Supplementary Material for "Deconfounding via Profiled Transfer Learning"

S.1 Detailed Proofs and Technique Lemmas

S.1.1 Proof of Theorem 1

Proof of Theorem 1. First, we define several events:

$$\left\{\left\|\frac{1}{n_t}X_t^\top X_t - \Sigma_X\right\|_\infty \le C_x\sqrt{\frac{\log p}{n_t}}\right\}, \tag{$\mathcal{G}_1$}$$

$$\left\{\left\|\frac{1}{n_s}\sum_{k=1}^K X_s^{(k)\top}X_s^{(k)} - \Sigma_X\right\|_\infty \le C_x\sqrt{\frac{\log p}{n_s}}\right\}, \tag{$\mathcal{G}_2$}$$

$$\left\{\left\|\frac{1}{n_s}\sum_{k=1}^K n_s^{(k)}\left(\frac{1}{n_s^{(k)}}X_s^{(k)\top}X_s^{(k)} - \Sigma_X\right)\left(\beta_s^{(k)} - \frac{1}{n_s}\sum_{k=1}^K n_s^{(k)}\beta_s^{(k)}\right)\right\|_\infty \le C_\delta s_\delta\sqrt{\frac{\log p}{n_s}}\right\}, \tag{$\mathcal{G}_3$}$$

$$\left\{\left\|\frac{1}{n_t}f(H_t)^\top H_t - \mathbb{E}_H f(H)^\top H\right\|_{2,\infty} \le C_H\sqrt{\frac{q\log p}{n_t}}\right\}, \tag{$\mathcal{G}_4$}$$

$$\left\{\left\|\frac{1}{n_s}\sum_{k=1}^K f(H_s^{(k)})^\top H_s^{(k)} - \mathbb{E}_H f(H)^\top H\right\|_{2,\infty} \le C_H\sqrt{\frac{q\log p}{n_s}}\right\}, \tag{$\mathcal{G}_5$}$$

$$\left\{\left\|\frac{1}{n_t}E_t^\top H_t\right\|_{2,\infty} \le C_H\sqrt{\frac{q\log p}{n_t}}\right\}, \tag{$\mathcal{G}_6$}$$

$$\left\{\left\|\frac{1}{n_s}\sum_{k=1}^K E_s^{(k)\top}H_s^{(k)}\right\|_{2,\infty} \le C_H\sqrt{\frac{q\log p}{n_s}}\right\}, \tag{$\mathcal{G}_7$}$$

$$\left\{\left\|\frac{1}{n_t}X_t^\top \varepsilon_t\right\|_\infty \le C_\varepsilon\sqrt{\frac{\log p}{n_t}}\right\}, \tag{$\mathcal{G}_8$}$$

$$\left\{\left\|\frac{1}{n_s}\sum_{k=1}^K X_s^{(k)\top}\varepsilon_s^{(k)}\right\|_\infty \le C_\varepsilon\sqrt{\frac{\log p}{n_s}}\right\}, \tag{$\mathcal{G}_9$}$$

Here we define the two-infinite norm $\|A\|_{2,\infty}$ of a matrix $A \in \mathbb{R}^{m_1\times m_2}$ as the maximum of the l_2 norm among the row vectors of A, i.e., $\|A\|_{2,\infty} = \max_{1\le i\le m_1}\|A_{i,:}\|_2$.

Define the event $\mathcal{G} = \cap_{j=1}^{9} \mathcal{G}_j$. From Lemma S.1, the event $\mathcal{G}$ happens with a probability

$$\mathbb{P}(\mathcal{G}) \geq 1 - p^{-c}.$$

Recall the definition of $\hat{Z}_s^{(k)}$ as

$$\hat{Z}_s^{(k)} = Y_s^{(k)} - X_s^{(k)} \hat{\beta}_{\text{init}} = X_s^{(k)} \left(\beta_s^{(k)} - \hat{\beta}_{\text{init}} \right) + H_s^{(k)} \phi_s^{(k)} + \varepsilon_s^{(k)}.$$

Define that $b_s^{(k)} = \left(X_s^{(k)\top} X_s^{(k)} \right)^{\dagger} X_s^{(k)\top} H_s^{(k)} \phi$, so

$$\begin{aligned}
& \left\| \frac{1}{n_t} X_t^\top \hat{Z}_t - \frac{1}{n_s} \sum_{k=1}^{K} X_s^{(k)\top} \hat{Z}_s^{(k)} \right\|_\infty \\
= & \min_{Z_t \in \mathbb{R}^{n_t}} \left\| \frac{1}{n_t} X_t^\top Z_t - \frac{1}{n_s} \sum_{k=1}^{K} X_s^{(k)\top} \hat{Z}_s^{(k)} \right\|_\infty \\
\leq & \min_{\omega \in \mathbb{R}^p} \left\| \frac{1}{n_t} X_t^\top X_t \omega - \frac{1}{n_s} \sum_{k=1}^{K} X_s^{(k)\top} \hat{Z}_s^{(k)} \right\|_\infty \\
= & \min_{\omega \in \mathbb{R}^p} \left\| \frac{1}{n_t} X_t^\top X_t \left(\beta_s - \hat{\beta}_{\text{init}} + \omega \right) - \frac{1}{n_s} \sum_{k=1}^{K} X_s^{(k)\top} \left[X_s^{(k)} \left(\beta_s^{(k)} - \hat{\beta}_{\text{init}} + b_s^{(k)} \right) - \varepsilon_s^{(k)} \right] \right\|_\infty,
\end{aligned}$$

where the first equality follows from the definition of $\hat{Z}_t$; the second inequality holds because the space $\{X_t \omega,\ \omega \in \mathbb{R}^p\}$ is a subset of $\mathbb{R}^{n_t}$; the third equality applies the definitions of $\hat{Z}_s^{(k)}$ and $b_s^{(k)}$, and involves a change of variable from ω to $\omega + \beta_s - \hat{\beta}_{\text{init}}$.

With the definition of β_s, we have

$$\beta_s^{(k)} - \hat{\beta}_{\text{init}} = \left(\beta_s^{(k)} - \beta_s \right) + \left(\beta_s - \hat{\beta}_{\text{init}} \right) = \left(\beta_s^{(k)} - \frac{1}{n_s} \sum_{k=1}^{K} n_s^{(k)} \beta_s^{(k)} \right) + \left(\beta_s - \hat{\beta}_{\text{init}} \right).$$

So

$$\left\|\frac{1}{n_t}X_t^\top \hat{Z}_t - \frac{1}{n_s}\sum_{k=1}^K X_s^{(k)\top}\hat{Z}_s^{(k)}\right\|_\infty \leq \left\|\left(\frac{1}{n_t}X_t^\top X_t - \frac{1}{n_s}\sum_{k=1}^K X_s^{(k)\top}X_s^{(k)}\right)\left(\beta_s - \hat{\beta}_{\text{init}}\right)\right\|_\infty \tag{B_1}$$

$$+ \left\|\frac{1}{n_s}\sum_{k=1}^K X_s^{(k)\top}X_s^{(k)}\left(\beta_s^{(k)} - \frac{1}{n_s}\sum_{k=1}^K n_s^{(k)}\beta_s^{(k)}\right)\right\|_\infty \tag{B_2}$$

$$+ \min_{\omega\in\mathbb{R}^p}\left\|\frac{1}{n_t}X_t^\top X_t\omega - \frac{1}{n_s}\sum_{k=1}^K X_s^{(k)\top}X_s^{(k)}b_s^{(k)}\right\|_\infty \tag{B_3}$$

$$+ \left\|\frac{1}{n_s}\sum_{k=1}^K X_s^{(k)\top}\varepsilon_s^{(k)}\right\|_\infty. \tag{B_4}$$

On the right-hand side of the above inequality, we denote the four terms as B_1, B_2, B_3, and B_4, respectively. Term B_1 quantifies the error due to the difference in empirical covariance matrices of covariates X between the target and source data. Term B_2 arises from variability across different source models. Term B_3 represents the transfer loss associated with the confounding effects from the source models to the target model. Term B_4 accounts for the noise in the source data.

Bound Term B_1

We have

$$\begin{aligned}
&\left\|\left(\frac{1}{n_t}\tilde{X}_t^\top X_t - \frac{1}{n_s}\sum_{k=1}^K X_s^{(k)\top}X_s^{(k)}\right)\left(\beta_s - \hat{\beta}_{\text{init}}\right)\right\|_\infty \\
\leq\ & \left\|\frac{1}{n_t}\tilde{X}_t^\top X_t - \frac{1}{n_s}\sum_{k=1}^K X_s^{(k)\top}X_s^{(k)}\right\|_\infty \left\|\beta_s - \hat{\beta}_{\text{init}}\right\|_1 \\
\leq\ & \left(\left\|\frac{1}{n_t}\tilde{X}_t^\top \tilde{X}_t - \Sigma_X\right\|_\infty + \left\|\frac{1}{n_s}\sum_{k=1}^K X_s^{(k)\top}X_s^{(k)} - \Sigma_X\right\|_\infty\right)\left\|\beta_s - \hat{\beta}_{\text{init}}\right\|_1 \\
\leq\ & \left(C_x\left(\sqrt{\frac{\log p}{n_t}} + \sqrt{\frac{\log p}{n_s}}\right)\right)\left\|\beta_s - \hat{\beta}_{\text{init}}\right\|_1 \\
\leq\ & C_s\sqrt{\frac{\log p}{n_t}}.
\end{aligned} \tag{S.1}$$

Here the third inequality is from $\mathcal{G}_1$, $\mathcal{G}_2$ and the last equality is from $n_t \ll n_s$.

Bound Term B_2

We have

$$
\begin{aligned}
&\left\| \frac{1}{n_s}\sum_{k=1}^K X_s^{(k)\top}X_s^{(k)}\left(\beta_s^{(k)}-\frac{1}{n_s}\sum_{k=1}^K n_s^{(k)}\beta_s^{(k)}\right)\right\|_\infty \\
=& \left\| \frac{1}{n_s}\sum_{k=1}^K n_s^{(k)}\left(\frac{1}{n_s^{(k)}}X_s^{(k)\top}X_s^{(k)}-\Sigma_X\right)\left(\beta_s^{(k)}-\frac{1}{n_s}\sum_{k=1}^K n_s^{(k)}\beta_s^{(k)}\right)\right\|_\infty \\
\leq& C_\delta s_\delta\sqrt{\frac{\log p}{n_s}} = o\left(\sqrt{\frac{\log p}{n_t}}\right).
\end{aligned} \tag{S.2}
$$

Here the first equality is from

$$
\sum_{k=1}^K n_s^{(k)}\left(\beta_s^{(k)}-\frac{1}{n_s}\sum_{k=1}^K n_s^{(k)}\beta_s^{(k)}\right)=0.
$$

The second inequality is from $\mathcal{G}_3$.

Bound Term B_3

We have

$$
\begin{aligned}
&\min_{\omega\in\mathbb{R}^p}\left\| \frac{1}{n_t}X_t^\top X_t\omega-\frac{1}{n_s}\sum_{k=1}^K X_s^{(k)\top}X_s^{(k)}b_s^{(k)}\right\|_\infty \\
\leq& \left\| \frac{1}{n_t}X_t^\top X_t b_t-\frac{1}{n_s}\sum_{k=1}^K X_s^{(k)\top}X_s^{(k)}b_s^{(k)}\right\|_\infty \\
=& \left\| \frac{1}{n_t}X_t^\top H_t\phi_t-\frac{1}{n_s}\sum_{k=1}^K X_s^{(k)\top}H_s^{(k)}\phi_s^{(k)}\right\|_\infty \\
\leq& \left\| \frac{1}{n_t}f(H_t)^\top H_t\phi_t-\frac{1}{n_s}\sum_{k=1}^K f(H_s^{(k)})^\top H_s^{(k)}\phi_s^{(k)}\right\|_\infty + \left\|\frac{1}{n_t}E_t^\top H_t\phi_t\right\|_\infty + \left\|\frac{1}{n_s}\sum_{k=1}^K E_s^{(k)\top}H_s^{(k)}\phi_s^{(k)}\right\|_\infty.
\end{aligned} \tag{S.3}
$$

Here the first inequality is because that we take the value $\omega=b_t$, the second equality is from the definition of b_t and $b_s^{(k)}$ and the third inequality uses the confounding structure $X=f(H)+E$.

For the first term in (S.3), we have

$$
\begin{aligned}
&\left\|\frac{1}{n_t} f(H_t)^\top H_t \phi_t - \frac{1}{n_s}\sum_{k=1}^K f(H_s^{(k)})^\top H_s^{(k)} \phi_s^{(k)}\right\|_\infty \\
\le& \left\|\left(\frac{1}{n_t} f(H_t)^\top H_t - \frac{1}{n_s}\sum_{k=1}^K f(H_s^{(k)})^\top H_s^{(k)}\right)\phi_t\right\|_\infty + \left\|\frac{1}{n_s}\sum_{k=1}^K f(H_s^{(k)})^\top H_s^{(k)}\left(\phi_s^{(k)} - \phi_t\right)\right\|_\infty \\
\le& \left\|\left(\frac{1}{n_t} f(H_t)^\top H_t - \mathbb{E}_H f(H)^\top H\right)\phi_t\right\|_\infty + \left\|\left(\frac{1}{n_s}\sum_{k=1}^K f(H_s^{(k)})^\top H_s^{(k)} - \mathbb{E}_H f(H)^\top H\right)\phi_t\right\|_\infty \\
&\qquad + \left\|\frac{1}{n_s}\sum_{k=1}^K f(H_s^{(k)})^\top H_s^{(k)}\left(\phi_s^{(k)} - \phi_t\right)\right\|_\infty \\
\le& C_H\left(\sqrt{\frac{q\log p}{n_t}} + \sqrt{\frac{q\log p}{n_s}}\right)\|\phi_t\|_2 + o\left(\sqrt{\frac{\log p}{n_t}}\right),
\end{aligned} \tag{S.4}
$$

with

$$
\begin{aligned}
\left\|\left(\frac{1}{n_t} f(H_t)^\top H_t - \mathbb{E}_H f(H)^\top H\right)\phi_t\right\|_\infty &\le \left\|\frac{1}{n_t} f(H_t)^\top H_t - \mathbb{E}_H f(H)^\top H\right\|_{2,\infty}\|\phi_t\|_2 \\
&\le C_H\sqrt{\frac{q\log p}{n_t}}\|\phi_t\|_2
\end{aligned}
$$

in the event $\mathcal{G}_4$,

$$
\begin{aligned}
&\left\|\left(\frac{1}{n_s}\sum_{k=1}^K f(H_s^{(k)})^\top H_s^{(k)} - \mathbb{E}_H f(H)^\top H\right)\phi_t\right\|_\infty \\
\le& \left\|\frac{1}{n_s}\sum_{k=1}^K f(H_s^{(k)})^\top H_s^{(k)} - \mathbb{E}_H f(H)^\top H\right\|_{2,\infty}\|\phi_t\|_2 \\
\le& C_H\sqrt{\frac{q\log p}{n_s}}\|\phi_t\|_2
\end{aligned}
$$

in the event $\mathcal{G}_5$ and

$$
\begin{aligned}
&\left\|\frac{1}{n_s}\sum_{k=1}^K f(H_s^{(k)})^\top H_s^{(k)}\left(\phi_s^{(k)} - \phi_t\right)\right\|_\infty \\
=& \left\|\frac{1}{n_s}\sum_{k=1}^K \left(f(H_s^{(k)})^\top H_s^{(k)} - \mathbb{E}_H f(H)^\top H\right)\left(\phi_s^{(k)} - \phi_t\right)\right\|_\infty + \left\|\mathbb{E}_H f(H)^\top H\left(\frac{1}{n_s}\sum_{k=1}^K n_s^{(k)}\phi_s^{(k)} - \phi_t\right)\right\|_\infty \\
\le& C_H\sqrt{\frac{q\log p}{n_s}}\max_k\|\phi_s^{(k)} - \phi_t\|_2 + C\sqrt{q}\sqrt{\frac{\log p}{n_t}},
\end{aligned}
$$

under Assumption 1 in the event $\mathcal{G}_5$. The other two terms in (S.3) satisfy that

$$\left\|\frac{1}{n_t}E_t^\top H_t\phi_t\right\|_\infty \le \left\|\frac{1}{n_t}E_t^\top H_t\right\|_{2,\infty}\|\phi_t\|_2 \le C_H\sqrt{\frac{q\log p}{n_t}}\|\phi_t\|_2 \tag{S.5}$$

in the event $\mathcal{G}_6$ and

$$\left\|\frac{1}{n_s}\sum_{k=1}^K E_s^{(k)\top}H_s^{(k)}\phi_s^{(k)}\right\|_\infty \le \left\|\frac{1}{n_s}\sum_{k=1}^K E_s^{(k)\top}H_s^{(k)}\right\|_{2,\infty}\|\phi_t\|_2 \le C_H\sqrt{\frac{q\log p}{n_s}}\|\phi_t\|_2 \tag{S.6}$$

in the event $\mathcal{G}_7$. With (S.4), (S.5) and (S.6), the term B_3 is controlled by

$$\begin{aligned}
&\min_{\omega\in\mathbb{R}^p}\left\|\frac{1}{n_t}X_t^\top X_t\omega - \frac{1}{n_s}\sum_{k=1}^K X_s^{(k)\top}X_s^{(k)}b_s^{(k)}\right\|_\infty \\
&\le 2C_H\left(\sqrt{\frac{q\log p}{n_t}}+\sqrt{\frac{q\log p}{n_s}}\right)\|\phi_t\|_2 + o\left(\sqrt{\frac{\log p}{n_t}}\right) \\
&= 2C_H\sqrt{\frac{q\log p}{n_t}}\|\phi_t\|_2 + o\left(\sqrt{\frac{\log p}{n_t}}\right).
\end{aligned} \tag{S.7}$$

Bound Term B_4

We have

$$\left\|\frac{1}{n_s}\sum_{k=1}^K X_s^{(k)\top}\varepsilon_s^{(k)}\right\|_\infty \le C_\varepsilon\sqrt{\frac{\log p}{n_s}} = o\left(\sqrt{\frac{\log p}{n_t}}\right). \tag{S.8}$$

in event $\mathcal{G}_9$.

With the bounds (S.1), (S.2), (S.7) and (S.8) for error terms B_1, B_2, B_3 and B_4, it holds that

$$\left\|\frac{1}{n_t}X_t^\top\hat{Z}_t - \frac{1}{n_s}\sum_{k=1}^K X_s^{(k)\top}\hat{Z}_s^{(k)}\right\|_\infty \le (2C_H\sqrt{q}+C_s)\sqrt{\frac{\log p}{n_t}}\|\phi_t\|_2 + o\left(\sqrt{\frac{\log p}{n_t}}\right). \tag{S.9}$$

Now we focus on the Lasso step $Y_t - X_t\hat{\beta}_{\text{init}} - \hat{Z}_t \sim X_t$ in target data. From the definition of $\hat{\eta}$, it holds that

$$\frac{1}{n_t}\left\|Y_t - X_t\hat{\beta}_{\text{init}} - \hat{Z}_t - X_t\hat{\eta}\right\|_2^2 + \lambda_t\|\hat{\eta}\|_1 \le \frac{1}{n_t}\left\|Y_t - X_t\hat{\beta}_{\text{init}} - \hat{Z}_t - X_t\eta\right\|_2^2 + \lambda_t\|\eta\|_1.$$

With $Y_t = X_t\beta_t + H_t\phi_t + \varepsilon_t = X_t(\beta_s + \eta_0) + H_t\phi_t + \varepsilon_t$, we have

$$
\begin{aligned}
&\frac{1}{n_t}\left\|X_t\left(\hat{\eta}-\eta_0\right)\right\|_2^2 \\
\leq\ & \lambda_t\left(\|\eta_0\|_1 - \|\hat{\eta}\|_1\right) - \frac{2}{n_t}\left(\eta_0-\hat{\eta}\right)^\top X_t^\top \left(X_t\left(\beta_s - \hat{\beta}_{\text{init}}\right) + H_t\phi_t - \hat{Z}_t\right) - \frac{2}{n_t}\left(\eta_0-\hat{\eta}\right)^\top X_t^\top \varepsilon_t \\
\leq\ & \lambda_t\left(\|\eta_0\|_1 - \|\hat{\eta}\|_1\right) + 2\left\|\eta_0-\hat{\eta}\right\|_1 \left\|\frac{1}{n_t}X_t^\top\left(X_t\left(\beta_s-\hat{\beta}_{\text{init}}\right) + H_t\phi_t - \hat{Z}_t\right)\right\|_\infty + 2\left\|\eta_0-\hat{\eta}\right\|_1\left\|\frac{1}{n_t}X_t^\top\varepsilon_t\right\|_\infty .
\end{aligned}
\tag{S.10}
$$

In the event $\mathcal{G}_8$, we have

$$
\left\|\frac{1}{n_t}X_t^\top\varepsilon_t\right\|_\infty \leq C_\varepsilon\sqrt{\frac{\log p}{n_t}}.
$$

For the second term in (S.10), we have

$$
\begin{aligned}
&\left\|\frac{1}{n_t}X_t^\top\left(X_t\left(\beta_s-\hat{\beta}_{\text{init}}\right) + H_t\phi_t - \hat{Z}_t\right)\right\|_\infty \\
\leq\ & \left\|\frac{1}{n_t}X_t^\top\left(X_t\left(\beta_s-\hat{\beta}_{\text{init}}\right) + H_t\phi_t\right) - \frac{1}{n_s}\sum_{k=1}^K X_s^{(k)\top}\hat{Z}_s^{(k)}\right\|_\infty + \left\|\frac{1}{n_s}\sum_{k=1}^K X_s^{(k)\top}\hat{Z}_s^{(k)} - \frac{1}{n_t}X_t^\top\hat{Z}_t\right\|_\infty \\
=\ & \left\|\frac{1}{n_t}X_t^\top\left(X_t\left(\beta_s-\hat{\beta}_{\text{init}}\right) + H_t\phi_t\right) - \frac{1}{n_s}\sum_{k=1}^K X_s^{(k)\top}\left(X_s^{(k)}\left(\beta_s^{(k)}-\hat{\beta}_{\text{init}}\right) + H_s^{(k)}\phi_s^{(k)} + \varepsilon_s^{(k)}\right)\right\|_\infty \\
+\ & \left\|\frac{1}{n_s}\sum_{k=1}^K X_s^{(k)\top}\hat{Z}_s^{(k)} - \frac{1}{n_t}X_t^\top\hat{Z}_t\right\|_\infty \\
\leq\ & \left\|\left(\frac{1}{n_t}X_t^\top X_t - \frac{1}{n_s}\sum_{k=1}^K X_s^{(k)\top}X_s^{(k)}\right)\left(\beta_s-\hat{\beta}_{\text{init}}\right)\right\|_\infty + \left\|\frac{1}{n_s}\sum_{k=1}^K X_s^{(k)\top}X_s^{(k)}\left(\beta_s^{(k)} - \frac{1}{n_s}\sum_{k=1}^K n_s^{(k)}\beta_s^{(k)}\right)\right\|_\infty \\
+\ & \left\|\frac{1}{n_t}X_t^\top H_t\phi_t - \frac{1}{n_s}\sum_{k=1}^K X_s^{(k)\top}H_s^{(k)}\phi_s^{(k)}\right\|_\infty + \left\|\frac{1}{n_s}\sum_{k=1}^K X_s^{(k)\top}\varepsilon_s^{(k)}\right\|_\infty + \left\|\frac{1}{n_s}\sum_{k=1}^K X_s^{(k)\top}\hat{Z}_s^{(k)} - \frac{1}{n_t}X_t^\top\hat{Z}_t\right\|_\infty .
\end{aligned}
$$

Here the second equality uses the definition of $\hat{Z}_s^{(k)}$ and the third inequality uses the definition of $\beta_s^{(k)}, 1 \leq k \leq K$. With (S.1), (S.2), (S.7), (S.8) and (S.9), it holds that

$$
\left\|\frac{1}{n_t}X_t^\top\left(X_t\left(\beta_s-\hat{\beta}_{\text{init}}\right) + H_t\phi_t - \hat{Z}_t\right)\right\|_\infty \leq \left(4C_H\sqrt{q} + 2C_s\right)\sqrt{\frac{\log p}{n_t}}\|\phi_t\|_2 + o\left(\sqrt{\frac{\log p}{n_t}}\right).
$$

So

$$
\begin{aligned}
\frac{1}{n_t}\left\|X_t\left(\hat{\eta}-\eta_0\right)\right\|_2^2 &\leq \lambda_t\left(\|\eta_0\|_1 - \|\hat{\eta}\|_1\right) + 2\left(4C_H\sqrt{q} + 2C_s + C_\varepsilon\right)\sqrt{\frac{\log p}{n_t}}\left\|\eta_0-\hat{\eta}\right\|_1 \\
&\leq \lambda_t\|\eta_0\|_1 - \lambda_t\|\hat{\eta}\|_1 + \frac{\lambda_t}{2}\|\hat{\eta}-\eta_0\|_1,
\end{aligned}
\tag{S.11}
$$

when $\lambda_t \geq (C_1\sqrt{q}\|\phi_t\|_2 + C_2)\sqrt{\log p/n_t} \geq 4\left(4C_H\sqrt{q}\|\phi_t\|_2 + 2C_s + C_\varepsilon\right)\sqrt{\log p/n_t}$ for some constants C_1, C_2.

Then we split the coefficient by those located in the support S_η of η_0 or not, so

$$
\begin{aligned}
&\lambda_t\|\eta_0\|_1 - \lambda_t\|\hat{\eta}\|_1 + \frac{\lambda_t}{2}\|\hat{\eta} - \eta_0\|_1 \\
&= \lambda_t\|(\eta_0)_{S_\eta}\|_1 - \lambda_t\|(\hat{\eta})_{S_\eta}\|_1 - \lambda_t\|(\hat{\eta})_{S_\eta^c}\|_1 + \frac{\lambda_t}{2}\|(\hat{\eta} - \eta_0)_{S_\eta}\|_1 + \frac{\lambda_t}{2}\|(\hat{\eta} - \eta_0)_{S_\eta^c}\|_1 \\
&\leq \frac{3}{2}\lambda_t\|(\hat{\eta} - \eta_0)_{S_\eta}\|_1 - \frac{1}{2}\lambda_t\|(\hat{\eta} - \eta_0)_{S_\eta^c}\|_1,
\end{aligned}
$$

where in the second inequality we use $\|(\eta_0)_{S_\eta}\|_1 - \|(\hat{\eta})_{S_\eta}\|_1 \leq \|(\hat{\eta} - \eta_0)_{S_\eta}\|_1$ and $\|(\hat{\eta})_{S_\eta^c}\|_1 = \|(\hat{\eta} - \eta_0)_{S_\eta^c}\|_1$.

So

$$
\frac{1}{n_t}\|X_t(\hat{\eta} - \eta_0)\|_2^2 \leq \frac{3}{2}\lambda_t\|(\hat{\eta} - \eta_0)_{S_\eta}\|_1 - \frac{1}{2}\lambda_t\|(\hat{\eta} - \eta_0)_{S_\eta^c}\|_1. \tag{S.12}
$$

This gives us that $3\|(\hat{\eta} - \eta_0)_{S_\eta}\|_1 - \|(\hat{\eta} - \eta_0)_{S_\eta^c}\|_1 \geq 0$. With restricted eigenvalue condition in Assumption1, we have

$$
\frac{1}{n_t}\|X_t(\hat{\eta} - \eta_0)\|_2^2 \geq C_r\|\hat{\eta} - \eta_0\|_2^2. \tag{S.13}
$$

With (S.12) and (S.13), it holds that

$$
C_r\left\|(\hat{\eta} - \eta_0)_{S_\eta}\right\|_2^2 \leq C_r\|\hat{\eta} - \eta_0\|_2^2 \leq \frac{3}{2}\lambda_t\|(\hat{\eta} - \eta_0)_{S_\eta}\|_1 \leq \frac{3}{2}\lambda_t\sqrt{s_\eta}\|(\hat{\eta} - \eta_0)_{S_\eta}\|_2.
$$

So

$$
\left\|(\hat{\eta} - \eta_0)_{S_\eta}\right\|_2 \leq \frac{3}{2C_r}\sqrt{s_\eta}\lambda_t, \quad \|(\hat{\eta} - \eta_0)_{S_\eta}\|_1 \leq \sqrt{s_\eta}\|(\hat{\eta} - \eta_0)_{S_\eta}\|_2 \leq \frac{3}{2C_r}s_\eta\lambda_t.
$$

and then

$$
\begin{aligned}
\|\hat{\eta} - \eta_0\|_1 &\leq \left\|(\hat{\eta} - \eta_0)_{S_\eta}\right\|_1 + \left\|(\hat{\eta} - \eta_0)_{S_\eta^c}\right\|_1 \leq 4\left\|(\hat{\eta} - \eta_0)_{S_\eta}\right\|_1 \leq \frac{6}{C_r}s_\eta\lambda_t, \\
\|\hat{\eta} - \eta_0\|_2 &\leq \left(\frac{3}{2C_t}\lambda_t\left\|(\hat{\eta} - \eta_0)_{S_\eta}\right\|_1\right)^{1/2} \leq \frac{3}{2C_r}\sqrt{s_\eta}\lambda_t, \\
\frac{1}{\sqrt{n_t}}\|X_t(\hat{\eta} - \eta_0)\|_2 &\leq \left(\frac{3}{2}\lambda_t\left\|(\hat{\eta} - \eta_0)_{S_\eta}\right\|_1\right)^{1/2} \leq \frac{3}{2\sqrt{C_r}}\sqrt{s_\eta}\lambda_t.
\end{aligned}
$$

□

S.1.2 Lemmas for Theorem 1

Lemma S.1 *Under Assumption 1, the event* $\mathcal{G} = \cap_{j=1}^{9} \mathcal{G}_j$ *happens with a probability*

$$\mathbb{P}(\mathcal{G}) \geq 1 - p^{-c},$$

for a constant c.

Proof of Lemma S.1. From Lemma S.2, we know that

$$\mathbb{P}(\mathcal{G}_1 \cap \mathcal{G}_2) \geq 1 - 6p^{-cC_x^2+2}.$$

From Lemma S.3, we know that

$$\mathbb{P}(\mathcal{G}_3) \geq 1 - 6p^{-cC_x^2/16+2}.$$

From Lemma S.4, we know that

$$\mathbb{P}(\mathcal{G}_4 \cap \mathcal{G}_5 \cap \mathcal{G}_6 \cap \mathcal{G}_7) \geq 1 - 8p^{-cC_H^2+1+\log q}.$$

From S.5, we know that

$$\mathbb{P}(\mathcal{G}_8 \cap \mathcal{G}_9) \geq 1 - 8p^{-cC_\varepsilon^2/(4\sigma^2)+1}.$$

So

$$\mathbb{P}(\mathcal{G}) \geq 1 - 12p^{-cC_x^2+2} - 6p^{-cC_x^2/16+2} - 8p^{-cC_H^2+1+\log q} - 8p^{-cC^2/(4\sigma^2)+1} \geq 1 - p^{-c},$$

for some constant c. □

Lemma S.2 *If* $E_i, H_i, i = 1, \cdots, n$ *are i.i.d. sampled from distributions for* E *and* H *that satisfy Assumption 1 and the function* f *satisfies the condition for* f *in Assumption 1, then with probability larger than* $1 - 6p^{-cC_x^2/16+2}$,

$$\left\| \frac{1}{n} X^\top X - \Sigma_X \right\|_\infty \leq C_x \sqrt{\frac{\log p}{n}}$$

for some constants c, C_x.

Proof of Lemma S.2. From Assumption 1, we know that $E_{\cdot,i}$, $H_{\cdot,i}$ and $f(H)_{\cdot,i}$ follows sub-Gaussian distributions. So $E_{\cdot,i}E_{\cdot,j}, f(H)_{\cdot,i}f(H)_{\cdot,j}, f(H)_{\cdot,i}E_{\cdot,j}$ are sub-exponential random variables and from Bernstein inequality,

$$\begin{aligned}
&\mathbb{P}\left(\max_{1\le i,j\le p}\left|\frac{1}{n}\sum_{l=1}^{n}E_{l,i}E_{l,j}-\mathbb{E}E_{\cdot,i}E_{\cdot,j}\right|\ge C\sqrt{\frac{\log p}{n}}\right)\\
\le&\sum_{1\le i,j\le p}\mathbb{P}\left(\left|\frac{1}{n}\sum_{l=1}^{n}E_{l,i}E_{l,j}-\mathbb{E}E_{\cdot,i}E_{\cdot,j}\right|\ge C\sqrt{\frac{\log p}{n}}\right)\\
\le&\, 2p^2\exp\left\{-cC^2\log p\right\}\le 2p^{-cC^2+2}.
\end{aligned}$$

Similarly, we have

$$\begin{aligned}
&\mathbb{P}\left(\max_{1\le i,j\le p}\left|\frac{1}{n}\sum_{l=1}^{n}f(H)_{l,i}f(H)_{l,j}-\mathbb{E}f(H)_{\cdot,i}f(H)_{\cdot,j}\right|\ge C\sqrt{\frac{\log p}{n}}\right)\le 2p^{-cC_x^2+2},\\
&\mathbb{P}\left(\max_{1\le i,j\le p}\left|\frac{1}{n}\sum_{l=1}^{n}f(H)_{l,i}E_{l,j}-\mathbb{E}f(H)_{\cdot,i}E_{\cdot,j}\right|\ge C\sqrt{\frac{\log p}{n}}\right)\le 2p^{-cC^2+2}.
\end{aligned}$$

So

$$\begin{aligned}
\mathbb{P}\left(\left\|\frac{1}{n}X^\top X-\Sigma_X\right\|_\infty\ge C_x\sqrt{\frac{\log p}{n}}\right)\le&\ \mathbb{P}\left(\left\|\frac{1}{n}E^\top E-\mathbb{E}E^\top E\right\|_\infty\ge\frac{C_x}{4}\sqrt{\frac{\log p}{n}}\right)\\
&+\mathbb{P}\left(\left\|\frac{1}{n}f(H)^\top H-\mathbb{E}f(H)^\top H\right\|_\infty\ge\frac{C_x}{4}\sqrt{\frac{\log p}{n}}\right)\\
&+\mathbb{P}\left(\left\|\frac{1}{n}f(H)^\top E-\mathbb{E}f(H)^\top E\right\|_\infty\ge\frac{C_x}{4}\sqrt{\frac{\log p}{n}}\right)\\
\le&\ 6p^{-cC_x^2/16+2}.
\end{aligned}$$

□

Lemma S.3 *If $E_i^{(k)}, H_i^{(k)}, i=1,\cdots,n$ are i.i.d. sampled from distributions for E and H that satisfy Assumption 1 and the function f satisfies the condition for f in Assumption 1, then with probability larger than $1-6p^{-cC_x^2/16+2}$,*

$$\left\|\frac{1}{n_s}\sum_{k=1}^{K}n_s^{(k)}\left(\frac{1}{n_s^{(k)}}X_s^{(k)\top}X_s^{(k)}-\Sigma_X\right)\left(\beta_s^{(k)}-\frac{1}{n_s}\sum_{k=1}^{K}n_s^{(k)}\beta_s^{(k)}\right)\right\|_\infty\le C_\delta s_\delta\sqrt{\frac{\log p}{n_s}},$$

for some constants c, C_δ.

Proof of Lemma S.3. From Assumption 1, we know that $E_{\cdot,i}$, $H_{\cdot,i}$ and $f(H)_{\cdot,i}$ follows sub-Gaussian distributions. So $E_{\cdot,i}E_{\cdot,j}, f(H)_{\cdot,i}f(H)_{\cdot,j}, f(H)_{\cdot,i}E_{\cdot,j}$ are sub-exponential random variables. So

$$\begin{aligned}
&\mathbb{P}\left(\left\|\frac{1}{n_s}\sum_{k=1}^{K} n_s^{(k)}\left(\frac{1}{n_s^{(k)}}X_s^{(k)\top}X_s^{(k)}-\Sigma_X\right)\left(\beta_s^{(k)}-\frac{1}{n_s}\sum_{k=1}^{K} n_s^{(k)}\beta_s^{(k)}\right)\right\|_\infty \geq C_\beta C_x s_\delta\sqrt{\frac{\log p}{n_s}}\right)\\
\leq\ &\mathbb{P}\left(\left\|\frac{1}{n_s}\sum_{k=1}^{K} n_s^{(k)}\left(\frac{1}{n_s^{(k)}}X_s^{(k)\top}X_s^{(k)}-\Sigma_X\right)\right\|_\infty \sum_{k=1}^{K}\frac{n_s^{(k)}}{n_s}\left\|\beta_s^{(k)}-\frac{1}{n_s}\sum_{k=1}^{K} n_s^{(k)}\beta_s^{(k)}\right\|_1 \geq C_\beta C_x s_\delta\sqrt{\frac{\log p}{n_s}}\right)\\
\leq\ &\mathbb{P}\left(\left\|\frac{1}{n_s}\sum_{k=1}^{K} n_s^{(k)}\left(\frac{1}{n_s^{(k)}}X_s^{(k)\top}X_s^{(k)}-\Sigma_X\right)\right\|_\infty \geq C_x s_\delta\sqrt{\frac{\log p}{n_s}}\right) \leq 6p^{-cC_x^2/16+2},
\end{aligned}$$

where we use Lemma S.2 for the last inequality. □

Lemma S.4 *If $E_i, H_i, i = 1, \cdots, n$ are i.i.d. sampled from distributions for E and H that satisfy Assumption 1 and the function f satisfies the condition for f in Assumption 1, then with probability larger than $1-4qp^{-cC_H^2+1}$,*

$$\begin{aligned}
\left\|\frac{1}{n}f(H)^\top H-\mathbb{E}f(H)^\top H\right\|_{2,\infty} &\leq C_H\sqrt{\frac{q\log p}{n}},\\
\left\|\frac{1}{n}E^\top H\right\|_{2,\infty} &\leq C_H\sqrt{\frac{q\log p}{n}}.
\end{aligned}$$

for some constants c, C_H.

Proof of Lemma S.4. From Assumption 1, we know that $E_{\cdot,i}$, $H_{\cdot,i}$ and $f(H)_{\cdot,i}$ follows sub-Gaussian distributions. So $f(H)_{\cdot,i}H_{\cdot,j}, E_{\cdot,i}H_{\cdot,j}$ are sub-exponential random variables and from Bernstein inequality,

$$\begin{aligned}
&\mathbb{P}\left(\max_{1\leq i\leq p, 1\leq j\leq q}\left|\frac{1}{n}\sum_{l=1}^{n} f(H)_{l,i}H_{l,j}-\mathbb{E}f(H)_{\cdot,i}H_{\cdot,j}\right| \geq C\sqrt{\frac{\log p}{n}}\right)\\
\leq &\sum_{1\leq i\leq p, 1\leq j\leq q}\mathbb{P}\left(\left|\frac{1}{n}\sum_{l=1}^{n} f(H)_{l,i}H_{l,j}-\mathbb{E}f(H)_{\cdot,i}H_{\cdot,j}\right| \geq C\sqrt{\frac{\log p}{n}}\right)\\
\leq\ &2pq\exp\left\{-cC^2\log p\right\} \leq 2qp^{-cC^2+1}.
\end{aligned}$$

Similarly, we have

$$\mathbb{P}\left(\max_{1\le i\le p,1\le j\le q}\left|\frac{1}{n}\sum_{l=1}^{n}E_{l,i}H_{l,j}\right|\ge C\sqrt{\frac{\log p}{n}}\right)\le 2p^{-cC_H^2+2}$$

So

$$\begin{aligned}&\mathbb{P}\left(\left\|\frac{1}{n}f(H)^\top H-\mathbb{E}f(H)^\top H\right\|_{2,\infty}\ge C_H\sqrt{\frac{q\log p}{n}}\right)\\ =\ &\max_{1\le i\le p}\mathbb{P}\left(\left(\sum_{j=1}^{q}\left(\frac{1}{n}\sum_{l=1}^{n}f(H)_{l,i}H_{l,j}-\mathbb{E}f(H)_{\cdot,i}H_{\cdot,j}\right)^2\right)^{1/2}\ge C_H\sqrt{\frac{\log p}{n}}\right)\\ \le\ &\mathbb{P}\left(\max_{1\le i\le p,1\le j\le q}\left|\frac{1}{n}\sum_{l=1}^{n}f(H)_{l,i}H_{l,j}-\mathbb{E}f(H)_{\cdot,i}H_{\cdot,j}\right|\ge C_H\sqrt{\frac{\log p}{n}}\right)\le 2qp^{-cC_H^2+1},\end{aligned}$$

and similarly

$$\mathbb{P}\left(\left\|\frac{1}{n}E^\top H\right\|_{2,\infty}\ge C_H\sqrt{\frac{q\log p}{n}}\right)\le 2qp^{-cC_H^2+1},$$

$\square$

Lemma S.5 *If $E_i, H_i, \varepsilon_i, i=1,\cdots,n$ are i.i.d. sampled from distributions for E, H and ε that satisfy Assumption 1 and the function f satisfies the condition for f in Assumption 1, then with probability larger than $1-4p^{-c_\varepsilon C^2/(4\sigma^2)+1}$,*

$$\left\|\frac{1}{n}X^\top\varepsilon\right\|_\infty\le C_\varepsilon\sqrt{\frac{\log p}{n}}$$

for some constants c, C_ε.

Proof of Lemma S.5. From Assumption 1, we know that $E_{\cdot,i}$, $H_{\cdot,i}$, $f(H)_{\cdot,i}$ and $\varepsilon_\cdot$ follows sub-Gaussian distributions. So $E_{\cdot,i}\varepsilon_\cdot, f(H)_{\cdot,i}\varepsilon_\cdot$ are sub-exponential random variables and from Bernstein inequality,

$$\begin{aligned}\mathbb{P}\left(\max_{1\le i\le p}\left|\frac{1}{n}\sum_{l=1}^{n}E_{l,i}\varepsilon_l\right|\ge C\sqrt{\frac{\log p}{n}}\right)&\le\sum_{1\le i\le p}\mathbb{P}\left(\left|\frac{1}{n}\sum_{l=1}^{n}E_{l,i}\varepsilon_l\right|\ge C\sqrt{\frac{\log p}{n}}\right)\\&\le 2p\exp\left\{-\frac{cC^2}{\sigma^2}\log p\right\}\le 2p^{-cC^2/\sigma^2+1}.\end{aligned}$$

Similarly, we have

$$\mathbb{P}\left(\max_{1\le i\le p}\left|\frac{1}{n}\sum_{l=1}^{n} f(H)_{l,i}\varepsilon_l\right| \ge C\sqrt{\frac{\log p}{n}}\right) \le 2p^{-cC^2/\sigma^2+1}.$$

So

$$\begin{aligned}\mathbb{P}\left(\left\|\frac{1}{n}X^\top\varepsilon\right\|_\infty \ge C_\varepsilon\sqrt{\frac{\log p}{n}}\right) \le & \mathbb{P}\left(\left\|\frac{1}{n}E^\top\varepsilon\right\|_\infty \ge \frac{C_\varepsilon}{2}\sqrt{\frac{\log p}{n}}\right) \\ & + \mathbb{P}\left(\left\|\frac{1}{n}f(H)^\top\varepsilon\right\|_\infty \ge \frac{C_\varepsilon}{2}\sqrt{\frac{\log p}{n}}\right) \\ & \le 4p^{-cC_\varepsilon^2/(4\sigma^2)+1}.\end{aligned}$$

$\square$

S.1.3 Proof of Proposition 1

Proof of Proposition 1. We consider the special case that $\beta_s^{(k)} = \beta_s = 0, 1 \le k \le K$ and the noise distribution is Gaussian distribution $N(0,\sigma^2)$. We then consider the case that $f = 0, \phi_t = 0$ and $X_t, X_s^{(k)}$ follow standard Gaussian distribution $N(0, I_p)$. Define the distribution of source data and target data as $P_s(\eta), P_t(\eta)$. For $\eta_1, \eta_2 \in \mathcal{B}_0(s_\eta)$, We know that

$$D_{KL}\left(P_s(\eta_1)\|P_s(\eta_2)\right) = 0, \quad D_{KL}\left(P_t(\eta_1)\|P_t(\eta_2)\right) = \frac{1}{2\sigma^2}\left\|X_t\left(\eta_1-\eta_2\right)\right\|_2^2.$$

Define a δ-packing of $\mathcal{B}_0(s_\eta)$ in metric ρ as $\{\eta_1,\cdots,\eta_M\} \subseteq \mathcal{B}_0(s_\eta)$ such that $\rho(\eta_i,\eta_j) > \delta$ for all $i \ne j$. We construct a packing of $\mathcal{B}_0(s_\eta)$ as $\mathcal{H}_{s_\eta}$ in which each $\eta \in \mathcal{H}_{s_\eta}$ satisfies $\|\eta\|_0 = s_\eta$ and any nonzero element of η is $\gamma \ne 0$. From the Varshamov-Gilbert bound, there exist such a $\mathcal{H}_{s_\eta}$ with

$$\left|\mathcal{H}_{s_\eta}\right| \ge \exp\left\{\frac{s_\eta}{2}\log\frac{p-s_\eta}{s_\eta/2}\right\}$$

satisfies that for any $\eta_i, \eta_j \in \mathcal{H}_{s_\eta}$,

$$\sum_{l=1}^{p} \mathbf{1}_{(\eta_i)_l \neq (\eta_j)_l} \geq \frac{s_\eta}{2}.$$

Hence $\|\eta_i - \eta_j\|_2 \geq \gamma\sqrt{s_\eta/2}$. Choose $\gamma = \delta\sqrt{2/s_\eta}$, we know that such $\mathcal{H}_{s_\eta}$ is a δ-packing of $\mathcal{B}_0(s_\eta)$ in l_2 norm.

So

$$D_{KL}\left(P_t(\eta_i)\|P_t(\eta_j)\right) = \frac{1}{2\sigma^2}\left\|X_t\left(\eta_i - \eta_j\right)\right\|_2^2 \leq \frac{2n_t\|\eta_i - \eta_j\|_2^2}{\sigma^2} \leq \frac{2n_t s_\eta \delta^2}{\sigma^2}.$$

With Fano's lemma, we know that

$$\begin{aligned}
\inf_{\hat{\eta}} \sup_{\eta \in \mathcal{B}_0(s_\eta)} \mathbb{P}\left(\|\hat{\eta} - \eta_0\|_2 \geq \delta\right) &\geq 1 - \frac{\max_{i,j} D_{KL}\left(P_t(\eta_i)\|P_t(\eta_j)\right) + \log 2}{\log\left|\mathcal{H}_{s_\eta}\right|} \\
&\geq 1 - \frac{2(2n_t\delta^2 + \log 2)}{s_\eta \log(2(p - s_\eta)/s_\eta)} \geq \frac{1}{2},
\end{aligned}$$

if we choose $\delta = c_0\sqrt{s_\eta \log p/n_t}$ for a small enough constant c_0.

Similarly, we have $\|\eta_i - \eta_j\|_1 \geq \gamma s_\eta/2$. Choose $\gamma = 2\delta/s_\eta$, we know that such $\mathcal{H}_{s_\eta}$ is a δ-packing of $\mathcal{B}_0(s_\eta)$ in l_1 norm.

So

$$D_{KL}\left(P_t(\eta_i)\|P_t(\eta_j)\right) \leq \frac{4n_t\|\eta_i - \eta_j\|_1^2}{\sigma^2} \leq \frac{4n_t\delta^2}{\sigma^2},$$

this is because $\|\eta_i - \eta_j\|_2 \leq \|\eta_i - \eta_j\|_1$. With Fano's lemma, we know that

$$\begin{aligned}
\inf_{\hat{\eta}} \sup_{\eta \in \mathcal{B}_0(s_\eta)} \mathbb{P}\left(\|\hat{\eta} - \eta_0\|_1 \geq \delta\right) &\geq 1 - \frac{\max_{i,j} D_{KL}\left(P_t(\eta_i)\|P_t(\eta_j)\right) + \log 2}{\log\left|\mathcal{H}_{s_\eta}\right|} \\
&\geq 1 - \frac{2(4n_t\delta^2 + \log 2)}{s_\eta \log(2(p - s_\eta)/s_\eta)} \geq \frac{1}{2},
\end{aligned}$$

if we choose $\delta = c_0\sqrt{s_\eta \log p/n_t}$. $\square$

S.1.4 Proof of Theorem 2

Proof of Theorem 2. We first define $\mathcal{V}_s = V_s \operatorname{diag}(\min(\tau_s/D_s, 1))V_s^\top$ and thus $\tilde{X}_s^{(k)} = X_s^{(k)}\mathcal{V}_s$ and $\tilde{X}_s^{(k)\top}\tilde{X}_s^{(k)} = \mathcal{V}_s^2 X_s^{(k)\top}X_s^{(k)}$. Define several events:

$$\left\{\left\|\frac{1}{n_s}\sum_{k=1}^K \tilde{X}_s^{(k)\top}\tilde{\varepsilon}_s^{(k)}\right\|_\infty \le C_\varepsilon\sqrt{\frac{\log p}{n_s}}\right\}, \tag{$\mathcal{G}_{10}$}$$

$$\left\{\left\|\frac{1}{n_s}\sum_{k=1}^K \tilde{X}_s^{(k)\top}\left(\tilde{H}_s^{(k)}\phi_s^{(k)} - \tilde{X}_s^{(k)}b_s\right)\right\|_\infty \le C_k C_\phi\sqrt{\frac{q\log p}{n_s}}\|\phi_t\|_2 + 2C_t\sqrt{\frac{p}{n_s}}\lambda_\Psi^{-1}\|\phi_t\|_2\right\}, \tag{$\mathcal{G}_{11}$}$$

$$\left\{\left\|\frac{1}{n_s}\sum_{k=1}^K \tilde{X}_s^{(k)\top}\tilde{X}_s^{(k)}\left(\beta_s^{(k)} - \frac{1}{n_s}\sum_{k=1}^K n_s^{(k)}\beta_s^{(k)}\right)\right\|_\infty \le C_k s_\delta\sqrt{\frac{\log p}{n_s}}\right\}, \tag{$\mathcal{G}_{12}$}$$

Define the event $\mathcal{G}' = \cap_{j=10}^{12}\mathcal{G}_j$. From Lemma S.6, the event $\mathcal{G}'$ happens with a probability

$$\mathbb{P}(\mathcal{G}') \ge 1 - p^{-c}.$$

From the definition of $\hat{\beta}_s$, we know that

$$\frac{1}{n_s}\sum_{k=1}^K \left\|\tilde{Y}_s^{(k)} - \tilde{X}_s^{(k)}\hat{\beta}_s\right\|_2^2 + \lambda_s\|\hat{\beta}_s\|_1 \le \frac{1}{n_s}\sum_{k=1}^K \left\|\tilde{Y}_s^{(k)} - \tilde{X}_s^{(k)}\beta_s\right\|_2^2 + \lambda_s\|\beta_s\|_1.$$

The source model is

$$\tilde{Y}_s^{(k)} = \tilde{X}_s^{(k)}\left(\beta_s^{(k)} - \frac{1}{n_s}\sum_{k=1}^K n_s^{(k)}\beta_s^{(k)} + \beta_s\right) + \tilde{H}_s^{(k)}\phi_s^{(k)} + \tilde{\varepsilon}_s^{(k)}.$$

So

$$\frac{1}{n_s}\left\|\tilde{X}_s\left(\hat{\beta}_s - \beta_s\right)\right\|_2^2 \le \lambda_s\left(\|\beta_s\|_1 - \|\hat{\beta}_s\|_1\right) - \frac{2}{n_s}\left(\beta_s - \hat{\beta}_s\right)^\top\left(\sum_{k=1}^K \tilde{X}_s^{(k)\top}\tilde{\varepsilon}_s^{(k)}\right) \tag{B_5}$$

$$- \frac{2}{n_s}\left(\beta_s - \hat{\beta}_s\right)^\top\left(\sum_{k=1}^K \tilde{X}_s^{(k)\top}\tilde{H}_s^{(k)}\phi_s^{(k)}\right) \tag{B_6}$$

$$- \frac{2}{n_s}\left(\beta_s - \hat{\beta}_s\right)^\top\left(\sum_{k=1}^K \tilde{X}_s^{(k)\top}\tilde{X}_s^{(k)}\left(\beta_s^{(k)} - \frac{1}{n_s}\sum_{k=1}^K n_s^{(k)}\beta_s^{(k)}\right)\right). \tag{B_7}$$

In the right hand of the above inequality, we named three terms B_5, B_6 and B_7. The term B_5 measures the noise influence. The term B_6 is caused by the hidden confounders. The term B_7 is due to the fluctuation of source models.

For term B_5, we have

$$\left|\frac{2}{n_s}\left(\beta_s-\hat{\beta}_s\right)^\top\left(\sum_{k=1}^K \tilde{X}_s^{(k)\top}\tilde{\varepsilon}_s^{(k)}\right)\right| \le \left\|\frac{2}{n_s}\sum_{k=1}^K \tilde{X}_s^{(k)\top}\tilde{\varepsilon}_s^{(k)}\right\|_\infty \left\|\beta_s-\hat{\beta}_s\right\|_1$$
$$\le 2C_\varepsilon\sqrt{\frac{\log p}{n_s}}\left\|\beta_s-\hat{\beta}_s\right\|_1, \tag{S.14}$$

in the event $\mathcal{G}_{10}$.

For term B_6, define that $b_s=\Sigma_X^{-1}\mathbb{E}[X^\top H]\phi_t$, we have

$$\left|\frac{2}{n_s}\left(\beta_s-\hat{\beta}_s\right)^\top\left(\sum_{k=1}^K \tilde{X}_s^{(k)\top}\tilde{H}_s^{(k)}\phi_s^{(k)}\right)\right| \le \left|\frac{2}{n_s}\left(\beta_s-\hat{\beta}_s\right)^\top\left(\sum_{k=1}^K \tilde{X}_s^{(k)\top}\tilde{X}_s^{(k)}b_s\right)\right|$$
$$+\left|\frac{2}{n_s}\left(\beta_s-\hat{\beta}_s\right)^\top\left(\sum_{k=1}^K \tilde{X}_s^{(k)\top}\left(\tilde{H}_s^{(k)}\phi_s^{(k)}-\tilde{X}_s^{(k)}b_s\right)\right)\right|.$$

The first term is bounded by

$$\left|\frac{2}{n_s}\left(\beta_s-\hat{\beta}_s\right)^\top\left(\sum_{k=1}^K \tilde{X}_s^{(k)\top}\tilde{X}_s^{(k)}b_s\right)\right| \le \left\|\frac{2}{n_s}\tilde{X}_s b_s\right\|_2\left\|\tilde{X}_s\left(\beta_s-\hat{\beta}_s\right)\right\|_2.$$

The second term is bounded by

$$\left|\frac{2}{n_s}\left(\beta_s-\hat{\beta}_s\right)^\top\left(\sum_{k=1}^K \tilde{X}_s^{(k)\top}\left(\tilde{H}_s^{(k)}\phi_s^{(k)}-\tilde{X}_s^{(k)}b_s\right)\right)\right|$$
$$\le \left\|\frac{2}{n_s}\sum_{k=1}^K \tilde{X}_s^{(k)\top}\left(\tilde{H}_s^{(k)}\phi_s^{(k)}-\tilde{X}_s^{(k)}b_s\right)\right\|_\infty\left\|\beta_s-\hat{\beta}_s\right\|_1$$
$$\le 2\left(C_kC_\phi\sqrt{\frac{q\log p}{n_s}}+2C_t\sqrt{\frac{p}{n_s}}\lambda_\Psi^{-1}\|\phi_t\|_2\right)\left\|\beta_s-\hat{\beta}_s\right\|_1,$$

in event $\mathcal{G}_{11}$. So

$$\left|\frac{2}{n_s}\left(\beta_s-\hat{\beta}_s\right)^\top\left(\sum_{k=1}^K \tilde{X}_s^{(k)\top}\tilde{H}_s^{(k)}\phi_s^{(k)}\right)\right|$$
$$\le \left\|\frac{2}{n_s}\tilde{X}_s b_s\right\|_2\left\|\tilde{X}_s\left(\beta_s-\hat{\beta}_s\right)\right\|_2+2\left(C_kC_\phi\sqrt{\frac{q\log p}{n_s}}+2C_t\sqrt{\frac{p}{n_s}}\lambda_\Psi^{-1}\|\phi_t\|_2\right)\left\|\beta_s-\hat{\beta}_s\right\|_1. \tag{S.15}$$

For term B_7, we have

$$
\begin{aligned}
&\left|\frac{2}{n_s}\left(\beta_s-\hat{\beta}_s\right)^{\top}\left(\sum_{k=1}^{K}\tilde{X}_s^{(k)\top}\tilde{X}_s^{(k)}\left(\beta_s^{(k)}-\frac{1}{n_s}\sum_{k=1}^{K}n_s^{(k)}\beta_s^{(k)}\right)\right)\right|\\
\leq\ &\left\|\frac{2}{n_s}\sum_{k=1}^{K}\tilde{X}_s^{(k)\top}\tilde{X}_s^{(k)}\left(\beta_s^{(k)}-\frac{1}{n_s}\sum_{k=1}^{K}n_s^{(k)}\beta_s^{(k)}\right)\right\|_{\infty}\left\|\beta_s-\hat{\beta}_s\right\|_1\\
\leq\ &2C_k s_\delta\sqrt{\frac{\log p}{n_s}}\left\|\beta_s-\hat{\beta}_s\right\|_1,
\end{aligned}
\tag{S.16}
$$

in event $\mathcal{G}_{12}$.

With the bounds (S.14),(S.15) and (S.16) or error terms B_5,B_6 and B_7, it holds that

$$
\begin{aligned}
\frac{1}{n_s}\left\|\tilde{X}_s\left(\hat{\beta}_s-\beta_s\right)\right\|_2^2\leq\ &\lambda_s\left(\|\beta_s\|_1-\|\hat{\beta}_s\|_1\right)\\
&+2\left((C_\varepsilon+C_\delta C_x+C_kC_\phi)\sqrt{\frac{\log p}{n_s}}+2C_t\sqrt{\frac{p}{n_s}}\lambda_\Psi^{-1}\|\phi_t\|_2\right)\left\|\beta_s-\hat{\beta}_s\right\|_1\\
&+\left\|\frac{2}{n_s}\tilde{X}_s b_s\right\|_2\left\|\tilde{X}_s\left(\beta_s-\hat{\beta}_s\right)\right\|_2.
\end{aligned}
$$

When $\lambda_s\geq 8\left((C_\varepsilon+C_ks_\delta+C_kC_\phi)\sqrt{\log p/n_s}+C_t\sqrt{p/n_s}\lambda_\Psi^{-1}\|\phi_t\|_2\right)$, it holds that

$$
\frac{1}{n_s}\left\|\tilde{X}_s\left(\hat{\beta}_s-\beta_s\right)\right\|_2^2\leq\lambda_s\|\beta_s\|_1-\lambda_s\|\hat{\beta}_s\|_1+\frac{\lambda_s}{4}\|\hat{\beta}_s-\beta_s\|_1+\left\|\frac{2}{n_s}\tilde{X}_s b_s\right\|_2\left\|\tilde{X}_s\left(\beta_s-\hat{\beta}_s\right)\right\|_2.
\tag{S.17}
$$

Case 1. If

$$
\left\|\frac{2}{n_s}\tilde{X}_s b_s\right\|_2\left\|\tilde{X}_s\left(\beta_s-\hat{\beta}_s\right)\right\|_2\leq\frac{\lambda_s}{4}\left\|(\hat{\beta}_s-\beta_s)_{S_s}\right\|_1,
$$

then (S.17) can be simplified as

$$
\begin{aligned}
&\frac{1}{n_s}\left\|\tilde{X}_s\left(\hat{\beta}_s-\beta_s\right)\right\|_2^2\\
\leq\ &\lambda_s\|\beta_s\|_1-\lambda_s\|\hat{\beta}_s\|_1+\frac{\lambda_s}{2}\|\hat{\beta}_s-\beta_s\|_1\\
=\ &\lambda_s\|(\beta_s)_{S_s}\|_1-\lambda_s\|(\hat{\beta}_s)_{S_s}\|_1-\lambda_s\|(\hat{\beta}_s)_{S_s^c}\|_1+\frac{\lambda_s}{2}\|(\hat{\beta}_s-\beta_s)_{S_s}\|_1+\frac{\lambda_s}{2}\|(\hat{\beta}_s-\beta_s)_{S_s^c}\|_1\\
\leq\ &\frac{3}{2}\lambda_s\|(\hat{\beta}_s-\beta_s)_{S_s}\|_1-\frac{1}{2}\lambda_s\|(\hat{\beta}_s-\beta_s)_{S_s^c}\|_1,
\end{aligned}
$$

where in the last inequality we use $\|(\beta_s)_{S_s}\|_1 - \|(\hat{\beta}_s)_{S_s}\|_1 \leq \|(\hat{\beta}_s - \beta_s)_{S_s}\|_1$ and $\|(\hat{\beta}_s)_{S_s^c}\|_1 = \|(\hat{\beta}_s - \beta_s)_{S_s^c}\|_1$.

This gives us that $3\|(\hat{\beta}_s - \beta_s)_{S_s}\|_1 - \|(\hat{\beta}_s - \beta_s)_{S_s^c}\|_1 \geq 0$. From the restricted eigenvalue condition in Assumption 2, we have

$$C_r\|\hat{\beta}_s - \beta_s\|_2^2 \leq \frac{1}{n_s}\left\|\tilde{X}_s\left(\hat{\beta}_s - \beta_s\right)\right\|_2^2 \leq \frac{3}{2}\lambda_s\|(\hat{\beta}_s - \beta_s)_{S_s}\|_1 \leq \frac{3}{2}\lambda_s\sqrt{s}\|(\hat{\beta}_s - \beta_s)_{S_s}\|_1.$$

So

$$\|(\hat{\beta}_s - \beta_s)_{S_s}\|_2 \leq \frac{3}{2C_r}\sqrt{s}\lambda_s, \quad \|(\hat{\beta}_s - \beta_s)_{S_s}\|_1 \leq \sqrt{s}\|(\hat{\beta}_s - \beta_s)_{S_s}\|_2 \leq \frac{3}{2C_r}s\lambda_s,$$

and then

$$\begin{aligned}
&\|\hat{\beta}_s - \beta_s\|_1 \leq 4\|(\hat{\beta}_s - \beta_s)_{S_s}\|_1 \leq \frac{6}{C_r}s\lambda_s, \\
&\|\hat{\beta}_s - \beta_s\|_2 \leq \left(\frac{3}{2C_r}\lambda_s\|(\hat{\beta}_s - \beta_s)_{S_s}\|_1\right)^{1/2} \leq \frac{3}{2C_t}\sqrt{s}\lambda_s.
\end{aligned} \tag{S.18}$$

Case 2. If

$$\left\|\frac{2}{n_s}\tilde{X}_s b_s\right\|_2 \left\|\tilde{X}_s\left(\beta_s - \hat{\beta}_s\right)\right\|_2 \geq \frac{\lambda_s}{4}\left\|(\hat{\beta}_s - \beta_s)_{S_s}\right\|_1.$$

then

$$\|(\hat{\beta}_s - \beta_s)_{S_s}\|_1 \leq \frac{8}{\lambda_s}\left\|\frac{1}{n_s}\tilde{X}_s b_s\right\|_2 \left\|\tilde{X}_s\left(\beta_s - \hat{\beta}_s\right)\right\|_2.$$

And (S.17) can be simplified as

$$\begin{aligned}
&\frac{1}{n_s}\left\|\tilde{X}_s\left(\hat{\beta}_s - \beta_s\right)\right\|_2^2 \\
&\leq \frac{5}{4}\lambda_s\|(\hat{\beta}_s - \beta_s)_{S_s}\|_1 - \frac{1}{4}\lambda_s\|(\hat{\beta}_s - \beta_s)_{S_s^c}\|_1 + \left\|\frac{2}{n_s}\tilde{X}_s b_s\right\|_2 \left\|\tilde{X}_s\left(\beta_s - \hat{\beta}_s\right)\right\|_2 \\
&\leq 10\left\|\frac{1}{n_s}\tilde{X}_s b_s\right\|_2 \left\|\tilde{X}_s\left(\beta_s - \hat{\beta}_s\right)\right\|_2 - \frac{1}{4}\lambda_s\|(\hat{\beta}_s - \beta_s)_{S_s^c}\|_1.
\end{aligned} \tag{S.19}$$

So

$$\frac{1}{n_s}\left\|\tilde{X}_s\left(\hat{\beta}_s - \beta_s\right)\right\|_2 \leq 10\left\|\frac{1}{n_s}\tilde{X}_s b_s\right\|_2.$$

The inequality (S.19) also gives us that

$$\|(\hat{\beta}_s - \beta_s)_{S_s^c}\|_1 \leq \frac{40}{\lambda_s} \left\| \frac{1}{n_s} \tilde{X}_s b_s \right\|_2 \left\| \tilde{X}_s \left(\beta_s - \hat{\beta}_s \right) \right\|_2 .$$

So

$$\begin{aligned}
\|\hat{\beta}_s - \beta_s\|_2 \leq \|\hat{\beta}_s - \beta_s\|_1 &\leq \|(\hat{\beta}_s - \beta_s)_{S_s}\|_1 + \|(\hat{\beta}_s - \beta_s)_{S_s^c}\|_1 \\
&\leq \frac{48}{\lambda_s} \left\| \frac{1}{n_s} \tilde{X}_s b_s \right\|_2 \left\| \tilde{X}_s \left(\beta_s - \hat{\beta}_s \right) \right\|_2 \\
&\leq \frac{480}{\lambda_s} \left\| \frac{1}{\sqrt{n_s}} \tilde{X}_s b_s \right\|_2^2 \qquad \text{(S.20)}
\end{aligned}$$

With (S.18) and (S.20), it holds for a constant C that

$$\begin{aligned}
\|\hat{\beta}_s - \beta_s\|_2 &\leq \frac{C}{C_r} \sqrt{s} \lambda_s + \frac{1}{\lambda_s} \left\| \frac{1}{\sqrt{n_s}} \tilde{X}_s b_s \right\|_2^2 , \\
\|\hat{\beta}_s - \beta_s\|_1 &\leq \frac{C}{C_r} s \lambda_s + \frac{1}{\lambda_s} \left\| \frac{1}{\sqrt{n_s}} \tilde{X}_s b_s \right\|_2^2 .
\end{aligned}$$

Finally, we know that from the definition of the trim transform,

$$\left\| \frac{1}{\sqrt{n_s}} \tilde{X}_s b_s \right\|_2^2 \leq C \frac{p}{n_s} \|\Sigma_X^{-1} \mathbb{E}[f(H)^\top H] \phi_t\|_2^2 \leq C \frac{p \|\phi_t\|_2^2}{n_s \lambda_\Psi^2},$$

with Assumption 2. □

S.1.5 Lemmas for Theorem 2

Lemma S.6 *Under Assumption 1, the event $\mathcal{G}' = \cap_{j=10}^{12} \mathcal{G}_j$ happens with a probability*

$$\mathbb{P}\left(\mathcal{G}'\right) \geq 1 - p^{-c},$$

for a constant c.

Proof of Lemma S.6. From Lemma S.7, we know that for some c_1

$$\mathbb{P}(\mathcal{G}_{10}) \geq 1 - p^{-c_1}.$$

From Lemma S.9, we know that for some c_2

$$\mathbb{P}(\mathcal{G}_{11}) \geq 1 - p^{-c_2}.$$

From Lemma S.8, we know that for some c_3

$$\mathbb{P}(\mathcal{G}_{12}) \geq 1 - p^{-c_3}.$$

So there exist some c

$$\mathbb{P}(\mathcal{G}') \geq 1 - p^{-c}.$$

$\square$

Lemma S.7 *If $E_i, H_i, i = 1, \cdots, n$ are i.i.d. sampled from distributions for E and H that satisfies Assumption 1 and the function f satisfies the condition for f in Assumption 1, then with probability larger than $1 - p^{-c}$,*

$$\left\| \frac{1}{n} \tilde{X}_s^{(k)\top} \tilde{\varepsilon}_s^{(k)} \right\|_\infty \leq C_\varepsilon \sqrt{\frac{\log p}{n_s}}.$$

for some constants c, C_ε.

Proof of Lemma S.7. Because that ε_s is independent to X_s and $\varepsilon_{s,\cdot,i}, 1 \leq i \leq n_s$ are mean zero sub-Gaussian random variables. We have

$$\mathbb{P}\left(\left\| \frac{1}{n} \tilde{X}_s^{(k)\top} \tilde{\varepsilon}_s^{(k)} \right\|_\infty \geq C_\varepsilon \sqrt{\frac{\log p}{n_s}} \right) \leq \sum_{i=1}^{p} \mathbb{P}\left(\left| \frac{1}{n} \sum_{j=1}^{n_s} \tilde{X}_{s,i,j}^{(k)\top} \tilde{\varepsilon}_{s,,j}^{(k)} \right| \geq C_\varepsilon \sqrt{\frac{\log p}{n_s}} \right) \leq p^{-c}.$$

$\square$

Lemma S.8 *If $E_i, H_i, i = 1, \cdots, n$ are i.i.d. sampled from distributions for E and H that satisfies Assumption 1 and the function f satisfies the condition for f in Assumption 1, then with probability larger than $1 - p^{-c}$,*

$$\left\| \frac{1}{n_s} \mathcal{V}_s^2 \sum_{k=1}^{K} X_s^{(k)\top} X_s^{(k)} \left(\beta_s^{(k)} - \frac{1}{n_s} \sum_{k=1}^{K} n_s^{(k)} \beta_s^{(k)} \right) \right\|_\infty \leq C_k s_\delta \sqrt{\frac{\log p}{n_s}},$$

for some constants c, C_k.

Proof of Lemma S.8. From Assumption 1, we know that $E_{s,\cdot,i}$, $H_{s,\cdot,i}$, $f(H_s)_{\cdot,i}$ and ε. follows sub-Gaussian distributions. So $X_s^\top X_s/n_s$ is a random matrix with $(X_s)_{i,\cdot}$ as i.i.d. sub-Gaussian random variables.

Consider the perturbation form: $X_s^\top X_s/n_s = \Sigma + \left(X_s^\top X_s/n_s - \Sigma\right) := \Sigma + E$. We know that

$$\|E\|_{op} \le \sqrt{p}\left\|X_s^\top X_s/n_s - \Sigma\right\|_\infty \le C_x\sqrt{\frac{p\log p}{n}}$$

holds with probability larger than $1-6p^{-cC_x^2/16+2}$ from Lemma S.2. From the Weyl theorem, we know that

$$\left|\hat{\lambda}_i - \lambda_i\right| \le \|E\|_{op} \le C_x\sqrt{\frac{p\log p}{n}}.$$

From the Devis-Kahan theorem in the infinity norm (Theorem 15 in Bhardwaj & Vu (2024)), if the eigen-gap $\Delta_i = \lambda_i - \lambda_{i+1}$ for those eigenvalues $\lambda_i, 1 \le i \le \tau$ is larger than $p^{1/2}n^{-1/4}\log^{3/4} p$, then

$$\begin{aligned}
\|\hat{v}_i - v_i\|_\infty &\le c\|V\|_\infty\left(\|\hat{v}_i - v_i\|_2 + \frac{\sqrt{\|E\|_\infty p}\log p}{\lambda_i}\right)\\
&\le c\|V\|_\infty\left(\frac{\|E\|_{op}}{\Delta_i} + \frac{\sqrt{\|E\|_\infty p}\log p}{\lambda_i}\right)\\
&\le c\sqrt{\frac{\log p}{p}}\left(n_s^{-1/4}\log^{-1/4} p + n_s^{-1/4}p^{1/2}\lambda_\tau^{-1}\log^{5/4} p\right) \le c\left(\frac{\log p}{n_s p^2}\right)^{1/4}.
\end{aligned}$$

Here the second inequality is the Devis-Kahan theorem in the l_2 norm, the last inequality using the Assumption 2.

Define $\mathcal{V} = V\,\mathrm{diag}(\tau_s/D_{1:\tau}, 1_{p-\tau})V^\top$ where $\Sigma_X = VDV^\top$. It holds that

$$
\begin{aligned}
\left\|\mathcal{V}_s^2\frac{1}{n_s}X_s^\top X_s - \mathcal{V}^2\Sigma_X\right\|_\infty &\le \left\|X_s^\top X_s - \Sigma_X\right\|_\infty + \left\|\left(I_p - \mathcal{V}_s^2\right)X_s^{(k)}X_s - \left(I_p - \mathcal{V}^2\right)\Sigma_X\right\|_\infty \\
&\le \left\|X_s^{(k)}X_s - \Sigma_X\right\|_\infty + \sum_{i=1}^{\tau}\left\|(\hat{\lambda}_i - \tau_s^2)\hat{v}_i\hat{v}_i^\top - (\lambda_i - \tau_s^2)v_iv_i^\top\right\|_\infty \\
&\le C_x\sqrt{\frac{\log p}{n_s}} + \sum_{i=1}^{\tau}\left(|\hat{\lambda}_i - \lambda_i|\,\|v_i\|_\infty^2 + \left|\hat{\lambda}_i - \tau_s^2\right|\,\|\hat{v}_i - v_i\|_\infty^2\right) \\
&\le C_x\sqrt{\frac{\log p}{n_s}} + C\left(\sum_{i=1}^{\tau}\sqrt{\frac{p\log p}{n_s}}\left(\sqrt{\frac{\log p}{p}}\right)^2\right. \\
&\qquad\qquad \left. + \sum_{i=1}^{\tau}\left(\sqrt{\frac{p\log p}{n_s}} + |\lambda_i - \tau_s^2|\right)\sqrt{\frac{\log p}{n_s p^2}}\right) \\
&\le C_k\sqrt{\frac{\log p}{n_s}}. \qquad\qquad (S.21)
\end{aligned}
$$

Here we use Assumption 2 and $\sum_{i=1}^{\tau}|\lambda_i - \tau_s^2| \le cp$. If the element eigen-gap condition is not hold, we only need the eigen gap $\Delta_\tau = \lambda_\tau - \lambda_{\tau+1}$ is larger than $p^{1/2}n^{-1/4}\log^{3/4}p$ holds for a τ. From Lemma S.13 in supplementary of Chen & Yao (2024) or Theorem 2 in Pensky (2024), we have the subspace version of Devis-kahan theorem in the two-to-infinity norm that

$$
\left\|\hat{V}_{\cdot,1:\tau} - V_{\cdot,1:\tau}R\right\|_{2,\infty} \le c\|V_{\cdot,1:\tau}\|_{2,\infty}\left(\frac{\|E\|_{op}}{\Delta_\tau} + \frac{\sqrt{\|E\|_\infty p}\log p}{\lambda_\tau}\right)
$$

holds for a rotation matrix $\mathcal{R} \in \mathbb{R}^{\tau\times\tau}$ and the inequality (S.21) still holds.

So

$$
\begin{aligned}
&\left\|\frac{1}{n_s}\mathcal{V}_s^2\sum_{k=1}^{K}X_s^{(k)\top}X_s^{(k)}\left(\beta_s^{(k)} - \frac{1}{n_s}\sum_{k=1}^{K}n_s^{(k)}\beta_s^{(k)}\right)\right\|_\infty \\
=\ &\left\|\frac{1}{n_s}\sum_{k=1}^{K}\left(n_s^{(k)}\mathcal{V}_s^2\left(\frac{1}{n_s^{(k)}}X_s^{(k)\top}X_s^{(k)} - \mathcal{V}^2\Sigma_X\right)\right)\left(\beta_s^{(k)} - \frac{1}{n_s}\sum_{k=1}^{K}n_s^{(k)}\beta_s^{(k)}\right)\right\|_\infty \\
\le\ &\left\|\frac{1}{n_s}\mathcal{V}_s^2\sum_{k=1}^{K}X_s^{(k)\top}X_s^{(k)} - \mathcal{V}^2\Sigma_X\right\|_\infty\sum_{k=1}^{K}\frac{n_s^{(k)}}{n_s}\left\|\beta_s^{(k)} - \frac{1}{n_s}\sum_{k=1}^{K}n_s^{(k)}\beta_s^{(k)}\right\|_1 \\
\le\ &C_k s_\delta\sqrt{\frac{\log p}{n_s}}.
\end{aligned}
$$

$\square$

Lemma S.9 *If $E_i, H_i, i = 1, \cdots, n$ are i.i.d. sampled from distributions for E and H that satisfies Assumption 1 and the function f satisfies the condition for f in Assumption 1, then with probability larger than $1 - p^{-c} - \exp(-cn)$,*

$$\left\| \frac{1}{n_s} \sum_{k=1}^{K} \tilde{X}_s^{(k)\top} \left(\tilde{H}_s^{(k)} \phi_s^{(k)} - \tilde{X}_s^{(k)} b_s \right) \right\|_\infty \leq C_k C_\phi \sqrt{\frac{q \log p}{n_s}} + 2C_t \sqrt{\frac{p}{n_s}} \lambda_\Psi^{-1} \|\phi_t\|_2.$$

for some constants c, C_k, C_t.

Proof of Lemma S.9. Recall the definition of trim transform $\mathcal{V}_s$. We know that $\mathcal{V}_s^2 \sum_{k=1}^K X_s^{(k)\top} X_s^{(k)} = \tilde{X}_s^{(k)\top} \tilde{X}_s^{(k)}$ and $\mathcal{V}_s^2 \sum_{k=1}^K X_s^{(k)\top} H_s^{(k)} = \tilde{X}_s^{(k)\top} \tilde{H}_s^{(k)}$. So

$$\begin{aligned}
&\left\| \frac{1}{n_s} \sum_{k=1}^{K} \tilde{X}_s^{(k)\top} \left(\tilde{H}_s^{(k)} \phi_s^{(k)} - \tilde{X}_s^{(k)} b_s \right) \right\|_\infty \\
= &\left\| \frac{1}{n_s} \mathcal{V}_s^2 \sum_{k=1}^{K} X_s^{(k)\top} \left(H_s^{(k)} \phi_s^{(k)} - X_s^{(k)} b_s \right) \right\|_\infty \\
\leq &\left\| \frac{1}{n_s} \mathcal{V}_s^2 \sum_{k=1}^{K} \left(X_s^{(k)\top} H_s^{(k)} - \mathbb{E}[f(H)^\top H] \right) \phi_t \right\|_\infty + \left\| \frac{1}{n_s} \mathcal{V}_s^2 \sum_{k=1}^{K} \left(\Sigma_X - X_s^{(k)\top} X_s^{(k)} \right) \Sigma_X^{-1} \mathbb{E}[f(H)^\top H] \phi_t \right\|_\infty.
\end{aligned}$$

For the first term, we use the sub-gaussian of $\mathcal{V}_s^2 X_{si}^{(k)}$ and $H_{si}^{(k)}$, so that

$$\left\| \frac{1}{n_s} \mathcal{V}_s^2 \sum_{k=1}^{K} \left(X_s^{(k)\top} H_s^{(k)} - \mathbb{E}[f(H)^\top H] \right) \phi_t \right\|_\infty \leq C_k \sqrt{\frac{\log p}{n_s}} \|\phi_t\|_2.$$

For the second term, we have

$$\begin{aligned}
&\left\| \frac{1}{n_s} \mathcal{V}_s^2 \sum_{k=1}^{K} \left(\Sigma_X - X_s^{(k)\top} X_s^{(k)} \right) \Sigma_X^{-1} \mathbb{E}[f(H)^\top H] \phi_t \right\|_\infty \\
\leq &\left\| \frac{1}{n_s} \mathcal{V}_s^2 \sum_{k=1}^{K} \left(\Sigma_X - X_s^{(k)\top} X_s^{(k)} \right) \right\|_{2,\infty} \left\| \Sigma_X^{-1} \mathbb{E}[f(H)^\top H] \phi_t \right\|_2 \\
\leq &C_t \sqrt{\frac{p}{n_s}} \lambda_\Psi^{-1} \|\phi_t\|_2.
\end{aligned}$$

So

$$\begin{aligned}
\left\| \frac{1}{n_s} \sum_{k=1}^{K} \tilde{X}_s^{(k)\top} \left(\tilde{H}_s^{(k)} \phi_s^{(k)} - \tilde{X}_s^{(k)} b_s \right) \right\|_\infty &\leq C_k \sqrt{\frac{\log p}{n_s}} \|\phi\|_2 + 2C_t \sqrt{\frac{p}{n_s}} \lambda_\Psi^{-1} \|\phi_t\|_2 \\
&\leq C_k C_\phi \sqrt{\frac{\log p}{n_s}} + 2C_t \sqrt{\frac{p}{n_s}} \lambda_\Psi^{-1} \|\phi_t\|_2.
\end{aligned}$$

$\square$

S.1.6 Proof of Theorem 3

Proof of Theorem 3. On the event $\mathcal{G} \cap \mathcal{G}'$, Theorem 1 gives

$$\|\hat{\eta} - \eta_0\|_1 \leq C_7 s_\eta \sqrt{\frac{\log p}{n_t}},$$

when choosing $\lambda_t = \sqrt{\log p / n_t}$. From Theorem 2, it holds that

$$\|\hat{\beta}_s - \beta_s\|_1 \leq C_6 \sqrt{\frac{p\|\phi_t\|_2^2}{n_s \lambda_\Psi^2}} \vee C_3 s \left(\sqrt{\frac{\log p}{n_s}} + \frac{\|\phi_t\|_2^2}{\lambda_\Psi} \right) \leq C_6 \sqrt{\frac{p\|\phi_t\|_2^2}{n_s \lambda_\Psi^2}} \vee C s \sqrt{\frac{\log p}{n_s}},$$

when choosing $\lambda_s = \sqrt{\log p / n_s} \vee \lambda_\Psi^{-1} \|\phi_t\|_2 \sqrt{p/n_s}$. So

$$\begin{aligned}
\|\hat{\beta}_t - \beta_t\|_1 \leq \|\hat{\eta} - \eta_0\|_1 + \|\hat{\beta}_s - \beta_s\|_1 &\leq C_8 \sqrt{\frac{p\|\phi_t\|_2^2}{n_s \lambda_\Psi^2}} \vee C s \sqrt{\frac{\log p}{n_s}} \vee C_7 s_\eta \sqrt{\frac{\log p}{n_t}} \\
&\leq C_6 \sqrt{\frac{p\|\phi_t\|_2^2}{n_s \lambda_\Psi^2}} \vee C_7 s_\eta \sqrt{\frac{\log p}{n_t}}.
\end{aligned}$$

The results for l_2 norm error is similar as

$$\|\hat{\beta}_t - \beta_t\|_2 \leq \|\hat{\eta} - \eta_0\|_2 + \|\hat{\beta}_s - \beta_s\|_2 \leq C_6 \sqrt{\frac{p\|\phi_t\|_2^2}{n_s \lambda_\Psi^2}} \vee C_7 \sqrt{\frac{s_\eta \log p}{n_t}}.$$

$\square$

S.1.7 Proof of Proposition 3

Proof of Proposition 3. We consider the special case that $\beta_s^{(k)} = \beta_s, 1 \leq k \leq K$ and the noise distribution is Gaussian distribution $N(0, \sigma^2)$. We then consider the case that $f(H) = H\Psi$ with $\lambda_1(\Psi) = \cdots = \lambda_q(\Psi) = \lambda_\Psi$.

Define the distribution of target data as $P_t(\beta_t, \phi_t)$ and the distribution of source data as $P_s(\beta_s, \phi_t)$. Define that

$$D_{KL}(P(\beta_{t1}) \| P(\beta_{t2})) = \min_{\substack{\beta_{t1}=\beta_1+\eta_1, \beta_{t2}=\beta_2+\eta_2 \\ \beta_1,\beta_2 \in \mathcal{B}_0(s), \eta_1,\eta_2 \in \mathcal{B}_0(s_\eta) \\ \phi_1,\phi_2 \in \mathcal{B}_2(c_\phi)}} D_{KL}(P_t(\beta_1+\eta_1, \phi_1) \| P_t(\beta_2+\eta_2, \phi_2))$$

$$+ D_{KL}(P_s(\beta_1, \phi_1) \| P_s(\beta_2, \phi_2)).$$

Similar to those in the proof of Proposition 1, we define the δ-packing of $\mathcal{B}_0(s_\eta + s)$ as $\mathcal{H}_{s_\eta+s}$ in l_2 norm. So

$$|\mathcal{H}_{s_\eta}| \geq \exp\left\{ \frac{s_\eta + s}{2} \log \frac{p - s_\eta - s}{(s_\eta + s)/2} \right\}.$$

From Fano's lemma,

$$\inf_{\hat{\beta}_t} \sup_{\substack{\beta_t = \beta + \eta \\ \eta \in \mathcal{B}_0(s_\eta), \beta \in \mathcal{B}_0(s)}} \mathbb{P}(\|\hat{\beta}_t - \beta_t\|_2 \geq \delta) \geq 1 - \frac{\max_{i,j} D_{KL}(P(\beta_{ti}) \| P(\beta_{tj})) + \log 2}{\log |\mathcal{H}_{s_\eta+s}|}.$$

Case 1. We consider $E = 0$ and H follows Gaussian distribution $N(0, I_q)$, then

$$P_t(\beta_t, \phi_t) = N\left(X_t\beta_t, (\|\phi_t\|_2^2 + \sigma^2) I_{n_t}\right),$$

$$P_s(\beta_s, \phi_t) = N\left(X_s\beta_s, (\|\phi_t\|_2^2 + \sigma^2) I_{n_s}\right).$$

So for any $\eta_1, \eta_2 \in \mathcal{B}_0((s_\eta)$, $\beta_1, \beta_2 \in \mathcal{B}_0(s)$ and $\phi_1, \phi_2 \in \mathcal{B}_2(c_\phi)$, we have

$$\begin{aligned} & D_{KL}(P_t(\beta_1+\eta_1, \phi_1) \| P_t(\beta_2+\eta_2, \phi_2)) \\ = & \frac{n_t}{2} \log \frac{\|\phi_2\|_2^2 + \sigma^2}{\|\phi_1\|_2^2 + \sigma^2} + \frac{\|X_t(\eta_1 - \eta_2 + (\beta_1 - \beta_2))\|_2^2 + n_t(\|\phi_1\|_2^2 - \|\phi_2\|_2^2)}{2(\sigma^2 + \|\phi_2\|_2^2)}. \end{aligned}$$

And

$$\begin{aligned} & D_{KL}(P_s(\beta_1, \phi_1) \| P_s(\beta_2, \phi_2)) \\ = & \frac{n_s}{2} \log \frac{\|\phi_2\|_2^2 + \sigma^2}{\|\phi_1\|_2^2 + \sigma^2} + \frac{\|X_s(\eta_1 - \eta_2 + (\beta_1 - \beta_2))\|_2^2 + n_s(\|\phi_1\|_2^2 - \|\phi_2\|_2^2)}{2(\sigma^2 + \|\phi_2\|_2^2)}. \end{aligned}$$

So

$$D_{KL}(P(\beta_{t1})\|P(\beta_{t2})) \leq \min_{\substack{\beta_{t1}=\beta_1+\eta_1,\beta_{t2}=\beta_2+\eta_2,\\ \beta_1,\beta_2\in\mathcal{B}_0(s)\\ \eta_1,\eta_2\in\mathcal{B}_0(s_\eta)\\ \|\phi_1\|_2=c_\phi,\phi_2=\phi_1}} D_{KL}(P_t(\beta_1+\eta_1,\phi_1)\|P_t(\beta_2+\eta_2,\phi_2))$$

$$\begin{aligned}
&+D_{KL}(P_s(\beta_1,\phi_1)\|P_s(\beta_2,\phi_2)) \\
&= \frac{\|X_t((\eta_1-\eta_2)+(\beta_1-\beta_2))\|_2^2+\|X_s(\beta_1-\beta_2)\|_2^2}{2(\sigma^2+\|\phi_2\|_2^2)} \\
&\leq \frac{C\lambda_\Psi^2 p^{-1}(n_t\|\eta_1-\eta_2+(\beta_1-\beta_2)\|_2^2+n_s\|\beta_1-\beta_2\|_2^2)}{(2(\sigma^2+\|\phi_2\|_2^2)}.
\end{aligned}$$

Choose $\delta = c_1 p^{1/2} n_s^{-1/2}\lambda_\Psi^{-1}C_\phi$, with $n_s \gg n_t$, then

$$D_{KL}(P(\beta_{t1})\|P(\beta_{t2})) \leq 3.$$

So

$$\inf_{\hat{\beta}_t} \sup_{\substack{\beta_t=\beta+\eta\\ \eta\in\mathcal{B}_0(s_\eta),\beta\in\mathcal{B}_0(s)}} \mathbb{P}\left(\|\hat{\beta}_t-\beta_t\|_2 \geq c_1\sqrt{\frac{p\|\phi_t\|_2^2}{n_s\lambda_\Psi^2}}\right) \geq 1-\frac{6+2\log 2}{(s_\eta+s)\log(2(p-s_\eta-s)/(s_\eta+s)} \geq \frac{1}{2}.$$

Case 2. We consider $H=0, c_\phi=0$ and E follows Gaussian distribution $N(0,I_p)$. Under these conditions, the problem reduces to transfer learning for high-dimensional linear regression. Theorem B in the supplementary materials of Li et al. (2022) provides the minimax lower bound as follows:

$$\inf_{\hat{\beta}_t} \sup_{\substack{\beta_t=\beta+\eta\\ \eta\in\mathcal{B}_0(s_\eta),\beta\in\mathcal{B}_0(s)}} \mathbb{P}\left(\|\hat{\beta}_t-\beta_t\|_2 \geq c_2\sqrt{\frac{s_\eta\log p}{n_t}} \vee \sqrt{\frac{s\log p}{n_s}}\right) \geq \frac{1}{2}.$$

With the assumption that $n_s \gg n_t$ and s_η, s are constants, finally we have

$$\inf_{\hat{\beta}_t} \sup_{\substack{\beta_t=\beta+\eta\\ \eta\in\mathcal{B}_0(s_\eta),\beta\in\mathcal{B}_0(s)}} \mathbb{P}\left(\|\hat{\beta}_t-\beta_t\|_2 \geq c_1\sqrt{\frac{p\|\phi_t\|_2^2}{n_s\lambda_\Psi^2}} \vee c_2\sqrt{\frac{s_\eta\log p}{n_t}}\right) \geq \frac{1}{2}.$$

□

S.1.8 Proof of Proposition 2 and 4

Proof of Proposition 2 and 4. We consider the special case that $\beta_s^{(k)} = \beta_s, 1 \le k \le K$ and the noise distribution is Gaussian distribution $N(0, \sigma^2)$. We then consider the case that $f(H) = H\Psi$ with $\lambda_1(\Psi) = \cdots = \lambda_q(\Psi) = \lambda_\Psi$.

Define the distribution of target data as $P_t(\eta, \beta_s, \phi_t)$. Also define that

$$D_{KL}(P_t(\eta_1)\|P_t(\eta_2)) = \min_{\beta_1,\beta_2\in\mathcal{B}_0(s),\phi_1,\phi_2\in\mathcal{B}_2(c_\phi)} D_{KL}(P_t(\eta_1,\beta_1,\phi_1)\|P_t(\eta_2,\beta_2,\phi_2)).$$

Similar to those in the proof of Proposition 1, we define the δ-packing of $\mathcal{B}_0(s_\eta)$ as $\mathcal{H}_{s_\eta}$ in l_2 norm. So

$$\left|\mathcal{H}_{s_\eta}\right| \ge \exp\left\{\frac{s_\eta}{2}\log\frac{p-s_\eta}{s_\eta/2}\right\}.$$

From Fano's lemma,

$$\inf_{\tilde\eta}\sup_{\eta\in\mathcal{B}_0(s_\eta)} \mathbb{P}(\|\tilde\eta-\eta_0\|_2 \ge \delta) \ge 1 - \frac{\max_{i,j} D_{KL}(P_t(\eta_i)\|P_t(\eta_j)) + \log 2}{\log\|\mathcal{H}_{s_\eta}\|}.$$

Case 1. We consider $E = 0$ and H follows Gaussian distribution $N(0, I_q)$, then

$$P_t(\eta,\beta_s,\phi_t) = N\left(X_t\left(\eta+\beta_s-\hat\beta_s\right), \left(\|\phi_t\|_2^2+\sigma^2\right)I_n\right).$$

So for $\eta_1,\eta_2\in\mathcal{B}_0(s_\eta)$, $\beta_1,\beta_2\in\mathcal{B}_0(s)$ and $\phi_1,\phi_2\in\mathcal{B}_2(c_\phi)$, we have

$$\begin{aligned}
&D_{KL}(P_t(\eta_1,\beta_1,\phi_1)\|P_t(\eta_2,\beta_2,\phi_2))\\
=&\frac{n_t}{2}\log\frac{\|\phi_2\|_2^2+\sigma^2}{\|\phi_1\|_2^2+\sigma^2} + \frac{\|X_t(\eta_1-\eta_2+(\beta_1-\hat\beta_1)-(\beta_2-\hat\beta_2))\|_2^2 + n_t(\|\phi_1\|_2^2-\|\phi_2\|_2^2)}{2(\sigma^2+\|\phi_2\|_2^2)}.
\end{aligned}$$

So

$$\begin{aligned}
D_{KL}(P_t(\eta_1)\|P_t(\eta_2)) &\le \min_{\beta_1,\beta_2\in\mathcal{B}_0(s),\|\phi_1\|_2=c_\phi,\phi_2=\phi_1} D_{KL}(P_t(\eta_1,\beta_1,\phi_1)\|P_t(\eta_2,\beta_2,\phi_2))\\
&= \frac{C\lambda_\Psi^2 p^{-1} n_t\|\eta_1-\eta_2+(\beta_1-\hat\beta_1)-(\beta_2-\hat\beta_2)\|_2^2}{2(\sigma^2+\|\phi_1\|_2^2)}.
\end{aligned}$$

Here the last inequality is from $X_t = H_t\Psi$ and $\eta_1-\eta_2+(\beta_1-\hat\beta_1)-(\beta_2-\hat\beta_2))$ is independent to $H_t\Psi$. So with probability larger than $1-p^{-c}$, the length of the projection for the direction vector of $\eta_1-\eta_2+(\beta_1-\hat\beta_1)-(\beta_2-\hat\beta_2)$ to the row space of Ψ can be bounded by $Cp^{-1/2}$.

Choose $\delta = c_1 p^{1/2} n_t^{-1/2}\lambda_\Psi^{-1} c_\phi$ for some constant c_1, then

$$\begin{aligned}
&\frac{C\lambda_\Psi^2 p^{-1} n_t \|\eta_1-\eta_2+(\beta_1-\hat\beta_1)-(\beta_2-\hat\beta_2)\|_2^2}{2(\sigma^2+\|\phi_1\|_2^2)}\\
\le\ &\frac{2C\lambda_\Psi^2 p^{-1} n_t\left(\delta^2+(C\sqrt{s}pc_\phi^2)/(n_s\lambda_\Psi^2)\right)}{2(\sigma^2+\|\phi_1\|_2^2)} \le 3.
\end{aligned}$$

The first inequality uses Theorem 2 and $n_s \gg n_t$. Then

$$D_{KL}(P_t(\eta_1)\|P_t(\eta_2)) \le 3.$$

So

$$\inf_{\tilde\eta}\sup_{\eta\in\mathcal{B}_0(s_\eta)} \mathbb{P}\left(\|\tilde\eta-\eta_0\|_2 \ge c_1\sqrt{\frac{p\|\phi_t\|_2^2}{n_t\lambda_\Psi^2}}\right) \ge 1-\frac{6+2\log 2}{s_\eta\log(2(p-s_\eta)/s_\eta)} \ge \frac{1}{2}. \tag{S.22}$$

Case 2. We consider $H=0, c_\phi=0$ and E follows Gaussian distribution $N(0,I_p)$, then

$$\begin{aligned}
D_{KL}(P_t(\eta_1)\|P_t(\eta_2)) &\le \frac{1}{2\sigma^2}\|X_t(\eta_1-\eta_2+(\beta_1-\hat\beta_1)-(\beta_2-\hat\beta_2))\|_2^2\\
&\le \frac{2n_t\|\eta_i-\eta_j+(\beta_1-\hat\beta_1)-(\beta_2-\hat\beta_2)\|_2^2}{\sigma^2}.
\end{aligned}$$

Choose $\delta = c_2\sqrt{s_\eta\log p/n_t}$, with $\|\beta_s-\hat\beta_s\|_2 \le c\sqrt{s\log p/n_s}$, we know that

$$D_{KL}(P_t(\eta_1)\|P_t(\eta_2)) \le \frac{c_2 s_\eta\log p}{\sigma^2}.$$

and then

$$\inf_{\tilde\eta}\sup_{\eta\in\mathcal{B}_0(s_\eta)} \mathbb{P}\left(\|\tilde\eta-\eta_0\|_2 \ge c_2\sqrt{\frac{s_\eta\log p}{n_t}}\right) \ge \frac{1}{2}. \tag{S.23}$$

With (S.22) and (S.23), we have

$$\inf_{\tilde\eta}\sup_{\eta\in\mathcal{B}_0(s_\eta)} \mathbb{P}\left(\|\tilde\eta-\eta_0\|_2 \ge c_1\sqrt{\frac{p\|\phi_t\|_2^2}{n_t\lambda_\Psi^2}} \vee c_2\sqrt{\frac{s_\eta\log p}{n_t}}\right) \ge \frac{1}{2}.$$

For the l_1 norm lower bound. The case 1 is the same as above proof. For case 2, we only need to choose $\delta = c_2\sqrt{s_\eta \log p/n_t}$. Because $\|\eta_i - \eta_j\|_2 \leq \|\eta_i - \eta_j\|_1$. So we still have

$$D_{KL}(P_t(\eta_1)\|P_t(\eta_2)) \leq \frac{c_2 s_\eta \log p}{\sigma^2}.$$

So

$$\inf_{\tilde{\eta}} \sup_{\eta \in \mathcal{B}_0(s_\eta)} \mathbb{P}\left(\|\tilde{\eta} - \eta_0\|_1 \geq c_1\sqrt{\frac{p\|\phi_t\|_2^2}{n_t\lambda_\Psi^2}} \vee c_2\sqrt{s_\eta \frac{\log p}{n_t}}\right) \geq \frac{1}{2}.$$

For the definition of $\tilde{\beta}_t = \tilde{\eta} + \tilde{\beta}_s$, we have

$$\mathbb{P}\left(\|\tilde{\beta}_t - \beta_t\|_2 \geq \delta\right) \geq \mathbb{P}\left(\|\tilde{\eta} - \eta_0\|_2 - \|\tilde{\beta}_s - \beta_s\|_2 \geq \delta\right) \geq \mathbb{P}\left(\|\tilde{\eta} - \eta_0\|_2 \geq \frac{\delta}{2}\right).$$

This is because $\delta \gg \|\tilde{\eta} - \eta_0\|_2$ from $n_s \gg n_t$. $\square$

S.1.9 Proof of Theorem 4.

Proof of Theorem 4. Recall the event $\mathcal{G}, \mathcal{G}'$ defined in the proof of Theorem 1 and 2, similarly define $\mathcal{G}^{(k)}$ with only the Kth source data are used. Formally, we define that

$$\left\{\left\|\frac{1}{n_s(k)} X_s^{(k)\top} X_s^{(k)} - \Sigma_X\right\|_\infty \leq C_x\sqrt{\frac{\log p}{n_s^{(k)}}}\right\}, \tag{$\mathcal{G}_2^{(k)}$}$$

$$\left\{\left\|\frac{1}{n_s^{(k)}} f(H_s^{(k)})^\top H_s^{(k)} - \mathbb{E}_H f(H)^\top H\right\|_{2,\infty} \leq C_H\sqrt{\frac{q\log p}{n_s^{(k)}}}\right\}, \tag{$\mathcal{G}_5^{(k)}$}$$

$$\left\{\left\|\frac{1}{n_s^{(k)}} E_s^{(k)\top} H_s^{(k)}\right\|_{2,\infty} \leq C_H\sqrt{\frac{q\log p}{n_s^{(k)}}}\right\}, \tag{$\mathcal{G}_7^{(k)}$}$$

$$\left\{\left\|\frac{1}{n_s^{(k)}} X_s^{(k)\top} \varepsilon_s^{(k)}\right\|_\infty \leq C_\varepsilon\sqrt{\frac{\log p}{n_s^{(k)}}}\right\}, \tag{$\mathcal{G}_9^{(k)}$}$$

and $\mathcal{G}^{(k)} = \mathcal{G}_2^{(k)} \cap \mathcal{G}_5^{(k)} \cap \mathcal{G}_7^{(k)} \cap \mathcal{G}_9^{(k)}$. Similarly to Lemma S.1 and S.6, it holds that

$$\mathbb{P}\left(\mathcal{G}^{(k)}\right) \geq 1 - p^{-c}.$$

With the definition of $v^{(k)}$, we have

$$
\begin{aligned}
v^{(k)} = & \left\| \frac{1}{n_t} X_t^\top \left(X_t \left(\beta_t - \hat{\beta}_{\text{init}}^{(k)} - \hat{\eta}^{(k)} \right) + H_t \phi_t + \varepsilon_t \right) \right. \\
& \qquad \left. - \frac{1}{n_s^{(k)}} X_s^{(k)\top} \hat{Z}_s^{(k)} + \left(\frac{1}{n_s^{(k)}} X_s^{(k)\top} \hat{Z}_s^{(k)} - \frac{1}{n_t} X_t^\top \hat{Z}_t^{(k)} \right) \right\|_\infty \\
\leq & \left\| \mathbb{E}[X^\top H] \left(\phi_t - \phi_s^{(k)} \right) \right\|_\infty \\
+ & \left\| \frac{1}{n_t} X_t^\top \left(X_t \left(\beta_s^{(k)} - \hat{\beta}_{\text{init}}^{(k)} \right) \right) - \frac{1}{n_s^{(k)}} X_s^{(k)\top} \hat{Z}_s^{(k)} + \frac{1}{n_s^{(k)}} X_s^{(k)\top} H_s^{(k)} \phi_s^{(k)} \right\|_\infty && (A_1^{(k)}) \\
+ & \left\| \frac{1}{n_t} X_t^\top \left(X_t \left(\beta_t - \beta_s^{(k)} - \hat{\eta}^{(k)} \right) \right) \right\|_\infty && (A_2^{(k)}) \\
+ & \left\| \frac{1}{n_t} X_t^\top \varepsilon_t \right\|_\infty && (A_3^{(k)}) \\
+ & \left\| \frac{1}{n_t} X_t^\top \hat{Z}_t^{(k)} - \frac{1}{n_s^{(k)}} X_s^{(k)\top} \hat{Z}_s^{(k)} \right\|_\infty && (A_4^{(k)}) \\
+ & \left\| \left(\frac{1}{n_t} X_t^\top H_t - \frac{1}{n_s^{(k)}} X_s^{(k)\top} H_s^{(k)} \right) \phi_s^{(k)} \right\|_\infty && (A_5^{(k)}) \\
+ & \left\| \left(\frac{1}{n_t} X_t^\top H_t - \mathbb{E}[X^\top H] \right) \left(\phi_t - \phi_s^{(k)} \right) \right\|_\infty . && (A_6^{(k)})
\end{aligned}
$$

For term $A_1^{(k)}$, it holds that

$$
\begin{aligned}
& \left\| \frac{1}{n_t} X_t^\top \left(X_t \left(\beta_s^{(k)} - \hat{\beta}_{\text{init}}^{(k)} \right) \right) - \frac{1}{n_s^{(k)}} X_s^{(k)\top} \hat{Z}_s^{(k)} + \frac{1}{n_s^{(k)}} X_s^{(k)\top} H_s^{(k)} \phi_s^{(k)} \right\|_\infty \\
= & \left\| \frac{1}{n_t} X_t^\top \left(X_t \left(\beta_s^{(k)} - \hat{\beta}_{\text{init}}^{(k)} \right) \right) - \frac{1}{n_s^{(k)}} X_s^{(k)\top} \left(X_s^{(k)} \left(\beta_s^{(k)} - \hat{\beta}_{\text{init}}^{(k)} \right) + \varepsilon_s^{(k)} \right) \right\|_\infty \\
\leq & \left\| \frac{1}{n_t} X_t^\top X_t - \Sigma_X \right\|_\infty \left\| \beta_s^{(k)} - \hat{\beta}_{\text{init}}^{(k)} \right\|_1 + \left\| \frac{1}{n_s^{(k)}} X_s^{(k)\top} X_s^{(k)} - \Sigma_X \right\|_\infty \left\| \beta_s^{(k)} - \hat{\beta}_{\text{init}}^{(k)} \right\|_1 \\
+ & \left\| \frac{1}{n_s^{(k)}} X_s^{(k)\top} \varepsilon_s^{(k)} \right\|_\infty \\
\leq & \, C_s \left(\sqrt{\frac{\log p}{n_t}} \right),
\end{aligned}
$$

in events $\mathcal{G}_1$,$\mathcal{G}_2^{(k)}$,$\mathcal{G}_9^{(k)}$.

For term $A_2^{(k)}$, it holds that

$$\begin{aligned}\left\|\frac{1}{n_t}X_t^\top X_t\left(\beta_t-\beta_s^{(k)}-\hat{\eta}^{(k)}\right)\right\|_\infty &\le \left(\left\|\frac{1}{n_t}X_t^\top X_t-\Sigma_X\right\|_\infty+\|\Sigma_X\|_\infty\right)\left\|\beta_t-\beta_s^{(k)}-\hat{\eta}^{(k)}\right\|_1\\ &\le C\left(\sqrt{\frac{p\|\phi_t-\phi_s^{(k)}\|_2^2}{n_t\lambda_\Psi^2}}\vee\left(\sqrt{q}\|\phi_s^{(k)}\|_2+1\right)s_\eta\sqrt{\frac{\log p}{n_t}}\right)\end{aligned}$$

Here we use Lemma S.10 in the last inequality.

For term $A_3^{(k)}$, it holds that

$$\left\|\frac{1}{n_t}X_t^\top\varepsilon_t\right\|_\infty\le C_\varepsilon\sqrt{\frac{\log p}{n_t}}$$

in event $\mathcal{G}_8$.

For term $A_4^{(k)}$, we have the decomposition

$$\begin{aligned}\left\|\frac{1}{n_t}X_t^\top\hat{Z}_t^{(k)}-\frac{1}{n_s^{(k)}}X_s^{(k)\top}\hat{Z}_s^{(k)}\right\|_\infty &\le \left\|\left(\frac{1}{n_t}X_t^\top X_t-X_s^{(k)\top}X_s^{(k)}\right)\left(\beta_s-\hat{\beta}_{\text{init}}^{(k)}\right)\right\|_\infty && (B_1^{(k)})\\ &+\left\|\frac{1}{n_s^{(k)}}X_s^{(k)\top}X_s^{(k)}\left(\beta_s^{(k)}-\beta_s^{(k)}\right)\right\|_\infty && (B_2^{(k)})\\ &+\min_{\omega\in\mathbb{R}^p}\left\|\frac{1}{n_t}X_t^\top X_t\omega-\frac{1}{n_s^{(k)}}X_s^{(k)\top}X_s^{(k)}b_s^{(k)}\right\|_\infty && (B_3^{(k)})\\ &+\left\|\frac{1}{n_s^{(k)}}X_s^{(k)\top}\varepsilon_s^{(k)}\right\|_\infty. && (B_4^{(k)})\end{aligned}$$

The term $B_1^{(k)}$,$B_2^{(k)}$ and $B_4^{(k)}$ can be bounded in the same way as for term B_1, B_2 and B_4 in Theorem 1.

For term $B_3^{(k)}$, in the event $\mathcal{G}_4$ and $\mathcal{G}_5^{(k)}$,

$$\begin{aligned}\min_{\omega\in\mathbb{R}^p}\left\|\frac{1}{n_t}X_t^\top X_t\omega-\frac{1}{n_s^{(k)}}X_s^{(k)\top}X_s^{(k)}b_s^{(k)}\right\|_\infty &\le \left\|\frac{1}{n_t}X_t^\top H_t\phi_s^{(k)}-\frac{1}{n_s^{(k)}}X_s^{(k)\top}H_s^{(k)}\phi_s^{(k)}\right\|_\infty\\ &\le\left\|\frac{1}{n_t}X_t^\top H_t-\mathbb{E}_Hf(H)^\top H\right\|_{2,\infty}\left\|\phi_s^{(k)}\right\|_2\\ &+\left\|\frac{1}{n_s^{(k)}}X_s^{(k)\top}H_s^{(k)}-\mathbb{E}_Hf(H)^\top H\right\|_{2,\infty}\left\|\phi_s^{(k)}\right\|_2\\ &\le C_H\left(\sqrt{\frac{q\log p}{n_t}}+\sqrt{\frac{q\log p}{n_s^{(k)}}}\right)\left\|\phi_s^{(k)}\right\|_2.\end{aligned}$$

Then with $n_s^{(k)} \gg n_t$, it holds that

$$\left\| \frac{1}{n_t} X_t^\top \hat{Z}_t^{(k)} - \frac{1}{n_s^{(k)}} X_s^{(k)\top} \hat{Z}_s^{(k)} \right\|_\infty \leq (2C_H\sqrt{q} + C_s)\sqrt{\frac{\log p}{n_t}} \|\phi_s^{(k)}\|_2 + o\left(\sqrt{\frac{\log p}{n_t}}\right).$$

For term $A_5^{(k)}$, it holds that

$$\left\| \left(\frac{1}{n_t} X_t^\top H_t - \frac{1}{n_s^{(k)}} X_s^{(k)\top} H_s^{(k)} \right) \phi_s^{(k)} \right\|_\infty \leq C\|\phi_s^{(k)}\|_2 \sqrt{\frac{q\log p}{n_t}},$$

in events $\mathcal{G}_4$,$\mathcal{G}_5^{(k)}$,$\mathcal{G}_6$,$\mathcal{G}_7^{(k)}$.

For term $A_6^{(k)}$, it holds that

$$\left\| \left(\frac{1}{n_t} X_t^\top H_t - \mathbb{E}[X^\top H] \right) \left(\phi_t - \phi_s^{(k)}\right) \right\|_\infty \leq C\|\phi_t - \phi_s^{(k)}\|_2 \sqrt{\frac{q\log p}{n_t}}.$$

in events $\mathcal{G}_4$,$\mathcal{G}_6$.

So the value $v^{(k)}$ can be upper bounded by

$$\begin{aligned}
v^{(k)} &\leq \left\| \mathbb{E}[X^\top H](\phi_t - \phi_s^{(k)}) \right\|_\infty + (C_1\sqrt{q}(\|\phi_s^{(k)}\|_2 + 1) + C_2)(1 + s_\eta)\sqrt{\frac{\log p}{n_t}} + C_3\sqrt{\frac{p\|\phi_t - \phi_s^{(k)}\|_2^2}{n_t\lambda_\Psi^2}} \\
&\leq \left(C_4 + C_3\sqrt{\frac{p}{n_t\lambda_\Psi^2}} \right) \|\phi_t - \phi_s^{(k)}\|_2 + (C_1\sqrt{q}(\|\phi_s^{(k)}\|_2 + 1) + C_2)(1 + s_\eta)\sqrt{\frac{\log p}{n_t}} \\
&\leq (C_4 + C_3C_\lambda)\|\phi_t - \phi_s^{(k)}\|_2 + (C_1\sqrt{q}(\|\phi_s^{(k)}\|_2 + 1) + C_2)(1 + s_\eta)\sqrt{\frac{\log p}{n_t}}
\end{aligned}$$

Similarly, the value $v^{(k)}$ can be lower bounded by

$$v^{(k)} \geq (C_4 - C_3C_\lambda)\|\phi_t - \phi_s^{(k)}\|_2 - (C_1\sqrt{q}(\|\phi_s^{(k)}\|_2 + 1) + C_2)(1 + s_\eta)\sqrt{\frac{\log p}{n_t}}.$$

From Assumption 3, there exist some k^* such that $\|\phi_t - \phi_s^{(k^*)}\|_2 \leq C_\phi\lambda_\Psi^2 p^{-1}\log p$. So

$$v^{(k^*)} \leq (C_4 + C_3C_\lambda)C_\phi\lambda_\Psi^2 p^{-1}\log p + (C_1\sqrt{q}(\|\phi_s^{(k)}\|_2 + 1) + C_2)(1 + s_\eta)\sqrt{\frac{\log p}{n_t}}.$$

Recall the definition of selection source group, for any $k \in \mathcal{I}(\rho)$, we have

$$v^{(k)} \leq \rho \min_{1\leq k\leq K} v^{(k)} \leq \rho v^{(k^*)}.$$

So

$$\|\phi_t - \phi_s^{(k)}\|_2 \le C\rho\lambda_\Psi^2 p^{-1}\log p + C\rho(\sqrt{q}\|\phi_t\|_2 + 1)s_\eta\sqrt{\frac{\log p}{n_t}}. \tag{S.24}$$

Now we study the properties of $\hat{\eta}^{(\mathcal{I}(\rho))}$ and $\hat{\beta}_t^{(\mathcal{I}(\rho))} = \hat{\eta}^{(\mathcal{I}(\rho))} + \hat{\beta}_s^{(\mathcal{I}(\rho))}$.

For $\hat{\eta}^{(\mathcal{I}(\rho))}$, we follow the proof of Lemma S.10. Define that $\beta_s^{(\mathcal{I}(\rho))}$ is the weighted center of source models in $\mathcal{I}(\rho)$. Because for any $k \in \mathcal{I}(\rho)$, the inequality (S.24) holds, we have

$$\begin{aligned}\|\hat{\eta}^{(\mathcal{I}(\rho))} - (\beta_t - \beta_s^{(\mathcal{I}(\rho))})\|_1 &\le C_3\sqrt{\frac{p\max_{k\in\mathcal{I}(\rho)}\|\phi_t - \phi_s^{(k)}\|_2^2}{n_t\lambda_\Psi^2}} \vee (C_1\sqrt{q}\|\phi_t\|_2 + C_2)s_\eta\sqrt{\frac{\log p}{n_t}} \\ &\le ((C_1\sqrt{q}\|\phi_t\|_2 + C_2)s_\eta + \rho)\sqrt{\frac{\log p}{n_t}},\end{aligned}$$

and

$$\|\hat{\eta}^{(\mathcal{I}(\rho))} - (\beta_t - \beta_s^{(\mathcal{I}(\rho))})\|_2 \le ((C_1\sqrt{q}\|\phi_t\|_2 + C_2)\sqrt{s_\eta} + \rho)\sqrt{\frac{\log p}{n_t}}.$$

For $\hat{\beta}_s^{(\mathcal{I}(\rho)}$, from Theorem 2, we have

$$\begin{aligned}\|\hat{\beta}_s^{(\mathcal{I}(\rho)} - \beta_s^{(\mathcal{I}(\rho))}\|_1 &\le C\sqrt{\frac{p\|\phi_t\|_2^2}{n_s^{(\mathcal{I}(\rho))}\lambda_\Psi^2}} \vee Cs\sqrt{\frac{\log p}{n_s^{(\mathcal{I}(\rho))}}}, \\ \|\hat{\beta}_s^{(\mathcal{I}(\rho)} - \beta_s^{(\mathcal{I}(\rho))}\|_2 &\le C\sqrt{\frac{p\|\phi_t\|_2^2}{n_s^{(\mathcal{I}(\rho))}\lambda_\Psi^2}} \vee C\sqrt{\frac{s\log p}{n_s^{(\mathcal{I}(\rho))}}}.\end{aligned}$$

Finally, it holds that

$$\begin{aligned}\|\hat{\beta}_t^{(\mathcal{I}(\rho))} - \beta_t\|_1 &\le C\sqrt{\frac{p\|\phi_t\|_2^2}{n_s^{(\mathcal{I}(\rho))}\lambda_\Psi^2}} \vee C(s_\eta + \rho)\sqrt{\frac{\log p}{n_t}}, \\ \|\hat{\beta}_t^{(\mathcal{I}(\rho))} - \beta_t\|_2 &\le C\sqrt{\frac{p\|\phi_t\|_2^2}{n_s^{(\mathcal{I}(\rho))}\lambda_\Psi^2}} \vee C(\sqrt{s_\eta} + \rho)\sqrt{\frac{\log p}{n_t}}.\end{aligned}$$

□

S.1.10 Lemmas for Theorem 4.

Lemma S.10 *Under the same conditions as in Theorem 4, with probability larger than* $1-p^{-c}$,

$$\left\|\hat{\eta}^{(k)}-\left(\beta_t-\beta_s^{(k)}\right)\right\|_1 \leq C s_\eta \lambda_t^{(k)}+\frac{1}{\lambda_t^{(k)}} \frac{p\|\phi_t-\phi_s^{(k)}\|_2^2}{n_t \lambda_\Psi^2},$$
$$\left\|\hat{\eta}^{(k)}-\left(\beta_t-\beta_s^{(k)}\right)\right\|_2 \leq C \sqrt{s_\eta} \lambda_t^{(k)}+\frac{1}{\lambda_t^{(k)}} \frac{p\|\phi_t-\phi_s^{(k)}\|_2^2}{n_t \lambda_\Psi^2},$$

where the tuning parameter $\lambda_t^{(k)}$ *satisfies that* $\lambda_t^{(k)} \geq (c_1\sqrt{q}\|\phi_s^{(k)}\|+c_2)\sqrt{\log p/n_t}+c_3\lambda_\Psi^{-1}\|\phi_t-\phi_s^{(k)}\|_2$ *and* C, c_1, c_2, c_3 *are some constants.*

Proof of Lemma S.10. Define that $\eta^{(k)}=\beta_t-\beta_s^{(k)}, \zeta^{(k)}=\phi_t-\phi_s^{(k)}$. Then define that $b^{(k)}=\Sigma_X^{-1}\mathbb{E}[X^\top H]\zeta^{(k)}$, $\mathcal{V}_t=V_t \operatorname{diag}(\min(\tau_t/D_t),1))V_t^\top$ and $\tilde{\mathcal{V}}_s^{(k)}=V_s^{(k)}\operatorname{diag}(\tau_t/D_{s,1:\tau}^{(k)}, 1_{p-\tau})V_s^{(k)\top}$. So $\tilde{X}_t=X_t\mathcal{V}_t$ and $\tilde{X}_t^\top\tilde{H}_t=\mathcal{V}_t^2X_t^\top H_t$. Define several events:

$$\left\{\left\|\left(\mathcal{V}_t^2\frac{1}{n_t}X_t^\top X_t-\mathcal{V}^2\Sigma_X\right)\left(\beta_s^{(k)}-\hat{\beta}_{\text{init}}^{(k)}\right)\right\|_\infty \leq C\sqrt{\frac{\log p}{n_t}}\right\}, \tag{$\mathcal{G}_{13}$}$$

$$\left\{\left\|\left(\tilde{\mathcal{V}}_s^{(k)2}\frac{1}{n_s^{(k)}}X_s^{(k)\top}X_s^{(k)}-\mathcal{V}^2\Sigma_X\right)\left(\beta_s^{(k)}-\hat{\beta}_{\text{init}}^{(k)}\right)\right\|_\infty \leq C\sqrt{\frac{\log p}{n_s^{(k)}}}\right\}, \tag{$\mathcal{G}_{13}^{(k)}$}$$

$$\left\{\left\|\left(\mathcal{V}_t^2\frac{1}{n_t}X_t^\top H_t-\mathcal{V}^2\mathbb{E}[X^\top H]\right)\phi_s^{(k)}\right\|_\infty \leq C\|\phi_s^{(k)}\|_2\sqrt{\frac{q\log p}{n_t}}\right\}, \tag{$\mathcal{G}_{14}$}$$

$$\left\{\left\|\left(\tilde{\mathcal{V}}_s^2\frac{1}{n_s^{(k)}}X_s^{(k)\top}H_s^{(k)}-\mathcal{V}^2\mathbb{E}[X^\top H]\right)\phi_s^{(k)}\right\|_\infty \leq C\|\phi_s^{(k)}\|_2\sqrt{\frac{q\log p}{n_s^{(k)}}}\right\}, \tag{$\mathcal{G}_{14}^{(k)}$}$$

$$\left\{\left\|\mathcal{V}_t^2\frac{1}{n_t}X_t^\top\varepsilon_t\right\|_\infty \leq C\sqrt{\frac{\log p}{n_t}}\right\}, \tag{$\mathcal{G}_{15}$}$$

$$\left\{\left\|\tilde{\mathcal{V}}_s^{(k)2}\frac{1}{n_s^{(k)}}X_s^{(k)\top}\varepsilon_s^{(k)}\right\|_\infty \leq C\sqrt{\frac{\log p}{n_s^{(k)}}}\right\}, \tag{$\mathcal{G}_{15}^{(k)}$}$$

$$\left\{\left\|\frac{1}{n_t}\tilde{X}_t^\top\left(\tilde{H}_t\zeta^{(k)}-\tilde{X}_tb^{(k)}\right)\right\|_\infty \leq C_t\frac{1}{\lambda_\Psi}\|\zeta^{(k)}\|_2\right\}. \tag{$\mathcal{G}_{16}$}$$

Define the event $\mathcal{G}^{(k)\prime}=\mathcal{G}_{13}\cap\mathcal{G}_{13}^{(k)}\cap\mathcal{G}_{14}\cap\mathcal{G}_{14}^{(k)}\cap\mathcal{G}_{15}^{(k)}\cap\mathcal{G}_{16}$. From Lemma S.11, the event $\mathcal{G}^{(k)\prime}$ happens with a probability

$$\mathbb{P}(\mathcal{G}^{(k)\prime}) \geq 1-p^{-c}.$$

Recall the definition of $\hat{\eta}^{(k)}$. We have the inequality from lasso,

$$\frac{1}{n_t}\left\|\tilde{Y}_t-\tilde{X}_t\hat{\beta}^{(k)}_{\text{init}}-\tilde{Z}^{(k)}_t-\tilde{X}_t\hat{\eta}^{(k)}\right\|_2^2+\lambda^{(k)}_t\|\hat{\eta}^{(k)}\|_1\le\frac{1}{n_t}\left\|\tilde{Y}_t-\tilde{X}_t\hat{\beta}^{(k)}_{\text{init}}-\tilde{Z}^{(k)}_t-\tilde{X}_t\eta\right\|_2^2+\lambda^{(k)}_t\|\eta\|_1.$$

With $\tilde{Y}_t=\tilde{X}_t\beta_t+\tilde{H}_t\phi_t+\tilde{\varepsilon}_t=\tilde{X}_t(\beta^{(k)}_s+\eta^{(k)})+\tilde{H}_t(\phi^{(k)}_s+\zeta^{(k)})+\tilde{\varepsilon}_t$, we have

$$\begin{aligned}\frac{1}{n_t}\left\|\tilde{X}_t(\hat{\eta}^{(k)}-\eta^{(k)})\right\|_2^2&\le\lambda^{(k)}_t(\|\eta^{(k)}\|_1-\|\hat{\eta}^{(k)}\|_1)-\frac{2}{n_t}(\hat{\eta}^{(k)}-\eta^{(k)})^\top\tilde{X}_t^\top\tilde{\varepsilon}_t\\&\quad-\frac{2}{n_t}(\hat{\eta}^{(k)}-\eta^{(k)})^\top\tilde{X}_t^\top\left(\tilde{X}_t\left(\beta^{(k)}_s-\hat{\beta}^{(k)}_{\text{init}}\right)+\tilde{H}_t\phi_t-\tilde{Z}^{(k)}_t\right)\\&\le\lambda^{(k)}_t(\|\eta^{(k)}\|_1-\|\hat{\eta}^{(k)}\|_1)+2\|\hat{\eta}^{(k)}-\eta^{(k)}\|_1\left\|\frac{1}{n_t}\tilde{X}_t^\top\tilde{\varepsilon}_t\right\|_\infty\\&\quad+2\|\hat{\eta}^{(k)}-\eta^{(k)}\|_1\left\|\frac{1}{n_t}\tilde{X}_t^\top\left(\tilde{X}_t\left(\beta^{(k)}_s-\hat{\beta}^{(k)}_{\text{init}}\right)+\tilde{H}_t\phi^{(k)}_s-\tilde{Z}^{(k)}_t\right)\right\|_\infty\\&\quad+2\|\hat{\eta}^{(k)}-\eta^{(k)}\|_1\left\|\frac{1}{n_t}\tilde{X}_t^\top\left(\tilde{H}_t\zeta^{(k)}-\tilde{X}_tb^{(k)}\right)\right\|_\infty\\&\quad+2\left\|\tilde{X}_t(\hat{\eta}^{(k)}-\eta^{(k)})\right\|_2\left\|\frac{1}{n_t}\tilde{X}_t\zeta^{(k)}\right\|_2.\end{aligned}$$

Consider the term

$$\begin{aligned}&\left\|\frac{1}{n_t}\tilde{X}_t^\top\left(\tilde{X}_t\left(\beta^{(k)}_s-\hat{\beta}^{(k)}_{\text{init}}\right)+\tilde{H}_t\phi^{(k)}_s-\tilde{Z}^{(k)}_t\right)\right\|_\infty\\\le&\left\|\mathcal{V}_t^2\left[\frac{1}{n_t}X_t^\top\left(X_t\left(\beta^{(k)}_s-\hat{\beta}^{(k)}_{\text{init}}\right)+H_t\phi^{(k)}_s\right)\right]-\tilde{\mathcal{V}}^{(k)2}_s\frac{1}{n^{(k)}_s}X^{(k)\top}_s\hat{Z}^{(k)}_s\right\|_\infty\\+&\left\|\mathcal{V}_t^2\frac{1}{n_t}X_t^\top\hat{Z}_t-\tilde{\mathcal{V}}^{(k)2}_s\frac{1}{n^{(k)}_s}X^{(k)\top}_s\hat{Z}^{(k)}_s\right\|_\infty.\end{aligned}\tag{S.25}$$

The first term in the right hand of (S.25) is

$$\begin{aligned}&\left\|\mathcal{V}_t^2\left[\frac{1}{n_t}X_t^\top\left(X_t\left(\beta^{(k)}_s-\hat{\beta}^{(k)}_{\text{init}}\right)+H_t\phi^{(k)}_s\right)\right]-\tilde{\mathcal{V}}^{(k)2}_s\frac{1}{n^{(k)}_s}X^{(k)\top}_s\hat{Z}^{(k)}_s\right\|_\infty\\=&\left\|\left(\mathcal{V}_t^2\frac{1}{n_t}X_t^\top X_t-\tilde{\mathcal{V}}^{(k)2}_s\frac{1}{n^{(k)}_s}X^{(k)\top}_sX^{(k)}_s\right)\left(\beta^{(k)}_s-\hat{\beta}^{(k)}_{\text{init}}\right)\right\|_\infty\\+&\left\|\left(\mathcal{V}_t^2\frac{1}{n_t}X_t^\top H_t-\tilde{\mathcal{V}}^{(k)2}_s\frac{1}{n^{(k)}_s}X^{(k)\top}_sH^{(k)}_s\right)\phi^{(k)}_s\right\|_\infty+\left\|\tilde{\mathcal{V}}^{(k)2}_s\frac{1}{n^{(k)}_s}X^{(k)\top}_s\varepsilon^{(k)}_s\right\|_\infty,\end{aligned}$$

in which the first term is bounded by $\mathcal{G}_{13}$ and $\mathcal{G}^{(k)}_{13}$, the second term is bounded by $\mathcal{G}_{14}$ and

$\mathcal{G}_{14}^{(k)}$, the third term is bounded by $\mathcal{G}_{15}^{(k)}$. The second term in the right hand of (S.25) is

$$\begin{aligned}\left\|\mathcal{V}_t^2\frac{1}{n_t}X_t^\top\hat{Z}_t-\tilde{\mathcal{V}}_s^{(k)2}\frac{1}{n_s^{(k)}}X_s^{(k)\top}\hat{Z}_s^{(k)}\right\|_\infty &\leq \left\|\left(\mathcal{V}_t^2\frac{1}{n_t}X_t^\top X_t-\tilde{\mathcal{V}}_s^{(k)2}\frac{1}{n_s^{(k)}}X_s^{(k)\top}X_s^{(k)}\right)\left(\beta_s^{(k)}-\hat{\beta}_{\text{init}}^{(k)}\right)\right\|_\infty\\ &+\left\|\left(\mathcal{V}_t^2\frac{1}{n_t}X_t^\top H_t-\mathcal{V}_t^2\frac{1}{n_s^{(k)}}X_s^{(k)\top}H_s^{(k)}\right)\phi_s^{(k)}\right\|_\infty\\ &+\left\|\tilde{\mathcal{V}}_s^{(k)2}\frac{1}{n_s^{(k)}}X_s^{(k)\top}\varepsilon_s^{(k)}\right\|_\infty,\end{aligned}$$

which is similarly bounded in the defined events. So,

$$\left\|\frac{1}{n_t}\tilde{X}_t^\top\left(\tilde{X}_t\left(\beta_s^{(k)}-\hat{\beta}_{\text{init}}^{(k)}\right)+\tilde{H}_t\phi_s^{(k)}-\tilde{Z}_t^{(k)}\right)\right\|_\infty\leq(C_1\sqrt{q}\|\phi_s^{(k)}\|_2+C_2)\sqrt{\frac{\log p}{n_t}}.$$

In event $\mathcal{G}^{(k)\prime}$, it holds that

$$\begin{aligned}\frac{1}{n_t}\left\|\tilde{X}_t(\hat{\eta}^{(k)}-\eta^{(k)})\right\|_2^2 &\leq \lambda_t^{(k)}(\|\eta^{(k)}\|_1-\|\hat{\eta}^{(k)}\|_1)\\ &+2\left((C_1\sqrt{q}\|\phi_s^{(k)}\|_2+C_2)\sqrt{\frac{\log p}{n_t}}+C_t\frac{1}{\lambda_\psi}\|\zeta^{(k)}\|_2\right)\|\hat{\eta}^{(k)}-\eta^{(k)}\|_1\\ &+2\left\|\tilde{X}_t(\hat{\eta}^{(k)}-\eta^{(k)})\right\|_2\left\|\frac{1}{n_t}\tilde{X}_t\zeta^{(k)}\right\|_2.\end{aligned}$$

When $\lambda_t^{(k)}\geq 8\left((C_1\sqrt{q}\|\phi_s^{(k)}\|_2+C_2)\sqrt{\log p/n_t}+C_t\lambda_\Psi^{-1}\|\zeta^{(k)}\|_2\right)$, it holds that

$$\frac{1}{n_t}\left\|\tilde{X}_t(\hat{\eta}^{(k)}-\eta^{(k)})\right\|_2^2\leq\lambda_t^{(k)}\|\eta^{(k)}\|_1-\lambda_t^{(k)}\|\hat{\eta}^{(k)}\|_1+\frac{\lambda_t^{(k)}}{4}\|\hat{\eta}^{(k)}-\eta^{(k)}\|_1+\left\|\frac{2}{n_t}\tilde{X}_t\zeta^{(k)}\right\|_2\left\|\tilde{X}_t(\hat{\eta}^{(k)}-\eta^{(k)})\right\|_2. \tag{S.26}$$

The inequality (S.26) is similar to (S.17). Follow the proof of Theorem 2, with Assumption 3.a., it holds that

$$\begin{aligned}\left\|\hat{\eta}^{(k)}-\eta^{(k)}\right\|_1 &\leq C_3 s_\eta\lambda_t^{(k)}+\frac{1}{\lambda_t^{(k)}}\frac{p\|\zeta^{(k)}\|_2^2}{n_t\lambda_\Psi^2},\\ \left\|\hat{\eta}^{(k)}-\eta^{(k)}\right\|_2 &\leq C_3\sqrt{s_\eta}\lambda_t^{(k)}+\frac{1}{\lambda_t^{(k)}}\frac{p\|\zeta^{(k)}\|_2^2}{n_t\lambda_\Psi^2}.\end{aligned}$$

$\square$

Lemma S.11 *Under Assumption 1, 2 and 3, the event* $\mathcal{G}^{(k)\prime} = \mathcal{G}_{13} \cap \mathcal{G}_{13}^{(k)} \cap \mathcal{G}_{14} \cap \mathcal{G}_{14}^{(k)} \cap \mathcal{G}_{15}^{(k)} \cap \mathcal{G}_{16}$ *happens with a probability*

$$\mathbb{P}(\mathcal{G}^{(k)\prime}) \geq 1 - p^{-c},$$

for a constant c.

Proof of Lemma S.11. The probabilities of events $\mathcal{G}_{13}$ and $\mathcal{G}_{13}^{(k)}$ are bounded by (S.21). The probabilities of events $\mathcal{G}_{14}$ and $\mathcal{G}_{14}^{(k)}$ are similarly bounded by the proof of (S.21) in Lemma S.8. The probabilities of events $\mathcal{G}_{15}$ and $\mathcal{G}_{15}^{(k)}$ are similarly bounded like Lemma S.7. The probability of event $\mathcal{G}_{16}$ is bounded similarly to Lemma S.9. □

S.2 Theoretical results for the fixed-dimensional scenario

S.2.1 Model shift estimation

We first give some general assumptions under the fixed-dimensional scenario.

Assumption S.1 *a. The errors* $E_{ti}, E_{si}^{(k)}$*, the confounders* $H_{ti}, H_{si}^{(k)}$ *and* $f(H_{ti}), f(H_{si})^{(k)}$ *are i.i.d. sub-Gaussian random variables.*

b. The noise terms $\varepsilon_{ti}, \varepsilon_{si}^{(k)}$ *are i.i.d. sub-Gaussian random variables with zero mean and variance* σ^2*.*

c. The smallest eigenvalue of $\mathrm{Cov}(X_{ti})$ *is larger than a positive constant* $C_r > 0$*.*

d. The confounders exert the similar effects on the target and source models, i.e.,

$$\left\| \frac{1}{n_s} \sum_{k=1}^{K} n_s^{(k)} \phi_s^{(k)} - \phi_t \right\|_2 \leq C\sqrt{\frac{1}{n}},$$

where C is a constant.

The Assumption S.1.a. to c. are general assumptions. The Assumption S.1.e. requires that the average confounding effect in the source models is similar to that in the target model, which is similar to Assumption 1.d. under the high-dimensional scenario.

Theorem S.1 *Under Assumption S.1, suppose the initial source estimator $\hat{\beta}_{init}$ for β_s satisfies $\|\hat{\beta}_{init} - \beta_s\|_1 \leq C_s$ for a constant C_s. Then, with probability at least $1 - p^{-c}$ for a constant c, the estimator $\hat{\eta}$ defined in (5) satisfies:*

$$\|\hat{\eta} - \eta_0\|_1 \lesssim s_\eta \lambda_t, \quad \|\hat{\eta} - \eta_0\|_2 \lesssim \sqrt{s_\eta}\lambda_t, \quad \frac{1}{\sqrt{n_t}}\|X_t(\hat{\eta} - \eta_0)\|_2 \lesssim \sqrt{s_\eta}\lambda_t.$$

Here the tuning parameter λ_t satisfies $\lambda_t \geq (C_1'\sqrt{q}\|\phi_t\|_2 + C_2)n_t^{-1/2}$ for some constants C_1', C_2'.

Proof of Theorem S.1. Note that now p is fixed which is not increasing with n. The logarithm term $\log p$ is now a constant and we still have the bound (S.11) that

$$\frac{1}{n_t}\|X_t(\hat{\eta} - \eta_0)\|_2^2 \leq \lambda_t\|\eta_0\|_1 - \lambda_t\|\hat{\eta}\|_1 + \frac{\lambda_t}{2}\|\hat{\eta} - \eta_0\|_1,$$

where we set $\lambda_t \geq C_1'\sqrt{q}\|\phi_t\|_2 + C_2')n_t^{-1/2}$ for some constants C_1', C_2'. With the lower bound C_r of the smallest eigenvalue of $\mathrm{Cov}(X_{ti})$, we know that

$$\begin{aligned}\frac{1}{n_t}\|X_t(\hat{\eta} - \eta_0)\|_2^2 &\geq C_r\|\hat{\eta} - \eta_0\|_2^2 - \left|(\hat{\eta} - \eta_0)'\left(\Sigma_X - \frac{1}{n_t}X_t'X_t\right)(\hat{\eta} - \eta_0)\right| \\ &\gtrsim C_r\left(1 - \mathcal{O}_p\left(\sqrt{\frac{p}{n_t}}\right)\right)\|\hat{\eta} - \eta_0\|_2^2 \gtrsim \|\hat{\eta} - \eta_0\|_2^2.\end{aligned}$$

The remain proof is similar to that of Theorem 1, noting that now we can treat $p, \log p$ as constants. □

Remark S.1 *Under the fixed-dimensional setting, the profiled residual transfer (7) can also be replaced by a similar procedure*

$$\hat{Z}_t^{fix} = \arg\min_{Z_t \in \mathbb{R}^{n_t}}\left\{\left\|\frac{1}{n_t}X_t^\top Z_t - \frac{1}{n_s}\sum_{k=1}^{K}X_s^{(k)\top}\hat{Z}_s^{(k)}\right\|_2^2 + \lambda_z\|Z_t\|_2^2\right\} \tag{S.27}$$

which employs the l_2 norm now for some $\lambda_z > 0$. One advantage of $\hat{Z}_t^{fix}$ is that it has a closed-form expression. A result analogous to Theorem S.1 can be derived when $\hat{Z}_t^{fix}$ is used as the estimator of the profiled target residual.

S.2.2 Target parameter estimation

Similar to the high-dimensional case, we estimate β_t as $\hat{\beta}_t = \hat{\eta} + \hat{\beta}_s$, where $\hat{\beta}_s$ is an estimator of the source parameter β_s. For simplicity, we focus on the single source scenario ($K = 1$) here and adopt the estimation method proposed by Tang et al. (2023), which involves a two-step regression procedure. In the first step, the estimator $\hat{\Psi}$ of Ψ is obtained using maximum likelihood estimation or principal component analysis, and a semi-orthogonal matrix $B_{\hat{\Psi}^\perp}$ is constructed, whose columns are orthogonal to the columns of $\hat{\Psi}$. Then the synthetic IV $X_s B_{\hat{\Psi}^\perp}$ is regressed out to obtain $\hat{X}_s$, and β_s is then estimated by

$$\hat{\beta}_s = \arg\min_{\beta \in \mathbb{R}^p} \frac{1}{n_s} \left\| Y_s - \hat{X}_s \beta \right\|_2^2, \quad \text{subject to } \|\beta\|_0 \leq s,$$

where s denotes a tuning parameter controlling sparsity.

Assumption S.2 *a. The function $f(H) = H\Psi$ for some fixed $\Psi \in \mathbb{R}^{q \times p}$. The errors E_{ti}, E_{si} and the confounders H_{ti}, H_{si} are mean zero Gaussian random variables with $\mathrm{Cov}(H_{\cdot i}) = I_q$ and $\mathrm{Cov}(E_{\cdot i}) = D$.*

b. The true parameter β_s has sparsity $s < p - q$.

c. There exist an estimator $\hat{\Psi}$ and an orthogonal matrix $O \in \mathbb{R}^{q \times q}$ such that $\|\hat{\Psi} - \Psi O\|_2 \leq a_{n_s}$ for some $a_{n_s} \to 0$ as $n_s \to \infty$.

Theorem S.2 *Under Assumptions S.1 and S.2, with probability at least $1 - p^{-c}$ for a constant c, the estimator $\hat{\beta}_t$ satisfies*

$$\|\hat{\beta}_t - \beta_t\|_1 \lesssim n_t^{-1/2} \vee a_{n_s} \vee n_s^{-1/2}, \quad \|\hat{\beta}_t - \beta_t\|_2 \lesssim n_t^{-1/2} \vee a_{n_s} \vee n_s^{-1/2},$$

when we set λ_t in the model shift estimation as $\lambda_t \asymp n_t^{-1/2}$.

Proof of Theorem S.2. Theorem 2 in Tang et al. (2023) shows that

$$\|\hat{\beta}_s - \beta_s\|_1 = \mathcal{O}_p(n_s^{-1/2})$$

when $a_{n_s} \leq C n_s^{-1/2}$ for some constant C. When $n_s^{1/2} a_{n_s} \to \infty$, we can use only $a_{n_s}^{-2}$ data in the source dataset. Still from Theorem 2 in Tang et al. (2023), we have

$$\|\hat{\beta}_s - \beta_s\|_1 = \mathcal{O}_p(a_{n_s}).$$

Combine this with Theorem S.1 and set $\lambda_t \asymp n_t^{-1/2}$, we have

$$\|\hat{\beta}_t - \beta_t\|_1 \lesssim n_t^{-1/2} \vee a_{n_s} \vee n_s^{-1/2}.$$

Due to the fixed dimensional p, this also holds for l_2 norm. □

Remark S.2 *When directly using the target data or applying classical transfer learning methods, the available upper bound is $n_t^{-1/2} \vee a_{n_t} \vee n_s^{-1/2}$. Compared with our approach, when the convergence of $\hat{\Psi}$ is slow, these methods are constrained by the slower rate a_{n_t}. In contrast, our method can still achieve the optimal rate of $n_t^{-1/2}$ once n_s is sufficiently large. This finding is consistent with results under the high-dimensional setting, further demonstrating the generality of the proposed profiled transfer learning framework.*

S.3 Implementation Details

In this section, we describe the estimation and implementation procedures. For the source estimator $\hat{\beta}_s$, we adopt the standard lasso with cross-fitting as implemented in the R package *glmnet* in the high-dimensional setting. In the fixed-p regime, we follow the approach proposed in Tang et al. (2023). The number of confounders is estimated via maximum likelihood under a Gaussian assumption, and the estimator $\hat{\beta}_s$ is obtained through a model selection procedure with l_0 constraints.

For the model shift estimator $\hat{\eta}$, we again apply the standard lasso to the pseudo-response $Y_t - X_t\hat{\beta}_{\text{init}} - \hat{Z}_t$. To implement the profiled residual transfer, the optimization problem

$$\hat{Z}_t = \operatorname*{argmin}_{Z_t \in \mathbb{R}^{n_t}} \left\{ \left\| \frac{1}{n_t} X_t^\top Z_t - \frac{1}{n_s} \sum_{k=1}^{K} X_s^{(k)\top} \hat{Z}_s^{(k)} \right\|_\infty \right\}$$

can be reformulated as a linear program:

$$\hat{z} = \operatorname*{argmin}_{0 \preceq z \in \mathbb{R}^{2n_t+1}} \alpha^\top z, \quad \text{s.t.}\, Az \leq b,$$

where

$$\alpha = \begin{pmatrix} 0_{2n_t} \\ 1 \end{pmatrix}, \quad A = \begin{pmatrix} \frac{1}{n_t} X_t^\top & -\frac{1}{n_t} X_t^\top & -1 \\ -\frac{1}{n_t} X_t^\top & \frac{1}{n_t} X_t^\top & -1 \end{pmatrix}, \quad b = \begin{pmatrix} \frac{1}{n_s} \sum_{k=1}^{K} X_s^{(k)\top} \hat{Z}_s^{(k)} \\ -\frac{1}{n_s} \sum_{k=1}^{K} X_s^{(k)\top} \hat{Z}_s^{(k)} \end{pmatrix}.$$

This problem can be efficiently solved using interior-point algorithms, yielding

$$\hat{Z}_t = \hat{z}_{[1:n_t]} - \hat{z}_{[n_t+1:2n_t]}.$$

S.4 Additional Simulation

S.4.1 Fixed-dimensional scenario

In parallel with the high-dimensional setting discussed in the main text, we conduct additional simulations for a fixed-dimensional scenario. We consider the linear confounding structure $X_l = H_l^\top \Psi + E_l$ for $l = s, t$ and generate the source and target data (X_{li}, Y_{li}) using the same mechanism as in the main text. The covariate dimension is fixed at $p = 10$, which is much smaller than the sample size.

In the single-source case, the true source coefficient is defined as $\beta_s = 2\sum_{i=1}^{3} e_i$ where $e_i \in \mathbb{R}^p$ is the vector with the i-th entry equals to one and all other entries equal to zero. The target coefficient is given by $\beta_t = \beta_s + \eta_0$ where the different $\eta_0 = -4e_1 + 2e_p$. In the multi-source case, the number of sources is set to $K = 3$ with a sample size $n_s^{(k)} = 200$ for $1 \leq k \leq 3$. The coefficient for the k-th source is defined as $\beta_s^{(k)} = 2\sum_{i=1}^{3} e_i + e_k$ and other settings are the identical with those in the single-source case.

For the fixed-dimensional setting, we estimate the source models using synthetic two-stage regularized regression (Tang et al. 2023), and apply the profiled residual transfer in equation (S.27), which is computationally more efficient in low-dimensional problems. We compare our method with two baselines. Baseline 1 directly estimates both the source and target models using the approach of Tang et al. (2023). Baseline 2 first estimates the source coefficient $\hat{\beta}_s$ using the same approach and then applies Tang et al. (2023) to the transferred target data $(X_t, Y_t - X_t\hat{\beta}_s)$.

Tables S.1 and S.2 show the performance of our method and the baseline approaches, demonstrating that our method also outperforms the alternatives in the fixed-dimensional setting.

Table S.1: The l_2 norm errors of different methods under the fixed-dimensional scenario.

p	Method	single-source		multi-source	
		η	β_t	η	β_t
10	Baseline 1	0.7757 (0.2517)	0.7841 (0.2430)	0.7741 (0.2361)	0.7786 (0.2309)
	Baseline 2	0.3606 (0.0025)	0.3694 (0.0028)	0.3571 (0.0025)	0.3706 (0.0028)
	ProTrans	**0.2794** (0.0019)	**0.3471** (0.0035)	**0.2767** (0.0018)	**0.3307** (0.0026)

S.4.2 Sensitivity analysis of the initial estimator

We show that our method is robust to the choice of the initial estimator $\hat{\beta}_{\text{init}}$, as the condition in Theorem 1 shown. Besides from the used lasso initial estimator in our simulations,

Table S.2: The l_1 norm errors of different methods under the fixed-dimensional scenario.

p	Method	single-source		multi-source	
		η	β_t	η	β_t
10	Baseline 1	1.5060 (0.2517)	1.5263 (0.2430)	1.4803 (1.1537)	1.4909 (1.1284)
	Baseline 2	0.6852 (0.0025)	0.7708 (0.0028)	0.6759 (0.0125)	0.7706 (0.0155)
	ProTrans	**0.4409** (0.0019)	**0.6477** (0.0035)	**0.4271** (0.0060)	**0.6036** (0.0154)

we consider some other initial estimators which are the perturbations of the lasso initial estimator. We consider two different perturbation forms, the first has the same sparsity pattern but perturbed non-zero values; the second has some new non-zero elements compared to the lasso initial estimator. In detail, in the first one, we add independent normal distribution perturbations with mean zero and standard deviation 0.5 in each non-zero position. In the second one, we randomly select 50 new position and add this perturbation to those selected position. Results are reported in the Tables S.3,S.4,S.5,S.6 as the average of 100 times repeat with other settings the same as the simulation in our main text. The results show that the performance of our method is indeed not sensitive to the selection of initial estimator.

S.4.3 Treatment effect of education on earnings (High-dimensional scenario)

In this section, we continue to analyze the same dataset examining the effect of education on earnings. Unlike the main text, where an initial screening step is used to select several important covariates, here we retain 151 covariates including race, marital status, willingness to work and job satisfaction and apply ProTrans in the high-dimensional setting. All other preprocessing procedures and experimental settings remain the same as in the main text. As in Wang & Tchetgen Tchetgen (2018), we treat age, IQ scores, and residence in

Table S.3: The l_2 norm errors of ProTrans under different initial estimators when varying dimension p.

p	Method	single-source η	single-source β_t	multi-source η	multi-source β_t
Scenario 1 (Linear)					
500	Lasso initial	0.2460 (0.0165)	0.2502 (0.0165)	0.1965 (0.0121)	0.2167 (0.0107)
	Perturbation 1	0.2353 (0.0161)	0.2441 (0.0158)	0.1952 (0.0128)	0.2175 (0.0110)
	Perturbation 2	0.2622 (0.0171)	0.2644 (0.0173)	0.2037 (0.0121)	0.2232 (0.0107)
1000	Lasso initial	0.2480 (0.0148)	0.2422 (0.0129)	0.2143 (0.0109)	0.2160 (0.0088)
	Perturbation 1	0.2361 (0.0136)	0.2350 (0.0123)	0.2076 (0.0102)	0.2119 (0.0081)
	Perturbation 2	0.2543 (0.0133)	0.2478 (0.0119)	0.2328 (0.0125)	0.2314 (0.0100)
1500	Lasso initial	0.2477 (0.0156)	0.2364 (0.0137)	0.1989 (0.0103)	0.1990 (0.0090)
	Perturbation 1	0.2463 (0.0165)	0.2376 (0.0146)	0.2016 (0.0105)	0.2019 (0.0092)
	Perturbation 2	0.2716 (0.0169)	0.2570 (0.0149)	0.2287 (0.0120)	0.2212 (0.0103)
2000	Lasso initial	0.2894 (0.0173)	0.2706 (0.0150)	0.2213 (0.0106)	0.2185 (0.0092)
	Perturbation 1	0.2778 (0.0172)	0.2617 (0.0150)	0.2153 (0.0095)	0.2140 (0.0081)
	Perturbation 2	0.3064 (0.0179)	0.2843 (0.0155)	0.2497 (0.0106)	0.2404 (0.0092)
Scenario 2 (Nonlinear)					
500	Lasso initial	0.6301 (0.0599)	0.7037 (0.0657)	0.4209 (0.0449)	0.5159 (0.0455)
	Perturbation 1	0.7418 (0.0722)	0.7969 (0.0751)	0.4504 (0.0379)	0.5381 (0.0401)
	Perturbation 2	0.7045 (0.0668)	0.7651 (0.0706)	0.4365 (0.0354)	0.5267 (0.0376)
1000	Lasso initial	0.6330 (0.0438)	0.6231 (0.0400)	0.5221 (0.0449)	0.5329 (0.0455)
	Perturbation 1	0.7689 (0.0640)	0.7386 (0.0570)	0.5624 (0.0457)	0.5637 (0.0386)
	Perturbation 2	0.7062 (0.0477)	0.6854 (0.0435)	0.5636 (0.0447)	0.5651 (0.0372)
1500	Lasso initial	0.6042 (0.0368)	0.5776 (0.0351)	0.5348 (0.0449)	0.5207 (0.0455)
	Perturbation 1	0.6675 (0.0406)	0.6267 (0.0376)	0.5936 (0.0399)	0.5683 (0.0359)
	Perturbation 2	0.6433 (0.0376)	0.6103 (0.0360)	0.5888 (0.0388)	0.5635 (0.0341)
2000	Lasso initial	0.6586 (0.0449)	0.6497 (0.0455)	0.5187 (0.1853)	0.5206 (0.1732)
	Perturbation 1	0.7067 (0.0500)	0.6881 (0.0499)	0.5784 (0.0369)	0.5683 (0.0343)
	Perturbation 2	0.7296 (0.0467)	0.7085 (0.0466)	0.6218 (0.0359)	0.6027 (0.0329)
Scenario 3 (Nonlinear)					
500	Lasso initial	1.0273 (0.0972)	1.1476 (0.1070)	0.6867 (0.0652)	0.8084 (0.0678)
	Perturbation 1	1.0565 (0.1060)	1.1794 (0.1124)	0.7099 (0.0656)	0.8169 (0.0676)
	Perturbation 2	1.0608 (0.1069)	1.1817 (0.1144)	0.6910 (0.0608)	0.8085 (0.0648)
1000	Lasso initial	0.8309 (0.0761)	0.8661 (0.0705)	0.6457 (0.0518)	0.6869 (0.0463)
	Perturbation 1	0.8712 (0.0783)	0.8981 (0.0715)	0.6570 (0.0520)	0.6884 (0.0460)
	Perturbation 2	0.8490 (0.0718)	0.8817 (0.0663)	0.6547 (0.0507)	0.6878 (0.0438)
1500	Lasso initial	0.7191 (0.0497)	0.7511 (0.0523)	0.6389 (0.0440)	0.6539 (0.0430)
	Perturbation 1	0.7548 (0.0494)	0.7787 (0.0517)	0.6660 (0.0444)	0.6742 (0.0432)
	Perturbation 2	0.7592 (0.0494)	0.7844 (0.0518)	0.6699 (0.0464)	0.6792 (0.0455)
2000	Lasso initial	0.8115 (0.0597)	0.8480 (0.0625)	0.6699 (0.0503)	0.6975 (0.0503)
	Perturbation 1	0.8320 (0.0621)	0.8636 (0.0618)	0.7126 (0.0538)	0.7326 (0.0529)
	Perturbation 2	0.8769 (0.0665)	0.9014 (0.0672)	0.7178 (0.0517)	0.7326 (0.0515)

Table S.4: The l_1 norm errors of ProTrans under different initial estimators when varying dimension p.

p	Method	single-source η	single-source β_t	multi-source η	multi-source β_t
Scenario 1 (Linear)					
500	Lasso initial	0.7323 (0.0302)	1.1414 (0.0396)	0.6824 (0.0244)	1.1273 (0.0318)
	Perturbation 1	0.7292 (0.0291)	1.1401 (0.0394)	0.6887 (0.0231)	1.1392 (0.0309)
	Perturbation 2	0.7788 (0.0305)	1.1830 (0.0399)	0.7063 (0.0229)	1.1501 (0.0301)
1000	Lasso initial	0.7811 (0.0230)	1.1831 (0.0322)	0.7450 (0.0297)	1.1530 (0.0320)
	Perturbation 1	0.7574 (0.0246)	1.1612 (0.0340)	0.7208 (0.0305)	1.1296 (0.0332)
	Perturbation 2	0.7880 (0.0285)	1.1921 (0.0372)	0.7621 (0.0271)	1.1686 (0.0298)
1500	Lasso initial	0.8181 (0.0366)	1.1786 (0.0400)	0.7390 (0.0273)	1.1138 (0.0306)
	Perturbation 1	0.8518 (0.0415)	1.2104 (0.0454)	0.7214 (0.0267)	1.0954 (0.0304)
	Perturbation 2	0.8659 (0.0333)	1.2228 (0.0362)	0.7611 (0.0251)	1.1349 (0.0283)
2000	Lasso initial	0.8555 (0.0285)	1.2593 (0.0337)	0.7716 (0.0183)	1.1731 (0.0235)
	Perturbation 1	0.8101 (0.0253)	1.2092 (0.0302)	0.7261 (0.0168)	1.1294 (0.0216)
	Perturbation 2	0.8675 (0.0246)	1.2678 (0.0299)	0.8024 (0.0227)	1.2039 (0.0265)
Scenario 2 (Nonlinear)					
500	Lasso initial	1.2316 (0.0764)	2.0804 (0.1216)	1.0156 (0.0416)	1.7158 (0.0573)
	Perturbation 1	1.2658 (0.0738)	2.1177 (0.1212)	1.0777 (0.0472)	1.7771 (0.0635)
	Perturbation 2	1.2654 (0.0734)	2.1172 (0.1189)	1.0371 (0.0405)	1.7338 (0.0552)
1000	Lasso initial	1.3256 (0.0549)	2.0054 (0.0810)	1.1863 (0.0424)	1.7965 (0.0502)
	Perturbation 1	1.3713 (0.0488)	2.0582 (0.0785)	1.2153 (0.0443)	1.8251 (0.0526)
	Perturbation 2	1.3655 (0.0494)	2.0482 (0.0824)	1.2516 (0.0492)	1.8600 (0.0564)
1500	Lasso initial	1.3296 (0.0616)	1.9504 (0.0813)	1.2620 (0.0567)	1.8184 (0.0653)
	Perturbation 1	1.4054 (0.0555)	2.0151 (0.0769)	1.2734 (0.0471)	1.8342 (0.0557)
	Perturbation 2	1.4143 (0.0652)	2.0256 (0.0837)	1.3371 (0.0551)	1.8920 (0.0623)
2000	Lasso initial	1.3190 (0.0449)	2.2082 (0.1429)	1.2690 (0.0394)	1.8669 (0.0512)
	Perturbation 1	1.3455 (0.0416)	2.2355 (0.1418)	1.3337 (0.0485)	1.9354 (0.0596)
	Perturbation 2	1.4505 (0.0543)	2.3338 (0.1450)	1.3725 (0.0430)	1.9744 (0.0523)
Scenario 3 (Nonlinear)					
500	Lasso initial	1.5494 (0.1107)	2.6031 (0.1526)	1.2525 (0.0601)	2.0885 (0.0828)
	Perturbation 1	1.4941 (0.0806)	2.5533 (0.1315)	1.2479 (0.0622)	2.0754 (0.0836)
	Perturbation 2	1.5111 (0.0912)	2.5750 (0.1424)	1.2406 (0.0552)	2.0697 (0.0780)
1000	Lasso initial	1.3993 (0.0621)	2.2363 (0.0907)	1.2847 (0.0638)	2.0212 (0.0773)
	Perturbation 1	1.3747 (0.0557)	2.2033 (0.0853)	1.2093 (0.0466)	1.9417 (0.0635)
	Perturbation 2	1.4546 (0.0729)	2.2881 (0.1001)	1.2559 (0.0481)	1.9860 (0.0616)
1500	Lasso initial	1.4337 (0.0781)	2.2278 (0.1074)	1.2024 (0.0520)	1.8801 (0.0694)
	Perturbation 1	1.4215 (0.0726)	2.2113 (0.1028)	1.2404 (0.0558)	1.9183 (0.0721)
	Perturbation 2	1.4849 (0.0897)	2.2750 (0.1153)	1.2856 (0.0578)	1.9601 (0.0750)
2000	Lasso initial	1.3983 (0.0547)	2.3879 (0.1489)	1.3113 (0.0521)	2.0372 (0.0683)
	Perturbation 1	1.4380 (0.0608)	2.4190 (0.1539)	1.3124 (0.0487)	2.0417 (0.0659)
	Perturbation 2	1.4785 (0.0618)	2.4651 (0.1519)	1.3576 (0.0586)	2.0806 (0.0717)

Table S.5: The l_2 norm errors of ProTrans under different initial estimators across various levels of confounding.

$\|\phi_t\|_2$	**Method**	**single-source**		**multi-source**	
		η	β_t	η	β_t
		Scenario 1 (Linear)			
1	Lasso initial	0.2477 (0.0156)	0.2364 (0.0137)	0.1989 (0.0103)	0.1990 (0.0090)
	Perturbation 1	0.2463 (0.0165)	0.2376 (0.0146)	0.2016 (0.0105)	0.2019 (0.0092)
	Perturbation 2	0.2716 (0.0169)	0.2570 (0.0149)	0.2287 (0.0120)	0.2212 (0.0103)
2	Lasso initial	0.5908 (0.0440)	0.5682 (0.0399)	0.3678 (0.0231)	0.3612 (0.0219)
	Perturbation 1	0.5956 (0.0477)	0.5740 (0.0435)	0.3754 (0.0235)	0.3689 (0.0223)
	Perturbation 2	0.5990 (0.0471)	0.5747 (0.0431)	0.3911 (0.0239)	0.3799 (0.0222)
3	Lasso initial	1.1383 (0.0919)	1.1047 (0.0846)	0.6171 (0.0426)	0.6081 (0.0409)
	Perturbation 1	1.1625 (0.1019)	1.1302 (0.0940)	0.6164 (0.0417)	0.6066 (0.0403)
	Perturbation 2	1.1503 (0.0976)	1.1140 (0.0897)	0.6404 (0.0444)	0.6258 (0.0424)
4	Lasso initial	1.9081 (0.1599)	1.8653 (0.1472)	0.9478 (0.0732)	0.9452 (0.0711)
	Perturbation 1	1.9230 (0.1634)	1.8834 (0.1518)	0.9974 (0.0747)	0.9786 (0.0718)
	Perturbation 2	1.9385 (0.1724)	1.8910 (0.1587)	0.9912 (0.0745)	0.9798 (0.0724)
		Scenario 2 (Nonlinear)			
1	Lasso initial	0.1251 (0.0061)	0.1197 (0.0050)	0.1495 (0.0215)	0.1621 (0.0239)
	Perturbation 1	0.1340 (0.0067)	0.1271 (0.0056)	0.1529 (0.0073)	0.1647 (0.0057)
	Perturbation 2	0.1669 (0.0077)	0.1542 (0.0064)	0.1926 (0.0105)	0.1942 (0.0081)
2	Lasso initial	0.1522 (0.0075)	0.1446 (0.0064)	0.1802 (0.0168)	0.1861 (0.0207)
	Perturbation 1	0.1574 (0.0078)	0.1490 (0.0066)	0.1868 (0.0089)	0.1914 (0.0073)
	Perturbation 2	0.1935 (0.0089)	0.1783 (0.0076)	0.2238 (0.0114)	0.2188 (0.0089)
3	Lasso initial	0.1855 (0.0094)	0.1771 (0.0082)	0.2179 (0.0135)	0.2194 (0.0179)
	Perturbation 1	0.1960 (0.0100)	0.1861 (0.0088)	0.2213 (0.0109)	0.2214 (0.0091)
	Perturbation 2	0.2280 (0.0104)	0.2111 (0.0090)	0.2611 (0.0134)	0.2514 (0.0105)
4	Lasso initial	0.2315 (0.0118)	0.2219 (0.0106)	0.2618 (0.0076)	0.2596 (0.0131)
	Perturbation 1	0.2409 (0.0126)	0.2299 (0.0115)	0.2753 (0.0147)	0.2693 (0.0127)
	Perturbation 2	0.2776 (0.0136)	0.2588 (0.0120)	0.3121 (0.0169)	0.2972 (0.0136)
		Scenario 3 (Nonlinear)			
1	Lasso initial	0.1283 (0.0060)	0.1223 (0.0050)	0.1344 (0.0068)	0.1447 (0.0055)
	Perturbation 1	0.1300 (0.0061)	0.1242 (0.0050)	0.1342 (0.0062)	0.1446 (0.0052)
	Perturbation 2	0.1530 (0.0072)	0.1425 (0.0062)	0.1926 (0.0105)	0.1942 (0.0081)
2	Lasso initial	0.1610 (0.0078)	0.1548 (0.0068)	0.1632 (0.0077)	0.1716 (0.0066)
	Perturbation 1	0.1620 (0.0085)	0.1562 (0.0074)	0.1654 (0.0075)	0.1730 (0.0064)
	Perturbation 2	0.1880 (0.0097)	0.1767 (0.0085)	0.2238 (0.0114)	0.2188 (0.0089)
3	Lasso initial	0.2146 (0.0114)	0.2099 (0.0106)	0.2072 (0.0102)	0.2134 (0.0092)
	Perturbation 1	0.2093 (0.0115)	0.2051 (0.0104)	0.2079 (0.0101)	0.2137 (0.0093)
	Perturbation 2	0.2410 (0.0132)	0.2302 (0.0119)	0.2611 (0.0134)	0.2514 (0.0105)
4	Lasso initial	0.2767 (0.0159)	0.2759 (0.0153)	0.2646 (0.0138)	0.2711 (0.0132)
	Perturbation 1	0.2771 (0.0167)	0.2769 (0.0158)	0.2681 (0.0141)	0.2738 (0.0134)
	Perturbation 2	0.3015 (0.0169)	0.2954 (0.0160)	0.3121 (0.0169)	0.2972 (0.0136)

Table S.6: The l_1 norm errors of ProTrans under different initial estimators across various levels of confounding.

$\|\phi_t\|_2$	**Method**	**single-source**		**multi-source**	
		η	β_t	η	β_t
		Scenario 1 (Linear)			
1	Lasso initial	0.8181 (0.0366)	1.1786 (0.0400)	0.7390 (0.0273)	1.1138 (0.0306)
	Perturbation 1	0.8518 (0.0415)	1.2104 (0.0454)	0.7214 (0.0267)	1.0954 (0.0304)
	Perturbation 2	0.8659 (0.0333)	1.2228 (0.0362)	0.7611 (0.0251)	1.1349 (0.0283)
2	Lasso initial	1.3203 (0.0689)	1.9486 (0.0817)	0.9822 (0.0400)	1.4702 (0.0519)
	Perturbation 1	1.3101 (0.0774)	1.9384 (0.0914)	0.9697 (0.0402)	1.4598 (0.0501)
	Perturbation 2	1.2709 (0.0583)	1.8969 (0.0728)	1.0309 (0.0440)	1.5198 (0.0530)
3	Lasso initial	1.8291 (0.1016)	2.8361 (0.1317)	1.3233 (0.0651)	2.0187 (0.0945)
	Perturbation 1	1.8058 (0.0966)	2.8125 (0.1300)	1.3173 (0.0707)	2.0116 (0.1022)
	Perturbation 2	1.8121 (0.0931)	2.8130 (0.1261)	1.3008 (0.0602)	1.9974 (0.0940)
4	Lasso initial	2.3206 (0.1298)	3.8045 (0.1841)	1.6963 (0.0919)	2.7348 (0.1659)
	Perturbation 1	2.4047 (0.1537)	3.8629 (0.2112)	1.6833 (0.0935)	2.7177 (0.1615)
	Perturbation 2	2.3219 (0.1276)	3.7928 (0.1887)	1.6905 (0.0929)	2.7284 (0.1683)
		Scenario 2 (Nonlinear)			
1	Lasso initial	0.5473 (0.0136)	0.8054 (0.0149)	0.6539 (0.0191)	1.0290 (0.0197)
	Perturbation 1	0.5797 (0.0174)	0.8350 (0.0180)	0.6593 (0.0191)	1.0335 (0.0192)
	Perturbation 2	0.6656 (0.0194)	0.9184 (0.0205)	0.7565 (0.0242)	1.1264 (0.0242)
2	Lasso initial	0.6268 (0.0195)	0.9081 (0.0215)	0.7320 (0.0235)	1.1097 (0.0241)
	Perturbation 1	0.6366 (0.0180)	0.9148 (0.0200)	0.7381 (0.0207)	1.1184 (0.0209)
	Perturbation 2	0.7240 (0.0223)	0.9985 (0.0241)	0.8363 (0.0266)	1.2133 (0.0263)
3	Lasso initial	0.7194 (0.0265)	1.0383 (0.0296)	0.8021 (0.0273)	1.1963 (0.0286)
	Perturbation 1	0.7212 (0.0237)	1.0383 (0.0272)	0.8092 (0.0243)	1.2042 (0.0257)
	Perturbation 2	0.8102 (0.0289)	1.1201 (0.0305)	0.9089 (0.0317)	1.3006 (0.0323)
4	Lasso initial	0.8304 (0.0328)	1.1898 (0.0368)	0.8735 (0.0288)	1.2914 (0.0319)
	Perturbation 1	0.8106 (0.0282)	1.1691 (0.0328)	0.9045 (0.0305)	1.3248 (0.0336)
	Perturbation 2	0.9092 (0.0374)	1.2627 (0.0405)	0.9719 (0.0322)	1.3874 (0.0341)
		Scenario 3 (Nonlinear)			
1	Lasso initial	0.5390 (0.0150)	0.8042 (0.0164)	0.5762 (0.0166)	0.9370 (0.0179)
	Perturbation 1	0.5540 (0.0162)	0.8172 (0.0172)	0.5800 (0.0168)	0.9390 (0.0176)
	Perturbation 2	0.6214 (0.0199)	0.8831 (0.0211)	0.7565 (0.0242)	1.1264 (0.0242)
2	Lasso initial	0.6292 (0.0213)	0.9325 (0.0229)	0.6455 (0.0202)	1.0244 (0.0221)
	Perturbation 1	0.6175 (0.0209)	0.9192 (0.0228)	0.6430 (0.0198)	1.0205 (0.0213)
	Perturbation 2	0.6764 (0.0225)	0.9763 (0.0249)	0.8363 (0.0266)	1.2133 (0.0263)
3	Lasso initial	0.7539 (0.0337)	1.1154 (0.0379)	0.7261 (0.0271)	1.1361 (0.0302)
	Perturbation 1	0.6955 (0.0214)	1.0554 (0.0265)	0.7158 (0.0235)	1.1229 (0.0263)
	Perturbation 2	0.7754 (0.0295)	1.1326 (0.0339)	0.9089 (0.0317)	1.3006 (0.0323)
4	Lasso initial	0.8445 (0.0354)	1.2770 (0.0426)	0.8008 (0.0290)	1.2549 (0.0348)
	Perturbation 1	0.8355 (0.0320)	1.2662 (0.0399)	0.8341 (0.0337)	1.2829 (0.0377)
	Perturbation 2	0.8890 (0.0378)	1.3190 (0.0454)	0.9719 (0.0322)	1.3874 (0.0341)

the South (all measured in 1966) as confounding variables. The high-dimensional linear transfer learning method proposed by Li et al. (2022) is applied to the deconfounded data to obtain the oracle estimate $\hat{\eta}_{\text{ora}}$ for $\eta_0 = \beta_t - \beta_s$, which serves as the gold standard. We compare the l_2-norm error between the estimated model-shift $\hat{\eta}$ and the gold standard $\hat{\eta}_{\text{ora}}$.

Table S.7: The l_2-norm errors for estimating η_0 and the expectation of the CATE $\tau(x) = x'\eta_0$ using ProTrans and baselines.

Methods	ProTrans	SingleTrim	TransTrim	SingleFarm	TransFarm
Error for η_0	0.0041	0.0156	0.0157	0.0171	0.0185
Error for $\mathbb{E}\tau(x)$	0.6025	0.9342	0.8373	1.7906	1.8794

The results, presented in Table S.7, indicate that ProTrans is least influenced by confounders compared to the baselines, and has a clear advantage towards treatment effect estimation at the presence of confounders. In all four baselines, covariates such as the amount of liabilities and the father's job level exhibit a significant negative effect on the earnings difference between the treatment and control groups. Conversely, variables related to working hours and the age of individuals living with the respondents show a positive effect. However, both ProTrans and the oracle deconfounded model indicate that these variables do not substantially affect the earnings difference. These variables appear as false positives in the baseline results due to the influence of hidden confounders, whereas ProTrans remains robust against such confounding effects.

S.4.4 Supplement numerical results

We report the performance under the settings $p = 500, 1500$ with other settings the same as that in the main text in Table S.8. The results for confounder level $\|\phi_t\|_2 = 1, 2, 3, 4$ are shown in Tables S.9 and S.10. We report the l_1 norm errors of ProTrans and the baseline methods in the high-dimensional simulation study, as shown in Tables S.11, S.12

and S.13, S.14 and S.15. We present the results for different choices of ρ in the real-data experiment "gene expression in tissues" in Table S.16. The results indicate that the method's performance is not highly sensitive to the choice of ρ, and selecting a reasonable value of ρ is sufficient for good performance.

Table S.8: The l_2 norm errors of different methods across various dimensions p.

p	Method	single-source		multi-source	
		η	β_t	η	β_t
Scenario 1 (linear)					
500	SingleTrim	0.5487 (0.0475)	0.6848 (0.0580)	0.4317 (0.0378)	0.5504 (0.0456)
	TransTrim	0.4336 (0.0333)	0.3932 (0.0305)	0.3640 (0.0249)	0.3407 (0.0228)
	SingleFarm	0.8095 (0.0419)	1.2055 (0.0551)	0.8065 (0.0381)	1.2223 (0.0494)
	TransFarm	**0.1712** (0.0161)	**0.2493** (0.0170)	0.2016 (0.0199)	0.2890 (0.0201)
	ProTrans	0.2460 (0.0165)	0.2502 (0.0165)	**0.1965** (0.0121)	**0.2167** (0.0107)
1500	SingleTrim	0.7046 (0.0477)	0.8852 (0.0574)	0.5500 (0.0335)	0.6990 (0.0410)
	TransTrim	0.4317 (0.0298)	0.3883 (0.0265)	0.3566 (0.0202)	0.3201 (0.0180)
	SingleFarm	0.9031 (0.0540)	1.2649 (0.0651)	0.8402 (0.0544)	1.2124 (0.0653)
	TransFarm	0.2876 (0.0225)	0.3106 (0.0205)	0.2525 (0.0206)	0.2870 (0.0186)
	ProTrans	**0.2477** (0.0156)	**0.2364** (0.0137)	**0.1989** (0.0103)	**0.1990** (0.0090)
Scenario 2 (nonlinear)					
500	SingleTrim	3.4026 (0.3320)	4.1600 (0.4037)	1.9179 (0.1755)	2.4167 (0.1625)
	TransTrim	2.4928 (0.2576)	2.3104 (0.2446)	1.6905 (0.1799)	1.5379 (0.1683)
	SingleFarm	6.9808 (0.3406)	8.3018 (0.3801)	8.9792 (0.5033)	10.757 (0.5450)
	TransFarm	1.9045 (0.1755)	2.0508 (0.1625)	2.3952 (0.1853)	2.5490 (0.1732)
	ProTrans	**0.6301** (0.0599)	**0.7037** (0.0657)	**0.4209** (0.0449)	**0.5159** (0.0455)
1500	SingleTrim	4.4333 (0.3664)	5.2496 (0.4290)	2.3954 (0.1666)	2.9493 (0.1554)
	TransTrim	2.5141 (0.2187)	2.2771 (0.1979)	1.6564 (0.1666)	1.4628 (0.1554)
	SingleFarm	7.8893 (0.5033)	9.3110 (0.5450)	10.097 (0.1853)	11.877 (0.1732)
	TransFarm	2.4598 (0.1666)	2.4086 (0.1554)	3.0365 (0.1853)	2.9568 (0.1732)
	ProTrans	**0.6042** (0.0368)	**0.5776** (0.0351)	**0.5348** (0.0449)	**0.5207** (0.0455)
Scenario 3 (nonlinear)					
500	SingleTrim	5.3515 (0.5104)	6.6901 (0.6328)	3.2439 (0.3362)	4.1267 (0.4137)
	TransTrim	3.9820 (0.4091)	3.6446 (0.3788)	2.7281 (0.2696)	2.4604 (0.2442)
	SingleFarm	3.5923 (0.2974)	4.9351 (0.3778)	4.1794 (0.3157)	5.6942 (0.3899)
	TransFarm	1.2493 (0.1489)	1.4569 (0.1434)	1.3989 (0.1610)	1.5589 (0.1496)
	ProTrans	**1.0273** (0.0972)	**1.1476** (0.1070)	**0.6867** (0.0652)	**0.8084** (0.0678)
1500	SingleTrim	7.1824 (0.5372)	8.6437 (0.6405)	4.2873 (0.3303)	5.2799 (0.3966)
	TransTrim	3.9866 (0.3298)	3.5776 (0.2942)	2.7010 (0.2115)	2.3886 (0.1908)
	SingleFarm	3.9399 (0.3144)	5.1576 (0.3745)	4.9658 (0.3983)	6.2776 (0.4465)
	TransFarm	1.6318 (0.1355)	1.6095 (0.1220)	1.7710 (0.1517)	1.7458 (0.1382)
	ProTrans	**0.7191** (0.0497)	**0.7511** (0.0523)	**0.6389** (0.0440)	**0.6539** (0.0430)

Table S.9: The l_2 norm errors of different methods across various levels of confounding.

$\|\phi_t\|_2$	Method	single-source η	single-source β_t	multi-source η	multi-source β_t
		Scenario 1 (linear)			
1	SingleTrim	0.7046 (0.0477)	0.8852 (0.0574)	0.5500 (0.0335)	0.6990 (0.0410)
	TransTrim	0.4317 (0.0298)	0.3883 (0.0265)	0.3566 (0.0202)	0.3201 (0.0180)
	SingleFarm	0.9031 (0.0540)	1.2649 (0.0651)	0.8402 (0.0544)	1.2124 (0.0653)
	TransFarm	0.2876 (0.0225)	0.3106 (0.0205)	0.2525 (0.0207)	0.2870 (0.0186)
	ProTrans	**0.2477** (0.0156)	**0.2364** (0.0137)	**0.1989** (0.0103)	**0.1990** (0.0090)
2	SingleTrim	2.1710 (0.0736)	2.6549 (0.0768)	1.5831 (0.1196)	1.9357 (0.1422)
	TransTrim	1.2523 (0.1039)	1.1359 (0.0932)	0.9060 (0.0636)	0.8151 (0.0585)
	SingleFarm	1.5738 (0.1019)	2.1868 (0.1291)	1.1739 (0.0755)	1.6664 (0.0931)
	TransFarm	0.6719 (0.0581)	0.6869 (0.0527)	0.4152 (0.0328)	0.4494 (0.0293)
	ProTrans	**0.5908** (0.0440)	**0.5682** (0.0399)	**0.3678** (0.0231)	**0.3612** (0.0219)
		Scenario 2 (nonlinear)			
1	SingleTrim	0.2632 (0.0125)	0.3358 (0.0145)	0.2342 (0.0212)	0.2958 (0.0229)
	TransTrim	0.1760 (0.0085)	0.1581 (0.0071)	0.1961 (0.0129)	0.1798 (0.0097)
	SingleFarm	1.8268 (0.1068)	2.1755 (0.1172)	9.5069 (0.0148)	10.9077 (0.0212)
	TransFarm	0.5314 (0.0313)	0.5228 (0.0304)	2.5265 (0.0861)	2.5000 (0.0884)
	ProTrans	**0.1251** (0.0061)	**0.1197** (0.0050)	**0.1495** (0.0215)	**0.1621** (0.0239)
2	SingleTrim	0.4518 (0.0255)	0.5641 (0.0303)	0.3297 (0.0437)	0.4182 (0.0481)
	TransTrim	0.2852 (0.0163)	0.2560 (0.0146)	0.2595 (0.0166)	0.2346 (0.0192)
	SingleFarm	1.8901 (0.1116)	2.2608 (0.1215)	9.5360 (0.0090)	10.9435 (0.0171)
	TransFarm	0.5665 (0.0337)	0.5551 (0.0320)	2.5589 (0.0818)	2.5279 (0.0845)
	ProTrans	**0.1522** (0.0075)	**0.1446** (0.0064)	**0.1802** (0.0168)	**0.1861** (0.0207)
		Scenario 3 (nonlinear)			
1	SingleTrim	0.3116 (0.0162)	0.4022 (0.0192)	0.2632 (0.0122)	0.3438 (0.0155)
	TransTrim	0.1945 (0.0099)	0.1755 (0.0085)	0.2033 (0.0087)	0.1865 (0.0971)
	SingleFarm	0.7001 (0.0548)	0.9190 (0.0633)	3.7867 (0.3223)	4.6118 (0.3630)
	TransFarm	0.2454 (0.0177)	0.2486 (0.0166)	1.1057 (0.1137)	1.1105 (0.1047)
	ProTrans	**0.1283** (0.0060)	**0.1223** (0.0050)	**0.1344** (0.0068)	**0.1447** (0.0055)
2	SingleTrim	0.6401 (0.0413)	0.8016 (0.0495)	0.4422 (0.0255)	0.5723 (0.0316)
	TransTrim	0.3757 (0.0231)	0.3382 (0.0209)	0.3110 (0.0157)	0.2792 (0.0136)
	SingleFarm	0.7932 (0.0580)	1.0473 (0.0676)	3.8741 (0.3293)	4.7238 (0.3685)
	TransFarm	0.2985 (0.0222)	0.2989 (0.0204)	1.1311 (0.1153)	1.1343 (0.1116)
	ProTrans	**0.1610** (0.0078)	**0.1548** (0.0068)	**0.1632** (0.0077)	**0.1716** (0.0066)

Table S.10: The l_2 norm errors of different methods across various levels of confounding (continued).

$\|\phi_t\|_2$	Method	single-source		multi-source	
		η	β_t	η	β_t
		Scenario 1 (linear)			
3	SingleTrim	4.6748 (0.3813)	5.6459 (0.4594)	3.3765 (0.2779)	4.0441 (0.3258)
	TransTrim	2.5785 (0.2299)	2.3449 (0.2054)	1.7611 (0.1383)	1.5987 (0.1274)
	SingleFarm	2.8351 (0.2094)	3.8500 (0.2690)	1.7447 (0.1218)	2.4160 (0.1514)
	TransFarm	1.3929 (0.1244)	1.3750 (0.1129)	0.7160 (0.0579)	0.7397 (0.0521)
	ProTrans	**1.1383** (0.0919)	**1.1047** (0.0846)	**0.6171** (0.0426)	**0.6081** (0.0409)
4	SingleTrim	7.7609 (0.5610)	9.3499 (0.6788)	5.8085 (0.4662)	6.8857 (0.5428)
	TransTrim	4.3785 (0.3877)	3.9669 (0.3437)	2.9219 (0.2039)	2.6616 (0.2532)
	SingleFarm	4.6142 (0.3589)	6.1206 (0.4514)	2.6300 (0.2039)	3.5476 (0.1677)
	TransFarm	2.4432 (0.2223)	2.3621 (0.1997)	1.1707 (0.0980)	1.1701 (0.0893)
	ProTrans	**1.9081** (0.1599)	**1.8653** (0.1472)	**0.9478** (0.0732)	**0.9452** (0.0711)
		Scenario 2 (nonlinear)			
3	SingleTrim	0.7740 (0.0522)	0.9505 (0.0623)	0.4912 (0.0648)	0.6218 (0.0706)
	TransTrim	0.4720 (0.0312)	0.4242 (0.0284)	0.3736 (0.0124)	0.3334 (0.0173)
	SingleFarm	1.9822 (0.1190)	2.3895 (0.1291)	9.5974 (0.0104)	11.0360 (0.0163)
	TransFarm	0.6331 (0.0380)	0.6170 (0.0355)	2.6139 (0.0840)	2.5783 (0.0870)
	ProTrans	**0.1855** (0.0094)	**0.1771** (0.0082)	**0.2179** (0.0135)	**0.2194** (0.0179)
4	SingleTrim	1.2471 (0.0926)	1.5130 (0.1099)	0.7252 (0.0973)	0.9103 (0.0997)
	TransTrim	0.7497 (0.0593)	0.6730 (0.0533)	0.5359 (0.0067)	0.4757 (0.0106)
	SingleFarm	2.0977 (0.1252)	2.5547 (0.1371)	9.6475 (0.0092)	11.1246 (0.0161)
	TransFarm	0.7074 (0.0457)	0.6853 (0.0417)	2.6714 (0.0799)	2.6287 (0.0833)
	ProTrans	**0.2315** (0.0118)	**0.2219** (0.0106)	**0.2618** (0.0076)	**0.2596** (0.0131)
		Scenario 3 (nonlinear)			
3	SingleTrim	1.2124 (0.0914)	1.4907 (0.1090)	0.7565 (0.0504)	0.9640 (0.0618)
	TransTrim	0.6878 (0.0491)	0.6188 (0.0444)	0.5055 (0.0308)	0.4491 (0.0274)
	SingleFarm	0.9351 (0.0672)	1.2418 (0.0793)	4.0261 (0.3482)	4.9209 (0.3859)
	TransFarm	0.3866 (0.0294)	0.3827 (0.0265)	1.1769 (0.1172)	1.1781 (0.1129)
	ProTrans	**0.2146** (0.0114)	**0.2099** (0.0106)	**0.2072** (0.0102)	**0.2134** (0.0092)
4	SingleTrim	1.9247 (0.1356)	2.3586 (0.1631)	1.1918 (0.0844)	1.5013 (0.1031)
	TransTrim	1.1118 (0.0824)	0.9923 (0.0732)	0.7865 (0.0529)	0.6962 (0.0473)
	SingleFarm	1.1541 (0.0874)	1.5302 (0.1038)	4.1611 (0.3548)	5.1095 (0.3931)
	TransFarm	0.5076 (0.0391)	0.4975 (0.0352)	1.2456 (0.1223)	1.2443 (0.1166)
	ProTrans	**0.2767** (0.0159)	**0.2759** (0.0153)	**0.2646** (0.0138)	**0.2711** (0.0132)

Table S.11: The l_1 norm errors of different methods across various dimensions p.

p	Method	single-source η	single-source β_t	multi-source η	multi-source β_t
Scenario 1 (linear)					
500	SingleTrim	1.8195 (0.0760)	2.1458 (0.0845)	1.6540 (0.0687)	1.9932 (0.0719)
	TransTrim	0.9743 (0.0440)	1.3645 (0.0522)	0.9516 (0.0480)	1.3527 (0.0496)
	SingleFarm	2.6029 (0.0836)	3.0369 (0.0832)	2.6500 (0.0997)	3.1352 (0.0963)
	TransFarm	0.6023 (0.0515)	1.1654 (0.0527)	0.8129 (0.0852)	1.4081 (0.0842)
	ProTrans	**0.7323** (0.0302)	**1.1414** (0.0396)	**0.6824** (0.0244)	**1.1273** (0.0318)
1000	SingleTrim	2.1605 (0.0795)	2.4789 (0.0825)	1.9753 (0.0769)	2.3055 (0.0788)
	TransTrim	0.9696 (0.0311)	1.3574 (0.0381)	0.9115 (0.0296)	1.2970 (0.0295)
	SingleFarm	2.7750 (0.0759)	3.2662 (0.0783)	2.8200 (0.0954)	3.3460 (0.0964)
	TransFarm	0.7553 (0.0521)	1.2817 (0.0523)	0.7364 (0.0449)	1.2742 (0.0414)
	ProTrans	**0.7811** (0.0230)	**1.1831** (0.0322)	**0.7450** (0.0297)	**1.1530** (0.0320)
Scenario 2 (nonlinear)					
500	SingleTrim	5.4301 (0.2968)	5.9343 (0.3099)	3.5403 (0.1891)	4.1391 (0.2056)
	TransTrim	2.4838 (0.1591)	3.3413 (0.1959)	1.8937 (0.1084)	2.5522 (0.1208)
	SingleFarm	8.4803 (0.2771)	8.1500 (0.2363)	8.5488 (0.2754)	8.8189 (0.2241)
	TransFarm	2.0412 (0.1235)	3.6249 (0.1792)	2.5111 (0.1922)	3.7706 (0.2546)
	ProTrans	**1.2316** (0.0764)	**2.0804** (0.1216)	**1.0156** (0.0416)	**1.7158** (0.0573)
1000	SingleTrim	5.5720 (0.2344)	6.0550 (0.2440)	3.6734 (0.1471)	4.2793 (0.1608)
	TransTrim	2.3638 (0.1345)	3.0491 (0.1607)	1.8010 (0.0782)	2.4194 (0.0880)
	SingleFarm	7.7187 (0.2203)	8.3268 (0.2189)	8.4242 (0.2040)	9.1910 (0.2050)
	TransFarm	2.3127 (0.1113)	3.0816 (0.1125)	2.5887 (0.1005)	3.3723 (0.1056)
	ProTrans	**1.3256** (0.0549)	**2.0054** (0.0810)	**1.1863** (0.0424)	**1.7965** (0.0502)
Scenario 3 (nonlinear)					
500	SingleTrim	6.5854 (0.3279)	7.2482 (0.3453)	4.6464 (0.2542)	5.3890 (0.2768)
	TransTrim	2.9871 (0.1730)	4.0056 (0.2109)	2.3197 (0.1332)	3.1194 (0.1529)
	SingleFarm	5.8293 (0.2610)	6.0358 (0.2333)	5.8078 (0.2357)	6.5672 (0.2471)
	TransFarm	1.6083 (0.1110)	3.1416 (0.1606)	1.9433 (0.1977)	2.9370 (0.1982)
	ProTrans	**1.5494** (0.1107)	**2.6031** (0.1526)	**1.2525** (0.0601)	**2.0885** (0.0828)
1000	SingleTrim	6.8711 (0.2805)	7.5761 (0.3010)	4.9216 (0.2109)	5.6592 (0.2298)
	TransTrim	2.8971 (0.1653)	3.8045 (0.2087)	2.2496 (0.1084)	2.9897 (0.1257)
	SingleFarm	5.1877 (0.2062)	5.9738 (0.2051)	5.9599 (0.2371)	6.8234 (0.2384)
	TransFarm	1.6692 (0.0771)	2.6243 (0.0997)	1.8514 (0.0827)	2.7316 (0.0902)
	ProTrans	**1.3993** (0.0621)	**2.2363** (0.0907)	**1.2847** (0.0638)	**2.0212** (0.0773)

Table S.12: The l_1 norm errors of different methods across various dimensions p (continued).

p	Method	single-source		multi-source	
		η	β_t	η	β_t
		Scenario 1 (linear)			
1500	SingleTrim	2.3097 (0.0907)	2.6143 (0.0942)	2.0223 (0.0736)	2.3402 (0.0768)
	TransTrim	0.9943 (0.0407)	1.3520 (0.0449)	0.9494 (0.0457)	1.3134 (0.0497)
	SingleFarm	2.6162 (0.0863)	3.0241 (0.0870)	2.5491 (0.0877)	2.9786 (0.0866)
	TransFarm	0.7665 (0.0311)	1.2127 (0.0336)	0.7505 (0.0374)	1.2175 (0.0382)
	ProTrans	**0.8181** (0.0366)	**1.1786** (0.0400)	**0.7390** (0.0273)	**1.1138** (0.0306)
2000	SingleTrim	2.5860 (0.1098)	2.8727 (0.1111)	2.2090 (0.0898)	2.5573 (0.0907)
	TransTrim	1.0220 (0.0354)	1.4569 (0.0475)	0.9604 (0.0305)	1.3573 (0.0362)
	SingleFarm	2.7064 (0.0880)	3.0435 (0.0816)	2.5504 (0.0861)	2.9596 (0.0847)
	TransFarm	0.8862 (0.0412)	1.3660 (0.0611)	0.8225 (0.0260)	1.2751 (0.0271)
	ProTrans	**0.8555** (0.0285)	**1.2593** (0.0337)	**0.7716** (0.0183)	**1.1731** (0.0235)
		Scenario 2 (nonlinear)			
1500	SingleTrim	5.9946 (0.2966)	6.3513 (0.2872)	3.8495 (0.1569)	4.3813 (0.1711)
	TransTrim	2.4263 (0.1410)	3.0659 (0.1604)	1.7852 (0.0803)	2.3472 (0.0961)
	SingleFarm	7.9122 (0.2061)	8.4375 (0.2077)	8.4092 (0.2205)	9.0614 (0.2231)
	TransFarm	2.5724 (0.1246)	3.2393 (0.1318)	2.5827 (0.1114)	3.2693 (0.1113)
	ProTrans	**1.3296** (0.0616)	**1.9504** (0.0813)	**1.2620** (0.0567)	**1.8184** (0.0653)
2000	SingleTrim	6.7726 (0.3288)	7.0187 (0.2791)	4.4771 (0.2073)	5.0657 (0.2198)
	TransTrim	2.4105 (0.1118)	3.2059 (0.1783)	1.8519 (0.0727)	2.4771 (0.0880)
	SingleFarm	8.1257 (0.2411)	8.6057 (0.2323)	9.1359 (0.2702)	9.8391 (0.2682)
	TransFarm	2.6521 (0.2081)	3.3874 (0.2075)	3.1329 (0.1681)	3.8588 (0.1690)
	ProTrans	**1.3190** (0.0449)	**2.2082** (0.1429)	**1.2690** (0.0394)	**1.8669** (0.0512)
		Scenario 3 (nonlinear)			
1500	SingleTrim	7.3034 (0.3405)	7.7163 (0.3126)	5.0637 (0.2061)	5.7136 (0.2264)
	TransTrim	2.9041 (0.1225)	3.6870 (0.1457)	2.2195 (0.0867)	2.8957 (0.1079)
	SingleFarm	5.1953 (0.2089)	5.8374 (0.2096)	5.8308 (0.2621)	6.5458 (0.2635)
	TransFarm	1.9394 (0.1180)	2.7472 (0.1281)	1.9292 (0.1017)	2.6916 (0.1029)
	ProTrans	**1.4337** (0.0781)	**2.2278** (0.1074)	**1.2024** (0.0520)	**1.8801** (0.0694)
2000	SingleTrim	7.9291 (0.3461)	8.4003 (0.3109)	5.7839 (0.2588)	6.5066 (0.2791)
	TransTrim	3.0132 (0.1282)	3.9696 (0.1976)	2.3438 (0.0953)	3.0940 (0.1175)
	SingleFarm	5.5249 (0.2224)	6.1933 (0.2221)	6.2591 (0.2416)	7.0178 (0.2403)
	TransFarm	1.8790 (0.0731)	2.7432 (0.0882)	2.2212 (0.0938)	3.0268 (0.0960)
	ProTrans	**1.3983** (0.0547)	**2.3879** (0.1489)	**1.3113** (0.0521)	**2.0372** (0.0683)

Table S.13: The l_1 norm errors of different methods across various levels of confounding.

$\|\phi_t\|_2$	**Method**	**single-source**		**multi-source**	
		η	β_t	η	β_t
		Scenario 1 (linear)			
1	SingleTrim	2.3097 (0.0906)	2.6143 (0.0942)	2.0223 (0.0736)	2.3402 (0.0768)
	TransTrim	0.9943 (0.0407)	1.3520 (0.0449)	0.9494 (0.0457)	1.3134 (0.0496)
	SingleFarm	2.6162 (0.0863)	3.0241 (0.0870)	2.5491 (0.0877)	2.9786 (0.0866)
	TransFarm	**0.7665** (0.0311)	1.2127 (0.0336)	0.7505 (0.0374)	1.2175 (0.0382)
	ProTrans	0.8181 (0.0366)	**1.1786** (0.0400)	**0.7390** (0.0273)	**1.1138** (0.0306)
2	SingleTrim	4.3591 (0.2045)	4.6880 (0.2096)	3.6653 (0.1655)	4.0422 (0.1694)
	TransTrim	1.7303 (0.0947)	2.3870 (0.1141)	1.4927 (0.0752)	2.0270 (0.0889)
	SingleFarm	3.5150 (0.1376)	3.9653 (0.1392)	2.9595 (0.0983)	3.4302 (0.0981)
	TransFarm	**1.1390** (0.0614)	**1.8335** (0.0763)	**0.9754** (0.0469)	1.5454 (0.0516)
	ProTrans	1.3203 (0.0689)	1.9486 (0.0817)	0.9822 (0.0400)	**1.4702** (0.0519)
		Scenario 2 (nonlinear)			
1	SingleTrim	1.3708 (0.0396)	1.6030 (0.0410)	1.1959 (0.0355)	1.4836 (0.0362)
	TransTrim	0.6438 (0.0231)	0.8915 (0.0230)	0.6710 (0.0167)	1.0053 (0.0182)
	SingleFarm	3.7394 (0.0929)	4.0293 (0.0941)	8.2207 (0.1937)	8.7239 (0.1991)
	TransFarm	1.1853 (0.0610)	1.4886 (0.0628)	2.4276 (0.1224)	2.9684 (0.1247)
	ProTrans	**0.5473** (0.0136)	**0.8054** (0.0149)	**0.6539** (0.0191)	**1.0290** (0.0197)
2	SingleTrim	1.8670 (0.0673)	2.1221 (0.0692)	1.4633 (0.0484)	1.7720 (0.0480)
	TransTrim	0.8439 (0.0404)	1.1185 (0.0421)	0.7706 (0.0242)	1.1252 (0.0261)
	SingleFarm	3.8145 (0.0933)	4.1184 (0.0936)	8.2703 (0.1975)	8.7748 (0.2027)
	TransFarm	1.2250 (0.0603)	1.5448 (0.0621)	2.4370 (0.1158)	2.9850 (0.1174)
	ProTrans	**0.6268** (0.0195)	**0.9081** (0.0215)	**0.7320** (0.0235)	**1.1097** (0.0241)
		Scenario 3 (nonlinear)			
1	SingleTrim	1.5015 (0.0450)	1.7433 (0.0464)	1.3010 (0.0376)	1.5960 (0.0390)
	TransTrim	0.6801 (0.0269)	0.9371 (0.0274)	0.6784 (0.0190)	1.0114 (0.0204)
	SingleFarm	2.2170 (0.0773)	2.5115 (0.0790)	5.0659 (0.2112)	5.5443 (0.2158)
	TransFarm	0.7293 (0.0290)	1.0400 (0.0297)	1.4899 (0.0958)	1.9981 (0.0971)
	ProTrans	**0.5390** (0.0150)	**0.8042** (0.0164)	**0.5762** (0.0166)	**0.9370** (0.0179)
2	SingleTrim	2.2096 (0.0809)	2.4896 (0.0839)	1.7089 (0.0561)	2.0344 (0.0591)
	TransTrim	0.9707 (0.0568)	1.2694 (0.0585)	0.8359 (0.0273)	1.1931 (0.0305)
	SingleFarm	2.3528 (0.0819)	2.6774 (0.0834)	5.1147 (0.2212)	5.6023 (0.2249)
	TransFarm	0.7869 (0.0275)	1.1292 (0.0291)	1.5231 (0.0884)	2.0496 (0.0903)
	ProTrans	**0.6292** (0.0213)	**0.9325** (0.0229)	**0.6455** (0.0202)	**1.0244** (0.0221)

Table S.14: The l_1 norm errors of different methods across various levels of confounding (continued).

$\|\phi_t\|_2$	Method	single-source		multi-source	
		η	β_t	η	β_t
		Scenario 1 (linear)			
3	SingleTrim	6.5309 (0.3030)	6.6745 (0.2814)	5.4460 (0.2483)	5.7809 (0.2522)
	TransTrim	2.4135 (0.1283)	3.4964 (0.1711)	2.1108 (0.1117)	2.9469 (0.1431)
	SingleFarm	4.6487 (0.1928)	5.1329 (0.1943)	3.6416 (0.1311)	4.1364 (0.1288)
	TransFarm	**1.6853** (0.0943)	**2.6815** (0.1225)	**1.2502** (0.0559)	**1.9744** (0.0726)
	ProTrans	1.8291 (0.1016)	2.8361 (0.1317)	1.3233 (0.0651)	2.0187 (0.0945)
4	SingleTrim	8.3423 (0.3818)	8.3820 (0.3232)	7.1110 (0.3115)	7.3811 (0.3084)
	TransTrim	3.1478 (0.1682)	4.6246 (0.2297)	2.7484 (0.1487)	3.9419 (0.2010)
	SingleFarm	5.9089 (0.2536)	6.3974 (0.2386)	4.5217 (0.1809)	4.9902 (0.1677)
	TransFarm	**2.2717** (0.1269)	**3.6233** (0.1789)	**1.5905** (0.0817)	**2.5151** (0.1150)
	ProTrans	2.3206 (0.1298)	3.8045 (0.1841)	1.6963 (0.0919)	2.7348 (0.1659)
		Scenario 2 (nonlinear)			
3	SingleTrim	2.4679 (0.0996)	2.7586 (0.1043)	1.7907 (0.0647)	2.1289 (0.0677)
	TransTrim	1.0816 (0.0597)	1.3941 (0.0630)	0.9154 (0.0353)	1.2941 (0.0397)
	SingleFarm	3.8923 (0.0941)	4.2200 (0.0935)	8.3001 (0.2020)	8.8153 (0.2062)
	TransFarm	1.2782 (0.0591)	1.6262 (0.0606)	2.4848 (0.1229)	3.0428 (0.1243)
	ProTrans	**0.7194** (0.0265)	**1.0383** (0.0296)	**0.8021** (0.0273)	**1.1963** (0.0286)
4	SingleTrim	3.0709 (0.1285)	3.4017 (0.1355)	2.1902 (0.0853)	2.5572 (0.0898)
	TransTrim	1.3476 (0.0800)	1.7053 (0.0843)	1.0811 (0.0482)	1.4910 (0.0543)
	SingleFarm	3.9919 (0.0958)	4.3549 (0.0959)	8.2977 (0.2078)	8.8288 (0.2117)
	TransFarm	1.2816 (0.0468)	1.6702 (0.0485)	2.4937 (0.1182)	3.0695 (0.1196)
	ProTrans	**0.8304** (0.0328)	**1.1898** (0.0368)	**0.8735** (0.0288)	**1.2914** (0.0319)
		Scenario 3 (nonlinear)			
3	SingleTrim	2.9733 (0.1154)	3.3088 (0.1219)	2.2306 (0.0838)	2.5937 (0.0886)
	TransTrim	1.2412 (0.0615)	1.6002 (0.0668)	1.0322 (0.0368)	1.4283 (0.0429)
	SingleFarm	2.5463 (0.0896)	2.9172 (0.0918)	5.1915 (0.2247)	5.7027 (0.2271)
	TransFarm	0.9096 (0.0396)	1.2972 (0.0432)	1.5191 (0.0815)	2.0693 (0.0835)
	ProTrans	**0.7539** (0.0337)	**1.1154** (0.0379)	**0.7261** (0.0271)	**1.1361** (0.0302)
4	SingleTrim	3.6797 (0.1406)	4.0797 (0.1503)	2.7830 (0.1090)	3.1936 (0.1169)
	TransTrim	1.5749 (0.0782)	2.0035 (0.0855)	1.2410 (0.0455)	1.6844 (0.0544)
	SingleFarm	2.7717 (0.0981)	3.1919 (0.1022)	5.3336 (0.2366)	5.8735 (0.2381)
	TransFarm	1.0428 (0.0564)	1.4960 (0.0626)	1.5691 (0.0794)	2.1503 (0.0809)
	ProTrans	**0.8445** (0.0354)	**1.2770** (0.0426)	**0.8008** (0.0290)	**1.2549** (0.0348)

Table S.15: The l_1 norm errors of source selected methods under different selection level ρ.

Scenario	$\boldsymbol{p}$	$\rho = 1$	$\rho = 1.2$	$\rho = 1.4$	$\rho = 1.6$	$\rho = 2$	$\rho = 2.5$	$\rho = 3$	$\rho = \infty$
Scenario 1	500	0.8944	0.8350	0.7887	0.7895	0.8091	0.8437	0.8671	0.9048
	1000	0.9846	0.9033	0.8629	0.8224	0.8441	0.8976	0.9124	0.9862
	1500	1.0281	0.9350	0.8639	0.8665	0.8779	0.9083	0.9483	1.0364
	2000	1.0357	0.9371	0.8674	0.8560	0.8847	0.9521	0.9923	1.0329
Scenario 2	500	3.9340	3.5863	3.5848	3.6725	4.0116	3.8292	4.3169	6.2596
	1000	3.6200	3.5017	3.4292	3.5948	3.5111	3.6850	3.9392	6.5671
	1500	3.5815	3.4722	3.3414	3.3652	3.3568	3.5381	3.9280	6.5039
	2000	3.7953	3.6762	3.4330	3.4883	3.6191	3.9827	4.2696	6.6199
Scenario 3	500	5.1863	4.8025	4.5209	4.5675	4.3928	4.7071	4.9702	6.0026
	1000	4.6869	4.5383	4.0640	3.9247	4.2190	4.0945	4.6076	6.1151
	1500	4.1760	4.0700	4.0169	3.9059	3.9513	4.1056	4.1599	6.0418
	2000	4.4679	3.9434	3.7892	3.5979	3.5911	3.9393	3.9653	6.1384

Table S.16: The l_2 norm errors for estimating β_t using the proposed methods with source selection $\rho = 1$ to 3 for different target tissues.

Tissue	Adipose	Artery	Blood	Heart	Lung	Muscle	Spleen	Testis
$\rho = 1$	0.1393	0.0975	0.2035	0.1041	0.2641	0.1180	0.1098	0.0844
$\rho = 1.02$	0.1393	0.0975	0.2035	0.1041	0.2641	0.1180	0.1098	0.0844
$\rho = 1.1$	0.1393	0.1179	0.2023	0.1134	0.2641	0.1180	0.1098	0.0902
$\rho = 1.5$	0.4580	0.1179	0.2023	0.1219	0.2916	0.1176	0.1076	0.2741
$\rho = 3$	0.1681	0.2143	0.2036	0.1086	0.2916	0.2866	0.2835	0.2072

References

Bhardwaj, A. & Vu, V. (2024), Matrix perturbation: Davis-kahan in the infinity norm, in 'Proceedings of the 2024 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)', SIAM, pp. 880–934.

Chen, Z. & Yao, F. (2024), 'Matrix completion via residual spectral matching', arXiv preprint arXiv:2412.10005 .

Li, S., Cai, T. T. & 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.

Pensky, M. (2024), 'Davis-kahan theorem in the two-to-infinity norm and its application to perfect clustering', arXiv preprint arXiv:2411.11728 .

Tang, D., Kong, D. & Wang, L. (2023), 'The synthetic instrument: From sparse association to sparse causation', arXiv preprint arXiv:2304.01098 .

Wang, L. & Tchetgen Tchetgen, E. (2018), 'Bounded, efficient and multiply robust estimation of average treatment effects using instrumental variables', Journal of the Royal Statistical Society Series B: Statistical Methodology **80**(3), 531–550.