Supplementary Materials to "Regression Analysis with individual-specific patterns of missing covariates"

Huazhen Lin, Wei Liu, Wei Lan

Center of Statistical Research and School of Statistics,

Southwestern University of Finance and Economics, Chengdu, Sichuan, China

In this supplementary material, we give proofs of theorem 1 to theorem 3 and some additional simulation results.

We first review and introduce the notation. For any $\boldsymbol{\beta} \in \Omega$, define

$$\widehat{E}_{j,P_i}(w;\boldsymbol{\beta}) = \frac{\sum_{i'=1}^{n} I(P_{i'} \supset P_{i'} \cup j) X_{i'j} K_h\left(\mathbf{X}_{i',P_i}^{\mathrm{T}} \beta_{P_i} - w\right)}{\sum_{i'=1}^{n} I(P_{i'} \supset P_i \cup j) K_h\left(\mathbf{X}_{i',P_i}^{\mathrm{T}} \beta_{P_i} - w\right)}, \quad (j \in \bar{P}_i)$$

and we can rewrite $\mathbf{Z}_i(\boldsymbol{\beta})$ and $\tilde{\mathbf{Z}}_i(\boldsymbol{\beta})$ used in the main text as

$$\mathbf{Z}_i(\boldsymbol{\beta}) = (X_{ij} I(j \in P_i) + \widehat{E}_{j,P_i}(\tilde{\mathbf{X}}_i^{\mathrm{T}}\boldsymbol{\beta};\boldsymbol{\beta}) I(j \notin P_i) : 1 \le j \le p)$$

and

$$\tilde{\mathbf{Z}}_i(\boldsymbol{\beta}) = (\widehat{E}_{j,P_i}(\tilde{\mathbf{X}}_i^{\mathrm{T}}\boldsymbol{\beta};\boldsymbol{\beta}) : 1 \le j \le p).$$

Recall

$$U(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^{n} \tilde{\mathbf{Z}}_i(\boldsymbol{\beta})\big\{Y_i - \mathbf{Z}_i(\boldsymbol{\beta})^{\mathrm{T}}\boldsymbol{\beta}\big\}, \quad u(\boldsymbol{\beta}) = E\left[\tilde{\mathbf{z}}_i(\boldsymbol{\beta})\big\{\mathbf{X}_i^{\mathrm{T}}\boldsymbol{\beta}_0 - \mathbf{z}_i(\boldsymbol{\beta})^{\mathrm{T}}\boldsymbol{\beta}\big\}\right].$$

Recall $e_{j,P}(w;\boldsymbol{\beta}) = E(X_{ij} \mid \mathbf{X}_{i,P}^{\mathrm{T}}\beta_P = w)$, $\tilde{\mathbf{z}}_i(\boldsymbol{\beta}) = (e_{j,P_i}(\mathbf{X}_{i,P_i}^{\mathrm{T}}\beta_{P_i};\boldsymbol{\beta}) : 1 \le j \le p)^{\mathrm{T}}$ and

$$\mathbf{z}_i(\boldsymbol{\beta}) = (X_{ij} I(j \in P_i) + e_{j,P_i}(\tilde{\mathbf{X}}_i^{\mathrm{T}}\boldsymbol{\beta};\boldsymbol{\beta}) I(j \notin P_i) : 1 \le j \le p)^{\mathrm{T}}.$$

Let $\ddot{e}_{k,P}(w;\boldsymbol{\beta}) = \partial^2 e_{k,P}(w;\boldsymbol{\beta})/\partial w^2, \nu_2 = \int v^2 K(v)\mathrm{d}v$ and $\dot{K}_h(v) = \frac{dK_h(v)}{dv}$.

1 Proof of Proposition 1

Without loss of generality, given P_i, we rewrite $\mathbf{X}_i = (\mathbf{X}_{i,P_i}^{\mathrm{T}}, \mathbf{X}_{i,\bar{P}_i}^{\mathrm{T}})^{\mathrm{T}}$ and $\boldsymbol{\beta}_0 = (\beta_{P_i,0}, \beta_{\bar{P}_i,0})^{\mathrm{T}}$, then $\mathbf{z}_i(\boldsymbol{\beta}_0) = \left(\mathbf{X}_{i,P_i}^{\mathrm{T}}, E(\mathbf{X}_{i,\bar{P}_i}^{\mathrm{T}} | \mathbf{X}_{i,P_i}^{\mathrm{T}} \beta_{P_i,0}, P_i)\right)^{\mathrm{T}}$. First, we prove that $\boldsymbol{\beta}_0$ is a solution of $u(\boldsymbol{\beta}) = 0$. Taking the expectation on both sides of (2.2) given $\mathbf{X}_{i,P_i}^{\mathrm{T}} \beta_{P_i,0}$, we have

$$E\left(Y_i \mid \mathbf{X}_{i,P_i}^{\mathrm{T}} \beta_{P_i,0}, P_i\right) = \mathbf{X}_{i,P_i}^{\mathrm{T}} \beta_{P_i,0} + E(\mathbf{X}_{i,\bar{P}_i}^{\mathrm{T}} \mid \mathbf{X}_{i,P_i}^{\mathrm{T}} \beta_{P_i,0}, P_i) \beta_{\bar{P}_i,0} = \mathbf{z}_i(\boldsymbol{\beta}_0)^{\mathrm{T}} \boldsymbol{\beta}_0.$$

By the condition that $E(\epsilon_i \mid \mathbf{X}_i, P_i) = 0$, we have $E\left(Y_i \mid \mathbf{X}_{i,P_i}^{\mathrm{T}} \beta_{P_i,0}, P_i\right) = E(\mathbf{X}_i^{\mathrm{T}} | \mathbf{X}_{i,P_i}^{\mathrm{T}} \beta_{P_i,0}, P_i) \boldsymbol{\beta}_0$. Therefore, we get $\left\{E(\mathbf{X}_i^{\mathrm{T}} | \mathbf{X}_{i,P_i}^{\mathrm{T}} \beta_{P_i,0}, P_i) - \mathbf{z}_i(\boldsymbol{\beta}_0)^{\mathrm{T}}\right\} \boldsymbol{\beta}_0 = 0$, which implies $E\left\{(\mathbf{X}_i^{\mathrm{T}} - \mathbf{z}_i(\boldsymbol{\beta}_0)^{\mathrm{T}}) | \mathbf{X}_{i,P_i}^{\mathrm{T}} \beta_{P_i,0}, P_i\right\} \boldsymbol{\beta}_0 = 0$. Noting that $\tilde{\mathbf{z}}_i(\boldsymbol{\beta}_0)$ is a function of $(\mathbf{X}_{i,P_i}^{\mathrm{T}} \beta_{P_i,0}, P_i)$, we obtain

$$u(\boldsymbol{\beta}_0) = E\left\{\tilde{\mathbf{z}}_i(\boldsymbol{\beta}_0)(\mathbf{X}_i^{\mathrm{T}} \boldsymbol{\beta}_0 - \mathbf{z}_i(\boldsymbol{\beta}_0)^{\mathrm{T}} \boldsymbol{\beta}_0)\right\} = 0.$$

Next, we prove that $u(\boldsymbol{\beta}) = 0$ has a unique solution. We assume there is another solution $\tilde{\boldsymbol{\beta}} = (\tilde{\beta}_{P_i}, \tilde{\beta}_{\bar{P}_i})^{\mathrm{T}} \in \Omega$ such that $u(\tilde{\boldsymbol{\beta}}) = 0$. By the simple calculation, we have

$$\begin{aligned} u(\tilde{\boldsymbol{\beta}}) &= \begin{pmatrix} E\left[E(\mathbf{X}_{iP_i} | \mathbf{X}_{iP_i}^{\mathrm{T}} \tilde{\beta}_{P_i}) E(\mathbf{X}_{iP_i}^{\mathrm{T}} | \mathbf{X}_{iP_i}^{\mathrm{T}} \tilde{\beta}_{P_i})\right] & E\left[E(\mathbf{X}_{iP_i} | \mathbf{X}_{iP_i}^{\mathrm{T}} \tilde{\beta}_{P_i}) E(\mathbf{X}_{i\bar{P}_i}^{\mathrm{T}} | \mathbf{X}_{iP_i}^{\mathrm{T}} \tilde{\beta}_{P_i})\right] \\ E\left[E(\mathbf{X}_{iP_i} | \mathbf{X}_{iP_i}^{\mathrm{T}} \tilde{\beta}_{P_i}) E(\mathbf{X}_{i\bar{P}_i}^{\mathrm{T}} | \mathbf{X}_{iP_i}^{\mathrm{T}} \tilde{\beta}_{P_i})\right] & E\left[E(\mathbf{X}_{i\bar{P}_i} | \mathbf{X}_{iP_i}^{\mathrm{T}} \tilde{\beta}_{P_i}) E(\mathbf{X}_{i\bar{P}_i}^{\mathrm{T}} | \mathbf{X}_{iP_i}^{\mathrm{T}} \tilde{\beta}_{P_i})\right] \end{pmatrix} (\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \\ &= S(\tilde{\boldsymbol{\beta}})(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) = \mathbf{0}. \end{aligned}$$

By Conditions (C2) and (C4) that $S(\tilde{\boldsymbol{\beta}})$ is invertible, we have $\tilde{\boldsymbol{\beta}} = \boldsymbol{\beta}_0$. Combining the existence and uniqueness of the solution, we conclude that $\boldsymbol{\beta}_0$ is the unique solution of $u(\boldsymbol{\beta}) = 0$.

2 Proof of Theorem 1

Proof. Decomposing $U(\boldsymbol{\beta})$, we have

$$\begin{aligned} U(\boldsymbol{\beta})-u(\boldsymbol{\beta}) &= \frac{1}{n}\sum_{i=1}^{n}\tilde{\mathbf{Z}}_i(\boldsymbol{\beta})\big\{Y_i-\mathbf{Z}_i(\boldsymbol{\beta})^{\mathrm{T}}\boldsymbol{\beta}\big\}-E\big\{\tilde{\mathbf{z}}_i(\boldsymbol{\beta})(\mathbf{X}_i^{\mathrm{T}}\boldsymbol{\beta}_0-\mathbf{z}_i(\boldsymbol{\beta})^{\mathrm{T}}\boldsymbol{\beta})\big\} \\ &= \frac{1}{n}\sum_{i=1}^{n}(\tilde{\mathbf{Z}}_i(\boldsymbol{\beta})-\tilde{\mathbf{z}}_i(\boldsymbol{\beta}))Y_i-\frac{1}{n}\sum_{i=1}^{n}(\tilde{\mathbf{Z}}_i(\boldsymbol{\beta})\mathbf{Z}_i(\boldsymbol{\beta})^T-\tilde{\mathbf{z}}_i(\boldsymbol{\beta})\mathbf{z}_i(\boldsymbol{\beta})^T)\boldsymbol{\beta} \\ &+ \frac{1}{n}\sum_{i=1}^{n}\big[\tilde{\mathbf{z}}_i(\boldsymbol{\beta})\big\{\mathbf{X}_i^{\mathrm{T}}\boldsymbol{\beta}_0-\mathbf{z}_i(\boldsymbol{\beta})^{\mathrm{T}}\boldsymbol{\beta}\big\}-E\big\{\tilde{\mathbf{z}}_i(\boldsymbol{\beta})(\mathbf{X}_i^{\mathrm{T}}\boldsymbol{\beta}_0-\mathbf{z}_i(\boldsymbol{\beta})^{\mathrm{T}}\boldsymbol{\beta})\big\}\big]+\frac{1}{n}\sum_{i=1}^{n}\tilde{\mathbf{z}}_i(\boldsymbol{\beta})\varepsilon_{i0} \\ &= I_1-I_2+I_3+I_4. \end{aligned}$$

Then, we can get

$$\sup_{\boldsymbol{\beta}\in\Omega}\|U(\boldsymbol{\beta})-u(\boldsymbol{\beta})\|_2\le\sup_{\boldsymbol{\beta}\in\Omega}\|I_1\|_2+\sup_{\boldsymbol{\beta}\in\Omega}\|I_2\|_2+\sup_{\boldsymbol{\beta}\in\Omega}\|I_3\|_2+\sup_{\boldsymbol{\beta}\in\Omega}\|I_4\|_2.$$

Moreover, $\sup_{\boldsymbol{\beta}\in\Omega}\|I_3\|_2=o_p(1)$ can be showed by the fact that function class $\{f_{\boldsymbol{\beta}}(\mathbf{X}_i)=\tilde{\mathbf{z}}_i(\boldsymbol{\beta})(\mathbf{X}_i^{\mathrm{T}}\boldsymbol{\beta}_0-\mathbf{z}_i(\boldsymbol{\beta})^{\mathrm{T}}\boldsymbol{\beta}),\boldsymbol{\beta}\in\Omega\}$ indexed by $\boldsymbol{\beta}$ is a GC class (van der Vaart, 1998) under Conditions (C2)–(C4). Similarly, we can get uniform convergence on I_4.

Next, we prove that $\sup_{\boldsymbol{\beta}\in\Omega}\|I_1\|_2=o_p(1)$ can be proved under Conditions (C1)-(C6). Note

$$\widehat{E}_{j,P}(w;\boldsymbol{\beta})-e_{j,P}(w;\boldsymbol{\beta})=\frac{\sum_{i'=1}^{n}I(P_{i'}\supset P\cup j)\big\{X_{i'j}-e_{j,P}(w;\boldsymbol{\beta})\big\}K_h\left(\mathbf{X}_{i',P}^{\mathrm{T}}\beta_P-w\right)}{\sum_{i'=1}^{n}I(P_{i'}\supset P\cup j)K_h\left(\mathbf{X}_{i',P}^{\mathrm{T}}\beta_P-w\right)}\equiv\frac{S_{n1,j}(w;\boldsymbol{\beta},P)}{S_{n0,j}(w;\boldsymbol{\beta},P)},\quad(2.1)$$

where $S_{nr,j}(w;\boldsymbol{\beta},P)=n^{-1}\sum_{i'=1}^{n}I(P_{i'}\supset P\cup j)\left\{X_{i'j}-e_{j,P}(w;\boldsymbol{\beta})\right\}^{\otimes r}K_h\left(\mathbf{X}_{i',P}^{\mathrm{T}}\beta_P-w\right)$ for $r=0,1$. By the arguments in the proof of Lemma 4 in Chen et al. (2010), we can show that

$$\sup_{w,\boldsymbol{\beta}}\|S_{n1,1}(w;\boldsymbol{\beta},P)\|_2=O_p(\frac{\sqrt{\log n}}{\sqrt{nh}}+h^2),\quad(2.2)$$

$$\sup_{w,\boldsymbol{\beta}}|S_{nr,0}(w;\boldsymbol{\beta},P)-C(P\cup j)f_P(w;\boldsymbol{\beta})|=O_p(\frac{\sqrt{\log n}}{\sqrt{nh}}+h^2),\quad(2.3)$$

where $f_P(w;\boldsymbol{\beta})$ is the density function of $\mathbf{X}_{i,P}^{\mathrm{T}}\beta_P$, and $C(P\cup j)=\Pr(P_i\supset P\cup j)$. Thus, we obtain

$\sup_{\boldsymbol{\beta}\in\Omega} \|I_1\|_2 = o_p(1)$. Similarly, we can get uniform convergence on I_2, thus we have $\sup_{\boldsymbol{\beta}\in\Omega} \|U(\boldsymbol{\beta}) - u(\boldsymbol{\beta})\|_2 = 0$.

From Proposition 1, we know that $\boldsymbol{\beta}_0$ is the unique solution to $u(\boldsymbol{\beta}) = 0$. Therefore, $U(\widehat{\boldsymbol{\beta}}) = 0$ and $u(\boldsymbol{\beta}_0) = 0$. We next use a proof by contradiction. Suppose $\widehat{\boldsymbol{\beta}} \not\to \boldsymbol{\beta}_0$ in probability. Because $\{\widehat{\boldsymbol{\beta}}\} \subseteq \Omega$ is a bounded sequence, there must exist a subsequence $\{\widehat{\boldsymbol{\beta}}_n\}$ such that $\widehat{\boldsymbol{\beta}}_n \to \boldsymbol{\beta}^* \neq \boldsymbol{\beta}_0$ in probability by the Bolzano–Weierstrass theorem. Because $U(\widehat{\boldsymbol{\beta}}_n) = 0$, we have

$$u(\boldsymbol{\beta}^*) = U(\widehat{\boldsymbol{\beta}}_n) - \left\{U(\widehat{\boldsymbol{\beta}}_n) - u(\widehat{\boldsymbol{\beta}}_n)\right\} - \left\{u(\widehat{\boldsymbol{\beta}}_n) - u(\boldsymbol{\beta}^*)\right\} = -\left\{U(\widehat{\boldsymbol{\beta}}_n) - u(\widehat{\boldsymbol{\beta}}_n)\right\} - \left\{u(\widehat{\boldsymbol{\beta}}_n) - u(\boldsymbol{\beta}^*)\right\}.$$

Furthermore, by $\sup_{\boldsymbol{\beta}\in\Omega} \|U(\boldsymbol{\beta}) - u(\boldsymbol{\beta})\|_2 = o_p(1)$ and the continuous mapping theorem, we conclude $u(\boldsymbol{\beta}^*) = 0$ with $\boldsymbol{\beta}^* \neq \boldsymbol{\beta}_0$, which contradicts that $u(\boldsymbol{\beta}) = 0$ has a unique solution $\boldsymbol{\beta}_0$. Therefore, Theorem 1 holds.

3 Proof of Theorem 2

3.1 A lemma

Lemma 1 *Under Conditions (C1), (C2), (C3) and (C5), we have*

$$\mathbf{Z}_i(\boldsymbol{\beta}_0) - \mathbf{z}_i(\boldsymbol{\beta}_0) = O_p\{(nh)^{-1/2} + h^2\}, \tag{3.1}$$

$$\tilde{\mathbf{Z}}_i(\boldsymbol{\beta}_0) - \tilde{\mathbf{z}}_i(\boldsymbol{\beta}_0) = O_p\{(nh)^{-1/2} + h^2\}, \tag{3.2}$$

$$\frac{1}{n}\sum_{i=1}^{n}\{\tilde{\mathbf{Z}}_i(\boldsymbol{\beta}_0) - \tilde{\mathbf{z}}_i(\boldsymbol{\beta}_0)\}\epsilon_{i0} = o_p(n^{-1/2}), \tag{3.3}$$

$$\frac{1}{n}\sum_{i=1}^{n}\{\tilde{\mathbf{Z}}_i(\boldsymbol{\beta}_0) - \tilde{\mathbf{z}}_i(\boldsymbol{\beta}_0)\}(\tilde{\mathbf{X}}_i^c - \tilde{\mathbf{e}}_i^c)^{\mathrm{T}}\boldsymbol{\beta}_0 = o_p(n^{-1/2}), \tag{3.4}$$

$$\frac{1}{n}\sum_{i=1}^{n}\{\tilde{\mathbf{Z}}_i(\boldsymbol{\beta}_0) - \tilde{\mathbf{z}}_i(\boldsymbol{\beta}_0)\}(\tilde{\mathbf{E}}_i^c - \tilde{\mathbf{e}}_i^c)^{\mathrm{T}}\boldsymbol{\beta}_0 = o_p(n^{-1/2}), \tag{3.5}$$

$$\frac{1}{n}\sum_{i=1}^{n}\{\tilde{\mathbf{Z}}_i(\boldsymbol{\beta}_0) - \tilde{\mathbf{z}}_i(\boldsymbol{\beta}_0)\}\{\mathbf{Z}_i(\boldsymbol{\beta}_0) - \mathbf{z}_i(\boldsymbol{\beta}_0)\}^{\mathrm{T}}\boldsymbol{\beta}_0 = o_p(n^{-1/2}), \tag{3.6}$$

where we recall

$$\tilde{\mathbf{X}}_i^c = (X_{ij} I(j \notin P_i), 1 \le j \le p)^{\mathrm{T}}, \quad \tilde{\mathbf{e}}_i^c = (e_{j,P_i}(W_{i0}) I(j \notin P_i), 1 \le j \le p)^{\mathrm{T}}, \tag{3.7}$$

and define

$$\tilde{\mathbf{E}}_i^c = (\widehat{E}_{j,P_i}(W_{i0}) I(j \notin P_i), 1 \le j \le p)^{\mathrm{T}}, \quad e_{j,P_i}(W_{i0}) = e_{j,P_i}(W_{i0}; \boldsymbol{\beta}_0), \quad \widehat{E}_{j,P_i}(W_{i0}) = \widehat{E}_{j,P_i}(W_{i0}; \boldsymbol{\beta}_0). \tag{3.8}$$

Proof. Equations (3.1), (3.2), (3.5) and (3.6) follow from (2.2) under Conditions (C1), (C2), (C3) and (C5). We prove (3.3), and omit the proof of (3.4) due to the similarity.

For simplicity, we define

$$\begin{aligned}
\widehat{f}_{j,P_i}(W_{i0}) &= \frac{1}{n-1} \sum_{i' \neq i} I(P_{i'} \supset P_i \cup j) K_h(X_{i',P_i} \beta_{P_i,0} - W_{i0}), \\
\widehat{r}_{j,P_i}(W_{i0}) &= \frac{1}{n-1} \sum_{i' \neq i} I(P_{i'} \supset P_i \cup j) K_h(X_{i',P_i} \beta_{P_i,0} - W_{i0}) X_{i'j}, \\
r_{j,P_i}(W_{i0}) &= C(P_i \cup j) f_{P_i}(W_{i0}; \boldsymbol{\beta}_0) e_{j,P_i}(W_{i0}; \boldsymbol{\beta}_0), \\
f_{j,P_i}(W_{i0}) &= C(P_i \cup j) f_{P_i}(W_{i0}; \boldsymbol{\beta}_0).
\end{aligned}$$

Using some simple algebra and the kernel theory (Horowitz, 1996), we have

$$\begin{aligned}
&\frac{1}{n} \sum_{i=1}^{n} \{\widehat{E}_{j,P_i}(W_{i0}; \boldsymbol{\beta}_0) - e_{j,P_i}(W_{i0}; \boldsymbol{\beta}_0)\} \epsilon_{i0} \\
= \; &\frac{1}{n} \sum_{i=1}^{n} \epsilon_{i0} \left\{ \frac{\widehat{r}_{j,P_i}(W_{i0}) - r_{j,P_i}(W_{i0})}{f_{j,P_i}(W_{i0})} \right\} - \frac{1}{n} \sum_{i=1}^{n} \epsilon_{i0} \left\{ \frac{r_{j,P_i}(W_{i0})(\widehat{f}_{j,P_i}(W_{i0}) - f_{j,P_i}(W_{i0}))}{f_{j,P_i}^2(W_{i0})} \right\} \\
&+ O_p\{h^4 + 1/(nh)\}.
\end{aligned} \tag{3.9}$$

Define $\tilde{\epsilon}_{ij0} = \epsilon_{i0} / f_{j,P_i}(W_{i0})$, and write $n^{-1} \sum_{i=1}^{n} \tilde{\epsilon}_{ij0} \widehat{r}_{j,P_i}(W_{i0})$ as a second-order U-statistic:

$$\frac{1}{n} \sum_{i=1}^{n} \tilde{\epsilon}_{ij0} \widehat{r}_{j,P_i}(W_{i0}) = \frac{1}{n(n-1)} \sum\sum_{i \neq i'} I(P_{i'} \supset P_i \cup j) K_h(X_{i',P_i} \beta_{P_i,0} - W_{i0}) \{\tilde{\epsilon}_{ij0} X_{i'j} + \tilde{\epsilon}_{i'0} X_{ij}\}.$$

Using Lemma 5.2.1.A of Serfling (1980, page 183), we have

$$\frac{1}{n}\sum_{i=1}^{n}\tilde{\epsilon}_{ij0}\widehat{r}_{j,P_i}(W_{i0})-\frac{1}{n}\sum_{i=1}^{n}\tilde{\epsilon}_{ij0}E\{I(P_{i'}\supset P_i\cup j)K_h(X_{i',P_i}\beta_{P_i,0}-W_{i0})X_{i'j}|W_{i0},P_i\}=O_p\{1/(nh^{1/2})\}, \tag{3.10}$$

because the left hand side is a degenerated U-statistic. Using the standard method to calculate the bias in nonparametric regression under Conditions (C1)–(C3), we have

$$\sup_{W_{i0}}|E\{I(P_{i'}\supset P_i\cup j)K_h(X_{i',P_i}\beta_{P_i,0}-W_{i0})X_{i'j}\mid W_{i0},P_i\}-r_{j,P_i}(W_{i0})|=O_p(h^2),$$

which implies

$$\frac{1}{n}\sum_{i=1}^{n}\tilde{\epsilon}_{ij0}[E\{I(P_{i'}\supset P_i\cup j)K_h(X_{i',P_i}\beta_{P_i,0}-W_{i0})X_{i'j}\mid W_{i0},P_i\}-r_{j,P_i}(W_{i0})]=O_p(n^{-1/2}h^2). \tag{3.11}$$

Combining (3.10), (3.11) and Condition (C5), we have

$$\frac{1}{n}\sum_{i=1}^{n}\epsilon_{i0}\left\{\frac{\widehat{r}_{j,P_i}(W_{i0})-r_{j,P_i}(W_{i0})}{f_{j,P_i}(W_{i0})}\right\}=O_p\{n^{-1}h^{-1/2}+h^2n^{-1/2}\}=o_p(n^{-1/2}). \tag{3.12}$$

Similarly, we have

$$\frac{1}{n}\sum_{i=1}^{n}\epsilon_{i0}\left\{\frac{r_{j,P_i}(W_{i0})(\widehat{f}_{j,P_i}(W_{i0})-f_{P_i}(W_{i0}))}{f^2_{j,P_i}(W_{i0})}\right\}=o_p(n^{-1/2}). \tag{3.13}$$

Submitting (3.12) and (3.13) into (3.9) and using Condition (C5), we complete the proof of (3.3).

3.2 Proof

We prove the theorem in four steps.

In step one, we prove

$$U(\widehat{\boldsymbol{\beta}})-U(\boldsymbol{\beta}_0)=-\frac{1}{n}\sum_{i=1}^{n}\tilde{\mathbf{Z}}_i(\boldsymbol{\beta}_0)\left\{\mathbf{Z}_i(\boldsymbol{\beta}_0)+\tilde{\mathbf{D}}_i^{\mathrm{T}}\boldsymbol{\beta}_0\right\}^{\mathrm{T}}\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\right)+o_p(\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\|), \tag{3.14}$$

where $\tilde{\mathbf{D}}_i$ is defined in Section 3.

We write

$$
\begin{aligned}
U(\widehat{\boldsymbol{\beta}})-U(\boldsymbol{\beta}_0) &= -\frac{1}{n}\sum_{i=1}^{n}\tilde{\mathbf{Z}}_i(\boldsymbol{\beta}_0)\mathbf{Z}_i(\boldsymbol{\beta}_0)^{\mathrm{T}}\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\right)-\frac{1}{n}\sum_{i=1}^{n}\tilde{\mathbf{Z}}_i(\boldsymbol{\beta}_0)\left\{\mathbf{Z}_i(\widehat{\boldsymbol{\beta}})-\mathbf{Z}_i(\boldsymbol{\beta}_0)\right\}^{\mathrm{T}}\boldsymbol{\beta}_0 \\
&\quad +o_p(\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\|). \qquad (3.15)
\end{aligned}
$$

Denote by $\beta_{P,0}$ and β_{k0} the true values of β_P and β_k, respectively. Recall $\tilde{\mathbf{X}}_{i,P}=(X_{ij}I(j\in P),1\leq j\leq p)^{\mathrm{T}}$. Using some algebra, we have

$$
\begin{aligned}
&\widehat{E}_{j,P_i}(\tilde{\mathbf{X}}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}};\widehat{\boldsymbol{\beta}})-\widehat{E}_{j,P_i}(\tilde{\mathbf{X}}_i^{\mathrm{T}}\boldsymbol{\beta}_0;\boldsymbol{\beta}_0) \\
= \; & \frac{n^{-1}\sum_{i'=1}^{n}I(P_{i'}\supset P_i\cup j)X_{i'j}\dot{K}_h\left(\mathbf{X}_{i',P_i}^{\mathrm{T}}\beta_{P_i,0}-\mathbf{X}_{i,P_i}^{\mathrm{T}}\beta_{P_i,0}\right)\tilde{\mathbf{X}}_{i',P_i}^{\mathrm{T}}\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\right)}{n^{-1}\sum_{i'=1}^{n}I(P_{i'}\supset P_i\cup j)K_h\left(\mathbf{X}_{i',P_i}^{\mathrm{T}}\widehat{\beta}_{P_i}-\mathbf{X}_{i,P_i}^{\mathrm{T}}\widehat{\beta}_{P_i}\right)} \\
& -\frac{n^{-1}\sum_{i'=1}^{n}I(P_{i'}\supset P_i\cup j)X_{i'j}\dot{K}_h\left(\mathbf{X}_{i',P_i}^{\mathrm{T}}\beta_{P_i,0}-\mathbf{X}_{i,P_i}^{\mathrm{T}}\beta_{P_i,0}\right)\tilde{\mathbf{X}}_{i}^{\mathrm{T}}\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\right)}{n^{-1}\sum_{i'=1}^{n}I(P_{i'}\supset P_i\cup j)K_h\left(\mathbf{X}_{i',P_i}^{\mathrm{T}}\widehat{\beta}_{P_i}-\mathbf{X}_{i,P_i}^{\mathrm{T}}\widehat{\beta}_{P_i}\right)} \\
& -\frac{1}{n}\sum_{i'=1}^{n}I(P_{i'}\supset P_i\cup j)X_{i'j}K_h\left(\mathbf{X}_{i',P_i}^{\mathrm{T}}\beta_{P_i,0}-\mathbf{X}_{i,P_i}^{\mathrm{T}}\beta_{P_i,0}\right) \\
& \quad\times\left\{\frac{n^{-1}\sum_{i'=1}^{n}I(P_{i'}\supset P_i\cup j)\dot{K}_h\left(\mathbf{X}_{i',P_i}^{\mathrm{T}}\beta_{P_i,0}-\mathbf{X}_{i,P_i}^{\mathrm{T}}\beta_{P_i,0}\right)\tilde{\mathbf{X}}_{i',P_i}^{\mathrm{T}}\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\right)}{\left[n^{-1}\sum_{i'=1}^{n}I(P_{i'}\supset P_i\cup j)K_h\left(\mathbf{X}_{i',P_i}^{\mathrm{T}}\beta_{P_i,0}-\mathbf{X}_{i,P_i}^{\mathrm{T}}\beta_{P_i,0}\right)\right]^2}\right\} \\
& +\frac{1}{n}\sum_{i'=1}^{n}I(P_{i'}\supset P_i\cup j)X_{i'j}K_h\left(\mathbf{X}_{i',P_i}^{\mathrm{T}}\beta_{P_i,0}-\mathbf{X}_{i,P_i}^{\mathrm{T}}\beta_{P_i,0}\right) \\
& \quad\times\left\{\frac{n^{-1}\sum_{i'=1}^{n}I(P_{i'}\supset P_i\cup j)\dot{K}_h\left(\mathbf{X}_{i',P_i}^{\mathrm{T}}\beta_{P_i,0}-\mathbf{X}_{i,P_i}^{\mathrm{T}}\beta_{P_i,0}\right)\tilde{\mathbf{X}}_{i}^{\mathrm{T}}\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\right)}{\left[n^{-1}\sum_{i'=1}^{n}I(P_{i'}\supset P_i\cup j)K_h\left(\mathbf{X}_{i',P_i}^{\mathrm{T}}\beta_{P_i,0}-\mathbf{X}_{i,P_i}^{\mathrm{T}}\beta_{P_i,0}\right)\right]^2}\right\}+o_p(\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\|)
\end{aligned}
$$

Using the kernel theory (Horowitz, 1996), we can show that

$$
\widehat{E}_{j,P_i}(\tilde{\mathbf{X}}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}};\widehat{\boldsymbol{\beta}})-\widehat{E}_{j,P_i}(\tilde{\mathbf{X}}_i^{\mathrm{T}}\boldsymbol{\beta}_0;\boldsymbol{\beta}_0)=\mathbf{D}_{ij}^{\mathrm{T}}\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\right)+o_p(\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\|). \qquad (3.16)
$$

Substituting (3.16) into (3.15), we prove (3.14).

In the second step, we derive the asymptotic expression of $n^{1/2}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)$.

From (3.14), we have

$$\begin{aligned} U(\widehat{\boldsymbol\beta}) - U(\boldsymbol\beta_0) &= -\frac{1}{n}\sum_{i=1}^{n}\tilde{\mathbf z}_i(\boldsymbol\beta_0)\left\{\mathbf z_i(\boldsymbol\beta_0)+\tilde{\mathbf D}_i^{\mathrm T}\boldsymbol\beta_0\right\}^{\mathrm T}\left(\widehat{\boldsymbol\beta}-\boldsymbol\beta_0\right) \\ &- \frac{1}{n}\sum_{i=1}^{n}\{\tilde{\mathbf Z}_i(\boldsymbol\beta_0)-\tilde{\mathbf z}_i(\boldsymbol\beta_0)\}\left\{\mathbf Z_i(\boldsymbol\beta_0)+\tilde{\mathbf D}_i^{\mathrm T}\boldsymbol\beta_0\right\}^{\mathrm T}\left(\widehat{\boldsymbol\beta}-\boldsymbol\beta_0\right) \\ &- \frac{1}{n}\sum_{i=1}^{n}\tilde{\mathbf z}_i(\boldsymbol\beta_0)\{\mathbf Z_i(\boldsymbol\beta_0)-\mathbf z_i(\boldsymbol\beta_0)\}'\left(\widehat{\boldsymbol\beta}-\boldsymbol\beta_0\right)+o_p(\|\widehat{\boldsymbol\beta}-\boldsymbol\beta_0\|) \\ &= -\frac{1}{n}\sum_{i=1}^{n}\tilde{\mathbf z}_i(\boldsymbol\beta_0)\left\{\mathbf z_i(\boldsymbol\beta_0)+\tilde{\mathbf D}_i^{\mathrm T}\boldsymbol\beta_0\right\}^{\mathrm T}\left(\widehat{\boldsymbol\beta}-\boldsymbol\beta_0\right)+o_p(\|\widehat{\boldsymbol\beta}-\boldsymbol\beta_0\|) \end{aligned} \tag{3.17}$$

where the last equality follows from (3.1) and (3.2) in Lemma 1. Because $U(\widehat{\boldsymbol\beta})=0$, we have

$$n^{1/2}(\widehat{\boldsymbol\beta}-\boldsymbol\beta_0)=\left(\frac{1}{n}\sum_{i=1}^{n}\tilde{\mathbf z}_i(\boldsymbol\beta_0)\left\{\mathbf z_i(\boldsymbol\beta_0)+\tilde{\mathbf D}_i^{\mathrm T}\boldsymbol\beta_0\right\}^{\mathrm T}\right)^{-1}\sqrt{n}U(\boldsymbol\beta_0)+o_p(1). \tag{3.18}$$

In the third step, we derive the asymptotic expression of $U(\boldsymbol\beta_0)$. Write

$$U(\boldsymbol\beta_0)=\frac{1}{n}\sum_{i=1}^{n}\tilde{\mathbf Z}_i(\boldsymbol\beta_0)\left\{\epsilon_{i0}+\left(\tilde{\mathbf X}_i^c-\tilde{\mathbf e}_i^c\right)^{\mathrm T}\boldsymbol\beta_0\right\}-\frac{1}{n}\sum_{i=1}^{n}\tilde{\mathbf Z}_i(\boldsymbol\beta_0)\left(\tilde{\mathbf E}_i^c-\tilde{\mathbf e}_i^c\right)^{\mathrm T}\boldsymbol\beta_0, \tag{3.19}$$

recalling the definitions in (3.7) and (3.8).

Using (2.1) and Lemma 4 in Chen et al. (2010), we have

$$\begin{aligned} \widehat{E}_{j,P_i}(W_{i0})-e_{j,P_i}(W_{i0}) &= -\frac{s_{1,j}(W_{i0};P_i)}{s_{0,j}(W_{i0};P_i)^2}\{S_{n0,j}(W_{i0};P_i)-s_{0,j}(W_{i0};P_i)\} \\ &\quad+\frac{1}{s_{0,j}(W_{i0};P_i)}S_{n1,j}(W_{i0};P_i)+O_p\left(\frac{1}{\sqrt{nh}}+h^2\right)^2, \end{aligned} \tag{3.20}$$

where $S_{nr,j}(W_{i0};P_i)=n^{-1}\sum_{i'=1}^{n}I(P_{i'}\supset P_i\cup j)\left\{X_{i'j}-e_{i',P_i}(W_{i0})\right\}^{\otimes r}K_h\left(\mathbf X_{i',P_i}^{\mathrm T}\beta_{P_i,0}-W_{i0}\right)$ for $r=0,1$. $s_{0,j}(W_{i0};P_i)=C(P_i\cup j;W_{i0})f_{P_i}(W_{i0})$, and $f_{P_i}(w)=f_{P_i}(w;\boldsymbol\beta_0)$. Using Lemmas 4 and 5 in Chen et al. (2010), we obtain

$$\frac{1}{n}\sum_{i=1}^{n}\tilde{\mathbf z}_i(\boldsymbol\beta_0)\left(\tilde{\mathbf E}_i^c-\tilde{\mathbf e}_i^c\right)^{\mathrm T}\boldsymbol\beta_0=O_p(h^2)+o_p(n^{-1/2}). \tag{3.21}$$

Substituting (3.3), (3.4), (3.5) and (3.21) into (3.19), we obtain

$$U(\boldsymbol{\beta}_0) = \frac{1}{n}\sum_{i=1}^{n} \tilde{\mathbf{z}}_i(\boldsymbol{\beta}_0)\left\{\epsilon_{i0} + \left(\tilde{\mathbf{X}}_i^c - \tilde{\mathbf{e}}_i^c\right)^{\mathrm{T}} \boldsymbol{\beta}_0\right\} + o_p(n^{-1/2}). \tag{3.22}$$

In the fourth step, based on (3.18) and (3.22), we can use the central limit theorem and Slutsky's theorem to obtain $n^{1/2}\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right) \to N(0, \boldsymbol{\Sigma}^{-1}\mathbf{A}(\boldsymbol{\Sigma}^{-1})^{\mathrm{T}})$, where

$$\boldsymbol{\Sigma} = E\left[\tilde{\mathbf{z}}_i(\boldsymbol{\beta}_0)\left\{\mathbf{z}_i(\boldsymbol{\beta}_0) + \tilde{\mathbf{D}}_i^{\mathrm{T}}\boldsymbol{\beta}_0\right\}^{\mathrm{T}}\right], \quad \mathbf{A} = \mathrm{var}\left[\tilde{\mathbf{z}}_i(\boldsymbol{\beta}_0)\left\{\epsilon_{i0} + \left(\tilde{\mathbf{X}}_i^c - \tilde{\mathbf{e}}_i^c\right)^{\mathrm{T}}\boldsymbol{\beta}_0\right\}\right].$$

4 Proof of Theorem 3

We use the proving strategy of Jin, Ying, and Wei (2001). Let $V_i \sim \mathrm{Binomial}(n, 1/n)$ with mean 1 and variance $(n-1)/n$ in our bootstrap setting. Let

$$W_n(\boldsymbol{\beta}) = \sqrt{n}U(\boldsymbol{\beta}) \quad \text{and} \quad W_n^*(\boldsymbol{\beta}) = n^{-1/2}\sum_{i=1}^{n} V_i\tilde{\mathbf{Z}}_i(\boldsymbol{\beta})\{Y_i - \mathbf{Z}_i(\boldsymbol{\beta})^{\mathrm{T}}\boldsymbol{\beta}\}.$$

Therefore, $\widehat{\boldsymbol{\beta}}^*$ is the solution of $W_n^*(\boldsymbol{\beta}) = 0$. Similar to the proof of Theorem 2, we can show that

$$W_n^*(\beta_1) - W_n^*(\beta_2) = n^{1/2}\boldsymbol{\Sigma}(\beta_1 - \beta_2) + o_p(\sqrt{n}\|\beta_1 - \beta_2\| + 1),$$

uniformly in $\|\beta_1 - \boldsymbol{\beta}_0\| \le d_n$ and $\|\beta_2 - \boldsymbol{\beta}_0\| \le d_n$, where $\{d_n\}$ is any sequence of positive random variables, converging to 0 almost surely, and $\boldsymbol{\Sigma}$ is defined in Theorem 2. Then, in the probability space of $\{V, X, Y\}$,

$$n^{1/2}(\widehat{\boldsymbol{\beta}}^* - \widehat{\boldsymbol{\beta}}) = -\boldsymbol{\Sigma}^{-1}W_n^*(\widehat{\boldsymbol{\beta}}) + o_p(1 + n^{1/2}\|\widehat{\boldsymbol{\beta}}^* - \widehat{\boldsymbol{\beta}}\|), \tag{4.1}$$

almost surely. It follows from Theorem 2 that the asymptotic distribution of $n^{1/2}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ is the same as that of $\boldsymbol{\Sigma}^{-1}W_n(\boldsymbol{\beta}_0)$. Thus, in view of (4.1), to show that for every realisation that of $n^{1/2}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$, it suffices to show that for every realisation of $\{X, Y\}$ the conditional distribution of $W_n^*(\widehat{\boldsymbol{\beta}})$ converges to normal with mean 0 and covariance matrix $\mathbf{A}$, which is the limiting distribution of $W_n(\boldsymbol{\beta}_0)$.

Since $\widehat{\boldsymbol{\beta}}$ is a solution of $W_n(\boldsymbol{\beta}) = 0$, it implies $W_n(\widehat{\boldsymbol{\beta}}) = 0$. Thus, up to an almost surely negligible term,

$$\begin{aligned} W_n^*(\widehat{\boldsymbol{\beta}}) &= n^{-1/2} \sum_{i=1}^{n} (V_i - 1) \tilde{\mathbf{Z}}_i(\widehat{\boldsymbol{\beta}}) \{Y_i - \mathbf{Z}_i(\widehat{\boldsymbol{\beta}})^{\mathrm{T}} \widehat{\boldsymbol{\beta}}\} \\ &= n^{-1/2} \sum_{i=1}^{n} (V_i - 1) \tilde{\mathbf{z}}_i(\boldsymbol{\beta}_0) \left\{ \epsilon_{i0} + \left(\tilde{\mathbf{X}}_i^c - \tilde{\mathbf{e}}_i^c \right)^{\mathrm{T}} \boldsymbol{\beta}_0 \right\} + o_p(1). \end{aligned} \tag{4.2}$$

According to the strong law of large number, the covariance matrix of (4.2) converges almost surely to $\mathbf{A}$. Moreover, by the usual multivariate central limit theorem (Serfling, 1980, page 30), for every realization of $\{X, Y\}$, the conditional distribution of (4.2) converges to normal with mean 0 and covariance matrix $\mathbf{A}$. Hence, $n^{1/2}(\widehat{\boldsymbol{\beta}}^* - \widehat{\boldsymbol{\beta}})$ has the same asymptotic distribution as that of $n^{1/2}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$.

5 Additional simulation results

We simulate data with $n = 500$, normal error, and missing proportion 0.5, and apply our method varying the bandwidth $h = cn^{-1/3}$ with $c \in \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16\}$. Figure 1 shows the average root mean squared error of the regression coefficients as a function of the bandwidth h, which confirms that our method is not sensitive to the choice of h when h is taken in a reasonable interval.

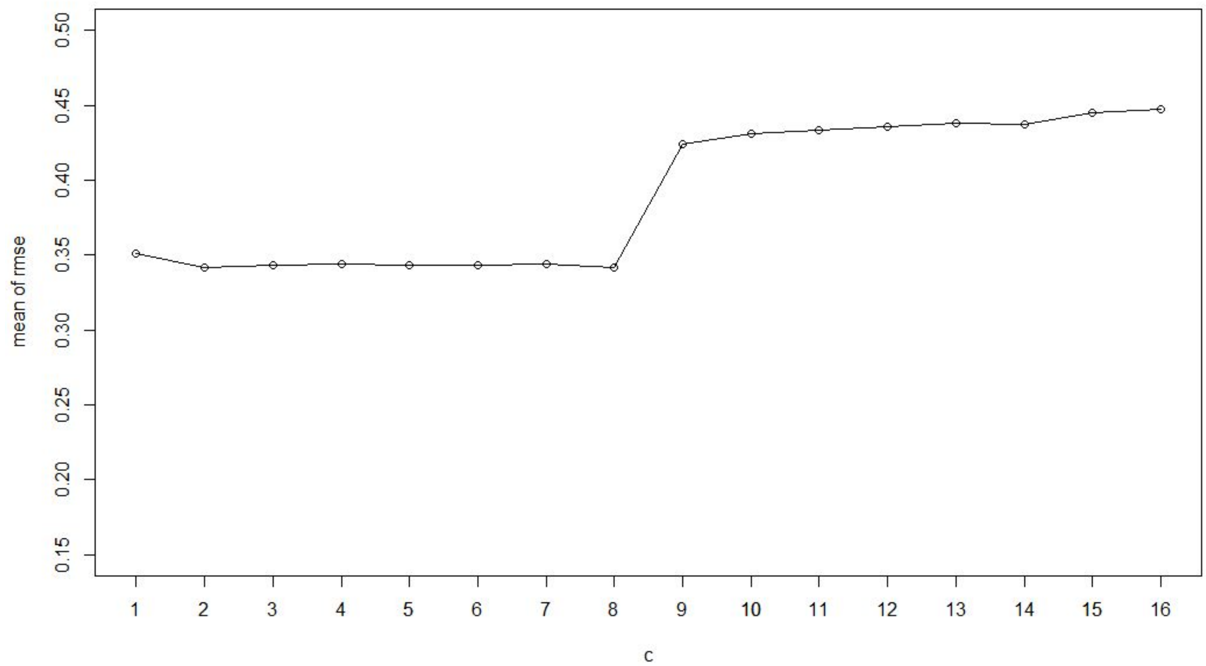

Figure 1: Sensitivity analysis of the bandwidth based on the simulated data with $n = 500$, normal error, and missing proportion 0.5

References

Chen, K., Guo, S., Sun, L. & Wang, J. (2010). Global Partial Likelihood for Nonparametric Proportional Hazards Models. *Journal of the American Statistical Association*, **105**, 750–760.

Horowitz, J. L. (1996). Semiparametric estimation of a regression model with an unknown transformation of the dependent variable. *Econometrica*, **64**, 103-137.

Jin, Z., Ying, Z., & Wei, L. J. (2001). A simple resampling method by perturbing the minimand. *Biometrika*, **88(2)**, 381-390.

Serfling, R. (1980). *Approximation theorems of mathematical statistics.* Wiley, New York.

van der Vaart, A. (1998). *Asymptotic Statistics.* Cambridge University Press.