Comput Stat (2016) 31:1107–1129
DOI 10.1007/s00180-015-0636-0

ORIGINAL PAPER

Polynomial spline estimation for partial functional linear regression models

Jianjun Zhou[1] · Zhao Chen[2] · Qingyan Peng[1]

Received: 27 June 2014 / Accepted: 9 December 2015 / Published online: 2 January 2016

Abstract Because of its orthogonality, interpretability and best representation, functional principal component analysis approach has been extensively used to estimate the slope function in the functional linear model. However, as a very popular smooth technique in nonparametric/semiparametric regression, polynomial spline method has received little attention in the functional data case. In this paper, we propose the polynomial spline method to estimate a partial functional linear model. Some asymptotic results are established, including asymptotic normality for the parameter vector and the global rate of convergence for the slope function. Finally, we evaluate the performance of our estimation method by some simulation studies.

Keywords Functional data analysis · Polynomial spline · Asymptotic normality · Rates of convergence

1 Introduction

With the development of technology of computation and measurement, scientists usually confront the data providing information about curves, surfaces or anything else varying with continuous variables. Such type of data structure, called functional data, attracted great interests in various fields. For example, in chemometrics the spectrometric data consists of hundreds of different wavelength spectra, fMRI data can recover the contours of invisible human organs and spatial data is used to study the topological,

✉ Jianjun Zhou
jjzhou@ynu.edu.cn

1 School of Mathematics and Statistics, Yunnan University, Kunming 650091, China

2 Department of Statistics, The Pennsylvania State University, Thomas building, University park, PA 16802, USA

geometric, or geographic properties of entities. Due to the infinite dimensionality and the strong mutual relations of predictors, the traditional multivariate statistical methods fail to analyze functional data. To overcome these problems, Ramsay and Dalzell (1991), Ramsay and Silverman (1997, 2005) introduced some fundamental models and tools for functional data analysis.

Regression analysis is very popular in statistical analysis. As an extension of ordinary linear models, Ramsay and Silverman (1997, 2005) introduced the functional linear model to model the relationship between a scalar response and a functional predictor. Further, Cardot et al. (1999), Cai and Hall (2006), Hall and Horowitz (2007) and Li and Hsing (2007) proposed estimation methods based on functional principal component analysis and investigated the asymptotic properties of the estimators. On the other hand, Cardot et al. (2003) and Crambes et al. (2009) employed penalized B-spline and smoothing spline to estimate the functional slope parameter. As an extension of nonparametric model, functional nonparametric regression was also studied in literatures. Kernel regression (Ferraty and Vieu 2006), local linear regression (Baíllo and Grané 2009) and K-nearest neighbours method (Burba et al. 2009) are used to deal with the functional nonparametric models.

In order to improve the power of prediction and interpretation of the functional regression model, some additional real-valued predictors could be introduced. There is some recent literature focusing on this situation. For example, Aneiros-Pérez and Vieu (2006) introduced a semi-functional partial linear regression model to predict the fat content of the chopped pure meat. Further, Aneiros-Pérez and Vieu (2008) extended this model to dependent data. Zhang et al. (2007) introduced the partial functional linear model to assess the effect of women's hormone on the total hip bone mineral density and Shin (2009) proposed a new estimation method based on functional principal component analysis. Cardot and Sarda (2008) generalized the functional linear model to a varying coefficient functional linear model in which an additional random variable influenced smoothly the functional coefficient. Zhou and Chen (2012) introduced a semi-functional linear model which combined the functional linear regression model and the nonparametric regression model.

In functional linear regression, because of its orthogonality, interpretability and best representation, functional principal component analysis approach has been extensively used to estimate the slope function (see Cardot et al. 1999; Cai and Hall 2006; Hall and Horowitz 2007; Li and Hsing 2007; Shin 2009). As a very popular smooth technique, polynomial spline or regression spline method can produce a smooth function estimate and can be operated easily, so it has received considerable attention in nonparametric/semiparametric regression (see Chen 1991; Stone 1994; Stone et al. 1997; Zhou et al. 1998; Huang 2003a, b; Huang et al. 2004a; Huang and Shen 2004b and so on). However, there is limited literature discussed the polynomial spline method in the functional data case. We only noted that Ramsay and Silverman (1997, 2005) applied polynomial spline to estimate the functional linear model, but they didn't investigate the asymptotic behaviors of the estimator.

In this paper, we focus on the polynomial spline estimators for the partial functional linear models. We employ the polynomial spline basis to approximate the functional coefficients. Using profile least squares technique, we obtain the optimal convergence rate and asymptotic normality for estimators of parameters. Based on these estimators, we also have the limiting distribution of Wald test statistic for linear hypothesis of parameters. The numerical studies indicate our proposed procedure can enjoy more smoothness for functional coefficients in finite samples.

The rest of the paper is organized as follows. In Sect. 2, we introduce the polynomial spline estimate for partial functional linear models. Section 3 investigates the asymptotic properties of estimators and discusses statical inference problem. Simulation studies are presented in Sect. 4. Conclusion and further research are given in Sect. 5. All technical details and proofs are given in the "Appendix".

2 Polynomial spline estimation

Let the observed data $(X_i, \mathbf{Z}_i, Y_i), i = 1, \ldots, n$, which are independent and identically distributed (i.i.d.), be generated from the following partial functional linear model

$$Y_i = \int_0^1 X_i(t)\alpha(t)dt + \mathbf{Z}_i^T\beta + \varepsilon_i, \quad i = 1, \ldots, n, \tag{1}$$

where Y_i and $\mathbf{Z}_i = (Z_{i1}, \ldots, Z_{ip})^T$ are the scalar response variable and the p-dimensional predictor vector, respectively. The predictor variable X_i is a random function valued in $H = L^2([0, 1])$, the Hilbert space containing square integrable functions defined on the unit interval. Let $\langle \phi, \varphi \rangle = \int_0^1 \phi(t)\varphi(t)dt$ denote the usual inner product of function ϕ and φ and let $\|\phi\| = \langle \phi, \phi \rangle^{1/2}$ denote the norm of H. The random errors ε_i are independent and identically distributed with mean 0 and finite variance σ^2 and are independent of $(X_i, \mathbf{Z}_i)$. Let β be an unknown p-dimensional parameter vector and $\alpha(t)$ be an unknown smoothing function belonging to H.

Before introducing the polynomial spline estimation, we recall simply the polynomial spline function. Let $k \geq 0$. The sequence $0 = t_0 < t_1 < \cdots < t_{N_n} < t_{N_n+1} = 1$ is a partition of interval [0, 1], which is called knot sequence. Suppose a function is a polynomial of degree k on each of the intervals $[t_i, t_{i+1}](i = 0, 1, \ldots, N_n)$, and it has $k - 1$ continuous derivatives for $k \geq 1$ on the interval [0, 1], then it is called spline function of degree k.

We next consider the polynomial spline estimate $\widehat{\alpha}$ of α. Let S_{k,N_n} be the space of polynomial splines defined on interval [0, 1] with degree k and N_n interior knots. The space S_{k,N_n} is a K_n-dimensional linear space, $K_n = N_n + k + 1$. From Theorem XII.1 of de Boor (2001), we can conclude that, if the slope function $\alpha(t)$ is sufficiently smooth, there is a spline function $a(t) \in S_{k,N_n}$ such that

$$\alpha(t) \approx a(t) = \sum_{s=1}^{K_n} b_s B_s(t), \tag{2}$$

where $B_j, j = 1, \ldots, K_n$ are the B-spline basis functions. Plugging the approximation (2) into model (1), we have

$$Y_i \approx \sum_{s=1}^{K_n} b_s \langle X_i, B_s \rangle + \mathbf{Z}_i^T\beta + \varepsilon_i, \quad i = 1, \ldots, n, \tag{3}$$

where the spline coefficient vector $b = (b_1, \ldots, b_{K_n})^T$ and parameter vector β are to be estimated. Then, the semiparametric estimation problem in model (1) turns into the ordinary parametric estimation problem.

Let the square loss function

$$l(\beta, b) = \sum_{i=1}^{n} \left(Y_i - \mathbf{Z}_i^T \beta - \sum_{s=1}^{K_n} b_s \langle X_i, B_s \rangle \right)^2. \tag{4}$$

The estimators of b and β can be obtained by minimizing (4). For ease to discuss the asymptotic properties, we apply the profile least squares procedure to estimate unknown spline coefficients and parameters. The estimators for β and b are given by

$$\widehat{\beta} = (\mathbf{Z}^T (I - A)\mathbf{Z})^{-1} \mathbf{Z}^T (I - A) Y, \quad \widehat{b} = (B^T B)^{-1} B^T (Y - \mathbf{Z}\widehat{\beta}), \tag{5}$$

where $Y = (Y_1, \ldots, Y_n)^T$, $\mathbf{Z} = (\mathbf{Z}_1, \ldots, \mathbf{Z}_n)^T$, $B = \left\{ \langle X_i, B_j \rangle \right\}_{\substack{i=1,\ldots,n \\ j=1,\ldots,K_n}}$ and $A = B(B^T B)^{-1} B^T$. Then, the polynomial spline estimator of $\alpha(t)$ and the estimator of σ^2 can be respectively defined by

$$\widehat{\alpha}(t) = \sum_{s=1}^{K_n} \widehat{b}_s B_s(t), \qquad \widehat{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^{n} \left(Y_i - \langle X_i, \widehat{\alpha} \rangle - \mathbf{Z}_i^T \widehat{\beta} \right)^2. \tag{6}$$

3 Asymptotic properties

In this section, we investigate the asymptotic properties of the polynomial spline estimators. For ease to discuss the asymptotic behaviors of our proposed estimators, the following notation is needed. For two sequences of positive numbers a_n and b_n, $a_n \lesssim b_n$ signifies a_n/b_n is uniformly bounded and $a_n \asymp b_n$ if $a_n \lesssim b_n$ and $b_n \lesssim a_n$. The covariance operator Γ of the random function X is defined as $\Gamma x(t) = \int_0^1 EX(t)X(s)x(s)ds$, $x \in H$. The norm $\|\cdot\|$ of a function $f \in C^{k+1}([0,1])$ is defined as $\|f\| = \left(\int_0^1 f(t)^2 dt \right)^{1/2}$.

In order to establish the theoretical properties of polynomial spline estimation, the following assumptions are required:

(C1) There are some positive constants M and $\frac{1}{4(k+1)} < r < \frac{1}{2}$ such that

$$h = \max_{j=0,\ldots,N_n} (t_{j+1} - t_j) \asymp n^{-r}, \quad K_n \asymp n^r, \quad h / \min_{j=0,\ldots,N_n} (t_{j+1} - t_j) \leq M.$$

(C2) $E||X||^4 < \infty$ and the eigenvalues of the covariance operator Γ of X are strictly positive.

(C3) $E|Z_{11}|^4 + \cdots + E|Z_{1p}|^4 + E|\varepsilon_1|^4 < \infty$.

(C4) For $j = 1, \ldots, p$, $E(Z_{1j}|X_1)$ is a continuous linear functional, that is, there exists a function $g_j \in H$ such that $E(Z_{1j}|X_1) = \langle X_1, g_j \rangle$. Further, we assume $g_j, j = 1, \ldots, p$ and slope function α are smooth enough, that is, $g_j \in C^{k+1}([0, 1]), \alpha \in C^{k+1}([0, 1])$.

(C5) Let $\eta_{1j} = Z_{1j} - E(Z_{1j}|X_1) = Z_{1j} - \langle X_1, g_j\rangle, j = 1, \ldots, p, \eta_1 = (\eta_{11}, \ldots, \eta_{1p})^T$. Furthermore, we assume that $\Sigma = E\eta_1\eta_1^T$ is a positive definite matrix.

Remark 1 Conditions (C1)–(C5) are very general in polynomial spline estimation and functional linear model. In fact, condition (C1) is similar to (3) in Zhou et al. (1998). For the number of spline basis K_n, the requirement is similar to (16) in Shin (2009). Condition (C2) is very common in functional linear model (see H1 and H2 in Cardot et al. 1999 and (12) in Shin 2009). However, we don't need additional assumption on eigenvalues of the covariance operator Γ like (14) in Shin (2009). Condition (C3) is similar to (11) in Aneiros-Pérez and Vieu (2006) and (17) in Shin (2009). Condition (C4) requires the dependence between the covariate $Z_{1j}, (j = 1, \ldots, p)$ and the random function X_1 is a continuous linear functional, which is a special case of conditional expectation operators $E(X_{ij}|T_i = t)$ in Aneiros-Pérez and Vieu (2006). Furthermore, to assure the validity of the polynomial spline estimation, we need a restricted smooth condition on each functional coefficient g_j and α. Condition (C5) is similar to (12) in Aneiros-Pérez and Vieu (2006) and (20) in Shin (2009).

Under the above assumption conditions, we have the following results.

Theorem 1 *If conditions (C1)–(C5) hold, as $n \to \infty$, we have*

$$\sqrt{n}(\widehat{\beta} - \beta) \xrightarrow{D} N(0, \sigma^2\Sigma^{-1}).$$

Theorem 2 *Suppose that conditions (C1)–(C5) are satisfied, then*

$$\|\widehat{\alpha} - \alpha\|^2 = O_p\Big(\frac{K_n}{n} + K_n^{-2(k+1)}\Big).$$

Remark 2 For the estimation of the parameter vector, Theorem 1 shows that the asymptotic result is similar to Theorem 1(i) in Aneiros-Pérez and Vieu (2006) and Theorem 3.1 in Shin (2009). For the estimation of functional coefficient, Theorem 2 indicates that, under smoother conditions (C^{k+1} in particular), the global convergence rate is similar to those given in Newey (1997) and Huang and Shen (2004b) in nonparametric regression setting, which shows that the existence of a random vector as a predictor does not change the rate of convergence of the estimated functional coefficient. Moreover, if we take $r = (a + 1)/(a + 2b)$ and $k = (2b - 1)/2(a + 1) - 1$, then we can obtain the same rate of convergence of the estimated functional coefficient as Shin (2009).

For the estimator of variance σ^2, we have the following theorem

Theorem 3 *If conditions (C1)–(C5) hold, then we have*

$$\sqrt{n}\left(\widehat{\sigma}_n^2 - \sigma^2\right) \xrightarrow{D} N(0, \Lambda^2),$$

where $\Lambda^2 = E(\varepsilon_1^2 - \sigma^2)^2$.

Further, let $\widehat{\Sigma}_n = n^{-1}\mathbf{Z}^T(I - A)\mathbf{Z}$. In the light of the above theorems, we can obtain the following corollary.

Corollary 1 *Under conditions (C1)–(C5), as* $n \to \infty$*, we have*

$$\widehat{\chi}^2_{n,p} = \frac{n}{\widehat{\sigma}^2_n}(\widehat{\beta} - \beta)^T \widehat{\Sigma}_n(\widehat{\beta} - \beta) \xrightarrow{D} \chi^2_p.$$

Remark 3 According to the corollary 1, we can obtain an approximate $(1-\gamma)$ asymptotic confidence region for parameter vector β, that is,

$$\left\{a \in R^p : \frac{n}{\widehat{\sigma}^2_n}(\widehat{\beta} - a)^T \widehat{\Sigma}_n(\widehat{\beta} - a) \le \chi^2_{p,1-\gamma}\right\}.$$

Also, we can get an approximate $(1 - \gamma)$ asymptotic confidence interval for every parameter β_j, $j = 1, \ldots, p$, that is,

$$\left[\widehat{\beta}_j + z_{\gamma/2}\frac{\widehat{\sigma}_n(\widehat{\Sigma}_n)^{-1}_{jj}}{\sqrt{n}}, \widehat{\beta}_j - z_{\gamma/2}\frac{\widehat{\sigma}_n(\widehat{\Sigma}_n)^{-1}_{jj}}{\sqrt{n}}\right],$$

where $\widehat{\sigma}_n(\widehat{\Sigma}_n)^{-1}_{jj}$ is the jth diagonal element of $\widehat{\sigma}_n(\widehat{\Sigma}_n)^{-1}$.

4 Simulation studies

In this section, we present some simulation results to illustrate the finite sample behaviors of the polynomial spline estimation and compare our method with the Shin (2009)'s.

4.1 Models for generating simulation data

In this subsection we specify four models to generate simulation data $\{(X_i, \mathbf{Z}_i, Y_i)\}_{i=1}^n$. In the first three models we take the same form as Lian (2011) to generate X_i, that is,

$$X_i = \sum_{j=1}^{50} \xi_{ij} j^{-1}\phi_j(t),$$

where $\phi_1(t) = 1, \phi_j(t) = \sqrt{2}\cos((j-1)\pi t)$ for $j \ge 2$ and ξ_{ij} is independent and identical distribution with $U[-\sqrt{3}, \sqrt{3}]$.

Model 1: $Y_i = 1.5Z_{i1} - Z_{i2} + 2Z_{i3} + \int_0^1 X_i(t)\alpha(t)dt + \varepsilon_i$, where $\mathbf{Z}_i = (Z_{i1}, Z_{i2}, Z_{i3})^T$ is from a multivariate normal distribution $N(0, \Phi)$ with covariance matrix $\Phi = [0.9, 0.2, 0.3; 0.2, 0.5, 0.1; 0.3, 0.1, 1]$. The functional coefficient

$\alpha(t) = \sum_{j=1}^{50} b_j\phi_j(t)$, where $b_1 = 0.5$, $b_j = 4j^{-2}$, for $j \geq 2$ and the error variable ε_i is $N(0, 1)$.

Model 2: $Y_i = 2Z_{i1} - Z_{i2} + \int_0^1 X_i(t)\alpha(t)dt + \varepsilon_i$, $Z_{i1} = \int_0^1 X_i(t)\alpha_1(t)dt + \varepsilon_{i1}$, $Z_{i2} = \int_0^1 X_i(t)\alpha_2(t)dt + \varepsilon_{i2}$, where the functional coefficient $\alpha(t)$ is similarly defined in **Model 1**. In addition, functional coefficients $\alpha_1(t) = \sum_{j=1}^{50} b_{1j}\phi_j(t)$ and $\alpha_2(t) = \sum_{j=1}^{50} b_{2j}\phi_j(t)$, where $b_{11} = 1, b_{21} = -0.5, b_{1j} = 2j^{-2}, b_{2j} = 3j^{-2}$ for $j \geq 2$. Random error variables ε_i and ε_{i1} are $N(0, 0.25)$ and ε_{i2} is $N(0, 0.64)$.

Model 3: $Y_i = 1.5Z_{i1} + 5Z_{i2} - 1.7Z_{i3} + \int_0^1 X_i(t)\alpha(t)dt + \varepsilon_i$, where $\mathbf{Z}_i = (Z_{i1}, Z_{i2}, Z_{i3})^T$ is from a multivariate normal distribution $N(0, \mathbf{I}_3)$. The functional coefficient is given by

$$\alpha(t) = 2\sin(0.5\pi t) + 4\sin(1.5\pi t) + 5\sin(2.5\pi t),$$

which is similar to example (a) in Cardot et al. (2003). The error variable ε_i is $N(0, 0.36)$.

Model 4: We take the same example in Shin (2009), that is,

$$Y_i = 2Z_{i1} - Z_{i2} + 1.5Z_{i3} + 5Z_{i4} - 1.7Z_{i5} + \int_0^1 X_i(t)\alpha(t)dt + \varepsilon_i,$$

where $X_i(t)$ is a standard Brownian motion and $\alpha(t) = \sqrt{2}\sin(\pi t/2) + 3\sqrt{2}\sin(3\pi t/2)$. Random vector $\mathbf{Z}_i = (Z_{i1}, Z_{i2}, Z_{i3}, Z_{i4}, Z_{i5})^T$ is from a multivariate normal distribution $N(0, \mathbf{I}_5)$, and error variable ε_i is $N(0, 1)$.

For practicality, the random functions $X_i(t)$ in Models 1–4 are all only observed at 100 equally spaced points on [0, 1].

4.2 Implementation

In this subsection, we specifically illustrate implementation of our method and Shin (2009)'s method. To implement Shin (2009)'s method, we need to turn discrete observation data of $X_i(t)$ into functional data objects. In this paper, we utilize the method mentioned in Chapter 4 of Ramsay et al. (2009) and choose 25 B-spline functions to build functional data. We also use pca.fd function mentioned in Chapter 7 of Ramsay et al. (2009) to carry out a functional principal components analysis. For our procedure, we have to choose the degrees of spline functions, the positions and the number of knots. Similarly to Huang and Shen (2004b), we choose B-spline basis with equally spaced knots and the fixed degree 2 in this paper. Then, we only need to select the number of the B-spline basis and eigenfunctions K_n. Many methods can be used to select K_n, for example, AIC (Akaike 1974), BIC (Schwarz 1978), "leave-one-subject-out" cross-validation (Rice and Silverman 1991) and modified multi-fold cross-validation (Cai et al. 2000). In this paper, we use "leave-one-subject-out" cross-validation technique to choose the number of B-spline basis and eigenfunctions. Specifically, we select K_n by minimizing the following cross-validation score:

$$CV(K_n) = \sum_{i=1}^{n} \left(Y_i - \langle X_i, \widehat{\alpha}^{-i}\rangle - \mathbf{Z}_i^T\widehat{\beta}^{-i}\right)^2,$$

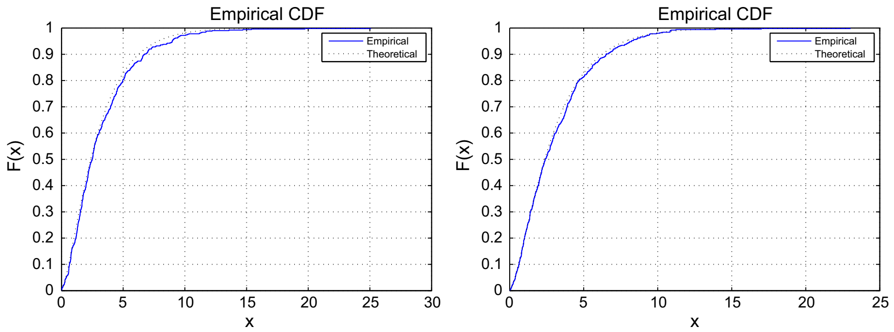

Fig. 1 The empirical distribution function (real line) of $\widehat{\chi}^2_{n,3}$ from 500 simulated samples under Model 1 with different sample sizes: (*left*) n = 100; (*right*) n = 500

where $\widehat{\alpha}^{-i}$ and $\widehat{\beta}^{-i}$ are estimators computed by deleting the ith observation $(X_i, \mathbf{Z}_i, Y_i)$. In our procedure, the number of B-spline basis ranges from 3 to 12 and the number of eigenfunctions ranges from 1 to 10. For the integrals involved in matrix B, we approximate them by the trapezoidal rule.

Two risk functions are used to assess the performances of our estimators and the Shin (2009)'s: the mean square prediction error of the response variable Y, which is similar to (26) in Cardot et al. (2003),

$$\text{MSPE} = n^{-1} \sum_{i=1}^{n} \left(\widehat{Y}_i - \mathbf{Z}_i^T \beta - \int_0^1 X_i(t)\alpha(t)dt \right)^2,$$

and the square-root of average squared error (RASE) of functional coefficient $\alpha(t)$, which is similar to (6) in Huang and Shen (2004b),

$$\text{RASE} = \left[n_{grid}^{-1} \sum_{k=1}^{n_{grid}} (\widehat{\alpha}(t_k) - \alpha(t_k))^2 \right]^{1/2},$$

where $\{t_k, k = 1, \ldots, n_{grid}\}$ are grid points chosen to be equally spaced on the interval [0, 1]. In this paper, the number of grid points $n_{grid} = 101$.

We use Matlab to implement our procedure. For each simulation model abovementioned, we consider two different sample sizes: $n = 100$ and $n = 500$, and each simulation experiment has been repeated 500 times.

4.3 Simulation results

In this subsection, we present some simulation results of 4 simulation models mentioned in (4.1). Notations $_s$ and $_p$ denote our estimation method and the Shin (2009)'s, respectively.

Figure 1 displays the empirical distribution function of $\widehat{\chi}^2_{n,3}$ from 500 simulated samples under Model 1. For Models 2–4, the empirical distribution functions of $\widehat{\chi}^2_{n,p}$ have same performance, so we omit them to save space. We can see from this figure that as the sample size n increases, the empirical distribution more and more approach

Table 1 The MSE of estimators $\widehat{\beta}$ and $\widehat{\sigma}_n^2$ under different models and sample sizes

	n = 100		n = 500		n = 100		n = 500	
	MSE_s	MSE_p	MSE_s	MSE_p	MSE_s	MSE_p	MSE_s	MSE_p
	Model 1				Model 2			
$\widehat{\beta}_1$	0.01463	0.01445	0.00301	0.00303	0.01107	0.01132	0.00187	0.00191
$\widehat{\beta}_2$	0.02436	0.02463	0.00464	0.00465	0.00452	0.00465	0.00083	0.00083
$\widehat{\beta}_3$	0.01126	0.01100	0.00212	0.00213	–	–	–	–
$\widehat{\sigma}_n^2$	0.02265	0.02274	0.00414	0.00413	0.00164	0.00170	0.00028	0.00029
	Model 3				Model 4			
$\widehat{\beta}_1$	0.00437	0.00446	0.00081	0.00081	0.01214	0.01207	0.00205	0.00206
$\widehat{\beta}_2$	0.00385	0.00384	0.00075	0.00076	0.01147	0.01145	0.00201	0.00202
$\widehat{\beta}_3$	0.00412	0.00417	0.00070	0.00071	0.01132	0.01105	0.00214	0.00214
$\widehat{\beta}_4$	–	–	–	–	0.01140	0.01123	0.00197	0.00196
$\widehat{\beta}_5$	–	–	–	–	0.01069	0.01077	0.00200	0.00120
$\widehat{\sigma}_n^2$	0.00375	0.00346	0.00058	0.00057	0.03013	0.02753	0.00482	0.00459

Table 2 The mean(sd) of RASE and MSPE for Models 1–4

		$RASE_s$	$RASE_p$	$MSPE_s$	$MSPE_p$
Model 1	n = 100	0.4120(0.1571)	0.4916(0.1104)	0.0614(0.0359)	0.0621(0.0335)
	n = 500	0.2138(0.0637)	0.3182(0.0640)	0.0127(0.0068)	0.0137(0.0070)
Model 2	n = 100	0.4366(0.1193)	0.5385(0.0786)	0.0182(0.0119)	0.0214(0.0121)
	n = 500	0.2399(0.0209)	0.3179(0.0117)	0.0041(0.0023)	0.0049(0.0022)
Model 3	n = 100	0.9257(0.4853)	1.0148(0.2983)	0.0426(0.0194)	0.0424(0.0168)
	n = 500	0.4509(0.2155)	0.5509(0.1274)	0.0088(0.0040)	0.0094(0.0034)
Model 4	n = 100	2.7417(3.3974)	1.4131(1.8058)	0.1036(0.0559)	0.0914(0.0539)
	n = 500	1.5595(1.5423)	0.5949(0.7420)	0.0232(0.0118)	0.0182(0.0109)

the theoretical distribution, which also reveals the validity of asymptotic normality in Sect. 3.

Table 1 summarizes mean squared errors (MSE) of estimators $\widehat{\beta}$ and $\widehat{\sigma}_n^2$ under Models 1–4. Table 2 presents the mean and standard deviation of RASE and MSPE to evaluate the performance of our estimation procedure. Figure 2 shows our estimate (dashed curve) and the Shin (2009)'s (dotted curve) from the typical samples which correspond to the minimum of $RASE_s$ and $RASE_p$ under Models 1–4, respectively. From these results in our simulation examples, we can know that the two estimation methods are very close for the parameter component β and σ. However, from the prediction perspective and the estimation effect of function coefficient $\alpha(t)$, if the functional coefficient $\alpha(t)$ can be expressed linear combination of eigenfunctions of covariance operator Γ, the Shin (2009)'s method is superior to ours, if not, our method seems to perform better than Shin (2009)' method. At the same time, the differences

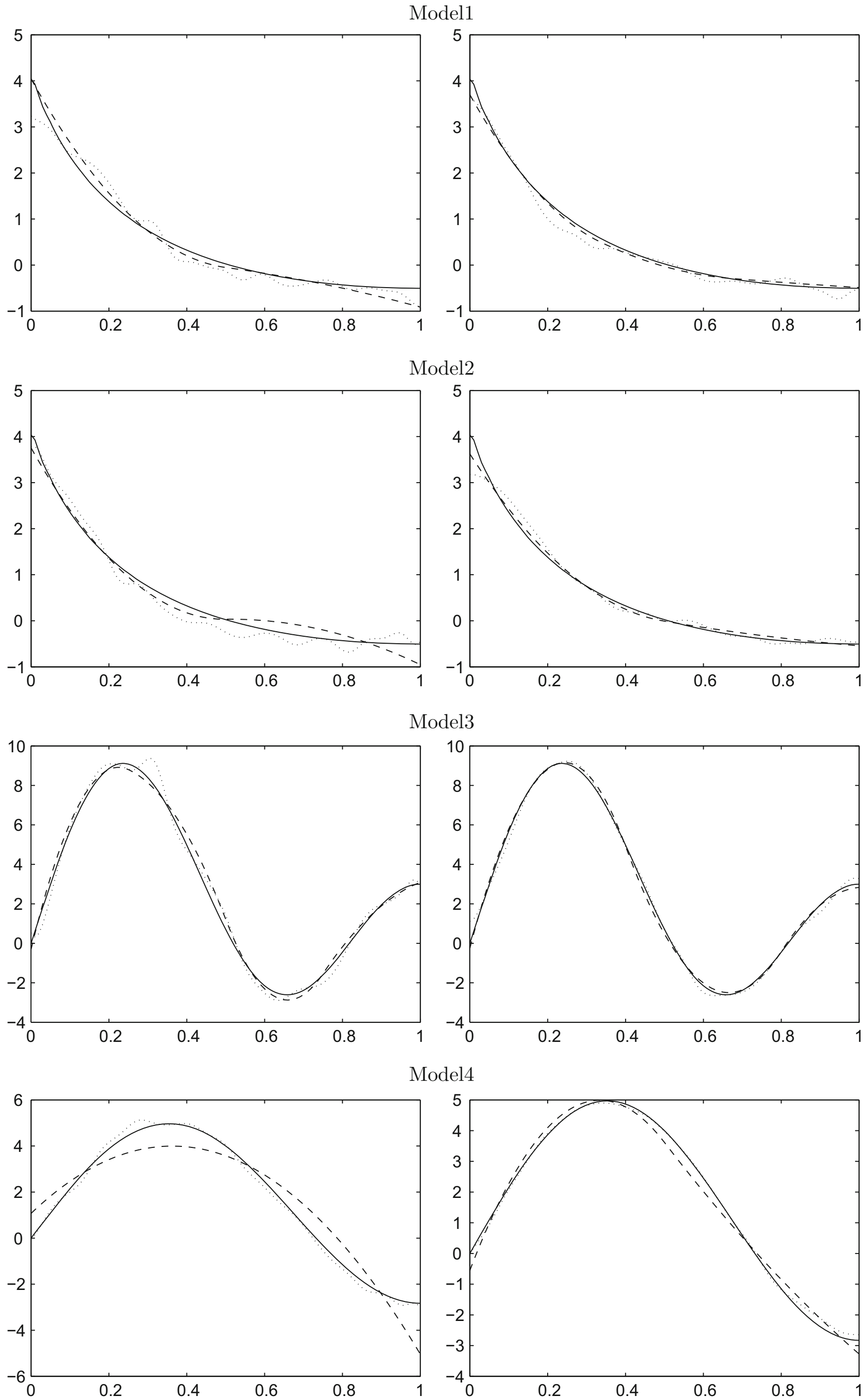

Fig. 2 The true $\alpha(t)$ (*solid curve*), the polynomial spline estimator (*dashed curve*) and the Shen (2009)'s (*dotted curve*) under different models and sample sizes: (*left*) n = 100; (*right*) n = 500

Table 3 The mean(sd) of running cpu time for Models 3–4

	Model 3		Model 4	
	n = 100	n = 500	n = 100	n = 500
Time_s	0.5647(0.1307)	78.7888(2.1625)	0.4171(0.1078)	57.7458(1.7945)
Time_p	0.7107(0.1477)	77.6594(2.4156)	0.4883(0.0962)	55.5548(1.8272)

between the two estimation methods become smaller and smaller as the sample size n increases.

Table 3 displays the mean and standard deviation of running cpu time in a Dell personal computer with Inter(R) Core(TM)2 Duo CPU. We seem to infer from Table 3 that our method is more computationally expedient at least in the examples studied when the sample size is small, while the Shin (2009)'s is more computationally expedient if the sample size is large.

5 Conclusion and further research

In this paper, we propose the polynomial spline estimation for the partial functional linear model. Some asymptotic results are established, including asymptotic normality for the parameter vector and the global rate of convergence for the functional coefficient. By simulation studies, we verify the validity of theoretical results. On the one hand, from the prediction perspective and the estimation effect of function coefficient $\alpha(t)$, we detect if the functional coefficient $\alpha(t)$ can be expressed linear combination of eigenfunctions of covariance operator Γ, the Shin (2009)'s method is superior to ours, if not, our method seems to perform better than Shin (2009)' method. While the differences between the two estimation methods become smaller and smaller as the sample size n increases. On the other hand, from computational time, we can draw a conclusion that our method is more computationally expedient at least in the examples studied when the sample size is small, while the Shin (2009)'s is more computationally expedient if the sample size is large. From our limited study, we only consider the functional predictor would be observed fully. However, we can usually obtain some sparsely discrete observations for each functional observation in practice. For this case, we can use smooth techniques to approximate the functional observations. And then, the polynomial spline method can also be used to estimate the partial functional linear model.

Acknowledgments The work was supported by National Nature Science Foundation of China (Grant Nos. 10961026, 11171293, 11225103, 11301464), the PH.D. Special Scientific Research Foundation of Chinese University (20115301110004), the Key Fund of Yunnan Province (Grant No. 2010CC003) and the Scientific Research Foundation of Yunnan Provincial Department of Education (No. 2013Y360). We are grateful to the referees and the editors for their constructive remarks that greatly improved the manuscript.

Appendix

In the appendix, we give the proofs of the theorems and corollary in Sect. 3.

Set $B_s = {K_n}^{1/2} N_s^b$, $s = 1, \ldots, K_n$, where N_s^b are the normalized B-splines. From the Theorem 4.2 of Chapter 5 of DeVore and Lorentz (1993), we have that for any spline function $\sum_{s=1}^{K_n} b_s B_s$, there are positive constants M_1 and M_2 such that

$$M_1\|b\|_2^2 \leq \int \left\{\sum_{s=1}^{K_n} b_s B_s\right\}^2 \leq M_2\|b\|_2^2, \tag{7}$$

where $\|\cdot\|_2$ is Euclidean norm. Let $\|r\|_\infty = \sup_{x\in[0,1]} |r(x)|$.

In order to prove the theorems, we need the following two lemmas.

Lemma 1 *If conditions (C1) and (C2) hold, then we have*

(i)

$$\sup_{a\in S_{k,N_n}} \left|\frac{\frac{1}{n}\sum_{i=1}^n \langle X_i, a\rangle^2}{E\langle X, a\rangle^2} - 1\right| = o_p(1).$$

(ii) there exists an interval $[M_3, M_4], 0 < M_3 < M_4 < \infty$ *such that as* $n \to \infty$,

$$P\left\{\text{all the eigenvalues of } \frac{1}{n}B^T B \text{ fall in } [M_3, M_4]\right\} \to 1.$$

Note that the Lemma 1 is a generalization of Lemma 1 and 2 in Huang and Shen (2004b) in functional data case. We give a brief proof in the following.

Proof (i) Let Γ_n denote the empirical versions of operator Γ, that is,

$$\Gamma_n x(t) = \frac{1}{n}\sum_{i=1}^n \langle X_i, x\rangle X_i(t), \quad x \in H, t \in [0, 1].$$

By the Cauchy–Schwarz inequality, condition (C2) and (28) in Cardot et al. (2003), we have

$$\begin{aligned}
\left|\frac{\frac{1}{n}\sum_{i=1}^n \langle X_i, a\rangle^2}{E\langle X, a\rangle^2} - 1\right| &= \left|\frac{\langle(\Gamma_n - \Gamma)a, a\rangle}{\langle\Gamma a, a\rangle}\right| \\
&\leq \frac{\|\Gamma_n - \Gamma\|_\infty \|a\|^2}{C\|a\|^2} \\
&= \frac{\|\Gamma_n - \Gamma\|_\infty}{C}.
\end{aligned}$$

Then for an arbitrary constant $\epsilon > 0$, by Lemma 5.2 in Cardot et al. (1999), we have

$$\begin{aligned}
P\left\{\sup_{a\in S_{k,N_n}} \left|\frac{\frac{1}{n}\sum_{i=1}^n \langle X_i, a\rangle^2}{E\langle X, a\rangle^2} - 1\right| > \epsilon\right\} &\leq P\left\{\|\Gamma_n - \Gamma\|_\infty > C\epsilon\right\} \\
&\leq \frac{E\|\Gamma_n - \Gamma\|_\infty^2}{C^2\epsilon^2} \\
&\leq \frac{E\|X\|^4}{nC^2\epsilon^2},
\end{aligned}$$

together with (C2), which gives the result.

(ii) Let $b = (b_1, \ldots, b_{K_n})^T$, $a = \sum_{s=1}^{K_n} b_s B_s$. It follows from (i) that except an event whose probability tends to zero as $n \to \infty$,

$$\frac{1}{n} b^T B^T B b = \frac{1}{n} \sum_{i=1}^{n} \left(\sum_{s=1}^{K_n} b_s \langle X_i, B_s \rangle \right)^2 \asymp E\langle X, a\rangle^2.$$

By the Cauchy–Schwarz inequality, (28) in Cardot et al. (2003) and (7),

$$E\langle X, a\rangle^2 \asymp \|a\|^2 \asymp \|b\|_2^2.$$

Thus, except an event whose probability tends to zero, $\frac{1}{n} b^T B^T B b \asymp \|b\|_2^2$, holds uniformly for all b, which yields the result. □

Lemma 2 *Under conditions (C1)–(C5), as $n \to \infty$, we have*

$$\frac{\mathbf{Z}^T (I - A)\mathbf{Z}}{n} \xrightarrow{P} \Sigma.$$

Proof Let $\mu_j(X_i) = E(Z_{ij}|X_i) = \langle X_i, g_j\rangle$, $\eta_{ij} = Z_{ij} - \mu_j(X_i)$,

$$\tilde{V_j} = \Big(\mu_j(X_1), \ldots, \mu_j(X_n)\Big)^T, \quad \tilde{\eta_j} = (\eta_{1j}, \ldots, \eta_{nj})^T, \quad j = 1, \ldots, p.$$

We also define $V = (\tilde{V}_1, \ldots, \tilde{V}_p)$, $\eta = (\tilde{\eta}_1, \ldots, \tilde{\eta}_p)$. Then, $\mathbf{Z} = \eta + V$ and

$$\begin{aligned}
\frac{\mathbf{Z}^T (I - A)\mathbf{Z}}{n} &= \frac{(\eta + V)^T (I - A)(\eta + V)}{n} \\
&= \frac{\eta^T (I - A)\eta}{n} + \frac{\eta^T (I - A)V}{n} + \frac{V^T (I - A)\eta}{n} + \frac{V^T (I - A)V}{n} \\
&= I_1 + I_2 + I_3 + I_4.
\end{aligned}$$

For the (j, l)th element of I_1

$$(I_1)_{jl} = \frac{\tilde{\eta_j}^T (I - A)\tilde{\eta_l}}{n} = \frac{\tilde{\eta_j}^T \tilde{\eta_l}}{n} - \frac{\tilde{\eta_j}^T A \tilde{\eta_l}}{n}, \quad j, l = 1, \ldots, p.$$

By independence and the Cauchy–Schwarz inequality, we have

$$\begin{aligned}
E\left\{ \frac{\sum_{i=1}^{n} [\eta_{ij}\eta_{il} - E(\eta_{ij}\eta_{il})]}{n} \right\}^2 &= \frac{E(\eta_{1j}\eta_{1l} - E\eta_{1j}\eta_{1l})^2}{n} \\
&\le \frac{E\eta_{1j}^2 \eta_{1l}^2}{n} \\
&\le \frac{(E\eta_{1j}^4)^{1/2} (E\eta_{1l}^4)^{1/2}}{n}.
\end{aligned}$$

Further, by C_r inequality and (C2)–(C4), we have

$$E(\eta_{1j})^4 = E\Big(Z_{1j} - \langle X_1, g_j\rangle\Big)^4 \le 8\Big(E|Z_{1j}|^4 + E|\langle X_1, g_j\rangle|^4\Big)$$
$$\le 8\Big(E|Z_{1j}|^4 + E\|X_1\|^4\|g_j\|^4\Big) < \infty, \quad j = 1, \ldots, p.$$

Thus,

$$\frac{\eta^T\eta}{n} \xrightarrow{P} \Sigma. \tag{8}$$

Note that $A \ge 0$, then we have

$$\left|\frac{\tilde{\eta_j}^T A\tilde{\eta_l}}{n}\right| \le \left|\frac{\tilde{\eta_j}^T A\tilde{\eta_j}}{n}\right|^{1/2}\left|\frac{\tilde{\eta_l}^T A\tilde{\eta_l}}{n}\right|^{1/2}.$$

By Lemma 1, we can know that except an event whose probability tends to zero,

$$\frac{\tilde{\eta_j}^T A\tilde{\eta_j}}{n} = \frac{\tilde{\eta_j}^T B(B^T B)^{-1}B^T\tilde{\eta_j}}{n} \asymp \frac{\tilde{\eta_j}^T BB^T\tilde{\eta_j}}{n^2}.$$

Also note that $E\langle X_i, B_s\rangle\eta_{ij} = E\langle X_i, B_s\rangle E(\eta_{ij}|X_i) = 0$. Then, by (7) and conditions (C2)–(C4), we have there exists a positive constant C such that

$$\begin{aligned} E\tilde{\eta_j}^T BB^T\tilde{\eta_j} &= E\left\{\sum_{s=1}^{K_n}\left[\sum_{i=1}^{n}\langle X_i, B_s\rangle\eta_{ij}\right]^2\right\} \\ &= n\sum_{s=1}^{K_n} E\langle X_1, B_s\rangle^2\eta_{1j}^2 \\ &\le n\sum_{s=1}^{K_n}\|B_s\|^2\big(E\|X_1\|^4\big)^{1/2}\big(E|\eta_{1j}|^4\big)^{1/2} \\ &\le CnK_n. \end{aligned}$$

Thus, for $j, l = 1, \ldots, p$,

$$\frac{\tilde{\eta_j}^T A\tilde{\eta_l}}{n} = O_p\Big(\frac{K_n}{n}\Big) = o_p(1),$$

which together with (8) yields

$$I_1 \xrightarrow{P} \Sigma. \tag{9}$$

For the (j,l)-th element of I_4, $j,l=1,\ldots,p$,

$$(I_4)_{jl} = \frac{\tilde{V_j}^T (I-A)\tilde{V_l}}{n},$$

by Cauchy–Schwartz inequality,

$$E\left|\tilde{V_j}^T (I-A)\tilde{V_l}\right| \le \left(E\tilde{V_j}^T (I-A)\tilde{V_j}\right)^{1/2}\left(E\tilde{V_l}^T (I-A)\tilde{V_l}\right)^{1/2}.$$

It follows from Theorem XII.1 of de Boor (2001) that there exist positive constant C_j and spline function $g_j^* \in S_{k,N_n}$, $j=1,\ldots,p$ such that

$$\|g_j - g_j^*\|_\infty \le C_j h^{k+1}.$$

Set $g_j^* = \sum_{s=1}^{K_n} b_{js}^* B_s$, $\quad b_j^* = (b_{j1}^*, \ldots, b_{jK_n}^*)^T$, $\quad j=1,\ldots,p$, then,

$$\tilde{V_j}^* = \left(\langle X_1, g_j^*\rangle, \ldots, \langle X_n, g_j^*\rangle\right)^T = Bb_j^*.$$

As A is an orthogonal projection matrix,

$$\begin{aligned} E\left|\tilde{V_j}^T (I-A)\tilde{V_j}\right| &= E\left|(I-A)\tilde{V_j}\right|^2 \le E\left|\tilde{V_j} - \tilde{V}_j^*\right|^2 \\ &\le nE\|X_1\|^2\|g_j - g_j^*\|^2 \lesssim nh^{2(k+1)}. \end{aligned}$$

From the above results and (C1), we have

$$\frac{\tilde{V_j}^T (I-A)\tilde{V_l}}{n} = O_p\left(h^{2(k+1)}\right) = o_p(1),$$

that is,

$$I_4 \xrightarrow{P} 0. \tag{10}$$

For the (j,l)-th element of I_2 and I_3, $j,l=1,\ldots,p$, we have

$$\begin{aligned} \frac{|\tilde{\eta_j}^T (I-A)\tilde{V_l}|}{n} &\le \frac{(\tilde{\eta_j}^T (I-A)\tilde{\eta_j})^{1/2}(\tilde{V_l}^T (I-A)\tilde{V_l})^{1/2}}{n}, \\ \frac{|\tilde{V_j}^T (I-A)\tilde{\eta_l}|}{n} &\le \frac{(\tilde{V_j}^T (I-A)\tilde{V_j})^{1/2}(\tilde{\eta_l}^T (I-A)\tilde{\eta_l})^{1/2}}{n}. \end{aligned}$$

Using (9) and (10), we can infer that

$$I_2 \xrightarrow{P} 0, \quad I_3 \xrightarrow{P} 0. \tag{11}$$

The combination of (9)–(11) allows us to finish the proof of Lemma 2. □

Proof of Theorem 1 Denote $\Phi = \Big(\langle X_1, \alpha\rangle, \ldots, \langle X_n, \alpha\rangle\Big)^T$, $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_1)^T$. Then, $Y = \mathbf{Z}\beta + \Phi + \varepsilon$. We can write

$$\begin{aligned}\sqrt{n}(\widehat{\beta} - \beta) &= \sqrt{n}\Big[\mathbf{Z}^T(I - A)\mathbf{Z}\Big]^{-1}\mathbf{Z}^T(I - A)\Phi + \sqrt{n}\Big[\mathbf{Z}^T(I - A)\mathbf{Z}\Big]^{-1}\mathbf{Z}^T(I - A)\varepsilon\\ &= \Delta_1 + \Delta_2.\end{aligned}$$

Observe that

$$\begin{aligned}\Delta_1 &= \Big[\frac{\mathbf{Z}^T(I - A)\mathbf{Z}}{n}\Big]^{-1} n^{-1/2}\mathbf{Z}^T(I - A)\Phi = \Big[\frac{\mathbf{Z}^T(I - A)\mathbf{Z}}{n}\Big]^{-1}\Delta_{11},\\ \Delta_2 &= \Big[\frac{\mathbf{Z}^T(I - A)\mathbf{Z}}{n}\Big]^{-1} n^{-1/2}\mathbf{Z}^T(I - A)\varepsilon = \Big[\frac{\mathbf{Z}^T(I - A)\mathbf{Z}}{n}\Big]^{-1}\Delta_{21}.\end{aligned}$$

For Δ_{11}, as $\mathbf{Z} = \eta + V$,

$$\Delta_{11} = n^{-1/2}\eta^T(I - A)\Phi + n^{-1/2}V^T(I - A)\Phi. \tag{12}$$

By (C4) and the Theorem XII.1 of de Boor (2001), we know that there is a spline function $\alpha^* = \sum_{s=1}^{K_n} b_s^* B_s \in S_{k,N_n}$ and positive constant C such that

$$\|\alpha - \alpha^*\|_\infty \le Ch^{k+1}. \tag{13}$$

Set $\Phi^* = (\langle X_1, \alpha^*\rangle, \ldots, \langle X_n, \alpha^*\rangle)^T$ and $b^* = (b_1^*, \ldots, b_{K_n}^*)^T$, we have $\Phi^* = Bb^*$. For $j = 1, \ldots, p$, by conditions (C1), (C2), (C4) and Theorem XII.1 of de Boor (2001), we can infer

$$\begin{aligned}E\Big|\tilde{V_j}^T(I - A)\Phi\Big| &= E\Big|(\tilde{V_j} - \tilde{V_j^*})^T(I - A)(\Phi - \Phi^*)\Big|\\ &\le E\Big\{\Big|(\tilde{V_j} - \tilde{V_j^*})^T(I - A)(\tilde{V_j} - \tilde{V_j^*})\Big|^{1/2}\Big|(\Phi - \Phi^*)^T(I - A)(\Phi - \Phi^*)\Big|^{1/2}\Big\}\\ &\le \Big(E\sum_{i=1}^n\langle X_i, g_j - g_j^*\rangle^2\Big)^{1/2}\Big(E\sum_{i=1}^n\langle X_i, \alpha - \alpha^*\rangle^2\Big)^{1/2}\\ &\lesssim nh^{2(k+1)}.\end{aligned}$$

Thus, by (C1) we have

$$n^{-1/2}V^T(I - A)\Phi = O_p(n^{1/2}h^{2(k+1)}) = o_p(1). \tag{14}$$

Observe that for $j = 1, \ldots, p$,

$$\Big|n^{-1/2}\tilde{\eta_j}^T(I - A)\Phi\Big| = \Big|n^{-1/2}\tilde{\eta_j}^T(I - A)(\Phi - \Phi^*)\Big|$$

$$
\begin{aligned}
&\leq \left|n^{-1/2}\tilde{\eta_j}^T(\Phi-\Phi^*)\right| + \left|n^{-1/2}\tilde{\eta_j}^T A(\Phi-\Phi^*)\right| \\
&\leq n^{-1/2}\left|\sum_{i=1}^{n}\eta_{ij}\langle X_i,\alpha-\alpha^*\rangle\right| \\
&\quad + n^{-1/2}\left|\tilde{\eta_j}^T A\tilde{\eta_j}\right|^{1/2}\left|\sum_{i=1}^{n}\langle X_i,\alpha-\alpha^*\rangle^2\right|^{1/2} \\
&\triangleq I_{j1}+I_{j2}.
\end{aligned}
$$

As $E\Big(\eta_{ij}\langle X_i,\alpha-\alpha^*\rangle\Big) = E\Big[\langle X_i,\alpha-\alpha^*\rangle E(\eta_{ij}|X_i)\Big] = 0$ and

$$
\begin{aligned}
E\left|\sum_{i=1}^{n}\eta_{ij}\langle X_i,\alpha-\alpha^*\rangle\right|^2 &= nE\eta_{1j}^2\langle X_1,\alpha-\alpha^*\rangle^2 \\
&\leq n\|\alpha-\alpha^*\|^2(E\eta_{1j}^4)^{1/2}(E\|X_1\|^4)^{1/2} \\
&\lesssim nh^{2(k+1)},
\end{aligned}
$$

we can infer

$$
I_{j1} = O_p(h^{k+1}) = o_p(1). \tag{15}
$$

Further, under the Lemma 1, (C1) and (13), we can show

$$
I_{j2} = O_p(K_n^{1/2}h^{k+1}) = O_p(n^{-r(k+\frac{1}{2})}) = o_p(1). \tag{16}
$$

By (12), (14)–(16) and Lemma 2, we have

$$
\Delta_1 \xrightarrow{P} 0. \tag{17}
$$

Δ_{21} can be expressed as

$$
\Delta_{21} = n^{-1/2}\eta^T(I-A)\varepsilon + n^{-1/2}V^T(I-A)\varepsilon \triangleq R_1+R_2.
$$

Let $\epsilon_i = \eta_i\varepsilon_i$. Since ε_i is independent of $(X_i, \mathbf{Z}_i)$ and $(X_i, \mathbf{Z}_i, Y_i)$ is i.i.d. sequence, the ϵ_i are i.i.d. random variables with $E\epsilon_i = 0$ and $Var(\epsilon_i) = \sigma^2\Sigma$.

Observe that

$$R_1 = n^{-1/2}\eta^T\varepsilon - n^{-1/2}\eta^T A\varepsilon = n^{-1/2}\sum_{i=1}^{n}\eta_i\varepsilon_i - n^{-1/2}\eta^T A\varepsilon$$
$$= n^{-1/2}\sum_{i=1}^{n}\epsilon_i - n^{-1/2}\eta^T A\varepsilon.$$

Then, by the central limit theorem,

$$n^{-1/2}\sum_{i=1}^{n}\epsilon_i \xrightarrow{D} N\big(0, \sigma^2\Sigma\big). \tag{18}$$

Also note that

$$\left|n^{-1/2}\tilde{\eta_j}^T A\varepsilon\right| \le n^{-1/2}\Big(\tilde{\eta_j}^T A\tilde{\eta}_j\Big)^{1/2}\Big(\varepsilon^T A\varepsilon\Big)^{1/2}.$$

Then, it follows from Lemma 1 that

$$\varepsilon^T A\varepsilon = \varepsilon^T B(B^T B)^{-1}B^T\varepsilon \asymp \frac{\varepsilon^T BB^T\varepsilon}{n}.$$

Since $E\langle X_i, B_s\rangle\varepsilon_i\langle X_j, B_s\rangle\varepsilon_j = 0, i \neq j$, we have

$$E\varepsilon^T BB^T\varepsilon = E\sum_{s=1}^{K_n}\left(\sum_{i=1}^{n}\varepsilon_i\langle X_i, B_s\rangle\right)^2 = \sum_{s=1}^{K_n}\sum_{i=1}^{n}E\varepsilon_i^2\langle X_i, B_s\rangle^2$$
$$\le n\sigma^2\sum_{s=1}^{K_n}E\|X_1\|^2\|B_s\|^2 \lesssim nK_n,$$

that is, $\varepsilon^T A\varepsilon = O_p(K_n)$. In addition, we can know from the proof of Lemma 2 that

$$\frac{\tilde{\eta_j}^T A\tilde{\eta}_j}{n} = O_p\Big(\frac{K_n}{n}\Big).$$

Thus,

$$n^{-1/2}\eta^T A\varepsilon = O_p(K_n n^{-1/2}) = o_p(1),$$

which, together with (18), yields

$$R_1 \xrightarrow{D} N\big(0, \sigma^2\Sigma\big). \tag{19}$$

For the jth element of R_2, $j = 1, \ldots, p$, we have

$$\left|n^{-1/2}\tilde{V_j}^T(I - A)\varepsilon\right| = \left|n^{-1/2}(\tilde{V}_j - \tilde{V}_j^*)^T(I - A)\varepsilon\right|$$
$$\le \left|n^{-1/2}(\tilde{V}_j - \tilde{V}_j^*)^T\varepsilon\right| + \left|n^{-1/2}(\tilde{V}_j - \tilde{V}_j^*)^T A\varepsilon\right|$$

$$\leq \left|n^{-1/2}(\tilde{V}_j - \tilde{V}_j^*)^T \varepsilon\right| + n^{-1/2}\left|(\tilde{V}_j - \tilde{V}_j^*)^T (\tilde{V}_j - \tilde{V}_j^*)\right|^{1/2} \left|\varepsilon A \varepsilon\right|^{1/2}$$
$$\triangleq J_{j1} + J_{j2}.$$

Since ε_i is independent $(X_i, \mathbf{Z}_i)$, we have

$$EJ_{j1}^2 = n^{-1}\sum_{i=1}^{n} E\varepsilon_i^2 \langle X_i, g_j - g_j^*\rangle^2 \leq \sigma^2 E\|X_1\|^2 \|g_j - g_j^*\|^2 \lesssim h^{2(k+1)}.$$

Then,

$$J_{j1} = O_p\big(h^{k+1}\big) = o_p(1).$$

Also, observe that

$$E\sum_{i=1}^{n}\langle X_i, g_j - g_j^*\rangle^2 = nE\langle X_1, g_j - g_j^*\rangle^2 \leq nE\|X_1\|^2\|g_j - g_j^*\|^2 \lesssim nh^{2k+2}.$$

Then, by (C1), we have

$$J_{j2} = O_p\big(K_n^{1/2} h^{k+1}\big) = O_p\big(n^{-r(k+1/2)}\big) = o_p(1).$$

From the above results, we can infer

$$R_2 \xrightarrow{P} 0. \tag{20}$$

Now, by Lemma 2, (17), (19), (20) and Slutsky theorem, we can obtain the Theorem 1. □

Proof of Theorem 2 Observe that

$$\widehat{b} = (B^T B)^{-1} B^T (Y - \mathbf{Z}\widehat{\beta}) = (B^T B)^{-1} B^T \big(\mathbf{Z}(\beta - \widehat{\beta}) + \Phi + \varepsilon\big).$$

Let $\tilde{Y} = \mathbf{Z}(\beta - \widehat{\beta}) + \Phi$. Denote $\tilde{b} = (B^T B)^{-1} B^T \tilde{Y}$ and $\tilde{\alpha}(t) = \sum_{s=1}^{K_n} \tilde{b}_s B_s(t)$, where $\tilde{b} = (\tilde{b_1}, \ldots, \tilde{b_{K_n}})^T$. Then, $\widehat{b} - \tilde{b} = (B^T B)^{-1} B^T \varepsilon$. By Lemma 1, we have

$$\|\widehat{b} - \tilde{b}\|_2^2 = \varepsilon^T B(B^T B)^{-1}(B^T B)^{-1} B^T \varepsilon \asymp \frac{\varepsilon^T B B^T \varepsilon}{n^2},$$

except on an event whose probability tends to zero as $n \to \infty$. Thus, by (7), we can infer

$$\|\widehat{\alpha} - \tilde{\alpha}\|^2 \asymp \|\widehat{b} - \tilde{b}\|_2^2 = O_p\Big(\frac{K_n}{n}\Big). \tag{21}$$

Also, it follows from the Theorem XII.1 of de Boor (2001) that there exists a spline function $\alpha^*(t) = \sum_{s=1}^{K_n} b_s^* B_s(t) \in S_{k,N_n}$ where $b^* = (b_1^*, \ldots, b_{K_n}^*)^T$ and constant $C > 0$ such that

$$\|\alpha^* - \alpha\| \le \|\alpha^* - \alpha\|_\infty \le Ch^{k+1}. \tag{22}$$

By the Theorem XII.1 of de Boor (2001) and (7), we have

$$\left\|\tilde{\alpha} - \alpha^*\right\|^2 \asymp \left\|\tilde{b} - b^*\right\|_2^2 \asymp \frac{(\tilde{b} - b^*)^T B^T B(\tilde{b} - b^*)}{n} = \frac{\|B\tilde{b} - Bb^*\|_2^2}{n}.$$

Observe that $B\tilde{b} = B(B^T B)^{-1} B^T \tilde{Y}$ and $B(B^T B)^{-1} B^T$ is an orthogonal projection matrix. Thus,

$$\begin{aligned}
\frac{\|B\tilde{b} - Bb^*\|_2^2}{n} &\le \frac{\|\tilde{Y} - Bb^*\|_2^2}{n} = \frac{\left\|\mathbf{Z}(\beta - \widehat{\beta}) + (\Phi - \Phi^*)\right\|_2^2}{n} \\
&\le \frac{2}{n}\left[\|\mathbf{Z}(\beta - \widehat{\beta})\|_2^2 + \|\Phi - \Phi^*\|_2^2\right] \\
&= \frac{2}{n}\sum_{i=1}^{n}\left\{\sum_{j=1}^{p} Z_{ij}(\beta_j - \widehat{\beta}_j)\right\}^2 + \frac{2}{n}\sum_{i=1}^{n}\langle X_i, \alpha - \alpha^*\rangle^2.
\end{aligned}$$

Applying (C2), (22) and the Cauchy–Schwarz inequality, we obtain that

$$E\langle X_1, \alpha - \alpha^*\rangle^2 \le E\|X_1\|^2\|\alpha - \alpha^*\|^2 \le Ch^{2(k+1)},$$

that is,

$$\frac{\sum_{i=1}^{n}\langle X_i, \alpha - \alpha^*\rangle^2}{n} = O_p\big(h^{2(k+1)}\big). \tag{23}$$

In addition, note that

$$\frac{1}{n}\sum_{i=1}^{n}\left\{\sum_{j=1}^{p} Z_{ij}(\beta_j - \widehat{\beta}_j)\right\}^2 \le p\sum_{j=1}^{p}\big(\beta_j - \widehat{\beta}_j\big)^2 \frac{\sum_{i=1}^{n} Z_{ij}^2}{n}.$$

Then, it follows from Theorem 1 and (C4) that

$$\frac{1}{n}\sum_{i=1}^{n}\left\{\sum_{j=1}^{p} Z_{ij}(\beta_j - \widehat{\beta}_j)\right\}^2 = O_p(n^{-1}), \tag{24}$$

which together with (23) yields

$$\|\tilde{\alpha} - \alpha^*\|^2 = O_p\big(n^{-1} + h^{2(k+1)}\big). \tag{25}$$

Further, we can infer that

$$\|\widehat{\alpha}-\alpha\|^2 \le 3\Big(\|\widehat{\alpha}-\tilde{\alpha}\|^2+\|\tilde{\alpha}-\alpha^*\|^2+\|\alpha^*-\alpha\|^2\Big). \tag{26}$$

Then, the combination of (21), (22), (25) and (26) allows us to complete the proof of Theorem 2. □

Proof of Theorem 3 We can write

$$\begin{aligned}
\sqrt{n}(\widehat{\sigma}_n^2-\sigma^2) &= \sqrt{n}\left\{\frac{\sum_{i=1}^n\big(Y_i-\langle X_i,\widehat{\alpha}\rangle-\mathbf{Z}_i^T\widehat{\beta}\big)^2}{n}-\sigma^2\right\}\\
&= \sqrt{n}\left\{\frac{\sum_{i=1}^n\big[\langle X_i,\alpha-\widehat{\alpha}\rangle+\mathbf{Z}_i^T(\beta-\widehat{\beta})+\varepsilon\big]^2}{n}-\sigma^2\right\}\\
&= n^{-1/2}\sum_{i=1}^n\langle X_i,\alpha-\widehat{\alpha}\rangle^2+n^{-1/2}\sum_{i=1}^n(\beta-\widehat{\beta})^T\mathbf{Z}_i\mathbf{Z}_i^T(\beta-\widehat{\beta})\\
&\quad+n^{-1/2}\sum_{i=1}^n(\varepsilon_i^2-\sigma^2)+2n^{-1/2}\sum_{i=1}^n\langle X_i,\alpha-\widehat{\alpha}\rangle\varepsilon_i\\
&\quad+2n^{-1/2}\sum_{i=1}^n\varepsilon_i\mathbf{Z}_i^T(\beta-\widehat{\beta})+2n^{-1/2}\sum_{i=1}^n\langle X_i,\alpha-\widehat{\alpha}\rangle\mathbf{Z}_i^T(\beta-\widehat{\beta})\\
&\triangleq R_{n1}+R_{n2}+R_{n3}+R_{n4}+R_{n5}+R_{n6}.
\end{aligned}$$

Observe that

$$R_{n1}=n^{-1/2}\sum_{i=1}^n\langle X_i,\alpha-\widehat{\alpha}\rangle^2 \le \|\alpha-\widehat{\alpha}\|^2 n^{-1/2}\sum_{i=1}^n\|X_i\|^2.$$

Then, by (C1), (C2), theorem 2, we have

$$R_{n1}=O_p\big(K_n n^{-1/2}+n^{1/2}h^{2(k+1)}\big)=O_p\big(n^{r-1/2}+n^{1/2-2r(k+1)}\big)=o_p(1). \tag{27}$$

It follows from (24) that

$$R_{n2}=O_p\big(n^{-1/2}\big)=o_p(1). \tag{28}$$

For R_{n3}, since $E(\varepsilon_1^2-\sigma^2)=0$ and $\Lambda^2=E(\varepsilon_1^2-\sigma^2)^2<\infty$, it follows from the central limit theorem that

$$R_{n3}\xrightarrow{D}N(0,\Lambda^2). \tag{29}$$

For R_{n4}, we have

$$|R_{n4}| = 2n^{1/2}\left|\langle \frac{\sum_{i=1}^{n} X_i\varepsilon_i}{n}, \alpha - \widehat{\alpha}\rangle\right| \le 2n^{1/2}\left\|\frac{\sum_{i=1}^{n} X_i\varepsilon_i}{n}\right\|\left\|\alpha - \widehat{\alpha}\right\|.$$

Then, applying (C1), (C2) and Theorem 2, we can obtain

$$R_{n4} = O_p\left(K_n^{1/2}n^{-1/2} + h^{k+1}\right) = o_p(1). \tag{30}$$

Note that

$$R_{n5} = 2n^{1/2}\frac{\sum_{i=1}^{n}\varepsilon_i \mathbf{Z}_i^T(\beta - \widehat{\beta})}{n} = 2n^{1/2}\sum_{j=1}^{p}(\beta_j - \widehat{\beta}_j)\frac{\sum_{i=1}^{n}\varepsilon_i Z_{ij}}{n}.$$

Thus, using (C3) and Theorem 1, we have

$$R_{n5} = O_p\left(n^{-1/2}\right) = o_p(1). \tag{31}$$

Also, observe that

$$\begin{aligned}|R_{n6}| &= 2n^{1/2}\frac{\left|\sum_{i=1}^{n}\langle X_i, \alpha - \widehat{\alpha}\rangle \mathbf{Z}_i^T(\beta - \widehat{\beta})\right|}{n} \\ &\le 2n^{1/2}\sum_{j=1}^{p}|\beta_j - \widehat{\beta}_j|\|\alpha - \widehat{\alpha}\|\frac{\sum_{i=1}^{n}\|X_i\||Z_{ij}|}{n}.\end{aligned}$$

Then, by (C1)–(C3), Theorem 1 and 2, we can get

$$R_{n6} = O_p\left(K_n^{1/2}n^{-1/2} + h^{k+1}\right) = o_p(1). \tag{32}$$

Finally, using (27)–(32), we can complete the proof of Theorem 3. □

Proof of Corollary 1 It follows form Theorem 1 that

$$\frac{\sqrt{n}}{\sigma}\Sigma^{\frac{1}{2}}(\widehat{\beta} - \beta) \xrightarrow{D} N(0, I_p).$$

Also, by Lemma 2 and Theorem 3, we have that

$$\widehat{\Sigma}_n \xrightarrow{P} \Sigma, \qquad \widehat{\sigma}_n^2 \xrightarrow{P} \sigma^2.$$

Then, by the Slutsky theorem, we obtain the Corollary 1. □

References

Akaike H (1974) A new look at the statistical model identification. IEEE Trans Autom Control 19:716–723
Aneiros-Pérez G, Vieu P (2006) Semi-functional partial linear regression. Stat Probab Lett 76:1102–1110
Aneiros-Pérez G, Vieu P (2008) Nonparametric time series prediction: a semi-functional partial linear modeling. J Multivar Anal 99:834–857
Baíllo A, Grané A (2009) Local linear regression for functional predictor and scalar response. J Multivar Anal 100:102–111
Burba F, Ferraty F, Vieu P (2009) K-nearest neighbour method in functional nonparametric regression. J Nonparametric Stat 21:453–469
Cai TT, Hall P (2006) Prediction in functional linear regression. Annals Stat 34:2159–2179
Cai Z, Fan J, Yao Q (2000) Functional-coefficient regression model for nonlinear time series. J Am Stat Assoc 95:941–956
Cardot H, Ferraty F, Sarda P (1999) Functional linear model. Stat Prob Lett 45:11–22
Cardot H, Ferraty F, Sarda P (2003) Spline estimators for the functional linear model. Stat Sin 13:571–591
Cardot H, Sarda P (2008) Varying-coefficient functional linear regression models. Commun Stat Theory Methods 37:3186–3203
Chen H (1991) Polynomial splines and nonparametric regression. J Nonparametric Stat 1:143–156
Crambes C, Kneip A, Sarda P (2009) Smoothing splines estimators for functional linear regression. Annals Stat 37:35–72
de Boor C (2001) A practical guide to splines. Springer, New York
DeVore RA, Lorentz GG (1993) Constructive approximation. Springer, Berlin
Ferraty F, Vieu P (2006) Nonparametric functional data analysis. Springer, New York
Huang JZ (2003a) Asymptotics for polynomial spline regression under weak conditions. Stat Probab Lett 65:207–216
Huang JZ (2003b) Local asymptotics for polynomial spline regression. Annals Stat 31:1600–1635
Huang JZ, Wu CO, Zhou L (2004a) Polynomial spline estimation and inference for varying coefficient model with longitudinal data. Stat Sin 14:763–788
Huang JZ, Shen H (2004b) Functional coefficient regression models for non-linear time series: a polynomial spline approach. Scand J Stat 31:515–534
Hall P, Horowitz J (2007) Methodology and convergence rates for functional linear regression. Annals Stat 35:70–91
Li Y, Hsing T (2007) On rates of convergence in functional linear regression. J Multivar Anal 98:1782–1804
Lian H (2011) Functional partial linear model. J Nonparametric Stat 23:115–128
Newey WK (1997) Convergence rates and asymptotic normality for series estimators. J Econom 79:147–168
Ramsay J, Dalzell C (1991) Some tools for functional data analysis. J R Stat Soc Ser B 53:539–572
Ramsay J, Silverman B (1997) Functional data analysis. Springer, New York
Ramsay J, Silverman B (2005) Functional data analysis, 2nd edn. Springer, New York
Ramsay J, Hooker G, Graves S (2009) Functional data analysis with R and Matlab. Springer, New York
Rice JA, Silverman BW (1991) Estimating the mean and covariance structure nonparametrically when the data are curves. J R Stat Soc Ser B 53:233–243
Shin H (2009) Partial functional linear regression. J Stat Plan Inference 139:3405–3418
Stone CJ (1994) The use of polynomial splines and their tensor products in multivariate function estimation (with discussion). Annals Stat 22:118–171
Stone CJ, Hansen M, Kooperberg C, Truong YK (1997) Polynomial splines and their tensor products in extended linear modeling (with discussion). Annals Stat 25:1371–1470
Schwarz G (1978) Estimating the dimension of a model. Annals Stat 6:461–464
Zhang D, Lin X, Sowers M (2007) Two-stage functional mixed models for evaluating the effect of longitudinal covariate profiles on a scalar outcome. Biometrics 63:351–362
Zhou S, Shen X, Wolfe DA (1998) Local asymptotics for regression splines and confidence regions. Annals Stat 26:1760–1782
Zhou J, Chen M (2012) Spline estimators for semi-functional linear model. Stat Probab Lett 82:505–513