Ann Inst Stat Math (2009) 61:47–84
DOI 10.1007/s10463-007-0137-1

Statistical estimation in partial linear models with covariate data missing at random

Qi-Hua Wang

Received: 22 March 2006 / Revised: 15 January 2007 / Published online: 7 July 2007

Abstract In this paper, we consider the partial linear model with the covariables missing at random. A model calibration approach and a weighting approach are developed to define the estimators of the parametric and nonparametric parts in the partial linear model, respectively. It is shown that the estimators for the parametric part are asymptotically normal and the estimators of $g(\cdot)$ converge to $g(\cdot)$ with an optimal convergent rate. Also, a comparison between the proposed estimators and the complete case estimator is made. A simulation study is conducted to compare the finite sample behaviors of these estimators based on bias and standard error.

Keywords Model calibration · Weighted estimator · Asymptotic normality

1 Introduction

Suppose that $\{(X_i, T_i, Y_i), 1 \leq i \leq n\}$ is a random sample generated from the following partial linear model

$$Y_i = X_i^\top \beta + g(T_i) + \epsilon_i, \quad i = 1, 2, \ldots, n, \tag{1}$$

Q.-H. Wang (✉)
Academy of Mathematics and Systems Science, Chinese Academy of Science,
Beijing 100080, China
e-mail: qhwang@amss.ac.cn

Q.-H. Wang
Department of Statistics and Actuarial Science, The University of Hong Kong,
Pokfulam, Hong Kong, China
e-mail: qhwang@hku.hk

where Y_i's are i.i.d. scalar response variates, X_i's are i.i.d. p-variate random covariate vectors and T_i's are i.i.d. scalar covariates 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_i's are random statistical errors. It is assumed that the errors $\epsilon_i' s$ are independent with conditional mean zero given the covariates.

(1) reduces to the linear regression model when $g(\cdot) = 0$. In many practical situation, the linear model is not complex enough to capture the underlying relation between the response variables and its associate covariates. Clearly, the partially linear models contain at least the linear models as a special case. Even if the model is linear, but we specify it as partially linear model. The resulting estimators based on the partially linear model are still consistent. Hence, the partially linear model is a flexible one and allows one to focus on particular variables that are thought to have very nonlinear effects. The partial linear model is semiparametric one since it contains both parametric and nonparametric components. It allows easier interpretation of the effect of each variable and may be preferred to a completely nonparametric regression because of the well known "curse of dimensionality". The partial linear model is a natural compromise between the linear model and the fully nonparametric model. It allows only some of the predictors to be modeled linearly, with others being modeled nonparametrically. The implicit asymmetry between the effects of X and T may be attractive when X consists of dummy or categorical variables, as in Stock (1989, 1991). This specification arises in various sample selection models that are popular in econometrics, see Ahn and Powell (1993), and Newey et al. (1990).

The partial linear model was introduced by Engle et al. (1986) to study the effect of weather on electricity demand. Speckman (1988) gave an application of the partially linear model to a mouthwash experiment. Schmalensee and Stoker (1999) used the partially linear model to analyze household gasoline consumption in the United States. Green and Siverman (1994) provided an example of the use of partially linear models, and compared their results with a classical approach employing blocking. In fact, the partially linear model has also been applied to many other fields such as biometrics, See Gray (1994), and has been studied extensively for complete data setting, see Heckman (1986), Rice (1986), Speckman (1988), Chen (1988), Robinson (1988), Chen and Shiau (1991), Schick (1996), Hamilton and Truong (1997), Severini and Staniswalis (1994), Wang and Jing (1999) and Härdle et al. (2000). Various estimators for β and $g(\cdot)$ were given by using different methods such as the kernel method, the polynomial method, the penalized spline method, the piecewise constant smooth method, projection method, the smoothing splines and the trigonometric series approach. These estimators of β are proved to be asymptotically normal with zero mean and covariance $\sigma^2\Sigma^{-1}$ under different conditions, where $\Sigma = E(X_1 - E[X_1|T_1])(X_1 - E[X_1|T_1])^\top$. Recently, Wang et al. (2004) developed semiparametric regression analysis approaches with missing response data.

In practice, covariates may be missing due to various reasons. If the covariate values are collected by a questionnaire or interview, non-response is a typical source for missing values. In retrospective studies covariate values are often collected on the basis of documents like hospital records. Imcompleteness of the documents causes missing values. In clinical trials, biochemical parameters are often used as covariates. The measurement of these parameters often requires a certain amount of blood, urine or tissue, which may not be available. More examples where missing data occur can be found in Vach (1994).

In the presence of missing data, the standard inference procedures cannot be applied directly. A simple method is to naively exclude subjects with missing covariates, then perform a regression analysis with the remaining data. This method is known as complete case analysis. However, it is well known that the complete case analysis can be biased when the data are not missing completely at random (see Little and Rubin 1987) and generally gives highly inefficient estimates. Thus to increase efficiency and reduce the bias, it is important to develop methods that incorporate the partially incomplete data into the analysis.

Missing covariate data problem has been paid considerably attention. Many methods to handling missing data have been suggested under missing at random (MAR) assumption. Two recent approaches are likelihood (Ibrahim 1990; Lipsitz and Ibrahim 1996) and weighted estimating equations (Robins et al. 1994; Lipsitz and Zbrabiw 1996; Wang et al. 1998; Lipsitz et al. 1999; Liang et al. 2004). Likelihood methods assume a joint parametric distribution for covariates and response. The weighted estimating equation (WEE) approach doesn't require any distribution assumptions. Estimation is based on the "complete-cases", that is, those with no missing data, with weighting inversely proportional to the probability that the covariates are observed.

In this paper, we suggest a model calibration method and a new weighted method to develop estimation theory in model (1) when the covariates X may be missing. That is, we develop estimation approaches for β and $g(\cdot)$ based on the following incomplete observations

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

from model (1), where the $Y_i's$ and $T_i's$ are observed completely and $\delta_i = 0$ if X_i is missing, otherwise $\delta_i = 1$.

Throughout this paper, we assume that X are missing at random (MAR). The MAR assumption implies that δ and X are conditionally independent given T and Y. That is, $P(\delta = 1|Y, X, T) = P(\delta = 1|Y, T)$. MAR is a common assumption for statistical analysis with missing data and is reasonable in many practical situations, see Little and Rubin (1987, Chap. 1). We define two estimators of β and $g(\cdot)$ by two different approaches. One of the approaches is first to calibrate model (1), and then use the kernel method and least square method. Another approach is to combine the inverse probability weighted estimating approach with kernel method. We compare the two approaches by simulation in terms of bias and standard error of the estimators. Asymptotic results for the two estimators of β and $g(\cdot)$ are derived, showing that the two proposed estimators of β are strongly consistent and weakly consistent with an optimal convergent rate.

This paper is organized as follows. We define the two estimators of β and $g(\cdot)$ and give the asymptotic properties in Sects. 2 and 3. In Sect. 4, we make a comparison between the proposed estimators and the complete case estimator. In Sect. 5, we compare these estimators by simulation. The proof of the main results are presented in the Appendix.

2 Model calibration based estimators and asymptotic properties

Let $\Delta(y,t) = P(\delta = 1|Y = y, T = t)$. Note that $E[\delta_i X_i/\triangle(Y_i, T_i)|X_i, Y_i, T_i] = X_i$ for $i = 1, 2, \ldots, n$ under MAR. This motivates me to define synthetic data $U_i = \delta_i X_i/\triangle(Y_i, T_i)$ such that (Y_i, U_i, T_i) follow a standard partial linear model for $i = 1, \ldots, n$. Then, the estimation approach for the standard partial linear model can be applied to estimation of β and $g(\cdot)$ if $\triangle(\cdot,\cdot)$ is a known function. Under MAR assumption, we have

$$\begin{aligned} &E\left[Y_i - U_i^\top \beta - g(T_i)|X_i, T_i\right] \\ &\quad = E\left\{E\left[Y_i - U_i^\top \beta - g(T_i)|X_i, Y_i, T_i\right]|X_i, T_i\right\} = 0, \end{aligned} \tag{2}$$

where $U_i = \frac{\delta_i X_i}{\triangle(Y_i, T_i)}$. This implies that the incomplete observed data follow the following models

$$Y_i = U_i^\top \beta + g(T_i) + e_i, \tag{3}$$

where e_i are i.i.d. random variables with conditional mean zero given covariables (X_i, T_i) for $i = 1, 2, \ldots, n$. By Speckman (1988), model (3) is equivalent to

$$Y_i - E[Y_i|T_i] = (U_i - E[U_i|T_i])^\top \beta + e_i. \tag{4}$$

Let $g_1(t) = E[X|T = t]$ and $g_2(t) = E[Y|T = t]$. Then, $g_1(T) = E[U|T = t]$. If $\triangle(\cdot,\cdot)$, $g_1(\cdot)$ and $g_2(\cdot)$ were known functions, the least square approach could be applied to (4) to define the least square estimate (LSE) of β to be

$$\widetilde{\beta}_{MC} = B_n^{-1} A_n,$$

where

$$B_n = \frac{1}{n}\sum_{i=1}^{n}[(U_i - g_1(T_i))(U_i - g_1(T_i))^\top]$$

and

$$A_n = \frac{1}{n}\sum_{i=1}^{n}(U_i - g_1(T_i))(Y_i - g_2(T_i)).$$

In practice, however, $\triangle(\cdot,\cdot)$, $g_1(\cdot)$ and $g_2(\cdot)$ are unknown. Naturally, one can define estimator of β to be $\widetilde{\beta}_{MC}$ with $\triangle(\cdot,\cdot)$, $g_1(\cdot)$ and $g_2(\cdot)$ in it replaced by their estimators. Let $K(\cdot)$ be a bivariate kernel function and h_n a bandwidth sequence tending to zero as $n \to \infty$. For simplicity, let $Z_i = (Y_i, T_i)$ for $i = 1, 2, \ldots, n$. Then, $\triangle(z)$ can be estimated by

$$\triangle_n(z) = \frac{\sum_{i=1}^n \delta_i K\left(\frac{z-Z_i}{h_n}\right)}{\sum_{i=1}^n K\left(\frac{z-Z_i}{h_n}\right)}.$$

Let $\omega(\cdot)$ be a kernel function and b_n a bandwidth sequence tending to zero as $n \to \infty$. Define the weights

$$W_{nj}(t) = \frac{\omega\left(\frac{t-T_j}{b_n}\right)}{\sum_{j=1}^n \omega\left(\frac{t-T_j}{b_n}\right)}.$$

Then $g_1(t)$ and $g_2(t)$ can be estimated consistently by $\widehat{g}_{1,n}(t) = \sum_{j=1}^n W_{nj}(t)\frac{\delta_j X_j}{\triangle_n(Z_j)}$ and $\widehat{g}_{2,n}(t) = \sum_{j=1}^n W_{nj}(t)Y_j$. Let $U_{in} = \frac{\delta_i X_i}{\triangle_n(Z_i)}$ for $i = 1, 2, \ldots, n$. We then can define the estimator of β, say $\widehat{\beta}_{MC}$, to be $\widetilde{\beta}_{MC}$ with $\triangle(\cdot, \cdot)$, $g_1(t)$ and $g_2(t)$ replaced by $\triangle_n(\cdot, \cdot)$, $\widehat{g}_{1,n}(t)$ and $\widehat{g}_{2,n}(t)$, respectively. That is

$$\widehat{\beta}_{MC} = \widehat{B}_n^{-1}\widehat{A}_n,$$

where $\widehat{A}_n$ and $\widehat{B}_n$ are A_n and B_n, respectively, with U_i, $g_1(\cdot)$ and $g_2(\cdot)$ replaced by U_{in}, $\widehat{g}_{1,n}(\cdot)$ and $\widehat{g}_{2,n}(\cdot)$.

Taking conditional expectation given T in (3), under MAR we have

$$g(t) = g_2(t) - g_1^\top(t)\beta. \tag{5}$$

This suggests that $g(\cdot)$ can be estimated by

$$\widehat{g}_{MC}(t) = \widehat{g}_{2,n}(t) - \widehat{g}_{1,n}^\top(t)\widehat{\beta}_{MC}.$$

Theorem 1 *Under the assumptions listed in Appendix A, we have*

$$\sqrt{n}(\widehat{\beta}_{MC} - \beta) \xrightarrow{\mathcal{L}} N(\mu_{MC}, V_{MC})$$

where $\mu_{MC} = -\Sigma^{-1}E\left[\frac{1-\triangle(Z)}{\triangle(Z)}XX^\top\right]\beta$ *and* $V_{MC} = \Sigma_{MC}^{-1}\Omega_{MC}\Sigma_{MC}^{-1}$ *with*

$$\begin{aligned}\Omega_{MC} = {} & E[(U - E[U|T])(U - E[U|T])^\top(Y - U^\top\beta - g(T))^2] \\ & -E\left[\frac{1-\triangle(Z)}{\triangle(Z)}XX^\top\right]\beta\beta^\top E\left[\frac{1-\triangle(Z)}{\triangle(Z)}XX^\top\right] \\ & +\Sigma_0\beta\beta^\top\Sigma_0 E\frac{1-\triangle(Z)}{\triangle(Z)} \\ & +2\left\{\left(E\left[\left(\frac{X}{\triangle(Z)} - E[X|T]\right)\left(Y - \frac{X^\top\beta}{\triangle(Z)} - g(T)\right)\right]\right.\right.\end{aligned}$$

$$+E\left[\frac{1-\triangle(Z)}{\triangle(Z)}XX^\top\right]\beta\Bigg)\beta^\top\Sigma_0\Bigg\},$$
$$\Sigma_{MC}=E[(U-E[U|T])(U-E[U|T])^\top]$$

and

$$\Sigma_0=E[(X-E[X|T])(X-E[X|T])^\top].$$

Remark 1 $\sqrt{n}(\widehat{\beta}_{MC}-\beta)$ is asymptotically biased. Clearly, the bias μ_{MC} closes to 0 when the response probability $\triangle(z)$ closes to 1. When $\triangle(z)=1$, the asymptotic bias is zero.

If $\widehat{\beta}^*$ is a consistent estimator of β, we can define a bias-corrected estimator as follows:

$$\widehat{\beta}^*_{MC}=\widehat{\beta}_{MC}-\frac{1}{\sqrt{n}}\widehat{\mu}_{MC}$$

where $\widehat{\mu}_{MC}=-\frac{\widehat{\Sigma}^{-1}_{MC}}{n}\sum_{i=1}^n\frac{1-\Delta_n(Z_i)}{\Delta_n(Z_i)}X_iX_i^\top\widehat{\beta}^*$ with

$$\widehat{\Sigma}_{MC}=\frac{1}{n}\sum_{i=1}^n(U_{in}-\widehat{g}_{1,n}(T_i))(U_{in}-\widehat{g}_{1,n}(T_i))^\top.$$

Theorem 2 *Under assumptions of Theorem 1, if $\widehat{\beta}^*$ is an $n^{\frac{1}{2}}$-consistent estimator of β we have*

$$\sqrt{n}(\widehat{\beta}^*_{MC}-\beta)\xrightarrow{\mathcal{L}}N(0,V_{MC}).$$

An example is to take $\widehat{\beta}^*$ in $\widehat{\beta}^*_{MC}$ to be $\widehat{\beta}_{MC}$ since $\widehat{\beta}_{MC}$ is a consistent estimator of β from Theorem 1. In Sect. 3, we will define a weighted estimator and then construct another bias-corrected estimator by taking $\widehat{\beta}^*_n$ in $\widehat{\beta}^*_{MC}$ to be the weighted one.

The asymptotic variance V_{MC} reduces to $V=\Sigma_0^{-1}\sigma^2$ if $\triangle(z)=1$. This is just the asymptotic variance when the data are observed completely. The asymptotic variance can be estimated consistently by combining the "plug in" method with sample moment method. An alternative is to use the jackknife method to define the asymptotic variance estimator.

Let $\widehat{\beta}^{(-i)}_{MC}$ be $\widehat{\beta}_{MC}$ based on $\{Y_j,X_j,\delta_j,T_j\}_{j\neq i}$ for $i=1,2,\dots,n$. Let J_{ni} be the jackknife pseudo-values. That is

$$J_{ni}=n\widehat{\beta}_{MC}-(n-1)\widehat{\beta}^{(-i)}_{MC},\quad i=1,2,\dots,n$$

Then, the jackknife variance estimator can be defined by

$$\widehat{V}_{MC,J}=\frac{1}{n}\sum_{i=1}^n(J_{ni}-\bar{J}_n)^2,$$

where $\bar{J}_n = n^{-1}\sum_{i=1}^n J_{ni}$. Under the conditions of Theorem 1, the variance estimator can be proved to be consistent.

Theorem 3 *Under conditions of Theorem 2, if $b_n = O(n^{-\frac{1}{3}})$ and $h_n = O(n^{-\frac{1}{6}})$ we have*

$$\widehat{g}^*_{MC}(t) - g(t) = O_p(n^{-\frac{1}{3}}),$$

where $\widehat{g}^*_{MC}(t)$ is $\widehat{g}_{MC}(t)$ with $\widehat{\beta}_{MC}$ replaced by $\widehat{\beta}^*_{MC}$.

3 Weighted estimators and asymptotic properties

In this section, we define an $n^{\frac{1}{2}}$-rate asymptotically normal estimator of β, say $\widehat{\beta}_W$, and a consistent estimator of $g(\cdot)$, say $\widehat{g}_W(t)$, with an optimal convergent rate by weighting approach. Based on $\widehat{\beta}_W$, we then define a model calibration bias-corrected estimator $\widehat{\beta}^*_{MC}$ with $n^{\frac{1}{2}}$-rate asymptotic normality.

Under MAR assumption, we have

$$\begin{aligned}
&\beta = E^{-1}[(X - E[X|T])(X - E[X|T])^\top]E[(X - E[X|T])(Y - E[Y|T])],\\
&E[(X - E[X|T])(X - E[X|T])^\top] = E\left[\frac{\delta}{\triangle(Z)}(X - E[X|T])(X - E[X|T])^\top\right],\\
&E[(X - E[X|T])(Y - E[Y|T])] = E\left[\frac{\delta}{\triangle(Z)}(X - E[X|T])(Y - E[Y|T])\right]
\end{aligned}$$

and

$$E[X|T] = E\left[\frac{\delta X}{\triangle(Z)}|T\right].$$

Combining the "plug in" method with sample moment method, β can be estimated by

$$\widehat{\beta}_W = \widetilde{B}_n^{-1}\widetilde{A}_n, \tag{6}$$

where

$$\widetilde{A}_n = \frac{1}{n}\sum_{i=1}^n \frac{\delta_i(X_i - \widehat{g}_{1,n}(T_i))(Y_i - \widehat{g}_{2,n}(T_i))}{\triangle_n(Z_i)}$$

and

$$\widetilde{B}_n = \frac{1}{n}\sum_{i=1}^n \frac{\delta_i(X_i - \widehat{g}_{1,n}(T_i))(X_i - \widehat{g}_{1,n}(T_i))}{\triangle_n(Z_i)}.$$

$g(\cdot)$ can then be estimated by

$$\widehat{g}_W(t) = \widehat{g}_{2,n}(t) - \widehat{g}_{1,n}(t)^\top \widehat{\beta}_W.$$

from (4).

For simplicity of stating the following theorem, let $M(z) = E[(X - E[X|T])(Y - X^\top\beta - g(T))|Z = z]$.

Theorem 4 *Under the assumptions of Theorem 1 and $(C.M)$ in Appendix B, we have*

$$\sqrt{n}(\widehat{\beta}_W - \beta) \xrightarrow{\mathcal{L}} N(0, V_W)$$

where $V_W = \Sigma_0^{-1}\Omega_W\Sigma_0^{-1}$ *with*

$$\Omega_W = E\left[\frac{(X - E[X|T])(X - E[X|T])^\top(Y - X^\top\beta - g(T))^2}{\triangle(Z)}\right] - E\left[\frac{M(Z)M^\top(Z)(1 - \triangle(Z))}{\triangle(Z)}\right]$$

It is noted that $\widehat{\beta}_W$ has the same asymptotic variance as $\widehat{\beta}_{\text{all}}$ defined by (5) in Liang et al. (2004). But $\widehat{\beta}_{\text{all}}$ is complicated for calculation.

Let $\widehat{\beta}_{MW}$ be $\widehat{\beta}^*_{MC}$ with $\widehat{\beta}^*$ taken to be $\widehat{\beta}_W$. By Theorem 4, $\widehat{\beta}_W$ is an $n^{\frac{1}{2}}$-consistent estimator of β. This together with Theorem 2 implies that $\widehat{\beta}_{MW}$ is asymptotically normal with mean 0 and variance V_{MC}. It is noted that the weighted estimation approach described above mainly uses the information contained in the complete case and use the addition information from $\{(Y_i, T_i) : \delta_i = 0\}$ by $\triangle_n(\cdot)$ only. A natural question is: Does $\widehat{\beta}_{MW}$ improve $\widehat{\beta}_W$ since $\widehat{\beta}_{MC}$ and hence $\widehat{\beta}_{MW}$ use more additional information from $\{(Y_i, T_i) : \delta_i = 0\}$? Unfortunately, it seems difficult to compare $\widehat{\beta}_W$ with $\widehat{\beta}_{MW}$ in terms of their asymptotic variances, V_W and V_{MC}. A simulation comparison will be made below for their finite sample properties. The simulation results show that $\widehat{\beta}_{MW}$ has less standard error.

Remark 2 Another alternative is to use the weighted estimation equation suggested by Robins et al. (1994). However, it can be shown that the estimator based on the extended estimation equation, say $\widehat{\beta}_{WEE}$, has the same asymptotic variance as $\widehat{\beta}_W$. Comparing to $\widehat{\beta}_{WEE}$, however, some obvious advantages of $\widehat{\beta}_W$ are that it is explicitly defined, easy to compute and does not require any iteration scheme.

It should be noted that the idea of developing the weighted approach here is similar to that of Horvitz and Thompson (1952), Robins et al. (1994) and Wang et al. (1998). However, it is innovative to develop the model calibration approach and combine it with the weighted method.

Remark 3 The asymptotic variance of $\widehat{\beta}_W$ can be estimated by "plug in" method and sample moment method. Also, jackknife method can be used to estimate the asymptotic variance as in Sect. 2. If $\triangle(z) = 1$, the asymptotic variance reduces to that of the estimator when data are observed completely.

Theorem 5 *Under assumptions of Theorem 4, if* $b_n = O(n^{-\frac{1}{3}})$ *and* $h_n = O(n^{-\frac{1}{6}})$ *we have*

$$\widehat{g}_W(t) - g(t) = O_p(n^{-\frac{1}{3}}).$$

Remark 4 The convergent rate for $\widehat{g}_W(t)$ is the same as the optimal convergent rate obtained by Stone (1980) in the case where data are observed completely. Let $\widehat{g}_{MW}(t)$ be $\widehat{g}^*_{MC}(t)$ with $\widehat{\beta}^*$ taken to be $\widehat{\beta}_W$. Then, $\widehat{g}_{MW}(t)$ has the same optimal convergent rate as $\widehat{g}_W(t)$.

Remark 5 Hong (1999) maked a discussion for the bandwidth selection problem in the partial linear model with complete observations. In the presence of missing data, the automatic bandwidth choice developed by Hong (1999) is applicable here by weighting. On the other hand, $\widehat{\beta}_W$ and $\widehat{\beta}_{MW}$ are global functionals and hence the $n^{\frac{1}{2}}$-rate asymptotic normality of the two estimators indicates that a proper choice of h_n and b_n specified in conditions (C.$h_n b_n$) and (C.h_n) depends only on the second order terms of the mean square error of the estimators. This implies that the selection of the bandwidth may not be so critical for estimating β.

4 Comparisons with complete case estimation approach

As pointed out before, one simple way to avoid the problem of missing data is to analyze only those subjects who are completely observed. This method is known as complete case analysis. In what follows, we use a complete case analysis to define the estimators.

Let $\omega_c(\cdot)$ be a kernel function and γ_n a bandwidth sequence tending to zero as $n \to \infty$, and define the weights

$$W_{nj,c}(t) = \frac{\omega_c\left(\frac{t-T_j}{\gamma_n}\right)}{\sum_{j=1}^n \delta_j \omega_c\left(\frac{t-T_j}{\gamma_n}\right)}.$$

Let $g^*_{1n}(t) = \sum_{j=1}^n W_{nj,c}(t)\delta_j X_j$ and $g^*_{2n}(t) = \sum_{j=1}^n W_{nj,c}(t)\delta_j Y_j$. The estimator of β is then defined by

$$\widehat{\beta}_C = \left[\sum_{i=1}^n \delta_i\{(X_i - g^*_{1n}(T_i))(X_i - g^*_{1n}(T_i))^\top\}\right]^{-1} \times \sum_{i=1}^n \delta_i\{(X_i - g^*_{1n}(T_i))(Y_i - g^*_{2n}(T_i))\}$$

based on the observed triples (X_i, T_i, Y_i) for $i \in \{i : \delta_i = 1\}$.

Under some mild conditions, $\widehat{\beta}_C$ can be proved to be asymptotically normal with mean

$$\mu_C = \Sigma_C^{-1} E[\triangle(Z)(X - g_1^*(T))(Y - g_2^*(T) - (X - g_1^*(T))^\top \beta)]$$

and variance $V_C = \Sigma_C^{-1}\Omega_C\Sigma_C^{-1}$, where $g_1^*(t) = \frac{E[\delta X|T=t]}{E[\delta|T=t]}$, $g_2^*(t) = \frac{E[Y|T=t]}{E[\delta|T=t]}$ and

$$\Omega_C = E[\triangle(Z)(X - g_1^*(T))(X - g_1^*(T))^\top (Y - g_2^*(T) - (X - g_1^*(T))^\top \beta)^2]$$

and

$$\Sigma_C = E[\triangle(Z)(X - g_1^*(T))(X - g_1^*(T))^\top].$$

Clearly, the asymptotic variance of $\widehat{\beta}_W$ is less than that of $\widehat{\beta}_C$ when $\triangle(z)$ is a constant. When $\triangle(z)$ is not a constant, it is hard to prove that $\widehat{\beta}_W$ has less asymptotic variance than $\widehat{\beta}_C$ although it is believed to be true. In this case, however, $\widehat{\beta}_C$ is an asymptotically biased estimator, and hence $\widehat{\beta}_C$ is not of practical interest. This implies that it might not so critical to compare its asymptotic variance with that of $\widehat{\beta}_W$ in such a case.

It is seen that the complete case analysis defines an asymptotic biased estimator with larger asymptotic variance than $\widehat{\beta}_W$ at least for the case where $\triangle(z)$ is a constant. When $\triangle(z)$ is a constant or $\triangle(z) = P(\delta = 1|Y, T) = P(\delta = 1|T)$, the asymptotic bias μ_C is zero.

5 Some simulation results

To illustrate the use of the proposed estimators and to compare their finite sample performance with the complete case analysis approach, we carried out a number of simulation study to calculate the bias and standard error of the estimators.

For each of n i.i.d. observations, a normally distributed covariate, $X \sim N(1, 1)$, a uniform distributed covariate, $T \sim U[0, 1]$, and a response variate Y from the partially linear model $Y = X\beta + 3.5(\exp(-(4T - 1).^2) + \exp(-(4T - 3).^2)) - 1.5 + \epsilon$ with $\beta = 1$ and $\epsilon \sim N(0, 1)$. The kernel function $K(\cdot)$ was taken to be $K(z) = K_1(y)K_2(t)$ with $K_1(y) = \frac{15}{16}(1 - y^2)^2$ if $|y| \le 1$, 0 otherwise and $K_2(t) = -\frac{15}{8}t^2 + \frac{9}{8}$ if $|t| \le 1$, 0 otherwise. h_n was taken to be $n^{-\frac{1}{5}}$. $\omega(u) = \omega_c(u) = -\frac{15}{8}u^2 + \frac{9}{8}$ if $|u| \le 1$, 0 otherwise. b_n and γ_n were taken to be $\frac{3}{2}n^{-\frac{1}{3}}$.

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

Case 1: $\Delta_1(y, t) = P(\delta = 1|Y = y, T = t) = \frac{1}{1+\exp\{-\ln(9)-0.1(y-\frac{16}{15})-0.2(t-0.5)\}}$;
Case 2: $\Delta_2(y, t) = P(\delta = 1|Y = y, T = t) = \frac{1}{1+\exp\{-\ln(3)-0.2(y-\frac{16}{15})-0.1(t-0.5)\}}$;
Case 3: $\Delta_3(y, t) = P(\delta = 1|Y = y, T = t) = \frac{1}{1+\exp\{-\ln(3/2)-0.2(y-\frac{16}{15})-0.2(t-0.5)\}}$.

Table 1 Biases of $\widehat{\beta}_W$, $\widehat{\beta}_{MC}$, $\widehat{\beta}_{MW}$ and $\widehat{\beta}_C$ under different missing functions $\Delta(x)$ and different sample sizes n

$P(x)$	n	$\widehat{\beta}_W$	$\widehat{\beta}_{MW}$	$\widehat{\beta}_{MC}$	$\widehat{\beta}_C$
	30	−0.0096	0.0147	−0.1625	−0.0145
$\Delta_1(x)$	60	−0.0026	0.0123	−0.1558	−0.0114
	120	0.0008	0.0107	−0.1502	0.0123
	200	−0.0005	0.0061	−0.1523	−0.0135
	30	−0.0128	0.0467	−0.3124	−0.0197
$\Delta_2(x)$	60	−0.0054	0.0284	0.3100	−0.0234
	120	−0.0025	0.0160	−0.3087	−0.0167
	200	−0.0038	0.0092	0.3096	−0.0179
	30	−0.0171	0.0760	−0.4757	−0.0269
$\Delta_3(x)$	60	−0.0068	−0.0459	−0.4731	−0.0180
	120	−0.0023	0.0296	−0.4694	−0.0181
	200	−0.0077	0.0151	−0.4696	−0.0185

Table 2 Standard errors of $\widehat{\beta}_W$, $\widehat{\beta}_{MC}$, $\widehat{\beta}_{MW}$ and $\widehat{\beta}_C$ under different missing functions $\Delta(x)$ and different sample sizes n

$P(x)$	n	$\widehat{\beta}_W$	$\widehat{\beta}_{MW}$	$\widehat{\beta}_{MC}$	$\widehat{\beta}_C$
	30	0.2297	0.2124	0.1918	0.2432
$\Delta_1(x)$	60	0.1587	0.1515	0.1281	0.1625
	120	0.1016	0.0983	0.0896	0.1074
	200	0.0761	0.0738	0.0686	0.0800
	30	0.2413	0.2230	0.2072	0.2685
$\Delta_2(x)$	60	0.1592	0.1453	0.1412	0.1706
	120	0.1079	0.0992	0.0946	0.1165
	200	0.0845	0.0753	0.0702	0.0888
	30	0.2724	0.2282	0.2264	0.3221
$\Delta_3(x)$	60	0.1816	0.1544	0.1509	0.2026
	120	0.1217	0.1024	0.0976	0.1334
	200	0.0924	0.0826	0.0738	0.1002

The average missing rates for the above three cases are approximately 0.10, 0.25 and 0.40 respectively. From the 2000 simulated values of $\widehat{\beta}_W$, $\widehat{\beta}_{MC}$, $\widehat{\beta}_{MW}$ and $\widehat{\beta}_C$, we calculated the biases, standard errors and jackknife variance estimators of the four estimators. These simulation results are reported in Tables 1, 2 and 3.

From Tables1, 2 and 3, we have the following observations:

(1) Biases of both $\widehat{\beta}_W$ and $\widehat{\beta}_{MW}$ decrease and are close to zero as sample size increases for every fixed missing rate. It seems that biases of $\widehat{\beta}_C$ and $\widehat{\beta}_{MC}$ do not decrease and are not close to zero as sample size increases. This verifies the theoretical results obtained in Sect. 2 and Sect. 4 and implies that it may not be of practical interest to consider $\widehat{\beta}_C$ and $\widehat{\beta}_{MC}$ in practice. Next, we compare $\widehat{\beta}_W$ and $\widehat{\beta}_{MW}$ only.

(2) $\widehat{\beta}_W$ has less bias than $\widehat{\beta}_{MW}$. But, it has larger standard error and Jackknife variance estimate.

Table 3 Jackknife variance estimators of $\widehat{\beta}_W$, $\widehat{\beta}_{MC}$, $\widehat{\beta}_{MW}$ and $\widehat{\beta}_C$ under different missing functions $\Delta(x)$ and different sample sizes n

$P(x)$	n	$\widehat{\beta}_W$	$\widehat{\beta}_{MW}$	$\widehat{\beta}_{MC}$	$\widehat{\beta}_C$
	30	0.0523	0.0446	0.0364	0.0585
$\Delta_1(x)$	60	0.0248	0.0223	0.0160	0.0256
	120	0.0101	0.0092	0.0075	0.0109
	200	0.0054	0.0050	0.0042	0.0078
	30	0.0576	0.0492	0.0425	0.0714
$\Delta_2(x)$	60	0.0245	0.0205	0.0193	0.0282
	120	0.0112	0.0093	0.0085	0.0129
	200	0.0066	0.0053	0.0046	0.0073
	30	0.0717	0.0516	0.0508	0.1027
$\Delta_3(x)$	60	0.0320	0.0231	0.0211	0.0398
	120	0.0141	0.0100	0.0090	0.0169
	200	0.0079	0.0063	0.0050	0.0093

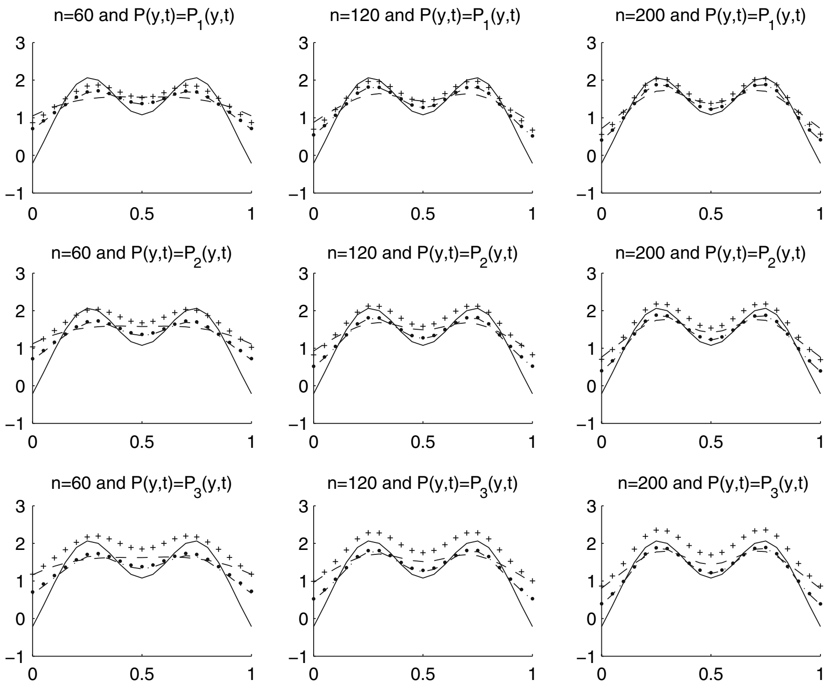

Fig. 1 Curve for $\widehat{g}_W(t)$, $\widehat{g}_{MC}$, $\widehat{g}_{MW}$, $\widehat{g}_C(t)$ and the true curve $g(t)$ under missing rate $\Delta_1(x)$ for sample size of n = 60, 120 and 200. *Dotted curve* is $\widehat{g}_W(t)$, plus curve is $\widehat{g}_{MC}(t)$, *dash-dotted curve* is $\widehat{g}_{MW}(t)$, *dashed curve* is $\widehat{g}_C(t)$ and *solid curve* is the true curve $g(t)$

Also, we calculated the simulated curves of $\widehat{g}_W(t)$, $\widehat{g}_{MC}(t)$, $\widehat{g}_{MW}(t)$ and $\widehat{g}_C(t)$ from the 2000 simulated values of them under the three different missing rates for sample size $n = 60$, 120 and 200 and compared them to the true curve. See Fig. 1.

From Fig. 1, we see that the curves $\widehat{g}_W(t)$, $\widehat{g}_{MW}(t)$ and $\widehat{g}_{MC}(t)$ capture the pattern of the true curve $g(t)$, and the curves of $\widehat{g}_W(t)$ and $\widehat{g}_{MW}(t)$ are closer to the true curve. $\widehat{g}_{MC}(t)$ depends missing rate heavily. $\widehat{g}_{MC}(t)$ is close to the true curve as the missing rate decreases. It seems that $\widehat{g}_C(t)$ performs poorly.

Acknowledgments The author thanks Peter Hall and two referees for their constructive comments and suggestions that led to significant improvements. 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).

Appendix A: Proofs of Theorems 1 and 2

(C.g): $g(\cdot)$, $g_{1r}(\cdot)$ and $g_2(\cdot)$ satisfy Lipschitz condition of order 1, i.e., $|g(t_1) - g(t_2)| \le c|t_1 - t_2|$ for some constant $c > 0$.

(C.ω): There exist constants $M_1 > 0$, $M_2 > 0$ and $\rho > 0$ such that

$$M_1 I[|u| \le \rho] \le \omega(u) \le M_2 I[|u| \le \rho].$$

(C.f). The density of Z, say $f(z)$, has bounded partial derivatives up to order $k(> 2)$ almost surely.

(C.ϵ): $\sup_{x,t} E[\epsilon^4 | X = x, T = t] < \infty$.

(C.X): $\sup_t E[\|X\|^4 | T = t] < \infty$, where $\|\cdot\|$ defines the Euclidean distance.

(C.T): The density of T, say $r(t)$, exists and satisfies

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

(C.b_n): $nb_n^2 \to \infty$ and $nb_n^4 \to 0$.

(C.$\triangle$): (i) $\triangle(z)$ has bounded partial derivatives up to order $k(> 2)$ almost surely.
(ii) $\inf_z \triangle(z) > 0$.

(C.K): (i) The kernel function K is a bounded kernel function with bounded support.
(ii) $K(\cdot)$ is a kernel of order $k(> 2)$.

(C.h_n): $nh_n^4 \to \infty$, $nh_nb_n^2 \to \infty$ and $nh_n^{2k+1} \to 0$ for $k > 2$.

(C.h_nb_n): $\frac{b_n}{h_n} \longrightarrow 0$ and $h_n^k / b_n \to 0$.

Remark 6 Conditions (C.g), (C.ω), (C.f) and (C.T) are standard conditions, which are commonly used in literature. See, e.g., Härdle et al. (2000), Speckman (1988) and Heckman (1986). Condition (C.$\triangle$) is used in literature on missing data analysis. See, e.g., Lipsitz et al. (1999) and Qi et al. (2005). Condition (C.K) is used in the investigation on some nonparametric kernel estimators. See, e.g., Prakasa Rao (1981) and Qi et al. (2005). An example for conditions (C.b_n), (C.h_n) and (C.h_nb_n) to be satisfied are $h_n = n^{-\frac{1}{6}}$ and $b_n = n^{-\frac{1}{3}}$.

Proof of Theorem 1. Clearly

$$\widehat{\beta}_{MC}-\beta=\widehat{B}_n^{-1}\widehat{C}_n, \tag{7}$$

where

$$\widehat{C}_n=\frac{1}{n}\sum_{i=1}^{n}\left[(U_{i,n}-\widehat{g}_{1,n}(T_i))(Y_i-\widehat{g}_{2,n}(T_i)-(U_{i,n}-\widehat{g}_{1,n}(T_i))^{\top}\beta\right],$$

with $U_{i,n}=\frac{\delta_i X_i}{\triangle_n(Z_i)}$ for $i=1,2,\ldots,n$. Next, we prove

$$\begin{aligned}\sqrt{n}\widehat{C}_n&=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(U_i-g_1(T_i))[Y_i-g_2(T_i)-(U_i-g_1(T_i))^{\top}\beta]\\&\quad+(\Sigma-E\left[\frac{1-\triangle(Z)}{\triangle(Z)}XX^{\top}\right]\beta)\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\frac{\delta_j-\triangle(Z_j)}{\triangle(Z_j)}+o_p(1).\end{aligned} \tag{8}$$

and

$$\widehat{B}_n-B_n\xrightarrow{p}0\quad\text{and}\quad B_n\xrightarrow{p}\Sigma. \tag{9}$$

Let $\widetilde{g}_{1,n}(t)=\sum_{j=1}^{n}W_{nj}(t)U_j$. Then

$$\sqrt{n}\widehat{C}_n=M_n+R_n+S_n+T_n, \tag{10}$$

where

$$\begin{aligned}M_n&=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(U_i-g_1(T_i))[Y_i-g_2(T_i)-(U_i-g_1(T_i))^{\top}\beta],\\R_n&=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(g_1(T_i)-\widetilde{g}_{1,n}(T_i))[Y_i-g_2(T_i)-(U_i-g_1(T_i)^{\top}\beta)],\\S_n&=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(U_i-\widetilde{g}_{1,n}(T_i))\{(g_2(T_i)-\widehat{g}_{2,n}(T_i))\\&\quad-[(U_{i,n}-\widehat{g}_{1,n}(T_i))^{\top}\beta-(U_i-g_1(T_i))^{\top}\beta]\end{aligned}$$

and

$$\begin{aligned}T_n=&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}[(U_{i,n}-\widehat{g}_{1,n}(T_i))-(U_i-\widetilde{g}_{1,n}(T_i))]\\&\times[Y_i-\widehat{g}_{2,n}(T_i)-(U_{i,n}-\widehat{g}_{1,n}(T_i))^{\top}\beta].\end{aligned}$$

It is clear

$$
\begin{aligned}
R_n &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sum_{j=1}^{n} W_{nj}(T_i)(g_1(T_i) - g_1(T_j))(Y_i - U_i^\top \beta - g(T_i)) \\
&\quad - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sum_{j=1}^{n} W_{nj}(T_i)(U_j - E[U_j|T_j])(Y_i - U_i^\top \beta - g(T_i)) \\
&:= R_{n1} + R_{n2}.
\end{aligned} \tag{11}
$$

Let $\zeta_{ni,s} = \sum_{j=1}^{n} W_{nj}(T_i)(g_{1s}(T_i) - g_{1s}(T_j))(Y_i - U_i^\top \beta - g(T_i))$, $s = 1, 2, \ldots, p$. Let $R_{n1,s}$ be the rth component of R_{n1}. From (2), it follows that

$$
E[\zeta_{nk,s}\zeta_{nl,s}] = E\{E[\zeta_{nk,s}\zeta_{nl,s}|T_1, T_2, \ldots, T_n, X_k, X_l]\} = 0. \tag{12}
$$

for $k \neq l$. This together with (C.g), (C.ω), (C.T) and ($C.\epsilon$) proves

$$
\begin{aligned}
ER_{n1,s}^2 &= \frac{1}{n} \sum_{i=1}^{n} E\zeta_{ni,s}^2 \\
&\leq \sum_{i=1}^{n} \sum_{j=1}^{n} E\{W_{nj}^2(T_i)(g_{1s}(T_i) - g_{1s}(T_j))^2 E[e_i^2|T_i]\} \\
&\leq cb_n^2 \sum_{i=1}^{n} \sum_{j=1}^{n} E\{W_{nj}^2(T_i) \left(\frac{T_i - T_j}{b_n}\right)^2 E[e_i^2|T_i]\} \\
&\leq cb_n^2 \sum_{i=1}^{n} \sum_{j=1}^{n} EW_{nj}^2(T_i) \leq c(n^2 b_n^2)(n^2 b_n)^{-1} \longrightarrow 0, \quad s = 1, 2, \ldots, p.
\end{aligned} \tag{13}
$$

by using C_r-inequality in the first inequality and using Lemma A.1 of Wang (1999) in the last inequality, where c is some constant. Next, c may be different constant in different place.

For R_{n2}, we have

$$
\begin{aligned}
R_{n2} &= -\frac{1}{\sqrt{n}} \sum_{i \neq j} W_{nj}(T_i)(U_j - E[U_j|T_j])(Y_i - U_i^\top \beta - g(T_i)) \\
&\quad - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} W_{ni}(T_i)(U_i - E[U_i|T_i])(Y_i - U_i^\top \beta - g(T_i)) \\
&:= R_{n2,1} + R_{n2,2}.
\end{aligned} \tag{14}
$$

By $(C.\epsilon)$, (C.X), (C.$\triangle$), (C.ω) and (C.T), we have

$$\begin{aligned}
E\|R_{n2,1}\|^2 &= \frac{1}{n}\sum_{i\neq j} E\|W_{nj}(T_i)(U_j - E[U_j|T_j])(Y_i - X_i^\top\beta - g(T_i))\|^2 \\
&= \frac{1}{n}\sum_{i\neq j} E\{W_{nj}^2(T_i)E[\|U_j \\
&\quad - E[U_j|T_j]\|^2|T_j]E[(Y_i - X_i^\top\beta - g(T_i))^2|T_i]\} \\
&\leq \frac{C}{n}\sum_{i\neq j} EW_{nj}^2(T_i) \leq c(nb_n)^{-1}. \qquad (15)
\end{aligned}$$

By (C.X) and (C.$\triangle$) and $(C.\epsilon)$, similar to (15) we have

$$\begin{aligned}
ER_{n2,2}^2 &= \sum_{i=1}^n E[W_{ni}^2(T_i)\|U_i - E[U_i|T_i]\|^2(Y_i - U_i^\top\beta - g(T_i))^2] \\
&\leq \sum_{i=1}^n E\{W_{ni}^2(T_i)\sup_t E^{\frac{1}{2}}[\|U_i\|^4|T=t]\sup_t E^{\frac{1}{2}}[Y^4|T=t]\} \\
&\leq C(nb_n)^{-1} \longrightarrow 0. \qquad (16)
\end{aligned}$$

By (14), (15) and (16), it follows that $R_{n2} = o_p(1)$. This together with (11) and (13) proves

$$R_n = o_p(1). \qquad (17)$$

For S_n, we have

$$S_n = S_{n,1} + S_{n,2} + S_{n,3}, \qquad (18)$$

where

$$\begin{aligned}
S_{n,1} &= \frac{1}{\sqrt{n}}\sum_{i=1}^n (U_i - E[U_i|T_i])\{(g_2(T_i) - \widehat{g}_{2,n}(T_i)) \\
&\quad - [(U_{i,n} - \widehat{g}_{1,n}(T_i))^\top\beta - (U_i - g_1(T_i))^\top\beta]\}, \\
S_{n,2} &= \frac{1}{\sqrt{n}}\sum_{i=1}^n\sum_{j=1}^n W_{nj}(T_i)(g_1(T_i) - g_1(T_j))\{g_2(T_i) - \widehat{g}_{2,n}(T_i) \\
&\quad - [(U_{i,n} - \widehat{g}_{1,n}(T_i))^\top\beta - (U_i - g_1(T_i))^\top\beta]\}
\end{aligned}$$

and

$$S_{n,3} = -\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sum_{j=1}^{n} W_{nj}(T_i)(U_j - E[U_j|T_j])\{g_2(T_i) - \widehat{g}_{2,n}(T_i) \\ -[(U_{i,n} - \widehat{g}_{1,n}(T_i))^\top \beta - (U_i - g_1(T_i))^\top \beta]\}.$$

Clearly

$$\begin{aligned} S_{n,1} &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (U_i - E[U_i|T_i]) \sum_{j=1}^{n} W_{nj}(T_i)(g_2(T_i) - g_2(T_j)) \\ &\quad + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (U_i - E[U_i|T_i]) \sum_{j=1}^{n} W_{nj}(T_i)(g_2(T_j) - Y_j) \\ &\quad - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (U_i - E[U_i|T_i])(U_{i,n} - U_i)^\top \beta \\ &\quad + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (U_i - E[U_i|T_i])(\widehat{g}_{1,n}(T_i) - g_1(T_i))^\top \beta \\ &:= S_{n1,1} + S_{n1,2} + S_{n1,3} + S_{n1,4}. \end{aligned} \tag{19}$$

Note that $S_{n1,1}$ is R_{n1} with $Y_i - X_i^\top \beta - g(T_i)$ and $g_1(\cdot)$ replaced by $U_i - E[U_i|T_i]$ and $g_2(\cdot)$ replaced, respectively, and hence has expression similar to R_{n1}. This implies that arguments similar to (13) can be used to prove

$$S_{n1,1} = o_p(1). \tag{20}$$

Clearly

$$\begin{aligned} S_{n1,2} &= \frac{1}{\sqrt{n}} \sum_{i \neq j} W_{nj}(T_i)(U_i - E[U_i|T_i])(g_2(T_j) - Y_j) \\ &\quad + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} W_{ni}(T_i)(U_i - E[U_i|T_i])(g_2(T_i) - Y_i) \\ &:= S_{n1,2}^{[1]} + S_{n1,2}^{[2]}. \end{aligned}$$

Again note that

$$\begin{aligned} &E[W_{nj}(T_i)(U_i - E[U_i|T_i])(g_2(T_j) - Y_j)] \\ &\quad = E\{E[W_{nj}(T_i)(U_i - E[U_i|T_i])(g_2(T_j) - Y_j)|T_i, T_j]\} = 0 \end{aligned}$$

for $i \neq j$, and $S_{n1,2}^{[1]}$ has expression similar to $R_{n2,1}$. Hence, similar arguments used in the proof of (15) can be used to prove $S_{n1,2}^{[1]} = o_p(1)$. By (4) and (16), we have

$$\begin{aligned} S_{n1,2}^{[2]} &= -\frac{1}{\sqrt{n}}\sum_{i=1}^{n} W_{ni}(T_i)(U_i - E[U_i|T_i])(U_i - E[U_i|T_i])^\top \beta + R_{n,22} \\ &= -\frac{1}{\sqrt{n}}\sum_{i=1}^{n} W_{ni}(T_i)(U_i - E[U_i|T_i])(U_i - E[U_i|T_i])^\top \beta + o_p(1), \end{aligned}$$

where $R_{n,22}$ is defined in (14). Hence

$$S_{n1,2} = -\frac{1}{\sqrt{n}}\sum_{i=1}^{n} W_{ni}(T_i)(U_i - E[U_i|T_i])(U_i - E[U_i|T_i])^\top \beta + o_p(1). \tag{21}$$

Recalling the definition of $\triangle_n(\cdot,\cdot)$, we have

$$\begin{aligned} S_{n1,3} \;=\; & \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(U_i - E[U_i|T_i])\frac{\frac{1}{nh_n^2}\sum_{j=1}^{n}(\delta_j - \triangle(Z_j))K\left(\frac{Z_i-Z_j}{h_n}\right)}{\triangle^2(Z_i)f_{\mathcal{Z}}(Z_i)}\delta_i X_i^\top \beta \\ & +\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(U_i - E[U_i|T_i])\frac{\frac{1}{nh_n^2}\sum_{j=1}^{n}(\triangle(Z_j) - \triangle(Z_i))K\left(\frac{Z_i-Z_j}{h_n}\right)}{\triangle^2(Z_i)f_{\mathcal{Z}}(Z_i)} \\ & \times \delta_i X_i^\top \beta + o_p(1) \\ := \; & S_{n1,3}^{[1]} + S_{n1,3}^{[2]} + o_p(1) \end{aligned} \tag{22}$$

Note that $\frac{\delta_i X_i}{\triangle(Z_i)} = U_i$ for $i = 1, 2, \ldots, n$, we get

$$\begin{aligned} S_{n1,3}^{[1]} \;=\; & \Sigma\beta\frac{1}{n^{\frac{3}{2}}h_n^2}\sum_{j=1}^{n}(\delta_j - \triangle(Z_j))\sum_{i=1}^{n}\frac{K\left(\frac{Z_i-Z_j}{h_n}\right)}{\triangle(Z_i)f_{\mathcal{Z}}(Z_i)} \\ & +\frac{1}{n^{\frac{3}{2}}h_n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\{[(U_i - E[U_i|T_i])(U_i - E[U_i|T_i])^\top \beta] \\ & -E[(U - E[U|T])(U - E[U|T])^\top \beta]\} \\ & \times(\delta_j - \triangle(Z_j))\frac{K\left(\frac{Z_i-Z_j}{h_n}\right)}{\triangle(Z_i)f_{\mathcal{Z}}(Z_i)} \\ & +\frac{1}{n^{\frac{3}{2}}h_n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}[(U_i - E[U_i|T_i])E^\top[U_i|T_i]\beta](\delta_j - \triangle(Z_j))\frac{K\left(\frac{Z_i-Z_j}{h_n}\right)}{\triangle(Z_i)f_{\mathcal{Z}}(Z_i)} \\ := \; & I_{n1} + I_{n2} + I_{n3}. \end{aligned} \tag{23}$$

By (C.K), (C.$\triangle$), (C.f), (C.X) and (C.h_n), using arguments similar to Wang and Rao (2002) it can be shown

$$I_{n1} = \Sigma\beta \frac{1}{\sqrt{n}h_n} \sum_{j=1}^{n} (\delta_j - \triangle(Z_j)) \int \frac{K\left(\frac{z-Z_j}{h_n}\right)}{\triangle(z)} \mathrm{d}z + o_p(1)$$
$$= \Sigma\beta \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \frac{\delta_j - \triangle(Z_j)}{\triangle(Z_j)} + o_p(1), \tag{24}$$

$I_{n2} = o_p(1)$, $I_{n3} = o_p(1)$ and $S_{n1,3}^{[2]} = o_p(1)$. This together with (22)–(24) proves that

$$S_{n1,3} = \Sigma\beta \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \frac{\delta_j - \triangle(Z_j)}{\triangle(Z_j)} + o_p(1). \tag{25}$$

For $S_{n1,4}$, we have

$$S_{n1,4} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (U_i - E[U_i|T_i]) \left[\sum_{j=1}^{n} W_{nj}(T_i) \left(\frac{\delta_j X_j}{\triangle_n(Z_j)} - \frac{\delta_j X_j}{\triangle(Z_j)} \right)^{\top} \beta \right]$$
$$+ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (U_i - E[U_i|T_i]) \sum_{j=1}^{n} W_{nj}(T_i) \left(\frac{\delta_j X_j}{\triangle(Z_j)} - g_1(T_j) \right)^{\top} \beta$$
$$+ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (U_i - E[U_i|T_i]) \sum_{j=1}^{n} W_{nj}(T_i) (g_1(T_j) - g_1(T_i))^{\top} \beta$$
$$:= S_{n1,4}^{[1]} + S_{n1,4}^{[2]} + S_{n1,4}^{[3]}. \tag{26}$$

It can be proved

$$S_{n1,4}^{[1]} = -\frac{1}{\sqrt{n}} \sum_{j=1}^{n} \left(\sum_{i=1}^{n} W_{nj}(T_i)(U_i - E[U_i|T_i]) \right)$$
$$\times \frac{U_j^{\top}\beta}{\triangle(Z_j)} (\triangle_n(Z_j) - \triangle(Z_j)) + o_p(1). \tag{27}$$

Clearly, the main term in (27) can be bounded by

$$\sup_z |\triangle_n(z) - \triangle(z)| Q_n, \tag{28}$$

where

$$Q_n = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \left\| \sum_{i=1}^{n} W_{nj}(T_i)(U_i - E[U_i|T_i]) \frac{U_j^{\top}\beta}{\triangle(Z_j)} \right\|. \tag{29}$$

By (C.$\triangle$) and (C.X), we have

$$
\begin{aligned}
EQ_n^2 &\le \sum_{j=1}^n E\|\frac{U_j^\top \beta}{\triangle(Z_j)}\sum_{i=1}^n W_{nj}(T_i)(U_i - E[U_i|T_i])\|^2 \\
&\le C\sum_{j=1}^n \left[\sum_{i=1,i\neq j}^n EW_{nj}^2(T_i)(U_i - E[U_i|T_i])^2\right] \\
&\quad + C\sum_{j=1}^n E\left[W_{nj}^2(T_j)\|(U_j - E[U_j|T_j])\frac{U_j^\top\beta}{\triangle(Z_j)}\|^2\right] \\
&\le Cb_n^{-1} + C(nb_n)^{-1}.
\end{aligned} \tag{30}
$$

This proves $Q_n = O(b_n^{-\frac{1}{2}})$. Hence, by (27), (28) and the fact

$$
\sup_z |\triangle_n(z) - \triangle(z)| = O_p((nh_n^2)^{-\frac{1}{2}}) + O_p(h_n^k), \tag{31}
$$

we have

$$
S_{n1,4}^{[1]} = o_p(1). \tag{32}
$$

as $nh_n^2 b_n \to \infty$ and $h_n^{2k}/b_n \to 0$, which are implied by (C.$h_n b_n$).

It is easy too see that

$$
\begin{aligned}
S_{n1,4}^{[2]} &= \frac{1}{\sqrt{n}}\sum_{i\neq j} W_{nj}(T_i)(U_i - E[U_i|T_i])(U_j - E[U_j|T_j])^\top\beta \\
&\quad + \frac{1}{\sqrt{n}}\sum_{i=1}^n W_{ni}(T_i)(U_i - E[U_i|T_i])(U_i - E[U_i|T_i])^\top\beta
\end{aligned} \tag{33}
$$

Using similar arguments to (15), it can be proved that the first term at the left hand side of the above equality is $o_p(1)$. Similar to (20), we have $S_{n1,4}^{[3]} = o_p(1)$. This together with (26), (32), (33) and (21) proves

$$
S_{n1,4} + S_{n1,2} = o_p(1). \tag{34}
$$

By (19), (20), (21), (25) and (34), it follows

$$
S_{n,1} = \Sigma\beta\frac{1}{\sqrt{n}}\sum_{j=1}^n \frac{\delta_j - \triangle(Z_j)}{\triangle(Z_j)} + o_p(1). \tag{35}
$$

For $S_{n,2}$, we have

$$\begin{aligned}
S_{n,2} &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{n}W_{nj}(T_i)(g_1(T_i)-g_1(T_j))\sum_{j=1}^{n}W_{nj}(T_i)(g_2(T_i)-g_2(T_j))\\
&\quad+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{n}W_{nj}(T_i)(g_1(T_i)-g_1(T_j))\sum_{j=1}^{n}W_{nj}(T_i)(g_2(T_j)-Y_j)\\
&\quad-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{n}W_{nj}(T_i)(g_1(T_i)-g_1(T_j))(U_{in}-U_i)^\top\beta\\
&\quad+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}W_{nj}(T_i)W_{nk}(T_i)(g_1(T_i)-g_1(T_j))(U_{kn}-U_k)^\top\beta\\
&\quad+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{n}W_{nj}(T_i)(g_1(T_i)-g_1(T_j))\sum_{k=1}^{n}W_{nk}(T_i)(U_k-g_1(T_k))^\top\beta\\
&\quad+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{n}W_{nj}(T_i)(g_1(T_i)-g_1(T_j))\sum_{k=1}^{n}W_{nk}(T_i)(g_1(T_k)-g_1(T_i))^\top\beta\\
&:= S_{n,21}+S_{n,22}+S_{n,23}+S_{n,24}+S_{n,25}+S_{n,26}. \qquad (36)
\end{aligned}$$

By (C.g), (C.ω) and (C.b_n), we get

$$\begin{aligned}
|S_{n,21}| &\le \frac{Cb_n^2}{\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{n}W_{nj}(T_i)\left|\frac{T_i-T_j}{b_n}\right|\sum_{j=1}^{n}W_{nj}(T_i)\frac{|T_i-T_j|}{b_n}\\
&\le C\sqrt{n}b_n^2 \longrightarrow 0. \qquad (37)
\end{aligned}$$

It is noted that $S_{n,26}$ is $S_{n,21}$ with $g_2(\cdot)$ replaced by $g_1^\top(\cdot)\beta$. Hence, similar arguments to (37) can be used to prove

$$S_{n,26} \le C\sqrt{n}b_n^2 \longrightarrow 0. \qquad (38)$$

By (C.g) and (C.ω), we have

$$\begin{aligned}
|S_{n,22}| &\le \frac{Cb_n}{\sqrt{n}}\left(\sum_{i=1}^{n}\sum_{j=1}^{n}W_{nj}(T_i)\left|\frac{T_i-T_j}{b_n}\right|\right)\left|\sum_{j=1}^{n}W_{nj}(T_i)(g_2(T_j)-Y_j)\right|\\
&\le \frac{Cb_n}{\sqrt{n}}\sum_{i=1}^{n}\left|\sum_{j=1}^{n}W_{nj}(T_i)(g_2(T_j)-Y_j)\right|.
\end{aligned}$$

Hence, by (C.ω) and $(C.\epsilon)$ we get

$$
\begin{aligned}
E|S_{n,22}|^2 &\leq \frac{Cb_n^2}{n} n \sum_{i=1}^{n} E\left[\sum_{j=1}^{n} W_{nj}(T_i)(g_2(T_j)-Y_j)\right]^2 \\
&\leq Cb_n^2 \sum_{i=1}^{n}\sum_{j=1}^{n} EW_{nj}^2(T_i)(g_2(T_j)-Y_j)^2 \leq Cb_n \longrightarrow 0.
\end{aligned}
$$

This proves

$$
S_{n,22} \xrightarrow{p} 0. \tag{39}
$$

Similarly, we have

$$
S_{n,25} \xrightarrow{p} 0. \tag{40}
$$

It is noted that

$$
S_{n,23} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n}\sum_{j=1}^{n} (g_1(T_i)-g_1(T_j)) \frac{\delta_i X_i^\top \beta}{\triangle^2(Z_i)} (\triangle_n(Z_i)-\triangle(Z_i)) + o_p(1). \tag{41}
$$

By (C.g), (C.$\triangle$), (C.ω), (C.$h_n b_n$) and (31), it follows that the main term in (41) is bounded by

$$
\begin{aligned}
&b_n \sup_z |\triangle_n(z)-\triangle(z)| \frac{1}{\sqrt{n}} \sum_{i=1}^{n}\sum_{j=1}^{n} W_{nj}(T_i) \left|\frac{T_i-T_j}{b_n}\right| |X_i^\top \beta| \\
&= O_p(\frac{b_n}{h_n}) + O_p(\sqrt{n} b_n h_n) = o_p(1).
\end{aligned} \tag{42}
$$

(41) and (42) together prove $S_{n,23} = o_p(1)$. Similarly, we have $S_{n,24} = o_p(1)$. This together with (36), (37), (38), (39) and (40) proves

$$
S_{n2} = o_p(1). \tag{43}
$$

For $S_{n,3}$, we have

$$
\begin{aligned}
S_{n,3} &= -\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{n} W_{nj}(T_i)(U_j - E[U_j|T_j])(g_2(T_i) - \widehat{g}_{2,n}(T_i)) \\
&\quad + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{n} W_{nj}(T_i)(U_j - E[U_j|T_j])(U_{i,n} - U_i) \\
&\quad - \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{n} W_{nj}(T_i)(U_j - E[U_j|T_j])(\widehat{g}_{1,n}(T_i) - g_1(T_i))^{\top}\beta \\
&:= S_{n,31} + S_{n,32} + S_{n,33}.
\end{aligned} \tag{44}
$$

Observe that

$$
\begin{aligned}
S_{n,31} &= -\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{n} W_{nj}(T_i)(U_j - E[U_j|T_j])\sum_{j=1}^{n} W_{nj}(T_i)(g_2(T_i) - g_2(T_j)) \\
&\quad + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{n} W_{nj}(T_i)(U_j - E[U_j|T_j])\sum_{j=1}^{n} W_{nj}(T_i)(Y_j - g_2(T_j)) \\
&:= S_{n,31}^{[1]} + S_{n,31}^{[2]}.
\end{aligned} \tag{45}
$$

By (C.g), (C.ω), (C.X) and (C.$\triangle$), we have

$$
\left|\sum_{j=1}^{n} W_{nj}(T_i)(g_2(T_i) - g_2(T_j))\right| \le Cb_n \sum_{j=1}^{n} W_{nj}(T_i)\left|\frac{T_i - T_j}{b_n}\right| \le Cb_n \tag{46}
$$

and

$$
\begin{aligned}
&E[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{n} W_{nj}(T_i)(U_j - E[U_j|T_j])]^2 \\
&\le C\sum_{i=1}^{n}\sum_{j=1}^{n} E\{W_{nj}^2(T_i)E[U_j^2|T_j]\} \le Cb_n^{-1}.
\end{aligned}
$$

This proves

$$
\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{n} W_{nj}(T_i)(U_j - E[U_j|T_j]) = O_p(b_n^{-\frac{1}{2}}). \tag{47}
$$

(46) and (47) together prove

$$
S_{n,31}^{[1]} = O_p(b_n^{\frac{1}{2}}). \tag{48}
$$

By Schwartz inequality, we have

$$\begin{aligned}
E\|S_{n,31}^{[2]}\| &\le \frac{1}{\sqrt{n}}\sum_{i=1}^{n} E^{\frac{1}{2}}\left\|\sum_{j=1}^{n} W_{nj}(T_i)(U_j - E[U_j|T_j]\right\|^2 \\
&\qquad \times E^{\frac{1}{2}}\left(\sum_{j=1}^{n} W_{nj}(T_i)(Y_j - E[Y_j|T_j]\right)^2 \\
&\le \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\sum_{j=1}^{n} E[W_{nj}^2(T_i)E[\|U_j\|^2|T_j]]\right)^{\frac{1}{2}} \\
&\qquad \times\left(\sum_{j=1}^{n} E[W_{nj}^2(T_i)E[\|U_j\|^2|T_j]\right)^{\frac{1}{2}} \\
&\le Cn^{-\frac{1}{2}}b_n^{-1} \longrightarrow 0 \qquad (49)
\end{aligned}$$

by (C.b_n). (45), (48) and (49) together prove

$$S_{n,31} \xrightarrow{p} 0. \qquad (50)$$

Similar to (32), we can prove

$$S_{n,32} \xrightarrow{p} 0. \qquad (51)$$

For $S_{n,33}$, we have

$$\begin{aligned}
S_{n,33} &= -\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{n} W_{nj}(T_i)(U_j - E[U_j|T_j]) \\
&\qquad \times\left[\sum_{j=1}^{n} W_{nj}(T_i)\left(\frac{\delta_j X_j}{\triangle_n(Z_j)} - \frac{\delta_j X_j}{\triangle(Z_j)}\right)^{\top}\beta\right] \\
&\quad -\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{n} W_{nj}(T_i)(U_j - E[U_j|T_j]) \\
&\qquad \times\left[\sum_{j=1}^{n} W_{nj}(T_i)\left(\frac{\delta_j X_j}{\triangle(Z_j)} - g_1(T_j)\right)^{\top}\beta\right] \\
&\quad -\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{n} W_{nj}(T_i)(U_j - E[U_j|T_j])(g_1(T_j) - g_1(T_i))^{\top}\beta \\
&:= S_{n,33}^{[1]} + S_{n,33}^{[2]} + S_{n,33}^{[3]} \qquad (52)
\end{aligned}$$

It is noted that

$$\begin{aligned}
\|S_{n,33}^{[1]}\| &\leq \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sum_{j=1}^{n} W_{nj}(T_i)(U_j - E[U_j|T_j]) \right. \\
&\quad \left. \times \left(\sum_{j=1}^{n} W_{nj}(T_i)\delta_j X_j^{\top} \beta \frac{\triangle(Z_j) - \triangle_n(Z_j)}{\triangle^2(Z_j)} \right) \right\| + o_p(1) \\
&\leq \frac{C}{\sqrt{n}} \left(\sum_{i=1}^{n} \left\| \sum_{j=1}^{n} W_{nj}(T_i)(U_j - E[U_j|T_j]) \right\| \right) \left(\sum_{j=1}^{n} W_{nj}(T_i)\|X_j\| \right) \\
&\quad \times \sup_z |\triangle_n(z) - \triangle(z)| + o_p(1). \qquad (53)
\end{aligned}$$

By (C.X), (C.$\triangle$), (C.ω) and (C.r), we have

$$\begin{aligned}
&E\left\{ \frac{1}{\sqrt{n}} \left(\sum_{i=1}^{n} \left\| \sum_{j=1}^{n} W_{nj}(T_i)(U_j - E[U_j|T_j]) \right\| \right) \left(\sum_{j=1}^{n} W_{nj}(T_i)\|X_j\| \right) \right\} \\
&\leq \frac{C}{\sqrt{n}} \sum_{i=1}^{n} E^{\frac{1}{2}} \left\| \sum_{j=1}^{n} W_{nj}(T_i)(U_j - E[U_j|T_j]) \right\|^2 E^{\frac{1}{2}} \left\| \sum_{j=1}^{n} W_{nj}(T_i) X_j \right\|^2 \\
&\leq \frac{C}{\sqrt{n}} \sum_{i=1}^{n} \left(\sum_{j=1}^{n} E W_{nj}^2(T_i) \right)^{\frac{1}{2}} \left(\sum_{j=1}^{n} W_{nj}^2(T_i) \right)^{\frac{1}{2}} \\
&\leq C b_n^{-1}
\end{aligned}$$

This together with (31) proves

$$\|S_{n,33}^{[1]}\| = o_p(1). \qquad (54)$$

Similar to (48) and (49), we have

$$S_{n,33}^{[2]} = o_p(1) \quad \text{and} \quad S_{n,33}^{[3]} = o_p(1). \qquad (55)$$

This together with (52), (54) and (55) proves

$$S_{n,33} = o_p(1). \qquad (56)$$

(44), (50), (51) and (57) together prove

$$S_{n,3} = o_p(1). \qquad (57)$$

By (18), (35), (43) and (57), we get

$$S_n = \Sigma\beta \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \frac{\delta_j - \triangle(Z_j)}{\triangle(Z_j)} + o_p(1). \tag{58}$$

It can be proved that

$$\begin{aligned} T_n &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (U_{in} - U_i)[Y_i - \widehat{g}_{2,n}(T_i) - (U_{in} - \widehat{g}_{1,n}(T_i))^\top \beta] \\ &\quad - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sum_{j=1}^{n} W_{nj}(T_i)(U_{jn} - U_j)(Y_i - \widehat{g}_{2,n}(T_i) - (U_{i,n} - \widehat{g}_{1,n}(T_i))^\top \beta] \\ &= T_{n1} + T_{n2}. \end{aligned} \tag{59}$$

For T_{n1}, we have

$$\begin{aligned} T_{n1} &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (U_{in} - U_i)(Y_i - g_2(T_i) - (U_i - g_1(T_i))^\top \beta) \\ &\quad + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (U_{in} - U_i)(g_2(T_i) - \widehat{g}_{2,n}(T_i)) \\ &\quad - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (U_{in} - U_i)(U_{in} - U_i)^\top \beta \\ &\quad + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (U_{in} - U_i)(\widehat{g}_{1,n}(T_i) - g_1(T_i))^\top \beta \\ &= T_{n1,1} + T_{n1,2} + T_{n1,3} + T_{n1,4}. \end{aligned} \tag{60}$$

Similar to (25), it can be proved

$$T_{n1,1} = -E\left[\frac{1 - \triangle(Z)}{\triangle(Z)} X X^\top\right] \beta \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \frac{\delta_j - \triangle(Z_j)}{\triangle(Z_j)} + o_p(1). \tag{61}$$

For $T_{n1,2}$, we have

$$\begin{aligned} T_{n1,2} &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\delta_i X_i(\triangle(Z_i) - \triangle_n(Z_i))}{\triangle^2(Z_i)} \sum_{j=1}^{n} W_{nj}(T_i)(g_2(T_i) - g_2(T_j)) \\ &\quad + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\delta_i X_i(\triangle(Z_i) - \triangle_n(Z_i))}{\triangle^2(Z_i)} \sum_{j=1}^{n} W_{nj}(T_i)(g_2(T_j) - Y_j) \\ &= T_{n1,2}^{[1]} + T_{n1,2}^{[2]}. \end{aligned} \tag{62}$$

By (31) and (C.$h_n b_n$), we have

$$\|T^{[1]}_{n1,2}\| \le C\sqrt{n} b_n \sup_z |\triangle_n(z) - \triangle(z)| \frac{1}{n}\sum_{i=1}^{n}\|X_i\|$$
$$= O_p(\frac{b_n}{h_n}) + O_p(\sqrt{n} b_n h_n^k) = o_p(1). \tag{63}$$

For $T^{[2]}_{n1,2}$, we have

$$\|T^{[2]}_{n1,2}\| \le \sup_z |\triangle_n(z) - \triangle(z)| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \|X_i\| |\sum_{j=1}^{n} W_{nj}(T_i)(g_2(T_j) - Y_j)|. \tag{64}$$

By (C.X) and ($c.\epsilon$), we have

$$\begin{aligned} &E\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\|X_i\|\sum_{j=1}^{n}W_{nj}(T_i)(g_2(T_j)-Y_j)\right]^2 \\ &\le c\sum_{i=1}^{n}E\left\{\|X_i\|^2\left(\sum_{j=1,j\ne i}^{n}W_{nj}(T_i)(g_2(T_j)-Y_j)\right)^2\right\} \\ &\quad + c\sum_{i=1}^{n}E[\|X_i\|^2 W_{ni}^2(T_i)] \\ &\le c\sum_{i=1}^{n}\sum_{j=1,j\ne i}^{n}E\{W_{nj}^2(T_i)E[Y_j^2|T_j]\} + C\sum_{i=1}^{n}E\{W_{ni}^2(T_i)E[\|X_i\|^2|T=t]\} \\ &= O(b_n^{-1}) + O((nb_n)^{-1}). \end{aligned} \tag{65}$$

This proves

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\|X_i\|\sum_{j=1}^{n}W_{nj}(T_i)(g_2(T_j)-T_j) = O_p(b_n^{-\frac{1}{2}}).$$

This together with (62), (63), (64) and (31) proves

$$\|T_{n1,2}\| = o_p(1). \tag{66}$$

by (C.$h_n b_n$) and (C.b_n). Similarly, it can be proved

$$T_{n1,4} = o_p(1). \tag{67}$$

By (31), standard but tedious arguments can be used to prove

$$T_{n1,3} = o_p(1) \quad \text{and} \quad T_{n2} = o_p(1). \tag{68}$$

(59), (60), (61), (62), (63), (66), (67) and (68) together prove

$$T_n = -E\left[\frac{1-\triangle(Z)}{\triangle(Z)} X X^\top\right]\beta \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \frac{\delta_j - \triangle(Z_j)}{\triangle(Z_j)} + o_p(1). \tag{69}$$

By (10), (17), (58) and (69), (8) is then proved.

By the fact

$$\sup_t \left| \sum_{j=1}^{n} W_{nj}(t) \frac{\delta_j X_j}{\triangle(Z_j)} - E\left[\frac{\delta X}{\triangle(Z)}\middle| T = t\right]\right| = O_p((nb_n)^{-\frac{1}{2}}) + O_p(b_n),$$

and (31), it can be prove that

$$\begin{aligned}
&\sup_t |g_1(t) - \widehat{g}_{1,n}(t)| \\
&\le \sup_t \left| \sum_{j=1}^{n} W_{nj}(t) \frac{\delta_j X_j}{\triangle(Z_j)} - E\left[\frac{\delta X}{\triangle(Z)} | T = t\right]\right| \\
&\quad + \sup_t \left| \sum_{j=1}^{n} W_{nj}(t) \frac{\delta_j X_j (\triangle(Z_j) - \triangle_n(Z_j))}{\triangle(Z_j)\triangle_n(Z_j)}\right| \\
&\le \sup_t \left| \sum_{j=1}^{n} W_{nj}(t) \frac{\delta_j X_j}{\triangle(Z_j)} - E\left[\frac{\delta X}{\triangle(Z)}\middle| T = t\right]\right| \\
&\quad + \sup_z |\triangle_n(z) - \triangle(z)| \left| \sum_{j=1}^{n} W_{nj}(t) \frac{\|\delta_j X_j\|}{\triangle(Z_j)\triangle_n(Z_j)}\right| \\
&= O_p((nb_n)^{-\frac{1}{2}}) + O_p(b_n) + O_p((nh_n^2)^{-\frac{1}{2}}) + O_p(h_n^k).
\end{aligned} \tag{70}$$

(31) and (70) together prove (9).

Note that $\Sigma - E\left[\frac{1-\triangle(Z)}{\triangle(Z)} X X^\top\right]\beta = \Sigma_0$ and

$$E[(U - g_1(T))(Y - g_2(T) - (U - g_1(T))^\top \beta] = -E\left[\frac{1-\triangle(Z)}{\triangle(Z)} X X^\top\right]\beta.$$

The central limit theorem can be applied to (8) and (9) to prove Theorem 1 by some simple calculations. □

Proof of Theorem 2. Theorem 2 is a direct result of Theorem 1 by noting

$$\sqrt{n}(\widehat{\mu}_{MC} - \mu_{MC}) = o_p(1).$$

□

Proof of Theorem 3. Observe

$$\begin{aligned}\widehat{g}_n(t) - g(t) = \widehat{g}_{2n}(t) - g_2(t) - (\widehat{g}_{1n}(t) - g_1(t))^\top \\ \beta - (\widehat{g}_{1n}(t) - g_1(t))^\top(\widehat{\beta}^* - \beta) - g_1^\top(t)(\widehat{\beta}^* - \beta).\end{aligned}$$

By (70) and the following fact

$$\sup_t |\widehat{g}_{2,n}(t) - g_2(t)| = O_p((nb_n)^{-\frac{1}{2}}) + O_p(b_n) \tag{71}$$

and the assumption that $\widehat{\beta}^* - \beta = O_p(n^{-\frac{1}{2}})$, we get

$$\widehat{g}^*_{MC}(t) - g(t) = O_p((nb_n)^{-\frac{1}{2}}) + O_p(b_n) + O_p((nh_n^2)^{-\frac{1}{2}}) + O_p(h_n^k) + O_p(n^{-\frac{1}{2}}).$$

Theorem 2 is then proved if $b_n = n^{-\frac{1}{3}}$ and $h_n = n^{-\frac{1}{6}}$. □

Appendix B: Proofs of Theorems 4 and 5

(C.M(z)): $M(z)$ has bounded partial derivatives up to order $k(> 2)$ almost surely.

Proof of Theorem 4. Clearly

$$\sqrt{n}(\widetilde{\beta}_n - \beta) = \sqrt{n}\widetilde{B}_n^{-1}[\widetilde{A}_n - \widetilde{B}_n\beta)] \tag{72}$$

and

$$\begin{aligned}&\sqrt{n}(\widetilde{A}_n - \widetilde{B}_n\beta)\\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{\delta_i(X_i - g_1(T_i))}{\triangle_n(Z_i)}(Y_i - g_2(T_i) - (X_i - g_1(T_i))^\top\beta)\\
&\quad + \frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{\delta_i(X_i - g_1(T_i))}{\triangle_n(Z_i)}(g_2(T_i) - \widehat{g}_{2,n}(T_i))\\
&\quad + \frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{\delta_i(X_i - g_1(T_i))}{\triangle_n(Z_i)}(\widehat{g}_{1,n}(T_i) - g_1(T_i))^\top\beta\\
&\quad + \frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{\delta_i(g_1(T_i) - \widehat{g}_{1,n}(T_i))}{\triangle_n(Z_i)}(Y_i - \widehat{g}_{2,n}(T_i) - (X_i - \widehat{g}_{1,n}(T_i))^\top\beta)\\
&:= E_n + F_n + G_n + H_n.\end{aligned} \tag{73}$$

Observe that

$$
\begin{aligned}
E_n &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\delta_i(X_i-g_1(T_i))\epsilon_i}{\triangle(Z_i)}+\frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i(X_i-g_1(T_i))(\triangle(Z_i)-\triangle_n(Z_i))\epsilon_i}{\triangle^2(Z_i)}\\
&\quad+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\delta_i(X_i-g_1(T_i)(\triangle(Z_i)-\triangle_n(Z_i))^2\epsilon_i}{\widehat{\triangle}_n(Z_i)\triangle^2(Z_i)}\\
&:= E_{n,1}+E_{n,2}+E_{n,3}.
\end{aligned}
\tag{74}
$$

For $E_{n,2}$, we have

$$
\begin{aligned}
E_{n,2} &= -\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\delta_i(X_i-g_1(T_i))\epsilon_i}{\triangle^2(Z_i)f_{\mathcal{Z}}(Z_i)}\frac{1}{nh_n^2}\sum_{j=1}^{n}(\delta_j-\triangle(Z_j))K\left(\frac{Z_i-Z_j}{h_n}\right)\\
&\quad-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\delta_i(X_i-g_1(T_i))\epsilon_i}{\triangle^2(Z_i)f_{\mathcal{Z}}(Z_i)}\frac{1}{nh_n^2}\sum_{j=1}^{n}(\triangle(Z_j)-\triangle(Z_i))K\left(\frac{Z_i-Z_j}{h_n}\right)+o_p(1)\\
&:= E_{n,21}+E_{n,22}+o_p(1).
\end{aligned}
\tag{75}
$$

Clearly

$$
\begin{aligned}
E_{n,21} &= -\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{M(Z_i)}{\triangle(Z_i)f_{\mathcal{Z}}(Z_i)}\frac{1}{nh_n^2}\sum_{j=1}^{n}(\delta_j-\triangle(Z_j))K\left(\frac{Z_i-Z_j}{h_n}\right)\\
&\quad-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\frac{\delta_i(X_i-g_1(T_i))\epsilon_i}{\triangle^2(Z_i)f_{\mathcal{Z}}(Z_i)}-E\left[\frac{\delta_i(X_i-g_1(T_i))\epsilon_i}{\triangle^2(Z_i)f_{\mathcal{Z}}(Z_i)}\Big|Z_i\right]\right\}\\
&\quad\times\frac{1}{nh_n^2}\sum_{j=1}^{n}(\delta_j-\triangle(Z_j))K\left(\frac{Z_i-Z_j}{h_n}\right)\\
&:= E_{n,21}^{[1]}+E_{n,21}^{[2]}
\end{aligned}
\tag{76}
$$

By (C.K)ii, (C.$\triangle$) and (C.M), similar to (24) we have

$$
E_{n2,1}^{[1]} = -\frac{1}{\sqrt{n}}\sum_{j=1}^{n}(\delta_j-\triangle(Z_j))\frac{M(Z_j)}{\triangle(Z_j)}+o_p(1).
\tag{77}
$$

For $E_{n,21}^{[2]}$, we have

$$
\begin{aligned}
E_{n,21}^{[2]} &= -\frac{1}{n^{\frac{3}{2}}h_n^2}\sum_{i\neq j}\left\{\left(\frac{\delta_i(X_i-g_1(T_i))\epsilon_i}{\triangle^2(Z_i)f_{\mathcal{Z}}(Z_i)}-E\left[\frac{\delta_i(X_i-g_1(T_i))\epsilon_i}{\triangle^2(Z_i)f_{\mathcal{Z}}(Z_i)}\Big|Z_i\right]\right)\right.\\
&\qquad\left.(\delta_j-\triangle(Z_j))K\left(\frac{Z_i-Z_j}{h_n}\right)\right\}+o_p(1).
\end{aligned}
\tag{78}
$$

Let M_n be the first term of (78). Then

$$\begin{aligned}
EM_n^2 &= \frac{1}{n^3h_n^4}\sum_{i\neq j} E\Big\{E[\|X_i - g_1(T_i)\|^2\epsilon_i^2|Z_i]\frac{K^2\left(\frac{Z_i-Z_j}{h_n}\right)}{\triangle^3(Z_i)f_{\mathcal{Z}}^2(Z_i)}(\triangle(Z_j)-\triangle^2(Z_j))\Big\}\\
&\leq \frac{c}{n^2h_n^2}\sum_{i=1}^{n} E\Big\{E[\|X_i - g_1(T_i)\|^2\epsilon_i^2|Z_i]\\
&\quad\times\frac{\int(\triangle(Z_i-h_nu)-\triangle^2(Z_i-h_nu))K^2(u)f_{\mathcal{Z}}(Z_i-h_nu)\,\mathrm{d}u}{\triangle^3(Z_i)f_{\mathcal{Z}}^2(Z_i)}\Big\}\\
&\leq \frac{c}{n^2h_n^2}\sum_{i=1}^{n} E[\|X_i - g_1(T_i)\|^2\epsilon_i^2\frac{1-\triangle(Z_i)}{\triangle^2(Z_i)f(Z_i)}]\int K^2(u)\,\mathrm{d}u\\
&\leq c(nh_n^2)^{-1}\longrightarrow 0.
\end{aligned}\tag{79}$$

This together with (78) proves

$$E_{n,21}^{[2]} = o_p(1).\tag{80}$$

By (C.X), $(C.\epsilon)$, (C.$\triangle$)ii, (C.f), we have

$$\begin{aligned}
E_{n,22} &= -\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\delta_i(X_i-g_1(T_i))\epsilon_i}{\triangle^2(Z_i)f_{\mathcal{Z}}(Z_i)}\\
&\qquad\int(\triangle(Z_i-h_nu)-\triangle(Z_i))K(u)f(Z_i-h_nu)\,du + o_p(1)\\
&\leq c\sqrt{n}h_n^k + o_p(1).
\end{aligned}\tag{81}$$

as $nh_n^{2k}\to 0$.

By (75), (76), (77), (80) and (81), we get

$$E_{n2} = \frac{1}{\sqrt{n}}\sum_{j=1}^{n}(\triangle(Z_j)-\delta_j)\frac{M(Z_j)}{\triangle(Z_j)} + o_p(1).\tag{82}$$

where $M(z)$ is as defined in Sect. 3.

For E_{n3}, we have

$$\begin{aligned}
E_{n3} &\leq \sqrt{n}\sup_z|\triangle(z)-\widehat{\triangle}_n(z)|^2\Big\{\frac{1}{n}\sum_{i=1}^{n}\frac{|\delta_i(X_i-g_1(T_i))\epsilon_i|}{\widehat{\triangle}_n(Z_i)}I[\widehat{\triangle}_n(Z_i)\geq\frac{1}{2}\triangle(Z_i)]\\
&\quad+\frac{1}{n}\sum_{i=1}^{n}\left|\frac{\delta_i(X_i-g_1(T_i))\epsilon_i}{\widehat{\triangle}_n(Z_i)}\right|I[\widehat{\triangle}_n(Z_i)<\frac{1}{2}\triangle(Z_i)]\Big\}
\end{aligned}\tag{83}$$

Clearly

$$\frac{1}{n}\sum_{i=1}^{n}\left|\frac{\delta_i(X_i-g_1(T_i))\epsilon_i}{\widehat{\triangle}_n(Z_i)}\right| I[\widehat{\triangle}_n(Z_i)\geq \frac{1}{2}\triangle(Z_i)] \leq \frac{2}{n}\sum_{i=1}^{n}\left|\frac{\delta_i(X_i-g_1(T_i))\epsilon_i}{\triangle(Z_i)}\right| = O_p(1). \tag{84}$$

For any $\epsilon > 0$, we have

$$\begin{aligned} &P\left(\frac{1}{n}\sum_{i=1}^{n}\left|\frac{\delta_i(X_i-g_1(T_i))\epsilon_i}{\widehat{\triangle}_n(Z_i)}\right| I[\widehat{\triangle}_n(Z_i) < \frac{1}{2}\triangle(Z_i)] > \epsilon\right) \\ &\leq P\left(\bigcup_{i=1}^{n}\{|\widehat{\triangle}_n(Z_i)-\triangle(Z_i)| > \frac{1}{2}\inf_z \triangle(z)\}\right) \\ &= P\left(\sup_z |\widehat{\triangle}_n(z)-\triangle(z)| > \frac{1}{2}\inf_z \triangle(z)\right) \longrightarrow 0. \end{aligned} \tag{85}$$

by (C.$\triangle$). (83), (84) and (85) together with (29) prove

$$E_{n3} = O_p(n^{-\frac{1}{2}}h_n^{-2}) + O_p(\sqrt{n}h_n^{2k}) = o_p(1). \tag{86}$$

as $nh_n^4 \to \infty$ and $nh_n^{4k} \to 0$, which are implied by (C.h_n). (74), (82) and (86) together prove

$$E_n = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\frac{\delta_i(X_i-g_1(T_i))\epsilon_i}{\triangle(Z_i)} + (\triangle(Z_i)-\delta_i)\frac{M(Z_i)}{\triangle(Z_i)}\right] + o_p(1). \tag{87}$$

For F_n, we have

$$\begin{aligned} F_n &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\delta_i(X_i-g_1(T_i))}{\triangle(Z_i)}\sum_{j=1}^{n}W_{nj}(T_i)(g_2(T_i)-g_2(T_j)) \\ &\quad+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\delta_i(X_i-g_1(T_i))}{\triangle(Z_i)}\sum_{j=1}^{n}W_{nj}(T_i)(g_2(T_j)-Y_j) \\ &\quad+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\delta_i(X_i-g_1(T_i))}{\widehat{\triangle}_n(Z_i)\triangle(Z_i)}(\triangle(Z_i)-\triangle_n(Z_i))(g_2(T_i)-\widehat{g}_{2,n}(T_i)) \\ &:= F_{n1}+F_{n2}+F_{n3}. \end{aligned} \tag{88}$$

Under MAR, we have

$$E\left[\frac{\delta(X-g_1(T)}{\triangle(Z)}\Big|T\right]=E\left\{E\left[\frac{\delta(X-g_1(T))}{\triangle(Z)}|Y,X,T\right]|T\right\}=E[(X-g_1(T))|T]=0.$$

Let $\xi_k = \frac{\delta_k(X_k - g_1(T_k))}{\triangle(Z_k)}$ and $\zeta_{nk} = \sum_{j=1}^{n} W_{nj}(T_k)(g_2(T_k) - g_2(T_j))$. Then,

$$E[\xi_k \xi_l \zeta_{nk} \zeta_{nj}] = E\{E[E(\xi_k | Y_k, T_k, X_k)|T_k]E[E(\xi_l | Y_l, T_l, X_l)|T_l]\zeta_{nk}\zeta_{nl}\} = 0.$$

This together with assumptions (C.$\triangle$), (C.X), (C.K), (C.g) and (C.b_n) proves

$$E\|F_{n1}\|^2 \le Cnb_n^2(n^2 b_n)^{-1} \le cb_n \longrightarrow 0.$$

by similar arguments to (13). This proves

$$F_{n1} = o_p(1). \tag{89}$$

By arguments similar to (13), (14), (15) and (16), it can be proved

$$F_{n2} = o_p(1). \tag{90}$$

Similar to (86), we have

$$F_{n3} = o_p(1) \tag{91}$$

by (31) and (71) as $nh_n^2 b_n \to \infty$, $\frac{h_n^{2k}}{b_n} \to 0$ and $\frac{b_n}{h_n} \to 0$, which are implied by (C.$h_n b_n$) and (C.h_n). This together with (88), (89), (89) and (91) proves

$$F_n = o_p(1). \tag{92}$$

For G_n, we have

$$\begin{aligned}
G_n &= \frac{1}{\sqrt{n}} \frac{\delta_i(X_i - g_1(T_i))}{\triangle(Z_i)} (g_1(T_i) - \widehat{g}_{1,n}(T_i))^\top \beta + o_p(1) \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\delta_i(X_i - g_1(T_i))}{\triangle(Z_i)} \sum_{j=1}^{n} W_{nj}(T_i)(g_1(T_i) - g_1(T_j))^\top \beta \\
&\quad + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\delta_i(X_i - g_1(T_i))}{\triangle(Z_i)} \sum_{j=1}^{n} W_{nj}(T_i)(g_1(T_j) - \frac{\delta_j X_j}{\triangle(Z_j)})^\top \beta \\
&\quad + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\delta_i(X_i - g_1(T_i))}{\triangle(Z_i)} \sum_{j=1}^{n} W_{nj}(T_i) \left(\frac{\delta_j X_j}{\triangle(Z_j)} - \frac{\delta_j X_j}{\triangle_n(Z_j)} \right)^\top \beta + o_p(1) \\
&= G_{n,1}^{[1]} + G_{n,1}^{[2]} + G_{n,1}^{[3]} + o_p(1).
\end{aligned} \tag{93}$$

Similar to the proofs of (13), (14), (15) and (16), we have

$$G_{n,1}^{[1]} = o_p(1) \quad \text{and} \quad G_{n,2}^{[2]} = o_p(1). \tag{94}$$

For $G_{n,1}^{[3]}$, using some arguments similar to (83) and (84) we have

$$\begin{aligned}\|G_{n,1}^{[3]}\| &= \|\sum_{j=1}^{n}\left\{\sum_{i=1}^{n}W_{nj}(T_i)\frac{\delta_i(X_i-g_1(T_i))}{\triangle(Z_i)}\right\}\frac{\delta_j X_j^\top\beta}{\triangle(Z_j)\triangle_n(Z_j)}(\triangle_n(Z_j)-\triangle(Z_j))\|\\ &\le \sup_z|\triangle_n(z)-\triangle(z)|\|\sum_{i=1}^{n}W_{nj}(T_i)\frac{\delta_i(X_i-g_1(T_i))}{\triangle(Z_i)}\frac{\delta_j X_j^\top\beta}{\triangle(Z_j)\triangle_n(Z_j)}\\ &\quad\times I[\triangle_n(Z_j)\ge\frac{1}{2}\triangle(Z_j)]\|+o_p(1)\end{aligned} \tag{95}$$

By (C.X), (C.$\triangle$), (C.ω), we have

$$\begin{aligned}&E\left\|\sum_{j=1}^{n}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\delta_i(X_i-g_1(T_i))}{\triangle(Z_i)}W_{nj}(T_i)\right|X_j^\top\beta\right\|^2\\ &\le n\sum_{j=1}^{n}E\left[\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\delta_i(X_i-g_1(T_i))}{\triangle(Z_i)}W_{nj}(T_i)\right\|^2(X_j^\top\beta)^2\right]\\ &\le\sum_{j=1}^{n}E\left[\sum_{i=1,i\ne j}^{n}E[\|X_i\|^2|T_i]W_{nj}^2(T_i)(X_j^\top\beta)^2\right]\\ &\quad+\sum_{j=1}^{n}E\left[\frac{\delta_j\|X_j-g_1(T_j)\|^2}{\triangle^2(Z_j)}W_{nj}^2(T_j)(X_j^\top\beta)^2\right]\\ &\le cn^2EW_{nj}^2(T_i)\le cb_n^{-1}.\end{aligned} \tag{96}$$

This proves

$$\left\{\sum_{i=1}^{n}W_{nj}(T_i)\frac{\delta_i(X_i-g_1(T_i))}{\triangle(Z_i)}\right\}\frac{\delta_j X_j^\top\beta}{\triangle(Z_j)\triangle_n(Z_j)}I\left[\triangle_n(Z_j)\ge\frac{1}{2}\triangle(Z_j)\right]=o_p(b_n^{-\frac{1}{2}}).$$

This together with (31) and (95) proves

$$G_{n,1}^{[3]}=o_p(1) \tag{97}$$

by (C.h_n) and (C.h_nb_n). By (93), (94) and (97), we have

$$G_n=o_p(1). \tag{98}$$

For H_n, we have

$$\begin{aligned}
H_n &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\delta_i(g_1(T_i)-\widehat{g}_{1,n}(T_i))\epsilon_i}{\triangle(Z_i)} \\
&\quad+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\delta_i(g_1(T_i)-\widehat{g}_{1,n}(T_i))(\triangle(Z_i)-\triangle_n(Z_i))\epsilon_i}{\triangle_n(Z_i)\triangle(Z_i)} \\
&\quad+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\delta_i(g_1(T_i)-\widehat{g}_{1,n}(T_i))}{\triangle_n(Z_i)}(g_2(T_i)-\widehat{g}_{2,n}(T_i)) \\
&\quad+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\delta_i(g_1(T_i)-\widehat{g}_{1,n}(T_i))}{\triangle_n(Z_i)}(\widehat{g}_{1,n}(T_i)-g_1(T_i))^{\top}\beta \\
&:= H_{n1}+H_{n2}+H_{n3}+H_{n4}.
\end{aligned}\tag{99}$$

For H_{n1}, we have

$$\begin{aligned}
H_{n1} &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\delta_i\epsilon_i}{\triangle(Z_i)}\sum_{j=1}^{n}W_{nj}(T_i)(g_1(T_i)-g_1(T_j)) \\
&\quad+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\delta_i\epsilon_i}{\triangle(Z_i)}\sum_{j=1}^{n}W_{nj}(T_i)\left(g_1(T_i)-\frac{\delta_jX_j}{\triangle(Z_j)}\right) \\
&\quad+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\delta_i\epsilon_i}{\triangle(Z_i)}\sum_{j=1}^{n}W_{nj}(T_i)\left(\frac{\delta_jX_j}{\triangle(Z_j)}-\frac{\delta_jX_j}{\triangle_n(Z_j)}\right) \\
&:= H_{n1,1}+H_{n1,2}+H_{n1,3}.
\end{aligned}\tag{100}$$

Similar to (89) and (89), we have

$$H_{n1,1}=o_p(1)\qquad H_{n1,2}=o_p(1).\tag{101}$$

For $H_{n1,3}$, we have

$$\begin{aligned}
H_{n1,3} &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\delta_i\epsilon_i}{\triangle(Z_i)}\sum_{j=1}^{n}W_{nj}(T_i)\frac{U_j}{\triangle(Z_j)}(\triangle_n(Z_j)-\triangle(Z_j))+o_p(1) \\
&= \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left(\sum_{i=1}^{n}\frac{\delta_i\epsilon_i}{\triangle(Z_i)}W_{nj}(T_i)\right)\frac{U_j}{\triangle(Z_j)}(\triangle_n(Z_j)-\triangle(Z_j))+o_p(1).
\end{aligned}\tag{102}$$

Clearly, the main term in the above formula can be bounded by

$$\sup_z|\triangle_n(z)-\triangle(z)|Q_n,\tag{103}$$

where

$$Q_n = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \left\| \sum_{i=1}^{n} \frac{\delta_i \epsilon_i}{\triangle(Z_i)} W_{nj}(T_i) \frac{U_j}{\triangle(Z_j)} \right\|$$

By (C.$\triangle$) and (C.X), we have

$$\begin{aligned} E\|Q_n\| &\le \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \left[E^{\frac{1}{2}} \| \frac{\delta_j X_j}{\triangle(Z_j)} \|^2 E^{\frac{1}{2}} \left(\sum_{i=1}^{n} W_{nj}(T_i) \frac{\delta_i \epsilon_i}{\triangle(Z_i)} \right)^2 \right] \\ &\le \frac{c}{\sqrt{n}} \sum_{j=1}^{n} \left(\sum_{i=1}^{n} E W_{nj}^2(T_i) \frac{\delta_i \epsilon_i^2}{\triangle^2(Z_i)} \right)^{\frac{1}{2}} \le c b_n^{-\frac{1}{2}}. \end{aligned}$$

This proves

$$Q_n = o_p(b_n^{-\frac{1}{2}}). \tag{104}$$

(31), (102), (103) and (104) together prove

$$H_{n1,3} = o_p(1) \tag{105}$$

as $nh_n^2 b_n \to \infty$ and $\frac{h_n^{2k}}{b_n} \to 0$, which is implied by (C.$h_n b_n$).

(100), (101) and (105) together prove

$$H_{n1} = o_p(1). \tag{106}$$

Using arguments similar to (83) and (84), by (31) and (70), we get

$$\begin{aligned} |H_{n2}| &\le \sqrt{n} \sup_t |\widehat{g}_{1,n}(t) - g_1(t)| \sup_z |\triangle_n(z) - \triangle(z)| \left(\frac{2}{n} \sum_{i=1}^{n} \frac{\delta_i |\epsilon_i|}{\triangle^2(Z_i)} + o_p(1) \right) \\ &= O_p(n^{-\frac{1}{2}} b_n^{-\frac{1}{2}} h_n^{-1}) + O_p(\frac{b_n}{h_n}) + O_p\left(\frac{h_n^k}{b_n^{\frac{1}{2}}} \right) + O_p(\sqrt{n} h_n^k b_n) \\ &\quad + O_p(n^{\frac{1}{2}} h_n^2) + O_p(\sqrt{n} h_n^{2k}). \end{aligned} \tag{107}$$

as $nh_n^4 \to \infty$, $nh_n^2 b_n \to \infty$, $nh_n^{4k} \to 0$, $nh_n^{2k} b_n^2 \to 0$, $\frac{h_n^{2k}}{b_n} \to 0$, $\frac{b_n}{h_n} \to 0$, which are implied by (C.h_n) and (C.$h_n b_n$).

Similarly, we have

$$H_{n3} = o_p(1) \quad \text{and} \quad H_{n4} = o_p(1). \tag{108}$$

(99), (106), (107) and (108) together prove

$$H_n = o_p(1). \tag{109}$$

By (73), (87), (92), (98) and (109), we get

$$\sqrt{n}(\widetilde{A}_n - \widetilde{B}_n\beta) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\frac{\delta_i(X_i - g_1(T_i))\epsilon_i}{\triangle(Z_i)} + (\triangle(Z_i) - \delta_i)\frac{M(Z_i)}{\triangle(Z_i)}\right] + o_p(1). \tag{110}$$

Central limit theorem can be used to prove

$$\sqrt{n}(\widetilde{A}_n - \widetilde{B}_n\beta) \xrightarrow{\mathcal{L}} N(0, \Omega_W), \tag{111}$$

where

$$\Omega_W = E\left[\frac{(X - E[X|T])(X - E[X|T])^\top(Y - X^\top\beta - g(T))^2}{\triangle(Z)}\right] \\ -E\left[\frac{M(z)M^\top(Z)(1 - \triangle(Z))}{\triangle(Z)}\right]$$

By (31), (70) and (71), it can be proved

$$\widetilde{B}_n \xrightarrow{p} \Sigma_W. \tag{112}$$

(111) and (112) together prove Theorem 4. □

Proof of Theorem 5. The proof is similar to that of Theorem 5. □

References

Ahn, H., Powell, J. L. (1993). Estimation of censored selection models with a nonparametric selection mechanism. *Journal of Econometrics*, *58*, 3–30.

Chen, H. (1988). Convergent rates for parametric components in a partly linear model. *The Annals of Statistics*, *16*, 136–146.

Chen, H., Shiau, H. J. (1991). A two-stage spline smoothing method for partially linear model. *Journal of Statistical Planning and Inference*, *27*, 187–201.

Engle, R., Granger, C., Rice, J., Weiss, A. (1986). Nonparametric estimates of the relation between weather and electricity sales. *Journal of the American Statistical Association*, *81*, 310–320.

Gray, R. (1994). Spline-based tests in survival analysis. *Biometrics*, *50*, 640–652.

Green, P., Siverman, B. W. (1994). Nonparametric regression and generalized linear model: a rougher's penalty approach. vol 58 of *Monographs on Statistics and Applied Probabilities*. New York: Chapman and Hall.

Hamilton, S. A., Truong, Y. K. (1997). Local linear estimation in partly linear models. *Journal of Multivariate Analysis*, *60*, 1–19.

Härdle, W., Liang, H., Gao, J. (2000). Partially linear models. Heidelberg, Physica-Verlag.

Heckman, N. (1986). Spline smoothing in partly linear models. *Journal of the Royal Statistical Society* B, *48*, 244–248.

Hong, S.-Y. (1999). Automatic bandwidth choice in a semiparametric regression model. *Statistica Sinica*, *9*, 775–794.

Horvitz, D. G., Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, *90*, 141–150.

Ibrahim, J. G. (1990). Incomplete data in generalized linear models. *Journal of the American Statistical Association*, *85*, 765–769.

Liang, H., Wang, S., Robins, J. M., Carroll, R. J. (2004). Estimation in partially linear models with missing covariates. *Journal of the American Statistical Association*. *99*, 357–367.

Lipsitz, S. R., Ibrahim, J. G. (1996). A conditional model for incomplete covariates in parametric regression models. *Biometrika*, *83*, 916–922.

Lipsitz, S. R., Ibrahim, J. G., Zhao, L. P. (1999). A weighted estimating equation for missing covariate data with properties similar to maximum likelihood. *Journal of the American Statistical Asssociation*, *94*, 1147–1160.

Little, R. J. A., Rubin, D. B. (1987). *Statistical analysis with missing data.* New York: Wiley.

Newey, W. K., Powell, J. L., Walker, J. R. (1990). Semiparametric estimation of selection models: some empirical results. *American Economic Review*, *80*, 324–328.

Prakasa Rao, B. L. S. (1981). *Nonparametric functional estimation*. Orlanda, Fla.: Academic Press.

Qi, L., Wang, C. Y., Prentice, R. L. (2005). Weighted estimators for proportional hazards regression with missing covariates. *Journal of the American Statistical Association*, *100*, 1250–1263.

Rice, J. (1986). Concergence rates for partially splined models. *Statistics & Probability Letters*, *4*, 203–208.

Robins, J. M., Rontnitzky, A., Zhao, L. P. (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of the Royal Statistical Society*, *89*, 846–866.

Robinson, P. M. (1988). Root n-consistent semiparametric regression. *Econometrica, 56*, 931–954.

Schick, A. (1996). Root-n-consistent in partly linear regression models. *Statistics and Probability Letters*, *28*, 353–358.

Schmalensee, R., Stoker, T. M. (1999). Household gaseline demand in the United States. *Econometrica*, *67*, 645–642.

Severini, T. A., Staniswalis, J. G. (1994). Quasilikelihood estimation in semiparametric models. *Journal of the American Statistical Asssociation*, *89*, 501–511.

Speckman, P. (1988). Kernel smoothing in partial linear models. *Journal of the Royal Statistical Society Ser B*, *50*, 413–436.

Stock, J. H. (1989). Nonparametric policy analysis. *Journal of the American Statistical Association*, *84*, 567–576.

Stock, J. H. (1991). Nonparametric policy analysis: an application to estimating hazardous waste cleanup benefits. In Barnett, Powell, Tauchen (Eds.) *Nonparametric and semiparametric methods in econometrics and statistics*, pp. 77–98. Cambridge, Cambridge University Press.

Stone, C. J. (1980). Optimal rates of convergence for nonparametric estimators. *The Annals of Statistics*, *8*, 1348–1360.

Vach, W. (1994). Logistic Regression with missing values in the covariates. *Lecture Notes in Statistics*, vol. 86. Heidelberg: Springer.

Wang, C. Y., Wang, S., Gutierrez, R. G. (1998). Local linear regression for generalized linear models with missing data. *The Annals of Statistics*, *26*, 1028–1050.

Wang, Q. H. (1999). Estimation of partial linear error-in-variables models with validation data. *Journal of Multivariate Analysis*, *69*, 30–64.

Wang, Q. H., Jing, B. Y. (1999). Empirical likelihood for partial linear model with fixed design. *Statistics & Probability Letters*, *41*, 425–433.

Wang, Q. H., Rao, J. N. K (2002). Empirical likelhood-based inference under imputation for missing response data. *The Annals of Statistics*, *30*(3), 345–358.

Wang, Q. H., Linton, O., Härdle, W. (2004). Semiparametric regression analysis with missing response at random. *Journal of the American Statistical Association*, *99*, 334–345.

Zhao, L. P., Lipsitz, S. R., Lew, D. (1996). Regression analysis with missing covariate data using estimating equations. *Biometrics*, *52*, 1165–1182.