Stat Papers (2016) 57:827–841
DOI 10.1007/s00362-015-0681-3

REGULAR ARTICLE

Varying coefficient partially functional linear regression models

Qing-Yan Peng[1] · **Jian-Jun Zhou**[1] · **Nian-Sheng Tang**[1]

Received: 7 February 2014 / Revised: 14 February 2015 / Published online: 1 April 2015

Abstract By relaxing the linearity assumption in partial functional linear regression models, we propose a varying coefficient partially functional linear regression model (VCPFLM), which includes varying coefficient regression models and functional linear regression models as its special cases. We study the problem of functional parameter estimation in a VCPFLM. The functional parameter is approximated by a polynomial spline, and the spline coefficients are estimated by the ordinary least squares method. Under some regular conditions, we obtain asymptotic properties of functional parameter estimators, including the global convergence rates and uniform convergence rates. Simulation studies are conducted to investigate the performance of the proposed methodologies.

Keywords Functional linear models · Global convergence rate · Polynomial spline · Uniform convergence rate · Varying coefficient model

1 Introduction

Functional data that both predictor and response are random functions are often encountered in meteorology, medicine, biology, economy and finance (Ramsay and Silverman 2005). Due to its flexibility and interpretability, functional regression analysis has received a lot of attention in past years. For example, see Cardot et al. (1999, 2003), Chiou et al. (2003), Ramsay and Silverman (2005), Yao et al. (2005), Cai and Hall (2006), Hall and Horowitz (2007), Ferraty and Vieu (2006) and Baíllo and Grané (2009).

✉ Nian-Sheng Tang
nstang@ynu.edu.cn

1 Department of Statistics, Yunnan University, Kunming 650091, People's Republic of China

To improve power of prediction and interpretation of functional regression models, some additional real-valued predictors were introduced into functional regression models, which leaded to some new functional linear regression models. For example, Aneiros-Pérez and Vieu (2006) proposed a semi-functional partial linear regression model by combining the feature of a linear model together with the methodology for nonparametric treatment of functional data; Aneiros-Pérez and Vieu (2008) presented an extended semi-functional partial linear regression model for dependent data; Aneiros-Pérez and Vieu (2011) further proposed a fully automatic estimation procedure in a partial linear model with functional data; Zhang et al. (2007) developed a partial functional linear model by incorporating a parametric linear regression into functional linear regression models; Wong et al. (2008) proposed a functional-coefficient partially linear regression model by combining nonparametric and functional-coefficient regression model; Dabo-Niang and Guillas (2010) introduced a functional semiparametric model in which a real-valued random variable was explained by the sum of an unknown linear combination of the components of a multivariate random variable and an unknown transformation of a functional random variable and the random error was autocorrelated; Lian (2011) considered a functional partial linear model by taking advantage of both parametric and nonparametric functional models; Lian (2012) proposed an empirical likelihood approach to nonparametric functional regression and semi-functional partially linear model; Zhou and Chen (2012) introduced a semi-functional linear model by combining the feature of a functional linear regression model and a nonparametric regression model.

To broaden the applicability of functional linear regression models, a varying-coefficient functional linear regression model (VCFLRM) by allowing the slope function to depend on some additional scalar covariates was also proposed by Cardot and Sarda (2008) and has only received a little attention in recent years. For example, Wu et al. (2010) discussed estimation of the slope function in a VCFLRM based on functional principal components for sparse and irregular data, and investigated the asymptotic properties of the proposed estimators; Müller and Sentürk (2011) presented a review of statistical inference on VCFLRM. Inspired by the work of Cardot and Sarda (2008) and Wu et al. (2010), we consider a varying-coefficient partially functional linear regression model (VCPFLRM) by relaxing the linearity assumption in Zhang et al. (2007), which is an extension of partial functional linear regression models and varying-coefficient functional linear regression models.

Polynomial spline is a very popular smoothing technique in a nonparametric regression, and it enables us to extend the standard methods for parametric models to nonparametric settings and is easy to implement in applications, hence it is employed to approximate functional coefficients in our considered VCPFLRM. Based on the polynomial spline approximations to functional coefficients, we first use the least squares approach to estimate parameters in polynomial spline approximations and then obtain estimations of functional coefficients. Under some regular conditions, we discuss the global and uniform convergence rates of the proposed estimators.

The rest of this paper is organized as follows. Section 2 describes varying coefficient partially functional linear regression models and presents the polynomial spline estimators of functional coefficients. In Sect. 3, we study asymptotic properties of the proposed estimators. Simulation studies are conducted to investigate the performance of the proposed methods in Sect. 4. Technique details are given in the Appendix.

2 Model and estimation

Let Y be a real-valued response variable defined on a probability space $(\Omega, \mathfrak{B}, \mathcal{P})$, let U and $Z = (Z_1, \ldots, Z_p)^T$ be one-dimensional and p-dimensional vectors of explanatory variables defined on the same probability space, respectively. Also, let $\{X(t): t \in \mathcal{T}\}$ be a zero mean, second-order (i.e., $E|X(t)|^2 < \infty$ for all $t \in \mathcal{T}$) stochastic process defined on the probability space $(\Omega, \mathfrak{B}, \mathcal{P})$ with sample paths in $L_2(\mathcal{T})$, which represents the Hilbert space containing square integrable functions defined on $\mathcal{T}$ with inner product $\langle x, y\rangle = \int_{\mathcal{T}} x(t)y(t)dt$ for $\forall x, y \in L_2(\mathcal{T})$ and norm $\|x\|_2 = \langle x, x\rangle^{1/2}$. We assume that the relationship between Y and (X, U, Z) is given by

$$Y = \int_{\mathcal{T}} X(t)a_0(t)dt + a_1(U)Z_1 + \cdots + a_p(U)Z_p + \varepsilon, \tag{1}$$

where $a_0(t)$ and $a_j(U)$'s are unknown smooth functions for $j = 1, \ldots, p$, and ε is a random error with mean zero and finite variance σ^2 and is independent of (X, U, Z). Without loss of generality, we assume $\mathcal{T} = [0, 1]$. Clearly, the above defined model includes varying-coefficient models and functional linear regression models, which correspond to the cases that $a_0(t) = 0$ and $a_j(U) = 0$ for $j = 1, \ldots, p$, respectively. Also, it includes partial functional linear regression models (Zhang et al. 2007) when $a_j(U) \equiv \beta_j$ for $j = 1, \ldots, p$. Hence, the above defined model (1) is an extension of partial functional linear regression models and varying-coefficient models, and is referred to as a varying-coefficient partially functional linear regression model.

Let the data set $\{X_i, U_i, Z_i, Y_i\}_{i=1}^n$ be n independent realizations of $\{X, U, Z, Y\}$ generated from model (1), i.e.,

$$Y_i = \int_0^1 X_i(t)a_0(t)dt + a_1(U_i)Z_{i1} + \cdots + a_p(U_i)Z_{ip} + \varepsilon_i \quad \text{for} \quad i = 1, \ldots, n, \tag{2}$$

where the random errors ε_i's are independent and identically distributed with $E\varepsilon_i = 0$ and $E\varepsilon_i^2 = \sigma^2$, and are independent of (X_i, U_i, Z_i).

In what follows, we consider estimations of functional parameters a_j's for $j = 0, 1, \ldots, p$. Let $S_{k_0,N_{0n}}$ be the space of polynomial splines on [0, 1] with degree k_0 and N_{0n} knots $u_{0,1}, \ldots, u_{0,N_{0n}}$ satisfying

$$0 = u_{0,0} < u_{0,1} < \cdots < u_{0,N_{0n}} < u_{0,N_{0n}+1} = 1, \tag{3}$$

and $S_{k_j,N_{jn}}$ $(j = 1, \ldots, p)$ be the space of polynomial splines on $[a, b]$, which is the compact support of the density of U, with degree k_j and N_{jn} knots $u_{j,1}, \ldots, u_{j,N_{jn}}$ satisfying

$$a = u_{j,0} < u_{j,1} < \cdots < u_{j,N_{jn}} < u_{j,N_{jn}+1} = b, \tag{4}$$

where the numbers of knots N_{jn} ($j = 0, 1, \ldots, p$) increase when sample size n increases. The space of polynomial splines $S_{k_j,N_{jn}}$ is a linear space of K_{jn}-dimension, where $K_{jn} \equiv N_{jn}+k_j+1$ for $j = 0, 1, \ldots, p$. Generally, we can choose the truncated power basis and B-spline function as a basis of the above defined linear space. Due to some good numerical properties of B-spline function, we use the B-spline basis to approximate functional parameters a_j as follows. More details for spline function and spline space can refer to de Boor (2001) and Schumaker (1981).

Following the arguments of de Boor (2001), if the unknown function a_j ($j = 0, 1, \ldots, p$) is sufficiently smooth, there exists a spline function $\bar{a}_j$ in the linear space $S_{k_j,N_{jn}}$ such that

$$a_j \approx \bar{a}_j = \sum_{s=1}^{K_{jn}} b_{js} B_{js}, \tag{5}$$

where B_{js} denotes the B-spline function in the linear space $S_{k_j,N_{jn}}$ for $j = 0, 1, \ldots, p$. Thus, the model (2) can be approximated by

$$Y_i \approx \int_0^1 X_i(t)\bar{a}_0(t)dt + \bar{a}_1(U_i)Z_{i1} + \cdots + \bar{a}_p(U_i)Z_{ip} + \varepsilon_i. \tag{6}$$

Denote

$$l(b) = \sum_{i=1}^{n} \left\{ Y_i - \sum_{s=1}^{K_{0n}} b_{0s} < X_i, B_{0s} > - \sum_{j=1}^{p} \sum_{s'=1}^{K_{jn}} b_{js'} B_{js'}(U_i) Z_{ij} \right\}^2, \tag{7}$$

where $b = (b_0^T, b_1^T, \ldots, b_p^T)^T$ with $b_j = (b_{j1}, \ldots, b_{jK_{jn}})^T$ for $j = 0, 1, \ldots, p$. By minimizing $l(b)$ given in Equation (7), we can obtain the least squares estimator $\widehat{b} = (\widehat{b}_0^T, \widehat{b}_1^T, \ldots, \widehat{b}_p^T)^T$ of b, where $\widehat{b}_j = (\widehat{b}_{j1}, \ldots, \widehat{b}_{jK_{jn}})^T$ for $j = 0, 1, \ldots, p$. Thus, the polynomial spline estimators of functional parameters a_j's are given by $\widehat{a}_j = \sum_{s=1}^{K_{jn}} \widehat{b}_{js} B_{js}$. In this case, we can also define an estimator of variance σ^2 as $\widehat{\sigma}_n^2 = n^{-1} \sum_{i=1}^{n} \{Y_i - < X_i, \widehat{a}_0 > - \sum_{j=1}^{p} \widehat{a}_j(U_i) Z_{ij}\}^2$.

3 Asymptotic properties

In this section, we investigate asymptotic properties of the above proposed estimators. For simplicity, we first introduce the following notation. For two sequences of positive numbers c_n and d_n, $c_n \lesssim d_n$ represents that c_n/d_n is uniformly bounded, and $c_n \asymp d_n$ if and only if $c_n \lesssim d_n$ and $d_n \lesssim c_n$. The covariance operator Γ of a random function X is defined as $\Gamma x(t) = \int_0^1 EX(t)X(s)x(s)ds$ for $x \in L_2(\mathcal{T})$. $\|\cdot\|_\infty$ denotes the super-norm of function on some region D, that is, $\|r\|_\infty = \sup_{x \in D} |r(x)|$. Denote

$\mathcal{K}_n = \max_{j\in\{0,1,\cdots,p\}} K_{jn}$, $h_j = \max_{l=0,\cdots,N_{jn}}(u_{j,l+1} - u_{j,l})$, and let $C^q([a,b])$ be the collection of all functions that are q times continuously differentiable on $[a,b]$.

To study asymptotic properties of the above proposed estimators, we here assume that the degrees k_j's are fixed and the numbers of knots N_{jn}'s depend on sample size n. In addition, we also require the following assumptions.

(A1) For the knot sequences given in Eqs. (3) and (4), there exists some positive constant C_1 such that

$$\max_{j=0,1,\cdots,p} \frac{\max_{l=0,\cdots,N_{jn}}(u_{j,l+1} - u_{j,l})}{\min_{l=0,\cdots,N_{jn}}(u_{j,l+1} - u_{j,l})} \le C_1.$$

Also, $\mathcal{K}_n \asymp n^r$ for $0 < r < 1/3$, and $h_j \asymp \mathcal{K}_n^{-1}$ for $j = 0, 1, \ldots, p$.

(A2) The density function $f_U(u)$ of random variable U has a compact support $D_u = [a,b]$ and $f_U(u)$ is bounded away from zero and infinity on D_u.

(A3) $a_0(t) \in C^q([0,1])$, $a_j(u) \in C^q([a,b])$ for $j = 1, \ldots, p$ and $1 < q \le k$, where $k = \min_{j=0,1,\ldots,p} k_j$.

(A4) $\|X\|_2 \le C_2 < \infty$ a.s., and there is a positive constant C_3 such that $\langle \Gamma a_0^*, a_0^* \rangle \ge C_3\|a_0^*\|^2$ for any $a_0^* \in S_{k_0,N_{0n}}$, where C_2 is a positive constant and C_3 does not depend on n.

(A5) The eigenvalues of $E(Z_i^* Z_i^{*T} | X_i = x, U_i = u)$ are uniformly bounded away from zero and infinity for all $(x,u) \in L_2(\mathcal{T}) \times D_u$, where $Z_i^* = (1, Z_{i1}, \ldots, Z_{ip})^T$.

(A6) For some $m_0 > 2$, $E|Z_{1j}|^{m_0} < \infty$ for $j = 1, \ldots, p$.

Remark 1 Assumption (A1) is similar to Eq. (3) of Zhou et al. (1998) and Assumption (C3) of Xue and Yang (2006). Also, $\mathcal{K}_n$ stands for the growth rate of the dimension of the spline spaces relative to sample size. Assumption (A2) is very common in nonparametric regression, for example, see Condition 1 of Stone (1985) and Condition 2 of Chen (1991). Assumption (A3) ensures that $a_0(t)$ and $a_j(U)$ for $j = 1, \ldots, p$ are sufficiently smooth so that they can be approximated by spline functions. Assumption (A4) is a stronger condition than that given in functional linear regression models. Assumption (A5) is a generalization of Condition (ii) of Huang and Shen (2004) and Assumption (C2) of Xue and Yang (2006). Assumption (A6) is similar to Condition (v) of Huang and Shen (2004).

Under Assumptions (A1)–(A6), we obtain the following global and uniform convergence rates of the polynomial spline estimators.

Theorem 1 *Suppose that Assumptions* (A1)–(A6) *hold. Then, we have*

$$\|\widehat{a}_j - a_j\|_2^2 = O_p\left(\frac{\mathcal{K}_n}{n}\right) + O_p\left(\mathcal{K}_n^{-2q}\right) \quad \text{for} \quad j = 0, 1, \ldots, p.$$

Theorem 2 *Under Assumptions* (A1)–(A6)*, we have*

$$||\widehat{a}_j - a_j||_\infty = O_p\left(\mathcal{K}_n n^{-1/2}\right) + O_p(\mathcal{K}_n^{1/2-q}) \quad \text{for} \quad j = 0, 1, \ldots, p.$$

Remark 2 Theorem 1 shows the global convergence rates of the polynomial spline estimators of $a_0(t)$ and $a_j(U)$ for $j = 1, \ldots, p$, which are similar to Theorem 1 of Huang and Shen (2004) and Newey (1997) and Theorem 3.2 of Huang et al. (2004). Particularly, when $\mathcal{K}_n \asymp n^{1/(1+2q)}$, we have $\|\widehat{a}_j - a_j\|_2^2 = O_p(\mathcal{K}_n^{-2q/(1+2q)})$, which is the optimal global convergence rate given in Stone (1982). The uniform convergence rate in Theorem 2 is the same as Theorem 7 of Newey (1997). The above results indicate that the existence of a random function does not affect the convergence rates of the polynomial spline estimators of functional coefficients.

4 Simulation study

Experiment 1 To investigate the finite sample performance of our proposed methodologies, we conducted the first simulation study. In this simulation study, we generated data $\{X_i, U_i, Z_i, Y_i\}_{i=1}^n$ from the following model

$$Y_i = \int_0^1 X_i(t) a_0(t) dt + a_1(U_i) Z_{i1} + a_2(U_i) Z_{i2} + \varepsilon_i \quad \text{for} \quad i = 1, \ldots, n.$$

For functional linear components, similar to Lian (2011), we took $a_0(t) = \sum_{j=1}^{50} \kappa_j \phi_j(t)$ and $X_i(t) = \sum_{j=1}^{50} \xi_{ij} \iota_j \phi_j(t)$ with $\kappa_1 = 0.5$ and $\kappa_j = 4/j^2$ for $j = 2, \ldots, 50$, $\phi_1(t) = 1$ and $\phi_j(t) = \sqrt{2}\cos((j-1)\pi t)$ for $j = 2, \ldots, 50$, and $\iota_j = 1/j$ and ξ_{ij} was independently and uniformly distributed on the interval $[-\sqrt{3}, \sqrt{3}]$ for $j = 1, \ldots, 50$. For varying coefficient components, we set $a_1(U) = 0.138 + (0.316 + 0.982U)\exp(-3.89U^2)$, $a_2(U) = -0.437 - (0.659 + 1.260U)\exp(-3.89U^2)$, and U_i, Z_{i1} and Z_{i2} were simulated from the uniform distribution on the interval $[-0.5, 0.5]$ (Wong et al. 2008). Random errors ε_i's were independently generated from the normal distribution with mean zero and variance 0.2^2, i.e., $\varepsilon_i \sim N(0, 0.2^2)$.

To implement our proposed methods, we took $k_0 = 2$, $k_1 = 3$ and $k_2 = 3$ in using B-spline functions to approximate functions $a_0(t)$, $a_1(U)$ and $a_2(U)$, respectively. For the knot positions, we can uniformly take knots on the considered interval of random variable t (or U) or take the sample quantiles of random variable t (or U) to be their corresponding knots. For simplicity, we uniformly took knots $u_{0,h}$ on the interval $[0, 1]$ (i.e., $u_{0,h} = h/(N_{0n} + 1)$) for $h = 0, 1, \ldots, N_{0n} + 1$, and knots $u_{j,h}$ on the interval $[-0.5, 0.5]$ (i.e., $u_{j,h} = -0.5 + h/(N_{jn} + 1)$) for $j = 1, 2$ and $h = 0, 1, \ldots, N_{jn} + 1$. Thus, selecting the numbers of knots is equivalent to choosing the numbers of B-spline functions K_{0n}, K_{1n} and K_{2n} when k_0, k_1 and k_2 are fixed. Generally, AIC, BIC, the "leave-one-out" cross-validation (Rice and Silverman 1991) and the modified multi-fold cross-validation (Cai et al. 2000) can be used to select the required numbers of B-spline functions. Here, we used the "leave-one-out" cross-validation technique to choose the numbers of B-spline functions. Following Rice and Silverman (1991), K_{0n}, K_{1n} and K_{2n} can be selected by minimizing the following cross-validation score:

Table 1 Sample means, medians and variances of RASE and MSEP

	$n = 200$			$n = 500$		
	Mean	Median	Var	Mean	Median	Var
RASE_0	0.1652	0.1251	0.0099	0.1156	0.0972	0.0038
RASE_1	0.1258	0.1153	0.0039	0.0758	0.0704	0.0010
RASE_2	0.1281	0.1166	0.0038	0.0767	0.0721	0.0009
RASE	0.4191	0.3933	0.0183	0.2681	0.2510	0.0061
MSEP	0.0039	0.0036	2.839×10^{-6}	0.0016	0.0015	0.430×10^{-7}

$$CV(K_{0n}, K_{1n}, K_{2n}) = \frac{1}{n}\sum_{i=1}^{n}\left\{Y_i - < X_i, \widehat{a_0}^{-i} > - \widehat{a_1}^{-i}(U_i)Z_{i1} - \widehat{a_2}^{-i}(U_i)Z_{i2}\right\}^2,$$

where $\widehat{a_0}^{-i}(t)$, $\widehat{a_1}^{-i}(U)$ and $\widehat{a_2}^{-i}(U)$ are respectively estimators of $a_0(t)$, $a_1(U)$ and $a_2(U)$, and are evaluated by deleting the ith observation $\{X_i, U_i, Z_i, Y_i\}$ from the full data set $\{(X_j, U_j, Z_j, Y_j) \colon j = 1, \ldots, n\}$.

To assess the performance of our proposed estimators, we computed the mean square error of prediction (MSEP) of response variable Y (Cardot et al. 2003), which is defined by

$$\text{MSEP} = \frac{1}{n}\sum_{i=1}^{n}\left\{\widehat{Y}_i - < X_i, a_0 > - a_1(U_i)Z_{i1} - a_2(U_i)Z_{i2}\right\}^2,$$

and the square-root of average squared error (RASE) of functional parameters $a_0(\cdot)$, $a_1(\cdot)$ and $a_2(\cdot)$ (Huang and Shen 2004), which is defined by

$$\text{RASE} = \sum_{j=0}^{2}\text{RASE}_j \quad \text{with} \quad \text{RASE}_j = \left\{\frac{1}{n_j}\sum_{h=1}^{n_j}(\widehat{a}_j(t_h) - a_j(t_h))^2\right\}^{1/2}$$

where t_h's are the regular grid points for $h = 1, \ldots, n_j$.

In this simulation study, we considered two different sample sizes: $n = 200$ and 500. For each sample size, results were obtained via 1000 replications. Table 1 presented sample means, medians and variances of RASE, RASE_j and MSEP. Figures 1, 2, and 3 displayed the polynomial spline estimates of $a_0(t)$, $a_1(U)$ and $a_2(U)$ for a special replication corresponding to minimum of RASE. From Table 1, we observed that the sample mean, median and variance of RASE, RASE_j and MSEP decrease as sample size increases. Also, from Figures 1, 2, and 3, we observed that estimates of $a_0(t)$, $a_1(U)$ and $a_2(U)$ become more and more accurate as sample size increases. All these findings showed that our proposed estimation procedure performs well under our considered settings.

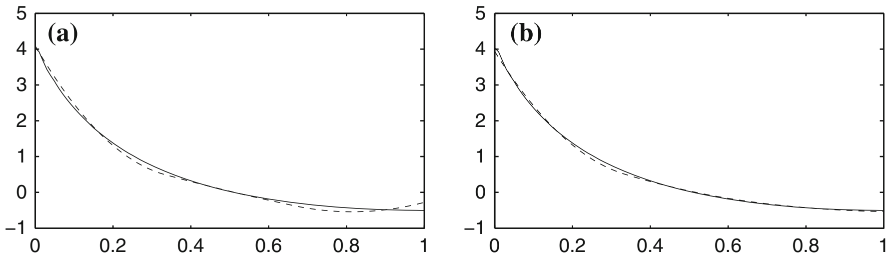

Fig. 1 The true $a_0(t)$ (*solid curve*) and its polynomial spline estimation $\widehat{a}_0(t)$ (*dash curve*) under sample sizes: **a** $n = 200$ and **b** $n = 500$.

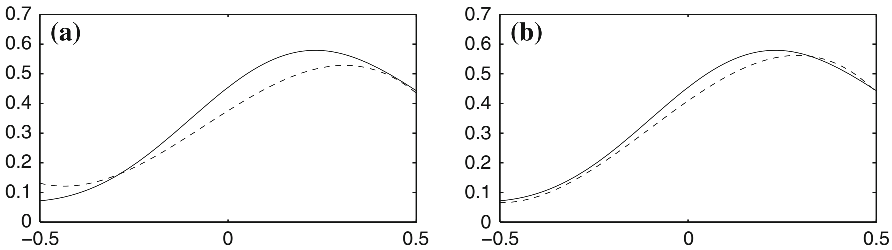

Fig. 2 The true $a_1(U)$ (*solid curve*) and its polynomial spline estimation $\widehat{a}_1(U)$ (*dash curve*) under sample sizes: **a** $n = 200$ and **b** $n = 500$.

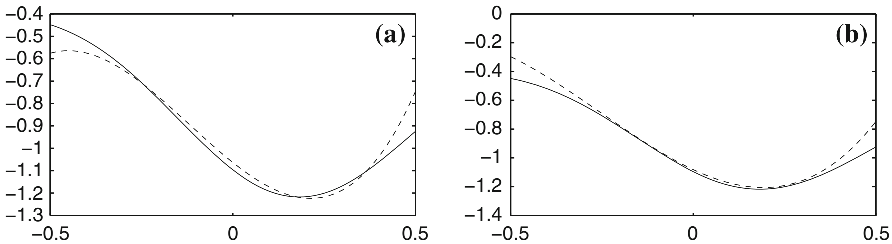

Fig. 3 The true $a_2(U)$ (*solid curve*) and its polynomial spline estimation $\widehat{a}_2(U)$ (*dash curve*) under sample sizes: **a** $n = 200$ and **b** $n = 500$.

Experiment 2 To compare the performance of our proposed estimators with those obtained from a partial functional linear model, we conducted the following two simulation studies in this experiment.

In the second simulation study, the observed data $\{X_i, U_i, Z_i, Y_i\}_{i=1}^n$ were generated from the following partial functional linear regression model (PFLRM):

$$Y_i = \int_0^1 X_i(t)a_0(t)dt + a_1(U)Z_{i1} + a_2(U)Z_{i2} + \varepsilon_i$$

with the same settings as given in the first simulation study except that $a_1(U)$ and $a_2(U)$ were taken to be $a_1(U) \equiv 1.5$ and $a_2(U) \equiv -1$ and $\varepsilon_i \overset{\text{i.i.d.}}{\sim} N(0, 0.5^2)$ for $i = 1, \ldots, n$.

Table 2 Mean and standard deviation (SD) of MSEP in experiment 2

Fitted Model	Simulation 2 (True model: PFLM)				Simulation 3 (True model: VCPFLM)			
	$n = 200$		$n = 500$		$n = 200$		$n = 500$	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
PFLM	0.0119	0.0062	0.0049	0.0023	0.1991	0.0170	0.1977	0.0105
VCPFLM	0.0252	0.0115	0.0106	0.0049	0.0244	0.0113	0.0105	0.0041

In the third simulation study, the observed data $\{X_i, U_i, Z_i, Y_i\}_{i=1}^n$ were generated from the following VCPFLRM:

$$Y_i = \int_0^1 X_i(t)a_0(t)dt + a_1(U_i)Z_{i1} + a_2(U_i)Z_{i2} + \varepsilon_i$$

with the same settings as given in the first simulation study except that $a_1(U)$ and $a_2(U)$ were taken to be $a_1(U) = \sin(2\pi U)$ and $a_2(U) = 4U(1-U)$, $U_i \overset{\text{i.i.d.}}{\sim} \text{Uniform}(0, 1)$, $Z_{ij} \overset{\text{i.i.d.}}{\sim} \text{Uniform}(-1, 1)$ for $j = 1$ and 2, and $\varepsilon_i \overset{\text{i.i.d.}}{\sim} N(0, 0.5^2)$ for $i = 1, \ldots, n$.

For each of the above two simulation studies, 500 data sets were generated and were fitted to PFLRM and our proposed VCPFLRM with the same selections of k_0, k_1, k_2 and knots as given in the first simulation study, respectively. Results corresponding to MSEP for $n = 200$ and 500 were presented in Table 2. From Table 2, we observed that (i) using our proposed VCPLRM to fit PFLRM data has the same performance as using PFLRM to fit PFLRM data in terms of their means and standard deviations of MSEP; (ii) using PFLRM to fit VCPLRM data may yield a relatively large mean value of MSEP, which indicated that our proposed VCPFLRM behaves better than PFLRM under misspecified functional linear models; (iii) increasing sample size can reduce standard deviation of MSEP.

Acknowledgments The authors are grateful to the Editor and two referees for their valuable suggestions that greatly improved the manuscript. The work was supported by National Nature Science Foundation of China (Grant Nos. 11225103, 11301464), Research Fund for the Doctoral Program of Higher Education of China (20115301110004) and the Scientific Research Foundation of Yunnan Provincial Department of Education (No. 2013Y360).

Appendix: Proofs of Theorems

Denote $B_{js} = K_{jn}^{1/2}\phi_{js}$, where ϕ_{js}'s are the normalized B-splines in the space $S_{k_j,N_{jn}}$ for $s = 1, \ldots, K_{jn}$ and $j = 0, 1, \ldots, p$. It follows from Theorem 4.2 of Chapter 5 in DeVore and Lorentz (1993) that for any spline function $\sum_{s=1}^{K_{jn}} b_{js}B_{js}$, there are positive constants M_1 and M_2 such that

$$M_1|b_j|_2^2 \le \int \left\{ \sum_{s=1}^{K_{jn}} b_{js} B_{js} \right\}^2 \le M_2|b_j|_2^2, \tag{8}$$

where $|\cdot|_2$ is Euclidean norm.

Define $\mathbf{B} = (\mathbf{X}, \mathbf{Z})$, where

$$\mathbf{X} = \begin{pmatrix} < X_1, B_{01} > & < X_1, B_{02} > & \dots & < X_1, B_{0K_{0n}} > \\ < X_2, B_{01} > & < X_2, B_{02} > & \dots & < X_2, B_{0K_{0n}} > \\ \vdots & \vdots & \ddots & \vdots \\ < X_n, B_{01} > & < X_n, B_{02} > & \dots & < X_n, B_{0K_{0n}} > \end{pmatrix},$$

$$\mathbf{Z} = \begin{pmatrix} B_{11}(U_1)Z_{11} & \dots & B_{1K_{1n}}(U_1)Z_{11} & \dots & B_{p1}(U_1)Z_{1p} & \dots & B_{pK_{pn}}(U_1)Z_{1p} \\ B_{11}(U_2)Z_{21} & \dots & B_{1K_{1n}}(U_2)Z_{21} & \dots & B_{p1}(U_2)Z_{2p} & \dots & B_{pK_{pn}}(U_2)Z_{2p} \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ B_{11}(U_n)Z_{n1} & \dots & B_{1K_{1n}}(U_n)Z_{n1} & \dots & B_{p1}(U_n)Z_{np} & \dots & B_{pK_{pn}}(U_n)Z_{np} \end{pmatrix}.$$

To prove Theorems 1 and 2, we require the following Lemmas.

Lemma 1 *If Assumptions* (A1)–(A6) *hold, we have*

$$\sup_{\substack{a_j \in S_{k_j, N_{jn}} \\ j=0,1,\dots,p}} \left| \frac{\frac{1}{n}\sum_{i=1}^n \left\{ < X_i, a_0 > + \sum_{j=1}^p a_j(U_i)Z_{ij} \right\}^2}{E(< X_1, a_0 > + \sum_{j=1}^p a_j(U_1)Z_{1j})^2} - 1 \right| = o_p(1).$$

Proof For an i.i.d. random variable sequence $\xi_1, \dots, \xi_n$, let $E_n(\xi_i) = \frac{1}{n}\sum_{i=1}^n \xi_i$. By Assumptions (A2)–(A5), we have

$$E\left(< X_1, a_0 > + \sum_{j=1}^p a_j(U_1)Z_{1j} \right)^2 \asymp E\left(< X_1, a_0 >^2 + \sum_{j=1}^p a_j^2(U_1) \right) \asymp \sum_{j=0}^p ||a_j||_2^2.$$

Consequently, we only need to prove that for arbitrary given $\eta > 0$, as $n \to \infty$, we have

$$\mathbb{I} = P\left\{ \sup_{\substack{a_j \in S_{k_j, N_{jn}} \\ j=0,1,\dots,p}} \frac{\left| (E_n - E)\left[< X_i, a_0 > + \sum_{j=1}^p a_j(U_i)Z_{ij} \right]^2 \right|}{\sum_{j=0}^p ||a_j||_2^2} > (p+1)\eta \right\} \to 0.$$

If $|(E_n - E) < X_i, a_0 >^2 | \le \eta ||a_0||_2^2$, $|(E_n - E)\left\{ < X_i, a_0 > a_j(U_i)Z_{ij} \right\}| \le \eta ||a_0||_2$ $||a_j||_2$ for $j = 1, \dots, p$, and $|(E_n - E)\{a_j(U_i)a_{j'}(U_i)Z_{ij}Z_{ij'}\}| \le \eta ||a_j||_2 ||a_{j'}||_2$ for $j', j = 1, \dots, p$, we obtain

$$|\,(E_n-E)\left\{< X_i, a_0 > + \sum_{j=1}^{p} a_j(U_i)Z_{ij}\right\}^2 | \le \eta \left(\sum_{j=0}^{p} ||a_j||_2)^2 \le (p+1\right) \eta \sum_{j=0}^{p} ||a_j||_2^2.$$

Thus, we have

$$\begin{aligned}
\mathbb{I} \le\; & P\left\{\sup_{a_0\in S_{k_0,N_{jn}}} \frac{|(E_n-E) < X_i, a_0 >^2 |}{||a_0||_2^2} > \eta\right\} \\
& + 2\sum_{j=1}^{p} P\left\{\sup_{\substack{a_j\in S_{k_j,N_{jn}}\\ a_0\in S_{k_0,N_{0n}}}} \frac{|(E_n-E)[< X_i, a_0 > a_j(U_i)Z_{ij}]|}{||a_0||_2||a_j||_2} > \eta\right\} \\
& + \sum_{j=1}^{p}\sum_{j'=1}^{p} P\left\{\sup_{\substack{a_j\in S_{k_j,N_{jn}}\\ a_{j'}\in S_{k_{j'},N_{j'n}}}} \frac{|(E_n-E)[a_j(U_i)a_{j'}(U_i)Z_{ij}Z_{ij'}]|}{||a_j||_2||a_{j'}||_2} > \eta\right\} \\
\triangleq\; & I_1 + 2I_2 + I_3.
\end{aligned}$$

For I_1, it follows from Lemma 5.2 of Cardot et al. (1999) that $I_1 \to 0$ as $n \to \infty$. Following the similar argument of Lemma 1 in Huang and Shen (2004), it is easily shown that $I_3 \to 0$ as $n \to \infty$. Consequently, we only need to prove that for $j = 1, \cdots, p$,

$$\mathbb{I}_j = P\left\{\sup_{\substack{a_j\in S_{k_j,N_{jn}}\\ a_0\in S_{k_0,N_{0n}}}} \frac{|(E_n-E)[< X_i, a_0 > a_j(U_i)Z_{ij}]|}{||a_0||_2||a_j||_2} > \eta\right\} \to 0 \text{ as } n \to \infty.$$

Note that $< X_i, a_0 > a_j(U_i)Z_{ij} = \sum_{s_0=1}^{K_{0n}} \sum_{s_j=1}^{K_{jn}} b_{0s_0}b_{js_j} < X_i, B_{0s_0} > B_{js_j}(U_i)Z_{ij}$ for $j = 1, \ldots, p$. Hence, if $|(E_n-E)\{< X_i, B_{0s_0} > B_{js_j}(U_i)Z_{ij}\}| \le \eta$ for $s_0 = 1, \ldots, K_{0n}$ and $s_j = 1, \ldots, K_{jn}$, it follows from the Cauchy-Schwarz inequality and Eq. (8) that

$$\begin{aligned}
|(E_n-E)[< X_i, a_0 > a_j(U_i)Z_{ij}]| &\le \eta \sum_{s_0=1}^{K_{0n}} \sum_{s_j=1}^{K_{jn}} |b_{0s_0}||b_{js_j}| \\
&\le K_{0n}^{1/2} K_{jn}^{1/2} (\sum_{s_0=1}^{K_{0n}} b_{0s_0}^2)^{1/2} (\sum_{s_j=1}^{K_{jn}} b_{js_j}^2)^{1/2} \\
&\le C\eta\mathcal{K}_n ||a_0||_2||a_j||_2.
\end{aligned}$$

Thus, we have

$$\mathbb{I}_j \leq \sum_{s_0=1}^{K_{0n}} \sum_{s_j=1}^{K_{jn}} P\left\{|(E_n - E)\left(< X_i, B_{0s_0} > B_{js_j}(U_i)Z_{ij}\right)| > \eta/C\mathcal{K}_n\right\}. \quad (9)$$

Denote $\tilde{Z}_{ij} = Z_{ij}I(|Z_{ij}| \leq n^{\delta})$ for $j = 1, \ldots, p$, and we assume $m_0 > \delta^{-1}$ with $\delta > 0$. It follows from condition (A6) that as $n \to \infty$, we have

$$P\{\exists\, i = 1, \ldots, n \text{ such that } Z_{ij} \neq \tilde{Z}_{ij}\} \leq \sum_{i=1}^{n} P\{|Z_{ij}| > n^{\delta}\} \leq \frac{E|Z_{1j}|^{m_0}}{n^{m_0\delta-1}} \to 0.$$

Combining condition (A1) and Eq. (9) yields

$$\mathbb{I}_j \lesssim n^{2r} \max_{\substack{s_0=1,\ldots,K_{0n},\\ s_j=1,\ldots,K_{jn}}} P\left\{|(E_n - E)\left(< X_i, B_{0s_0} > B_{js_j}(U_i)Z_{ij}\right)| > \eta/C\mathcal{K}_n\right\}.$$

From Lemma A.8 of Ferraty and Vieu (2006), we have

$$\mathbb{I}_j \lesssim n^{2r}\exp(-C\eta^2 n^{1-(2\delta+3r)}).$$

Since $\delta^{-1} < m_0$ and $0 < r < 1/3$, we can always find $\delta > 0$ and $r > 0$ such that $2\delta + 3r < 1$. Hence, as $n \to \infty$, we have $\mathbb{I}_j \to 0$ for $j = 1, \ldots, p$. Combining the above equations leads to Lemma 1. □

Lemma 2 *If Assumptions* (A1)–(A6) *hold, there is an interval* $[M_3, M_4]$ *with* $0 < M_3 < M_4$ *such that as* $n \to \infty$*, we have*

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

Proof The proof of Lemma 2 is similar to that given in Lemma 2 of Huang and Shen (2004). Hence, we here omit it. □

Lemma 2 shows that the convergence rate of estimator $\widehat{b}$ does not depend on the eigenvalues of the covariance operator Γ of X. Thus, it follows from Cardot et al. (2003) that the convergence rate of our proposed estimator can attain the nonparametric convergence rate.

Proof of Theorem 1 Denote $\tilde{Y}_i =< X_i, a_0 > + \sum_{j=1}^{p} a_j(U_i)Z_{ij}$ and $\tilde{Y} = (\tilde{Y}_1, \cdots, \tilde{Y}_n)^T$. Let $\tilde{b} = (\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T\tilde{Y}$, where $\tilde{b} = (\tilde{b}_0^T, \tilde{b}_1^T, \ldots, \tilde{b}_p^T)^T$ with $\tilde{b}_j = (\tilde{b}_{j1}, \ldots, \tilde{b}_{jK_{jn}})^T$ for $j = 0, 1, \ldots, p$. Denote $\tilde{a}_j = \sum_{s=1}^{K_{jn}} \tilde{b}_{js}B_{js}$ and $\varepsilon = (\varepsilon_1, \cdots, \varepsilon_n)^T$. Under the above notation, it follows from Lemma 2 that

$E|\widehat{b}-\tilde{b}|^2 = E(\varepsilon^T \mathbf{B}(\mathbf{B}^T\mathbf{B})^{-1}(\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T\varepsilon) = \frac{\sigma^2}{n}E(\mathrm{tr}(\frac{1}{n}\mathbf{B}^T\mathbf{B})^{-1}) \lesssim \mathcal{K}_n/n$. Hence, it follows from Eq. (8) that

$$\sum_{j=0}^{p} ||\widehat{a}_j - \tilde{a}_j||_2^2 \asymp |\widehat{b}-\tilde{b}|^2 = O_p(\frac{\mathcal{K}_n}{n}). \tag{10}$$

Again, it follows from condition (A1) and Theorem XII.1 of de Boor (2001) that for $j = 0, 1, \ldots, p$, there exist spline function $a_j^* \in S_{k_j, N_{jn}}$ and constant $C_j > 0$ such that

$$||a_j^* - a_j||_\infty \le C_j h_j^q \lesssim \mathcal{K}_n^{-q}. \tag{11}$$

Let $b^* = (b_0^{*T}, b_1^{*T}, \ldots, b_p^{*T})^T$ with $b_j^* = (b_{j1}^*, \ldots, b_{jK_{jn}}^*)^T$, and $a_j^* = \sum_{s=1}^{K_{jn}} b_{js}^* B_{js}$ for $j = 0, 1, \ldots, p$. It follows from Equation (8) and Lemma 2 that $\sum_{j=0}^{p} ||a_j^* - \tilde{a}_j||_2^2 \asymp |b^* - \tilde{b}|^2 \asymp \frac{1}{n}(\tilde{b}-b^*)^T\mathbf{B}^T\mathbf{B}(\tilde{b}-b^*)$ a.s.. Since $\mathbf{B}(\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T$ is an orthogonal projection matrix, we have

$$\begin{aligned}\frac{1}{n}(\tilde{b}-b^*)^T\mathbf{B}^T\mathbf{B}(\tilde{b}-b^*) &\le \frac{1}{n}|\tilde{Y} - \mathbf{B}b^*|^2 \\ &\le \frac{1}{n}\sum_{i=1}^{n}\{< X_i, a_0 - a_0^* > + \sum_{j=1}^{p}[a_j(U_i) - a_j^*(U_i)]Z_{ij}\}^2.\end{aligned}$$

By Assumptions (A2)–(A4) and Eq. (11), we obtain

$$E\{< X_1, a_0 - a_0^* > + \sum_{j=1}^{p}(a_j(U_1) - a_j^*(U_1))Z_{1j}\}^2 \asymp \sum_{j=0}^{p}||a_j - a_j^*||_2^2 = O_p(\mathcal{K}_n^{-2q}). \tag{12}$$

For $j = 0, 1, \ldots, p$, we can obtain

$$||\widehat{a}_j - a_j||_2^2 \le 3(||\widehat{a}_j - \tilde{a}_j||_2^2 + ||\tilde{a}_j - a_j^*||_2^2 + ||a_j^* - a_j||_2^2). \tag{13}$$

Combining Eqs. (10)–(13) yields Theorem 1. □

Proof of Theorem 2 For $j = 0, 1, \ldots, p$, we have

$$||\widehat{a}_j - a_j||_\infty \le ||\widehat{a}_j - \tilde{a}_j||_\infty + ||\tilde{a}_j - a_j^*||_\infty + ||a_j^* - a_j||_\infty, \tag{14}$$

where $\tilde{a}_j$ and a_j^* are defined in the proof of Theorem 1. Also, it follows from Huang et al. (2004) that there is a constant $M > 0$ such that

$$||g_j||_\infty \le M\sqrt{K_{jn}}||g_j||_2 \tag{15}$$

for $g_j \in S_{k_j,N_{jn}}$ $(j = 0, 1, \ldots, p)$. Hence, by condition (A1), (10), (13) and (15), we obtain

$$||\widehat{a}_j - \tilde{a}_j||_\infty \leq M\sqrt{K_{jn}}||\widehat{a}_j - \tilde{a}_j||_2 = O_p(\mathcal{K}_n n^{-1/2}),$$

$$||\tilde{a}_j - a_j^*||_\infty \leq M\sqrt{K_{jn}}||\tilde{a}_j - a_j^*||_2 = O_p(\mathcal{K}_n^{1/2-q}).$$

Again, it follows from Eq. (11) that $||a_j^* - a_j||_\infty = O(\mathcal{K}_n^{-q}) = o(\mathcal{K}_n^{1/2-q})$. Therefor, combining the above equations leads to Theorem 2. □

References

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
Aneiros-Pérez G, Vieu P (2011) Automatic estimation procedure in partial linear model with functional data. Stat Pap 52:751–771
Baíllo A, Grané A (2009) Local linear regression for functional predictor and scalar response. J Multivar Anal 100:102–111
Cai TT, Hall P (2006) Prediction in functional linear regression. Ann 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 Probab 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 37:3186–3203
Chen H (1991) Polynomial splines and nonparametric regression. J Nonparametric Stat 1:143–156
Chiou JM, Müller H, Wang J (2003) Functional quasi-likelihood regression models with smooth random effects. J R Stat Soc B 65:405–423
Dabo-Niang S, Guillas S (2010) Functional semiparametric partially linear model with autoregression errors. J Multivar Anal 101:307–315
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
Hall P, Horowitz J (2007) Methodology and convergence rates for functional linear regression. Ann Stat 35:70–91
Huang JZ, Wu CO, Zhou L (2004) Polynomial spline estimation and inference for varying coefficient model with longitudinal data. Stat Sin 14:763–788
Huang JZ, Shen H (2004) Functional coefficient regression models for non-linear time series: a polynomial spline approach. Scand J Stat 31:515–534
Lian H (2011) Functional partial linear model. J Nonparametric Stat 23:115–128
Lian H (2012) Empirical likelihood confidence intervals for nonparametric functional data analysis. J Stat Plan Inference 142:1669–1677
Müller HG, Sentürk D (2011) Functional varying coefficient models. Recent advances in functional data analysis and related topics. In: Ferraty F (ed) Contributions to statistics. Springer, Berlin Heidelbeg, pp 225–230
Newey WK (1997) Convergence rates and asymptotic normality for series estimators. J Econ 79:147–168
Ramsay J, Silverman B (2005) Functional data analysis, 2nd edn. 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
Schumaker LL (1981) Splines function: basic theory. Wiley, New York
Stone CJ (1982) Optimal global rates of convergence for nonparametric regression. Ann Stat 10:1040–1053
Stone CJ (1985) Additive regression and other nonparametric models. Ann Stat 13:689–705

Wong H, Zhang RQ, Ip W, Li GY (2008) Functional-coefficient partially linear regression model. J Multivariate Anal 99:278–305

Wu Y, Fan J, Müller H (2010) Varying-coefficient functional linear regression. Bernoulli 16:730–758

Xue L, Yang L (2006) Additive coefficient modeling via polynomial spline. Stat Sin 16:1423–1446

Yao F, Müller HG, Wang JL (2005) Functional linear regression analysis for longitudinal data. Ann Stat 33:2873–2903

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. Ann Stat 26:1760–1782

Zhou J, Chen M (2012) Spline estimators for semi-functional linear model. Stat Probab Lett 82:505–513