

Wang, Qihua; Zhang, Tao; Härdle, Wolfgang Karl

**Working Paper**

## An extended single index model with missing response at random

SFB 649 Discussion Paper, No. 2014-003

**Provided in Cooperation with:**

Collaborative Research Center 649: Economic Risk, Humboldt University Berlin

*Suggested Citation:* Wang, Qihua; Zhang, Tao; Härdle, Wolfgang Karl (2014) : An extended single index model with missing response at random, SFB 649 Discussion Paper, No. 2014-003, Humboldt University of Berlin, Collaborative Research Center 649 - Economic Risk, Berlin

This Version is available at:

<https://hdl.handle.net/10419/91583>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

SFB 649 Discussion Paper 2014-003

# An Extended Single Index Model with Missing Response at Random

Qihua Wang\*

Tao Zhang\*\*

Wolfgang Karl Härdle\*\*\*



\* Shenzhen University, China; Chinese Academy of Sciences, China

\*\* Guangxi University of Technology, China

\*\*\* Humboldt-Universität zu Berlin, Germany and C.A.S.E Center for Applied Statistics and Economics, Germany

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".

<http://sfb649.wiwi.hu-berlin.de>

ISSN 1860-5664

SFB 649, Humboldt-Universität zu Berlin  
Spandauer Straße 1, D-10178 Berlin



SFB 649 ECONOMIC RISK BERLIN

# An Extended Single Index Model with Missing Response at Random

Qihua Wang<sup>1,3</sup>, Tao Zhang<sup>2</sup> and Wolfgang Karl Härdle<sup>4,5</sup>

<sup>1</sup>Institute of Statistical Science, Shenzhen University, Shenzhen 518060, China

<sup>2</sup>College of Science, Guangxi University of Technology, Liuzhou, China

<sup>3</sup> Academy of Mathematics and Systems Sciences, Chinese Academy of Sciences  
Beijing 100190, China

<sup>4</sup> School of Business, Singapore Management University,  
50 Stamford Road, Singapore 178899

and

<sup>5</sup> C.A.S.E Center for Applied Statistics and Economics, Humboldt-Universität zu Berlin  
Berlin 10099, Germany

## Abstract

An extended single-index model is considered when responses are missing at random. A three-step estimation procedure is developed to define an estimator for the single index parameter vector by a joint estimating equation. The proposed estimator is shown to be asymptotically normal. An iterative scheme for computing this estimator is proposed. This algorithm only involves one-dimensional nonparametric smoothers, thereby avoiding the data sparsity problem caused by high model dimensionality. Some simulation study is conducted to investigate the finite sample performances of the proposed estimators.

*Key words:* Missing data; Estimating equations; Single-index models; Asymptotic normality.

*AMS(2000) subject classifications.* 62J99; 62E20

# 1 Introduction

The single-index model has been paid considerable attention recently because it is useful in several areas of science such as econometrics, biostatistics, finance and so on. The single-index model (SIM), which is investigated extensively, is of the following form

$$Y = g(\beta^\top X) + \varepsilon, \quad (1.1)$$

where  $Y$  is the univariate response and  $X$  is a  $d$ -dimensional covariable vector,  $\beta$  is an unknown index parameter vector of interest, the function  $g(\cdot)$  is an unknown link function, and  $E(\varepsilon|X) = 0$ . The SIM provides dimension reduction in the sense that, if one can estimate the index  $\beta$  efficiently, the univariate index  $\beta^\top X$  serves as a covariable to estimate the nonparametric link  $g(\cdot)$ . Much effort has been devoted to estimating the index  $\beta$  efficiently. Hall (1989), Zhu and Fang (1992) considered a projection pursuit framework. Härdle et al. (1993) employed the kernel smoothing method to study the model (1.1), and gave an empirical rule for bandwidth selection. Ichimura (1993) studied the properties of a semi-parametric least-squares estimator in a general single-index model. Ichimura (1987) showed that the parameter vector  $\beta$  can be estimated root- $n$  consistently. Härdle et al.(1993) and Hristache et al.(2001) obtained a  $\sqrt{n}$  consistent estimator of the index vector  $\beta$  using the average derivative method. The technology of sliced inverse regression can also be used to achieve  $\sqrt{n}$  consistent estimator, see Li (1991) and Zhu (1996).

Let  $(Y_i, X_i)$  denote the observed values with  $Y_i$  being the response variable and  $X_i$  being the vector of  $d$  explanatory variables. In this paper we consider an extended single index model (ESIM) which specifies the relationship of the mean and variance of  $Y_i$  as follows

$$E(Y_i|X_i) = \mu\{g(\beta^\top X_i)\}, \text{Var}(Y_i|X_i) = \sigma^2 V\{g(\beta^\top X_i)\}, \quad (1.2)$$

where  $\mu(\cdot)$  is a known monotonic function,  $V(\cdot)$  is a known covariance function,  $g(\cdot)$  is an unknown univariate link function and  $\beta$  is an unknown index vector which belongs to the parameter space  $\Theta = \{\beta = (\beta_1, \dots, \beta_d)^\top : \|\beta\| = 1, \beta_1 > 0, \beta \in \mathbb{R}^d\}$ . Cui, Härdle and Zhu (2011) developed a method of estimating function (EFM) to study the ESIM. They investigated the efficiency and computation of the estimates for the ESIM, and obtained the

asymptotic properties of the EFM. However, the existing work is for the case where data are observed fully.

In practice, some responses may be missing, by design (as in two-stage studies) or by circumstance. For example, the response  $Y'$ s may be very expensive to measure and only part of  $Y'$ s are available. Another example is that the  $Y'$ 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. Missing data issues have been investigated extensively. See, e.g., Rosenbaum and Rubin (1983), Robins et al. (1994), Robins et al. (1995), Wang et al. (2002), Wang et al. (2004) and among others. To the best of our knowledge, the literature is reduced to just a few recent papers for the single-index models (1.1) with  $\mu\{g(\beta^\top X_i)\} = g(\beta^\top X_i)$  and  $V\{g(\beta^\top X_i)\} = 1$  for missing data. For this special case, Wang et al. (2010) derived semi-parametric nonlinear least squares estimators by incorporating missing mechanism into the least-squares loss function proposed by Härdle et al.(1993) and minimizing the loss function with respect to the bandwidth and the parameters simultaneously. They obtained the central limit theorem(CLT), the law of the iterated logarithm(LIL) for the estimator of  $\beta$ , and the optimal convergence rate for the estimator of  $g(\cdot)$ . However, the computational burden of solving the minimization problem is very high when the dimension of explanatory variable vector is large.

In this paper, we extend the EFM due to Cui, Härdle and Zhu (2011) to the missing response case for estimating both  $\beta$  and  $g(\cdot)$  in model (1.2). That is, we consider the case where some  $Y$  -values may be missing and  $X$  is observed completely. The data we observe are

$$\{(Y_i, \delta_i, X_i)\}_{i=1}^n$$

where  $\delta_i = 0$  if  $Y_i$  is missing, otherwise  $\delta_i = 1$ . Throughout this paper, it is assumed that  $Y$  is missing at random (MAR). The MAR assumption implies that  $\delta$  and  $Y$  are conditionally independent given  $X$ . That is,  $P(\delta = 1|Y, X) = P(\delta = 1|X)$ . MAR is a common assumption for statistical analysis with missing data and is reasonable in many practical situations, see Little and Rubin (2002).

In this paper, we develop a three-steps estimating approach for estimating both  $\beta$  and

$g(\cdot)$  by extending the EFM due to Cui, Härdle and Zhu (2011) to the missing response problem. Unlike the two-step estimating approach of Cui, Härdle and Zhu (2011), the three-steps estimating approach can define an estimator of  $g(\cdot)$ . For the estimating approach, the estimating function system only involves one-dimensional nonparametric smoothers, thereby avoiding the data sparsity problem caused by high dimensionality. Firstly, unlike the method proposed by Wang et al.(2010) for the special case of the ESIM where the minimization is difficult to implement when  $d$  is large, our method is easy to implement. Secondly, unlike the method proposed by Wang et al.(2010) where the methodology can only be applied to the case of homogeneous errors, our method can apply to the case of heterogeneous errors. Hence, the proposed methodology based on model (1.2) has more wide application and much more flexible framework. Cui, Härdle and Zhu (2011) define the estimator of  $\beta$  only when data are observed fully. However, we define the estimators of both  $\beta$  and  $g(\cdot)$  and investigate their asymptotic properties with data missing.

This paper is organized as follows. In Section 2, we describe the estimating procedures. In Section 3, we establish the asymptotic theory for the proposed procedure. Some simulation studies are provided in Section 4. In Section 5, we analyze a real data set to illustrate the proposed procedures and all proofs are included in Section 6.

## 2 Three-Step Estimation

We develop the following three-step approach to define the estimators of  $\beta$  and  $g(\cdot)$ , respectively.

Step 1: We use the nonparametric fusion-refinement (FR) approach to get the initial estimate of  $\beta$ , denoted by  $\tilde{\beta}$  with  $\|\tilde{\beta}\| = 1$ , see Ding and Wang (2011).

Step 2: Define the estimator of  $g(\cdot)$  and  $g'(\cdot)$ .

Note that under MAR, we have

$$\mu\{g(t)\} = \mathbf{E}[\delta Y | \beta^\top X = t] / \mathbf{E}[\delta | \beta^\top X = t].$$

We then may obtain an initial estimator of  $\mu\{g(t)\}$

$$\mu\{\tilde{g}(t)\} = (\sum_{j=1}^n \delta_j Y_j H_{h_n}(t - \tilde{\beta}^\top X_j)) / (\sum_{j=1}^n \delta_j H_{h_n}(t - \tilde{\beta}^\top X_j)),$$

where  $H(\cdot)$  is a kernel function with support on  $(-1, 1)$ ,  $h_n$  is a bandwidth sequence and  $H_{h_n}(\cdot) = H(\cdot/h_n)$ .

Denote by  $\alpha_0$  and  $\alpha_1$  the values of  $g(\cdot)$  and  $g'(\cdot)$  evaluating at  $\beta^\top x$ , respectively. The local linear approximation for  $g(\beta^\top X)$  in a neighborhood of  $\beta^\top x$  is  $g_0(\beta^\top X) = \alpha_0 + \alpha_1(\beta^\top X - \beta^\top x)$ . The estimators  $G(\beta^\top x) \stackrel{\text{def}}{=} (g(\beta^\top x), g'(\beta^\top x))$  are obtained by solving the kernel estimating equations:

$$\begin{aligned} & \sum_{j=1}^n K_{b_n}(\tilde{\beta}^\top X_j - \beta^\top x) \mu' \{g_0(\tilde{\beta}^\top X_j)\} V^{-1} \{g_0(\tilde{\beta}^\top X_j)\} \\ & \times [\delta_j Y_j + (1 - \delta_j) \mu \{\tilde{g}(\tilde{\beta}^\top X_j)\} - \mu \{g_0(\tilde{\beta}^\top X_j)\}] = 0, \\ & \sum_{j=1}^n (\tilde{\beta}^\top X_j - \beta^\top x) K_{b_n}(\tilde{\beta}^\top X_j - \beta^\top x) \mu' \{g_0(\tilde{\beta}^\top X_j)\} V^{-1} \{g_0(\tilde{\beta}^\top X_j)\} \\ & \times [\delta_j Y_j + (1 - \delta_j) \mu \{\tilde{g}(\tilde{\beta}^\top X_j)\} - \mu \{g_0(\tilde{\beta}^\top X_j)\}] = 0 \end{aligned} \tag{2.1}$$

where  $K_{b_n}(\cdot)$  is the symmetric kernel density function satisfying  $K_{b_n}(\cdot) = K(\cdot/b_n)$  and  $b_n$  is a bandwidth, with respect to  $\alpha_0$  and  $\alpha_1$ , yielding  $\widehat{G}(\beta^\top x) = (\widehat{g}(\beta^\top x), \widehat{g}'(\beta^\top x)) = (\widehat{\alpha}_0, \widehat{\alpha}_1)$ .

Step 3: Obtain the estimator of  $\beta$ . Similar to Cui et al (2011), by eliminating  $\beta_1$ , the parameter space  $\Theta$  can be rearranged to the form  $\Theta = \{(1 - \sum_{r=2}^d \beta_r^2)^{1/2}, \beta_2, \dots, \beta_d\}^\top : \sum_{r=2}^d \beta_r^2 < 1\}$ .

We turn to the estimation of  $\beta \in \Theta$ . First, we estimate  $\beta^{(1)} = (\beta_2, \dots, \beta_d)$ , which can be obtained by solving the following equation

$$\sum_{j=1}^n [\partial \mu \{\widehat{g}(\beta^\top X_j)\} / \partial \beta^{(1)}] V^{-1} \{\widehat{g}(\beta^\top X_j)\} [\delta_j Y_j + (1 - \delta_j) \mu \{\tilde{g}(\beta^\top X_j)\} - \mu \{\widehat{g}(\beta^\top X_j)\}] = 0. \tag{2.2}$$

The solution is defined as  $\tilde{\beta}^{(1)}$  and hence we obtain  $\tilde{\beta}$  by the transformation. Repeat Steps 2 and 3 until convergence and hence we can obtain the estimate of  $\beta^{(1)}$  and  $\beta$ ,  $\widehat{\beta}^{(1)}$  and  $\widehat{\beta}$  say, respectively.

### 3 Asymptotic theory

To establish asymptotic theory, we firstly give some notations. Let  $q_1(z, y) = \mu'(z) V^{-1}(z) \{y - \mu(z)\}$ ,  $\rho_l(z) = \{\mu'(z)\}^l V^{-1}(z)$ ,  $\pi(X) = \text{P}(\delta = 1 | X)$ . Let

$$\gamma_j = \int t^j K(t) dt \text{ and } \nu_j = \int t^j K^2(t) dt, \quad j = 1, 2, \dots$$

and  $S = \begin{pmatrix} \gamma_0 & 0 \\ 0 & \gamma_2 \end{pmatrix}$ ,  $S^* = \begin{pmatrix} \nu_0 & \nu_1 \\ \nu_1 & \nu_2 \end{pmatrix}$ . Denote by  $\beta^0 = (\beta_1^0, \beta^{(1)0\top})^\top$  the true values of  $\beta = (\beta_1, \beta^{(1)\top})^\top$ . Denote  $J = \frac{\partial \beta}{\partial \beta^{(1)}}$  be the Jacobian matrix of size  $d \times (d-1)$  with

$$J = \begin{pmatrix} -\beta^{(1)\top} / \sqrt{1 - \|\beta^{(1)}\|^2} \\ I_{d-1} \end{pmatrix}.$$

Denote  $C = (1 - \delta)J^\top \mathbf{E}\{X|\beta^\top X\} + J^\top (X - \mathbf{E}\{X|\beta^\top X\})g' \{(\beta^\top X)\}$ . Let

$$\begin{aligned} A &= \mathbf{E}[\rho_2\{g(\beta^\top X)\}C^\top C], \\ B &= \mathbf{E}[\delta\rho_2\{g(\beta^\top X)\}\sigma^2 C^\top C]. \end{aligned}$$

We are ready to present the asymptotic results of the proposed estimators. The proofs of the theorem are provided in Section 6.

**Theorem 3.1** Suppose that conditions (a) – (f) hold in Section 6, if  $nb_n^4 \rightarrow 0$ ,  $nh_n^4 \rightarrow 0$ ,  $nh_n^2/\log(1/h_n) \rightarrow \infty$  and  $nb_n^2h_n^2 \rightarrow 0$ , then

$$\sqrt{n}(\widehat{\beta}^{(1)} - \beta^{(1)0}) \xrightarrow{\mathcal{L}} \mathbf{N}_{d-1}(0, \Omega),$$

where  $\Omega = A^{-1}BA^{-1}|_{\beta^{(1)}=\beta^{(1)0}}$ .

**Remark 3.1** When  $\delta = 1$ , the asymptotic co-variance matrix reduces to that of Cui, Härdle and Zhu (2011).

To define a consistent estimator of the asymptotic variance, a natural way is first to define estimators of  $h(t) = \mathbf{E}\{X|\beta^\top X\}$  using the local linear estimate as

$$\widehat{h}(t) = \sum_{i=1}^n b_i(t)X_i / \sum_{i=1}^n b_i(t).$$

where  $b_i(t) = K_{b_n}(\widehat{\beta}^\top X_i - t)\{S_{n,2}(t) - (\widehat{\beta}^\top X_i - t)S_{n,1}(t)\}$  and  $S_{n,k}(t) = K_{b_n}(\widehat{\beta}^\top X_i - t)(\widehat{\beta}^\top X_i - t)^k$ ,  $k = 1, 2$ . Let  $\widehat{C}_i = (1 - \delta_i)J^\top \widehat{h}(\widehat{\beta}^\top X_i) + J^\top (X_i - \widehat{h}(\widehat{\beta}^\top X_i))\widehat{g}' \{(\widehat{\beta}^\top X_i)\}$ . Then the asymptotic variance  $\Omega$  can be estimated as

$$\begin{aligned} \widehat{\Omega} &= [n^{-1} \sum_{i=1}^n \rho_2\{\widehat{g}(\widehat{\beta}^\top X_i)\}\widehat{C}_i\widehat{C}_i^\top]^{-1} \\ &\quad \times \{n^{-1} \sum_{i=1}^n \delta_i q_1^2[\widehat{g}(\widehat{\beta}^\top X_i), Y_i]\widehat{C}_i\widehat{C}_i^\top\} [n^{-1} \sum_{i=1}^n \rho_2\{\widehat{g}(\widehat{\beta}^\top X_i)\}\widehat{C}_i\widehat{C}_i^\top]^{-1}. \end{aligned}$$



**Remark 3.2** If  $\mu\{g(\beta^\top X)\} = g(\beta^\top X)$ ,  $\sigma^2 V\{g(\beta^\top X)\} = \sigma^2$ , then the matrix  $\Omega$  in Theorem 3.1 reduces to

$$A^{-1}BA^{-1} = \mathbf{E}\left\{\{(1-\delta)J^\top \mathbf{E}(X|\beta^\top X) + J^\top(X - \mathbf{E}(X^\top|\beta^\top X))\}[g'(\beta^\top X)]\right\} \\ \times \left\{\{(1-\delta)J^\top \mathbf{E}(X|\beta^\top X) + J^\top(X - \mathbf{E}(X^\top|\beta^\top X))\}[g'(\beta^\top X)]\}^\top \sigma^2\right\}.$$

The asymptotic normality of  $\widehat{\beta} = (\widehat{\beta}_1, \widehat{\beta}^{(1)\top})^\top$  follows from Theorem 3.1 with a simple application of the multivariate delta-method, since  $\widehat{\beta}_1 = \sqrt{1 - \|\widehat{\beta}^{(1)\top}\|}$ .

**Corollary 3.2** Under the conditions of Theorem 3.1, we have

$$\sqrt{n}(\widehat{\beta} - \beta^0) \xrightarrow{\mathcal{L}} \mathbf{N}_{d-1}(0, \Lambda_{\beta^0}),$$

where  $\Lambda_{\beta^0} = J\Omega J^\top|_{\beta=\beta^0}$ .

Using the plug in method, the asymptotic variance  $\Lambda_{\beta^0}$  can be estimated by  $\widehat{J}\widehat{\Omega}\widehat{J}^\top$ , where  $\widehat{J}$  is  $J$  with  $\beta$  replaced by  $\widehat{\beta}$ .

**Theorem 3.3** Suppose that conditions of Theorem 3.1 hold, we have

$$\sqrt{nb_n}(\widehat{g}(\widehat{\beta}^\top X) - g(\beta^{0\top} X) - \frac{\mu^{(2)}\{g(\beta^\top x)\}}{2}e_1 S^{-1} U b_n^2) \xrightarrow{\mathcal{L}} \mathbf{N}(0, \Lambda_1),$$

where  $U = (\mu_2, \mu_3)$ ,  $e_1 = (1, 0)$  and  $\Lambda_1 = \frac{\sigma^2}{\pi(x)\rho_2\{g(\beta^\top x)\}f_{\beta^\top x}(\beta^\top x)}e_1 S^{-1} S^* S^{-1}$ .

Let  $Z_i^* = (1, \frac{\widehat{\beta}^\top X_i - t}{b_n})^\top$ . The asymptotic variance  $\Lambda_1$  can be estimated by

$$\widehat{\Lambda}_1 = e_1 [n^{-1} \sum_{i=1}^n \delta_i q_2 [\mu\{\widehat{g}(\widehat{\beta}^\top X_i)\}, Y_i] Z_i^* Z_i^{*\top} K_{b_n}(\widehat{\beta}^\top X_i - t)]^{-1} \\ \times [n^{-1} \sum_{i=1}^n \delta_i q_1^2 [\mu\{\widehat{g}(\widehat{\beta}^\top X_i)\}, Y_i] Z_i^* Z_i^{*\top} K_{b_n}^2(\widehat{\beta}^\top X_i - t)] \\ \times [n^{-1} \sum_{i=1}^n \delta_i q_2 [\mu\{\widehat{g}(\widehat{\beta}^\top X_i)\}, Y_i] Z_i^* Z_i^{*\top} K_{b_n}(\widehat{\beta}^\top X_i - t)]^{-1}$$

**Remark 3.2.** The choice of bandwidth is a very important topic in nonparametric regression estimation. For the semiparametric problem considered here, the  $n^{1/2}$ -rate asymptotic normality of the proposed estimators of the global parameter vectors  $\beta$  implies that a proper choice of the bandwidths depends only on the second order term of the mean square errors of these estimators. Therefore the selection of bandwidths might be not so critical if one is only interested in estimation of  $\beta$ . However, the estimators of  $g(\cdot)$  depend the choice of the bandwidth heavily. The popular cross-validation method such as cross-validation, gen-

eralized cross-validation (GCV) and the rule of thumb can be used to select the optimal bandwidth for the estimator of  $g(\cdot)$ . Here, we recommend using GCV to determine the optimal bandwidth.

## 4 Simulation studies

We conducted some Monte Carlo simulation studies to evaluate the performance of the proposed estimators for finite samples.

In our simulation, kernel functions  $H(\cdot)$  and  $K(\cdot)$  were taken as Gaussian kernel. As pointed out in Remark 3.2, the selection of bandwidths is not so critical if one is only interested in estimation of the parametric part. In the following simulation study, the bandwidths were directly taken to be  $h_n = n^{-2/5}$  and  $b_n = n^{-1/3}$  which satisfy the conditions in the above theorems.

**Example 1.** To compare the proposed method with Wang *et al* (2010), we first consider the following *simple* single-index model

$$Y = (X^\top \beta)^2 + \varepsilon, \quad (4.1)$$

where  $X$  is generated from  $\mathbf{N}_d(2, I)$  for  $d = 50$ ,  $\varepsilon \sim \mathbf{N}(0, 0.2)$ , the true parameter is  $\beta = (2/\sqrt{5}, 1/\sqrt{5}, 0, \dots, 0)$ . Take the missing mechanism:

$$\text{logit}\{\mathbf{P}(\delta = 1|Y, X)\} = \gamma^\top X + c_0, \quad (4.2)$$

where  $\text{logit}(a) = \log\{a/(1-a)\}$ ,  $\gamma = (\sqrt{2}/4, \dots, \sqrt{2}/4, 0, c_1)^\top / \sqrt{1+c_1^2}$ ,  $c_0$  is a constant to control missing proportion and  $c_1$  is a constant to control the distance between  $\gamma$  and  $\beta$ . The number of replications is 500. The size of the sample was taken to be  $n=100, 200$  and  $400$ , respectively.

The proposed estimator  $\hat{\beta}$  is compared to  $\hat{\beta}^{wang}$  of Wang, et al (2010) and the complete case (CC) estimator (denoted by  $\hat{\beta}^{cc}$ ), ignoring the missing data. We compute the average absolute bias (AB) which is defined by

$$\text{AB} = \frac{1}{500} \sum_{i=1}^{500} \left( \frac{1}{d} \sum_{s=1}^d |\hat{\beta}_{n,s}^i - \beta_s| \right),$$

where  $\widehat{\beta}_{n,s}^i$  is the  $s$ th component of  $\widehat{\beta}_n^i$  and  $\widehat{\beta}_n^i$  is one of  $\widehat{\beta}$ ,  $\widehat{\beta}^{wang}$  and  $\widehat{\beta}^{cc}$  at the  $i$ th run. We also compute the square root of the trace of the standard covariance matrix (SRTSC) which is defined by

$$\text{SRTSC} = \sqrt{\frac{1}{499} \sum_{i=1}^{500} \left\{ \frac{1}{d} (\widehat{\beta}_n^i - \bar{\beta}) (\widehat{\beta}_n^i - \bar{\beta})^\top \right\}},$$

where  $\bar{\beta} = \frac{1}{500} \sum_{i=1}^{500} \widehat{\beta}_n^i$ . The results of AB and SRTSC for  $\widehat{\beta}$ ,  $\widehat{\beta}^{wang}$ ,  $\widehat{\beta}^{cc}$  with about 25% and 50% missing proportions are reported in Table 1.

*insert Table 1 about here*

Several observations can be made from Tables 1. Firstly, we can see that AB and SRTSC of all estimators decrease as the sample size increases as expected. Secondly, we also see that  $\widehat{\beta}$  clearer outperforms  $\widehat{\beta}^{wang}$  and  $\widehat{\beta}^{cc}$  in terms of AB and SRTSC. This shows that the proposed method for the simple single index model improves the method due to Wang, et al (2010) although the proposed methods are suggested for the *extended* single index model. AB and SRTSC increase with the missing rate increasing for all the estimators.

**Example 2.** In this study, we consider the following the extended single index model

$$\begin{aligned} \text{E}(Y|X) &= \exp\{g(\beta^\top X)\}, & g(\beta^\top X) &= \sin(X^\top \beta) \\ \text{Var}(Y|X) &= \sigma^2, & \sigma &= 0.2. \end{aligned} \tag{4.3}$$

The true parameter is  $\beta = (2/\sqrt{5}, 1/\sqrt{5}, 0, \dots, 0)$ ,  $X$  is generated from  $\text{N}_d(2, I)$  for  $d = 50$ ,  $\varepsilon \sim \text{N}(0, 0.04)$  and the missing mechanism follows the model (4.2). We calculated AB and SRTSC for  $\widehat{\beta}$  and  $\widehat{\beta}^{cc}$  where  $\mu(\cdot) = \exp(\cdot)$  in (1.2). At the same time, AB and SRTSC for  $\widehat{\beta}^{wang}$  are also computed where we treated model (4.3) as a simply single index model. For each sample size of  $n=100, 200$  and  $400, 500$  replications were calculated. The simulation results are summarized in Tables 2.

*insert Table 2 about here*

In this setting, we also compare the AB and SRTSC of  $\widehat{\beta}$ ,  $\widehat{\beta}^{cc}$  and  $\widehat{\beta}^{wang}$ . From Table 2, the similar observations to Example 1 can be found except that  $\widehat{\beta}$  have more obviously

advantage than  $\hat{\beta}^{cc}$ . This shows that the proposed method is more attractive than  $\hat{\beta}^{cc}$  for the extended single index model.

**Example 3.** To illustrate the adaptivity of our algorithm to heterogeneous errors, we consider model (4.1),

$$\begin{aligned} E(Y|X) &= \{g(\beta^\top X)\}^2, & g(\beta^\top X) &= X^\top \beta \\ \text{Var}(Y|X) &= \sigma^2 \exp\{\frac{\sqrt{5}}{7}g(\beta^\top X)\}, & \sigma^2 &= 1. \end{aligned} \quad (4.4)$$

where the true parameter is  $\beta = (2/\sqrt{5}, 1/\sqrt{5}, 0, \dots, 0)$ ,  $X$  is generated from  $\mathbf{N}_d(2, I)$  for  $d = 50$  and the missing mechanism follows the model (4.2). We calculated AB and SRTSC for  $\hat{\beta}$  and  $\hat{\beta}^{cc}$ . For each sample size of  $n=100, 200$  and  $400$ ,  $500$  replications were calculated. The simulation results are also summarized in Tables 3.

For the heteroscedastic setting,  $\hat{\beta}^{wang}$  cannot be calculated and hence we compare  $\hat{\beta}$  with  $\hat{\beta}^{cc}$  only. From Table 3, the similar observations to Example 1 can be found except that  $\hat{\beta}$  have more obviously advantage than  $\hat{\beta}^{cc}$ .

*insert Table 3 about here*

## 5 Real data analysis

ACTG 175 data have been studied by some authors (see, e.g., Hammer et al., 1996; Davidian et al., 2005; Ding and Wang 2011; Hu et al., 2010). In an HIV clinical trial, 2139 HIV positive patients were involved. The patients were randomized into four arms to receive monotherapy (ZDV) or combined therapy (ADV+didanosine, ZDV+zalcitabine, and didanosine). We apply the proposed methods to this data set. The response  $Y = I(\text{"the CD4 count at } 96 \pm 5 \text{ weeks"} \geq 300)$ . The predictors  $X$  are six baseline characteristics: age, weight, CD4 counts at baseline and  $20 \pm 5$  weeks, CD8 counts at baseline and  $20 \pm 5$  weeks. Let  $T$  denote the received therapy, i.e.,  $T = 1$  if receiving combined therapy, and  $T = 0$  otherwise. Among the 746 patients, there were 473 patients with observations in  $Y$ , including 105 patients receiving monotherapy and 368 patients receiving other therapies, and due to death and dropout there were 273 patients with missing observations in  $Y$ , including 74 patients with  $T = 0$  and 199 patients with  $T = 1$ . All the patients had predictors  $X$  observed.

The single-index model will be used to model the relationship between the CD4 count at  $96 \pm 5$  weeks and the relevant 6 predictors  $X = (X_1, \dots, X_6)^\top$ :

$$P(\text{"the CD4 count at } 96 \pm 5 \text{ weeks"} \geq 300 | X) = \exp\{g(\beta^\top X)\} / [1 + \exp\{g(\beta^\top X)\}], \quad (5.1)$$

where  $\beta = (\beta_1, \dots, \beta_6)^\top$ . We first focused on the subset of data labeled by  $T = 0$ . we can obtain the estimator  $\hat{\beta}$  by the proposed method. The estimator  $\hat{\beta}$  is  $(0.1289, 0.9195, 0.0161, 0.3546, -0.0677)^\top$ . For the subset of data labeled by  $T = 1$ , we can also obtain  $\hat{\beta} = (0.1927, -0.9792, -0.0058, -0.0079, 0.0582, 0.0244)^\top$ .

As one can see from two estimates, 'weight' has the larger positive influence when patients receive combined therapy. On the contrary, there is a negative influence when patients receive monotherapy for proposing method. 'Age' has the positive influence in the two setting, this is true because resistance become more and more weak with increasing age.

We also plot the scatter plot of the estimated single index  $\hat{g}(\hat{\beta}^\top X)$  against  $\hat{\beta}^\top X$  in the setting of  $T = 0$  and  $T = 1$ , respectively. The scatter plot suggests a curvature relationship between the response and covariates. The pattern is displayed in Fig 1 and Fig 2, respectively.

*insert Figure 1 about here*

*insert Figure 2 about here*

It is seen that there is a nonlinear trend. Therefore, using the model (5.1) in the regression is perhaps more appropriate than using the internally linear model

$$P(\text{"the CD4 count at } 96 \pm 5 \text{ weeks"} \geq 300 | X) = \exp(\beta^\top X) / \{1 + \exp(\beta^\top X)\}. \quad (5.2)$$

## 6 Technical Assumptions and Proofs

### 6.1 Technical Assumptions

In order to prove the asymptotic normality of the estimators, we first introduce some regularity conditions.

(a)  $\mu(\cdot)$ ,  $V(\cdot)$  and  $g(\cdot)$  have two bounded and continuous derivatives.  $V(\cdot)$  is uniformly bounded and bounded away from 0.

(b) Assume that  $\partial q(z, y)/\partial z < 0$  for  $z \in \mathbb{R}$  and  $y$  in the range of the response variable.

(c) Define the block partition of matrix  $\Omega$  as follows:

$$\Omega = \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix}$$

where  $\Omega_{11}$  is a positive constant,  $\Omega_{12}$  is a  $(d-1)$ -dimensional row vector,  $\Omega_{21}$  is a  $(d-1)$ -dimensional column vector and  $\Omega_{22}$  is a  $(d-1) \times (d-1)$  nonnegative definite matrix. The largest eigenvalues of  $\Omega_{22}$  is bounded away from infinity.

(d) The density function of  $X$  has a continuous second derivative on its support  $A$ . The density function  $f_{\beta^\top X}(\beta^\top X)$  of random variable  $\beta^\top X$  is bounded away from 0 on  $T_\beta$  and satisfies the Lipschitz condition of order 1 on  $T_\beta$ , where  $T_\beta = \{\beta^\top X : X \in T\}$  and  $T$  is the compact support set of  $X$ .

(e) The kernel  $K(\cdot)$  is a bounded and symmetric density function with a bounded derivative, and satisfies

$$\int_{-\infty}^{+\infty} |t|^2 K(t) dt < \infty,$$

$H(\cdot)$  is a bounded kernel function of order 2 with bounded support.

(f)  $\pi(\cdot) > 0$  and  $\mu(\cdot) \neq 0$ .

## 6.2 Proofs of Theorems

In order to prove the asymptotic normality of the estimators, we first introduce several lemmas.

**Lemma 1.** Let  $(x_1; y_1), \dots, (x_n; y_n)$  be i.i.d random vectors, where the  $y_i$ 's are univariate random variables. Assume that  $\mathbf{E}|y|^r < \infty$  and  $\sup_x \int |y|^r p(x, y) < \infty$ , where  $p$  denotes the joint density of  $(x, y)$ . Let  $K$  be a bounded positive function with bounded support, satisfying the Lipschitz condition. Then

$$\sup_x |n^{-1} \sum_{i=1}^n \{K_h(x_i - x)y_i - \mathbf{E}[K_h(x_i - x)y_i]\}| = \mathcal{O}_p\left\{\left(\frac{-\log h}{nh}\right)^{1/2}\right\}$$

provided that  $n^{2\varepsilon-1}h \rightarrow \infty$  for some  $\varepsilon < 1 - r^{-1}$ .

Lemma 1 is a direct result of Mack and Silverman (1982), which is also cited by many papers on kernel method. In what follows, we give an important lemma which derives the asymptotic structure of  $g$  which will be used to get the asymptotic property of parameters.

**Lemma 2.** Suppose that conditions of Theorem 3.1 hold,  $\widehat{G}(\beta^\top x)$  and  $G(\beta^\top x)$  are defined in Step 2 of Section 2, then

$$\sqrt{nb_n}(H_2\{\widehat{G}(\beta^\top x) - G(\beta^\top x) - \lambda(x)\}) \xrightarrow{\mathcal{L}} \mathbf{N}(0, \frac{\sigma^2}{\pi(x)\rho_2\{g(\beta^\top x)\}f_{\beta^\top x}(\beta^\top x)}S^{-1}S^*S^{-1})$$

where  $H_2 = \text{diag}(1, b_n)$  and

$$\lambda(x) = \frac{\mu^{(2)}\{g(\beta^\top x)\}}{2}S^{-1}Ub_n^2.$$

**Proof** Let  $\widehat{Y}_i^* = \delta_i Y_i + (1 - \delta_i)\mu(\widehat{g}\{\beta^\top X_i\})$  and  $Y_i^* = \delta_i Y_i + (1 - \delta_i)\mu(g\{\beta^\top X_i\})$ . Note that solution  $\widehat{G}(\beta^\top x)$  of the estimating equation defined in (2.1) can be obtained by maximizing the quasi-likelihood:

$$\ell(\alpha_0, \alpha_1) \stackrel{\text{def}}{=} \sum_{i=1}^n Q[\mu\{\alpha_0 + \alpha_1(\beta^\top X_i - \beta^\top x)\}, \widehat{Y}_i^*]K_{b_n}(\beta^\top X_i - \beta^\top x)$$

with respect to  $(\alpha_0, \alpha_1)$ , where  $Q(\mu, y) = \int_y^\mu \frac{y-s}{V\{\mu^{-1}(s)\}}$  and  $\mu^{-1}(\cdot)$  is the inverse function of  $\mu(\cdot)$ .

Let  $q_l(z, y) = \frac{\partial^l}{\partial z^l} Q[\mu(z), y]$ ,  $l = 1, 2, 3$ , then  $q_1[z, y] = \{y - \mu(z)\}\rho_1(z)$  and  $q_2[z, y] = \{y - \mu(z)\}\rho_1'(z) - \rho_2(z)$ , where  $\rho_l(z) = \mu^{(l)}\{z\}V^{-1}\{z\}$ .

Denote  $\widehat{\alpha}^* = \sqrt{nb_n}H_2(\widehat{g}(\beta^\top x) - g(\beta^\top x), \widehat{g}'(\beta^\top x) - g'(\beta^\top x))$ ,  $\alpha^* = \sqrt{nb_n}H_2(\alpha_0 - g(\beta^\top x), \alpha_1 - g'(\beta^\top x))$  and  $\bar{\alpha}_i(\beta^\top x) = \alpha_0 + \alpha_1(\beta^\top X_i - \beta^\top x)$  and  $X_i^* = (1, \frac{\beta^\top X_i - \beta^\top x}{b_n})$ . Then  $\widehat{\alpha}^*$  is the solution of the following normalized function

$$\ell(\alpha^*) = b_n \sum_{i=1}^n \{Q[\mu\{\bar{\alpha}_i(\beta^\top x) + \frac{1}{\sqrt{nb_n}}\alpha^{*\top} X_i^*\}, \widehat{Y}_i^*] - Q[\mu\{\bar{\alpha}_i(\beta^\top x)\}, \widehat{Y}_i^*]\}K_{b_n}(\beta^\top X_i - \beta^\top x).$$

By Taylor expansion, we have

$$\ell(\alpha^*) = V_n^\top \alpha^* + \frac{1}{2} \alpha^{*\top} B_n \alpha^* (1 + o_p(1)), \quad (6.1)$$

where

$$\begin{aligned} V_n &= \sqrt{\frac{b_n}{n}} \sum_{i=1}^n q_1[\bar{\alpha}_i(\beta^\top x), \hat{Y}_i^*] X_i^* K_{b_n}(\beta^\top X_i - \beta^\top x), \\ B_n &= \frac{1}{n} \sum_{i=1}^n q_2[\bar{\alpha}_i(\beta^\top x), \hat{Y}_i^*] X_i^* X_i^{*\top} K_{b_n}(\beta^\top X_i - \beta^\top x) \end{aligned}$$

According to the definition of  $q_2[x, y]$ , we have

$$\begin{aligned} & q_2(\bar{\alpha}_i, \hat{Y}_i^*) - q_2(\bar{\alpha}_i, Y_i^*) \\ &= \rho_1(\bar{\alpha}_i)(\hat{Y}_i^* - Y_i^*) \\ &= (1 - \delta_i) \rho_1(\bar{\alpha}_i) (\mu\{\tilde{g}(\beta^\top X_i)\} - \mu\{g(\beta^\top X_i)\}). \end{aligned} \quad (6.2)$$

It can be observed

$$\begin{aligned} & \mu\{\tilde{g}(\beta^\top X_i)\} - \mu\{g(\beta^\top X_i)\} \\ &= \{\sum_{j=1}^n \delta_j H_{h_n}(\beta^\top X_i - \beta^\top X_j)\}^{-1} \{\sum_{j=1}^n \delta_j Y_j H_{h_n}(\beta^\top X_i - \beta^\top X_j)\} - \mu\{g(\beta^\top X_i)\} \\ &= [\sum_{j=1}^n \delta_j Y_j H_{h_n}(\beta^\top X_i - \beta^\top X_j) - \sum_{j=1}^n \delta_j \mu\{g(\beta^\top X_i)\} H_{h_n}(\beta^\top X_i - \beta^\top X_j)] \\ & \quad \times \{\sum_{j=1}^n \delta_j H_{h_n}(\beta^\top X_i - \beta^\top X_j)\}^{-1} \\ &= \sum_{j=1}^n \delta_j [Y_j - \mu\{g(\beta^\top X_j)\}] H_{h_n}(\beta^\top X_i - \beta^\top X_j) \{\sum_{j=1}^n \delta_j H_{h_n}(\beta^\top X_i - \beta^\top X_j)\}^{-1} \\ & \quad - \sum_{j=1}^n \delta_j [\mu\{g(\beta^\top X_j)\} - \mu\{g(\beta^\top X_i)\}] H_{h_n}(\beta^\top X_i - \beta^\top X_j) \{\sum_{j=1}^n \delta_j H_{h_n}(\beta^\top X_i - \beta^\top X_j)\}^{-1} \\ &= \sum_{j=1}^n \delta_j q_1[\mu\{g(\beta^\top X_j)\}, Y_j] \{\rho_1\{g(\beta^\top X_j)\}\}^{-1} H_{h_n}(\beta^\top X_i - \beta^\top X_j) \\ & \quad \times \{nh_n \pi(X_i) f_{\beta^\top X_i}(\beta^\top X_i)\}^{-1} + o_p(1). \end{aligned} \quad (6.3)$$



By (6.2) and (6.3), we have

$$\begin{aligned}
B_n &= \frac{1}{n} \sum_{i=1}^n q_2 \{ \bar{\alpha}_i(\beta^\top x), Y_i^* \} X_i^* X_i^{*\top} K_{b_n}(\beta^\top X_i - \beta^\top x) + \mathcal{O}_p(1) \\
&= E[q_2[\bar{\alpha}_1(\beta^\top x), Y_1^*] X_1^* X_1^{*\top} K_{b_n}(\beta^\top X_1 - \beta^\top x)] + \mathcal{O}_p(1) \\
&= -\rho_2 \{ g(\beta^\top x) \} f_{\beta^\top x}(\beta^\top x) S + \mathcal{O}_p(1).
\end{aligned} \tag{6.4}$$

For  $V_n$ , we have

$$\begin{aligned}
V_n &= \sqrt{\frac{b_n}{n}} \sum_{i=1}^n q_1[\bar{\alpha}_i(\beta^\top x), Y_i^*] X_i^* K_{b_n}(\beta^\top X_i - \beta^\top x) \\
&\quad + \sqrt{\frac{b_n}{n}} \sum_{i=1}^n [q_1(\bar{\alpha}_i(\beta^\top x), \hat{Y}_i^*) - q_1(\bar{\alpha}_i(\beta^\top x), Y_i^*)] X_i^* K_{b_n}(\beta^\top X_i - \beta^\top x) \\
&\stackrel{\text{def}}{=} V_{n1} + V_{n2}.
\end{aligned} \tag{6.5}$$

For  $V_{n2}$ , by Taylor expansion, we have

$$\begin{aligned}
V_{n2} &= \sqrt{\frac{b_n}{n}} \sum_{i=1}^n (1 - \delta_i) \rho_1 \{ g(\bar{\alpha}_i(\beta^\top x)) \} [\mu \{ \tilde{g}(\beta^\top X_i) \} - \mu \{ g(\beta^\top X_i) \}] X_i^* K_{b_n}(\beta^\top X_i - \beta^\top x) \\
&\quad + \mathcal{O}_p(\sqrt{nb_n}) |\mu \{ \tilde{g}(\beta^\top X_i) \} - \mu \{ g(\beta^\top X_i) \}|^2 \\
&= \sqrt{\frac{b_n}{n}} \sum_{i=1}^n \frac{(1 - \delta_i) \rho_1 \{ g(\bar{\alpha}_i(\beta^\top x)) \}}{nh_n f_{\beta^\top X_i}(\beta^\top X_i) \pi(X_i)} \\
&\quad \times \sum_{j=1}^n \delta_j \frac{q_1[\mu \{ g(\beta^\top X_j) \}, Y_j]}{\rho_1 \{ g(\beta^\top X_j) \}} X_j^* K_{h_n}(\beta^\top X_j - \beta^\top X_i) K_{b_n}(\beta^\top X_i - \beta^\top x) + \mathcal{O}_p(1) \\
&= \sqrt{\frac{b_n}{n}} \sum_{i=1}^n \delta_i q_1[\mu \{ g(\beta^\top X_i) \}, Y_i] \frac{(1 - \pi(X_i) \rho_1 \{ g(\bar{\alpha}_i(\beta^\top x)) \})}{\pi(X_i) \rho_1 \{ g(\beta^\top X_i) \}} X_i^* K_{b_n}(\beta^\top X_i - \beta^\top x) + \mathcal{O}_p(1) \\
&\equiv T_n + \mathcal{O}_p(1).
\end{aligned} \tag{6.6}$$

By (6.1), (6.4), (6.5) and (6.6), we have

$$\ell(\alpha^*) = (V_{n1} + T_n)^\top \alpha^* - \alpha^* B \alpha^* / 2 + \mathcal{O}_p(1). \tag{6.7}$$

According to quadratic approximation lemma, we obtain

$$\hat{\alpha}^* = B^{-1}(V_{n1} + T_n) + \mathcal{O}_p(1). \tag{6.8}$$

It is easy to show that

$$\begin{aligned}
& \mathbf{E}(V_{n1} + T_n) \\
&= \sqrt{\frac{b_n}{n}} \mathbf{E}[q_1\{\bar{\alpha}_1(\beta^\top x), Y_1^*\} X_1^* K_{b_n}(\beta^\top X_1 - \beta^\top x)] \\
&\quad + \sqrt{\frac{b_n}{n}} \mathbf{E}[\delta_1 q_1[\mu\{g(\beta^\top X_1)\}, Y_1] \frac{(1-\pi(X_1)\rho_1\{g(\bar{\alpha}_1(\beta^\top x))\})}{\pi(X_1\rho_1\{g(\beta^\top x)\})} X_1^* K_{b_n}(\beta^\top X_1 - \beta^\top x)] \\
&= \sqrt{\frac{b_n}{n}} \frac{\mu^{(2)}\{g(\beta^\top x)\} b_n^2}{2} \rho_2\{g(\beta^\top x)\} f_{\beta^\top x}(\beta^\top x) U\{1 + o(1)\}.
\end{aligned}$$

and

$$\begin{aligned}
& \mathbf{Var}(V_{n1} + T_n) \\
&= b_n \mathbf{Var}[q_1\{\bar{\alpha}_1(\beta^\top x), Y_1^*\} X_1^* K_{b_n}(\beta^\top X_1 - \beta^\top x)] \\
&\quad + b_n \mathbf{Var}[\delta_1 q_1[\mu\{g(\beta^\top X_1)\}, Y_1] \frac{(1-\pi(X_1)\rho_1\{g(\bar{\alpha}_1(\beta^\top x))\})}{\pi(X_1\rho_1\{g(\beta^\top x)\})} X_1^* K_{b_n}(\beta^\top X_1 - \beta^\top x)] \\
&= \frac{\sigma^2 \rho_2\{g(\beta^\top x)\} f_{\beta^\top x}(\beta^\top x)}{\pi(x)} S^* \{1 + o(1)\}.
\end{aligned} \tag{6.9}$$

Since  $V_{n1} + V_{n2}$  is a sum of i.i.d. random vectors and Liapounov's condition is satisfied, thus proof is completed.

**Lemma 3.** Suppose that conditions of Theorem 3.1 hold, then

$$\frac{\partial \hat{g}(\beta^\top x)}{\partial \beta^{(1)}} \xrightarrow{p} C,$$

where  $C$  is defined in Section 3.

Proof The first equation of (2.1) can be decomposed as

$$\begin{aligned}
0 &= n^{-1} \sum_{j=1}^n K_{b_n}(\beta^\top X_j - t) \mu' \{ \hat{\alpha}_0 + \hat{\alpha}_1(\beta^\top X_j - \beta^\top x) \} V^{-1} \{ \hat{\alpha}_0 + \hat{\alpha}_1(\beta^\top X_j - \beta^\top x) \} \\
&\quad \times [\delta_j Y_j + (1 - \delta_j) \mu \{ \tilde{g}(\beta^\top X_j) \} - \mu \{ \hat{\alpha}_0 + \hat{\alpha}_1(\beta^\top X_j - \beta^\top x) \}] \\
&= n^{-1} \sum_{j=1}^n K_{b_n}(\beta^\top X_j - t) \mu' \{ \hat{\alpha}_0 + \hat{\alpha}_1(\beta^\top X_j - \beta^\top x) \} V^{-1} \{ \hat{\alpha}_0 + \hat{\alpha}_1(\beta^\top X_j - \beta^\top x) \} \\
&\quad \times [\delta_j Y_j + (1 - \delta_j) \mu \{ g(\beta^\top X_j) \} - \mu \{ \hat{\alpha}_0 + \hat{\alpha}_1(\beta^\top X_j - \beta^\top x) \}] + o_p(1) \\
&= n^{-1} \sum_{j=1}^n \delta_j K_{b_n}(\beta^\top X_j - t) \mu' \{ \hat{\alpha}_0 + \hat{\alpha}_1(\beta^\top X_j - \beta^\top x) \} V^{-1} \{ \hat{\alpha}_0 + \hat{\alpha}_1(\beta^\top X_j - \beta^\top x) \} \\
&\quad \times [Y_j - \mu \{ \hat{\alpha}_0 + \hat{\alpha}_1(\beta^\top X_j - \beta^\top x) \}] \\
&\quad + n^{-1} \sum_{j=1}^n (1 - \delta_j) K_{b_n}(\beta^\top X_j - t) \mu' \{ \hat{\alpha}_0 + \hat{\alpha}_1(\beta^\top X_j - \beta^\top x) \} V^{-1} \{ \hat{\alpha}_0 + \hat{\alpha}_1(\beta^\top X_j - \beta^\top x) \} \\
&\quad \times [\mu \{ g(\beta^\top X_j) \} - \mu \{ \hat{\alpha}_0 + \hat{\alpha}_1(\beta^\top X_j - \beta^\top x) \}] + o_p(1) \\
&\stackrel{\text{def}}{=} D_1 + D_2 + o_p(1).
\end{aligned} \tag{6.10}$$

For  $D_1$ , taking derivatives with respect to  $\beta^{(1)}$ , we have

$$\begin{aligned}
\frac{\partial D_1}{\partial \beta^{(1)}} &= n^{-1} \sum_{j=1}^n \delta_j K'_{b_n}(\beta^\top X_j - \beta^\top x) J^\top(X_j - x) \mu' \{ \hat{\alpha}_0 + \hat{\alpha}_1(\beta^\top X_j - \beta^\top x) \} \\
&\quad \times V^{-1} \{ \hat{\alpha}_0 + \hat{\alpha}_1(\beta^\top X_j - \beta^\top x) \} [Y_j - \mu \{ \hat{\alpha}_0 + \hat{\alpha}_1(\beta^\top X_j - \beta^\top x) \}] \\
&+ n^{-1} \sum_{j=1}^n \delta_j K_{b_n}(\beta^\top X_j - \beta^\top x) \mu'' \{ \hat{\alpha}_0 + \hat{\alpha}_1(\beta^\top X_j - \beta^\top x) \} \\
&\quad \times [\frac{\partial \hat{\alpha}_0}{\partial \beta^{(1)}} + \frac{\partial \hat{\alpha}_1}{\partial \beta^{(1)}} \beta^{(1)}(\beta^\top X_j - \beta^\top x) + \hat{\alpha}_1 J^\top(X_j - x)] \\
&\quad \times V^{-1} \{ \hat{\alpha}_0 + \hat{\alpha}_1(\beta^\top X_j - \beta^\top x) \} [Y_j - \mu \{ \hat{\alpha}_0 + \hat{\alpha}_1(\beta^\top X_j - \beta^\top x) \}] \\
&+ n^{-1} \sum_{j=1}^n \delta_j K_{b_n}(\beta^\top X_j - \beta^\top x) \mu' \{ \hat{\alpha}_0 + \hat{\alpha}_1(\beta^\top X_j - \beta^\top x) \} (V^{-1})' \{ \hat{\alpha}_0 + \hat{\alpha}_1(\beta^\top X_j - \beta^\top x) \} \\
&\quad \times [\frac{\partial \hat{\alpha}_0}{\partial \beta^{(1)}} + \frac{\partial \hat{\alpha}_1}{\partial \beta^{(1)}} \beta^{(1)}(\beta^\top X_j - \beta^\top x) + \hat{\alpha}_1 J^\top(X_j - x)] \\
&\quad \times [Y_j - \mu \{ \hat{\alpha}_0 + \hat{\alpha}_1(\beta^\top X_j - \beta^\top x) \}] \\
&- n^{-1} \sum_{j=1}^n \delta_j K_{b_n}(\beta^\top X_j - \beta^\top x) [\mu' \{ \hat{\alpha}_0 + \hat{\alpha}_1(\beta^\top X_j - \beta^\top x) \}]^2 \\
&\quad \times [\frac{\partial \hat{\alpha}_0}{\partial \beta^{(1)}} + \frac{\partial \hat{\alpha}_1}{\partial \beta^{(1)}} \beta^{(1)}(\beta^\top X_j - \beta^\top x) + \hat{\alpha}_1 J^\top(X_j - x)] \\
&\quad \times V^{-1} \{ \hat{\alpha}_0 + \hat{\alpha}_1(\beta^\top X_j - \beta^\top x) \} \\
&\stackrel{\text{def}}{=} F_1 + F_2 + F_3 - F_4.
\end{aligned}$$

We decompose  $F_1$  as follows:

$$\begin{aligned}
F_1 = & n^{-1} \sum_{j=1}^n \delta_j K'_{b_n}(\beta^\top X_j - \beta^\top x) J^\top(X_j - x) \mu' \{g(\beta^{0\top} X_j)\} V^{-1} \{g(\beta^{0\top} X_j)\} [Y_j - \mu \{g(\beta^{0\top} X_j)\}] \\
& + n^{-1} \sum_{j=1}^n \delta_j K'_{b_n}(\beta^\top X_j - \beta^\top x) J^\top(X_j - x) [\mu' \{\hat{\alpha}_0 + \hat{\alpha}_1(\beta^\top X_j - \beta^\top x)\} - \mu' \{g(\beta^{0\top} X_j)\}] \\
& \quad \times V^{-1} \{g(\beta^{0\top} X_j)\} [Y_j - \mu \{g(\beta^{0\top} X_j)\}] \\
& + n^{-1} \sum_{j=1}^n \delta_j K'_{b_n}(\beta^\top X_j - \beta^\top x) J^\top(X_j - x) \mu' \{g(\beta^{0\top} X_j)\} \\
& \quad \times V^{-1} \{g(\beta^{0\top} X_j)\} [\mu \{\hat{\alpha}_0 + \hat{\alpha}_1(\beta^\top X_j - \beta^\top x)\} - \mu \{g(\beta^{0\top} X_j)\}] \\
& + n^{-1} \sum_{j=1}^n \delta_j K'_{b_n}(\beta^\top X_j - \beta^\top x) J^\top(X_j - x) [\mu' \{\hat{\alpha}_0 + \hat{\alpha}_1(\beta^\top X_j - \beta^\top x)\} - \mu' \{g(\beta^{0\top} X_j)\}] \\
& \quad \times V^{-1} \{g(\beta^{0\top} X_j)\} [\mu \{\hat{\alpha}_0 + \hat{\alpha}_1(\beta^\top X_j - \beta^\top x)\} - \mu \{g(\beta^{0\top} X_j)\}] \\
& + n^{-1} \sum_{j=1}^n \delta_j K'_{b_n}(\beta^\top X_j - \beta^\top x) J^\top(X_j - x) \mu' \{g(\beta^{0\top} X_j)\} \\
& \quad \times [V^{-1} \{\hat{\alpha}_0 + \hat{\alpha}_1(\beta^\top X_j - \beta^\top x)\} - V^{-1} \{g(\beta^{0\top} X_j)\}] [Y_j - \mu \{g(\beta^{0\top} X_j)\}] \\
& + n^{-1} \sum_{j=1}^n \delta_j K'_{b_n}(\beta^\top X_j - \beta^\top x) J^\top(X_j - x) [\mu' \{\hat{\alpha}_0 + \hat{\alpha}_1(\beta^\top X_j - \beta^\top x)\} - \mu' \{g(\beta^{0\top} X_j)\}] \\
& \quad \times [V^{-1} \{\hat{\alpha}_0 + \hat{\alpha}_1(\beta^\top X_j - \beta^\top x)\} - V^{-1} \{g(\beta^{0\top} X_j)\}] [Y_j - \mu \{g(\beta^{0\top} X_j)\}] \\
& + n^{-1} \sum_{j=1}^n \delta_j K'_{b_n}(\beta^\top X_j - \beta^\top x) J^\top(X_j - x) \mu' \{g(\beta^{0\top} X_j)\} \\
& \quad \times [V^{-1} \{\hat{\alpha}_0 + \hat{\alpha}_1(\beta^\top X_j - \beta^\top x)\} - V^{-1} \{g(\beta^{0\top} X_j)\}] \\
& \quad \times [\mu \{\hat{\alpha}_0 + \hat{\alpha}_1(\beta^\top X_j - \beta^\top x)\} - \mu \{g(\beta^{0\top} X_j)\}] \\
& + n^{-1} \sum_{j=1}^n \delta_j K'_{b_n}(\beta^\top X_j - \beta^\top x) J^\top(X_j - x) [\mu' \{\hat{\alpha}_0 + \hat{\alpha}_1(\beta^\top X_j - \beta^\top x)\} - \mu' \{g(\beta^{0\top} X_j)\}] \\
& \quad \times [V^{-1} \{\hat{\alpha}_0 + \hat{\alpha}_1(\beta^\top X_j - \beta^\top x)\} - V^{-1} \{g(\beta^{0\top} X_j)\}] \\
& \quad \times [\mu \{\hat{\alpha}_0 + \hat{\alpha}_1(\beta^\top X_j - \beta^\top x)\} - \mu \{g(\beta^{0\top} X_j)\}] \\
& \stackrel{\text{def}}{=} \sum_{i=1}^8 F_{1i}.
\end{aligned}$$

Noting that  $\int u^k K^{(1)}(u) du = 0$  when  $k$  is an even number and using the arguments similar to the proof of Theorem 5.2 in Ichimura (1993), we have  $F_{1i} = o_p(1)$  for  $k = 1, \dots, 8$ .

Similarly, we can show that  $F_2 = o_p(1)$  and  $F_3 = o_p(1)$  under Conditions (a), (d) and

(e). Further, we also can show

$$\begin{aligned}
& n^{-1} \sum_{j=1}^n \delta_j K_{b_n}(\beta^\top X_j - \beta^\top x) [\mu' \{\hat{\alpha}_0 + \hat{\alpha}_1(\beta^\top X_j - \beta^\top x)\}]^2 \\
& \quad \times \frac{\partial \hat{\alpha}_1}{\partial \beta^{(1)}}(\beta^\top X_j - \beta^\top x) V^{-1} \{\hat{\alpha}_0 + \hat{\alpha}_1(\beta^\top X_j - \beta^\top x)\} \\
& = \mathcal{O}_p(1).
\end{aligned}$$

According to Lemma 1, we obtain

$$\begin{aligned}
& n^{-1} \sum_{j=1}^n \delta_j K_{b_n}(\beta^\top X_j - \beta^\top x) [\mu' \{\hat{\alpha}_0 + \hat{\alpha}_1(\beta^\top X_j - \beta^\top x)\}]^2 V^{-1} \{\hat{\alpha}_0 + \hat{\alpha}_1(\beta^\top X_j - \beta^\top x)\} \\
& = \mathbf{E} \{ \delta [\mu' \{g(\beta^\top X)\}]^2 V^{-1} \{g(\beta^\top X)\} | \beta^\top X = u \} f(u) \{1 + \mathcal{O}_p(1)\}.
\end{aligned}$$

and

$$\begin{aligned}
& n^{-1} \sum_{j=1}^n \delta_j K_{b_n}(\beta^\top X_j - \beta^\top x) [\mu' \{\hat{\alpha}_0 + \hat{\alpha}_1(\beta^\top X_j - \beta^\top x)\}]^2 J^\top X_j V^{-1} \{\hat{\alpha}_0 + \hat{\alpha}_1(\beta^\top X_j - \beta^\top x)\} \\
& = \mathbf{E} \{ \delta [\mu' \{g(\beta^\top X)\}]^2 V^{-1} \{g(\beta^\top X)\} J^\top X | \beta^\top X = u \} f(u) \{1 + \mathcal{O}_p(1)\}.
\end{aligned}$$

Then, we have

$$\begin{aligned}
D_1 = & - \mathbf{E} \{ \delta [\mu' \{g(\beta^\top X)\}]^2 V^{-1} \{g(\beta^\top X)\} | \beta^\top X = u \} f(u) \frac{\partial \hat{\alpha}_0}{\partial \beta^{(1)}} \\
& - \mathbf{E} \{ \delta [\mu' \{g(\beta^\top X)\}]^2 V^{-1} \{g(\beta^\top X)\} J^\top X | \beta^\top X = u \} f(u) \hat{\alpha}_1 \\
& + \mathbf{E} \{ \delta [\mu' \{g(\beta^\top X)\}]^2 V^{-1} \{g(\beta^\top X)\} J^\top x | \beta^\top X = u \} f(u) \hat{\alpha}_1 + \mathcal{O}_p(1).
\end{aligned} \tag{6.11}$$

For  $D_2$ , similarly taking derivatives with respect to  $\beta^{(1)}$ , we have

$$\begin{aligned}
\frac{\partial D_2}{\partial \beta^{(1)}} = & n^{-1} \sum_{j=1}^n (1 - \delta_j) K'_{b_n}(\beta^\top X_j - \beta^\top x) J^\top(X_j - x) \mu' \{ \hat{\alpha}_0 + \hat{\alpha}_1(\beta^\top X_j - \beta^\top x) \} \\
& \times V^{-1} \{ \hat{\alpha}_0 + \hat{\alpha}_1(\beta^\top X_j - \beta^\top x) \} [ \mu \{ g(\beta^\top X_j) \} - \mu \{ \hat{\alpha}_0 + \hat{\alpha}_1(\beta^\top X_j - \beta^\top x) \} ] \\
& + n^{-1} \sum_{j=1}^n (1 - \delta_j) K_{b_n}(\beta^\top X_j - \beta^\top x) \mu'' \{ \hat{\alpha}_0 + \hat{\alpha}_1(\beta^\top X_j - \beta^\top x) \} \\
& \times [ \frac{\partial \hat{\alpha}_0}{\partial \beta^{(1)}} + \frac{\partial \hat{\alpha}_1}{\partial \beta^{(1)}} \beta^{(1)} (\beta^\top X_j - \beta^\top x) + \hat{\alpha}_1 J^\top(X_j - x) ] \\
& \times V^{-1} \{ \hat{\alpha}_0 + \hat{\alpha}_1(\beta^\top X_j - \beta^\top x) \} [ \mu \{ g(\beta^\top X_j) \} - \mu \{ \hat{\alpha}_0 + \hat{\alpha}_1(\beta^\top X_j - \beta^\top x) \} ] \\
& + n^{-1} \sum_{j=1}^n (1 - \delta_j) K_{b_n}(\beta^\top X_j - \beta^\top x) \mu' \{ \hat{\alpha}_0 + \hat{\alpha}_1(\beta^\top X_j - \beta^\top x) \} (V^{-1})' \\
& \times \{ \hat{\alpha}_0 + \hat{\alpha}_1(\beta^\top X_j - \beta^\top x) \} [ \frac{\partial \hat{\alpha}_0}{\partial \beta^{(1)}} + \frac{\partial \hat{\alpha}_1}{\partial \beta^{(1)}} \beta^{(1)} (\beta^\top X_j - \beta^\top x) + \hat{\alpha}_1 J^\top(X_j - x) ] \\
& \times [ \mu \{ g(\beta^\top X_j) \} - \mu \{ \hat{\alpha}_0 + \hat{\alpha}_1(\beta^\top X_j - \beta^\top x) \} ] \\
& - n^{-1} \sum_{j=1}^n (1 - \delta_j) K_{b_n}(\beta^\top X_j - \beta^\top x) [ \mu' \{ \hat{\alpha}_0 + \hat{\alpha}_1(\beta^\top X_j - \beta^\top x) \} ]^2 \\
& \times [ \frac{\partial \hat{\alpha}_0}{\partial \beta^{(1)}} + \frac{\partial \hat{\alpha}_1}{\partial \beta^{(1)}} \beta^{(1)} (\beta^\top X_j - \beta^\top x) + \hat{\alpha}_1 J^\top(X_j - x) ] V^{-1} \{ \hat{\alpha}_0 + \hat{\alpha}_1(\beta^\top X_j - \beta^\top x) \} \\
& + n^{-1} \sum_{j=1}^n (1 - \delta_j) K_{b_n}(\beta^\top X_j - \beta^\top x) \mu' \{ \hat{\alpha}_0 + \hat{\alpha}_1(\beta^\top X_j - \beta^\top x) \} \\
& \times \mu' \{ g(\beta^\top X_j) \} J^\top X_j V^{-1} \{ \hat{\alpha}_0 + \hat{\alpha}_1(\beta^\top X_j - \beta^\top x) \} \\
\stackrel{\text{def}}{=} & R_1 + R_2 + R_3 - R_4 + R_5.
\end{aligned}$$

Using the arguments similar to  $F_1$ , we can obtain  $R_j = o_p(1)$  for  $j = 1, 2, 3$ . Again, according to Lemma 1, we also obtain

$$\begin{aligned}
R_4 = & \mathbf{E} \{ (1 - \delta) [ \mu' \{ g(\beta^\top X) \} ]^2 V^{-1} \{ g(\beta^\top X) \} | \beta^\top X = u \} f(u) \frac{\partial \hat{\alpha}_0}{\partial \beta^{(1)}} \\
& + \mathbf{E} \{ (1 - \delta) [ \mu' \{ g(\beta^\top X) \} ]^2 V^{-1} \{ g(\beta^\top X) \} J^\top X | \beta^\top X = u \} f(u) \hat{\alpha}_1 \\
& - \mathbf{E} \{ (1 - \delta) [ \mu' \{ g(\beta^\top X) \} ]^2 V^{-1} \{ g(\beta^\top X) \} J^\top x | \beta^\top X = u \} f(u) \hat{\alpha}_1 + o_p(1).
\end{aligned}$$

and

$$R_5 = \mathbf{E} \{ (1 - \delta) [ \mu' \{ g(\beta^\top X) \} ]^2 V^{-1} \{ g(\beta^\top X) \} J^\top X | \beta^\top X = u \} f(u) + o_p(1).$$

Then, we obtain

$$\begin{aligned}
D_2 = & -\mathbb{E}\{(1-\delta)[\mu'\{g(\beta^\top X)\}]^2 V^{-1}\{g(\beta^\top X)\}|\beta^\top X = u\}f(u)\frac{\partial\hat{\alpha}_0}{\partial\beta^{(1)}} \\
& -\mathbb{E}\{(1-\delta)[\mu'\{g(\beta^\top X)\}]^2 V^{-1}\{g(\beta^\top X)\}J^\top X|\beta^\top X = u\}f(u)\hat{\alpha}_1 \\
& +\mathbb{E}\{(1-\delta)[\mu'\{g(\beta^\top X)\}]^2 V^{-1}\{g(\beta^\top X)\}J^\top x|\beta^\top X = u\}f(u)\hat{\alpha}_1 \\
& +\mathbb{E}\{(1-\delta)[\mu'\{g(\beta^\top X)\}]^2 V^{-1}\{g(\beta^\top X)\}J^\top X|\beta^\top X = u\}f(u) + o_p(1).
\end{aligned} \tag{6.12}$$

Combining the (6.10), (6.11) and (6.12), we obtain

$$\frac{\partial\hat{\alpha}_0}{\partial\beta^{(1)}} = \frac{\partial\hat{g}(\beta^\top x)}{\partial\beta^{(1)}} \xrightarrow{p} C.$$

The proof is completed.

**Proof of Theorem 3.1** By Taylor expansion, we have

$$\begin{aligned}
& [\mu\{\hat{g}(\hat{\beta}^\top X_j)\} - \mu\{g(\beta^\top X_j)\}] \\
& = \mu'\{g(\beta^\top X_j)\}\{\hat{g}(\hat{\beta}^\top X_j) - g(\beta^\top X_j)\} + o_p(1) \\
& = \mu'\{g(\beta^\top X_j)\}\{\hat{g}(\hat{\beta}^\top X_j) - \hat{g}(\beta^\top X_j) + \hat{g}(\beta^\top X_j) - g(\beta^\top X_j)\} + o_p(1) \\
& = \mu'\{g(\beta^\top X_j)\}\frac{\partial\hat{g}(\beta^\top X_j)}{\partial\beta^{(1)}}(\hat{\beta}^{(1)} - \beta^{(1)}) \\
& \quad + \mu'\{g(\beta^\top X_j)\}\{\hat{g}(\beta^\top X_j) - g(\beta^\top X_j)\} + o_p(1).
\end{aligned}$$

Let

$$\begin{aligned}
Z_j & = [\mu'\{g(\beta^\top X_j)\}\frac{\partial\hat{g}(\beta^\top X_j)}{\partial\beta^{(1)}}]^\top V^{-1}\{g(\beta^\top X_j)\}[\mu'\{g(\beta^\top X_j)\}\frac{\partial\hat{g}(\beta^\top X_j)}{\partial\beta^{(1)}}], \\
Q_j & = [\mu'\{g(\beta^\top X_j)\}\frac{\partial\hat{g}(\beta^\top X_j)}{\partial\beta^{(1)}}]^\top V^{-1}\{g(\beta^\top X_j)\}\mu'\{g(\beta^\top X_j)\}\{\hat{g}(\beta^\top X_j) - g(\beta^\top X_j)\}.
\end{aligned}$$

Then, it can be observed

$$\begin{aligned}
& n^{-1} \sum_{j=1}^n Z_j \sqrt{n} (\widehat{\beta}^{(1)} - \beta^{(1)}) \\
&= n^{-1/2} \sum_{j=1}^n [\mu' \{g(\beta^\top X_j)\} \frac{\partial \widehat{g}(\beta^\top X_j)}{\partial \beta^{(1)}}]^\top V^{-1} \{g(\beta^\top X_j)\} \\
&\quad \times [\mu \{\widehat{g}(\widehat{\beta}^\top X_j)\} - \mu \{g(\beta^\top X_j)\}] - n^{-1/2} \sum_{j=1}^n Q_j + \mathcal{o}_p(1). \\
&= n^{-1/2} \sum_{j=1}^n \{[\mu' \{g(\beta^\top X_j)\} \frac{\partial \widehat{g}(\beta^\top X_j)}{\partial \beta^{(1)}}]^\top V^{-1} \{g(\beta^\top X_j)\} \{\delta_j [\mu \{\widehat{g}(\widehat{\beta}^{(1)\top} X_j)\} - \mu \{g(\beta^{(1)\top} X_j)\}] \\
&\quad + (1 - \delta_j) [\mu \{\widehat{g}(\widehat{\beta}^{(1)\top} X_j)\} - \mu \{g(\beta^{(1)\top} X_j)\}]\} - n^{-1/2} \sum_{j=1}^n Q_j + \mathcal{o}_p(1).
\end{aligned} \tag{6.13}$$

By Lemma 2 and some tedious calculations, we have

$$\begin{aligned}
& n^{-1/2} \sum_{j=1}^n \{[\mu' \{\widehat{g}(\beta^\top X_j)\} - \mu' \{g(\beta^\top X_j)\}] \frac{\partial \widehat{g}(\beta^\top X_j)}{\partial \beta^{(1)}}]^\top V^{-1} \{g(\beta^\top X_j)\} \\
&\quad \times \{\delta_j [Y_j - \mu \{\widehat{g}(\widehat{\beta}^\top X_j)\}] + (1 - \delta_j) [\mu \{\widehat{g}(\beta^\top X_j)\} - \mu \{\widehat{g}(\widehat{\beta}^\top X_j)\}]\} = \mathcal{o}_p(1).
\end{aligned} \tag{6.14}$$

By (6.13) and (6.14), we have

$$\begin{aligned}
\sqrt{n} (\widehat{\beta}^{(1)} - \beta^{(1)}) &= Z^{-1} \{n^{-1/2} \sum_{j=1}^n \mu' \{\widehat{g}(\beta^\top X_j)\} \frac{\partial \widehat{g}(\beta^\top X_j)}{\partial \beta^{(1)}} V^{-1} \{g(\beta^\top X_j)\} \\
&\quad \times \{\delta_j [Y_i - \mu \{g(\beta^\top X_j)\}] + (1 - \delta_j) [\mu \{\widehat{g}(\beta^\top X_j)\} - \mu \{g(\beta^\top X_j)\}]\} \\
&\quad - n^{-1/2} \sum_{j=1}^n Q_j + \mathcal{o}_p(1).
\end{aligned} \tag{6.15}$$

By Lemma 1 and Lemma 2, we can obtain

$$n^{-1/2} \sum_{j=1}^n Q_j = \mathcal{o}_p(1). \tag{6.16}$$

By condition (a), (6.15) and (6.16), we obtain

$$\begin{aligned}
\sqrt{n} (\widehat{\beta}^{(1)} - \beta^{(1)}) &= Z^{-1} n^{-1/2} \sum_{j=1}^n \mu' \{\widehat{g}(\beta^\top X_j)\} \frac{\partial \widehat{g}(\beta^\top X_j)}{\partial \beta^{(1)}} V^{-1} \{g(\beta^\top X_j)\} \\
&\quad \times \{\delta_j [Y_i - \mu \{g(\beta^\top X_j)\}]\} + \mathcal{o}_p(1).
\end{aligned} \tag{6.17}$$

Theorem 3.1 follows directly from Lemma 3. The proof is completed.



**Proof of Theorem 3.3** By Theorem 3.1, we know that  $\widehat{\beta}$  is a root- $n$  consistent estimator of  $\beta^0$ . Then, using the arguments similar to the proof of Proposition 1 (iii) in Cui et al (2011), we have that

$$\sqrt{nb_n}(\widehat{g}(\widehat{\beta}^\top X) - \widehat{g}(\beta^{0\top} X)) = o_p(1).$$

According to Lemma 2, we know

$$\begin{aligned} & \sqrt{nb_n}\{\widehat{g}(\beta^{0\top} X) - g(\beta^{0\top} X) - \frac{\mu^{(2)}\{g(\beta^\top x)\}}{2}e_1 S^{-1} U b_n^2\} \\ & \xrightarrow{\mathcal{L}} \mathbf{N}(0, \frac{\rho_2\{g(\beta^\top x)\}}{\pi(x)f_{\beta^\top x}(\beta^\top x)} S^{-1} S^* S^{-1}), \end{aligned}$$

Therefore, we have

$$\begin{aligned} & \sqrt{nb_n}\{\widehat{g}(\widehat{\beta}^\top X) - g(\beta^{0\top} X) - \frac{\mu^{(2)}\{g(\beta^\top x)\}}{2}e_1 S^{-1} U b_n^2\} \\ & = \sqrt{nb_n}\{\widehat{g}(\widehat{\beta}^\top X) - \widehat{g}(\beta^{0\top} X)\} \\ & \quad + \sqrt{nb_n}[\widehat{g}(\beta^{0\top} X) - \mu\{g(\beta^{0\top} X) - \frac{\mu^{(2)}\{g(\beta^\top x)\}}{2}e_1 S^{-1} U b_n^2\}] \\ & \xrightarrow{\mathcal{L}} \mathbf{N}(0, \frac{\rho_2\{g(\beta^\top x)\}}{\pi(x)f_{\beta^\top x}(\beta^\top x)} S^{-1} S^* S^{-1}). \end{aligned}$$

The proof is completed.

Acknowledgements. Wang's research was supported by the National Science Fund for Distinguished Young Scholars in China (10725106), the National Natural Science Foundation of China (General program 11171331 and Key program 11331011), a grant from the Key Lab of Random Complex Structure and Data Science, CAS and the Natural Science Foundation of SZU. Härdle's research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".

## References

- Carroll, R. J., Ruppert, D., Welsh, A. H., (1998). Local estimating equations. *J. Amer. Statist. Assoc.*, 93:214-227.
- Cui, X., Härdle, W., Zhu, L. X. (2011). The EFM approach for single-index models. *Ann. Statist.*, 39:1658–1688.
- Ding, X. B., Wang, Q. H. (2011). Fusion-Refinement procedure for dimension reduction with missing response at random *J. Amer. Statist. Assoc.*, 106:1193–1207.
- Fan, J., Gijbels, I. (1996). *Local polynomial modeling and its applications*. London: Chapman and Hall.
- Gelman, A., Carlin, J. B., Stern, H. S., Rubin, D. B.(1995). *Bayesian Data Analysis*. London: Chapman and Hall
- Hall, P. (1989). On projection pursuit regression. *Ann. Statist.*, 17:573–588.
- Härdle, W., Hall, P., Ichimura, H. (1993). Optimal smoothing in single-index models *Ann. Statist.*, 21:157–178.
- Härdle, W., Tsybakov, A. B.(1993). How sensitive are average derivatives. *Journal of Econometrics*, 58:31–48.
- Hristache, M., Juditsky, A., Spokoiny, V. (2001). Direct estimation of the index coefficient in a single-index model. *Ann. Statist.*, 29:595–623.
- Hu, Z. H., Follmann, D. A., and Qin, J. (2010). Semiparametric dimension reduction estimation for mean response with missing data. *Biometrika*, 97:305–319.
- Ichimura, H. (1993). Semiparametric least squares (SLS) and weighted SLS estimation of single-index models. *Journal of Econometrics*, 58:72–120.
- Ichimura, H. (1987). *Estimation of single index models*. Ph.D. Dissertation, Dept. Economics, MIT.
- Mack, Y., Silverman, B. (1982). Weak and strong uniform consistency of kernel regression estimates. *Probability Theory and Related Fields*, 61:405–415.

- Robins, J. M., Rotnitzky, A., Zhao L. P. (1994). Estimation of regression coefficients when some regressors are not always observed. *J. Amer. Statist. Assoc.*, 89:846–866.
- Robins, J. M., Rotnitzky, A., Zhao L. P. (1995). Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *J. Amer. Statist. Assoc.*, 90:106–121.
- Rosenbaum, P. R., Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70:41–55.
- Wang, Q. H., Rao, J. N. K. (2002). Empirical likelihood-based inference under imputation for missing response data. *The Annals of Statist.*, 30:896–924.
- Wang, Q. H., Linton, O., Härdle, W. (2004). Semiparametric regression analysis with missing response at random. *J. Amer. Statist. Assoc.*, 99:334–345.
- Li, K. C. (1991). Sliced inverse regression for dimension reduction. *J. Amer. Statist. Assoc.*, 86:316–342.
- Wang, Y. H., Shen, J. S., He, S. Y., Wang, Q. H. (2010). Estimation of single index model with missing response at random. *J. Stat. Plan. Inference*, 140:1671–1690.
- Zhu, L. X., Fang, K. T. (1992). On projection pursuit approximation for nonparametric regression. In *P.S. Sen, I.A. Salama (Eds.), Proceedings of Order Statistics and Nonparametrics: Theory and Applications*, 455–469.
- Zhu, L. X., Fang, K. T. (1996). Asymptotics for kernel estimate of sliced inverse regression. *Ann. Statist.*, 14:1053–1068.

Table 1: AB and SRTSC of  $\hat{\beta}$ ,  $\hat{\beta}^{cc}$  and  $\hat{\beta}^{wang}$  with different missing proportion and different sample sizes.

AB				
n	p	$\hat{\beta}$	$\hat{\beta}^{cc}$	$\hat{\beta}^{wang}$
100	$p = 0.25$	0.0247	0.0967	0.1014
	$p = 0.50$	0.0476	0.1256	0.1277
200	$p = 0.25$	0.0104	0.0405	0.0413
	$p = 0.50$	0.0203	0.0737	0.0209
400	$p = 0.25$	0.0063	0.0062	0.0114
	$p = 0.50$	0.0096	0.0247	0.0102
SRTSC				
100	$p = 0.25$	0.0305	0.0932	0.1092
	$p = 0.50$	0.0572	0.0808	0.1090
200	$p = 0.25$	0.0133	0.0784	0.0873
	$p = 0.50$	0.0253	0.0934	0.0775
400	$p = 0.25$	0.0086	0.0271	0.0494
	$p = 0.50$	0.0125	0.0623	0.0439

Table 2: AB and SRTSC of  $\hat{\beta}$ ,  $\hat{\beta}^{cc}$  and  $\hat{\beta}^{wang}$  with different missing proportion and different sample sizes.

AB				
n	p	$\hat{\beta}$	$\hat{\beta}^{cc}$	$\hat{\beta}^{wang}$
100	$p = 0.25$	0.0347	0.1273	0.1044
	$p = 0.50$	0.0685	0.1366	0.1262
200	$p = 0.25$	0.0106	0.0784	0.0184
	$p = 0.50$	0.0232	0.1144	0.0730
400	$p = 0.25$	0.0064	0.0203	0.0059
	$p = 0.50$	0.0095	0.0535	0.0092
SRTSC				
100	$p = 0.25$	0.0442	0.0813	0.1267
	$p = 0.50$	0.0775	0.0722	0.1371
200	$p = 0.25$	0.0138	0.0904	0.0581
	$p = 0.50$	0.0305	0.0861	0.1070
400	$p = 0.25$	0.0090	0.0203	0.0296
	$p = 0.50$	0.0125	0.0831	0.0438

Table 3: AB and SRTSC of  $\hat{\beta}$  and  $\hat{\beta}^{cc}$  with different missing proportion and different sample sizes.

AB			
n	p	$\hat{\beta}$	$\hat{\beta}^{cc}$
100	$p = 0.25$	0.0500	0.1214
	$p = 0.50$	0.0718	0.1335
200	$p = 0.25$	0.0325	0.0829
	$p = 0.50$	0.0443	0.1106
400	$p = 0.25$	0.0227	0.0360
	$p = 0.50$	0.0294	0.0630
SRTSC			
100	$p = 0.25$	0.0567	0.0873
	$p = 0.50$	0.0779	0.0759
200	$p = 0.25$	0.0382	0.0916
	$p = 0.50$	0.0504	0.0913
400	$p = 0.25$	0.0263	0.0556
	$p = 0.50$	0.0340	0.0824

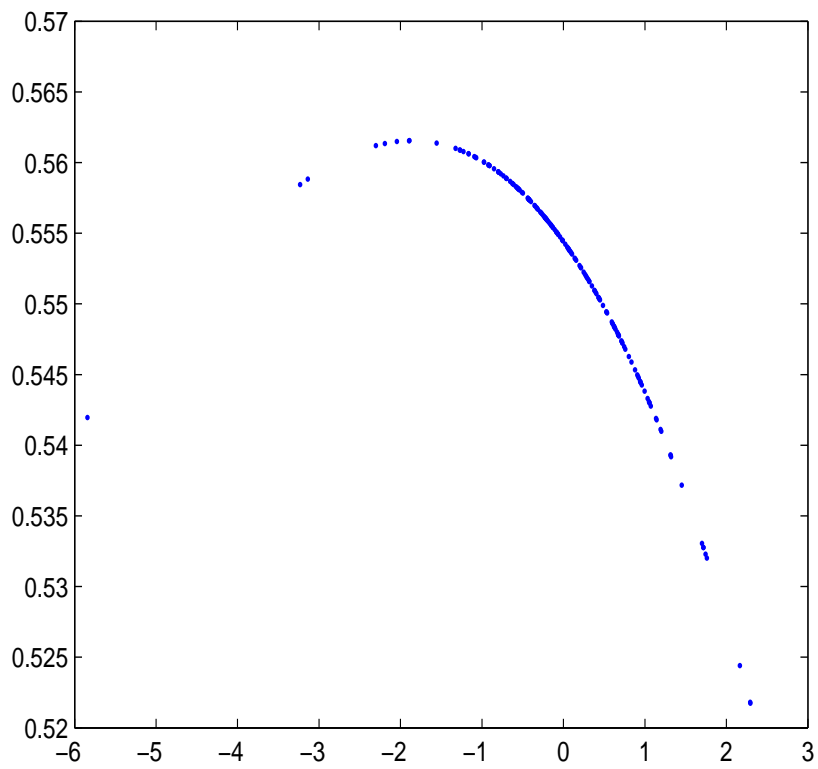


Figure 1: the scatter plot of the estimated single index  $\hat{g}(\hat{\beta}^\top X)$  against  $\hat{\beta}^\top X$  in the setting of  $T = 0$ .

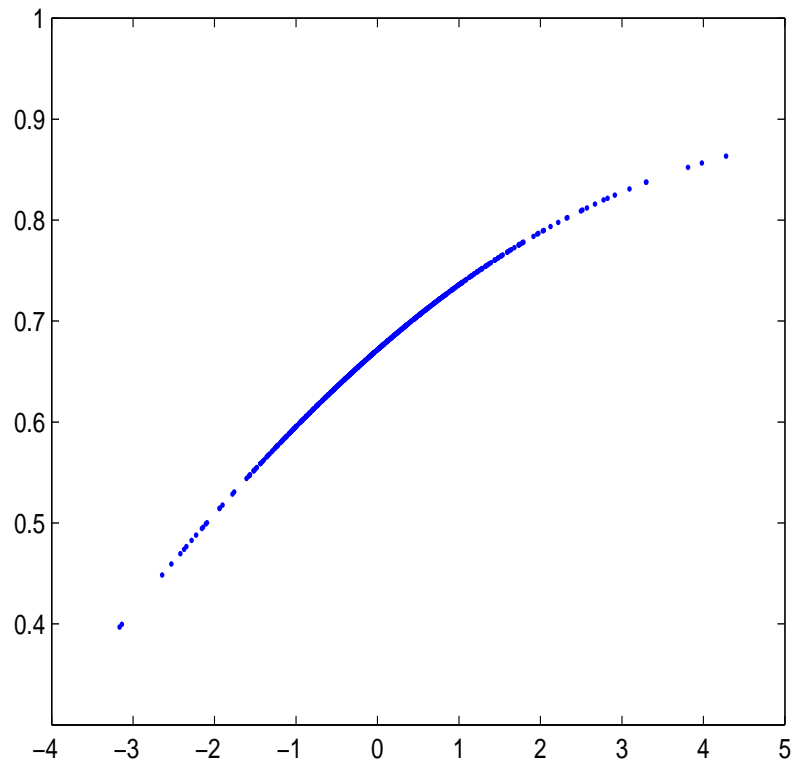


Figure 2: the scatter plot of the estimated single index  $\hat{g}(\hat{\beta}^\top X)$  against  $\hat{\beta}^\top X$  in the setting of  $T = 1$ .



## SFB 649 Discussion Paper Series 2014

For a complete list of Discussion Papers published by the SFB 649, please visit <http://sfb649.wiwi.hu-berlin.de>.

- 001 "Principal Component Analysis in an Asymmetric Norm" by Ngoc Mai Tran, Maria Osipenko and Wolfgang Karl Härdle, January 2014.
- 002 "A Simultaneous Confidence Corridor for Varying Coefficient Regression with Sparse Functional Data" by Lijie Gu, Li Wang, Wolfgang Karl Härdle and Lijian Yang, January 2014.
- 003 "An Extended Single Index Model with Missing Response at Random" by Qihua Wang, Tao Zhang, Wolfgang Karl Härdle, January 2014.

**SFB 649, Spandauer Straße 1, D-10178 Berlin**  
**<http://sfb649.wiwi.hu-berlin.de>**

This research was supported by the Deutsche  
Forschungsgemeinschaft through the SFB 649 "Economic Risk".

