Ann Inst Stat Math (2013) 65:237–267
DOI 10.1007/s10463-012-0371-z

Partial linear single index models with distortion measurement errors

Jun Zhang · Yao Yu · Li-Xing Zhu · Hua Liang

Received: 22 August 2011 / Revised: 9 April 2012 / Published online: 20 July 2012

Abstract We study partial linear single index models when the response and the covariates in the parametric part are measured with errors and distorted by unknown functions of commonly observable confounding variables, and propose a semiparametric covariate-adjusted estimation procedure. We apply the minimum average variance estimation method to estimate the parameters of interest. This is different from all existing covariate-adjusted methods in the literature. Asymptotic properties of the proposed estimators are established. Moreover, we also study variable selection by adopting the coordinate-independent sparse estimation to select all relevant but distorted covariates in the parametric part. We show that the resulting sparse estimators can exclude all irrelevant covariates with probability approaching one. A simulation study is conducted to evaluate the performance of the proposed methods and a real data set is analyzed for illustration.

Keywords Coordinate-independent sparse estimation (CISE) · Covariate adjusted · Dimension reduction · Distorting function · Minimum average variance estimation (MAVE) · Measurement errors · Single index · Sparse principle component (SPC)

J. Zhang
Shenzhen-Hong Kong Joint Centre for Applied Statistics Research, Shenzhen University, Shenzhen 518060, China

Y. Yu · H. Liang (✉)
Department of Biostatistics and Computational Biology, University of Rochester, Rochester, NY 14642, USA
e-mail: Hua_Liang@URMC.Rochester.edu; hliang@bst.rochester.edu

L.-X. Zhu
Department of Mathematics, Hong Kong Baptist University, Kowloon Tong, Hong Kong, China

1 Introduction

Various semiparametric regression models have been proposed to relax model assumptions imposed on traditional parametric models for dealing with complex real data. Important semiparametric regression models include partial linear models (Wahba 1984; Härdle et al. 2000), additive models (Härdle et al. 2004), partial linear single index models (Liang et al. 2010; Carroll et al. 1997; Yu and Ruppert 2002; Xia and Härdle 2006), and varying-coefficient models (Hastie and Tibshirani 1993; Fan and Zhang 1999). In this article, we focus on partial linear single index models (PLSiM), which allow retaining the ease of interpretation of parameters in multiple linear regression and the flexibility of the single index model, and can be expressed as

$$Y = \boldsymbol{X}^{\tau}\boldsymbol{\beta}_0 + g(\boldsymbol{Z}^{\tau}\boldsymbol{\theta}_0) + \varepsilon, \tag{1}$$

where "τ" denotes the transport operation throughout this paper, $(\boldsymbol{X}^{\tau}, \boldsymbol{Z}^{\tau})^{\tau} \in R^p \times R^q$, $(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0)$ is an unknown vector in $R^p \times R^q$, ε is the error term with mean zero and finite variance, and $g(\cdot)$ is an unknown univariate function. For the sake of identifiability, we assume, without loss of generality, that $\boldsymbol{\theta}_0$ is a unit vector and its first component is positive, i.e., the parameter space of $\boldsymbol{\theta}_0$ is $\boldsymbol{\Theta} = \{\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_q)^{\tau}, \|\boldsymbol{\theta}\|_2 = 1, \theta_1 > 0, \boldsymbol{\theta} \in R^q\}$; here, $\|\cdot\|_2$ stands for the Euclidean norm.

PLSiM are quite general and cover two important special cases, i.e., when the dimension of $\boldsymbol{Z}$ is one, (1) become partial linear models (PLM). Relevant studies for PLM include Chen (1988), Heckman (1986) and Speckman (1988). Härdle et al. (2000) gave a comprehensive review for PLM. When $\boldsymbol{\beta}_0 = 0$, (1) are single index models. There are various estimation procedures for single index models. See, for example, Härdle et al. (1993), Ichimura (1993), Horowitz and Härdle (1996) among others. Horowitz (2009) enlists various examples and illustrates the usefulness of the single index model.

The estimation for parameters $\boldsymbol{\beta}_0$ and $\boldsymbol{\theta}_0$ in (1) was once studied by Carroll et al. (1997) using a backfitting algorithm. Yu and Ruppert (2002) proposed a penalized spline estimation procedure. Xia and Härdle (2006) integrated the dimension reduction idea and the minimum average variance estimation (MAVE, Xia et al. 2002) for model (1). More recently, Wang et al. (2010) proposed a dimension reduction-based estimation procedure under the mild assumption that covariate $\boldsymbol{X}$ has a dimension reduction structure on the covariate $\boldsymbol{Z}$. Their procedure requires no iteration and results in a more efficient estimator of $\boldsymbol{\theta}_0$ than those obtained by Härdle et al. (1993) and Carroll et al. (1997). Moreover, Liang et al. (2010) proposed a profile least squares estimation procedure.

It is well known that both prediction and inference need to be assessed during the process of data analysis. Variable selection is then often the most important aspect regarding accuracy of the working model. The early developments of variable selection include Akaike's information criterion (AIC, Akaike 1973), Bayes information criterion (BIC, Schwarz 1978), and Mallows' C_p (Mallows 1973). However, these variable selection procedures encounter the problem of intensive computation and the lack of stability as argued by Breiman (1996). To overcome these drawbacks, Tibshirani (1996) proposed the least absolute shrinkage and selection operator (LASSO), and Fan and Li (2001) proposed the smoothing clipped absolute deviation penalty (SCAD). LASSO and SCAD procedures can select variables and estimate the corresponding nonzero coefficients simultaneously. One distinguishing feature of the SCAD procedure is that it can estimate the coefficients of the selected variables with an oracle property. That is, the resulting estimators perform asymptotically as efficient as if the true model were known. Recently, variable selection for semiparametric models has received attention. Related works include Li and Liang (2008) for the generalized varying-coefficient partially linear model, Liang et al. (2010) for PLSiM and Wang et al. (2011) for the generalized additive partial linear model. However, these works focus on directly observed data.

In many applications, variables may not be directly observed but with certain contamination. This is common in many disciplines, such as health science and medicine research. Observations with measurement errors must be handled delicately to make valid inferences. Carroll et al. (2006) gave a comprehensive survey on measurement errors. With measurement errors, variable selection becomes much more complicated. Liang and Li (2009) developed two variable selection procedures, penalized least squares and penalized quantile regression for PLM with additive measurement errors, and observed that if measurement errors were ignored, some variable selection procedures might falsely choose variables and result in the final model biased. Thus, measurement error should be taken into account in variable selection procedures to avoid bias and false statistical inference.

In this paper, both response Y and covariate $\boldsymbol{X}$ are distorted by certain multiplicative distorting functions. Formally:

$$\tilde{Y} = \phi(U)Y, \quad \tilde{\boldsymbol{X}} = \psi(U)\boldsymbol{X}, \tag{2}$$

$U \perp\!\!\!\perp (Y, \boldsymbol{Z}, \boldsymbol{X})$, where $\perp\!\!\!\perp$ indicates independence, U is an observed continuous scalar confounding variable, $\psi(U)$ is a $p \times p$ diagonal matrix, and $\operatorname{diag}\big(\psi_1(U), \ldots, \psi_p(U)\big)$, where $\phi(\cdot)$ and $\psi_r(\cdot)$ denote the unknown continuous distorting functions. The diagonal form of $\psi(U)$ implies that the confounding variable U distorts each component of the unobserved covariate $\boldsymbol{X}$ in a multiplicative fashion (Şentürk and Müller 2005; Şentürk and Müller 2006, 2009). This scenario is not uncommon in biomedical and health-related studies. The collected data are often needed to adjust for some measures like body mass index, body surface area, height or weight. For instance, in a study of the relationship between fibrinogen and serum transferrin levels among haemodialysis patients, Kaysen et al. (2002) realized that the fibrinogen and serum transferrin levels should be normalized by dividing both response and predictors by the body mass index (BMI). This implies a multiplicative fashion of the relationship between unobserved primary variables and confounding variable. Şentürk and Müller (2005); Şentürk and Müller (2006, 2009) suggested that the confounding variable affects the primary variables through flexible multiplicative unknown functions. Such normalization by division of the general distorting functions may reduce non-negligible bias and lead to consistent estimators of the parameters of interest.

Describing the relationship between unobserved primary variables and confounding variable by a varying coefficient model, Şentürk and Müller investigated some parametric models, such as the linear model (Şentürk and Müller 2005; Şentürk and Müller 2006) and the generalized linear model (Şentürk and Müller 2009) using a binning method for fitting the varying coefficient model. Recently, Cui et al. (2009) proposed a direct plug-in estimation procedure for nonlinear regression. The direct plug-in estimation procedure first estimates the distorting functions $\phi(\cdot)$ and $\psi_r(\cdot)$s by using the local linear regression, and then estimates the unobserved predictors and response, namely, $\hat{\boldsymbol{X}} = \hat{\psi}^{-1}\tilde{\mathbf{X}}$, $\hat{Y} = \tilde{Y}/\hat{\phi}$, respectively. Further estimation is then implemented on the estimated predictors and response. A key feature of this direct plug-in method is that it has a more potential application and can be easily adopted in linear, nonlinear, generalized linear models or other semiparametric models, whereas the binning technique used by Şentürk and Müller (2005); Şentürk and Müller (2006, 2009) is designed for linear or generalized linear covariate-adjusted models.

As mentioned above, the semiparametric models have more flexibility to handle underlying models that are unknown. However, the estimation in covariate-adjusted semiparametric models is very challenging. This is partially owing to substantial difficulties in fitting covariate-adjusted semiparametric models: the lack of directly observed data and the need of estimating some infinite-dimensional parameters in semiparametric models. In particular, the root-n consistency and asymptotic normality are more difficult to establish than those for parametric models. Furthermore, variable selection for covariate-adjusted data needs to be developed, a fundamental problem that has not been addressed in the literature.

We first investigate the estimation procedure of covariate-adjusted semiparametric models, namely, covariate-adjusted partial linear single index model (CAPLSiM). We propose a new estimation procedure based on MAVE (Xia et al. 2002). Our goal is to estimate the unknown parameters $(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0)$ consistently based on the observed and confounded data $\{\tilde{Y}_i, \tilde{\boldsymbol{X}}_i, \boldsymbol{Z}_i, U_i\}_{i=1}^n$, and to further establish asymptotic normality for the proposed estimators. Moreover, studying variable selection, we propose a sparse principle component (SPC) analysis based on the recently developed coordinate-independent sparse estimation (CISE, Chen et al. 2010). We demonstrate that the resulting SPC-based solution is selection consistent. The final sparse estimator of $\boldsymbol{\beta}_0$ is closely related to the first SPC. We conduct Monte Carlo simulation experiments to assess the performance of the proposed procedures. Our simulation results show that the proposed procedures perform well both in estimation and variable selection.

The paper is organized as follows. In Sect. 2, we propose the estimation procedure for the parameters $\boldsymbol{\beta}_0$ and $\boldsymbol{\theta}_0$, introduce the algorithms for computing their estimators, and present asymptotic properties of the resulting estimators. In Sect. 3, we provide an algorithm for variable selection and give theoretical properties. In Sect. 4, we report the results of a simulation study and the results of an analysis of a diabetes data set. All the technical proofs of the asymptotic results are given in the Appendix.

2 Estimation and main results

From model (2), we re-write PLSiM (1) into the following CAPLSiM:

$$\tilde{Y} = \tilde{\boldsymbol{X}}^{\tau}\boldsymbol{\beta}_0(U) + \phi(U)g(\boldsymbol{Z}^{\tau}\boldsymbol{\theta}_0) + \phi(U)\varepsilon, \tag{3}$$

where $\boldsymbol{\beta}_0(U) = \left(\boldsymbol{\beta}_{0,1}\phi(U)/\psi_1(U), \boldsymbol{\beta}_{0,2}\phi(U)/\psi_2(U), \ldots, \boldsymbol{\beta}_{0,p}\phi(U)/\psi_p(U)\right)^{\tau}$.

Let $\{(\tilde{\boldsymbol{X}}_i, \boldsymbol{Z}_i, U_i, \tilde{Y}_i)\}_{i=1}^n$ be an $i.i.d.$ sample from $(\tilde{\boldsymbol{X}}, \boldsymbol{Z}, U, \tilde{Y})$. When $(\boldsymbol{Z}_i^\tau \boldsymbol{\theta}_0, U_i)$ close to $(z^\tau \boldsymbol{\theta}_0, u)$, we have an approximation:

$$\phi(U)g(\boldsymbol{Z}^\tau \boldsymbol{\theta}_0) \approx \phi(u)g(z^\tau \boldsymbol{\theta}_0) + \phi(u)g'(z^\tau \boldsymbol{\theta}_0)\boldsymbol{\theta}_0^\tau(\boldsymbol{Z}_i - z) + \phi'(u)g(z^\tau \boldsymbol{\theta}_0)(U_i - u).$$

Then, the resulting estimators of $\boldsymbol{\theta}_0$, $\{\boldsymbol{\beta}_0(U_i)\}_{i=1}^n$, $\{\phi(U_i)g(\boldsymbol{Z}_i^\tau \boldsymbol{\theta}_0)\}_{i=1}^n$, $\{\phi(U_i)g'(\boldsymbol{Z}_i^\tau \boldsymbol{\theta}_0)\}_{i=1}^n$ and $\{\phi'(U_i)g(\boldsymbol{Z}_i^\tau \boldsymbol{\theta}_0)\}_{i=1}^n$ are the minimizers with respect to $\boldsymbol{\theta}$ and a_j, b_j, d_{1j} d_{2j} for $j = 1, \ldots, n$; that is,

$$\arg \min_{\theta^\tau \theta = 1} \sum_{j=1}^n \sum_{i=1}^n \left\{\tilde{Y}_i - \tilde{\boldsymbol{X}}_i^\tau b_j - a_j - d_{1j}\boldsymbol{\theta}^\tau(\boldsymbol{Z}_i - \boldsymbol{Z}_j) - d_{2j}(U_i - U_j)\right\}^2 \omega_i(\boldsymbol{Z}_j, U_j), \tag{4}$$

where $\omega_i(\boldsymbol{Z}_j, U_j)$ are some kernel weight functions. We shall discuss the choices of the kernel weight functions in Sect. 2.1. Given $\boldsymbol{\theta}$, (4) is a local linear smoothing estimation procedure. At the same time, given $(a_j, b_j, d_{1j}, d_{2j})$, (4) is a weighted least squares problem for $\boldsymbol{\theta}$. Consequently, minimization of (4) can be solved by an iterative procedure; that is, iteratively estimating nonparametric component $(a_j, b_j, d_{1j}, d_{2j})$ and parameter component $\boldsymbol{\theta}$. Note that the above minimizer estimation procedure is similar to MAVE (Xia et al. 2002). Minimizing (4) is a quadratic programming, which can be solved easily with simple expressions. We describe the algorithm in the next subsection.

2.1 Algorithm for estimation

2.1.1 An initial estimator of $\boldsymbol{\theta}_0$

Note that U is independent of $\boldsymbol{Z}$, so we can choose kernel function $\omega_i(\boldsymbol{Z}_j, U_j)$ as follows. $\omega_i(\boldsymbol{Z}_j, U_j) = I_n(\boldsymbol{Z}_j, U_j)H_{h_1}(\boldsymbol{Z}_i - \boldsymbol{Z}_j)L_{h_2}(U_i - U_j)/\sum_{i=1}^n H_{h_1}(\boldsymbol{Z}_i - \boldsymbol{Z}_j)L_{h_2}(U_i - U_j)$, where $H(\cdot)$ is a q-variate multivariate density function and $H_{h_1}(\cdot) = h_1^{-q}H(\cdot/h_1)$, $L(\cdot)$ is a univariate density function and $L_{h_2}(\cdot) = h_2^{-1}L(\cdot/h_2)$, h_1, h_2 are bandwidths and $I_n(\boldsymbol{Z}_j, U_j) = I\{\frac{1}{n}\sum_{i=1}^n H_{h_1}(\boldsymbol{Z}_i - \boldsymbol{Z}_j)L_{h_2}(U_i - U_j) > c_0 c_1\}$ for two constants c_0 and c_1 given in Assumption (A2); $I\{\cdot\}$ is the indictor function. $I_n(\boldsymbol{Z}_j, U_j)$ is used to remove the effect of the boundary points in the support of $\boldsymbol{Z}$ and U. Similar technique was employed in Xia and Härdle (2006). The q-variate multivariate density function $H(\cdot)$ still faces the "curse of dimensionality", but it suffices to provide us a consistent initial estimator for $\boldsymbol{\theta}_0$.

Write $\mu_j = d_{1j}\boldsymbol{\theta}$, $\boldsymbol{M}_{ij} = \left(1, \tilde{\boldsymbol{X}}_i^\tau, (U_i - U_j), (\boldsymbol{Z}_i - \boldsymbol{Z}_j)^\tau\right)^\tau$ and $\omega_{ij} = \omega_i(\boldsymbol{Z}_j, U_j)$. Minimizing (4) with respect to $\{a_j, b_j, d_{2j}, \mu_j\}$, we obtain that, for $j = 1, \ldots, n$,

$$\left(\hat{a}_j, \hat{b}_j, \hat{d}_{2j}, \hat{\mu}_j\right)^\tau = \left\{\sum_{i=1}^n \boldsymbol{M}_{ij}\boldsymbol{M}_{ij}^\tau \omega_{ij}\right\}^{-1} \sum_{i=1}^n \boldsymbol{M}_{ij}\tilde{Y}_i\omega_{ij}.$$

Furthermore, let $\hat{\boldsymbol{V}}_n = (\hat{\mu}_1, \hat{\mu}_2, \ldots, \hat{\mu}_n)$ and $\hat{\varsigma}$ be the eigenvector of $\hat{\boldsymbol{V}}_n \hat{\boldsymbol{V}}_n^\tau / n$ corresponding to its largest eigenvalue. Recalling that the first component of $\boldsymbol{\theta}_0$ is positive, we define the initial estimator of $\boldsymbol{\theta}_0$ as $\hat{\boldsymbol{\theta}}_{0,ini} = \text{sign}(\hat{\varsigma}_1)\hat{\varsigma}$, where $\hat{\varsigma}_1$ is the first component of $\hat{\varsigma}$. The reason for choosing the first eigenvector $\hat{\varsigma}$ is that $\hat{\mu}_j - \phi(U_j) g'(\boldsymbol{Z}_j^\tau \boldsymbol{\theta}_0)\boldsymbol{\theta}_0 \xrightarrow{P} 0$ as n goes to infinity. In other words, $\hat{\mu}_i$ is proportional to $\boldsymbol{\theta}_0$ in probability, then the first eigenvector $\hat{\varsigma}$ is also proportional to $\boldsymbol{\theta}_0$ in probability. The theoretical justification of this conclusion will be presented in Theorem 1.

2.1.2 Estimator of $\boldsymbol{\theta}_0$

To enhance the estimation efficiency, we update the kernel function $\omega_i(\boldsymbol{Z}_j, U_j)$ as $I_n^\theta(\boldsymbol{Z}_j, U_j) K_h(\boldsymbol{Z}_i^\tau \boldsymbol{\theta} - \boldsymbol{Z}_j^\tau \boldsymbol{\theta}) K_h(U_i - U_j) / \sum_{i=1}^n K_h(\boldsymbol{Z}_i^\tau \boldsymbol{\theta} - \boldsymbol{Z}_j^\tau \boldsymbol{\theta}) K_h(U_i - U_j)$, where $K(\cdot)$ is a univariate density function, h is a bandwidth, and $I_n^\theta(\boldsymbol{Z}_j, U_j) = I\{\frac{1}{n}\sum_{i=1}^n K_h(\boldsymbol{Z}_i^\tau \boldsymbol{\theta} - \boldsymbol{Z}_j^\tau \boldsymbol{\theta}) K_h(U_i - U_j) > c_0 c_1\}$. We implement the idea in the following steps.

Step 1. Given $\boldsymbol{\theta}^*$, let $\boldsymbol{M}_{ij,\theta^*} = \left(1, \tilde{\boldsymbol{X}}_i^\tau, (\boldsymbol{Z}_i - \boldsymbol{Z}_j)^\tau \boldsymbol{\theta}^*, (U_i - U_j)\right)^\tau$ and $\omega_{ij}^{\theta^*} = I_n^{\theta^*}(\boldsymbol{Z}_j, U_j) K_h(\boldsymbol{Z}_i^\tau \boldsymbol{\theta}^* - \boldsymbol{Z}_j^\tau \boldsymbol{\theta}^*) K_h(U_i - U_j) / \sum_{i=1}^n K_h(\boldsymbol{Z}_i^\tau \boldsymbol{\theta}^* - \boldsymbol{Z}_j^\tau \boldsymbol{\theta}^*) K_h (U_i - U_j)$. We obtain

$$\left(a_{j,\theta^*}, b_{j,\theta^*}, d_{1j,\theta^*}, d_{2j,\theta^*}\right)^\tau = \left\{\sum_{i=1}^n \boldsymbol{M}_{ij,\theta^*} \boldsymbol{M}_{ij,\theta^*}^\tau \omega_{ij}^{\theta^*}\right\}^{-1} \sum_{i=1}^n \boldsymbol{M}_{ij,\theta^*} \tilde{Y}_i \omega_{ij}^{\theta^*}. \tag{5}$$

Step 2. Using $\left(a_{j,\theta^*}, b_{j,\theta^*}, d_{1j,\theta^*}, d_{2j,\theta^*}, \boldsymbol{\theta}^*\right)^\tau$ from (5), and letting $\tilde{Y}_{ij}^* = \tilde{Y}_i - a_{j,\theta^*} - \tilde{\boldsymbol{X}}_i^\tau b_{1j,\theta^*} - d_{2j,\theta^*}(U_i - U_j)$, we have

$$\begin{aligned}\boldsymbol{\theta} = &\left\{\sum_{j=1}^n \sum_{i=1}^n d_{1j,\theta^*}^2 (\boldsymbol{Z}_i - \boldsymbol{Z}_j)(\boldsymbol{Z}_i - \boldsymbol{Z}_j)^\tau \omega_{ij}^{\theta^*}\right\}^{-1} \\ &\times \sum_{j=1}^n \sum_{i=1}^n d_{1j,\theta^*}(\boldsymbol{Z}_i - \boldsymbol{Z}_j)\tilde{Y}_{ij}^* \omega_{ij}^{\theta^*}.\end{aligned} \tag{6}$$

Standardize $\boldsymbol{\theta} = \text{sign}(\theta_1)\boldsymbol{\theta}/\|\boldsymbol{\theta}\|_2$, where θ_1 is the first element of $\boldsymbol{\theta}$ and $\text{sign}(\cdot)$ is the sign function.

We first use the initial estimator $\hat{\boldsymbol{\theta}}_{0,ini}$ as $\boldsymbol{\theta}^*$ to obtain $(a_{j,\theta^*}, b_{j,\theta^*}, d_{1j,\theta^*}, d_{2j,\theta^*})^\tau$ through (5), then update the estimator of $\boldsymbol{\theta}_0$ through (6). The final estimator $\hat{\boldsymbol{\theta}}_0$ can be obtained iteratively between (5) and (6) until convergence.

2.1.3 Estimator of $\boldsymbol{\beta}_0$

Under the following identifiability condition that was suggested by Şentürk and Müller (2005); Şentürk and Müller (2006)

$$E\phi(U) = 1, \quad E\psi(U) = \mathbf{I}_p, \tag{7}$$

where $\mathbf{I}_p$ is an $p \times p$ identical matrix, and the distorting effects vanish with no average distortion, namely, $E\tilde{Y} = EY$ and $E\tilde{\boldsymbol{X}} = E\boldsymbol{X}$. From (3) and the assumption of $U \perp\!\!\!\perp (\boldsymbol{X}, Y)$, Şentürk and Müller (2005) noted that, in the population level, $E\{\boldsymbol{\beta}_{0,r}(U)\tilde{X}_r\} = \boldsymbol{\beta}_{0,r}EX_r = \boldsymbol{\beta}_{0,r}E\tilde{X}_r$ for $r = 1, \ldots, p$. Under Condition (A4) in the Appendix that EX_rs are nonzero, we know that

$$\boldsymbol{\beta}_{0,r} = E\boldsymbol{\beta}_{0,r}(U)\tilde{X}_r / E\tilde{X}_r. \tag{8}$$

Motivated from (8), different from the binning method proposed by Şentürk and Müller (2005) to estimate $\boldsymbol{\beta}_{0,r}$s, we propose the following estimators

$$\hat{\boldsymbol{\beta}}_{0,r} = \sum_{j=1}^{n} \tilde{\boldsymbol{X}}_{j,r}\hat{b}_{j,\hat{\theta}_0,r} \Big/ \sum_{j=1}^{n} \tilde{\boldsymbol{X}}_{j,r}, \tag{9}$$

where $\hat{b}_{j,\hat{\theta}_0,r}$ is the r-th component of $\hat{b}_{j,\hat{\theta}_0}$, which is obtained from steps 1 and 2, and $\tilde{\boldsymbol{X}}_{j,r}$ is the r-th components of $\tilde{\boldsymbol{X}}_j$ for $j = 1, \ldots, n$ and $r = 1, \ldots, p$.

2.2 Asymptotic results for $\hat{\boldsymbol{\beta}}_0$ and $\hat{\boldsymbol{\theta}}_0$

We now present the sampling property of our proposed estimators, whose proofs are given in the Appendix. The first theorem establishes the consistency of the initial estimator $\hat{\boldsymbol{\theta}}_{0,ini}$.

Theorem 1 *Under Conditions* (A1)–(A3) *and* (A5), *if* $h_1 \to 0$, $h_2 \to 0$, $h_2^3/h_1 \to 0$ *and* $nh_1^{q+2}h_2/\log n \to \infty$ *as* $n \to \infty$, *then* $\boldsymbol{\theta}_{0,ini} \xrightarrow{P} \boldsymbol{\theta}_0$.

Theorem 1 indicates that the multi-dimensional kernel $H(\cdot)$ can ensure obtaining a consistent estimator of $\boldsymbol{\theta}_0$, although the associated bandwidth has a slower rate than that of the optimal bandwidth. As advocated by Xia and Härdle (2006), the MAVE procedure coupling with such an initial estimator would eventually provide a root-n consistent estimator of $\boldsymbol{\theta}_0$.

In what follows, $\mathbf{A}^{\otimes 2} = \mathbf{A}\mathbf{A}^\tau$ for any vector or matrix $\mathbf{A}$. Denote $\xi_\theta(t) = E(\boldsymbol{X}|\boldsymbol{Z}^\tau\boldsymbol{\theta} = t) = \big(\xi_{\theta,1}(t), \ldots, \xi_{\theta,p}(t)\big)$ where $\xi_{\theta,r}(t) = E(X_r|\boldsymbol{Z}^\tau\boldsymbol{\theta} = t)$, $\varsigma_\theta(t) = E\big(\mathbf{Z}|\boldsymbol{Z}^\tau\boldsymbol{\theta} = t\big)$, $\boldsymbol{S}_\theta(t) = E\big(\boldsymbol{X}^{\otimes 2}|\boldsymbol{Z}^\tau\boldsymbol{\theta} = t\big)$, $\mathcal{T}_\theta(t) = E\big(\mathbf{Z}\boldsymbol{X}^\tau|\boldsymbol{Z}^\tau\boldsymbol{\theta} = t\big)$, $\mathcal{W}_\theta(t) = \boldsymbol{S}_\theta(t) - \xi_\theta^{\otimes 2}(t)$, $\mathcal{V}_\theta(t) = E\big(\mathbf{Z}^{\otimes 2}|\boldsymbol{Z}^\tau\boldsymbol{\theta} = t\big) - \varsigma_\theta^{\otimes 2}(t)$, $F_\theta(\boldsymbol{Z}^\tau\boldsymbol{\theta}) = \operatorname{diag}\left(\frac{\xi_{\theta,1}(\boldsymbol{Z}^\tau\boldsymbol{\theta})}{EX_1}, \ldots, \frac{\xi_{\theta,p}(\boldsymbol{Z}^\tau\boldsymbol{\theta})}{EX_p}\right)$. Moreover, $L_{\beta_0} = \operatorname{diag}\left(\frac{\boldsymbol{\beta}_{0,1}}{EX_1}, \ldots, \frac{\boldsymbol{\beta}_{0,p}}{EX_p}\right)$, $\mathcal{D}_{\theta_0} = E\Big[\phi(U)^2\big\{g'(\boldsymbol{Z}^\tau\boldsymbol{\theta}_0)\big\}^2\mathcal{V}_{\theta_0}(\boldsymbol{Z}^\tau\boldsymbol{\theta}_0)$ $\mathcal{I}^{\theta_0}(\boldsymbol{Z}, U)\Big]$, where $\mathcal{I}^{\theta_0}(\boldsymbol{Z}, U) = I\{(\boldsymbol{Z}, U) \in \mathcal{Z} \times \mathcal{U}$, such that $f_{Z^\tau\theta_0}(\boldsymbol{Z}^\tau\theta_0)f_U(U) > c_0c_1\}$. We now present the asymptotic distribution of $\hat{\boldsymbol{\theta}}_0$.

Theorem 2 *Under Conditions* (A1)–(A8)*, if* $h \to 0$, $nh^2 \to \infty$, $nh^6 \to 0$ *as* $n \to \infty$, $\sqrt{n}(\hat{\boldsymbol{\beta}}_0 - \boldsymbol{\beta}_0)$ *is asymptotically normal with mean* $\mathbf{0}_p$ *and covariance*

$$\Sigma_{\beta_0} = L_{\beta_0} E\Big\{\big(\phi(U)\mathbf{I}_p - \psi(U)\big)\big\{E\boldsymbol{X}^{\otimes 2}\big\}\big(\phi(U)\mathbf{I}_p - \psi(U)\big)\Big\} L_{\beta_0} \\ +\sigma^2 E\Big\{\phi^2(U) F_{\theta_0}(\mathbf{Z}^\tau\boldsymbol{\theta}_0)\mathcal{W}_{\theta_0}^{-1}(\mathbf{Z}^\tau\boldsymbol{\theta}_0) F_{\theta_0}(\mathbf{Z}^\tau\boldsymbol{\theta}_0)\mathcal{I}^{\theta_0}(\mathbf{Z}_j, U_j)\Big\}.$$

Remark 1 From the expressions of F_{θ_0} and L_{θ_0}, we need Condition (A8) that $EX_r s$ are all nonzeros. This is because, in the population level, the estimator of $\boldsymbol{\beta}_{0,r}$ is obtained as $\boldsymbol{\beta}_{0,r} = E\{\tilde{X}_r\boldsymbol{\beta}_{0,r}(U)\}/EX_r$, $r = 1, \ldots, p$. This assumption was also imposed in Şentürk and Müller (2005); Şentürk and Müller (2006) for their binning method, and Cui et al. (2009) for their direct estimation method.

We next study the asymptotic distributions of $\hat{\boldsymbol{\theta}}_0$.

Theorem 3 *Under the conditions of Theorem 2,* $\sqrt{n}(\hat{\boldsymbol{\theta}}_0 - \boldsymbol{\theta}_0)$ *is asymptotically normal with mean* $\mathbf{0}_q$ *and the covariance matrix* Σ_{θ_0}*:*

$$\Sigma_{\theta_0} = \frac{\sigma^2}{4}\mathcal{D}_{\theta_0}^{-} E\Big[\phi^2(U)\Big\{\big(\mathcal{T}_{\theta_0}(\mathbf{Z}^\tau\boldsymbol{\theta}_0) - \varsigma_{\theta_0}(\mathbf{Z}^\tau\boldsymbol{\theta}_0)\xi_{\theta_0}^\tau(\mathbf{Z}^\tau\boldsymbol{\theta}_0)\big)\mathcal{W}_{\theta_0}^{-1}(\mathbf{Z}^\tau\boldsymbol{\theta}_0)\big(\boldsymbol{X} - \xi_{\theta_0}(\mathbf{Z}^\tau\boldsymbol{\theta}_0)\big) \\ +\big(\varsigma_{\theta_0}(\mathbf{Z}^\tau\boldsymbol{\theta}_0) - \mathbf{Z}\big)\Big\} g'(\mathbf{Z}^\tau\boldsymbol{\theta}_0)\mathcal{I}^{\theta_0}(\mathbf{Z}, U)\Big]^{\otimes 2}\mathcal{D}_{\theta_0}^{-},$$

where $\mathcal{D}_{\theta_0}^{-}$ *is the generalized inverse of* $\mathcal{D}_{\theta_0}$*.*

Remark 2 Since the term $\mathcal{V}_{\theta_0}(t)$ is the conditional variance of the covariate vector $\boldsymbol{Z}$ given a linear constraint in $\mathbf{Z}$, $\boldsymbol{\theta}_0^\tau\mathcal{V}_{\theta_0}(\mathbf{Z}^\tau\boldsymbol{\theta}_0)\boldsymbol{\theta}_0 = 0$. So the rank of $\mathcal{D}_{\theta_0}$ and Σ_{θ_0} equals to $q - 1$ instead of q, and these two matrices are not invertible.

3 Variable selection for $\boldsymbol{\beta}_0$

From (3), we have $\boldsymbol{\beta}_0(U) = \Xi(U)\boldsymbol{\beta}_0$, with, $\Xi(U) = \mathrm{diag}\left(\frac{\phi(U)}{\psi_1(U)}, \frac{\phi(U)}{\psi_2(U)}, \ldots, \frac{\phi(U)}{\psi_p(U)}\right)$. In other words, $\boldsymbol{\beta}_{0,r}(u) = \boldsymbol{\beta}_{0,r}\phi(U)/\psi_r(u)$. Under Condition (A7), imposed on the distorting functions $\phi(\cdot)$ and $\psi_r(\cdot)$s, we know that $\phi(u)/\psi_r(u)$s are nonzero on the support of U; i.e., $\boldsymbol{\beta}_{0,r}(u) = 0$ if and only if $\boldsymbol{\beta}_{0,r} = 0$. This means selecting the nonzero elements of $\boldsymbol{\beta}_0$ is equivalent to selecting nonzero components of $\boldsymbol{\beta}_0(u)$ or $\pm\boldsymbol{\beta}_0(u)/\|\boldsymbol{\beta}_0(u)\|$. In this context, selection of variables is equivalent to identifying the first sparse principal component (SPC) of $\boldsymbol{\beta}_0(u)\boldsymbol{\beta}_0(u)^\tau$ for each u on the support of U. This motivates us to solve the variable selection problem in CAPLSiM through finding the first SPC of $E\{\boldsymbol{\beta}_0(U)^{\otimes 2}\}$.

To get insights into the estimator of the first SPC of $E\{\boldsymbol{\beta}_0(U)^{\otimes 2}\}$, we consider an estimator of $E\{\boldsymbol{\beta}_0(U)^{\otimes 2}\}$, constructed as

$$\hat{E}\Big\{\boldsymbol{\beta}_0(U)^{\otimes 2}\Big\} = \frac{1}{n}\sum_{j=1}^n \hat{b}_{j,\hat{\theta}_0}^{\otimes 2},$$

where $\{\hat{b}_{1,\hat{\theta}_0}, \ldots, \hat{b}_{n,\hat{\theta}_0}\}$ are the estimates obtained from steps 1 and 2. Let $\hat{\eta}_n$ be an estimator of the first principal component of $E\{\boldsymbol{\beta}_0(U)^{\otimes 2}\}$ obtained by solving the following eigen-decomposition equation:

$$\left\{\frac{1}{n}\sum_{j=1}^{n}\hat{b}_{j,\hat{\theta}_0}^{\otimes 2}\right\}\hat{\eta}_n = \hat{\lambda}\hat{\eta}_n, \tag{10}$$

where $\hat{\lambda}$ is the largest eigenvalue of $\frac{1}{n}\sum_{j=1}^{n}\hat{b}_{j,\hat{\theta}_0}^{\otimes 2}$. Note that the eigen-decomposition (10) can be solved by minimizing a least square objective function (Li 2007):

$$\hat{\eta}_n = \arg\min_{v}\sum_{i=1}^{p}\|m_i - vv^{\tau}m_i\|^2 \text{ subject to } v^{\tau}v = 1, \tag{11}$$

where v is a $p \times 1$ vector, and m_i is the ith column of $\left\{\frac{1}{n}\sum_{j=1}^{n}\hat{b}_{j,\hat{\theta}_0}^{\otimes 2}\right\}^{1/2}$ for $i = 1, \ldots, p$, which satisfies $\left\{\frac{1}{n}\sum_{j=1}^{n}\hat{b}_{j,\hat{\theta}_0}^{\otimes 2}\right\}^{1/2}\left\{\frac{1}{n}\sum_{j=1}^{n}\hat{b}_{j,\hat{\theta}_0}^{\otimes 2}\right\}^{1/2} = \frac{1}{n}\sum_{j=1}^{n}\hat{b}_{j,\hat{\theta}_0}^{\otimes 2}$. To get a sparse solution of $\hat{\eta}_n$, we adopt the CISE idea proposed by Chen et al. (2010) as follows.

$$\breve{\eta}_n = \arg\min_{v}\left\{\sum_{i=1}^{p}\|m_i - vv^{\tau}m_i\|^2 + \sum_{r=1}^{p}\alpha_r|v_r|\right\}, \tag{12}$$

where $\alpha_r \geq 0$ serves as penalty parameters. In this paper, we use absolute value function $|\cdot|$ to achieve variable selection. As claimed in Chen et al. (2010), one may choose any positive convex function which is non-differentiable at the zero. The feature of non-differentiability at zero results in shrinking some elements of $\breve{\eta}_n$ to zero and consequently achieving selection of variables. The choice of turning parameters α_rs is discussed in Example 2 of Sect. 4. It is worth noting that the turning parameters α_rs are not necessarily the same for all coefficients to keep important variables in the final model. The resulting estimator $\breve{\eta}_n$ is the first SPC estimator of $E\{\boldsymbol{\beta}_0(U)^{\otimes 2}\}$. After we obtain the sparse solution $\breve{\eta}_n$ through (12), the resulting sparse estimator of $\boldsymbol{\beta}_0$ can be obtained as, for $r = 1, \ldots, p$,

$$\hat{\boldsymbol{\beta}}_{0,r}^{s} = \sum_{j=1}^{n}\tilde{X}_{j,r}\hat{b}_{j,\hat{\theta}_0,r}I\{\breve{\eta}_{n,r} \neq 0\}\Big/\sum_{j=1}^{n}\tilde{X}_{j,r}, \tag{13}$$

where $I\{\cdot\}$ is the indicator function and $\breve{\eta}_{n,r}$ is the r-th component of $\breve{\eta}_n$.

3.1 Algorithm for variable selection

With some algebraic derivations, the minimization of objective function (12) is equivalent to minimization of the objective function:

$$\mathcal{L} = -v \left\{ \frac{1}{n} \sum_{i=1}^{n} \hat{b}_{j,\hat{\theta}_0}^{\otimes 2} \right\} v + \sum_{r=1}^{p} \alpha_r |v_r|.$$

This can be achieved by using the algorithm for CISE proposed by Chen et al. (2010). First, we adopt the local quadratic approximation proposed by Fan and Li (2001) to overcome the non-differentiability of the absolute function $|\cdot|$.

Define $\rho(v) = \sum_{r=1}^{p} \alpha_r |v_r|$. The un-constrained first derivative of $\rho(v)$ with respect to the nonzero $p \times 1$ vector v is $\frac{\partial \rho(v)}{\partial v} = \text{diag}\left(\frac{\alpha_1}{|v_1|}, \ldots, \frac{\alpha_p}{|v_p|}\right) v$. Following the idea of Fan and Li (2001), the first derivative of $\rho(v)$ around $v^{(0)}$ can be approximately obtained by

$$\frac{\partial \rho(v)}{\partial v} \approx \text{diag}\left(\frac{\alpha_1}{\left|v_1^{(0)}\right|}, \ldots, \frac{\alpha_p}{\left|v_p^{(0)}\right|}\right) v \overset{\text{def}}{=} \mathrm{N}^{(0)} v. \tag{14}$$

The second-order Taylor expansion entails that, for some constant c_0^*,

$$\rho(v) \approx \frac{1}{2} v^\tau \mathrm{N}^{(0)} v + c_0^*. \tag{15}$$

Next, find $v^{(1)}$ by minimizing the objective function:

$$\mathcal{L}^{(1)} = -v^\tau \left\{ \frac{1}{n} \sum_{i=1}^{n} \hat{b}_{j,\hat{\theta}_0}^{\otimes 2} \right\} v + \frac{1}{2} v^\tau \mathrm{N}^{(0)} v. \tag{16}$$

In fact, the minimizing problem of $\mathcal{L}^{(1)}$ in (16) can be easily solved by the eigen-decomposition problem, that is, the solution of $v^{(1)}$ is the first eigenvector of $\left\{ \frac{1}{n} \sum_{i=1}^{n} \hat{b}_{j,\hat{\theta}_0}^{\otimes 2} \right\} - \frac{1}{2}\mathrm{N}^{(0)}$ corresponding to its largest eigenvalue. Next, let $v^{(1)}$ be the starting value and update $\rho(v)$ in (15). Iterate (14), (15) and (16) until we find the solution for v. Let δ be a pre-specified small positive constant (e.g. $\delta = 10^{-6}$), during the iteration. If $\left|v_r^{(k)}\right| < \delta$, then the r-th element of v is removed. As for the starting value $v^{(0)}$, we suggest the use of the estimator $\hat{\eta}_n$ from (10). The theoretical results for $\breve{\eta}_n$ are given in Theorem 4.

3.2 Theoretical properties

Let $\mathcal{A} = \left\{r : \boldsymbol{\beta}_{0,r} \neq 0\right\}$. $\mathcal{A}_n = \left\{r : \breve{\eta}_{n,r} \neq 0\right\}$. In other words, $\mathcal{A}$ is the set corresponding to the relevant variables of $\boldsymbol{X}$, and $\mathcal{A}_n$ represent the set corresponding to the variables of $\boldsymbol{X}$ selected out by the first SPC eigenvector $\breve{\eta}_n$.

Theorem 4 *In addition to the conditions of Theorem 2, we have*

(a) $\frac{1}{n} \sum_{j=1}^{n} \hat{b}_{j,\hat{\theta}_0}^{\otimes 2} - E\left\{\boldsymbol{\beta}_0(U)^{\otimes 2}\right\} = O_P(n^{-1/2})$.

(b) *Moreover, if we let* $\alpha_r = \alpha_0 |\hat{\eta}_{n,r}|^{-\varpi}$ *for some* $\varpi > 0$, *the turning parameter* α_0 *satisfies* $\alpha_0 \to 0$ *and* $\alpha_0 n^{\varpi/2} \to \infty$, *then* $P\left(\mathcal{A}_n = \mathcal{A}\right) \to 1$.

Remark 3 Theorem 4(a) indicates that the estimator $\frac{1}{n}\sum_{j=1}^{n}\hat{b}_{j,\hat{\theta}_0}^{\otimes 2}$ is a root-n consistent estimator of $E\{\boldsymbol{\beta}_0(U)^{\otimes 2}\}$. By the perturbation theory (Kato 1983), we know that the eigenvector $\hat{\eta}_n$ corresponding to its largest eigenvalue $\hat{\lambda}$ is also root-n consistent. Thus, we can use $\hat{\eta}_n$ as the starting value $v^{(0)}$ in Sect. 3.1. The foregoing algorithm works very fast basad on our numerical experience.

The reason for taking $\hat{\eta}_n$ as a starting value can be intuitively explained as follows. Let η be the first principle component of $E\left\{\boldsymbol{\beta}(U)^{\otimes 2}\right\}$. Suppose the true value $\boldsymbol{\beta}_{0,r} = 0$, then the r-th row and r-th column of $E\left\{\boldsymbol{\beta}(U)^{\otimes 2}\right\}$ are zero. This means the r-th element of η is zero, i.e., $\eta_r = 0$. As a consequence, using Theorem 4(a), we know that the r-th element of $\hat{\eta}_n$ satisfies $\hat{\eta}_{n,r} = \eta_r + O_P(n^{-1/2}) = O_P(n^{-1/2})$. Thus, as $\alpha_0 n^{\varpi/2} \to \infty$, we have $\alpha_r \xrightarrow{P} \infty$ and α_r will penalize $\breve{\eta}_{n,r}$ to 0. On the other hand, if $\boldsymbol{\beta}_{0,r'}$ is not zero and the corresponding eigenvector $\eta_{r'}$ is not zero, we have $\hat{\eta}_{n,r'} = \eta_{r'} + O_P(n^{-1/2}) = O_P(1)$. As $\alpha_0 \to 0$, we have that $\alpha_{r'} \xrightarrow{P} 0$ and $\alpha_{r'}$ will not penalize $\breve{\eta}_{n,r'}$ asymptotically. Hence, we choose $\hat{\eta}_n$ as our starting value to adaptively shrink η.

4 Numerical studies

In this section, we conduct simulation studies to assess the performance of the proposed methods. To estimate $(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0)$, the *Algorithm* proposed in Sect. 2.1 is adopted. We also investigate the finite sample performance of variable selection procedure in Sect. 3.1. We then apply our methods to analyze a real data set from a diabetes study.

4.1 Simulation studies

Example 1 We generated 500 data sets consisting of $n = 100, 200, 300$ and 400 observations, respectively, from the model:

$$Y = \boldsymbol{X}^{\tau}\boldsymbol{\beta}_0 + \sin(\boldsymbol{Z}^{\tau}\boldsymbol{\theta}_0) + \varepsilon, \tag{17}$$

where $\boldsymbol{\beta}_0 = (3.0, 1.5, 1.5, 0.5, 0.5)^{\tau}$, $\boldsymbol{\theta}_0 = (1, 1, 1)/\sqrt{3}$. The model error ε follows $N(0, 0.1^2)$ and is independent of $(\boldsymbol{X}, \boldsymbol{Z})$. The covariates $(\boldsymbol{X}, \boldsymbol{Z})^{\tau}$ follow $N_8\big((\mu_X, \mu_Z)^{\tau}, \Sigma\big)$ with $\mu_X = (5, \ldots, 5)^{\tau}$, $\mu_Z = (0, 0, 0)^{\tau}$ and $\Sigma = \left(0.5^{|i-j|}\right)_{1\le i,j\le 8}$. The confounding variable U was drawn from a Uniform(1, 6), the distortion function for the response Y is $\phi(U) = \frac{(U+1)^2}{22.3333}$, and those for the predictors $\boldsymbol{X}$ are $\psi_r(U) = \frac{U+r}{3.5+r}$, $r = 1, \ldots, p$. The constants in the distorting functions were chosen to ensure identifiability (7).

To obtain an initial estimator $\hat{\boldsymbol{\theta}}_{0,ini}$, we used the standard multivariate normal density function as the multivariate kernel $H(\cdot)$ and Epanechnikov kernel $L(t) = 0.75(1-t^2)_+$. As known, the optimal bandwidth for the multivariate setting is difficult to choose. In this simulation, we used $h_1 = n^{-1/7}$ and $h_2 = n^{-1/5}$, which meet the requirement on the bandwidths in Theorem 1. We verified that the resulting

Table 1 Simulation results for Example 1

	$\hat{\boldsymbol{\beta}}_{0,1}$	$\hat{\boldsymbol{\beta}}_{0,2}$	$\hat{\boldsymbol{\beta}}_{0,3}$	$\hat{\boldsymbol{\beta}}_{0,4}$	$\hat{\boldsymbol{\beta}}_{0,5}$
$n = 100$					
Bias	0.0110	0.0006	−0.0014	−0.0085	0.0015
SE	0.1094	0.0921	0.0910	0.0739	0.0832
$n = 200$					
Bias	−0.0003	−0.0012	−0.0087	−0.0043	−0.0044
SE	0.0721	0.0480	0.0525	0.0419	0.0376
$n = 300$					
Bias	0.0075	−0.0013	−0.0023	−0.0006	−0.0021
SE	0.0544	0.0390	0.0405	0.0274	0.0274
$n = 400$					
Bias	0.0090	0.0045	0.0004	−0.0010	−0.0011
SE	0.0467	0.0344	0.0354	0.0216	0.0213

The average bias and associated standard error for $\boldsymbol{\beta}_0$ in model (17)

initial values $\hat{\boldsymbol{\theta}}_{0,ini}$ were stable when we shifted several values around the selected bandwidths. After obtaining $\hat{\boldsymbol{\theta}}_{0,ini}$, we further obtained the final estimators of $(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0)$ by using the algorithm given in Sects. 2.1.2 and 2.1.3. We can use the Epanechnikov kernel $K(t) = 0.75(1 - t^2)_+$. Note that the "optimal" bandwidth or order $n^{-1/5}$ we used in this stage satisfies the conditions of Theorems 2 and 3. We used the following leave-one-out cross-validation to select bandwidth h.

Bandwidth selection For a given θ, define $\mathrm{CV}(h) = n^{-1}\sum_{j=1}^{n}\{\tilde{Y}_j - \tilde{\boldsymbol{X}}_j^{\tau} b_{j,\theta}^{\setminus j} - a_{j,\theta}^{\setminus j}\}^2$, where $a_{j,\theta}^{\setminus j}$ and $b_{j,\theta}^{\setminus j}$ are obtained through $\left(a_{j,\theta}^{\setminus j}, b_{j,\theta}^{\setminus j}, d_{1j,\theta}^{\setminus j}, d_{2j,\theta}^{\setminus j}\right)^{\tau} = \left(\sum_{i=1,i\neq j}^{n} \boldsymbol{M}_{ij,\theta}\boldsymbol{M}_{ij,\theta}^{\tau}\omega_{ij}^{\theta}\right)^{-1} \sum_{i=1,i\neq j}^{n} \boldsymbol{M}_{ij,\theta}\tilde{Y}_i\omega_{ij}^{\theta}$. The bandwidth h in each step is chosen as $\arg\min_h \mathrm{CV}(h)$.

The average bias and associated standard errors of $(\hat{\boldsymbol{\beta}}_0, \hat{\boldsymbol{\theta}}_0)$ are reported in Tables 1 and 2, respectively. The estimated $\hat{\boldsymbol{\beta}}_0$ are close to the true value $\boldsymbol{\beta}_0$, and the estimated values of single-index $\hat{\boldsymbol{\theta}}_0$ are also close to the true value $\boldsymbol{\theta}_0$ as sample size n increases. Moreover, larger sample sizes lead to smaller standard errors of $(\hat{\boldsymbol{\beta}}_0, \hat{\boldsymbol{\theta}}_0)$. For the estimation of $\boldsymbol{\theta}_0$ in Table 2, we present the mean and standard errors of the angle (in radians) between $\hat{\boldsymbol{\theta}}_0$ and $\boldsymbol{\theta}_0$ in Table 2. As expected, the mean and standard errors of the angle become smaller with the sample size. Both tables indicate that our estimation procedure performs well.

Example 2 In this example, we simulated 500 realizations, each consisting of $n = 300$, 400, and 500 random samples from model (17) with $\boldsymbol{\beta}_0 = (3, 2, 1.5, 0.2, 0.3, 0.15, 0, 0, 0, 0, 0, 0)^{\tau}$ and $\boldsymbol{\theta}_0 = (1, 1, 1)/\sqrt{3}$. The covariates $(\boldsymbol{X}, \boldsymbol{Z})^{\tau}$ follow normal distribution $N_{15}\left((\mu_X, \mu_Z)^{\tau}, \Sigma\right)$ with $\mu_X = (5, \ldots, 5)^{\tau}$, $\mu_Z = (0, 0, 0)^{\tau}$, and $\Sigma = \left(0.5^{|i-j|}\right)_{1\leq i,j\leq 15}$. The covariates $\boldsymbol{X}$ have 12 elements and the first 6 covariates of $\boldsymbol{X}$ are relevant to the model. We considered the following two cases:

Table 2 Simulation results for Example 1

	$\hat{\boldsymbol{\theta}}_{0,1}$	$\hat{\boldsymbol{\theta}}_{0,2}$	$\hat{\boldsymbol{\theta}}_{0,3}$	$\arccos(\hat{\boldsymbol{\theta}}_0, \boldsymbol{\theta}_0)$
$n = 100$				
Bias	−0.0162	−0.0706	−0.0457	0.2848
SE	0.1573	0.2672	0.2229	0.3393
$n = 200$				
Bias	0.0030	−0.0121	0.0013	0.0811
SE	0.0547	0.0588	0.0465	0.0468
$n = 300$				
Bias	0.0027	−0.0051	−0.0004	0.0493
SE	0.0318	0.0363	0.0288	0.0274
$n = 400$				
Bias	0.0008	−0.0013	−0.0013	0.0375
SE	0.0240	0.0264	0.0233	0.0202

The average bias and associated standard error for $\boldsymbol{\theta}_0$ in model (17)

Case 1. ε follows normal distribution $N(0, \sigma^2)$ with $\sigma = 0.1$, and is independent of $(\boldsymbol{X}, \boldsymbol{Z})$.

Case 2. ε follows $N\big(0, \sigma^2 \times (|\boldsymbol{X}_1| + |\boldsymbol{Z}_1|)\big)$ with $\sigma = 0.1$, where $\boldsymbol{X}_1$, $\boldsymbol{Z}_1$ are the first element of $\boldsymbol{X}$, $\boldsymbol{Z}$, respectively. In this setting, ε is correlated to $(\boldsymbol{X}, \boldsymbol{Z})$.

Choice of the penalty parameters α_0 and ϖ We used $\alpha_r = \alpha_0 |\hat{\eta}_{n,r}|^{-\varpi}$ in Theorem 4, where $\hat{\eta}_{n,r}$ is the r-th component of $\hat{\eta}_n$, defined in (10). (α, ϖ) are positive tuning parameters to be selected by minimizing the BIC-type criterion (Chen et al. 2010):

$$f(\alpha_0, \varpi) = -\breve{\eta}^{\tau}_{n(\alpha_0,\varpi)} \left(\frac{1}{n} \sum_{j=1}^{n} \hat{b}^{\otimes 2}_{j,\hat{\theta}_0} \right) \breve{\eta}_{n(\alpha_0,\varpi)} + \frac{\log n}{n} (N_{(\alpha_0,\varpi)} - 1), \qquad (18)$$

where $\breve{\eta}_{n(\alpha_0,\varpi)}$ is the estimator through (12) for given (α_0, ϖ), $N_{(\alpha_0,\varpi)}$ denotes the number of nonzero elements of $\breve{\eta}_{n(\alpha_0,\varpi)}$, and $\log n/n$ is the BIC-type factor suggested by Li (2007). The minimization of (18) can be solved by a two-dimensional grid search. In our simulation, the range of (α_0, ϖ) was selected to be wide enough so that the minimizer of $f(\alpha_0, \varpi)$ was approximately at the center of the range, and 20×20 grid points were set over the range of (α_0, ϖ).

In Table 3, we present summary statistics—ω_{u,β_0}, ω_{c,β_0}, ω_{o,β_0}, and Medse_{β_0} to assess how well our proposed method works. ω_{u,β_0}, ω_{c,β_0}, ω_{o,β_0} are the fractions of underfitted, correctly fitted and overfitted. In the case of overfitted, the labels "1", "2" and "$\geq$ 3" are the fraction of models including 1, 2 and more than 2 irrelevant covariates. Medse_{β_0} stands for the median of square error $\left\| \hat{\boldsymbol{\beta}}^s_0 - \boldsymbol{\beta}_0 \right\|^2$, where $\hat{\boldsymbol{\beta}}^s_0$ is defined in (13). The label "C_{β_0}" denotes the average number of the zero coefficients that were correctly set to zero, and the label "IN_{β_0}" denotes the average number of the nonzero coefficients that were incorrectly set to zero.

Table 3 Simulation results for Example 2: underfitted-ω_{u,β_0}, correctly fitted-ω_{c,β_0}, overfitted-ω_{o,β_0}, median of square error-Medse$_{\beta_0}$, zero coefficients of $\boldsymbol{\beta}_0$ that were correctly set to zero-C$_{\beta_0}$, and nonzero coefficients of $\boldsymbol{\beta}_0$ that were incorrectly set to zero-IN$_{\beta_0}$

n	ω_{u,β_0} (%)	ω_{c,β_0} (%)	ω_{o,β_0} (%)			Medse$_{\beta_0}$	No of zeros	
			"1" (%)	"2" (%)	"$\geq$ 3"(%)		C$_{\beta_0}$	IN$_{\beta_0}$
Case 1.								
300	13.00	75.00	12.00	0.00	0.00	0.0133	5.840	0.160
400	0.20	85.80	13.80	0.20	0.00	0.0064	5.858	0.002
500	0.00	92.60	7.40	0.00	0.00	0.0047	5.926	0.000
Case 2.								
300	4.80	74.60	18.60	2.60	0.00	0.0112	5.746	0.050
400	0.20	84.40	14.00	1.20	0.20	0.0080	5.828	0.002
500	0.00	89.20	10.40	0.40	0.00	0.0049	5.888	0.000

From Table 3, we can see that the first SPC $\breve{\eta}_n$ successfully distinguish zero and nonzeros of $\boldsymbol{\beta}_0$ in both two cases. The values of "C$_{\beta_0}$" and "IN$_{\beta_0}$" are close to the true value 6 and 0, respectively. Overall, the proportion of the model correctly fitted is over 70 %, and the proportion of underfitted and overfitted models are around 10 and 20 %. In particular, when the sample size increases to 500, the proportion of correctly fitted model is close to 90 % in both the cases. This indicates that our method can indeed identify the true model consistently. Furthermore, the Medse$_{\beta_0}$ decreases quickly with the sample size either in homogeneous or heteroscedastic error.

4.2 An empirical example

We apply the methods to analyze a data set from a diabetes study as an illustration. In this data set, there are 442 observations for diabetes patients with response variable Y (a quantitative measurement of disease progression one year after baseline) and covariates: age, sex, body mass index (BMI), average blood pressure (BP), and six blood serum measurements about total cholesterol, density level, tension glaucoma level, and glucose concentration denoted by TC, LDL, HDL, TCH, LTG, and GLU, respectively. This data set was analyzed by Efron et al. (2004) via the least angle regression (LARS). They used a linear regression model and "LARS" algorithm to select important factors (covariates) in disease progression. To avoid model misspecification, we applied PLSiM to analyze this data set. Here, we considered BMI as the potential confounding variable U, the six blood serum measurements and sex as $\boldsymbol{X}$, and age and BP as $\boldsymbol{Z}$.

The final estimate is $\hat{\boldsymbol{\theta}}_0 = (0.5010, -0.8655)$. Thus, the single index is 0.5010 age -0.8655 BP. In addition, the first SPC estimate $\breve{\eta}_n$ is (0.2502, 0.0344, -0.0366, -0.0144, 0.2327, -0.9384, 0.0033). This indicates that all seven elements of $\boldsymbol{X}$ should be kept in the final model. The estimates of $\boldsymbol{\beta}_0$ are (-24.4754, 1.4202, -1.3503, -2.5827, -9.3805, 33.7842, 0.8867). Finally, we used the local linear smoother

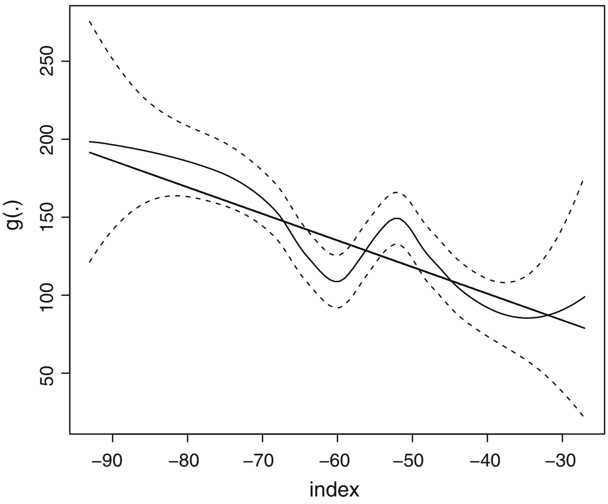

Fig. 1 The local linear estimator of $g(\cdot)$ (*solid line*) against estimated index 0.5010 age −0.8655 BP, along with the associated 95 % pointwise confidence intervals (*dotted lines*), and a linear fitting (*straight line*)

method (Fan and Gijbels 1996) to estimate $g(\cdot)$ based on the "synthesis" data $\left\{\tilde{Y}_j, \tilde{\boldsymbol{X}}_j^\tau \hat{b}_{j,\hat{\theta}_0}, \boldsymbol{Z}_j^\tau \hat{\boldsymbol{\theta}}_0\right\}_{j=1}^n$, and presented the estimated $\hat{g}(\cdot)$ in Fig. 1 along with the 95 % pointwise confidence intervals. For illustrative purpose, we fitted a linear regression for $\left\{\tilde{Y}_j - \tilde{\boldsymbol{X}}_j^\tau \hat{b}_{j,\hat{\theta}_0}, \boldsymbol{Z}_j^\tau \hat{\boldsymbol{\theta}}_0\right\}_{j=1}^n$ and displayed the straight line in Fig. 1, which is not encapsulated in the band. We therefore considered that the a nonlinear pattern of $g(\cdot)$ is more proper for this data set. We further used the test proposed by Stute and Zhu (2005) to check whether $\hat{g}(\cdot)$ has appropriately fitted this data set. The associated value of the test statistic is 1.4905 with p value 0.1361. This indicates that the single-index model $g(\cdot)$ is appropriate for this data set.

A Appendix

In this Appendix, we present the conditions, prepare several preliminary lemmas, and give the proofs of the main results.

A.1 Conditions

The following are the regularity conditions for our asymptotic results.

(A1) The functions $\xi_\theta(t)$, $\varsigma_\theta(t)$, $\boldsymbol{S}_\theta(t)$, $\mathcal{T}_\theta(t)$, $\mathcal{W}_\theta(t)$, $\mathcal{V}_\theta(t)$ defined in Sect. 2.2 are three times continuously differentiable with respect to t. Their third derivatives are uniformly Lipschitz continuous on $\mathcal{C} = \{t = z^\tau \boldsymbol{\theta} : z \in \mathcal{Z} \subset R^q, \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$, where $\mathcal{Z}$ is a compact support set. Furthermore, $\mathcal{W}_\theta(t)$ is invertible on $\mathcal{C}$, and the functions $\xi(z)$, $\boldsymbol{S}(z)$ defined in Lemma 1 have three times bounded derivatives on $\mathcal{Z}$.

(A2) With Probability 1, $\boldsymbol{Z}$ lies in the compact set $\mathcal{Z}$, such that the marginal density functions $f_Z(z)$ of $\boldsymbol{Z}$ and $f_{Z^\tau\theta}(z^\tau\boldsymbol{\theta})$ of $\boldsymbol{Z}^\tau\boldsymbol{\theta}$ for any $\|\boldsymbol{\theta}\| = 1$ have three times bounded derivatives; For some positive constant c_0, regions $\{z : f_Z(z) > c_0\}$ and $\{z : f_{Z^\tau\theta}(z^\tau\boldsymbol{\theta}) > c_0\}$ for all $\|\boldsymbol{\theta}\| = 1$ are nonempty. Moreover, there exists a positive constant c_1 such that the density function $f_U(u) > c_1$ on $\mathcal{U}$: the support of U.

(A3) $g(\cdot)$ has bounded, continuous third-order derivative, and is not constant on the support $\mathcal{C}$.

(A4) $E\varepsilon^4 < \infty$, and the covariance matrix of $(\boldsymbol{X}^\tau, \boldsymbol{Z}^\tau)^\tau$ is positive finite. Furthermore, $EX_r^4 < \infty, r = 1, \ldots, p$, $EZ_l^4 < \infty, l = 1, \ldots, q$.

(A5) Σ_{β_0} defined in Theorem 2 is positive definite with finite elements.

(A6) The multivariate kernel functions $H(\cdot)$ is a continuous and symmetric multivariate density function with bounded derivative and bounded support satisfying $\int z_r^2 H(z_1, \ldots, z_q)dz_r \neq 0$ and $\int |z_r|^j H(z_1, \ldots, z_q)dz_r < \infty$ for $j = 1, 2, 3, r = 1, \ldots, q$. The kernel functions $L(\cdot)$, $K(\cdot)$ are univariate continuous and symmetric density functions with bounded derivative and bounded support satisfying that $\int t^2 L(t)dt \neq 0$, $\int t^2 K(t)dt \neq 0$, and $\int |t|^j L(t)dt < \infty$, $\int |t|^j K(t)dt < \infty$ for $j = 1, 2, 3$. Moreover, the second derivative of $K(\cdot)$ is bounded on R^1.

(A7) The distorting functions $\phi(u)$, $\psi_r(u)$s have three bounded and continuous derivatives and are not equal to zero on the support of u. Moreover, $E\phi^4(U) < \infty$ and $E\psi_r^2(U) < \infty$.

(A8) $EX_r, r = 1, \ldots, p$, are bounded away from 0.

Condition (A1) is a mild smoothness condition on the involved functions. This condition is needed for the higher-order Taylor expansion, and entails the root-n consistency of $\boldsymbol{\beta}_0$ and $\boldsymbol{\theta}_0$. Condition (A2) entails the density functions $f_Z(z)$, $f_{Z^\tau\theta}(z^\tau\boldsymbol{\theta})$ and $f_U(u)$ positive, and guarantees the denominators involved in the weight functions ω_{ij} and ω_{ij}^θ in Algorithms are not equal to 0, as long as n is large enough. Condition (A3) is a mild smoothness condition on the function $g(\cdot)$, also used by Liang et al. (2010) and Xia and Härdle (2006). Conditions (A4)–(A5) are essential for the asymptotic results of the estimators of $\boldsymbol{\beta}_0$ and $\boldsymbol{\theta}_0$. Condition (A6) is a common assumption for nonparametric kernel smoothing. Condition (A7) is the condition imposed on the distorting functions. Analogous to condition (A1), the first statement of Condition (A7) is related to smoothness of the $\phi(u)$ and $\psi_r(u)$s. The second statement of Condition (A7) is a common condition on $\phi(u)$ and $\psi_r(u)$s in the covariate-adjusted models (Şentürk and Müller 2005; Şentürk and Müller 2006, 2009). In particular, this condition ensures the availability of variable selection of $\boldsymbol{\beta}_0$ by way of selecting nonzero component of $\boldsymbol{\beta}_0(u)$. The finiteness of $E\phi^4(U)$ is used for asymptotic results of Σ_{θ_0}. Moreover, $E\psi_r^2(U) < \infty$ guarantees $E\{\boldsymbol{\beta}_0(U)\boldsymbol{\beta}_0^\tau(U)\} < \infty$. Condition (A8) is necessary in the study of covariate-adjusted models. In order to estimate $\boldsymbol{\beta}_0$, EX_r is used in the denominator of the proposed estimators in the population level; see (8). This technical condition is also needed in the binning method (Şentürk and Müller 2005; Şentürk and Müller 2006, 2009), and the direct plug-in estimation (Zhang et al. 2012; Cui et al. 2009).

A.2 Technical lemmas

Lemma 1 *Let $(S_1, T_1), \ldots, (S_n, T_n)$ be i.i.d. random vectors, where T_is are scalar random variables and S_is are l-dimensional random vectors. Assume further that $E|T|^r < \infty$ and the $\sup_s \int |t|^r f(s,t)dt < \infty$, where f denote the joint density function of (S, T). Let $K^o(\cdot)$ be a bounded positive l-dimensional kernel function with bounded support, satisfying Lipschitz condition. Then*

$$\sup_{s\in\mathcal{D}} \left| n^{-1} \sum_{i=1}^{n} \{K^o_{h_*}(S_i - s)T_i - E[K^o_{h_*}(S_i - s)T_i]\} \right| = O_P\left(\{nh_*^l/\log(n)\}^{-1/2} \right),$$

provided that the bandwidth $h_* \to 0$ and $n^{2\epsilon-1}h_*^l \to \infty$ for some $\epsilon < 1 - r^{-1}$.

Proof This follows a direct result of Mack and Silverman (1982).

□

We introduce the following notation: $\zeta_L = \int v^2 L(v)\mathrm{d}v$, $\rho_r = \int z_r^2 H(z_1, \ldots, z_q)\, dz_1 \ldots dz_q$ for $r = 1, \ldots, q$, $\kappa = \mathrm{diag}(\rho_1, \ldots, \rho_q)$. For $j = 1, \ldots, n$, $\xi(\boldsymbol{Z}_j) = \big(\xi_1(\boldsymbol{Z}_j), \ldots, \xi_p(\boldsymbol{Z}_j)\big)^\tau$ where $\xi_r(\cdot) = E(X_r|\boldsymbol{Z} = \cdot)$ for $r = 1, \ldots, p$, $\boldsymbol{S}(\boldsymbol{Z}_j) = \big(s_{l,t}(\boldsymbol{Z}_j)\big)_{1\le l,t\le p}$ with $s_{l,t}(\cdot) = E(X_l X_t|\boldsymbol{Z} = \cdot)$.

Moreover, let $I(\boldsymbol{Z}_j, U_j) = I\{f_Z(\boldsymbol{Z}_j) f_U(U_j) > c_0c_1\}$, $I^\theta(\boldsymbol{Z}_j, U_j) = I\{f_{Z^\tau\theta}(\boldsymbol{Z}_j^\tau\boldsymbol{\theta}) f_U(U_j) > c_0c_1\}$, and $I^{\theta_0}(\boldsymbol{Z}_j, U_j) = I\{f_{Z^\tau\theta_0}(\boldsymbol{Z}_j^\tau\boldsymbol{\theta}_0) f_U(U_j) > c_0c_1\}$. Here $I\{\cdot\}$ is the indicator function, $f_Z(\cdot)$ is the marginal density of $\boldsymbol{Z}$, $f_{Z^\tau\theta}(\cdot)$, $f_{Z^\tau\theta_0}(\cdot)$ is the marginal density function of $\boldsymbol{Z}^\tau\boldsymbol{\theta}$, $\boldsymbol{Z}^\tau\boldsymbol{\theta}_0$, respectively. Let $U_{ij} = U_i - U_j$, $\boldsymbol{Z}_{ij} = \boldsymbol{Z}_i - \boldsymbol{Z}_j$ in the following, and

$$\omega_{ij} = \frac{I_n(\boldsymbol{Z}_j, U_j) H_{h_1}(\boldsymbol{Z}_{ij}) L_{h_2}(U_{ij})}{\sum_{i=1}^n H_{h_1}(\boldsymbol{Z}_{ij}) L_{h_2}(U_{ij})}, \quad \omega_{ij}^\theta = \frac{I_n^\theta(\boldsymbol{Z}_j, U_j) K_h(\boldsymbol{Z}_{ij}^\tau) K_h(U_{ij})}{\sum_{i=1}^n K_h(\boldsymbol{Z}_{ij}^\tau\theta) K_h(U_{ij})},$$

where $I_n(\boldsymbol{Z}_j, U_j) = I\{\frac{1}{n}\sum_{i=1}^n H_{h_1}(\boldsymbol{Z}_{ij}) L_{h_2}(U_{ij}) > c_0c_1\}$, $I_n^\theta(\boldsymbol{Z}_j, U_j) = I\{\frac{1}{n}\sum_{i=1}^n K_h(\boldsymbol{Z}_{ij}^\tau\boldsymbol{\theta})$
$K_h(U_{ij}) > c_0c_1\}$. Analogous to ω_{ij}^θ, we further define $\omega_{ij}^{\hat\theta_0}$ and $\omega_{ij}^{\theta_0}$ by substituting $\boldsymbol{\theta}$ with $\hat{\boldsymbol{\theta}}_0$ and $\boldsymbol{\theta}_0$.

Let $\delta_n = \max\{h_1^4, h_2^4\} + \log n/nh_1^q h_2$, we have the following asymptotic results.

Lemma 2 *Under Conditions* (A1)–(A3) *and* (A6)–(A7)*, uniformly in j, we have*

$$\sum_{i=1}^n \tilde{X}_i \omega_{ij} = \psi(U_j)\xi(\boldsymbol{Z}_j) I(\boldsymbol{Z}_j, U_j) + O_P\big(\delta_n^{1/2}\big) 1_p,$$

$$\sum_{i=1}^n \tilde{X}_i U_{ij} \omega_{ij} = O_P\big(h_2^2 + h_2\delta_n^{1/2}\big) 1_p,$$

$$\sum_{i=1}^n U_{ij}^2 \omega_{ij} = h_2^2 I(\boldsymbol{Z}_j, U_j)\big\{\zeta_L + O_P\big(\delta_n^{1/2}\big)\big\}, \quad \sum_{i=1}^n U_{ij}\omega_{ij} = O_P\big(h_2^2 + h_2\delta_n^{1/2}\big),$$

$$\sum_{i=1}^{n} \tilde{X}_i \tilde{X}_i^\tau \omega_{ij} = \psi(U_j) S(Z_j) \psi(U_j) I(Z_j, U_j) + O_P\left(\delta_n^{1/2}\right) 1_p 1_p^\tau,$$

$$\sum_{i=1}^{n} Z_{ij}^{\otimes 2} \omega_{ij} = h_1^2 I(Z_j, U_j)\left\{\kappa + O_P\left(\delta_n^{1/2}\right) 1_q 1_q^\tau\right\},$$

$$\sum_{i=1}^{n} U_{ij} Z_{ij} \omega_{ij} = O_P\left(h_1^2 h_2^2 + h_1 h_2 \delta_n^{1/2}\right) 1_q,$$

$$\sum_{i=1}^{n} Z_{ij} U_{ij}^2 \omega_{ij} = O_P\left(h_1^2 h_2^2 + h_1 h_2^2 \delta_n^{1/2} 1_q\right),$$

$$\sum_{i=1}^{n} Z_{ij} (Z_{ij}^\tau \boldsymbol{\theta}_0)^2 \omega_{ij} = O_P\left(h_1^4 + \delta_n^{1/2} h_1^3\right) 1_q,$$

$$\sum_{i=1}^{n} \tilde{X}_i Z_{ij}^\tau \omega_{ij} = O_P\left(h_1^2 + h_1 \delta_n^{1/2}\right) 1_p 1_q^\tau,$$

$$\sum_{i=1}^{n} Z_{ij} (Z_{ij}^\tau \theta_0) U_{ij} \omega_{ij} = O_P\left(h_1^2 h_2^2 + h_1^2 h_2 \delta_n^{1/2}\right) 1_q,$$

$$\sum_{i=1}^{n} Z_{ij} \omega_{ij} = O_P\left(h_1^2 + h_1 \delta_n^{1/2}\right) 1_q, \quad \sum_{i=1}^{n} \omega_{ij} Z_{ij} \tilde{X}_i^\tau \left\{\Xi(U_i) - \Xi(U_j)\right\} = O_P\left(h_1^2 h_2^2 + h_1 h_2 \delta_n^{1/2}\right) 1_q.$$

Proof We only prove the last equation. The rest can be justified in a similar way and thus we omit their proofs. Let $Z_{sij} = Z_{si} - Z_{sj}$, $s = 1, \ldots, q$. By the definition of ω_{ij}, we have that

$$\begin{aligned}
&\sum_{i=1}^{n} \omega_{ij} Z_{sij} \tilde{X}_{ri} \left\{\phi(U_i)/\psi_r(U_i) - \phi(U_j)/\psi_r(U_j)\right\} \\
&= \frac{\frac{1}{n}\sum_{i=1}^{n} H_{h_1}(Z_{ij}) L_{h_2}(U_{ij}) Z_{sij} X_{ri} \left\{\phi(U_i) - \frac{\phi(U_j)}{\psi_r(U_j)} \psi_r(U_i)\right\} I_n(Z_j, U_j)}{\frac{1}{n}\sum_{i=1}^{n} H_{h_1}(Z_{ij}) L_{h_2}(U_{ij})} \\
&\stackrel{\text{def}}{=} \frac{V_{n1} I_n(Z_j, U_j)}{V_{n2}}.
\end{aligned}$$

Step 1. Consider the denominator V_{n2}. By directly using Lemma 1, we have

$$\begin{aligned}
&\sup_{(z,u)\in \mathcal{Z}\times\mathcal{U}} \left| n^{-1} \sum_{i=1}^{n} [H_{h_1}(Z_i - z) L_{h_2}(U_i - u) - E\{H_{h_1}(Z_i - z) L_{h_2}(U_i - u)\}] \right| \\
&\quad = O_P\left(\sqrt{\log n / n h_1^q h_2}\right), \qquad (19)
\end{aligned}$$

Using the technique of change-of-variable and Taylor expansion, we have

$$E[H_{h_1}(\mathbf{Z}_i - z)L_{h_2}(U_i - u)] = f_Z(z)f_U(u) + O(h_1^2 + h_2^2). \tag{20}$$

Together with (19) and (20), it follows that $V_{n2} = f_Z(\mathbf{Z}_j)f_U(U_j) + O_P(\delta_n^{1/2})$.

Step 2. Consider V_{n1}. Similar to the analysis for V_{n2}, by using Lemma 1, we only need to calculate the following mean function. Firstly, note that $\int H(\omega)\omega_s d\omega = 0$. It follows that

$$\begin{aligned} h_1^{-2}E\{H_{h_1}(\mathbf{Z}-z)(Z_s-z_s)\xi_r(z)\} &= h_1^{-1}\int H(\omega)\omega_s\{\xi_r(z)+h_1\omega^\tau\partial\xi_r(z)/\partial z+O(h_1^2)\} \\ &\times\{f_Z(z)+h_1\omega^\tau\partial f_Z(z)/\partial z+O(h_1^2)\}d\omega = O(1+h_1^2). \end{aligned} \tag{21}$$

Secondly, similar to (21), by the Taylor expansion of $\phi(u+h_2t)$, $\psi_r(u+h_2t)$ and $\int L(t)tdt = 0$, we have $E\left[L_{h_2}(U-u)\{\phi(U)-\phi(u)\psi_r(U)/\psi_r(u)\}\right] = O(h_2^2 + h_2^4)$. Thus, $U \perp\!\!\!\perp \mathbf{Z}$ entails that

$$\begin{aligned} E\Big[(nh_1h_2)^{-1}\sum_{i=1}^n H_{h_1}(\mathbf{Z}_i - z)L_{h_2}(U_i - u)(Z_{si} - z_s)X_{ri}\{\phi(U_i) \\ - \phi(u)\psi_r(U_i)/\psi_r(u)\}\Big] = O(h_1h_2 + h_1^3h_2 + h_1h_2^3). \end{aligned} \tag{22}$$

Together with Lemma 1 and (22), we have $V_{n1} = O_P\Big(h_1^2h_2^2 + h_1^4h_2^2 + h_1^2h_2^4 + h_1h_2\sqrt{\log n/nh_1^q h_2}\Big)$. From (20) and Lemma 1, as $h_1 \to 0$, $h_2 \to 0$, $\frac{nh_1^q h_2}{\log n} \to \infty$, we have $\frac{1}{n}\sum_{i=1}^n H_{h_1}(\mathbf{Z}_{ij})L_{h_2}(U_{ij}) \xrightarrow{P} f_Z(\mathbf{Z}_j)f_U(U_j)$, so $I_n(\mathbf{Z}_j, U_j) \xrightarrow{P} I(\mathbf{Z}_j, U_j)$ and $V_{n2} = f_Z(\mathbf{Z}_j)f_U(U_j) + O_P(\delta_n^{1/2})$. As a consequence, together with the asymptotic results of V_{n1} and V_{n2}, we complete the proof.

Let $\delta_\theta = \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|$ and $\delta_n' = h^4 + \frac{\log n}{nh^2}$. Denote $\zeta_K = \int t^2K(t)dt$.

Lemma 3 *Under Conditions* (A1)–(A3) *and* (A6)–(A7)*, uniformly in j, we have*

$$\sum_{i=1}^n U_{ij}^2\omega_{ij}^\theta = h^2\{\zeta_K + O_P(\delta_\theta + \delta_n^{'1/2})\}I^{\theta_0}(\mathbf{Z}_j, U_j),$$

$$\sum_{i=1}^n U_{ij}\omega_{ij}^\theta = O_P\big(h^2 + h^2\delta_\theta + h\delta_n^{'1/2}\big),$$

$$\sum_{i=1}^n \tilde{\mathbf{X}}_i\omega_{ij}^\theta = \psi(U_j)\xi_{\theta_0}(\mathbf{Z}_j^\tau\boldsymbol{\theta}_0)I^{\theta_0}(\mathbf{Z}_j, U_j) + O_P(\delta_\theta + \delta_n^{'1/2}),$$

$$\sum_{i=1}^n \tilde{\mathbf{X}}_iU_{ij}\omega_{ij}^\theta = O_P\big(h^2+h^2\delta_\theta+h\delta_n^{'1/2}\big), \quad \sum_{i=1}^n \tilde{\mathbf{X}}_i\mathbf{Z}_{ij}^\tau\boldsymbol{\theta}\omega_{ij}^\theta = O_P\big(h^2+h^2\delta_\theta+h\delta_n^{'1/2}\big),$$

$$\sum_{i=1}^{n} \tilde{X}_i \tilde{X}_i^\tau \omega_{ij}^\theta = \psi(U_j) S_{\theta_0}(Z_j^\tau \theta_0) \psi(U_j) I^{\theta_0}(Z_j, U_j) + O_P(\delta_\theta + \delta_n^{'1/2}),$$

$$\sum_{i=1}^{n} Z_{ij}^\tau \theta \omega_{ij}^\theta = O_P\left(h^2 + h^2\delta_\theta + h\delta_n^{'1/2}\right), \quad \sum_{i=1}^{n} U_{ij} Z_{ij}^\tau \theta \omega_{ij}^\theta = O_P\left(h^4 + h^4\delta_\theta + h^2\delta_n^{'1/2}\right),$$

$$\sum_{i=1}^{n} \theta^\tau Z_{ij}^{\otimes 2} \theta \omega_{ij}^\theta = h^2\left\{\zeta_K + O_P(\delta_\theta + \delta_n^{'1/2})\right\} I^{\theta_0}(Z_j, U_j).$$

Proof The proof is similar to those of Lemma 2. We omit the details. □

A.3 Proof of Theorem 1

Denote $\Upsilon_j = \left(\hat{a}_j, \hat{b}_j^\tau, \hat{d}_{2j}, \hat{\mu}_j^\tau\right)^\tau - \left\{\phi(U_j) g(Z_j^\tau \theta_0), \beta_0(U_j)^\tau, \phi'(U_j) g(Z_j^\tau \theta_0), \phi(U_j) g'(Z_j^\tau \theta_0) \theta_0^\tau\right\}^\tau$ and $\pi = (0_{q\times(p+2)}, \mathbf{I}_q)$. We only need to show that $\pi \Upsilon_j = o_P(1)$, $j = 1, \ldots, n$. Let

$$\begin{aligned}\Delta\tilde{Y}_{ij} &= \tilde{X}_i^\tau\left\{\tilde{\beta}_0(U_i) - \tilde{\beta}_0(U_j)\right\} + \left\{\phi(U_i) g(Z_i^\tau \theta_0) - \phi(U_j) g(Z_j^\tau \theta_0)\right. \\ &\quad \left. - \phi'(U_j) g(Z_j^\tau \theta_0) U_{ij} - \phi(U_j) g'(Z_j^\tau \theta_0) Z_{ij}^\tau \theta_0\right\} + \phi(U_i)\varepsilon_i \\ &= \Delta\tilde{Y}_{ij}^{(1)} + \Delta\tilde{Y}_{ij}^{(2)} + \Delta\tilde{Y}_{ij}^{(3)}.\end{aligned}$$

In fact, we have

$$\Upsilon_j = \left\{\sum_{i=1}^{n} M_{ij} M_{ij}^\tau \omega_{ij}\right\}^{-1} \sum_{i=1}^{n} M_{ij}\left\{\Delta\tilde{Y}_{ij}^{(1)} + \Delta\tilde{Y}_{ij}^{(2)} + \Delta\tilde{Y}_{ij}^{(3)}\right\}\omega_{ij} \overset{\text{def}}{=} \Upsilon_j^{(1)} + \Upsilon_j^{(2)} + \Upsilon_j^{(3)}.$$

In the following, we show that the three terms $\pi \Upsilon_j^{(1)}$, $\pi \Upsilon_j^{(2)}$ and $\pi \Upsilon_j^{(3)}$ are $o_P(1)$.

Step 1.1 Investigate $\sum_{i=1}^{n} M_{ij} M_{ij}^\tau \omega_{ij}$. Let $\Gamma_n = \text{diag}\left(\mathbf{I}_{p+1}, 1/h_2, 1/h_1 \times \mathbf{I}_q\right)$. Lemma 2 entails

$$\sum_{i=1}^{n} \left\{\Gamma_n M_{ij} M_{ij}^\tau \Gamma_n\right\}\omega_{ij} = \aleph_j I(Z_j, U_j) + O_P(\delta_n^{1/2}) 1_{p+q+2} 1_{p+q+2}^\tau, \quad \text{where,}$$

$\aleph_j = \text{diag}(\Delta_j, \zeta_L, \kappa)$ with $\Delta_j = \begin{pmatrix} 1 & \xi^\tau(Z_j)\psi(U_j) \\ \psi(U_j)\xi(Z_j) & \psi(U_j) S(Z_j) \psi(U_j) \end{pmatrix}$. Thus, we have

$$\sum_{i=1}^{n} M_{ij} M_{ij}^\tau \omega_{ij} = \Gamma_n^{-1}\left\{\aleph_j I(Z_j, U_j) + O_P(\delta_n^{1/2}) 1_{p+q+2} 1_{p+q+2}^\tau\right\}\Gamma_n^{-1}. \quad (23)$$

Step 1.2 Consider $\Upsilon_j^{(1)}$. Using (23), Lemma 2 and the definition of $\boldsymbol{\pi}$, we can have

$$\begin{aligned}\boldsymbol{\pi}\Upsilon_j^{(1)} &= h_1^{-1}\kappa^{-1}I(\mathbf{Z}_j, U_j)\sum_{i=1}^{n}\omega_{ij}(\mathbf{Z}_{ij}/h_1)[\tilde{\boldsymbol{X}}_i^{\tau}\{\Xi(U_i)-\Xi(U_j)\}\boldsymbol{\beta}_0]\\ &\quad\times\left(1+O_P\left(\delta_n^{1/2}\right)\right)\\ &= O_P\left(h_2^2+\delta_n^{1/2}h_2/h_1\right)1_q. \end{aligned} \tag{24}$$

Step 1.3 Consider $\Upsilon_j^{(2)}$. By using Taylor expansion and Lemma 2, we obtain that

$$\begin{aligned}\boldsymbol{\pi}\Upsilon_j^{(2)} &= \frac{I(\mathbf{Z}_j, U_j)}{h_1\kappa}\sum_{i=1}^{n}\omega_{ij}(\mathbf{Z}_{ij}/h_1)\Big[\frac{1}{2}\phi(U_j)g''(\mathbf{Z}^{\tau}\boldsymbol{\theta}_0)\{\mathbf{Z}_{ij}^{\tau}\boldsymbol{\theta}_0\}^2\\ &\quad+\phi'(U_j)g'(\mathbf{Z}^{\tau}\boldsymbol{\theta}_0)U_{ij}\mathbf{Z}_{ij}^{\tau}\boldsymbol{\theta}_0+\frac{1}{2}\phi''(U_j)g(\mathbf{Z}^{\tau}\boldsymbol{\theta}_0)U_{ij}^2\Big]\\ &\quad\times\left(1+O_P\left(\delta_n^{1/2}\right)\right) = O_P\left(h_1^2+h_2^2+\delta_n^{1/2}h_2^2/h_1\right)1_q. \end{aligned} \tag{25}$$

Step 1.4 Consider $\Upsilon_j^{(3)}$. Note that $E\left\{\frac{1}{n}\sum_{i=1}^{n}H_{h_1}(\mathbf{Z}_{ij})L_{h_2}(U_{ij})\mathbf{Z}_{ij}\phi(U_i)\varepsilon_i\right\}=0$, and, for $s=1,\ldots,q$, $E\left[\left\{n^{-1}\sum_{i=1}^{n}H_{h_1}(\mathbf{Z}_{ij})L_{h_2}(U_{ij})Z_{sij}\phi(U_i)\varepsilon_i\right\}^2\middle|\mathbf{Z}_j\right]=O_P\left(\sqrt{h_1^2/nh_1^q h_2}\right)$. Thus, we have

$$\boldsymbol{\pi}\Upsilon_j^{(3)} = \frac{I(\mathbf{Z}_j, U_j)}{h_1\kappa}\sum_{i=1}^{n}\omega_{ij}(\mathbf{Z}_{ij}/h_1)\phi(U_i)\varepsilon_i(1+O_P(\delta_n^{1/2})) = O_P\left(\sqrt{1/nh_1^{q+2}h_2}\right). \tag{26}$$

Thus, from (24), (25) and (26), if $h_1\to 0, h_2\to 0, h_2^3/h_1\to 0$ and $nh_1^{q+2}h_2/\log n\to\infty$, then $\boldsymbol{\pi}\Upsilon_j^{(1)}$, $\boldsymbol{\pi}\Upsilon_j^{(2)}$ and $\boldsymbol{\pi}\Upsilon_j^{(3)}$ are all $o_P(1)$. As a consequence, we have $\boldsymbol{\pi}\Upsilon_j = o_P(1)$ and then $\hat{\mu}_j-\phi(U_j)g'(\mathbf{Z}_j^{\tau}\boldsymbol{\theta}_0)\boldsymbol{\theta}_0 \xrightarrow{P} 0$. Denote $c_{n0}=\frac{1}{n}\sum_{j=1}^{n}\left(\phi(U_j)g'(\mathbf{Z}_j^{\tau}\boldsymbol{\theta}_0)\right)^2$. We have

$$n^{-1}\hat{\boldsymbol{V}}_n\hat{\boldsymbol{V}}_n^{\tau}-c_{n0}\boldsymbol{\theta}_0\boldsymbol{\theta}_0^{\tau}\xrightarrow{P}\mathbf{0}_{q\times q}. \tag{27}$$

As the first eigenvector of $\boldsymbol{\theta}_0\boldsymbol{\theta}_0^{\tau}$ corresponding to its largest eigenvalue is $\boldsymbol{\theta}_0$ or $-\boldsymbol{\theta}_0$, consequently (27) and the perturbation theory (Kato 1983) yield $\hat{\boldsymbol{\theta}}_{0,ini}\xrightarrow{P}\boldsymbol{\theta}_0$. We complete the proof. □

A.4 Proof of Theorem 2

We split the proof into several steps to enhance the readability. Write

$$\begin{aligned}\Upsilon_{j,\hat{\theta}_0} &\stackrel{\text{def}}{=} \left(\hat{a}_{j,\hat{\theta}_0}, \hat{b}_{j,\hat{\theta}_0}^{\tau}, \hat{d}_{1j,\hat{\theta}_0}, \hat{d}_{2j,\hat{\theta}_0}\right)^{\tau}-\left\{\phi(U_j)g(\mathbf{Z}_j^{\tau}\boldsymbol{\theta}_0), \tilde{\boldsymbol{\beta}}_0(U_j)^{\tau},\right.\\ &\qquad\left.\phi(U_j)g'(\mathbf{Z}_j^{\tau}\boldsymbol{\theta}_0), \phi'(U_j)g(\mathbf{Z}_j^{\tau}\boldsymbol{\theta}_0)\right\}^{\tau}.\end{aligned}$$

Similar to handling Υ_j, we decompose $\Upsilon_{j,\hat{\theta}_0}$ as follows

$$\Upsilon_{j,\hat{\theta}_0} = \left(\sum_{i=1}^{n} \boldsymbol{M}_{ij,\hat{\theta}_0}\boldsymbol{M}^{\tau}_{ij,\hat{\theta}_0}\omega_{ij}^{\hat{\theta}_0}\right)^{-1} \sum_{i=1}^{n} \boldsymbol{M}_{ij,\hat{\theta}_0}\{\Delta\tilde{Y}^{(1)}_{ij} + \Delta\tilde{Y}^{(2)}_{ij,\hat{\theta}_0} + \Delta\tilde{Y}^{(3)}_{ij}\}\omega_{ij}^{\hat{\theta}_0}$$
$$\stackrel{\text{def}}{=} \Upsilon^{(1)}_{j,\hat{\theta}_0} + \Upsilon^{(2)}_{j,\hat{\theta}_0} + \Upsilon^{(3)}_{j,\hat{\theta}_0},$$

where $\Delta\tilde{Y}^{(2)}_{ij,\hat{\theta}_0} = \phi(U_i)g(\boldsymbol{Z}_i^{\tau}\boldsymbol{\theta}_0) - \phi(U_j)g(\boldsymbol{Z}_j^{\tau}\boldsymbol{\theta}_0) - \phi'(U_j)g(\boldsymbol{Z}_j^{\tau}\boldsymbol{\theta}_0)U_{ij} - \phi(U_j)g'(\boldsymbol{Z}_j^{\tau}\boldsymbol{\theta}_0)\boldsymbol{Z}_{ij}^{\tau}\hat{\boldsymbol{\theta}}_0$. Let $\boldsymbol{\pi}^* = (0_{p\times 1}, \mathbf{I}_p, 0_{p\times 1}, 0_{p\times 1})$, thus $\hat{b}_{j,\hat{\theta}_0} - \tilde{\boldsymbol{\beta}}_0(U_j) = \boldsymbol{\pi}^*\Upsilon_{j,\hat{\theta}_0} = \boldsymbol{\pi}^*\Upsilon^{(1)}_{j,\hat{\theta}_0} + \boldsymbol{\pi}^*\Upsilon^{(2)}_{j,\hat{\theta}_0} + \boldsymbol{\pi}^*\Upsilon^{(3)}_{j,\hat{\theta}_0}$. Let e_r be a $p \times 1$ vector with the rth position 1 and 0 elsewhere, $r = 1, \ldots, p$. We can further have

$$\hat{b}_{j,\hat{\theta}_0,r} - \tilde{\boldsymbol{\beta}}_{0,r}(U_j) = e_r^{\tau}\{\hat{b}_{j,\hat{\theta}_0} - \tilde{\boldsymbol{\beta}}_0(U_j)\} = e_r^{\tau}\left(\boldsymbol{\pi}^*\Upsilon^{(1)}_{j,\hat{\theta}_0} + \boldsymbol{\pi}^*\Upsilon^{(2)}_{j,\hat{\theta}_0} + \boldsymbol{\pi}^*\Upsilon^{(3)}_{j,\hat{\theta}_0}\right). \tag{28}$$

Let $\bar{\tilde{X}}_r = n^{-1}\sum_{j=1}^{n}\tilde{\boldsymbol{X}}_{j,r}$. Recall that $\tilde{\boldsymbol{\beta}}_{0,r}(U_j) = \boldsymbol{\beta}_{0,r}\frac{\phi(U_j)}{\psi_r(U_j)}$, $\tilde{\boldsymbol{X}}_{j,r} = \boldsymbol{X}_{j,r}\psi_r(U_j)$ and $\hat{\boldsymbol{\beta}}_{0,r} = n^{-1}\sum_{j=1}^{n}\tilde{\boldsymbol{X}}_{j,r}\hat{b}_{j,\hat{\theta}_0,r}/\bar{\tilde{X}}_r$. So that,

$$\hat{\boldsymbol{\beta}}_{0,r} - \boldsymbol{\beta}_{0,r} = n^{-1}\sum_{j=1}^{n}\tilde{\boldsymbol{X}}_{j,r}\{\tilde{\boldsymbol{\beta}}_{0,r}(U_j) - \boldsymbol{\beta}_{0,r}\}/\bar{\tilde{X}}_r$$
$$+n^{-1}\sum_{j=1}^{n}\tilde{\boldsymbol{X}}_{j,r}\{\hat{b}_{j,\hat{\theta}_0,r} - \tilde{\boldsymbol{\beta}}_{0,r}(U_j)\}/\bar{\tilde{X}}_r$$
$$\stackrel{\text{def}}{=} A^{(1)}_{\beta_0} + A^{(2)}_{\beta_0}.$$

Step 2.1 Consider $A^{(2)}_{\beta_0}$. Using expression (28), we re-write $A^{(2)}_{\beta_0}$ as follows.

$$A^{(2)}_{\beta_0} = e_r^{\tau}\frac{1}{n}\sum_{j=1}^{n}\tilde{\boldsymbol{X}}_{j,r}\boldsymbol{\pi}^*\Upsilon^{(1)}_{j,\hat{\theta}_0}/\bar{\tilde{X}}_r + e_r^{\tau}\frac{1}{n}\sum_{j=1}^{n}\tilde{\boldsymbol{X}}_{j,r}\boldsymbol{\pi}^*\Upsilon^{(2)}_{j,\hat{\theta}_0}/\bar{\tilde{X}}_r$$
$$+e_r^{\tau}\frac{1}{n}\sum_{j=1}^{n}\tilde{X}_{jr}\boldsymbol{\pi}^*\Upsilon^{(3)}_{j,\hat{\theta}_0}/\bar{\tilde{X}}_r$$
$$\stackrel{\text{def}}{=} A^{(2)}_{\beta_0}[1] + A^{(2)}_{\beta_0}[2] + A^{(2)}_{\beta_0}[3].$$

Let $\Gamma_{n,h} = \text{diag}\big(\mathbf{I}_{p+1}, 1/h, 1/h\big)$. By using Lemma 3, we obtain that $\sum_{i=1}^{n}\Big(\Gamma_{n,h}\boldsymbol{M}_{ij,\hat{\theta}_0}\boldsymbol{M}^{\tau}_{ij,\hat{\theta}_0}\Gamma_{n,h}\Big)\omega_{ij}^{\hat{\theta}_0} = \aleph_{j,\theta_0}I^{\theta_0}(\boldsymbol{Z}_j, U_j) + O_P\big(\delta_{\hat{\theta}_0} + \delta_n'^{1/2}\big)1_{p+3}1^{\tau}_{p+3}$, where $\aleph_{j,\theta_0} = \text{diag}(\Delta_{j,\theta_0}, \zeta_K\mathbf{I}_2)$ with

$$\Delta_{j,\theta_0} = \begin{pmatrix} 1 & \xi^{\tau}_{\theta_0}(\boldsymbol{Z}^{\tau}_j\boldsymbol{\theta}_0)\psi(U_j) \\ \psi(U_j)\xi_{\theta_0}(\boldsymbol{Z}^{\tau}_j\boldsymbol{\theta}_0) & \psi(U_j)\boldsymbol{S}^{\theta_0}(\boldsymbol{Z}^{\tau}_j\boldsymbol{\theta}_0)\psi(U_j) \end{pmatrix}.$$

Thus, we have $\sum_{i=1}^n \boldsymbol{M}_{ij,\hat\theta_0}\boldsymbol{M}^{\tau}_{ij,\hat\theta_0}\omega^{\hat\theta_0}_{ij} = \Gamma^{-1}_{n,h}\Big\{\aleph_{j,\theta_0} I^{\theta_0}(\boldsymbol{Z}_j, U_j) + O_P(\delta_{\hat\theta_0} + \delta'^{1/2}_n)1_{p+3} 1^{\tau}_{p+3}\Big\}\Gamma^{-1}_{n,h}$.

Step 2.1.1 Consider $A^{(2)}_{\beta_0}[1]$. Recalling the definition of $\boldsymbol{\pi}^*$, as $h \to 0$, $nh^2 \to \infty$ and $\hat{\boldsymbol{\theta}}_0 \xrightarrow{P} \boldsymbol{\theta}_0$, we have $O_P\big(\delta_{\hat\theta_0} + \delta'^{1/2}_n\big) = o_P(1)$, and $A^{(2)}_{\beta_0}[1] = e^{\tau}_r \frac{1}{n}\sum_{j=1}^n \tilde{\boldsymbol{X}}_{j,r}(0_{p\times 1}, \mathbf{I}_p)$ $\big\{\Delta^{-1}_{j,\theta_0} + o_P(1)\big\}\sum_{i=1}^n (1, \tilde{\boldsymbol{X}}^{\tau}_i)^{\tau}\big[\tilde{\boldsymbol{X}}^{\tau}_i\{\tilde{\boldsymbol{\beta}}_0(U_i) - \tilde{\boldsymbol{\beta}}_0(U_j)\}\big]\omega^{\hat\theta_0}_{ij}/\bar{\tilde{X}}_r$. Note that

$$\Delta^{-1}_{j,\theta_0} = \begin{pmatrix} 1 + \xi^{\tau}_{\theta_0}(\boldsymbol{Z}^{\tau}_j\boldsymbol{\theta}_0)\mathcal{W}^{-1}_{\theta_0}(\boldsymbol{Z}^{\tau}_j\boldsymbol{\theta}_0)\xi_{\theta_0}(\boldsymbol{Z}^{\tau}_j\boldsymbol{\theta}_0), & -\xi^{\tau}_{\theta_0}(\boldsymbol{Z}^{\tau}_j\boldsymbol{\theta}_0)\mathcal{W}^{-1}_{\theta_0}(\boldsymbol{Z}^{\tau}_j\boldsymbol{\theta}_0)\psi(U_j)^{-1} \\ -\psi(U_j)^{-1}\mathcal{W}^{-1}_{\theta_0}(\boldsymbol{Z}^{\tau}_j\boldsymbol{\theta}_0)\xi_{\theta_0}(\boldsymbol{Z}^{\tau}_j\boldsymbol{\theta}_0), & \psi(U_j)^{-1}\mathcal{W}^{-1}_{\theta_0}(\boldsymbol{Z}^{\tau}_j\boldsymbol{\theta}_0)\psi(U_j)^{-1} \end{pmatrix}. \tag{29}$$

From (29), we have

$$\begin{aligned} A^{(2)}_{\beta_0}[1] &= e^{\tau}_r\Bigg[\frac{1}{n}\sum_{j=1}^n\sum_{i=1}^n \tilde{\boldsymbol{X}}_{j,r}\psi(U_j)^{-1}\mathcal{W}^{-1}_{\theta_0}(\boldsymbol{Z}^{\tau}_j\boldsymbol{\theta}_0)\psi(U_j)^{-1}\tilde{\boldsymbol{X}}_i\tilde{\boldsymbol{X}}^{\tau}_i\{\tilde{\boldsymbol{\beta}}_0(U_i) - \tilde{\boldsymbol{\beta}}_0(U_j)\}\omega^{\hat\theta_0}_{ij} \\ &\quad - \frac{1}{n}\sum_{j=1}^n\sum_{i=1}^n \tilde{\boldsymbol{X}}_{j,r}\psi(U_j)^{-1}\mathcal{W}^{-1}_{\theta_0}(\boldsymbol{Z}^{\tau}_j\boldsymbol{\theta}_0)\xi_{\theta_0}(\boldsymbol{Z}^{\tau}_j\boldsymbol{\theta}_0)\tilde{\boldsymbol{X}}^{\tau}_i\{\tilde{\boldsymbol{\beta}}_0(U_i) - \tilde{\boldsymbol{\beta}}_0(U_j)\}\omega^{\hat\theta_0}_{ij}\Bigg]\Big/\bar{\tilde{X}}_r \\ &\overset{\text{def}}{=} e^{\tau}_r(B^{[1]}_{\beta_0} - B^{[2]}_{\beta_0})/\bar{\tilde{X}}_r. \end{aligned}$$

Similar to V_{n2} in Lemma 2, $I^{\theta}_n(\boldsymbol{Z}_j, U_j) \xrightarrow{P} I^{\theta_0}(\boldsymbol{Z}_j, U_j)$. For $e^{\tau}_r B^{[2]}_{\beta_0}$, we use projection of U-statistics (See Serfling 1980, Section 5.3.1). For any $\boldsymbol{\theta} \in \big\{\boldsymbol{\theta} : \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|_2 \le c_0 n^{-1/2}\big\}$,

$$\begin{aligned} &\frac{e^{\tau}_r}{n}\sum_{j=1}^n\sum_{i=1}^n \tilde{\boldsymbol{X}}_{j,r}\psi(U_j)^{-1}\mathcal{W}^{-1}_{\theta}(\boldsymbol{Z}^{\tau}_j\boldsymbol{\theta})\xi_{\theta}(\boldsymbol{Z}^{\tau}_j\boldsymbol{\theta})\tilde{\boldsymbol{X}}^{\tau}_i\{\tilde{\boldsymbol{\beta}}_0(U_i) - \tilde{\boldsymbol{\beta}}_0(U_j)\}\omega^{\theta}_{ij} \\ &= \sum_{i=1}^n \frac{e^{\tau}_r}{2n}\mathcal{W}^{-1}_{\theta}(\boldsymbol{Z}^{\tau}_i\boldsymbol{\theta})\xi_{\theta}(\boldsymbol{Z}^{\tau}_i\boldsymbol{\theta})\big\{\boldsymbol{X}_{i,r}\xi^{\tau}_{\theta}(\boldsymbol{Z}^{\tau}_i\boldsymbol{\theta}) - \xi_{\theta,r}(\boldsymbol{Z}^{\tau}_i\boldsymbol{\theta})\boldsymbol{X}^{\tau}_i\big\}\psi(U_i)\Xi'(U_i)\boldsymbol{\beta}_0 \\ &\quad \times\{f'_U(U_i)/f_U(U_i)\}h^2 I^{\theta_0}(\boldsymbol{Z}_j, U_j) + O_P\big(h^3\big) + o_P\big(n^{-1/2}\big) \\ &= O_P\big(n^{-1/2}h^2 + h^3\big), \end{aligned} \tag{30}$$

where $\Xi'(u) = \text{diag}\Big\{\Big(\frac{\phi(u)}{\psi_1(u)}\Big)', \ldots, \Big(\frac{\phi(u)}{\psi_1(u)}\Big)'\Big\}$. The fact that $E\big\{\boldsymbol{X}_{i,r}\xi^{\tau}_{\theta}(\boldsymbol{Z}^{\tau}_i\boldsymbol{\theta}) - \xi_{\theta,r}(\boldsymbol{Z}^{\tau}_i\boldsymbol{\theta})\boldsymbol{X}^{\tau}_i|\boldsymbol{Z}^{\tau}_i\boldsymbol{\theta}\big\} = 0$ was used in the last step above. If $h \to 0$ and $nh^6 \to 0$, the right-hand side of (30) is $o_P\big(n^{-1/2}\big)$. By using the Taylor expansion for $\mathcal{W}^{-1}_{\theta_0}(\boldsymbol{Z}^{\tau}_j\boldsymbol{\theta}_0)\xi_{\theta_0}(\boldsymbol{Z}^{\tau}_j\boldsymbol{\theta}_0)$ around $\hat{\boldsymbol{\theta}}_0$, together with (30), we have $B^{[2]}_{\beta_0} = o_P(n^{-1/2})$.

Using an analysis similar to $B^{[2]}_{\beta_0}$, we can show that $B^{[1]}_{\beta_0} = o_P(n^{-1/2})$. Using the fact that $\bar{\tilde{X}}_r = \frac{1}{n}\sum_{j=1}^{n} \tilde{X}_{j,r} = EX_r, a.s.$, we obtain that $A^{(2)}_{\beta_0}[1] = o_P(n^{-1/2})$.

Step 2.1.2 Consider $A^{(2)}_{\beta_0}[2]$ and $A^{(2)}_{\beta_0}[3]$. Similar to $A^{(2)}_{\beta_0}[1]$, we can have $A^{(2)}_{\beta_0}[2] = o_P(n^{-1/2})$.

Together with $\bar{\tilde{X}}_r - EX_r = O_P(n^{-1/2})$, $A^{(2)}_{\beta_0}[1] = o_P(n^{-1/2})$, $A^{(2)}_{\beta_0}[2] = o_P(n^{-1/2})$, and using the projection of U-statistics to $A^{(2)}_{\beta_0}[3]$, we can have

$$A^{(2)}_{\beta_0} = \frac{e_r^\tau}{nEX_r}\sum_{j=1}^{n} \phi(U_j)\xi_{\theta_0,r}(\mathbf{Z}_j^\tau\boldsymbol{\theta}_0)\mathcal{W}^{-1}_{\theta_0}(\mathbf{Z}_j^\tau\boldsymbol{\theta}_0)\big\{\boldsymbol{X}_j - \xi_{\theta_0}(\mathbf{Z}_j^\tau\boldsymbol{\theta}_0)\big\} I^{\theta_0}(\mathbf{Z}_j, U_j)\varepsilon_i + o_P(n^{-1/2}). \tag{31}$$

Step 2.2 From expression $A^{(1)}_{\beta_0}$ and the asymptotic expression (31), we have

$$\begin{aligned}\hat{\boldsymbol{\beta}}_{0,r} - \boldsymbol{\beta}_{0,r} &= \frac{1}{nEX_r}\sum_{j=1}^{n} \tilde{\boldsymbol{X}}_{j,r}\{\tilde{\boldsymbol{\beta}}_{0,r}(U_j) - \boldsymbol{\beta}_{0,r}\} \\ &+ \frac{e_r^\tau}{nEX_r}\sum_{j=1}^{n} \phi(U_j)\xi_{\theta_0,r}(\mathbf{Z}_j^\tau\boldsymbol{\theta}_0)\mathcal{W}^{-1}_{\theta_0}(\mathbf{Z}_j^\tau\boldsymbol{\theta}_0)\big\{\boldsymbol{X}_j - \xi_{\theta_0}(\mathbf{Z}_j^\tau\boldsymbol{\theta}_0)\big\} I^{\theta_0}(\mathbf{Z}_j, U_j)\varepsilon_i \\ &+ o_P(n^{-1/2}).\end{aligned} \tag{32}$$

Write $F_{i,\theta_0}(\mathbf{Z}_i^\tau\boldsymbol{\theta}_0) = \operatorname{diag}\big(\xi_{\theta_0,1}(\mathbf{Z}_i^\tau\boldsymbol{\theta}_0)/EX_1, \ldots, \xi_{\theta_0,p}(\mathbf{Z}_i^\tau\boldsymbol{\theta}_0)/EX_p\big)$. It follows from (32) that

$$\begin{aligned}\sqrt{n}(\hat{\boldsymbol{\beta}}_0 - \boldsymbol{\beta}_0) = \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\Big[& L_{\beta_0}\{\phi(U_j)\mathbf{I}_p - \psi(U_j)\}\boldsymbol{X}_j + \phi(U_j)F_{j,\theta_0}(\mathbf{Z}_j^\tau\boldsymbol{\theta}_0)\mathcal{W}^{-1}_{\theta_0}(\mathbf{Z}_j^\tau\boldsymbol{\theta}_0) \\ & \times\big\{\boldsymbol{X}_j - \xi_{\theta_0}(\mathbf{Z}_j^\tau\boldsymbol{\theta}_0)\big\} I^{\theta_0}(\mathbf{Z}_j, U_j)\varepsilon_j\Big] + o_P(1) \xrightarrow{L} N(\mathbf{0}_p, \Sigma_{\beta_0}),\end{aligned}$$

where L_{β_0} is defined in Theorem 2. We complete the proof. □

A.5 Proof of Theorem 3

Let $e_1 = \big(1, 0_{1\times(p+2)}\big)^\tau$, $e_2 = \big(0_{1\times(p+1)}, 1, 0\big)^\tau$ and $e_3 = \big(0_{1\times(p+2)}, 1\big)^\tau$, and let $\boldsymbol{\pi}^{**} = \big(e_1, e_2, e_3\big)^\tau$. We have $\boldsymbol{\pi}^{**}\Upsilon_{j,\hat{\theta}_0} = \big\{\hat{a}_{j,\hat{\theta}_0} - \phi(U_j)g(\mathbf{Z}_j^\tau\boldsymbol{\theta}_0), \hat{d}_{1j,\hat{\theta}_0} - \phi(U_j)g'(\mathbf{Z}_j^\tau\boldsymbol{\theta}_0), \hat{d}_{2j,\hat{\theta}_0} - \phi'(U_j)g(\mathbf{Z}_j^\tau\boldsymbol{\theta}_0)\big\}^\tau$, and

$$\hat{a}_{j,\hat{\theta}_0} - \phi(U_j)g(\mathbf{Z}_j^\tau\boldsymbol{\theta}_0) = \big\{1 + \xi^\tau_{\theta_0}(\mathbf{Z}_j^\tau\boldsymbol{\theta}_0)\mathcal{W}^{-1}_{\theta_0}(\mathbf{Z}_j^\tau\boldsymbol{\theta}_0)\xi_{\theta_0}(\mathbf{Z}_j^\tau\boldsymbol{\theta}_0)\big\} \times\sum_{i=1}^{n}\Big\{\Delta\tilde{Y}^{(1)}_{ij} + \Delta\tilde{Y}^{(2)}_{ij,\hat{\theta}_0} + \Delta\tilde{Y}^{(3)}_{ij}\Big\}\omega^{\hat{\theta}_0}_{ij}, \tag{33}$$

$$\hat{d}_{1j,\hat{\theta}_0} - \phi(U_j)g'(\mathbf{Z}_j^\tau\boldsymbol{\theta}_0) = \zeta_K^{-1}\sum_{i=1}^n(\mathbf{Z}_{ij}^\tau\hat{\boldsymbol{\theta}}_0/h^2)\Big\{\Delta\tilde{Y}_{ij}^{(1)} + \Delta\tilde{Y}_{ij,\hat{\theta}_0}^{(2)} + \Delta\tilde{Y}_{ij}^{(3)}\Big\}\omega_{ij}^{\hat{\theta}_0}, \tag{34}$$

$$\hat{d}_{2j,\hat{\theta}_0} - \phi'(U_j)g(\mathbf{Z}_j^\tau\boldsymbol{\theta}_0) = \zeta_K^{-1}\sum_{i=1}^n(U_{ij}/h^2)\Big\{\Delta\tilde{Y}_{ij}^{(1)} + \Delta\tilde{Y}_{ij,\hat{\theta}_0}^{(2)} + \Delta\tilde{Y}_{ij}^{(3)}\Big\}\omega_{ij}^{\hat{\theta}_0}. \tag{35}$$

Thus, letting $\tilde{Y}_{ij}^{**} = \tilde{Y}_i - \hat{a}_{j,\hat{\theta}_0} - \tilde{\boldsymbol{X}}_i^\tau\hat{b}_{j,\hat{\theta}_0} - \hat{d}_{2j,\hat{\theta}_0}U_{ij}$, we have

$$n^{-1}\sum_{i=1}^n\sum_{j=1}^n\hat{d}_{1j,\hat{\theta}_0}\mathbf{Z}_{ij}\tilde{Y}_{ij}^{**}\omega_{ij}^{\hat{\theta}_0} = n^{-1}\sum_{i=1}^n\sum_{j=1}^n\hat{d}_{1j,\hat{\theta}_0}^2\mathbf{Z}_{ij}^{\otimes 2}\boldsymbol{\theta}_0 + D_{n1} + D_{n2}, \tag{36}$$

where $D_{n1} = \frac{1}{n}\sum_{j=1}^n\sum_{i=1}^n\hat{d}_{1j,\hat{\theta}_0}\mathbf{Z}_{ij}\phi(U_i)\varepsilon_i\omega_{ij}^{\hat{\theta}_0}$ and

$$\begin{aligned}
D_{n2} &= n^{-1}\sum_{j=1}^n\sum_{i=1}^n\Big[\big\{\phi(U_j)g(\mathbf{Z}_j^\tau\boldsymbol{\theta}_0) - \hat{a}_{j,\hat{\theta}_0}\big\} + \tilde{\boldsymbol{X}}_i^\tau\big\{\tilde{\boldsymbol{\beta}}_0(U_i) - \tilde{\boldsymbol{\beta}}_0(U_j)\big\}\\
&\quad + \tilde{\boldsymbol{X}}_i^\tau\big\{\tilde{\boldsymbol{\beta}}_0(U_j) - \hat{b}_{j,\hat{\theta}_0}\big\} + \big\{\phi(U_i)g(\mathbf{Z}_i^\tau\boldsymbol{\theta}_0) - \phi(U_j)g(\mathbf{Z}_j^\tau\boldsymbol{\theta}_0)\\
&\quad - \hat{d}_{2j,\hat{\theta}_0}U_{ij} - \hat{d}_{1j,\hat{\theta}_0}\mathbf{Z}_{ij}^\tau\boldsymbol{\theta}_0\big\}\Big]\hat{d}_{1j,\hat{\theta}_0}\mathbf{Z}_{ij}\omega_{ij}^{\hat{\theta}_0}\\
&\stackrel{\text{def}}{=} \sum_{t=1}^4 D_{n2}[t].
\end{aligned}$$

From (36), we have $\sqrt{n}\big(\hat{\boldsymbol{\theta}}_0 - \boldsymbol{\theta}_0\big) = \sqrt{n}\Big\{\frac{1}{n}\sum_{i=1}^n\sum_{j=1}^n\hat{d}_{1j,\hat{\theta}_0}^2\mathbf{Z}_{ij}^{\otimes 2}\omega_{ij}^{\hat{\theta}_0}\Big\}^-(D_{n1}+D_{n2})$.

Step 3.1 Investigate $\frac{1}{n}\sum_{i=1}^n\sum_{j=1}^n\hat{d}_{1j,\hat{\theta}_0}^2\mathbf{Z}_{ij}^{\otimes 2}\omega_{ij}^{\hat{\theta}_0}$. First, we derive the order of $\hat{d}_{1j,\hat{\theta}_0} - \phi(U_j)g'(\mathbf{Z}_j^\tau\boldsymbol{\theta}_0)$. From (34), similar to (25), we use the results of Lemma 3 and obtain that

$$\zeta^{-1}\sum_{i=1}^n(\mathbf{Z}_{ij}^\tau\hat{\boldsymbol{\theta}}_0/h^2)\big\{\Delta\tilde{Y}_{ij}^{(1)} + \Delta\tilde{Y}_{ij}^{(2)}\big\}\omega_{ij}^{\hat{\theta}_0} = O_P\big(h^2\delta_{\hat{\theta}_0} + \delta_n'^{1/2}\big). \tag{37}$$

$E\big[h^{-1}K'\big(\mathbf{Z}_{ij}^\tau\boldsymbol{\theta}_0/h\big)K_h(U_{ij})\mathbf{Z}_{ij}\big|\mathbf{Z}_j^\tau\boldsymbol{\theta}_0, U_i\big] = \big\{f_{Z^\tau\theta_0}(\mathbf{Z}_j^\tau\boldsymbol{\theta}_0)f_U(U_j)\varsigma_{\theta_0}'(\mathbf{Z}_j^\tau\boldsymbol{\theta}_0)\int tK'(t)dt\big\}h + O_P(h^3)$, then, Lemma 1 and the Taylor expansion of $K_h\big(\mathbf{Z}_{ij}^\tau\hat{\boldsymbol{\theta}}_0\big)$ at $\boldsymbol{\theta}_0$ entail that

$$\begin{aligned}
n^{-1}\sum_{i=1}^n K_h\big(\mathbf{Z}_{ij}^\tau\hat{\boldsymbol{\theta}}_0\big)K_h(U_{ij}) &= f_{Z^\tau\theta_0}(\mathbf{Z}_j^\tau\boldsymbol{\theta}_0)f_U(U_j)\\
&\quad + (nh)^{-1}\sum_{i=1}^n K_h'\big(\mathbf{Z}_{ij}^\tau\boldsymbol{\theta}_0\big)K_h(U_{ij})(\mathbf{Z}_{ij}^\tau\delta_{\hat{\theta}_0})\\
&\quad + O_P\big(\delta_n'^{1/2}\big) = f_{Z^\tau\theta_0}(\mathbf{Z}_j^\tau\boldsymbol{\theta}_0)f_U(U_j) + O_P\big(\delta_n'^{1/2} + \delta_{\hat{\theta}_0}\big).
\end{aligned} \tag{38}$$

Moreover, similar to the analysis of Theorem 3 (Fan and Gijbels 1996, pp. 101–103), we have $(nh^2)^{-1}\sum_{i=1}^{n}\boldsymbol{Z}_{ij}^{\tau}\boldsymbol{\theta}_0\phi(U_i)\varepsilon_i K_h\big(\boldsymbol{Z}_{ij}^{\tau}\boldsymbol{\theta}_0\big)K_h(U_{ij}) = O_P(\sqrt{1/nh^4})$, and we can also have that $(nh^2)^{-1}\sum_{i=1}^{n}\boldsymbol{Z}_{ij}\phi(U_i)\varepsilon_i K_h\big(\boldsymbol{Z}_{ij}^{\tau}\boldsymbol{\theta}_0\big)K_h(U_{ij}) = O_P(\sqrt{1/nh^4})$. Together with (38), we can have

$$\zeta_K^{-1}\sum_{i=1}^{n}(\boldsymbol{Z}_{ij}\hat{\boldsymbol{\theta}}_0/h^2)\Delta\tilde{Y}_{ij}^{(3)}\omega_{ij}^{\hat{\theta}_0} = O_P(\sqrt{1/nh^4}) = O_P\big(\delta_n^{'1/2}/h\big). \tag{39}$$

(37) and (39) indicate that $\hat{d}_{1j,\hat{\theta}_0} - \phi(U_j)g'(\boldsymbol{Z}_j^{\tau}\boldsymbol{\theta}_0) = O_P\big(\delta_{\hat{\theta}_0} + \delta_n^{'1/2}/h\big)$. As a consequence, the condition $h \to 0$, $nh^4/\log n \to \infty$ and $\hat{\boldsymbol{\theta}}_0 \xrightarrow{P} \boldsymbol{\theta}_0$ ensure that $\hat{d}_{1j,\hat{\theta}_0} - \phi(U_j)g'(\boldsymbol{Z}_j^{\tau}\boldsymbol{\theta}_0) = o_P(1)$ for $j = 1,\ldots,n$. By using the projection of U-statistics, we have $\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\big\{\phi(U_j)g'(\boldsymbol{Z}_j^{\tau}\boldsymbol{\theta})\big\}^2\boldsymbol{Z}_{ij}^{\otimes 2}\omega_{ij}^{\theta} = \frac{2}{n}\sum_{j=1}^{n}\big\{\phi(U_j)g'(\boldsymbol{Z}_j^{\tau}\boldsymbol{\theta})\big\}^2\mathcal{V}_{\theta}(\boldsymbol{Z}_j^{\tau}\boldsymbol{\theta})I_n^{\theta}(\boldsymbol{Z}_j, U_j) + o_P(1)$. Together with (38), we obtain $I_n^{\hat{\theta}}(\boldsymbol{Z}_j, U_j) \xrightarrow{P} I^{\theta_0}(\boldsymbol{Z}_j, U_j)$ and

$$n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n}\hat{d}_{1j,\hat{\theta}_0}^2\boldsymbol{Z}_{ij}^{\otimes 2}\omega_{ij}^{\hat{\theta}_0} \xrightarrow{P} 2E\big[\phi(U)^2\big\{g'(\boldsymbol{Z}^{\tau}\boldsymbol{\theta}_0)\big\}^2\mathcal{V}(\boldsymbol{Z}^{\tau}\boldsymbol{\theta}_0)\mathcal{I}^{\theta_0}(\boldsymbol{Z}, U)\big] \overset{\text{def}}{=} \mathcal{D}_{\theta_0}. \tag{40}$$

Step 3.2 Consider D_{n1}. Similar to the analysis of (40), the projection of U-statistics and Taylor expansion at $\boldsymbol{\theta}_0$ entail that

$$D_{n1} = n^{-1}\sum_{j=1}^{n}I^{\theta_0}(\boldsymbol{Z}_j, U_j)g'(\boldsymbol{Z}_j^{\tau}\boldsymbol{\theta}_0)\phi^2(U_j)\big\{\varsigma_{\theta_0}(\boldsymbol{Z}_j^{\tau}\boldsymbol{\theta}_0) - \boldsymbol{Z}_j\big\}\varepsilon_j + o_P(n^{-1/2}).$$

Step 3.3 Consider D_{n2} in several sub-steps. In the following, we show that $D_{n2}[3]$ asymptotically follows a normal distribution, and further show that $D_{n2}[1]$, $D_{n2}[2]$ and $D_{n2}[4]$ are all $o_P(n^{-1/2})$.

Step 3.3.1 Deal with $D_{n2}[3]$. Using $\hat{d}_{1j,\hat{\theta}_0} - \phi(U_j)g'(\boldsymbol{Z}_j^{\tau}\boldsymbol{\theta}_0) = O_P\big(\delta_{\hat{\theta}_0} + \delta_n^{'1/2}\big)$ in step 3.1, we have

$$\begin{aligned} D_{n2}[3] &= n^{-1}\sum_{j=1}^{n}\sum_{i=1}^{n}\phi(U_j)g'(\boldsymbol{Z}_j^{\tau}\boldsymbol{\theta}_0)\boldsymbol{Z}_{ij}\tilde{\boldsymbol{X}}_i^{\tau}\{\boldsymbol{\beta}_0(U_j) - \hat{b}_{j,\hat{\theta}_0}\}\omega_{ij}^{\hat{\theta}_0} + D_{n2}^R[3] \\ &\overset{\text{def}}{=} D_{n2}^M[3] + D_{n2}^R[3]. \end{aligned}$$

From Lemma 3, we know that $\sum_{i=1}^{n}\boldsymbol{Z}_{ij}\tilde{\boldsymbol{X}}_i\omega_{ij}^{\theta} = \big\{\mathcal{T}_{\theta}(\boldsymbol{Z}_j^{\tau}\boldsymbol{\theta}) - \boldsymbol{Z}_j\xi_{\theta}^{\tau}(\boldsymbol{Z}_j^{\tau}\boldsymbol{\theta})\big\}\psi(U_j)I^{\theta}(\boldsymbol{Z}_j, U_j) + O_P(\delta_n^{'1/2})$. Moreover, using arguments similar to those for $A_{\beta_0}^{(2)}[1]$ in step 2.1.1, we have

$$n^{-1}\sum_{j=1}^{n} I^{\theta_0}(\boldsymbol{Z}_j,U_j)\phi(U_j)g'(\boldsymbol{Z}_j^{\tau}\boldsymbol{\theta}_0)\{\mathcal{T}_{\theta_0}(\boldsymbol{Z}_j^{\tau}\boldsymbol{\theta}_0)-\boldsymbol{Z}_j\xi_{\theta_0}^{\tau}(\boldsymbol{Z}_j^{\tau}\boldsymbol{\theta}_0)\}\psi(U_j)$$
$$\times\{\boldsymbol{\pi}^*\Upsilon_{j,\hat{\theta}_0}^{(1)}+\boldsymbol{\pi}^*\Upsilon_{j,\hat{\theta}_0}^{(2)}\}\times\big(\mathbf{I}_q+O_P(\delta_{\hat{\theta}_0})+O_P(\delta_n^{'1/2})\big)=o_P(n^{-1/2}). \quad (41)$$

Thus, together with (28) and (41), similar to $A_{\beta_0}^{(2)}[3]$ in step 2.1.3, we can obtain that

$$D_{n2}^{M}[3]=n^{-1}\sum_{j=1}^{n} I^{\theta_0}(\boldsymbol{Z}_j,U_j)\phi^2(U_j)g'(\boldsymbol{Z}_j^{\tau}\boldsymbol{\theta}_0)\Big\{\mathcal{T}_{\theta_0}(\boldsymbol{Z}_j^{\tau}\boldsymbol{\theta})-\varsigma_{\theta_0}(\boldsymbol{Z}_j^{\tau}\boldsymbol{\theta}_0)\xi_{\theta}^{\tau}(\boldsymbol{Z}_j^{\tau}\boldsymbol{\theta})\Big\}$$
$$\times\mathcal{W}_{\theta_0}^{-1}(\boldsymbol{Z}_j^{\tau}\boldsymbol{\theta}_0)\{\boldsymbol{X}_j-\xi_{\theta_0}(\boldsymbol{Z}_j^{\tau}\boldsymbol{\theta}_0)\}\varepsilon_j+o_P(n^{-1/2}). \quad (42)$$

Furthermore, analogous to (41), we can have $D_{n2}^{R}[3]=o_P(n^{-1/2})$.

Step 3.3.2 Deal with $D_{n2}[1]$. Note that $D_{n2}[1]=\frac{1}{n}\sum_{j=1}^{n}\hat{d}_{1j,\hat{\theta}_0}\{\phi(U_j)g(\boldsymbol{Z}_j^{\tau}\boldsymbol{\theta}_0)-\hat{a}_{j,\hat{\theta}_0}\}\sum_{i=1}^{n}\boldsymbol{Z}_{ij}\omega_{ij}^{\hat{\theta}_0}$, and $\sum_{i=1}^{n}(\boldsymbol{Z}_i-\boldsymbol{Z}_j)\omega_{ij}^{\hat{\theta}_0}=\big(\varsigma_{\theta_0}(\boldsymbol{Z}_j^{\tau}\boldsymbol{\theta}_0)-\boldsymbol{Z}_j\big)I^{\theta_0}(\boldsymbol{Z}_j,U_j)+O_P\big(\delta_{\hat{\theta}_0}+\delta_n^{'1/2}\big)$ and $\hat{d}_{1j,\hat{\theta}_0}-\phi(U_j)g'(\boldsymbol{Z}_j^{\tau}\boldsymbol{\theta}_0)=O_P\big(\delta_{\hat{\theta}_0}+\delta_n^{'1/2}/h\big)$. By using (33), we can have

$$D_{n2}[1]=n^{-1}\sum_{j=1}^{n}\phi(U_j)g'(\boldsymbol{Z}_j^{\tau}\boldsymbol{\theta}_0)\{\varsigma_{\theta_0}(\boldsymbol{Z}_j^{\tau}\boldsymbol{\theta}_0)-\boldsymbol{Z}_j\}$$
$$\times\Big\{1+\xi_{\theta_0}^{\tau}(\boldsymbol{Z}_j^{\tau}\boldsymbol{\theta}_0)\mathcal{W}_{\theta_0}^{-1}(\boldsymbol{Z}_j^{\tau}\boldsymbol{\theta}_0)\xi_{\theta_0}(\boldsymbol{Z}_j^{\tau}\boldsymbol{\theta}_0)\Big\}$$
$$\times\sum_{i=1}^{n}\Delta\tilde{Y}_{ij}^{(3)}\omega_{ij}^{\hat{\theta}_0}\times\big(1+O_P\big(\delta_{\hat{\theta}_0}+\delta_n^{'1/2}/h\big)\big).$$

By Lemma 1, similar to (38), we can have $\sum_{i=1}^{n}\Delta\tilde{Y}_{ij}^{(3)}\omega_{ij}^{\hat{\theta}_0}=O_P(\delta_n^{'1/2}+\delta_{\hat{\theta}_0})$. Together with $E[\{\varsigma_{\theta_0}(\boldsymbol{Z}_j^{\tau}\boldsymbol{\theta}_0)-\boldsymbol{Z}_j\}I^{\theta_0}(\boldsymbol{Z}_j,U_j)]=0$, we have $D_{n2}[1]=O_P(n^{-1/2})\times O_P(\delta_n^{'1/2}+\delta_{\hat{\theta}_0})=o_P(n^{-1/2})$.

Step 3.3.3 Deal with $D_{n2}[2]$. Similar to $D_{n2}[1]$ in step 3.3.2 and $B_{\beta_0}^{[2]}$ in step 2.1.1, we can have

$$D_{n2}[2]=n^{-1}\sum_{j=1}^{n}\sum_{i=1}^{n}\phi(U_j)g'(\boldsymbol{Z}_j^{\tau}\boldsymbol{\theta}_0)\boldsymbol{Z}_{ij}\tilde{\boldsymbol{X}}_i^{\tau}\{\tilde{\boldsymbol{\beta}}_0(U_i)-\tilde{\boldsymbol{\beta}}_0(U_j)\}\omega_{ij}^{\hat{\theta}_0}$$
$$\times\big(1+O_P\big(\delta_{\hat{\theta}_0}+\delta_n^{'1/2}/h\big)\big)$$
$$=O_P(h^3+n^{-1/2})\times\big(1+O_P\big(\delta_{\hat{\theta}_0}+\delta_n^{'1/2}/h\big)\big). \quad (43)$$

Thus, from (43), as $nh^6\to 0$, we have $D_{n2}[2]=o_P(n^{-1/2})$.

Step 3.4 Show that $D_{n2}[4]=o_P(n^{-1/2})$. Note that

$$\big\{\phi(U_i)g(\boldsymbol{Z}_i^{\tau}\boldsymbol{\theta}_0)-\phi(U_j)g(\boldsymbol{Z}_j^{\tau}\boldsymbol{\theta}_0)\big\}-\hat{d}_{2j,\hat{\theta}_0}U_{ij}-\hat{d}_{1j,\hat{\theta}_0}\boldsymbol{Z}_{ij}^{\tau}\boldsymbol{\theta}_0\overset{\text{def}}{=}J_{ij,1}-J_{ij,2}-J_{ij,3}.$$

Thus, $D_{n2}[4] = \frac{1}{n}\sum_{j=1}^{n}\sum_{i=1}^{n}\hat{d}_{1j,\hat{\theta}_0}\boldsymbol{Z}_{ij}\left(J_{ij,1} - J_{ij,2} - J_{ij,3}\right)\omega_{ij}^{\hat{\theta}_0}$. Similar to (43), as $nh^6 \to 0$, we have $n^{-1}\sum_{j=1}^{n}\sum_{i=1}^{n}\hat{d}_{1j,\hat{\theta}_0}\boldsymbol{Z}_{ij}J_{ij,1}\omega_{ij}^{\hat{\theta}_0} = o_P(n^{-1/2})$. We next consider $\frac{1}{n}\sum_{j=1}^{n}\sum_{i=1}^{n}\hat{d}_{1j,\hat{\theta}_0}\boldsymbol{Z}_{ij}J_{ij,2}\omega_{ij}^{\hat{\theta}_0}$. Similar to (30), we have $n^{-1}\sum_{j=1}^{n}\sum_{i=1}^{n}\hat{d}_{1j,\hat{\theta}_0}(\boldsymbol{Z}_i - \boldsymbol{Z}_j)J_{ij,2}\omega_{ij}^{\hat{\theta}_0} = o_P(n^{-1/2})$. Moreover, we can also obtain $n^{-1}\sum_{j=1}^{n}\sum_{i=1}^{n}\hat{d}_{1j,\hat{\theta}_0}\boldsymbol{Z}_{ij}$ $J_{ij,3}\omega_{ij}^{\hat{\theta}_0} = o_P(n^{-1/2})$. Thus, we obtain that $D_{n2}[4] = o_P(n^{-1/2})$.

A combination of the results, $D_{n2}[1] = o_P(n^{-1/2})$ in step 3.3.2, $D_{n2}[2] = o_P(n^{-1/2})$ in step 3.3.3, $D_{n2}[4] = o_P(n^{-1/2})$ in step 3.3.4, and (42) of $D_{n2}[3]$ in step 3.3.1, indicates that

$$\begin{aligned} D_{n2} = n^{-1}\sum_{j=1}^{n} I^{\theta_0}(\boldsymbol{Z}_j, U_j)\phi^2(U_j)g'(\boldsymbol{Z}_j^{\tau}\boldsymbol{\theta}_0)\big\{\mathcal{T}_{\theta_0}(\boldsymbol{Z}_j^{\tau}\boldsymbol{\theta}_0) - \varsigma_{\theta_0}(\boldsymbol{Z}_j^{\tau}\boldsymbol{\theta}_0)\xi_{\theta_0}^{\tau}(\boldsymbol{Z}_j^{\tau}\boldsymbol{\theta}_0)\big\} \\ \times\mathcal{W}_{\theta_0}^{-1}(\boldsymbol{Z}_j^{\tau}\boldsymbol{\theta}_0)\big\{\boldsymbol{X}_j - \xi_{\theta_0}(\boldsymbol{Z}_j^{\tau}\boldsymbol{\theta}_0)\big\}\varepsilon_j + o_P(n^{-1/2}). \end{aligned} \tag{44}$$

From the asymptotic results of step 3.1, D_{n1} in step 3.2 and D_{n2} in step 3.3, we have

$$\begin{aligned} \sqrt{n}\Big(\hat{\boldsymbol{\theta}}_0 - \boldsymbol{\theta}_0\Big) = \mathcal{D}_{\theta_0}^{-}\frac{1}{\sqrt{n}}\sum_{j=1}^{n} I^{\theta_0}(\boldsymbol{Z}_j, U_j)\phi^2(U_j)g'(\boldsymbol{Z}_j^{\tau}\boldsymbol{\theta}_0)\Big[\Big(\varsigma_{\theta_0}(\boldsymbol{Z}_j^{\tau}\boldsymbol{\theta}_0) - \boldsymbol{Z}_j\Big) \\ +\Big\{\mathcal{T}_{\theta_0}(\boldsymbol{Z}_j^{\tau}\boldsymbol{\theta}_0) - \varsigma_{\theta_0}(\boldsymbol{Z}_j^{\tau}\boldsymbol{\theta}_0)\xi_{\theta_0}^{\tau}(\boldsymbol{Z}_j^{\tau}\boldsymbol{\theta}_0)\Big\}\mathcal{W}_{\theta_0}^{-1}(\boldsymbol{Z}_j^{\tau}\boldsymbol{\theta}_0) \\ \times\Big\{\boldsymbol{X}_i - \xi_{\theta_0}(\boldsymbol{Z}_j^{\tau}\boldsymbol{\theta}_0)\Big\}\Big]\varepsilon_j + o_P(n^{-1/2}) \xrightarrow{L} N\Big(0_{q\times 1}, \Sigma_{\theta_0}\Big). \end{aligned}$$

Theorem holds as claimed. □

A.6 Proof of Theorem 4

Step 4.1 To prove the theorem, we first show $\frac{1}{n}\sum_{j=1}^{n}\hat{b}_{j,\hat{\theta}_0}^{\otimes 2} - E\boldsymbol{\beta}_0(U)^{\otimes 2} = O_P(n^{-1/2})$.

$$\begin{aligned} n^{-1}\sum_{j=1}^{n}\hat{b}_{j,\hat{\theta}_0}^{\otimes 2} - E\boldsymbol{\beta}_0(U)^{\otimes 2} &= n^{-1}\sum_{j=1}^{n}\big\{\hat{b}_{j,\hat{\theta}_0} - \boldsymbol{\beta}_0(U_j)\big\}\boldsymbol{\beta}_0^{\tau}(U_j) \\ &+n^{-1}\sum_{j=1}^{n}\boldsymbol{\beta}_0(U_j)\big\{\hat{b}_{j,\hat{\theta}_0} - \boldsymbol{\beta}_0(U_j)\big\}^{\tau} \\ &+\left\{n^{-1}\sum_{j=1}^{n}\boldsymbol{\beta}_0(U_j)^{\otimes 2} - E\boldsymbol{\beta}_0(U)^{\otimes 2}\right\} \\ &+n^{-1}\sum_{j=1}^{n}\big\{\hat{b}_{j,\hat{\theta}_0} - \boldsymbol{\beta}_0(U_j)\big\}^{\otimes 2} \\ &\stackrel{\text{def}}{=} M_{n1} + M_{n2} + M_{n3} + M_{n4}. \end{aligned}$$

Similar to $A^{(2)}_{\beta_0}$ in step 2.1, we can show that $M_{n1} = O_P(n^{-1/2})$, $M_{n2} = O_P(n^{-1/2})$. Furthermore, the elements of M_{n3} are *i.i.d* random variables. The weak law of large numbers means $M_{n3} = O_P(n^{-1/2})$. In a similar way to D_{n2} in step 3.3, we can show $M_{n4} = o_P(n^{-1/2})$. Thus, $\frac{1}{n}\sum_{j=1}^{n}\hat{b}^{\otimes 2}_{j,\hat{\theta}_0} - E\boldsymbol{\beta}_0(U)^{\otimes 2} = O_P(n^{-1/2})$.

Step 4.2 Without loss of generality, we assume that the first p_0 predictors $\{X_1, X_2, \ldots, X_{p_0}\}$ are relevant. In other words, $\{\boldsymbol{\beta}_{0,1}, \boldsymbol{\beta}_{0,2}, \ldots, \boldsymbol{\beta}_{0,p_0}\}$ are nonzero. We further assume that all the elements of the first principal of $E\boldsymbol{\beta}_0(U)^{\otimes 2}$ are nonzero. In such a setting, we first show that there exists an $r > p_0$, such that $\breve{\eta}_{n,r} = 0$. Suppose that $\breve{\eta}_{n,r}$ are nonzero for $r > p_0$, then, by the proof of Theorem 2 in Chen et al. (2010), $\breve{\eta}_n$ satisfies the following equation:

$$2N_n\left\{n^{-1}\sum_{j=1}^{n}\hat{b}^{\otimes 2}_{j,\hat{\theta}_0}\right\}\breve{\eta}_n = N_n\varrho_n, \tag{45}$$

where $N_n = \mathbf{I}_p - \breve{\eta}_n\breve{\eta}_n^{\tau}$ and $\varrho_n = \left(\alpha_1\text{sign}(\breve{\eta}_{n,1}), \alpha_2\text{sign}(\breve{\eta}_{n,2}), \ldots, \alpha_p\text{sign}(\breve{\eta}_{n,p})\right)^{\tau}$. Note that $\breve{\eta}_n$ is the eigenvector of $\breve{\eta}_n\breve{\eta}_n^{\tau}$ corresponding to its eigenvalue 1. Let $(l_1, l_2, \ldots, l_{p-1})$ be the eigenvectors of $\breve{\eta}_n\breve{\eta}_n^{\tau}$ corresponding to its eigenvalues 0. Thus, $(\breve{\eta}_n, l_1, \ldots, l_{p-1})$ is an independent basis of the space R^p. Using this independent basis, there exists two sequences of constants $\{a_r\}_{r=0}^{p-1}$ and $\{a'_r\}_{r=0}^{p-1}$ such that $\{\frac{1}{n}\sum_{j=1}^{n}\hat{b}^{\otimes 2}_{j,\hat{\theta}_0}\}\breve{\eta}_n = a_0\breve{\eta}_n + \sum_{r=1}^{p-1}a_r l_r$ and $\varrho_n = a'_0\breve{\eta}_n + \sum_{r=1}^{p-1}a'_r l_r$. Combining them with (45), we have $2\left\{n^{-1}\sum_{j=1}^{n}\hat{b}^{\otimes 2}_{j,\hat{\theta}_0}\right\}\breve{\eta}_n - 2a_0\breve{\eta}_n = \varrho_n - a'_0\breve{\eta}_n$, from which and the expression of ϱ_n,

$$\breve{\eta}_n^{\tau}\varrho_n = \sum_{r=1}^{p_0}\alpha_r|\breve{\eta}_{n,r}| + \sum_{r=p_0+1}^{p}\alpha_r|\breve{\eta}_{n,r}| = 2\breve{\eta}_n^{\tau}\left\{n^{-1}\sum_{j=1}^{n}\hat{b}^{\otimes 2}_{j,\hat{\theta}_0}\right\}\breve{\eta}_n - (2a_0 - a'_0). \tag{46}$$

Let $\lambda_{\max}\left(\frac{1}{n}\sum_{j=1}^{n}\hat{b}^{\otimes 2}_{j,\hat{\theta}_0}\right)$, $\lambda_{\max}\left(E\boldsymbol{\beta}_0(U)^{\otimes 2}\right)$ be the largest eigenvalues of $\frac{1}{n}\sum_{j=1}^{n}\hat{b}^{\otimes 2}_{j,\hat{\theta}_0}$ and $E\boldsymbol{\beta}_0(U)^{\otimes 2}$, respectively. From the result of step 4.1, $\frac{1}{n}\sum_{j=1}^{n}\hat{b}^{\otimes 2}_{j,\hat{\theta}_0} - E\boldsymbol{\beta}_0(U)^{\otimes 2} = O_P(n^{-1/2})$, which means

$$\left|\lambda_{\max}\left(n^{-1}\sum_{j=1}^{n}\hat{b}^{\otimes 2}_{j,\hat{\theta}_0}\right) - \lambda_{\max}\left(E\boldsymbol{\beta}_0(U)^{\otimes 2}\right)\right| = O_P(n^{-1/2}). \tag{47}$$

Thus, (47) entails that $\breve{\eta}_n^{\tau}\left\{\frac{1}{n}\sum_{j=1}^{n}\hat{b}^{\otimes 2}_{j,\hat{\theta}_0}\right\}\breve{\eta}_n \le \lambda_{\max}\left(\frac{1}{n}\sum_{j=1}^{n}\hat{b}^{\otimes 2}_{j,\hat{\theta}_0}\right) = O_P(1)$. Together with the assumption that $\alpha_0 \to 0$ and that $\alpha_r = \alpha_0|\hat{\eta}_{n,r}|^{-\varpi} = O_P(\alpha_0) = o_P(1)$ for $r \le p_0$. We have that $\max_{r\le p_0}\{\alpha_r\} \to 0$. From (46), we further obtain that

$$\sum_{r=p_0+1}^{p} \alpha_r |\breve{\eta}_{n,r}| = 2\breve{\eta}_n^{\tau} \left\{ n^{-1} \sum_{j=1}^{n} \hat{b}_{j,\hat{\theta}_0}^{\otimes 2} \right\} \breve{\eta}_n - (2a_0 - a_0') + o_P(1). \tag{48}$$

Thus, (48) together with the assumption that $\alpha_0 n^{\varpi/2} \to \infty$, then we obtain that $\alpha_r = \alpha_0 |\hat{\eta}_{n,r}|^{-\varpi} \xrightarrow{P} \infty$, $\min_{r>p_0}\{\alpha_r\} \xrightarrow{P} \infty$. Then, together with (47), we obtain that

$$\begin{aligned}
\min_{r>p_0} \{|\breve{\eta}_{n,r}|\} &\leq \frac{\sum_{r=p_0+1}^{p} \alpha_r |\breve{\eta}_{n,r}|}{\min_{r>p_0}\{\alpha_r\}} = O_P\Big(2\breve{\eta}_n^{\tau} \Big\{ n^{-1} \sum_{j=1}^{n} \hat{b}_{j,\hat{\theta}_0}^{\otimes 2} \Big\} \breve{\eta}_n \\
&\quad - (2a_0 - a_0') + o_P(1) \Big) / \alpha_0 n^{\varpi/2} \\
&= O_P\Big(\lambda_{\max}\big(E\boldsymbol{\beta}_0(U)^{\otimes 2} \big) - (2a_0 - a_0') + o_P(1) \Big) / \alpha_0 n^{\varpi/2} = o_P(1).
\end{aligned}$$

Because we have assumed above that all the $\breve{\eta}_{n,r}$ are nonzero for $r > p_0$, it leads to $1 = \text{sign}\Big(\min_{r>p_0}\{|\breve{\eta}_{n,r}|\} \Big) = \text{sign}\big(o_P(1) \big) \xrightarrow{P} 0$, a contradiction. This means that there exists at least $r_0 > p_0$ such that $\breve{\eta}_{n,r_0} = 0$ with probability 1. We can further confirm that $\breve{\eta}_{n,r} = 0$ for all $r > p_0$ with probability 1. This confirmation can be proved in a way similar to the proof of Theorem 2 in Chen et al. (2010). □

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

References

Akaike, H. (1973). Maximum likelihood identification of Gaussian autoregressive moving average models. *Biometrika, 60*, 255–265.

Breiman, L. (1996). Heuristics of instability and stabilization in model selection. *The Annals of Statistics, 24*, 2350–2383.

Carroll, R., Fan, J., Gijbels, I., Wand, M. P. (1997). Generalized partially linear single-index models. *Journal of the American Statistical Association, 92*, 477–489.

Carroll, R. J., Ruppert, D., Stefanski, L. A., Crainiceanu, C. M. (2006). *Nonlinear measurement error models: A modern perspective* (2nd ed.). New York: Chapman and Hall.

Chen, H. (1988). Convergence rates for parametric components in a partly linear model. *The Annals of Statistics, 16*, 136–146.

Chen, X., Zou, C., Cook, R. D. (2010). Coordinate-independent sparse sufficient dimension reduction and variable selection. *The Annals of Statistics, 38*, 3696–3723.

Cui, X., Guo, W., Lin, L., Zhu, L. (2009). Covariate-adjusted nonlinear regression. *The Annals of Statistics, 37*, 1839–1870.

Efron, B., Hastie, T., Johnstone, I., Tibshirani, R. (2004). Least angle regression. *The Annals of Statistics, 32*, 407–499.

Fan, J., Gijbels, I. (1996). *Local polynomial modelling and its applications*. London: Chapman & Hall.

Fan, J., Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association, 96*, 1348–1360.

Fan, J., Zhang, W. (1999). Statistical estimation in varying coefficient models. *The Annals of Statistics, 27*, 1491–1518.

Härdle, W., Hall, P., Ichimura, H. (1993). Optimal smoothing in single-index models. *The Annals of Statistics, 21*, 157–178.

Härdle, W., Liang, H., Gao, J. T. (2000). *Partially linear models*. Heidelberg: Springer Physica.

Härdle, W., Müller, M., Sperlich, S., Werwatz, A. (2004). *Nonparametric and semiparametric models*. New York: Springer-Verlag.

Hastie, T., Tibshirani, R. (1993). Varying-coefficient models. *Journal of the Royal Statistical Society, Series B, 55*, 757–796.

Heckman, N. E. (1986). Spline smoothing in partly linear models. *Journal of the Royal Statistical Society, Series B, 48*, 244–248.

Horowitz, J. L. (2009). *Semiparametric and nonparametric methods in econometrics*. New York: Springer.

Horowitz, J. L., Härdle, W. (1996). Direct semiparametric estimation of single-index models with discrete covariates. *Journal of the American Statistical Association, 91*, 1632–1640.

Ichimura, H. (1993). Semiparametric least squares (SLS) and weighted SLS estimation of single-index models. *Journal of Econometrics, 58*, 71–120.

Kato, T. (1983). *Perturbation theory for linear operation* (2nd ed.). New York: Springer.

Kaysen, G. A., Dubin, J. A., Müller, H.-G., Mitch, W. E., Rosales, L. M., Nathan W. Levin the Hemo Study Group (2002). Relationships among inflammation nutrition and physiologic mechanisms establishing albumin levels in hemodialysis patients. *Kidney International, 61*, 2240–2249.

Li, L. (2007). Sparse sufficient dimension reduction. *Biometrika, 94*, 603–613.

Li, R., Liang, H. (2008). Variable selection in semiparametric regression modeling. *The Annals of Statistics, 36*, 261–286.

Liang, H., Li, R. (2009). Variable selection for partially linear models with measurement errors. *Journal of the American Statistical Association, 104*, 234–248.

Liang, H., Liu, X., Li, R., Tsai, C. L. (2010). Estimation and testing for partially linear single-index models. *The Annals of Statistics, 38*, 3811–3836.

Mack, Y. P., Silverman, B. W. (1982). Weak and strong uniform consistency of kernel regression estimates. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete, 61*, 405–415.

Mallows, C. L. (1973). Some comments on C_P. *Technometrics, 15*, 661–675.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics, 6*, 461–464.

Şentürk, D., Müller, H.-G. (2005). Covariate-adjusted regression. *Biometrika, 92*, 75–89.

Şentürk, D., Müller, H.-G. (2006). Inference for covariate adjusted regression via varying coefficient models. *The Annals of Statistics, 34*, 654–679.

Şentürk, D., Müller, H.-G. (2009). Covariate-adjusted generalized linear models. *Biometrika, 96*, 357–370.

Serfling, R. J. (1980). *Approximation theorems of mathematical statistics*. New York: Wiley.

Speckman, P. E. (1988). Kernel smoothing in partial linear models. *Journal of the Royal Statistical Society, Series B, 50*, 413–436.

Stute, W., Zhu, L.-X. (2005). Nonparametric checks for single-index models. *The Annals of Statistics, 33*, 1048–1083.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B, 58*, 267–288.

Wahba, G. (1984). Partial spline models for the semiparametric estimation of functions of several variables. *Statistical Analyses for Time Series*, Institute of Statistical Mathematics, Tokyo (pp. 319–329). Japan-US Joint Seminar.

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

Wang, L., Liu, X., Liang, H., Carroll, R. J. (2011). Estimation and variable selection for generalized additive partial linear models. *The Annals of Statistics, 39*, 1827–1851.

Xia, Y., Härdle, W. (2006). Semi-parametric estimation of partially linear single-index models. *Journal of Multivariate Analysis, 97*, 1162–1184.

Xia, Y., Tong, H., Li, W. K., Zhu, L.-X. (2002). An adaptive estimation of dimension reduction space. *Journal of the Royal Statistical Society, Series B, 64*, 363–410.

Yu, Y., Ruppert, D. (2002). Penalized spline estimation for partially linear single-index models. *Journal of the American Statistical Association, 97*, 1042–1054.

Zhang, J., Zhu, L.-P., Zhu, L.-X. (2012). On a dimension reduction regression with covariate adjustment. *Journal of Multivariate Analysis, 104*, 39–55.