Available online at www.sciencedirect.com
ScienceDirect
Journal of Multivariate Analysis 98 (2007) 1470–1493
Journal of Multivariate Analysis
www.elsevier.com/locate/jmva

Estimation in partially linear models with missing responses at random

Qihua Wang[a, b, *], Zhihua Sun[a]

[a]*Academy of Mathematics and Systems Science, Chinese Academy of Science, Beijing 100080, P.R. China*
[b]*Department of Statistics and Actuarial Science, The University of Hong Kong, Pokfulam Road, Hong Kong*

Received 11 November 2005
Available online 29 November 2006

Abstract

A partially linear model is considered when the responses are missing at random. Imputation, semiparametric regression surrogate and inverse marginal probability weighted approaches are developed to estimate the regression coefficients and the nonparametric function, respectively. All the proposed estimators for the regression coefficients are shown to be asymptotically normal, and the estimators for the nonparametric function are proved to converge at an optimal rate. A simulation study is conducted to compare the finite sample behavior of the proposed estimators.

AMS 2000 subject classification: Primary 62J99; secondary 62E20

Keywords: Imputation estimator; Regression surrogate estimator; Inverse marginal probability weighted estimator; Asymptotic normality

1. Introduction

Consider the partial linear model

$$Y = X^{\tau}\beta + g(T) + \varepsilon, \tag{1.1}$$

where Y is a scalar response variate, X is a p-variate random covariate vector and T is a scalar covariate taking values in [0, 1], and where β is a $p \times 1$ column vector of unknown regression parameter, $g(\cdot)$ is an unknown measurable function on [0, 1] and ε is a random statistical error with $E[\varepsilon|X, T] = 0$.

* Corresponding author. Fax: +852 28589041.
E-mail address: qhwang@amss.ac.cn (Q. Wang).

0047-259X/$ - see front matter
doi:10.1016/j.jmva.2006.10.003

Model (1.1) has gained much attention in recent years. Speckman [17] gave an application of the partially linear model to a mouthwash experiment. Schmalensee and Stoker [16] used the partially linear model to analyze household gasoline consumption in the United States. Green and Silverman [5] provided an example of the use of partially linear models, and compared their results with a classical approach. Zeger and Diggle [23] used a semiparametric mixed model to analyze the CD4 cell count in HIV seroconverters where $g(\cdot)$ is estimated by a kernel smoother. Hu et al. [10] studied the profile-kernel and backfitting methods for the model. The partially linear model has been applied in various fields such as biometrics, see Gray [4], econometrics, see Ahn and Powell [1], and so on. The model has been studied extensively for complete data setting, see Heckman [8], Rice [13], Speckman [17], Robinson [15] among others.

In practice, some response variables may be missing, by design (as in two-stage studies) or by happenstance. For example, the response Y's may be very expensive to measure and only part of Y's are available. Another example is that Y's represent the responses to a set of questions and some sampled individuals refuse to supply the desired information. Actually, missingness of responses is very common in opinion polls, market research surveys, mail enquiries, social-economic investigations, medical studies and other scientific experiments. Wang et al. [21] developed inference tools for the mean of Y in model (1.1) with missing response data.

In this paper, we develop some approaches of estimating β and $g(\cdot)$ with responses missing. Suppose we obtain a random sample of incomplete data

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

from model (1.1), where $\delta_i = 0$ if Y_i is missing, otherwise $\delta_i = 1$. Throughout this paper, we assume that Y is missing at random (MAR). The MAR assumption implies that δ and Y are conditionally independent given X and T. That is, $p(\delta = 1|Y, X, T) = p(\delta = 1|X, T)$. MAR is a common assumption for statistical analysis with missing data and is reasonable in many practical situations; see Little and Rubin [11].

To deal with missing data, one method is to impute a plausible value for each missing datum and then analyze the results as if they are complete. In regression problems, commonly used imputation approaches include linear regression imputation [7] , nonparametric kernel regression imputation [3,22], semiparametric regression imputation [21], among others. Wang et al. [21] developed semiparametric imputation approach to estimate the mean of Y. We here extend the method to the estimation of β and $g(\cdot)$.

It is interesting to note that Matloff [12] verified that if the form of regression is known and only characterized by some unknown parameter, the method of replacing the responses by the estimated regression values outperforms that of using the observed responses directly for the estimation of means. Motivated by Matloff [12], we develop a so-called semiparametric regression surrogate approach. This method is just to use the estimated semiparametric regression values instead of the corresponding response values to define estimators whether the responses are observed or not. Our research results also verify that the semiparametric regression surrogate approach indeed works well. Similar methods are also used by Cheng [3] and Wang et al. [21], where the methods are also competitive.

It is well known that the inverse probability weighted approach is another popular method to handle missing data. The inverse weighted approach has gained considerable attention to missing data problems. See Zhao, Lipsitz and Lew [24], Wang et al. [19], Robins, Rotnitzky and Zhao [14] and Wang, Lindon and Härdle [21]. For missing response problems, the inverse probability weighted approach usually depends on high-dimensional smoothing for estimating the completely unknown propensity score function, and hence the well known "curse of dimensionality" may restrict the use of this estimator. Wang et al. [21] suggested an inverse marginal probability weighted method to estimate the mean of Y, which avoids the problem of "curse of dimensionality". Furthermore, it is shown that the resulting estimator has a credible "double robustness" property. This motivates us to employ the inverse marginal probability weighted method to estimate β and $g(\cdot)$.

The rest of this paper is organized as follows. In Section 2, we define imputation estimators of β and $g(\cdot)$, and investigate their asymptotic properties. In Sections 3 and 4, we develop a semiparametric regression surrogate method and an inverse marginal probability weighted method to estimate β and $g(\cdot)$, and investigate their asymptotic properties, respectively. In Section 5, we conduct a simulation study to compare the finite sample properties of these suggested estimators. The proofs of the main results are presented in the appendix.

2. Imputation estimators and asymptotic properties

Let $Z=(X,T)$, $\sigma^2(Z)=E(\varepsilon^2|Z)$, $\Delta(z)=P(\delta=1|Z=z)$ and $\Delta_t(t)=P(\delta=1|T=t)$. Let $U_i^{[\mathrm{I}]}=\delta_i Y_i+(1-\delta_i)(X_i^\tau\beta+g(T_i))$, that is, $U_i^{[\mathrm{I}]}=Y_i$ if $\delta_i=1$, otherwise, $U_i^{[\mathrm{I}]}=X_i^\tau\beta+g(T_i)$. By MAR assumption, we have $E[U^{[\mathrm{I}]}|Z]=E[\delta Y+(1-\delta)(X^\tau\beta+g(T))]|Z)=X^\tau\beta+g(T)$. This implies

$$U_i^{[\mathrm{I}]}=X_i^\tau\beta+g(T_i)+e_i, \tag{2.1}$$

where $E[e_i|Z_i]=0$. This is just the form of the standard partial linear model. Let

$$\omega_{ni}(t)=\frac{M\left(\frac{t-T_i}{b_n}\right)}{\sum_{i=1}^n M\left(\frac{t-T_i}{b_n}\right)},$$

where $M(\cdot)$ is a kernel function and b_n is a bandwidth sequence. Standard approach can be used to define the following estimator of β:

$$\tilde{\beta}_{\mathrm{I}}=\left[\sum_{i=1}^n(X_i-\tilde{g}_{1n}(T_i))(X_i-\tilde{g}_{1n}(T_i))^\tau\right]^{-1}\sum_{i=1}^n(X_i-\tilde{g}_{1n}(T_i))(U_i^{[\mathrm{I}]}-\tilde{g}_{2n}^{[\mathrm{I}]}(t)), \tag{2.2}$$

where $\tilde{g}_{1n}(t)$ and $\tilde{g}_{2n}^{[\mathrm{I}]}(t)$ are, respectively, given by

$$\tilde{g}_{1n}(t)=\sum_{i=1}^n\omega_{ni}(t)X_i,\quad \tilde{g}_{2n}^{[\mathrm{I}]}(t)=\sum_{i=1}^n\omega_{ni}(t)U_i^{[\mathrm{I}]}. \tag{2.3}$$

Let

$$\omega_{nj}^{\mathrm{C}}(t)=\frac{K\left(\frac{t-T_j}{h_n}\right)}{\sum_{j=1}^n\delta_j K\left(\frac{t-T_j}{h_n}\right)},$$

where $K(\cdot)$ is a kernel function and h_n is a bandwidth sequence. Clearly, $U_i^{[\mathrm{I}]}$ contains unknown β and $g(T_i)$. Hence $\tilde{\beta}_{\mathrm{I}}$ is not a true estimator. Naturally, we replace $U_i^{[\mathrm{I}]}$ by

$$U_{ni}^{[\mathrm{I}]}=\delta_i Y_i+(1-\delta_i)(X_i^\tau\hat{\beta}_{\mathrm{C}}+g_n^{\mathrm{C}}(T_i)) \tag{2.4}$$

in (2.2) and denote the corresponding estimator by $\hat{\beta}_{\rm I}$, where $\hat{\beta}_{\rm C}$ and $g_n^{\rm C}(T_i)$ are given, respectively, by

$$\hat{\beta}_{\rm C} = \left[\sum_{i=1}^n \delta_i (X_i - g_{1n}^{\rm C}(T_i))(X_i - g_{1n}^{\rm C}(T_i))^\tau\right]^{-1} \sum_{i=1}^n \delta_i (X_i - g_{1n}^{\rm C}(T_i))(Y_i - g_{2n}^{\rm C}(T_i)) \tag{2.5}$$

and

$$g_n^{\rm C}(t) = g_{2n}^{\rm C}(t) - g_{1n}^{\rm C}(t)^\tau \hat{\beta}_{\rm C}, \tag{2.6}$$

where

$$g_{1n}^{\rm C}(t) = \sum_{j=1}^n \delta_j \omega_{nj}^{\rm C}(t) X_j, \quad g_{2n}^{\rm C}(t) = \sum_{j=1}^n \delta_j \omega_{nj}^{\rm C}(t) Y_j. \tag{2.7}$$

Let $g_1(t) = E[X|T=t]$ and $g_2(t) = E[Y|T=t] = E[U^{[\rm I]}|T=t]$. From (2.1), by taking expectation of T, we have

$$g(t) = g_2(t) - g_1(t)^\tau \beta. \tag{2.8}$$

Then, $g(t)$ can be estimated by

$$\hat{g}_n^{[\rm I]}(t) = g_{2n}^{[\rm I]}(t) - g_{1n}(t)^\tau \hat{\beta}_I, \tag{2.9}$$

where $g_{1n}(t)$ is $\widetilde{g}_{1n}(t)$ and $g_{2n}^{[\rm I]}(t)$ is $\widetilde{g}_{2n}^{[\rm I]}(t)$ with $U_i^{[\rm I]}$ replaced by $U_{ni}^{[\rm I]}$ for $i = 1, 2, \ldots, n$.

Denote $\check{X} = X - E(X|T)$ and $\widetilde{X} = X - \frac{E(\delta X|T)}{E(\delta|T)}$. Let

$$\Sigma_0 = E[\Delta(Z)\widetilde{X}\widetilde{X}^\tau], \quad \Sigma_1 = E[\check{X}\check{X}^\tau], \quad \Sigma_2 = E[(1-\Delta(Z))\check{X}\widetilde{X}^\tau].$$

Theorem 2.1. *Under all the assumptions listed in appendix, except* (b)(i) *and* (c)(iii), *we have*

$$\sqrt{n}(\hat{\beta}_{\rm I} - \beta) \xrightarrow{L} N(0, \Sigma_1^{-1} V_{\rm I} \Sigma_1^{-1}),$$

where

$$V_{\rm I} = (\Sigma_2 + \Sigma_0)\Sigma_0^{-1} E[\Delta(Z)\widetilde{X}\widetilde{X}^\top \sigma^2(Z)])\Sigma_0^{-1}(\Sigma_2 + \Sigma_0).$$

If δ_i is independent of X_i given T_i, by simple computation, the asymptotic variance of $\hat{\beta}_{\rm I}$ reduces to $\Sigma_{01}^{-1} E[\Delta_t(T)\check{X}\check{X}^\tau \sigma^2(Z)]\Sigma_{01}^{-1}$, where $\Sigma_{01} = E[\Delta_t(T)\check{X}\check{X}^\top]$. Furthermore, if $\Delta(\cdot)$ and hence $\Delta_t(\cdot)$ equal to a constant a, i.e. under the assumption of missing completely at random, it is easy to see that the asymptotic variance reduces to $\frac{1}{a}\Sigma_1^{-1} E[\check{X}\check{X}^\tau \sigma^2(Z)]\Sigma_1^{-1}$. Specifically, if $\Delta(Z) = 1$, the asymptotic variance is $\Sigma_1^{-1} E[\check{X}\check{X}^\tau \sigma^2(Z)]\Sigma_1^{-1}$, which is just the asymptotic variance of the standard estimator when the data are observed completely (See [2]).

To define a consistent estimator of the asymptotic variance, a natural way is first to define estimators of $\Delta(z)$, $\sigma^2(z)$, $E[X|T]$, $E[\delta X|T]$ and $E[\delta|T]$ using kernel regression method and then define a consistent estimator by combining sample moment approach and "plug in" method. However, this method may not provide a good estimator of the asymptotic variance in high dimensions. Kernel smoothing can be avoided because $\Delta(z)$ and $\sigma^2(z)$ only enter in the numerator and hence can be replaced by the indicator function or squared residuals where appropriate. For example, Σ_0 can be estimated consistently by

$$\widehat{\Sigma}_{0n} = \frac{1}{n}\sum_{i=1}^{n} \delta_i (X_i - g_{1n}^{\mathrm{C}}(T_i))(X_i - g_{1n}^{\mathrm{C}}(T_i))^{\top},$$

where $g_{1n}^{\mathrm{C}}(t)$ is defined in (2.7).

Theorem 2.2. *Under conditions of Theorem* 2.1, *if* $b_n = O_p(n^{-\frac{1}{3}})$ *and* $h_n = O_p(n^{-\frac{1}{3}})$, *we have*

$$\widehat{g}_n^{[\mathrm{I}]}(t) - g(t) = O_p(n^{-\frac{1}{3}}).$$

The proofs of Theorems 2.1 and 2.2 are given in the Appendix. Theorem 2.2 shows that $\widehat{g}_n^{[\mathrm{I}]}(t)$ attains the optimal convergence rate of nonparametric kernel regression estimator. See Stone [18].

3. Semiparametric regression surrogate estimators and asymptotic properties

In this section, we develop a so-called semiparametric regression surrogate approach. This method uses estimated semiparametric regression values instead of the corresponding response values to define estimators, whether the responses are observed or not. Let

$$U_{ni}^{[\mathrm{R}]} = X_i^{\tau}\widehat{\beta}_{\mathrm{C}} + g_n^{\mathrm{C}}(T_i). \tag{3.1}$$

The semiparametric regression surrogate estimator of β, written $\widehat{\beta}_{\mathrm{R}}$, can be defined to be $\widehat{\beta}_{\mathrm{I}}$ with $U_{ni}^{[\mathrm{I}]}$ in it replaced by $U_{ni}^{[\mathrm{R}]}$ for $i = 1, 2, \ldots, n$. The estimator of $g(\cdot)$, written $\hat{g}_n^{[\mathrm{R}]}(\cdot)$, can be defined to be $\widehat{g}_n^{[\mathrm{I}]}(\cdot)$ with $U_{ni}^{[\mathrm{I}]}$ and $\widehat{\beta}_{\mathrm{I}}$ in it replaced by $U_{ni}^{[\mathrm{R}]}$ and $\widehat{\beta}_{\mathrm{R}}$, respectively.

Theorem 3.1. *Under the assumptions of Theorem* 2.1, *we have*

$$\sqrt{n}(\hat{\beta}_{\mathrm{R}} - \beta) \xrightarrow{L} N(0, \Sigma_1^{-1} V_{\mathrm{R}} \Sigma_1^{-1}),$$

where

$$V_{\mathrm{R}} = \Sigma_1\, \Sigma_0^{-1}\, E[\widetilde{X}\widetilde{X}^{\tau}\Delta(Z)\sigma^2(Z)]\Sigma_0^{-1}\Sigma_1.$$

It is interesting to note that $\widehat{\beta}_{\mathrm{R}}$ has the same asymptotic variance as $\widehat{\beta}_{\mathrm{I}}$. This can be seen under the MAR condition by noting

$$\begin{aligned}\Sigma_0 + \Sigma_2 = {} & E[(X - E[X|T])(X - E[X|T])] \\ & + E\left[(1+\Delta(Z))(X - E[X|T])\left(E[X|T] - \frac{E[\delta X|T]}{E[\delta|T]}\right)^{\top}\right] \\ & + E\left[\Delta(Z)\left(E[X|T] - \frac{E[\delta X|T]}{E[\delta|T]}\right)\left(E[X|T] - \frac{E[\delta X|T]}{E[\delta|T]}\right)^{\top}\right] \\ = {} & \Sigma_1 + E\left[\delta\left(X - \frac{E[\delta X|T]}{E[\delta|T]}\right)\left(E[X|T] - \frac{E[\delta X|T]}{E[\delta|T]}\right)^{\top}\right] = \Sigma_1,\end{aligned}$$

where Σ_0, Σ_1 and Σ_2 are defined in Section 2.

Theorem 3.2. *Under conditions of Theorem* 3.1, *if* $b_n = O_p(n^{-\frac{1}{3}})$ *and* $h_n = O_p(n^{-\frac{1}{3}})$, *we have*

$$\hat{g}_n^{[\mathrm{R}]}(t) - g(t) = O_p(n^{-\frac{1}{3}}).$$

The proof of Theorems 3.1 and 3.2 are presented in appendix.

4. Inverse marginal probability weighted estimators and asymptotic properties

We note that under the MAR condition,

$$E\left[\frac{\delta_i}{\Delta(Z_i)}Y_i + \left(1 - \frac{\delta_i}{\Delta(Z_i)}\right)(X_i^\tau\beta + g(T_i))|Z_i\right] = X_i^\tau\beta + g(T_i).$$

Similar to Section 2, one can use the above equation to estimate β and $g(\cdot)$. But this method concerns the nonparametric regression estimator of $\Delta(z)$ and hence the well known "curse of dimensionality" problem may occur if the dimension of X is high. Motivated by Wang et al. [21], we use the inverse marginal probability weighted approach. Let

$$U_i^{[\mathrm{IP}]} = \frac{\delta_i}{\Delta_t(T_i)}Y_i + \left(1 - \frac{\delta_i}{\Delta_t(T_i)}\right)(X_i^\tau\beta + g(T_i)) \tag{4.1}$$

and taking expectation of Z, we have $E(U_i^{[\mathrm{IP}]}|Z_i) = X_i^\tau\beta + g(T_i)$. Hence

$$U_i^{[\mathrm{IP}]} = X_i^\tau\beta + g(T_i) + \eta_i, \tag{4.2}$$

where $\eta_i's$ satisfy $E[\eta_i|Z_i] = 0$. Let

$$\widetilde{\omega}_{ni}(t) = \frac{\Omega\left(\frac{t-T_i}{\gamma_n}\right)}{\sum_{j=1}^n \Omega\left(\frac{t-T_j}{\gamma_n}\right)},$$

where $\Omega(\cdot)$ is a kernel function and γ_n is a bandwidth sequence. Formula (4.2) is a standard partial linear model. Hence, similar to Section 2, the inverse marginal probability weighted estimator of β, say $\widehat{\beta}_{\mathrm{IP}}$, can be defined to be $\widehat{\beta}_{\mathrm{I}}$ with $U_{ni}^{[\mathrm{I}]}$ replaced by $U_{ni}^{[\mathrm{IP}]}$, and the estimator of $g(\cdot)$, $\hat{g}_n^{[\mathrm{IP}]}(t)$, can be defined to be $\widehat{g}_n^{[\mathrm{I}]}(\cdot)$ with $U_{ni}^{[\mathrm{I}]}$ and $\widehat{\beta}_{\mathrm{I}}$ replaced by $U_{ni}^{[\mathrm{IP}]}$ and $\widehat{\beta}_{\mathrm{IP}}$, where

$$U_{ni}^{[\mathrm{IP}]} = \frac{\delta_i}{\hat{\Delta}_t(T_i)}Y_i + \left(1 - \frac{\delta_i}{\hat{\Delta}_t(T_i)}\right)(X_i^\tau\hat{\beta}_{\mathrm{C}} + g_n^{\mathrm{C}}(T_i))$$

with

$$\hat{\Delta}_t(T_i) = \sum_{i=1}^n \widetilde{\omega}_{ni}(t)\delta_i.$$

Let

$$L(T) = \frac{\Sigma_0}{\Delta_t(T)} + E\left(\left(1 - \frac{\delta}{\Delta_t(T)}\right)(X - g_1(T))(X - g_1^{\mathrm{C}}(T))^\top\right).$$

Theorem 4.1. *Under all the assumptions listed in appendix, we have*

$$\sqrt{n}(\hat{\beta}_{\mathrm{IP}} - \beta) \xrightarrow{L} N(0, \Sigma_1^{-1} V_{\mathrm{IP}} \Sigma_1^{-1}),$$

where

$$V_{\mathrm{IP}} = E\Big\{ L(T)\Sigma_0^{-1}(X - g_1^{\mathrm{C}}(T))(X - g_1^{\mathrm{C}}(T))^\top \Sigma_0^{-1} L(T)\Delta(Z)\sigma^2(Z) \Big\}.$$

In theory, it seems difficult to compare the asymptotic variance of $\widehat{\beta}_{\mathrm{IP}}$ with that of $\widehat{\beta}_{\mathrm{I}}$ and $\widehat{\beta}_{\mathrm{R}}$. We will make a simulation comparison between them. Next, we discuss some special cases. If δ_i is independent of X_i given T_i, the asymptotic variance reduces to $\Sigma_1^{-1} E\left[\check{X}\check{X}^\top \frac{\sigma^2(Z)}{\Delta_t(T)} \right] \Sigma_1^{-1}$. Under MCAR, the asymptotic variance is the same as that of $\widehat{\beta}_{\mathrm{I}}$ and $\widehat{\beta}_{\mathrm{R}}$. In the special case of $\Delta(Z) = 1$, the asymptotic variance reduces to that of the standard estimator due to Chen [2] with data observed completely.

The asymptotic variance can be estimated by the method similar to that used in the estimating of the asymptotic variance of $\widehat{\beta}_{\mathrm{I}}$.

Theorem 4.2. *Under conditions of Theorem* 4.1, *if* $b_n = O(n^{-\frac{1}{3}})$, $h_n = O(n^{-\frac{1}{3}})$ *and* $\gamma_n = O(n^{-\frac{1}{3}})$, *we have*

$$\hat{g}_n^{[\mathrm{IP}]}(t) - g(t) = O_p(n^{-\frac{1}{3}}).$$

5. Bandwidth selection

It is well known that an important issue in applying kernel regression estimate is the selection of an appropriate bandwidth sequence. This issue has been extensively studied in the context of nonparametric regression. One of bandwidth selection rules is the delete-one cross-validation rule. Hong [9] extend this method to the partially linear regression setting. Here, we further extend this method to the partially linear regression problem when responses are MAR. It is noted that our estimators involve two or three bandwidths. Hence, it is somewhat complicated to select appropriate bandwidths for our estimators. We state the procedure in the following three steps:

(1) Select h_n by minimizing

$$CV_1(h_n) = \frac{1}{n}\sum_{i=1}^{n} \delta_i (Y_i - X_i^\top \widehat{\beta}_{\mathrm{C}} - g_{n,-i}^{\mathrm{C}}(T_i))^2$$

where $g_{n,-i}^{\mathrm{C}}(\cdot)$ is a "leave one out" version of $g_n^{\mathrm{C}}(\cdot)$.

(ii) Select γ_n by minimizing

$$CV_2(\gamma_n) = \frac{1}{n}\sum_{i=1}^{n} (\delta_i - \widehat{\Delta}_{t,-i}(T_i))^2,$$

where $\widehat{\Delta}_{t,-i}(\cdot)$ is a "leave one out" version of $\widehat{\Delta}_t(\cdot)$.

(iii) After obtaining h_n and γ_n, we choose b_n to minimize

$$CV_3(b_n) = \frac{1}{n}\sum_{i=1}^{n} (U_{ni} - X_i^\top \widehat{\beta}_n - g_{n,-i}(T_i))^2,$$

Table 1
Biases of $\hat{\beta}_{\text{I}}$, $\hat{\beta}_{\text{R}}$, $\hat{\beta}_{\text{IP}}$, $\hat{\beta}_{\text{C}}$ and $\hat{\beta}_{\text{full}}$ with different missing functions $\Delta(z)$ and different sample sizes

$\Delta(z)$	n	$\hat{\beta}_{\text{I}}$	$\hat{\beta}_{\text{R}}$	$\hat{\beta}_{\text{IP}}$	$\hat{\beta}_{\text{C}}$	$\hat{\beta}_{\text{full}}$
$\Delta_1(z)$	30	0.0017	0.0010	0.0007	0.0027	0.0016
	60	0.0015	0.0014	0.0023	−0.0016	0.0017
	120	−0.0005	0.0001	0.0006	−0.0009	0.0001
	200	−0.0002	−0.0004	0.0001	−0.0020	−0.0007
$\Delta_2(z)$	30	−0.0042	−0.0045	−0.0053	−0.0087	−0.0021
	60	−0.0032	−0.0039	−0.0022	−0.0049	0.0022
	120	−0.0011	−0.0013	−0.0010	−0.0021	−0.0017
	200	0.0007	0.0007	0.0007	0.0008	0.0007
$\Delta_3(z)$	30	−0.0050	−0.0053	−0.0094	−0.0074	−0.0018
	60	0.0047	0.0049	0.0058	0.0056	−0.0036
	120	−0.0028	−0.0026	−0.0033	−0.0028	0.0007
	200	−0.0012	−0.0011	−0.0015	0.0007	0.0004

Table 2
Standard errors (SE) of $\hat{\beta}_{\text{I}}$, $\hat{\beta}_{\text{R}}$, $\hat{\beta}_{\text{IP}}$, $\hat{\beta}_{\text{C}}$ and $\hat{\beta}_{\text{full}}$ with different missing functions $\Delta(z)$ and different sample sizes

$\Delta(z)$	n	$\hat{\beta}_{\text{I}}$	$\hat{\beta}_{\text{R}}$	$\hat{\beta}_{\text{IP}}$	$\hat{\beta}_{\text{C}}$	$\hat{\beta}_{\text{full}}$
$\Delta_1(z)$	30	0.2332	0.2356	0.2332	0.2689	0.2168
	60	0.1516	0.1529	0.1556	0.1681	0.1405
	120	0.1008	0.1014	0.1008	0.1064	0.0944
	200	0.0787	0.0791	0.0791	0.0819	0.0745
$\Delta_2(z)$	30	0.2802	0.2847	0.2803	0.3231	0.2156
	60	0.1765	0.1797	0.1836	0.1973	0.1414
	120	0.1144	0.1153	0.1149	0.1211	0.0963
	200	0.0875	0.0881	0.0878	0.0914	0.0747
$\Delta_3(z)$	30	0.4330	0.4385	0.4171	0.4788	0.2224
	60	0.2376	0.2410	0.2384	0.2574	0.1413
	120	0.1490	0.1508	0.1493	0.1574	0.0981
	200	0.1072	0.1082	0.1070	0.1129	0.0753

where $g_{n,-i}(\cdot)$ is a "leave one out" version of $g_n(\cdot)$, $g_n(\cdot)$ denotes one of $\widehat{g}_n^{[\text{I}]}(t)$, $\widehat{g}_n^{[\text{R}]}(t)$ and $\widehat{g}_{\text{IP}}(t)$ and U_{ni} denotes one of $U_{ni}^{[\text{I}]}$, $U_{ni}^{[\text{R}]}$ and $U_{ni}^{[\text{IP}]}$ for $i = 1, 2, \cdots, n$.

On the other hand, we should point out that the selection of bandwidths is not so critical if one is only interested in estimation of the parametric part. This can be seen from the following arguments. The fact that β is a global functional and hence the $n^{1/2}$-rate asymptotic normality of $\widehat{\beta}_{\text{I}}$, $\widehat{\beta}_{\text{R}}$ and $\widehat{\beta}_{\text{IP}}$ implies that a proper choice of the bandwidths specified in conditions (g) and (h) depends only on the second order term of the mean square errors of $\widehat{\beta}_{\text{I}}$, $\widehat{\beta}_{\text{R}}$ and $\widehat{\beta}_{\text{IP}}$.

6. Simulation

To understand the finite sample behaviors of the proposed methods, we conducted a simulation study to compare their finite sample properties.

Table 3
MSE of $\hat{\beta}_{\rm I}$, $\hat{\beta}_{\rm R}$, $\hat{\beta}_{\rm IP}$, $\hat{\beta}_{\rm C}$ and $\hat{\beta}_{\rm full}$ with different missing functions $\Delta(z)$ and different sample sizes

$\Delta(z)$	n	$\hat{\beta}_{\rm I}$	$\hat{\beta}_{\rm R}$	$\hat{\beta}_{\rm IP}$	$\hat{\beta}_{\rm C}$	$\hat{\beta}_{\rm full}$
$\Delta_1(z)$	30	0.0543	0.0554	0.0543	0.0723	0.0470
	60	0.0229	0.0234	0.0242	0.0282	0.0197
	120	0.0102	0.0103	0.0113	0.0154	0.0089
	200	0.0062	0.0063	0.0063	0.0067	0.0055
$\Delta_2(z)$	30	0.0785	0.0810	0.0856	0.1044	0.0465
	60	0.0312	0.0323	0.0337	0.0389	0.0200
	120	0.0131	0.0133	0.0132	0.0147	0.0093
	200	0.0076	0.0078	0.0077	0.0084	0.0056
$\Delta_3(z)$	30	0.1874	0.1922	0.1740	0.2292	0.0494
	60	0.0564	0.0580	0.0568	0.0662	0.0200
	120	0.0222	0.0227	0.0223	0.0248	0.0096
	200	0.0115	0.0117	0.0114	0.0128	0.0057

Table 4
Mean integrated square error (MISE) of $\hat{g}_n^{[\rm I]}(t)$, $\hat{g}_n^{[\rm R]}(t)$, $\hat{g}_n^{[\rm IP]}(t)$, $\hat{g}_n^{\rm C}(t)$ and $g_n^{\rm full}(t)$ with different missing functions $\Delta(z)$ and different sample sizes

$\Delta(z)$	n	$\hat{g}_n^{[\rm I]}(t)$	$\hat{g}_n^{[\rm R]}(t)$	$\hat{g}_n^{[\rm IP]}(t)$	$\hat{g}_n^{\rm C}(t)$	$g_n^{\rm full}(t)$
$\Delta_1(z)$	30	0.3124	0.3074	0.3138	0.5810	0.2810
	60	0.1694	0.1665	0.1810	0.3611	0.1507
	120	0.0906	0.0887	0.0909	0.2021	0.0816
	200	0.0606	0.0590	0.0609	0.1375	0.0551
$\Delta_2(z)$	30	0.3981	0.3945	0.4029	0.7089	0.2824
	60	0.2104	0.2073	0.2274	0.4354	0.1513
	120	0.1137	0.1112	0.1151	0.2531	0.0830
	200	0.0741	0.0724	0.0752	0.1706	0.0549
$\Delta_3(z)$	30	0.5753	0.5744	0.5810	0.9128	0.2853
	60	0.2862	0.2834	0.2972	0.5602	0.1490
	120	0.1555	0.1529	0.1590	0.3385	0.0836
	200	0.0982	0.0962	0.1005	0.2256	0.0550

The simulation used the model $Y = \beta^\tau X + g(T) + \varepsilon$ with X and T simulated from the normal distribution with mean 1 and variance 1 and the uniform distribution $U[0, 1]$, respectively, and ε generated from the standard normal distribution, where $\beta = 1.5$, $g(t) = (\sin(2\pi t^2))^{\frac{1}{3}}$ if $t \in [0, 1]$, $g(t) = 0$ otherwise. The kernel function $K(\cdot)$ was taken to be $K(t) = \frac{15}{16}(1 - t^2)^2$, if $|t| \leqslant 1$, 0, otherwise, $M(\cdot)$ to be $M(t) = \frac{15}{16}(1 - 2t^2 + t^4)$, if $|t| \leqslant 1$, 0, otherwise, and $\Omega(\cdot)$ to be $\Omega(t) = -\frac{15}{8}t^2 + \frac{9}{8}$, if $|t| \leqslant 1$, 0, otherwise. The bandwidths b_n, h_n and γ_n were taken to be $\frac{2}{5}n^{-7/24}$, $\frac{1}{5}n^{-1/3}$ and $\frac{4}{5}n^{-1/3}$, which satisfy the conditions (g) and (h), respectively. We did not use the bandwidth selection method suggested in Section 5 since it is time consuming for calculation and one is mainly interested in estimation of the parametric part in the partial linear model.

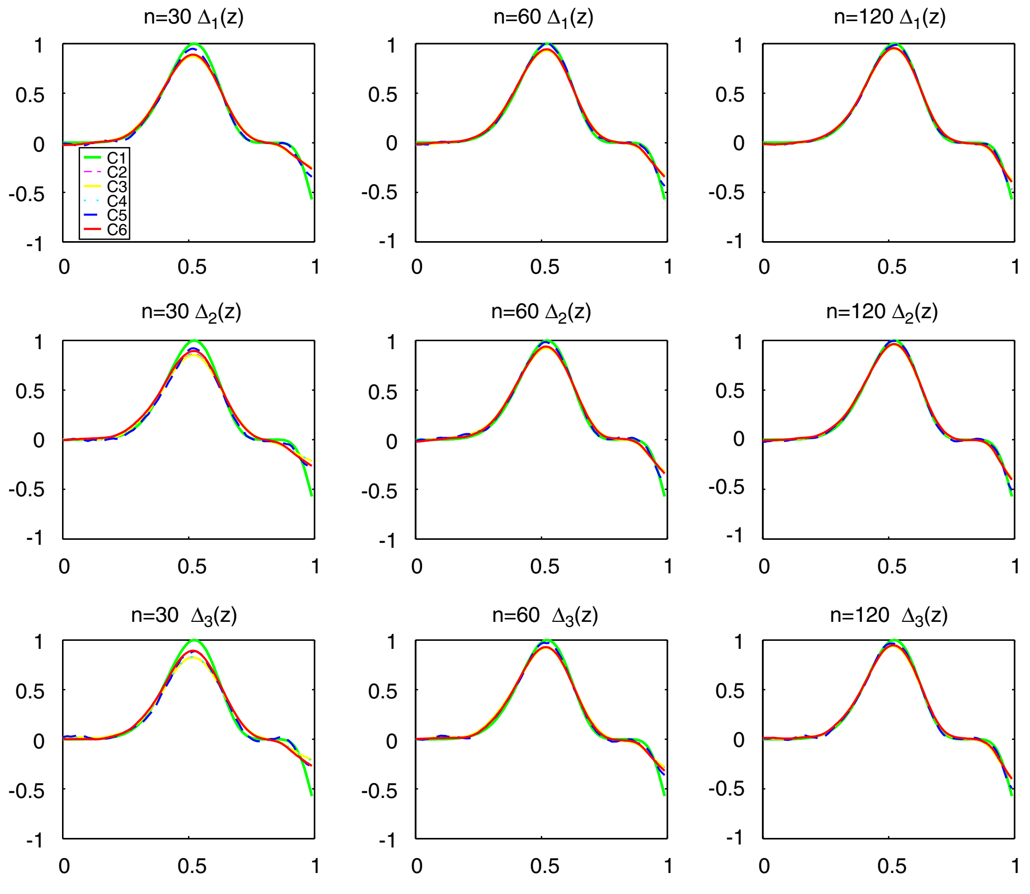

Fig. 1. Simulated curves of $\widehat{g}_n^{[I]}(t)$, $\widehat{g}_n^{[R]}(t)$, $\widehat{g}_n^{[IP]}(t)$, $g_n^{\text{full}}(t)$ and $\widehat{g}_n^{C}(t)$ with different missing functions $\Delta(z)$ and different sample sizes.

Based on the above model, we considered the following three response probability functions: $\Delta(z) = P(\delta = 1|X = x, T = t)$ under the MAR assumption.

We generated, respectively, 2000 Monte Carlo random samples of size $n = 30, 60, 120$ and 200 for the following three cases, respectively.

Case 1: $\Delta_1(z) = P(\delta = 1|X = x, T = t) = 0.8 + 0.2(|x-1| + |t-0.5|)$ if $|x-1| + |t-0.5| \leqslant 1$, and $= 0.90$ elsewhere.

Case 2: $\Delta_2(z) = P(\delta = 1|X = x, T = t) = 0.9 - 0.2(|x-1| + |t-0.5|)$ if $|x-1| + |t-0.5| \leqslant 1.5$, and $= 0.80$ elsewhere.

Case 3: $\Delta_3(z) = P(\delta = 1|X = x, T = t) = 0.8 - 0.2(|x-1| + |t-0.5|)$if $|x-1| + |t-0.5| \leqslant 1$, and $= 0.50$ elsewhere.

For the above three cases, the mean response rates are $E\Delta_1(z) \approx 0.90$, $E\Delta_2(z) \approx 0.75$ and $E\Delta_3(z) \approx 0.60$. From the 2000 simulated values of $\widehat{\beta}_{\mathrm{I}}$, $\widehat{\beta}_{\mathrm{R}}$, $\widehat{\beta}_{\mathrm{IP}}$, $\widehat{\beta}_{\mathrm{C}}$ and $\widehat{\beta}_{\text{full}}$, we calculated the biases, standard errors (SEs) and MSE of these estimators, where $\widehat{\beta}_{\mathrm{C}}$ denotes the complete case (CC) estimator which is defined by simply ignoring the missing data and $\widehat{\beta}_{\text{full}}$ denotes the standard estimator when data are observed completely. $\widehat{\beta}_{\text{full}}$ is practically unachievable, but it can serve as a gold standard. These simulated results are reported in Tables 1–3 respectively. From the 2000 simulated values of $\widehat{g}_n^{[I]}(t)$, $\widehat{g}_n^{[R]}(t)$, $\widehat{g}_n^{[IP]}(t)$, $g_n^{C}(t)$ and $g_n^{full}(\cdot)$, we calculated the mean integrated square error (MISE) and plotted the simulated curves. The result was reported in Table 4 and Fig. 1.

From Tables 1–3, all the proposed estimators of β have similar bias, SE and MSE and hence perform similarly. Generally, the bias, SE and MSE of $\widehat{\beta}_I$, $\widehat{\beta}_R$ and $\widehat{\beta}_{IP}$ are only slightly greater than $\widehat{\beta}_{full}$, the gold standard, and hence the proposed estimators of β perform well. From Tables 1–3, $\widehat{\beta}_I$, $\widehat{\beta}_R$ and $\widehat{\beta}_{IP}$ perform better than $\widehat{\beta}_C$. From Table 4, the proposed estimators $\widehat{g}_n^{[I]}(t)$, $\widehat{g}_n^{[R]}(t)$ and $\widehat{g}_n^{[IP]}(t)$ outperform $\widehat{g}_n^{C}(t)$, the CC estimator for $g(\cdot)$, in terms of MISE. It is also noted that $\widehat{g}_n^{[IP]}(t)$ has uniformly slightly larger MISE than $\widehat{g}_n^{[I]}(t)$ and $\widehat{g}_n^{[R]}(t)$, and $\widehat{\beta}_{IP}$ has more complicated variance structure and requires estimating of the marginal propensity score function $\pi(\cdot)$. Hence, one may prefer the imputation estimator and regression surrogate estimator to the inverse marginal probability weighted one.

Appendix A. Proofs of Theorems

We begin this section by listing the conditions needed in the proofs of all the theorems.

(a) (i) $E[\check{X}\check{X}^\tau]$ is a positive definite matrix.
(ii) $E[\Delta(Z)\widetilde{X}\widetilde{X}^\tau]$ is a positive definite matrix.
(b) (i) $\inf_t \Delta_t(T) > 0$.
(ii) $\Delta_t(\cdot)$ has bounded partial derivatives up to order 2.
(c) (i) $K(\cdot)$ is a bounded kernel function of order 2 with bounded support.
(ii) $M(\cdot)$ is a bounded kernel function of order 2 with bounded support.
(iii) $\Omega(\cdot)$ is a bounded kernel function of order 2 with bounded support.
(d) (i) $g_1(\cdot)$ and $g_2(\cdot)$ have bounded derivatives up to order 2.
(ii) $g_1^C(\cdot)$ and $g_2^C(\cdot)$ have bounded derivatives up to order 2.
(e) (i) $\sup_{x,t} E[Y^2|X = x, T = t] < \infty$,
(ii) $\sup_t E[\|X\|^2|T = t] < \infty$.
(f) The density of T, say $f_t(T)$, exists and has bounded derivatives up to order 2 and satisfies

$$0 < \inf_{t\in[0,1]} f_T(t) \leqslant \sup_{t\in[0,1]} f_T(t) < \infty.$$

(g) $nb_nh_n \longrightarrow \infty$; $nh_n^4 \longrightarrow 0$, $nb_n^4 \to 0$ and $\frac{h_n^2}{b_n} \to 0$.
(h) $n\gamma_n \to \infty$ and $n\gamma_n^4 \to 0$.

Remark. Condition (b)(i) is reasonable since it assumes that the response probability function is bounded from 0. Condition (f) is a commonly used assumption in the context of partially linear regression. See, e.g., [6]. Other conditions are some usual assumptions.

For the sake of convenience, we denote by c the general constant whose value may be different at each appearance.

Lemma A.1. *Under Assumptions* (a)(ii), (b)(ii), (c)(i), (d)(ii), (e) *and* (f), *if* $nh_n \to \infty$ *we have*

$$\sqrt{n}(\hat{\beta}_C - \beta) \xrightarrow{L} N(0, \Sigma_0^{-1} V_C \Sigma_0^{-1}),$$

where

$$V_{\mathrm{C}} = E[\Delta(Z)\widetilde{X}\widetilde{X}^{\tau}\sigma^2(Z)].$$

Proof. Wang et al. [21] has shown that

$$\sqrt{n}(\hat{\beta}_{\mathrm{C}} - \beta) = \frac{\Sigma_0^{-1}}{\sqrt{n}} \sum_{i=1}^{n} [X_i - g_1^{\mathrm{C}}(T_i)]\delta_i \varepsilon_i + o_p(1), \tag{A.1}$$

where $g_1^{\mathrm{C}}(t) = E[\delta X | T = t]/E[\delta | T = t]$. By central limit theorem, the lemma is then proved. □

Proof of Theorem 2.1. Let

$$\sqrt{n}(\hat{\beta}_{\mathrm{I}} - \beta) = B_n^{-1} A_n,$$

where

$$B_n = \frac{1}{n}\sum_{i=1}^{n}(X_i - g_{1n}(T_i))(X_i - g_{1n}(T_i))^{\tau}$$

and

$$A_n = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i - g_{1n}(T_i))[U_{ni}^{[\mathrm{I}]} - g_{2n}^{[\mathrm{I}]}(T_i) - (X_i - g_{1n}(T_i))^{\tau}\beta].$$

Observe that

$$\begin{aligned} B_n &= \frac{1}{n}\sum_{i=1}^{n}(X_i - g_{1n}(T_i))(X_i - g_{1n}(T_i))^{\tau} \\ &= \frac{1}{n}\sum_{i=1}^{n}(X_i - g_1(T_i))(X_i - g_1(T_i))^{\tau} + \frac{2}{n}\sum_{i=1}^{n}(X_i - g_1(T_i))(g_1(T_i) - g_{1n}(T_i))^{\tau} \\ &\quad + \frac{1}{n}\sum_{i=1}^{n}(g_1(T_i) - g_{1n}(T_i))(g_1(T_i) - g_{1n}(T_i))^{\tau} \\ &:= B_{n1} + B_{n2} + B_{n3}. \end{aligned} \tag{A.2}$$

By the law of large numbers, we have

$$B_{n1} \xrightarrow{P} \Sigma_1. \tag{A.3}$$

Let $B(s,m)$ denote the (s,m)th element of some matrix B and X_{is}, $g_{1s}(t)$, $g_{1ns}(t)$ the sth element of x_i, $g_1(t)$ and $g_{1n}(t)$, respectively, for $i = 1, 2, \ldots, n$, $s = 1, 2, \ldots, p$. For B_{n2}, we have

$$|B_{n2}(s,m)| \leqslant \sup_t |g_{1nm}(t) - g_{1m}(t)| \frac{2}{n}\sum_{i=1}^{n}|X_{is} - g_1(T_{is})| \xrightarrow{p} 0. \tag{A.4}$$

by conditions (d)(i), (c)(ii) and (e)(ii). Similarly, it can be shown that $B_{n3} \xrightarrow{p} 0$. This together with (A.2), (A.3) and (A.4) yields

$$B_n \xrightarrow{P} \Sigma_1. \tag{A.5}$$

Next we verify that

$$A_n = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}[X_i - g_1^{\mathrm{C}}(T_i)]\delta_i\varepsilon_i$$
$$+E[(1-\Delta(Z_1))(X_1 - g_1(T_1))(X_1 - g_1^{\mathrm{C}}(T_1))^{\tau}]\frac{\Sigma_0^{-1}}{\sqrt{n}}\sum_{i=1}^{n}(X_i - g_1^{\mathrm{C}}(T_i))\delta_i\varepsilon_i$$
$$+o_p(1). \tag{A.6}$$

For A_n, we have

$$A_n = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i - g_1(T_i))[\delta_i Y_i + (1-\delta_i)(X_i^{\tau}\hat{\beta}_{\mathrm{C}} + g_n^{\mathrm{C}}(T_i)) - g_{2n}^{[\mathrm{I}]}(T_i)$$
$$-(X_i - g_{1n}(T_i))^{\tau}\beta] + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(g_1(T_i) - g_{1n}(T_i))$$
$$\times[\delta_i Y_i + (1-\delta_i)(X_i^{\tau}\hat{\beta}_{\mathrm{C}} + g_n^{\mathrm{C}}(T_i)) - g_{2n}^{[\mathrm{I}]}(T_i) - (X_i - g_{1n}(T_i))^{\tau}\beta]$$
$$:= A_{n1} + A_{n2}. \tag{A.7}$$

Further, we have

$$A_{n1} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i - g_1(T_i))[\delta_i Y_i + (1-\delta_i)(X_i^{\tau}\beta + g(T_i)) - g_2(T_i) - (X_i - g_1(T_i))^{\tau}\beta]$$
$$+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i - g_1(T_i))(1-\delta_i)(X_i - g_{1n}^{\mathrm{C}}(T_i))^{\tau}(\widehat{\beta}_{\mathrm{C}} - \beta)$$
$$+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i - g_1(T_i))(1-\delta_i)(g_{n0}^{\mathrm{C}}(T_i) - g(T_i))$$
$$+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i - g_1(T_i))(g_2(T_i) - g_{2n}^{[\mathrm{I}]}(T_i))$$
$$+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i - g_1(T_i))(g_{1n}(T_i) - g_1(T_i))^{\tau}\beta$$
$$:= A_{n11} + A_{n12} + A_{n13} + A_{n14} + A_{n15}, \tag{A.8}$$

where $g_{n0}^{\mathrm{C}}(t) = g_{2n}^{\mathrm{C}}(t) - g_{1n}^{\mathrm{C}}(t)^{\top}\beta$. By the fact $g(t) = g_2(t) - g_1^{\tau}(t)\beta$, it follows that

$$A_{n11} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i - g_1(T_i))\delta_i\varepsilon_i. \tag{A.9}$$

Clearly, the law of large numbers and (A.1) can be used to get

$$A_{n12} = \left[\frac{1}{n}\sum_{i=1}^{n}(1-\delta_i)(X_i - g_1(T_i))(X_i - g_{1n}^{\mathrm{C}}(T_i))^{\tau}\right][\sqrt{n}(\hat{\beta}_{\mathrm{C}} - \beta)]$$
$$= E[(1-\Delta(Z))(X - g_1(T))(X - g_1^{\mathrm{C}}(T_i))^{\tau}]\frac{\Sigma_0^{-1}}{\sqrt{n}}\sum_{j=1}^{n}(X_j - g_1^{\mathrm{C}}(T_j))\delta_j\varepsilon_j$$
$$+o_p(1) \tag{A.10}$$

by assumptions (a)(ii), (b)(ii), (c), (d)(ii), (e) and (f). For A_{n13}, we have

$$A_{n13} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - g_1(T_i))(1-\delta_i) \frac{\sum_{j=1}^{n} \delta_j (Y_j - X_j^{\tau}\beta - g(T_i)) K\left(\frac{T_i - T_j}{h_n}\right)}{\sum_{j=1}^{n} \delta_j K\left(\frac{T_i - T_j}{h_n}\right)}$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - g_1(T_i))(1-\delta_i) \frac{\sum_{j=1}^{n} \delta_j (Y_j - X_j^{\tau}\beta - g(T_j)) K\left(\frac{T_i - T_j}{h_n}\right)}{nh_n \Delta_t(T_i) f_t(T_i)}$$

$$+ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - g_1(T_i))(1-\delta_i) \frac{\sum_{j=1}^{n} \delta_j (g(T_j) - g(T_i)) K\left(\frac{T_i - T_j}{h_n}\right)}{nh_n \Delta_t(T_i) f_t(T_i)} + o_p(1)$$

$$= A_{n131} + A_{n132} + o_p(1) \tag{A.11}$$

by (f)(ii) and (b)(ii).

By conditions (b)(ii), (c)(i), (d) and (f)(ii), we obtain

$$A_{n131} = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \delta_j \varepsilon_j \frac{1}{nh_n} \sum_{i=1}^{n} \frac{E[(X_i - g_1(T_i))(1-\delta_i)|T_i]}{\Delta_t(T_i) f_t(T_i)} K\left(\frac{T_i - T_j}{h_n}\right) + o_p(1)$$

$$= \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \delta_j \varepsilon_j \frac{E[(X_j - g_1(T_j))(1-\delta_j)|T_j]}{\Delta(T_j)} + o_p(1)$$

$$= -\frac{1}{\sqrt{n}} \sum_{j=1}^{n} \delta_j \varepsilon_j \frac{E[(X_j - g_1(T_j))\delta_j|T_j]}{\Delta(T_j)} + o_p(1). \tag{A.12}$$

Assumptions (e)(ii), (b)(i), (c)(i), (d)(i) and (f) can be used to prove that

$$\|A_{n132}\| = \frac{1}{\sqrt{n}h_n} \left\| \sum_{i=1}^{n} \frac{(X_i - g_1(T_i))(1-\delta_i)}{\Delta_t(T_i) f_t(T_i)} \frac{1}{n} \sum_{j=1}^{n} \delta_j (g(T_j) - g(T_i)) K\left(\frac{T_i - T_j}{h_n}\right) \right\|$$

$$\leqslant \frac{1}{\sqrt{n}h_n} \left\| \sum_{i=1}^{n} \frac{(X_i - g_1(T_i))(1-\delta_i)}{\Delta_t(T_i) f_t(T_i)} \int \Delta_t(t)(g(t) \right.$$

$$\left. - g(T_i)) K\left(\frac{T_i - t}{h_n}\right) f_t(t)\, \mathrm{d}t \right\| + o_p(1)$$

$$\leqslant \frac{ch_n^2}{\sqrt{n}} \sum_{i=1}^{n} \|X_i - g_1(T_i)\| + o_p(1) = o_p(1) \tag{A.13}$$

as $nh_n^4 \to 0$. By (A.11), (A.12) and (A.13), we have

$$A_{n13} = -\frac{1}{\sqrt{n}} \sum_{j=1}^{n} \delta_j \varepsilon_j \frac{E[(X_j - g_1(T_j))\delta_j|T_j]}{\Delta(T_j)} + o_p(1). \tag{A.14}$$

For A_{n14}, we have

$$
\begin{aligned}
A_{n14} &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i - g_1(T_i))\sum_{j=1}^{n}\omega_{nj}(T_i)\{g_2(T_i) - \delta_j Y_j - (1-\delta_j)(X_j^{\tau}\widehat{\beta}_{\mathrm{C}} + g_n^{\mathrm{C}}(T_j))\}\\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i - g_1(T_i))\sum_{j=1}^{n}\omega_{nj}(T_i)[g_2(T_i) - g_2(T_j)]\\
&\quad + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i - g_1(T_i))\sum_{j=1}^{n}\omega_{nj}(T_i)(g_2(T_j) - Y_j)\\
&\quad + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i - g_1(T_i))\sum_{j=1}^{n}\omega_{nj}(T_i)(1-\delta_j)(Y_j - X_j^{\tau}\beta - g(T_j))\\
&\quad + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i - g_1(T_i))\sum_{j=1}^{n}\omega_{nj}(T_i)(1-\delta_j)X_j^{\tau}(\beta - \hat{\beta}_{\mathrm{C}})\\
&\quad + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i - g_1(T_i))\sum_{j=1}^{n}\omega_{nj}(T_i)(1-\delta_j)(g_n^{\mathrm{C}}(T_j) - g(T_j))\\
&:= A_{n141} + A_{n142} + A_{n143} + A_{n144} + A_{n145}.
\end{aligned}
\tag{A.15}
$$

By arguments similar to those used in the analysis of A_{132}, we can show that $A_{n141} = o_p(1)$. Similar to (A.12), it is easy to get $A_{n142} = o_p(1)$ and $A_{n143} = o_p(1)$. By the fact that $\widehat{\beta}_{\mathrm{C}} - \beta = O_p(n^{-\frac{1}{2}})$, it is easy to verify that $A_{n144} = o_p(1)$. To obtain $A_{n14} = o_p(1)$, it remains to prove $A_{n145} = o_p(1)$. Observe that

$$
\begin{aligned}
|A_{n145}| &\leqslant \left|\frac{1}{\sqrt{n}}\sum_{j=1}^{n}(1-\delta_j)(g_n^{\mathrm{C}}(T_j) - g(T_j))\sum_{i=1}^{n}\omega_{nj}(T_i)(X_i - g_1(T_i))\right|\\
&\leqslant \sup_t |g_n^{\mathrm{C}}(t) - g(t)| \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left|\sum_{j=1}^{n}\omega_{nj}(T_i)(X_i - g_1(T_i))\right|.
\end{aligned}
\tag{A.16}
$$

By Wang and Li [20] and conditions (c)(ii), (e) and (f), we have

$$
E\left[\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left|\sum_{i=1}^{n}\omega_{nj}(T_i)(X_i - g_1(T_i))\right|\right]^2 \leqslant c\sum_{j=1}^{n}\sum_{i}^{n} E\omega_{nj}^2(T_i) = O(b_n^{-1}). \tag{A.17}
$$

This together with (A.16) and the following fact:

$$
\sup_t |g_n^{\mathrm{C}}(t) - g(t)| = O_P((nh_n)^{-\frac{1}{2}}) + O_P(h_n)
$$

yields $A_{n145} = o_p(1)$ by condition (g). This proves

$$
A_{n14} = o_p(1). \tag{A.18}
$$

Using arguments similar to that used in the proof of (A.14), we have

$$A_{n15} = o_p(1) \tag{A.19}$$

Note that $E[(X_1 - g_1(T_1))\delta_1|T_1]/\Delta(T_1) = g_1^{\mathrm{C}}(T_1)$ under MAR assumption. By combining (A.8)–(A.10), (A.14), (A.18) and (A.19), it follows that

$$\begin{aligned} A_{n1} = & \frac{1}{\sqrt{n}}\sum_{i=1}^{n}[X_i - g_1^{\mathrm{C}}(T_i)]\delta_i\varepsilon_i \\ & + E[(1-\Delta(Z_1))(X_1 - g_1(T_1))(X_1 - g_1^{\mathrm{C}}(T_1))^\tau]\frac{\Sigma_0^{-1}}{\sqrt{n}}\sum_{i=1}^{n}(X_i - g_1^{\mathrm{C}}(T_i))\delta_i\varepsilon_i \\ & + o_p(1). \end{aligned} \tag{A.20}$$

For A_{n2}, we have

$$\begin{aligned} A_{n2} = & \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(g_1(T_i) - g_{1n}(T_i))[\delta_i Y_i + (1-\delta_i)(X_i^\tau\beta + g(T_i)) - g_2(T_i) \\ & -(X_i - g_1(T_i))^\tau\beta] + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(g_1(T_i) - g_{1n}(T_i))(1-\delta_i)X_i^\tau(\hat{\beta}_{\mathrm{C}} - \beta) \\ & + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(g_1(T_i) - g_{1n}(T_i))(1-\delta_i)(g_n^{\mathrm{C}}(T_i) - g(T_i)) \\ & + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(g_1(T_i) - g_{1n}(T_i))(g_2(T_i) - g_{2n}^{[\mathrm{I}]}(T_i)) \\ & + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(g_1(T_i) - g_{1n}(T_i))(g_{1n}(T_i) - g_1(T_i))^\tau\beta \\ := & A_{n21} + A_{n22} + A_{n23} + A_{n24} + A_{n25}. \end{aligned} \tag{A.21}$$

Similarly to A_{131}, it can be shown that

$$\begin{aligned} A_{n21} &= \frac{1}{\sqrt{n}}\sum_{j=1}^{n}(X_j - g_1(T_j))E[\delta_j\varepsilon_j|T_j] + o_p(1) \\ &= o_p(1). \end{aligned} \tag{A.22}$$

For A_{n22}, we have

$$\| A_{n22} \| \leqslant \sqrt{n} \| \hat{\beta}_{\mathrm{C}} - \beta \| \sup_t \| g_1(t) - g_{1n}(t) \| \frac{1}{n}\sum_{i=1}^{n} \| X_i \| = o_p(1). \tag{A.23}$$

Hence

$$A_{n22} = o_p(1). \tag{A.24}$$

By a similar method, it can be demonstrated that

$$A_{n23} = o_p(1),\ A_{n24} = o_p(1),\ A_{n25} = o_p(1). \tag{A.25}$$

From (A.21)–(A.25), we have

$$A_{n2} = o_p(1). \tag{A.26}$$

Combining (A.7), (A.20) and (A.26), we prove (A.6). This together with central limit theorem proves Theorem 2.1 by (A.3) and Lemma A.1.

Proof of Theorem 2.2. By the definition of $\hat{g}_n(t)$, we have

$$\hat{g}_n^{[\mathrm{I}]}(t) - g(t) = g_{n2}^{[\mathrm{I}]}(t) - g_2(t) - (g_{n1}(t) - g_1(t))^\tau(\hat{\beta}_{\mathrm{I}} - \beta) - g_1(t)^\tau(\hat{\beta}_{\mathrm{I}} - \beta) \\ -(g_{n1}(t) - g_1(t))^\tau \beta. \tag{A.27}$$

First, we investigate $g_{n2}^{[\mathrm{I}]}(t) - g_2(t)$. Recalling the definition of $g_{n2}^{[\mathrm{I}]}(t)$, we have

$$\begin{aligned} g_{n2}^{[\mathrm{I}]}(t) - g_2(t) &= \sum_{i=1}^n \omega_{ni}(t) U_{ni}^{[\mathrm{I}]} - g_2(t) \\ &= \sum_{i=1}^n \omega_{ni}(t)[\delta_i Y_i + (1-\delta_i)(X_i^\tau \hat{\beta}_{\mathrm{C}} + g_n^{\mathrm{C}}(T_i)) - g_2(t)] \\ &= \sum_{i=1}^n \omega_{ni}(t)(U_i^{[\mathrm{I}]} - g_2(t)) + \sum_{i=1}^n \omega_{ni}(t)(1-\delta_i) X_i^\tau(\hat{\beta}_{\mathrm{C}} - \beta) \\ &\quad + \sum_{i=1}^n \omega_{ni}(t)(1-\delta_i)(g_n^{\mathrm{C}}(t) - g(t)). \end{aligned} \tag{A.28}$$

Note that $E[U_i^{[\mathrm{I}]}|T_i = t] = g_2(t)$ and $E[|(1-\delta_i)X_i||T_i] < \infty$. Hence, standard kernel regression theory gives

$$\sup_t \left| \sum_{i=1}^n \omega_{ni}(t)(U_i^{[\mathrm{I}]} - g_2(t)) \right| = O_P((nb_n)^{-\frac{1}{2}}) + O_P(b_n), \tag{A.29}$$

$$\sup_t |g_n^{\mathrm{C}}(t) - g(t)| = O_p((nh_n)^{-1}) + O_p(h_n), \tag{A.30}$$

$$\sup_t |g_{n1}(t) - g_1(t)| = O_P((nb_n)^{-\frac{1}{2}}) + O_P(b_n) \tag{A.31}$$

and $\sum_{i=1}^n \omega_{ni}(t)(1-\delta_i)X_i = O_P(1)$ and $\sum_{i=1}^n \omega_{ni}(t)(1-\delta_i) = O_P(1)$. This is together with (A.27) and (A.28), the facts $\widehat{\beta}_{\mathrm{C}} - \beta = O_p(n^{-\frac{1}{2}})$ and $\widehat{\beta}_{\mathrm{I}} - \beta = O_p(n^{-\frac{1}{3}})$ yields

$$\begin{aligned} \sup_t |\hat{g}_n^{[\mathrm{I}]}(t) - g(t)| &= O_P((nb_n)^{-\frac{1}{2}}) + O_P(b_n) + O_P((nh_n)^{-\frac{1}{2}}) + O_P(h_n) \\ &\quad + [O_P((nb_n)^{-\frac{1}{2}}) + O_P(b_n)] O_P(n^{-\frac{1}{2}}) + O_P(n^{-\frac{1}{2}}) \\ &\quad + O_P((nb_n)^{-\frac{1}{2}}) + O_P(b_n) \\ &= O_P((nb_n)^{-\frac{1}{2}}) + O_P(b_n) + O_P((nh_n)^{-\frac{1}{2}}) + O_P(h_n). \end{aligned} \tag{A.32}$$

Theorem 2.2 is then proved if $b_n = n^{-\frac{1}{3}}$ and $h_n = n^{-\frac{1}{3}}$. □

We can show Theorems 3.2 and 4.2 using similar arguments.

Next we prove Theorems 3.1 and 4.1.

Proof of Theorem 3.1. . Let

$$\sqrt{n}(\hat{\beta}_{\mathrm{R}} - \beta) = B_n^{-1} C_n,$$

where

$$B_n = \frac{1}{n} \sum_{i=1}^{n} (X_i - g_{1n}(T_i))(X_i - g_{1n}(T_i))^{\tau}$$

and

$$C_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - g_{1n}(T_i))[X_i^{\tau} \hat{\beta}_{\mathrm{C}} + g_n^{\mathrm{C}}(T_i) - g_{2n}^{[\mathrm{R}]}(T_i) - (X_i - g_{1n}(T_i))^{\tau} \beta].$$

It is shown in Theorem 1 that $B_n \xrightarrow{P} \Sigma_1$. Next we will demonstrate that

$$C_n = E[(X_1 - g_1(T_1))(X_1 - g_1(T_1))^{\tau}] \frac{\Sigma_0^{-1}}{\sqrt{n}} \sum_{j=1}^{n} (X_j - g_1^{\mathrm{C}}(T_j)) \delta_j \varepsilon_j + o_p(1). \tag{A.33}$$

For C_n, it is easy to get

$$\begin{aligned} C_n &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - g_1(T_i))[X_i^{\tau} \hat{\beta}_{\mathrm{C}} + g_n^{\mathrm{C}}(T_i) - g_{2n}^{[\mathrm{R}]}(T_i) - (X_i - g_{1n}(T_i))^{\tau} \beta] \\ &\quad + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (g_1(T_i) - g_{1n}(T_i))[X_i^{\tau} \hat{\beta}_{\mathrm{C}} + g_n^{\mathrm{C}}(T_i) - g_{2n}^{[\mathrm{R}]}(T_i) - (X_i - g_{1n}(T_i))^{\tau} \beta] \\ &:= C_{n1} + C_{n2}. \end{aligned} \tag{A.34}$$

Notice that $g(t) = g_2(t) - g_1(t)^{\tau} \beta$ and then we have

$$\begin{aligned} C_{n1} &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - g_1(T_i)) X_i^{\tau} (\hat{\beta}_{\mathrm{C}} - \beta) + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - g_1(T_i))(g_n^{\mathrm{C}}(T_i) - g(T_i)) \\ &\quad + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - g_1(T_i))(g_2(T_i) - g_{2n}^{[\mathrm{R}]}(T_i)) \\ &\quad + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - g_1(T_i))(g_{1n}(T_i) - g_1(T_i))^{\tau} \beta \\ &:= C_{n11} + C_{n12} + C_{n13} + C_{n14}. \end{aligned} \tag{A.35}$$

By (A.1) and the law of large numbers, it follows that

$$C_{n11} = E[(X - g_1(T)) X^{\tau}] \frac{\Sigma_0^{-1}}{\sqrt{n}} \sum_{j=1}^{n} (X_j - g_1^{\mathrm{C}}(T_j)) \delta_j \varepsilon_j + o_p(1). \tag{A.36}$$

For C_{n12}, we have

$$
\begin{aligned}
C_{n12} &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i - g_1(T_i))[g_{2n}^{\mathrm{C}}(T_i) - g_{1n}^{\mathrm{C}}(T_i)^{\tau}\hat{\beta}_{\mathrm{C}} - g_2^{\mathrm{C}}(T_i) + g_1^{\mathrm{C}}(T_i)^{\tau}\beta] \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i - g_1(T_i))(g_{2n}^{\mathrm{C}}(T_i) - g_2^{\mathrm{C}}(T_i)) \\
&\quad + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i - g_1(T_i))g_1^{\mathrm{C}}(T_i)^{\tau}(\beta - \hat{\beta}_{\mathrm{C}}) \\
&\quad + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i - g_1(T_i))(g_1^{\mathrm{C}}(T_i) - g_{1n}^{\mathrm{C}}(T_i)^{\tau}\beta \\
&\quad + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i - g_1(T_i))(g_1^{\mathrm{C}}(T_i) - g_{1n}^{\mathrm{C}}(T_i)^{\tau}(\hat{\beta}_{\mathrm{C}} - \beta) \\
&:= C_{n121} + C_{n122} + C_{n123} + C_{n124}.
\end{aligned} \tag{A.37}
$$

Using similar arguments as in the analysis of the terms A_{n12}, A_{n13}, A_{n14} and A_{n15}, it can be verified that $C_{n12i} = o_p(1), i = 1, 2, 3, 4$. Hence by (A.37), it follows that $C_{n12} = o_p(1)$. Similar to A_{n14}, we can obtain $C_{n13} = o_p(1)$. Notice that C_{n14}, is just the same as A_{n15}. By (A.19), we have $C_{n14} = o_p(1)$. This together with (A.35) and (A.36) proves

$$
\begin{aligned}
C_{n1} &= E[(X - g_1(T))X^{\tau}]\frac{\Sigma_0^{-1}}{\sqrt{n}}\sum_{j=1}^{n}(X_j - g_1^{\mathrm{C}}(T_j))\delta_j\varepsilon_j + o_p(1) \\
&= E[(X - g_1(T))(X - g_1(T))^{\tau}]\frac{\Sigma_0^{-1}}{\sqrt{n}}\sum_{j=1}^{n}(X_j - g_1^{\mathrm{C}}(T_j))\delta_j\varepsilon_j + o_p(1).
\end{aligned} \tag{A.38}
$$

For C_{n2}, similarly to the proof of $A_{n2} = o_p(1)$, it can be shown that $C_{n2} = o_p(1)$. This, together with (A.34) and (A.38), has proved (A.33). By the central limit theorem, Lemma A.1 and assumption (a), Theorem 3.1 is then proved. □

Proof of Theorem 4.1. Let

$$\sqrt{n}(\hat{\beta}_{\mathrm{IP}} - \beta) = B_n^{-1}D_n,$$

where

$$B_n = \frac{1}{n}\sum_{i=1}^{n}(X_i - g_{1n}(T_i))(X_i - g_{1n}(T_i))^{\tau}$$

and

$$D_n = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i - g_{1n}(T_i))\left[(U_{ni}^{[\mathrm{IP}]} - g_{2n}^{[\mathrm{IP}]}(T_i)) - (X_i - g_{1n}(T_i))^{\tau}\beta\right],$$

where $g_{1n}(\cdot)$ is defined in Section 2 and $g_{2n}^{[IP]}(t) = \sum_{i=1}^{n} \omega_{ni}(t) U_{ni}^{[IP]}$. Recalling that $U_{ni}^{[IP]} = \frac{\delta_i}{\hat{\Delta}_t(T_i)} Y_i + (1 - \frac{\delta_i}{\hat{\Delta}_t(T_i)})(X_i^\tau \hat{\beta}_C + g_n^C(T_i))$, by some simple computations, we have

$$\begin{aligned} D_n &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - g_{1n}(T_i)) \frac{\delta_i}{\hat{\Delta}_t(T_i)} [Y_i - (X_i^\tau \hat{\beta}_C + g_n^C(T_i))] \\ &\quad + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - g_{1n}(T_i)) [(X_i^\tau \hat{\beta}_C + g_n^C(T_i)) - g_{2n}^{[IP]}(T_i) - (X_i - g_{1n}(T_i))^\tau \beta] \\ &:= D_{n1} + D_{n2}. \end{aligned} \tag{A.39}$$

Observe

$$\begin{aligned} D_{n1} &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - g_1(T_i)) \frac{\delta_i}{\hat{\Delta}_t(T_i)} [Y_i - (X_i^\tau \hat{\beta}_C + g_n^C(T_i))] \\ &\quad + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (g_1(T_i) - g_{1n}(T_i)) \frac{\delta_i}{\hat{\Delta}_t(T_i)} [Y_i - (X_i^\tau \hat{\beta}_C + g_n^C(T_i))] \\ &:= D_{n11} + D_{n12}. \end{aligned} \tag{A.40}$$

For D_{n11}, we have

$$\begin{aligned} D_{n11} &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - g_1(T_i)) \frac{\delta_i \varepsilon_i}{\Delta_t(T_i)} \\ &\quad + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - g_1(T_i)) \left(\frac{\delta_i}{\hat{\Delta}_t(T_i)} - \frac{\delta_i}{\Delta_t(T_i)} \right) \varepsilon_i \\ &\quad - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - g_1(T_i)) \frac{\delta_i}{\Delta_t(T_i)} X_i^\tau (\hat{\beta}_C - \beta) \\ &\quad - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - g_1(T_i)) \left(\frac{\delta_i}{\hat{\Delta}_t(T_i)} - \frac{\delta_i}{\Delta_t(T_i)} \right) X_i^\tau (\hat{\beta}_C - \beta) \\ &\quad - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - g_1(T_i)) \frac{\delta_i}{\Delta_t(T_i)} (g_n^C(T_i) - g(T_i)) \\ &\quad - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - g_1(T_i)) \left(\frac{\delta_i}{\hat{\Delta}_t(T_i)} - \frac{\delta_i}{\Delta_t(T_i)} \right) (g_n^C(T_i) - g(T_i)) \\ &:= D_{n111} + D_{n112} + D_{n113} + D_{n114} + D_{n115} + D_{n116}. \end{aligned} \tag{A.41}$$

By assumption (b), (c)(iii) and (d), we have

$$
\begin{aligned}
D_{n112} &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i - g_1(T_i))\frac{\Delta_t(T_i)-\hat{\Delta}_t(T_i)}{\Delta^2(T_i)}\delta_i\varepsilon_i + o_p(1)\\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i - g_1(T_i))\delta_i\varepsilon_i\frac{\sum_{j=1}^{n}(\Delta(T_j)-\delta_j)\Omega\left(\frac{T_i-T_j}{\gamma_n}\right)}{nb_n\Delta^2(T_i)f_t(T_i)} + o_p(1)\\
&= \frac{1}{\sqrt{n}}\sum_{j=1}^{n}(\Delta(T_j)-\delta_j)\frac{1}{nb_n}\sum_{i=1}^{n}\frac{(X_i - g_1(T_i))\delta_i\varepsilon_i\Omega\left(\frac{T_i-T_j}{\gamma_n}\right)}{\Delta^2(T_i)f_t(T_i)} + o_p(1)\\
&= \frac{1}{\sqrt{n}}\sum_{j=1}^{n}(\Delta(T_j)-\delta_j)\frac{1}{nb_n}\sum_{i=1}^{n}\frac{E[(X_i - g_1(T_i))\delta_i\varepsilon_i|T_i]\Omega\left(\frac{T_i-T_j}{\gamma_n}\right)}{\Delta^2(T_i)f_t(T_i)} + o_p(1)\\
&= \frac{1}{\sqrt{n}}\sum_{j=1}^{n}(\Delta(T_j)-\delta_j)\frac{E[(X_j - g_1(T_j))\delta_j\varepsilon_j|T_j]}{\Delta^2(T_j)} + o_p(1) = o_p(1)
\end{aligned}
\tag{A.42}
$$

by noting $E[(X-g_1(T))\delta\varepsilon|T]=0$ under MAR assumption. For D_{n113}, by the law of large numbers, we have

$$
D_{n113} = -E\left[\frac{\delta_1}{\Delta(T_1)}(X_1 - g_1(T_1))X_1^{\tau}\right][\sqrt{n}(\hat{\beta}_{\mathrm{C}}-\beta)] + o_p(1). \tag{A.43}
$$

Similar to (A.23), we can verify

$$
D_{n14} = o_p(1), \quad D_{n16} = o_p(1). \tag{A.44}
$$

Observe

$$
\begin{aligned}
D_{n115} =& -\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i - g_1(T_i))\frac{\delta_i}{\Delta_t(T_i)}(g_{2n}^{\mathrm{C}}(T_i) - g_2^{\mathrm{C}}(T_i))\\
&-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i - g_1(T_i))\frac{\delta_i}{\Delta_t(T_i)}g_1^{\mathrm{C}}(T_i)^{\tau}(\beta-\hat{\beta}_{\mathrm{C}})\\
&-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i - g_1(T_i))\frac{\delta_i}{\Delta_t(T_i)}(g_1^{\mathrm{C}}(T_i) - g_{1n}^{\mathrm{C}}(T_i))^{\tau}\beta\\
&-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i - g_1(T_i))\frac{\delta_i}{\Delta_t(T_i)}(g_1^{\mathrm{C}}(T_i) - g_{1n}^{\mathrm{C}}(T_i))^{\tau}(\hat{\beta}_{\mathrm{C}}-\beta)\\
:=& D_{n1151} + D_{n1152} + D_{n1153} + D_{n1154}.
\end{aligned}
\tag{A.45}
$$

Similar to A_{n131}, we obtain

$$D_{n1151} = -\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\delta_j(Y_j - g_2^{\mathrm{C}}(T_j))\frac{E[(X_j - g_1(T_j))\delta_j/\Delta_t(T_j)|T_j]}{\Delta_t(T_j)} + o_p(1), \quad \text{(A.46)}$$

$$D_{n1152} = E\left[(X_1 - g_1(T_1))\frac{\delta_1}{\Delta_t(T_1)}g_1^{\mathrm{C}}(T_1)^{\tau}\right][\sqrt{n}(\hat{\beta}_{\mathrm{C}} - \beta)] + o_p(1), \quad \text{(A.47)}$$

$$D_{n1153} = \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\delta_j(X_j - g_1^{\mathrm{C}}(T_j))^{\tau}\beta\frac{E[(X_j - g_1(T_j))\delta_j/\Delta_t(T_j)|T_j]}{\Delta_t(T_j)} + o_p(1) \quad \text{(A.48)}$$

and

$$D_{n1154} = o_p(1). \quad \text{(A.49)}$$

By (A.45)–(A.49), it can be shown that

$$D_{n115} = -\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\frac{\delta_j\varepsilon_j}{\Delta_t(T_j)}E[(X_j - g_1(T_j))\delta_j/\Delta_t(T_j)|T_j]$$
$$+E[(X_1 - g_1(T_1))\frac{\delta_i}{\Delta_t(T_i)}g_1^{\mathrm{C}}(T_1)^{\tau}][\sqrt{n}(\hat{\beta}_{\mathrm{C}} - \beta)] + o_p(1). \quad \text{(A.50)}$$

From (A.41)–(A.44) and (A.50), we have

$$D_{n11} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i - g_1^{\mathrm{C}}(T_i))\frac{\delta_i\varepsilon_i}{\Delta_t(T_i)}$$
$$-E\left[\frac{\delta_i}{\Delta_t(T_i)}(X_1 - g_1(T_1))(X_1 - g_1^{\mathrm{C}}(T_1)^{\tau})\right][\sqrt{n}(\hat{\beta}_{\mathrm{C}} - \beta)] + o_p(1). \quad \text{(A.51)}$$

Similarly to the proof of $A_{n2} = o_p(1)$, it can be shown that $D_{n12} = o_p(1)$. This together with (A.40) and (A.51) demonstrates that

$$D_{n1} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i - g_1^{\mathrm{C}}(T_i))\frac{\delta_i\varepsilon_i}{\Delta_t(T_i)}$$
$$-E\left[\frac{\delta_i}{\Delta_t(T_i)}(X_1 - g_1(T_1))(X_1 - g_1^{\mathrm{C}}(T_1))^{\tau}\right][\sqrt{n}(\hat{\beta}_{\mathrm{C}} - \beta)] + o_p(1). \quad \text{(A.52)}$$

Recalling the definitions of $g_{2n}^{[\mathrm{R}]}(\cdot)$ and $g_{2n}^{[\mathrm{IP}]}(\cdot)$, it is direct to verify that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i - g_{1n}(T_i))(g_{2n}^{[\mathrm{R}]}(T_i) - g_{2n}^{[\mathrm{IP}]}(T_i)) = o_p(1). \quad \text{(A.53)}$$

This proves

$$D_{n2} = C_n + o_p(1) \tag{A.54}$$

$$= E[(X - g_1(T))(X - g_1(T_1))^\tau]\frac{\Sigma_0^{-1}}{\sqrt{n}}\sum_{j=1}^{n}(X_j - g_1^{\mathrm{C}}(T_j))\delta_j\varepsilon_j + o_p(1), \tag{A.55}$$

where C_n is defined in the proof of Theorem 3.1.

From (A.39), (A.52) and (A.54), we have

$$D_n = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i - g_1^{\mathrm{C}}(T_i))\frac{\delta_i\varepsilon_i}{\Delta_t(T_i)}$$
$$+E\left[\left(1 - \frac{\delta_i}{\Delta_t(T_i)}\right)(X_1 - g_1(T_1))(X_1 - g_1^{\mathrm{C}}(T_1)^\tau)\right]\frac{\Sigma_0^{-1}}{\sqrt{n}}\sum_{j=1}^{n}(X_j - g_1^{\mathrm{C}}(T_j))\delta_j\varepsilon_j$$
$$+o_p(1). \tag{A.56}$$

By the central limit theorem and Lemma A.1, Theorem 4.1 is then proved. □

Acknowledgments

The research was supported by the National Natural Science Foundation of China (Key Grant: 10231030; General Grant: 10671198) and a grant from the Research Grants Council of the Hong Kong, China (HKU 7050/06P).

References

[1] H. Ahn, J.L. Powell, Estimation of censored selection model with a nonparametric model, J. Econometrics 58 (1997) 3–30.
[2] H. Chen, Convergent rates for parametric components in partly linear model, Ann. Statist. 16 (1988) 136–146.
[3] P.E. Cheng, Nonparametric estimation of mean functionals with data missing at random, J. Amer. Statist. Assoc. 89 (1994) 81–87.
[4] R. Gray, Spline-based tests in survival analysis, Biometrics 50 (1994) 640–652.
[5] P.J. Green, B.W. Silverman, Nonparametric Regression and Generalized Linear Models: a Roughness Penalty Approach, Chapman & Hall, London, 1994.
[6] W. Härdle, H. Liang, J.T. Gao, Partially Linear Models, Physica-Verlag, Heidelberg, 2000.
[7] M.J.R. Healy, M. Westmacott, Missing values in experiments analysis on automatic computers, Appl. Statist. 5 (1956) 203–206.
[8] N. Heckman, Spline smoothing in partly linear models, J.R. Statist. Soc. Ser. B 48 (1986) 244–248.
[9] S.Y. Hong, Automatic bandwidth choice in a semiparametric regression model, Statist. Sinica 9 (1999) 775–794.
[10] Z.H. Hu, N. Wang, R.J. Carroll, Profile-kernel versus backfitting in the partially linear models for longitudinal/cluster data, Biometrika 91 (2004) 251–262.
[11] R.J.A. Little, D.B. Rubin, Statistical Analysis with Missing Data, Wiley, New York, 1987.
[12] N.S. Matloff, Use of regression functions for improved estimation of means, Biometrika 68 (1981) 685–689.
[13] J. Rice, Concergence rates for partially splined models, Statist. Probab. Lett. 4 (1986) 203–208.
[14] J.M. Robins, A. Rotnitzky, L.P. Zhao, Estimation of regression coefficients when some regressors are not always observed, J. Amer. Statist. Assoc. 89 (1994) 846–866.
[15] P.M. Robinson, Root n-consistent semiparametric regression, Econometrica 56 (1988) 931–954.
[16] R. Schmalensee, T.M. Stoker, Household gasoline demand in the United States, Econometrica 67 (1999) 645–662.
[17] P. Speckman, Kernel smoothing in partial linear models, J.R. Statist. Soc. Ser. B 50 (1988) 413–436.
[18] C.J. Stone, Optimal rates of convergence for nonparametric estimators, Ann. Statist. 8 (1980) 1348–1360.
[19] C.Y. Wang, S.J. Wang, L.P. Zhao, S.T. Ou, Weighted semiparametric estimation in regression analysis regression with missing covariates data, J. Amer. Statist. Assoc. 92 (1997) 512–525.

[20] Q.H. Wang, G. Li, Empirical likelihood semiparametric regression analysis under random censorship, J. Multivariate Anal. 83 (2002) 469–486.
[21] Q.H. Wang, O. Lindon, W. Härdle, Semiparametric regression analysis with missing response at random, J. Amer. Statist. Assoc. 99 (2004) 334–345.
[22] Q.H. Wang, J.N.K. Rao, Empirical likelihood-based inference under imputation for missing response data, Ann. Statist. 30 (2002) 345–358.
[23] S.L. Zeger, P.J. Diggle, Semiparametric models for longitudinal data with application to CD4 cell numbers in HIV seroconverters, Biometrics 50 (1994) 689–699.
[24] L.P. Zhao, S. Lipsitz, D. Lew, Regression analysis with missing covariate data using estimating equations, Biometrics 52 (1996) 1165–1182.