Journal of Multivariate Analysis 112 (2012) 1–23

Contents lists available at SciVerse ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Nonlinear models with measurement errors subject to single-indexed distortion☆

Jun Zhang, Li-Xing Zhu, Hua Liang *

Shenzhen-Hong Kong Joint Research Centre for Applied Statistics, Shenzhen University, Shenzhen, 518060, China
Department of Mathematics, Hong Kong Baptist University, Kowloon Tong, Hong Kong, China
Department of Biostatistics and Computational Biology, University of Rochester, Rochester, NY 14642, USA

ARTICLE INFO

Article history:
Received 20 December 2011
Available online 6 June 2012

AMS subject classifications:
62G08
62G20
62H12

Keywords:
Covariate-adjusted regression
Error-prone
Empirical likelihood
Estimating equation function
Local linear smoothing
Measurement errors models
Single index
Distorting function

ABSTRACT

We study nonlinear regression models whose both response and predictors are measured with errors and distorted as single-index models of some observable confounding variables, and propose a multicovariate-adjusted procedure. We first examine the relationship between the observed primary variables (observed response and observed predictors) and the confounding variables by appropriately estimating the single index. We then develop a semiparametric profile nonlinear least square estimation procedure for the parameters of interest after we calibrate the error-prone response and predictors. Asymptotic properties of the proposed estimators are established. To avoid estimating the asymptotic covariance matrix that contains the infinite-dimensional nuisance distorting functions and the single index, and to improve the accuracy of the proposed estimation, we also propose an empirical likelihood-based statistic, which is shown to be asymptotically chi-squared. A simulation study is conducted to evaluate the performance of the proposed methods and a real dataset is analyzed as an illustration.

1. Introduction

Consider the covariate-adjusted model

$$\begin{cases} Y = f(\mathbf{X}, \boldsymbol{\beta}) + \varepsilon, \\ \tilde{Y} = \phi(\theta^\tau U)Y, \\ \tilde{\mathbf{X}} = \psi(\theta^\tau U)\mathbf{X}, \end{cases} \tag{1}$$

where Y is the unobservable response, $\mathbf{X} = (X_1, X_2, \ldots, X_p)^\tau$ is a unobservable continuous predictor vector (the superscript τ denotes the transpose operator throughout this paper), $f(\cdot, \cdot)$ is a known continuous nonlinear function, $\boldsymbol{\beta}$ is an unknown $q \times 1$ parameter vector on a compact parameter space $B \subset R^q$, $\tilde{Y}$ and $\tilde{\mathbf{X}}$ are the observed response and predictors, θ is an unknown index vector, U is a confounding variable, and $\psi(\cdot)$ is a $p \times p$ diagonal matrix diag $(\psi_1(\cdot), \ldots, \psi_p(\cdot))$, where $\phi(\cdot)$ and $\psi_r(\cdot)$ are unknown continuous distorting functions. The diagonal form of $\psi(\cdot)$ indicates that the confounding

☆ Zhang's research was supported by the NSFC grant 11101157, China. Zhu's research was supported by a RGC grant from the Research Grants Council of Hong Kong, Hong Kong, China. Liang's research was partially supported by NSF grant DMS-1007167. This work was done when the first author visited the third author. The authors thank referees for their valuable comments and suggestions.

* Corresponding author at: Department of Biostatistics and Computational Biology, University of Rochester, Rochester, NY 14642, USA.
E-mail addresses: HLiang@bst.rochester.edu, hliang@bst.rochester.edu (H. Liang).

0047-259X/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmva.2012.05.012

variables distort each component of the unobserved predictors $\mathbf{X}$ in a multiplicative fashion. The confounding variable U is independent of $(\mathbf{X}, Y)$. Note that both Y and $\mathbf{X}$ are unobservable. These are essentially measurement error models.

There is substantial literature on nonlinear models with measurement errors. See [3] for a comprehensive survey, in which they systematically summarized the results for the cases when the components of $\mathbf{X}$ are measured with errors. In this paper, we study another class of measurement errors models, in which both the response and predictors are distorted by confounding variables. This occurrence is not uncommon in biomedical research and health-related studies. For instance, in a study of the relationship between the fibrinogen and serum transferrin levels among hemodialysis patients, Kaysen et al. [10] realized that the body mass index (BMI) generally has an influence on the fibrinogen and serum transferrin levels and may contaminate these variables. Thus, they suggested a calibration approach in which the response variable and predictors were simply divided by the confounding variable BMI. This implies a multiplicative fashion of the relationship between the unobserved primary variables and the confounding variable. Nevertheless, the exact relationship between the confounding variable and primary variables of interest is hardly known in practice, and the method of simply dividing the confounding variable itself to the variables of interest to estimate the original response Y and predictors $\mathbf{X}$ may cause non-negligible bias and lead to an inconsistent estimator of the parameter $\boldsymbol{\beta}$. As a remedy, Şentürk and Müller [19] suggested that the confounding variable BMI affects the primary variables through flexible multiplicative unknown functions and studied a linear covariate-adjusted model with emphasis on regression. Şentürk and Müller [20] further studied that linear covariate-adjusted model with a one-dimensional confounding variable in a setting in which the observed $\tilde{Y}$ and $\tilde{\mathbf{X}}$ are related through a varying coefficient model, using the binning method. Nguyen and Şentürk [14] then studied Şentürk and Müller's model with multi-dimensional confounding variables in the same setting as Şentürk and Müller [20] for the connection of $\tilde{Y}$ and $\tilde{\mathbf{X}}$. The authors modeled their distortion functions by single-index models and used a hybrid backfitting algorithm to simultaneously estimate the unknown single-index and varying coefficient functions. The final estimator of the major parameter is a weighted-average of the estimated coefficient functions. However, they did not provide theoretical justification for their approach. More recently, Cui et al. [4] studied nonlinear models with a one-dimensional confounding variable. They used the traditional nonparametric regression to obtain estimators of the distortion functions, say $\hat{\phi}(\cdot)$ and $\hat{\psi}(\cdot)$. Then they calibrated $\mathbf{X}$ and Y by $\hat{\psi}^{-1}\tilde{\mathbf{X}}$ and $\tilde{Y}/\hat{\phi}$, respectively, and engaged estimation by using these calibrated quantities. Zhang et al. [27] further examined this direct-plug-in method to the semiparametric models incorporating dimension reduction techniques. It is worth mentioning that the direct plug-in method can be easily adopted in linear, nonlinear, generalized linear, and semi-parametric models, while the transformation technique used by Şentürk and Müller [20,21] is designed for linear or generalized linear covariate-adjusted models.

In this paper we further investigate nonlinear covariate-adjusted models and allow the confounding variables to be multi-dimensional. We estimate the single-index θ using the recently developed estimating function method (EFM) by Cui et al. [5] because this method is more efficient than its competitors in the literature, is easy to implement, and is not sensitive to initial values. We then derive profile nonlinear least squares estimators of $\boldsymbol{\beta}$, establish asymptotic normality for the proposed estimators and correct a technical error in Lemma A.1 of [4], which plays a critical role in the proofs of their main theoretical results. As the asymptotic covariance matrix of the estimators of $\boldsymbol{\beta}$ contains several unknown components in a very complex structure, it may not be convenient for statistical inference-based on the normal approximation in practice. We therefore also propose an empirical likelihood based statistic, which is shown to be asymptotically chi-squared distributed and can be conveniently used to construct confidence regions.

The paper is organized as follows. In Section 2, we describe the estimation procedure for the single index θ and the parameter $\boldsymbol{\beta}$, present the asymptotic results, develop an empirical log-likelihood ratio statistic for the parameter $\boldsymbol{\beta}$, and show that the ratio statistic has an asymptotic chi-squared distribution. In Section 3, we report the results of a simulation study and an analysis of a diabetes study. All of the technical proofs of the asymptotic results are given in Appendix A.

2. Methodology and large sample properties

2.1. Estimating the single index θ

The parameter space of θ is assumed, without loss of generality, to be $\Theta = \{\theta = (\theta_1, \theta_2, \ldots, \theta_d)^\tau : \|\theta\| = 1,\ \theta_1 > 0,\ \theta \in R^d\}$. By re-parametrization, the parameter space Θ can be written as, after eliminating θ_1, a $(d-1)$-dimensional space $\{((1-\sum_{l=2}^d \theta_l^2)^{1/2}, \theta_2, \ldots, \theta_d)^\tau : \sum_{l=2}^d \theta_l^2 < 1\}$. The surface of the unit ball in R^d with $\|\theta\| = 1$ is transformed to the interior of the unit ball in R^{d-1} ($\sum_{l=2}^d \theta_l^2 < 1$). A variety of estimation methods for single-index models have been proposed in the literature. See [12,26,24,28,2] for more details.

Recall that U is independent of $(Y, \mathbf{X})$. The conditional mean and variance of $(\tilde{Y}, \tilde{\mathbf{X}})$ given U can be expressed as follows:

$$\boldsymbol{E}(\tilde{Y}|U) = \phi(\theta^\tau U)\boldsymbol{E}Y, \qquad \mathrm{Var}(\tilde{Y}|U) = \mathrm{Var}(Y)\phi^2(\theta^\tau U), \tag{2}$$

$$\boldsymbol{E}(\tilde{X}_r|U) = \psi_r(\theta^\tau U)\boldsymbol{E}X_r, \qquad \mathrm{Var}(\tilde{X}_r|U) = \mathrm{Var}(X_r)\psi_r^2(\theta^\tau U), \tag{3}$$

for $r = 1, \ldots, p$. (2) and (3) indicate that the conditional mean and variance contain the single-index θ and the unknown distorting functions $\phi(\cdot)$ and $\psi_r(\cdot)$, respectively.

Suppose that $\{(\tilde{\mathbf{X}}_i, \tilde{Y}_i, U_i), i = 1, \ldots, n\}$ is the i.i.d. random sample from $(\tilde{X}, \tilde{Y}, U)$. For any fixed θ, we use local linear regression to estimate $\phi(\cdot)$, $\psi_r(\cdot)$, and $\phi'(\cdot)$, $\psi_r'(\cdot)$. Let h denote the bandwidth, $K(\cdot)$ be the kernel function satisfying the conditions given in Appendix A, and $K_h(\cdot) = h^{-1}K(\cdot/h)$. For each t in a neighborhood of $\theta^\tau u$, we approximate $\phi(\theta^\tau u)$ and $\psi_r(\theta^\tau u)$ as follows. $\phi_a(\theta^\tau u) := \gamma_0 + \gamma_1(\theta^\tau u - t)$, and $\psi_{ar}(\theta^\tau u) := \gamma_{0r} + \gamma_{1r}(\theta^\tau u - t)$. The estimators $\hat{\phi}(t)$, $\hat{\phi}'(t)$, and $\hat{\psi}_r(t)$, $\hat{\psi}_r'(t)$ are obtained by solving the following $p + 1$ *locally estimating functions* with respect to (γ_0, γ_1) and $(\gamma_{r0}, \gamma_{r1})$ for $r = 1, \ldots, p$,

$$\begin{cases} \sum_{i=1}^{n} K_h(\theta^\tau U_i - t)[\phi_a^2(\theta^\tau U_i)]^{-1}(\tilde{Y}_i - \bar{\bar{Y}}\phi_a(\theta^\tau U_i)) = 0, \\ \sum_{i=1}^{n} K_h(\theta^\tau U_i - t)(\theta^\tau U_i - t)[\phi_a^2(\theta^\tau U_i)]^{-1}(\tilde{Y}_i - \bar{\bar{Y}}\phi_a(\theta^\tau U_i)) = 0, \end{cases} \tag{4}$$

$$\begin{cases} \sum_{i=1}^{n} K_h(\theta^\tau U_i - t)[\psi_{ar}^2(\theta^\tau U_i)]^{-1}(\tilde{X}_{ri} - \bar{\bar{X}}_r\psi_{ar}(\theta^\tau U_i)) = 0, \\ \sum_{i=1}^{n} K_h(\theta^\tau U_i - t)(\theta^\tau U_i - t)[\psi_{ar}^2(\theta^\tau U_i)]^{-1}(\tilde{X}_{ri} - \bar{\bar{X}}_r\psi_{ar}(\theta^\tau U_i)) = 0, \end{cases} \tag{5}$$

where $\bar{\bar{Y}} = \frac{1}{n}\sum_{i=1}^{n} \tilde{Y}_i$ and $\bar{\bar{X}}_r = \frac{1}{n}\sum_{i=1}^{n} \tilde{X}_{ri}$. They are the estimators of the unknown quantities $\boldsymbol{E}Y$ and $\boldsymbol{E}X_r$, respectively. As Şentürk and Müller [19,20] suggested, for response Y and predictors $\mathbf{X}$, the distorting functions satisfy

$$\boldsymbol{E}\phi(\theta^\tau U) = 1, \qquad \boldsymbol{E}\psi(\theta^\tau U) = I_p, \tag{6}$$

where I_p is an $p \times p$ identical matrix. The identifiability condition (6) ensures that the distorting effect vanishes at the population level, namely, $\boldsymbol{E}Y = \boldsymbol{E}\tilde{Y}$ and $\boldsymbol{E}\mathbf{X} = \boldsymbol{E}\tilde{\mathbf{X}}$. Thus, we can estimate the unknown quantities $\boldsymbol{E}Y$ and $\boldsymbol{E}X_r$ by the sample mean of $\{\tilde{Y}_i, \tilde{X}_{1i}, \ldots, \tilde{X}_{pi}\}_{i=1}^n$. Having estimated (γ_0, γ_1), $(\gamma_{r0}, \gamma_{r1})$ at t as $(\hat{\gamma}_0, \hat{\gamma}_1)$, $(\hat{\gamma}_{r0}, \hat{\gamma}_{r1})$ through Eqs. (4) and (5), the local linear estimators of $\phi(t)$, $\phi'(t)$, $\psi_r(t)$, and $\psi_r'(t)$ are $\hat{\phi}(t) = \hat{\gamma}_0$, $\hat{\phi}'(t) = \hat{\gamma}_1$, $\hat{\psi}_r(t) = \hat{\gamma}_{r0}$, and $\hat{\phi}_r'(t) = \hat{\gamma}_{r1}$, respectively.

We now proceed to estimation of $\theta \in \Theta$. If $\phi(\cdot)$ and $\psi_r(\cdot)$ were known, we can formulate quasi-likelihood estimating equations from (2) and (3) for a single index θ as follows.

$$\sum_{i=1}^{n} \left(\frac{\partial\psi_r(\theta^\tau U_i)}{\partial\theta^{(1)}}\right)[\psi_r^2(\theta^\tau U_i)]^{-1}(\tilde{X}_{ri} - \bar{\bar{X}}_r\psi_r(\theta^\tau U_i)) = 0, \quad \text{for } r = 1, \ldots, p, \tag{7}$$

$$\sum_{i=1}^{n} \left(\frac{\partial\phi(\theta^\tau U_i)}{\partial\theta^{(1)}}\right)[\phi^2(\theta^\tau U_i)]^{-1}(\tilde{Y}_i - \bar{\bar{Y}}\phi(\theta^\tau U_i)) = 0. \tag{8}$$

By substituting ψ_r, ϕ and the derivatives by their estimators obtained from (4) and (5), and a direct calculation, we have the estimating equations for θ as follows.

$$\begin{aligned} \hat{\mathbf{\Phi}}_r(\theta^{(1)}) &\triangleq \bar{\bar{X}}_r \sum_{i=1}^{n} J^\tau \hat{\psi}_r'(\theta^\tau U_i)(U_i - \hat{s}(\theta^\tau U_i))[\hat{\psi}_r^2(\theta^\tau U_i)]^{-1}(\tilde{X}_{ri} - \bar{\bar{X}}_r\hat{\psi}_r(\theta^\tau U_i)) \\ &= 0, \quad \text{for } r = 1, \ldots, p, \end{aligned} \tag{9}$$

$$\hat{\mathbf{\Phi}}_{p+1}(\theta^{(1)}) \triangleq \bar{\bar{Y}} \sum_{i=1}^{n} J^\tau \hat{\phi}'(\theta^\tau U_i)(U_i - \hat{s}(\theta^\tau U_i))[\hat{\phi}^2(\theta^\tau U_i)]^{-1}(\tilde{Y}_i - \bar{\bar{Y}}\hat{\phi}(\theta^\tau U_i)) = 0, \tag{10}$$

in which $J = \partial\theta/\partial\theta^{(1)}$ is the Jacobian matrix of size $d \times (d-1)$; that is,

$$J = \begin{pmatrix} -\theta^{(1)\tau}/\sqrt{1 - \|\theta^{(1)}\|^2} \\ I_{d-1} \end{pmatrix}, \tag{11}$$

and $\hat{s}(t)$ is the local linear estimator of $s(t) = \boldsymbol{E}(U|\theta^\tau U = t) = (s_1(t), \ldots, s_d(t))^\tau$, defined as $\hat{s}(t) = \sum_{i=1}^{n} b_i(t)U_i / \sum_{i=1}^{n} b_i(t)$, where $b_i(t) = K_h(\theta^\tau U_i - t)[S_{n,2}(t) - (\theta^\tau U_i - t)S_{n,1}(t)]$, and $S_{n,j} = \sum_{i=1}^{n} K_h(\theta^\tau U_i - t)(\theta^\tau U_i - t)^j$, $j = 1, 2$. The estimation procedure for θ through (4), (5) and (9), (10) is called the estimating function method (EFM) by Cui et al. [5]. It is worth pointing out that the population versions of (9) and (10), $\mathbf{\Phi}_r(\theta^{(1)})$ and $\mathbf{\Phi}_{p+1}(\theta^{(1)})$ [See (A.7) and (A.8) in Appendix A], satisfy the second Bartlett identity; that is, for any θ and $r = 1, \ldots, p+1$,

$$E\mathbf{\Phi}_r(\theta^{(1)})\mathbf{\Phi}_r{}^\tau(\theta^{(1)}) = -E\left\{\frac{\partial\mathbf{\Phi}_r(\theta^{(1)})}{\partial\theta^{(1)}}\right\}. \tag{12}$$

This feature ensures that the proposed estimators of θ are possibly semiparametrically efficient. See [5] for a detailed discussion.

Based on the conclusion drawn by Cui et al. [5], each equation of (9) and (10) can derive a root-n consistent estimator of $\theta^{(1)}$. Thus, we obtain $p+1$ root-n consistent estimators of $\theta^{(1)}$. Denote by $\hat{\theta}^{(1)}[r]$ the solution of the r-th equation $\hat{\mathbf{\Phi}}_r(\hat{\theta}^{(1)}[r]) = 0$. We then define the resulting estimator of $\theta^{(1)}$ as

$$\hat{\theta}^{(1)} = \frac{1}{p+1}\sum_{r=1}^{p+1}\hat{\theta}^{(1)}[r]. \tag{13}$$

Finally, we apply the equation $\theta_1 = \sqrt{1-\|\theta^{(1)}\|^2}$ to estimate θ_1 by

$$\hat{\theta}_1 = \sqrt{1-\|\hat{\theta}^{(1)}\|^2}, \tag{14}$$

and the final estimator of θ is $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}^{(1)})^\tau$.

2.2. Estimation of $\boldsymbol{\beta}$

From the identifiability condition given in (6) and Assumption (A5), we know

$$\boldsymbol{E}\left\{\frac{\tilde{Y}}{\boldsymbol{E}Y}\middle|\theta^\tau U\right\} = \phi(\theta^\tau U), \qquad \boldsymbol{E}\left\{\frac{\tilde{X}_r}{\boldsymbol{E}X_r}\middle|\theta^\tau U\right\} = \psi_r(\theta^\tau U).$$

The local linear estimators of $\phi(\cdot)$ and $\psi_r(\cdot)$ are then obtained by substituting θ with $\hat{\theta}$. That is,

$$\hat{\phi}_b(t) = \frac{\sum_{i=1}^n r_i(t,\hat{\theta})\tilde{Y}_i}{\sum_{i=1}^n r_i(t,\hat{\theta})\bar{\tilde{Y}}}, \qquad \hat{\psi}_{br}(t) = \frac{\sum_{i=1}^n r_i(t,\hat{\theta})\tilde{X}_{ri}}{\sum_{i=1}^n r_i(t,\hat{\theta})\bar{\tilde{X}}_r}, \quad r = 1,\ldots,p, \tag{15}$$

where $r_i(t,\hat{\theta}) = L_{h_1}(\hat{\theta}^\tau U_i - t)[Q_{n,2}(t,\hat{\theta}) - (\hat{\theta}^\tau U_i - t)Q_{n,1}(t,\hat{\theta})]$, $Q_{n,j}(t,\hat{\theta}) = \sum_{i=1}^n L_{h_1}(\hat{\theta}^\tau U_i - t)(\hat{\theta}^\tau U_i - t)^j$ for $j = 1, 2$, $L_{h_1}(\cdot) = h_1^{-1}L(\cdot/h_1)$, with the kernel function $L(\cdot)$ satisfying the conditions in Appendix A, and h_1 being a bandwidth. Thus, the "synthesis" data $\{\hat{Y}_i, \hat{X}_{1i},\ldots,\hat{X}_{pi}\}_{i=1}^n$, after substituting the unobservable response and predictors $\{Y_i, X_{i1},\ldots,X_{ip}\}_{i=1}^n$, can be obtained as

$$\hat{Y}_i = \frac{\tilde{Y}_i}{\hat{\phi}_b(\hat{\theta}^\tau U_i)}, \qquad \hat{X}_{ri} = \frac{\tilde{X}_{ri}}{\hat{\psi}_{br}(\hat{\theta}^\tau U_i)}, \tag{16}$$

for $r = 1,\ldots,p$ and $i = 1,\ldots,n$.

The nonlinear least squares estimators $\hat{\boldsymbol{\beta}}$ are defined as the solution of the q equations

$$\sum_{i=1}^n (\hat{Y}_i - f(\hat{X}_i,\boldsymbol{\beta}))\frac{\partial f(\hat{X}_i,\boldsymbol{\beta})}{\partial \beta_k} = 0, \quad \text{for } k = 1,\ldots,q, \tag{17}$$

where $\partial f(\cdot,\boldsymbol{\beta})/\partial\boldsymbol{\beta}_k$ is the partial derivative of f with respect to $\boldsymbol{\beta}_k$. When (17) has no closed-form solution, one may iteratively solve these equations.

2.3. Large sample properties of the estimators

We now present the asymptotic normality of the estimators $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}^{(1)})^\tau$ and $\hat{\boldsymbol{\beta}}$. We introduce the following notation: $A^{\otimes 2} = AA^\tau$ for any matrix or vector A, and $\breve{U} = U - E(U|\theta^\tau U)$. Without loss of generality, we assume that $\hat{\theta}^{(1)}$ belongs to a $\sqrt{n}$-neighborhood of $\theta^{(1)}$, i.e., $\hat{\theta}^{(1)} \in \{\theta^{(1)'} : \|\theta^{(1)'} - \theta^{(1)}\| \le C_0 n^{-1/2}\}$ for some positive constant C_0. This assumption is feasible because we can find such an $\sqrt{n}$ initial estimator of $\hat{\theta}^{(1)}$ by using existing methods for single-index models. See for example [5,9,8,7,17]. We have the following asymptotic results.

Theorem 1. *Assume that Conditions* (A1)–(A3), (A4)(i), *and* (A5)–(A6) *in* Appendix A *are satisfied. Then* $\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{L} N_d(0, J\boldsymbol{\Sigma}_\theta J^\tau)$, *where* J *is given in* (11) *and*

$$\begin{aligned}\boldsymbol{\Sigma}_\theta = {} & \frac{1}{(p+1)^2}\boldsymbol{E}\left\{\sum_{s=1}^p\sum_{t=1}^p (\boldsymbol{E}X_s\boldsymbol{E}X_t)\Gamma_s^{-1}J^\tau\breve{U}^{\otimes 2}J\Gamma_t^{-1}\frac{\psi_s'(\theta^\tau U)\psi_t'(\theta^\tau U)}{\psi_s(\theta^\tau U)\psi_t(\theta^\tau U)}\mathrm{Cov}(X_s, X_t)\right\} \\ & + \frac{1}{(p+1)^2}\boldsymbol{E}\left\{\sum_{s=1}^p (\boldsymbol{E}Y\boldsymbol{E}X_s)(\Gamma_s^{-1}J^\tau\breve{U}^{\otimes 2}J\Gamma_{p+1}^{-1} + \Gamma_{p+1}^{-1}J^\tau\breve{U}^{\otimes 2}J\Gamma_s^{-1})\right. \\ & \left. \times \frac{\psi_s'(\theta^\tau U)\phi'(\theta^\tau U)}{\psi_s(\theta^\tau U)\phi(\theta^\tau U)}\mathrm{Cov}(X_s, Y)\right\} + \frac{1}{(p+1)^2}\mathrm{Var}(Y)\Gamma_{p+1}^{-1},\end{aligned}$$

with

$$\Gamma_r = (\boldsymbol{E}X_r)^2 \boldsymbol{E}\left(\frac{\psi_r'(\theta^\tau U)}{\psi_r(\theta^\tau U)}\right)^2 J^\tau \breve{U}^{\otimes 2} J \quad \text{for } 1 \le r \le p, \quad \text{and} \quad \Gamma_{p+1} = (\boldsymbol{E}Y)^2 \boldsymbol{E}\left(\frac{\phi'(\theta^\tau U)}{\phi(\theta^\tau U)}\right)^2 J^\tau \breve{U}^{\otimes 2} J. \tag{18}$$

Theorem 2. *Assume that Conditions* (A1)–(A3), (A4)(ii), *and* (A7)–(A10) *hold. When* $\|\hat{\theta} - \theta\| = O_P(n^{-1/2})$, *we have* (i) $\hat{\boldsymbol{\beta}}$ *converge in probability to the true value* $\boldsymbol{\beta}$, *and* (ii) $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{L} N_q(\mathbf{0}, \boldsymbol{\Sigma})$. *Here* $\boldsymbol{\Sigma} = \boldsymbol{\Lambda}^{-1}\boldsymbol{\Omega}\boldsymbol{\Lambda}^{-1}$. *The* (s, t)*th entry of* $\boldsymbol{\Lambda}$ *and* $\boldsymbol{\Omega}$ *equals* $\boldsymbol{\Lambda}(s, t) = \boldsymbol{E}\frac{\partial f(\mathbf{X},\boldsymbol{\beta})}{\partial \boldsymbol{\beta}_s}\frac{\partial f(\mathbf{X},\boldsymbol{\beta})}{\partial \boldsymbol{\beta}_t}$ *and* $\boldsymbol{\Omega}(s, t) = \sigma^2\boldsymbol{\Lambda}(s, t) + \boldsymbol{\Upsilon}(s, t)$, *respectively, where*

$$\begin{aligned}
\boldsymbol{\Upsilon}(s, t) &= \mathrm{Var}\left(\frac{\tilde{Y} - Y}{\boldsymbol{E}Y}\right)\boldsymbol{E}\left(Y\frac{\partial f(\mathbf{X}, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}_s}\right)\boldsymbol{E}\left(Y\frac{\partial f(\mathbf{X}, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}_t}\right) \\
&\quad + \sum_{r=1}^{p}\sum_{l=1}^{p}\mathrm{Cov}\left(\frac{\tilde{X}_r - X_r}{\boldsymbol{E}X_r}, \frac{\tilde{X}_l - X_l}{\boldsymbol{E}X_l}\right)\boldsymbol{E}\left(X_r\frac{\partial f(\mathbf{X}, \boldsymbol{\beta})}{\partial X_r}\frac{\partial f(\mathbf{X}, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}_s}\right)\boldsymbol{E}\left(X_l\frac{\partial f(\mathbf{X}, \boldsymbol{\beta})}{\partial X_l}\frac{\partial f(\mathbf{X}, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}_t}\right) \\
&\quad - \sum_{r=1}^{p}\mathrm{Cov}\left(\frac{\tilde{X}_r - X_r}{\boldsymbol{E}X_r}, \frac{\tilde{Y} - Y}{\boldsymbol{E}Y}\right)\left\{\boldsymbol{E}\left(X_r\frac{\partial f(\mathbf{X}, \boldsymbol{\beta})}{\partial X_r}\frac{\partial f(\mathbf{X}, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}_s}\right)\right. \\
&\quad \left.\times\ \boldsymbol{E}\left(Y\frac{\partial f(\mathbf{X}, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}_t}\right) + \boldsymbol{E}\left(X_r\frac{\partial f(\mathbf{X}, \boldsymbol{\beta})}{\partial X_r}\frac{\partial f(\mathbf{X}, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}_t}\right)\boldsymbol{E}\left(Y\frac{\partial f(\mathbf{X}, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}_s}\right)\right\}.
\end{aligned}$$

Remark 1. In the asymptotic variance $\boldsymbol{\Sigma} = \sigma^2\boldsymbol{\Lambda}^{-1} + \boldsymbol{\Lambda}^{-1}\boldsymbol{\Upsilon}\boldsymbol{\Lambda}^{-1}$, we can observe that the first term $\sigma^2\boldsymbol{\Lambda}^{-1}$ is the usual asymptotic covariance matrix of the nonlinear least squares estimator when the data are observed without distortion, i.e., $\phi(\cdot) = 1$ and $\psi_r(\cdot) = 1$. $\boldsymbol{\Lambda}^{-1}\boldsymbol{\Upsilon}\boldsymbol{\Lambda}^{-1}$ is an extra term due to the distortion in the covariate.

Remark 2. For the linear covariate-adjusted model with a one-dimensional confounding variable proposed by Şentürk and Müller [20], i.e., $f(\mathbf{X}, \boldsymbol{\beta}) = \boldsymbol{\beta}_0 + \sum_{r=1}^{p}\boldsymbol{\beta}_r X_r$ and $\theta \equiv 1$, we estimate the unobserved Y_i and $\{X_{1i}, \ldots, X_{pi}\}_{i=1}^n$ by $\hat{Y}_i = \tilde{Y}_i/\hat{\phi}(U_i)$ and $\hat{X}_{ri} = \tilde{X}_{ri}/\hat{\psi}_r(U_i)$, where $\hat{\phi}(\cdot)$ and $\hat{\psi}_r(\cdot)$ are the local linear estimators of $\hat{\phi}(\cdot)$ and $\hat{\psi}_r(\cdot)$:

$$\hat{\phi}(u) = \frac{\sum_{i=1}^{n} v_i(u)\tilde{Y}_i}{\sum_{i=1}^{n} v_i(u)\bar{\tilde{Y}}}, \qquad \hat{\psi}_r(u) = \frac{\sum_{i=1}^{n} v_i(u)\tilde{X}_{ri}}{\sum_{i=1}^{n} v_i(u)\bar{\tilde{X}}_r}, \quad \text{for } r = 1, \ldots, p,$$

where $v_i(u) = L_{h_1}(U_i - u)[Q_{n,2}(u) - (U_i - u)Q_{n,1}(u)]$ with $Q_{n,j}(u) = \sum_{i=1}^{n} L_{h_1}(U_i - u)(U_i - u)^j$ for $j = 1, 2$.

The estimating Eq. (17) for the linear covariate-adjusted model can be simplified as:

$$\widehat{\mathbf{G}}_n^k(\boldsymbol{\beta}) = \sum_{i=1}^{n}\left(\hat{Y}_i - \sum_{r=0}^{p}\hat{X}_{ri}\boldsymbol{\beta}_r\right)\hat{X}_{ki} = 0, \quad \text{for } k = 0, \ldots, p, \tag{19}$$

in which $\hat{X}_{0i} = X_{0i} = 1$ for $i = 1, \ldots, n$. Thus, the estimating Eq. (19) reduces to a classical linear regression. Denote the solution of the estimating Eq. (19) by $\hat{\boldsymbol{\beta}}_{\mathrm{LS}}$.

Corollary 1. *Under the conditions of Theorem 2, we have* $\sqrt{n}(\hat{\boldsymbol{\beta}}_{LS} - \boldsymbol{\beta}) \xrightarrow{L} N_{p+1}(\mathbf{0}, \boldsymbol{\Sigma}_{LS})$, *where* $\boldsymbol{\Sigma}_{LS} = \sigma^2\boldsymbol{\Lambda}_{LS}^{-1} + \boldsymbol{\Upsilon}_{LS}$ *with the* (s, t)*th elements of* $\boldsymbol{\Lambda}_{LS}$ *and* $\boldsymbol{\Upsilon}_{LS}$ *being* $\boldsymbol{\Lambda}_{LS}(s, t) = \boldsymbol{E}X_sX_t$, *and*

$$\begin{aligned}
\boldsymbol{\Upsilon}_{LS}(s, t) = &\left\{\frac{\boldsymbol{E}Y^2}{(\boldsymbol{E}Y)^2}\mathrm{Var}(\phi(U)) + \frac{\boldsymbol{E}X_sX_t}{\boldsymbol{E}X_s\boldsymbol{E}X_t}\mathrm{Cov}(\psi_s(U), \psi_t(U))\right. \\
&\left. - \frac{\boldsymbol{E}X_sY}{\boldsymbol{E}Y\boldsymbol{E}X_s}\mathrm{Cov}(\phi(U), \psi_s(U)) - \frac{\boldsymbol{E}X_tY}{\boldsymbol{E}Y\boldsymbol{E}X_t}\mathrm{Cov}(\phi(U), \psi_t(U))\right\}\boldsymbol{\beta}_s\boldsymbol{\beta}_t,
\end{aligned}$$

for $0 \le s \le t \le p$, $X_0 = 1$ *and* $\psi_0(\cdot) \equiv 1$.

Corollary 1 indicates that the asymptotic variance of $\sqrt{n}(\hat{\boldsymbol{\beta}}_{\mathrm{LS},r} - \boldsymbol{\beta}_r)$ equals to $\sigma_r^2 = \sigma^2(\boldsymbol{\Lambda}_{\mathrm{LS}}^{-1})(r, r) + \boldsymbol{\Upsilon}_{\mathrm{LS}}(r, r)$ for $0 \le r \le p$. Note that the asymptotic variance of $\hat{\boldsymbol{\beta}}_{\mathrm{LS}}$ proposed by Şentürk and Müller [20] can be expressed as:

$$\breve{\sigma}_r^2 = \sigma^2(\boldsymbol{\Lambda}_{\mathrm{LS}}^{-1})(r, r) + \sigma^2(\boldsymbol{\Lambda}_{\mathrm{LS}}^{-1})(r, r)\mathrm{Var}(\phi(U)) + \boldsymbol{\beta}_r^2\frac{\boldsymbol{E}X_r^2}{(\boldsymbol{E}X_r)^2}\mathrm{Var}(\phi(U) - \psi_r(U)),$$

for $0 \le k \le p$ with $X_0 = 1$, $\psi_0(\cdot) \equiv 1$.

We now compare the asymptotic variance σ_r^2 with $\breve{\sigma}_r^2$. Write $\mathbf{\Lambda}_{\mathrm{LS}} = (\Lambda_{\mathrm{LS},0}, \Lambda_{\mathrm{LS},1}, \ldots, \Lambda_{\mathrm{LS},p})$ with $\Lambda_{\mathrm{LS},k}$ being a $(p+1)$-dimensional column vector. e_r is a $(p+1)$-vector with 1 in the $(r+1)$th position and 0 elsewhere for $r = 0 \sim p$.

Corollary 2. *$\sigma_r^2 \le \breve{\sigma}_r^2$ if and only if $\boldsymbol{\beta}$ lies in the set $\{\boldsymbol{\beta}^\tau \mathbf{D}_r \boldsymbol{\beta} \le 0, 0 \le r \le p\}$, where*

$$\mathbf{D}_r = \boldsymbol{\beta}_r^2 \left\{ \mathbf{\Lambda}_{LS} - \frac{(\Lambda_{LS,r}\Lambda_{LS,0}^\tau + \Lambda_{LS,0}\Lambda_{LS,r}^\tau)}{\boldsymbol{E}X_r} \frac{\mathrm{Cov}(\phi(U), \psi_r(U))}{\mathrm{Var}(\phi(U))} - \frac{\boldsymbol{E}X_r^2}{(\boldsymbol{E}X_r)^2} \Lambda_{LS,0}\Lambda_{LS,0}^\tau \right.$$
$$\left. + \, 2\frac{\boldsymbol{E}X_r^2}{(\boldsymbol{E}X_r)^2} \frac{\mathrm{Cov}(\phi(U), \psi_r(U))}{\mathrm{Var}(\phi(U))} \Lambda_{LS,0}\Lambda_{LS,0}^\tau \right\} + \sigma^2\{e_r e_r^\tau - \mathbf{\Lambda}_{LS}^{-1}(r,r)\Lambda_{LS,0}\Lambda_{LS,0}^\tau\},$$

and the matrices $\mathbf{D}_r$'s are symmetric with at least one negative eigenvalue.

Remark 3. When $f(\mathbf{X}, \boldsymbol{\beta})$ in model (1) is linear as studied by Nguyen and Şentürk [14], i.e., $f(\mathbf{X}, \boldsymbol{\beta}) = \boldsymbol{\beta}_0 + \sum_{r=1}^p X_r \boldsymbol{\beta}_r$, using the arguments similar to Theorem 2 and Corollary 1, we have $\sqrt{n}(\hat{\boldsymbol{\beta}}_{\mathrm{LS}} - \boldsymbol{\beta}) \xrightarrow{L} N_{p+1}(\mathbf{0}, \mathbf{\Sigma}'_{\mathrm{LS}})$, where $\mathbf{\Sigma}'_{\mathrm{LS}} = \sigma^2 \mathbf{\Lambda}_{\mathrm{LS}}^{-1} + \mathbf{\Upsilon}'_{\mathrm{LS}}$. $\mathbf{\Upsilon}'_{\mathrm{LS}}$ is the same as $\mathbf{\Upsilon}_{\mathrm{LS}}$ except U is replaced by $\theta^\tau U$ in each element of $\mathbf{\Upsilon}_{\mathrm{LS}}$.

Its proof is similar to the proofs of Theorem 2 and Corollary 1 and thus is omitted.

2.4. Inference based on empirical likelihood

Based on the covariance matrix given in Theorem 2, one may estimate each of its unknown elements and give a confidence region for $\boldsymbol{\beta}$; i.e., $I_{\alpha,NOR} = \{\boldsymbol{\beta}' : n(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}')^\tau \hat{\mathbf{\Sigma}}^{-1} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}') \le c_\alpha\}$, where $\hat{\mathbf{\Sigma}}$ is a plug-in estimator of $\mathbf{\Sigma}$. Although we can easily confirm that the estimator $\hat{\mathbf{\Sigma}}$ is consistent under mild assumptions, its finite-sample behavior is certainly affected by the need to plug in several estimated terms. Furthermore, the confidence region derived by this procedure is based on a normal approximation, which may not be *precise* in small samples. As an alternative, the empirical likelihood (EL) principle [18,15] is preferable due to its attractive features: improvement of the confidence region, increased accuracy of coverage because of using auxiliary information, easy implementation, avoidance of estimating variances, and studentizing automatically. Therefore in this section, we study inference based on the EL principle.

We introduce an auxiliary random vector $\varpi_{n,i}(\boldsymbol{\beta}') = (\varpi_{n,i}^1(\boldsymbol{\beta}'), \ldots, \varpi_{n,i}^q(\boldsymbol{\beta}'))^\tau$ with

$$\varpi_{n,i}^s(\boldsymbol{\beta}') = (Y_i - f(\mathbf{X}_i, \boldsymbol{\beta}')) \frac{\partial f(\mathbf{X}_i, \boldsymbol{\beta}')}{\partial \boldsymbol{\beta}'_s}.$$

Note that $\boldsymbol{E}\varpi_{n,i}(\boldsymbol{\beta}') = 0$ for $\boldsymbol{\beta}' = \boldsymbol{\beta}$. Then an empirical log-likelihood ratio function is defined as $l_n(\boldsymbol{\beta}') = -2\max \{\sum_{i=1}^n \log(np_i) : p_i \ge 0, \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i \varpi_{n,i}(\boldsymbol{\beta}') = 0\}$. Because the response and predictors are distorted and unobservable, this empirical log-likelihood ratio function cannot be used directly. Instead, we plug $\{\hat{Y}_i, \hat{X}_{i1}, \ldots, \hat{X}_{ip}\}_{i=1}^n$ into $l_n(\boldsymbol{\beta}')$ and an adjusted EL ratio function can be obtained as

$$\hat{l}_n(\boldsymbol{\beta}') = -2\max\left\{\sum_{i=1}^n \log(np_i) : p_i \ge 0, \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i \hat{\varpi}_{n,i}(\boldsymbol{\beta}') = 0\right\}, \tag{20}$$

where $\hat{\varpi}_{n,i}^s(\boldsymbol{\beta}') = (\hat{Y}_i - f(\hat{\mathbf{X}}_i, \boldsymbol{\beta}'))\partial f(\hat{\mathbf{X}}_i, \boldsymbol{\beta}')/\partial \boldsymbol{\beta}'_s$ for $s = 1, \ldots, q$.

By the Lagrange multiplier method, $\hat{l}_n(\boldsymbol{\beta}')$ can be represented as

$$\hat{l}_n(\boldsymbol{\beta}') = 2\sum_{i=1}^n \log\{1 + \lambda^\tau \hat{\varpi}_{n,i}(\boldsymbol{\beta}')\},$$

where λ is determined by

$$\frac{1}{n}\sum_{i=1}^n \frac{\hat{\varpi}_{n,i}(\boldsymbol{\beta}')}{1 + \lambda^\tau \hat{\varpi}_{n,i}(\boldsymbol{\beta}')} = 0.$$

Theorem 3. *Suppose that Conditions* (A1)–(A10) *hold. Then $\hat{l}_n(\boldsymbol{\beta})$ converges to a chi-squared distribution with p degrees of freedom.*

Based on Theorem 3, a confidence region of $\boldsymbol{\beta}$ can be given as $I_{\alpha,EL} = \{\boldsymbol{\beta}' : \hat{l}_n(\boldsymbol{\beta}') \le c_\alpha\}$, where c_α denotes the α quantile of the chi-squared distribution. It is worth mentioning that our EL-based statistic has a standard chi-squared distribution and is free of the infinite-dimensional nuisance parameters $\phi(\cdot)$ and $\psi_r(\cdot)$. For their plug-in estimators, neither bias correction is needed as done by Zhu and Xue [28] for single-index models. This property makes this statistic easy to implement and is computationally efficient.

Table 1
Simulation study. The estimated mean and associated standard error for case 1.

		β_1	β_2	θ_1	θ_2	θ_3
$n = 300$	*Bias*	0.0048	0.0249	0.0269	0.0317	−0.0361
	SE	0.0308	0.0668	0.0968	0.0746	0.0588
$n = 400$	*Bias*	0.0050	0.0227	0.0192	0.0297	−0.0283
	SE	0.0247	0.0554	0.0869	0.0695	0.0575
$n = 500$	*Bias*	0.0039	0.0149	0.0208	0.0282	−0.0260
	SE	0.0241	0.0533	0.0761	0.0646	0.0534
$n = 600$	*Bias*	0.0040	0.0132	0.0154	0.0280	−0.0223
	SE	0.0206	0.0431	0.0695	0.0657	0.0520

Table 2
Simulation study. The estimated mean and associated standard error for case 2.

		β_1	β_2	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6
$n = 300$	*Bias*	−0.0017	−0.0016	−0.0099	0.0194	0.0088	−0.0045	−0.0015	−0.0271
	SE	0.0264	0.0220	0.0939	0.0711	0.0679	0.0564	0.0522	0.0480
$n = 400$	*Bias*	0.0008	−0.0007	−0.0014	0.0126	0.0062	0.0064	−0.0029	−0.0264
	SE	0.0219	0.0186	0.0767	0.0637	0.0582	0.0528	0.0520	0.0473
$n = 500$	*Bias*	−0.0001	−0.0009	−0.0055	0.0124	0.0071	0.0021	−0.0059	−0.0177
	SE	0.0203	0.0176	0.0691	0.0573	0.0552	0.0501	0.0470	0.0424
$n = 600$	*Bias*	0.0040	−0.0004	−0.0011	0.0134	0.0023	0.0039	−0.0082	−0.0126
	SE	0.0175	0.0160	0.0664	0.0531	0.0480	0.0477	0.0424	0.0389

3. Numerical studies

In this section, we conduct a simulation study to assess the performance of the proposed method and report a real data analysis. We choose Epanechnikov kernel function $L(t) = K(t) = 0.75(1 - t^2)_+$ and use the leave-one-out cross-validation to select the optimal bandwidths. To estimate θ, the *fixed point iterative algorithm* proposed by Cui et al. [5] is adopted as it is easy to implement and not sensitive to the initial value of θ. Having the estimators of θ, we calibrate the distorted Y and $\mathbf{X}$ by (16), and then obtain the estimated values $\hat{\boldsymbol{\beta}}$ based on (17).

3.1. A simulation study

We generated 500 datasets consisting of $n = 300$, 400, 500, and 600 observations, respectively, from the model:

$$Y = \sin(\boldsymbol{\beta}_1 X_1) + (2 + X_2)^{\boldsymbol{\beta}_2} + \varepsilon, \tag{21}$$

where $\boldsymbol{\beta}_1 = 1$, $\boldsymbol{\beta}_2 = 0.5$. The model error ε follows $N(0, 0.5^2)$ and the predictors $(X_1, X_2)^\tau$ follow $N_2(\mu_{\mathbf{X}}, \Sigma_{\mathbf{X}})$ with $\mu_{\mathbf{X}} = (2, 2)^\tau$ and

$$\Sigma_{\mathbf{X}} = \begin{pmatrix} 1 & 0.1 \\ 0.1 & 0.25 \end{pmatrix}.$$

The distorting functions are $\phi(\theta^\tau U) = (2 + \theta^\tau U)^2 / C_Y$, $\psi_1(\theta^\tau U) = (1.5 + \theta^\tau U)/C_{X_1}$, and $\psi_2(\theta^\tau U) = (1 + (\theta^\tau U)^2)/C_{X_2}$. The constants C_Y, C_{X_1}, and C_{X_2} in the distorting functions are chosen to ensure identifiability (6). In this simulation example, we took the initial value $\theta_{initial} = (1, 1, \ldots, 1)^\tau / \sqrt{d}$ and stop the iterations when $\max_{1 \le i \le d} |\theta_{new,i} - \theta_{old,i}| \le 0.001$.

*Case*1. The single-index θ was chosen as $(2, 3, 4)/\sqrt{29}$, and U follows $N_3(\mu_U, \Sigma_U)$ with $\mu_U = (4, 5, 6)^\tau$, and

$$\Sigma_U = \begin{pmatrix} 1 & 0.4 & -0.2 \\ 0.4 & 1 & 0.3 \\ -0.2 & 0.3 & 1 \end{pmatrix}.$$

The constants (C_Y, C_{X_1}, C_{X_2}) equal $(116.3869, 10.2277, 78.4761)$. We truncated $\theta^\tau U$ into the interval [0.1576 12.2557] to satisfy Condition (A2); i.e., the distorting functions $\phi(\theta^\tau U)$, $\psi_1(\theta^\tau U)$, and $\psi_2(\theta^\tau U)$ are nonzero in this interval.

*Case*2. The single-index θ was chosen as $(1, 2, 3, 4, 5, 6)/\sqrt{91}$, U follows $N_6(\mu_U, \Sigma_U)$ with $\mu_U = (3, 3, 3, 3, 3, 3)^\tau$, and $\Sigma_U = (\sigma_{ij})$ with $\sigma_{ij} = 0.5^{|i-j|}$. The constants (C_Y, C_{X_1}, C_{X_2}) equal $(76.1915, 8.1042, 46.7747)$. We truncated $\theta^\tau U$ into the interval [0.4323 12.7761] to satisfy Condition (A2).

The bias and the associated standard errors are reported in Tables 1 and 2. It is seen that the estimated values of $(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$ are close to the true value $(1, 0.5)$, and the estimated values of the single-index $\hat{\theta}$ are also close to the true value θ as the sample

Table 3
Simulation study. The coverage probabilities of the confidence regions for $(\beta_1, \beta_2)^\tau$ with nominal level 95%.

Method	$n = 300$	$n = 400$	$n = 500$	$n = 600$
	Case 1			
Empirical likelihood (%)	91.2	93.4	94.2	94.6
Normal approximation (%)	85.8	89.8	92.2	93.9
	Case 2			
Empirical likelihood (%)	92.0	93.1	94.0	94.5
Normal approximation (%)	90.7	90.1	93.5	94.2

size n increases. The coverage probabilities are presented in Table 3, from which we can see that the coverage probabilities based on the EL approach are uniformly closer to the nominal level than those based on the normal approximation approach.

We also conducted one simulation run with a sample size of 400 to give the confidence region of $(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$, based on both the normal approximation and the EL-based approach, and delineate them in Figs. B.1 and B.2. The area based on the EL approach is smaller than the one based on normal approximation. This indicates that the EL approach has a better numerical performance and is superior to the normal approximation one.

3.2. An empirical example

We applied our method to study the Pima Indian diabetes data for an illustration. This dataset has been analyzed by Nguyen and Şentürk [14]. They investigated the relationship between plasma glucose concentration (GLU) and diastolic blood pressure using a linear regression model, and suggested that body mass index and triceps skin-fold thickness are confounding variables. We investigated the relationship between GLU and 2-h serum insulin (SER), which is of particular interest as the normal utilization of glucose can be ruined by abnormal insulin action with high levels of insulin, especially for the patients with diabetes mellitus Type 2. Hans et al. [6] found that there is a significant correlation between glucose concentrations and BMI. Carmina et al. [1] once noticed that SER is significantly correlated with BMI. More recently, Mohamed et al. [13] found that SER is also correlated with triceps skin-fold thickness (SFT). We therefore feel that the BMI and SFT of the body configuration may affect the response, GLU, and the predictor, SER, and therefore treat BMI and SFT as confounding variables in this data analysis.

We removed 14 outliers that include measurements of GLU or SER being zeros and SER measurements being smaller than 30 or larger than 600, which is not possible in practice. We therefore had 380 observations for the data analysis. We chose the initial value $\theta_{initial} = (1/\sqrt{5}, 2/\sqrt{5})$ and stopped iterations if $\max_{1\le i\le d} |\theta_{new,i} - \theta_{old,i}| \le 0.005$. The final estimator is $\hat{\theta} = (0.7579,\ 0.6524)$. Thus, the confounding single-index variable in this dataset is estimated as 0.7579SFT+0.6524BMI. Based on this estimated single-index covariate, the estimates of the distorting function $\phi(\cdot)$ and $\psi(\cdot)$ can be obtained through (15). To see whether the confounding variable has an impact on the response GLU as well as the predictor SER, we presented the patterns of $\hat{\phi}(u)$ and $\hat{\psi}(u)$ in Fig. B.4. Two plots indicate that $\phi(u)$ and $\psi(u)$ are not linear, suggesting the distortion effect of the single index $\theta_1\text{SFT} + \theta_2\text{BMI}$ on GLU and SER.

We used the estimated single-index 0.7579SFT + 0.6524BMI and estimation procedure (16) to obtain estimated values of GLU and SER. These intermediate estimated values are displayed in Fig. B.3, in which we depict the local linear smoothing curve (thin solid line) and the 95% pointwise confidence band. As an illustrative purpose, we also fitted a linear regression for this dataset and display the straight line in Fig. B.3, which is not encapsulated in the band. In what follows, we used the following nonlinear model for this data analysis, which is commonly used to depict the pattern in pharmacokinetic modeling glucose concentration [11]:

$$\text{GLU} = f(\text{SER}, \boldsymbol{\beta}) = (\boldsymbol{\beta}_1 + \boldsymbol{\beta}_2\text{SER})/(\boldsymbol{\beta}_3 + \text{SER}). \tag{22}$$

In the same line as in Section 3.1, we obtain the estimated values of $(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3)$ as $(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3) = (4441.10, 182.02, 96.19)^\tau$. The corresponding 95% asymptotic and EL-based confidence intervals(CIs) of the parameters $(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3)$ are (−2180, 11063), (161.57, 202.47), (20.15, 172.25), and (4000.6, 4908.7), (178.01,186.19), (91.44, 101.03), respectively. The marginal EL-based confidence interval is calculated using (20) by treating the estimated values of the remaining parameters as the true values. The estimated values indicate that the GLU will be stable around 182, with large SER values. It is worth pointing out that the CIs based on the normal approximation method are substantially wider than those based on the empirical likelihood method. The CI of β_1 based on the normal approximation method contains 0, while its EL-based CI excludes 0. This leads to two controversial conclusions. Recalling the performance of the two methods in the simulation experiment, we prefer the conclusion based on the EL procedure. The fitted nonlinear curve along with 95% pointwise confidence intervals, is displayed in Fig. B.5, which properly captures the nonlinear pattern of the GLU.

To assess how well this model captures the curve, we used the test of Stute et al. [23] to check whether the model (22) is adequate or not. The associated value of the test statistic is 0.0497 with a p-value of 0.9790. This indicates that model (22) is appropriate to fit this dataset. We also fitted model (22) for the original data. The estimated values of $(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3)$ equal

$(-7.3082 \times 10^4, 1.3108 \times 10^3, 436.30)^\tau$, which indicates that when the SER becomes large, the GLU will be stable, around 1.3108×10^3. This is not true in practice. Furthermore, the mean of residual square error based on the naive methods is 6.9356×10^3, while the mean of the residual square error based on the proposed method remarkably decreases to 521. Therefore, the confounding variables do have a substantial impact on the improvement of model fitting. To make a comparison, we refit the model using the data after removing zero GLU and SER. The resulting estimated values of $(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3)$ are $(9970.20,\ 192.31,\ 152.83)^\tau$, and the EL-based confidence intervals (CIs) of these parameters are (9385, 10584), (187.42, 197.33), (147.48, 185.30). The mean of the residual square error in this context increases to 534.66. So the extremely large or small SER measurements make the EL-based intervals wider. For instance, the interval length of $\boldsymbol{\beta}_3$ is 37, which is three times longer than that obtained after removal of 18 unusual observations.

Appendix A

In this appendix, we present the conditions and give the proofs of the main results. The necessary lemmas for the following proofs are given in Appendix B.

A.1. Conditions

The following are the regularity conditions for our asymptotic results.

(A1) The density function $f_{\theta^\tau U}(\theta^\tau u)$ of the random variable $\theta^\tau U$ is bounded away from 0 and satisfies the Lipschitz condition of order 1 on $\Omega_\theta = \{\theta^\tau u : u \in \mathcal{U}\}$, and $\mathcal{U}$ is a compact support set of U.

(A2) $\phi(\cdot), \psi_r(\cdot), s(\cdot) = \boldsymbol{E}(U|\theta^\tau U = \cdot)$ have three bounded and continuous derivatives. Moreover, $\phi(\theta^\tau u)$ and $\psi_r(\theta^\tau u)$ are nonzero on Ω_θ.

(A3) The kernel functions $K(\cdot)$ and $L(\cdot)$ are symmetric about zero and have bounded derivatives. Furthermore, $L(\cdot)$ satisfies a Lipschitz condition on R^1, and $\int_{-\infty}^{\infty} u^2 K(u)du \neq 0$, $\int_{-\infty}^{\infty} |u|^j K(u)du < \infty$, $\int_{-\infty}^{\infty} u^2 L(u)du \neq 0$, $\int_{-\infty}^{\infty} |u|^j L(u)du < \infty$, for $j = 1, 2, \ldots$.

(A4) As $n \to \infty$, the bandwidths h and h_1 satisfy:
 (i) $h \to 0$, $nh^4 \to \infty$, and $nh^6 \to 0$.
 (ii) $h_1 \to 0$, $\frac{(\log n)^2}{nh_1^2} \to 0$, $nh_1^4 \to 0$, and $nh_1^2 \to \infty$.

(A5) $\boldsymbol{E}Y$ and $\boldsymbol{E}\,(X_r)$, $r = 1, \ldots, p$ are bounded away from 0.

(A6) Γ_r $(r = 1, \ldots, p+1)$ defined in (18) are positive definite.

(A7) For $l_1, l_2, l_3, l_4 = 0, 1, 2$, $l_1 + l_2 + l_3 + l_4 \leq 3$, $1 \leq s_1, s_2 \leq p$ and $1 \leq t_1, t_2 \leq q$, the partial derivatives

$$\frac{\partial^{l_1+l_2+l_3+l_4} f(\mathbf{X}, \boldsymbol{\beta}')}{\partial^{l_1} \boldsymbol{\beta}'_{t_1} \partial^{l_2} \boldsymbol{\beta}'_{t_2} \partial^{l_3} X_{s_1} \partial^{l_4} X_{s_2}}$$

exist, and

$$\left| \frac{\partial^{l_1+l_2+l_3+l_4} f(\mathbf{X}, \boldsymbol{\beta}')}{\partial^{l_1} \boldsymbol{\beta}'_{t_1} \partial^{l_2} \boldsymbol{\beta}'_{t_2} \partial^{l_3} X_{s_1} \partial^{l_4} X_{s_2}} \right| \leq C, \quad \text{when } l_3 + l_4 \geq 1,$$

for some positive constant C and

$$\boldsymbol{E} \left\{ \sup_{\beta'} \left| \frac{\partial^{l_1+l_2+l_3+l_4} f(\mathbf{X}, \boldsymbol{\beta}')}{\partial^{l_1} \boldsymbol{\beta}'_{t_1} \partial^{l_2} \boldsymbol{\beta}'_{t_2} \partial^{l_3} X_{s_1} \partial^{l_4} X_{s_2}} \right| \right\} < \infty, \quad \text{when } 1 \leq l_1 + l_2 \leq 2, \text{ and } l_3 + l_4 = 0.$$

(A8) $\boldsymbol{E}\varepsilon = 0$ and $\boldsymbol{E}\varepsilon^4 < \infty$, and the covariance matrix of $\mathbf{X}$ is positive and finite.

(A9) $\boldsymbol{\Lambda}$ defined in Theorem 2 is a positive definite matrix with finite elements.

(A10) $\boldsymbol{E}[f(\mathbf{X}, \boldsymbol{\beta}') - f(\mathbf{X}, \boldsymbol{\beta})]^2$ admits one unique minimum at $\boldsymbol{\beta}' = \boldsymbol{\beta}$.

Condition (A1) ensures the density function $f_{\theta^\tau U}(\cdot)$ is positive, which implies that the denominators involved in the nonparametric estimators are bounded away from 0. Condition (A2) is a mild smoothness condition on the involved functions. The absolute values of $\phi(\theta^\tau u)$ and $\psi_r(\theta^\tau u)$ are above zero on the set Ω_θ, which ensures that the denominators involved in the estimating equation of the EFM approach are not equal to zero. Condition (A3) is commonly imposed in nonparametric regression literature. The Gaussian kernel and quadratic kernel satisfy this condition. Condition (A4) is required for asymptotic normality of the estimators $\hat{\theta}$ and $\hat{\boldsymbol{\beta}}$. Condition (A5) is necessary in the study of covariate-adjusted models, see [20,4,27]. Condition (A6) is generally true. Conditions (A7)–(A10) are essential for the asymptotic results of nonlinear least square estimators. See more details in [25].

Throughout the appendix, $Z_n = O_P(a_n)$ means that $a_n^{-1} Z_n$ is bounded in probability. When the variances of the mean-zero random variables Z_n are finite, we can easily show that $Z_n = O_P(\sqrt{\boldsymbol{E}(Z_n^2)})$. This fact will often be used later.

A.2. Proof of Theorem 1

We complete the proof in three steps.

Step 1. We have that

$$\hat{\boldsymbol{\Phi}}_r(\theta^{(1)}) - \bar{\bar{X}}_r \sum_{i=1}^{n} \left(\frac{\partial \hat{\psi}_r(\theta^\tau U_i)}{\partial \theta^{(1)}} \right) [\hat{\psi}_r^2(\theta^\tau U_i)]^{-1} (\tilde{X}_{ri} - \bar{\bar{X}}_r \hat{\psi}_r(\theta^\tau U_i)) = o_P(\sqrt{n}). \tag{A.1}$$

$$\hat{\boldsymbol{\Phi}}_{p+1}(\theta^{(1)}) - \bar{\bar{Y}} \sum_{i=1}^{n} \left(\frac{\partial \hat{\phi}(\theta^\tau U_i)}{\partial \theta^{(1)}} \right) [\hat{\phi}^2(\theta^\tau U_i)]^{-1} (\tilde{Y}_i - \bar{\bar{Y}} \hat{\phi}(\theta^\tau U_i)) = o_P(\sqrt{n}). \tag{A.2}$$

The proofs of (A.1) and (A.2) are similar to the proof of (2.6) in [5]. We omit the details.

Step 2. We prove the following statements.

$$\frac{\partial \hat{\boldsymbol{\Phi}}_r(\theta^{(1)})}{\partial \theta^{(1)}} + \sum_{i=1}^{n} \left(\frac{\partial \hat{\psi}_r(\theta^\tau U_i)}{\partial \theta^{(1)}} \right) [\hat{\psi}_r^2(\theta^\tau U_i)]^{-1} (\bar{\bar{X}}_r)^2 \left(\frac{\partial \hat{\psi}_r(\theta^\tau U_i)}{\partial \theta^{(1)}} \right)^\tau = o_P(n). \tag{A.3}$$

$$\frac{\partial \hat{\boldsymbol{\Phi}}_{p+1}(\theta^{(1)})}{\partial \theta^{(1)}} + \sum_{i=1}^{n} \left(\frac{\partial \hat{\phi}(\theta^\tau U_i)}{\partial \theta^{(1)}} \right) [\hat{\phi}^2(\theta^\tau U_i)]^{-1} (\bar{\bar{Y}})^2 \left(\frac{\partial \hat{\phi}(\theta^\tau U_i)}{\partial \theta^{(1)}} \right)^\tau = o_P(n). \tag{A.4}$$

We only prove (A.4). A direct use of Proposition 1(iii) in [5] and the assumption on the bandwidth (A4)(i) yield

$$\frac{\partial \hat{\boldsymbol{\Phi}}_r(\theta^{(1)})}{\partial \theta^{(1)}} + (\boldsymbol{E}X_r)^2 \sum_{i=1}^{n} \left(\frac{\psi_r'(\theta^\tau U_i)}{\psi_r(\theta^\tau U_i)} \right)^2 J^\tau \breve{U}_i^{\otimes 2} J = o_P(n), \tag{A.5}$$

$$\frac{\partial \hat{\boldsymbol{\Phi}}_{p+1}(\theta^{(1)})}{\partial \theta^{(1)}} + (\boldsymbol{E}Y)^2 \sum_{i=1}^{n} \left(\frac{\phi'(\theta^\tau U_i)}{\phi(\theta^\tau U_i)} \right)^2 J^\tau \breve{U}_i^{\otimes 2} J = o_P(n). \tag{A.6}$$

Furthermore, (A.5) and (A.6) imply that

$$\frac{\partial \hat{\boldsymbol{\Phi}}_r(\theta^{(1)})}{\partial \theta^{(1)}} - \frac{\partial \boldsymbol{\Phi}_r(\theta^{(1)})}{\partial \theta^{(1)}} = o_P(\sqrt{n}), \quad \text{and} \quad \frac{\partial \hat{\boldsymbol{\Phi}}_{p+1}(\theta^{(1)})}{\partial \theta^{(1)}} - \frac{\partial \boldsymbol{\Phi}_{p+1}(\theta^{(1)})}{\partial \theta^{(1)}} = o_P(\sqrt{n}),$$

where

$$\boldsymbol{\Phi}_r(\theta^{(1)}) = \boldsymbol{E}X_r \sum_{i=1}^{n} J^\tau \psi_r'(\theta^\tau U_i) \breve{U}_i [\psi_r^2(\theta^\tau U_i)]^{-1} \{\tilde{X}_{ri} - \boldsymbol{E}X_r \psi_r(\theta^\tau U_i)\}, \tag{A.7}$$

$$\boldsymbol{\Phi}_{p+1}(\theta^{(1)}) = \boldsymbol{E}Y \sum_{i=1}^{n} J^\tau \phi'(\theta^\tau U_i) \breve{U}_i [\phi^2(\theta^\tau U_i)]^{-1} \{\tilde{Y}_i - \boldsymbol{E}Y \phi(\theta^\tau U_i)\}. \tag{A.8}$$

Similar to the derivation of (A.27) in [5], we find that the proof of $\frac{\partial \hat{\boldsymbol{\Phi}}_{p+1}(\theta^{(1)})}{\partial \theta^{(1)}} - \frac{\partial \boldsymbol{\Phi}_{p+1}(\theta^{(1)})}{\partial \theta^{(1)}} = o_P(\sqrt{n})$ or $\frac{\partial \hat{\boldsymbol{\Phi}}_r(\theta)}{\partial \theta^{(1)}} - \frac{\partial \boldsymbol{\Phi}_r(\theta^{(1)})}{\partial \theta^{(1)}} = o_P(\sqrt{n})$ is equivalent to proving that $\hat{\boldsymbol{\Phi}}_{p+1}(\theta^{(1)}) - \boldsymbol{\Phi}_{p+1}(\theta^{(1)}) = o_P(\sqrt{n})$ or $\hat{\boldsymbol{\Phi}}_r(\theta^{(1)}) - \boldsymbol{\Phi}_r(\theta^{(1)}) = o_P(\sqrt{n})$. The desired result can be proved by following the proof of (2.7) in [5].

Step 3. From expressions (A.2) and (A.4), we have

$$\sqrt{n}(\hat{\theta}^{(1)}[p+1] - \theta^{(1)}) = \left[\frac{\partial \hat{\boldsymbol{\Phi}}_{p+1}(\theta^{(1)})}{\partial \theta^{(1)}} \right]^{-1} \sqrt{n} \boldsymbol{\Phi}_{p+1}(\theta^{(1)}) + o_P(1),$$

$$\sqrt{n}(\hat{\theta}^{(1)}[r] - \theta^{(1)}) = \left[\frac{\partial \hat{\boldsymbol{\Phi}}_r(\theta^{(1)})}{\partial \theta^{(1)}} \right]^{-1} \sqrt{n} \boldsymbol{\Phi}_r(\theta^{(1)}) + o_P(1).$$

Then

$$\begin{aligned} \sqrt{n}(\hat{\theta}^{(1)} - \theta^{(1)}) &= \frac{1}{p+1} \sum_{j=1}^{p+1} \left[\frac{1}{n} \frac{\partial \hat{\boldsymbol{\Phi}}_j(\theta^{(1)})}{\partial \theta^{(1)}} \right]^{-1} \frac{1}{\sqrt{n}} \boldsymbol{\Phi}_j(\theta^{(1)}) + o_P(1) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{1}{p+1} \left\{ \sum_{r=1}^{p} \boldsymbol{E}X_r \Gamma_r^{-1} J^\tau \breve{U}_i \frac{\psi_r'(\theta^\tau U_i)}{\psi_r^2(\theta^\tau U_i)} [\tilde{X}_{ri} - \boldsymbol{E}X_r \psi_r(\theta^\tau U_i)] \right. \\ &\quad \left. + \boldsymbol{E}Y \Gamma_{p+1}^{-1} J^\tau \breve{U}_i \frac{\phi'(\theta^\tau U_i)}{\phi^2(\theta^\tau U_i)} [\tilde{Y}_i - \boldsymbol{E}Y \phi(\theta^\tau U_i)] \right\} + o_P(1). \end{aligned}$$

Thus, $\sqrt{n}(\hat{\theta}^{(1)} - \theta^{(1)})$ converges to $N_{d-1}(0, \Sigma_\theta)$ in distribution by a direct calculation. Accordingly the asymptotic normality of $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}^{(1)})^\tau$ follows from these arguments along with the Delta-method. We therefore complete the proof of Theorem 1.

A.3. Proof of Theorem 2

We omit the proof of statement (i), as it is similar to the proof of Theorem 1 in [4] and Lemma 1 in [25].

We now prove the statement of (ii) in the following. By the mean-value theorem to $\widehat{\mathbf{G}}_n(\boldsymbol{\beta})$, we have, for $\boldsymbol{\beta}^*$ lying between $\boldsymbol{\beta}$ and $\hat{\boldsymbol{\beta}}$,

$$0 = \frac{1}{n}\widehat{\mathbf{G}}_n(\hat{\boldsymbol{\beta}}) = \frac{1}{n}\widehat{\mathbf{G}}_n(\boldsymbol{\beta}) + \frac{1}{n}\frac{\partial \widehat{\mathbf{G}}_n(\boldsymbol{\beta}^*)}{\partial \boldsymbol{\beta}}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}).$$

Then

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = \left\{\frac{1}{n}\frac{\partial \widehat{\mathbf{G}}_n(\boldsymbol{\beta}^*)}{\partial \boldsymbol{\beta}}\right\}^{-1} \frac{1}{\sqrt{n}}\widehat{\mathbf{G}}_n(\boldsymbol{\beta}),$$

with the (s, t) element of matrix $\frac{1}{n}\frac{\partial \widehat{\mathbf{G}}_n(\boldsymbol{\beta}^*)}{\partial \boldsymbol{\beta}}$:

$$\frac{1}{n}\frac{\partial \widehat{\mathbf{G}}_n(\boldsymbol{\beta}^*)}{\partial \boldsymbol{\beta}}(s, t) = -\frac{1}{n}\sum_{i=1}^{n}\frac{\partial f(\hat{\mathbf{X}}_i, \boldsymbol{\beta}^*)}{\partial \boldsymbol{\beta}_s}\frac{\partial f(\hat{\mathbf{X}}_i, \boldsymbol{\beta}^*)}{\partial \boldsymbol{\beta}_t} + \frac{1}{n}\sum_{i=1}^{n}(\hat{Y}_i - f(\hat{\mathbf{X}}_i, \boldsymbol{\beta}^*))\frac{\partial^2 f(\hat{\mathbf{X}}_i, \boldsymbol{\beta}^*)}{\partial \boldsymbol{\beta}_s \partial \boldsymbol{\beta}_t}.$$

Similarly as the proof of Lemma B.3, we find that the second term on the right-hand side is $o_P(1)$; that is,

$$\left\{\frac{1}{n}\sum_{i=1}^{n}(\hat{Y}_i - f(\hat{\mathbf{X}}_i, \boldsymbol{\beta}^*))\frac{\partial^2 f(\hat{\mathbf{X}}_i, \boldsymbol{\beta}^*)}{\partial \boldsymbol{\beta}_s \partial \boldsymbol{\beta}_k}\right\}^2 = o_P(1).$$

Now we consider the first term $-\frac{1}{n}\sum_{i=1}^{n}\frac{\partial f(\hat{\mathbf{X}}_i, \boldsymbol{\beta}^*)}{\partial \boldsymbol{\beta}_s}\frac{\partial f(\hat{\mathbf{X}}_i, \boldsymbol{\beta}^*)}{\partial \boldsymbol{\beta}_k}$. Similarly to the proof of Lemma B.3, we also can obtain that

$$-\frac{1}{n}\sum_{i=1}^{n}\frac{\partial f(\hat{\mathbf{X}}_i, \boldsymbol{\beta}^*)}{\partial \boldsymbol{\beta}_s}\frac{\partial f(\hat{\mathbf{X}}_i, \boldsymbol{\beta}^*)}{\partial \boldsymbol{\beta}_k} \xrightarrow{P} -\boldsymbol{E}\frac{\partial f(\mathbf{X}, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}_s}\frac{\partial f(\mathbf{X}, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}_k}.$$

Thus, $\frac{1}{n}\frac{\partial \widehat{\mathbf{G}}_n(\boldsymbol{\beta}^*)}{\partial \boldsymbol{\beta}}(s, t) \xrightarrow{P} \boldsymbol{E}\frac{\partial f(\mathbf{X}, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}_s}\frac{\partial f(\mathbf{X}, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}_k}$. Using Lemma B.3, we have

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = \left\{\frac{1}{n}\frac{\partial \widehat{\mathbf{G}}_n(\boldsymbol{\beta}^*)}{\partial \boldsymbol{\beta}}\right\}^{-1} \frac{1}{\sqrt{n}}\widehat{\mathbf{G}}_n(\boldsymbol{\beta}) = \boldsymbol{\Lambda}^{-1}\frac{1}{\sqrt{n}}\mathbf{F}_n(\boldsymbol{\beta}) + o_P(1) \xrightarrow{L} N_q(\mathbf{0}, \boldsymbol{\Sigma}).$$

A.4. Proof of Corollary 1

Write $\mathbf{X}^{(n)} = (X_0^{(n)}, X_1^{(n)}, \ldots, X_p^{(n)})^\tau$ with $X_0^{(n)} = (1, 1, \ldots, 1)^\tau$ and $X_s^{(n)} = (X_{s1}, \ldots, X_{sn})^\tau$ for $1 \le s \le p$, $Y = (Y_1, \ldots, Y_n)^\tau$ and $\hat{\mathbf{X}}^{(n)}$ and $\hat{Y}$ the estimators of unobservable $\mathbf{X}^{(n)}$ and Y, i.e., $\hat{\mathbf{X}}^{(n)} = (\hat{X}_0^{(n)}, \hat{X}_1^{(n)}, \ldots, \hat{X}_p^{(n)})$ with $\hat{X}_0^{(n)} = (1, 1, \ldots, 1)^\tau$ and $\hat{X}_s^{(n)} = (\hat{X}_{s1}, \ldots, \hat{X}_{sn})^\tau$, $\hat{Y} = (\hat{Y}_1, \ldots, \hat{Y}_n)^\tau$. Thus, the estimator $\hat{\boldsymbol{\beta}}_{\mathrm{LS}}$ is the LS one, that is, $\hat{\boldsymbol{\beta}}_{\mathrm{LS}} = \{\mathbf{X}^{(n)\tau}\mathbf{X}^{(n)}\}^{-1}\mathbf{X}^{(n)\tau}\hat{Y}$. We represent $\hat{\boldsymbol{\beta}}_{\mathrm{LS}}$ as $\{\mathbf{X}^{(n)\tau}\mathbf{X}^{(n)} + \Delta_{\mathbf{X}^{(n)}}\}^{-1}\{\mathbf{X}^{(n)\tau}Y + \Delta_{(\mathbf{X}^{(n)}, Y)}\}$, where $\Delta_{(\mathbf{X}^{(n)}, Y)} = \mathbf{X}^{(n)\tau}(\hat{Y} - Y) + (\hat{\mathbf{X}}^{(n)} - \mathbf{X}^{(n)})^\tau Y + (\hat{\mathbf{X}}^{(n)} - \mathbf{X}^{(n)})^\tau(\hat{Y} - Y)$ and $\Delta_{\mathbf{X}^{(n)}} = \mathbf{X}^{(n)\tau}(\hat{\mathbf{X}}^{(n)} - \mathbf{X}^{(n)}) + (\hat{\mathbf{X}}^{(n)} - \mathbf{X}^{(n)})^\tau\mathbf{X}^{(n)} + (\hat{\mathbf{X}}^{(n)} - \mathbf{X}^{(n)})^\tau(\hat{\mathbf{X}}^{(n)} - \mathbf{X}^{(n)})$.

With a similar analysis to the proof of Lemma B.1, we can show that $\sup_{u\in\mathcal{U}}|\hat{\phi}_l(u) - \phi_l(u)| = O_P(\frac{\log n}{nh_1})^{1/2}$ and $\sup_{u\in\mathcal{U}}|\hat{\psi}_{lr}(u) - \psi_r(u)| = O_P(\frac{\log n}{nh_1})^{1/2}$. Thus, by Condition (A8), we obtain that

$$\Delta_{(\mathbf{X}^{(n)}, Y)} = \mathbf{X}^{(n)\tau}(\hat{Y} - Y) + (\hat{\mathbf{X}}^{(n)} - \mathbf{X}^{(n)})^\tau Y + O_P\left(\frac{\log n}{nh_1}\right),$$

$$\Delta_{\mathbf{X}^{(n)}} = \mathbf{X}^{(n)\tau}(\hat{\mathbf{X}}^{(n)} - \mathbf{X}^{(n)}) + (\hat{\mathbf{X}}^{(n)} - \mathbf{X}^{(n)})^\tau\mathbf{X}^{(n)} + O_P\left(\frac{\log n}{nh_1}\right),$$

$$\{\mathbf{X}^{(n)\tau}\mathbf{X}^{(n)} + \Delta_{\mathbf{X}^{(n)}}\}^{-1} - \{\mathbf{X}^{(n)\tau}\mathbf{X}^{(n)}\}^{-1} = -\{\mathbf{X}^{(n)\tau}\mathbf{X}^{(n)}\}^{-1}\Delta_{\mathbf{X}^{(n)}}\{\mathbf{X}^{(n)\tau}\mathbf{X}^{(n)}\}^{-1} + O_P\left(\frac{\log n}{nh_1}\right).$$

As a consequence, we have

$$\left\{\frac{1}{n}\mathbf{X}^{(n)\tau}\mathbf{X}^{(n)}\right\}\sqrt{n}(\hat{\boldsymbol\beta}_{\mathrm{LS}}-\boldsymbol\beta) = \left\{\frac{1}{\sqrt{n}}\mathbf{X}^{(n)\tau}\varepsilon - \frac{1}{\sqrt{n}}\mathbf{X}^{(n)\tau}(\hat{\mathbf{X}}^{(n)}-\mathbf{X}^{(n)})\boldsymbol\beta - \frac{1}{\sqrt{n}}(\hat{\mathbf{X}}^{(n)}-\mathbf{X}^{(n)})^\tau\mathbf{X}^{(n)}\boldsymbol\beta \right.$$
$$\left. + \frac{1}{\sqrt{n}}\mathbf{X}^{(n)\tau}(\hat{Y}-Y) + \frac{1}{\sqrt{n}}(\hat{\mathbf{X}}^{(n)}-\mathbf{X}^{(n)})^\tau Y\right\} + O_P\left(\frac{1}{n}\mathbf{X}^{(n)\tau}\varepsilon\right) + O_P\left(\sqrt{\frac{\log^2 n}{nh_1^2}}\right).$$

Using Lemma B.2, we know that the s-th element of $\frac{1}{\sqrt{n}}\mathbf{X}^{(n)\tau}(\hat{\mathbf{X}}^{(n)}-\mathbf{X}^{(n)})\boldsymbol\beta$ has the following asymptotic expression:

$$\left[\frac{1}{\sqrt{n}}\mathbf{X}^{(n)\tau}(\hat{\mathbf{X}}^{(n)}-\mathbf{X}^{(n)})\boldsymbol\beta\right]_s = \sum_{r=1}^{p}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\tilde{X}_{ri}-X_{ri})\frac{\boldsymbol{E}(X_rX_s)}{\boldsymbol{E}X_r}\boldsymbol\beta_r + o_P(1).$$

Similarly,

$$\left[\frac{1}{\sqrt{n}}(\hat{\mathbf{X}}^{(n)}-\mathbf{X}^{(n)})^\tau\mathbf{X}^{(n)}\boldsymbol\beta\right]_s = \sum_{r=0}^{p}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\tilde{X}_{si}-X_{si})\frac{\boldsymbol{E}(X_rX_s)}{\boldsymbol{E}X_s}\boldsymbol\beta_r + o_P(1),$$

and

$$\left[\frac{1}{\sqrt{n}}\mathbf{X}^{(n)\tau}(\hat{Y}-Y)\right]_s = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\tilde{Y}_i-Y_i)\frac{\boldsymbol{E}(YX_s)}{\boldsymbol{E}Y} + o_P(1).$$

Furthermore, $\frac{1}{\sqrt{n}}(\hat{\mathbf{X}}_n-\mathbf{X}_n)^\tau Y = \frac{1}{\sqrt{n}}(\hat{\mathbf{X}}_n-\mathbf{X}_n)^\tau\mathbf{X}_n\boldsymbol\beta + \frac{1}{\sqrt{n}}(\hat{\mathbf{X}}_n-\mathbf{X}_n)^\tau\varepsilon$. With a similar analysis to Lemma B.2, we can obtain that $\frac{1}{\sqrt{n}}(\hat{\mathbf{X}}_n-\mathbf{X}_n)^\tau\varepsilon = o_P(1)$. Noting that $\boldsymbol{E}\mathbf{X}^\tau Y = \boldsymbol{E}\mathbf{X}^\tau\mathbf{X}\boldsymbol\beta$, we then have

$$\left[\frac{1}{\sqrt{n}}(\hat{\mathbf{X}}^{(n)}-\mathbf{X}^{(n)})^\tau Y\right]_s = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\tilde{X}_{si}-X_{si})\frac{\boldsymbol{E}(YX_s)}{\boldsymbol{E}X_s} + o_P(1).$$

Note that $X_{0i}=1$ for $i=1,\ldots,n$. Thus,

$$\left\{\frac{1}{n}\mathbf{X}^{(n)\tau}\mathbf{X}^{(n)}\right\}\sqrt{n}(\hat{\boldsymbol\beta}_{\mathrm{LS}}-\boldsymbol\beta) = \frac{1}{\sqrt{n}}\{\mathbf{X}^{(n)\tau}\varepsilon - \mathbf{X}^{(n)\tau}(\hat{\mathbf{X}}^{(n)}-\mathbf{X}^{(n)})\boldsymbol\beta + \mathbf{X}^{(n)\tau}(\hat{Y}-Y)\} + o_P(1)$$
$$= \begin{pmatrix} \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\varepsilon_iX_{0i} + (\tilde{Y}_i-Y_i)\frac{\boldsymbol{E}(YX_0)}{\boldsymbol{E}Y} - \sum_{r=1}^{p}(\tilde{X}_{ri}-X_{ri})\frac{\boldsymbol{E}(X_rX_0)}{\boldsymbol{E}X_r}\boldsymbol\beta_r\right\} \\ \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\varepsilon_iX_{1i} + (\tilde{Y}_i-Y_i)\frac{\boldsymbol{E}(YX_1)}{\boldsymbol{E}Y} - \sum_{r=1}^{p}(\tilde{X}_{ri}-X_{ri})\frac{\boldsymbol{E}(X_rX_1)}{\boldsymbol{E}X_r}\boldsymbol\beta_r\right\} \\ \vdots \\ \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\varepsilon_iX_{pi} + (\tilde{Y}_i-Y_i)\frac{\boldsymbol{E}(YX_p)}{\boldsymbol{E}Y} - \sum_{r=1}^{p}(\tilde{X}_{ri}-X_{ri})\frac{\boldsymbol{E}(X_rX_p)}{\boldsymbol{E}X_r}\boldsymbol\beta_r\right\} \end{pmatrix} + o_P(1).$$

Recall that $\boldsymbol\Lambda_{\mathrm{LS}}(s,t) = \boldsymbol{E}X_sX_t$ for $0\le s,\ t\le p$ ($X_0=1$). Thus, $\frac{1}{n}\mathbf{X}^{(n)\tau}\mathbf{X}^{(n)} \xrightarrow{a.s.} \boldsymbol\Lambda_{\mathrm{LS}}$. It is easy to show $\sqrt{n}(\hat{\boldsymbol\beta}_{\mathrm{LS}}-\boldsymbol\beta) \xrightarrow{L} N_{p+1}(\mathbf{0},\boldsymbol\Sigma_{\mathrm{LS}})$, where $\boldsymbol\Sigma_{\mathrm{LS}} = \boldsymbol\Lambda_{\mathrm{LS}}^{-1}\boldsymbol\Omega_{\mathrm{LS}}\boldsymbol\Lambda_{\mathrm{LS}}^{-1}$ with $\boldsymbol\Omega_{\mathrm{LS}} = \sigma^2\boldsymbol\Lambda_{\mathrm{LS}}+\boldsymbol\Omega'_{\mathrm{LS}}$. Note that $Y=\sum_{r=0}^{p}X_r\boldsymbol\beta_r+\varepsilon$. Therefore, $\boldsymbol\Omega'_{\mathrm{LS}} = \boldsymbol\Lambda_{\mathrm{LS}}\boldsymbol\Upsilon_{\mathrm{LS}}\boldsymbol\Lambda_{\mathrm{LS}}$. Setting $X_0=1$ and $\psi_0(\cdot)\equiv 1$, we complete the proof of Corollary 1.

A.5. Proof of Corollary 2

Recall that $\boldsymbol\Lambda_{\mathrm{LS}} = (\Lambda_{\mathrm{LS},0},\Lambda_{\mathrm{LS},1},\ldots,\Lambda_{\mathrm{LS},p})$. Thus, we have the following expressions:

$$\boldsymbol{E}Y^2 = \boldsymbol\beta^\tau\boldsymbol\Lambda_{\mathrm{LS}}\boldsymbol\beta + \sigma^2, \qquad \boldsymbol{E}Y = \Lambda_{\mathrm{LS},0}^\tau\boldsymbol\beta = \boldsymbol\beta^\tau\Lambda_{\mathrm{LS},0} \tag{A.9}$$

$$\boldsymbol{E}YX_r = \Lambda_{\mathrm{LS},r}^\tau\boldsymbol\beta = \boldsymbol\beta^\tau\Lambda_{\mathrm{LS},r}, \qquad \boldsymbol{E}YX_r\boldsymbol{E}Y = \boldsymbol\beta^\tau\Lambda_{\mathrm{LS},0}\Lambda_{\mathrm{LS},r}^\tau\boldsymbol\beta = \boldsymbol\beta^\tau\Lambda_{\mathrm{LS},r}\Lambda_{\mathrm{LS},0}^\tau\boldsymbol\beta. \tag{A.10}$$

It is seen that $\sigma_r^2 \le \breve{\sigma}_r^2$ if and only if $\boldsymbol\Upsilon_{\mathrm{LS}}(r,r) \le \sigma^2\boldsymbol\Lambda_{\mathrm{LS}}^{-1}(r,r)\mathrm{Var}(\phi(U)) + \boldsymbol\beta_r^2\frac{\boldsymbol{E}X_r^2}{(\boldsymbol{E}X_r)^2}\mathrm{Var}(\phi(U)-\psi_r(U))$. This is equivalent to the following inequality:

$$\boldsymbol{E}Y^2\boldsymbol\beta_r^2 - \frac{2\boldsymbol{E}X_rY\boldsymbol{E}Y}{\boldsymbol{E}X_r}\frac{\mathrm{Cov}(\phi(U),\psi_r(U))}{\mathrm{Var}(\phi(U))}\boldsymbol\beta_r^2$$
$$\le \left\{\sigma^2\boldsymbol\Lambda_{\mathrm{LS}}^{-1}(r,r) + \boldsymbol\beta_r^2\frac{\boldsymbol{E}X_r^2}{(\boldsymbol{E}X_r)^2} - 2\boldsymbol\beta_r^2\frac{\boldsymbol{E}X_r^2}{(\boldsymbol{E}X_r)^2}\frac{\mathrm{Cov}(\phi(U),\psi_r(U))}{\mathrm{Var}(\phi(U))}\right\}(\boldsymbol{E}Y)^2.$$

Plugging (A.9) and (A.10) into this inequality, we have

$$\boldsymbol{\beta}^\tau\{\boldsymbol{\Lambda}_{\mathrm{LS}}\boldsymbol{\beta}_r^2+\sigma^2 e_r e_r^\tau\}\boldsymbol{\beta}-\boldsymbol{\beta}^\tau\left\{\frac{(\Lambda_{\mathrm{LS},r}\Lambda_{\mathrm{LS},0}^\tau+\Lambda_{\mathrm{LS},0}\Lambda_{\mathrm{LS},r}^\tau)}{\boldsymbol{E}X_r}\frac{\mathrm{Cov}(\phi(U),\psi_r(U))}{\mathrm{Var}(\phi(U))}\boldsymbol{\beta}_r^2\right\}\boldsymbol{\beta}$$
$$\le\boldsymbol{\beta}^\tau\left\{\sigma^2\boldsymbol{\Lambda}_{\mathrm{LS}}^{-1}(r,r)+\boldsymbol{\beta}_r^2\frac{\boldsymbol{E}X_r^2}{(\boldsymbol{E}X_r)^2}-2\boldsymbol{\beta}_r^2\frac{\boldsymbol{E}X_r^2}{(\boldsymbol{E}X_r)^2}\frac{\mathrm{Cov}(\phi(U),\psi_r(U))}{\mathrm{Var}(\phi(U))}\right\}\Lambda_{\mathrm{LS},0}\Lambda_{\mathrm{LS},0}^\tau\boldsymbol{\beta}.$$

Then $\sigma_r^2\le\breve{\sigma}_r^2$ if and only if $\boldsymbol{\beta}^\tau\mathbf{D}_r\boldsymbol{\beta}\le 0$.

Next, we prove the second assertion of Corollary 2.

- If $\boldsymbol{\beta}_r=0$, then $\mathbf{D}_r=\sigma^2 e_r e_r^\tau-\sigma^2\boldsymbol{\Lambda}_{\mathrm{LS}}^{-1}(r,r)\Lambda_{\mathrm{LS},0}\Lambda_{\mathrm{LS},0}^\tau$. For any symmetric $q\times q$ matrix A, the maximum eigenvalue and minimum eigenvalue of A have the following inequality: $\lambda_{\min}(A)\le A(s,s)\le\lambda_{\max}(A)$ with $1\le s\le q$. Note that e_r is a $(p+1)$-vector with 1 in the $(r+1)$th position and 0 elsewhere for $r=0\sim p$. Then,

$$\lambda_{\min}(\mathbf{D}_r)\le\mathbf{D}_r(1,1)=\sigma^2(e_r e_r^\tau)(1,1)-\sigma^2\boldsymbol{\Lambda}_{\mathrm{LS}}^{-1}(r,r)(\Lambda_{\mathrm{LS},0}\Lambda_{\mathrm{LS},0}^\tau)(1,1)$$
$$=\sigma^2(I(r=0)-\boldsymbol{\Lambda}_{\mathrm{LS}}^{-1}(r,r)).$$

When $r=0$, we have $\boldsymbol{\Lambda}_{\mathrm{LS}}^{-1}(0,0)=1/\{1-(\boldsymbol{E}\mathbf{X})\mathbf{M}_X^{-1}(\boldsymbol{E}\mathbf{X})^\tau\}$, where $\mathbf{M}_X=\boldsymbol{E}(\mathbf{X}\mathbf{X}^\tau)$, which is a positive matrix by Assumption (A8). Since $\boldsymbol{\Lambda}_{\mathrm{LS}}^{-1}(0,0)$ is the asymptotic variance of $\hat{\boldsymbol{\beta}}_0$ for the linear regression model $f(\mathbf{X},\boldsymbol{\beta})=\boldsymbol{\beta}_0+\sum_{r=1}^p\boldsymbol{\beta}_r X_r$. By the assumptions of Corollary 1, $\boldsymbol{\Lambda}_{\mathrm{LS}}$ is a positive definite matrix. Thus, $\boldsymbol{\Lambda}_{\mathrm{LS}}^{-1}(0,0)\ge\lambda_{\min}(\boldsymbol{\Lambda}_{\mathrm{LS}}^{-1})=1/\lambda_{\max}(\boldsymbol{\Lambda}_{\mathrm{LS}})>0$; i.e., $\boldsymbol{\Lambda}_{\mathrm{LS}}^{-1}(0,0)$ is a positive constant. Then

$$\lambda_{\min}(\mathbf{D}_0)\le-\frac{\sigma^2(\boldsymbol{E}\mathbf{X})\mathbf{M}_X^{-1}(\boldsymbol{E}\mathbf{X})^\tau}{1-(\boldsymbol{E}\mathbf{X})\mathbf{M}_X^{-1}(\boldsymbol{E}\mathbf{X})^\tau}\le 0.$$

When $1\le r\le p$, we know $\lambda_{\min}(\mathbf{D}_r)\le-\sigma^2\boldsymbol{\Lambda}_{\mathrm{LS}}^{-1}(r,r)\le 0$ due to the fact that $\boldsymbol{\Lambda}_{\mathrm{LS}}^{-1}(r,r)\ge\lambda_{\min}(\boldsymbol{\Lambda}_{\mathrm{LS}}^{-1})=1/\lambda_{\max}(\boldsymbol{\Lambda}_{\mathrm{LS}})>0$.

- If $\boldsymbol{\beta}_r\ne 0$, then we have

$$\frac{\lambda_{\min}(\mathbf{D}_r)}{\boldsymbol{\beta}_r^2}\le 1+\frac{\sigma^2 I(r=0)}{\boldsymbol{\beta}_r^2}-2\frac{\mathrm{Cov}(\phi(U),\psi_r(U))}{\mathrm{Var}(\phi(U))}I(r\ne 0)-\frac{\sigma^2\boldsymbol{\Lambda}_{\mathrm{LS}}^{-1}(r,r)}{\boldsymbol{\beta}_r^2}$$
$$-\frac{\boldsymbol{E}X_r^2}{(\boldsymbol{E}X_r)^2}+2\frac{\boldsymbol{E}X_r^2}{(\boldsymbol{E}X_r)^2}\frac{\mathrm{Cov}(\phi(U),\psi_r(U))}{\mathrm{Var}(\phi(U))}I(r\ne 0).$$

When $r=0$, $X_r=1$ and $\boldsymbol{E}X_r^2=(\boldsymbol{E}X_r)^2$. Then

$$\frac{\lambda_{\min}(\mathbf{D}_0)}{\boldsymbol{\beta}_r^2}\le\frac{\sigma^2}{\boldsymbol{\beta}_0^2}-\frac{\sigma^2\boldsymbol{\Lambda}_{\mathrm{LS}}^{-1}(0,0)}{\boldsymbol{\beta}_0^2}=-\frac{\sigma^2(\boldsymbol{E}\mathbf{X})\mathbf{M}_X^{-1}(\boldsymbol{E}\mathbf{X})^\tau}{\boldsymbol{\beta}_0^2(1-(\boldsymbol{E}\mathbf{X})\mathbf{M}_X^{-1}(\boldsymbol{E}\mathbf{X})^\tau)}\le 0.$$

When $1\le r\le p$, we have

$$\frac{\lambda_{\min}(\mathbf{D}_r)}{\boldsymbol{\beta}_r^2}\le 1-2\frac{\mathrm{Cov}(\phi(U),\psi_r(U))}{\mathrm{Var}(\phi(U))}-\frac{\sigma^2\boldsymbol{\Lambda}_{\mathrm{LS}}^{-1}(r,r)}{\boldsymbol{\beta}_r^2}-\frac{\boldsymbol{E}X_r^2}{(\boldsymbol{E}X_r)^2}+2\frac{\boldsymbol{E}X_r^2}{(\boldsymbol{E}X_r)^2}\frac{\mathrm{Cov}(\phi(U),\psi_r(U))}{\mathrm{Var}(\phi(U))}$$
$$=-\frac{\sigma^2\boldsymbol{\Lambda}_{\mathrm{LS}}^{-1}(r,r)}{\boldsymbol{\beta}_r^2}+\frac{\mathrm{Var}(X_r)}{(\boldsymbol{E}X_r)^2}\left\{\frac{\boldsymbol{E}\phi(U)\psi_r(U)-\boldsymbol{E}\phi(U)^2-1}{\mathrm{Var}(\phi(U))}\right\}.$$

We find that if the distorting functions satisfy $\boldsymbol{E}\phi(U)\psi_r(U)\le 1+\boldsymbol{E}\phi(U)^2$, then $\lambda_{\min}(\mathbf{D}_r)\le 0$, which entails the region of $\boldsymbol{\beta}$ confined by the conditions of Corollary 2 is not empty, and thus we complete the proof of Corollary 2.

A.6. Proof of Theorem 3

For $1\le s\le q$, decompose $\hat{\varpi}_{n,i}(\boldsymbol{\beta})_s$ into the following terms:

$$\hat{\varpi}_{n,i}^s(\boldsymbol{\beta})=(Y_i-f(\mathbf{X}_i,\boldsymbol{\beta}))\frac{\partial f(\mathbf{X}_i,\boldsymbol{\beta})}{\partial\boldsymbol{\beta}_s}+E_{n,i1}^s+E_{n,i2}^s+E_{n,i3}^s+E_{n,i4}^s+E_{n,i5}^s,$$

where

$$E_{n,i1}^s=(\hat{Y}_i-Y_i)\frac{\partial f(\mathbf{X}_i,\boldsymbol{\beta})}{\partial\boldsymbol{\beta}_s},$$

$$E_{n,i2}^s=(Y_i-f(\mathbf{X}_i,\boldsymbol{\beta}))\left(\frac{\partial f(\hat{\mathbf{X}}_i,\boldsymbol{\beta})}{\partial\boldsymbol{\beta}_s}-\frac{\partial f(\mathbf{X}_i,\boldsymbol{\beta})}{\partial\boldsymbol{\beta}_s}\right),$$

$$E^s_{n,i3} = (f(\mathbf{X}_i, \boldsymbol{\beta}) - f(\hat{\mathbf{X}}_i, \boldsymbol{\beta}))\frac{\partial f(\mathbf{X}_i, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}_s},$$

$$E^s_{n,i4} = (\hat{Y}_i - Y_i)\left(\frac{\partial f(\hat{\mathbf{X}}_i, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}_s} - \frac{\partial f(\mathbf{X}_i, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}_s}\right),$$

$$E^s_{n,i5} = (f(\mathbf{X}_i, \boldsymbol{\beta}) - f(\hat{\mathbf{X}}_i, \boldsymbol{\beta}))\left(\frac{\partial f(\hat{\mathbf{X}}_i, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}_s} - \frac{\partial f(\mathbf{X}_i, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}_s}\right).$$

To prove Theorem 3, we need to show that

$$\max_{1\le i\le n} |\hat{\varpi}^s_{n,i}(\boldsymbol{\beta})| = o_P(n^{1/2}).$$

First, we consider the argument $\max_{1\le i\le n} |(Y_i - f(\mathbf{X}_i, \boldsymbol{\beta}))\frac{\partial f(\mathbf{X}_i,\boldsymbol{\beta})}{\partial \boldsymbol{\beta}_s}|$. For any sequence of $i.i.d.$ random variables $\{V_i, 1 \le i \le n\}$ and $\boldsymbol{E}V^2 \le \infty$, we have $\max_{1\le i\le n} \frac{|V_i|}{\sqrt{n}} \to 0$, a.s. Together with Conditions (A7) and (A8), we have

$$\max_{1\le i\le n} \left|(Y_i - f(\mathbf{X}_i, \boldsymbol{\beta}))\frac{\partial f(\mathbf{X}_i, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}_s}\right| = o_P(n^{1/2}).$$

Next, for $E^s_{n,i1}$, directly using Lemma B.1 and Condition (A2), we have

$$|E^s_{n,i1}| = \frac{|\phi(\theta^\tau U_i) - \hat{\phi}_b(\hat{\theta}^\tau U_i, \hat{\theta})|}{|\hat{\phi}_b(\hat{\theta}^\tau U_i, \hat{\theta})|}\left|Y_i \frac{\partial f(\mathbf{X}_i, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}_s}\right| \le C_1 O_P\left(\sqrt{\frac{\log n}{nh_1}}\right)\left|Y_i \frac{\partial f(\mathbf{X}_i, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}_s}\right|$$

for some positive constant C_1. Conditions (A7) and (A8) entail that $\boldsymbol{E}|Y\frac{\partial f(\mathbf{X},\boldsymbol{\beta})}{\partial \boldsymbol{\beta}_s}|^2 < \infty$. Thus, it is easily seen that if Condition (A4)(ii) holds we have $\max_{1\le i\le n} |E^s_{n,i1}| = o_P(n^{1/2})$.

Similar to the proof of Lemma B.3, we apply Taylor expansion to $\frac{\partial f(\hat{\mathbf{X}}_i,\boldsymbol{\beta})}{\partial \boldsymbol{\beta}_s} - \frac{\partial f(\mathbf{X}_i,\boldsymbol{\beta})}{\partial \boldsymbol{\beta}_s}$ with respect to $\mathbf{X}_i$. Using Lemma B.1 again, we obtain $\max_{1\le i\le n} |E^s_{n,ij}| = o_P(n^{1/2})$ for $j = 2, 3, 4, 5$. Next, following the same argument for (2.14) as [16] and Lemma B.2 entails $\lambda = O_P(n^{1/2})$. Thus, $\max_{1\le i\le n} |\lambda^\tau \hat{\varpi}^s_{n,i}(\boldsymbol{\beta})| = o_P(1)$.

Note that $\log(1+t) \doteq t - \frac{1}{2}t^2$ for sufficiently small t, we have

$$\hat{l}(\boldsymbol{\beta}) = 2\sum_{i=1}^n \left(\lambda^\tau \hat{\varpi}^s_{n,i}(\boldsymbol{\beta}) - \frac{1}{2}\{\lambda^\tau \hat{\varpi}^s_{n,i}(\boldsymbol{\beta})\}^2\right) + o_P(1). \tag{A.11}$$

Due to the fact that λ satisfies the following equation,

$$\frac{1}{n}\sum_{i=1}^n \frac{\hat{\varpi}_{n,i}(\boldsymbol{\beta})}{1 + \lambda^\tau \hat{\varpi}_{n,i}(\boldsymbol{\beta})} = 0.$$

Furthermore,

$$\begin{aligned} 0 &= \frac{1}{n}\sum_{i=1}^n \frac{\hat{\varpi}_{n,i}(\boldsymbol{\beta})}{1 + \lambda^\tau \hat{\varpi}_{n,i}(\boldsymbol{\beta})} \\ &= \frac{1}{n}\sum_{i=1}^n \hat{\varpi}_{n,i}(\boldsymbol{\beta}) - \frac{1}{n}\sum_{i=1}^n \hat{\varpi}_{n,i}(\boldsymbol{\beta})\hat{\varpi}_{n,i}(\boldsymbol{\beta})^\tau \lambda + \frac{1}{n}\sum_{i=1}^n \frac{\hat{\varpi}_{n,i}(\boldsymbol{\beta})\{\lambda^\tau \hat{\varpi}^s_{n,i}(\boldsymbol{\beta})\}^2}{1 + \lambda^\tau \hat{\varpi}_{n,i}(\boldsymbol{\beta})}. \end{aligned} \tag{A.12}$$

This equation and $\max_{1\le i\le n} |\lambda^\tau \hat{\varpi}^s_{n,i}(\boldsymbol{\beta})| = o_P(1)$ entail that

$$\lambda = \left(\frac{1}{n}\sum_{i=1}^n \hat{\varpi}_{n,i}(\boldsymbol{\beta})\hat{\varpi}_{n,i}(\boldsymbol{\beta})^\tau\right)^{-1} \frac{1}{n}\sum_{i=1}^n \hat{\varpi}_{n,i}(\boldsymbol{\beta}) + o_P(n^{-1/2}). \tag{A.13}$$

Plugging the asymptotic expression (A.13) to (A.11), we have

$$\hat{l}(\boldsymbol{\beta}) = n\left(\frac{1}{n}\sum_{i=1}^n \hat{\varpi}_{n,i}(\boldsymbol{\beta})\right)^\tau \left(\frac{1}{n}\sum_{i=1}^n \hat{\varpi}_{n,i}(\boldsymbol{\beta})\hat{\varpi}_{n,i}(\boldsymbol{\beta})^\tau\right)^{-1} \left(\frac{1}{n}\sum_{i=1}^n \hat{\varpi}_{n,i}(\boldsymbol{\beta})\right) + o_P(1).$$

Applying Lemma B.2 to $E^s_{n,ij}$ for $j = 1, \ldots, 5$, and similarly to the proof of Theorem 2, we obtain that

$$\hat{l}(\boldsymbol{\beta}) = n\left(\frac{1}{n}\sum_{i=1}^n \kappa_{n,i}(\boldsymbol{\beta})\right)^\tau \left(\frac{1}{n}\sum_{i=1}^n \kappa_{n,i}(\boldsymbol{\beta})\kappa_{n,i}(\boldsymbol{\beta})^\tau\right)^{-1} \left(\frac{1}{n}\sum_{i=1}^n \kappa_{n,i}(\boldsymbol{\beta})\right) + o_P(1),$$

where $\kappa_{n,i}(\boldsymbol{\beta})$'s are i.i.d. q-dimensional random vectors with zero mean. Theorem 3 follows from the central limit theorem and the Slutsky theorem.

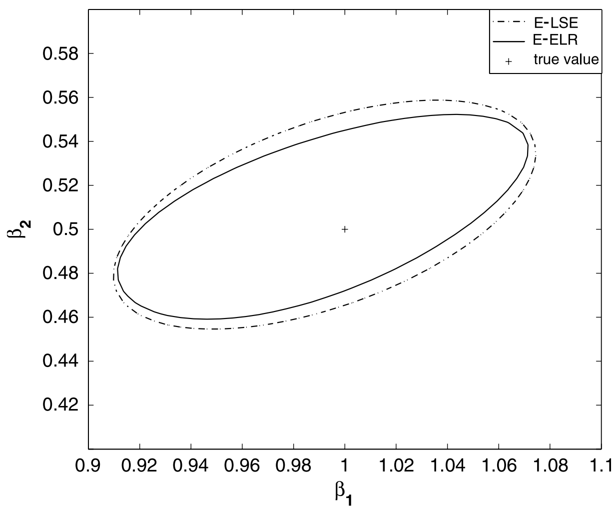

Fig. B.1. Simulation study. Confidence regions for case 1. The solid and dash-dotted lines correspond to the empirical likelihood and normal approximation methods, respectively. The plus denotes the true value.

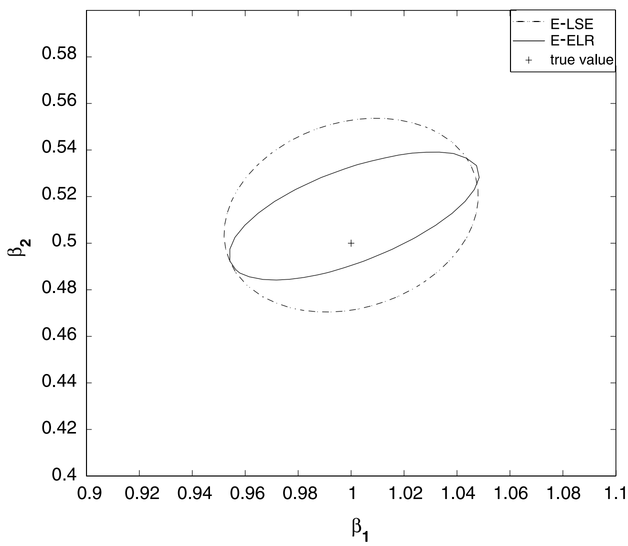

Fig. B.2. Simulation study. Confidence regions for case 2. The solid and dash-dotted lines correspond to the empirical likelihood and normal approximation methods, respectively. The plus denotes the true value.

Appendix B. Technical lemmas

The technical lemmas are used in the proofs of Theorems 1–3 in the paper.

Lemma B.1. *Suppose that Conditions* (A1)–(A5) *hold. Let* $B_n = \{(\theta', u), (\theta', u) \in \Theta \times \mathcal{U}, \|\theta' - \theta\| \le cn^{-1/2}\}$ *for a constant* $c > 0$*. Then*

$$\sup_{(\theta',u')\in B_n} |\hat{\phi}_b(\theta'^{\tau} u) - \phi(\theta^{\tau} u)| = O_P\left(\left(\frac{\log n}{nh_1}\right)^{1/2}\right), \tag{B.1}$$

$$\sup_{(\theta',u)\in B_n} |\hat{\psi}_{br}(\theta'^{\tau} u) - \psi_r(\theta^{\tau} u)| = O_P\left(\left(\frac{\log n}{nh_1}\right)^{1/2}\right), \quad r = 1, \ldots, p. \tag{B.2}$$

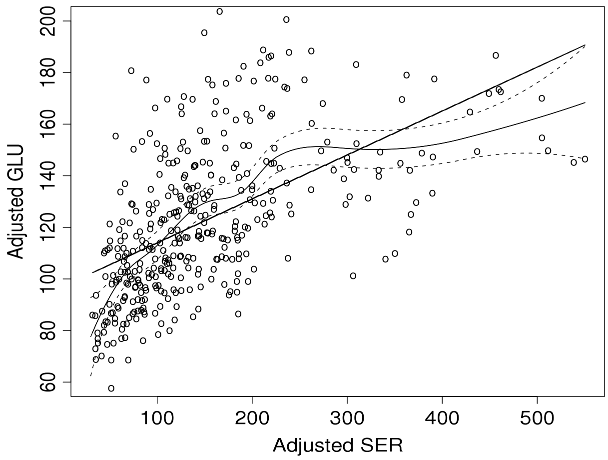

Fig. B.3. The scatterplot of adjusted GLU against adjusted SER (points), and the local linear estimators (thin solid line) along with the 95% pointwise confidence intervals (dotted lines) and a linear fitting (straight line).

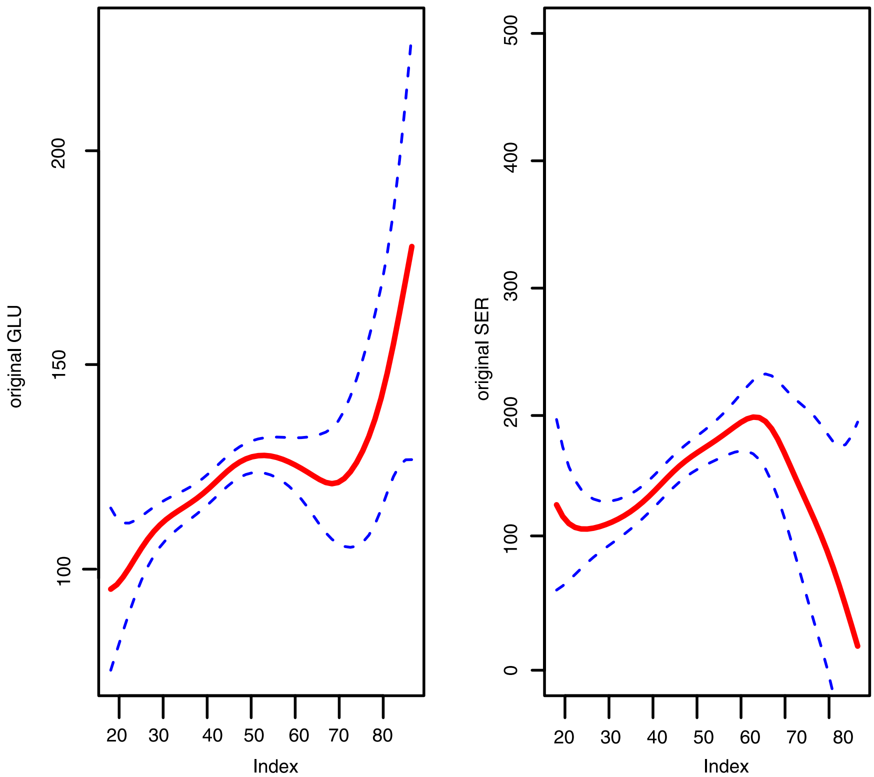

Fig. B.4. The local linear estimators of original GLU (the left panel) and original SER (the right panel) against the estimated single-index 0.7579SFT + 0.6524BMI and the 95% pointwise confidence intervals (dotted lines).

Proof. We only prove (B.1), and can complete the proof of (B.2) in a similar way.

Because $\boldsymbol{E}(\tilde{Y}|\theta^\tau U) = (\boldsymbol{E}Y)\phi(\theta^\tau U)$, we write the following model

$$\tilde{Y}_i = \boldsymbol{E}Y\phi(\theta^\tau U_i) + \eta_i, \quad \text{for } i = 1\ldots, n, \tag{B.3}$$

where $\{\eta_1, \ldots, \eta_n\}$ are $i.i.d.$ random variables with zero mean and finite variance σ_1^2 and independent of $\{U_1, \ldots, U_n\}$. From expression (10), if we denote that $C_{ni}(t, \theta') = \frac{r_i(t,\theta')}{\sum_{i=1}^n r_i(t,\theta')}$, then we have

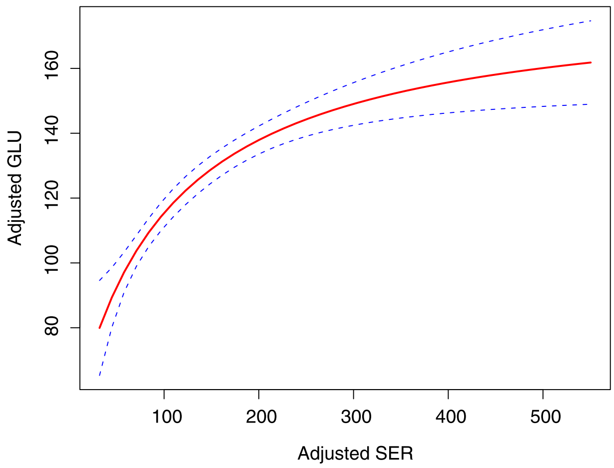

Fig. B.5. The estimated curve of adjusted SER against adjusted GLU and the associated 95% pointwise confidence intervals (dotted lines).

$$\begin{aligned}|\hat{\phi}_b(\theta'^{\tau}u)-\phi(\theta^{\tau}u)| &= \frac{1}{\bar{\tilde{Y}}}\left|\sum_{i=1}^{n}C_{ni}(t,\theta')\tilde{Y}_i-\bar{\tilde{Y}}\phi(\theta^{\tau}u)\right| \\ &\le \frac{1}{\bar{\tilde{Y}}}\left|\sum_{i=1}^{n}C_{ni}(t,\theta')\tilde{Y}_i-(\boldsymbol{E}Y)\phi(\theta'^{\tau}u)\right|+\frac{\boldsymbol{E}Y}{\bar{\tilde{Y}}}|\phi(\theta'^{\tau}u)-\phi(\theta^{\tau}u)|+\frac{1}{\bar{\tilde{Y}}}|\bar{\tilde{Y}}-\boldsymbol{E}Y\parallel\phi(\theta^{\tau}u)|.\end{aligned}$$

Note that $\bar{\tilde{Y}}-\boldsymbol{E}Y=O_P(n^{-1/2})$, $\bar{\tilde{Y}}=O_P(1)$ and $\sup_{(\theta,u)\in B_n}|\phi(\theta'^{\tau}u)-\phi(\theta^{\tau}u)|=O(n^{-1/2})$. It suffices to prove

$$\sup_{(\theta',u)\in B_n}\left|\sum_{i=1}^{n}C_{ni}(t,\theta')\tilde{Y}_i-(\boldsymbol{E}Y)\phi(\theta'^{\tau}u)\right|=O_P\left(\left(\frac{\log n}{nh_1}\right)^{1/2}\right). \tag{B.4}$$

Applying this to (B.3), similarly to Lemma 4 of [24], we have

$$\boldsymbol{E}\left|\sum_{i=1}^{n}C_{ni}(t,\theta')\tilde{Y}_i-(\boldsymbol{E}Y)\phi(\theta'^{\tau}u)\right|^2\le c\boldsymbol{E}\left|\sum_{i=1}^{n}C_{ni}(t,\theta')\phi(\theta'^{\tau}U_i)-\phi(\theta'^{\tau}u)\right|^2+c\boldsymbol{E}\sum_{i=1}^{n}C_{ni}^2(t,\theta')\sigma_1^2+O\left(\frac{1}{n}\right). \tag{B.5}$$

Directly using Lemmas A.2 and A.3 of [24], we obtain that

$$\boldsymbol{E}\left|\sum_{i=1}^{n}C_{ni}(t,\theta')\tilde{Y}_i-(\boldsymbol{E}Y)\phi(\theta'^{\tau}u)\right|^2\le ch_1^4+c\frac{1}{nh_1}. \tag{B.6}$$

Given a $M>0$, by Chebyshev's inequality, we have

$$\begin{aligned}P\left(\left|\sum_{i=1}^{n}C_{ni}(t,\theta')\tilde{Y}_i-(\boldsymbol{E}Y)\phi(\theta'^{\tau}u)\right|>\frac{M}{2}\left(\frac{\log n}{nh_1}\right)^{1/2}\right) &\le \frac{4(nh_1)}{M^2(\log n)}\boldsymbol{E}\left|\sum_{i=1}^{n}C_{ni}(t,\theta')\tilde{Y}_i-(\boldsymbol{E}Y)\phi(\theta'^{\tau}u)\right|^2 \\ &\le cM^{-2}(nh_1^5+(\log n)^{-1}).\end{aligned}$$

We choose an M large enough so that $cM^{-2}(nh_1^5+(\log n)^{-1})\le\frac{1}{2}$. Using Lemma A.1 of [24], we obtain

$$\begin{aligned}&P\left(\sup_{(\theta',u)\in B_n}\left|\sum_{i=1}^{n}C_{ni}(t,\theta')\tilde{Y}_i-(\boldsymbol{E}Y)\phi(\theta'^{\tau}u)\right|>\frac{M}{2}\left(\frac{\log n}{nh_1}\right)^{1/2}\right) \\ &\le cn^{2pa}M^{-2p}\boldsymbol{E}\left\{\sup_{(\theta',u)\in B_n}2\exp\left(\frac{-M^2\log n/(128nh_1)}{\sum_{i=1}^{n}(C_{ni}(t,\theta')\tilde{Y}_i-(\boldsymbol{E}Y)\phi(\theta'^{\tau}u))^2}\right)\wedge 1\right\}.\end{aligned}$$

(B.5) and (B.6) imply that $\sum_{i=1}^n (C_{ni}(t,\theta')\tilde{Y}_i - (\boldsymbol{E}Y)\phi(\theta'^\tau u))^2 = O_P(h_1^4 + \frac{1}{nh_1})$. It follows that

$$P\left(\sup_{(\theta',u)\in B_n}\left|\sum_{i=1}^n C_{ni}(t,\theta')\tilde{Y}_i - (\boldsymbol{E}Y)\phi(\theta'^\tau u)\right| > \frac{M}{2}\left(\frac{\log n}{nh_1}\right)^{1/2}\right)$$
$$\le cn^{2pa}M^{-2p}\exp\left(\frac{(-M^2\log n)}{128(nh_1^5+1)}\right) \to 0, \quad \text{for large enough } M.$$

As a result,

$$\sup_{(\theta',u)\in B_n}\left|\sum_{i=1}^n C_{ni}(t,\theta')\tilde{Y}_i - (\boldsymbol{E}Y)\phi(\theta'^\tau u)\right| = O_P\left(\left(\frac{\log n}{nh_1}\right)^{1/2}\right)$$

and

$$\sup_{(\theta',u)\in B_n}|\hat{\phi}_b(\theta'^\tau u,\theta') - \phi(\theta^\tau u)| = O_P\left(\left(\frac{\log n}{nh_1}\right)^{1/2}\right).$$

We complete the proof of Lemma B.1. □

Lemma B.2. *Suppose that Conditions* (A1)–(A5) *hold. Let* $T(x)$ *be a continuous function satisfying* $\boldsymbol{E}T^2(\mathbf{X}) < \infty$. *Then, we have the following asymptotic representation, for* $r = 1,\ldots,p$,

$$\frac{1}{n}\sum_{i=1}^n(\hat{Y}_i - Y_i)T(\mathbf{X}_i) = \frac{1}{n}\sum_{i=1}^n(\tilde{Y}_i - Y_i)\frac{\boldsymbol{E}(YT(\mathbf{X}))}{\boldsymbol{E}Y} + o_P(n^{-1/2}), \tag{B.7}$$

$$\frac{1}{n}\sum_{i=1}^n(\hat{X}_{ri} - X_{ri})T(\mathbf{X}_i) = \frac{1}{n}\sum_{i=1}^n(\tilde{X}_{ri} - X_{ri})\frac{\boldsymbol{E}(X_rT(\mathbf{X}))}{\boldsymbol{E}X_r} + o_P(n^{-1/2}). \tag{B.8}$$

Remark 4. The results in Lemma B.2 are different from what [4] obtained in their Lemma A.1, where they had two redundant terms: $\frac{1}{2n}\sum_{i=1}^n(Y_i - \boldsymbol{E}Y)\boldsymbol{E}(YT(\mathbf{X}))/\boldsymbol{E}Y$, $\frac{1}{2n}\sum_{i=1}^n(X_{ri} - \boldsymbol{E}X_r)\boldsymbol{E}(X_rT(\mathbf{X}))/\boldsymbol{E}X_r$, because they erroneously used a result for the U-statistic. Note that $\mathrm{Cov}((\tilde{Y} - Y)(Y - \boldsymbol{E}Y)) = 0$, and $\mathrm{Cov}((\tilde{X}_r - X_r)(X_r - \boldsymbol{E}X_r)) = 0$. All their results based on their Lemma A.1 need to be modified accordingly and the asymptotic covariance of their estimators $\hat{\boldsymbol{\beta}}$ should be smaller in the sense of the semi-positive definite.

Proof. We only prove (B.7). The proof of (B.8) is similar. Decompose $\frac{1}{n}\sum_{i=1}^n(Y_i - \hat{Y}_i)T(\mathbf{X}_i)$ as

$$\frac{1}{n}\sum_{i=1}^n\frac{Y_iT(\mathbf{X}_i)}{\phi(\theta^\tau U_i)}[\phi(\theta^\tau U_i) - \hat{\phi}_b(\theta^\tau U_i)] + \frac{1}{n}\sum_{i=1}^n\frac{Y_iT(\mathbf{X}_i)}{\phi(\theta^\tau U_i)}[\hat{\phi}_b(\theta^\tau U_i) - \hat{\phi}_b(\hat{\theta}^\tau U_i)]$$
$$+\frac{1}{n}\sum_{i=1}^n Y_iT(\mathbf{X}_i)\phi(\theta^\tau U_i)\frac{(\phi(\theta^\tau U_i) - \hat{\phi}_b(\theta^\tau U_i))(\hat{\phi}_b(\theta^\tau U_i) - \hat{\phi}_b(\hat{\theta}^\tau U_i))}{\phi(\theta^\tau U_i)\hat{\phi}_b(\hat{\theta}^\tau U_i)}$$
$$=: I_{n1} + I_{n2} + I_{n3}.$$

We will evaluate I_{n1}, I_{n2}, and I_{n3} in the following three steps.

Step 1. We prove that

$$I_{n1} = \frac{1}{n}\sum_{i=1}^n(\tilde{Y}_i - Y_i)\frac{\boldsymbol{E}(YT(\mathbf{X}))}{\boldsymbol{E}Y} + o_P(n^{-1/2}). \tag{B.9}$$

Note that

$$I_{n1} = \frac{1}{n}\sum_{i=1}^n Y_iT(\mathbf{X}_i) - \frac{1}{n}\sum_{i=1}^n\frac{Y_iT(\mathbf{X}_i)}{\phi(\theta^\tau U_i)}\hat{\phi}_b(\theta^\tau U_i).$$

From expression (10), it is easily seen that

$$\frac{1}{n}\sum_{i=1}^n\frac{Y_iT(\mathbf{X}_i)}{\phi(\theta^\tau U_i)}\hat{\phi}_b(\theta^\tau U_i) = \frac{1}{n}\sum_{i=1}^n\frac{Y_iT(\mathbf{X}_i)}{\phi(\theta^\tau U_i)}\left[\frac{\sum_{j=1}^n L_{h_1}(\theta^\tau U_j - \theta^\tau U_i)\tilde{Y}_j}{Q_{n0}(\theta^\tau U_i,\theta)\bar{\bar{Y}}}\right]$$

$$-\frac{1}{n}\sum_{i=1}^{n}\frac{Y_iT(\mathbf{X}_i)}{\phi(\theta^\tau U_i)}\left[\frac{\sum_{j=1}^{n}L_{h_1}(\theta^\tau U_j-\theta^\tau U_i)\tilde{Y}_j}{Q_{n0}^2(\theta^\tau U_i,\theta)Q_{n2}(\theta^\tau U_i,\theta)-Q_{n1}^2(\theta^\tau U_i,\theta)Q_{n0}(\theta^\tau U_i,\theta)}\right]\frac{1}{\bar{\tilde{Y}}}$$

$$=: I_{n1}^{(1)}-I_{n1}^{(2)}.$$

Note that $I_{n1}^{(1)}$ can further be expressed as the summand of $I_{n1}^{(1)R_1}$, $I_{n1}^{(1)R_2}$ and $I_{n1}^{(1)R_3}$, where

$$I_{n1}^{(1)R_1}=\frac{1}{n}\sum_{i=1}^{n}\frac{Y_iT(\mathbf{X}_i)}{\phi(\theta^\tau U_i)}\left[\frac{\frac{1}{n}\sum_{j=1}^{n}L_{h_1}(\theta^\tau U_j-\theta^\tau U_i)\tilde{Y}_j}{f_{\theta^\tau U}(\theta^\tau U_i)\boldsymbol{E}Y}\right],$$

$$I_{n1}^{(1)R_2}=\frac{1}{n}\sum_{i=1}^{n}\frac{Y_iT(\mathbf{X}_i)}{\phi(\theta^\tau U_i)}\left[\frac{\frac{1}{n}\sum_{j=1}^{n}L_{h_1}(\theta^\tau U_j-\theta^\tau U_i)\tilde{Y}_j}{f_{\theta^\tau U}(\theta^\tau U_i)\boldsymbol{E}Y}\right]\left[\frac{\boldsymbol{E}Y-\bar{\tilde{Y}}}{\bar{\tilde{Y}}}\right],$$

$$I_{n1}^{(1)R_3}=\frac{1}{n}\sum_{i=1}^{n}\frac{Y_iT(\mathbf{X}_i)}{\phi(\theta^\tau U_i)}\left[\frac{\frac{1}{n}\sum_{j=1}^{n}L_{h_1}(\theta^\tau U_j-\theta^\tau U_i)\tilde{Y}_j}{f_{\theta^\tau U}(\theta^\tau U_i)}\right]\left[\frac{f_{\theta^\tau U}(\theta^\tau U_i)-\frac{1}{n}Q_{n0}(\theta^\tau U_i,\theta)}{\frac{1}{n}Q_{n0}(\theta^\tau U_i,\theta)}\right].$$

Consider term $I_{n1}^{(1)R_1}$. We know $\frac{1}{n^2h_1}\sum_{i=1}^{n}\frac{Y_i^2T(\mathbf{X}_i)L(0)}{f_{\theta^\tau U}(\theta^\tau U_i)\boldsymbol{E}Y}=o_P(n^{-1/2})$ by the law of large numbers and Assumption (A4)(ii) imposed on the bandwidth. The summation for $i\neq j$ of $I_{n1}^{(1)R_1}$ is a standard U-statistic with a varying kernel with bandwidth h_1; that is,

$$\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\frac{Y_i\tilde{Y}_jT(\mathbf{X}_i)L_{h_1}(\theta^\tau U_j-\theta^\tau U_i)}{\phi(\theta^\tau U_i)f_{\theta^\tau U}(\theta^\tau U_i)\boldsymbol{E}Y}=\frac{2a_n}{n(n-1)}\sum_{1\le i<j\le n}H((\mathbf{X}_i,Y_i,U_i),(\mathbf{X}_j,Y_j,U_j)),$$

where $a_n=(n-1)/n$. Recall that $L(\cdot)$ is a symmetric function and the symmetric U-statistic kernel is $H(\cdot,\cdot)$; that is,

$$H((\mathbf{X}_i,Y_i,U_i),(\mathbf{X}_j,Y_j,U_j))=\frac{1}{2}L_{h_1}(\theta^\tau U_j-\theta^\tau U_i)\left[\frac{Y_iT(\mathbf{X}_i)\tilde{Y}_j}{\phi(\theta^\tau U_i)f_{\theta^\tau U}(\theta^\tau U_i)\boldsymbol{E}Y}+\frac{Y_jT(\mathbf{X}_j)\tilde{Y}_i}{\phi(\theta^\tau U_j)f_{\theta^\tau U}(\theta^\tau U_j)\boldsymbol{E}Y}\right].$$

Using the projection of the U-statistic and seeing more details in Section 5.3.1 of [22], we obtain that

$$\frac{2}{n(n-1)}\sum_{1\le i<j\le n}H((\mathbf{X}_i,Y_i,U_i),(\mathbf{X}_j,Y_j,U_j))-\boldsymbol{E}H((\mathbf{X}_1,Y_1,U_1),(\mathbf{X}_2,Y_2,U_2))$$
$$=\frac{2}{n}\sum_{i=1}^{n}H^*(\mathbf{X}_i,Y_i,U_i)+o_P(n^{-1/2}),$$

with

$$\begin{aligned}H^*(\mathbf{X}_i,Y_i,U_i)&=\frac{1}{2}\boldsymbol{E}\left[\frac{L_{h_1}(\theta^\tau U_j-\theta^\tau U_i)Y_iT(\mathbf{X}_i)\tilde{Y}_j}{\phi(\theta^\tau U_i)f_{\theta^\tau U}(\theta^\tau U_i)\boldsymbol{E}Y}\middle|(\mathbf{X}_i,Y_i,U_i)\right]\\&\quad+\frac{1}{2}\boldsymbol{E}\left[\frac{L_{h_1}(\theta^\tau U_j-\theta^\tau U_i)Y_jT(\mathbf{X}_j)\tilde{Y}_i}{\phi(\theta^\tau U_j)f_{\theta^\tau U}(\theta^\tau U_j)\boldsymbol{E}Y}\middle|(\mathbf{X}_i,Y_i,U_i)\right]-\boldsymbol{E}H((\mathbf{X}_1,Y_1,U_1),(\mathbf{X}_2,Y_2,U_2))\\&=\frac{1}{2}\frac{Y_iT(\mathbf{X}_i)}{\phi(\theta^\tau U_i)f_{\theta^\tau U}(\theta^\tau U_i)}\boldsymbol{E}[L_{h_1}(\theta^\tau U_j-\theta^\tau U_i)\phi(\theta^\tau U_j)|U_i]\\&\quad+\frac{1}{2}\tilde{Y}_i\frac{\boldsymbol{E}YT(\mathbf{X})}{\boldsymbol{E}Y}\boldsymbol{E}\left[\frac{L_{h_1}(\theta^\tau U_j-\theta^\tau U_i)}{\phi(\theta^\tau U_j)f_{\theta^\tau U}(\theta^\tau U_j)}\middle|U_i\right]-\boldsymbol{E}H((\mathbf{X}_1,Y_1,U_1),(\mathbf{X}_2,Y_2,U_2)).\end{aligned}$$

We can verify that

$$\boldsymbol{E}[L_{h_1}(\theta^\tau U_j-\theta^\tau U_i)\phi(\theta^\tau U_j)|U_i]=f_{\theta^\tau U}(\theta^\tau U_i)\phi(\theta^\tau U_i)+O_P(h_1^2),$$
$$\boldsymbol{E}\left[\left.\frac{L_{h_1}(\theta^\tau U_j-\theta^\tau U_i)}{\phi(\theta^\tau U_j)f_{\theta^\tau U}(\theta^\tau U_j)}\right|U_i\right]=\frac{1}{\phi(\theta^\tau U_i)}+O_P(h_1^2),$$
$$\boldsymbol{E}H((\mathbf{X}_1,Y_1,U_1),(\mathbf{X}_2,Y_2,U_2))=\boldsymbol{E}YT(\mathbf{X})+O(h_1^2).$$

Thus, we have

$$H^*(\mathbf{X}_i,Y_i,U_i)=\frac{1}{2}Y_iT(\mathbf{X}_i)+\frac{1}{2}Y_i\frac{\boldsymbol{E}YT(\mathbf{X})}{\boldsymbol{E}Y}-\boldsymbol{E}YT(\mathbf{X})+O_P(h_1^2).$$

Furthermore, $O_P(h_1^2)=o_P(n^{-1/2})$ when $nh_1^4\to 0$. Then

$$\frac{2}{n(n-1)}\sum_{1\le i<j\le n}H((\mathbf{X}_i,Y_i,U_i),(\mathbf{X}_j,Y_j,U_j))-\boldsymbol{E}H((\mathbf{X}_1,Y_1,U_1),(\mathbf{X}_2,Y_2,U_2))$$
$$=\frac{1}{n}\sum_{i=1}^n\left\{Y_iT(\mathbf{X}_i)+Y_i\frac{\boldsymbol{E}YT(\mathbf{X})}{\boldsymbol{E}Y}-2\boldsymbol{E}YT(\mathbf{X})\right\}+o_P(n^{-1/2}).$$

As a result,

$$I_{n1}^{(1)R_1}=\frac{1}{n}\sum_{i=1}^n\left\{Y_iT(\mathbf{X}_i)+Y_i\frac{\boldsymbol{E}YT(\mathbf{X})}{\boldsymbol{E}Y}-2\boldsymbol{E}YT(\mathbf{X})\right\}+\boldsymbol{E}YT(\mathbf{X})+o_P(n^{-1/2}).$$

Note that $I_{n1}^{(1)R_2}=I_{n1}^{(1)R_1}(\boldsymbol{E}Y-\bar{\tilde{Y}})/\bar{\tilde{Y}}$ and we have shown that $I_{n1}^{(1)R_1}=\boldsymbol{E}YT(\mathbf{X})+o_P(1)$. It follows that

$$I_{n1}^{(1)R_2}=\frac{\boldsymbol{E}YT(\mathbf{X})}{\boldsymbol{E}Y}\frac{1}{n}\sum_{i=1}^n(\boldsymbol{E}Y-\tilde{Y}_i)+o_P(n^{-1/2}).$$

The third term $I_{n1}^{(1)R_3}$ has the following asymptotic expansion.

$$I_{n1}^{(1)R_3}=\frac{1}{n}\sum_{i=1}^n\frac{Y_iT(\mathbf{X}_i)}{\phi(\theta^\tau U_i)f_{\theta^\tau U}^2(\theta^\tau U_i)}\frac{1}{n^2}\sum_{j=1}^n\sum_{s=1}^n L_{h_1}(\theta^\tau U_j-\theta^\tau U_i)\tilde{Y}_j(L_{h_1}(\theta^\tau U_s-\theta^\tau U_i)-f_{\theta^\tau U}(\theta^\tau U_i)).$$

Using a similar analysis to the derivation of the expression $I_{n1}^{(1)R_1}$, we have

$$\frac{1}{n^2}\sum_{j=1}^n\sum_{s=1}^n L_{h_1}(\theta^\tau U_j-\theta^\tau U_i)\tilde{Y}_j(L_{h_1}(\theta^\tau U_s-\theta^\tau U_i)-f_{\theta^\tau U}(\theta^\tau U_i))=O_P\left(\frac{1}{nh_1}\right).$$

Thus, we have $I_{n1}^{(1)R_3}=o_P(n^{-1/2})$. A combination of $I_{n1}^{(1)R_1}$, $I_{n1}^{(1)R_2}$ and $I_{n1}^{(1)R_3}$ yields

$$I_{n1}^{(1)}=\frac{1}{n}\sum_{i=1}^n Y_iT(\mathbf{X}_i)+\frac{1}{n}\sum_{i=1}^n(Y_i-\tilde{Y}_i)\frac{\boldsymbol{E}YT(\mathbf{X})}{\boldsymbol{E}Y}+o_P(n^{-1/2}).$$

In a way analogous to the proof of $I_{n1}^{(1)R_1}=O_P(n^{-1/2})$, we can also prove that $I_{n1}^{(2)}=o_P(n^{-1/2})$ and complete the proof of (B.9).

Step 2. We prove $I_{n2}=o_P(n^{-1/2})$.

$$I_{n2}=\frac{1}{n}\sum_{i=1}^n\frac{Y_iT(\mathbf{X}_i)}{\phi(\theta^\tau U_i)}[\hat{\phi}_b(\theta^\tau U_i)-\hat{\phi}_b(\hat{\theta}^\tau U_i)].$$

First we represent $\hat{\phi}_b(\theta^\tau u)-\hat{\phi}_b(\hat{\theta}^\tau u)$ as follows.

$$\left\{\frac{\sum_{i=1}^n L_{h_1}(\theta^\tau U_i-\theta^\tau u)\tilde{Y}_i}{Q_{n0}(\theta^\tau u,\theta)-Q_{n1}^2(\theta^\tau u,\theta)/Q_{n2}(\theta^\tau u,\theta)}-\frac{\sum_{i=1}^n L_{h_1}(\hat{\theta}^\tau U_i-\hat{\theta}^\tau u)\tilde{Y}_i}{Q_{n0}(\hat{\theta}^\tau u,\hat{\theta})-Q_{n1}^2(\hat{\theta}^\tau u,\hat{\theta})/Q_{n2}(\hat{\theta}^\tau u,\hat{\theta})}\right\}$$

$$+\left\{\frac{\sum_{i=1}^{n} L_{h_1}(\hat\theta^\tau U_i-\hat\theta^\tau u)(\hat\theta^\tau U_i-\hat\theta^\tau u)\tilde Y_i\times Q_{n1}(\hat\theta^\tau u,\hat\theta)}{Q_{n0}(\hat\theta^\tau u,\hat\theta)Q_{n2}(\hat\theta^\tau u,\hat\theta)-Q_{n1}^2(\hat\theta^\tau u,\hat\theta)}\right.$$
$$\left.-\frac{\sum_{i=1}^{n} L_{h_1}(\theta^\tau U_i-\theta^\tau u)(\theta^\tau U_i-\theta^\tau u)\tilde Y_i\times Q_{n1}(\theta^\tau u,\theta)}{Q_{n0}(\theta^\tau u,\theta)Q_{n2}(\theta^\tau u,\theta)-Q_{n1}^2(\theta^\tau u,\theta)}\right\}$$
$$=: D_{n2}^{(1)}(\theta^\tau u,\hat\theta^\tau u)+D_{n2}^{(2)}(\theta^\tau u,\hat\theta^\tau u).$$

Recall that $\hat\theta-\theta=O_P(n^{-1/2})$. By Taylor expansion, we have

$$\begin{aligned}
&L_{h_1}(\hat\theta^\tau U_i-\hat\theta^\tau u)L_{h_1}(\theta^\tau U_j-\theta^\tau u)-L_{h_1}(\hat\theta^\tau U_j-\hat\theta^\tau u)L_{h_1}(\theta^\tau U_i-\theta^\tau u)\\
&\quad=(\hat\theta-\theta)^\tau\left(\frac{U_i-u}{h_1}\right)L'_{h_1}(\theta^\tau U_i-\theta^\tau u)L_{h_1}(\theta^\tau U_j-\theta^\tau u)\\
&\quad-(\hat\theta-\theta)^\tau\left(\frac{U_j-u}{h_1}\right)L'_{h_1}(\theta^\tau U_j-\theta^\tau u)L_{h_1}(\theta^\tau U_i-\theta^\tau u)+O_P\left(\frac{1}{n}\right).
\end{aligned}\tag{B.10}$$

We then have,

$$\begin{aligned}
D_{n2}^{(1)R_1}(\theta^\tau u,\hat\theta^\tau u)&\triangleq\frac{Q_{n0}(\hat\theta^\tau u,\hat\theta)}{n}\frac{1}{n}\sum_{i=1}^{n}L_{h_1}(\theta^\tau U_j-\theta^\tau u)\tilde Y_i-\frac{Q_{n0}(\theta^\tau u,\theta)}{n}\frac{1}{n}\sum_{i=1}^{n}L_{h_1}(\hat\theta^\tau U_i-\hat\theta^\tau u)\tilde Y_i\\
&=\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}(\hat\theta-\theta)^\tau\left(\frac{U_i-u}{h_1}\right)L'_{h_1}(\theta^\tau U_i-\theta^\tau u)L_{h_1}(\theta^\tau U_j-\theta^\tau u)\tilde Y_i\\
&\quad-\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}(\hat\theta-\theta)^\tau\left(\frac{U_j-u}{h_1}\right)L'_{h_1}(\theta^\tau U_j-\theta^\tau u)L_{h_1}(\theta^\tau U_i-\theta^\tau u)\tilde Y_j+O_P\left(\frac{1}{n}\right)\\
&=(\hat\theta-\theta)O_P(h_1)+O_P\left(\frac{1}{n}\right)=o_P(n^{-1/2}).
\end{aligned}$$

The same argument implies that

$$\begin{aligned}
D_{n2}^{(1)R_2}(\theta^\tau u,\hat\theta^\tau u)&\triangleq\frac{1}{n}\sum_{i=1}^{n}L_{h_1}(\hat\theta^\tau U_i-\hat\theta u)\tilde Y_i\frac{(Q_{n1}(\theta^\tau u,\theta)/nh_1)^2}{Q_{n2}(\theta^\tau u,\theta)/nh_1^2}\\
&\quad-\frac{1}{n}\sum_{i=1}^{n}L_{h_1}(\theta^\tau U_i-\theta^\tau u)\tilde Y_i\frac{(Q_{n1}(\hat\theta^\tau u,\hat\theta)/nh_1)^2}{Q_{n2}(\hat\theta^\tau u,\hat\theta)/nh_1^2}=o_P(n^{-1/2}).
\end{aligned}$$

Thus, $D_{n2}^{(1)}(\theta^\tau u,\hat\theta^\tau u)=o_P(n^{-1/2})$. In the same way we can prove that $D_{n2}^{(2)}(\theta^\tau u,\hat\theta^\tau u)=O_P(n^{-1/2})$. These arguments, along with a direct calculation, indicate $I_{n2}=O_P(n^{-1/2})$.

Step 3. We now consider the last term I_{n3}.

$$\begin{aligned}
I_{n3}&=\frac{1}{n}\sum_{i=1}^{n}Y_iT(\mathbf{X}_i)\phi(\theta^\tau U_i)\frac{(\phi(\theta^\tau U_i)-\hat\phi_b(\hat\theta^\tau U_i))(\hat\phi_b(\theta^\tau U_i)-\hat\phi_b(\hat\theta^\tau U_i))}{\phi(\theta^\tau U_i)\hat\phi_b(\hat\theta^\tau U_i)}\\
&\quad+\frac{1}{n}\sum_{i=1}^{n}Y_iT(\mathbf{X}_i)\phi(\theta^\tau U_i)\frac{(\hat\phi_b(\theta^\tau U_i)-\hat\phi_b(\hat\theta^\tau U_i))^2}{\phi(\theta^\tau U_i)\hat\phi_b(\hat\theta^\tau U_i,\hat\theta)}\\
&\triangleq I_{n3}^{(1)}+I_{n3}^{(2)}.
\end{aligned}$$

Applying Lemma B.1 and the results obtained in *Step* 2, we obtain by Cauchy–Schwarz inequality

$$|I_{n3}^{(1)}|^2\le O_P\left\{\left[\frac{1}{n}\sum_{i=1}^{n}(Y_iT(\mathbf{X}_i)\phi(\theta^\tau U_i))^2(\hat\phi_b(\theta^\tau U_i)-\hat\phi_b(\hat\theta^\tau U_i))^2\right]\frac{\log n}{nh_1}\right\}=o_P(n^{-1}).$$

Similarly, $|I_{n3}^{(2)}|^2 = o_P(n^{-1})$. Thus, we have $I_{n3} = o_P(n^{-1/2})$. Together with $I_{n2} = o_P(n^{-1/2})$ and the asymptotic expression of I_{n1}, we conclude that

$$\frac{1}{n}\sum_{i=1}^{n}(\hat{Y}_i - Y_i)T(\mathbf{X}_i) = \frac{1}{n}\sum_{i=1}^{n}(\tilde{Y}_i - Y_i)\frac{\boldsymbol{E}(YT(\mathbf{X}))}{\boldsymbol{E}Y} + o_P(n^{-1/2}). \quad \square$$

Lemma B.3. *Suppose that Conditions* (A1)–(A5) *and* (A7) *hold. We have, for* $1 \le k \le p$,

$$n^{-1}\widehat{\mathbf{G}}_n^k(\boldsymbol{\beta}) = n^{-1}\mathbf{F}_n^k(\boldsymbol{\beta}) + o_P(n^{-1/2}) \tag{B.11}$$

$$n^{-1}\widehat{\mathbf{G}}_n(\boldsymbol{\beta})\widehat{\mathbf{G}}_n^\tau(\boldsymbol{\beta}) = n^{-1}\mathbf{F}_n(\boldsymbol{\beta})\mathbf{F}_n^\tau(\boldsymbol{\beta}) + o_P(1), \tag{B.12}$$

where $\widehat{\mathbf{G}}_n(\boldsymbol{\beta}) = (\widehat{\mathbf{G}}_n^1(\boldsymbol{\beta}), \ldots, \widehat{\mathbf{G}}_n^q(\boldsymbol{\beta}))^\tau$, *and* $\mathbf{F}_n(\boldsymbol{\beta}) = (\mathbf{F}_n^1(\boldsymbol{\beta}), \ldots, \mathbf{F}_n^q(\boldsymbol{\beta}))^\tau$ *with*

$$\begin{aligned}\mathbf{F}_n^k(\boldsymbol{\beta}) &= \sum_{i=1}^{n}\varepsilon_i\frac{\partial f(\mathbf{X}_i, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}_k} + \sum_{i=1}^{n}(\tilde{Y}_i - Y_i)\boldsymbol{E}\left(\frac{Y\partial f(\mathbf{X}, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}_k}\right)\Big/\boldsymbol{E}Y \\ &\quad - \sum_{i=1}^{n}\sum_{r=1}^{p}(\tilde{X}_{ri} - X_{ri})\left(\boldsymbol{E}X_r\frac{\partial f(\mathbf{X}, \boldsymbol{\beta})}{\partial X_r}\frac{\partial f(\mathbf{X}, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}_k}\right)\Big/\boldsymbol{E}X_r.\end{aligned} \tag{B.13}$$

Proof. By Taylor expansion, we represent $\widehat{\mathbf{G}}_n^k(\boldsymbol{\beta})$ as $\widehat{\mathbf{G}}_n^k(\boldsymbol{\beta})_1 + \widehat{\mathbf{G}}_n^k(\boldsymbol{\beta})_2 + \widehat{\mathbf{G}}_n^k(\boldsymbol{\beta})_3$ with

$$\begin{aligned}\widehat{\mathbf{G}}_n^k(\boldsymbol{\beta})_1 &= \sum_{i=1}^{n}\varepsilon_i\frac{\partial f(\mathbf{X}_i, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}_k} + \sum_{i=1}^{n}(\hat{Y}_i - Y_i)\frac{\partial f(\mathbf{X}_i, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}_k} - \sum_{i=1}^{n}\sum_{r=1}^{p}(\hat{X}_{ri} - X_{ri})\frac{\partial f(\mathbf{X}_i, \boldsymbol{\beta})}{\partial X_{ri}}\frac{\partial f(\mathbf{X}_i, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}_k}, \\ \widehat{\mathbf{G}}_n^k(\boldsymbol{\beta})_2 &= \frac{1}{2}\sum_{i=1}^{n}\sum_{r=1}^{p}\sum_{t=1}^{p}(\hat{X}_{ri} - X_{ri})(\hat{X}_{ti} - X_{ti})\frac{\partial f(\mathbf{X}_i, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}_k}\frac{\partial^2 f(\mathbf{X}_i^*, \boldsymbol{\beta})}{\partial X_{ri}\partial X_{ti}} \\ &\quad + \frac{1}{2}\sum_{i=1}^{n}\sum_{1\le r,t,l\le p}(\hat{X}_{ri} - X_{ri})(\hat{X}_{ti} - X_{ti})(\hat{X}_{li} - X_{li})\frac{\partial^2 f(\mathbf{X}_i^{**}, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}_k\partial X_{li}}\frac{\partial^2 f(\mathbf{X}_i^*, \boldsymbol{\beta})}{\partial X_{ri}\partial X_{ti}}, \\ \widehat{\mathbf{G}}_n^k(\boldsymbol{\beta})_3 &= \sum_{i=1}^{n}\sum_{r=1}^{p}\varepsilon_i(\hat{X}_{ri} - X_{ri})\frac{\partial^2 f(\mathbf{X}_i^{**}, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}_k\partial X_{ri}} + \sum_{i=1}^{n}\sum_{r=1}^{p}\varepsilon_i(\hat{X}_{ri} - X_{ri})(\hat{Y}_i - Y_i)\frac{\partial^2 f(\mathbf{X}_i^{**}, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}_k\partial X_{ri}} \\ &\quad + \sum_{i=1}^{n}\sum_{r=1}^{p}\sum_{t=1}^{p}(\hat{X}_{ri} - X_{ri})(\hat{X}_{ti} - X_{ti})\frac{\partial f(\mathbf{X}_i, \boldsymbol{\beta})}{\partial X_{ri}}\frac{\partial^2 f(\mathbf{X}_i^{**}, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}_k\partial X_{ti}},\end{aligned}$$

where $\mathbf{X}_i^{**} = (X_{1i}^{**}, \ldots, X_{pi}^{**})$ and $\mathbf{X}_i^* = (X_{1i}^*, \ldots, X_{pi}^*)$ are two points between $\mathbf{X}_i$ and $\hat{\mathbf{X}}_i$. Applying Lemma B.2 to $\widehat{\mathbf{G}}_n^k(\boldsymbol{\beta})_1$ with $T(\mathbf{X}) = \frac{\partial f(\mathbf{X}, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}_k}$, and $T(\mathbf{X}) = \frac{\partial f(\mathbf{X}, \boldsymbol{\beta})}{\partial X_{ri}}\frac{\partial f(\mathbf{X}, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}_k}$, we obtain that

$$\frac{1}{n}\widehat{\mathbf{G}}_n^k(\boldsymbol{\beta})_1 = \frac{1}{n}\mathbf{F}_n^k(\boldsymbol{\beta}) + o_P(n^{-1/2}).$$

We now prove that both $\frac{1}{n}\widehat{\mathbf{G}}_n^k(\boldsymbol{\beta})_2$ and $\frac{1}{n}\widehat{\mathbf{G}}_n^k(\boldsymbol{\beta})_3$ are $o_P(n^{-1/2})$. Applying Cauchy–Schwarz inequality and Assumption (A7), the first term of $\frac{1}{n}\widehat{\mathbf{G}}_n^k(\boldsymbol{\beta})_2$ is bounded by

$$\begin{aligned}&\left|\frac{1}{2n}\sum_{i=1}^{n}\sum_{r=1}^{p}\sum_{t=1}^{p}(\hat{X}_{ri} - X_{ri})(\hat{X}_{ti} - X_{ti})\frac{\partial f(\mathbf{X}_i, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}_k}\frac{\partial^2 f(\mathbf{X}_i^*, \boldsymbol{\beta})}{\partial X_{ri}\partial X_{ti}}\right| \\ &\quad \le \frac{C}{2}\sum_{r=1}^{p}\sum_{t=1}^{p}\left\{\frac{1}{n}\sum_{i=1}^{n}(\hat{X}_{ri} - X_{ri})^2(\hat{X}_{ti} - X_{ti})^2\right\}^{1/2}\left\{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{\partial f(\mathbf{X}_i, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}_k}\right)^2\right\}^{1/2}.\end{aligned}$$

Using Lemma B.1, we know $(\hat{X}_{ri} - X_{ri})^2 = X_{ri}^2(\frac{\psi_r(\theta^\tau U_i) - \hat{\psi}_{br}(\hat{\theta}^\tau U_i, \hat{\theta})}{\hat{\psi}_{br}(\hat{\theta}^\tau U_i, \hat{\theta})})^2 = O_P(\frac{\log n}{nh_1})$. Using a similar analysis to the second term of $\frac{1}{n}\widehat{\mathbf{G}}_n^k(\boldsymbol{\beta})_2$, we know $\frac{1}{n}\widehat{\mathbf{G}}_n^k(\boldsymbol{\beta})_2 = O_P(\frac{\log n}{nh_1}) = o_P(n^{-1/2})$.

The first term in the expression of $\frac{1}{n}\widehat{\mathbf{G}}_n^k(\boldsymbol{\beta})_3$ is bounded by

$$\begin{aligned}
&\left\{\frac{1}{n}\sum_{i=1}^{n}\sum_{r=1}^{p}\varepsilon_i(\hat{X}_{ri}-X_{ri})\frac{\partial^2 f(\mathbf{X}_i^{**},\boldsymbol{\beta})}{\partial\boldsymbol{\beta}_k\partial X_{ri}}\right\}^2\\
&\le p^2C^2\sum_{r=1}^{p}\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i^2\frac{1}{n}\sum_{i=1}^{n}X_{ri}^2\left(\frac{\psi_r(\theta^\tau U_i)-\hat{\psi}_{br}(\hat{\theta}^\tau U_i,\hat{\theta})}{\hat{\psi}_{br}(\hat{\theta}^\tau U_i,\hat{\theta})}\right)^2\\
&= O_P\left(\frac{\log n}{nh_1}\right)=o_P(n^{-1/2}).
\end{aligned}$$

With a similar analysis for the first term in the expression of $\frac{1}{n}\widehat{\mathbf{G}}_n^k(\boldsymbol{\beta})_3$, the second and third terms of $\frac{1}{n}\widehat{\mathbf{G}}_n^k(\boldsymbol{\beta})_3$ are $o_P(n^{-1/2})$. We complete the proof of (B.11). The proof of (B.12) follows (B.11) directly. □

References

[1] E. Carmina, F. Gonzalez, A. Vidali, F. Stanczyk, M. Ferin, R. Lobo, The contributions of oestrogen and growth factors to increased adrenal androgen secretion in polycystic ovary syndrome, Human Reproduction 14 (1999) 307–311.
[2] R. Carroll, J. Fan, I. Gijbels, M.P. Wand, Generalized partially linear single-index models, Journal of the American Statistical Association 92 (1997) 477–489.
[3] R.J. Carroll, D. Ruppert, L.A. Stefanski, C.M. Crainiceanu, Nonlinear Measurement Error Models, A Modern Perspective, second ed., Chapman and Hall, New York, 2006.
[4] X. Cui, W. Guo, L. Lin, L. Zhu, Covariate-adjusted nonlinear regression, The Annals of Statistics 37 (2009) 1839–1870.
[5] X. Cui, W. Härdle, L. Zhu, Generalized single-index models: the EFM approach, The Annals of Statistics 39 (2011) 1658–1688.
[6] P. Hans, A. Vanthuyne, P.Y. Dewandre, J.F. Brichant, V. Bonhomme, Blood glucose concentration profile after 10 mg dexamethasone in non-diabetic and type 2 diabetic patients undergoing abdominal surgery, British Journal of Anaesthesia 97 (2006) 164–170.
[7] W. Härdle, P. Hall, H. Ichimura, Optimal smoothing in single-index models, The Annals of Statistics 21 (1993) 157–178.
[8] W. Härdle, T. Stoker, Investigating smooth multiple regression by the method of average derivatives, Journal of the American Statistical Association 84 (1989) 986–995.
[9] J.L. Horowitz, W. Härdle, Direct semiparametric estimation of single-index models with discrete covariates, Journal of the American Statistical Association 91 (1996) 1632–1640.
[10] G.A. Kaysen, J.A. Dubin, H.-G. Müller, W.E. Mitch, L.M. Rosales, W. Nathan, Levin the Hemo Study Group, Relationships among inflammation nutrition and physiologic mechanisms establishing albumin levels in hemodialysis patients, Kidney International 61 (2002) 2240–2249.
[11] S. Lee, K. Kwon, Pharmacokinetic-pharmacodynamic modeling for the relationship between glucose-lowering effect and plasma concentration of metformin in volunteers, Archives of Pharmacal Research 27 (2004) 806–810.
[12] H. Liang, X. Liu, R. Li, C.L. Tsai, Estimation and testing for partially linear single-index models, The Annals of Statistics 38 (2010) 3811–3836.
[13] M.H. Mohamed, G.I. Gad, H.Y. Ibrahim, M.S.E. Shemi, M.F. Moustafa, S.H. Atef, N.M. Ramadan, S.M.E. Saeid, Clinical research cord blood resistin and adiponectin in term newborns of diabetic mothers, Archives of Medical Science 6 (2010) 558–566.
[14] D.V. Nguyen, D. Şentürk, Multicovariate-adjusted regression models, Journal of Statistical Computation and Simulation 78 (2008) 813–827.
[15] A.B. Owen, Empirical Likelihood, Chapman and Hall/CRC, London, 2001.
[16] A. Owen, Empirical likelihood for linear models, The Annals of Statistics 19 (1991) 1725–1747.
[17] J.L. Powell, J.H. Stock, T.M. Stoker, Semiparametric estimation of index coefficients, Econometrica 57 (1989) 1403–1430.
[18] J. Qin, J. Lawless, Empirical likelihood and general estimating equations, The Annals of Statistics 22 (1994) 300–325.
[19] D. Şentürk, H.-G. Müller, Covariate-adjusted regression, Biometrika 92 (2005) 75–89.
[20] D. Şentürk, H.-G. Müller, Inference for covariate adjusted regression via varying coefficient models, The Annals of Statistics 34 (2006) 654–679.
[21] D. Şentürk, H.-G. Müller, Covariate-adjusted generalized linear models, Biometrika 96 (2009) 357–370.
[22] R.J. Serfling, Approximation Theorems of Mathematical Statistics, John Wiley & Sons Inc., New York, 1980.
[23] W. Stute, S. Thies, L.-X. Zhu, Model checks for regression: an innovation process approach, The Annals of Statistics 26 (1998) 1916–1934.
[24] J.-L. Wang, L. Xue, L. Zhu, Y.S. Chong, Estimation for a partial-linear single-index model, The Annals of Statistics 38 (2010) 246–274.
[25] C.-F. Wu, Asymptotic theory of nonlinear least squares estimation, The Annals of Statistics 9 (1981) 501–513.
[26] Y. Yu, D. Ruppert, Penalized spline estimation for partially linear single-index models, Journal of the American Statistical Association 97 (2002) 1042–1054.
[27] J. Zhang, L. Zhu, L. Zhu, On a dimension reduction regression with covariate adjustment, Journal of Multivariate Analysis 104 (2012) 39–55.
[28] L. Zhu, L. Xue, Empirical likelihood confidence regions in a partially linear single-index model, Journal of the Royal Statistical Society. Series B. Statistical Methodology 68 (2006) 549–570.