

Least Squares Parameter Estimation for Sparse Functional Varying Coefficient Model

Behdad Mostafaiy

*Department of Statistics, Faculty of Mathematical Science,
Shahid Beheshti University, Tehran, Iran;*

B_mostafaiy@sbu.ac.ir

Mohammad Reza Faridrohani

*Department of Statistics, Faculty of Mathematical Science,
Shahid Beheshti University, Tehran, Iran;*

m_faridrohani@sbu.ac.ir

Received 15 June 2015

Accepted 12 June 2016

Abstract

In the present paper, we study functional varying coefficient model in which both the response and the predictor are functions. We give estimates of the intercept and the slope functions in the case that the observations are sparse and noise-contaminated longitudinal data by using least squares representation of the model parameters. To estimate the parameter functions involved in the representation, we use a regularization method in some reproducing kernel Hilbert spaces. As we will see, our procedure is easy to implement. Also, we obtain the convergence rates of the estimators in the L^2 -sense. These convergence rates establish that the procedure performs well, especially, when sampling frequency or sample size increases.

Keywords: Functional varying coefficient model; Longitudinal data Analysis; Rate of convergence; Regularization; Reproducing kernel Hilbert space; Sparsity.

2000 Mathematics Subject Classification: 62G05, 34D05

1. Introduction

Nowadays, due to development of science and technology, it is possible to collect data which are infinite dimensional. Functional data analysis deals with such observations. See Ramsay and Silverman (2002; 2005) for an overview of methods and applications of functional data analysis. See also Ferraty and Vieu (2006) and, Horváth and Kokoszka (2012).

An extension of parametric regression models, commonly known as varying coefficient models, were introduced by Cleveland et al. (1991) and, Hastie and Tibshirani (1993). These models allow its regression coefficients to vary in respect to some predictors of interest. A collection of available methods is provided by Fan and Zhang (2008). The estimation approaches given in Fan and Zhang (2008) are based on polynomial spline, smoothing splines and local polynomial smoothing. Other references include Hoover et al. (1998), Wu et al. (1998), Kauermann and Tutz (1999), Wu and Chiang (2000), Chiang et al. (2001) and Huang et al. (2002; 2004). Most of these approaches are not applicable for sparse designs. In the case of sparse designs,

see for example Şentürk and Müller (2008), Noh and Park (2010), Şentürk and Müller (2010), Sentürk and Nguyen (2011), Chiou et al. (2012), and Şentürk et al. (2013).

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

$$Y(t) = \beta_0(t) + \beta_1(t)X(t) + Z(t), \quad t \in T \tag{1.1}$$

where $T \subset \mathbb{R}$ is a compact set, $Y(t)$ and $X(t)$ are square integrable response and predictor processes, respectively, $\beta_0(t)$ and $\beta_1(t)$ are smoothed intercept and slope functions, respectively, and $Z(t)$ is a mean zero random process, independent of $X(t)$. We want to estimate functions $\beta_0(t)$ and $\beta_1(t)$ in the situation that the observations are sparse and irregular longitudinal data and combined with some measurement errors. Following Yao et al. (2005), we model this situation as follows. Let U_{ij} and V_{ij} denote the observations of the random functions X_i and Y_i respectively at the random times T_{ij} , contaminated with measurement errors ε_{ij} and ϵ_{ij} respectively, which are assumed to be independent and identically distributed with means zero and variances σ_X^2 and σ_Y^2 respectively, and independent of the random functions. We may represent the observed data as

$$\begin{aligned} U_{ij} &= X_i(T_{ij}) + \varepsilon_{ij}, & j = 1, \dots, m_i; & \quad i = 1, \dots, n, \\ V_{ij} &= Y_i(T_{ij}) + \epsilon_{ij}, & j = 1, \dots, m_i; & \quad i = 1, \dots, n. \end{aligned} \tag{1.2}$$

Consider the population least squares representation for $(\beta_0(t), \beta_1(t))$ in model (1.1) as

$$(\beta_0(t), \beta_1(t)) = \arg \min_{(\theta_0(t), \theta_1(t))} \{E[Y(t) - \theta_0(t) - \theta_1(t)X(t)]^2\}.$$

The solutions are

$$\beta_0(t) = \mu_Y(t) - \beta_1(t)\mu_X(t), \quad \beta_1(t) = \frac{C_{XY}(t, t)}{C_{XX}(t, t)} \tag{1.3}$$

where $\mu_X(t) = E[X(t)]$ and $\mu_Y(t) = E[Y(t)]$ are mean functions of X and Y respectively, and $C_{XX}(s, t) = \text{cov}(X(s), X(t))$ and $C_{XY}(s, t) = \text{cov}(X(s), Y(t))$ are respectively covariance function of X and cross-covariance function of X and Y .

In the present paper, we use some regularization methods to estimate the parameter functions involved in representation (1.3). In addition, by assuming that the sample paths of X and Y are α times differentiable, we obtain convergence rates of the estimators. The obtained results show that our proposed method performs well.

The paper is organized as follows. In section 2, we review some basic facts of RKHS which are important concepts in the sequel. Section 3 contains our estimation methods and, finally, convergence rates of our estimators are obtained under some certain regularity assumptions in section 4.

2. Reproducing Kernel Hilbert Spaces

In this section, we will review, without proof, some basic facts about reproducing kernel Hilbert spaces (RKHSs). Verification of these results as well as more detailed discussions of RKHS theory can be found, for example, in Aronszajn (1950) and Wahba (1990).

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

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

When (i) and (ii) hold, \mathcal{H} is said to be an RKHS with reproducing kernel K . Relation (i) is called the reproducing property of K . If K is a reproducing kernel then it is symmetric, nonnegative definite and unique.

Conversely, if K is a nonnegative definite function on $T \times T$, a unique RKHS of functions on T with K as its reproducing kernel can be constructed.

In the sequel, we will denote the Hilbert space with reproducing kernel K by $\mathcal{H}(K)$. Also the correspond inner product and norm will be denoted by $\langle \cdot, \cdot \rangle_{\mathcal{H}(K)}$ and $\|\cdot\|_{\mathcal{H}(K)}$, respectively.

According to properties of inner product and by using properties (i) and (ii), for any $N, N' \in \mathbb{N}$, $\alpha_1, \dots, \alpha_N, \alpha'_1, \dots, \alpha'_{N'} \in \mathbb{R}$ and $t_1, \dots, t_N, t_1^{me}, \dots, t'_{N'} \in T$, we have

$$\begin{aligned} \left\langle \sum_{i=1}^N \alpha_i K(\cdot, t_i), \sum_{j=1}^{N'} \alpha'_j K(\cdot, t'_j) \right\rangle_{\mathcal{H}(K)} &= \sum_{i=1}^N \sum_{j=1}^{N'} \alpha_i \alpha'_j \langle K(\cdot, t_i), K(\cdot, t'_j) \rangle_{\mathcal{H}(K)} \\ &= \sum_{i=1}^N \sum_{j=1}^{N'} \alpha_i \alpha'_j K(t_i, t'_j). \end{aligned} \tag{2.1}$$

Now, let \mathcal{H}_1 and \mathcal{H}_2 be two Hilbert spaces of real-valued functions with, respectively, inner products $\langle \cdot, \cdot \rangle_{\mathcal{H}_1}$ and $\langle \cdot, \cdot \rangle_{\mathcal{H}_2}$. The tensor product space of \mathcal{H}_1 and \mathcal{H}_2 , $\mathcal{H}_1 \otimes \mathcal{H}_2$, is defined in the following fashion. For $f_1 \in \mathcal{H}_1$ and $g_1 \in \mathcal{H}_2$ define the bilinear form $f_1 \otimes g_1: \mathcal{H}_1 \times \mathcal{H}_2 \rightarrow \mathbb{R}$ by

$$[f_1 \otimes g_1](f_2, g_2) = \langle f_1, f_2 \rangle_{\mathcal{H}_1} \langle g_1, g_2 \rangle_{\mathcal{H}_2}, \quad (f_2, g_2) \in \mathcal{H}_1 \times \mathcal{H}_2.$$

Let \mathcal{H}_0 be the set of all finite linear combinations of such bilinear forms. An inner product, $\langle \cdot, \cdot \rangle_{\mathcal{H}_0}$ on \mathcal{H}_0 can be defined as follows:

$$\langle f_1 \otimes g_1, f_2 \otimes g_2 \rangle_{\mathcal{H}_0} = \langle f_1, f_2 \rangle_{\mathcal{H}_1} \langle g_1, g_2 \rangle_{\mathcal{H}_2}, \tag{2.2}$$

where $f_1, f_2 \in \mathcal{H}_1$ and $g_1, g_2 \in \mathcal{H}_2$. Then the tensor product space $\mathcal{H}_1 \otimes \mathcal{H}_2$ is defined as the completion of \mathcal{H}_0 under the inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}_0}$ defined in (2) and it is shown that it is a Hilbert space.

For an RKHS $\mathcal{H}(K)$, consider the tensor product space $\mathcal{H}(K \otimes K) := \mathcal{H}(K) \otimes \mathcal{H}(K)$. It can be shown that $\mathcal{H}(K \otimes K)$ is an RKHS with reproducing kernel

$$(K \otimes K)((s_1, t_1), (s_2, t_2)) = K(s_1, s_2)K(t_1, t_2).$$

To define our regularization method, we use RKHS.

3. Estimation Procedures

In the present section, we estimate the parameter functions involved in (1.3). Suppose the sample paths of random functions X and Y belong to the same, without loss of generality, RKHS $\mathcal{H}(K)$ almost surely, such that

$$E \|X\|_{\mathcal{H}(K)}^2 < \infty, \quad E \|Y\|_{\mathcal{H}(K)}^2 < \infty. \tag{3.1}$$

A result in Cai and Yuan (2010) shows that if condition (3.1) holds then $\mu_X, \mu_Y \in \mathcal{H}(K)$ and $C_{XX} \in \mathcal{H}(K \otimes K)$. In addition, Mostafaiy et. al. (2014) have shown condition (3.1) implies that $C_{XY} \in \mathcal{H}(K \otimes K)$. These results let us to define some regularization methods to estimate the parameter functions.

First, we estimate mean function of X . Notice that, mean function of Y can be estimated in the same way by using V_{ij} 's instead of U_{ij} 's in the following procedure.

Consider the following regularization method:

$$\hat{\mu}_X = \arg \min_{f \in \mathcal{H}(K)} \{ \ell_1(f) + \lambda_1 \|f\|_{\mathcal{H}(K)}^2 \} \tag{3.2}$$

where

$$\ell_1(f) = \frac{1}{n} \sum_{i=1}^n \frac{1}{m_i} \sum_{j=1}^{m_i} (U_{ij} - f(T_{ij}))^2$$

and $\lambda_1 \geq 0$ is a smoothing parameter that balances the fidelity to the data measured by ℓ_1 and the smoothness of the estimate measured by the squared RKHS norm.

According to the so-called representer lemma (see Wahba 1990), the solution of the minimization problem (3.2) can be expressed as

$$\hat{\mu}_{X,a}(t) = \sum_{i=1}^n \sum_{j=1}^{m_i} a_{ij} K(t, T_{ij}) \tag{3.3}$$

for some coefficient vector $\mathbf{a} = [a_{11}, \dots, a_{1m_1}, a_{21}, \dots, a_{nm_n}]' \in \mathbb{R}^{N_1}$ with $N_1 = \sum_{i=1}^n m_i$. This result demonstrates that although the minimization in (3.2) is taken over an infinite dimensional space, the solution can be found in a finite dimensional subspace, and it suffices to evaluate coefficient vector \mathbf{a} in (3.3). To do this, let \mathbf{P} be an $N_1 \times N_1$ matrix defined by

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

where

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

By using (2.1), we have

$$\begin{aligned} \|\hat{\mu}_{X,a}\|_{\mathcal{H}(K)}^2 &= \left\langle \sum_{i_1=1}^n \sum_{j_1=1}^{m_{i_1}} a_{i_1 j_1} K(t, T_{i_1 j_1}), \sum_{i_2=1}^n \sum_{j_2=1}^{m_{i_2}} a_{i_2 j_2} K(t, T_{i_2 j_2}) \right\rangle_{\mathcal{H}(K)} \\ &= \sum_{i_1=1}^n \sum_{j_1=1}^{m_{i_1}} \sum_{i_2=1}^n \sum_{j_2=1}^{m_{i_2}} a_{i_1 j_1} a_{i_2 j_2} K(T_{i_1 j_1}, T_{i_2 j_2}) \\ &= \mathbf{a}' \mathbf{P} \mathbf{a} \end{aligned}$$

Notice that

$$\ell_1(\hat{\mu}_{X,a}) = \|\mathbf{U} - \mathbf{P} \mathbf{a}\|_{\ell^2}^2,$$

where ℓ^2 is the space of square summable sequences and $\mathbf{U} = [U_{11}, \dots, U_{1m_1}, U_{21}, \dots, U_{nm_n}]'$.

Therefore we have

$$\ell_1(\hat{\mu}_{X,a}) + \|\hat{\mu}_{X,a}\|_{\mathcal{H}(K)}^2 = \|\mathbf{U} - \mathbf{P} \mathbf{a}\|_{\ell^2}^2 + \mathbf{a}' \mathbf{P} \mathbf{a} \tag{3.4}$$

The right hand side of (4) is quadratic in \mathbf{a} and so it is not hard to see that the minimizer of it is

$$\mathbf{a} = (\mathbf{P} + N\lambda_1 \mathbf{I})^{-1} \mathbf{U} \tag{3.5}$$

where \mathbf{I} is an $N_1 \times N_1$ identity matrix.

To estimate C_{XX} and C_{XY} , let

$$\begin{aligned} C_{XX,i}(T_{ij_1}, T_{ij_2}) &= [U_{ij_1} - \hat{\mu}_X(T_{ij_1})][U_{ij_2} - \hat{\mu}_X(T_{ij_2})], \\ C_{XY,i}(T_{ij_1}, T_{ij_2}) &= [U_{ij_1} - \hat{\mu}_X(T_{ij_1})][V_{ij_2} - \hat{\mu}_Y(T_{ij_2})] \end{aligned}$$

and define the following regularization methods:

$$\hat{C}_{XX} = \arg \min_{C \in \mathcal{H}(K) \otimes \mathcal{H}(K)} \{ \ell_2(C) + \lambda_2 \| C \|_{\mathcal{H}(K) \otimes \mathcal{H}(K)}^2 \},$$

$$\hat{C}_{XY} = \arg \min_{C \in \mathcal{H}(K) \otimes \mathcal{H}(K)} \{ \ell_3(C) + \lambda_3 \| C \|_{\mathcal{H}(K) \otimes \mathcal{H}(K)}^2 \}$$

where

$$\ell_2(C) = \frac{1}{n} \sum_{i=1}^n \frac{1}{m_i(m_i - 1)} \sum_{1 \leq j_1 \neq j_2 \leq m_i} \{ C_{XX,i}(T_{ij_1}, T_{ij_2}) - C(T_{ij_1}, T_{ij_2}) \}^2,$$

$$\ell_3(C) = \frac{1}{n} \sum_{i=1}^n \frac{1}{m_i^2} \sum_{j_1=1}^{m_i} \sum_{j_2=1}^{m_i} \{ C_{XY,i}(T_{ij_1}, T_{ij_2}) - C(T_{ij_1}, T_{ij_2}) \}^2$$

and λ_2 and λ_3 are smoothing parameters. By representer lemma, we have

$$\hat{C}_{XX}(s, t) = \sum_{i=1}^n \sum_{j_1=1}^{m_i} \sum_{j_2=1}^{m_i} b_{ij_1 j_2} K(s, T_{ij_1}) K(t, T_{ij_2}),$$

for some $\mathbf{b} = [b_{111}, \dots, b_{11m_1}, b_{121}, \dots, b_{nm_n m_n}] \in \mathbb{R}^{N_2}$ with $N_2 = \sum_{i=1}^n m_i^2$, and

$$\hat{C}_{XY}(s, t) = \sum_{i=1}^n \sum_{j_1=1}^{m_i} \sum_{j_2=1}^{m_i} c_{ij_1 j_2} K(s, T_{ij_1}) K(t, T_{ij_2}),$$

for some $\mathbf{c} = [c_{111}, \dots, c_{11m_1}, c_{121}, \dots, c_{nm_n m_n}] \in \mathbb{R}^{N_2}$.

Coefficient vectors \mathbf{b} and \mathbf{c} can be evaluated in the same manner as coefficient vector \mathbf{a} in the estimation procedure of μ_X .

Now, from (1.3) the plug-in estimates of β_0 and β_1 are given by

$$\hat{\beta}_1(t) = \frac{\hat{C}_{XY}(t, t)}{\hat{C}_{XX}(t, t)}, \quad \hat{\beta}_0(t) = \hat{\mu}_Y(t) - \hat{\beta}_1(t) \hat{\mu}_X(t). \tag{3.6}$$

Notice that the above estimators depend on smoothing parameters λ_1 , λ_2 and λ_3 . In practice these smoothing parameters can be chosen by some optimization criteria such as cross-validation.

4. Rates of Convergence

In this section, we obtain convergence rates of the estimators in terms of integrated squared errors. To do this, we assume the following conditions. These conditions also have been adopted by Cai and Yuan (2010). Let $\mathcal{F}(\alpha; M, c)$ be the collection of probability measures defined on (X, Y) such that

- The sample paths of α time differentiable X and Y belong to $\mathcal{H}(K)$ almost surely and there exists constant $M > 0$ such that $E[\| X \|_{\mathcal{H}(K)}^2] < M$ and $E[\| Y \|_{\mathcal{H}(K)}^2] < M$.
- K is a Mercer kernel with eigenvalues, ρ_k , satisfying $\rho_k \asymp k^{-2\alpha}$, where for two positive sequences r_n and s_n , $r_n \asymp s_n$ means that r_k/s_k is bounded away from 0 and ∞ as $k \rightarrow \infty$, that is,

$$0 < \liminf_{k \rightarrow \infty} (r_k/s_k) \leq \limsup_{k \rightarrow \infty} (r_k/s_k) < \infty.$$

- For Z being either X or Y , there exists a constant $c > 0$ such that

$$E[Z^4(t)] \leq c (E[Z^2(t)])^2 \quad \text{for all } t \in T,$$

and

$$E \left[\int_T Z(t) f(t) dt \right]^4 \leq c \left(E \left[\int_T Z(t) f(t) dt \right]^2 \right)^2,$$

for any $f \in L^2(T)$.

The condition (a) imposes smoothness of the processes X and Y . The boundedness requirement $E[\|X\|_{\mathcal{H}(K)}^2] < M$ and $E[\|Y\|_{\mathcal{H}(K)}^2] < M$ in (a) is a technical condition. The condition (b) guaranties the smoothness of the kernel function K . As shown in Micchelli and Wahba (1981), the Sobolev space \mathcal{W}_2^α satisfies the condition (b). Finally, the condition (c) concerns the fourth moment of a linear functional of Z . This condition is satisfied with $c = 3$ for a Gaussian process because $\int_T Z(t) f(t) dt$ is normally distributed.

Let m be the harmonic mean of m_1, \dots, m_n , that is,

$$m := \left(\frac{1}{n} \sum_{i=1}^n \frac{1}{m_i} \right)^{-1}.$$

For the smoothing parameters we assume

$$\lambda_1 = (nm)^{-\frac{2\alpha}{2\alpha+1}}, \quad \lambda_2, \lambda_3 = \left(\frac{\log(n)}{nm} \right)^{\frac{2\alpha}{2\alpha+1}}. \quad (4.1)$$

Given a sequence c_n of positive constants, we shall use $O_p(c_n)$ to denote random variable R_n which satisfies

$$\lim_{D \rightarrow \infty} \limsup_{n \rightarrow \infty} \sup_{F \in \mathcal{F}(\alpha; M, c)} P_F \{ |R_n| > Dc_n \} = 0.$$

The following result gives the rates of convergence for the regularized estimators $\hat{\mu}_X$ and $\hat{\mu}_Y$.

Theorem 1. Assume that T_{ij} 's are independent and identically distributed with a density bounded away from zero on T . Suppose the smoothing parameter λ_1 satisfies (4.1). Then

$$\| \hat{\mu}_Z - \mu_Z \|_{L^2}^2 = O_p \left(n^{-1} + (nm)^{-\frac{2\alpha}{2\alpha+1}} \right), \quad \text{for } Z = X \text{ or } Y. \quad (4.2)$$

Proof. The proof is given in Cai and Yuan (2011). □

The following theorem gives convergence rates of \hat{C}_{XX} and \hat{C}_{XY} .

Theorem 2. Assume that $E(\epsilon^4) < \infty$, $E(\epsilon^4) < \infty$ and T_{ij} 's are independent and identically distributed with a density bounded away from zero on T . Suppose the smoothing parameters λ_1, λ_2 and λ_3 satisfy (4.1). Then

$$\| \hat{C}_{XX} - C_{XX} \|_{L^2}^2 = O_p \left(\frac{1}{n} + \left(\frac{\log(n)}{nm} \right)^{\frac{2\alpha}{2\alpha+1}} \right)$$

and

$$\| \hat{C}_{XY} - C_{XY} \|_{L^2}^2 = O_p \left(\frac{1}{n} + \left(\frac{\log(n)}{nm} \right)^{\frac{2\alpha}{2\alpha+1}} \right).$$

Proof. The proof is given in Cai and Yuan (2010), and Mostafaiy et. al. (2014). □

The convergence rates of $\hat{\beta}_0$ and $\hat{\beta}_1$ are established in the following theorem.

Theorem 3. Under the conditions of Theorem 2, the varying coefficient estimators (3.6) satisfy

$$\| \hat{\beta}_k - \beta_k \|_{L^2}^2 = O_p \left(\frac{1}{n} + \left(\frac{\log(n)}{nm} \right)^{\frac{2\alpha}{2\alpha+1}} \right), \quad \text{for } k = 0, 1.$$

Proof. The proof follows from (3.6) and Theorems 1, 2. □

5. Conclusions and Extensions

We have shown that the convergence rates for the proposed estimators of the slope and the intercept functions in the L^2 -sense are $\frac{1}{n} + \left(\frac{\log(n)}{nm} \right)^{\frac{2\alpha}{2\alpha+1}}$. When m is small, that is $m \ll n^{\frac{1}{2\alpha}} \log n$, the convergence rates are determined by factor $\left(\frac{\log(n)}{nm} \right)^{\frac{2\alpha}{2\alpha+1}}$ which depend jointly on the values of both m and n . On the other hand, when m is large, that is $m \gg n^{\frac{1}{2\alpha}} \log n$, the rates turn to $\frac{1}{n}$ regardless of m .

In the present paper, we have focused on stochastic processes defined on a compact subset T of the real line \mathbb{R} . We note that this assumption can be extended for more general compact domains. For example, for $T = [0, 1]^r$, it is not hard to show that the convergence rates become $\frac{1}{n} + \left(\frac{\log(n)}{nm} \right)^{\frac{2\alpha}{2\alpha+r}}$.

References

- [1] N. Aronszajn, Theory of reproducing kernel, *Transaction of American Mathematical Society* **68** (1950) 337-404.
- [2] T. Cai and M. Yuan, Nonparametric covariance estimation for functional and longitudinal data, Technical Report (University of Pennsylvania and Georgia Institute of Technology, 2010).
- [3] T. Cai and M. Yuan, Optimal estimation of the mean function based on discretely sampled functional data: Phase transition, *The Annals of Statistics* **39** (2011) 2330-2355.
- [4] C. Chiang, J. A. Rice and C. O. Wu, Smoothing spline estimation for varying coefficient models with repeatedly measured dependent variables, *Journal of the American Statistical Association* **96** (2001) 605-617.
- [5] J. M. Chiou, Y. Ma and C. L. Tsai, Functional random effect time-varying coefficient model for longitudinal data. *Stat* **1** (2012) 75-89.
- [6] W. S. Cleveland, E. Grosse, and W. M. Shyu, Local regression models. *Statistical Models in S* (Chambers, J. M. and Hastie, T. J., eds, Wadsworth and Books, Pacic Grove. 1991) pp. 309-376.
- [7] J. Fan and W. Zhang, Statistical methods with varying coefficient models, *Statistics and Its Interface* **1** (2008) 179-195.
- [8] F. Ferraty and P. Vieu, *Nonparametric Functional Data Analysis: Methods, Theory, Applications and Implementations* (Springer, New York, 2006).
- [9] T. J. Hastie and R. J. Tibshirani, Varying-coefficient models, *Journal of the Royal Statistical Society, B* **55** (1993) 757-796.
- [10] D. R. Hoover, J. A. Rice, C. O. Wu and L.-P. Yang, Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data, *Biometrika*, **85** (1998) 809-822.
- [11] L. Horvath and P. Kokoszka, *Inference for Functional Data with Applications*. Springer, New York (2012).
- [12] J. Z. Huang, C. O. Wu and L. Zhou, Varying-coefficient models and basis function approximations for the analysis of repeated measurements *Biometrika* **89** (2002) 111-128.
- [13] J. Z. Huang, C. O. Wu and L. Zhou, Polynomial spline estimation and inference for varying coefficient models with longitudinal data, *Statistica Sinica* **14** (2004) 763-788.
- [14] G. Kauermann and G. Tutz, On model diagnostics using varying coefficient models, *Biometrika* **86** (1999) 119-128.
- [15] C. Micchelli and G. Wahba, Design problems for optimal surface interpolation. In *Approximation Theory and Applications* (Z. Ziegler, ed.). (Academic Press, New York, 1981), pp. 329-347.
- [16] B. Mostafaiy, M. FridRohani, and S. Chenouri, Sparse Functional Linear Regression and Singular Functions by Reproducing Kernel Hilbert Space Methods, Manuscript (2014).
- [17] H. S. Noh and B. U. Park, Sparse varying coefficient models for longitudinal data, *Statistica Sinica* **20** (2010) 1183-1202.
- [18] J. O. Ramsay and B. W. Silverman, *Applied Functional Data Analysis: Methods and Case Studies* (Springer, New York, 2002).

- [19] J. O. Ramsay and B. W. Silverman, *Functional Data Analysis*, 2nd Edition. (Springer, New York, 2005).
- [20] D. Senturk and H. G. Muller, Generalized varying coefficient models for longitudinal data, *Biometrika* **95** (2008) 653-666.
- [21] D. Senturk and H. G. Muller, Functional varying coefficient models for longitudinal data, *Journal of the American Statistical Association*, **105** (2010) 1256-1264.
- [22] D. Senturk and D. V. Nguyen, Vary coefficient models for sparse noise- contaminated longitudinal data, *Statistica Sinica* **21** (2011) 1831-1856.
- [23] D. Senturk, L. S. Dalrymple, S. M. Mohammed, G. A. Kaysen and D. V. Nguyen, Modeling time-varying effects with generalized and unsynchronized longitudinal data, *Statistics in Medicine* **32** (2013) 2971-2987.
- [24] G. Wahba, *Spline models for observational data*. (SIAM, Philadelphia, 1990).
- [25] C. O. Wu, C. T. Chiang and D. R. Hoover, Asymptotic confidence regions for kernel smoothing of a varying-coefficient model with longitudinal data, *Journal of the American Statistical Association* **93** (1998)1388-1402.
- [26] C. O. Wu and C. T. Chiang, Kernel smoothing on varying coefficient models with longitudinal dependent variable, *Statistica Sinica* **10** (2000) 433-456.
- [27] F. Yao, H. G. Muller, and J. Wang, Functional data analysis for sparse longitudinal data, *Journal of the American Statistical Association* **100** (2005) 577-590.