

---

---

## AN RKHS FRAMEWORK FOR SPARSE FUNCTIONAL VARYING COEFFICIENT MODEL

---

---

Authors: BEHDAD MOSTAFAIY

– Department of Statistics, Faculty of Mathematical Science,  
Shahid Beheshti University, Iran  
behdad.mostafaiy@gmail.com

MOHAMMAD REZA FARIDROHANI

– Department of Statistics, Faculty of Mathematical Science,  
Shahid Beheshti University, Iran  
m\_faridrohani@sbu.ac.ir

S. MOHAMMAD E. HOSSEININASAB

– Department of Statistics, Faculty of Mathematical Science,  
Shahid Beheshti University, Iran  
m\_hosseininasab@sbu.ac.ir

Received: July 2014

Revised: November 2014

Accepted: February 2015

Abstract:

- We study functional varying coefficient model in which both the response and the predictor are functions of a common variable such as time. We demonstrate the estimation of the slope function for the case of sparse and noise-contaminated longitudinal data. So far, a few methods have been introduced based on varying coefficient model. To estimate the slope function, we consider a regularization method using a reproducing kernel Hilbert space framework. Despite the generality of the regularization method, the procedure is easy to implement. Our numerical results show that the introduced procedure performs well in some senses.

Key-Words:

- *functional varying coefficient model; regularization; reproducing kernel Hilbert space; sparsity.*

AMS Subject Classification:

- 62G05, 62P10.



---

## 1. INTRODUCTION

---

Due to rapid development of science and technology, it is possible to collect data that are naturally functions. This type of data, that are referred as functional data, has many applications in various fields of science, including, for example, environmental science, chemometrics, engineering, biomedical studies, public health, and econometrics. Functional data analysis deals with situations in which the individual observed data are infinite-dimensional, such as curves. See Ramsay and Silverman (2002, 2005) and Ferraty and Vieu (2006) for comprehensive discussions on methods and applications for functional data.

Functional linear model is one of the most useful methods to explore the relationship between two sets of observations. There are various types of functional linear model that have been widely studied in the literature. In this paper, we consider a functional linear model where one observe a random sample  $\{(X_i, Y_i) : i = 1, 2, \dots, n\}$  corresponds to functional varying coefficient model, i.e.,

$$(1.1) \quad Y(t) = \alpha(t) + \beta(t)X(t) + Z(t),$$

where  $\alpha$  and  $\beta$  are smoothed functions, and  $Z(t)$  is a noise term with zero mean and finite variance. Without loss of generality, we assume that  $E[X(t)] = E[Y(t)] = 0$ , then the functional linear model (1.1) becomes,

$$(1.2) \quad Y(t) = \beta(t)X(t) + Z(t).$$

Relation (1.2) models  $Y$  via  $X$  pointwisely, and allows  $\beta$  to vary with time. Fan and Zhang (2008) have provided a review of statistical methods proposed for various varying coefficient models according to three approaches. These approaches are based on polynomial spline, smoothing splines and local polynomial smoothing. See also Wu *et al.* (1998), Huang *et al.* (2002, 2004), Hoover *et al.* (1998), Chiang *et al.* (2001), Wu and Chiang (2000), and Kauermann and Tutz (1999). Fan and Zhang (1999), Wang and Xia (2009), and Lin and Ying (2001) applied another approaches for varying coefficient models. Most of these papers did not examine sparse and irregular designs and face some problems in implementing these designs.

In many experiments though, for example most longitudinal studies, the functional trajectories of the involved smooth random processes are not directly observable. In these cases, the observed data are noisy, sparse and irregularly spaced measurements of these trajectories.

Following the notation in Yao *et al.* (2005a), let  $U_{ij}$  and  $V_{ij}$  the  $j$ th observations of the random trajectories  $X_i(\cdot)$  and  $Y_i(\cdot)$  at a random time points  $T_{ij}$ , respectively, where  $T_{ij}$  are independently drawn from a distribution on compact domain  $T \subset \mathbb{R}$ . Assume that  $U_{ij}$  and  $V_{ij}$  are contaminated with measurement errors  $\varepsilon_{ij}$  and  $\epsilon_{ij}$ , respectively. These errors are assumed to be i.i.d. with mean

zero and finite variance  $\sigma_X^2$  for  $\varepsilon_{ij}$  and  $\sigma_Y^2$  for  $\epsilon_{ij}$ . Therefore, the models may be represented in the following forms:

$$(1.3) \quad \begin{aligned} U_{ij} &= X_i(T_{ij}) + \varepsilon_{ij}, & j &= 1, \dots, m; & i &= 1, \dots, n, \\ V_{ij} &= Y_i(T_{ij}) + \epsilon_{ij}, & j &= 1, \dots, m; & i &= 1, \dots, n. \end{aligned}$$

Functional data analysis of model (1.3) has been extended by Yao et. al (2005a, 2005b). See also Li and Hsing (2010), and Yang et. al (2011). Şentürk and Müller (2010), and Şentürk and Nguyen (2011) have considered functional varying coefficient in model (1.1). The model given in Şentürk and Müller (2010) is a model with one covariate process that incorporates a history index. Their estimation approach is based on least square estimation. Şentürk and Nguyen (2011) have studied a model with error-prone time-dependent variables and time-invariant covariates. They used covariance representation techniques to estimate the slope function. More references that studied varying coefficient models for model (1.3) include Şentürk and Müller (2008), Noh and Park (2010), Chiou et al. (2012), and Şentürk et al. (2013).

In this paper, we assume that the slope function  $\beta$  belongs to a reproducing kernel Hilbert space (RKHS)  $\mathcal{H}$ , and investigate the regularization method for estimating  $\beta$ . By simulation, we show that our estimation method perform well as sampling frequency and sample size increase. We do our simulation study in two different settings. One is when locations are same and equidistant for all curves, that is,  $T_{1j} = T_{2j} = \dots = T_{nj} = \frac{2j}{2m+1}$  for all  $j = 1, 2, \dots, m$ . Another setting is when  $T_{ij}$  are independently sampled from  $T$ . These settings are referred as *common design* and *stochastic design*, respectively (see Cai and Yuan (2011)).

The paper is organized as follows. In section 2, our estimation procedure is introduced. The numerical results are given in Section 3. Section 4 collects the obtained results and discusses possible extensions of our work.

---

## 2. ESTIMATION PROCEDURE

---

In this section, we introduce a regularization method for estimating the slope function  $\beta$  using a reproducing kernel Hilbert space (RKHS) framework. First, we review some basic facts of RKHS. A Hilbert space  $\mathcal{H}$  of functions on a set  $T$  with inner product  $\langle \cdot, \cdot \rangle_{\mathcal{H}}$  is called an RKHS if there exists a bivariate function  $K(\cdot, \cdot)$  on  $T \times T$  such that for every  $t \in T$  and  $f \in \mathcal{H}$ ,

- (i)  $K(\cdot, t) \in \mathcal{H}$ ,
- (ii)  $f(t) = \langle f, K(\cdot, t) \rangle_{\mathcal{H}}$ .

Relation (ii) is termed the reproducing property of  $K$ , and  $K$  is called reproducing kernel of  $\mathcal{H}$ . Every reproducing kernel determine unique RKHS. In addition, an RKHS has unique reproducing kernel. For any  $s_1, \dots, s_{m'}, s'_1, \dots, s'_{n'} \in T$  and  $a_1, \dots, a_{m'}, b_1, \dots, b_{n'} \in \mathbb{R}$ , we have

$$(2.1) \quad \left\langle \sum_{i=1}^{m'} a_i K(\cdot, s_i), \sum_{j=1}^{n'} b_j K(\cdot, s'_j) \right\rangle_{\mathcal{H}} = \sum_{i=1}^{m'} \sum_{j=1}^{n'} a_i b_j K(s_i, s'_j)$$

More details on RKHS can be found in Aronszajn (1950), Berlinet and Thomas-Agnan (2004) and Wahba (1990).

Now, we investigate the method of regularization to estimate  $\beta$ . We assume that  $\beta \in \mathcal{H}(K)$ , where  $\mathcal{H}(K)$  is an RKHS with reproducing kernel  $K$ . We estimate  $\beta$  via

$$(2.2) \quad \hat{\beta}_\lambda = \arg \min_{\beta \in \mathcal{H}(K)} \left\{ \ell_{mn}(\beta) + \lambda \|\beta\|_{\mathcal{H}(K)}^2 \right\}$$

where

$$\ell_{mn}(\beta) = \frac{1}{nm} \sum_{i=1}^n \sum_{j=1}^m (V_{ij} - U_{ij}\beta(T_{ij}))^2$$

and  $\lambda > 0$  is tuning parameter that control tradeoff between fidelity to the data measured by  $\ell_{mn}$  and smoothness of the solution measured by RKHS norm.

**Remark 1.** We can define minimization problem (2.2) in more general sense. For example, one may replace  $\|\beta\|_{\mathcal{H}(K)}^2$  by  $J(\beta)$  and then define

$$\hat{\beta}_\lambda = \arg \min_{\beta \in \mathcal{H}(K)} \{ \ell_{mn}(\beta) + \lambda J(\beta) \}$$

where the penalty functional  $J$  is a squared semi-norm on  $\mathcal{H}(K)$  such that the null space

$$\mathcal{H}_0(K) = \{g \in \mathcal{H}(K) : J(g) = 0\}$$

be a finite dimensional linear subspace of  $\mathcal{H}(K)$ .

The representer theorem gives the solution of regularization problem (2.2) in a finite dimensional subspace, although it is taken over an infinite dimensional subspace (see Wahba 1990).

**Theorem 1.** Consider minimization problem (2.2), then there exist constants  $a_{ij}$ ,  $i = 1, \dots, n$ ,  $j = 1, \dots, m$ , such that

$$\hat{\beta}_\lambda(t) = \sum_{i=1}^n \sum_{j=1}^m a_{ij} K(t, T_{ij}).$$

The proof of this Theorem is similar to that of Theorem 1.3.1 in Wahba (1990) and so we omit it.

In order to see how calculate this estimate, let  $T := [0, 1]$  and  $\mathcal{H} := \mathcal{W}_2^r$  where  $\mathcal{W}_2^r$  is the  $r$ th order Sobolev–Hilbert space:

$$\mathcal{W}_2^r = \left\{ g : [0, 1] \rightarrow \mathbb{R} \mid g, g^{(1)}, \dots, g^{(r-1)} \text{ are absolutely continuous} \right. \\ \left. \text{and } g^{(r)} \in L^2([0, 1]) \right\}.$$

Sobolev spaces have many applications in nonparametric function estimation. The smoothness of a function that belongs to some Sobolev spaces is guaranteed by existing its derivatives in some orders. To further study about Sobolev spaces see, for example, Adams (1975). There are various norms that we can equip to  $\mathcal{W}_2^r$  so that  $\mathcal{W}_2^r$  be an RKHS (see Berliet and Thomas-Agnan (2004)). If we endow  $\mathcal{W}_2^r$  with squared norm  $\|g\|_{\mathcal{W}_2^r}^2 = \sum_{k=0}^{r-1} (f g^{(k)})^2 + \int [g^{(r)}]^2$ , then it is an RKHS with reproducing kernel

$$K_r(s, t) = \frac{1}{(r!)^2} B_r(s)B_r(t) + \frac{(-1)^{r-1}}{(2r)!} B_{2r}(|s - t|),$$

where  $B_r(\cdot)$  is the  $r$ th Bernoulli polynomial. By Theorem 1, it suffices to consider  $\beta$  of the following form:

$$\beta(t) = \sum_{i=1}^n \sum_{j=1}^m a_{ij} K_r(t, T_{ij})$$

for some  $\mathbf{a} = [a_{11}, \dots, a_{1m}, a_{21}, \dots, a_{nm}]' \in \mathbb{R}^{nm}$ . Using equation (2.1) yields

$$\|\beta\|_{\mathcal{W}_2^r}^2 = \sum_{i_1=1}^n \sum_{j_1=1}^m \sum_{i_2=1}^n \sum_{j_2=1}^m a_{i_2 j_2} a_{i_1 j_1} K_r(T_{i_1 j_1}, T_{i_2 j_2}) \\ = \mathbf{a}' \mathbf{P} \mathbf{a}$$

where

$$\mathbf{P} = \begin{pmatrix} \mathbf{P}_{11} & \mathbf{P}_{12} & \mathbf{P}_{13} & \cdots & \mathbf{P}_{1n} \\ \mathbf{P}_{21} & \mathbf{P}_{22} & \mathbf{P}_{23} & \cdots & \mathbf{P}_{2n} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{P}_{n1} & \mathbf{P}_{n2} & \mathbf{P}_{n3} & \cdots & \mathbf{P}_{nn} \end{pmatrix}$$

and

$$\mathbf{P}_{i_1 i_2} = [K_r(T_{i_1 j_1}, T_{i_2 j_2})]_{1 \leq j_1, j_2 \leq m}, \quad 1 \leq i_1, i_2 \leq n$$

Define  $\mathbf{U} = [U_{11}, \dots, U_{1m}, U_{21}, \dots, U_{nm}]'$  and  $\mathbf{V} = [V_{11}, \dots, V_{1m}, V_{21}, \dots, V_{nm}]'$  then

$$(2.3) \quad \ell_{mn}(\beta) + \lambda \|\beta\|_{\mathcal{W}_2^r}^2 = \frac{1}{nm} \|\mathbf{V} - \mathbf{U} \circ (\mathbf{P} \mathbf{a})\|_{\ell_2}^2 + \lambda \mathbf{a}' \mathbf{P} \mathbf{a},$$

where  $\mathbf{A} \circ \mathbf{B}$  is the Hadamard product of two matrices  $\mathbf{A}$  and  $\mathbf{B}$ . So finding minimizer of left hand side of (2.3) over  $\mathcal{W}_2^r$  is equivalent to finding a vector

$\mathbf{a} \in \mathbb{R}^{nm}$  which minimizes right hand side of (2.3). Let  $\mathbf{Q}$  be an  $nm \times nm$  matrix such that

$$(\text{the } i\text{th column of } \mathbf{Q}) = (\text{the } i\text{th column of } \mathbf{P}) \circ (\mathbf{U} \circ \mathbf{U}), \quad i = 1, 2, \dots, nm$$

It can be seen that the minimizer of (2.3) is

$$\mathbf{a} = (\mathbf{Q} + nm\lambda\mathbf{I})^{-1} (\mathbf{U} \circ \mathbf{V})$$

---

### 3. SIMULATION STUDY

---

In our simulation study, we carried out a set of simulation studies to emphasize the practical implementation of our methodology. Let true slope function be

$$\beta(t) = \sum_{k=1}^{50} \zeta_k \phi_k(t), \quad t \in [0, 1],$$

where  $\zeta_1 = 0.3$ ,  $\phi_1(t) = 1$ ,  $\zeta_k = 4(-1)^{k+1}k^{-2}$  and  $\phi_k(t) = \cos((k-1)\pi t)$  for  $k \geq 1$ . It is clear that this function belongs to the second order Sobolev space ( $r = 2$ ). Random functions  $X_i$ 's are generated independently as follows:

$$X(t) = \sqrt{2} \sin(\pi t) \xi_1 + \sqrt{2} \cos(\pi t) \xi_2, \quad t \in [0, 1],$$

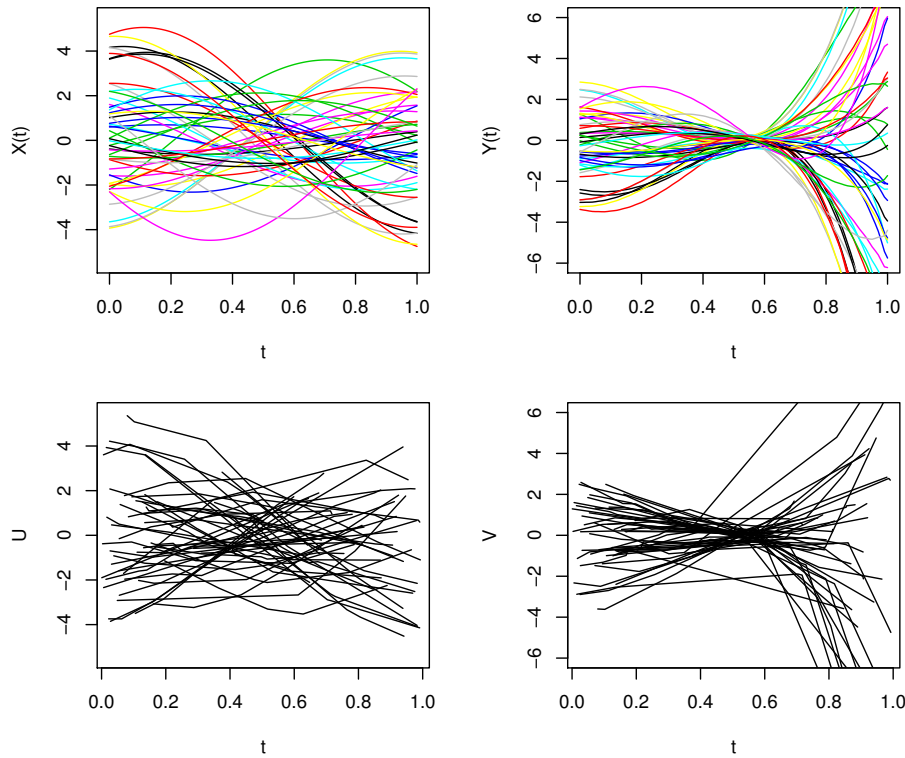
where  $\xi_1$  and  $\xi_2$  are independent random variables with  $\xi_i \sim N(0, i)$ ,  $i = 1, 2$ . The response trajectories are generated according to model (1.2) with

$$Z(t) = \sqrt{2} \sin(\pi t) Z_1 + \sqrt{2} \cos(\pi t) Z_2, \quad t \in [0, 1],$$

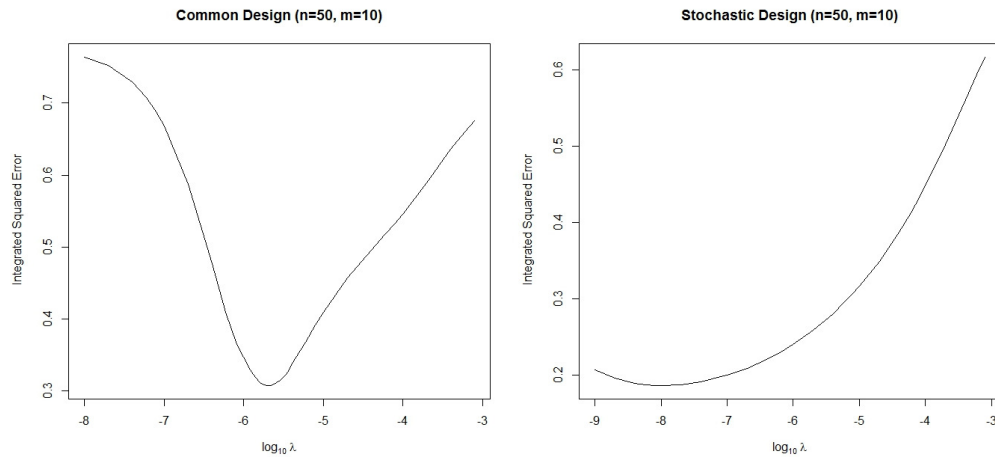
where  $Z_1$  and  $Z_2$  are i.i.d. random variables from  $N(0, 0.1)$ . Design points are selected based on common or random design. Noisy observations of each curve obtain according to model (1.3) in each curve.

The fifty curves from  $X(t)$  and  $Y(t)$  were given in the top panels of Figure 1, the left panel for  $X(t)$  and the right panel for  $Y(t)$ . The lower panels of Figure 1 shows the observed data for  $m = 5$  random design points based on stochastic design, the left panel for  $U$  and the right panel for  $V$ .

We use integrated squared error,  $\|\hat{\beta}_\lambda - \beta\|_{L_2}^2 = \int_0^1 (\hat{\beta}_\lambda(t) - \beta(t))^2 dt$ , to assess goodness of fit of the model. The integrated squared error,  $\|\hat{\beta}_\lambda - \beta\|_{L_2}^2$ , as a function of smoothing parameter  $\lambda$  is shown in Figure 2 for both designs, the right panel for stochastic design, and the left panel for common design. The best choice for smoothing parameter is the value of  $\lambda$  that minimizes  $\|\hat{\beta}_\lambda - \beta\|_{L_2}^2$ .



**Figure 1:** The top panels give 50 simulated curves, the left panel for  $X$  and the right panel for  $Y$ . Noisy observations at 5 random location based on stochastic design are shown in the lower panels, the left panel for  $U$  and the right panel for  $V$ .



**Figure 2:** Sensitivity of integrated squared error,  $\|\hat{\beta}_\lambda - \beta\|_{L_2}^2$ , with respect to smoothing parameter  $\lambda$  for both designs. The right panel for stochastic design, and the left panel for common design.

We calculated  $\|\hat{\beta}_\lambda - \beta\|_{L_2}^2$  for different combinations of  $n \in \{25, 50, 100, 200\}$  and  $m \in \{3, 5, 10, 20\}$ . Table 1 presents the obtained value of the smoothing



parameter for each simulated data set. As we see in Table 1, the smoothing parameter for common design is much greater than the smoothing parameter for stochastic design. This is because, in the stochastic design we observe random functions  $X$  and  $Y$  in many different points over whole of  $n$  samples, while in the common design, we observe random functions  $X$  and  $Y$  in only  $m$  equidistant points. In addition, the results of a Monte Carlo approximations of  $\|\hat{\beta}_\lambda - \beta\|_{L_2}^2$  for common and stochastic design are reported in Tables 2 and 3 respectively. It can be seen from both Tables that the averaged integrated squared error and variance of estimated slope function decrease as either  $m$  or  $n$  increases. On the other hand, the values of  $\|\hat{\beta}_\lambda - \beta\|_{L_2}^2$  for stochastic design is smaller than that is for common design. These results imply that stochastic design has better performance than common design.

**Table 1:** The value of smoothing parameter for common and stochastic design.

Type of design	$n$	$m$				
		3	5	10	15	20
common design	25	$8 \times 10^{-5}$	$2.5 \times 10^{-5}$	$3.5 \times 10^{-6}$	$5 \times 10^{-7}$	$10^{-7}$
stochastic design		$5 \times 10^{-7}$	$10^{-7}$	$5 \times 10^{-8}$	$2 \times 10^{-8}$	$10^{-8}$
common design	50	$7.5 \times 10^{-5}$	$2 \times 10^{-5}$	$3 \times 10^{-6}$	$5 \times 10^{-7}$	$9 \times 10^{-8}$
stochastic design		$10^{-7}$	$5 \times 10^{-8}$	$1.5 \times 10^{-8}$	$4 \times 10^{-9}$	$2 \times 10^{-9}$
common design	100	$7 \times 10^{-5}$	$1.5 \times 10^{-5}$	$2.5 \times 10^{-6}$	$4.5 \times 10^{-7}$	$8.5 \times 10^{-8}$
stochastic design		$2.5 \times 10^{-8}$	$7.5 \times 10^{-9}$	$2.5 \times 10^{-9}$	$10^{-9}$	$5.5 \times 10^{-10}$
common design	200	$6.5 \times 10^{-5}$	$10^{-5}$	$2 \times 10^{-6}$	$4 \times 10^{-7}$	$8 \times 10^{-8}$
stochastic design		$7.5 \times 10^{-9}$	$2.5 \times 10^{-9}$	$6.5 \times 10^{-10}$	$4.5 \times 10^{-10}$	$3.5 \times 10^{-10}$

**Table 2:** Averaged integrated squared error  $\|\hat{\beta}_\lambda - \beta\|_{L_2}^2$  and variance of  $\hat{\beta}_\lambda$  (in the parentheses) for common design.

$n$	$m$				
	3	5	10	15	20
25	0.7683 (0.0091)	0.4967 (0.0085)	0.3167 (0.0074)	0.2498 (0.0064)	0.2136 (0.0051)
50	0.7628 (0.0050)	0.4882 (0.0041)	0.3112 (0.0033)	0.2458 (0.0031)	0.2108 (0.0027)
100	0.7596 (0.0024)	0.4854 (0.0019)	0.3085 (0.0015)	0.2436 (0.0013)	0.2098 (0.0011)
200	0.7538 (0.0012)	0.4827 (0.0009)	0.3029 (0.0007)	0.2400 (0.0006)	0.2054 (0.0004)

**Table 3:** Averaged integrated squared error  $\|\hat{\beta}_\lambda - \beta\|_{L_2}^2$  and variance of  $\hat{\beta}_\lambda$  (in the parentheses) for stochastic design.

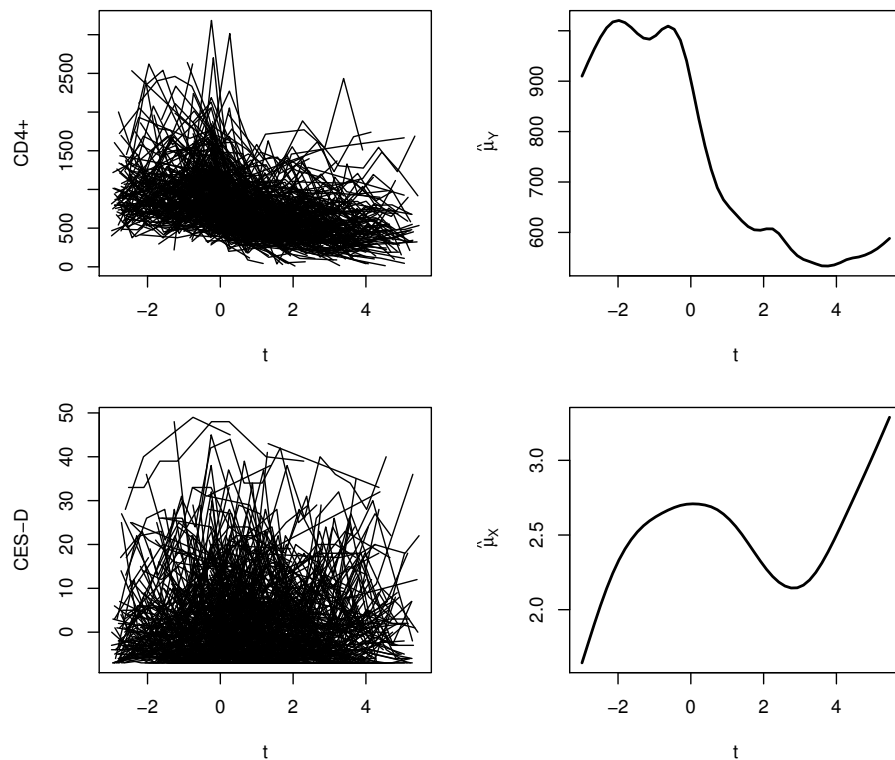
$n$	$m$				
	3	5	10	15	20
25	0.3435 (0.1335)	0.2866 (0.1111)	0.2324 (0.0673)	0.1944 (0.0406)	0.1902 (0.0354)
50	0.2660 (0.0902)	0.2316 (0.0667)	0.1858 (0.0413)	0.1794 (0.0313)	0.1676 (0.0305)
100	0.2111 (0.0569)	0.1812 (0.0392)	0.1609 (0.0258)	0.1528 (0.0214)	0.1487 (0.0172)
200	0.1776 (0.0366)	0.1577 (0.0245)	0.1437 (0.0165)	0.1365 (0.0111)	0.1312 (0.0084)

---

#### 4. APPLICATION

---

The human immune deficiency virus (HIV) attacks immune cells called CD4+ and leads to AIDS. CD4+ cells are a specific kind of white blood cell and are a necessary part of the immune system. They lead the attack against infections. The CD4+ cell count measures the number of CD4+ cells in a sample of blood. CD4+ cell counts are reported as the number of cells in a cubic millimetre of blood. A normal CD4+ cell count is around 1100 cells per cubic millimetre of blood. The CD4+ cell counts can vary time to time. When someone is infected with HIV the number of CD4+ cells they have goes down. So an infected person's CD4+ cell number can be used to monitor disease progression.



**Figure 3:** The top panel provides observed individual trajectories and the smooth estimate of the mean function for CD4+ cell counts. The bottom panel includes observed individual trajectories and the smooth estimate of the mean function for CES-D scores.

The CES-D scale is a short self-report scale designed to measure depressive symptomatology during the past week. A higher score indicates greater depressive symptoms. It is interesting to explore whether there is an association between

depressive symptoms and CD4+ cell counts over time. The data, reported by Kaslow *et al.* (1987), recorded CD4+ cell counts, CES-D scores and other variables over time for a total of 369 infected men enrolled in the Multicenter AIDS Cohort Study. The measurements were scheduled at each half-yearly visit. But because of missing appointments among other factors, the actual measurement times are random, irregular and sparse. For both CD4+ and CES-D the number of observations ranged from 1 to 12, with a median of 6 measurements per subject, yielding a total of 2376 records.

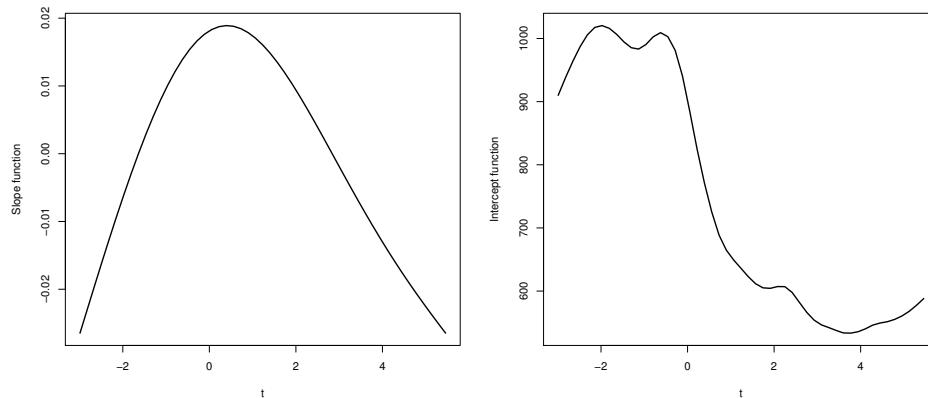
In this dataset, both the CD4+ cell counts and CES-D scores are considered as functions of time since seroconversion (time when HIV becomes detectable). We model the response process CD4+ cell counts and the predictor process CES-D scores via functional varying coefficient model (1.1). Individual trajectories of CD4+ cell counts and CES-D scores are shown in Figure 3, along with the smooth estimated mean functions of CD4+ cell counts and CES-D scores. The estimated mean function of CD4+ cell counts shows a drastic decreasing from seroconversion to around 2 years after seroconversion. Also the estimated mean function of CES-D scores is decreasing in this period.

We used 5-fold cross-validation to choose the smoothing parameter  $\lambda$ . The procedure is as follows. Divide the data into 5 roughly equal parts at random. For each part, fit the model with parameter  $\lambda$  to the remaining 4 parts. Let  $\hat{\beta}_\lambda^{(-k)}$  be the estimated slope function by dropping the  $k$ th part,  $k = 1, 2, \dots, 5$ . The cross-validation error is given by

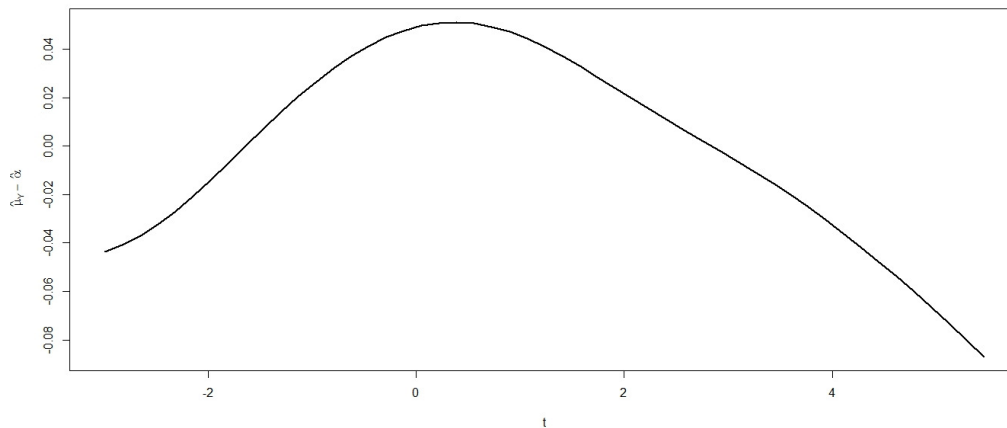
$$CV(\lambda) = \frac{1}{5} \sum_{k=1}^5 \sum_{i \in k\text{th part}} \frac{1}{m_i} \sum_{j=1}^{m_i} \left( V_{ij} - \hat{\mu}_Y(T_{ij}) - [U_{ij} - \hat{\mu}_X(T_{ij})] \hat{\beta}_\lambda^{(-k)}(T_{ij}) \right)^2,$$

where  $m_i$  is number of measurements for  $i$ th subject and,  $\hat{\mu}_X(t)$  and  $\hat{\mu}_Y(t)$  are the smooth estimates of mean function for CD4+ cell counts and CES-D scores respectively. To estimate the mean function under sparse and irregular designs, we refer the readers to Yao *et al.* (2005a), Li and Hsing (2010), and Cai and Yuan (2011). Now calculate  $CV(\lambda)$  for different values of  $\lambda$  and choose the optimal value of  $\lambda$  as the minimal of  $CV(\lambda)$ . Here we obtained  $\lambda = 100$ .

The estimated slope function  $\hat{\beta}$  and intercept function  $\hat{\alpha}$  are displayed in Figure 4, where we used  $\hat{\alpha}(t) = \hat{\mu}_Y(t) - \hat{\mu}_X(t)\hat{\beta}(t)$  to estimate the intercept function. Since the value of CES-D and the estimated slope function are small with respect to the value CD4+ cell counts, the shapes of  $\hat{\mu}_Y(t)$  and  $\hat{\alpha}(t)$  are obtained almost similar. In Figure 5, we provided difference  $\hat{\mu}_Y(t) - \hat{\alpha}(t)$ . By comparing Figures 4 and 5 we see that there is a minor association between CD4+ cell counts and CES-D scores in earlier and later times. In addition the association in other times is negligible.



**Figure 4:** The left panel shows the estimated slope function and the right panel displays the estimated intercept function.



**Figure 5:** The difference  $\hat{\mu}_Y(t) - \hat{\alpha}(t)$ .

---

## 5. CONCLUSIONS AND EXTENSIONS

---

We have presented a regularization method to estimate the slope function in functional varying coefficient model using an RKHS approach. Our procedure is easy to implement in the numerical scheme and do not need resorting some numerical techniques to compute the slope function. As we saw in the simulation study, increasing either  $m$  or  $n$  leads to improved estimates, in the sense of integrated squared error and variance. In this paper, we have assumed that all sampling points on each curve are same. We note that this assumption is not necessary and we may have different sampling points on each curves. Let  $m_i$  be

the sampling frequency on  $i$ th curve. It is suffice to define

$$\hat{\beta}_\lambda = \arg \min_{\beta \in \mathcal{H}} \{ \ell_{mn}(\beta) + \lambda \|\beta\|_{\mathcal{H}}^2 \}$$

where

$$\ell_{mn}(\beta) = \frac{1}{n} \sum_{i=1}^n \frac{1}{m_i} \sum_{j=1}^{m_i} (V_{ij} - U_{ij}\beta(T_{ij}))^2$$

and then use the given procedure with some mild modifications.

Obtaining rates of convergence and studying optimality of the estimators, in some sense, are interesting problems in nonparametric function estimation. Şentürk and Müller (2010) have given rate of convergence for functional varying coefficient model with sparse and noise-contaminated data in the supremum of absolute error sense but they have not studied optimality of their estimators. Another interesting problem is estimating derivatives of  $\beta(t)$  in this model. These ideas will be explored in future works.

---

## ACKNOWLEDGMENTS

---

The authors are grateful to the referee for valuable comments and suggestions that improved the paper.

---

## REFERENCES

---

- [1] ADAMS, R.A. (1975). *Sobolev Spaces*, Academic Press, New York.
- [2] ARONSAJN, N. (1950). Theory of reproducing kernel, *Transactions of the American Mathematical Society*, **68**, 337–404.
- [3] BERLINET, A. and THOMAS-AGNAN, C. (2004). *Reproducing Kernel Hilbert Spaces in Probability and Statistics*, Kluwer-Academic Publishers, Dordrecht.
- [4] CAI, T. and YUAN, M. (2011). Optimal estimation of the mean function based on discretely sampled functional data: Phase transition, *The Annals of Statistics*, **39**, 2330–2355.
- [5] CHIANG, C.; RICE J.A. and WU, C.O. (2001). Smoothing spline estimation for varying coefficient models with repeatedly measured dependent variables, *Journal of the American Statistical Association*, **96**, 605–617.
- [6] CHIOU, J.M.; MA, Y. and TSAI, C.L. (2012). Functional random effect time-varying coefficient model for longitudinal data, *Stat.*, **1**, 75–89.

- [7] FAN, J. and ZHANG, W. (1999). Statistical estimation in varying coefficient models, *The Annals of Statistics*, **27**, 1491–1518.
- [8] FAN, J. and ZHANG, W. (2008). Statistical methods with varying coefficient models, *Statistics and Its Interface*, **1**, 179–195.
- [9] FERRATY, F. and VIEU, P. (2006). *Nonparametric Functional Data Analysis: Theory and Practice*, Springer, New York.
- [10] HOOVER, D.R.; RICE, J.A.; WU, C.O. and YANG, L.-P. (1998). Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data, *Biometrika*, **85**, 809–822.
- [11] HUANG, J.Z.; WU, C.O. and ZHOU, L. (2002). Varying-coefficient models and basis function approximations for the analysis of repeated measurements, *Biometrika*, **89**, 111–128.
- [12] HUANG, J.Z.; WU, C.O. and ZHOU, L. (2004). Polynomial spline estimation and inference for varying coefficient models with longitudinal data, *Statistica Sinica*, **14**, 763–788.
- [13] KASLOW, R.; OSTROW, D.; DETELS, R.; PHAIR, J.; POLK, B. and RINALDO, C. (1987). The multicenter AIDS cohort study: Rationale, organization and selected characteristics of the participants, *American Journal of Epidemiology*, **129**, 310–318.
- [14] KAUERMANN, G. and TUTZ, G. (1999). On model diagnostics using varying coefficient models, *Biometrika*, **86**, 119–128.
- [15] LI, Y. and HSING, T. (2010). Uniform convergence rates for nonparametric regression and principal component analysis in functional/longitudinal data, *The Annals of Statistics*, **38**, 3321–3351.
- [16] LIN, D.Y. and YING, Z. (2001). Semiparametric and nonparametric regression analysis for longitudinal data, *Journal American Statistical Association*, **96**, 103–126.
- [17] NOH, H.S. and PARK, B.U. (2010). Sparse varying coefficient models for longitudinal data, *Statistica Sinica*, **20**, 1183–1202.
- [18] RAMSAY, J.O. and SILVERMAN, B.W. (2002). *Applied Functional Data Analysis: Methods and Case Studies*, Springer, New York.
- [19] RAMSAY, J.O. and SILVERMAN, B.W. (2005). *Functional Data Analysis, 2nd Edition*, Springer, New York.
- [20] SENTÜRK, D. and MÜLLER, H.G. (2008). Generalized varying coefficient models for longitudinal data, *Biometrika*, **95**, 653–666.
- [21] SENTÜRK, D. and MÜLLER, H.G. (2010). Functional varying coefficient models for longitudinal data, *Journal of the American Statistical Association*, **105**, 1256–1264.
- [22] SENTÜRK, D. and NGUYEN, D.V. (2011). Varying coefficient models for sparse noise-contaminated longitudinal data, *Statistica Sinica*, **21**, 1831–1856.
- [23] SENTÜRK, D.; DALRYMPLE, L.S.; MOHAMMED, S.M.; KAYSEN, G.A. and NGUYEN, D.V. (2013). Modeling time-varying effects with generalized and unsynchronized longitudinal data, *Statistics in Medicine*, **32**, 2971–2987.
- [24] WAHBA, G. (1990). *Spline models for observational data*, SIAM, Philadelphia.

- [25] WANG, H. and XIA, C. (2009). Shrinkage estimation of the varying coefficient model, *Journal of the American Statistical Association*, **104**, 747–757.
- [26] WU, C.O.; CHIANG, C.T. and HOOVER, D.R. (1998). Asymptotic confidence regions for kernel smoothing of a varying coefficient models with longitudinal data, *Journal of the American Statistical Association*, **93**, 1388–1402.
- [27] WU, C.O. and CHIANG, C.T. (2000). Kernel smoothing on varying coefficient models with longitudinal dependent variable, *Statistica Sinica*, **10**, 433–456.
- [28] YANG, W.; MÜLLER, H.G. and STADTMÜLLER, U. (2011). Functional singular component analysis, *Journal of the Royal Statistical Society Series B Statistical Methodology*, **73**, 303–324.
- [29] YAO, F.; MÜLLER, H.G. and WANG, J. (2005a). Functional data analysis for sparse longitudinal data, *Journal of the American Statistical Association*, **100**, 577–590.
- [30] YAO, F.; MÜLLER, H.G. and WANG, J.L. (2005b). Functional linear regression analysis for longitudinal data, *The Annals of Statistics*, **33**, 2873–2903.