Mathematics

http://www.aimspress.com/journal/Math

AIMS Mathematics, 6(7): 7125–7152.
DOI:10.3934/math.2021418
Received: 20 January 2021
Accepted: 08 April 2021
Published: 28 April 2021

Research article

Empirical likelihood for varying coefficient partially nonlinear model with missing responses

Liqi Xia[1], Xiuli Wang[1,*], Peixin Zhao[2] and Yunquan Song[3]

[1] School of Mathematics and Statistics, Shandong Normal University, Jinan 250358, China

[2] College of Mathematics and Statistics, Chongqing Technology and Business University, Chongqing 400067, China

[3] College of Science, China University of Petroleum, Qingdao 266580, China

* **Correspondence:** Email: wxlmath@163.com.

Abstract: In this paper, we consider the statistical inferences for varying coefficient partially nonlinear model with missing responses. Firstly, we employ the profile nonlinear least squares estimation based on the weighted imputation method to estimate the unknown parameter and the nonparametric function, meanwhile the asymptotic normality of the resulting estimators is proved. Secondly, we consider empirical likelihood inferences based on the weighted imputation method for the unknown parameter and nonparametric function, and propose an empirical log-likelihood ratio function for the unknown parameter vector in the nonlinear function and a residual-adjusted empirical log-likelihood ratio function for the nonparametric component, meanwhile construct relevant confidence regions. Thirdly, the response mean estimation is also studied. In addition, simulation studies are conducted to examine the finite sample performance of our methods, and the empirical likelihood approach based on the weighted imputation method (IEL) is further applied to a real data example.

Keywords: varying coefficient partially nonlinear model; profile nonlinear least squares estimation; weighted imputation; missing responses; empirical likelihood inferences; confidence region
Mathematics Subject Classification: 62G05, 62G20

1. Introduction

Recently, the varying coefficient partially nonlinear model (VCPNLM) has received a great deal of attention because of the flexibility of the semiparametric models, it has the following form:

$$Y = \mathbf{X}^T \boldsymbol{\theta}(U) + g(\mathbf{Z}, \boldsymbol{\beta}) + \varepsilon,$$

where Y is a response variable, $\mathbf{X}$, U and $\mathbf{Z}$ are the covariates, $\mathbf{X} \in R^q, U \in R$, $\mathbf{Z} \in R^r$, $\boldsymbol{\theta}(\cdot) = (\theta_1(\cdot), \cdots, \theta_q(\cdot))^T$ is a q-dimensional vector of unknown coefficient functions, $g(\cdot,\cdot)$ is a given nonlinear function with unknown parameter vector $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^T$, the covariate $\mathbf{Z}$ doesn't necessarily have the same dimension as the unknown parameter vector $\boldsymbol{\beta}$ in $g(\mathbf{Z}, \boldsymbol{\beta})$, ε is the random error with $E(\varepsilon \mid \mathbf{X}, U, \mathbf{Z}) = 0$ and $Var(\varepsilon \mid \mathbf{X}, U, \mathbf{Z}) = \sigma^2$.

This model which was first proposed by Li and Mei [1] contains many important submodels such as varying coefficient model, nonlinear model, partially nonlinear model and varying coefficient partially linear model, and it is valuable and meaningful to do some statistical researches on VCPNLMs. In Li and Mei [1], they proposed a profile nonlinear least squares estimation approach for the parameter vector β and coefficient function vector $\boldsymbol{\theta}(U)$. Yang and Yang [2] proposed a two-stage estimation by employing an orthogonality-projection-based method for parametric coefficient estimation, then developed a variable selection procedure for the coefficient functions based on smooth-threshold estimating equations. Qian and Huang [3] considered the corrected profile least squares estimation procedure with measurement errors for the nonparametric part, and a generalized likelihood ratio test to check whether the coefficient functions are a constant or not. Zhou, Zhao and Wang [4] applied the empirical likelihood technique to obtain the confidence regions for the unknown parameter vector and the nonparametric vector. In Jiang, Ji and Xie [5], a robust estimation procedure based on exponential squared loss function was proposed. Xiao and Chen [6] developed a bias-corrected profile least squares estimation procedure with additive measurement errors and the likelihood ratio test approach for VCPNLM. Recently, Wang, Zhao and Du [7] studied statistical inferences for this model with missing covariates, first they proposed an inverse probability weighted profile nonlinear least squares technique for estimating the unknown parameter and the nonparametric function, then they considered empirical likelihood inferences for the unknown parameter and nonparametric function.

To our best knowledge, there is no literature to study the inferences for VCPNLM with missing responses. In fact, missing responses is common in opinion polls, mail enquiries, market research surveys, medical studies, socioeconomic investigations and other scientific experiments. In such circumstances, the early method to deal with missing responses at random is complete data method, please refer to Little and Rubin [8]. However, this method may lose plenty of information by simply ignoring the missing data. To improve the accuracy of the estimators, a natural method is to impute a value for each missing response in order to achieve a full data set, see Cheng [9], Wang and Rao [10, 11], Wang, Linton and Hardle [12], but not confined these. In Xue [13, 14], the weighted imputation method was used to construct weighted-corrected empirical likelihood ratio for the parameters of interest when the regression model is linear and nonparametric. Further, Wang, Chen and Lin [15] generalized the weighted corrected method to estimating equation and defined the weighted-corrected estimating function, then proved an empirical log-likelihood ratio based on the weighted-corrected estimating function to be a standard chi-square distribution asymptotically under some suitable conditions. Inspired by Xue [13, 14] and Wang, Chen and Lin [15], in this paper we extend the weighted imputation method to VCPNLM with missing responses. Firstly, the profile nonlinear least squares inference of the complete data method and the asymptotic normality properties of the resulting estimators are finished, then based on estimators of the complete data, we impute a value for each missing response by the inverse probability weighted method, the asymptotic normality and confidence regions are established. In addition, due to the advantages of the empirical likelihood method introduced by Owen [16, 17] in constructing confidence region, we apply empirical likelihood method to the imputed data, the confidence regions for the parameter and the response mean, the point-wise confidence intervals for coefficient function are also obtained. From simulation studies, we find that the empirical likelihood based on the weighted imputation method outperforms the profile nonlinear least squares estimation based on the weighted imputation method.

The rest of this paper is organized as follows. In Section 2, the profile nonlinear least squares estimation of the complete data method and the weighted imputation method for β and $\boldsymbol{\theta}(U)$ are proposed. In Section 3, empirical likelihood inferences based on the weighted imputation method for the unknown parameter and nonparametric function are suggested. The response mean estimation is considered in Section 4. In Section 5, some simulation studies and real data analysis are conducted to assess the performance of two methods. Discussion is made in Section 6 and technical proofs are provided in Section 7.

2. The estimation based on the profile nonlinear least squares

Suppose that $\{Y_i, \mathbf{X}_i, U_i, \mathbf{Z}_i, i = 1, 2, \cdots, n\}$ is a random sample from the following varying coefficient partially nonlinear model

$$Y = \mathbf{X}^T \boldsymbol{\theta}(U) + g(\mathbf{Z}, \boldsymbol{\beta}) + \varepsilon,$$

where $\{(\mathbf{X}_i, U_i, \mathbf{Z}_i), i = 1, 2, \ldots, n\}$ is completely observable, the response Y_i is missing at random (MAR). Introducing the indicator variable δ_i, when Y_i is observed, then $\delta_i = 1$, otherwise $\delta_i = 0$.

The response probability, also called the propensity score or selection probability function under MAR assumption, is given by

$$P(\delta_i = 1 | Y_i, \mathbf{X}_i, U_i, \mathbf{Z}_i) = P(\delta_i = 1 | \mathbf{X}_i, U_i, \mathbf{Z}_i) = \pi(\mathbf{X}_i, U_i, \mathbf{Z}_i).$$

It is essential to estimate the completely unknown propensity score function before we construct estimators for $\boldsymbol{\beta}$ and $\boldsymbol{\theta}(\cdot)$. We posit the following logistic model for $\pi(\mathbf{X}_i, U_i, \mathbf{Z}_i)$, that is

$$\pi(\mathbf{V}_i, \omega) = \frac{\exp(\omega_0 + \boldsymbol{\omega}_1^T \mathbf{X}_i + \omega_2 U_i + \boldsymbol{\omega}_3^T \mathbf{Z}_i)}{1 + \exp(\omega_0 + \boldsymbol{\omega}_1^T \mathbf{X}_i + \omega_2 U_i + \boldsymbol{\omega}_3^T \mathbf{Z}_i)},$$

where $\mathbf{V}_i = (\mathbf{X}_i^T, U_i, \mathbf{Z}_i^T)^T$, and $\boldsymbol{\omega} = (\omega_0, \boldsymbol{\omega}_1^T, \omega_2, \boldsymbol{\omega}_3^T)^T$ is an unknown parameter vector. Naturally, $\boldsymbol{\omega}$ can be estimated by maximizing the log-likelihood function:

$$lnL(\boldsymbol{\omega}) = \sum_{i=1}^{i=n} \{\delta_i ln\pi(\mathbf{V}_i, \boldsymbol{\omega}) + (1 - \delta_i) ln[1 - \pi(\mathbf{V}_i, \boldsymbol{\omega})]\},$$

then the estimator of $\pi(\mathbf{V}_i, \boldsymbol{\omega})$ is $\pi(\mathbf{V}_i, \hat{\omega})$, for details, please refer to Chen, Feng and Xue [18].

2.1. The least squares estimation of the complete data method

The following model

$$\delta_i Y_i = \delta_i \mathbf{X}_i^T \theta(U_i) + \delta_i g(\mathbf{Z}_i, \boldsymbol{\beta}) + \delta_i \varepsilon_i, \tag{2.1}$$

is complete case, for u in a small neighborhood of u_0, applying the Taylor expansion for $\theta_j(u)$ gives $\theta_j(u) \approx \theta_j(u_0) + \theta'_j(u_0)(u - u_0)$, $j = 1, 2, \cdots, q$. Suppose that $\boldsymbol{\beta}$ is known in advance, we employ the local linear method and minimize the following objective function to estimate the varying coefficient functions $\theta(u)$,

$$\sum_{i=1}^{n} \delta_i \{Y_i - g(\mathbf{Z}_i, \boldsymbol{\beta}) - \sum_{j=1}^{q} [\theta_j(u_0) + \theta'_j(u_0)(U_i - u_0)] X_{ij}\}^2 K_{1,h_1}(U_i - u_0), \tag{2.2}$$

where $K_{1,h_1}(\cdot) = K(\cdot/h_1)/h_1$ is the kernel function with bandwidth h_1.

Denote $\mathbf{Y} = (Y_1, \cdots, Y_n)^T, \mathbf{g}(\mathbf{Z}, \boldsymbol{\beta}) = (g(\mathbf{Z}_1, \boldsymbol{\beta}), \cdots, g(\mathbf{Z}_n, \boldsymbol{\beta}))^T, \mathbf{M} = (\mathbf{X}_1^T \boldsymbol{\theta}(U_1), \cdots, \mathbf{X}_n^T \boldsymbol{\theta}(U_n))^T$, $\boldsymbol{\Psi}(u_0) = (\boldsymbol{\theta}^T(u_0), h_1 \boldsymbol{\theta}'^T(u_0))^T, \mathbf{W}_1(u_0) = diag(K_{1,h_1}(U_1 - u_0), \cdots, K_{1,h_1}(U_n - u_0)), \boldsymbol{\Delta} = diag(\delta_1, \cdots, \delta_n)$, and

$$\mathbf{X}_{h_1}(u_0) = \begin{pmatrix} \mathbf{X}_1^T & h_1^{-1}(U_1 - u_0)\mathbf{X}_1^T \\ \vdots & \vdots \\ \mathbf{X}_n^T & h_1^{-1}(U_1 - u_0)\mathbf{X}_n^T \end{pmatrix}.$$

Then the solution of the weighted least squares problem (2.2) can be expressed as

$$\hat{\boldsymbol{\Psi}}_c(u_0, \boldsymbol{\beta}) = [\mathbf{X}_{h_1}^T(u_0)\mathbf{W}_1(u_0)\boldsymbol{\Delta}\mathbf{X}_{h_1}(u_0)]^{-1}\mathbf{X}_{h_1}^T(u_0)\mathbf{W}_1(u_0)\boldsymbol{\Delta}[\mathbf{Y} - \mathbf{g}(\mathbf{Z}, \boldsymbol{\beta})],$$

the estimator of the coefficient function vector $\boldsymbol{\theta}(u)$ at u_0 is

$$\hat{\boldsymbol{\theta}}_c(u_0, \boldsymbol{\beta}) = (\mathbf{I}_{q\times q}, \mathbf{0}_{q\times q})[\mathbf{X}_{h_1}^T(u_0)\mathbf{W}_1(u_0)\boldsymbol{\Delta}\mathbf{X}_{h_1}(u_0)]^{-1}\mathbf{X}_{h_1}^T(u_0)\mathbf{W}_1(u_0)\boldsymbol{\Delta}[\mathbf{Y} - \mathbf{g}(\mathbf{Z}, \boldsymbol{\beta})].$$

Let

$$\hat{\mathbf{M}} = \begin{pmatrix} \mathbf{X}_1^T \hat{\boldsymbol{\theta}}_c(U_1, \boldsymbol{\beta}) \\ \vdots \\ \mathbf{X}_n^T \hat{\boldsymbol{\theta}}_c(U_n, \boldsymbol{\beta}) \end{pmatrix} \hat{=} \mathbf{S}_{h_1}[\mathbf{Y} - \mathbf{g}(\mathbf{Z}, \boldsymbol{\beta})],$$

where

$$\mathbf{S}_{h_1} = \begin{pmatrix} (\mathbf{X}_1^T \quad \mathbf{0}_{1\times q})[\mathbf{X}_{h_1}^T(U_1)\mathbf{W}_1(U_1)\boldsymbol{\Delta}\mathbf{X}_{h_1}(U_1)]^{-1}\mathbf{X}_{h_1}^T(U_1)\mathbf{W}_1(U_1)\boldsymbol{\Delta} \\ \vdots \\ (\mathbf{X}_n^T \quad \mathbf{0}_{1\times q})[\mathbf{X}_{h_1}^T(U_n)\mathbf{W}_1(U_n)\boldsymbol{\Delta}\mathbf{X}_{h_1}(U_n)]^{-1}\mathbf{X}_{h_1}^T(U_n)\mathbf{W}_1(U_n)\boldsymbol{\Delta} \end{pmatrix},$$

then, minimizing the following profile nonlinear least squares function with respect to $\boldsymbol{\beta}$

$$\begin{aligned} Q_c(\boldsymbol{\beta}) &= \sum_{i=1}^{n} \delta_i [Y_i - \mathbf{X}_i^T \hat{\boldsymbol{\theta}}_c(U_i, \boldsymbol{\beta}) - g(\mathbf{Z}_i, \boldsymbol{\beta})]^2 \\ &= [\mathbf{Y} - \mathbf{g}(\mathbf{Z}, \boldsymbol{\beta})]^T (\mathbf{I}_n - \mathbf{S}_{h_1})^T \boldsymbol{\Delta} (\mathbf{I}_n - \mathbf{S}_{h_1})[\mathbf{Y} - \mathbf{g}(\mathbf{Z}, \boldsymbol{\beta})] \end{aligned}$$

yields the profile nonlinear least squares estimator $\hat{\boldsymbol{\beta}}_c$ of the complete data method. Then, substituting $\hat{\boldsymbol{\beta}}_c$ into $\hat{\boldsymbol{\theta}}_c(u_0, \boldsymbol{\beta})$, we obtain the estimator $\hat{\boldsymbol{\theta}}_c(u_0, \hat{\boldsymbol{\beta}}_c)$ of $\hat{\boldsymbol{\theta}}_c(u_0, \boldsymbol{\beta})$.

We define some necessary notations as follows:

$$\mu_{1,j} = \int \mu^j K_{1,h_1}(u)du, \quad \nu_{1,j} = \int \mu^j K_{1,h_1}^2(u)du, \quad \mathbf{g}'(\mathbf{Z}, \boldsymbol{\beta}) = \partial \mathbf{g}(\mathbf{Z}, \boldsymbol{\beta})/\partial \boldsymbol{\beta},$$

$$\mathbf{\Gamma}(u) = E(\pi(\mathbf{V})\mathbf{X}\mathbf{X}^T|U = u), \quad \mathbf{\Phi}(u) = E[\pi(\mathbf{V})\mathbf{X}\mathbf{g}'(\mathbf{Z}, \boldsymbol{\beta}_0)^T|U = u].$$

Theorem 2.1.1 Suppose that the conditions $C1 - C8$ in Section 7 hold, and $\boldsymbol{\beta}_0$ is the true parameter vector of $\boldsymbol{\beta}$, then

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_c - \boldsymbol{\beta}_0) \xrightarrow{D} N(\mathbf{0}, \sigma^2\mathbf{\Sigma}^{-1}),$$

where "$\xrightarrow{D}$" stands for the convergence in distribution, $\mathbf{\Sigma} = E[\pi(\mathbf{V})\mathbf{g}'(\mathbf{Z}, \boldsymbol{\beta}_0)\mathbf{g}'(\mathbf{Z}, \boldsymbol{\beta}_0)^T] - E[\mathbf{\Phi}(U)^T\mathbf{\Gamma}^{-1}(U)\mathbf{\Phi}(U)]$.

Theorem 2.1.2 Suppose that the conditions $C1 - C8$ in Section 7 hold. For any $u_0 \in \Omega$, we have

$$\sqrt{nh_1}\left[\hat{\mathbf{\Psi}}_c(u_0, \hat{\boldsymbol{\beta}}_c) - \mathbf{\Psi}_0(u_0) - \frac{1}{2}h_1^2\mu_{1,2}\begin{pmatrix}\boldsymbol{\theta}_0''(u_0)\\ 0\end{pmatrix}\right] \xrightarrow{D} N(\mathbf{0}, \mathbf{\Sigma}_1),$$

where $\mathbf{\Sigma}_1 = \sigma^2 f^{-1}(u_0)\mathbf{\Gamma}^{-1}(u_0) \otimes \begin{pmatrix}\nu_{1,0} & 0\\ 0 & \nu_{1,2}\mu_{1,2}^{-2}\end{pmatrix}$, and $\otimes$ denotes the Kronecker product. In particular, we have

$$\sqrt{nh_1}[\hat{\boldsymbol{\theta}}_c(u_0, \hat{\boldsymbol{\beta}}_c) - \boldsymbol{\theta}_0(u_0) - \frac{1}{2}h_1^2\mu_{1,2}\boldsymbol{\theta}_0''(u_0)] \xrightarrow{D} N(\mathbf{0}, \sigma^2\nu_{1,0}f^{-1}(u_0)\mathbf{\Gamma}^{-1}(u_0)).$$

2.2. *The least squares estimation based on the weighted imputation method*

Although the implementation of the complete case method is simple, it may result in an inefficient estimator and the loss of a great of information by simply ignoring the missing data. In this section, we introduce a weighted imputation method to deal with the problems.

Let

$$\breve{Y}_i = \frac{\delta_i}{\hat{\pi}(\mathbf{V}_i)}Y_i + (1 - \frac{\delta_i}{\hat{\pi}(\mathbf{V}_i)})[\mathbf{X}_i^T\hat{\boldsymbol{\theta}}_c(U_i, \hat{\boldsymbol{\beta}}_c) + g(\mathbf{Z}_i, \hat{\boldsymbol{\beta}}_c)], \tag{2.3}$$

where $\hat{\pi}(\mathbf{V}_i)$ denotes $\pi(\mathbf{V}_i, \hat{\omega})$, the estimator of $\pi(\mathbf{V}_i, \boldsymbol{\omega})$. According to the profile nonlinear least squares method similar to section 2.1, we can obtain

$$\hat{\mathbf{\Psi}}_I(u_0, \boldsymbol{\beta}) = [\mathbf{X}_{h_2}^T(u_0)\mathbf{W}_2(u_0)\mathbf{X}_{h_2}(u_0)]^{-1}\mathbf{X}_{h_2}^T(u_0)\mathbf{W}_2(u_0)[\breve{\mathbf{Y}} - \mathbf{g}(\mathbf{Z}, \boldsymbol{\beta})],$$

$$\hat{\boldsymbol{\theta}}_I(u_0, \boldsymbol{\beta}) = (\mathbf{I}_{q\times q}, \mathbf{0}_{q\times q})[\mathbf{X}_{h_2}^T(u_0)\mathbf{W}_2(u_0)\mathbf{X}_{h_2}(u_0)]^{-1}\mathbf{X}_{h_2}^T(u_0)\mathbf{W}_2(u_0)[\breve{\mathbf{Y}} - \mathbf{g}(\mathbf{Z}, \boldsymbol{\beta})],$$

where
$\mathbf{X}_{h_2}(u_0) = \begin{pmatrix}\mathbf{X}_1^T & h_2^{-1}(U_1 - u_0)\mathbf{X}_1^T\\ \vdots & \vdots\\ \mathbf{X}_n^T & h_2^{-1}(U_1 - u_0)\mathbf{X}_n^T\end{pmatrix}$, $\breve{\mathbf{Y}} = (\breve{Y}_1, \cdots, \breve{Y}_n)^T$, $\mathbf{W}_2(u_0) = \mathrm{diag}(K_{2,h_2}(U_1 - u_0), \cdots, K_{2,h_2}(U_n - u_0))$, $K_{2,h_2}(\cdot) = K(\cdot/h_2)/h_2$ is the kernel function with bandwidth h_2. Minimizing the following function with respect to $\boldsymbol{\beta}$

$$Q(\boldsymbol{\beta}) = \sum_{i=1}^{n}[\breve{Y}_i - \mathbf{X}_i^T\hat{\boldsymbol{\theta}}_I(U_i, \boldsymbol{\beta}) - g(\mathbf{Z}_i, \boldsymbol{\beta})]^2,$$

we can obtain the profile nonlinear least squares estimator $\hat{\boldsymbol{\beta}}_I$ of the weighted imputation method. Then, substituting $\hat{\boldsymbol{\beta}}_I$ into $\hat{\boldsymbol{\theta}}_I(u_0, \boldsymbol{\beta})$ gives the estimator $\hat{\boldsymbol{\theta}}_I(u_0, \hat{\boldsymbol{\beta}}_I)$ of $\hat{\boldsymbol{\theta}}_I(u_0, \boldsymbol{\beta})$.

Denote $\mathbf{A}^{\otimes 2} = \mathbf{A}\mathbf{A}^T$ for a vector or matrix $\mathbf{A}$, $\mu_{2,j} = \int \mu^j K_{2,h_2}(u)du$, $\nu_{2,j} = \int \mu^j K_{2,h_2}^2(u)du$,

$$\mathbf{\Gamma}^*(u) = E(\mathbf{X}\mathbf{X}^T|U = u), \quad \mathbf{\Phi}^*(u) = E[\mathbf{X}\mathbf{g}'(\mathbf{Z}, \boldsymbol{\beta})^T|U = u].$$

The following Theorem 2.2.1 and Theorem 2.2.2 give the asymptotic normality of $\hat{\boldsymbol{\beta}}_I$ and $\hat{\boldsymbol{\theta}}_I(u_0, \hat{\boldsymbol{\beta}}_I)$, respectively.

Theorem 2.2.1 Suppose that the conditions $C1 - C8$ in Section 7 hold and $\boldsymbol{\beta}_0$ is the true parameter vector of $\boldsymbol{\beta}$, then

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_I - \boldsymbol{\beta}_0) \xrightarrow{D} N(\mathbf{0}, \boldsymbol{\Sigma}_3^{-1}\boldsymbol{\Sigma}_4\boldsymbol{\Sigma}_3^{-1}),$$

where $\boldsymbol{\Sigma}_3 = E[\mathbf{g}'(\mathbf{Z}, \boldsymbol{\beta}_0)\mathbf{g}'(\mathbf{Z}, \boldsymbol{\beta}_0)^T] - E[\boldsymbol{\Phi}^*(U)^T\boldsymbol{\Gamma}^*(U)^{-1}\boldsymbol{\Phi}^*(U)]$, $\boldsymbol{\Sigma}_4 = E[\pi^{-1}(\mathbf{V})\boldsymbol{\varepsilon}^2\mathbf{g}'(\mathbf{Z}, \boldsymbol{\beta}_0)\mathbf{g}'(\mathbf{Z}, \boldsymbol{\beta}_0)^T] - E[\pi^{-1}(\mathbf{V})\boldsymbol{\varepsilon}^2\boldsymbol{\Phi}^*(U)^T\boldsymbol{\Gamma}^*(U)^{-1}\boldsymbol{\Phi}^*(U)]$.

Remark 1: As a consequence of Theorem 2.2.1, we can construct the normal approximated $1 - \alpha$ confidence region based on the weighted imputation method (INA) for $\boldsymbol{\beta}$:

$$C_{INA}(\alpha) = \{\boldsymbol{\beta} : n(\hat{\boldsymbol{\beta}}_I - \boldsymbol{\beta})^T(\hat{\boldsymbol{\Sigma}}_3^{-1}\hat{\boldsymbol{\Sigma}}_4\hat{\boldsymbol{\Sigma}}_3^{-1})^{-1}(\hat{\boldsymbol{\beta}}_I - \boldsymbol{\beta}) \le \chi^2_{1-\alpha}(p)\},$$

where $\hat{\boldsymbol{\Sigma}}_3 = \frac{1}{n}\mathbf{g}'(\mathbf{Z}, \hat{\boldsymbol{\beta}}_I)^T(\mathbf{I}_n - \mathbf{S}_{h_2})^T(\mathbf{I}_n - \mathbf{S}_{h_2})\mathbf{g}'(\mathbf{Z}, \hat{\boldsymbol{\beta}}_I)$, $\hat{\boldsymbol{\Sigma}}_4 = \frac{1}{n}[\mathbf{g}'(\mathbf{Z}, \hat{\boldsymbol{\beta}}_I)^T(\mathbf{I}_n - \mathbf{S}_{h_2})^T(\mathbf{I}_n - \mathbf{S}_{h_2})(\check{\mathbf{Y}} - \mathbf{g}(\mathbf{Z}, \hat{\boldsymbol{\beta}}_I))]^{\otimes 2}$ and

$$\mathbf{S}_{h_2} = \begin{pmatrix} (\mathbf{X}_1^T \quad \mathbf{0}_{1\times q})[\mathbf{X}_{h_2}^T(U_1)\mathbf{W}_2(U_1)\mathbf{X}_{h_2}(U_1)]^{-1}\mathbf{X}_{h_2}^T(U_1)\mathbf{W}_2(U_1) \\ \vdots \\ (\mathbf{X}_n^T \quad \mathbf{0}_{1\times q})[\mathbf{X}_{h_2}^T(U_n)\mathbf{W}_2(U_n)\mathbf{X}_{h_2}(U_n)]^{-1}\mathbf{X}_{h_2}^T(U_n)\mathbf{W}_2(U_n) \end{pmatrix}.$$

Theorem 2.2.2 Suppose that the conditions $C1 - C8$ in Section 7 hold. For any $u_0 \in \Omega$, we have

$$\sqrt{nh_2}\left[\hat{\boldsymbol{\Psi}}_I(u_0, \hat{\boldsymbol{\beta}}_I) - \boldsymbol{\Psi}_0(u_0) - \frac{1}{2}h_2^2\mu_{2,2}\begin{pmatrix}\boldsymbol{\theta}_0''(u_0) \\ 0\end{pmatrix}\right] \xrightarrow{D} N(\mathbf{0}, \boldsymbol{\Sigma}_5),$$

where $\boldsymbol{\Sigma}_5 = f^{-1}(u_0)\boldsymbol{\Gamma}^*(u_0)^{-1}E[\pi^{-1}(\mathbf{V})\boldsymbol{\varepsilon}^2\mathbf{X}\mathbf{X}^T|U = u_0]\boldsymbol{\Gamma}^*(u_0)^{-1} \otimes \begin{pmatrix}\nu_{2,0} & 0 \\ 0 & \nu_{2,2}\mu_{2,2}^{-2}\end{pmatrix}$. In particular, we have

$$\sqrt{nh_2}[\hat{\boldsymbol{\theta}}_I(u_0) - \boldsymbol{\theta}_0(u_0) - \frac{1}{2}h_2^2\mu_{2,2}\boldsymbol{\theta}_0''(u_0)] \xrightarrow{D} N(\mathbf{0}, f^{-1}(u_0)\boldsymbol{\Gamma}^*(u_0)^{-1}\nu_{2,0}E[\pi^{-1}(\mathbf{V})\boldsymbol{\varepsilon}^2\mathbf{X}\mathbf{X}^T|U = u_0]\boldsymbol{\Gamma}^*(u_0)^{-1}),$$

where $\hat{\boldsymbol{\theta}}_I(u_0)$ denotes $\hat{\boldsymbol{\theta}}_I(u_0, \hat{\boldsymbol{\beta}}_I)$.

3. The empirical likelihood based on the weighted imputation method

3.1. The empirical likelihood for the unknown parametric vector

Let

$$\frac{\partial Q(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = -2\sum_{i=1}^{n} \widetilde{\mathbf{g}'}(\mathbf{Z}_i, \boldsymbol{\beta})[\tilde{Y}_i - \widetilde{g}(\mathbf{Z}_i, \boldsymbol{\beta})],$$

where $\tilde{Y}_i = \check{Y}_i - \sum_{k=1}^{n} S_{h_2,ik}\check{Y}_k$, $\widetilde{g}(\mathbf{Z}_i, \boldsymbol{\beta}) = g(\mathbf{Z}_i, \boldsymbol{\beta}) - \sum_{k=1}^{n} S_{h_2,ik}g(\mathbf{Z}_k, \boldsymbol{\beta})$, $\widetilde{\mathbf{g}'}(\mathbf{Z}_i, \boldsymbol{\beta}) = \mathbf{g}'(\mathbf{Z}_i, \boldsymbol{\beta}) - \sum_{k=1}^{n} S_{h_2,ik}\mathbf{g}'(\mathbf{Z}_k, \boldsymbol{\beta})$, $\mathbf{g}'(\mathbf{Z}_i, \boldsymbol{\beta}) = \frac{\partial g(\mathbf{Z}_i, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}}$, $S_{h_2,ik}$ is the (i, k)-th element of matrix $\mathbf{S}_{h_2}$. Thus the auxiliary random vector of $\boldsymbol{\beta}$ can be introduced as

$$\boldsymbol{\eta}_i(\boldsymbol{\beta}) = \widetilde{g}'(\mathbf{Z}_i, \boldsymbol{\beta})[\widetilde{Y}_i - \widetilde{g}(\mathbf{Z}_i, \boldsymbol{\beta})].$$

Therefore, we can define the empirical log-likelihood ratio function for $\boldsymbol{\beta}$ as follows

$$L(\boldsymbol{\beta}) = \max\{\sum_{i=1}^{n} \log(np_i)|p_i \ge 0, \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i\boldsymbol{\eta}_i(\boldsymbol{\beta}) = \mathbf{0}\}.$$

By using the Lagrange multiplier method, $L(\boldsymbol{\beta})$) can be expressed as

$$L(\boldsymbol{\beta}) = -\sum_{i=1}^{n} \log(1 + \boldsymbol{\lambda}^T \boldsymbol{\eta}_i(\boldsymbol{\beta})),$$

where $\boldsymbol{\lambda}$ is the lagrange multiplier and satisfies $\frac{1}{n}\sum_{i=1}^{n} \frac{\boldsymbol{\eta}_i(\boldsymbol{\beta})}{1+\boldsymbol{\lambda}^T\boldsymbol{\eta}_i(\boldsymbol{\beta})} = 0$. The maximum empirical likelihood estimator (MELE) $\hat{\boldsymbol{\beta}}_E$ of $\boldsymbol{\beta}$ can be obtained by

$$\hat{\boldsymbol{\beta}}_E = \arg\max_{\boldsymbol{\beta}} L(\boldsymbol{\beta}).$$

The following Theorem 3.1.1 gives the asymptotic behavior of $-2L(\boldsymbol{\beta})$.

Theorem 3.1.1 Suppose that the conditions $C1 - C8$ in Section 7 hold, and $\boldsymbol{\beta}_0$ is the true parameter vector of $\boldsymbol{\beta}$, then

$$-2L(\boldsymbol{\beta}_0) \xrightarrow{D} \chi^2(p).$$

Remark 2: As a consequence of Theorem 3.1.1, for any $0 < \alpha < 1$, the empirical likelihood confidence region of $\boldsymbol{\beta}$ based on the weighted imputation method (IEL) can be constructed as

$$C_{IEL}(\alpha) = \{\boldsymbol{\beta} : -2L(\boldsymbol{\beta}) \leq \chi^2_{1-\alpha}(p)\}.$$

The following Theorem 3.1.2 gives the asymptotic behavior of $\hat{\boldsymbol{\beta}}_E$, which shows that $\hat{\boldsymbol{\beta}}_E$ is asymptotically equivalent to $\hat{\boldsymbol{\beta}}_I$ obtained in Theorem 2.2.1.

Theorem 3.1.2 Suppose that the conditions $C1 - C8$ hold in Section 7, and $\boldsymbol{\beta}_0$ is the true parameter vector of $\boldsymbol{\beta}$, then

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_E - \boldsymbol{\beta}_0) \xrightarrow{D} N(\mathbf{0}, \boldsymbol{\Sigma}_3^{-1}\boldsymbol{\Sigma}_4\boldsymbol{\Sigma}_3^{-1}).$$

3.2. The empirical likelihood for the nonparametric vector

If $\boldsymbol{\beta}$ is known in advance, we notice $E\{\mathbf{X}[\check{\mathbf{Y}} - \mathbf{g}(\mathbf{Z}, \boldsymbol{\beta}) - \mathbf{X}^T\boldsymbol{\theta}(u)]\}f(u) = 0$, where $f(u)$ is the density function of U. Thus, an auxiliary random vector for $\boldsymbol{\theta}(u)$ can be given by

$$\boldsymbol{\zeta}_i(\boldsymbol{\theta}(u)) = \mathbf{X}_i[\check{Y}_i - g(\mathbf{Z}_i, \boldsymbol{\beta}) - \mathbf{X}_i^T\boldsymbol{\theta}(u)]K_{2,h_2}(U_i - u).$$

Noting that $E[\boldsymbol{\zeta}_i(\boldsymbol{\theta}(u))] = o(1), i = 1, \cdots, n$, the empirical log-likelihood ratio for $\boldsymbol{\theta}(u)$ can be introduced by

$$l(\boldsymbol{\theta}(u)) = \max\{\sum_{i=1}^{n} \log(np_i)|p_i \geq 0, \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i\boldsymbol{\zeta}_i(\boldsymbol{\theta}(u)) = \mathbf{0}\}.$$

Because $\boldsymbol{\beta}$ is unknown, we can replace $\boldsymbol{\beta}$ with its MELE $\hat{\boldsymbol{\beta}}_E$, then obtain an estimated auxiliary random vector

$$\tilde{\boldsymbol{\zeta}}_i(\boldsymbol{\theta}(u)) = \mathbf{X}_i[\check{Y}_i - g(\mathbf{Z}_i, \hat{\boldsymbol{\beta}}_E) - \mathbf{X}_i^T\boldsymbol{\theta}(u)]K_{2,h_2}(U_i - u).$$

The corresponding empirical log-likelihood ratio for $\boldsymbol{\theta}(u)$ is

$$\tilde{l}(\boldsymbol{\theta}(u)) = \max\{\sum_{i=1}^{n} \log(np_i)|p_i \geq 0, \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i\tilde{\boldsymbol{\zeta}}_i(\boldsymbol{\theta}(u)) = \mathbf{0}\}.$$

But the distribution of $\tilde{l}(\boldsymbol{\theta}(u))$ is not a standard chi-square distribution except under the case of undersmoothing. Following Zhou, Zhao and Wang [4], we also propose the residual adjusted auxiliary random vector as follows

$$\hat{\boldsymbol{\zeta}}_i(\boldsymbol{\theta}(u)) = \mathbf{X}_i\{\check{Y}_i - g(\mathbf{Z}_i,\hat{\boldsymbol{\beta}}_E) - \mathbf{X}_i^T\boldsymbol{\theta}(u) - \mathbf{X}_i^T[\hat{\boldsymbol{\theta}}_I(U_i,\hat{\boldsymbol{\beta}}_I) - \hat{\boldsymbol{\theta}}_I(u,\hat{\boldsymbol{\beta}}_I)]\}K_{2,h_2}(U_i - u),$$

where $\hat{\boldsymbol{\theta}}_I(\cdot)$ is the profile nonlinear least squares estimator of the weighted imputation method. Therefore, a residual adjusted estimated empirical log-likelihood ratio for $\boldsymbol{\theta}(u)$ can be defined as

$$\hat{l}(\boldsymbol{\theta}(u)) = \max\{\sum_{i=1}^n \log(np_i)|p_i \geq 0, \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i\hat{\boldsymbol{\zeta}}_i(\boldsymbol{\theta}(u)) = \mathbf{0}\}.$$

By using the Lagrange multiplier method, $\hat{l}(\boldsymbol{\theta}(u))$ can be expressed as

$$\hat{l}(\boldsymbol{\theta}(u)) = -\sum_{i=1}^n \log(1 + \boldsymbol{\gamma}^T\hat{\boldsymbol{\zeta}}_i(\boldsymbol{\theta}(u))),$$

where $\boldsymbol{\gamma}$ is the solution of the following equation

$$\frac{1}{n}\sum_{i=1}^n \frac{\hat{\boldsymbol{\zeta}}_i(\boldsymbol{\theta}(u))}{1 + \boldsymbol{\gamma}^T\hat{\boldsymbol{\zeta}}_i(\boldsymbol{\theta}(u))} = 0.$$

Theorem 3.2.1 Suppose that the conditions $C1 - C8$ in Section 7 hold, $\boldsymbol{\theta}_0(u)$ is the true function of $\boldsymbol{\theta}(u)$, then

$$-2\hat{l}(\boldsymbol{\theta}_0(u)) \xrightarrow{D} \chi^2(q).$$

Remark 3: As a consequence of Theorem 3.2.1, for any $0 < \alpha < 1$, the empirical likelihood confidence region of $\boldsymbol{\theta}(u)$ based on the weighted imputation method (IEL) can be constructed as

$$C_{IEL}(\alpha) = \{\boldsymbol{\theta}(u) : -2\hat{l}(\boldsymbol{\theta}(u)) \leq \chi^2_{1-\alpha}(q)\}.$$

4. The response mean estimation

4.1. The normal approximated confidence interval of the response mean

Sometimes $E(Y) = \mu$ is the parameter we're interested in. We propose the weighted imputation method to estimate the response mean, so μ is estimated by

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^n\{\frac{\delta_i}{\hat{\pi}(\mathbf{V}_i)}Y_i + (1 - \frac{\delta_i}{\hat{\pi}(\mathbf{V}_i)})[\mathbf{X}_i^T\hat{\boldsymbol{\theta}}_c(U_i,\hat{\boldsymbol{\beta}}_c) + g(\mathbf{Z}_i,\hat{\boldsymbol{\beta}}_c)]\}.$$

The asymptotic normality of $\hat{\mu}$ is given in the following theorem.

Theorem 4.1.1 Suppose that the conditions $C1 - C8$ hold in Section 7, μ_0 is the true parameter of μ, then

$$\sqrt{n}(\hat{\mu} - \mu_0) \xrightarrow{D} N(0, \Lambda),$$

where $\Lambda = E[\frac{\sigma^2}{\pi(\mathbf{V})}] + Var[\mathbf{X}^T\boldsymbol{\theta}(U) + g(\mathbf{Z},\boldsymbol{\beta})], \mathbf{V} = (\mathbf{X}, U, \mathbf{Z})^T, Var(\boldsymbol{\varepsilon}|\mathbf{V}) = \sigma^2$.

Remark 4: As a consequence of Theorem 4.1.1, we can construct the normal approximated $1-\alpha$ confidence interval based on the weighted imputation method (INA) for μ :

$$C_{INA}(\alpha)=\left(\hat{\mu}-z_{1-\frac{\alpha}{2}}\sqrt{\frac{\hat{\Lambda}}{n}},\hat{\mu}+z_{1-\frac{\alpha}{2}}\sqrt{\frac{\hat{\Lambda}}{n}}\right),$$

where $\hat{\Lambda}=\frac{1}{n}\sum(\breve{Y}_i-\hat{\mu})^2$, $\breve{Y}_i=\frac{\delta_i}{\hat{\pi}(\mathbf{V}_i)}Y_i+(1-\frac{\delta_i}{\hat{\pi}(\mathbf{V}_i)})[\mathbf{X}_i^T\hat{\boldsymbol{\theta}}_c(U_i,\hat{\boldsymbol{\beta}}_c)+g(\mathbf{Z}_i,\hat{\boldsymbol{\beta}}_c)]$.

4.2. The empirical likelihood confidence interval of the response mean

To construct the empirical likelihood ratio of μ, we introduce the auxiliary random vector

$$Y_i^*=\frac{\delta_i}{\pi(\mathbf{V}_i)}Y_i+(1-\frac{\delta_i}{\pi(\mathbf{V}_i)})[\mathbf{X}_i^T\boldsymbol{\theta}_c(U_i,\boldsymbol{\beta}_c)+g(\mathbf{Z}_i,\boldsymbol{\beta}_c)].$$

When the response is missing at random, we have $E(Y_i^*)=\mu$. Since Y_i^* is unknown, we need to replace it by its estimators

$$\breve{Y}_i=\frac{\delta_i}{\hat{\pi}(\mathbf{V}_i)}Y_i+(1-\frac{\delta_i}{\hat{\pi}(\mathbf{V}_i)})[\mathbf{X}_i^T\hat{\boldsymbol{\theta}}_c(U_i,\hat{\boldsymbol{\beta}}_c)+g(\mathbf{Z}_i,\hat{\boldsymbol{\beta}}_c)].$$

Then we can define the empirical log-likelihood ratio function for μ as follows

$$R(\mu)=\max\{\sum_{i=1}^n\log(np_i)|p_i\geq 0,\sum_{i=1}^n p_i=1,\sum_{i=1}^n p_i\breve{Y}_i=\mu\}.$$

By using the Lagrange multiplier method, $R(\mu)$ can be expressed as

$$R(\mu)=-\sum_{i=1}^n\log(1+\rho(\breve{Y}_i-\mu)),$$

where $\rho=\rho(\mu)$ is the lagrange multiplier and satisfies $\frac{1}{n}\sum_{i=1}^n\frac{\breve{Y}_i-\mu}{1+\rho(\breve{Y}_i-\mu)}=0$.

The following Theorem 4.2.1 gives the asymptotic distribution of $-2R(\mu)$.

Theorem 4.2.1 Suppose that the conditions $C1-C8$ in Section 7 hold, μ_0 is the true parameter of μ, then

$$-2R(\mu_0)\xrightarrow{D}\chi^2(1).$$

Remark 5: As a consequence of Theorem 4.2.1, for any $0<\alpha<1$, the empirical likelihood confidence interval of μ based on the weighted imputation method (IEL) can be constructed as

$$C_{IEL}(\alpha)=\{\mu:-2R(\mu)\leq\chi^2_{1-\alpha}(1)\}.$$

5. Simulation study

5.1. Numerical simulation

We conducted simulation studies to assess the performance of the estimation methods in Sections 2, 3 and 4. Assume the data come from the following varying coefficient partially nonlinear model, which is also considered in Li and Mei [1], Zhou, Zhao and Wang [4], Wang, Zhao and Du [7],

$$Y=X\theta(U)+g(Z_1,Z_2;\beta_1,\beta_2)+\varepsilon,$$

where $\theta(U) = \sin(2\pi U)$ and $g(Z_1, Z_2; \beta_1, \beta_2) = \exp\{Z_1\beta_1 + Z_2\beta_2\}$ with $\beta_1 = 1$, $\beta_2 = 2$, $X \sim N(0,1)$, $U \sim U(0,1)$, $Z_1 \sim U(0,1)$, $Z_2 \sim U(0,1)$, $\varepsilon \sim N(0,0.5)$. Throughout this section, the kernel function is taken as the Epanechnikov kernel $K(u) = 0.75(1-u^2)_+$, the bandwidth h_1 and h_2 are taken by the leave-one-sample-out method in the cross-validation criterion, namely, we can get the bandwidth by minimizing the following formula,

$$CV(h) = \frac{1}{n}\sum_{i=1}^{n}[Y_i - g(Z_i, \tilde{\beta}^{[i]}) - X_i^T\tilde{\theta}^{[i]}(U_i)]^2,$$

where $\tilde{\beta}^{[i]}$, $\tilde{\theta}^{[i]}(U_i)$ are the leave-one-out estimators of β and $\theta(\cdot)$, respectively, which are computed with the data, but the i-th observation is deleted. Three selection probability functions are chosen as:

$$\pi_1(x,u,z) = P(\delta = 1|X = x, U = u, Z = (z_1, z_2)) = [1 + exp(-1 + 0.8x - 1.5u + 0.5z_1 - z_2)]^{-1},$$

$$\pi_2(x,u,z) = P(\delta = 1|X = x, U = u, Z = (z_1, z_2)) = [1 + exp(-1 + 0.8x - 1.7u + 2z_1 - z_2)]^{-1},$$

$$\pi_3(x,u,z) = P(\delta = 1|X = x, U = u, Z = (z_1, z_2)) = [1 + exp(-0.7 - x + 0.5u - 0.3z_1 + 0.25z_2)]^{-1}.$$

The missing data are generated by the method described in Section 4 in Wang, Chen and Lin [15], the average missing probability of Y for above three cases is about 15%, 25%, 40%. In the following, the sample size n is taken to be 100, 200, 300, respectively, and 1000 replications for each case.

5.1.1. The performance of parametric confidence regions

For parameter components, we consider two confidence regions of $\boldsymbol{\beta} = (\beta_1, \beta_2)^T$ with nominal level 95%: the normal approximated confidence region of the weighted imputation method (INA) based on Theorem 2.2.1, and the empirical likelihood confidence region of the weighted imputation method (IEL) based on Theorem 3.1.1. The confidence regions of $\boldsymbol{\beta} = (\beta_1, \beta_2)^T$ for above three selection probability functions are presented in Figure 1.

From Figure 1, we have the following conclusions for parametric confidence regions. In the case of the same missing rate, the confidence regions of IEL and INA will become smaller with the increase of sample size, but the confidence regions of IEL are always smaller than that of INA. In addition, in the case of the same sample size, the confidence regions of IEL and INA will be larger with the increase of the missing rate, but the IEL confidence regions always keep smaller than that of INA. In one word, the results in Figure 1 illustrate that the IEL method performs better than the INA method.

5.1.2. The performance of nonparametric confidence regions

For the nonparametric component $\theta(u)$, we also consider two confidence regions with nominal level 95% using two methods: the normal approximated confidence region (INA) based on Theorem 2.2.2 , and the empirical likelihood confidence region (IEL) based on Theorem 3.2.1. The pointwise confidence intervals for $\theta(u)$ are shown in Figure 2.

From Figure 2, we can obtain that, for the same missing rate and sample size, the associated pointwise confidence intervals of IEL method have uniformly shorter average lengths than that of the INA method for $\theta(u)$, which is similar to the parametric components.

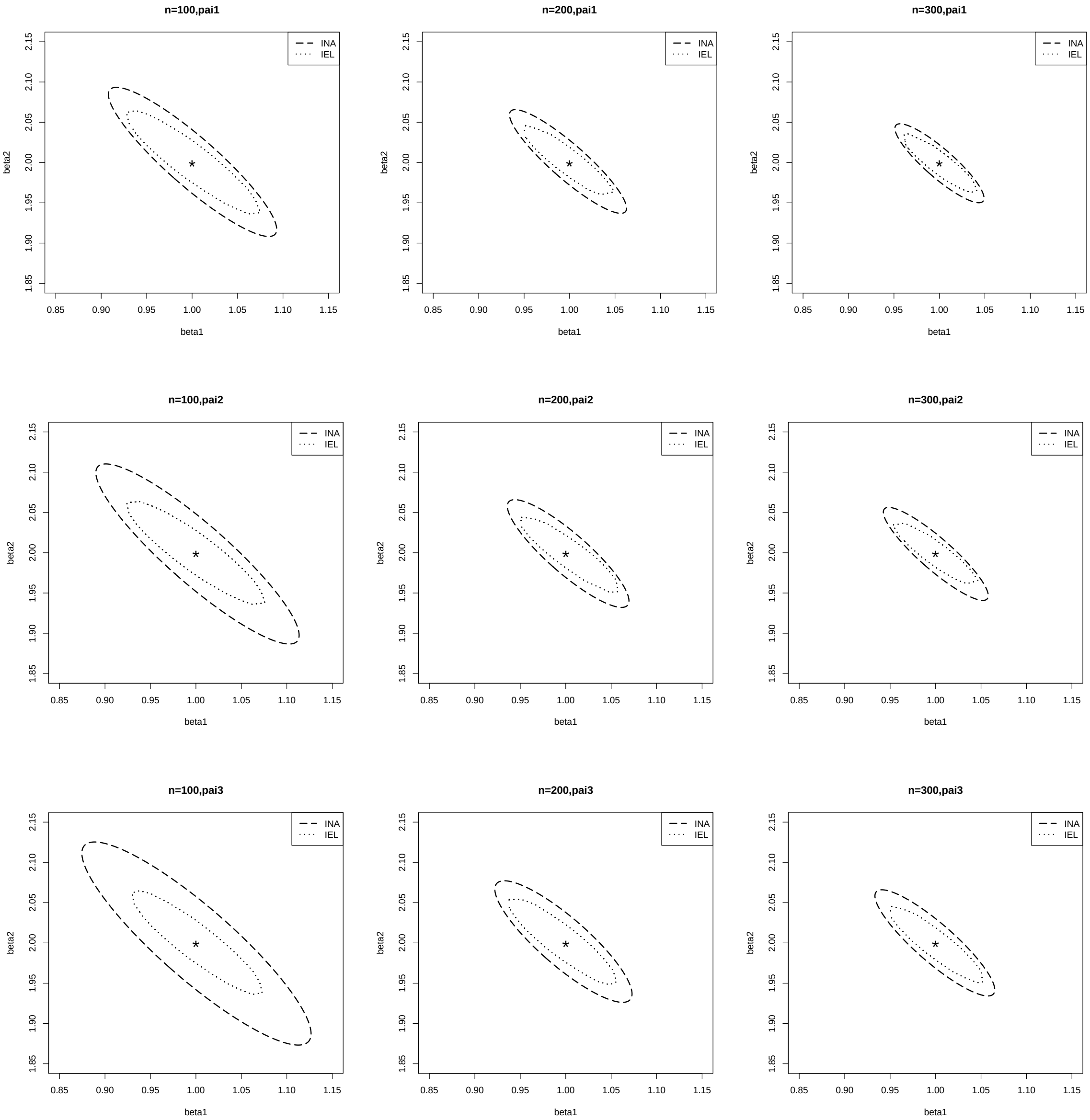

Figure 1. The 95% confidence regions for the parametric component based on the INA method (dashed curves) and IEL method (dotted curves) for three selection probability functions and sample sizes.

5.1.3. The performance of the response mean

For the response mean, we also consider two methods: the normal approximated confidence intervals based on Theorem 4.1.1, and the empirical likelihood confidence intervals based on Theorem 4.2.1. The average lengths and empirical coverage probabilities of the confidence intervals with a nominal level $1 - \alpha = 0.95$ for above three selection probability functions and samples are presented in Table 1.

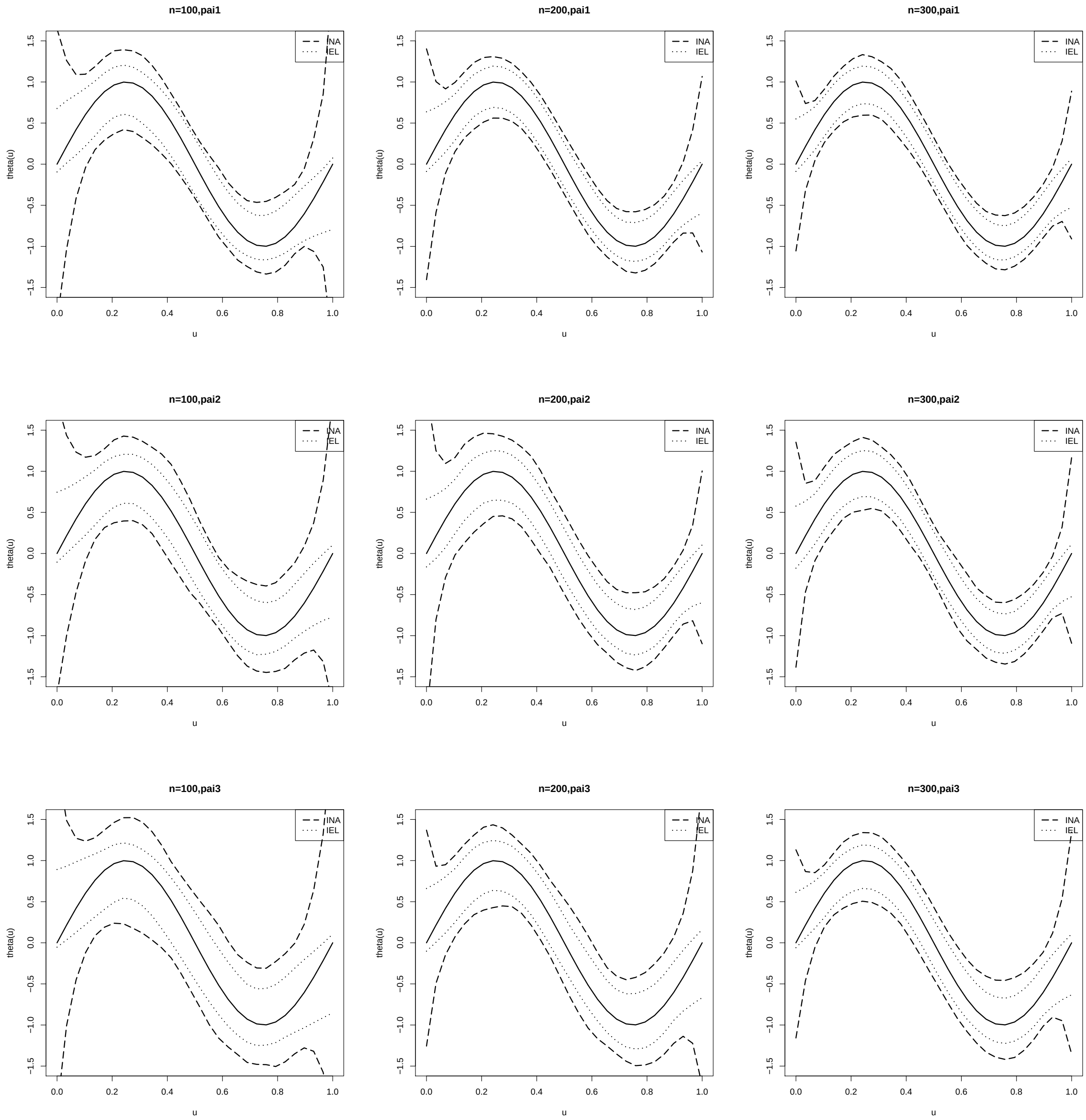

Figure 2. The 95% pointwise confidence intervals for the nonparametric components based on the INA method (dashed curves) and IEL method (dotted curves) for three selection probability functions and sample sizes, where the solid curve is the real curve.

From Table 1, we can have the following findings:

(1) When the missing rate and sample size are the same, the average lengths of the confidence intervals under the IEL method are always shorter than that under the INA method, and the IEL method maintains a higher coverage probabilities than the INA method;

(2) In the case of the same missing rate, with the increase of the sample size, the lengths of the confidence intervals of INA and IEL methods will become shorter and the coverage probabilities will become larger, but IEL method has always maintained a shorter confidence intervals and higher coverage probabilities than INA method;

Table 1. The 95% confidence intervals average lengths (AL) and coverage probabilities (CP) for the response mean based on the INA method and IEL method for three selection probability functions and sample sizes.

		AL (Average lengths)		CP (Coverage probabilities)	
		INA	IEL	INA	IEL
	n = 100	1.4443	1.2189	0.944	0.945
15% missing rate	n = 200	1.0235	0.9896	0.948	0.954
	n = 300	0.8344	0.8303	0.957	0.960
	n = 100	1.4512	1.2222	0.940	0.942
25% missing rate	n = 200	1.0298	0.9934	0.947	0.951
	n = 300	0.8437	0.8378	0.953	0.955
	n = 100	1.4611	1.2264	0.936	0.939
40% missing rate	n = 200	1.0424	1.0029	0.945	0.950
	n = 300	0.8508	0.8442	0.946	0.952

(3) In the case of the same sample size, with the increase of the missing rate, the average lengths of the confidence intervals of INA and IEL will become larger and the coverage rate will become smaller, however, IEL method also keeps shorter confidence interval and higher coverage than INA method.

The simulation results show that IEL method has better advantages than INA method in terms of average lengths and coverage probabilities of confidence intervals. In addition, in the case of the same sample size, with the change of the missing rate, the average lengths of the confidence intervals of INA and IEL only slightly changes, which shows that the weighted imputation method is very effective.

From Figures 1, 2 and Table 1, we can conclude that IEL method outperforms INA method, so we propose IEL method to the following real data example.

5.2. A real data example

In this section, we illustrate the application of the proposed method by analyzing the Boston housing price data which consists of 506 observations from Boston Standard Metropolitan Statistical Area in 1970. This data have also been studied in Li and Mei [1], Zhou, Zhao and Wang [4], and Wang, Zhao and Du [7]. To illustrate the proposed method, we analyze this data by using the following varying coefficient partially nonlinear model:

$$Y = \theta_0(U) + X_1\theta_1(U) + X_2\theta_2(U) + X_3\theta_3(U) + X_4\theta_4(U) + \exp(Z_1\beta_1 + Z_2\beta_2) + \varepsilon,$$

where the response variable Y denotes the median value of owner-occupied homes in \$1000, X_1 denotes the per capita crime rate by town, X_2 denotes the average number of rooms per dwelling, X_3 denotes the full value property tax per \$10,000, X_4 denotes the nitrogen oxide concentration in parts per 10 million, Z_1 denotes the pupil−teacher ratio by town school district, Z_2 denotes the proportion of owner-occupied homes built prior to 1940, and U denotes the square root of the proportion of population that is in the lower status.

We assume that the missing rate of Y is about 20% , and δ is randomly generated in simulation by the method mentioned in Section 4 in Wang, Chen and Lin [15], the kernel function is taken as the Epanechnikov kernel $K(u) = 0.75(1 - u^2)_+$, the bandwidth h_1 and h_2 are also taken by the leave-one-sample-out method in the cross-validation criterion, and the run number is 1000.

Based on our IEL method, the estimator of $\boldsymbol{\beta} = (\beta_1, \beta_2)^T$ is $(-0.3029, 0.0598)^T$, whose signs is consistent with that obtained in Zhou, Zhao and Wang [4], but the estimator value is slightly different because 20% Y is imputed by the weighted imputation method. The confidence region of $\boldsymbol{\beta} = (\beta_1, \beta_2)^T$ and the pointwise confidence intervals for the nonparametric component $\theta_0(u) \sim \theta_4(u)$ with nominal level 95% are shown in Figures 3 and 4, respectively. It can be seen from Figure 4 that, the estimated coefficient functions for each nonparametric components based on IEL are very similar to the Figure 5 in Zhou, Zhao and Wang [4], however, the latter was obtained without missing data. We also calculated the 95% confidence interval of the response mean, it is (21.7773, 23.3864), whose length is 1.6090, and the estimator of the response mean is 22.5528. This further confirms that the proposed IEL method with missing responses is preferable in real data analysis.

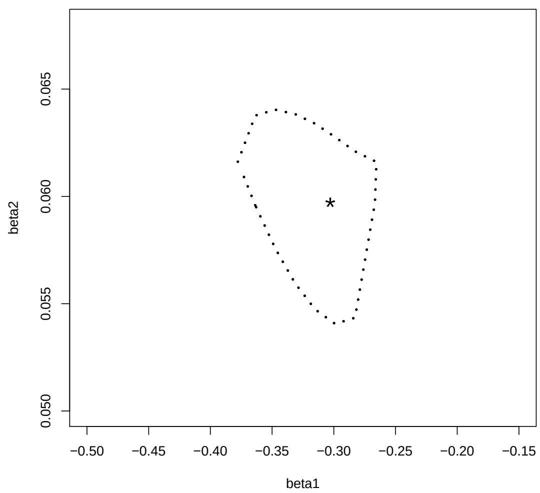

Figure 3. Application to Boston housing price data with 20% Y missing. The 95% confidence region for the parametric components (β_1, β_2).

6. Discussion

In this article, we have proposed a weighted imputation method based on the profile nonlinear least squares and empirical likelihood to the varying coefficient partially nonlinear model with missing responses. The asymptotic properties of our proposals have been obtained under certain conditions. Our simulation studies reveal that empirical likelihood of the weighted imputation method has better advantages than the profile nonlinear least squares.

At last, we put forward some further research topics. First, in this paper, we only focus on estimation of VCPNLM, variable selection and model averaging procedures with missing response are also the

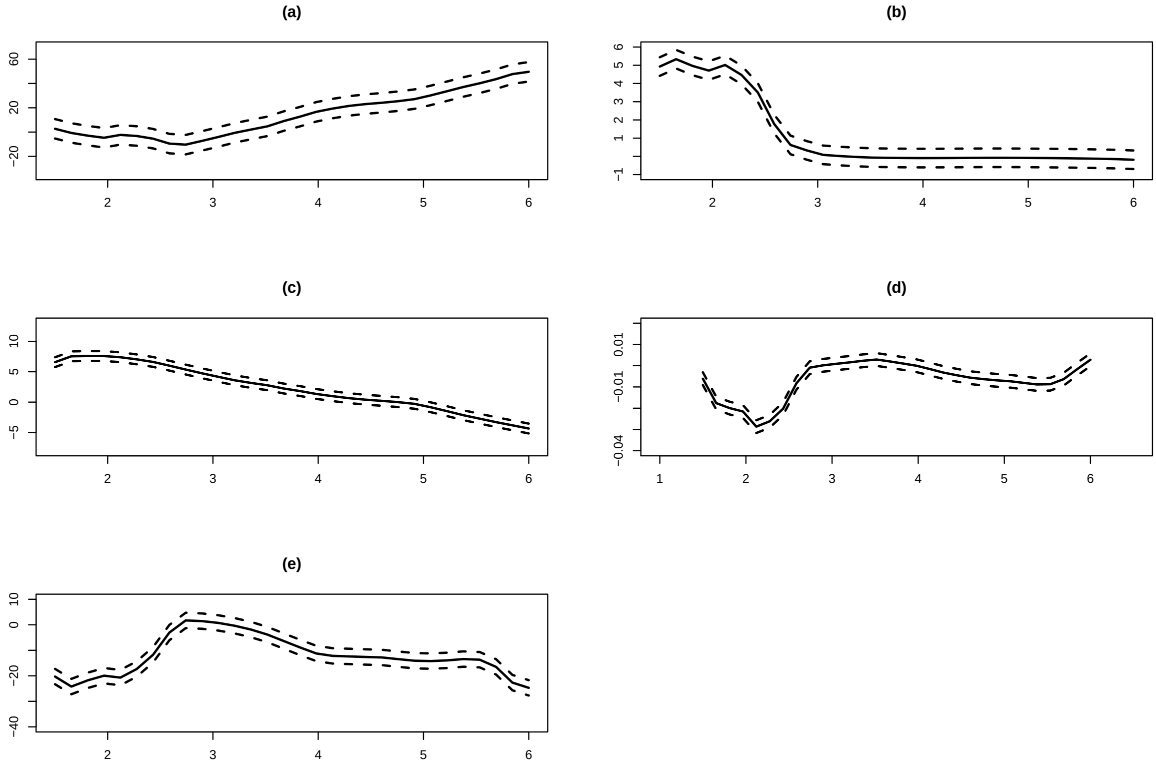

Figure 4. Application to Boston housing price data with 20% Y missing. The 95% pointwise confidence intervals for the nonparametric component $\theta_0(u) \sim \theta_4(u)$, which are presented in figures $(a) \sim (e)$ respectively. Here, the solid curve is the estimated coefficient functions, and the dashed curve is the corresponding 95% pointwise confidence intervals.

directions for future work. Second, when the missing mechanism is not missing at random, or the data has a heavy tailed or skewed distribution, how to use the method proposed in Wang, Song and Lin [21] and Wang, Song and Zhang [22] to VCPNLM is deserved further study. Third, the logistic model for the selection probability function is assumed in our article, when the selection probability is misspecified, how to utilizes the covariate information suggested in Sun, Luan and Jiang [23] to derive a robust estimation of the selection probability is worth further study. At last, how to generalize our method to optimal reinsurance problems of Fang, Cheng and Qu [24] and Fang, Wang, Liu and Li [25] is also an interesting topic.

7. Proofs

The Assumptions required in this paper are following, which can also be found in Li and Mei [1], Zhou, Zhao and Wang [4], Xiao and Chen [6], Wang, Zhao and Du [7] and other missing data literatures.

$C1$: The covariate U has a bounded support Ω, its density function $f(u)$ is Lipschitz continuous and bounded away from 0 on its support Ω.

$C2$: All the coefficient functions $\{\theta_j(U), j = 1, \cdots, q\}$ have continuous second derivatives in $U \in \Omega$.

$C3$: For any $\mathbf{Z}$, $\mathbf{g}(\mathbf{Z},\boldsymbol{\beta})$ is a continuous function of $\boldsymbol{\beta}$, and the second derivatives of $\mathbf{g}(\mathbf{Z},\boldsymbol{\beta})$ with respect to $\boldsymbol{\beta}$ are all continuous.

$C4$: There is an $s > 2.5$ such that $E(\|\mathbf{X}\|^{2s}) < \infty$ and $E(\|\mathbf{g}'(\mathbf{Z},\boldsymbol{\beta})\|^{2s}) < \infty$, and for some $0 < \delta < 2 - s^{-1}$, there is $n^{2\delta-1}h_1 \to \infty$, $n^{2\delta-1}h_2 \to \infty$, where$\|\cdot\|$ represents Euclidean norm.

$C5$: The matrices $E[\mathbf{g}'(\mathbf{Z},\boldsymbol{\beta})^{\otimes 2}]$, $E[E(\mathbf{g}'(\mathbf{Z},\boldsymbol{\beta})|U)^{\otimes 2}]$ and $E[(Vech\{\boldsymbol{g}''(\mathbf{Z},\boldsymbol{\beta})\}|U)^{\otimes 2}]$ are all bounded in a neighborhood of $\boldsymbol{\beta}$. $E\|\mathbf{g}'(\mathbf{Z},\boldsymbol{\beta})\|^4 < \infty$ and $E\|Vech\boldsymbol{g}''(\mathbf{Z},\boldsymbol{\beta})\|^4 < \infty$.

$C6$: The kernel functions $K_1(\cdot)$ and $K_2(\cdot)$ are symmetric density functions with compact support. Furthermore, they both satisfy the Lipschitz condition. The functions $u^3K_i(u)$ and $u^3K_i'(u)$ are bounded and $\int u^4K_i(u)du < \infty, i = 1,2$. The bandwidths h_i satisfy $nh_i^8 \to 0$, $nh_i^2/(logn)^2 \to \infty$, $h_i \to 0$, $nh_i \to \infty, i = 1,2$, as $n \to \infty$.

$C7$: The $q \times q$ matrix $\boldsymbol{\Gamma}(U)$ and $\boldsymbol{\Gamma}^*(U)$ are nonsingular for each $U \in \Omega$. The matrices $\boldsymbol{\Gamma}(U)$, $\boldsymbol{\Gamma}^*(U)$, $\boldsymbol{\Gamma}(U)^{-1}$ $\boldsymbol{\Gamma}^*(U)^{-1}$, $\boldsymbol{\Phi}(U)$, $\boldsymbol{\Phi}^*(U)$ are all Lipschitz continuous.

$C8$: Under the support of $(\mathbf{X}_i, U_i, \mathbf{Z}_i)$, $\pi(\cdot)$ is bounded away from 0 and has a continuous two order partial derivative.

Lemma 1. Let $(X_1, Y_1), \cdots, (X_n, Y_n)$ be i.i.d. random vectors, where Y_i is scalar random variable. Further assume that $E|Y_i|^s < \infty$ and $\sup_x \int |y|^s f(x,y)dy < \infty$, where $f(.,.)$ denotes the joint density of (X, Y). Let $K(\cdot)$ be a bounded positive function with a bounded support, satisfying the Lipschitz condition. Give that $n^{2\delta-1}h \to \infty$ for some $\delta < 1 - s^{-1}$, then

$$\sup_x |\frac{1}{n}\sum_{i=1}^n \{K_h(X_i - x)Y_i - E[K_h(X_i - x)Y_i]\}| = O_p(\{\frac{log(1/h)}{nh}\}^{\frac{1}{2}}).$$

Proof: This Lemma follows from the result that was obtained by Mack and Silverman [19].

Lemma 2. Suppose that Conditions $C1 - C8$ hold, then as $n \longrightarrow \infty$, we have

$$\frac{1}{n}\mathbf{X}_{h_1}(u)^T\mathbf{W}_1(u)\boldsymbol{\Delta}\mathbf{X}_{h_1}(u) = f(u)\boldsymbol{\Gamma}(u) \otimes \begin{pmatrix}1 & 0\\ 0 & \mu_{1,2}\end{pmatrix}[1 + O_p(c_n)], \tag{7.1}$$

$$\frac{1}{n}\mathbf{X}_{h_1}(u)^T\mathbf{W}_1(u)\boldsymbol{\Delta}\mathbf{g}'(\mathbf{Z},\boldsymbol{\beta}_0) = f(u)\boldsymbol{\Phi}(u) \otimes (1,0)^T[1 + O_p(c_n)], \tag{7.2}$$

$$\frac{1}{n}\mathbf{X}_{h_1}(u)^T\mathbf{W}_1(u)\boldsymbol{\Delta}M_0 = \boldsymbol{\Gamma}(u)\boldsymbol{\theta}_0(u)f(u) \otimes (1,0)^T[1 + O_p(c_n)], \tag{7.3}$$

$$\frac{1}{n}\mathbf{X}_{h_1}(u)^T\mathbf{W}_1(u)\boldsymbol{\Delta}\boldsymbol{\varepsilon} = O_p(\{\frac{log(1/h_1)}{nh_1}\}^{\frac{1}{2}}), \tag{7.4}$$

where $c_n = h_1^2 + \{\frac{log(1/h_1)}{nh_1}\}^{\frac{1}{2}}$.

Proof: It is easy to obtain the following result

$$\frac{1}{n}\mathbf{X}_{h_1}(u)^T\mathbf{W}_1(u)\boldsymbol{\Delta}\mathbf{X}_{h_1}(u) = \begin{pmatrix}\frac{1}{n}\sum_{i=1}^n \delta_i\mathbf{X}_i\mathbf{X}_i^TK_{1,h_1}(U_i - u) & \frac{1}{n}\sum_{i=1}^n \delta_i\mathbf{X}_i\mathbf{X}_i^T(\frac{U_i-u}{h_1})K_{1,h_1}(U_i - u)\\ \frac{1}{n}\sum_{i=1}^n \delta_i\mathbf{X}_i\mathbf{X}_i^T(\frac{U_i-u}{h_1})K_{1,h_1}(U_i - u) & \frac{1}{n}\sum_{i=1}^n \delta_i\mathbf{X}_i\mathbf{X}_i^T(\frac{U_i-u}{h_1})^2K_{1,h_1}(U_i - u)\end{pmatrix}. \tag{7.5}$$

By Lemma 1 and the law of iterated expectations, we can easily derive

$$\frac{1}{n}\sum_{i=1}^n \delta_i\mathbf{X}_i\mathbf{X}_i^TK_{1,h_1}(U_i - u) = \boldsymbol{\Gamma}(u)f(u) + O(h_1^2) + O_p(\{\frac{log(1/h_1)}{nh_1}\}^{\frac{1}{2}}) = \boldsymbol{\Gamma}(u)f(u)[1 + O_p(c_n)]. \tag{7.6}$$

Similar to the proof of Eq (7.6), we have

$$\frac{1}{n}\sum_{i=1}^{n}\delta_i\mathbf{X}_i\mathbf{X}_i^T(\frac{U_i-u}{h_1})K_{1,h_1}(U_i-u)=O(h_1^2)+O_p(\{\frac{log(1/h_1)}{nh_1}\}^{\frac{1}{2}}),$$

$$\frac{1}{n}\sum_{i=1}^{n}\delta_i\mathbf{X}_i\mathbf{X}_i^T(\frac{U_i-u}{h_1})^2K_{1,h_1}(U_i-u)=\mathbf{\Gamma}(u)f(u)\mu_{1,2}[1+O_p(c_n)].$$

Hence, Eq (7.1) have been proofed. Similarly, Eqs (7.2), (7.3) and (7.4) can be proved.

Lemma 3. Suppose that Conditions $C1-C8$ hold, we have

$$\frac{1}{n}\mathbf{g}'(\mathbf{Z},\boldsymbol{\beta}_0)^T(\mathbf{I}_n-\mathbf{S}_{h_1})^T\mathbf{\Delta}(\mathbf{I}_n-\mathbf{S}_{h_1})\mathbf{g}'(\mathbf{Z},\boldsymbol{\beta}_0)=\mathbf{\Sigma}\{1+o_p(1)\},\tag{7.7}$$

where $\mathbf{\Sigma}=E[\pi(\mathbf{V})\mathbf{g}'(\mathbf{Z},\boldsymbol{\beta}_0)\mathbf{g}'(\mathbf{Z},\boldsymbol{\beta}_0)^T]-E[\mathbf{\Phi}(U)^T\mathbf{\Gamma}(U)^{-1}\mathbf{\Phi}(U)]$,

$$\frac{1}{n}\mathbf{g}'(\mathbf{Z},\boldsymbol{\beta}_0)^T(\mathbf{I}_n-\mathbf{S}_{h_1})^T\mathbf{\Delta}(\mathbf{I}_n-\mathbf{S}_{h_1})[\mathbf{Y}-\mathbf{g}(\mathbf{Z},\boldsymbol{\beta}_0)]=\xi_n+O_p(c_n^2),\tag{7.8}$$

where $\boldsymbol{\xi}_n=\frac{1}{n}\sum_{i=1}^{n}\delta_i\{\mathbf{g}'(\mathbf{Z}_i,\boldsymbol{\beta}_0)-E[\mathbf{g}'(\mathbf{Z}_i,\boldsymbol{\beta}_0)\mathbf{X}_i^T|U=U_i]\mathbf{\Gamma}(U_i)^{-1}\mathbf{X}_i\}\varepsilon_i$.

Proof: By combining (7.1) with (7.2), we deduce that

$$(\mathbf{X}^T,0)[\mathbf{X}_{h_1}(u)^T\mathbf{W}_1(u)\mathbf{\Delta}\mathbf{X}_{h_1}(u)]^{-1}\mathbf{X}_{h_1}(u)^T\mathbf{W}_1(u)\mathbf{\Delta}\mathbf{g}'(\mathbf{Z},\boldsymbol{\beta}_0)=\mathbf{X}^T\mathbf{\Gamma}(U)^{-1}\mathbf{\Phi}(u)[1+O_p(c_n)].$$

By the weak law of large numbers and above formula, we have

$$\frac{1}{n}\mathbf{g}'(\mathbf{Z},\boldsymbol{\beta}_0)^T(\mathbf{I}_n-\mathbf{S}_{h_1})^T\mathbf{\Delta}(\mathbf{I}_n-\mathbf{S}_{h_1})\mathbf{g}'(\mathbf{Z},\boldsymbol{\beta}_0)=\mathbf{\Sigma}(1+o_p(1))+O_p(c_n^2)E[\mathbf{\Phi}(U)^T\mathbf{\Gamma}(U)^{-1}\mathbf{\Phi}(U)][1+o_p(1)].$$

Under the condition $C6$, $c_n^2=o_p(1)$, hence, Eq (7.7) holds.

Next we prove the Eq (7.8), let $\mathbf{M}_0=(\mathbf{X}_1^T\boldsymbol{\theta}_0(U_1),\cdots,\mathbf{X}_n^T\boldsymbol{\theta}_0(U_n))^T$, with the help of Eqs (7.1) and (7.3), we have

$$(\mathbf{X}^T,0)[\mathbf{X}_{h_1}(u)^T\mathbf{W}_1(u)\mathbf{\Delta}\mathbf{X}_{h_1}(u)]^{-1}\mathbf{X}_{h_1}(u)^T\mathbf{W}_1(u)\mathbf{\Delta}\mathbf{M}_0=\mathbf{X}^T\boldsymbol{\theta}_0(u)[1+O_p(c_n)],$$

so

$$\frac{1}{n}\mathbf{g}'(\boldsymbol{\beta}_0)^T(\mathbf{I}_n-\mathbf{S}_{h_1})^T\mathbf{\Delta}(\mathbf{I}_n-\mathbf{S}_{h_1})\mathbf{M}_0=E[\mathbf{\Phi}(U)^T\boldsymbol{\theta}_0(U)][1+o_p(1)]O_p(c_n^2)=O_p(c_n^2).$$

By Eqs (7.1), (7.4) and $E(\boldsymbol{\varepsilon}|\mathbf{V})=0$, we can derive

$$(\mathbf{X}^T,0)[\mathbf{X}_{h_1}(u)^T\mathbf{W}_1(u)\mathbf{\Delta}\mathbf{X}_{h_1}(u)]^{-1}\mathbf{X}_{h_1}(u)^T\mathbf{W}_1(u)\mathbf{\Delta}\boldsymbol{\varepsilon}=O_p(c_n).$$

Then

$$\frac{1}{n}\mathbf{g}'(\mathbf{Z},\boldsymbol{\beta}_0)^T(\mathbf{I}_n-\mathbf{S}_{h_1})^T\mathbf{\Delta}(\mathbf{I}_n-\mathbf{S}_{h_1})\boldsymbol{\varepsilon}=\boldsymbol{\xi}_n+O_p(c_n^2).$$

So

$$\frac{1}{n}\mathbf{g}'(\mathbf{Z},\boldsymbol{\beta}_0)^T(\mathbf{I}_n-\mathbf{S}_{h_1})^T\mathbf{\Delta}(\mathbf{I}_n-\mathbf{S}_{h_1})[\mathbf{Y}-\mathbf{g}(\mathbf{Z},\boldsymbol{\beta}_0)]=\frac{1}{n}\mathbf{g}'(\mathbf{Z},\boldsymbol{\beta}_0)^T(\mathbf{I}_n-\mathbf{S}_{h_1})^T\mathbf{\Delta}(\mathbf{I}_n-\mathbf{S}_{h_1})(\mathbf{M}_0+\boldsymbol{\varepsilon})=\boldsymbol{\xi}_n+O_p(c_n^2).$$

Hence Eq (7.8) is completed.

The proof of Theorem 2.1.1. We first prove the consistency of $\hat{\boldsymbol{\beta}}_c$. It suffices to prove that for any sufficiently small a,

$$\lim_{n\to\infty} P[\sup_{\|\mathbf{t}\|=a} Q_c(\boldsymbol{\beta}_0+\mathbf{t}) > Q_c(\boldsymbol{\beta}_0)] = 1. \tag{7.9}$$

With Taylor expansion to $Q_c(\boldsymbol{\beta}_0)$, we have $Q_c(\boldsymbol{\beta}_0+\mathbf{t}) - Q_c(\boldsymbol{\beta}_0) = \mathbf{Q}_c'(\boldsymbol{\beta}_0)^T\mathbf{t} + \frac{1}{2}\mathbf{t}^T\mathbf{Q}_c''(\boldsymbol{\beta}^*)\mathbf{t}$, where $\boldsymbol{\beta}^* \in (\boldsymbol{\beta}_0, \boldsymbol{\beta}_0+\mathbf{t})$, $\mathbf{Q}_c'(\boldsymbol{\beta}_0)$ is the first derivative of $Q_c(\boldsymbol{\beta})$ at $\boldsymbol{\beta}_0$, and $\mathbf{Q}_c''(\boldsymbol{\beta}^*)$ is the second derivative at $\boldsymbol{\beta}^*$, then we can prove $\mathbf{t}^T\mathbf{Q}_c''(\boldsymbol{\beta}^*)\mathbf{t} = 2n\{\mathbf{t}^T\boldsymbol{\Sigma}t + O_p(\|\mathbf{t}\|^3)\}$, this proof process is similar to the proof of Theorem 1 in Li and Mei [1], we omit it here. Next we prove the asymptotic normality of $\hat{\boldsymbol{\beta}}_c$.
Note that

$$Q_c(\boldsymbol{\beta}) = \sum_{i=1}^n \delta_i[Y_i - \mathbf{X}_i^T\hat{\boldsymbol{\theta}}_c(U_i,\boldsymbol{\beta}) - g(\mathbf{Z}_i,\boldsymbol{\beta})]^2 = [\mathbf{Y}-\mathbf{g}(\mathbf{Z},\boldsymbol{\beta})]^T(\mathbf{I}_n-\mathbf{S}_{h_1})^T\boldsymbol{\Delta}(\mathbf{I}_n-\mathbf{S}_{h_1})[\mathbf{Y}-\mathbf{g}(\mathbf{Z},\boldsymbol{\beta})],$$

and

$$\mathbf{Q}_c'(\boldsymbol{\beta}_0) = -2\mathbf{g}'(\mathbf{Z},\boldsymbol{\beta}_0)^T(\mathbf{I}_n-\mathbf{S}_{h_1})^T\boldsymbol{\Delta}(\mathbf{I}_n-\mathbf{S}_{h_1})[\mathbf{Y}-\mathbf{g}(\mathbf{Z},\boldsymbol{\beta}_0)],$$

by Eq (7.8), we can show

$$\frac{1}{n}\mathbf{Q}_c'(\boldsymbol{\beta}_0) = -\frac{2}{n}\mathbf{g}'(\mathbf{Z},\boldsymbol{\beta}_0)^T(\mathbf{I}_n-\mathbf{S}_{h_1})^T\boldsymbol{\Delta}(\mathbf{I}_n-\mathbf{S}_{h_1})[\mathbf{Y}-\mathbf{g}(\mathbf{Z},\boldsymbol{\beta}_0)] = -2\boldsymbol{\xi}_n + O_p(c_n^2).$$

It follows from the Taylor's expansion that

$$\mathbf{0} = \mathbf{Q}_c'(\hat{\boldsymbol{\beta}}_c) = \mathbf{Q}_c'(\boldsymbol{\beta}_0) + \mathbf{Q}_c''(\boldsymbol{\beta}^*)(\hat{\boldsymbol{\beta}}_c - \boldsymbol{\beta}_0), \quad \frac{1}{2n}\mathbf{Q}_c''(\boldsymbol{\beta}^*) = \boldsymbol{\Sigma}\{1+o_p(1)\}.$$

Thus, $\sqrt{n}\boldsymbol{\Sigma}\{1+o_p(1)\}(\hat{\boldsymbol{\beta}}_c-\boldsymbol{\beta}_0) = \sqrt{n}\boldsymbol{\xi}_n + O_p(\sqrt{n}c_n^2) = \sqrt{n}\boldsymbol{\xi}_n + o_p(1)$.
Note that, $\sqrt{n}\boldsymbol{\xi}_n = \frac{1}{\sqrt{n}}\sum_{i=1}^n \delta_i\{\mathbf{g}'(\mathbf{Z}_i,\boldsymbol{\beta}_0) - E[\mathbf{g}'(\mathbf{Z}_i,\boldsymbol{\beta}_0)\mathbf{X}_i^T|U=u_i]\boldsymbol{\Gamma}(U_i)^{-1}\mathbf{X}_i\}\varepsilon_i$.
By the Slutsky theorem and the central limit theorem, we have

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_c-\boldsymbol{\beta}_0) \xrightarrow{D} N(\mathbf{0},\sigma^2\boldsymbol{\Sigma}^{-1}).$$

The proof of Theorem 2.1.2. By the definition of $\hat{\boldsymbol{\Psi}}_c(u_0,\hat{\boldsymbol{\beta}}_c)$, we obtain

$$\begin{aligned}
\hat{\boldsymbol{\Psi}}_c(u_0,\boldsymbol{\beta}_c) &= [\mathbf{X}_{h_1}^T(u_0)\mathbf{W}_1(u_0)\boldsymbol{\Delta}\mathbf{X}_{h_1}(u_0)]^{-1}\mathbf{X}_{h_1}^T(u_0)\mathbf{W}_1(u_0)\boldsymbol{\Delta}[\mathbf{Y}-\mathbf{g}(\mathbf{Z},\hat{\boldsymbol{\beta}}_c)]\\
&= [\mathbf{X}_{h_1}^T(u_0)\mathbf{W}_1(u_0)\boldsymbol{\Delta}\mathbf{X}_{h_1}(u_0)]^{-1}\mathbf{X}_{h_1}^T(u_0)\mathbf{W}_1(u_0)\boldsymbol{\Delta}\{\mathbf{M}_0+\boldsymbol{\varepsilon}-[\mathbf{g}(\mathbf{Z},\hat{\boldsymbol{\beta}}_c)-\mathbf{g}(\mathbf{Z},\boldsymbol{\beta}_0)]\}\\
&= [\mathbf{X}_{h_1}^T(u_0)\mathbf{W}_1(u_0)\boldsymbol{\Delta}\mathbf{X}_{h_1}(u_0)]^{-1}\mathbf{X}_{h_1}^T(u_0)\mathbf{W}_1(u_0)\boldsymbol{\Delta}\mathbf{M}_0\\
&+ [\mathbf{X}_{h_1}^T(u_0)\mathbf{W}_1(u_0)\boldsymbol{\Delta}\mathbf{X}_{h_1}(u_0)]^{-1}\mathbf{X}_{h_1}^T(u_0)\mathbf{W}_1(u_0)\boldsymbol{\Delta}\boldsymbol{\varepsilon}\\
&- [\mathbf{X}_{h_1}^T(u_0)\mathbf{W}_1(u_0)\boldsymbol{\Delta}\mathbf{X}_{h_1}(u_0)]^{-1}\mathbf{X}_{h_1}^T(u_0)\mathbf{W}_1(u_0)\boldsymbol{\Delta}[\mathbf{g}(\mathbf{Z},\hat{\boldsymbol{\beta}}_c)-\mathbf{g}(\mathbf{Z},\boldsymbol{\beta}_0)]\\
&\hat{=} I_1+I_2-I_3.
\end{aligned}$$

First, we calculate I_3,

$$\begin{aligned}
I_3 &= [\mathbf{X}_{h_1}^T(u_0)\mathbf{W}_1(u_0)\boldsymbol{\Delta}\mathbf{X}_{h_1}(u_0)]^{-1}\mathbf{X}_{h_1}^T(u_0)\mathbf{W}_1(u_0)\boldsymbol{\Delta}[\mathbf{g}'(\mathbf{Z},\boldsymbol{\beta}_0)(\hat{\boldsymbol{\beta}}_c-\boldsymbol{\beta}_0)+O_p\|\hat{\boldsymbol{\beta}}_c-\boldsymbol{\beta}_0\|^2]\\
&= \boldsymbol{\Gamma}^{-1}(u_0)\{\boldsymbol{\Phi}(u_0)\otimes(1,0)^T + E(\mathbf{X}|U=u_0)O_p(n^{-\frac{1}{2}})\}[1+O_p(c_n)]O_p(n^{-\frac{1}{2}})
\end{aligned}$$

$$= O_p(n^{-\frac{1}{2}}).$$

Next we consider I_1, since,

$\mathbf{X}_{h_1}^T(u_0)\mathbf{W}_1(u_0)\mathbf{\Delta}\mathbf{M}_0$

$$\begin{aligned}
&= \mathbf{X}_{h_1}^T(u_0)\mathbf{W}_1(u_0)\mathbf{\Delta}\left(\begin{pmatrix}\mathbf{X}_1^T\boldsymbol{\theta}_0(u_0) + (U_1 - u_0)\mathbf{X}_1^T\boldsymbol{\theta}_0'(u_0) + \frac{1}{2}(U_1 - u_0)^2\mathbf{X}_1^T\boldsymbol{\theta}_0''(u_0)\\ \vdots \\ \mathbf{X}_n^T\boldsymbol{\theta}_0(u_0) + (U_n - u_0)\mathbf{X}_n^T\boldsymbol{\theta}_0'(u_0) + \frac{1}{2}(U_n - u_0)^2\mathbf{X}_n^T\boldsymbol{\theta}_0''(u_0)\end{pmatrix} + o(h_1^2)\right)\\
&= \mathbf{X}_{h_1}^T(u_0)\mathbf{W}_1(u_0)\mathbf{\Delta}\mathbf{X}_{h_1}(u_0)\mathbf{\Psi}_0(u_0) + \frac{1}{2}h_1^2\mathbf{X}_{h_1}^T(u_0)\mathbf{W}_1(u_0)\mathbf{\Delta}\mathbf{A}(u_0) + \mathbf{X}_{h_1}^T(u_0)\mathbf{W}_1(u_0)\mathbf{\Delta}\mathbf{I}_n o(h_1^2),
\end{aligned}$$

where

$$\mathbf{A}(u_0) = \begin{pmatrix}(\frac{U_1-u_0}{h_1})^2\mathbf{X}_1^T\boldsymbol{\theta}_0''(u_0)\\ \vdots \\ (\frac{U_1-u_0}{h_1})\mathbf{X}_n^T\boldsymbol{\theta}_0''(u_0)\end{pmatrix},$$

and

$$\mathbf{X}_{h_1}^T(u_0)\mathbf{W}_1(u_0)\mathbf{\Delta}\mathbf{A}(u_0) = nf(u_0)\mathbf{\Gamma}(u_0)\otimes(\mu_{1,2},0)^T[1 + O_p(c_n)]\boldsymbol{\theta}''(u_0),$$

$$\mathbf{X}_{h_1}^T(u_0)\mathbf{W}_1(u_0)\mathbf{\Delta}\mathbf{I}_n o(h_1^2) = nf(u_0)E(\pi(\mathbf{V})\mathbf{X}|U = u_0)\otimes(1,0)^T[1 + O_p(c_n)]o(h_1^2).$$

Then,

$$\begin{aligned}
I_1 &= [\mathbf{X}_{h_1}^T(u_0)\mathbf{W}_1(u_0)\mathbf{\Delta}\mathbf{X}_{h_1}(u_0)]^{-1}\mathbf{X}_{h_1}^T(u_0)\mathbf{W}_1(u_0)\mathbf{\Delta}\mathbf{X}_{h_1}(u_0)\mathbf{\Psi}_0(u_0)\\
&+ \frac{1}{2}h_1^2[\mathbf{X}_{h_1}^T(u_0)\mathbf{W}_1(u_0)\mathbf{\Delta}\mathbf{X}_{h_1}(u_0)]^{-1}nf(u_0)\mathbf{\Gamma}(u_0)\otimes(\mu_{1,2},0)^T[1 + O_p(c_n)]\boldsymbol{\theta}''(u_0)\\
&+ [\mathbf{X}_{h_1}^T(u_0)\mathbf{W}_1(u_0)\mathbf{\Delta}\mathbf{X}_{h_1}(u_0)]^{-1}nf(u_0)E(\pi(\mathbf{V})\mathbf{X}|U = u_0)\otimes(1,0)^T[1 + O_p(c_n)]o(h_1^2)\\
&= \mathbf{\Psi}_0(u_0) + \frac{1}{2}h_1^2\mu_{1,2}\begin{pmatrix}\boldsymbol{\theta}_0''(u_0)\\ 0\end{pmatrix} + o_p(h_1^2).
\end{aligned}$$

Hence, $\hat{\mathbf{\Psi}}_c(u_0,\hat{\boldsymbol{\beta}}_c) = \mathbf{\Psi}_0(u_0) + \frac{1}{2}h_1^2\mu_{1,2}\begin{pmatrix}\boldsymbol{\theta}_0''(u_0)\\ 0\end{pmatrix} + o(h_1^2) + I_2 + O_p(n^{-\frac{1}{2}})$.

Namely, $\sqrt{nh_1}\left[\hat{\mathbf{\Psi}}_c(u_0,\hat{\boldsymbol{\beta}}_c) - \mathbf{\Psi}_0(u_0) - \frac{1}{2}h_1^2\mu_{1,2}\begin{pmatrix}\boldsymbol{\theta}_0''(u_0)\\ 0\end{pmatrix}\right] = \sqrt{nh_1}I_2 + O_p(\sqrt{nh_1^5} + \sqrt{h_1})$.

It is easy to derive

$$I_2 = \frac{1}{n}f^{-1}(u_0)\mathbf{\Gamma}^{-1}(u_0)\otimes\begin{pmatrix}1 & 0\\ 0 & \mu_{1,2}^{-1}\end{pmatrix}[1 + O_p(c_n)]^{-1}\mathbf{X}_{h_1}^T(u_0)\mathbf{W}_1(u_0)\mathbf{\Delta}\boldsymbol{\varepsilon}.$$

Since $\frac{1}{n}\mathbf{X}_{h_1}^T(u_0)\mathbf{W}_1(u_0)\mathbf{\Delta}\boldsymbol{\varepsilon} = \begin{pmatrix}\frac{1}{n}\sum_{i=1}^n\delta_i\varepsilon_i\mathbf{X}_iK_{1,h_1}(U_i - u_0)\\ \frac{1}{n}\sum_{i=1}^n\delta_i\varepsilon_i\mathbf{X}_i\frac{U_i-u_0}{h_1}K_{1,h_1}(U_i - u_0)\end{pmatrix}$,

and $\sqrt{nh_1}\frac{1}{n}\mathbf{X}_{h_1}^T(u_0)\mathbf{W}_1(u_0)\mathbf{\Delta}\boldsymbol{\varepsilon} \xrightarrow{D} N(\mathbf{0},\mathbf{\Sigma}^*)$, where $\mathbf{\Sigma}^* = \sigma^2\mathbf{\Gamma}(u_0)f(u_0)\otimes\begin{pmatrix}\nu_{1,0} & 0\\ 0 & \nu_{1,2}\end{pmatrix}$.

Lastly, we can obtain

$$\sqrt{nh_1}I_2 = f^{-1}(u_0)\mathbf{\Gamma}^{-1}(u_0)\otimes\begin{pmatrix}1 & 0\\ 0 & \mu_{1,2}^{-1}\end{pmatrix}[1+O_p(c_n)]^{-1}\sqrt{nh_1}\frac{1}{n}\mathbf{X}_{h_1}^T(u_0)\mathbf{W}_1(u_0)\mathbf{\Delta}\boldsymbol{\varepsilon}$$
$$\xrightarrow{D} N(\mathbf{0},\sigma^2 f^{-1}(u_0)\mathbf{\Gamma}^{-1}(u_0)\otimes\begin{pmatrix}\nu_{1,0} & 0\\ 0 & \nu_{1,2}\mu_{1,2}^{-2}\end{pmatrix}).$$

Now, the proof of Theorem 2.1.2 is completed.

Lemma 4. Suppose that Conditions $C1-C8$ hold, then as $n\longrightarrow\infty$, we have

$$\frac{1}{n}\mathbf{X}_{h_2}(u)^T\mathbf{W}_2(u)\mathbf{X}_{h_2}(u) = f(u)\mathbf{\Gamma}^*(u)\otimes\begin{pmatrix}1 & 0\\ 0 & \mu_{2,2}\end{pmatrix}[1+O_p(c_n^*)], \tag{7.10}$$

$$\frac{1}{n}\mathbf{X}_{h_2}(u)^T\mathbf{W}_2(u)\mathbf{g}'(\mathbf{Z},\boldsymbol{\beta}_0) = f(u)\mathbf{\Phi}^*(u)\otimes(1,0)^T[1+O_p(c_n^*)], \tag{7.11}$$

$$\frac{1}{n}\mathbf{X}_{h_2}(u)^T\mathbf{W}_2(u)\mathbf{M}_0 = \mathbf{\Gamma}^*(u)\boldsymbol{\theta}_0(u)f(u)\otimes(1,0)^T[1+O_p(c_n^*)], \tag{7.12}$$

$$\frac{1}{n}\mathbf{X}_{h_2}(u)^T\mathbf{W}_2(u)\boldsymbol{\varepsilon} = O_p(\{\frac{log(1/h_2)}{nh_2}\}^{\frac{1}{2}}), \tag{7.13}$$

$$\frac{1}{n}\mathbf{X}_{h_2}(u)^T\mathbf{W}_2(u)\mathbf{\Delta}^*\boldsymbol{\varepsilon} = O_p(\{\frac{log(1/h_2)}{nh_2}\}^{\frac{1}{2}}), \tag{7.14}$$

where $c_n^* = h_2^2+\{\frac{log(1/h_2)}{nh_2}\}^{\frac{1}{2}}$.

Proof: The proof of Lemma 4 can be done in a similar way to Li and Mei [1], and we omit it here.

Lemma 5. Let $T_1,\cdots,T_n$ be independent and identically distributed random variables. If $E|T_i|^s$ is bounded for $s>1$, then $\max_{1\le i\le n}|T_i| = o(n^{1/s})a.s.$.

Proof: The proof of Lemma 5 can be referred to Shi and Lau [20].

Lemma 6. Let $\boldsymbol{\tau}_i = (1,\mathbf{X}_i^T,U_i,\mathbf{Z}_i^T)^T$, $\lambda_{\min}$ denotes the minimum eigenvalue of $\sum_{i=1}^n\boldsymbol{\tau}_i\boldsymbol{\tau}_i^T$ under the condition that $\sup_{i\ge1}\|\boldsymbol{\tau}_i\|<\infty$ and $\lambda_{\min}\to\infty$, the quasi-likelihood estimation $\hat{\boldsymbol{\omega}} = (\hat{\omega}_0,\hat{\boldsymbol{\omega}}_1^T,\hat{\omega}_2,\hat{\boldsymbol{\omega}}_3^T)^T$ of $\boldsymbol{\omega} = (\omega_0,\boldsymbol{\omega}_1^T,\omega_2,\boldsymbol{\omega}_3^T)^T$ satisfies

$$\sqrt{n}(\hat{\boldsymbol{\omega}}-\boldsymbol{\omega}) = \mathbf{A}^{-1}n^{-\frac{1}{2}}\sum_{i=1}^n\boldsymbol{\tau}_i(\delta_i-\pi_i)+o_p(1),$$

where $\mathbf{A} = E[\boldsymbol{\tau}_1\boldsymbol{\tau}_1^T\pi_1(1-\pi_1)]$.

Proof: This Lemma can be found in Chen, Feng and Xue [18].

Lemma 7. Let $\mathbf{\Delta}^* = diag\{\frac{\delta_1}{\pi(\mathbf{V}_1)},\cdots,\frac{\delta_n}{\pi(\mathbf{V}_n)}\}$, $\hat{\mathbf{\Delta}}^* = diag\{\frac{\delta_1}{\hat{\pi}(\mathbf{V}_1)},\cdots,\frac{\delta_n}{\hat{\pi}(\mathbf{V}_n)}\}$, then

$$\|\hat{\mathbf{\Delta}}^*-\mathbf{\Delta}^*\| = o_p(n^{-\frac{1}{2}+\frac{1}{2s}}).$$

Proof: With the help of Lemma 5, Lemma 6 and Chen, Feng and Xue [18], Lemma 7 is easily proved, we omit the details here.

Lemma 8. Suppose that Conditions $C1-C8$ hold, we have

$$\frac{1}{n}\mathbf{g}'(\mathbf{Z},\boldsymbol{\beta}_0)^T(\mathbf{I}_n-\mathbf{S}_{h_2})^T(\mathbf{I}_n-\mathbf{S}_{h_2})\mathbf{g}'(\mathbf{Z},\boldsymbol{\beta}_0) = \mathbf{\Sigma}_3\{1+o_p(1)\}, \tag{7.15}$$

where $\mathbf{\Sigma}_3 = E[\mathbf{g}'(\mathbf{Z},\boldsymbol{\beta}_0)\mathbf{g}'(\mathbf{Z},\boldsymbol{\beta}_0)^T] - E[\mathbf{\Phi}^*(U)^T\mathbf{\Gamma}^*(U)^{-1}\mathbf{\Phi}^*(U)]$,

$$\frac{1}{n}\mathbf{g}'(\mathbf{Z},\boldsymbol{\beta}_0)^T(\mathbf{I}_n - \mathbf{S}_{h_2})^T(\mathbf{I}_n - \mathbf{S}_{h_2})[\check{\mathbf{Y}} - \mathbf{g}(\mathbf{Z},\boldsymbol{\beta}_0)] = \boldsymbol{\xi}_n^* + O_p((c_n^*)^2). \tag{7.16}$$

where $\boldsymbol{\xi}_n^* = \frac{1}{n}\sum_{i=1}^n \frac{\delta_i}{\pi(\mathbf{V}_i)}\{\mathbf{g}'(\mathbf{Z}_i,\boldsymbol{\beta}_0) - E[\mathbf{g}'(\mathbf{Z}_i,\boldsymbol{\beta}_0)\mathbf{X}_i^T|U = U_i]\mathbf{\Gamma}^*(U_i)^{-1}\mathbf{X}_i\}\varepsilon_i$.

Proof: Eq (7.15) is similar to the proof of Eq (7.7), we omit the details here.

Next we consider (7.16), let $\hat{\mathbf{M}}_c = \begin{pmatrix} \mathbf{X}_1^T\hat{\boldsymbol{\theta}}_c(U_1,\hat{\boldsymbol{\beta}}_c) \\ \vdots \\ \mathbf{X}_n^T\hat{\boldsymbol{\theta}}_c(U_n,\hat{\boldsymbol{\beta}}_c) \end{pmatrix}$,

then,

$\frac{1}{n}\mathbf{g}'(\mathbf{Z},\boldsymbol{\beta}_0)^T(\mathbf{I}_n - \mathbf{S}_{h_2})^T(\mathbf{I}_n - \mathbf{S}_{h_2})[\check{\mathbf{Y}} - \mathbf{g}(\mathbf{Z},\boldsymbol{\beta}_0)]$

$$\begin{aligned}
&= \frac{1}{n}\mathbf{g}'(\mathbf{Z},\boldsymbol{\beta}_0)^T(\mathbf{I}_n - \mathbf{S}_{h_2})^T(\mathbf{I}_n - \mathbf{S}_{h_2})\{\mathbf{M}_0 + \hat{\mathbf{\Delta}}^*\boldsymbol{\varepsilon} + (\mathbf{I}_n - \hat{\mathbf{\Delta}}^*)[\mathbf{g}(\mathbf{Z},\hat{\boldsymbol{\beta}}_c) - \mathbf{g}(\mathbf{Z},\boldsymbol{\beta}_0)] + (\mathbf{I}_n - \hat{\mathbf{\Delta}}^*)(\hat{\mathbf{M}}_c - \mathbf{M}_0)\} \\
&= \frac{1}{n}\mathbf{g}'(\mathbf{Z},\boldsymbol{\beta}_0)^T(\mathbf{I}_n - \mathbf{S}_{h_2})^T(\mathbf{I}_n - \mathbf{S}_{h_2})\mathbf{M}_0 \\
&+ \frac{1}{n}\mathbf{g}'(\mathbf{Z},\boldsymbol{\beta}_0)^T(\mathbf{I}_n - \mathbf{S}_{h_2})^T(\mathbf{I}_n - \mathbf{S}_{h_2})\hat{\mathbf{\Delta}}^*\boldsymbol{\varepsilon} \\
&+ \frac{1}{n}\mathbf{g}'(\mathbf{Z},\boldsymbol{\beta}_0)^T(\mathbf{I}_n - \mathbf{S}_{h_2})^T(\mathbf{I}_n - \mathbf{S}_{h_2})(\mathbf{I}_n - \hat{\mathbf{\Delta}}^*)[\mathbf{g}(\mathbf{Z},\hat{\boldsymbol{\beta}}_c) - \mathbf{g}(\mathbf{Z},\boldsymbol{\beta}_0)] \\
&+ \frac{1}{n}\mathbf{g}'(\mathbf{Z},\boldsymbol{\beta}_0)^T(\mathbf{I}_n - \mathbf{S}_{h_2})^T(\mathbf{I}_n - \mathbf{S}_{h_2})(\mathbf{I}_n - \hat{\mathbf{\Delta}}^*)(\hat{\mathbf{M}}_c - \mathbf{M}_0) \\
&\hat{=} A_1 + A_2 + A_3 + A_4.
\end{aligned}$$

According to the similar proof to Eq (7.8), we obtain $A_1 = O_p((c_n^*)^2)$.
It is easy to calculate

$$\begin{aligned}
A_2 &= \frac{1}{n}\mathbf{g}'(\mathbf{Z},\boldsymbol{\beta}_0)^T(\mathbf{I}_n - \mathbf{S}_{h_2})^T(\mathbf{I}_n - \mathbf{S}_{h_2})(\hat{\mathbf{\Delta}}^* - \mathbf{\Delta}^*)\boldsymbol{\varepsilon} \\
&+ \frac{1}{n}\mathbf{g}'(\mathbf{Z},\boldsymbol{\beta}_0)^T(\mathbf{I}_n - \mathbf{S}_{h_2})^T(\mathbf{I}_n - \mathbf{S}_{h_2})\mathbf{\Delta}^*\boldsymbol{\varepsilon} \\
&\hat{=} A_{21} + A_{22}.
\end{aligned}$$

By Lemma 7 and condition $C4$,

$$\|A_{21}\| = \frac{1}{n}\|\mathbf{g}'(\mathbf{Z},\boldsymbol{\beta}_0)^T(\mathbf{I}_n - \mathbf{S}_{h_2})^T(\mathbf{I}_n - \mathbf{S}_{h_2})(\hat{\mathbf{\Delta}}^* - \mathbf{\Delta}^*)\boldsymbol{\varepsilon}\| = o_p(n^{-\frac{1}{2}}).$$

Using the similar arguments to the proof of Eq (7.8), we have

$$(\mathbf{X}^T, 0)[\mathbf{X}_{h_2}(u)^T\mathbf{W}_2(u)\mathbf{X}_{h_2}(u)]^{-1}\mathbf{X}_{h_2}(u)^T\mathbf{W}_2(u)\mathbf{\Delta}^*\boldsymbol{\varepsilon} = O_p(c_n^*).$$

Then, $A_{22} = \boldsymbol{\xi}_n^* + O_p((c_n^*)^2)$. Thus $A_2 = A_{21} + A_{22} = \boldsymbol{\xi}_n^* + O_p((c_n^*)^2)$.

Next, we consider A_3, with the help of the proof method of I_3 in the Theorem 2.1.2, we show

$$A_3 = \frac{1}{n}\mathbf{g}'(\mathbf{Z},\boldsymbol{\beta}_0)^T(\mathbf{I}_n - \mathbf{S}_{h_2})^T(\mathbf{I}_n - \mathbf{S}_{h_2})(\mathbf{I}_n - \mathbf{\Delta}^*)[\mathbf{g}(\mathbf{Z},\hat{\boldsymbol{\beta}}_c) - \mathbf{g}(\mathbf{Z},\boldsymbol{\beta}_0)]$$

$$
\begin{aligned}
&- \frac{1}{n}\mathbf{g}'(\mathbf{Z},\boldsymbol{\beta}_0)^T(\mathbf{I}_n-\mathbf{S}_{h_2})^T(\mathbf{I}_n-\mathbf{S}_{h_2})(\hat{\boldsymbol{\Delta}}^*-\boldsymbol{\Delta}^*)[\mathbf{g}(\mathbf{Z},\hat{\boldsymbol{\beta}}_c)-\mathbf{g}(\mathbf{Z},\boldsymbol{\beta}_0)]\\
&= O_p(n^{-\frac{1}{2}}).
\end{aligned}
$$

In the end, we derive

$$
\begin{aligned}
A_4 &= \frac{1}{n}\mathbf{g}'(\mathbf{Z},\boldsymbol{\beta}_0)^T(\mathbf{I}_n-\mathbf{S}_{h_2})^T(\mathbf{I}_n-\mathbf{S}_{h_2})(\mathbf{I}_n-\boldsymbol{\Delta}^*)\mathbf{S}_{h_1}[\mathbf{g}(\mathbf{Z},\boldsymbol{\beta}_0)-\mathbf{g}(\mathbf{Z},\hat{\boldsymbol{\beta}}_c)]\\
&- \frac{1}{n}\mathbf{g}'(\mathbf{Z},\boldsymbol{\beta}_0)^T(\mathbf{I}_n-\mathbf{S}_{h_2})^T(\mathbf{I}_n-\mathbf{S}_{h_2})(\hat{\boldsymbol{\Delta}}^*-\boldsymbol{\Delta}^*)\mathbf{S}_{h_1}[\mathbf{g}(\mathbf{Z},\boldsymbol{\beta}_0)-\mathbf{g}(\mathbf{Z},\hat{\boldsymbol{\beta}}_c)]\\
&= O_p(n^{-\frac{1}{2}}).
\end{aligned}
$$

Now, Eq (7.16) holds.

The proof of Theorem 2.2.1. Similar to the proof of Theorem 2.1.1 and the proof of Theorem 1 in Li and Mei [1], we have $Q(\boldsymbol{\beta}_0+\mathbf{t})-Q(\boldsymbol{\beta}_0)=\mathbf{Q}'(\boldsymbol{\beta}_0)^T\mathbf{t}+\frac{1}{2}\mathbf{t}^T\mathbf{Q}''(\boldsymbol{\beta}^*)\mathbf{t}$, where $\boldsymbol{\beta}^*\in(\boldsymbol{\beta}_0,\boldsymbol{\beta}_0+\mathbf{t})$. Due to $Q(\boldsymbol{\beta})=\sum_{i=1}^n[\breve{Y}_i-\mathbf{X}_i^T\hat{\boldsymbol{\theta}}(U_i,\boldsymbol{\beta})-g(\mathbf{Z}_i,\boldsymbol{\beta})]^2=[\breve{\mathbf{Y}}-\mathbf{g}(\mathbf{Z},\boldsymbol{\beta})]^T(\mathbf{I}_n-\mathbf{S}_{h_2})^T(\mathbf{I}_n-\mathbf{S}_{h_2})[\breve{\mathbf{Y}}-\mathbf{g}(\mathbf{Z},\boldsymbol{\beta})]$, and $\mathbf{Q}'(\boldsymbol{\beta}_0)=-2\mathbf{g}'(\mathbf{Z},\boldsymbol{\beta}_0)^T(\mathbf{I}_n-\mathbf{S}_{h_2})^T(\mathbf{I}_n-\mathbf{S}_{h_2})[\breve{\mathbf{Y}}-\mathbf{g}(\mathbf{Z},\boldsymbol{\beta}_0)]$, by Lemma 8, we can show

$$
\frac{1}{n}\mathbf{Q}'(\boldsymbol{\beta}_0)=-\frac{2}{n}\mathbf{g}'(\mathbf{Z},\boldsymbol{\beta}_0)^T(\mathbf{I}_n-\mathbf{S}_{h_2})^T(\mathbf{I}_n-\mathbf{S}_{h_2})[\breve{\mathbf{Y}}-\mathbf{g}(\mathbf{Z},\boldsymbol{\beta}_0)]=-2\boldsymbol{\xi}_n^*+O_p((c_n^*)^2).
$$

With $\mathbf{0}=\mathbf{Q}'(\hat{\boldsymbol{\beta}}_I)=\mathbf{Q}'(\boldsymbol{\beta}_0)+\mathbf{Q}''(\boldsymbol{\beta}^*)^T(\hat{\boldsymbol{\beta}}_I-\boldsymbol{\beta}_0)$ and $\frac{1}{2n}\mathbf{Q}''(\boldsymbol{\beta}^*)=\boldsymbol{\Sigma}_3\{1+o_p(1)\}$ in hand, we have $\sqrt{n}\boldsymbol{\Sigma}_3(1+o_p(1))(\hat{\boldsymbol{\beta}}_I-\boldsymbol{\beta}_0)=\sqrt{n}\boldsymbol{\xi}_n^*+O_p(\sqrt{n}(c_n^*)^2)=\sqrt{n}\boldsymbol{\xi}_n^*+o_p(1)$.

Lastly, by the Slutsky theorem and the central limit theorem, we have

$$
\sqrt{n}(\hat{\boldsymbol{\beta}}_I-\boldsymbol{\beta}_0)\xrightarrow{D}N(\mathbf{0},\boldsymbol{\Sigma}_3^{-1}\boldsymbol{\Sigma}_4\boldsymbol{\Sigma}_3^{-1}).
$$

The proof of Theorem 2.2.2. After a series of calculations, we can obtain

$$
\begin{aligned}
\hat{\boldsymbol{\Psi}}_I(u_0,\hat{\boldsymbol{\beta}}_I) &= [\mathbf{X}_{h_2}(u_0)^T\mathbf{W}_2(u_0)\mathbf{X}_{h_2}(u_0)]^{-1}\mathbf{X}_{h_2}(u_0)^T\mathbf{W}_2(u_0)[\breve{\mathbf{Y}}-\boldsymbol{g}(\mathbf{Z},\hat{\boldsymbol{\beta}}_I)]\\
&= [\mathbf{X}_{h_2}(u_0)^T\mathbf{W}_2(u_0)\mathbf{X}_{h_2}(u_0)]^{-1}\mathbf{X}_{h_2}(u_0)^T\mathbf{W}_2(u_0)\mathbf{M}_0\\
&+ [\mathbf{X}_{h_2}(u_0)^T\mathbf{W}_2(u_0)\mathbf{X}_{h_2}(u_0)]^{-1}\mathbf{X}_{h_2}(u_0)^T\mathbf{W}_2(u_0)\hat{\boldsymbol{\Delta}}^*\boldsymbol{\varepsilon}\\
&+ [\mathbf{X}_{h_2}(u_0)^T\mathbf{W}_2(u_0)\mathbf{X}_{h_2}(u_0)]^{-1}\mathbf{X}_{h_2}(u_0)^T\mathbf{W}_2(u_0)(\mathbf{I}_n-\hat{\boldsymbol{\Delta}}^*)(\mathbf{I}_n-\mathbf{S}_{h_1})[\boldsymbol{g}(\mathbf{Z},\hat{\boldsymbol{\beta}}_c)-\boldsymbol{g}(\mathbf{Z},\boldsymbol{\beta}_0)]\\
&+ [\mathbf{X}_{h_2}(u_0)^T\mathbf{W}_2(u_0)\mathbf{X}_{h_2}(u_0)]^{-1}\mathbf{X}_{h_2}(u_0)^T\mathbf{W}_2(u_0)[\boldsymbol{g}(\mathbf{Z},\hat{\boldsymbol{\beta}}_I)-\boldsymbol{g}(\mathbf{Z},\boldsymbol{\beta}_0)]\\
&\hat{=} B_1+B_2+B_3+B_4.
\end{aligned}
$$

Similar to the proof of I_1 in Theorem 2.1.2, we have

$$
B_1=\boldsymbol{\Psi}_0(u_0)+\frac{1}{2}h_2^2\mu_{2,2}\begin{pmatrix}\boldsymbol{\theta}_0''(u_0)\\0\end{pmatrix}+o(h_2^2).
$$

It is also easy for us to deduce

$$
\begin{aligned}
B_2 &= [\mathbf{X}_{h_2}(u_0)^T\mathbf{W}_2(u_0)\mathbf{X}_{h_2}(u_0)]^{-1}\mathbf{X}_{h_2}(u_0)^T\mathbf{W}_2(u_0)(\hat{\boldsymbol{\Delta}}^*-\boldsymbol{\Delta}^*)\boldsymbol{\varepsilon}\\
&+ [\mathbf{X}_{h_2}(u_0)^T\mathbf{W}_2(u_0)\mathbf{X}_{h_2}(u_0)]^{-1}\mathbf{X}_{h_2}(u_0)^T\mathbf{W}_2(u_0)\boldsymbol{\Delta}^*\boldsymbol{\varepsilon}
\end{aligned}
$$

$$\hat{=} \quad B_{21} + B_{22},$$

by Lemma 7 and Theorem 2.1.2, we can obtain

$$B_{21} = o_p(n^{-\frac{1}{2}+\frac{1}{2s}}), B_{22} = \frac{1}{n} f^{-1}(u_0)\mathbf{\Gamma}^*(u_0)^{-1} \otimes \begin{pmatrix} 1 & 0 \\ 0 & \mu_{2,2}^{-1} \end{pmatrix} [1 + O_p(c_n^*)]^{-1} \mathbf{X}_{h_2}^T(u_0)\mathbf{W}_2(u_0)\mathbf{\Delta}^* \boldsymbol{\varepsilon}.$$

According to the central limit theorem, we get

$$\sqrt{nh_2}\frac{1}{n}\mathbf{X}_{h_2}^T(u_0)\mathbf{W}_2(u_0)\mathbf{\Delta}^*\boldsymbol{\varepsilon} = \sqrt{nh_2}\begin{pmatrix} \frac{1}{n}\sum_{i=1}^n \frac{\delta_i}{\pi(\mathbf{V}_i)}\varepsilon_i \mathbf{X}_i K_{2,h_2}(U_i - u_0) \\ \frac{1}{n}\sum_{i=1}^n \frac{\delta_i}{\pi(\mathbf{V}_i)}\varepsilon_i \mathbf{X}_i \frac{U_i - u_0}{h_2} K_{2,h_2}(U_i - u_0) \end{pmatrix} \xrightarrow{D} N(\mathbf{0}, \mathbf{\Sigma}^{**}),$$

where

$$\mathbf{\Sigma}^{**} = E(\boldsymbol{\varepsilon}^2\pi^{-1}(\mathbf{V})\mathbf{X}\mathbf{X}^T | U = u_0) f(u_0) \otimes \begin{pmatrix} \nu_{2,0} & 0 \\ 0 & \nu_{2,2} \end{pmatrix}.$$

So,

$$\sqrt{nh_2}B_2 \xrightarrow{D} N(\mathbf{0}, f^{-1}(u_0)\mathbf{\Gamma}^*(u_0)^{-1}E(\boldsymbol{\varepsilon}^2\pi^{-1}(\mathbf{V})\mathbf{X}\mathbf{X}^T | U = u_0)\mathbf{\Gamma}^*(u_0)^{-1} \otimes \begin{pmatrix} \nu_{2,0} & 0 \\ 0 & \nu_{2,2}\mu_{2,2}^{-2} \end{pmatrix}).$$

Next, we can calculate $B_3 = O_p(n^{-\frac{1}{2}})$, and $B_4 = O_p(n^{-\frac{1}{2}})$. With B_1, B_2, B_3 and B_4 in hand, we can get the following equation

$$\hat{\mathbf{\Psi}}_I(u_0, \hat{\boldsymbol{\beta}}_I) = \mathbf{\Psi}_0(u_0) + \frac{1}{2}h_2^2\mu_{2,2}\begin{pmatrix} \boldsymbol{\theta}_0''(u_0) \\ 0 \end{pmatrix} + B_2 + O_p(h_2^2 + n^{-\frac{1}{2}}).$$

So,

$$\begin{aligned} &\sqrt{nh_2}\left[\hat{\mathbf{\Psi}}_I(u_0, \hat{\boldsymbol{\beta}}_I) - \mathbf{\Psi}_0(u_0) - \frac{1}{2}h_2^2\mu_{2,2}\begin{pmatrix} \boldsymbol{\theta}_0''(u_0) \\ 0 \end{pmatrix}\right] \\ = &\quad \sqrt{nh_2}B_2 + O_p(\sqrt{nh_2^5} + \sqrt{h_2}) \\ \xrightarrow{D} &\quad N(\mathbf{0}, f^{-1}(u_0)\mathbf{\Gamma}^*(u_0)^{-1}E(\boldsymbol{\varepsilon}^2\pi^{-1}(\mathbf{V})\mathbf{X}\mathbf{X}^T | U = u_0)\mathbf{\Gamma}^*(u_0)^{-1} \otimes \begin{pmatrix} \nu_{2,0} & 0 \\ 0 & \nu_{2,2}\mu_{2,2}^{-2} \end{pmatrix}). \end{aligned}$$

Now, the proof of Theorem 2.2.2 is completed.

Lemma 9. Suppose that Conditions $C1 - C8$ hold, we have

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n \boldsymbol{\eta}_i(\boldsymbol{\beta}_0) \xrightarrow{D} N(\mathbf{0}, \mathbf{\Sigma}_4), \tag{7.17}$$

$$\frac{1}{n}\sum_{i=1}^n \boldsymbol{\eta}_i(\boldsymbol{\beta}_0)\boldsymbol{\eta}_i(\boldsymbol{\beta}_0)^T - \mathbf{\Sigma}_4 = o_p(1), \tag{7.18}$$

$$\max_{1\le i\le n} \|\boldsymbol{\eta}_i(\boldsymbol{\beta}_0)\| = o_p(n^{\frac{1}{2}}), \tag{7.19}$$

$$\|\boldsymbol{\lambda}\| = O_p(n^{-\frac{1}{2}}), \tag{7.20}$$

$$\frac{1}{n}\sum_{i=1}^{n}\|\boldsymbol{\eta}_i(\boldsymbol{\beta}_0)\|^3 = o_p(n^{\frac{1}{2}}), \tag{7.21}$$

where $\boldsymbol{\Sigma}_4 = E[\pi^{-1}(\mathbf{V})\boldsymbol{\varepsilon}^2\mathbf{g}'(\mathbf{Z},\boldsymbol{\beta}_0)\mathbf{g}'(\mathbf{Z},\boldsymbol{\beta}_0)^T] - E[\pi^{-1}(\mathbf{V})\boldsymbol{\varepsilon}^2\boldsymbol{\Phi}^*(U)^T\boldsymbol{\Gamma}^*(U)^{-1}\boldsymbol{\Phi}^*(U)]$.

Proof: First, we prove (7.17), by Lemma 8, we can get

$$\begin{aligned}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\boldsymbol{\eta}_i(\boldsymbol{\beta}_0) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{\widetilde{\mathbf{g}'}(\mathbf{Z}_i,\boldsymbol{\beta}_0)[\widetilde{Y}_i - \widetilde{g}(\mathbf{Z}_i,\boldsymbol{\beta}_0)]\} \\ &= \frac{1}{\sqrt{n}}\mathbf{g}'(\mathbf{Z},\boldsymbol{\beta}_0)^T(\mathbf{I}_n - \mathbf{S}_{h_2})^T(\mathbf{I}_n - \mathbf{S}_{h_2})[\check{\mathbf{Y}} - \mathbf{g}(\mathbf{Z},\boldsymbol{\beta}_0)] \\ &= \sqrt{n}[\boldsymbol{\xi}_n^* + O_p((c_n^*)^2)].\end{aligned}$$

Combing $E(\boldsymbol{\xi}_n^*) = 0$ and
$Cov(\boldsymbol{\xi}_n^*) = \frac{1}{n}\{E[\pi^{-1}(\mathbf{V})\boldsymbol{\varepsilon}^2\mathbf{g}'(\mathbf{Z},\boldsymbol{\beta}_0)\mathbf{g}'(\mathbf{Z},\boldsymbol{\beta}_0)^T] - E[\pi^{-1}(\mathbf{V})\boldsymbol{\varepsilon}^2\boldsymbol{\Phi}^*(U)^T\boldsymbol{\Gamma}^*(U)^{-1}\boldsymbol{\Phi}^*(U)]\}$, (7.17) is proved. (7.18) is also easy to be proved, we omit it here.

Next we prove (7.19), obviously, we can show

$$\max_{1\le i\le n}\|\boldsymbol{\eta}_i(\boldsymbol{\beta}_0)\| = \max_{1\le i\le n}\|\tilde{g}'(\mathbf{Z}_i,\boldsymbol{\beta}_0)\| + \max_{1\le i\le n}\|\widetilde{Y}_i - \tilde{g}(\mathbf{Z}_i,\boldsymbol{\beta}_0)\|.$$

With condition $C4$, we have

$$\max_{1\le i\le n}\|\tilde{g}'(\mathbf{Z}_i,\boldsymbol{\beta}_0)\| = o(n^{\frac{1}{2s}}), \quad \max_{1\le i\le n}\|\widetilde{Y}_i - \tilde{g}(\mathbf{Z}_i,\boldsymbol{\beta}_0)\| = \max_{1\le i\le n}\|\varepsilon_i\| = o(n^{\frac{1}{2s}}).$$

Thus, (7.19) holds. By (7.17), (7.18) and owen [16, 17], we can prove (7.20). By (7.18) and (7.19), we can get (7.21).

The proof of Theorem 3.1.1 and Theorem 3.1.2 are similar to the proof of Theorem 2.1 and Theorem 2.2 in Zhou, Zhao and Wang [4] respectively with Lemma 9 in hand, so here we omit the details.

Lemma 10. Suppose that Conditions $C1 - C8$ hold, and $\boldsymbol{\theta}_0(u)$ is the true coefficient function of $\boldsymbol{\theta}(u)$, we have

$$\sqrt{\frac{h_2}{n}}\sum_{i=1}^{n}\hat{\boldsymbol{\zeta}}_i(\boldsymbol{\theta}_0(u)) \xrightarrow{D} N(\mathbf{0},\mathbf{B}), \tag{7.22}$$

$$\frac{h_2}{n}\sum_{i=1}^{n}\hat{\boldsymbol{\zeta}}_i(\boldsymbol{\theta}_0(u))\hat{\boldsymbol{\zeta}}_i(\boldsymbol{\theta}_0(u))^T - \mathbf{B} = o_p(1), \tag{7.23}$$

$$\max_{1\le i\le n}\|\hat{\boldsymbol{\zeta}}_i(\boldsymbol{\theta}_0(u))\| = o_p(\sqrt{\frac{n}{h_2}}), \tag{7.24}$$

$$\|\boldsymbol{\gamma}\| = O_p(\sqrt{\frac{h_2}{n}}), \tag{7.25}$$

$$\frac{1}{n}\sum_{i=1}^{n}\|\hat{\boldsymbol{\zeta}}_i(\boldsymbol{\theta}_0(u))\|^3 = o_p(\sqrt{\frac{n}{h_2}}), \tag{7.26}$$

where $\mathbf{B} = E(\boldsymbol{\varepsilon}^2\pi^{-1}(\mathbf{V})\mathbf{X}\mathbf{X}^T|U = u_0)f(u_0)\nu_{2,0}$.

Proof: First, we prove (7.22), we can show

$$\begin{aligned}
\sqrt{\frac{h_2}{n}}\sum_{i=1}^{n}\hat{\zeta}_i(\boldsymbol{\theta}_0(u)) &= \sqrt{\frac{h_2}{n}}\sum_{i=1}^{n}\mathbf{X}_i\{\breve{Y}_i - g(\mathbf{Z}_i,\hat{\boldsymbol{\beta}}_E) - \mathbf{X}_i^T\boldsymbol{\theta}_0(u) - \mathbf{X}_i[\hat{\boldsymbol{\theta}}_I(U_i,\hat{\boldsymbol{\beta}}_I) - \hat{\boldsymbol{\theta}}_I(u,\hat{\boldsymbol{\beta}}_I)]\}K_{2,h_2}(U_i - u) \\
&= \sqrt{\frac{h_2}{n}}\sum_{i=1}^{n}\frac{\delta_i}{\hat{\pi}(\mathbf{V}_i)}\mathbf{X}_i\varepsilon_i K_{2,h_2}(U_i - u) \\
&+ \sqrt{\frac{h_2}{n}}\sum_{i=1}^{n}\mathbf{X}_i[1 - \frac{\delta_i}{\hat{\pi}(\mathbf{V}_i)}][\mathbf{X}_i^T\hat{\boldsymbol{\theta}}_c(U_i,\hat{\boldsymbol{\beta}}_c) - \mathbf{X}_i^T\boldsymbol{\theta}_0(U_i)]K_{2,h_2}(U_i - u) \\
&+ \sqrt{\frac{h_2}{n}}\sum_{i=1}^{n}\mathbf{X}_i\{[1 - \frac{\delta_i}{\hat{\pi}(\mathbf{V}_i)}][g(\mathbf{Z}_i,\hat{\boldsymbol{\beta}}_c) - g(\mathbf{Z}_i,\boldsymbol{\beta}_0)] - [g(\mathbf{Z}_i,\hat{\boldsymbol{\beta}}_E) - g(\mathbf{Z}_i,\boldsymbol{\beta}_0)]\}K_{2,h_2}(U_i - u) \\
&+ \sqrt{\frac{h_2}{n}}\sum_{i=1}^{n}\mathbf{X}_i\mathbf{X}_i^T\{\boldsymbol{\theta}_0(U_i) - \boldsymbol{\theta}_0(u) - [\hat{\boldsymbol{\theta}}_I(U_i,\hat{\boldsymbol{\beta}}_I) - \hat{\boldsymbol{\theta}}_I(u,\hat{\boldsymbol{\beta}}_I)]\}K_{2,h_2}(U_i - u) \\
&\hat{=} D_1 + D_2 + D_3 + D_4.
\end{aligned}$$

By Lemma 7, we can get

$$\begin{aligned}
D_1 &= \sqrt{\frac{h_2}{n}}\sum_{i=1}^{n}[\frac{\delta_i}{\pi(\mathbf{V}_i)} + o_p(n^{-\frac{1}{2}+\frac{1}{2s}})]\mathbf{X}_i\varepsilon_i K_{2,h_2}(U_i - u) \\
&= D_{11} + o_p\left(n^{\frac{1}{2s}}\left(\frac{log(1/h_2)}{n}\right)^{\frac{1}{2}}\right).
\end{aligned}$$

With $E(D_{11}) = 0$, $Cov(D_{11}) = E(\boldsymbol{\varepsilon}^2\pi^{-1}(\mathbf{V})\mathbf{X}\mathbf{X}^T|U = u_0)f(u_0)\nu_{2,0}$, and the central limit theorem, we can deduce $D_1 \xrightarrow{D} N(\mathbf{0},\mathbf{B})$. Next, we calculate D_2 and D_3,

$$D_2 = \sqrt{\frac{h_2}{n}}\sum_{i=1}^{n}\mathbf{X}_i[1 - \frac{\delta_i}{\pi(\mathbf{V}_i)} + o_p(n^{-\frac{1}{2}+\frac{1}{2s}})]\mathbf{X}_i^T O_p(n^{-\frac{1}{2}})K_{2,h_2}(U_i - u) = O_p\left(\frac{log(1/h_2)}{n}\right)^{\frac{1}{2}},$$

$$D_3 = \sqrt{\frac{h_2}{n}}\sum_{i=1}^{n}\mathbf{X}_i\{[1 - \frac{\delta_i}{\pi(\mathbf{V}_i)} + o_p(n^{-\frac{1}{2}+\frac{1}{2s}})]O_p(n^{-\frac{1}{2}}) - O_p(n^{-\frac{1}{2}})\}K_{2,h_2}(U_i - u) = O_p\left(\frac{log(1/h_2)}{n}\right)^{\frac{1}{2}}.$$

Applying Taylor expansion to $\boldsymbol{\theta}_0(U_i)$ and $\hat{\boldsymbol{\theta}}_I(U_i)$ at u, respectively, we have

$$\boldsymbol{\theta}_0(U_i) - \boldsymbol{\theta}_0(u) - [\hat{\boldsymbol{\theta}}_I(U_i,\hat{\boldsymbol{\beta}}_I) - \hat{\boldsymbol{\theta}}_I(u,\hat{\boldsymbol{\beta}}_I)] = [\boldsymbol{\theta}_0'(u) - \hat{\boldsymbol{\theta}}_I'(u,\hat{\boldsymbol{\beta}}_I)](U_i - u) + o_p(U_i - u).$$

With $\sqrt{\frac{h_2}{n}}\sum_{i=1}^{n}(U_i - u)K_{2,h_2}(U_i - u) = O_p(1)$ and $\boldsymbol{\theta}_0'(u) - \hat{\boldsymbol{\theta}}_I'(u,\hat{\boldsymbol{\beta}}_I) = o_p(1)$, then, we obtain $D_4 = o_p(1)$. Finally, we have

$$\sqrt{\frac{h_2}{n}}\sum_{i=1}^{n}\hat{\zeta}_i(\boldsymbol{\theta}_0(u)) = D_{11} + o_p\left(n^{\frac{1}{2s}}\left(\frac{log(1/h_2)}{n}\right)^{\frac{1}{2}}\right) + O_p\left(\frac{log(1/h_2)}{n}\right)^{\frac{1}{2}} + O_p\left(\frac{log(1/h_2)}{n}\right)^{\frac{1}{2}} + o_p(1)$$

$$\xrightarrow{D} N(\mathbf{0}, \mathbf{B}).$$

With (7.22) in hand, (7.23), (7.24), (7.25) and (7.26) can be also easily proved, we omit the details here. According to the same method as the proof of Theorem 3.1.1, together with Lemma 10, we can easily prove Theorem 3.2.1, we also omit the details here.

The proof of Theorem 4.1.1 First, we can calculate

$$\begin{aligned}
\sqrt{n}(\hat{\mu}-\mu_0) &= \sqrt{n}\{\frac{1}{n}\sum_{i=1}^{n}\{\frac{\delta_i}{\hat{\pi}(\mathbf{V}_i)}Y_i+(1-\frac{\delta_i}{\hat{\pi}(\mathbf{V}_i)})[\mathbf{X}_i^T\hat{\boldsymbol{\theta}}_c(U_i,\hat{\boldsymbol{\beta}}_c)+g(\mathbf{Z}_i,\hat{\boldsymbol{\beta}}_c)]\}-\mu_0\} \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{\frac{\delta_i}{\hat{\pi}(\mathbf{V}_i)}\varepsilon_i+\mathbf{X}_i^T\boldsymbol{\theta}(U_i)+g(\mathbf{Z}_i,\boldsymbol{\beta})-\mu_0\} \\
&+ \frac{1}{\sqrt{n}}\sum_{i=1}^{n}[\frac{\delta_i}{\hat{\pi}(\mathbf{V}_i)}-\frac{\delta_i}{\pi(\mathbf{V}_i)}]\varepsilon_i \\
&+ \frac{1}{\sqrt{n}}\sum_{i=1}^{n}[1-\frac{\delta_i}{\hat{\pi}(\mathbf{V}_i)}][\mathbf{X}_i^T\hat{\boldsymbol{\theta}}_c(U_i,\hat{\boldsymbol{\beta}}_c)-\mathbf{X}_i^T\boldsymbol{\theta}(U_i)] \\
&+ \frac{1}{\sqrt{n}}\sum_{i=1}^{n}[1-\frac{\delta_i}{\hat{\pi}(\mathbf{V}_i)}][g(\mathbf{Z}_i,\hat{\boldsymbol{\beta}}_c)-g(\mathbf{Z}_i,\boldsymbol{\beta})] \\
&\hat{=} S_1+S_2+S_3+S_4.
\end{aligned}$$

By the central limit theorem, we have $S_1 \xrightarrow{D} N(0,\Lambda)$.
According to Lemma 7, we obtain

$$S_2=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}[\frac{\delta_i}{\hat{\pi}(\mathbf{V}_i)}-\frac{\delta_i}{\pi(\mathbf{V}_i)}]\varepsilon_i=o_p(n^{-\frac{1}{2}+\frac{1}{2s}}),$$

$$\begin{aligned}
S_3 &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}[1-\frac{\delta_i}{\pi(\mathbf{V}_i)}][\mathbf{X}_i^T\hat{\boldsymbol{\theta}}_c(U_i,\hat{\boldsymbol{\beta}}_c)-\mathbf{X}_i^T\boldsymbol{\theta}(U_i)]+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}[\frac{\delta_i}{\hat{\pi}(\mathbf{V}_i)}-\frac{\delta_i}{\pi(\mathbf{V}_i)}][\mathbf{X}_i^T\hat{\boldsymbol{\theta}}_c(U_i,\hat{\boldsymbol{\beta}}_c)-\mathbf{X}_i^T\boldsymbol{\theta}(U_i)] \\
&= O_p(n^{-\frac{1}{2}})+o_p(n^{-\frac{1}{2}+\frac{1}{2s}}) \\
&= O_p(n^{-\frac{1}{2}}),
\end{aligned}$$

$$S_4=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}[1-\frac{\delta_i}{\hat{\pi}(\mathbf{V}_i)}][g(\mathbf{Z}_i,\hat{\boldsymbol{\beta}}_c)-g(\mathbf{Z}_i,\boldsymbol{\beta})]=O_p(n^{-\frac{1}{2}}).$$

Now, the proof of Theorem 4.1.1 is completed.

The proof of Theorem 4.2.1 By Theorem 4.1.1, it is easy to derive

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\breve{Y}_i-\mu_0)\xrightarrow{D} N(0,\Lambda),$$

$$\frac{1}{n}\sum_{i=1}^{n}(\breve{Y}_i-\mu_0)^2-\Lambda=o_p(1),$$

$$\max_{1\le i\le n}|\check{Y}_i-\mu_0| = o_p(n^{\frac{1}{2}}).$$

With the help of Owen [16, 17], we can get

$$|\rho| = O_p(n^{-\frac{1}{2}}),$$

$$\frac{1}{n}\sum_{i=1}^{n}(\check{Y}_i-\mu_0)^3 = o_p(n^{\frac{1}{2}}).$$

Using the similar mehtod to the proof of Theorem 2.1 in Zhou, Zhao and Wang [4], Theorem 4.2.1 can be also easily proved, we omit the details here.

Acknowledgments

Wang's researches are supported by National Natural Science Foundation (NNSF) projects (11971265) of China. Zhao's researches are supported by the National Social Science Foundation of China (Grant No.18BTJ035) and the Natural Science Foundation of Chongqing (Grant No.cstc2020jcyj-msxmX0006). Song's researches are supported by Natural Science Foundation (NSF) project (ZR2019MA016) of Shandong Province of China, the Fundamental Research Funds for the Central Universities (No.20CX05012A).

Conflict of interest

The authors declare that there is no conflict of interest.

References

1. T. Z. Li, C. L. Mei, Estimation and inference for varyig coefficient partially nonlinear models, *J. Stat. Plann Inference,* **143** (2013), 2023–2037.
2. J. Yang, H. Yang, Smooth-threshold estimating equations for varying coefficient partially nonlinear models based on orthogonality-projection method, *J. Comput. Appl. Math.,* **302** (2016), 24–37.
3. Y. Y. Qian, Z. S. Huang, Statistical inference for a varying-coefficient partially nonlinear model with measurement errors, *Stat. Methodol.,* **32** (2016), 122–130.
4. X. S. Zhou, P. X. Zhao, X. L. Wang, Empirical likelihood inferences for varying coefficient partially nonlinear models, *J. Appl. Stat.,* **44** (2017), 474–492.
5. Y. L. Jiang, Q. H. Ji, B. J. Xie, Robust estimation for the varying coefficient partially nonlinear models, *J. Comput. Appl. Math.,* **326** (2017), 31–43.
6. Y. T. Xiao, Z. S. Chen, Bias-corrected estimations in varying-coefficient partially nonlinear models with measurement error in the nonparametric part, *J. Appl. Stat.,* **45** (2018), 586–603.
7. X. L. Wang, P. X. Zhao, H. Y. Du, Statistical inferences for varying coefficient partially nonlinear model with missing covariates, *Commun. Stat. Theory Methods,* (2019).
8. R. J. A. Little, D. B. Rubin, *Statistical analysis with missing data*, 2 Eds., New York: Wiley, 2002.

9. P. E. Cheng, Nonparametric estimation of mean functional with data missing at random, *J. Am. Stat. Assoc.,* **89** (1994), 81–87.
10. Q. H. Wang, J. N. K. Rao, Empirical likelihood for linear regression models under imputation for missing responses, *Can. J. Stat.,* **29** (2001), 596–608.
11. Q. H. Wang, J. N. K. Rao, Empirical likelihood-based inference under imputation for missing response data, *Ann. Stat.,* **30** (2002), 896–924.
12. Q. H. Wang, O. Linton, W. Härdle, Semiparametric regression analysis with missing response at random, *J. Am. Stat. Assoc.,* **99** (2004), 334–345.
13. L. G. Xue, Empirical likelihood for linear models with missing responses, *J. Multivar. Anal.,* **100** (2009), 1353–1366.
14. L. G. Xue, Empirical likelihood confidence intervals for response mean with data missing at Random, *Scand. J. Stat.,* **36** (2009), 671–685.
15. X. L. Wang, F. Chen, L. Lin, Empirical likelihood inference for estimating equation with missing data, *Sci. China Math.,* **56** (2013), 1233–1245.
16. A. Owen, Empirical likelihood ratio confidence intervals for a single functional, *Biometrika,* **75** (1988), 237–249.
17. A. Owen, Empirical likelihood ratio confidence regions, *Ann. Stat.,* **18** (1990), 90–120.
18. P. P. Chen, S. Y. Feng, L. G. Xue, Statistical inference for semiparametric varying coefficient partially linear model with missing data (in Chinese), *Acta Math. Sci.,* **35** (2015), 345–358.
19. Y. P. Mack, B. W. Silverman, Weak and strong uniform consistency of kernel regression estimates, *Z. Wahrsch. Verw. Gebiete.,* **61** (1982), 405–415.
20. J. Shi, T. S. Lau, Empirical Likelihood for partially linear models, *Multivar. Anal.,* **72** (2000), 132–148.
21. X. L. Wang, Y. Q. Song, L. Lin, Handling estimating equation with nonignorably missing data based on SIR algorithm, *J. Comput. Appl. Math.,* **326** (2017), 62–70.
22. X. L. Wang, Y. Q. Song, S. X. Zhang, An efficient estimation for the parameter in additive partially linear models with missing covariates, *J. Korean Stat. Soc.,* **49** (2020), 779–801.
23. H. M. Sun, Y. H. Luan, J. M. Jiang, A new classified mixed model predictor, *J. Stat. Plann. Inference,* **207** (2020), 45–54.
24. Y. Fang, G. Cheng, Z. F. Qu, Optimal reinsurance for both an insurer and a reinsurer under general premium principles, *AIMS Math.,* **5** (2020), 3231–3255.
25. Y. Fang, X. Wang, H. L. Liu, T. Li, Pareto-optimal reinsurance for both the insurer and the reinsurer with general premium principles, *Commun. Stat. Theory Methods,* **48** (2019), 6134–6154.