Mathematics

Research article

Robust estimation for varying-coefficient partially linear measurement error model with auxiliary instrumental variables

Yanting Xiao∗**and Wanying Dong**

Department of Applied Mathematics, Xi'an University of Technology, Xi'an, Shaanxi 710048, China

* **Correspondence:** Email: xiaoyanting03@163.com.

Abstract: We study the varying-coefficient partially linear model when some linear covariates are not observed, but their auxiliary instrumental variables are available. Combining the calibrated error-prone covariates and modal regression, we present a two-stage efficient estimation procedure, which is robust against outliers or heavy-tail error distributions. Asymptotic properties of the resulting estimators are established. Performance of our proposed estimation procedure is illustrated through some numerous simulations and a real example. And the results confirm that the proposed methods are satisfactory.

1. Introduction

Varying-coefficient partially linear model is one of the most popular semiparametric regression models. It takes the form as

$$Y = X^T \Theta + Z^T \alpha(U) + \varepsilon, \tag{1.1}$$

where Y is the response variable, X, Z and U are the associated covariates, Θ is a p-dimensional vector of unknown parameters, $\alpha(.)$ is a q-dimensional vector of nonparametric function, ε is model error with $E(\varepsilon|X, Z, U) = 0$ and $\text{Cov}(\varepsilon|X, Z, U) = \sigma^2$. By combining the flexibility of nonparametric model and the adaptability of parametric model, model (1.1) is quite general which covers many important models as special cases, such as liner model, varying-coefficient model, partially linear model and others. However, due to the low accuracy of instrument equipment, the imperfection of measurement technique, and the high price or time cost, it is often impossible to observe accurate data in the experiments, but the data with errors-in-variables(EV) [1]. If one simply ignores the measurement error, known as naive method, it will result in unreliable statistical inference, such as

biased estimators, loss of efficiency, decreasing power of hypothesis testing. According to the diversity of errors in variables, the treatment of measurement errors differs for various types.

Early research focused on the study of simple additive measurement errors in the linear component. Instead of true variable X, we can observe η with $\eta = X + e$ and e is the measurement error, which is independent of (X, Z, U, ε). You and Chen [2] proposed a modified profile least squares procedure and Wang et al. [3] applied empirical likelihood inference to study model (1.1) with measurement errors in the linear part. On the other hand, measurement error in the nonparametric part in model (1.1) may occur in reality. The surrogate variable η is observed by $\eta = Z + e$ with the measurement error e. Feng and Xue [4] proposed a local bias-corrected restricted profile least squares estimators and Fan et al. [5] applied some auxiliary information to construct empirical log-likelihood ratios for both parameter and nonparametric functions. It is noted that in the above literature, we assumed that the relationship between the unobservable variable and the observed EV data is simple additive. Furthermore, it is necessary that in the statistical inference, the variance of measurement error is known before. Even if the variance of error is unavailable in practice, we can use the estimators to replace it by repeatedly measuring actually observation data η.

However, the additive measurement error structure may not describe the complexity of experimental data and sometimes there are no repeated observations to estimate the variance of measurement error. In this paper, we do not specify any model structure of the measurement error, and do not require the information of the variance of measurement error. Furthermore, repeated measurement data are not necessary. We assume that some variables are unobserved directly, but auxiliary information is available to remit the true variable. Specifically, the unobserved variable and the observed surrogate variable with measurement error is connected by a nonparametric structure with an instrument variable. Let $X = (\xi^T, W^T)^T$, where ξ is a $p_1 \times 1$ unobserved directly vector and W is a vector of the remaining exactly observed components. We assume that the true variable ξ is related to the observed measurement variable η and auxiliary variable V through the nonparametric structure

$$\xi = E(\eta|V) = \xi(V). \tag{1.2}$$

From assumption (1.2), model (1.1) with error-prone linear covariates implies that

$$\begin{cases} Y = \xi^T\beta + W^T\theta + Z^T\alpha(U) + \varepsilon \\ \eta = \xi(V) + e \end{cases} \tag{1.3}$$

with $\Theta = (\beta^T, \theta^T)^T$, $X = (\xi^T, W^T)^T$ and e is the measurement error with $E(e|V) = \mathbf{0}$ and the covariance matrix Σ_e. Therefore, the error model structures of ξ is a special cases of the additive error models. For the sake of simplicity, we assume that the auxiliary variables V is scalar. This structure of measurement error is more general, introduced by Zhou and Liang [6], which includes denoise linear model and rational expectation model and others. Zhou and Liang [6] proposed a regression calibration technique to reduce the bias due to mismeasurement and then developed the profile least-square based estimation procedure for parametric and nonparametric components.

Meanwhile, this structure of error-prone covariates has been extensively studied by researchers for other semiparametric models. For the varying-coefficient models with error-prone covariates, Zhao and Xue [7] constructed confidence interval for the varying-coefficient functions using instrumental variable-based empirical likelihood inference and Xu et al. [8] presented an instrumental variable

type local polynomial estimator. Zhang et al. [9] proposed the estimation and variable selection for partial linear single index model when some linear covariates are not observed, but their ancillary variables are available. Huang and Ding [10] studied a new partially linear error-in-variable models with error-prone covariates in the parametric part and mismeasured covariates in the nonparametric part, simultaneously. Sun et al. [11] developed an estimation procedure for semiparametric varying-coefficient partially linear model when both response and part of covariates are measured with error by functions of some auxiliary variables.

In this paper, we propose two step estimation procedure for model (1.3). In the first step, we give a bias correction for the error-prone covariate by using ancillary information and applying locally linear technique. Then the covariates are corrected based on instrumental variables. In the second stage, instead of profile least squares estimation, we propose a robust estimation procedure. It is remarkable that the least squares technique may be sensitive to outliers and inefficient for many non-normal error distributions. Modal regression, introduced by Yao et al. [12], is a powerful tool for revealing the underlying relationship between the response and corresponding covariates in the presence of outlier or non-normal error distribution, and it has been successfully applied to various semiparametric models (Zhao et al. [13], Yang and Yang [14] and Yang et al. [15], Lv et al. [16], Yu et al. [17] and others). This technique has some good properties, such as easy to implement, robust to outliers, full asymptotic efficiency under the non-normal error distribution. Motivated by these, we extend modal regression in the second stage to estimate the parameter and nonparametric function.

The rest of this paper is organized as follows. Section 2 introduces a two-stage robust instrumental variables-based modal regression estimator and Section 3 establishes its theoretical properties. In section 4, we discuss the selection of bandwidth and the specific estimation algorithm. Simulation studies are conducted to evaluate the performances of the proposed estimation procedure in Section 5. A real data analysis is illustrated in section 6 to show the effective of the proposed estimation procedures. We make our concluding remarks in Section 7 and leave the proofs of the main Theorems in Section 8.

2. Robust estimation procedure

In this section, we propose a two-stage robust estimation procedure for model (1.3). Firstly, the local linear smoothing is adapted to calibrate the variable ξ by using auxiliary instrument variables η and V. Secondly, based on B-spline basis functions and modal regression, the final estimators of the parametric and nonparametric components are obtained.

2.1. Covariate calibration

Suppose that $\{(Y_i, \xi_i, W_i, Z_i, U_i, \eta_i, V_i)\}_{i=1}^n$ is the independent and identically distributed sample from $(Y, \xi, W, Z, U, \eta, V)$. The varying-coefficient partially linear model with error-prone in linear covariates has the following form,

$$\begin{cases} Y_i = \xi_i^T \beta + W_i^T \theta + Z_i^T \alpha(U_i) + \varepsilon_i \\ \eta_i = \xi(V_i) + e_i \end{cases}. \tag{2.1}$$

Firstly, we need to calibrate the covariate ξ_i, which is not observed in model (2.1). Let η_{ik} be the kth entry of vector η_i, $i = 1, \ldots, n$. To estimate $\xi_k(v)$, the kth component of $\xi(v)$, we employ the local

linear smoothing technique [18] to minimize

$$\sum_{i=1}^{n} \{\eta_{ik} - a_{0k} - a_{1k}(V_i - v)\}^2 L_{b_k}(V_i - v), \tag{2.2}$$

where $L_b(.) = L(./b)/b$ is the kernel function, and $b = b_k$ for $k = 1, \ldots, p_1$ is the bandwidth for the kth component of $\boldsymbol{\xi}$. The local linear estimator of $\xi_k(v), k = 1, \ldots, p_1$, denoted by $\hat{\xi}_k(v)$,

$$\hat{\xi}_k(v) = \hat{a}_{0k} = (1, 0)(V_v^T W_v V_v)^{-1} V_v^T W_v \boldsymbol{\eta}^{(k)}, \tag{2.3}$$

where $\boldsymbol{\eta}^{(k)} = (\eta_{1k}, \eta_{2k}, \ldots, \eta_{nk})^T$, $W_v = \text{diag}(L_{b_k}(V_1 - v), L_{b_k}(V_2 - v), \ldots, L_{b_k}(V_n - v))$, and V_v is defined in the following,

$$V_v = \begin{pmatrix} 1 & V_1 - v \\ 1 & V_2 - v \\ \vdots & \vdots \\ 1 & V_n - v \end{pmatrix}.$$

Therefore, unobserved covariate $\boldsymbol{\xi}_i$ can be replaced by its local linear estimator $\hat{\boldsymbol{\xi}}(V_i)$, abbreviated as $\hat{\boldsymbol{\xi}}_i$. Then the robust estimators of parameters and nonparametric function can be constructed in the following procedure.

2.2. Modal regression estimator

Following Schumaker [19], we can replace the nonparametric functions $\alpha_k(u), k = 1, 2, \ldots, q$ through their basis function approximations. Let $\boldsymbol{B}(u) = (B_1(u), \ldots, B_L(u))^T$ be the B-spline basis function with the order of M, where $L = K + M$ and K is the number of interior knots. Then, the nonparametric function $\alpha_k(u)$ can be approximated by

$$\alpha_k(u) \approx \boldsymbol{B}(u)^T \boldsymbol{\gamma}_k, k = 1, 2, \ldots, q, \tag{2.4}$$

where $\boldsymbol{\gamma}_k = (\gamma_{k1}, \gamma_{k2}, \ldots, \gamma_{kL})^T$. Then model (2.4) is substituted into the model (2.1), we can obtain

$$Y_i \approx \boldsymbol{\xi}_i^T \boldsymbol{\beta} + W_i^T \boldsymbol{\theta} + \boldsymbol{\Pi}_i^T \boldsymbol{\gamma} + \varepsilon_i, i = 1, \ldots, n, \tag{2.5}$$

where $\boldsymbol{\Pi}_i = I_q \otimes \boldsymbol{B}(U_i) \cdot \boldsymbol{Z}_i$ and $\boldsymbol{\gamma} = (\gamma_1^T, \gamma_2^T, \ldots, \gamma_q^T)^T$. Model (2.5) is a standard linear regression model, and it contains three parametric vectors $\boldsymbol{\beta}$, $\boldsymbol{\theta}$ and $\boldsymbol{\gamma}$. Each function $\alpha_k(.)$ in model (2.1) is characterized by $\boldsymbol{\gamma}_k$ in model (2.4).

Based on the thought of modal regression [12], the estimators of the parameters $\boldsymbol{\beta}$, $\boldsymbol{\theta}$ and $\boldsymbol{\gamma}$ are obtained through maximizing the following objective function,

$$L(\boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{\gamma}) = \sum_{i=1}^{n} \phi_h(Y_i - \boldsymbol{\xi}_i^T \boldsymbol{\beta} - W_i^T \boldsymbol{\theta} - \boldsymbol{\Pi}_i^T \boldsymbol{\gamma}), \tag{2.6}$$

where $\phi_h(.) = \phi(./h)/h$, $\phi(.)$ is the kernel function and h is a bandwidth. The value of bandwidth h determines the level of the robustness of the estimator, then the specific selection of optimal bandwidth h is described in the next section.

However, we can not maximize (2.6) directly due to the unobserved variable ξ_i. Instead, we substitute the local linear estimator $\hat{\xi}_i$ with the true variable ξ_i. Further, the calibrated expression is rewrite as

$$\hat{L}(\boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{\gamma}) = \sum_{i=1}^{n} \phi_h(Y_i - \hat{\xi}_i^T \boldsymbol{\beta} - W_i^T \boldsymbol{\theta} - \boldsymbol{\Pi}_i^T \boldsymbol{\gamma}). \tag{2.7}$$

Therefore, we get the spline modal regression estimators $\hat{\boldsymbol{\beta}}$, $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\gamma}}$ by maximizing the objective function (2.7). Then, the corresponding estimator of coefficient function $\alpha_k(u)$ is obtained by $\hat{\alpha}_k(u) = \boldsymbol{B}(u)^T \hat{\boldsymbol{\gamma}}_k$, $k = 1, 2, \ldots, q$.

3. Theoretical properties

In this section, we discuss the asymptotic properties of the resulting estimators. Denote $\boldsymbol{\Theta}_0 = (\boldsymbol{\beta}_0^T, \boldsymbol{\theta}_0^T)^T$ and $\alpha_0(.)$ to be the true value of $\boldsymbol{\Theta} = (\boldsymbol{\beta}^T, \boldsymbol{\theta}^T)^T$ and $\alpha(.)$ in model (1.1). Correspondingly, $\boldsymbol{\gamma}_{0j}$ is the best approximation coefficient of $\alpha_{0j}(u)$ in the B-spline space. Let $F(\boldsymbol{x}, \boldsymbol{z}, u, h) = \mathrm{E}(\phi_h''(\varepsilon)|\boldsymbol{X} = \boldsymbol{x}, \boldsymbol{Z} = \boldsymbol{z}, U = u)$ and $G(\boldsymbol{x}, \boldsymbol{z}, u, h) = \mathrm{E}(\phi_h'(\varepsilon)^2|\boldsymbol{X} = \boldsymbol{x}, \boldsymbol{Z} = \boldsymbol{z}, U = u)$.

We begin with the following assumption conditions required to derive the main results. These conditions are quite mild and can be easily satisfied.

A1: The random variable U has a bounded support $\mathcal{U}$. Its density function $f_u(.)$ is Lipschitz continuous and bounded away from 0 and infinite on its support. The density function of random function V, $f_v(.)$ is continuously differentiable and bounded away from 0 and infinite on its support $\mathcal{V}$.

A2: The nonparametric function $\alpha_k(u), k = 1, \ldots, q$ and $\xi_k(v), k = 1, \ldots, p_1$ have a continuous second derivative.

A3: The matrix $E(XX^T|U)$ is nonsingular for each $U \in \mathcal{U}$. All elements of the matrices $E(XX^T|U)$, $E(XX^T|U)^{-1}$ and $E(XZ^T|U)^{-1}$ are Lipschitz continuous.

A4: The kernel functions $L(.)$ and $\phi(.)$ are density functions with compact supports.

A5: There exists an $s > 0$ such that $\mathrm{E}\|\mathbf{X}\|^{2s} < \infty$ and for some $r < 2 - s^{-1}$ such that $n^{2r-1}h \to \infty$, $n^{2r-1}b_k \to \infty$ and $nhb_k^4 \to 0$ as $n \to \infty$, $k = 1, 2, \ldots, p_1$.

A6: $nh^8 \to 0$ and $nh^2/(\log n)^2 \to \infty$.

A7: Let $c_1, \ldots, c_K$ be the interior knots on $[0, 1]$. Set $c_0 = 0$, $c_{K+1} = 1$ and $h_i = c_i - c_{i-1}$. Then, there exists a constant C_0 such that

$$\frac{\max\{h_i\}}{\min\{h_i\}} \le C_0, \qquad \max\{|h_{i+1} - h_i|\} = o(K^{-1}).$$

A8: $F(\boldsymbol{x}, \boldsymbol{z}, u, h)$ and $G(\boldsymbol{x}, \boldsymbol{z}, u, h)$ are continuous with respect to $(\boldsymbol{x}, \boldsymbol{z}, u)$. In addition, $F(\boldsymbol{x}, \boldsymbol{z}, u, h) < 0$ for any $h > 0$.

A9: $\mathrm{E}(\phi_h'(\varepsilon)|\boldsymbol{x}, \boldsymbol{z}, u) = 0$ and $\mathrm{E}(\phi_h''(\varepsilon)^2|\boldsymbol{x}, \boldsymbol{z}, u)$, $\mathrm{E}(\phi_h'(\varepsilon)^3|\boldsymbol{x}, \boldsymbol{z}, u)$ and $\mathrm{E}(\phi_h'''(\varepsilon)|\boldsymbol{x}, \boldsymbol{z}, u)$ are continuous with respect to $\boldsymbol{x}$, $\boldsymbol{z}$ and u.

The following Theorem 1 gives the consistency of spline modal estimators $\hat{\boldsymbol{\Theta}} = (\hat{\boldsymbol{\beta}}^T, \hat{\boldsymbol{\theta}}^T)^T$ and $\hat{\alpha}_k(u), k = 1, \ldots, q$.

Theorem 1. *Suppose that the assumptions A1-A9 hold, and the number of knots $K = O(n^{1/(2r+1)})$. Then we have*

$$(i)\|\hat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}_0\| = O_p(n^{-\frac{r}{2r+1}}),$$

$$(ii) \|\hat{\alpha}_k(u) - \alpha_{0k}(u)\| = O_p(n^{-\frac{r}{2r+1}}), k = 1, \ldots, q.$$

The following Theorem 2 shows the asymptotic normality for the parameter estimator $\hat{\mathbf{\Theta}}$.

Let $\Phi(U) = \mathrm{E}(F(X, Z, U, h)\mathbf{\Pi X}^T|U)$, and $\Gamma(U) = \mathrm{E}(F(X, Z, U, h)\mathbf{\Pi\Pi}^T|U)$. Denote $\tilde{\mathbf{X}}(\mathbf{Z}, \mathbf{X}, U) = \mathbf{X} - \Phi(U)^T\Gamma^{-1}(U)\mathbf{\Pi}$.

Theorem 2. *Suppose that the assumptions A1-A9 hold, and the number of knots $K = O(n^{1/(2r+1)})$, then we have*

$$\sqrt{n}(\hat{\mathbf{\Theta}} - \mathbf{\Theta}_0) \xrightarrow{d} N(\mathbf{0}, \Sigma^{-1}\Omega\Sigma^{-1}),$$

where $\Sigma = \mathrm{E}(F(X, Z, U, h)\tilde{\mathbf{X}}\tilde{\mathbf{X}}^T)$ and $\Omega = \mathrm{E}(G(X, Z, U, h)\tilde{\mathbf{X}}\tilde{\mathbf{X}}^T) + \boldsymbol{\beta}_0^T\Sigma_e\boldsymbol{\beta}_0\mathrm{E}(F(X, Z, U, h)^2\mathrm{E}(\tilde{X}\tilde{X}^T)).$

4. Bandwidth selection and estimation algorithm

In this section, we discuss the selection of bandwidth and estimation procedure based on the modal expectation maximization(MEM) algorithm proposed by Li et al. [20].

4.1. Bandwidth selection in practice

In order to obtain the spline modal estimators $\hat{\boldsymbol{\beta}}$, $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\gamma}}$ by maximizing function (2.7), we need to select an appropriate bandwidth h due to the fact that its value affects directly the robustness of the estimators. According to the thought of Yao et al. [12], it is easy to show that the ratio of the asymptotic variance of our proposed estimator to that of the least square B-spline estimator is given by

$$R(h) = \frac{\hat{G}(h)}{\hat{F}^2(h)\hat{\sigma}^2}, \tag{4.1}$$

where $\hat{F}(h) = \frac{1}{n}\sum_{i=1}^n \phi_h''(\hat{\varepsilon}_i)$, $\hat{G}(h) = \frac{1}{n}\sum_{i=1}^n \phi_h'(\hat{\varepsilon}_i)^2$, $\hat{\varepsilon}_i = Y_i - \hat{\boldsymbol{\xi}}_i^T\boldsymbol{\beta}^{(0)} - \boldsymbol{W}_i^T\boldsymbol{\theta}^{(0)} - \boldsymbol{\Pi}_i^T\boldsymbol{\gamma}^{(0)}$, the parameters $\boldsymbol{\beta}^{(0)}$, $\boldsymbol{\theta}^{(0)}$ and $\boldsymbol{\gamma}^{(0)}$ are the initial estimators of the parametric vectors $\boldsymbol{\beta}$, $\boldsymbol{\theta}$ and $\boldsymbol{\gamma}$ via the least square estimation procedure, and $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n \hat{\varepsilon}_i^2$.

The optimal bandwidth h is calculated by minimizing (4.1). In the actual calculation, we use the grid search method to choose h, and the possible grids points for h can be $h = 0.5\hat{\sigma} \times 1.02^j$, $j = 0, 1, \ldots, 100$ (See Yao et al. [12] and Lv et al. [16]).

4.2. The MEM algorithm for parameters

In this subsection, the following modal expectation maximization(MEM) algorithm by Li et al. [20] is adopted to obtain the estimators of the parameters $\mathbf{\Theta}$ and $\boldsymbol{\gamma}$. Let $\boldsymbol{Q}_i = (\hat{\boldsymbol{\xi}}_i^T, \boldsymbol{W}_i^T, \boldsymbol{\Pi}_i^T)^T$, $\Xi = (\mathbf{\Theta}^T, \boldsymbol{\gamma}^T)^T$. Set $\Xi^{(0)} = (\mathbf{\Theta}^{(0)T}, \boldsymbol{\gamma}^{(0)T})^T$, where $\mathbf{\Theta}^{(0)}$ and $\boldsymbol{\gamma}^{(0)}$ are initial estimators of $\mathbf{\Theta}$ and $\boldsymbol{\gamma}$, respectively with $m = 0$.

Step1.$(\mathbf{E} - \mathbf{step})$: We update $\pi(i|\Xi^{(m)})$ by

$$\pi(i|\Xi^{(m)}) = \frac{\phi_h(Y_i - \boldsymbol{Q}_i^T\Xi^{(m)})}{\sum_{i=1}^n \phi_h(Y_i - \boldsymbol{Q}_i^T\Xi^{(m)})}, i = 1, 2, \ldots, n.$$

Step2.$(\mathbf{M} - \mathbf{step})$: Update $\Xi^{(m+1)}$ by

$$\Xi^{(m+1)} = \arg\max_{\Xi} \sum_{i=1}^n (\pi(i|\Xi^{(m)}) \log \phi_h(Y_i - \boldsymbol{Q}_i^T\Xi))$$

$$= (\mathbf{Q}^T \omega \mathbf{Q})^{-1} \mathbf{Q}^T \omega \mathbf{Y}$$

where $\mathbf{Q} = (\mathbf{Q}_1, \mathbf{Q}_2, \ldots, \mathbf{Q}_n)^T$, $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_n)^T$, ω is $n \times n$ diagonal matrix with $\omega = diag(\pi(1|\mathbf{\Xi}^{(m)}), \pi(2|\mathbf{\Xi}^{(m)}), \ldots, \pi(n|\mathbf{\Xi}^{(m)}))$.

Step3: Iterate the E-step and M-step until the algorithm converges. We denote the final estimator of $\mathbf{\Xi}$ as $\hat{\mathbf{\Xi}}$. Then, the parameters for $\mathbf{\Theta}$ and $\boldsymbol{\gamma}$ based on modal regression are $\hat{\mathbf{\Theta}} = (\mathbf{I}_{p \times p}, \mathbf{0}_{p \times qL}) \hat{\mathbf{\Xi}}$ and $\hat{\boldsymbol{\gamma}} = (\mathbf{0}_{qL \times p}, \mathbf{I}_{qL \times qL}) \hat{\mathbf{\Xi}}$. Then, the corresponding estimator of $\alpha_k(.)$ is $\hat{\alpha}_k(u) = \boldsymbol{B}(u)^T \hat{\boldsymbol{\gamma}}_k, k = 1, 2, \ldots, q$.

5. Simulation studies

In this section, we conduct the following simulation to assess the finite sample performance of the proposed estimation procedure. For this purpose, we consider the following varying-coefficient partially linear model with error-prone linear covariates,

$$\begin{cases} Y_i = \boldsymbol{\xi}_i^T \boldsymbol{\beta} + \boldsymbol{W}_i^T \boldsymbol{\theta} + \boldsymbol{Z}_i^T \boldsymbol{\alpha}(U_i) + \varepsilon_i \\ \boldsymbol{\eta}_i = \boldsymbol{\xi}(V_i) + \boldsymbol{e}_i \end{cases} \tag{5.1}$$

where the parametric vector $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3)^T = (3, 1.8, 0.8)^T$, $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)^T = (2, 0.5, -1.5)^T$, the nonparametric function $\boldsymbol{\alpha}(U) = (\alpha_1(U), \alpha_2(U))^T$ with $\alpha_1(U) = 3.5(\exp(-(4U-1)^2) + \exp(-(4U-3)^2)) - 1.5$, $\alpha_2(U) = \cos(2\pi U)$ and $U_i \sim U(0, 1)$, the covariates $\boldsymbol{W}_i$ are multivariate normal distribution with mean $\mathbf{0}$ and covariance matrix $\Sigma_1 = (\sigma_{ij})_{1 \le i, j \le 3}$ with $\sigma_{ij} = 0.5^{|i-j|}$ for $i, j = 1, 2, 3$ and the covariates $\boldsymbol{Z}_i$ are multivariate normal distribution with mean $\mathbf{0}$ and covariance matrix $\Sigma_2 = (\sigma_{ij})_{1 \le i, j \le 2}$ with $\sigma_{ij} = 0.5^{|i-j|}$ for $i, j = 1, 2$. The following three different model errors: (1) the normal distribution: $\varepsilon_i \sim N(0, 1)$; (2) the t distribution: $\varepsilon_i \sim t(3)$; (3) the mixed normal distribution(MN): $\varepsilon_i \sim 0.9N(0, 1) + 0.1N(0, 9^2)$ are considered.

The variables $\boldsymbol{\xi}_i$ in model (5.1) are the function of the auxiliary instrumental variables V_i. It is noted that $\boldsymbol{\xi}_i$ are not observed, and the observed variables are $\boldsymbol{\eta}_i$ with $\boldsymbol{\eta}_i = \boldsymbol{\xi}(V_i) + \boldsymbol{e}_i$, and $\xi_1(V_i) = V_i(1 - V_i)$, $\xi_2(V_i) = 2\sin(V_i)$, $\xi_3(V_i) = 3V_i - 2\cos(V_i)$ with $V_i \sim N(2, 1.5)$. Furthermore, the measurement error $\boldsymbol{e}_i$ is independent with V_i and $(\boldsymbol{W}_i, \boldsymbol{Z}_i, \varepsilon_i)$ and $\boldsymbol{e}_i \sim N(\mathbf{0}, I_3 \sigma_e^2)$, where $\sigma_e^2 = 0.4^2$ and $\sigma_e^2 = 0.6^2$ represent the different level of the measurement error.

In the simulation experiments, we choose the Epanechnikov kernel function $L(t) = \frac{3}{4}(1 - t^2)_+$, the bandwidth $b = \hat{\sigma}_v n^{-1/3}$ is adopted in local linear smoothing according to Zhou and Liang [6], where $\hat{\sigma}_v$ is sample root variance of the variable V_i. We use the standard normal density for $\phi(t)$ in the modal regression and the bandwidth h is selected through minimizing (4.1). We use the cubic B-spline basis function in our simulation. The sample size n is 100, 250 and 500, and run simulation is 200.

To evaluate the performance of the proposed estimation procedure in the parametric component, we compared the modal regression (MR) estimators to the profile least squares (PLS) estimators proposed by Zhou and Liang [6]. Three types of estimators include the naive estimators in which $\boldsymbol{\xi}_i$ is simply replaced by $\boldsymbol{\eta}_i$ (denoted by MR-N and PLS-N), our proposed calibrated estimators (denoted by MR-C and PLS-C), and the benchmark estimators which assumes that $\boldsymbol{\xi}_i$ is known before (denoted by MR-B and PLS-B). Tables 1–6 show the bias and standard deviation (denoted by SD in the brackets) of the estimators of parameters $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$.

Table 1. The bias and SD of estimators for parameters with $\varepsilon_i \sim N(0, 1)$ and $\sigma_e^2 = 0.4^2$.

n	Method	β_1	β_2	β_3	θ_1	θ_2	θ_3
100	PLS-N	0.0968(0.1254)	0.4061(0.2673)	0.0720(0.0932)	0.2575(0.3178)	0.2848(0.3505)	0.2375(0.3044)
	PLS-C	0.0873(0.1124)	0.1989(0.2506)	0.0676(0.0874)	0.1366(0.1692)	0.1514(0.1897)	0.1310(0.1672)
	PLS-B	0.0401(0.0535)	0.0950(0.1176)	0.0323(0.0420)	0.0978(0.1220)	0.1099(0.1383)	0.1007(0.1258)
	MR-N	0.0981(0.1278)	0.4060(0.2694)	0.0732(0.0949)	0.2566(0.3197)	0.2898(0.3554)	0.2386(0.3100)
	MR-C	0.0873(0.1120)	0.1993(0.2508)	0.0678(0.0877)	0.1343(0.1673)	0.1511(0.1887)	0.1278(0.1646)
	MR-B	0.0403(0.0538)	0.0959(0.1200)	0.0326(0.0423)	0.0978(0.1223)	0.1091(0.1374)	0.1012(0.1274)
250	PLS-N	0.0555(0.0697)	0.4147(0.1537)	0.0434(0.0531)	0.1382(0.1738)	0.1578(0.1991)	0.1513(0.1844)
	PLS-C	0.0491(0.0619)	0.1223(0.1494)	0.0415(0.0508)	0.0704(0.0895)	0.0806(0.1009)	0.0672(0.0846)
	PLS-B	0.0201(0.0247)	0.0482(0.0630)	0.0161(0.0197)	0.0605(0.0764)	0.0706(0.0863)	0.0578(0.0728)
	MR-N	0.0555(0.0700)	0.4146(0.1553)	0.0430(0.0532)	0.1389(0.1751)	0.1592(0.1995)	0.1522(0.1846)
	MR-C	0.0509(0.0635)	0.1219(0.1496)	0.0426(0.0522)	0.0687(0.0879)	0.0787(0.0985)	0.0651(0.0827)
	MR-B	0.0198(0.0244)	0.0481(0.0629)	0.0160(0.0196)	0.0606(0.0768)	0.0703(0.0860)	0.0577(0.0726)
500	PLS-N	0.0336(0.0414)	0.4169(0.1076)	0.0294(0.0362)	0.0907(0.1138)	0.0992(0.1327)	0.0863(0.1102)
	PLS-C	0.0331(0.0407)	0.0794(0.0979)	0.0285(0.0360)	0.0430(0.0548)	0.0522(0.0657)	0.0474(0.0602)
	PLS-B	0.0140(0.0185)	0.0354(0.0446)	0.0121(0.0156)	0.0381(0.0479)	0.0423(0.0529)	0.0413(0.0501)
	MR-N	0.0356(0.0439)	0.4160(0.1084)	0.0284(0.0358)	0.0917(0.1141)	0.0990(0.1321)	0.0866(0.1100)
	MR-C	0.0342(0.0430)	0.0801(0.1002)	0.0288(0.0365)	0.0423(0.0538)	0.0503(0.0636)	0.0468(0.0595)
	MR-B	0.0140(0.0186)	0.0355(0.0448)	0.0120(0.0156)	0.0380(0.0482)	0.0422(0.0525)	0.0409(0.0499)

Table 2. The bias and SD of estimators for parameters with $\varepsilon_i \sim t(3)$ and $\sigma_e^2 = 0.4^2$.

n	Method	β_1	β_2	β_3	θ_1	θ_2	θ_3
100	PLS-N	0.1293(0.1696)	0.4702(0.3694)	0.0974(0.1255)	0.2974(0.3728)	0.3108(0.3959)	0.3037(0.3752)
	PLS-C	0.1235(0.1647)	0.2756(0.3634)	0.0933(0.1222)	0.2064(0.2619)	0.2196(0.2780)	0.1866(0.2446)
	PLS-B	0.0695(0.0971)	0.1596(0.2031)	0.0523(0.0724)	0.1894(0.2359)	0.1943(0.2477)	0.1636(0.2198)
	MR-N	0.1243(0.1601)	0.4601(0.3593)	0.0952(0.1203)	0.2863(0.3609)	0.3063(0.3919)	0.2907(0.3620)
	MR-C	0.1165(0.1539)	0.2559(0.3449)	0.0866(0.1140)	0.1701(0.2162)	0.1994(0.2512)	0.1656(0.2151)
	MR-B	0.0575(0.0788)	0.1289(0.1646)	0.0435(0.0583)	0.1407(0.1771)	0.1611(0.2071)	0.1387(0.1799)
250	PLS-N	0.0597(0.0799)	0.4552(0.1749)	0.0492(0.0631)	0.1725(0.2117)	0.1855(0.2373)	0.1416(0.1792)
	PLS-C	0.0587(0.0774)	0.1463(0.1822)	0.0460(0.0598)	0.1183(0.1503)	0.1274(0.1622)	0.0930(0.1172)
	PLS-B	0.0330(0.0438)	0.0900(0.1133)	0.0275(0.0371)	0.1107(0.1417)	0.1150(0.1464)	0.0906(0.1147)
	MR-N	0.0590(0.0801)	0.4515(0.1737)	0.0465(0.0607)	0.1680(0.2030)	0.1803(0.2276)	0.1401(0.1754)
	MR-C	0.0569(0.0771)	0.1392(0.1740)	0.0445(0.0580)	0.0978(0.1222)	0.1064(0.1323)	0.0786(0.0967)
	MR-B	0.0271(0.0363)	0.0720(0.0891)	0.0219(0.0289)	0.0873(0.1085)	0.0936(0.1136)	0.0763(0.0932)
500	PLS-N	0.0387(0.0502)	0.4226(0.1295)	0.0314(0.0403)	0.1189(0.1520)	0.1288(0.1676)	0.1196(0.1506)
	PLS-C	0.0391(0.0501)	0.0990(0.1209)	0.0319(0.0410)	0.0766(0.0984)	0.0791(0.1034)	0.0732(0.0926)
	PLS-B	0.0247(0.0330)	0.0620(0.0783)	0.0211(0.0271)	0.0725(0.0923)	0.0795(0.1029)	0.0707(0.0887)
	MR-N	0.0372(0.0482)	0.4217(0.1235)	0.0311(0.0392)	0.1149(0.1455)	0.1270(0.1609)	0.1140(0.1440)
	MR-C	0.0369(0.0482)	0.0895(0.1131)	0.0307(0.0391)	0.0576(0.0723)	0.0602(0.0726)	0.0563(0.0695)
	MR-B	0.0183(0.0237)	0.0447(0.0562)	0.0153(0.0195)	0.0523(0.0642)	0.0575(0.0708)	0.0549(0.0658)

Table 3. The bias and SD of estimators for parameters with $\varepsilon_i \sim 0.9N(0,1) + 0.1N(0,9^2)$ and $\sigma_e^2 = 0.4^2$.

n	Method	β_1	β_2	β_3	θ_1	θ_2	θ_3
100	PLS-N	0.0892(0.1162)	0.4152(0.2756)	0.0736(0.0929)	0.2603(0.3168)	0.2910(0.3690)	0.2545(0.3084)
	PLS-C	0.0880(0.1111)	0.2164(0.2648)	0.0723(0.0895)	0.1606(0.1992)	0.1714(0.2129)	0.1411(0.1759)
	PLS-B	0.0481(0.0607)	0.1223(0.1553)	0.0387(0.0494)	0.1322(0.1674)	0.1460(0.1791)	0.1230(0.1534)
	MR-N	0.0912(0.1196)	0.4180(0.2790)	0.0750(0.0952)	0.2612(0.3159)	0.2906(0.3706)	0.2549(0.3085)
	MR-C	0.0894(0.1142)	0.2174(0.2669)	0.0740(0.0918)	0.1583(0.1966)	0.1695(0.2093)	0.1397(0.1728)
	MR-B	0.0485(0.0619)	0.1261(0.1587)	0.0391(0.0502)	0.1318(0.1681)	0.1427(0.1771)	0.1229(0.1539)
250	PLS-N	0.0530(0.0678)	0.4115(0.1638)	0.0453(0.0561)	0.1524(0.1866)	0.1593(0.1966)	0.1501(0.1845)
	PLS-C	0.0494(0.0639)	0.1367(0.1542)	0.0413(0.0513)	0.0823(0.1056)	0.1001(0.1214)	0.0858(0.1080)
	PLS-B	0.0257(0.0332)	0.0633(0.0806)	0.0213(0.0262)	0.0719(0.0914)	0.0870(0.1089)	0.0726(0.0919)
	MR-N	0.0534(0.0683)	0.4126(0.1653)	0.0458(0.0567)	0.1511(0.1863)	0.1628(0.1999)	0.1506(0.1855)
	MR-C	0.0509(0.0650)	0.1245(0.1533)	0.0419(0.0519)	0.0802(0.1028)	0.0972(0.1198)	0.0835(0.1059)
	MR-B	0.0256(0.0331)	0.0628(0.0796)	0.0210(0.0259)	0.0720(0.0914)	0.0861(0.1087)	0.0732(0.0928)
500	PLS-N	0.0356(0.0447)	0.4292(0.1118)	0.0294(0.0365)	0.1121(0.1379)	0.1210(0.1494)	0.1116(0.1416)
	PLS-C	0.0358(0.0472)	0.0938(0.1208)	0.0292(0.0375)	0.0641(0.0794)	0.0728(0.0882)	0.0573(0.0706)
	PLS-B	0.0183(0.0232)	0.0467(0.0587)	0.0147(0.0183)	0.0566(0.0713)	0.0651(0.0799)	0.0570(0.0700)
	MR-N	0.0351(0.0444)	0.4283(0.1116)	0.0291(0.0363)	0.1114(0.1369)	0.1200(0.1488)	0.1118(0.1412)
	MR-C	0.0361(0.0474)	0.0937(0.1208)	0.0294(0.0377)	0.0639(0.0790)	0.0722(0.0877)	0.0574(0.0710)
	MR-B	0.0184(0.0232)	0.0466(0.0584)	0.0147(0.0183)	0.0572(0.0717)	0.0649(0.0798)	0.0566(0.0697)

Table 4. The bias and SD of estimators for parameters with $\varepsilon_i \sim N(0,1)$ and $\sigma_e^2 = 0.6^2$.

n	Method	β_1	β_2	β_3	θ_1	θ_2	θ_3
100	PLS-N	0.1217(0.1674)	0.8211(0.3693)	0.1017(0.1307)	0.3291(0.4148)	0.3761(0.4737)	0.3621(0.4432)
	PLS-C	0.1211(0.1618)	0.3084(0.3572)	0.0985(0.1255)	0.1615(0.2025)	0.2043(0.2483)	0.1738(0.2214)
	PLS-B	0.0404(0.0527)	0.1000(0.1244)	0.0308(0.0400)	0.1117(0.1383)	0.1234(0.1538)	0.1090(0.1354)
	MR-N	0.1215(0.1687)	0.8229(0.3756)	0.1014(0.1314)	0.3309(0.4147)	0.3833(0.4790)	0.3675(0.4441)
	MR-C	0.1204(0.1626)	0.3057(0.3604)	0.0993(0.1265)	0.1579(0.2001)	0.1960(0.2375)	0.1669(0.2099)
	MR-B	0.0400(0.0522)	0.1002(0.1250)	0.0305(0.0395)	0.1144(0.1410)	0.1236(0.1539)	0.1093(0.1359)
250	PLS-N	0.0683(0.0862)	0.8608(0.2111)	0.0558(0.0702)	0.2053(0.2566)	0.2269(0.2799)	0.2065(0.2601)
	PLS-C	0.0632(0.0757)	0.1773(0.1828)	0.0512(0.0625)	0.0893(0.1089)	0.0997(0.1238)	0.0932(0.1136)
	PLS-B	0.0223(0.0288)	0.0572(0.0730)	0.0185(0.0234)	0.0619(0.0771)	0.0686(0.0902)	0.0675(0.0833)
	MR-N	0.0729(0.0918)	0.8556(0.2181)	0.0592(0.0742)	0.2065(0.2597)	0.2274(0.2810)	0.2081(0.2587)
	MR-C	0.0633(0.0755)	0.1805(0.1825)	0.0504(0.0621)	0.0862(0.1060)	0.0937(0.1191)	0.0890(0.1100)
	MR-B	0.0221(0.0287)	0.0579(0.0733)	0.0185(0.0233)	0.0619(0.0766)	0.0688(0.0901)	0.0684(0.0841)
500	PLS-N	0.0498(0.0621)	0.8526(0.1587)	0.0425(0.0524)	0.1469(0.1885)	0.1590(0.1984)	0.1416(0.1750)
	PLS-C	0.0441(0.0556)	0.1301(0.1288)	0.0369(0.0462)	0.0539(0.0687)	0.0596(0.0743)	0.0522(0.0651)
	PLS-B	0.0134(0.0173)	0.0341(0.0443)	0.0108(0.0140)	0.0425(0.0544)	0.0432(0.0547)	0.0396(0.0496)
	MR-N	0.0521(0.0660)	0.8482(0.1621)	0.0445(0.0545)	0.1461(0.1892)	0.1608(0.2009)	0.1427(0.1766)
	MR-C	0.0444(0.0559)	0.1312(0.1293)	0.0373(0.0463)	0.0520(0.0659)	0.0553(0.0704)	0.0504(0.0616)
	MR-B	0.0136(0.0175)	0.0342(0.0446)	0.0110(0.0142)	0.0425(0.0547)	0.0428(0.0545)	0.0393(0.0496)

Table 5. The bias and SD of estimators for parameters with $\varepsilon_i \sim t(3)$ and $\sigma_e^2 = 0.6^2$.

n	Method	β_1	β_2	β_3	θ_1	θ_2	θ_3
100	PLS-N	0.1384(0.1850)	0.9086(0.4104)	0.1090(0.1430)	0.3517(0.4390)	0.3971(0.5007)	0.3568(0.4438)
	PLS-C	0.1410(0.1865)	0.3495(0.4005)	0.1003(0.1405)	0.2166(0.2795)	0.2469(0.3142)	0.1964(0.2492)
	PLS-B	0.0626(0.0852)	0.1594(0.2041)	0.0497(0.0647)	0.1716(0.2279)	0.1822(0.2269)	0.1483(0.1881)
	MR-N	0.1383(0.1879)	0.9015(0.4100)	0.1087(0.1442)	0.3507(0.4361)	0.3932(0.4986)	0.3649(0.4525)
	MR-C	0.1364(0.1803)	0.3319(0.4000)	0.1002(0.1401)	0.1924(0.2482)	0.2105(0.2742)	0.1820(0.2244)
	MR-B	0.0517(0.0678)	0.1249(0.1612)	0.0403(0.0515)	0.1432(0.1872)	0.1517(0.1866)	0.1207(0.1545)
250	PLS-N	0.0758(0.0989)	0.8488(0.2265)	0.0576(0.0752)	0.2146(0.2580)	0.2090(0.2628)	0.2111(0.2694)
	PLS-C	0.0715(0.0970)	0.1893(0.2166)	0.0558(0.0737)	0.1046(0.1289)	0.1225(0.1543)	0.1192(0.1493)
	PLS-B	0.0345(0.0457)	0.0867(0.1080)	0.0294(0.0378)	0.0922(0.1112)	0.1137(0.1436)	0.1101(0.1387)
	MR-N	0.0794(0.1009)	0.8474(0.2279)	0.0587(0.0746)	0.2121(0.2554)	0.2071(0.2591)	0.2104(0.2671)
	MR-C	0.0706(0.0945)	0.1882(0.2129)	0.0544(0.0712)	0.0896(0.1150)	0.1037(0.1317)	0.0949(0.1224)
	MR-B	0.0282(0.0369)	0.0663(0.0859)	0.0240(0.0304)	0.0714(0.0871)	0.0893(0.1107)	0.0835(0.1045)
500	PLS-N	0.0576(0.0726)	0.8684(0.1735)	0.0498(0.0608)	0.1636(0.2044)	0.1627(0.2072)	0.1606(0.1995)
	PLS-C	0.0528(0.0659)	0.1501(0.1497)	0.0445(0.0543)	0.0827(0.0993)	0.0904(0.1149)	0.0814(0.1020)
	PLS-B	0.0201(0.0269)	0.0561(0.0714)	0.0173(0.0227)	0.0715(0.0860)	0.0795(0.1010)	0.0706(0.0904)
	MR-N	0.0569(0.0734)	0.8616(0.1718)	0.0486(0.0605)	0.1616(0.2011)	0.1583(0.2025)	0.1570(0.1957)
	MR-C	0.0521(0.0649)	0.1430(0.1466)	0.0434(0.0531)	0.0638(0.0790)	0.0734(0.0948)	0.0610(0.0770)
	MR-B	0.0157(0.0202)	0.0412(0.0534)	0.0134(0.0170)	0.0518(0.0642)	0.0588(0.0760)	0.0509(0.0653)

Table 6. The bias and SD of estimators for parameters with $\varepsilon_i \sim 0.9N(0, 1) + 0.1N(0, 9^2)$ and $\sigma_e^2 = 0.6^2$.

n	Method	β_1	β_2	β_3	θ_1	θ_2	θ_3
100	PLS-N	0.1450(0.1896)	0.8040(0.3938)	0.1147(0.1464)	0.3471(0.4449)	0.4040(0.4977)	0.3230(0.4182)
	PLS-C	0.1380(0.1807)	0.3125(0.3985)	0.1115(0.1423)	0.1928(0.2454)	0.2038(0.2620)	0.1786(0.2295)
	PLS-B	0.0482(0.0635)	0.1215(0.1497)	0.0396(0.0500)	0.1493(0.1915)	0.1373(0.1706)	0.1261(0.1580)
	MR-N	0.1446(0.1895)	0.8032(0.3971)	0.1144(0.1463)	0.3529(0.4570)	0.4026(0.5025)	0.3301(0.4231)
	MR-C	0.1367(0.1818)	0.3082(0.3990)	0.1114(0.1423)	0.1912(0.2449)	0.1997(0.2536)	0.1748(0.2230)
	MR-B	0.0485(0.0636)	0.1210(0.1503)	0.0397(0.0502)	0.1497(0.1925)	0.1373(0.1701)	0.1278(0.1588)
250	PLS-N	0.0748(0.0947)	0.8816(0.2180)	0.0599(0.0768)	0.1931(0.2383)	0.2098(0.2724)	0.2029(0.2583)
	PLS-C	0.0705(0.0927)	0.1883(0.2257)	0.0583(0.0760)	0.1005(0.1260)	0.1055(0.1341)	0.0990(0.1251)
	PLS-B	0.0259(0.0320)	0.0698(0.0886)	0.0204(0.0253)	0.0825(0.1023)	0.0815(0.1028)	0.0753(0.0938)
	MR-N	0.0745(0.0948)	0.8796(0.2183)	0.0595(0.0763)	0.1928(0.2384)	0.2123(0.2748)	0.2066(0.2606)
	MR-C	0.0739(0.0953)	0.1846(0.2228)	0.0509(0.0747)	0.0965(0.1215)	0.1013(0.1277)	0.0941(0.1200)
	MR-B	0.0259(0.0319)	0.0692(0.0883)	0.0205(0.0253)	0.0829(0.1025)	0.0808(0.1019)	0.0759(0.0945)
500	PLS-N	0.0508(0.0625)	0.8576(0.1426)	0.0385(0.0479)	0.1423(0.1796)	0.1671(0.2068)	0.1374(0.1751)
	PLS-C	0.0444(0.0570)	0.1320(0.1548)	0.0378(0.0475)	0.0650(0.0818)	0.0730(0.0891)	0.0581(0.0748)
	PLS-B	0.0181(0.0231)	0.0462(0.0577)	0.0152(0.0191)	0.0555(0.0684)	0.0616(0.0778)	0.0605(0.0768)
	MR-N	0.0507(0.0625)	0.8534(0.1428)	0.0381(0.0477)	0.1433(0.1809)	0.1662(0.2057)	0.1376(0.1747)
	MR-C	0.0422(0.0561)	0.1316(0.1522)	0.0400(0.0493)	0.0648(0.0814)	0.0705(0.0868)	0.0570(0.0743)
	MR-B	0.0181(0.0231)	0.0434(0.0540)	0.0145(0.0182)	0.0470(0.0600)	0.0582(0.0733)	0.0554(0.0684)

According to Tables 1–6, we draw the following conclusions: (1) The estimators of parameters are obtained under six estimation approaches. Their bias and SD become smaller with the increasing of the sample size in all situations. This indicates that all estimation procedures are reasonable. (2) The bias and SD of the PLS estimator are smaller than MR estimator while the model error yielding to normal distribution. However, when the model error is the t or MN distribution, the proposed MR estimator gives smaller bias and SD than the PLS estimator. It illustrates that the PLS estimator is sensitive to the non normal distributed model error. (3) Compared the estimators among MR-N, MR-C and MR-B, we find that our proposed MR-C estimator is superior to MR-N estimator and it performs just a little bit worse than MR-B. And the PLS-C estimator has the similar performance to the MR-C estimator, which demonstrates that it is necessary to calibrate measurement errors in the calculation and we should not ignore the effect of the measurement errors. (4) Given the sample size, with the increase of the measurement error, the SD of PLS-C and MR-C become smaller. This reveals that the calibration of the error-prone covaritates is effective.

Furthermore, to evaluate the estimation results on the nonparametric functions, we present the fitting curve of $\alpha_1(u)$ and $\alpha_2(u)$. Figures 1 and 2 show the plot of estimated curve with the dotted, the dashed and the solid lines which denote the PLS-C estimator, MR-C estimator and the real function, respectively. We only give the results of the measurement error $\sigma_e^2 = 0.4^2$ and $\sigma_e^2 = 0.6^2$ under $n = 250$ with model error $\varepsilon_i \sim t(3)$. From Figures 1 and 2, we can see that two estimation procedures of PLS-C and MR-C estimator are all closed to the real curve, and the MR-C is much closer to the real curve. This indicates that our proposed estimation procedure is more robust when the data follows non normal distributions.
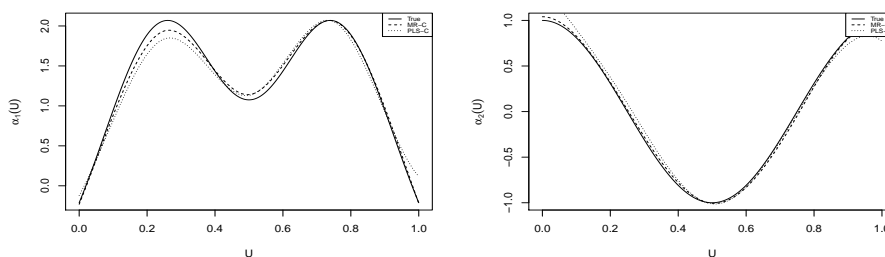

Figure 1. The estimated curves of the nonparametric functions with $t(3)$ and $\sigma_e^2 = 0.4^2$.
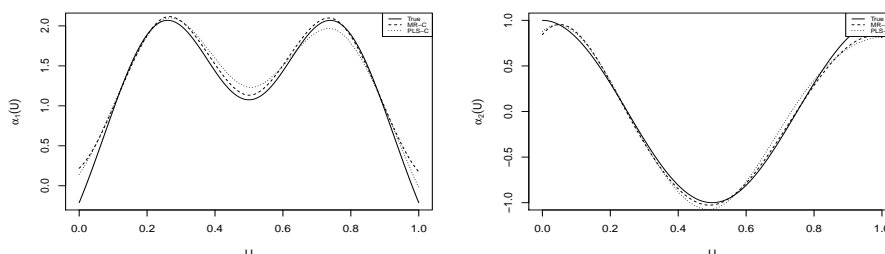

Figure 2. The estimated curves of the nonparametric functions with $t(3)$ and $\sigma_e^2 = 0.6^2$.

In addition, we apply the root of mean square error(RMSE) to illustrate the effectiveness of the

nonparametric estimators, and the form of the RMSE is

$$RMSE = \sqrt{\frac{1}{M} \sum_{j=1}^{2} \sum_{i=1}^{M} [\hat{\alpha}_j(u_i) - \alpha_j(u_i)]^2},$$

where $M = 200$, u_i is the equidistant points on $(0, 1)$. The boxplots of the RMSE for the nonparametric functions are given in Figures 3 and 4 by PLS-C and MR-C estimators when the measurement error is $\sigma_e^2 = 0.4^2$ and $\sigma_e^2 = 0.6^2$, respectively, and the sample size $n = 250$ and $n = 500$ under the model error $\varepsilon_i \sim t(3)$. From Figures 3 and 4, the MR-C estimator performs smaller RMSE than PLS-C estimator. It further confirms that our proposed estimation procedure is robust when the model error is non normal. Meanwhile, the RMSEs of two procedures are decreasing as the sample size increases.
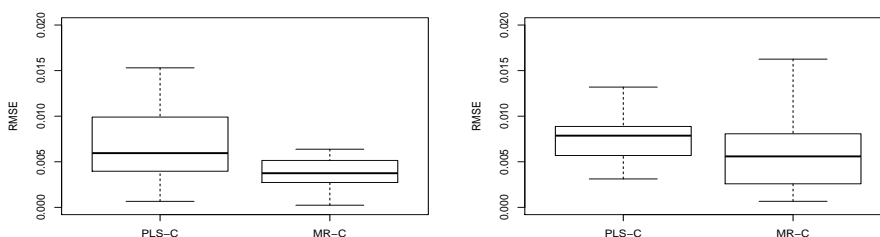
Figure 3. The boxplots of the RMSE for the nonparametric functions with $\sigma_e^2 = 0.4^2$(left) and $\sigma_e^2 = 0.6^2$(right) under $n = 250$.
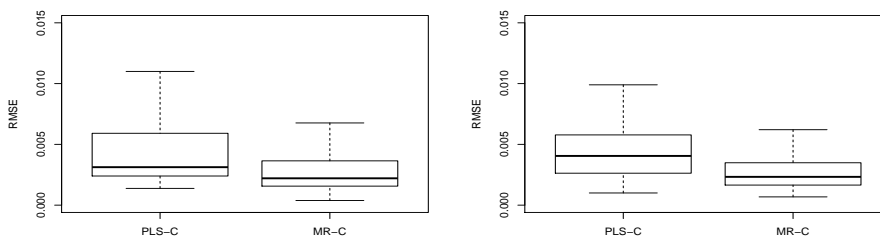

Figure 4. The boxplots of the RMSE for the nonparametric functions with $\sigma_e^2 = 0.4^2$(left) and $\sigma_e^2 = 0.6^2$(right) under $n = 500$.

6. A real example

We adopt a diabetes data to our proposed estimation procedure in this section. The data set contains 10 baseline variables, including age, sex, body mass index, average blood pressure, six blood serum measurements obtained from 442 observations, and the response of interest, a quantitative measurement of disease progression 1 year after baseline. A detailed description of each variable can be seen by Efron et al. [21]. Similar to Sun et al. [11], the following varying-coefficient partially linear measurement error model is given as

$$Y = \boldsymbol{\xi}^T \boldsymbol{\beta} + \mathbf{W}^T \boldsymbol{\theta} + \mathbf{Z}^T \boldsymbol{\alpha}(U) + \varepsilon. \tag{6.1}$$

Low-density level (ldl) and high-density level(hdl) are the covariates ξ which are measured with error. The exactly observed covariates $\mathbf{W}$ are triglycerides(tc), total cholesterol(tch) and tension glaucoma level(ltg). The covariates $\mathbf{Z}$ are the glucose concentration(glu) and average blood pressure(bp) and the variable U is age. Further, the variable body mass index(bmi) is considered as the auxiliary instrumental variable V.

We presented two methods with PLS-C and IMR-C to estimate parameter and nonparametric functions in model (6.1). The parametric estimating results are given in Table 7. From Table 7, the calibrated methods based on auxiliary instrumental variables are available in practice and the parametric estimators with PLS-C and MR-C have similar effects on response variable. In addition, the fitting curves of the coefficient functions are given in Figure 5, where the solid line and the dashed line describe the PLS-C method and the MR-C method, respectively. From Figure 5, we can see that the fitting trend of the nonparametric functions with PLS-C is close to MR-C. In a word, our proposed estimation procedures are effective in both parametric and nonparametric aspects.

Table 7. The estimators of the parameters $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ in real example.

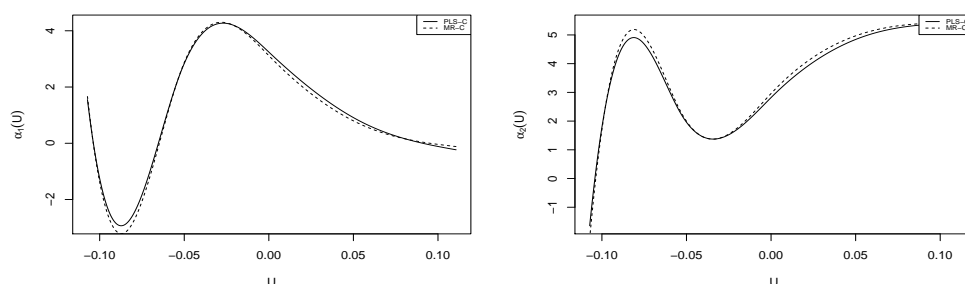|	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_3$
PLS-C	0.0513	-12.4701	-3.4992	2.3424	8.0382
MR-C	-0.0196	-12.4499	-3.5284	2.2832	8.1563

Figure 5. The fitting curves of the nonparametric functions for $\alpha_1(.)$ and $\alpha_2(.)$.

7. Conclusions

We have presented a robust estimation procedure for varying-coefficient partially linear model when some covariates in parametric part are error-prone. We assume that some linear covariates in the model can not be observed directly, and do not specify the model structure of measurement error. Firstly, we calibrate the covariates with the auxiliary instrumental variable by the local linear smoothing technique. Then with the B-spline basis function and modal regression, the components of parameters and coefficient functions are simultaneously estimated. The proposed estimation procedure not only can attenuate the effect of the measurement errors, but also is robust against outliers or heavy-tail distributed data. Under some mild conditions, the consistency and the asymptotic normality of the resulting estimators are established. We also give the strategy to choose optimal bandwidth and specific algorithm in practice. Some simulation studies and a real data analysis are carried out to demonstrate

the proposed estimation methods are satisfactory.

Acknowledgments

Conflict of interest

There are no conflicts of interest as declared by the authors.

References

1. W. A. Fuller, *Measurement error models*, New York: Wiley, 1987.

2. J. H. You, G. M. Chen, Estimation of a semiparametric varying-coefficient partially linear errors-in-variables model, *J. Multivariate Anal.*, **97** (2006), 324–341. https://doi.org/10.1016/j.jmva.2005.03.002

3. X. L. Wang, G. R. Li, L. Lin, Empirical likelihood inference for semi-parametric varying-coefficient partially linear EV models, *Metrika*, **73** (2011), 171–185. https://doi.org/10.1007/s00184-009-0271-2

4. S. Y. Feng, L. G. Xue, Bias-corrected statistical inference for partially linear varying coefficient errors-in-variables models with restricted condition, *Ann. Inst. Stat. Math.*, **66** (2014), 121–140. https://doi.org/10.1007/s10463-013-0407-z

5. G. L. Fan, H. X. Xu, Z. S. Huang, Empirical likelihood for semivarying coefficient model with measurement error in the nonparametric part, *ASTA Adv. Stat. Anal.*, **100** (2016), 21–41. https://doi.org/10.1007/s10182-015-0247-7

6. Y. Zhou, H. Liang, Statistical inference for semiparametric varying-coefficient partially linear models with error-prone linear covariates, *Ann. Statist.*, **37** (2009), 427–458. https://doi.org/10.1214/07-AOS561

7. P. X. Zhao, L. G. Xue, Instrumental variable-based empirical likelihood inferences for varying-coefficient models with error-prone covariates, *J. Appl. Stat.*, **40** (2013), 380–396. https://doi.org/10.1080/02664763.2012.744810

8. Y. Q. Xu, X. L. Li, G. M. Chen, Estimation and inference for varying-coefficient regression models with error-prone covariates, *J. Syst. Sci. Complex.*, **27** (2014), 1263–1285. https://doi.org/10.1007/s11424-014-3014-z

9. J. Zhang, X. G. Wang, Y. Yu, Y. J. Gai, Estimation and variable selection in partial linear single index models with error-prone linear covariates, *Statistics*, **48** (2014), 1048–1070. https://doi.org/10.1080/02331888.2013.800519

10. Z. S. Huang, H. Y. Ding, Statistical estimation for partially linear error-in-variable models with error-prone covariates, *Commun. Stat.-Simul. Comput.*, **46** (2017), 6559–6573. https://doi.org/10.1080/03610918.2016.1208233

11. Z. H. Sun, Y. F. Jiang, X. Ye, Improved statistical inference on semiparametric varying-coefficient partially linear measurement error model, *J. Nonparametr. Stat.*, **31** (2019), 549–566. https://doi.org/10.1080/10485252.2019.1603383

12. W. X. Yao, B. G. Lindsay, R. Z. Li, Local modal regression, *J. Nonparametr. Stat.*, **24** (2012), 647–663. https://doi.org/10.1080/10485252.2012.678848

13. W. H. Zhao, R. Q. Zhang, J. C. Liu, Y. Z. Lv, Robust and efficient variable selection for semiparametric partially linear varying coefficient model based on modal regression, *Ann. Inst. Stat. Math.*, **66** (2014), 165–191. https://doi.org/10.1007/s10463-013-0410-4

14. H. Yang, J. Yang, A robust and efficient estimation and variable selection method for partially linear single-index models, *J. Multivariate Anal.*, **129** (2014), 227–242. https://doi.org/10.1016/j.jmva.2014.04.024

15. H. Yang, J. Lv, C. H. Guo, Robust estimation and variable selection for varying-coefficient single-index models based on modal regression, *Commun. Stat.-Theor. Meth.*, **45** (2016), 4048–4067. https://doi.org/10.1080/03610926.2014.915043

16. J. Lv, H. Yang, C. H. Guo, Variable selection in partially linear additive models for modal regression, *Commun. Stat.-Simul. Comput.*, **46** (2017), 5646–5665. https://doi.org/10.1080/03610918.2016.1171346

17. P. Yu, Z. Y. Zhu, J. H. Shi, X. K. Ai, Robust estimation for partial functional linear regression model based on modal regression, *J. Syst. Sci. Complex.*, **33** (2022), 527–544. https://doi.org/10.1007/s11424-020-8217-x

18. J. Fan, R. Gijbels, *Local polynomial modelling and its applications*, London: Chapman & Hall, 1996.

19. L. L. Schumaker, *Spline function*, New York: Wiley, 1981.

20. J. Li, S. Ray, B. G. Lindsay, A nonparametric statistical approach to clustering via mode identification, *J. Mach. Learn. Res.*, **8** (2007), 1687–1723. https://doi.org/10.1007/s10846-007-9145-x

21. B. Efron, T. Hastie, I. Johnstone, R. Tibshirani, Least angle regression, *Ann. Statist.*, **32** (2004), 407–499. https://doi.org/10.1214/009053604000000067

Appendix: Proofs of theorems

Lemma 1. *Under the Assumptions of the theoretical properties in Section 3, we have*

$$\hat{\xi}_k(v) - \xi_k(v) = \frac{\mu_{L_2}}{2}\xi_k^{(2)}(v)b_k^2 + \frac{1}{nf_V(v)}\sum_{i=1}^{n}L_{b_k}(V_i - v)e_{ik} + o(b_k^2 + \frac{\log b_k^{-1}}{\sqrt{nb_k}})$$

uniformly on $v \in V$, where V is the compact support of v and e_{ik} is the kth entry of e_i. $\mu_{L_2} = \int v^2 L(v)dv$ and $f_V(v)$ is the density function of V.

Proof. This result is directly followed by Zhou and Liang [6]. □

Proof of Theorem 1. Let $\delta = n^{-r/(2r+1)}$. Define $\boldsymbol{\Theta} = \boldsymbol{\Theta}_0 + \delta \mathbf{v}_1$ and $\boldsymbol{\gamma} = \boldsymbol{\gamma}_0 + \delta \mathbf{v}_2$, and $\mathbf{v} = (\mathbf{v}_1^T, \mathbf{v}_2^T)^T$, where $\boldsymbol{\Theta}_0 = (\boldsymbol{\beta}_0^T, \boldsymbol{\theta}_0^T)^T$ and $\boldsymbol{\gamma}_0$ are the true values of $\boldsymbol{\Theta}$ and $\boldsymbol{\gamma}$, $\mathbf{v}_1$ and $\mathbf{v}_2$ are vector with p and qL dimension, respectively.

We show that, for any given $\varepsilon > 0$, there exists a large enough constant C such that

$$P\{\sup_{\|\mathbf{v}\|=C} \hat{L}(\boldsymbol{\Theta}, \boldsymbol{\gamma}) < \hat{L}(\boldsymbol{\Theta}_0, \boldsymbol{\gamma}_0)\} \le 1 - \varepsilon, \tag{7.1}$$

where $\hat{L}(\boldsymbol{\Theta}, \boldsymbol{\gamma})$ is defined in (2.7).

Let $R_k(u) = \alpha_k(u) - \boldsymbol{B}(u)^T \boldsymbol{\gamma}_{0k}$, $\boldsymbol{R}(u) = (R_1(u), R_2(u), \ldots, R_q(u))^T$. Adopting Taylor expansion to (2.7), we have

$$
\begin{aligned}
\hat{L}(\boldsymbol{\Theta}, \boldsymbol{\gamma}) - \hat{L}(\boldsymbol{\Theta}_0, \boldsymbol{\gamma}_0) &= \sum_{i=1}^{n} -\delta \phi_h'(\varepsilon_i + \boldsymbol{Z}_i^T \boldsymbol{R}(U_i) + (\boldsymbol{\xi}_i - \hat{\boldsymbol{\xi}}_i)^T \boldsymbol{\beta}_0)(\hat{\mathbf{X}}_i^T \mathbf{v}_1 + \boldsymbol{\Pi}_i^T \mathbf{v}_2) \\
&\quad + \frac{1}{2} \sum_{i=1}^{n} \delta^2 \phi_h''(\varepsilon_i + \boldsymbol{Z}_i^T \boldsymbol{R}(U_i) + (\boldsymbol{\xi}_i - \hat{\boldsymbol{\xi}}_i)^T \boldsymbol{\beta}_0)(\hat{\mathbf{X}}_i^T \mathbf{v}_1 + \boldsymbol{\Pi}_i^T \mathbf{v}_2)^2 \\
&\quad - \frac{1}{6} \sum_{i=1}^{n} \delta^3 \phi_h'''(\zeta_i)(\hat{\mathbf{X}}_i^T \mathbf{v}_1 + \boldsymbol{\Pi}_i^T \mathbf{v}_2)^3 \\
&= J_1 + J_2 + J_3 \tag{7.2}
\end{aligned}
$$

where $\hat{\mathbf{X}}_i = (\hat{\boldsymbol{\xi}}_i^T, \boldsymbol{W}_i^T)^T$, ζ_i lies in $\varepsilon_i + \boldsymbol{Z}_i^T \boldsymbol{R}(U_i) + (\boldsymbol{\xi}_i - \hat{\boldsymbol{\xi}}_i)^T \boldsymbol{\beta}_0$ and $\varepsilon_i + \boldsymbol{Z}_i^T \boldsymbol{R}(U_i) + (\boldsymbol{\xi}_i - \hat{\boldsymbol{\xi}}_i)^T \boldsymbol{\beta}_0 - \delta(\hat{\mathbf{X}}_i^T \mathbf{v}_1 + \boldsymbol{\Pi}_i^T \mathbf{v}_2)$.

Using Taylor expanding J_1 around ε_i, we have

$$
\begin{aligned}
J_1 &= \delta \sum_{i=1}^{n} \{\phi_h'(\varepsilon_i) + \phi_h''(\varepsilon_i)\boldsymbol{Z}_i^T \boldsymbol{R}(U_i) + \phi_h''(\varepsilon_i)(\boldsymbol{\xi}_i - \hat{\boldsymbol{\xi}}_i)^T \boldsymbol{\beta}_0 \\
&\quad + \phi_h'''(\varepsilon_i^*)[\boldsymbol{Z}_i^T \boldsymbol{R}(U_i) + (\boldsymbol{\xi}_i - \hat{\boldsymbol{\xi}}_i)^T \boldsymbol{\beta}_0]^2\}(\hat{\mathbf{X}}_i^T \mathbf{v}_1 + \boldsymbol{\Pi}_i^T \mathbf{v}_2) \tag{7.3}
\end{aligned}
$$

where ε_i^* is between ε_i and $\varepsilon_i + \boldsymbol{Z}_i^T \boldsymbol{R}(U_i) + (\boldsymbol{\xi}_i - \hat{\boldsymbol{\xi}}_i)^T \boldsymbol{\beta}_0$.

By Corollary 6.21 in Schumaker [19], we have $\|\boldsymbol{R}(U_i)\| = O_p(K^{-r}) = O_p(\delta)$. In addition, $\hat{\boldsymbol{\xi}}_i$ is the local linear estimators of $\boldsymbol{\xi}_i$, invoking condition A5 and A9 and the simple calculation yields

$$J_1 = O_p(n\delta K^{-r})\|\mathbf{v}\| = O_p(n\delta^2\|\mathbf{v}\|). \tag{7.4}$$

Similarly, we also have

$$J_2 = E(F(\mathbf{X}, \mathbf{Z}, U, h))O_p(n\delta^2\|\mathbf{v}\|^2). \tag{7.5}$$

Therefore, by choosing a sufficient large C, J_2 dominates J_1 uniformly $\|\mathbf{v}\| = C$. Similarly, we can prove that $J_3 = O_p(n\delta^3\|\mathbf{v}\|^3)$. Since $\delta \to 0$, it follows that $\delta\|\mathbf{v}\| \to 0$ with $\|\mathbf{v}\| = C$, which lead to $J_3 = o_p(J_2)$. Therefor, J_3 is also dominated by J_2 in $\|\mathbf{v}\| = C$.

Hence, Eq (7.1) holds and there exists local maximizers $\hat{\boldsymbol{\Theta}}$ and $\hat{\boldsymbol{\gamma}}$ such that $\|\hat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}_0\| = O_p(\delta)$ and $\|\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0\| = O_p(\delta)$. This completes the proof of part (i) in Theorem 1.

$$\|\hat{\alpha}_k(u) - \alpha_{0k}(u)\|^2 = \int_0^1 (\boldsymbol{B}(u)^T \hat{\boldsymbol{\gamma}}_k - \alpha_k(u))^2 du$$

$$= \int_0^1 [\boldsymbol{B}(u)^T \hat{\boldsymbol{\gamma}}_k - \boldsymbol{B}(u)^T \boldsymbol{\gamma}_{0k} - R_k(u)]^2 du$$

$$\leq 2 \int_0^1 [\boldsymbol{B}(u)^T (\hat{\boldsymbol{\gamma}}_k - \boldsymbol{\gamma}_{0k})]^2 du + 2 \int_0^1 R_k(u)^2 du$$

$$= 2(\hat{\boldsymbol{\gamma}}_k - \boldsymbol{\gamma}_{0k})^T \mathbf{H}(\hat{\boldsymbol{\gamma}}_k - \boldsymbol{\gamma}_{0k}) + 2 \int_0^1 R_k(u)^2 du \qquad (7.6)$$

where $\mathbf{H} = \int_0^1 \boldsymbol{B}(u)\boldsymbol{B}(u)^T du$.

Then invoking $\|\mathbf{H}\| = O(1)$ and $\|\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0\| = O_p(\delta)$, we have $(\hat{\boldsymbol{\gamma}}_k - \boldsymbol{\gamma}_{0k})^T \mathbf{H}(\hat{\boldsymbol{\gamma}}_k - \boldsymbol{\gamma}_{0k}) = O_p(\delta^2)$. Moreover, we can get that $\int_0^1 R_k(u)^2 du = O_p(\delta^2)$. Then $\|\hat{\alpha}_k(u) - \alpha_{0k}(u)\| = O_p(\delta)$ is obtained. This completes the Theorem 1. □

Proof of Theorem 2. Since $(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\gamma}})$ is the solution of maximizing the objective function (2.7), we have

$$-\frac{\partial \hat{L}(\boldsymbol{\Theta}, \boldsymbol{\gamma})}{\partial \boldsymbol{\Theta}}|_{\boldsymbol{\Theta} = \hat{\boldsymbol{\Theta}}, \boldsymbol{\gamma} = \hat{\boldsymbol{\gamma}}} = \sum_{i=1}^n \mathbf{X}_i \phi_h'(Y_i - \hat{\boldsymbol{\xi}}_i^T \hat{\boldsymbol{\beta}} - \mathbf{W}_i^T \hat{\boldsymbol{\theta}} - \boldsymbol{\Pi}_i^T \hat{\boldsymbol{\gamma}}) = 0$$

$$(7.7)$$

and

$$-\frac{\partial \hat{L}(\boldsymbol{\Theta}, \boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}}|_{\boldsymbol{\Theta} = \hat{\boldsymbol{\Theta}}, \boldsymbol{\gamma} = \hat{\boldsymbol{\gamma}}} = \sum_{i=1}^n \boldsymbol{\Pi}_i \phi_h'(Y_i - \hat{\boldsymbol{\xi}}_i^T \hat{\boldsymbol{\beta}} - \mathbf{W}_i^T \hat{\boldsymbol{\theta}} - \boldsymbol{\Pi}_i^T \hat{\boldsymbol{\gamma}}) = 0.$$

$$(7.8)$$

Note that

$$Y_i - \hat{\boldsymbol{\xi}}_i^T \hat{\boldsymbol{\beta}} - \mathbf{W}_i^T \hat{\boldsymbol{\theta}} - \boldsymbol{\Pi}_i^T \hat{\boldsymbol{\gamma}}$$
$$= \varepsilon_i + \boldsymbol{\xi}_i^T (\boldsymbol{\beta}_0 - \hat{\boldsymbol{\beta}}) + \mathbf{W}_i^T (\boldsymbol{\theta}_0 - \hat{\boldsymbol{\theta}}) + \boldsymbol{\Pi}_i^T (\boldsymbol{\gamma}_0 - \hat{\boldsymbol{\gamma}})$$
$$+ (\boldsymbol{\xi}_i - \hat{\boldsymbol{\xi}}_i)^T \boldsymbol{\beta}_0 + (\hat{\boldsymbol{\xi}}_i - \boldsymbol{\xi}_i)^T (\boldsymbol{\beta}_0 - \hat{\boldsymbol{\beta}}) + \mathbf{Z}_i^T \boldsymbol{R}(U_i).$$

$$(7.9)$$

Thus, from (7.9) and Taylor expression, we have

$$\phi_h'(Y_i - \hat{\boldsymbol{\xi}}_i^T \hat{\boldsymbol{\beta}} - \mathbf{W}_i^T \hat{\boldsymbol{\theta}} - \boldsymbol{\Pi}_i^T \hat{\boldsymbol{\gamma}})$$
$$= \phi_h'(\varepsilon_i) + \phi_h''(\varepsilon_i)\boldsymbol{\xi}_i^T(\boldsymbol{\beta}_0 - \hat{\boldsymbol{\beta}}) + \phi_h''(\varepsilon_i)\mathbf{W}_i^T(\boldsymbol{\theta}_0 - \hat{\boldsymbol{\theta}}) + \phi_h''(\varepsilon_i)\boldsymbol{\Pi}_i^T(\boldsymbol{\gamma}_0 - \hat{\boldsymbol{\gamma}})$$
$$+ \phi_h''(\varepsilon_i)(\boldsymbol{\xi}_i - \hat{\boldsymbol{\xi}}_i)^T \boldsymbol{\beta}_0 + \phi_h''(\varepsilon_i)\mathbf{Z}_i^T \boldsymbol{R}(U_i) + o_p(1). \qquad (7.10)$$

Thus, from Lemma A.1 and (7.10), we can obtain that (7.8) has the following asymptotic expression,

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)\frac{1}{n}\sum_{i=1}^n \phi_h''(\varepsilon_i)\boldsymbol{\xi}_i^T \boldsymbol{\Pi}_i + \sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)\frac{1}{n}\sum_{i=1}^n \phi_h''(\varepsilon_i)\mathbf{W}_i^T \boldsymbol{\Pi}_i$$

$$+ \sqrt{n}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0)\frac{1}{n}\sum_{i=1}^n \phi_h''(\varepsilon_i)\boldsymbol{\Pi}_i^T \boldsymbol{\Pi}_i$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \phi_h{}'(\varepsilon_i)\mathbf{\Pi}_i + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \phi_h{}''(\varepsilon_i)\mathbf{Z}_i^T \mathbf{R}(U_i)\mathbf{\Pi}_i + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \boldsymbol{\beta}_0^T b^2 \mu_{L_2}/2\phi_h{}''(\varepsilon_i)\boldsymbol{\xi}^{(2)}(V_i)\mathbf{\Pi}_i$$

$$+ \quad \frac{1}{n} \sum_{i=1}^{n} \phi_h{}''(\varepsilon_i)\mathbf{\Pi}_i \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \boldsymbol{\beta}_0^T f_V^{-1}(V_i)L_b(V_j - V_i)\boldsymbol{e}_j + o_p(1). \tag{7.11}$$

By the condition A2 and A6, and some calculations based on (7.11), it follows that

$$\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0 = [\Gamma_n + o_p(1)]^{-1}(\Lambda_n - \Phi_n(\hat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}_0)), \tag{7.12}$$

where

$$\Gamma_n = \frac{1}{n} \sum_{i=1}^{n} \phi_h''(\varepsilon_i)\mathbf{\Pi}_i\mathbf{\Pi}_i^T, \quad \Phi_n = \frac{1}{n} \sum_{i=1}^{n} \phi_h''(\varepsilon_i)\mathbf{\Pi}_i\mathbf{X}_i^T,$$

$$\Lambda_n = \frac{1}{n} \sum_{i=1}^{n} \mathbf{\Pi}_i\{\phi_h'(\varepsilon_i) + \phi_h''(\varepsilon_i)[\mathbf{Z}_i^T \mathbf{R}(U_i) + \boldsymbol{\beta}_0^T b^2 \mu_{L_2}/2\boldsymbol{\xi}^{(2)}(V_i) + \sum_{j=1}^{n} \boldsymbol{\beta}_0^T f_V^{-1}(V_i)L_b(V_j - V_i)\boldsymbol{e}_j]\}.$$

Then, by the law of large numbers, we have $\Gamma_n \xrightarrow{p} \Gamma$ and $\Phi_n \xrightarrow{p} \Phi$, where $\Gamma = E[F(\boldsymbol{X}, \boldsymbol{Z}, U, h)\Pi\Pi^T]$ and $\Phi = E[F(\boldsymbol{X}, \boldsymbol{Z}, U, h)\Pi X^T]$.

Substituting (7.12) into (7.7), and similar to (7.11), we obtain that

$$\frac{1}{n} \sum_{i=1}^{n} \phi_h''(\varepsilon_i)\mathbf{X}_i(\mathbf{X}_i - \Phi_n^T\Gamma_n^{-1}\mathbf{\Pi}_i)^T(\hat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}_0)$$

$$= \quad \frac{1}{n} \sum_{i=1}^{n} \mathbf{X}_i[\phi_h'(\varepsilon_i) + \phi_h''(\varepsilon_i)\mathbf{Z}_i^T \mathbf{R}(U_i) - \phi_h''(\varepsilon_i)\mathbf{\Pi}_i^T[\Gamma_n + o_p(1)]^{-1}\Lambda_n + \phi_h''(\varepsilon_i)(\boldsymbol{\xi}_i - \hat{\boldsymbol{\xi}}_i)^T\boldsymbol{\beta}_0].$$

$$\tag{7.13}$$

Note that

$$\frac{1}{n} \sum_{i=1}^{n} \phi_h''(\varepsilon_i)\Phi_n^T\Gamma_n^{-1}\mathbf{\Pi}_i(\mathbf{X}_i - \Phi_n^T\Gamma_n^{-1}\mathbf{\Pi}_i)^T = 0$$

and

$$\frac{1}{n} \sum_{i=1}^{n} \Phi_n^T\Gamma_n^{-1}\mathbf{\Pi}_i\{\phi_h'(\varepsilon_i) + \phi_h''(\varepsilon_i)[\mathbf{Z}_i^T \mathbf{R}(U_i) - \mathbf{\Pi}_i^T\Gamma_n^{-1}\Lambda_n + (\boldsymbol{\xi}_i - \hat{\boldsymbol{\xi}}_i)^T\boldsymbol{\beta}_0]\} = 0.$$

Let $\tilde{\mathbf{X}}_i = \mathbf{X}_i - \Phi_n^T\Gamma_n^{-1}\mathbf{\Pi}_i$, (7.13) can be expressed by

$$[\frac{1}{n} \sum_{i=1}^{n} \phi_h''(\varepsilon_i)\tilde{\mathbf{X}}_i\tilde{\mathbf{X}}_i^T + o_p(1)] \sqrt{n}(\hat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}_0)$$

$$= \quad \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \phi_h'(\varepsilon_i)\tilde{\mathbf{X}}_i + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \phi_h''(\varepsilon_i)\tilde{\mathbf{X}}_i\mathbf{Z}_i^T \mathbf{R}(U_i)$$

$$- \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \phi_h''(\varepsilon_i) \tilde{\mathbf{X}}_i \boldsymbol{\Pi}_i^T \Gamma_n^{-1} \Lambda_n + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \phi_h''(\varepsilon_i) \tilde{\mathbf{X}}_i (\boldsymbol{\xi}_i - \hat{\boldsymbol{\xi}}_i)^T \boldsymbol{\beta}_0 + o_p(1)$$
$$= I_1 + I_2 + I_3 + I_4. \tag{7.14}$$

Since $\frac{1}{n} \sum_{i=1}^{n} \phi_h''(\varepsilon_i) \tilde{\mathbf{X}}_i \boldsymbol{\Pi}_i^T = \frac{1}{n} \sum_{i=1}^{n} \phi_h''(\varepsilon_i) \mathbf{X}_i \boldsymbol{\Pi}_i^T - \Phi_n^T \Gamma_n^{-1} \frac{1}{n} \sum_{i=1}^{n} \phi_h''(\varepsilon_i) \boldsymbol{\Pi}_i \boldsymbol{\Pi}_i^T = 0$, then $I_3 = 0$. Together with $\|\boldsymbol{R}(U_i)\| = O(K^{-r}) = o(1)$, a simple calculation yields $I_2 = o_p(1)$.

For I_4,

$$I_4 = \frac{\sqrt{n} b^2 \mu_{L_2}}{2n} \sum_{i=1}^{n} \tilde{\mathbf{X}}_i \phi_h''(\varepsilon_i) \boldsymbol{\beta}_0^T \boldsymbol{\xi}^{(2)}(V_i)$$
$$+ \frac{1}{n^{3/2}} \sum_{i=1}^{n} \sum_{j=1}^{n} \tilde{\mathbf{X}}_i \phi_h''(\varepsilon_i) \boldsymbol{\beta}_0^T f_V^{-1}(V_i) L_b(V_j - V_i) \boldsymbol{e}_j$$
$$= I_{41} + I_{42}.$$

As $nb^4 \to 0$, we have $I_{41} = o_p(1)$. Then through the similar argument of Zhang et al. [9], we have

$$I_{42} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} E(\tilde{\mathbf{X}}_i) \phi_h''(\varepsilon_i) \boldsymbol{\beta}_0^T \boldsymbol{e}_i.$$

Then, (7.14) can be written as

$$\frac{1}{n} \sum_{i=1}^{n} \phi_h''(\varepsilon_i) \tilde{\mathbf{X}}_i \tilde{\mathbf{X}}_i^T \sqrt{n}(\hat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}_0)$$
$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \phi_h'(\varepsilon_i) \tilde{\mathbf{X}}_i + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \phi_h''(\varepsilon_i) \tilde{\mathbf{X}}_i (\boldsymbol{\xi}_i - \hat{\boldsymbol{\xi}}_i)^T \boldsymbol{\beta}_0 + o_p(1). \tag{7.15}$$

By the Slutsky's theorem and (7.15), it follows that

$$\sqrt{n}(\hat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}_0) \xrightarrow{d} N(0, \boldsymbol{\Sigma}^{-1} \boldsymbol{\Omega} \boldsymbol{\Sigma}^{-1})$$

where $\boldsymbol{\Omega} = E(G(X, Z, U, h) \tilde{\mathbf{X}} \tilde{\mathbf{X}}^T) + \boldsymbol{\beta}_0^T \Sigma_e \boldsymbol{\beta}_0 E(F(X, Z, U, h)^2 E(\tilde{\mathbf{X}} \tilde{\mathbf{X}}^T))$. This completes the proof of Theorem 2.

$\square$