Supplementary Material for
"Two-Step Estimation of Incomplete Information Social Interaction Models with Sample Selection"

Tadao Hoshino
School of Management, Tokyo University of Science
E-mail: thoshino@rs.tus.ac.jp

(Not for publication)

1 Proofs of Lemmas

Proof of Lemma A.1. (i) By the mean value expansion,

$$\begin{aligned}\left\|\widehat{P}_n^{L\prime}D_n/\sqrt{n}-P_n^{L\prime}D_n/\sqrt{n}\right\| &= \left\|\dot{P}_n^L\left(\overline{V}_n\right)'\operatorname{diag}\left[W_n\left(\widehat{\gamma}_n-\gamma_0\right)\right]D_n/\sqrt{n}\right\| \\ &= O_P(1)\cdot\left\|\dot{P}_n^L\left(\overline{V}_n\right)'D_n/n\right\| \\ &= O_P(\zeta_1(L)/\sqrt{n})=o_P(1)\end{aligned}$$

as $n\to\infty$ by Assumptions 2.2 and 5.2, where $\dot{P}_n^L\left(\overline{V}_n\right)=\left(\partial_v p^L(\bar{v}_1),...,\partial_v p^L(\bar{v}_n)\right)'$ with $\overline{V}_n=(\bar{v}_1,...,\bar{v}_n)'$, and $\bar{v}_i\in[\widehat{v}_i,v_i]$ for each $i=1,...,n$.

(ii) We obtain, by the independence,

$$\begin{aligned}\mathrm{E}\left\|P_n^{L\prime}D_nP_n^L/n-P_n^{L\prime}\overline{D}_nP_n^L/n\right\|^2 &= \mathrm{E}\left\|n^{-1}\sum_{i=1}^n p^L(v_i)p^L(v_i)'\left(d_i-\mathrm{E}\left[d_i\right]\right)\right\|^2 \\ &= n^{-2}\sum_{k=1}^L\sum_{j=1}^L\mathrm{E}\left(\sum_{i=1}^n p_j(v_i)p_k(v_i)\left(d_i-\mathrm{E}[d_i]\right)\right)^2 \\ &= n^{-2}\sum_{k=1}^L\sum_{j=1}^L\sum_{i=1}^n\sum_{\ell=1}^n p_j(v_i)p_k(v_i)p_j(v_\ell)p_k(v_\ell)\mathrm{E}\left[\left(d_i-\mathrm{E}\left[d_i\right]\right)\left(d_\ell-\mathrm{E}\left[d_\ell\right]\right)\right] \\ &= n^{-2}\sum_{k=1}^L\sum_{j=1}^L\sum_{i=1}^n p_j^2(v_i)p_k^2(v_i)\mathrm{E}\left[d_i\right]\left(1-\mathrm{E}\left[d_i\right]\right) \\ &\leq n^{-2}\sum_{k=1}^L\sum_{j=1}^L\sum_{i=1}^n p_j^2(v_i)p_k^2(v_i) \\ &= O(\zeta_0^2(L)L/n)=o(1).\end{aligned}$$

Therefore, the result follows from Markov's inequality with Assumption 5.2.

(iii) Using the same argument as in the proof of Lemma A.1 in Su and Jin (2012), we obtain

$$\begin{aligned}\tau_{\max}\left(P_n^{L\prime}D_nP_n^L/n\right) &= \max_{\|\varkappa\|=1}\left\{\varkappa'\left(P_n^{L\prime}\overline{D}_nP_n^L/n\right)\varkappa+\varkappa'\left(P_n^{L\prime}D_nP_n^L/n-P_n^{L\prime}\overline{D}_nP_n^L/n\right)\varkappa\right\} \\ &\leq \tau_{\max}\left(P_n^{L\prime}\overline{D}_nP_n^L/n\right)+\left\|P_n^{L\prime}D_nP_n^L/n-P_n^{L\prime}\overline{D}_nP_n^L/n\right\| \\ &\leq \bar{c}_P+o_P(1)\end{aligned}$$

and

$$\begin{aligned}\tau_{\min}\left(P_n^{L\prime}D_nP_n^L/n\right) &= \min_{\|\varkappa\|=1}\left\{\varkappa'\left(P_n^{L\prime}\overline{D}_nP_n^L/n\right)\varkappa+\varkappa'\left(P_n^{L\prime}D_nP_n^L/n-P_n^{L\prime}\overline{D}_nP_n^L/n\right)\varkappa\right\} \\ &\geq \tau_{\min}\left(P_n^{L\prime}\overline{D}_nP_n^L/n\right)-\left\|P_n^{L\prime}D_nP_n^L/n-P_n^{L\prime}\overline{D}_nP_n^L/n\right\| \\ &\geq \underline{c}_d\underline{c}_P-o_P(1)\end{aligned}$$

for sufficiently large n by Assumption 6 and the result (ii).

(iv) By the triangle inequality,

$$\begin{aligned}\left\|\widehat{P}_n^{L\prime}D_n\widehat{P}_n^L/n - P_n^{L\prime}D_nP_n^L/n\right\| &\leq \left\|\widehat{P}_n^{L\prime}D_n/\sqrt{n} - P_n^{L\prime}D_n/\sqrt{n}\right\|^2 + 2\left\|\left[\widehat{P}_n^{L\prime}D_n - P_n^{L\prime}D_n\right]P_n^L/n\right\| \\ &\leq 2\left\|\left[\widehat{P}_n^{L\prime}D_n - P_n^{L\prime}D_n\right]P_n^L/n\right\| + O_P(\zeta_1^2(L)/n)\end{aligned}$$

by the result (i). Furthermore, by Assumption 6.2 and (i),

$$\begin{aligned}&\left\|\left[\widehat{P}_n^{L\prime}D_n - P_n^{L\prime}D_n\right]P_n^L/n\right\|^2 \\ =\ & \mathrm{tr}\left\{\left[\widehat{P}_n^{L\prime}D_n/\sqrt{n} - P_n^{L\prime}D_n/\sqrt{n}\right]\left(P_n^LP_n^{L\prime}/n\right)\left[\widehat{P}_n^{L\prime}D_n/\sqrt{n} - P_n^{L\prime}D_n/\sqrt{n}\right]'\right\} \\ \leq\ & \bar{c}_P\left\|\widehat{P}_n^{L\prime}D_n/\sqrt{n} - P_n^{L\prime}D_n/\sqrt{n}\right\|^2 = O_P(\zeta_1^2(L)/n)\end{aligned}$$

as $n \to \infty$. This implies (iv).

The proof of (v) is similar to that of (iii) with the use of (iv), and is omitted.

(vi) Denote the ith element of $H_n(\theta)$ as $h_i(\theta)$. Because matrix $S_n(\lambda)$ is uniformly bounded in absolute value in row sums for any $\lambda \in \mathcal{L} \subset (-1,1)$ (see, e.g., Lemma A.2 in Lee, 2004), the $h_i(\theta)$'s are also uniformly bounded in absolute value for any $\theta \in \mathcal{L} \times \mathcal{B}$. Then, we obtain

$$\sup_{\theta\in\mathcal{L}\times\mathcal{B}}\left\|D_nU_n(\theta)/\sqrt{n}\right\| \leq \sup_{\theta\in\mathcal{L}\times\mathcal{B}}\left\|D_nH_n(\theta)/\sqrt{n}\right\| + \left\|D_n\mathcal{E}_n/\sqrt{n}\right\| = O_P(1). \tag{1.1}$$

Hence,

$$\sup_{\theta\in\mathcal{L}\times\mathcal{B}}\left\|\left(P_n^{L\prime}D_n - \widehat{P}_n^{L\prime}D_n\right)U_n(\theta)/n\right\| \leq \left\|\widehat{P}_n^{L\prime}D_n/\sqrt{n} - P_n^{L\prime}D_n/\sqrt{n}\right\| \cdot \sup_{\theta\in\mathcal{L}\times\mathcal{B}}\left\|D_nU_n(\theta)/\sqrt{n}\right\| = o_P(1)$$

by (i).

(vii) Noting that the eigenvalues of an idempotent matrix are at most one, for sufficiently large n,

$$\begin{aligned}\sup_{\theta\in\mathcal{L}\times\mathcal{B}}\left\|U_n(\theta)'\mathcal{A}_n\right\| &= \sup_{\theta\in\mathcal{L}\times\mathcal{B}}\left\{U_n(\theta)'D_nP_n^L\left[P_n^{L\prime}D_nP_n^L\right]^{-2}P_n^{L\prime}D_nU_n(\theta)\right\}^{1/2} \\ &\leq \left[\underline{c}_d\underline{c}_P - o_P(1)\right]^{-1/2}\sup_{\theta\in\mathcal{L}\times\mathcal{B}}\left\|D_nU_n(\theta)/\sqrt{n}\right\| = O_P(1)\end{aligned}$$

by (iii) and (1.1). The proof of (viii) is similar to that of (vii) with the use of (v) and (1.1), and is omitted. ■

Proof of Lemma A.2. Write

$$Q_n(\theta) - Q_n^*(\theta) = Q_n(\theta) - Q_{n,1}(\theta) + Q_{n,1}(\theta) - Q_n^*(\theta),$$

where

$$Q_{n,1}(\theta) = \frac{1}{2n}U_n(\theta)'D_n\left[I_n - \mathcal{P}_n\right]D_nU_n(\theta).$$

First, observe that

$$\begin{aligned}\left|Q_n(\theta) - Q_{n,1}(\theta)\right| &= \left|U_n(\theta)'\mathcal{P}_nU_n(\theta)/(2n) - U_n(\theta)'\widehat{\mathcal{P}}_nU_n(\theta)/(2n)\right| \\ &= \left|U_n(\theta)'\mathcal{A}_nP_n^{L\prime}D_nU_n(\theta)/(2n) - U_n(\theta)'\widehat{\mathcal{A}}_n\widehat{P}_n^{L\prime}D_nU_n(\theta)/(2n)\right| \\ &\leq \left|U_n(\theta)'\mathcal{A}_n\left(\widehat{P}_n^{L\prime}D_n\widehat{P}_n^L - P_n^{L\prime}D_nP_n^L\right)\widehat{\mathcal{A}}_n'U_n(\theta)/(2n)\right| \\ &\quad + \left|U_n(\theta)'\mathcal{A}_n\left(P_n^{L\prime}D_n - \widehat{P}_n^{L\prime}D_n\right)U_n(\theta)/(2n)\right| + \left|U_n(\theta)'\widehat{\mathcal{A}}_n\left(\widehat{P}_n^{L\prime}D_n - P_n^{L\prime}D_n\right)U_n(\theta)/(2n)\right| \\ &\leq \left\|U_n(\theta)'\mathcal{A}_n\right\|\left\|\widehat{P}_n^{L\prime}D_n\widehat{P}_n^L/n - P_n^{L\prime}D_nP_n^L/n\right\|\left\|\widehat{\mathcal{A}}_n'U_n(\theta)\right\| \\ &\quad + \left(\left\|U_n(\theta)'\mathcal{A}_n\right\| + \left\|U_n(\theta)'\widehat{\mathcal{A}}_n\right\|\right)\left\|\left(P_n^{L\prime}D_n - \widehat{P}_n^{L\prime}D_n\right)U_n(\theta)/n\right\| = o_P(1)\end{aligned} \tag{1.2}$$

uniformly in $\theta \in \mathcal{L} \times \mathcal{B}$ by Lemma A.1 (iv), (vi), (vii), and (viii).

Next, let

$$\begin{aligned}
\delta_{n1}(\theta) &\equiv \frac{1}{n}\left[U_n(\theta)' D_n U_n(\theta) - \mathrm{E}\left(U_n(\theta)' D_n U_n(\theta)\right)\right] \\
\delta_{n2}(\theta) &\equiv \frac{1}{n}\left[U_n(\theta)' \mathcal{P}_n U_n(\theta) - \mathrm{E}\left(U_n(\theta)' D_n P_n^L\right)\left(P_n^{L\prime} \overline{D}_n P_n^L\right)^{-1} \mathrm{E}\left(P_n^{L\prime} D_n U_n(\theta)\right)\right]
\end{aligned}$$

so that

$$Q_{n,1}(\theta) - Q_n^*(\theta) = \frac{1}{2}\delta_{n1}(\theta) - \frac{1}{2}\delta_{n2}(\theta).$$

Below, we show that $\delta_{n1}(\theta) = o_P(1)$ uniformly in $\theta \in \mathcal{L} \times \mathcal{B}$:

$$\begin{aligned}
\delta_{n1}(\theta) &= H_n(\theta)' \left\{D_n - \mathrm{E}\left[D_n\right]\right\} H_n(\theta)/n + 2H_n(\theta)' \left\{D_n \mathcal{E}_n - \mathrm{E}\left[D_n \mathcal{E}_n\right]\right\}/n + \left(\mathcal{E}_n' D_n \mathcal{E}_n/n - \mathrm{E}\left[\mathcal{E}_n' D_n \mathcal{E}_n/n\right]\right) \\
&= n^{-1}\sum_{i=1}^n h_i^2(\theta)\left(d_i - \mathrm{E}[d_i]\right) + 2n^{-1}\sum_{i=1}^n h_i(\theta)\left(d_i \varepsilon_i - \mathrm{E}\left[d_i \varepsilon_i\right]\right) + \left(\mathcal{E}_n' D_n \mathcal{E}_n/n - \mathrm{E}\left[\mathcal{E}_n' D_n \mathcal{E}_n/n\right]\right) \\
&= T_{n1}(\theta) + T_{n2}(\theta) + T_{n3}, \text{ say.}
\end{aligned}$$

Recalling that $h_i(\theta)$'s are uniformly bounded in absolute value for any $\theta \in \mathcal{L} \times \mathcal{B}$, it is straightforward to see $T_{n1}(\theta) = o_P(1)$ and $T_{n2}(\theta) = o_P(1)$ uniformly in $\theta \in \mathcal{L} \times \mathcal{B}$. In addition, $T_{n3} = o_P(1)$ can be shown by Chebyshev's inequality. Specifically, observe that, by the independence,

$$\begin{aligned}
\mathrm{E}\left[\mathcal{E}_n' D_n \mathcal{E}_n\right] &= \mathrm{E}\left[\sum_{i=1}^n d_i \varepsilon_i^2\right] = \sum_{i=1}^n \mathrm{E}[d_i]\sigma_i^2 \\
\mathrm{E}\left[\left(\mathcal{E}_n' D_n \mathcal{E}_n\right)^2\right] &= \mathrm{E}\left[\sum_{i=1}^n d_i \varepsilon_i^4 + \sum_{i=1}^n \sum_{j \neq i}^n d_i \varepsilon_i^2 d_j \varepsilon_j^2\right] \\
&= \mathrm{tr}\left(\mathrm{E}\left[D_n\right]\Sigma_{n,4}\right) + \sum_{i=1}^n \sum_{j \neq i}^n \mathrm{E}\left[d_i\right]\mathrm{E}[d_j]\sigma_i^2 \sigma_j^2,
\end{aligned}$$

where $\Sigma_{n,4} = \mathrm{diag}\left(\mathrm{E}\left[\varepsilon_i^4 | d_i = 1\right], ..., \mathrm{E}\left[\varepsilon_j^4 | d_j = 1\right]\right)$, which exists by Assumption 3.1. Therefore,

$$\begin{aligned}
\mathrm{Var}\left(\mathcal{E}_n' D_n \mathcal{E}_n - \mathrm{E}\left[\mathcal{E}_n' D_n \mathcal{E}_n\right]\right) &= \mathrm{E}\left[\left(\mathcal{E}_n' D_n \mathcal{E}_n\right)^2\right] - \left(\mathrm{E}\left[\mathcal{E}_n' D_n \mathcal{E}_n\right]\right)^2 \\
&= \mathrm{tr}\left(\mathrm{E}\left[D_n\right]\Sigma_{n,4}\right) + \sum_{i=1}^n \sum_{j \neq i}^n \mathrm{E}\left[d_i\right]\mathrm{E}\left[d_j\right]\sigma_i^2\sigma_j^2 - \sum_{i=1}^n \sum_{j=1}^n \mathrm{E}\left[d_i\right]\mathrm{E}\left[d_j\right]\sigma_i^2\sigma_j^2 \\
&= \mathrm{tr}\left(\mathrm{E}\left[D_n\right]\Sigma_{n,4} - \mathrm{E}\left[D_n\right]^2 \Sigma_n^2\right) = O(n),
\end{aligned}$$

where $\Sigma_n \equiv \mathrm{diag}(\sigma_1^2, ..., \sigma_n^2)$. Thus, we obtain $\mathrm{Var}\left(\mathcal{E}_n' D_n \mathcal{E}_n/n - \mathrm{E}\left[\mathcal{E}_n' D_n \mathcal{E}_n/n\right]\right) = o(1)$.

Finally, we prove $\delta_{n2}(\theta) = o_P(1)$ uniformly in $\theta \in \mathcal{L} \times \mathcal{B}$ by showing that

$$\begin{aligned}
\left\|\left(P_n^{L\prime} D_n P_n^L/n\right)^{-1} - \left(P_n^{L\prime} \overline{D}_n P_n^L/n\right)^{-1}\right\| &= o_P(1) \text{ and} \\
\left\|P_n^{L\prime} D_n U_n(\theta)/n - \mathrm{E}\left(P_n^{L\prime} D_n U_n(\theta)/n\right)\right\| &= o_P(1) \text{ uniformly in } \theta \in \mathcal{L} \times \mathcal{B}.
\end{aligned}$$

For the first part, we have already shown that $\left\|P_n^{L\prime} D_n P_n^L/n - P_n^{L\prime} \overline{D}_n P_n^L/n\right\| = o_P(1)$ in Lemma A.1 (ii), which implies the result because

$$\begin{aligned}
&\left\|\left(P_n^{L\prime} D_n P_n^L/n\right)^{-1} - \left(P_n^{L\prime} \overline{D}_n P_n^L/n\right)^{-1}\right\| \qquad (1.3) \\
= &\left\|\left(P_n^{L\prime} D_n P_n^L/n\right)^{-1}\left(P_n^{L\prime} \overline{D}_n P_n^L/n - P_n^{L\prime} D_n P_n^L/n\right)\left(P_n^{L\prime} \overline{D}_n P_n^L/n\right)^{-1}\right\| \\
\leq &\; O_P(1) \cdot \left\|P_n^{L\prime} \overline{D}_n P_n^L/n - P_n^{L\prime} D_n P_n^L/n\right\| = o_P(1)
\end{aligned}$$

by Lemma A.1 (ii) and (iii).

For the second part, by the triangle inequality,

$$\left\| P_n^{L\prime} D_n U_n(\theta)/n - \mathrm{E}\left(P_n^{L\prime} D_n U_n(\theta)/n\right)\right\| \leq \left\| n^{-1}\sum_{i=1}^{n} p^L(v_i)\left(d_i\varepsilon_i - \mathrm{E}\left[d_i\varepsilon_i\right]\right)\right\| + \left\| n^{-1}\sum_{i=1}^{n} p^L(v_i) h_i(\theta)\left(d_i - \mathrm{E}[d_i]\right)\right\|.$$

Observe that, by the independence,

$$\begin{aligned} \mathrm{E}\left\| n^{-1}\sum_{i=1}^{n} p^L(v_i)\left(d_i\varepsilon_i - \mathrm{E}\left[d_i\varepsilon_i\right]\right)\right\|^2 &= n^{-2}\sum_{k=1}^{L}\mathrm{E}\left(\sum_{i=1}^{n} p_k(v_i)\left(d_i\varepsilon_i - \mathrm{E}\left[d_i\varepsilon_i\right]\right)\right)^2 \\ &\leq cn^{-2}\sum_{k=1}^{L}\sum_{i=1}^{n} p_k^2(v_i) = O\left(L/n\right) = o(1). \end{aligned} \tag{1.4}$$

Then, it follows that $\left\| n^{-1}\sum_{i=1}^{n} p^L(v_i)\left(d_i\varepsilon_i - \mathrm{E}\left[d_i\varepsilon_i\right]\right)\right\| = o_P(1)$ by Markov's inequality. Similarly, we can show that $\left\| n^{-1}\sum_{i=1}^{n} p^L(v_i)h_i(\theta)\left(d_i - \mathrm{E}\left[d_i\right]\right)\right\| = o_P(1)$ uniformly in $\theta \in \mathcal{L}\times\mathcal{B}$ by the fact that the $h_i(\theta)$'s are uniformly bounded in absolute value for any $\theta \in \mathcal{L}\times\mathcal{B}$. ■

Proof of Lemma A.3. Because $\Gamma_n'\left[I_n - \widehat{\mathcal{P}}_n\right] D_n \widehat{P}_n^L B_{0L} = \mathbf{0}_{(k_1+k_2+2)}$ and $I_n - \widehat{\mathcal{P}}_n$ is idempotent,

$$\begin{aligned} & \left\|\Gamma_n'\left[I_n - \widehat{\mathcal{P}}_n\right] D_n \Pi_0(\widehat{V}_n)/\sqrt{n}\right\|^2 \\ = & \left\|\Gamma_n'\left[I_n - \widehat{\mathcal{P}}_n\right] D_n\left(\Pi_0(\widehat{V}_n) - \widehat{P}_n^L B_{0L}\right)/\sqrt{n}\right\|^2 \\ = & \left(\Pi_0(\widehat{V}_n) - \widehat{P}_n^L B_{0L}\right)' D_n\left[I_n - \widehat{\mathcal{P}}_n\right]\left(\Gamma_n\Gamma_n'/n\right)\left[I_n - \widehat{\mathcal{P}}_n\right] D_n\left(\Pi_0(\widehat{V}_n) - \widehat{P}_n^L B_{0L}\right) \\ \leq & c\left\|\Pi_0(\widehat{V}_n) - \widehat{P}_n^L B_{0L}\right\|^2 = O(nL^{-2\rho}) \end{aligned}$$

as $n \to \infty$. Thus, $\left\|\Gamma_n'\left[I_n - \widehat{\mathcal{P}}_n\right] D_n\Pi_0(\widehat{V}_n)/\sqrt{n}\right\| = O(n^{1/2}L^{-\rho}) = o(1)$ by Assumption 5.1. ■

Proof of Lemma A.4. Write

$$\begin{aligned} & \Gamma_n'\left[I_n - \widehat{\mathcal{P}}_n\right] D_n\widetilde{\mathcal{E}}_n/\sqrt{n} - \Gamma_n^{\perp\prime} D_n\widetilde{\mathcal{E}}_n/\sqrt{n} \\ = & \Gamma_n'\left[\mathcal{P}_n - \widehat{\mathcal{P}}_n\right] D_n\widetilde{\mathcal{E}}_n/\sqrt{n} + \left\{\Gamma_n'\overline{D}_n P_n^L\left(P_n^{L\prime}\overline{D}_n P_n^L\right)^{-1} P_n^{L\prime} D_n\widetilde{\mathcal{E}}_n/\sqrt{n} - \Gamma_n'\mathcal{P}_n D_n\widetilde{\mathcal{E}}_n/\sqrt{n}\right\} \\ = & R_{n1} + R_{n2}, \text{ say.} \end{aligned}$$

Observe that

$$\begin{aligned} \mathrm{E}\left[\left\|R_{n1}\right\|^2|D_n\right] &= n^{-1}\mathrm{tr}\left(\Gamma_n'\left[\mathcal{P}_n - \widehat{\mathcal{P}}_n\right] D_n\mathrm{E}\left[\widetilde{\mathcal{E}}_n\widetilde{\mathcal{E}}_n'|D_n\right] D_n\left[\mathcal{P}_n - \widehat{\mathcal{P}}_n\right]\Gamma_n\right) \\ &\leq \bar{c}_{\varepsilon,2} n^{-1}\mathrm{tr}\left(\Gamma_n'\left[\mathcal{P}_n - \widehat{\mathcal{P}}_n\right]\left[\mathcal{P}_n - \widehat{\mathcal{P}}_n\right]\Gamma_n\right). \end{aligned}$$

Using the same argument as in (1.2), we can show both $\left\|n^{-1}\Gamma_n'\left[\mathcal{P}_n - \widehat{\mathcal{P}}_n\right]\mathcal{P}_n\Gamma_n\right\|$ and $\left\|n^{-1}\Gamma_n'\left[\mathcal{P}_n - \widehat{\mathcal{P}}_n\right]\widehat{\mathcal{P}}_n\Gamma_n\right\|$ are of order $o_P(1)$. Thus, $n^{-1}\mathrm{tr}\left(\Gamma_n'\left[\mathcal{P}_n - \widehat{\mathcal{P}}_n\right]\left[\mathcal{P}_n - \widehat{\mathcal{P}}_n\right]\Gamma_n\right) = o_P(1)$, implying that $\left\|R_{n1}\right\| = o_P(1)$ by Markov's inequality.

Next, by the triangle inequality,

$$\begin{aligned} \left\|R_{n2}\right\| \leq & \left\|\left(\Gamma_n'\overline{D}_n P_n^L - \Gamma_n' D_n P_n^L\right)\left(P_n^{L\prime}\overline{D}_n P_n^L\right)^{-1} P_n^{L\prime} D_n\widetilde{\mathcal{E}}_n/\sqrt{n}\right\| \\ & + \left\|\Gamma_n' D_n P_n^L\left[\left(P_n^{L\prime}\overline{D}_n P_n^L\right)^{-1} - \left(P_n^{L\prime} D_n P_n^L\right)^{-1}\right] P_n^{L\prime} D_n\widetilde{\mathcal{E}}_n/\sqrt{n}\right\| \\ = & \left\|R_{n21}\right\| + \left\|R_{n22}\right\|, \text{ say.} \end{aligned}$$

Noting that $\left\|\Gamma_n' D_n P_n^L/n - \Gamma_n' \overline{D}_n P_n^L/n\right\| = o_P(1)$ by a similar argument in (1.4),

$$\mathrm{E}\left[\|R_{n21}\|^2 \,|D_n\right] \leq c \cdot \left\|\Gamma_n' D_n P_n^L/n - \Gamma_n' \overline{D}_n P_n^L/n\right\|^2 = o_P(1)$$

for sufficiently large n. In addition, because we have already shown that $\left\|\left(P_n^{L\prime} D_n P_n^L\right)^{-1} - \left(P_n^{L\prime} \overline{D}_n P_n^L\right)^{-1}\right\| = o_P(1)$ in (1.3), similarly

$$\mathrm{E}\left[\|R_{n22}\|^2 \,|D_n\right] \leq c \cdot \left\|\left(P_n^{L\prime} \overline{D}_n P_n^L/n\right)^{-1} - \left(P_n^{L\prime} D_n P_n^L/n\right)^{-1}\right\|^2 = o_P(1)$$

for sufficiently large n. Finally, by Markov's inequality, $\|R_{n2}\| = o_P(1)$, which yields the desired result. ■

Proof of Lemma A.5. By the second-order Taylor expansion, as $n \to \infty$,

$$\begin{aligned}
\Gamma_n' \left[I_n - \widehat{\mathcal{P}}_n\right] \left(D_n \Pi_0(\widehat{V}_n) - D_n \Pi_0(V_n)\right)/\sqrt{n} &= \Gamma_n' \left[I_n - \widehat{\mathcal{P}}_n\right] D_n \dot{\Pi}_0(V_n) W_n \left[\widehat{\gamma}_n - \gamma_0\right]/\sqrt{n} \\
&\quad + \frac{1}{2}\Gamma_n' \left[I_n - \widehat{\mathcal{P}}_n\right] D_n \ddot{\Pi}_0(\overline{V}_n) \left\{W_n \left[\widehat{\gamma}_n - \gamma_0\right]\right\}^2/\sqrt{n} \\
&= \left[\Gamma_n' \left[I_n - \widehat{\mathcal{P}}_n\right] D_n \dot{\Pi}_0(V_n) W_n/n\right] \psi_n + o_P(1),
\end{aligned}$$

where $\ddot{\Pi}_0(\overline{V}_n) = \mathrm{diag}\left\{\partial_v^2 \pi_0(\bar{v}_1), ..., \partial_v^2 \pi_0(\bar{v}_n)\right\}$ with $\bar{v}_i \in [\widehat{v}_i, v_i]$ for $i = 1, ..., n$, and the second equality follows from Assumption 2.2. Then, using similar arguments to the proof of Lemma A.4, we obtain the desired result. ■

Proof of Lemma A.6. First, observe that

$$\begin{aligned}
\widehat{B}_{nL} &= \left[\widehat{P}_n^{L\prime} D_n \widehat{P}_n^L\right]^{-1} \widehat{P}_n^{L\prime} \left(D_n Y_n - D_n S_n(\widehat{\lambda}_n) Z_n \widehat{\beta}_n\right) \\
&= B_{0L} + \widehat{\mathcal{A}}_n' \Gamma_n(\widehat{\lambda}_n)(\theta_0 - \widehat{\theta}_n) + \widehat{\mathcal{A}}_n' \left(\Pi_0(V_n) - \widehat{P}_n^L B_{0L}\right) + \widehat{\mathcal{A}}_n' \widetilde{\mathcal{E}}_n \\
&= B_{0L} + M_{n1} + M_{n2} + M_{n3}, \text{ say.}
\end{aligned}$$

By the $\sqrt{n}$-consistency of $\widehat{\theta}_n$,

$$\begin{aligned}
\|M_{n1}\|^2 &\leq (\theta_0 - \widehat{\theta}_n)' \Gamma_n(\widehat{\lambda}_n)' D_n \widehat{P}_n^L \left[\widehat{P}_n^{L\prime} D_n \widehat{P}_n^L\right]^{-2} \widehat{P}_n^{L\prime} D_n \Gamma_n(\widehat{\lambda}_n)(\theta_0 - \widehat{\theta}_n) \\
&= \left[\underline{c}_d \underline{c}_P - o_P(1)\right]^{-1} (\theta_0 - \widehat{\theta}_n)' \left(\Gamma_n(\widehat{\lambda}_n)' \Gamma_n(\widehat{\lambda}_n)/n\right)(\theta_0 - \widehat{\theta}_n) = O_P(1/n)
\end{aligned}$$

as $n \to \infty$. Thus, $\|M_{n1}\| = O_P(1/\sqrt{n})$. For M_{n2}, by the mean value expansion and the same argument as in the proof of Lemma A.3, we obtain

$$\begin{aligned}
\|M_{n2}\| &\leq \left\|\widehat{\mathcal{A}}_n' \dot{\Pi}_0(\overline{V}_n) W_n \left[\widehat{\gamma}_n - \gamma_0\right]\right\| + \left\|\widehat{\mathcal{A}}_n' \left(\Pi_0(\widehat{V}_n) - \widehat{P}_n^L B_{0L}\right)\right\| \\
&= O_P(1/\sqrt{n}) + O(L^{-\rho}).
\end{aligned}$$

In addition, by Markov's inequality, we can readily show that $\|M_{n3}\| = O_P(\sqrt{L}/\sqrt{n})$. Then, the result follows from the triangle inequality. ■

2 Sketch of the Proof of Theorem 4.3

In Theorem 4.3 of this paper, we showed that the NLS estimator $\widehat{\theta}_{n,-1}$ for the model with group-specific effects has an asymptotic normality result similar to the one in Theorem 4.1. In order to establish this result, we modify Assumption 5 and 7, as follows.

Assumption 5′

Assumption 5 holds and, in addition, $L^2/n \to 0$ as $n \to \infty$.

Assumption 6′

1. $0 < \underline{c}_R < \liminf_{n\to\infty} \tau_{\min}\left(\overline{R}_n\right)$ and **2**. Assumption 6.2 holds.

Assumption 7′

1. $\Psi_1^\dagger \equiv \lim_{n\to\infty} \Psi_{n1}^\dagger$ and $\Psi_2^\dagger \equiv \lim_{n\to\infty} \Psi_{n2}^\dagger$ exist and are positive definite and **2**. $\digamma(\lambda) \equiv \lim_{n\to\infty} \digamma_n(\lambda)$ exists and is positive definite uniformly in $\lambda \in \mathcal{L}$, where

$$\begin{aligned}
\digamma_n(\lambda) &\equiv \Gamma_{n,-1}^{\perp}(\lambda)'\overline{R}_n\Gamma_{n,-1}^{\perp}(\lambda)/n \\
\Gamma_{n,-1}(\lambda) &\equiv S_n(\lambda)\left[G_nS_n(\lambda_0)Z_{n,-1}\beta_{0,-1}, Z_{n,-1}\right] \\
\Gamma_{n,-1}^{\perp}(\lambda) &\equiv \Gamma_{n,-1}(\lambda) - P_n^L\left(P_n^{L\prime}\overline{R}_nP_n^L\right)^{-1}P_n^{L\prime}\overline{R}_n\Gamma_{n,-1}(\lambda).
\end{aligned}$$

Note that Assumption 5′ is automatically satisfied by Assumption 5.2 for, for example, splines such that $\zeta_0(L) = O(L^{1/2})$. Then, replacing D_n and $\overline{D}_n$ with R_n and $\overline{R}_n$, respectively, we can obtain the desired result following the same line as Theorem 4.1. Most parts of the proof are similar to those given in the paper and above and can be omitted, so we only point out the differences.

First, we show the result which corresponds to Lemma A.1 (ii); namely,

$$\left\|P_n^{L\prime}R_nP_n^L/n - P_n^{L\prime}\overline{R}_nP_n^L/n\right\| = o_P(1). \tag{2.1}$$

Observe that, by definition, we can write

$$\begin{aligned}
P_n^{L\prime}R_nP_n^L/n &= \frac{1}{n}\sum_{i=1}^n\sum_{j\in\mathcal{N}_{r(i)}} p^L(v_i)\rho_{n,ij}p^L(v_j)' \\
&= \frac{1}{n}\sum_{i=1}^n p^L(v_i)p^L(v_i)'d_i - \frac{1}{n}\sum_{i=1}^n\sum_{j\in\mathcal{N}_{r(i)}} p^L(v_i)p^L(v_j)'e_{i,j},
\end{aligned}$$

where $\rho_{n,ij}$ is the (i,j)th element of R_n, and $e_{i,j} \equiv (\sum_{j\in\mathcal{N}_{r(i)}} d_j)^{-1}d_id_j$. Hence, by the triangle inequality,

$$\begin{aligned}
\left\|P_n^{L\prime}R_nP_n^L/n - P_n^{L\prime}\overline{R}_nP_n^L/n\right\| \leq{}& \left\|\frac{1}{n}\sum_{i=1}^n p^L(v_i)p^L(v_i)'(d_i - \mathrm{E}[d_i])\right\| \\
&+ \left\|\frac{1}{n}\sum_{i=1}^n\sum_{j\in\mathcal{N}_{r(i)}} p^L(v_i)p^L(v_j)'(e_{i,j} - \mathrm{E}[e_{i,j}])\right\|.
\end{aligned}$$

As we have shown above, the first term on the right-hand side is of order $o_P(1)$. For the second term, recalling that the size of each group is fixed at a constant number (i.e., there exists a constant $c_r > 0$ such that $\max_{1\leq i\leq n} n_{r(i)} < c_r$ uniformly in n),

$$\begin{aligned}
&\mathrm{E}\left\|\frac{1}{n}\sum_{i=1}^n\sum_{j\in\mathcal{N}_{r(i)}} p^L(v_i)p^L(v_j)'(e_{i,j} - \mathrm{E}[e_{i,j}])\right\|^2 \qquad (2.2)\\
={}& n^{-2}\sum_{k=1}^L\sum_{\ell=1}^L \mathrm{E}\left(\sum_{i=1}^n\sum_{j\in\mathcal{N}_{r(i)}} p_k(v_i)p_\ell(v_j)(e_{i,j} - \mathrm{E}[e_{i,j}])\right)^2 \\
={}& n^{-2}\sum_{k=1}^L\sum_{\ell=1}^L\sum_{i=1}^n\sum_{h=1}^n\sum_{j\in\mathcal{N}_{r(i)}}\sum_{j'\in\mathcal{N}_{r(h)}} p_k(v_i)p_\ell(v_j)p_k(v_h)p_\ell(v_{j'})\mathrm{E}[(e_{i,j} - \mathrm{E}[e_{i,j}])(e_{h,j'} - \mathrm{E}[e_{h,j'}])] \\
={}& n^{-2}\sum_{k=1}^L\sum_{\ell=1}^L\sum_{i=1}^n\sum_{h\in\mathcal{N}_{r(i)}}\sum_{j\in\mathcal{N}_{r(i)}}\sum_{j'\in\mathcal{N}_{r(h)}} p_k(v_i)p_\ell(v_j)p_k(v_h)p_\ell(v_{j'})\mathrm{E}[(e_{i,j} - \mathrm{E}[e_{i,j}])(e_{h,j'} - \mathrm{E}[e_{h,j'}])] \\
\leq{}& cn^{-2}\sum_{k=1}^L\sum_{\ell=1}^L\sum_{i=1}^n\sum_{h\in\mathcal{N}_{r(i)}}\sum_{j\in\mathcal{N}_{r(i)}}\sum_{j'\in\mathcal{N}_{r(i)}} p_k(v_i)p_\ell(v_j)p_k(v_h)p_\ell(v_{j'}).
\end{aligned}$$

Here, note that all (h, j, j') belong to the same group as i. Let

$$
\begin{aligned}
A_{n1} &\equiv n^{-2} \sum_{k=1}^{L} \sum_{\ell=1}^{L} \sum_{i=1}^{n} p_k^2(v_i) p_\ell^2(v_i) \\
A_{n2} &\equiv n^{-2} \sum_{k=1}^{L} \sum_{\ell=1}^{L} \sum_{i=1}^{n} \sum_{j' \in \mathcal{N}_{r(i)} \setminus \{i\}} p_k^2(v_i) p_\ell(v_i) p_\ell(v_{j'}) \\
A_{n3} &\equiv n^{-2} \sum_{k=1}^{L} \sum_{\ell=1}^{L} \sum_{i=1}^{n} \sum_{j \in \mathcal{N}_{r(i)} \setminus \{i\}} \sum_{j' \in \mathcal{N}_{r(i)} \setminus \{i\}} p_k^2(v_i) p_\ell(v_j) p_\ell(v_{j'}) \\
A_{n4} &\equiv n^{-2} \sum_{k=1}^{L} \sum_{\ell=1}^{L} \sum_{i=1}^{n} \sum_{h \in \mathcal{N}_{r(i)} \setminus \{i\}} \sum_{j \in \mathcal{N}_{r(i)} \setminus \{i\}} \sum_{j' \in \mathcal{N}_{r(i)} \setminus \{i\}} p_k(v_i) p_\ell(v_j) p_k(v_h) p_\ell(v_{j'}).
\end{aligned}
$$

As already shown in the proof of Lemma A.1 (ii), we have $A_{n1} = O(\zeta_0^2(L)L/n) = o(1)$. Similarly, for A_{n2},

$$
\begin{aligned}
A_{n2} &\leq n^{-2} \sum_{k=1}^{L} \sum_{\ell=1}^{L} \sum_{i=1}^{n} p_k^2(v_i) \cdot |p_\ell(v_i)| \sum_{j' \in \mathcal{N}_{r(i)} \setminus \{i\}} |p_\ell(v_{j'})| \\
&\leq O(c_r) n^{-2} \sum_{k=1}^{L} \sum_{\ell=1}^{L} \sum_{i=1}^{n} p_k^2(v_i) \cdot |p_\ell(v_i)| \\
&= O(\zeta_0^2(L)L/n) = o(1).
\end{aligned}
$$

Further, we have $A_{n3} = O(\zeta_0^2(L)L/n) = o(1)$ and $A_{n4} = O(L^2/n) = o(1)$, implying that the left-hand side of (2.2) is $o(1)$. Thus, by Markov's inequality, we obtain (2.1).

Next, as in Lemma A.2, we need to confirm that $\mathcal{E}_n' R_n \mathcal{E}_n / n - \mathrm{E}\left[\mathcal{E}_n' R_n \mathcal{E}_n / n\right] = o(1)$. This can be proved by using an analogous argument to Lemmas A.2 and A.3 of Lin and Lee (2010).

Finally, in the proof of asymptotic normality, we show, with a slight abuse of notation,

$$
\sum_{j=1}^{n} \mathrm{E}\left[\xi_{1j}^4\right] = o(1), \quad \text{and} \quad \sum_{j=1}^{n} \mathrm{E}\left[\xi_{2j}^4\right] = o(1)
$$

so that

$$
\sum_{j=1}^{n} (\xi_{1j} + \xi_{2j}) = n^{-1/2} \mathbf{c}' \Psi_{n1}^{\dagger -1/2} \Gamma_{n,-1}^{\perp\prime} R_n \widetilde{\mathcal{E}}_n \xrightarrow{d} N(0,1),
$$

where

$$
\begin{aligned}
\xi_{1j} &\equiv n^{-1/2} \mathbf{c}' \Psi_{n1}^{\dagger -1/2} \sum_{i=1}^{n} \varsigma_{i,-1} \rho_{n,ij} \widetilde{\varepsilon}_j, \\
\xi_{2j} &\equiv -n^{-1/2} \mathbf{c}' \Psi_{n1}^{\dagger -1/2} \Gamma_n' \overline{D}_n P_n^L \left(P_n^{L\prime} \overline{D}_n P_n^L \right)^{-1} \sum_{i=1}^{n} p^L(v_i) \rho_{n,ij} \widetilde{\varepsilon}_j.
\end{aligned}
$$

Note that, since $\varsigma_{i,-1}$ is uniformly bounded and the column sums of R_n is uniformly bounded in absolute value (as long as each group has at least one respondent with $d = 1$), we have $\sum_{i=1}^{n} |\varsigma_{i,-1} \rho_{n,ij}| < \infty$ uniformly in n. Hence,

$$
\begin{aligned}
\sum_{j=1}^{n} \mathrm{E}\left[\xi_{1j}^4\right] &= n^{-2} \sum_{j=1}^{n} \mathrm{E}\mathrm{E}\left[\left(\mathbf{c}' \Psi_{n1}^{\dagger -1/2} \sum_{i=1}^{n} \varsigma_{i,-1} \rho_{n,ij} \right)^4 \widetilde{\varepsilon}_j^4 \middle| D_n \right] \\
&\leq c n^{-2} \sum_{j=1}^{n} \mathrm{E}\left\{ \mathbf{c}' \Psi_{n1}^{\dagger -1} \Psi_{n1}^{\dagger -1} \mathbf{c} \cdot \mathrm{E}\left[\widetilde{\varepsilon}_j^4 | d_j \right] \right\} = O(n^{-1}) = o(1).
\end{aligned}
$$

Similarly, we can show $\sum_{j=1}^{n} \mathrm{E}\left[\xi_{2j}^4\right] = o(1)$.

3 Supplementary Tables

Below, we provide the descriptive statistics for the dataset used and supplementary empirical results in the above paper. Table 1 presents the descriptive statistics for our data, Table 2 shows the estimation results for the first-step estimation by the standard probit and Klein and Spady's (1993) SML, and Table 3 reports the results of the estimation of the GPA equation by the IMR and SSS models for the case when the two health status variables, Health and Absence, are excluded from the regressors.

Table 1: Descriptive statistics

	Mean	Median	Std. Dev.	Min.	Max.
d	0.8695	1	0.3369	0	1
GPA	2.8438	3	0.7780	1	4
Age	14.8631	15	0.8365	11	19
Male	0.4712	0	0.4992	0	1
White	0.7115	1	0.4531	0	1
Hispanic	0.5492	0	1.8769	0	8
Black	0.1781	0	0.3826	0	1
Asian	0.0516	0	0.2213	0	1
Live with both parents	0.7363	1	0.4406	0	1
Mother's education (< 9)	0.1033	0	0.3043	0	1
Mother's education (≥ 16)	0.2576	0	0.4373	0	1
Mother's job (professional)	0.2995	0	0.4581	0	1
Mother's job (unemployed)	0.0482	0	0.2142	0	1
Father's education (< 9)	0.0922	0	0.2894	0	1
Father's education (≥ 16)	0.2358	0	0.4245	0	1
Father's job (professional)	0.2436	0	0.4293	0	1
Father's job (unemployed)	0.0415	0	0.1995	0	1
Academic club	0.2273	0	0.4191	0	1
Sports club	0.6434	1	0.4790	0	1
No club	0.1427	0	0.3498	0	1
Delinquency	1.8459	1.3744	1.6098	0	9.5090
Health	2.1287	2	0.9357	1	5
Absence	0.4943	0	0.6355	0	4

$*$: The descriptive statistics for GPA are calculated over those with $d = 1$.
Sample size: 6721 (3302: 9th graders, 3419: 10th graders)

Table 2: Estimation results for the selection equation

	Probit		Klein and Spady*	
	Coefficient	t-value	Coefficient	t-value
Intercept	3.2836	6.1690	-	-
Own Effects				
Neg. Age	0.1273	4.0026	1.0000	-
Male	-0.0043	-0.0881	-0.0018	-0.1894
White	0.1801	2.4712	0.0466	3.0430
Hispanic	-0.0192	-1.8523	-0.0014	-0.6202
Black	0.0564	0.6237	0.0141	0.7587
Asian	0.0423	0.3689	0.0114	0.5094
Live with both parents	0.1475	2.8179	0.0220	2.0100
Mother's education (<9)	0.0602	0.8648	0.0090	0.6064
Mother's education (≥ 16)	0.1391	2.1769	0.0172	1.3755
Mother's job (professional)	0.0879	1.6439	0.0177	1.6537
Mother's job (unemployed)	-0.1184	-1.2835	-0.0172	-0.9024
Father's education (<9)	0.0363	0.4624	0.0082	0.5017
Father's education (≥ 16)	-0.0116	-0.1712	-0.0021	-0.1584
Father's job (professional)	0.0618	0.9622	0.0051	0.4088
Father's job (unemployed)	-0.1054	-0.9923	-0.0150	-0.6988
Academic club	0.2544	4.2995	0.0321	2.8296
Sports club	-0.0186	-0.3325	0.0012	0.1047
No club	-0.1076	-1.5528	-0.0220	-1.5102
Delinquency	0.0001	0.0096	-0.0004	-0.1307
Health	-0.0410	-1.7165	-0.0080	-1.7059
Absence	-0.0581	-1.7116	-0.0150	-2.1526
10th grade dummy	-0.1502	-0.5061	0.0032	0.0858
Contextual Effects				
Age	-0.0036	-0.3181	-0.0031	-1.3163
Male	0.0233	0.2906	0.0043	0.2672
White	0.0966	0.7830	0.0246	0.9396
Hispanic	-0.0138	-0.7126	-0.0023	-0.5717
Black	0.0381	0.2679	0.0201	0.6917
Asian	-0.2920	-1.6273	-0.0744	-2.1104
Live with both parents	0.1764	1.8620	0.0446	2.2741
Mother's education (<9)	-0.1139	-0.9381	-0.0209	-0.8257
Mother's education (≥ 16)	0.1329	1.2138	0.0240	1.1372
Mother's job (professional)	0.1906	1.9695	0.0292	1.5117
Mother's job (unemployed)	0.1362	0.7874	0.0243	0.7017
Father's education (<9)	-0.0439	-0.3303	-0.0058	-0.2120
Father's education (≥ 16)	0.0776	0.6348	0.0158	0.6781
Father's job (professional)	0.0245	0.2182	-0.0018	-0.0822
Father's job (unemployed)	-0.0132	-0.0717	-0.0106	-0.2879
Academic club	0.0094	0.0987	-0.0091	-0.5070
Sports club	0.1566	1.6978	0.0404	2.1920
No club	0.0130	0.1031	0.0021	0.0787
Delinquency	-0.0456	-1.8786	-0.0084	-1.6872
Sample size	6721		6721	

Note: The results for the school-specific dummies are omitted to save space.
$*$: The coefficient of negative age (Neg. Age) is set to one for identification.

Table 3: Estimation results when the health variables are excluded from the GPA equation

Group-specific effects	No				Yes			
	IMR model*		SSS model*		IMR model		SSS model	
	Coef.	t-value	Coef.	t-value	Coef.	t-value	Coef.	t-value
Intercept	4.2580	15.2265	3.3020	3.0151				
Endogeneous Effect	0.3450	5.9117	0.3940	4.7499	0.5660	4.9870	0.6274	6.9272
Own Effects								
Age	-0.1154	-5.8334	-0.4061	-1.8679	-0.0581	-2.5911	-0.0816	-4.3190
Male	-0.1279	-5.1513	-0.1286	-6.1635	-0.1534	-6.0798	-0.1545	-6.9703
White	0.0716	1.6221	0.0765	2.0622	0.0961	2.3601	0.0979	2.4013
Hispanic	-0.0222	-3.4842	-0.0214	-4.0982	-0.0137	-2.0847	-0.0165	-2.9593
Black	-0.0986	-1.8107	-0.1041	-2.3222	0.0078	0.1129	0.0126	0.1979
Asian	0.3222	5.0255	0.2996	5.8169	0.2071	2.8758	0.2130	3.3561
Live with both parents	0.0774	2.5756	0.0777	3.1456	0.0688	2.1819	0.0909	3.4854
Mother's education (< 9)	-0.0781	-1.8716	-0.0795	-2.2663	-0.0988	-2.2451	-0.0936	-2.4297
Mother's education (≥ 16)	0.1254	4.3520	0.1284	5.2429	0.0822	2.7650	0.0977	3.9159
Mother's job (professional)	0.0371	1.4768	0.0406	1.8763	0.0131	0.5109	0.0210	0.9543
Mother's job (unemployed)	-0.0453	-0.8366	-0.0519	-1.1782	-0.0265	-0.4744	-0.0454	-0.9661
Father's education (< 9)	-0.0099	-0.2324	-0.0059	-0.1635	-0.0309	-0.7057	-0.0207	-0.5417
Father's education (≥ 16)	0.0883	3.0054	0.0885	3.5554	0.0518	1.7546	0.0482	1.9199
Father's job (professional)	0.1133	4.0584	0.1127	4.7156	0.0615	2.1683	0.0689	2.8361
Father's job (unemployed)	-0.0365	-0.6539	-0.0397	-0.8481	-0.0401	-0.6991	-0.0580	-1.2001
Academic club	0.1933	7.0493	0.1975	8.4003	0.1194	4.0222	0.1435	6.1584
Sports club	0.0751	2.6718	0.0766	3.1830	0.0439	1.4984	0.0433	1.6928
No club	-0.0988	-2.4578	-0.1000	-2.9652	-0.0252	-0.5971	-0.0408	-1.1292
Delinquency	-0.0888	-11.5775	-0.0912	-14.0129	-0.0745	-9.6298	-0.0710	-10.4665
10th grade dummy	0.1017	2.6680	0.1213	3.2293				
Contextual Effects								
Age	-0.0554	-5.0827	-0.0123	-0.8798	0.0020	0.0707	0.0199	0.7268
Male	0.1737	4.5161	0.1848	5.6997	0.1772	4.9572	0.1812	5.6110
White	-0.0040	-0.0606	-0.0157	-0.2947	0.0390	0.5904	0.0251	0.4232
Hispanic	0.0002	0.0202	0.0022	0.2574	-0.0085	-0.8453	-0.0065	-0.7173
Black	0.0515	0.6686	0.0581	0.9035	0.0211	0.2461	0.0263	0.3332
Asian	-0.1140	-1.1578	-0.1420	-1.7748	0.0797	0.7383	0.0358	0.3797
Live with both parents	-0.0685	-1.4216	-0.0683	-1.6664	-0.0238	-0.5104	-0.0193	-0.4534
Mother's education (< 9)	0.0741	1.1047	0.0826	1.4731	0.0169	0.2467	0.0116	0.1890
Mother's education (≥ 16)	-0.0409	-0.8964	-0.0492	-1.2241	-0.0193	-0.4025	-0.0204	-0.4751
Mother's job (professional)	0.0825	2.0040	0.0831	2.3812	0.0478	1.1371	0.0583	1.5511
Mother's job (unemployed)	0.1288	1.4864	0.1284	1.7649	-0.0232	-0.2602	0.0110	0.1381
Father's education (< 9)	-0.0276	-0.4173	-0.0492	-0.8812	-0.0838	-1.2458	-0.0818	-1.3581
Father's education (≥ 16)	0.1045	2.1227	0.0991	2.2678	0.0535	1.4398	0.0580	1.8678
Father's job (professional)	-0.0383	-0.8313	-0.0459	-1.1338	0.0163	0.3369	0.0026	0.0605
Father's job (unemployed)	-0.0298	-0.3434	-0.0250	-0.3386	-0.0676	-0.7437	-0.0525	-0.6534
Academic club	0.0539	1.2993	0.0356	0.9126	0.0899	1.7570	0.0591	1.2865
Sports club	-0.0123	-0.2915	-0.0099	-0.2669	-0.0159	-0.3921	-0.0037	-0.1009
No club	-0.0189	-0.2927	0.0021	0.0384	-0.0346	-0.5602	-0.0160	-0.2892
Delinquency	-0.0214	-1.5513	-0.0136	-0.8763	-0.0082	-0.4405	-0.0092	-0.5710
Sample size	6721		6721		5689		5689	

$*$ School specific-dummy variables are included, but are not displayed to save space.

4 Additional Monte Carlo experiment

In this section, we numerically examine the estimation accuracy of the intercept term in a model with isolated individuals. As discussed in Section 3.1, the (regular) identification of the intercept α_0 is not possible if $\lambda_0 = 0$; and even when λ_0 is not zero but close to zero, the identification is weak. The DGP considered in this analysis is as follows:

$$\begin{aligned} d_i \cdot y_i &= d_i \cdot \left(\frac{\lambda_0}{n(i)} \sum_{j \in \mathcal{C}_i} \mathrm{E}\left[y_j\right] + x_i \beta_{01} + \frac{1}{n(i)} \sum_{j \in \mathcal{C}_i} x_j \beta_{02} + \alpha_0 + \varepsilon_i \right) \\ d_i &= 1(\gamma_{01} + x_i \gamma_{02} + w_i \gamma_{03} + \eta_i \geq 0), \end{aligned}$$

for $i = 1, ..., n$, where, x_i's $\sim$ i.i.d. $N(0,1)$, w_i's $\sim$ i.i.d. $N(0,1)$, η_i's $\sim$ i.i.d. $N(0,1)$, and $\varepsilon_i = \sqrt{0.3} \cdot \eta_i + \sqrt{1 - 0.3} \cdot \varepsilon_{1i}$ with ε_{1i}'s $\sim$ i.i.d. $N(0,1)$. The true values for the parameters are as follows: $\beta_{01} = 1$, $\beta_{02} = 1$, $\gamma_{01} = 0.5$, $\gamma_{02} = 0.5$, $\gamma_{03} = 0.5$, and $\alpha_0 = 1$. For this DGP, we consider three sample sizes: $n \in \{200, 400, 800\}$. The size of each group n_r is set to either 20, 30, or 50 in a way such that $n_1 = 20$, $n_2 = 30$, $n_3 = 50$, $n_4 = 20$, ... and so on. For each group r, the interaction matrix G_r is defined according to a rook-contiguity spatial weight matrix. In other words, we randomly allocate n_r units on the lattice of $5 \times t$ such that $5t = n_r$, and set $j \in \mathcal{C}_i$ if i and j are rook-contiguous on the lattice. Furthermore, we randomly select $n_r/5$ individuals within each group and reset their $\mathcal{C}_i$ to empty to identify α_0. Then, we compute the average bias for estimating α_0 with different λ_0 values ranging from 0.1 to 0.9, that is, $\lambda_0 \in \{0.1, 0.2, ..., 0.9\}$. The first-step estimation of $(\gamma_{01}, \gamma_{02}, \gamma_{03})$ is carried out simply using the standard probit estimator. The number of Monte Carlo repetitions is set to 500 for each case.

The results of this simulation analysis are summarized in Figure 1, where the x-axis represents the value of λ_0 and the y-axis is the average bias. As predicted by the theory, when λ_0 is close to zero, α_0 cannot be estimated correctly regardless of sample size. In addition, we observe that the identification power is heavily dependent on the sample size: the larger the sample size, the higher the power. This result may reflect the fact that, in our DGP, we can observe a larger number of isolated individuals as the sample size increases, and thus more variations in the value of $(I_n - \lambda_0 G_n)^{-1}$. For example, when $\lambda_0 = 0.4$, it is possible to almost precisely estimate the intercept if we have $n = 800$, while the estimates of the intercept for $n = 200$ are still largely biased. When $n = 200$, it seems to require λ_0 to be at least 0.6 for precise estimation; the case with $n = 400$ is in between.

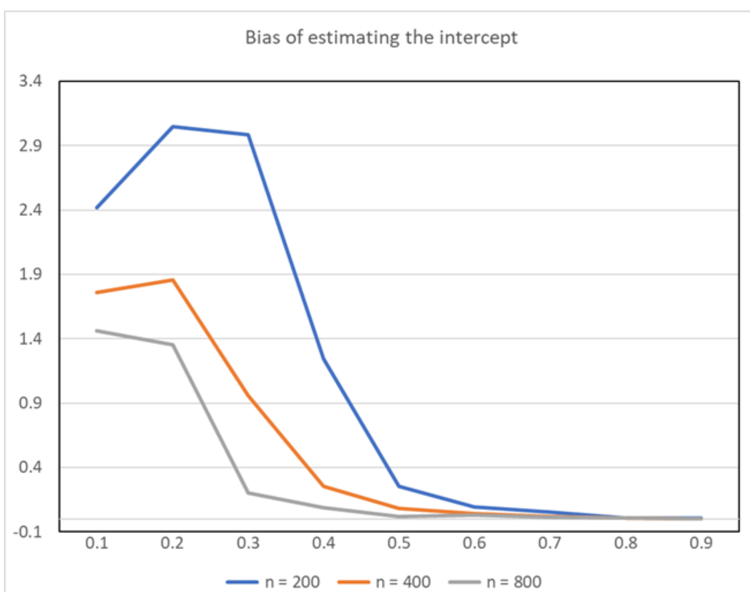

Figure 1: Bias of estimating the intercept

References

Klein, R. W., and Spady, R. H. (1993) An efficient semiparametric estimator for binary response models. Econometrica, 61, 387-421.

Lee, L. F. (2004) Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. Econometrica, 72, 1899-1925.

Lin, X., and Lee, L. F. (2010) GMM estimation of spatial autoregressive models with unknown heteroskedasticity. Journal of Econometrics, 157, 34-52.

Su, L., and Jin, S. (2012) Sieve estimation of panel data models with cross section dependence. Journal of Econometrics, 169, 34-47.