FUNCTIONAL VARYING COEFFICIENT MODEL WITH TIME-INDEPENDENT COVARIATE AND LONGITUDINAL RESPONSE

Behdad Mostafaiy and Mohammad Reza Faridrohani
Department of Statistics, Shahid Beheshti University

Abstract: In this paper, we consider functional varying coefficient model in present of a time invariant covariate for sparse longitudinal data contaminated with some measurement errors. We propose a regularization method to estimate the slope function based on a reproducing kernel Hilbert space approach. As we will see, our procedure is easy to implement. Our simulation results show that the procedure performs well, especially when either sampling frequency or sample size increases. Applications of our method are illustrated in an analysis of a longitudinal CD4+ count dataset from an HIV study.

Key words: CD4+ count, functional varying coefficient model, longitudinal data analysis, reproducing kernel Hilbert space, sparsity

1. Introduction

In functional data, unlike multivariate data, the observations are naturally curves. In fact they are independent and identically distributed realizations of a stochastic process. See Ramsay and Silverman (2002, 2005) for an overview of methods and applications. See also Ferraty and Vieu (2006) and, Horváth and Kokoszka (2012). In many experiments realizations of involved trajectories are not directly observable. Instead, the observed data are obtained at discrete location points. These type of data are usually sparsely and irregularly sampled on random time points and are noise-contaminated. The aforementioned situation often occurs in many longitudinal experiments, for example in the most of biological, biomedical and medical studies.

Varying coefficient models which are extension of parametric regression models, became popular after the works of Cleveland et al. (1991) and, Hastie and Tibshirani (1993). These models have been extensionally studied in the literature. Most of existing approach are based on polynomial spline, smoothing splines and local polynomial smoothing. See for example Hoover et al. (1998), Wu et al. (1998), Kauermann and Tutz (1999), Chiang et al. (2001), Wu and Chiang (2000) and Huang et al. (2002, 2004). See Fan and Zhang (2008) to review the collection of approaches and applications up-to that date. In the case of sparse designs and noisy measurement, we refer the readers to Sentürk and Müller (2008), Noh and Park (2010), Sentürk and Müller (2010), Sentürk and Nguyen (2011), Chiou et al. (2012), and Sentürk et al. (2013).

In this paper, we consider the following functional varying coefficient model

$$Y(t) = \alpha_0(t) + \beta_0(t)X + Z(t), \qquad t \in \mathcal{T} \tag{1}$$

where $\mathcal{T} \subset \mathbb{R}$ is a compact set, $Y(t)$ is a square integrable response with mean function $\mu_Y(t)$ and X is a random variable with finite variance σ_X^2. Also, $\alpha_0(t)$ and $\beta_0(t)$ are smoothed intercept and slope functions, respectively, and $Z(t)$ is a mean zero random processes, independent of X. We want to estimate $\alpha_0(t)$ and $\beta_0(t)$ in the situation that the observations are sparse and irregular longitudinal data and contaminated with some measurement errors. Let V_{ij} denote the observations of the random function Y_i at the random times T_{ij}, contaminated with measurement errors ε_{ij} which are assumed to be independent and identically distributed with means zero and finite variance, and independent of random function Y. Then the model that we consider is

$$V_{ij} = Y_i(T_{ij}) + \varepsilon_{ij}, \qquad j = 1, \dots, m; \qquad i = 1, \dots, n. \qquad (2)$$

The requirement that the number of time points in each curve are equal is not essential and is placed for easy illustration of the methodology.

In the present paper, we use a reproducing kernel Hilbert space (RKHS) framework for estimating the slope function. We assume that the slope function β_0 resides in an RKHS, a subspace of the collection of square integrable functions on $\mathcal{T}$. Then we investigate the method of regularization for estimating β_0. By obtaining an estimate of β_0, say $\hat{\beta}_0$, we estimate the intercept function α_0 via

$$\hat{\alpha}_0(t) = \hat{\mu}_Y(t) - \bar{X}\hat{\beta}_0(t), \qquad t \in \mathcal{T},$$

where $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$ and $\hat{\mu}_Y(t)$ is a smoothed estimate of $\mu_Y(t)$. Under the random design setup discussed above, the mean function estimate $\hat{\mu}_Y(t)$ can be obtained by, for example, any of the methods discussed in Yao et al. (2005), Li and Hsing (2010), and Cai and Yuan (2011).

The paper is organized as follows. In section 2, we introduce our methodology for estimating the slope function β_0. Section 3 presents results of some simulation studies. An application to a longitudinal CD4+ count dataset are demonstrated in section 4. Finally, concluding remarks and possible extensions of our work are presented in Section 5.

2. Estimation Method

In this section, we introduce a regularization method for estimating the slope function. First, we present the general definitions and common properties of RKHS required in this paper. Verification of these results as well as more detailed discussions of RKHS theory can be found, for example, in Aronszajn (1950).

Let $\mathcal{H}$ be a Hilbert space of functions on some set $\mathcal{T}$ and denote by $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ the inner product in $\mathcal{H}$. A bivariate function K on $\mathcal{T} \times \mathcal{T}$ is said to be a reproducing kernel for $\mathcal{H}$ if

(i) $K(\cdot, t) \in \mathcal{H}$ for every $t \in \mathcal{T}$, and

(ii) $f(t) = \langle f, K(\cdot, t) \rangle_{\mathcal{H}}$ for every $t \in \mathcal{T}$ and $f \in \mathcal{H}$.

When (i) and (ii) hold, $\mathcal{H}$ is said to be a reproducing kernel Hilbert space (RKHS) with reproducing kernel K. Property (ii) is called the reproducing property. If K is a reproducing

kernel of $\mathcal{H}$ then, it can be shown that, K is the unique reproducing kernel and K is a symmetric and positive definite function: that is, $K(s, t) = K(t, s)$ for $s, t \in \mathcal{T}$ and for any real $b_1, \ldots, b_N$ and $s_1, \ldots, s_N \in \mathcal{T}$,

$$\sum_{i=1}^{N} \sum_{j=1}^{N} b_i b_j K(s_i, s_j) \geq 0.$$

Conversely, if K is a symmetric and positive definite function on $\mathcal{T} \times \mathcal{T}$, a unique RHKS of functions on $\mathcal{T}$ with K as the reproducing kernel can be constructed. The notation $\mathcal{H}(K)$ will be used to denote the RKHS having the reproducing kernel K.

Assume that the reproducing kernel K is square integrable, that is,

$$\int_{\mathcal{T} \times \mathcal{T}} K^2(s, t) ds dt < \infty.$$

Then Mercer's theorem (see Riesz and Sz.-Nagy 1955) states that K admits the following eigenvalue decomposition:

$$K(s, t) = \sum_{k \geq 1} \rho_k \phi_k(s) \phi_k(t) \tag{3}$$

where $\rho_1, \rho_2, \ldots$ are constants and $\phi_1, \phi_2, \ldots$ are orthonormal basis for $L^2(\mathcal{T})$, that is

$$\langle \phi_j, \phi_k \rangle_{L^2(\mathcal{T})} = \int_{\mathcal{T}} \phi_j(t) \phi_k(t) dt = \delta_{jk},$$

where δ is Kronecker's delta. From Lemma 1.1.1 in Wahba (1990), we know that any squared integrable function f on $\mathcal{T}$ belongs to $\mathcal{H}(K)$ if and only if

$$\parallel f \parallel_{\mathcal{H}(K)}^2 = \sum_{k \geq 1} \rho_k^{-1} f_k^2 < \infty, \tag{4}$$

where $k = 1, 2, \ldots$ for $f_k = \langle f, \phi_k \rangle_{L^2(\mathcal{T})}$.

Now, consider functional varying coefficient model (1). We investigate the method of regularization for estimating β_0. We assume that the slope function β_0 resides in an RKHS $\mathcal{H}(K)$, a subspace of the collection of square integrable functions on $\mathcal{T}$. The method of regularization estimates β_0 by

$$\hat{\beta}_\lambda = \underset{\beta \in \mathcal{H}(K)}{argmin} \{ \ell_{nm}(\beta) + \lambda \parallel \beta \parallel_{\mathcal{H}(K)}^2 \} \tag{5}$$

where

$$\ell_{nm}(\beta) = \frac{1}{nm} \sum_{i=1}^{n} \sum_{j=1}^{m} \left(V_{ij} - X_i \beta(T_{ij}) \right)^2 \tag{6}$$

measures how well β fits the data and $\parallel \beta \parallel_{\mathcal{H}(K)}^2$ is an RKHS norm that measures smoothness of β, and $\lambda \geq 0$ is a tuning parameter that balances the fidelity to the data and the smoothness.

Remark 1 In general, ℓ_{nm} is chosen such that it is convex in β and $E[\ell_{nm}]$ is uniquely minimized by β_0.

Let $\mathbf{K}$ be an $nm \times nm$ matrix defined by

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

where

$$
\mathbf{K}_{i_1 i_2} = \left[K(T_{i_1 j_1}, T_{i_2 j_2}) \right]_{1 \leq j_1 \leq m, 1 \leq j_2 \leq m}, \qquad 1 \leq i_1, i_2 \leq n.
$$

Define nm dimensional vectors $\mathbf{X}$ and $\mathbf{V}$ respectively as $\mathbf{X} = [X_1 \mathbf{1}'_m, \ldots, X_n \mathbf{1}'_m]'$ where $\mathbf{1}_m$ is the m dimensional vector with all elements equal to 1 and $\mathbf{V} = [V_{11}, \ldots, V_{1m}, V_{21}, \ldots, V_{nm}]'$. Suppose $\mathbf{Q}$ be an $nm \times nm$ matrix such that

(the ith column of $\mathbf{Q}$) = (the ith column of $\mathbf{K}$) $\circ$ $(\mathbf{X} \circ \mathbf{X})$), $\qquad i = 1, 2, \ldots, nm$

where $\mathbf{A} \circ \mathbf{B}$ is the Hadamard product of two matrices $\mathbf{A}$ and $\mathbf{B}$.

In the following theorem, we give solution of the minimization problem (5).

Theorem 1 The function $\hat{\beta}_\lambda$ minimizing the regularized empirical error

$$
REE(\beta) = \frac{1}{nm} \sum_{i=1}^{n} \sum_{j=1}^{m} \left(V_{ij} - X_i \beta(T_{ij}) \right)^2 + \lambda \parallel \beta \parallel^2_{\mathcal{H}(K)}
$$

over $\beta \in \mathcal{H}(K)$, may be expressed as

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

where $\mathbf{a} = [a_{11}, \ldots, a_{1m}, a_{21}, \ldots, a_{nm}]'$ is the unique solution of the well-posed linear system in $\mathbb{R}^{nm}$

$$
(\mathbf{Q} + nm\lambda \mathbf{I})\mathbf{a} = \mathbf{X} \circ \mathbf{V},
$$

with $\mathbf{I}$ is the $nm \times nm$ identity matrix.

Proof. For any $\beta \in \mathcal{H}(K)$, write $\beta(t) = \sum_{k \geq 1} c_k \phi_k(t)$ where $c_k = \langle \beta, \phi_k \rangle_{L^2(\mathcal{T})}$. From (4), we have $\parallel \beta \parallel^2_{\mathcal{H}(K)} = \sum_{k \geq 1} \rho_k^{-1} c_k^2$.

For any $k \geq 1$, we have

$$
\frac{\partial REE}{\partial c_k} = \frac{-2}{nm} \sum_{i=1}^{n} \sum_{j=1}^{m} X_i [V_{ij} - X_i \beta(T_{ij})] \phi_k(T_{ij}) + 2\lambda \frac{c_k}{\rho_k}.
$$

If β is a minimizer of REE, then for each k, we must have $\frac{\partial REE}{\partial c_k} = 0$ or equivalently

$$c_k = \rho_k \sum_{i=1}^{n} \sum_{j=1}^{m} a_{ij}\phi_k(T_{ij})$$

where $a_{ij} = \frac{X_i[V_{ij}-X_i\beta(T_{ij})]}{nm\lambda}$. Thus,

$$\hat{\beta}(t) = \sum_{k=1}^{\infty} c_k\phi_k(t) = \sum_{k=1}^{\infty} \rho_k \sum_{i=1}^{n} \sum_{j=1}^{m} a_{ij}\phi_k(T_{ij})\phi_k(t)$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{m} a_{ij} \sum_{k=1}^{\infty} \rho_k\phi_k(T_{ij})\phi_k(t) = \sum_{i=1}^{n} \sum_{j=1}^{m} a_{ij}K(t,T_{ij}),$$

where the last equality follows from (3). Replacing $\beta(t)$ in the definition of a_{ij} above to obtain

$$a_{ij} = \frac{X_i[V_{ij} - X_i \sum_{\ell=1}^{n} \sum_{k=1}^{m} a_{\ell k}K(T_{ij}, T_{\ell k})]}{nm\lambda}.$$

By multiplying both sides in $nm\lambda$ and writing the results in matrix form, we obtain $(\mathbf{Q} + nm\lambda\mathbf{I})\mathbf{a} = \mathbf{X} \circ \mathbf{V}$. This system is well-posed since $\mathbf{Q}$ is positive definite and the addition of a positive definite matrix and identity matrix is strictly positive definite.

Remark 2 *We can define minimization problem (0.5) in more general sense. For example, one may replace the RKHS norm $\| \beta \|_{\mathcal{H}(K)}^2$ by penalty functional $J(\beta)$ and then define*

$$\hat{\beta}_\lambda = \underset{\beta\in H(K)}{argmin}\{\ell_{mn}(\beta) + \lambda J(\beta)\}$$

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 $\mathcal{H}(K)$. In this case, let $\{\xi_1, \dots, \xi_N\}$ be an orthonormal basis for $\mathcal{H}_0(K)$, where $N = \dim(\mathcal{H}_0(K))$. It can be shown that , there exist $\mathbf{d} = [d_1, \dots, d_N]' \in \mathbb{R}^N$ and $\mathbf{a} = [a_{11}, \dots, a_{1m}, a_{21}, \dots, a_{nm}]' \in \mathbb{R}^{nm}$ such that

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

3. Simulation Studies

To demonstrate the performance of the proposed estimator, we carried out a set of simulation studies with different combinations of sampling frequency and sample size. Let $\mathcal{T} = [0,1]$. The true slope function β_0 is fixed as

$$\beta_0(t) = cos(\pi t) + sin(\pi t), \qquad t \in [0,1].$$

It is clear that, $\beta_0 \in \mathcal{W}_2^2$ where $\mathcal{W}_2^r$ is the rth order Sobolev-Hilbert space:

$$\mathcal{W}_2^r = \{g:[0,1] \to \mathbb{R}|g, g^{(1)}, \dots, g^{(r-1)} \quad are \quad absolutely \quad continuous \quad and \quad g^{(r)} \in L^2([0,1])\}$$

Under the squared norm $\| g \|_{\mathcal{W}_2^r}^2 = \sum_{k=0}^{r-1} \left(\int_0^1 g^{(k)}(t) \, dt\right)^2 + \int_0^1 [g^{(r)}(t)]^2 \, dt$, the Hilbert space $\mathcal{W}_2^r$ 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 rth Bernoulli polynomial. Abramowitz and Stegun (1965) give the following formula for Bernoulli polynomial

$$B_{2q}(t) = (-1)^{q-1} 2(2q)! \sum_{v=1}^{\infty} \frac{cos(2\pi v t)}{(2\pi v)^{2q}}.$$

The covariate X is generated from $N(0,2)$. The response trajectories were generated according to model (0.1) with

$$Z(t) = sin(\pi t)Z_1 + cos(\pi t)Z_2, \qquad t \in [0,1]$$

where Z_1 and Z_2 are $i.i.d$ random variables from $N(0,0.1)$. From each generated function, m random locations were uniformly sampled from $[0,1]$ and so the noisy observations are obtained following model (2) with $\varepsilon \sim N(0,0.1)$.
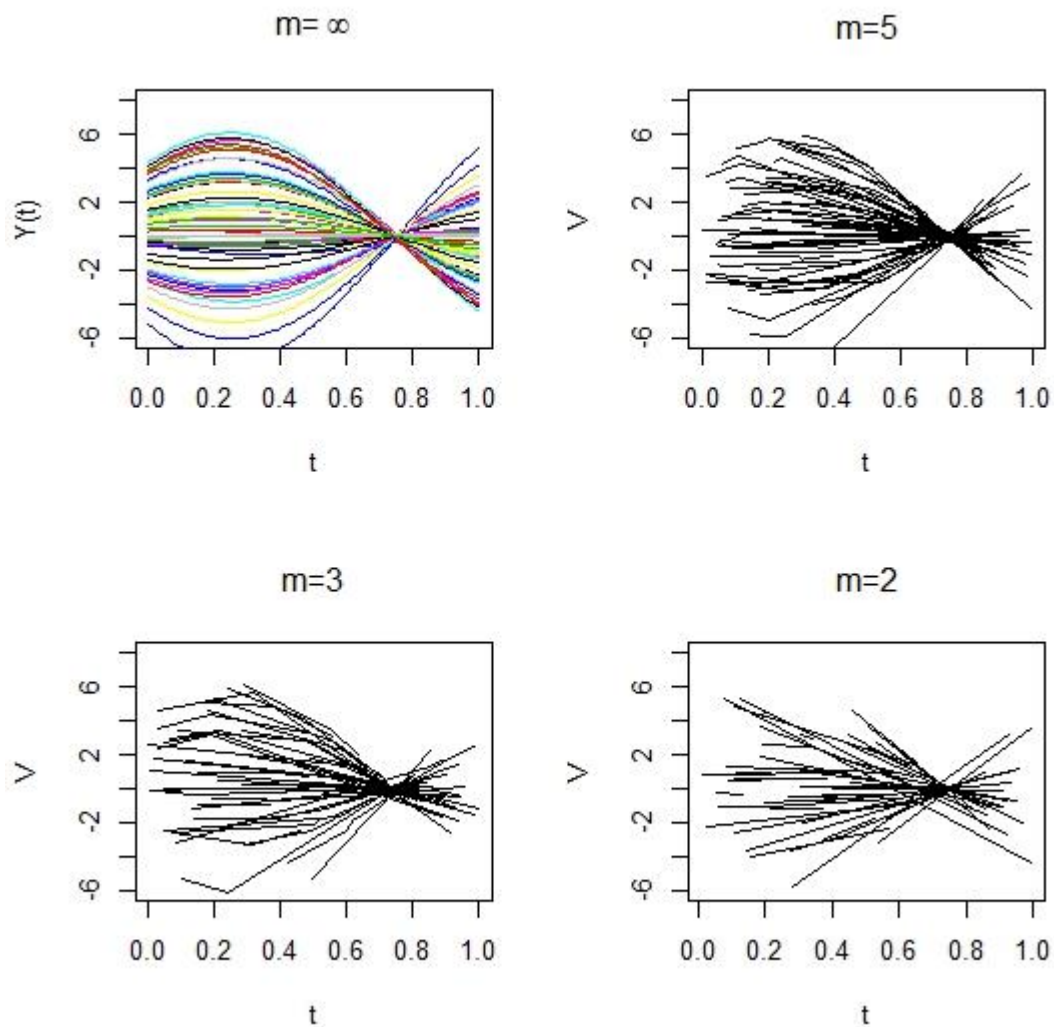
Figure 1: A typical simulated dataset: The top left panel shows $n = 50$ simulated functions for $Y(t)$. Noisy observations at $m = 5, 3$ and 2 random locations on each curve are given in the top right, the bottom left and the bottom right panels, respectively.

Figure 1 provides a visual inspection of a typically simulated data. The fifty curves were given in the top left panel of Figure 1. Other panels of Figure 1 show the observed data for $m = 5, 3$ and 2, where observations from the same curve are connected together.
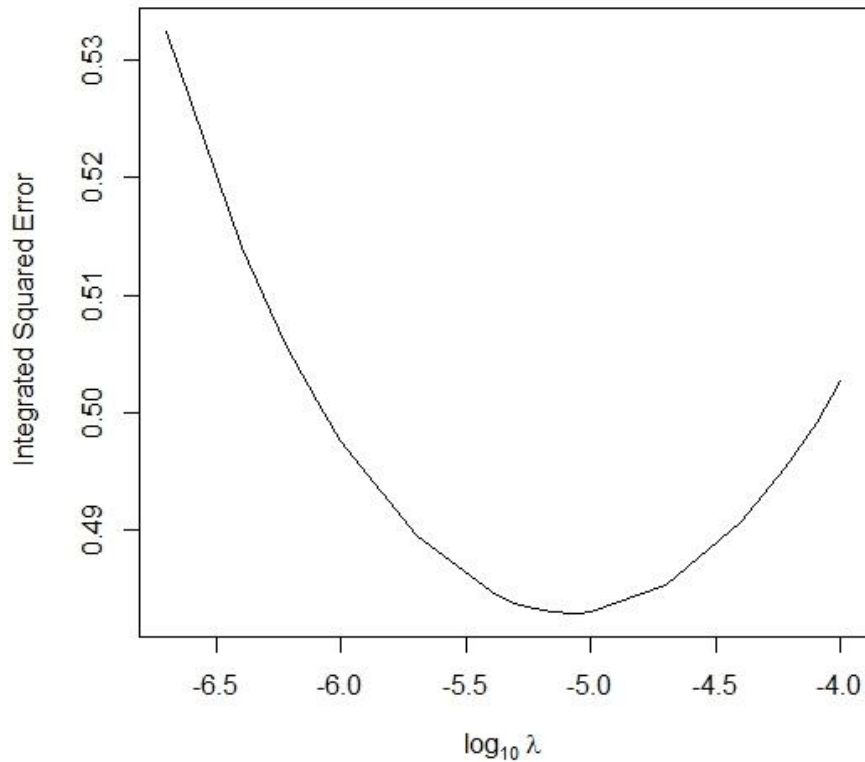
Figure 2: The values of the integrated squared error, $\| \hat{\beta}_\lambda - \beta_0 \|^2_{L^2(\mathcal{T})}$ as a function of λ with $n = 50$ and $m = 5$.

We estimate β based on these generated data according to Theorem 1. The integrated squared error, $\| \hat{\beta}_\lambda - \beta_0 \|^2_{L^2[0,1]} = \int_0^1 (\hat{\beta}_\lambda(t) - \beta_0(t))^2 dt$, as a function of the tuning parameter λ is given in Figure 2. The smoothing parameter is set to yield the smallest integrated squared error and therefore reflect the best performance of the estimating procedure.

The next numerical experiment intends to demonstrate the effect of sample size n and sampling frequency m. We repeat the experiment with varying combinations of $n = 25, 50, 100$ or 200, and $m = 1, 3$ or 5. The obtained value of the smoothing parameter, for each simulated data set, is reported in Table 1. As we see in the Table, the value of smoothing parameters decreases whenever either n or m increases. That is because, when number of observations increase we just need a little smoothing. By 200 runs replications the average of integrated squared errors and variance of $\hat{\beta}$ for each of the settings are reported in Table 2. It can be seen from the table that increasing either m or n leads to improved estimates, whereas such

improvement is more visible for small values of m. In addition, the average integrated squared error and the variance of estimated slope function decrease as either m or n increases.

Table 1: The value of smoothing parameter for different values of m and n

	$m = 1$	$m = 3$	$m = 5$
$n = 25$	4×10^{-4}	6×10^{-5}	2.5×10^{-5}
$n = 50$	2×10^{-4}	2×10^{-5}	8×10^{-6}
$n = 100$	6×10^{-5}	5×10^{-6}	2.5×10^{-6}
$n = 200$	2×10^{-5}	10^{-6}	5×10^{-7}

Table 2: Averaged integrated squared error $\| \widehat{\beta}_\lambda - \beta_0 \|_{L_2}^2$ for various combinations of m and n over 200 runs. The numbers in the parentheses are the variances

	$m = 1$	$m = 3$	$m = 5$
$n = 25$	0.6353 (0.1028)	0.5417 (0.0656)	0.5164 (0.0542)
$n = 50$	0.5788 (0.0741)	0.5055 (0.0476)	0.4830 (0.0363)
$n = 100$	0.5274 (0.0529)	0.4755 (0.0339)	0.4579 (0.0226)
$n = 200$	0.4978 (0.0392)	0.4554 (0.0254)	0.4440 (0.0183)

4. Application

CD4+ cells are a type of white blood cell that fights infection. CD4+ cells move throughout body, helping to identify and destroy germs such as bacteria and viruses. It is a well known fact that the human immune deficiency virus (HIV) causes AIDS by attacking CD4+ cells. HIV infects components of CD4+ cells. It directly and indirectly destroys CD4+ cells. The CD4+ count measures the number of CD4+ cells in a sample of blood. CD4+ counts are reported as the number of cells in a cubic millimeter of blood. A normal CD4+ count is around 1100 cells per cubic millimetre of blood. In general, HIV disease is progressing if the CD4+ count is going down. This means the immune system is getting weaker and HIV infected persons are more likely to get sick. In some persons, CD4+ counts can drop dramatically, even going down to zero.

The dataset considered here is from the Multicenter AIDS Cohort Study. The data contains CD4+ cell counts for a total of 369 infected men along with other variables (see Kaslow et al, 1987). In this study, the patients were scheduled to have their measurements made twice a year, however some subjects missed some of their scheduled visits. So, the actual times of measurement is random, irregular and sparse. The number of measurements for each individual varies from 1 to 12 yielding a total of 2376 records.

The objective here is to evaluate the effects of centred age at HIV infection on CD4+ counts based on model (1). In this dataset, centred age at HIV infection is time-independent covariate variable X and CD4+ counts are considered as response functions $Y(t)$ of time since seroconversion (time when HIV becomes detectable). Figure 3 displays individual trajectory of CD4+ count for the included 369 subjects along with the smooth estimated mean function. As we see in the figure, mean of CD+ counts decreases to below 1100 cells when time runs up.
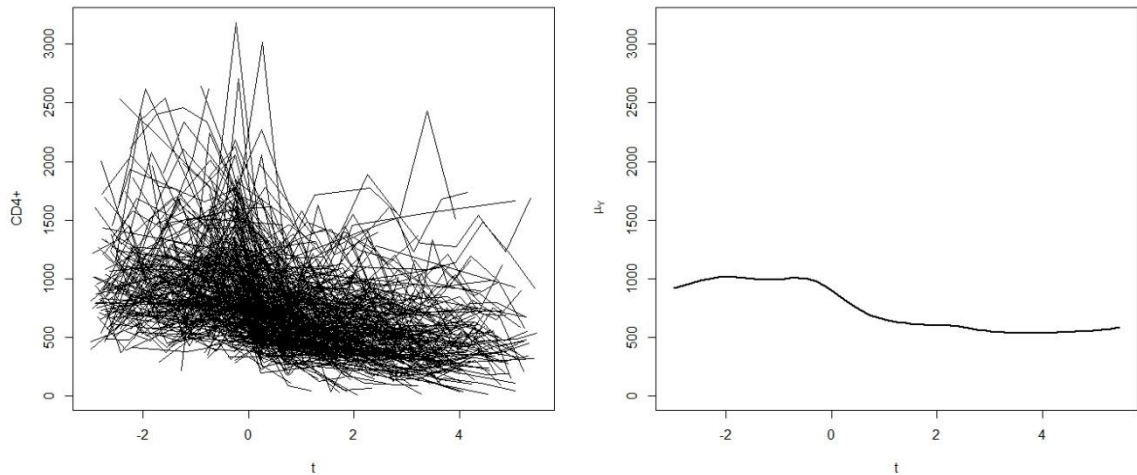

Figure 3: Left panel: Observed individual trajectories of CD4+ count. Right panel: the smooth estimate of the mean function for CD4+ count.

We calculate the smoothing parameter of $\hat{\beta}_\lambda$ using the following K-fold cross-validation. The 369 subjects is randomly partitioned into K roughly equal part. Of the K parts, a single part is retained as the validation data for testing the model, and the remaining $K - 1$ parts are used as training data. For each $k = 1, 2, \dots, K$, we fit the model with parameter λ to the other $K - 1$ parts, giving $\hat{\beta}_\lambda^{-k}$. Then we compute its error in predicting the kth part as follows:

$$E_k(\lambda) = \sum_{i \in ith\ part} \frac{1}{m_i} \sum_{j=1}^{m} \left(V_{ij} - \hat{\mu}_Y(T_{ij}) - (X_i - \bar{X})\hat{\beta}_\lambda^{-k}(T_{ij}) \right)^2.$$

The cross-validation error is defined by

$$CV(\lambda) = \frac{1}{K} \sum_{k=1}^{K} E_k(\lambda).$$

We compute $CV(\lambda)$ for many values of λ and choose the value of λ that makes $CV(\lambda)$ smallest. The estimated smoothing parameter $\lambda = 0.1$ was obtained by 5-fold cross-validation.

The estimated slope function, $\hat{\beta}_\lambda$ is shown in the left panel of Figure 4. The right panel of Figure 4 shows the estimated mean surface $E[Y(t)|X]$. The estimated slope function, $\hat{\beta}_\lambda$ initially increases and then after seroconversion decreases rapidly. Note that the surface of Figure 4 shows that the age at HIV infection has significant effect on the mean of CD4+ cell numbers. From this surface, we observe that for older peoples the means of CD4+ cell numbers are increasing function before seroconversion and decreasing after that. In contrast, for younger peoples these means function decrease until one year after seroconversion and then increase slowly. But for middle age the means of CD4+ cell numbers fluctuate around 1000 cells before seroconversion and during the second year after seroconversion decrease rapidly, and then increase slowly. In general, for older people, when time runs up, HIV infections spread faster. On the other hand, the situation for younger people before seroconversion is normal.
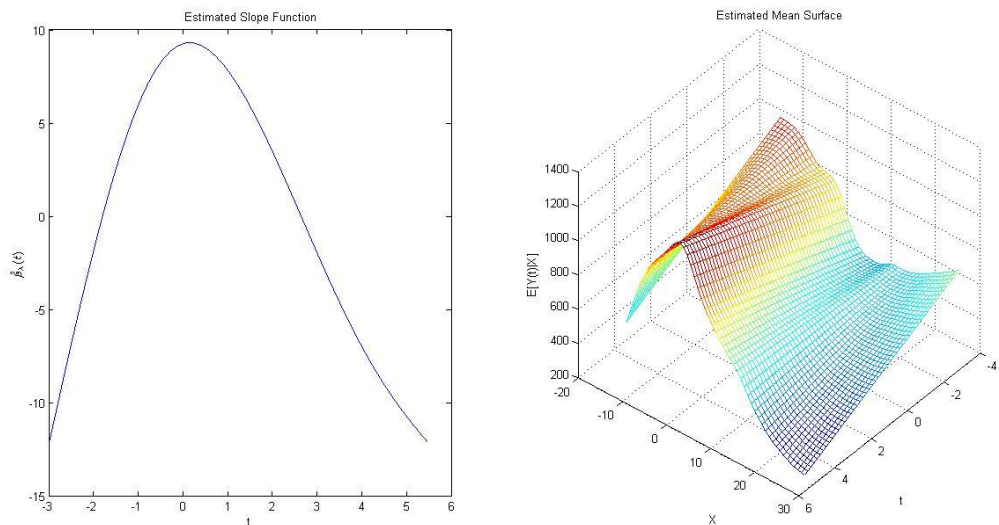

Figure 4: Left panel: Estimated slope function. Right panel: Estimated $E[Y(t)|X]$. for CD4+ count.

5. Discussions

We have developed a regularization method for slope function estimation in functional varying coefficient model with longitudinal response data and scalar covariate. Since longitudinal

data arise in many fields of science, the proposed method can be applicable in these fields. We note that although the methodology is proposed for irregular and sparsely sampled functional data, such as longitudinal data, it can be also applied for regularly and densely sampled functional data. Our simulation study shows that the procedure performs well. Also the estimator of the slope function introduced in closed form that makes it computationally very tractable.

The theoretical properties of the proposed estimator can be studied. For example, one can study convergence rate and asymptotically distribution of the slope function estimator. On the other hand, the existence of only one covariate is the limitation of our proposed model, that can be resolved by extension the model to the case with combination of longitudinal and scaler covariates. These ideas will be explored in future works.

References

[1] Abramowitz, M. and Stegun, I. (1965). *Handbook of mathematical functions with formulas, graphs and mathematical tables*. U. S. government printing office, Washington, D.C.

[2] Aronszajn, N. (1950). Theory of reproducing kernel. *Transactions of the American Mathematical Society* **68**, 337-404.

[3] 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.

[4] 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.

[5] Chiou, J. M., Ma, Y. and Tsai, C. L. (2012). Functional random effect time-varying coefficient model for longitudinal data. *Stat* **1**, 75-89.

[6] Cleveland, W. S., Grosse, E. and Shyu, W. M. (1991). Local regression models. Statistical Models in S (Chambers, J. M. and Hastie, T. J., eds), 309-376. Wadsworth and Books, Pacic Grove.

[7] Fan, J., and Zhang, W. (2008). Statistical methods with varying coefficient models. *Statistics and Its Interface* **1**, 179-195.

[8] Ferraty, F. and Vieu, P. (2006), *Nonparametric Functional Data Analysis: Methods, Theory, Applications and Implementations*. Springer, New York.

[9] Hastie, T. J. and Tibshirani, R. J. (1993). Varying-coefficient models. *Journal of the Royal Statistical Society B* **55**, 757-796.

[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] Horvàth, L. and Kokoszka, P. (2012). *Inference for Functional Data with Applications*. Springer, New York

[12] 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.

[13] 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.

[14] 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.

[15] Kauermann, G. and Tutz, G. (1999). On model diagnostics using varying coefficient models. *Biometrika* **86**, 119-128.

[16] 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.

[17] Pepe, M. S., (2000). Receiver operating characteristic methodology, *Journal of the American Statistical Association* **95**, 308-11.

[18] Noh, H. S. and Park, B. U. (2010). Sparse varying coefficient models for longitudinal data. *Statistica Sinica* **20**, 1183-1202.

[19] Ramsay, J. O. and Silverman, B. W. (2002). *Applied Functional Data Analysis: Methods and Case Studies*. Springer, New York.

[20] Ramsay, J. O. and Silverman, B. W. (2005), *Functional Data Analysis*, 2nd Edition. Springer, New York.

[21] Riesz, F. and Sz.-Nagy, B. (1955). *Functional Analysis*. Dover Publications, New York.

[22] Sentürk, D. and Müller, H. G. (2008). Generalized varying coefficient models for longitudinal data. *Biometrika* **95**, 653-666.

[23] Sentürk, D. and Müller, H. G (2010). Functional varying coefficient models for longitudinal data. *Journal of the American Statistical Association* **105**, 1256-1264.

[24] Sentürk, D. and Nguyen, D. V. (2011). Vary coefficient models for sparse noise-contaminated longitudinal data. *Statistica Sinica* **21**, 1831-1856.

[25] 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.

[26] Wahba, G. (1990). *Spline models for observational data*, SIAM, Philadelphia.

[27] Wu, C. O., Chiang, C. T. and Hoover, D. R. (1998). Asymptotic confidence regions for kernel smoothing of a varying-coefficient model with longitudinal data. *Journal of the American Statistical Association* **93**, 1388-1402.

[28] Wu, C. O. and Chiang, C. T. (2000). Kernel smoothing on varying coefficient models with longitudinal dependent variable. *Statistica Sinica* **10**, 433-456.

[29] Yao, F., Müller, H. G. and Wang, J. (2005). Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association* **100**, 577-590.

Behdad Mostafaiy

Department of Statistics

Shahid Beheshti University

Tehran 19834, Iran

behdad.mostafaiy@gmail.com

Mohammad Reza Faridrohani

Department of Statistics

Shahid Beheshti University

Tehran 19834, Iran

m_faridrohani@sbu.ac.ir