

Effects of measurement error on a class of single-index varying coefficient regression models

Jianhong Shi¹ · Qian Yang¹ · Xiongya Li² ·
Weixing Song²

Received: 14 January 2016 / Accepted: 5 April 2017
© Springer-Verlag Berlin Heidelberg 2017

Abstract This paper investigates the estimation in a class of single-index varying coefficient regression model when some covariates are contaminated with measurement errors. A bias-corrected least square procedure based on the observed data is proposed. By replacing the nonparametric single index part with a local linear approximation, an iterative algorithm for estimating the index parameter is proposed. More importantly, a special case is identified in which the naive procedure provides consistent estimates for the single index parameters. Large sample properties of the proposed estimators are established. The finite sample performance of the proposed estimators are evaluated by simulation studies.

Keywords Bias-corrected estimate · Local linear smoothing · Consistency · Asymptotic normality

1 Introduction

Single index modeling, or the projection pursuit regression modeling has been proven to be one of the efficient venues to eliminate the curse of dimensionality, often encountered in nonparametric smoothing, while keeping the interpretability as in some parametric settings such as the linear regression modeling. Due to its wide application in economics, finance and biomedical studies, single index modeling has been received much attention and studied extensively in the past two decades. To accommodate more

✉ Weixing Song
weixing@ksu.edu

¹ College of Mathematics and Computer Sciences, Shanxi Normal University,
Linfen 041004, China

² Kansas State University, Manhattan, KS 66503, USA

complicated scenarios, many extensions of the single index modeling has been made. One of such extensions is the single index varying coefficient model which has the following form

$$Y = g^\tau(\beta^\tau Z)X + \varepsilon, \quad (1.1)$$

where g is a $q \times 1$ vector of functions, the p -dimensional vector β is called the single index parameter, Z and X are p -dimensional and q -dimensional covariates, respectively, and ε is the random error. For the sake of identifiability, we have to assume that the single index parameter β satisfies $\|\beta\| = 1$, and one of its components to be positive, where $\|\cdot\|$ denotes the Euclidean norm. Model (1.1) is general enough to include many other important statistical models. When $q = 1$ and $X = 1$, model (1.1) reduces to the single index model. The identifiability of the single index model is investigated by [Ichimura \(1993\)](#) and [Horowitz \(1998\)](#), many estimation procedures are proposed in literature, including the semiparametric least square and the maximum likelihood estimate, the method of average derivatives, sliced inverse regression, penalized spline estimate, as well as the estimating equation and the empirical likelihood method. See [Ichimura \(1993\)](#), [Härdle et al. \(1993\)](#), [Delecroix et al. \(2003\)](#), [Stocker \(1986\)](#), [Li \(1989, 1991\)](#), [Yu and Ruppert \(2002\)](#), [Xue and Zhu \(2006\)](#) among others; when $p = 1$, $\beta = 1$, and X is a vector of 1's, model (1.1) becomes the renowned additive model. An extensive introduction of the additive model can be found in [Hastie and Tibshirani \(1993\)](#), [Xia and Li \(1999\)](#) considered a more general form of (1.1) where the linear combination $\beta^\tau Z$ is replaced with a known parametric function. [Xue and Pang \(2013\)](#) proposed a semiparametric least square estimate for the single index parameter in model (1.1), as well as a local linear smoother for the nonparametric function g . An empirical likelihood confidence region for β is constructed in [Xue and Wang \(2012\)](#).

Often times, the covariates in a regression model are measured with certain amount of error. Simply using the error contaminated value replacing the true value are most likely to introduce nonnegligible bias to the classical estimation and results in inefficient testing procedures, if no particular bias correction techniques are adopted. A wide spectral of discussions on measurement error modeling are well documented in several monographs by [Fuller \(2006\)](#), [Cheng and Ness \(1999\)](#), [Carroll et al. \(2012\)](#) and the references therein. Many real world examples from agriculture, bioassay, medical and nutritional studies can be found in [Carroll et al. \(2012\)](#).

In this paper, we will consider a bias-corrected estimation procedure in model (1.1) when X is measured with an additive error, that is, instead of observing X , we can observe $W = X + U$, where U is the measurement error with zero mean vector and known covariance matrix Σ_u . It seems that a known Σ_u is a rather strict assumption, but this assumption indeed is commonly used in the measurement error literature due to some identifiability considerations. If, by any chance, this assumption cannot be satisfied, then replicated measurements on X would allow us to construct an estimate for Σ_u to be used in the estimation procedure. For the sake of brevity, we simply assume the covariance matrix Σ_u is known throughout this paper.

The current set up assumes that we can observe Z completely, and X is contaminated with measurement error, this implies that Z and X do not have identical components.

Therefore, we will not encounter the identifiability issue when the nonparametric function g takes some specific forms, as discussed in [Fan et al. \(2003\)](#), [Xue and Pang \(2013\)](#). In addition to [Xue and Pang \(2013\)](#), another three papers related to our current work include [Liang and Wang \(2005\)](#), [Li et al. \(2011\)](#), [You et al. \(2006\)](#). They either deal with the varying-coefficient models with error-prone covariates or partial linear single index model with measurement error in linear part.

The paper is organized as follows. The bias-corrected estimation procedure for β and the local linear smoother for g is introduced in Sect. 2. In this section, we also identified a special case in which the naive estimation procedure can also provide consistent estimates for the single index parameter β ; the large sample properties of the proposed estimates, as well as the needed technical assumptions, are presented in Sect. 3; Sect. 4 consists of a simulation study, and the proofs of the main results are postponed to Sect. 5.

In the sequel, for any vector or matrix a , a^τ denotes its transposition.

2 Bias-corrected estimation

We start this section with a rather interesting, if not striking, phenomenon related to the naive estimation procedure associated with the model under consideration. When X can be observed directly, according to [Xue and Pang \(2013\)](#), we may estimate β and g by exploiting the following equation

$$\sum_{i=1}^n [Y_i - g^\tau(\beta^\tau Z_i) X_i] \dot{g}^\tau(\beta^\tau Z_i) X_i J_{\beta(1)}^\tau Z_i w(\beta^\tau Z_i) = 0,$$

where w is a weight function which will be discussed more later. When X is not available, the naive method for estimating the unknown β and g is to replace X with W in the above equation and consider

$$\sum_{i=1}^n [Y_i - g^\tau(\beta^\tau Z_i) W_i] \dot{g}^\tau(\beta^\tau Z_i) W_i J_{\beta(1)}^\tau Z_i w(\beta^\tau Z_i) = 0.$$

It almost becomes a stereotype that the naive procedure often leads to biased estimates. However, this may not be the case for the current model. To see this, note that

$$E(Y|Z, W) = E[g^\tau(\beta^\tau Z)X + \varepsilon|Z, W] = g^\tau(\beta^\tau Z)E[X|Z, W]. \tag{2.1}$$

If $E[X|Z, W] = AW$ for some matrix A , then since g is unknown, so is $A^\tau g(\beta^\tau Z)$. Therefore, g and $A^\tau g(\beta^\tau Z)$ act equivalently in model (1.1). Accordingly, the naive procedure, or the least square procedure based on the regression model (2.1) also provides consistent estimates for β . An example such that $E[X|Z, W] = AW$ can be obtained when X, U and Z are independent, $X \sim N(0, \Sigma_x)$ and $U \sim N(0, \Sigma_u)$, respectively. In fact, the independence and normality implies $E(X|Z, W) = E(X|W) = \Sigma_x(\Sigma_x + \Sigma_u)^{-1}W$.

Note that even though the naive procedure could produce consistent estimates for β , it may not lead to consistent estimates for g and σ^2 . It is likely that if

$E(X|Z, W) \neq AW$ for any matrix A , biased estimates for β may be obtained by the naive procedure. However, we tried several distributions for U (for example, Laplace distribution, t -distribution, even some skewed distribution such as the centered exponential distribution), which leads to nonlinear structures for the conditional expectation, and found out that the naive estimates for the index parameters are all close to the bias-corrected ones. This puzzling fact might be partly explained by the self-adjustment to the measurement error when estimating the nonparametric functions, but the certainty of this claim needs a rigorous justification and a further study is necessary. That said, the bias-corrected estimation procedure proposed in the following will surely leads to consistent estimates for all the unknown parameters, including the nonparametric components g .

2.1 Estimation procedure

Assume that U , ε , and (Z, X) are independent. $\{(Y_i, Z_i, W_i)\}_{i=1}^n$ is a sample of size n from the following varying coefficient single index model with X being measured with error:

$$Y = g^\tau(\beta^\tau Z)X + \varepsilon, \quad W = X + U. \tag{2.2}$$

By the independence assumption, it is easy to show that

$$E[Y - g^\tau(\beta^\tau Z)W]^2 = E[Y - g^\tau(\beta^\tau Z)X]^2 + E[g^\tau(\beta^\tau Z)\Sigma_u g(\beta^\tau Z)].$$

When $\beta = \beta_0$, the true value of β , we have

$$E[Y - g^\tau(\beta^\tau Z)W]^2 = \sigma^2 + E[g^\tau(\beta^\tau Z)\Sigma_u g(\beta^\tau Z)]. \tag{2.3}$$

Therefore, if g is known, one can estimate β by the following nonlinear least squares procedure

$$\tilde{\beta}_n(g) = \operatorname{argmin}_{\beta \in \mathcal{B}} \sum_{i=1}^n \left([Y_i - g^\tau(\beta^\tau Z_i)W_i]^2 - g^\tau(\beta^\tau Z_i)\Sigma_u g(\beta^\tau Z_i) \right),$$

where \mathcal{B} is the parameter space of β . Suppose g is differentiable. By taking derivatives with respect to β , and set the derivatives equal to 0, then $\tilde{\beta}_n(g)$ is the solution of the following equation

$$\sum_{i=1}^n ([Y_i - g^\tau(\beta^\tau Z_i)W_i]\dot{g}^\tau(\beta^\tau Z_i)W_i + g^\tau(\beta^\tau Z_i)\Sigma_u \dot{g}(\beta^\tau Z_i)) J_{\beta^{(1)}}^\tau Z_i = 0 \tag{2.4}$$

subject to $\|\beta\| = 1$. Without loss of generality, we assume that the first component of β is positive. Then $J_{\beta^{(1)}}$ has the following form:

$$J_{\beta^{(1)}} = \begin{pmatrix} \frac{-\beta^{(1)\tau}}{\sqrt{1-\|\beta^{(1)}\|^2}} \\ I_{p-1} \end{pmatrix}$$

which is the Jacobian matrix $\partial\beta/\partial\beta^{(1)}$, where $\beta^{(1)}$ is the $p - 1$ -dimensional vector obtained by excluding the first component from β . In the measurement error free case, Xue and Pang (2013) suggest that, to control the boundary effect when estimating g and its derivative \dot{g} , one can eliminate some “extreme” X values from calculation. Applying the similar idea to our case, instead of solving (2.4), we can estimate β using the solution, still denoted as $\tilde{\beta}_n(g)$, of the following modified equation:

$$\sum_{i=1}^n ([Y_i - g^\tau(\beta^\tau Z_i)] \dot{g}^\tau(\beta^\tau Z_i) W_i + g^\tau(\beta^\tau Z_i) \Sigma_u \dot{g}(\beta^\tau Z_i)) J_{\beta^{(1)}}^\tau Z_i w(\beta^\tau Z_i) = 0, \tag{2.5}$$

where $w(\cdot)$ is a weight function with bounded support which will be specified later. In fact, the choice of the weight function may affect the efficiency of the resulting estimates and the ideal approach is to choose the weight function to minimize the asymptotic variance of the estimate. However, since the question of choosing optimal weigh function is not our research priority in the current work, so we will not pursue this possibility in the paper.

As a matter of fact, $\tilde{\beta}_n(g)$ is not a real estimator since the nonparametric function g is unknown. In the following, we shall use local linear smoothing technique to estimate the function g . For this purpose, suppose the j th component of g is $g_j(\cdot)$ and $g_j(\cdot)$ is sufficiently smooth. For any value V in the neighborhood of v , both from the domain of $g_j(\cdot)$, we can approximate $g_j(V)$ by a linear function of the form $g_j(V) = g_j(v) + \dot{g}_j(v)(V - v) := a_j + b_j(V - v)$. For the sake of brevity, denote $L_{i,j}(\beta, v) = a_j + b_j(\beta^\tau Z_i - v)$ and $\mathbf{a} = (a_1, \dots, a_q)$, $\mathbf{b} = (b_1, \dots, b_q)$. At a fixed point $\beta = \beta_0$, one can estimate the a_j 's and b_j 's at $V = v$ by minimizing the following weighted squares

$$(\tilde{\mathbf{a}}, \tilde{\mathbf{b}}) = \operatorname{argmin}_{\mathbf{a}, \mathbf{b}} \sum_{i=1}^n \left\{ \left[Y_i - \sum_{j=1}^q L_{i,j}(\beta_0, v) W_{ij} \right]^2 - \sum_{j=1}^q \sum_{k=1}^q \sigma_{jk} L_{i,j}(\beta_0, v) L_{i,k}(\beta_0, v) \right\} K_h(\beta_0^\tau Z_i - v),$$

where $\sigma_{i,j}$ is the (i, j) th element of Σ_u , K is a kernel density function, h is a bandwidth depending on the sample size n . For $j = 0, 1, 2, k = 0, 1$, denote

$$S_{n,j}(v; \beta_0) = \sum_{i=1}^n K_h(\beta_0^\tau Z_i - v) \left(\frac{\beta_0^\tau Z_i - v}{h} \right)^j (W_i W_i^\tau - \Sigma_u),$$

$$b_{n,k}(v; \beta_0) = \sum_{i=1}^n K_h(\beta_0^\tau Z_i - v) \left(\frac{\beta_0^\tau Z_i - v}{h} \right)^k W_i Y_i.$$

Then a routine least square procedure leads to

$$\begin{pmatrix} \tilde{\mathbf{a}} \\ h\tilde{\mathbf{b}} \end{pmatrix} = \begin{pmatrix} S_{n,0}(v; \beta_0) & S_{n,1}(v; \beta_0) \\ S_{n,1}(v; \beta_0) & S_{n,2}(v; \beta_0) \end{pmatrix}^{-1} \begin{pmatrix} b_{n,0}(v; \beta_0) \\ b_{n,1}(v; \beta_0) \end{pmatrix}$$

Note that $\tilde{\mathbf{a}}, \tilde{\mathbf{b}}$ are functions of β_0 and v . Thus, $g(v)$ can be estimated by $\hat{g}(v, \beta_0) = \tilde{\mathbf{a}}$, and $\dot{g}(v)$ can be estimated by $\hat{\dot{g}}(v; \beta_0) = h\tilde{\mathbf{b}}$. Finally, we can estimate β by the solution of the following equation:

$$\sum_{i=1}^n ([Y_i - \hat{g}^\tau(\beta^\tau Z_i, \beta_0)W_i] \hat{\dot{g}}^\tau(\beta^\tau Z_i, \beta_0) W_i + g^\tau(\beta^\tau Z_i, \beta_0) \Sigma_u \hat{\dot{g}}(\beta^\tau Z_i, \beta_0)) J_{\beta^{(1)}}^\tau Z_i w(\beta^\tau Z_i, \beta_0) = 0 \tag{2.6}$$

subject to $\|\beta\| = 1$. Denote the solution as $\hat{\beta}_n = \tilde{\beta}_n(\hat{g})$. Thus $g(\cdot)$ can be estimated by $\hat{g}(v; \hat{\beta}_n)$. From (2.3), σ^2 can be estimated by

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n \left\{ [Y_i - \hat{g}^\tau(\hat{\beta}_n^\tau Z_i, \hat{\beta}_n)W_i]^2 - \hat{g}^\tau(\hat{\beta}_n^\tau Z_i, \hat{\beta}_n) \Sigma_u \hat{\dot{g}}(\hat{\beta}_n^\tau Z_i, \hat{\beta}_n) \right\}. \tag{2.7}$$

The above arguments indeed provide an iterated algorithm to implement the proposed estimation procedure.

-
- Step 1:* Choose an initial value of β_0 with $\|\beta_0\| = 1$;
 - Step 2:* Estimate $g(\cdot)$ by $\tilde{\mathbf{a}}$ with $\beta = \beta_0$;
 - Step 3:* Estimate β by $\tilde{\beta}_n$ using (2.6);
 - Step 4:* Set $\hat{\beta}_n$ as the initial value, repeat Step 1 to Step 3 until convergence. A commonly used convergence criterion is to stop the iteration whenever the L_1 -distance between the last two iterations is less than a pre-specified threshold;
 - Step 5:* Estimate $g(\cdot)$ by $\hat{g}(v, \hat{\beta}_n)$, and $\hat{\sigma}^2$ by (2.7).
-

2.2 Some technical considerations

In this section, we explore some technical considerations when implementing the above algorithm.

Although the algorithm is well aligned with the methodology development, certain modification can speed up the computation. Similar to Xue and Pang (2013) in error-free case, we can estimate β by the solution to the following equations:

$$\sum_{i=1}^n ([Y_i - \hat{g}^\tau(\beta^\tau Z_i, \beta)W_i] \hat{\dot{g}}^\tau(\beta^\tau Z_i, \beta) W_i + g^\tau(\beta^\tau Z_i, \beta) \Sigma_u \hat{\dot{g}}(\beta^\tau Z_i, \beta)) J_{\beta^{(1)}}^\tau Z_i w(\beta^\tau Z_i, \beta) = 0 \tag{2.8}$$

subject to the constraint $\|\beta\| = 1$. By doing this, we successfully circumvent the iteration steps.

In the error-free case, [Xue and Pang \(2013\)](#) indicated that, to solve the similar equation, a Fisher scoring type of Newton–Raphson algorithm might be needed. However, no details are disclosed. For the sake of completeness, we briefly sketch a Newton–Raphson type algorithm to solve (2.8) here.

For convenience, use $T_n(\beta)$ to denote the left hand side of Eq. (2.8). Reparametrizing $T_n(\beta)$ as a function of $\beta^{(1)}$, the solution $\hat{\beta}_n^{(1)}$ satisfies $T_n(\hat{\beta}_n^{(1)}) = 0$. By Taylor expansion at $\beta_0^{(1)}$, the presumed true value of $\beta^{(1)}$, we obtain

$$0 = T_n(\hat{\beta}_n^{(1)}) \approx T_n(\beta_0^{(1)}) + \dot{T}_n(\beta_0^{(1)})(\hat{\beta}_n^{(1)} - \beta_0^{(1)}).$$

Hence an iterated algorithm for solving $\hat{\beta}_n^{(1)}$ can be built upon

$$\hat{\beta}_n^{(1)} = \beta_0^{(1)} - \dot{T}_n^{-1}(\beta_0^{(1)})T_n(\beta_0^{(1)}).$$

Note that

$$\dot{T}_n(\beta) = - \sum_{i=1}^n [\dot{g}^\tau(\beta^\tau Z_i)(W_i W_i^\tau - \Sigma_u)\dot{g}(\beta^\tau Z_i)] J_{\beta^{(1)}}^\tau Z_i Z_i J_{\beta^{(1)}} w(\beta^\tau Z_i) + R_n,$$

where the remainder term R_n consists several sums whose summands either include the residual $Y_i - g^\tau(\beta^\tau Z_i)X_i$ or $U_i W_i^\tau - \Sigma_u$. After dividing R_n by n , all these averages converge to 0 when $\beta = \beta_0$, the true value of β . Therefore, we can only use the first term, denoted by $\dot{T}_{n1}(\beta)$, as an approximation of $\dot{T}_n(\beta)$. That is, the iteration is based upon

$$\hat{\beta}_n^{(1)} = \beta_0^{(1)} - \dot{T}_{n1}^{-1}(\beta_0^{(1)})T_n(\beta_0^{(1)}).$$

In fact, Several R-functions can be used to directly solve the Eq. (2.8) when p is small, for example, the `unroot` and `nlm` functions are very efficient and effective when $p = 2$ or 3, 4.

It is well known that the bandwidth selection plays a crucial role in nonparametric smoothing. The estimation procedure (2.8) involves both g and its derivative \dot{g} . If the same bandwidth is used, the estimate \hat{g} has a slower convergence rate than $\hat{\dot{g}}$, which may slow down the convergence of $\hat{\beta}$. To address this issue, different bandwidths may be used. In the error-free case, [Xue and Pang \(2013\)](#) suggests to use bandwidth $\hat{h} = h_{\text{opt}} n^{-1/20} (\log n)^{-1/2}$ in estimating g , and h_{opt} in estimating \dot{g} , where h_{opt} is determined by a multi-fold cross validation procedure proposed by [Cai et al. \(2000\)](#). A rule-of-thumb selection for h_{opt} is to choose the ones proportional to $n^{-1/5}$.

It is also known that the nonlinear minimization procedure is very sensitive to the initial values, in particular, if the nonlinear target function has multiple modes. We suggest to use the naive estimate as the initial values, that is, using the solution to the following equation

$$\sum_{i=1}^n [Y_i - \hat{g}^\tau(\beta^\tau Z_i, \beta) W_i] \dot{g}^\tau(\beta^\tau Z_i, \beta) W_i J_{\beta^{(1)}}^\tau Z_i w(\beta^\tau Z_i, \beta) = 0 \tag{2.9}$$

to initiate the iteration process. However, solving (2.9) may need an extra iteration, and a new initial values is needed. It is noted that the above target function is exactly the same as the error-free case discussed in Xue and Pang (2013), so the initial value may be obtained by fitting the data with a generalized linear regression model. When p is large, finding a proper initial value in nonlinear minimization procedure remains an open question.

3 Main results

The following technical conditions are needed to derive the large sample properties of the proposed estimates. It is noted that most of the conditions are similar to those stated in Xue and Pang (2013), except for some notations defined differently due to the measurement error.

- (C1) The density function $f(t)$ of $\beta^\tau Z$ is bounded away from zero for all $t \in \mathcal{I}$ and β in a neighborhood of β_0 . Further more, The density function $f(t)$ is Lipschitz continuous of order 1 on \mathcal{I} .
- (C2) The weight function w is supported on \mathcal{I} .
- (C3) The functions $g_j(t)$, $1 \leq j \leq q$, have bounded, continuous, second order derivatives on I , where g_j is the j -th component of g .
- (C4) The bandwidths h and h_1 satisfy $nh^2/\log^2 n \rightarrow \infty$, $nh^4 \log n \rightarrow 0$, $nhh_1^3/\log^2 n \rightarrow \infty$, $nh_1^5 = O(1)$, $h_1^2 \log n/h \rightarrow 0$, where h is the bandwidth when estimating g , and h_1 is the bandwidth when estimating \dot{g} .
- (C5) The kernel function K is symmetric bounded probability density function supported on $(-1, 1)$.
- (C6) $D(t) = E(XX^\tau | \beta_0^\tau Z = t)$ is a positive definite matrix, and every component of $D(t)$ and $C(t) = E(VX^\tau | \beta_0^\tau Z = t)$ are Lipschitz continuous with order 1 on \mathcal{I} , where $V = J_{\beta^{(v)}}^\tau ZX' \dot{g}_0(\beta_0^\tau Z) w(\beta_0^\tau Z)$.
- (C7) The conditional density functions $\beta_0^\tau Z$ given W is bounded.

The uniform convergence of the local linear estimates \hat{g} and $\hat{\dot{g}}$ is stated in the following theorem.

Theorem 3.1 *Suppose that condition (C1)–(C3) and (C5)–(C7) hold. If the same bandwidth h is used in estimating g and \dot{g} , $h \rightarrow 0$, $\log n/(nh^3) \rightarrow 0$, then almost surely,*

$$\begin{aligned} \sup_{u \in \mathcal{I}, \beta \in \mathcal{B}_n} \|\hat{g}(u; \beta) - g_0(u)\| &= O\left(\sqrt{\frac{\log n}{nh}} + h^2\right), \\ \sup_{u \in \mathcal{I}, \beta \in \mathcal{B}_n} \|\hat{\dot{g}}(u; \beta) - \dot{g}_0(u)\| &= O\left(\sqrt{\frac{\log n}{nh^3}} + h\right), \end{aligned}$$

where $\mathcal{B}_n = \{\beta : \|\beta - \beta_0\| \leq cn^{-1/2}\}$ for some constant c .

More discussion on \mathcal{B}_n can be found in [Xue and Pang \(2013\)](#). Note that the convergence rates of \hat{g} and \hat{g}^τ are the same as in the error free case. The following theorem is about the asymptotic normality of the proposed local linear estimates.

Theorem 3.2 *Suppose that condition (C1)–(C3) and (C5)–(C7) hold. Then almost surely,*

$$\sqrt{nh} \begin{pmatrix} \sqrt{R(K)/f(v)} H^{1/2}(v) D^{-1}(v) [\tilde{\mathbf{a}} - g_0(v) - 2^{-1} h^2 \mu_2(K) D(v) \ddot{g}_0(v) f(v)] \\ \sqrt{\int x^2 K^2(x) dx / f(v) \mu_2^2(K) H^{1/2}(v) D^{-1}(v) [h\tilde{\mathbf{b}} - h\dot{g}_0(v)]} \end{pmatrix} \implies N(0, I),$$

where

$$H(v) = \sigma_\varepsilon^2 D(v) + g_0^\tau(v) \Sigma_u g_0(v) D(v) + \sigma_\varepsilon^2 \Sigma_u + E g_0^\tau(v) U U^\tau g_0(v) U U^\tau - \Sigma_u g_0(v) g_0^\tau(v) \Sigma_u,$$

and $D(v)$ is defined in Condition (C6).

Note that when $\Sigma_u = 0$ or the variable X can be observed directly, the above results reduce to the error free case discussed in [Xue and Pang \(2013\)](#).

For any $\beta \in \mathcal{B}_n$, from the proof of Theorem 3.2, the following corollary can be derived,

Corollary 3.1 *Under the conditions of Theorem 3.2, for all $\beta \in \mathcal{B}_n$,*

$$\hat{g}(v; \beta) - g_0(v) = \frac{D^{-1}(v)}{nf(v)} \sum_{i=1}^n K_h(\beta_0^\tau Z_i - v) [W_i(\varepsilon_i - U_i^\tau g_0(v)) + \Sigma_u g_0(v)] + \frac{h^2}{2} \mu_2(K) \ddot{g}_0(v) + c_n,$$

where $c_n = o_p(n^{-1/2}) + o_p(h^2)$.

The asymptotic normality of $\hat{\beta}_n$ is summarized in the following theorem.

Theorem 3.3 *Suppose the conditions (C1)–(C8) hold. Then*

$$\sqrt{n}(\hat{\beta}_n - \beta_0) \implies N(0, J_{\beta_0} B^{-1} A B^{-1} J_{\beta_0}^\tau),$$

where

$$A = \text{Var}((\varepsilon - \beta_0^\tau g_0(\beta_0^\tau Z)) \zeta W + \zeta \Sigma_u g_0(\beta_0^\tau Z)),$$

$$B = E [V V^\tau - V X^\tau \dot{g}_0(\beta_0^\tau Z) E(Z^\tau | \beta_0^\tau Z) J_{\beta_0^\tau}],$$

and V is defined in (C6), $\zeta = J_{\beta_0}^\tau Z \dot{g}_0^\tau(\beta_0^\tau Z) w(\beta_0^\tau Z) - C(\beta_0^\tau Z) D^{-1}(\beta_0^\tau Z)$.

The matrix B is the same as the one in [Xue and Pang \(2013\)](#), while extra terms can be found in the matrix A , indicating that the asymptotic variances of the proposed estimate become larger due to the extra uncertainty caused by the measurement errors. Clearly, when measurement error vanishes, the result stated in the above theorem is reduced to the error free case.

Finally, the consistency of the estimate of $\hat{\sigma}^2$ defined in (2.7) is summarized in the following theorem.

Theorem 3.4 *Suppose the conditions (C1)–(C8) hold. Then $\hat{\sigma}^2 \rightarrow \sigma^2$ in probability as $n \rightarrow \infty$.*

4 Numerical studies

In this section, a simulation study is conducted to evaluate the finite sample performance of the proposed methodology. We also conduct a sensitivity study by applying the proposed estimation procedure to the Boston housing data set.

4.1 Simulation study

The data generating procedure is similar to the one used in [Xue and Pang \(2013\)](#) except the measurement error structure is introduced to the X variables in model (1.1), and some true values for the unknown parameters. To be specific, in this simulation study, the data are generated from the single-index varying coefficient model with three components

$$Y_i = g_0(\beta_0^\tau Z_i)X_{i0} + g_1(\beta_0^\tau Z_i)X_{i1} + g_2(\beta_0^\tau Z_i)X_{i2} + \varepsilon_i,$$

where the true values $\beta_0 = (\beta_{01}, \beta_{02})^\tau = (1/5, \sqrt{24}/5)^\tau$, $Z_i = (Z_{i1}, Z_{i2})^\tau$, $i = 1, 2, \dots, n$. Instead of observing X_i 's, we observe $W_{ik} = X_{ik} + U_{ik}$, $k = 0, 1, 2$. In the simulation, $Z_i = (Z_{i1}, Z_{i2})^\tau$ are generated from uniform distribution on $[-1, 1]^2$, $(X_{i0}, X_{i1}, X_{i2})^\tau$ follows a multivariate normal distribution with mean vector 0, and $\text{Var}(X_{i0}) = \text{Var}(X_{i1}) = \text{Var}(X_{i2}) = 1$, and correlation coefficients being all 0. The measurement error $U_i = (U_{i0}, U_{i1}, U_{i2})$ are also normally distributed with mean vector 0 and $\text{Var}(U_{i0}) = \text{Var}(U_{i1}) = \text{Var}(U_{i2}) = 0.3^2$, and correlation coefficients being all 0. The random error $\varepsilon_i \sim N(0, 1)$. The coefficient functions are $g_0(u) = \exp(-u)$, $g_1(u) = 3u^3$, and $g_2(u) = 5 \cos(\pi u)$.

Various sample sizes, $n = 100, 300, 400, 500, 800$ are used in the simulation. The kernel function for the local smoother is chosen to be the Epanechnikov kernel for which the condition (C5) is satisfied, and the bandwidths are chosen according to the rules stated in Sect. 2.2. We didn't use any bandwidth selection procedure such as MCV in this simulation study, instead, $h = an^{-1/5}$ with different a -values are tried to investigate the sensitivity of the proposed estimation procedures related to the bandwidth selection. For each scenario, the estimation procedure is repeated 200 times. The mean square errors (MSE) of these 200 estimated β -values are calculated.

For the three nonparametric components, we plot the average curves of all the 200 estimated curves at 100 equally spaced values of v from $[-1, 1]$.

The iterated algorithm provided in Sect. 2.1 has a simpler form for the current data generating model. In fact, the parameter $\beta = (\beta_1, \beta_2)$ can be reparameterized as $\beta^{(1)} = (\sqrt{1 - \beta_2^2}, \beta_2)$, so the Jacobian matrix $J_{\beta^{(1)}} = (-\beta_2/\sqrt{1 - \beta_2^2}, 1)^\tau$. Therefore,

$$T_n(\beta) = \sum_{i=1}^n ([Y_i - a_{1i}W_{i0} - a_{2i}W_{i1} - a_{3i}W_{i2}][b_{1i}W_{i0} + b_{2i}W_{i1} + b_{3i}W_{i2}] + \sigma_{1,1}[a_{1i}b_{1i} + a_{2i}b_{2i} + a_{3i}b_{3i}] + 2\sigma_{12}[a_{1i}b_{2i} + a_{1i}b_{3i} + a_{2i}b_{3i}]) \cdot (-\beta_2 Z_{i1}/\sqrt{1 - \beta_2^2} + Z_{i2})w(\beta^\tau Z_i)$$

and the first derivative of $T_n(\beta)$, now only a function of the univariate β_2 , has the form of

$$\dot{T}_n(\beta) = - \sum_{i=1}^n \left([b_{1i}W_{i0} + b_{2i}W_{i1} + b_{3i}W_{i2}]^2 - \sigma_{1,1}[b_{1i}b_{1i} + b_{2i}b_{2i} + b_{3i}b_{3i}] - 2\sigma_{12}[b_{1i}b_{2i} + b_{1i}b_{3i} + b_{2i}b_{3i}] \right) \cdot \left(\frac{\beta_2^2 Z_{i1}^2}{1 - \beta_2^2} - \frac{2\beta_2 Z_{i1}Z_{i2}}{\sqrt{1 - \beta_2^2}} + Z_{i2}^2 \right) w(\beta^\tau Z_i).$$

In the simulation, the iteration will be terminated whenever the L_1 -distance between the last two iterations is less than 10^{-6} . The Epanechnikov kernel function is used in the local linear smoother, and similar to Xue and Pang (2013), the weight function is taken to be the indicator function $[-(1 + \sqrt{24})/5, (1 + \sqrt{24})/5]$.

Table 1 reports the biases and MSEs of the bias-corrected and the naive estimates. No any evidence shows that the measurement error introduces the bias to the single index parameter estimates, but it does affect the estimate of the variance ε . Given the reason discussed in Sect. 2.2, this is not beyond our expectation. The simulation is repeated several times and we found out that occasionally, the biases from the bias-corrected estimates are slightly larger than those of the naive estimates, but we believe this is due to the randomness rather than an intrinsic property. Figure 1 shows the histogram of β -estimates from the proposed bias-corrected procedure, the solid curve is a kernel density estimate drawn using R embedded function. Note that although the majority of the estimated values is centered around the true values, the

Table 1 Biases and MSEs of bias-corrected and naive estimates

	Bias-corrected			Naive		
	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\sigma}^2$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\sigma}^2$
Bias	0.0010	-0.0002	-0.1126	0.0046	-0.0013	1.1028
MSE	0.0003	0.0000	0.0364	0.0002	0.0000	1.3600

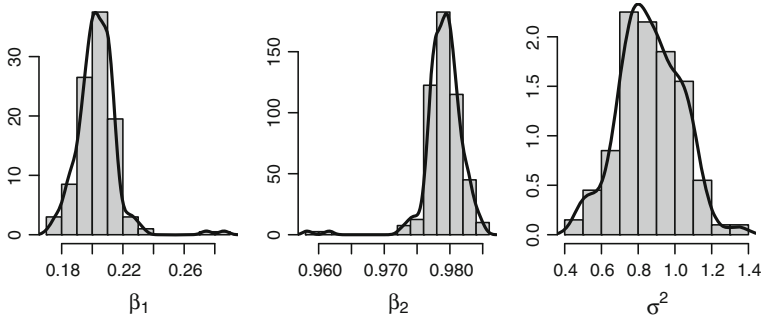


Fig. 1 Parameter estimates (bias-corrected, $n = 500$)

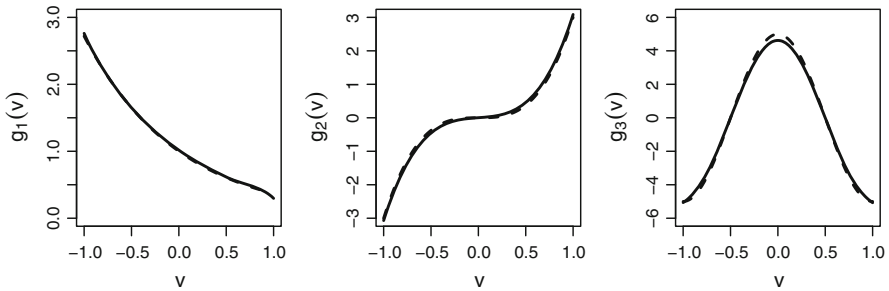


Fig. 2 Nonparametric estimates (bias-corrected, $n = 500$)

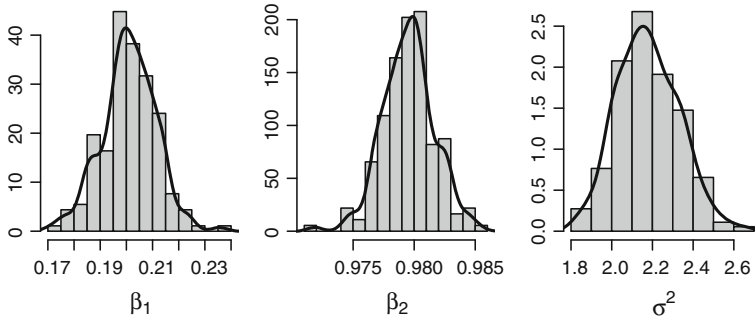


Fig. 3 Parameter estimates (naive, $n = 500$)

potential outliers do indicate the algorithm may not be very stable. Figure 2 shows the estimated curves for three nonparametric functions. In each plot, the dashed line is the true function curve, and the solid line is the fitted curve. Overall, the estimated curves fit the true functions very well except for some small regions. Figure 3 shows the histogram of β -estimates from the naive estimation procedure. Similar to the findings in Table 1, it seems that the measurement error has little effect when estimating the single index parameter. However, non-negligible and systematic bias is evident when estimating the variance parameter σ^2 . Figure 4 shows the estimated curves for three nonparametric functions using the naive procedure. Although the estimated curve for

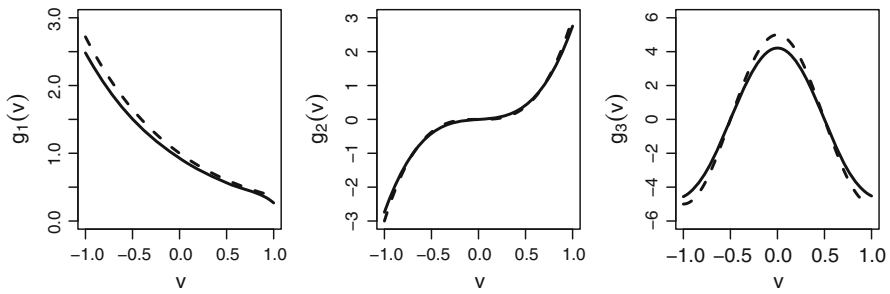


Fig. 4 Nonparametric estimates (naive, $n = 500$)

Table 2 Biases and MSEs of bias-corrected and naive estimates, correlated case

	Bias-corrected			Naive		
	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\sigma}^2$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\sigma}^2$
Bias	0.0008	-0.0002	-0.1439	0.0003	-0.0001	0.5380
MSE	0.0001	0.0000	0.0119	0.0001	0.0000	0.2993

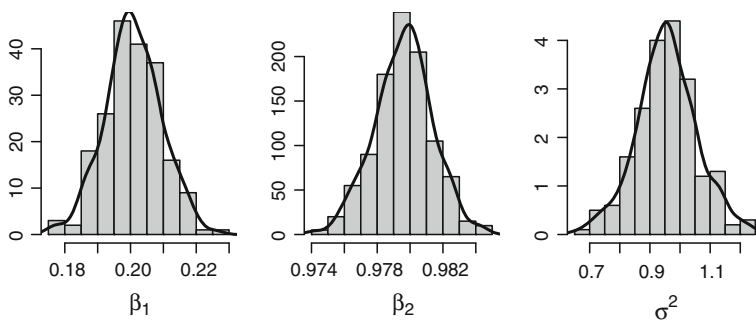


Fig. 5 Parameter estimates (bias-corrected, $n = 500$)

the second nonparametric component fits the true curve very well, the bias resulted from the measurement error on the other two estimated curves is obvious.

We also tried a correlated case in which the X -vector has a multivariate normal distribution with mean 0 and $\text{Var}(X_{i0}) = \text{Var}(X_{i1}) = \text{Var}(X_{i2}) = 1$, and covariance being all 0.5. The measurement error U -vector is also normally distributed with mean vector 0 and $\text{Var}(U_{i0}) = \text{Var}(U_{i1}) = \text{Var}(U_{i2}) = 0.2^2$, and covariance being all 0.1². Similar patterns as in Table 1 and Figs. 1, 2, 3 and 4 are obtained. For example, Table 2 reports the biases and MSEs of the estimated parameters. Figures 5, 6, 7 and 8 show the distribution of the estimated parameter values and the nonparametric curves.

The simulation results reported above is for $a = 1$. We also tried other values ($a = 0.5, 0.8, 1.2$) and the simulation results are pretty stable. The sample sizes affect the simulation results substantially. In particular, when the sample size gets smaller, the bias-corrected procedure has more chance to produce unreasonable solutions, while

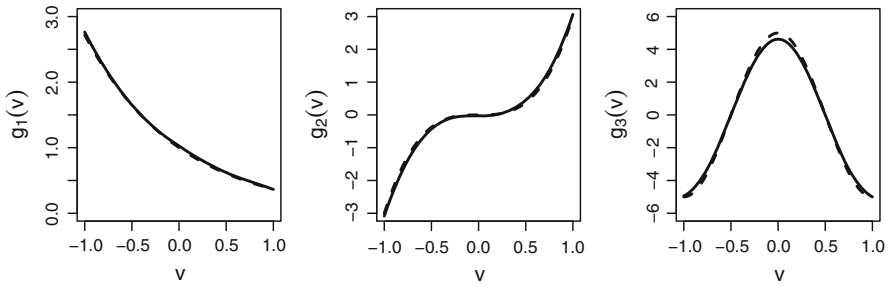


Fig. 6 Nonparametric estimates (bias-corrected, $n = 500$)

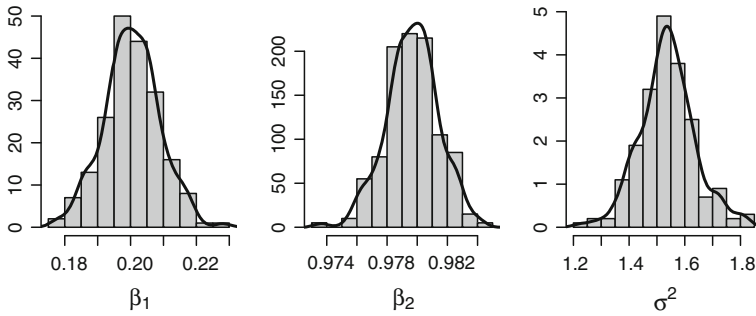


Fig. 7 Parameter estimates (naive, $n = 500$)

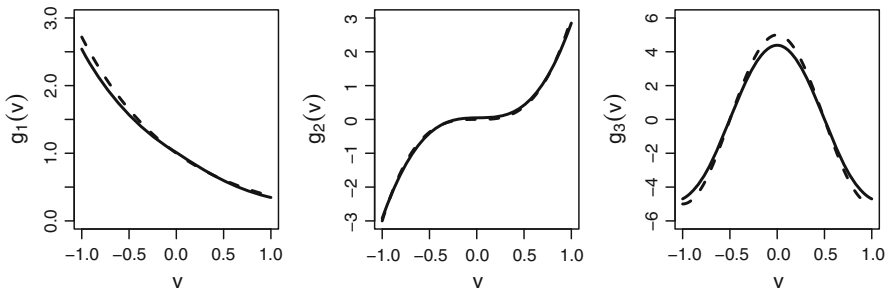


Fig. 8 Nonparametric estimates (naive, $n = 500$)

the naive procedure seems to work well. However, this by no means implies the bias-corrected procedure is inferior to the naive procedure, nevertheless, this indeed tells us more efficient algorithm should be designed for solving the Eq. (2.6).

4.2 Sensitivity study: Boston housing data

In this subsection, we conduct a sensitivity study using the well-known Boston housing data set, which contains the housing information in Boston, Massachusetts and it has been used extensively in the literature to investigate the relationship between MEDV

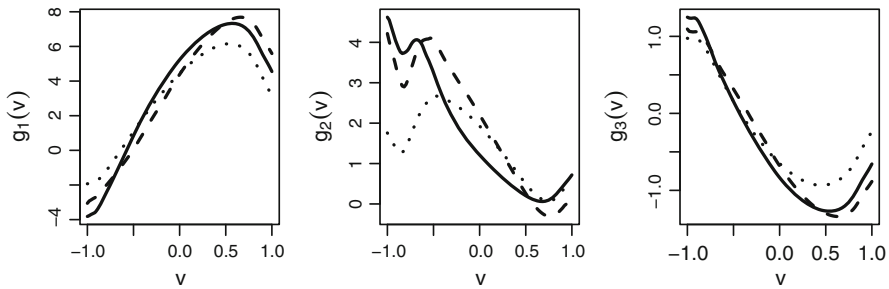


Fig. 9 Boston housing data

(the median value of owner-occupied homes) and other thirteen covariates, the number of cases is 506. For more details on the data set, see [Harrison and Rubinfeld \(1978\)](#).

We will fit a single index varying coefficient regression model to the Boston housing data. The response variable Y is chosen to be MEDV, and five other variables will be treated as covariates, they are RM (the average number of rooms per dwelling), DIS (the weighted distances to five Boston employment centers), PTRATIO (pupil-teacher ratio by town), B (a transformed value of the proportion of blacks by town), and LSTAT (lower status of the population). To ensure convergence, we transform the data from variables B and LSTAT into values in $[-1, 1]$ by simple linear transformation. For simplicity, we denote RM, DIS, PTRATIO, and the transformed values of B, LSTAT as X_0, X_1, X_2 , and Z_1, Z_2 , respectively.

To see how sensitive the fitting results to the magnitude of the measurement errors, we add three independent normal random errors U_0, U_1, U_2 with mean 0 and variance σ_u^2 to X_0, X_1, X_2 , respectively. That is, the surrogates are $W_0 = X_0 + U_0, W_1 = X_1 + U_1$, and $W_2 = X_2 + U_2$. Several values of σ_u^2 around 0.5^2 are tried in the study, only the case of $\sigma_u^2 = 0.5^2$ are reported here due to the similarity of the fitting results.

Three scenarios are considered, estimation using the true values from X -covariates, bias-corrected estimation, and naive estimation. The estimates of the index parameter (β_1, β_2) are $(0.9121, 0.4101)$, $(0.9139, 0.4060)$, and $(0.9191, 0.3940)$. These almost identical results are well aligned with the findings from the simulation study, and once again remind us that if the error-contaminated covariates are tangled with the nonparametric parts, the effects on index parameter estimates might be negligible. Figure 9 shows the plots of the nonparametric functions g_1, g_2 and g_3 , with solid lines for the estimates using the true X -values, dashed lines for the bias-corrected estimates, and dotted lines for the naive estimates. Clearly, comparing with the dotted lines, the dashed lines are more closer to the solid lines for all three nonparametric functions.

5 Proofs of main results

The following lemma is needed to facilitate the proof of the asymptotic normality of $\hat{\beta}_n$, which is similar to the Lemma A.2 in [Xia and Li \(1999\)](#) and Lemma 1 in [Xue and](#)

Pang (2013). Since we only deal with the i.i.d. case, so the proof is less involved, and this also can be seen from the simpler conditions imposed on the model.

Lemma 5.1 *Suppose $\varphi(t)$ is a bounded function and has bounded derivatives on the closed interval $[a, b]$, and $\{(U_i, \xi_i)\}$ are i.i.d. with $E|U|^r < \infty$ and $[E|\xi|^r]^{1/r} \leq c < \infty$ for all $r > 1$, and the conditional density functions $f_{\xi|U}$ is bounded. Then if (C5) holds, then*

$$\sup_{u \in [a, b]} \left| \frac{1}{n} \sum_{i=1}^n [\psi_i(u)] - E\psi_1(u) \right| = O\left(\sqrt{\frac{\log n}{nh}}\right), \quad a.s.$$

where $\psi_i(u) = K_h((U_i - u))\varphi((U_i - u)/h)\xi_i, i = 1, 2, \dots, n$.

Proof of Lemma 5.1 Denote $\xi_{in}(u) = n^{-1}\psi_i(u) - E[n^{-1}\psi_i(u)]$. Then $n^{-1} \sum_{i=1}^n [\psi_i(u) - E\psi_1(u)]$ is simply the $\sum_{i=1}^n \xi_{in}(u)$. Denote $\mu(u) = E(\xi|U = u)$, and $\tau^2(u) = E(\xi^2|U = u)$. Then

$$\begin{aligned} E[\psi_i(u)] &= EK_h((U - u))\varphi(U)\xi = \frac{1}{h} \int K\left(\frac{v - u}{h}\right)\varphi\left(\frac{v - u}{h}\right)\mu(v)f(v)dv \\ &= \mu(u)f(u) \int K(v)\phi(v)dv + h[\mu(u)\dot{f}(u) + \dot{\mu}(u)\phi(u)] \\ &\quad \times \int vK(v)\phi(v)dv + U(h^2). \end{aligned}$$

where $U(a_n)$ denotes the uniform order, and

$$\begin{aligned} E[\psi_i(u)]^2 &= EK_h^2((U - u))\varphi^2((U - u)/h)\xi^2 \\ &= \frac{1}{h^2} \int K^2\left(\frac{v - u}{h}\right)\varphi^2\left(\frac{v - u}{h}\right)\tau^2(v)f(v)dv \\ &= h^{-1}\tau^2(u)f(u) \int K^2(v)\phi^2(v)dv + u(h^{-1}). \end{aligned}$$

Therefore, $\text{Var}(\psi_i(u)) = h^{-1}\tau^2(u)f(u) \int K^2(v)\phi^2(v)dv + u(h^{-1})$ and

$$\text{Var}\left(\sum_{i=1}^n \xi_{in}(u)\right) = \frac{1}{n}\text{Var}(\psi_i(u)) = \frac{1}{nh}\tau^2(u)f(u) \int K^2(v)\phi^2(v)dv + U\left(\frac{1}{nh}\right).$$

The compact support of K implies the boundedness of $K(u)\phi(u)$. Therefore, for any $r > 2$,

$$E|\xi_{in}(u)|^r \leq r! \left(\frac{c}{nh}\right)^{r-2} E|\xi_{in}(u)|^2$$

for some constant c not depending on n and x . Thus, ξ_{in} satisfies the Cramér’s condition. For any positive constant d , by Bernstein inequality,

$$P \left\{ \frac{|\sum_{i=1}^n \xi_{in}(u)|}{\sqrt{\sum_{i=1}^n E |\xi_{in}(u)|^2}} > d\sqrt{n} \right\} \leq 2 \exp \left\{ -\frac{d^2 \log n}{4(1 + 2dcn^{-1/2}h^{-1/2}\sqrt{\log n})} \right\}.$$

By the assumption on h , we can see that, for any $u \in [a, b]$, for n large enough,

$$P \left\{ \frac{|\sum_{i=1}^n \xi_{in}(u)|}{\sqrt{\sum_{i=1}^n E |\xi_{in}(u)|^2}} > d\sqrt{n} \right\} \leq 2 \exp \left\{ -\frac{d^2 \log n}{8} \right\} = \frac{2}{n^8}.$$

by taking $d = 8$. Denote $C(f, \tau, \phi, K) = 8\sqrt{R(K\phi)\tau^2(u)f(u)}$ and $R(K\phi) = \int K^2(v)\phi^2(v)dv < \infty$, then for n large enough,

$$P \left(\left| \sum_{i=1}^n \xi_{in}(u) \right| > C(f, \tau, \phi, K)\sqrt{\frac{\log n}{nh}} \right) \leq \frac{2}{n^8}.$$

To bound the sum uniformly for all $u \in [a, b]$, we discretize by equally spaced $a = u_0 < u_1 < \dots < u_N = b$, $N = n^4$. Easy to see that

$$P \left\{ \max_{j=0}^N \left| \sum_{i=1}^n \xi_{in}(u_j) \right| > C(f, \tau, \phi, K)\sqrt{\frac{\log(n)}{nh}} \right\} \leq \frac{2(N + 1)}{n^8} < \frac{3}{n^4}.$$

Hence

$$\sum_{n=1}^{\infty} P \left\{ \max_{j=0}^N \left| \sum_{i=1}^n \xi_{in}(u_j) \right| > C(f, \tau, \phi, K)\sqrt{\frac{\log(n)}{nh}} \right\} \leq \sum_{n=1}^{\infty} \frac{3}{n^4} < \infty.$$

Borel–Cantelli lemma implies that

$$\max_{j=0}^N \left| \sum_{i=1}^n \xi_{in}(u_j) \right| = O \left(\sqrt{\frac{\log(n)}{nh}} \right).$$

Then if K is Lipschitz continuous,

$$\begin{aligned} \sup_{u \in [a, b]} \left| \sum_{i=1}^n \xi_{in}(u) \right| &\leq \max_{j=0}^N \left| \sum_{i=1}^n \xi_{in}(u_j) \right| + \max_{j=0}^{N-1} \sup_{u \in [u_j, u_{j+1}]} \left| \sum_{i=1}^n \xi_{in}(u_j) - \sum_{i=1}^n \xi_{in}(u) \right| \\ &= O \left(\sqrt{\frac{\log(n)}{nh}} \right) + \max_{j=0}^{N-1} \sup_{u \in [u_j, u_{j+1}]} \left| \sum_{i=1}^n \xi_{in}(u_j) - \sum_{i=1}^n \xi_{in}(u) \right| \\ &\leq O \left(\sqrt{\frac{\log(n)}{nh}} \right) + \frac{c}{n^4} \cdot \frac{1}{h^2} = O \left(\sqrt{\frac{\log(n)}{nh}} \right) \end{aligned}$$

for some constant $c > 0$. This completes the proof. □

Proof of Theorem 3.1 Without loss of generality, we shall assume that all Z_i 's are bounded. First we derive some asymptotic results for $S_{nj}(v; \beta)$ and $b_{n,k}(v; \beta)$. Without loss of generality, we assume the dimension of W to be 1. The general cases can be argued elementwise. For $j = 0, 1, 2$, denote $\phi(t) = t^j$, $\xi_i = W_i^2 - \Sigma_u$ or $W_i Y_i$, $U_i = \beta^\tau Z_i$. Applying Lemma 5.1 leads to

$$\sup_{u \in \mathcal{I}, \beta \in \mathcal{B}_n} |S_{n,j}(u; \beta) - ES_{n,j}(u; \beta)| = O\left(\sqrt{\frac{\log n}{nh}}\right), \quad \text{a.s.}$$

$$\sup_{u \in \mathcal{I}, \beta \in \mathcal{B}_n} |b_{n,j}(u; \beta) - Eb_{n,j}(u; \beta)| = O\left(\sqrt{\frac{\log n}{nh}}\right), \quad \text{a.s.}$$

Denote $\eta_\beta(u) = E(W^2|\beta^\tau Z = u) - \Sigma_u$, then we have $ES_{n0}(u; \beta) = \eta_\beta(u) f(u) + U(h^2)$, $ES_{n1}(u; \beta) = h[\eta_\beta(u)\dot{f}(u) + \dot{\eta}_\beta(u)f(u)]\mu_2(K) + U(h^2)$, $ES_{n2}(u; \beta) = \eta_\beta(u)f(u)\mu_2(K) + U(h^2)$. Denote $\eta_{g\beta}(u) = E[WY|\beta^\tau Z = u]$, then $Eb_{n,0}(u, \beta) = \eta_{g\beta}(u)f(u) + U(h^2)$ and $Eb_{n,1}(u, \beta) = h[\eta_{g\beta}(u)\dot{f}(u) + \dot{\eta}_{g\beta}(u)f(u)]\mu_2(K) + U(h^2)$. Therefore,

$$\begin{aligned} & \begin{pmatrix} ES_{n,0}(u; \beta) & ES_{n,1}(u; \beta) \\ ES_{n,1}(u; \beta) & ES_{n,12}(u; \beta) \end{pmatrix}^{-1} \begin{pmatrix} Eb_{n,0}(u; \beta) \\ Eb_{n,1}(u; \beta) \end{pmatrix} \\ &= \frac{1}{\eta_\beta^2(u)f^2(u)\mu_2(K) + U(h^2)} \\ & \quad \begin{pmatrix} \eta_\beta(u)\eta_{g\beta}(u)f^2(u)\mu_2(K) + U(h^2) \\ h[\eta_\beta(u)\dot{\eta}_{g\beta}(u) - \dot{\eta}_\beta(u)\eta_{g\beta}(u)]\mu_2(K)f^2(u) + U(h^2) \end{pmatrix}. \end{aligned}$$

Since $E(X|\beta_0^\tau Z = u)$ and $E(X^2|\beta_0^\tau Z = u)$ are Lipschitz continuous with order 1, so for any u such that $|u - u_0| = O(1/\sqrt{n})$ we have $\sup_{u:|u-u_0|} |\eta_{\beta_0}(u) - \eta_{\beta_0}(u_0)| = O(1/\sqrt{n})$ and $\sup_{u:|u-u_0|} |\eta_{g\beta_0}(u) - \eta_{g\beta_0}(u_0)| = O(1/\sqrt{n})$, where

$$\begin{aligned} \eta_{g\beta_0}(u) &= E[YW|\beta_0^\tau Z = u] = E[(g_0(\beta_0^\tau Z)X + \varepsilon)(X + U)|\beta_0^\tau Z = u] \\ &= g_0(u)E[X^2|\beta_0^\tau Z = u]. \end{aligned}$$

Note that $\eta_\beta(u) - \eta_{\beta_0}(u) = E[X^2|\beta_0^\tau Z = u + (\beta_0 - \beta)^\tau Z] - E[X^2|\beta_0^\tau Z = u]$. Therefore,

$$\sup_{u \in \mathcal{I}, \beta \in \mathcal{B}_n} |\eta_\beta(u) - \eta_{\beta_0}(u)| \leq c \cdot \sup_{\beta \in \mathcal{B}_n} |(\beta_0 - \beta)^\tau Z| = O(n^{-1/2}).$$

Similar result holds for $\eta_{g\beta}(u)$. These imply that

$$\begin{aligned} & \begin{pmatrix} ES_{n,0}(u; \beta) & ES_{n,1}(u; \beta) \\ ES_{n,1}(u; \beta) & ES_{n,12}(u; \beta) \end{pmatrix}^{-1} \begin{pmatrix} Eb_{n,0}(u; \beta) \\ Eb_{n,1}(u; \beta) \end{pmatrix} \\ &= \frac{1}{\eta_{\beta_0}^2(u)f^2(u)\mu_2(K)} \begin{pmatrix} \eta_{\beta_0}(u)\eta_{g\beta_0}(u)f^2(u)\mu_2(K) \\ h[\eta_{\beta_0}(u)\dot{\eta}_{g\beta_0}(u) - \dot{\eta}_{\beta_0}(u)\eta_{g\beta_0}(u)]\mu_2(K)f^2(u) \end{pmatrix} \\ & \quad + U(h^2) + U(n^{-1/2}). \end{aligned}$$

Calculation shows that

$$\frac{1}{\eta_{\beta_0}^2(u) f^2(u) \mu_2(K)} \left(\begin{array}{c} \eta_{\beta_0}(u) \eta_{g\beta_0}(u) f^2(u) \mu_2(K) \\ h[\eta_{\beta_0}(u) \dot{\eta}_{g\beta_0}(u) - \dot{\eta}_{\beta_0}(u) \eta_{g\beta_0}(u)] \mu_2(K) f^2(u) \end{array} \right) = \begin{pmatrix} g_0(u) \\ h \dot{g}_0(u) \end{pmatrix}.$$

This concludes the proof of Theorem 3.1. □

The following Theorem provides an asymptotic expansion of the nonparametric estimate \hat{g} , which will be used in the proof of our main result. As a by-product, the asymptotic normality of the local linear estimate can be derived easily.

Proof of Theorem 3.2 First we write $(\tilde{\mathbf{a}}^\tau, h\tilde{\mathbf{b}}^\tau)^\tau - (g_0(v)^\tau, hg'_0(v)^\tau)^\tau$ as

$$\begin{pmatrix} S_{n,0}(v; \beta_0) & S_{n,1}(v; \beta_0) \\ S_{n,1}(v; \beta_0) & S_{n,2}(v; \beta_0) \end{pmatrix}^{-1} \left[\begin{pmatrix} b_{n,0}(v; \beta_0) \\ b_{n,1}(v; \beta_0) \end{pmatrix} - \begin{pmatrix} S_{n,0}(v; \beta_0) & S_{n,1}(v; \beta_0) \\ S_{n,1}(v; \beta_0) & S_{n,2}(v; \beta_0) \end{pmatrix} \begin{pmatrix} g_0(v) \\ hg'_0(v) \end{pmatrix} \right] \tag{5.1}$$

For the inverse matrix, note that $D(v) = E[XX^\tau | \beta_0^\tau Z = v] = E[WW^\tau - \Sigma_u | \beta_0^\tau Z = v]$ by the independence of the measurement error U from random entities. So, routine argument shows that

$$\begin{pmatrix} S_{n,0}(v; \beta_0) & S_{n,1}(v; \beta_0) \\ S_{n,1}(v; \beta_0) & S_{n,2}(v; \beta_0) \end{pmatrix}^{-1} = f^{-1}(v) \begin{pmatrix} D(v) & 0 \\ 0 & \mu_2(K)D(v) + u_p(1) \end{pmatrix}^{-1}. \tag{5.2}$$

Note the second factor of the right hand side of (5.1) can be written as

$$\begin{pmatrix} b_{n,0}(v; \beta_0) - S_{n,0}(v; \beta_0)g_0(v) - S_{n,1}(v; \beta_0)hg'_0(v) \\ b_{n,1}(v; \beta_0) - S_{n,1}(v; \beta_0)g_0(v) - S_{n,2}(v; \beta_0)hg'_0(v) \end{pmatrix} = \begin{pmatrix} Q_{n0}(v; \beta_0) \\ Q_{n1}(v; \beta_0) \end{pmatrix}.$$

Let's consider the first element in the above vector. A straightforward calculation shows that

$$\begin{aligned} Q_{n0}(v, \beta_0) &= \frac{1}{n} \sum_{i=1}^n K_h(\beta_0^\tau Z_i - v) X_i [Y_i - g_0^\tau(v) X_i - (\beta_0^\tau Z_i - v) g'_0(v) X_i] \\ &+ \frac{1}{n} \sum_{i=1}^n K_h(\beta_0^\tau Z_i - v) U_i [Y_i - g_0^\tau(v) X_i - (\beta_0^\tau Z_i - v) g'_0(v) X_i] \\ &- \frac{1}{n} \sum_{i=1}^n K_h(\beta_0^\tau Z_i - v) X_i U_i^\tau [g_0(v) + (\beta_0^\tau Z_i - v) g'_0(v)] \\ &- \frac{1}{n} \sum_{i=1}^n K_h(\beta_0^\tau Z_i - v) [(U_i U_i^\tau - \Sigma_u) g_0(v) \\ &+ (\beta_0^\tau Z_i - v) (U_i U_i^\tau - \Sigma_u) g'_0(v)]. \end{aligned}$$

Note that

$$Y_i - g_0^\tau(v)X_i - (\beta_0^\tau Z_i - v)g_0^{\prime\tau}(v)X_i = \varepsilon_i + \frac{1}{2}(\beta_0^\tau Z_i - v)^2 \ddot{g}^\tau(\tilde{v})X_i,$$

where \tilde{v} is between $\beta_0^\tau Z_i$ and v . Therefore

$$\begin{aligned} Q_{n0}(v, \beta_0) &= \frac{1}{n} \sum_{i=1}^n K_h(\beta_0^\tau Z_i - v)W_i \varepsilon_i \\ &\quad + \frac{1}{2n} \sum_{i=1}^n K_h(\beta_0^\tau Z_i - v)W_i(\beta_0^\tau Z_i - v)^2 \ddot{g}^\tau(\tilde{v})X_i \\ &\quad - \frac{1}{n} \sum_{i=1}^n K_h(\beta_0^\tau Z_i - v)X_i U_i^\tau [g_0(v) + (\beta_0^\tau Z_i - v)g_0'(v)] \\ &\quad - \frac{1}{n} \sum_{i=1}^n K_h(\beta_0^\tau Z_i - v)[(U_i U_i^\tau - \Sigma_u)g_0(v) \\ &\quad + (\beta_0^\tau Z_i - v)(U_i U_i^\tau - \Sigma_u)g_0'(v)]. \end{aligned}$$

The bounded continuity of \ddot{g} implies that

$$\frac{1}{2n} \sum_{i=1}^n K_h(\beta_0^\tau Z_i - v)W_i(\beta_0^\tau Z_i - v)^2 \ddot{g}^\tau(\tilde{v})X_i = \frac{h^2}{2} \mu_2(K)D(v)\ddot{g}(v)f(v) + o(h^2).$$

Also, note that

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n K_h(\beta_0^\tau Z_i - v)X_i U_i^\tau (\beta_0^\tau Z_i - v)g_0'(v) &= O_p\left(\frac{h}{n}\right), \\ \frac{1}{n} \sum_{i=1}^n K_h(\beta_0^\tau Z_i - v)(\beta_0^\tau Z_i - v)(U_i U_i^\tau - \Sigma_u)g_0'(v) &= O_p\left(\frac{h}{n}\right). \end{aligned}$$

Therefore, we have

$$\begin{aligned} Q_{n0}(v, \beta_0) &= \frac{1}{n} \sum_{i=1}^n K_h(\beta_0^\tau Z_i - v)W_i \varepsilon_i + \frac{h^2}{2} \mu_2(K)D(v)\ddot{g}(v)f(v) \\ &\quad - \frac{1}{n} \sum_{i=1}^n K_h(\beta_0^\tau Z_i - v)X_i U_i^\tau g_0(v) \\ &\quad - \frac{1}{n} \sum_{i=1}^n K_h(\beta_0^\tau Z_i - v)(U_i U_i^\tau - \Sigma_u)g_0(v) \\ &\quad + o_p(h^2) + O_p\left(\frac{h}{n}\right) \end{aligned}$$

$$\begin{aligned}
 &= \frac{1}{n} \sum_{i=1}^n K_h(\beta_0^\tau Z_i - v) [W_i(\varepsilon_i - U_i^\tau g_0(v)) + \Sigma_u g_0(v)] \\
 &\quad + \frac{h^2}{2} \mu_2(K) D(v) \ddot{g}(v) f(v) \\
 &\quad + o_p(h^2) + O_p\left(\frac{h}{n}\right).
 \end{aligned}$$

Similarly, one can obtain

$$\begin{aligned}
 Q_{n1}(v, \beta_0) &= \frac{1}{nh} \sum_{i=1}^n K_h(\beta_0^\tau Z_i - v)(\beta_0^\tau Z_i - v) [W_i(\varepsilon_i - U_i^\tau g_0(v)) + \Sigma_u g_0(v)] \\
 &\quad + o_p(h^2) + O_p\left(\frac{h}{n}\right).
 \end{aligned}$$

Let

$$\begin{aligned}
 H(v) &= \sigma_\varepsilon^2 D(v) + g_0^\tau(v) \Sigma_u g_0(v) D(v) + \sigma_\varepsilon^2 \Sigma_u + E g_0^\tau(v) U U^\tau g_0(v) U U^\tau \\
 &\quad - \Sigma_u g_0(v) g_0^\tau(v) \Sigma_u.
 \end{aligned}$$

Using Crámer-Wold devise and central limit theorem, we can show that

$$\begin{aligned}
 &\sqrt{nh} \begin{pmatrix} \tilde{\mathbf{a}} - g_0(v) - 2^{-1} h^2 \mu_2(K) D(v) \ddot{g}(v) f(v) \\ h \tilde{\mathbf{b}} - h g_0'(v) \end{pmatrix} \\
 &\implies f^{-1}(v) \begin{pmatrix} D(v) & 0 \\ 0 & \mu_2(K) D(v) + u_p(1) \end{pmatrix}^{-1} \\
 &\quad \cdot N \left(\begin{pmatrix} R(K) f(v) H(v) & 0 \\ 0 & \int x^2 K^2(x) dx f(v) H(v) \end{pmatrix} \right).
 \end{aligned}$$

Theorem 3.2 can be proved after some algebra. □

Proof of Theorem 3.3 The proof is similar to the proof of Theorem 2 in Xue and Pang (2013), so only the differences are presented here for the sake of brevity. Denote

$$\begin{aligned}
 g_0(\beta^\tau Z, \beta) &= E[g_0(\beta_0^\tau Z) | \beta^\tau Z], \quad \dot{g}_0(\beta^\tau Z, \beta) = E[\dot{g}_0(\beta_0^\tau Z) | \beta^\tau Z], \\
 T(g, \beta) &= E[(Y - g^\tau(\beta^\tau Z) W) \dot{g}^\tau(\beta^\tau Z) W + g^\tau(\beta^\tau Z) \Sigma_u \dot{g}(\beta^\tau Z)] J_{\beta(r)} Z w(\beta^\tau Z)
 \end{aligned}$$

and $T_n(g, \beta)$ is the empirical version of $T(g, \beta)$. Also, denote

$$\begin{aligned}
 \tau(g, \beta) &= E[(Y - g_0(\beta^\tau Z; \beta)^\tau W)(\dot{g}(\beta^\tau Z; \beta) - \dot{g}_0(\beta^\tau Z; \beta))^\tau W J_{\beta(r)} Z w(\beta^\tau Z)] \\
 &\quad - E[(g(\beta^\tau Z; \beta) - g_0(\beta^\tau Z; \beta))^\tau W W^\tau \dot{g}_0^\tau(\beta^\tau Z; \beta) J_{\beta(r)} Z w(\beta^\tau Z)] \\
 &\quad + E[(g(\beta^\tau Z; \beta) - g_0(\beta^\tau Z; \beta))^\tau \Sigma_U \dot{g}_0(\beta^\tau Z; \beta) J_{\beta(r)} Z w(\beta^\tau Z)] \\
 &\quad + E[g_0^\tau(\beta^\tau Z; \beta) \Sigma_U (\dot{g}(\beta^\tau Z; \beta) - \dot{g}_0(\beta^\tau Z; \beta)) J_{\beta(r)} Z w(\beta^\tau Z)].
 \end{aligned}$$

Then $T_n(\hat{g}, \beta)$ can be written as the following four terms

$$\begin{aligned} S_{n1}(\beta) &= T_n(\hat{g}, \beta) - T(\hat{g}, \beta) - T_n(g_0, \beta_0), \\ S_{n2}(\beta) &= T(\hat{g}, \beta) - T(g_0, \beta) - \tau(\hat{g}, \beta), \\ S_{n3}(\beta) &= \tau(\hat{g}, \beta) - \tau(\hat{g}, \beta_0), \quad S_{n4}(\beta_0) = T_n(g_0, \beta_0) + \tau(\hat{g}, \beta_0). \end{aligned}$$

The following lemma is needed to facilitate the proof of Theorem 3.3.

Lemma 5.2 *Suppose that the condition (C1)–(C6) hold. then*

$$\sup_{\beta \in \mathcal{B}_n} \|S_{n1}(\beta)\| = o_p(n^{-1/2}), \tag{5.3}$$

$$\sup_{\beta \in \mathcal{B}_n} \|S_{n2}(\beta)\| = o_p(n^{-1/2}), \tag{5.4}$$

$$\sup_{\beta \in \mathcal{B}_n} \|S_{n3}(\beta)\| = o_p(n^{-1/2}), \tag{5.5}$$

$$\sqrt{n}S_{n4}(\beta_0) \implies N(0, \sigma^2 A), \tag{5.6}$$

where Σ is defined in Theorem 3.3.

Proof Similar to [Xue and Pang \(2013\)](#), let $r_n(g, \beta) = \sqrt{n}(T_n(g, \beta) - T(g, \beta))$. Since $T(g_0, \beta_0) = 0$, so $S_{n1}(\beta) = n^{-1/2}[r_n(\hat{g}, \beta) - r_n(g_0, \beta_0)]$. By checking the conditions of Theorem 1 in [Doukhan et al. \(1995\)](#), one can show the stochastic equicontinuity of the empirical process $\{r_n(g, \beta) : g \in \mathcal{G}_1, \beta \in \mathcal{B}_1\}$. Therefore, $r_n(g, \beta) - r(g_0, \beta_0) = o_p(1)$ which implies (5.3).

To prove (5.4), first note that

$$\begin{aligned} S_{n2} &= -E[(\hat{g}(\beta^\tau Z; \beta) - g_0(\beta^\tau Z; \beta))^\tau (WW^\tau - \Sigma_u)(\hat{g}(\beta^\tau Z; \beta) \\ &\quad - \dot{g}_0(\beta^\tau Z; \beta))]J_{\beta^{(r)}}Zw(\beta^\tau Z). \end{aligned}$$

Therefore, from Theorem 3.1, we have

$$\sqrt{n} \sup_{\beta \in \mathcal{B}_n} \|S_{n2}(\beta)\| \leq \sqrt{n} \cdot O\left(\sqrt{\frac{\log n}{nh}} + h^2\right) \cdot O\left(\sqrt{\frac{\log n}{nh^3}} + h_1\right) = o_p(1).$$

To show (5.5), first we have

$$\begin{aligned} &\tau(\hat{g}, \beta) - \tau(\hat{g}, \beta_0) \\ &= E[(g_0(\beta_0^\tau Z; \beta_0)) - g_0(\beta^\tau Z; \beta)]^\tau W(\hat{g}(\beta^\tau Z; \beta) - \dot{g}_0(\beta^\tau Z; \beta))^\tau WJ_{\beta^{(r)}} \\ &\quad Zw(\beta^\tau Z) \\ &\quad - E[(\hat{g}(\beta^\tau Z; \beta) - g_0(\beta^\tau Z; \beta))^\tau WW^\tau \dot{g}_0^\tau(\beta^\tau Z; \beta)]J_{\beta^{(r)}}Zw(\beta^\tau Z) \\ &\quad + E[(\hat{g}(\beta^\tau Z; \beta) - g_0(\beta^\tau Z; \beta))^\tau \Sigma_U \dot{g}_0(\beta^\tau Z; \beta)]J_{\beta^{(r)}}Zw(\beta^\tau Z) \\ &\quad + E[g_0^\tau(\beta^\tau Z; \beta)\Sigma_U(\hat{g}(\beta^\tau Z; \beta) - \dot{g}_0(\beta^\tau Z; \beta))]J_{\beta^{(r)}}Zw(\beta^\tau Z) \end{aligned}$$

$$\begin{aligned}
 & - E[(\hat{g}(\beta_0^\tau Z; \beta_0) - g_0(\beta_0^\tau Z; \beta_0))^\tau W W^\tau \dot{g}_0^\tau(\beta_0^\tau Z; \beta_0) J_{\beta_0^{(r)}} J_{\beta_0^{(r)}} Z w(\beta_0^\tau Z)] \\
 & + E[(\hat{g}(\beta_0^\tau Z; \beta_0) - g_0(\beta_0^\tau Z; \beta_0))^\tau \Sigma_U \dot{g}_0^\tau(\beta_0^\tau Z; \beta_0) J_{\beta_0^{(r)}} J_{\beta_0^{(r)}} Z w(\beta_0^\tau Z)] \\
 & + E[g_0^\tau(\beta_0^\tau Z; \beta_0) \Sigma_U (\hat{g}(\beta_0^\tau Z; \beta_0) - \dot{g}_0(\beta_0^\tau Z; \beta_0)) J_{\beta_0^{(r)}} J_{\beta_0^{(r)}} Z w(\beta_0^\tau Z)].
 \end{aligned}$$

For sufficiently large n and all $\beta \in \mathcal{B}_n$, we have $\|J_{\beta^{(r)}} Z w(\beta^\tau Z) - J_{\beta_0^{(r)}} Z w(\beta_0^\tau Z)\| = O(1/\sqrt{n})$, and $\|\dot{g}_0^\tau(\beta^\tau Z; \beta) - \dot{g}_0^\tau(\beta_0^\tau Z; \beta_0)\| = O(1/\sqrt{n})$, $\|\dot{g}_0^\tau(\beta^\tau Z; \beta) - \dot{g}_0^\tau(\beta_0^\tau Z; \beta_0)\| = O(1/\sqrt{n})$ uniformly. By Theorem 3.1, we can obtain

$$\begin{aligned}
 & \tau(\hat{g}, \beta) - \tau(\hat{g}, \beta_0) \\
 & = -E[(\hat{g}(\beta^\tau Z; \beta) - g_0(\beta^\tau Z; \beta))^\tau (W W^\tau - \Sigma_u) \dot{g}_0^\tau(\beta_0^\tau Z; \beta_0) J_{\beta_0^{(r)}} J_{\beta_0^{(r)}} Z w(\beta_0^\tau Z)] \\
 & + E[g_0^\tau(\beta^\tau Z; \beta) \Sigma_U (\hat{g}(\beta^\tau Z; \beta) - \dot{g}_0(\beta^\tau Z; \beta)) J_{\beta_0^{(r)}} J_{\beta_0^{(r)}} Z w(\beta_0^\tau Z)] \\
 & + E[(\hat{g}(\beta_0^\tau Z; \beta_0) - g_0(\beta_0^\tau Z; \beta_0))^\tau (W W^\tau - \Sigma_u) \dot{g}_0^\tau(\beta_0^\tau Z; \beta_0) J_{\beta_0^{(r)}} J_{\beta_0^{(r)}} Z w(\beta_0^\tau Z)] \\
 & - E[g_0^\tau(\beta_0^\tau Z; \beta_0) \Sigma_U (\hat{g}(\beta_0^\tau Z; \beta_0) - \dot{g}_0(\beta_0^\tau Z; \beta_0)) J_{\beta_0^{(r)}} J_{\beta_0^{(r)}} Z w(\beta_0^\tau Z)].
 \end{aligned}$$

Thus, (5.5) can be derived by noting that

$$\begin{aligned}
 & (\hat{g}(\beta^\tau Z; \beta) - g_0(\beta^\tau Z; \beta)) - (\hat{g}(\beta_0^\tau Z; \beta_0) - g_0(\beta_0^\tau Z; \beta_0)) = o(1/\sqrt{n}), \\
 & (\hat{g}(\beta^\tau Z; \beta) - \dot{g}_0(\beta^\tau Z; \beta)) - (\hat{g}(\beta_0^\tau Z; \beta_0) - \dot{g}_0(\beta_0^\tau Z; \beta_0)) = o(1/\sqrt{n}).
 \end{aligned}$$

From the definition of $\tau(g, \beta)$ and the independence of ε, U and (X, Z) , we can obtain

$$\tau(\hat{g}, \beta_0) = -E(\hat{g}(\beta_0^\tau Z; \beta_0) - g_0(\beta_0^\tau Z))^\tau X X^\tau \dot{g}_0(\beta_0^\tau Z) J_{\beta_0} Z w(\beta_0^\tau Z).$$

Then by the Corollary 3.1 and the notation $C(v)$ defined in Condition (C6), together with the smooth conditions on $g_0, D(v)$ and $C(v)$ from (C2), (C6), we have

$$\begin{aligned}
 \tau(\hat{g}, \beta_0) & = -EC(\beta_0^\tau Z)(\hat{g}(\beta_0^\tau Z; \beta_0) - g_0(\beta_0^\tau Z)) = - \int C(v)[\hat{g}(v; \beta_0) - g_0(v)]dv \\
 & = -\frac{1}{n} \sum_{i=1}^n C(\beta_0^\tau Z_i) D^{-1}(\beta_0^\tau Z_i) [W_i(\varepsilon_i - U_i^\tau g_0(\beta_0^\tau Z_i)) + \Sigma_u g_0(\beta_0^\tau Z_i)] \\
 & \quad - \frac{h^2}{2} \mu_2(K) E[C(\beta_0^\tau Z) g_0''(\beta_0^\tau Z)] + c_n.
 \end{aligned}$$

Denote $\zeta_i = J_{\beta_0}^\tau Z_i \dot{g}_0^\tau(\beta_0^\tau Z_i) w(\beta_0^\tau Z_i) - C(\beta_0^\tau Z_i) D^{-1}(\beta_0^\tau Z_i)$, and $\xi_i = \zeta_i W_i$. Then, we have

$$S_{n4}(\beta_0) = \frac{1}{n} \sum_{i=1}^n [(\varepsilon_i - U_i^\tau g_0(\beta_0^\tau Z_i))\xi_i + \zeta_i \Sigma_u g_0(\beta_0^\tau Z_i)] - \frac{h^2}{2} \mu_2(K) E[C(\beta_0^\tau Z)\ddot{g}_0(\beta_0^\tau Z)] + c_n.$$

Therefore, by classical central limit theorem and $nh^4 \rightarrow 0$, we have

$$\sqrt{n}S_{n4}(\beta_0) = \frac{1}{n} \sum_{i=1}^n [(\varepsilon_i - U_i^\tau g_0(\beta_0^\tau Z_i))\xi_i + \zeta_i \Sigma_u g_0(\beta_0^\tau Z_i)] + o_p(1) \implies N(0, A),$$

where $A = \text{Var}((\varepsilon - U^\tau g_0(\beta_0^\tau Z))\xi + \zeta \Sigma_u g_0(\beta_0^\tau Z))$. □

Now, let's return to the proof of Theorem 3.3. By Taylor expansion, $T(g_0, \beta_0) = 0$, and $\beta^{(r)} - \beta_0^{(r)} = O(n^{-1/2})$, we can obtain

$$T(g_0, \beta^{(r)}) = -B(\beta^{(r)} - \beta_0^{(r)}) + o(n^{-1/2}), \tag{5.7}$$

where $B = E[-VV^\tau + VX^\tau \dot{g}_0(\beta_0^\tau Z)E(Z^\tau | \beta_0^\tau Z)J_{\beta(r)}]$. In fact, from the independence of U and X, Z , we have

$$\begin{aligned} T(g_0, \beta^{(r)}) &= E[(Y - g_0^\tau(\beta^\tau Z)W) \dot{g}_0^\tau(\beta^\tau Z)W + g_0^\tau(\beta^\tau Z)\Sigma_u \dot{g}_0(\beta^\tau Z)] J_{\beta(r)} Zw(\beta^\tau Z) \\ &= E\{[g_0^\tau(\beta_0^\tau Z)X - E[g_0^\tau(\beta_0^\tau Z)|\beta^\tau Z]W] \dot{g}_0^\tau(\beta^\tau Z)W + g_0^\tau(\beta^\tau Z)\Sigma_u \dot{g}_0(\beta^\tau Z)\} J_{\beta(r)} Zw(\beta^\tau Z) \\ &= E\{[g_0^\tau(\beta_0^\tau Z)X - g_0^\tau(\beta^\tau Z)X + g_0^\tau(\beta^\tau Z)X - E[g_0^\tau(\beta_0^\tau Z)|\beta^\tau Z](X + U)] \\ &\quad E[\dot{g}_0^\tau(\beta_0^\tau Z)|\beta^\tau Z](X + U) + E[g_0^\tau(\beta_0^\tau Z)|\beta^\tau Z]\Sigma_u E[\dot{g}_0^\tau(\beta_0^\tau Z)|\beta^\tau Z]\} J_{\beta(r)} Zw(\beta^\tau Z) \\ &= E\{[g_0^\tau(\beta_0^\tau Z)X - g_0^\tau(\beta^\tau Z)X + g_0^\tau(\beta^\tau Z)X - E[g_0^\tau(\beta_0^\tau Z)|\beta^\tau Z]X] \\ &\quad E[\dot{g}_0^\tau(\beta_0^\tau Z)|\beta^\tau Z]X\} J_{\beta(r)} Zw(\beta^\tau Z). \end{aligned}$$

Note that

$$[g_0^\tau(\beta^\tau Z) - g_0^\tau(\beta_0^\tau Z)]X = \dot{g}_0^\tau(\beta_0^\tau Z)XZ' J_{\beta(r)}(\beta^{(r)} - \beta_0^{(r)}) + o_p(n^{-1/2}),$$

this, together with the continuity of $E[\dot{g}_0^\tau(\beta_0^\tau Z)|\beta^\tau Z]$ with respect to β , we have, uniformly for $\beta^{(r)} \in \mathcal{B}_n^{(r)}$,

$$T(g_0, \beta^{(r)}) = E[-VV^\tau + VX^\tau \dot{g}_0(\beta_0^\tau Z)E(Z^\tau | \beta_0^\tau Z)J_{\beta(r)}] (\beta^{(r)} - \beta_0^{(r)}) + o_p(n^{-1/2}),$$

which is exactly (5.7). The rest of the proof is similar to that of Theorem 2 in [Xue and Pang \(2013\)](#). □

The proof of Theorem 3.4 is a consequence of the consistency of the local linear estimate of g and single index estimate of β . Details are omitted here for the sake of brevity.

Acknowledgements The authors would like to thank editors and reviewers for their constructive comments which greatly improve the presentation of the paper; Weixing Song's research is supported by the grants from NSF DMS 1205276.

References

- Cai ZW, Fan JQ, Li RZ (2000) Efficient estimation and inferences for varying coefficient models. *J Am Stat Assoc* 95:888–902
- Carroll RJ, Ruppert D, Stefanski LA, Crainiceanu CM (2012) *Measurement error in nonlinear models: a modern perspective*. CRC Press, Boca Raton
- Cheng CL, Ness JWV (1999) *Statistical regression with measurement error*. Wiley, New York
- Delecroix M, Härdle W, Hristache M (2003) Efficient estimation in conditional single-index regression. *J Multivar Anal* 86:213–226
- Doukhan P, Massart P, Rio E (1995) Invariance principles for absolutely regular empirical processes. *Ann Inst Henri Poincaré Probab Stat* 31:393–427
- Fan J, Yao Q, Cai Z (2003) Adaptive varying-coefficient linear models. *J R Stat Soc B* 65:57–80
- Fuller WA (2006) *Measurement error models*. Wiley-Interscience, New York
- Härdle W, Hall P, Ichimura H (1993) Optimal smoothing in single-index models. *Ann Stat* 21:157–178
- Harrison D, Rubinfeld DL (1978) Hedonic prices and the demand for clean air. *J Environ Econ Manag* 5:81–102
- Hastie TJ, Tibshirani R (1993) Varying-coefficient models. *J R Stat Soc B* 55:757–796
- Horowitz JL (1998) *Semiparametric methods in econometrics*. Springer, London
- Ichimura H (1993) Semiparametric least squares (SLS) and weighted SLS estimation of single-index models. *J Econom* 58:71–120
- Li KC (1989) Data visualization with SIR: a transformation based projection pursuit method. *UCLA Stat Ser* 24:1989
- Li KC (1991) Sliced inverse regression for dimension reduction (with discussion). *J Am Stat Assoc* 86:316–342
- Li X, You J, Zhou Y (2011) Statistical inference for varying coefficient models with error-prone covariates. *J Stat Comput Simul* 81:1755–1771
- Liang H, Wang N (2005) Partially linear single-index measurement error models. *Stat Sin* 15:99–116
- Stocker TM (1986) Consistent estimation of scaled coefficients. *Econometrica* 54:1461–1481
- Xia YC, Li WK (1999) On single-index coefficient regression models. *J Am Stat Assoc* 94:1275–1285
- Xue LG, Zhu LX (2006) Empirical likelihood for single-index model. *J Multivar Anal* 97:1295–1312
- Xue LG, Wang QH (2012) Empirical likelihood for single-index varyingcoefficient models. *Bernoulli* 18:836–856
- Xue L, Pang Z (2013) Statistical inference for a single-index varying coefficient model. *Stat Comput* 23:589–599
- You J, Zhou Y, Chen G (2006) Corrected local polynomial estimation in varying-coefficient models with measurement errors. *Can J Stat* 34:391–410
- Yu Y, Ruppert D (2002) Penalized spline estimation for partially linear single-index models. *J Am Stat Assoc* 97:1042–1054