Open Mathematics

Research Article

Xing Xiang* and Wanrong Liu

Empirical likelihood confidence regions of the parameters in a partially single-index varying-coefficient model

https://doi.org/10.1515/math-2019-0059
Received February 9, 2018; accepted May 21, 2019

Abstract: In this paper, we investigate a partially single-index varying-coefficient model, and suggest two empirical log-likelihood ratio statistics for the unknown parameters in the model. The first statistic is asymptotically distributed as a weighted sum of independent chi-square variables under some mild conditions. It is proved that another statistic, with adjustment factor, is asymptotically standard chi-square under some suitable conditions. These useful statistics could be used to construct the confidence regions of the parameters. A simulation study indicates that, with the increase of sample size, the coverage probability of the confidence region constructed by us gradually approaches the theoretical value.

Keywords: confidence region, empirical likelihood, partially single-index varying-coefficient model, chi-square distribution

MSC: 62-xx

1 Introduction

Consider a partially single-index varying-coefficient model of the form

$$Y = g_0^\tau(\beta_0^\tau U)X + \theta_0^\tau Z + \varepsilon, \tag{1.1}$$

where $(U, X, Z) \in R^p \times R^d \times R^q$ is a vector of covariates, Y is the response variable, β_0 is a $p \times 1$ vector of unknown parameters, θ_0 is a $q \times 1$ vector of regression coefficient, $g_0(\cdot)$ is a $d \times 1$ vector of unknown functions and ε is a random error with $E(\varepsilon \mid U, X, Z) = 0$ and $Var(\varepsilon \mid U, X, Z) = \sigma^2$. Assume that ε and (U, X, Z) are independent. In order to make sure the identifiability, it is often assumed that $\|\beta_0\| = 1$, where $\|\cdot\|$ denotes the Euclidean metric.

Feng and Xue[1] considered the problem of model detection and estimation for single-index varying-coefficient model, they identified the true model structure and obtained a new semiparametric model, which is partially linear single-index varying-coefficient model. As we all know, the optimal parametric estimation rate is $n^{-1/2}$ and the optimal nonparametric estimation rate is $n^{-2/5}$, if we treat a parametric component as a nonparametric component, the problems of data overfitting and efficiency loss will occur.

Model (1.1), may include crossproduct terms of some components of X and Z, is easily interpreted in real applications because it has the features of the partial linear model, the single-index model and the varying-coefficient model, which make the model more general. Model (1.1) takes many other regression models as special cases, such as linear model, partial linear model, varying-coefficient model, single-index model,

***Corresponding Author: Xing Xiang:** Hunan Normal University Changsha, China, E-mail: chichu0219@hnu.edu.cn
Wanrong Liu: Hunan Normal University Changsha, China, E-mail: liuwr08@163.com

partial linear single-index model, single-index varying-coefficient model, etc. The linear component $\theta_0^\tau Z$ provides a simple summary of covariates effects which are of the main scientific interest. The index $\beta_0^\tau U$ enables us to simplify the treatment of the multiple auxiliary variables, and the functions $g_0(\cdot)$s enrich model flexibility. It is well known that in order to construct the confidence region of (β_0, θ_0) by using the normal approximation method, it is necessary to construct the embedded estimation of the asymptotic variance of the corresponding estimator, which includes the estimation of parametric and non-parametric components. The empirical likelihood method avoids this shortcoming and its structure does not include estimation of parameter (β_0, θ_0). In this paper, we can construct an empirical likelihood ratio function for (β_0, θ_0) by assuming $g_0(\cdot)$ and its derivative $\dot{g}_0(\cdot)$ to be known functions.

As far as we know, there is not much literature on this model by using empirical likelihood method, although it has been applied to varieties of models. In this paper, we consider the problem of a method of constructing confidence regions for (β_0, θ_0), since the empirical likelihood method, which is introduced by Owen[2, 3], has many advantages. For example, it does not require the construction of a pivotal quantity, and it does not impose a priori constraint on the shape of the region. Owen[4] proved the empirical log-likelihood ratio is asymptotically a standard chi-square variable when he applied the empirical likelihood to linear regression model, so that it can be applied to constructing the confidence region of the regression parameter. There are studies related to empirical likelihood, such as Wang and Rao[5], Wang, Linton and Härdle[6], Xue and Zhu[7–9], Zhu and Xue[10], You and Zhou[11], Qin and Zhang[12], Stute, Xue and Zhang[13], Xue[14, 15],Wang et al.[16], Huang and Zhang[17], Wang and Xue[18], Lian[19], Xiao[20], Zhou, Zhao and Wang[21], Fang, Liu and Lu[22], Arteaga-Molina and Rodriguez-Poo[23], among others.

The rest of this article is organized as follows. In Section 2, we give an estimated empirical likelihood ratio, and investigate the asymptotic properties of the proposed estimators. In Section 3, we give an adjusted empirical log-likelihood and derive its asymptotic distribution. Section 4 reports a simulation study. Proofs of theorems and lemmas are postponed in Appendix A and Appendix B, respectively.

2 Main results

2.1 Methodology

Suppose that $\{(Y_i, U_i, X_i, Z_i); 1 \le i \le n\}$ is an i.i.d sample from model (1.1), that is

$$Y_i = g_0^\tau(\beta_0^\tau U_i)X_i + \theta_0^\tau Z_i + \varepsilon_i, \qquad i = 1, \cdots, n, \tag{2.1}$$

where ε_is are i.i.d. random errors with mean 0 and finite variance σ^2. Assume that ε_is are independent of $\{U_i, X_i, Z_i; 1 \le i \le n\}$.

Our primary interest is to construct the confidence region of (β_0, θ_0). In order to construct an empirical likelihood ratio function for (β_0, θ_0), we introduce an auxiliary random vector

$$\eta_i(\beta, \theta) = [Y_i - g_0^\tau(\beta^\tau U_i)X_i - \theta^\tau Z_i]w(\beta^\tau U_i)(\dot{g}_0^\tau(\beta^\tau U_i)X_i U_i^\tau, Z_i^\tau)^\tau. \tag{2.2}$$

where $\dot{g}_0(\cdot)$ denotes the derivative of the function vector $g_0(\cdot)$, and $w(\cdot)$ is a bounded weight function with a bounded support $\mathcal{T}_w$. In order to control the boundary effect in the estimations of $g_0(\cdot)$ and $\dot{g}_0(\cdot)$, it is necessary to introduce this function here. To convenience, we take $w(\cdot)$ the indicator function of the set $\mathcal{T}_w$. Hence, the problem of testing whether (β, θ) is the true parameter is equivalent to testing whether $E[\eta_i(\beta, \theta)] = 0$ if (β, θ) is the true parameter. By Owen[2], this can be done by using the empirical likelihood. That is, we can define the profile empirical likelihood ratio function as follows

$$L_n(\beta, \theta) = max\{\prod_{i=1}^{n}(np_i) \mid p_i \ge 0, \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i\eta_i(\beta, \theta) = 0\}. \tag{2.3}$$

It can be shown that $-2\log L_n(\beta_0, \theta_0)$ is asymptotically chi-squared with $p + q$ degrees of freedom. However, $L_n(\beta, \theta)$ cannot directly be used to make statistical inference of (β_0, θ_0) because $L_n(\beta, \theta)$ contains the unknowns $g(\cdot)$ and $\dot{g}(\cdot)$. A common approach is to replace $g(\cdot)$ and $\dot{g}(\cdot)$ in $L_n(\beta,\theta)$ by their estimators and define an estimated empirical likelihood function.

When (β,θ) is known, model (1.1) can be treated as a varying-coefficient partially linear regression model. Then, we can use the methodology of profile least square to estimate $g_0(\cdot)$ and $\dot{g}_0(\cdot)$. We estimate the vector functions $g_0(\cdot)$ and $\dot{g}_0(\cdot)$ via the local linear regression technique. The local linear estimators for $g_0(t)$ and $\dot{g}_0(t)$ are defined as $\tilde{g}(t;\beta_0,\theta_0)=\tilde{a}$ and $\tilde{\dot{g}}(t;\beta_0,\theta_0)=\tilde{b}$ at fixed point (β_0,θ_0), where $\tilde{a}$ and $\tilde{b}$ minimize the sum of weighted squares

$$\sum_{i=1}^{n}[Y_i-\theta_0^\tau Z_i-\{a+b(\beta_0^\tau U_i-t)\}^\tau X_i]^2K_h(\beta_0^\tau U_i-t), \tag{2.4}$$

where $K_h(\cdot)=h^{-1}K(\cdot/h)$, $K(\cdot)$ is a kernel function, and $h=h_n$ is a bandwidth sequence that decreases to 0 as n increases to ∞. Simple calculation yields

$$(\tilde{g}(t;\beta_0),h\tilde{\dot{g}}(t;\beta_0))^\tau=(\xi_{0,\beta_0}^\tau W_{0,\beta_0}\xi_{0,\beta_0})^{-1}\xi_{0,\beta_0}^\tau W_{0,\beta_0}(Y-Z\theta_0), \tag{2.5}$$

where

$$\xi_{0,\beta_0}=\begin{pmatrix} X_1^\tau & X_1^\tau(\beta_0^\tau U_1-t)/h\\ \vdots & \vdots\\ X_n^\tau & X_n^\tau(\beta_0^\tau U_n-t)/h\end{pmatrix},\ W_{0,\beta_0}=diag\{K_h(\beta_0^\tau U_1-t),\cdots,K_h(\beta_0^\tau U_n-t)\},$$

$Y=(Y_1,\cdots,Y_n)^\tau$ and $Z=(Z_1,\cdots,Z_n)^\tau$.

Let $\tilde{g}(t;\beta,h)$ and $\tilde{\dot{g}}(t;\beta,h)$ denote the estimators of $g(t)$ and $\dot{g}(t)$ with the bandwidths h and $h_1=h_{1n}$ respectively. Therefore, let $\hat{\eta}_i(\beta,\theta)$ denote $\eta_i(\beta,\theta)$ with $g(\beta^\tau U_i)$ and $\dot{g}(\beta^\tau U_i)$ replaced by $\tilde{g}(\beta^\tau U_i;\beta)$ and $\tilde{\dot{g}}(\beta^\tau U_i;\beta)$ respectively for $i=1,\cdots,n$. Then an estimated empirical log-likelihood ratio function is defined as

$$\hat{l}(\beta,\theta)=-2max\{\sum_{i=1}^{n}\log(np_i)\mid p_i\geq 0,\sum_{i=1}^{n}p_i=1,\sum_{i=1}^{n}p_i\hat{\eta}_i(\beta,\theta)=0\}. \tag{2.6}$$

By the Lagrange multiplier method, $-2\log\hat{L}(\beta,\theta)$ can be represented as

$$\hat{l}(\beta,\theta)=2\sum_{i=1}^{n}\log(1+\lambda^\tau\hat{\eta}_i(\beta,\theta)), \tag{2.7}$$

where $\lambda=\lambda(\beta,\theta)$ is determined by

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\hat{\eta}_i(\beta,\theta)}{1+\lambda^\tau\hat{\eta}_i(\beta,\theta)}=0. \tag{2.8}$$

Firstly, we write $G_\beta=(g_0^\tau(\beta^\tau U_1)X_1,\cdots,g_0^\tau(\beta^\tau U_n)X_n)^\tau$. Hence, we can derive a estimator of G_β by (2.5), which is

$$\tilde{G}_{\beta_0}=\begin{pmatrix}[X_1^\tau,0_d](\xi_{1,\beta_0}^\tau W_{1,\beta_0}\xi_{1,\beta_0})^{-1}\xi_{1,\beta_0}^\tau W_{1,\beta_0}\\ \vdots\\ [X_n^\tau,0_d](\xi_{n,\beta_0}^\tau W_{1,\beta_0}\xi_{n,\beta_0})^{-1}\xi_{n,\beta_0}^\tau W_{n,\beta_0}\end{pmatrix}(Y-Z\theta_0) \tag{2.9}$$

$$\triangleq S_{\beta_0}(Y-Z\theta_0) \tag{2.10}$$

here 0_d denotes the d-dimensional zero vector, let ξ_{i,β_0} and W_{i,β_0} denote ξ_{0,β_0} and W_{0,β_0} with t replaced by $\beta_0^\tau U_i$ for $i=1,\cdots,n$. Hence we get an approximate linear model as follows

$$(I_n-S_{\beta_0})Y=(I_n-S_{\beta_0})Z\theta_0+\varepsilon, \tag{2.11}$$

here I_n denotes the nth identity matrix.

Secondly, we can use the least square theory to obtain

$$\breve{\theta} = \{Z^\tau (I_n - S_{\beta_0})^\tau (I_n - S_{\beta_0}) Z\}^{-1} Z^\tau (I_n - S_{\beta_0})^\tau (I_n - S_{\beta_0}) Y, \tag{2.12}$$

we can get the estimators of $g(\cdot)$ and $\dot{g}(\cdot)$ at $t = \beta_0^\tau u$ by substituting (2.12) into (2.5) as follows

$$(\breve{g}(t;\beta_0), h\breve{\dot{g}}(t;\beta_0))^\tau = (\xi_{0,\beta_0}^\tau W_{0,\beta_0} \xi_{0,\beta_0})^{-1} \xi_{0,\beta_0}^\tau W_{0,\beta_0} (Y - Z\breve{\theta}). \tag{2.13}$$

Thirdly, noticed that (2.13) and (2.12) are based on a known β_0. Under the condition $\|\beta\| = 1$, we have a estimator $\hat{\beta}$ of β_0 by minimize the following equation that

$$N(\beta) = \sum_{i=1}^{n} \{Y_i - \breve{g}^\tau(\beta^\tau U_i) X_i - \breve{\theta}^\tau Z_i\}^2. \tag{2.14}$$

Obviously, solving (2.14) is equivalent to solving the following equation under the condition $\|\beta\| = 1$ that

$$\sum_{i=1}^{n} \{Y_i - \breve{g}^\tau(\beta^\tau U_i) X_i - \breve{\theta}^\tau Z_i\} \breve{\dot{g}}^\tau(\beta^\tau U_i) X_i U_i w(\beta^\tau U_i) = 0. \tag{2.15}$$

Finally, we get the final estimators $\hat{\theta}$, $\hat{g}(\cdot)$ and $\hat{\dot{g}}(\cdot)$ with β_0 replaced by $\hat{\beta}$ in $\breve{\theta}$, $\breve{g}(\cdot)$ and $\breve{\dot{g}}(\cdot)$, respectively.

2.2 Asymptotic properties

Let $\mathcal{B}_n = \{\beta \in \mathcal{B} : \|\beta - \beta_0\| \le c_0 n^{-1/2}\}$ and $\Theta_n = \{\theta \in \Theta : \|\theta - \theta_0\| \le c_0 n^{-1/2}\}$ for some positive constant c_0. To obtain the asymptotic distribution of $-2\log \hat{L}(\beta_0, \theta_0)$, we give a set of conditions first. These conditions are not very hard to satisfy, similar restrictions were also made by Härdle, Hall and Ichimura[24], Xia and Li[25], Wang and Xue[18], Xue and Pang[26].

(C1) The density function $f(t)$ of $\beta^\tau U$ is bounded away from zero for $t \in \mathcal{T}_w$ and β near β_0, and satisfies the Lipschitz condition of order 1 on $\mathcal{T}_w$, where $\mathcal{T}_w$ is the support of $w(t)$.
(C2) The functions $g_j(t)$, $1 \le j \le q$, have continuous second derivatives on $\mathcal{T}_w$, where $g_j(t)$ are the jth components of $g_0(t)$.
(C3) $E(\|U\|^6) \le \infty$, $E(\|X\|^6) \le \infty$,$E(\|Z\|^6) \le \infty$, $E(|\varepsilon|^6) \le \infty$.
(C4) The bandwidths satisfy that $h \to 0$, $nh^2/\log^2 n \to \infty$, $nh^4 \log n = o_P(1)$, $nh^8 = o_P(1)$, $nhh_1^3/\log^2 n \to \infty$, $h_1 \sim n^{-1/5}$.
(C5) The kernel $K(\cdot)$ is a symmetric probability density function with a bounded support and satisfies the Lipschitz condition of order 1 and $\int t^2 K(t)dt \neq 0$.
(C6) The matrix $D(t) = E(XX^\tau | \beta_0^\tau U = t)$ is positive definite, and each entry of $D(t)$, $D^{-1}(t)$ and $C(t) = E(XZ^\tau | \beta_0^\tau U = t)$ satisfies the Lipschitz condition of order 1 on $\mathcal{T}_w$, where $\Lambda = w(\beta_0^\tau U)(\dot{g}_0^\tau(\beta_0^\tau U) X U^\tau, Z^\tau)^\tau$.
(C7) The matrix $B(\beta_0, \theta_0) = E(\Lambda \Lambda^\tau)$ is positive definite.

The following theorem shows that $-2\log \hat{L}(\beta_0, \theta_0)$ is asymptotically distributed as a weighted sum of independent χ_1^2 variables.

Theorem 2.1. *Suppose conditions (C1)-(C7) hold. If (β_0, θ_0) is the true value of the parameter, then*

$$-2\log \hat{L}(\beta_0, \theta_0) \xrightarrow{D} w_1 \chi_{1,1}^2 + \cdots + w_{p+q} \chi_{1,p+q}^2,$$

where $\xrightarrow{D}$ denotes the convergence in distribution, the weights w_j $(1 \le j \le p+q)$ are the eigenvalues of $V(\beta_0, \theta_0) = B^{-1}(\beta_0, \theta_0) A(\beta_0, \theta_0)$, and $\{\chi_{1,i}^2 1 \le i \le p+q\}$ are the independent χ_1^2 variables. Where

$$A(\beta_0, \theta_0) = B(\beta_0, \theta_0) - E[C(\beta_0^\tau U) D^{-1}(\beta_0^\tau U) C^\tau(\beta_0^\tau U)]. \tag{2.16}$$

In order to apply Theorem(2.1) to construct a confidence region for (β_0, θ_0), we have to estimate the unknown weights w_is consistently. By the plug-in method, $A(\beta_0, \theta_0)$ and $B(\beta_0, \theta_0)$ can be estimated consistently by

$$\hat{A}(\hat{\beta}, \hat{\theta}) = \frac{1}{n} \sum_{i=1}^{n} [\hat{\Lambda}_i \hat{\Lambda}_i^\tau - \hat{C}(\hat{\beta}^\tau U_i) \hat{D}^{-1}(\hat{\beta}^\tau U_i) \hat{C}^\tau(\hat{\beta}^\tau U_i)], \tag{2.17}$$

and

$$\hat{B}(\hat{\beta}, \hat{\theta}) = \frac{1}{n} \sum_{i=1}^{n} \hat{\Lambda}_i \hat{\Lambda}_i^\tau, \tag{2.18}$$

respectively, where $(\hat{\beta}, \hat{\theta})$ is defined in the above, and $\hat{\Lambda}_i = w(\hat{\beta}^\tau U_i)(\hat{g}'^\tau(\hat{\beta}^\tau U_i) X_i U_i^\tau, Z_i^\tau)^\tau$, $\hat{C}(\cdot) = \sum_{i=1}^n W_{ni}(\cdot) \hat{\Lambda}_i X_i^\tau$ and $\hat{D}(\cdot) = \sum_{i=1}^n W_{ni}(\cdot) X_i X_i^\tau$ with

$$W_{ni}(\cdot) = K_1(\tfrac{\hat{\beta}^\tau U_i - \cdot}{b_n}) / \textstyle\sum_{k=1}^n K_1(\tfrac{\hat{\beta}^\tau U_k - \cdot}{b_n}),$$

where $K_1(\cdot)$ is a kernel function, and b_n is a bandwidth with $0 < b_n \to 0$.

This means that $\hat{w}_j$, the eigenvalues of $\hat{V}(\hat{\beta}, \hat{\theta}) = \hat{B}^{-1}(\hat{\beta}, \hat{\theta}) \hat{A}(\hat{\beta}, \hat{\theta})$, consistently estimate w_j for $j = 1, \cdots, p+q$. Let $\hat{c}_{1-\alpha}$ be the $1-\alpha$ quantile of the conditional distribution of the weighted sum $\hat{s} = \hat{w}_1 \chi_{1,1}^2 + \cdots + \hat{w}_{p+q} \chi_{1,p+q}^2$ given the data. Then we can define an approximate $1 - \alpha$ confidence region for (β_0, θ_0) as

$$\mathcal{R}_{eel}(\alpha) = \{(\beta, \theta) \in \mathcal{B} \times \Theta : -2 \log \hat{L}(\beta, \theta) \le \hat{c}_{1-\alpha}\}.$$

In practice, we can calculate the conditional distribution of the weighted sum $\hat{s}$, given the sample $\{(Y_i, U_i, X_i, Z_i), 1 \le i \le n\}$ by using Monte Carlo simulations, by repeatedly generating independent samples $\chi_{1,1}^2, \cdots, \chi_{1,p+q}^2$ from the χ_1^2 distribution.

3 Adjusted empirical likelihood

When we use Theorem (2.1) to construct confidence regions of (β, θ), the weights w_is need to be estimated, the accuracy of confidence region is decreased. Let $\rho(\beta_0, \theta_0) = (p + q)/tr\{V_0(\beta_0, \theta_0)\}$, where $V_0(\beta_0, \theta_0)$ is defined in Theorem (2.1). By Rao and Scott[27], the distribution of $\rho(\beta_0, \theta_0) \sum_{i=1}^{p+q} w_i \chi_{1,i}^2$ can be approximated by χ_{p+q}^2, which is a standard chi-square distribution with $p + q$ degrees of freedom. Therefore, an improved Rao-Scott adjusted empirical log-likelihood ratio can be defined as

$$\hat{l}(\beta, \theta) = \hat{\rho}(\beta, \theta)\{-2 \log \hat{L}(\beta, \theta)\}. \tag{3.1}$$

However, the accuracy of this approximation depends on the w_is. Xue and Wang[28] proposed another adjusted empirical log-likelihood. By using an approximate result in the above, the adjustment technique is developed by Wang and Rao[5]. Note that $\hat{\rho}(\beta, \theta)$ can be written as

$$\hat{\rho}(\beta, \theta) = \frac{tr\{\hat{A}^-(\beta, \theta) \hat{A}(\beta, \theta)\}}{tr\{\hat{B}^{-1}(\beta, \theta) \hat{A}(\beta, \theta)\}},$$

where A^- represents a generalized inverse of matrix A. By examining the asymptotic expansion of $-2 \log \hat{L}(\beta, \theta)$, similar to Xue and Wang[28], we can define an adjustment factor

$$\hat{r}(\beta, \theta) = \frac{tr\{\hat{A}^-(\beta, \theta) \hat{\Sigma}(\beta, \theta)\}}{tr\{\hat{B}^{-1}(\beta, \theta) \hat{\Sigma}(\beta, \theta)\}},$$

where $\hat{\Sigma}(\beta, \theta) = \{\sum_{i=1}^n \hat{\eta}_i(\beta, \theta)\}\{\sum_{i=1}^n \hat{\eta}_i(\beta, \theta)\}^\tau$. The adjusted empirical log-likelihood ratio is defined by

$$\hat{l}(\beta, \theta) = \hat{r}(\beta, \theta)\{-2 \log \hat{L}(\beta, \theta)\}. \tag{3.2}$$

The following theorem shows that the adjusted empirical log-likelihood ratio is asymptotically distributed as standard chi-square.

Theorem 3.1. *Suppose that conditions (C1)-(C7) hold. Then*

$$\hat{l}(\beta_0, \theta_0) \xrightarrow{D} \chi^2_{p+q}.$$

Invoking Theorem (3.1), $\hat{l}(\beta, \theta)$ can be used to construct an approximate confidence region for (β_0, θ_0). Thereby, we can obtain the confidence region of (β, θ)

$$\mathcal{R}_{ael}(\alpha) = \{(\beta, \theta) \in \mathcal{B} \times \Theta : \hat{l}(\beta, \theta) \le c_{1-\alpha}\},$$

where $P(\chi^2_{p+q} \le c_{1-\alpha}) = 1 - \alpha$.

To apply Theorem (3.1) to construct a confidence region for (β_0, θ_0), we only need to estimate the adjustment factor $\hat{r}(\beta, \theta)$ by replacing (β, θ) by $(\hat{\beta}, \hat{\theta})$. The value of $-2 \log \hat{L}(\beta, \theta)$ is not depend on the estimation of (β, θ). In practice, we can calculate the numerical value of $-2 \log \hat{L}(\beta, \theta)$ by using the package in R(see 'emplik', http://cran.r-project.org/web/packages/emplik/).

4 Simulation study

Consider the regression model

$$Y_i = g_1(\beta_0^\tau U_i)X_{i1} + g_2(\beta_0^\tau U_i)X_{i2} + \theta_0^\tau Z_{i1} + \theta_0^\tau Z_{i2} + \varepsilon_i, \quad i = 1, \cdots, n, \tag{4.1}$$

where $\beta_0 = (\sqrt{2}/2, \sqrt{2}/2)^\tau$, $\theta_0 = (2, 1.2)^\tau$, $\varepsilon_i \sim N(0, 0.3^2)$. The sample $\{U_i = (U_{i1}, U_{i2})^\tau; 1 \le i \le n\}$ was generated from a bivariate uniform distribution $[-1, 1]^2$ with independent components, X_{i1}, X_{i2} were generated from a normal distribution $N(0, 0.8^2)$ and $\{Z_i = (Z_{i1}, Z_{i2})^\tau; 1 \le i \le n\}$ was generated from a bivariate normal distribution $N(0, \Sigma)$ with $Cov(Z_{ik}, Z_{il}) = 2 \times 0.5^{|k-l|}$, $k, l = 1, 2$. In model (4.1), the coefficient functions are $g_1(t) = sin(t - 2)$ and $g_2(t) = 1 - t^2$.

We used a Epanechnikov kernel $K(t) = 0.75(1 - t^2)_+$ and took the weight function $w(t) = I_{[-\sqrt{2}, \sqrt{2}]}(t)$. The bandwidths $\hat{h} = \hat{h}_1 n^{-1/25} (\log n)^{-1/2}$ and $\hat{h}_1 = \hat{h}_{opt}$ respectively, where $\hat{h}_{opt}$ was an optimal bandwidth by using the generalized cross validation(GCV). It's not hard to see that the bandwidth $\hat{h}$ satisfies condition (C4).

Table 1 The coverage probabilities of the confidence regions on (β, θ) when the nominal level is 0.95

n	Resampling times	Coverage probabilities
90	500	0.924
120	500	0.934
150	500	0.942
170	500	0.940
190	500	0.950

From Table 1, we can see that the probability of coverage increases with n until it approaches the theoretical value 0.95.

5 Appendices

5.1 Appendix A: Proofs of theorems

Proof of Theorem 2.1. Applying the Taylor expansion to (2.7) and utilizing Lemma (5.4), we obtain that

$$-2 \log \hat{L}(\beta_0, \theta_0) = -\sum_{i=1}^{n} \{\lambda^\tau \hat{\eta}_i(\beta_0, \theta_0) - \frac{1}{2}[\lambda^\tau \hat{\eta}_i(\beta_0, \theta_0)]^2\} + o_P(1), \tag{T.1}$$

Employing (2.8), we have

$$\begin{aligned}0 &= \frac{1}{n}\sum_{i=1}^{n}\frac{\hat{\eta}_i(\beta_0,\theta_0)}{1+\lambda^\tau\hat{\eta}_i(\beta_0,\theta_0)}\\ &= \frac{1}{n}\sum_{i=1}^{n}\hat{\eta}_i(\beta_0,\theta_0) - \frac{1}{n}\sum_{i=1}^{n}\hat{\eta}_i(\beta_0,\theta_0)\hat{\eta}_i^\tau(\beta_0,\theta_0)\lambda\\ &\quad + \frac{1}{n}\sum_{i=1}^{n}\frac{\hat{\eta}_i(\beta_0,\theta_0)[\lambda^\tau\hat{\eta}_i(\beta_0,\theta_0)]^2}{1+\lambda^\tau\hat{\eta}_i(\beta_0,\theta_0)}.\end{aligned}$$

The application of Lemma (5.4) yields that

$$\sum_{i=1}^{n}[\lambda^\tau\hat{\eta}_i(\beta_0,\theta_0)]^2 = \sum_{i=1}^{n}\lambda^\tau\hat{\eta}_i(\beta_0,\theta_0) + o_P(1),$$

and

$$\lambda = [\sum_{i=1}^{n}\hat{\eta}_i(\beta_0,\theta_0)\hat{\eta}_i^\tau(\beta_0,\theta_0)]^{-1}\sum_{i=1}^{n}\hat{\eta}_i(\beta_0,\theta_0) + o_P(n^{-1/2}).$$

This together with (T.1) proves that

$$-2\log\hat{L}(\beta_0,\theta_0) = nQ_n^\tau(\breve{g},\beta_0,\theta_0)R_n^{-1}(\beta_0,\theta_0)Q_n(\breve{g},\beta_0,\theta_0) + o_P(1). \tag{T.2}$$

Let $\hat{l}(\beta_0,\theta_0) = -2\log\hat{L}(\beta_0,\theta_0)$, and from (A.4.2) we obtain

$$\hat{l}(\beta_0,\theta_0) = [(\sigma^2A)^{-1/2}\sqrt{n}Q_n(\breve{g},\beta_0,\theta_0)]^\tau V(\beta_0,\theta_0)[(\sigma^2A)^{-1/2}\sqrt{n}Q_n(\breve{g},\beta_0,\theta_0)] + o_P(1), \tag{T.3}$$

where $V(\beta_0,\theta_0) = A^{1/2}(\beta_0,\theta_0)B^{-1}(\beta_0,\theta_0)A^{1/2}(\beta_0,\theta_0)$ is defined in Theorem(2.1). Let $\tilde{V} = diag(w_1,\cdots,w_{p+q})$, where w_j, $1 \le i \le p+q$, are the eigenvalues of $V(\beta_0,\theta_0)$. Then there exists an orthogonal matrix H such that $H^\tau\tilde{V}H = V(\beta_0,\theta_0)$. Thus,we have

$$Q_n(\breve{g},\beta,\theta) = J_1(\breve{g},\beta,\theta) + J_2(\breve{g},\beta) + J_3(\breve{g},\beta,\theta) + J_4(\breve{g},\beta_0,\theta_0) + Q(g_0,\beta,\theta). \tag{T.4}$$

From (T.4) and Lemma (5.3), we have

$$Q_n(\breve{g},\beta_0,\theta_0) = J_4(\breve{g},\beta_0,\theta_0) + o_P(n^{-1/2}).$$

From (A.3.4) we derive that

$$H\{\sigma^2A^-(\beta_0,\theta_0)\}^{-1/2}Q_n(\breve{g},\beta_0,\theta_0) \xrightarrow{D} N(0,I_{p+q}), \tag{T.5}$$

where I_{p+q} is the $(p+q)\times(p+q)$ identity matrix. Results (T.5) and (T.4) together prove Theorem (2.1). □

Proof of Theorem 3.1. Note that $\hat{A}(\beta_0,\theta_0) \xrightarrow{P} A(\beta_0,\theta_0)$ and $\hat{B}(\beta_0,\theta_0) \xrightarrow{P} B(\beta_0,\theta_0)$. By the expansion of (3.2) and

$$\log\hat{L}(\beta,\theta) = -\frac{n}{2}Q_n^\tau(\breve{g},\beta,\theta)\{\sigma^2B(\beta,\theta)\}^{-1}Q_n(\breve{g},\beta,\theta) + o_P(1),$$

we obtain

$$\hat{l}_{ael}(\beta_0,\theta_0) = nQ_n^\tau(\breve{g},\beta_0,\theta_0)\{\sigma^{-2}A^-(\beta_0,\theta_0)\}^{-1}Q_n(\breve{g},\beta_0,\theta_0) + o_P(1). \tag{T.6}$$

Hence, (T.6) proves Theorem (3.1). □

5.2 Appendix B: Proofs of lemmas

Lemma 5.1. *Suppose conditions (C1)-(C3), (C5) and (C6) hold. Then*

$$\sup_{t\in\mathcal{T}_w,\beta\in\mathcal{B}_n} \|\breve{g}(t;\beta)-g_0(t)\| = O_P(h^2+\{\frac{\log n}{nh}\}^{1/2}), \tag{5.1}$$

and

$$\sup_{t\in\mathcal{T}_w,\beta\in\mathcal{B}_n} \|\breve{\dot{g}}(t;\beta)-\dot{g}_0(t)\| = O_P(h+\{\frac{\log n}{nh^3}\}^{1/2}). \tag{5.2}$$

Proof of Lemma 5.1. This lemma is a direct extension of known results in nonparametric function estimation, we can find its proof in Wang and Xue[18], they used the result of Theorem 2 in Einmahl and Mason[29], we omit the detail here. □

To make formulations more concise, we give some notations here. Denote $\mathcal{G}=\{g:\mathcal{T}_w\times\mathcal{B}\mapsto R^d\}$, $\|g\|_{\mathcal{G}}=\sup_{t\in\mathcal{T}_w,\beta\in\mathcal{B}}\|g(t;\beta)\|$. From Lemma (5.1), we have $\|\breve{g}-g_0\|_{\mathcal{G}}=o_P(1)$ and $\|\breve{\dot{g}}-\dot{g}_0\|_{\mathcal{G}}=o_P(1)$. Hence we can assume that g lies in $\mathcal{G}_\delta$ with $\delta=\delta_n\to 0$ and $\delta>0$, where

$$\mathcal{G}_\delta=\{g\in\mathcal{G}:\|g-g_0\|_{\mathcal{G}}\le\delta, \|\dot{g}-\dot{g}_0\|_{\mathcal{G}}\le\delta\}, \tag{A.1.1}$$

$$Q(g,\beta,\theta)=E[\{Y-\theta^\tau Z-g^\tau(\beta^\tau U;\beta)X\}(\dot{g}^\tau(\beta^\tau U;\beta)XU^\tau, Z^\tau)^\tau w(\beta^\tau U)], \tag{A.1.2}$$

$$Q_n(g,\beta,\theta)=\frac{1}{n}\sum_{i=1}^{n}\{Y_i-\theta^\tau Z_i-g^\tau(\beta^\tau U_i;\beta)X_i\}(\dot{g}^\tau(\beta^\tau U_i;\beta)X_iU_i^\tau, Z_i^\tau)^\tau w(\beta^\tau U_i). \tag{A.1.3}$$

Lemma 5.2. *Suppose conditions (C1)-(C6) hold. Then*

$$\sqrt{n}(\breve{\theta}-\theta_0)\xrightarrow{D} N(0,\sigma^2 M_\theta^{-1}), \tag{5.3}$$

where $M_\theta=E[\{Z-C^\tau(\beta_0^\tau U)D^{-1}(\beta_0^\tau U)X\}\{Z-C^\tau(\beta_0^\tau U)D^{-1}(\beta_0^\tau U)X\}^\tau]$, *and* $C(\cdot)$, $D(\cdot)$ *are defined in condition C6.*

Proof of Lemma 5.2. From (2.9) and (2.12), we have

$$\begin{aligned}&\sqrt{n}(\breve{\theta}-\theta_0)\\&=\sqrt{n}\{Z^\tau(I_n-S_\beta)^\tau(I_n-S_\beta)Z\}^{-1}Z^\tau(I_n-S_\beta)^\tau(I_n-S_\beta)(Y-Z\theta_0)\\&=\{\frac{1}{n}Z^\tau(I_n-S_\beta)^\tau(I_n-S_\beta)Z\}^{-1}\frac{1}{\sqrt{n}}Z^\tau(I_n-S_\beta)^\tau(I_n-S_\beta)(G_\beta-\varepsilon).\end{aligned}$$

Let's consider the three terms on the right-side hand of the equation given in the above.
First, we show that

$$\frac{1}{\sqrt{n}}Z^\tau(I_n-S_\beta)^\tau(I_n-S_\beta)\varepsilon\xrightarrow{D} N(0,\sigma^2 M_\theta). \tag{A.2.1}$$

We have

$$\frac{1}{n}\xi_{0,\beta}^\tau W_{0,\beta}\xi_{0,\beta}=\begin{pmatrix}\frac{1}{n}\sum_{i=1}^n K_h(\beta^\tau U_i-t)X_iX_i^\tau & \frac{1}{n}\sum_{i=1}^n K_h(\beta^\tau U_i-t)X_iX_i^\tau(\frac{\beta^\tau U_i-t}{h})\\ \frac{1}{n}\sum_{i=1}^n K_h(\beta^\tau U_i-t)X_iX_i^\tau(\frac{\beta^\tau U_i-t}{h}) & \frac{1}{n}\sum_{i=1}^n K_h(\beta^\tau U_i-t)X_iX_i^\tau(\frac{\beta^\tau U_i-t}{h})^2\end{pmatrix}, \tag{5.4}$$

$$\frac{1}{n}\xi_{0,\beta}^\tau W_{0,\beta}Z=\begin{pmatrix}\frac{1}{n}\sum_{i=1}^n K_h(\beta^\tau U_i-t)X_iZ_i^\tau\\ \frac{1}{n}\sum_{i=1}^n K_h(\beta^\tau U_i-t)X_iZ_i^\tau(\beta^\tau U_i-t)\end{pmatrix}, \tag{5.5}$$

$$\frac{1}{n}\xi_{0,\beta}^\tau W_{0,\beta}\varepsilon=\begin{pmatrix}\frac{1}{n}\sum_{i=1}^n K_h(\beta^\tau U_i-t)X_i\varepsilon_i\\ \frac{1}{n}\sum_{i=1}^n K_h(\beta^\tau U_i-t)X_i\varepsilon_i(\beta^\tau U_i-t)\end{pmatrix}. \tag{5.6}$$

Note that each entry of the above matrices has a standard kernel estimation form, hence

$$\frac{1}{n}\xi_{0,\beta}^{\tau}W_{0,\beta}\xi_{0,\beta} = f(t)D(t)\otimes diag\{1,\mu_2\} + O_P(h^2+\{\log n/nh\}^{1/2}), \tag{5.7}$$

$$\frac{1}{n}\xi_{0,\beta}^{\tau}W_{0,\beta}Z = f(t)C(t)\otimes \{1,0\}^{\tau} + O_P(h^2+\{\log n/nh\}^{1/2}), \tag{5.8}$$

$$\frac{1}{n}\xi_{0,\beta}^{\tau}W_{0,\beta}\varepsilon = O_P(h^2+\{\log n/nh\}^{1/2}), \tag{5.9}$$

uniformly for $t\in\mathcal{T}_w$ and $\beta\in\mathcal{B}_n$, here $\otimes$ denotes the Kronecker product.
Let $T_i = Z_i - ([X_i, 0_d](\xi_{i,\beta}^{\tau}W_{i,\beta}\xi_{i,\beta})^{-1}\xi_{i,\beta}^{\tau}W_{i,\beta}Z)^{\tau}$, now we have

$$\begin{aligned}
&\frac{1}{n}Z^{\tau}(I_n-S_\beta)^{\tau}(I_n-S_\beta)\varepsilon\\
&=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}T_i[\varepsilon_i-([X_i,0_d](\xi_{i,\beta}^{\tau}W_{i,\beta}\xi_{i,\beta})^{-1}\xi_{i,\beta}^{\tau}W_{i,\beta}\varepsilon)]\\
&=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}T_i\varepsilon_i-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}T_i[X_i,0_d](\xi_{i,\beta}^{\tau}W_{i,\beta}\xi_{i,\beta})^{-1}\xi_{i,\beta}^{\tau}W_{i,\beta}\varepsilon.
\end{aligned}$$

Since we have $\|\beta-\beta_0\| = O_P(n^{-1/2})$ when $\beta\in\mathcal{B}_n$. Then, $\sup_{\beta\in\mathcal{B}_n}\|\xi_{0,\beta}^{\tau}W_{0,\beta}\xi_{0,\beta}-\xi_{0,\beta_0}^{\tau}W_{0,\beta_0}\xi_{0,\beta_0}\| = O_P(n^{1/2})$ and $\sup_{\beta\in\mathcal{B}_n}\|\xi_{0,\beta}^{\tau}W_{0,\beta}Z-\xi_{0,\beta_0}^{\tau}W_{0,\beta_0}Z\| = O_P(n^{1/2})$. By (5.7), (5.8), (5.9) and the results in the above, we have

$$T_i = Z_i - C^{\tau}(\beta_0^{\tau}U_i)D^{-1}(\beta_0^{\tau}U_i)X_i + O_P(n^{-1/2}+h^2+\{\log n/nh\}^{1/2}), \tag{5.10}$$

$$[X_i,0_d](\xi_{i,\beta}^{\tau}W_{i,\beta}\xi_{i,\beta})^{-1}\xi_{i,\beta}^{\tau}W_{i,\beta}Z = \{C^{\tau}(\beta^{\tau}U_i)D^{-1}(\beta^{\tau}U_i)X_i\}^{\tau} + O_P(h^2+\{\log n/nh\}^{1/2}), \tag{5.11}$$

$$[X_i,0_d](\xi_{i,\beta}^{\tau}W_{i,\beta}\xi_{i,\beta})^{-1}\xi_{i,\beta}^{\tau}W_{i,\beta}\varepsilon = O_P(h^2+\{\log n/nh\}^{1/2}). \tag{5.12}$$

Then, using the Law of large Numbers, it's not hard to obtain

$$E[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}T_i\varepsilon_i] = 0,$$

$$E[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}T_i([X_i,0_d](\xi_{i,\beta}^{\tau}W_{i,\beta}\xi_{i,\beta})^{-1}\xi_{i,\beta}^{\tau}W_{i,\beta}\varepsilon)] = 0,$$

$$E[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}T_i\varepsilon_i]^2 = \sigma^2E[\{Z-C^{\tau}(\beta_0^{\tau}U)D^{-1}(\beta_0^{\tau}U)X\}\{Z-C^{\tau}(\beta_0^{\tau}U)D^{-1}(\beta_0^{\tau}U)X\}^{\tau}] + o(1).$$

Also, we have

$$\begin{aligned}
&E\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}T_i([X_i,0_d](\xi_{i,\beta}^{\tau}W_{i,\beta}\xi_{i,\beta})^{-1}\xi_{i,\beta}^{\tau}W_{i,\beta}\varepsilon\|^2\\
&\le E\|T_i\|^2E\|T_i([X_i,0_d](\xi_{i,\beta}^{\tau}W_{i,\beta}\xi_{i,\beta})^{-1}\xi_{i,\beta}^{\tau}W_{i,\beta}\varepsilon\|^2.
\end{aligned}$$

Then, we can derive that

$$E\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}[X_i,0_d](\xi_{i,\beta}^{\tau}W_{i,\beta}\xi_{i,\beta})^{-1}\xi_{i,\beta}^{\tau}W_{i,\beta}\varepsilon\}^2 = o(1).$$

Using the results in the above, we can prove Lemma(5.2) by applying the central limit theorem and Slutsky's theorem.

Second, we will show that

$$\frac{1}{\sqrt{n}}Z^\tau(I_n - S_\beta)^\tau(I_n - S_\beta)G_\beta = o_P(1). \quad (A.2.2)$$

Similar to (5.4) and (5.7), we have

$$\begin{aligned}&\frac{1}{n}\xi_{0,\beta}^\tau W_{0,\beta}G_\beta\\ &= \begin{pmatrix} \frac{1}{n}\sum_{i=1}^n K_h(\beta^\tau U_i - t)g_0^\tau(\beta^\tau U_i)X_iX_i^\tau \\ \frac{1}{n}\sum_{i=1}^n K_h(\beta^\tau U_i - t)g_0^\tau(\beta^\tau U_i)X_iX_i^\tau(\frac{\beta^\tau U_i - t}{h})\end{pmatrix}\\ &= f(t)D(t)g_0(t)\otimes[1,0]^\tau + O_P(h^2 + \{\log n/nh\}^{1/2}),\end{aligned}$$

uniformly for $t \in \mathcal{T}_w$ and $\beta \in \mathcal{B}_n$. Then, we have

$$[X_i, 0_d](\xi_{i,\beta}^\tau W_{i,\beta}\xi_{i,\beta})^{-1}\xi_{i,\beta}^\tau W_{i,\beta}G_\beta = \{g_0^\tau(\beta_0^\tau U_i)X_i\}^\tau + O_P(h^2 + \{\log n/nh\}^{1/2}).$$

So, we have the result as followed

$$\begin{aligned}&\frac{1}{\sqrt{n}}Z^\tau(I_n - S_\beta)^\tau(I_n - S_\beta)G_\beta\\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^n T_i\{g^\tau(\beta^\tau U_i)X_i - [X_i, 0_d](\xi_{i,\beta}^\tau W_{i,\beta}\xi_{i,\beta})^{-1}\xi_{i,\beta}^\tau W_{i,\beta}G_\beta\}\\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^n\{g_0^\tau(\beta_0^\tau U_i)X_i\}^\tau O_P(h^2 + \{\log n/nh\}^{1/2})(Z_i - C^\tau(\beta_0^\tau U_i)D^{-1}(\beta_0^\tau U_i)X_i)\\ &= O_P(\sqrt{n}[h^2 + \{\log n/nh\}^{1/2}]^2).\end{aligned}$$

Checking the condition (C6), it's obvious that

$$\lim_{n\to\infty}\sqrt{n}(h^2 + \{\log n/nh\}^{1/2})^2 = 0.$$

Hence, (A.2.2) is proved.

Third, obviously,

$$\begin{aligned}&\frac{1}{\sqrt{n}}Z^\tau(I_n - S_\beta)^\tau(I_n - S_\beta)Z\\ &= E[\{Z_i - C^\tau(\beta_0^\tau U_i)D^{-1}(\beta_0^\tau U_i)X_i\}\{Z_i - C^\tau(\beta_0^\tau U_i)D^{-1}(\beta_0^\tau U_i)X_i\}^\tau] + O_P(\frac{1}{n} + h^2 + \{\log n/nh\}^{1/2}),\end{aligned}$$

uniformly for $t \in \mathcal{T}_w$ and $\beta \in \mathcal{B}_n$. By the Law of large Numbers, we have

$$\begin{aligned}&\frac{1}{\sqrt{n}}Z^\tau(I_n - S_\beta)^\tau(I_n - S_\beta)Z\\ &= E[\{Z - C^\tau(\beta_0^\tau U)D^{-1}(\beta_0^\tau U)X\}\{Z - C^\tau(\beta_0^\tau U)D^{-1}(\beta_0^\tau U)X\}^\tau] + o_P(1). \quad (A.2.3)\end{aligned}$$

Hence, (A.2.1), (A.2.2) and (A.2.3) are together to prove Lemma (5.2). □

Lemma 5.3. *Suppose that conditions (C1)-(C6) hold. Then*

$$\sup_{(g,\beta,\theta)\in\mathcal{G}_\delta\times\mathcal{B}_n\times\Theta_n}\|J_1(g,\beta,\theta)\| = o_P(n^{-1/2}), \quad (A.3.1)$$

$$\sup_{\beta\in\mathcal{B}_n}\|J_2(\breve{g},\beta)\| = o_P(n^{-1/2}), \quad (A.3.2)$$

$$\sup_{(g,\beta,\theta)\in\mathcal{G}_\delta\times\mathcal{B}_n\times\Theta_n}\|J_3(g,\beta,\theta)\| = o(n^{-1/2}), \quad (A.3.3)$$

$$\sqrt{n}J_4(\breve{g},\beta,\theta)\xrightarrow{D}N(0,\sigma^2A(\beta_0,\theta_0)), \quad (A.3.4)$$

where $A(\beta_0, \theta_0)$ is defined in (2.16),

$$\begin{aligned} J_1(g,\beta,\theta) &= Q_n(g,\beta,\theta) - Q(g,\beta,\theta) - Q_n(g_0,\beta_0,\theta_0),\\ J_2(g,\beta) &= Q(g,\beta,\theta) - Q(g_0,\beta,\theta)\\ &\quad - \varpi(g_0(\beta^\tau U;\beta),\beta)\{g(\beta^\tau U;\beta) - g_0(\beta^\tau U;\beta)\},\\ J_3(g,\beta,\theta) &= \varpi(g_0(\beta^\tau U;\beta),\beta)\{g(\beta^\tau U;\beta) - g_0(\beta^\tau U;\beta)\}\\ &\quad - \varpi(g_0(\beta^\tau U;\beta),\beta_0)\{g(\beta^\tau U;\beta) - g_0(\beta^\tau U;\beta)\},\\ J_4(g,\beta_0,\theta) &= Q_n(g_0,\beta_0,\theta_0) + \varpi(g_0(\beta^\tau U;\beta),\beta_0)\{g(\beta^\tau U;\beta) - g_0(\beta^\tau U;\beta)\}. \end{aligned}$$

Proof of Lemma 5.3. To prove (A.3.1). Denote $r_n(g,\beta,\theta) = \sqrt{n}\{Q_n(g,\beta,\theta) - Q(g,\beta,\theta)\}$. Noting that $Q(g_0,\beta_0,\theta_0) = 0$, we have

$$J_1(g,\beta,\theta) = n^{-1/2}\{r_n(g,\beta,\theta) - r_n(g_0,\beta_0,\theta_0)\}. \tag{A.3.5}$$

It can be shown that the empirical process $\{r_n(g,\beta,\theta) : g \in \mathcal{G}_1, \beta \in \mathcal{B}_1, \theta \in \Theta_1\}$ has the stochastic equicontinuity, where $\mathcal{B}_1 = \{\beta \in \mathcal{B} : \|\beta - \beta_0\| \le 1\}$, $\Theta_1 = \{\theta \in \Theta : \|\theta - \theta_0\| \le 1\}$ and $\mathcal{G}_1$ is defined in (A.1.1) with δ=1. The equicontinuity is sufficient for proof of (A.3.1) since $\delta < 1$ for large enough n.

To prove (A.3.2), define the functional derivative $\varpi(g_0(\cdot;\beta),\beta)$ of $Q(g,\beta,\theta)$ with respect to $g(\cdot;\beta)$ at $g_0(\cdot;\beta)$ at the direction $g(\cdot;\beta) - g_0(\cdot;\beta)$ by

$$\begin{aligned} &\varpi(g_0(\cdot;\beta),\beta)\{g(\cdot;\beta) - g_0(\cdot;\beta)\}\\ &= \lim_{\tau\to 0}[Q(g_0(\cdot;\beta) + \tau(g(\cdot;\beta) - g_0(\cdot;\beta)),\beta,\theta) - Q(g_0(\cdot;\beta),\beta,\theta)]\cdot\frac{1}{\tau}. \end{aligned}$$

Let $g_0(\beta^\tau U;\beta) = E\{g_0(\beta_0^\tau U)|\beta^\tau U\}$ and $\dot{g}_0(\beta^\tau U;\beta) = E\{\dot{g}_0(\beta_0^\tau U)|\beta^\tau U\}$. Obviously, $g_0(\beta_0^\tau U;\beta_0) = g_0(\beta_0^\tau U)$ and $\dot{g}_0(\beta_0^\tau U;\beta_0) = \dot{g}_0(\beta_0^\tau U)$. Then, some elementary calculation yields that

$$\begin{aligned} J_2(g,\beta) = &- E[\{g(\beta^\tau U;\beta) - g_0(\beta^\tau U;\beta)\}^\tau X\\ &\times(\{\dot{g}(\beta^\tau U;\beta) - \dot{g}_0(\beta^\tau U;\beta)\}^\tau X U^\tau, \mathbf{0}_q^\tau)^\tau w(\beta^\tau U)], \end{aligned}$$

here $\mathbf{0}_q$ represents the q-dimensional zero vector. Note that the lower q components of $J_2(g,\beta)$ are $\mathbf{0}$, so we just have to consider about the upper p components of $J_2(g,\beta)$. Denote

$$\begin{aligned} J_{2,1}(g,\beta) = &- E[\{g(\beta^\tau U;\beta) - g_0(\beta^\tau U;\beta)\}^\tau X\\ &\times\{\dot{g}(\beta^\tau U;\beta) - \dot{g}_0(\beta^\tau U;\beta)\}^\tau X U w(\beta^\tau U)]. \end{aligned} \tag{A.3.6}$$

Similarly to the proof of Lemma 2 in Xue and Wang[28], for any p-dimension vector ω, we have

$$\begin{aligned} \omega^\tau J_{2,1}(\breve{g},\beta) = &-\int\{\breve{g}(t;\beta) - g_0(t)\}^\tau\mu_\omega(t)\\ &\times\{\dot{\breve{g}}(t;\beta) - \dot{g}_0(t)\}w(t)f(t)dt + o_P(n^{-1/2}), \end{aligned}$$

where $\mu_\omega(t) = E\{X\omega^\tau U X^\tau|\beta^\tau U = t\}$, and $f(t)$ is the probability density of $\beta^\tau U$. By using the standard argument of nonparametric estimation, it's not hard to prove

$$\breve{g}(t;\beta) - g_0(t) = D^{-1}(t)\{f(t)\}^{-1}\phi_n(t;\beta) + O_P(n^{-1/2} + h^2 + \{\log n/nh\}^{1/2}),$$

uniformly for $t \in \mathcal{T}_w$ and $\beta \in \mathcal{B}_n$, where

$$\phi_n(t;\beta) = \frac{1}{n}\sum_{i=1}^{n}\{Y_i - g_0^\tau(\beta^\tau U_i)X_i\}X_i K_h(\beta^\tau U_i - t).$$

Hence, we can derive that

$$\omega^\tau J_{2,1}(\breve{g},\beta) = -\int\{D^{-1}(t)\phi_n(t;\beta)\}^\tau\mu_\omega(t)\{\dot{\breve{g}}(t;\beta) - \dot{g}_0(t)\}dt + O_P(n^{-1/2} + h^2 + \{\log n/nh\}^{1/2})$$

$$= -n^{-1/2}\{\gamma_n(\breve{g},\beta)-\gamma_n(\dot{g}_0,\beta)\}+O_P(n^{-1/2}+h^2+\{\log n/nh\}^{1/2}),$$

where $\gamma_n(\dot{g},\beta)=n^{-1/2}\sum_{i=1}^n \varepsilon_i w(\beta^\tau U_i)X_i^\tau D^{-1}(\beta^\tau U_i)\mu_\omega(\beta^\tau U_i)\dot{g}(\beta^\tau U_i;\beta)$. Similar to the proof of (A.3.5), we can use the empirical process techniques and show that the stochastic equicontinuity of $\gamma_n(\dot{g},\beta)$, and hence $\|\gamma_n(\breve{g},\beta)-\gamma_n(\dot{g}_0,\beta)\|=o_P(1)$. By checking the condition (C4), the proof of (A.3.2) is complete.

We now prove (A.3.3). For the convenience, let $J_{3,1}(g,\beta,\theta)$ and $J_{3,2}(g,\beta,\theta)$ denote the upper p components and the lower q components of $J_3(g,\beta,\theta)$ respectively. Hence, we have

$$\begin{aligned}J_{3,1}(g,\beta,\theta)=&-E[(\theta_0-\theta)^\tau Z[\dot{g}(\beta^\tau U;\beta)-\dot{g}_0(\beta^\tau U;\beta)]^\tau XUw(\beta^\tau U)]\\&+E[\{(\theta_0-\theta)^\tau Z[\dot{g}(\beta^\tau U;\beta_0)-\dot{g}_0(\beta^\tau U;\beta_0)]^\tau XUw(\beta^\tau U)]\\&+E[\{g(\beta^\tau U;\beta)-g_0(\beta^\tau U;\beta)\}^\tau X\dot{g}_0^\tau(\beta^\tau U;\beta)XUw(\beta^\tau U)]\\&-E[\{g(\beta^\tau U;\beta_0)-g_0(\beta^\tau U;\beta_0)\}^\tau X\dot{g}_0^\tau(\beta^\tau U;\beta_0)XUw(\beta^\tau U)],\end{aligned}$$

and

$$J_{3,2}(g,\beta,\theta)=E[\{g(\beta^\tau U;\beta)-g_0(\beta^\tau U;\beta)+g_0(\beta^\tau U;\beta_0)-g(\beta^\tau U;\beta_0)\}^\tau XZw(\beta^\tau U)]. \tag{5.13}$$

Check the condition (C6), and denote $\psi(\dot{g}_0,\beta)=\dot{g}_0^\tau(\beta^\tau U;\beta)XUw(\beta^\tau U)$ and $\varphi(g,\beta)=\{g(\beta^\tau U;\beta)-g_0(\beta^\tau U;\beta)\}^\tau X$. Then, we have

$$\begin{aligned}J_{3,1}(g,\beta,\theta)&=E[\varphi(g,\beta)\psi(\dot{g}_0,\beta)]-E[\varphi(g,\beta_0)\psi(\dot{g}_0,\beta_0)]+o_P(n^{-1/2})\\&=E[\{\varphi(g,\beta)-\varphi(g,\beta_0)\}\psi(\dot{g}_0,\beta)]+E[\varphi(g,\beta_0)\{\psi(\dot{g},\beta)-\psi(\dot{g}_0,\beta_0)\}]+o_P(n^{-1/2})\\&\triangleq J_{3,1a}(g,\beta)+J_{3,1b}(g,\beta)+o_P(n^{-1/2}).\end{aligned}$$

By condition (C2), we have

$$\begin{aligned}&\|\varphi(g,\beta)-\varphi(g,\beta_0)\|\\&\quad=\|\{g(\beta^\tau U;\beta)-g_0(\beta^\tau U;\beta)+g_0(\beta^\tau U;\beta_0)-g(\beta^\tau U;\beta_0)\}^\tau X\|\\&\quad=\{\dot{g}(\beta_1^\tau U;\beta_1)-\dot{g}_0(\beta_2^\tau U)\}^\tau X(\beta-\beta_0)\{U-E[U|\beta_0^\tau U]\}\|\\&\quad\le\|\dot{g}-\dot{g}_0\|_{\mathcal{G}}\|\beta-\beta_0\|(\|U-E[U|\beta_0^\tau U]\|)(\|X\|),\end{aligned}$$

where β_1 and β_2 are between β and β_0, and $\|\psi(\dot{g}_0,\beta)\|\le c(\|X\|)(\|U\|)$. Now, we have $\|J_{3,1a}(g,\beta)\|=o(n^{-1/2})$, uniformly for $g\in\mathcal{G}_\delta$ and $\beta\in\mathcal{B}_n$, $\theta\in\Theta_n$. Similarly, we can prove $\|J_{3,1b}(g,\beta)\|=o(n^{-1/2})$ and $J_{3,2}(g,\beta)=o(n^{-1/2})$, uniformly for $g\in\mathcal{G}_\delta$ and $\beta\in\mathcal{B}_n$, $\theta\in\Theta_n$.

Finally, we prove (A.3.4). Using the dominated convergence theorem ,we can obtain

$$\begin{aligned}&\varpi(g_0(\beta^\tau U),\beta_0)\{g(\beta_0^\tau U;\beta_0)-g_0(\beta_0^\tau U)\}\\&\quad=-E[\{g(\beta_0^\tau U;\beta_0)-g_0(\beta_0^\tau U)\}^\tau X(\dot{g}_0^\tau(\beta_0^\tau U)XU^\tau,Z^\tau)^\tau w(\beta^\tau U)]\\&\qquad+E[\{(\theta_0-\theta)^\tau Z\}([\dot{g}(\beta_0^\tau U;\beta_0)-\dot{g}_0(\beta_0^\tau U)]^\tau XU^\tau,\mathbf{0}_q^\tau)^\tau w(\beta^\tau U)]\\&\quad=-\int C(t)\{\breve{g}(t,\beta_0)-g_0(t)\}f(t)dt+o_P(n^{-1/2}+h^2+\{\log n/nh\}^{1/2})\\&\quad=-\frac{1}{n}\sum_{i=1}^n\varepsilon_i C(\beta_0^\tau U_i)D^{-1}(\beta_0^\tau U_i)X_i+o_P(n^{-1/2}+h^2+\{\log n/nh\}^{1/2}).\end{aligned}$$

This together with (A.1.3) proves that

$$J_4(\breve{g},\beta_0,\theta_0)=\frac{1}{n}\varepsilon_i\zeta_i+o_P(n^{-1/2}+h^2+\{\log n/nh\}^{1/2}),$$

where $\zeta_i=w(\beta_0^\tau U_i)\dot{g}_0^\tau(\beta_0^\tau U_i)X_iU_i-C(\beta_0^\tau U_i)D^{-1}(\beta_0^\tau U_i)X_i$. By the central limit theorem and Slutsky's theorem, we have

$$\sqrt{n}J_4(\breve{g},\beta_0,\theta_0)\xrightarrow{D}N(0,\sigma^2A(\beta_0,\theta_0)). \tag{A.3.7}$$

Therefore, the proof of Lemma (5.3) is complete. □

Lemma 5.4. *Suppose that conditions (C1)-(C6) hold. Then*

$$\sup_{(\beta,\theta)\in\mathcal{B}_n\times\Theta_n} \|Q_n(\breve{g},\beta,\theta)\| = O_P(n^{-1/2}), \tag{A.4.1}$$

$$\sup_{(\beta,\theta)\in\mathcal{B}_n\times\Theta_n} \|R_n(\beta,\theta)-\sigma^2 B(\beta_0,\theta_0)\| = o_P(1), \tag{A.4.2}$$

$$\sup_{(\beta,\theta)\in\mathcal{B}_n\times\Theta_n} \max_{1\le i\le n}\|\hat{\eta}_i(\beta,\theta)\| = o_P(n^{1/2}), \tag{A.4.3}$$

$$\sup_{(\beta,\theta)\in\mathcal{B}_n\times\Theta_n} \|\lambda(\beta,\theta)\| = o_P(n^{-1/2}), \tag{A.4.4}$$

where $Q_n(\breve{g},\beta,\theta)=\frac{1}{n}\sum_{i=1}^n\{Y_i-\breve{g}^\tau(\beta^\tau U_i;\beta)X_i-\theta^\tau Z_i\}w(\beta^\tau U_i)(\breve{g}^\tau(\beta^\tau U_i;\beta)X_iU_i^\tau, Z_i^\tau)^\tau$, $R_n(\beta,\theta)=\frac{1}{n}\sum_{i=1}^n\hat{\eta}_i(\beta,\theta)\hat{\eta}_i^\tau(\beta,\theta)$ *and* $B(\beta_0,\theta_0)$ *is defined in condition (C7) and* $\hat{\eta}_i$ *is defined in (2.7).*

Proof of Lemma 5.4. By Lemma (5.3) and (T.4), note that $Q_n(\breve{g},\breve{\beta},\breve{\theta})=0$ and $Q(g_0,\beta_0,\theta_0)=0$, we can prove (A.4.1). To prove(A.4.2), let

$$\begin{aligned}R_{ni}(\beta,\theta) &= \hat{\eta}_i(\beta,\theta)-\eta_i(\beta_0,\theta_0)\\
&= \varepsilon_i[w(\beta^\tau U_i;\beta)-w(\beta_0^\tau U_i)](\dot{g}_0^\tau(\beta_0^\tau U_i)X_iU_i^\tau, Z_i^\tau)^\tau\\
&\quad+\varepsilon_i[w(\beta^\tau U_i;\beta)-w(\beta_0^\tau U_i)]\{[\dot{\breve{g}}(\beta^\tau U_i;\beta)-\dot{g}_0(\beta_0^\tau U_i)]^\tau X_iU_i^\tau, 0^\tau\}^\tau\\
&\quad+\varepsilon_i w(\beta^\tau U_i;\beta)\{[\dot{\breve{g}}(\beta^\tau U_i;\beta)-\dot{g}_0(\beta_0^\tau U_i)]^\tau X_iU_i^\tau, 0^\tau\}^\tau\\
&\quad+[g_0(\beta_0^\tau U_i)-\breve{g}(\beta^\tau U_i;\beta)]^\tau X_i w(\beta^\tau U_i;\beta)(\dot{g}_0^\tau(\beta_0^\tau U_i)X_iU_i^\tau, Z_i^\tau)^\tau\\
&\quad+(\theta_0-\theta)^\tau Z_i w(\beta^\tau U_i;\beta)(\dot{g}_0^\tau(\beta_0^\tau U_i)X_iU_i^\tau, Z_i^\tau)^\tau\\
&\quad+[g_0(\beta_0^\tau U_i)-\breve{g}(\beta^\tau U_i;\beta)]^\tau X_i w(\beta^\tau U_i;\beta)\{[\dot{\breve{g}}(\beta^\tau U_i;\beta)-\dot{g}_0(\beta_0^\tau U_i)]^\tau X_iU_i^\tau, 0^\tau\}^\tau\\
&\quad+(\theta_0-\theta)^\tau Z_i w(\beta^\tau U_i;\beta)\{[\dot{\breve{g}}(\beta^\tau U_i;\beta)-\dot{g}_0(\beta_0^\tau U_i)]^\tau X_iU_i^\tau, 0^\tau\}^\tau,\end{aligned}$$

where $\eta_i(\cdot)$ is defined in (2.2), hence

$$\begin{aligned}R_n(\beta,\theta) &= \frac{1}{n}\sum_{i=1}^n\eta_i(\beta_0,\theta_0)\eta_i^\tau(\beta_0,\theta_0)+\frac{1}{n}\sum_{i=1}^n R_{ni}(\beta,\theta)R_{ni}^\tau(\beta,\theta)\\
&\quad+\frac{1}{n}\sum_{i=1}^n\eta_i(\beta_0,\theta_0)R_{ni}^\tau(\beta,\theta)+\frac{1}{n}\sum_{i=1}^n R_{ni}(\beta,\theta)\eta_i^\tau(\beta_0,\theta_0)\\
&\triangleq M_1(\beta_0,\theta_0)+M_2(\beta,\theta)+M_3(\beta,\theta)+M_4(\beta,\theta).\end{aligned}$$

By the law of large numbers, we have $M_1(\beta_0,\theta_0)\xrightarrow{P}\sigma^2B(\beta_0,\theta_0)$. In order to prove (A.4.2), we only need to prove that $M_l(\beta,\theta)\xrightarrow{P}0$ uniformly for (β,θ), $l=2,3,4$.

Let $M_{2,st}(\beta,\theta)$ denote the (s,t) element of $M_2(\beta,\theta)$, and $R_{ni,s}(\beta,\theta)$ denote the sth component of $R_{ni}(\beta,\theta)$. By Cauchy-Schwarz inequality,we have

$$|M_{2,st}(\beta,\theta)|\le\Big(\frac{1}{n}\sum_{i=1}^n R_{ni,s}^2(\beta,\theta)\Big)^{1/2}\Big(\frac{1}{n}\sum_{i=1}^n R_{ni,t}^2(\beta,\theta)\Big)^{1/2}. \tag{A.4.5}$$

Since the lower q components of $R_{ni,s}(\beta,\theta)$ are zero, we need only to consider the upper p components of $R_{ni,s}(\beta,\theta)$. It can be shown by some elementary calculation that

$$\frac{1}{n}\sum_{i=1}^n R_{ni,s}^2(\beta,\theta)\xrightarrow{P}0,$$

uniformly for $(\beta,\theta)\in\mathcal{B}_n\times\Theta_n$. By this, we can prove that $M_2(\beta,\theta)\xrightarrow{P}0$, uniformly for $(\beta,\theta)\in\mathcal{B}_n\times\Theta_n$. Similarly, it can be shown that $M_3(\beta,\theta)\xrightarrow{P}0$ and $M_4(\beta,\theta)\xrightarrow{P}0$ uniformly for $(\beta,\theta)\in\mathcal{B}_n\times\Theta_n$. Hence, this proves (A.4.2).

Similar to the proof of (A.4.2), we can derive (A.4.3) and (A.4.4), we omit the detail here. □

References

[1] Feng S., Xue L., Model detection and estimation for single-index varying coefficient model, *Journal of Multivariate Analysis*, 2015, 139, 227-244

[2] Owen A.B., Empirical likelihood ratio confidence intervals for a single functional,*Biometrika*, 1988, 75 (2), 237-249

[3] Owen A., Empirical likelihood ratio confidence regions, *The Annals of Statistics*, 1990, 18 (1), 90-120

[4] Owen A., Empirical likelihood for linear models, *The Annals of Statistics*, 1991, 1725-1747

[5] Wang Q., Rao J.N., Empirical likelihood-based inference in linear errors-in-covariables models with validation data, *Biometrika*, 2002, 89 (2), 345-358

[6] Wang Q., Linton O., Härdle W., Semiparametric regression analysis with missing response at random, *Journal of the American Statistical Association*, 2004, 99 (466), 334-345

[7] Xue L. G., Zhu L., Empirical likelihood for single-index models, *Journal of Multivariate Analysis*, 2006, 97 (6), 1295-1312

[8] Xue L., Zhu L., Empirical likelihood for a varying coefficient model with longitudinal data, *Journal of the American Statistical Association*, 2007, 102 (478), 642-654

[9] Xue L., Zhu L., Empirical likelihood semiparametric regression analysis for longitudinal data, *Biometrika*, 2007, 94(4), 921-937

[10] Zhu L., Xue L., Empirical likelihood confidence regions in a partially linear single-index model, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 2006, 68 (3), 549-570

[11] You J., Zhou Y., Empirical likelihood for semiparametric varying-coefficient partially linear regression models, *Statistics & Probability Letters*, 2006, 76 (4), 412-422

[12] Qin J., Zhang B., Empirical-likelihood-based inference in missing response problems and its application in observational studies, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 2007, 69 (1), 101-122

[13] Stute W., Xue L., Zhu L., Empirical likelihood inference in nonlinear errors-in-covariables models with validation data, *Journal of the American Statistical Association*, 2007, 102 (477), 332-346

[14] Xue L., Empirical likelihood confidence intervals for response mean with data missing at random, *Scandinavian Journal of Statistics*, 2009, 36 (4), 671-685

[15] Xue L., Empirical likelihood for linear models with missing responses, *Journal of Multivariate Analysis*, 2009, 100 (7), 1353-1366

[16] Wang J.L., Xue L., Zhu L., Chong Y.S., Estimation for a partial-linear single-index model, *The Annals of statistics*, 2010, 38 (1), 246-274

[17] Huang Z., Zhang R., Efficient empirical-likelihood-based inferences for the single-index model, *Journal of Multivariate Analysis*, 2011, 102 (5), 937-947

[18] Wang Q., Xue L. (2011). Statistical inference in partially-varying-coefficient single-index model, *Journal of Multivariate Analysis*, 2011, 102 (1), 1-19

[19] Lian H. (2012). Empirical likelihood confidence intervals for nonparametric functional data analysis, *Journal of Statistical Planning and Inference*, 2012, 142 (7), 1669-1677

[20] Xiao Y., Tian Z., Li F., Empirical likelihood-based inference for parameter and nonparametric function in partially nonlinear models. *Journal of the Korean Statistical Society*, 2014, 43 (3), 367-379

[21] Zhou X., Zhao P., Wang X., Empirical likelihood inferences for varying coefficient partially nonlinear models, *Journal of Applied Statistics*, 2017, 4 (3), 474-492

[22] Fang J., Liu W., Lu X., Empirical likelihood for heteroscedastic partially linear single-index models with growing dimensional data, *Metrika*, 2018, 81 (3), 255-281

[23] Arteaga-Molina L.A., Rodriguez-Poo J.M., Empirical likelihood based inference for fixed effects varying coefficient panel data models, *Journal of Statistical Planning and Inference*, 2018, 196, 144-162

[24] Härdle W., Hall P., Ichimura H., Optimal smoothing in single-index models, *The annals of Statistics*, 1993, 157-178

[25] Xia Y., Li W.K., On single-index coefficient regression models, *Journal of the American Statistical Association*, 1999, 94 (448), 1275-1285

[26] Xue L., Pang Z., Statistical inference for a single-index varying-coefficient model, *Statistics and Computing*, 2013, 23 (5), 589-599

[27] Rao J.N., Scott A.J., The analysis of categorical data from complex sample surveys: chi-squared tests for goodness of fit and independence in two-way tables, *Journal of the American statistical association*, 1981, 76 (374), 221-230

[28] Xue L., Wang Q., Empirical likelihood for single-index varying-coefficient models, 2012,*Bernoulli*, 18 (3), 836-856

[29] Einmahl U., Mason D.M., Uniform in bandwidth consistency of kernel-type function estimators, *The Annals of Statistics*, 2005, 33 (3), 1380-1403