Supplementary Material

A Pairwise Strategic Network Formation Model with Group Heterogeneity: With an Application to International Travel

Tadao Hoshino, Waseda University. Email: thoshino@waseda.jp

A Notations

Variables and parameters

$$\begin{aligned}
W_{i,j} &\equiv (Z_{i,j}^\top, \chi_{n,i}^\top, \chi_{n,j}^\top)^\top \\
\theta &\equiv (\beta^\top, \alpha, \rho)^\top \\
\mathbf{A} &\equiv (A_1, \dots, A_n)^\top, \quad \mathbf{A}_{-1} \equiv (A_2, \dots, A_n)^\top, \quad \mathbf{B} \equiv (B_1, \dots, B_n)^\top \\
\boldsymbol{\gamma} &\equiv (\mathbf{A}^\top, \mathbf{B}^\top)^\top, \quad \boldsymbol{\gamma}_{-1} \equiv (\mathbf{A}_{-1}^\top \mathbf{B}^\top)^\top \\
\Pi &\equiv (\beta^\top, \boldsymbol{\gamma}^\top)^\top \\
\pi_{i,j} &\equiv Z_{i,j}^\top \beta + A_i + B_j = W_{i,j}^\top \Pi \\
\delta &\equiv (\theta^\top, a_2, \dots, a_{K^A}, b_1, \dots, b_{K^B})^\top.
\end{aligned}$$

Functions and derivatives Throughout this appendix, for notational simplicity, we denote $\partial_a g(a) = \partial g(a)/(\partial a)$, $\partial^2_{ab} g(a,b) = \partial^2 g(a,b)/(\partial a \partial b)$, and so fourth.

$$\begin{aligned}
P_{i,j}(\theta, \boldsymbol{\gamma}) &\equiv F(W_{i,j}^\top \Pi) - H(W_{i,j}^\top \Pi, W_{j,i}^\top \Pi + \alpha; \rho) \\
\ell_{i,j}(\theta, \boldsymbol{\gamma}) &\equiv y_{i,j} \ln P_{i,j}(\theta, \boldsymbol{\gamma}) + y_{j,i} \ln P_{j,i}(\theta, \boldsymbol{\gamma}) + (1 - y_{i,j} - y_{j,i}) \ln(1 - P_{i,j}(\theta, \boldsymbol{\gamma}) - P_{j,i}(\theta, \boldsymbol{\gamma})) \\
\mathcal{L}_n(\theta, \boldsymbol{\gamma}) &\equiv \frac{2}{N} \sum_{i=1}^n \sum_{j>i} \ell_{i,j}(\theta, \boldsymbol{\gamma}) \\
p_{1,i,j}(\theta, \boldsymbol{\gamma}) &\equiv \partial_{\pi_{i,j}} P_{i,j}(\theta, \boldsymbol{\gamma}) = f(W_{i,j}^\top \Pi) - H_1(W_{i,j}^\top \Pi, W_{j,i}^\top \Pi + \alpha; \rho) \\
p_{2,i,j}(\theta, \boldsymbol{\gamma}) &\equiv \partial_{\pi_{j,i}} P_{i,j}(\theta, \boldsymbol{\gamma}) = -H_2(W_{i,j}^\top \Pi, W_{j,i}^\top \Pi + \alpha; \rho) \\
H_\rho(\cdot, \cdot, ; \rho) &\equiv \partial_\rho H(\cdot, \cdot; \rho),
\end{aligned}$$

where f is the derivative of F, and $H_l(\cdot, \cdot; \rho)$ is the derivative of $H(\cdot, \cdot; \rho)$ with respect to the l-th argument.

$$\partial_{A_k} \pi_{i,j} = \mathbf{1}\{i = k\}, \;\; \partial_{A_k} \pi_{j,i} = \mathbf{1}\{j = k\}, \;\; \partial_{B_k} \pi_{i,j} = \mathbf{1}\{j = k\}, \;\; \partial_{B_k} \pi_{j,i} = \mathbf{1}\{i = k\}$$

$$\begin{aligned}
\partial_{A_k} P_{i,j}(\theta, \boldsymbol{\gamma}) &= p_{1,i,j}(\theta, \boldsymbol{\gamma}) \mathbf{1}\{i = k\} + p_{2,i,j}(\theta, \boldsymbol{\gamma}) \mathbf{1}\{j = k\} \\
\partial_{B_k} P_{i,j}(\theta, \boldsymbol{\gamma}) &= p_{1,i,j}(\theta, \boldsymbol{\gamma}) \mathbf{1}\{j = k\} + p_{2,i,j}(\theta, \boldsymbol{\gamma}) \mathbf{1}\{i = k\} \\
s_{1,i,j}(\theta, \boldsymbol{\gamma}) \equiv \partial_{\pi_{i,j}} \ell_{i,j}(\theta, \boldsymbol{\gamma}) &= \frac{y_{i,j} p_{1,i,j}(\theta, \boldsymbol{\gamma})}{P_{i,j}(\theta, \boldsymbol{\gamma})} + \frac{y_{j,i} p_{2,j,i}(\theta, \boldsymbol{\gamma})}{P_{j,i}(\theta, \boldsymbol{\gamma})} - \frac{(1 - y_{i,j} - y_{j,i})[p_{1,i,j}(\theta, \boldsymbol{\gamma}) + p_{2,j,i}(\theta, \boldsymbol{\gamma})]}{1 - P_{i,j}(\theta, \boldsymbol{\gamma}) - P_{j,i}(\theta, \boldsymbol{\gamma})} \\
s_{2,i,j}(\theta, \boldsymbol{\gamma}) \equiv \partial_{\pi_{j,i}} \ell_{i,j}(\theta, \boldsymbol{\gamma}) &= \frac{y_{i,j} p_{2,i,j}(\theta, \boldsymbol{\gamma})}{P_{i,j}(\theta, \boldsymbol{\gamma})} + \frac{y_{j,i} p_{1,j,i}(\theta, \boldsymbol{\gamma})}{P_{j,i}(\theta, \boldsymbol{\gamma})} - \frac{(1 - y_{i,j} - y_{j,i})[p_{2,i,j}(\theta, \boldsymbol{\gamma}) + p_{1,j,i}(\theta, \boldsymbol{\gamma})]}{1 - P_{i,j}(\theta, \boldsymbol{\gamma}) - P_{j,i}(\theta, \boldsymbol{\gamma})}
\end{aligned}$$

$$
\begin{aligned}
s_{i,j}^{A_k}(\theta,\boldsymbol{\gamma}) &\equiv \partial_{A_k}\ell_{i,j}(\theta,\boldsymbol{\gamma}) = s_{1,i,j}(\theta,\boldsymbol{\gamma})\mathbf{1}\{i=k\} + s_{2,i,j}(\theta,\boldsymbol{\gamma})\mathbf{1}\{j=k\} \\
s_{i,j}^{B_k}(\theta,\boldsymbol{\gamma}) &\equiv \partial_{B_k}\ell_{i,j}(\theta,\boldsymbol{\gamma}) = s_{1,i,j}(\theta,\boldsymbol{\gamma})\mathbf{1}\{j=k\} + s_{2,i,j}(\theta,\boldsymbol{\gamma})\mathbf{1}\{i=k\} \\
s_{i,j}^{\mathbf{A}}(\theta,\boldsymbol{\gamma}) &\equiv \partial_{\mathbf{A}_{-1}}\ell_{i,j}(\theta,\boldsymbol{\gamma}) = s_{1,i,j}(\theta,\boldsymbol{\gamma})\chi_{n,i,-1} + s_{2,i,j}(\theta,\boldsymbol{\gamma})\chi_{n,j,-1} \\
s_{i,j}^{\mathbf{B}}(\theta,\boldsymbol{\gamma}) &\equiv \partial_{\mathbf{B}}\ell_{i,j}(\theta,\boldsymbol{\gamma}) = s_{1,i,j}(\theta,\boldsymbol{\gamma})\chi_{n,j} + s_{2,i,j}(\theta,\boldsymbol{\gamma})\chi_{n,i} \\
\xi_{i,j}(\theta,\boldsymbol{\gamma}) &\equiv \partial_{\theta}P_{i,j}(\theta,\boldsymbol{\gamma}) = \left[p_{1,i,j}(\theta,\boldsymbol{\gamma})Z_{i,j}^{\top} + p_{2,i,j}(\theta,\boldsymbol{\gamma})Z_{j,i}^{\top},\ p_{2,i,j}(\theta,\boldsymbol{\gamma}),\ -H_{\rho}(W_{i,j}^{\top}\Pi, W_{j,i}^{\top}\Pi+\alpha;\rho)\right]^{\top} \\
s_{i,j}^{\theta}(\theta,\boldsymbol{\gamma}) &\equiv \partial_{\theta}\ell_{i,j}(\theta,\boldsymbol{\gamma}) = \frac{y_{i,j}\xi_{i,j}(\theta,\boldsymbol{\gamma})}{P_{i,j}(\theta,\boldsymbol{\gamma})} + \frac{y_{j,i}\xi_{j,i}(\theta,\boldsymbol{\gamma})}{P_{j,i}(\theta,\boldsymbol{\gamma})} - \frac{(1-y_{i,j}-y_{j,i})[\xi_{i,j}(\theta,\boldsymbol{\gamma})+\xi_{j,i}(\theta,\boldsymbol{\gamma})]}{1-P_{i,j}(\theta,\boldsymbol{\gamma})-P_{j,i}(\theta,\boldsymbol{\gamma})},
\end{aligned}
$$

where $\chi_{n,i,-1}$ and $\chi_{n,j,-1}$ are $(n-1)\times 1$ vectors defined by removing the first element of $\chi_{n,i}$ and $\chi_{n,j}$, respectively. Using these notations, we write

$$
\begin{aligned}
\mathcal{S}_{n,\theta}(\theta,\boldsymbol{\gamma}) &\equiv \partial_{\theta}\mathcal{L}_n(\theta,\boldsymbol{\gamma}) = \frac{2}{N}\sum_{i=1}^{n}\sum_{j>i} s_{i,j}^{\theta}(\theta,\boldsymbol{\gamma}) \\
\mathcal{S}_{n,\boldsymbol{\gamma}}(\theta,\boldsymbol{\gamma}) &\equiv \partial_{\boldsymbol{\gamma}_{-1}}\mathcal{L}_n(\theta,\boldsymbol{\gamma}) = \frac{2}{N}\sum_{i=1}^{n}\sum_{j>i}\left[s_{i,j}^{\mathbf{A}}(\theta,\boldsymbol{\gamma})^{\top},\ s_{i,j}^{\mathbf{B}}(\theta,\boldsymbol{\gamma})^{\top}\right]^{\top}.
\end{aligned}
$$

Further, writing $\ell_{i,j}(\delta) \equiv \ell_{i,j}\left(\theta, \left\{\sum_{k=1}^{K^A} a_k\cdot\mathbf{1}\{i\in\mathcal{C}_{0,k}^{A}\}\right\}, \left\{\sum_{k=1}^{K^B} b_k\cdot\mathbf{1}\{i\in\mathcal{C}_{0,k}^{B}\}\right\}\right)$ so that $\mathcal{L}_n(\delta) = \frac{2}{N}\sum_{i=1}^{n}\sum_{j>i}\ell_{i,j}(\delta)$, we define

$$
\begin{aligned}
s_{i,j}^{a_k}(\delta) &\equiv \partial_{a_k}\ell_{i,j}(\delta) = s_{1,i,j}(\delta)\mathbf{1}\{i\in\mathcal{C}_{0,k}^{A}\} + s_{2,i,j}(\delta)\mathbf{1}\{j\in\mathcal{C}_{0,k}^{A}\} \\
s_{i,j}^{b_k}(\delta) &\equiv \partial_{b_k}\ell_{i,j}(\delta) = s_{1,i,j}(\delta)\mathbf{1}\{j\in\mathcal{C}_{0,k}^{B}\} + s_{2,i,j}(\delta)\mathbf{1}\{i\in\mathcal{C}_{0,k}^{B}\} \\
s_{i,j}^{\delta}(\delta) &\equiv \partial_{\delta}\ell_{i,j}(\delta) = \left[s_{i,j}^{\theta}(\delta)^{\top},\ s_{i,j}^{a_2}(\delta),\ldots,s_{i,j}^{a_{K^A}}(\delta), s_{i,j}^{b_1}(\delta),\ldots,s_{i,j}^{b_{K^B}}(\delta)\right]^{\top},
\end{aligned}
$$

where the definitions of $s_{1,i,j}(\delta)$, $s_{2,i,j}(\delta)$, and $s_{i,j}^{\theta}(\delta)$ should be clear from the context.

Hessian matrix Define

$$
\begin{aligned}
h_{11,i,j}(\theta,\boldsymbol{\gamma}) &\equiv \partial_{\pi_{i,j}}s_{1,i,j}(\theta,\boldsymbol{\gamma}) \\
h_{12,i,j}(\theta,\boldsymbol{\gamma}) &\equiv \partial_{\pi_{j,i}}s_{1,i,j}(\theta,\boldsymbol{\gamma}) = \partial_{\pi_{i,j}}s_{2,i,j}(\theta,\boldsymbol{\gamma}) \\
h_{22,i,j}(\theta,\boldsymbol{\gamma}) &\equiv \partial_{\pi_{j,i}}s_{2,i,j}(\theta,\boldsymbol{\gamma}).
\end{aligned}
$$

It is easy to see that

$$
\begin{aligned}
\mathbb{E}h_{11,i,j}(\theta_0,\boldsymbol{\gamma}_0) &= -\frac{p_{1,i,j}^2}{P_{i,j}} - \frac{p_{2,j,i}^2}{P_{j,i}} - \frac{[p_{1,i,j}+p_{2,j,i}]^2}{1-P_{i,j}-P_{j,i}} \\
\mathbb{E}h_{12,i,j}(\theta_0,\boldsymbol{\gamma}_0) &= -\frac{p_{1,i,j}p_{2,i,j}}{P_{i,j}} - \frac{p_{1,j,i}p_{2,j,i}}{P_{j,i}} - \frac{[p_{1,i,j}+p_{2,j,i}][p_{2,i,j}+p_{1,j,i}]}{1-P_{i,j}-P_{j,i}},
\end{aligned}
\tag{A.1}
$$

where we have used $p_{1,i,j}$, $p_{2,i,j}$, and $P_{i,j}$ to denote $p_{1,i,j}(\theta_0,\boldsymbol{\gamma}_0)$, $p_{2,i,j}(\theta_0,\boldsymbol{\gamma}_0)$, and $P_{i,j}(\theta_0,\boldsymbol{\gamma}_0)$, respectively, for simplicity. Hereinafter, when the dependence on the parameters $(\theta,\boldsymbol{\gamma})$ is suppressed, it means that the functions are evaluated at the true value $(\theta_0,\boldsymbol{\gamma}_0)$.

Note that, since $\ell_{i,j}(\theta,\boldsymbol{\gamma}) = \ell_{j,i}(\theta,\boldsymbol{\gamma})$, we have $h_{11,i,j}(\theta,\boldsymbol{\gamma}) = \partial^2_{\pi_{i,j}\pi_{i,j}}\ell_{i,j}(\theta,\boldsymbol{\gamma}) = \partial^2_{\pi_{i,j}\pi_{i,j}}\ell_{j,i}(\theta,\boldsymbol{\gamma}) = h_{22,j,i}(\theta,\boldsymbol{\gamma})$ and $h_{12,i,j}(\theta,\boldsymbol{\gamma}) = h_{12,j,i}(\theta,\boldsymbol{\gamma})$. By tedious calculations, we have

$$\partial^2_{A_l A_k}\mathcal{L}_n(\theta,\boldsymbol{\gamma}) = \frac{2}{N}\sum_{j\neq k} h_{11,k,j}(\theta,\boldsymbol{\gamma})\mathbf{1}\{l=k\} + \frac{2}{N}h_{12,l,k}(\theta,\boldsymbol{\gamma})\mathbf{1}\{l\neq k\} \ \ \text{(for } l,k\geq 2\text{)}$$
$$\partial^2_{B_l B_k}\mathcal{L}_n(\theta,\boldsymbol{\gamma}) = \frac{2}{N}\sum_{j\neq k} h_{11,j,k}(\theta,\boldsymbol{\gamma})\mathbf{1}\{l=k\} + \frac{2}{N}h_{12,l,k}(\theta,\boldsymbol{\gamma})\mathbf{1}\{l\neq k\} \ \ \text{(for } l,k\geq 1\text{)}$$
$$\partial^2_{A_l B_k}\mathcal{L}_n(\theta,\boldsymbol{\gamma}) = \frac{2}{N}\sum_{j\neq k} h_{12,k,j}(\theta,\boldsymbol{\gamma})\mathbf{1}\{l=k\} + \frac{2}{N}h_{11,l,k}(\theta,\boldsymbol{\gamma})\mathbf{1}\{l\neq k\} \ \ \text{(for } l\geq 2,\ k\geq 1\text{)}.$$

Hence,

$$\begin{aligned}
\mathcal{H}_{n,\mathbf{AA}}(\theta,\boldsymbol{\gamma}) &\equiv \partial^2_{\mathbf{A}_{-1}\mathbf{A}_{-1}^\top}\mathcal{L}_n(\theta,\boldsymbol{\gamma}) = \frac{2}{N}\begin{pmatrix} \sum_{j\neq 2} h_{11,2,j}(\theta,\boldsymbol{\gamma}) & \cdots & h_{12,2,n}(\theta,\boldsymbol{\gamma}) \\ \vdots & \ddots & \vdots \\ h_{12,n,2}(\theta,\boldsymbol{\gamma}) & \cdots & \sum_{j\neq n} h_{11,n,j}(\theta,\boldsymbol{\gamma}) \end{pmatrix} \\
\mathcal{H}_{n,\mathbf{BB}}(\theta,\boldsymbol{\gamma}) &\equiv \partial^2_{\mathbf{BB}^\top}\mathcal{L}_n(\theta,\boldsymbol{\gamma}) = \frac{2}{N}\begin{pmatrix} \sum_{j\neq 1} h_{11,j,1}(\theta,\boldsymbol{\gamma}) & \cdots & h_{12,1,n}(\theta,\boldsymbol{\gamma}) \\ \vdots & \ddots & \vdots \\ h_{12,n,1}(\theta,\boldsymbol{\gamma}) & \cdots & \sum_{j\neq n} h_{11,j,n}(\theta,\boldsymbol{\gamma}) \end{pmatrix} \\
\mathcal{H}_{n,\mathbf{AB}}(\theta,\boldsymbol{\gamma}) &\equiv \partial^2_{\mathbf{A}_{-1}\mathbf{B}^\top}\mathcal{L}_n(\theta,\boldsymbol{\gamma}) = \frac{2}{N}\begin{pmatrix} h_{11,2,1}(\theta,\boldsymbol{\gamma}) & \sum_{j\neq 2} h_{12,2,j}(\theta,\boldsymbol{\gamma}) & \cdots & h_{11,2,n}(\theta,\boldsymbol{\gamma}) \\ \vdots & \vdots & \ddots & \vdots \\ h_{11,n,1}(\theta,\boldsymbol{\gamma}) & h_{11,n,2}(\theta,\boldsymbol{\gamma}) & \cdots & \sum_{j\neq n} h_{12,n,j}(\theta,\boldsymbol{\gamma}) \end{pmatrix} \\
\mathcal{H}_{n,\boldsymbol{\gamma}\boldsymbol{\gamma}}(\theta,\boldsymbol{\gamma}) &= \begin{pmatrix} \mathcal{H}_{n,\mathbf{AA}}(\theta,\boldsymbol{\gamma}) & \mathcal{H}_{n,\mathbf{AB}}(\theta,\boldsymbol{\gamma}) \\ \mathcal{H}_{n,\mathbf{BA}}(\theta,\boldsymbol{\gamma}) & \mathcal{H}_{n,\mathbf{BB}}(\theta,\boldsymbol{\gamma}) \end{pmatrix}
\end{aligned} \tag{A.2}$$

B Proofs of Theorems

Proof of Theorem 2.1

(i) We first confirm that the true parameter vector $(\theta_0,\boldsymbol{\gamma}_0)$ is a maximizer of $\mathbb{E}\mathcal{L}_n(\theta,\boldsymbol{\gamma})$. We can observe that

$$\begin{aligned}
\mathbb{E}\mathcal{L}_n(\theta,\boldsymbol{\gamma}) - \mathbb{E}\mathcal{L}_n(\theta_0,\boldsymbol{\gamma}_0) &= \frac{1}{N}\sum_{i=1}^n\sum_{j\neq i}\mathbb{E}\left\{\ln\left[P_{i,j}(\theta,\boldsymbol{\gamma})^{y_{i,j}}P_{j,i}(\theta,\boldsymbol{\gamma})^{y_{j,i}}[1-P_{i,j}(\theta,\boldsymbol{\gamma})-P_{j,i}(\theta,\boldsymbol{\gamma})]^{1-y_{i,j}-y_{j,i}}\right]\right. \\
&\qquad \left. - \ln\left[P_{i,j}^{y_{i,j}}P_{j,i}^{y_{j,i}}[1-P_{i,j}-P_{j,i}]^{1-y_{i,j}-y_{j,i}}\right]\right\} \\
&= \frac{1}{N}\sum_{i=1}^n\sum_{j\neq i}\mathbb{E}\left\{\ln\frac{P_{i,j}(\theta,\boldsymbol{\gamma})^{y_{i,j}}P_{j,i}(\theta,\boldsymbol{\gamma})^{y_{j,i}}[1-P_{i,j}(\theta,\boldsymbol{\gamma})-P_{j,i}(\theta,\boldsymbol{\gamma})]^{1-y_{i,j}-y_{j,i}}}{P_{i,j}^{y_{i,j}}P_{j,i}^{y_{j,i}}[1-P_{i,j}-P_{j,i}]^{1-y_{i,j}-y_{j,i}}}\right\} \\
&\leq \frac{1}{N}\sum_{i=1}^n\sum_{j\neq i}\ln\mathbb{E}\left\{\frac{P_{i,j}(\theta,\boldsymbol{\gamma})^{y_{i,j}}P_{j,i}(\theta,\boldsymbol{\gamma})^{y_{j,i}}[1-P_{i,j}(\theta,\boldsymbol{\gamma})-P_{j,i}(\theta,\boldsymbol{\gamma})]^{1-y_{i,j}-y_{j,i}}}{P_{i,j}^{y_{i,j}}P_{j,i}^{y_{j,i}}[1-P_{i,j}-P_{j,i}]^{1-y_{i,j}-y_{j,i}}}\right\},
\end{aligned} \tag{B.1}$$

where the last inequality follows from Jensen's inequality. Further,

$$\mathbb{E}\left\{\frac{P_{i,j}(\theta,\boldsymbol{\gamma})^{y_{i,j}}P_{j,i}(\theta,\boldsymbol{\gamma})^{y_{j,i}}[1-P_{i,j}(\theta,\boldsymbol{\gamma})-P_{j,i}(\theta,\boldsymbol{\gamma})]^{1-y_{i,j}-y_{j,i}}}{P_{i,j}^{y_{i,j}}P_{j,i}^{y_{j,i}}[1-P_{i,j}-P_{j,i}]^{1-y_{i,j}-y_{j,i}}}\right\}$$
$$=\mathbb{E}[y_{i,j}]\frac{P_{i,j}(\theta,\boldsymbol{\gamma})}{P_{i,j}}+\mathbb{E}[y_{j,i}]\frac{P_{j,i}(\theta,\boldsymbol{\gamma})}{P_{j,i}}+\mathbb{E}[1-y_{i,j}-y_{j,i}]\frac{1-P_{i,j}(\theta,\boldsymbol{\gamma})-P_{j,i}(\theta,\boldsymbol{\gamma})}{1-P_{i,j}-P_{j,i}}=1,$$

implying that the left-hand side term of (B.1) is less than or at most equal to zero for any given $(\theta,\boldsymbol{\gamma})$. Then, since ρ_0 is known, it is sufficient to show that

$$\frac{1}{N}\sum_{i=1}^{n}\sum_{j\neq i}\mathbf{1}\left\{P_{i,j}((\beta^\top,\alpha,\rho_0)^\top,\boldsymbol{\gamma})\neq P_{i,j}(\theta_0,\boldsymbol{\gamma}_0)\right\}>0$$

for all sufficiently large n and for all $(\beta,\alpha,\boldsymbol{\gamma})\in\mathcal{B}\times\mathcal{A}\times\mathbb{C}_n$ such that $(\beta,\alpha,\boldsymbol{\gamma})\neq(\beta_0,\alpha_0,\boldsymbol{\gamma}_0)$. The existence of pairs satisfying $P_{i,j}((\beta^\top,\alpha,\rho_0)^\top,\boldsymbol{\gamma})\neq P_{i,j}(\theta_0,\boldsymbol{\gamma}_0)$ contributes to a non-negligible difference between $\mathbb{E}\mathcal{L}_n(\theta,\boldsymbol{\gamma})$ and $\mathbb{E}\mathcal{L}_n(\theta_0,\boldsymbol{\gamma}_0)$, allowing us to distinguish $(\theta,\boldsymbol{\gamma})$ and $(\theta_0,\boldsymbol{\gamma}_0)$. Here, by Assumptions 2.1(i) and (ii), $F(a)-H(a,b;\rho_0)$ is strictly increasing in a and decreasing in b, respectively. Therefore, we have

$$W_{i,j}^\top\Pi>W_{i,j}^\top\Pi_0,\ \ W_{j,i}^\top\Pi+\alpha<W_{j,i}^\top\Pi_0+\alpha_0\Longrightarrow P_{i,j}((\beta^\top,\alpha,\rho_0)^\top,\boldsymbol{\gamma})>P_{i,j}(\theta_0,\boldsymbol{\gamma}_0)$$
$$W_{i,j}^\top\Pi<W_{i,j}^\top\Pi_0,\ \ W_{j,i}^\top\Pi+\alpha>W_{j,i}^\top\Pi_0+\alpha_0\Longrightarrow P_{i,j}((\beta^\top,\alpha,\rho_0)^\top,\boldsymbol{\gamma})<P_{i,j}(\theta_0,\boldsymbol{\gamma}_0).$$

Then, Assumption 2.4 gives the desired result.

(ii) Since $(\theta_0,\boldsymbol{\gamma}_0)$ is a maximizer of $\mathbb{E}\mathcal{L}_n(\theta,\boldsymbol{\gamma})$ as confirmed above, ρ_0 must be a maximizer of $\mathcal{L}_n^*(\rho)$, where the definition of $\mathcal{L}_n^*(\rho)$ can be found in the statement of the theorem, and $(\widetilde{\beta}_0(\rho_0),\widetilde{\alpha}_0(\rho_0),\widetilde{\boldsymbol{\gamma}}_0(\rho_0))=(\beta_0,\alpha_0,\boldsymbol{\gamma}_0)$ holds. For all $\rho\in\mathcal{R}$, $\mathbb{E}\mathcal{L}_n((\beta^\top,\alpha,\rho)^\top,\boldsymbol{\gamma})-\mathbb{E}\mathcal{L}_n((\widetilde{\beta}_0(\rho)^\top,\widetilde{\alpha}_0(\rho),\rho)^\top,\widetilde{\boldsymbol{\gamma}}_0(\rho))\leq 0$ holds by definition. Then, by the same argument as in the proof of (i), we can identify $(\widetilde{\beta}_0(\rho),\widetilde{\alpha}_0(\rho),\widetilde{\boldsymbol{\gamma}}_0(\rho))$ uniquely for all $\rho\in\mathcal{R}$ as Assumption 2.4 is independent of the value of ρ. Thus, if ρ_0 is identified as a unique maximizer of $\mathcal{L}_n^*(\rho)$, all the parameters of the model are identified. □

Proof of Theorem 3.1

(i) First, note that Assumptions 2.1–2.3 imply that there exist constants $\kappa_1,\kappa_2\in(0,1)$ such that $P_{i,j}(\theta,\boldsymbol{\gamma})\in(\kappa_1,1-\kappa_1)$ and $1-P_{i,j}(\theta,\boldsymbol{\gamma})-P_{j,i}(\theta,\boldsymbol{\gamma})\in(\kappa_2,1-\kappa_2)$ for all possible parameter values. Observe that

$$\begin{aligned}&\mathcal{L}_n(\theta,\boldsymbol{\gamma})-\mathbb{E}\mathcal{L}_n(\theta,\boldsymbol{\gamma})\\&=\frac{1}{N}\sum_{i=1}^{n}\sum_{j\neq i}\left[2(y_{i,j}-\mathbb{E}y_{i,j})\ln P_{i,j}(\theta,\boldsymbol{\gamma})-2(y_{i,j}-\mathbb{E}y_{i,j})\ln\left(1-P_{i,j}(\theta,\boldsymbol{\gamma})-P_{j,i}(\theta,\boldsymbol{\gamma})\right)\right]\\&=\frac{2}{N}\sum_{i=1}^{n}\sum_{j\neq i}(y_{i,j}-\mathbb{E}y_{i,j})\psi_{i,j}(\theta,\boldsymbol{\gamma}),\end{aligned}$$

where $\psi_{i,j}(\theta,\boldsymbol{\gamma})\equiv\ln\left[P_{i,j}(\theta,\boldsymbol{\gamma})/\left(1-P_{i,j}(\theta,\boldsymbol{\gamma})-P_{j,i}(\theta,\boldsymbol{\gamma})\right)\right]$. Let $\bar{\psi}\equiv\ln((1-\kappa_1)/\kappa_2)$, so that

$$-(1-\kappa_1)\bar{\psi}<(y_{i,j}-\mathbb{E}y_{i,j})\psi_{i,j}(\theta,\boldsymbol{\gamma})<(1-\kappa_1)\bar{\psi},$$

where the inequalities are uniform in $(\theta, \boldsymbol{\gamma}) \in \Theta \times \mathbb{C}_n$. By the triangle inequality,

$$|\mathcal{L}_n(\theta, \boldsymbol{\gamma}) - \mathbb{E}\mathcal{L}_n(\theta, \boldsymbol{\gamma})| \leq \frac{2}{n}\sum_{i=1}^{n}\left|\frac{1}{n-1}\sum_{j\neq i}(y_{i,j} - \mathbb{E}y_{i,j})\psi_{i,j}(\theta, \boldsymbol{\gamma})\right|.$$

Further, by Hoeffding's inequality,

$$\begin{aligned}\Pr\left(\left|\frac{1}{n-1}\sum_{j\neq i}(y_{i,j} - \mathbb{E}y_{i,j})\psi_{i,j}(\theta, \boldsymbol{\gamma})\right| > t\right) &\leq 2\exp\left(-\frac{2(n-1)^2t^2}{\sum_{j\neq i}(2(1-\kappa_1)\bar{\psi})^2}\right) \\ &= 2\exp\left(-\frac{(n-1)t^2}{2(1-\kappa_1)^2\bar{\psi}^2}\right).\end{aligned}$$

Hence, Boole's inequality gives

$$\Pr\left(\max_{1\leq i\leq n}\left|\frac{1}{n-1}\sum_{j\neq i}(y_{i,j} - \mathbb{E}y_{i,j})\psi_{i,j}(\theta, \boldsymbol{\gamma})\right| > t\right) \leq 2n\exp\left(-\frac{(n-1)t^2}{2(1-\kappa_1)^2\bar{\psi}^2}\right).$$

Setting $t = C\sqrt{\ln n/n}$ for a sufficiently large constant $C > 0$, we have

$$\begin{aligned}2n\exp\left(-\frac{(n-1)t^2}{2(1-\kappa_1)^2\bar{\psi}^2}\right) &= 2n\exp\left(-\frac{n-1}{2(1-\kappa_1)^2\bar{\psi}^2}\frac{C^2\ln n}{n}\right) \\ &= 2\exp\left(\ln n - \left(\frac{C^2(n-1)/n}{2(1-\kappa_1)^2\bar{\psi}^2}\right)\ln n\right) \to 0 \quad \text{as} \quad n\to\infty.\end{aligned}$$

This implies that

$$\sup_{(\theta,\boldsymbol{\gamma})\,\in\,\Theta\times\mathbb{C}_n}|\mathcal{L}_n(\theta, \boldsymbol{\gamma}) - \mathbb{E}\mathcal{L}_n(\theta, \boldsymbol{\gamma})| = O_P\left(\sqrt{\frac{\ln n}{n}}\right). \tag{B.2}$$

Then, with Assumption 3.1, the rest of the proof follows from the same argument as in the proof of Theorem 2 in Graham (2017). □

(ii) (iii) We prove the result by contradiction. Suppose that there exists a positive constant c such that

$$\max\left\{\frac{1}{n}\sum_{i=1}^{n}\left|\widehat{A}_{n,i} - A_{0,i}\right|,\ \frac{1}{n}\sum_{i=1}^{n}\left|\widehat{B}_{n,i} - B_{0,i}\right|\right\} \geq c > 0$$

w.p.a.1. This implies that there is a non-vanishing potion of observations with either or both $\widehat{A}_{n,i}$ and $\widehat{B}_{n,i}$ being not in the neighborhood of $A_{0,i}$ and $B_{0,i}$, respectively. Therefore, by Assumption 3.1, there exist a constant $\eta(c) > 0$ and $n(c) < \infty$ such that

$$4\eta(c) < \mathbb{E}\mathcal{L}_n(\theta_0, \mathbf{A}_0, \mathbf{B}_0) - \mathbb{E}\mathcal{L}_n(\theta_0, \widehat{\mathbf{A}}_n, \widehat{\mathbf{B}}_n) \tag{B.3}$$

for all $n \geq n(c)$. Here, $\mathbb{E}\mathcal{L}_n(\theta_0, \widehat{\mathbf{A}}_n, \widehat{\mathbf{B}}_n)$ is read as $\mathbb{E}\mathcal{L}_n(\theta_0, \mathbf{A}, \mathbf{B})|_{(\mathbf{A},\mathbf{B})=(\widehat{\mathbf{A}}_n,\widehat{\mathbf{B}}_n)}$; the same applies to what follows.

Note that (B.2) implies that

$$\mathbb{E}\mathcal{L}_n(\theta_0, \mathbf{A}_0, \mathbf{B}_0) < \mathcal{L}_n(\theta_0, \mathbf{A}_0, \mathbf{B}_0) + \eta(c) \tag{B.4}$$

w.p.a.1. By the definition of the ML estimator,

$$\mathcal{L}_n(\theta_0, \mathbf{A}_0, \mathbf{B}_0) < \mathcal{L}_n(\widehat{\theta}_n, \widehat{\mathbf{A}}_n, \widehat{\mathbf{B}}_n) + \eta(c). \tag{B.5}$$

In addition, by the continuous mapping theorem and result (i), we have

$$\mathcal{L}_n(\widehat{\theta}_n, \widehat{\mathbf{A}}_n, \widehat{\mathbf{B}}_n) < \mathcal{L}_n(\theta_0, \widehat{\mathbf{A}}_n, \widehat{\mathbf{B}}_n) + \eta(c) \tag{B.6}$$

w.p.a.1. Now, combining the inequalities (B.3)–(B.6) gives

$$\begin{aligned}
\mathbb{E}\mathcal{L}_n(\theta_0, \widehat{\mathbf{A}}_n, \widehat{\mathbf{B}}_n) &< \mathbb{E}\mathcal{L}_n(\theta_0, \mathbf{A}_0, \mathbf{B}_0) - 4\eta(c) \\
&< \mathcal{L}_n(\theta_0, \mathbf{A}_0, \mathbf{B}_0) - 3\eta(c) \\
&< \mathcal{L}_n(\widehat{\theta}_n, \widehat{\mathbf{A}}_n, \widehat{\mathbf{B}}_n) - 2\eta(c) \\
&< \mathcal{L}_n(\theta_0, \widehat{\mathbf{A}}_n, \widehat{\mathbf{B}}_n) - \eta(c)
\end{aligned}$$

w.p.a.1. The last line implies that $\eta(c) < \mathcal{L}_n(\theta_0, \widehat{\mathbf{A}}_n, \widehat{\mathbf{B}}_n) - \mathbb{E}\mathcal{L}_n(\theta_0, \widehat{\mathbf{A}}_n, \widehat{\mathbf{B}}_n)$ w.p.a.1; however, this contradicts with (B.2). Hence, as the choice of c is arbitrary, we obtain the desired result. □

(iv) Note that, for each i ($i \neq 1$), it holds that

$$\widehat{A}_{n,i} = \underset{A_i \in \mathbb{A}}{\operatorname{argmax}} \, \mathcal{L}_n(\widehat{\theta}_n, A_i, \widehat{\mathbf{A}}_{n,-i}, \widehat{\mathbf{B}}_n), \qquad A_{0,i} = \underset{A_i \in \mathbb{A}}{\operatorname{argmax}} \, \mathbb{E}\mathcal{L}_n(\theta_0, A_i, \mathbf{A}_{0,-i}, \mathbf{B}_0),$$

where $\mathbf{A}_{-i} \equiv (A_1, \ldots, A_{i-1}, A_{i+1}, \ldots, A_n)^\top$, and $\mathcal{L}_n(\theta, A_i, \mathbf{A}_{-i}, \mathbf{B}) = \mathcal{L}_n(\theta, \mathbf{A}, \mathbf{B})$. Pick any $c > 0$, and let $\mathbb{A}_i^c \equiv \{A \in \mathbb{A} : |A - A_{0,i}| \geq c\}$. Define $\varepsilon_n(c)$ as follows:

$$\varepsilon_n(c) \equiv \min_{2 \leq i \leq n} \left[\mathbb{E}\mathcal{L}_n(\theta_0, A_{0,i}, \mathbf{A}_{0,-i}, \mathbf{B}_0) - \max_{A_i \in \mathbb{A}_i^c} \mathbb{E}\mathcal{L}_n(\theta_0, A_i, \mathbf{A}_{0,-i}, \mathbf{B}_0) \right].$$

By Assumption 3.1, there exists $n(c) < \infty$ such that $\varepsilon_n(c)$ is strictly larger than zero for all $n \geq n(c)$. By the definition of $\widehat{A}_{n,i}$, we have

$$\mathcal{L}_n(\widehat{\theta}_n, \widehat{A}_{n,i}, \widehat{\mathbf{A}}_{n,-i}, \widehat{\mathbf{B}}_n) > \mathcal{L}_n(\widehat{\theta}_n, A_{0,i}, \widehat{\mathbf{A}}_{n,-i}, \widehat{\mathbf{B}}_n) - \varepsilon_n(c)/5. \tag{B.7}$$

By the triangle inequality,

$$\begin{aligned}
&\left| \mathcal{L}_n(\widehat{\theta}_n, A_i, \widehat{\mathbf{A}}_{n,-i}, \widehat{\mathbf{B}}_n) - \mathcal{L}_n(\theta_0, A_i, \mathbf{A}_{0,-i}, \mathbf{B}_0) \right| \\
&\leq \left| \mathcal{L}_n(\widehat{\theta}_n, A_i, \widehat{\mathbf{A}}_{n,-i}, \widehat{\mathbf{B}}_n) - \mathcal{L}_n(\theta_0, A_i, \widehat{\mathbf{A}}_{n,-i}, \widehat{\mathbf{B}}_n) \right| + \left| \mathcal{L}_n(\theta_0, A_i, \widehat{\mathbf{A}}_{n,-i}, \widehat{\mathbf{B}}_n) - \mathcal{L}_n(\theta_0, A_i, \mathbf{A}_{0,-i}, \widehat{\mathbf{B}}_n) \right| \\
&\quad + \left| \mathcal{L}_n(\theta_0, A_i, \mathbf{A}_{0,-i}, \widehat{\mathbf{B}}_n) - \mathcal{L}_n(\theta_0, A_i, \mathbf{A}_{0,-i}, \mathbf{B}_0) \right|
\end{aligned}$$

$$\leq \left|\partial_{\mathbf{A}_{-i}^{\top}}\mathcal{L}_n(\theta_0, A_i, \bar{\mathbf{A}}_{n,-i}, \widehat{\mathbf{B}}_n)[\widehat{\mathbf{A}}_{n,-i} - \mathbf{A}_{0,-i}]\right| + \left|\partial_{\mathbf{B}^{\top}}\mathcal{L}_n(\theta_0, A_i, \mathbf{A}_{0,-i}, \bar{\mathbf{B}}_n)[\widehat{\mathbf{B}}_n - \mathbf{B}_0]\right| + o_P(1),$$

where the second inequality follows from the mean value expansion and result (i), $\bar{\mathbf{A}}_{n,-i} \in [\widehat{\mathbf{A}}_{n,-i}, \mathbf{A}_{0,-i}]$, and $\bar{\mathbf{B}}_n \in [\widehat{\mathbf{B}}_n, \mathbf{B}_0]$. Here, the first term on the right-hand side has the following form:

$$\begin{aligned}
&\partial_{\mathbf{A}_{-k}^{\top}}\mathcal{L}_n(\theta_0, A_k, \bar{\mathbf{A}}_{n,-k}, \widehat{\mathbf{B}}_n)[\widehat{\mathbf{A}}_{n,-k} - \mathbf{A}_{0,-k}] \\
&= \frac{2}{N}\sum_{i=1}^{n}\sum_{j>i}\partial_{\mathbf{A}_{-k}^{\top}}\ell_{i,j}(\theta_0, A_k, \bar{\mathbf{A}}_{n,-k}, \widehat{\mathbf{B}}_n)[\widehat{\mathbf{A}}_{n,-k} - \mathbf{A}_{0,-k}] \\
&= \frac{2}{N}\sum_{i=1}^{n}\sum_{j>i}s_{1,i,j}(\theta_0, A_k, \bar{\mathbf{A}}_{n,-k}, \widehat{\mathbf{B}}_n)\chi_{n,i,-k}^{\top}[\widehat{\mathbf{A}}_{n,-k} - \mathbf{A}_{0,-k}] + \frac{2}{N}\sum_{i=1}^{n}\sum_{j>i}s_{2,i,j}(\theta_0, A_k, \bar{\mathbf{A}}_{n,-k}, \widehat{\mathbf{B}}_n)\chi_{n,j,-k}^{\top}[\widehat{\mathbf{A}}_{n,-k} - \mathbf{A}_{0,-k}] \\
&= \frac{2}{N}\sum_{i=1}^{n}\sum_{j>i}s_{1,i,j}(\theta_0, A_k, \bar{\mathbf{A}}_{n,-k}, \widehat{\mathbf{B}}_n)[\widehat{A}_{n,i} - A_{0,i}]\mathbf{1}\{i \neq k\} + \frac{2}{N}\sum_{i=1}^{n}\sum_{j>i}s_{2,i,j}(\theta_0, A_k, \bar{\mathbf{A}}_{n,-k}, \widehat{\mathbf{B}}_n)[\widehat{A}_{n,j} - A_{0,j}]\mathbf{1}\{j \neq k\} \\
&= \frac{2}{N}\sum_{i=1}^{n}\sum_{j>i}s_{1,i,j}(\theta_0, A_k, \bar{\mathbf{A}}_{n,-k}, \widehat{\mathbf{B}}_n)[\widehat{A}_{n,i} - A_{0,i}]\mathbf{1}\{i \neq k\} + \frac{2}{N}\sum_{j=1}^{n}\sum_{i>j}s_{1,i,j}(\theta_0, A_k, \bar{\mathbf{A}}_{n,-k}, \widehat{\mathbf{B}}_n)[\widehat{A}_{n,i} - A_{0,i}]\mathbf{1}\{i \neq k\} \\
&= \frac{2}{N}\sum_{i=1}^{n}\sum_{j\neq i}s_{1,i,j}(\theta_0, A_k, \bar{\mathbf{A}}_{n,-k}, \widehat{\mathbf{B}}_n)[\widehat{A}_{n,i} - A_{0,i}]\mathbf{1}\{i \neq k\},
\end{aligned}$$

where $\chi_{n,i,-k}$ and $\chi_{n,j,-k}$ are $(n-1)\times 1$ vectors defined by removing the k-th element of $\chi_{n,i}$ and $\chi_{n,j}$, respectively, and the last equality holds because $s_{2,i,j}(\theta, \mathbf{A}, \mathbf{B}) = \partial_{\pi_{j,i}}\ell_{i,j}(\theta, \mathbf{A}, \mathbf{B}) = \partial_{\pi_{j,i}}\ell_{j,i}(\theta, \mathbf{A}, \mathbf{B}) = s_{1,j,i}(\theta, \mathbf{A}, \mathbf{B})$. Then, for a constant $c > 0$ independent of A_k and k, we have

$$\left|\partial_{\mathbf{A}_{-k}^{\top}}\mathcal{L}_n(\theta_0, A_k, \bar{\mathbf{A}}_{n,-k}, \widehat{\mathbf{B}}_n)[\widehat{\mathbf{A}}_{n,-k} - \mathbf{A}_{0,-k}]\right| \leq \frac{c}{n}\sum_{i=1}^{n}\left|\widehat{A}_{n,i} - A_{0,i}\right| = o_P(1)$$

by result (ii). Based on the same argument, we can also show that $\left|\partial_{\mathbf{B}^{\top}}\mathcal{L}_n(\theta_0, A_i, \mathbf{A}_{0,-i}, \bar{\mathbf{B}}_n)[\widehat{\mathbf{B}}_n - \mathbf{B}_0]\right| = o_P(1)$, implying that

$$\left|\mathcal{L}_n(\widehat{\theta}_n, A_i, \widehat{\mathbf{A}}_{n,-i}, \widehat{\mathbf{B}}_n) - \mathcal{L}_n(\theta_0, A_i, \mathbf{A}_{0,-i}, \mathbf{B}_0)\right| = o_P(1)$$

uniformly in $A_i \in \mathbb{A}$ and i. Similarly, we can show that $\left|\mathbb{E}\mathcal{L}_n(\widehat{\theta}_n, A_i, \widehat{\mathbf{A}}_{n,-i}, \widehat{\mathbf{B}}_n) - \mathbb{E}\mathcal{L}_n(\theta_0, A_i, \mathbf{A}_{0,-i}, \mathbf{B}_0)\right| = o_P(1)$. Hence, the following inequalities hold w.p.a.1:

$$\mathcal{L}_n(\theta_0, A_i, \mathbf{A}_{0,-i}, \mathbf{B}_0) > \mathcal{L}_n(\widehat{\theta}_n, A_i, \widehat{\mathbf{A}}_{n,-i}, \widehat{\mathbf{B}}_n) - \varepsilon_n(c)/5 \tag{B.8}$$

$$\mathbb{E}\mathcal{L}_n(\widehat{\theta}_n, A_i, \widehat{\mathbf{A}}_{n,-i}, \widehat{\mathbf{B}}_n) > \mathbb{E}\mathcal{L}_n(\theta_0, A_i, \mathbf{A}_{0,-i}, \mathbf{B}_0) - \varepsilon_n(c)/5 \tag{B.9}$$

uniformly in $A_i \in \mathbb{A}$ and i. In addition, (B.2) implies that

$$\mathbb{E}\mathcal{L}_n(\theta_0, \widehat{A}_{n,i}, \mathbf{A}_{0,-i}, \mathbf{B}_0) > \mathcal{L}_n(\theta_0, \widehat{A}_{n,i}, \mathbf{A}_{0,-i}, \mathbf{B}_0) - \varepsilon_n(c)/5 \tag{B.10}$$

$$\mathcal{L}_n(\widehat{\theta}_n, A_{0,i}, \widehat{\mathbf{A}}_{0,-i}, \widehat{\mathbf{B}}_n) > \mathbb{E}\mathcal{L}_n(\widehat{\theta}_n, A_{0,i}, \widehat{\mathbf{A}}_{0,-i}, \widehat{\mathbf{B}}_n) - \varepsilon_n(c)/5 \tag{B.11}$$

w.p.a.1. Then, combining the inequalities (B.7)–(B.11) yields

$$
\begin{aligned}
\mathbb{E}\mathcal{L}_n(\theta_0, \widehat{A}_{n,i}, \mathbf{A}_{0,-i}, \mathbf{B}_0) &> \mathcal{L}_n(\theta_0, \widehat{A}_{n,i}, \mathbf{A}_{0,-i}, \mathbf{B}_0) - \varepsilon_n(c)/5 \\
&> \mathcal{L}_n(\widehat{\theta}_n, \widehat{A}_{n,i}, \widehat{\mathbf{A}}_{n,-i}, \widehat{\mathbf{B}}_n) - 2\varepsilon_n(c)/5 \\
&> \mathcal{L}_n(\widehat{\theta}_n, A_{0,i}, \widehat{\mathbf{A}}_{n,-i}, \widehat{\mathbf{B}}_n) - 3\varepsilon_n(c)/5 \\
&> \mathbb{E}\mathcal{L}_n(\widehat{\theta}_n, A_{0,i}, \widehat{\mathbf{A}}_{n,-i}, \widehat{\mathbf{B}}_n) - 4\varepsilon_n(c)/5 \\
&> \mathbb{E}\mathcal{L}_n(\theta_0, A_{0,i}, \mathbf{A}_{0,-i}, \mathbf{B}_0) - \varepsilon_n(c) \\
&= \max_{A_i \in \mathbb{A}_i^c} \mathbb{E}\mathcal{L}_n(\theta_0, A_i, \mathbf{A}_{0,-i}, \mathbf{B}_0) \\
&\quad + \underbrace{\left[\mathbb{E}\mathcal{L}_n(\theta_0, A_{0,i}, \mathbf{A}_{0,-i}, \mathbf{B}_0) - \max_{A_i \in \mathbb{A}_i^c} \mathbb{E}\mathcal{L}_n(\theta_0, A_i, \mathbf{A}_{0,-i}, \mathbf{B}_0)\right] - \varepsilon_n(c)}_{\geq 0} \\
&\geq \max_{A_i \in \mathbb{A}_i^c} \mathbb{E}\mathcal{L}_n(\theta_0, A_i, \mathbf{A}_{0,-i}, \mathbf{B}_0)
\end{aligned}
$$

w.p.a.1 for all i. The last line implies that $\widehat{A}_{n,i} \notin \mathbb{A}_i^c$. As the choice of c is arbitrary, this further implies that $\max_{1\leq i\leq n} |\widehat{A}_{n,i} - A_{0,i}| \xrightarrow{p} 0$. Analogously, we can also show that $\max_{1\leq i\leq n} |\widehat{B}_{n,i} - B_{0,i}| \xrightarrow{p} 0$. □

Lemma B.1. For any $(\theta, \boldsymbol{\gamma}) \in \Theta \times \mathbb{C}_n$ such that $||\theta - \theta_0|| = o(1)$ and $||\boldsymbol{\gamma} - \boldsymbol{\gamma}_0||_\infty = o(1)$,

(i) $\max_{1\leq k\leq n} \left|\frac{1}{n-1}\sum_{j\neq k}(h_{11,k,j}(\theta, \boldsymbol{\gamma}) - \mathbb{E}h_{11,k,j}(\theta_0, \boldsymbol{\gamma}_0))\right| = o_P(1)$,

(ii) $\max_{1\leq k\leq n} \left|\frac{1}{n-1}\sum_{j\neq k}(h_{11,j,k}(\theta, \boldsymbol{\gamma}) - \mathbb{E}h_{11,j,k}(\theta_0, \boldsymbol{\gamma}_0))\right| = o_P(1)$.

Proof. We only prove (i) since (ii) is completely analogous. By the triangle inequality,

$$
\begin{aligned}
\left|\frac{1}{n-1}\sum_{j\neq k}(h_{11,k,j}(\theta, \boldsymbol{\gamma}) - \mathbb{E}h_{11,k,j}(\theta_0, \boldsymbol{\gamma}_0))\right| &\leq \left|\frac{1}{n-1}\sum_{j\neq k}(h_{11,k,j}(\theta, \boldsymbol{\gamma}) - h_{11,k,j}(\theta_0, \boldsymbol{\gamma}_0))\right| \\
&\quad + \left|\frac{1}{n-1}\sum_{j\neq k}(h_{11,k,j}(\theta_0, \boldsymbol{\gamma}_0) - \mathbb{E}h_{11,k,j}(\theta_0, \boldsymbol{\gamma}_0))\right|.
\end{aligned}
\tag{B.12}
$$

With Assumption 2.1(iii), the mean value expansion gives

$$
\begin{aligned}
h_{11,k,j}(\theta, \boldsymbol{\gamma}) - h_{11,k,j}(\theta_0, \boldsymbol{\gamma}_0) &= h_{11,k,j}(\theta, \boldsymbol{\gamma}) - h_{11,k,j}(\theta_0, \boldsymbol{\gamma}) + h_{11,k,j}(\theta_0, \boldsymbol{\gamma}) - h_{11,k,j}(\theta_0, \boldsymbol{\gamma}_0) \\
&= \partial_{\theta^\top} h_{11,k,j}(\bar{\theta}, \boldsymbol{\gamma})[\theta - \theta_0] + \partial_{\mathbf{A}_{-1}^\top} h_{11,k,j}(\theta, \bar{\boldsymbol{\gamma}})[\mathbf{A}_{-1} - \mathbf{A}_{0,-1}] + \partial_{\mathbf{B}^\top} h_{11,k,j}(\theta, \bar{\boldsymbol{\gamma}})[\mathbf{B} - \mathbf{B}_0],
\end{aligned}
$$

where $\bar{\theta} \in [\theta, \theta_0]$, and $\bar{\boldsymbol{\gamma}} \in [\boldsymbol{\gamma}, \boldsymbol{\gamma}_0]$. Further, letting $h_{111,k,j}(\theta, \boldsymbol{\gamma}) \equiv \partial_{\pi_{k,j}} h_{11,k,j}(\theta, \boldsymbol{\gamma})$ and $h_{112,k,j}(\theta, \mathbf{A}) \equiv \partial_{\pi_{j,k}} h_{11,k,j}(\theta, \boldsymbol{\gamma})$, we have

$$
\begin{aligned}
\partial_{\mathbf{A}_{-1}^\top} h_{11,k,j}(\theta, \bar{\boldsymbol{\gamma}})[\mathbf{A}_{-1} - \mathbf{A}_{0,-1}] &= h_{111,k,j}(\theta, \bar{\boldsymbol{\gamma}})\chi_{n,k,-1}^\top[\mathbf{A}_{-1} - \mathbf{A}_{0,-1}] + h_{112,k,j}(\theta, \bar{\gamma})\chi_{n,j,-1}^\top[\mathbf{A}_{-1} - \mathbf{A}_{0,-1}] \\
\partial_{\mathbf{B}^\top} h_{11,k,j}(\theta, \bar{\boldsymbol{\gamma}})[\mathbf{B} - \mathbf{B}_0] &= h_{111,k,j}(\theta, \bar{\boldsymbol{\gamma}})\chi_{n,j}^\top[\mathbf{B} - \mathbf{B}_0] + h_{112,k,j}(\theta, \bar{\gamma})\chi_{n,k}^\top[\mathbf{B} - \mathbf{B}_0].
\end{aligned}
$$

Then, for some large constant $c > 0$,

$$
\begin{aligned}
&\left|\frac{1}{n-1}\sum_{j\neq k}\left(h_{11,k,j}(\theta,\boldsymbol{\gamma}) - h_{11,k,j}(\theta_0,\boldsymbol{\gamma}_0)\right)\right| \\
&\leq \left|\frac{1}{n-1}\sum_{j\neq k}\partial_{\theta^\top} h_{11,k,j}(\bar{\theta},\boldsymbol{\gamma})[\theta-\theta_0]\right| \\
&\quad + \left|\frac{1}{n-1}\sum_{j\neq k} h_{111,k,j}(\theta,\bar{\gamma})\chi_{n,k,-1}^\top[\mathbf{A}_{-1}-\mathbf{A}_{0,-1}]\right| + \left|\frac{1}{n-1}\sum_{j\neq k} h_{112,k,j}(\theta,\bar{\gamma})\chi_{n,j,-1}^\top[\mathbf{A}_{-1}-\mathbf{A}_{0,-1}]\right| \\
&\quad + \left|\frac{1}{n-1}\sum_{j\neq k} h_{111,k,j}(\theta,\bar{\gamma})\chi_{n,j}^\top[\mathbf{B}-\mathbf{B}_0]\right| + \left|\frac{1}{n-1}\sum_{j\neq k} h_{112,k,j}(\theta,\bar{\gamma})\chi_{n,k}^\top[\mathbf{B}-\mathbf{B}_0]\right| \\
&\leq c\left\|\theta-\theta_0\right\| + 2c\left\|\mathbf{A}-\mathbf{A}_0\right\|_\infty + 2c\left\|\mathbf{B}-\mathbf{B}_0\right\|_\infty.
\end{aligned}
$$

As the right-hand side term in the last line is independent of k, we have $\left|\frac{1}{n-1}\sum_{j\neq k}\left(h_{11,k,j}(\theta,\boldsymbol{\gamma}) - h_{11,k,j}(\theta_0,\boldsymbol{\gamma}_0)\right)\right| = o(1)$ for all k for any (θ,γ) such that $||\theta-\theta_0|| = o(1)$ and $||\boldsymbol{\gamma}-\boldsymbol{\gamma}_0||_\infty = o(1)$.

For the second term of (B.12), note that the random components involved in $h_{11,k,j}(\theta_0,\boldsymbol{\gamma}_0)$ are only $(y_{k,j}, y_{j,k})$ and, thus, that $\{h_{11,k,j}(\theta_0,\boldsymbol{\gamma}_0)\}_{j\neq k}$ are independent by Assumption 2.1(ii). Further, as $h_{11,k,j}(\theta_0,\boldsymbol{\gamma}_0)$ is uniformly bounded, using Hoeffding's and Boole's inequalities similarly as above, we can show that

$$
\max_{1\leq k\leq n}\left|\frac{1}{n-1}\sum_{j\neq k}\left(h_{11,k,j}(\theta_0,\boldsymbol{\gamma}_0) - \mathbb{E}h_{11,k,j}(\theta_0,\boldsymbol{\gamma}_0)\right)\right| = O_P\left(\sqrt{\frac{\ln n}{n}}\right).
$$

This completes the proof. □

Lemma B.2. (i) $\|\widetilde{\gamma}_n(\theta) - \widetilde{\gamma}_0(\theta_0)\|_\infty = o_P(1)$ for any $\theta\in\Theta$ such that $||\theta-\theta_0|| = o(1)$,

(ii) $\frac{1}{n}\sum_{i=1}^n\left|\widetilde{A}_{n,i}(\theta_0) - \widetilde{A}_{0,i}(\theta_0)\right| = O_P(n^{-1/2})$,

(iii) $\frac{1}{n}\sum_{i=1}^n\left|\widetilde{B}_{n,i}(\theta_0) - \widetilde{B}_{0,i}(\theta_0)\right| = O_P(n^{-1/2})$,

(iv) $\|\widetilde{\gamma}_n(\theta_0) - \widetilde{\gamma}_0(\theta_0)\|_\infty = O_P(\sqrt{\ln n/n})$.

Proof. (i) By the triangle inequality,

$$
\|\widetilde{\gamma}_n(\theta) - \widetilde{\gamma}_0(\theta_0)\|_\infty \leq \|\widetilde{\gamma}_n(\theta) - \widetilde{\gamma}_0(\theta)\|_\infty + \|\widetilde{\gamma}_0(\theta) - \widetilde{\gamma}_0(\theta_0)\|_\infty.
$$

For the first term on the right-hand side, the same argument as in the proof of Theorem 3.1(iv) achieves $||\widetilde{\gamma}_n(\theta) - \widetilde{\gamma}_0(\theta)||_\infty = o_P(1)$ for any θ in the neighborhood of θ_0 under Assumption 3.2. For the second term, Assumption 3.2 and Berge's theorem implies that every element of $\widetilde{\gamma}_0(\theta)$ is continuous in the neighborhood of θ_0 (see, e.g., Corollary A4.8, Kreps, 2012). Thus, $||\widetilde{\gamma}_0(\theta) - \widetilde{\gamma}_0(\theta_0)||_\infty = o(1)$ holds.

(ii) (iii) By the first-order condition and the mean value expansion,

$$\begin{aligned}\mathbf{0}_{(2n-1)\times 1} &= n\cdot\mathcal{S}_{n,\boldsymbol{\gamma}}(\theta_0,\widetilde{\boldsymbol{\gamma}}_n(\theta_0)) \\ &= n\cdot\mathcal{S}_{n,\boldsymbol{\gamma}} - (-n\cdot\mathcal{H}_{n,\boldsymbol{\gamma}\boldsymbol{\gamma}}(\theta_0,\bar{\boldsymbol{\gamma}}_n))\,[\widetilde{\boldsymbol{\gamma}}_{n,-1}(\theta_0)-\widetilde{\boldsymbol{\gamma}}_{0,-1}(\theta_0)],\end{aligned}$$

where $\bar{\boldsymbol{\gamma}}_n \in [\widetilde{\boldsymbol{\gamma}}_n(\theta_0),\widetilde{\boldsymbol{\gamma}}_0(\theta_0)]$. Then, by result (i) and Assumption 3.3(i), we have

$$\widetilde{\boldsymbol{\gamma}}_{n,-1}(\theta_0)-\widetilde{\boldsymbol{\gamma}}_{0,-1}(\theta_0) = (-n\cdot\mathcal{H}_{n,\boldsymbol{\gamma}\boldsymbol{\gamma}}(\theta_0,\bar{\boldsymbol{\gamma}}_n))^{-1}\,n\cdot\mathcal{S}_{n,\boldsymbol{\gamma}} \tag{B.13}$$

w.p.a.1; thus, $||\widetilde{\boldsymbol{\gamma}}_n(\theta_0)-\widetilde{\boldsymbol{\gamma}}_0(\theta_0)|| = ||\widetilde{\boldsymbol{\gamma}}_{n,-1}(\theta_0)-\widetilde{\boldsymbol{\gamma}}_{0,-1}(\theta_0)|| \le O_P(1)\cdot||n\cdot\mathcal{S}_{n,\boldsymbol{\gamma}}||$.

Further, observe that

$$\begin{aligned}\mathbb{E}||n\cdot\mathcal{S}_{n,\boldsymbol{\gamma}}||^2 &= \frac{4}{(n-1)^2}\sum_{i=1}^{n}\sum_{j>i}\sum_{k=1}^{n}\sum_{l>k}\mathbb{E}\left[s_{i,j}^{\mathbf{A}}(\theta_0,\boldsymbol{\gamma}_0)^\top s_{k,l}^{\mathbf{A}}(\theta_0,\boldsymbol{\gamma}_0)+s_{i,j}^{\mathbf{B}}(\theta_0,\boldsymbol{\gamma}_0)^\top s_{k,l}^{\mathbf{B}}(\theta_0,\boldsymbol{\gamma}_0)\right] \\ &= \frac{4}{(n-1)^2}\sum_{i=1}^{n}\sum_{j>i}\sum_{k=1}^{n}\sum_{l>k}\mathbb{E}\left[(s_{1,i,j}\chi_{n,i,-1}^\top+s_{2,i,j}\chi_{n,j,-1}^\top)(s_{1,k,l}\chi_{n,k,-1}+s_{2,k,l}\chi_{n,l,-1})\right] \\ &\quad+\frac{4}{(n-1)^2}\sum_{i=1}^{n}\sum_{j>i}\sum_{k=1}^{n}\sum_{l>k}\mathbb{E}\left[(s_{1,i,j}\chi_{n,j}^\top+s_{2,i,j}\chi_{n,i}^\top)(s_{1,k,l}\chi_{n,l}+s_{2,k,l}\chi_{n,k})\right] \\ &= \frac{4}{(n-1)^2}\sum_{i=1}^{n}\sum_{j>i}\mathbb{E}\left[(s_{1,i,j}\chi_{n,i,-1}^\top+s_{2,i,j}\chi_{n,j,-1}^\top)(s_{1,i,j}\chi_{n,i,-1}+s_{2,i,j}\chi_{n,j,-1})\right] \\ &\quad+\frac{4}{(n-1)^2}\sum_{i=1}^{n}\sum_{j>i}\mathbb{E}\left[(s_{1,i,j}\chi_{n,i,-1}^\top+s_{2,i,j}\chi_{n,j,-1}^\top)(s_{1,j,i}\chi_{n,j,-1}+s_{2,j,i}\chi_{n,i,-1})\right] \\ &\quad+\frac{4}{(n-1)^2}\sum_{i=1}^{n}\sum_{j>i}\mathbb{E}\left[(s_{1,i,j}\chi_{n,j}^\top+s_{2,i,j}\chi_{n,i}^\top)(s_{1,i,j}\chi_{n,j}+s_{2,i,j}\chi_{n,i})\right] \\ &\quad+\frac{4}{(n-1)^2}\sum_{i=1}^{n}\sum_{j>i}\mathbb{E}\left[(s_{1,i,j}\chi_{n,j}^\top+s_{2,i,j}\chi_{n,i}^\top)(s_{1,j,i}\chi_{n,i}+s_{2,j,i}\chi_{n,j})\right] = O(1)\end{aligned}$$

by Assumption 2.1(ii). Then, it holds that $||n\cdot\mathcal{S}_{n,\boldsymbol{\gamma}}|| = O_P(1)$ by Markov's inequality; thus, we have $||\widetilde{\boldsymbol{\gamma}}_n(\theta_0)-\widetilde{\boldsymbol{\gamma}}_0(\theta_0)|| = O_P(1)$. Finally, by the basic norm inequality, it holds that

$$\sum_{i=1}^{n}|\widetilde{A}_{n,i}(\theta_0)-\widetilde{A}_{0,i}(\theta_0)|+\sum_{i=1}^{n}|\widetilde{B}_{n,i}(\theta_0)-\widetilde{B}_{0,i}(\theta_0)| \le \sqrt{2n-1}||\widetilde{\boldsymbol{\gamma}}_n(\theta_0)-\widetilde{\boldsymbol{\gamma}}_0(\theta_0)|| = O_P(\sqrt{n}),$$

which gives the desired result.

(iv) Let $\boldsymbol{\gamma}_{-k} = (\mathbf{A}_{-k}^\top,\mathbf{B}^\top)^\top$ for $k\neq 1$ and write $\mathcal{L}_n(\theta,A_k,\boldsymbol{\gamma}_{-k})$ as $\mathcal{L}_n(\theta,\boldsymbol{\gamma})$. By the first-order condition and mean value expansion,

$$\begin{aligned}0 &= n\cdot\partial_{A_k}\mathcal{L}_n(\theta_0,\widetilde{A}_{n,k}(\theta_0),\widetilde{\boldsymbol{\gamma}}_{n,-k}(\theta_0)) \\ &= \frac{2}{n-1}\sum_{i=1}^{n}\sum_{j>i}s_{i,j}^{A_k}+n\cdot\partial_{A_k}\mathcal{L}_n(\theta_0,\widetilde{A}_{n,k}(\theta_0),\widetilde{\boldsymbol{\gamma}}_{n,-k}(\theta_0))-n\cdot\partial_{A_k}\mathcal{L}_n(\theta_0,\widetilde{A}_{0,k}(\theta_0),\widetilde{\boldsymbol{\gamma}}_{n,-k}(\theta_0)) \\ &\quad+n\cdot\partial_{A_k}\mathcal{L}_n(\theta_0,A_{0,k},\widetilde{\boldsymbol{\gamma}}_{n,-k}(\theta_0))-n\cdot\partial_{A_k}\mathcal{L}_n(\theta_0,A_{0,k},\widetilde{\boldsymbol{\gamma}}_{0,-k}(\theta_0))\end{aligned}$$

$$= \frac{2}{n-1}\sum_{i=1}^{n}\sum_{j>i} s_{i,j}^{A_k} + \frac{2}{n-1}\sum_{j\neq k} h_{11,k,j}(\theta_0, \bar{A}_{n,k}, \widetilde{\boldsymbol{\gamma}}_{n,-k}(\theta_0))[\widetilde{A}_{n,k}(\theta_0) - \widetilde{A}_{0,k}(\theta_0)]$$

$$+ \frac{2}{n-1}\sum_{i=1}^{n}\sum_{j>i} \partial_{\boldsymbol{\gamma}_{-k}^{\top}} s_{i,j}^{A_k}(\theta_0, A_{0,k}, \bar{\boldsymbol{\gamma}}_{n,-k})[\widetilde{\boldsymbol{\gamma}}_{n,-k}(\theta_0) - \widetilde{\boldsymbol{\gamma}}_{0,-k}(\theta_0)],$$

where $\bar{A}_{n,k} \in [\widetilde{A}_{n,k}(\theta_0), \widetilde{A}_{0,k}(\theta_0)]$, and $\bar{\boldsymbol{\gamma}}_{n,-k} \in [\widetilde{\boldsymbol{\gamma}}_{n,-k}(\theta_0), \widetilde{\boldsymbol{\gamma}}_{0,-k}(\theta_0)]$. In view of (A.1), Lemma B.1(i), and result (i) imply that $\frac{2}{n-1}\sum_{j\neq k} h_{11,k,j}(\theta_0, \bar{A}_{n,k}, \widetilde{\boldsymbol{\gamma}}_{n,-k}(\theta_0))$ is bounded and away from zero w.p.a.1 uniformly in k. Then,

$$\left|\widetilde{A}_{n,k}(\theta_0) - \widetilde{A}_{0,k}(\theta_0)\right| \le (c + o_p(1))\left\{|T_{1,n,k}| + |T_{2,n,k}|\right\}$$

for some $c > 0$, where

$$T_{1,n,k} \equiv \frac{2}{n-1}\sum_{i=1}^{n}\sum_{j>i} s_{i,j}^{A_k},$$

$$T_{2,n,k} \equiv \frac{2}{n-1}\sum_{i=1}^{n}\sum_{j>i} \partial_{\boldsymbol{\gamma}_{-k}^{\top}} s_{i,j}^{A_k}(\theta_0, A_{0,k}, \bar{\boldsymbol{\gamma}}_{n,-k})[\widetilde{\boldsymbol{\gamma}}_{n,-k}(\theta_0) - \widetilde{\boldsymbol{\gamma}}_{0,-k}(\theta_0)]$$

$$= \frac{2}{n-1}\sum_{i=1}^{n}\sum_{j>i} \partial_{\mathbf{A}_{-k}^{\top}} s_{i,j}^{A_k}(\theta_0, A_{0,k}, \bar{\boldsymbol{\gamma}}_{n,-k})[\widetilde{\mathbf{A}}_{n,-k}(\theta_0) - \widetilde{\mathbf{A}}_{0,-k}(\theta_0)]$$

$$+ \frac{2}{n-1}\sum_{i=1}^{n}\sum_{j>i} \partial_{\mathbf{B}^{\top}} s_{i,j}^{A_k}(\theta_0, A_{0,k}, \bar{\boldsymbol{\gamma}}_{n,-k})[\widetilde{\mathbf{B}}_{n}(\theta_0) - \widetilde{\mathbf{B}}_{0}(\theta_0)]$$

$$\equiv T_{21,n,k} + T_{22,n,k}, \quad \text{say.}$$

First, observe that

$$T_{1,n,k} = \frac{2}{n-1}\sum_{i=1}^{n}\sum_{j>i} s_{1,i,j}\mathbf{1}\{i=k\} + \frac{2}{n-1}\sum_{i=1}^{n}\sum_{j>i} s_{2,i,j}\mathbf{1}\{j=k\}$$

$$= \frac{2}{n-1}\sum_{j>k} s_{1,k,j} + \frac{2}{n-1}\sum_{j<k} s_{2,j,k}$$

$$= \frac{2}{n-1}\sum_{j\neq k} s_{1,k,j},$$

where the last equality holds because $s_{1,k,j} = s_{2,j,k}$. Clearly, $s_{1,k,j}$ is uniformly bounded and $\mathbb{E}s_{1,k,j} = 0$. Then, with Assumption 2.1(ii), we can show that

$$\max_{1\le k\le n} |T_{1,n,k}| = O_P(\sqrt{\ln n/n}) \tag{B.14}$$

by Hoeffding's and Boole's inequalities.

Next, observe that there exist bounded functions $s_{11,i,j}(\theta,\boldsymbol{\gamma}) \equiv \partial_{\pi_{i,j}} s_{1,i,j}(\theta,\boldsymbol{\gamma})$, $s_{12,i,j}(\theta,\boldsymbol{\gamma}) \equiv \partial_{\pi_{j,i}} s_{1,i,j}(\theta,\boldsymbol{\gamma})$, $s_{21,i,j}(\theta,\boldsymbol{\gamma}) \equiv \partial_{\pi_{i,j}} s_{2,i,j}(\theta,\boldsymbol{\gamma})$, and $s_{22,i,j}(\theta,\boldsymbol{\gamma}) \equiv \partial_{\pi_{j,i}} s_{2,i,j}(\theta,\boldsymbol{\gamma})$, satisfying

$$\sum_{i=1}^{n}\sum_{j>i} \partial_{\mathbf{A}_{-k}^{\top}} s_{i,j}^{A_k}(\theta,\boldsymbol{\gamma}) = \sum_{i=1}^{n}\sum_{j>i} s_{11,i,j}(\theta,\boldsymbol{\gamma})\mathbf{1}\{i=k\}\chi_{n,i,-k}^{\top} + \sum_{i=1}^{n}\sum_{j>i} s_{12,i,j}(\theta,\boldsymbol{\gamma})\mathbf{1}\{i=k\}\chi_{n,j,-k}^{\top}$$

$$
\begin{aligned}
&+\sum_{i=1}^{n}\sum_{j>i} s_{21,i,j}(\theta,\boldsymbol{\gamma})\mathbf{1}\{j=k\}\chi_{n,i,-k}^{\top}+\sum_{i=1}^{n}\sum_{j>i} s_{22,i,j}(\theta,\boldsymbol{\gamma})\mathbf{1}\{j=k\}\chi_{n,j,-k}^{\top}\\
&=\sum_{j>k} s_{12,k,j}(\theta,\boldsymbol{\gamma})\chi_{n,j,-k}^{\top}+\sum_{j<k} s_{21,j,k}(\theta,\boldsymbol{\gamma})\chi_{n,j,-k}^{\top}\\
&=\sum_{j\neq k} s_{12,k,j}(\theta,\boldsymbol{\gamma})\chi_{n,j,-k}^{\top},
\end{aligned}
$$

where the second equality follows since $\mathbf{1}\{i=k\}\chi_{n,i,-k}$ is only a vector of zeros. Hence, for some $c>0$, the triangle inequality gives

$$
\begin{aligned}
|T_{21,n,k}| &= \left|\frac{2}{n-1}\sum_{j\neq k} s_{12,k,j}(\theta_0,A_{0,k},\bar{\boldsymbol{\gamma}}_{n,-k})\chi_{n,j,-k}^{\top}[\widetilde{\mathbf{A}}_{n,-k}(\theta_0)-\widetilde{\mathbf{A}}_{0,-k}(\theta_0)]\right|\\
&\le \frac{c}{n-1}\sum_{j\neq k}\left|\chi_{n,j,-k}^{\top}[\widetilde{\mathbf{A}}_{n,-k}(\theta_0)-\widetilde{\mathbf{A}}_{0,-k}(\theta_0)]\right|\\
&= \frac{c}{n-1}\sum_{j\neq k}\left|\widetilde{A}_{n,j}(\theta_0)-\widetilde{A}_{0,j}(\theta_0)\right| \le \frac{c}{n-1}\sum_{j=1}^{n}\left|\widetilde{A}_{n,j}(\theta_0)-\widetilde{A}_{0,j}(\theta_0)\right| = O_P(n^{-1/2})
\end{aligned}
$$

by result (ii). Note that the last inequality is independent of k. Analogously, we can show that $|T_{22,n,k}|=O_P(n^{-1/2})$ uniformly in k by result (iii). Thus, we have

$$
\max_{1\le k\le n}|T_{2,n,k}| = O_P(n^{-1/2}). \tag{B.15}
$$

Combining (B.14) and (B.15) yields $\max_{1\le k\le n}|\widetilde{A}_{n,k}(\theta_0)-\widetilde{A}_{0,k}(\theta_0)| = O_P(\sqrt{\ln n/n})+O_P(n^{-1/2})$.

Similarly as above, using Lemma B.1(ii), we can also show that $\max_{1\le k\le n}|\widetilde{B}_{n,k}(\theta_0)-\widetilde{B}_{0,k}(\theta_0)| = O_P(\sqrt{\ln n/n})+O_P(n^{-1/2})$. This completes the proof. □

Lemma B.3. $||\widehat{\theta}_n-\theta_0|| = O_P(n^{-1/2})$.

Proof. Applying the implicit function theorem to $\mathcal{S}_{n,\boldsymbol{\gamma}}(\theta,\widetilde{\boldsymbol{\gamma}}_n(\theta)) = \partial_{\boldsymbol{\gamma}_{-1}}\mathcal{L}_n(\theta,\widetilde{\boldsymbol{\gamma}}_n(\theta)) = \mathbf{0}_{(2n-1)\times 1}$ for $\theta\in\Theta$ yields

$$
\begin{aligned}
\mathbf{0}_{(2n-1)\times(d_z+2)} &= \partial_{\theta^{\top}}\mathcal{S}_{n,\boldsymbol{\gamma}}(\theta,\widetilde{\boldsymbol{\gamma}}_n(\theta))\\
&= \mathcal{H}_{n,\boldsymbol{\gamma}\theta}(\theta,\widetilde{\boldsymbol{\gamma}}_n(\theta)) + \mathcal{H}_{n,\boldsymbol{\gamma}\boldsymbol{\gamma}}(\theta,\widetilde{\boldsymbol{\gamma}}_n(\theta))\partial_{\theta^{\top}}\widetilde{\boldsymbol{\gamma}}_n(\theta)\\
\Longrightarrow \partial_{\theta^{\top}}\widetilde{\boldsymbol{\gamma}}_n(\theta) &= -\left[n\cdot\mathcal{H}_{n,\boldsymbol{\gamma}\boldsymbol{\gamma}}(\theta,\widetilde{\boldsymbol{\gamma}}_n(\theta))\right]^{-1} n\cdot\mathcal{H}_{n,\boldsymbol{\gamma}\theta}(\theta,\widetilde{\boldsymbol{\gamma}}_n(\theta)),
\end{aligned}
$$

where the right-hand side exists w.p.a.1 for θ in a neighborhood of θ_0 by Lemma B.2(i) and Assumption 3.3(i). Then, by the second-order Taylor expansion,

$$
\begin{aligned}
0 \le \mathcal{L}_n(\widehat{\theta}_n,\widetilde{\boldsymbol{\gamma}}_n(\widehat{\theta}_n)) - \mathcal{L}_n(\theta_0,\widetilde{\boldsymbol{\gamma}}_n(\theta_0)) &= \mathcal{S}_{n,\theta}(\theta_0,\widetilde{\boldsymbol{\gamma}}_n(\theta_0))^{\top}(\widehat{\theta}_n-\theta_0)\\
&\quad + \frac{1}{2}(\widehat{\theta}_n-\theta_0)^{\top}\left[\mathcal{H}_{n,\theta\theta}(\bar{\theta}_n,\widetilde{\boldsymbol{\gamma}}_n(\bar{\theta}_n)) + \mathcal{H}_{n,\theta\boldsymbol{\gamma}}(\bar{\theta}_n,\widetilde{\boldsymbol{\gamma}}_n(\bar{\theta}_n))\{\partial_{\theta^{\top}}\widetilde{\boldsymbol{\gamma}}_n(\bar{\theta}_n)\}\right](\widehat{\theta}_n-\theta_0)\\
&\le \|\mathcal{S}_{n,\theta}(\theta_0,\widetilde{\boldsymbol{\gamma}}_n(\theta_0))\|\cdot\left\|\widehat{\theta}_n-\theta_0\right\| - \frac{1}{2}\lambda_{\min}\left(-\mathcal{I}_{n,\theta\theta}(\bar{\theta}_n,\widetilde{\boldsymbol{\gamma}}_n(\bar{\theta}_n))\right)\cdot\left\|\widehat{\theta}_n-\theta_0\right\|^2\\
&\Longrightarrow \left\|\widehat{\theta}_n-\theta_0\right\| \le \frac{2\left\|\mathcal{S}_{n,\theta}(\theta_0,\widetilde{\boldsymbol{\gamma}}_n(\theta_0))\right\|}{\lambda_{\min}\left(-\mathcal{I}_{n,\theta\theta}(\bar{\theta}_n,\widetilde{\boldsymbol{\gamma}}_n(\bar{\theta}_n))\right)},
\end{aligned}
$$

where $\bar{\theta}_n \in [\widehat{\theta}_n, \theta_0]$. Here, since Theorem 3.1(i) and Lemma B.2(i) imply that $\widetilde{\gamma}_n(\bar{\theta}_n)$ is uniformly consistent for γ_0, $\lambda_{\min}\left(-\mathcal{I}_{n,\theta\theta}(\bar{\theta}_n, \widetilde{\gamma}_n(\bar{\theta}_n))\right) > c_\theta$ w.p.a.1 by Assumption 3.3(ii). Thus, it suffices to show that $||\mathcal{S}_{n,\theta}(\theta_0, \widetilde{\gamma}_n(\theta_0))|| = O_P(n^{-1/2})$.

We decompose $\mathcal{S}_{n,\theta}(\theta_0, \widetilde{\gamma}_n(\theta_0))$ into the following two terms:

$$\begin{aligned}\mathcal{S}_{n,\theta}(\theta_0, \widetilde{\gamma}_n(\theta_0)) &= \frac{2}{N}\sum_{i=1}^{n}\sum_{j>i} s^{\theta}_{i,j}(\theta_0, \gamma_0) + \frac{2}{N}\sum_{i=1}^{n}\sum_{j>i}[s^{\theta}_{i,j}(\theta_0, \widetilde{\gamma}_n(\theta_0)) - s^{\theta}_{i,j}(\theta_0, \widetilde{\gamma}_0(\theta_0))] \\ &\equiv s^{\theta}_{1,n} + s^{\theta}_{2,n}, \quad \text{say.}\end{aligned}$$

Since $\mathbb{E}s^{\theta}_{1,n} = \mathbf{0}_{(d_z+2)\times 1}$, by Assumption 2.1(ii),

$$\begin{aligned}\text{Var}\left[s^{\theta}_{1,n}\right] &= \frac{4}{N^2}\sum_{i=1}^{n}\sum_{j>i}\sum_{k=1}^{n}\sum_{l>k}\mathbb{E}[s^{\theta}_{i,j}s^{\theta\top}_{k,l}] \\ &= \frac{4}{N^2}\sum_{i=1}^{n}\sum_{j>i}\mathbb{E}[s^{\theta}_{i,j}s^{\theta\top}_{i,j} + s^{\theta}_{i,j}s^{\theta\top}_{j,i}] = O(n^{-2}).\end{aligned}$$

Thus, by Chebyshev's inequality, $||s^{\theta}_{1,n}|| = O_P(n^{-1})$. For $s^{\theta}_{2,n}$, observe that there exist bounded functions $s^{\theta}_{1,i,j}(\theta, \gamma) \equiv \partial_{\pi_{i,j}} s^{\theta}_{i,j}(\theta, \gamma)$ and $s^{\theta}_{2,i,j}(\theta, \gamma) \equiv \partial_{\pi_{j,i}} s^{\theta}_{i,j}(\theta, \gamma)$, such that

$$\begin{aligned}&s^{\theta}_{i,j}(\theta_0, \widetilde{\gamma}_n(\theta_0)) - s^{\theta}_{i,j}(\theta_0, \widetilde{\gamma}_0(\theta_0)) \\ &= \partial_{\mathbf{A}^{\top}_{-1}} s^{\theta}_{i,j}(\theta_0, \bar{\gamma}_n)[\widetilde{\mathbf{A}}_{n,-1}(\theta_0) - \widetilde{\mathbf{A}}_{0,-1}(\theta_0)] + \partial_{\mathbf{B}^{\top}} s^{\theta}_{i,j}(\theta_0, \bar{\gamma}_n)[\widetilde{\mathbf{B}}_n(\theta_0) - \widetilde{\mathbf{B}}_0(\theta_0)] \\ &= s^{\theta}_{1,i,j}(\theta_0, \bar{\gamma}_n)\chi^{\top}_{n,i,-1}[\widetilde{\mathbf{A}}_{n,-1}(\theta_0) - \widetilde{\mathbf{A}}_{0,-1}(\theta_0)] + s^{\theta}_{2,i,j}(\theta_0, \bar{\gamma}_n)\chi^{\top}_{n,j,-1}[\widetilde{\mathbf{A}}_{n,-1}(\theta_0) - \widetilde{\mathbf{A}}_{0,-1}(\theta_0)] \qquad \text{(B.16)} \\ &\quad + s^{\theta}_{1,i,j}(\theta_0, \bar{\gamma}_n)\chi^{\top}_{n,j}[\widetilde{\mathbf{B}}_n(\theta_0) - \widetilde{\mathbf{B}}_0(\theta_0)] + s^{\theta}_{2,i,j}(\theta_0, \bar{\gamma}_n)\chi^{\top}_{n,i}[\widetilde{\mathbf{B}}_n(\theta_0) - \widetilde{\mathbf{B}}_0(\theta_0)] \\ &\equiv t_{1,i,j} + t_{2,i,j} + t_{3,i,j} + t_{4,i,j}, \quad \text{say,}\end{aligned}$$

where $\bar{\gamma}_n \in [\widetilde{\gamma}_n(\theta_0), \widetilde{\gamma}_0(\theta_0)]$. For some $c > 0$,

$$\begin{aligned}\left\|\frac{2}{N}\sum_{i=1}^{n}\sum_{j>i} t_{1,i,j}\right\| &\leq \frac{c}{N}\sum_{i=1}^{n}\sum_{j>i}\left|\chi^{\top}_{n,i,-1}[\widetilde{\mathbf{A}}_{n,-1}(\theta_0) - \widetilde{\mathbf{A}}_{0,-1}(\theta_0)]\right| \\ &= \frac{c}{n}\sum_{i=2}^{n}\left|\widetilde{A}_{n,i}(\theta_0) - \widetilde{A}_{0,i}(\theta_0)\right| = O_P(n^{-1/2})\end{aligned}$$

by Lemma B.2(ii). Similarly, it is straightforward to see that $||\frac{2}{N}\sum_{i=1}^{n}\sum_{j>i} t_{2,i,j}|| = O_P(n^{-1/2})$ and that $||\frac{2}{N}\sum_{i=1}^{n}\sum_{j>i} t_{3,i,j}|| = O_P(n^{-1/2})$ and $||\frac{2}{N}\sum_{i=1}^{n}\sum_{j>i} t_{4,i,j}|| = O_P(n^{-1/2})$ by Lemma B.2(iii). Hence, we have $||s^{\theta}_{2,n}|| = O_P(n^{-1/2})$, and this completes the proof. □

Proof of Theorem 3.2

(i) (ii) By the first-order condition and the mean value expansion,

$$\mathbf{0}_{(2n-1)\times 1} = n \cdot \mathcal{S}_{n,\gamma}(\widehat{\theta}_n, \widehat{\gamma}_n)$$

$$= n \cdot \mathcal{S}_{n,\boldsymbol{\gamma}} + n \cdot \mathcal{S}_{n,\boldsymbol{\gamma}}(\widehat{\theta}_n, \widehat{\boldsymbol{\gamma}}_n) - n \cdot \mathcal{S}_{n,\boldsymbol{\gamma}}(\theta_0, \widehat{\boldsymbol{\gamma}}_n) + n \cdot \mathcal{S}_{n,\boldsymbol{\gamma}}(\theta_0, \widehat{\boldsymbol{\gamma}}_n) - n \cdot \mathcal{S}_{n,\boldsymbol{\gamma}}(\theta_0, \boldsymbol{\gamma}_0)$$
$$= n \cdot \mathcal{S}_{n,\boldsymbol{\gamma}} + n \cdot \partial_{\theta^\top} \mathcal{S}_{n,\boldsymbol{\gamma}}(\bar{\theta}_n, \widehat{\boldsymbol{\gamma}}_n)[\widehat{\theta}_n - \theta_0] - (-n \cdot \mathcal{H}_{n,\boldsymbol{\gamma}\boldsymbol{\gamma}}(\theta_0, \bar{\boldsymbol{\gamma}}_n)) [\widehat{\boldsymbol{\gamma}}_{n,-1} - \boldsymbol{\gamma}_{0,-1}],$$

where $\bar{\theta}_n \in [\widehat{\theta}_n, \theta_0]$, and $\bar{\boldsymbol{\gamma}}_n \in [\widehat{\boldsymbol{\gamma}}_n, \boldsymbol{\gamma}_0]$. Thus, under Assumption 3.3(i),

$$\widehat{\boldsymbol{\gamma}}_{n,-1} - \boldsymbol{\gamma}_{0,-1} = (-n \cdot \mathcal{H}_{n,\boldsymbol{\gamma}\boldsymbol{\gamma}}(\theta_0, \bar{\boldsymbol{\gamma}}_n))^{-1} n \cdot \mathcal{S}_{n,\boldsymbol{\gamma}} + (-n \cdot \mathcal{H}_{n,\boldsymbol{\gamma}\boldsymbol{\gamma}}(\theta_0, \bar{\boldsymbol{\gamma}}_n))^{-1} n \cdot \partial_{\theta^\top} \mathcal{S}_{n,\boldsymbol{\gamma}}(\bar{\theta}_n, \widehat{\boldsymbol{\gamma}}_n)[\widehat{\theta}_n - \theta_0].$$

As shown in the proof of Lemma B.2(ii)–(iii), $||n \cdot \mathcal{S}_{n,\boldsymbol{\gamma}}|| = O_P(1)$. For the second term, noting that $\partial_{\theta^\top} \mathcal{S}_{n,\boldsymbol{\gamma}}(\theta, \boldsymbol{\gamma}) = \frac{2}{N} \sum_{i=1}^n \sum_{j>i} [\partial_{\mathbf{A}_{-1}^\top} s_{i,j}^\theta(\theta, \boldsymbol{\gamma}),\ \partial_{\mathbf{B}^\top} s_{i,j}^\theta(\theta, \boldsymbol{\gamma})]^\top$. Hence,

$$\begin{aligned}
\|n \cdot \partial_{\theta^\top} \mathcal{S}_{n,\boldsymbol{\gamma}}(\theta, \boldsymbol{\gamma})\|^2 &= n^2 \cdot \text{tr} \left\{ [\partial_{\theta^\top} \mathcal{S}_{n,\boldsymbol{\gamma}}(\theta, \boldsymbol{\gamma})]^\top \partial_{\theta^\top} \mathcal{S}_{n,\boldsymbol{\gamma}}(\theta, \boldsymbol{\gamma}) \right\} \\
&= \text{tr} \left\{ \frac{4}{(n-1)^2} \sum_{i=1}^n \sum_{j>i} \sum_{k=1}^n \sum_{l>k} \left[\partial_{\mathbf{A}_{-1}^\top} s_{i,j}^\theta(\theta, \boldsymbol{\gamma}) \right] \left[\partial_{\mathbf{A}_{-1}^\top} s_{k,l}^\theta(\theta, \boldsymbol{\gamma}) \right]^\top \right\} \\
&\quad + \text{tr} \left\{ \frac{4}{(n-1)^2} \sum_{i=1}^n \sum_{j>i} \sum_{k=1}^n \sum_{l>k} \left[\partial_{\mathbf{B}^\top} s_{i,j}^\theta(\theta, \boldsymbol{\gamma}) \right] \left[\partial_{\mathbf{B}^\top} s_{k,l}^\theta(\theta, \boldsymbol{\gamma}) \right]^\top \right\} \\
&\equiv u_n^{\mathbf{A}}(\theta, \boldsymbol{\gamma}) + u_n^{\mathbf{B}}(\theta, \boldsymbol{\gamma}), \quad \text{say.}
\end{aligned}$$

Further,

$$\begin{aligned}
&u_n^{\mathbf{A}}(\theta, \boldsymbol{\gamma}) \\
&= \frac{4}{(n-1)^2} \sum_{i=1}^n \sum_{j>i} \sum_{k=1}^n \sum_{l>k} \text{tr} \left\{ \left[s_{1,i,j}^\theta(\theta, \boldsymbol{\gamma}) \chi_{n,i,-1}^\top + s_{2,i,j}^\theta(\theta, \boldsymbol{\gamma}) \chi_{n,j,-1}^\top \right] \left[\chi_{n,k,-1} s_{1,k,l}^\theta(\theta, \boldsymbol{\gamma})^\top + \chi_{n,l,-1} s_{2,k,l}^\theta(\theta, \boldsymbol{\gamma})^\top \right] \right\} \\
&= \frac{4}{(n-1)^2} \sum_{i=1}^n \sum_{j>i} \sum_{k=1}^n \sum_{l>k} \left[\chi_{n,i,-1}^\top \chi_{n,k,-1} \cdot s_{1,k,l}^\theta(\theta, \boldsymbol{\gamma})^\top s_{1,i,j}^\theta(\theta, \boldsymbol{\gamma}) + \chi_{n,j,-1}^\top \chi_{n,k,-1} \cdot s_{1,k,l}^\theta(\theta, \boldsymbol{\gamma})^\top s_{2,i,j}^\theta(\theta, \boldsymbol{\gamma}) \right. \\
&\qquad \left. + \chi_{n,i,-1}^\top \chi_{n,l,-1} \cdot s_{2,k,l}^\theta(\theta, \boldsymbol{\gamma})^\top s_{1,i,j}^\theta(\theta, \boldsymbol{\gamma}) + \chi_{n,j,-1}^\top \chi_{n,l,-1} \cdot s_{2,k,l}^\theta(\theta, \boldsymbol{\gamma})^\top s_{2,i,j}^\theta(\theta, \boldsymbol{\gamma}) \right].
\end{aligned}$$

Noting that $\chi_{n,i,-1}^\top \chi_{n,k,-1} = \mathbf{1}\{i = k > 1\}$, we have

$$\begin{aligned}
\frac{4}{(n-1)^2} \sum_{i=1}^n \sum_{j>i} \sum_{k=1}^n \sum_{l>k} \chi_{n,i,-1}^\top \chi_{n,k,-1} \cdot s_{1,k,l}^\theta(\theta, \boldsymbol{\gamma})^\top s_{1,i,j}^\theta(\theta, \boldsymbol{\gamma}) &= \frac{4}{(n-1)^2} \sum_{i=2}^n \sum_{j>i} \sum_{l>i} s_{1,i,l}^\theta(\theta, \boldsymbol{\gamma})^\top s_{1,i,j}^\theta(\theta, \boldsymbol{\gamma}) \\
&= O(n)
\end{aligned}$$

for any $(\theta, \boldsymbol{\gamma}) \in \Theta \times \mathbb{C}_n$. Applying the same discussion to the other terms, we obtain $u_n^{\mathbf{A}}(\theta, \boldsymbol{\gamma}) = O(n)$. By the same argument, we can easily show that $u_n^{\mathbf{B}}(\theta, \boldsymbol{\gamma}) = O(n)$ for any $(\theta, \boldsymbol{\gamma}) \in \Theta \times \mathbb{C}_n$. Then, combined with Lemma B.3, we obtain $||n \cdot \partial_{\theta^\top} \mathcal{S}_{n,\boldsymbol{\gamma}}(\bar{\theta}_n, \widehat{\boldsymbol{\gamma}}_n)[\widehat{\theta}_n - \theta_0]|| \leq ||n \cdot \partial_{\theta^\top} \mathcal{S}_{n,\boldsymbol{\gamma}}(\bar{\theta}_n, \widehat{\boldsymbol{\gamma}}_n)|| \cdot ||\widehat{\theta}_n - \theta_0|| = O_P(1)$.

From these results, under Assumption 3.3(i), it holds that $||\widehat{\boldsymbol{\gamma}}_n - \boldsymbol{\gamma}_0|| = O_P(1)$. Finally, we obtain the desired result by the basic norm inequality, as in the proof of Lemma B.2(ii)–(iii).

(iii) Recall that, as in the proof of Lemma B.2(iv), we write $\mathcal{L}_n(\theta, A_k, \boldsymbol{\gamma}_{-k}) = \mathcal{L}_n(\theta, \boldsymbol{\gamma})$ for $k \neq 1$. By the first-order condition and mean value expansion,

$$
\begin{aligned}
0 &= n\cdot\partial_{A_k}\mathcal{L}_n(\widehat{\theta}_n,\widehat{A}_{n,k},\widehat{\boldsymbol{\gamma}}_{n,-k}) \\
&= \frac{2}{n-1}\sum_{i=1}^{n}\sum_{j>i}s_{i,j}^{A_k} + n\cdot\partial_{A_k}\mathcal{L}_n(\widehat{\theta}_n,\widehat{A}_{n,k},\widehat{\boldsymbol{\gamma}}_{n,-k}) - n\cdot\partial_{A_k}\mathcal{L}_n(\widehat{\theta}_n,A_{0,k},\widehat{\boldsymbol{\gamma}}_{n,-k}) \\
&\quad + n\cdot\partial_{A_k}\mathcal{L}_n(\widehat{\theta}_n,A_{0,k},\widehat{\boldsymbol{\gamma}}_{n,-k}) - n\cdot\partial_{A_k}\mathcal{L}_n(\theta_0,A_{0,k},\widehat{\boldsymbol{\gamma}}_{n,-k}) + n\cdot\partial_{A_k}\mathcal{L}_n(\theta_0,A_{0,k},\widehat{\boldsymbol{\gamma}}_{n,-k}) - n\cdot\partial_{A_k}\mathcal{L}_n(\theta_0,A_{0,k},\boldsymbol{\gamma}_{0,-k}) \\
&= \frac{2}{n-1}\sum_{i=1}^{n}\sum_{j>i}s_{i,j}^{A_k} + \frac{2}{n-1}\sum_{j\neq k}h_{11,k,j}(\widehat{\theta}_n,\bar{A}_{n,k},\widehat{\boldsymbol{\gamma}}_{n,-k})[\widehat{A}_{n,k}-A_{0,k}] \\
&\quad + \frac{2}{n-1}\sum_{i=1}^{n}\sum_{j>i}\partial_{\theta^\top}s_{i,j}^{A_k}(\bar{\theta}_n,A_{0,k},\widehat{\boldsymbol{\gamma}}_{n,-k})[\widehat{\theta}_n-\theta_0] + \frac{2}{n-1}\sum_{i=1}^{n}\sum_{j>i}\partial_{\boldsymbol{\gamma}_{-k}^\top}s_{i,j}^{A_k}(\theta_0,A_{0,k},\bar{\boldsymbol{\gamma}}_{n,-k})[\widehat{\boldsymbol{\gamma}}_{n,-k}-\boldsymbol{\gamma}_{0,-k}],
\end{aligned}
$$

where $\bar{\theta}_n\in[\widehat{\theta}_n,\theta_0]$, $\bar{A}_{n,k}\in[\widehat{A}_{n,k},A_{0,k}]$, and $\bar{\boldsymbol{\gamma}}_{n,-k}\in[\widehat{\boldsymbol{\gamma}}_{n,-k},\boldsymbol{\gamma}_{0,-k}]$. Then, similar to the proof of Lemma B.2(iv), we have

$$
\left|\widehat{A}_{n,k}-A_{0,k}\right| \le (c+o_p(1))\left\{|T_{1,n,k}|+|T_{2,n,k}|+|T_{3,n,k}|\right\}
$$

for some $c>0$, where

$$
\begin{aligned}
T_{1,n,k} &\equiv \frac{2}{n-1}\sum_{i=1}^{n}\sum_{j>i}s_{i,j}^{A_k}, \quad T_{2,n,k} \equiv \frac{2}{n-1}\sum_{i=1}^{n}\sum_{j>i}\partial_{\boldsymbol{\gamma}_{-k}^\top}s_{i,j}^{A_k}(\theta_0,A_{0,k},\bar{\boldsymbol{\gamma}}_{n,-k})[\widehat{\boldsymbol{\gamma}}_{n,-k}-\boldsymbol{\gamma}_{0,-k}], \\
T_{3,n,k} &\equiv \frac{2}{n-1}\sum_{i=1}^{n}\sum_{j>i}\partial_{\theta^\top}s_{i,j}^{A_k}(\bar{\theta}_n,A_{0,k},\widehat{\boldsymbol{\gamma}}_{n,-k})[\widehat{\theta}_n-\theta_0].
\end{aligned}
$$

As shown in (B.14) and (B.15), $\max_{1\le k\le n}|T_{1,n,k}| = O_P(\sqrt{\ln n/n})$ and $\max_{1\le k\le n}|T_{2,n,k}| = O_P(n^{-1/2})$. For $T_{3,n,k}$, observe that

$$
\begin{aligned}
\frac{2}{n-1}\sum_{i=1}^{n}\sum_{j>i}\partial_{\theta^\top}s_{i,j}^{A_k}(\theta,\boldsymbol{\gamma}) &= \frac{2}{n-1}\sum_{i=1}^{n}\sum_{j>i}\left[\mathbf{1}\{i=k\}s_{1,i,j}^{\theta}(\theta,\boldsymbol{\gamma})^\top + \mathbf{1}\{j=k\}s_{2,i,j}^{\theta}(\theta,\boldsymbol{\gamma})^\top\right] \\
&= \frac{2}{n-1}\sum_{j>k}s_{1,k,j}^{\theta}(\theta,\boldsymbol{\gamma})^\top + \frac{2}{n-1}\sum_{i<k}s_{2,i,k}^{\theta}(\theta,\boldsymbol{\gamma})^\top.
\end{aligned}
$$

Hence, clearly, $\max_{1\le k\le n}||\frac{2}{n-1}\sum_{i=1}^{n}\sum_{j>i}\partial_{\theta^\top}s_{i,j}^{A_k}(\theta,\boldsymbol{\gamma})|| = O(1)$ for any $(\theta,\boldsymbol{\gamma})\in\Theta\times\mathbb{C}_n$. Then, with Lemma B.3,

$$
\max_{1\le k\le n}|T_{3,n,k}| = O_P(n^{-1/2}).
$$

Combining these results yields $\max_{1\le k\le n}|\widehat{A}_{n,k}-A_{0,k}| = O_P(\sqrt{\ln n/n}) + O_P(n^{-1/2})$.

Similarly as above, we can also show that $\max_{1\le k\le n}|\widehat{B}_{n,k}-B_{0,k}| = O_P(\sqrt{\ln n/n}) + O_P(n^{-1/2})$. This completes the proof. □

Proof of Theorem 3.3

To simplify the discussion, we focus on the estimation of $\mathcal{C}_0^A$ with $K^A=3$ only. The other cases can be proved analogously. In addition, for notational simplicity, we omit the superscript A in this proof.

Now, let

$$u_{n,i} \equiv \widehat{A}_{n,i} - A_{0,i} \text{ for } i = 1, \dots, n.$$

In particular, $u_{n,1} = 0$ holds by the normalization. In accordance with the ordering $\widehat{A}_{n,(1)} \leq \cdots \leq \widehat{A}_{n,(n)}$, we permutate $A_{0,i}$'s and obtain $\{A_{0,(i)}\}$. By Theorem 3.2(iii), we have $\max_{1\leq i\leq n} |u_{n,i}| = O_P(\sqrt{\ln n/n})$. Hence, w.p.a.1, the sequence $\{A_{0,(i)}\}$ contains two "true" break points $(t_1^0, t_2^0) \equiv (t_{n,1}^0, t_{n,2}^0)$ in the following manner:

$$A_{0,(i)} = \begin{cases} a_{0,1} & \text{if } 1 \leq i \leq t_1^0 \\ a_{0,2} & \text{if } t_1^0 + 1 \leq i \leq t_2^0 \\ a_{0,3} & \text{if } t_2^0 + 1 \leq i \leq n. \end{cases}$$

We can assume, without loss of generality, that $\widehat{S}_{1,n}(t_1^0) < \widehat{S}_{1,n}(t_2^0)$. If $\widehat{S}_{1,n}(t_1^0) > \widehat{S}_{1,n}(t_2^0)$, by reversing the order of $\{\widehat{A}_{n,(i)}\}$ and re-labeling the break points appropriately, we can prove the theorem completely analogously. Recall that $\widehat{t}_1 = \operatorname{argmin}_{1\leq\kappa<n} \widehat{S}_{1,n}(\kappa)$. We first show that $\Pr(\widehat{t}_1 = t_1^0) \to 1$ by demonstrating that (i) $\Pr(\widehat{t}_1 < t_1^0) \to 0$, (ii) $\Pr(t_1^0 < \widehat{t}_1 \leq t_2^0) \to 0$, and (iii) $\Pr(t_2^0 < \widehat{t}_1) \to 0$.

(i) For a given $m < t_1^0$, we have

$$\bar{A}_{n,1,m} = \frac{1}{m}\sum_{l=1}^{m}(A_{0,(l)} + u_{n,(l)}) = a_{0,1} + \bar{u}_{n,1,m}$$
$$\bar{A}_{n,m+1,n} = \frac{1}{n-m}\sum_{l=m+1}^{n}(A_{0,(l)} + u_{n,(l)}) = \frac{(t_1^0 - m)a_{0,1}}{n-m} + \frac{(t_2^0 - t_1^0)a_{0,2}}{n-m} + \frac{(n - t_2^0)a_{0,3}}{n-m} + \bar{u}_{n,m+1,n},$$

where $\bar{u}_{n,1,m} = m^{-1}\sum_{l=1}^{m} u_{n,(l)}$, and $\bar{u}_{n,m+1,n} = (n-m)^{-1}\sum_{l=m+1}^{n} u_{n,(l)}$. Hence, since $\widehat{A}_{n,(l)} - \bar{A}_{n,1,m} = u_{n,(l)} - \bar{u}_{n,1,m}$ for $l \leq m$, we have

$$\widehat{\Delta}(1, m) = \sum_{l=1}^{m}(u_{n,(l)} - \bar{u}_{n,1,m})^2.$$

Similarly, since

$$\widehat{A}_{n,(l)} - \bar{A}_{n,m+1,n} = \begin{cases} a_{1m} + u_{n,(l)} - \bar{u}_{n,m+1,n} & \text{if } m + 1 \leq l \leq t_1^0 \\ a_{2m} + u_{n,(l)} - \bar{u}_{n,m+1,n} & \text{if } t_1^0 + 1 \leq l \leq t_2^0 \\ a_{3m} + u_{n,(l)} - \bar{u}_{n,m+1,n} & \text{if } t_2^0 + 1 \leq l \leq n, \end{cases}$$

where $a_{1m} \equiv \frac{(t_2^0 - t_1^0)(a_{0,1} - a_{0,2})}{n-m} + \frac{(n - t_2^0)(a_{0,1} - a_{0,3})}{n-m}$, $a_{2m} \equiv \frac{(t_1^0 - m)(a_{0,2} - a_{0,1})}{n-m} + \frac{(n - t_2^0)(a_{0,2} - a_{0,3})}{n-m}$, and $a_{3m} \equiv \frac{(t_1^0 - m)(a_{0,3} - a_{0,1})}{n-m} + \frac{(t_2^0 - t_1^0)(a_{0,3} - a_{0,2})}{n-m}$, we have

$$\widehat{\Delta}(m + 1, n) = (t_1^0 - m)a_{1m}^2 + (t_2^0 - t_1^0)a_{2m}^2 + (n - t_2^0)a_{3m}^2 + \sum_{l=m+1}^{n}(u_{n,(l)} - \bar{u}_{n,m+1,n})^2$$

$$+ 2a_{1m} \sum_{l=m+1}^{t_1^0} (u_{n,(l)} - \bar{u}_{n,m+1,n}) + 2a_{2m} \sum_{l=t_1^0+1}^{t_2^0} (u_{n,(l)} - \bar{u}_{n,m+1,n}) + 2a_{3m} \sum_{l=t_2^0+1}^{n} (u_{n,(l)} - \bar{u}_{n,m+1,n}).$$

Then, it holds that $\widehat{S}_{1,n}(m) = \frac{1}{n}\left(\widehat{\Delta}(1,m) + \widehat{\Delta}(m+1,n)\right) = \mu_1(m) + r_1(m)$, where

$$\begin{aligned}
\mu_1(m) &\equiv \frac{t_1^0 - m}{n} a_{1m}^2 + \frac{t_2^0 - t_1^0}{n} a_{2m}^2 + \frac{n - t_2^0}{n} a_{3m}^2 \\
r_1(m) &\equiv \frac{1}{n}\left[\sum_{l=1}^{m}(u_{n,(l)} - \bar{u}_{n,1,m})^2 + \sum_{l=m+1}^{n}(u_{n,(l)} - \bar{u}_{n,m+1,n})^2\right] \\
&\quad + \frac{2a_{1m}}{n} \sum_{l=m+1}^{t_1^0} (u_{n,(l)} - \bar{u}_{n,m+1,n}) + \frac{2a_{2m}}{n} \sum_{l=t_1^0+1}^{t_2^0} (u_{n,(l)} - \bar{u}_{n,m+1,n}) + \frac{2a_{3m}}{n} \sum_{l=t_2^0+1}^{n} (u_{n,(l)} - \bar{u}_{n,m+1,n}).
\end{aligned}$$

By similar calculation, we can observe $\widehat{S}_{1,n}(t_1^0) = \frac{1}{n}\left(\widehat{\Delta}(1,t_1^0) + \widehat{\Delta}(t_1^0+1,n)\right) = \mu_1(t_1^0) + r_1(t_1^0)$, where

$$\begin{aligned}
\widehat{\Delta}(1,t_1^0) &= \sum_{l=1}^{t_1^0}(u_{n,(l)} - \bar{u}_{n,1,t_1^0})^2 \\
\widehat{\Delta}(t_1^0+1,n) &= (t_2^0 - t_1^0) a_{2t_1^0}^2 + (n - t_2^0) a_{3t_1^0}^2 + \sum_{l=t_1^0+1}^{n}(u_{n,(l)} - \bar{u}_{n,t_1^0+1,n})^2 \\
&\quad + 2a_{2t_1^0} \sum_{l=t_1^0+1}^{t_2^0} (u_{n,(l)} - \bar{u}_{n,t_1^0+1,n}) + 2a_{3t_1^0} \sum_{l=t_2^0+1}^{n} (u_{n,(l)} - \bar{u}_{n,t_1^0+1,n}),
\end{aligned}$$

with $a_{2t_1^0} \equiv \frac{(n-t_2^0)(a_{0,2}-a_{0,3})}{n-t_1^0}$, and $a_{3t_1^0} \equiv \frac{(t_2^0-t_1^0)(a_{0,3}-a_{0,2})}{n-t_1^0}$, and

$$\begin{aligned}
\mu_1(t_1^0) &\equiv \frac{t_2^0 - t_1^0}{n} a_{2t_1^0}^2 + \frac{n - t_2^0}{n} a_{3t_1^0}^2 \\
r_1(t_1^0) &\equiv \frac{1}{n}\left[\sum_{l=1}^{t_1^0}(u_{n,(l)} - \bar{u}_{n,1,t_1^0})^2 + \sum_{l=t_1^0+1}^{n}(u_{n,(l)} - \bar{u}_{n,t_1^0+1,n})^2\right] \\
&\quad + \frac{2a_{2t_1^0}}{n} \sum_{l=t_1^0+1}^{t_2^0} (u_{n,(l)} - \bar{u}_{n,t_1^0+1,n}) + \frac{2a_{3t_1^0}}{n} \sum_{l=t_2^0+1}^{n} (u_{n,(l)} - \bar{u}_{n,t_1^0+1,n}).
\end{aligned}$$

By tedious calculation, we can find that

$$\mu_1(m) - \mu_1(t_1^0) = \frac{(t_1^0 - m)}{n(1-m/n)(1-t_1^0/n)}\left[\left(1 - \frac{t_1^0}{n}\right)(a_{0,1} - a_{0,2}) + \left(1 - \frac{t_2^0}{n}\right)(a_{0,2} - a_{0,3})\right]^2.$$

Note that $\mu_1(m) - \mu_1(t_1^0) > 0$ for any $m < t_1^0$ under Assumption 3.4(i). Meanwhile,

$$r_1(m) - r_1(t_1^0) = \frac{1}{n}\left[\sum_{l=1}^{m}(u_{n,(l)} - \bar{u}_{n,1,m})^2 - \sum_{l=1}^{t_1^0}(u_{n,(l)} - \bar{u}_{n,1,t_1^0})^2\right]$$

$$
\begin{aligned}
&+\frac{1}{n}\left[\sum_{l=m+1}^{n}(u_{n,(l)}-\bar{u}_{n,m+1,n})^2-\sum_{l=t_1^0+1}^{n}(u_{n,(l)}-\bar{u}_{n,t_1^0+1,n})^2\right]\\
&+\frac{2a_{1m}}{n}\sum_{l=m+1}^{t_1^0}(u_{n,(l)}-\bar{u}_{n,m+1,n})\\
&+\frac{2}{n}\left(a_{2m}\sum_{l=t_1^0+1}^{t_2^0}(u_{n,(l)}-\bar{u}_{n,m+1,n})-a_{2t_1^0}\sum_{l=t_1^0+1}^{t_2^0}(u_{n,(l)}-\bar{u}_{n,t_1^0+1,n})\right)\\
&+\frac{2}{n}\left(a_{3m}\sum_{l=t_2^0+1}^{n}(u_{n,(l)}-\bar{u}_{n,m+1,n})-a_{3t_1^0}n\sum_{l=t_2^0+1}^{n}(u_{n,(l)}-\bar{u}_{n,t_1^0+1,n})\right).
\end{aligned}
$$

Then, by carefully examining each term of the right hand,[1] we can find that $r_1(m)-r_1(t_1^0)=\frac{t_1^0-m}{n}\cdot\{O_P(\ln n/n)+O_P(\sqrt{\ln n/n})\}$. Hence, we have

$$
\begin{aligned}
\widehat{S}_{1,n}(m)-\widehat{S}_{1,n}(t_1^0)=&\underbrace{\mu_1(m)-\mu_1(t_1^0)}_{=n^{-1}(t_1^0-m)\cdot c_n\ (c_n>0)}+\underbrace{r_1(m)-r_1(t_1^0)}_{n^{-1}(t_1^0-m)\cdot o_P(1)}\\
\geq&\ 0
\end{aligned}
$$

w.p.a.1 for any $m<t_1^0$. Thus, since $\widehat{t}_1$ is a minimizer of $\widehat{S}_{1,n}(m)$,

$$
\begin{aligned}
\Pr(\widehat{t}_1<t_1^0)&=\Pr(\widehat{S}_{1,n}(\widehat{t}_1)-\widehat{S}_{1,n}(t_1^0)\geq 0,\ \widehat{t}_1<t_1^0)+\Pr(\widehat{S}_{1,n}(\widehat{t}_1)-\widehat{S}_{1,n}(t_1^0)<0,\ \widehat{t}_1<t_1^0)\\
&=\Pr(\widehat{S}_{1,n}(\widehat{t}_1)-\widehat{S}_{1,n}(t_1^0)<0,\ \widehat{t}_1<t_1^0)\\
&\leq\Pr(\widehat{S}_{1,n}(m)-\widehat{S}_{1,n}(t_1^0)<0,\ m<t_1^0\ \text{ for some } m)\to 0
\end{aligned}
$$

as $n\to\infty$. Thus, (i) follows.

(ii) For a given $t_1^0<m\leq t_2^0$, we have

$$
\begin{aligned}
\bar{A}_{n,1,m}&=\frac{1}{m}\sum_{l=1}^{m}(A_{0,(l)}+u_{n,(l)})=\frac{t_1^0}{m}a_{0,1}+\frac{m-t_1^0}{m}a_{0,2}+\bar{u}_{n,1,m}\\
\bar{A}_{n,m+1,n}&=\frac{1}{n-m}\sum_{l=m+1}^{n}(A_{0,(l)}+u_{n,(l)})=\frac{(t_2^0-m)a_{0,2}}{n-m}+\frac{(n-t_2^0)a_{0,3}}{n-m}+\bar{u}_{n,m+1,n}.
\end{aligned}
$$

Hence, since

$$
\widehat{A}_{n,(l)}-\bar{A}_{n,1,m}=\begin{cases}a_{1m}+u_{n,(l)}-\bar{u}_{n,1,m} & \text{if }\ 1\leq l\leq t_1^0\\ a_{2m}+u_{n,(l)}-\bar{u}_{n,1,m} & \text{if }\ t_1^0+1\leq l\leq m,\end{cases}
$$

where $a_{1m}\equiv\frac{(m-t_1^0)(a_{0,1}-a_{0,2})}{m}$, and $a_{2m}\equiv\frac{t_1^0(a_{0,2}-a_{0,1})}{m}$, we have

$$
\widehat{\Delta}(1,m)=t_1^0a_{1m}^2+(m-t_1^0)a_{2m}^2+\sum_{l=1}^{m}(u_{n,(l)}-\bar{u}_{n,1,m})^2
$$

[1]For this calculation, see pp. 8–9 in Supplementary Material for Wang and Su (2020).

$$+ 2a_{1m}\sum_{l=1}^{t_1^0}(u_{n,(l)} - \bar{u}_{n,1,m}) + 2a_{2m}\sum_{l=t_1^0+1}^{m}(u_{n,(l)} - \bar{u}_{n,1,m}).$$

Similarly, since

$$\widehat{A}_{n,(l)} - \bar{A}_{n,m+1,n} = \begin{cases} a_{3m} + u_{n,(l)} - \bar{u}_{n,m+1,n} & \text{if } m+1 \leq l \leq t_2^0 \\ a_{4m} + u_{n,(l)} - \bar{u}_{n,m+1,n} & \text{if } t_2^0+1 \leq l \leq n, \end{cases}$$

where $a_{3m} \equiv \frac{(n-t_2^0)(a_{0,2}-a_{0,3})}{n-m}$, and $a_{4m} \equiv \frac{(t_2^0-m)(a_{0,3}-a_{0,2})}{n-m}$, we have

$$\widehat{\Delta}(m+1,n) = (t_2^0 - m)a_{3m}^2 + (n-t_2^0)a_{4m}^2 + \sum_{l=m+1}^{n}(u_{n,(l)} - \bar{u}_{n,m+1,n})^2$$

$$+ 2a_{3m}\sum_{l=m+1}^{t_2^0}(u_{n,(l)} - \bar{u}_{n,m+1,n}) + 2a_{4m}\sum_{l=t_2^0+1}^{n}(u_{n,(l)} - \bar{u}_{n,m+1,n}).$$

Then, it holds that $\widehat{S}_{1,n}(m) = \frac{1}{n}\left(\widehat{\Delta}(1,m) + \widehat{\Delta}(m+1,n)\right) = \mu_2(m) + r_2(m)$, where

$$\begin{aligned}
\mu_2(m) &\equiv \frac{t_1^0}{n}a_{1m}^2 + \frac{m-t_1^0}{n}a_{2m}^2 + \frac{t_2^0-m}{n}a_{3m}^2 + \frac{n-t_2^0}{n}a_{4m}^2 \\
r_2(m) &\equiv \frac{1}{n}\left[\sum_{l=1}^{m}(u_{n,(l)} - \bar{u}_{n,1,m})^2 + \sum_{l=m+1}^{n}(u_{n,(l)} - \bar{u}_{n,m+1,n})^2\right] \\
&\quad + \frac{2a_{1m}}{n}\sum_{l=1}^{t_1^0}(u_{n,(l)} - \bar{u}_{n,1,m}) + \frac{2a_{2m}}{n}\sum_{l=t_1^0+1}^{m}(u_{n,(l)} - \bar{u}_{n,1,m}) \\
&\quad + \frac{2a_{3m}}{n}\sum_{l=m+1}^{t_2^0}(u_{n,(l)} - \bar{u}_{n,m+1,n}) + \frac{2a_{4m}}{n}\sum_{l=t_2^0+1}^{n}(u_{n,(l)} - \bar{u}_{n,m+1,n}).
\end{aligned}$$

To be consistent with the above notations, we re-write $\widehat{S}_{1,n}(t_1^0) = \frac{1}{n}\left(\widehat{\Delta}(1,t_1^0) + \widehat{\Delta}(t_1^0+1,n)\right) = \mu_2(t_1^0) + r_2(t_1^0)$, where

$$\begin{aligned}
\mu_2(t_1^0) &\equiv \frac{t_2^0-t_1^0}{n}a_{3t_1^0}^2 + \frac{n-t_2^0}{n}a_{4t_1^0}^2 \\
r_2(t_1^0) &\equiv \frac{1}{n}\left[\sum_{l=1}^{t_1^0}(u_{n,(l)} - \bar{u}_{n,1,t_1^0})^2 + \sum_{l=t_1^0+1}^{n}(u_{n,(l)} - \bar{u}_{n,t_1^0+1,n})^2\right] \\
&\quad + \frac{2a_{3t_1^0}}{n}\sum_{l=t_1^0+1}^{t_2^0}(u_{n,(l)} - \bar{u}_{n,t_1^0+1,n}) + \frac{2a_{4t_1^0}}{n}\sum_{l=t_2^0+1}^{n}(u_{n,(l)} - \bar{u}_{n,t_1^0+1,n}),
\end{aligned}$$

with $a_{3t_1^0} \equiv \frac{(n-t_2^0)(a_{0,2}-a_{0,3})}{n-t_1^0}$, and $a_{4t_1^0} \equiv \frac{(t_2^0-t_1^0)(a_{0,3}-a_{0,2})}{n-t_1^0}$. By tedious calculation, we can find that

$$\mu_2(m) - \mu_2(t_1^0) = \frac{m-t_1^0}{n}\left[\frac{t_1^0}{m}(a_{0,1}-a_{0,2})^2 - \frac{(n-t_2^0)^2}{(n-m)(n-t_1^0)}(a_{0,2}-a_{0,3})^2\right]$$

$$\geq \frac{m-t_1^0}{n}\left[\frac{t_1^0}{m}(a_{0,1}-a_{0,2})^2-\frac{t_2^0(n-t_2^0)}{m(n-t_1^0)}(a_{0,2}-a_{0,3})^2\right]$$
$$=\frac{m-t_1^0}{n}\frac{t_2^0}{m}\left[\frac{t_1^0}{t_2^0}(a_{0,1}-a_{0,2})^2-\frac{n-t_2^0}{n-t_1^0}(a_{0,2}-a_{0,3})^2\right],$$

where the inequality follows because $(n-t_2^0)/(n-m)\leq t_2^0/m$. Further, it can be shown that $r_2(m)-r_2(t_1^0)=n^{-1}(m-t_1^0)\cdot o_P(1)$ uniformly in $t_1^0<m\leq t_2^0$. Here, recall that we have assumed $\widehat{S}_{1,n}(t_1^0)<\widehat{S}_{1,n}(t_2^0)$. Through straightforward calculation, we can find that

$$0<\widehat{S}_{1,n}(t_2^0)-\widehat{S}_{1,n}(t_1^0)=\frac{t_1^0}{n}a_{1t_2^0}^2+\frac{t_2^0-t_1^0}{n}a_{2t_2^0}^2-\frac{t_2^0-t_1^0}{n}a_{3t_1^0}^2-\frac{n-t_2^0}{n}a_{4t_1^0}^2+\frac{t_2^0-t_1^0}{n}\cdot o_P(1)$$
$$=\frac{t_2^0-t_1^0}{n}\left[\frac{t_1^0}{t_2^0}(a_{0,1}-a_{0,2})^2-\frac{n-t_2^0}{n-t_1^0}(a_{0,2}-a_{0,3})^2\right]+\frac{t_2^0-t_1^0}{n}\cdot o_P(1),$$

where $a_{1t_2^0}\equiv\frac{(t_2^0-t_1^0)(a_{0,1}-a_{0,2})}{t_2^0}$, and $a_{2t_2^0}\equiv\frac{t_1^0(a_{0,2}-a_{0,1})}{t_2^0}$. This implies that

$$\frac{t_1^0}{t_2^0}(a_{0,1}-a_{0,2})^2-\frac{n-t_2^0}{n-t_1^0}(a_{0,2}-a_{0,3})^2\to\frac{\tau_1}{\tau_1+\tau_2}(a_{0,1}-a_{0,2})^2-\frac{\tau_3}{\tau_2+\tau_3}(a_{0,2}-a_{0,3})^2>0$$

by Assumption 3.4(ii). Hence, $\mu_2(m)-\mu_2(t_2^0)>0$ holds for any $t_1^0<m\leq t_2^0$ for sufficiently large n. Then, combining these results implies that $\widehat{S}_{1,n}(m)-\widehat{S}_{1,n}(t_1^0)>0$ w.p.a.1 uniformly in $t_1^0<m\leq t_2^0$, leading to the desired result:

$$\Pr(t_1^0<\hat{t}_1\leq t_2^0)=\Pr(\widehat{S}_{1,n}(\hat{t}_1)-\widehat{S}_{1,n}(t_1^0)\geq 0,\ t_1^0<\hat{t}_1\leq t_2^0)+\Pr(\widehat{S}_{1,n}(\hat{t}_1)-\widehat{S}_{1,n}(t_1^0)<0,\ t_1^0<\hat{t}_1\leq t_2^0)$$
$$=\Pr(\widehat{S}_{1,n}(\hat{t}_1)-\widehat{S}_{1,n}(t_1^0)<0,\ t_1^0<\hat{t}_1\leq t_2^0)$$
$$\leq\Pr(\widehat{S}_{1,n}(m)-\widehat{S}_{1,n}(t_1^0)<0,\ t_1^0<m\leq t_2^0\ \text{ for some } m)\to 0.$$

Then, (ii) also follows. For case (iii), since this case is symmetric to (i), we can safely omit the proof.

Once the first break point is obtained, we can partition $\{\widehat{A}_{n,(i)}\}$ into two subregions $\{\widehat{A}_{n,(i)}\}_{i=1}^{\hat{t}_1}$ and $\{\widehat{A}_{n,(i)}\}_{i=\hat{t}_1+1}^{n}$. We next estimate in which group the second break point exists. Given the above consistency result, w.p.a.1, we have $\widehat{S}_{1,\hat{t}_1}(\hat{t}_1)=\widehat{S}_{1,t_1^0}(t_1^0)$ and $\widehat{S}_{\hat{t}_1+1,n}(n)=\widehat{S}_{t_1^0+1,n}(n)$. Since $t_1^0<t_2^0$, it suffices to show that $\Pr(\widehat{S}_{1,t_1^0}(t_1^0)<\widehat{S}_{t_1^0+1,n}(n))\to 1$. We can observe that $\widehat{S}_{1,t_1^0}(t_1^0)=\frac{1}{t_1^0}\widehat{\Delta}(1,t_1^0)=r_{t_1^0}^*$ and $\widehat{S}_{t_1^0+1,n}(n)=\frac{1}{n-t_1^0}\widehat{\Delta}(t_1^0+1,n)=\mu_n^*+r_n^*$, where

$$r_{t_1^0}^*\equiv\frac{1}{t_1^0}\sum_{l=1}^{t_1^0}(u_{n,(l)}-\bar{u}_{n,1,t_1^0})^2$$
$$\mu_n^*\equiv\frac{t_2^0-t_1^0}{n-t_1^0}(a_{1n}^*)^2+\frac{n-t_2^0}{n-t_1^0}(a_{2n}^*)^2$$
$$r_n^*\equiv\frac{1}{n-t_1^0}\sum_{l=t_1^0+1}^{n}(u_{n,(l)}-\bar{u}_{n,t_1^0+1,n})^2+\frac{2a_{1n}^*}{n-t_1^0}\sum_{l=t_1^0+1}^{t_2^0}(u_{n,(l)}-\bar{u}_{n,t_1^0+1,n})+\frac{2a_{2n}^*}{n-t_1^0}\sum_{l=t_2^0+1}^{n}(u_{n,(l)}-\bar{u}_{n,t_1^0+1,n}),$$

with $a_{1n}^*\equiv\frac{(n-t_2^0)(a_{0,2}-a_{0,3})}{n-t_1^0}$, and $a_{2n}^*\equiv\frac{(t_2^0-t_1^0)(a_{0,3}-a_{0,2})}{n-t_1^0}$. Through some calculation, we can find that

$$\mu_n^*=\frac{(t_2^0-t_1^0)(n-t_2^0)}{(n-t_1^0)^2}(a_{0,2}-a_{0,3})^2\to\frac{\tau_2\tau_3}{(\tau_2+\tau_3)^2}(a_{0,2}-a_{0,3})^2>0$$

under Assumptions 3.4(i) and (ii). It can be easily seen that $r_n^* - r_{t_1^0}^* = o_P(1)$. Hence, $\widehat{S}_{t_1^0+1,n}(n) - \widehat{S}_{1,t_1^0}(t_1^0) > 0$ holds w.p.a.1, as desired. Then, once the subset $\{\widehat{A}_{n,(i)}\}_{i=\widehat{t}_1+1}^n$ is selected for the detection of the second break point, it can be estimated by $\widehat{t}_2 = \mathrm{argmin}_{\widehat{t}_1+1\le\kappa<n} \widehat{S}_{\widehat{t}_1+1,n}(\kappa)$. The consistency of $\widehat{t}_2$ for t_2^0 follows from the same argument as above. □

Proof of Theorem 3.4

The consistency of $\widehat{\delta}_n^{\mathrm{oracle}}$ is straightforward from Theorem 3.1. By the first-order condition, Taylor expansion, and the law of large numbers, we have

$$\sqrt{\frac{N}{2}}(\widehat{\delta}_n^{\mathrm{oracle}} - \delta_0) = -\mathbb{E}\left[\partial^2_{\delta\delta^\top}\mathcal{L}_n(\delta_0)\right]^{-1}\sqrt{\frac{2}{N}}\sum_{i=1}^n\sum_{j>i} s_{i,j}^\delta(\delta_0) + o_P(1).$$

Note that, under the assumptions made, $\{s_{i,j}^\delta(\delta_0)\}$ are uniformly bounded, and $\sum_{i=1}^n\sum_{j>i} s_{i,j}^\delta(\delta_0)$ is a sum of independent random variables. Thus, the asymptotic normality result follows from the central limit theorem for bounded random variables (see, e.g., Example 27.4 in Billingsley (2012)).

The asymptotic distributional equivalence result is straightforward from

$$\begin{aligned}\Pr\left(\sqrt{\frac{N}{2}}(\widehat{\delta}_n - \delta_0) \in C\right) &= \Pr\left(\sqrt{\frac{N}{2}}(\widehat{\delta}_n - \delta_0) \in C,\ (\widehat{\mathcal{C}}_n^A, \widehat{\mathcal{C}}_n^B) = (\mathcal{C}_0^A, \mathcal{C}_0^B)\right)\\ &\quad + \Pr\left(\sqrt{\frac{N}{2}}(\widehat{\delta}_n - \delta_0) \in C,\ (\widehat{\mathcal{C}}_n^A, \widehat{\mathcal{C}}_n^B) \neq (\mathcal{C}_0^A, \mathcal{C}_0^B)\right)\\ &= \Pr\left(\sqrt{\frac{N}{2}}(\widehat{\delta}_n^{\mathrm{oracle}} - \delta_0) \in C\right) + o(1)\end{aligned}$$

for any $C \subseteq \mathbb{R}^{d_z + K^A + K^B + 1}$. □

C Supplementary Materials

C.1 Explicit form of $\partial^2\mathcal{L}_n^*(\rho)/(\partial\rho)^2$

For notational simplicity, let $\mathbf{C} = (\beta^\top, \alpha, \boldsymbol{\gamma}^\top)^\top$ and $\widetilde{\mathbf{C}}_0(\rho) = (\widetilde{\beta}_0(\rho)^\top, \widetilde{\alpha}_0(\rho), \widetilde{\boldsymbol{\gamma}}_0(\rho)^\top)^\top$, and write $\mathcal{L}_n(\theta, \boldsymbol{\gamma})$ equivalently as $\mathcal{L}_n(\rho, \mathbf{C})$, so that $\widetilde{\mathbf{C}}_0(\rho) = \mathrm{argmax}_{\mathbf{C}\in\mathcal{B}\times\mathcal{A}\times\mathbb{C}_n}\mathcal{L}_n(\rho, \mathbf{C})$ and $\mathcal{L}_n^*(\rho) = \mathbb{E}\mathcal{L}_n(\rho, \widetilde{\mathbf{C}}_0(\rho))$. By the chain rule,

$$\partial_\rho\mathcal{L}_n^*(\rho) = \partial_{\rho_1}\mathbb{E}\mathcal{L}_n(\rho, \widetilde{\mathbf{C}}_0(\rho)) + \partial_{\mathbf{C}^\top}\mathbb{E}\mathcal{L}_n(\rho, \widetilde{\mathbf{C}}_0(\rho))\partial_\rho\widetilde{\mathbf{C}}_0(\rho) = \partial_{\rho_1}\mathbb{E}\mathcal{L}_n(\rho, \widetilde{\mathbf{C}}_0(\rho)),$$

where the second equality holds for any $\rho \in \mathcal{R}$ by the first-order condition for $\widetilde{\mathbf{C}}_0(\rho)$, and we have denoted ∂_{ρ_1} to stand for the partial derivative with respect to the "first" ρ. Then, we also have

$$\partial^2_{\rho\rho}\mathcal{L}_n^*(\rho) = \partial^2_{\rho_1\rho_1}\mathbb{E}\mathcal{L}_n(\rho, \widetilde{\mathbf{C}}_0(\rho)) + \partial^2_{\rho_1\mathbf{C}^\top}\mathbb{E}\mathcal{L}_n(\rho, \widetilde{\mathbf{C}}_0(\rho))\partial_\rho\widetilde{\mathbf{C}}_0(\rho).$$

In addition, applying the implicit function theorem to $\partial_{\mathbf{C}}\mathbb{E}\mathcal{L}_n(\rho, \widetilde{\mathbf{C}}_0(\rho)) = \mathbf{0}_{(d_z+2n)\times 1}$ for $\rho \in \mathcal{R}$ yields

$$\mathbf{0}_{(d_z+2n)\times 1} = \partial^2_{\mathbf{C}\rho}\mathbb{E}\mathcal{L}_n(\rho, \widetilde{\mathbf{C}}_0(\rho)) = \partial^2_{\mathbf{C}\rho_1}\mathbb{E}\mathcal{L}_n(\rho, \widetilde{\mathbf{C}}_0(\rho)) + \partial^2_{\mathbf{C}\mathbf{C}^\top}\mathbb{E}\mathcal{L}_n(\rho, \widetilde{\mathbf{C}}_0(\rho))\partial_\rho\widetilde{\mathbf{C}}_0(\rho)$$
$$\Longrightarrow \partial_\rho\widetilde{\mathbf{C}}_0(\rho) = -\left[\partial^2_{\mathbf{C}\mathbf{C}^\top}\mathbb{E}\mathcal{L}_n(\rho, \widetilde{\mathbf{C}}_0(\rho))\right]^{-1}\partial^2_{\mathbf{C}\rho_1}\mathbb{E}\mathcal{L}_n(\rho, \widetilde{\mathbf{C}}_0(\rho)).$$

Therefore, we obtain

$$\partial^2_{\rho\rho}\mathcal{L}^*_n(\rho) = \partial^2_{\rho_1\rho_1}\mathbb{E}\mathcal{L}_n(\rho, \widetilde{\mathbf{C}}_0(\rho)) - \partial^2_{\rho_1\mathbf{C}^\top}\mathbb{E}\mathcal{L}_n(\rho, \widetilde{\mathbf{C}}_0(\rho))\left[\partial^2_{\mathbf{C}\mathbf{C}^\top}\mathbb{E}\mathcal{L}_n(\rho, \widetilde{\mathbf{C}}_0(\rho))\right]^{-1}\partial^2_{\mathbf{C}\rho_1}\mathbb{E}\mathcal{L}_n(\rho, \widetilde{\mathbf{C}}_0(\rho)).$$

Note that since $\partial^2_{\mathbf{C}\mathbf{C}^\top}\mathbb{E}\mathcal{L}_n(\rho, \widetilde{\mathbf{C}}_0(\rho))$ is negative semidefinite by the second-order condition of maximization, the second term on the right-hand side is non-positive. Thus, to ensure that $\partial^2_{\rho\rho}\mathcal{L}^*_n(\rho)$ is strictly negative, $\partial^2_{\rho_1\rho_1}\mathbb{E}\mathcal{L}_n(\rho, \widetilde{\mathbf{C}}_0(\rho))$ must also be negative and larger in magnitude than the second term.

C.2 Supplementary tables for Section 4

Table C.1: Simulation Results **DGP 1**: Estimation of Common Parameters

			α_0		$\beta_{0,1}$		$\beta_{0,2}$		ρ_0	
r	n	Estimator	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
0.4	54	1stML	0.000	0.212	-0.136	0.380	0.223	0.255	0.044	0.181
		BS	0.073	0.167	-0.071	0.178	0.109	0.149	-0.079	0.173
		BS4	0.053	0.178	-0.115	0.290	0.181	0.215	-0.025	0.168
		repBS4	0.051	0.179	-0.118	0.295	0.186	0.218	-0.022	0.166
		KM	0.074	0.167	-0.072	0.177	0.110	0.149	-0.080	0.173
		KM4	0.048	0.186	-0.119	0.293	0.186	0.218	-0.020	0.170
		Oracle	-0.006	0.126	-0.004	0.073	0.011	0.089	0.009	0.118
	72	1stML	0.021	0.163	-0.121	0.285	0.159	0.181	0.010	0.118
		BS	0.068	0.131	-0.066	0.142	0.081	0.111	-0.080	0.134
		BS4	0.056	0.133	-0.103	0.226	0.132	0.156	-0.038	0.120
		repBS4	0.055	0.131	-0.105	0.231	0.134	0.158	-0.034	0.118
		KM	0.068	0.131	-0.065	0.142	0.080	0.111	-0.080	0.134
		KM4	0.056	0.132	-0.105	0.230	0.135	0.158	-0.035	0.118
		Oracle	0.005	0.101	-0.004	0.055	0.007	0.070	-0.004	0.091
0.8	54	1stML	0.016	0.239	-0.142	0.369	0.230	0.262	0.037	0.184
		BS	0.060	0.163	-0.042	0.172	0.065	0.120	-0.119	0.190
		BS4	0.048	0.193	-0.103	0.287	0.166	0.200	-0.034	0.167
		repBS4	0.050	0.178	-0.105	0.291	0.170	0.204	-0.029	0.166
		KM	0.059	0.165	-0.042	0.172	0.064	0.119	-0.119	0.192
		KM4	0.052	0.181	-0.105	0.287	0.171	0.206	-0.031	0.168
		Oracle	0.002	0.131	-0.008	0.071	0.015	0.087	0.001	0.120
	72	1stML	0.038	0.163	-0.113	0.293	0.161	0.183	0.005	0.128
		BS	0.055	0.125	-0.037	0.141	0.042	0.083	-0.101	0.152
		BS4	0.054	0.135	-0.086	0.235	0.116	0.142	-0.044	0.130
		repBS4	0.055	0.135	-0.088	0.238	0.119	0.144	-0.043	0.128
		KM	0.054	0.124	-0.037	0.141	0.042	0.083	-0.100	0.151
		KM4	0.055	0.143	-0.087	0.236	0.119	0.144	-0.045	0.130
		Oracle	0.010	0.097	-0.008	0.054	0.010	0.067	-0.009	0.092

Note. 1stML: the initial ML estimator, BS: the three-step ML estimator based on the BS method with no repartitions, BS4: the BS estimator with $K^A = K^B = 4$, repBS4: the repartitioned BS4 estimator, KM: the three-step ML estimator based on the k-means method, KM4: the KM estimator with $K^A = K^B = 4$, Oracle: the oracle estimator based on the true group membership.

Table C.2: Simulation Results **DGP 2**: Estimation of Common Parameters

			α_0		$\beta_{0,1}$		$\beta_{0,2}$		ρ_0	
r	n	Estimator	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
0.4	54	1stML	-0.009	0.251	-0.145	0.402	0.218	0.251	0.040	0.180
		BS	0.053	0.163	-0.089	0.257	0.123	0.162	-0.074	0.172
		repBS	0.053	0.161	-0.102	0.277	0.139	0.174	-0.061	0.161
		BS4	0.042	0.174	-0.111	0.305	0.164	0.199	-0.031	0.167
		repBS4	0.038	0.186	-0.112	0.311	0.168	0.203	-0.026	0.168
		KM	0.052	0.165	-0.103	0.274	0.141	0.176	-0.059	0.164
		KM4	0.038	0.176	-0.111	0.309	0.165	0.201	-0.023	0.167
		Oracle	-0.005	0.123	-0.004	0.073	0.013	0.089	0.006	0.114
	72	1stML	0.028	0.157	-0.116	0.272	0.160	0.180	0.007	0.114
		BS	0.051	0.128	-0.063	0.171	0.085	0.113	-0.075	0.134
		repBS	0.050	0.128	-0.071	0.186	0.098	0.124	-0.066	0.130
		BS4	0.050	0.128	-0.090	0.219	0.122	0.144	-0.041	0.117
		repBS4	0.050	0.126	-0.093	0.222	0.124	0.146	-0.039	0.115
		KM	0.049	0.130	-0.073	0.189	0.098	0.124	-0.064	0.130
		KM4	0.050	0.128	-0.095	0.222	0.124	0.146	-0.040	0.118
		Oracle	0.001	0.099	-0.004	0.050	0.006	0.065	-0.003	0.090
0.8	54	1stML	0.016	0.230	-0.155	0.364	0.218	0.251	0.047	0.202
		BS	0.024	0.189	-0.045	0.235	0.050	0.117	-0.112	0.205
		repBS	0.037	0.177	-0.064	0.247	0.077	0.134	-0.095	0.193
		BS4	0.007	0.226	-0.088	0.283	0.120	0.164	-0.042	0.182
		repBS4	0.020	0.202	-0.093	0.289	0.126	0.169	-0.039	0.183
		KM	0.043	0.173	-0.068	0.249	0.082	0.135	-0.098	0.198
		KM4	0.023	0.194	-0.091	0.285	0.123	0.165	-0.041	0.178
		Oracle	0.010	0.129	-0.011	0.072	0.018	0.090	-0.002	0.130
	72	1stML	0.040	0.168	-0.105	0.275	0.158	0.179	0.015	0.132
		BS	0.008	0.134	-0.019	0.169	0.012	0.074	-0.097	0.156
		repBS	0.030	0.117	-0.037	0.179	0.041	0.083	-0.080	0.142
		BS4	0.033	0.136	-0.059	0.215	0.082	0.112	-0.048	0.134
		repBS4	0.038	0.130	-0.062	0.216	0.085	0.114	-0.047	0.132
		KM	0.030	0.117	-0.040	0.179	0.042	0.084	-0.077	0.140
		KM4	0.037	0.141	-0.064	0.214	0.085	0.114	-0.051	0.131
		Oracle	0.000	0.091	-0.003	0.054	0.004	0.067	0.001	0.089

Note. 1stML: the initial ML estimator, BS: the three-step ML estimator based on the BS method with no repartitions, repBS: the repartitioned BS estimator, BS4: the BS estimator with $K^A = K^B = 4$, repBS4: the repartitioned BS4 estimator, KM: the three-step ML estimator based on the k-means method, KM4: the KM estimator with $K^A = K^B = 4$, Oracle: the oracle estimator based on the true group membership.

Table C.3: Simulation Results **DGP 3**: Estimation of Common Parameters

		α_0		$\beta_{0,1}$		$\beta_{0,2}$		ρ_0	
n	Estimator	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
54	1stML	0.035	0.243	-0.135	0.344	0.214	0.256	0.046	0.236
	BS2	-0.092	0.342	0.077	0.176	-0.123	0.153	-0.223	0.284
	BS4	-0.304	0.648	-0.053	0.253	0.079	0.143	-0.064	0.204
	repBS4	-0.308	0.655	-0.059	0.258	0.087	0.145	-0.060	0.201
	KM2	-0.103	0.354	0.076	0.175	-0.124	0.152	-0.219	0.278
	KM4	-0.284	0.632	-0.055	0.256	0.082	0.141	-0.062	0.206
72	1stML	0.037	0.165	-0.105	0.279	0.153	0.176	0.022	0.137
	BS2	-0.136	0.382	0.085	0.160	-0.141	0.154	-0.204	0.242
	BS4	-0.118	0.425	-0.037	0.216	0.047	0.090	-0.062	0.142
	repBS4	-0.097	0.397	-0.038	0.216	0.050	0.092	-0.061	0.143
	KM2	-0.144	0.395	0.085	0.160	-0.141	0.154	-0.205	0.244
	KM4	-0.095	0.391	-0.035	0.213	0.046	0.089	-0.062	0.142

Note. 1stML: the initial ML estimator, BS2: the three-step ML estimator based on the BS method with $K^A = K^B = 2$, BS4: the BS estimator with $K^A = K^B = 4$, repBS4: the repartitioned BS4 estimator, KM2: the three-step ML estimator based on the k-means method with $K^A = K^B = 2$, KM4: the KM estimator with $K^A = K^B = 4$.

Table C.4: Simulation Results: Estimation of Group Specific Effects

			Est.	BS		repBS		KM		Oracle	
r	DGP	n		Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
0.4	1	54	$a_{0,2}$	0.230	0.258			0.225	0.243	-0.001	0.077
			$b_{0,1}$	-0.267	0.335			-0.263	0.320	-0.002	0.114
			$b_{0,2}$	-0.036	0.196			-0.033	0.172	0.006	0.100
		72	$a_{0,2}$	0.157	0.170			0.156	0.170	0.002	0.057
			$b_{0,1}$	-0.190	0.228			-0.190	0.228	-0.007	0.088
			$b_{0,2}$	-0.028	0.119			-0.029	0.118	-0.004	0.073
	2	54	$a_{0,2}$	0.251	0.295	0.278	0.317	0.258	0.294	0.000	0.093
			$a_{0,3}$	0.488	0.533	0.555	0.598	0.530	0.565	0.009	0.096
			$b_{0,1}$	-0.469	0.639	-0.561	0.666	-0.555	0.619	-0.003	0.139
			$b_{0,2}$	-0.229	0.482	-0.276	0.413	-0.290	0.380	-0.003	0.127
			$b_{0,3}$	0.016	0.447	0.004	0.333	-0.017	0.261	0.000	0.114
		72	$a_{0,2}$	0.157	0.185	0.184	0.202	0.184	0.202	0.000	0.069
			$a_{0,3}$	0.334	0.383	0.387	0.429	0.377	0.398	0.001	0.071
			$b_{0,1}$	-0.346	0.492	-0.407	0.459	-0.395	0.450	-0.005	0.094
			$b_{0,2}$	-0.182	0.398	-0.218	0.315	-0.209	0.282	-0.006	0.088
			$b_{0,3}$	-0.010	0.346	-0.013	0.267	-0.019	0.189	0.002	0.077
0.8	1	54	$a_{0,2}$	0.091	0.217			0.082	0.138	0.005	0.086
			$b_{0,1}$	-0.092	0.212			-0.089	0.210	-0.010	0.130
			$b_{0,2}$	0.005	0.236			0.002	0.178	0.005	0.102
		72	$a_{0,2}$	0.042	0.083			0.041	0.084	0.008	0.064
			$b_{0,1}$	-0.061	0.139			-0.063	0.140	-0.013	0.097
			$b_{0,2}$	-0.016	0.117			-0.018	0.118	-0.007	0.077
	2	54	$a_{0,2}$	0.095	0.235	0.142	0.231	0.130	0.224	0.007	0.096
			$a_{0,3}$	0.249	0.609	0.340	0.679	0.330	0.646	0.023	0.131
			$b_{0,1}$	-0.115	0.405	-0.222	0.373	-0.266	0.401	-0.021	0.170
			$b_{0,2}$	-0.006	0.465	-0.065	0.301	-0.112	0.334	-0.011	0.135
			$b_{0,3}$	0.137	0.671	0.147	0.650	0.076	0.511	-0.004	0.117
		72	$a_{0,2}$	0.030	0.159	0.075	0.140	0.080	0.132	0.007	0.070
			$a_{0,3}$	0.055	0.242	0.151	0.273	0.162	0.273	0.010	0.088
			$b_{0,1}$	0.005	0.261	-0.109	0.218	-0.149	0.249	-0.010	0.113
			$b_{0,2}$	0.042	0.338	-0.036	0.204	-0.076	0.204	-0.007	0.090
			$b_{0,3}$	0.053	0.431	0.023	0.342	-0.014	0.157	-0.008	0.082

Note. BS: the three-step ML estimator based on the BS method with correctly chosen (K^A, K^B), repBS: the repartitioned BS estimator, KM: the three-step ML estimator based on the k-means method with correctly chosen (K^A, K^B), Oracle: the oracle estimator based on the true group membership.

Table C.5: Simulation Results: Group Membership Estimation

			Correct classification ratio			
			DGP 1		DGP 2	
r	n	Estimator	$\mathcal{C}_0^A$	$\mathcal{C}_0^B$	$\mathcal{C}_0^A$	$\mathcal{C}_0^B$
0.4	54	BS	0.679	0.680	0.527	0.529
		repBS			0.550	0.552
		KM	0.681	0.679	0.551	0.552
	72	BS	0.714	0.712	0.565	0.566
		repBS			0.602	0.602
		KM	0.713	0.711	0.601	0.604
0.8	54	BS	0.826	0.830	0.668	0.678
		repBS			0.738	0.753
		KM	0.825	0.828	0.749	0.756
	72	BS	0.875	0.870	0.735	0.731
		repBS			0.819	0.817
		KM	0.876	0.871	0.825	0.821

Note. BS: the three-step ML estimator based on the BS method with no repartitions, repBS: the repartitioned BS estimator, KM: the three-step ML estimator based on the k-means method.

C.3 A finite-mixture approach

C.3.1 An ML estimator

In the finite-mixture framework, we introduce the proportions (probabilities) of the latent groups as additional estimation parameters, rather than identifying the specific group to which each agent belongs. Here, as in Assumption 3.4(ii), let $\tau^A_{0,k}$ be the true probability of belonging to group $C^A_{0,k}$ for $k = 1, \ldots, K^A$, and $\tau^B_{0,k}$ be the probability of group $C^B_{0,k}$ for $k = 1, \ldots, K^B$. Further, by writing $\vartheta \equiv (\delta^\top, \tau^A_1, \ldots, \tau^A_{K^A}, \tau^B_1, \ldots, \tau^B_{K^B})^\top$, the log-likelihood function for the mixture model can be given by

$$\mathcal{L}^{\text{mix}}_n(\vartheta) \equiv \frac{2}{N}\sum_{i=1}^{n}\sum_{j>i}\left[y_{i,j}\ln P^{\text{mix}}_{i,j}(\vartheta) + y_{j,i}\ln P^{\text{mix}}_{j,i}(\vartheta) + (1 - y_{i,j} - y_{j,i})\ln(1 - P^{\text{mix}}_{i,j}(\vartheta) - P^{\text{mix}}_{j,i}(\vartheta))\right],$$

where

$$P^{\text{mix}}_{i,j}(\vartheta) \equiv \sum_{k_{j,2}=1}^{K^B}\sum_{k_{j,1}=1}^{K^A}\sum_{k_{i,2}=1}^{K^B}\sum_{k_{i,1}=1}^{K^A}\tau^B_{k_{j,2}}\tau^A_{k_{j,1}}\tau^B_{k_{i,2}}\tau^A_{k_{i,1}}\left[F(Z^\top_{i,j}\beta + a_{k_{i,1}} + b_{k_{j,2}}) \right.$$
$$\left. - H(Z^\top_{i,j}\beta + a_{k_{i,1}} + b_{k_{j,2}}; Z^\top_{j,i}\beta + a_{k_{j,1}} + b_{k_{i,2}} + \alpha; \rho)\right].$$

As shown above, the marginal probability for each pair of agents is represented as a mixture of $(K^AK^B)^2$ probabilities; in other words, there are potentially $(K^AK^B)^2$ different patterns for each "pair's" group membership. It is important to note that the above log-likelihood function corresponds to a model that is more general than the model in (2.1), in that the agents' sender and receiver effects may not be fixed but vary depending on their partners.

In principle, we can estimate all model parameters, including the group shares, by directly maximizing $\mathcal{L}^{\text{mix}}_n(\vartheta)$ (under the constraints $\sum_k^{K^A}\tau^A_k = 1$ and $\sum_k^{K^B}\tau^B_k = 1$). However, because the log-likelihood function is highly non-convex, there is a huge computational complexity, as numerically demonstrated below. The computational problem becomes more serious when K^A and K^B are larger than two. To mitigate such computational difficulty, alternative optimization methods, such as the EM-algorithm or Bayesian method, should be considered. The identification problem should be also a huge concern. However, these are out of scope of this study.

C.3.2 A numerical experiment

In this subsection, we conduct a small numerical analysis of the performance of the above-mentioned mixture approach. Considering the computational cost, we focus only on DGP 1 in Section 4 with $n = 54$ and $r \in \{0.4, 0.8\}$. The number of Monte Carlo repetitions is set to 500.

Table C.6 summarizes the results of the simulation. As can be observed from the table, the finite mixture method clearly underperforms compared with our proposed method. Particularly, the estimated heterogeneity parameters have a large bias; increasing the value of r (the gap between the heterogeneity parameters) does not significantly reduce the bias. After carefully inspecting the parameter estimates, we found that the estimator often falls into an extreme local solution, which could be one reason for this poor performance. Therefore, as mentioned above, it is necessary to examine the identifiability of these parameters more carefully and develop a more computationally reliable optimization procedure to improve the performance of the finite-mixture ML estimator.

Table C.6: Simulation Results: Finite-mixture Approach

		α_0		$\beta_{0,1}$		$\beta_{0,2}$		ρ_0		$a_{0,2}$		$b_{0,1}$		$b_{0,2}$	
r	Est.	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
0.4	FM	0.13	0.30	-0.21	0.33	0.31	0.48	0.22	0.29	0.56	1.24	-0.64	1.23	-0.14	1.16
	BS	0.07	0.17	-0.07	0.18	0.11	0.15	-0.08	0.17	0.23	0.26	-0.27	0.34	-0.04	0.20
0.8	FM	0.14	0.36	-0.21	0.37	0.31	0.53	0.17	0.31	0.64	1.50	-0.58	1.34	-0.12	1.34
	BS	0.06	0.16	-0.04	0.17	0.07	0.12	-0.12	0.19	0.09	0.22	-0.09	0.21	0.01	0.24

Note. FM: the finite-mixture ML estimator, BS: the three-step ML estimator based on the BS method.

C.4 Bias-corrected ML estimator

C.4.1 Derivation of the bias term

As shown in the proof of Lemma B.3, we find that $\partial_{\theta^\top}\widetilde{\gamma}_n(\theta) = -\left[n \cdot \mathcal{H}_{n,\gamma\gamma}(\theta, \widetilde{\gamma}_n(\theta))\right]^{-1} n \cdot \mathcal{H}_{n,\gamma\theta}(\theta, \widetilde{\gamma}_n(\theta))$. Moreover, by the first order condition and mean value expansion,

$$\begin{aligned}
&\mathbf{0}_{(d_z+2)\times 1} = \mathcal{S}_{n,\theta}(\widehat{\theta}_n, \widetilde{\gamma}_n(\widehat{\theta}_n)) = \mathcal{S}_{n,\theta}(\theta_0, \widetilde{\gamma}_n(\theta_0)) + \mathcal{I}_{n,\theta\theta}(\bar{\theta}_n, \widetilde{\gamma}_n(\bar{\theta}_n))[\widehat{\theta}_n - \theta_0] \\
&\Longrightarrow \widehat{\theta}_n - \theta_0 = \left[\mathcal{I}_{n,\theta\theta}(\bar{\theta}_n, \widetilde{\gamma}_n(\bar{\theta}_n))\right]^{-1} \mathcal{S}_{n,\theta}(\theta_0, \widetilde{\gamma}_n(\theta_0)).
\end{aligned}$$

where $\bar{\theta}_n \in [\widehat{\theta}_n, \theta_0]$. Therefore, letting $\mathcal{I}_* \equiv \text{plim}_{n\to\infty}\mathcal{I}_{n,\theta\theta}(\theta_0, \boldsymbol{\gamma}_0)$, we have

$$\sqrt{\frac{N}{2}}(\widehat{\theta}_n - \theta_0) = \mathcal{I}_*^{-1}\left[\sqrt{\frac{2}{N}}\sum_{i=1}^{n}\sum_{j>i} s_{i,j}^{\theta}(\theta_0, \widetilde{\gamma}_n(\theta_0))\right] + o_P(1)$$

In the following, we follow the proof of Theorem 4 in Graham (2017). Specifically, by writing $\boldsymbol{\gamma}_{-1} = (\gamma_2, \gamma_3, \ldots, \gamma_{2n})^\top$ and suppressing the dependence of $\widetilde{\gamma}_n(\theta_0)$ on θ_0, a third-order Taylor expansion yields

$$\begin{aligned}
\sqrt{\frac{2}{N}}\sum_{i=1}^{n}\sum_{j>i} s_{i,j}^{\theta}(\theta_0, \widetilde{\gamma}_n) &= \sqrt{\frac{2}{N}}\sum_{i=1}^{n}\sum_{j>i}\left\{s_{i,j}^{\theta}(\theta_0, \boldsymbol{\gamma}_0) + \partial_{\boldsymbol{\gamma}_{-1}^\top} s_{i,j}^{\theta}(\theta_0, \boldsymbol{\gamma}_0)[\widetilde{\boldsymbol{\gamma}}_{n,-1} - \boldsymbol{\gamma}_{0,-1}]\right\} \\
&+ \frac{1}{2}\sqrt{\frac{2}{N}}\sum_{k=2}^{2n}[\widetilde{\gamma}_{n,k} - \gamma_{0,k}]\sum_{i=1}^{n}\sum_{j>i}\partial^2_{\gamma_k\boldsymbol{\gamma}_{-1}^\top} s_{i,j}^{\theta}(\theta_0, \boldsymbol{\gamma}_0)[\widetilde{\boldsymbol{\gamma}}_{n,-1} - \boldsymbol{\gamma}_{0,-1}] \\
&+ \frac{1}{6}\sqrt{\frac{2}{N}}\sum_{l=2}^{2n}\sum_{k=2}^{2n}[\widetilde{\gamma}_{n,l} - \gamma_{0,l}][\widetilde{\gamma}_{n,k} - \gamma_{0,k}]\sum_{i=1}^{n}\sum_{j>i}\partial^3_{\gamma_l\gamma_k\boldsymbol{\gamma}_{-1}^\top} s_{i,j}^{\theta}(\theta_0, \bar{\boldsymbol{\gamma}}_n)[\widetilde{\boldsymbol{\gamma}}_{n,-1} - \boldsymbol{\gamma}_{0,-1}],
\end{aligned} \tag{C.1}$$

where $\bar{\boldsymbol{\gamma}}_n \in [\widetilde{\boldsymbol{\gamma}}_n(\theta_0), \boldsymbol{\gamma}_0]$. First, after tedious calculations, we can show that the third term of (C.1) is of order $O_P\left(\frac{(\ln n)^{3/2}}{\sqrt{n}}\right) = o_P(1)$ by using Lemma B.2(iv). Next, the second term is a non-vanishing bias term. We define the asymptotic bias as

$$\text{bias}_0 \equiv \underset{n\to\infty}{\text{plim}}\sqrt{\frac{1}{2N}}\sum_{k=2}^{2n}[\widetilde{\gamma}_{n,k} - \gamma_{0,k}]\sum_{i=1}^{n}\sum_{j>i}\partial^2_{\gamma_k\boldsymbol{\gamma}_{-1}^\top} s_{i,j}^{\theta}(\theta_0, \boldsymbol{\gamma}_0)[\widetilde{\boldsymbol{\gamma}}_{n,-1} - \boldsymbol{\gamma}_{0,-1}]. \tag{C.2}$$

Lastly, the first term of (C.1) is the term to which a central limit theorem is applied. By integrating these results, we obtain

$$\sqrt{\frac{N}{2}}(\widehat{\theta}_n - \theta_0) = \mathcal{I}_*^{-1}\text{bias}_0 + \mathcal{I}_*^{-1}\left[\sqrt{\frac{2}{N}}\sum_{i=1}^{n}\sum_{j>i}\left\{s_{i,j}^{\theta}(\theta_0,\gamma_0) + \partial_{\gamma_{-1}^\top}s_{i,j}^{\theta}(\theta_0,\gamma_0)[\widetilde{\gamma}_{n,-1} - \gamma_{0,-1}]\right\}\right] + o_P(1).$$

Then, $\sqrt{\frac{N}{2}}(\widehat{\theta}_n - \theta_0)$ is asymptotically normally distributed with mean $\mathcal{I}_*^{-1}\text{bias}_0$ and covariance matrix $\mathcal{I}_*^{-1}\mathcal{V}\mathcal{I}_*^{-1}$, where $\mathcal{V}$ is the asymptotic variance of the first term in (C.1). Unfortunately, because our model is more complicated than those in Graham (2017) and Yan *et al.* (2019), it would be difficult to derive an analytically simpler form for the bias term.

C.4.2 A numerical experiment

In this subsection, we conduct a small numerical analysis of the bias-corrected ML estimator. As stated above, computing the bias term directly from definition (C.2) is difficult. Thus, we alternatively consider using a bootstrap method to estimate the bias in a similar way to Kim and Sun (2016). That is, letting T be the number of bootstrap iterations, we compute $\widehat{\text{bias}}_{n,T} \equiv \frac{1}{T}\sum_{t=1}^{T}(\widehat{\theta}_n^{(t)} - \widehat{\theta}_n)$ as a proxy for $\sqrt{\frac{2}{N}}\mathcal{I}_*^{-1}\text{bias}_0$, and define the bias-corrected ML estimator as $\widehat{\theta}_n^{bc} \equiv \widehat{\theta}_n - \widehat{\text{bias}}_{n,T}$, where $\widehat{\theta}_n^{(t)}$ means that it is obtained from the t-th bootstrap sample.

In Table C.7, we report the performance of the bias-corrected and initial ML estimators for GDP 1 in Section 4 with $n \in \{54, 72\}$ and $r = 0.4$, except that we fix $\rho_0 = 0$ to speed up the computation. To further reduce the computational burden, we set $T = 50$ only, and the number of Monte Carlo repetitions to 200. The results show that although the number of bootstrap iterations is not large, the bias-corrected ML estimator works satisfactorily. Therefore, if only the common parameters are of interest, using this estimator would be a reasonable alternative.

Table C.7: Simulation Results: Bias-Corrected ML Estimator

		α_0		$\beta_{0,1}$		$\beta_{0,2}$	
n	Estimator	Bias	RMSE	Bias	RMSE	Bias	RMSE
54	1stML	-0.010	0.080	-0.132	0.326	0.190	0.207
	BCML	0.058	0.096	0.025	0.268	-0.040	0.081
72	1stML	-0.005	0.058	-0.060	0.251	0.126	0.138
	BCML	0.036	0.066	0.036	0.227	-0.021	0.053

Note. 1stML: the initial ML estimator, BCML: the bias-corrected ML estimator.

C.5 Supplementary tables for Section 5

Table C.8: In-degree and Out-degree

Country	In-degree	Out-degree	(cont')			(cont')		
Armenia	21	46						
Australia	43	6	Japan	48	14	Papua New Guinea	18	21
Azerbaijan	24	24	Jordan	13	41	Philippines	22	38
Bahrain	33	22	Kazakhstan	29	34	Qatar	29	31
Bangladesh	6	39	Kiribati	22	15	Russia	37	18
Belarus	26	33	Kuwait	35	23	Saudi Arabia	31	19
Bhutan	15	2	Kyrgyzstan	24	33	Singapore	47	34
Brunei	43	26	Laos	17	49	South Korea	49	25
Cambodia	15	54	Latvia	40	19	Sri Lanka	7	53
China	24	9	Lebanon	13	34	Tajikistan	23	36
Cyprus	41	19	Lithuania	31	18	Thailand	31	37
Estonia	42	19	Malaysia	46	46	Tonga	23	21
Fiji	23	29	Moldova	26	28	Turkey	35	27
Georgia	29	32	Mongolia	21	20	UAE	44	20
Hong Kong	43	33	Myanmar	13	16	Ukraine	36	29
India	13	3	Nauru	21	7	Uzbekistan	22	28
Indonesia	25	51	Nepal	9	55	Vanuatu	27	33
Iran	10	48	New Zealand	45	18	Viet Nam	16	14
Iraq	5	5	Oman	34	5	Yemen	7	9
Israel	33	21	Pakistan	5	21			

Table C.9: Summary Statistics

	Mean	Std. Dev.	Min.	Max.	# Observations
gdp_pc_i	14.4409	17.3406	0.8071	67.5714	57
$free_i$	4.0789	1.9291	1	7	57
$export_{ij}$	2.8562	2.9874	0	12.6580	3,192
$import_{ij}$	2.9219	2.9773	0	12.5461	3,192

Table C.10: BIC Model Selection

		K^B					
		2	3	4	5	6	7
K^A	2	2276.659	2078.159	2122.551	2102.056	2104.882	2117.091
	3	2033.604	1935.613	1909.670	1903.809	1899.412	1909.794
	4	1985.267	1883.387	1837.424	1834.551	1823.058	1848.088
	5	1931.314	1837.387	1798.743	1789.363	1787.481	1795.286
	6	1918.471	1808.645	1775.658	1769.636	1765.887	1768.170
	7	1897.173	1807.099	1777.586	1766.433	1761.725	1764.802

References

Billingsley, P., 2012. *Probability and Measure: Anniversary Edition*, John Wiley & Sons.

Graham, B.S., 2017. An econometric model of network formation with degree heterogeneity, *Econometrica*, 85 (4), 1033–1063.

Kim, M.S. and Sun, Y., 2016. Bootstrap and k-step bootstrap bias corrections for the fixed effects estimator in nonlinear panel data models, *Econometric Theory*, 32 (6), 1523–1568.

Kreps, D.M., 2012. *Microeconomic Foundations I: Choice and Competitive Markets*, Princeton University Press.

Wang, W. and Su, L., 2020. Identifying latent group structures in nonlinear panels, *Journal of Econometrics*.

Yan, T., Jiang, B., Fienberg, S.E., and Leng, C., 2019. Statistical inference in a directed network model with covariates, *Journal of the American Statistical Association*, 114 (526), 857–868.