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

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

\[ Y = g_0(U)X + \theta_0^T Z + \epsilon, \]

where \((U, X, Z) \in \mathbb{R}^p \times \mathbb{R}^d \times \mathbb{R}^q\) is a vector of covariates, \(Y\) is the response variable, \(\beta_0\) is a \(p \times 1\) vector of unknown parameters, \(\theta_0\) is a \(q \times 1\) vector of regression coefficient, \(g_0(\cdot)\) is a \(d \times 1\) vector of unknown functions and \(\epsilon\) is a random error with \(E(\epsilon \mid U, X, Z) = 0\) and \(\text{Var}(\epsilon \mid U, X, Z) = \sigma^2\). Assume that \(\epsilon\) 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

© 2019 Xiang and Liu, published by De Gruyter. This work is licensed under the Creative Commons Attribution 4.0 License.
partial linear single-index model, single-index varying-coefficient model, etc. The linear component \( \theta_0^T Z \) provides a simple summary of covariates effects which are of the main scientific interest. The index \( \beta_0^T U \) 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 \((\beta_0, \theta_0)\) 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 \((\beta_0, \theta_0)\). In this paper, we can construct an empirical likelihood ratio function for \((\beta_0, \theta_0)\) by assuming \( g_0(\cdot) \) and its derivative \( \bar{g}_0(\cdot) \) 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 estimating the empirical likelihood ratio and investigating 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 \( \{(Y_i, U_i, X_i, Z_i); 1 \leq i \leq n\} \) is an i.i.d sample from model (1.1), that is

\[
Y_i = g_0(\beta_0^T U_i)X_i + \theta_0^T Z_i + \epsilon_i, \quad i = 1, \ldots, n, \tag{2.1}
\]

where \( \epsilon_i \)s are i.i.d. random errors with mean 0 and finite variance \( \sigma^2 \). Assume that \( \epsilon_i \)s are independent of \( \{U_i, X_i, Z_i; 1 \leq i \leq n\} \).

Our primary interest is to construct the confidence region of \((\beta_0, \theta_0)\). In order to construct an empirical likelihood ratio function for \((\beta_0, \theta_0)\), we introduce an auxiliary random vector

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

where \( \bar{g}_0(\cdot) \) 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 \( \bar{g}_0(\cdot) \), 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 \((\beta, \theta)\) is the true parameter is equivalent to testing whether \( E[\eta_i(\beta, \theta)] = 0 \) if \((\beta, \theta)\) 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

\[
L_n(\beta, \theta) = \max \left\{ \prod_{i=1}^{n} (np_i) \mid p_i \geq 0, \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i \eta_i(\beta, \theta) = 0 \right\}. \tag{2.3}
\]

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 \((\beta_0, \theta_0)\) 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 \( \beta_n(\beta, \theta) \) by their estimators and define an estimated empirical likelihood function.

When \( (\beta, \theta) \) 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 \( g(t; \beta_0, \theta_0) \) and \( \dot{g}_0(t) \) are defined as \( \hat{g}(t; \beta_0, \theta_0) = \bar{a} \) and \( \hat{g}(t; \beta_0, \theta_0) = \bar{b} \) at fixed point \((\beta_0, \theta_0)\), where \( \bar{a} \) and \( \bar{b} \) minimize the sum of weighted squares

\[
\sum_{i=1}^{n} [Y_i - \theta_0^T Z_i - \{a + b(\beta_i^T U_i - t)\}^T X_i]^2 K_h(\beta_i^T U_i - t),
\]

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 \( \infty \). Simple calculation yields

\[
(\hat{g}(t; \beta_0), h\hat{g}(t; \beta_0))^T = (\xi_0^T W_{0, \beta_0} \xi_0, \theta_0)^{-1}\xi_0^T W_{0, \beta_0}(Y - Z0),
\]

where

\[
\xi_{0, \beta_0} = \begin{pmatrix}
X_1^T X_1(\beta_0^T U_1 - t)/h \\
\vdots \\
X_n^T X_n(\beta_0^T U_n - t)/h
\end{pmatrix},
\]

\[
W_{0, \beta_0} = \text{diag}\{K_h(\beta_0^T U_1 - t), \ldots, K_h(\beta_0^T U_n - t)\},
\]

\[
Y = (Y_1, \ldots, Y_n)^T \text{ and } Z = (Z_1, \ldots, Z_n)^T.
\]

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

\[
\hat{l}(\beta, \theta) = -2\max\{\sum_{i=1}^{n} \log(\eta_i) \mid \eta_i \geq 0, \sum_{i=1}^{n} \eta_i = 1, \sum_{i=1}^{n} \eta_i \hat{n}_i(\beta, \theta) = 0\}.
\]

By the Lagrange multiplier method, 

\[
\hat{l}(\beta, \theta) = 2 \sum_{i=1}^{n} \log(1 + \lambda^T \hat{n}_i(\beta, \theta)),
\]

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

\[
\frac{1}{n} \sum_{i=1}^{n} \frac{\hat{n}_i(\beta, \theta)}{1 + \lambda^T \hat{n}_i(\beta, \theta)} = 0.
\]

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

\[
\tilde{G}_\beta = \begin{pmatrix}
[X_1^T, 0_d](\xi_{1, \beta_0}^T W_{1, \beta_0} \xi_{1, \beta_0})^{-1} \xi_{1, \beta_0}^T W_{1, \beta_0} & \ldots \\
\vdots & \vdots \\
[X_n^T, 0_d](\xi_{n, \beta_0}^T W_{n, \beta_0} \xi_{n, \beta_0})^{-1} \xi_{n, \beta_0}^T W_{n, \beta_0}
\end{pmatrix}
\]

\[
\dot{S}_{\beta_0}(Y - Z0).
\]

here \( 0_d \) denotes the \( d \)-dimensional zero vector, let \( \xi_{i, \beta_0} \) and \( W_{i, \beta_0} \) denote \( \dot{\xi}_{0, \beta_0} \) and \( W_{0, \beta_0} \) with \( t \) replaced by \( \beta_i^T U_i \) for \( i = 1, \ldots, n \). Hence we get an approximate linear model as follows

\[
(I_n - \tilde{S}_{\beta_0})Y = (I_n - S_{\beta_0})Z0 + \varepsilon.
\]

here \( I_n \) denotes the \( n \)-th identity matrix.
Secondly, we can use the least square theory to obtain
\[ \hat{\beta} = (Z^T (I_n - S_{\beta_0}) Z)^{-1} Z^T (I_n - S_{\beta_0}) Y, \]
we can get the estimators of \( g(\cdot) \) and \( \hat{g}(\cdot) \) at \( t = \beta_0^* U \) by substituting (2.12) into (2.5) as follows
\[ (\hat{g}(t; \beta_0), \hat{h}(t; \beta_0)) = (\xi_{0, \beta_0} W_{0, \beta_0} \xi_{0, \beta_0})^{-1} \xi_{0, \beta_0} W_{0, \beta_0} (Y - Z \hat{\beta}). \]

Thirdly, noticed that (2.13) and (2.12) are based on a known \( \beta_0 \). Under the condition \( ||\beta|| = 1 \), we have an estimator \( \hat{\beta} \) of \( \beta_0 \) by minimizing the following equation that
\[ N(\beta) = \sum_{i=1}^{n} \{Y_i - \hat{g}(\beta^T U_i) X_i - \hat{\beta}^T Z_i\}^2. \]  

Obviously, solving (2.14) is equivalent to solving the following equation under the condition \( ||\beta|| = 1 \) that
\[ \sum_{i=1}^{n} \{Y_i - \hat{g}(\beta^T U_i) X_i - \hat{\beta}^T Z_i\} \hat{g}(\beta^T U_i) X_i U_i w(\beta^T U_i) = 0. \]

Finally, we get the final estimators \( \hat{\beta}, \hat{g}(\cdot) \) and \( \hat{g}(\cdot) \) with \( \beta_0 \) replaced by \( \hat{\beta} \) in \( \hat{\beta}, \hat{g}(\cdot) \) and \( \hat{g}(\cdot) \), respectively.

### 2.2 Asymptotic properties

Let \( \mathcal{B}_n = \{ \beta \in \mathcal{B} : ||\beta - \beta_0|| \leq c_0 n^{-1/2} \} \) and \( \Theta_n = \{ \theta \in \Theta : ||\theta - \theta_0|| \leq c_0 n^{-1/2} \} \) 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 \( \beta^T U \) is bounded away from zero for \( t \in \mathcal{T}_w \) and \( \beta \) near \( \beta_0 \), 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 \( g_j(t), 1 \leq j \leq q \), have continuous second derivatives on \( \mathcal{T}_w \), where \( g_j(t) \) are the \( j \)th components of \( g_0(t) \).

(C3) \( E(||U||^6) < \infty, E(||X||^6) < \infty, E(||Z||^6) < \infty, E(|e|^6) < \infty \).

(C4) The bandwidths satisfy that \( h \to 0, nh^2 / \log n \to \infty, nh^4 \log n = o_P(1), nh^6 = o_P(1), nh^{3/2} / \log^2 n \to \infty, h_1 \sim n^{-1/5} \).

(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 \( D(t) = E(XX^T | \beta_0^T U = t) \) is positive definite, and each entry of \( D(t) \), \( D^{-1}(t) \) and \( C(t) = E(XZ^T | \beta_0^T U = t) \) satisfies the Lipschitz condition of order 1 on \( \mathcal{T}_w \), where \( A = w(\beta_0^T U) g_0^T (\beta_0^T U) XU^T, Z^T \).

(C7) The matrix \( B(\beta_0, \theta_0) = E(\Lambda \Lambda^T) \) 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 \( \chi^2_1 \) variables.

**Theorem 2.1.** Suppose conditions (C1)-(C7) hold. If \( (\beta_0, \theta_0) \) is the true value of the parameter, then
\[ -2 \log \hat{L}(\beta_0, \theta_0) \xrightarrow{D} w_1 \chi^2_1 + \cdots + w_{p+q} \chi^2_1, \]
where \( \xrightarrow{D} \) denotes the convergence in distribution, the weights \( w_j (1 \leq j \leq p + q) \) are the eigenvalues of \( V(\beta_0, \theta_0) = B^{-1}(\beta_0, \theta_0) A(\beta_0, \theta_0), \) and \( (\chi^2_1, 1 \leq i \leq p + q) \) are the independent \( \chi^2_1 \) variables. Where
\[ A(\beta_0, \theta_0) = B(\beta_0, \theta_0) - E[C(\beta_0^T U) D^{-1}(\beta_0^T U) C^T(\beta_0^T U)]. \]
In order to apply Theorem (2.1) to construct a confidence region for \((\beta_0, \theta_0)\), we have to estimate the unknown weights \(w_i\) consistently. By the plug-in method, \(A(\beta_0, \theta_0)\) and \(B(\beta_0, \theta_0)\) can be estimated consistently by

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

(2.17)

and

\[
\hat{B}(\hat{\beta}, \hat{\theta}) = \frac{1}{n} \sum_{i=1}^{n} \hat{\Lambda}_i \hat{\Lambda}_i^T,
\]

(2.18)

respectively, where \((\hat{\beta}, \hat{\theta})\) is defined in the above, and \(\hat{\Lambda}_i = w(\hat{\beta}^T U_i) \hat{\xi}(\hat{\beta}^T U_i) X_i \hat{V}_i^T\) and \(\hat{D}(\cdot) = \sum_{i=1}^{n} W_{ni}(\cdot) X_i X_i^T\) with

\[
W_{ni}(\cdot) = K_1(\hat{\beta}^T U_{ni}) \sum_{k=1}^{n} K_1(\hat{\beta}^T U_{nk} - \cdot),
\]

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

This means that \(\hat{\theta}_i\), the eigenvalues of \(\hat{\Sigma}(\hat{\beta}, \hat{\theta}) = \hat{B}^{-1}(\hat{\beta}, \hat{\theta}) \hat{\Lambda}(\hat{\beta}, \hat{\theta})\), consistently estimate \(\theta_i\) for \(j = 1, \cdots, p + q\). Let \(s_{1-a}\) be the \(1 - a\) quantile of the conditional distribution of the weighted sum \(\hat{s} = \hat{w}_1 \chi^2_{1,1} + \cdots + \hat{w}_{p+q} \chi^2_{1,p+q}\) given the data. Then we can define an approximate \(1 - a\) confidence region for \((\beta_0, \theta_0)\) as

\[
\mathcal{R}_{eel}(a) = \{ (\beta, \theta) \in \mathcal{B} \times \Theta : -2 \log \hat{L}(\beta, \theta) \leq s_{1-a} \}.
\]

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

### 3 Adjusted empirical likelihood

When we use Theorem (2.1) to construct confidence regions of \((\beta, \theta)\), the weights \(w_i\) need to be estimated, the accuracy of confidence region is decreased. Let \(\rho(\beta_0, \theta_0) = (p + q) / \text{tr}(V_0(\beta_0, \theta_0))\), where \(V_0(\beta_0, \theta_0)\) is defined in Theorem (2.1). By Rao and Scott[27], the distribution of \(\rho(\beta_0, \theta_0) \sum_{i=1}^{n} w_i \chi^2_{1,i}\) can be approximated by \(\chi^2_{p+q}\), 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

\[
\hat{L}(\beta, \theta) = \hat{\rho}(\beta, \theta) (-2 \log \hat{L}(\beta, \theta)).
\]

(3.1)

However, the accuracy of this approximation depends on the \(w_i\)'s. 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

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

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

\[
\hat{\tau}(\beta, \theta) = \frac{\text{tr}(\hat{A}^- (\beta, \theta) \hat{S}(\beta, \theta))}{\text{tr}(\hat{B}^{-1}(\beta, \theta) \hat{S}(\beta, \theta))},
\]

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

\[
\hat{L}(\beta, \theta) = \hat{\tau}(\beta, \theta) (-2 \log \hat{L}(\beta, \theta)).
\]

(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 \((\beta_0, \theta_0)\).

Thereby, we can obtain the confidence region of \((\beta, \theta)\)
\[ \mathfrak{R}_{\text{eff}}(\alpha) = \{ (\beta, \theta) \in \mathbb{B} \times \Theta : \hat{l}(\beta, \theta) \leq c_{1-\alpha} \}, \]
where \( P(\chi^2_{P+q} \leq c_{1-\alpha}) = 1 - \alpha \).

To apply Theorem (3.1) to construct a confidence region for \((\beta_0, \theta_0)\), we only need to estimate the adjustment factor \( \hat{h}(\beta, \theta) \) by replacing \((\beta, \theta)\) by \((\hat{\beta}, \hat{\theta})\). The value of \(-2 \log \hat{l}(\beta, \theta)\) is not depend on the estimation of \((\beta, \theta)\).

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

where \( \beta_0 = (\sqrt{2}/2, \sqrt{2}/2)^T \), \( \theta_0 = (2, 1.2)^T \), \( \varepsilon_i \sim N(0, 0.3^2) \). The sample \( \{U_i = (U_{i1}, U_{i2})^T; 1 \leq i \leq n\} \) 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 \( \{Z_i = (Z_{i1}, Z_{i2})^T; 1 \leq i \leq n\} \) was generated from a bivariate normal distribution \( N(0, \Sigma) \) with \( \text{Cov}(Z_{ik}, Z_{il}) = 2 \times 0.5^{|k-l|}, k, l = 1, 2 \). 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 \( w(t) = 1_{[-\sqrt{2}, \sqrt{2}]}(t) \). The bandwidths \( \hat{h} = \hat{h}_1 n^{-1/25} (\log n)^{-1/2} \) and \( \hat{h}_1 = \hat{h}_{\text{opt}} \) respectively, where \( \hat{h}_{\text{opt}} \) 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).

| 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
\[ -2 \log \hat{l}(\beta_0, \theta_0) = - \sum_{i=1}^{n} \{ \lambda^T \hat{h}_i(\beta_0, \theta_0) - \frac{1}{2} [\lambda^T \hat{h}_i(\beta_0, \theta_0)]^2 \} + o_p(1), \quad (T.1) \]
Employing (2.8), we have

\[
0 = \frac{1}{n} \sum_{i=1}^{n} \frac{\hat{h}_i(\beta_0, \theta_0)}{1 + \lambda^T \hat{h}_i(\beta_0, \theta_0)} = \frac{1}{n} \sum_{i=1}^{n} \hat{h}_i(\beta_0, \theta_0) - \frac{1}{n} \sum_{i=1}^{n} \hat{h}_i(\beta_0, \theta_0)\hat{h}_i^T(\beta_0, \theta_0)\lambda + \frac{1}{n} \sum_{i=1}^{n} \hat{h}_i(\beta_0, \theta_0)[\lambda^T \hat{h}_i(\beta_0, \theta_0)]^2.
\]

The application of Lemma (5.4) yields that

\[
\sum_{i=1}^{n} [\lambda^T \hat{h}_i(\beta_0, \theta_0)]^2 = n \lambda^T \hat{h}_i(\beta_0, \theta_0) + o_p(1),
\]

and

\[
\lambda = \left[ \sum_{i=1}^{n} \hat{h}_i(\beta_0, \theta_0)\hat{h}_i^T(\beta_0, \theta_0) \right]^{-1} \sum_{i=1}^{n} \hat{h}_i(\beta_0, \theta_0) + o_p(n^{-1/2}).
\]

This together with (T.1) proves that

\[
-2 \log \hat{L}(\beta_0, \theta_0) = nQ_n^* (\tilde{g}, \beta_0, \theta_0)R_n^{-1} (\beta_0, \theta_0)Q_n (\tilde{g}, \beta_0, \theta_0) + o_p(1).
\]

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

\[
\hat{L}(\beta_0, \theta_0) = [(\sigma^2 A)^{-1/2} \sqrt{n}Q_n (\tilde{g}, \beta_0, \theta_0)]^T V(\beta_0, \theta_0)((\sigma^2 A)^{-1/2} \sqrt{n}Q_n (\tilde{g}, \beta_0, \theta_0)) + o_p(1),
\]

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 \( \tilde{V} = \text{diag}(w_1, \ldots, w_{p+q}) \), where \( w_i, 1 \leq i \leq p + q \), are the eigenvalues of \( V(\beta_0, \theta_0) \). Then there exists an orthogonal matrix \( H \) such that \( H^T \tilde{V}H = V(\beta_0, \theta_0) \). Thus, we have

\[
Q_n (\tilde{g}, \beta, \theta) = J_1 (\tilde{g}, \beta, \theta) + J_2 (\tilde{g}, \beta) + J_3 (\tilde{g}, \beta, \theta) + J_4 (\tilde{g}, \beta, \theta) + Q(g_0, \beta, \theta).
\]

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

\[
Q_n (\tilde{g}, \beta_0, \theta_0) = J_4 (\tilde{g}, \beta_0, \theta_0) + o_p(n^{-1/2}).
\]

From (A.3.4) we derive that

\[
H(\sigma^2 A^{-1}(\beta_0, \theta_0))^{-1/2} Q_n (\tilde{g}, \beta_0, \theta_0) \xrightarrow{D} N(0, I_{p+q}),
\]

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

\[
\log \hat{L}(\beta, \theta) = -\frac{n}{2} Q_n^* (\tilde{g}, \beta, \theta) (\sigma^2 B(\beta, \theta))^{-1} Q_n (\tilde{g}, \beta, \theta) + o_p(1),
\]

we obtain

\[
\hat{1}_{det}(\beta_0, \theta_0) = nQ_n^* (\tilde{g}, \beta_0, \theta_0) (\sigma^2 A^{-1}(\beta_0, \theta_0))^{-1} Q_n (\tilde{g}, \beta_0, \theta_0) + o_p(1).
\]

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

\[ \sup_{t \in \mathcal{T} \cap B_n} \| \hat{g}(t; \beta) - g_0(t) \| = O_p(h^2 + \{ \frac{\log n}{nh} \}^{1/2}), \]  

(5.1)

and

\[ \sup_{t \in \mathcal{T} \cap B_n} \| \hat{g}(t; \beta) - \hat{g}_0(t) \| = O_p(h + \{ \frac{\log n}{nh^2} \}^{1/2}). \]  

(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 \( \mathcal{G} = \{ g : \mathcal{T} \times B \rightarrow \mathbb{R}^d \}, \| g \|_G = \sup_{t \in \mathcal{T} \cap B_n} \| g(t; \beta) \|. \) From Lemma (5.1), we have \( \| \hat{g} - g_0 \|_G = o_p(1) \) and \( \| \hat{g} - \hat{g}_0 \|_G = o_p(1). \) Hence we can assume that \( g \) lies in \( \mathcal{G}_\delta \) with \( \delta = \delta_n \rightarrow 0 \) and \( \delta > 0, \) where

\[ \mathcal{G}_\delta = \{ g \in \mathcal{G} : \| g - g_0 \|_G \leq \delta, \| \hat{g} - g_0 \|_G \leq \delta \}, \]

(A.1.1)

\[ Q(g, \beta, \theta) = E\{ Y - \theta^T g^2(\beta^T U; \beta)X_i(\beta^T U; \beta)X_i^T, Z^T Xw(\beta^T U) \}, \]

(A.2.1)

\[ Q_n(g, \beta, \theta) = \frac{1}{n} \sum_{i=1}^n \{ Y_i - \theta^T g^2(\beta^T U; \beta)X_i(\beta^T U; \beta)X_i^T, Z^T Xw(\beta^T U) \}. \]

(A.1.3)

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

\[ \sqrt{n}(\hat{\beta} - \beta_0) \overset{D}{\rightarrow} N(0, \sigma^2 M_\beta^{-1}), \]

(5.3)

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

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

\[ \sqrt{n}(\hat{\beta} - \beta_0) = \sqrt{n} Z^T (I_n - \hat{S}_n)^{-1} Z Y_\beta - \beta_0), \]

(5.4)

\[ \begin{split} & = \frac{1}{n} \sum_{i=1}^n \{ Y_i - \theta^T g^2(\beta^T U; \beta)X_i(\beta^T U; \beta)X_i^T, Z^T Xw(\beta^T U) \} \overset{D}{\rightarrow} N(0, \sigma^2 M_\beta). \end{split} \]

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

\[ \frac{1}{\sqrt{n}} \{ Z^T (I_n - \hat{S}_n)^{-1} Z \} \overset{D}{\rightarrow} N(0, \sigma^2 M_\beta). \]

(A.2.1)

We have

\[ \frac{1}{n} \xi_{0, \beta}^T W_{0, \beta} \xi_{0, \beta} = \left( \begin{array}{c} \frac{1}{n} \sum_{i=1}^n K_h(\beta^T U_i - t)X_iX_i^T \frac{1}{n} \sum_{i=1}^n K_h(\beta^T U_i - t)X_iX_i^T (\frac{\beta^T U_i - t}{n}) \frac{1}{n} \sum_{i=1}^n K_h(\beta^T U_i - t)X_iX_i^T (\frac{\beta^T U_i - t}{n}^2) \end{array} \right), \]

(5.4)

\[ \frac{1}{n} \xi_{0, \beta}^T W_{0, \beta} Z = \left( \begin{array}{c} \frac{1}{n} \sum_{i=1}^n K_h(\beta^T U_i - t)X_iZ_i \frac{1}{n} \sum_{i=1}^n K_h(\beta^T U_i - t)X_iZ_i (\beta^T U_i - t) \end{array} \right), \]

(5.5)

\[ \frac{1}{n} \xi_{0, \beta}^T W_{0, \beta} \varepsilon = \left( \begin{array}{c} \frac{1}{n} \sum_{i=1}^n K_h(\beta^T U_i - t)X_i \varepsilon_i \frac{1}{n} \sum_{i=1}^n K_h(\beta^T U_i - t)X_i \varepsilon_i (\beta^T U_i - t) \end{array} \right). \]

(5.6)
Note that each entry of the above matrices has a standard kernel estimation form, hence

\[ \frac{1}{n} \xi^T W_{0, \beta} = f(t) D(t) \otimes \text{diag}(1, \mu_2) + O_p(h^2 + \{\log n/nh\}^{1/2}), \]  
(5.7)

\[ \frac{1}{n} \xi^T W_{0, \beta} Z = f(t) C(t) \otimes \{1, 0\}^T + O_p(h^2 + \{\log n/nh\}^{1/2}), \]  
(5.8)

\[ \frac{1}{n} \xi^T W_{0, \beta} \varepsilon = O_p(h^2 + \{\log n/nh\}^{1/2}), \]  
(5.9)

uniformly for \( t \in T_w \) and \( \beta \in \mathcal{B}_n \), here \( \otimes \) denotes the Kronecker product.

Let \( T_i = Z_i - ([X_i, O_d](\xi^T W_{1, \beta} \xi_i, \beta)^{-1} \xi^T W_{1, \beta} Z)^T \), now we have

\[
\begin{align*}
\frac{1}{n} T_i^T (I_n - S_{\beta})^T (I_n - S_{\beta}) \varepsilon &= \frac{1}{\sqrt{n}} \sum_{i=1}^n T_i [e_i - ([X_i, O_d](\xi^T W_{1, \beta} \xi_i, \beta)^{-1} \xi^T W_{1, \beta} \varepsilon)] \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^n T_i e_i - \frac{1}{\sqrt{n}} \sum_{i=1}^n T_i [X_i, O_d](\xi^T W_{1, \beta} \xi_i, \beta)^{-1} \xi^T W_{1, \beta} \varepsilon.
\end{align*}
\]

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

\[ T_i = Z_i - C^T (\beta_0^T U_1) D^{-1}(\beta_0^T U_1) X_i + O_p(n^{-1/2} + h^2 + \{\log n/nh\}^{1/2}), \]  
(5.10)

\[ [X_i, O_d](\xi^T W_{1, \beta} \xi_i, \beta)^{-1} \xi^T W_{1, \beta} Z = \{C^T (\beta^T U_1) D^{-1}(\beta^T U_1) X_i\}^T + O_p(h^2 + \{\log n/nh\}^{1/2}), \]  
(5.11)

\[ [X_i, O_d](\xi^T W_{1, \beta} \xi_i, \beta)^{-1} \xi^T W_{1, \beta} \varepsilon = O_p(h^2 + \{\log n/nh\}^{1/2}). \]  
(5.12)

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

\[
E\left[ \frac{1}{\sqrt{n}} \sum_{i=1}^n T_i e_i \right] = 0,
\]

\[
E\left[ \frac{1}{\sqrt{n}} \sum_{i=1}^n T_i [X_i, O_d](\xi^T W_{1, \beta} \xi_i, \beta)^{-1} \xi^T W_{1, \beta} \varepsilon] \right] = 0,
\]

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

Also, we have

\[
E\left[ \frac{1}{\sqrt{n}} \sum_{i=1}^n T_i [X_i, O_d](\xi^T W_{1, \beta} \xi_i, \beta)^{-1} \xi^T W_{1, \beta} \varepsilon] \right]^2 \leq E\|T_i\|^2 E\|T_i [X_i, O_d](\xi^T W_{1, \beta} \xi_i, \beta)^{-1} \xi^T W_{1, \beta} \varepsilon\|^2.
\]

Then, we can derive that

\[
E\left( \frac{1}{\sqrt{n}} \sum_{i=1}^n [X_i, O_d](\xi^T W_{1, \beta} \xi_i, \beta)^{-1} \xi^T W_{1, \beta} \varepsilon \right)^2 = o(1).
\]

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
\[
\frac{1}{\sqrt{n}} Z^T (I_n - S_\beta)^T (I_n - S_\beta) G_\beta = o_P(1).
\]
(A.2.2)

Similar to (5.4) and (5.7), we have
\[
\frac{1}{n} \xi_{i,\beta} W_{i,\beta} G_\beta
= \frac{1}{n} \sum_{i=1}^n K_h(\beta^T U_i - t) g_0'(\beta^T U_i) X_i X_i^T
= \left( \frac{1}{n} \sum_{i=1}^n K_h(\beta^T U_i - t) g_0'(\beta^T U_i) X_i X_i^T \right)
= f(t)D(t)g_0(t) \otimes [1, 0]^T + O_P(h^2 + \{\log n/nh\}^{1/2}),
\]
uniformly for \( t \in T_w \) and \( \beta \in \mathcal{B}_n \). Then, we have
\[
[X_i, O_d](\xi_{i,\beta} W_{i,\beta} \tilde{\xi}_{i,\beta})^{-1} \xi_{i,\beta} W_{i,\beta} G_\beta
= \{g_0'(\beta_0^T U_i)X_i\}^T + O_P(h^2 + \{\log n/nh\}^{1/2}).
\]

So, we have the result as followed
\[
\frac{1}{\sqrt{n}} Z^T (I_n - S_\beta)^T (I_n - S_\beta) G_\beta
= \frac{1}{n} \sum_{i=1}^n T_i (g^T(\beta^T U_i) X_i - [X_i, O_d](\xi_{i,\beta} W_{i,\beta} \tilde{\xi}_{i,\beta})^{-1} \xi_{i,\beta} W_{i,\beta} G_\beta)
= \frac{1}{n} \sum_{i=1}^n \{g_0'(\beta_0^T U_i)X_i\}^T O_P(h^2 + \{\log n/nh\}^{1/2})(Z_i - C^T(\beta_0^T U_i)D^{-1}(\beta_0^T U_i)X_i)
= O_P(\sqrt{n}[h^2 + \{\log n/nh\}^{1/2}]).
\]

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,
\[
\frac{1}{\sqrt{n}} Z^T (I_n - S_\beta)^T (I_n - S_\beta) Z
= E[(Z_i - C^T(\beta_0^T U_i)D^{-1}(\beta_0^T U_i)X_i)\{Z_i - C^T(\beta_0^T U_i)D^{-1}(\beta_0^T U_i)X_i\}^T] + O_P(\frac{1}{n} h^2 + \{\log n/nh\}^{1/2}),
\]
uniformly for \( t \in T_w \) and \( \beta \in \mathcal{B}_n \). By the Law of large Numbers, we have
\[
\frac{1}{\sqrt{n}} Z^T (I_n - S_\beta)^T (I_n - S_\beta) Z
= E[(Z - C^T(\beta_0^T U)D^{-1}(\beta_0^T U)X)\{Z - C^T(\beta_0^T U)D^{-1}(\beta_0^T U)X\}^T] + o_P(1).
\]
(A.2.3)

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

**Lemma 5.3.** Suppose that conditions (C1)-(C6) hold. Then
\[
\sup_{(g, \beta, \theta) \in S_\mathfrak{a} \times \mathcal{B}_n \times \Theta_n} \|J_1(g, \beta, \theta)\| = o_P(n^{-1/2}),
\]
(A.3.1)
\[
\sup_{\beta \in \mathcal{B}_n} \|J_2(\tilde{g}, \beta)\| = o_P(n^{-1/2}),
\]
(A.3.2)
\[
\sup_{(g, \beta, \theta) \in S_\mathfrak{a} \times \mathcal{B}_n \times \Theta_n} \|J_3(g, \beta, \theta)\| = o(n^{-1/2}),
\]
(A.3.3)
\[
\sqrt{n}J_4(\tilde{g}, \beta, \theta) \overset{D}{\rightarrow} N(0, \sigma^2 A(\beta_0, \theta_0)),
\]
(A.3.4)
where \( A(\beta_0, \theta_0) \) is defined in (2.16),

\[
\begin{align*}
J_1(g, \beta, \theta) &= Q_n(g, \beta, \theta) - Q(g, \beta, \theta) - Q_n(g_0, \beta_0, \theta_0), \\
J_2(g, \beta) &= Q(g, \beta, \theta) - Q(g_0, \beta, \theta) \\
&- \omega(g_0(\beta'; \beta), (g(\beta'; \beta) - g_0(\beta'; \beta))), \\
J_3(g, \beta, \theta) &= \omega(g_0(\beta'; \beta), (g(\beta'; \beta) - g_0(\beta'; \beta)) \\
&- \omega(g_0(\beta'; \beta), \beta_0)(g(\beta'; \beta) - g_0(\beta'; \beta)), \\
J_4(g, \beta, \theta) &= Q_n(g_0, \beta_0, \theta_0) + \omega(g_0(\beta'; \beta), \beta_0)(g(\beta'; \beta) - g_0(\beta'; \beta)).
\end{align*}
\]

**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 \( \{ r_n(g, \beta, \theta) : g \in \mathcal{S}_1, \beta \in \mathcal{B}_1, \theta \in \Theta_1 \} \) has the stochastic equicontinuity, where \( \mathcal{B}_1 = \{ \beta \in \mathcal{B} : \| \beta - \beta_0 \| \leq 1 \} \), \( \Theta_1 = \{ \theta \in \Theta : \| \theta - \theta_0 \| \leq 1 \} \) and \( \mathcal{S}_1 \) is defined in (A.1.1) with \( \delta = 1 \). 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 \( \omega(g_0(\cdot; \beta), \beta) \) of \( Q(g, \beta, \theta) \) with respect to \( g(\cdot; \beta) \) at the direction \( g(\cdot; \beta) - g_0(\cdot; \beta) \) by

\[
\omega(g_0(\cdot; \beta), \beta)(g(\cdot; \beta) - g_0(\cdot; \beta)) \\
= \lim_{t \to 0} \left[ Q(g_0(\cdot; \beta) + \tau(g(\cdot; \beta) - g_0(\cdot; \beta)), \beta, \theta) \right. \\
- Q(g(\cdot; \beta), \beta, \theta) \cdot \frac{1}{\tau}.
\]

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

\[
J_2(g, \beta) = -E\{ (g(\beta'; \beta) - g_0(\beta; \beta))\}^T X \\
\times \{ (\tilde{g}(\beta'; \beta) - \tilde{g}_0(\beta; \beta))\}^T X\mu(t),
\]

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

\[
J_{2,1}(g, \beta) = -E\{ (g(\beta'; \beta) - g_0(\beta; \beta))\}^T X \\
\times \{ (\tilde{g}(\beta'; \beta) - \tilde{g}_0(\beta; \beta))\}^T X\mu(t), \tag{A.3.6}
\]

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

\[
\omega^T J_{2,1}(g, \beta) = - \int \{ \tilde{g}(t; \beta) - \tilde{g}_0(t) \}^T \mu(t) \\\n\times \{ \tilde{g}(t; \beta) - \tilde{g}_0(t) \} w(t) f(t) dt + O_p(n^{-1/2}),
\]

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

\[
\tilde{g}(t; \beta) - g_0(t) = D^{-1}(t) f(t)^{-1} \phi_n(t; \beta) + O_p(n^{-1/2} + h^2 + \log n/nh)^{1/2},
\]

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

\[
\phi_n(t; \beta) = \frac{1}{n} \sum_{i=1}^n \{ Y_i - g_0^i(\beta' U_i)X_i \} X_i K_h(\beta' U_i - t).
\]

Hence, we can derive that

\[
\omega^T J_{2,1}(g, \beta) = - \int \{ D^{-1}(t) \phi_n(t; \beta) \}^T \mu(t) \{ \tilde{g}(t; \beta) - \tilde{g}_0(t) \} dt + O_p(n^{-1/2} + h^2 + \log n/nh)^{1/2})
\]
where $\gamma_\beta(\hat{g}, \beta) = n^{-1/2} \sum_{i=1}^{n} e_i w(\beta^T U_i) X_i D^{-1}(\beta^T U_i) \mu_\beta(\beta^T U_i) \hat{g}(\beta^T U_i)$. Similar to the proof of (A.3.5), we can use the empirical process techniques and show that the stochastic equicontinuity of $\gamma_\beta(\hat{g}, \beta)$, and hence $\|\gamma_\beta(\hat{g}, \beta) - \gamma_\beta(\hat{g}_0, \beta)\| = o_P(1)$. 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

$$J_{3,1}(g, \beta, \theta) = -E[(\theta_0 - \theta)^T Z [g(\beta^T U; \beta) - \hat{g}_0(\beta^T U; \beta)]^T X U w(\beta^T U)] + E[(\theta_0 - \theta)^T Z [\hat{g}(\beta^T U; \beta) - \hat{g}_0(\beta^T U; \beta)]^T X U w(\beta^T U)]$$

$$+ E[(g(\beta^T U; \beta) - \hat{g}(\beta^T U; \beta))^T X \hat{g}(\beta^T U; \beta) X U w(\beta^T U)] - E[(g(\beta^T U; \beta) - \hat{g}(\beta^T U; \beta))^T X \hat{g}_0(\beta^T U; \beta) X U w(\beta^T U)],$$

and

$$J_{3,2}(g, \beta, \theta) = E[(g(\beta^T U; \beta) - \hat{g}_0(\beta^T U; \beta) X U w(\beta^T U))] \varphi(g, \beta) = \{g(\beta^T U; \beta) - \hat{g}_0(\beta^T U; \beta)\}^T X.$$ Then, we have

$$J_{3,1}(g, \beta, \theta) = -E[E[\varphi(g, \beta) \psi(\hat{g}_0, \beta)] - E[\varphi(g, \beta_0) \psi(\hat{g}_0, \beta_0)] + o_P(n^{-1/2})$$

$$= E[\{\varphi(g, \beta) - \varphi(g, \beta_0)\} \psi(\hat{g}_0, \beta_0)] + o_P(n^{-1/2})$$

$$\leq J_{3,1a}(g, \beta) + J_{3,1b}(g, \beta) + o_P(n^{-1/2}).$$

By condition (C2), we have

$$||\varphi(g, \beta) - \varphi(g, \beta_0)||$$

$$= ||\{g(\beta^T U; \beta) - \hat{g}_0(\beta^T U; \beta) X U w(\beta^T U)\}^T X||$$

$$\leq ||\hat{g} - \hat{g}_0|| \|\beta - \beta_0\| ||U|| X||,$$

where $\beta_1$ and $\beta_2$ are between $\beta$ and $\beta_0$, and $||\psi(\hat{g}_0, \beta)|| \leq c(||X|| ||U||)$. Now, we have $||J_{3,1a}(g, \beta)|| = o(n^{-1/2})$, uniformly for $g \in G_\delta$ 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 $g \in G_\delta$ and $\beta \in \mathcal{B}_n$, $\theta \in \Theta_n$.

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

$$\omega(g(\beta^T U), \beta_0) \{g(\beta^T U; \beta) - \hat{g}_0(\beta^T U)\}$$

$$= -E[\{g(\beta^T U; \beta) - \hat{g}_0(\beta^T U)\}^T X \hat{g}_0(\beta^T U) X U^T, Z] w(\beta^T U)]$$

$$+ E[(\theta_0 - \theta)^T Z [\hat{g}(\beta^T U; \beta) - \hat{g}_0(\beta^T U)]^T X U^T, \theta_0^T Z] w(\beta^T U)]$$

$$= - \int C(t) [\hat{g}(t, \beta_0) - \hat{g}(t) f(t) dt + o_p(n^{-1/2} + h^2 + ||\log n/nh||^{1/2})$$

$$= - \frac{1}{n} \sum_{i=1}^{n} e_i C(\beta_0^T U_i) D^{-1}(\beta_0^T U_i) X_i + o_P(n^{-1/2} + h^2 + ||\log n/nh||^{1/2}).$$

This together with (A.1.3) proves that

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

where $\zeta_i = w(\beta_0^T U_i) \hat{g}_0(\beta_0^T U_i) X_i U_i - C(\beta_0^T U_i) D^{-1}(\beta_0^T U_i) X_i$. By the central limit theorem and Slutsky’s theorem, we have

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

Therefore, the proof of Lemma (5.3) is complete. □
Lemma 5.4. Suppose that conditions (C1)-(C6) hold. Then

\[
\sup_{(\beta, \theta) \in \mathcal{B}_n \times \Theta_n} \| Q_n(\tilde{g}, \beta, \theta) \| = O_P(n^{-1/2}),
\]

\[
\sup_{(\beta, \theta) \in \mathcal{B}_n \times \Theta_n} \| R_n(\beta, \theta) - \sigma^2 B(\beta_0, \theta_0) \| = o_P(1),
\]

\[
\sup_{(\beta, \theta) \in \mathcal{B}_n \times \Theta_n} \max_{1 \leq l \leq n} \| \hat{\eta}_l(\beta, \theta) \| = o_P(n^{1/2}),
\]

\[
\sup_{(\beta, \theta) \in \mathcal{B}_n \times \Theta_n} \| A(\beta, \theta) \| = o_P(n^{-1/2}),
\]

where \( Q_n(\tilde{g}, \beta, \theta) = \frac{1}{n} \sum_{i=1}^n \{ Y_i - \tilde{g}(\beta U_i; \beta) X_i - \theta Z_i \} w(\beta U_i) \tilde{g}(\beta U_i; \beta) X_i U_i^T, \)

\( R_n(\beta, \theta) = \frac{1}{n} \sum_{i=1}^n \hat{\eta}_i(\beta, \theta) \hat{\eta}_i(\beta, \theta) \) 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 \( Q_n(\tilde{g}, \beta, \theta) = 0 \) and \( Q(g_0, \beta_0, \theta_0) = 0 \), we can prove (A.4.1). To prove (A.4.2), let

\[
R_{ni}(\beta, \theta) = \tilde{\eta}_i(\beta, \theta) - \eta(\beta_0, \theta_0)
\]

\[
= \epsilon_1 [w(\beta U_i; \beta) - w(\beta_0 U_i)] \tilde{g}(\beta_0 U_i; \beta) X_i U_i^T, Z_i^T
\]

\[
+ \epsilon_1 [w(\beta U_i; \beta) - w(\beta_0 U_i)] \{ \tilde{g}(\beta_0 U_i; \beta) - g_0(\beta_0 U_i) \} X_i U_i^T, O^T
\]

\[
+ [g_0(\beta_0 U_i) - \tilde{g}(\beta_0 U_i; \beta)] X_i w(\beta_0 U_i; \beta)(\beta_0^2 U_i) X_i U_i^T, Z_i^T
\]

\[
+ (\theta_0 - \theta) X_i w(\beta_0 U_i; \beta)(\beta_0^2 U_i) X_i U_i^T, Z_i^T
\]

\[
+ [g_0(\beta_0 U_i) - \tilde{g}(\beta_0 U_i; \beta)] X_i w(\beta_0 U_i; \beta)(\beta_0^2 U_i) X_i U_i^T, O^T
\]

\[
+ (\theta_0 - \theta) X_i w(\beta_0 U_i; \beta)(\beta_0^2 U_i) X_i U_i^T, O^T
\]

where \( \eta(\cdot) \) is defined in (2.2), hence

\[
R_n(\beta, \theta) = \frac{1}{n} \sum_{i=1}^n \eta_i(\beta_0, \theta_0) \eta^*_i(\beta_0, \theta_0) + \frac{1}{n} \sum_{i=1}^n R_{ni}(\beta, \theta) R^*_n(\beta, \theta)
\]

\[
+ \frac{1}{n} \sum_{i=1}^n \eta_i(\beta_0, \theta_0) R^*_l(\beta, \theta) + \frac{1}{n} \sum_{i=1}^n R_{ni}(\beta, \theta) \eta^*_i(\beta_0, \theta_0)
\]

\[
\triangleq M_1(\beta_0, \theta_0) + M_2(\beta, \theta) + M_3(\beta, \theta) + M_4(\beta, \theta).
\]

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

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

\[
|M_{2, st}(\beta, \theta)| \leq \left( \frac{1}{n} \sum_{i=1}^n R^2_{nl,s}(\beta, \theta) \right)^{1/2} \left( \frac{1}{n} \sum_{i=1}^n R^2_{nl,t}(\beta, \theta) \right)^{1/2}.
\]

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

\[
\frac{1}{n} \sum_{i=1}^n R^2_{nl,s}(\beta, \theta) \xrightarrow{p} 0,
\]

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. \(\square\)
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-variables 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*, 2006, 99 (666), 334-345

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

[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*, 2016, 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