Acta Mathematicae Applicatae Sinica, English Series
Vol. 31, No. 3 (2015) 823–840
DOI: 10.1007/s10255-015-0508-y
http://www.ApplMath.com.cn & www.SpringerLink.com

*Acta Mathematicae Applicatae Sinica, English Series*


# Empirical Likelihood-based Inferences in Varying Coefficient Models with Missing Data

**Xiao-hui LIU**[1,2]

[1] School of Statistics, Jiangxi University of Finance and Economics, Nanchang 330013, China
[2] Research Center of Applied Statistics, Jiangxi University of Finance and Economics, Nanchang 330013, China
(E-mail: csuliuxh912@gmail.com)

**Abstract** In this paper, we consider the empirical likelihood-based inferences for varying coefficient models $Y = X^\tau\alpha(U) + \varepsilon$ when $X$ are subject to missing at random. Based on the inverse probability-weighted idea, a class of empirical log-likelihood ratios, as well as two maximum empirical likelihood estimators, are developed for $\alpha(u)$. The resulting statistics are shown to have standard chi-squared or normal distributions asymptotically. Simulation studies are also constructed to illustrate the finite sample properties of the proposed statistics.

**Keywords** varying coefficient models; missing at random; empirical likelihood; maximum empirical likelihood estimator

**2000 MR Subject Classification** 62G05, 62G20

## 1 Introduction

Missing data are frequently encountered in many statistical applications due to various reasons[12]. To handle the missing data, the current practice is only using the complete subjects and ignoring those with missing values, known as complete case analysis (CCA). However, it is well known that, in the presence of missing data, CCA can not only lose efficiency, but also generate considerable bias, especially when the missing mechanism depends on the outcome variables; see [11,19] for more details. Therefore, it is important to develop some new methods which can take the partially incomplete data into account.

In this paper, we are interested in the following varying-coefficient model

$$Y = X^\tau\alpha(U) + \varepsilon. \tag{1}$$

where $Y$ is a response variable, $X$ a $p$-variate random covariate vector, $U$ a scalar covariate, and $\alpha(\cdot) = (\alpha_1(\cdot), \alpha_2(\cdot), \cdots, \alpha_p(\cdot))^\tau$ an unknown vector of some smooth functions. The model error $\varepsilon$ satisfies $E(\varepsilon|X, U) = 0$ and $E(\varepsilon^2|X, U) < \infty$. $A^\tau$ denotes the transposition of a matrix $A$. We focus mainly on the case that the covariates $X$ may be missing at random (MAR). That is, the available incomplete data are

$$(\delta_i, X_i, Y_i, U_i), \qquad i = 1, 2, \cdots, n,$$

where $\delta_i = 0$ if the $X_i$ is missing, otherwise $\delta_i = 1$. $\delta_i$ satisfies that $P(\delta_i = 1|X_i, Y_i, U_i) = P(\delta_i = 1|Y_i, U_i) = \pi(Z_i) = \pi_i$ with $Z_i = (Y_i, U_i)$. MAR is commonly assumed in the literature; see for example [9,11,12,18,22,26].

Manuscript received July 23, 2012. Revised June 2, 2015.
This work was supported in part by NSF of China (No.11461029), NSF of Jiangxi Province (No.20142BAB211014), and YSFP of Jiangxi provincial education department (No. GJJ14350).

As introduced in [7], Model (1) provides a natural and useful extension of the classical linear regression model by allowing the regression coefficients to depend on certain covariates. Due to the flexibility, this model has been intensively studied by many authors in the past decades. For example, [4] considered a two-step estimating procedure for Model (1) when the coefficient functions have possibly different degrees of smoothness. [1] developed an efficient estimating procedure for this model in the framework of generalized linear models. When $X$ may be missing at random, [26] proposed a locally weighted estimator for this model based on the inverse probability-weighted idea[8]. For more details, we refer the readers to [5]. Nevertheless, to our best knowledge, there is few research in the literature concerning the empirical likelihood-based inferences in Model (1). What we know is the work by [28]. However, their method is not directly applicable in the setting of missing data.

In this paper, we propose three locally weighted empirical log-likelihood ratio (ELR) statistics relying on the inverse probability-weighted idea. The first statistic uses an auxiliary random vector similar to [26]. For this naive statistic, it is shown that the Wilks' theorem for ELR is no longer available due to the mismatch between the variance of the quantity $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\widehat{\mathcal{H}}_i$ and the probability limit of $\frac{1}{n}\sum_{i=1}^{n}\widehat{\mathcal{H}}_i\widehat{\mathcal{H}}_i^{\tau}$ (see Section 2). A similar phenomenon has also been found by [20] in a more general sense. To resolve this problem, we define an another new ELR which has a standard $\chi^2$-distribution asymptotically, inspired by [10] and [18]. However, since the range of the corresponding bandwidth $h$ does not contain the optimal bandwidth, undersmoothing becomes necessary. Aim at avoiding the undersmoothing, we also propose a residual-adjusted ELR motivated by [28]. Furthermore, from the first two ELR's, two maximum empirical likelihood estimators are developed and shown to be asymptotically equivalent to those of [26].

Empirical likelihood is first introduced by [14,15]. Our motivation of using EL is that, although EL is computer-intensive, it is a powerful tool for statistical inferences due to it involves no explicit variance estimation, which is difficult especially when missing data are present[26], and it produces confidence regions with natural shape and orientation. There are many literature concerning the EL method. See, for example, [15,17,20,24,28] among others. Many of the early results are summarized in [16], and the updated results can be found in [2].

The rest of this paper is organized as follows. In Section 2, we present three empirical-likelihood-based statistics for Model (1) with missing at random covariates $X$, and derive their asymptotic distributions. Two maximum empirical likelihood estimators (MELE) are also developed. In Section 3, we construct some simulation studies to illustrate the finite sample properties of the proposed statistics. In Section 4, we conclude this paper with a brief discussion. The technical details of the proofs of the main results are provided in the Appendix.

## 2 Methodology and Main Results

In this section, three locally weighted empirical log-likelihood ratio statistics are suggested relying on the inverse probability-weighted idea. Two maximum empirical likelihood estimators are also defined as by-products.

### 2.1 A Naive Locally Weighted Empirical Log-likelihood Ratio

Motivated by [26] and [28], we have the following observation:

$$E\Big\{\frac{\delta_i}{\pi(Z_i)}(Y_i - X_i^{\tau}\alpha(U_i))X_i\Big| U_i = u\Big\}\gamma(u) = 0, \qquad i = 1, 2, \cdots, n, \tag{2}$$

under the assumption of MAR, where $\gamma(u)$ denotes the density function of $U_1$. Using this, an auxiliary random vector can be defined as follows:

$$\mathcal{H}_i(\alpha(u)) = \Big\{\frac{\delta_i}{\pi(Z_i)}(Y_i - X_i^\tau \alpha(u))X_i\Big\} K_h(U_i - u),$$

$i = 1, 2, \cdots, n$, where $K_h(\cdot) = K(\cdot/h)$ and $h$ is the bandwidth. For the sake of convenience, in the sequel we drop the arguments $\alpha(u)$ and $Z_i$ from $\mathcal{H}_i(\alpha(u))$ and $\pi(Z_i)$, respectively.

Note that $\{\mathcal{H}_i\}_{i=1}^n$ are independent and satisfy $E\mathcal{H}_i = 0$ if and only if $\alpha(u)$ is the true parameter. By Owen (1991)???, a naive locally weighted ELR for $\alpha(u)$ can be defined as

$$\mathcal{L}_w(\alpha(u)) = -2\max\Big\{\sum_{i=1}^n \log(np_i)\Big| p_i \ge 0,\ \sum_{i=1}^n p_i = 1,\ \sum_{i=1}^n p_i\mathcal{H}_i = 0\Big\}.$$

Since $\mathcal{L}_w(\alpha(u))$ contains an unknown function $\pi(\cdot)$, it can not be utilized directly in the statistical inferences for $\alpha(u)$. A natural idea to solve this problem is to replace $\pi(\cdot)$ with its estimator, namely[22],

$$\widehat{\pi}_i := \widehat{\pi}(Z_i) = \sum_{j=1}^n \delta_i \mathcal{K}_1\Big(\frac{Z_i - Z_j}{b}\Big)\Big/\sum_{j=1}^n \mathcal{K}_1\Big(\frac{Z_i - Z_j}{b}\Big).$$

The corresponding estimated ELR is then as follows:

$$\widehat{\mathcal{L}}_w(\alpha(u)) = -2\mathrm{max}\Big\{\sum_{i=1}^n \log(np_i)\Big| p_i \ge 0,\ \sum_{i=1}^n p_i = 1,\ \sum_{i=1}^n p_i\widehat{\mathcal{H}}_i = 0\Big\}.$$

Assume that 0 lies inside in the convex hull of $\widehat{\mathcal{H}}_1, \cdots, \widehat{\mathcal{H}}_n$. By the Lagrange multiplier method, $\widehat{\mathcal{L}}_w(\alpha(u))$ can be represented as

$$\widehat{\mathcal{L}}_w(\alpha(u)) = 2\sum_{i=1}^n \log(1 + \lambda^\tau \widehat{\mathcal{H}}_i), \tag{3}$$

where $\lambda$ is a $p \times 1$ vector given as the solution to

$$\frac{1}{n}\sum_{i=1}^n \frac{\widehat{\mathcal{H}}_i}{1 + \lambda^\tau \widehat{\mathcal{H}}_i} = 0. \tag{4}$$

The following theorem gives the asymptotic distribution of $\widehat{\mathcal{L}}_w(\alpha(u))$.

**Theorem 2.1.** *Suppose that Assumptions A2-A6 hold (see the Appendix), and that $nh \to \infty$ and $nh^5 \to 0$ as $n \to \infty$. If $\alpha(u)$ is the true parameter, then we have*

$$\widehat{\mathcal{L}}_w(\alpha(u)) \to^L w_1\chi_{1,1}^2 + w_2\chi_{1,2}^2 + \cdots + w_p\chi_{1,p}^2,$$

*where $\to^L$ denotes the convergence in distribution, $\{\chi_{1,i}^2,\ 1 \le i \le p\}$ are the independent $\chi_1^2$ variables, and $w_i$'s are the eigenvalues of $\Sigma(u) = \Omega_2(u)^{-1}\Omega_1(u)$, which will be specified in the Appendix.*

In order to utilize Theorem 2.1 in practice, one needs to estimate the unknown weights $w_i$'s consistently. Denote

$$\widehat{\mathcal{R}}_i(\widehat{\alpha}(u)) = \Big\{\frac{\delta_i}{\widehat{\pi}_i}(Y_i - X_i^\tau \widehat{\alpha}(u))X_i + \frac{\widehat{\pi}_i - \delta_i}{\widehat{\pi}_i}\widehat{\Phi}_u(Z_i, \widehat{\alpha}(u))\Big\} K_h(U_i - u),$$

where $\widehat{\alpha}(u)$ denotes a consistent estimator of $\alpha(u)$ such as the locally weighted estimator proposed by [26], $\widehat{\Phi}_u(z, \alpha(u)) = Y\widehat{g}_1(z) - \widehat{g}_2(z)\alpha(u)$, $\widehat{g}_1(z)$ and $\widehat{g}_2(z)$ denote the Horvitz-Thompson (HT) bivariate local linear estimators of $g_1(z) = E(X|Z = z)$ and $g_2(z) = E(XX^\tau|Z = z)$, respectively. By adopting [11], $\widehat{g}_i(\cdot)$ $(i = 1, 2)$ converge to $g_i(\cdot)$ at order $n^{\frac{1}{4}}$ uniformly. Then it is easy to show that $\widehat{\Sigma}(u) = \widehat{\Omega}_2(u)^{-1}\widehat{\Omega}_1(u)$ is a consistent estimator of $\Sigma(u)$ (see Lemma A.3), where

$$\widehat{\Omega}_1(u) = \frac{1}{n}\sum_{i=1}^{n} \widehat{\mathcal{R}}_i(\widehat{\alpha}(u))(\widehat{\mathcal{R}}_i(\widehat{\alpha}(u)))^\tau,$$
$$\widehat{\Omega}_2(u) = \frac{1}{n}\sum_{i=1}^{n} \widehat{\mathcal{H}}_i(\widehat{\alpha}(u))(\widehat{\mathcal{H}}_i(\widehat{\alpha}(u)))^\tau.$$

Therefore, one can take the eigenvalues of $\widehat{\Sigma}(u)$ to be the consistent estimators of $w_i$'s.

However, the accuracy of this procedure depends on the values of $w_i$'s. Along the line of [23], we will give an adjusted ELR, which is exactly standard chi-squared distributed asymptotically. Denote

$$\widehat{r}(\widehat{\alpha}(u)) = \mathrm{tr}\{\widehat{\Omega}_1^{-1}(u)\widehat{V}(u)\}/\mathrm{tr}\{\widehat{\Omega}_2^{-1}(u)\widehat{V}(u)\},$$

where

$$\widehat{V}(u) = \Big(\sum_{i=1}^{n} \widehat{\mathcal{H}}_i(\widehat{\alpha}(u))\Big)\Big(\sum_{i=1}^{n} \widehat{\mathcal{H}}_i(\widehat{\alpha}(u))\Big)^\tau.$$

Then an adjusted ELR can be defined as follows

$$\widehat{\mathcal{L}}_{ad}(\alpha(u)) = \widehat{r}(\alpha(u)) \times \widehat{\mathcal{L}}_w(\alpha(u)),$$

where $\widehat{r}(\alpha(u))$ is obtained from $\widehat{r}(\widehat{\alpha}(u))$ by replacing $\widehat{\alpha}(u)$ with $\alpha(u)$. Similar to the proof of Theorem 2 in [23], we can show that

**Theorem 2.2.** *If $\alpha(u)$ is the true parameter, then under the same assumptions of Theorem 2.1, we have*

$$\widehat{\mathcal{L}}_{ad}(\alpha(u)) \longrightarrow^L \chi^2_p.$$

*where $\chi^2_p$ is a chi-squared variable with p degrees of freedom.*

Furthermore, from Theorem 2.1, a MELE of $\alpha(u)$, say $\widehat{\alpha}_w(u)$, can be defined by minimizing $\widehat{\mathcal{L}}_w(\alpha(u))$. Write

$$\widehat{\mathcal{A}}_{1n}(u) = \frac{1}{nh}\sum_{i=1}^{n} \frac{\delta_i}{\widehat{\pi}_i} X_i X_i^\tau K_h(U_i - u),$$
$$\widehat{\mathcal{B}}_{1n}(u) = \frac{1}{nh}\sum_{i=1}^{n} \frac{\delta_i}{\widehat{\pi}_i} X_i Y_i K_h(U_i - u).$$

Assume that the matrix $\widehat{\mathcal{A}}_{1n}(u)$ is invertible, then it is easy to show that

$$\widehat{\alpha}_w(u) = \widehat{\mathcal{A}}_{1n}^{-1}(u)\widehat{\mathcal{B}}_{1n}(u) + o_p((nh)^{-1/2}).$$

This further implies that the following theorem.

**Theorem 2.3.** *Under the Assumptions of A1–A6, we have*

$$\sqrt{nh}\{\widehat{\alpha}_w(u) - \alpha(u) - b_1(u)\} \longrightarrow^L N\Big(0, \frac{v_0}{\gamma(u)}\Psi(u)^{-1}\Omega_1(u)\Psi(u)^{-1}\Big),$$

*where*

$$v_0 = \int K^2(t)dt, \qquad \Psi(u) = E(XX^\tau | U = u), \qquad b_1(u) = \frac{1}{2}h^2\mu_2\alpha''(u) + o_p(h^2)$$

*with* $\mu_2 = \int t^2 K(t)dt$.

This theorem can be shown similarly as that of Theorem 3 in [26]. In other words, this ELME is asymptotically equivalent to the locally weighted estimator proposed by [26].

### 2.2 A $R$-type Locally Weighted Empirical Log-likelihood Ratio

To overcome the mismatch problem of the first ELR, inspired by [10] and [18], we suggest an another new ELR, named $R$-type locally weighted ELR, by employing the following constrain

$$E\Big\{\frac{\delta_i}{\pi_i}(Y_i - X_i^\tau \alpha(U_i))X_i + \Big(1 - \frac{\delta_i}{\pi_i}\Big)\Phi_u(Z_i)\Big| U_i = u\Big\}\gamma(u) = 0, \qquad i = 1, 2, \cdots, n, \tag{5}$$

and auxiliary random vector

$$\mathcal{R}_i(\alpha(u)) = \Big\{\frac{\delta_i}{\pi_i}(Y_i - X_i^\tau \alpha(u))X_i + \Big(1 - \frac{\delta_i}{\pi_i}\Big)\Phi_u(Z_i)\Big\}K_h(U_i - u),$$

$i = 1,\ 2,\ \cdots,\ n$, where

$$\Phi_u(Z_i) = E((Y_i - X_i^\tau \alpha(u))X_i | Z_i = z) = Y_i g_1(Z_i) - g_2(Z_i)\alpha(u).$$

Similar to Section 2.1, by replacing the unknown functions, i.e. $\pi(\cdot)$, $g_1(\cdot)$ and $g_2(\cdot)$, with their estimators, i.e. $\widehat{\pi}(\cdot)$, $\widehat{g}_1(\cdot)$, and $\widehat{g}_2(\cdot)$, respectively, we obtain the following estimated empirical log-likelihood ratio

$$\widehat{\mathcal{L}}_r(\alpha(u)) = -2\max\Big\{\sum_{i=1}^n \log(np_i)\Big| p_i \geq 0,\ \sum_{i=1}^n p_i = 1,\ \sum_{i=1}^n p_i\widehat{\mathcal{R}}_i = 0\Big\}.$$

Here $\widehat{\mathcal{R}}_i = \widehat{\mathcal{R}}_i(\alpha(u))$. The following theorem states the asymptotic distribution of $\widehat{\mathcal{L}}_r(\alpha(u))$.

**Theorem 2.4.** *If $\alpha(u)$ is the true parameter, then under the same assumptions of Theorem 2.1, we have $\widehat{\mathcal{L}}_r(\alpha(u)) \to^L \chi^2_p$.*

As a byproduct, an another MELE of $\alpha(u)$, write $\widehat{\alpha}_r(u)$, can be defined by minimizing $\widehat{\mathcal{L}}_r(\alpha(u))$. Denote

$$\widehat{\mathcal{A}}_{2n}(u) = \frac{1}{nh}\sum_{i=1}^n \Big\{\frac{\delta_i}{\widehat{\pi}_i}X_iX_i^\tau + \frac{\widehat{\pi}_i - \delta_i}{\widehat{\pi}_i}\widehat{g}_2(Z_i)\Big\}K_h(U_i - u),$$
$$\widehat{\mathcal{B}}_{2n}(u) = \frac{1}{nh}\sum_{i=1}^n \Big\{\frac{\delta_i}{\widehat{\pi}_i}X_iY_i + \frac{\widehat{\pi}_i - \delta_i}{\widehat{\pi}_i}\widehat{g}_1(Z_i)Y_i\Big\}K_h(U_i - u).$$

Assume that $\widehat{\mathcal{A}}_{2n}(u)$ is invertible. Similar to Section 2.1, we have

$$\widehat{\alpha}_r(u) = \widehat{\mathcal{A}}_{2n}^{-1}(u)\widehat{\mathcal{B}}_{2n}(u) + o_p((nh)^{-1/2}).$$

The following theorem of $\widehat{\alpha}_r(u)$ can be shown similarly as Theorem 4 of [26].

**Theorem 2.5.** *Under the same assumptions of Theorem 2.3, we have*

$$\sqrt{nh}\{\widehat{\alpha}_r(u) - \alpha(u) - b_1(u)\} \longrightarrow^L N\Big(0,\ \frac{v_0}{\gamma(r)}\Psi(u)^{-1}\Omega_1(u)\Psi(u)^{-1}\Big).$$

**Remark 2.1.** The results above show that, although constrain (2) has an advantage over (5) in terms of establishing estimators (see also Remark 1 of [26]), (5) is more suitable for constructing empirical likelihood based regions.

### 2.3 A Residual-adjusted Locally Weighted Empirical Log-likelihood Ratio

Although Theorem 2.4 has removed the mismatch problem, the range of $h$ is within the interval $(c_1 n^{-1/2}, c_2 n^{-1/5})$ for some positive constants $c_1$ and $c_2$, which does not contain the optimal bandwidth $h_0 = n^{-1/5}$. Similar to [28], we propose a residual-adjusted locally weighted ELR relying on Section 2.2.

Let $\mathcal{E}_i := \mathcal{E}_i(\alpha(u)) = \mathcal{R}_i(\alpha(u)) - \phi_i(u) \ \ (i = 1, 2, \cdots, n)$, where

$$\phi_i(u) = \Big\{\frac{\delta_i}{\widehat{\pi}_i} X_i X_i^\tau + \Big(1 - \frac{\delta_i}{\widehat{\pi}_i}\Big)\widehat{g}_2(Z_i)\Big\}(\widehat{\alpha}_r(U_i) - \widehat{\alpha}_r(u))K_h(U_i - u).$$

with $\widehat{\alpha}_r(u)$ being the MELE given in Section 2.2. Clearly, $\mathcal{E}_i(\alpha(u))$'s are adjustments of $\mathcal{R}_i(\alpha(u))$'s. Similarly, by substituting the unknowns, we obtain the following estimated residual-adjusted locally weighted ELR

$$\widehat{\mathcal{L}}_{\mathcal{E}}(\alpha(u)) = -2\max\Big\{\sum_{i=1}^n \log(np_i)\Big| \, p_i \ge 0, \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i \widehat{\mathcal{E}}_i = 0\Big\}.$$

The asymptotic result of $\widehat{\mathcal{L}}_{\mathcal{E}}(\alpha(u))$ is provided as follows.

**Theorem 2.6.** *Under the Assumptions of A1–A6, we have*

$$\widehat{\mathcal{L}}_{\mathcal{E}}(\alpha(u)) \to^L \chi_p^2.$$

*Let $\chi_p^2(1-\theta)$ be the $1-\theta$ quantile of $\chi_p^2$ $(0 < \theta < 1)$. Using the results of Theorems 2.2, 2.4 and 2.6, an approximate $1-\theta$ pointwise confidence region for $\alpha(u)$ can be given as*

$$\varphi_\theta(\alpha(u)) = \big\{\beta \in R^p \mid \widehat{l}(\beta) \le \chi_p^2(1-\theta)\big\},$$

*where $\widehat{l}$ denotes $\widehat{\mathcal{L}}_{ad}$, $\widehat{\mathcal{L}}_R$ or $\widehat{\mathcal{L}}_{\mathcal{E}}$, respectively.*

### 2.4 Partial Profile Empirical Log-likelihood Ratio

In order to construct the pointwise confidence interval for a component, say $\alpha_k(u)$, of $\alpha(u)$, we utilize the partial profile empirical likelihood method, and define an estimated empirical log-likelihood ratio as follows

$$\widehat{\mathcal{L}}_{\mathcal{E},k}(\alpha_k(u)) = 2\sum_{i=1}^n \log(1 + \lambda^\tau \widehat{\mathcal{E}}_{i,k}(\alpha_k(u))),$$

where $1 \le k \le p$, $\widehat{\mathcal{E}}_{i,k}(\alpha_k(u)) = e_k^\tau \widehat{\mathcal{A}}_{2n}^{-1}(u)\widehat{\mathcal{E}}_i(\widehat{\alpha}_r^1(u), \cdots, \widehat{\alpha}_r^{k-1}(u), \alpha_k(u), \widehat{\alpha}_r^{k+1}(u), \cdots, \widehat{\alpha}_r^p(u))$, $e_k$ is a $p$-dimensional vector with $k$-th component 1, and $\widehat{\alpha}_r^j(u) = e_j^\tau \widehat{\alpha}_r(u)$ is the $j$-th component of $\widehat{\alpha}_r(u)$. Similar to [28], we have

**Theorem 2.7.** *Under the same assumptions of Theorem 2.6, we have*

$$\widehat{\mathcal{L}}_{\mathcal{E},k}(\alpha_k(u)) \longrightarrow^L \chi_1^2.$$

*This theorem implies that an approximate* $1-\theta$ *confidence interval of* $\alpha_k(u)$ *can be defined as follows:*

$$\varphi_{\theta,k}(\alpha_k(u)) = \big\{\beta \in R^1 \big| \widehat{\mathcal{L}}_{\mathcal{E},k}(\beta) \le \chi_1^2(1-\theta)\big\}.$$

## 3 Simulation Studies

To investigate the finite sample properties of the proposed methods, some simulation studies are constructed in this section.

Suppose that

$$Y = \alpha_1(U)X_1 + \alpha_2(U)X_2 + \varepsilon,$$

where $\alpha_1(u) = \sin(2\pi u)$, $\alpha_2(u) = \exp(-(3u-1)^2)$, $X_1$ and $X_2$ are uniformly distributed over $[-1,1]$, $U$ subjects to uniform distribution over $[0,1]$ and the model error $\varepsilon$, independent of $U$ and $X$, follows the normal distribution $N(0,\, 0.5^2)$.

Three methods are compared: (a) CCA, (b) normal approximation (NA), see [26], and (c) resident-adjusted locally weighted empirical log-likelihood ratio (RAELR). The average lengths and coverage probabilities of the pointwise intervals, with a nominal level $1-\theta = 95\%$, are computed based on 1000 simulations. We choose the following two missing data mechanisms:

**Case I.** $\pi_1(y,t) = \exp(1+0.15y+0.2t)/(1+\exp(1+0.15y+0.2y))$.

**Case II.** $\pi_2(y,t) = \exp(0.5+0.15y+0.2t)/(1+\exp(0.5+0.15y+0.2t))$.

For each case, the sample size is $n = 200$. The kernel function $\mathcal{K}(\cdot)$ is taken to be $\mathcal{K}(z) = \mathcal{K}_1(y)\mathcal{K}_2(u)$ with

$$\mathcal{K}_i(t) = 15/16(1-t^2)^2 I(|t| \le 1), \qquad i = 1,2.$$

The bandwidth $b_{CV}$ is selected by minimizing

$$CV_1(b) = \frac{1}{n}\sum_{i=1}^{n}(\delta_i - \widehat{\pi}^{(-i)}(Z_i))^2,$$

where $\widehat{\pi}^{(-i)}(z)$ is a "delete one out" version of $\widehat{\pi}(z)$. While the kernel function $K(u)$ is taken to be $K(u) = 0.75(1-u^2)I(|u|<1)$ with $h_{CV}$ selected by minimizing

$$CV_2(h) = \frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i}{\widehat{\pi}_i}[Y_i - X_i^\tau \widehat{\alpha}_w^{(-i)}(U_i,h)]^2,$$

where $\widehat{\alpha}_w^{(-i)}(u,h)$ is a "delete one out" version of MELE $\widehat{\alpha}_w(u)$. Furthermore, we choose $n^{-\frac{1}{6}}$ as the bandwidth for estimating $g_1(\cdot)$ and $g_2(\cdot)$.

To construct a confidence region for $\alpha(u)$ by using the NA method, we first estimate the asymptotic covariance matrix $\Sigma_1(u)$ of the locally weighted estimator given in [26]. That is,

$$\widehat{\Sigma}_1(u) = \frac{v_0}{\widehat{\gamma}(u)}\widehat{\Psi}(u)^{-1}\widehat{\Omega}_1(u)\widehat{\Psi}(u)^{-1},$$

where

$$\widehat{\gamma}(u) = \frac{1}{n}\sum_{i=1}^{n}K_{h_{CV}}(U_i-u), \qquad \widehat{\Psi}(u) = \frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i}{\widehat{\pi}_i}X_iX_i^\tau K_{h_{CV}}(U_i-u).$$

The bias term $\alpha''(u)$ is estimated by using the method proposed in Remark 2 of [26]. For the quasi-kernel function, we set $v_0 = 5/7$, $\mu_2 = 1/7$.

The average missing rates of these two cases are approximately 0.25 and 0.354, respectively. Figure 1, corresponding to Case (I), reports the 95% pointwise average confidence intervals over [0, 1], as well as the average lengths and coverage probabilities of these intervals, for $\alpha_1(u)$ and $\alpha_2(u)$. While Figure 2 reports analogous results corresponding to Case (Ⅱ). The dashed, dash-dotted and plus curve correspond to RAELR, CCA and NA, respectively.

From Figures 1 and 2, we can see that RAELR and CCA give much narrower confidence pointwise intervals than NA, although NA can yield higher coverage probabilities. Compared with RAELR, CCA performs poorly since its pointwise intervals may not cover the true curve with a high probability, especially when the curvature is larger; see, for example, 2 (e) of Figure 2. Furthermore, as the missing rate increases, the lengths of the pointwise intervals increase and the coverage probabilities decrease for both $\alpha_1(u)$ and $\alpha_2(u)$. This implies that the missing rate also has an impact on the performance of the proposed methods.

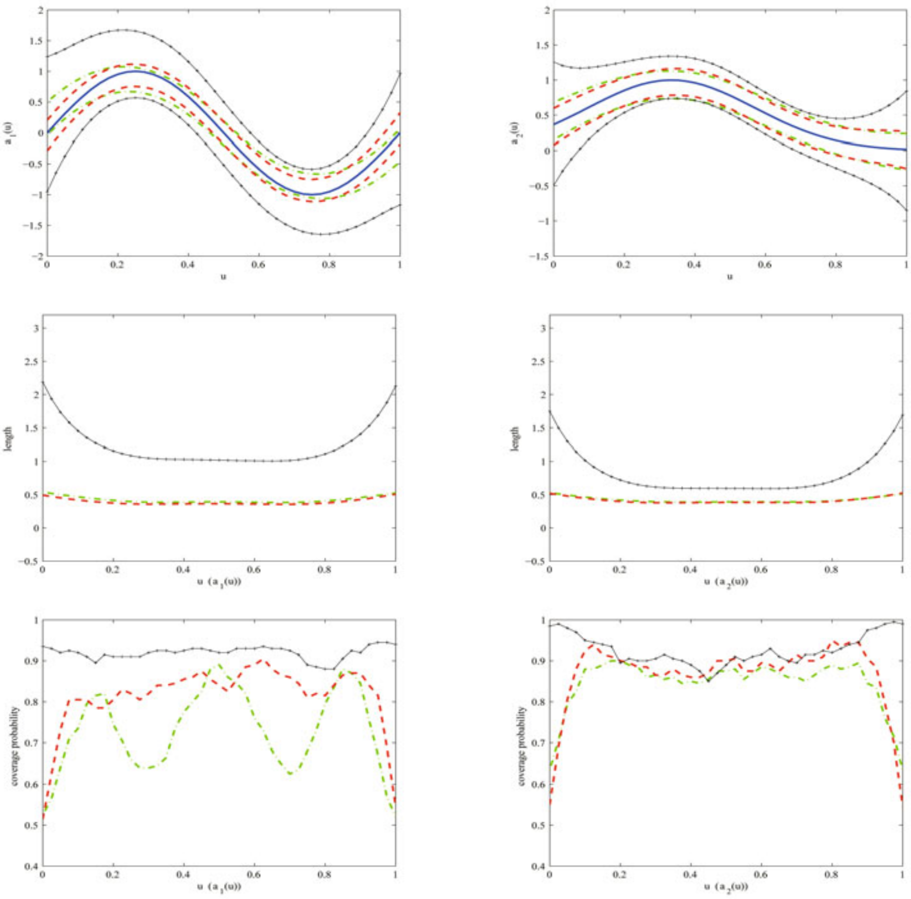

**Figure 1.** The 95% Pointwise Average Confidence Intervals, and the Corresponding Average Lengths and Coverage Probabilities of $\alpha_i(u)$ $(i = 1, 2)$ for Case (I). 1(a), 1(c), 1(e) Correspond to $\alpha_1(u)$, the Others Correspond to $\alpha_2(u)$.

To provide more information on the comparison of these three methods, we also consider the asymptotic confidence regions of $(\alpha_1(u), \alpha_2(u))$. Here, we only consider the asymptotic confidence regions at $u = 0.2$, because the other case with a different $u$ can be investigated similarly. The simulation results are reported in Figure 3, which shows that RAELR performs best among these three methods. Therefore, we recommend RAELR for constructing confidence pointswise regions/intervals for Model (1.1) when there are missing covariates.

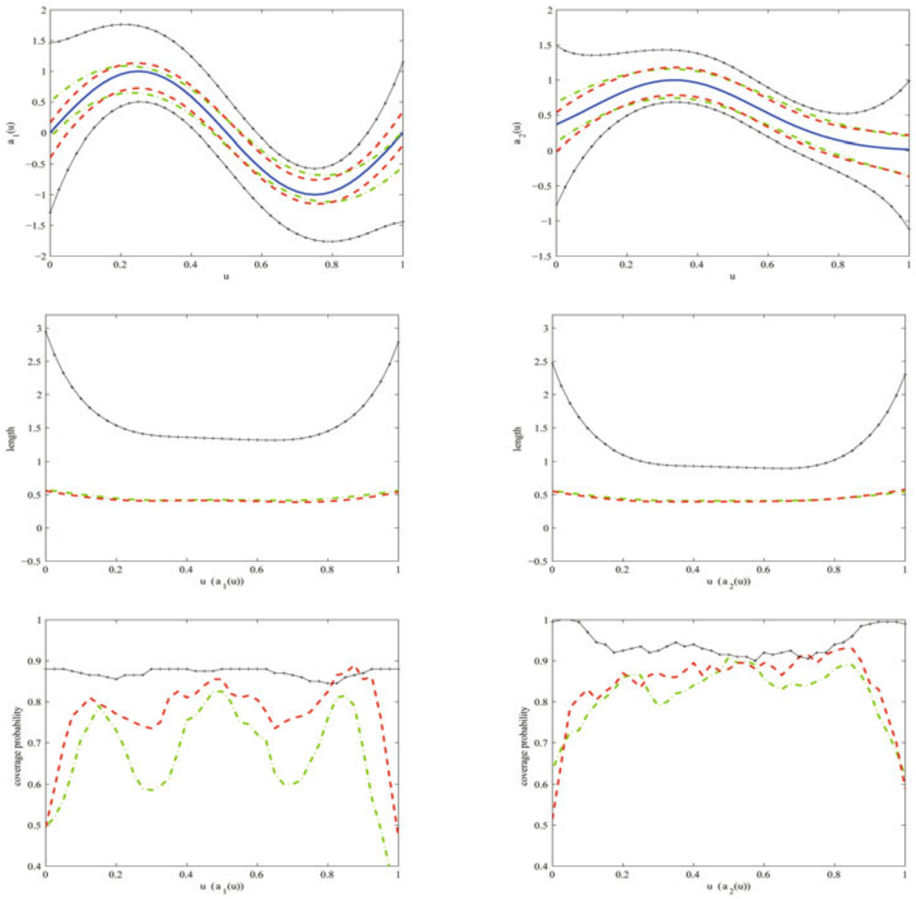

**Figure 2.** The 95% Pointwise Average Confidence Intervals, and the Corresponding Average Lengths and Coverage Probabilities of $\alpha_i(u)$ $(i = 1, 2)$ for Case (I). 2(a), 2(c), 2(e) Correspond to $\alpha_1(u)$, the Others Correspond to $\alpha_2(u)$.

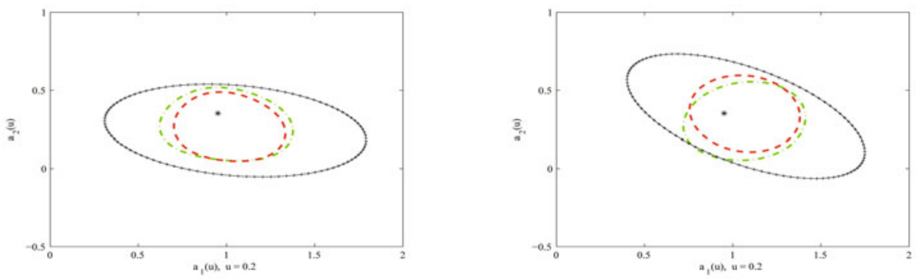

**Figure 3.** The 95% Pointwise Confidence Regions of $(\alpha_1(u), \alpha_2(u))$ with $u = 0.2$, where 3(a) and (3b) Correspond to Case (I) and (II), Respectively.

## 4 Concluding Discussions

In this paper, we applied the empirical likelihood method to the varying coefficient models when the covariates $X$ may be missing at random. A class of empirical log-likelihood ratios for $\alpha(u)$ have been proposed relying on the locally weighted estimating equations and the nonparametric version of the Wilks' theorem has also been derived. So the confidence regions for the nonparametric part $\alpha(u)$ with asymptotically correct coverage probabilities can be constructed. In addition, we also obtained the asymptotic normality of the maximum empirical likelihood estimators of $\alpha(u)$. Interesting works for further researches include applying the empirical likelihood method to inferences for $\alpha(u)$ when the responses $Y$ may be missing at random, and developing variable selection procedures for such models with missing data, since the existing procedures (see, for example, [21]) can not be used directly any more when missing data are present.

## Appendix: Proofs of the Main Results

For the sake of convenience, let $c$ be a positive constant which may be a different value at each appearance throughout this paper. To derive the main results, we need the following technical assumptions.

**A1**. The bandwidth satisfies $h = cn^{-1/5}$ for a constant $c > 0$.

**A2**. The kernel $K(\cdot)$ is a bounded and symmetric probability density function, and satisfies $\int u^4 K(u)du < \infty$.

**A3**. The density, say $\gamma(u)$, of $U$ is bounded away from 0, and has continuous first derivatives at $u$. The density $\pi(\cdot)$ has bounded partial derivatives up to order $k(> 2)$ almost surely, with $\inf_z \pi(z) > 0$.

**A4**. $\alpha(\cdot)$, $g_1(\cdot)$, $g_2(\cdot)$ and $\Psi(\cdot)$ are twice continuously differentiable. Furthermore, assume $\alpha_j''(u) \neq 0$, for $j = 1, \cdots, p$, and $\Psi(\cdot)$ is a $p \times p$ positive definite matrix for any given $u$.

**A5**. $b$ satisfies $nb^4 \to \infty$ and $nbh^2 \to \infty$, $h/b \to 0$, $b^k/h \to 0$ and $nb^{2k+1} \to 0$ for $k > 2$.

**A6**. $\sup_u E[\|X\|^4|U = u] < \infty$, where $\|\cdot\|$ denotes the Euclidean distance. The model errors satisfy $\sup\limits_i(E\varepsilon_i^4) < \infty$.

Assumption A1, A2, A4, A6 are regular and often seen in the literature; see [26] and [28]. A3 and A5 are usually utilized when missing data are present (see [22]). An example for A5 to be satisfied is that $h = n^{-1/5}$ (or $n^{-1/4}$) and $b = n^{-1/6}$.

The proofs of Theorems 2.1–2.5 rely on the following lemmas.

**Lemma 1** (Abel). *Let $\{a_i\}_{i=1}^n$, $\{b_i\}_{i=1}^n$ be two sequences of real numbers, $S_k = \sum\limits_{i=1}^k a_i$. Then*

$$\max_{1\le i\le n}\Big|\sum_{i=1}^n a_i b_i\Big| \le c \max_{1\le i\le n}|b_i| \max_{1\le i\le n}|S_i|.$$

**Lemma 2.** *Let $v_i$ be i.i.d. r.v.s. with $Ev_i = 0$ and $Ev_i^2 < \infty$. Then for any permutation $(j_1, \cdots, j_n)$ of $(1, \cdots, n)$, we have*

$$\max_{1\le k\le n}\Big|\sum_{i=1}^k v_{j_i}\Big| = O_p(n^{\frac{1}{2}}\log n).$$

This lemma comes from [6].

**Lemma 3.** *Under the same assumptions of Theorem 2.1, we have*

$$\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\widehat{\mathcal{H}}_i \longrightarrow^{L} N(0,\, v(u)\Omega_1(u)),$$

*where* $v(u)=\gamma(u)v_0$,

$$\Omega_1(u)=E\Big(\frac{1}{\pi}\{X\varepsilon\}^{\otimes 2}+\frac{\pi-1}{\pi}E(X\varepsilon|Z)^{\otimes 2}\big|U=u\Big).$$

*Proof.* Clearly

$$\begin{aligned}\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\widehat{\mathcal{H}}_i =&\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\frac{\delta_i}{\widehat{\pi}_i}X_i\varepsilon_iK_h(U_i-u)\\ &+\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\frac{\delta_i}{\widehat{\pi}_i}X_iX_i^{\tau}\Gamma_i(u)K_h(U_i-u)\\ =&A_1+A_2.\end{aligned}$$

Hereafter, we denote $\Gamma_i(u)=\alpha(U_i)-\alpha(u)$ and $\Gamma_{i,k}(u)=\alpha_k(U_i)-\alpha_k(u)$ for convenience.

Similar to the proof of Theorem 4 in [22], we can prove that

$$A_1=\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\Big\{\frac{\delta_i}{\pi_i}X_i\varepsilon_i+\frac{\pi_i-\delta_i}{\pi_i}E(X\varepsilon|Z_i)\Big\}K_h(U_i-u)+o_p(1). \tag{6}$$

Next, for term $A_2$, we have

$$\begin{aligned}A_2 =&\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\frac{\delta_i}{\pi_i}X_iX_i^{\tau}\Gamma_i(u)K_h(U_i-u)\\ &+\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\frac{\delta_i}{\pi_i^2}(\pi_i-\widehat{\pi}_i)X_iX_i^{\tau}\Gamma_i(u)K_h(U_i-u)\\ &+\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\frac{\delta_i}{\pi_i^2\widehat{\pi}_i}(\pi_i-\widehat{\pi}_i)^2X_iX_i^{\tau}\Gamma_i(u)K_h(U_i-u)\\ =&A_{21}+A_{22}+A_{23},\end{aligned}$$

Since we have

$$\begin{aligned}E(A_{21,s}^2) =&\frac{1}{nh}E\Big(\sum_{i=1}^{n}\Big\{\frac{\delta_i}{\pi_i^2}X_{i,s}^2\Big(\sum_{k=1}^{p}X_{i,k}\Gamma_{i,k}(u)\Big)^2K_h^2(U_i-u)\Big\}\Big)\\ &+\frac{1}{nh}E\Big(\sum_{i\neq j}\Big\{\frac{\delta_i}{\pi_i}X_{i,s}\Big(\sum_{k=1}^{p}X_{i,k}\Gamma_{i,k}(u)\Big)K_h(U_i-u)\Big\}\\ &\times\Big\{\frac{\delta_j}{\pi_j}X_{j,s}\Big(\sum_{k=1}^{p}X_{j,k}\Gamma_{j,k}(u)\Big)K_h(U_j-u)\Big\}\Big)\\ =&A_{21,s}^{[1]}+A_{21,s}^{[2]},\end{aligned}$$

for any $1 \le s \le p$. Here $A_{21,s}$, $X_{i,s}$ and $\alpha_s(\cdot)$ denote the $s$-th component of $A_{21}$, $X_i$ and $\alpha(\cdot)$ respectively. Then similar to the proof of Lemma 1 in [27], one can show that

$$\begin{aligned} A_{21,s}^{[1]} \le & \frac{1}{\inf_z^2 \pi(z)} \frac{1}{nh} \sum_{i=1}^n E\Big(X_{i,s}^2\Big(\sum_{k=1}^p X_{i,k}\Gamma_{i,k}(u)\Big)^2 K_h^2(U_i-u)\Big) \\ = & \frac{c}{nh} \cdot nh^3 \cdot \{1+o_p(1)\} \longrightarrow 0 \end{aligned}$$

and

$$\begin{aligned} A_{21,s}^{[2]} = & \frac{1}{nh} \sum_{i \ne j}^n \Big(\sum_{k=1}^p \int \Psi_{s,k}(t)(\alpha_k(t)-\alpha_k(u))K\Big(\frac{t-u}{h}\Big)\gamma(t)dt\Big)^2 \\ \le & \frac{1}{nh} \cdot n^2 \cdot (ch^3)^2 \cdot \{1+o_p(1)\} \longrightarrow 0, \end{aligned}$$

where $\Psi_{s,k}(u)$ denotes the $(s,k)$-th component of $\Psi(u)$. This proves

$$A_{21} = o_p(1). \tag{7}$$

For term $A_{22}$, since

$$\begin{aligned} A_{22} = & \frac{1}{\sqrt{nh}} \sum_{i=1}^n \frac{\delta_i X_i X_i^\tau \Gamma_i(u) K_h(U_i-u)}{nb^2\pi_i^2 f_{\mathcal{Z}}(Z_i)} \sum_{j=1}^n (\pi_j-\delta_j)\mathcal{K}_b(Z_j-Z_i) \\ & + \frac{1}{\sqrt{nh}} \sum_{i=1}^n \frac{\delta_i X_i X_i^\tau \Gamma_i(u) K_h(U_i-u)}{nb^2\pi_i^2 f_{\mathcal{Z}}(Z_i)} \sum_{j=1}^n (\pi_i-\pi_j)\mathcal{K}_b(Z_j-Z_i) + o_p(1) \\ = & A_{22}^{[1]} + A_{22}^{[2]} + o_p(1). \end{aligned}$$

By interchanging the summation, we obtain

$$A_{22}^{[1]} = \frac{1}{\sqrt{nh}} \sum_{j=1}^n \frac{\pi_j-\delta_j}{\pi_j} E(XX^\tau|Z_j)\Gamma_j(u)K_h(U_j-u) + o_p(1).$$

Then following a similar fashion of (7), we have $A_{22}^{[1]} = o_p(1)$. Next, similar to (81) (see [22, p.77]), we obtain $A_{22}^{[2]} = o_p(1)$. This proves

$$A_{22} = o_p(1). \tag{8}$$

By the fact that $\sup_z\{|\pi-\widehat{\pi}|\} = O_p((nb^2)^{-\frac{1}{2}}) + O_p(b^k)$, we have

$$|A_{23,k}| \le C\sqrt{nh}(\sup_z\{|\pi-\widehat{\pi}|\})^2 \frac{1}{nh} \sum_{i=1}^n |X_{i,k}| \cdot \|X_i\| \cdot \|\Gamma_i(u)\| \cdot K_h(U_i-u) = o_p(1).$$

This, together with (6)–(8), proves

$$\frac{1}{\sqrt{nh}} \sum_{i=1}^n \widehat{\mathcal{H}}_i = \frac{1}{\sqrt{nh}} \sum_{i=1}^n \Big\{\frac{\delta_i}{\pi_i} X_i\varepsilon_i + \frac{\pi_i-\delta_i}{\pi_i} E(X\varepsilon|Z_i)\Big\} K_h(U_i-u) + o_p(1).$$

Finally, this lemma follows immediately by using the central limit theorem. □

**Lemma 4.** *Under the same assumptions of Lemma 3, we have*

$$\frac{1}{nh}\sum_{i=1}^{n}\widehat{\mathcal{H}}_i\widehat{\mathcal{H}}_i^{\tau} \to^p v(u)\Omega_2(u),$$

*where* $\to^p$ *denotes the convergence in probability, and*

$$\Omega_2(u) = E\Big(\frac{1}{\pi}\{X\varepsilon\}^{\otimes 2}\Big|U=u\Big).$$

*Proof.* Clearly

$$\begin{aligned}
&\frac{1}{nh}\sum_{i=1}^{n}\widehat{\mathcal{H}}_{1i}\widehat{\mathcal{H}}_{1i}^{\tau}\\
=&\frac{1}{nh}\sum_{i=1}^{n}\frac{\delta_i}{\widehat{\pi}_i^2}X_iX_i^{\tau}\varepsilon_i^2K_h^2(U_i-u)+\frac{1}{nh}\sum_{i=1}^{n}\frac{\delta_i}{\widehat{\pi}_i^2}X_iX_i^{\tau}(X_i^{\tau}\Gamma_i(u))^2K_h^2(U_i-u)\\
&+\frac{2}{nh}\sum_{i=1}^{n}\frac{\delta_i}{\widehat{\pi}_i^2}X_iX_i^{\tau}\varepsilon_iX_i^{\tau}\Gamma_i(u)K_h^2(U_i-u)\\
=&B_1+B_2+B_3,
\end{aligned}$$

where

$$\begin{aligned}
B_1=&\frac{1}{nh}\sum_{i=1}^{n}\frac{\delta_i}{\pi_i^2}X_iX_i^{\tau}\varepsilon_i^2K_h^2(U_i-u)+\frac{1}{nh}\sum_{i=1}^{n}\left\{\frac{1}{\widehat{\pi}_i^2}-\frac{1}{\pi_i^2}\right\}X_iX_i^{\tau}\varepsilon_i^2K_h^2(U_i-u)\\
=&B_{11}+B_{12}.
\end{aligned}$$

Note that

$$\begin{aligned}
|B_{12}^{[k,l]}|=&\Big|\frac{1}{nh}\sum_{i=1}^{n}X_{i,k}X_{i,l}\varepsilon_i^2K_h^2(U_i-u)\frac{(\pi_i+\widehat{\pi}_i)(\pi_i-\widehat{\pi}_i)}{\pi_i^2\widehat{\pi}_i^2}\Big|\\
\le&\sup_z|\pi_i-\widehat{\pi}_i|\,\frac{1}{(\inf_z\pi)^4}\frac{1}{nh}\sum_{i=1}^{n}|X_{i,k}X_{i,l}|\varepsilon_i^2K_h^2(U_i-u)\\
=&o_p(1),
\end{aligned}$$

which implies that

$$B_1=\frac{1}{nh}\sum_{i=1}^{n}\frac{\delta_i}{\pi_i^2}X_iX_i^{\tau}\varepsilon_i^2K_h^2(U_i-u)+o_p(1).$$

For term $B_2$, we have

$$\begin{aligned}
B_2=&\frac{1}{nh}\sum_{i=1}^{n}\frac{\delta_i}{\pi_i^2}X_iX_i^{\tau}(X_i^{\tau}\Gamma_i(u))^2K_h^2(U_i-u)+o_p(1)\\
\le&c\cdot\sup_t\|\alpha'(t)\|\cdot\int E(\|X\|^4|U=t)(t-u)^2K^2\Big(\frac{t-u}{h}\Big)dt\\
\le&ch^2\int s^2K^2(s)ds\longrightarrow 0.
\end{aligned}$$

Similarly, we can show $B_3=o_p(1)$. Finally, this lemma follows immediately by using the law of large number. □

**Lemma 5.** *Under the same assumptions of Lemma 3, we have*

$$\max_{1\le i\le n}\|\widehat{\mathcal{H}}_i\| = o_p((nh)^{\frac{1}{2}}) \qquad \textit{and} \qquad \lambda = O_p((nh)^{-\frac{1}{2}}).$$

*Proof.* Denote $\mathcal{U}_i(u) = X_i\varepsilon_i K_h(U_i-u)$, $\mathcal{V}_i(u) = X_iX_i^\tau\Gamma_i(u)K_h(U_i-u)$ and

$$\mathcal{F}_n = \Big\{\max_{1\le i\le n}\Big|\frac{1}{\widehat{\pi}_i}\Big| \ge \frac{1}{\frac{1}{2}\inf_z \pi}\Big\}.$$

By Assumption A3, it is easy to show that

$$P(\mathbf{I}_{\{\mathcal{F}_n\}}=1) \le P\Big(\bigcup_{i=1}^n |\widehat{\pi}_i-\pi_i| \ge \frac{1}{2}\inf_z \pi\Big) \le P\Big(\sup_z |\widehat{\pi}_i-\pi_i| \ge \frac{1}{2}\inf_z \pi\Big) \longrightarrow 0.$$

Then by using arguments as Lemma 1 of [28], we can obtain

$$\begin{aligned}
&P\Big(\max_{1\le i\le n}\Big\|\frac{\delta_i}{\widehat{\pi}_i}\mathcal{U}_i(u)\Big\| > c(nh)^{\frac{1}{2}}\Big)\\
\le& P\big(\mathbf{I}_{\{\mathcal{F}_n\}}=1\big) + P\Big(\max_{1\le i\le n}\Big\|\frac{\delta_i}{\widehat{\pi}_i}\mathcal{U}_i(u)\Big\| > c(nh)^{\frac{1}{2}}, \mathbf{I}_{\{\mathcal{F}_n\}}=0\Big)\\
\le& P\big(\mathbf{I}_{\{\mathcal{F}_n\}}=1\big) + P\Big(\max_{1\le i\le n}\|\mathcal{U}_i(u)\| > c(nh)^{\frac{1}{2}}\Big) \longrightarrow 0.
\end{aligned}$$

This leads to

$$\max_{1\le i\le n}\Big\|\frac{\delta_i}{\widehat{\pi}_i}\mathcal{U}_i(u)\Big\| \le o_p((nh)^{1/2}).$$

Similarly, we obtain

$$\max_{1\le i\le n}\Big\|\frac{\delta_i}{\widehat{\pi}_i}\mathcal{V}_i(u)\Big\| \le o_p((nh)^{1/2}).$$

This completes the first part. Since the proof of the second part follows a similar fashion as (2.14) in [15], so we omit it.

*Proof of Theorem 2.1.* By applying Taylor expansion to (3), we obtain

$$\widehat{\mathcal{L}}_w(\alpha(u)) = 2\sum_{i=1}^N \Big[\lambda^\tau\widehat{\mathcal{H}}_i - \frac{1}{2}(\lambda^\tau\widehat{\mathcal{H}}_i)^2\Big] + o_p(1).$$

By (4), it follows that

$$0 = \frac{1}{nh}\sum_{i=1}^n \frac{\widehat{\mathcal{H}}_i}{1+\lambda^\tau\widehat{\mathcal{H}}_i} = \frac{1}{nh}\sum_{i=1}^n \widehat{\mathcal{H}}_i - \frac{1}{nh}\sum_{i=1}^n \widehat{\mathcal{H}}_i\widehat{\mathcal{H}}_i^\tau\lambda + \frac{1}{nh}\sum_{i=1}^n \frac{\widehat{\mathcal{H}}_i(\lambda^\tau\widehat{\mathcal{H}}_i)^2}{1+\lambda^\tau\widehat{\mathcal{H}}_i}.$$

The application of Lemmas 3–5 yields that

$$\sum_{i=1}^n(\lambda^\tau\widehat{\mathcal{H}}_i)^2 = \sum_{i=1}^n \lambda^\tau\widehat{\mathcal{H}}_i + o_p(1), \qquad \lambda = \Big(\sum_{i=1}^n \widehat{\mathcal{H}}_i\widehat{\mathcal{H}}_i\Big)^{-1}\sum_{i=1}^n \widehat{\mathcal{H}}_i + o_p(n^{-\frac{1}{2}}).$$

Then, this together with (10) leads to

$$\begin{aligned}
\widehat{\mathcal{L}}_w(\alpha(u)) =& \Big(\frac{1}{\sqrt{nh}}\sum_{i=1}^n \widehat{\mathcal{H}}_i\Big)^\tau\Big(\frac{1}{nh}\sum_{i=1}^n \widehat{\mathcal{H}}_i\widehat{\mathcal{H}}_i^\tau\Big)^{-1}\Big(\frac{1}{\sqrt{nh}}\sum_{i=1}^n \widehat{\mathcal{H}}_i\Big) + o_p(1)\\
=& \widetilde{\mathbf{H}}_n^\tau \times \mathcal{M}_1(u) \times \widetilde{\mathbf{H}}_n + o_p(1),
\end{aligned}$$

where

$$\widetilde{\mathbf{H}}_n = (v(u)\Omega_1(u))^{-\frac{1}{2}} \frac{1}{\sqrt{nh}} \sum_{i=1}^{n} \widehat{\mathcal{H}}_i$$

and $\mathcal{M}_1 = (\Omega_1(u))^{\frac{1}{2}}(\Omega_2(u))^{-1}(\Omega_1(u))^{\frac{1}{2}}$.

Denote $\mathcal{D}_0 = \mathrm{diag}\{w_1, w_2, \cdots, w_p\}$, where $w_i \ \ (1 \le i \le p)$ denote the eigenvalues of matrix $\mathcal{M}_1$. Then there exists an orthogonal matrix $Q$ such that $Q^\tau \mathcal{M}_1 Q = \mathcal{D}_0$. Thus

$$\widehat{\mathcal{L}}_w(\alpha(u)) = (Q^\tau \times \widetilde{\mathbf{H}}_n)^\tau \times \mathcal{D}_0 \times (Q^\tau \times \widetilde{\mathbf{H}}_n) + o_p(1).$$

From Lemma 3, we have

$$Q^\tau (v(u)\Omega_1(u))^{-\frac{1}{2}} \frac{1}{\sqrt{nh}} \sum_{i=1}^{n} \widehat{\mathcal{H}}_i^\tau \longrightarrow N(0, I_p).$$

Theorem 2.1 now follows immediately. □

**Lemma 6.** it Under the same assumptions of Lemma 3, we have

$$\frac{1}{\sqrt{nh}} \sum_{i=1}^{n} \widehat{\mathcal{R}}_i \to^L N(0, v(u)\Omega_1(u)),$$

*Proof.* Recall that

$$\begin{aligned}
&\frac{1}{\sqrt{nh}} \sum_{i=1}^{n} \Big(1 - \frac{\delta_i}{\widehat{\pi}_i}\Big)(Y_i \widehat{g}_1(Z_i) - \widehat{g}_2(Z_i)\alpha(u))K_h(U_i - u) \\
=&\frac{1}{\sqrt{nh}} \sum_{i=1}^{n} \Big(1 - \frac{\delta_i}{\widehat{\pi}_i}\Big) \Phi_u(Z_i) K_h(U_i - u) \\
&+ \frac{1}{\sqrt{nh}} \sum_{i=1}^{n} \Big(1 - \frac{\delta_i}{\widehat{\pi}_i}\Big) Y_i(\widehat{g}_1(Z_i) - g_1(Z_i))K_h(U_i - u) \\
&+ \frac{1}{\sqrt{nh}} \sum_{i=1}^{n} \Big(1 - \frac{\delta_i}{\widehat{\pi}_i}\Big)(\widehat{g}_2(Z_i) - g_2(Z_i))\alpha(u)K_h(U_i - u) \\
=&C_1 + C_2 + C_3.
\end{aligned}$$

Similar to (6), we show that

$$\begin{aligned}
C_1 =&\frac{1}{\sqrt{nh}} \sum_{i=1}^{n} \Phi_u(Z_i)K_h(U_i - u) - \frac{1}{\sqrt{nh}} \sum_{i=1}^{n} \Big\{ \frac{\delta_i}{\pi_i} \Phi_u(Z_i) K_h(U_i - u) \\
&+ \Big(1 - \frac{\delta_i}{\pi_i}\Big) E(\Phi_u(Z)|Z_i)K_h(U_i - u)\Big\} + o_p(1) = o_p(1).
\end{aligned}$$

Next,

$$\begin{aligned}
C_2 =&\frac{1}{\sqrt{nh}} \sum_{i=1}^{n} \Big(1 - \frac{\delta_i}{\pi_i}\Big) Y_i(\widehat{g}_1(Z_i) - g_1(Z_i))K_h(U_i - u) \\
&+ \frac{1}{\sqrt{nh}} \sum_{i=1}^{n} \frac{\widehat{\pi}_i - \pi_i}{\pi_i \widehat{\pi}_i} \delta_i Y_i(\widehat{g}_1(Z_i) - g_1(Z_i))K_h(U_i - u) \\
=&C_{21} + C_{22}.
\end{aligned}$$

For term $C_{21}$, by the fact that $\sup_z |\widehat{g}_1(z) - g_1(z)| = o_p(n^{-\frac{1}{4}})$ (see [11]), a direct use of Lemmas 1 and 2 yields that

$$\begin{aligned}|C_{21}| \le & c(nh)^{-\frac{1}{2}} \cdot \sup_z |\widehat{g}_1(z) - g_1(z)| \cdot \max_{1\le k\le n} \Big| \sum_{j=1}^{k} \Big(1 - \frac{\delta_{j_i}}{\pi_{j_i}}\Big) Y_{j_i} \Big| \\ = & c(nh)^{-\frac{1}{2}} \cdot o_p(n^{-\frac{1}{4}}) \cdot n^{\frac{1}{2}} \log n \longrightarrow 0.\end{aligned}$$

Similar to (9), we have $C_{22} = o_p(1)$. This proves $C_2 = o_p(1)$. Similarly, we can show $C_3 = o_p(1)$.

Finally, we obtain

$$\frac{1}{\sqrt{nh}} \sum_{i=1}^{n} \widehat{\mathcal{R}}_i = \frac{1}{\sqrt{nh}} \sum_{i=1}^{n} \widehat{\mathcal{H}}_i + o_p(1).$$

By Lemma 3, this lemma follows immediately. □

**Lemma 7.** *Under the same assumptions of Lemma 3, we have*

$$\frac{1}{nh} \sum_{i=1}^{n} \widehat{\mathcal{R}}_i \widehat{\mathcal{R}}_i^\tau \longrightarrow^p v(u)\Omega_1(u).$$

*Proof.* By using similar arguments as that of Lemma 4 and (11), we show that

$$\begin{aligned}\frac{1}{nh} \sum_{i=1}^{n} \widehat{\mathcal{R}}_i \widehat{\mathcal{R}}_i^\tau = & \frac{1}{nh} \sum_{i=1}^{n} \Big\{ \frac{\delta_i}{\pi_i^2} X_i X_i^\tau \varepsilon_i^2 + 2\frac{\delta_i}{\pi_i}\Big(1 - \frac{\delta_i}{\pi_i}\Big) X_i \varepsilon_i E(X\varepsilon|Z_i) \\ & + \Big(1 - \frac{\delta_i}{\pi_i}\Big)^2 E(X\varepsilon|Z_i)^{\otimes 2} \Big\} K_h^2(U_i - u) + o_p(1).\end{aligned}$$

Then a simple derivation leads to this lemma. □

*Proof of Theorem 2.4.* Based on Lemmas 6 and 7, this theorem can be showed similar to Theorem 2.1. □

*Proof of Theorem 2.6.* Using the Taylor expansion to $\alpha(U_i) - \alpha(u)$ and $\widehat{\alpha}_r(U_i) - \widehat{\alpha}_r(u)$ at $u$, some simple calculations lead to

$$\alpha(U_i) - \alpha(u) - (\widehat{\alpha}_r(U_i) - \widehat{\alpha}_r(u)) = (\alpha'(u) - \widehat{\alpha}_r'(u))(U_i - u) + O_p(U_i - u)^2.$$

□

By Assumptions A1–A6, we show that

$$\alpha'(u) - \widehat{\alpha}_r'(u) \longrightarrow^p 0, \qquad \frac{1}{\sqrt{nh}} \sum_{i=1}^{n} \frac{\delta_i}{\pi_i} (U_i - u)^l K_h(U_i - u) = O_p(1)$$

and

$$\frac{1}{\sqrt{nh}} \sum_{i=1}^{n} \frac{\pi_i - \delta_i}{\pi_i} (U_i - u)^l K_h(U_i - u) = o_p(1), \qquad l = 1, 2.$$

Note that

$$\frac{1}{\sqrt{nh}} \sum_{i=1}^{n} \widehat{\mathcal{E}}_i = A_1 + C_1 + C_2 + C_3 + \Big(A_2 - \frac{1}{\sqrt{nh}} \sum_{i=1}^{n} \phi_i(u)\Big).$$

Then similar to Lemma A.8 of [29], we can prove that

$$A_2 - \frac{1}{\sqrt{nh}} \sum_{i=1}^{n} \phi_i(u) = o_p(1).$$

This, combined with Lemmas 3 and 6, proves that

$$\frac{1}{\sqrt{nh}} \sum_{i=1}^{n} \widehat{\mathcal{E}}_i = \frac{1}{\sqrt{nh}} \sum_{i=1}^{n} \widehat{\mathcal{H}}_i + o_p(1).$$

The rest of the proof follows a similar fashion as that of Theorems 2.1 and 2.4, so we omit it. □

*Proof of Theorem 2.7.* For simplicity, write $\widetilde{\alpha}_k^*(u) = (\widehat{\alpha}_r^1(u), \cdots, \widehat{\alpha}_r^{k-1}(u), \alpha_k(u), \widehat{\alpha}_r^{k+1}(u), \cdots, \widehat{\alpha}_r^p(u))^\tau$. Obviously,

$$\begin{aligned}\widehat{\mathcal{E}}_{i,k}(\alpha_k(u)) =& e_k^\tau \widehat{\mathcal{A}}_{2n}^{-1}(u)\Big(\widehat{\mathcal{E}}_i(\alpha(u)) - \Big(\frac{\delta_i}{\widehat{\pi}} X_i X_i^\tau + \Big(1 - \frac{\delta_i}{\widehat{\pi}}\Big)\widehat{g}_2(Z_i)\Big) \\ & \times K_h(U_i - u)(\widetilde{\alpha}_k^*(u) - \alpha(u))\Big) \\ =& e_k^\tau \widehat{\mathcal{A}}_{2n}^{-1}(u)\widehat{\mathcal{E}}_i(\alpha(u)) - \mathcal{W}_i.\end{aligned}$$

By Theorem 2.6, we have

$$e_k^\tau \widehat{\mathcal{A}}_{2n}^{-1}(u) \frac{1}{\sqrt{nh}} \sum_{i=1}^{n} \widehat{\mathcal{E}}_i(\alpha(u)) \to^L N(0, \sigma_u^2)$$

with $\sigma_u^2 = v^{-1}(u) e_k^\tau \Omega_1^{-1}(u) e_k$. □

Next,

$$\begin{aligned}\frac{1}{\sqrt{nh}} \sum_{i=1}^{n} \mathcal{W}_i =& e_k^\tau \widehat{\mathcal{A}}_{2n}^{-1}(u) \times \frac{1}{\sqrt{nh}} \sum_{i=1}^{n} \Big(\frac{\delta_i}{\widehat{\pi}} X_i X_i^\tau + \Big(1 - \frac{\delta_i}{\widehat{\pi}}\Big)\widehat{g}_2(Z_i)\Big) \\ & \times K_h(U_i - u)(\widetilde{\alpha}_k^*(u) - \alpha(u)) \\ =& e_k^\tau \widehat{\mathcal{A}}_{2n}^{-1}(u) \cdot \widehat{\mathcal{A}}_{2n}(u) \cdot \sqrt{nh} \cdot (\widetilde{\alpha}_k^*(u) - \alpha(u)) = 0.\end{aligned}$$

Then we obtain that

$$\frac{1}{\sqrt{nh}} \sum_{i=1}^{n} \widehat{\mathcal{E}}_{i,k}(\alpha_k(u)) \longrightarrow^L N(0, \sigma_u^2).$$

Next, as Lemma A.2 of [3], a tedious but technically not very challenging derivation yields that

$$\|\widehat{\alpha}_r(u) - \alpha(u)\| = O_p\big(\log(1/h)/nh)^{1/2} + h^2\big) + O_p\big((nb^2)^{-1/2} + b^k\big).$$

Using this, it is easy to show that

$$\frac{1}{nh} \sum_{i=1}^{n} \widehat{\mathcal{E}}_{i,k}(\alpha_k(u)) \widehat{\mathcal{E}}_{i,k}^\tau(\alpha_k(u)) \to^p \sigma_u^2.$$

The rest of the proof is similar to that of Theorem 2.6. □

**Acknowledgements.** This work was partially done during the author's visit to the Department of Statistics and Probability in Michigan State University as a joint Ph D student. He thanks his co-advisor, Professor Yi-jun Zuo, for great help and the department for good studying conditions. The author thanks the editor, an associate editor and one anonymous reviewer, whose helpful comments greatly improved the presentation of this paper.

## References

[1] Cai, Z.W., Fan, J.Q., Li, R.Z. Efficient estimation and inferences for varying-coefficient models. *Journal of the American Statistical Association*, 95: 888–902 (2000)
[2] Chen S.X., Keilegom, I.V. A review on empirical likelihood methods for regression. Test 18, 415–447 (1990)
[3] Fan, J.Q., Huang, T. Profile likelihood inferences on semiparametric varying-coefficient partially linear models. *Bernoulli*, 11(6): 1031–1057 (2005)
[4] Fan, J.Q., Zhang, W. Statistical estimation in varying coefficient models. *The Annals of Statistics*, 27: 1491–1518 (1999)
[5] Fan, J.Q., Zhang, W. Statistical methods with varying coefficient models. *Statistics and Its Interface*, 1: 179–195 (2008)
[6] Gao, J.T. Asymptotic theory for partly linear models. *Communications in Statistics: Theory and Methods*, 24: 1985–2010 (1995)
[7] Hastie, T.J., Tibshirani. R.J. Varying-coefficient models (with discussion). *Journal of the Royal Statistical Society* (Serial B), 55: 757–796 (1993)
[8] Horvitz, D.G., Thompson, D.J. A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47: 663–685 (1952)
[9] Liang, H. Generalized partially linear models with missing covariates. *Journal of Multivariate Analysis*, 99, 880–895 (2008)
[10] Liang, H., Qin, Y.S. Empirical likelihood-based inference for partially linear models with missing covariates. *Australian & New Zealand Journal of Statistics*, 50(4): 347–359 (2008)
[11] Liang, H., Wang, S.J., Robbins, J.M., Carroll, R.J. Estimation in partially linear models with missing covariates. *Journal of the American Statistical Association*, 99: 357–367 (2004)
[12] Little, R.J.A., Rubin, D.B. Statistical analysis with missing data, 2nd ed. Wiley, Hoboken, NJ, 2002
[13] Mack, Y., Silverman, B. Weak and strong uniform consistency of kernel regression estimates. *Probability Theory and Related Fields*, 61(3): 405–415 (1982)
[14] Owen, A.B. Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, 75: 237–249 (1988)
[15] Owen, A.B. Empirical likelihood ratio confidence regions. *The Annals of Statistics*, 18: 90–120 (1990)
[16] Owen, A.B. Empirical likelihood. Chapman & Hall, London, 2001
[17] Qin, J., Lawless, J. Empirical likelihood and general estimating equations. *The Annals of Statitics*, 22: 300–325 (1994)
[18] Robins, J.M., Rotnitzky, A., Zhao, L.P. Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association*, 89: 846–866 (1994)
[19] Tsiatis, A.A. Semiparametric Theory and Missing Data. Springer, 2006
[20] Wang, D., Chen, S.X. Empirical likelihood for estimating equations with missing values. *The Annals of Statitics*, 37: 490–517 (2009)
[21] Wang, H.S., Xia, Y.C. Shrinkage estimation of the varying coefficient model. *Journal of the American Statistical Association*, 104: 747–757 (2009)
[22] Wang, Q.H. Statistical estimation in partial linear models with covariate data missing at random. *Annals of the Institute of Statistical Mathematics*, 61: 47–84 (2009)
[23] Wang, Q.H., Rao, J.N.K. Empirical likelihood-based inference in linear errors-in-covariables models with validation data. *Biometrika*, 89(2): 345–358 (2002)
[24] Wang, Q.H., Rao, J.N.K. Empirical likelihood-based inference under imputation for missing response data. *The Annals of Statitics*, 30: 896–924 (2002)
[25] Wang, Q.H., Qin, Y.S. Empirical likelihood confidence bands for distribution functions with missing responses. *Journal of Statistical Planning and Inference*, 140: 2778–2789 (2010)
[26] Wong, H., Guo, S.J., Chen, M., IP, W.C. On locally weighted estimation and hypothesis testing of varying-coefficient models with missing covariates. *Journal of Statistical Planning and Inference*, 139: 2933–2951 (2009)
[27] Wu, C.O., Chiang, C.T., Hoover, D.R. Asymptotic confidence regions for kernel smoothing of a varying-coefficient model with longitudinal data. *Journal of the American Statistical Association*, 93: 1388–1402 (1998)
[28] Xue, L.G., Zhu, L.X. Empirical likelihood for a varying coefficient model with longitudinal data. *Journal of the American Statistical Association*, 102: 642–654 (2007)
[29] Zhao, P.X., Xue, L.G. Empirical likelihood inferences for semiparametric varying-coefficient partially linear errors-in-variables models with longitudinal data. *Journal of Nonparametric Statistics*, 21(7): 907–923 (2009)