Statistica Sinica: Supplement

MULTILAYER NETWORK REGRESSION WITH EIGENVECTOR CENTRALITY AND COMMUNITY STRUCTURE

Fudan University and Columbia University

Supplementary Material

This document provides supplementary material for the paper on *Multilayer network regression with eigenvector centrality and community structure*. The supplementary materials consist of three sections. Section S1 provides theoretical supplements, including proofs of main theorems, analysis under unknown community structure, and discussion of key assumptions. Section S2 presents additional simulation results, comparisons with alternative models, and sensitivity analyses. Section S3 offers further details on the real-data application using WIOD, including variable definitions, estimation results, and comparisons of centrality measures. This supplementary material supports the main findings and methodology presented in the paper.

S1 Theoretical supplements

In this section, we provides theoretical supplements, including complete proofs of the main theorems, additional theoretical analysis for the case where community information is unknown, and a detailed discussion of key assumptions used in the paper.

To begin with, we introduce the following notation for projection matrices used throughout the analysis: $P_X := X(X^\top X)^{-1}X^\top$, $P_C := C(C^\top C)^{-1}C^\top$, and $P_{\hat{Z}} := \hat{Z}(\hat{Z}^\top \hat{Z})^{-1}\hat{Z}^\top$. These projection matrices play a central role in characterizing the properties of the estimators and their asymptotic behavior.

S1.1 Proof of Column Full-Rank Implication

Claim: If $\sigma_{\min}((I_{\mathsf{N}} - P_X)V) \geq l_{\mathsf{N}} > 0$, then $\mathbf{W}_1 = (X, C)$ is column full-rank.

Proof. Assume $\mathbf{W}_1 = [X \quad C]$ is *not* column full-rank. Then there exists a non-zero vector $\theta = [\theta_X^\top \quad \theta_C^\top]^\top \neq 0$ such that:

$$X\theta_X + C\theta_C = 0.$$

If $\theta_C = 0$, then $X\theta_X = 0$. Since X is column full-rank (by $\mathsf{N} > \mathsf{P} + \mathsf{L}$), this implies $\theta_X = 0$, contradicting $\theta \neq 0$. Thus, $\theta_C \neq 0$.

Rearranging the equation:

$$C\theta_C = -X\theta_X.$$

Projecting both sides onto the orthogonal complement of X:

$$(I_{\mathsf{N}} - P_X)C\theta_C = (I_{\mathsf{N}} - P_X)(-X\theta_X) = 0,$$

where we used $(I_\mathsf{N} - P_X)X = 0$. Substituting $C = a_\mathsf{N} V$:

$$(I_\mathsf{N} - P_X)V\theta_C = \frac{1}{a_\mathsf{N}}(I_\mathsf{N} - P_X)C\theta_C = 0.$$

This implies:

$$\sigma_{\min}\left((I_\mathsf{N} - P_X)V\right) \leq \|(I_\mathsf{N} - P_X)V\theta_C\|_2/\|\theta_C\|_2 = 0,$$

which contradicts $\sigma_{\min}\left((I_\mathsf{N} - P_X)V\right) \geq l_\mathsf{N} > 0$. Therefore, $\mathbf{W}_1$ must be column full-rank. □

S1.2 Three important Lemmas

Lemma S1.1. *(Davis and Kahan, 1970)Recall the network model in (2.8). Let $\delta := \lambda_1 - \lambda_2$ be the spectral gap between the largest and second largest eigenvalues of B_0. Suppose $\widetilde{u}_1$ and u_1 are the top eigenvectors of B and B_0, respectively. Then we have*

$$\|\widetilde{u}_1 - u_1\|_2 = O\left(\frac{\|E_0\|_2}{\delta}\right),$$

where $\|E_0\|_2 = \max_{\|u\|_2 \leq 1} \|E_0 u\|_2$ denotes the matrix operator norm.

Lemma S1.1 requires $\delta \gg \|E_0\|_2$ for $\widetilde{u}_1$ to converge to u_1. In our framework, this result directly translates to the estimation error bound between the noisy and true centrality measures. Specifically, we derive the following explicit rate for $\hat{C}$:

Lemma S1.2. *Under Assumptions 1-3, we have*

$$\mathbb{E}\left[\|\hat{C} - C\|_F^2\right] = O\left(\frac{a_{\mathsf{N}}^2 \mathsf{NL}}{\delta^2}\right). \tag{S1.1}$$

Proof. Under Assumption 2 and the setting of Lemma S1.1, we have

$$\begin{aligned}\|\hat{C} - C\|_F^2 &= \|\mathrm{vec}(\hat{C}) - \mathrm{vec}(C)\|_2^2 \\ &= a_{\mathsf{N}}^2 \left\|\widetilde{u}_1 - u_1\right\|_2^2,\end{aligned}$$

and from Lemma S1.1 we see that $\left\|\widetilde{u}_1 - u_1\right\|_2 = O\left(\frac{\|E_0\|_2}{\delta}\right)$. Therefore, $\left\|\widetilde{u}_1 - u_1\right\|_2^2 \le c\frac{\|E_0\|_2^2}{\delta^2}$ a.s. for some positive constant c. Therefore, we have

$$\mathbb{E}\left[\|\hat{C} - C\|_F^2\right] \le c a_{\mathsf{N}}^2 \frac{\mathbb{E}\left[\|E_0\|_2^2\right]}{\delta^2}.$$

Under Assumption 1 where $\mathbb{E}\left[\|E_0\|_2^2\right] = O(\mathsf{NL})$, we obtain

$$\mathbb{E}\left[\|\hat{C} - C\|_F^2\right] = O(\frac{a_{\mathsf{N}}^2 \mathsf{NL}}{\delta^2}).$$

□

In what follows, Lemma S1.3 is a powerful tool, which we now explain. It is used in the proofs of Theorem 4 and 5.

Lemma S1.3. *Suppose A and B are positive semi-definite matrices with the same size $n \times n$. Then we have $tr(AB) \le tr(A)tr(B)$.*

Proof. From Cauchy-Schwarz inequality, we have

$$\mathrm{tr}(AB) \le \|A\|_F \|B\|_F = \sqrt{\mathrm{tr}(A^2)}\sqrt{\mathrm{tr}(B^2)}.$$

Denote the eigenvalues of A as $\nu_i \geq 0, i = 1, \cdots n$, the eigenvalues of B as $\mu_i \geq 0, i = 1, \cdots, n$. Then $\sqrt{\mathrm{tr}(A^2)} = \sqrt{\sum \nu_i^2} \leq \sum \nu_i = \mathrm{tr}(A)$, and similarly we have $\sqrt{\mathrm{tr}(B^2)} = \sqrt{\sum \mu_i^2} \leq \sum \mu_i = \mathrm{tr}(B)$. Finally, we have $\mathrm{tr}(AB) \leq \mathrm{tr}(A)\mathrm{tr}(B)$ and the proof is complete. □

S1.3 Proof of Theorem 1

(i) Let $\mathbf{W}_1 = (X, C)$ and $\beta = (\beta_X^\top, \beta_C^\top)^\top$, then the OLS estimator is

$$\hat{\beta}^{(ols)} = \arg\min_{\beta_X, \beta_C} \|y - X\beta_X - C\beta_C\|_2^2 .$$

Define also that $\mathbb{L} := \|y - X\beta_X - C\beta_C\|_2^2$, then setting the partial derivatives of all the parameters as zero leads to

$$\frac{\partial \mathbb{L}}{\partial \beta_X} = -\frac{2}{\mathsf{N}} X^\top (y - X\beta_X - C\beta_C) = 0,$$
$$\frac{\partial \mathbb{L}}{\partial \beta_C} = -\frac{2}{\mathsf{N}} C^\top (y - X\beta_X - C\beta_C) = 0,$$

which gives

$$X^\top X \hat{\beta}_X^{(ols)} = X^\top (y - C\hat{\beta}_C^{(ols)}),$$
$$C^\top C \hat{\beta}_C^{(ols)} = C^\top (y - X\hat{\beta}_X^{(ols)}).$$

This further implies

$$\begin{aligned}
\hat{\beta}_X^{(ols)} &= (X^\top (I_{\mathsf{N}} - P_C) X)^{-1} X^\top (I_{\mathsf{N}} - P_C) y \\
&= (X^\top (I_{\mathsf{N}} - P_C) X)^{-1} X^\top (I_{\mathsf{N}} - P_C)(X\beta_X + C\beta_C + \varepsilon) \\
&= \beta_X + (X^\top (I_{\mathsf{N}} - P_C) X)^{-1} X^\top (I_{\mathsf{N}} - P_C)(C\beta_C + \varepsilon),
\end{aligned}$$

and

$$\begin{aligned}\hat{\beta}_C^{(ols)} &= (C^\top(I_\mathsf{N} - P_X)C)^{-1}C^\top(I_\mathsf{N} - P_X)y \\ &= (C^\top(I_\mathsf{N} - P_X)C)^{-1}C^\top(I_\mathsf{N} - P_X)(X\beta_X + C\beta_C + \varepsilon) \\ &= \beta_C + (C^\top(I_\mathsf{N} - P_X)C)^{-1}C^\top(I_\mathsf{N} - P_X)(X\beta_X + \varepsilon).\end{aligned}$$

Note that the projection matrices P_C and P_X satisfy $(I_\mathsf{N} - P_C)C = 0$ and $(I_\mathsf{N} - P_X)X = 0$, we then have

$$\begin{aligned}\hat{\beta}_X^{(ols)} - \beta_X = (X^\top(I_\mathsf{N} - P_C)X)^{-1}X^\top(I_\mathsf{N} - P_C)\varepsilon, \\ \hat{\beta}_C^{(ols)} - \beta_C = (C^\top(I_\mathsf{N} - P_X)C)^{-1}C^\top(I_\mathsf{N} - P_X)\varepsilon.\end{aligned} \tag{S1.2}$$

Also, since $\mathbf{W}_1$ is column full rank, from the inverse formula for the partitioned matrix, we see that

$$\begin{aligned}(\mathbf{W}_1^\top \mathbf{W}_1)^{-1} &= \begin{pmatrix} X^\top X & X^\top C \\ C^\top X & C^\top C \end{pmatrix}^{-1} \\ &= \begin{pmatrix} (X^\top(I_\mathsf{N} - P_C)X)^{-1} & *_1 \\ *_2 & (C^\top(I_\mathsf{N} - P_X)C)^{-1} \end{pmatrix},\end{aligned}$$

where

$$\begin{aligned}*_1 &= -(X^\top(I_\mathsf{N} - P_C)X)^{-1}X^\top C(C^\top C)^{-1} \\ &= -(X^\top X)^{-1}X^\top C(C^\top(I_\mathsf{N} - P_X)C)^{-1},\end{aligned}$$

$$\begin{aligned}*_2 &= -(C^\top(I_\mathsf{N} - P_X)C)^{-1}C^\top X(X^\top X)^{-1} \\ &= -(C^\top C)^{-1}C^\top X(X^\top(I_\mathsf{N} - P_C)X)^{-1},\end{aligned}$$

and

$$(X^\top(I_\mathsf{N}-P_C)X)^{-1} =(X^\top X)^{-1} + (X^\top X)^{-1}X^\top C(C^\top(I_\mathsf{N}-P_X)C)^{-1}C^\top X(X^\top X)^{-1},$$

$$(C^\top(I_\mathsf{N}-P_X)C)^{-1} =(C^\top C)^{-1} + (C^\top C)^{-1}C^\top X(X^\top(I_\mathsf{N}-P_C)X)^{-1}X^\top C(C^\top C)^{-1}.$$

Here, both $X^\top(I_\mathsf{N}-P_C)X$ and $C^\top(I_\mathsf{N}-P_X)C$ are symmetric and positive definite.

Next, we consider the asymptotic behavior of

$$\begin{aligned}\hat{\beta}_X^{(ols)} - \beta_X &= (X^\top(I_\mathsf{N}-P_C)X)^{-1}X^\top(I_\mathsf{N}-P_C)\varepsilon \\ &= (\frac{1}{\mathsf{N}}X^\top(I_\mathsf{N}-P_C)X)^{-1}\frac{1}{\mathsf{N}}X^\top(I_\mathsf{N}-P_C)\varepsilon.\end{aligned}$$

We start by showing

$$\frac{1}{\mathsf{N}}X^\top P_C X \xrightarrow{L_1} 0. \tag{S1.3}$$

Note that the projection matrix P_C is idempotent and $\mathrm{rank}(P_C) = \mathrm{tr}(P_C) = \mathsf{L}$. Therefore, for a fixed C, there exists an orthogonal matrix $J = (J_{ij})_{i,j=1}^{\mathsf{N}}$ such that

$$JP_CJ^\top = \begin{bmatrix} I_\mathsf{L} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}_{\mathsf{N}\times\mathsf{N}}. \tag{S1.4}$$

Here we consider each individual element of $\frac{1}{\mathsf{N}}X^\top P_C X$. Denote $X = [X_1, \cdots, X_P]$, and we have

$$\frac{1}{\mathsf{N}}X^\top P_C X = \left(\frac{1}{\mathsf{N}}X_i^\top P_C X_j\right)_{i,j=1}^{\mathsf{P}}.$$

Denote the conditional expectation on C as $\mathbb{E}^C[\cdot] := \mathbb{E}[\cdot|C]$. By (S1.4), we have

$$
\begin{aligned}
\mathbb{E}^C\left[\left|\frac{1}{\mathsf{N}}X_i^\top P_C X_j\right|\right] &= \mathbb{E}^C\left[\left|\frac{1}{\mathsf{N}}X_i^\top J^\top \begin{bmatrix} I_{\mathsf{L}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} J X_j\right|\right] \\
&= \mathbb{E}^C\left[\left|\frac{1}{\mathsf{N}}\sum_{k=1}^{\mathsf{L}}\left(\sum_{l=1}^{\mathsf{N}} J_{kl}X_{li}\right)\left(\sum_{l=1}^{\mathsf{N}} J_{kl}X_{lj}\right)\right|\right] \\
&\leq \frac{1}{\mathsf{N}}\sum_{k=1}^{\mathsf{L}}\mathbb{E}^C\left[\left|\left(\sum_{l=1}^{\mathsf{N}} J_{kl}X_{li}\right)\left(\sum_{l=1}^{\mathsf{N}} J_{kl}X_{lj}\right)\right|\right] \\
&\leq \frac{1}{\mathsf{N}}\sum_{k=1}^{\mathsf{L}}\left(\mathbb{E}^C\left|\sum_{l=1}^{\mathsf{N}} J_{kl}X_{li}\right|^2\right)^{\frac{1}{2}}\left(\mathbb{E}^C\left|\sum_{l=1}^{\mathsf{N}} J_{kl}X_{lj}\right|^2\right)^{\frac{1}{2}},
\end{aligned}
\tag{S1.5}
$$

where the last inequality follows from the Cauchy-Schwartz inequality.

Since J is orthogonal, we have $\sum_{l=1}^{\mathsf{L}} J_{kl}^2 = 1$, for $k \in \{1, \cdots, \mathsf{N}\}$. Since $\mathbb{E}[X_{ij}|C, E_0] = 0$ and $\mathbb{E}[X_{ij}^2|C, E_0] < \infty$ for $1 \leq i \leq \mathsf{N}, 1 \leq j \leq \mathsf{P}$, we obtain

$$
\begin{aligned}
\mathbb{E}^C\left|\sum_{l=1}^{\mathsf{N}} J_{kl}X_{li}\right|^2 &= \mathbb{E}^C\left(\sum_{l=1}^{\mathsf{N}} J_{kl}^2 X_{li}^2 + \sum_{l\neq l'} J_{kl}X_{li}J_{kl'}X_{l'i}\right) \\
&= \sum_{l=1}^{\mathsf{N}} J_{kl}^2\mathbb{E}^C X_{li}^2 + \sum_{l\neq l'} J_{kl}J_{kl'}\mathbb{E}^C\left[X_{li}X_{l'i}\right] \\
&= \mathbb{E}^C X_{1i}^2 + \sum_{l\neq l'} J_{kl}J_{kl'}\cdot 0 \\
&= \mathbb{E}^C X_{1i}^2 < \infty,
\end{aligned}
$$

as X_{li} and $X_{l'i}$ are independent for $l \neq l'$. Therefore, by (S1.5), we have as

$\mathsf{N} \longrightarrow \infty$, $\mathsf{L}/\mathsf{N} \to 0$ and

$$\begin{aligned}\mathbb{E}^C\left[\left|\frac{1}{\mathsf{N}}X_i^\top P_C X_j\right|\right] &\le \frac{1}{\mathsf{N}}\sum_{k=1}^{\mathsf{L}}\sqrt{\mathbb{E}^C X_{1i}^2 \mathbb{E}^C X_{1j}^2} \\ &= \frac{\mathsf{L}}{\mathsf{N}}\sqrt{\mathbb{E}^C X_{1i}^2 \mathbb{E}^C X_{1j}^2} \longrightarrow 0, \forall i,j \in \{1,\cdots,\mathsf{P}\}, \quad \text{(S1.6)}\end{aligned}$$

which further gives

$$\begin{aligned}\mathbb{E}\left[\left|\frac{1}{\mathsf{N}}X_i^\top P_C X_j\right|\right] &= \mathbb{E}\left[\mathbb{E}^C\left[\left|\frac{1}{\mathsf{N}}X_i^\top P_C X_j\right|\right]\right] \\ &\le \mathbb{E}\left[\frac{\mathsf{L}}{\mathsf{N}}\sqrt{\mathbb{E}^C X_{1i}^2 \mathbb{E}^C X_{1j}^2}\right] \\ &= \frac{\mathsf{L}}{\mathsf{N}}\sqrt{\mathbb{E}^C X_{1i}^2 \mathbb{E}^C X_{1j}^2} \longrightarrow 0, \quad \forall i,j \in \{1,\cdots,\mathsf{P}\},\end{aligned}$$

and proves (S1.3). Also, (S1.3) implies

$$\frac{1}{\mathsf{N}}X^\top P_C X \xrightarrow{P} 0. \quad \text{(S1.7)}$$

By the law of large numbers, we have

$$\frac{1}{\mathsf{N}}X^\top X \xrightarrow{P} V_X, \quad \text{(S1.8)}$$

where V_X is a deterministic and nonsingular diagonal matrix. From (S1.7) and (S1.8) we also obtain

$$\frac{1}{\mathsf{N}}X^\top (I - P_C) X \xrightarrow{P} V_X. \quad \text{(S1.9)}$$

Applying the continuous mapping theorem to (S1.9) we conclude

$$(\frac{1}{\mathsf{N}}X^\top (I - P_C) X)^{-1} \xrightarrow{P} V_X^{-1}. \quad \text{(S1.10)}$$

Now we consider the asymptotic normality of

$$\sqrt{\mathsf{N}}(\hat{\beta}_X^{(ols)} - \beta_X) = (\frac{1}{\mathsf{N}} X^\top (I_\mathsf{N} - P_C) X)^{-1} \frac{1}{\sqrt{\mathsf{N}}} X^\top (I_\mathsf{N} - P_C)\varepsilon.$$

By (S1.3), we have

$$\mathbb{E}\left[\left\| \frac{1}{\sqrt{\mathsf{N}}} X^\top P_C \varepsilon \right\|_2^2\right] = \frac{\sigma_y^2}{\mathsf{N}} \mathrm{tr}\left(\mathbb{E}\left[X^\top P_C X\right]\right) \to 0,$$

which $\frac{1}{\sqrt{\mathsf{N}}} X^\top P_C \varepsilon \xrightarrow{P} 0$. Also, for $\frac{1}{\sqrt{\mathsf{N}}} X^\top \varepsilon$, the central limit theorem gives that

$$\frac{1}{\sqrt{\mathsf{N}}} X^\top \varepsilon \xrightarrow{d} \mathcal{N}(0, \sigma_y^2 V_X).$$

Hence, we arrive at the asymptotic normality result:

$$\sqrt{\mathsf{N}}(\hat{\beta}_X^{(ols)} - \beta_X) \xrightarrow{d} \mathcal{N}(0, \sigma_y^2 V_X^{-1}).$$

(ii) To prove the consistency of $\hat{\beta}_C^{(ols)} - \beta_C$, we first point out that

$$\hat{\beta}_C^{(ols)} - \beta_C = (C^\top (I_\mathsf{N} - P_X) C)^{-1} C^\top (I_\mathsf{N} - P_X)\varepsilon,$$

and examine the ℓ_2-norm of $\hat{\beta}_C^{(ols)} - \beta_C$ as follows:

$$\begin{aligned}
\mathbb{E}\left[\left\| \hat{\beta}_C^{(ols)} - \beta_C \right\|_2^2\right] &= \mathbb{E}\left[\varepsilon^\top (I_\mathsf{N} - P_X) C (C^\top (I_\mathsf{N} - P_X) C)^{-2} C^\top (I_\mathsf{N} - P_X)\varepsilon\right] \\
&= \mathbb{E}\left[\mathrm{tr}\left(\varepsilon^\top (I_\mathsf{N} - P_X) C (C^\top (I_\mathsf{N} - P_X) C)^{-2} C^\top (I_\mathsf{N} - P_X)\varepsilon\right)\right] \\
&= \mathrm{tr}\left(\mathbb{E}\left[\varepsilon\varepsilon^\top (I_\mathsf{N} - P_X) C (C^\top (I_\mathsf{N} - P_X) C)^{-2} C^\top (I_\mathsf{N} - P_X)\right]\right) \\
&= \sigma_y^2 \mathrm{tr}\left(\mathbb{E}\left[(I_\mathsf{N} - P_X) C (C^\top (I_\mathsf{N} - P_X) C)^{-2} C^\top (I_\mathsf{N} - P_X)\right]\right)
\end{aligned}$$

$$= \sigma_y^2 \mathbb{E}\left[\operatorname{tr}\left((I_\mathsf{N} - P_X)C(C^\top(I_\mathsf{N} - P_X)C)^{-2}C^\top(I_\mathsf{N} - P_X)\right)\right]$$

$$= \sigma_y^2 \mathbb{E}\left[\operatorname{tr}\left(C^\top(I_\mathsf{N} - P_X)C(C^\top(I_\mathsf{N} - P_X)C)^{-2}\right)\right]$$

$$= \sigma_y^2 \mathbb{E}\left[\operatorname{tr}\left((C^\top(I_\mathsf{N} - P_X)C)^{-1}\right)\right].$$

Note that $C^\top(I_\mathsf{N} - P_X)C$ is positive definite, and we denote its eigenvalues as $\mu_1 \geq \cdots \geq \mu_\mathsf{L} > 0$. Then we have

$$\operatorname{tr}\left((C^\top(I_\mathsf{N} - P_X)C)^{-1}\right) = \sum_{i=1}^{\mathsf{L}} \frac{1}{\mu_i} \leq \frac{\mathsf{L}}{\mu_\mathsf{L}} = \frac{\mathsf{L}}{a_\mathsf{N}^2 \sigma_{\min}^2((I_\mathsf{N} - P_X)V)} \leq \frac{\mathsf{L}}{a_\mathsf{N}^2 l_\mathsf{N}^2}.$$

Thus, as $\mathsf{N} \to \infty$,

$$\mathbb{E}\left[\left\|\hat{\beta}_C^{(ols)} - \beta_C\right\|_2^2\right] = \sigma_y^2 \mathbb{E}\left[\operatorname{tr}\left((C^\top(I_\mathsf{N} - P_X)C)^{-1}\right)\right] \leq \sigma_y^2 \mathbb{E}\left[\frac{\mathsf{L}}{a_\mathsf{N}^2 l_\mathsf{N}^2}\right] = \frac{\sigma_y^2 \mathsf{L}}{a_\mathsf{N}^2 l_\mathsf{N}^2} \to 0,$$

thereby verifying the consistency of $\hat{\beta}_C^{(ols)}$. □

S1.4 Proof of Theorem 2:

Similar to the calculation of (S1.2), we have

$$\tilde{\beta}_X^{(ols)} - \beta_X = (X^\top(I_\mathsf{N} - P_Z)X)^{-1}X^\top(I_\mathsf{N} - P_Z)\varepsilon,$$

$$\tilde{\beta}_Z^{(ols)} - \beta_Z = (Z^\top(I_\mathsf{N} - P_X)Z)^{-1}Z^\top(I_\mathsf{N} - P_X)\varepsilon. \tag{S1.11}$$

Applying a similar proof strategy to $\tilde{\beta}_X^{(ols)}$ gives its consistency and asymptotic normality. Thus, we only need to consider $\tilde{\beta}_Z^{(ols)}$, and divide the proof into three steps:

1. Show that $\frac{1}{\mathsf{N}} Z^\top P_X Z \xrightarrow{P} 0$.

2. For $a_{\mathsf{N}} = \sqrt{\mathsf{NL}}$, show that there exists a constant $m > 0$ such that $\frac{1}{\mathsf{N}} \|Z\|_2^2 \geq m$ a.s..

3. Show that $\frac{1}{\mathsf{N}} Z^\top (I_{\mathsf{N}} - P_X)\varepsilon \xrightarrow{P} 0$.

With the above three steps, we conclude that as $\mathsf{N} \to \infty$

$$
\begin{aligned}
&\left|\tilde{\beta}_Z^{(ols)} - \beta_Z\right| \\
&= \left|(Z^\top (I - P_X) Z)^{-1} Z^\top (I_{\mathsf{N}} - P_X)\varepsilon\right| \\
&= \left(\frac{1}{\mathsf{N}} Z^\top (I - P_X) Z\right)^{-1} \left|\frac{1}{\mathsf{N}} Z^\top (I_{\mathsf{N}} - P_X)\varepsilon\right| \\
&\leq \left(m - \frac{1}{\mathsf{N}} Z^\top P_X Z\right)^{-1} \left|\frac{1}{\mathsf{N}} Z^\top (I_{\mathsf{N}} - P_X)\varepsilon\right| \\
&\xrightarrow{P} m^{-1} \cdot 0 = 0,
\end{aligned}
$$

showing the consistency of $\tilde{\beta}_Z^{(ols)}$.

Step 1: We start the proof by showing $\frac{1}{\mathsf{N}} Z^\top P_X Z \xrightarrow{P} 0$. Note that

$$
\frac{1}{\mathsf{N}} Z^\top P_X Z = \frac{1}{\mathsf{N}} Z^\top X \left(\frac{1}{\mathsf{N}} X^\top X\right)^{-1} \frac{1}{\mathsf{N}} X^\top Z,
$$

where

$$
\frac{1}{\mathsf{N}} X^\top Z = \begin{bmatrix} \frac{1}{\mathsf{N}} \sum_{i=1}^{\mathsf{N}} X_{i1} Z_i \\ \vdots \\ \frac{1}{\mathsf{N}} \sum_{i=1}^{\mathsf{N}} X_{i\mathsf{P}} Z_i \end{bmatrix}.
$$

Recall from the definition of Z (2.10) that Z represents the estimated community-based centrality of nodes, and nodes within the same community share the same value for the centrality measure Z. Here we rewrite Z corresponding to R communities by $\{Z^{(1)}, \cdots, Z^{(\mathsf{R})}\}$ i.e. $Z = [Z^{(c_1)}, \cdots, Z^{(c_\mathsf{N})}]^\top$ where $Z^{(c_i)}$ denotes the centrality of community c_i with $c_i \in \{1, \cdots, \mathsf{R}\}$ being the community label of node i. From the definitions of Z and U, we observe that

$$\mathsf{L}\sum_{k=1}^{\mathsf{N}} Z_k = \mathsf{L}\sum_{r=1}^{\mathsf{R}} \mathsf{N}_r Z^{(r)} = \sum_{i,j} C_{ij}, \tag{S1.12}$$

which further gives

$$\mathsf{N}_r Z^{(r)} \leq \sum_{r=1}^{\mathsf{R}} \mathsf{N}_r Z^{(r)} = \frac{1}{\mathsf{L}}\sum_{i,j} C_{ij} \leq \frac{1}{\mathsf{L}}\sqrt{\mathsf{NL}\sum_{i,j} C_{ij}^2} = \frac{\sqrt{\mathsf{N}a_\mathsf{N}^2}}{\sqrt{\mathsf{L}}} = \mathsf{N}. \tag{S1.13}$$

From Assumption 4, we have $\min_i \frac{\mathsf{N}_i}{\mathsf{N}} > \epsilon$, thus

$$Z^{(r)} \leq \frac{\mathsf{N}}{\mathsf{N}_r} \leq \max_r \frac{\mathsf{N}}{\mathsf{N}_r} < \frac{1}{\epsilon}. \tag{S1.14}$$

Now we prove that $\frac{1}{\mathsf{N}}X^\top Z \xrightarrow{P} 0$. From (S1.14), entrywisely we have

$$\begin{aligned}
\left|\frac{1}{\mathsf{N}}\sum_{i=1}^{\mathsf{N}} X_{ij} Z_i\right| &= \left|\frac{1}{\mathsf{N}}\sum_{r=1}^{\mathsf{R}} Z^{(r)} \sum_{i=1}^{\mathsf{N}_r} X_{1i}\right| \\
&\leq \sum_{r=1}^{\mathsf{R}} Z^{(r)} \left|\frac{1}{\mathsf{N}}\sum_{i=1}^{\mathsf{N}_r} X_{ij}\right| \\
&\leq \sum_{r=1}^{\mathsf{R}} \frac{1}{\epsilon} \left|\frac{1}{\mathsf{N}}\sum_{i=1}^{\mathsf{N}_r} X_{ij}\right| \xrightarrow{P} 0
\end{aligned}$$

if $\frac{1}{\mathsf{N}}\sum_{i=1}^{\mathsf{N}_r} X_{ij} \xrightarrow{P} 0$. So now we only need to prove $\frac{1}{\mathsf{N}}\sum_{i=1}^{\mathsf{N}_r} X_{ij} \xrightarrow{P} 0$. Consider the second moment of $\frac{1}{\mathsf{N}}\sum_{i=1}^{\mathsf{N}_r} X_{ij}$:

$$\begin{aligned}\mathbb{E}\left[\left(\frac{1}{\mathsf{N}}\sum_{i=1}^{\mathsf{N}_r} X_{ij}\right)^2\right] &= \mathbb{E}\left[\frac{1}{\mathsf{N}^2}\mathbb{E}^{\mathsf{N}_r}\left[\left(\sum_{i=1}^{\mathsf{N}_r} X_{ij}\right)^2\right]\right] \\ &= \frac{1}{\mathsf{N}^2}\mathbb{E}\left[\mathsf{N}_r \mathbb{E}X_{ij}^2\right] \\ &= \frac{1}{\mathsf{N}}\mathbb{E}\left[\frac{\mathsf{N}_r}{\mathsf{N}}\mathbb{E}X_{ij}^2\right] \le \frac{1}{\mathsf{N}}\mathbb{E}X_{ij}^2 \to 0\end{aligned}$$

where $\frac{\mathsf{N}_r}{\mathsf{N}} \le 1$. Thus $\frac{1}{\mathsf{N}}\sum_{i=1}^{\mathsf{N}_r} X_{ij} \xrightarrow{L_2} 0$ and $\frac{1}{\mathsf{N}}\sum_{i=1}^{\mathsf{N}_r} X_{ij} \xrightarrow{P} 0$, which implies

$$\frac{1}{\mathsf{N}} X^\top Z \xrightarrow{P} 0 \tag{S1.15}$$

and

$$\frac{1}{\mathsf{N}} Z^\top P_X Z \xrightarrow{P} 0 \cdot V_X^{-1} \cdot 0 = 0. \tag{S1.16}$$

Step 2: Now we consider $\frac{1}{\mathsf{N}}Z^\top Z$. Since $C_{ij} > 0$, with Cauchy-Schwarz inequality, we have

$$\frac{1}{\mathsf{N}}\|Z\|_2^2 = \frac{1}{\mathsf{N}^2}\sum_{k=1}^{\mathsf{N}} Z_k^2 \sum_{k=1}^{\mathsf{N}} 1 \ge \frac{1}{\mathsf{N}^2}\left(\sum_{k=1}^{\mathsf{N}} Z_k\right)^2.$$

Combining with equation (S1.12) and Assumption 4, we see that

$$\frac{1}{\mathsf{N}}\|Z\|_2^2 \ge \frac{1}{\mathsf{N}^2}\left(\frac{1}{\mathsf{L}}\sum_{i,j} C_{ij}\right)^2 \ge \frac{1}{\mathsf{N}^2}\frac{1}{\mathsf{L}^2}(\mathsf{N}\min_{1\le i\le \mathsf{N}}\sum_{j=1}^{\mathsf{L}} C_{ij})^2 = \frac{1}{\mathsf{L}^2}\min_{1\le i\le \mathsf{N}}\|C_i\|_1^2 \asymp \frac{a_{\mathsf{N}}^2}{\mathsf{NL}} \asymp 1.$$

Hence, with $a_{\mathsf{N}} \asymp \sqrt{\mathsf{NL}}$, there exists a constant $m > 0$ such that

$$\frac{1}{\mathsf{N}}\|Z\|_2^2 \ge m, \quad a.s.. \tag{S1.17}$$

From the results of **Step 1** and **Step 2**, by (S1.16) and (S1.17), we see that for N sufficiently large,

$$\begin{aligned}\frac{1}{\mathsf{N}}Z^\top(I_\mathsf{N} - P_X)Z &= \frac{1}{\mathsf{N}}Z^\top Z - \frac{1}{\mathsf{N}}Z^\top P_X Z \\ &\geq m - \frac{1}{\mathsf{N}}Z^\top P_X Z > 0,\end{aligned}$$

which gives

$$\left(\frac{1}{\mathsf{N}}Z^\top(I_\mathsf{N} - P_X)Z\right)^{-1} \leq \left(m - \frac{1}{\mathsf{N}}Z^\top P_X Z\right)^{-1}. \tag{S1.18}$$

Step 3: Now we consider the behavior of

$$\frac{1}{\mathsf{N}}Z^\top(I_\mathsf{N} - P_X)\varepsilon = \frac{1}{\mathsf{N}}Z^\top\varepsilon - \frac{1}{\mathsf{N}}Z^\top P_X\varepsilon. \tag{S1.19}$$

For the first part of RHS of (S1.19), $\frac{1}{\mathsf{N}}Z^\top\varepsilon \xrightarrow{P} 0$ follows from the arguments showing $\frac{1}{\mathsf{N}}X^\top Z \xrightarrow{P} 0$. Moreover, we have $\frac{1}{\mathsf{N}}X^\top Z \xrightarrow{L_2} 0$:

$$\begin{aligned}\mathbb{E}\left[\left\|\frac{1}{\mathsf{N}}Z^\top\varepsilon\right\|_2^2\right] &= \frac{1}{\mathsf{N}^2}\mathbb{E}\left[\varepsilon^\top ZZ^\top\varepsilon\right] \\ &= \frac{\sigma_y^2}{\mathsf{N}^2}\mathbb{E}\left[\|Z\|_2^2\right].\end{aligned} \tag{S1.20}$$

And then we need to calculate $\mathbb{E}\left[\|Z\|_2^2\right]$. The randomness of Z arises from both eigenvector centrality and community structure. Therefore, we need to consider $\|Z\|_2^2$ from a different perspective here. From the definition of Z in (2.10), we have

$$\|Z\|_2^2 = \frac{1}{\mathsf{L}^2}\|U\,1_\mathsf{L}\,\|_2^2 \leq \frac{1}{\mathsf{L}}\|U\|_F^2.$$

By the definition of U in (2.9), we have

$$\begin{aligned}\|U\|_F^2 &= \left\|S(S^\top S)^{-1}S^\top C\right\|_F^2 \\ &\leq \left\|S(S^\top S)^{-1}S^\top\right\|_F^2 \|C\|_F^2 \\ &= \mathrm{tr}(S(S^\top S)^{-1}S^\top)\|C\|_F^2 \\ &= \mathrm{tr}(I_{\mathsf{R}})\|C\|_F^2 = \mathsf{R}\|C\|_F^2.\end{aligned}$$

Hence, with Assumption 4, we have the following upper bound for $\|Z\|_2^2$:

$$\|Z\|_2^2 \leq \frac{1}{\mathsf{L}}\|U\|_F^2 \leq \frac{\mathsf{R}}{\mathsf{L}}\|C\|_F^2. \tag{S1.21}$$

Then taking expectations on both sides of (S1.21) gives

$$\mathbb{E}\left[\|Z\|_2^2\right] \leq \frac{\mathsf{R}}{\mathsf{L}}\mathbb{E}\left[\|C\|_F^2\right] = O\left(\frac{a_{\mathsf{N}}^2}{\mathsf{L}}\right), \tag{S1.22}$$

and with (S1.22), we have as $\mathsf{N} \to \infty$

$$\mathbb{E}\left[\left\|\frac{1}{\mathsf{N}}Z^\top\varepsilon\right\|_2^2\right] = \frac{\sigma_y^2}{\mathsf{N}^2}\mathbb{E}\left[\|Z\|_2^2\right] \leq \frac{\sigma_y^2}{\mathsf{N}^2}O\left(\frac{a_{\mathsf{N}}^2}{\mathsf{L}}\right) \to 0, \tag{S1.23}$$

i.e. $\frac{1}{\mathsf{N}}Z^\top\varepsilon \xrightarrow{L_2} 0$.

For the second part of RHS of (S1.19), using (S1.15) and the law of large numbers, we have

$$\frac{1}{\mathsf{N}}Z^\top P_X\varepsilon = \frac{1}{\mathsf{N}}Z^\top X\left(\frac{1}{\mathsf{N}}X^\top X\right)^{-1}\frac{1}{\mathsf{N}}X^\top\varepsilon \xrightarrow{P} 0 \cdot V_X^{-1} \cdot 0 = 0, \tag{S1.24}$$

so that $\frac{1}{\mathsf{N}}Z^\top P_X\varepsilon \xrightarrow{P} 0$.

Therefore, combining (S1.23), and (S1.24), we conclude that $\frac{1}{\mathsf{N}}Z^\top(I_{\mathsf{N}} - P_X)\varepsilon \xrightarrow{P} 0$ and completes the proof of Step 3.

Now we turn to the asymptotic normality of $\tilde{\beta}_Z^{(ols)}$.

The OLS estimator is:

$$\tilde{\beta}_Z^{(ols)} - \beta_Z = \left(Z^\top (I_\mathsf{N} - P_X) Z\right)^{-1} Z^\top (I_\mathsf{N} - P_X)\varepsilon.$$

Normalizing by $\sqrt{Z^\top Z/\sigma_y^2}$, we have

$$\sqrt{\frac{Z^\top Z}{\sigma_y^2}}\left(\tilde{\beta}_Z^{(ols)} - \beta_Z\right) = \underbrace{\frac{Z^\top Z}{\mathsf{N}}\left(\frac{Z^\top (I_\mathsf{N} - P_X) Z}{\mathsf{N}}\right)^{-1}}_{(A)} \cdot \underbrace{\frac{Z^\top \varepsilon}{\sqrt{Z^\top Z \sigma_y^2}}}_{(B)} - \frac{Z^\top Z}{\mathsf{N}}\left(\frac{Z^\top (I_\mathsf{N} - P_X) Z}{\mathsf{N}}\right)^{-1} \cdot \underbrace{\frac{Z^\top P_X \varepsilon}{\sqrt{Z^\top Z \sigma_y^2}}}_{(C)}.$$

We only need to prove that (A) $\xrightarrow{P} 1$, (B) $\xrightarrow{d} \mathcal{N}(0,1)$ and (C) $\xrightarrow{d} 0$. For the term (A), with (S1.16), we obtain

$$(A) = \frac{Z^\top (I_\mathsf{N} - P_X) Z}{\mathsf{N}}\left(\frac{Z^\top Z}{\mathsf{N}}\right)^{-1} \xrightarrow{P} 1.$$

The term (B) is specified by

$$(B) = \frac{Z^\top \varepsilon}{\sqrt{Z^\top Z \sigma_y^2}} = \sum_{i=1}^{\mathsf{N}} \frac{Z_i \varepsilon_i}{\sqrt{Z^\top Z \sigma_y^2}}.$$

Define $\mathcal{W}_{\mathsf{N},i} = \frac{Z_i \varepsilon_i}{\sqrt{Z^\top Z \sigma_y^2}}$, then:

$$\mathbb{E}\left[\mathcal{W}_{\mathsf{N},i} \,|\, Z\right] = 0, \quad \sum_{i=1}^{\mathsf{N}} \mathbb{E}\left[\mathcal{W}_{\mathsf{N},i}^2 \,|\, Z\right] = \frac{\sum_{i=1}^{\mathsf{N}} Z_i^2 \sigma_y^2}{Z^\top Z \sigma_y^2} = 1.$$

To apply the conditional central limit theorem (Yuan et al., 2014), we need to show

$$\forall \epsilon > 0, \quad \sum_{i=1}^{\mathsf{N}} \mathbb{E}\left[\mathcal{W}_{\mathsf{N},i}^2 1_{\{|\mathcal{W}_{\mathsf{N},i}|>\epsilon\}} \,|\, Z\right] \xrightarrow{a.s.} 0.$$

For any $\epsilon_0 > 0$,

$$1_{\{|\mathcal{W}_{\mathsf{N},i}|>\epsilon_0\}} \leq \frac{\mathcal{W}_{\mathsf{N},i}^2}{\epsilon_0^2}.$$

Thus,

$$\mathbb{E}\left[\mathcal{W}_{\mathsf{N},i}^2 1_{\{|\mathcal{W}_{\mathsf{N},i}|>\epsilon_0\}} \,|\, Z\right] \leq \frac{\mathbb{E}\left[\mathcal{W}_{\mathsf{N},i}^4 \,|\, Z\right]}{\epsilon_0^2}.$$

From Assumption 1, $\mathbb{E}[\varepsilon_i^4] \leq k_0$, then

$$\mathbb{E}\left[\mathcal{W}_{\mathsf{N},i}^4 \,|\, Z\right] = \frac{Z_i^4 \mathbb{E}[\varepsilon_i^4]}{(Z^\top Z \sigma_y^2)^2} \leq \frac{k_0 Z_i^4}{(Z^\top Z \sigma_y^2)^2}.$$

Sum over i:

$$\sum_{i=1}^{\mathsf{N}} \mathbb{E}\left[\mathcal{W}_{\mathsf{N},i}^4 \,|\, Z\right] \leq \frac{k_0 \sum_{i=1}^{\mathsf{N}} Z_i^4}{(Z^\top Z \sigma_y^2)^2}.$$

Since

$$\sum_{i=1}^{\mathsf{N}} Z_i^4 \leq \left(\max_{1\leq i\leq \mathsf{N}} Z_i^2\right) \sum_{i=1}^{\mathsf{N}} Z_i^2 = \left(\max_{1\leq i\leq \mathsf{N}} Z_i^2\right) Z^\top Z,$$

we have

$$\frac{k_0 \sum_{i=1}^{\mathsf{N}} Z_i^4}{(Z^\top Z \sigma_y^2)^2} \leq \frac{k_0 \left(\max_{1\leq i\leq \mathsf{N}} Z_i^2\right)}{Z^\top Z \sigma_y^4} = \frac{k_0}{\sigma_y^4} \frac{\left(\max_{1\leq i\leq \mathsf{N}} Z_i^2\right)}{Z^\top Z}.$$

From Assumption 4, for each community r, $\mathsf{N}_r \geq \epsilon \mathsf{N}$. The condition $Z^\top Z >$ $\mathsf{N}_r Z_i^2$ implies:

$$Z_i^2 < \frac{Z^\top Z}{\mathsf{N}_r} \quad \forall i \in \text{community } r.$$

Thus, there exists r_0 such that $\max_{1\le i\le \mathsf{N}} Z_i^2 < \frac{Z^\top Z}{\mathsf{N}_{r_0}}$. Further,

$$\frac{k_0 \sum_{i=1}^{\mathsf{N}} Z_i^4}{(Z^\top Z \sigma_y^2)^2} \le \frac{k_0}{\sigma_y^4} \frac{(\max_{1\le i\le \mathsf{N}} Z_i^2)}{Z^\top Z} < \frac{k_0}{\sigma_y^4}\frac{1}{\mathsf{N}_{r_0}} \le \frac{k_0}{\sigma_y^4}\frac{1}{\epsilon \mathsf{N}} \xrightarrow{a.s.} 0.$$

Finally, we have

$$\sum_{i=1}^{\mathsf{N}} \mathbb{E}\left[\mathcal{W}_{\mathsf{N},i}^2 1_{\{|\mathcal{W}_{\mathsf{N},i}|>\epsilon_0\}} \,|\, Z\right] \le \frac{k_0}{\sigma_y^4}\frac{1}{\epsilon \mathsf{N}} \xrightarrow{a.s.} 0,$$

which implies that the Lindeberg condition holds almost surely given Z. By the conditional Lindeberg-Feller CLT, we obtain the convergence of the conditional distribution over Z, i.e.,

$$\forall t \in \mathbb{R}, \quad \mathbb{P}\left(\frac{Z^\top \varepsilon}{\sqrt{Z^\top Z \sigma_y^2}} \le t \,\middle|\, Z \right) \xrightarrow{P} \Phi(t).$$

Here, convergence in probability of the random probability measures means convergence in probability in the space $PM(\mathbb{R})$ of probability measures on $\mathbb{R}$ metrized by weak convergence. This implies that for almost every realization of Z, the conditional distribution converges to the standard normal distribution as $\mathsf{N} \to \infty$.

To extend this to the unconditional distribution, we integrate over Z:

$$\mathbb{P}\left(\frac{Z^\top \varepsilon}{\sqrt{Z^\top Z \sigma_y^2}} \in A \right) = \mathbb{E}\left[\mathbb{P}\left(\frac{Z^\top \varepsilon}{\sqrt{Z^\top Z \sigma_y^2}} \in A \,\middle|\, Z \right)\right],$$

for any measurable set A. Since the conditional probability $\mathbb{P}(\cdot|Z)$ is bounded by 1 (i.e., $0 \le \mathbb{P}(\cdot|Z) \le 1$ for all Z), the Dominated Convergence Theorem

(DCT) justifies interchanging the limit and expectation:

$$\lim_{\mathsf{N}\to\infty} \mathbb{P}\left(\frac{Z^\top \varepsilon}{\sqrt{Z^\top Z \sigma_y^2}} \in A\right) = \mathbb{E}\left[\lim_{\mathsf{N}\to\infty} \mathbb{P}\left(\frac{Z^\top \varepsilon}{\sqrt{Z^\top Z \sigma_y^2}} \in A \,\middle|\, Z\right)\right] = \Phi(A).$$

Thus, the unconditional distribution converges weakly:

$$\frac{Z^\top \varepsilon}{\sqrt{Z^\top Z \sigma_y^2}} \xrightarrow{d} \mathcal{N}(0,1). \tag{S1.25}$$

For the term (C), we have

$$\frac{Z^\top P_X \varepsilon}{\sqrt{Z^\top Z \sigma_y^2}} = \frac{Z^\top X (X^\top X)^{-1} X^\top \varepsilon}{\sqrt{Z^\top Z \sigma_y^2}} = \frac{Z^\top X}{\sqrt{Z^\top Z \mathsf{N}}} \sqrt{\frac{\mathsf{N}}{\sigma_y^2}} (X^\top X)^{-1} X^\top \varepsilon$$

where combining (S1.17) and $\frac{Z^\top X}{\mathsf{N}} \xrightarrow{P} 0$, we have

$$\frac{Z^\top X}{\sqrt{Z^\top Z \mathsf{N}}} = \sqrt{\frac{\mathsf{N}}{Z^\top Z}} \frac{Z^\top X}{\mathsf{N}} \xrightarrow{P} 0.$$

Moreover, from the central limit theorem, we obtain

$$\sqrt{\frac{\mathsf{N}}{\sigma_y^2}} (X^\top X)^{-1} X^\top \varepsilon \xrightarrow{d} \mathcal{N}(0, V_X^{-1}).$$

By Slutsky's Theorem, we finally obtain term (C) converge to 0 in distribution.

Combine the results of terms (A), (B) and (C) above, we get

$$\sqrt{\frac{Z^\top Z}{\sigma_y^2}} \left(\tilde{\beta}_Z^{(ols)} - \beta_Z\right) \xrightarrow{d} \mathcal{N}(0,1).$$

□

S1.5 Proof of Theorem 3

We start from the two-stage estimator defined by regressing y on the augmented regressor matrix $\hat{\mathbf{W}}_1 = (X, \hat{C})$. When $\hat{\mathbf{W}}_1$ is not of full column rank due to measurement errors in $\hat{C}$, the usual inverse $(\hat{\mathbf{W}}_1^\top \hat{\mathbf{W}}_1)^{-1}$ does not exist.

Instead, we use the Moore–Penrose pseudoinverse to define the projection matrix onto the column space of $\hat{C}$:

$$P_{\hat{C}} := \hat{C}(\hat{C}^\top \hat{C})^+ \hat{C}^\top,$$

which is always well-defined regardless of the rank of $\hat{C}$.

Then, by applying the Frisch–Waugh–Lovell theorem(Lovell, 1963; Frisch and Waugh, 1933) the estimator for β_X can be expressed as the coefficient from regressing y on X after projecting out the effect of $\hat{C}$:

$$\hat{\beta}_X = (X^\top (I_\mathsf{N} - P_{\hat{C}})X)^{-1} X^\top (I_\mathsf{N} - P_{\hat{C}})y.$$

Since we assume $y = X\beta_X + C\beta_C + \varepsilon$, we have for $\delta_C := C - \hat{C}$,

$$\hat{\beta}_X = \beta_X + (X^\top (I_\mathsf{N} - P_{\hat{C}})X)^{-1} X^\top (I_\mathsf{N} - P_{\hat{C}})(\delta_C \beta_C + \varepsilon). \tag{S1.26}$$

We now prove the consistency of $\hat{\beta}_X$, and observe from (S1.26) that

$$\begin{aligned}\hat{\beta}_X - \beta_X &= (X^\top (I_\mathsf{N} - P_{\hat{C}})X)^{-1} X^\top (I_\mathsf{N} - P_{\hat{C}})[\delta_C \beta_C + \varepsilon] \\ &= (\frac{1}{\mathsf{N}} X^\top (I_\mathsf{N} - P_{\hat{C}})X)^{-1} \left[\frac{1}{\mathsf{N}} X^\top (I_\mathsf{N} - P_{\hat{C}})\delta_C \beta_C + \frac{1}{\mathsf{N}} X^\top (I_\mathsf{N} - P_{\hat{C}})\varepsilon\right].\end{aligned}$$

Hence, as long as we justify the three convergence results below:

1. $(\frac{1}{\mathsf{N}}X^\top(I-P_{\hat{C}})X)^{-1} \xrightarrow{P} V_X^{-1}$,

2. $\frac{1}{\mathsf{N}}X^\top(I_{\mathsf{N}}-P_{\hat{C}})\varepsilon \xrightarrow{P} 0$,

3. $\frac{1}{\mathsf{N}}X^\top(I_{\mathsf{N}}-P_{\hat{C}})\delta_C\beta_C \xrightarrow{P} 0$,

we are able to obtain the consistency of $\hat{\beta}_X$.

Step 1: Using a similar proof strategy as for (S1.3) gives

$$\frac{1}{\mathsf{N}}X^\top P_{\hat{C}}X \xrightarrow{L_1} 0, \tag{S1.27}$$

which combined with the law of large numbers leads to

$$\frac{1}{\mathsf{N}}X^\top(I-P_{\hat{C}})X \xrightarrow{P} V_X. \tag{S1.28}$$

Applying the continuous mapping theorem to (S1.28), we obtain

$$(\frac{1}{\mathsf{N}}X^\top(I-P_{\hat{C}})X)^{-1} \xrightarrow{P} V_X^{-1}. \tag{S1.29}$$

Step 2: Now we show that $\frac{1}{\mathsf{N}}X^\top(I-P_{\hat{C}})\varepsilon \xrightarrow{P} 0$. Consider the ℓ_2-norm:

$$\begin{aligned}
\mathbb{E}\left[\left\|\frac{1}{\mathsf{N}}X^\top P_{\hat{C}}\varepsilon\right\|_2^2\right] &= \frac{1}{\mathsf{N}^2}\mathbb{E}\left[\varepsilon^\top P_{\hat{C}}XX^\top P_{\hat{C}}\varepsilon\right] \\
&= \frac{1}{\mathsf{N}^2}\mathbb{E}\left[\mathrm{tr}(\varepsilon^\top P_{\hat{C}}XX^\top P_{\hat{C}}\varepsilon)\right] \\
&= \frac{1}{\mathsf{N}^2}\mathbb{E}\left[\mathrm{tr}(\varepsilon\varepsilon^\top P_{\hat{C}}XX^\top P_{\hat{C}})\right] \\
&= \frac{1}{\mathsf{N}^2}\sigma_y^2\mathbb{E}\left[\mathrm{tr}(P_{\hat{C}}XX^\top P_{\hat{C}})\right]
\end{aligned}$$

$$= \frac{\sigma_y^2}{\mathsf{N}^2}\mathbb{E}\left[\mathrm{tr}(X^\top P_{\hat{C}} X)\right]$$
$$= \frac{\sigma_y^2}{\mathsf{N}}\mathrm{tr}\left(\mathbb{E}\left[\frac{1}{\mathsf{N}}X^\top P_{\hat{C}} X\right]\right)$$
$$\leq \frac{\sigma_y^2}{\mathsf{N}}\mathrm{tr}\left(\mathbb{E}\left[\left|\frac{1}{\mathsf{N}}X^\top P_{\hat{C}} X\right|\right]\right) \to 0,$$

where the convergence is given by (S1.27) . Therefore, $\frac{1}{\mathsf{N}}X^\top P_{\hat{C}}\varepsilon \xrightarrow{L_2} 0$ and

$$\frac{1}{\mathsf{N}}X^\top P_{\hat{C}}\varepsilon \xrightarrow{P} 0. \tag{S1.30}$$

Then combining the law of large numbers with (S1.30) gives

$$\frac{1}{\mathsf{N}}X^\top (I_{\mathsf{N}} - P_{\hat{C}})\varepsilon \xrightarrow{P} 0. \tag{S1.31}$$

Step 3: For $\frac{1}{\mathsf{N}}X^\top(I_{\mathsf{N}} - P_{\hat{C}})\delta_C\beta_C$, we again consider its ℓ_2-norm:

$$\mathbb{E}\left[\left\|\frac{1}{\mathsf{N}}X^\top(I_{\mathsf{N}} - P_{\hat{C}})\delta_C\beta_C\right\|_2^2\right] = \frac{1}{\mathsf{N}^2}\mathbb{E}\left[\beta_C^\top\delta_C^\top(I_n - P_{\hat{C}})XX^\top(I_{\mathsf{N}} - P_{\hat{C}})\delta_C\beta_C\right]$$
$$= \frac{1}{\mathsf{N}^2}\mathbb{E}\left[\mathrm{tr}(\beta_C^\top\delta_C^\top(I_n - P_{\hat{C}})XX^\top(I_{\mathsf{N}} - P_{\hat{C}})\delta_C\beta_C)\right]$$
$$= \frac{1}{\mathsf{N}^2}\mathbb{E}\left[\mathrm{tr}(\beta_C\beta_C^\top\delta_C^\top(I_n - P_{\hat{C}})XX^\top(I_{\mathsf{N}} - P_{\hat{C}})\delta_C)\right]$$

and applying Lemma S1.3 gives

$$\leq \frac{1}{\mathsf{N}^2}\mathbb{E}\left[\mathrm{tr}(\beta_C\beta_C^\top)\mathrm{tr}(\delta_C^\top(I_n - P_{\hat{C}})XX^\top(I_{\mathsf{N}} - P_{\hat{C}})\delta_C)\right]$$
$$= \frac{1}{\mathsf{N}^2}\mathrm{tr}(\beta_C\beta_C^\top)\mathbb{E}\left[\mathrm{tr}(\delta_C^\top(I_n - P_{\hat{C}})XX^\top(I_{\mathsf{N}} - P_{\hat{C}})\delta_C)\right]$$
$$= \frac{1}{\mathsf{N}^2}\mathrm{tr}(\beta_C\beta_C^\top)\mathbb{E}\left[\mathrm{tr}(\delta_C\delta_C^\top(I_n - P_{\hat{C}})XX^\top(I_{\mathsf{N}} - P_{\hat{C}}))\right];$$

since $I_{\mathsf{N}} - P_{\hat{C}}$ is idempotent, applying Lemma S1.3 again leads to

$$\leq \frac{1}{\mathsf{N}^2}\mathrm{tr}(\beta_C\beta_C^\top)\mathbb{E}\left[\mathrm{tr}(\delta_C\delta_C^\top)\mathrm{tr}((I_n - P_{\hat{C}})XX^\top(I_{\mathsf{N}} - P_{\hat{C}}))\right]$$

$$
\begin{aligned}
&= \frac{1}{\mathsf{N}^2}\mathrm{tr}(\beta_C\beta_C^\top)\mathbb{E}\left[\|\delta_C\|_F^2\mathrm{tr}(X^\top(I_n - P_{\hat{C}})X)\right] \\
&= \mathrm{tr}(\beta_C\beta_C^\top)\left(\frac{1}{\mathsf{N}^2}\mathbb{E}\left[\|\delta_C\|_F^2\mathrm{tr}(X^\top X)\right] - \frac{1}{\mathsf{N}^2}\mathbb{E}\left[\|\delta_C\|_F^2\mathrm{tr}(X^\top P_{\hat{C}}X)\right]\right).
\end{aligned}
$$

Next, since $\mathbb{E}^{C,E_0}[\frac{1}{\mathsf{N}}X_i^\top X_i] < \infty$, for $i = 1, \cdots, \mathsf{P}$, we then have

$$
\begin{aligned}
\frac{1}{\mathsf{N}^2}\mathbb{E}\left[\|\delta_C\|_F^2\mathrm{tr}(X^\top X)\right] &= \frac{1}{\mathsf{N}}\mathbb{E}\left[\|\delta_C\|_F^2\mathbb{E}^{C,E_0}\left[\mathrm{tr}(\frac{1}{\mathsf{N}}X^\top X)\right]\right] \\
&= \frac{1}{\mathsf{N}}\mathbb{E}\left[\|\delta_C\|_F^2\mathrm{tr}\left(\mathbb{E}^{C,E_0}\left[\frac{1}{\mathsf{N}}X^\top X\right]\right)\right] \\
&= \frac{1}{\mathsf{N}}\mathbb{E}\left[\|\delta_C\|_F^2\sum_{i=1}^{\mathsf{P}}\mathbb{E}^{C,E_0}\left[\frac{1}{\mathsf{N}}X_i^\top X_i\right]\right].
\end{aligned}
$$

Also, we see from Lemma S1.2 that $\mathbb{E}\left[\|\delta_C\|_F^2\right] = O(\frac{a_{\mathsf{N}}^2\mathsf{NL}}{\delta^2})$, so

$$
\frac{1}{\mathsf{N}^2}\mathbb{E}\left[\|\delta_C\|_F^2\mathrm{tr}(X^\top X)\right] = \frac{1}{\mathsf{N}}O(\frac{a_{\mathsf{N}}^2\mathsf{NL}}{\delta^2}) = O(\frac{a_{\mathsf{N}}^2\mathsf{L}}{\delta^2}). \tag{S1.32}
$$

Similar to (S1.6), we have for $i, j = 1, \cdots, \mathsf{P}$,

$$
\mathbb{E}^{\hat{C}}\left[\left|\frac{1}{\mathsf{N}}X_i^\top P_{\hat{C}}X_j\right|\right] = \mathbb{E}^{C,E_0}\left[\left|\frac{1}{\mathsf{N}}X_i^\top P_{\hat{C}}X_j\right|\right] \le \frac{\mathsf{L}}{\mathsf{N}}\sqrt{\mathbb{E}^{C,E_0}X_{1i}^2\mathbb{E}^{C,E_0}X_{1j}^2} = O(\frac{\mathsf{L}}{\mathsf{N}}),
$$

then

$$
\begin{aligned}
\frac{1}{\mathsf{N}^2}\mathbb{E}\left[\|\delta_C\|_F^2\mathrm{tr}(X^\top P_{\hat{C}}X)\right] &= \frac{1}{\mathsf{N}^2}\mathbb{E}\left[\|\delta_C\|_F^2\mathbb{E}^{C,E_0}\left[\mathrm{tr}(X^\top P_{\hat{C}}X)\right]\right] \\
&= \mathbb{E}\left[\frac{1}{\mathsf{N}}\|\delta_C\|_F^2\mathbb{E}^{C,E_0}\left[\mathrm{tr}(\frac{1}{\mathsf{N}}X^\top P_{\hat{C}}X)\right]\right] \\
&= \mathbb{E}\left[\frac{1}{\mathsf{N}}\|\delta_C\|_F^2\mathbb{E}^{C,E_0}\left[\mathrm{tr}(\frac{1}{\mathsf{N}}X^\top P_{\hat{C}}X)\right]\right] \\
&= \mathbb{E}\left[\frac{1}{\mathsf{N}}\|\delta_C\|_F^2 O\left(\frac{\mathsf{L}}{\mathsf{N}}\right)\right] \\
&= \frac{1}{\mathsf{N}}\mathbb{E}\left[\|\delta_C\|_F^2\right]O\left(\frac{\mathsf{L}}{\mathsf{N}}\right).
\end{aligned}
$$

Additionally, we obtain from Lemma S1.2 that

$$\frac{1}{\mathsf{N}^2}\mathbb{E}\left[\|\delta_C\|_F^2 \mathrm{tr}(X^\top P_{\hat{C}} X)\right] = \frac{1}{\mathsf{N}}\mathbb{E}\left[\|\delta_C\|_F^2\right] O\left(\frac{\mathsf{L}}{\mathsf{N}}\right) = \frac{1}{\mathsf{N}}O(\frac{a_{\mathsf{N}}^2\mathsf{N}\mathsf{L}}{\delta^2})O\left(\frac{\mathsf{L}}{\mathsf{N}}\right) = O(\frac{a_{\mathsf{N}}^2\mathsf{L}^2}{\mathsf{N}\delta^2}). \tag{S1.33}$$

Combining (S1.32) and (S1.33), we have

$$\frac{1}{\mathsf{N}^2}\mathbb{E}\left[\|\delta_C\|_F^2 \mathrm{tr}(X^\top X)\right] - \frac{1}{\mathsf{N}^2}\mathbb{E}\left[\|\delta_C\|_F^2 \mathrm{tr}(X^\top P_{\hat{C}} X)\right] = O(\frac{a_{\mathsf{N}}^2\mathsf{L}}{\delta^2}) - O(\frac{a_{\mathsf{N}}^2\mathsf{L}}{\delta^2}\frac{\mathsf{L}}{\mathsf{N}}) = O(\frac{a_{\mathsf{N}}^2\mathsf{L}}{\delta^2}).$$

Provided Assumption 4 holds, then as $\mathsf{N} \to \infty$,

$$\begin{aligned}\mathbb{E}\left[\left\|\frac{1}{\mathsf{N}}X^\top(I_{\mathsf{N}} - P_{\hat{C}})\delta_C\beta_C\right\|_2^2\right] &\le \mathrm{tr}(\beta_C\beta_C^\top)\left(\frac{1}{\mathsf{N}^2}\mathbb{E}\left[\|\delta_C\|_F^2 \mathrm{tr}(X^\top X)\right] - \frac{1}{\mathsf{N}^2}\mathbb{E}\left[\|\delta_C\|_F^2 \mathrm{tr}(X^\top P_{\hat{C}} X)\right]\right) \\ &= \mathrm{tr}(\beta_C\beta_C^\top)O(\frac{a_{\mathsf{N}}^2\mathsf{L}}{\delta^2}) \to 0,\end{aligned}$$

which implies $\frac{1}{\mathsf{N}}X^\top(I_{\mathsf{N}} - P_{\hat{C}})\delta_C\beta_C \xrightarrow{L_2} 0$, so that $\frac{1}{\mathsf{N}}X^\top(I_{\mathsf{N}} - P_{\hat{C}})\delta_C\beta_C \xrightarrow{P} 0$.

This completes the proof of Step 3.

□

S1.6 Proof of Theorem 4

Similar to the calculation procedure of (S1.26) in Theorem 4, we denote $\delta_Z = Z - \hat{Z}$, and

$$\tilde{\beta} = \begin{pmatrix} \tilde{\beta}_X \\ \tilde{\beta}_Z \end{pmatrix} = \begin{pmatrix} (X^\top(I_{\mathsf{N}} - P_{\hat{Z}})X)^{-1}X^\top(I_{\mathsf{N}} - P_{\hat{Z}})y \\ (\hat{Z}^\top(I_{\mathsf{N}} - P_X)\hat{Z})^{-1}\hat{Z}^\top(I_{\mathsf{N}} - P_X)y \end{pmatrix}.$$

From the regression model $y = X\beta_X + Z\beta_Z + \varepsilon$, we have

$$\tilde{\beta} = \begin{pmatrix} \tilde{\beta}_X \\ \tilde{\beta}_Z \end{pmatrix} = \beta + \begin{pmatrix} (X^\top(I_{\mathsf{N}} - P_{\hat{Z}})X)^{-1}X^\top(I_{\mathsf{N}} - P_{\hat{Z}})[\delta_Z\beta_Z + \varepsilon] \\ (\hat{Z}^\top(I_{\mathsf{N}} - P_X)\hat{Z})^{-1}\hat{Z}^\top(I_{\mathsf{N}} - P_X)[\delta_Z\beta_Z + \varepsilon] \end{pmatrix}.$$

First, we show the consistency of $\tilde{\beta}_X$. Since

$$\tilde{\beta}_X - \beta_X = (X^\top(I_{\mathsf{N}} - P_{\hat{Z}})X)^{-1}X^\top(I_{\mathsf{N}} - P_{\hat{Z}})[\delta_Z\beta_Z + \varepsilon],$$

then similar to the proof of (S1.29) and (S1.31), we have

$$(\frac{1}{\mathsf{N}}X^\top(I_{\mathsf{N}} - P_{\hat{Z}})X)^{-1} \xrightarrow{P} V_X^{-1}, \tag{S1.34}$$

and

$$\frac{1}{\mathsf{N}}X^\top(I_{\mathsf{N}} - P_{\hat{Z}})\varepsilon \xrightarrow{P} 0. \tag{S1.35}$$

Thus, it suffices to prove

$$\frac{1}{\mathsf{N}}X^\top(I_{\mathsf{N}} - P_{\hat{Z}})\delta_Z\beta_Z \xrightarrow{P} 0. \tag{S1.36}$$

Here we prove (S1.36) by computing the ℓ_2-norm of $\frac{1}{\mathsf{N}}X^\top(I_{\mathsf{N}} - P_{\hat{Z}})\delta_Z\beta_Z$:

$$\begin{aligned}\mathbb{E}\left[\left\|\frac{1}{\mathsf{N}}X^\top(I_{\mathsf{N}} - P_{\hat{Z}})\delta_Z\beta_Z\right\|_2^2\right] = &\frac{1}{\mathsf{N}^2}\mathbb{E}\left[\beta_Z^\top\delta_Z^\top(I_{\mathsf{N}} - P_{\hat{Z}})XX^\top(I_{\mathsf{N}} - P_{\hat{Z}})\delta_Z\beta_Z\right] \\ =&\beta_Z^2\frac{1}{\mathsf{N}^2}\mathbb{E}\left[\mathrm{tr}(\delta_Z^\top(I_{\mathsf{N}} - P_{\hat{Z}})XX^\top(I_{\mathsf{N}} - P_{\hat{Z}})\delta_Z)\right] \\ =&\beta_Z^2\frac{1}{\mathsf{N}^2}\mathbb{E}\left[\mathrm{tr}(\delta_Z\delta_Z^\top(I_{\mathsf{N}} - P_{\hat{Z}})XX^\top(I_{\mathsf{N}} - P_{\hat{Z}}))\right],\end{aligned}$$

and by Lemma S1.3, we have the upper bound

$$\leq\beta_Z^2\frac{1}{\mathsf{N}^2}\mathbb{E}\left[\mathrm{tr}(\delta_Z\delta_Z^\top)\mathrm{tr}((I_{\mathsf{N}} - P_{\hat{Z}})XX^\top(I_{\mathsf{N}} - P_{\hat{Z}}))\right]$$

$$
\begin{aligned}
&=\beta_Z^2 \frac{1}{\mathsf{N}^2}\mathbb{E}\left[\operatorname{tr}(\delta_Z\delta_Z^\top)\operatorname{tr}((I_\mathsf{N}-P_{\hat{Z}})XX^\top)\right] \\
&=\beta_Z^2 \frac{1}{\mathsf{N}^2}\mathbb{E}\left[\|\delta_Z\|_F^2\operatorname{tr}(X^\top(I_\mathsf{N}-P_{\hat{Z}})X)\right] \\
&=\beta_Z^2 \frac{1}{\mathsf{N}^2}(\mathbb{E}\left[\|\delta_Z\|_F^2\operatorname{tr}(X^\top X)\right]-\mathbb{E}\left[\|\delta_Z\|_F^2\operatorname{tr}(X^\top P_{\hat{Z}}X)\right]).
\end{aligned} \tag{S1.37}
$$

Now we first calculate $\mathbb{E}\left[\|\delta_Z\|_F^2\right]$. From the definition of $\hat{Z}$, we see that

$$
\|\delta_Z\|_F^2 = \frac{1}{\mathsf{L}^2}\|(\hat{U}-U)\,1_\mathsf{L}\|_2^2 \le \frac{1}{\mathsf{L}}\|\hat{U}-U\|_F^2.
$$

Similar to (S1.21), replacing U and C with $\hat{U}-U$, and $\hat{C}-C$ respectively yields

$$
\begin{aligned}
\|\delta_Z\|_F^2 &\le \frac{1}{\mathsf{L}}\|\hat{U}-U\|_F^2 \\
&\le \frac{1}{\mathsf{L}}\left\|S(S^\top S)^{-1}S^\top\right\|_F^2\left\|\hat{C}-C\right\|_2^2 \\
&= \frac{\mathsf{R}}{\mathsf{L}}\left\|\hat{C}-C\right\|_2^2 = \frac{\mathsf{R}}{\mathsf{L}}\left\|\hat{C}-C\right\|_F^2.
\end{aligned} \tag{S1.38}
$$

Then by Lemma S1.2, taking expectations on both sides of (S1.38) gives

$$
\begin{aligned}
\mathbb{E}\left[\|\delta_Z\|_F^2\right] &\le \frac{\mathsf{R}}{\mathsf{L}}\mathbb{E}\left[\left\|\hat{C}-C\right\|_F^2\right] \\
&= O(\frac{a_\mathsf{N}^2\mathsf{N}}{\delta^2}).
\end{aligned} \tag{S1.39}
$$

Given the upper bound in (S1.39), we return to the ℓ_2-norm of $\frac{1}{\mathsf{N}}X^\top(I_\mathsf{N}-P_{\hat{Z}})\delta_Z\beta_Z$ as in (S1.37). Since Z and $\hat{Z}$ are independent of X, we have

$$
\frac{1}{\mathsf{N}^2}\mathbb{E}\left[\|\delta_Z\|_F^2\operatorname{tr}(X^\top X)\right] = \mathbb{E}\left[\frac{1}{\mathsf{N}}\|\delta_Z\|_F^2\operatorname{tr}\left(\frac{1}{\mathsf{N}}X^\top X\right)\right]
$$

$$= \mathbb{E}\left[\frac{1}{\mathsf{N}}\|\delta_Z\|_F^2 \mathbb{E}^{Z,\hat{Z}}\left[\mathrm{tr}\left(\frac{1}{\mathsf{N}}X^\top X\right)\right]\right]$$
$$= \mathbb{E}\left[\frac{1}{\mathsf{N}}\|\delta_Z\|_F^2 \mathbb{E}^{C,\hat{C}}\left[\mathrm{tr}\left(\frac{1}{\mathsf{N}}X^\top X\right)\right]\right]$$

Since $\mathbb{E}^{C,E_0}\left[\frac{1}{\mathsf{N}}X_i^\top X_i\right] < \infty$, then

$$\frac{1}{\mathsf{N}^2}\mathbb{E}\left[\|\delta_Z\|_F^2 \mathrm{tr}(X^\top X)\right] = O\left(\frac{a_{\mathsf{N}}^2}{\delta^2}\right). \tag{S1.40}$$

In addition, we see that

$$\begin{aligned}\frac{1}{\mathsf{N}^2}\mathbb{E}\left[\|\delta_Z\|_F^2 \mathrm{tr}(X^\top P_{\hat{Z}}X)\right] &= \frac{1}{\mathsf{N}}\mathbb{E}\left[\|\delta_Z\|_F^2 \mathbb{E}^{\hat{Z},Z}\left[\mathrm{tr}\left(\frac{1}{\mathsf{N}}X^\top P_{\hat{Z}}X\right)\right]\right] \\ &= \frac{1}{\mathsf{N}}\mathbb{E}\left[\|\delta_Z\|_F^2 \mathrm{tr}\left(\mathbb{E}^{Z,\hat{Z}}\left[\frac{1}{\mathsf{N}}X^\top P_{\hat{Z}}X\right]\right)\right] \\ &= \frac{1}{\mathsf{N}}\mathbb{E}\left[\|\delta_Z\|_F^2 \sum_{i=1}^{\mathsf{P}}\left(\mathbb{E}^{Z,\hat{Z}}\left[\frac{1}{\mathsf{N}}X_i^\top P_{\hat{Z}}X_i\right]\right)\right]\end{aligned}$$

and $\mathbb{E}^{Z,\hat{Z}}\left[\frac{1}{\mathsf{N}}X_i^\top P_{\hat{Z}}X_j\right] = O(\frac{1}{\mathsf{N}})$ gives

$$\begin{aligned}&= \frac{1}{\mathsf{N}}\mathbb{E}\left[\|\delta_Z\|_F^2 O(\frac{1}{\mathsf{N}})\right] \\ &= \frac{1}{\mathsf{N}}O(\frac{a_{\mathsf{N}}^2\mathsf{N}}{\delta^2}\frac{1}{\mathsf{N}}) = O(\frac{a_{\mathsf{N}}^2}{\mathsf{N}\delta^2}).\end{aligned}$$

Hence, when Assumption 4 holds,

$$\begin{aligned}&\mathbb{E}\left[\left\|\frac{1}{\mathsf{N}}X^\top(I_{\mathsf{N}} - P_{\hat{Z}})\delta_Z\beta_Z\right\|_2^2\right] \\ \le&\beta_Z^2\frac{1}{\mathsf{N}^2}\left(\mathbb{E}\left[\|\delta_Z\|_F^2 \mathrm{tr}(X^\top X)\right] - \mathbb{E}\left[\|\delta_Z\|_F^2 \mathrm{tr}(X^\top P_{\hat{Z}}X)\right]\right) \\ =&O(\frac{a_{\mathsf{N}}^2}{\delta^2}) - O(\frac{a_{\mathsf{N}}^2}{\mathsf{N}\delta^2}) = O(\frac{a_{\mathsf{N}}^2}{\delta^2}) \to 0,\end{aligned}$$

which shows $\frac{1}{\mathsf{N}}X^\top(I_\mathsf{N}-P_{\hat{Z}})\delta_Z\beta_Z \xrightarrow{L_2} 0$ and then $\frac{1}{\mathsf{N}}X^\top(I_\mathsf{N}-P_{\hat{Z}})\delta_Z\beta_Z \xrightarrow{P} 0$, completing the proof of (S1.36). Finally, combining (S1.34), (S1.35) and (S1.36), we obtain

$$\begin{aligned}
\tilde{\beta}_X - \beta_X &= (X^\top(I_\mathsf{N}-P_{\hat{Z}})X)^{-1}X^\top(I_\mathsf{N}-P_{\hat{Z}})[\delta_Z\beta_Z+\varepsilon] \\
&= (\frac{1}{\mathsf{N}}X^\top(I_\mathsf{N}-P_{\hat{Z}})X)^{-1}\frac{1}{\mathsf{N}}X^\top(I_\mathsf{N}-P_{\hat{Z}})[\delta_Z\beta_Z+\varepsilon] \\
&\xrightarrow{P} V_X^{-1}(0+0) = 0,
\end{aligned}$$

showing the consistency of $\tilde{\beta}_X$.

Now we consider the consistency of $\tilde{\beta}_Z$. Note that

$$\tilde{\beta}_Z - \beta_Z = (\hat{Z}^\top(I-P_X)\hat{Z})^{-1}\hat{Z}^\top(I_\mathsf{N}-P_X)[\delta_Z\beta_Z+\varepsilon],$$

and we divide the proof of the consistency of $\tilde{\beta}_Z - \beta_Z$ to 2 steps:

- **Step 1:** Prove that with N sufficiently large, we have

$$\begin{aligned}
\left(\frac{1}{\mathsf{N}}\hat{Z}^\top(I_\mathsf{N}-P_X)\hat{Z}\right)^{-1} &\le \left(m - \frac{2}{\mathsf{N}}\frac{\mathsf{R}}{\mathsf{L}}a_\mathsf{N}^2\left\|\tilde{u}_1-u_1\right\|_2 - \frac{1}{\mathsf{N}}\hat{Z}^\top P_X\hat{Z}\right)^{-1} \\
&\xrightarrow{P} m^{-1}
\end{aligned}$$

  where m is a positive constant.

- **Step 2:** Prove that $\frac{1}{\mathsf{N}}\hat{Z}^\top(I_\mathsf{N}-P_X)[\delta_Z\beta_Z+\varepsilon] \xrightarrow{P} 0$.

Assembling the results of **Step 1** and **Step 2**, we obtain that as $\mathsf{N}\to\infty$,

$$\left|\tilde{\beta}_Z - \beta_Z\right|$$

$$
\begin{aligned}
&= \left| (\hat{Z}^\top (I - P_X)\hat{Z})^{-1}\hat{Z}^\top (I_{\mathsf{N}} - P_X)[\delta_Z \beta_Z + \varepsilon] \right| \\
&= \left(\frac{1}{\mathsf{N}} \hat{Z}^\top (I - P_X)\hat{Z} \right)^{-1} \left| \frac{1}{\mathsf{N}} \hat{Z}^\top (I_{\mathsf{N}} - P_X)[\delta_Z \beta_Z + \varepsilon] \right| \\
&\leq \left(m - \frac{2}{\mathsf{N}} \frac{\mathsf{R}}{\mathsf{L}} a_{\mathsf{N}}^2 \left\| \tilde{u}_1 - u_1 \right\|_2 - \frac{1}{\mathsf{N}} \hat{Z}^\top P_X \hat{Z} \right)^{-1} \left| \frac{1}{\mathsf{N}} \hat{Z}^\top (I_{\mathsf{N}} - P_X)[\delta_Z \beta_Z + \varepsilon] \right| \\
&\xrightarrow{P} m^{-1} \cdot 0 = 0,
\end{aligned}
$$

which gives the consistency of $\tilde{\beta}_Z$.

Step 1: Given the analogous properties of Z and $\hat{Z}$, similar to (S1.15) and (S1.16), we have

$$
\frac{1}{\mathsf{N}} \hat{Z}^\top X \xrightarrow{P} 0. \tag{S1.41}
$$

and

$$
\frac{1}{\mathsf{N}} \hat{Z}^\top P_X \hat{Z} \xrightarrow{P} 0. \tag{S1.42}
$$

Then we focus on $\frac{1}{\mathsf{N}} \hat{Z}^\top \hat{Z}$. Since $\frac{1}{\mathsf{N}} \delta_Z^\top \delta_Z \geq 0$, we see that

$$
\frac{1}{\mathsf{N}} \hat{Z}^\top \hat{Z} = \frac{1}{\mathsf{N}} Z^\top Z - \frac{2}{\mathsf{N}} Z^\top \delta_Z + \frac{1}{\mathsf{N}} \delta_Z^\top \delta_Z \geq \frac{1}{\mathsf{N}} Z^\top Z - \frac{2}{\mathsf{N}} Z^\top \delta_Z. \tag{S1.43}
$$

Then with the upper bound of $\|Z\|_2^2$ and $\|\delta_Z\|_2^2$ derived in (S1.21) and (S1.38), we have

$$
\begin{aligned}
\frac{2}{\mathsf{N}} Z^\top \delta_Z \leq \frac{2}{\mathsf{N}} \|Z\|_2 \|\delta_Z\|_2 &\leq \frac{2}{\mathsf{N}} \frac{\mathsf{R}}{\mathsf{L}} \|C\|_F \left\| \hat{C} - C \right\|_F \\
&\leq \frac{2}{\mathsf{N}} \frac{\mathsf{R}}{\mathsf{L}} \|C\|_F \left\| \hat{C} - C \right\|_F \\
&= \frac{2}{\mathsf{N}} \frac{\mathsf{R}}{\mathsf{L}} a_{\mathsf{N}}^2 \left\| \tilde{u}_1 - u_1 \right\|_2.
\end{aligned} \tag{S1.44}
$$

Plugging (S1.17) and (S1.44) into (S1.43), we see that

$$\frac{1}{\mathsf{N}}\hat{Z}^\top\hat{Z} \geq m - \frac{2}{\mathsf{N}}\frac{\mathsf{R}}{\mathsf{L}}a_{\mathsf{N}}^2 \left\|\tilde{u}_1 - u_1\right\|_2, \tag{S1.45}$$

where $\tilde{u}_1$ and u_1 are as defined in Lemma S1.1. To derive a positive lower bound for $\frac{1}{\mathsf{N}}\hat{Z}^\top\hat{Z}$, i.e. show that the RHS of (S1.45) is greater than zero for N sufficiently large, we need to show that

$$\frac{2}{\mathsf{N}}\frac{\mathsf{R}}{\mathsf{L}}a_{\mathsf{N}}^2 \left\|\tilde{u}_1 - u_1\right\|_2 \xrightarrow{P} 0.$$

Since $\mathbb{E}\left[\frac{2}{\mathsf{N}}\frac{\mathsf{R}}{\mathsf{L}}a_{\mathsf{N}}^2 \left\|\tilde{u}_1 - u_1\right\|_2\right] = O(\frac{a_{\mathsf{N}}^2}{\sqrt{\mathsf{N}\mathsf{L}}\delta}) = O(\frac{a_{\mathsf{N}}}{\delta}) \to 0$, we have

$$\frac{2}{\mathsf{N}}\frac{\mathsf{R}}{\mathsf{L}}a_{\mathsf{N}}^2 \left\|\tilde{u}_1 - u_1\right\|_2 \xrightarrow{P} 0. \tag{S1.46}$$

Then with (S1.42), (S1.45) and (S1.46), for N large enough, we have

$$\begin{aligned}\frac{1}{\mathsf{N}}\hat{Z}^\top(I_{\mathsf{N}} - P_X)\hat{Z} &= \frac{1}{\mathsf{N}}\hat{Z}^\top\hat{Z} - \frac{1}{\mathsf{N}}\hat{Z}^\top P_X\hat{Z}\\ &\geq m - \frac{2}{\mathsf{N}}\frac{\mathsf{R}}{\mathsf{L}}a_{\mathsf{N}}^2 \left\|\tilde{u}_1 - u_1\right\|_2 - \frac{1}{\mathsf{N}}\hat{Z}^\top P_X\hat{Z} > 0,\end{aligned}$$

which implies

$$\begin{aligned}\left(\frac{1}{\mathsf{N}}\hat{Z}^\top(I_{\mathsf{N}} - P_X)\hat{Z}\right)^{-1} &\leq \left(m - \frac{2}{\mathsf{N}}\frac{\mathsf{R}}{\mathsf{L}}a_{\mathsf{N}}^2 \left\|\tilde{u}_1 - u_1\right\|_2 - \frac{1}{\mathsf{N}}\hat{Z}^\top P_X\hat{Z}\right)^{-1}\\ &\xrightarrow{P} m^{-1}.\end{aligned} \tag{S1.47}$$

Step 2: Now we consider

$$\frac{1}{\mathsf{N}}\hat{Z}^\top(I_{\mathsf{N}} - P_X)[\delta_Z\beta_Z + \varepsilon] = \frac{1}{\mathsf{N}}\hat{Z}^\top\delta_Z\beta_Z - \frac{1}{\mathsf{N}}\hat{Z}^\top P_X\delta_Z\beta_Z + \frac{1}{\mathsf{N}}\hat{Z}^\top\varepsilon - \frac{1}{\mathsf{N}}\hat{Z}^\top P_X\varepsilon. \tag{S1.48}$$

For the first part of RHS of (S1.48), the Cauchy-Schwarz inequality gives the upper bound that

$$\mathbb{E}\left[\left|\frac{1}{\mathsf{N}}\hat{Z}^\top\delta_Z\beta_Z\right|\right] = \frac{\beta_Z}{\mathsf{N}}\mathbb{E}\left[\left|\hat{Z}^\top\delta_Z\right|\right] \leq \frac{\beta_Z}{\mathsf{N}}\left(\mathbb{E}\left[\left\|\hat{Z}\right\|_2^2\right]\right)^{\frac{1}{2}}\left(\mathbb{E}\left[\|\delta_Z\|_2^2\right]\right)^{\frac{1}{2}}.$$

Similar to the calculation leading to (S1.21), replacing Z and C with $\hat{Z}$ and $\hat{C}$ respectively, we have

$$\|\hat{Z}\|_2^2 \leq \frac{\mathsf{R}}{\mathsf{L}}\|\hat{C}\|_F^2, \tag{S1.49}$$

and

$$\mathbb{E}\left[\|\hat{Z}\|_2^2\right] \leq \mathbb{E}\left[\frac{\mathsf{R}}{\mathsf{L}}\|\hat{C}\|_F^2\right] = O(\frac{a_{\mathsf{N}}^2}{\mathsf{L}}). \tag{S1.50}$$

Then combining (S1.39) and (S1.50), provided Assumption 4 holds, we have as $\mathsf{N} \to 0$,

$$\mathbb{E}\left[\left|\frac{1}{\mathsf{N}}\hat{Z}^\top\delta_Z\beta_Z\right|\right] \leq \frac{\beta_Z}{\mathsf{N}}O(\frac{a_{\mathsf{N}}}{\sqrt{\mathsf{L}}})O(\frac{a_{\mathsf{N}}\sqrt{\mathsf{N}}}{\delta}) = O(\frac{\sqrt{\mathsf{N}}}{\sqrt{\mathsf{L}}\delta}) \to 0, \tag{S1.51}$$

i.e. $\frac{1}{\mathsf{N}}\hat{Z}^\top\delta_Z\beta_Z \xrightarrow{L_1} 0$.

For the second part in the RHS of (S1.48), also using the Cauchy-Schwarz inequality and Lemma S1.3, we have the following upper bound:

$$\mathbb{E}\left[\left|\frac{1}{\mathsf{N}}\hat{Z}^\top P_X\delta_Z\beta_Z\right|\right] = \frac{\beta_Z}{\mathsf{N}}\mathbb{E}\left[\left|\hat{Z}^\top P_X\delta_Z\right|\right] \leq \frac{\beta_Z}{\mathsf{N}}\left(\mathbb{E}\left[\left\|P_X\hat{Z}\right\|_2^2\right]\right)^{\frac{1}{2}}\left(\mathbb{E}\left[\|\delta_Z\|_2^2\right]\right)^{\frac{1}{2}}.$$

From (S1.50), we see that

$$\mathbb{E}\left[\left\|P_X\hat{Z}\right\|_2^2\right]=\mathbb{E}\left[\hat{Z}^\top P_X\hat{Z}\right]=\mathbb{E}\left[\mathrm{tr}\left(\hat{Z}^\top P_X\hat{Z}\right)\right]=\mathbb{E}\left[\mathrm{tr}\left(\hat{Z}\hat{Z}^\top P_X\right)\right]$$
$$\leq\mathbb{E}\left[\mathrm{tr}\left(\hat{Z}\hat{Z}^\top\right)\mathrm{tr}\left(P_X\right)\right]=P\mathbb{E}\left[\|\hat{Z}\|_2^2\right]\leq O(\frac{a_{\mathsf{N}}^2}{\mathsf{L}}),$$

and as $\mathsf{N}\to\infty$,

$$\mathbb{E}\left[\left|\frac{1}{\mathsf{N}}\hat{Z}^\top P_X\delta_Z\beta_Z\right|\right]=\frac{\beta_Z}{\mathsf{N}}O(\frac{a_{\mathsf{N}}}{\sqrt{\mathsf{L}}})O(\frac{a_{\mathsf{N}}\sqrt{\mathsf{N}}}{\delta})=O\left(\frac{\sqrt{\mathsf{N}}}{\delta\sqrt{\mathsf{L}}}\right)\to 0. \quad \text{(S1.52)}$$

For the third part in the RHS of (S1.48), also from Lemma S1.3 and (S1.50) we have as $\mathsf{N}\to\infty$,

$$\mathbb{E}\left[\left\|\frac{1}{\mathsf{N}}\hat{Z}^\top\varepsilon\right\|_2^2\right]=\frac{1}{\mathsf{N}^2}\mathbb{E}\left[\varepsilon^\top\hat{Z}\hat{Z}^\top\varepsilon\right]$$
$$=\frac{\sigma_y^2}{\mathsf{N}^2}\mathbb{E}\left[\|\hat{Z}\|_2^2\right]=O(\frac{a_{\mathsf{N}}^2}{\mathsf{N}^2\mathsf{L}})\to 0. \quad \text{(S1.53)}$$

Finally, for the fourth part in the RHS of (S1.48), with (S1.41) and the law of large numbers, we have

$$\frac{1}{\mathsf{N}}\hat{Z}^\top P_X\varepsilon=\frac{1}{\mathsf{N}}\hat{Z}^\top X\left(\frac{1}{\mathsf{N}}X^\top X\right)^{-1}\frac{1}{\mathsf{N}}X^\top\varepsilon\xrightarrow{P}0\cdot V_X^{-1}\cdot 0=0. \quad \text{(S1.54)}$$

By synthesizing equations (S1.51), (S1.52), (S1.53) and (S1.54), we can demonstrate the convergence of (S1.48) which completes the proof of Step 2.

□

S1.7 Proof of Theorem 5

Let $\hat{Z} = Z - \delta_Z$. The OLS estimator is:

$$\tilde{\beta}_Z - \beta_Z = \left(\hat{Z}^\top (I_\mathsf{N} - P_X)\hat{Z}\right)^{-1} \hat{Z}^\top (I_\mathsf{N} - P_X)(\varepsilon + \delta_Z \beta_Z).$$

Normalizing by $\sqrt{Z^\top Z/\sigma_y^2}$, we have

$$\sqrt{\frac{Z^\top Z}{\sigma_y^2}}\left(\tilde{\beta}_Z - \beta_Z\right) = \underbrace{\frac{Z^\top Z}{\mathsf{N}}\left(\frac{\hat{Z}^\top (I_\mathsf{N} - P_X)\hat{Z}}{\mathsf{N}}\right)^{-1}}_{(D)} \underbrace{\frac{\hat{Z}^\top (I_\mathsf{N} - P_X)\varepsilon}{\sqrt{Z^\top Z \sigma_y^2}}}_{(E)}$$
$$+ \frac{Z^\top Z}{\mathsf{N}}\left(\frac{\hat{Z}^\top (I_\mathsf{N} - P_X)\hat{Z}}{\mathsf{N}}\right)^{-1} \underbrace{\frac{\hat{Z}^\top (I_\mathsf{N} - P_X)\delta_Z \beta_Z}{\sqrt{Z^\top Z \sigma_y^2}}}_{(F)}.$$

Then we only need to prove that (D) $\xrightarrow{P} 1$, (E) $\xrightarrow{d} \mathcal{N}(0,1)$ and (F) $\xrightarrow{P} 0$.

For the term (D), we notice that

$$\begin{aligned}\frac{\hat{Z}^\top \hat{Z}}{\mathsf{N}}(\frac{1}{\mathsf{N}}\hat{Z}^\top (I - P_X)\hat{Z})^{-1} &= \left(\frac{\hat{Z}^\top \hat{Z}}{\mathsf{N}} - \frac{\hat{Z}^\top P_X \hat{Z}}{\mathsf{N}}\right)(\frac{1}{\mathsf{N}}\hat{Z}^\top (I - P_X)\hat{Z})^{-1} \\ &\quad + \frac{\hat{Z}^\top P_X \hat{Z}}{\mathsf{N}}(\frac{1}{\mathsf{N}}\hat{Z}^\top (I - P_X)\hat{Z})^{-1} \\ &= 1 + \frac{\hat{Z}^\top P_X \hat{Z}}{\mathsf{N}}(\frac{1}{\mathsf{N}}\hat{Z}^\top (I - P_X)\hat{Z})^{-1}.\end{aligned}$$

From (S1.42) and Assumption 5, we know $\frac{1}{\mathsf{N}}\hat{Z}^\top P_X \hat{Z} \xrightarrow{P} 0$. Also from (S1.47), there exists $m > 0$ such that for N large enough, $\left(\frac{1}{\mathsf{N}}\hat{Z}^\top (I_\mathsf{N} - P_X)\hat{Z}\right)^{-1} \leq m^{-1}$ with probability 1. Thus we have $\frac{\hat{Z}^\top P_X \hat{Z}}{\mathsf{N}}(\frac{1}{\mathsf{N}}\hat{Z}^\top (I - P_X)\hat{Z})^{-1} \xrightarrow{P} 0$ and $\frac{\hat{Z}^\top \hat{Z}}{\mathsf{N}}(\frac{1}{\mathsf{N}}\hat{Z}^\top (I - P_X)\hat{Z})^{-1} \xrightarrow{P} 1$ is proved.

Then we show that $\frac{Z^\top Z}{\hat{Z}^\top \hat{Z}} \xrightarrow{P} 1$. Notice that

$$\begin{aligned}\frac{\hat{Z}^\top \hat{Z}}{Z^\top Z} &= \frac{Z^\top Z - 2\delta_Z^\top Z + \|\delta_Z\|_2^2}{Z^\top Z} \\ &= 1 + \frac{\|\delta_Z\|_2^2 - 2\delta_Z^\top Z}{\|Z\|_2^2} \\ &= 1 + \frac{(\|\delta_Z\|_2^2 - 2\delta_Z^\top Z)/\mathsf{N}}{\|Z\|_2^2/\mathsf{N}}.\end{aligned}$$

From (S1.39), $\|\delta_Z\|_2^2/\mathsf{N} \xrightarrow{P} 0$. With (S1.45) and (S1.46), $\delta_Z^\top Z/\mathsf{N} \xrightarrow{P} 0$. From (S1.17), we know that $\|Z\|_2^2/\mathsf{N}$ is lower bounded by a positive constant m. Thus we have $\frac{(\|\delta_Z\|_2^2 - 2\delta_Z^\top Z)/\mathsf{N}}{\|Z\|_2^2/\mathsf{N}} \xrightarrow{P} 0$ and $\frac{\hat{Z}^\top \hat{Z}}{Z^\top Z} \xrightarrow{P} 1$. Therefore, the term (D) converges to 1 in probability.

Then we turn to analyze term (E). Notice that

$$\begin{aligned}\frac{\hat{Z}^\top (I_\mathsf{N} - P_X)\varepsilon}{\sqrt{Z^\top Z \sigma_y^2}} &= \frac{(Z - \delta_Z)^\top (I_\mathsf{N} - P_X)\varepsilon}{\sqrt{Z^\top Z \sigma_y^2}} \\ &= \frac{Z^\top (I_\mathsf{N} - P_X)\varepsilon}{\sqrt{Z^\top Z \sigma_y^2}} - \frac{\delta_Z^\top (I_\mathsf{N} - P_X)\varepsilon}{\sqrt{Z^\top Z \sigma_y^2}} \\ &= \frac{Z^\top \varepsilon}{\sqrt{Z^\top Z \sigma_y^2}} - \frac{Z^\top P_X \varepsilon}{\sqrt{Z^\top Z \sigma_y^2}} - \frac{\delta_Z^\top (I_\mathsf{N} - P_X)\varepsilon}{\sqrt{Z^\top Z \sigma_y^2}}. \qquad \text{(S1.55)}\end{aligned}$$

From the arguments of the convergence in the noiseless case, we know that

$$\frac{Z^\top \varepsilon}{\sqrt{Z^\top Z \sigma_y^2}} - \frac{Z^\top P_X \varepsilon}{\sqrt{Z^\top Z \sigma_y^2}} \xrightarrow{d} \mathcal{N}(0, 1).$$

Consider

$$\frac{\delta_Z^\top (I_\mathsf{N} - P_X)\varepsilon}{\sqrt{Z^\top Z \sigma_y^2}} = \frac{\delta_Z^\top \varepsilon}{\sqrt{Z^\top Z \sigma_y^2}} - \frac{\delta_Z^\top P_X \varepsilon}{\sqrt{Z^\top Z \sigma_y^2}}$$

where

$$\mathbb{E}\left[\left\|\frac{1}{\sqrt{\mathsf{N}}}\delta_Z^\top\varepsilon\right\|^2\right] \leq \frac{\sigma_y^2}{\mathsf{N}}\mathbb{E}\left[\|\delta_Z\|_F^2\right] = O(\frac{a_\mathsf{N}^2}{\delta^2}) \to 0 \tag{S1.56}$$

and

$$\mathbb{E}\left[\left\|\frac{1}{\sqrt{\mathsf{N}}}\delta_Z^\top P_X\varepsilon\right\|^2\right] \leq \frac{P\sigma_y^2}{\mathsf{N}}\mathbb{E}\left[\|\delta_Z\|_F^2\right] = O(\frac{a_\mathsf{N}^2}{\delta^2}) \to 0,$$

with (S1.17), we have

$$\frac{\delta_Z^\top(I_\mathsf{N} - P_X)\varepsilon}{\sqrt{Z^\top Z\sigma_y^2}} = \frac{\sqrt{\mathsf{N}}}{\sqrt{Z^\top Z\sigma_y^2}}\frac{\delta_Z^\top(I_\mathsf{N} - P_X)\varepsilon}{\sqrt{\mathsf{N}}} \xrightarrow{P} 0.$$

Thus, the term (E) converges to $\mathcal{N}(0,1)$ in distribution.

For the term (F), with Cauchy-Schwarz inequality,

$$\frac{\hat{Z}^\top(I_\mathsf{N} - P_X)\delta_Z}{\sqrt{Z^\top Z\sigma_y^2}} = \frac{\sqrt{\mathsf{N}}}{\sqrt{Z^\top Z\sigma_y^2}}\frac{\hat{Z}^\top(I_\mathsf{N} - P_X)\delta_Z}{\sqrt{\mathsf{N}}},$$

where

$$\left|\frac{\hat{Z}^\top(I_\mathsf{N} - P_X)\delta_Z}{\sqrt{\mathsf{N}}}\right| \leq \left|\frac{\hat{Z}^\top\delta_Z}{\sqrt{\mathsf{N}}}\right| + \left|\frac{\hat{Z}^\top P_X\delta_Z}{\sqrt{\mathsf{N}}}\right|.$$

Consider the expectation below

$$\mathbb{E}\left[\left|\frac{\hat{Z}^\top\delta_Z}{\sqrt{\mathsf{N}}}\right|\right] \leq \frac{1}{\sqrt{\mathsf{N}}}\mathbb{E}\left[\|\hat{Z}\|\|\delta_Z\|\right] \leq \frac{1}{\sqrt{\mathsf{N}}}\sqrt{\mathbb{E}[\|\hat{Z}\|^2]\mathbb{E}[\|\delta_Z\|^2]}.$$

Combining (S1.39) and (S1.50),

$$\frac{1}{\sqrt{\mathsf{N}}}\sqrt{\mathbb{E}[\|\hat{Z}\|^2]\mathbb{E}[\|\delta_Z\|^2]} \leq \frac{1}{\sqrt{\mathsf{N}}}O(\frac{a_\mathsf{N}}{\sqrt{\mathsf{L}}})O(\frac{a_\mathsf{N}\sqrt{\mathsf{N}}}{\delta}) = O(\frac{a_\mathsf{N}^2}{\delta\sqrt{\mathsf{L}}}) \to 0.$$

Also,

$$\left|\frac{\hat{Z}^\top P_X\delta_Z}{\sqrt{\mathsf{N}}}\right| = \left|\frac{\hat{Z}^\top X}{\sqrt{\mathsf{N}}}(\frac{1}{\mathsf{N}}X^\top X)^{-1}\frac{1}{\mathsf{N}}X^\top\delta_Z\right|.$$

Similar to the arguments of (S1.25), we have $\frac{\hat{Z}^\top X}{\sqrt{\mathsf{N}}}$ converges to a normal distribution in distribution. With (S1.40),

$$\mathbb{E}\left[\left\|\frac{1}{\mathsf{N}}X^\top \delta_Z\right\|_2^2\right] = O(\frac{a_{\mathsf{N}}^2}{\delta^2}) \to 0$$

Together with (S1.8), we have $\frac{\hat{Z}^\top P_X \delta_Z}{\sqrt{\mathsf{N}}} \xrightarrow{P} 0$. Finally, we obtain the term (F) converges to 0 in probability.

Combine the results above that term (D) converges to 1 in probability, term (E) converge to the standard normal distribution in distribution, and term (F) converges to 0 in probability, we have

$$\sqrt{\frac{Z^\top Z}{\sigma_y^2}}\left(\tilde{\beta}_Z - \beta_Z\right) \xrightarrow{d} \mathcal{N}(0,1).$$

Asymptotic Normality of $\tilde{\beta}_X$:

Recall that the OLS estimator is:

$$\tilde{\beta}_X = \left(X^\top (I_{\mathsf{N}} - P_{\hat{Z}})X\right)^{-1} X^\top (I_{\mathsf{N}} - P_{\hat{Z}})y.$$

We have:

$$\sqrt{\mathsf{N}}(\tilde{\beta}_X - \beta_X) = \left(\frac{X^\top (I_{\mathsf{N}} - P_{\hat{Z}})X}{\mathsf{N}}\right)^{-1} \frac{X^\top (I_{\mathsf{N}} - P_{\hat{Z}})(\varepsilon + \delta_Z \beta_Z)}{\sqrt{\mathsf{N}}}.$$

Define:

$$(A) := \left(\frac{X^\top (I_{\mathsf{N}} - P_{\hat{Z}})X}{\mathsf{N}}\right)^{-1}, \quad (B) := \frac{X^\top (I_{\mathsf{N}} - P_{\hat{Z}})\varepsilon}{\sqrt{\mathsf{N}}}, \quad (C) := \frac{X^\top (I_{\mathsf{N}} - P_{\hat{Z}})\delta_Z \beta_Z}{\sqrt{\mathsf{N}}}.$$

From (S1.34), $(A) \xrightarrow{P} V_X^{-1}$.

As for (B), we have

$$\frac{X^\top (I_{\mathsf{N}} - P_{\hat{Z}})\varepsilon}{\sqrt{\mathsf{N}}} = \frac{X^\top \varepsilon}{\sqrt{\mathsf{N}}} - \frac{X^\top P_{\hat{Z}}\varepsilon}{\sqrt{\mathsf{N}}}.$$

From CLT, $\frac{X^\top \varepsilon}{\sqrt{\mathsf{N}}} \xrightarrow{d} \mathcal{N}(0, \sigma_y^2 V_X)$. Now we turn to $\frac{X^\top P_{\hat{Z}}\varepsilon}{\sqrt{\mathsf{N}}}$.

$$\begin{aligned}\frac{X^\top P_{\hat{Z}}\varepsilon}{\sqrt{\mathsf{N}}} &= \frac{X^\top \hat{Z}(\hat{Z}^\top \hat{Z})^{-1}\hat{Z}^\top \varepsilon}{\sqrt{\mathsf{N}}} \\ &= \frac{X^\top \hat{Z}}{\mathsf{N}} \left(\frac{\hat{Z}^\top \hat{Z}}{\mathsf{N}}\right)^{-1/2} \frac{\hat{Z}^\top \varepsilon}{\sqrt{\hat{Z}^\top \hat{Z}}} \\ &= \frac{X^\top \hat{Z}}{\mathsf{N}} \left(\frac{\hat{Z}^\top \hat{Z}}{\mathsf{N}}\right)^{-1/2} \frac{\sqrt{Z^\top Z}}{\sqrt{\hat{Z}^\top \hat{Z}}} \frac{(Z - \delta_Z)^\top \varepsilon}{\sqrt{Z^\top Z}}.\end{aligned}$$

With (S1.41), (S1.45), $\frac{\hat{Z}^\top \hat{Z}}{Z^\top Z} \xrightarrow{P} 1$ and (S1.56), we have $\frac{X^\top P_{\hat{Z}}\varepsilon}{\sqrt{\mathsf{N}}} \xrightarrow{d} 0$. Thus, $(B) \xrightarrow{d} \mathcal{N}(0, \sigma_y^2 V_X)$.

We now prove that $(C) \xrightarrow{P} 0$. when Assumption 4 holds,

$$\begin{aligned}&\mathbb{E}\left[\left\|\frac{1}{\mathsf{N}} X^\top (I_{\mathsf{N}} - P_{\hat{Z}})\delta_Z \beta_Z\right\|_2^2\right] \\ \le&\beta_Z^2 \frac{1}{\mathsf{N}^2}\left(\mathbb{E}\left[\|\delta_Z\|_F^2 \mathrm{tr}(X^\top X)\right] - \mathbb{E}\left[\|\delta_Z\|_F^2 \mathrm{tr}(X^\top P_{\hat{Z}} X)\right]\right) \\ =&O(\frac{a_{\mathsf{N}}^2}{\delta^2}) - O(\frac{a_{\mathsf{N}}^2}{\mathsf{N}\delta^2}) = O(\frac{a_{\mathsf{N}}^2}{\delta^2}) \to 0,\end{aligned}$$

which means $(C) \xrightarrow{P} 0$.

By Slutsky's Theorem:

$$\sqrt{\mathsf{N}}(\tilde{\beta}_X - \beta_X) \xrightarrow{d} V_X^{-1} \cdot \mathcal{N}(0, \sigma_y^2 V_X) = \mathcal{N}(0, \sigma_y^2 V_X^{-1}).$$

□

S1.8 Community Detection and Error Propagation in CC-MNetR

In this section, we introduce a practical estimation strategy for settings where the community structure is unknown. We also provide sufficient conditions under which the resulting estimation error does not affect the consistency and asymptotic normality results established in the main text.

Community Estimation Procedure

When community structure is not predefined in the data, we recommend the following practical methodology:

Step 1: Network Aggregation

Construct the mean adjacency matrix across layers as

$$\bar{A} = \frac{1}{\mathsf{L}} \sum_{\ell=1}^{\mathsf{L}} A^{\ell}.$$

This consolidates connectivity patterns while preserving consistent structural features across layers.

Step 2: Spectral Clustering

Apply spectral clustering to $\bar{A}$ - a well-established community detection method for network data (von Luxburg, 2007). For multilayer networks with consistent community structure, this approach is supported by Han et al. (2015), who demonstrate its application to aggregated networks.

While Han et al. (2015)'s theoretical guarantees (Theorem 1) are derived under specific conditions:

- Stationary ergodic layer variations;
- Identifiable community structure (the expected cross-layer connectivity matrix $M = \mathbb{E}[\bar{P}]$ has distinct rows) *within the SBM framework*.

The methodology fundamentally operates on **graph topology** rather than generative mechanisms, as evidenced by its algorithmic foundation in minimizing the normalized cut objective (Shi and Malik, 1997):

$$\min_{Y} \operatorname{Tr}\left(Y^\top (D - \bar{A}) Y\right) \quad \text{s.t.} \quad Y^\top D Y = I$$

where D is the degree matrix of $\bar{A}$. This formulation depends solely on connectivity patterns in the aggregated adjacency matrix $\bar{A}$, independent of underlying data-generating processes. This topological basis is further reinforced by empirical validation in Han et al. (2015), where spectral clustering successfully extracted communities from Bluetooth proximity networks exhibiting non-SBM temporal dynamics (Sec. 5.3) and multi-relational networks with heterogeneous semantic layers (Sec. 5.4), both deviating from strict SBM assumptions.

Error Propagation Analysis

When community assignments are estimated, we define the estimated community-based centrality $\hat{Z}_{comm}$ analogously to Z as

$$\hat{Z}_{comm} = \frac{1}{\mathsf{L}} \hat{S}(\hat{S}^\top \hat{S})^{-1} (\hat{S}^\top C) \mathbf{1}_{\mathsf{L}}$$

where

- $\hat{S}$ is the estimated $\mathsf{N} \times \mathsf{R}$ community assignment matrix (each row has one 1 at the estimated community);
- $(\hat{S}^\top \hat{S})^{-1} = \mathrm{diag}(1/\hat{\mathsf{N}}_1, \ldots, 1/\hat{\mathsf{N}}_{\mathsf{R}})$;
- $\hat{\mathsf{N}}_r = \sum_{i=1}^{\mathsf{N}} \hat{S}_{ir}$ is the estimated size of community r.

Define community misassignment rate η as

$$\eta := \frac{1}{\mathsf{N}} \sum_{i=1}^{\mathsf{N}} \mathbb{I}\{c_i \neq \hat{c}_i\}$$

where c_i denotes the true label of node i and $\hat{c}_i$ represents the estimated label of node i.

Step 1: **Community matrix norms.** $\|S\|_F = \sqrt{\sum_{i=1}^{\mathsf{N}} \|S_i\|_2^2} = \sqrt{\mathsf{N}}$ since each row has one 1 and others 0. Similarly $\|\hat{S}\|_F = \sqrt{\mathsf{N}}$.

Step 2: **Community size error.** For N_r (size of community r) and estimate $\hat{\mathsf{N}}_r$, we have

$$|\hat{\mathsf{N}}_r - \mathsf{N}_r| \leq \eta \mathsf{N}.$$

Suppose the misassignment rate η vanishes as $\mathsf{N} \to \infty$, there exists N big enough so that $\eta < \epsilon/2$ where ϵ denotes lower bound of the proportion of all communities. Then with $\mathsf{N}_r \geq \epsilon \mathsf{N}$, we obtain $\hat{\mathsf{N}}_r \geq \epsilon \mathsf{N} - \eta \mathsf{N} \geq \epsilon \mathsf{N}/2$. Naturally, we have

$$\|(\hat{S}^\top \hat{S})^{-1} - (S^\top S)^{-1}\|_F^2 = \sum_{r=1}^{\mathsf{R}} \left(\frac{1}{\hat{\mathsf{N}}_r} - \frac{1}{\mathsf{N}_r} \right)^2 \leq \mathsf{R} \left(\frac{\eta \mathsf{N}}{(\epsilon \mathsf{N}/2)^2} \right)^2 = \frac{4\mathsf{R}\eta^2}{\epsilon^4 \mathsf{N}^2}$$

and

$$\boxed{\|(\hat{S}^\top \hat{S})^{-1} - (S^\top S)^{-1}\|_F = O\left(\frac{\eta}{\mathsf{N}}\right).}$$

Step 3: **Community aggregation error.**

We want to bound the error in community-aggregated features

$$\|(\hat{S}^\top C) - (S^\top C)\|_F$$

where $\hat{S}$ is the estimated community assignment matrix, S is the true assignment matrix (both $\mathsf{N} \times \mathsf{R}$), and C is the $\mathsf{N} \times \mathsf{L}$ centrality matrix.

(a) **Matrix element expression**

For community r and feature ℓ, the element-wise difference is

$$[(\hat{S}^\top C) - (S^\top C)]_{r\ell} = \sum_{i=1}^{\mathsf{N}} (\hat{S}_{ir} C_{i\ell} - S_{ir} C_{i\ell}),$$

which can be rewritten as

$$\sum_{i=1}^{\mathsf{N}} C_{i\ell} (\hat{S}_{ir} - S_{ir}).$$

(b) **Restrict to misassigned nodes**

Let $\mathcal{M}$ be the misassigned nodes set. For correctly assigned nodes, $\hat{S}_{ir} = S_{ir}$, so their contributions cancel out. Then the error comes only from misassigned nodes in $\mathcal{M}$ ($|\mathcal{M}| = \eta \mathsf{N}$) which shows

$$[(\hat{S}^\top C) - (S^\top C)]_{r\ell} = \sum_{i \in \mathcal{M}} C_{i\ell}(\hat{S}_{ir} - S_{ir}).$$

(c) **Element-wise bound**

Taking absolute values and using the triangle inequality, we get

$$\left| \sum_{i \in \mathcal{M}} C_{i\ell}(\hat{S}_{ir} - S_{ir}) \right| \leq \sum_{i \in \mathcal{M}} |C_{i\ell}||\hat{S}_{ir} - S_{ir}|.$$

Since $|\hat{S}_{ir} - S_{ir}| \leq 1$ (as both are indicator functions), we have

$$\left| [(\hat{S}^\top C) - (S^\top C)]_{r\ell} \right| \leq \sum_{i \in \mathcal{M}} |C_{i\ell}|.$$

(d) **Frobenius norm squared**

The squared Frobenius norm sums over all communities and features

$$\|(\hat{S}^\top C) - (S^\top C)\|_F^2 = \sum_{r=1}^{\mathsf{R}} \sum_{\ell=1}^{\mathsf{L}} \left| [(\hat{S}^\top C) - (S^\top C)]_{r\ell} \right|^2.$$

Using the bound from Step 3 we obtain

$$\|(\hat{S}^\top C) - (S^\top C)\|_F^2 \leq \sum_{r=1}^{\mathsf{R}} \sum_{\ell=1}^{\mathsf{L}} \left(\sum_{i \in \mathcal{M}} |C_{i\ell}| \right)^2.$$

(e) **Factorize sums**

Since there are R communities and L features, we can write

$$\sum_{r=1}^{\mathsf{R}}\sum_{\ell=1}^{\mathsf{L}}\left(\sum_{i\in\mathcal{M}}|C_{i\ell}|\right)^2 = \mathsf{R}\sum_{\ell=1}^{\mathsf{L}}\left(\sum_{i\in\mathcal{M}}|C_{i\ell}|\right)^2$$

(f) **Uniform bound on features**:

Since $|C_{i\ell}| \le a_{\mathsf{N}}$ for all i, ℓ (where a_{N} may depend on N), then

$$\sum_{i\in\mathcal{M}}|C_{i\ell}| \le \sum_{i\in\mathcal{M}} a_{\mathsf{N}} = \eta\mathsf{N}a_{\mathsf{N}},$$

and

$$\left(\sum_{i\in\mathcal{M}}|C_{i\ell}|\right)^2 \le (\eta\mathsf{N}a_{\mathsf{N}})^2.$$

(g) **Final bound for Frobenius norm squared**

Substituting back into the norm expression we get

$$\|(\hat{S}^\top C)-(S^\top C)\|_F^2 \le \mathsf{R}\sum_{\ell=1}^{\mathsf{L}}(\eta\mathsf{N}a_{\mathsf{N}})^2 = \mathsf{RL}(\eta\mathsf{N}a_{\mathsf{N}})^2 = \eta^2\mathsf{N}^2a_{\mathsf{N}}^2\mathsf{RL}.$$

(h) **Taking square root**

The Frobenius norm is

$$\|(\hat{S}^\top C) - (S^\top C)\|_F \le \sqrt{\eta^2\mathsf{N}^2a_{\mathsf{N}}^2\mathsf{RL}} = \eta\mathsf{N}a_{\mathsf{N}}\sqrt{\mathsf{RL}}.$$

Finally, we have

$$\boxed{\|\hat{S}^\top C - S^\top C\|_F = O\left(\eta\mathsf{N}a_{\mathsf{N}}\sqrt{\mathsf{L}}\right).}$$

Step 4: Z **estimation error.** Using the revised definition:

$$
\begin{aligned}
\|\hat{Z}_{comm} - Z\|_2 &\leq \frac{1}{\sqrt{\mathsf{L}}} \left\| \hat{S}(\hat{S}^\top \hat{S})^{-1}(\hat{S}^\top C) - S(S^\top S)^{-1}(S^\top C) \right\|_F \\
&\leq \frac{1}{\sqrt{\mathsf{L}}} \left(\left\|\hat{S}\right\|_F \left\|(\hat{S}^\top \hat{S})^{-1}\right\|_2 \left\|(\hat{S}^\top C) - (S^\top C)\right\|_F \right. \\
&\left. + \left\|\hat{S}\right\|_F \left\|(\hat{S}^\top \hat{S})^{-1} - (S^\top S)^{-1}\right\|_F \|S^\top C\|_F + \left\|\hat{S} - S\right\|_F \|(S^\top S)^{-1}\|_2 \|S^\top C\|_F \right) \\
&= O\left(\frac{1}{\sqrt{\mathsf{L}}} \left(\sqrt{\mathsf{N}} \frac{1}{\epsilon \mathsf{N}} \eta \mathsf{N} a_{\mathsf{N}} \sqrt{\mathsf{L}} + \sqrt{\mathsf{N}} \frac{\eta}{\mathsf{N}} a_{\mathsf{N}} \sqrt{\mathsf{N}} + \sqrt{\eta \mathsf{N}} \frac{1}{\epsilon \mathsf{N}} a_{\mathsf{N}} \sqrt{\mathsf{N}} \right) \right) \\
&= O\left(a_{\mathsf{N}} \eta \sqrt{\mathsf{N}} \right) + O\left(a_{\mathsf{N}} \frac{\eta}{\sqrt{\mathsf{L}}} \right) + O\left(a_{\mathsf{N}} \sqrt{\frac{\eta}{\mathsf{L}}} \right) \\
&= O\left(a_{\mathsf{N}} \eta \sqrt{\mathsf{N}} \right) + O\left(a_{\mathsf{N}} \sqrt{\frac{\eta}{\mathsf{L}}} \right)
\end{aligned}
$$

Step 5: **Regression coefficient impact.**

For CC-MNetR estimator $\hat{\beta}_{comm} = (\hat{\mathbf{W}}_3^\top \hat{\mathbf{W}}_3)^{-1} \hat{\mathbf{W}}_3^\top y$ where $\hat{\mathbf{W}}_3 = (X, \hat{Z}_{comm})$, we have

$$
\hat{\beta}_{comm} = \begin{pmatrix} \hat{\beta}_{X,comm} \\ \hat{\beta}_{Z,comm} \end{pmatrix} = \beta + \begin{pmatrix} (X^\top (I_{\mathsf{N}} - P_{\hat{Z}_{comm}}) X)^{-1} X^\top (I_{\mathsf{N}} - P_{\hat{Z}_{comm}}) [\hat{\delta}_Z \beta_Z + \varepsilon] \\ (\hat{Z}_{comm}^\top (I_{\mathsf{N}} - P_X) \hat{Z}_{comm})^{-1} \hat{Z}_{comm}^\top (I_{\mathsf{N}} - P_X) [\hat{\delta}_Z \beta_Z + \varepsilon] \end{pmatrix}.
$$

where $\hat{\delta}_Z = \hat{Z}_{comm} - Z$ and $P_{\hat{Z}_{comm}} = \hat{Z}_{comm}(\hat{Z}_{comm}^\top \hat{Z}_{comm})^{-1} \hat{Z}_{comm}^\top$.

Then we proceed to discuss the additional conditions required to ensure the consistency of the estimator.

First, we consider $\hat{\beta}_{Z,comm}$.

$$
\begin{aligned}
&\mathbb{E} \left\| \hat{\beta}_{Z,comm} - \beta_Z \right\|^2 \\
\leq &\mathbb{E} \left\| (\frac{1}{\mathsf{N}} \hat{Z}_{comm}^\top (I_{\mathsf{N}} - P_X) \hat{Z}_{comm})^{-1} \frac{1}{\mathsf{N}} \hat{Z}_{comm}^\top (I_{\mathsf{N}} - P_X) [\hat{\delta}_Z \beta_Z + \varepsilon] \right\|^2
\end{aligned}
$$

$$\leq \mathbb{E}\left\|(\frac{1}{\mathsf{N}}\hat{Z}_{comm}^\top(I_\mathsf{N}-P_X)\hat{Z}_{comm})^{-1}\frac{1}{\mathsf{N}}\hat{Z}_{comm}^\top(I_\mathsf{N}-P_X)\hat{\delta}_Z\beta_Z\right\|^2 + \mathbb{E}\left\|(\frac{1}{\mathsf{N}}\hat{Z}_{comm}^\top(I_\mathsf{N}-P_X)\hat{Z}_{comm})^{-1}\frac{1}{\mathsf{N}}\hat{Z}_{comm}^\top(I_\mathsf{N}-P_X)\varepsilon\right\|^2$$

where

$$\begin{aligned}
&\frac{1}{\mathsf{N}}(\hat{Z}_{comm}^\top(I_\mathsf{N}-P_X)\hat{Z}_{comm})\\
=&\frac{1}{\mathsf{N}}\hat{Z}_{comm}^\top\hat{Z}_{comm}-\frac{1}{\mathsf{N}}\hat{Z}_{comm}^\top P_X\hat{Z}_{comm}\\
\geq&\frac{1}{\mathsf{N}}Z^\top Z-\frac{2}{\mathsf{N}}Z^\top\hat{\delta}_Z-\frac{1}{\mathsf{N}}\hat{Z}_{comm}^\top P_X\hat{Z}_{comm}.
\end{aligned}$$

From (S1.15) and (S1.16) we know that $\frac{1}{\mathsf{N}}\hat{Z}_{comm}^\top P_X\hat{Z}_{comm}\xrightarrow{P}0$. From (S1.17), with $a_\mathsf{N}=\sqrt{\mathsf{NL}}$, there exists a constant $m>0$ such that $\frac{1}{\mathsf{N}}Z^\top Z\geq m$, a.s.. Similar to (S1.42), we have

$$\begin{aligned}
\frac{2}{\mathsf{N}}Z^\top\hat{\delta}_Z\leq\frac{2}{\mathsf{N}}\|Z\|_2\|\hat{\delta}_Z\|_2\leq\frac{2\|C\|_F}{\mathsf{NL}}\left(O\left(a_\mathsf{N}\eta\sqrt{\mathsf{N}}\right)+O\left(a_\mathsf{N}\sqrt{\frac{\eta}{\mathsf{L}}}\right)\right)\\
=O\left(\frac{a_\mathsf{N}^2\eta}{\sqrt{\mathsf{NL}}}\right)+O\left(\frac{a_\mathsf{N}^2}{\mathsf{NL}}\sqrt{\frac{\eta}{\mathsf{L}}}\right).
\end{aligned}$$

When $a_\mathsf{N}\asymp\sqrt{\mathsf{NL}}$, naturally we have $O\left(\frac{a_\mathsf{N}^2\eta}{\sqrt{\mathsf{NL}}}\right)+O\left(\frac{a_\mathsf{N}^2}{\mathsf{NL}}\sqrt{\frac{\eta}{\mathsf{L}}}\right)=O\left(\eta\sqrt{\mathsf{N}}\right)+O\left(\sqrt{\frac{\eta}{\mathsf{L}}}\right)$. If $\eta\sqrt{\mathsf{N}}=o(1)$, then $\frac{2}{\mathsf{N}}Z^\top\hat{\delta}_Z\to0$ as $\mathsf{N}\to\infty$. Consequently, as $\mathsf{N}\to\infty$, we have

$$\left(\frac{1}{\mathsf{N}}(\hat{Z}_{comm}^\top(I_\mathsf{N}-P_X)\hat{Z}_{comm})\right)^{-1}\leq\left(m-\frac{2}{\mathsf{N}}Z^\top\hat{\delta}_Z-\frac{1}{\mathsf{N}}\hat{Z}_{comm}^\top P_X\hat{Z}_{comm}\right)^{-1}\to m^{-1}.$$

Thus,

$$
\begin{aligned}
&\mathbb{E}\left\|\hat{\beta}_{Z,comm}-\beta_Z\right\|^2\\
\leq&\mathbb{E}\left\|(\frac{1}{\mathsf{N}}\hat{Z}_{comm}^\top(I_{\mathsf{N}}-P_X)\hat{Z}_{comm})^{-1}\frac{1}{\mathsf{N}}\hat{Z}_{comm}^\top(I_{\mathsf{N}}-P_X)\hat{\delta}_Z\beta_Z\right\|^2\\
&\quad+\mathbb{E}\left\|(\frac{1}{\mathsf{N}}\hat{Z}_{comm}^\top(I_{\mathsf{N}}-P_X)\hat{Z}_{comm})^{-1}\frac{1}{\mathsf{N}}\hat{Z}_{comm}^\top(I_{\mathsf{N}}-P_X)\varepsilon\right\|^2\\
\leq&m^{-1}\left(\mathbb{E}\left\|\frac{1}{\mathsf{N}}\hat{Z}_{comm}^\top(I_{\mathsf{N}}-P_X)\hat{\delta}_Z\beta_Z\right\|^2+\mathbb{E}\left\|\frac{1}{\mathsf{N}}\hat{Z}_{comm}^\top(I_{\mathsf{N}}-P_X)\varepsilon\right\|^2\right)\\
\leq&m^{-1}\left(\mathbb{E}\left[\left(\frac{1}{\mathsf{N}}\|\hat{Z}_{comm}\|_2\|\hat{\delta}_Z\|_2\beta_Z\right)^2\right]+\frac{\sigma_y^2}{\mathsf{N}^2}\mathbb{E}\left[\|(I_{\mathsf{N}}-P_X)\hat{Z}_{comm}\|_F^2\right]\right).
\end{aligned}
$$

Here, from (S1.48), when $a_{\mathsf{N}}\asymp\sqrt{\mathsf{NL}}$, we have

$$
\begin{aligned}
\left(\frac{1}{\mathsf{N}}\|\hat{Z}_{comm}\|_2\|\hat{\delta}_Z\|_2\beta_Z\right)^2&\leq\frac{1}{\mathsf{N}^2}\|\hat{Z}_{comm}\|_2^2\|\hat{\delta}_Z\|_2^2\beta_Z^2\\
&=O(\frac{1}{\mathsf{N}^2}\frac{a_{\mathsf{N}}^2}{\mathsf{L}}(O(a_{\mathsf{N}}^2\eta^2\mathsf{N})+O(a_{\mathsf{N}}^2\frac{\eta}{\mathsf{L}})))\\
&=O(\eta^2\mathsf{NL})+O(\eta)
\end{aligned}
$$

and

$$
\frac{\sigma_y^2}{\mathsf{N}^2}\mathbb{E}\left[\|(I_{\mathsf{N}}-P_X)\hat{Z}_{comm}\|_F^2\right]\leq\frac{\sigma_y^2}{\mathsf{N}^2}\mathbb{E}\left[\|\hat{Z}_{comm}\|_F^2\right]=O(\frac{a_{\mathsf{N}}^2}{\mathsf{N}^2\mathsf{L}})=O(\frac{1}{\mathsf{N}}).
$$

If $\eta=o(1)$ and $\eta^2\mathsf{NL}=o(1)$, then $\mathbb{E}\left\|\hat{\beta}_{Z,comm}-\beta_Z\right\|^2\to 0$.

Then, we consider $\hat{\beta}_{X,comm}$. The main difference between $\tilde{\beta}_X$ and $\hat{\beta}_{X,comm}$ exists in the difference between δ_Z and $\hat{\delta}_Z$. From (S1.38), we have

$$
\mathbb{E}\left[\|\delta_Z\|^2\right]=O(\frac{a_{\mathsf{N}}^2\mathsf{N}}{\delta^2}).
$$

And from above, we know that

$$\mathbb{E}\left[\|\hat{\delta}_Z\|^2\right] = O\left(a_{\mathsf{N}}\eta\sqrt{\mathsf{N}}\right) + O\left(a_{\mathsf{N}}\sqrt{\frac{\eta}{\mathsf{L}}}\right).$$

Since $\eta^2\mathsf{NL} = o(1)$, $\mathbb{E}\left[\|\hat{\delta}_Z\|^2\right] = O\left(a_{\mathsf{N}}\eta\sqrt{\mathsf{N}}\right) + O\left(a_{\mathsf{N}}\sqrt{\frac{\eta}{\mathsf{L}}}\right)$. If $\frac{\delta^2\eta}{a_{\mathsf{N}}\sqrt{\mathsf{N}}} = o(1)$ and $\frac{\delta^2\sqrt{\eta}}{a_{\mathsf{N}}\mathsf{N}\sqrt{\mathsf{L}}} = o(1)$, then the order of $\mathbb{E}\left[\|\hat{\delta}_Z\|^2\right]$ is lower that the order of $\mathbb{E}\left[\|\delta_Z\|^2\right]$. Therefore, the entire derivation holds.

Now we present the additional assumptions required for the regression coefficients to remain consistent when community information is unknown and obtained through estimation. Given $a_{\mathsf{N}} \asymp \sqrt{\mathsf{NL}}$, CC-MNetR maintains consistency when

1. $\eta = o(1)$: Community misassignment rate vanishes asymptotically.
2. $\eta = o(1/\sqrt{\mathsf{N}})$: Cumulative misassignment (i.e. $\eta\mathsf{N}$) grows slower than $\sqrt{\mathsf{N}}$.
3. $\eta = o(1/\sqrt{\mathsf{NL}})$: Cumulative misassignment (i.e. $\eta\mathsf{N}$) grows slower than $\sqrt{\frac{\mathsf{N}}{\mathsf{L}}}$.
4. $\dfrac{\delta^2\eta}{\mathsf{N}\sqrt{\mathsf{L}}} \to 0$ and $\dfrac{\delta^2\sqrt{\eta}}{\mathsf{N}^{3/2}\mathsf{L}} \to 0$: Spectral gap interaction conditions.

Conditions 3 automatically satisfy Condition 1 and 2. To sum up, $\eta = o(1/\sqrt{\mathsf{NL}})$, $\dfrac{\delta^2\eta}{\mathsf{N}\sqrt{\mathsf{L}}} \to 0$, and $\dfrac{\delta^2\sqrt{\eta}}{\mathsf{N}^{3/2}\mathsf{L}} \to 0$ guarantee consistency of $\hat{\beta}_{comm}$.

S1.9 Discussion on Assumption 5

This section discusses when Assumption 5 and the regularity conditions in Theorem 5 are expected to be satisfied or potentially violated in practical settings. Specifically, the spectral gap condition is now stated as:

- $\frac{a_{\mathsf{N}}\sqrt{\mathsf{L}}}{\delta} \to 0, \quad$ with $a_{\mathsf{N}} \asymp \sqrt{\mathsf{NL}}$, in Assumption 5.

- $\frac{a_{\mathsf{N}}\sqrt{\mathsf{N}}}{\delta} \to 0, \quad$ with $a_{\mathsf{N}} \asymp \sqrt{\mathsf{NL}}$, in Theorem 5.

These conditions accommodate a broader class of multilayer networks beyond the fixed-L setting. In particular, consistency requires the spectral gap δ to grow faster than $\sqrt{\mathsf{N}}\mathsf{L}$, while asymptotic normality requires δ to grow faster than $\sqrt{\mathsf{L}}\mathsf{N}$.

This formulation more accurately reflects the underlying complexity when layers are large or heterogeneous, and it enables a meaningful comparison across network types with varying density and coupling structures.

In what follows, we rigorously analyze whether this updated condition is satisfied under several representative multilayer network models.

Case 1: Dense multilayer networks ($\delta = \Theta(\mathsf{NL})$)

Setup: Consider a multilayer network where each layer has dense intra-layer connections, and layers are coupled via uniform inter-layer connections. The adjacency matrix is defined as:

$$B_0 = \begin{bmatrix} p(\mathbf{1}\mathbf{1}^\top - I) & r\mathbf{1}\mathbf{1}^\top & \cdots \\ r\mathbf{1}\mathbf{1}^\top & p(\mathbf{1}\mathbf{1}^\top - I) & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix},$$

where B_0 is the supra-adjacency matrix of the multilayer network, $p > 0$ denotes the intra-layer coupling strength (connections within the same layer), $r > 0$ denotes the inter-layer coupling strength (connections between different layers), $\mathbf{1}$ is an N-dimensional all-ones vector, and I is the $\mathsf{N} \times \mathsf{N}$ identity matrix.

Eigenvalue derivation: To understand the spectral gap $\delta = \lambda_1 - \lambda_2$, we analyze the two largest eigenvalues of B_0 using their corresponding eigenvectors:

1. *Leading eigenvalue* λ_1:

Let $\mathbf{v}$ be an NL-dimensional all-ones vector. Then

$$B_0\mathbf{v} = \begin{bmatrix} p(\mathsf{N}-1)\mathbf{1}_\mathsf{N} + r(\mathsf{N}(\mathsf{L}-1))\mathbf{1}_\mathsf{N} \\ \vdots \\ p(\mathsf{N}-1)\mathbf{1}_\mathsf{N} + r(\mathsf{N}(\mathsf{L}-1))\mathbf{1}_\mathsf{N} \end{bmatrix} = [p(\mathsf{N}-1) + r\mathsf{N}(\mathsf{L}-1)]\mathbf{v}$$

So $\lambda_1 \geq p(\mathsf{N}-1) + r\mathsf{N}(\mathsf{L}-1) = \Theta(\mathsf{NL})$.

2. *Second eigenvalue* λ_2:

Let $\mathbf{u} = \begin{bmatrix} \mathbf{1}_{\mathsf{N}} \\ -\mathbf{1}_{\mathsf{N}} \\ \mathbf{0}_{(\mathsf{L}-2)\mathsf{N}} \end{bmatrix}$ denote a contrast vector.

$$B_0\mathbf{u} = \begin{bmatrix} p(\mathsf{N}-1)\mathbf{1}_{\mathsf{N}} - r\mathsf{N}\mathbf{1}_{\mathsf{N}} \\ -p(\mathsf{N}-1)\mathbf{1}_{\mathsf{N}} + r\mathsf{N}\mathbf{1}_{\mathsf{N}} \\ \mathbf{0}_{(\mathsf{L}-2)\mathsf{N}} \end{bmatrix} = [p(\mathsf{N}-1) - r\mathsf{N}]\mathbf{u}$$

So $\lambda_2 \geq p(\mathsf{N}-1) - r\mathsf{N}$.

3.*Exact eigenvalues:* The full spectrum of B_0 consists of:

- $\lambda_1 = p(\mathsf{N}-1) + r\mathsf{N}(\mathsf{L}-1)$, multiplicity 1;
- $\lambda_2 = p(\mathsf{N}-1) - r\mathsf{N}$, multiplicity $\mathsf{L}-1$;
- $\mathsf{L}(\mathsf{N}-1)$ eigenvalues equal to $-p$.

This confirms that λ_1 and λ_2 are indeed the two largest eigenvalues of B_0, and the spectral gap satisfies

$$\delta = \lambda_1 - \lambda_2 = 2r\mathsf{N}(\mathsf{L}-1) = \Theta(\mathsf{NL}).$$

With $\delta = \Theta(\mathsf{NL})$ and $a_{\mathsf{N}} \asymp \sqrt{\mathsf{NL}}$, the condition $\frac{a_{\mathsf{N}}\sqrt{\mathsf{L}}}{\delta} \to 0$ reduces to $\sqrt{\frac{1}{\mathsf{N}}} \to 0$. This holds naturally. The condition $\frac{a_{\mathsf{N}}\sqrt{\mathsf{N}}}{\delta} \to 0$ reduces to $\sqrt{\frac{1}{\mathsf{L}}} \to 0$. This holds if the number of layers L grows with the number of nodes N (e.g., $\mathsf{L} = O(\mathsf{N}^{\alpha})$ with $\alpha < 1$), ensuring asymptotic normality of our estimates.

The block-constant model used in our spectral gap derivation provides a convenient analytical form, but real-world multilayer networks are rarely so regular. In practice, intra- and inter-layer connections often exhibit heterogeneity, with approximate block structure or community patterns.

Consider the supra-adjacency matrix $B \in \mathbb{R}^{\mathsf{NL}\times\mathsf{NL}}$ as a *single* random graph on NL nodes. Assume its expected version B_0 satisfies:

1. The minimum expected degree $\delta_{\min} \geq c_1\mathsf{NL}$ for some constant $c_1 > 0$
2. The maximum expected degree $\Delta \leq c_2\mathsf{NL}$ for some constant $c_2 > 0$
3. The spectral gap of B_0 satisfies $\delta_0 := \lambda_1(B_0) - \lambda_2(B_0) \geq c_3\mathsf{NL}$ for some constant $c_3 > 0$

These conditions imply B_0 is dense (e.g., analogous to an Erdős-Rényi graph with $p = \Theta(1)$) and has a large spectral gap. Applying Graham and Radcliffe (2011)'s Theorem 1 directly to the flattened matrix B:

$$\|\lambda_i(B) - \lambda_i(B_0)\| \leq \sqrt{4\Delta \ln(2\mathsf{NL}/\epsilon)} = O\left(\sqrt{\mathsf{NL}\ln(\mathsf{NL})}\right)$$

with probability $\geq 1-\epsilon$ when $\Delta > \frac{4}{9}\ln(2\mathsf{NL}/\epsilon)$. Consequently:

$$\delta := \lambda_1(B) - \lambda_2(B) \geq \delta_0 - O\left(\sqrt{\mathsf{NL}\ln(\mathsf{NL})}\right) = \Theta(\mathsf{NL}) - O\left(\sqrt{\mathsf{NL}\ln(\mathsf{NL})}\right)$$

which dominates to $\delta = \Theta(\mathsf{NL})$ for large NL. The key ratio then scales as:

$$\frac{a_{\mathsf{N}}\sqrt{\mathsf{L}}}{\delta} = \frac{\sqrt{\mathsf{NL}}\sqrt{\mathsf{L}}}{\Theta(\mathsf{NL})} = O\left(\frac{1}{\sqrt{\mathsf{N}}}\right) \to 0, \quad \frac{a_{\mathsf{N}}\sqrt{\mathsf{N}}}{\delta} = \frac{\sqrt{\mathsf{NL}}\sqrt{\mathsf{N}}}{\Theta(\mathsf{NL})} = O\left(\frac{1}{\sqrt{\mathsf{L}}}\right) \to 0$$

This holds as $\mathsf{N} \to \infty$. Such conditions arise when:

- *Intra-layer connections:* Each layer is dense (e.g., ER graphs with $p = \Theta(1)$)
- *Inter-layer connections:* Uniform and non-vanishing coupling (e.g., $P_{\alpha\beta}^{\text{inter}} = \Theta(1)$ for $\alpha \neq \beta$)

Thus, our theoretical results apply more broadly to general dense multilayer networks that exhibit strong global connectivity but may not follow perfectly uniform patterns.

Case 2: Sparse multilayer networks ($\delta = o(\mathsf{N})$)

Setup: Consider a multilayer stochastic block model (SBM) where each layer is sparse—i.e., the expected degree per node is bounded or grows slowly with N. For instance, each layer follows an SBM with K blocks and intra-/inter-block connection probabilities on the order of $O(1/\mathsf{N})$. Inter-layer coupling is either absent or weak ($r = o(1)$).

In this setting, the expected supra-adjacency matrix B_0 has leading eigenvalue $\lambda_1 = O(1)$, and the second eigenvalue may be close (e.g., due to weak community separation or weak inter-layer coupling). The gap satisfies

$$\delta = \lambda_1 - \lambda_2 = o(\mathsf{N}).$$

With $a_{\mathsf{N}} \asymp \sqrt{\mathsf{NL}}$ and $\delta = o(\mathsf{N})$, the condition

$$\frac{a_{\mathsf{N}}\sqrt{\mathsf{L}}}{\delta} \asymp \frac{\sqrt{\mathsf{N}}\mathsf{L}}{\delta} \not\to 0$$

generally fails. Therefore, Assumption 5 is violated in sparse multilayer settings, and Theorems 3 and 4 (which rely on accurate centrality recovery under measurement error) may no longer hold. This highlights the necessity of sufficient network density for the theoretical guarantees to apply.

Case 3: Weakly coupled layers ($\delta = \Theta(\mathsf{N})$)

Setup: Suppose each layer is a dense graph (e.g., complete or Erdős–Rényi with $p = \Theta(1)$), but inter-layer coupling is very weak or vanishing, i.e., $r = o(1)$ or zero. The supra-adjacency matrix then has a block-diagonal structure, or near block-diagonal.

Since layers are weakly coupled, the spectrum of B_0 is close to that of the block-diagonal matrix:

$$B_0 \approx \mathrm{diag}(B^{(1)}, \dots, B^{(\mathsf{L})}).$$

Each $B^{(\ell)}$ contributes a top eigenvalue $\lambda_1^{(\ell)} = \Theta(\mathsf{N})$, and due to near independence, these top eigenvalues are nearly degenerate. Thus, the gap between the largest and second-largest eigenvalues of the full B_0 becomes:

$$\delta = \lambda_1 - \lambda_2 = \Theta(\mathsf{N}).$$

With $a_{\mathsf{N}} \asymp \sqrt{\mathsf{NL}}$ and $\delta = \Theta(\mathsf{N})$, we have

$$\frac{a_{\mathsf{N}}\sqrt{\mathsf{L}}}{\delta} \asymp \sqrt{\mathsf{L}^2/\mathsf{N}}.$$

This implies the condition holds only if $\mathsf{L} = o(\mathsf{N}^{1/2})$. Hence, when the number of layers grows too fast (e.g., $\mathsf{L} \asymp \mathsf{N}^{1/2}$ or more), or inter-layer connections are extremely weak, Assumption 5 fails. This illustrates the sensitivity of our results to the strength of inter-layer coupling.

S2 Simulation supplements

This section contains simulation results, including extended plots omitted from the main text, additional simulations comparing our method with alternative models, and sensitivity analyses evaluating robustness under various settings. Specifically, the section is organized as follows:

- **Section S2.1** provides supplementary visualizations (boxplots and Q-Q plots) for the simulation results in the main text. These plots offer further insights into the consistency and asymptotic normality of the proposed estimators.

- **Section S2.2** compares our proposed methods (C-MNetR and CC-MNetR) with **Regression with Community Fixed Effects (RCFE)**. This comparison evaluates whether including community dummies directly in the regression captures the latent structural effects as effectively as our centrality-based approaches.

- **Section S2.3** benchmarks **CC-MNetR against an Aggregated Centrality baseline**, which uses node-level eigenvector centrality computed from a flattened network. This comparison focuses on the structural interpretability.

- **Section S2.4** conducts a **sensitivity analysis under varying noise levels**, examining how different magnitudes of measurement error affect the performance of CC-MNetR. This analysis illustrates the robustness of our method in more challenging, noisy environments.

Together, these results demonstrate the statistical reliability and structural advantages of CC-MNetR, especially in multilayer networks subject to measurement error.

S2.1 Boxplots and QQ-plots of coefficients for C-MNetR/CC-MNetR in Simulation part:

In this section, we present the plots of $\hat{\beta}^{(ols)}$, $\tilde{\beta}^{(ols)}$, $\hat{\beta}$, and $\tilde{\beta}$.

Noiseless case

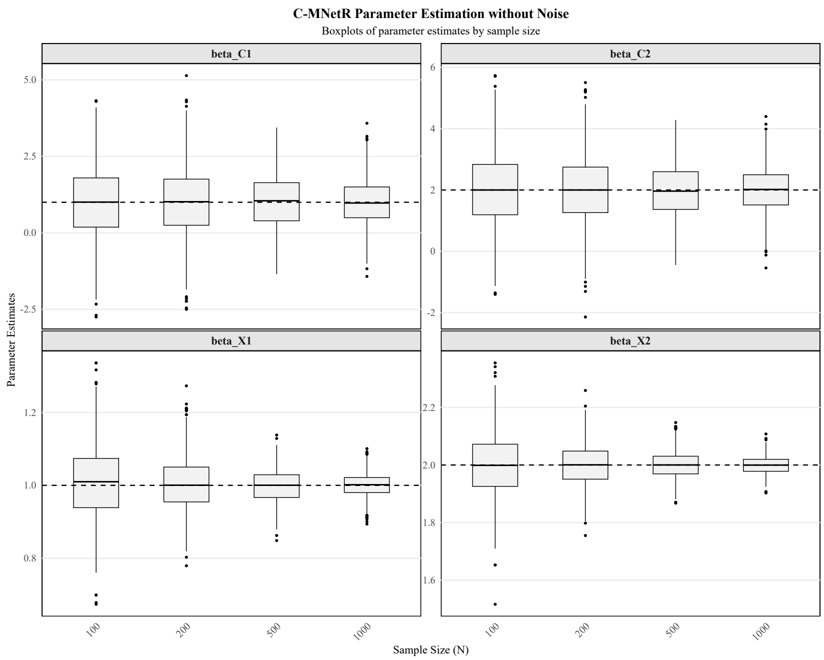


Figure 1: Boxplots of the coefficient estimates for the C-MNetR model across different sample sizes without measurement error.

The boxplots in Figure 1 and Figure 2 illustrate the distribution of the coefficient estimates for C-MNetR and CC-MNetR across different sample sizes. As the sample size increases, the estimates for all coefficients become increasingly concentrated around their true values, demonstrating improved consistency. Notably, in Figure 2, the spread of the estimates decreases with larger sample sizes, indicating reduced variability. This highlights the effectiveness of CC-MNetR in producing more reliable and stable coefficient estimates as the sample size grows.

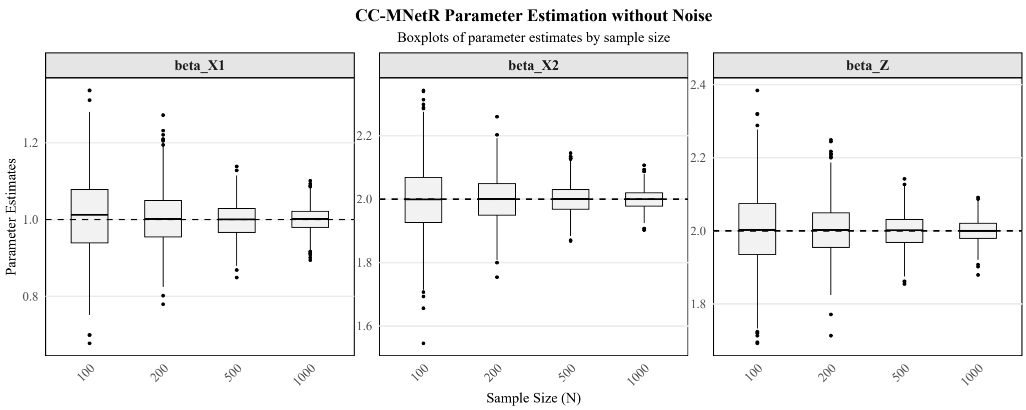


Figure 2: Boxplots of the coefficient estimates for the CC-MNetR model across different sample sizes without measurement error.

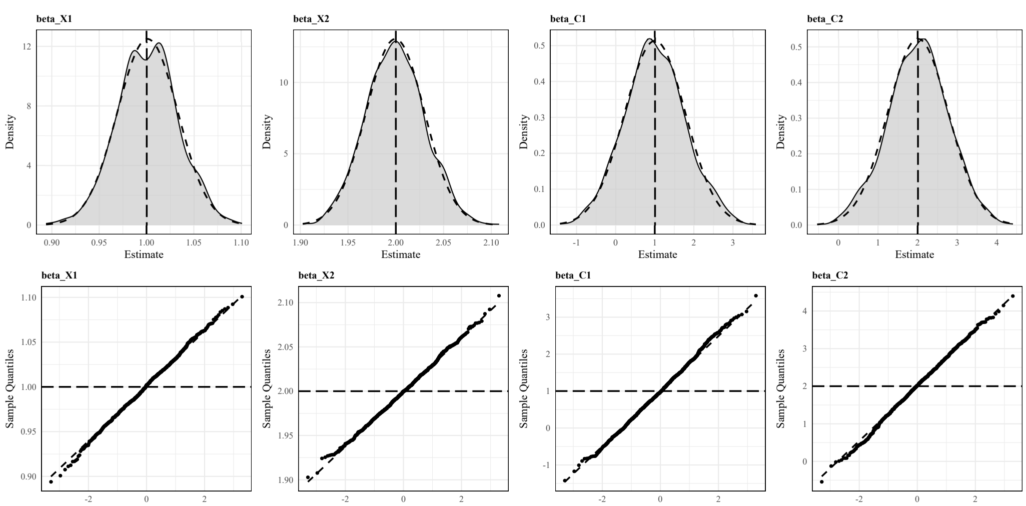


Figure 3: QQ-plot of the coefficient estimates for the C-MNetR model across different sample sizes without measurement error when $\mathsf{N} = 1000$.

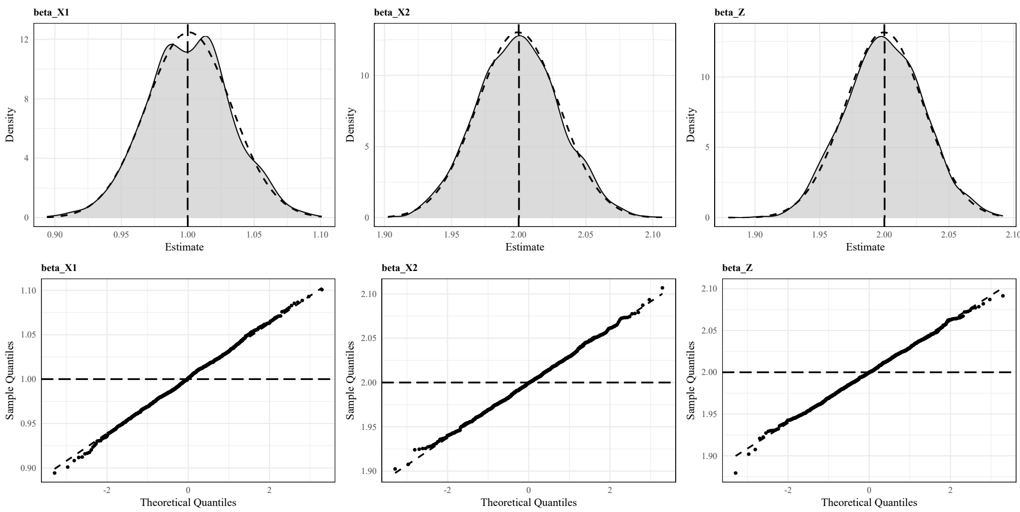


Figure 4: QQ-plot of the coefficient estimates for the CC-MNetR model without measurement error when $\mathsf{N} = 1000$.

Figure 3-4 show the coefficient estimates for C-MNetR and CC-MNetR when the sample size is $\mathsf{N} = 1000$. Both QQ-plots indicate that the coefficient estimates follow a normal distribution.

Noisy case

This subsection presents the plots of coefficient estimates for the C-MNetR and CC-MNetR models with measurement error.

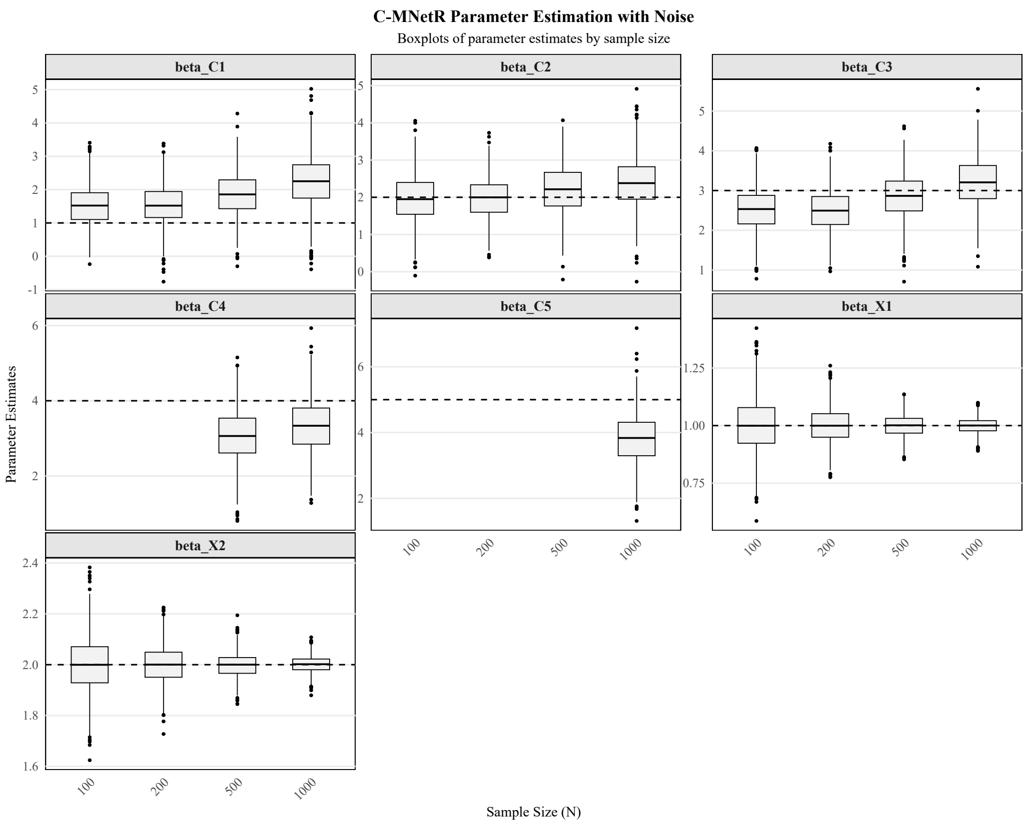


Figure 5: Boxplots of the coefficient estimates for the C-MNetR model across different sample sizes with measurement error. ($\sigma_b = 1$.)

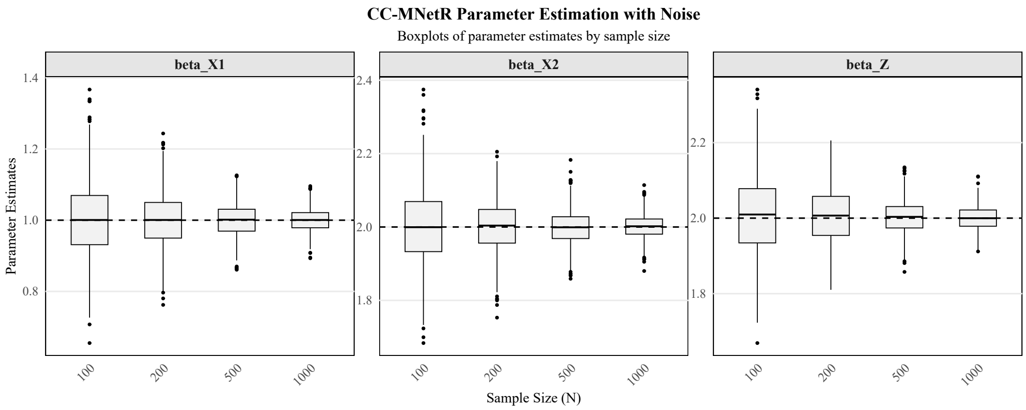


Figure 6: Boxplots of the coefficient estimates for the CC-MNetR model across different sample sizes with measurement error. ($\sigma_b = 1$.)

The boxplots in Figure 5 and Figure 6 illustrate the distribution of the coefficient estimates for C-MNetR and CC-MNetR across different sample sizes.

For C-MNetR, as shown in Figure 5, $\hat{\beta}_X$ is consistent under Assumption 5, but $\hat{\beta}_C$ exhibits persistent bias when $a_{\mathsf{N}} = \sqrt{\mathsf{NL}}$, even as N increases.

For CC-MNetR, as the sample size increases, the estimates for all coefficients become increasingly concentrated around their true values, demonstrating improved consistency. The boxplots are shown in Figure 6.

Figure 7-8 show the QQ-plots of coefficient estimates for C-MNetR and CC-MNetR when the sample size is $\mathsf{N} = 1000$. Both QQ-plots indicate that the coefficient estimates follow a normal distribution.

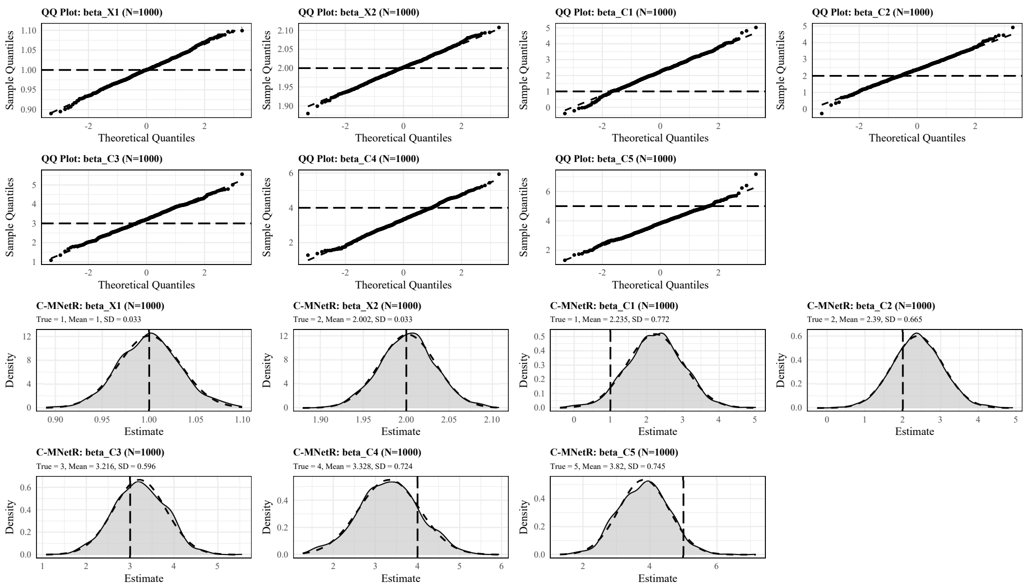


Figure 7: QQ-plot of the coefficient estimates for the C-MNetR model across different sample sizes with measurement error when $\mathsf{N} = 1000$.

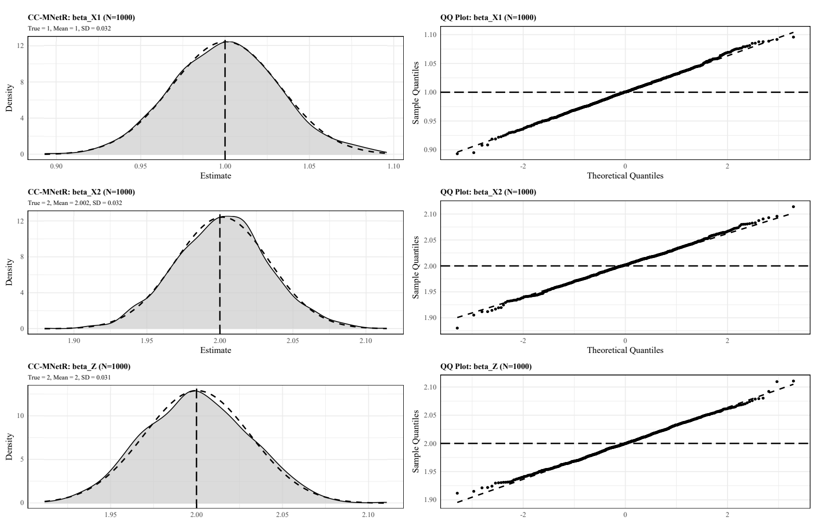


Figure 8: QQ-plot of the coefficient estimates for the CC-MNetR model with measurement error when $\mathsf{N} = 1000$.

These results highlight the robustness of the CC-MNetR model compared to the C-MNetR model, particularly in scenarios involving measurement error.

S2.2 Simulation results of comparison between RCEF and C-MNetR/CC-MNetR:

In this section, we show the simulation results of both Regression with Community Fixed Effects (RCEF) and C-MNetR/CC-MNetR, which demonstrate that RCEF does not exhibit strong consistency properties under the same conditions when C-MNetR performs great, let alone compared to the much better-performing CC-MNetR.

- **Compared to C-MNetR:** First, we have conducted a detailed comparison in the absence of measurement error between our method, C-MNetR:

  $$y = X\beta_X + C\beta_C + \varepsilon$$

  and method RCFE:

  $$y = X\beta_X + C\beta_C + S\beta_S + \varepsilon$$

  where S is the community label matrix. Results in both Table 4.1 (c) in the manuscript and Figure 9 show the consistency of $\hat{\beta}_C^{(ols)}$ with

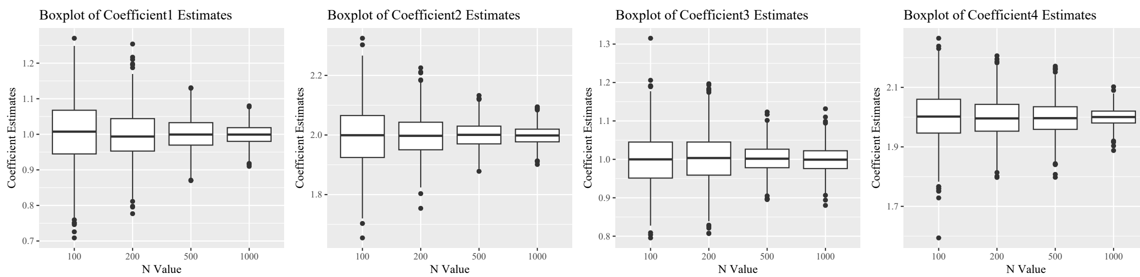


Figure 9: Boxplot of $n = 1000$ Estimates of 4 Coefficients $\beta = (\beta_{X_1}, \beta_{X_2}, \beta_{C_1}, \beta_{C_2})$ in **C-MNetR** when $a_{\mathsf{N}} = \mathsf{N}$ without measurement error. $\hat{\beta}_C$ shows consistency in this case.

$a_{\mathsf{N}} = \mathsf{N}$ in the absence of measurement error. But in Figure 10 with the same condition, $\hat{\beta}_S$ lacks consistency.

- **Compared to CC-MNetR:** Then, we have conducted a detailed comparison in the absence of measurement error between our method, CC-MNetR:

$$y = X\beta_X + Z\beta_Z + \varepsilon$$

and method RCFE:

$$y = X\beta_X + C\beta_C + S\beta_S + \varepsilon$$

where S is the community label matrix. Our findings indicate that directly regressing on community labels to obtain community fixed effects does not achieve consistency.

Specifically, as is shown in Figure 10, the standard deviation of the estimators does not decrease with increasing N. In contrast, as shown

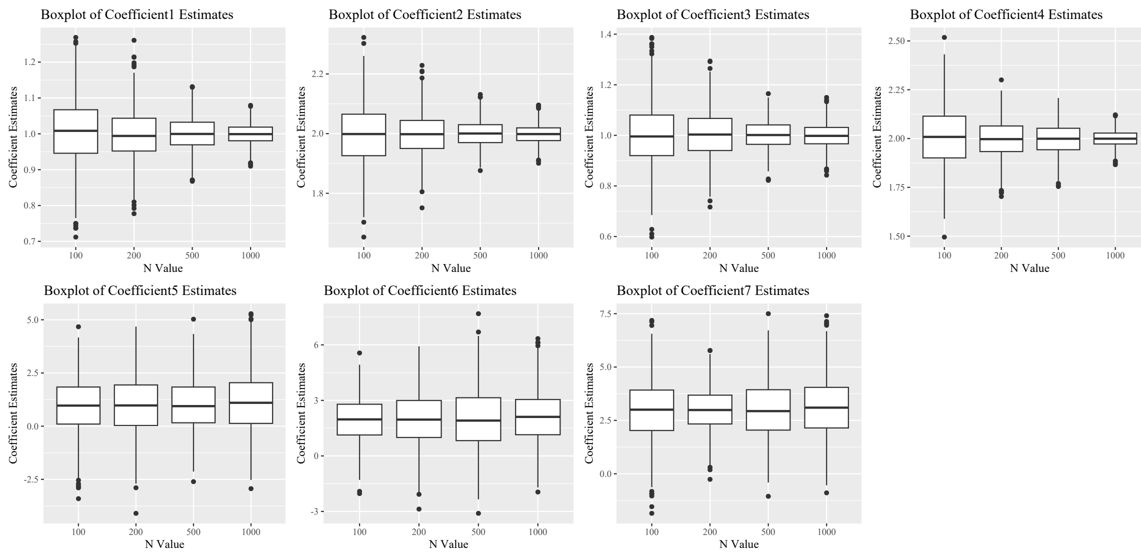


Figure 10: Boxplot of $n = 1000$ Estimates of 7 Coefficients $\beta = (\beta_{X_1}, \beta_{X_2}, \beta_{C_1}, \beta_{C_2}, \beta_{S_1}, \beta_{S_2},\ \beta_{S_3})$ in **RCFE** when $a_{\mathsf{N}} = \mathsf{N}$ without measurement error. Even though the order of a_{N} is already large, $\hat{\beta}_S$ still lacks consistency.

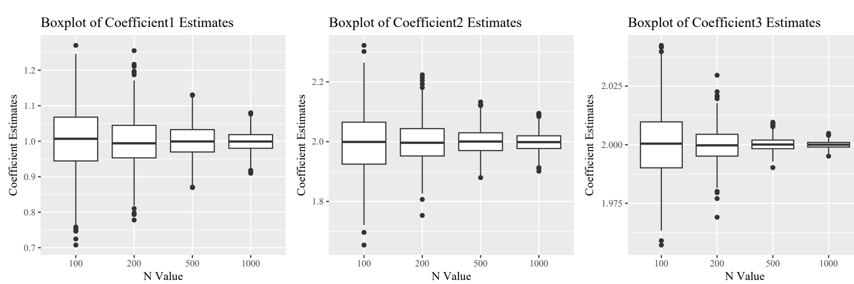


Figure 11: Boxplot of $n = 1000$ Estimates of 3 Coefficients $\beta = (\beta_{X_1}, \beta_{X_2}, \beta_Z)$ in **CC-MNetR** when $a_{\mathsf{N}} = \mathsf{N}$ without measurement error. Compared to $\hat{\beta}_S$, $\hat{\beta}_Z$ shows better consistency.

in Figure 11, simulation results of $\hat{\beta}_Z^{(ols)}$ show consistency properties since our CC-MNetR method addresses the inconsistency of $\hat{\beta}_C^{(ols)}$ by incorporating restricted community structure into the centrality measure.

S2.3 Simulation results of comparison between CC-MNetR and Aggregated Centrality baseline

We now design a simulation experiment to compare the performance of CC-MNetR against two benchmarks: (i) an oracle model using true community labels, and (ii) a natural baseline that uses node-level eigenvector centrality computed from the aggregated adjacency matrix. Although all three models show robust estimation behavior in terms of variance, their interpretability and structural validity differ substantially, especially when the outcome variable is generated from latent group-level effects.

We compare the following methods:

- **Model 1 (Oracle model):** Uses the true community label matrix S as regressors.

$$y = X\beta_X + S\beta_S + \varepsilon,$$

  This serves as a gold standard benchmark, enabling direct estimation of group-level fixed effects.

- **Model 2 (CC-MNetR):** Compute node-layer centrality using the full supra-adjacency matrix, then aggregate those values using known community structure to obtain a community-specific scalar regressor Z. This approach incorporates both layer-level variation and group structure.

- **Model 3 (Aggregated baseline):** Flatten the multilayer network into a single-layer network by averaging adjacency matrices across layers, then compute eigenvector centrality $EC_{agg,i}$ for each node i.

$$y = X\beta_X + EC_{agg}\beta_{EC_{agg}} + \varepsilon.$$

  This approach ignores inter-layer heterogeneity and assumes node importance is stable across layers.

Figure 12 shows the boxplots of coefficient estimates for the Oracle model, where both covariate and community effects are recovered accurately with increasing sample size N. This confirms the consistency and stability of the estimator when true community structure is directly used.

Figure 13 presents estimates from CC-MNetR, illustrating that the community-level centrality regressor Z effectively captures the group-level structural effects. The estimator exhibits similar stability and convergence behavior to the Oracle model, underscoring that our method preserves

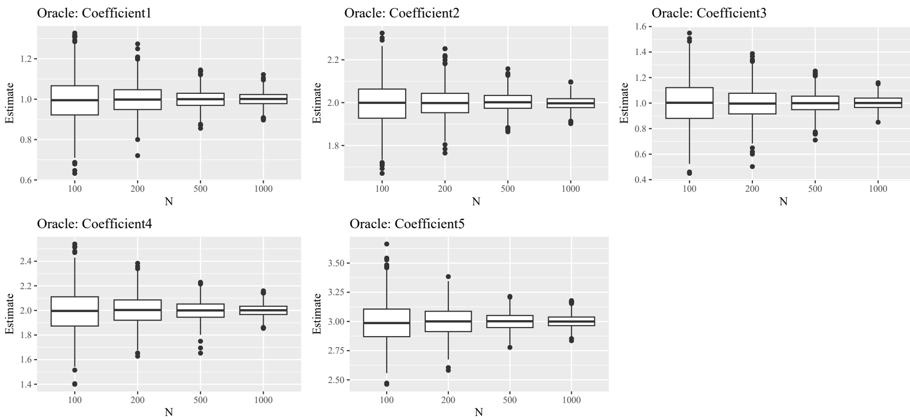


Figure 12: Boxplot of $n = 1000$ estimates of coefficients $\beta = (\beta_{X_1}, \beta_{X_2}, \beta_{S_1}, \beta_{S_2}, \beta_{S_3})$ in the **Oracle model** with regression on community label matrix S, when $a_{\mathsf{N}} = \sqrt{\mathsf{NL}}$ without measurement error. Estimators recover both covariate and community fixed effects consistently.

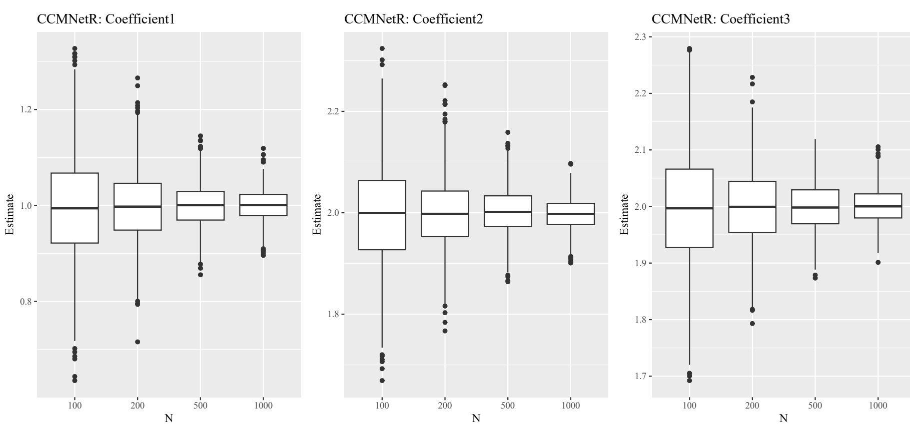


Figure 13: Boxplot of $n = 1000$ estimates of coefficients $\beta = (\beta_{X_1}, \beta_{X_2}, \beta_Z)$ in **CC-MNetR** when $a_{\mathsf{N}} = \sqrt{\mathsf{NL}}$ without measurement error. Compared to the Oracle model, the constructed community-level centrality regressor Z yields stable and consistent estimates.

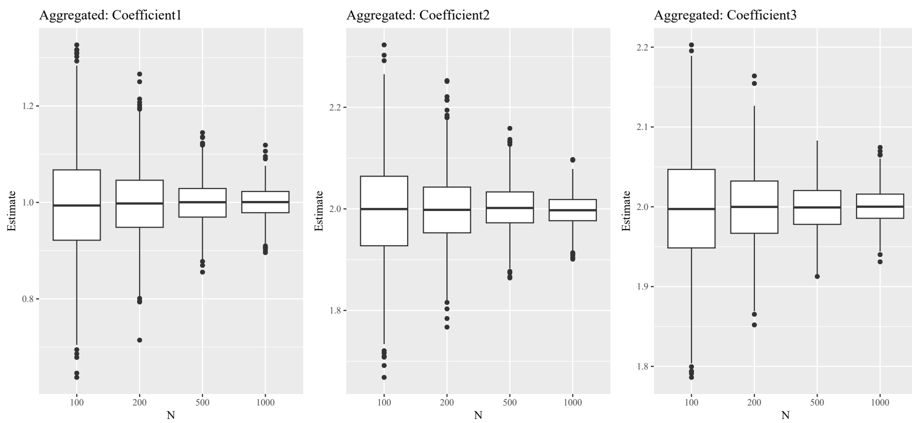


Figure 14: Boxplot of $n = 1000$ estimates of coefficients $\beta = (\beta_{X_1}, \beta_{X_2}, \beta_{EC_{agg}})$ in the **Aggregated baseline model** when $a_{\mathsf{N}} = \sqrt{\mathsf{NL}}$ without measurement error. The aggregated eigenvector centrality regressor EC_{agg} shows higher variability and less precise estimation than CC-MNetR.

meaningful cross-layer and community information.

Figure 14 depicts results for the Aggregated baseline model. While this method also demonstrates robustness and some consistency, the variability in coefficient estimates tends to be higher than CC-MNetR, especially for the network centrality coefficient. This can be attributed to the loss of inter-layer heterogeneity and the compression of multi-layer information into a single aggregated centrality measure, which dilutes the distinct community signals.

Overall, while all models are technically consistent in large samples, only CC-MNetR preserves interpretable community-level structural signals.

Unlike the Oracle model, which treats communities as discrete and unranked categories, CC-MNetR offers a meaningful scalar summary that reflects each community's importance in the network. And unlike the Aggregated baseline, CC-MNetR respects both group and layer structure, avoiding bias due to structural compression. This makes it particularly suitable for applications where community effects are both collective and structurally embedded.

S2.4 Sensitivity of CC-MNetR to Network Measurement Noise

To assess the robustness of the proposed CC-MNetR estimator in the presence of imperfect or noisy network data, we conduct a detailed sensitivity analysis with respect to measurement error. Specifically, we simulate multilayer networks with known ground-truth structure and add Gaussian noise of varying standard deviations σ_b (ranging from 0.5 to 5.0) to the network adjacency matrices to mimic network-level measurement error.

For each noise level, we recompute the centrality-based covariates from the noisy networks and refit the CC-MNetR model. The regression coefficients are held fixed across settings, allowing for a direct comparison of estimation performance. We evaluate robustness in terms of the distribution of estimated coefficients and their average mean squared error (MSE) across repetitions. All simulations are conducted across multiple sample sizes, enabling analysis of how data availability interacts with noise levels.

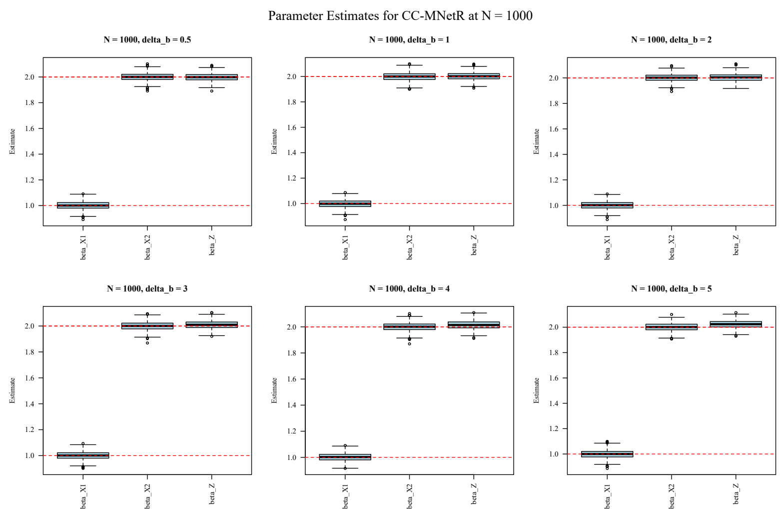
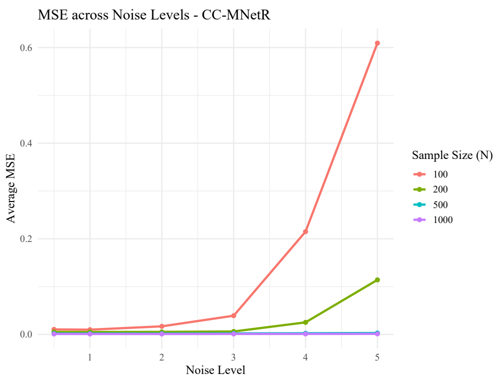
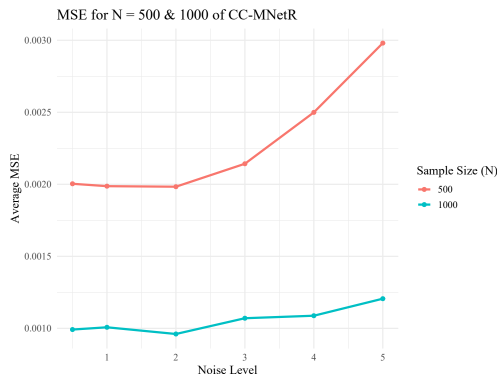


Figure 15: Boxplots of CC-MNetR coefficient estimates at $\mathsf{N} = 1000$ across different noise levels. Each subplot corresponds to a different standard deviation of the Gaussian perturbation added to the network structure. Red dashed lines indicate the true parameter values. The estimator remains stable under moderate noise and degrades smoothly with increasing variance.

Figures 15 and 16 provide empirical support for the theoretical insights of Theorem 5. In particular, they illustrate that the CC-MNetR estimator degrades gracefully under increasing noise, especially when the sample size is moderate to large and the network retains sufficient spectral structure. These diagnostics may aid practitioners in interpreting estimation reliability and understanding the potential effects of unobserved or suspected measurement error in applied settings.



(a) All sample sizes (N = 100, 200, 500, 1000). Larger MSE for small N dominates the scale.



(b) Zoom-in: only N = 500 and 1000. Fine-grained view of stable estimates.

Figure 16: Mean squared error (MSE) of CC-MNetR coefficient estimates across different noise levels σ_b. The left panel shows all sample sizes; the right panel zooms in on N = 500 and 1000 for better resolution in moderate noise settings.

S3 Real data analysis supplements

This section presents further details on the real data application based on the World Input-Output Database (WIOD), including variable descriptions, full estimation results, and a comparison and interpretation of different centrality measures.

S3.1 Comparison with Aggregated Network Centrality

To evaluate the effectiveness of our proposed multilayer community-level centrality measure, we compare it with a baseline approach that computes eigenvector centrality on a single-layer aggregated network. Specifically, the aggregated network is constructed by extracting only the diagonal intra-layer blocks from each country's input-output matrix, rescaling them individually, and then averaging across all layers. This results in a single-layer adjacency matrix that retains domestic production structure but discards cross-country interactions. Eigenvector centrality is computed for each node, and then averaged within each community to yield community-level scores.

Figures 17 and 18 present the community-level centrality scores obtained from the CC-MNetR method and the aggregated network approach, respectively, for the year 2014. Notably, our method ranks *Construction*, *Manufacturing*, and *Wholesale and retail trade; repair of motor vehicles and motorcycle* as the top three most central communities in the global production network. In contrast, the aggregated method identifies *Construction*, *Real Estate Activities*, and *Electricity, Gas, Steam and Air Conditioning Supply* as the most central sectors.

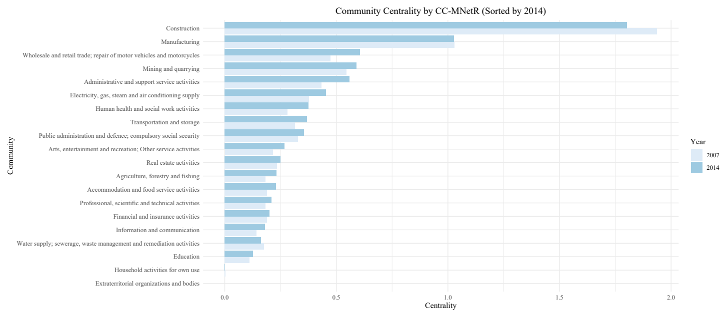


Figure 17: Community-level centrality based on CC-MNetR method (sorted by 2014).

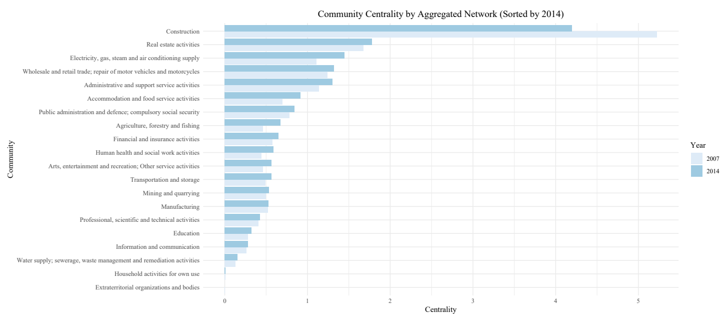


Figure 18: Community-level centrality based on Aggregated Network (sorted by 2014).

The observed discrepancy between the two rankings reflects fundamental differences in how the methods capture economic influence across global production networks. The CC-MNetR method leverages the multilayer structure of the data and explicitly incorporates cross-country input-output linkages. As a result, it highlights communities like *Manufacturing* and *Wholesale and retail trade*, which are deeply integrated into international supply chains and play vital roles in facilitating intermediate goods exchange and cross-border commerce. These sectors benefit from inter-layer dependencies that boost their structural centrality in the global context.

By contrast, the aggregated method flattens the multilayer network into a purely domestic snapshot, ignoring inter-country flows and treating each node as isolated within its national boundary. This simplification can elevate the apparent centrality of sectors such as *Real Estate Activities* and *Electricity, Gas, Steam and Air Conditioning Supply*, which may be important within individual countries but are less involved in international production networks. Consequently, the aggregated approach may overlook globally strategic industries that operate through extensive cross-national connections.

These findings underscore the added value of our community-based multilayer approach. By preserving and leveraging inter-country linkages, CC-MNetR yields a more accurate and interpretable assessment of each sector's central role in the world economy.

S3.2 Details of variables in SEA dataset:

The variable details of the SEA dataset are summarized in Table 1, and their relationships are visualized in the scatterplot (Figure 19).

Values	Description
GO	Gross output by industry at current basic prices (in millions of national currency)
II	Intermediate inputs at current purchasers' prices (in millions of national currency)
VA	Gross value added at current basic prices (in millions of national currency)
EMP	Number of persons engaged (thousands)
EMPE	Number of employees (thousands)
H_EMPE	Total hours worked by employees (millions)
COMP	Compensation of employees (in millions of national currency)
LAB	Labour compensation (in millions of national currency)
CAP	Capital compensation (in millions of national currency)
K	Nominal capital stock (in millions of national currency)

Table 1: Descriptions of 10 variables contained in SEA.

Covariate Diagnostics and Multicollinearity

We provide additional details on covariate diagnostics to support the regression analysis in Section 5. The 10 variables in the SEA dataset are listed in Table 1. After excluding *Intermediate Input* (II), which directly overlaps with the construction of centrality scores, we assess multicollinearity among the remaining 9 variables using two tools:

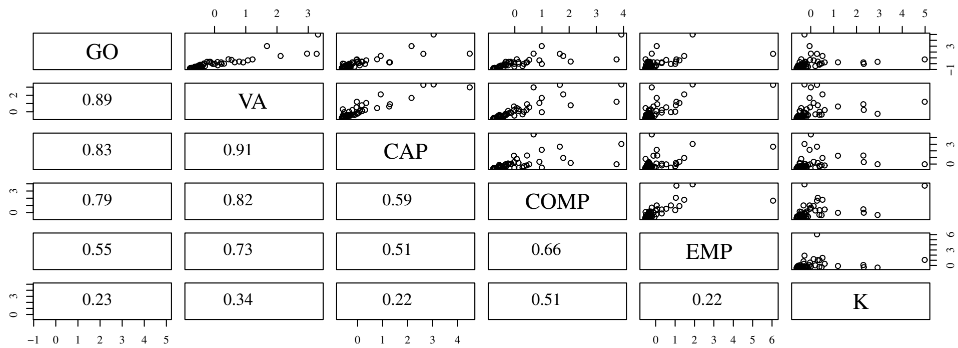


Figure 19: Scatterplot matrix of 6 variables and the lower panel of this matrix denotes the correlation coefficient between variables.

Variance Inflation Factor (VIF). We compute the VIF for each covariate X_i, defined as

$$\mathrm{VIF}(X_i) = \frac{1}{1 - R_i^2},$$

where R_i^2 is obtained from a regression of X_i on all other covariates. Table 2 below shows the VIFs calculated using the 2014 dataset. Variables with VIFs exceeding 5—namely VA (38.80), CAP (16.34), and COMP (7.03)—are considered problematic due to strong collinearity. We retain only VA among them for its interpretability and relevance.

Variable	VA	CAP	COMP	EMP	K
VIF	38.80	16.34	7.03	3.44	1.41

Table 2: VIFs of 5 selected variables based on the 2014 data.

Correlation between Z and X. To further assess the distinct contribution of the network-based centrality measure Z, we compute its empirical correlations with the remaining regressors. The centrality scores Z are weakly correlated with all covariates used in the regression model, with absolute correlations below 0.2. This indicates that Z captures structural information from the multilayer network that is not reflected in conventional production-side covariates. Therefore, the inclusion of Z provides complementary variation that enhances both the robustness and the interpretability of the model.

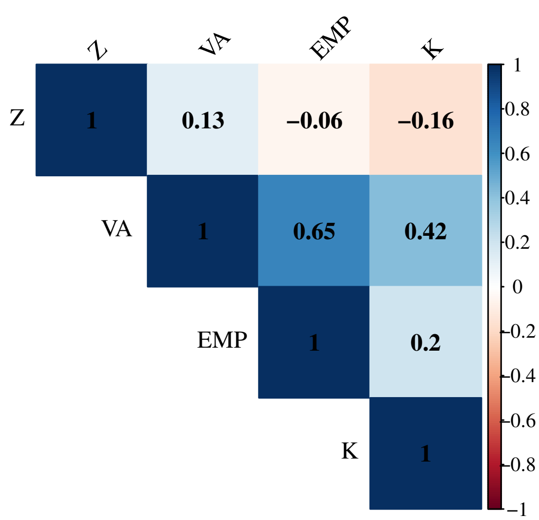


Figure 20: Correlation matrix of covariates and centrality score Z (2007 data).

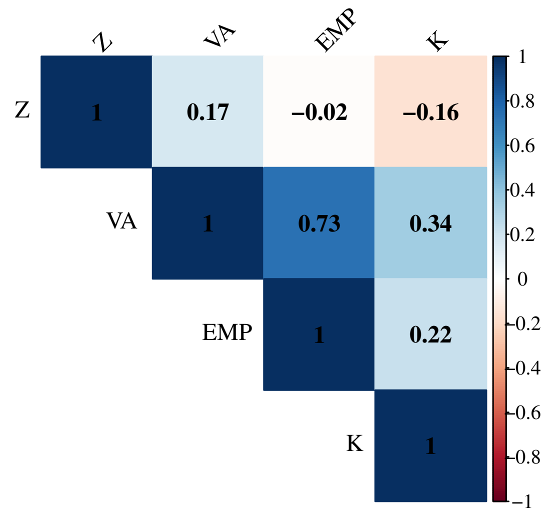


Figure 21: Correlation matrix of covariates and centrality score Z (2014 data).

Figure 20-21 present the correlation matrix of the variables included in the final regression. These diagnostics validate the choice of covariates and the stability of the regression model presented in the main text.

S3.3 Details of sectors:

According to ISIC Rev.4, industries in WIOD release 2016 are shown in Table 3.

No.	Industry	Description	Community
1	A01	Crop and animal production, hunting and related service activities	Agriculture, forestry and fishing

No.	**Industry**	**Description**	**Community**
2	A02	Forestry and logging	Agriculture, forestry and fishing
3	A03	Fishing and aquaculture	Agriculture, forestry and fishing
4	B	Mining and quarrying	Mining and quarrying
5	C10-C12	Manufacture of food products, beverages and tobacco products	Manufacturing
6	C13-C15	Manufacture of textiles, wearing apparel and leather products	Manufacturing
7	C16	Manufacture of wood and of products of wood and cork, except furniture; etc.	Manufacturing
8	C17	Manufacture of paper and paper products	Manufacturing
9	C18	Printing and reproduction of recorded media	Manufacturing

No.	Industry	Description	Community
10	C19	Manufacture of coke and refined petroleum products	Manufacturing
11	C20	Manufacture of chemicals and chemical products	Manufacturing
12	C21	Manufacture of basic pharmaceutical products and pharmaceutical preparations	Manufacturing
13	C22	Manufacture of rubber and plastic products	Manufacturing
14	C23	Manufacture of other non-metallic mineral products	Manufacturing
15	C24	Manufacture of basic metals	Manufacturing
16	C25	Manufacture of fabricated metal products, except machinery and equipment	Manufacturing
17	C26	Manufacture of computer, electronic and optical products	Manufacturing

No.	Industry	Description	Community
18	C27	Manufacture of electrical equipment	Manufacturing
19	C28	Manufacture of machinery and equipment n.e.c.	Manufacturing
20	C29	Manufacture of motor vehicles, trailers and semi-trailers	Manufacturing
21	C30	Manufacture of other transport equipment	Manufacturing
22	C31-C32	Manufacture of furniture; other manufacturing	Manufacturing
23	C33	Repair and installation of machinery and equipment	Manufacturing
24	D	Electricity, gas, steam and air conditioning supply	Electricity, gas, steam and air conditioning supply
25	E36	Water collection, treatment and supply	Water supply; sewerage, waste management and remediation activities

No.	Industry	Description	Community
26	E37-E39	Sewerage; waste collection, treatment and disposal activities; materials recovery; etc.	Water supply; sewerage, waste management and remediation activities
27	F	Construction	Construction
28	G45	Wholesale and retail trade and repair of motor vehicles and motorcycles	Wholesale and retail trade; repair of motor vehicles and motorcycles
29	G46	Wholesale trade, except of motor vehicles and motorcycles	Wholesale and retail trade; repair of motor vehicles and motorcycles
30	G47	Retail trade, except of motor vehicles and motorcycles	Wholesale and retail trade; repair of motor vehicles and motorcycles
31	H49	Land transport and transport via pipelines	Transportation and storage
32	H50	Water transport	Transportation and storage
33	H51	Air transport	Transportation and storage
34	H52	Warehousing and support activities for transportation	Transportation and storage

No.	Industry	Description	Community
35	H53	Postal and courier activities	Transportation and storage
36	I	Accommodation and food service activities	Accommodation and food service activities
37	J58	Publishing activities	Information and communication
38	J59-J60	Motion picture, video and television program production, sound recording and music publishing activities; etc.	Information and communication
39	J61	Telecommunications	Information and communication
40	J62-J63	Computer programming, consultancy and related activities; information service activities	Information and communication
41	K64	Financial service activities, except insurance and pension funding	Financial and insurance activities

No.	Industry	Description	Community
42	K65	Insurance, reinsurance and pension funding, except compulsory social security	Financial and insurance activities
43	K66	Activities auxiliary to financial services and insurance activities	Financial and insurance activities
44	L	Real estate activities	Real estate activities
45	M69-M70	Legal and accounting activities; activities of head offices; management consultancy activities	Professional, scientific and technical activities
46	M71	Architectural and engineering activities; technical testing and analysis	Professional, scientific and technical activities
47	M72	Scientific research and development	Professional, scientific and technical activities
48	M73	Advertising and market research	Professional, scientific and technical activities

No.	Industry	Description	Community
49	M74-M75	Other professional, scientific and technical activities; veterinary activities	Professional, scientific and technical activities
50	N	Rental and leasing activities, Employment activities, Travel services, security and services to buildings	Administrative and support service activities
51	O	Public administration and defence; compulsory social security	Public administration and defence; compulsory social security
52	P	Education	Education
53	Q	Human health and social work activities	Human health and social work activities
54	R-S	Creative, Arts, Sports, Recreation and entertainment activities and all other personal service activities	Arts, entertainment and recreation; Other service activities

No.	Industry	Description	Community
55	T	Activities of households as employers; undifferentiated goods- and services-producing activities of households for own use	Activities of households as employers; undifferentiated goods- and services-producing activities of households for own use
56	U	Activities of extra-territorial organizations and bodies	Activities of extraterritorial organizations and bodies

Table 3: 56 sectors and their corresponding communities in WIOD release 2016

S3.4 Regression results

Table 4 shows the estimated results of regression models for 2007 and 2014 WIOD tables.

Bibliography

Davis, C. and W. M. Kahan (1970). The rotation of eigenvectors by a perturbation. iii. *SIAM Journal on Numerical Analysis 7*(1), 1–46.

Frisch, R. and F. V. Waugh (1933). Partial time regressions as compared with individual trends. *Econometrica 1*, 387.

Graham, F. C. and M. Radcliffe (2011). On the spectra of general random graphs. *The Electronic Journal of Combinatorics 18*, 215–215.

Han, Q., K. Xu, and E. Airoldi (2015). Consistent estimation of dynamic and multi-layer block models. In *Proceedings of the 32nd International Conference on Machine Learning*, pp. 1511–1520. PMLR.

Lovell, M. C. (1963). Seasonal adjustment of economic time series and multiple regression analysis. *Journal of the American Statistical Association 58*, 993–1010.

Shi, J. and J. Malik (1997). Normalized cuts and image segmentation. *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 731–737.

von Luxburg, U. (2007). A tutorial on spectral clustering. *Statistics and Computing 17*, 395–416.

Yuan, D.-M., L.-R. Wei, and L. Lei (2014). Conditional central limit theorems for a sequence of conditional independent random variables. *Journal of the Korean Mathematical Society 51*(1), 1–15.

Variable	Estimate	Std Error	F value	p-value	Significance
Z	0.7018	0.1136	84.68	2.03e-12	***
VA	0.8905	0.0690	334.04	<2.2e-16	***
EMP	-0.0648	0.0625	1.14	0.2915	
K	0.0057	0.0528	0.01	0.9141	
Intercept	-0.3778	0.0763			

(a) Results for 2007

Variable	Estimate	Std Error	F value	p-value	Significance
Z	0.6112	0.1319	68.08	5.90e-11	**
VA	0.9517	0.0811	269.99	<2.2e-16	***
EMP	-0.1385	0.0759	3.26	0.0769	·
K	-0.0155	0.0555	0.08	0.7809	
Intercept	-0.3446	0.0896			

(b) Results for 2014

Table 4: Estimated coefficients, standard errors, and ANOVA results for 2007 and 2014.

Significance levels: *** ($p < 0.001$), ** ($p < 0.01$), * ($p < 0.05$), · ($p < 0.1$).