Statistics and Probability Letters 94 (2014) 119–127

Contents lists available at ScienceDirect

Statistics and Probability Letters

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

A robust and efficient estimation method for single-index varying-coefficient models

Hu Yang, Chaohui Guo *, Jing Lv

College of Mathematics and Statistics, Chongqing University, Chongqing, 401331, China

ARTICLE INFO

Article history:
Received 29 December 2013
Received in revised form 8 June 2014
Accepted 10 July 2014
Available online 18 July 2014

Keywords:
Asymptotic normality
Local linear smoothing
Modal regression
Robust estimation
Single-index varying-coefficient models

ABSTRACT

A new estimation procedure based on modal regression is proposed for single-index varying-coefficient models. The proposed method achieves better robustness and efficiency than that of Xue and Pang (2013). We establish the asymptotic normalities of proposed estimators and evaluate the performance of the proposed method by a numerical simulation.

1. Introduction

We consider the following single-index varying-coefficient model (SIVCM)

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

where $Y \in \mathbb{R}$, $(\mathbf{X}, \mathbf{Z}) \in \mathbb{R}^p \times \mathbb{R}^q$ with $Z_{i1} = 1$, $\boldsymbol{\beta}_0$ are unknown parameters, $\mathbf{g}_0(.)$ is a $q \times 1$ vector of unknown univariate smoothing functions and ε is the random error with $E(\varepsilon|\mathbf{X}, \mathbf{Z}) = 0$. For the reason of identifiability, we generally assume $\|\boldsymbol{\beta}_0\| = 1$ and $\beta_{01} > 0$, where $\|.\|$ denotes the Euclidean metric.

The SIVCM includes many existing important statistical models. As special examples of the SIVCM, the single-index model with $Z = 1$ and $q = 1$ and the varying-coefficient model with $p = 1$ and $\beta = 1$, which were investigated by many scholars (see Härdle et al., 1993; Chang et al., 2010; Hastie and Tibshirani, 1993; Fan et al., 2003 and Xue and Zhu, 2006). Due to its flexibility and interpretability, in recent years, many authors paid attention to the estimation of the SIVCM which has experienced rapid developments in both theory and methodology. Xue and Wang (2012) presented an empirical likelihood ratio estimator. Xue and Pang (2013) proposed new estimating equations to obtain a more efficient estimator of parametric component in the SIVCM, where the nonparametric component is estimated by using local linear smoothing. Huang and Zhang (2013) proposed a profile empirical likelihood inference based on new and simple estimating equations for the SIVCM.

However, these mentioned estimation procedures above are expected to be very sensitive to outliers or/and heavy-tailed random errors, which result in reducing their efficiency for many commonly used non-normal errors. Recently, researchers devote themselves to exploring an alternative robust estimation method to overcome this problem. For example, Yao et al. (2012), Zhang et al. (2013) and Liu et al. (2013) investigated a new estimation method based on modal regression for nonparametric models, semiparametric partially linear varying coefficient models and single-index models respectively.

* Corresponding author.
E-mail address: guochaohui2010@126.com (C. Guo).

http://dx.doi.org/10.1016/j.spl.2014.07.011

The estimation method mentioned above not only has very good robustness in the presence of outliers or heavy-tail error distributions, but also can achieve full asymptotic efficiency under the normal error distribution. This fact motivates us to extend the modal regression method to the SIVCM. In this paper, we first estimate the nonparametric component using local modal smoothing, then we construct an estimator of parametric component by using robust estimating equations.

The rest of the paper is organized as follows. In Section 2, we propose the modal regression method for the SIVCM and present an EM-type algorithm for the proposed estimators. In Section 3, the asymptotic normalities of the proposed estimators are established. We discuss the selection of bandwidths in Section 4. In Section 5, a Monte Carlo simulation study is used to illustrate the proposed methodology. All the regularity conditions and the technical proofs are provided in the Appendix.

2. Methodology

2.1. Modal regression

Suppose that $\{(\mathbf{X}_i, \mathbf{Z}_i, Y_i),\ 1 \le i \le n\}$ is a sample from (1). Our main interest is to estimate $\boldsymbol{\beta}_0$ and $\mathbf{g}_0(.)$ in (1). First, for given $\boldsymbol{\beta}$ with $\|\boldsymbol{\beta}\| = 1$ and $\beta_1 > 0$, we use the local linear polynomial to approximate the unknown nonparametric functions $g_{0j}(.),\ j = 1, \ldots, q$ for v in a small neighborhood of u

$$g_{0j}(v) \approx g_{0j}(u) + {g'}_{0j}(u)(v-u) \triangleq a_j + b_j(v-u). \tag{2}$$

Let $\mathbf{a} = (a_1, \ldots, a_q)^T$, $\mathbf{b} = (b_1, \ldots, b_q)^T$ and $\left\{\hat{\boldsymbol{a}}, \hat{\boldsymbol{b}}\right\}$ be the maximizer of local modal function

$$\frac{1}{n}\sum_{i=1}^{n} \phi_{h_2}\left[Y_i - \sum_{j=1}^{q}\left\{a_j + b_j(\boldsymbol{\beta}^T\mathbf{X}_i - u)\right\} Z_{ij}\right] K_{h_1}(\boldsymbol{\beta}^T\mathbf{X}_i - u), \tag{3}$$

where $K_{h_1}(.) = K(./h_1)/h_1$, $\phi_{h_2}(t) = h_2^{-1}\phi(t/h_2)$, together with that $K(.), \phi(\cdot)$ are kernel functions and h_1, h_2 are bandwidths. In order to simplify the calculation, we use the standard normal density for $\phi(t)$ throughout this paper. Moreover, we denote the estimations of $\mathbf{g}_0(u)$ and $\mathbf{g}_0'(u)$ as $\hat{\mathbf{g}}(u; \boldsymbol{\beta})$ and $\hat{\mathbf{g}}'(u; \boldsymbol{\beta})$ at fixed point $\boldsymbol{\beta}$, where $\mathbf{g}_0'(.)$ stands for the first derivative of the function vector $\mathbf{g}_0(.)$, $\hat{\mathbf{g}}(u; \boldsymbol{\beta}) = (\hat{g}_1(u; \boldsymbol{\beta}), \ldots, \hat{g}_q(u; \boldsymbol{\beta}))^T$ with $\hat{g}_j(u; \boldsymbol{\beta}) = \hat{a}_j$ and $\hat{\mathbf{g}}'(u; \boldsymbol{\beta}) = (\hat{g}_1'(u; \boldsymbol{\beta}), \ldots, \hat{g}_q'(u; \boldsymbol{\beta}))^T$ with $\hat{g}_j'(u; \boldsymbol{\beta}) = \hat{b}_j$ for $j = 1, \ldots, q$.

Then combine with the estimators $\hat{\mathbf{g}}(u; \boldsymbol{\beta})$ and $\hat{\mathbf{g}}'(u; \boldsymbol{\beta})$, we can achieve an efficient estimate of the global parameter $\boldsymbol{\beta}_0$ by maximizing the following modal function

$$L_n(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^{n} \phi_{h_3}\left[Y_i - \hat{\mathbf{g}}^T(\boldsymbol{\beta}^T\boldsymbol{X}_i; \boldsymbol{\beta})\mathbf{Z}_i\right], \tag{4}$$

subject to the constraint $\|\boldsymbol{\beta}\| = 1$ and $\beta_1 > 0$.

The parameter space $\boldsymbol{\Omega} = \{\boldsymbol{\beta} \in R^p : \|\boldsymbol{\beta}\| = 1, \beta_1 > 0\}$ means that $\boldsymbol{\beta}$ is on the boundary of a unit ball, so the function $\mathbf{g}(\boldsymbol{\beta}^T\boldsymbol{X}_i)$ is not differentiable at point $\boldsymbol{\beta}$. To handle this constraint, we employ the "remove-one-component" method to transfer the restricted modal regression (4) to the unrestricted modal regression. Specifically, for $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^T$, let $\boldsymbol{\beta}^{(1)} = \left(\beta_2, \ldots, \beta_p\right)^T$ be a $p-1$ dimensional vector by removing the 1st component β_1 in $\boldsymbol{\beta}$. Then $\boldsymbol{\beta}$ can be rewritten as $\boldsymbol{\beta} = \beta(\boldsymbol{\beta}^{(1)}) = (\sqrt{1 - \|\boldsymbol{\beta}^{(1)}\|^2}, \boldsymbol{\beta}^{(1)T})^T$, $\|\boldsymbol{\beta}^{(1)}\| < 1$. Thus, $\boldsymbol{\beta}$ is infinitely differentiable with respect to $\boldsymbol{\beta}^{(1)}$ and the Jacobian matrix is

$$J = \frac{\partial \boldsymbol{\beta}}{\partial \boldsymbol{\beta}^{(1)}} = \begin{pmatrix} -\boldsymbol{\beta}^{(1)T} \Big/ \sqrt{1 - \|\boldsymbol{\beta}^{(1)}\|^2} \\ I_{p-1} \end{pmatrix}, \tag{5}$$

where I_p is the $p \times p$ identity matrix.

After this re-parametrization, we can easily find that $L_n(\boldsymbol{\beta}) = L_n(\beta(\boldsymbol{\beta}^{(1)}))$. Then, we can obtain an estimator $\hat{\boldsymbol{\beta}}^{(1)}$ of $\boldsymbol{\beta}_0^{(1)}$ by maximizing $L_n(\beta(\boldsymbol{\beta}^{(1)}))$, which is equivalent to solve the following robust estimating equations (REE)

$$\frac{1}{n}\sum_{i=1}^{n} \phi_{h_3}'\left[Y_i - \hat{\mathbf{g}}^T(\boldsymbol{\beta}^T\mathbf{X}_i; \boldsymbol{\beta}, h_1)\mathbf{Z}_i\right] \hat{\mathbf{g}}'^{\,T}(\boldsymbol{\beta}^T\mathbf{X}_i; \boldsymbol{\beta}, h_4)\mathbf{Z}_i\mathbf{J}^T\mathbf{X}_i = \mathbf{0}. \tag{6}$$

We use two different bandwidths h_1 and h_4 to accommodate different convergence rates for the estimates of $\mathbf{g}_0$ and its derivative $\mathbf{g}_0'$ such that the estimate $\hat{\boldsymbol{\beta}}^{(1)}$ has $\sqrt{n}$ convergence rate. Then, the estimator $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}_0$ can be obtained via a transform of the estimator $\hat{\boldsymbol{\beta}}^{(1)}$.

After obtaining the $\sqrt{n}$ consistent estimate $\hat{\boldsymbol{\beta}}$, we can define the final estimator of $\mathbf{g}_0(u)$ as $\hat{g}_j(u) = \hat{g}_j(u, \hat{\boldsymbol{\beta}})$, where $\hat{g}_j(u, \hat{\boldsymbol{\beta}}) = \hat{a}_j, j = 1, \ldots, q$ is the solution of maximizing (3). In practice, we use a data-driven method to select all the above four bandwidths $h_1, \ldots, h_4$ and detailed discussion about bandwidths are specified in Sections 4 and 5.

2.2. Computation

We first discuss the modified EM algorithm which is similar to that of Liu et al. (2013). For given u and $\hat{\boldsymbol{\beta}}$, maximizing (3) with respect to $\boldsymbol{\theta} = (\boldsymbol{a}^T, \boldsymbol{b}^T)^T$ is equivalent to

$$\log \sum_{i=1}^{n} f(\boldsymbol{X}_i, Y_i, \boldsymbol{\theta}) \triangleq \log \sum_{i=1}^{n} \phi_{h_2}\left[Y_i - \sum_{j=1}^{q}\left\{a_j + b_j(\boldsymbol{\beta}^T\boldsymbol{X}_i - u)\right\} Z_{ij}\right] K_{h_1}(\boldsymbol{\beta}^T\boldsymbol{X}_i - u),$$

Let $\{\pi_i\}_{i=1}^n$ be a discrete distribution, satisfying $\sum_{i=1}^n \pi_i = 1$ and $\pi_i \geq 0$. By Jensen's inequality, we have

$$\log \sum_{i=1}^{n} f(\boldsymbol{X}_i, Y_i, \boldsymbol{\theta}) = \log \sum_{i=1}^{n} \frac{f(\boldsymbol{X}_i, Y_i, \boldsymbol{\theta})}{\pi_i}\pi_i \geq \sum_{i=1}^{n} \pi_i \log \frac{f(\boldsymbol{X}_i, Y_i, \boldsymbol{\theta})}{\pi_i}. \tag{7}$$

When $\frac{f(\boldsymbol{X}_i, Y_i, \boldsymbol{\theta})}{\pi_i}$ is constant for $i = 1, \ldots, n$, the equality holds. Because of $\sum_{i=1}^n \pi_i = 1$, so we have

$$\pi_i = \frac{K_{h_1}(\hat{\boldsymbol{\beta}}^T\mathbf{X}_i - u)\phi_{h_2}\left\{Y_i - \sum_{j=1}^{q}\left\{a_j + b_j(\hat{\boldsymbol{\beta}}^T\mathbf{X}_i - u)\right\} Z_{ij}\right\}}{\sum_{i=1}^{n} K_{h_1}(\hat{\boldsymbol{\beta}}^T\mathbf{X}_i - u)\phi_{h_2}\left\{Y_i - \sum_{j=1}^{q}\left\{a_j + b_j(\hat{\boldsymbol{\beta}}^T\mathbf{X}_i - u)\right\} Z_{ij}\right\}}.$$

This is the E-step of the algorithm, (7) gives a lower-bound of objective function, then, we only need to maximize this lower-bound in the M-step with respect to $\boldsymbol{\theta} = (\boldsymbol{a}^T, \boldsymbol{b}^T)^T$. With the same argument as the traditional EM algorithm, the ascending property of the proposed EM algorithm can be obtained which ensures the convergence of the algorithm.
Now, we outline the previous estimation methodology as follows:

Step 0. Give an initial value $\hat{\boldsymbol{\beta}}$ with $\left\|\hat{\boldsymbol{\beta}}\right\| = 1$ and $\hat{\beta}_1 > 0$.

Step 1. Obtain $\hat{\boldsymbol{a}}(u; \hat{\boldsymbol{\beta}}), \hat{\boldsymbol{b}}(u; \hat{\boldsymbol{\beta}})$ by updating $\boldsymbol{a}, \boldsymbol{b}$ based on (3) with $\hat{\boldsymbol{\beta}}$ replacing $\boldsymbol{\beta}$. Let $\boldsymbol{\theta}^{(0)} = (\mathbf{a}^{(0)T}, \mathbf{b}^{(0)T})^T$ be the initial estimation and set $m = 0$.

(E-step): We first update $\pi(i|\theta^{(m)})$ by

$$\pi(i|\theta^{(m)}) = \frac{K_{h_1}(\hat{\boldsymbol{\beta}}^T\mathbf{X}_i - u)\phi_{h_2}\left\{Y_i - \sum_{j=1}^{q}\left\{a_j^{(m)} + b_j^{(m)}(\hat{\boldsymbol{\beta}}^T\mathbf{X}_i - u)\right\} Z_{ij}\right\}}{\sum_{i=1}^{n} K_{h_1}(\hat{\boldsymbol{\beta}}^T\mathbf{X}_i - u)\phi_{h_2}\left\{Y_i - \sum_{j=1}^{q}\left\{a_j^{(m)} + b_j^{(m)}(\hat{\boldsymbol{\beta}}^T\mathbf{X}_i - u)\right\} Z_{ij}\right\}}$$

for $i = 1, \ldots, n$.
(M-step): Then, we update $\boldsymbol{\theta}^{(m+1)}$ by

$$\begin{aligned}\boldsymbol{\theta}^{(m+1)} &= \arg\max_{\boldsymbol{\theta}} \frac{1}{n}\sum_{i=1}^{n}\left\{\pi(i|\boldsymbol{\theta}^{(m)}) \log \phi_{h_2}\left[Y_i - \{\mathbf{a} + \mathbf{b}(\hat{\boldsymbol{\beta}}^T\mathbf{X}_i - u)\}^T\mathbf{Z}_i\right]\right\} \\ &= \left(\mathbf{Z}^{*T}\mathbf{W}\mathbf{Z}^*\right)^{-1}\mathbf{Z}^{*T}\mathbf{W}\mathbf{Y},\end{aligned} \tag{8}$$

where $\mathbf{Z}^* = (\mathbf{Z}_1^*, \ldots, \mathbf{Z}_n^*)^T$ with $\mathbf{Z}_i^* = \left(\mathbf{Z}_i^T, (\hat{\boldsymbol{\beta}}^T\mathbf{X}_i - u)\mathbf{Z}_i^T\right)^T$, $\mathbf{Y} = (Y_1, \ldots, Y_n)^T$ and W is an $n \times n$ diagonal matrix whose diagonal elements are $\pi(i|\boldsymbol{\theta}^{(m)})$s. Iterate the E-step and M-step until the algorithm converges.
Step 2. For given $\hat{\boldsymbol{g}}, \hat{\boldsymbol{g}}'$, we use Fisher scoring method to solve the estimating equation (6). Let $\boldsymbol{\beta}^{(0)} = \hat{\boldsymbol{\beta}}$ be the initial estimator and set $m = 0$. We update $\pi^*(i|\boldsymbol{\beta}^{(m)})$ by

$$\pi^*(i|\boldsymbol{\beta}^{(m)}) = \frac{\phi_{h_3}(Y_i - \hat{\boldsymbol{g}}^T(\boldsymbol{\beta}^{(m)T}\boldsymbol{X}_i; \boldsymbol{\beta}^{(m)}, h_1)Z_i)}{\sum_{i=1}^{n}\phi_{h_3}(Y_i - \hat{\boldsymbol{g}}^T(\boldsymbol{\beta}^{(m)T}\boldsymbol{X}; \boldsymbol{\beta}^{(m)}, h_1)Z_i)}, \quad \text{for } i = 1, \ldots, n. \tag{9}$$

With the same argument as Step 1, maximizing (4) is equivalent to maximize

$$L_n(\beta(\boldsymbol{\beta}^{(1)})) = \frac{1}{n}\sum_{i=1}^{n}\pi^*(i|\boldsymbol{\beta}^{(m)}) \log \phi_{h_3}\left[Y_i - \hat{\boldsymbol{g}}^T(\boldsymbol{\beta}^T\boldsymbol{X}_i; \boldsymbol{\beta})\boldsymbol{Z}_i\right].$$

And equivalently, the resulting estimator, say $\hat{\boldsymbol\beta}^{(1)}$ is the solution of the following estimating equations by removing some irrelevant terms

$$\sum_{i=1}^{n}\pi^*(i|\boldsymbol\beta^{(m)})\left[Y_i-\hat{\mathbf{g}}^T(\boldsymbol\beta^T\mathbf{X}_i;\boldsymbol\beta,h_1)\mathbf{Z}_i\right]\hat{\mathbf{g}}'^{\,T}(\boldsymbol\beta^T\mathbf{X}_i;\boldsymbol\beta,h_4)\mathbf{Z}_i\mathbf{J}^T\mathbf{X}_i=\mathbf{0}.$$

With the help of the Fisher's scoring method, we can find the estimate of $\boldsymbol\beta^{(1)}$ using the iterative procedure

$$(\boldsymbol\beta^{(1)})^{(m+1)}=(\boldsymbol\beta^{(1)})^{(m)}+\left(\mathbf{X}^{*T}\mathbf{W}^*\mathbf{X}^*\right)^{-1}\mathbf{X}^{*T}\mathbf{W}^*\mathbf{Y}^*, \tag{10}$$

where $\mathbf{X}^*=(\mathbf{X}_1^*,\dots,\mathbf{X}_n^*)^T$ with $\mathbf{X}_i^*=\hat{\mathbf{g}}'^{\,T}(\boldsymbol\beta^{(m)}\mathbf{X}_i;\boldsymbol\beta^{(m)},h_4)\mathbf{Z}_i\boldsymbol{J}^T\boldsymbol{X}_i$, $\mathbf{Y}^*=(Y_1^*,\dots,Y_n^*)^T$ with $Y_i^*=Y_i-\hat{\mathbf{g}}^T(\boldsymbol\beta^{(m)}\mathbf{X}_i;\boldsymbol\beta^{(m)},h_1)\boldsymbol{Z}_i$ and $\boldsymbol{W}^*$ is an $n\times n$ diagonal matrix whose diagonal elements are $\pi^*(i|\boldsymbol\beta^{(m)})$s.

Step 3. Repeat Steps 1 and 2 until convergence. We obtain the estimate $\hat{\boldsymbol\beta}$.

Step 4. With the $\sqrt{n}$ consistent estimate $\hat{\boldsymbol\beta}$ obtained from Step 3, the final estimate of $\mathbf{g}_0(.)$ is $\hat{\mathbf{g}}(u;\hat{\boldsymbol\beta})$ which can be obtained by carrying out Step 1.

Remark 1. If $\max|\hat{\boldsymbol\theta}^{(m+1)}-\hat{\boldsymbol\theta}^{(m)}|$ and $\max|(\hat{\boldsymbol\beta}^{(1)})^{(m+1)}-(\hat{\boldsymbol\beta}^{(1)})^{(m)}|$ are smaller than a cutoff value $\epsilon>0$ (such as 10^{-6}), then we stop the iteration. Our numerical experience in Section 5 indicates that the convergence criterion is met usually in a few iterations.

3. Theoretical properties

In this section, we study the large sampling properties of estimators. Define $f_U(.)$ as the marginal density function of $U=\boldsymbol\beta_0^T\mathbf{X}$. We assume that the kernel density function $K(.)$ is symmetric. Let $\mu_j=\int u^jK(u)du$ and $\nu_j=\int u^jK^2(u)du$, $j=0,1,2,\dots$, and $F(u,\mathbf{z},h)=E\left\{\phi''_h(\varepsilon)\mid U=u,\mathbf{Z}=\mathbf{z}\right\}$, $G(u,\mathbf{z},h)=E\left\{\phi'_h(\varepsilon)^2\mid U=u,\mathbf{Z}=\mathbf{z}\right\}$.

Theorem 1. *Assume the regularity conditions given in the Appendix hold, and that h_2 and h_3 are constant and independent with n, we have*

$$\sqrt{n}(\hat{\boldsymbol\beta}-\boldsymbol\beta_0)\xrightarrow{d}N(\mathbf{0},\mathbf{J}\mathbf{Q}^-\boldsymbol\Sigma\mathbf{Q}^-\mathbf{J}^T), \tag{11}$$

where $\mathbf{Q}^-$ is a generalized inverse of $\mathbf{Q}$,

$$\mathbf{Q}(h_2,h_3)\equiv\mathbf{Q}=E(\phi''_{h_3}(\varepsilon)\mathbf{V}\mathbf{V}^T)-E\left\{\frac{\mathbf{C}(U)\phi''_{h_3}(\varepsilon)E\left[\mathbf{g}'_0(U)\phi''_{h_2}(\varepsilon)\mathbf{X}^T\mathbf{J}|U,\boldsymbol{Z}\right]}{F(U,\mathbf{Z},h_2)}\right\},$$

$\boldsymbol\Sigma(h_2,h_3)\equiv\boldsymbol\Sigma=-E\left(\phi'_{h_2}(\varepsilon)E\{\mathbf{C}(U)\mathbf{D}^{-1}(U)\mathbf{Z}\phi''_{h_3}(\varepsilon)|U,\mathbf{Z}\}/F(U,\mathbf{Z},h_2)\right)^{\otimes 2}+E\left\{\phi'^2_{h_3}(\varepsilon)\mathbf{V}\mathbf{V}^T\right\}$ *and $\mathbf{V}$, $\mathbf{D}(u)$ and $\mathbf{C}(u)$ are defined in condition* (C4) *in the Appendix.*

Theorem 2. *Under the regularity conditions given in the Appendix, together with the root-n consistent estimator $\hat{\boldsymbol\beta}$ of $\boldsymbol\beta_0$ and that h_2 is constant and does not depend on n, we have*

$$\sqrt{nh_1}\left[\hat{\mathbf{g}}(u;\hat{\boldsymbol\beta})-\mathbf{g}_0(u)-\frac{1}{2}h_1^2\mu_2\mathbf{g}''_0(u)\right]\xrightarrow{d}N\left(\mathbf{0},\frac{\nu_0G(u,\mathbf{z},h_2)D^{-1}(u)}{f_U(u)F(u,\mathbf{z},h_2)^2}\right). \tag{12}$$

4. Asymptotic bandwidth

4.1. Asymptotic bandwidth for nonparametric part

Based on Theorem 2 and the nonparametric's asymptotic variance of the estimating equations (EE) given in Xue and Pang (2013), we can show that the ratio of the asymptotic variance of our proposed estimator to that of their estimator is defined by

$$r(u,\mathbf{z},h_2)\triangleq\frac{G(u,\mathbf{z},h_2)F^{-2}(u,\mathbf{z},h_2)}{\mathrm{Var}(\varepsilon|U=u,\mathbf{Z}=\mathbf{z})}. \tag{13}$$

Therefore, the ideal choice of h_2 is

$$h_{2\mathrm{opt}}=\arg\min_{h_2}r(u,\mathbf{z},h_2)=\arg\min_{h_2}G(u,\mathbf{z},h_2)F^{-2}(u,\mathbf{z},h_2). \tag{14}$$

From (14), we can see that $h_{2\text{opt}}$ is independent with n and only depends on the error distribution of ε given U and $\mathbf{Z}$. Moreover, we can obtain the asymptotic bandwidth $h_{1\text{opt}}$ by minimizing the mean square error (MSE) of $\hat{\boldsymbol{g}}(u;\boldsymbol{\beta})$ based on (12) and have

$$h_{1\text{opt}} = r(u, \mathbf{z}, h_{2\text{opt}})^{1/5} h_{\text{EE}}, \tag{15}$$

where h_{EE} is the asymptotic optimal bandwidth for the EE estimator (Xue and Pang, 2013). Furthermore, we can obtain the nonparametric asymptotic relative efficiency (ARE) between our proposed REE estimator with $h_{1\text{opt}}$ and $h_{2\text{opt}}$ and the EE estimator with h_{EE} is

$$\text{ARE} = \frac{\text{MSE(EE)}}{\text{MSE(REE)}} = r(u, \mathbf{z}, h_{2\text{opt}})^{-4/5}, \tag{16}$$

where $\text{MSE(REE)} = \text{Bias}(\hat{\boldsymbol{g}}(u;\boldsymbol{\beta}))^T \text{Bias}(\hat{\boldsymbol{g}}(u;\boldsymbol{\beta})) + \text{tr}(\text{Var}(\hat{\boldsymbol{g}}(u;\boldsymbol{\beta})))$.

4.2. Asymptotic bandwidth for parametric part

If we assume $h_2 = h_3$, $\mathbf{Q}$ and $\boldsymbol{\Sigma}$ defined in Theorem 1 become $\mathbf{Q}(h_3, h_3) \equiv \mathbf{Q} = E\Big\{\phi''_{h_3}(\varepsilon)(\mathbf{g}_0'^T(U)\mathbf{Z})^2\mathbf{J}^T\big(\mathbf{X} - E\big\{\mathbf{X}\phi''_{h_3}(\varepsilon)|U,\mathbf{Z}\big\}/F(U,\mathbf{Z},h_3)\big)^{\otimes 2}\mathbf{J}\Big\}$ and $\boldsymbol{\Sigma}(h_3,h_3) \equiv \boldsymbol{\Sigma} = -E\big(\phi'_{h_3}(\varepsilon)E\{\mathbf{C}(U)\mathbf{D}^{-1}(U)\mathbf{Z}\phi''_{h_3}(\varepsilon)|U,\mathbf{Z}\}/F(U,\mathbf{Z},h_3)\big)^{\otimes 2} + E\big\{\phi'^2_{h_3}(\varepsilon)\mathbf{V}\mathbf{V}^T\big\}$. Furthermore, if we assume that ε is independent of $\mathbf{X}$ and $\mathbf{Z}$, then $\mathbf{Q} = E\{\phi''_{h_3}(\varepsilon)\}E\Big\{(\mathbf{g}_0'^T(U)\mathbf{Z})^2\mathbf{J}^T\big(\mathbf{X} - E(\mathbf{X}|U)\big)^{\otimes 2}\mathbf{J}\Big\}$ and $\boldsymbol{\Sigma} = E\big\{\phi'^2_{h_3}(\varepsilon)\big\}\Big(E\{\mathbf{V}\mathbf{V}^T\} - E\{\mathbf{C}(U)\mathbf{D}^{-1}(U)\mathbf{C}^T(U)\}\Big)$. Therefore, we have the following results.

Corollary 1. *If $q = 1$ and $Z = 1$, then model (1) becomes single-index models. Under the same conditions as in Theorem 1, together with $h_2 = h_3$ and that ε is independent of $\boldsymbol{X}$ and Z, we have*

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \xrightarrow{d} N\left(\mathbf{0}, \frac{G(h_3)}{F(h_3)^2}\mathbf{J}\mathbf{Q}_1^-\mathbf{J}^T\right), \tag{17}$$

where $\mathbf{Q}_1 = E\left\{(g_0'(U))^2\mathbf{J}^T\,(\mathbf{X} - E(\mathbf{X}|U))^{\otimes 2}\,\mathbf{J}\right\}$, $F(h_3) = E\{\phi''_{h_3}(\varepsilon)\}$ *and* $G(h_3) = E\{\phi'_{h_3}(\varepsilon)^2\}$.

Remark 2. For single-index models, let $\hat{\boldsymbol{\beta}}^*$ be the estimator of $\boldsymbol{\beta}_0$ obtained from Liu et al. (2013). From Theorem 1 of their paper, together with the condition that ε is independent of $\mathbf{X}$ and Z and $h_2 = h_3$, the asymptotic covariance matrix of $\hat{\boldsymbol{\beta}}^*$ becomes $\frac{G(h_3)}{F(h_3)^2}\mathbf{Q}_2^-$, where $\mathbf{Q}_2 = E\left\{(g_0'(U))^2\,(\mathbf{X} - E(\mathbf{X}|U))^{\otimes 2}\right\}$. By Corollary 1 of Cui et al. (2011), it is easy to see that $\mathbf{Q}_2^- - \mathbf{J}\mathbf{Q}_1^-\mathbf{J}^T$ is a semi-positive definite matrix. This implies that the covariance matrix given in the Corollary 1 is smaller than $\frac{G(h_2)}{F(h_2)^2}\mathbf{Q}_2^-$. Therefore, Corollary 1 demonstrates that our estimator for $\boldsymbol{\beta}_0$ is more efficient than that of Liu et al. (2013) since we use the "remove-one component" method.

From Xue and Pang (2013), we see that the asymptotic variance of the EE estimator is $\mathbf{J}^T\mathbf{A}_1^{-1}\mathbf{A}_0\mathbf{A}_1^{-1}\mathbf{J}$, where $\mathbf{A}_1 = E\left\{(\mathbf{g}_0'^T(U)\mathbf{Z})^2\mathbf{J}^T\,(\mathbf{X} - E(\mathbf{X}|U))^{\otimes 2}\,\mathbf{J}\right\}$ and $\mathbf{A}_0 = E\{\varepsilon^2\mathbf{V}\mathbf{V}^T\} - E\{\varepsilon^2\mathbf{C}(U)\mathbf{D}^{-1}(U)\mathbf{C}^T(U)\}$.

Theorem 3. *Under the conditions given in Theorem 1, if $h_2 = h_3$, we have*

$$\inf_{h_3}\mathbf{Q}^-(h_3,h_3)\boldsymbol{\Sigma}(h_3,h_3)\mathbf{Q}^-(h_3,h_3) \le \mathbf{A}_1^{-1}\mathbf{A}_0\mathbf{A}_1^{-1} \tag{18}$$

and the equality holds true if we assume that ε is normal and independent of $\mathbf{X}$ and $\mathbf{Z}$. Meaning that the REE estimator $\hat{\boldsymbol{\beta}}$ is more efficient than the EE estimator.

If ε is independent of $\mathbf{X}$ and $\mathbf{Z}$ and $h_2 = h_3$, the ratio of the asymptotic variance of the REE estimator to that of the EE estimator is defined by

$$R(h_3) \triangleq \frac{\text{tr}(\mathbf{Q}^-\boldsymbol{\Sigma}\mathbf{Q}^-)}{\text{tr}\left(\mathbf{A}_1^{-1}\mathbf{A}_0\mathbf{A}_1^{-1}\right)} = \frac{G(h_3)}{F^2(h_3)\sigma^2}, \tag{19}$$

where σ^2 is the variance of ε. Then, the ideal choices of h_3 is given by

$$h_{3\text{opt}} = \arg\min_{h_3} G(h_3)F^{-2}(h_3). \tag{20}$$

5. Simulation study

5.1. Bandwidth selection in practice

In our simulation study, we assume the error ε is independent of $\mathbf{X}$ and $\mathbf{Z}$. Then, the estimates of $F(h_2)$ and $G(h_2)$ are defined by

$$\hat{F}(h_2) = \frac{1}{n}\sum_{i=1}^{n} \phi''_{h_2}(\hat{\varepsilon}_i) \quad \text{and} \quad \hat{G}(h_2) = \frac{1}{n}\sum_{i=1}^{n} \left\{\phi'_{h_2}(\hat{\varepsilon}_i)\right\}^2,$$

respectively. Then $r(h)$ can be estimated by $\hat{r}(h_2) = \hat{G}(h_2)\hat{F}^{-2}(h_2)/\hat{\sigma}^2$, where $\hat{\varepsilon}_i = Y_i - \hat{\boldsymbol{g}}(\hat{\boldsymbol{\beta}}^T\boldsymbol{X}_i)\boldsymbol{Z}_i$, $\hat{\boldsymbol{\beta}}$, $\hat{\boldsymbol{g}}$ and $\hat{\sigma}$ are estimated based on the pilot estimates. Then, we can easily find $h_{2\text{opt}}$ to minimize $\hat{r}(h_2)$ by using the grid search method. According to the advice of Yao et al. (2012), the possible grids points for h_2 are $h_2 = 0.5\hat{\sigma} \times 1.02^j$, $j = 0, 1, \ldots, k$, for some constant k (i.e., $k = 100$). After finding $h_{2\text{opt}}$, we can estimate $r(h_2)^{1/5}$ by $\hat{r}(h_{2\text{opt}})^{1/5}$. Let h_{opt} be the optimal bandwidth for estimating $\hat{\boldsymbol{g}(\cdot)}$, then, $\hat{h}_{\text{opt}} = \hat{h}_{4opt} = \hat{r}(h_{2\text{opt}})^{1/5}\hat{h}_{\text{EE}}$ if the optimal bandwidth h_{EE} is selected by 5-fold cross-validation. Similarly, we can obtain the estimator of $h_{3\text{opt}}$ based on (20). When calculating the estimator $\hat{\boldsymbol{\beta}}$, a recommended bandwidth for h_1 is $\hat{h}_{1\text{opt}} = \hat{h}_{\text{opt}} n^{-1/20}(\log n)^{-1/2}$, similar strategies are adopted by Xue and Pang (2013).

Remark 3. Let T be the full dataset, $T - T^v$ and T^v be the cross-validation training and test set, respectively, for $v = 1, \ldots, 5$. Given $\hat{\boldsymbol{\beta}}$, for each h and v, we find the estimator $\hat{\boldsymbol{g}}(u; \hat{\boldsymbol{\beta}})$ of $\mathbf{g}_0(u)$ using the training set $T - T^v$, then, we can form the fivefold cross-validation criterion as

$$CV(h) = \sum_{v=1}^{5} \sum_{(Y_i, \boldsymbol{X}_i)\in T^v} (Y_i - \hat{\boldsymbol{g}}^T(\hat{\boldsymbol{\beta}}^T\mathbf{X}_i; \hat{\boldsymbol{\beta}}, h)\mathbf{Z}_i)^2.$$

By using the grid search method, we find a $\hat{h} = \hat{h}_{\text{EE}}$ that minimizes the $CV(h)$.

5.2. Simulation

We use Gauss kernel for $K(.)$ and conduct a simulation study to assess the performance of the proposed estimators from two aspects, one is comparison of efficiency between our method (REE) and the EE (Xue and Pang, 2013), another is robustness of the proposed method with the existence of outliers and heavy tailed distribution. We generated data from the following model

$$Y_i = g_{01}\left(\boldsymbol{\beta}_0^T\boldsymbol{X}_i\right) + g_{01}\left(\boldsymbol{\beta}_0^T\boldsymbol{X}_i\right)Z_{i1} + g_{02}\left(\boldsymbol{\beta}_0^T\boldsymbol{X}_i\right)Z_{i2} + 0.6\varepsilon_i$$

where $\boldsymbol{\beta}_0 = (\beta_{01}, \beta_{02}) = \left(1/2, \sqrt{3}/2\right)^T$, $\mathbf{X}_i = (X_{i1}, X_{i2})^T$ are independent random vectors uniformly distributed on $[-1, 1]^2$, (Z_{i1}, Z_{i2}) are jointly standard normal distribution with correlation coefficient 0.5. The coefficient functions are $g_{01}(u) = \exp(-u)$, $g_{02}(u) = 3t^2$ and $g_{03}(u) = 5\cos(\pi u)$. For each dataset, we consider three error distributions: $N(0, 1)$, $t(3)$ and contaminated normal (CN) $0.95N(0, 1) + 0.05N(0, 100)$. The 5% data from $N(0, 100)$ are most likely to be outliers. The performances of $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{g}}$ are assessed by the absolute deviation (AD), $\text{AD}(\hat{\beta}_k) = \left|\hat{\beta}_k - \beta_{0k}\right|$, $k = 1, 2$ and the square root of average square errors (RASE), $\text{RASE}(\hat{g}_j) = \sqrt{\frac{1}{n_{\text{grid}}}\sum_{i=1}^{n_{\text{grid}}}\left\|\hat{g}_j(t_i) - g_{0j}(t_i)\right\|^2}$, $j = 1, 2, 3$. Here, $\left\{t_i,\ i = 1, \ldots, n_{\text{grid}}\right\}$ are the grid points at which the functions $\hat{\boldsymbol{g}}$ are evaluated. A total of 200 replications with $n = 100$, 200 are performed and the average and standard deviation of AD and RASE are reported in Table 1.

From Table 1 we can see that the estimates based on REE have great gain of efficiency and robustness over estimates based on EE for non-normal distribution errors. Moreover, the REE method performs no worse than the EE even for the normal error case, especially for $n = 200$. In addition, for the given error distribution, all the ADs and RASEs decrease with the sample size n increasing. These findings agree with our asymptotic properties. To conclude, the REE estimator is better than or at least as well as the EE estimator.

Acknowledgments

The authors are very grateful to anonymous referees for their detailed comments on the earlier version of the manuscript, which led to a much improved paper. This work is supported by the National Natural Science Foundation of China (Grant No. 11171361) and Ph.D. Programs Foundation of Ministry of Education of China (Grant No. 20110191110033).

Appendix

Our basis assumptions are as follows:

(C1) The functions $g_{0j}(u)$, $j = 1, \ldots, q$ have a continuous second derivative.

Table 1
Simulation results of the absolute deviation (AD) ($\times 10^2$), square root of average square errors (RASE) ($\times 10^2$) and their corresponding standard deviations (in parentheses $\times$ 10^2).

Error distribution	n	Method	Parametric part		Nonparametric part		
			β_{01}	β_{02}	g_0	g_1	g_2
$N(0, 1)$	100	EE	1.155 (1.401)	0.659 (0.733)	6.349 (5.819)	16.99 (23.04)	32.18 (28.21)
		REE	0.962 (0.850)	0.561 (0.505)	10.77 (27.48)	29.29 (68.49)	30.11 (48.16)
	200	EE	0.626 (0.931)	0.358 (0.490)	3.205 (2.630)	4.617 (3.615)	11.35 (8.133)
		REE	0.493 (0.362)	0.285 (0.211)	3.811 (2.886)	5.728 (4.148)	8.830 (6.758)
t_3	100	EE	2.681 (3.623)	1.612 (2.503)	13.71 (16.19)	22.97 (28.42)	51.70 (53.59)
		REE	1.193 (1.030)	0.698 (0.613)	13.38 (15.71)	21.93 (25.75)	30.83 (21.60)
	200	EE	1.225 (1.836)	0.691 (0.953)	6.439 (7.523)	14.08 (30.45)	22.67 (23.73)
		REE	0.822 (0.681)	0.478 (0.405)	6.002 (7.053)	12.69 (27.54)	15.98 (16.26)
CN	100	EE	2.537 (3.862)	1.457 (1.896)	20.91 (24.28)	30.28 (27.99)	60.17 (56.20)
		REE	1.177 (0.950)	0.689 (0.570)	13.97 (23.74)	21.55 (20.09)	38.49 (43.44)
	200	EE	2.357 (3.543)	1.279 (1.633)	9.213 (9.590)	15.69 (14.73)	32.68 (43.24)
		REE	0.707 (0.520)	0.411 (0.307)	4.784 (6.297)	9.583 (11.82)	15.47 (16.43)

(C2) The marginal density function of $U, f_U(u)$ is positive and has a continuous second derivative.
(C3) The kernel $K(u)$ is a symmetric density function with bounded support and satisfies the Lipschitz condition.
(C4) The matrix $\mathbf{D}(u) = E(\mathbf{Z}\mathbf{Z}^T|U = u)$ and $\mathbf{C}(u) = E(\mathbf{V}\mathbf{Z}^T|U = u)$ are continuous with respect to u. Furthermore, for given u, $\mathbf{D}(u)$ is positive definite, where $\mathbf{V} = \mathbf{J}^T\mathbf{X}\mathbf{g}_0'^T(U)\mathbf{Z}$.
(C5) $F(u, \mathbf{z}, h)$ and $G(u, \mathbf{z}, h)$ are continuous with respect to $\mathbf{z}$, u.
(C6) $F(u, \mathbf{z}, h) < 0$ for any $h > 0$.
(C7) $E\left(\phi'_h(\varepsilon)|U = u, \mathbf{Z} = \mathbf{z}\right) = 0$ and $E\left(\phi''_h(\varepsilon)^2|U = u, \mathbf{Z} = \mathbf{z}\right)$, $E\left(\phi'_h(\varepsilon)^3|U = u, \mathbf{Z} = \mathbf{z}\right)$ and $E\left(\phi'''_h(\varepsilon)|U = u, \mathbf{Z} = \mathbf{z}\right)$ are continuous with respect to $\mathbf{z}$, u.
(C8) $nh_1^2/\log^2 n \to \infty$, $nh_1^4 \log n \to 0$, $nh_1h_4^3/\log^2 n \to \infty$, $nh_4^5 = O(1)$.

Lemma 1. *Assume that conditions* (C1)–(C8) *hold, we have* $\frac{1}{n}\sum_{i=1}^n K_{h_1}(U_i - u)\phi''_{h_2}(\varepsilon_i)(\frac{U_i-u}{h_1})^j = F(u, \mathbf{z}, h_2)f_U(u)\mu_j + o_p(1)$, *and* $\frac{1}{n}\sum_{i=1}^n K_{h_1}(U_i-u)\phi''_{h_2}(\varepsilon_i)r_i(\frac{U_i-u}{h_1})^j = \frac{1}{2}h_1^2F(u, \mathbf{z}, h_2)f_U(u)\mathbf{g}''_0(u)\mu_{j+2}+o_p(h_1^2)$, *where* $\mathbf{r}_i = \mathbf{g}_0(U_i)-\mathbf{g}_0(u)-\mathbf{g}'_0(u)(U_i-u)$.

Proof of Lemma 1. The proofs are similar to those of Lemma 1 of Yao et al. (2012) based on the fact $\xi = E(\xi)+O_p(\sqrt{\mathrm{var}(\xi)})$. Hence, we omit it here.

Proof of Theorem 1. Let $U_i = \boldsymbol{\beta}_0^T\mathbf{X}_i$, $\hat{U}_i = \hat{\boldsymbol{\beta}}^T\mathbf{X}_i$, $\mathbf{Z}_i^* = \begin{pmatrix}\mathbf{Z}_i\\ \mathbf{Z}_i(U_i-u)/h_1\end{pmatrix}$, $\boldsymbol{\theta}_0 = \begin{pmatrix}\mathbf{g}_0(u)\\ h_1\mathbf{g}'_0(u)\end{pmatrix}$ and $\hat{\boldsymbol{\theta}} = \begin{pmatrix}\hat{\mathbf{a}}(u;\hat{\boldsymbol{\beta}})\\ h_1\hat{\mathbf{b}}(u;\hat{\boldsymbol{\beta}})\end{pmatrix}$. For a fixed u and the $\sqrt{n}$-consistent estimator $\hat{\boldsymbol{\beta}}$, the vector $\left(\hat{\mathbf{a}}^T(u;\hat{\boldsymbol{\beta}}), h_1\hat{\mathbf{b}}^T(u;\hat{\boldsymbol{\beta}})\right)^T$ maximizes (3) meaning that $\hat{\boldsymbol{\theta}}$ satisfies the following equation

$$\frac{1}{n}\sum_{i=1}^n \hat{\mathbf{Z}}_i^* K_{h_1}(\hat{U}_i - u)\phi'_{h_2}(Y_i - \hat{\boldsymbol{\theta}}^T\hat{\mathbf{Z}}_i^*) = \mathbf{0}, \tag{A.1}$$

where $\hat{\mathbf{Z}}_i^* = (\mathbf{Z}_i^T, \mathbf{Z}_i^T(\hat{U}_i - u)/h_1)^T$. Using $\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\| = O_p(n^{-1/2})$ and Taylor expansion by eliminating higher order terms, we obtain

$$\begin{aligned}\mathbf{0} = {} & \frac{1}{n}\sum_{i=1}^n K_{h_1}(U_i - u)\mathbf{Z}_i^*\phi'_{h_2}\left[Y_i - \mathbf{g}_0^T(u)\mathbf{Z}_i - \mathbf{g}_0'^T(u)\mathbf{Z}_i(U_i-u)\right] \\ & - \mathbf{B}_{n1}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) - \mathbf{B}_{n2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) + o_p(n^{-1/2}) + O_p(h_1^2),\end{aligned} \tag{A.2}$$

where $\mathbf{B}_{n1} = \frac{1}{n}\sum_{i=1}^n K_{h_1}(U_i - u)\mathbf{Z}_i^*\mathbf{Z}_i^{*T}\phi''_{h_2}\left[Y_i - \mathbf{g}_0^T(u)\mathbf{Z}_i - \mathbf{g}_0'^T(u)\mathbf{Z}_i(U_i-u)\right]$ and $\mathbf{B}_{n2} = \frac{1}{n}\sum_{i=1}^n K_{h_1}(U_i - u)\mathbf{Z}_i^*\mathbf{Z}_i^T\mathbf{g}_0'(u)\mathbf{X}_i^T$ $\phi''_{h_2}\left\{Y_i - \mathbf{g}_0^T(u)\mathbf{Z}_i - \mathbf{g}_0'^T(u)\mathbf{Z}_i \times (U_i-u)\right\}$. Solving (A.2) and using Lemma 1, we obtain

$$\begin{aligned}\hat{\mathbf{a}}(u;\hat{\boldsymbol{\beta}}) - \mathbf{g}_0(u) = {} & \frac{1}{n}\frac{\sum_{i=1}^n K_{h_1}(U_i-u)\phi'_{h_2}(\varepsilon_i)\mathbf{D}^{-1}(U_i)\mathbf{Z}_i}{\{f_U(u)F(u,z,h_2)\}} - \frac{E\left\{\mathbf{g}'_0(U)\phi''_{h_2}(\varepsilon)\mathbf{X}^T|U=u, \boldsymbol{Z}=\boldsymbol{z}.\right\}}{F(u,\mathbf{z},h_2)}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \\ & + o_p(n^{-1/2}) + O_p(h_1^2).\end{aligned} \tag{A.3}$$

Since $\hat{\boldsymbol{\beta}}$ is the solution of estimating equation (6), that is,

$$\mathbf{0} = \frac{1}{n}\sum_{i=1}^n \phi'_{h_3}\left[Y_i - \hat{\mathbf{g}}^T(\beta(\hat{\boldsymbol{\beta}}^{(1)})^T\mathbf{X}_i;\hat{\boldsymbol{\beta}})\mathbf{Z}_i\right]\hat{\mathbf{g}}'(\beta(\hat{\boldsymbol{\beta}}^{(1)})^T\mathbf{X}_i;\hat{\boldsymbol{\beta}})\mathbf{Z}_i\mathbf{J}^T\mathbf{X}_i. \tag{A.4}$$

By Taylor expansion,

$$\mathbf{0} = \left\{ \frac{1}{n}\sum_{i=1}^{n} \phi'_{h_3}(\varepsilon_i)\mathbf{g}_0'^{T}(\boldsymbol{\beta}_0^T\mathbf{X}_i)\mathbf{Z}_i\mathbf{J}^T\mathbf{X}_i - \frac{1}{n}\sum_{i=1}^{n} \phi''_{h_3}(\varepsilon_i)\left[\hat{\mathbf{g}}(\hat{\boldsymbol{\beta}}^T\mathbf{X}_i;\hat{\boldsymbol{\beta}}) - \mathbf{g}_0(\boldsymbol{\beta}_0^T\mathbf{X}_i)\right]^T \mathbf{Z}_i\mathbf{g}_0'^{T}(\boldsymbol{\beta}_0^T\mathbf{X}_i)\mathbf{Z}_i\mathbf{J}^T\mathbf{X}_i \right\} \times [1 + o_p(1)]. \tag{A.5}$$

Note that

$$\begin{aligned} \hat{\mathbf{g}}(\hat{\boldsymbol{\beta}}^T\mathbf{X}_i;\hat{\boldsymbol{\beta}}) - \mathbf{g}_0(\boldsymbol{\beta}_0^T\mathbf{X}_i) &= \hat{\mathbf{g}}(\hat{\boldsymbol{\beta}}^T\mathbf{X}_i;\hat{\boldsymbol{\beta}}) - \hat{\mathbf{g}}(\boldsymbol{\beta}_0^T\mathbf{X}_i;\hat{\boldsymbol{\beta}}) + \hat{\mathbf{g}}(\boldsymbol{\beta}_0^T\mathbf{X}_i;\hat{\boldsymbol{\beta}}) - \mathbf{g}_0(\boldsymbol{\beta}_0^T\mathbf{X}_i) \\ &= \hat{\mathbf{g}}'(\boldsymbol{\beta}_0^T\mathbf{X}_i;\hat{\boldsymbol{\beta}})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)^T\mathbf{X}_i + \hat{\mathbf{g}}(\boldsymbol{\beta}_0^T\mathbf{X}_i;\hat{\boldsymbol{\beta}}) - \mathbf{g}_0(\boldsymbol{\beta}_0^T\mathbf{X}_i) + o_p(n^{-1/2}) \\ &= \mathbf{g}'_0(\boldsymbol{\beta}_0^T\mathbf{X}_i)(\hat{\boldsymbol{\beta}}^{(1)} - \boldsymbol{\beta}_0^{(1)})^T\mathbf{J}^T\mathbf{X}_i + \hat{\mathbf{g}}(\boldsymbol{\beta}_0^T\mathbf{X}_i;\hat{\boldsymbol{\beta}}) - \mathbf{g}_0(\boldsymbol{\beta}_0^T\mathbf{X}_i) + o_p(n^{-1/2}). \end{aligned} \tag{A.6}$$

Substituting (A.6) into (A.5), we have

$$\begin{aligned} 0 = &\frac{1}{n}\sum_{i=1}^{n} \phi'_{h_3}(\varepsilon_i)\mathbf{g}_0'^{T}(\boldsymbol{\beta}_0^T\mathbf{X}_i)\mathbf{Z}_i\mathbf{J}^T\mathbf{X}_i - \frac{1}{n}\sum_{i=1}^{n} \phi''_{h_3}(\varepsilon_i)\left(\mathbf{g}_0'^{T}(\boldsymbol{\beta}_0^T\mathbf{X}_i)\mathbf{Z}_i\right)^2\mathbf{J}^T\mathbf{X}_i\mathbf{X}_i^T\mathbf{J}(\hat{\boldsymbol{\beta}}^{(1)} - \boldsymbol{\beta}_0^{(1)}) \\ &- \frac{1}{n}\sum_{i=1}^{n} \phi''_{h_3}(\varepsilon_i)\mathbf{g}_0'^{T}(\boldsymbol{\beta}_0^T\mathbf{X}_i)\mathbf{Z}_i\mathbf{J}^T\mathbf{X}_i\mathbf{Z}_i^T\left(\hat{\mathbf{g}}(\boldsymbol{\beta}_0^T\mathbf{X}_i;\hat{\boldsymbol{\beta}}) - \mathbf{g}_0(\boldsymbol{\beta}_0^T\mathbf{X}_i)\right) + o_p(1). \end{aligned} \tag{A.7}$$

Substituting (A.3) into (A.7), we obtain

$$\begin{aligned} 0 = &\frac{1}{n}\sum_{i=1}^{n} \phi'_{h_3}(\varepsilon_i)\mathbf{g}_0'^{T}(U_i)\mathbf{Z}_i\mathbf{J}^T\mathbf{X}_i - \mathbf{Q}(\hat{\boldsymbol{\beta}}^{(1)} - \boldsymbol{\beta}_0^{(1)}) - \frac{1}{n}\sum_{i=1}^{n} \phi''_{h_3}(\varepsilon_i)\mathbf{g}_0'^{T}(U_i)\mathbf{Z}_i\mathbf{J}^T\mathbf{X}_i\mathbf{Z}_i^T \\ &\times \frac{1}{n}\sum_{j=1}^{n} \frac{\mathbf{D}^{-1}(U_i)\mathbf{Z}_iK_{h_1}(U_j - U_i)\phi'_{h_2}(\varepsilon_j)}{f_U(U_i)F(U_i,\mathbf{Z}_i,h_2)} + o_p(1). \end{aligned} \tag{A.8}$$

Interchanging the summations, the third term on the right-hand side of (A.8) is

$$-\frac{1}{n}\sum_{i=1}^{n} \phi'_{h_2}(\varepsilon_i)\frac{1}{n}\sum_{j=1}^{n} \frac{K_{h_1}(U_j - U_i)\phi''_{h_3}(\varepsilon_j)\mathbf{g}_0'^{T}(U_j)\mathbf{Z}_j\mathbf{J}^T\mathbf{X}_i\mathbf{Z}_i^T\mathbf{D}^{-1}(U_i)\mathbf{Z}_i}{f_U(U_j)F(U_j,\mathbf{Z}_j,h_2)},$$

which is equivalent asymptotically to

$$-\frac{1}{n}\sum_{i=1}^{n} \phi'_{h_2}(\varepsilon_i)\frac{E\left\{\phi''_{h_3}(\varepsilon)\mathbf{D}^{-1}(U)\mathbf{J}^T\mathbf{X}\mathbf{g}_0'^{T}(U)\mathbf{Z}\mathbf{Z}^T\mathbf{D}^{-1}(U)\mathbf{Z}|U_i,\mathbf{Z}_i\right\}}{F(U_i,\mathbf{Z}_i,h_2)}. \tag{A.9}$$

Thus, combining (A.8) and (A.9), we obtain

$$\begin{aligned} \mathbf{Q}\sqrt{n}(\hat{\boldsymbol{\beta}}^{(1)} - \boldsymbol{\beta}_0^{(1)}) = n^{-1/2}\sum_{i=1}^{n} &\Big\{\phi'_{h_3}(\varepsilon_i)\mathbf{J}^T\mathbf{X}_i\mathbf{g}_0'^{T}(U_i)\mathbf{Z}_i \\ &- \phi'_{h_2}(\varepsilon_i)\frac{E\left\{\phi''_{h_3}(\varepsilon)\mathbf{J}^T\mathbf{X}\mathbf{g}_0'^{T}(U)\mathbf{Z}\mathbf{Z}^T\mathbf{D}^{-1}(U)\mathbf{Z}|U_i,\mathbf{Z}_i\right\}}{F(U_i,\mathbf{Z}_i,h_2)}\Big\}. \end{aligned} \tag{A.10}$$

It now follows from Eq. (5) that $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) = \sqrt{n}\mathbf{J}(\hat{\boldsymbol{\beta}}^{(1)} - \boldsymbol{\beta}_0^{(1)}) + O_p(n^{-1/2})$. Then, using the central limiting theorem and Slutsky theorem, we complete the proof of Theorem 1. □

Proof of Theorem 2. Note that $\sqrt{nh_1}\left(\hat{\mathbf{g}}(u;\hat{\boldsymbol{\beta}}) - \mathbf{g}_0(u)\right) = \sqrt{nh_1}\left(\hat{\mathbf{g}}(u;\hat{\boldsymbol{\beta}}) - \hat{\mathbf{g}}(u;\boldsymbol{\beta}_0)\right) + \sqrt{nh_1}\left(\hat{\mathbf{g}}(u;\boldsymbol{\beta}_0) - \mathbf{g}_0(u)\right)$ where $\hat{\mathbf{g}}(u;\boldsymbol{\beta}_0)$ is a local linear estimator of $\mathbf{g}_0(.)$ if $\boldsymbol{\beta}_0$ is a known constant. Similar to Theorems 1 and 3 in Zhang et al. (2013), we obtain

$$\sqrt{nh_1}\left[\hat{\mathbf{g}}(u;\boldsymbol{\beta}_0) - \mathbf{g}_0(u) - \frac{1}{2}h_1^2\mu_2\mathbf{g}''_0(u)\right] \xrightarrow{d} N\left(0, \frac{\nu_0 G(u,\mathbf{z},h_2)\mathbf{D}^{-1}(u)}{f_U(u)F(u,\mathbf{z},h_2)^2}\right).$$

By the root-n consistency estimate $\hat{\boldsymbol{\beta}}$ which is obtained from Theorem 1 and using the same argument as Theorem 1 in Yao et al. (2012), we can show that $\sqrt{nh_1}(\hat{\mathbf{g}}(u;\hat{\boldsymbol{\beta}}) - \hat{\mathbf{g}}(u;\boldsymbol{\beta}_0)) = o_p(1)$. Thus, we complete the proof of Theorem 2. □

Proof of Theorem 3. Using the same argument as Theorem 3 of Liu et al. (2013), we have $h_3^3 E\left\{\phi''_{h_3}(\varepsilon)|U=u, \mathbf{Z}=\mathbf{z}\right\} = -\phi(0)+o(1)$ and $h_3^6 E\left\{\phi'_{h_3}(\varepsilon)^2|U=u, \mathbf{Z}=\mathbf{z}\right\} = \phi^2(0)\mathrm{Var}(\varepsilon|U=u, \mathbf{Z}=\mathbf{z})+o(1)$ where $h_3 \to \infty$. Then, if $h_2=h_3$, let $\boldsymbol{\Sigma}(h_3)=\boldsymbol{\Sigma}(h_3,h_3)$ and $\mathbf{Q}(h_3)=\mathbf{Q}(h_3,h_3)$, we have

$$\begin{aligned}\boldsymbol{\Sigma}(h_3) &= E\left\{\phi_{h_3}'^2(\varepsilon)\boldsymbol{V}\boldsymbol{V}^T\right\} - E\left(\frac{\phi'_{h_3}(\varepsilon)E\{\mathbf{C}(U)\mathbf{D}^{-1}(U)\mathbf{Z}\phi''_{h_3}(\varepsilon)|U,\boldsymbol{Z}\}}{F(U,\mathbf{Z},h_3)}\right)^{\otimes 2}\\ &= E\left\{E\left(\phi_{h_3}'^2(\varepsilon)|U,\boldsymbol{Z}\right)\mathbf{V}\mathbf{V}^T\right\} - E\left(E\left(\phi_{h_3}'^2(\varepsilon)|U,\boldsymbol{Z}\right)\frac{E\{\mathbf{C}(U)\mathbf{D}^{-1}(U)\mathbf{Z}\phi''_{h_3}(\varepsilon)|U,\boldsymbol{Z}\}}{F(U,\mathbf{Z},h_3)}\right)^{\otimes 2}\\ &= h_3^{-6}\phi^2(0)E\left\{\varepsilon^2\left(\mathbf{V}\mathbf{V}^T-\mathbf{C}(U)\mathbf{D}^{-1}(U)\mathbf{C}^T(U)\right)\right\}+o(1),\end{aligned}$$

and

$$\begin{aligned}\mathbf{Q}(h_3) &= E\left\{(\phi''_{h_3}(\varepsilon)\mathbf{g}_0'^T(U)\mathbf{Z})^2\mathbf{J}^T\left(\frac{\mathbf{X}-E\{\mathbf{X}\phi''_{h_3}(\varepsilon)|U,\boldsymbol{Z}\}}{F(U,\mathbf{Z},h_3)}\right)^{\otimes 2}\mathbf{J}\right\}\\ &= -h_3^{-3}\phi(0)E\left\{\left(\mathbf{g}_0'^T(U)\mathbf{Z}\right)^2\mathbf{J}^T\left(\mathbf{X}-E\{\mathbf{X}|U,\boldsymbol{Z}\}\right)^{\otimes 2}\mathbf{J}\right\}+o_p(1)\end{aligned}$$

as $h_3 \to \infty$. Thus, $\lim_{h_3\to\infty}\mathbf{Q}^-(h_3)\boldsymbol{\Sigma}(h_3)\mathbf{Q}^-(h_3) = \mathbf{A}_1^-\mathbf{A}_0\mathbf{A}_1^-$ and we have $\inf_{h_3}\mathbf{Q}^-(h_3)\boldsymbol{\Sigma}(h_3)\mathbf{Q}^-(h_3) \le \inf_{h_3\to\infty}\mathbf{Q}^-(h_3)\boldsymbol{\Sigma}(h_3)\mathbf{Q}^-(h_3) = \mathbf{A}_1^{-1}\mathbf{A}_0\mathbf{A}_1^{-1}$. If ε is normal and independent of $\mathbf{X}$ and $\mathbf{Z}$, we can get $\inf_{h_3}\frac{E\{\phi'_{h_3}(\varepsilon)^2\}}{\sigma^2}E\{\phi''_{h_3}(\varepsilon)\}^2 = 1$ by Theorem 3 of Liu et al. (2013), where σ^2 is the variance of ε. Thus, $\inf_{h_3}\mathbf{Q}^-(h_3,h_3)\boldsymbol{\Sigma}(h_3,h_3)\mathbf{Q}^-(h_3,h_3) = \mathbf{A}_1^{-1}\mathbf{A}_0\mathbf{A}_1^{-1}$ holds. □

References

Chang, Z., Xue, L., Zhu, L., 2010. On an asymptotically more efficient estimation of the single-index model. J. Multivariate Anal. 101, 1898–1901.
Cui, X., Hädle, W.K., Zhu, L., 2011. The EFM approach for single-index models. Ann. Statist. 39, 1658–1688.
Fan, J., Yao, Q.W., Cai, Z.W., 2003. Adaptive varying-coefficient linear models. J. R. Stat. Soc. Ser. B 65, 57–80.
Härdle, W., Hall, P., Ichimura, H., 1993. Optimal smoothing in single-index models. Ann. Statist. 21, 157–178.
Hastie, T.J., Tibshirani, R., 1993. Varying-coefficient models. J. R. Stat. Soc. Ser. B 55, 757–796.
Huang, Z., Zhang, R., 2013. Profile empirical-likelihood inferences for the single-index-coefficient regression model. Stat. Comput. 23, 455–465.
Liu, J., Zhang, R., Zhao, W., Lv, Y., 2013. A robust and efficient estimation method for single index models. J. Multivariate Anal. 122, 226–238.
Xue, L., Pang, Z., 2013. Statistical inference for a single-index varying-coefficient model. Stat. Comput. 23, 589–599.
Xue, L., Wang, Q., 2012. Empirical likelihood for single-index varying-coefficient models. Bernoulli 18, 836–856.
Xue, L., Zhu, L., 2006. Empirical likelihood for single-index model. J. Multivariate Anal. 97, 1295–1312.
Yao, W., Lindsay, B., Li, R., 2012. Local modal regression. J. Nonparametr. Stat. 24, 647–663.
Zhang, R., Zhao, W., Liu, J., 2013. Robust estimation and variable selection for semiparametric partially linear varying coefficient model based on modal regression. J. Nonparametr. Stat. 25, 523–544.