Partially linear single-index model with missing responses at random

Peng Lai\textsuperscript{a,b}, Qihua Wang\textsuperscript{a,c,*}

\textsuperscript{a} Academy of Mathematics and Systems Science, Chinese Academy of Science, Beijing 100190, China
\textsuperscript{b} College of Math & Physics, Nanjing University of Information Science & Technology, Nanjing 210044, China
\textsuperscript{c} Department of Mathematics and Statistics, Yunnan University, Kunming, 650091, China

\begin{abstract}
This paper considers semiparametric partially linear single-index model with missing responses at random. Imputation approach is developed to estimate the regression coefficients, single-index coefficients and the nonparametric function, respectively. The imputation estimators for the regression coefficients and single-index coefficients are obtained by a stepwise approach. These estimators are shown to be asymptotically normal, and the estimator for the nonparametric function is proved to be asymptotically normal at any fixed point. The bandwidth problem is also considered in this paper, a delete-one cross validation method is used to select the optimal bandwidth. A simulation study is conducted to evaluate the proposed methods.
\end{abstract}

\section{Introduction}

Regression analysis is one of useful techniques in statistics. It is used in many different areas. However, in regression analysis, dimensionality and collinearity are two big difficulties which have received much attention. In order to deal with the high dimensional problems and avoid the so-called “curse of dimensionality”, many dimension reduction methods have been considered and many parametric or semiparametric models have been proposed for different practical situations. Examples include additive model (Breiman and Friedman, 1985; Hastie and Tibshirani, 1990), partially linear model (Green and Silverman, 1994; Liang, 2006; You et al., 2007), varying-coefficient model (Hastie and Tibshirani, 1993), generalized partially linear single-index model (Carroll et al., 1997), extended partially linear single-index model (Xia et al., 1999), semiparametric varying-coefficient partially linear model (Fan and Huang, 2005), varying-coefficient single-index model (Wong et al., 2008), and so on. Partially linear single-index model is an important alternative to the classical linear model by putting a nonparametric component into the model. That is

\begin{equation}
Y = \theta^T Z + g(X^T \beta) + e, \tag{1.1}
\end{equation}

where $Y$ is a scalar response variate, $(X, Z) \in \mathbb{R}^p \times \mathbb{R}^q$, $g(\cdot)$ is an unknown univariate link function, $e$ is the random error with $E(e|X, Z) = 0$ and $\text{var}(e|X, Z) = \sigma^2$. $(\beta, \theta)$ is an unknown vector in $\mathbb{R}^p \times \mathbb{R}^q$ with $\|\beta\| = 1$ (where $\| \cdot \|$ denotes the Euclidean metric). Assume the true parameter vector is $(\beta_0, \theta_0)$.
This model has gained much attention in recent years. Carroll et al. (1997) estimate the nonparametric function and all parameters by minimizing two different loss functions with under-smoothing. Xia and Härdle (2006) propose a minimum average variance estimation method and an iterative algorithm to get the consistent estimators of the parametric vector and nonparametric function without under-smoothing. They apply this model to credit scoring, circulatory and respiratory problems in Hong Kong and the research of complicated effect of weather conditions on health problems. Zhu and Xue (2006) propose an empirical likelihood method to construct confidence regions for the parameters in the partially linear single-index model. They propose a bias correction method to avoid under-smoothing, and prove that the adjusted empirical log-likelihood is asymptotically standard $\chi^2$ distributed. However, the existing work is based on complete data set.

In practice, response variable may be missing, by design or by happenstance. For example, maybe it is too expensive to measure the response $Y$’s and only part of $Y$’s are available. Missing response data are commonly found in opinion polls, market research surveys, social investigations, medical studies and other disciplines. In the past decade years, missing data analysis problems have been investigated extensively.

In this paper, we propose an imputation approach to estimate $\beta, \theta$ and $g(\cdot)$ with response missing. Suppose that we obtain a random sample of incomplete data

$$(Y_i, \delta_i, X_i, Z_i), \quad i = 1, 2, \ldots, n,$$

from model (1.1), where $\delta_i = 0$ if $Y_i$ is missing, $\delta_i = 1$ otherwise. Throughout this paper we assume that $Y$ is missing at random (MAR). That is, $p(\delta = 1|Y, X, Z) = p(\delta = 1|X, Z)$. MAR is a common assumption for statistical analysis with missing data and it is reasonable in many practical situations; see Little and Rubin (1987). In order to define the imputation estimators of $\beta, \theta$ and $g(\cdot)$, the initial estimators are defined firstly. We compare the initial estimators and imputation estimators by simulation in terms of bias and mean square errors of the estimators. Asymptotic results for the imputation estimators are derived, respectively. Details can be found in the Appendix.

The rest of this paper is organized as follows. In Section 2, we use the imputation method to estimate the relative parameters and unknown function. In Section 3, some asymptotic results are presented. In Section 4, the bandwidth selection problem is considered. In Section 5, a simulation study is conducted to compare the biases and mean square errors of our proposed estimators with complete case estimators.

2. Methodology

In this section, we define initial estimators and imputation estimators. With the incomplete observations \{$(Y_i, \delta_i, X_i, Z_i); 1 \leq i \leq n$\}, similar to Wang et al. (2004) multiplying the observation indicator on both side of (1.1) and taking conditional expectations given $X^T \beta$, we have

$$E[\delta_i Y_i | X_i^T \beta] = E[\delta_i Z_i | X_i^T \beta]^T \theta + E[\delta_i | X_i^T \beta]g(X_i^T \beta).$$

(2.1)

Similarly, we can obtain

$$E[\delta_i | X_i^T \beta]Y_i = E[\delta_i | X_i^T \beta]Z_i - E[\delta_i | X_i^T \beta]g(X_i^T \beta) + E[\delta_i | X_i^T \beta] \epsilon_i.$$  

(2.2)

Therefore, we have

$$\delta_i p_1(X_i^T \beta) Y_i - \delta_i m_1(X_i^T \beta) = \delta_i [p_1(X_i^T \beta) Z_i - m_2(X_i^T \beta)]^T \theta + \delta_i \gamma_i,$$

(2.3)

where $p_1(t) = E[\delta | X^T \beta = t], m_1(t) = E[Y | X^T \beta = t], m_2(t) = E[Z | X^T \beta = t], \gamma_i = p_1(X_i^T \beta) \epsilon_i$.

If $\beta$ is known and $p_1(t), m_1(t), m_2(t)$ are known functions, according to (2.3), the standard approach can be used to define the estimator of $\theta$. However, $p_1(t), m_1(t), m_2(t), \beta$ are unknown, we need to replace them with their estimators to obtain the estimator of $\theta$. For a fixed $\beta$, the estimator $\hat{\theta}_n$ of $\theta$ can be defined as

$$\hat{\theta}_n = \left\{ \frac{1}{n} \sum_{i=1}^{n} \delta_i [\hat{p}_1(X_i^T \beta) Z_i - \hat{m}_2(X_i^T \beta)]^T \hat{p}_1(X_i^T \beta) Z_i - \hat{m}_2(X_i^T \beta)] \right\}^{-1} \left\{ \frac{1}{n} \sum_{i=1}^{n} (\hat{p}_1(X_i^T \beta) Y_i - \hat{m}_1(X_i^T \beta) \hat{p}_1(X_i^T \beta) Z_i - \hat{m}_2(X_i^T \beta)) \delta_i \right\},$$

(2.4)

where

$$\hat{m}_1(t) = \frac{\sum_{i=1}^{n} \delta_i Y_i K_h \left( \frac{X_i^T \beta - t}{h_1} \right)}{\sum_{i=1}^{n} K_h \left( \frac{X_i^T \beta - t}{h_1} \right)}, \quad \hat{m}_2(t) = \frac{\sum_{i=1}^{n} \delta_i Z_i K_h \left( \frac{X_i^T \beta - t}{h_1} \right)}{\sum_{i=1}^{n} K_h \left( \frac{X_i^T \beta - t}{h_1} \right)}, \quad \hat{p}_1(t) = \frac{\sum_{i=1}^{n} \delta_i K_h \left( \frac{X_i^T \beta - t}{h_1} \right)}{\sum_{i=1}^{n} K_h \left( \frac{X_i^T \beta - t}{h_1} \right)},$$

where $K_h(\cdot)$ is a kernel function and $h_1 = h_1(n)$ is a bandwidth with $0 < h_1$ and $h_1 \to 0$.

For any fixed $\beta$, based on $\hat{\theta}_n$, we can define the initial estimators of $g(t)$ and $g'(t)$, say $\hat{g}_n(t)$ and $\hat{g'}_n(t)$, by minimizing

$$\min_{a(t), b(t)} \sum_{i=1}^{n} (Y_i - Z_i \hat{\theta}_n - a(t) - b(t)(X_i^T \beta - t))^2 K_{h_1}(X_i^T \beta - t) \delta_i.$$
Let $\tilde{\alpha}(t), \tilde{b}(t)$ be the solutions of the above optimal equation. Then, let $\tilde{g}_n(t) = \tilde{\alpha}(t)\tilde{g}_n(t) - \tilde{b}(t)$, through direct calculation, we have

$$
\begin{pmatrix}
\tilde{g}_n(t) \\
h_2\tilde{g}_n'(t)
\end{pmatrix} = \frac{1}{n} \sum_{i=1}^{n} \left( \frac{1}{n} B_0^T W_i B_0 \right)^{-1} B_0 \delta_i K_{h_2}(X_i^T \beta - t)(Y_i - Z_i^T \tilde{b}_n),
$$

(2.5)

where

$$
B_0 = \begin{pmatrix}
B_{10}^T \\
\vdots \\
B_{n0}^T
\end{pmatrix}, \quad W_1 = \begin{pmatrix}
\delta_1 K_{h_2}(X_1^T \beta - t) \\
\vdots \\
\delta_n K_{h_2}(X_n^T \beta - t)
\end{pmatrix},
$$

with $K_{h_2}(\cdot)$ being a kernel function.

The estimators $\tilde{\alpha}_n, \tilde{g}_n(\cdot)$ and $\tilde{g}_n'(\cdot)$ are derived in the case where $\beta$ is fixed. However, in practice, we do not know the true value of $\beta$. We need to estimate it by minimizing the sum of square errors

$$
L_{\beta}(\beta) = \sum_{i=1}^{n} \delta_i (Y_i - Z_i^T \tilde{b}_n - \tilde{g}_n(X_i^T \beta))^2,
$$

(2.6)

under the condition that $\|\beta\| = 1$, say $\tilde{\beta}_{af}$. Replacing $\beta$ with $\tilde{\beta}_{af}$ in $\tilde{\alpha}_n, \tilde{g}_n(\cdot)$ and $\tilde{g}_n'(\cdot)$, we obtain $\tilde{\alpha}_{af}, \tilde{g}_{af}(\cdot)$ and $\tilde{g}_{af}'(\cdot)$. The estimation procedure for obtaining $\tilde{\beta}_{af}, \tilde{\alpha}_{af}$ and $\tilde{g}_{af}(\cdot)$ can be decomposed as an iterative process.

Step 1: Select an initial value $\tilde{\beta}^{(0)}$, for example, the complete case estimator by the method of Xia and Härdle (2006).

Let $\tilde{\beta}_{af}^{(0)} = \tilde{\beta}^{(0)}$.

Step 2: Calculate $\tilde{\alpha}_n^{(k)}, \tilde{g}_n^{(k)}(\cdot)$ and $\tilde{g}_n^{(k)}(\cdot)$ based on (2.4) and (2.5) with $\beta = \tilde{\beta}_{af}^{(k)}$.

Step 3: Newton Raphson algorithm is used to solve the equation $\sum_{i=1}^{n} \delta_i (Y_i - Z_i^T \tilde{b}_n^{(k)} - \tilde{g}_n^{(k)}(X_i^T \beta))\tilde{g}_n'(X_i^T \beta)X_i = 0$ on $\beta$.

Let $\tilde{\beta}_{af}^{(k+1)}$ be the solution, which minimizes (2.6). Let $\tilde{\beta}_{af}^{(k+1)} = \tilde{\beta}_{af}^{(k+1)} / \|\tilde{\beta}_{af}^{(k+1)}\|$.

Step 4: Continue Steps 2 and 3 until convergence, and obtain the iterative estimate $\tilde{\beta}_{af}$.

From (2.5) with $\beta = \tilde{\beta}_{af}$, we get

$$
\tilde{\alpha}_{af}(\cdot) = \tilde{g}_{af}(\cdot),
$$

(2.7)

with $E[\tilde{g}_n(Z_t)] = 0$.

By (2.7), it follows that

$$
E[Y_i|Z_t^T \beta] = E[Z_t^T \beta] \theta + g(X_i^T \beta).
$$

Therefore, we have

$$
\tilde{Y}_i = m_{10}(X_i^T \beta) = [Z_t - m_{20}(X_i^T \beta)]^T \theta + e_i
$$

where $m_{10}(t) = E[\tilde{Y}^T \beta = t]$ and $m_{20}(t) = E[Z^T \beta = t]$. We can obtain the least square estimator of $\theta$:

$$
\hat{\theta}_{af} = \left[ \frac{1}{n} \sum_{i=1}^{n} (Z_t - m_{20}(X_i^T \beta))(Z_t - m_{20}(X_i^T \beta))^T \right]^{-1} \left[ \frac{1}{n} \sum_{i=1}^{n} (\tilde{Y}_i - m_{10}(X_i^T \beta))(Z_t - m_{20}(X_i^T \beta))^T \right].
$$

(2.9)

Replacing $\tilde{Y}_i$ with

$$
\tilde{Y}_i = \delta_i Y_i + (1 - \delta_i)(\tilde{\alpha}_{af}(Z_t) + \tilde{g}_{af}(X_i^T \tilde{\beta}_{af})),
$$

(2.10)

and replacing $m_{10}(t), m_{20}(t)$ with

$$
\hat{m}_{10}(t) = \frac{\sum_{i=1}^{n} \tilde{Y}_i K_{3}(X_i^T \beta - t)}{\sum_{i=1}^{n} K_{3}(X_i^T \beta - t)}, \quad \hat{m}_{20}(t) = \frac{\sum_{i=1}^{n} Z_t K_{3}(X_i^T \beta - t)}{\sum_{i=1}^{n} K_{3}(X_i^T \beta - t)}.
$$
we can define an imputation estimator of $\theta$ by

$$
\hat{\theta}_{nl} = \left( \frac{1}{n} \sum_{i=1}^{n} (Z_i - \hat{m}_{20}(X_i^T \beta))(Z_i - \hat{m}_{20}(X_i^T \beta))^T \right)^{-1} \left( \frac{1}{n} \sum_{i=1}^{n} (\hat{Y}_i - \hat{m}_{10}(X_i^T \beta))(\hat{Y}_i - \hat{m}_{10}(X_i^T \beta))^T \right),
$$

(2.11)

where $K(\cdot)$ is a kernel function and $h_3 = h_3(n)$ is a bandwidth with $0 < h_3$ and $h_3 \to 0$.

Similarly, the imputation estimators of $g(t)$ and $g(t)$ can be obtained by

$$
\min_{a(t),b(t)} \sum_{i=1}^{n} (\hat{Y}_i - Z_i^T \hat{\theta}_{nl} - a_i(t) - b_i(t)(X_i^T \beta - t))^2 K_{h_4}(X_i^T \beta - t).
$$

(2.12)

They are

$$
\begin{pmatrix}
\hat{g}_{nl}(t) \\
\hat{h}_{nl}(t)
\end{pmatrix} = \frac{1}{n} \sum_{i=1}^{n} \left( \frac{1}{B^T W_{4} B} \right)^{-1} B_i K_{h_4}(X_i^T \beta - t)(\hat{Y}_i - Z_i^T \hat{\theta}_{nl}).
$$

(2.13)

where

$$
B = \begin{pmatrix}
B_1^T \\
\vdots \\
B_n^T
\end{pmatrix},
W_4 = \begin{pmatrix}
K_{h_4}(X_1^T \beta - t) & \cdots & K_{h_4}(X_n^T \beta - t)
\end{pmatrix},
B_i = \left( \begin{array}{c}
X_i^T \beta - t \\
\frac{1}{h_4}
\end{array} \right), i = 1,2,\ldots,n
$$

and $K_{h_4}(\cdot) = h_4^{-1} K_{h_4}(\cdot/h_4), 0 < h_4 \to 0$,

with $K_{h_4}(\cdot)$ being a kernel function.

The above estimators of $\theta$ and $g(\cdot)$ depend on the unknown parameter $\beta$. It is nature to replace $\beta$ with its estimator $\hat{\beta}_{nl}$.

It can be obtained by minimizing the sum of square errors

$$
L_2(\hat{\beta}) = \sum_{i=1}^{n} (\hat{Y}_i - Z_i^T \hat{\theta}_{nl} - \hat{g}_{nl}(X_i^T \beta))^2,
$$

(2.14)

under the condition that $\|\beta\| = 1$.

We then define the imputation estimators of $\theta$ and $g(\cdot)$ by replacing $\beta$ with $\hat{\beta}_{nl}$ in $\hat{\theta}_{nl}, \hat{g}_{nl}(\cdot)$, and obtain the final imputation estimators $\hat{\theta}_{nl}$ and $\hat{g}_{nl}(\cdot)$. The algorithm is same as that of calculating the initial estimators.

3. Asymptotic properties

With probability 1, $X$ and $Z$ lie in compact sets $A_X$ and $A_Z$, respectively. Let $B = \{ \beta \in R^p : \|\beta\| = 1 \}$, then $\beta_0$ is an inner point of the compact set $B$. Denote $B_0 = \{ \beta \in B : \|\beta - \beta_0\| \leq Cn^{-1/2} \}$ for any positive constant $C$. To state the asymptotic properties, the following conditions are needed.

1. The density function of $T = X^T \beta$, $f(t)$, is compactly supported, Lipschitz continuous, bounded away from zero on $T$, where $T = \{ t = X^T \beta, x \in A_X, \beta \in B_n \}$. The conditional density functions of $X$ and $Z$ given $T=t$, denoted by $f_{X|T}(x|t)$ and $f_{Z|T}(z|t)$, respectively, have bounded and continuous derivatives up to order 2 on $T$.

2. The functions $g(t)$, $m_1(t)$, $m_2(t)$, $m_{10}(t)$ and $m_{20}(t)$ have bounded and continuous derivatives up to order 2 on $T$, where $m_{2j}(t)$ and $m_{20}(t)$ are the $j$th components of $m_{2j}(t)$ and $m_{2o}(t)$, respectively, $j = 1,\ldots,q$.

3. $E[\hat{\beta}_0 = 1|X^T \beta = t, Z = z] = p(t,z)$ has bounded partial derivatives up to order 2 and is bounded away from zero on $T \times A_Z$.

4. The kernel $K(\cdot), i = 1,2,3,4$, are bounded symmetrical density functions with support $(-1,1)$ and satisfy the Lipschitz condition.

5. $\theta \in \Theta$, where $\Theta$ is a bounded closure set.

6. $0 < h_i \to 0$ and $n^{1-\gamma} h_i \to \infty$ for some $\gamma < 1 - s^{-1}$, $s > 2$.

7. The matrix $S = E[\partial p_1(X^T \beta_0 Z - m_2(X^T \beta_0))]$, $\hat{S} = E[(Z - m_{20}(X^T \beta_0))]$, $\hat{F}_4 = E[(g'(X^T \beta_0))^2 XX^T]$ are positive definite, where $B^{\otimes 2} = BB^T$.

8. The function $E(Z|X^T \beta = t), E(ZZ^T|X^T \beta = t)$ and $E((ZZ^T)^{1/2}|X^T \beta = t)$ are Lipschitz continuous, where $A = B$ is the Hadamard product of matrix of $A$ and $B$.


The proposed imputation estimators in Section 2 have the following properties.

**Theorem 3.1.** Suppose that conditions 1–9 are satisfied. Let $\log n/nh_1 \to 0, \log n/nh_1^2 \to 0, nh_1^3 \to \infty, nh_2^3 \to \infty, nh_3 \to \infty, nh_2 h_3 \to \infty, nh_3 h_4 \to \infty, nh_4 \to 0, nh_4 \to 0, nh_4^2 \to 0$. Then, we have

$$
\sqrt{n}(\hat{\theta}_{nl} - \theta_0) \overset{D}{\to} N(0,\sigma^2 \Sigma),
$$

where $\Sigma = E(D_3(X^T \beta_0 Z)D_4(X^T \beta_0 Z))$ is given in (A.10) of the Appendix.
Theorem 3.2. Suppose the conditions of Theorem 3.1 hold. Then,

\[ \sqrt{n}h_2(g_{nf}(t) - g(t)) \xrightarrow{D} N\left(0, \frac{p_1(t)\gamma_2(K_d)\sigma^2}{f_1(t)} \right), \]

where \( \gamma_2(K_d) = \int K_d^2(u) \, du \).

Theorem 3.3. Suppose the conditions of Theorem 3.2 hold. We have

\[ \sqrt{n}(\hat{\beta}_nl - \beta_0) \xrightarrow{D} N\left(0, \sigma^2\hat{F}_l^{-1}\Sigma_1\hat{F}_l^{-1} \right), \]

where \( \hat{F}_l \) and \( \Sigma_1 \) are given in (A.15).

4. Bandwidth selection

The choices of bandwidths are rather crucial, and an automatic bandwidth choice procedure is of both theoretical and practical interest (see 

Hart and Vieu, 1990). In this paper, the least-squares delete-one cross-validation (CV) method is used to select bandwidths. It is noted that our estimators involve four bandwidths. We state the selection process in the following two steps:

1. Select \( h_1 \) and \( h_2 \) through

\[ (h_1, h_2) = \arg\min_{h_1, h_2} \frac{1}{n} \sum_{i=1}^{n} (Y_i - Z_i^T \hat{\theta}_{nf}^{(-i)} - \hat{g}_{nf}^{(-i)}(X_i \hat{\beta}_{nf}^{(-i)}))^2 \delta_i, \]

where \( \hat{\theta}_{nf}^{(-i)}, \hat{g}_{nf}^{(-i)}(\cdot), \hat{\beta}_{nf}^{(-i)} \) are the “leave-one-out” version of \( \theta_{nf}, g_{nf}(\cdot), \beta_{nf} \), respectively.

2. Select \( h_3 \) and \( h_4 \) through

\[ (h_3, h_4) = \arg\min_{h_3, h_4} \frac{1}{n} \sum_{i=1}^{n} (Y_i - Z_i^T \hat{\theta}_{nl}^{(-i)} - \hat{g}_{nl}^{(-i)}(X_i \hat{\beta}_{nl}^{(-i)}))^2, \]

where \( \hat{\theta}_{nl}^{(-i)}, \hat{g}_{nl}^{(-i)}(\cdot), \hat{\beta}_{nl}^{(-i)} \) are the “leave-one-out” version of \( \theta_{nl}, g_{nl}(\cdot), \beta_{nl} \), respectively.

However, \( h_i, i=1,2,3,4 \), obtained from (4.1) and (4.2) may not satisfy the conditions imposed in the theorems for \( h_i \). But, we could restrict the choice range of \( h_i \) to minimize (4.1) and (4.2). Assuming \( h_i = c_0n_i^{-\alpha}, i=1,2,3,4 \), and using the conditions of theorems, we can obtain the optimal bandwidth according to (4.1) and (4.2) by choosing \( c_i, i=1,2,3,4 \), from \( \mathcal{C} = \{ 0 > c_1 > -\frac{1}{2}, -\frac{4}{3} > c_2 > -\frac{1}{2}, \frac{2}{3} > c_3 > -1-c_2, c_2 > c_4 > -\frac{1}{2} \} \), where \( c_0 \) is a fixed constant. The greedy grid search algorithm can be used to calculate the optimal bandwidths.

5. Numerical examples

To compare the proposed estimators with other competing estimators, we first give some notations. Let \( \hat{\beta}_C, \hat{\theta}_C, \hat{\gamma}_C \) be the CC estimators which are defined by simply ignoring the missing data, and let \( \hat{\beta}_F, \hat{\theta}_F, \hat{\gamma}_F \) be estimators based on full data, which are often used as golden rule.

In our simulation, the kernel functions were taken to be \( K_0(x) = \frac{1}{2}(1-x^2) \) for \( |x| \leq 1 \), 0 otherwise. \( i=1,2,3,4 \). The simulation used the model \( Y = g(X^T \beta) + Z^T \theta + \epsilon \) with \( X \) and \( Z \) simulated from the uniform distribution \( U[0,1] \) and the normal distribution \( N(0,1) \).

Table 1

<table>
<thead>
<tr>
<th>( p(x) )</th>
<th>( n )</th>
<th>( \hat{\beta}_{nf} )</th>
<th>( \hat{\theta}_{nf} )</th>
<th>( \hat{\beta}_{ml} )</th>
<th>( \hat{\theta}_{ml} )</th>
<th>( \hat{\beta}_C )</th>
<th>( \hat{\theta}_C )</th>
<th>( \hat{\gamma}_C )</th>
<th>( \hat{\beta}_F )</th>
<th>( \hat{\theta}_F )</th>
<th>( \hat{\gamma}_F )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( p_1(x) )</td>
<td>40</td>
<td>0.1541</td>
<td>-0.0964</td>
<td>-0.0939</td>
<td>0.1519</td>
<td>-0.0955</td>
<td>-0.0916</td>
<td>0.1527</td>
<td>-0.0949</td>
<td>-0.0934</td>
<td>0.1505</td>
</tr>
<tr>
<td>60</td>
<td>0.1144</td>
<td>-0.0770</td>
<td>-0.0566</td>
<td>0.1128</td>
<td>-0.0756</td>
<td>-0.0558</td>
<td>0.1155</td>
<td>-0.0790</td>
<td>-0.0563</td>
<td>0.1183</td>
<td>-0.0695</td>
</tr>
<tr>
<td>80</td>
<td>0.0906</td>
<td>-0.0460</td>
<td>-0.0563</td>
<td>0.0894</td>
<td>-0.0447</td>
<td>-0.0560</td>
<td>0.0915</td>
<td>-0.0468</td>
<td>-0.0566</td>
<td>0.1013</td>
<td>-0.0551</td>
</tr>
<tr>
<td>100</td>
<td>0.0809</td>
<td>-0.0476</td>
<td>-0.0426</td>
<td>0.0794</td>
<td>-0.0465</td>
<td>-0.0417</td>
<td>0.0816</td>
<td>-0.0479</td>
<td>-0.0430</td>
<td>0.0737</td>
<td>-0.0437</td>
</tr>
<tr>
<td>( p_2(x) )</td>
<td>40</td>
<td>0.2292</td>
<td>-0.1559</td>
<td>-0.1628</td>
<td>0.2276</td>
<td>-0.1549</td>
<td>-0.1608</td>
<td>0.2291</td>
<td>-0.1548</td>
<td>-0.1638</td>
<td>0.1615</td>
</tr>
<tr>
<td>60</td>
<td>0.1233</td>
<td>-0.0711</td>
<td>-0.0745</td>
<td>0.1226</td>
<td>-0.0703</td>
<td>-0.0743</td>
<td>0.1223</td>
<td>-0.0717</td>
<td>-0.0726</td>
<td>0.1196</td>
<td>-0.0781</td>
</tr>
<tr>
<td>80</td>
<td>0.1069</td>
<td>-0.0662</td>
<td>-0.0572</td>
<td>0.1061</td>
<td>-0.0660</td>
<td>-0.0564</td>
<td>0.1044</td>
<td>-0.0683</td>
<td>-0.0519</td>
<td>0.0996</td>
<td>-0.0570</td>
</tr>
<tr>
<td>100</td>
<td>0.1089</td>
<td>-0.0632</td>
<td>-0.0629</td>
<td>0.1085</td>
<td>-0.0629</td>
<td>-0.0627</td>
<td>0.1082</td>
<td>-0.0634</td>
<td>-0.0618</td>
<td>0.0728</td>
<td>-0.0336</td>
</tr>
<tr>
<td>( p_3(x) )</td>
<td>40</td>
<td>0.2065</td>
<td>-0.1368</td>
<td>-0.1398</td>
<td>0.2050</td>
<td>-0.1362</td>
<td>-0.1377</td>
<td>0.2057</td>
<td>-0.1365</td>
<td>-0.1386</td>
<td>0.1765</td>
</tr>
<tr>
<td>60</td>
<td>0.1669</td>
<td>-0.0934</td>
<td>-0.1170</td>
<td>0.1661</td>
<td>-0.0926</td>
<td>-0.1166</td>
<td>0.1620</td>
<td>-0.0908</td>
<td>-0.1120</td>
<td>0.1198</td>
<td>-0.0702</td>
</tr>
<tr>
<td>80</td>
<td>0.1211</td>
<td>-0.0787</td>
<td>-0.0641</td>
<td>0.1208</td>
<td>-0.0795</td>
<td>-0.0663</td>
<td>0.1205</td>
<td>-0.0735</td>
<td>-0.0643</td>
<td>0.0989</td>
<td>-0.0568</td>
</tr>
<tr>
<td>100</td>
<td>0.1233</td>
<td>-0.0725</td>
<td>-0.0731</td>
<td>0.1231</td>
<td>-0.0720</td>
<td>-0.0734</td>
<td>0.1240</td>
<td>-0.0745</td>
<td>-0.0722</td>
<td>0.0742</td>
<td>-0.0405</td>
</tr>
</tbody>
</table>
distribution with mean 1 and variance 1, respectively, and \( \varepsilon \) generated from the normal distribution with mean 0 and variance 0.01, where \( \beta = (1/\sqrt{3})(1,1,1)^T, g(t) = \sin(\pi(t^2 - A)/(B - A)), \theta = 1.5, A = 0.3912, B = 1.3409. 

Based on this model, we considered the following three response probability functions \( p(x,z) = P(\delta = 1 | X = x, Z = z) = \exp(x^T \phi + \varphi z) / (1 + \exp(x^T \phi + \varphi z)), \) under the MAR assumption.

Case 1: \( \phi = (0.7, 0.7, 0.7)^T, \varphi = 0.85; \)
Case 2: \( \phi = (0.1, 0.1, 0.1)^T, \varphi = 0.52; \)
Case 3: \( \phi = (-0.3, -0.3, -0.3)^T, \varphi = 0.5. \)

For the above cases, the mean response rates are \( E[p_1(x,z)] \approx 0.8, E[p_2(x,z)] \approx 0.6 \) and \( E[p_3(x,z)] \approx 0.45, \) where \( p_1(x,z), p_2(x,z) \) and \( p_3(x,z) \) are the response probability functions for Cases 1, 2 and 3, respectively. For our simulation model, in the

### Table 2
MSE of \( \hat{\beta}_{af}, \hat{\beta}_{ad}, \hat{\beta}_{C} \) and \( \hat{\beta}_{f}. \)

<table>
<thead>
<tr>
<th>( p(x,z) )</th>
<th>( n )</th>
<th>( \hat{\beta}_{af} )</th>
<th>( \hat{\beta}_{ad} )</th>
<th>( \hat{\beta}<em>{af}, \hat{\beta}</em>{ad}, \hat{\beta}_{C} )</th>
<th>( \hat{\beta}_{f} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( p_1(x,z) )</td>
<td>40</td>
<td>0.0628</td>
<td>0.1691</td>
<td>0.1647</td>
<td>0.0619</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>0.0388</td>
<td>0.1336</td>
<td>0.1133</td>
<td>0.0365</td>
</tr>
<tr>
<td></td>
<td>80</td>
<td>0.0167</td>
<td>0.0892</td>
<td>0.1007</td>
<td>0.0164</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.0130</td>
<td>0.0867</td>
<td>0.0779</td>
<td>0.0141</td>
</tr>
<tr>
<td>( p_2(x,z) )</td>
<td>40</td>
<td>0.0850</td>
<td>0.1997</td>
<td>0.2062</td>
<td>0.0856</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>0.0363</td>
<td>0.1231</td>
<td>0.1274</td>
<td>0.0361</td>
</tr>
<tr>
<td></td>
<td>80</td>
<td>0.0218</td>
<td>0.1155</td>
<td>0.1029</td>
<td>0.0218</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.0229</td>
<td>0.1035</td>
<td>0.1011</td>
<td>0.0227</td>
</tr>
<tr>
<td>( p_3(x,z) )</td>
<td>40</td>
<td>0.0844</td>
<td>0.1977</td>
<td>0.2066</td>
<td>0.0837</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>0.0512</td>
<td>0.1463</td>
<td>0.1678</td>
<td>0.0507</td>
</tr>
<tr>
<td></td>
<td>80</td>
<td>0.0352</td>
<td>0.1328</td>
<td>0.1150</td>
<td>0.0351</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.0282</td>
<td>0.1092</td>
<td>0.1099</td>
<td>0.0279</td>
</tr>
</tbody>
</table>

### Table 3
Biases of \( \hat{\beta}_{af}, \hat{\beta}_{ad}, \hat{\beta}_{C} \) and \( \hat{\beta}_{f}. \)

<table>
<thead>
<tr>
<th>( p(x,z) )</th>
<th>( n )</th>
<th>( \hat{\beta}_{af} )</th>
<th>( \hat{\beta}_{ad} )</th>
<th>( \hat{\beta}_{C} )</th>
<th>( \hat{\beta}_{f} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( p_1(x,z) )</td>
<td>40</td>
<td>-0.0005</td>
<td>0.0009</td>
<td>-0.0002</td>
<td>-0.0017</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>-0.0039</td>
<td>0.0024</td>
<td>-0.0044</td>
<td>-0.0030</td>
</tr>
<tr>
<td></td>
<td>80</td>
<td>0.0026</td>
<td>0.0083</td>
<td>0.0019</td>
<td>0.0012</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>-0.0007</td>
<td>0.0042</td>
<td>0.0008</td>
<td>-0.0004</td>
</tr>
<tr>
<td>( p_2(x,z) )</td>
<td>40</td>
<td>-0.0027</td>
<td>0.0020</td>
<td>-0.0027</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>0.0008</td>
<td>-0.0083</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td>80</td>
<td>0.0008</td>
<td>0.0078</td>
<td>0.0002</td>
<td>0.0008</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>-0.0020</td>
<td>0.0021</td>
<td>-0.0023</td>
<td>-0.0015</td>
</tr>
<tr>
<td>( p_3(x,z) )</td>
<td>40</td>
<td>0.0115</td>
<td>0.0180</td>
<td>0.0113</td>
<td>0.0024</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>0.0013</td>
<td>0.0079</td>
<td>0.0016</td>
<td>-0.0023</td>
</tr>
<tr>
<td></td>
<td>80</td>
<td>0.0003</td>
<td>0.0070</td>
<td>0.0007</td>
<td>-0.0009</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.0002</td>
<td>0.0046</td>
<td>-0.0008</td>
<td>-0.0006</td>
</tr>
</tbody>
</table>

### Table 4
MSE of \( \hat{\beta}_{af}, \hat{\beta}_{ad}, \hat{\beta}_{C} \) and \( \hat{\beta}_{f}. \)

<table>
<thead>
<tr>
<th>( p(x,z) )</th>
<th>( n )</th>
<th>( \hat{\beta}_{af} )</th>
<th>( \hat{\beta}_{ad} )</th>
<th>( \hat{\beta}_{C} )</th>
<th>( \hat{\beta}_{f} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( p_1(x,z) )</td>
<td>40</td>
<td>0.0055</td>
<td>0.0058</td>
<td>0.0055</td>
<td>0.0034</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>0.0020</td>
<td>0.0020</td>
<td>0.0020</td>
<td>0.0017</td>
</tr>
<tr>
<td></td>
<td>80</td>
<td>0.0019</td>
<td>0.0014</td>
<td>0.0018</td>
<td>0.0009</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.0008</td>
<td>0.0008</td>
<td>0.0007</td>
<td>0.0005</td>
</tr>
<tr>
<td>( p_2(x,z) )</td>
<td>40</td>
<td>0.0106</td>
<td>0.0103</td>
<td>0.0099</td>
<td>0.0039</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>0.0033</td>
<td>0.0034</td>
<td>0.0031</td>
<td>0.0017</td>
</tr>
<tr>
<td></td>
<td>80</td>
<td>0.0032</td>
<td>0.0026</td>
<td>0.0033</td>
<td>0.0010</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.0015</td>
<td>0.0015</td>
<td>0.0016</td>
<td>0.0006</td>
</tr>
<tr>
<td>( p_3(x,z) )</td>
<td>40</td>
<td>0.0213</td>
<td>0.0165</td>
<td>0.0210</td>
<td>0.0035</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>0.0073</td>
<td>0.0065</td>
<td>0.0075</td>
<td>0.0018</td>
</tr>
<tr>
<td></td>
<td>80</td>
<td>0.0029</td>
<td>0.0020</td>
<td>0.0025</td>
<td>0.0008</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.0032</td>
<td>0.0025</td>
<td>0.0032</td>
<td>0.0006</td>
</tr>
</tbody>
</table>
above three cases, we generated, respectively, 1000 Monte Carlo random samples of size $n=40, 60, 80$ and 100. For the 1000 simulated values of $\hat{\theta}_{af}, \hat{\theta}_{nf}, \hat{\theta}_{al}, \hat{\theta}_{c}, \hat{\theta}_{F}$ and $\hat{\theta}_{F}$, the biases and MSE of these estimators were calculated. The bandwidths $h_i, i=1,2,3,4$, were obtained by the bandwidth selection method suggested in Section 5. These simulated results are reported in Tables 1–4. From the 1000 simulated values of $\hat{\gamma}_{af}(\cdot), \hat{\gamma}_{nf}(\cdot), \hat{\gamma}_{c}(\cdot)$ and $\hat{\gamma}_{F}(\cdot)$, we plotted the simulated

Table 5

<table>
<thead>
<tr>
<th>$p(x,z)$</th>
<th>$n$</th>
<th>$\hat{\gamma}_{af}(\cdot)$</th>
<th>$\hat{\gamma}_{nf}(\cdot)$</th>
<th>$\hat{\gamma}_{c}(\cdot)$</th>
<th>$\hat{\gamma}_{F}(\cdot)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p_1(xz)$</td>
<td>40</td>
<td>0.0744</td>
<td>0.0550</td>
<td>0.0740</td>
<td>0.0452</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>0.0434</td>
<td>0.0311</td>
<td>0.0436</td>
<td>0.0258</td>
</tr>
<tr>
<td></td>
<td>80</td>
<td>0.0401</td>
<td>0.0225</td>
<td>0.0401</td>
<td>0.0180</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.0240</td>
<td>0.0168</td>
<td>0.0241</td>
<td>0.0125</td>
</tr>
<tr>
<td>$p_2(xz)$</td>
<td>40</td>
<td>0.1052</td>
<td>0.1084</td>
<td>0.1052</td>
<td>0.0480</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>0.0481</td>
<td>0.0434</td>
<td>0.0479</td>
<td>0.0261</td>
</tr>
<tr>
<td></td>
<td>80</td>
<td>0.0431</td>
<td>0.0339</td>
<td>0.0428</td>
<td>0.0181</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.0218</td>
<td>0.0247</td>
<td>0.0217</td>
<td>0.0124</td>
</tr>
<tr>
<td>$p_3(xz)$</td>
<td>40</td>
<td>0.1124</td>
<td>0.1112</td>
<td>0.1117</td>
<td>0.0531</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>0.0615</td>
<td>0.0647</td>
<td>0.0598</td>
<td>0.0262</td>
</tr>
<tr>
<td></td>
<td>80</td>
<td>0.0455</td>
<td>0.0443</td>
<td>0.0454</td>
<td>0.0178</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.0266</td>
<td>0.0320</td>
<td>0.0267</td>
<td>0.0124</td>
</tr>
</tbody>
</table>

Fig. 1. The big dot lines are the true curve $g_0(t)$; the solid lines are estimated curves $\hat{\gamma}_{af}(t)$; the dot lines are estimated curves $\hat{\gamma}_{nf}(t)$; the dashed lines are the estimated curves $\hat{\gamma}_{c}(t)$; the dot-dashed lines are the full data estimated curves $\hat{\gamma}_{F}(t)$, where $t = x^T\beta$. 
curves on the inner points and calculated the relative mean integrated square error (MISE). The results are reported in Table 5 and Fig. 1.

From Tables 1–4, generally, \( \hat{\beta}_C \) has the biggest bias and MSE. \( \hat{\beta}_{nl} \) has smaller bias and MSE than \( \hat{\beta}_{nl} \) and \( \hat{\beta}_C \), but larger than \( \hat{\beta}_r \). Furthermore, all the estimators of \( \hat{\theta} \) have small bias and MSE. Generally, \( \hat{\beta}_{nl} \) has smaller MSE than others. From Tables 1–4, we can see that with the sample size increasing, the bias and MSE of these estimators decrease when the missing probability \( p(\delta) = 1|X,z \) is fixed. Moreover, with the average response probability increasing, the bias and MSE of these estimators decrease when the sample size is fixed. From Table 5, in general, the proposed estimator \( \tilde{g}_{ml}(X^T\beta) \) outperforms \( \tilde{g}_m(X^T\beta) \) in terms of MISE.

Fig. 1 presented the estimated curves of \( g(t) \) when \( p(\delta) = 1|X,z \) = \( p_1(\delta) = 1|X,z \), \( p_2(\delta) = 1|X,z \) and \( p_3(\delta) = 1|X,z \), \( n = 40, 60, 80 \) and 100, respectively. From Fig. 1, we can see our proposed estimated curves are closer to the true one in general. Furthermore, with the sample size increasing, the MISE of the estimated curves of \( g(t) \) decreases when the missing probability is fixed. Furthermore, with the missing probability increasing, the MISE of the estimated curves of \( g(t) \) decreases when the sample size is fixed.

Acknowledgements

Qihua Wang’s research was supported by the National Science Fund for Distinguished Young Scholars in China (10725106), the National Natural Science Foundation of China (10671198), the National Science Fund for Creative Research Groups in China, the two grants from the Key Lab of Random Complex Structure and Data Science, CAS and Yunnan University.

Appendix A

In order to prove the asymptotic properties of the proposed estimators, we first prove some lemmas for the sake of convenience to prove the theorems.

Lemma A.1. Under the assumptions of Section 3, as \( \log n/nh_1 \rightarrow 0 \) and \( nh_1^2 \rightarrow \infty \),

\[
\hat{\theta}_n - \theta_0 = \frac{1}{n} \sum_{i=1}^{n} S^{-1} e_i \delta p_1(t_i)|p_1(t_i)|Z_i - m_2(t_i)| + o_p \left( \frac{1}{\sqrt{n}} \right),
\]

with \( S = E(\delta p_1(X^T\beta_0)|Z_i - m_2(X^T\beta_0))^2 \).

Proof. Recalling the definition of \( \hat{\theta}_n \) in (2.4), by some algebra we have

\[
\hat{\theta}_n = \frac{1}{n} \sum_{i=1}^{n} S^{-1} \delta [\tilde{p}_1(t_i)Y_i - p_1(t_i)Y_i + p_1(t_i)Y_i - m_1(t_i) + m_1(t_i) - \tilde{m}_1(t_i)] [\tilde{p}_1(t_i)Z_i - p_1(t_i)Z_i + p_1(t_i)Z_i - m_2(t_i) + m_2(t_i) - \tilde{m}_2(t_i)] \delta_i,
\]

where \( S_n = (1/n) \sum_{i=1}^{n} \delta [\tilde{p}_1(X_i^T\beta_0)|Z_i - m_2(X_i^T\beta_0)|] \tilde{p}_1(X_i^T\beta_0)Z_i - m_2(X_i^T\beta_0)ı^T t_i = X_i^T \beta_0.

By Mark and Silverman (1982) and Theorems 1 and 2 in Einmahl and Mason (2005), we can prove

\[
\sup_{\theta \in \Theta} \| \hat{\phi}(X_i^T\beta, \theta) - \phi(X_i^T\beta, \theta) \| = o_p \left( \frac{\log n}{nh_1} \right)^{1/2} + h_1^2,
\]

\[
E(\hat{\phi}_j(X_i^T\beta, \beta) - \phi_j(X_i^T\beta, \beta))^2 = O((nh_1)^{-1} + h_1^2),
\]

where \( \phi(\cdot) \) defines one of \( m_1(\cdot), m_2(\cdot) \) and \( p_1(\cdot), \phi_j(\cdot) \) and \( \tilde{\phi}_j(\cdot) \) denote the \( j \)th components of \( \phi(\cdot) \) and \( \hat{\phi}(\cdot) \), respectively. If \( \phi_j(\cdot), \tilde{\phi}_j(\cdot) \) denote \( m_2(\cdot) \) and \( \tilde{m}_2(\cdot) \), respectively, \( j = 1, \ldots, q \), \( j = 1 \) otherwise. By (A.1) and (A.2), we have

\[
S_n = \frac{1}{n} \sum_{i=1}^{n} \delta [\tilde{p}_1(t_i)Z_i - m_2(t_i)] [\tilde{p}_1(t_i)Z_i - m_2(t_i)] \delta_i^T
\]

\[
= \frac{1}{n} \sum_{i=1}^{n} \delta [\tilde{p}_1(t_i)Z_i - m_2(t_i)]^2 + o_p(1)
\]

\[
= E(\delta p_1(X_i^T\beta_0)|Z_i - m_2(X_i^T\beta_0))^2 + o_p(1) = S + o_p(1).
\]

This together with (A.2) and \( nh_1^2 \rightarrow \infty, \log n/nh_1 \rightarrow 0 \) proves

\[
\hat{\theta}_n - \theta_0 = \frac{1}{n} \sum_{i=1}^{n} S^{-1} e_i \delta p_1(t_i)|p_1(t_i)|Z_i - m_2(t_i)| + o_p \left( \frac{1}{\sqrt{n}} \right),
\]

\( t_i = X_i^T \beta \). \( \square \)
Lemma A.2. Under the assumptions of Lemma A.1,
\[ g_n(t) - g(t) = \frac{1}{2} \mu_2(K_2) g''(t) h_2^2 + \frac{1}{n} \sum_{i=1}^{n} \frac{1}{p_i(t)f_i(t)} \delta_i K_{h_1}(X_i^T \beta - t) \epsilon_i + o_p \left( \frac{1}{\sqrt{n} h_2} \right) \]
and
\[ h_2(g_n'(t) - g'(t)) = \frac{1}{n} \frac{1}{p_i(t)f_i(t) \mu_2(K_2)} \sum_{i=1}^{n} \delta_i K_{h_1}(X_i^T \beta - t) \frac{X_i^T \beta - t}{h_2} \epsilon_i + o_p \left( \frac{1}{\sqrt{n} h_2} \right). \]

Proof. From (2.5), it follows
\[
\begin{pmatrix}
\hat{g}_n(t) \\
\hat{h}_2 \hat{g}_n'(t)
\end{pmatrix} -
\begin{pmatrix}
g(t) \\
h_2 g'(t)
\end{pmatrix} = \left\{ \frac{1}{n} B_0^T W_1 B_0 \right\}^{-1} \frac{1}{n} \sum_{i=1}^{n} \left[ \delta_i K_{h_1}(X_i^T \beta - t) B_0 \left\{ \frac{1}{2} \left( \frac{X_i^T \beta - t}{h_2} \right)^2 g''(t) + \epsilon_i + o_p(h_2^2) \right\} \right].
\]
(Note that
\[
\frac{1}{n} B_0^T W_1 B_0 = p_1(t)f_1(t) \begin{pmatrix} 1 & 0 \\ 0 & \mu_2(K_2) \end{pmatrix} (1 + o_p(1)), \quad \mu_2(K_2) = \int u^2 K_2(u) \, du.
\]
Combining (A.4)-(A.6), Lemma A.2 is then proved.

Lemma A.3. Under the assumptions of Lemma A.1, if \( nh_2^4 \to 0 \),
\[ \hat{\beta}_{nf} - \beta_0 = \frac{1}{n} \sum_{i=1}^{n} \hat{E}_i, \quad \hat{E}_i = E(\delta g'(X_i^T \beta_0) X_i^T X_i), \]
where
\[ E_i = E(\delta g'(X_i^T \beta_0) X_i^T X_i), \]
\[ G(X_i^T \beta_0, X_i) = g'(X_i^T \beta_0) X_i - E(\delta g'(X_i^T \beta_0) X_i), \]
\[ Q(X_i^T \beta_0, Z_i) = \delta_i p_1(X_i^T \beta_0)p_1(X_i^T \beta_0)Z_i - m_2(X_i^T \beta_0). \]
\[ P_{\beta_0} = I - \beta_0 \beta_0^T \quad \text{and} \quad P_{\beta_0} \quad \text{is the generalized inverse matrix of} \quad P_{\beta_0}. \]

Proof. From (2.6), with \( \lambda \) being the Lagrange multiplier, \( \hat{\beta}_{nf} \) is the solution of
\[ \frac{1}{n} \sum_{i=1}^{n} \delta_i (Y_i - Z_i^T \hat{\beta}_{nf} - \hat{g}_{nf}(X_i^T \hat{\beta}_{nf})) \hat{g}'_{nf}(X_i^T \hat{\beta}_{nf})X_i - \lambda \hat{\beta}_{nf} = 0. \]

It can be rewritten as
\[ \frac{1}{n} \sum_{i=1}^{n} \delta_i (Y_i - Z_i^T \hat{\beta}_{nf} - \hat{g}_{nf}(X_i^T \hat{\beta}_{nf})) \hat{g}'_{nf}(X_i^T \hat{\beta}_{nf})X_i + \frac{1}{n} \sum_{i=1}^{n} \delta_i (g'(X_i^T \beta_0) - \hat{g}_{nf}(X_i^T \hat{\beta}_{nf})) \hat{g}'_{nf}(X_i^T \hat{\beta}_{nf})X_i + \frac{1}{n} \sum_{i=1}^{n} \delta_i \epsilon_i \hat{g}'_{nf}(X_i^T \hat{\beta}_{nf})X_i - \lambda \hat{\beta}_{nf} = 0. \]

Using Lemmas A.1 and A.2 with \( \hat{\beta}_{nf} \in B_0 \) and \( nh_2^4 \to 0 \), by (A.7) it is obtained
\[
\frac{1}{n} \sum_{i=1}^{n} \delta_i (g'(X_i^T \beta_0))^2 X_i^T (\hat{\beta}_{nf} - \beta_0)
\]
\[ = \frac{1}{n} \sum_{i=1}^{n} \delta_i g'(X_i^T \beta_0)(X_i - E(\delta g'(X_i^T \beta_0) X_i)) + \frac{1}{n} \sum_{i=1}^{n} E(\delta g'(X_i^T \beta_0) X_i^T) Q(X_i^T \beta_0, Z_i) \delta_i \epsilon_i + \lambda \hat{\beta}_{nf} + o_p \left( \frac{1}{\sqrt{n}} \right).
\]
\[ \begin{aligned}
\frac{1}{n} \sum_{i=1}^{n} \left| G(X_i^T \beta_0, X_i) - E(\delta g'(X_i^T \beta_0) X_i^T) Q(X_i^T \beta_0, Z_i) \right| \delta_i \epsilon_i + \lambda \hat{\beta}_{nf} + o_p \left( \frac{1}{\sqrt{n}} \right).
\end{aligned}
\]
\[ G(X_i^T \beta_0, X_i) = g'(X_i^T \beta_0)(X_i - E(\delta g'(X_i^T \beta_0) X_i)), \]
\[ Q(X_i^T \beta_0, Z_i) = S^{-1} p_1(X_i^T \beta_0)p_1(X_i^T \beta_0)Z_i - m_2(X_i^T \beta_0). \]
Referring to the method of Carroll et al. (1997), multiplying the above equation by \( I - \beta_0 \beta_0^T \), together with

\[
\lambda = \frac{1}{n} \sum_{i=1}^{n} \delta_i (Y_i - Z_i^T \tilde{\beta}_{af} - \hat{g}_{af}(X_i^T \tilde{\beta}_{af})) \tilde{g}_{af}(X_i^T \tilde{\beta}_{af}) X_i^T \tilde{\beta}_{af} = o_p(1),
\]

we have

\[
(I - \beta_0 \beta_0^T) \frac{1}{n} \sum_{i=1}^{n} \delta_i (g(X_i^T \beta_0))^2 X_i^T (\hat{\beta}_{af} - \beta_0) = -(I - \beta_0 \beta_0^T) \frac{1}{n} \sum_{j=1}^{n} (g(X_j^T \beta_0, X_j) - E(\delta g(X^T \beta_0)XZ^T)Q(X_j^T \beta_0, Z_j) ) \delta_j \epsilon_j + o_p\left(\frac{1}{\sqrt{n}}\right).
\]

This implies

\[
\hat{\beta}_{af} - \beta_0 = -\frac{1}{n} \sum_{j=1}^{n} \tilde{E}_1^{-1} P_{\beta_0}^T (g(X_j^T \beta_0, X_j) - \tilde{E}_2 Q(X_j^T \beta_0, Z_j) ) \delta_j \epsilon_j + o_p\left(\frac{1}{\sqrt{n}}\right),
\]

where

\[
\tilde{E}_1 = E(\delta g(X^T \beta_0))^2 XX^T), \quad \tilde{E}_2 = E(\delta g(X^T \beta_0)XZ^T),
\]

\( P_{\beta_0} = I - \beta_0 \beta_0^T \) and \( P_{\beta_0} \) is the generalized inverse matrix of \( P_{\beta_0} \). □

**Proof of Theorem 3.1.** For any \( \beta \in \mathcal{B}_n \), by (2.11), similar to the proof of Lemma A.1, we have

\[
\hat{\beta}_{af} = \hat{S}_n^{-1} \frac{1}{n} \sum_{i=1}^{n} (Y_i - m_{10}(X_i^T \beta_0))(Z_i - m_{20}(X_i^T \beta_0)) + \hat{S}_n^{-1} \frac{1}{n} \sum_{i=1}^{n} (Y_i - m_{10}(X_i^T \beta_0))(m_{20}(X_i^T \beta_0) - \hat{m}_{20}(X_i^T \beta))
\]

\[
+ \hat{S}_n^{-1} \frac{1}{n} \sum_{i=1}^{n} (m_{10}(X_i^T \beta_0) - \hat{m}_{10}(X_i^T \beta))(Z_i - m_{20}(X_i^T \beta_0)) + \hat{S}_n^{-1} \frac{1}{n} \sum_{i=1}^{n} (m_{10}(X_i^T \beta_0) - \hat{m}_{10}(X_i^T \beta))(m_{20}(X_i^T \beta_0) - \hat{m}_{20}(X_i^T \beta)), \quad (A.8)
\]

where \( \hat{S}_n = (1/n) \sum_{i=1}^{n} (Z_i - \hat{m}_{20}(X_i^T \beta))(Z_i - m_{20}(X_i^T \beta))^T \). It can be proved

\[
\bar{S} = \hat{S} + o_p(1), \quad \text{and } \bar{S} = E[(Z - m_{20}(X^T \beta))^2].
\]

Using Lemmas A.1-A.3 and (A.2), with \( nh_i h_i \to \infty, h_i^3 / h_i \to 0 \), from (A.8) it follows:

\[
\hat{\beta}_{af} - \beta_0 = \hat{S}_n^{-1} \frac{1}{n} \sum_{j=1}^{n} \frac{1}{P_j (X_j^T \beta_0)} D_0 (X_j^T \beta_0) \delta_j \epsilon_j - \hat{S}_n^{-1} \frac{1}{n} \sum_{j=1}^{n} D_j \tilde{E}_1^{-1} P_{\beta_0}^T P_{\beta_0} (g(X_j^T \beta_0, X_j) - \tilde{E}_2 Q(X_j^T \beta_0, Z_j)) \delta_j \epsilon_j
\]

\[
+ \hat{S}_n^{-1} \frac{1}{n} \sum_{j=1}^{n} D_2 Q(X_j^T \beta_0, Z_j) \delta_j \epsilon_j + \hat{S}_n^{-1} \frac{1}{n} \sum_{j=1}^{n} (Z_j - m_{20}(X_j^T \beta_0)) \delta_j \epsilon_j + o_p\left(\frac{1}{\sqrt{n}}\right),
\]

where

\[
D_0 (X_j^T \beta_0) = E[(Z - m_{20}(X_j^T \beta_0))X_j^T \beta_0],
\]

\[
D_1 = E[(1 - \delta)(X_j^T \beta_0)(Z - m_{20}(X_j^T \beta_0))X_j^T],
\]

\[
D_2 = E[(1 - \delta)(Z - m_{20}(X_j^T \beta_0))Z_j^T].
\]

By applying the central limit theorem to (A.9), we obtain

\[
\sqrt{n}(\hat{\beta}_{af} - \beta_0) \xrightarrow{D} N(0, \Sigma),
\]

where

\[
\Sigma = E[\delta D_0(X_j^T \beta_0, Z_j)D_1^T(X_j^T \beta_0, Z_j)]. \quad (A.10)
\]

Replacing \( \beta \) by \( \hat{\beta}_{af} \) in \( \hat{\beta}_{af} \), we complete the proof of Theorem 3.1. □

**Proof of Theorem 3.2.** By (2.13), we have

\[
\hat{g}_{af}(t, \beta) - g(t) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{f_i(t)} K_{n_i}(X_i^T \beta - t)(\theta_0 - \hat{\theta}_{af})^T Z_i \delta_i + \frac{1}{n} \sum_{i=1}^{n} \frac{1}{f_i(t)} K_{n_i}(X_i^T \beta - t)(g(X_i^T \beta_0) - \hat{g}_{af}(X_i^T \hat{\theta}_{af})) \delta_i
\]
By the central limit theorem, it follows:
\[
\frac{1}{n} \sum_{i=1}^{n} \left( \frac{1}{n} \sum_{l=1}^{n} K_h(X_l^T \beta - t) \bar{\delta}_i \bar{e}_l + \frac{1}{n} \sum_{l=1}^{n} K_h(X_l^T \beta - \bar{t}) \bar{Z}_l (\bar{\theta}_{nl} - \theta_0) + \frac{1}{n} \sum_{l=1}^{n} K_h(X_l^T \beta - t) (\bar{g}_{nl}(X_l^T \bar{\beta}_{nl}) - g(X_l^T \beta_0)) \right) + \frac{1}{n} \sum_{l=1}^{n} K_h(X_l^T \beta - t) \left( \frac{1}{2} \left( \frac{X_l^T \beta - t}{\sigma^2} \right)^2 g''(t) \right) + \frac{1}{n} \sum_{l=1}^{n} K_h(X_l^T \beta - t) (\bar{\theta}_{nl} - \theta_0)^T Z_l + o_p \left( \frac{1}{\sqrt{n \sigma^4}} \right). \tag{A.11}
\]

By \(h_4/h_2 \to 0\), \(n \to \infty\), together with Lemmas A.1–A.3 and Theorem 3.1, we obtain
\[
\bar{g}_{nl}(t, \beta) - g(t) = \frac{1}{n} \sum_{l=1}^{n} K_h(X_l^T \beta - t) \bar{\delta}_i \bar{e}_l + o_p \left( \frac{1}{\sqrt{n \sigma^4}} \right). \tag{A.12}
\]

By the central limit theorem, it follows:
\[
\sqrt{n \sigma^4} (\bar{g}_{nl}(t, \beta) - g(t)) \xrightarrow{D} N \left( 0, \frac{D}{f_1(t)} \right),
\]
where \(\gamma_2(K_h) = \int K_h^2(u) \, du\).

Therefore, with \(\beta\) replaced by \(\hat{\beta}_{nl}\) in \(\bar{g}_{nl}(\cdot)\), we can obtain the result of Theorem 3.2. \(\square\)

**Proof of Theorem 3.3.** The estimated equation can be written as
\[
\frac{1}{n} \sum_{i=1}^{n} (Y_i - \bar{\bar{X}}_n^T \bar{\beta}_{nl}) \bar{\delta}_i \bar{e}_l = 0.
\]

By some algebra, we have
\[
\frac{1}{n} \sum_{i=1}^{n} (\bar{\bar{X}}_n^T \bar{\beta}_{nl}) \bar{\delta}_i \bar{e}_l + \frac{1}{n} \sum_{i=1}^{n} (1 - \bar{\delta}_i) (\bar{\bar{X}}_n^T \bar{\beta}_{nl}) \bar{Z}_i \bar{g}_{nl}(X_i^T \bar{\beta}_{nl}) X_i + \frac{1}{n} \sum_{i=1}^{n} (1 - \bar{\delta}_i)(\bar{g}_{nl}(X_i^T \bar{\beta}_{nl}) - g(X_i^T \beta_0)) \bar{g}_{nl}(X_i^T \bar{\beta}_{nl}) X_i
\]
\[
+ \frac{1}{n} \sum_{i=1}^{n} \left( g(X_i^T \beta_0) - \bar{g}_{nl}(X_i^T \bar{\beta}_{nl}) \right) \bar{g}_{nl}(X_i^T \bar{\beta}_{nl}) X_i + \frac{1}{n} \sum_{i=1}^{n} \bar{\delta}_i \bar{g}_{nl}(X_i^T \bar{\beta}_{nl}) X_i \bar{e}_l - \lambda \bar{\beta}_{nl} = 0. \tag{A.13}
\]

With the conditions of Theorem 3.1 satisfied, we rewrite (A.13) as
\[
\hat{F}_4(\bar{\beta}_{nl} - \beta_0) = \frac{1}{n} \sum_{i=1}^{n} \hat{F}_1 \bar{D}_3(X_i^T \beta_0, Z_i) \bar{\delta}_l \bar{e}_j + \frac{1}{n} \sum_{i=1}^{n} \hat{F}_2 Q(X_i^T \beta_0, Z_i) \bar{\delta}_l \bar{e}_j + \frac{1}{n} \sum_{i=1}^{n} \frac{E(1 - \bar{\delta}_i) X_i^T \beta_0}{p_1(X_i^T \beta_0)} g(X_i^T \beta_0) \bar{\delta}_l \bar{e}_j
\]
\[
- \frac{1}{n} \sum_{i=1}^{n} \hat{F}_3 \hat{E}_1^{-1} P_{\beta_0} P_{\beta_0} (g(X_i^T \beta_0, X_l) - \hat{E}_2 Q(X_i^T \beta_0, Z_i)) \bar{\delta}_l \bar{e}_j - \lambda \bar{\beta}_{nl}
\]
\[
= \frac{1}{n} \sum_{i=1}^{n} \hat{F}(X_i^T \beta_0, Z_i) \bar{\delta}_l \bar{e}_j - \lambda \bar{\beta}_{nl} + o_p \left( \frac{1}{\sqrt{n \sigma^4}} \right). \tag{A.14}
\]

where
\[
\hat{F}_1 = E(g(X_i^T \beta_0) X_i Z_i), \quad \hat{F}_2 = E(1 - \bar{\delta}_i) g(X_i^T \beta_0) X_i Z_i T),
\]
\[
\hat{F}_3 = E(1 - \bar{\delta}_i) g(X_i^T \beta_0)^2 X_i Z_i T), \quad \hat{F}_4 = E(g(X_i^T \beta_0)^2 X_i Z_i T),
\]
\[
\hat{F}(X_i^T \beta_0, Z_i) = \hat{F}_1 \bar{D}_3(X_i^T \beta_0, Z_i) + \hat{F}_2 Q(X_i^T \beta_0, Z_i) + \frac{E(1 - \bar{\delta}_i) X_i^T \beta_0}{p_1(X_i^T \beta_0)} g(X_i^T \beta_0) - \hat{F}_3 \hat{E}_1^{-1} P_{\beta_0} P_{\beta_0} (g(X_i^T \beta_0, X_l) - \hat{E}_2 Q(X_i^T \beta_0, Z_i)). \tag{A.15}
\]

Multiplying \(1 - \bar{\delta}_i \beta_0^0\) on both sides of (A.14) and applying the CLT, similar to the proof of Lemma A.3, we have
\[
\sqrt{n \sigma^4} (\bar{\beta}_{nl} - \beta_0) \xrightarrow{D} N(0, \sigma^2 \hat{F}_4^{-1} \Sigma_1 \hat{F}_4^{-1}),
\]
where
\[
\Sigma_1 = P_{\beta_0} E(\bar{\delta}_l \hat{F}(X_i^T \beta_0, Z_i) \hat{F}^T(X_i^T \beta_0, Z_i))P_{\beta_0}^{-1}. \quad \square
\]

**References**


