{\mathcal{A}} h$ and

$$\|\hat{\boldsymbol{v}}^{\mathcal{A}}\|_2 \leq \|\hat{\boldsymbol{v}}^{\mathcal{A}}\|_1 \leq 6C_\Sigma^{\mathcal{A}} h.$$

Meanwhile, from Lemma 2, we have

$$\begin{aligned} \phi_1 \|\hat{\boldsymbol{v}}^{\mathcal{A}}\|_2^2 - \rho \|\mathbf{C}^\top \hat{\boldsymbol{v}}^{\mathcal{A}}\|_2^2 - c_1 \sqrt{\frac{\log p}{n_0}} \|\hat{\boldsymbol{v}}^{\mathcal{A}}\|_1^2 &\lesssim \frac{1}{4}(\hat{\boldsymbol{v}}^{\mathcal{A}})^\top \hat{\boldsymbol{\Sigma}}^{(0)} \hat{\boldsymbol{v}}^{\mathcal{A}} \\ &\leq 3\lambda_\delta \|\boldsymbol{\delta}^{\mathcal{A}}\|_1 \\ &\leq 3\lambda_\delta C_\Sigma^{\mathcal{A}} h, \end{aligned}$$

with probability at least $1 - \exp(-c \log p)$, and $\mathbf{C}^\top \hat{\boldsymbol{v}}^{\mathcal{A}} = \mathbf{0}$ and $C_\Sigma h(\rho/p + \sqrt{\log p / n_0}) = o(1)$ as derived form Assumption 3, we have

$$\|\hat{\boldsymbol{v}}^{\mathcal{A}}\|_2^2 \lesssim \lambda_\delta C_\Sigma h.$$

On the other hand, if $(\hat{\boldsymbol{u}}^{\mathcal{A}})^\top \hat{\boldsymbol{\Sigma}}^{(0)} \hat{\boldsymbol{u}}^{\mathcal{A}} \geq \lambda_\delta \|\boldsymbol{\delta}^{\mathcal{A}}\|_1$, we have

$$0 \leq \frac{1}{4}(\hat{\boldsymbol{v}}^{\mathcal{A}})^\top \hat{\boldsymbol{\Sigma}}^{(0)} \hat{\boldsymbol{v}}^{\mathcal{A}} \leq 3(\hat{\boldsymbol{u}}^{\mathcal{A}})^\top \hat{\boldsymbol{\Sigma}}^{(0)} \hat{\boldsymbol{u}}^{\mathcal{A}} - \frac{1}{2}\lambda_\delta \|\hat{\boldsymbol{v}}^{\mathcal{A}}\|_1.$$

then

$$\|\hat{\boldsymbol{v}}^{\mathcal{A}}\|_1 \leq \frac{6}{\lambda_\delta}(\hat{\boldsymbol{u}}^{\mathcal{A}})^\top \hat{\boldsymbol{\Sigma}}^{(0)} \hat{\boldsymbol{u}}^{\mathcal{A}} \text{and } (\hat{\boldsymbol{v}}^{\mathcal{A}})^\top \hat{\boldsymbol{\Sigma}}^{(0)} \hat{\boldsymbol{v}}^{\mathcal{A}} \leq 12(\hat{\boldsymbol{u}}^{\mathcal{A}})^\top \hat{\boldsymbol{\Sigma}}^{(0)} \hat{\boldsymbol{u}}^{\mathcal{A}}. \tag{S2.6}$$

Together with the second statement in E_2, (S2.6) and part 1, we arrive at

$$\|\hat{\boldsymbol{v}}^{\mathcal{A}}\|_1 \lesssim \frac{\lambda_w(s\lambda_w + C_\Sigma^{\mathcal{A}} h)}{\lambda_\delta}.$$

Using the second inequality in (S2.6) and Lemma 2, we have

$$\begin{aligned}
\phi_1 \|\hat{\boldsymbol{v}}^{\mathcal{A}}\|_2^2 - \rho \|\mathbf{C}^\top \hat{\boldsymbol{v}}^{\mathcal{A}}\|_2^2 - c_1 \sqrt{\frac{\log p}{n_0}} \|\hat{\boldsymbol{v}}^{\mathcal{A}}\|_1^2 &\lesssim \frac{1}{4}(\hat{\boldsymbol{v}}^{\mathcal{A}})^\top \hat{\boldsymbol{\Sigma}}^{(0)} \hat{\boldsymbol{v}}^{\mathcal{A}} \\
&\leq 3(\hat{\boldsymbol{u}}^{\mathcal{A}})^\top \hat{\boldsymbol{\Sigma}}^{(0)} \hat{\boldsymbol{u}}^{\mathcal{A}} \\
&\lesssim \|\hat{\boldsymbol{u}}^{\mathcal{A}}\|_2^2 \\
&\lesssim \lambda_w(s\lambda_w + C_\Sigma^{\mathcal{A}} h)
\end{aligned}$$

with probability at least $1 - \exp(-c \log p)$ and $\mathbf{C}^\top \hat{\boldsymbol{v}}^{\mathcal{A}} = \mathbf{0}$. Thus, we have

$$(\hat{\boldsymbol{u}}^{\mathcal{A}})^\top \hat{\boldsymbol{\Sigma}}^{(0)} \hat{\boldsymbol{u}}^{\mathcal{A}} \vee \|\hat{\boldsymbol{v}}^{\mathcal{A}}\|_2^2 \lesssim \lambda_w(s\lambda_w + C_\Sigma^{\mathcal{A}} h).$$

Part 3. Since $\boldsymbol{w}^{(k)} = \boldsymbol{\beta}^* - \boldsymbol{\delta}^{(k)}$ and $\boldsymbol{w}^{\mathcal{A}} = \boldsymbol{\beta}^* - \boldsymbol{\delta}^{\mathcal{A}}$, then

$$
\begin{aligned}
&\left\| \frac{1}{n_{\mathcal{A}} + n_0} \sum_{k \in \{0\} \bigcup \mathcal{A}} (\mathbf{Z}^{(k)})^\top (\boldsymbol{y}^{(k)} - \mathbf{Z}^{(k)} \boldsymbol{w}^{\mathcal{A}}) + \tilde{\mathbf{C}} \tilde{\boldsymbol{r}} \right\|_\infty \\
&= \left\| \frac{1}{n_{\mathcal{A}} + n_0} \sum_{k \in \{0\} \bigcup \mathcal{A}} (\mathbf{Z}^{(k)})^\top \boldsymbol{\varepsilon}^{(k)} + \sum_{k \in \{0\} \bigcup \mathcal{A}} \alpha_k \hat{\boldsymbol{\Sigma}}^{(k)} \boldsymbol{w}^{(k)} - \hat{\boldsymbol{\Sigma}} \boldsymbol{w}^{\mathcal{A}} + \tilde{\mathbf{C}} \tilde{\boldsymbol{r}} \right\|_\infty \\
&= \left\| \frac{1}{n_{\mathcal{A}} + n_0} \sum_{k \in \{0\} \bigcup \mathcal{A}} (\mathbf{Z}^{(k)})^\top \boldsymbol{\varepsilon}^{(k)} - \sum_{k \in \{0\} \bigcup \mathcal{A}} \alpha_k \hat{\boldsymbol{\Sigma}}^{(k)} \boldsymbol{\delta}^{(k)} - \sum_{k \in \{0\} \bigcup \mathcal{A}} \alpha_k \boldsymbol{\Sigma}^{(k)} \boldsymbol{\delta}^{(k)} \right. \\
&\left. + \sum_{k \in \{0\} \bigcup \mathcal{A}} \alpha_k \boldsymbol{\Sigma}^{(k)} \boldsymbol{\delta}^{(k)} + \hat{\boldsymbol{\Sigma}} \boldsymbol{\delta}^{\mathcal{A}} + \tilde{\mathbf{C}} \tilde{\boldsymbol{r}} \right\|_\infty \\
&\leq \left\| \frac{1}{n_{\mathcal{A}} + n_0} \sum_{k \in \{0\} \bigcup \mathcal{A}} (\mathbf{Z}^{(k)})^\top \varepsilon^{(k)} - \sum_{k \in \{0\} \bigcup \mathcal{A}} \alpha_k (\hat{\boldsymbol{\Sigma}}^{(k)} - \boldsymbol{\Sigma}^{(k)}) \boldsymbol{\delta}^{(k)} \right\|_\infty \\
&+ \left\| - \sum_{k \in \{0\} \bigcup \mathcal{A}} \alpha_k \boldsymbol{\Sigma}^{(k)} \boldsymbol{\delta}^{(k)} + \hat{\boldsymbol{\Sigma}} \boldsymbol{\delta}^{\mathcal{A}} + \tilde{\mathbf{C}} \tilde{\boldsymbol{r}} \right\|_\infty .
\end{aligned}
$$

Since $\boldsymbol{\delta}^{\mathcal{A}} = \boldsymbol{\Omega} \sum_{k \in \{0\} \bigcup \mathcal{A}} \alpha_k \boldsymbol{\Sigma}^{(k)} \boldsymbol{\delta}^{(k)}$, then

$$
\begin{aligned}
&\left\| - \sum_{k \in \{0\} \bigcup \mathcal{A}} \alpha_k \boldsymbol{\Sigma}^{(k)} \boldsymbol{\delta}^{(k)} + \hat{\boldsymbol{\Sigma}} \boldsymbol{\delta}^{\mathcal{A}} + \tilde{\mathbf{C}} \tilde{\boldsymbol{r}} \right\|_\infty \\
&= \left\| - \sum_{k \in \{0\} \bigcup \mathcal{A}} \alpha_k \boldsymbol{\Sigma}^{(k)} \boldsymbol{\delta}^{(k)} + \boldsymbol{\Sigma} \boldsymbol{\delta}^{\mathcal{A}} - \boldsymbol{\Sigma} \boldsymbol{\delta}^{\mathcal{A}} + \hat{\boldsymbol{\Sigma}} \boldsymbol{\delta}^{\mathcal{A}} + \tilde{\mathbf{C}} \tilde{\boldsymbol{r}} \right\|_\infty \\
&= \left\| - \sum_{k \in \{0\} \bigcup \mathcal{A}} \alpha_k \boldsymbol{\Sigma}^{(k)} \boldsymbol{\delta}^{(k)} + \boldsymbol{\Sigma} \boldsymbol{\Omega}_c \sum_{k \in \{0\} \bigcup \mathcal{A}} \alpha_k \boldsymbol{\Sigma}^{(k)} \boldsymbol{\delta}^{(k)} - (\boldsymbol{\Sigma} - \hat{\boldsymbol{\Sigma}}) \boldsymbol{\delta}^{\mathcal{A}} + \tilde{\mathbf{C}} \tilde{\boldsymbol{r}} \right\|_\infty \\
&= \left\| - \sum_{k \in \{0\} \bigcup \mathcal{A}} \alpha_k \boldsymbol{\Sigma}^{(k)} \boldsymbol{\delta}^{(k)} + (\mathbf{I}_p - \tilde{\mathbf{C}} \tilde{\mathbf{C}}^\top) \sum_{k \in \{0\} \bigcup \mathcal{A}} \alpha_k \boldsymbol{\Sigma}^{(k)} \boldsymbol{\delta}^{(k)} - (\boldsymbol{\Sigma} - \hat{\boldsymbol{\Sigma}}) \boldsymbol{\delta}^{\mathcal{A}} + \tilde{\mathbf{C}} \tilde{\boldsymbol{r}} \right\|_\infty \\
&= \left\| (\boldsymbol{\Sigma} - \hat{\boldsymbol{\Sigma}}) \boldsymbol{\delta}^{\mathcal{A}} \right\|_\infty .
\end{aligned}
$$

By Assumption 1 and 2, and Lemma 1 , $\boldsymbol{z}_i^{(k)\top} \boldsymbol{\varepsilon}_i^{(k)}$ is sub-exponential, and hence due to

similar analysis of Lemma 1 in Li et al. (2022), for $\log p = o(n_{\mathcal{A}} + n_0)$, with probability at least $1 - \exp\{-c \log p\}$, we can get

$$\left\| \frac{1}{n_{\mathcal{A}} + n_0} \sum_{k \in \{0\} \bigcup \mathcal{A}} (\mathbf{Z}^{(k)})^\top \boldsymbol{\varepsilon}^{(k)} \right\|_\infty \leq c_1' \sqrt{\frac{\log p}{n_{\mathcal{A}} + n_0}},$$
$$\left\| \sum_{k \in \{0\} \bigcup \mathcal{A}} \alpha_k (\hat{\boldsymbol{\Sigma}}^{(k)} - \boldsymbol{\Sigma}^{(k)}) \boldsymbol{\delta}^{(k)} \right\|_\infty \leq c_1'' \sqrt{\frac{\log p}{n_{\mathcal{A}} + n_0}},$$
$$\left\| \boldsymbol{\Sigma} \boldsymbol{\delta}^{\mathcal{A}} - \hat{\boldsymbol{\Sigma}} \boldsymbol{\delta}^{\mathcal{A}} \right\|_\infty \leq c_1''' \sqrt{\frac{\log p}{n_{\mathcal{A}} + n_0}}.$$

Hence, take $\lambda_w = 2max\{c_1', c_1'', c_1'''\} \sqrt{\frac{\log p}{n_{\mathcal{A}} + n_0}}$, with probability at least $1 - \exp\{-c \log p\}$, such that

$$\left\| \frac{1}{n_{\mathcal{A}} + n_0} \sum_{k \in \{0\} \bigcup \mathcal{A}} (\mathbf{Z}^{(k)})^\top (\boldsymbol{y}^{(k)} - \mathbf{Z}^{(k)} \boldsymbol{w}^{\mathcal{A}}) + \tilde{\mathbf{C}} \tilde{\boldsymbol{r}} \right\|_\infty \leq \frac{\lambda_w}{2}.$$

we can get $P(E_1) \geq 1 - \exp\{-c \log p\}$. By the sub-Gaussian property of $\boldsymbol{z}_i^{(k)\top} \boldsymbol{\varepsilon}_i^{(k)}$, we have

$$Pr\left(\frac{1}{n_0} \|(\mathbf{Z}^{(0)})^\top \boldsymbol{\varepsilon}^{(0)}\|_\infty \geq c \sqrt{\frac{\log p}{n_0}} \right) \leq \exp\{-c_1 \log p\}$$

for some large enough constant c_1. As $\mathbf{Z}^{(0)}$ is independent of $\hat{\boldsymbol{u}}^{\mathcal{A}}$, hence we have

$$Pr\left(\frac{1}{n_0} \|\mathbf{Z}^{(0)} \hat{\boldsymbol{u}}^{\mathcal{A}}\|_2^2 \leq 2 \lambda_{max}(\boldsymbol{\Sigma}^{(0)}) \|\hat{\boldsymbol{u}}^{\mathcal{A}}\|_2^2 \right) \leq \exp\{-c_1 n_0\}.$$

Hence we can get $Pr(E_2) \geq 1 - \exp\{-c \log p\}$ and further leads to $Pr(E_1 \bigcap E_2) \to 1$ as

$p \to \infty$. Finally, we can get the desired result by utilizing the above result and the fact that

$$\begin{aligned}\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\|_2 &= \|(\hat{\boldsymbol{w}}^{\mathcal{A}} + \hat{\boldsymbol{\delta}}^{\mathcal{A}}) - (\boldsymbol{w}^{\mathcal{A}} + \boldsymbol{\delta}^{\mathcal{A}})\|_2 \\ &\leq \|\hat{\boldsymbol{u}}^{\mathcal{A}}\|_2 + \|\hat{\boldsymbol{v}}^{\mathcal{A}}\|_2.\end{aligned}$$

S3 Proof of Theorem 2

The proof of Theorem 2 follows the same outline as the proof of Theorem 4 inTian and Feng (2023) and Theorem 4.2 in Jin et al. (2024). We list a few useful lemmas first, and these lemmas will be used for the proof of Theorem 2. Recall that the average of squared prediction error on the target testing data $\mathcal{I}^c$ is

$$\hat{Q}(\boldsymbol{\beta}) = \frac{2}{n_0} \sum_{i \in \mathcal{I}^c} (\boldsymbol{y}_i^{(0)} - \boldsymbol{z}_i^{(0)} \boldsymbol{\beta})^2.$$

The corresponding population version of $\hat{Q}(\boldsymbol{\beta})$ is

$$Q(\boldsymbol{\beta}) = \mathbb{E}\{(\boldsymbol{y}^{(0)} - \mathbf{Z}^{(0)}\boldsymbol{\beta})^\top \boldsymbol{y}^{(0)} - \mathbf{Z}^{(0)}\boldsymbol{\beta})\},$$

where the expectation is taken with respect to the target distribution. Consider the following population version of $\hat{\boldsymbol{w}}^{(0,k)}$ which define in step 3 in Algorithm 4(Trans-Coda-Lasso) by running Step 1 in Algorithm 1 with $(\mathbf{Z}_{\mathcal{I}}^{(0)}, \boldsymbol{y}_{\mathcal{I}}^{(0)}) \bigcup (\mathbf{Z}^{(k)}, \boldsymbol{y}^{(k)})$, which is the solution $\boldsymbol{w}^{(0,k)}$ of the following constrained optimization problem

$$\min_{w} \sum_{j \in \{0,k\}} \alpha_j \mathbb{E}\left[(\boldsymbol{y}^{(j)} - \mathbf{Z}^{(j)}\boldsymbol{w})^\top (\boldsymbol{y}^{(j)} - \mathbf{Z}^{(j)}\boldsymbol{w})\right], \quad \text{subject to } \mathbf{C}^\top \boldsymbol{w} = \mathbf{0},$$

where $\alpha_0 = \frac{n_0/2}{n_k+n_0/2}$ and $\alpha_k = \frac{n_k}{n_k+n_0/2}$ for $k = 1, \ldots, K$ and $\boldsymbol{w}^{(0,k)}$ can be expressed as a linear transformation of the true parameter $\boldsymbol{w}^{*(k)}$ and $\boldsymbol{\beta}^*$, that is,

$$\boldsymbol{w}^{(0,k)} = \boldsymbol{\Omega}_c^{(0,k)} \sum_{j\in\{0,k\}} \alpha_j \boldsymbol{\Sigma}^{(j)} \boldsymbol{w}^{*(j)},$$

where $\boldsymbol{\Omega}_c^{(0,k)}$ is the Moore-Penrose inverse of $\boldsymbol{\Sigma}^{(0,k)} = \sum_{j=0,k} \alpha_j \boldsymbol{\Sigma}^{(j)}$ and $\boldsymbol{w}^{*(0)} = \boldsymbol{\beta}^*$. Which indicates that $\boldsymbol{w}^{(0,k)}$ is a linear transform of $\boldsymbol{\beta}^*$ and $\boldsymbol{w}^{*(k)}$.

Lemma 3. *(Lemma S1.4 in Jin et al. (2024)) Under Assumption 2, we can have*

$$\sup_{k\in\mathcal{A}} Q(\boldsymbol{w}^{(0,k)}) - Q(\boldsymbol{\beta}^*) = O(\sup_{k\in\mathcal{A}} \|\boldsymbol{w}^{(0,k)} - \boldsymbol{\beta}^*\|_2^2) = O(h^2)$$

Proof of Lemma 3. According to the fact that

$$Q'(\boldsymbol{\beta}^*) = -\mathbb{E}\{2\mathbf{Z}^{(0)\top}(\boldsymbol{y}^{(0)} - \mathbf{Z}^{(0)}\boldsymbol{\beta}^*)\} = 0,$$

for any give $t \in (0, 1)$, we have

$$\begin{aligned} Q'(\boldsymbol{\beta}^* + t(\boldsymbol{w}^{(0,k)} - \boldsymbol{\beta}^*)) &= -\mathbb{E}\{2\mathbf{Z}^{(0)\top}(\boldsymbol{y}^{(0)} - \mathbf{Z}^{(0)}(\boldsymbol{\beta}^* + t(\boldsymbol{w}^{(0,k)} - \boldsymbol{\beta}^*)))\} \\ &= 2t\mathbb{E}(\mathbf{Z}^{(0)\top}\mathbf{Z}^{(0)})(\boldsymbol{w}^{(0,k)} - \boldsymbol{\beta}^*) \\ &= 2t\boldsymbol{\Sigma}^{(0)}(\boldsymbol{w}^{(0,k)} - \boldsymbol{\beta}^*). \end{aligned}$$

Then we can get

$$\begin{aligned} Q(\boldsymbol{w}^{(0,k)}) - Q(\boldsymbol{\beta}^*) &= 2t(\boldsymbol{w}^{(0,k)} - \boldsymbol{\beta}^*)^\top \boldsymbol{\Sigma}^{(0)} (\boldsymbol{w}^{(0,k)} - \boldsymbol{\beta}^*) \\ &= 2t(\boldsymbol{w}^{(0,k)} - \boldsymbol{\beta}^*)^\top (\boldsymbol{\Sigma}^{(0)} + \rho \tilde{\mathbf{C}} \tilde{\mathbf{C}}^\top)(\boldsymbol{w}^{(0,k)} - \boldsymbol{\beta}^*) \\ &= O(\|\boldsymbol{w}^{(0,k)} - \boldsymbol{\beta}^*\|_2^2) \end{aligned}$$

where the last second step is due to $\tilde{\mathbf{C}}^\top \boldsymbol{w}^{(0,k)} = \mathbf{0}$ and $\tilde{\mathbf{C}}^\top \boldsymbol{\beta}^* = \mathbf{0}$. The result can then be reached by using $\ell_1 - \ell_2$ norm inequality and $\boldsymbol{w}^{(0,k)}$ is a linear transform of $\boldsymbol{\beta}^*$ and $\boldsymbol{w}^{*(k)}$. □

Lemma 4. *Recall $s^* = s \vee s'$ and $h^* = C_\Sigma^{\mathcal{A}} h \vee h'$. Denote $\underline{n} = \min_{k=1}^K n_k$, there exist η such that $\eta^2 = o(n_0)$,*

$$\kappa_1 = \sqrt{\frac{\log p}{n_0}} \left(s^* \sqrt{\frac{\log p}{n_0 + \underline{n}}} + h^* \right) + \sqrt{s^*} \left(1 + \sqrt{\frac{\eta^2}{n_0}} \right) \left(s^* \frac{\log p}{n_0 + \underline{n}} + h * \sqrt{\frac{\log p}{n_0 + \underline{n}}} \right)^{\frac{1}{2}}$$

and

$$\kappa_2 = s^* \sqrt{\frac{\log p}{n_0}}.$$

Under assumptions 1-5 we have

$$\begin{aligned} & Pr\left(\sup_{k \neq 0} |\hat{Q}(\hat{\boldsymbol{w}}^{(0,k)}) - \hat{Q}(\boldsymbol{w}^{(0,k)})| \lesssim \kappa_1 \right) \\ & \geq 1 - \tilde{c}_1 \exp\{-c_1 \eta^2\} - \tilde{c}_2 \exp\{-c_2 \log p\} \end{aligned}$$

and

$$\begin{aligned} & Pr\left(|\hat{Q}(\hat{\boldsymbol{\beta}}_{\mathcal{I}}^{(0)}) - \hat{Q}(\boldsymbol{\beta}^*)| \lesssim \kappa_2 \right) \\ & \geq 1 - \tilde{c}_2 \exp(-c_2 \log p) \end{aligned}$$

with positive constants c_1, $\tilde{c}_1$,$\tilde{c}_2$ *and* $\tilde{c}_2$.

Proof of Lemma 4. Similar to the proofs of Lemma 7 in Tian and Feng (2023) for the linear model and Lemma S1.5 in Jin et al. (2024), we have for each $k \neq 0$,

$$
\begin{aligned}
&|\hat{Q}(\hat{\boldsymbol{w}}^{(0,k)}) - \hat{Q}(\boldsymbol{w}^{(0,k)})| \\
&\leq \frac{2}{n_0}|\sum_{i\in\mathcal{I}^c} \varepsilon_i^{(0)} \boldsymbol{z}_i^{(0)\top}(\hat{\boldsymbol{w}}^{(0,k)} - \boldsymbol{w}^{(0,k)})| + \frac{2}{n_0}|\sum_{i\in\mathcal{I}^c} \boldsymbol{z}_i^{(0)\top}\boldsymbol{\beta}^*\boldsymbol{z}_i^{(0)\top}(\hat{\boldsymbol{w}}^{(0,k)} - \boldsymbol{w}^{(0,k)})| \\
&+ \frac{1}{n_0}|\sum_{i\in\mathcal{I}^c} \boldsymbol{z}_i^{(0)\top}(\hat{\boldsymbol{w}}^{(0,k)} + \boldsymbol{w}^{(0,k)})(\hat{\boldsymbol{w}}^{(0,k)} - \boldsymbol{w}^{(0,k)})\boldsymbol{z}_i^{(0)}| \qquad \text{(S3.1)}
\end{aligned}
$$

For the first term on the right hand side of (S3.1),

$$
\begin{aligned}
&\frac{2}{n_0}|\sum_{i\in\mathcal{I}^c} \varepsilon_i^{(0)} \boldsymbol{z}_i^{(0)\top}(\hat{\boldsymbol{w}}^{(0,k)} - \boldsymbol{w}^{(0,k)})| \\
&\leq \|\frac{2}{n_0}\sum_{i\in\mathcal{I}^c} \varepsilon_i^{(0)} \boldsymbol{z}_i^{(0)\top}\|_\infty \|\hat{\boldsymbol{w}}^{(0,k)} - \boldsymbol{w}^{(0,k)}\|_1 \\
&\lesssim \sqrt{\frac{\log p}{n_0}}\|\hat{\boldsymbol{w}}^{(0,k)} - \boldsymbol{w}^{(0,k)}\|_1 \qquad \text{(S3.2)}
\end{aligned}
$$

with probability at least $1-\tilde{c}_2\exp\{-c_2\log p\}$, where the last inequality is due to Assumption 1 and the sub-exponential property of $z_i^{(0)}\varepsilon_i^{(0)}$. For the second and third terms on the right hand side of (S3.1), we consider that $z_i^{(0)}$s are independent and sub-Gaussian distribution, we can assert that the centralized version of

$$
\frac{2}{n_0}\sum_{i\in\mathcal{I}^c} \boldsymbol{z}_i^{(0)\top}\boldsymbol{\beta}^*\boldsymbol{z}_i^{(0)\top}(\hat{\boldsymbol{w}}^{(0,k)} - \boldsymbol{w}^{(0,k)})
$$

and

$$
\frac{1}{n_0}\sum_{i\in\mathcal{I}^c} \boldsymbol{z}_i^{(0)\top}(\hat{\boldsymbol{w}}^{(0,k)} + \boldsymbol{w}^{(0,k)})(\hat{\boldsymbol{w}}^{(0,k)} - \boldsymbol{w}^{(0,k)})^\top\boldsymbol{z}_i^{(0)}
$$

are both sub-exponential with mean at most multiplicative to $\|\boldsymbol{w}^{(0,k)}\|_2\|\hat{\boldsymbol{w}}^{(0,k)}-\boldsymbol{w}^{(0,k)}\|_2$ according to Vershynin (2018), then based on the same analysis in the proof of Lemma S1.5 in Jin et al. (2024) and Lemma 7 in Tian and Feng (2023), by tail bounds and union bounds, we can get

$$\sup_{k\neq 0}\max\left(\frac{2}{n_0}|\sum_{i\in\mathcal{I}^c}\boldsymbol{z}_i^{(0)\top}\boldsymbol{\beta}^*\boldsymbol{z}_i^{(0)\top}(\hat{\boldsymbol{w}}^{(0,k)}-\boldsymbol{w}^{(0,k)})|,\frac{1}{n_0}\sum_{i\in\mathcal{I}^c}\boldsymbol{z}_i^{(0)\top}(\hat{\boldsymbol{w}}^{(0,k)}+\boldsymbol{w}^{(0,k)})(\hat{\boldsymbol{w}}^{(0,k)}-\boldsymbol{w}^{(0,k)})^\top\boldsymbol{z}_i^{(0)}\right)$$
$$\lesssim\left(1+\sqrt{\frac{\eta^2}{n_0}}\right)\sup_{k\neq 0}\|\boldsymbol{w}^{(0,k)}\|_2\|\hat{\boldsymbol{w}}^{(0,k)}-\boldsymbol{w}^{(0,k)}\|_2$$

with probability at least $1-\tilde{c}_1\exp\{-c_1\eta^2\}$. Hence, we have

$$\begin{aligned}&\sup_{k\neq 0}|\hat{Q}(\hat{\boldsymbol{w}}^{(0,k)})-\hat{Q}(\boldsymbol{w}^{(0,k)})|\\&\lesssim\sqrt{\frac{\log p}{n_0}}\|\hat{\boldsymbol{w}}^{(0,k)}-\boldsymbol{w}^{(0,k)}\|_1+\left(1+\sqrt{\frac{\eta^2}{n_0}}\right)\sup_{k\neq 0}\|\boldsymbol{w}^{(0,k)}\|_2\|\hat{\boldsymbol{w}}^{(0,k)}-\boldsymbol{w}^{(0,k)}\|_2\end{aligned}$$

with probability at least $1-\tilde{c}_1\exp\{-c_1\eta^2\}-\tilde{c}_2\exp\{-c_2\log p\}$. According to the first part in the proof of Theorem 1, with $n'=\min_{k\in\mathcal{A}}n_k$, we have

$$\begin{aligned}&Pr\left(\sup_{k\in\mathcal{A}}\|\hat{\boldsymbol{w}}^{(0,k)}-\boldsymbol{w}^{(0,k)}\|_2^2\lesssim s\frac{\log p}{n_0+n'}+C_\Sigma^{\mathcal{A}}h\sqrt{\frac{\log p}{n_0+n'}}\right)\\&\geq 1-\exp\{-c\log p\}\end{aligned}$$

and

$$\begin{aligned}&Pr\left(\sup_{k\in\mathcal{A}}\|\hat{\boldsymbol{w}}^{(0,k)}-\boldsymbol{w}^{(0,k)}\|_1\lesssim s\sqrt{\frac{\log p}{n_0+n'}}+C_\Sigma^{\mathcal{A}}h\right)\\&\geq 1-\exp\{-c\log p\}\end{aligned}$$

For all $k \in \mathcal{A}^c$, combining Assumption 4 with the similar analysis of the first part in the proof of Theorem 1, we have

$$Pr\left(\sup_{k\in\mathcal{A}^c} \|\hat{\boldsymbol{w}}^{(0,k)} - \boldsymbol{w}^{(0,k)}\|_2^2 \lesssim s'\frac{\log p}{n_0+n''} + h'\sqrt{\frac{\log p}{n_0+n''}}\right)$$
$$\geq 1 - \exp\{-c\log p\}$$

and

$$Pr\left(\sup_{k\in\mathcal{A}^c} \|\hat{\boldsymbol{w}}^{(0,k)} - \boldsymbol{w}^{(0,k)}\|_1 \lesssim s'\sqrt{\frac{\log p}{n_0+n''}} + h'\right)$$
$$\geq 1 - \exp\{-c\log p\}$$

where $n'' = \min_{k\in\mathcal{A}^c} n_k$. Therefore, the desired result holds true by combining the above two inequality with $\sup_{k\neq 0} \|w^{(0,k)}\| \lesssim \sqrt{s \vee s'}$ from Assumption 4. The result for the target estimator holds true similarly.

□

Lemma 5. *Under Assumption 1-5 we have*

$$Pr\left(\sup_{k\neq 0} |\hat{Q}(\boldsymbol{w}^{(0,k)}) - Q(\boldsymbol{w}^{(0,k)})| \vee |\hat{Q}(\boldsymbol{\beta}^*) - Q(\boldsymbol{\beta}^*)| \lesssim \kappa_2\right)$$
$$\geq 1 - \tilde{c}_1\exp\{-c_1\eta^2\} - \tilde{c}_2\exp\{-c_2\log p\}$$

with some positive constants c_1 ,$\tilde{c}_1$, c_2 *and* $\tilde{c}_2$.

Proof of Lemma 5. Similar to the proof of Lemma 8 inTian and Feng (2023), we can get

$$\begin{aligned}
&|\hat{Q}(\boldsymbol{w}^{(0,k)}) - Q(\boldsymbol{w}^{(0,k)})| \\
&\leq \frac{2}{n_0}|\sum_{i\in\mathcal{I}^c}\varepsilon_i^{(0)}\boldsymbol{z}_i^{(0)\top}\boldsymbol{\beta}^*| + \frac{2}{n_0}|\sum_{i\in\mathcal{I}^c}(\boldsymbol{z}_i^{(0)\top}\boldsymbol{w}^{(0,k)})^2 - E(\boldsymbol{z}_i^{(0)\top}\boldsymbol{w}^{(0,k)})^2| \\
&+ \frac{2}{n_0}|\sum_{i\in\mathcal{I}^c}\boldsymbol{z}_i^{(0)\top}\boldsymbol{\beta}^*\boldsymbol{z}_i^{(0)\top}\boldsymbol{w}^{(0,k)} - E(\boldsymbol{z}_i^{(0)\top}\boldsymbol{\beta}^*\boldsymbol{z}_i^{(0)\top}\boldsymbol{w}^{(0,k)})| \qquad \text{(S3.3)}
\end{aligned}$$

For the first term of the right hand side of (S3.3), we have

$$\frac{2}{n_0}|\sum_{i\in\mathcal{I}^c}\varepsilon_i^{(0)}\boldsymbol{z}_i^{(0)\top}\boldsymbol{\beta}^*| \leq \|\frac{2}{n_0}\sum_{i\in\mathcal{I}^c}\varepsilon_i^{(0)}\boldsymbol{z}_i^{(0)\top}\|_\infty\|\boldsymbol{\beta}^*\|_1 \lesssim \sqrt{\frac{\log p}{n_0}}\|\boldsymbol{\beta}^*\|_1 \qquad \text{(S3.4)}$$

with probability at least $1-\tilde{c}_2\exp\{-c_2\log p\}$. On the other hand ,

$$\frac{1}{n_0}\sum_{i\in\mathcal{I}^c}(\boldsymbol{z}_i^{(0)\top}\boldsymbol{w}^{(0,k)})^2$$

and

$$\frac{1}{n_0}\sum_{i\in\mathcal{I}^c}\boldsymbol{z}_i^{(0)\top}\boldsymbol{\beta}^*\boldsymbol{z}_i^{(0)\top}\boldsymbol{w}^{(0,k)}$$

are sub-exponential with their means at most multiplicative to $\|\boldsymbol{w}^{(0,k)}\|_2^2$ and $\|\boldsymbol{\beta}^*\|_2\|\boldsymbol{w}^{(0,k)}\|_2$ respectively, according to Vershynin (2018). Hence we use Bernstein's inequality to get

$$\sup_{k\neq 0}\frac{1}{n_0}|\sum_{i\in\mathcal{I}^c}(\boldsymbol{z}_i^{(0)\top}\boldsymbol{w}^{(0,k)})^2 - E(\boldsymbol{z}_i^{(0)\top}\boldsymbol{w}^{(0,k)})^2| \lesssim \sqrt{\frac{\eta^2}{n_0}}\sup_{k\neq 0}\|\boldsymbol{w}^{(0,k)}\|_2^2 \qquad \text{(S3.5)}$$

and

$$\sup_{k\neq 0}\frac{1}{n_0}|\sum_{i\in\mathcal{I}^c}\boldsymbol{z}_i^{(0)\top}\boldsymbol{\beta}^*\boldsymbol{z}_i^{(0)\top}\boldsymbol{w}^{(0,k)} - E(\boldsymbol{z}_i^{(0)\top}\boldsymbol{\beta}^*\boldsymbol{z}_i^{(0)\top}\boldsymbol{w}^{(0,k)})| \lesssim \sqrt{\frac{\eta^2}{n_0}}\|\boldsymbol{\beta}^*\|_2\sup_{k\neq 0}\|\boldsymbol{w}^{(0,k)}\|_2 \qquad \text{(S3.6)}$$

with probability at least $1-\tilde{c}_1\exp\{-c_1\eta^2\}$. Combining (S3.4),(S3.5) and (S3.6) and noticing that $\|\boldsymbol{\beta}^*\|_1$ is in level of s, $\|\boldsymbol{\beta}^*\|_2$ and $\|\boldsymbol{w}^{(0,k)}\|_2$ are in level of $\sqrt{s\vee s'}$ under Assumption 4, we can get the result. The conclusion holds similarly for the target estimator. □

We now prove Theorem 2. Firstly, based on the decomposition

$$\begin{aligned}
&\sup_{k\in\mathcal{A}}|\hat{Q}(\hat{\boldsymbol{w}}^{(0,k)})-\hat{Q}(\hat{\boldsymbol{\beta}}_{\mathcal{I}}^{(0)})|\\
\leq&\sup_{k\in\mathcal{A}}|\hat{Q}(\hat{\boldsymbol{w}}^{(0,k)})-\hat{Q}(\boldsymbol{w}^{(0,k)})|+|\hat{Q}(\hat{\boldsymbol{\beta}}_{\mathcal{I}}^{(0)})-\hat{Q}(\boldsymbol{\beta}^*)|\\
&+\sup_{k\in\mathcal{A}}|\hat{Q}(\boldsymbol{w}^{(0,k)})-\hat{Q}(\boldsymbol{\beta}^*)-Q(\boldsymbol{w}^{(0,k)})+Q(\boldsymbol{\beta}^*)|\\
&+\sup_{k\in\mathcal{A}}|Q(\boldsymbol{w}^{(0,k)})-Q(\boldsymbol{\beta}^*)| \qquad \text{(S3.7)}
\end{aligned}$$

and Lemma 4, we can bound the first and second term of the right hand side as

$$\begin{aligned}
&\sup_{k\in\mathcal{A}}|\hat{Q}(\hat{\boldsymbol{w}}^{(0,k)})-\hat{Q}(\boldsymbol{w}^{(0,k)})|+|\hat{Q}(\hat{\boldsymbol{\beta}}_{\mathcal{I}}^{(0)})-\hat{Q}(\boldsymbol{\beta}^*)|\\
\leq&\sup_{k\neq 0}|\hat{Q}(\hat{\boldsymbol{w}}^{(0,k)})-\hat{Q}(\boldsymbol{w}^{(0,k)})|+|\hat{Q}(\hat{\boldsymbol{\beta}}_{\mathcal{I}}^{(0)})-\hat{Q}(\boldsymbol{\beta}^*)|\\
\lesssim&\kappa_1+\kappa_2
\end{aligned}$$

with probability at least $1-\tilde{c}_1\exp\{-c_1\eta^2\}-\tilde{c}_2\exp\{-c_2\log p\}$ .Meanwhile, the fourth term on the right side of (S3.7) can be bounded by Lemma 3 as

$$\sup_{k\in\mathcal{A}}|Q(\boldsymbol{w}^{(0,k)})-Q(\boldsymbol{\beta}^*)|=O(h^2).$$

The third term on the right side of (S3.7) can be bounded by Lemma 5 as

$$\begin{aligned}
&\sup_{k\in\mathcal{A}}|\hat{Q}(\boldsymbol{w}^{(0,k)})-\hat{Q}(\boldsymbol{\beta}^*)-Q(\boldsymbol{w}^{(0,k)})+Q(\boldsymbol{\beta}^*)|\\
\leq&\sup_{k\in\mathcal{A}}|\hat{Q}(\boldsymbol{w}^{(0,k)})-Q(\boldsymbol{w}^{(0,k)})|+\sup_{k\in\mathcal{A}}|\hat{Q}(\boldsymbol{\beta}^*)-Q(\boldsymbol{\beta}^*)|\\
\leq&\sup_{k\neq 0}|\hat{Q}(\boldsymbol{w}^{(0,k)})-Q(\boldsymbol{w}^{(0,k)})|+\sup_{k\neq 0}|\hat{Q}(\boldsymbol{\beta}^*)-Q(\boldsymbol{\beta}^*)|\\
\lesssim&\kappa_2
\end{aligned}$$

wit probability at least $1-\tilde{c}_1\exp\{-c_1\eta^2\}-\tilde{c}_2\exp\{-c_2\log p\}$. Hence, considering that $\kappa=\kappa_1\vee\kappa_2$, we can get

$$\sup_{k\in\mathcal{A}}|\hat{Q}(\hat{\boldsymbol{w}}^{(0,k)})-\hat{Q}(\hat{\boldsymbol{\beta}}_{\mathcal{I}}^{(0)})|\lesssim h^2+2\kappa$$

with probability at least $1-\tilde{c}_1\exp\{-c_1\eta^2\}-\tilde{c}_2\exp\{-c_2\log p\}$ for some positive constants c_1 ,$\tilde{c}_1$, c_2 and $\tilde{c}_2$.. Hence, the constant c_ε mentioned in Assumption 5 and the fact that $|\hat{Q}(\hat{\boldsymbol{\beta}}_{\mathcal{I}}^{(0)})-Q(\boldsymbol{\beta}^*)|\lesssim 2\kappa_2$ form the combination of Lemma 4 and 5, we can get

$$\sup_{k\in\mathcal{A}}|\hat{Q}(\hat{\boldsymbol{w}}^{(0,k)})-\hat{Q}(\hat{\boldsymbol{\beta}}_{\mathcal{I}}^{(0)})|\lesssim c_\varepsilon\hat{Q}(\hat{\boldsymbol{\beta}}_{\mathcal{I}}^{(0)})$$

with probability at least $1-\tilde{c}_1\exp\{-c_1\eta^2\}-\tilde{c}_2\exp\{-c_2\log p\}$. We can further have

$$\sup_{k\in\mathcal{A}}\hat{Q}(\hat{\boldsymbol{w}}^{(0,k)})\leq(1+c_\varepsilon)\hat{Q}(\hat{\boldsymbol{\beta}}_{\mathcal{I}}^{(0)}).$$

Secondly, we have

$$\begin{aligned}
&\inf_{k\in\mathcal{A}^c} |\hat{Q}(\hat{\boldsymbol{w}}^{(0,k)}) - \hat{Q}(\hat{\boldsymbol{\beta}}_{\mathcal{I}}^{(0)})| \\
\geq &\inf_{k\in\mathcal{A}^c} |Q(\boldsymbol{w}^{(0,k)}) - Q(\boldsymbol{\beta}^*)| \\
&- \sup_{k\in\mathcal{A}^c} |\hat{Q}(\hat{\boldsymbol{w}}^{(0,k)}) - \hat{Q}(\boldsymbol{w}^{(0,k)})| - \sup_{k\in\mathcal{A}^c} |\hat{Q}(\hat{\boldsymbol{\beta}}_{\mathcal{I}}^{(0)}) - \hat{Q}(\boldsymbol{\beta}^*)| \\
&- \sup_{k\in\mathcal{A}^c} |\hat{Q}(\boldsymbol{w}^{(0,k)}) - \hat{Q}(\boldsymbol{\beta}^*) - Q(\boldsymbol{w}^{(0,k)}) + Q(\boldsymbol{\beta}^*)| \\
\geq &\inf_{k\in\mathcal{A}^c} |Q(\boldsymbol{w}^{(0,k)}) - Q(\boldsymbol{\beta}^*)| - 3\kappa
\end{aligned}$$

with probability at least $1 - \tilde{c}_1 \exp\{-c_1\eta^2\} - \tilde{c}_2 \exp\{-c_2 \log p\}$. According to $Q(\boldsymbol{\beta})$ being quadratic in $\boldsymbol{\beta}$, a second-order Taylor expansion around $\boldsymbol{\beta}^*$ together with $\nabla Q(\boldsymbol{\beta}^*) = 0$ yields $Q\big(\boldsymbol{w}^{(0,k)}\big) - Q(\boldsymbol{\beta}^*) = \big(\boldsymbol{w}^{(0,k)} - \boldsymbol{\beta}^*\big)^\top \boldsymbol{\Sigma}^{(0)} \big(\boldsymbol{w}^{(0,k)} - \boldsymbol{\beta}^*\big)$. Since $\tilde{\mathbf{C}}^\top(\boldsymbol{w}^{(0,k)} - \boldsymbol{\beta}^*) = 0$, we further have $\big(\boldsymbol{w}^{(0,k)} - \boldsymbol{\beta}^*\big)^\top \boldsymbol{\Sigma}^{(0)} \big(\boldsymbol{w}^{(0,k)} - \boldsymbol{\beta}^*\big) = \big(\boldsymbol{w}^{(0,k)} - \boldsymbol{\beta}^*\big)^\top \big(\boldsymbol{\Sigma}^{(0)} + \rho\tilde{\mathbf{C}}\tilde{\mathbf{C}}^\top\big) \big(\boldsymbol{w}^{(0,k)} - \boldsymbol{\beta}^*\big)$. Because $\rho > 0$ and $\boldsymbol{\Sigma}^{(0)} + \rho\tilde{\mathbf{C}}\tilde{\mathbf{C}}^\top$ is invertible, it is positive definite. Denote $\lambda_{\min}$ is the minimum eigenvalue of the matrix $\boldsymbol{\Sigma}^{(0)} + \rho\tilde{\mathbf{C}}\tilde{\mathbf{C}}^\top$ and thus

$$\begin{aligned}
&\inf_{k\in\mathcal{A}^c} |\hat{Q}(\hat{\boldsymbol{w}}^{(0,k)}) - \hat{Q}(\hat{\boldsymbol{\beta}}_{\mathcal{I}}^{(0)})| \\
\geq &\lambda_{\min} \inf_{k\in\mathcal{A}^c} \|\boldsymbol{w}^{(0,k)} - \boldsymbol{\beta}^*\|^2 - 3\kappa
\end{aligned}$$

Since $w^{(0,k)} = \boldsymbol{\Omega}_c^{(0,k)}(A_k \boldsymbol{w}^{*(k)} + A_0\boldsymbol{\beta}^*)$, where $A_k = \alpha_k \boldsymbol{\Sigma}^{(k)}$, we can get

$$\begin{aligned}
\boldsymbol{w}^{(0,k)} - \boldsymbol{\beta}^* &= \boldsymbol{\Omega}_c^{(0,k)}(A_k \boldsymbol{w}^{*(k)} + A_0\boldsymbol{\beta}^*) - \boldsymbol{\beta}^* \\
&= \boldsymbol{\Omega}_c^{(0,k)}(A_k \boldsymbol{w}^{*(k)} + A_0\boldsymbol{\beta}^*) - \boldsymbol{\Omega}^{(0,k)}(A_k + A_0)\boldsymbol{\beta}^* \\
&= \boldsymbol{\Omega}_c^{(0,k)} A_k (\boldsymbol{w}^{*(k)} - \boldsymbol{\beta}^*)
\end{aligned}$$

The second line is valid due to $\boldsymbol{\Omega}_c^{(0,k)}(A_k + A_0) = \boldsymbol{I}_p - \tilde{\mathbf{C}}\tilde{\mathbf{C}}^\top$, $\tilde{\mathbf{C}}^\top\boldsymbol{\beta}^* = 0$ and $\tilde{\mathbf{C}}^\top\boldsymbol{w}^{*(k)} = 0$. Hence

$$\begin{aligned}\alpha_k^{-1}\boldsymbol{\Omega}_c^{(k)}\boldsymbol{\Sigma}^{(0,k)}(\boldsymbol{w}^{(0,k)} - \boldsymbol{\beta}^*) &= \alpha_k^{-1}\boldsymbol{\Omega}_c^{(k)}\boldsymbol{\Sigma}^{(0,k)}\boldsymbol{\Omega}_c^{(0,k)}A_k(\boldsymbol{w}^{*(k)} - \boldsymbol{\beta}^*) \\ &= \boldsymbol{\Omega}_c^{(k)}\mathbf{G}\boldsymbol{\Sigma}^{(k)}(\boldsymbol{w}^{*(k)} - \boldsymbol{\beta}^*) \\ &= \mathbf{G}(\boldsymbol{w}^{*(k)} - \boldsymbol{\beta}^*) \\ &= \boldsymbol{w}^{*(k)} - \boldsymbol{\beta}^*,\end{aligned}$$

where $\mathbf{G} = \boldsymbol{I}_p - \tilde{\mathbf{C}}\tilde{\mathbf{C}}^\top$ and the second line is valid due to $\boldsymbol{\Omega}_c^{(k)}\mathbf{G} = \boldsymbol{\Omega}_c^{(k)}$. Therefore, we have

$$\|\boldsymbol{w}^{*(k)} - \boldsymbol{\beta}^*\|_2^2 \leq \|\alpha_k^{-1}\boldsymbol{\Omega}_c^{(k)}\boldsymbol{\Sigma}^{(0,k)}\|_2^2\|\boldsymbol{w}^{(0,k)} - \boldsymbol{\beta}^*\|_2^2,$$

Denote $\kappa_3 = \max_{k\in\mathcal{A}^c}\|\alpha_k^{-1}\boldsymbol{\Omega}_c^{(k)}\boldsymbol{\Sigma}^{(0,k)}\|_2^2$, we can get

$$\|\boldsymbol{w}^{*(k)} - \boldsymbol{\beta}^*\|_2^2 \leq \kappa_3\|\boldsymbol{w}^{(0,k)} - \boldsymbol{\beta}^*\|_2^2,$$

and under Assumption 5, we have

$$\|\boldsymbol{w}^{(0,k)} - \boldsymbol{\beta}^*\|_2^2 \geq \lambda_{min}^{-1}\{c_\epsilon Q(\boldsymbol{\beta}^*) + 4\kappa\},$$

and

$$\begin{aligned}
&\inf_{k\in\mathcal{A}^c} |\hat{Q}(\hat{\boldsymbol{w}}^{(0,k)}) - \hat{Q}(\hat{\boldsymbol{\beta}}_{\mathcal{I}}^{(0)})| \\
\geq &\lambda_{\min} \inf_{k\in\mathcal{A}^c} \|\boldsymbol{w}^{(0,k)} - \boldsymbol{\beta}^*\|^2 - 3\kappa \\
\geq &c_\varepsilon Q(\boldsymbol{\beta}^*) + \kappa \\
\geq &c_\varepsilon \hat{Q}(\hat{\boldsymbol{\beta}}_{\mathcal{I}}^{(0)})
\end{aligned}$$

with probability at least $1-\tilde{c}_1\exp\{-c_1\eta^2\}-\tilde{c}_2\exp\{-c_2\log p\}$, where the last line holds due to $|\hat{Q}(\hat{\boldsymbol{\beta}}^{(0)}) - Q(\boldsymbol{\beta}^*_{\mathcal{I}})| \lesssim \kappa_2$ from the combination of Lemma 4 and 5. Therefore, we can get

$$\inf_{k\in\mathcal{A}^c} \hat{Q}(\hat{\boldsymbol{w}}^{(0,k)}) \geq (1+c_\varepsilon)\hat{Q}(\hat{\boldsymbol{\beta}}_{\mathcal{I}}^{(0)}).$$

with probability at least $1-\tilde{c}_1\exp\{-c_1\eta^2\}-\tilde{c}_2\exp\{-c_2\log p\}$. Finally, we can get our desired result, since

$$\begin{aligned}
&Pr(\hat{\mathcal{A}} \neq \mathcal{A}) \\
\leq &Pr\left(\left\{\inf_{k\in\mathcal{A}^c} \hat{Q}(\hat{\boldsymbol{w}}^{(0,k)}) < (1+c_\varepsilon)\hat{Q}(\hat{\boldsymbol{\beta}}_{\mathcal{I}}^{(0)})\right\} \cup \left\{\sup_{k\in\mathcal{A}} \hat{Q}(\hat{\boldsymbol{w}}^{(0,k)}) > (1+c_\varepsilon)\hat{Q}(\hat{\boldsymbol{\beta}}_{\mathcal{I}}^{(0)})\right\}\right) \\
\leq &Pr\left(\left\{\inf_{k\in\mathcal{A}^c} \hat{Q}(\hat{\boldsymbol{w}}^{(0,k)}) < (1+c_\varepsilon)\hat{Q}(\hat{\boldsymbol{\beta}}_{\mathcal{I}}^{(0)})\right\}\right) + Pr\left(\left\{\sup_{k\in\mathcal{A}} \hat{Q}(\hat{\boldsymbol{w}}^{(0,k)}) > (1+c_\varepsilon)\hat{Q}(\hat{\boldsymbol{\beta}}_{\mathcal{I}}^{(0)})\right\}\right) \\
\leq &\tilde{c}_1\exp\{-c_1\eta^2\} + \tilde{c}_2\exp\{-c_2\log p\}.
\end{aligned}$$

S4 Additional comparison with other constrained methods

S4.1 Comparison with the Constrained Trans-Fusion benchmark

To further examine whether a constrained transfer-learning strategy could help relax the practical difficulty associated with the condition $C_\Sigma^A < \infty$, we conduct additional simulations comparing our method with a constrained version of the Trans-Fusion estimator. Prior work suggests that, in unconstrained settings, both the transfer-learning framework in Li et al. (2024) and the Trans-Fusion estimator in He et al. (2024) can alleviate this difficulty to some extent. In addition, the numerical results in He et al. (2024) suggest that Trans-Fusion is relatively robust to covariate shift. Since the original Trans-Fusion estimator does not incorporate the linear coefficient constraints required in compositional regression, we adapt it by solving

$$\begin{aligned}&\left((\hat{\boldsymbol{w}}^{(0)})^\top,\ (\hat{\boldsymbol{w}}^{(1)})^\top,\ \ldots,\ (\hat{\boldsymbol{w}}^{(K)})^\top\right)^\top \\ &\in \arg\min\left\{\frac{1}{2N}\sum_{k=0}^{K}\left\|y^{(k)}-\mathbf{Z}^{(k)}\boldsymbol{w}^{(k)}\right\|_2^2+\lambda_0\left(\|\boldsymbol{w}^{(0)}\|_1+\sum_{k=1}^{K}a_k\|\boldsymbol{w}^{(k)}-\boldsymbol{w}^{(0)}\|_1\right)\right\},\end{aligned}$$

subject to $\mathbf{C}^\top\boldsymbol{w}^{(k)}=\mathbf{0}$ for $k=0,1,\ldots,K$, and setting $\hat{\boldsymbol{\beta}}=\frac{n_0}{N}\hat{\boldsymbol{w}}^{(0)}+\sum_{k=1}^{K}\frac{n_k}{N}\hat{\boldsymbol{w}}^{(k)}$ as the estimator for $\boldsymbol{\beta}^{(0)}$, where $N=\sum_{k=0}^{K}n_k$.

We consider two types of covariate generation mechanisms: (i) directly generated Gaussian covariates and (ii) generated compositional covariates. We focus on one homogeneous covariance structure and two heterogeneous covariance structures, as well as settings in which all sources are informative and settings in which informative and non-informative sources coexist.

In this simulation, we consider only the 200-dimensional compositional covariates in the

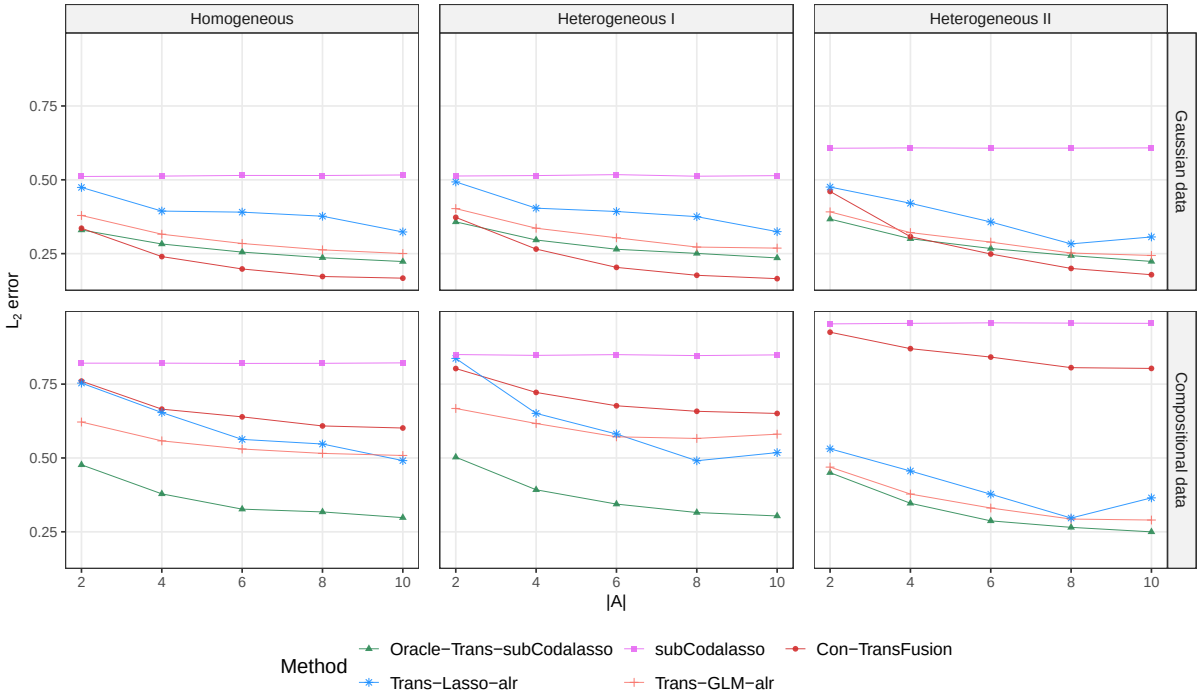

Figure 1: The average ℓ_2-estimation error for the Gaussian and compositional data using five methods with three different covariate distributions, and $K = |\mathcal{A}|$=2,4,6,8,10, all informative sources. For directly generated Gaussian covariates, we use the analogous differencing transformation $\boldsymbol{z}_j - \boldsymbol{z}_p$ to denote alr-based methods.

model, i.e. $(q', q) = (0, 200)$. We use the same data-generating mechanisms as in the previous simulations for compositional covariates. The Gaussian data, in our setting, corresponds to $\mathbf{Z}^{(k)} = \mathbf{U}^{(k)}$. We consider model shift in fixed setting ($d = 4$, Scenario I)and covariate shift with Homogeneous design in $\rho = 0.2$ and the new Heterogeneous design as following (named Heterogeneous design I) and Heterogeneous design in Section 4.1 in main text (named Heterogeneous design II).

- **Heterogeneous design I:** Each $\boldsymbol{u}_i^{(k)} \sim N(\boldsymbol{\nu}, \boldsymbol{\Sigma}_U^{(k)})$, where $\boldsymbol{\Sigma}_U^{(k)} = (\rho^{|i-j|})$ with $\rho = 0.2$ for $k = 0$ and $\rho = 0.3$ for $k = 1, \ldots, K$,.

We first consider the setting in which all source datasets are informative. The number of informative sources is set to $2, 4, 6, 8$, and 10, respectively, and each scenario is repeated 50 times.

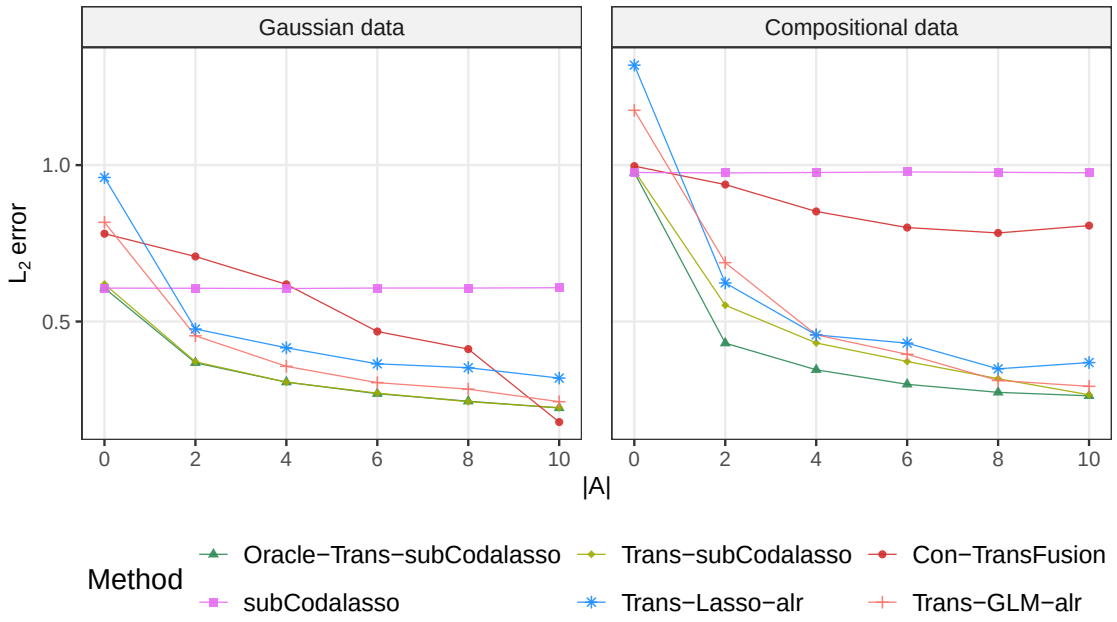

Figure 2: The average ℓ_2-estimation error for the Gaussian and compositional data using six methods with Heterogeneous design II and $K = 10$, number of informative sources $|\mathcal{A}|$=2,4,6,8,10.

Figure 1 reports the ℓ_2 estimation error under Gaussian and compositional designs when all source datasets are informative, for one homogeneous design and two heterogeneous designs. Under the Gaussian design, the constrained Trans-Fusion benchmark outperforms our method in the settings considered. In contrast, under the compositional design, our method performs better in both the homogeneous and heterogeneous settings. A possible reason is that the closure operation and log-contrast structure in compositional regression change the effective source–target discrepancy, making methods tailored to the compositional setting more advantageous.

We next consider a mixed-source scenario in which the total number of auxiliary sources is fixed at 10 and the number of informative sources varies from 0 to 10 in increments of 2. Figure 2 reports the results under Heterogeneous Design II.

Figure 2 shows that, when informative and non-informative sources coexist, the constrained Trans-Fusion benchmark performs worse than our method even under directly gen-

erated Gaussian covariates. This indicates that our method is relatively more robust in the presence of mixed informative and non-informative sources, including in the compositional setting. A possible reason is that our two-step transfer procedure can benefit from an explicit source-selection step, whereas one-step procedures directly regularize the discrepancy between the target and source domains.

S4.2 Comparison with constrained ℓ_0-based methods

In addition to our ℓ_1-based approach, we further examine ℓ_0-based alternatives, and consider two linearly constrained ℓ_0-type competitors: a constrained extension of SDAR (Huang et al., 2018) and a hard-thresholding constrained ℓ_0 method, inspired by MIO-based best subset selection approaches (Bertsimas et al., 2016). In our implementation, the linear equality constraint is handled through an augmented-Lagrangian scheme, while sparsity is enforced by coordinate-wise hard-thresholding updates. Section S4.2.1 describes the two algorithms. The simulation settings, numerical results, and analysis are presented in Sections S4.2.2 and S4.2.3.

S4.2.1 Algorithms

We briefly describe the two constrained ℓ_0-type competitors considered in the additional simulations. The first method is a constrained extension of SDAR, where the refitting step is modified to accommodate the linear equality constraint. The second method is a hard-thresholding constrained ℓ_0 algorithm based on an augmented-Lagrangian formulation.

The first competitor is a constrained extension of SDAR. The standard SDAR algorithm alternates between support detection and root finding. To incorporate the compositional linear constraint, we retain the SDAR active-set selection rule and replace the ordinary least-squares refitting step with a constrained least-squares problem on the selected active

set. Specifically, after an active set is identified, the coefficients on this active set are obtained by solving the corresponding KKT system under the linear equality constraint, with zero rows of $C_{\mathcal{A}^t}$ removed when necessary. The detailed procedure is summarized in Algorithm S3.

Algorithm S3 Constrained SDAR algorithm

Require: $Z, y, C, w_e, \lambda_0, T, \epsilon, k$
Ensure: $\widehat{\beta}$
  Initialize $\beta^0 = 0$, $r^0 = y - Z\beta^0$, $d^0 = Z^\top r^0/n$
  Set $\tau_j = w_{e,j}\sqrt{2\lambda_0}$, $\mathcal{U} = \{j : w_{e,j} = 0\}$, $\mathcal{P} = \{1, \ldots, p\} \setminus \mathcal{U}$, $\mathcal{A}^{\text{old}} = \emptyset$
  for $t = 0, 1, \ldots, T-1$ **do**
    $s^t = \beta^t + d^t$
    Select active set $\mathcal{A}^t = \mathcal{U} \cup \{j \in \mathcal{P} : |s_j^t| \geq \tau_j\}$, retaining only the top k penalized indices if necessary. Let $\mathcal{I}^t = (\mathcal{A}^t)^c$.
    Refit on $\mathcal{A}^t$ by solving $\widehat{\beta}_{\mathcal{A}^t} = \arg\min_b \frac{1}{2n}\|y - Z_{\mathcal{A}^t} b\|_2^2$ s.t. $C_{\mathcal{A}^t} b = 0$, using the corresponding KKT system, after removing zero rows of $C_{\mathcal{A}^t}$ when necessary.
    Set $\beta_{\mathcal{A}^t}^{t+1} = \widehat{\beta}_{\mathcal{A}^t}$, $\beta_{\mathcal{I}^t}^{t+1} = 0$.
    Update $r^{t+1} = y - Z\beta^{t+1}$, $d^{t+1} = Z^\top r^{t+1}/n$.
    if $\mathcal{A}^t = \mathcal{A}^{\text{old}}$ or $\|\beta^{t+1} - \beta^t\|_2 < \epsilon$ **then**
      break
    end if
    $\mathcal{A}^{\text{old}} = \mathcal{A}^t$
  end for
  return $\widehat{\beta} = \beta^{t+1}$

The second competitor is a hard-thresholding constrained ℓ_0 method. This method directly targets the linearly constrained ℓ_0-penalized problem through an augmented-Lagrangian framework. The linear equality constraint is handled by dual-variable and penalty updates, while sparsity is enforced through coordinate-wise hard-thresholding updates. The detailed procedure is summarized in Algorithm S4.

S4.2.2 Simulation settings

In this simulation study, the data are generated in the same manner as described in Section 4 of the main text. Here, we consider a model that involves only compositional data, and the covariance matrix is given by $\mathbf{\Sigma}_U^{(k)} = (0.5^{|j_1 - j_2|})$ for $j_1, j_2 = 1, \ldots, p$, $k = 0, \ldots, K$. To investigate the impact of the source, the coefficient vector $\boldsymbol{\theta}^{*(k)}$ for the source is set as follows. $\boldsymbol{\theta}^{*(k)}$ share the same grouping constraints as $\boldsymbol{\theta}^{*(0)}$. Specifically, for each source model, we randomly select one group and randomly perturb two components within the selected

Algorithm S4 Hard-thresholding constrained ℓ_0 algorithm (HTC-ℓ_0)

Require: Z_t, y, C, w_e,λ_0, T, ϵ, μ, ν.
Ensure: $\widehat{\beta}$.
  Compute $Z_z = Z_t^\top Z_t/n$, $z_y = Z_t^\top y/n$, $C_c = CC^\top$.
  Let $d_z = \mathrm{diag}(Z_z)$ and $d_c = \mathrm{diag}(C_c)$; set diagonal entries of Z_z and C_c to zero.
  Initialize $\beta^0 = 0$, $\mathrm{ssq}^0 = \sqrt{\|y - Z_t\beta^0\|_2^2/n}$, $\lambda = w_e\sqrt{2\lambda_0}$.
  for $m = 0, 1, 2, \ldots, T-1$ **do**
    $\beta^{\mathrm{old}} = \beta^m$, $\mathrm{ssq}^{\mathrm{old}} = \mathrm{ssq}^m$.
    Initialize $\psi = 0$.
    for $t = 0, 1, 2, \ldots, T-1$ **do**
      $\widetilde{\beta} = \beta^m$.
      for $j = 1, 2, \ldots, p$ **do**
        Compute $d_j = z_{y,j} - \sum_{\ell=1}^p Z_{z,j\ell}\widetilde{\beta}_\ell - \mu\left(\sum_{\ell=1}^p C_{c,j\ell}^\top\widetilde{\beta}_\ell + \sum_{r=1}^k C_{jr}^\top\psi_r\right)$.
        Update $\widetilde{\beta}_j = \frac{\mathcal{H}(d_j,\lambda_j)}{d_{x,j}+\mu d_{c,j}+10^{-8}}$, where $\mathcal{H}(u,\lambda) = \begin{cases} u, & |u| \geq \lambda, \\ 0, & |u| < \lambda. \end{cases}$
      end for
      $\mu \leftarrow \nu\mu$, $\psi \leftarrow \psi + C^\top\widetilde{\beta}$, $\mathrm{err} = \|\widetilde{\beta} - \beta^m\|_2$, $\beta^m \leftarrow \widetilde{\beta}$.
      if $\mathrm{err} < \epsilon$ **then**
        break
      end if
    end for
    $\mathrm{ssq}^{m+1} = \sqrt{\|y - Z_t\beta^m\|_2^2/n}$.
    $\mathrm{err}_\beta = \|\beta^m - \beta^{\mathrm{old}}\|_2$.
    if $\mathrm{err}_\beta < \epsilon$ **and** $|\mathrm{ssq}^{m+1} - \mathrm{ssq}^{\mathrm{old}}| < \epsilon$ **then**
      break
    end if
  end for
  return $\widehat{\beta} = \beta^m$.

group. For informative sources, the perturbations are generated from a uniform distribution on $(0, 0.05)$. When the grouped subcompositional structure is ignored, we instead randomly select two components from the entire composition and perturb them. In all settings, the true coefficient vector $\boldsymbol{\theta}^{*(0)}$ is 100-dimensional, and its nonzero entries are specified as follows:

- **Regime I: strong signals and sparse coefficients** For subcompositional model, as in Shi et al. (2016), the target coefficient $\boldsymbol{\theta}^{*(0)}$ has nonzero entries at positions $1, 2, 3, 6, 11, 13$, and 16, with values $1, -0.8, 0.4, -0.6, -1.5, 1.2$, and 0.3, respectively; all other entries are zero. The regression coefficient $\boldsymbol{\theta}^{*(0)}$ used in the simulation satisfies the following 3 linear constraints:

$$\sum_{j=1}^{10}\theta_j^*(0) = 0, \qquad \sum_{j=11}^{16}\theta_j^{*(0)} = 0, \qquad \sum_{j=17}^{p}\theta_j^{*(0)} = 0.$$

For single compositional model, as in Lin et al. (2014), the target coefficient $\boldsymbol{w}^{*(0)}$ has nonzero entries at positions $1, 2, 3, 6, 7$, and 8, with values $1, -0.8, 0.6, -1.5, -0.5$, and 1.2, respectively; all other entries are zero.

- **Regime II: weak signals and dense coefficients** For subcompositional model, the nonzero entries of $\boldsymbol{\theta}^{*(0)}$ are located at positions $1, \ldots, 16$ and $25, \ldots, 40$. The coefficients at positions $1, \ldots, 16$ are

  $(0.14, -0.07, 0.03, -0.19, 0.11, -0.04, 0.08, -0.16, 0.05, 0.12, -0.33, 0.27, -0.41, 0.18, 0.46, -0.24)$,

  and those at positions $25, \ldots, 40$ repeat the same pattern; all remaining entries are zero. The regression coefficient $\boldsymbol{w}^{(0)}$ used in the simulation satisfies the following 3 linear constraints:

$$\sum_{j=1}^{20} \theta_j^{*(0)} = 0, \qquad \sum_{j=21}^{40} \theta_j^{*(0)} = 0, \qquad \sum_{j=41}^{p} \theta_j^{*(0)} = 0.$$

  For the single compositional model, $\boldsymbol{\theta}^{*(0)}$ is nonzero only in the first 16 positions, with the same values as in the first 16 components of the subcompositional model.

On the other hand, in this simulation, the candidate values of the tuning parameter are taken from $[c_{\min}\lambda_{\text{init}}, c_{\max}\lambda_{\text{init}}]$, where λ_{init} is a data-adaptive initial value defined by $\lambda_{\text{init}} = \frac{\|X^\top y\|_\infty}{(n_0+n_{\mathcal{A}})\sqrt{\log(G+1)}}$, and $c_{\min}$ and $c_{\max}$ are pre-specified constants. A logarithmically spaced grid with $R = 50$ values is used, and the final value is chosen by cross-validation. To assess sensitivity, we further consider three choices of $(c_{\min}, c_{\max})$: (0.005, 0.5), (0.05, 0.25), and (0.1, 1).

To evaluate estimation accuracy, we calculate two metrics based on $\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*$ for each method: the ℓ_2 estimation error (L_2-error), defined as:L_2-error $= \|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*\|_2$. In addition, we compute the number of false positives and the number of false negatives, where positives and negatives refer to nonzero and zero coefficients, respectively.

S4.2.3 Simulation results and analysis

Tables 1–2 show that under the strong signals and sparse coefficient settings, the constrained ℓ_0-type methods generally achieve better estimation accuracy than the proposed linearly constrained ℓ_1-penalized method, with the constrained extension of SDAR often performing best but at a substantially higher computational cost. The constrained hard-thresholding ℓ_0 method is computationally faster and could outperform the proposed method when the number of informative source datasets was small. In contrast, Tables 3–4 show that under the weak signals and dense coefficient settings, the performance of the constrained ℓ_0-type methods deteriorates markedly, especially for the constrained SDAR extension, while the proposed method remains more stable. Tables 5–6 further indicate that the constrained hard-thresholding ℓ_0 method is more sensitive to the choice of $(c_{\min}, c_{\max})$, particularly in the weak-signal settings.

In terms of variable selection, our proposed method has a smaller number of false negatives, which is important for compositional data analysis because missing truly relevant components can be problematic. Therefore, while the constrained ℓ_0-type methods can be advantageous in strong-signal and sparse settings, the proposed linearly constrained ℓ_1-penalized transfer learning method provides a more stable and robust choice across different signal and sparsity regimes.

Table 1: Results for the single compositional model with strong signals and sparse coefficients

$\lvert\mathcal{A}\rvert$	Model	Metric, mean(sd)				
		Runtime (s)	L_2-error	FN	FP	Predicition error
2	subCodalasso	2.84(0.18)	1.03(0.21)	0.20(0.47)	32.02(11.92)	2.99(1.34)
	Oracle-Trans-subCodalasso	5.94(0.35)	0.62(0.14)	0.01(0.10)	23.32(11.18)	2.04(4.08)
	Constrained SDAR	284.09(60.95)	0.32(0.24)	0.19(0.51)	0.48(1.56)	1.73(0.36)
	HTC-ℓ_0	4.82(0.38)	0.47(0.28)	0.55(0.78)	0.64(1.51)	1.82(0.40)
	Trans-Lasso-alr	0.12(0.02)	0.60(0.12)	0.00(0.00)	13.35(10.66)	1.83(0.31)
	Trans-GLM-alr	0.54(0.08)	3.25(0.12)	1.87(0.39)	10.49(5.33)	7.04(3.20)
4	subCodalasso	2.79(0.15)	1.03(0.22)	0.20(0.47)	31.98(11.94)	2.99(1.34)
	Oracle-Trans-subCodalasso	6.23(0.38)	0.48(0.11)	0.00(0.00)	21.86(10.69)	1.90(0.42)
	Constrained SDAR	259.39(60.91)	0.21(0.18)	0.05(0.26)	0.40(1.60)	1.66(0.34)
	HTC-ℓ_0	4.96(0.32)	0.29(0.22)	0.24(0.43)	0.19(0.73)	1.69(0.34)
	Trans-Lasso-alr	0.10(0.01)	0.48(0.09)	0.00(0.00)	10.81(7.06)	1.75(0.28)
	Trans-GLM-alr	0.50(0.07)	3.34(0.10)	1.86(0.35)	9.17(4.60)	7.28(1.19)
8	subCodalasso	2.28(0.09)	1.03(0.22)	0.20(0.47)	31.89(11.85)	2.99(1.34)
	Oracle-Trans-subCodalasso	5.72(0.25)	0.37(0.11)	0.00(0.00)	20.02(10.70)	1.77(0.37)
	Constrained SDAR	186.50(41.22)	0.16(0.15)	0.02(0.14)	0.36(1.64)	1.64(0.34)
	HTC-ℓ_0	4.56(0.28)	0.18(0.17)	0.04(0.20)	0.33(1.29)	1.65(0.34)
	Trans-Lasso-alr	0.06(0.00)	0.42(0.10)	0.00(0.00)	10.90(7.09)	1.72(0.30)
	Trans-GLM-alr	0.16(0.02)	3.42(0.09)	1.77(0.42)	8.54(6.28)	7.54(1.24)

Table 2: Results for the subcompositional model with strong signals and sparse coefficients

$\lvert\mathcal{A}\rvert$	Model	Metric, mean(sd)				
		Runtime (s)	L_2-error	FN	FP	Predicition error
2	subCodalasso	2.69(0.13)	0.78(0.18)	0.17(0.40)	34.64(11.48)	1.84(0.64)
	Oracle-Trans-subCodalasso	5.32(0.25)	0.50(0.10)	0.00(0.00)	23.50(10.14)	1.38(0.31)
	Constrained SDAR	396.44(40.60)	0.33(0.19)	0.44(0.59)	1.91(3.08)	1.28(0.33)
	HTC-ℓ_0	3.63(0.28)	0.28(0.18)	0.30(0.46)	0.59(1.60)	1.23(0.27)
	Trans-Lasso-alr	0.05(0.01)	0.50(0.10)	0.00(0.00)	14.25(10.32)	1.30(0.25)
	Trans-GLM-alr	0.17(0.02)	3.37(0.07)	3.54(0.64)	13.51(5.23)	6.87(1.09)
4	subCodalasso	3.76(0.18)	0.78(0.19)	0.17(0.40)	34.79(11.57)	1.83(0.64)
	Oracle-Trans-subCodalasso	8.03(0.41)	0.40(0.09)	0.00(0.00)	22.58(10.70)	1.30(0.27)
	Constrained SDAR	850.27(96.99)	0.21(0.16)	0.19(0.42)	1.05(2.39)	1.18(0.25)
	HTC-ℓ_0	5.95(0.37)	0.19(0.14)	0.11(0.31)	0.29(1.17)	1.18(0.26)
	Trans-Lasso-alr	0.13(0.01)	0.41(0.08)	0.00(0.00)	11.32(7.19)	1.23(0.22)
	Trans-GLM-alr	0.60(0.08)	3.43(0.07)	3.58(0.52)	11.64(4.36)	7.02(1.09)
8	subCodalasso	3.37(0.18)	0.78(0.18)	0.17(0.40)	34.94(11.46)	1.83(0.64)
	Oracle-Trans-subCodalasso	8.30(0.40)	0.31(0.07)	0.00(0.00)	21.47(10.21)	1.22(0.24)
	Constrained SDAR	606.13(79.62)	0.16(0.13)	0.09(0.29)	0.99(2.48)	1.16(0.24)
	HTC-ℓ_0	6.16(0.31)	0.13(0.12)	0.01(0.10)	0.31(1.18)	1.16(0.25)
	Trans-Lasso-alr	0.09(0.01)	0.36(0.09)	0.00(0.00)	11.56(7.55)	1.21(0.22)
	Trans-GLM-alr	0.55(0.09)	3.49(0.07)	3.53(0.69)	11.42(5.54)	7.20(1.16)

S5 Additional numerical results

S5.1 More details about the implemetation of numerical experiments

All experiments in this paper are conducted in R. The two constrained optimization problems in our transfer learning framework can be solved in practice using coding techniques similar to Shi et al. (2016) and Mishra and Müller (2022). All methods involving the selection of the tuning parameter λ are implemented using 10-fold cross-validation(CV), including

Table 3: Results for the single compositional model with weak signals and dense coefficients

$\lvert\mathcal{A}\rvert$	Model	Metric, mean(sd)				
		Runtime (s)	L_2-error	FN	FP	Predicition error
2	subCodalasso	3.92(0.14)	0.31(0.08)	0.82(1.27)	62.30(11.37)	0.13(0.06)
	Oracle-Trans-subCodalasso	7.89(0.33)	0.16(0.04)	0.19(0.42)	58.59(11.77)	0.07(0.02)
	Constrained SDAR	279.31(67.70)	0.22(0.04)	5.94(1.14)	0.47(1.31)	0.09(0.03)
	HTC-ℓ_0	4.08(0.24)	0.16(0.04)	4.36(1.31)	0.35(0.74)	0.06(0.02)
	Trans-Lasso-alr	0.15(0.02)	0.15(0.02)	0.67(0.74)	33.57(10.92)	0.06(0.01)
	Trans-GLM-alr	0.41(0.05)	1.43(0.02)	2.31(0.94)	20.32(6.41)	0.99(0.15)
4	subCodalasso	3.83(0.12)	0.31(0.08)	0.81(1.27)	62.35(11.38)	0.13(0.06)
	Oracle-Trans-subCodalasso	8.09(0.25)	0.13(0.03)	0.08(0.34)	58.96(9.99)	0.06(0.01)
	Constrained SDAR	327.98(75.18)	0.21(0.04)	5.86(1.06)	0.50(1.59)	0.09(0.03)
	HTC-ℓ_0	4.63(0.25)	0.14(0.03)	4.06(0.85)	0.12(0.52)	0.06(0.02)
	Trans-Lasso-alr	0.13(0.01)	0.12(0.02)	0.47(0.64)	30.63(9.25)	0.06(0.01)
	Trans-GLM-alr	0.39(0.05)	1.44(0.02)	2.00(0.72)	17.13(5.00)	1.03(0.15)
8	subCodalasso	3.98(0.13)	0.31(0.08)	0.76(1.14)	62.28(11.41)	0.13(0.06)
	Oracle-Trans-subCodalasso	9.88(0.33)	0.10(0.02)	0.01(0.10)	57.72(8.84)	0.05(0.01)
	Constrained SDAR	450.63(82.90)	0.20(0.03)	5.62(0.96)	0.83(2.07)	0.08(0.02)
	HTC-ℓ_0_old	5.91(0.23)	0.12(0.02)	3.69(0.75)	0.07(0.43)	0.05(0.01)
	Trans-Lasso-alr	0.13(0.01)	0.11(0.02)	0.12(0.33)	32.27(9.80)	0.05(0.01)
	Trans-GLM-alr	0.39(0.05)	1.47(0.01)	1.46(0.56)	16.62(5.43)	1.07(0.16)

Table 4: Results for the subcompositional model with weak signals and dense coefficients

$\lvert A\rvert$	Model	Metric, mean(sd)				
		Runtime (s)	L_2-error	FN	FP	Predicition error
2	subCodalasso	2.92(0.05)	0.54(0.11)	1.86(2.14)	54.04(8.14)	0.30(0.11)
	Oracle-Trans-subCodalasso	5.81(0.08)	0.26(0.05)	0.61(0.83)	52.68(5.44)	0.13(0.03)
	Constrained SDAR	367.05(55.15)	0.71(0.12)	19.45(3.24)	1.92(2.71)	0.40(0.17)
	HTC-ℓ_0	3.00(0.17)	0.27(0.06)	8.55(2.50)	3.20(2.77)	0.14(0.03)
	Trans-Lasso-alr	0.07(0.01)	0.26(0.04)	1.56(1.47)	41.45(9.02)	0.14(0.03)
	Trans-GLM-alr	0.19(0.02)	2.04(0.03)	5.35(1.89)	25.33(5.80)	2.31(0.34)
4	subCodalasso	3.13(0.37)	0.54(0.11)	1.79(2.14)	53.94(8.16)	0.30(0.11)
	Oracle-Trans-subCodalasso	6.60(0.82)	0.20(0.04)	0.19(0.44)	53.60(5.14)	0.12(0.02)
	Constrained SDAR	449.91(69.28)	0.67(0.11)	19.53(2.86)	2.08(2.82)	0.36(0.13)
	HTC-ℓ_0	3.59(0.55)	0.20(0.05)	7.33(1.95)	1.68(1.67)	0.12(0.02)
	Trans-Lasso-alr	0.08(0.02)	0.23(0.04)	0.73(0.99)	42.77(9.19)	0.13(0.02)
	Trans-GLM-alr	0.20(0.05)	2.07(0.03)	3.82(1.53)	24.77(5.54)	2.37(0.35)
8	subCodalasso	3.18(0.43)	0.54(0.11)	1.81(2.13)	53.92(8.02)	0.30(0.11)
	Oracle-Trans-subCodalasso	7.43(1.05)	0.15(0.02)	0.05(0.22)	55.14(4.48)	0.11(0.02)
	Constrained SDAR	544.32(98.19)	0.66(0.09)	19.10(3.16)	1.92(2.56)	0.34(0.09)
	HTC-ℓ_0	4.37(0.65)	0.17(0.03)	6.87(1.52)	0.47(1.14)	0.11(0.02)
	Trans-GLM-alr	0.08(0.02)	0.22(0.04)	0.17(0.38)	44.76(8.34)	0.13(0.02)
	Trans-GLM-alr	0.21(0.05)	2.10(0.02)	2.54(1.08)	23.53(5.18)	2.46(0.36)

the selection of initial parameters and the main experiments. The parameter λ_ω is chosen according to the `lambda.1se` criterion, which selects the largest tuning parameter whose cross-validation error lies within one standard error of the minimum cross-validation error. This criterion is used to obtain a more stable and sparse estimator. The parameter λ_δ is chosen according to the `lambda.min` criterion, which selects the tuning parameter that minimizes the cross-validation error and helps mitigate the bias induced by the preceding transferring step. The threshold parameter ϵ_0, which controls the source-domain selection

Table 5: Results for weak signals and dense coefficients in the single compositional model under different ranges of initial values for the tuning parameter, reported as mean (sd)

$\|A\|$	Model	$(c_{\min}, c_{\max}) = (0.05, 0.25)$				$(c_{\min}, c_{\max}) = (0.1, 1)$			
		L_2-error	FN	FP	Predicition error	L_2-error	FN	FP	Predicition error
2	subCodalasso	0.32(0.08)	1.00(1.08)	57.54(9.88)	0.13(0.07)	0.39(0.14)	1.99(2.32)	45.15(11.61)	0.16(0.07)
	Oracle-Trans-subCodalasso	0.20(0.03)	0.32(0.53)	53.90(6.97)	0.08(0.02)	0.29(0.03)	1.32(0.89)	35.20(9.54)	0.04(0.02)
	HTC-ℓ_0	0.58(0.10)	11.97(1.51)	0.07(0.26)	0.22(0.09)	0.72(0.09)	14.00(1.11)	0.09(0.32)	0.30(0.08)
	Trans-Lasso-alr	0.30(0.03)	4.98(0.75)	1.09(0.29)	0.08(0.02)	0.52(0.04)	8.79(1.10)	1.00(0.00)	0.17(0.03)
	Trans-GLM-alr	1.30(0.04)	4.89(1.05)	7.33(4.47)	0.80(0.13)	1.16(0.05)	7.31(1.08)	5.52(3.06)	0.64(0.11)
4	subCodalasso	0.32(0.07)	0.96(1.01)	58.44(9.27)	0.13(0.07)	0.39(0.10)	1.95(2.33)	46.13(11.92)	0.15(0.07)
	Oracle-Trans-subCodalasso	0.17(0.02)	0.25(0.52)	52.75(7.36)	0.08(0.03)	0.26(0.03)	1.21(0.74)	31.67(10.96)	0.09(0.02)
	HTC-ℓ_0	0.57(0.10)	11.82(1.53)	0.06(0.24)	0.21(0.09)	0.72(0.08)	14.01(0.93)	0.09(0.29)	0.30(0.07)
	Trans-Lasso-alr	0.28(0.02)	4.88(0.61)	1.00(0.00)	0.08(0.01)	0.50(0.04)	8.31(0.91)	1.00(0.00)	0.15(0.03)
	Trans-GLM-alr	1.32(0.03)	4.55(1.05)	6.57(4.85)	0.83(0.13)	1.18(0.04)	6.83(0.95)	5.33(2.97)	0.66(0.11)
8	subCodalasso	0.31(0.07)	0.96(1.01)	58.49(9.42)	0.13(0.07)	0.38(0.09)	1.79(1.91)	46.57(11.05)	0.15(0.08)
	Oracle-Trans-subCodalasso	0.18(0.02)	0.10(0.30)	42.52(8.02)	0.07(0.02)	0.25(0.02)	0.96(0.70)	23.93(11.47)	0.08(0.02)
	HTC-ℓ_0	0.56(0.09)	11.75(1.40)	0.02(0.14)	0.20(0.07)	0.71(0.07)	14.08(0.80)	0.06(0.28)	0.29(0.06)
	Trans-Lasso-alr	0.28(0.02)	4.95(0.39)	1.00(0.00)	0.08(0.01)	0.50(0.03)	8.37(0.86)	1.00(0.00)	0.16(0.03)
	Trans-GLM-alr	1.33(0.02)	4.45(0.96)	6.86(5.15)	0.84(0.13)	1.18(0.04)	6.75(1.01)	5.12(2.87)	0.66(0.11)

Table 6: Results for weak signals and dense coefficients in the subcompositional model under different ranges of initial values for the tuning parameter, reported as mean (sd)

$\lvert A\rvert$	Model	$(c_{\min}, c_{\max}) = (0.05, 0.25)$				$(c_{\min}, c_{\max}) = (0.1, 1)$			
		L_2-error	FN	FP	Predicition error	L_2-error	FN	FP	Predicition error
2	subCodalasso	0.53(0.10)	1.76(1.57)	53.40(6.40)	0.30(0.11)	0.59(0.11)	2.83(2.49)	48.45(6.93)	0.33(0.13)
	Oracle.Codalasso	0.29(0.04)	0.66(0.74)	53.56(4.32)	0.16(0.05)	0.39(0.05)	2.06(1.36)	39.68(6.30)	0.17(0.04)
	HTC-ℓ_0	0.61(0.14)	19.00(2.84)	0.24(0.53)	0.30(0.10)	0.88(0.11)	24.84(2.42)	0.03(0.17)	0.47(0.14)
	Trans-Lasso-alr	0.32(0.04)	6.78(1.81)	8.22(3.34)	0.14(0.03)	0.52(0.07)	12.12(1.63)	1.74(1.12)	0.21(0.05)
	Trans-GLM-alr	1.96(0.04)	7.39(1.83)	16.46(3.50)	2.15(0.33)	1.82(0.06)	11.44(2.00)	10.35(3.19)	1.90(0.30)
4	subCodalasso	0.53(0.10)	1.70(1.57)	53.59(6.53)	0.30(0.11)	0.59(0.11)	2.73(2.39)	48.50(7.12)	0.33(0.13)
	Oracle.Codalasso	0.25(0.03)	0.32(0.57)	53.23(4.26)	0.14(0.04)	0.36(0.04)	1.63(1.19)	35.02(6.86)	0.15(0.03)
	HTC-ℓ_0	0.61(0.13)	19.19(2.68)	0.09(0.32)	0.28(0.09)	0.84(0.09)	24.35(2.11)	0.10(0.48)	0.42(0.11)
	Trans-Lasso-alr	0.29(0.03)	6.50(1.71)	4.06(1.11)	0.13(0.02)	0.49(0.06)	11.90(1.51)	1.12(0.41)	0.19(0.04)
	Trans-GLM-alr	1.98(0.04)	6.77(1.60)	13.75(4.60)	2.19(0.33)	1.85(0.06)	10.48(1.88)	9.48(4.60)	1.95(0.30)
8	subCodalasso	0.53(0.10)	1.69(1.56)	53.65(6.60)	0.30(0.11)	0.58(0.11)	2.70(2.49)	49.07(7.04)	0.32(0.13)
	Oracle.Codalasso	0.22(0.02)	0.10(0.33)	51.66(3.91)	0.14(0.04)	0.33(0.03)	1.00(0.83)	30.59(6.00)	0.14(0.03)
	HTC-ℓ_0	0.56(0.12)	18.24(2.32)	0.17(0.51)	0.26(0.08)	0.80(0.09)	23.51(2.12)	0.05(0.26)	0.39(0.09)
	Trans-Lasso-alr	0.26(0.03)	5.99(1.59)	1.69(1.08)	0.12(0.02)	0.46(0.04)	11.39(1.21)	1.01(0.10)	0.18(0.03)
	Trans-GLM-alr	2.00(0.03)	6.11(1.25)	11.36(5.24)	2.24(0.33)	1.87(0.05)	9.68(1.80)	8.44(5.17)	1.99(0.31)

step in Trans-subCodalasso, is fixed at 0.01 in the simulation studies to strictly regulate the inclusion of source domains and reduce the potential risk of negative transfer. In the real data analysis, ϵ_0 is selected via grid search over the candidate set $\{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 1\}$, and the value yielding the smallest validation prediction error is used. The implementations of Trans-GLM-alr and Trans-Lasso-alr are consistent with those proposed by Tian and Feng (2023) and Li et al. (2022).

S5.2 Further results for simulation study

S5.2.1 Performance evaluation under different source-domain settings

In this part, we first report the ℓ_2 estimation errors for the cases $(q', q) = (0, 100)$, $(10, 50)$ and $(10, 200)$.The results are presented in Figures 3–11. Overall, the results are qualitatively similar to those under the setting $(q', q) = (10, 100)$ in the main text, and the proposed method consistently exhibits distinct advantages over the comparison methods under almost all settings.

It is worth noting that, similar to the case where $(q', q) = (10, 100)$, the pooled method also performs well under the homogeneous setting and Scenario II when the number of informative source domains $|\mathcal{A}|$ exceeds 10. However, such obvious superiority of the pooled method is not observed for $q' = 0$ and $(q', q) = (10, 50)$. This indicates that as the dimension of variables increases, pooling all source domains can still achieve competitive performance when non-informative sources do not deviate substantially from the target domain.

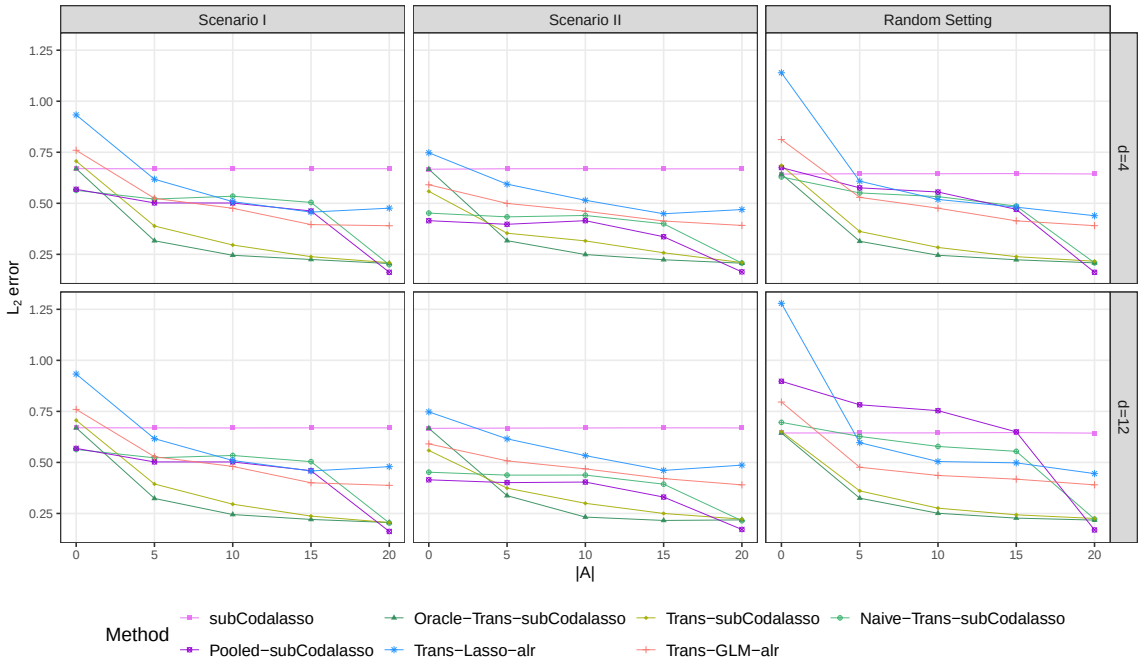

Figure 3: The average L_2-error of the seven methods under different settings for coefficient vector with homogeneous covariance matrices ($\rho = 0.2$, $(q', q) = (0, 100)$).

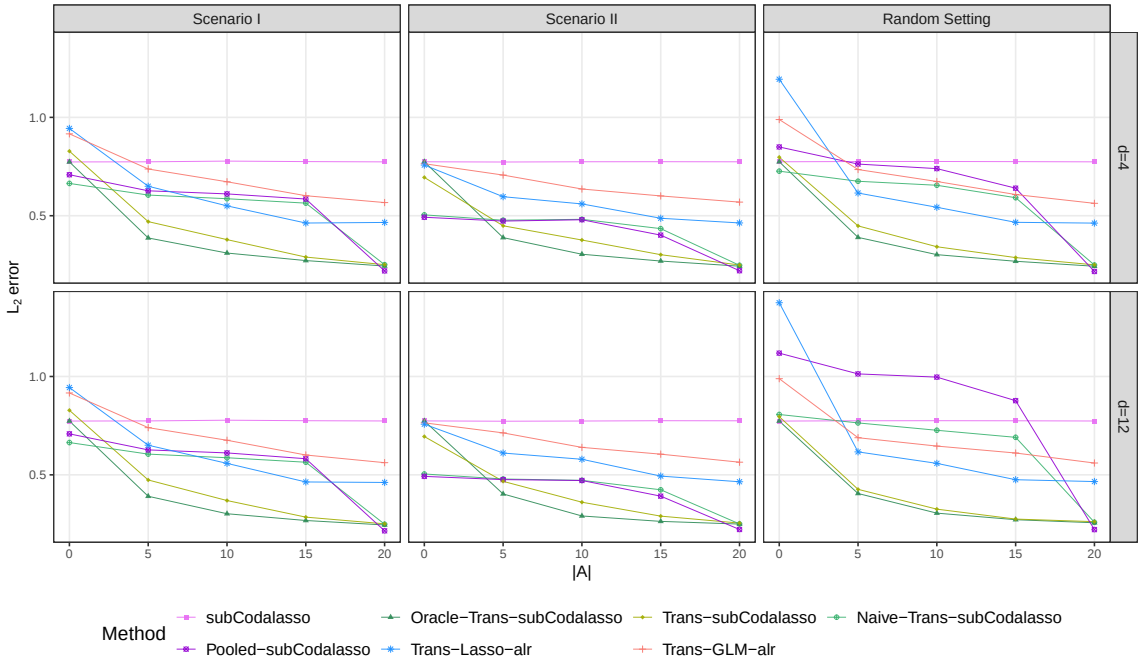

Figure 4: The average L_2-error of the seven methods under different settings for coefficient vector with homogeneous covariance matrices ($\rho = 0.5$, $(q', q) = (0, 100)$).

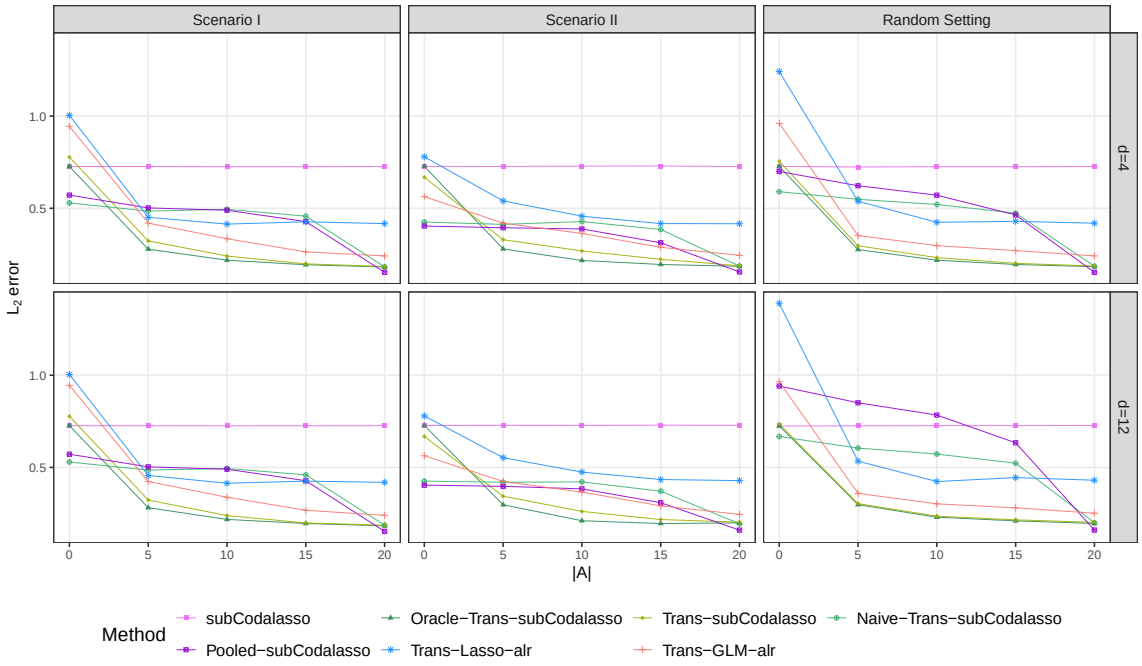

Figure 5: The average L_2-error of the seven methods under different settings for coefficient vector with heterogenous covariance matrices $(q', q) = (0, 100)$.

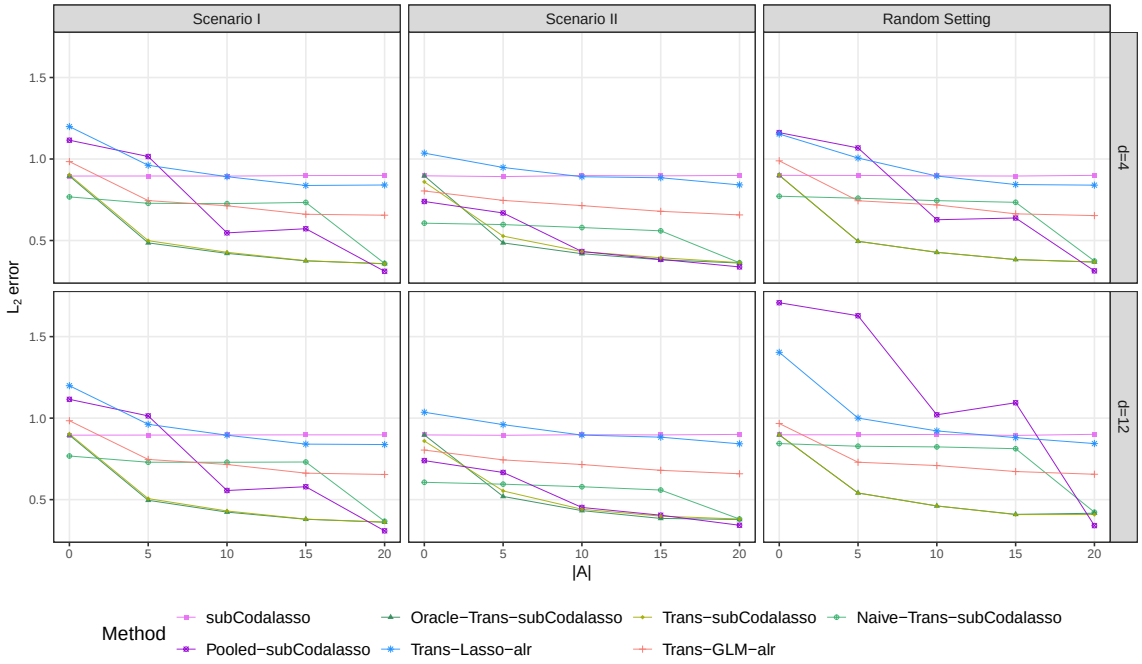

Figure 6: The average L_2-error of the seven methods under different settings for coefficient vector with homogeneous covariance matrices $(\rho = 0.2,\ (q', q) = (10, 50))$.

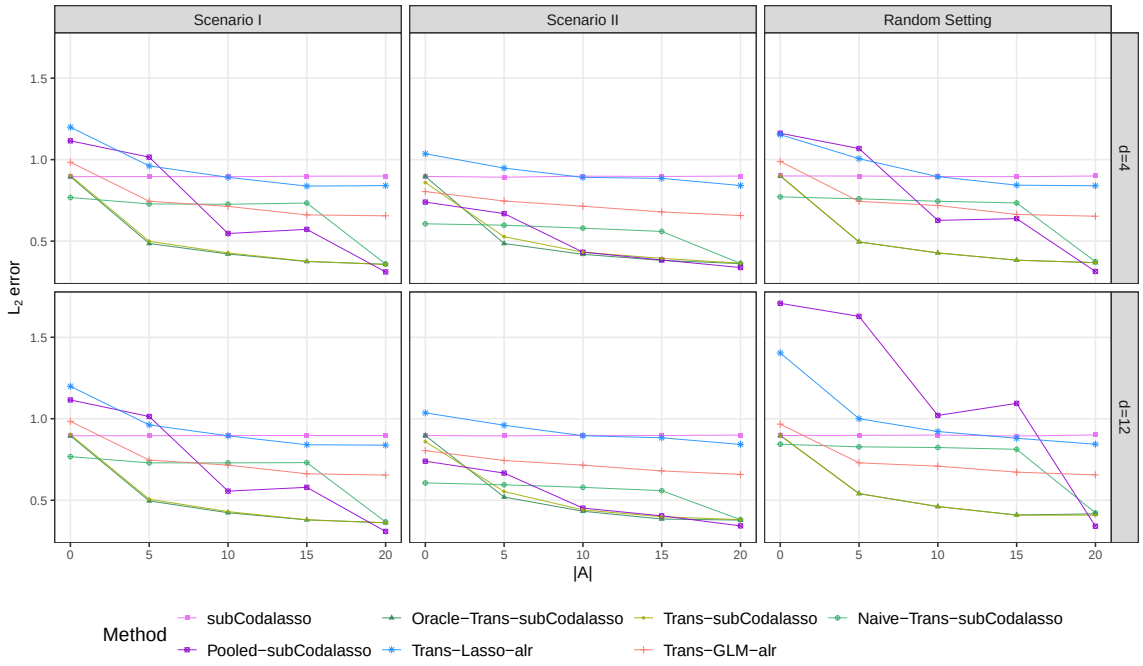

Figure 7: The average L_2-error of the seven methods under different settings for coefficient vector with homogeneous covariance matrices ($\rho = 0.5$, $(q', q) = (10, 50)$).

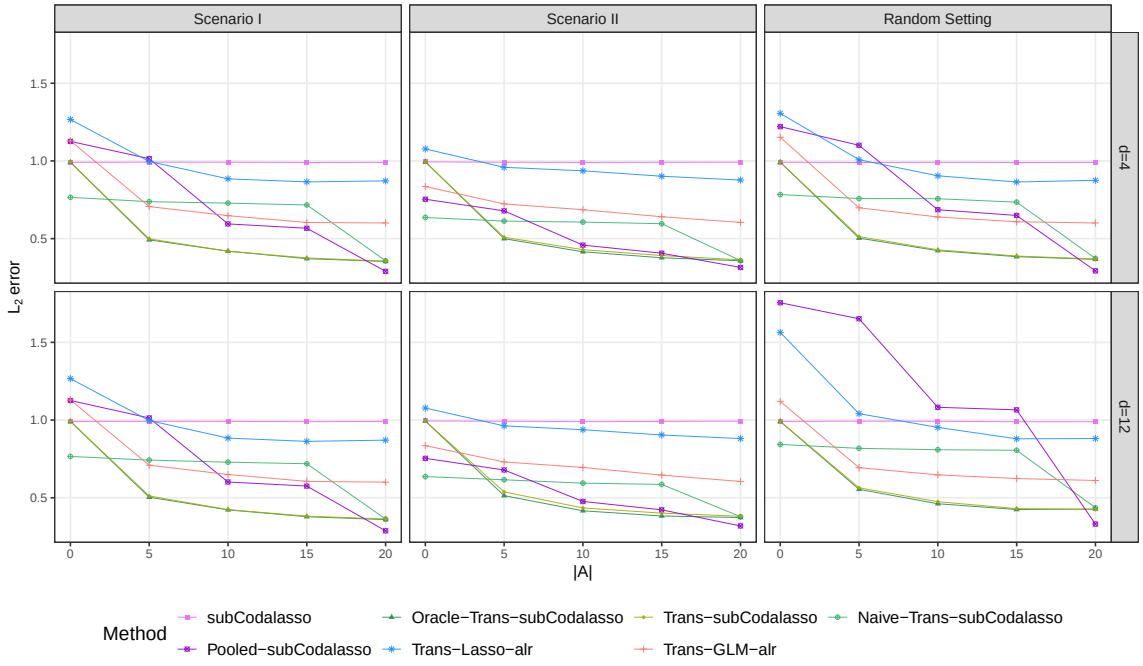

Figure 8: The average L_2-error of the seven methods under different settings for coefficient vector with heterogenous covariance matrices $(q', q) = (10, 50)$.

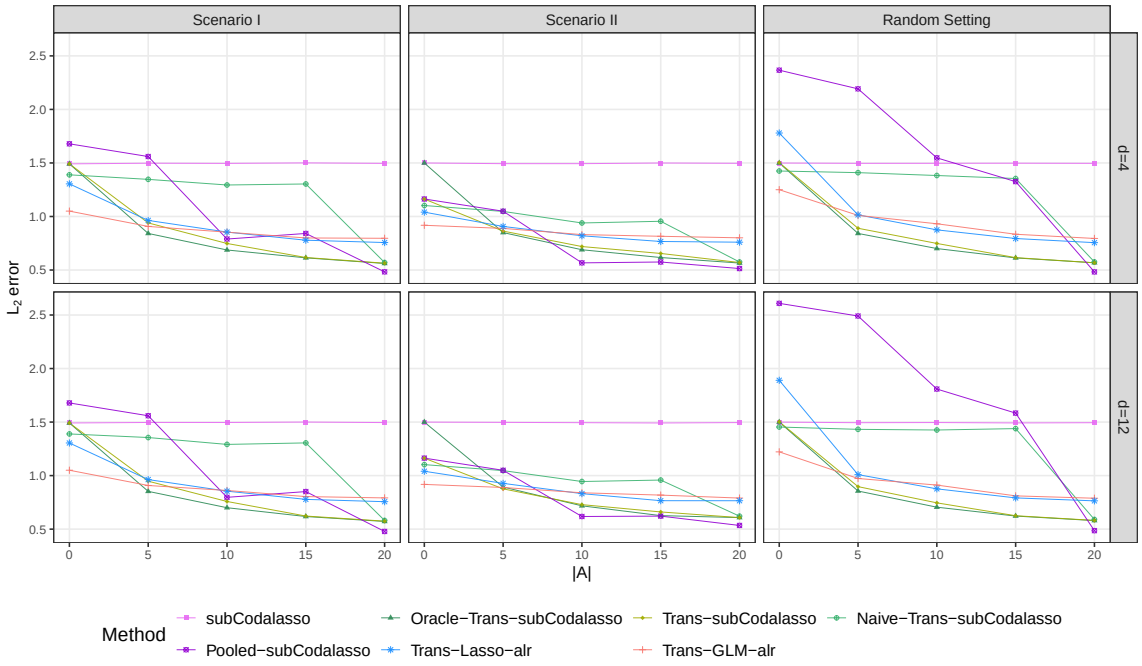

Figure 9: The average L_2-error of the seven methods under different settings for coefficient vector with homogeneous covariance matrices ($\rho = 0.2$, $(q', q) = (10, 200)$).

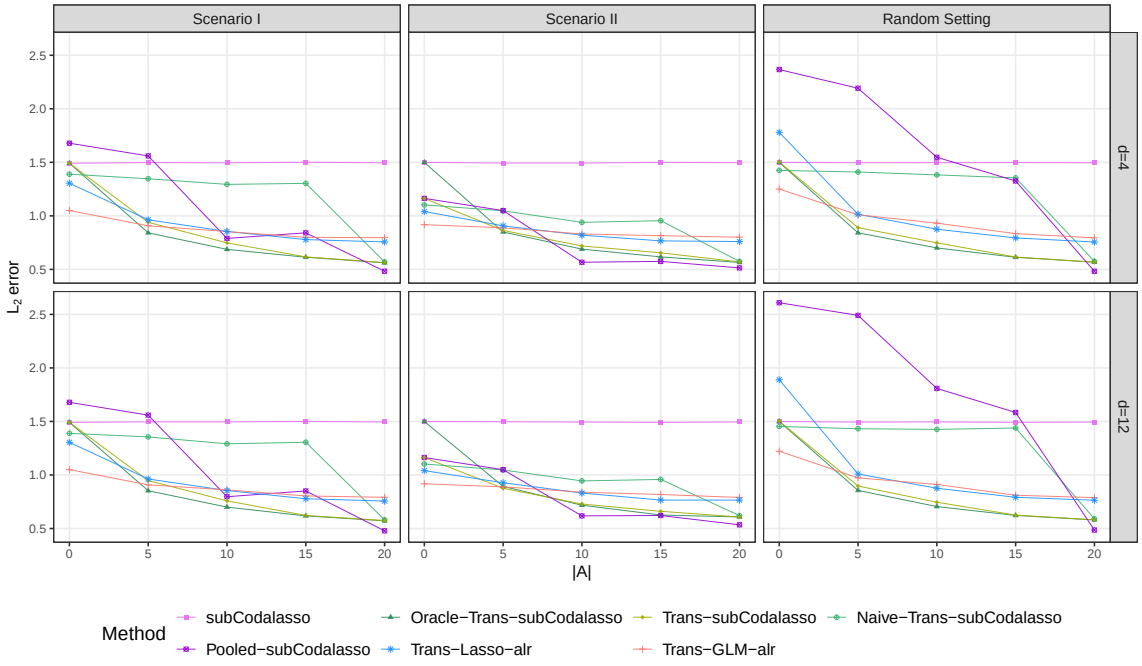

Figure 10: The average L_2-error of the seven methods under different settings for coefficient vector with homogeneous covariance matrices ($\rho = 0.5$, $(q', q) = (10, 200)$).

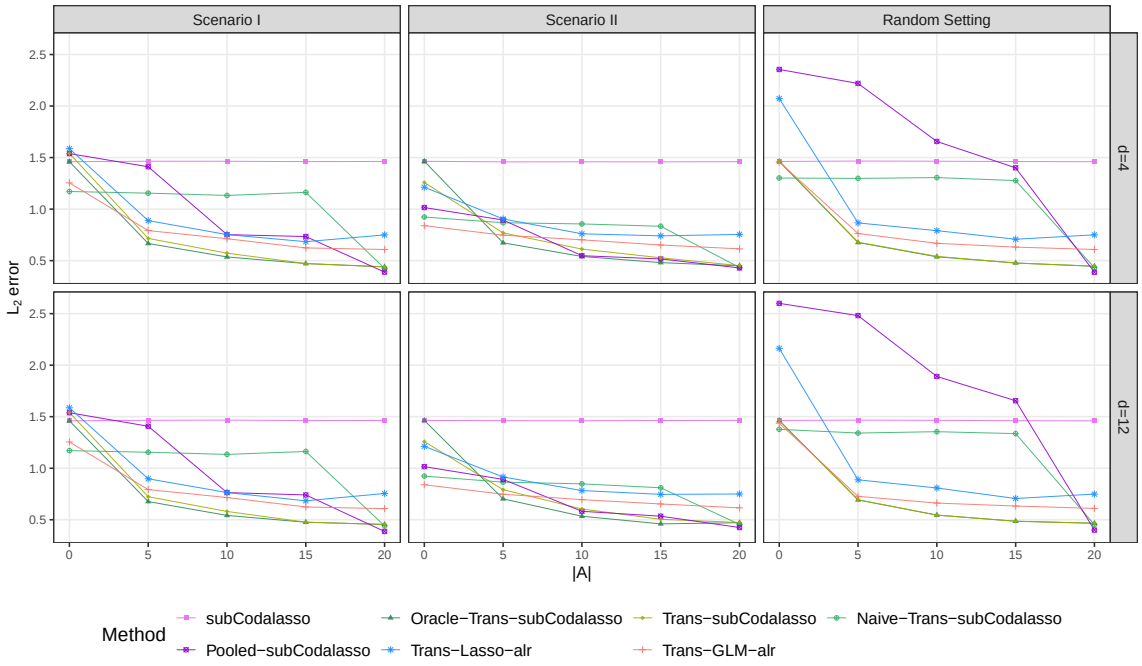

Figure 11: The average L_2-error of the seven methods under different settings for coefficient vector with heterogenous covariance matrices $(q', q) = (10, 200)$.

Next, we turn our attention to the performance of prediction error. We first focus on the average prediction error (PE) of seven methods under the random setting of coefficient vector with $(q', q) = (10, 100)$ based on 100 replications, as summarized in Figure 12. Under this scenario, Trans-Lasso-alr and the pooled method produce larger prediction errors than subCodalasso, which utilizes only the target domain, with the only exception of the pooled method when all source domains are informative. In contrast, the proposed (Oracle) Trans-subCodalasso exhibits consistently favorable performance and strong robustness across diverse covariate shift settings and perturbation magnitudes. Notably, under the heterogeneous design, Trans-GLM-alr achieves prediction performance comparable to that of our method, yet it yields substantially higher ℓ_2 estimation errors under identical conditions.

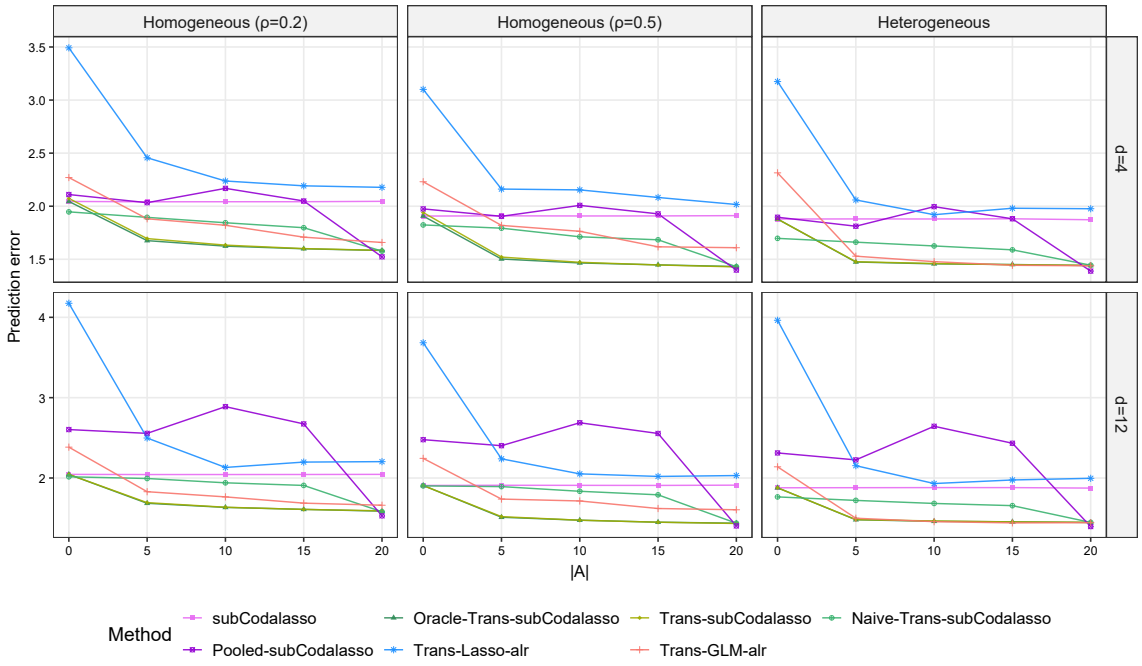

Figure 12: The average prediction error of seven methods under random setting for coefficient vector and $(q', q) = (10, 100)$ based on 100 replicates.

Tables 7-10 show the prediction errors for the cases $(q', q) = (0, 100)$,$(q', q) = (10, 100)$, $(10, 50)$ and $(10, 200)$. It can be observed from the tables that (Oracle) Trans-sub-Codalasso outperforms other methods in almost all settings.

Tables 11-17 show the true positive rate (TPR) and false discovery rate (FDR) for the cases $(q', q) = (0, 100)$,$(q', q) = (10, 100)$, $(10, 50)$ and $(10, 200)$. The results under various settings are essentially consistent with the analysis in the main text.

S5.2.2 Effect of compositional covariates constraints and non-compositional covariate exclusion

In this part,we take the dimension of non-compositional data $q' = 10$ and the dimension of compositional data $q = 100$. We set $K = 10$ and vary the number of informative sources as $|\mathcal{A}| \in \{0, 2, 4, 6, 8, 10\}$. We consider model shift in random setting ($d = 4$) and covariate shift in Homogenous design with $\rho = 0.5$ and Heterogenous design.

Table 7: Means and standard errors (in parentheses) of prediction error for seven methods with $(q', q) = (0, 100)$ based on 100 replicates.

Setting	d	Model	Scenario I			Scenario II			Random setting		
			$\|\mathcal{A}\| = 5$	$\|\mathcal{A}\| = 10$	$\|\mathcal{A}\| = 15$	$\|\mathcal{A}\| = 5$	$\|\mathcal{A}\| = 10$	$\|\mathcal{A}\| = 15$	$\|\mathcal{A}\| = 5$	$\|\mathcal{A}\| = 10$	$\|\mathcal{A}\| = 15$
Homogeneous ($\rho = 0.2$)	d=4	sC	1.89 (0.49)	1.88 (0.49)	1.89 (0.49)	1.89 (0.49)	1.89 (0.49)	1.88 (0.49)	1.88 (0.68)	1.88 (0.68)	1.88 (0.68)
		OTC	1.36 (0.26)	1.29 (0.24)	1.27 (0.24)	1.35 (0.26)	1.29 (0.24)	1.27 (0.24)	1.35 (0.25)	1.28 (0.24)	1.26 (0.23)
		TC	1.40 (0.29)	1.33 (0.27)	1.28 (0.24)	1.36 (0.27)	1.34 (0.27)	1.30 (0.25)	1.39 (0.28)	1.31 (0.28)	1.27 (0.24)
		NTC	1.72 (0.42)	1.76 (0.43)	1.73 (0.44)	1.51 (0.32)	1.58 (0.34)	1.55 (0.33)	1.72 (0.39)	1.75 (0.45)	1.70 (0.42)
		PC	2.33 (0.52)	2.66 (0.52)	2.59 (0.47)	1.65 (0.36)	1.78 (0.37)	1.69 (0.33)	2.37 (0.59)	2.60 (0.61)	2.37 (0.53)
		TL-a	1.67 (0.56)	1.54 (0.43)	1.51 (0.42)	1.65 (0.61)	1.56 (0.56)	1.55 (0.55)	1.77 (1.13)	1.59 (0.54)	1.59 (0.65)
		TG-a	1.42 (0.29)	1.38 (0.26)	1.32 (0.25)	1.41 (0.27)	1.37 (0.28)	1.34 (0.25)	1.44 (0.27)	1.39 (0.25)	1.33 (0.22)
	d=12	sC	1.89 (0.49)	1.89 (0.49)	1.89 (0.49)	1.89 (0.49)	1.88 (0.49)	1.89 (0.49)	1.88 (0.68)	1.87 (0.68)	1.88 (0.68)
		OTC	1.37 (0.27)	1.28 (0.24)	1.27 (0.23)	1.37 (0.26)	1.26 (0.24)	1.25 (0.24)	1.37 (0.26)	1.28 (0.24)	1.26 (0.23)
		TC	1.40 (0.29)	1.32 (0.27)	1.28 (0.24)	1.38 (0.27)	1.31 (0.26)	1.29 (0.25)	1.41 (0.34)	1.30 (0.29)	1.27 (0.25)
		NTC	1.72 (0.42)	1.76 (0.43)	1.73 (0.44)	1.51 (0.32)	1.56 (0.33)	1.53 (0.31)	1.85 (0.51)	1.82 (0.52)	1.79 (0.53)
		PC	2.33 (0.51)	2.66 (0.52)	2.57 (0.47)	1.65 (0.36)	1.74 (0.36)	1.65 (0.33)	3.54 (0.97)	3.89 (1.04)	3.52 (0.84)
		TL-a	1.68 (0.57)	1.55 (0.43)	1.50 (0.42)	1.70 (0.67)	1.59 (0.62)	1.57 (0.66)	1.72 (0.78)	1.68 (1.34)	1.60 (0.65)
		TG-a	1.43 (0.29)	1.38 (0.26)	1.32 (0.24)	1.42 (0.28)	1.38 (0.28)	1.34 (0.25)	1.38 (0.24)	1.35 (0.23)	1.33 (0.22)
Homogeneous ($\rho = 0.5$)	d=4	sC	1.78 (0.60)	1.78 (0.60)	1.78 (0.60)	1.77 (0.60)	1.78 (0.60)	1.78 (0.60)	1.78 (0.60)	1.78 (0.60)	1.78 (0.60)
		OTC	1.21 (0.26)	1.16 (0.24)	1.14 (0.24)	1.21 (0.26)	1.16 (0.24)	1.14 (0.24)	1.22 (0.27)	1.15 (0.24)	1.14 (0.24)
		TC	1.26 (0.27)	1.19 (0.25)	1.15 (0.24)	1.23 (0.27)	1.20 (0.26)	1.16 (0.24)	1.26 (0.29)	1.18 (0.25)	1.15 (0.24)
		NTC	1.57 (0.41)	1.60 (0.50)	1.57 (0.49)	1.36 (0.30)	1.41 (0.33)	1.38 (0.31)	1.61 (0.45)	1.64 (0.54)	1.57 (0.50)
		PC	2.11 (0.55)	2.42 (0.61)	2.37 (0.51)	1.46 (0.32)	1.61 (0.33)	1.54 (0.30)	2.20 (0.59)	2.42 (0.59)	2.24 (0.50)
		TL-a	1.46 (0.44)	1.36 (0.38)	1.30 (0.32)	1.38 (0.39)	1.40 (0.50)	1.34 (0.44)	1.42 (0.33)	1.35 (0.36)	1.30 (0.35)
		TG-a	1.36 (0.25)	1.31 (0.26)	1.26 (0.27)	1.35 (0.28)	1.30 (0.28)	1.27 (0.27)	1.37 (0.30)	1.31 (0.27)	1.26 (0.26)
	d=12	sC	1.78 (0.60)	1.78 (0.60)	1.78 (0.60)	1.77 (0.60)	1.78 (0.60)	1.78 (0.60)	1.78 (0.60)	1.78 (0.60)	1.78 (0.60)
		OTC	1.22 (0.27)	1.15 (0.24)	1.13 (0.24)	1.23 (0.27)	1.13 (0.23)	1.13 (0.24)	1.23 (0.27)	1.15 (0.24)	1.13 (0.24)
		TC	1.26 (0.28)	1.18 (0.25)	1.14 (0.24)	1.25 (0.26)	1.18 (0.25)	1.15 (0.24)	1.24 (0.27)	1.16 (0.24)	1.13 (0.24)
		NTC	1.57 (0.40)	1.60 (0.50)	1.57 (0.49)	1.37 (0.31)	1.39 (0.32)	1.37 (0.30)	1.77 (0.58)	1.76 (0.63)	1.71 (0.60)
		PC	2.11 (0.55)	2.42 (0.60)	2.36 (0.51)	1.47 (0.32)	1.57 (0.33)	1.51 (0.30)	3.27 (0.98)	3.66 (0.96)	3.40 (0.78)
		TL-a	1.46 (0.45)	1.37 (0.39)	1.30 (0.34)	1.42 (0.40)	1.41 (0.53)	1.35 (0.45)	1.43 (0.36)	1.36 (0.35)	1.32 (0.36)
		TG-a	1.36 (0.25)	1.31 (0.26)	1.25 (0.26)	1.36 (0.28)	1.30 (0.27)	1.27 (0.26)	1.32 (0.27)	1.29 (0.26)	1.26 (0.25)
Heterogeneous	d=4	sC	1.68 (0.52)	1.68 (0.53)	1.68 (0.53)	1.68 (0.53)	1.68 (0.53)	1.68 (0.53)	1.68 (0.53)	1.68 (0.53)	1.68 (0.53)
		OTC	1.25 (0.24)	1.20 (0.23)	1.18 (0.22)	1.24 (0.24)	1.20 (0.23)	1.19 (0.22)	1.24 (0.24)	1.20 (0.23)	1.18 (0.23)
		TC	1.27 (0.27)	1.22 (0.24)	1.19 (0.23)	1.25 (0.24)	1.23 (0.24)	1.21 (0.24)	1.25 (0.24)	1.21 (0.23)	1.19 (0.23)
		NTC	1.48 (0.39)	1.52 (0.40)	1.50 (0.41)	1.38 (0.27)	1.42 (0.30)	1.40 (0.30)	1.55 (0.47)	1.53 (0.43)	1.49 (0.38)
		PC	2.46 (0.56)	2.73 (0.62)	2.61 (0.50)	1.59 (0.30)	1.71 (0.33)	1.64 (0.30)	2.85 (0.85)	2.86 (0.85)	2.47 (0.60)
		TL-a	1.39 (0.48)	1.36 (0.50)	1.38 (0.51)	1.44 (0.52)	1.38 (0.48)	1.36 (0.51)	1.46 (0.46)	1.36 (0.47)	1.37 (0.49)
		TG-a	1.23 (0.23)	1.20 (0.25)	1.16 (0.23)	1.24 (0.23)	1.21 (0.23)	1.17 (0.23)	1.19 (0.22)	1.17 (0.22)	1.16 (0.23)
	d=12	sC	1.68 (0.52)	1.68 (0.53)	1.68 (0.53)	1.68 (0.53)	1.68 (0.53)	1.68 (0.53)	1.68 (0.53)	1.68 (0.53)	1.68 (0.53)
		OTC	1.25 (0.24)	1.20 (0.23)	1.19 (0.23)	1.25 (0.24)	1.18 (0.23)	1.18 (0.22)	1.25 (0.25)	1.20 (0.23)	1.19 (0.23)
		TC	1.27 (0.26)	1.21 (0.24)	1.19 (0.23)	1.27 (0.24)	1.22 (0.24)	1.20 (0.24)	1.25 (0.26)	1.20 (0.24)	1.19 (0.24)
		NTC	1.48 (0.38)	1.52 (0.40)	1.50 (0.40)	1.38 (0.27)	1.41 (0.29)	1.38 (0.28)	1.62 (0.59)	1.60 (0.53)	1.55 (0.48)
		PC	2.46 (0.56)	2.73 (0.62)	2.60 (0.50)	1.60 (0.30)	1.67 (0.32)	1.59 (0.27)	4.40 (1.37)	4.36 (1.29)	3.71 (0.99)
		TL-a	1.39 (0.48)	1.36 (0.48)	1.38 (0.51)	1.46 (0.51)	1.38 (0.48)	1.36 (0.48)	1.48 (0.49)	1.36 (0.44)	1.38 (0.50)
		TG-a	1.23 (0.23)	1.20 (0.25)	1.16 (0.23)	1.24 (0.22)	1.21 (0.22)	1.18 (0.23)	1.19 (0.23)	1.17 (0.23)	1.16 (0.23)

Note: sC: subCodalasso; OTC: Oracle-Trans-subCodalasso; TC: Trans-subCodalasso; NTC: Naive-Trans-subCodalasso; PC: Pooled-subCodalasso; TL-a: Trans-Lasso-alr; TG-a: Trans-GLM-alr.

Table 8: Means and standard errors (in parentheses) of prediction error for seven methods with $(q', q) = (10, 100)$ based on 100 replicates.

Setting	d	Model	Scenario I			Scenario II			Random setting		
			$\lvert\mathcal{A}\rvert = 5$	$\lvert\mathcal{A}\rvert = 10$	$\lvert\mathcal{A}\rvert = 15$	$\lvert\mathcal{A}\rvert = 5$	$\lvert\mathcal{A}\rvert = 10$	$\lvert\mathcal{A}\rvert = 15$	$\lvert\mathcal{A}\rvert = 5$	$\lvert\mathcal{A}\rvert = 10$	$\lvert\mathcal{A}\rvert = 15$
Homogeneous ($\rho = 0.2$)	d=4	sC	2.04 (0.35)	2.04 (0.35)	2.04 (0.35)	2.04 (0.34)	2.04 (0.34)	2.04 (0.35)	2.04 (0.35)	2.04 (0.34)	2.04 (0.35)
		OTC	1.68 (0.26)	1.62 (0.25)	1.60 (0.25)	1.67 (0.25)	1.62 (0.25)	1.60 (0.24)	1.68 (0.26)	1.62 (0.25)	1.60 (0.24)
		TC	1.70 (0.27)	1.64 (0.25)	1.60 (0.25)	1.70 (0.28)	1.64 (0.26)	1.61 (0.25)	1.69 (0.28)	1.63 (0.26)	1.60 (0.24)
		NTC	1.85 (0.30)	1.75 (0.27)	1.75 (0.27)	1.74 (0.27)	1.69 (0.26)	1.65 (0.25)	1.89 (0.31)	1.84 (0.31)	1.80 (0.29)
		PC	1.90 (0.31)	2.21 (0.37)	2.20 (0.37)	1.69 (0.26)	1.81 (0.30)	1.73 (0.28)	2.03 (0.42)	2.17 (0.43)	2.05 (0.37)
		TL-a	2.23 (0.70)	2.11 (0.61)	2.21 (0.70)	2.23 (0.67)	2.18 (0.82)	2.20 (0.81)	2.46 (0.83)	2.24 (0.58)	2.19 (0.71)
		TG-a	1.84 (0.31)	1.79 (0.28)	1.70 (0.26)	1.80 (0.28)	1.75 (0.28)	1.70 (0.26)	1.88 (0.32)	1.82 (0.30)	1.71 (0.26)
	d=12	sC	2.04 (0.35)	2.04 (0.35)	2.04 (0.35)	2.04 (0.34)	2.04 (0.34)	2.04 (0.35)	2.04 (0.35)	2.04 (0.34)	2.04 (0.35)
		OTC	1.67 (0.26)	1.63 (0.25)	1.60 (0.25)	1.69 (0.26)	1.63 (0.25)	1.61 (0.25)	1.69 (0.26)	1.63 (0.25)	1.61 (0.25)
		TC	1.70 (0.27)	1.64 (0.25)	1.60 (0.25)	1.71 (0.28)	1.65 (0.26)	1.62 (0.25)	1.69 (0.26)	1.64 (0.25)	1.61 (0.24)
		NTC	1.85 (0.30)	1.76 (0.27)	1.75 (0.27)	1.74 (0.27)	1.69 (0.26)	1.65 (0.25)	1.99 (0.33)	1.94 (0.32)	1.91 (0.32)
		PC	1.90 (0.32)	2.20 (0.37)	2.18 (0.37)	1.69 (0.26)	1.75 (0.29)	1.70 (0.27)	2.56 (0.63)	2.89 (0.64)	2.67 (0.54)
		TL-a	2.25 (0.73)	2.12 (0.61)	2.23 (0.71)	2.25 (0.67)	2.16 (0.82)	2.16 (0.79)	2.50 (0.86)	2.13 (0.63)	2.20 (0.73)
		TG-a	1.84 (0.30)	1.79 (0.28)	1.70 (0.26)	1.81 (0.28)	1.75 (0.27)	1.70 (0.26)	1.83 (0.32)	1.76 (0.29)	1.69 (0.26)
Homogeneous ($\rho = 0.5$)	d=4	sC	1.91 (0.39)	1.91 (0.39)	1.91 (0.39)	1.91 (0.39)	1.91 (0.39)	1.91 (0.39)	1.91 (0.39)	1.91 (0.39)	1.91 (0.39)
		OTC	1.50 (0.27)	1.46 (0.27)	1.45 (0.27)	1.51 (0.27)	1.47 (0.27)	1.45 (0.27)	1.50 (0.27)	1.47 (0.27)	1.45 (0.26)
		TC	1.53 (0.28)	1.48 (0.27)	1.44 (0.26)	1.54 (0.28)	1.48 (0.28)	1.45 (0.27)	1.52 (0.28)	1.47 (0.27)	1.45 (0.26)
		NTC	1.71 (0.34)	1.60 (0.30)	1.58 (0.29)	1.58 (0.31)	1.54 (0.30)	1.50 (0.28)	1.79 (0.35)	1.71 (0.32)	1.68 (0.32)
		PC	1.76 (0.33)	2.08 (0.36)	2.06 (0.35)	1.57 (0.30)	1.68 (0.32)	1.60 (0.30)	1.90 (0.37)	2.01 (0.37)	1.93 (0.35)
		TL-a	2.07 (0.69)	2.00 (0.67)	2.08 (0.69)	2.01 (0.63)	2.01 (0.74)	2.05 (0.60)	2.16 (0.62)	2.15 (0.85)	2.08 (0.69)
		TG-a	1.74 (0.34)	1.68 (0.31)	1.61 (0.30)	1.71 (0.31)	1.66 (0.30)	1.62 (0.30)	1.82 (0.35)	1.76 (0.33)	1.62 (0.31)
	d=12	sC	1.91 (0.39)	1.91 (0.39)	1.91 (0.39)	1.91 (0.39)	1.91 (0.39)	1.91 (0.39)	1.91 (0.39)	1.91 (0.39)	1.91 (0.39)
		OTC	1.50 (0.27)	1.46 (0.27)	1.45 (0.27)	1.53 (0.28)	1.47 (0.27)	1.45 (0.27)	1.51 (0.27)	1.47 (0.27)	1.45 (0.27)
		TC	1.53 (0.28)	1.48 (0.27)	1.44 (0.26)	1.55 (0.29)	1.49 (0.28)	1.46 (0.27)	1.52 (0.27)	1.47 (0.27)	1.45 (0.27)
		NTC	1.71 (0.34)	1.60 (0.30)	1.58 (0.29)	1.58 (0.31)	1.54 (0.29)	1.51 (0.28)	1.89 (0.39)	1.83 (0.36)	1.79 (0.37)
		PC	1.76 (0.33)	2.07 (0.36)	2.05 (0.35)	1.57 (0.30)	1.62 (0.31)	1.57 (0.29)	2.40 (0.55)	2.69 (0.53)	2.55 (0.47)
		TL-a	2.09 (0.76)	1.99 (0.64)	2.09 (0.69)	2.07 (0.64)	2.01 (0.81)	2.01 (0.55)	2.24 (0.69)	2.05 (0.58)	2.02 (0.63)
		TG-a	1.75 (0.34)	1.68 (0.31)	1.61 (0.30)	1.71 (0.31)	1.67 (0.30)	1.63 (0.30)	1.74 (0.35)	1.71 (0.33)	1.62 (0.31)
Heterogeneous	d=4	sC	1.88 (0.41)	1.88 (0.41)	1.88 (0.41)	1.88 (0.41)	1.88 (0.41)	1.88 (0.41)	1.88 (0.41)	1.88 (0.41)	1.88 (0.40)
		OTC	1.47 (0.32)	1.46 (0.32)	1.45 (0.31)	1.47 (0.32)	1.46 (0.32)	1.45 (0.32)	1.47 (0.32)	1.46 (0.32)	1.45 (0.33)
		TC	1.49 (0.32)	1.47 (0.33)	1.45 (0.32)	1.50 (0.30)	1.47 (0.32)	1.45 (0.32)	1.48 (0.31)	1.46 (0.32)	1.45 (0.32)
		NTC	1.62 (0.34)	1.59 (0.35)	1.57 (0.35)	1.54 (0.32)	1.53 (0.32)	1.50 (0.33)	1.66 (0.35)	1.63 (0.35)	1.59 (0.34)
		PC	1.70 (0.37)	2.11 (0.45)	2.06 (0.42)	1.52 (0.30)	1.69 (0.35)	1.61 (0.34)	1.81 (0.40)	2.00 (0.44)	1.88 (0.39)
		TL-a	2.02 (0.74)	1.88 (0.72)	2.01 (0.80)	2.04 (0.87)	2.00 (0.90)	1.95 (0.69)	2.06 (0.64)	1.92 (0.62)	1.98 (0.73)
		TG-a	1.56 (0.33)	1.50 (0.33)	1.46 (0.33)	1.53 (0.32)	1.51 (0.33)	1.48 (0.32)	1.53 (0.35)	1.48 (0.33)	1.44 (0.32)
	d=12	sC	1.88 (0.41)	1.88 (0.41)	1.88 (0.41)	1.88 (0.41)	1.88 (0.41)	1.88 (0.41)	1.88 (0.41)	1.88 (0.41)	1.88 (0.41)
		OTC	1.47 (0.32)	1.46 (0.32)	1.45 (0.31)	1.48 (0.32)	1.46 (0.33)	1.45 (0.31)	1.48 (0.32)	1.46 (0.32)	1.45 (0.33)
		TC	1.49 (0.32)	1.47 (0.33)	1.45 (0.32)	1.50 (0.30)	1.48 (0.33)	1.45 (0.32)	1.48 (0.31)	1.46 (0.32)	1.45 (0.32)
		NTC	1.62 (0.34)	1.59 (0.35)	1.56 (0.35)	1.54 (0.32)	1.52 (0.32)	1.50 (0.33)	1.72 (0.37)	1.68 (0.35)	1.66 (0.35)
		PC	1.70 (0.37)	2.10 (0.45)	2.05 (0.41)	1.52 (0.30)	1.63 (0.34)	1.57 (0.33)	2.23 (0.53)	2.64 (0.60)	2.43 (0.51)
		TL-a	2.03 (0.74)	1.89 (0.74)	1.99 (0.80)	2.03 (0.88)	1.97 (0.92)	1.93 (0.73)	2.15 (0.89)	1.93 (0.74)	1.98 (0.78)
		TG-a	1.56 (0.32)	1.51 (0.33)	1.46 (0.32)	1.52 (0.32)	1.51 (0.34)	1.48 (0.32)	1.50 (0.33)	1.46 (0.32)	1.44 (0.31)

Note: sC: subCodalasso; OTC: Oracle-Trans-subCodalasso; TC: Trans-subCodalasso; NTC: Naive-Trans-subCodalasso; PC: Pooled-subCodalasso; TL-a: Trans-Lasso-alr; TG-a: Trans-GLM-alr.

Table 9: Means and standard errors (in parentheses) of prediction error for seven methods with $(q', q) = (10, 50)$ based on 100 replicates.

Setting	d	Model	Scenario I			Scenario II			Random setting		
			$\lvert\mathcal{A}\rvert = 5$	$\lvert\mathcal{A}\rvert = 10$	$\lvert\mathcal{A}\rvert = 15$	$\lvert\mathcal{A}\rvert = 5$	$\lvert\mathcal{A}\rvert = 10$	$\lvert\mathcal{A}\rvert = 15$	$\lvert\mathcal{A}\rvert = 5$	$\lvert\mathcal{A}\rvert = 10$	$\lvert\mathcal{A}\rvert = 15$
Homogeneous ($\rho = 0.2$)	d=4	sC	1.56 (0.27)	1.56 (0.26)	1.56 (0.26)	1.56 (0.26)	1.56 (0.26)	1.56 (0.26)	1.56 (0.26)	1.56 (0.26)	1.56 (0.26)
		OTC	1.33 (0.21)	1.30 (0.20)	1.28 (0.20)	1.33 (0.21)	1.30 (0.20)	1.28 (0.20)	1.33 (0.21)	1.30 (0.21)	1.28 (0.20)
		TC	1.34 (0.21)	1.30 (0.20)	1.28 (0.20)	1.35 (0.20)	1.31 (0.20)	1.29 (0.20)	1.33 (0.21)	1.30 (0.21)	1.28 (0.20)
		NTC	1.44 (0.23)	1.39 (0.20)	1.39 (0.21)	1.38 (0.22)	1.36 (0.21)	1.33 (0.21)	1.45 (0.23)	1.41 (0.22)	1.41 (0.21)
		PC	1.56 (0.26)	1.75 (0.29)	1.74 (0.28)	1.39 (0.22)	1.46 (0.24)	1.40 (0.23)	1.59 (0.29)	1.72 (0.30)	1.71 (0.29)
		TL-a	1.91 (0.85)	1.86 (0.90)	1.84 (0.82)	1.94 (0.69)	1.84 (0.81)	1.87 (0.72)	2.03 (1.00)	1.84 (0.71)	1.82 (0.73)
		TG-a	1.44 (0.23)	1.41 (0.24)	1.35 (0.22)	1.45 (0.25)	1.41 (0.23)	1.38 (0.23)	1.45 (0.25)	1.40 (0.22)	1.35 (0.22)
	d=12	sC	1.56 (0.27)	1.56 (0.26)	1.56 (0.26)	1.56 (0.27)	1.56 (0.26)	1.56 (0.26)	1.56 (0.26)	1.56 (0.26)	1.56 (0.26)
		OTC	1.33 (0.21)	1.30 (0.21)	1.28 (0.20)	1.34 (0.21)	1.30 (0.21)	1.29 (0.20)	1.35 (0.22)	1.32 (0.21)	1.29 (0.21)
		TC	1.34 (0.21)	1.30 (0.21)	1.28 (0.20)	1.36 (0.21)	1.31 (0.20)	1.29 (0.20)	1.35 (0.22)	1.32 (0.21)	1.30 (0.21)
		NTC	1.44 (0.23)	1.39 (0.21)	1.39 (0.21)	1.38 (0.22)	1.36 (0.21)	1.33 (0.21)	1.52 (0.25)	1.49 (0.24)	1.49 (0.23)
		PC	1.56 (0.26)	1.75 (0.29)	1.73 (0.28)	1.39 (0.22)	1.42 (0.23)	1.38 (0.23)	2.31 (0.53)	2.76 (0.54)	2.75 (0.51)
		TL-a	1.91 (0.84)	1.86 (0.89)	1.82 (0.74)	1.94 (0.70)	1.81 (0.73)	1.82 (0.60)	2.01 (0.81)	1.84 (0.70)	1.82 (0.70)
		TG-a	1.45 (0.24)	1.41 (0.24)	1.35 (0.22)	1.45 (0.25)	1.41 (0.23)	1.38 (0.23)	1.43 (0.24)	1.40 (0.23)	1.36 (0.22)
Homogeneous ($\rho = 0.5$)	d=4	sC	1.45 (0.27)	1.45 (0.27)	1.45 (0.27)	1.45 (0.27)	1.45 (0.27)	1.45 (0.27)	1.45 (0.27)	1.45 (0.27)	1.45 (0.27)
		OTC	1.21 (0.21)	1.18 (0.20)	1.16 (0.20)	1.21 (0.21)	1.18 (0.20)	1.17 (0.21)	1.21 (0.21)	1.18 (0.20)	1.16 (0.21)
		TC	1.22 (0.21)	1.18 (0.21)	1.16 (0.21)	1.24 (0.21)	1.19 (0.21)	1.17 (0.21)	1.22 (0.21)	1.18 (0.21)	1.16 (0.21)
		NTC	1.34 (0.24)	1.27 (0.21)	1.27 (0.22)	1.26 (0.22)	1.23 (0.20)	1.20 (0.20)	1.35 (0.24)	1.29 (0.22)	1.28 (0.23)
		PC	1.43 (0.26)	1.63 (0.27)	1.60 (0.26)	1.28 (0.21)	1.34 (0.22)	1.28 (0.21)	1.44 (0.26)	1.58 (0.26)	1.56 (0.26)
		TL-a	1.71 (0.44)	1.84 (1.12)	1.73 (0.62)	1.71 (0.70)	1.76 (0.92)	1.71 (0.56)	1.82 (0.92)	1.67 (0.50)	1.67 (0.52)
		TG-a	1.37 (0.25)	1.32 (0.23)	1.27 (0.23)	1.33 (0.23)	1.31 (0.24)	1.28 (0.23)	1.38 (0.27)	1.32 (0.25)	1.27 (0.24)
	d=12	sC	1.45 (0.27)	1.45 (0.27)	1.45 (0.27)	1.45 (0.27)	1.45 (0.27)	1.45 (0.27)	1.45 (0.27)	1.45 (0.27)	1.45 (0.27)
		OTC	1.22 (0.21)	1.18 (0.20)	1.16 (0.20)	1.23 (0.21)	1.19 (0.21)	1.17 (0.21)	1.22 (0.21)	1.20 (0.21)	1.17 (0.21)
		TC	1.23 (0.21)	1.18 (0.21)	1.16 (0.21)	1.24 (0.20)	1.19 (0.20)	1.17 (0.21)	1.23 (0.21)	1.20 (0.20)	1.18 (0.21)
		NTC	1.34 (0.24)	1.27 (0.21)	1.27 (0.22)	1.26 (0.22)	1.23 (0.20)	1.21 (0.20)	1.43 (0.28)	1.38 (0.26)	1.37 (0.26)
		PC	1.43 (0.26)	1.62 (0.27)	1.60 (0.26)	1.28 (0.21)	1.30 (0.22)	1.26 (0.21)	2.09 (0.44)	2.59 (0.45)	2.54 (0.44)
		TL-a	1.72 (0.45)	1.84 (1.12)	1.72 (0.61)	1.74 (0.67)	1.75 (0.91)	1.70 (0.54)	1.88 (0.64)	1.68 (0.55)	1.73 (0.77)
		TG-a	1.37 (0.26)	1.32 (0.23)	1.27 (0.24)	1.34 (0.24)	1.31 (0.24)	1.28 (0.23)	1.34 (0.26)	1.31 (0.24)	1.27 (0.23)
Heterogeneous	d=4	sC	1.45 (0.30)	1.44 (0.30)	1.44 (0.30)	1.44 (0.30)	1.44 (0.30)	1.44 (0.30)	1.44 (0.30)	1.44 (0.30)	1.44 (0.30)
		OTC	1.18 (0.21)	1.16 (0.21)	1.15 (0.21)	1.17 (0.21)	1.16 (0.21)	1.15 (0.21)	1.17 (0.21)	1.16 (0.21)	1.15 (0.21)
		TC	1.18 (0.21)	1.16 (0.21)	1.15 (0.21)	1.18 (0.21)	1.16 (0.21)	1.15 (0.21)	1.18 (0.21)	1.16 (0.21)	1.15 (0.21)
		NTC	1.27 (0.22)	1.25 (0.23)	1.23 (0.22)	1.22 (0.21)	1.21 (0.21)	1.19 (0.20)	1.28 (0.22)	1.26 (0.22)	1.24 (0.22)
		PC	1.38 (0.25)	1.62 (0.27)	1.57 (0.23)	1.24 (0.21)	1.31 (0.21)	1.26 (0.20)	1.42 (0.24)	1.58 (0.25)	1.54 (0.23)
		TL-a	1.81 (0.83)	1.67 (0.71)	1.62 (0.48)	1.70 (0.63)	1.76 (0.84)	1.75 (0.77)	1.70 (0.70)	1.62 (0.50)	1.64 (0.58)
		TG-a	1.23 (0.24)	1.18 (0.22)	1.15 (0.22)	1.24 (0.22)	1.21 (0.22)	1.17 (0.22)	1.22 (0.24)	1.17 (0.22)	1.15 (0.22)
	d=12	sC	1.44 (0.30)	1.45 (0.30)	1.44 (0.30)	1.45 (0.30)	1.44 (0.30)	1.44 (0.30)	1.44 (0.30)	1.44 (0.30)	1.44 (0.30)
		OTC	1.18 (0.21)	1.16 (0.21)	1.15 (0.21)	1.18 (0.21)	1.16 (0.21)	1.15 (0.21)	1.19 (0.21)	1.17 (0.21)	1.16 (0.21)
		TC	1.18 (0.21)	1.16 (0.21)	1.15 (0.21)	1.19 (0.21)	1.16 (0.21)	1.15 (0.21)	1.19 (0.21)	1.17 (0.21)	1.16 (0.21)
		NTC	1.28 (0.23)	1.26 (0.23)	1.23 (0.22)	1.22 (0.21)	1.21 (0.21)	1.19 (0.20)	1.31 (0.24)	1.30 (0.23)	1.29 (0.23)
		PC	1.38 (0.25)	1.61 (0.27)	1.57 (0.24)	1.24 (0.21)	1.28 (0.21)	1.24 (0.20)	2.05 (0.38)	2.60 (0.44)	2.52 (0.39)
		TL-a	1.79 (0.79)	1.67 (0.69)	1.63 (0.53)	1.70 (0.63)	1.75 (0.85)	1.73 (0.78)	1.75 (0.67)	1.65 (0.53)	1.62 (0.56)
		TG-a	1.23 (0.24)	1.18 (0.22)	1.15 (0.22)	1.24 (0.22)	1.21 (0.22)	1.17 (0.22)	1.21 (0.23)	1.17 (0.22)	1.16 (0.22)

Note: sC: subCodalasso; OTC: Oracle-Trans-subCodalasso; TC: Trans-subCodalasso; NTC: Naive-Trans-subCodalasso; PC: Pooled-subCodalasso; TL-a: Trans-Lasso-alr; TG-a: Trans-GLM-alr.

Table 10: Means and standard errors (in parentheses) of Prediction error for seven methods with $(q', q) = (10, 200)$ based on 100 replicates.

Setting	d	Model	Scenario I			Scenario II			Random setting		
			$\lvert\mathcal{A}\rvert = 5$	$\lvert\mathcal{A}\rvert = 10$	$\lvert\mathcal{A}\rvert = 15$	$\lvert\mathcal{A}\rvert = 5$	$\lvert\mathcal{A}\rvert = 10$	$\lvert\mathcal{A}\rvert = 15$	$\lvert\mathcal{A}\rvert = 5$	$\lvert\mathcal{A}\rvert = 10$	$\lvert\mathcal{A}\rvert = 15$
Homogeneous ($\rho = 0.2$)	d=4	sC	2.59 (0.61)	2.59 (0.61)	2.59 (0.61)	2.59 (0.61)	2.59 (0.61)	2.60 (0.61)	2.59 (0.61)	2.59 (0.61)	2.59 (0.62)
		OTC	2.00 (0.42)	1.93 (0.40)	1.91 (0.40)	2.00 (0.42)	1.94 (0.40)	1.92 (0.40)	2.00 (0.42)	1.94 (0.40)	1.91 (0.40)
		TC	2.07 (0.47)	1.97 (0.43)	1.91 (0.40)	2.03 (0.42)	1.96 (0.41)	1.92 (0.41)	2.04 (0.44)	1.97 (0.41)	1.92 (0.40)
		NTC	2.24 (0.46)	2.12 (0.42)	2.12 (0.44)	2.10 (0.44)	2.04 (0.43)	2.00 (0.43)	2.42 (0.50)	2.38 (0.49)	2.28 (0.45)
		PC	2.35 (0.49)	2.84 (0.52)	2.76 (0.50)	2.03 (0.38)	2.19 (0.41)	2.09 (0.39)	2.87 (0.68)	2.84 (0.63)	2.57 (0.55)
		TL-a	2.62 (0.81)	2.48 (0.80)	2.56 (0.92)	2.62 (0.80)	2.44 (0.69)	2.45 (0.66)	2.96 (1.06)	2.59 (0.79)	2.56 (0.87)
		TG-a	2.22 (0.49)	2.15 (0.43)	2.08 (0.43)	2.22 (0.48)	2.13 (0.40)	2.11 (0.41)	2.40 (0.59)	2.28 (0.54)	2.12 (0.45)
	d=12	sC	2.59 (0.61)	2.59 (0.61)	2.59 (0.61)	2.59 (0.61)	2.59 (0.62)	2.59 (0.61)	2.59 (0.61)	2.59 (0.61)	2.59 (0.62)
		OTC	2.00 (0.42)	1.94 (0.40)	1.92 (0.40)	2.03 (0.41)	1.95 (0.41)	1.92 (0.39)	2.00 (0.42)	1.94 (0.40)	1.92 (0.40)
		TC	2.08 (0.47)	1.98 (0.43)	1.92 (0.40)	2.03 (0.42)	1.96 (0.42)	1.93 (0.40)	2.04 (0.45)	1.97 (0.43)	1.92 (0.40)
		NTC	2.25 (0.46)	2.12 (0.42)	2.12 (0.44)	2.10 (0.44)	2.04 (0.44)	2.01 (0.43)	2.46 (0.49)	2.44 (0.50)	2.39 (0.49)
		PC	2.35 (0.49)	2.82 (0.51)	2.74 (0.50)	2.03 (0.38)	2.11 (0.40)	2.05 (0.39)	3.33 (0.85)	3.39 (0.77)	3.00 (0.67)
		TL-a	2.62 (0.80)	2.46 (0.83)	2.54 (0.92)	2.66 (0.87)	2.48 (0.79)	2.45 (0.69)	3.04 (1.02)	2.60 (0.75)	2.48 (0.75)
		TG-a	2.22 (0.49)	2.16 (0.46)	2.08 (0.43)	2.22 (0.48)	2.15 (0.44)	2.11 (0.40)	2.36 (0.58)	2.24 (0.52)	2.09 (0.44)
Homogeneous ($\rho = 0.5$)	d=4	sC	2.37 (0.52)	2.38 (0.51)	2.38 (0.53)	2.38 (0.52)	2.38 (0.52)	2.38 (0.52)	2.38 (0.52)	2.38 (0.52)	2.38 (0.52)
		OTC	1.83 (0.34)	1.77 (0.33)	1.76 (0.33)	1.83 (0.34)	1.78 (0.33)	1.76 (0.33)	1.83 (0.34)	1.77 (0.33)	1.76 (0.33)
		TC	1.91 (0.38)	1.81 (0.34)	1.76 (0.32)	1.87 (0.36)	1.79 (0.35)	1.76 (0.33)	1.87 (0.37)	1.80 (0.34)	1.76 (0.33)
		NTC	2.13 (0.41)	1.98 (0.38)	1.96 (0.36)	1.95 (0.35)	1.87 (0.35)	1.83 (0.34)	2.30 (0.49)	2.20 (0.45)	2.10 (0.41)
		PC	2.17 (0.41)	2.70 (0.49)	2.62 (0.46)	1.88 (0.33)	2.05 (0.38)	1.95 (0.36)	2.58 (0.56)	2.62 (0.59)	2.36 (0.48)
		TL-a	2.58 (0.89)	2.45 (0.76)	2.49 (0.70)	2.50 (0.74)	2.41 (0.67)	2.42 (0.63)	2.71 (0.78)	2.42 (0.64)	2.42 (0.62)
		TG-a	2.15 (0.44)	2.03 (0.37)	1.97 (0.35)	2.08 (0.37)	2.01 (0.36)	1.98 (0.35)	2.36 (0.53)	2.21 (0.45)	2.05 (0.47)
	d=12	sC	2.38 (0.52)	2.38 (0.51)	2.38 (0.53)	2.38 (0.52)	2.38 (0.52)	2.38 (0.52)	2.45 (0.50)	2.44 (0.50)	2.44 (0.50)
		OTC	1.83 (0.34)	1.78 (0.33)	1.76 (0.33)	1.85 (0.34)	1.79 (0.33)	1.77 (0.33)	1.85 (0.36)	1.80 (0.36)	1.79 (0.36)
		TC	1.91 (0.38)	1.82 (0.34)	1.76 (0.32)	1.87 (0.36)	1.81 (0.34)	1.77 (0.33)	1.85 (0.36)	1.80 (0.35)	1.79 (0.36)
		NTC	2.13 (0.41)	1.98 (0.38)	1.96 (0.37)	1.95 (0.35)	1.87 (0.35)	1.83 (0.35)	2.20 (0.41)	2.18 (0.40)	2.09 (0.40)
		PC	2.17 (0.41)	2.68 (0.49)	2.60 (0.46)	1.88 (0.34)	1.98 (0.37)	1.90 (0.35)	2.86 (0.66)	3.09 (0.73)	2.79 (0.65)
		TL-a	2.60 (0.91)	2.43 (0.74)	2.47 (0.67)	2.55 (0.78)	2.38 (0.70)	2.40 (0.68)	2.60 (0.90)	2.22 (0.52)	2.20 (0.61)
		TG-a	2.15 (0.44)	2.04 (0.37)	1.97 (0.35)	2.09 (0.37)	2.01 (0.36)	1.99 (0.35)	1.86 (0.38)	1.80 (0.37)	1.78 (0.35)
Heterogeneous	d=4	sC	2.44 (0.50)	2.44 (0.50)	2.44 (0.50)	2.44 (0.50)	2.44 (0.50)	2.44 (0.50)	2.45 (0.50)	2.44 (0.50)	2.44 (0.50)
		OTC	1.84 (0.35)	1.79 (0.35)	1.78 (0.36)	1.84 (0.36)	1.80 (0.36)	1.78 (0.36)	1.84 (0.36)	1.80 (0.35)	1.78 (0.35)
		TC	1.86 (0.35)	1.81 (0.35)	1.78 (0.36)	1.87 (0.36)	1.82 (0.36)	1.79 (0.36)	1.84 (0.36)	1.79 (0.36)	1.78 (0.35)
		NTC	2.00 (0.40)	1.94 (0.37)	1.90 (0.35)	1.89 (0.37)	1.87 (0.37)	1.83 (0.36)	2.14 (0.41)	2.12 (0.39)	2.04 (0.39)
		PC	2.07 (0.44)	2.72 (0.55)	2.61 (0.50)	1.83 (0.36)	2.08 (0.43)	1.98 (0.40)	2.53 (0.55)	2.64 (0.60)	2.41 (0.52)
		TL-a	2.28 (0.58)	2.13 (0.44)	2.17 (0.53)	2.40 (0.66)	2.24 (0.54)	2.24 (0.49)	2.57 (0.90)	2.20 (0.51)	2.19 (0.60)
		TG-a	1.90 (0.36)	1.84 (0.36)	1.77 (0.34)	1.87 (0.36)	1.84 (0.35)	1.79 (0.34)	1.89 (0.39)	1.81 (0.36)	1.77 (0.34)
	d=12	sC	2.44 (0.50)	2.44 (0.50)	2.44 (0.50)	2.44 (0.50)	2.44 (0.50)	2.45 (0.50)	2.45 (0.50)	2.44 (0.50)	2.44 (0.50)
		OTC	1.84 (0.35)	1.80 (0.35)	1.78 (0.36)	1.85 (0.36)	1.80 (0.35)	1.78 (0.35)	1.85 (0.36)	1.80 (0.36)	1.79 (0.36)
		TC	1.86 (0.35)	1.81 (0.35)	1.78 (0.36)	1.89 (0.39)	1.82 (0.35)	1.79 (0.35)	1.85 (0.36)	1.80 (0.35)	1.79 (0.36)
		NTC	2.00 (0.40)	1.94 (0.37)	1.90 (0.35)	1.89 (0.37)	1.87 (0.36)	1.84 (0.36)	2.20 (0.41)	2.18 (0.40)	2.09 (0.40)
		PC	2.07 (0.44)	2.70 (0.55)	2.59 (0.50)	1.84 (0.36)	2.00 (0.42)	1.93 (0.39)	2.86 (0.66)	3.09 (0.73)	2.79 (0.65)
		TL-a	2.28 (0.57)	2.13 (0.46)	2.15 (0.51)	2.42 (0.64)	2.25 (0.56)	2.23 (0.48)	2.60 (0.90)	2.22 (0.52)	2.20 (0.61)
		TG-a	1.90 (0.36)	1.84 (0.36)	1.77 (0.34)	1.87 (0.35)	1.82 (0.35)	1.79 (0.34)	1.86 (0.38)	1.80 (0.37)	1.78 (0.35)

Note: sC: subCodalasso; OTC: Oracle-Trans-subCodalasso; TC: Trans-subCodalasso; NTC: Naive-Trans-subCodalasso; PC: Pooled-subCodalasso; TL-a: Trans-Lasso-alr; TG-a: Trans-GLM-alr.

Table 11: Means and standard errors (in parentheses) of true positive rate (TPR) and false discovery rate (FDR) for seven methods with $(q', q) = (0, 100)$ based on 100 replicates.

Type	Scenario	Method	TPR						FDR					
			$d = 4$			$d = 12$			$d = 4$			$d = 12$		
			$\lvert\mathcal{A}\rvert = 5$	$\lvert\mathcal{A}\rvert = 10$	$\lvert\mathcal{A}\rvert = 15$	$\lvert\mathcal{A}\rvert = 5$	$\lvert\mathcal{A}\rvert = 10$	$\lvert\mathcal{A}\rvert = 15$	$\lvert\mathcal{A}\rvert = 5$	$\lvert\mathcal{A}\rvert = 10$	$\lvert\mathcal{A}\rvert = 15$	$\lvert\mathcal{A}\rvert = 5$	$\lvert\mathcal{A}\rvert = 10$	$\lvert\mathcal{A}\rvert = 15$
Homogeneous ($\rho = 0.2$)	scenario I	sC	0.99(0.04)	0.99(0.04)	0.99(0.04)	0.99(0.04)	0.99(0.04)	0.99(0.04)	0.83(0.08)	0.83(0.08)	0.83(0.08)	0.83(0.08)	0.83(0.08)	0.83(0.09)
		OTC	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.60(0.22)	0.53(0.22)	0.48(0.24)	0.60(0.22)	0.55(0.22)	0.50(0.23)
		TC	1.00(0.02)	1.00(0.00)	1.00(0.00)	1.00(0.02)	1.00(0.00)	1.00(0.00)	0.63(0.19)	0.51(0.24)	0.47(0.25)	0.63(0.20)	0.53(0.23)	0.48(0.24)
		NTC	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.79(0.12)	0.78(0.14)	0.76(0.13)	0.79(0.11)	0.78(0.14)	0.76(0.13)
		PC	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.54(0.16)	0.34(0.22)	0.10(0.15)	0.54(0.16)	0.34(0.22)	0.11(0.15)
		TL-a	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.90(0.09)	0.88(0.11)	0.82(0.20)	0.90(0.10)	0.88(0.11)	0.82(0.20)
		TG-a	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.50(0.22)	0.35(0.25)	0.19(0.25)	0.50(0.22)	0.35(0.25)	0.19(0.25)
	scenario II	sC	0.99(0.04)	0.99(0.04)	0.99(0.04)	0.99(0.04)	0.99(0.04)	0.99(0.04)	0.83(0.08)	0.83(0.08)	0.83(0.08)	0.83(0.08)	0.83(0.08)	0.83(0.08)
		OTC	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.61(0.21)	0.57(0.22)	0.52(0.23)	0.66(0.19)	0.67(0.15)	0.54(0.22)
		TC	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.65(0.19)	0.54(0.23)	0.46(0.27)	0.66(0.17)	0.61(0.20)	0.51(0.24)
		NTC	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.72(0.15)	0.67(0.19)	0.65(0.21)	0.72(0.14)	0.68(0.19)	0.64(0.22)
		PC	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.54(0.14)	0.40(0.16)	0.34(0.17)	0.56(0.13)	0.39(0.15)	0.34(0.19)
		TL-a	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.88(0.13)	0.86(0.13)	0.80(0.21)	0.89(0.13)	0.87(0.11)	0.81(0.20)
		TG-a	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.51(0.18)	0.40(0.21)	0.25(0.26)	0.51(0.18)	0.40(0.23)	0.26(0.26)
	Random Setting	sC	1.00(0.02)	1.00(0.02)	1.00(0.02)	1.00(0.02)	1.00(0.02)	1.00(0.02)	0.83(0.08)	0.83(0.08)	0.83(0.08)	0.83(0.08)	0.83(0.08)	0.83(0.09)
		OTC	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.58(0.21)	0.54(0.22)	0.49(0.25)	0.61(0.20)	0.61(0.19)	0.55(0.22)
		TC	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.01)	1.00(0.00)	1.00(0.00)	0.58(0.22)	0.53(0.23)	0.47(0.25)	0.61(0.20)	0.60(0.19)	0.53(0.22)
		NTC	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.01)	1.00(0.01)	1.00(0.00)	0.82(0.07)	0.78(0.13)	0.74(0.15)	0.84(0.07)	0.80(0.10)	0.78(0.14)
		PC	0.99(0.03)	1.00(0.03)	1.00(0.01)	0.95(0.08)	0.96(0.07)	0.99(0.03)	0.63(0.16)	0.43(0.21)	0.25(0.21)	0.61(0.17)	0.42(0.21)	0.23(0.21)
		TL-a	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.89(0.12)	0.90(0.10)	0.87(0.12)	0.83(0.18)	0.87(0.14)	0.88(0.12)
		TG-a	1.00(0.02)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.43(0.25)	0.32(0.27)	0.26(0.26)	0.39(0.23)	0.30(0.25)	0.26(0.27)
Homogeneous ($\rho = 0.5$)	scenario I	sC	0.98(0.06)	0.98(0.06)	0.98(0.06)	0.98(0.06)	0.98(0.06)	0.98(0.06)	0.81(0.10)	0.82(0.09)	0.81(0.10)	0.81(0.10)	0.81(0.09)	0.81(0.10)
		OTC	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.61(0.17)	0.52(0.22)	0.50(0.24)	0.62(0.18)	0.56(0.20)	0.52(0.23)
		TC	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.63(0.17)	0.52(0.24)	0.49(0.23)	0.64(0.17)	0.54(0.22)	0.52(0.22)
		NTC	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.78(0.11)	0.75(0.14)	0.69(0.21)	0.79(0.11)	0.75(0.14)	0.69(0.21)
		PC	1.00(0.02)	1.00(0.02)	1.00(0.01)	1.00(0.02)	1.00(0.02)	1.00(0.01)	0.61(0.13)	0.44(0.18)	0.17(0.18)	0.61(0.13)	0.43(0.17)	0.17(0.18)
		TL-a	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.91(0.05)	0.89(0.08)	0.80(0.19)	0.92(0.02)	0.89(0.08)	0.81(0.20)
		TG-a	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.59(0.17)	0.43(0.23)	0.22(0.27)	0.59(0.16)	0.43(0.22)	0.22(0.27)
	scenario II	sC	0.98(0.06)	0.98(0.06)	0.98(0.06)	0.98(0.06)	0.98(0.06)	0.98(0.06)	0.81(0.09)	0.81(0.10)	0.81(0.09)	0.82(0.09)	0.81(0.10)	0.81(0.09)
		OTC	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.63(0.16)	0.54(0.21)	0.51(0.23)	0.66(0.16)	0.66(0.16)	0.57(0.21)
		TC	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.66(0.16)	0.56(0.22)	0.48(0.25)	0.66(0.17)	0.61(0.20)	0.55(0.22)
		NTC	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.74(0.11)	0.70(0.16)	0.66(0.19)	0.75(0.11)	0.69(0.16)	0.65(0.20)
		PC	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.64(0.12)	0.49(0.18)	0.42(0.19)	0.65(0.12)	0.50(0.18)	0.43(0.20)
		TL-a	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.89(0.06)	0.85(0.13)	0.76(0.23)	0.90(0.06)	0.86(0.12)	0.78(0.21)
		TG-a	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.02)	1.00(0.00)	1.00(0.00)	0.55(0.19)	0.45(0.23)	0.32(0.28)	0.58(0.18)	0.46(0.23)	0.33(0.26)
	Random Setting	sC	0.98(0.06)	0.98(0.06)	0.98(0.06)	0.98(0.06)	0.98(0.06)	0.98(0.06)	0.81(0.10)	0.82(0.09)	0.81(0.10)	0.81(0.10)	0.81(0.09)	0.81(0.10)
		OTC	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.61(0.18)	0.56(0.21)	0.52(0.23)	0.63(0.17)	0.62(0.18)	0.57(0.21)
		TC	1.00(0.01)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.61(0.18)	0.56(0.21)	0.53(0.21)	0.63(0.17)	0.61(0.18)	0.56(0.21)
		NTC	0.99(0.03)	1.00(0.01)	1.00(0.00)	0.99(0.05)	0.99(0.03)	1.00(0.02)	0.79(0.11)	0.76(0.16)	0.69(0.22)	0.83(0.08)	0.80(0.10)	0.77(0.13)
		PC	0.97(0.06)	0.98(0.05)	0.99(0.03)	0.90(0.11)	0.90(0.09)	0.92(0.07)	0.66(0.13)	0.48(0.20)	0.27(0.20)	0.65(0.15)	0.45(0.23)	0.25(0.20)
		TL-a	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.02)	1.00(0.00)	1.00(0.00)	0.88(0.11)	0.90(0.05)	0.85(0.14)	0.81(0.18)	0.90(0.07)	0.86(0.14)
		TG-a	0.99(0.05)	1.00(0.02)	1.00(0.00)	0.99(0.05)	0.99(0.03)	1.00(0.00)	0.46(0.25)	0.34(0.28)	0.23(0.28)	0.44(0.23)	0.33(0.25)	0.24(0.28)
Heterogeneous	scenario I	sC	0.98(0.06)	0.98(0.06)	0.98(0.06)	0.98(0.06)	0.98(0.06)	0.98(0.06)	0.82(0.06)	0.83(0.05)	0.83(0.05)	0.82(0.06)	0.82(0.05)	0.83(0.05)
		OTC	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.59(0.22)	0.48(0.26)	0.40(0.25)	0.58(0.22)	0.50(0.25)	0.40(0.25)
		TC	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.62(0.22)	0.48(0.27)	0.39(0.25)	0.62(0.22)	0.50(0.24)	0.41(0.24)
		NTC	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.79(0.09)	0.78(0.13)	0.75(0.16)	0.79(0.09)	0.78(0.13)	0.75(0.16)
		PC	1.00(0.02)	1.00(0.00)	1.00(0.00)	1.00(0.02)	1.00(0.00)	1.00(0.00)	0.52(0.19)	0.41(0.22)	0.17(0.20)	0.52(0.19)	0.41(0.22)	0.17(0.19)
		TL-a	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.90(0.11)	0.89(0.08)	0.83(0.18)	0.90(0.11)	0.89(0.08)	0.83(0.18)
		TG-a	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.48(0.19)	0.38(0.22)	0.26(0.24)	0.48(0.19)	0.37(0.21)	0.27(0.24)
	scenario II	sC	0.98(0.06)	0.98(0.06)	0.98(0.06)	0.98(0.06)	0.98(0.06)	0.98(0.06)	0.83(0.06)	0.82(0.06)	0.83(0.05)	0.83(0.06)	0.83(0.06)	0.83(0.05)
		OTC	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.64(0.19)	0.56(0.23)	0.46(0.26)	0.66(0.17)	0.61(0.19)	0.50(0.24)
		TC	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.70(0.16)	0.57(0.22)	0.44(0.28)	0.71(0.16)	0.61(0.19)	0.45(0.27)
		NTC	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.75(0.12)	0.75(0.13)	0.72(0.17)	0.76(0.12)	0.73(0.14)	0.72(0.17)
		PC	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.60(0.15)	0.51(0.16)	0.45(0.19)	0.61(0.13)	0.50(0.16)	0.45(0.20)
		TL-a	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.91(0.04)	0.88(0.07)	0.81(0.19)	0.91(0.04)	0.89(0.06)	0.83(0.17)
		TG-a	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.57(0.15)	0.53(0.16)	0.37(0.22)	0.58(0.14)	0.53(0.17)	0.41(0.20)
	Random Setting	sC	0.98(0.06)	0.98(0.06)	0.98(0.06)	0.98(0.06)	0.98(0.06)	0.98(0.06)	0.83(0.05)	0.83(0.05)	0.83(0.05)	0.83(0.05)	0.82(0.06)	0.83(0.05)
		OTC	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.59(0.21)	0.51(0.24)	0.41(0.25)	0.62(0.22)	0.55(0.23)	0.46(0.25)
		TC	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.60(0.22)	0.50(0.24)	0.39(0.25)	0.62(0.22)	0.56(0.22)	0.47(0.25)
		NTC	1.00(0.01)	1.00(0.00)	1.00(0.00)	1.00(0.01)	1.00(0.01)	1.00(0.00)	0.81(0.07)	0.79(0.09)	0.76(0.13)	0.81(0.07)	0.81(0.07)	0.78(0.09)
		PC	0.99(0.03)	1.00(0.03)	1.00(0.00)	0.93(0.10)	0.96(0.08)	1.00(0.03)	0.57(0.18)	0.46(0.23)	0.34(0.22)	0.54(0.21)	0.45(0.24)	0.32(0.22)
		TL-a	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.86(0.15)	0.90(0.07)	0.88(0.10)	0.78(0.25)	0.87(0.12)	0.86(0.16)
		TG-a	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.39(0.20)	0.31(0.22)	0.26(0.24)	0.40(0.20)	0.34(0.21)	0.31(0.24)

Note: sC: subCodalasso; OTC: Oracle-Trans-subCodalasso; TC: Trans-subCodalasso; NTC: Naive-Trans-subCodalasso; PC: Pooled-subCodalasso; TL-a: Trans-Lasso-alr; TG-a: Trans-GLM-alr.

Table 12: Means and standard errors (in parentheses) of true positive rate (TPR) and false discovery rate (FDR) for seven methods with $(q', q) = (10, 100)$ based on 100 replicates (All Covariates).

Type	Scenario	Method	TPR						FDR					
			$d = 4$			$d = 12$			$d = 4$			$d = 12$		
			$\lvert\mathcal{A}\rvert = 5$	$\lvert\mathcal{A}\rvert = 10$	$\lvert\mathcal{A}\rvert = 15$	$\lvert\mathcal{A}\rvert = 5$	$\lvert\mathcal{A}\rvert = 10$	$\lvert\mathcal{A}\rvert = 15$	$\lvert\mathcal{A}\rvert = 5$	$\lvert\mathcal{A}\rvert = 10$	$\lvert\mathcal{A}\rvert = 15$	$\lvert\mathcal{A}\rvert = 5$	$\lvert\mathcal{A}\rvert = 10$	$\lvert\mathcal{A}\rvert = 15$
Homogeneous ($\rho = 0.2$)	scenario I	sC	0.94(0.06)	0.94(0.07)	0.94(0.07)	0.94(0.06)	0.94(0.07)	0.94(0.07)	0.84(0.05)	0.84(0.06)	0.84(0.05)	0.84(0.05)	0.84(0.06)	0.84(0.05)
		OTC	0.91(0.04)	0.91(0.04)	0.91(0.04)	0.91(0.04)	0.91(0.04)	0.91(0.04)	0.42(0.26)	0.33(0.28)	0.25(0.28)	0.41(0.26)	0.33(0.28)	0.26(0.28)
		TC	0.91(0.05)	0.91(0.04)	0.91(0.04)	0.91(0.04)	0.91(0.04)	0.91(0.04)	0.46(0.25)	0.34(0.28)	0.26(0.28)	0.46(0.25)	0.34(0.28)	0.27(0.28)
		NTC	0.90(0.05)	0.91(0.04)	0.90(0.04)	0.90(0.05)	0.91(0.04)	0.90(0.04)	0.56(0.26)	0.53(0.23)	0.38(0.33)	0.56(0.26)	0.52(0.24)	0.38(0.33)
		PC	0.88(0.03)	0.89(0.02)	0.89(0.01)	0.88(0.03)	0.89(0.01)	0.89(0.01)	0.25(0.19)	0.28(0.18)	0.03(0.07)	0.25(0.19)	0.27(0.18)	0.03(0.06)
		TL-a	0.83(0.07)	0.84(0.06)	0.84(0.06)	0.83(0.07)	0.84(0.06)	0.84(0.06)	0.82(0.18)	0.82(0.10)	0.65(0.26)	0.82(0.17)	0.83(0.11)	0.66(0.25)
		TG-a	0.72(0.08)	0.74(0.08)	0.79(0.03)	0.72(0.08)	0.74(0.08)	0.79(0.03)	0.50(0.25)	0.36(0.29)	0.27(0.27)	0.50(0.25)	0.38(0.29)	0.27(0.27)
	scenario II	sC	0.94(0.06)	0.94(0.06)	0.94(0.07)	0.94(0.07)	0.94(0.07)	0.94(0.07)	0.84(0.05)	0.84(0.05)	0.84(0.05)	0.84(0.06)	0.84(0.06)	0.84(0.05)
		OTC	0.91(0.04)	0.91(0.04)	0.90(0.04)	0.91(0.04)	0.91(0.04)	0.91(0.04)	0.44(0.26)	0.35(0.28)	0.27(0.28)	0.43(0.25)	0.35(0.28)	0.26(0.30)
		TC	0.90(0.04)	0.90(0.03)	0.90(0.04)	0.90(0.04)	0.90(0.04)	0.90(0.04)	0.48(0.25)	0.40(0.24)	0.28(0.28)	0.49(0.24)	0.38(0.27)	0.26(0.30)
		NTC	0.90(0.04)	0.90(0.04)	0.90(0.03)	0.90(0.04)	0.90(0.04)	0.90(0.04)	0.51(0.23)	0.49(0.21)	0.43(0.23)	0.52(0.23)	0.48(0.22)	0.40(0.26)
		PC	0.89(0.00)	0.89(0.01)	0.89(0.02)	0.89(0.02)	0.89(0.01)	0.89(0.02)	0.30(0.14)	0.35(0.14)	0.27(0.14)	0.33(0.16)	0.32(0.14)	0.23(0.14)
		TL-a	0.84(0.06)	0.83(0.06)	0.84(0.06)	0.84(0.06)	0.83(0.06)	0.84(0.06)	0.81(0.12)	0.78(0.17)	0.64(0.27)	0.81(0.15)	0.78(0.18)	0.65(0.27)
		TG-a	0.79(0.04)	0.79(0.04)	0.79(0.04)	0.79(0.05)	0.79(0.04)	0.79(0.03)	0.44(0.23)	0.40(0.23)	0.34(0.26)	0.45(0.23)	0.37(0.24)	0.32(0.25)
	Random Setting	sC	0.94(0.07)	0.94(0.07)	0.94(0.07)	0.94(0.07)	0.94(0.07)	0.94(0.07)	0.84(0.06)	0.84(0.06)	0.84(0.05)	0.84(0.06)	0.84(0.06)	0.84(0.06)
		OTC	0.91(0.04)	0.91(0.04)	0.91(0.04)	0.91(0.04)	0.91(0.04)	0.91(0.04)	0.42(0.26)	0.33(0.28)	0.26(0.28)	0.45(0.25)	0.36(0.27)	0.30(0.28)
		TC	0.91(0.04)	0.91(0.04)	0.91(0.04)	0.91(0.04)	0.91(0.04)	0.91(0.04)	0.43(0.27)	0.33(0.28)	0.28(0.28)	0.46(0.25)	0.36(0.27)	0.31(0.28)
		NTC	0.91(0.05)	0.91(0.04)	0.91(0.04)	0.90(0.07)	0.90(0.06)	0.91(0.05)	0.62(0.23)	0.55(0.26)	0.44(0.33)	0.70(0.17)	0.64(0.23)	0.54(0.31)
		PC	0.86(0.05)	0.88(0.02)	0.89(0.00)	0.78(0.09)	0.83(0.08)	0.86(0.05)	0.28(0.23)	0.20(0.20)	0.07(0.13)	0.24(0.24)	0.19(0.20)	0.06(0.13)
		TL-a	0.81(0.09)	0.84(0.06)	0.84(0.06)	0.80(0.09)	0.84(0.07)	0.84(0.06)	0.76(0.25)	0.80(0.20)	0.68(0.26)	0.67(0.31)	0.78(0.21)	0.72(0.22)
		TG-a	0.74(0.08)	0.75(0.08)	0.78(0.04)	0.75(0.08)	0.77(0.06)	0.79(0.03)	0.49(0.27)	0.45(0.29)	0.26(0.27)	0.39(0.31)	0.37(0.30)	0.29(0.27)
Homogeneous ($\rho = 0.5$)	scenario I	sC	0.92(0.09)	0.92(0.09)	0.92(0.09)	0.92(0.09)	0.92(0.09)	0.92(0.09)	0.83(0.07)	0.83(0.06)	0.83(0.06)	0.83(0.07)	0.83(0.06)	0.83(0.07)
		OTC	0.91(0.04)	0.91(0.05)	0.91(0.04)	0.91(0.04)	0.92(0.05)	0.91(0.04)	0.48(0.20)	0.39(0.23)	0.30(0.23)	0.48(0.19)	0.40(0.22)	0.30(0.22)
		TC	0.91(0.04)	0.91(0.04)	0.91(0.04)	0.91(0.04)	0.91(0.05)	0.91(0.04)	0.50(0.20)	0.39(0.21)	0.29(0.22)	0.50(0.19)	0.39(0.20)	0.30(0.21)
		NTC	0.90(0.05)	0.90(0.04)	0.90(0.04)	0.90(0.05)	0.90(0.04)	0.90(0.04)	0.62(0.17)	0.54(0.20)	0.35(0.28)	0.62(0.17)	0.54(0.21)	0.36(0.27)
		PC	0.86(0.06)	0.89(0.02)	0.88(0.03)	0.86(0.06)	0.89(0.02)	0.88(0.03)	0.41(0.16)	0.39(0.19)	0.11(0.15)	0.41(0.16)	0.39(0.19)	0.11(0.15)
		TL-a	0.84(0.06)	0.83(0.06)	0.84(0.06)	0.84(0.06)	0.83(0.06)	0.84(0.06)	0.84(0.10)	0.79(0.13)	0.60(0.27)	0.84(0.09)	0.79(0.12)	0.60(0.27)
		TG-a	0.69(0.10)	0.71(0.10)	0.76(0.07)	0.69(0.10)	0.71(0.10)	0.75(0.07)	0.48(0.24)	0.40(0.30)	0.29(0.29)	0.48(0.24)	0.40(0.30)	0.30(0.29)
	scenario II	sC	0.92(0.09)	0.92(0.09)	0.92(0.09)	0.92(0.09)	0.92(0.09)	0.92(0.09)	0.83(0.07)	0.83(0.07)	0.83(0.06)	0.83(0.07)	0.83(0.07)	0.83(0.06)
		OTC	0.90(0.04)	0.91(0.04)	0.91(0.04)	0.91(0.04)	0.91(0.04)	0.90(0.04)	0.50(0.19)	0.40(0.22)	0.30(0.23)	0.52(0.19)	0.42(0.22)	0.34(0.23)
		TC	0.91(0.04)	0.91(0.04)	0.91(0.04)	0.91(0.04)	0.90(0.04)	0.90(0.04)	0.58(0.17)	0.46(0.23)	0.30(0.23)	0.58(0.19)	0.46(0.24)	0.33(0.24)
		NTC	0.90(0.03)	0.90(0.04)	0.90(0.04)	0.90(0.04)	0.90(0.04)	0.90(0.04)	0.59(0.14)	0.56(0.16)	0.43(0.18)	0.60(0.15)	0.55(0.16)	0.38(0.23)
		PC	0.89(0.01)	0.89(0.02)	0.90(0.03)	0.89(0.02)	0.89(0.02)	0.90(0.03)	0.48(0.14)	0.47(0.14)	0.35(0.16)	0.50(0.14)	0.47(0.14)	0.28(0.20)
		TL-a	0.83(0.07)	0.82(0.05)	0.83(0.06)	0.83(0.06)	0.82(0.06)	0.83(0.06)	0.77(0.18)	0.72(0.17)	0.55(0.28)	0.76(0.20)	0.73(0.18)	0.56(0.28)
		TG-a	0.76(0.08)	0.77(0.06)	0.76(0.06)	0.76(0.08)	0.76(0.07)	0.77(0.06)	0.48(0.25)	0.42(0.27)	0.37(0.28)	0.50(0.24)	0.43(0.27)	0.36(0.29)
	Random Setting	sC	0.92(0.09)	0.92(0.09)	0.92(0.09)	0.92(0.09)	0.92(0.09)	0.92(0.09)	0.83(0.06)	0.83(0.07)	0.83(0.07)	0.83(0.07)	0.83(0.07)	0.83(0.07)
		OTC	0.91(0.05)	0.91(0.05)	0.91(0.04)	0.91(0.04)	0.91(0.04)	0.91(0.05)	0.50(0.19)	0.39(0.23)	0.31(0.22)	0.52(0.18)	0.43(0.21)	0.34(0.22)
		TC	0.91(0.05)	0.91(0.05)	0.91(0.04)	0.91(0.04)	0.91(0.04)	0.91(0.05)	0.51(0.20)	0.39(0.22)	0.31(0.21)	0.52(0.19)	0.43(0.20)	0.36(0.22)
		NTC	0.89(0.06)	0.89(0.05)	0.90(0.05)	0.88(0.09)	0.89(0.07)	0.88(0.07)	0.67(0.17)	0.58(0.22)	0.45(0.28)	0.76(0.11)	0.69(0.18)	0.63(0.21)
		PC	0.81(0.07)	0.85(0.06)	0.86(0.05)	0.71(0.11)	0.78(0.08)	0.79(0.04)	0.41(0.20)	0.33(0.20)	0.11(0.13)	0.37(0.22)	0.32(0.19)	0.10(0.13)
		TL-a	0.83(0.08)	0.84(0.06)	0.84(0.06)	0.80(0.09)	0.83(0.06)	0.84(0.06)	0.82(0.14)	0.81(0.15)	0.67(0.26)	0.70(0.29)	0.77(0.23)	0.71(0.23)
		TG-a	0.70(0.10)	0.71(0.10)	0.75(0.07)	0.72(0.09)	0.73(0.09)	0.75(0.07)	0.50(0.28)	0.44(0.30)	0.30(0.30)	0.40(0.31)	0.42(0.31)	0.31(0.31)
Heterogeneous	scenario I	sC	0.92(0.09)	0.93(0.08)	0.93(0.08)	0.92(0.09)	0.93(0.08)	0.93(0.08)	0.84(0.06)	0.84(0.06)	0.84(0.06)	0.84(0.06)	0.84(0.06)	0.84(0.06)
		OTC	0.90(0.03)	0.90(0.04)	0.90(0.04)	0.90(0.03)	0.90(0.04)	0.90(0.03)	0.50(0.22)	0.43(0.26)	0.35(0.27)	0.49(0.24)	0.43(0.26)	0.35(0.27)
		TC	0.90(0.03)	0.90(0.04)	0.90(0.04)	0.90(0.04)	0.90(0.04)	0.90(0.03)	0.52(0.23)	0.44(0.26)	0.35(0.27)	0.51(0.23)	0.44(0.26)	0.35(0.27)
		NTC	0.90(0.04)	0.90(0.04)	0.90(0.04)	0.90(0.04)	0.90(0.04)	0.90(0.04)	0.59(0.21)	0.59(0.21)	0.43(0.29)	0.59(0.21)	0.59(0.21)	0.44(0.28)
		PC	0.88(0.02)	0.89(0.01)	0.89(0.01)	0.88(0.02)	0.89(0.01)	0.89(0.01)	0.33(0.22)	0.40(0.20)	0.11(0.14)	0.32(0.22)	0.39(0.21)	0.11(0.13)
		TL-a	0.85(0.07)	0.85(0.06)	0.85(0.06)	0.85(0.07)	0.85(0.07)	0.85(0.06)	0.83(0.13)	0.81(0.15)	0.68(0.24)	0.83(0.13)	0.80(0.17)	0.68(0.24)
		TG-a	0.76(0.06)	0.77(0.04)	0.78(0.02)	0.76(0.06)	0.77(0.04)	0.78(0.02)	0.51(0.21)	0.41(0.23)	0.31(0.20)	0.50(0.21)	0.39(0.24)	0.30(0.20)
	scenario II	sC	0.92(0.08)	0.92(0.09)	0.92(0.09)	0.92(0.08)	0.92(0.08)	0.92(0.09)	0.84(0.06)	0.84(0.06)	0.84(0.06)	0.84(0.06)	0.84(0.06)	0.84(0.06)
		OTC	0.90(0.03)	0.90(0.04)	0.90(0.04)	0.90(0.04)	0.90(0.04)	0.90(0.03)	0.55(0.21)	0.48(0.23)	0.38(0.27)	0.53(0.21)	0.44(0.26)	0.33(0.28)
		TC	0.90(0.04)	0.90(0.03)	0.90(0.04)	0.90(0.04)	0.90(0.03)	0.90(0.03)	0.59(0.19)	0.50(0.22)	0.37(0.26)	0.57(0.20)	0.46(0.24)	0.33(0.28)
		NTC	0.90(0.03)	0.90(0.04)	0.90(0.04)	0.90(0.03)	0.90(0.04)	0.90(0.04)	0.61(0.19)	0.61(0.17)	0.56(0.20)	0.62(0.18)	0.59(0.18)	0.54(0.21)
		PC	0.89(0.01)	0.89(0.00)	0.89(0.02)	0.89(0.01)	0.89(0.00)	0.89(0.02)	0.46(0.17)	0.46(0.17)	0.40(0.17)	0.47(0.17)	0.43(0.17)	0.37(0.18)
		TL-a	0.86(0.06)	0.84(0.07)	0.84(0.06)	0.86(0.06)	0.84(0.07)	0.84(0.06)	0.83(0.12)	0.75(0.18)	0.65(0.24)	0.84(0.11)	0.77(0.17)	0.67(0.24)
		TG-a	0.78(0.03)	0.78(0.02)	0.78(0.03)	0.79(0.03)	0.78(0.02)	0.79(0.03)	0.54(0.17)	0.51(0.16)	0.43(0.20)	0.55(0.17)	0.51(0.16)	0.45(0.19)
	Random Setting	sC	0.93(0.08)	0.92(0.08)	0.92(0.08)	0.92(0.08)	0.92(0.08)	0.92(0.08)	0.84(0.06)	0.84(0.06)	0.84(0.06)	0.84(0.06)	0.84(0.06)	0.84(0.06)
		OTC	0.90(0.03)	0.90(0.04)	0.90(0.03)	0.90(0.03)	0.90(0.03)	0.90(0.03)	0.51(0.22)	0.43(0.26)	0.35(0.27)	0.54(0.22)	0.47(0.25)	0.39(0.26)
		TC	0.90(0.03)	0.90(0.04)	0.90(0.03)	0.90(0.03)	0.90(0.03)	0.90(0.03)	0.51(0.22)	0.43(0.26)	0.35(0.26)	0.54(0.22)	0.47(0.25)	0.39(0.26)
		NTC	0.91(0.04)	0.91(0.04)	0.90(0.04)	0.91(0.06)	0.91(0.05)	0.91(0.04)	0.63(0.19)	0.60(0.20)	0.49(0.26)	0.68(0.17)	0.66(0.17)	0.58(0.23)
		PC	0.86(0.05)	0.88(0.03)	0.89(0.01)	0.78(0.10)	0.83(0.08)	0.88(0.04)	0.37(0.22)	0.33(0.23)	0.19(0.19)	0.33(0.24)	0.30(0.22)	0.15(0.19)
		TL-a	0.82(0.08)	0.86(0.05)	0.86(0.05)	0.82(0.07)	0.85(0.05)	0.85(0.05)	0.68(0.31)	0.80(0.20)	0.78(0.18)	0.62(0.31)	0.73(0.24)	0.77(0.18)
		TG-a	0.77(0.05)	0.78(0.03)	0.78(0.02)	0.79(0.03)	0.78(0.03)	0.78(0.02)	0.41(0.22)	0.37(0.23)	0.31(0.20)	0.41(0.21)	0.37(0.24)	0.30(0.20)

Note: sC: subCodalasso; OTC: Oracle-Trans-subCodalasso; TC: Trans-subCodalasso; NTC: Naive-Trans-subCodalasso; PC: Pooled-subCodalasso; TL-a: Trans-Lasso-alr; TG-a: Trans-GLM-alr.

Table 13: Means and standard errors (in parentheses) of true positive rate (TPR) and false discovery rate (FDR) for seven methods with $(q', q) = (10, 100)$ based on 100 replicates (Compositional Covariates).

Type	Scenario	Method	TPR						FDR					
			$d = 4$			$d = 12$			$d = 4$			$d = 12$		
			$\lvert\mathcal{A}\rvert = 5$	$\lvert\mathcal{A}\rvert = 10$	$\lvert\mathcal{A}\rvert = 15$	$\lvert\mathcal{A}\rvert = 5$	$\lvert\mathcal{A}\rvert = 10$	$\lvert\mathcal{A}\rvert = 15$	$\lvert\mathcal{A}\rvert = 5$	$\lvert\mathcal{A}\rvert = 10$	$\lvert\mathcal{A}\rvert = 15$	$\lvert\mathcal{A}\rvert = 5$	$\lvert\mathcal{A}\rvert = 10$	$\lvert\mathcal{A}\rvert = 15$
Homogeneous ($\rho = 0.2$)	scenario I	sC	0.99(0.04)	0.99(0.04)	0.99(0.04)	0.99(0.04)	0.99(0.04)	0.99(0.04)	0.86(0.05)	0.86(0.05)	0.86(0.05)	0.86(0.05)	0.86(0.05)	0.86(0.05)
		OTC	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.43(0.27)	0.34(0.30)	0.26(0.29)	0.43(0.27)	0.34(0.29)	0.27(0.29)
		TC	1.00(0.01)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.47(0.26)	0.35(0.29)	0.26(0.30)	0.48(0.26)	0.35(0.29)	0.27(0.29)
		NTC	0.99(0.03)	1.00(0.00)	1.00(0.00)	0.99(0.03)	1.00(0.00)	1.00(0.00)	0.57(0.26)	0.55(0.23)	0.39(0.34)	0.57(0.26)	0.54(0.24)	0.39(0.34)
		PC	0.99(0.04)	1.00(0.00)	1.00(0.01)	0.99(0.04)	1.00(0.00)	1.00(0.01)	0.27(0.20)	0.30(0.19)	0.04(0.08)	0.27(0.20)	0.29(0.19)	0.04(0.07)
		TL-a	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.86(0.16)	0.81(0.19)	0.63(0.33)	0.86(0.16)	0.81(0.19)	0.62(0.34)
		TG-a	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.52(0.19)	0.38(0.24)	0.26(0.25)	0.53(0.19)	0.38(0.24)	0.26(0.24)
	scenario II	sC	0.99(0.04)	0.99(0.04)	0.99(0.05)	0.99(0.04)	0.99(0.04)	0.99(0.04)	0.86(0.05)	0.86(0.05)	0.86(0.05)	0.85(0.06)	0.85(0.06)	0.86(0.05)
		OTC	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.46(0.26)	0.36(0.29)	0.28(0.28)	0.44(0.26)	0.36(0.29)	0.27(0.31)
		TC	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.50(0.25)	0.42(0.24)	0.29(0.29)	0.51(0.25)	0.39(0.28)	0.27(0.31)
		NTC	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.53(0.23)	0.52(0.21)	0.45(0.23)	0.54(0.23)	0.50(0.22)	0.42(0.26)
		PC	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.33(0.14)	0.38(0.14)	0.29(0.15)	0.35(0.16)	0.35(0.14)	0.25(0.15)
		TL-a	1.00(0.00)	1.00(0.00)	1.00(0.02)	1.00(0.00)	1.00(0.00)	1.00(0.02)	0.81(0.19)	0.78(0.22)	0.60(0.33)	0.82(0.19)	0.79(0.21)	0.63(0.31)
		TG-a	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.53(0.17)	0.42(0.20)	0.36(0.22)	0.55(0.17)	0.42(0.19)	0.35(0.22)
	Random Setting	sC	0.99(0.04)	0.99(0.04)	0.99(0.04)	0.99(0.04)	0.99(0.04)	0.99(0.04)	0.85(0.05)	0.86(0.05)	0.86(0.05)	0.85(0.06)	0.85(0.06)	0.85(0.05)
		OTC	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.44(0.27)	0.34(0.29)	0.27(0.29)	0.47(0.26)	0.37(0.28)	0.31(0.29)
		TC	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.44(0.27)	0.34(0.29)	0.29(0.29)	0.47(0.26)	0.37(0.28)	0.32(0.29)
		NTC	1.00(0.02)	1.00(0.00)	1.00(0.00)	0.97(0.06)	0.99(0.05)	1.00(0.02)	0.63(0.23)	0.57(0.26)	0.45(0.34)	0.72(0.17)	0.66(0.23)	0.55(0.32)
		PC	0.97(0.07)	0.99(0.03)	1.00(0.00)	0.85(0.12)	0.92(0.10)	0.96(0.07)	0.30(0.24)	0.21(0.21)	0.08(0.14)	0.26(0.25)	0.20(0.21)	0.06(0.14)
		TL-a	1.00(0.02)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.84(0.21)	0.84(0.19)	0.69(0.29)	0.81(0.22)	0.84(0.17)	0.73(0.28)
		TG-a	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.04)	1.00(0.00)	1.00(0.00)	0.51(0.22)	0.36(0.27)	0.26(0.24)	0.42(0.24)	0.34(0.25)	0.24(0.22)
Homogeneous ($\rho = 0.5$)	scenario I	sC	0.97(0.08)	0.97(0.07)	0.97(0.08)	0.97(0.08)	0.97(0.08)	0.97(0.08)	0.85(0.06)	0.85(0.06)	0.85(0.06)	0.85(0.06)	0.85(0.06)	0.85(0.06)
		OTC	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.49(0.21)	0.40(0.24)	0.31(0.24)	0.50(0.20)	0.41(0.23)	0.32(0.23)
		TC	1.00(0.01)	1.00(0.00)	1.00(0.00)	1.00(0.01)	1.00(0.00)	1.00(0.00)	0.51(0.21)	0.41(0.22)	0.31(0.24)	0.52(0.20)	0.41(0.21)	0.32(0.23)
		NTC	0.99(0.04)	1.00(0.01)	1.00(0.02)	0.99(0.04)	1.00(0.01)	1.00(0.02)	0.64(0.17)	0.56(0.20)	0.37(0.28)	0.64(0.17)	0.56(0.21)	0.37(0.28)
		PC	0.96(0.07)	1.00(0.02)	0.99(0.04)	0.96(0.07)	1.00(0.02)	0.99(0.04)	0.44(0.16)	0.42(0.19)	0.12(0.16)	0.44(0.16)	0.41(0.19)	0.12(0.16)
		TL-a	1.00(0.03)	0.99(0.05)	1.00(0.04)	1.00(0.03)	0.99(0.05)	1.00(0.04)	0.83(0.20)	0.77(0.23)	0.54(0.36)	0.83(0.20)	0.77(0.22)	0.55(0.37)
		TG-a	0.99(0.03)	1.00(0.00)	1.00(0.00)	0.99(0.04)	1.00(0.00)	1.00(0.00)	0.54(0.19)	0.43(0.21)	0.24(0.24)	0.54(0.19)	0.43(0.22)	0.24(0.24)
	scenario II	sC	0.97(0.08)	0.97(0.08)	0.97(0.08)	0.97(0.08)	0.97(0.08)	0.97(0.08)	0.85(0.06)	0.85(0.06)	0.85(0.06)	0.85(0.06)	0.85(0.06)	0.85(0.06)
		OTC	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.52(0.20)	0.41(0.23)	0.31(0.24)	0.54(0.19)	0.43(0.23)	0.35(0.24)
		TC	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.60(0.17)	0.47(0.23)	0.31(0.24)	0.60(0.19)	0.47(0.24)	0.34(0.25)
		NTC	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.61(0.14)	0.58(0.16)	0.45(0.19)	0.62(0.14)	0.57(0.16)	0.40(0.24)
		PC	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.50(0.13)	0.50(0.14)	0.37(0.17)	0.52(0.14)	0.50(0.14)	0.30(0.21)
		TL-a	1.00(0.03)	0.99(0.05)	0.99(0.05)	1.00(0.00)	0.99(0.04)	1.00(0.04)	0.80(0.19)	0.66(0.28)	0.57(0.34)	0.81(0.18)	0.67(0.29)	0.61(0.32)
		TG-a	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.51(0.18)	0.45(0.19)	0.34(0.23)	0.54(0.16)	0.45(0.19)	0.35(0.22)
	Random Setting	sC	0.97(0.08)	0.97(0.08)	0.97(0.07)	0.97(0.08)	0.97(0.08)	0.97(0.07)	0.85(0.06)	0.85(0.06)	0.85(0.06)	0.85(0.06)	0.85(0.06)	0.85(0.06)
		OTC	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.52(0.19)	0.40(0.23)	0.32(0.23)	0.54(0.19)	0.45(0.22)	0.36(0.23)
		TC	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.53(0.20)	0.40(0.23)	0.32(0.22)	0.54(0.19)	0.45(0.21)	0.37(0.24)
		NTC	0.97(0.06)	0.98(0.05)	0.99(0.04)	0.95(0.09)	0.96(0.06)	0.96(0.06)	0.69(0.16)	0.60(0.23)	0.46(0.28)	0.78(0.11)	0.70(0.17)	0.65(0.21)
		PC	0.90(0.09)	0.95(0.07)	0.96(0.06)	0.77(0.14)	0.86(0.10)	0.87(0.06)	0.44(0.21)	0.35(0.21)	0.12(0.14)	0.40(0.23)	0.35(0.20)	0.11(0.14)
		TL-a	0.99(0.05)	1.00(0.03)	1.00(0.03)	0.98(0.07)	0.99(0.05)	1.00(0.02)	0.84(0.18)	0.82(0.18)	0.65(0.33)	0.81(0.23)	0.82(0.20)	0.67(0.30)
		TG-a	0.97(0.07)	0.99(0.04)	1.00(0.00)	0.98(0.06)	1.00(0.04)	1.00(0.00)	0.52(0.23)	0.41(0.25)	0.25(0.24)	0.47(0.22)	0.39(0.25)	0.26(0.22)
Heterogeneous	scenario I	sC	0.97(0.07)	0.97(0.06)	0.97(0.06)	0.97(0.07)	0.97(0.06)	0.97(0.06)	0.86(0.05)	0.85(0.06)	0.85(0.06)	0.86(0.05)	0.86(0.06)	0.85(0.06)
		OTC	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.51(0.24)	0.44(0.26)	0.36(0.27)	0.50(0.25)	0.45(0.26)	0.36(0.27)
		TC	1.00(0.01)	1.00(0.00)	1.00(0.00)	1.00(0.01)	1.00(0.00)	1.00(0.00)	0.53(0.24)	0.46(0.27)	0.36(0.28)	0.52(0.25)	0.46(0.26)	0.36(0.27)
		NTC	1.00(0.01)	1.00(0.00)	1.00(0.00)	1.00(0.01)	1.00(0.00)	1.00(0.00)	0.61(0.22)	0.61(0.21)	0.45(0.29)	0.61(0.21)	0.61(0.21)	0.45(0.29)
		PC	0.99(0.03)	1.00(0.01)	1.00(0.00)	0.99(0.03)	1.00(0.01)	1.00(0.00)	0.35(0.23)	0.42(0.21)	0.12(0.15)	0.35(0.23)	0.41(0.21)	0.12(0.15)
		TL-a	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.88(0.13)	0.84(0.16)	0.69(0.30)	0.88(0.13)	0.84(0.17)	0.68(0.30)
		TG-a	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.48(0.18)	0.41(0.20)	0.37(0.21)	0.49(0.17)	0.41(0.20)	0.38(0.21)
	scenario II	sC	0.97(0.07)	0.97(0.07)	0.97(0.07)	0.97(0.07)	0.97(0.07)	0.97(0.07)	0.85(0.06)	0.86(0.06)	0.85(0.06)	0.85(0.06)	0.86(0.06)	0.85(0.06)
		OTC	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.57(0.22)	0.50(0.24)	0.39(0.27)	0.54(0.22)	0.45(0.26)	0.34(0.29)
		TC	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.01)	1.00(0.00)	1.00(0.00)	0.61(0.20)	0.52(0.23)	0.38(0.27)	0.59(0.21)	0.48(0.25)	0.34(0.29)
		NTC	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.63(0.18)	0.63(0.17)	0.58(0.20)	0.64(0.18)	0.61(0.17)	0.56(0.21)
		PC	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.49(0.17)	0.49(0.17)	0.43(0.18)	0.50(0.17)	0.46(0.17)	0.40(0.19)
		TL-a	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.87(0.11)	0.79(0.19)	0.69(0.30)	0.88(0.11)	0.79(0.22)	0.70(0.29)
		TG-a	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.56(0.13)	0.53(0.15)	0.45(0.17)	0.57(0.12)	0.52(0.16)	0.46(0.17)
	Random Setting	sC	0.98(0.06)	0.97(0.07)	0.97(0.06)	0.97(0.06)	0.97(0.07)	0.97(0.06)	0.85(0.06)	0.85(0.06)	0.85(0.06)	0.85(0.06)	0.85(0.06)	0.85(0.06)
		OTC	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.52(0.24)	0.45(0.27)	0.36(0.27)	0.55(0.22)	0.49(0.25)	0.40(0.27)
		TC	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.52(0.24)	0.45(0.27)	0.36(0.27)	0.55(0.22)	0.49(0.25)	0.40(0.26)
		NTC	1.00(0.01)	1.00(0.00)	1.00(0.00)	0.99(0.04)	1.00(0.02)	1.00(0.01)	0.65(0.19)	0.62(0.20)	0.50(0.26)	0.70(0.17)	0.68(0.17)	0.60(0.23)
		PC	0.97(0.06)	0.99(0.03)	1.00(0.01)	0.86(0.13)	0.92(0.10)	0.99(0.05)	0.40(0.23)	0.35(0.24)	0.21(0.20)	0.35(0.26)	0.32(0.24)	0.16(0.20)
		TL-a	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.82(0.22)	0.82(0.21)	0.75(0.28)	0.75(0.25)	0.83(0.18)	0.78(0.23)
		TG-a	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.43(0.21)	0.38(0.21)	0.35(0.21)	0.45(0.19)	0.42(0.18)	0.38(0.20)

Note: sC: subCodalasso; OTC: Oracle-Trans-subCodalasso; TC: Trans-subCodalasso; NTC: Naive-Trans-subCodalasso; PC: Pooled-subCodalasso; TL-a: Trans-Lasso-alr; TG-a: Trans-GLM-alr.

Table 14: Means and standard errors (in parentheses) of true positive rate (TPR) and false discovery rate (FDR) for seven methods with $(q', q) = (10, 50)$ based on 100 replicates (All Covariates).

Type	Scenario	Method	TPR						FDR					
			$d = 4$			$d = 12$			$d = 4$			$d = 12$		
			$\lvert\mathcal{A}\rvert = 5$	$\lvert\mathcal{A}\rvert = 10$	$\lvert\mathcal{A}\rvert = 15$	$\lvert\mathcal{A}\rvert = 5$	$\lvert\mathcal{A}\rvert = 10$	$\lvert\mathcal{A}\rvert = 15$	$\lvert\mathcal{A}\rvert = 5$	$\lvert\mathcal{A}\rvert = 10$	$\lvert\mathcal{A}\rvert = 15$	$\lvert\mathcal{A}\rvert = 5$	$\lvert\mathcal{A}\rvert = 10$	$\lvert\mathcal{A}\rvert = 15$
Homogeneous ($\rho = 0.2$)	scenario I	sC	0.94(0.06)	0.94(0.06)	0.94(0.06)	0.94(0.06)	0.94(0.06)	0.94(0.06)	0.75(0.04)	0.75(0.04)	0.75(0.05)	0.75(0.04)	0.75(0.04)	0.75(0.05)
		OTC	0.91(0.05)	0.91(0.05)	0.91(0.05)	0.91(0.05)	0.91(0.05)	0.92(0.05)	0.33(0.22)	0.27(0.22)	0.22(0.23)	0.33(0.22)	0.26(0.22)	0.24(0.23)
		TC	0.91(0.05)	0.91(0.05)	0.91(0.05)	0.91(0.05)	0.91(0.05)	0.91(0.05)	0.35(0.23)	0.27(0.23)	0.22(0.23)	0.36(0.23)	0.27(0.22)	0.23(0.23)
		NTC	0.92(0.05)	0.91(0.05)	0.91(0.05)	0.92(0.05)	0.91(0.05)	0.91(0.05)	0.56(0.17)	0.54(0.18)	0.37(0.26)	0.56(0.17)	0.53(0.18)	0.37(0.26)
		PC	0.89(0.02)	0.89(0.01)	0.89(0.00)	0.88(0.02)	0.89(0.01)	0.89(0.00)	0.30(0.18)	0.37(0.20)	0.04(0.09)	0.30(0.18)	0.36(0.19)	0.04(0.09)
		TL-a	0.86(0.06)	0.86(0.05)	0.87(0.05)	0.86(0.06)	0.86(0.05)	0.87(0.04)	0.83(0.04)	0.81(0.06)	0.70(0.18)	0.82(0.07)	0.80(0.06)	0.70(0.18)
		TG-a	0.77(0.07)	0.77(0.07)	0.80(0.04)	0.77(0.07)	0.77(0.07)	0.80(0.04)	0.40(0.26)	0.33(0.27)	0.26(0.27)	0.41(0.26)	0.33(0.27)	0.26(0.26)
	scenario II	sC	0.94(0.06)	0.94(0.06)	0.94(0.06)	0.94(0.06)	0.94(0.06)	0.94(0.06)	0.75(0.04)	0.75(0.05)	0.75(0.05)	0.75(0.05)	0.75(0.05)	0.75(0.05)
		OTC	0.92(0.05)	0.91(0.04)	0.91(0.04)	0.91(0.05)	0.91(0.04)	0.91(0.04)	0.38(0.22)	0.30(0.22)	0.24(0.23)	0.35(0.22)	0.29(0.23)	0.23(0.24)
		TC	0.91(0.04)	0.91(0.04)	0.91(0.04)	0.91(0.04)	0.90(0.04)	0.91(0.04)	0.44(0.20)	0.35(0.22)	0.24(0.23)	0.42(0.22)	0.33(0.24)	0.24(0.25)
		NTC	0.92(0.05)	0.91(0.05)	0.91(0.04)	0.92(0.05)	0.91(0.04)	0.91(0.04)	0.54(0.17)	0.53(0.17)	0.44(0.19)	0.55(0.17)	0.51(0.17)	0.41(0.20)
		PC	0.89(0.02)	0.90(0.03)	0.89(0.02)	0.90(0.03)	0.90(0.03)	0.89(0.02)	0.38(0.15)	0.43(0.14)	0.33(0.15)	0.41(0.15)	0.40(0.15)	0.29(0.16)
		TL-a	0.86(0.05)	0.86(0.05)	0.86(0.06)	0.86(0.05)	0.86(0.05)	0.86(0.05)	0.80(0.09)	0.78(0.09)	0.68(0.18)	0.80(0.09)	0.78(0.10)	0.69(0.18)
		TG-a	0.80(0.04)	0.80(0.04)	0.80(0.04)	0.80(0.04)	0.80(0.04)	0.80(0.04)	0.43(0.21)	0.38(0.24)	0.32(0.25)	0.43(0.21)	0.38(0.24)	0.31(0.24)
	Random Setting	sC	0.94(0.06)	0.94(0.06)	0.94(0.06)	0.94(0.06)	0.94(0.06)	0.94(0.06)	0.75(0.05)	0.75(0.05)	0.75(0.05)	0.75(0.05)	0.75(0.05)	0.75(0.04)
		OTC	0.91(0.05)	0.91(0.05)	0.91(0.05)	0.92(0.05)	0.91(0.04)	0.91(0.04)	0.34(0.22)	0.27(0.22)	0.25(0.23)	0.39(0.20)	0.33(0.21)	0.31(0.20)
		TC	0.91(0.05)	0.91(0.04)	0.91(0.05)	0.92(0.05)	0.91(0.04)	0.91(0.05)	0.35(0.22)	0.27(0.21)	0.25(0.23)	0.40(0.21)	0.34(0.21)	0.31(0.21)
		NTC	0.92(0.05)	0.91(0.05)	0.92(0.05)	0.92(0.06)	0.92(0.05)	0.92(0.05)	0.57(0.17)	0.51(0.21)	0.42(0.26)	0.67(0.11)	0.64(0.12)	0.57(0.18)
		PC	0.89(0.02)	0.89(0.01)	0.89(0.01)	0.76(0.11)	0.85(0.07)	0.84(0.06)	0.27(0.18)	0.29(0.21)	0.05(0.11)	0.24(0.20)	0.29(0.19)	0.02(0.05)
		TL-a	0.86(0.06)	0.86(0.05)	0.86(0.05)	0.85(0.07)	0.86(0.07)	0.87(0.04)	0.82(0.09)	0.82(0.06)	0.73(0.16)	0.77(0.15)	0.77(0.16)	0.78(0.11)
		TG-a	0.77(0.06)	0.78(0.06)	0.80(0.04)	0.79(0.05)	0.80(0.04)	0.80(0.04)	0.38(0.27)	0.29(0.27)	0.27(0.27)	0.29(0.25)	0.30(0.27)	0.29(0.27)
Homogeneous ($\rho = 0.5$)	scenario I	sC	0.95(0.06)	0.96(0.06)	0.95(0.06)	0.95(0.06)	0.96(0.06)	0.95(0.06)	0.76(0.07)	0.76(0.06)	0.76(0.06)	0.76(0.06)	0.76(0.06)	0.76(0.06)
		OTC	0.92(0.05)	0.92(0.05)	0.92(0.05)	0.92(0.05)	0.93(0.05)	0.92(0.05)	0.40(0.19)	0.31(0.20)	0.26(0.20)	0.41(0.19)	0.31(0.19)	0.28(0.20)
		TC	0.91(0.05)	0.92(0.05)	0.92(0.05)	0.92(0.05)	0.93(0.05)	0.93(0.05)	0.43(0.20)	0.33(0.20)	0.26(0.20)	0.43(0.19)	0.32(0.19)	0.29(0.20)
		NTC	0.92(0.05)	0.91(0.04)	0.91(0.04)	0.92(0.05)	0.91(0.04)	0.91(0.04)	0.61(0.15)	0.58(0.16)	0.40(0.23)	0.61(0.15)	0.57(0.16)	0.40(0.23)
		PC	0.88(0.05)	0.90(0.03)	0.89(0.03)	0.88(0.05)	0.90(0.03)	0.89(0.03)	0.44(0.13)	0.46(0.18)	0.18(0.17)	0.45(0.13)	0.45(0.18)	0.18(0.17)
		TL-a	0.87(0.05)	0.87(0.05)	0.86(0.05)	0.87(0.05)	0.87(0.05)	0.86(0.05)	0.83(0.06)	0.80(0.08)	0.70(0.18)	0.82(0.07)	0.79(0.08)	0.71(0.17)
		TG-a	0.75(0.08)	0.75(0.07)	0.79(0.04)	0.75(0.08)	0.75(0.07)	0.79(0.04)	0.43(0.23)	0.37(0.26)	0.29(0.30)	0.44(0.22)	0.37(0.26)	0.30(0.30)
	scenario II	sC	0.95(0.06)	0.95(0.06)	0.95(0.06)	0.95(0.06)	0.95(0.06)	0.95(0.06)	0.76(0.06)	0.76(0.06)	0.76(0.07)	0.76(0.06)	0.76(0.07)	0.76(0.06)
		OTC	0.91(0.05)	0.91(0.04)	0.91(0.05)	0.92(0.05)	0.91(0.04)	0.92(0.05)	0.45(0.18)	0.36(0.20)	0.28(0.20)	0.45(0.20)	0.39(0.18)	0.29(0.21)
		TC	0.91(0.05)	0.91(0.05)	0.91(0.05)	0.91(0.05)	0.91(0.04)	0.91(0.04)	0.52(0.18)	0.41(0.20)	0.29(0.21)	0.51(0.19)	0.44(0.17)	0.31(0.21)
		NTC	0.91(0.04)	0.92(0.05)	0.91(0.05)	0.91(0.04)	0.92(0.05)	0.91(0.04)	0.59(0.13)	0.61(0.13)	0.51(0.16)	0.60(0.13)	0.59(0.13)	0.48(0.17)
		PC	0.90(0.03)	0.90(0.04)	0.90(0.03)	0.90(0.03)	0.90(0.04)	0.90(0.03)	0.48(0.11)	0.52(0.14)	0.42(0.14)	0.50(0.11)	0.50(0.13)	0.38(0.16)
		TL-a	0.86(0.06)	0.85(0.05)	0.86(0.05)	0.86(0.06)	0.85(0.06)	0.86(0.05)	0.79(0.11)	0.73(0.17)	0.68(0.19)	0.79(0.11)	0.74(0.17)	0.69(0.20)
		TG-a	0.79(0.05)	0.80(0.04)	0.79(0.03)	0.78(0.05)	0.80(0.04)	0.79(0.04)	0.44(0.21)	0.41(0.23)	0.31(0.27)	0.44(0.22)	0.40(0.23)	0.33(0.26)
	Random Setting	sC	0.95(0.06)	0.95(0.06)	0.95(0.06)	0.95(0.06)	0.95(0.06)	0.95(0.07)	0.76(0.07)	0.76(0.06)	0.76(0.06)	0.76(0.07)	0.76(0.07)	0.76(0.06)
		OTC	0.92(0.05)	0.92(0.05)	0.92(0.05)	0.92(0.05)	0.92(0.05)	0.92(0.05)	0.41(0.19)	0.32(0.19)	0.29(0.20)	0.45(0.18)	0.41(0.19)	0.37(0.19)
		TC	0.92(0.05)	0.92(0.05)	0.92(0.05)	0.92(0.05)	0.92(0.05)	0.92(0.05)	0.42(0.19)	0.33(0.20)	0.29(0.20)	0.47(0.18)	0.42(0.20)	0.36(0.19)
		NTC	0.91(0.05)	0.91(0.05)	0.91(0.04)	0.92(0.07)	0.92(0.06)	0.91(0.06)	0.63(0.11)	0.60(0.13)	0.42(0.22)	0.71(0.11)	0.68(0.11)	0.59(0.21)
		PC	0.87(0.05)	0.89(0.03)	0.89(0.03)	0.75(0.09)	0.81(0.08)	0.80(0.04)	0.48(0.12)	0.45(0.14)	0.19(0.15)	0.45(0.16)	0.42(0.18)	0.11(0.14)
		TL-a	0.86(0.05)	0.86(0.05)	0.86(0.05)	0.85(0.06)	0.87(0.05)	0.87(0.04)	0.82(0.10)	0.80(0.11)	0.72(0.17)	0.77(0.16)	0.81(0.13)	0.76(0.13)
		TG-a	0.76(0.08)	0.78(0.06)	0.80(0.04)	0.78(0.05)	0.79(0.05)	0.79(0.04)	0.40(0.26)	0.37(0.27)	0.32(0.30)	0.36(0.26)	0.34(0.27)	0.32(0.28)
Heterogeneous	scenario I	sC	0.95(0.06)	0.95(0.06)	0.95(0.06)	0.95(0.06)	0.95(0.06)	0.95(0.06)	0.76(0.08)	0.76(0.07)	0.76(0.07)	0.76(0.08)	0.76(0.07)	0.76(0.07)
		OTC	0.91(0.05)	0.91(0.04)	0.91(0.04)	0.92(0.05)	0.91(0.04)	0.91(0.04)	0.44(0.19)	0.36(0.23)	0.30(0.23)	0.44(0.19)	0.37(0.22)	0.32(0.23)
		TC	0.91(0.05)	0.91(0.04)	0.91(0.04)	0.92(0.05)	0.91(0.04)	0.91(0.05)	0.44(0.20)	0.36(0.23)	0.30(0.23)	0.44(0.19)	0.37(0.22)	0.32(0.23)
		NTC	0.91(0.05)	0.92(0.05)	0.91(0.04)	0.91(0.05)	0.92(0.05)	0.91(0.04)	0.58(0.14)	0.57(0.15)	0.42(0.24)	0.58(0.14)	0.56(0.15)	0.42(0.24)
		PC	0.88(0.03)	0.89(0.01)	0.89(0.00)	0.88(0.03)	0.89(0.01)	0.89(0.00)	0.34(0.18)	0.37(0.17)	0.14(0.14)	0.34(0.19)	0.36(0.18)	0.13(0.14)
		TL-a	0.87(0.05)	0.86(0.05)	0.86(0.06)	0.87(0.05)	0.86(0.05)	0.86(0.06)	0.80(0.15)	0.80(0.12)	0.74(0.18)	0.80(0.14)	0.81(0.09)	0.73(0.19)
		TG-a	0.78(0.05)	0.78(0.04)	0.79(0.03)	0.78(0.05)	0.78(0.03)	0.79(0.03)	0.37(0.20)	0.27(0.22)	0.21(0.22)	0.37(0.21)	0.27(0.23)	0.20(0.22)
	scenario II	sC	0.95(0.06)	0.95(0.06)	0.95(0.06)	0.95(0.06)	0.95(0.06)	0.96(0.06)	0.76(0.07)	0.76(0.07)	0.76(0.07)	0.76(0.07)	0.76(0.07)	0.76(0.08)
		OTC	0.91(0.04)	0.91(0.04)	0.91(0.04)	0.91(0.05)	0.91(0.04)	0.91(0.04)	0.48(0.19)	0.43(0.20)	0.35(0.23)	0.45(0.22)	0.38(0.23)	0.32(0.25)
		TC	0.91(0.05)	0.91(0.04)	0.91(0.04)	0.92(0.05)	0.91(0.04)	0.91(0.04)	0.54(0.16)	0.47(0.19)	0.36(0.22)	0.54(0.19)	0.42(0.22)	0.32(0.25)
		NTC	0.91(0.05)	0.91(0.04)	0.91(0.04)	0.91(0.05)	0.91(0.05)	0.91(0.05)	0.58(0.15)	0.60(0.14)	0.53(0.18)	0.58(0.15)	0.60(0.14)	0.52(0.17)
		PC	0.89(0.02)	0.89(0.02)	0.90(0.03)	0.89(0.02)	0.89(0.02)	0.90(0.03)	0.44(0.16)	0.51(0.15)	0.41(0.18)	0.46(0.16)	0.49(0.16)	0.38(0.17)
		TL-a	0.86(0.06)	0.85(0.06)	0.85(0.06)	0.86(0.06)	0.86(0.06)	0.86(0.06)	0.81(0.09)	0.79(0.13)	0.72(0.19)	0.81(0.08)	0.79(0.12)	0.74(0.17)
		TG-a	0.79(0.04)	0.79(0.03)	0.79(0.03)	0.79(0.04)	0.79(0.03)	0.79(0.03)	0.47(0.15)	0.42(0.19)	0.27(0.22)	0.48(0.15)	0.41(0.19)	0.30(0.22)
	Random Setting	sC	0.95(0.06)	0.95(0.06)	0.95(0.06)	0.96(0.06)	0.95(0.06)	0.95(0.06)	0.76(0.07)	0.76(0.08)	0.76(0.07)	0.76(0.08)	0.76(0.07)	0.76(0.07)
		OTC	0.92(0.05)	0.91(0.05)	0.91(0.04)	0.91(0.05)	0.91(0.04)	0.91(0.04)	0.45(0.18)	0.37(0.22)	0.32(0.23)	0.49(0.20)	0.40(0.22)	0.38(0.22)
		TC	0.92(0.05)	0.91(0.05)	0.91(0.04)	0.91(0.05)	0.91(0.04)	0.91(0.04)	0.45(0.18)	0.37(0.22)	0.33(0.23)	0.50(0.20)	0.43(0.23)	0.41(0.22)
		NTC	0.91(0.05)	0.91(0.04)	0.91(0.05)	0.92(0.06)	0.92(0.05)	0.92(0.05)	0.57(0.15)	0.52(0.19)	0.41(0.24)	0.63(0.12)	0.61(0.13)	0.54(0.18)
		PC	0.88(0.02)	0.89(0.01)	0.89(0.00)	0.78(0.10)	0.84(0.08)	0.85(0.07)	0.30(0.17)	0.30(0.17)	0.10(0.13)	0.23(0.19)	0.28(0.18)	0.06(0.10)
		TL-a	0.87(0.05)	0.87(0.04)	0.86(0.06)	0.86(0.06)	0.87(0.04)	0.86(0.06)	0.81(0.11)	0.82(0.09)	0.75(0.17)	0.74(0.21)	0.80(0.13)	0.78(0.14)
		TG-a	0.79(0.05)	0.79(0.03)	0.79(0.03)	0.79(0.04)	0.79(0.03)	0.79(0.03)	0.34(0.22)	0.24(0.22)	0.20(0.22)	0.36(0.21)	0.25(0.21)	0.23(0.23)

Note: sC: subCodalasso; OTC: Oracle-Trans-subCodalasso; TC: Trans-subCodalasso; NTC: Naive-Trans-subCodalasso; PC: Pooled-subCodalasso; TL-a: Trans-Lasso-alr; TG-a: Trans-GLM-alr.

Table 15: Means and standard errors (in parentheses) of true positive rate (TPR) and false discovery rate (FDR) for seven methods with $(q', q) = (10, 50)$ based on 100 replicates (Compositional Covariates).

Type	Scenario	Method	TPR						FDR					
			$d = 4$			$d = 12$			$d = 4$			$d = 12$		
			$\lvert\mathcal{A}\rvert = 5$	$\lvert\mathcal{A}\rvert = 10$	$\lvert\mathcal{A}\rvert = 15$	$\lvert\mathcal{A}\rvert = 5$	$\lvert\mathcal{A}\rvert = 10$	$\lvert\mathcal{A}\rvert = 15$	$\lvert\mathcal{A}\rvert = 5$	$\lvert\mathcal{A}\rvert = 10$	$\lvert\mathcal{A}\rvert = 15$	$\lvert\mathcal{A}\rvert = 5$	$\lvert\mathcal{A}\rvert = 10$	$\lvert\mathcal{A}\rvert = 15$
Homogeneous ($\rho = 0.2$)	scenario I	sC	0.99(0.04)	0.99(0.04)	0.99(0.04)	0.99(0.04)	0.99(0.04)	0.99(0.04)	0.76(0.04)	0.76(0.04)	0.76(0.05)	0.76(0.04)	0.76(0.04)	0.76(0.05)
		OTC	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.33(0.23)	0.27(0.23)	0.22(0.24)	0.33(0.23)	0.27(0.22)	0.25(0.24)
		TC	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.35(0.24)	0.27(0.23)	0.22(0.24)	0.36(0.24)	0.27(0.22)	0.24(0.24)
		NTC	1.00(0.01)	1.00(0.00)	1.00(0.00)	1.00(0.01)	1.00(0.00)	1.00(0.00)	0.57(0.19)	0.56(0.18)	0.36(0.27)	0.57(0.18)	0.55(0.19)	0.37(0.27)
		PC	0.99(0.03)	1.00(0.00)	1.00(0.00)	0.99(0.03)	1.00(0.00)	1.00(0.00)	0.31(0.19)	0.39(0.20)	0.04(0.09)	0.32(0.19)	0.38(0.20)	0.04(0.09)
		TL-a	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.84(0.05)	0.81(0.10)	0.72(0.22)	0.83(0.05)	0.81(0.08)	0.72(0.21)
		TG-a	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.44(0.19)	0.34(0.21)	0.29(0.21)	0.44(0.18)	0.35(0.20)	0.29(0.21)
	scenario II	sC	0.99(0.04)	0.99(0.04)	0.99(0.04)	0.99(0.04)	0.99(0.04)	0.99(0.04)	0.76(0.05)	0.76(0.05)	0.76(0.05)	0.76(0.05)	0.76(0.05)	0.76(0.05)
		OTC	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.38(0.22)	0.30(0.22)	0.25(0.23)	0.35(0.23)	0.30(0.23)	0.23(0.24)
		TC	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.45(0.21)	0.36(0.23)	0.24(0.24)	0.43(0.23)	0.33(0.24)	0.24(0.25)
		NTC	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.56(0.17)	0.55(0.16)	0.46(0.20)	0.57(0.17)	0.53(0.17)	0.43(0.21)
		PC	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.40(0.16)	0.45(0.14)	0.35(0.15)	0.43(0.16)	0.42(0.15)	0.31(0.16)
		TL-a	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.82(0.09)	0.80(0.09)	0.73(0.18)	0.82(0.08)	0.82(0.07)	0.75(0.16)
		TG-a	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.51(0.16)	0.45(0.19)	0.36(0.19)	0.52(0.16)	0.44(0.18)	0.36(0.19)
	Random Setting	sC	0.99(0.04)	0.99(0.04)	0.99(0.04)	0.99(0.04)	0.99(0.04)	0.99(0.04)	0.76(0.05)	0.76(0.05)	0.76(0.05)	0.76(0.05)	0.76(0.05)	0.76(0.04)
		OTC	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.34(0.23)	0.28(0.22)	0.25(0.24)	0.39(0.21)	0.34(0.21)	0.32(0.21)
		TC	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.35(0.23)	0.27(0.22)	0.26(0.24)	0.40(0.21)	0.35(0.21)	0.32(0.21)
		NTC	1.00(0.01)	1.00(0.00)	1.00(0.00)	0.99(0.04)	1.00(0.01)	1.00(0.02)	0.58(0.17)	0.53(0.21)	0.42(0.27)	0.68(0.11)	0.65(0.13)	0.57(0.19)
		PC	1.00(0.02)	1.00(0.00)	1.00(0.00)	0.83(0.14)	0.95(0.09)	0.94(0.08)	0.29(0.19)	0.31(0.22)	0.05(0.11)	0.26(0.21)	0.32(0.20)	0.02(0.06)
		TL-a	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.84(0.05)	0.81(0.12)	0.75(0.17)	0.80(0.12)	0.82(0.10)	0.77(0.17)
		TG-a	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.44(0.18)	0.36(0.21)	0.29(0.21)	0.40(0.17)	0.39(0.20)	0.31(0.19)
Homogeneous ($\rho = 0.5$)	scenario I	sC	0.99(0.03)	0.99(0.03)	0.99(0.03)	0.99(0.03)	0.99(0.03)	0.99(0.03)	0.77(0.07)	0.77(0.07)	0.77(0.06)	0.77(0.07)	0.77(0.06)	0.77(0.06)
		OTC	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.40(0.20)	0.32(0.21)	0.27(0.21)	0.42(0.20)	0.32(0.20)	0.30(0.21)
		TC	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.43(0.21)	0.33(0.21)	0.27(0.21)	0.44(0.20)	0.33(0.20)	0.30(0.20)
		NTC	1.00(0.02)	1.00(0.00)	1.00(0.00)	1.00(0.02)	1.00(0.00)	1.00(0.00)	0.62(0.15)	0.59(0.16)	0.41(0.23)	0.63(0.15)	0.59(0.16)	0.41(0.24)
		PC	0.98(0.05)	1.00(0.02)	1.00(0.02)	0.98(0.05)	1.00(0.02)	1.00(0.02)	0.47(0.13)	0.48(0.18)	0.19(0.18)	0.47(0.13)	0.47(0.18)	0.19(0.17)
		TL-a	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.83(0.07)	0.81(0.11)	0.67(0.26)	0.83(0.07)	0.81(0.11)	0.67(0.26)
		TG-a	1.00(0.02)	1.00(0.00)	1.00(0.00)	1.00(0.02)	1.00(0.00)	1.00(0.00)	0.54(0.16)	0.43(0.20)	0.29(0.21)	0.55(0.16)	0.44(0.19)	0.30(0.21)
	scenario II	sC	0.99(0.03)	0.99(0.03)	0.99(0.03)	0.99(0.03)	0.99(0.03)	0.99(0.03)	0.77(0.06)	0.77(0.06)	0.77(0.07)	0.77(0.07)	0.77(0.07)	0.77(0.06)
		OTC	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.45(0.19)	0.37(0.20)	0.29(0.21)	0.46(0.21)	0.41(0.19)	0.29(0.22)
		TC	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.54(0.18)	0.42(0.20)	0.30(0.21)	0.53(0.19)	0.45(0.18)	0.32(0.22)
		NTC	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.61(0.13)	0.63(0.12)	0.54(0.16)	0.62(0.12)	0.61(0.13)	0.50(0.17)
		PC	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.51(0.11)	0.54(0.14)	0.45(0.15)	0.53(0.11)	0.52(0.13)	0.41(0.16)
		TL-a	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.80(0.14)	0.78(0.15)	0.73(0.17)	0.80(0.14)	0.78(0.15)	0.75(0.16)
		TG-a	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.56(0.13)	0.47(0.17)	0.38(0.20)	0.58(0.12)	0.49(0.15)	0.38(0.19)
	Random Setting	sC	0.99(0.03)	0.99(0.03)	0.99(0.03)	0.99(0.03)	0.99(0.03)	0.99(0.04)	0.77(0.07)	0.77(0.07)	0.77(0.06)	0.77(0.07)	0.77(0.07)	0.77(0.06)
		OTC	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.42(0.20)	0.33(0.20)	0.30(0.20)	0.46(0.20)	0.43(0.20)	0.39(0.19)
		TC	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.01)	1.00(0.00)	1.00(0.00)	0.42(0.20)	0.34(0.20)	0.31(0.20)	0.48(0.19)	0.44(0.21)	0.38(0.19)
		NTC	1.00(0.02)	1.00(0.00)	1.00(0.00)	0.98(0.05)	0.99(0.03)	0.98(0.05)	0.65(0.10)	0.61(0.12)	0.43(0.22)	0.72(0.11)	0.70(0.11)	0.60(0.21)
		PC	0.98(0.05)	0.99(0.03)	1.00(0.02)	0.82(0.12)	0.89(0.09)	0.88(0.06)	0.51(0.13)	0.48(0.15)	0.20(0.16)	0.48(0.16)	0.45(0.19)	0.12(0.15)
		TL-a	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.02)	1.00(0.00)	1.00(0.00)	0.83(0.07)	0.83(0.07)	0.70(0.21)	0.82(0.10)	0.83(0.04)	0.74(0.19)
		TG-a	1.00(0.02)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.53(0.19)	0.43(0.19)	0.31(0.20)	0.44(0.19)	0.43(0.19)	0.36(0.19)
Heterogeneous	scenario I	sC	0.99(0.03)	0.99(0.03)	0.99(0.03)	0.99(0.03)	0.99(0.03)	0.99(0.03)	0.77(0.08)	0.77(0.08)	0.77(0.08)	0.77(0.08)	0.77(0.08)	0.77(0.08)
		OTC	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.45(0.19)	0.37(0.23)	0.31(0.23)	0.46(0.19)	0.38(0.22)	0.33(0.23)
		TC	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.45(0.20)	0.38(0.23)	0.31(0.24)	0.45(0.20)	0.38(0.22)	0.33(0.23)
		NTC	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.58(0.15)	0.58(0.14)	0.43(0.24)	0.58(0.15)	0.58(0.15)	0.43(0.24)
		PC	0.99(0.03)	1.00(0.00)	1.00(0.00)	0.99(0.03)	1.00(0.00)	1.00(0.00)	0.36(0.19)	0.40(0.18)	0.15(0.15)	0.36(0.19)	0.38(0.18)	0.15(0.15)
		TL-a	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.84(0.06)	0.82(0.09)	0.73(0.21)	0.84(0.06)	0.82(0.12)	0.73(0.22)
		TG-a	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.44(0.18)	0.35(0.17)	0.29(0.15)	0.44(0.18)	0.36(0.16)	0.30(0.14)
	scenario II	sC	0.99(0.03)	0.99(0.03)	0.99(0.03)	0.99(0.03)	0.99(0.03)	0.99(0.03)	0.77(0.08)	0.77(0.08)	0.77(0.08)	0.77(0.08)	0.77(0.08)	0.77(0.08)
		OTC	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.50(0.19)	0.44(0.21)	0.37(0.24)	0.47(0.23)	0.40(0.24)	0.33(0.25)
		TC	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.56(0.17)	0.49(0.19)	0.37(0.23)	0.55(0.20)	0.44(0.22)	0.33(0.25)
		NTC	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.60(0.15)	0.62(0.13)	0.55(0.17)	0.60(0.15)	0.62(0.13)	0.54(0.17)
		PC	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.47(0.16)	0.54(0.15)	0.43(0.18)	0.49(0.16)	0.51(0.16)	0.41(0.18)
		TL-a	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.83(0.09)	0.81(0.11)	0.77(0.15)	0.83(0.08)	0.82(0.11)	0.78(0.15)
		TG-a	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.55(0.12)	0.49(0.15)	0.39(0.16)	0.56(0.11)	0.51(0.13)	0.39(0.16)
	Random Setting	sC	0.99(0.03)	0.99(0.03)	0.99(0.03)	0.99(0.03)	0.99(0.03)	0.99(0.03)	0.77(0.08)	0.77(0.08)	0.77(0.08)	0.77(0.09)	0.77(0.08)	0.77(0.08)
		OTC	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.46(0.19)	0.39(0.22)	0.34(0.23)	0.50(0.20)	0.42(0.23)	0.39(0.22)
		TC	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.46(0.19)	0.39(0.22)	0.34(0.23)	0.51(0.20)	0.45(0.23)	0.43(0.23)
		NTC	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.02)	1.00(0.02)	1.00(0.01)	0.58(0.16)	0.53(0.19)	0.41(0.25)	0.64(0.13)	0.62(0.14)	0.54(0.19)
		PC	0.99(0.03)	1.00(0.01)	1.00(0.00)	0.87(0.13)	0.94(0.10)	0.95(0.09)	0.32(0.18)	0.32(0.18)	0.11(0.14)	0.26(0.20)	0.30(0.19)	0.06(0.11)
		TL-a	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.84(0.08)	0.84(0.04)	0.76(0.19)	0.81(0.13)	0.84(0.08)	0.80(0.10)
		TG-a	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.42(0.18)	0.34(0.16)	0.31(0.15)	0.43(0.16)	0.39(0.15)	0.33(0.15)

Note: sC: subCodalasso; OTC: Oracle-Trans-subCodalasso; TC: Trans-subCodalasso; NTC: Naive-Trans-subCodalasso; PC: Pooled-subCodalasso; TL-a: Trans-Lasso-alr; TG-a: Trans-GLM-alr.

Table 16: Means and standard errors (in parentheses) of true positive rate (TPR) and false discovery rate (FDR) for seven methods with $(q', q) = (10, 200)$ based on 100 replicates (All Covariates).

Type	Scenario	Method	TPR						FDR					
			$d = 4$			$d = 12$			$d = 4$			$d = 12$		
			$\lvert\mathcal{A}\rvert = 5$	$\lvert\mathcal{A}\rvert = 10$	$\lvert\mathcal{A}\rvert = 15$	$\lvert\mathcal{A}\rvert = 5$	$\lvert\mathcal{A}\rvert = 10$	$\lvert\mathcal{A}\rvert = 15$	$\lvert\mathcal{A}\rvert = 5$	$\lvert\mathcal{A}\rvert = 10$	$\lvert\mathcal{A}\rvert = 15$	$\lvert\mathcal{A}\rvert = 5$	$\lvert\mathcal{A}\rvert = 10$	$\lvert\mathcal{A}\rvert = 15$
Homogeneous ($\rho = 0.2$)	scenario I	sC	0.93(0.07)	0.93(0.08)	0.93(0.07)	0.93(0.07)	0.93(0.08)	0.93(0.07)	0.90(0.06)	0.90(0.06)	0.90(0.06)	0.90(0.06)	0.90(0.06)	0.90(0.06)
		OTC	0.91(0.04)	0.91(0.04)	0.92(0.05)	0.91(0.05)	0.91(0.04)	0.91(0.05)	0.49(0.25)	0.39(0.29)	0.36(0.33)	0.51(0.24)	0.40(0.29)	0.37(0.32)
		TC	0.90(0.04)	0.90(0.04)	0.91(0.05)	0.91(0.05)	0.91(0.04)	0.91(0.05)	0.54(0.25)	0.41(0.29)	0.36(0.32)	0.55(0.25)	0.42(0.29)	0.36(0.32)
		NTC	0.90(0.06)	0.90(0.04)	0.91(0.04)	0.90(0.06)	0.90(0.04)	0.91(0.04)	0.63(0.26)	0.53(0.28)	0.46(0.34)	0.62(0.26)	0.53(0.29)	0.46(0.34)
		PC	0.86(0.05)	0.89(0.02)	0.89(0.00)	0.86(0.05)	0.89(0.02)	0.89(0.00)	0.17(0.18)	0.21(0.17)	0.04(0.09)	0.17(0.18)	0.21(0.17)	0.04(0.09)
		TL-a	0.80(0.08)	0.81(0.08)	0.82(0.07)	0.79(0.09)	0.80(0.08)	0.83(0.06)	0.82(0.16)	0.74(0.22)	0.62(0.31)	0.80(0.18)	0.75(0.22)	0.60(0.33)
		TG-a	0.68(0.09)	0.70(0.08)	0.75(0.07)	0.68(0.09)	0.69(0.09)	0.75(0.07)	0.49(0.29)	0.43(0.31)	0.36(0.31)	0.49(0.29)	0.44(0.31)	0.37(0.31)
	scenario II	sC	0.93(0.08)	0.93(0.07)	0.93(0.07)	0.93(0.08)	0.93(0.07)	0.93(0.07)	0.90(0.06)	0.90(0.07)	0.90(0.06)	0.90(0.06)	0.90(0.07)	0.90(0.05)
		OTC	0.91(0.04)	0.91(0.04)	0.91(0.05)	0.91(0.04)	0.91(0.04)	0.91(0.04)	0.51(0.25)	0.42(0.28)	0.37(0.33)	0.52(0.26)	0.43(0.29)	0.35(0.33)
		TC	0.91(0.04)	0.90(0.04)	0.91(0.04)	0.91(0.04)	0.90(0.04)	0.91(0.04)	0.54(0.24)	0.47(0.27)	0.36(0.32)	0.55(0.25)	0.45(0.29)	0.34(0.32)
		NTC	0.91(0.04)	0.91(0.04)	0.90(0.04)	0.91(0.04)	0.91(0.04)	0.91(0.04)	0.55(0.27)	0.54(0.25)	0.45(0.30)	0.54(0.27)	0.53(0.26)	0.45(0.30)
		PC	0.89(0.02)	0.89(0.02)	0.89(0.00)	0.89(0.02)	0.89(0.02)	0.89(0.00)	0.25(0.13)	0.31(0.15)	0.20(0.15)	0.26(0.13)	0.29(0.16)	0.17(0.14)
		TL-a	0.82(0.07)	0.82(0.07)	0.81(0.06)	0.82(0.07)	0.82(0.07)	0.81(0.07)	0.72(0.26)	0.63(0.28)	0.48(0.36)	0.73(0.26)	0.62(0.30)	0.50(0.35)
		TG-a	0.76(0.07)	0.78(0.05)	0.77(0.06)	0.75(0.07)	0.77(0.06)	0.77(0.07)	0.46(0.25)	0.45(0.26)	0.40(0.28)	0.46(0.26)	0.45(0.27)	0.41(0.27)
	Random Setting	sC	0.93(0.07)	0.93(0.07)	0.93(0.07)	0.93(0.07)	0.93(0.08)	0.93(0.07)	0.90(0.06)	0.90(0.06)	0.90(0.06)	0.90(0.06)	0.90(0.06)	0.90(0.06)
		OTC	0.91(0.04)	0.91(0.04)	0.91(0.05)	0.91(0.04)	0.91(0.04)	0.91(0.04)	0.51(0.24)	0.40(0.30)	0.36(0.32)	0.49(0.26)	0.39(0.29)	0.37(0.32)
		TC	0.91(0.05)	0.91(0.04)	0.91(0.05)	0.91(0.05)	0.91(0.04)	0.91(0.04)	0.53(0.25)	0.41(0.31)	0.36(0.32)	0.51(0.27)	0.40(0.30)	0.36(0.32)
		NTC	0.91(0.07)	0.91(0.06)	0.91(0.05)	0.90(0.08)	0.91(0.07)	0.90(0.07)	0.78(0.17)	0.71(0.24)	0.58(0.33)	0.80(0.16)	0.77(0.18)	0.69(0.28)
		PC	0.77(0.10)	0.82(0.08)	0.87(0.04)	0.70(0.11)	0.76(0.10)	0.83(0.07)	0.15(0.21)	0.09(0.17)	0.04(0.13)	0.14(0.21)	0.08(0.16)	0.03(0.10)
		TL-a	0.76(0.12)	0.80(0.09)	0.81(0.06)	0.74(0.12)	0.79(0.09)	0.82(0.06)	0.62(0.37)	0.72(0.27)	0.57(0.32)	0.61(0.37)	0.62(0.34)	0.64(0.33)
		TG-a	0.68(0.10)	0.68(0.10)	0.73(0.09)	0.70(0.09)	0.69(0.10)	0.75(0.07)	0.54(0.30)	0.49(0.30)	0.40(0.32)	0.51(0.32)	0.49(0.31)	0.36(0.32)
Homogeneous ($\rho = 0.5$)	scenario I	sC	0.89(0.11)	0.89(0.11)	0.89(0.11)	0.89(0.11)	0.89(0.11)	0.89(0.11)	0.88(0.06)	0.88(0.06)	0.87(0.07)	0.88(0.06)	0.88(0.06)	0.87(0.07)
		OTC	0.91(0.04)	0.91(0.05)	0.92(0.05)	0.91(0.04)	0.91(0.05)	0.92(0.05)	0.61(0.18)	0.51(0.21)	0.43(0.25)	0.61(0.18)	0.50(0.21)	0.44(0.24)
		TC	0.89(0.06)	0.91(0.04)	0.92(0.05)	0.89(0.06)	0.91(0.04)	0.92(0.05)	0.61(0.19)	0.53(0.22)	0.43(0.25)	0.61(0.18)	0.53(0.21)	0.44(0.24)
		NTC	0.87(0.08)	0.89(0.05)	0.89(0.05)	0.87(0.08)	0.89(0.05)	0.89(0.05)	0.63(0.22)	0.50(0.26)	0.37(0.32)	0.62(0.22)	0.50(0.26)	0.38(0.32)
		PC	0.82(0.07)	0.87(0.04)	0.87(0.05)	0.81(0.07)	0.87(0.04)	0.87(0.05)	0.34(0.16)	0.31(0.19)	0.07(0.11)	0.34(0.16)	0.31(0.19)	0.08(0.11)
		TL-a	0.77(0.10)	0.78(0.08)	0.82(0.06)	0.77(0.10)	0.79(0.08)	0.82(0.06)	0.77(0.19)	0.68(0.25)	0.51(0.32)	0.77(0.20)	0.69(0.26)	0.50(0.32)
		TG-a	0.63(0.09)	0.63(0.08)	0.64(0.09)	0.63(0.09)	0.63(0.08)	0.64(0.09)	0.53(0.26)	0.40(0.31)	0.33(0.32)	0.52(0.25)	0.41(0.31)	0.34(0.31)
	scenario II	sC	0.89(0.11)	0.89(0.11)	0.89(0.11)	0.89(0.11)	0.89(0.11)	0.89(0.11)	0.88(0.06)	0.88(0.07)	0.88(0.07)	0.88(0.06)	0.88(0.06)	0.88(0.07)
		OTC	0.91(0.04)	0.91(0.04)	0.91(0.05)	0.90(0.05)	0.91(0.04)	0.91(0.04)	0.60(0.19)	0.52(0.22)	0.44(0.25)	0.59(0.19)	0.52(0.21)	0.42(0.27)
		TC	0.90(0.04)	0.91(0.04)	0.91(0.05)	0.90(0.04)	0.91(0.04)	0.91(0.04)	0.56(0.22)	0.49(0.25)	0.42(0.26)	0.57(0.21)	0.48(0.25)	0.43(0.27)
		NTC	0.90(0.04)	0.90(0.03)	0.90(0.03)	0.90(0.04)	0.90(0.03)	0.90(0.03)	0.57(0.22)	0.57(0.22)	0.48(0.26)	0.56(0.22)	0.56(0.21)	0.45(0.27)
		PC	0.89(0.02)	0.89(0.02)	0.89(0.01)	0.89(0.03)	0.89(0.01)	0.89(0.01)	0.38(0.18)	0.41(0.18)	0.29(0.21)	0.39(0.17)	0.40(0.18)	0.25(0.21)
		TL-a	0.81(0.07)	0.81(0.06)	0.81(0.06)	0.82(0.07)	0.80(0.07)	0.81(0.07)	0.69(0.25)	0.59(0.28)	0.44(0.33)	0.71(0.24)	0.61(0.29)	0.46(0.34)
		TG-a	0.68(0.09)	0.68(0.09)	0.67(0.09)	0.67(0.10)	0.67(0.09)	0.66(0.09)	0.46(0.28)	0.41(0.29)	0.38(0.30)	0.46(0.28)	0.40(0.29)	0.40(0.30)
	Random Setting	sC	0.89(0.11)	0.89(0.11)	0.89(0.11)	0.87(0.11)	0.87(0.11)	0.87(0.11)	0.88(0.06)	0.88(0.06)	0.88(0.07)	0.86(0.08)	0.86(0.08)	0.86(0.08)
		OTC	0.91(0.04)	0.91(0.04)	0.92(0.05)	0.91(0.04)	0.90(0.03)	0.90(0.04)	0.61(0.18)	0.50(0.21)	0.44(0.24)	0.69(0.16)	0.56(0.22)	0.49(0.25)
		TC	0.91(0.04)	0.91(0.04)	0.92(0.05)	0.91(0.04)	0.90(0.03)	0.90(0.04)	0.62(0.19)	0.50(0.22)	0.44(0.24)	0.70(0.16)	0.56(0.22)	0.49(0.25)
		NTC	0.85(0.12)	0.86(0.09)	0.87(0.07)	0.88(0.07)	0.88(0.07)	0.90(0.05)	0.76(0.18)	0.68(0.21)	0.58(0.27)	0.74(0.17)	0.67(0.23)	0.59(0.27)
		PC	0.72(0.12)	0.78(0.08)	0.82(0.06)	0.72(0.11)	0.77(0.10)	0.86(0.06)	0.38(0.26)	0.28(0.21)	0.22(0.21)	0.23(0.22)	0.13(0.17)	0.10(0.15)
		TL-a	0.73(0.12)	0.78(0.10)	0.82(0.05)	0.77(0.12)	0.82(0.08)	0.84(0.06)	0.58(0.37)	0.65(0.32)	0.64(0.26)	0.44(0.38)	0.53(0.38)	0.66(0.29)
		TG-a	0.63(0.10)	0.64(0.10)	0.64(0.09)	0.77(0.05)	0.77(0.05)	0.78(0.02)	0.59(0.25)	0.55(0.27)	0.41(0.32)	0.50(0.18)	0.42(0.22)	0.41(0.2)
Heterogeneous	scenario I	sC	0.87(0.11)	0.87(0.11)	0.87(0.11)	0.87(0.11)	0.87(0.11)	0.87(0.11)	0.86(0.08)	0.86(0.08)	0.87(0.07)	0.86(0.08)	0.86(0.08)	0.87(0.08)
		OTC	0.91(0.04)	0.90(0.03)	0.90(0.03)	0.91(0.04)	0.90(0.03)	0.90(0.03)	0.67(0.18)	0.55(0.22)	0.48(0.25)	0.67(0.18)	0.54(0.23)	0.48(0.26)
		TC	0.91(0.04)	0.90(0.03)	0.90(0.03)	0.91(0.04)	0.90(0.03)	0.90(0.03)	0.70(0.17)	0.57(0.21)	0.48(0.25)	0.70(0.16)	0.56(0.22)	0.48(0.26)
		NTC	0.90(0.04)	0.90(0.03)	0.90(0.03)	0.90(0.04)	0.90(0.03)	0.90(0.03)	0.60(0.24)	0.59(0.23)	0.47(0.30)	0.60(0.24)	0.59(0.23)	0.46(0.30)
		PC	0.88(0.03)	0.89(0.00)	0.89(0.01)	0.88(0.03)	0.89(0.00)	0.89(0.01)	0.33(0.20)	0.36(0.22)	0.17(0.15)	0.33(0.20)	0.36(0.22)	0.17(0.15)
		TL-a	0.85(0.06)	0.85(0.05)	0.84(0.06)	0.84(0.07)	0.86(0.05)	0.84(0.06)	0.83(0.18)	0.84(0.15)	0.61(0.30)	0.80(0.22)	0.84(0.15)	0.62(0.30)
		TG-a	0.74(0.06)	0.77(0.03)	0.78(0.02)	0.74(0.06)	0.77(0.03)	0.78(0.02)	0.52(0.21)	0.46(0.21)	0.40(0.19)	0.52(0.21)	0.46(0.21)	0.40(0.19)
	scenario II	sC	0.88(0.11)	0.88(0.11)	0.87(0.11)	0.87(0.11)	0.87(0.11)	0.87(0.11)	0.86(0.08)	0.86(0.08)	0.87(0.07)	0.86(0.08)	0.86(0.08)	0.86(0.08)
		OTC	0.91(0.04)	0.90(0.03)	0.90(0.03)	0.90(0.04)	0.90(0.03)	0.90(0.03)	0.70(0.16)	0.59(0.21)	0.50(0.25)	0.64(0.19)	0.50(0.27)	0.42(0.29)
		TC	0.90(0.03)	0.90(0.03)	0.90(0.03)	0.90(0.04)	0.90(0.03)	0.90(0.03)	0.68(0.18)	0.60(0.21)	0.50(0.24)	0.68(0.18)	0.55(0.25)	0.42(0.29)
		NTC	0.90(0.03)	0.90(0.03)	0.90(0.03)	0.90(0.03)	0.90(0.03)	0.90(0.03)	0.60(0.21)	0.61(0.21)	0.58(0.20)	0.62(0.19)	0.59(0.22)	0.54(0.24)
		PC	0.89(0.01)	0.89(0.00)	0.89(0.00)	0.89(0.01)	0.89(0.00)	0.89(0.01)	0.47(0.18)	0.47(0.16)	0.41(0.16)	0.49(0.18)	0.43(0.16)	0.37(0.17)
		TL-a	0.84(0.07)	0.84(0.07)	0.83(0.07)	0.85(0.05)	0.83(0.07)	0.83(0.07)	0.83(0.19)	0.76(0.21)	0.64(0.29)	0.83(0.16)	0.76(0.21)	0.65(0.29)
		TG-a	0.78(0.02)	0.78(0.02)	0.78(0.02)	0.78(0.02)	0.78(0.02)	0.78(0.02)	0.59(0.14)	0.57(0.13)	0.47(0.18)	0.60(0.14)	0.54(0.15)	0.48(0.17)
	Random Setting	sC	0.87(0.10)	0.87(0.10)	0.87(0.11)	0.87(0.11)	0.87(0.11)	0.87(0.11)	0.86(0.08)	0.86(0.08)	0.86(0.08)	0.86(0.08)	0.86(0.08)	0.86(0.08)
		OTC	0.91(0.04)	0.90(0.03)	0.90(0.04)	0.91(0.04)	0.90(0.03)	0.90(0.04)	0.67(0.18)	0.55(0.23)	0.46(0.27)	0.69(0.16)	0.56(0.22)	0.49(0.25)
		TC	0.91(0.04)	0.90(0.03)	0.90(0.04)	0.91(0.04)	0.90(0.03)	0.90(0.04)	0.67(0.18)	0.55(0.23)	0.46(0.27)	0.70(0.16)	0.56(0.22)	0.49(0.25)
		NTC	0.89(0.06)	0.90(0.05)	0.90(0.04)	0.88(0.07)	0.88(0.07)	0.90(0.05)	0.68(0.22)	0.63(0.24)	0.52(0.30)	0.74(0.17)	0.67(0.23)	0.59(0.27)
		PC	0.80(0.09)	0.84(0.07)	0.88(0.02)	0.72(0.11)	0.77(0.10)	0.86(0.06)	0.26(0.21)	0.17(0.18)	0.12(0.17)	0.23(0.22)	0.13(0.17)	0.10(0.15)
		TL-a	0.79(0.10)	0.82(0.08)	0.84(0.06)	0.77(0.12)	0.82(0.08)	0.84(0.06)	0.54(0.38)	0.62(0.35)	0.68(0.30)	0.44(0.38)	0.53(0.38)	0.66(0.29)
		TG-a	0.75(0.07)	0.77(0.04)	0.78(0.02)	0.77(0.05)	0.77(0.05)	0.78(0.02)	0.50(0.20)	0.42(0.22)	0.39(0.20)	0.50(0.18)	0.42(0.22)	0.41(0.20)

Note: sC: subCodalasso; OTC: Oracle-Trans-subCodalasso; TC: Trans-subCodalasso; NTC: Naive-Trans-subCodalasso; PC: Pooled-subCodalasso; TL-a: Trans-Lasso-alr; TG-a: Trans-GLM-alr.

Table 17: Means and standard errors (in parentheses) of true positive rate (TPR) and false discovery rate (FDR) for seven methods with $(q', q) = (10, 200)$ based on 100 replicates (Compositional Covariates).

Type	Scenario	Method	TPR						FDR					
			$d=4$			$d=12$			$d=4$			$d=12$		
			$\lvert\mathcal{A}\rvert=5$	$\lvert\mathcal{A}\rvert=10$	$\lvert\mathcal{A}\rvert=15$	$\lvert\mathcal{A}\rvert=5$	$\lvert\mathcal{A}\rvert=10$	$\lvert\mathcal{A}\rvert=15$	$\lvert\mathcal{A}\rvert=5$	$\lvert\mathcal{A}\rvert=10$	$\lvert\mathcal{A}\rvert=15$	$\lvert\mathcal{A}\rvert=5$	$\lvert\mathcal{A}\rvert=10$	$\lvert\mathcal{A}\rvert=15$
Homogeneous ($\rho=0.2$)	scenario I	sC	0.98(0.05)	0.98(0.05)	0.98(0.05)	0.98(0.05)	0.98(0.05)	0.98(0.05)	0.91(0.06)	0.91(0.06)	0.91(0.06)	0.91(0.05)	0.91(0.06)	0.91(0.06)
		OTC	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.51(0.26)	0.40(0.30)	0.37(0.34)	0.53(0.24)	0.42(0.30)	0.38(0.33)
		TC	1.00(0.02)	1.00(0.00)	1.00(0.00)	1.00(0.02)	1.00(0.00)	1.00(0.00)	0.56(0.25)	0.42(0.30)	0.37(0.33)	0.57(0.25)	0.43(0.30)	0.37(0.33)
		NTC	0.99(0.04)	1.00(0.01)	1.00(0.00)	0.99(0.04)	1.00(0.01)	1.00(0.00)	0.65(0.26)	0.54(0.29)	0.48(0.35)	0.65(0.26)	0.55(0.29)	0.48(0.35)
		PC	0.96(0.07)	1.00(0.02)	1.00(0.00)	0.96(0.07)	1.00(0.02)	1.00(0.00)	0.18(0.19)	0.23(0.18)	0.04(0.10)	0.18(0.19)	0.23(0.18)	0.04(0.09)
		TL-a	1.00(0.03)	0.99(0.07)	0.98(0.09)	1.00(0.03)	0.99(0.07)	0.98(0.09)	0.85(0.22)	0.81(0.25)	0.54(0.37)	0.85(0.22)	0.81(0.25)	0.55(0.38)
		TG-a	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.54(0.22)	0.38(0.26)	0.29(0.29)	0.54(0.22)	0.40(0.27)	0.30(0.29)
	scenario II	sC	0.98(0.05)	0.98(0.05)	0.98(0.05)	0.98(0.05)	0.98(0.05)	0.98(0.05)	0.91(0.05)	0.91(0.06)	0.91(0.05)	0.91(0.05)	0.91(0.06)	0.91(0.05)
		OTC	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.53(0.25)	0.44(0.29)	0.38(0.33)	0.54(0.27)	0.45(0.30)	0.35(0.34)
		TC	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.01)	1.00(0.00)	1.00(0.00)	0.56(0.24)	0.49(0.27)	0.38(0.33)	0.57(0.26)	0.47(0.30)	0.35(0.34)
		NTC	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.56(0.27)	0.57(0.24)	0.47(0.30)	0.56(0.27)	0.55(0.26)	0.47(0.30)
		PC	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.27(0.14)	0.34(0.16)	0.21(0.16)	0.28(0.14)	0.31(0.16)	0.19(0.15)
		TL-a	0.99(0.05)	1.00(0.05)	1.00(0.00)	0.99(0.06)	1.00(0.03)	1.00(0.00)	0.76(0.29)	0.66(0.32)	0.52(0.39)	0.76(0.30)	0.68(0.33)	0.57(0.38)
		TG-a	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.47(0.24)	0.45(0.24)	0.39(0.28)	0.49(0.24)	0.44(0.24)	0.39(0.27)
	Random Setting	sC	0.98(0.05)	0.98(0.05)	0.98(0.05)	0.98(0.05)	0.98(0.05)	0.98(0.05)	0.91(0.05)	0.91(0.05)	0.91(0.06)	0.91(0.05)	0.91(0.06)	0.91(0.05)
		OTC	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.53(0.25)	0.41(0.31)	0.38(0.33)	0.52(0.27)	0.41(0.30)	0.38(0.33)
		TC	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.01)	1.00(0.00)	1.00(0.00)	0.55(0.26)	0.43(0.32)	0.37(0.33)	0.53(0.28)	0.42(0.31)	0.38(0.33)
		NTC	0.98(0.05)	0.98(0.04)	1.00(0.02)	0.97(0.06)	0.98(0.06)	0.98(0.05)	0.79(0.16)	0.73(0.23)	0.60(0.34)	0.82(0.15)	0.79(0.18)	0.70(0.28)
		PC	0.85(0.12)	0.91(0.10)	0.97(0.06)	0.75(0.14)	0.83(0.13)	0.92(0.09)	0.17(0.22)	0.10(0.18)	0.04(0.14)	0.15(0.22)	0.08(0.17)	0.03(0.11)
		TL-a	1.00(0.02)	1.00(0.00)	1.00(0.05)	0.99(0.04)	1.00(0.03)	1.00(0.00)	0.81(0.25)	0.84(0.24)	0.65(0.37)	0.77(0.28)	0.76(0.29)	0.73(0.31)
		TG-a	0.98(0.06)	1.00(0.02)	1.00(0.00)	1.00(0.02)	1.00(0.02)	1.00(0.00)	0.52(0.28)	0.42(0.31)	0.31(0.29)	0.50(0.28)	0.39(0.31)	0.32(0.29)
Homogeneous ($\rho=0.5$)	scenario I	sC	0.93(0.10)	0.93(0.10)	0.93(0.10)	0.93(0.10)	0.93(0.10)	0.93(0.10)	0.89(0.06)	0.89(0.06)	0.89(0.06)	0.89(0.06)	0.89(0.05)	0.89(0.06)
		OTC	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.63(0.18)	0.53(0.22)	0.45(0.26)	0.63(0.18)	0.53(0.21)	0.47(0.25)
		TC	0.99(0.04)	1.00(0.00)	1.00(0.00)	0.99(0.04)	1.00(0.00)	1.00(0.00)	0.64(0.18)	0.56(0.21)	0.45(0.26)	0.64(0.18)	0.56(0.21)	0.47(0.25)
		NTC	0.95(0.07)	0.98(0.05)	0.98(0.04)	0.95(0.07)	0.98(0.05)	0.98(0.05)	0.65(0.22)	0.53(0.26)	0.38(0.33)	0.65(0.22)	0.52(0.26)	0.39(0.33)
		PC	0.91(0.09)	0.98(0.05)	0.97(0.06)	0.90(0.09)	0.98(0.05)	0.97(0.06)	0.37(0.17)	0.34(0.20)	0.08(0.11)	0.36(0.17)	0.34(0.20)	0.08(0.11)
		TL-a	0.98(0.07)	0.97(0.09)	0.95(0.12)	0.98(0.08)	0.97(0.09)	0.95(0.12)	0.77(0.27)	0.62(0.36)	0.40(0.42)	0.76(0.28)	0.63(0.36)	0.41(0.41)
		TG-a	0.96(0.08)	0.99(0.05)	1.00(0.00)	0.95(0.08)	0.99(0.04)	1.00(0.00)	0.57(0.18)	0.41(0.22)	0.27(0.28)	0.57(0.18)	0.42(0.22)	0.27(0.28)
	scenario II	sC	0.94(0.10)	0.94(0.10)	0.94(0.10)	0.94(0.10)	0.94(0.10)	0.94(0.10)	0.89(0.06)	0.89(0.06)	0.89(0.06)	0.89(0.06)	0.89(0.06)	0.89(0.06)
		OTC	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.02)	1.00(0.00)	1.00(0.00)	0.63(0.19)	0.55(0.22)	0.46(0.26)	0.62(0.19)	0.55(0.21)	0.44(0.28)
		TC	1.00(0.01)	1.00(0.00)	1.00(0.00)	1.00(0.02)	1.00(0.00)	1.00(0.00)	0.58(0.22)	0.51(0.26)	0.44(0.27)	0.59(0.22)	0.50(0.26)	0.45(0.27)
		NTC	1.00(0.02)	1.00(0.00)	1.00(0.00)	1.00(0.02)	1.00(0.00)	1.00(0.00)	0.59(0.22)	0.59(0.21)	0.50(0.26)	0.58(0.22)	0.58(0.21)	0.47(0.28)
		PC	1.00(0.02)	1.00(0.00)	1.00(0.00)	1.00(0.02)	1.00(0.00)	1.00(0.00)	0.41(0.18)	0.43(0.18)	0.31(0.21)	0.42(0.18)	0.43(0.18)	0.26(0.22)
		TL-a	0.97(0.10)	0.96(0.11)	0.97(0.09)	0.97(0.10)	0.96(0.11)	0.97(0.09)	0.66(0.34)	0.48(0.36)	0.37(0.39)	0.66(0.35)	0.51(0.36)	0.40(0.39)
		TG-a	1.00(0.02)	1.00(0.00)	1.00(0.00)	1.00(0.02)	1.00(0.00)	1.00(0.00)	0.46(0.21)	0.40(0.22)	0.31(0.24)	0.47(0.21)	0.41(0.22)	0.32(0.25)
	Random Setting	sC	0.93(0.10)	0.93(0.10)	0.94(0.10)	0.93(0.10)	0.93(0.10)	0.93(0.10)	0.89(0.06)	0.89(0.06)	0.89(0.06)	0.88(0.08)	0.88(0.08)	0.88(0.07)
		OTC	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.63(0.18)	0.52(0.21)	0.46(0.25)	0.72(0.15)	0.59(0.22)	0.51(0.26)
		TC	1.00(0.01)	1.00(0.01)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.64(0.18)	0.52(0.22)	0.46(0.25)	0.72(0.15)	0.58(0.22)	0.51(0.26)
		NTC	0.91(0.12)	0.93(0.09)	0.95(0.07)	0.96(0.07)	0.97(0.07)	0.99(0.03)	0.79(0.17)	0.70(0.20)	0.60(0.28)	0.76(0.17)	0.69(0.22)	0.61(0.27)
		PC	0.77(0.14)	0.85(0.10)	0.91(0.07)	0.79(0.14)	0.85(0.13)	0.96(0.07)	0.40(0.27)	0.31(0.22)	0.24(0.22)	0.25(0.24)	0.14(0.18)	0.11(0.16)
		TL-a	0.96(0.10)	0.98(0.07)	0.96(0.11)	1.00(0.00)	1.00(0.05)	1.00(0.00)	0.77(0.29)	0.77(0.29)	0.56(0.39)	0.63(0.36)	0.67(0.36)	0.61(0.38)
		TG-a	0.93(0.10)	0.93(0.11)	0.98(0.05)	0.99(0.03)	1.00(0.00)	1.00(0.00)	0.57(0.25)	0.43(0.28)	0.29(0.30)	0.49(0.18)	0.42(0.21)	0.39(0.22)
Heterogeneous	scenario I	sC	0.93(0.10)	0.93(0.10)	0.93(0.10)	0.93(0.10)	0.92(0.10)	0.93(0.10)	0.88(0.08)	0.88(0.08)	0.88(0.07)	0.88(0.08)	0.88(0.08)	0.88(0.08)
		OTC	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.70(0.18)	0.58(0.22)	0.50(0.26)	0.70(0.17)	0.57(0.24)	0.50(0.27)
		TC	1.00(0.01)	1.00(0.00)	1.00(0.00)	1.00(0.01)	1.00(0.00)	1.00(0.00)	0.72(0.17)	0.59(0.21)	0.50(0.26)	0.72(0.16)	0.58(0.22)	0.50(0.27)
		NTC	0.99(0.03)	1.00(0.00)	1.00(0.00)	0.99(0.03)	1.00(0.00)	1.00(0.00)	0.63(0.24)	0.62(0.22)	0.49(0.30)	0.63(0.24)	0.61(0.23)	0.48(0.30)
		PC	0.99(0.04)	1.00(0.00)	1.00(0.00)	0.99(0.04)	1.00(0.00)	1.00(0.00)	0.36(0.21)	0.39(0.23)	0.19(0.16)	0.35(0.21)	0.38(0.23)	0.18(0.16)
		TL-a	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.81(0.24)	0.75(0.27)	0.51(0.39)	0.82(0.23)	0.73(0.28)	0.51(0.39)
		TG-a	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.54(0.16)	0.42(0.21)	0.37(0.23)	0.53(0.16)	0.42(0.21)	0.37(0.23)
	scenario II	sC	0.93(0.10)	0.93(0.10)	0.93(0.10)	0.93(0.10)	0.93(0.10)	0.93(0.10)	0.88(0.07)	0.88(0.08)	0.88(0.07)	0.88(0.08)	0.88(0.08)	0.88(0.08)
		OTC	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.72(0.15)	0.61(0.21)	0.53(0.25)	0.67(0.19)	0.52(0.27)	0.44(0.29)
		TC	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.04)	1.00(0.00)	1.00(0.00)	0.70(0.18)	0.63(0.21)	0.53(0.24)	0.71(0.18)	0.58(0.25)	0.44(0.30)
		NTC	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.63(0.21)	0.64(0.20)	0.60(0.20)	0.65(0.19)	0.61(0.21)	0.57(0.24)
		PC	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.49(0.18)	0.50(0.16)	0.44(0.16)	0.52(0.18)	0.46(0.16)	0.39(0.17)
		TL-a	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.77(0.29)	0.69(0.31)	0.58(0.39)	0.78(0.28)	0.72(0.29)	0.63(0.38)
		TG-a	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.59(0.16)	0.55(0.17)	0.46(0.20)	0.60(0.14)	0.54(0.17)	0.47(0.19)
	Random Setting	sC	0.93(0.10)	0.93(0.10)	0.92(0.10)	0.93(0.10)	0.93(0.10)	0.93(0.10)	0.88(0.08)	0.88(0.08)	0.88(0.08)	0.88(0.08)	0.88(0.08)	0.88(0.07)
		OTC	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.70(0.17)	0.57(0.23)	0.49(0.27)	0.72(0.15)	0.59(0.22)	0.51(0.26)
		TC	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	1.00(0.00)	0.70(0.17)	0.57(0.23)	0.49(0.27)	0.72(0.15)	0.58(0.22)	0.51(0.26)
		NTC	0.98(0.05)	0.99(0.04)	1.00(0.01)	0.96(0.07)	0.97(0.07)	0.99(0.03)	0.70(0.21)	0.66(0.24)	0.54(0.31)	0.76(0.17)	0.69(0.22)	0.61(0.27)
		PC	0.88(0.11)	0.94(0.09)	0.99(0.03)	0.79(0.14)	0.85(0.13)	0.96(0.07)	0.28(0.23)	0.19(0.19)	0.13(0.18)	0.25(0.24)	0.14(0.18)	0.11(0.16)
		TL-a	1.00(0.00)	1.00(0.05)	1.00(0.00)	1.00(0.00)	1.00(0.05)	1.00(0.00)	0.67(0.33)	0.66(0.37)	0.64(0.36)	0.63(0.36)	0.67(0.36)	0.61(0.38)
		TG-a	1.00(0.03)	1.00(0.00)	1.00(0.00)	0.99(0.03)	1.00(0.00)	1.00(0.00)	0.48(0.19)	0.41(0.22)	0.36(0.23)	0.49(0.18)	0.42(0.21)	0.39(0.22)

Note: sC: subCodalasso; OTC: Oracle-Trans-subCodalasso; TC: Trans-subCodalasso; NTC: Naive-Trans-subCodalasso; PC: Pooled-subCodalasso; TL-a: Trans-Lasso-alr; TG-a: Trans-GLM-alr.

Figures 13 illustrates the ℓ_2 estimation error and prediction error of different methods under single constraint and multiple constraints, respectively. It can be observed that the correct multiple linear constraints yield smaller ℓ_2 estimation errors and prediction errors than the single constraint case. This indicates that incorporating correct grouped constraints improves both estimation precision and prediction accuracy. Notably, the proposed (Oracle) Trans-subCodalasso outperforms other competing methods in most cases even under a single constraint.

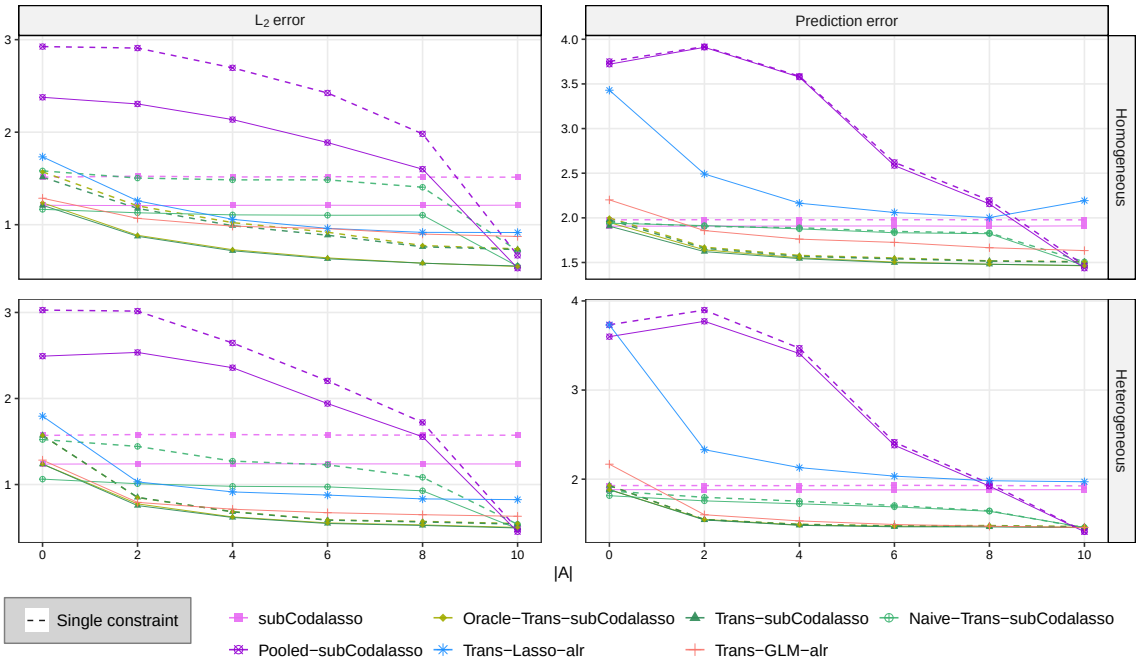

Figure 13: The average ℓ_2 estimation error and prediction error of single-constraint and multi-comstraints.

To further evaluate the impact of additional non-compositional covariates, we compare the proposed method and competing methods using $(\boldsymbol{y}^{(k)}, \mathbf{N}^{(k)}, \tilde{\tilde{\mathbf{Z}}}^{(k)})$ and only $(\boldsymbol{y}^{(k)}, \tilde{\tilde{\mathbf{Z}}}^{(k)})$.The results are shown in Figure 14. In the true data generating mechanism, both non-compositional variables $\mathbf{N}^{(k)}$ and compositional variables $\tilde{\tilde{\mathbf{Z}}}^{(k)}$ contribute to the response variable. Accordingly, neglecting $\mathbf{N}^{(k)}$ generally leads to omitted variable bias and loss of valid predictive information. Nevertheless, different methods possess distinct adaptability to the joint struc-

ture of $\mathbf{N}^{(k)}$ and $\tilde{\tilde{\mathbf{Z}}}^{(k)}$, resulting in disparate performance after excluding $\mathbf{N}^{(k)}$.

As shown in Figure 14 , the full-model-based (Oracle) Trans-subCodalasso performs the best under almost all settings. For the pooled method and Trans-Lasso/GLM-alr, both coefficient estimation errors and prediction errors decline when $\mathbf{N}^{(k)}$ is neglected. This indicates that such methods struggle to stably implement joint modeling of non-compositional and compositional variables. Removing $\mathbf{N}^{(k)}$ reduces model complexity and mitigates effect confounding as well as estimation variance, whose benefits outweigh the information loss caused by omitting $\mathbf{N}^{(k)}$. As for subCodalasso, excluding $\mathbf{N}^{(k)}$ reduces estimation errors yet elevates prediction errors. It demonstrates that excluding $\mathbf{N}^{(k)}$ improves the stability of parameter estimation, but meanwhile discards the valid predictive information related to the response contained in $\mathbf{N}^{(k)}$, thereby deteriorating the generalization prediction performance.

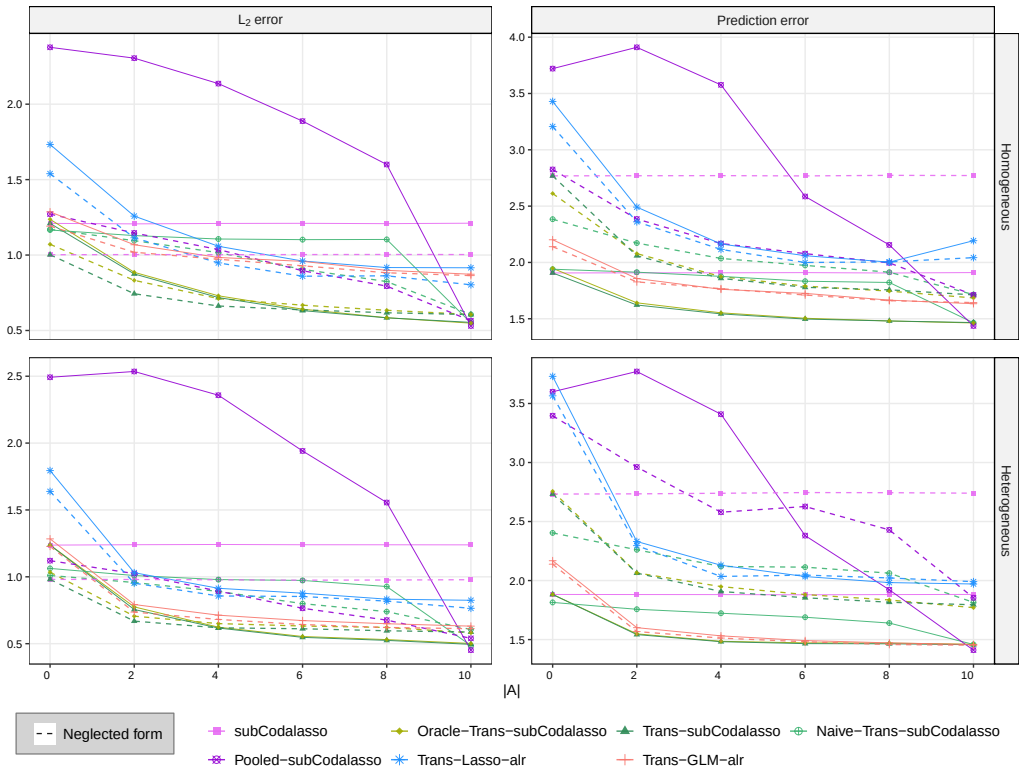

Figure 14: The average ℓ_2 estimation error and prediction error of full model and the model without non-compositional covariates.

In comparison, (Oracle) Trans-subCodalasso performs similarly to subCodalasso that only utilizes target source domains when the number of valid sources is small. As the number of valid sources increases, however, both estimation error and prediction error rise after excluding $\mathbf{N}^{(k)}$, which demonstrates that our method can effectively capture and exploit the true effects of $\mathbf{N}^{(k)}$. Eliminating $\mathbf{N}^{(k)}$ will introduce omitted variable bias and impair the predictive ability of the model. Therefore, the proposed method is more consistent with the actual data structure and achieves better overall performance under the full model containing both $\mathbf{N}^{(k)}$ and $\tilde{\tilde{\mathbf{Z}}}^{(k)}$.

S5.2.3 Effect of target-domain sample size

In this part,we take the dimensional of non-composition data $q' = 10$ and the dimension of compositional data $q = 100$, $K = 10$ and $|\mathcal{A}| \in \{0, 2, 4, 6, 8, 10\}$.We consider model shift in random setting ($d = 4$) and covariate shift in Homogenous design with $\rho = 0.2$. Figure 15 shows the ℓ_2 estimation errors as the target sample size varies over $\{50, 100, 200, 300, 400\}$, while each source sample size is fixed at 100.

Overall, the ℓ_2 estimation errors and prediction errors of all methods decline with the growth of target sample size. The proposed (Oracle) Trans-subCodalasso maintains stable and excellent performance, except in the case of small target sample size $n_0 = 50$ when $|\mathcal{A}| = 0$. Notably, under the same auxiliary source settings, as the target sample size increases, the performance of subCodalasso gradually approaches that of the oracle method. This indicates that when the target sample size is sufficiently large, the auxiliary effect of external source domains diminishes. When $|\mathcal{A}| = 0$, by comparing the gap between the pooled method and the Naive Trans-subCodalasso, it can be observed that the advantage of using the target domain for bias correction gradually increases as the target sample size

grows.

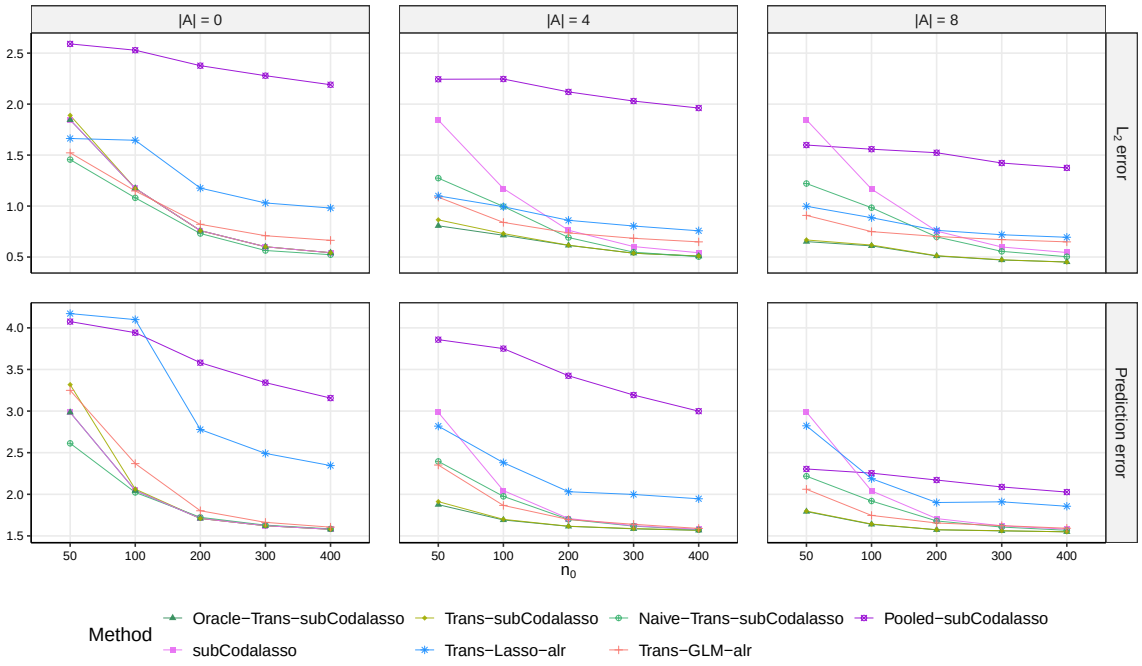

Figure 15: The average ℓ_2 estimation error and pediction error using seven methods with different n_0.

S5.3 Further results for the real-data study

S5.3.1 Dataset similarity analysis

To quantify the similarity among the four subgroup datasets, we conduct additional comparisons in terms of demographic distributions, microbial composition, and associations between demographic and compositional variables. Table 18 summarizes the distributions of BMI and age across the four subgroup datasets. The mean BMI values are broadly comparable across groups, whereas the normal groups are slightly older than the UC groups.

Table 18: Summary statistics of BMI and age across the four subgroup datasets. Values are presented as mean ± standard deviation.

Dataset	n	BMI (mean ± sd)	Age (mean ± sd)
UC female	72	23.79 ± 3.76	43.15 ± 10.81
Normal female	318	25.19 ± 5.88	47.86 ± 13.14
Normal male	310	24.97 ± 4.86	46.08 ± 15.48
UC male	39	25.97 ± 4.19	43.62 ± 10.43

We further assess microbial similarity using two measures: the pairwise CLR center distance and the pairwise Spearman correlation of genus prevalence profiles. As shown in Figure 16, the genus prevalence profiles are highly correlated across the four subgroup datasets, with pairwise correlations ranging from 0.87 to 0.98, suggesting substantial similarity in the overall genus occurrence patterns. Meanwhile, the CLR center distances indicate that the average microbial compositions are not identical across groups, especially for comparisons involving UC female.

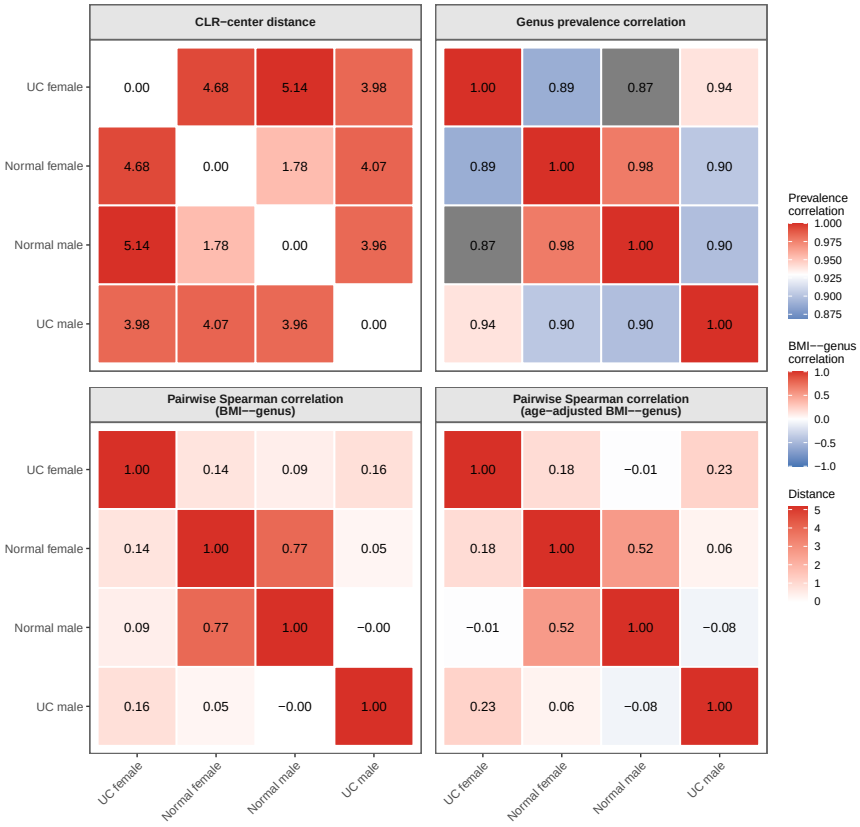

Figure 16: Similarity analysis among the four subgroup datasets.

We also compare the BMI-related microbial association patterns across subgroup datasets. For each subgroup dataset, we first compute an association vector consisting of the Spearman correlations between BMI and the 81 genera. To account for the potential effect of age, we further construct an age-adjusted association vector by residualizing both BMI and

each genus abundance with respect to age and then calculating the Spearman correlations between the corresponding residuals. As shown in Figure 16, the BMI-related association patterns are more heterogeneous than the genus prevalence profiles. The strongest similarity is observed between normal female and normal male, whereas comparisons involving UC groups generally show weaker correlations. After age adjustment, a similar pattern remains.

Overall, these results indicate that the four subgroup datasets are related but not fully homogeneous. The high genus prevalence correlations suggest the presence of shared microbial information, supporting the use of transfer learning. At the same time, the observed differences in CLR centers and marginal BMI–genus association patterns suggest that information borrowing should be performed adaptively rather than by directly pooling all subgroup datasets. This motivates the use of Trans-subCodaLasso with transferable source detection in the subsequent analysis.

S5.3.2 Source selection results

To further interpret the prediction results, we examine the source-selection behavior of Trans-subCodalasso across target subgroups and modeling settings.

As shown in Table 19, the complete set of all candidate sources is most frequently selected in all settings, with selection frequencies ranging from 0.53 to 0.79. This helps explain why Trans-subCodalasso often achieves similar performance to naive Trans-subCodalasso, since the adaptive method frequently uses all available sources. Empty-set selection is uncommon overall, suggesting that the method usually identifies at least one useful source population. UC male shows relatively higher empty-set frequencies, around 0.11–0.15 across settings, which may be partly related to the absence of a clearly close source population in the similarity heatmaps (Figure 16). However, UC female also shows some variation in

Table 19: Source-selection summary of Trans-subCodalasso across target and model settings

Target	Model setting	Empty-set freq.	Most frequent source set	Selection freq.
UC female	Subcomposition + age	0.04	All	0.66
	Subcomposition	0.11	All	0.68
	Composition + age	0.03	All	0.70
	Composition	0.07	All	0.69
Normal female	Subcomposition + age	0.09	All	0.68
	Subcomposition	0.02	All	0.79
	Composition + age	0.08	All	0.53
	Composition	0.05	All	0.66
Normal male	Subcomposition + age	0.00	All	0.69
	Subcomposition	0.04	All	0.79
	Composition + age	0.01	All	0.62
	Composition	0.02	All	0.67
UC male	Subcomposition + age	0.15	All	0.70
	Subcomposition	0.14	All	0.64
	Composition + age	0.11	All	0.68
	Composition	0.15	All	0.67

Note: "Empty-set freq." denotes the proportion of repetitions in which no source subgroup was selected by Trans-subCodalasso. "Most frequent source set" denotes the exact combination of source subgroups that was selected most often for a given target subgroup and data setting. "All" indicates that all available candidate source subgroups were selected. "Selection freq." denotes the selection frequency of this most frequent source set.

empty-set selection, ranging from 0.03 to 0.11 depending on the data setting. This indicates that heatmap-based subgroup similarity alone cannot fully explain source-selection behavior. Other factors, including target sample size, predictive-signal stability and covariate adjustment, may also affect whether transfer learning is beneficial. Overall, these results suggest that transfer learning is generally useful, but that transferability varies across target subgroups and modeling settings.

S5.3.3 Variable selection results

Beyond prediction performance, we analyze the microbial features selected by the proposed method to assess its interpretability. For each target subgroup and modeling setting, genera are ranked by selection probability, with ties broken by the mean absolute coefficient magnitude. The top 10 genera selected by Trans-subCodalasso for each target subgroup and model setting are reported in Table 22.

The results indicate that *Subdoligranulum* is the most stable genus, being selected across all target subgroups and modeling settings. Several other genera, including *Faecalibacterium*, *Coprococcus*, *Oscillibacter*, *Collinsella*, and *Coprobacter*, are repeatedly selected in most

settings, suggesting a recurrent core set of BMI-related microbial signals.

Some of these repeatedly selected genera are biologically interpretable. In particular, *Faecalibacterium*, *Coprococcus*, and *Subdoligranulum* have been reported as butyrate-producing or butyrate-associated taxa, whereas *Oscillibacter* has been linked to valerate production, thereby potentially influencing host energy metabolism (**???**). These genera also belong to the phylum Firmicutes, which has been increasingly recognized for its role in dietary fiber fermentation, short-chain fatty acid production, and host metabolic homeostasis (**?**). Moreover, previous studies have reported that reduced abundance of *Faecalibacterium* is associated with obesity, and a specific *Faecalibacterium* strain is found to be negatively associated with BMI and to alleviate high-fat-diet-induced obesity in experimental models (**?**). Given the established role of short-chain fatty acids in host energy homeostasis, inflammation, insulin sensitivity, and metabolic regulation (**??**), these genera represent biologically plausible microbial features associated with BMI-related phenotypes.

Nevertheless, the selected variables are not identical across target subgroups and model settings. Subgroup-specific selection patterns exist between healthy and disease-related subgroups. Results with and without age adjustment are largely consistent in core microbial signatures, though age adjustment alters the selection preference of certain taxa. Similarly, subcompositional and compositional model settings share considerable overlapping signals, while still exhibiting slight differences in microbial feature selection. These results suggest that, in addition to a stable core set of BMI-related microbial signals, there may also be target- and setting-specific microbial associations, especially in disease-related subgroups.

To further compare the agreement among methods, we add a Jaccard similarity analysis of the selected genera. For each method, we define the selected set as the genera with selection

Table 20: Jaccard similarity of selected genera between Trans-subCodalasso and competing methods under different model settings.

Target	Model setting	SubCoda-lasso	Naive-Trans-subCodalasso	Pooled-subCodalasso	Trans-GLM-alr	Trans-Lasso-alr
UC female	Subcomposition + age	0.21	0.96	0.73	0.10	0.19
	Subcomposition	0.50	1.00	0.48	0.20	0.31
	Composition + age	0.24	1.00	0.71	0.20	0.36
	Composition	0.46	1.00	0.48	0.37	0.28
Normal female	Subcomposition + age	0.71	0.96	0.64	0.08	0.11
	Subcomposition	0.71	1.00	0.65	0.20	0.23
	Composition + age	0.76	0.95	0.50	0.31	0.37
	Composition	0.65	1.00	0.61	0.29	0.23
Normal male	Subcomposition + age	0.45	0.86	0.75	0.12	0.12
	Subcomposition	0.52	1.00	0.78	0.12	0.13
	Composition + age	0.47	0.93	0.73	0.21	0.32
	Composition	0.48	0.94	0.77	0.23	0.27
UC male	Subcomposition + age	0.92	0.95	0.37	0.53	0.89
	Subcomposition	0.29	0.95	0.76	0.09	0.25
	Composition + age	0.90	0.93	0.36	0.21	0.89
	Composition	0.40	0.94	0.67	0.27	0.25

Note: The selected set is defined as the genera with selection frequency greater than 0.5 over 100 repetitions. The Jaccard similarity between two selected sets A and B is defined as $J(A, B) = \frac{|A \cap B|}{|A \cup B|}$, where a larger value indicates a higher degree of overlap.

frequency greater than 0.5 over 100 repetitions. We then compute the Jaccard similarity between the selected set of each competing method and that of Trans-subCodaLasso. The results are summarized in Table 20.

The results in Table 20 show that Naive Trans-subCodaLasso has the highest overlap with Trans-subCodalasso, with Jaccard similarities ranging from 0.86 to 1.00. This indicates that the two methods select highly similar genera, which is consistent with the source-selection results that all candidate sources are frequently selected by Trans-subCodalasso. Pooled-subCodalasso shows moderate overlap, suggesting that directly pooling all available data may change the selected genera. SubCodalasso exhibits target-dependent overlap, indicating that source information can affect variable selection compared with using the target data alone. In contrast, the alr-based methods generally show lower overlap, especially Trans-GLM-alr, suggesting that different compositional modeling strategies may lead to different selected genera.

Table 21: Means and standard errors (in parentheses) of prediction errors for six methods under different model settings based on 100 replicates.

Target	Model setting	subCoda Lasso	NaiveTrans-subCodalasso	Pooled-subCodalasso	Trans-subCodalasso	Trans-GLM-alr	Trans-Lasso-alr
UC female	Subcomposition + age	12.78 (2.85)	11.64 (2.78)	12.90 (2.32)	11.56 (2.63)	22.18 (6.24)	23.87 (6.83)
	Subcomposition	12.89 (3.07)	11.90 (2.86)	15.36 (3.00)	11.86 (2.77)	30.76 (7.94)	44.38 (13.92)
	Composition + age	12.86 (3.03)	11.54 (2.67)	13.04 (2.40)	11.52 (2.64)	23.05 (4.67)	23.60 (7.97)
	Composition	13.04 (3.21)	12.10 (2.91)	15.55 (3.10)	11.88 (2.81)	27.05 (7.86)	39.05 (14.68)
Normal female	Subcomposition + age	28.34 (3.66)	28.13 (3.58)	29.43 (3.68)	28.07 (3.58)	33.93 (4.08)	35.31 (4.75)
	Subcomposition	31.38 (3.69)	31.03 (3.58)	32.14 (3.84)	31.00 (3.65)	54.65 (9.88)	55.25 (9.14)
	Composition + age	28.32 (3.69)	28.03 (3.62)	29.25 (3.67)	27.98 (3.65)	37.34 (4.85)	36.68 (5.15)
	Composition	31.71 (3.84)	31.40 (3.65)	32.16 (3.85)	31.29 (3.68)	50.65 (9.12)	53.04 (8.57)
Normal male	Subcomposition + age	17.81 (2.58)	17.49 (2.51)	18.04 (2.64)	17.46 (2.53)	33.05 (4.31)	34.95 (4.72)
	Subcomposition	20.32 (2.95)	20.06 (2.78)	20.63 (2.96)	20.04 (2.83)	52.54 (11.11)	56.90 (11.94)
	Composition + age	17.95 (2.57)	17.50 (2.50)	17.98 (2.62)	17.49 (2.52)	35.23 (4.67)	34.92 (7.03)
	Composition	20.56 (2.92)	20.27 (2.78)	20.71 (2.96)	20.24 (2.85)	47.41 (9.99)	52.18 (11.03)
UC male	Subcomposition + age	19.68 (6.92)	20.46 (7.04)	17.80 (7.73)	19.80 (7.00)	49.37 (18.23)	64.50 (37.87)
	Subcomposition	18.27 (6.32)	18.33 (6.39)	16.93 (6.23)	17.84 (5.89)	51.54 (17.17)	66.01 (33.44)
	Composition + age	21.44 (8.88)	20.63 (8.47)	18.16 (7.94)	20.11 (8.97)	53.24 (22.16)	55.17 (23.76)
	Composition	19.28 (7.17)	18.80 (7.28)	17.11 (6.20)	18.38 (6.80)	30.82 (14.40)	45.83 (20.70)

Bibliography

Bertsekas, D. (1996). *Constrained Optimization and Lagrange Multiplier Methods.* Athena Scientific.

Bertsimas, D., King, A., and Mazumder, R. (2016). Best subset selection via a modern optimization lens. *The Annals of Statistics*, 44(2):813 – 852.

He, Y., Li, Q., Hu, Q., and Liu, L. (2022). Transfer learning in high-dimensional semi-parametric graphical models with application to brain connectivity analysis. *Statistics in Medicine*, 41(21):4112–4129.

He, Z., Sun, Y., Liu, J., and Li, R. (2024). Transfusion: Covariate-shift robust transfer learning for high-dimensional regression. In Dasgupta, S., Mandt, S., and Li, Y., editors, *INTERNATIONAL CONFERENCE ON ARTIFICIAL INTELLIGENCE AND STATISTICS, VOL 238*, volume 238 of *Proceedings of Machine Learning Research*. 27th International Conference on Artificial Intelligence and Statistics (AISTATS), Valencia, SPAIN, MAY 02-04, 2024.

Table 22: Top 10 selected genera by Trans-subCodalasso under different model settings.

Target	Model setting	Top 10 selected genera by Trans-subCodaLasso
UC female	Subcomposition + age	*Subdoligranulum*; *Faecalibacterium*; *Oscillibacter*; *Coprococcus*; *Parasutterella*; *Burkholderiales_noname*; *Prevotella*; *Escherichia*; *Parabacteroides*; *Coprobacter*
	Subcomposition	*Subdoligranulum*; *Faecalibacterium*; *Erysipelotrichaceae_noname*; *Klebsiella*; *Burkholderiales_noname*; *Sutterella*; *Parabacteroides*; *Oscillibacter*; *Collinsella*; *Coprococcus*
	Composition + age	*Subdoligranulum*; *Faecalibacterium*; *Burkholderiales_noname*; *Oscillibacter*; *Coprococcus*; *Rothia*; *Coprobacter*; *Streptococcus*; *Eggerthella*; *Erysipelotrichaceae_noname*
	Composition	*Subdoligranulum*; *Faecalibacterium*; *Erysipelotrichaceae_noname*; *Burkholderiales_noname*; *Oscillibacter*; *Coprococcus*; *Eggerthella*; *Rothia*; *Pseudoflavonifractor*; *Coprobacter*
Normal female	Subcomposition + age	*Oscillibacter*; *Subdoligranulum*; *Burkholderiales_noname*; *Collinsella*; *Streptococcus*; *Parasutterella*; *Coprococcus*; *Lactococcus*; *Dorea*; *Faecalibacterium*
	Subcomposition	*Subdoligranulum*; *Faecalibacterium*; *Collinsella*; *Eggerthella*; *Oscillibacter*; *Coprococcus*; *Burkholderiales_noname*; *Sutterella*; *Coprobacter*; *Erysipelotrichaceae_noname*
	Composition + age	*Subdoligranulum*; *Oscillibacter*; *Rothia*; *Burkholderiales_noname*; *Coprobacter*; *Coprococcus*; *Parasutterella*; *Cellulophaga*; *Faecalibacterium*; *Dorea*
	Composition	*Subdoligranulum*; *Coprococcus*; *Eggerthella*; *Oscillibacter*; *Faecalibacterium*; *Collinsella*; *Coprobacter*; *Rothia*; *Burkholderiales_noname*; *Sutterella*
Normal male	Subcomposition + age	*Faecalibacterium*; *Subdoligranulum*; *Coprococcus*; *Oscillibacter*; *Streptococcus*; *Burkholderiales_noname*; *Collinsella*; *Eggerthella*; *Coprobacter*; *Parasutterella*
	Subcomposition	*Faecalibacterium*; *Subdoligranulum*; *Coprococcus*; *Collinsella*; *Burkholderiales_noname*; *Eggerthella*; *Oscillibacter*; *Pseudoflavonifractor*; *Parabacteroides*; *Coprobacter*
	Composition + age	*Subdoligranulum*; *Coprococcus*; *Rothia*; *Faecalibacterium*; *Burkholderiales_noname*; *Oscillibacter*; *Coprobacter*; *Eggerthella*; *Streptococcus*; *Collinsella*
	Composition	*Faecalibacterium*; *Subdoligranulum*; *Coprococcus*; *Eggerthella*; *Burkholderiales_noname*; *Oscillibacter*; *Rothia*; *Coprobacter*; *Collinsella*; *Parabacteroides*
UC male	Subcomposition + age	*Megamonas*; *Gemella*; *Erysipelotrichaceae_noname*; *Dialister*; *Collinsella*; *Barnesiella*; *Streptococcus*; *Subdoligranulum*; *Parvimonas*; *Phascolarctobacterium*
	Subcomposition	*Erysipelotrichaceae_noname*; *Streptococcus*; *Collinsella*; *Subdoligranulum*; *Faecalibacterium*; *Coprobacter*; *Coprococcus*; *Burkholderiales_noname*; *Parabacteroides*; *Oscillibacter*
	Composition + age	*Erysipelotrichaceae_noname*; *Megamonas*; *Enterococcus*; *Streptococcus*; *Dialister*; *Collinsella*; *Coprobacter*; *Parvimonas*; *Subdoligranulum*; *Coprococcus*
	Composition	*Erysipelotrichaceae_noname*; *Subdoligranulum*; *Faecalibacterium*; *Burkholderiales_noname*; *Collinsella*; *Coprobacter*; *Oscillibacter*; *Rothia*; *Coprococcus*; *Eggerthella*

Note: For each target subgroup and model setting, genera were ranked by their selection probabilities, with ties broken by the mean absolute coefficient magnitude.

Huang, J., Jiao, Y., Liu, Y., and Lu, X. (2018). A constructive approach to l_0 penalized regression. *Journal of Machine Learning Research*, 19(10):1–37.

Jin, J., Yan, J., Aseltine, R. H., and Chen, K. (2024). Transfer learning with large-scale quantile regression. *Technometrics*, 66(3):381–393.

Li, S., Cai, T. T., and Li, H. (2022). Transfer Learning for High-Dimensional Linear Regression: Prediction, Estimation and Minimax Optimality. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 84(1):149–173.

Li, S., Zhang, L., Cai, T. T., and Li, H. (2024). Estimation and inference for high-dimensional generalized linear models with knowledge transfer. *Journal of the American Statistical Association*, 119(546):1274–1285. PMID: 38948492.

Lin, W., Shi, P., Feng, R., and Li, H. (2014). Variable selection in regression with compositional covariates. *Biometrika*, 101(4):785–797.

Mishra, A. and Müller, C. L. (2022). Robust regression with compositional covariates. *Computational Statistics & Data Analysis*, 165:107315.

Shi, P., Zhang, A., and Li, H. (2016). Regression analysis for microbiome compositional data. *The Annals of Applied Statistics*, 10:1019–1040.

Tian, Y. and Feng, Y. (2023). Transfer learning under high-dimensional generalized linear models. *Journal of the American Statistical Association*, 118(544):2684–2697.

Vershynin, R. (2018). *High-Dimensional Probability: An Introduction with Applications in Data Science*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.

Zhang, S., Wang, H., and Lin, W. (2025). Care: Large precision matrix estimation for compositional data. *Journal of the American Statistical Association*, 120(549):305–317.