Supplementary Information for "Partial Replacement Imputation Estimation for Partially Linear Models with Complex Missing Pattern Covariates"

Zishu Zhan[1,2], Xiangjie Li[3] and Jingxiao Zhang[1,2]

September 25, 2022

[1] *School of Statistics, Renmin University of China*

[2] *Center for Applied Statistics, Renmin University of China*

[3] *Changping Laboratory*

We give proofs of the theoretical results and some additional numerical results that support our conclusions in the following supplementary information.

1 Proofs of the theoretical results

We give some notation. Let $Z_{ij} = X_{ij}$, $\boldsymbol{Z}_i = (Z_{ij} : j = p+1, \cdots, p+q)^\top$, $\boldsymbol{Z}_{i,B_i} = \boldsymbol{X}_{i,B_i}$, $\boldsymbol{\theta}_0 = (\boldsymbol{1}^\top, \boldsymbol{\beta}_0^\top)^\top$, $\boldsymbol{\theta} = (\boldsymbol{1}^\top, \boldsymbol{\beta}^\top)^\top$, $\boldsymbol{\gamma} = (\boldsymbol{b}^\top, \boldsymbol{\beta}^\top)^\top$, $\dot{\boldsymbol{\gamma}}_\beta = (\dot{\boldsymbol{b}}^\top, \boldsymbol{\beta}^\top)^\top$, $\dot{\boldsymbol{\gamma}} = (\dot{\boldsymbol{b}}^\top, \dot{\boldsymbol{\beta}}^\top)^\top$, $\hat{\boldsymbol{\gamma}} = (\hat{\boldsymbol{b}}^\top, \hat{\boldsymbol{\beta}}^\top)^\top$, $\boldsymbol{\gamma}_0 = (\dot{\boldsymbol{b}}^\top, \boldsymbol{\beta}_0^\top)^\top$, $\tilde{\boldsymbol{Z}}_i = (Z_{ij}I(j \in B_i) + \hat{W}_{ij}I(j \in \bar{B}_i) : j = p+1, p+2, \cdots, p+q)^\top$,

$\tilde{\boldsymbol{z}}_i = (Z_{ij}I(j \in B_i) + W_{ij}I(j \in \bar{B}_i) : j = p+1, p+2, \cdots, p+q)^\top$, $\tilde{\boldsymbol{a}}_j(X_{ij}) = (a_{j,l}(X_{ij})I(j \in A_i) + E(a_{j,l}|\boldsymbol{X}_{i,A_i}, \boldsymbol{Z}_{i,B_i})I(j \in \bar{A}_i) : l = 1, 2, \cdots, L)^\top$, $\tilde{\boldsymbol{\alpha}}_j(X_{ij}) = (a_{j,l}(X_{ij})I(j \in A_i) + \hat{a}_{j,l}I(j \in \bar{A}_i) : l = 1, 2, \cdots, L)^\top$, $\tilde{\boldsymbol{\alpha}}(\boldsymbol{X}_i) = (\tilde{\boldsymbol{\alpha}}_1(X_{i1})^\top, \cdots, \tilde{\boldsymbol{\alpha}}_p(X_{ip})^\top)^\top$, $\tilde{\boldsymbol{a}}(\boldsymbol{X}_i) = (\tilde{\boldsymbol{a}}_1(X_{i1})^\top, \cdots, \tilde{\boldsymbol{a}}_p(X_{ip})^\top)^\top$,

$g(\boldsymbol{X}_i) = (g_1(X_{i1}), \cdots, g_p(X_{ip}))^\top$, $\tilde{g}(\boldsymbol{X}_i) = \big(g_j(X_{ij})I(j \in A_i) + V_{ij}I(j \in \bar{A}_i) : j = 1, 2, \cdots, p\big)^\top$, $\boldsymbol{L}_i = (g(\boldsymbol{X}_i)^\top, \boldsymbol{Z}_i^\top)^\top$, $\tilde{\boldsymbol{l}}_i = (\tilde{g}(\boldsymbol{X}_i)^\top, \tilde{\boldsymbol{z}}_i^\top)^\top$, $\boldsymbol{F}_i = (\boldsymbol{a}(\boldsymbol{X}_i)^\top, \boldsymbol{Z}_i^\top)^\top$, $\tilde{\boldsymbol{f}}_i = (\tilde{\boldsymbol{a}}(\boldsymbol{X}_i)^\top, \tilde{\boldsymbol{z}}_i^\top)^\top$. Noting that the dimension of $\boldsymbol{1}$ is allowed to vary with p and it is more proper to write $\boldsymbol{1}_p$. However, we generally ignore the dimensions of $\boldsymbol{1}$ vectors for notational convenience, and denote it by $\boldsymbol{1}$. If necessary, we will write their dimensions explicitly.

1.1 Proof of Lemma 1

Firstly, we prove that $u(\boldsymbol{\theta}_0) = u(\boldsymbol{\beta}_0) = \mathbf{0}$, where $u(\boldsymbol{\theta}) = E\left\{\tilde{\boldsymbol{l}}_i\left(L_i^\top \boldsymbol{\theta}_0 - \tilde{\boldsymbol{l}}_i^\top \boldsymbol{\theta}\right)\right\}$, $u(\boldsymbol{\beta}) = E\left\{\tilde{\boldsymbol{z}}_i\left(\boldsymbol{Z}_i^\top \boldsymbol{\beta}_0 - \tilde{\boldsymbol{z}}_i^\top \boldsymbol{\beta}\right)\right\}$.

We have

$$E(Y_i|\boldsymbol{X}_{i,A_i}, \boldsymbol{Z}_{i,B_i}) = \sum_{j\in A_i} g_j(X_{ij}) + \sum_{j\in \bar{A}_i} V_{ij} + \sum_{j\in B_i} Z_{ij}\beta_j + \sum_{j\in \bar{B}_i} W_{ij}\beta_j = \tilde{\boldsymbol{l}}_i^\top \boldsymbol{\theta}_0.$$

Also we have $E(Y_i|\boldsymbol{X}_{i,A_i}, \boldsymbol{Z}_{i,B_i}) = E(\boldsymbol{L}_i^\top|\boldsymbol{X}_{i,A_i}, \boldsymbol{Z}_{i,B_i})\boldsymbol{\theta}_0$.

Hence $E\left(\boldsymbol{L}_i^\top - \tilde{\boldsymbol{l}}_i^\top|\boldsymbol{X}_{i,A_i}, \boldsymbol{Z}_{i,B_i}\right)\boldsymbol{\theta}_0 = E\left(\boldsymbol{Z}_i^\top - \tilde{\boldsymbol{z}}_i^\top|\boldsymbol{X}_{i,A_i}, \boldsymbol{Z}_{i,B_i}\right)\boldsymbol{\beta}_0 = 0$ and then $E\left\{\tilde{\boldsymbol{l}}_i\left(\boldsymbol{L}_i^\top \boldsymbol{\theta}_0 - \tilde{\boldsymbol{l}}_i^\top \boldsymbol{\theta}_0\right)\right\} = 0$, $E\left\{\tilde{\boldsymbol{z}}_i\left(\boldsymbol{Z}_i^\top \boldsymbol{\beta}_0 - \tilde{\boldsymbol{z}}_i^\top \boldsymbol{\beta}_0\right)\right\} = 0$ due to the reason that $\tilde{\boldsymbol{l}}_i$ and $\tilde{\boldsymbol{z}}_i$ are the functions of $\left(\boldsymbol{X}_{i,A_i}^\top, \boldsymbol{Z}_{i,B_i}^\top, A_i, B_i\right)$.

Next, we prove that $u(\boldsymbol{\beta}) = 0$ has a unique solution. We assume there is another solution $\tilde{\boldsymbol{\beta}} \in \mathcal{B}$ such that $u(\tilde{\boldsymbol{\beta}}) = 0$. By the simple calculation, we have

$$\begin{aligned} u(\tilde{\boldsymbol{\beta}}) - u(\boldsymbol{\beta}_0) &= E\left\{\tilde{\boldsymbol{z}}_i(\boldsymbol{Z}_i^\top \boldsymbol{\beta}_0 - \tilde{\boldsymbol{z}}_i^\top \tilde{\boldsymbol{\beta}})\right\} - E\left\{\tilde{\boldsymbol{z}}_i(\boldsymbol{Z}_i^\top \boldsymbol{\beta}_0 - \tilde{\boldsymbol{z}}_i^\top \boldsymbol{\beta}_0)\right\} \\ &= -E\left\{\tilde{\boldsymbol{z}}_i \tilde{\boldsymbol{z}}_i^\top\right\}(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) = 0. \end{aligned}$$

For $E\left\{\tilde{\boldsymbol{z}}_i \tilde{\boldsymbol{z}}_i^\top\right\}$ is invertible, we can obtain $\tilde{\boldsymbol{\beta}} = \boldsymbol{\beta}_0$ and then $\boldsymbol{\beta}_0$ is the unique solution of $u(\boldsymbol{\beta}) = 0$.

Noting that $\dot{U}(\boldsymbol{\beta}) - u(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^n \tilde{\boldsymbol{Z}}_i\{Y_i - \tilde{\boldsymbol{F}}_i^\top \dot{\boldsymbol{\gamma}}_\beta\} - E\left\{\tilde{\boldsymbol{z}}_i(\boldsymbol{Z}_i^\top \boldsymbol{\beta}_0 - \tilde{\boldsymbol{z}}_i^\top \boldsymbol{\beta})\right\}$, then we rewrite

$$\dot{U}(\boldsymbol{\beta}) - u(\boldsymbol{\beta}) = U_1 - U_2 + U_3 + U_4 + U_5,$$

where

$$\begin{aligned} U_1 &= \frac{1}{n}\sum_{i=1}^n \left(\tilde{\boldsymbol{Z}}_i - \tilde{\boldsymbol{z}}_i\right) Y_i, \\ U_2 &= \frac{1}{n}\sum_{i=1}^n \left(\tilde{\boldsymbol{Z}}_i \tilde{\boldsymbol{F}}_i^\top - \tilde{\boldsymbol{z}}_i \tilde{\boldsymbol{f}}_i^\top\right)\dot{\boldsymbol{\gamma}}_\beta, \\ U_3 &= \frac{1}{n}\sum_{i=1}^n \tilde{\boldsymbol{z}}_i(\tilde{\boldsymbol{l}}_i^\top \boldsymbol{\theta} - \tilde{\boldsymbol{f}}_i^\top \dot{\boldsymbol{\gamma}}_\beta), \\ U_4 &= \frac{1}{n}\sum_{i=1}^n \{\tilde{\boldsymbol{z}}_i(\boldsymbol{L}_i^\top \boldsymbol{\theta}_0 - \tilde{\boldsymbol{l}}_i^\top \boldsymbol{\theta})\} - E\{\tilde{\boldsymbol{z}}_i(\boldsymbol{L}_i^\top \boldsymbol{\theta}_0 - \tilde{l}_i^\top \boldsymbol{\theta})\}, \\ U_5 &= \frac{1}{n}\sum_{i=1}^n \tilde{\boldsymbol{z}}_i \varepsilon_{i0}. \end{aligned}$$

Next, we will show that

(i) $\sup_{\boldsymbol{\beta}\in\mathcal{B}} \|U_1\| = o_p(1)$

(ii) $\sup_{\boldsymbol{\beta}\in\mathcal{B}} \|U_2\| = o_p(1)$

(iii) $\sup_{\boldsymbol{\beta}\in\mathcal{B}} \|U_3\| = o_p(1)$

(iv) $\sup_{\boldsymbol{\beta}\in\mathcal{B}} \|U_4\| = o_p(1)$

(v) $\sup_{\boldsymbol{\beta}\in\mathcal{B}} \|U_5\| = o_p(1)$

<u>(i)</u> For $\left(j \in \bar{B}_i\right)$.

$$W_{ij} - \hat{W}_{ij} = \frac{\sum_{i'=1}^{n} I(C_{i'} \supset \{C_i \cup j\})\{Z_{i'j} - W_{ij}\}\mathbf{K}_{\mathbf{H}}(\boldsymbol{U}_{i',C_i} - \boldsymbol{U}_{i,C_i})}{\sum_{i'=1}^{n} I(C_{i'} \supset \{C_i \cup j\})\mathbf{K}_{\mathbf{H}}(\boldsymbol{U}_{i',C_i} - \boldsymbol{U}_{i,C_i})} = \frac{W_{ij1}}{W_{ij2}},$$

By the arguments in [3] and [4], define $d_1 = |C_i| < p + q$, we have that

$$\sup_{j,\boldsymbol{U}_{i,C_i},C_i} \|W_{ij1}\| = O_p\left(\sqrt{\frac{\ln n}{nh^{d_1}}} + h^2\right),$$

$$\sup_{j,\boldsymbol{U}_{i,C_i},C_i} \|W_{ij2} - P(C_{i'} \supset \{C_i \cup j\}) f_{C_i}(\boldsymbol{U}_{i,C_i})\| = O_p\left(\sqrt{\frac{\ln n}{nh^{d_1}}} + h^2\right),$$

$$\sup_{j,\boldsymbol{U}_{i,C_i},C_i} \|\hat{W}_{ij} - W_{ij}\| = o_p(1). \tag{1}$$

and

$$\left\| \frac{1}{n}\sum_{i=1}^{n} \left(\tilde{\boldsymbol{Z}}_i - \tilde{\boldsymbol{z}}_i\right) Y_i \right\| \leq \frac{1}{n}\sum_{i=1}^{n} q\|Y_i\| \sup_{j,\boldsymbol{U}_{i,C_i},C_i} \|\hat{W}_{ij} - W_{ij}\| = o_p(1). \tag{2}$$

Hence, we have $\sup_{\boldsymbol{\beta}\in\mathcal{B}} \|U_1\| = o_p(1)$.

<u>(ii)</u> Analogous to the proof of (i), we can get that $\sup_{j,\boldsymbol{U}_{i,C_i},C_i} \|\hat{V}_{ij} - V_{ij}\| = o_p(1)$. Also, we get the following decomposition

$$\begin{aligned} U_2 &= \frac{1}{n}\sum_{i=1}^{n} \left(\tilde{\boldsymbol{Z}}_i - \tilde{\boldsymbol{z}}_i\right) \tilde{\boldsymbol{f}}_i^{\top} \dot{\boldsymbol{\gamma}}_{\beta} + \frac{1}{n}\sum_{i=1}^{n} \tilde{\boldsymbol{Z}}_i \left(\tilde{\boldsymbol{F}}_i^{\top} \dot{\boldsymbol{\gamma}}_{\beta} - \tilde{\boldsymbol{f}}_i^{\top} \dot{\boldsymbol{\gamma}}_{\beta}\right) \\ &= U_{21} + U_{22}. \end{aligned}$$

where

$$\|U_{21}\| \leq \frac{1}{n}\sum_{i=1}^{n} q\|\tilde{\boldsymbol{f}}_i^{\top} \dot{\gamma}_{\boldsymbol{\beta}}\| \sup_{j,\boldsymbol{U}_{i,C_i},C_i} \|\hat{W}_{ij} - W_{ij}\| = o_p(1),$$

and there exists a constant $0 < \bar{C} < \infty$ such that

$$\begin{aligned}
\|U_{22}\| &\leq \frac{1}{n}\sum_{i=1}^{n} Lp\|\tilde{\boldsymbol{Z}}_i\| \sup_{j,\boldsymbol{U}_{i,C_i},C_i} \|\hat{W}_{ij} - W_{ij}\| \\
&+ \frac{1}{n}\sum_{i=1}^{n} Lp\bar{C}\|\tilde{\boldsymbol{Z}}_i\| \sup_{j,\boldsymbol{U}_{i,C_i},C_i} \|\hat{V}_{ij} - V_{ij}\| = o_p(1).
\end{aligned}$$

Thus, $\sup_{\boldsymbol{\beta}\in\mathcal{B}} \|U_2\| \leq \sup_{\boldsymbol{\beta}\in\mathcal{B}} \|U_{21}\| + \sup_{\boldsymbol{\beta}\in\mathcal{B}} \|U_{22}\| = o_p(1)$.

(iii) By simple calculation, we have the following results

$$\begin{aligned}
U_3 &= \frac{1}{n}\sum_{i=1}^{n} \tilde{\boldsymbol{z}}_i \left(\tilde{g}(\boldsymbol{X}_i)^\top \boldsymbol{1} - \tilde{\boldsymbol{a}}(\boldsymbol{X}_i)^\top \dot{\boldsymbol{b}}\right) \\
&= E\left[\frac{1}{n}\sum_{i=1}^{n} \tilde{\boldsymbol{z}}_i (\tilde{g}(\boldsymbol{X}_i)^\top \boldsymbol{1} - \tilde{\boldsymbol{a}}(\boldsymbol{X}_i)^\top \dot{\boldsymbol{b}}) \Big| \boldsymbol{X}_{i,A_i}, \boldsymbol{Z}_{i,B_i}\right].
\end{aligned}$$

By the Assumption (A7) and spline theory mentioned in the main paper, we have

$$\begin{aligned}
\left\|\frac{1}{n}\sum_{i=1}^{n} \tilde{\boldsymbol{z}}_i (\tilde{g}(\boldsymbol{X}_i)^\top \boldsymbol{1} - \tilde{\boldsymbol{a}}(\boldsymbol{X}_i)^\top \dot{\boldsymbol{b}})\right\| &\leq \frac{1}{n}\sum_{i=1}^{n} \|\tilde{\boldsymbol{z}}_i\| \|\tilde{g}(\boldsymbol{X}_i)^\top \boldsymbol{1} - \tilde{\boldsymbol{a}}(\boldsymbol{X}_i)^\top \dot{\boldsymbol{b}}\|_\infty \\
&= O_p(L^{-(r+v)}) = o_p(1).
\end{aligned}$$

Thus, $\sup_{\boldsymbol{\beta}\in\mathcal{B}} \|U_3\| = o_p(1)$.

(iv) We consider the function class $\mathcal{F} = \{f(\boldsymbol{\beta}) = \tilde{\boldsymbol{z}}_i(\boldsymbol{L}_i^\top \boldsymbol{\theta}_0 - \tilde{\boldsymbol{l}}_i^\top \boldsymbol{\theta}) : \boldsymbol{\beta} \in \mathcal{B}\}$. By Theorem 19.1 of [7], we conclude that $\mathcal{F}$ is a Glivenko-Cantelli class. And it follows from Glivenko-Cantelli theorem that $\sup_{\boldsymbol{\beta}\in\mathcal{B}} \|U_4\| = o_p(1)$.

(v) Similar to the proof of (iv), we can also obtain $\sup_{\boldsymbol{\beta}\in\mathcal{B}} \|U_5\| = o_p(1)$.

Thus, we complete the proof of $\sup_{\boldsymbol{\beta}\in\mathcal{B}} \|\dot{U}(\boldsymbol{\beta}) - u(\boldsymbol{\beta})\| = o_p(1)$. Similar to [4], using the fact that $\boldsymbol{\beta}_0$ is the unique solution to $u(\boldsymbol{\beta})$ and the continuous mapping theorem, we can conclude that $\dot{\boldsymbol{\beta}} \to \boldsymbol{\beta}_0$ in probability as $n \to \infty$.

1.2 Proof of Lemma 2

$$\dot{U}(\dot{\boldsymbol{\beta}}) - \dot{U}(\boldsymbol{\beta}_0) = \frac{1}{n}\sum_{i=1}^{n} \tilde{\boldsymbol{Z}}_i\{Y_i - \tilde{\boldsymbol{F}}_i^\top \dot{\boldsymbol{\gamma}}\} - \frac{1}{n}\sum_{i=1}^{n} \tilde{\boldsymbol{Z}}_i\{Y_i - \tilde{\boldsymbol{F}}_i^\top \boldsymbol{\gamma}_0\}$$

$$
\begin{aligned}
&= -\frac{1}{n}\sum_{i=1}^{n}\tilde{\boldsymbol{z}}_i\tilde{\boldsymbol{z}}_i^\top(\dot{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)\\
&- \frac{1}{n}\sum_{i=1}^{n}\{\tilde{\boldsymbol{Z}}_i-\tilde{\boldsymbol{z}}_i\}\tilde{\boldsymbol{Z}}_i^\top(\dot{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)\\
&- \frac{1}{n}\sum_{i=1}^{n}\tilde{\boldsymbol{z}}_i(\tilde{\boldsymbol{Z}}_i-\tilde{\boldsymbol{z}}_i)^\top(\dot{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)\\
&= -\frac{1}{n}\sum_{i=1}^{n}\tilde{\boldsymbol{z}}_i\tilde{\boldsymbol{z}}_i^\top(\dot{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)+o_p(\|\dot{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\|),
\end{aligned}
$$

where the last equality follows from (1) and (2) in the proof of Lemma 1. By the fact $\dot{U}(\dot{\boldsymbol{\beta}})=0$, we have

$$
\sqrt{n}(\dot{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)=\left(\frac{1}{n}\sum_{i=1}^{n}\tilde{\boldsymbol{z}}_i\tilde{\boldsymbol{z}}_i^\top\right)^{-1}\sqrt{n}\dot{U}(\boldsymbol{\beta}_0)+o_p(1).
$$

$$
\begin{aligned}
\dot{U}(\boldsymbol{\beta}_0) &= \frac{1}{n}\sum_{i=1}^{n}\tilde{\boldsymbol{Z}}_i\left(\varepsilon_{i0}+\boldsymbol{L}_i^\top\boldsymbol{\theta}_0-\tilde{\boldsymbol{F}}_i^\top\boldsymbol{\gamma}_0\right)\\
&= \frac{1}{n}\sum_{i=1}^{n}\tilde{\boldsymbol{Z}}_i\left(\varepsilon_{i0}+\boldsymbol{L}_i^\top\boldsymbol{\theta}_0-\tilde{\boldsymbol{l}}_i^\top\boldsymbol{\theta}_0\right)+\frac{1}{n}\sum_{i=1}^{n}\tilde{\boldsymbol{Z}}_i\left(\tilde{\boldsymbol{l}}_i^\top\boldsymbol{\theta}_0-\tilde{\boldsymbol{f}}_i^\top\boldsymbol{\gamma}_0\right)-\frac{1}{n}\sum_{i=1}^{n}\tilde{\boldsymbol{Z}}_i\left(\tilde{\boldsymbol{F}}_i^\top\boldsymbol{\gamma}_0-\tilde{\boldsymbol{f}}_i^\top\boldsymbol{\gamma}_0\right)\\
&= \frac{1}{n}\sum_{i=1}^{n}\tilde{\boldsymbol{z}}_i\left(\varepsilon_{i0}+\boldsymbol{L}_i^\top\boldsymbol{\theta}_0-\tilde{\boldsymbol{l}}_i^\top\boldsymbol{\theta}_0\right)+\frac{1}{n}\sum_{i=1}^{n}\{\tilde{\boldsymbol{Z}}_i-\tilde{\boldsymbol{z}}_i\}\left(\varepsilon_{i0}+\boldsymbol{L}_i^\top\boldsymbol{\theta}_0-\tilde{\boldsymbol{l}}_i^\top\boldsymbol{\theta}_0\right)\\
&+ \frac{1}{n}\sum_{i=1}^{n}\tilde{\boldsymbol{z}}_i\left(\tilde{l}_i^\top\boldsymbol{\theta}_0-\tilde{f}_i^\top\boldsymbol{\gamma}_0\right)+\frac{1}{n}\sum_{i=1}^{n}\{\tilde{\boldsymbol{Z}}_i-\tilde{\boldsymbol{z}}_i\}\left(\tilde{l}_i^\top\boldsymbol{\theta}_0-\tilde{f}_i^\top\boldsymbol{\gamma}_0\right)\\
&- \frac{1}{n}\sum_{i=1}^{n}\tilde{\boldsymbol{z}}_i\left(\tilde{\boldsymbol{F}}_i^\top\boldsymbol{\gamma}_0-\tilde{\boldsymbol{f}}_i^\top\boldsymbol{\gamma}_0\right)-\frac{1}{n}\sum_{i=1}^{n}\{\tilde{\boldsymbol{Z}}_i-\tilde{\boldsymbol{z}}_i\}\left(\tilde{\boldsymbol{F}}_i^\top\boldsymbol{\gamma}_0-\tilde{\boldsymbol{f}}_i^\top\boldsymbol{\gamma}_0\right)\\
&= \frac{1}{n}\sum_{i=1}^{n}\tilde{\boldsymbol{z}}_i\left(\varepsilon_{i0}+\boldsymbol{L}_i^\top\boldsymbol{\theta}_0-\tilde{\boldsymbol{l}}_i^\top\boldsymbol{\theta}_0\right)+o_p(n^{-1/2}).
\end{aligned}
$$

The last equality follows from the Lemma 1 in [4], Assumption (9) and Assumption (10) in the main paper. Hence, we can obtain $\sqrt{n}(\dot{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)\to N(\mathbf{0},\mathbf{A}^{-1}\boldsymbol{\Sigma}(\mathbf{A}^{-1})^\top)$ in distribution as $n\to\infty$, where

$$
\mathbf{A}=E(\tilde{\boldsymbol{z}}_i\tilde{\boldsymbol{z}}_i^\top),\quad \boldsymbol{\Sigma}=\mathrm{Var}\left\{\tilde{\boldsymbol{z}}_i\left(\varepsilon_{i0}+\boldsymbol{L}_i^\top\boldsymbol{\theta}_0-\tilde{\boldsymbol{l}}_i^\top\boldsymbol{\theta}_0\right)\right\}.
$$

1.3 Proof of Theorem 1

Noting that

$$
\hat{U}(\hat{\boldsymbol{\gamma}})-\hat{U}(\dot{\boldsymbol{\gamma}})=\left.\frac{\partial\hat{U}(\boldsymbol{\gamma})}{\partial\boldsymbol{\gamma}}\right|_{\boldsymbol{\gamma}=\bar{\boldsymbol{\gamma}}}(\hat{\boldsymbol{\gamma}}-\dot{\boldsymbol{\gamma}}),
$$

where $\hat{U}(\hat{\boldsymbol{\gamma}}) = 1/n\sum_{i=1}^n \tilde{\boldsymbol{F}}_i(Y_i - \tilde{\boldsymbol{F}}_i^\top \hat{\boldsymbol{\gamma}})$, $\hat{U}(\dot{\boldsymbol{\gamma}}) = 1/n\sum_{i=1}^n \tilde{\boldsymbol{F}}_i(Y_i - \tilde{\boldsymbol{F}}_i^\top \dot{\boldsymbol{\gamma}})$ and $\bar{\boldsymbol{\gamma}}$ is between $\hat{\boldsymbol{\gamma}}$ and $\dot{\boldsymbol{\gamma}}$. Thus, by the fact that $\hat{U}(\hat{\boldsymbol{\gamma}}) = 0$, we have

$$\hat{\boldsymbol{\gamma}} - \dot{\boldsymbol{\gamma}} = -\left(\left.\frac{\partial \hat{U}(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}}\right|_{\boldsymbol{\gamma}=\bar{\boldsymbol{\gamma}}}\right)^{-1} \hat{U}(\dot{\boldsymbol{\gamma}}).$$

We rewrite

$$\hat{U}(\dot{\boldsymbol{\gamma}}) = \left\{\hat{U}(\boldsymbol{b})^\top, \hat{U}(\boldsymbol{\beta})^\top\right\}^\top \Big|_{\boldsymbol{\gamma}=\dot{\boldsymbol{\gamma}}},$$

where

$$\hat{U}(\boldsymbol{b})\Big|_{\boldsymbol{\gamma}=\dot{\boldsymbol{\gamma}}} = \frac{1}{n}\sum_{i=1}^n \tilde{\boldsymbol{\alpha}}(\boldsymbol{X}_i)(Y_i - \tilde{\boldsymbol{F}}_i^\top \dot{\boldsymbol{\gamma}}), \quad \hat{U}(\boldsymbol{\beta})\Big|_{\boldsymbol{\gamma}=\dot{\boldsymbol{\gamma}}} = \frac{1}{n}\sum_{i=1}^n \tilde{\boldsymbol{Z}}_i(Y_i - \tilde{\boldsymbol{F}}_i^\top \dot{\boldsymbol{\gamma}}).$$

By the results in the proof of Lemma 1, we have

$$\frac{1}{n}\sum_{i=1}^n \tilde{\boldsymbol{Z}}_i(Y_i - \tilde{\boldsymbol{F}}_i^\top \dot{\boldsymbol{\gamma}}) = E\{\tilde{\boldsymbol{z}}_i(\boldsymbol{Z}_i^\top \boldsymbol{\beta}_0 - \tilde{\boldsymbol{z}}_i^\top \boldsymbol{\beta}_0)\} - E\{\tilde{\boldsymbol{z}}_i \tilde{\boldsymbol{z}}_i^\top\}(\dot{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) = o_p(1).$$

Following the proof of Lemma 1 above and applying the proof skills of Lemma A.3 in [5], we can also obtain

$$\frac{1}{n}\sum_{i=1}^n \tilde{\boldsymbol{\alpha}}(\boldsymbol{X}_i)(Y_i - \tilde{\boldsymbol{F}}_i^\top \dot{\boldsymbol{\gamma}}) = o_p(1).$$

According to Lemma A.2 in [5], $\left\|\left(\left.\frac{\partial \hat{U}(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}}\right|_{\boldsymbol{\gamma}=\bar{\boldsymbol{\gamma}}}\right)^{-1}\right\| = O_p(1)$. Hence,

$$\|\hat{\boldsymbol{\gamma}} - \dot{\boldsymbol{\gamma}}\| \le \left\|\left(\left.\frac{\partial \hat{U}(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}}\right|_{\boldsymbol{\gamma}=\bar{\boldsymbol{\gamma}}}\right)^{-1}\right\| \|\hat{U}(\dot{\boldsymbol{\gamma}})\| = o_p(1).$$

Then the theorem is proved.

1.4 Proof of Lemma 3

Define $CP(\boldsymbol{w}) = \|\boldsymbol{Y}_S - \dot{\boldsymbol{\mu}}_S\|^2 + 2\mathrm{tr}\{\mathbf{P}_S(\boldsymbol{w})\Sigma_S\}$, where $\Sigma_S = \mathrm{Var}(\boldsymbol{\varepsilon}_S)$, $\boldsymbol{\varepsilon}_S = (\varepsilon_1, \cdots, \varepsilon_{n_0})^\top$, $\mathbf{P}_S(\boldsymbol{w}) = \sum_{j=1}^{p+q} w_j \mathbf{P}_{S(j)}$ and $\hat{\boldsymbol{w}}^{cp} = \arg\min_{\boldsymbol{w}\in\mathcal{H}} CP(\boldsymbol{w})$. Following the proof of Lemma 1 in [6] and the second result of Theorem 2 of [8], under assumptions (A9) and (A10), we can obtain that

$$\frac{\dot{L}_n(\hat{\boldsymbol{w}}^{cp})}{\inf_{\boldsymbol{w}\in\mathcal{H}} \dot{L}_n(\boldsymbol{w})} \xrightarrow{p} 1 \text{ and } \sup_{\boldsymbol{w}\in\mathcal{H}} \left|\frac{\dot{L}_n(\boldsymbol{w})}{\dot{R}_n(\boldsymbol{w})} - 1\right| \xrightarrow{p} 0. \tag{3}$$

We define $\mathbf{P}^{cv}_{S(j)} = \mathbf{D}_{(j)}(\mathbf{P}_{S(j)} - \mathrm{I}_{n_0}) + \mathrm{I}_{n_0}$, $\mathbf{P}^{cv}_S(\boldsymbol{w}) = \sum_{j=1}^{p+q} w_j \mathbf{P}^{cv}_{S(j)}$ and $\hat{\boldsymbol{\mu}}^{cv}(\boldsymbol{w}) = \mathbf{P}^{cv}_S(\boldsymbol{w})\boldsymbol{Y}_S$. Note that $CV(\boldsymbol{w}) = \|\boldsymbol{Y}_S - \hat{\boldsymbol{\mu}}^{cv}(\boldsymbol{w})\|^2$ and $\mathrm{tr}\{\mathbf{P}^{cv}_S(\boldsymbol{w})\Sigma_S\} = 0$. Hence $CV(\boldsymbol{w})$, which can be expressed as the form of $CP(\boldsymbol{w})$, that is

$$CV(\boldsymbol{w}) = \|\boldsymbol{Y}_S - \hat{\boldsymbol{\mu}}^{cv}(\boldsymbol{w})\|^2 + 2\mathrm{tr}\{\mathbf{P}^{cv}_S(\boldsymbol{w})\Sigma_S)\}.$$

We also define

$$\dot{L}^{cv}_n(\boldsymbol{w}) = \|\boldsymbol{\mu}_S - \hat{\boldsymbol{\mu}}^{cv}_S(\boldsymbol{w})\|^2 \text{ and } \dot{R}^{cv}_n(\boldsymbol{w}) = E\{\dot{L}^{cv}_n(\boldsymbol{w})|\boldsymbol{X}_1, \cdots, \boldsymbol{X}_n\}.$$

To prove Lemma 3, according to the proof of Lemma 1 in [1], it suffices to show that

$$\sup_{\boldsymbol{w}\in\mathcal{H}} \left| \frac{\dot{R}^{cv}_n(\boldsymbol{w})}{\dot{R}_n(\boldsymbol{w})} - 1 \right| \xrightarrow{p} 0. \tag{4}$$

Following the proof of [1] with assumptions (A10) and (A11), we can obtain

$$\begin{aligned} \dot{R}^{cv}_n(\boldsymbol{w}) &= \|(\mathbf{I}_{n_0} - \mathbf{P}^{cv}_S(\boldsymbol{w}))\boldsymbol{\mu}_S\|^2 + \mathrm{tr}\{\mathbf{P}^{cv}_S(\boldsymbol{w})^2\Sigma_S\} \\ &= \dot{R}_n(\boldsymbol{w}) + \|(\mathbf{I}_{n_0} - \mathbf{P}_S)\boldsymbol{\mu}_S\|^2 o_p(1) + \mathrm{tr}\{[\mathbf{P}^{cv}_S(\boldsymbol{w})^2 - \mathbf{P}_S(\boldsymbol{w})^2]\Sigma_S\}. \end{aligned}$$

Thus

$$\left| \frac{\dot{R}^{cv}_n(\boldsymbol{w})}{\dot{R}_n(\boldsymbol{w})} - 1 \right| \le o_p(1) + \sup_{i\ge1} \sigma_i^2 \frac{\mathrm{tr}\,|\mathbf{P}^{cv}_S(\boldsymbol{w}) - \mathbf{P}_S(\boldsymbol{w})|}{\dot{R}_n(\boldsymbol{w})}$$

in which

$$\begin{aligned} &\mathrm{tr}\left|\mathbf{P}^{cv}_S(\boldsymbol{w})^2 - \mathbf{P}_S(\boldsymbol{w})^2\right| / \dot{R}_n(\boldsymbol{w}) \\ &\le \mathrm{tr}\left\{\mathbf{P}_S(\boldsymbol{w})^2\right\} o_p(1)/\dot{R}_n(\boldsymbol{w}) + 2\sum_{j=1}^{p+q}\sum_{j'=1}^{p+q} w_j w_{j'}(p+q+J_n+d)o_p(1)/\dot{R}_n(\boldsymbol{w}) \\ &+ \sum_{j=1}^{p+q}\sum_{j'=1}^{p+q} w_j w_{j'} \sum_{i=1}^{n_0} \frac{h^{ii}_{(j)}}{1-h^{ii}_{(j)}} \frac{h^{ii}_{(j')}}{1-h^{ii}_{(j')}} \Big/ \dot{R}_n(\boldsymbol{w}) \\ &= E_1 + E_2 + E_3. \end{aligned}$$

Recall that $\dot{\zeta}_n = \inf_{\boldsymbol{w}\in\mathcal{H}} \dot{R}_n(\boldsymbol{w})$. Thus, we have

$$E_1 \le \frac{\mathrm{tr}\left\{\mathbf{P}_S(\boldsymbol{w})^2\right\} o_p(1)}{\mathrm{tr}\left\{\mathbf{P}_S(\boldsymbol{w})^2\right\} \inf_{i\ge1}\sigma_i^2} = o_p(1),$$

$$E_2 \le 2(p+q+J_n+d)/\dot{\zeta}_n = O(1),$$

and

$$E_3 \le n_0 \left\{ \frac{\sup_j \lambda_{\max}\left(\mathbf{P}_{S(j)}\right)}{1-\sup_j \lambda_{\max}\left(\mathbf{P}_{S(j)}\right)} \right\}^2 \Big/ \dot{\zeta}_n \le n_0 \left(\frac{\kappa_2 n_0^{-1}(p+q+J_n+d)}{1-\kappa_2 n_0^{-1}(p+q+J_n+d)} \right)^2 \Big/ \dot{\zeta}_n$$

$$= \left(\frac{\kappa_2}{1-\kappa_2 n_0^{-1}(p+q+J_n+d)} \right)^2 \frac{p+q+J_n+d}{n_0} \frac{p+q+J_n+d}{\dot{\zeta}_n} = o(1)$$

Hence, we complete the proof of (4). According to the proof of Theorem 2.1 in [9], under assumption (A10), we can obtain

$$(p+q)(\dot{\zeta}_n^{cv})^{-2\kappa_1} \sum_{j=1}^{p+q} \{\dot{R}_n^{cv}(\boldsymbol{w}_j^0)\}^{\kappa_1} \xrightarrow{a.s.} 0, \quad \dot{\zeta}_n^{cv} = \inf_{\boldsymbol{w}\in\mathcal{H}} \dot{R}_n^{cv}(\boldsymbol{w})$$

Then we have

$$\frac{\dot{L}_n^{cv}(\boldsymbol{w})}{\inf_{\boldsymbol{w}\in\mathcal{H}} \dot{L}_n^{cv}(\boldsymbol{w})} \xrightarrow{p} 1 \quad \text{and} \quad \sup_{\boldsymbol{w}\in\mathcal{H}} \left| \frac{\dot{L}_n^{cv}(\boldsymbol{w})}{\dot{R}_n^{cv}(\boldsymbol{w})} - 1 \right| \xrightarrow{p} 0 \tag{5}$$

By the arguments of equation (3), we have $\dot{L}_n(\hat{\boldsymbol{w}}^{cv})/\dot{R}_n(\hat{\boldsymbol{w}}^{cv}) \xrightarrow{p} 1$ and $\dot{\zeta}_n/\inf_{\boldsymbol{w}\in\mathcal{H}} \dot{L}_n(\boldsymbol{w}) \xrightarrow{p} 1$, by equation (4), we have $\dot{R}_n(\hat{\boldsymbol{w}}^{cv})/\dot{R}_n^{cv}(\hat{\boldsymbol{w}}^{cv}) \xrightarrow{p} 1$ and $\dot{\zeta}_n^{cv}/\dot{\zeta}_n \xrightarrow{p} 1$. Also, by equation (5), we can obtain $\dot{R}_n^{cv}(\hat{\boldsymbol{w}}^{cv})/\dot{L}_n^{cv}(\hat{\boldsymbol{w}}^{cv}) \xrightarrow{p} 1$, $\dot{L}_n^{cv}(\hat{\boldsymbol{w}}^{cv})/\inf_{\boldsymbol{w}\in\mathcal{H}} \dot{L}_n^{cv}(\boldsymbol{w}) \xrightarrow{p} 1$ and $\inf_{\boldsymbol{w}\in\mathcal{H}} \dot{L}_n^{cv}(\boldsymbol{w})/\dot{\zeta}_n^{cv} \xrightarrow{p} 1$. Hence, the conclusion in Lemma 3 holds.

1.5 Proof of Theorem 2

To prove Theorem 2, follows from Lemma 3, we just need to prove that $\sup_{\boldsymbol{w}\in\mathcal{H}} |\{\dot{L}_n(\boldsymbol{w})-L_n(\boldsymbol{w})\}/\dot{L}_n(\boldsymbol{w})| \xrightarrow{p} 0$. Actually,

$$\begin{aligned}
&\sup_{\boldsymbol{w}\in\mathcal{H}} \left| \frac{\dot{L}_n(\boldsymbol{w})-L_n(\boldsymbol{w})}{\dot{L}_n(\boldsymbol{w})} \right| \\
&= \sup_{\boldsymbol{w}\in\mathcal{H}} \left| \frac{\|\hat{\boldsymbol{\mu}}_S(\boldsymbol{w})-\dot{\boldsymbol{\mu}}_S(\boldsymbol{w})\|^2 + 2\{\boldsymbol{\mu}_S-\hat{\boldsymbol{\mu}}_S(\boldsymbol{w})\}^\top \{\hat{\boldsymbol{\mu}}_S(\boldsymbol{w})-\dot{\boldsymbol{\mu}}_S(\boldsymbol{w})\}}{\dot{L}_n(\boldsymbol{w})} \right| \\
&\le \sup_{\boldsymbol{w}\in\mathcal{H}} \frac{\|\hat{\boldsymbol{\mu}}_S(\boldsymbol{w})-\dot{\boldsymbol{\mu}}_S(\boldsymbol{w})\|^2}{\dot{L}_n(\boldsymbol{w})} + 2\sup_{\boldsymbol{w}\in\mathcal{H}} \left| \frac{\{\boldsymbol{\mu}_S-\hat{\boldsymbol{\mu}}_S(\boldsymbol{w})\}^\top \{\hat{\boldsymbol{\mu}}_S(\boldsymbol{w})-\dot{\boldsymbol{\mu}}_S(\boldsymbol{w})\}}{\dot{L}_n(\boldsymbol{w})} \right| \\
&\le \sup_{\boldsymbol{w}\in\mathcal{H}} \frac{\|\hat{\boldsymbol{\mu}}_S(\boldsymbol{w})-\dot{\boldsymbol{\mu}}_S(\boldsymbol{w})\|^2}{\dot{L}_n(\boldsymbol{w})} + 2\sup_{\boldsymbol{w}\in\mathcal{H}} \sqrt{\frac{L_n(\boldsymbol{w})}{\dot{L}_n(\boldsymbol{w})}} \sup_{\boldsymbol{w}\in\mathcal{H}} \sqrt{\frac{\|\hat{\boldsymbol{\mu}}_S(\boldsymbol{w})-\dot{\boldsymbol{\mu}}_S(\boldsymbol{w})\|^2}{\dot{L}_n(\boldsymbol{w})}} \\
&\le \sup_{\boldsymbol{w}\in\mathcal{H}} \frac{\|\hat{\boldsymbol{\mu}}_S(\boldsymbol{w})-\dot{\boldsymbol{\mu}}_S(\boldsymbol{w})\|^2}{\dot{L}_n(\boldsymbol{w})} + 2\sqrt{\sup_{\boldsymbol{w}\in\mathcal{H}} \left| \frac{\dot{L}_n(\boldsymbol{w})-L_n(\boldsymbol{w})}{\dot{L}_n(\boldsymbol{w})} \right| + 1}\sqrt{\sup_{\boldsymbol{w}\in\mathcal{H}} \frac{\|\hat{\boldsymbol{\mu}}_S(\boldsymbol{w})-\dot{\boldsymbol{\mu}}_S(\boldsymbol{w})\|^2}{\dot{L}_n(\boldsymbol{w})}}.
\end{aligned} \tag{6}$$

Hence, it is sufficient to show that $\sup_{\boldsymbol{w}\in\mathcal{H}} \frac{\|\hat{\boldsymbol{\mu}}_S(\boldsymbol{w})-\dot{\boldsymbol{\mu}}_S(\boldsymbol{w}))\|^2}{\dot{L}_n(\boldsymbol{w})} \xrightarrow{p} 0$.

By (3), we have

$$
\begin{aligned}
&\sup_{\boldsymbol{w}\in\mathcal{H}} \frac{\left\|\hat{\boldsymbol{\mu}}_S(\boldsymbol{w})-\dot{\boldsymbol{\mu}}_S(\boldsymbol{w})\right\|^2}{\dot{L}_n(\boldsymbol{w})}\\
&\le \sup_{\boldsymbol{w}\in\mathcal{H}} \left\|\hat{\boldsymbol{\mu}}_S(\boldsymbol{w})-\dot{\boldsymbol{\mu}}_S(\boldsymbol{w})\right\|^2 \dot{\zeta}_n^{-1} O_p(1)\\
&= \sup_{\boldsymbol{w}\in\mathcal{H}} \left\|\sum_{j=1}^{p+q} w_j \mathbf{G}_{S(j)}\left(\hat{\boldsymbol{\gamma}}_{(j)}-\dot{\boldsymbol{\gamma}}_{(j)}\right)\right\|^2 \dot{\zeta}_n^{-1} O_p(1)\\
&\le \sup_{\boldsymbol{w}\in\mathcal{H}} \sum_{j=1}^{p+q} w_j \left\|\mathbf{G}_{S(j)}\left(\hat{\boldsymbol{\gamma}}_{(j)}-\dot{\boldsymbol{\gamma}}_{(j)}\right)\right\|^2 \dot{\zeta}_n^{-1} O_p(1)\\
&\le \max_{1\le j\le p+q} \left\|\mathbf{G}_{S(j)}\left(\hat{\boldsymbol{\gamma}}_{(j)}-\dot{\boldsymbol{\gamma}}_{(j)}\right)\right\|^2 \dot{\zeta}_n^{-1} O_p(1)\\
&= \max_{1\le j\le p+q} \left(\hat{\boldsymbol{\gamma}}_{(j)}-\dot{\boldsymbol{\gamma}}_{(j)}\right)^\top \mathbf{G}_{S(j)}^\top \mathbf{G}_{S(j)} \left(\hat{\boldsymbol{\gamma}}_{(j)}-\dot{\boldsymbol{\gamma}}_{(j)}\right) \dot{\zeta}_n^{-1} O_p(1).\\
&\le \max_{1\le j\le p+q} \lambda_{\max}\left(n_0^{-1}\mathbf{G}_{S(j)}^\top \mathbf{G}_{S(j)}\right) \left\|\hat{\boldsymbol{\gamma}}_{(j)}-\dot{\boldsymbol{\gamma}}_{(j)}\right\|^2 n_0 \dot{\zeta}_n^{-1} O_p(1),
\end{aligned}
$$

where $\dot{\boldsymbol{\gamma}}_{(j)} = (\mathbf{G}_{S(j)}^\top \mathbf{G}_{S(j)})^{-1}\mathbf{G}_{S(j)}^\top \boldsymbol{Y}_S$, $\hat{\boldsymbol{\gamma}}_{(j)} = (\tilde{\mathbf{G}}_{(j)}^\top \tilde{\mathbf{G}}_{(j)})^{-1}\tilde{\mathbf{G}}_{(j)}^\top \boldsymbol{Y}$ and $\tilde{\mathbf{G}}_{(j)} = (\tilde{\boldsymbol{G}}_{1(j)},\cdots,\tilde{\boldsymbol{G}}_{n(j)})^\top$, which has been defined in the main paper. It follows from above that the desired Theorem 2 holds if we can prove that $\left\|\hat{\boldsymbol{\gamma}}_{(j)}-\dot{\boldsymbol{\gamma}}_{(j)}\right\|^2 = O_p(n^{-1/2})$. Acually, the j-th approximating model mentioned above is

$$
Y_i = \mu_{i(j)} + \varepsilon_{i(j)} = g_j(X_{ij}) + \sum_{k\ne j}^{p+q} X_{ik}\beta_k + \varepsilon_{i(j)}, \quad i=1,\cdots,n. \tag{7}
$$

To estimate $g_j(X_{ij})$ in (7), we employ the basis approximation, and the model (7) can be approximated by

$$
Y_i = \boldsymbol{G}_{i(j)}^\top \boldsymbol{\gamma}_{(j)}^\star + \varepsilon_{i(j)}, \quad i=1,2,\cdots,n.
$$

where $\boldsymbol{G}_{i(j)} = \left(\boldsymbol{a}_j(X_{ij})^\top, \boldsymbol{X}_{i(-j)}^\top\right)^\top$, $\boldsymbol{X}_{i(-j)} = (X_{ij} : j=1,\cdots,j-1,j+1,\cdots,p+q)^\top$, $\boldsymbol{\gamma}_{(j)}^\star = (\boldsymbol{b}_j^{\star\top}, \boldsymbol{\beta}_{(-j)}^{\star\top})^\top$, $\boldsymbol{b}_j^\star = (b_{j,l}^\star : l=1,\cdots,L)^\top$, $\boldsymbol{\beta}_{(-j)}^\star = (\beta_j^\star : j=1,\cdots,j-1,j+1,\cdots,p+q)^\top$ and $\boldsymbol{\gamma}_{(j)}^\star$ denoting the pseudo-true parameter vector.

We give some notation which may be useful in the following process. Denote $\tilde{\boldsymbol{g}}_{i(j)} = (\tilde{\boldsymbol{a}}_{(j)}(X_{ij})^\top, \tilde{\boldsymbol{x}}_{i(-j)}^\top)^\top$, $\tilde{\boldsymbol{a}}_{(j)}(X_{ij}) = (a_{j,l}(X_{ij})I(j\in C_i) + E(a_{j,l}|\boldsymbol{X}_{i,C_i})I(j\notin C_i) : l=1,\cdots,L)^\top$, $\tilde{\boldsymbol{x}}_{i(-j)} = (X_{ij}I(j\in C_i) + W_{ij}I(j\notin C_i) : j=1,\cdots,j-1,j+1,\cdots,p+q)^\top$. Recall that $\tilde{\boldsymbol{G}}_{i(j)} = (\tilde{\boldsymbol{\alpha}}_j(X_{ij})^\top, \tilde{\boldsymbol{X}}_{i(-j)}^\top)^\top$, $\tilde{\boldsymbol{\alpha}}_j(X_{ij}) = (a_{j,l}(X_{ij})I(j\in C_i) + \hat{a}_{j,l}I(j\notin C_i) : l=1,\cdots,L)^\top$, $\tilde{\boldsymbol{X}}_{i(-j)} = (X_{ij}I(j\in C_i) + \hat{W}_{ij}I(j\notin C_i) : j=1,\cdots,j-1,j+1,\cdots,p+q)^\top$.

Next, we prove that $u(\boldsymbol{\gamma}_{(j)}^{\star}) = 0$, where $u(\boldsymbol{\gamma}_{(j)}) = E\left\{\tilde{\boldsymbol{g}}_{i(j)}\left(\boldsymbol{G}_{i(j)}^{\top}\boldsymbol{\gamma}_{(j)}^{\star} - \tilde{\boldsymbol{g}}_{i(j)}^{\top}\boldsymbol{\gamma}_{(j)}\right)\right\}$.

Note that

$$E(Y_i|\boldsymbol{X}_{i,C_i}) = \tilde{\boldsymbol{g}}_{i(j)}^{\top}\boldsymbol{\gamma}_{(j)}^{\star},$$

This in cojunction with the following equation

$$E(Y_i|\boldsymbol{X}_{i,C_i}) = E(\boldsymbol{G}_{i(j)}^{\top}|\boldsymbol{X}_{i,C_i})\boldsymbol{\gamma}_{(j)}^{\star},$$

we have $E\left(\boldsymbol{G}_{i(j)}^{\top} - \tilde{\boldsymbol{g}}_{i(j)}^{\top}|\boldsymbol{X}_{i,C_i}\right)\boldsymbol{\gamma}_{(j)}^{\star} = 0$ and then $E\left\{\tilde{\boldsymbol{g}}_{i(j)}\left(\boldsymbol{G}_{i(j)}^{\top}\boldsymbol{\gamma}_{(j)}^{\star} - \tilde{\boldsymbol{g}}_{i(j)}^{\top}\boldsymbol{\gamma}_{(j)}^{\star}\right)\right\} = 0$. Using similar arguments in Lemma 1, we can conclude that $\boldsymbol{\gamma}_{(j)}^{\star}$ is the unique solution of $u(\boldsymbol{\gamma}_{(j)}) = 0$. Let $\hat{\gamma}_{(j)}$ be the estimator as a solution of the estimation equation $U(\boldsymbol{\gamma}_{(j)}) = \frac{1}{n}\sum_{i=1}^{n}\tilde{\boldsymbol{G}}_{i(j)}\left\{Y_i - \tilde{\boldsymbol{G}}_{i(j)}^{\top}\boldsymbol{\gamma}_{(j)}\right\} = 0$.

Noting that

$$\begin{aligned}
U(\boldsymbol{\gamma}_{(j)}) - u(\boldsymbol{\gamma}_{(j)}) &= \frac{1}{n}\sum_{i=1}^{n}\tilde{\boldsymbol{G}}_{i(j)}\left\{Y_i - \tilde{\boldsymbol{G}}_{i(j)}^{\top}\boldsymbol{\gamma}_{(j)}\right\} - E\{\tilde{\boldsymbol{g}}_{i(j)}(\boldsymbol{G}_{i(j)}^{\top}\boldsymbol{\gamma}_{(j)}^{\star} - \tilde{\boldsymbol{g}}_{i(j)}^{\top}\boldsymbol{\gamma}_{(j)})\} \\
&= \frac{1}{n}\sum_{i=1}^{n}\left(\tilde{\boldsymbol{G}}_{i(j)} - \tilde{\boldsymbol{g}}_{i(j)}\right)Y_i - \frac{1}{n}\sum_{i=1}^{n}\left(\tilde{\boldsymbol{G}}_{i(j)}\tilde{\boldsymbol{G}}_{i(j)}^{\top} - \tilde{\boldsymbol{g}}_{i(j)}\tilde{\boldsymbol{g}}_{i(j)}^{\top}\right)\boldsymbol{\gamma}_{(j)} + \frac{1}{n}\sum_{i=1}^{n}\tilde{\boldsymbol{g}}_{i(j)}\varepsilon_{i(j)} \\
&+ \frac{1}{n}\sum_{i=1}^{n}\tilde{\boldsymbol{g}}_{i(j)}\left\{\boldsymbol{G}_{i(j)}^{\top}\boldsymbol{\gamma}_{(j)}^{\star} - \tilde{\boldsymbol{g}}_{i(j)}^{\top}\boldsymbol{\gamma}_{(j)}\right\} - E\left\{\tilde{\boldsymbol{g}}_{i(j)}\left(\boldsymbol{G}_{i(j)}^{\top}\boldsymbol{\gamma}_{(j)}^{\star} - \tilde{\boldsymbol{g}}_{i(j)}^{\top}\boldsymbol{\gamma}_{(j)}\right)\right\}.
\end{aligned}$$

Analogous to the proof of Lemma 1, for a given j, we can obtain that $\|U(\boldsymbol{\gamma}_{(j)}) - u(\boldsymbol{\gamma}_{(j)})\| = o_p(1)$, which yields that $\hat{\gamma}_{(j)} \xrightarrow{p} \boldsymbol{\gamma}_{(j)}^{\star}$.

Also, using the similar arguments in Lemma 2, we have

$$U(\hat{\gamma}_{(j)}) - U(\boldsymbol{\gamma}_{(j)}^{\star}) = -\frac{1}{n}\sum_{i=1}^{n}\tilde{\boldsymbol{g}}_{i(j)}\tilde{\boldsymbol{g}}_{i(j)}^{\top}\left(\hat{\gamma}_{(j)} - \boldsymbol{\gamma}_{(j)}^{\star}\right) + o_p(\|\hat{\gamma}_{(j)} - \boldsymbol{\gamma}_{(j)}^{\star}\|),$$

which implies $\sqrt{n}\left(\hat{\gamma}_{(j)} - \boldsymbol{\gamma}_{(j)}^{\star}\right) = \left(\frac{1}{n}\sum_{i=1}^{n}\tilde{\boldsymbol{g}}_{i(j)}\tilde{\boldsymbol{g}}_{i(j)}^{\top}\right)^{-1}\sqrt{n}U(\boldsymbol{\gamma}_{(j)}^{\star}) + o_p(1)$, in which

$$\begin{aligned}
U(\boldsymbol{\gamma}_{(j)}^{\star}) &= \frac{1}{n}\sum_{i=1}^{n}\tilde{\boldsymbol{G}}_{i(j)}\left(\varepsilon_{i(j)} + \boldsymbol{G}_{i(j)}^{\top}\boldsymbol{\gamma}_{(j)}^{\star} - \tilde{\boldsymbol{G}}_{i(j)}^{\top}\boldsymbol{\gamma}_{(j)}^{\star}\right) \\
&= \frac{1}{n}\sum_{i=1}^{n}\tilde{\boldsymbol{g}}_{i(j)}\left(\varepsilon_{i(j)} + \boldsymbol{G}_{i(j)}^{\top}\boldsymbol{\gamma}_{(j)}^{\star} - \tilde{\boldsymbol{g}}_{i(j)}^{\top}\boldsymbol{\gamma}_{(j)}^{\star}\right) + o_p(n^{-1/2}).
\end{aligned}$$

Then, we have $\sqrt{n}(\hat{\gamma}_{(j)} - \boldsymbol{\gamma}_{(j)}^{\star}) \to N(\mathbf{0}, \tilde{\mathbf{A}}^{-1}\tilde{\mathbf{\Sigma}}(\tilde{\mathbf{A}}^{-1})^{\top})$ in distribution as $n \to \infty$, where $\tilde{\mathbf{A}} = E(\tilde{\boldsymbol{g}}_{i(j)}\tilde{\boldsymbol{g}}_{i(j)}^{\top})$ and $\tilde{\mathbf{\Sigma}} = \mathrm{Var}\left\{\tilde{\boldsymbol{g}}_{i(j)}\left(\varepsilon_{i(j)} + \boldsymbol{G}_{i(j)}^{\top}\boldsymbol{\gamma}_{(j)}^{\star} - \tilde{\boldsymbol{g}}_{i(j)}^{\top}\boldsymbol{\gamma}_{(j)}^{\star}\right)\right\}$.

It implies that $\left\|\hat{\boldsymbol{\gamma}}_{(j)}-\boldsymbol{\gamma}^{\star}_{(j)}\right\|=O_p(n^{-1/2})$. For we obtain $\dot{\boldsymbol{\gamma}}_{(j)}$ using the complete case data with sample size n_0, we can easily show that $\left\|\dot{\boldsymbol{\gamma}}_{(j)}-\boldsymbol{\gamma}^{\star}_{(j)}\right\|=O_p(n^{-1/2})$. Therefore, we can obtain $\left\|\hat{\boldsymbol{\gamma}}_{(j)}-\dot{\boldsymbol{\gamma}}_{(j)}\right\|=O_p(n^{-1/2})$, and we complete the proof of Theorem 2.

2 Additional numerical results

In this case, we compare the computing time of PRIME, PRIME-MA, ILSE, SSL, ML, MI and CC. For the sake of fairness, we record the time that the methods mentioned above cost in Scenario 1 with $\rho=0.6$, $R^2=0.7$ and heteroscedastic error in a laptop with 2.4 GHz Intel Core i5 and 8GB of RAM. We repeat the simulation $N=100$ times. The results are shown in Table 1 with different sample sizes and missing rates. PRIME, ILSE and SSL are more time-consuming than other methods. Among these three methods, Selecting scale parameters with a grid-search over the prespecified bounded region force SSL to spend more time in estimation than PRIME and ILSE. Our proposed new PRIME-MA strategy is computationally more demanding due to the use of multiple candidates. ML and CC take less time but have low accuracy. For applications where the method is used to quickly obtain an estimation result in a time-sensitive manner, it is perhaps enough or even preferred to use, e.g., MI. However, for applications where accuracy is more important, the extra computation cost of PRIME/PRIME-MA may be worthwhile, especially when the correct model structure of the partially linear model is unknown.

Table 1: Time Cost

MR	n	MI	PRIME	ILSE	SSL	ML	CC	PRIME-MA
60%	200	0.535	1.333	1.130	1.510	0.040	0.010	8.110
	400	0.681	3.861	3.214	4.724	0.013	0.012	22.121
85%	200	0.595	1.392	1.759	1.811	0.035	0.034	8.310
	400	0.758	4.036	4.037	4.750	0.041	0.036	24.019

References

1. Fang F, Lan W, Tong J and Shao J (2019) Model averaging for prediction with fragmentary data.

Journal of Business and Economic Statistics 37(3): 517–527.

2. Le Q, Sarlós T and Smola A (2013) Fastfood: approximating kernel expansions in loglinear time. Proceedings of the 30th International Conference on International Conference on Machine Learning.

3. Li Y (2011) Efficient semiparametric regression for longitudinal data with nonparametric covariance estimation. *Biometrika* : 355–370.

4. Lin H, Liu W and Lan W (2019) Regression analysis with individual-specific patterns of missing covariates. *Journal of Business and Economic Statistics* 19(3): 231–253.

5. Liu X, Wang L and Liang H (2011) Estimation and variable selection for semiparametric additive partial linear models. *Statistica Sinica* 21(3): 1225.

6. Xia X (2021) Model averaging prediction for nonparametric varying-coefficient models with b-spline smoothing. *Statistical Papers* 62(6): 2885-2905.

7. Van der Vaart AW (2000) *Asymptotic statistics.* Cambridge University press.

8. Whittle P (1960) Bounds for the moments of linear and quadratic forms in independent variables. *Theory of Probability & Its Applications* 5(3): 302–305.

9. Zhang X (2013) Model averaging with covariates that are missing completely at random. *Economics Letters* 121(3): 360–363.