Supplementary Material for Network Varying Coefficient Model

The supplementary material consists of nine sections. Section S.1 introduces six lemmas. Sections S.2–S.4 demonstrate Theorems 1–3, respectively. Section S.5 presents additional discussion on Condition (C3) and discusses a related issue on collinearity. Section S.6 investigates the extension of the proposed model to accommodate the heterophilic network. Section S.7 introduces two competing methods used in our simulations. Sections S.8 and S.9 present additional data description and results in simulation studies and real data analysis, respectively.

S.1 Useful Lemmas

To prove the theorems in Sections 2 and 3, we introduce the following six useful lemmas. Lemma 1 is borrowed from Theorem 1.1 of Rudelson and Vershynin (2013), and Lemmas 2 and 3 are borrowed from Lemmas 31 and 28 of Ma et al. (2020), respectively. Then, we present the proofs of Lemmas 4 and 5. Since Lemma 6 can be demonstrated by employing similar techniques to those used in proving Lemma 5, its proof is omitted.

Lemma 1. *Let $\tilde{\varepsilon} = (\tilde{\epsilon}_1, \cdots, \tilde{\epsilon}_n)^\top \in \mathbb{R}^n$, where the $\tilde{\epsilon}_i$s for $i = 1, \cdots, n$ are independent and satisfy $\mathrm{E}(\tilde{\epsilon}_i) = 0$, and $\sup_h h^{-1/2}\left\{\mathrm{E}\left(|\tilde{\epsilon}_i|^h\right)\right\}^{1/h} < \varphi$ for a finite positive constant φ. Let U be an arbitrary $n \times n$ matrix. Then, for any $t > 0$, $Pr\left(|\tilde{\varepsilon}^\top U\tilde{\varepsilon} - \mathrm{E}\left(\tilde{\varepsilon}^\top U\tilde{\varepsilon}\right)| > t\right) \le 2\exp\left\{-c_1 \min\{t^2\varphi^{-4}\|U\|_F^{-2}, t\varphi^{-2}\|U\|_2^{-1}\}\right\}$, where c_1 is a finite positive constant.*

Lemma 2. *Let $A = (a_{ij}) \in \mathbb{R}^{n\times n}$ be the symmetric adjacency matrix of a random graph on n nodes, where a_{ij}, for any $i < j$, is independent and $a_{ii} = 0$. Let $\mathrm{E}(a_{ij}) = P_{ij}$ for all $i \ne j$, and $P_{ii} \in [0, 1]$. Then, for any $r > 0$, there exists a constant $C_A = C_A(r)$ such that $\|A - P\|_2 \le C_A\sqrt{n}$ with a probability of at least $1 - n^{-r}$.*

Lemma 3. *For any arbitrary matrices Z_1 and $Z_2 \in \mathbb{R}^{n\times k}$, we obtain $\min_{O:O^\top O=OO^\top=I_k} \|Z_1 - Z_2O\|_F^2 \le \frac{1}{2(\sqrt{2}-1)\lambda_k\left(Z_2^\top Z_2\right)}\|Z_1Z_1^\top - Z_2Z_2^\top\|_F^2$.*

Lemma 4. *Under Conditions (C1)–(C5), we have*

$$\min_{O:O^\top O=OO^\top=I_k} \|\hat{Z} - Z_0O\|_F^2 \le \tilde{c} := \frac{2c''^2}{\left(\sqrt{2}-1\right)c^{*2}\tau_3},$$

with a probability of $1-2\exp(-c_1n)-n^{-r}$, where $c^* = \min\left\{1/\delta, 2^{-1}\exp(M_2)/(1+\exp(M_2))^2\right\}$ and $c'' = \max\left\{\sqrt{(2k+4)}C_A, 2\delta^{-1}\sqrt{(\varphi^2+1)\sigma^2}\right\}$. Here, δ is a scale parameter defined in equation (5), M_2 is defined in Condition (C2), σ^2 is the variance of the random error, and c_1, C_A and r were defined in Lemmas 1 and 2.

Proof: We prove this lemma in two steps. Step I obtains an upper bound of the estimation error of Θ_0, and Step II shows the upper bound of $\min_{O:O^\top O=OO^\top=I_k} \|\hat{Z} - Z_0O\|_F^2$.

Step I: Since $\big(\hat{Z},\hat{\alpha},\hat{B},\hat{\gamma}\big)=\operatorname{argmin}_{Z,\alpha,B,\gamma}L\big(Z,\alpha,B,\gamma\big)$, we have that $L\big(\hat{Z},\hat{\alpha},\hat{B},\hat{\gamma}\big)-L\big(Z_0,\alpha_0,B_0,\gamma_0\big)\le 0$. Then, employing Taylor's expansion, we obtain

$$
\begin{aligned}
&L\left(\hat{Z},\hat{\alpha},\hat{B},\hat{\gamma}\right)-L\left(Z_0,\alpha_0,B_0,\gamma_0\right)=\sum_{i=1}^{n}\left(y_i-x_i^\top\hat{\beta}_i\right)^2/\delta-\sum_{i=1}^{n}\left(y_i-x_i^\top\beta_{0,i}\right)^2/\delta\\
&-\sum_{i,j}^{n}\left\{A_{ij}\hat{\Theta}_{ij}-\log\left(1+\exp\left(\hat{\Theta}_{ij}\right)\right)\right\}+\sum_{i,j}^{n}\left\{A_{ij}\Theta_{0,ij}-\log\left(1+\exp\left(\Theta_{0,ij}\right)\right)\right\}\\
=&\frac{1}{\delta}\sum_{i=1}^{n}\left\{\left(\hat{\beta}_i-\beta_{0,i}\right)^\top x_ix_i^\top\left(\hat{\beta}_i-\beta_{0,i}\right)-2\left(y_i-x_i^\top\beta_{0,i}\right)x_i^\top\left(\hat{\beta}_i-\beta_{0,i}\right)\right\}\\
&-\sum_{i,j}^{n}\left\{\left(A_{ij}-P_{0,ij}\right)\left(\hat{\Theta}_{ij}-\Theta_{0,ij}\right)-\frac{1}{2}\breve{P}_{ij}\left(1-\breve{P}_{ij}\right)\left(\hat{\Theta}_{ij}-\Theta_{0,ij}\right)^2\right\}\le 0,
\end{aligned}
$$

where $\breve{P}_{ij}=1/\big\{1+\exp\big(-\breve{\Theta}_{ij}\big)\big\}$, and $\breve{\Theta}_{ij}$ lies between $\hat{\Theta}_{ij}$ and $\Theta_{0,ij}$. Accordingly, we obtain

$$
\begin{aligned}
&\frac{1}{\delta}\sum_{i=1}^{n}\big(\hat{\beta}_i-\beta_{0,i}\big)^\top x_ix_i^\top\big(\hat{\beta}_i-\beta_{0,i}\big)+\sum_{i,j}^{n}\frac{1}{2}\breve{P}_{ij}\big(1-\breve{P}_{ij}\big)\big(\hat{\Theta}_{ij}-\Theta_{0,ij}\big)^2\\
\le&\sum_{i,j}^{n}\big(A_{ij}-P_{0,ij}\big)\big(\hat{\Theta}_{ij}-\Theta_{0,ij}\big)+2\delta^{-1}\sum_{i=1}^{n}\big(y_i-x_i^\top\beta_{0,i}\big)x_i^\top\big(\hat{\beta}_i-\beta_{0,i}\big).
\end{aligned}
$$

By Condition (C2), we have $1/\left\{1+\exp\left(M_2\right)\right\}\le\breve{P}_{ij}\le\exp\left(M_2\right)/\left\{1+\exp\left(M_2\right)\right\}$. In addition, denote $c^*=\min\big\{1/\delta,2^{-1}\exp\left(M_2\right)/\left(1+\exp\left(M_2\right)\right)^2\big\}$. We then have

$$
\begin{aligned}
&c^*\Big\{\|\hat{\Theta}-\Theta_0\|_F^2+\sum_{i=1}^{n}\left(\hat{\beta}_i-\beta_{0,i}\right)^\top x_ix_i^\top\left(\hat{\beta}_i-\beta_{0,i}\right)\Big\}\\
\le&\operatorname{tr}\Big\{\big(A-P_0\big)^\top\big(\hat{\Theta}-\Theta_0\big)\Big\}+2\delta^{-1}\sum_{i=1}^{n}\big(y_i-x_i^\top\beta_{0,i}\big)x_i^\top\big(\hat{\beta}_i-\beta_{0,i}\big)\\
\le&\operatorname{tr}\Big\{\big(A-P_0\big)^\top\big(\hat{\Theta}-\Theta_0\big)\Big\}+2\delta^{-1}\big(\sum_{i=1}^{n}\epsilon_i^2\big)^{1/2}\Big\{\sum_{i=1}^{n}\big(\hat{\beta}_i-\beta_{0,i}\big)^\top x_ix_i^\top\big(\hat{\beta}_i-\beta_{0,i}\big)\Big\}^{1/2}\\
\doteq&\mathcal{I}_1+\mathcal{I}_2,
\end{aligned}
$$

where $\epsilon_i=y_i-x_i^\top\beta_{0,i}$ for $i=1,\cdots,n$. The above result can be used to obtain the bound of the estimation error of Θ_0. We next consider $\mathcal{I}_1$ and $\mathcal{I}_2$ separately.

After algebraic calculations, the first term $\mathcal{I}_1$ satisfies $\operatorname{tr}\big\{\big(A-P_0\big)^\top\big(\hat{\Theta}-\Theta_0\big)\big\}\le\|A-P_0\|_2\operatorname{rank}^{1/2}\big(\hat{\Theta}-\Theta_0\big)\|\hat{\Theta}-\Theta_0\|_F\le\sqrt{2k+4}\|A-P_0\|_2\|\hat{\Theta}-\Theta_0\|_F$. By Lemma 2, with a probability of $1-n^{-r}$, we have $\|A-P_0\|_2\le C_A\sqrt{n}$. This, together with the above result, implies that $\mathcal{I}_1\le\sqrt{\left(2k+4\right)n}C_A\|\hat{\Theta}-\Theta_0\|_F$, with a probability of $1-n^{-r}$.

We next consider the second part $\mathcal{I}_2$. According to Lemma 1, with a probability of $1-2\exp\left(-c_1n\right)$, we have $\sqrt{\sum_{i=1}^n\epsilon_i^2}\le\sqrt{\left(\varphi^2+1\right)n\sigma^2}$. Denote $\mathcal{O}=\sqrt{\sum_{i=1}^n\left(\hat{\beta}_i-\beta_{0,i}\right)^\top x_ix_i^\top\big(\hat{\beta}_i-\beta_{0,i}\big)}$. Accordingly, we have $\mathcal{I}_2\le2\delta^{-1}\mathcal{O}\sqrt{\left(\varphi^2+1\right)n\sigma^2}$, with a probability of $1-2\exp\big(-c_1n\big)$.

Combining the above results, we have that, with a probability of $1-2\exp(-c_1 n)-n^{-r}$,

$$c^*\Big\{\|\hat{\Theta}-\Theta_0\|_F^2+\sum_{i=1}^n\left(\hat{\beta}_i-\beta_{0,i}\right)^\top x_i x_i^\top\left(\hat{\beta}_i-\beta_{0,i}\right)\Big\}=c^*\Big(\|\hat{\Theta}-\Theta_0\|_F^2+\mathcal{O}^2\Big)\leq c^{''}\sqrt{n}\Big(\|\hat{\Theta}-\Theta_0\|_F+\mathcal{O}\Big), \tag{S.1}$$

where $c^{''}=\max\left\{\sqrt{(2k+4)}C_A, 2/\delta\sqrt{(\varphi^2+1)\,\sigma^2}\right\}$. By the Cauchy-Schwarz inequality, we have $\|\hat{\Theta}-\Theta_0\|_F^2+\mathcal{O}^2\geq(\|\hat{\Theta}-\Theta_0\|_F+\mathcal{O})^2/2$. This, together with (S.1), leads to $\|\hat{\Theta}-\Theta_0\|_F+\mathcal{O}\leq 2c^{''}/c^*\sqrt{n}$. Accordingly, $\|\hat{\Theta}-\Theta_0\|_F\leq\|\hat{\Theta}-\Theta_0\|_F+\mathcal{O}\leq 2c^{''}/c^*\sqrt{n}$ with probability $1-2\exp(-c_1 n)-n^{-r}$, which completes the proof of Step I.

Step II: Define $\Theta=\mathcal{A}+\mathcal{Z}$, $\mathcal{A}=\alpha\mathbf{1}_n^\top+\mathbf{1}_n\alpha^\top$ and $\mathcal{Z}=ZZ^\top$. After simple calculations, we obtain that

$$\|\hat{\Theta}-\Theta_0\|_F^2=\|\hat{\mathcal{A}}-\mathcal{A}_0+\hat{\mathcal{Z}}-\mathcal{Z}_0|_F^2=\|\hat{\mathcal{A}}-\mathcal{A}_0\|_F^2+\|\hat{\mathcal{Z}}-\mathcal{Z}_0\|_F^2+2\mathrm{tr}\big\{\big(\hat{\mathcal{A}}-\mathcal{A}_0\big)\big(\hat{\mathcal{Z}}-\mathcal{Z}_0\big)\big\}.$$

By the identification condition that $J_n Z_0=Z_0$ and $J_n\hat{Z}=\hat{Z}$, we have $\mathbf{1}_n^\top Z_0=\mathbf{1}_n^\top\hat{Z}=0$. This implies that

$$\mathrm{tr}\big\{\big(\hat{\mathcal{A}}-\mathcal{A}_0\big)\big(\hat{\mathcal{Z}}-\mathcal{Z}_0\big)\big\}=\mathrm{tr}\Big\{\big(\hat{\alpha}-\alpha_0\big)\mathbf{1}_n^\top\big(\hat{Z}\hat{Z}^\top-Z_0Z_0^\top\big)+\big(\hat{\alpha}-\alpha_0\big)\big(\hat{Z}\hat{Z}^\top-Z_0Z_0^\top\big)\mathbf{1}_n\Big\}=0.$$

Accordingly, $\|\hat{\Theta}-\Theta_0\|_F^2=\|\hat{\mathcal{A}}-\mathcal{A}_0\|_F^2+\|\hat{\mathcal{Z}}-\mathcal{Z}_0\|_F^2$, which indicates that $\|\hat{Z}\hat{Z}^\top-Z_0Z_0^\top\|_F^2=\|\hat{\mathcal{Z}}-\mathcal{Z}_0\|_F^2\leq 4c^{''2}/c^{*2}n$ with a probability of $1-2\exp(-c_1 n)-n^{-r}$. This, together with Lemma 3, leads to

$$\min_{O:O^\top O=OO^\top=I_k}\|\hat{Z}-Z_0O\|_F^2\leq\frac{1}{2\left(\sqrt{2}-1\right)\lambda_k\left(Z_0^\top Z_0\right)}\|\hat{Z}\hat{Z}^\top-Z_0Z_0^\top\|_F^2\leq\frac{2c^{''2}}{\left(\sqrt{2}-1\right)c^{*2}\tau_3},$$

with a probability of $1-2\exp(-c_1 n)-n^{-r}$, which completes the entire proof.

Before introducing Lemma 5, let $\hat{O}=\mathrm{argmin}_{O:O^\top O=OO^\top=I_k}\|\hat{Z}-Z_0O\|_F^2$ and $\tilde{Z}:=(\tilde{z}_1,\cdots,\tilde{z}_n)^\top=\hat{Z}\hat{O}^\top$, where $\tilde{z}_i=(\tilde{z}_{i1},\cdots,\tilde{z}_{ik})^\top\in\mathbb{R}^k$ and $\tilde{B}=\hat{B}\hat{O}^\top$. To ease notation, we denote $\tilde{W}=\left(\mathrm{vec}\left(x_1\tilde{z}_1^\top\right),\cdots,\mathrm{vec}\left(x_n\tilde{z}_n^\top\right)\right)^\top$ and $\tilde{H}=\Big(\tilde{W},X\Big)_{n\times p(k+1)}$. In addition, let $z_{0i}\in\mathbb{R}^k$ be the true latent vector of i and $z_{0,il}$ be the l-th element of z_{0i}.

Lemma 5. *Assume that Conditions (C1)–(C5) hold and $p<(M_1^2\tilde{c})^{-1}(\sqrt{\tau_1+\tau_2}-\sqrt{\tau_2})^2n$, where $\tilde{c}$ was defined in Lemma 4, M_1 was defined in Condition (C1), and τ_1 and τ_2 were defined in Condition (C3). We then have that (i) $\|\tilde{H}-H_0\|_F^2\leq\tilde{c}M_1^2p$, and (ii) there exist two finite positive constants $\tau_1^*\leq\tau_1$ and $\tau_2^*\leq\tau_1+\tau_2$ such that $\tau_1^*<\lambda_{p(k+1)}\Big(n^{-1}\tilde{H}^\top\tilde{H}\Big)\leq\lambda_1\Big(n^{-1}\tilde{H}^\top\tilde{H}\Big)<\tau_2^*$ with a probability of $1-2\exp(-c_1 n)-n^{-r}$.*

Proof: By Lemma 4, we have

$$\begin{aligned}\|\tilde{H}-H_0\|_F^2&=\sum_{i=1}^n\|\big(\mathrm{vec}^\top(x_i\tilde{z}_i^\top),x_i^\top\big)^\top-\big(\mathrm{vec}^\top(x_iz_{0i}^\top),x_i^\top\big)^\top\|_2^2\\&=\sum_{i=1}^n\sum_{j=1}^p\sum_{l=1}^k x_{ij}^2\left(\tilde{z}_{il}-z_{0,il}\right)^2\leq M_1^2p\|\tilde{Z}-Z_0\|_F^2\leq M_1^2p\tilde{c},\end{aligned}\tag{S.2}$$

which completes Lemma 5 (i).

We next prove Lemma 5 (ii). By the Weyl inequality, we have

$$\begin{aligned}&\lambda_{p(k+1)}(n^{-1}H_0^\top H_0) - \big|\lambda_1\big(n^{-1}H_0^\top H_0 - n^{-1}\tilde{H}^\top\tilde{H}\big)\big| \le \lambda_{p(k+1)}(n^{-1}\tilde{H}^\top\tilde{H})\\ \le&\lambda_1(n^{-1}\tilde{H}^\top\tilde{H}) \le \lambda_1(n^{-1}H_0H_0^\top) + \big|\lambda_1\big(n^{-1}H_0^\top H_0 - n^{-1}\tilde{H}^\top\tilde{H}\big)\big|.\end{aligned}$$

By Condition (C3), it suffices to evaluate the bound of $\lambda_1(n^{-1}H_0^\top H_0 - n^{-1}\tilde{H}^\top\tilde{H})$ to complete the proof. After simple calculations, we have

$$\begin{aligned}\big|\lambda_1(n^{-1}H_0^\top H_0 - n^{-1}\tilde{H}^\top\tilde{H})\big| =& n^{-1}\|H_0^\top H_0 - \tilde{H}^\top\tilde{H}\|_2\\ =& n^{-1}\|(H_0-\tilde{H})^\top H_0 + (H_0-\tilde{H})^\top(H_0-\tilde{H}) + H_0^\top(H_0-\tilde{H})\|_2\\ \le& n^{-1}\big(2\|H_0\|_2\|H_0-\tilde{H}\|_2 + \|H_0-\tilde{H}\|_2^2\big)\\ \le& n^{-1}\big(2\sqrt{M_1^2p\tilde{c}\tau_2 n} + M_1^2p\tilde{c}\big) < \tau_1,\end{aligned}$$

where the fist inequality is due to the Cauchy-Schwarz inequality, and the last inequality is by the result of Lemma 5 (i). Thus, there exist $0 < \tau_1^* < \tau_1$ and $\tau_2^* < \tau_1 + \tau_2$ such that $\tau_1^* < \lambda_{p(k+1)}(n^{-1}\tilde{H}^\top\tilde{H}) \le \lambda_1(n^{-1}\tilde{H}^\top\tilde{H}) < \tau_2^*$, with a probability of $1 - 2\exp(-c_1 n) - n^{-r}$, which completes the entire proof.

Before presenting Lemma 6, denote S_1 and S_2 as the sets of nonzero rows in B_0 and nonzero elements in γ_0, respectively, and let s_1 and s_2 be their corresponding numbers of elements in S_1 and S_2. Let $x_{iS_1} = (x_{ij} : j \in S_1)$ be the subvector of x_i associated with S_1. Define $W^o = \big(\mathrm{vec}\big(x_{1S_1}\hat{z}_1^{o\top}\big), \cdots, \mathrm{vec}\big(x_{nS_1}\hat{z}_n^{o\top}\big)\big)^\top$ and $H^o = (W^o, X_{.S_2})_{n\times(s_1k+s_2)}$, where $\hat{Z}^o = (\hat{z}_1^o, \cdots, \hat{z}_n^o)^\top$ was defined in (S.9) and $X_{.S_2} = (X_{.j} : j \in S_2)$ is the submatrix of X corresponding to S_2, where $X_{.j}$ is the j-th column X. In addition, let $B_{S_1.} = (B_{i.} : i \in S_1)$ be the submatrix of B, and $\gamma_{S_2} = (\gamma_j : j \in S_2)$ be the vector of γ associated with S_2. Denote $\theta_S = (\mathrm{vec}^\top(B_{S_1.}), \gamma_{S_2}^\top)^\top$ and H_S be the submatrix of H corresponding to θ_S. Analogously, $\theta_{0,S}$, $H_{0,S}$, and $H_{0,(j)}$ can be defined under Z_0 and H_0.

Lemma 6. *Assume that Conditions (C1)–(C5) hold and* $s_1 < (M_1^2\tilde{c})^{-1}(\sqrt{\tau_1+\tau_2} - \sqrt{\tau_2})^2 n$, *where* $\tilde{c}$ *was defined in Lemma 4,* M_1 *was defined in Condition (C1), and* τ_1 *and* τ_2 *were defined in Condition (C3). We then have that (i)* $\|H^o - H_{0,S}\|_F^2 \le \tilde{c}M_1^2 s_1$, *and (ii) there exist finite positive constants* $\tau_1^{'} \le \tau_1$ *and* $\tau_2^{'} \le \tau_1 + \tau_2$ *such that* $\tau_1^{'} < \lambda_{p(k+1)}(n^{-1}H^{o\top}H^o) \le \lambda_1(n^{-1}H^{o\top}H^o) < \tau_2^{'}$ *with a probability of* $1 - 2\exp(-c_1 n) - n^{-r}$.

S.2 Proof of Theorem 1

Let $\tilde{\theta} = (\mathrm{vec}^\top(\tilde{B}), \hat{\gamma})^\top$, and denote its corresponding true parameter θ_0. To prove this theorem, we first obtain an upper bound of $\|\tilde{\theta} - \theta_0\|_2^2$. Then, using the fact that $\min_{O:O^\top O = O^\top O = I_k}\|\hat{B} - B_0O\|_F \le \|\tilde{\theta} - \theta_0\|_2$ and $\|\hat{\gamma} - \gamma_0\|_2 \le \|\tilde{\theta} - \theta_0\|_2$, we can complete the proof of Theorem 1. Note that $\tilde{\theta}$ minimizes the mean squared loss in equation (5) for given $\tilde{Z}$ defined above

Lemma 5. Then, by the first-order condition, we have $\tilde{\theta} = (\tilde{H}^\top \tilde{H})^{-1}\tilde{H}^\top Y$. Subsequently,

$$\begin{aligned}\|\tilde{\theta} - \theta_0\|_2 &= \|(\tilde{H}^\top \tilde{H})^{-1}\tilde{H}^\top (H_0\theta_0 + \varepsilon) - \theta_0\|_2 \\ &= \|(\tilde{H}^\top \tilde{H})^{-1}\tilde{H}^\top (H_0 - \tilde{H})\theta_0 + (\tilde{H}^\top \tilde{H})^{-1}\tilde{H}^\top \varepsilon\|_2 \\ &\le \|(\tilde{H}^\top \tilde{H})^{-1}\tilde{H}^\top (H_0 - \tilde{H})\theta_0\|_2 + \|(\tilde{H}^\top \tilde{H})^{-1}\tilde{H}^\top \varepsilon\|_2 \doteq I_1 + I_2.\end{aligned}$$

We next evaluate I_1 and I_2 separately. By Lemma 5 (ii), we have that

$$\begin{aligned}I_1 &\le \|n^{-1}(n^{-1}\tilde{H}^\top \tilde{H})^{-1}\tilde{H}^\top (H_0 - \tilde{H})\|_2\|\theta_0\|_2 \\ &\le \frac{1}{\tau_1^* n}\|\tilde{H}\|_2\|\tilde{H} - H_0\|_2\|\theta_0\|_2 \le \frac{\sqrt{\tau_2^*}}{\tau_1^*\sqrt{n}}\|\tilde{H} - H_0\|_2\|\theta_0\|_2. \end{aligned} \tag{S.3}$$

In addition, by Lemma 5 (i), we obtain that $\|\tilde{H} - H_0\|_F^2 \le \tilde{c}M_1^2 p$, with a probability of $1 - 2\exp(-c_1 n) - n^{-r}$. This result, together with (S.3) and $\|\tilde{H} - H_0\|_2 \le \|\tilde{H} - H_0\|_F$, leads to

$$I_1 \le \tau_1^{*-1} n^{-1/2} M_1\sqrt{\tau_2^* \tilde{c} p}\|\theta_0\|_2. \tag{S.4}$$

As for I_2, we have

$$I_2 = \|n^{-1}(n^{-1}\tilde{H}^\top \tilde{H})^{-1}\tilde{H}^\top \varepsilon\|_2 \le \frac{1}{\tau_1^* n}\|\tilde{H}^\top \varepsilon\|_2 \le \frac{1}{\tau_1^* n}\big(\|\tilde{H} - H_0\|_2\|\varepsilon\|_2 + \|H_0^\top \varepsilon\|_2\big), \tag{S.5}$$

where the first inequality is due to Lemma 5 (ii) and the second inequality follows from the triangle inequality. As given above, $\|\tilde{H} - H_0\|_2 \le \|\tilde{H} - H_0\|_F \le \sqrt{\tilde{c}M_1^2 p}$, with a probability of $1 - 2\exp(-c_1 n) - n^{-r}$. In addition, by Lemma 1, we obtain that

$$Pr\big(\|\varepsilon\|_2 \ge \sqrt{(\varphi^2 + 1)n\sigma^2}\big) \le 2\exp(-c_1 n) \text{ and} \tag{S.6}$$

$$Pr\Big(\|H_0^\top \varepsilon\|_2 \ge \sqrt{(\mathrm{tr}(H_0 H_0^\top) + t)\sigma^2}\Big) \le 2\exp\Big[-c_1 \min\big\{t^2\varphi^{-4}\|H_0H_0^\top\|_F^{-2}, t\varphi^{-2}\|H_0H_0^\top\|_2^{-1}\big\}\Big].$$

Moreover, by Condition (C3), $\mathrm{tr}\left(H_0H_0^\top\right) \le np\,(k+1)\,\tau_2$, $\|H_0H_0^\top\|_F^2 = n^2\mathrm{tr}\big\{\left(n^{-1}H_0^\top H_0\right)^2\big\} \le n^2 p\,(k+1)\,\tau_2^2$, and $\|H_0H_0^\top\|_2 = n\lambda_1\left(n^{-1}H_0^\top H_0\right) \le n\tau_2$. Let $t = c_t np\,(k+1)\,\tau_2\varphi^2$ for $c_t > 1$. Then, we obtain

$$\min\big\{t^2\varphi^{-4}\|H_0H_0^\top\|_F^{-2}, t\varphi^{-2}\|H_0H_0^\top\|_2^{-1}\big\} \ge c_t p\,(k+1)\,,$$

which immediately leads to

$$Pr\left(\|H_0^\top \varepsilon\|_2 \ge \sqrt{(c_t\varphi^2 + 1)\,\tau_2 np\,(k+1)\,\sigma^2}\right) \le 2\exp\left\{-c_1 c_t p\,(k+1)\right\}. \tag{S.7}$$

This, together with (S.5) and (S.6), leads to

$$I_2 \le \tau_1^{*-1} n^{-1/2}\Big\{M_1\sqrt{p\tilde{c}\,(\varphi^2 + 1)\,\sigma^2} + \sqrt{(c_t\varphi^2 + 1)\,\tau_2 p\,(k+1)\,\sigma^2}\Big\}. \tag{S.8}$$

Combining the results in (S.4) and (S.8), we obtain that

$$\|\tilde{\theta}-\theta_0\|_2 \leq \frac{1}{\tau_1^*\sqrt{n}}\left\{\left(\sqrt{\tau_2^*}\|\theta_0\|_2+\sqrt{(\varphi^2+1)\,\sigma^2}\right)M_1\sqrt{p\tilde{c}}+\sqrt{(c_t\varphi^2+1)\,\tau_2 p\,(k+1)\,\sigma^2}\right\}$$

with a probability of $1-2\exp\{-c_1c_tp\,(k+1)\}-2\exp(-c_1n)-n^{-r}$. Since $\min_{O:O^\top O=O^\top O=I_k}\|\hat{B}-B_0O\|_F \leq \|\tilde{\theta}-\theta_0\|_2$ and $\|\hat{\gamma}-\gamma_0\|_2 \leq \|\tilde{\theta}-\theta_0\|_2$, we have completed the proof of Theorem 1.

S.3 Proof of Theorem 2

By the definition of $\hat{\mathcal{T}}$, we have

$$\begin{aligned}
\|\hat{\mathcal{T}}-\mathcal{T}_0\|_F &\leq \|\tilde{Z}\tilde{B}^\top - Z_0B_0^\top\|_F + \|\mathbf{1}_n\hat{\gamma}^\top-\mathbf{1}_n\gamma_0^\top\|_F\\
&\leq \|\tilde{Z}\tilde{B}^\top - Z_0B_0^\top\|_F + \sqrt{n}\|\hat{\gamma}-\gamma_0\|_2\\
&\leq \|\tilde{Z}\tilde{B}^\top - Z_0\tilde{B}^\top + Z_0\tilde{B}^\top - Z_0B_0^\top\|_F + \sqrt{n}\|\hat{\gamma}-\gamma_0\|_2\\
&\leq \|\tilde{B}\|_F\|\tilde{Z}-Z_0\|_F + \|Z_0\|_F\|\tilde{B}-B_0\|_F + \sqrt{n}\|\hat{\gamma}-\gamma_0\|_2\\
&\doteq \breve{I}_1+\breve{I}_2+\breve{I}_3.
\end{aligned}$$

We study the above three parts $\breve{I}_1$, $\breve{I}_2$ and $\breve{I}_3$ separately.

By Lemma 4 and the proof of Theorem 1, we obtain

$$\breve{I}_1 \leq \left(\|B_0\|_F+\|\tilde{B}-B_0\|_F\right)\|\tilde{Z}-Z_0\|_F \leq \left(\|B_0\|_F + C_\theta/\sqrt{n}\right)\tilde{c}^{1/2}$$

with a probability of $1-2\exp\{-c_1c_tp\,(k+1)\}-2\exp(-c_1n)-n^{-r}$, where C_θ was defined in Theorem 1. In addition, by Condition (C4) and the proof of Theorem 1, we have $\breve{I}_2 \leq \sqrt{\tau_4}C_\theta$ and $\breve{I}_3 \leq C_\theta$, with a probability of $1-2\exp\{-c_1c_tp\,(k+1)\}-2\exp(-c_1n)-n^{-r}$. The above results imply that

$$\|\hat{\mathcal{T}}-\mathcal{T}_0\|_F \leq \left(\|B_0\|_F + C_\theta/\sqrt{n}\right)\tilde{c}^{1/2} + (\sqrt{\tau_4}+1)\,C_\theta,$$

with a probability of $1-2\exp\{-c_1c_tp\,(k+1)\}-2\exp(-c_1n)-n^{-r}$, which completes the proof.

S.4 Proof of Theorem 3

To prove this theorem, we first introduce some notation. Let $Q\,(Z,\alpha,\theta)$ represent the objective function to be minimized in equation (8), and it is

$$Q\,(Z,\alpha,\theta) = L(Z,\alpha,B,\gamma) + 2n\delta^{-1}\sum_{j=1}^{p}\Big\{\rho(\|B_{j,\cdot}\|_2,\mu)+\rho(|\gamma_j|,\mu)\Big\},$$

where $\theta=\left(\mathrm{vec}^\top(B),\gamma^\top\right)^\top$. Define $\theta_{(j)}=B_{j,\cdot}$ and $\theta_{(j+p)}=\gamma_j$ for $j=1,\cdots,p$. Let $H_{(j)}$ be the submatix of H corresponding to $\theta_{(j)}$. In addition, define $\breve{\theta}_{(j)}$s such that $\breve{\theta}_{(j)}=\theta_{(j)}$ if

$j \in S$ and $\breve{\theta}_{(j)} = 0$ otherwise, and $\breve{\theta}$ is the corresponding version of θ. Denote $\hat{\alpha}^o, \hat{\theta}^o$ and $\hat{Z}^o$ as the associated oracle estimators, that is

$$(\hat{Z}^o, \hat{\alpha}^o, \hat{\theta}^o) = \text{argmin}_{(Z,\alpha,\theta)\in\mathcal{F}} Q(Z, \alpha, \theta) \text{ subject to } \theta_{(j)} = 0 \text{ if } j \notin S. \tag{S.9}$$

The oracle estimates of $\theta_{(j)}$, B, and γ are denoted by $\hat{\theta}^o_{(j)}$, $\hat{B}^o$, and $\hat{\gamma}^o$, respectively, where $j = 1, \cdots, 2p$. Let $\hat{O}^o = \text{argmin}_{O:O^\top O = OO^\top = I_k} \|\hat{Z}^o - Z_0 O\|_F^2$. For the sake of simplicity, we assume $\hat{O}^o = I_k$ and the orthogonal transformations corresponding to all estimates are the identity matrix. We next prove Theorem 3 in two steps. In step I, we show that $\hat{\mathcal{T}}^o = \hat{Z}^o \hat{B}^{o\top} + \mathbf{1}_n \hat{\gamma}^{o\top}$ satisfies $\|\hat{\mathcal{T}}^o - \mathcal{T}_0\|_F = O_p(s)$. In step II, we prove that $(\hat{Z}^o, \hat{\alpha}^o, \hat{\theta}^o)$ is a local solution of equation (8).

Step I: Employing the similar techniques as those used in the proof of Lemma 4 and Theorem 1, with a probability not less than $1-2\exp(-c_1 n) - n^{-r} - 2\exp\{-c_1 c_t (s_1 k + s_2)\}$, we have that $\|\hat{Z}^o - Z_0\|_F^2 \le \tilde{c}$ and $\|\hat{\theta}^o - \theta_0\|_2 \le C'_\theta n^{-1/2}$, where $C'_\theta = \tau_1^{'-1}\Big[\big(\sqrt{\tau'_2}\,(\|B_0\|_F^2 + \|\gamma_0\|_2^2)^{1/2} + \sqrt{(\varphi^2+1)\,\sigma^2}\big) M_1 \sqrt{s_1 \tilde{c}} + \sqrt{\tau_2\,(c_t\varphi^2+1)\,(s_1 k + s_2)\sigma^2}\Big]$, τ'_1 and τ'_2 were defined in Lemma 6, $\tilde{c}$ was defined in Lemma 4, and c_t can be any constant greater than 1. In addition, applying the similar techniques to those used in proving Theorem 2, together with the above results, we obtain that

$$\begin{aligned}
\|\hat{\mathcal{T}}^o - \mathcal{T}_0\|_F &\le \|\hat{Z}^o\hat{B}^{o\top} - Z_0 B_0^\top\|_F + \|\mathbf{1}_n\hat{\gamma}^{o\top} - \mathbf{1}_n\gamma_0^\top\|_F \\
&\le \|\hat{Z}^o\hat{B}^{o\top} - Z_0 B_0^\top\|_F + \sqrt{n}\|\hat{\gamma}^o - \gamma_0\|_2 \\
&\le \|\hat{Z}^o\hat{B}^{o\top} - Z_0\hat{B}^{o\top} + Z_0\hat{B}^{o\top} - Z_0 B_0^\top\|_F + \sqrt{n}\|\hat{\gamma}^o - \gamma_0\|_2 \\
&\le \|\hat{B}^o\|_F\|\hat{Z}^o - Z_0\|_F + \|Z_0\|_F\|\hat{B}^o - B_0\|_F + \sqrt{n}\|\hat{\gamma}^o - \gamma_0\|_2 \\
&\le \left(\|B_0\|_F + C'_\theta/\sqrt{n}\right)\tilde{c}^{1/2} + (\sqrt{\tau_4}+1)C'_\theta,
\end{aligned}$$

with probability tending to 1 as $\{n, s_1, s_2\} \to \infty$. Since $k < \infty$ by definition, we then have $\|B_0\|_F = O(\sqrt{s_1})$ and $C'_\theta = O_p(\sqrt{(s_1+s_2)s_1})$, which immediately leads to $\|\hat{\mathcal{T}}^o - \mathcal{T}_0\|_F = O_p(s_1 + s_2) = O_p(s)$. This completes the proof of Step I.

Step II: Define a neighborhood of the true parameters θ_0, α_0 and Z_0 as

$$\mathcal{F}^o = \mathcal{F} \cap \Big\{Z, \alpha, \theta : \|Z - Z_0\|_F^2 \le \tilde{c}, \|\theta - \theta_0\|_2 \le C'_\theta n^{-1/2}\Big\}.$$

Let event $\mathbb{E}_{v1}$ be $(\hat{Z}^o, \hat{\alpha}^o, \hat{\theta}^o) \in \mathcal{F}^o$. According to the result of Step I, the probability of $\mathbb{E}_{v1}$ is not less than $1 - 2\exp(-c_1 n) - n^{-r} - 2\exp\{-c_1 c_t (s_1 k + s_2)\}$. To prove that $(\hat{Z}^o, \hat{\alpha}^o, \hat{\theta}^o)$ is a local solution of equation (8), it suffices to show the following two parts:

Part (i). For any $(Z, \alpha, \theta) \in \mathcal{F}^o$,

$$Q(Z, \alpha, \breve{\theta}) \ge Q(\hat{Z}^o, \hat{\alpha}^o, \hat{\theta}^o) \text{ over the event } \mathbb{E}_{v1};$$

Part (ii). For any $(Z, \alpha, \theta) \in \mathcal{F}^o$, there exists an event $\mathbb{E}_{v2}$ with high probability, such that

$$Q(Z, \alpha, \theta) \ge Q(Z, \alpha, \breve{\theta}) \text{ over the event } \mathbb{E}_{v1} \cap \mathbb{E}_{v2}.$$

To prove **Part (i)**, we have that

$$
\begin{aligned}
& Q(Z,\alpha,\breve{\theta}) - Q(\hat{Z}^o,\hat{\alpha}^o,\hat{\theta}^o) \\
=& L(Z,\alpha,\breve{\theta}) - L(\hat{Z}^o,\hat{\alpha}^o,\hat{\theta}^o) + 2n\delta^{-1}\sum_{j=1}^{2p}\rho(\|\breve{\theta}_{(j)}\|_2,\mu) - 2n\delta^{-1}\sum_{j=1}^{p}\rho(\|\hat{\theta}^o_{(j)}\|_2,\mu) \\
\doteq& \mathbb{M}_1 + \mathbb{M}_2, \qquad \text{(S.10)}
\end{aligned}
$$

where $\mathbb{M}_1 = L(Z,\alpha,\breve{\theta}) - L(\hat{Z}^o,\hat{\alpha}^o,\hat{\theta}^o)$ and $\mathbb{M}_2 = 2n\delta^{-1}\sum_{j=1}^{2p}\rho(\|\breve{\theta}_{(j)}\|_2,\mu) - 2n\delta^{-1}\sum_{j=1}^{p}\rho(\|\hat{\theta}^o_{(j)}\|_2,\mu)$. We next study $\mathbb{M}_1$ and $\mathbb{M}_2$ separately.

By the definition of oracle estimators, we have that $\mathbb{M}_1 \geq 0$. As for $\mathbb{M}_2$, for any $j \notin S$, $\breve{\theta}_{(j)} = \hat{\theta}^o_{(j)} = 0$, which leads to $\rho(\|\breve{\theta}_{(j)}\|_2,\mu) = \rho(\|\hat{\theta}^o_{(j)}\|_2,\mu) = 0$. In addition, for any $j \in S$, by Condition (C6), $\|\breve{\theta}_{(j)}\|_2 \geq \|\theta_{0,(j)}\|_2 - \|\theta_{0,(j)} - \breve{\theta}_{(j)}\|_2 > \kappa\mu$. As a result, $\rho(\|\breve{\theta}_{(j)}\|_2,\mu) = \mu\int_0^{\|\breve{\theta}_{(j)}\|_2}(1 - x/(\mu\kappa))_+ d_x = \mu^2\kappa/2$. Analogously, we can show that, over $\mathbb{E}_{v1}$, $\|\hat{\theta}^o_{(j)}\|_2 \geq \|\theta_{0,(j)}\|_2 - \|\theta_{0,(j)} - \hat{\theta}^o_{(j)}\|_2 > \kappa\mu$. Then, we obtain that $\rho(\|\hat{\theta}^o_{(j)}\|_2,\mu) = \mu^2\kappa/2$, which immediately leads to $\mathbb{M}_2 = 0$. This, together with $\mathbb{M}_1 \geq 0$ and (S.10), completes the proof of **Part (i)**.

We next demonstrate **Part (ii)**. Define $S_\theta = \{j : \theta_{(j)} \neq 0\} \cap S^c$. By Taylor's expansion, we have that

$$
\begin{aligned}
& Q(Z,\alpha,\theta) - Q(Z,\alpha,\breve{\theta}) \\
=& L(Z,\alpha,\theta) - L(Z,\alpha,\breve{\theta}) + 2n\delta^{-1}\sum_{j\in S_\theta}\rho\left(\|\theta_{(j)}\|_2,\mu\right) \\
=& 2\sum_{j\in S_\theta}\left(Y - H\theta^*\right)^\top H_{(j)}(\theta_{(j)} - \breve{\theta}_{(j)}) + 2n\delta^{-1}\dot{\rho}\left(\|\theta^*_{(j)}\|_2,\mu\right)\frac{\theta^*_{(j)}}{\|\theta^*_{(j)}\|_2}(\theta_{(j)} - \breve{\theta}_{(j)}) \\
=& 2\sum_{j\in S_\theta}\left(Y - H\theta^*\right)^\top H_{(j)}(\theta_{(j)} - \breve{\theta}_{(j)}) + 2n\delta^{-1}\sum_{j\in S_\theta}\dot{\rho}\left(\|\theta^*_{(j)}\|_2,\mu,\right)\|\theta_{(j)} - \breve{\theta}_{(j)}\|_2 \\
\doteq& 2\bar{M}_1 + 2\bar{M}_2,
\end{aligned}
$$

where $\theta^* = \iota\theta + (1-\iota)\breve{\theta}$ for some $\iota \in (0,1)$, and $\dot{\rho}(\cdot)$ is the derivative function of $\rho(\cdot)$.

We evaluate $\bar{M}_1$ and $\bar{M}_2$ separately. By the definition of MCP, $\dot{\rho}(x,\mu) \geq 0$ for any $x > 0$. This implies that $\bar{M}_2 \geq 0$. Furthermore, we obtain

$$
\bar{M}_2 = n\delta^{-1}\sum_{j\in S_\theta}\dot{\rho}\left(\|\theta^*_{(j)}\|_2,\mu\right)\|\theta_{(j)} - \breve{\theta}_{(j)}\|_2 = n\delta^{-1}\sum_{j\in S_\theta}\mu\left(1 - \|\theta^*_{(j)}\|_2\,(\mu\kappa)^{-1}\right)_+\|\theta_{(j)} - \breve{\theta}_{(j)}\|_2.
$$

Note that $\theta^* = \iota\theta + (1-\iota)\breve{\theta}$ for some $\iota \in (0,1)$, which leads to $\|\theta^*_{(j)}\|_2 \leq \iota\|\theta_{(j)}\|_2 + (1-\iota)\|\breve{\theta}_{(j)}\|_2$. By the definition of $\breve{\theta}$, we have $\|\breve{\theta}_{(j)}\|_2 \leq \|\theta_{(j)}\|_2$. Thus, $\|\theta^*_{(j)}\|_2 \leq \|\theta_{(j)}\|_2$. For $j \in S_\theta$, we have $\theta_{0,(j)} = 0$. For any $(Z,\alpha,\theta) \in \mathcal{F}^o$, we have $\max_{j\in S_\theta}\|\theta_{(j)}\|_2 = \max_{j\in S_\theta}\|\theta_{(j)} - \theta_{0(j)}\|_2 \leq \|\theta - \theta_0\|_2 \leq C'_\theta n^{-1/2}$. Thus, $\max_{j\in S_\theta}\|\theta^*_{(j)}\|_2 \leq C'_\theta n^{-1/2}$. By Condition (C6), we can obtain $C'_\theta n^{-1/2} \leq C_\mu\mu\kappa$ and $C_\mu < 1$. Accordingly, we have

$$
\bar{M}_2 \geq (n/\delta)\,(1 - C_\mu)\mu\sum_{j\in S_\theta}\|\theta_{(j)} - \breve{\theta}_{(j)}\|_2 > 0.
$$

As for $\bar{M}_1$, we employ the Hölding inequality and obtain that

$$
\begin{aligned}
|\bar{M}_1| &\leq \sum_{j\in S_\theta} \|H_{(j)}^\top (Y - H\theta^*)\|_2 \|\theta_{(j)} - \breve{\theta}_{(j)}\|_2 \\
&\leq \sum_{j\in S_\theta} \|H_{(j)}^\top (\varepsilon + H_0\theta_0 - H\theta_0 + H\theta_0 - H\theta^*)\|_2 \|\theta_{(j)} - \breve{\theta}_{(j)}\|_2 \\
&\leq \sum_{j\in S_\theta} \big(\|H_{(j)}^\top \varepsilon\|_2 + \|H_{(j)}\|_2 \|H_{0,S} - H_S\|_2 \|\theta_{0,S}\|_2 + \|H_{(j)}\|_2 \|H\|_2 \|\theta_0 - \theta^*\|_2\big) \|\theta_{(j)} - \breve{\theta}_{(j)}\|_2.
\end{aligned}
$$

Since $(Z, \alpha, \theta) \in \mathcal{F}^o$, we have

$$
\max_{j\in S_\theta} \|H_{(j)} - H_{0,(j)}\|_2 \leq \max_{j\in S_\theta} \Big\{ \sum_{i=1}^{n} \sum_{l=1}^{k} x_{ij}^2 (z_{il} - z_{0,il})^2 \Big\}^{1/2} \leq M_1 \|Z - Z_0\|_F \leq M_1 \tilde{c}^{1/2}.
$$

In addition, Condition (C3) implies that $\max_j \|H_{0,(j)}\|_2 \leq \sqrt{\tau_2 n}$. Accordingly, $\max_j \|H_{(j)}\|_2 \leq \max_j \|H_{0,(j)} - H_{(j)}\|_2 + \max_j \|H_{0,(j)}\|_2 \leq M_1 \tilde{c}^{1/2} + \sqrt{\tau_2 n}$. Analogously, we obtain that

$$
\|H_{0,S} - H_S\|_2 \leq M_1 \sqrt{s_1 \tilde{c}}, \text{ and } \|H\|_2 \leq M_1 \sqrt{p\tilde{c}} + \sqrt{\tau_2 n}.
$$

Combining the above results, we then have

$$
\begin{aligned}
|\bar{M}_1| &\leq \sum_{j\in S_\theta} \Big[\|H_{(j)}^\top \varepsilon\|_2 + \big(\sqrt{\tau_2 n} + M_1 \tilde{c}^{1/2}\big) \Big\{ M_1 \sqrt{s_1 \tilde{c}} \|\theta_0\|_2 + n^{-1/2} C_\theta' (M_1 \sqrt{p\tilde{c}} + \sqrt{\tau_2 n}) \Big\} \Big] \|\theta_{(j)} - \breve{\theta}_{(j)}\|_2 \\
&= \sum_{j\in S_\theta} \Xi_j \|\theta_{(j)} - \breve{\theta}_{(j)}\|_2.
\end{aligned}
$$

Subsequently, we define the event $\mathbb{E}_{v2} = \{\max_{j\in S_\theta} J_j \leq \tilde{\mu}\sigma\}$, where $J_j = \|H_{(j)}^\top \varepsilon\|_2$ and $\tilde{\mu} = (n/\delta)(1 - C_\mu)\mu/\sigma - (\sqrt{\tau_2 n} + M_1 \tilde{c}^{1/2})\{M_1 \sqrt{s_1} \tilde{c}^{1/2} \|\theta_0\|_2 + n^{-1/2} C_\theta' (M_1 \sqrt{p\tilde{c}} + \sqrt{\tau_2 n})\}/\sigma$ was defined in Condition (C7). Over $\mathbb{E}_{v2}$,

$$
\bar{M}_1 + \bar{M}_2 \geq \bar{M}_2 - |\bar{M}_1| \geq \sum_{j\in S_\theta} ((n/\delta)(1 - C_\mu)\mu - \Xi_j) \|\theta_{(j)} - \breve{\theta}_{(j)}\|_2 \geq 0.
$$

Accordingly, over $\mathbb{E}_{v2} \cap \mathbb{E}_{v1}$, we have that $Q(Z, \alpha, \theta) \geq Q(Z, \alpha, \breve{\theta})$, which completes the proof of **Part (ii)**.

We finally show the bound of the probability of $\mathbb{E}_{v2}$. For J_j, we have

$$
\begin{aligned}
Pr(\max_{j\in S_\theta} J_j > \tilde{\mu}\sigma) \leq & Pr\left[\max_{j\in S_\theta} \left\{ \|H_{(j)} - H_{0,(j)}\|_2 \|\varepsilon\|_2 + \|H_{0,(j)}^\top \varepsilon\|_2 \right\} \geq \tilde{\mu}\sigma \right] \\
\leq & Pr\left[\max_{j\in S_\theta} \|H_{(j)} - H_{0,(j)}\|_2 \|\varepsilon\|_2 \geq 2^{-1} \tilde{\mu}\sigma \right] \\
& + Pr\left[\max_{j\in S_\theta} \|H_{0,(j)}^\top \varepsilon\|_2 \geq 2^{-1} \tilde{\mu}\sigma \right]. \qquad \text{(S.11)}
\end{aligned}
$$

For the first part of (S.11), we have

$$
\begin{aligned}
&Pr\Big[\max_{j\in S_\theta}\|H_{(j)}-H_{0,(j)}\|_2\|\varepsilon\|_2 \geq 2^{-1}\tilde{\mu}\sigma\Big]\\
\leq &Pr\big\{\|\varepsilon\|_2 > (2M_1\tilde{c}^{1/2})^{-1}\tilde{\mu}\sigma\big\}\\
= &Pr\Big\{\|\varepsilon/\sigma\|_2^2 - n > (2M_1\tilde{c}^{1/2})^{-2}\tilde{\mu}^2 - n\Big\}.
\end{aligned}
$$

Let $t=(2M_1\tilde{c}^{1/2})^{-2}\tilde{\mu}^2-n$ in Lemma 1. Then, by Lemma 1, we have

$$
Pr\big\{\|\varepsilon\|_2 > (2M_1\tilde{c}^{1/2})^{-1}\tilde{\mu}\sigma\big\} \leq 2\exp\Big\{-c_1\min(t^2\varphi^{-4}n^{-1}, t\varphi^{-2})\Big\}.
$$

Note that $\tilde{\mu}^2 > 4M_1^2\tilde{c}\,(\varphi^2+1)\,n$, we then have $Pr\big\{\|\varepsilon\|_2 > (2M_1\tilde{c}^{1/2})^{-1}\tilde{\mu}\sigma\big\} \leq 2\exp\{-c_1 n\}$, which immediately leads to

$$
Pr\Big[\max_{j\in S_\theta}\|H_{(j)}-H_{0,(j)}\|_2\|\varepsilon\|_2 \geq 2^{-1}\tilde{\mu}\sigma\Big] \leq 2\exp\{-c_1 n\}. \tag{S.12}
$$

For the second part of (S.11), we have

$$
Pr\Big[\max_{j\in S_\theta}\|H_{0,(j)}^\top\varepsilon\|_2 \geq 1/2\tilde{\mu}\sigma\Big] \leq \sum_{j=1}^{p} Pr\big(\varepsilon^\top H_{0,(j)}H_{0,(j)}^\top\varepsilon \geq 1/4\tilde{\mu}^2\sigma^2\big).
$$

By Condition (C3), we have that $\|H_{0,(j)}H_{0,(j)}^\top\|_F^2 \leq n^2k\tau_2^2$, $\|H_{0,(j)}H_{0,(j)}^\top\|_2 \leq n\tau_2$, and $\mathrm{tr}\big(H_{0,(j)}H_{0,(j)}^\top\big) \leq kn\tau_2$. Then, we obtain

$$
\begin{aligned}
&Pr\left[\left(\varepsilon^\top/\sigma\right)H_{0,(j)}H_{0,(j)}^\top\left(\varepsilon/\sigma\right) - \mathrm{tr}\left(H_{0,(j)}H_{0,(j)}^\top\right) \geq 1/4\tilde{\mu}^2 - \mathrm{tr}\left(H_{0,(j)}H_{0,(j)}^\top\right)\right]\\
\leq &Pr\left[\left(\varepsilon^\top/\sigma\right)H_{0,(j)}H_{0,(j)}^\top\left(\varepsilon/\sigma\right) - \mathrm{tr}\left(H_{0,(j)}H_{0,(j)}^\top\right) \geq 1/4\tilde{\mu}^2 - k\tau_2 n\right].
\end{aligned}
$$

Set $t=\tilde{\mu}^2/4-k\tau_2 n$ in Lemma 1. Then, by Lemma 1, we have

$$
\begin{aligned}
&Pr\left[\left(\varepsilon^\top/\sigma\right)H_{0,(j)}H_{0,(j)}^\top\left(\varepsilon/\sigma\right) - \mathrm{tr}\left(H_{0,(j)}H_{0,(j)}^\top\right) \geq 1/4\tilde{\mu}^2 - k\tau_2 n\right]\\
\leq &2\exp\Big\{-c_1\min\left(t^2\varphi^{-4}\|H_{0,(j)}H_{0,(j)}^\top\|_F^{-2}, t\varphi^{-2}\|H_{0,(j)}H_{0,(j)}^\top\|_2^{-1}\right)\Big\}.
\end{aligned}
$$

Since $\tilde{\mu}^2 > 4\left(2\varphi^2\log p/(c_1k)+1\right)k\tau_2 n$, we then have that

$$
Pr\Big[\max_{j\in S_\theta}\|H_{0,(j)}^\top\varepsilon\|_2 \geq \tilde{\mu}\sigma/2\Big] \leq 2p\exp\Big\{-c_1(\frac{1}{4}\tilde{\mu}^2-k\tau_2 n)\varphi^{-2}(n\tau_2)^{-1}\Big\} \leq 2p^{-1}. \tag{S.13}
$$

By (S.11), (S.12) and (S.13), we have that the probability of $\mathbb{E}_{v2}$ is not less than $1-2\exp(-c_1n)-2p^{-1}$, and the probability of $\mathbb{E}_{v1}\cap\mathbb{E}_{v2}$ is not less than $1-4\exp(-c_1n)-n^{-r}-2\exp\{-c_1c_t(s_1k+s_2)\}-2p^{-1}$. Combining the above results, we have completed the entire proof.

S.5 Additional discussion on Condition (C3)

To ensure the identifiability of the regression model, we introduce $\lambda_{p(k+1)}\left(n^{-1}H_0^\top H_0\right) > \tau_1$ in the first part of Condition (C3), where τ_1 is a finite positive constant. This section establishes the relationship between Condition (C3) and the collinearity between X and Z_0.

Define $\breve{z}_i = (1, z_{0i}^\top)^\top$. After straightforward calculations, we have $\lambda_{p(k+1)}\left(n^{-1}H_0^\top H_0\right) > \tau_1$, which is equivalent to $\lambda_{p(k+1)}\left\{\frac{1}{n}\sum_{i=1}^n\left((x_i x_i^\top)\otimes(\breve{z}_i\breve{z}_i^\top)\right)\right\} > \tau_1$. For the sake of simplicity and to illustrate collinearity, we assume that (x_i, z_{0i}) are independent and identically distributed random vectors. To characterize the dependence between x_i and z_{0i}, we assume that

$$x_i = f_z\left(z_{0i}\right) + \xi_i, \text{ for } i = 1, \cdots, n, \tag{S.14}$$

where $f_z : \mathbb{R}^k \to \mathbb{R}^p$ is an unknown function, $\xi_i = (\xi_{i1}, \cdots, \xi_{ip})^\top$ has mean zero and covariance Σ_ξ, and the ξ_is are independent of the z_{0i}s. Model (S.14) includes the following two special cases.

Case (a): If $f_z\left(z_{0i}\right) = C_z z_{0i}$ for some matrix $C_z \in \mathbb{R}^{p\times k}$, then model (S.14) is the same as that in Binkiewicz et al. (2017) except that z_{0i} is the location of node i rather than the membership vector.

Case (b): Consider the case that there are repeated measures in the nodal covariates and latent positions. Let $v_i \in \mathbb{R}^l$ be the repeated measure with $l \le k$. Assume that $z_{0i} = \mathbb{C}_z v_i + \xi_{zi}$ and $x_i = \mathbb{C}_x v_i + \xi_{xi}$, where $\mathbb{C}_z \in \mathbb{R}^{k\times l}$ and $\mathbb{C}_x \in \mathbb{R}^{p\times l}$ are two matrices. Without loss of generality, we assume that the first l rows of $\mathbb{C}_z$, $\mathbb{C}_{z,1:l,\cdot}$, are invertible. Denote $\mathbb{C}_v = (\mathbb{C}_{z,1:l,\cdot}^{-1}, 0) \in \mathbb{R}^{l\times k}$. We then have $x_i = \mathbb{C}_x\mathbb{C}_v z_{0i} - \mathbb{C}_x\mathbb{C}_v\xi_{zi} + \xi_{x_i}$. Let $f_z(z_{0i}) = \mathbb{C}_x\mathbb{C}_v z_{0i}$ and $\xi_i = -\mathbb{C}_x\mathbb{C}_v\xi_{zi} + \xi_{x_i}$. As a result, Case (b) is a special case of (S.14).

By the law of large numbers, $\frac{1}{n}\sum_{i=1}^n\left\{\left(x_i x_i^\top\right)\otimes\left(\breve{z}_i\breve{z}_i^\top\right)\right\}$ converges to $E\left\{\left(x_i x_i^\top \otimes\left(\breve{z}_i\breve{z}_i^\top\right)\right\}\right)$. After algebraic simplification, we obtain that

$$\begin{aligned}
E\left\{(x_i x_i^\top)\otimes(\breve{z}_i\breve{z}_i^\top)\right\} =& E\left\{(f_z(z_{0i})f_z(z_{0i})^\top)\otimes(\breve{z}_i\breve{z}_i^\top)\right\} + E\left\{(f(z_{0i})\xi_i^\top)\otimes(\breve{z}_i\breve{z}_i^\top)\right\} \\
&+ E\left\{(\xi_i f(z_{0i})^\top)\otimes(\breve{z}_i\breve{z}_i^\top)\right\} + E\left\{(\xi_i\xi_i^\top)\otimes(\breve{z}_i\breve{z}_i^\top)\right\} \\
=& E\left\{(f(z_{0i})f(z_{0i})^\top)\otimes(\breve{z}_i\breve{z}_i^\top)\right\} + E\left\{(\xi_i\xi_i^\top)\otimes(\breve{z}_i\breve{z}_i^\top)\right\} \\
=& E\left\{(f(z_{0i})f(z_{0i})^\top)\otimes(\breve{z}_i\breve{z}_i^\top)\right\} + \Sigma_\xi\otimes\tilde{\Sigma},
\end{aligned}$$

where

$$\tilde{\Sigma} = \begin{pmatrix} 1 & E(z_{0i}^\top) \\ E(z_{0i}) & E(z_{0i}z_{0i}^\top) \end{pmatrix}.$$

Since $E(\{(f(z_{0i})f(z_{0i})^\top)\otimes(\breve{z}_i\breve{z}_i^\top)\})$ is positive semidefinite, the assumption that Σ_ξ and $\tilde{\Sigma}$ are positive definite is sufficient to ensure the positive definiteness of $E(\{(x_i x_i^\top)\otimes(\breve{z}_i\breve{z}_i^\top)\})$. As a result, this assumption is sufficient to ensure the first part of Condition (C3), which indicates that X cannot be fully explained by the latent vector Z_0.

S.6 Heterophilic Networks

As suggested by an anonymous referee, the proposed model can be extended to accommodate heterophilic networks. Suppose that the network is generated via the following model with indefinite inner products:

$$\mathrm{logit}\left(P_{ij}\right) := \Theta_{ij} = \alpha_i + \alpha_j + z_i^\top I_{k_1,k_2} z_j, \tag{S.15}$$

where α_i, α_j, z_i, and z_j are defined as those in equation (1) of the manuscript, and $I_{k_1,k_2} = \mathrm{diag}\left(I_{k_1}, -I_{k_2}\right)$ for some constants k_1 and k_2 with $k_1 + k_2 = k$. Equation (S.15) is similar to the model of Rubin-Delanchy et al. (2022), although they considered P_{ij} rather than $\mathrm{logit}(P_{ij})$ in (S.15). This model indicates that the probability of nodes i and j being connected increases along with the similarity of the first k_1 elements of z_i and z_j (i.e., the inner product of z_i and z_j), and decreases along with the similarity of their last k_2 elements.

We next adopt model (2) in the manuscript to construct the relationship between y_i and x_i as follows:

$$y_i = x_i^\top \beta_i + \epsilon_i = x_i^\top (Bz_i + \gamma) + \epsilon_i \tag{S.16}$$

for $i = 1, \cdots, n$, where z_i was defined in Model (S.15), B is a factor-loading matrix related to z_i, γ is a coefficient vector that does not change with the "locations", and $\mathcal{T} = (\beta_1, \cdots, \beta_n)^\top \in \mathbb{R}^{n\times p}$ is the regression coefficient matrix. To ensure identifiability of models (S.15) and (S.16), we assume that $(I_n - J_n)Z = 0$. Accordingly, Z is identifiable up to an indefinite orthogonal transformation. This means the probability matrix of the network remains the same if Z is replaced with ZM, where $MI_{k_1,k_2}M^\top = I_{k_1,k_2}$ and $M \in \mathbb{R}^{k\times k}$. Since the column space of $(\mathbf{1}_n, Z)$ and $(\mathbf{1}_n, ZM)$ are the same, the indefinite orthogonal transformation does not affect the regression coefficient matrix. As a result, the pairs (Z, B) and $(ZM, BM^{-1\top})$ yield identical probability and regression coefficient matrices. Consequently, the estimation procedure proposed in Section 2.3 is applicable for models (S.15) and (S.16). Please note that the i-th row of B and that of $BM^{-1\top}$ are different under the l_2 norm. Thus, the variable selection procedure proposed in the paper is not suitable for this model. Additional conditions are required to ensure the applicability of of variable selections. That needs further investigation.

As for the interpretation of coefficients, we note that the coefficient matrix $\mathcal{T}$ depends on three components, i.e., z_is, B and γ. The latent variables z_is can be viewed as the unobservable "location" of the nodes, and they affect the network connectivity in different ways. For example, the similarity of the first k_1 variables promotes connections between the nodes, while the others deter them. In addition, the factor-loading matrix B characterizes the relationship between regression coefficients β_i and latent "locations" z_is. If $B \neq 0$, the regression coefficients depend on latent "locations". Note that the difference between $\beta_{i_1 j}$ and $\beta_{i_2 j}$ is $|B_{j,\cdot}\left(z_{i_1} - z_{i_2}\right)|$, where $B_{j,\cdot}^\top$ is the j-th row of B. If $B_{j,\cdot} = 0$, there is no interaction between the network locations and the j-th covariate. Lastly, γ is a coefficient vector that does not change with latent "locations", and it is a classical homogeneous regression coefficient vector when $B = 0$.

S.7 Introduction of Two Competing Methods

The finite mixture model and network lasso method have been used for comparison in our simulation studies. This section presents a brief introduction of these two competing methods.

The finite mixture model (Leisch, 2004) assumes that the data consists of K groups with homogeneous regression coefficients within each group. The density function of y_i given x_i is expressed as:

$$f\left(y_i|x_i\right) = \sum_{k=1}^{K} \pi_k N(x_i^\top \beta_k, \sigma_k^2), \tag{S.17}$$

where π_k is the probability that y_i comes from the k-th group for $k = 1, \cdots, K$, and β_k and σ_k^2 are the regression coefficient vector and the error variance, respectively, for the k-th group. The parameters including π_k, β_k and σ_k^2 for $k = 1, \cdots, K$ in model (S.17) are estimated using the R package "flexmix". The number of groups K is selected via the BIC-type criterion.

The network lasso method (Hallac et al., 2015) assumes that

$$y_i = x_i^\top \beta_i + \epsilon_i, \text{ for } i = 1, \cdots, n, \tag{S.18}$$

where β_i is the coefficient vector, and ϵ_i is the random noise with mean 0 and variance σ^2. The coefficient matrix $\beta = (\beta_1, \cdots, \beta_n)^\top$ is estimated by

$$\hat{\beta} = \operatorname{argmin}_{\beta_i, i=1,\cdots,n} \sum_{i=1}^{n} (y_i - x_i^\top \beta_i)^2 + \mu_l \sum_{i<j} a_{ij} \|\beta_i - \beta_j\|_2,$$

where $\hat{\beta} = (\hat{\beta}_1, \cdots, \hat{\beta}_n)^\top$, μ_l is a tuning parameter and it can be chosen by the BIC criterion, $\|\cdot\|_2$ is the ℓ_2 norm function, and a_{ij} is the (i, j)-th element in the adjacency matrix A. The alternating direction method of multipliers (ADMM) algorithm is employed to calculate $\hat{\beta}$.

S.8 Additional Simulation Results

S.8.1 Additional results based on the settings in Section 4

This subsection presents additional simulation results based on those settings described in Section 4. Tables S.1-S.2 present the results of Num_k and CT when the random errors are normally distributed, where Num_k is the value of $\hat{k}$ selected by BIC in (6), and CT is the proportion of the true value of k being selected. These two tables indicate that all Num_ks are equal to 2 and all CTs are 1. As a result, the BIC in (6) can consistently select the dimension of the latent space. Tables S.3-S.8 report the simulation results when the random errors are generated from a mixture normal distribution, while Tables S.9–S.14 report the simulation results when the random errors are simulated from the standardized exponential distribution. The results in Tables S.3-S.14 are qualitatively similar to those in Tables 1-4 and Tables S.1-S.2. This suggests that our method is robust against different types of error distributions.

Table S.1: The selection of latent space dimension under the low-dimensional setting with $k = 2$ and standard normal random errors. The upper panel displays the means of the Num_ks along with the standard deviations in parentheses, while the lower panel displays CT.

		$B_{0,jl} \sim U(0.25, 0.5)$				$B_{0,jl} \sim U(0.5, 1)$			
		$z_{0,ij} \sim N(0, 4)$		$z_{0,ij} \sim U(-3, 3)$		$z_{0,ij} \sim N(0, 4)$		$z_{0,ij} \sim U(-3, 3)$	
p	n	NVCM	TNVC	NVCM	TNVC	NVCM	TNVC	NVCM	TNVC
					Num_k				
5	300	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0
		(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
	500	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0
		(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
	800	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0
		(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
10	300	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0
		(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
	500	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0
		(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
					CT				
5	300	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
	500	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
	800	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
10	300	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
	500	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
	800	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0

Table S.2: The selection of latent space dimension under the high-dimensional setting with $k = 2$ and standard normal random errors. The upper panel displays the means of the Num_ks along with the standard deviations in parentheses, while the lower panel displays CT.

	$p = 50$				$p = 100$			
	$z_{0,ij} \sim N(0, 4)$		$z_{0,ij} \sim U(-3, 3)$		$z_{0,ij} \sim N(0, 4)$		$z_{0,ij} \sim U(-3, 3)$	
n	SNVC	TSNVC	SNVC	TSNVC	SNVC	TSNVC	SNVC	TSNVC
				Num_k				
500	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
800	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
				CT				
500	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
800	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0

Table S.3: The means of the $\mathrm{ME}_{\mathcal{T}}$s along with the standard deviations in parentheses, which are obtained under the low-dimensional setting with $k = 2$ and mixture normal random errors.

p	n	NVCM	TNVC	NLM	FMM	CLR	NVCM	TNVC	NLM	FMM	CLR
		$B_{0,jl} \sim U(0.25, 0.5)$					$B_{0,jl} \sim U(0.5, 1)$				
		$z_{0,ij} \sim N(0, 4)$									
5	300	0.040	0.052	0.867	0.575	0.879	0.025	0.050	1.022	0.525	1.004
		(0.010)	(0.010)	(0.040)	(0.377)	(0.039)	(0.005)	(0.007)	(0.035)	(0.192)	(0.036)
	500	0.033	0.044	0.849	0.478	0.870	0.022	0.044	0.995	0.475	0.990
		(0.005)	(0.005)	(0.023)	(0.113)	(0.022)	(0.004)	(0.004)	(0.020)	(0.180)	(0.026)
	800	0.023	0.027	0.824	0.380	0.844	0.018	0.028	0.955	0.344	0.973
		(0.003)	(0.003)	(0.013)	(0.044)	(0.013)	(0.002)	(0.002)	(0.016)	(0.049)	(0.015)
10	300	0.042	0.065	0.908	1.011	0.924	0.028	0.063	1.098	1.045	1.085
		(0.009)	(0.011)	(0.060)	(0.620)	(0.065)	(0.005)	(0.011)	(0.063)	(0.773)	(0.068)
	500	0.032	0.050	0.859	0.658	0.885	0.020	0.053	1.032	0.729	1.031
		(0.006)	(0.007)	(0.040)	(0.376)	(0.040)	(0.004)	(0.005)	(0.038)	(0.356)	(0.047)
	800	0.022	0.030	0.845	0.435	0.867	0.015	0.032	0.978	0.459	1.002
		(0.003)	(0.003)	(0.022)	(0.084)	(0.022)	(0.002)	(0.003)	(0.025)	(0.090)	(0.025)
		$z_{0,ij} \sim U(-3, 3)$									
5	300	0.027	0.030	0.823	0.483	0.840	0.018	0.024	1.017	0.449	0.997
		(0.009)	(0.009)	(0.043)	(0.153)	(0.030)	(0.004)	(0.004)	(0.043)	(0.086)	(0.048)
	500	0.017	0.020	0.788	0.413	0.826	0.011	0.016	0.953	0.353	0.974
		(0.005)	(0.005)	(0.019)	(0.071)	(0.018)	(0.002)	(0.002)	(0.031)	(0.048)	(0.026)
	800	0.012	0.013	0.786	0.366	0.812	0.008	0.011	0.931	0.307	0.961
		(0.003)	(0.003)	(0.016)	(0.048)	(0.017)	(0.001)	(0.001)	(0.017)	(0.042)	(0.017)
10	300	0.032	0.036	0.869	0.712	0.885	0.019	0.028	1.098	0.900	1.080
		(0.007)	(0.008)	(0.057)	(0.301)	(0.056)	(0.004)	(0.005)	(0.075)	(0.810)	(0.084)
	500	0.018	0.023	0.811	0.457	0.854	0.011	0.018	0.980	0.485	1.018
		(0.004)	(0.003)	(0.035)	(0.095)	(0.035)	(0.002)	(0.002)	(0.038)	(0.182)	(0.044)
	800	0.012	0.015	0.796	0.376	0.825	0.008	0.013	0.946	0.375	0.981
		(0.002)	(0.002)	(0.022)	(0.064)	(0.022)	(0.001)	(0.001)	(0.023)	(0.052)	(0.023)

Table S.4: The means of the ME_γs and the ME_Bs along with the standard deviations in parentheses, which are obtained under the low-dimensional setting with $k = 2$ and mixture normal random errors.

p	n	NVCM	TNVC	NVCM	TNVC	NVCM	TNVC	NVCM	TNVC
		ME_γ		ME_B		ME_γ		ME_B	
		$B_{0,jl} \sim U(0.25, 0.5)$				$B_{0,jl} \sim U(0.5, 1)$			
		$z_{0,ij} \sim N(0, 4)$							
5	300	0.026	0.030	0.037	0.038	0.032	0.065	0.023	0.029
		(0.021)	(0.022)	(0.013)	(0.016)	(0.028)	(0.049)	(0.006)	(0.010)
	500	0.018	0.021	0.032	0.037	0.022	0.042	0.021	0.032
		(0.012)	(0.014)	(0.008)	(0.012)	(0.019)	(0.031)	(0.004)	(0.009)
	800	0.010	0.011	0.023	0.025	0.013	0.019	0.017	0.021
		(0.007)	(0.007)	(0.006)	(0.007)	(0.011)	(0.016)	(0.003)	(0.005)
10	300	0.036	0.053	0.040	0.052	0.051	0.110	0.038	0.044
		(0.022)	(0.030)	(0.010)	(0.016)	(0.027)	(0.054)	(0.013)	(0.017)
	500	0.019	0.026	0.034	0.046	0.031	0.083	0.019	0.039
		(0.012)	(0.014)	(0.009)	(0.014)	(0.020)	(0.045)	(0.004)	(0.011)
	800	0.010	0.013	0.022	0.027	0.016	0.036	0.015	0.026
		(0.006)	(0.007)	(0.004)	(0.006)	(0.009)	(0.022)	(0.002)	(0.005)
		$z_{0,ij} \sim U(-3, 3)$							
5	300	0.024	0.025	0.017	0.016	0.030	0.036	0.010	0.008
		(0.021)	(0.021)	(0.010)	(0.010)	(0.023)	(0.027)	(0.005)	(0.004)
	500	0.013	0.013	0.014	0.014	0.016	0.019	0.007	0.008
		(0.010)	(0.010)	(0.007)	(0.007)	(0.013)	(0.015)	(0.002)	(0.003)
	800	0.008	0.008	0.009	0.009	0.010	0.012	0.005	0.006
		(0.007)	(0.007)	(0.003)	(0.003)	(0.008)	(0.009)	(0.001)	(0.002)
10	300	0.031	0.035	0.023	0.022	0.041	0.059	0.020	0.011
		(0.017)	(0.018)	(0.008)	(0.008)	(0.023)	(0.030)	(0.007)	(0.004)
	500	0.015	0.017	0.016	0.018	0.019	0.030	0.010	0.010
		(0.008)	(0.009)	(0.005)	(0.005)	(0.011)	(0.016)	(0.004)	(0.003)
	800	0.009	0.010	0.010	0.011	0.012	0.016	0.005	0.007
		(0.004)	(0.005)	(0.003)	(0.003)	(0.006)	(0.009)	(0.001)	(0.002)

Table S.5: The means of the $\mathrm{ME}_\mathcal{T}$s along with the standard deviations in parentheses, which are obtained under the high-dimensional setting with $k = 2$ and mixture normal random errors.

p	n	SNVC	TSNVC	NVCM	TNVC	SNVC	TSNVC	NVCM	TNVC
		$z_{0,ij} \sim N(0, 4)$				$z_{0,ij} \sim U(-3, 3)$			
50	500	0.052	0.100	0.113	0.176	0.028	0.033	0.116	0.122
		(0.014)	(0.021)	(0.015)	(0.033)	(0.011)	(0.012)	(0.025)	(0.025)
	800	0.036	0.049	0.084	0.107	0.016	0.018	0.059	0.062
		(0.006)	(0.010)	(0.011)	(0.016)	(0.004)	(0.004)	(0.010)	(0.010)
100	500	0.060	0.111	0.389	0.507	0.029	0.036	0.431	0.389
		(0.018)	(0.029)	(0.058)	(0.083)	(0.015)	(0.016)	(0.074)	(0.066)
	800	0.037	0.051	0.183	0.225	0.014	0.017	0.126	0.133
		(0.007)	(0.012)	(0.021)	(0.028)	(0.003)	(0.004)	(0.016)	(0.016)

Table S.6: The means of the four model selection measures (TPR_B, TPR_γ, FDR_B and FDR_γ) along with the standard deviations in parentheses, which are obtained under the high-dimensional setting with mixture normal random errors.

p	n	TPR_B		TPR_γ		FDR_B		FDR_γ	
		SNVC	TSNVC	SNVC	TSNVC	SNVC	TSNVC	SNVC	TSNVC
					$z_{0,ij} \sim N(0,4)$				
50	500	1.000	1.000	0.936	0.769	0.102	0.094	0.022	0.029
		(0.000)	(0.000)	(0.094)	(0.145)	(0.099)	(0.111)	(0.046)	(0.056)
	800	1.000	1.000	0.997	0.985	0.043	0.075	0.006	0.010
		(0.000)	(0.000)	(0.017)	(0.044)	(0.060)	(0.081)	(0.024)	(0.029)
100	500	1.000	0.999	0.919	0.767	0.114	0.130	0.023	0.033
		(0.000)	(0.010)	(0.106)	(0.142)	(0.113)	(0.130)	(0.049)	(0.063)
	800	1.000	1.000	0.995	0.986	0.065	0.114	0.007	0.017
		(0.000)	(0.000)	(0.022)	(0.045)	(0.079)	(0.110)	(0.027)	(0.041)
					$z_{0,ij} \sim U(-3,3)$				
50	500	1.000	1.000	0.988	0.984	0.035	0.060	0.013	0.021
		(0.000)	(0.000)	(0.038)	(0.042)	(0.055)	(0.072)	(0.034)	(0.043)
	800	1.000	1.000	0.999	1.000	0.005	0.020	0.003	0.006
		(0.000)	(0.000)	(0.010)	(0.000)	(0.022)	(0.042)	(0.021)	(0.023)
100	500	1.000	1.000	0.980	0.970	0.045	0.075	0.022	0.035
		(0.000)	(0.000)	(0.055)	(0.064)	(0.071)	(0.086)	(0.046)	(0.057)
	800	1.000	1.000	1.000	0.999	0.012	0.025	0.003	0.008
		(0.000)	(0.000)	(0.000)	(0.010)	(0.031)	(0.049)	(0.021)	(0.026)

Table S.7: The selection of latent space dimension under the low-dimensional setting with $k = 2$ and mixture normal random errors. The upper panel displays the means of the Num_ks along with the standard deviations in parentheses, while the lower panel displays CT.

		$B_{0,jl} \sim U(0.25, 0.5)$				$B_{0,jl} \sim U(0.5, 1)$			
		$z_{0,ij} \sim N(0,4)$		$z_{0,ij} \sim U(-3,3)$		$z_{0,ij} \sim N(0,4)$		$z_{0,ij} \sim U(-3,3)$	
p	n	NVCM	TNVC	NVCM	TNVC	NVCM	TNVC	NVCM	TNVC
					Num_k				
5	300	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0
		(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
	500	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0
		(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
	800	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0
		(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
10	300	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0
		(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
	500	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0
		(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
					CT				
5	300	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
	500	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
	800	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
10	300	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
	500	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
	800	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0

Table S.8: The selection of latent space dimension under the high-dimensional setting with $k = 2$ and mixture normal random errors. The upper panel displays the means of the Num_ks along with the standard deviations in parentheses, while the lower panel displays CT.

	$p = 50$				$p = 100$			
	$z_{0,ij} \sim N(0,4)$		$z_{0,ij} \sim U(-3,3)$		$z_{0,ij} \sim N(0,4)$		$z_{0,ij} \sim U(-3,3)$	
n	SNVC	TSNVC	SNVC	TSNVC	SNVC	TSNVC	SNVC	TSNVC
				Num_k				
500	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
800	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
				CT				
500	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
800	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0

Table S.9: The means of the $\mathrm{ME}_{\mathcal{T}}$s along with the standard deviations in parentheses, which are obtained under the low-dimensional setting with $k = 2$ and standardized exponential random errors.

p	n	NVCM	TNVC	NLM	FMM	CLR	NVCM	TNVC	NLM	FMM	CLR
		$B_{0,jl} \sim U(0.25, 0.5)$					$B_{0,jl} \sim U(0.5, 1)$				
		$z_{0,ij} \sim N(0, 4)$									
5	300	0.040	0.052	0.864	0.538	0.877	0.025	0.051	1.022	0.529	1.005
		(0.008)	(0.009)	(0.042)	(0.150)	(0.037)	(0.004)	(0.006)	(0.037)	(0.183)	(0.038)
	500	0.032	0.043	0.847	0.493	0.869	0.022	0.044	0.994	0.482	0.990
		(0.004)	(0.005)	(0.021)	(0.139)	(0.021)	(0.002)	(0.004)	(0.019)	(0.153)	(0.025)
	800	0.023	0.027	0.824	0.388	0.843	0.018	0.028	0.955	0.345	0.972
		(0.002)	(0.003)	(0.012)	(0.043)	(0.013)	(0.001)	(0.002)	(0.016)	(0.041)	(0.015)
10	300	0.045	0.067	0.912	0.975	0.926	0.028	0.064	1.098	1.334	1.083
		(0.009)	(0.012)	(0.062)	(0.556)	(0.067)	(0.006)	(0.010)	(0.063)	(1.023)	(0.067)
	500	0.033	0.051	0.859	0.647	0.885	0.020	0.052	1.033	0.683	1.030
		(0.005)	(0.007)	(0.039)	(0.270)	(0.039)	(0.003)	(0.005)	(0.037)	(0.248)	(0.046)
	800	0.023	0.030	0.844	0.447	0.866	0.015	0.032	0.978	0.454	1.002
		(0.003)	(0.003)	(0.022)	(0.093)	(0.022)	(0.001)	(0.003)	(0.025)	(0.083)	(0.025)
		$z_{0,ij} \sim U(-3, 3)$									
5	300	0.028	0.030	0.827	0.475	0.845	0.017	0.023	1.016	0.455	0.996
		(0.007)	(0.007)	(0.044)	(0.096)	(0.036)	(0.003)	(0.004)	(0.044)	(0.142)	(0.047)
	500	0.018	0.020	0.788	0.412	0.826	0.011	0.016	0.951	0.347	0.974
		(0.004)	(0.004)	(0.018)	(0.064)	(0.017)	(0.001)	(0.002)	(0.032)	(0.042)	(0.026)
	800	0.012	0.013	0.786	0.365	0.812	0.008	0.011	0.931	0.306	0.961
		(0.002)	(0.002)	(0.017)	(0.030)	(0.017)	(0.001)	(0.001)	(0.019)	(0.034)	(0.018)
10	300	0.032	0.036	0.867	0.748	0.883	0.020	0.029	1.099	0.882	1.082
		(0.007)	(0.008)	(0.056)	(0.571)	(0.053)	(0.003)	(0.005)	(0.078)	(1.030)	(0.088)
	500	0.019	0.022	0.810	0.477	0.854	0.011	0.019	0.980	0.506	1.017
		(0.004)	(0.004)	(0.032)	(0.143)	(0.032)	(0.002)	(0.002)	(0.036)	(0.208)	(0.043)
	800	0.013	0.015	0.796	0.384	0.825	0.008	0.013	0.946	0.379	0.981
		(0.002)	(0.002)	(0.023)	(0.088)	(0.022)	(0.001)	(0.001)	(0.023)	(0.051)	(0.023)

Table S.10: The means of the ME_γs and the ME_Bs along with the standard deviations in parentheses, which are obtained under the low-dimensional setting with $k = 2$ and standardized exponential random errors.

p	n	NVCM	TNVC	NVCM	TNVC	NVCM	TNVC	NVCM	TNVC
		ME_γ		ME_B		ME_γ		ME_B	
		$B_{0,jl} \sim U(0.25, 0.5)$				$B_{0,jl} \sim U(0.5, 1)$			
		$z_{0,ij} \sim N(0, 4)$							
5	300	0.029	0.033	0.037	0.038	0.035	0.067	0.023	0.029
		(0.021)	(0.023)	(0.013)	(0.014)	(0.024)	(0.047)	(0.006)	(0.011)
	500	0.016	0.020	0.032	0.037	0.020	0.041	0.021	0.032
		(0.012)	(0.014)	(0.007)	(0.012)	(0.013)	(0.027)	(0.004)	(0.009)
	800	0.008	0.009	0.023	0.025	0.012	0.017	0.016	0.021
		(0.006)	(0.007)	(0.006)	(0.007)	(0.008)	(0.013)	(0.003)	(0.005)
10	300	0.038	0.053	0.045	0.056	0.055	0.120	0.036	0.044
		(0.021)	(0.030)	(0.013)	(0.018)	(0.033)	(0.055)	(0.013)	(0.016)
	500	0.019	0.026	0.034	0.047	0.028	0.075	0.019	0.039
		(0.009)	(0.013)	(0.007)	(0.015)	(0.017)	(0.042)	(0.003)	(0.011)
	800	0.012	0.015	0.023	0.027	0.017	0.034	0.015	0.025
		(0.007)	(0.009)	(0.004)	(0.006)	(0.008)	(0.020)	(0.002)	(0.006)
		$z_{0,ij} \sim U(-3, 3)$							
5	300	0.024	0.026	0.018	0.017	0.030	0.036	0.010	0.007
		(0.016)	(0.017)	(0.008)	(0.008)	(0.022)	(0.028)	(0.005)	(0.004)
	500	0.013	0.013	0.014	0.014	0.015	0.017	0.007	0.009
		(0.010)	(0.011)	(0.005)	(0.005)	(0.009)	(0.011)	(0.002)	(0.003)
	800	0.009	0.009	0.009	0.009	0.010	0.011	0.005	0.006
		(0.006)	(0.006)	(0.003)	(0.003)	(0.007)	(0.009)	(0.001)	(0.002)
10	300	0.029	0.034	0.024	0.023	0.045	0.068	0.020	0.012
		(0.016)	(0.019)	(0.008)	(0.009)	(0.023)	(0.034)	(0.008)	(0.005)
	500	0.015	0.016	0.016	0.018	0.020	0.030	0.009	0.010
		(0.010)	(0.011)	(0.005)	(0.005)	(0.010)	(0.015)	(0.003)	(0.003)
	800	0.010	0.011	0.010	0.011	0.011	0.015	0.005	0.007
		(0.007)	(0.007)	(0.002)	(0.003)	(0.006)	(0.008)	(0.001)	(0.002)

Table S.11: The means of the $\text{ME}_\mathcal{T}$s along with the standard deviations in parentheses, which are obtained under the high-dimensional setting with $k = 2$ and standardized exponential random errors.

p	n	SNVC	TSNVC	NVCM	TNVC	SNVC	TSNVC	NVCM	TNVC
		$z_{0,ij} \sim N(0, 4)$				$z_{0,ij} \sim U(-3, 3)$			
50	500	0.044	0.076	0.120	0.162	0.025	0.033	0.104	0.113
		(0.015)	(0.022)	(0.017)	(0.022)	(0.007)	(0.011)	(0.020)	(0.019)
	800	0.029	0.039	0.074	0.093	0.014	0.016	0.052	0.056
		(0.004)	(0.008)	(0.010)	(0.013)	(0.002)	(0.002)	(0.007)	(0.007)
100	500	0.049	0.091	0.398	0.479	0.031	0.040	0.444	0.406
		(0.016)	(0.026)	(0.066)	(0.067)	(0.016)	(0.017)	(0.077)	(0.070)
	800	0.030	0.041	0.168	0.201	0.015	0.017	0.137	0.141
		(0.006)	(0.010)	(0.022)	(0.028)	(0.003)	(0.004)	(0.016)	(0.016)

Table S.12: The means of the four model selection measures (TPR_B, TPR_γ, FDR_B and FDR_γ) along with the standard deviations in parentheses, which are obtained under the high-dimensional setting with standardized exponential random errors.

p	n	TPR_B		TPR_γ		FDR_B		FDR_γ	
		SNVC	TSNVC	SNVC	TSNVC	SNVC	TSNVC	SNVC	TSNVC
					$z_{0,ij} \sim N(0,4)$				
50	500	1.000	1.000	0.964	0.882	0.073	0.090	0.020	0.021
		(0.000)	(0.000)	(0.075)	(0.115)	(0.085)	(0.094)	(0.044)	(0.050)
	800	1.000	1.000	1.000	0.994	0.029	0.054	0.009	0.013
		(0.000)	(0.000)	(0.000)	(0.024)	(0.059)	(0.075)	(0.029)	(0.034)
100	500	1.000	1.000	0.924	0.788	0.094	0.104	0.026	0.028
		(0.000)	(0.000)	(0.092)	(0.146)	(0.101)	(0.122)	(0.051)	(0.067)
	800	1.000	1.000	0.996	0.989	0.040	0.099	0.008	0.024
		(0.000)	(0.000)	(0.020)	(0.037)	(0.067)	(0.100)	(0.026)	(0.047)
					$z_{0,ij} \sim U(-3,3)$				
50	500	1.000	1.000	0.994	0.983	0.038	0.053	0.019	0.019
		(0.000)	(0.000)	(0.024)	(0.043)	(0.065)	(0.073)	(0.043)	(0.043)
	800	1.000	1.000	1.000	1.000	0.005	0.016	0.004	0.005
		(0.000)	(0.000)	(0.000)	(0.000)	(0.024)	(0.040)	(0.018)	(0.020)
100	500	1.000	1.000	0.982	0.965	0.053	0.086	0.029	0.035
		(0.000)	(0.000)	(0.056)	(0.066)	(0.070)	(0.077)	(0.059)	(0.063)
	800	1.000	1.000	1.000	0.999	0.005	0.012	0.006	0.010
		(0.000)	(0.000)	(0.000)	(0.010)	(0.022)	(0.033)	(0.023)	(0.031)

Table S.13: The selection of latent space dimension under the low-dimensional setting with $k = 2$ and standardized exponential random errors. The upper panel displays the means of the Num_ks along with the standard deviations in parentheses, while the lower panel displays CT.

		$B_{0,jl} \sim U(0.25, 0.5)$				$B_{0,jl} \sim U(-3, 3)$			
		$z_{0,ij} \sim N(0, 4)$		$z_{0,ij} \sim U(-3, 3)$		$z_{0,ij} \sim N(0, 4)$		$z_{0,ij} \sim U(-3, 3)$	
p	n	NVCM	TNVC	NVCM	TNVC	NVCM	TNVC	NVCM	TNVC
					Num_k				
5	300	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0
		(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
	500	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0
		(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
	800	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0
		(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
10	300	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0
		(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
	500	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0
		(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
					CT				
5	300	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
	500	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
	800	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
10	300	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
	500	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
	800	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0

Table S.14: The selection of latent space dimension under the high-dimensional setting with $k = 2$ and standardized exponential random errors. The upper panel displays the means of the Num_ks along with the standard deviations in parentheses, while the lower panel displays CT.

	$p = 50$				$p = 100$			
	$z_{0,ij} \sim N(0, 4)$		$z_{0,ij} \sim U(-3, 3)$		$z_{0,ij} \sim N(0, 4)$		$z_{0,ij} \sim U(-3, 3)$	
n	SNVC	TSNVC	SNVC	TSNVC	SNVC	TSNVC	SNVC	TSNVC
				Num_k				
500	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
800	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
				CT				
500	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
800	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0

S.8.2 Additional simulation studies for collinearity

In this subsection, we explore the effect of collinearity on the performance of our proposed method. To assess the effect of collinearity, we compared NVCM with the other four methods, TNVC, NLM, FMM, and CLR, used in simulation studies. The parameters and random errors are the same as those generated from the low dimensional setting in Section 4. As for generating the covariates, we consider the model $X = Z_0 C_x + \mathcal{E}$, where the elements of $C_x \in \mathbb{R}^{k\times p}$ are independently generated from a standard normal distribution, and the elements of $\mathcal{E} \in \mathbb{R}^{n\times p}$ are independently generated from $N(0, \varrho)$ with $\varrho = 0.25$, 0.5, and 1. As a result, the degree of collinearity increases as ϱ decreases.

Tables S.15–S.20 present simulation results and show the following findings. As ϱ decreases, the NVCM's performance deteriorates as expected. For example, Table S.15 indicates that $\mathrm{ME}_{\mathcal{T}}$s of NVCM are 0.037, 0.064, and 0.177 when $\varrho = 1$, 0.5, and 0.25, respectively, under the setting $p = 10$, $n = 300$, $z_{0,ij} \sim N(0, 4)$, and $B_{0,ij} \sim U(0.25, 0.5)$. Furthermore, the robustness of NVCM against collinearity is either comparable or superior to TNVC and outperforms NLM, FMM, and CLR.

Table S.15: The means of the $\mathrm{ME}_{\mathcal{T}}$s along with the standard deviations in parentheses, which are obtained under the setting that $X = Z_0 C_x + \mathcal{E}$ with $\varrho = 1$ and standard normal random errors.

p	n	NVCM	TNVC	NLM	FMM	CLR	NVCM	TNVC	NLM	FMM	CLR
				$B_{0,jl} \sim U(0.25, 0.5)$					$B_{0,jl} \sim U(0.5, 1)$		
						$z_{0,ij} \sim N(0, 4)$					
5	300	0.035	0.040	0.771	0.766	0.794	0.028	0.043	0.955	0.847	0.949
		(0.004)	(0.004)	(0.012)	(0.079)	(0.009)	(0.003)	(0.003)	(0.009)	(0.108)	(0.009)
	500	0.031	0.039	0.834	0.727	0.841	0.024	0.041	0.970	0.805	0.966
		(0.002)	(0.002)	(0.015)	(0.073)	(0.012)	(0.002)	(0.002)	(0.011)	(0.108)	(0.011)
	800	0.020	0.025	0.835	0.559	0.844	0.016	0.027	0.969	0.573	0.963
		(0.001)	(0.001)	(0.016)	(0.063)	(0.009)	(0.001)	(0.001)	(0.008)	(0.077)	(0.008)
10	300	0.037	0.077	0.939	1.032	0.974	0.032	0.082	1.068	1.287	1.134
		(0.007)	(0.015)	(0.081)	(0.319)	(0.086)	(0.005)	(0.013)	(0.086)	(0.799)	(0.092)
	500	0.025	0.054	0.905	0.866	0.908	0.015	0.058	1.033	0.954	1.046
		(0.003)	(0.007)	(0.046)	(0.309)	(0.047)	(0.002)	(0.007)	(0.052)	(0.227)	(0.055)
	800	0.017	0.032	0.885	0.557	0.882	0.011	0.034	1.020	0.609	1.019
		(0.002)	(0.003)	(0.033)	(0.182)	(0.034)	(0.001)	(0.003)	(0.041)	(0.062)	(0.042)
						$z_{0,ij} \sim U(-3, 3)$					
5	300	0.020	0.021	0.751	0.726	0.774	0.016	0.019	0.945	0.768	0.940
		(0.004)	(0.004)	(0.013)	(0.091)	(0.012)	(0.002)	(0.002)	(0.009)	(0.122)	(0.010)
	500	0.014	0.015	0.743	0.658	0.763	0.011	0.014	0.922	0.719	0.934
		(0.002)	(0.002)	(0.007)	(0.084)	(0.007)	(0.001)	(0.001)	(0.016)	(0.093)	(0.008)
	800	0.009	0.010	0.775	0.666	0.794	0.008	0.010	0.935	0.675	0.944
		(0.001)	(0.001)	(0.006)	(0.045)	(0.006)	(0.001)	(0.001)	(0.013)	(0.058)	(0.006)
10	300	0.026	0.028	0.809	0.617	0.824	0.018	0.027	1.000	0.735	1.020
		(0.006)	(0.005)	(0.041)	(0.141)	(0.043)	(0.003)	(0.004)	(0.046)	(0.139)	(0.047)
	500	0.014	0.018	0.800	0.623	0.826	0.010	0.017	0.982	0.703	0.990
		(0.002)	(0.002)	(0.033)	(0.057)	(0.029)	(0.001)	(0.002)	(0.031)	(0.074)	(0.029)
	800	0.010	0.012	0.768	0.639	0.806	0.007	0.012	0.963	0.741	0.975
		(0.001)	(0.001)	(0.015)	(0.061)	(0.015)	(0.001)	(0.001)	(0.025)	(0.075)	(0.022)

Table S.16: The means of the $\mathrm{ME}_{\mathcal{T}}$s along with the standard deviations in parentheses, which are obtained under the setting that $X = Z_0 C_x + \mathcal{E}$ with $\varrho = 0.5$ and standard normal random errors.

p	n	NVCM	TNVC	NLM	FMM	CLR	NVCM	TNVC	NLM	FMM	CLR
		$B_{0,jl} \sim U(0.25, 0.5)$					$B_{0,jl} \sim U(0.5, 1)$				
		$z_{0,ij} \sim N(0, 4)$									
5	300	0.043	0.047	0.783	0.844	0.797	0.035	0.046	0.951	0.912	0.952
		(0.007)	(0.007)	(0.012)	(0.079)	(0.012)	(0.004)	(0.004)	(0.011)	(0.098)	(0.011)
	500	0.036	0.043	0.837	0.772	0.849	0.028	0.043	0.975	0.807	0.975
		(0.004)	(0.004)	(0.021)	(0.074)	(0.018)	(0.003)	(0.003)	(0.018)	(0.103)	(0.017)
	800	0.022	0.027	0.838	0.567	0.850	0.017	0.028	0.971	0.573	0.969
		(0.002)	(0.002)	(0.017)	(0.056)	(0.014)	(0.001)	(0.002)	(0.012)	(0.063)	(0.014)
10	300	0.064	0.114	1.067	1.455	1.114	0.048	0.117	1.223	1.669	1.290
		(0.019)	(0.028)	(0.151)	(0.835)	(0.157)	(0.013)	(0.024)	(0.162)	(1.025)	(0.168)
	500	0.034	0.070	0.976	0.963	0.981	0.019	0.073	1.110	1.061	1.127
		(0.006)	(0.012)	(0.084)	(0.212)	(0.087)	(0.003)	(0.012)	(0.093)	(0.156)	(0.099)
	800	0.022	0.040	0.936	0.593	0.938	0.014	0.041	1.076	0.678	1.080
		(0.004)	(0.006)	(0.058)	(0.073)	(0.063)	(0.002)	(0.006)	(0.070)	(0.070)	(0.077)
		$z_{0,ij} \sim U(-3, 3)$									
5	300	0.028	0.028	0.764	0.801	0.777	0.019	0.022	0.936	0.842	0.942
		(0.007)	(0.007)	(0.014)	(0.084)	(0.014)	(0.002)	(0.003)	(0.012)	(0.091)	(0.011)
	500	0.019	0.019	0.754	0.758	0.765	0.014	0.016	0.923	0.778	0.936
		(0.004)	(0.004)	(0.009)	(0.058)	(0.009)	(0.001)	(0.001)	(0.011)	(0.083)	(0.010)
	800	0.012	0.012	0.784	0.692	0.796	0.009	0.011	0.934	0.713	0.946
		(0.002)	(0.002)	(0.008)	(0.040)	(0.008)	(0.001)	(0.001)	(0.009)	(0.047)	(0.008)
10	300	0.045	0.043	0.858	0.661	0.891	0.028	0.036	1.066	0.736	1.099
		(0.014)	(0.009)	(0.075)	(0.141)	(0.079)	(0.007)	(0.007)	(0.082)	(0.103)	(0.085)
	500	0.021	0.025	0.823	0.637	0.863	0.013	0.021	0.991	0.726	1.033
		(0.005)	(0.004)	(0.049)	(0.048)	(0.049)	(0.002)	(0.003)	(0.047)	(0.060)	(0.049)
	800	0.014	0.016	0.798	0.715	0.832	0.008	0.014	0.971	0.821	1.008
		(0.002)	(0.002)	(0.027)	(0.051)	(0.027)	(0.001)	(0.001)	(0.037)	(0.056)	(0.039)

Table S.17: The means of the ME$_{\mathcal{T}}$s along with the standard deviations in parentheses, which are obtained under the setting that $X = Z_0 C_x + \mathcal{E}$ with $\varrho = 0.25$ and standard normal random errors.

p	n	NVCM	TNVC	NLM	FMM	CLR	NVCM	TNVC	NLM	FMM	CLR
		$B_{0,jl} \sim U(0.25, 0.5)$					$B_{0,jl} \sim U(0.5, 1)$				
		$z_{0,ij} \sim N(0, 4)$									
5	300	0.059	0.059	0.795	0.873	0.804	0.044	0.052	0.953	0.948	0.958
		(0.015)	(0.012)	(0.017)	(0.094)	(0.017)	(0.007)	(0.006)	(0.016)	(0.076)	(0.016)
	500	0.045	0.050	0.852	0.804	0.866	0.033	0.047	0.985	0.811	0.992
		(0.008)	(0.008)	(0.031)	(0.074)	(0.031)	(0.004)	(0.004)	(0.030)	(0.107)	(0.031)
	800	0.027	0.032	0.851	0.569	0.863	0.019	0.030	0.978	0.582	0.981
		(0.004)	(0.004)	(0.026)	(0.052)	(0.025)	(0.002)	(0.003)	(0.022)	(0.063)	(0.024)
10	300	0.177	0.189	1.339	2.273	1.394	0.103	0.188	1.530	2.427	1.600
		(0.112)	(0.052)	(0.292)	(1.762)	(0.297)	(0.040)	(0.047)	(0.314)	(1.513)	(0.318)
	500	0.054	0.104	1.116	1.317	1.128	0.031	0.103	1.265	1.446	1.290
		(0.014)	(0.023)	(0.160)	(0.396)	(0.167)	(0.013)	(0.023)	(0.178)	(0.583)	(0.186)
	800	0.034	0.057	1.035	0.718	1.051	0.022	0.056	1.185	0.787	1.203
		(0.008)	(0.011)	(0.115)	(0.123)	(0.122)	(0.005)	(0.012)	(0.135)	(0.119)	(0.146)
		$z_{0,ij} \sim U(-3, 3)$									
5	300	0.042	0.041	0.776	0.884	0.783	0.025	0.026	0.938	0.940	0.945
		(0.014)	(0.014)	(0.018)	(0.120)	(0.018)	(0.004)	(0.005)	(0.014)	(0.090)	(0.014)
	500	0.028	0.027	0.763	0.850	0.770	0.017	0.018	0.930	0.855	0.939
		(0.008)	(0.008)	(0.013)	(0.071)	(0.013)	(0.002)	(0.002)	(0.014)	(0.082)	(0.014)
	800	0.016	0.016	0.792	0.667	0.801	0.011	0.013	0.940	0.715	0.951
		(0.004)	(0.004)	(0.012)	(0.042)	(0.012)	(0.001)	(0.002)	(0.012)	(0.046)	(0.012)
10	300	0.177	0.071	0.984	0.737	1.024	0.058	0.055	1.211	0.829	1.257
		(0.196)	(0.017)	(0.147)	(0.196)	(0.150)	(0.021)	(0.012)	(0.157)	(0.177)	(0.161)
	500	0.036	0.039	0.897	0.673	0.938	0.022	0.029	1.068	0.781	1.118
		(0.010)	(0.008)	(0.088)	(0.062)	(0.088)	(0.005)	(0.005)	(0.088)	(0.069)	(0.089)
	800	0.021	0.024	0.853	0.784	0.885	0.012	0.018	1.034	0.923	1.073
		(0.005)	(0.004)	(0.051)	(0.069)	(0.050)	(0.002)	(0.003)	(0.070)	(0.094)	(0.072)

Table S.18: The means of the ME_γs and the ME_Bs along with the standard deviations in parentheses, which are obtained under the setting that $X = Z_0C_x + \mathcal{E}$ with $\varrho = 1$ and standard normal random errors.

p	n	NVCM	TNVC	NVCM	TNVC	NVCM	TNVC	NVCM	TNVC
		ME_γ		ME_B		ME_γ		ME_B	
		$B_{0,jl} \sim U(0.25, 0.5)$				$B_{0,jl} \sim U(0.5, 1)$			
		$z_{0,ij} \sim N(0, 4)$							
5	300	0.011	0.012	0.029	0.027	0.014	0.017	0.022	0.024
		(0.008)	(0.008)	(0.011)	(0.011)	(0.013)	(0.013)	(0.007)	(0.011)
	500	0.007	0.009	0.027	0.032	0.009	0.015	0.017	0.029
		(0.005)	(0.006)	(0.008)	(0.011)	(0.007)	(0.012)	(0.005)	(0.009)
	800	0.004	0.005	0.020	0.023	0.005	0.010	0.014	0.022
		(0.003)	(0.004)	(0.006)	(0.007)	(0.004)	(0.009)	(0.002)	(0.005)
10	300	0.022	0.049	0.041	0.071	0.039	0.142	0.053	0.064
		(0.011)	(0.024)	(0.015)	(0.027)	(0.019)	(0.072)	(0.015)	(0.022)
	500	0.012	0.026	0.027	0.054	0.015	0.072	0.017	0.048
		(0.006)	(0.013)	(0.007)	(0.018)	(0.007)	(0.034)	(0.005)	(0.018)
	800	0.007	0.013	0.019	0.032	0.008	0.031	0.013	0.029
		(0.003)	(0.006)	(0.005)	(0.010)	(0.004)	(0.016)	(0.002)	(0.008)
		$z_{0,ij} \sim U(-3, 3)$							
5	300	0.009	0.009	0.010	0.008	0.010	0.011	0.005	0.004
		(0.006)	(0.006)	(0.006)	(0.005)	(0.008)	(0.008)	(0.003)	(0.002)
	500	0.006	0.006	0.011	0.010	0.006	0.006	0.006	0.006
		(0.004)	(0.005)	(0.005)	(0.005)	(0.004)	(0.004)	(0.002)	(0.002)
	800	0.003	0.003	0.007	0.007	0.004	0.005	0.005	0.005
		(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)	(0.001)	(0.002)
10	300	0.018	0.019	0.019	0.016	0.024	0.031	0.019	0.011
		(0.010)	(0.010)	(0.007)	(0.006)	(0.010)	(0.016)	(0.008)	(0.004)
	500	0.009	0.010	0.011	0.012	0.012	0.017	0.007	0.009
		(0.005)	(0.005)	(0.004)	(0.004)	(0.005)	(0.009)	(0.003)	(0.003)
	800	0.005	0.006	0.008	0.008	0.007	0.009	0.004	0.006
		(0.003)	(0.003)	(0.002)	(0.002)	(0.003)	(0.004)	(0.001)	(0.002)

Table S.19: The means of the ME_γs and the ME_Bs along with the standard deviations in parentheses, which are obtained under the setting that $X = Z_0C_x + \mathcal{E}$ with $\varrho = 0.5$ and standard normal random errors.

p	n	NVCM	TNVC	NVCM	TNVC	NVCM	TNVC	NVCM	TNVC
		ME_γ		ME_B		ME_γ		ME_B	
		$B_{0,jl} \sim U(0.25, 0.5)$				$B_{0,jl} \sim U(0.5, 1)$			
		$z_{0,ij} \sim N(0, 4)$							
5	300	0.020	0.021	0.037	0.033	0.024	0.027	0.026	0.027
		(0.015)	(0.015)	(0.015)	(0.014)	(0.022)	(0.021)	(0.010)	(0.013)
	500	0.014	0.015	0.032	0.037	0.016	0.025	0.019	0.031
		(0.010)	(0.011)	(0.011)	(0.014)	(0.013)	(0.020)	(0.006)	(0.010)
	800	0.008	0.009	0.022	0.025	0.010	0.016	0.015	0.023
		(0.006)	(0.006)	(0.008)	(0.009)	(0.009)	(0.015)	(0.003)	(0.007)
10	300	0.044	0.091	0.079	0.115	0.071	0.257	0.076	0.100
		(0.023)	(0.045)	(0.037)	(0.043)	(0.035)	(0.126)	(0.029)	(0.035)
	500	0.023	0.048	0.038	0.074	0.029	0.129	0.022	0.065
		(0.012)	(0.025)	(0.011)	(0.026)	(0.014)	(0.060)	(0.007)	(0.026)
	800	0.014	0.024	0.026	0.042	0.016	0.058	0.017	0.036
		(0.007)	(0.012)	(0.008)	(0.014)	(0.008)	(0.030)	(0.003)	(0.011)
		$z_{0,ij} \sim U(-3, 3)$							
5	300	0.017	0.017	0.017	0.014	0.018	0.019	0.007	0.005
		(0.013)	(0.012)	(0.009)	(0.008)	(0.014)	(0.015)	(0.004)	(0.003)
	500	0.011	0.011	0.015	0.014	0.011	0.011	0.008	0.007
		(0.008)	(0.008)	(0.008)	(0.007)	(0.007)	(0.007)	(0.003)	(0.003)
	800	0.006	0.006	0.009	0.009	0.008	0.009	0.006	0.006
		(0.005)	(0.005)	(0.005)	(0.005)	(0.007)	(0.007)	(0.002)	(0.002)
10	300	0.036	0.036	0.038	0.029	0.048	0.055	0.027	0.018
		(0.020)	(0.019)	(0.018)	(0.010)	(0.020)	(0.028)	(0.013)	(0.007)
	500	0.018	0.019	0.018	0.019	0.023	0.030	0.010	0.012
		(0.009)	(0.009)	(0.007)	(0.006)	(0.011)	(0.016)	(0.004)	(0.004)
	800	0.010	0.011	0.011	0.012	0.014	0.017	0.006	0.008
		(0.005)	(0.005)	(0.003)	(0.003)	(0.007)	(0.007)	(0.002)	(0.003)

Table S.20: The means of the ME_γs and the ME_Bs along with the standard deviations in parentheses, which are obtained under the setting that $X = Z_0 C_x + \mathcal{E}$ with $\varrho = 0.25$ and standard normal random errors.

p	n	NVCM	TNVC	NVCM	TNVC	NVCM	TNVC	NVCM	TNVC
		ME_γ		ME_B		ME_γ		ME_B	
		$B_{0,jl} \sim U(0.25, 0.5)$				$B_{0,jl} \sim U(0.5, 1)$			
		$z_{0,ij} \sim N(0, 4)$							
5	300	0.039	0.039	0.054	0.046	0.045	0.047	0.033	0.033
		(0.028)	(0.028)	(0.025)	(0.021)	(0.041)	(0.039)	(0.017)	(0.016)
	500	0.026	0.028	0.041	0.045	0.030	0.043	0.023	0.035
		(0.018)	(0.020)	(0.017)	(0.019)	(0.024)	(0.037)	(0.008)	(0.013)
	800	0.015	0.017	0.027	0.030	0.019	0.029	0.017	0.026
		(0.011)	(0.012)	(0.011)	(0.013)	(0.017)	(0.027)	(0.004)	(0.008)
10	300	0.101	0.173	0.259	0.202	0.154	0.486	0.162	0.173
		(0.057)	(0.086)	(0.206)	(0.076)	(0.074)	(0.234)	(0.069)	(0.062)
	500	0.047	0.092	0.063	0.114	0.059	0.244	0.036	0.098
		(0.024)	(0.048)	(0.022)	(0.042)	(0.028)	(0.113)	(0.023)	(0.042)
	800	0.028	0.046	0.042	0.060	0.031	0.114	0.029	0.051
		(0.013)	(0.023)	(0.015)	(0.021)	(0.016)	(0.059)	(0.007)	(0.018)
		$z_{0,ij} \sim U(-3, 3)$							
5	300	0.032	0.032	0.030	0.026	0.035	0.035	0.010	0.008
		(0.026)	(0.025)	(0.018)	(0.016)	(0.027)	(0.027)	(0.006)	(0.005)
	500	0.020	0.020	0.024	0.021	0.020	0.020	0.010	0.009
		(0.016)	(0.016)	(0.013)	(0.011)	(0.012)	(0.013)	(0.004)	(0.004)
	800	0.012	0.012	0.014	0.013	0.016	0.016	0.007	0.007
		(0.010)	(0.010)	(0.007)	(0.007)	(0.013)	(0.013)	(0.003)	(0.003)
10	300	0.090	0.069	0.200	0.055	0.109	0.102	0.060	0.034
		(0.066)	(0.036)	(0.253)	(0.019)	(0.051)	(0.051)	(0.029)	(0.013)
	500	0.035	0.037	0.033	0.033	0.047	0.057	0.018	0.019
		(0.018)	(0.018)	(0.014)	(0.011)	(0.022)	(0.030)	(0.007)	(0.007)
	800	0.020	0.021	0.020	0.020	0.027	0.032	0.009	0.011
		(0.010)	(0.010)	(0.006)	(0.006)	(0.013)	(0.014)	(0.003)	(0.004)

S.9 Additional Data Description and Results of Data Analysis

S.9.1 Ten financial ratios

The detailed description of the 10 financial ratios is provided in Table S.21.

Table S.21: Definitions of the 10 financial ratios.

Covariates	Description
X_1	book to market ratio
X_2	earnings per share/stock price
X_3	firm equity/interest-bearing debt
X_4	fixed assets ratio
X_5	total operating revenue
X_6	net debt
X_7	interest-free liabilities
X_8	operating cash flow/net debt
X_9	price-to-book ratio
X_{10}	working capital

S.9.2 Stock groups

To better understand the varying relationships between stock returns and ten financial ratios, the stocks are grouped based on their correspondingly estimated coefficients $\hat{\beta}_i$ classified by K-means. The number of groups is determined by the elbow method, which yields three groups. In sum, the number of stocks in each group is 283, 134, and 383, respectively. For each individual group, Figure S.1 depicts the boxplots of the coefficients for the intercept and ten financial ratios. Moreover, Figure S.1 supports the findings of Figure 2. That is identifying the distributional difference in regression coefficients among the three groups, especially for the intercept and X_8.

S.9.3 Estimation stability

We employ the approach of Li et al. (2020) to evaluate the stability of the estimates of B. We generate adjacency matrices $\tilde{A}$ with 90% edges being randomly selected from those of the true adjacency matrix A. Given $\tilde{A}$, we fit models (1) and (2) to estimate B. This procedure was repeated 50 times, and Figure S.2 depicts the estimates of B. The results indicate that the estimation of B is robust to small changes in the network structure.

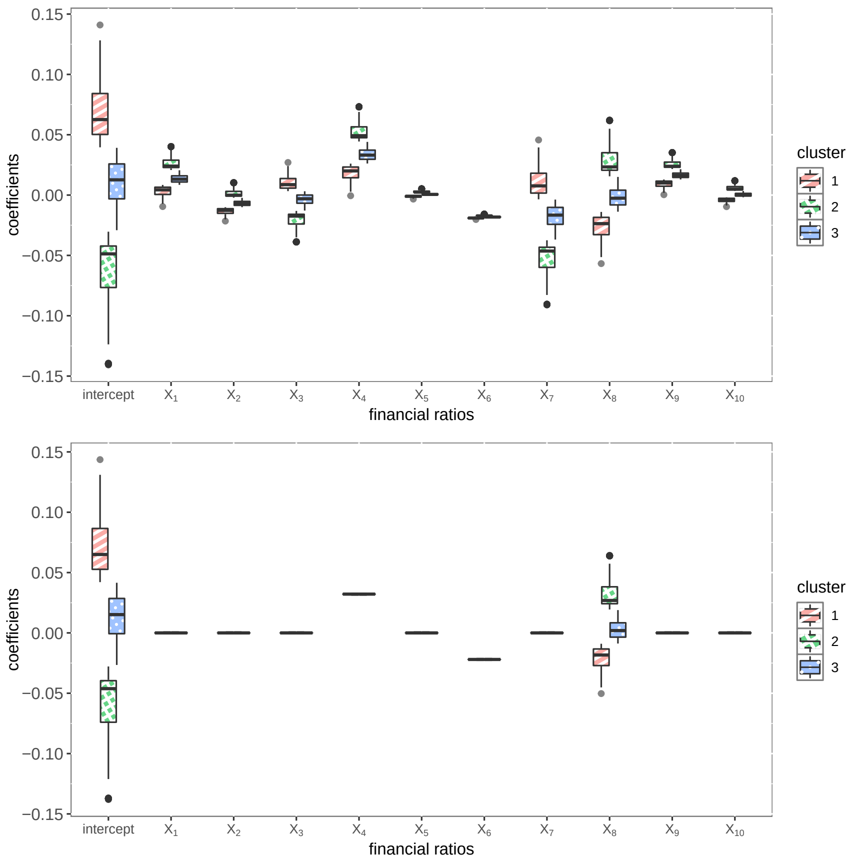

Figure S.1: Boxplots of the estimated regression coefficients for the intercept and 10 financial ratios in three groups separately, which are obtained by fitting all the stocks via NVCM and SNVC. The upper and lower panels depict plots for NVCM and SNVC, respectively. For each covariate, the three boxplots from left to right correspond to the first to the third groups.

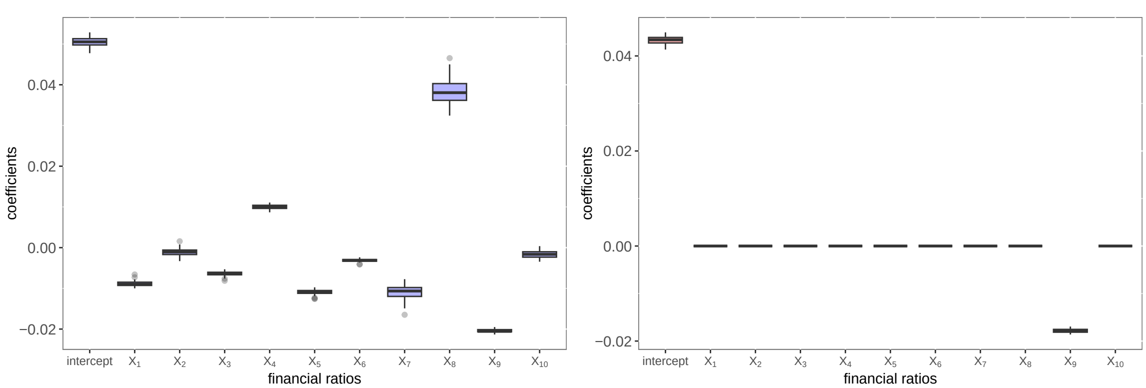

Figure S.2: Boxplots of the estimators of B for the intercept and 10 financial ratios. The left and right panels correspond to the estimates obtained using NVCM and SNVC, respectively.

References

Binkiewicz, N., Vogelstein, J. T. and Rohe, K. (2017). "Covariate-assisted spectral clustering," *Biometrika*, 104, 361-377.

Hallac, D., Leskovec, J. and Boyd, S. (2015). "Network lasso: clustering and optimization in large graphs," *In Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining,* 387-396.

Leisch, F. (2004). "Flexmix: A general framework for finite mixture models and latent glass regression in R," *Journal of Statistical Software*, 11, 1-18.

Li, C. and Li, H. (2008). "Network-constrained regularization and variable selection for analysis of genomic data," *Bioinformatics*, 24, 1175-1182.

Li, T., Levina, E. and Zhu, J. (2020). "Network cross-validation by edge sampling," *Biometrika*, 107, 257-276.

Ma, Z., Ma, Z. and Yuan, H. (2020). "Universal latent space model fitting for large networks with edge covariates," *Journal of Machine Learning Research*, 21, 1-67.

Rubin-Delanchy, P., Cape, J., Tang, M. and Priebe, C. E. (2022). "A statistical interpretation of spectral embedding: the generalised random dot product graph," *Journal of the Royal Statistical Society Series B: Statistical Methodology,* 84, 1446-1473.

Rudelson, M. and Vershynin, R. (2013). "Hanson-wright inequality and sub-gaussian concentration," *Electronic Communications in Probability,* 18, 1-9.