Statistics and Its Interface Volume 14 (2021) 359–371

Estimation and inference for covariate adjusted partially functional linear regression models

Zhiqiang Jiang, Zhensheng Huang*, and Hanbing Zhu

In this paper, we introduce covariate adjusted partially functional linear regression models, in which both the response and the covariates in the non-functional linear component can only be observed after being distorted by some multiplicative factors. We first estimate the distorting functions by nonparametrically regressing the response variables and covariates on the distorting covariate, and then the estimators of the slope function and the partially linear coefficient are obtained using the estimated response variables and covariates and functional principal component analysis based on corrected profile least-squares. We establish the asymptotic properties of the proposed estimators. In addition, using empirical likelihood and functional principal component analysis, we construct confidence intervals and bands for the coefficient parameters and the slope function, respectively. Finally, some simulation studies and an empirical analysis of a real dataset are conducted to illustrate the finite sample performance of the proposed method.

AMS 2000 subject classifications: Primary 62G08; secondary 62G20.
Keywords and phrases: Confidence region, Corrected profile least-squares, Covariate adjusted regression, Functional data, Functional principal component analysis, Partially functional linear models.

1. INTRODUCTION

With the progress in technology of data collection and storage, data are increasingly being collected in the form of smooth curves over time or other random variables taking values in a continuous domain. Such data are called functional data. As a result, in many applications, the response variable of interest is often related to both vector-valued and functional predictors. [19] used the term "hybrid data" to describe the situations when both vector-valued and function-valued predictors are of interest for modeling. One of the most basic and popular regression models with hybrid data is the following partially functional linear regression model

$$Y = \int_{\mathcal{J}} X(t)\gamma(t)dt + Z^\top \beta + \varepsilon, \tag{1}$$

where Y is a scalar continuous response, $X(\cdot)$ is a functional predictor defined on a bounded closed interval $\mathcal{J}$, $Z = (Z_1, \ldots, Z_p)^\top$ is a p-dimensional vector of scalar covariates, and ε is a random error with zero mean and finite variance σ^2.

Since [29] proposed the model (1), many scholars have studied this model. For instance, [24] proposed estimators for the parameters of model (1) and gave asymptotic properties of the estimators. [25] considered the prediction of model (1), and they proved that the two methods they considered achieved the same convergence rate of the mean squared prediction error under respective assumptions. [9] proposed a class of partially functional linear models to take into account multiple functional and ultra-high dimensional scalar predictors and identify important features. [26] considered the estimation and variable selection for model (1), and they derived the asymptotic distribution of the estimator of the vector of slope parameters and established oracle properties of penalized regression estimators.

All of the aforementioned articles on partially functional linear regression model (1) focused on the situation where covariate vector Z and response Y are all precisely observable, which is sometimes not the case in real applications. In fact, empirical datasets are often contaminated by measurement errors ([13, 14]), and the problem of covariates with measurement errors is one of the most fundamental issues in empirical data analysis in many fields such as medicine, engineering and economics. [30] considered the estimation and testing problems for partial functional linear regression models when scalar covariates are measured with errors. They assumed that covariance matrix of measurement errors is known. However, it is often that prior knowledge of covariance matrix of measurement errors may not exist, hence covariance matrix of measurement errors is unlikely to be known in practice. In addition, they only considered an additive measurement errors model. Indeed, in many applications, response and covariates are often subject to multiplicative distortion measurement errors. For example, [8] considered body mass index (BMI) as a distorting covariate when they analyzed the hemodialysis hemodialysis patients data.

There are a lot of literatures about distortion measurement errors. [20] introduced covariate adjusted regression for situations where both plasma fibrinogen concentration and serum transferrin level for 69 haemodialysis patients are

*Corresponding author.

thought to be influenced in a multiplicative fashion by BMI. They transformed the covariate adjusted linear model into a varying coefficient model, then obtained consistent estimators of parameters based on equidistant binning. Following this estimation procedure, [21] proposed covariate adjusted correlation analysis to target the correlation between house price and different influencing factors for Boston house price data, where proportion of population of lower educational status (LSTAT) are distorting covariates. See also, [22, 23]. But, this method is difficult to apply directly to the nonlinear model, because it fully utilizes and completely depends on the assumptions of linear structure in covariate adjusted regression. [4] applied nonparametric regression to obtain the estimators of the distorting functions, and then employed the least squares method to obtain the estimations of the parameter vector. Following this method, [3] proposed the covariate adjusted nonlinear regression and constructed empirical likelihood-based confidence regions for the unknown parameters. [11] studied covariate adjusted partially linear regression models, and apply their methods to explore the problem of calcium deficiency, where body surface area (BSA) have influence on variables of interest. [5] proposed varying coefficient partially nonlinear models with distorted measurement errors, and established the asymptotic properties of the resulting estimators.

The aforementioned work on the covariate adjusted regression has been restricted to vector-valued data. To our best knowledge, there is no literature which study the covariate adjusted regression with functional data. In order to fill this gap, we propose the following covariate adjusted partially functional linear regression model

$$(2) \qquad \begin{cases} Y = \int_{\mathcal{J}} X(t)\gamma(t)dt + Z^\top\beta + \varepsilon, \\ \widetilde{Y} = \psi(V)Y, \\ \widetilde{Z}_r = \phi_r(V)Z_r, \quad r = 1, \dots, p, \end{cases}$$

where Y and $Z = (Z_1, ..., Z_p)^\top$ are not observable, the distorted variables $\widetilde{Y}$ and $\widetilde{Z}_r$ are the observed response variable and covariates, respectively, the distorting covariate V is observable and independent of $(Y, X(\cdot), Z)$, $\psi(\cdot)$ and $\phi_r(\cdot)$ are unknown continuous multiplicative distortion functions. For notational simplicity, we suppose that $\mathcal{J} = [0, 1]$, and Y, Z and X are centered. In this paper, we first obtain the estimates of $\psi(V)$ and $\phi_r(V)$ by the estimation method proposed by [3]. Then, we calculate the bias-corrected estimates $(\hat{Y}, \hat{Z}_r)$ under the identifiability condition and obtain the profile least squares estimators of the parameter vector β and the coefficient function $\gamma(\cdot)$. Although the above estimation procedure in principle is applicable, the root n-consistency and asymptotic normality are very difficult to study. Hence, we propose using an empirical likelihood (EL) based confidence region that avoids estimating the limiting variance and has better accuracy. EL introduced by [15, 16] has become a widely used method to analyze regression type inference problems. Because the confidence intervals, constructed by EL, are range preserving and transformation respecting. Moreover, the shape of the confidence regions is determined completely by the data itself. See [17] for an overview.

Our motivation comes from the investigation of a real estate dataset which was collected from the statistical yearbooks and report on government work of various cities, real estate market reports and statistical bulletins on national economic and social development in China. Since it takes many years of savings for the average resident to buy a house, we choose the average annual income of the residents as the functional covariate. The scalar covariates of primary interests include urban category (UC), urban population (UP), urban GDP (GDP), urbanization rate (UR), urban comprehensive competitiveness (UCC), urban livability index (ULI) and urban development index (UDI). [26] analysed the real estate data set in China in 2016 to explore the relation between urban housing price (UHP) and their influencing factors by model (1). We update this data set from 2016 to 2017. On the one hand, we think that the UR should be a influencing factor because the higher UR of a city means the higher development level and the higher UHP of the city. On the other hand, inspired by the Boston housing price data in which the LSTAT is a distorting covariate (see [5, 6, 12, 21]), we think that UR may be a distorting variable because it affects some of the interesting variables in the data. So, we explore this data by model (2). The real data analysis in Section 5 proved that the model proposed in this paper fits the data better.

The rest of the paper is organized as follows. Section 2 proposes the estimation procedure and the corresponding asymptotic properties. Empirical likelihood-based confidence region for regression coefficient of non-functional predictors and the confidence bands centered at FPCA-based estimator for the slope function are established in Section 3. Simulation studies are carried out to illustrate our proposed approach in Section 4. Analysis of a real dataset is presented in Section 5. The proofs of the theorems are provided in Appendix A.

2. METHODOLOGY AND MAIN RESULTS

As in [20], we assume that the mean distorting effect vanishes, that is,

$$E[\psi(V)] = 1, \quad E[\phi_r(V)] = 1, \quad r = 1, \dots, p.$$

Let $\{\widetilde{Y}_i, \widetilde{Z}_i, V_i, X_i\}_{i=1}^n$ be an independent identically distributed (i.i.d.) random sample which comes from the model in (2), i.e.,

$$(3) \qquad \begin{cases} Y_i = \int_0^1 X_i(t)\gamma(t)dt + Z_i^\top\beta + \varepsilon_i, \\ \widetilde{Y}_i = \psi(V_i)Y_i, \\ \widetilde{Z}_{ir} = \phi_r(V_i)Z_{ir}, \quad r = 1, \dots, p; \ i = 1, \dots, n. \end{cases}$$

Let $Y = (Y_1, ..., Y_n)^\top$, $\widetilde{Y} = (\widetilde{Y}_1, ..., \widetilde{Y}_n)^\top$, $Z_r = (Z_{1r}, ..., Z_{nr})^\top$ and $\widetilde{Z}_r = (\widetilde{Z}_{1r}, ..., \widetilde{Z}_{nr})^\top$, as is introduced by [3], we have

$$\psi(V)=\frac{E(\widetilde{Y}|V)}{EY},\quad \phi_r(V)=\frac{E(\widetilde{Z}_r|V)}{EZ_r},\quad r=1,\dots,p.$$

By kernel estimation, we can obtain the estimators for $\psi(V)$ and $\phi(V)$ as follows:

$$\hat{\psi}(v)=\frac{\frac{1}{nh_0}\sum_{i=1}^n K(\frac{v-V_i}{h_0})\widetilde{Y}_i}{\frac{1}{nh_0}\sum_{i=1}^n K(\frac{v-V_i}{h_0})}\times\frac{1}{\bar{\bar{\widetilde{Y}}}}\triangleq\frac{\hat{g}_Y(v)}{\hat{p}(v)}\times\frac{1}{\bar{\bar{\widetilde{Y}}}},\tag{4}$$

$$\hat{\phi}_r(v)=\frac{\frac{1}{nh_r}\sum_{i=1}^n K(\frac{v-V_i}{h_r})\widetilde{Z}_{ir}}{\frac{1}{nh_r}\sum_{i=1}^n K(\frac{v-V_i}{h_r})}\times\frac{1}{\bar{\bar{\widetilde{Z}}}_r}\triangleq\frac{\hat{g}_r(v)}{\hat{p}(v)}\times\frac{1}{\bar{\bar{\widetilde{Z}}}_r},\tag{5}$$

where $\bar{\bar{\widetilde{Y}}}=\frac{1}{n}\sum_{i=1}^n\widetilde{Y}_i$, $\bar{\bar{\widetilde{Z}}}_r=\frac{1}{n}\sum_{i=1}^n\widetilde{Z}_{ir}$, $K(\cdot)$ is a kernel function and h_r $(r=0,1,\dots,p)$ are bandwidth depending on n. Then, we have

$$\hat{Y}_i=\frac{\widetilde{Y}_i}{\hat{\psi}(V_i)},\quad \hat{Z}_{ir}=\frac{\widetilde{Z}_{ir}}{\hat{\phi}_r(V_i)}.\tag{6}$$

Based on (3) and (6), we may get the following approximate formula:

$$\hat{Y}_i\approx\int_0^1 X_i(t)\gamma(t)dt+\hat{Z}_i^\top\beta+\varepsilon_i.\tag{7}$$

Denote the covariance function of $X(\cdot)$ by $K_X(s,t)$. Mercer's theorem implies that there exists a complete orthonormal basis $\{\varphi_k\}$ in $L^2(T)$ and a nonincreasing sequence of nonnegative eigenvalues $\{\nu_k\}$ such that $K_X(s,t)=\sum_{k=1}^\infty\nu_k\varphi_k(s)\varphi_k(t)$ with $\sum_{k=1}^\infty\nu_k<\infty$. We further assume that $\nu_1>\nu_2>\cdots>0$. Then we can get the Karhunen-Loéve expression of $X(t)$, $X(t)=\sum_{i=1}^\infty U_i\varphi_i(t)$, where the U_i are uncorrelated random variables with mean 0 and variance $E[U_i^2]=\nu_i$. Every function in L^2 all have Karhunen-Loéve expression, therefore model (7) can be approximated as:

$$\hat{Y}_i\doteq\sum_{j=1}^m\gamma_j\langle X_i,\varphi_j\rangle+\hat{Z}_i^\top\beta+\varepsilon_i,\tag{8}$$

where $\gamma_j=\langle\gamma,\varphi_j\rangle$, m is large enough. In fact, in order to estimate β and $\gamma_j, j=1,\cdots,m$, we must get the substitute $\hat{\varphi}$ of φ. To this end, we consider the empirical approximation of covariance function $X(\cdot)$ as follows,

$$\hat{K}_X(s,t)=\frac{1}{n}\sum_{i=1}^n X_i(s)X_i(t)=\sum_{i=1}^\infty\hat{\nu}_i\hat{\varphi}_i(s)\hat{\varphi}_i(t),$$

where $(\hat{\nu}_i,\hat{\varphi}_i)$ are pairs of eigenvalue and eigenfunction for $\hat{K}_X$ and $\hat{\nu}_1\ \hat{\nu}_2\ \cdots\geq 0$. We use $(\hat{\nu}_i,\hat{\varphi}_i)$ as the estimator of (ν_i,φ_i), $i=1,\dots,m$. Then (8) can be written as

$$\hat{Y}\doteq U_m\gamma+\hat{Z}\beta+\varepsilon,$$

where $\hat{Y}=(\hat{Y}_1,...,\hat{Y}_n)^\top$, $U_m=\{\langle X_i,\hat{\varphi}_j\rangle\}_{1\leq i\leq n,1\leq j\leq m}$, $\gamma=(\gamma_1,\dots,\gamma_m)^\top$, $\hat{Z}=(\hat{Z}_1,\dots,\hat{Z}_n)^\top$. The β and γ can be estimated by minimizing

$$(\hat{Y}-U_m\gamma-\hat{Z}\beta)^\top(\hat{Y}-U_m\gamma-\hat{Z}\beta)$$

over β and γ. For $m+p\leq n$, the corrected profile least-squares estimator (CPLSE) for β is given by

$$\hat{\beta}=\left[\hat{Z}^\top(I-S_m)\hat{Z}\right]^{-1}\hat{Z}^\top(I-S_m)\hat{Y},$$

where $S_m=U_m(U_m^\top U_m)^{-1}U_m^\top$. Therefore,

$$\hat{\gamma}=(U_m^\top U_m)^{-1}U_m^\top(\hat{Y}-\hat{Z}\hat{\beta}).$$

In the following, we adopt the notations of [24], and define

$$\hat{K}_{\hat{Z}X}=\frac{1}{n}\sum_{i=1}^n\hat{Z}_iX_i(t),\qquad \hat{K}_{X\hat{Z}}=\frac{1}{n}\sum_{i=1}^n X_i(t)\hat{Z}_i,$$

$$\hat{K}_{\hat{Z}}=\frac{1}{n}\sum_{i=1}^n\hat{Z}_i\hat{Z}_i^T,\ \hat{K}_{\hat{Z}\hat{Y}}=\frac{1}{n}\sum_{i=1}^n\hat{Z}_i\hat{Y}_i,\ \hat{K}_{\hat{Y}X}=\frac{1}{n}\sum_{i=1}^n\hat{Y}_iX_i(t).$$

By the simple calculation, it can be shown that $\hat{\beta}$ is equal to

$$\hat{\beta}=\Big(\hat{K}_{\hat{Z}}-\sum_{j=1}^m\frac{\langle\hat{K}_{\hat{Z}X},\hat{\varphi}_j\rangle\langle\hat{K}_{X\hat{Z}},\hat{\varphi}_j\rangle}{\hat{\nu}_j}\Big)^{-1}\Big(\hat{K}_{\hat{Z}\hat{Y}}-\sum_{j=1}^m\frac{\langle\hat{K}_{\hat{Z}X},\hat{\varphi}_j\rangle\langle\hat{K}_{\hat{Y}X},\hat{\varphi}_j\rangle}{\hat{\nu}_j}\Big),$$

where

$$\begin{aligned}\langle\hat{K}_{WX},\hat{\varphi}_j\rangle=&\frac{1}{n}\sum_{i=1}^n W_i\langle X_i,\hat{\varphi}_j\rangle,\langle\hat{K}_{XW},\hat{\varphi}_j\rangle\\=&\frac{1}{n}\sum_{i=1}^n W_i^\top\langle X_i,\hat{\varphi}_j\rangle,\langle\hat{K}_{YX},\hat{\varphi}_j\rangle\\=&\frac{1}{n}\sum_{i=1}^n Y_i^\top\langle X_i,\hat{\varphi}_j\rangle.\end{aligned}$$

Similarly, the estimator of γ is given by $\hat{\gamma}=(\hat{\gamma}_1,\hat{\gamma}_2,\cdots,\hat{\gamma}_m)^\top$ with $\hat{\gamma}_j=\frac{\langle\hat{K}_{\hat{Y}X}-\hat{K}_{\hat{Z}X}^\top\hat{\beta},\hat{\varphi}_j\rangle}{\hat{\nu}_j}$, for $j=1,\cdots,m$, then

$$\hat{\gamma}(t)=\sum_{i=1}^m\hat{\gamma}_j\hat{\varphi}_j(t).$$

Next we present the asymptotic behavior of $\hat{\beta}$.

Theorem 2.1. *Assume that the conditions (C1)-(C12) in the Appendix A hold, we have*

$$\sqrt{n}(\hat{\beta}-\beta)\xrightarrow{d}N(0,\Sigma),$$

where $\Sigma = \sigma^2 B^{-1} + \frac{1}{4}\Lambda + \Gamma$, *B is defined as in (C8) and for* $1 \le r, k \le p$,

$$\Lambda(r,k) = \beta_r\beta_k E\Big[\Big(\frac{Y-EY}{EY} - \frac{Z_r - EZ_r}{EZ_r}\Big)\Big(\frac{Y-EY}{EY} - \frac{Z_k - EZ_k}{EZ_k}\Big)\Big],$$

$$\Gamma(r,k) = \beta_r\beta_k E\Big[\Big(\frac{\widetilde{Y}-Y}{EY} - \frac{\widetilde{Z}_r - Z_r}{EZ_r}\Big)\Big(\frac{\widetilde{Y}-Y}{EY} - \frac{\widetilde{Z}_k - Z_k}{EZ_k}\Big)\Big].$$

Theorem 2.1 states that the estimator of β in (2) is asymptotically normal. The second and third terms of Σ are caused by the distortions. If the variables Y and Z_r can be directly observed, where $\psi = 1$ and $\phi = 1$, then its asymptotic covariance matrix is $\sigma^2 B^{-1}$, which was calculated by [24].

Theorem 2.2. *Assume that the conditions (C1)-(C12) in the Appendix A hold, we have*

$$\|\hat{\gamma} - \gamma\|^2 = O_p(n^{-(2b-1)/(a+2b)}).$$

Theorem 2.2 is similar to the result of [24], which indicates that the calibrated variables have little effect on the asymptotic normality of $\hat{\gamma}$. In the estimation procedure, the optimal bandwidths of h_s can be selected adaptively by leave-one-curve-out cross-validation.

3. CONSTRUCTION OF CONFIDENCE REGIONS

3.1 Empirical likelihood based confidence region for β

In this section, we discuss the construction of confidence region for β. Normal approximation (NA) is a natural approach, because we have obtained the asymptotic normality in the above section. However, as is shown in the Theorem 2.1, the matrix Σ is rather complex and includes several unknown components to be estimated. If we use normal approximation to construct a confidence region for β, we need a plug-in estimation that involves the estimation for many unknown components. So, we propose empirical likelihood to construct confidence region for β.

To construct a confidence region for β, we first introduce an auxiliary random variable $\eta_{ni}(\beta) = Z_i[Y_i - \beta^\top Z_i - (U_i^*)^\top \hat{\gamma}_Z]$, where $\hat{\gamma}_Z = (U_m^\top U_m)^{-1} U_m^\top (Y - Z\beta)$. Similar to [15], an empirical log-likelihood ratio can be defined as

$$l(\beta) = -2\max\Big\{\sum_{i=1}^{n}\log n\omega_i : \omega_i \ge 0, \sum_{i=1}^{n}\omega_i = 1, \sum_{i=1}^{n}\omega_i\eta_{ni}(\beta) = 0\Big\}.$$

However, Y_i and Z_i in $l(\beta)$ are unobservable. A natural method is to replace them by their estimators $\hat{Y}_i$ and $\hat{Z}_i$, respectively. Then, the adjusted empirical log-likelihood ratio is

$$\hat{l}(\beta) = -2\max\Big\{\sum_{i=1}^{n}\log n\omega_i : \omega_i \ge 0, \sum_{i=1}^{n}\omega_i = 1, \sum_{i=1}^{n}\omega_i\hat{\eta}_{ni}(\beta) = 0\Big\},$$

where $\hat{\eta}_{ni}(\beta) = \hat{Z}_i[\hat{Y}_i - \beta^\top \hat{Z}_i - (U_i^*)^\top \hat{\gamma}_{\hat{Z}}]$ and $\hat{\gamma}_{\hat{Z}} = (U_m^\top U_m)^{-1} U_m^\top (\hat{Y} - \hat{Z}\beta)$. By the Lagrange multiplier method, we obtain

$$\hat{l}(\beta) = 2\sum_{i=1}^{n}\log\Big\{1 + \lambda^\top \hat{\eta}_{ni}(\beta)\Big\}, \tag{9}$$

where $\lambda \in R^p$ is the root of

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\hat{\eta}_{ni}(\beta)}{1 + \lambda^\top\hat{\eta}_{ni}(\beta)} = 0. \tag{10}$$

Theorem 3.1. *Assume that the conditions (C1)-(C12) in the Appendix A hold. If the β is the true value of the parameter value, we have*

$$\hat{l}(\beta) \xrightarrow{d} \chi_p^2.$$

Based on the Theorem 3.1, an empirical likelihood confidence region for β with nominal confidence level α is

$$I_\alpha(\beta) = \{\beta : \hat{l}(\beta) \le c_\alpha\},$$

where c_α satisfies $P(\chi_p^2 \le c_\alpha) = 1 - \alpha$.

3.2 Confidence band for the slope function

For a given $\tau \in (0,1)$, a confidence band for $\gamma(t)$ with level $1 - \tau$ is a collection of random intervals $\mathcal{C} = \{\mathcal{C}(t) = [\ell(t), u(t)] : t \in [0,1]\}$ such that

$$P\{\gamma(t) \in [\ell(t), u(t)] \text{ for all } t \in [0,1]\} \ge 1 - \tau. \tag{11}$$

As [7] point out the requirement (11) is too stringent. So, in this paper, we adopt a less demanding requirement proposed by [7] to construct confidence bands for $\gamma(t)$. For a given $\tau \in (0,1)$, let

$$\hat{c}_n(1-\tau) = \text{conditional}(1-\tau)\text{-quantile of } \sqrt{\sum_{j=1}^{m}\eta_j/\hat{\nu}_j} \text{ given } X_1^n,$$

where $X_1^n = \{X_1, \ldots, X_n\}$, $\eta_1, \ldots, \eta_m$ are independent $\chi^2(1)$ random variables independent from X_1^n, and $\hat{c}_n(1-\tau)$ can be computed by simulations in practice. Then, the confidence band for $\gamma(t)$ is as follows

$$\hat{\mathcal{C}}(t) = \left[\hat{\gamma}(t) - \frac{\hat{\sigma}\hat{c}_n(1-\tau_1)}{\sqrt{n}}\sqrt{\frac{1}{\tau_2\mathfrak{L}(I)}}, \hat{\gamma}(t) + \frac{\hat{\sigma}\hat{c}_n(1-\tau_1)}{\sqrt{n}}\sqrt{\frac{1}{\tau_2\mathfrak{L}(I)}}\right], \tag{12}$$

where τ_1, τ_2 are constants in $(0,1)$, $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\{Y_i - Z_i^\top\hat{\beta} - \int_0^1 X_i(t)\hat{\gamma}(t)dt\}^2$, and $\mathfrak{L}(\cdot)$ denotes the Lebesgue measure.

Theorem 3.2. *Assume that the conditions (C1)-(C12) in the Appendix A hold. For given $\tau_1, \tau_2 \in (0,1)$, if $\varepsilon \sim N(0, \sigma^2)$ and $m^{a+2b-1}/n \to \infty$, then we have*

$$\mathbb{P}\left\{\|\hat{\gamma} - \gamma\| \le \hat{\sigma}\hat{c}_n(1-\tau_1)/\sqrt{n}\right\} = 1 - \tau_1 + o(1), \tag{13}$$

Table 1. The bias, variance and MSE of the estimators for β

	$n=100$			$n=200$			$n=500$		
	bias	variance	MSE	bias	variance	MSE	bias	variance	MSE
$\breve{\beta}_1$ $(\times 10^{-4})$	-335.45	57.29	68.54	-316.52	29.40	39.42	348.48	11.27	23.41
$\breve{\beta}_2$ $(\times 10^{-4})$	213.45	65.50	70.06	294.42	31.57	40.24	307.96	14.17	23.65
$\hat{\beta}_1$ $(\times 10^{-4})$	66.16	14.90	15.69	79.37	14.81	15.11	46.15	12.9	15.04
$\hat{\beta}_2$ $(\times 10^{-4})$	-53.12	14.28	14.27	71.37	15.27	14.77	38.02	12.15	13.59
$\tilde{\beta}_1$ $(\times 10^{-4})$	13.51	8.81	8.83	0.65	3.36	3.36	-3.82	1.20	1.20
$\tilde{\beta}_2$ $(\times 10^{-4})$	-4.01	8.31	8.31	-13.81	3.94	3.96	9.85	1.22	1.23

Table 2. The integrated squared bias, integrated variance and MISE of the estimators for $\gamma(t)$

	$n=100$			$n=200$			$n=500$		
	bias^2	variance	MISE	bias^2	variance	MISE	bias^2	variance	MISE
$\breve{\gamma}$ $(\times 10^{-2})$	6.00	40.07	46.07	6.08	23.54	29.62	5.85	11.41	17.26
$\hat{\gamma}$ $(\times 10^{-2})$	3.51	18.27	21.78	3.86	10.29	14.15	3.71	8.39	12.11
$\tilde{\gamma}$ $(\times 10^{-2})$	1.55	5.25	6.81	1.66	4.26	5.92	1.74	4.14	5.88

and the confidence band $\hat{\mathcal{C}}(t)$ for $\gamma(t)$ satisfies

$$(14)\quad \mathbb{P}\Big\{\mathfrak{L}\Big(\Big\{t\in[0,1]:\gamma(t)\notin\hat{\mathcal{C}}(t)\Big\}\Big)\le\tau_2\mathfrak{L}(\mathcal{I})\Big\}\ge 1-\tau_1+o(1).$$

Furthermore, the width of the confidence band $\widehat{\mathcal{C}}(t)$ is $O_P(n^{-(2b-1)/(a+2b)})$.

Note that (14) means that with probability at least $1-\tau_1+o(1)$, the proportion of the set $[0,1]$ at which $\gamma(t)$ is not covered by $\hat{\mathcal{C}}$ is at most τ_2.

4. SIMULATION STUDY

In this section, we carry out some simulations to illustrate the theories of Section 2 and 3. Here, we assume that the kernel function $K(t)=\frac{15}{32}(3-7t^2)(1-t^2)_+$. The tuning parameter m is selected adaptively by leave-one-curve-out cross-validation (see [24]).

Consider the following covariate adjusted partially functional linear model:

$$(15)\qquad \begin{cases} Y=\int_0^1 X(t)\gamma(t)dt+Z^\top\beta+\varepsilon,\\ \widetilde{Y}=\psi(V)Y,\\ \widetilde{Z}_r=\phi_r(V)Z_r,\quad r=1,2.\end{cases}$$

where $\beta=[-0.75,0.5]^T$, $Z_1\sim N(0,1)$, $Z_2\sim N(0,1)$, the covariates $V\sim N(4,1)$, $\psi(V)=(V+1)^2/26$, $\phi_1(V)=(V+1)/5$, $\phi_2(V)=(V+11)/15$ and the model error $\varepsilon\sim N(0,1)$. We take $\gamma(t)=\sum_{j=1}^{50}\gamma_j\varphi_j(t)$ and $X_i(t)=\sum_{j=1}^{50}\nu_{ij}\varphi_j(t)$, where $\gamma_1=0.3$ and $\gamma_j=4(-1)^{j+1}j^{-2}$, $j\ge 2$; $\varphi_1(t)\equiv 1$ and $\varphi_j(t)=2^{1/2}\cos((j-1)\pi t)$, $j\ge 2$; the ν_{ij}'s are uniformly distributed on $[-3^{1/2},3^{1/2}]$. In order to construct the confidence band (CB) of $\gamma(t)$, we take $\tau_1=0.9$ and $\tau_2=0.9$ in (12). The results given below are based on 500 replications. The sample size is $n=100$, $n=200$ and $n=500$, respectively.

In order to show the performance of the $\hat{\beta}$ (proposed in this paper), we compare it with other estimators: the naive estimator (NE) $\breve{\beta}$ (neglecting the distortion measurement error) and the benchmark estimator (BE) $\tilde{\beta}$ (Y and Z_r can be observed exactly). The bias, variance and mean squared error (MSE) of the estimators of β are reported in Table 1. In Table 2, the estimator of $\gamma(t)$ is assessed by the integrated squared bias (bias^2), integrated variance (var) and mean integrated squared error (MISE). In Fig. 1, we show the estimators of slope function and their 90% confidence bands based on $n=500$. In order to show the performance of EL, some representative coverage probabilities (CP) on (β_1,β_2) with the nominal level $1-\alpha=0.95$ are reported in Table 3. Meanwhile, confidence regions for (β_1,β_2) are presented in Figs. 2.

Table 3. The CP of the confidence regions on $(\beta_1,\beta_2)^T$

Method	n=100	n=200	n=500
Empirical likelihood	0.9140	0.9320	0.9480
Normal approximation	0.8950	0.9100	0.9270

From these simulation results, we draw the following conclusions. In Tables 1 and 2, the bias (integrated squared bias), variance (integrated variance) and mean squared error (mean integrated squared error) of estimators based on our proposed method is closer to the results of benchmark estimators than that of naive estimators. Fig. 1 demonstrates that $\hat{\gamma}(t)$ outperforms $\breve{\gamma}(t)$. From Table 3, it can see that the EL approach outperforms the NA method due to the EL approach has slightly higher CP than the NA method. At the same time, Table 3 indicates that the CP is close to 95% as n increases both for the EL and NA methods. Further, in Fig. 2, the confidence regions based on the EL approach are smaller than that of the NA method for all cases.

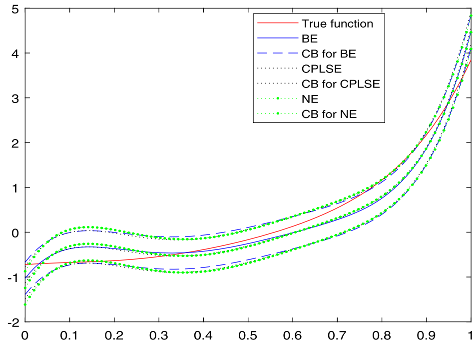

Figure 1. The estimators of the slope function based on $n = 500$, and their 90% confidence bands.

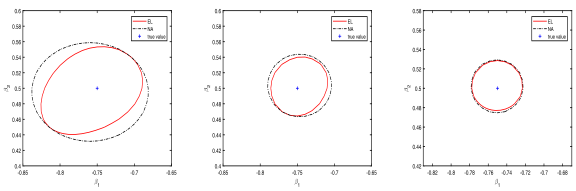

Figure 2. The 95% confidence regions for (β_1, β_2) based on $n = 100$, $n = 200$, $n = 500$.

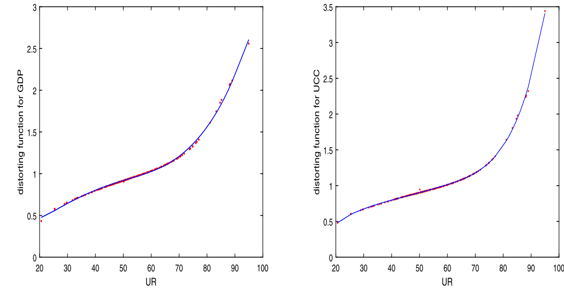

Figure 3. The estimated curves of distorting function for UHP, GDP and UCC.

5. REAL DATA ANALYSIS

We apply our method to a real estate data set which was collected from the statistical yearbooks and report on government work of various cities, real estate market reports and statistical bulletins on national economic and social development in China. It includes the real estate data for 190 second-, third- and fourth-tier cities in China according to the full list of China's urban classification in 2017. In this data set, there are the average annual income of urban residents from 2008 to 2017, and the other data are based on 2017. The response variable is urban housing price (UHP). Since it takes many years of savings for the average resident to buy a house, we choose the average annual income of the residents as the functional covariate (see [26]). Let $X_i^*(t)$ denote the average annual income of the residents of the i-th city for the year t and $X_i(t) = X_i^*(t) - \bar{X}^*(t)$, where $\bar{X}^*(t) = \frac{1}{190}\sum_{i=1}^{190} X_i^*(t)$. The scalar covariates of primary interests include urban category (UC), urban population (UP), urban GDP (GDP), urbanization rate (UR), urban comprehensive competitiveness (UCC), urban livability index (ULI) and urban development index (UDI). Let $Y_i = \log(\mathrm{UHP}_i)$, (Z_{i1}, Z_{i2}) stand for $\mathrm{UC_i}$ ($Z_{i1} = 1$ and $Z_{i2} = 0$ stand for second-tier city, $Z_{i1} = 0$ and $Z_{i2} = 1$ stand for third-tier city, and $Z_{i1} = 0$ and $Z_{i2} = 0$ stand for fourth-tier city), $Z_{i3} = \mathrm{UP}_i$, $Z_{i4} = \mathrm{GDP}_i$, $Z_{i5} = \mathrm{UCC}_i$, $Z_{i6} = \mathrm{ULI}_i$, $Z_{i7} = \mathrm{UDI}_i$ and $Z_{i8} = \sqrt{\mathrm{UR}_i}$. We note that among these variables the data of some variables such as Z_3 and Z_4 are very large, whereas those of some variables such as Z_7 are small. For this purpose, for each data of these variables, we first make the following modification: let z_{i3}, $i = 1, \cdots, 190$ be the observations of Z_3. Let $Z_{i3} = z_{i3}/\max z_{i3}$, $i = 1, \cdots, 190$, so that the maximum of modified data of the variable Z_3 is 1. The data of the variables $Z_4, \cdots, Z_7$ are modified in a similar fashion. We construct the following models:

$$Y_i = \int_0^{10} X_i(t)\gamma(t)dt + \sum_{q=1}^{7} Z_{iq}^\top \beta_q + \varepsilon_i, \tag{16}$$

$$Y_i = \int_0^{10} X_i(t)\gamma(t)dt + \sum_{q=1}^{8} Z_{iq}^\top \beta_q + \varepsilon_i, \tag{17}$$

$$\begin{cases} Y_i = \int_0^{10} X_i(t)\gamma(t)dt + \sum\limits_{q=1}^{8} Z_{iq}^\top \beta_q + \varepsilon_i, \\ \widetilde{Y}_i = \psi(V_i)Y_i, \\ \widetilde{Z}_{iq} = \phi_q(V_i)Z_{iq}, \quad q = 1, \ldots, 8, \end{cases} \tag{18}$$

where V is UR. Model (16) is the model considered in [26], in which UR was not taken into account. In model (17), we add $\sqrt{\mathrm{UR}}$ to model (16). Furthermore, in model (18), we take UR as a distorting variable.

To make the analysis results comparable, the kernel function is taken as $K(t) = \frac{15}{32}(3 - 7t^2)(1 - t^2)_+$, $\tau_1 = 0.9$ and $\tau_2 = 0.9$. Firstly, we present the estimated curves of $\hat{\psi}(V)$, $\hat{\phi}_4(V)$ and $\hat{\phi}_5(V)$ in Fig. 3 (the other estimated curves are similar and not shown here due to space limitations). Then, the estimates of β_q and the residual standard deviation σ for all models are given in Table 4. Table 5 gives 95% confidence intervals for the scalar covariates based on the model (16), (17) and (18). At the same time, the estimated regression parameter functions together with their 90% confidence bands based on the model (16), (17) and (18) are plotted in Fig. 4.

From Fig. 3, we can see that the estimated curves of $\hat{\psi}(V)$, $\hat{\phi}_4(V)$ and $\hat{\phi}_5(V)$ are not horizontal (the other estimated curves are similar), which gives the evidence that the distorting covariate UR has a connection with UHP, UC, UP, GDP, ULI, UCC and UDI. We can observe from Table 4 that the residual standard deviation σ of model (17) is smaller than that of model (16), and the residual

Table 4. *The estimates of $\beta_1, \cdots, \beta_8$ and the residual standard derivation σ for the model (16), (17) and (18)*

Model	σ	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8
(5.1)	2.5827	-5.2065	-0.8387	11.1700	0.4847	8.4524	3.2297	9.1355	\
(5.2)	0.8928	-0.8069	-0.0926	2.8370	1.4489	-2.3289	0.0041	1.048	1.1496
(5.3)	0.5114	0.2032	0.1234	0.5320	0.0425	-0.1071	0.0101	0.3671	1.1756

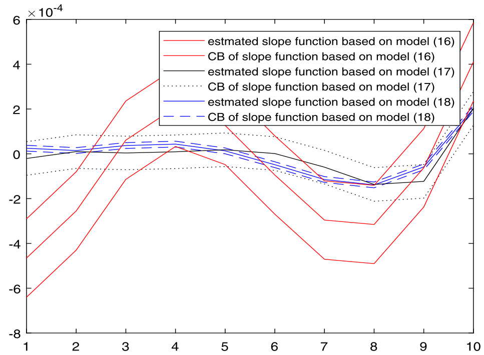

Figure 4. The estimated regression parameter functions, 90% confidence bands for average annual income of the residents based on model (16), (17) and (18).

standard deviation σ of model (18) is smaller than that of model (17). Therefore, the assumption that UR is a influencing factor is appropriate. Also, the assumption that UR is a confounder contaminating the variables UHP, UC, UP, GDP, UCC, ULI and UDI is reasonable. It can be seen from Fig. 4 that the confidence bands based on model (18) are smaller than the ones based on model (16) and model (17). Table 5 indicates that the model (18) gives smaller intervals and imposes symmetry on the confidence intervals. In general, for this dataset, the model (18) has better performance than that of model (17) and model (16).

Table 5. *The 95% confidence intervals based on model (16), (17) and (18) for the real estate data set*

Model	Confidence intervals		
	(16)	(17)	(18)
β_1	[-4.1340,-0.7992]	[-1.4586,-0.0262]	[-0.1515,0.4907]
β_2	[-1.3362,0.4439]	[-0.5768,0.1777]	[-0.2261,0.2219]
β_3	[7.1195,14.1861]	[1.5860,4.7562]	[-0.6414,1.5832]
β_4	[-5.7798,2.9926]	[-0.9348,2.7933]	[-1.6849,1.2006]
β_5	[4.7840,15.4240]	[-3.3558,1.4089]	[-1.2389,2.6388]
β_6	[3.1632,9.4517]	[-0.1731,2.5836]	[-0.3891,1.3632]
β_7	[2.2094,9.3607]	[-0.4858,2.6133]	[-0.4752,1.5608]
β_8	\	[0.9730,1.1197]	[1.0814,1.1755]

APPENDIX A. PROOF OF THE MAIN RESULTS

In order to prove the main results, we first list some regular conditions used in this paper.

(C1) The square integrable random function X satisfies $E\|X\|^4 < \infty$.

(C2) For each j, $E[U_j^4] \le C\lambda_j^2$ for some constants C.

(C3) For the eigenvalues λ_j, there exist some constants C and $a > 1$ such that $C^{-1}j^{-a} \le \lambda_j \le Cj^{-a}$, $\lambda_j - \lambda_{j+1} \ge Cj^{-a-1}$, $j \ge 1$.

(C4) For the Fourier coefficients γ_j, there exist some constants C and b, where $b > a/2 + 1$, such that $|\gamma_j| \le Cj^{-b}$, $j \ge 1$.

(C5) For the tuning parameter m, we suppose that $m \sim n^{1/(a+2b)}$, where the notation $a_n \sim b_n$ means that there exist constants $0 < L < M < \infty$ such that $L \le a_n/b_n \le M$.

(C6) $E\|X\|^4_{R^p} < \infty$, where $\|A\|_{R^p} = (A^\top A)^{1/2}$ is the usual Euclidean norm in R^p.

(C7) There exist some constants C such that for each k, $|\langle K_{Z_k X}, \varphi_j \rangle| \le Cj^{-(b+a)}$, $j \ge 1$.

(C8) $B = K_Z - \sum_{j=1}^{\infty} \frac{\langle K_{ZX}, \varphi_j\rangle \langle K_{XZ}, \varphi_j\rangle}{\nu_j}$ is a positive definite matrix.

(C9) The continuous kernel function $K(\cdot)$ has the following properties:

  (1) The support of $K(\cdot)$ is the interval $[-1, 1]$;

  (2) $K(\cdot)$ is symmetric about zero;

  (3) $\int_{-1}^{1} K(v)dv = 1$, $\int_{-1}^{1} v^i K(v)dv = 0$, $i = 1, 2, 3$.

(C10) All $g_r(v) = \phi_r(v)p(v)$, $1 \le r \le p$, $g_Y(v) = \psi(v)p(v)$, $\phi_r(v)$, $\psi(v)$ and $p(v)$ (density of V) are greater than a positive constant and are differential and their derivatives satisfy the following condition: there exists a neighborhood of the origin, say Δ, and a constant $c > 0$ such that, for any $\delta \in \Delta$,

$$|g_r^{(3)}(v+\delta) - g_r^{(3)}(v)| \le c|\delta|, \ \ 1 \le r \le p,$$
$$|g_Y^{(3)}(v+\delta) - g_Y^{(3)}(v)| \le c|\delta|,$$
$$|p^{(3)}(v+\delta) - p^{(3)}(v)| \le c|\delta|.$$

(C11) In the limit as $n \to \infty$, the bandwidth h_j, $j = 0, 1, \ldots, p$ varies in the range between $O(n^{-\frac{1}{4}} \log n)$ and $O(n^{-\frac{1}{8}})$.

(C12) EY and EZ_r are all bounded away from zero, $1 \le r \le p$, $E[Y]^2 < \infty$ and $E[Z_r]^2 < \infty$.

Lemma A.1. *Suppose that (C1)-(C12) hold, we can obtain the following asymptotic representations:*

(i) $\frac{1}{\sqrt{n}}[\hat{Z}^\top(I-S_m) - Z^\top(I-S_m)](\hat{Y}-Y) = o_p(1)$.

(ii) $\frac{1}{\sqrt{n}}Z^\top(I-S_m)(\hat{Y}-Y) = \frac{1}{2}(\frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{Y_i-EY}{EY})B\beta$

$+ (\frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{\widetilde{Y}_i-Y_i}{EY})B\beta + o_p(1)$.

(iii) $\frac{1}{\sqrt{n}}Z^\top(I-S_m)[\hat{Z}-Z]$

$= \frac{1}{2}B\text{diag}\{\frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{Z_{i1}-EZ_1}{EZ_1}, \frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{Z_{i2}-EZ_2}{EZ_2}, \ldots, \frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{Z_{ip}-EZ_p}{EZ_p}\}$

$+B\text{diag}\{\frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{\widetilde{Z}_{i1}-Z_{i1}}{EZ_1}, \frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{\widetilde{Z}_{i2}-Z_{i2}}{EZ_2}, \ldots, \frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{\widetilde{Z}_{ip}-Z_{ip}}{EZ_p}\} + o_p(1)$.

Proof of Lemma A.1. Proof of (i). Noting that $S_m = U_m(U_m^\top U_m)^{-1}U_m^\top$, we have $(I-S_m)^\top(I-S_m) = I-S_m$ and $(I-S_m)^2 = I-S_m$. Then, by the properties of $n\times n$ matrix, we can know that the maximum eigenvalue of matrix $I-S_m$ is bounded. The r-th element of $\frac{1}{\sqrt{n}}[\hat{Z}^\top(I-S_m) - Z^\top(I-S_m)](\hat{Y}-Y)$ is

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\Big[\hat{Z}_r^\top(I-S_m) - Z_r^\top(I-S_m)\Big](\hat{Y}-Y)\\
=&\frac{1}{\sqrt{n}}(\hat{Z}_r^\top - Z_r^\top)(I-S_m)(\hat{Y}-Y)\\
\le& C\frac{1}{\sqrt{n}}(\hat{Z}_r^\top - Z_r^\top)(\hat{Y}-Y) = o_p(1).
\end{aligned}$$

The last equation is obtained due to the fact that $\frac{1}{\sqrt{n}}(\hat{Z}^\top - Z^\top)(\hat{Y}-Y) = o_p(1)$ holds (see [4]).

Proof of (ii). Note that

$$\begin{aligned}
&\frac{1}{\sqrt{n}}Z^\top(I-S_m)(\hat{Y}-Y)\\
=&\frac{1}{\sqrt{n}}\sum_{i=1}^n(Z_i - \sum_{j=1}^m \frac{\langle\hat{K}_{ZX},\hat{\varphi}_j\rangle\langle X_i,\hat{\varphi}_j\rangle}{\hat{\nu}_j})(\hat{Y}_i - Y_i)\\
=&\frac{1}{\sqrt{n}}\sum_{i=1}^n(Z_i - \sum_{j=1}^m \frac{\langle\hat{K}_{ZX},\hat{\varphi}_j\rangle\langle X_i,\hat{\varphi}_j\rangle}{\hat{\nu}_j})\hat{Y}_i\\
&-\frac{1}{\sqrt{n}}\sum_{i=1}^n(Z_i - \sum_{j=1}^m \frac{\langle\hat{K}_{ZX},\hat{\varphi}_j\rangle\langle X_i,\hat{\varphi}_j\rangle}{\hat{\nu}_j})Y_i.
\end{aligned}$$

Thus, it suffices to deal with the quantity $\frac{1}{\sqrt{n}}\sum_{i=1}^n\Big(Z_i - \sum_{j=1}^m \frac{\langle\hat{K}_{ZX},\hat{\varphi}_j\rangle\langle X_i,\hat{\varphi}_j\rangle}{\hat{\nu}_j}\Big)\hat{Y}_i$. By (4) and (5), we have

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^n\Big(Z_i - \sum_{j=1}^m \frac{\langle\hat{K}_{ZX},\hat{\varphi}_j\rangle\langle X_i,\hat{\varphi}_j\rangle}{\hat{\nu}_j}\Big)\hat{Y}_i\\
=&\frac{1}{\sqrt{n}}\sum_{i=1}^n\Big(Z_i - \sum_{j=1}^m \frac{\langle\hat{K}_{ZX},\hat{\varphi}_j\rangle\langle X_i,\hat{\varphi}_j\rangle}{\hat{\nu}_j}\Big)\widetilde{Y}_i\Big[\frac{\hat{p}(v_i)}{\hat{g}_Y(v_i)}\Big]\bar{\widetilde{Y}}\\
=&\frac{1}{\sqrt{n}}\sum_{i=1}^n\Big(Z_i - \sum_{j=1}^m \frac{\langle\hat{K}_{ZX},\hat{\varphi}_j\rangle\langle X_i,\hat{\varphi}_j\rangle}{\hat{\nu}_j}\Big)EY\\
&+\frac{1}{\sqrt{n}}\sum_{i=1}^n\Big(Z_i - \sum_{j=1}^m \frac{\langle\hat{K}_{ZX},\hat{\varphi}_j\rangle\langle X_i,\hat{\varphi}_j\rangle}{\hat{\nu}_j}\Big)\widetilde{Y}_i\Big[\frac{\hat{p}(v_i)}{\hat{g}_Y(v_i)}\Big](\bar{\widetilde{Y}} - EY)\\
\triangleq& L_1 + L_2.
\end{aligned}$$

First, we consider L_1. The proof is divided into three steps.

Step 1. Show that

$$L_1 = L_{11} - L_{12} + L_{13} + o_p(1), \tag{19}$$

where

$$\begin{aligned}
L_{11} &= \frac{1}{\sqrt{n}}\sum_{i=1}^n\Big(Z_i - \sum_{j=1}^m \frac{\langle\hat{K}_{ZX},\hat{\varphi}_j\rangle\langle X_i,\hat{\varphi}_j\rangle}{\hat{\nu}_j}\Big)Y_i,\\
L_{12} &= \frac{1}{\sqrt{n}}\sum_{i=1}^n\Big(Z_i - \sum_{j=1}^m \frac{\langle\hat{K}_{ZX},\hat{\varphi}_j\rangle\langle X_i,\hat{\varphi}_j\rangle}{\hat{\nu}_j}\Big)Y_i\frac{\hat{g}_Y(v_i)}{g_Y(v_i)},\\
L_{13} &= \frac{1}{\sqrt{n}}\sum_{i=1}^n\Big(Z_i - \sum_{j=1}^m \frac{\langle\hat{K}_{ZX},\hat{\varphi}_j\rangle\langle X_i,\hat{\varphi}_j\rangle}{\hat{\nu}_j}\Big)Y_i\frac{\hat{p}(v_i)}{p(v_i)}.
\end{aligned}$$

Noting that $\psi(v) = \frac{g_Y(v)}{p(v)EY}$, we have

$$\begin{aligned}
L_1 =& \frac{1}{\sqrt{n}}\sum_{i=1}^n\Big(Z_i - \sum_{j=1}^m \frac{\langle\hat{K}_{ZX},\hat{\varphi}_j\rangle\langle X_i,\hat{\varphi}_j\rangle}{\hat{\nu}_j}\Big)Y_i\frac{g_Y(V_i)}{p(V_i)}\frac{\hat{p}(V_i)}{\hat{g}_Y(V_i)}\\
=& \frac{1}{\sqrt{n}}\sum_{i=1}^n\Big(Z_i - \sum_{j=1}^m \frac{\langle\hat{K}_{ZX},\hat{\varphi}_j\rangle\langle X_i,\hat{\varphi}_j\rangle}{\hat{\nu}_j}\Big)Y_i\frac{g_Y(V_i)}{p(V_i)}\frac{p(V_i)}{g_Y(V_i)}\Big[1 - \frac{\hat{g}_Y(V_i)}{g_Y(V_i)} + \frac{\hat{p}(V_i)}{p(V_i)}\Big]\\
&+ L_1^{R_1} - L_1^{R_2}\\
=& L_{11} - L_{12} + L_{13} + \cdots + L_1^{R_1} - L_1^{R_2},
\end{aligned}$$

where

$$\begin{aligned}
L_1^{R_1} &= \frac{1}{\sqrt{n}}\sum_{i=1}^n\Big(Z_i - \sum_{j=1}^m \frac{\langle\hat{K}_{ZX},\hat{\varphi}_j\rangle\langle X_i,\hat{\varphi}_j\rangle}{\hat{\nu}_j}\Big)Y_i\frac{[\hat{g}_Y(V_i) - g_Y(V_i)]^2}{g_Y(V_i)\hat{g}_Y(V_i)},\\
L_1^{R_2} &= \frac{1}{\sqrt{n}}\sum_{i=1}^n\Big(Z_i - \sum_{j=1}^m \frac{\langle\hat{K}_{ZX},\hat{\varphi}_j\rangle\langle X_i,\hat{\varphi}_j\rangle}{\hat{\nu}_j}\Big)Y_i\\
&\quad\times\frac{[\hat{g}_Y(V_i) - g_Y(V_i)][\hat{p}(V_i) - p(V_i)]}{p(V_i)\hat{g}_Y(V_i)}.
\end{aligned}$$

Invoking Theorem 2.1.8 of [18], Lemma 3 of [31], conditions (C11)-(C12) and Law of Large Numbers (LLN) for $\frac{1}{\sqrt{n}}\sum_{i=1}^n\Big(Z_i - \sum_{j=1}^m \frac{\langle\hat{K}_{ZX},\hat{\varphi}_j\rangle\langle X_i,\hat{\varphi}_j\rangle}{\hat{\nu}_j}\Big)Y_i$, we have

$$\begin{aligned}
L_1^{R_1} =& O_p\Big[\Big(h^4 + n^{-1/2}h^{-1}\log n\Big)^2\\
&\times\frac{1}{\sqrt{n}}\sum_{i=1}^n\Big(Z_i - \sum_{j=1}^m \frac{\langle\hat{K}_{ZX},\hat{\varphi}_j\rangle\langle X_i,\hat{\varphi}_j\rangle}{\hat{\nu}_j}\Big)Y_i\Big]\\
=& o_p(1).
\end{aligned}$$

Similar to $L_1^{R_1}$, it can yield that $L_1^{R_2} = o_p(1)$.

Step 2. Show that

$$L_{12} = \frac{1}{2}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Big(Z_i - \sum_{j=1}^{m}\frac{\langle \hat{K}_{ZX}, \hat{\varphi}_j\rangle\langle X_i, \hat{\varphi}_j\rangle}{\hat{\nu}_j}\Big)Y_i + \frac{1}{2}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} Y_i \frac{E[Z^\top(I-S_m)Y]}{EY} + o_p(1), \tag{20}$$

$$L_{13} = \frac{1}{2}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Big(Z_i - \sum_{j=1}^{m}\frac{\langle \hat{K}_{ZX}, \hat{\varphi}_j\rangle\langle X_i, \hat{\varphi}_j\rangle}{\hat{\nu}_j}\Big)Y_i + \frac{1}{2}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} E[Z^\top(I-S_m)Y] + o_p(1). \tag{21}$$

For L_{12}, applying the similar arguments used by [3], we have the following asymptotic representation:

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Big(Z_i - \sum_{j=1}^{m}\frac{\langle \hat{K}_{ZX}, \hat{\varphi}_j\rangle\langle X_i, \hat{\varphi}_j\rangle}{\hat{\nu}_j}\Big)Y_i\frac{\hat{g}_Y(V_i)}{g_Y(V_i)}\\
=&\frac{1}{2}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{1}{h_0}\int\Big(Z_i - \sum_{j=1}^{m}\frac{\langle \hat{K}_{ZX}, \hat{\varphi}_j\rangle\langle X_i, \hat{\varphi}_j\rangle}{\hat{\nu}_j}\Big)Y_i\frac{1}{g_Y(V_i)}\\
&\times K\Big(\frac{V_i - v}{h_0}\Big)y\psi(v)p_{Y,V}(y,v)dydv\\
&+\frac{1}{2}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{1}{h_0}\int Y_i\psi(V_i)\frac{1}{g_Y(v)}K\Big(\frac{V_i - v}{h_0}\Big)tp_{T,V}(t,v)dtdv\\
&+o_p(1),
\end{aligned}$$

where $p_{Y,V}(y,v)$ is the density function of (Y,V), $T = Z^T(I-S_m)Y$ and $p_{T,V}(t,v)$ is the density of (T,V). Note that Y and V are independent, T and U are independent. By conditions (C10)-(C12), we have

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Big(Z_i - \sum_{j=1}^{m}\frac{\langle \hat{K}_{ZX}, \hat{\varphi}_j\rangle\langle X_i, \hat{\varphi}_j\rangle}{\hat{\nu}_j}\Big)Y_i\frac{\hat{g}_Y(V_i)}{g_Y(V_i)}\\
=&\frac{1}{2}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Big(Z_i - \sum_{j=1}^{m}\frac{\langle \hat{K}_{ZX}, \hat{\varphi}_j\rangle\langle X_i, \hat{\varphi}_j\rangle}{\hat{\nu}_j}\Big)Y_i\frac{EY}{g_Y(V_i)}\\
&\times\frac{1}{h_0}\int K\Big(\frac{V_i - v}{h_0}\Big)\psi(v)p_V(v)dv\\
&+\frac{1}{2}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}Y_i\psi(V_i)E[Z^\top(I-S_m)Y]\\
&\times\frac{1}{h_0}\int\frac{p_V(v)}{g_Y(v)}K\Big(\frac{V_i - v}{h_0}\Big)dv\\
&+o_p(1).
\end{aligned}$$

The result of L_{13} can be similarly proved. Combining (19) with (20) and (21), we have

$$\begin{aligned}
L_1 =&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Big\{\Big(Z_i - \sum_{j=1}^{m}\frac{\langle \hat{K}_{ZX}, \hat{\varphi}_j\rangle\langle X_i, \hat{\varphi}_j\rangle}{\hat{\nu}_j}\Big)Y_i\\
&-E[Z^\top(I-S_m)Y]\Big\}\\
&-\frac{1}{2}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(Y_i - EY)\frac{E[Z^\top(I-S_m)Y]}{EY}\\
&+\sqrt{n}E[Z^\top(I-S_m)Y] + o_p(1).
\end{aligned}\tag{22}$$

Step 3. Show that

$$L_2 = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\widetilde{Y}_i - EY)\frac{E[Z^\top(I-S_m)Y]}{EY} + o_p(1). \tag{23}$$

From (23) and the definition of L_2, we have

$$\begin{aligned}
L_2 =&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\bar{\widetilde{Y}} - EY}{EY}L_1\\
=&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Big\{\Big(Z_i - \sum_{j=1}^{m}\frac{\langle \hat{K}_{ZX}, \hat{\varphi}_j\rangle\langle X_i, \hat{\varphi}_j\rangle}{\hat{\nu}_j}\Big)Y_i\\
&-E[Z^\top(I-S_m)Y]\Big\}\frac{\bar{\widetilde{Y}} - EY}{EY}\\
&-\frac{1}{2}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(Y_i - EY)\frac{E[Z^\top(I-S_m)Y]}{EY}\frac{\bar{\widetilde{Y}} - EY}{EY}\\
&+\sqrt{n}E[Z^\top(I-S_m)Y]\frac{\bar{\widetilde{Y}} - EY}{EY} + o_p(1)\\
=&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\widetilde{Y}_i - EY)\frac{E[Z^\top(I-S_m)Y]}{EY} + o_p(1),
\end{aligned}$$

where the last equality is obtained by applying LLN to $\frac{\bar{\widetilde{Y}} - EY}{EY}$, $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(Y_i - EY)$ and $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Big[\Big(Z_i - \sum_{j=1}^{m}\frac{\langle \hat{K}_{ZX}, \hat{\varphi}_j\rangle\langle X_i, \hat{\varphi}_j\rangle}{\hat{\nu}_j}\Big)Y_i - E[Z^\top(I-S_m)Y]\Big]$. Finally, by (22) and (23), the desired result is easy to arrive at.

Proof of (iii). Similar to the proof of (ii), we can get it. ■

Lemma A.2. *Suppose that (C1)-(C12) hold, we have*

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\hat{Z}^\top(I-S_m)\hat{Y} - \frac{1}{\sqrt{n}}Z^\top(I-S_m)Y\\
=&\frac{1}{\sqrt{n}}\Big[\hat{Z}^\top(I-S_m) - Z^\top(I-S_m)\Big]Y\\
&+\frac{1}{\sqrt{n}}Z^\top(I-S_m)(\hat{Y} - Y) + o_p(1)
\end{aligned}\tag{24}$$

and

$$\begin{aligned}
&\Big[\frac{1}{n}\hat{Z}^\top(I-S_m)\hat{Z}\Big]^{-1} - \Big[\frac{1}{n}Z^\top(I-S_m)Z\Big]^{-1}\\
=&-\Big[\frac{1}{n}Z^\top(I-S_m)Z\Big]^{-1}\Big\{\frac{1}{n}[\hat{Z}^\top(I-S_m) - Z^\top(I-S_m)]Z\Big\}\\
&\times\Big[\frac{1}{n}Z^\top(I-S_m)Z\Big]^{-1}\\
&-\Big[\frac{1}{n}Z^\top(I-S_m)Z\Big]^{-1}\Big\{\frac{1}{n}Z^\top(I-S_m)[\hat{Z} - Z]\Big\}
\end{aligned}\tag{25}$$

$$\times \left[\frac{1}{n}Z^\top(I-S_m)Z\right]^{-1} + o_p\left(n^{-\frac{1}{2}}\right).$$

Proof of Lemma A.2. For (24), by Lemma A.1 (i), we have

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\hat{Z}^\top(I-S_m)\hat{Y} - \frac{1}{\sqrt{n}}Z^\top(I-S_m)Y \\
=&\frac{1}{\sqrt{n}}\left[\hat{Z}^\top(I-S_m) - Z^\top(I-S_m)\right]Y + \frac{1}{\sqrt{n}}Z^\top(I-S_m)(\hat{Y}-Y) \\
&+ \frac{1}{\sqrt{n}}\left[\hat{Z}^\top(I-S_m) - Z^\top(I-S_m)\right](\hat{Y}-Y) \\
=&\frac{1}{\sqrt{n}}\left[\hat{Z}^\top(I-S_m) - Z^\top(I-S_m)\right]Y + \frac{1}{\sqrt{n}}Z^\top(I-S_m)(\hat{Y}-Y) \\
&+ o_p(1).
\end{aligned}$$

For (25), by simple calculation, we have

$$\begin{aligned}
&\left[\frac{1}{n}\hat{Z}^\top(I-S_m)\hat{Z}\right]^{-1} - \left[\frac{1}{n}Z^\top(I-S_m)Z\right]^{-1} \\
=&-\left[\frac{1}{n}Z^\top(I-S_m)Z\right]^{-1}\left\{\frac{1}{n}[\hat{Z}^\top(I-S_m) - Z^\top(I-S_m)]Z\right\} \\
&\times \left[\frac{1}{n}Z^\top(I-S_m)Z\right]^{-1} \\
&- \left[\frac{1}{n}Z^\top(I-S_m)Z\right]^{-1}\left[\frac{1}{n}Z^\top(I-S_m)(\hat{Z}-Z)\right] \\
&\times \left[\frac{1}{n}Z^\top(I-S_m)Z\right]^{-1} \\
&- \left[\frac{1}{n}Z^\top(I-S_m)Z\right]^{-1}\Big\{\frac{1}{n}[\hat{Z}^\top(I-S_m) \\
&- Z^\top(I-S_m)](\hat{Z}-Z)\Big\}\left[\frac{1}{n}Z^\top(I-S_m)Z\right]^{-1} \\
&- \left\{\left[\frac{1}{n}\hat{Z}^\top(I-S_m)\hat{Z}\right]^{-1} - \left[\frac{1}{n}Z^\top(I-S_m)Z\right]^{-1}\right\} \\
&\times \left[\frac{1}{n}\hat{Z}^\top(I-S_m)\hat{Z} - \frac{1}{n}Z^\top(I-S_m)Z\right]\left[\frac{1}{n}Z^\top(I-S_m)Z\right]^{-1}.
\end{aligned}$$

[24] has shown that $\frac{1}{n}Z^\top(I-S_m)Z = \hat{B} \xrightarrow{P} B$. Similar to the proof of Lemma A.1 (i), we can get that $\frac{1}{n}[\hat{Z}^\top(I-S_m) - Z^\top(I-S_m)](\hat{Z}-Z) = o_p(n^{-\frac{1}{2}})$. Therefore, the third term has the same order. Similar to (24), it is easy to see that $\frac{1}{n}\hat{Z}^\top(I-S_m)\hat{Z} - \frac{1}{n}Z^\top(I-S_m)Z = O_p(n^{-\frac{1}{2}})$, then the third term also has the order of $o_p(n^{-\frac{1}{2}})$. ∎

Lemma A.3. *Suppose that (C1)-(C12) hold, we have*

$$\frac{1}{n}\hat{\eta}_n(\beta) = \frac{1}{n}R_n(\beta) + o_p(n^{-1/2}) \tag{26}$$

and

$$\frac{1}{n}\hat{\eta}_n(\beta)\hat{\eta}_n^\top(\beta) = \frac{1}{n}R_n(\beta)R_n^\top(\beta) + o_p(1), \tag{27}$$

where $\hat{\eta}_n(\beta) = \sum_{i=1}^n \hat{Z}_i[\hat{Y}_i - \beta^\top\hat{Z}_i - (U_i^*)^\top\hat{\gamma}_{\hat{Z}}]$ *and*

$$\begin{aligned}
&R_n(\beta) \\
=&\sum_{i=1}^n Z_i[Y_i - \beta^\top Z_i - (U_i^*)^\top\hat{\gamma}_Z] \\
&+ \frac{1}{2}\Big(\sum_{i=1}^n \frac{Y_i - EY}{EY}\Big)B\beta + \Big(\sum_{i=1}^n \frac{\widetilde{Y}_i - Y_i}{EY}\Big)B\beta \\
&- \frac{1}{2}B\mathrm{diag}\Big\{\sum_{i=1}^n \frac{Z_{i1} - EZ_1}{EZ_1}, \cdots, \frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{Z_{ip} - EZ_p}{EZ_p}\Big\}\beta \\
&- B\mathrm{diag}\Big\{\sum_{i=1}^n \frac{\widetilde{Z}_{i1} - Z_{i1}}{EZ_1}, \cdots, \sum_{i=1}^n \frac{\widetilde{Z}_{ip} - Z_{ip}}{EZ_p}\Big\}\beta.
\end{aligned}$$

Proof of Lemma A.3. Proof of (26). By simple calculation, we have

$$\begin{aligned}
\hat{\eta}_n(\beta) =& \sum_{i=1}^n \Big(Z_i - \sum_{j=1}^m \frac{\langle \hat{K}_{ZX}, \hat{\varphi}_j\rangle\langle X_i, \hat{\varphi}_j\rangle}{\hat{\nu}_j}\Big)(Y_i - Z_i\beta) \\
&+ \sum_{i=1}^n \Big(Z_i - \sum_{j=1}^m \frac{\langle \hat{K}_{ZX}, \hat{\varphi}_j\rangle\langle X_i, \hat{\varphi}_j\rangle}{\hat{\nu}_j}\Big)(\hat{Y}_i - Y_i) \\
&- \sum_{i=1}^n \Big(Z_i - \sum_{j=1}^m \frac{\langle \hat{K}_{ZX}, \hat{\varphi}_j\rangle\langle X_i, \hat{\varphi}_j\rangle}{\hat{\nu}_j}\Big)(\hat{Z}_i\beta - Z_i\beta) \\
&+ \sum_{i=1}^n \Big(\hat{Z}_i - Z_i - \sum_{j=1}^m \frac{\langle \hat{K}_{\hat{Z}X} - \hat{K}_{ZX}, \hat{\varphi}_j\rangle\langle X_i, \hat{\varphi}_j\rangle}{\hat{\nu}_j}\Big)(\hat{Y}_i - Y_i) \\
&+ \sum_{i=1}^n \Big(\hat{Z}_i - Z_i - \sum_{j=1}^m \frac{\langle \hat{K}_{\hat{Z}X} - \hat{K}_{ZX}, \hat{\varphi}_j\rangle\langle X_i, \hat{\varphi}_j\rangle}{\hat{\nu}_j}\Big)Y_i \\
&- \sum_{i=1}^n \Big(\hat{Z}_i - Z_i - \sum_{j=1}^m \frac{\langle \hat{K}_{\hat{Z}X} - \hat{K}_{ZX}, \hat{\varphi}_j\rangle\langle X_i, \hat{\varphi}_j\rangle}{\hat{\nu}_j}\Big)(\hat{Z}_i - Z_i) \\
&- \sum_{i=1}^n \Big(\hat{Z}_i - Z_i - \sum_{j=1}^m \frac{\langle \hat{K}_{\hat{Z}X} - \hat{K}_{ZX}, \hat{\varphi}_j\rangle\langle X_i, \hat{\varphi}_j\rangle}{\hat{\nu}_j}\Big)(\hat{Z}_i - Z_i).
\end{aligned}$$

For D_1, applying Lemma A.1 (ii)-(iii), we have

$$\begin{aligned}
D_1 =& \sum_{i=1}^n Z_i[Y_i - \beta^\top Z_i - (U_i^*)^\top\hat{\gamma}_Z] \\
&+ \frac{1}{2}\Big(\sum_{i=1}^n \frac{Y_i - EY}{EY}\Big)B\beta + \Big(\sum_{i=1}^n \frac{\widetilde{Y}_i - Y_i}{EY}\Big)B\beta \\
&- \frac{1}{2}B\mathrm{diag}\Big\{\sum_{i=1}^n \frac{Z_{i1} - EZ_1}{EZ_1}, \cdots, \frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{Z_{ip} - EZ_p}{EZ_p}\Big\}\beta \\
&- B\mathrm{diag}\Big\{\sum_{i=1}^n \frac{\widetilde{Z}_{i1} - Z_{i1}}{EZ_1}, \cdots, \sum_{i=1}^n \frac{\widetilde{Z}_{ip} - Z_{ip}}{EZ_p}\Big\}\beta + o_p\Big(n^{1/2}\Big).
\end{aligned}$$

For D_2, similar to proof Lemma A.1 (ii)-(iii), and invoking Lemma A.1 (i), we can get $D_2 = o_p(n^{1/2})$. Then, the desired result can be obtained.

Proof of (27). Note that

$$\frac{1}{n}\hat{\eta}_n(\beta)\hat{\eta}_n^\top(\beta) = \frac{1}{n}R_n(\beta)R_n^\top(\beta) + \frac{1}{n}\Big[\hat{\eta}_n(\beta) - R_n(\beta)\Big]R_n^\top(\beta) + \frac{1}{n}R_n(\beta)\Big[\hat{\eta}_n(\beta) - R_n(\beta)\Big]^\top + \frac{1}{n}\Big[\hat{\eta}_n(\beta) - R_n(\beta)\Big]\Big[\hat{\eta}_n(\beta) - R_n(\beta))\Big]^\top.$$

Together with (26) and $R_n(\beta) = O_p(n^{1/2})$, the proof can be completed. ■

Proof of Theorem 2.1. By the definition of $\hat{\beta}$ and β, we know that

$$\begin{aligned}
\hat{\beta} =& [\hat{Z}^\top(I - S_m)\hat{Z}]^{-1}\hat{Z}^\top(I - S_m)\hat{Y} \\
=& \Big\{\Big[\frac{1}{n}\hat{Z}^\top(I - S_m)\hat{Z}\Big]^{-1} - \Big[\frac{1}{n}Z^\top(I - S_m)Z\Big]^{-1}\Big\} \\
&\times \frac{1}{n}\Big[\hat{Z}^\top(I - S_m) - Z^\top(I - S_m)\Big](\hat{Y} - Y) \\
&+ \Big\{\Big[\frac{1}{n}\hat{Z}^\top(I - S_m)\hat{Z}\Big]^{-1} - \Big[\frac{1}{n}Z^\top(I - S_m)Z\Big]^{-1}\Big\} \\
&\times \frac{1}{n}Z^\top(I - S_m)(\hat{Y} - Y) \\
&+ \Big\{\Big[\frac{1}{n}\hat{Z}^\top(I - S_m)\hat{Z}\Big]^{-1} - \Big[\frac{1}{n}Z^\top(I - S_m)Z\Big]^{-1}\Big\} \\
&\times \frac{1}{n}\Big[\hat{Z}^\top(I - S_m) - Z^\top(I - S_m)\Big]Y \\
&+ \Big\{\Big[\frac{1}{n}\hat{Z}^\top(I - S_m)\hat{Z}\Big]^{-1} - \Big[\frac{1}{n}Z^\top(I - S_m)Z\Big]^{-1}\Big\} \\
&\times \frac{1}{n}Z^\top(I - S_m)Y \\
&+ \Big[\frac{1}{n}Z^\top(I - S_m)Z\Big]^{-1}\frac{1}{n}\Big[\hat{Z}^\top(I - S_m) - Z^\top(I - S_m)\Big](\hat{Y} - Y) \\
&+ \Big[\frac{1}{n}Z^\top(I - S_m)Z\Big]^{-1}\frac{1}{n}Z^\top(I - S_m)(\hat{Y} - Y) \\
&+ \Big[\frac{1}{n}Z^\top(I - S_m)Z\Big]^{-1}\frac{1}{n}\Big[\hat{Z}^\top(I - S_m) - Z^\top(I - S_m)\Big]Y \\
&+ \Big[\frac{1}{n}Z^\top(I - S_m)Z\Big]^{-1}\frac{1}{n}Z^\top(I - S_m)Y,
\end{aligned}$$

and

$$\begin{aligned}
\beta =& \Big[\frac{1}{n}Z^\top(I - S_m)Z\Big]^{-1}\frac{1}{n}Z^\top(I - S_m)Z\beta \\
=& \Big[\frac{1}{n}Z^\top(I - S_m)Z\Big]^{-1}\frac{1}{n}Z^\top(I - S_m)[Y - \langle\gamma(t), X(t)\rangle - \varepsilon] \\
=& \Big[\frac{1}{n}Z^\top(I - S_m)Z\Big]^{-1}\frac{1}{n}Z^\top(I - S_m)Y \\
&- \Big[\frac{1}{n}Z^\top(I - S_m)Z\Big]^{-1}\frac{1}{n}Z^\top(I - S_m)[\langle\gamma(t), X(t)\rangle + \varepsilon].
\end{aligned}$$

From Theorem 3.1 in [24], Lemma A.1 and Lemma A.2, it follows that

$$\begin{aligned}
&\sqrt{n}(\hat{\beta} - \beta) \\
=& \Big[\frac{1}{n}Z^\top(I - S_m)Z\Big]^{-1}\frac{1}{n}Z^\top(I - S_m)[\langle\gamma(t), X(t)\rangle + \varepsilon] \\
&+ \Big[\frac{1}{n}Z^\top(I - S_m)Z\Big]^{-1}\frac{1}{n}[\hat{Z}^\top(I - S_m) - Z^\top(I - S_m)]Y \\
&+ \Big[\frac{1}{n}Z^\top(I - S_m)Z\Big]^{-1}\frac{1}{n}Z^\top(I - S_m)(\hat{Y} - Y) \\
&+ \Big\{\Big[\frac{1}{n}\hat{Z}^\top(I - S_m)\hat{Z}\Big]^{-1} - \Big[\frac{1}{n}Z^\top(I - S_m)Z\Big]^{-1}\Big\} \\
&\times \frac{1}{n}Z^\top(I - S_m)Y + o_p(1) \\
=& \Big[\frac{1}{n}Z^\top(I - S_m)Z\Big]^{-1}\frac{1}{n}Z^\top(I - S_m)[\langle\gamma(t), X(t)\rangle + \varepsilon] \\
&+ \Big[\frac{1}{n}Z^\top(I - S_m)Z\Big]^{-1}\frac{1}{n}Z^\top(I - S_m)(\hat{Y} - Y) \\
&- \Big[\frac{1}{n}Z^\top(I - S_m)Z\Big]^{-1}\frac{1}{n}[\hat{Z}^\top(I - S_m) - Z^\top(I - S_m)]\beta + o_p(1).
\end{aligned}$$

Obviously, Theorem 2.1 holds. ■

Proof of Theorem 2.2. By the definition of $\hat{\gamma}$ and γ, we know that

$$\begin{aligned}
&\|\hat{\gamma} - \gamma\|^2 \\
=& \Big\|\sum_{j=1}^m \hat{\gamma}_j\hat{\varphi}_j - \sum_{j=1}^\infty \gamma_j\varphi_j\Big\|^2 \\
\le& 2\Big\|\sum_{j=1}^m \hat{\gamma}_j\hat{\varphi}_j - \sum_{j=1}^m \gamma_j\varphi_j\Big\|^2 + 2\Big\|\sum_{j=m+1}^\infty \gamma_j\varphi_j\Big\|^2 \\
\le& 4\Big\|\sum_{j=1}^m (\hat{\gamma}_j - \gamma_j)\hat{\varphi}_j\Big\|^2 + 4\Big\|\sum_{j=1}^m \gamma_j(\hat{\varphi}_j - \varphi_j)\Big\|^2 + 2\sum_{j=m+1}^\infty \gamma_j^2 \\
\le& 4\sum_{j=1}^m (\hat{\gamma}_j - \gamma_j)^2 + 8m\sum_{j=1}^m \gamma_j^2\|\hat{\varphi}_j - \varphi_j\|^2 + 2\sum_{j=m+1}^\infty \gamma_j^2 \\
\triangleq& 4A_1 + 8A_2 + 2A_3.
\end{aligned}$$

[24] has shown that $A_2 = O_p(n^{-1}m) = o_p\Big(n^{-(2b-1)/(a+2b)}\Big)$ and $A_3 = O(m^{-(2b-1)}) = O(n^{-(2b-1)/(a+2b)})$. Following the argument in the proof of Theorem 3.2 in [24], we can get $A_1 = O_p(n^{-(2b-1)/(a+2b)})$. Therefore, we can complete the proof of Theorem 2.2. ■

Proof of Theorem 3.1. Let $\hat{\gamma}_{\hat{Z}} = (U_m^\top U_m)^{-1}U_m(\hat{Y} - \hat{Z}\beta)$ and $\hat{\gamma}_Z = (U_m^\top U_m)^{-1}U_m(Y - Z\beta)$. By simple calculation, we have

$$\begin{aligned}
\hat{\eta}_{ni}(\beta) =& \hat{Z}_i[\hat{Y}_i - \beta^\top\hat{Z}_i - (U_i^*)^\top\hat{\gamma}_{\hat{Z}}] \\
=& \Big(\hat{Z}_i - \sum_{j=1}^m \frac{\langle\hat{K}_{\hat{Z}X}, \hat{\varphi}_j\rangle\langle X_i, \hat{\varphi}_j\rangle}{\hat{\nu}_j}\Big)(\hat{Y}_i - \hat{Z}_i\beta) \\
=& \Big(Z_i - \sum_{j=1}^m \frac{\langle\hat{K}_{ZX}, \hat{\varphi}_j\rangle\langle X_i, \hat{\varphi}_j\rangle}{\hat{\nu}_j}\Big)(Y_i - Z_i\beta) \\
&+ \Big(Z_i - \sum_{j=1}^m \frac{\langle\hat{K}_{ZX}, \hat{\varphi}_j\rangle\langle X_i, \hat{\varphi}_j\rangle}{\hat{\nu}_j}\Big)(\hat{Y}_i - Y_i)
\end{aligned}$$

$$+\Big(\hat{Z}_i - Z_i - \sum_{j=1}^{m} \frac{\langle \hat{K}_{\hat{Z}X} - \hat{K}_{ZX}, \hat{\varphi}_j\rangle\langle X_i, \hat{\varphi}_j\rangle}{\hat{\nu}_j}\Big)(\hat{Y}_i - \hat{Z}_i\beta)$$

$$- \Big(Z_i - \sum_{j=1}^{m} \frac{\langle \hat{K}_{ZX}, \hat{\varphi}_j\rangle\langle X_i, \hat{\varphi}_j\rangle}{\hat{\nu}_j}\Big)(\hat{Z}_i\beta - Z_i\beta)$$

$$\triangleq Q_{i1} + Q_{i2} + Q_{i3} - Q_{i4}.$$

Firstly, we show that

$$\max_{1\le i\le n} \|\hat{\eta}_{ni}(\beta)\| = o_p(n^{1/2}). \tag{28}$$

For Q_{i1}, we have

$$\begin{aligned}
Q_{i1} =& Z_i[Y_i - \beta^\top Z_i - (U_i^*)^\top \hat{\gamma}_Z] \\
=& \Big(Z_i - \sum_{j=1}^{m} \frac{\langle \hat{K}_{ZX}, \hat{\varphi}_j\rangle\langle X_i, \hat{\varphi}_j\rangle}{\hat{\nu}_j}\Big)(\langle \gamma, X_i\rangle + \varepsilon_i) \\
=& \Big(Z_i - \sum_{j=1}^{m} \frac{\langle \hat{K}_{ZX}, \hat{\varphi}_j\rangle\langle X_i, \hat{\varphi}_j\rangle}{\hat{\nu}_j}\Big)\langle \gamma, X_i\rangle \\
&+ \Big(Z_i - \sum_{j=1}^{\infty} \frac{\langle \hat{K}_{ZX}, \hat{\varphi}_j\rangle\langle X_i, \hat{\varphi}_j\rangle}{\hat{\nu}_j}\Big)\varepsilon_i \\
&+ \Big(\sum_{j=1}^{\infty} \frac{\langle \hat{K}_{ZX}, \hat{\varphi}_j\rangle\langle X_i, \hat{\varphi}_j\rangle}{\hat{\nu}_j} - \sum_{j=1}^{m} \frac{\langle \hat{K}_{ZX}, \hat{\varphi}_j\rangle\langle X_i, \hat{\varphi}_j\rangle}{\hat{\nu}_j}\Big)\varepsilon_i \\
\triangleq& Q_{i11} + Q_{i12} + Q_{i13}.
\end{aligned}$$

By the argument in [24], we have $\|Q_{i11}\| = o_p(n^{-1/2})$ and $\|Q_{i12}\| = o_p(n^{-1/2})$. Noting that $E(Q_{i13}) = 0$ and $Var(Q_{i13}) = \sigma^2 B < \infty$, then invoking Lemma 3 of [16], we have $\|Q_{i13}\| = o_p(n^{1/2})$. Therefore, $\max_{1\le i\le n} \|E_{i1}\| = o_p(n^{1/2})$. By Lemma A.1, we have $E_{i2} = o_p(n^{1/2})$, $E_{i3} = o_p(n^{1/2})$ and $E_{i4} = o_p(n^{1/2})$.

Secondly, applying the results of Lemma A.3 and following the same arguments as were used in the proof of expression (2.14) in [16], we have

$$\lambda = O_p(n^{-1/2}). \tag{29}$$

Finally, we will finish the proof of the Theorem 3.1.

Applying Taylor expansion to (9), and invoking (28), (29) and Lemma A.3, we can verify that

$$\hat{l}(\beta) = 2\sum_{i=1}^{n} \Big\{\lambda^\top \hat{\eta}_{ni}(\beta) - \frac{1}{2}[\lambda^\top \hat{\eta}_{ni}(\beta)]^2\Big\} + o_p(1). \tag{30}$$

By (10), it follows that

$$\begin{aligned}
0 =& \frac{1}{n}\sum_{i=1}^{n} \frac{\hat{\eta}_{ni}(\beta)}{1 + \lambda^\top \hat{\eta}_{ni}(\beta)} \\
=& \frac{1}{n}\sum_{i=1}^{n} \hat{\eta}_{ni}(\beta) - \frac{1}{n}\sum_{i=1}^{n} \hat{\eta}_{ni}(\beta)\hat{\eta}_{ni}^\top(\beta)\lambda \\
&+ \frac{1}{n}\sum_{i=1}^{n} \frac{\hat{\eta}_{ni}(\beta)[\lambda^\top \hat{\eta}_{ni}(\beta)]^2}{1 + \lambda^\top \hat{\eta}_{ni}(\beta)}.
\end{aligned}$$

The application of (28), (29) and Lemma A.3 yields

$$\lambda = \Big[\sum_{i=1}^{n} \hat{\eta}_{ni}(\beta)\hat{\eta}_{ni}^\top(\beta)\Big]^{-1} \sum_{i=1}^{n} \hat{\eta}_{ni}(\beta) + o_p(n^{-1/2}). \tag{31}$$

By (30) and (31), we have

$$\hat{l}(\beta) = n\Big[\frac{1}{n}\sum_{i=1}^{n} \hat{\eta}_{ni}(\beta)\Big]^\top \Big[\frac{1}{n}\sum_{i=1}^{n} \hat{\eta}_{ni}(\beta)\hat{\eta}_{ni}^\top(\beta)\Big]^{-1}\Big[\frac{1}{n}\sum_{i=1}^{n} \hat{\eta}_{ni}(\beta)\Big] + o_p(1). \tag{32}$$

From Lemma A.3 and (32), we have

$$\hat{l}(\beta) = R_n^\top(\beta)[R_n(\beta)R_n^\top(\beta)]^{-1}R_n(\beta) + o_p(1).$$

Since $R_n(\beta)$ is a sum of independent and identically distributed random variables, the result of Theorem 3.1 is immediately achieved by central limit theorems. ■

Proof of Theorem 3.2. Following [7], we can see it suffices to show that (13) holds. Let $\breve{\gamma}_j = \int_0^1 \gamma(t)\hat{\varphi}_j(t)dt$, and $\breve{\gamma} = (\breve{\gamma}_1, ..., \breve{\gamma}_m)^\top$, then we have

$$\begin{aligned}
\hat{\gamma} - \breve{\gamma} =& (U_m^\top U_m)^{-1}U_m^\top(\hat{Y} - \hat{Z}\hat{\beta} - U_m\breve{\gamma}) \\
=& (U_m^\top U_m)^{-1}U_m^\top(Y - Z\hat{\beta} - U_m\breve{\gamma}) \\
&+ (U_m^\top U_m)^{-1}U_m^\top[\hat{Y} - Y + (\hat{Z} - Z)\hat{\beta}].
\end{aligned}$$

By the proof of Lamma A.1 and the Theorem 7 in [28], we have

$$\|\hat{\gamma} - \breve{\gamma}\|^2 = \|\varepsilon U_m(U_m^\top U_m)^{-1}\hat{\varphi}\|^2 + o_p(n^{-\frac{2b-1}{a+2b}}).$$

Then, by the same discussion in [28], we can prove (13). This completes the proof of Theorem 3.2. ■

ACKNOWLEDGEMENTS

This research was supported by the National Natural Science Foundation of China (Grant Nos. 11471160, 11101114), the National Statistical Science Research Major Program of China (Grant No. 2018LD01), the Fundamental Research Funds for the Central Universities (Grant No. 30920130111015), sponsored by Qing Lan Project and Graduate Research Innovation Project of the Faculty Economics and Management, ECNU(2018FEM-BCKZD004).

Received 15 November 2019

REFERENCES

[1] Cardot, H., Ferraty, F. and Sarda, P. (2003). Spline estimators for the functional linear model. *Statistica Sinica* **13** 571–591. MR1997162

[2] Cardot, H. and Johannes, J. (2010). Thresholding projection estimators in functional linear models. *Journal of Multivariate Analysis* **101** 395–408. MR2564349

[3] Cui, X., Guo, W., Lin, L. and Zhu, L. (2009). Covariate-adjusted nonlinear regression. *The Annals of Statistics* **37** 1839–1870. MR2533473

[4] Cui, X., Zhu, L. and Lin, L. (2007). A direct estimation for covariate-adjusted regression. (Manuscript)

[5] Dai, S. and Huang, Z. (2019). Estimation for varying coefficient partially nonlinear models with distorted measurement errors. *Journal of the Korean Statistical Society* **48** 117–133. MR3926976

[6] Delaigle, A., Hall, P. and Zhou, W. (2016). Nonparametric covariate-adjusted regression. *The Annals of Statistics* **44** 2190–2220. MR3546448

[7] Imaizumi, M. and Kato, K. (2019). A simple method to construct confidence bands in functional linear regression. *Statistica Sinica* **29** 2055–2081. MR3970347

[8] Kaysen, G., Dubin, J., Müller, H., Mitch, W., Rosales, L. and Levin, N. (2003). Relationships among inflammation nutrition and physiologic mechanisms establishing albumin levels in hemodialysis patients. *Kidney Internet* **61** 2240–2249.

[9] Kong, D., Xue, K., Yao, F. and Zhang, H. (2016). Partially functional linear regression in high dimensions. *Biometrika* **103** 147–159. MR3465827

[10] Lee, E. and Park, B. (2012). Sparse estimation in functional linear regression. *Journal of Multivariate Analysis* **105** 1–17. MR2877499

[11] Li, F., Lin, L. and Cui, X. (2010). Covariate-adjusted partially linear regression models. *Communications in Statistics-Theory and Methods* **39** 1054–1074. MR2745361

[12] Li, X., Du, J., Li, G. and Fan, M. (2014). Variable selection for covariate adjusted regression model. *Journal of Systems Science and Complexity* **27** 1227–1246. MR3284348

[13] Liang, H. (2009). Generalized partially linear mixed-effects models incorporating mismeasured covariates. *The Institute of Statistical Mathematics* **61** 27–46. MR2481027

[14] Lin, X. and Carroll, R. (2000). Nonparametric function estimation for clustered data when the predictor is measured without/with error. *Journal of the American Statistical Association* **95** 520–534. MR1803170

[15] Owen, A. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika* **75** 237–249. MR0946049

[16] Owen, A. (1990). Empirical likelihood ratio confidence regions. *The Annals of Statistics* **18** 90–120. MR1041387

[17] Owen, A. (2001). *Empirical Likelihood*. Chapman and Hall, CRC.

[18] Rao, B. (1983). *Nonparametric Functional Estimation*. Academic Press, New York. MR0740865

[19] Ramsay, J. and Silverman, B. (1997). *Functional Data Analysis*. Springer, New York. MR2168993

[20] Sentürk, D. and Müller, H. (2005a). Covariate adjusted regression. *Biometrika* **92** 75–89. MR2158611

[21] Sentürk, D. and Müller, H. (2005b). Covariate adjusted correlation analysis via varying coefficient models. *Scandinavian Journal of Statistics* **32** 365–383. MR2204625

[22] Sentürk, D. and Müller, H. (2006). Inference for covariate adjusted regression via varying coefficient models. *The Annals of Statistics* **34** 654–679. MR2281880

[23] Sentürk, D. and Nguyen, D. (2009). Asymptotic Properties of Covariate-adjusted Regression with Correlated Errors. *Statistics and Probability Letters* **79** 1175–1180. MR2519000

[24] Shin, H. (2009). Partial functional linear regression. *Journal of Statistical Planning and Inference* **139** 3405–3418. MR2549090

[25] Shin, H. and Lee, M. (2012). On prediction rate in partial functional linear regression. *Journal of Multivariate Analysis* **103** 93–106. MR2823711

[26] Tang, Q. and Jin, P. (2019). Estimation and variable selection for partial functional linear regression. *AStA Advances in Statistical Analysis* **103** 475–501. MR4029635

[27] Wang, G., Su, Y. and Shu, L. (2016). One-day-ahead daily power forecasting of photovoltaic systems based on partial functional linear regression models. *Renewable energy* **96** 469–478.

[28] Xu, W., Ding, H., Zhang, R. and Liang, H. (2020). Estimation and inference in partially functional linear regression with multiple functional covariates. *Journal of Statistical Planning and Inference* **209** 44–61. MR4096253

[29] Zhang, D., Lin, X. and Sowers, M. (2007). Assessing the effects of reproductive hormone profiles on bone mineral density using functional two-stage mixed models. *Biometrics* **63** 351–362. MR2370793

[30] Zhu, H., Zhang, R., Yu, Z., Lian, H. and Liu, Y. (2019). Estimation and testing for partially functional linear errors-in-variables models. *Journal of Multivariate Analysis* **170** 296–314. MR3913042

[31] Zhu, L. and Fang, K. (1996). Asymptotics for kernel estimate of sliced inverse regression. *The Annals of Statistics* **24** 1053–1068. MR1401836

Zhiqiang Jiang
School of Science
Nanjing University of Science and Technology
China
E-mail address: zqjiang61@126.com

Zhensheng Huang
School of Science
Nanjing University of Science and Technology
China
E-mail address: stahzs@126.com

Hanbing Zhu
School of Statistics
Key Laboratory of Advanced Theory and Application in Statistics and Data Science-MOE
East China Normal University
China
E-mail address: zhuhbecnu@163.com