FUNCTIONAL INVERSE REGRESSION AND REPRODUCING KERNEL HILBERT SPACE

A Dissertation

by

HAOBO REN

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

August 2005

Major Subject: Statistics

FUNCTIONAL INVERSE REGRESSION AND REPRODUCING KERNEL HILBERT SPACE

A Dissertation

by

HAOBO REN

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Chair of Committee,	Tailen Hsing
Committee Members,	Raymond J. Carroll
	Suojin Wang
	Qi Li
Head of Department,	Simon J. Sheather

August 2005

Major Subject: Statistics

ABSTRACT

Functional Inverse Regression and Reproducing Kernel
Hilbert Space. (August 2005)
Haobo Ren, B.S., Peking University;
M.S., Peking University
Chair of Advisory Committee: Dr. Tailen Hsing

The basic philosophy of Functional Data Analysis (FDA) is to think of the observed data functions as elements of a possibly infinite-dimensional function space. Most of the current research topics on FDA focus on advancing theoretical tools and extending existing multivariate techniques to accommodate the infinite-dimensional nature of data. This dissertation reports contributions on both fronts, where a unifying inverse regression theory for both the multivariate setting (Li 1991) and functional data from a Reproducing Kernel Hilbert Space (RKHS) prospective is developed.

We proposed a functional multiple-index model which models a real response variable as a function of a few predictor variables called indices. These indices are random elements of the Hilbert space spanned by a second order stochastic process and they constitute the so-called Effective Dimensional Reduction Space (EDRS). To conduct inference on the EDRS, we discovered a fundamental result which reveals the geometrical association between the EDRS and the RKHS of the process. Two inverse regression procedures, a "slicing" approach and a kernel approach, were introduced to estimate the counterpart of the EDRS in the RKHS. Further the estimate of the EDRS was achieved via the transformation from the RKHS to the original Hilbert space. To construct an asymptotic theory, we introduced an isometric mapping from the empirical RKHS to the theoretical RKHS, which can be used to measure the distance between the estimator and the target. Some general

computational issues of FDA were discussed, which led to the smoothed versions of the functional inverse regression methods. Simulation studies were performed to evaluate the performance of the inference procedures and applications to biological and chemometrical data analysis were illustrated.

To my wife, Caixia Zhao

and in memory of my father, Xizheng Ren

ACKNOWLEDGMENTS

I want to express my sincerest appreciation to my advisor, Dr. Tailen Hsing, who brought me into the exciting world of functional data analysis, showed me the amazing idea of inverse regression and continuously supported me to complete my Ph.D. degree during the past few years. His broad knowledge and strong intuition in probability and statistics, and foremost his great personality have set an example to me for my future career.

Special thanks are due to all my committee members: Dr. Raymond J. Carroll, who encouraged me on my theoretical studies during my preliminary exam and his workshop in 2004, Dr. Qi Li, who discussed with me many interesting problems in statistics and econometrics, and Dr. Suojin Wang, who taught me the theory of sampling, resampling and asymptotics. I gratefully acknowledge Dr. Michael Sherman for his willingness and time to substitute for Dr. Carroll, to attend and participate in my defense. Dr. Sherman introduced me to spatial statistics in the fall of 2000, which was instrumental in preparing for my thesis. Further I wish to mention Dr. Randy L. Eubank, who introduced the beautiful theory of reproducing kernel Hilbert space to the seminar of functional data analysis in the spring of 2003, which I attended.

I am indebted to my extended family including my relatives living in China for their patience and understanding while I was studying in Texas A&M University. My greatest debt is to my wife, who helped me type this thesis and continued to provide love and support through this process.

TABLE OF CONTENTS

Page

CHAPTER Page

LIST OF FIGURES

FIGURE Page

CHAPTER I

INTRODUCTION

Stochastic statistics, longitudinal data analysis (LDA) and functional data analysis (FDA), are three closely related areas and comprise a trilogy in modern statistics. Stochastic statistics, including time series analysis and spatial statistics, makes inference based on a long observed trajectory of a random process. Longitudinal data often refer to many short series of records, for which various parametric models combined with nonparametric smoothing are the major approaches. Compared to stochastic statistics and LDA, FDA is a more general area of research which is nonparametric in nature, while parametric modeling can be done in abstract function spaces to capture the functional features of the data.

FDA is largely motivated by the emergence of an abundance of functional data. With the rapid development of accurate instruments, measurement could be taken continuously over a period of time to produce data in functional form. Statisticians tend to compare functional data with longitudinal data, and view functional data as densely observed longitudinal data, similar to the generalization from repeated measurement model to longitudinal data analysis. However, the concept of functional data has brought more profound, creative and revolutionary ideas to statistics.

The basic philosophy of FDA is to think of each data function as a single observational unit due to the precise and frequent sampling procedure, although in reality it is only possible to observe the function at a finite number of grid points. As a result, FDA should

---

This dissertation follows the style and format of the *Journal of the American Statistical Association*.

be considered in function spaces which are likely to be infinite-dimensional. This leads to considerations which are substantially different from those of the traditional multivariate analysis. Another aspect of FDA is to treat every observation element as a sample path from a specific stochastic process, hence stochastic inference is relevant.

Most of the current research on FDA focuses on two directions: to create advanced theoretical tools and to adapt existing multivariate techniques to infinite-dimensional functional data. Reproducing Kernel Hilbert Space (RKHS) is an effective and profound device for statistical analysis involving infinite dimensional data objects, while inverse regression (IR) is a renowned and brilliant idea in multivariate analysis, both of which are recently introduced to the FDA context. This research will focus on the functional inverse regression (FIR) from the RKHS prospective.

This dissertation is organized as follows. Chapter II will give a literature review on FDA, IR and RKHS, respectively. In Chapter III, we present a functional multiple-index model and explore its probabilistic structure. The estimation framework within RKHS is described and the associated asymptotic theory is constructed in Chapter IV. Chapter V investigates the transformation from the RKHS to the Hilbert space of the stochastic process. Chapter VI will discuss some computational issues in FDA and propose the smoothed versions of the estimation. Empirical studies including simulation and data analysis are reported in Chapter VII. Chapter VIII gives a brief conclusion.

CHAPTER II

LITERATURE REVIEW

This research comprises three components: Functional Data Analysis (FDA), Inverse Regression (IR) and Reproducing Kernel Hilbert Space (RKHS). In this chapter, we shall make an overview of current research topics on FDA and IR and recall the history of RKHS.

2.1 FDA

More and more data nowadays can be easily collected in the form of curves or images. Statistical methods for analyzing such data are termed "functional data analysis", a term coined by Ramsay and Dalzell (1991). The basic philosophy of FDA is to think of each observed curve or image as a single observation rather than a collection of individual observations. In terms of the different philosophy and methodology on how to treat these data, the current FDA research can be largely classified into three schools: English, French and Stochastic. The relationship among FDA, longitudinal data analysis and time series analysis will also be discussed in this review.

2.1.1 English School

The methodologies of the English school are represented by Ramsay and Silverman (1997) with the applications illustrated in Ramsay and Silverman (2002). Smoothing techniques dominate this school. Each sampling element is viewed as a smooth function, and the key step is to convert raw, discretely observed, data into genuine functional elements by various smoothing techniques including basis function (Fourier, wavelets and B-splines), localized smoothing (kernel and local polynomial regression) and roughness penalty or regularization approach (smoothing splines). There is a vast literature in the field of statistical smoothing,

see the monographs by Eubank (1999), Green and Silverman (1994), Simonoff (1996) and Wand and Jones (1995) for an overview of these nonparametric techniques.

So far, considerations on processing and displaying of functional data have focused on pattern or structure search and registration (Kneip and Gassar 1992, Gassar and Kneip 1995, Ramsay and Li 1997, Liu and Muller 2003, and Rossi, Delannay, Conan-Guez and Verleysen 2005), the estimation of mean and covariance structure (Rice and Silverman 1991), principal component analysis (Silverman 1996), canonical correlation analysis (Leugrans, Moyeed and Silverman 1993) and linear discriminant analysis (Hastie, Buja and Tibshirani 1995).

As in multivariate data analysis, various types of linear models are the most-studied topics. Obviously, there are many different varieties of functional linear models since we can entertain a number of different combinations of functional and scalar components in both the response and predictor variables. The simplest and most-studied one has a scalar response and functional predictor. Hall and Horowitz (2004) and Cai and Hall (2005) discussed the large sample properties of functional linear regression. Crambes (2005) proposed total least square approach for functional linear measurement error models. James (2002) and Müller and Stadtmüller (2004) provided some extensions of generalized linear models and quasi-likelihood method to functional predictors. Escabias, Aguilear and Valderrama (2004) introduced the functional principal component logistic regression.

Recently, James and Siverman (2005) introduced an adaptive functional model which extends generalized linear models, generalized additive models and projection pursuit regression to handle functional predictors. It takes the form

$$g(E(y|x)) = \beta_0 + \sum_{k=1}^{r} f_k \left(\int x(t)\beta_k(t)\,\mathrm{d}t \right),$$

where x is the predicting curve, y is an exponentially distributed scalar response, respectively, g is the link function, β_k's are coefficient functions, and f_k's are the suitably smooth

curves as in additive models or projection pursuit regression models. A penalized maximum likelihood estimation approach is used to fit both f_k's and β_k's.

On the other hand, for works involving both functional response and predictors, Malfait and Ramsay (2003) considered the historical linear model

$$y(t) = \alpha(t) + \int_0^S x(s)\beta(s,t)\,ds + \varepsilon(t),$$

where $x(s), s \in [0,S]$ and $y(t), t \in [0,T]$ are the exploratory and response curves, respectively, ε is the error process and β is the bivariate regression coefficient function. They applied the finite element method to estimate β.

2.1.2 French School

The name of FDA is twofold: the data is functional and the analytical method uses functional analysis. The French FDA school applies functional analysis extensively. The basic observational unit is treated more abstractly as an element in a function space and most functional analysis concepts and tools such as operator theory can help form the mathematical foundation of functional data analysis.

Dauxois, Pousse and Romain (1982) was one of the pioneering works in FDA which constructed an asymptotic theory for functional principal component analysis by using pure analysis and measure-theoretic language. Fine (2003) explored the similar theory for canonical correlation analysis in a Hilbert space using operator and tensor approach. Bosq (1991) proposed a first order Hilbertian autoregressive model in H, where H is a real separable Hilbert space equipped with norm $\|\cdot\|$ and inner product $\langle\cdot,\cdot\rangle$, to describe the dynamics of a sequence $\{X_t\}$ of H-valued random variables such that

$$X_t = \rho(X_{t-1}) + \varepsilon_t, \qquad t \in \mathbf{Z},$$

where $\{\varepsilon_t\}$ is an H-white noise, which means that ε_t's are H-valued, identically and independently distributed (i.i.d.) random variables. It is assumed that $0 < E\|\varepsilon_t\|^2 < \infty$ and

$E\varepsilon_t = 0$, and ρ is a symmetric compact linear operator on H with $\|\rho\|_{B(H)} < 1$, where $B(H)$ is the space of linear bounded operator on H. Additionally, assume that $E\|X_0\|^4 < \infty$.

Define the covariance operator C of X_0 by $C = E(X_0 \otimes X_0)$, and the cross-covariance operator D of (X_0, X_1) by $D = E(X_0 \otimes X_1)$, where $\otimes$ is the tensor product in H, meaning for $a, b \in H$, $a \otimes b \in B(H)$, such that

$$a \otimes b(f) = \langle a, f \rangle b, \qquad \forall f \in H.$$

By the strict stationarity one can derive that $D = \rho C$.

To estimate ρ, the inverse of C must be treated properly since it could be either non-existing or unbounded. The following projection method was then developed in the paper.

Let $(\lambda_k, \nu_k), k \in \mathbf{N}$, be the eigenvalues and eigenfunctions of C arranged in the order of $\lambda_1 \geq \lambda_2 \geq \cdots \geq 0$. Define the subspace of H, $V_K = \text{span}(\nu_1, \cdots, \nu_K)$, and the associated projector by

$$\Pi^K = \underset{V_K}{Proj} = \sum_{k=1}^{K} \nu_k \otimes \nu_k.$$

The projected covariance and cross-covariance are $C_{(K)} = \Pi^K C \Pi^K$ and $D_{(K)} = \Pi^K D \Pi^K$, respectively. We then consider an estimator $\rho_K = D_{(K)} C_{(K)}^{-1}$ of ρ.

This idea was borrowed directly by Cardot, Ferraty and Sarda (1999) to estimate Ψ in the functional linear model

$$Y = \Psi(X) + \varepsilon,$$

where Y and ε are real random variables, X is a H-valued random variable with $E\|X\|^2 < \infty$, and Ψ is a real linear continuous functional on H. This approach also formed the basis of functional principal component regression. In the follow-up studies, Cardot, Ferraty and Sarda (2003) addressed the computational implementation by penalized B-spline, and Cardot, Ferraty, Mas, and Sarda (2003) discussed the testing of hypothesis of $H_0 : \Psi = 0$

with certain computing issues addressed in Cardot, Goia and Sarda (2004). Goia (2003) also considered the model selection problem.

Among other topics, Cardot (2000) investigated the smoothing effect in functional principal component analysis, Cardot, Crambes and Sarda (2004) proposed the functional quantile regression, Cardot and Sarda (2005) studied the functional generalized linear model, and Cueva, Febrero and Fraiman (2002) handled the functional linear model with functional response.

2.1.3 Stochastic School

The stochastic school treats each functional sample unit as a realization from a random process. Strictly speaking, this is different from classical stochastic statistics (Rao 2000) in which the inference is based on only one realization of a stochastic process. In FDA, the covariance function is crucial for many analyses, therefore the Karhunen–Loéve expansion (Ash and Gardner 1975) is one of the most useful tools in this approach.

Karhunen–Loéve expansion: Let $\{X(t), t \in [a,b]\}$ be a L^2 process ($E(X(t)^2 < \infty$ for all $t \in [a,b]$) with zero mean and continuous covariance K. Let $\{e_n, n = 1,2,\ldots\}$ be an orthonormal basis for the space spanned by the eigenfunctions of the nonzero eigenvalues of the integral operator associated with K, with e_n taken as an eigenvector corresponding to the eigenvalue λ_n. Then

$$X(t) = \sum_{n=1}^{\infty} Z_n e_n(t), \qquad t \in [a,b],$$

where $Z_n = \int_a^b X(t)e_n(t)dt$, and the Z_n are orthogonal random variables with $E(Z_n) = 0$ and $E(Z_n^2) = \lambda_n$. The series converges in L^2 to $X(t)$, uniformly in t.

Huang, Quek and Phoon (2001) studied the performance of the truncated Karhunen–Loéve expansion method in the simulation of a stochastic process. Yao, Müller, Clifford, Dueker, Follett, Lin, Buchholz and Vogel (2003) used a smoothed version of the truncated

Karhunen–Loéve expansion to represent each sampled curve. The functional canonical correlation analysis was implemented by He, Müller and Wang (2002) and an application could be found in He, Müller and Wang (2004). Preda and Saporta (2005a) considered functional partial least squares and also gave an application in Preda and Saporta (2005b).

All schools focus on the adaptation of standard multivariate techniques to FDA. So far, principal component analysis (Silverman 1996), canonical correlation analysis (Leurgans, Moyeed and Silverman 1993 and He et al. 2002), and linear model (Cardot et al. 1999) have been successfully considered in functional data analysis context. In a nutshell, the approach of the English school is very effective in the practical data analysis, while the approach of the French school gives functional data analysis a sound theoretical basis. It will be beneficial to study FDA from all three perspectives.

2.1.4 A Trilogy in Modern Statistics

FDA, LDA and time series analysis (TSA, or more generally, the stochastic statistics, which includes spatial statistics) constitute a trilogy of modern statistics. In terms of data structure, time series data are collected usually as a long series of observations, longitudinal data are measured repeatedly over time giving rise to many short time series, while functional data have more general forms. Theoretically, LDA and TSA are in the field of parametric statistics, while the essence of functional data is infinite-dimensional and hence nonparametric statistics has a large role in FDA.

TSA focuses on modeling the dependence and prediction using mainly parametric approaches; an example is the classical autoregressive moving average (ARMA) model. In LDA, the challenge is the apparent nonstationarity of the repeated measurements for each subject. With the assumption of independence among the subjects, the inferences about the common covariance matrix can be achieved by borrowing strength across many subjects, which is the idea of pooled TSA.

A heuristic and insightful comparison on LDA and FDA from the smoothing perspective could be seen in Rice (2004). Many FDA studies are longitudinal-data driven. This is analogous to the relation between repeated measures model and LDA. The typical linear model in LDA could be viewed as a functional linear model with both functional response and covariates,

$$Y(t) = \beta(t)^T \mathbf{X}(t) + \varepsilon(t),$$

which has been studied extensively in FDA. Specifically, the mixed effects model from LDA is transplanted to functional linear model and combined with smoothing or basis splines from the FDA side, which generates a powerful tool for both LDA and FDA. Brumback and Rice (1998) and Rice and Wu (2001) made contribution in this direction, Chiou, Müller and Wang (2003) and Guo (2002) studied the functional mixed effect models in more depth. It becomes a trend that more FDA techniques are introduced to LDA, for example, functional principal component analysis has been applied to LDA (Besse, Cardot and Ferraty 1997, James, Hastie and Sugar 2000 and Yao, Müller and Wang 2005) and a functional multiplicative effects model was proposed to study longitudinal behavior in biology science (Chiou, Müller, Wang and Carey 2003). Between TSA and FDA, the works in Bosq (1991, 2000) introduced some advanced techniques in TSA to FDA. Laukaitis and Račkauskas (2002) illustrated an application of functional autoregressive model in financial time series data.

2.2 IR

IR has been an active research topic for about fifteen years since its introduction in Li (1991). Recently, functional data analysts started working on it. In this section, the research of both multivariate and functional IR will be reviewed.

2.2.1 Multivariate Multiple-Index Model and IR

The seminal paper Li (1991) proposed the following multiple-index model which is a regression-type semiparametric model,

$$y = f(\beta_1'\mathbf{x}, \beta_2'\mathbf{x}, \ldots, \beta_p'\mathbf{x}, \varepsilon)$$

where $\mathbf{x}$ and β_i's belong to $\mathbf{R}^d$, ε and $\mathbf{x}$ are independent of each other, $1 \leq p \leq d$ and $f : \mathbf{R}^{p+1} \mapsto \mathbf{R}$. Call each $\beta_i'x$ an index, β_i the index coefficient vector, and f the link function. The number of indices, coefficient vectors and the link function are all unknown.

One important implication of the model is that the projection of the d-dim explanatory variable onto the p-dim subspace

$$B = \text{span}(\beta_1'\mathbf{x}, \beta_2'\mathbf{x}, \ldots, \beta_p'\mathbf{x}),$$

captures all we need to know about y, or, in probability language, y and $\mathbf{x}$ are independent of each other given $(\beta_1'\mathbf{x}, \beta_2'\mathbf{x}, \ldots, \beta_p'\mathbf{x})$. The central goal of the model is to estimate the so-called effective dimension-reduction (EDR) space, $\text{span}(\beta_1, \beta_2, \ldots, \beta_p)$.

There are many papers on multiple-index type models. Härdle and Stocker (1989) investigated the method of average derivative estimate (ADE), which has been developed in a series of works; for example, Donkers and Schafgans (2003) used outer product of derivatives, and Hristache, Juditsky, Polzehl and Spokoiny (2001) proposed an iterative improvement. The single-index model ($p = 1$) attracted the most attention. Among many works Hristache, Juditsky and Spokoiny (2001) developed ADE for single-index model, Naik and Tsai (2000) studied the performance of partial least square method, and Yu and Ruppert (2002) used penalized splines for partially linear single-index model.

On the other hand, to extend the idea of EDR space, Cook (1998) defined B as the dimension-reduction subspace, and then further developed the concept of central subspace in theoretical and graphical tools.

To explore the geometrical structure of the multiple-index model, Li (1991) added a crucial condition,

$$\mathrm{E}(\mathbf{b}'\mathbf{x}|\beta_1'\mathbf{x},\beta_2'\mathbf{x},\ldots,\beta_p'\mathbf{x}) \in \mathrm{span}\{\beta_1'\mathbf{x},\beta_2'\mathbf{x},\ldots,\beta_p'\mathbf{x}\}, \qquad \text{for} \quad \forall \mathbf{b} \in \mathbf{R}^d.$$

See Hall and Li (1993) for more detailed discussion of the condition. For a model satisfying this condition, the space $\mathrm{span}(\Sigma\beta_1, \Sigma\beta_2, \Sigma\ldots, \beta_p)$ contains the centered IR curve $\mathrm{E}(\mathbf{x}|y) - \mathrm{E}(\mathbf{x})$, where $\Sigma = \mathrm{Cov(x)}$. This implies that the principal component analysis or eigen-decomposition of Δ can achieve the estimation of the EDR space. Then how to estimate Δ or $\Sigma - \Delta$ becomes the central problem, where $\Delta = \mathrm{Cov}[\mathrm{E}(\mathbf{x}|y)]$.

Li (1991) proposed the now well-known sliced IR (SIR) which can be proceeded by five steps based on data $(\mathbf{x}_i, y_i)$, $i = 1, \cdots, n$,

1. Center and standardize $\mathbf{x}$, $\mathbf{z}_i = \hat{\Sigma}_n^{-1/2}(\mathbf{x}_i - \bar{\mathbf{x}})$, $i = 1, \cdots, n$, where $\hat{\Sigma}_n$ is the estimate of Σ.

2. Divide the range of y into S slices, $I_1, \cdots, I_S$.

3. Estimate $E(\mathbf{z}|y)$ by sliced mean, $\bar{\mathbf{z}}_s = \frac{1}{n_s}\sum_{i=1}^{n} \mathbf{z}_i I(y_i \in I_s)$, $s = 1, \cdots, S$, where $n_s = \sum_{i=1}^{n} I(y_i \in I_s)$,

4. Estimate $Cov(E(\mathbf{z}|\mathbf{y}))$ by the weighted covariance matrix, $\hat{V} = \frac{1}{n}\sum_{s=1}^{S} n_s \bar{\mathbf{z}}_s \bar{\mathbf{z}}_s'$.

5. Implement the principal component analysis of $\hat{V}$.

Duan and Li (1991) and Li (1997) presented more delicate results for analyzing single-index regression by SIR, Hsing and Carroll (1992) and Zhu and Ng (1995) derived the large sample properties of SIR based on $\Sigma - \Delta$, Chen and Li (1998) illustrated the features of SIR. Schott (1994), Ferré(1997 and 1998) discussed the determination of the number of indices in SIR. Carroll and Li (1992) applied SIR to measurement error models, and Becker and Fried (2002) made a direct use of SIR in high-dimensional time series analysis.

For data analysis, He, Fang and Xu (2003) analyzed mass spectra data by combining SIR and classification tree, Gannoun, Girard, Guinot and Saracco (2004) combined SIR and a kernel estimation of conditional quantile to estimate reference curves in clinical studies.

For other IR approaches, Zhu and Fang (1996) proposed a kernel regression to estimate $E(\mathbf{x}|\mathbf{y})$ and also gave an asymptotic result. Gather, Hilker and Becker (2002) evaluated the sensitivity of SIR to outliers and Gather, Hilker and Becker (2001) provided a robust version of SIR. Fung, He, Liu and Shi (2002) implemented SIR by *B*-spline and canonical correlation to estimate Δ. Bura and Cook (2001) introduced a parametric IR which fitted $E(\mathbf{x}|y)$ via a multivariate linear regression. Naik and Tsai (2005) proposed the constrained IR in the presence of linear constraints on parameters. A more innovative idea based on SIR was in Xia, Tong, Li and Zhu (2002) which generated a minimum average conditional variance estimation inspired by the SIR, ADE and local linear smoothers.

A complementary method of SIR termed sliced average variance estimates (SAVE) was introduced in Cook and Weisberg (1991). Its asymptotic theory was provided in Gannoun and Saracco (2003) and an application to microarray data was reported in Bura and Pfeiffer (2003).

2.2.2 Functional Inverse Regression

In the functional SIR proposed in Ferré and Yao (2003) under setting of the French school, the model, assumption and procedures were all parallel to Li (1991) by translating the terms from the linear algebra to functional analysis. The model is

$$Y = f\left(\langle \beta_1, X\rangle_H, \langle \beta_2, X\rangle_H, \ldots, \langle \beta_p, X\rangle_H, \varepsilon\right),$$

where β's and X belong to $H = L^2([a,b])$ with $\mathrm{E}(\|X\|^2) < \infty$. With the assumption

$$\mathrm{E}(\langle \beta, X\rangle_H | \langle \beta_1, X\rangle_H, \ldots, \langle \beta_p, X\rangle_H \in \mathrm{span}\{\langle \beta_1, X\rangle_H, \ldots, \langle \beta_p, X\rangle_H\} \qquad \text{for} \quad \forall \beta \in \mathrm{H},$$

the spectral decomposition of $\Gamma_X^{-1}\Gamma_{\mathrm{E}(X|Y)}$ can estimate $\mathrm{span}\{\beta_1,\ldots,\beta_p\}$, where Γ_Z is the covariance operator of a stochastic process $Z \in \mathcal{H}$. The SIR procedure can be implemented in a way that is similar to the multivariate case to estimate $\Gamma_{\mathrm{E}(X|Y)}$. However, Γ_X^{-1} is problematic due to its unboundedness. The projection technique in Bosq (1991) was then applied here.

Another functional IR method under the same setting as above was considered by Amato, Antoniadis and Feis (2004), in which $\mathrm{E}(X|Y)$ was estimated by wavelet smoothing.

Both Li, Aragon, Shedden and Agnan (2003) and Setodji and Cook (2004) applied IR to the functional response and multivariate input model:

$$y(t) = g(\beta_1(t)'\mathbf{x},\ldots,\beta_p(t)'\mathbf{x}) + \varepsilon(t).$$

The former used basis presentation and the latter used *k*-means approach for slicing. These two papers originally intended to extend the univariate IR to multivariate response data, which had been done in Hsing (1999) by nearest neighbor IR.

2.3 RKHS with Applications in Probability and Statistics

RKHS methods have been employed by probabilists and statisticians for at least fifty-five years. With the applications done by of the machine learning community, it has become an active topic of research in statistics. Recent works also bring strong evidences of its promising role in FDA.

The theory of RKHS was originated from complex analysis, developed in integral equations and bounded-value problems, and matured into the present form in the landmark paper Aronszajn (1950). It was introduced into the probability world by Loéve (1948), which built the famous Loéve's isometry. Parzen (1959) introduced RKHS to statisticians; Parzen (1961a, 1961b) solved several crucial problems in signal analysis using this powerful tool, providing a convincing evidence of the relevance of RKHS in time series analysis

and general stochastic inference. See Weinert (1982) for a collection of these papers. Barton and Poor (1990) and Nuzman and Poor (2001) gave two applications to robust signal analysis and self-similar processes, respectively. Wahba's well-known work in the 1970's formulated the mathematical foundation of the nonparametric smoothing with spline functions using RKHS (Wahba 1990). Gu (2002) developed tensor product smoothing splines by further application of RKHS. Vapnik's statistical learning theory (Vapnik 1995) including the support vector machine revitalized the interest of RKHS in nonparametric statistics; the kernel-based algorithms are among the most exciting research topics in both signal analysis and statistics today. For other application of RKHS, see Berlinet and Thomas–Agnan (2004).

In FDA context, the use of RKHS can be seen throughout Ramsay and Dalzell (1991). The function space H is partitioned into a direct sum of subspaces H_1 and H_2 for two linear operators, L and B, such that $H_1 = \ker(L)$, $H_2 = \ker(B)$ and $H = H_1 \oplus H_2$, where H_1 contains the structural components and is usually of finite dimension, H_2 contains the residual components and is usually infinite dimensional. Based on this partitioning, further analyses including the representation of the discrete observation by a function in H was accomplished by RKHS or Wahba's spline theory.

Eubank and Hsing (2005) brought RKHS to FDA under the setting of the stochastic school. By Loéve's isometry and tensor product of RKHS, both the concept and the computations of canonical correlations were extended to general stochastic processes.

In the remainder of the dissertation, we will present a new functional multiple-index model and propose inverse regression approaches from the RKHS perspective. This methodology is closer in spirit to the practice of the stochastic school, and at the same time it uses the functional analysis language from the French school, and incorporates smoothing techniques from the English school.

CHAPTER III

A FUNCTIONAL MULTIPLE-INDEX MODEL

In this chapter, a multiple-index model related to a second order stochastic process will be proposed in the first section. To explore the geometrical structure of the model, we will give some basic facts of the reproducing kernel Hilbert space (RKHS) in Section 3.2. Then Section 3.3 provides the relationship between the RKHS and inverse regression.

3.1 A Second Order Multiple-Index Model

Let $\{X_t, t \in T\}$ be a real-valued zero-mean, second order stochastic process defined on some probability space $(\Omega, \mathcal{F}, P)$, where the index set T is assumed to be a separable metric space. Here and elsewhere, a second-order process or L^2 process refers to a process such that $\mathrm{E}(|X_t|^2) < \infty$ for all $t \in T$. See Ash and Gardner (1975) for background knowledge on second-order stochastic processes.

Define the Hilbert space L_X^2 generated by the process in the following way. First let $\mathrm{span}\{X_t, t \in T\}$, be the set of finite linear combinations of random variables of the form X_{t_i} for $t_i \in T$. Further let $\overline{\mathrm{span}}\{X_t, t \in T\}$ be the closure of $\mathrm{span}\{X_t, t \in T\}$ in $L^2(\Omega, \mathcal{F}, P)$. Then L_X^2 is defined to be the space containing elements in $\overline{\mathrm{span}}\{X_t, t \in T\}$ and equipped with the inner product

$$\langle U, V \rangle_{L_X^2} = \mathrm{E}(UV).$$

Thus, L_X^2 contains the random variables attainable by linear operations on X_t and their L^2 limits.

Let us now define the following conditions in which $\xi_1, \ldots, \xi_p$ are elements in L_X^2 and Y is a real random variable.

(IR1) Y and X are conditionally independent of each other given $\xi_1, \ldots, \xi_p$.

(IR2) For any $\xi \in L_X^2$,

$$\mathrm{E}(\xi|\xi_1, \ldots, \xi_p) \in \mathrm{span}\{\xi_1, \ldots, \xi_p\} \quad \text{a.s..}$$

A particularly relevant situation for which (IR1) holds is the multiple-index model

$$Y = f(\xi_1, \ldots, \xi_p, \varepsilon), \tag{3.1}$$

where ε is a random error independent of the process $\{X_t\}$, each $\xi_i' x$ is called an index and f the link function. The number of indices, indices and the link function are all unknown.

Condition (IR2) holds if the joint distribution of any finite collection of elements from L_X^2 is spherically symmetric, which would be the case if, for instance, $\{X_t\}$ is a Gaussian process.

The goal of the model is to estimate the so-called effective dimension-reduction (EDR) space, a subspace of L_X^2,

$$L_{X,e}^2 = \mathrm{span}\{\xi_1, \ldots, \xi_p\} \subset L_X^2,$$

which is equipped with the inner products of L_X^2.

So far, the set-up is analogous to Li (1991), however, we need more advanced theory for the RKHS to explore the structure of this model. Before doing so, we will briefly review the notion of RKHS.

3.2 Some Facts of RKHS

The general definition and common properties of an RKHS will be presented in this section, which could also be found in Aronszajn (1950), Pazern (1959), Kailath (1971), Wahba (1990), Lukić and Beden (2001) (henceforth LB).

Definition III.1. A Hilbert space H is said to be a RKHS with reproducing kernel K, if each element of H is a function defined on some set T, and there is a bivariate function K on $T \times T$, having the following two properties:

(1) For all $t \in T$, $K(\cdot,t) \in H$

(2) For all $t \in T$ and $f \in H$, $f(t) = \langle f, K(\cdot,t) \rangle_H$

There are three components of a RKHS: the index set T, the Hilbert space H of functions, i.e. $H \subset \mathbf{R}^T$, and the kernel function with the specific reproducing property in (2). We use the triple notation (T,H,K) or $H(K,T)$ to denote a RKHS.

There are many nice properties of RKHS, mostly due to the existence of kernel. The first two properties are related to the alternative characterization and denseness of RKHS.

Property III.1. (Alternative Characterization)

In the RKHS, $H(K,T)$, every evaluation functional is continuous, which means that for all $t \in T$, the functional e_t with $e_t(f) = f(t)$ for all $f \in H(K,T)$, satisfies $|e_t(f)| \leq M_t \| f \|_H$ for some $M_t \in [0,\infty)$.

Property III.2. (Denseness)

$$H(K,T) = \overline{\text{span}}\{K(\cdot,t), t \in T\}.$$

The following properties are related to the continuity and smoothness of the function.

Property III.3. (Pointwise Continuity)

If $\{ f_n, n \in \mathbf{N}, f \} \subset H(K,T)$, then

$$\| f_n - f \|_H \xrightarrow[n\to\infty]{} 0 \Longrightarrow f_n \to f \quad \text{for each point of } T, \text{ as } n \to \infty.$$

$$\langle f_n, a \rangle_H \xrightarrow[n\to\infty]{} \langle f, a \rangle_H, \quad \forall a \in H \Longrightarrow f_n \to f \quad \text{for each point of } T, \text{ as } n \to \infty.$$

Property III.4. (Smoothness)

If K is a continuous bivariate function on $T \times T$, then any $f \in H(K,T)$ is continuous on T.

The following properties are related to the kernel.

Property III.5. (Uniqueness)

If K is the reproducing kernel of a RKHS, then K is non-negative definite and unique. Conversely, if K is a non-negative definite bivariate function on $T \times T$, a unique RKHS of real-valued function on T with K as its reproducing kernel can be constructed.

Definition III.2. (Non-negative and Positive Definite Functions)

A symmetric real-valued bivariate function K defined on T is said to be non-negative definite if for $\forall n \in \mathbf{N}$, $\{a_1,\ldots,a_n\} \subset \mathbf{R}$, $\{t_1,\ldots,t_n\} \in T$,

$$\sum_{i,j=1}^{n} a_i a_j K(t_i,t_j) \geq 0,$$

and positive definite if the equality holds only when $a_1 = a_2 = \ldots = a_n = 0$. We shall use $K \geq 0$ and $K > 0$ to denote that K is non-negative and positive definite, respectively.

Property III.6. (Sum of Reproducing Kernels)

The direct sum of two spaces (T,H_1,K_1) and (T,H_2,K_2) is also a RKHS. Denote it as (T,H,K), where $H = \{f = f_1 + f_2, f_i \in H_i, i = 1,2\}$ and $K = K_1 + K_2$ with norm defined by

$$\forall f \in H, \|f\|_H^2 = \min_{\substack{f=f_1+f_2 \\ f_1 \in H_1, f_2 \in H_2}} (\|f_1\|_{H_1}^2 + \|f_2\|_{H_2}^2).$$

Property III.7. (Difference of Reproducing Kernels)

For two spaces (T,H_1,K_1) and (T,H_2,K_2), assume that $K_2 - K_1 \geq 0$, then $H_2 \supset H_1$ and there exists a unique linear operator $L : H_2 \mapsto H_1$ such that for $\forall f \in H_2$ and $g \in H_1$,

$$\langle f,g \rangle_{H_2} = \langle Lf,g \rangle_{H_1},$$

L satisfies

$$LK_2(\cdot,t) = K_1(\cdot,t), \qquad \forall t \in T,$$

and L is a bounded self-adjoint and positive operator.

Definition III.3. (Dominance and nuclear dominance) Under the assumption of Property III.7, we say that K_2 dominates K_1 if $K_2 - K_1 \geq 0$, and denote it by $K_2 \geq K_1$. The operator L is called the dominance operator. If L is nuclear, i.e., it is trace-class operator, we say K_2 n-dominates K_1, denote this by $K_2 \gg K_1$ and L is called the nuclear dominance.

The following properties are related to the index set T.

Property III.8.

$$f \in H(K,T) \iff \sup_S \sup_{a_i} \frac{|\sum_i a_i f(t_i)|^2}{\sum_i \sum_j a_i a_j K(t_i,t_j)} < \infty, \tag{3.2}$$

where the suprema are taken over all $S = \{t_1,\ldots,t_n\} \in \mathcal{F} = \{S \subset T : S \text{ is finite}\}$ and all real $a_1,\ldots,a_n$, with n arbitrary, such that the denominator in (3.2) is not zero.

Property III.9. Let T be finite, and let the matrix determined by kernel K is nonsingular. Then $\forall f,g \in H(K,T)$:

$$\begin{aligned} \|f\|_K^2 &= \sum_{t,s\in T} f(t) f(s) K^{-1}(t,s), \\ \langle f,g\rangle_K &= \sum_{t,s\in T} f(t) g(s) K^{-1}(t,s), \end{aligned}$$

where K^{-1} is the inverse of the matrix K.

Let T be an index set and $T_1 \subset T$. For any f defined on T, let $f|_{T_1}$ stand for the restriction of f to the subset of T_1.

Property III.10. (Restriction of Index Set)

Suppose $T_1 \subset T$, let $H_1 = \{f_1 = f|_{T_1} : f \in H\}$ and $K_1 = K|_{T_1\times T_1}$, then $H_1 = H(K_1,T_1)$ is a RKHS and

$$\|f_1\|_{H_1} = \min_{\substack{f\in H \\ f|_{T_1}=f_1}} \|f\|_H, \text{ for all } f_1 \in H_1.$$

Property III.11. (Approximation) Let T be either a countable or separable metric space, $H(K,T)$ be a RKHS with $K>0$ and $\{T_n, n\in\mathbf{N}\}$ be a sequence of subsets of T that is monotone increasing and $\overline{\cup_{i=1}^{\infty}T_n}=T$, we can define a sequence of spaces $H_n=H(K_n,T_n)$ with reproducing kernels $K_n=K|_{T_n\times T_n}$, then

(1) For any function f defined on T,

$$\|f_n\|_{H_n}\le\|f_{n+1}\|_{H_{n+1}},$$

where $f_n=f|_{T_n}$;

(2) For any $f\in H(K,T)$,

$$\|f\|_H=\lim_{n\to\infty}\|f_n\|_{H_n};$$

(3) For any f which is continuous on T, then

$$\lim_{n\to\infty}\|f_n\|_{H_n}<\infty\Longrightarrow f\in H(K,T).$$

Property III.12. (Separability)

A topological structure of the index set could be induced by the kernel. For the RKHS (T,H,K), denote $K_s=K(\cdot,s)$, then $d_K(s,t)=\|K_s-K_t\|$ defines a pseudo-metric on T, and we have

(1) $K>0$ implies that d_K is a metric on T.

(2) If d_K is a metric on T, then

(a) $\forall f\in H(K,T)$ is d_K-continuous.

(b) (T,d_K) is separable iff $H(K,T)$ is separable.

The last result concerns with the situation where both kernel and index set are varying. The following definition and property are excerpted form LB.

Definition III.4. (Hamel Basis and Hamel Set)

A set v_α of linearly independent vectors in a vector space V is called a Hamel basis in V if $\{v_\alpha\}$ spans V. Let V be the vector space spanned by $\{K_t, t \in T\}$, a set $T_0 \subseteq T$ such that $\{K_t, t \in T_0\}$ is a Hamel basis of V will be called an K-Hamel subset of T.

Property III.13. Let $H(R,T)$ be separable and let the kernel K be such that $R \gg K$ with dominance operator L. Let T_0 be an R-Hamel subset of T and $S_0 = \{s_1, \ldots\}$ be a d_K-dense subset of T_0. Denote by K_n and R_n the matrices obtained by restricting the kernels K and R to the set $\{s_1, \ldots, s_n\} \subset S_0$. Then

$$\mathrm{Tr}(L) = \mathrm{Tr}(K_n R_n^{-1}),$$

where Tr(L) is the trace of operator of L.

The proofs of above properties could be found in Aronszajn (1950), Pazern (1959) and LB.

3.3 Probabilistic Substructure—Between IR and RKHS

The method of embedding of an abstract space into some RKHS is used extensively to stochastic processes. The central result is Loéve's isometry. Consider a real-valued zero-mean L^2 process $\{X_t, t \in T\}$ with

$$R(s,t) = \mathrm{E}(X_s X_t) \quad \text{and} \quad R_t(\cdot) = R(t,\cdot),\ s,t \in T.$$

Then we can define the RKHS generated by $\{X_t, t \in T\}$, which is just the RKHS with reproducing kernel R:

$$\mathcal{H}_X = \overline{\mathrm{span}}\{R_t, t \in T\} = \text{the RKHS of } \{X_t, t \in T\} = H(R,T).$$

The inner product of $H(R,T)$ is

$$\langle f, g \rangle_{\mathcal{H}_X} = \sum_{i,j=1}^{\infty} a_i b_j R(s_i, t_j), \text{ for } f = \sum_{i=1}^{\infty} a_i R_{s_i} \text{ and } g = \sum_{j=1}^{\infty} b_j R_{t_j} \in \mathcal{H}_X,$$

which, by the reproducing property, satisfies

$$\langle f, R_t \rangle_{\mathcal{H}_X} = f(t),\ t \in T.$$

Comparing this to the Hilbert space generated by the process, L_X^2, defined in Section 3.1, we see that L_X^2 is isometrically isomorphic to $\mathcal{H}_X$, that is, there exists a one-to-one correspondence between the two spaces that preserves the inner products of the two spaces. Let Ψ be the corresponding isometry that maps L_X^2 to $\mathcal{H}_X$ with

$$\Psi(X_t) = R_t,\ t \in T.$$

It follows that $\Psi(\eta)(t) = E(\eta X_t)$, for $\eta \in L_X^2$ and $t \in T$. This mapping is called Loéve's isometry, which provides a duality between a stochastic process and its RKHS (Wahba 1990).

By this mapping, define the counterpart of EDRS in $\mathcal{H}_X$, a subspace of $\mathcal{H}_X$,

$$\mathcal{H}_{X,e} = \Psi_X(L_{X,e}^2) = \text{span}\{\Psi(\xi_1), \ldots, \Psi(\xi_p)\} \subset \mathcal{H}_X,$$

which is equipped with the inner products of $\mathcal{H}_X$ and we simply call it the EDRS in RKHS.

Motivated by the main theorem in Li (1991), we propose the following conjecture:

The sample paths of conditional process, $\mathrm{E}(X|Y) \in \mathcal{H}_{X,e}$ a.s.

At first glance, this may be proved using the same type of arguments as those in Li (1991) or Ferré and Yao (2003). However, we do not believe that this is the case. It turns that a more thorough study of the relationship between a RKHS and the sample paths of a stochastic process is required in the proof.

Parzen (1963) observed that almost all the sample paths of X lie outside $\mathcal{H}_X$ if T is an infinite separable metric space and the R is continuous on $T \times T$. Driscoll (1973) was the first paper to investigate the RKHS structure of the sample paths of a Gaussian process and gave sufficient conditions for the sample paths falling into a RKHS. Nearly

thirty years later, LB summarized and generalized this category of problems. The following development is inspired by their results.

Defining the following conditions,

(P1) Let R be a continuous positive kernel on $T \times T$;

(P2) The sample paths of $\mathrm{E}(X|Y)$ are continuous with probability one.

we have,

Theorem III.1. Assume that (P1), (P2), (IR1), and (IR2) hold. Then

$$\mathrm{E}(X|Y) \in \mathcal{H}_{X,e} \text{ a.s.}$$

Proof. To prove the theorem, a lemma is first stated.

Lemma III.1. Let S be a separable metric space. Let $\{U_s, s \in S\}$ be a L_2-process on probability space $(\Omega, \mathcal{F}, P)$ with mean function u and continuous covariance kernel K_1. Assume that almost all the sample paths of U are continuous on S. Let K_2 be a continuous positive kernel on $S \times S$ such that with $K_2 \gg K_1$ and $u \in H(K_2)$. Then $P[U \in H(K_2)] = 1$.

PROOF OF LEMMA. Denote the metric of S by d and let S_0 be a countable dense subset of S in d. Define

$$d_{K_2}(s,t) = \|K_2(s,\cdot) - K_2(t,\cdot)\|_{K_2}.$$

Since $K_2 > 0$, it follows from Property III.12(1) that d_{K_2} is a metric on S. For any $s \in S$, let s_n be a sequence of elements in S_0 which converges to s in S in the metric d. Then, by the reproducing property,

$$d^2_{K_2}(s_n,s) = \|K_2(s_n,\cdot) - K_2(s,\cdot)\|^2_{K_2} = K_2(s_n,s_n) - 2K_2(s_n,s) + K_s(s,s) \to 0$$

by the continuity of K_2. This shows that S_0 is also dense in the metric d_{K_2} so that S is also d_{K_2}-separable. Thus, by Property III.12(2b), $\mathcal{H}(K_2)$ is separable.

Now enumerate the elements of S_0 and let S_n be the collections of the first n elements according to the enumeration. Then S_n is monotone increasing and $\lim_{n\to\infty} S_n = S_0$. Define $f = U - u$, $K_{1,n} = K_1|_{S_n\times S_n}$, $K_{2,n} = K_2|_{S_n\times S_n}$, $f_n = f|_{S_n}$, and $U_n = U|_{S_n}$. Observe that

$$\begin{aligned}\mathrm{E}[\|f_n\|^2_{K_{2,n}}] &= \mathrm{E}(f_n' K_{2,n}^{-1} f_n) = \mathrm{E}(\mathrm{tr}(f_n' K_{2,n}^{-1} f_n)) = \mathrm{E}(\mathrm{tr}(f_n f_n' K_{2,n}^{-1})) \\ &= \mathrm{tr}[\mathrm{E}(f_n f_n') K_{2,n}^{-1}] = \mathrm{tr}[\mathrm{cov}(f_n) K_{2,n}^{-1}] = \mathrm{tr}(K_{1,n} K_{2,n}^{-1}).\end{aligned}$$

Since $\|f_n\|_{K_{2,n}}$ is monotone by (1) of Property III.11, it follows from the monotone convergence theorem that

$$\mathrm{E}[\lim_{n\to\infty} \|f_n\|^2_{K_{2,n}}] = \lim_{n\to\infty} \mathrm{tr}(K_{1,n} K_{2,n}^{-1}). \tag{3.3}$$

Note that since K_2 is nonsingular, T itself is an K_2-Hamel subset of T. Let L be the dominance operator for $\mathcal{H}(K_2)$ over $\mathcal{H}(K_1)$. By Property III.13 and the assumption $K_2 \gg K_1$, we have

$$\lim_{n\to\infty} \mathrm{tr}(K_{1,n} K_{2,n}^{-1}) = \mathrm{tr}(L) < \infty.$$

It then follows from (3.3) that

$$\lim_{n\to\infty} \|f_n\|^2_{K_{2,n}} < \infty \quad \text{a.s.}$$

By (3) of Property III.11, this implies that $f \in \mathcal{H}(K_2)$ a.s., and completes the proof.

The proof of III.1 is accomplished in five steps as

1. Let $\mathcal{H}_{\mathrm{E}(X|Y)}$ denote the RKHS of the process $\{\mathrm{E}(X_t|Y), t \in T\}$, which is well-defined.

   Since

   $$\mathrm{E}[\mathrm{E}(X_t|Y)] = \mathrm{E}(X_t) = 0,$$

   and

   $$\mathrm{Var}(\mathrm{E}(X_t|Y)) \le \mathrm{Var}(X_t) < \infty,\ t \in T,$$

it follows that $\{\mathrm{E}(X(t)|Y), t \in T\}$ is also a zero-mean, second order stochastic process, with covariance function

$$K(s,t) := \mathrm{Cov}\left(\mathrm{E}(X_s|Y), \mathrm{E}(X_t|Y)\right),\ s,t \in T.$$

Thus, define $L^2_{\mathrm{E}(X|Y)}$ and the RKHS $\mathcal{H}_{\mathrm{E}(X|Y)}$ in the usual way.

2. Verify that $\dim(\mathcal{H}_{\mathrm{E}(X|Y)}) \leq p$.

By definition

$$L^2_{\mathrm{E}(X|Y)} = \overline{\mathrm{span}}\{\mathrm{E}(X_t|Y), t \in T\},$$

and by (IR1) and (IR2),

$$\begin{aligned} \mathrm{E}(X_t|Y) &= \mathrm{E}(\mathrm{E}(X_t|Y, \xi_1, \ldots, \xi_p)|Y) \\ &= \mathrm{E}(\mathrm{E}(X_t|\xi_1, \ldots, \xi_p)|Y) \\ &= \sum_{i=1}^{p} c_{i,t}\mathrm{E}(\xi_i|Y) \quad \text{a.s.} \end{aligned}$$

for some constants $c_{i,t}$. It follows that

$$\mathrm{E}\left(X_t|Y\right) \in \mathrm{span}\{\mathrm{E}(\xi_i|Y),\ i = 1, \ldots, p\}.$$

Consequently,

$$L^2_{\mathrm{E}(X|Y)} \subseteq \mathrm{span}\{\mathrm{E}(\xi_i|Y),\ i = 1, \ldots, p\},$$

and hence

$$\dim(L^2_{\mathrm{E}(X|Y)}) = \dim(\mathcal{H}_{\mathrm{E}(X|Y)}) \leq p.$$

3. Verify that there exists a dominance operator of $\mathcal{H}_X$ over $\mathcal{H}_{\mathrm{E}(X|Y)}$.

Let $\mathbf{t} = (t_1, \ldots, t_m)' \in T^m$ and $\mathbf{a} = (a_1, \ldots, a_m) \in \mathbf{R}^m, m = 1, 2, \ldots$. Writing $\mathbf{X} = (X_{t_1}, \ldots, X_{t_m})'$, we have

$$\operatorname{var}(\mathbf{a}'\mathbf{X}) = \operatorname{var}(E(\mathbf{a}'\mathbf{X}|Y)) + E(\operatorname{var}(\mathbf{a}'\mathbf{X}|Y)).$$

Thus,

$$\mathbf{a}'(R_m - K_m)\mathbf{a} = E(\operatorname{var}(\mathbf{a}'\mathbf{X}|Y)) \geq 0,$$

where

$$R_m = \{R(t_i, t_j)\}_{i,j=\overline{1,m}} \text{ and } K_m = \{K(t_i, t_j)\}_{i,j=\overline{1,m}}.$$

This implies that $R - K \geq 0$, and we conclude that

$$\mathcal{H}_X \supseteq \mathcal{H}_{\mathrm{E}(X|Y)}.$$

Further, by Definition III.3, there exists a dominance operator

$$L : \mathcal{H}_X \to \mathcal{H}_{\mathrm{E}(X|Y)}$$

such that

$$\langle f, g \rangle_{\mathcal{H}_X} = \langle Lf, g \rangle_{\mathcal{H}_{\mathrm{E}(X|Y)}} \quad \text{for all } f \in \mathcal{H}_X \text{ and } g \in \mathcal{H}_{\mathrm{E}(X|Y)}.$$

4. Verify that $\mathrm{E}(X|Y) \in \mathcal{H}_X$ a.s.

Combining steps 2 and 3, we can conclude that the dominance operator L is a finite rank operator, hence L is a nuclear operator with $\operatorname{tr}(L) < \infty$. Since T is separable, and the continuity of R implies the L^2-continuity of $\{X_t, t \in T\}$, by the L^2 convergence property of conditional expectation, $\{\mathrm{E}(X(t)|Y), t \in T\}$ is also L^2-continuous and so K is continuous. It then follows from Lemma III.1 that

$$\mathrm{E}(X|Y) \in \mathcal{H}_X.$$

5. Finally, prove that $\mathrm{E}(X|Y) \in \mathcal{H}_{X,e}$ a.s..

We will show that

$$\langle \mathrm{E}(X|Y), h \rangle_{\mathcal{H}_X} = 0$$

for any $h \in \mathcal{H}_X$ such that

$$\langle h, \Psi(\xi_i) \rangle_{\mathcal{H}_X} = 0,\ 1 \le i \le p. \tag{3.4}$$

Let $\xi = \Psi^{-1}(h) \in L^2_X$. If $h = R_t$, then $\Psi^{-1}(h) = X_t$. By the reproducing kernel property,

$$\langle \mathrm{E}(X|Y), h \rangle_{\mathcal{H}_X} = \mathrm{E}(X|Y)(t) = \mathrm{E}(X_t|Y) = \mathrm{E}(\xi|Y).$$

In general if $h = \sum_{i=1}^{\infty} d_i R_{t_i}$, then $\xi = \sum_{i=1}^{\infty} d_i X_{t_i}$ and

$$\langle \mathrm{E}(X|Y), h \rangle_{\mathcal{H}_X} = \sum_{i=1}^{\infty} d_i \langle \mathrm{E}(X|Y), R_{t_i} \rangle_{\mathcal{H}_X} = \sum_{i=1}^{\infty} d_i \mathrm{E}(X_{t_i}|Y) = \mathrm{E}(\xi|Y).$$

By the properties of conditional expectation and (IR1),

$$\mathrm{E}(\xi|Y) = \mathrm{E}\left(\mathrm{E}(\xi|\xi_1, \ldots, \xi_p, Y)|Y\right) = \mathrm{E}\left(\mathrm{E}(\xi|\xi_1, \ldots, \xi_p)|Y\right).$$

It suffices to show that the above righthand side equals 0, which we now do. Since by (IR2),

$$\mathrm{E}(\xi|\xi_1, \ldots, \xi_p) = \sum_{i=1}^{p} c_i \xi_i$$

for some $c_i, 1 \le i \le p$, we have

$$\begin{aligned} \mathrm{E}\left(E^2(\xi|\xi_1, \ldots, \xi_p)\right) &= \mathrm{E}\left(\sum_{i=1}^{p} c_i \xi_i \mathrm{E}(\xi|\xi_1, \ldots, \xi_p)\right) \\ &= \mathrm{E}\left(\sum_{i=1}^{p} c_i E(\xi\xi_i|\xi_1, \ldots, \xi_p)\right) = \sum_{i=1}^{p} c_i \mathrm{E}(\xi\xi_i), \end{aligned}$$

which is equal to

$$\sum_{i=1}^{p} c_i \langle \Psi(\xi_i), \Psi(\xi) \rangle_{\mathcal{H}_X} = \sum_{i=1}^{p} c_i \langle \Psi(\xi_i), h \rangle_{\mathcal{H}_X} = 0$$

by (3.4). Then

$$\mathrm{E}(\xi | \xi_1, \ldots, \xi_p) = 0$$

which implies

$$\mathrm{E}\left(\mathrm{E}(\xi | \xi_1, \ldots, \xi_p) | Y\right) = 0.$$

Hence the proof is complete.

♢

To apply the theorem to estimate the EDRS in RKHS, $\mathcal{H}_{X,e}$, we need to derive some corollaries from the theorem. By Theorem 3.1 of LB, since the sample paths of $\mathrm{E}(X|Y)$ belong to $\mathcal{H}_X$, then $R \geq K$ implies that the covariance operator of $\mathrm{E}(X|Y)$ is well-defined and just equals to the dominance operator L, i.e.,

$$L = \mathrm{E}\left(\mathrm{E}(X|Y) \otimes_{\mathcal{H}_X} \mathrm{E}(X|Y)\right),$$

which is defined by

$$Lf = \mathrm{E}(\langle \mathrm{E}(X|Y), f \rangle_{\mathcal{H}_X} \mathrm{E}(X|Y)), \quad \text{for } f \in \mathcal{H}_X.$$

Since $R \gg K$, L is a nuclear operator meaning a trace-class, symmetric and a non-negative operator. There are more properties of L:

Corollary III.1. L is degenerate in any direction orthogonal to $\mathcal{H}_{X,e}$.

Proof. It follows from the theorem that

$$\langle s, \mathrm{E}(X|Y) \rangle_{\mathcal{H}_X} = 0$$

for all $s \in \mathcal{H}_X$ such that

$$\langle s, \Psi_X(\xi_i)\rangle_{\mathcal{H}_X} = 0, 1 \leq i \leq p.$$

Hence

$$\begin{aligned} 0 &= \mathrm{E}(\langle s, \mathrm{E}(X|Y)\rangle^2_{\mathcal{H}_X}) \\ &= \mathrm{E}(\langle s, \mathrm{E}(X|Y)\rangle_{\mathcal{H}_X} \langle s, \mathrm{E}(X|Y)\rangle_{\mathcal{H}_X}) \\ &= \mathrm{E}(\langle s, \langle s, \mathrm{E}(X|Y)\rangle_{\mathcal{H}_X} \mathrm{E}(X|Y)\rangle_{\mathcal{H}_X}) \\ &= \langle s, E(\langle s, \mathrm{E}(X|Y)\rangle_{\mathcal{H}_X} \mathrm{E}(X|Y))\rangle_{\mathcal{H}_X} \\ &= \langle s, Ls\rangle_{\mathcal{H}_X}. \end{aligned}$$

♢

Corollary III.2. For the range of L, we have

$$Im(L) \subset \mathcal{H}_{X,e},$$

where $Im(L)$ is the range of L.

Proof. For any $f \in \mathcal{H}_X$ and $h \perp \mathcal{H}_{X,e}$,

$$\begin{aligned} \langle Lf, h\rangle_{\mathcal{H}_X} &= \mathrm{E}\left(\langle \mathrm{E}(X|Y), f\rangle_{\mathcal{H}_X} \langle \mathrm{E}(X|Y), h\rangle_{\mathcal{H}_X}\right) \\ &= \mathrm{E}\left(\langle \mathrm{E}(X|Y), f\rangle_{\mathcal{H}_X} \cdot 0\right) \\ &= 0. \end{aligned}$$

♢

In order to introduce functional sliced inverse regression in $\mathcal{H}_X$, we need another operator,

$$\tilde{L} = \sum_{s=1}^{S} p_s E(X|Y \in I_s) \otimes_{\mathcal{H}_X} E(X|Y \in I_s),$$

where $Im(Y) = \oplus_{s=1}^{S} I_s$ and $p_s = Pr(Y \in I_s), s = 1, \cdots, S$.

Corollary III.3. For the sliced conditional sample path, if the sample paths of $E(X|Y \in I_s)$ are continuous with probability one, we have

$$E(X|Y \in I_s) \in \mathcal{H}_X, \quad \text{a.s.}$$

Proof. Let K and J be the kernels of $E(X|Y)$ and $E(X|Y \in I_s)$, respectively. Note that $E(X|Y \in I_s) = E[E(X|Y)|Y \in I_s]$. Then $K - J \geq 0$, which implies $H(K) \supset H(J)$. Since $dim(H(J)) \leq dim(H(K)) \leq p$, the dominance mapping from $H(K)$ to $H(J)$ is nuclear, by Lemma III.1,

$$E(X|Y \in I_s) \in H(K) \subset \mathcal{H}_X.$$

$\diamondsuit$

Corollary III.4. $\tilde{L}$ is degenerated in any direction orthogonal to $\mathcal{H}_{X,e}$.

Proof. Since

$$\langle s, \mathrm{E}(X|Y) \rangle_{\mathcal{H}_X} = 0$$

for all $s \in \mathcal{H}_X$ such that

$$\langle s, \Psi_X(\xi_i) \rangle_{\mathcal{H}_X} = 0, 1 \leq i \leq p.$$

$$\begin{aligned}
\langle \tilde{L}s, s \rangle_{\mathcal{H}_X} &= \sum_{s=1}^{S} p_s \langle E(X|Y \in I_s), s \rangle^2_{\mathcal{H}_X} \\
&= \sum_{s=1}^{S} p_s \langle E[E(X|Y)|Y \in I_s], s \rangle^2_{\mathcal{H}_X} \\
&= \sum_{s=1}^{S} p_s E^2(\langle E(X|Y), s \rangle_{\mathcal{H}_X} | Y \in I_s) \\
&= 0.
\end{aligned}$$

$\diamondsuit$

Corollary III.5. For the range of $\tilde{L}$,

$$Im(\tilde{L}) \subset \mathcal{H}_{X,e}.$$

Proof. For any $f \in \mathcal{H}_X$ and $h \perp \mathcal{H}_{X,e}$,

$$\begin{aligned}
\langle \tilde{L} f, h \rangle_{\mathcal{H}_X} &= \sum_{s=1}^{S} p_s \langle \mathrm{E}(X|Y \in I_s), f \rangle_{\mathcal{H}_X} \langle E(X|Y \in I_s), h \rangle_{\mathcal{H}_X} \\
&= \sum_{s=1}^{S} p_s \langle \mathrm{E}(X|Y \in I_s), f \rangle_{\mathcal{H}_X} E(\langle E(X|Y), h \rangle_{\mathcal{H}_X} | Y \in I_s) \\
&= 0.
\end{aligned}$$

♢

CHAPTER IV

STATISTICAL FRAMEWORK IN RKHS

Theorem III.1 reveals the geometrical or probabilistic structure of the functional multiple-index model and facilitates the statistical inference based on inverse regression in the RKHS of the process. We will give the matrix version and computation related to the covariance operator in Section 4.1, two estimation procedures in Section 4.2, and the asymptotic results in Section 4.3.

4.1 Discretization

There are two reasons to consider discretization. Firstly, the covariance operators defined in most Hilbert spaces cannot be calculated directly. Secondly, in functional data analysis it is always the case that $\{X(t), t \in T\}$ is observed only on a discrete set of values of T, say $T_q = \{t_1, \ldots, t_q\}$. To apply the matrix language, we simply consider the rectangle design, which means observing $\mathbf{X} = (X(t_1), \ldots, X(t_q))'$ where $t_1 < \ldots < t_q$, and $|t_{i+1} - t_i|$, $i = 1, \ldots, q-1$, might be unequal but the same for all curves. This rectangle sampling scheme can be achieved by any representation method of the original sampling functional units. Then the following computation gives the matrix version for L. For $\mathbf{f} \in H(R_q, T_q)$,

$$\begin{aligned} L_q\mathbf{f} &= E\left(E(\mathbf{X}|Y) \otimes_{R_q} E(\mathbf{X}|Y)\right)\mathbf{f} \\ &= E\left(E(\mathbf{X}|Y)\langle E(\mathbf{X}|Y), \mathbf{f}\rangle_{R_q}\right) \\ &= E\left(E(\mathbf{X}|Y)E(\mathbf{X}|Y)'R_q^{-1}\mathbf{f}\right) \\ &= [\mathrm{Cov}(E(\mathbf{X}|Y))R_q^{-1}]\mathbf{f}. \end{aligned}$$

Hence $L_q = \mathrm{Cov}(E(\mathbf{X}|Y))R_q^{-1}$, where $R_q = \{\mathrm{cov}(X(t_i), X(t_j))\}_{i,j=1,\ldots,q}$.

Spectral decomposition in the discrete setting can be carried on by applying matrix optimization:

$$\begin{aligned}
& \max_{\|\mathbf{f}\|_{R_q}=1} \langle E\left(E(\mathbf{X}|Y)\otimes_{R_q} E(\mathbf{X}|Y)\right)\mathbf{f},\mathbf{f}\rangle_{R_q} \\
= & \max \frac{\langle \mathrm{Cov}(E(\mathbf{X}|Y))R_q^{-1}\mathbf{f},\mathbf{f}\rangle_{R_q}}{\mathbf{f}'R_q^{-1}\mathbf{f}} \\
= & \max \frac{\mathbf{f}'R_q^{-1}\mathrm{Cov}(E(\mathbf{X}|Y))R_q^{-1}\mathbf{f}}{\mathbf{f}'R_q^{-1}\mathbf{f}} \\
= & \max \frac{\mathbf{g}'\mathrm{Cov}(E(\mathbf{X}|Y))R_q^{-1}\mathbf{g}}{\mathbf{g}'\mathbf{g}} \\
= & \max_{\|\mathbf{g}\|=1} \mathbf{g}'\mathrm{Cov}(E(\mathbf{X}|Y))R_q^{-1}\mathbf{g}.
\end{aligned}$$

Thus, the spectral decomposition of L_q can be achieved by the eigen-decomposition for the matrix $\mathrm{Cov}(E(\mathbf{X}|Y))R_q^{-1}$. The above calculation also implies

$$\operatorname*{argmax}_{\|\mathbf{f}\|_{R_q}=1} \langle E\left(E(\mathbf{X}|Y)\otimes_{R_q} E(\mathbf{X}|Y)\right)\mathbf{f},\mathbf{f}\rangle_{R_q} = R_q^{1/2}\operatorname*{argmax}_{\|\mathbf{g}\|=1} \mathbf{g}'\mathrm{Cov}(E(\mathbf{X}|Y))R_q^{-1}\mathbf{g}. \qquad (4.1)$$

In data analysis, we will also encounter the problem that the covariance kernel is unknown, so that we have to estimate it in some way.

4.2 IR Procedures in RKHS

We have two approaches to estimate $\mathcal{H}_{X,e}$:

If $Im(\tilde{L}) = \mathcal{H}_{X,e}$, the functional sliced inverse regression (FSIR) which generalizes Li (1991) can be used to estimate $\tilde{L}$.

If $Im(L) = \mathcal{H}_{X,e}$, we can implement the functional kernel inverse regression (FKIR), the extension of Zhu and Fang (1996) to estimate L.

Then the spectral decomposition of the estimated operators gives an estimate of the EDR space in RKHS.

Let $(\mathbf{X}_i, Y_i)$, $i = 1, \ldots, n$, be a discrete sample of (X, Y). Estimate the covariance matrix of $\mathbf{X}$ by $\hat{R}_{q,n}$. Then the FSIR algorithm is

Step 0. Centering $\mathbf{X}_i$'s.

Step 1. S-partition the range of Y to form $\{I_s, s = 1, \ldots, S\}$.

Step 2. $n_s = \sum_{i=1}^{n} I(Y_i \in I_s)$, $\hat{p}_{s,n} = \frac{n_s}{n}$, and $\hat{\mu}_{s,q,n} = \frac{1}{n_s} \sum_{i=1}^{n} \mathbf{X}_i I(Y_i \in I_s)$, $s = 1, \ldots, S$.

Step 3. $\hat{\tilde{L}}_{q,n} = \sum_{s=1}^{S} \hat{p}_{s,n} \left(\hat{\mu}_{s,q,n} \otimes_{\hat{R}_{q,n}} \hat{\mu}_{s,q,n} \right)$. Where $\mathbf{a} \otimes_{\hat{R}_{q,n}} \mathbf{a} = \mathbf{a}\mathbf{a}' \hat{R}_{q,n}^{-1}$.

Step 4. Implement the eigen-decomposition of $\hat{\tilde{L}}_{q,n}$.

As the FSIR directly estimates the operator $\tilde{L}$, the FKIR algorithm first estimates the conditional expectation $E(X|Y)$ by kernel smoothing then the covariance of it:

Step 0. Centering $\mathbf{X}_i$'s.

Step 1. Choose a kernel function and calculate

$$\hat{\mathbf{g}}_{q,n}(y) = \frac{1}{nh} \sum_{i=1}^{n} \mathbf{X}_i K\left(\frac{Y_i - y}{h} \right)$$

and

$$\hat{f}(y) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{Y_i - y}{h} \right).$$

Step 2. Choose a small positive number b and compute

$$\hat{f}_b(y) = max(b, \hat{f}(y)).$$

Step 3. Implement the Waston-Nadaraya type kernel estimation

$$\hat{\mathbf{m}}_{b,q,n}(y) = \frac{\hat{\mathbf{g}}_{q,n}(y)}{\hat{f}_b(y)},$$

where the bandwidth could be selected by cross-validation procedure.

Step 4. $\hat{L}_{q,n} = \frac{1}{n} \sum_{i=1}^{n} \hat{\mathbf{m}}_{b,q,n}(Y_i) \otimes_{\hat{R}_{q,n}} \hat{\mathbf{m}}_{b,q,n}(Y_i)$.

Step 5. Implement the eigen-decomposition of $\hat{L}_{q,n}$.

4.3 Asymptotic Studies in RKHS

To prove the consistency of the estimators introduced in the previous section is a challenging task. In the literature of functional data analysis, asymptotic results based on partially observed functional data have been rarely considered. For the present topic, under the assumption that each curve or sample unit is observed completely, in papers from the French school a number of asymptotic results were proved as the sample size n goes to infinity. We cannot try this approach, since even if observing the whole curves, i.e., the complete sample, $\{(X_i,Y_i), i=1,\ldots,n\}$, we cannot prove that the estimator function, $\hat{\mu}_{s,n} = \frac{1}{n}\sum_{i=1}^{n} X_i I(Y_i \in I_s)$ from FSIR or $\hat{g}_n(y) = \frac{1}{nh}\sum_{i=1}^{n} X_i K\left(\frac{Y_i - y}{h}\right)$ from FKIR falls into the RKHS of X. Actually, we have the following more general result.

Proposition IV.1. Let $X_1,\cdots,X_n$ be a sample from a zero-mean Gaussian process on a real interval T, with a continuous positive covariance kernel R. The sample paths of X are continuous on T, a.s. Define $Y = \sum_{i=1}^{n} c_i X_i$, where $\sum_{i=1}^{n} c_i^2 > 0$. Then $Y \notin H(R)$ a.s.

Proof. Obviously, Y is a Gaussian process with continuous sample paths, and it is trivial to know that its covariance kernel is $K = \sum_{i=1}^{n} c_i^2 R$. To apply Theorem 3 in Driscoll (1973), let $T_q = \{t_1,\cdots,t_q\}$ be a increasing series satisfying $\overline{\cup T_q} = T$. Define the restrictions $K_q = K|_{T_q\times T_q}$ and $R_q = R|_{T_q\times T_q}$, then

$$\begin{aligned}
\lim_{q\to\infty} tr(K_q R_q^{-1}) &= \lim_{q\to\infty} tr(\sum_{i=1}^{n} c_i^2 R_q R_q^{-1}) \\
&= \lim_{q\to\infty} q\sum_{i=1}^{n} c_i^2 \\
&= \infty.
\end{aligned}$$

This completes the proof. ♢

On the other hand, in most classical works on application of RKHS in statistics, asymptotic results are about the limiting behavior when number of observational points,

q, tends to infinity, for just one curve and with the known covariance kernel. In the following, we will attempt to handle these two problems. We will introduce a very useful mapping, A_q, and by using it, we can project the estimator into the RKHS to study the distance between the estimator and the target.

In the following, we will use the symbols $H(R) = H(R,T) = \mathcal{H}_X$ interchangeably. Let $T_q = \{t_1, \cdots, t_q\} \subset T$ be the observation locations. Denote $f_S = f|_S =$ the restriction of f on $S \subset T$ for a function f defined on T, and for convenience, define $f_q = f_{T_q}$ and the corresponding restriction of the RKHS, $H_q = H(R_q, T_q)$.

In the following we will prove the consistency of both FSIR and FKIR in the RKHS. We first provide some lemmas on some basic properties for Hilbert spaces and more results of RKHS. The first two lemmas are two general results from Hilbert space theory.

Lemma IV.1. In a Hilbert space H, $a, b \in H$, then $\|a \otimes b\|_{B(H)} = \|a\|\|b\|$.

Proof.

$$\begin{aligned} \|a \otimes b\|_{B(H)} &= \sup_{\|x\|=1} \|a \otimes b(x)\| \\ &= \sup_{\|x\|=1} \|\langle a, x\rangle b\| \\ &= (\sup_{\|x\|=1} |\langle a, x\rangle|)\|b\| \\ &= \|a\|\|b\|. \end{aligned}$$

$\diamondsuit$

Lemma IV.2. In a Hilbert space H, $\{x_n\}, \{y_n\} \subset H$ and $x, y \in H$, then we have the following results:

1. $x_n \rightharpoonup x \Longleftrightarrow$

   (1) $\|x_n\|$ is bounded;

   (2) $\exists M \subset H, \overline{M} = H, \forall m \in M, \langle x_n, m\rangle \to \langle x, m\rangle$.

2. $x_n \to x \Longleftrightarrow$

   (1) $x_n \rightharpoonup x$;

   (2)$\|x_n\| \to \|x\|$.

3. $x_n \to x, y_n \to y \Longrightarrow \|x_n \otimes y_n - x \otimes y\|_{B(H)} \to 0$.

Proof. 1. See Zhang and Lin (1987).

2. See Parzen (1959).

3. By Lemma IV.1 and the fact that $\|x_n\|$ is uniformly bounded, hence

$$\begin{aligned} & \|x_n \otimes y_n - x \otimes y\|_{B(H)} \\ = \; & \|x_n \otimes (y_n - y) + (x_n - x) \otimes y\|_{B(H)} \\ \leq \; & \|x_n \otimes (y_n - y)\|_{B(H)} + \|(x_n - x) \otimes y\|_{B(H)} \\ = \; & \|x_n\|\|y_n - y\| + \|x_n - x\|\|y\| \to 0. \end{aligned}$$

♢

The next two lemmas state two useful results for the RKHS. The first is for continuous kernels and another is about discrete kernels.

Lemma IV.3. If R is a continuous non-negative kernel on $T \times T$ and $\overline{T_0} = T$ then

$$H(R,T) = \overline{\text{span}\{R_t, t \in T_0\}}.$$

Proof. By Lemma 1 in Driscoll (1973), since R is a continuous, every function in $H(R,T)$ is continuous on T. If a function $g \in H(R,T)$ satisfies the condition $\langle g, R_t \rangle = 0$ for $\forall t \in T_0$, which implies $g(t) = 0$, for $\forall t \in T_0$, then due to continuity and dense property, we know $g(t) = 0$ for $\forall t \in T$, this completes the proof by Lemma 1a in Parzen (1959). ♢

Lemma IV.4. If T is finite with q elements and R a positive kernel on $T \times T$, i.e., $R \in \mathbf{R}^{q\times q}$, then $H(R,T) = \mathbf{R}^q$.

Proof. First, we will prove the following result of matrix computation. Let $\mathbf{A} = \{a_{ij}\} \in \mathbf{R}^{n\times n}$, $\mathbf{A} > 0$, $T = \{t_1,\ldots,t_q\} \subset \{1,\ldots,n\}$, $\mathbf{b} = (b_1,\ldots,b_n)' \in \mathbf{R}^n$. Define $\mathbf{A_T} = \mathbf{A}|_T = \{a_{t_i,t_j}, i,j \in \{1,\ldots,q\}\}$ and $\mathbf{b_T} = \mathbf{b}|_T = (b_{t_1},\ldots,b_{t_q})'$, for any $T_1 \subset T_2 \subset \{1,\ldots,n\}$, we have

$$\mathbf{b_{T_1}}'\mathbf{A_{T_1}}^{-1}\mathbf{b_{T_1}} \leq \mathbf{b_{T_2}}'\mathbf{A_{T_2}}^{-1}\mathbf{b_{T_2}}$$

The following facts will be useful. Let

$$\mathbf{A} = \begin{pmatrix} \mathbf{A_{11}} & \mathbf{A_{12}} \\ \mathbf{A_{21}} & \mathbf{A_{22}} \end{pmatrix},$$

assume that $\mathbf{A_{11}}$ is nonsingular, we can define $\mathbf{A_{22.1}} = (\mathbf{A}/\mathbf{A_{11}}) = \mathbf{A_{22}} - \mathbf{A_{21}}\mathbf{A_{11}}^{-1}\mathbf{A_{12}}$. It is easy to show that

(1) If $\mathbf{A}$ is symmetric, then $\mathbf{A} > 0$ is equivalent to that $\mathbf{A_{11}} > 0$ and $(\mathbf{A}/\mathbf{A_{11}}) > 0$.

(2) If $\mathbf{A}$ is symmetric and $\mathbf{A_{11}} > 0$, then $\mathbf{A} \geq 0$ is equivalent to $(\mathbf{A}/\mathbf{A_{11}}) > 0$.

Let $\mathbf{b_i} = \mathbf{b_{T_i}}$, $\mathbf{A_i} = \mathbf{A_{T_i}}$, $i = 1,2$. After appropriate arrangement, $\mathbf{b_2} = \begin{pmatrix} \mathbf{b_1} \\ \mathbf{c} \end{pmatrix}$ and $\mathbf{A_2} = \begin{pmatrix} \mathbf{A_1} & \mathbf{B} \\ \mathbf{B}' & \mathbf{0} \end{pmatrix}$, then

$$\begin{aligned} \mathbf{A_2}^{-1} &= \begin{pmatrix} \mathbf{A_1}^{-1} + \mathbf{F}\mathbf{E}^{-1}\mathbf{F}' & -\mathbf{F}\mathbf{E}^{-1} \\ -\mathbf{E}^{-1}\mathbf{F}' & \mathbf{E}^{-1} \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{A_1}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} + \begin{pmatrix} \mathbf{F}\mathbf{E}^{-1}\mathbf{F}' & -\mathbf{F}\mathbf{E}^{-1} \\ -\mathbf{E}^{-1}\mathbf{F}' & \mathbf{E}^{-1} \end{pmatrix}, \end{aligned}$$

where $\mathbf{E} = \mathbf{D} - \mathbf{B}'\mathbf{A_1}^{-1}\mathbf{B}$, $\mathbf{F} = \mathbf{A_1}^{-1}\mathbf{B}$.

$\mathbf{A_2} > 0$ implies that $\mathbf{E} = (\mathbf{A_2}/\mathbf{A_1}) > 0$. Defining

$$\mathbf{G} = \begin{pmatrix} \mathbf{F}\mathbf{E}^{-1}\mathbf{F}' & -\mathbf{F}\mathbf{E}^{-1} \\ -\mathbf{E}^{-1}\mathbf{F}' & \mathbf{E}^{-1} \end{pmatrix} = \begin{pmatrix} \mathbf{G_{11}} & \mathbf{G_{12}} \\ \mathbf{G_{21}} & \mathbf{G_{22}} \end{pmatrix},$$

$\mathbf{G_{11}} > 0$ and $(\mathbf{G}/\mathbf{G_{11}}) = \mathbf{G_{22}} - \mathbf{G_{21}}\mathbf{G_{11}}^{-1}\mathbf{G_{12}} = \mathbf{F}\mathbf{E}^{-1}\mathbf{F}' - \mathbf{F}\mathbf{E}^{-1}\mathbf{E}\mathbf{E}^{-1}\mathbf{F}' = \mathbf{0}$ implies that $\mathbf{G} \geq 0$.

$$\begin{aligned} \mathbf{b_2}'\mathbf{A_2}^{-1}\mathbf{b_2} &= \begin{pmatrix} \mathbf{b_1}' & \mathbf{c}' \end{pmatrix} + \left[\begin{pmatrix} \mathbf{A_1}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} + \mathbf{G} \right] \begin{pmatrix} \mathbf{b_1} \\ \mathbf{c} \end{pmatrix} \\ &= \mathbf{b_1}'\mathbf{A_1}^{-1}\mathbf{b_1} + \begin{pmatrix} \mathbf{b_1}' & \mathbf{c} \end{pmatrix} \mathbf{G} \begin{pmatrix} \mathbf{b_1} \\ \mathbf{c} \end{pmatrix}, \end{aligned}$$

Hence, $\mathbf{b_2}'\mathbf{A_2}^{-1}\mathbf{b_2} - \mathbf{b_1}'\mathbf{A_1}^{-1}\mathbf{b_1} = \begin{pmatrix} \mathbf{b_1}' & \mathbf{c} \end{pmatrix} \mathbf{G} \begin{pmatrix} \mathbf{b_1} \\ \mathbf{c} \end{pmatrix} \geq 0$.

Now we can use Property III.8. Let $\mathcal{F} = \{S \subset T\}$, $(a_1, \ldots, a_n) \in \mathbf{R}^n$, and $\sum_i \sum_j a_i a_j R(t_i, t_j) \neq 0$.

$\forall f \in \mathbf{R}^n$, $T = \{1, \ldots, n\}$, and $S \subset T$,

$$\sup_{\{a_1, \ldots, a_n\}} \frac{(\sum_{i \in S} a_i f_i)^2}{\sum_{i,j \in S} a_i a_j R(i,j)} = \frac{(\mathbf{a}'\mathbf{f}|_S)^2}{\mathbf{a}' R|_S \mathbf{a}} = \mathbf{f}|_S' R|_S^{-1} \mathbf{f}|_S,$$

Hence,

$$\sup_{S \in \mathcal{F}} \mathbf{f}|_S' R|_S^{-1} \mathbf{f}|_S = \mathbf{f}' R^{-1} \mathbf{f} < \infty.$$

$\diamondsuit$

Lemma IV.4 implies that the RKHS defined on finite index set is the whole Euclidean space but with possibly different topological structure. In the following three statements we will introduce a useful isometric mapping and give some properties.

Lemma IV.5. Let R be a kernel on $T \times T$. Given $T_q = \{t_1, \cdots, t_q\} \subset T$, the restrictions on R is $R_q = R|_{T_q \times T_q}$. Denote $\hat{H}_q = H(R_q, T_q)$, and $\tilde{H}_q = \text{span}\{R_t = R(t,\cdot), t \in T_q\} \subset H(R,T)$. Then $\hat{H}_q$ and $\tilde{H}_q$ are isometric isomorphic. Let A_q be the isometric isomorphism from $\hat{H}_q$ to $\tilde{H}_q$, if $R > 0$, A_q can be expressed explicitly.

Proof. See Theorem 6C in Parzen (1959) for proof of the congruence. For $\forall \mathbf{f} \in \hat{H}_q$, $\mathbf{f} \in \text{span}\{\mathbf{R}_t, t \in T_q\}$, where $\mathbf{R}_t = (R(t,t_1), \cdots, R(t,t_q))'$, $t \in T_q$, so $\mathbf{f}$ can be represented as $\mathbf{f} = \sum_{i=1}^{q} \mathbf{R}_{t_i} c_i$. Since $R > 0$ implies $R_q > 0$, we can solve that

$$\mathbf{c} = (c_1, \cdots, c_q)' = R_q^{-1}\mathbf{f},$$

Hence it is easy to show that

$$A_g(\mathbf{f}) = \sum_{i=1}^{q} R_{t_i} c_i = (R(t_1,\cdot), \cdots, R(t_q,\cdot)) R_q^{-1}\mathbf{f}.$$

$\diamondsuit$

An interesting fact can be derived from the proof of Lemma IV.5. Since $\mathbf{f} = R_q R_q^{-1}\mathbf{f}$, we know $(R(t_1,t), \cdots, R(t_q,t)) R_q^{-1}\mathbf{f} = f(t)$, if $t \in T_q$. This implies the trivial property, $A_q(f|_{T_q})|_{T_q} = f|_{T_q}$ for any $f \in H(R,T)$. Now we prove an important proposition.

Proposition IV.2. Let R be a continuous positive kernel on $T \times T$, which generates the RKHS, $H = H(R,T)$. Given $T_q = \{t_1, \cdots, t_q\} \subset T$, the restrictions on R is $R_q = R|_{T_q \times T_q}$, with $\hat{H}_q = H(R_q, T_q)$, $\tilde{H}_q = \text{span}\{R_t = R(t,\cdot), t \in T_q\} \subset H$. Let A_q be the isometric isomorphism from $\hat{H}_q$ to $\tilde{H}_q$. If $T_q \nearrow T_0$ and $\overline{T_0} = T$. For $a \in H$ and a is continuous on T, defining $a_q := a|_{T_q}$, we have

$$A_q(a_q) \to a \quad \text{in } H \quad \text{as } q \to \infty.$$

Proof. By Lemma IV.3, $H(R,T) = \overline{\text{span}\{R_t, t \in T_0\}}$. For $\forall t \in T_0$, $\exists Q$ and $\forall q > Q$, $t \in Tq$. Then

$$\langle A_q(a_q) - a, R_t \rangle_R = A_a(a_q)(t) - a(t) = 0,$$

which implies $A_q(a_q) \rightharpoonup a$ in span$\{R_t, t \in T_0\}$. By Property III.11,

$$\|A_q(a_q)\|_R = \|a_q\|_{R_q} \leq \|a\|_R,$$

so by Lemma IV.2, we know $A_q(a_q) \rightharpoonup a$ in H. Refer to Theorem 6E in Parzen (1959),

$$\|A_q(a_q)\|_R = \|a_q\|_{R_q} \nearrow \|a\|_R.$$

We complete the proof. $\diamondsuit$

A more insightful proposition is about the eigen-analysis properties of A_q. In the following, let $\lambda(L)$ and $\Lambda(L)$ denote respectively, one representative eigenvalue and the corresponding eigenvector (or eigenfunction) of the operator (or matrix) L. And let $(\lambda, \Lambda) = eigen(L)$ be the paired eigen-decomposition.

Proposition IV.3. Under the same conditions as Proposition IV.2.

(1) Let S be an index set, $\{c_s, s \in S\}$ is a real consequence, $\{a_s, s \in S\}, \{b_s, s \in S\} \subset H(R,T)$ and $a_{s,q}, b_{s,q}$ are the restrictions of a_s, b_s on T_q, respectively, then we have

$$\begin{aligned} & (\lambda, \Lambda) \in eigen\left(\sum c_s A_q(a_{s,q}) \otimes_R A_q(b_{s,q})\right) \\ \Longrightarrow \quad & (\lambda, \Lambda_q) \in eigen\left(\sum c_s a_{s,q} \otimes_{R_q} b_{s,q}\right) \\ \Longrightarrow \quad & (\lambda, A_q(\Lambda_q)) \in eigen\left(\sum c_s A_q(a_{s,q}) \otimes_R A_q(b_{s,q})\right). \end{aligned}$$

(2) Let W and Z be two $H(R,T)-$ valued random variables, then under the existence of expectation, we have,

$$\begin{aligned} & (\lambda, \Lambda_q) \in eigen(E(W_q \otimes_{R_q} Z_q)) \\ \Longrightarrow \quad & (\lambda, A_q(\Lambda_q)) \in eigen(E(A_q(W_q) \otimes_R A_q(Z_q))). \end{aligned}$$

Proof. (1)*Proof of Proposition IV.3 (1).*

$$(\sum c_s A_q(a_{s,q}) \otimes_R A_q(b_{s,q}))(\Lambda) = \lambda\Lambda$$

$$
\begin{aligned}
&\Longrightarrow \quad \sum c_s \langle A_q(a_{s,q}), \Lambda \rangle_R A_q(b_{s,q}) = \lambda\Lambda \\
&\Longrightarrow \quad \sum c_s \langle (R(t_1,\cdot),\cdots,R(t_q,\cdot))R_q^{-1} a_{s,q}, \Lambda \rangle_R A_q(b_{s,q}) = \lambda\Lambda \\
&\Longrightarrow \quad \sum c_s (\Lambda(t_1),\cdots,\Lambda(t_q))R_q^{-1} a_{s,q} A_q(b_{s,q}) = \lambda\Lambda \\
&\Longrightarrow \quad \sum c_s \langle a_{s,q}, \Lambda_q \rangle_{R_q} A_q(b_{s,q}) = \lambda\Lambda \\
&\Longrightarrow \quad \sum c_s \langle a_{s,q}, \Lambda_q \rangle_{R_q} A_q(b_{s,q})|_{T_q} = \lambda\Lambda|_{T_q} \\
&\Longrightarrow \quad \sum c_s \langle a_{s,q}, \Lambda_q \rangle_{R_q} b_{s,q} = \lambda\Lambda_q \\
&\Longrightarrow \quad (\sum c_s a_{s,q} \otimes_{R_q} b_{s,q})(\Lambda_q) = \lambda\Lambda_q.
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
&\quad (\sum c_s a_{s,q} \otimes_{R_q} b_{s,q})(\Lambda_q) = \lambda\Lambda_q \\
&\Longrightarrow \quad \sum c_s \langle A_q(a_{s,q}), A_q(\Lambda_q) \rangle_R b_{s,q} = \lambda\Lambda_q \\
&\Longrightarrow \quad (R(t_1,\cdot),\cdots,R(t_q,\cdot))R_q^{-1} \sum c_s \langle A_q(a_{s,q}), A_q(\Lambda_q) \rangle_R b_{s,q} \\
&\qquad = \lambda(R(t_1,\cdot),\cdots,R(t_q,\cdot))R_q^{-1}\Lambda_q \\
&\Longrightarrow \quad \sum c_s \langle A_q(a_{s,q}), A_q(\Lambda_q) \rangle_R A_q(b_{s,q}) = \lambda A_q(\Lambda_q).
\end{aligned}
$$

(2) *Proof of Proposition IV.3 (2).*

$$
\begin{aligned}
&\quad E(\langle W_q, \Lambda_q \rangle_{R_q})(Z_q) = \lambda\Lambda_q \\
&\Longrightarrow \quad E(\langle W_q, \Lambda_q \rangle_{R_q} Z_q) = \lambda\Lambda_q \\
&\Longrightarrow \quad (R(t_1,\cdot),\cdots,R(t_q,\cdot))R_q^{-1} E(\langle W_q, \Lambda_q \rangle_{R_q} Z_q) = \lambda(R(t_1,\cdot),\cdots,R(t_q,\cdot))R_q^{-1}\Lambda_q \\
&\Longrightarrow \quad E(\langle A_q(W_q), A_q(\Lambda_q) \rangle_R)(A_q(Z_q)) = \lambda A_q(\Lambda_q) \\
&\Longrightarrow \quad E(A_q(W_q) \otimes_R A_q(Z_q))(A_q(\Lambda_q)) = \lambda A_q(\Lambda_q).
\end{aligned}
$$

♢

The next lemma is an extended version of dominated convergence theorem in a Hilbert space.

Lemma IV.6. In a Hilbert space H, $Z_m \to Z$ a.s., and $\|Z_m\| \leq Y$ a.s., where Y is a real random variable with $E(Y^2) < \infty$, then

$$\lim_{m\to\infty} E\|Z_m \otimes Z_m - Z \otimes Z\|_{B(H)} = 0,$$

hence,

$$\lim_{m\to\infty} \|E(Z_m \otimes Z_m) - E(Z \otimes Z)\|_{B(H)} = 0.$$

Proof. By Lemma IV.2, $Z_m \otimes Z_m \to Z \otimes Z$ a.s. in $B(H)$. By Lemma IV.1, $\|Z_m \otimes Z_m\|_{B(H)} \leq Y^2$, and $\|Z \otimes Z\|_{B(H)} \leq Y^2$, so $E(Z_m \otimes Z_m) \in B(H)$ and $E(Z \otimes Z) \in B(H)$. By applying dominated convergence theorem (Bosq, 2000) in $B(H)$, we have $\lim_{m\to\infty} E\|Z_m \otimes Z_m - Z \otimes Z\|_{B(H)} = 0$, and by the triangle inequality (Bosq, 2000),

$$\|E(Z_m \otimes Z_m) - E(Z \otimes Z)\|_{B(H)} \leq E\|Z_m \otimes Z_m - Z \otimes Z\|_{B(H)}.$$

♢

The last Lemma is a simple result about the convergence of a double sequence.

Lemma IV.7. Let (M, d) be a metric space, $\{a_{q,n}\} \subset M$ and $a \in M$, if

$$\lim_{q\to\infty} \lim_{n\to\infty} a_{q,n} = a,$$

then there exists $n = n(q) \to \infty$, as $q \to \infty$, such that

$$\lim_{q\to\infty} a_{q,n(q)} = a.$$

Proof. It is easy to verify by the definition of the repeated limit of double sequences. ♢

Now back to our estimation. Finishing the eigen-analysis of the estimated operator, $\hat{\tilde{L}}_{q,n}$ in FSIR, or $\hat{L}_{q,n}$ from the FKIR, we intend to compare them with the eigen-elements from the true operator, $\tilde{L}$ in FSIR, or L from the FKIR. Since the eigenvectors from the

estimated operators are in the discrete RKHS, $H(R_q, T_q)$, by Lemma IV.5 the following two main theorems show how A_q works.

The first theorem reveals the strong consistency of the FSIR algorithm.

Theorem IV.1. Under the same notation and assumptions in Proposition IV.2, for fixed number of slices, assume that $\{\lambda(\tilde{L})\}$ are distinct and positive, then there exists $n = n(q) \to \infty$ as $q \to \infty$ such that

$$\lim_{q\to\infty} |\lambda(\hat{\tilde{L}}_{q,n}) - \lambda(\tilde{L})| = 0 \quad \text{a.s.},$$

$$\lim_{q\to\infty} \|A_q(\Lambda(\hat{\tilde{L}}_{q,n})) - \Lambda(\tilde{L})\|_R = 0 \quad \text{a.s.}.$$

Proof. Define $\mu_{s,q} = E(X_q|Y \in I_s)$ and $\mu_s = E(X|Y \in I_s)$, so $\tilde{L} = \sum_{s=1}^S p_s \mu_s \otimes_R \mu_s$ and let $\tilde{\tilde{L}} = \sum_{s=1}^S p_s \mu_{s,q} \otimes_{R_q} \mu_{s,q}$.

$$\begin{aligned} & |\lambda(\hat{\tilde{L}}_{q,n}) - \lambda(\tilde{L})| \\ \leq\ & |\lambda(\hat{\tilde{L}}_{q,n}) - \lambda(\tilde{\tilde{L}})| + |\lambda(\tilde{\tilde{L}}) - \lambda(\tilde{L})| \\ =:\ & (1) + (2). \end{aligned}$$

Due to Proposition IV.2, IV.3 and Lemma 3.1 in Bosq (1991),

$$\begin{aligned} (2) \quad =\ & |\lambda(\sum_{s=1}^S p_s A_q(\mu_{s,q}) \otimes_R A_q(\mu_{s,q})) - \lambda(\tilde{L})| \\ \leq\ & \|\sum_{s=1}^S p_s A_q(\mu_{s,q}) \otimes_R A_q(\mu_{s,q}) - \sum_{s=1}^S p_s \mu_s \otimes_R \mu_s\|_{B(H(R))} \\ \leq\ & \sum_{s=1}^S p_s \|A_q(\mu_{s,q}) \otimes_R A_q(\mu_{s,q}) - \mu_s \otimes_R \mu_s\|_{B(H(R))} \\ \to\ & 0, \quad \text{as} \quad q \to \infty, \end{aligned}$$

by Proposition IV.2.

For each q, $(1) \to 0$, as $n = n(q) \to \infty$, which is the result from multivariate SIR (Li, 1991). Hence,

$$\lim_{q\to\infty} \lim_{n\to\infty} |\lambda(\hat{\tilde{L}}_{q,n}) - \lambda(\tilde{L})| = 0 \quad \text{a.s.},$$

and by Lemma IV.7 the convergence of the eigenvalues is verified.

$$\begin{aligned} & \|A_q(\Lambda(\hat{\tilde{L}}_{q,n})) - \Lambda(\tilde{L})\|_R \\ \leq \; & \|A_q(\Lambda(\hat{\tilde{L}}_{q,n})) - \Lambda(\tilde{\tilde{L}})\|_R + \|\Lambda(\tilde{\tilde{L}}) - \Lambda(\tilde{L})\|_R \\ =: \; & (3) + (4). \end{aligned}$$

Also due to Proposition IV.2, IV.3 and Lemma 3.1 in Bosq (1991), there exists a constant c, such that,

$$\begin{aligned} (4) \quad = \quad & \|\Lambda(\sum_{s=1}^{S} p_s A_q(\mu_{s,q}) \otimes_R A_q(\mu_{s,q})) - \Lambda(\tilde{L})\|_R \\ \leq \quad & c\|\sum_{s=1}^{S} p_s A_q(\mu_{s,q}) \otimes_R A_q(\mu_{s,q}) - \sum_{s=1}^{S} p_s \mu_s \otimes_R \mu_s\|_{B(H(R))} \\ \to \quad & 0, \quad \text{as} \quad q \to \infty. \end{aligned}$$

For each q,

$$\begin{aligned} (3) \quad = \quad & \|\Lambda(\hat{\tilde{L}}_{q,n}) - \Lambda(\tilde{\tilde{L}})\|_{R_q} \\ \leq \quad & |R_q^{-\frac{1}{2}}(\Lambda(\hat{\tilde{L}}_{q,n}) - \Lambda(\tilde{\tilde{L}}))|_{E_q} \\ \to \quad & 0, \quad \text{a.s.}, \quad \text{as} \quad n = n(q) \to \infty, \end{aligned}$$

which is the result from multivariate SIR (Li, 1991). Hence,

$$\lim_{q\to\infty} \lim_{n\to\infty} \|A_q(\Lambda(\hat{\tilde{L}}_{q,n})) - \Lambda(\tilde{L})\|_R = 0 \quad \text{a.s..}$$

and by Lemma IV.7 the convergence of the eigenfunctions is verified. ♢

The second main theorem reveals the weak consistency of the FKIR estimation.

Theorem IV.2. Under the same notation and assumptions in Proposition IV.2, for fixed number of slices, assume that $\{\lambda(L)\}$ are distinct and positive, then there exists $n = n(q) \to \infty$ as $q \to \infty$ such that

$$\lim_{q\to\infty} |\lambda(\hat{L}_{q,n}) - \lambda(L)| = 0 \quad \text{in probability,}$$

$$\lim_{q\to\infty} \|A_q(\Lambda(\hat{L}_{q,n})) - \Lambda(L)\|_R = 0 \quad \text{in probability.}$$

Proof. Define $\mu = E(X|Y)$ and $\mu_q = E(X_q|Y)$, so $L = E(\mu \otimes_R \mu)$ and let $L_q = E(\mu_q \otimes_{R_q} \mu_q)$.

$$\begin{aligned} & |\lambda(\hat{L}_{q,n}) - \lambda(L)| \\ \leq\ & |\lambda(\hat{L}_{q,n}) - \lambda(L_q)| + |\lambda(L_q) - \lambda(L)| \\ =:\ & (1) + (2). \end{aligned}$$

Due to Proposition IV.2, IV.3 and Lemma 3.1 in Bosq (1991),

$$\begin{aligned} (2) \quad &= \quad |\lambda(E(A_q(\mu_q) \otimes_R A_q(\mu_q)) - \lambda(E(\mu \otimes_R \mu))| \\ &\leq \quad \|E(A_q(\mu_q) \otimes_R A_q(\mu_q)) - E(\mu \otimes_R \mu)\|_{B(H(R))} \\ &\to \quad 0, \quad \text{as} \quad q \to \infty. \end{aligned}$$

For each q, $(1) \to 0$, in probability, as $n = n(q) \to \infty$, which is the result from multivariate KIR (Zhu and Fang, 1996). Hence,

$$\lim_{q\to\infty} \lim_{n\to\infty} |\lambda(\hat{L}_{q,n}) - \lambda(L)| = 0 \quad \text{in probability,}$$

and by IV.7 the convergence of the eigenvalues is verified.

$$\begin{aligned} & \|A_q(\Lambda(\hat{L}_{q,n})) - \Lambda(L)\|_R \\ \leq\ & \|A_q(\Lambda(\hat{L}_{q,n})) - A_q(\Lambda(L_q))\|_R + \|A_q(\Lambda(L_q)) - \Lambda(L)\|_R \\ =:\ & (3) + (4). \end{aligned}$$

Also due to Proposition IV.2, IV.3 and Lemma 3.1 in Bosq (1991), there exists a constant c, such that,

$$\begin{aligned} (4) \quad &= \quad \|\Lambda(E(A_q(\mu_q) \otimes_R A_q(\mu_q))) - \Lambda(L)\|_R \\ &\leq \quad c\|E(A_q(\mu_q) \otimes_R A_q(\mu_q)) - E(\mu \otimes_R \mu)\|_{B(H(R))} \end{aligned}$$

$$\rightarrow \quad 0, \quad \text{as} \quad q \rightarrow \infty.$$

For each q,

$$\begin{aligned} (3) &= \|\Lambda(\hat{L}_{q,n}) - \Lambda(L_q)\|_{R_q} \\ &\leq |R_q^{-\frac{1}{2}}(\Lambda(\hat{L}_{q,n}) - \Lambda(L))|_{E_q} \\ &\rightarrow 0, \quad \text{in probability,} \quad \text{as} \quad n = n(q) \rightarrow \infty, \end{aligned}$$

which is the result from multivariate KIR (Zhu and Fang, 1996). Hence,

$$\lim_{q\rightarrow\infty} \lim_{n\rightarrow\infty} \|A_q(\Lambda(\hat{L}_{q,n})) - \Lambda(L)\|_R = 0 \quad \text{in probability.}$$

and by Lemma IV.7 the convergence of the eigenfunctions is verified. ♢

The major problem in asymptotic theory of estimators in RKHS is the space sensitivity, which means that the RKHS inner products depend on the kernel. The isometric isomorphism A_q we explored is very helpful in this situation.

CHAPTER V

INVERSE OF LOÉVE'S ISOMETRY – TRANSFORMATION FROM RKHS TO L^2 SPACE

After the EDRS in RKHS is estimated, we need to transform it to the original EDRS in L_X^2. This relates to a fundamental problem of how to numerically compute the inverse of the Loéve's isometry, or the back-transformation, Ψ^{-1}, from $H(R)$ to L_X^2. Parzen (1961) proposed an approximation and an iterative algorithm to derive the uniformly minimum variance unbiased linear estimate of functional of the mean for a process. Weiner (1965) improved Parzen's algorithm by gradient method. Both approaches assumed, as in the classical application of RKHS theory, that the kernel is known and the whole curve in RKHS is observed. We will describe their methods and provide a more convenient and direct approach for handling both the discrete observations and the empirical kernels.

5.1 Parzen's Methods

Let L_X^2 and $\mathcal{H}_X = H(R,T)$ be the Hilbert space and RKHS generated by L^2 process $\{X_t, t \in T\}$ with $\mathrm{E}X_t = 0$ and $R(t,s) = E(X_t X_s)$ for all $t,s \in T = [a,b]$, with R being a known continuous positive definite function. Let Ψ be Loéve's isometry from L_X^2 to $\mathcal{H}_X$. For $h \in \mathcal{H}_X$, to calculate $\langle h,h\rangle_R$ and $\Psi^{-1}(h) \in L_X^2$, Parzen (1961) gave the following approximation by truncation.

Let $C(a,b)$ be the space of continuous function on $[a,b]$ and let $\{(\lambda_n,\phi_n), n \in \mathbf{N}\}$ be the eigenvalues and eigenfunctions of the kernel R arranged in decreasing order $\lambda_1 \geq \lambda_2 \geq \ldots \geq 0$. By defining

$$H_n(\cdot) = \sum_{k=1}^{n} \phi_k(\cdot)\frac{1}{\lambda_k}\int_a^b h(s)\phi_k(s)\,\mathrm{d}s \in C(a,b),$$

one can verify that

$$\int_a^b \int_a^b H_n(s)R(s,t)H_n(t)\,\mathrm{d}s\,\mathrm{d}t = \sum_{k=1}^{n} \frac{1}{\lambda_k} \left| \int_a^b h(t)\phi_k(t)\,\mathrm{d}t \right|^2, \tag{5.1}$$

$$\int_a^b H_n(t)X(t)\,\mathrm{d}t = \sum_{k=1}^{n} \frac{1}{\lambda_k} \int_a^b h(s)\phi_k(s)\,\mathrm{d}s \int_a^b X(t)\phi_k(t)\,\mathrm{d}t, \tag{5.2}$$

and further

$$\lim_{n\to\infty} \mathrm{E}\left[\left| \Psi^{-1}(h) - \int_a^b H_n(t)X(t)\,\mathrm{d}t \right|^2 \right] = 0,$$

and

$$\langle h,h\rangle_R = \lim_{n\to\infty} \int_a^b \int_a^b H_n(s)R(s,t)H_n(t)\,\mathrm{d}s\,\mathrm{d}t.$$

Hence, the right-hand sides of (5.1) and (5.2) can approximate $\langle h,h\rangle_R$ and $\Psi^{-1}(h)$, respectively.

Parzen (1961) also provided an iterative algorithm to calculate H_n.

$$\begin{cases} H_0(t) = 1 \\ H_{n+1} = H_n - \alpha(LH_n - h), \qquad n \geq 1, \end{cases}$$

where α is a constant such as $\alpha \in (0, \frac{2}{M}]$ with $M > \lambda_1$ and L is the transformation

$$Lf(\cdot) = \int_a^b f(s)R(s,\cdot)\,\mathrm{d}s,\ f \in C[a,b].$$

By the reproducing property,

$$\Psi^{-1}(Lf) = \int_a^b f(t)X(t)\,\mathrm{d}t$$

and

$$\langle Lf, Lf\rangle_R = \int_a^b \int_a^b f(s)R(s,t)f(t)\,\mathrm{d}s\,\mathrm{d}t.$$

It could be verified

$$E(|\Psi^{-1}(h)-\Psi^{-1}(LH_n)|^2)=\|h-LH_n\|_R^2\to 0,\ \text{as } n\to\infty,$$

hence

$$\langle h,h\rangle_R=\lim_{n\to\infty}\langle LH_n,LH_n\rangle_R.$$

5.2 Weiner's Improved Iteration

Another approximation was proposed by Weiner (1965) which improved the iterative algorithm described in Section 5.1 by gradient method. Let G be a Hilbert space of functions defined on T with a computationally convenient inner product. For $g\in G$, define the transformation

$$Ag(t)=\langle g(\cdot),R(\cdot,t)\rangle_G.$$

Choose an arbitrary $H_0\in H(R)$, then iterate the following starting from $n=0$:

$$\begin{aligned} r_n &= h-AH_n,\\ a_n &= \frac{\langle r_n,r_n\rangle_G}{\langle r_n,Ar_n\rangle_G},\\ H_{n+1} &= H_n+a_nr_n. \end{aligned}$$

Weiner (1965) proved, under mild conditions, that $\|r_n\|_R^2\searrow 0$, which implies that

$$\lim_{n\to\infty}E|\Psi^{-1}(h)-\langle H_n,X\rangle_G|\to 0,$$

and

$$\langle h,h\rangle_R=\lim_{n\to\infty}\langle H_n,AH_n\rangle_G.$$

The drawback of this algorithm is that there is no clear stopping rule because the norm of $H(R)$ is unknown.

5.3 Direct Approach

Both Parzen (1961) and Weiner (1965) assumed that the covariance function R is known. When R is unknown but could be estimated from the data, it seems plausible to use the estimated covariance in the procedures above. In the following, we address this problem. We provide here, under the discrete setting, a simpler way considering both discretization of the sample and estimation of the unknown kernel, then the consistence results will be verified.

Under the same setting as Section 4.1, $\{X(t), t \in T\}$ is observed only on a discrete set of values of T, say $T_q = \{t_1, \ldots, t_q\}$. Let ξ be an element in L_X^2, and correspondingly, $\eta = \Psi(\xi) \in \mathcal{H}_X$. Under the same conditions in Proposition IV.2, we observe η at T_q, hence $\eta_q = \eta|_{T_q} \in H(R_q, T_q)$, where $R_q = R|_{T_q \times T_q}$.

If R is given, define the approximation of the back-transformation at T_q,

$$\Psi_q^{-1}(\cdot) = R_q^{-1}\cdot,$$

and the corresponding back-transformed element,

$$\hat{\xi}_q = \Psi_q^{-1}(\eta_q).$$

Under a typical FDA setting, let $\{\mathbf{X}_i, i = 1, \ldots, n\}$, be a discrete sample of X. Based on the discrete sample, suppose that η_q could be estimated by $\hat{\eta}_{q,n}$ from the sample, which satisfies for each q, $\lim_{n\to\infty} \hat{\eta}_{q,n} = \eta_q$, in probability, and R_q is estimated by $\hat{R}_{q,n}$, satisfying that

$$\lim_{n\to\infty} \hat{R}_{q,n} = R_q, \text{ in probability.}$$

Defining, respectively,

$$\hat{\Psi}_{q,n}^{-1}(\cdot) = \hat{R}_{q,n}^{-1}\cdot,$$

and

$$\hat{\xi}_{q,n} = \hat{\Psi}_{q,n}^{-1}(\hat{\eta}_{q,n}),$$

we have,

Theorem V.1. Under the above conditions,

$$\lim_{q\to\infty} \|\hat{\xi}'_q(X(t_1),\cdots,X(t_q))' - \xi\|_{L_X^2} = 0 \tag{5.3}$$

and there exists $n = n(q) \to \infty$ as $q \to \infty$ such that

$$\lim_{q\to\infty} \|\hat{\xi}'_{q,n}(X(t_1),\cdots,X(t_q))' - \xi\|_{L_X^2} = 0 \quad \text{in probability.} \tag{5.4}$$

Proof. By Proposition IV.2,

$$\begin{aligned} & \|(X(t_1),\cdots,X(t_q))\Psi_q^{-1}(\eta_q) - \xi\|_{L_X^2} \\ = \;& \|(R(t_1,\cdot),\cdots,R(t_q,\cdot))R_q^{-1}(\eta_q) - \Psi(\xi)\|_R \\ = \;& \|A_q(\eta_q) - \eta\|_R \\ \to \;& 0, \quad \text{as} \quad q \to \infty, \end{aligned}$$

which verifies (5.3).

$$\begin{aligned} & \|(X(t_1),\cdots,X(t_q))\hat{\xi}_{q,n} - \xi\|_{L_X^2} \\ \leq \;& \|(X(t_1),\cdots,X(t_q))\hat{\xi}_{q,n} - (X(t_1),\cdots,X(t_q))\hat{\xi}_q\|_{L_X^2} \\ & + \|(X(t_1),\cdots,X(t_q))\hat{\xi}_q - \xi\|_{L_X^2} \\ =: \;& (1) + (2). \end{aligned}$$

For each q,

$$\begin{aligned} (1) \;=\; & \|(X(t_1),\cdots,X(t_q))(\hat{R}_{q,n}^{-1}\hat{\eta}_{q,n} - R_q^{-1}\eta_q)\|_{L_X^2} \\ \leq \; & \|\sum_{k=1}^{q} a_{q,n}^k X(t_k)\|_{L_X^2} \end{aligned}$$

$$\begin{aligned} &\leq \sum_{k=1}^{q} |a_{q,n}^k| \|X(t_k)\|_{L_X^2} \\ &\to 0, \quad \text{as} \quad n \to \infty, \end{aligned}$$

where

$$(a_{q,n}^k, k = 1, \cdots, q)' = \hat{R}_{q,n}^{-1} \hat{\eta}_{q,n} - R_q^{-1} \eta_q \to \mathbf{0} \quad \text{in probability.}$$

And $(2) \to 0$ is just verified by (5.3), which completes the proof of (5.4). $\diamondsuit$

With this transformation, the whole inverse regression approach is the same as the multivariate case in Li (1991). Considering the norms, since $\|\eta_q\|_{R_q}^2 = \eta_q' R_q^{-1} \eta_q$ and $\|\hat{\eta}_{q,n}\|_{\hat{R}_{q,n}}^2 = \hat{\eta}_{q,n}' \hat{R}_{q,n}^{-1} \hat{\eta}_{q,n}$, by properties of RKHS and similar arguments to the above theorems, we have

$$\left| \|\eta_q\|_{R_q}^2 - \|\eta\|_R \right| \to 0, \qquad \text{as } q \to \infty,$$

and there exists $n = n(q) \to \infty$ as $q \to \infty$ such that

$$\left| \|\hat{\eta}_{q,n}\|_{\hat{R}_{q,n}}^2 - \|\eta\|_R \right| \to 0, \qquad \text{as } q \to \infty.$$

CHAPTER VI

COMPUTATIONAL ISSUES IN FDA

Computational statistics is as equally important as theoretical and applied statistics since the implementation of modern data analytic tools is heavily dependent on computing. In this chapter, two general computational issues will be discussed in FDA setting. In the first section, smoothing will be reviewed and the smoothed versions of FSIR and FKIR will be proposed. Section 7.2 is about the use of generalized inverse in FDA.

6.1 Smoothing

Smoothing is an important issue in FDA. Most FDA is carried out in infinite-dimensional function spaces, hence the nature of FDA is nonparametric. As a result, nonparametric smoothing techniques are highly relevant for a successful FDA procedure.

Even in the French school, Cardot (2000) incorporated smoothing in the functional principal component analysis using B-spline, and verified the benefit of smoothing. A more general functional regression model called functional nonparametric regression with application to TSA was presented in Ferraty and Vieu (2004). This model is defined as

$$Y_i = r(X_i) + \varepsilon_i, i = 1, \cdots, n;$$

where Y_i is the real response, the explanatory variable X_i belongs an abstract space with semi-metric $d(\cdot,\cdot)$. The nonparametric estimate is then defined by

$$\hat{r}(x) = \frac{\sum Y_i K(d(x,X_i)/h)}{\sum K(d(x,X_i)/h)},$$

with smoothing kernel function K and smoothing parameter h. Ferraty and Vieu (2002) proposed the algorithms with applications in spectrometric data.

When it comes to smoothing functional data, the following considerations are important. The first is whether the smoothing procedure is carried out based on data from individual curve or all curves. The second is whether the smoothing is done before, during, or after the formal analysis. Ramsay and Silverman (1997) followed a more traditional approach to smooth each curve individually as a part of the pre-processing, while some other works, for example, Rice and Silverman (1991), simultaneously implemented smoothing and analysis. Rice and Silverman (1991) estimated the mean curve using penalized least squares, where the amount of smoothing was determined by a cross validation criterion leaving out a whole individual curve at a time. Rice (2004) also discussed this issue for smoothing multiple curves.

A functional dataset could be observed in the following three formats:

(Format 1.) Rectangle with equal-space design: $\{x_i(t_j), j = 1,\ldots,m, i = 1,\ldots,n\}$, where $|t_{j+1} - t_j|$ are equal for $j = 1,\ldots,m-1$.

(Format 2.) Rectangle with general-space design: $\{x_i(t_j), j = 1,\ldots,m, i = 1,\ldots,n\}$.

(Format 3.) General design: $\{x_i(t_{ij}), j = 1,\ldots,m_i, i = 1,\ldots,n\}$.

Format 1 gives the data a matrix-structure, Format 2 a table-structure, and Format 3 a list-structure. Multivariate data analysis can be applied to Format 1 directly, but smoothing is useful for all three situations, such as for interpolation of data at unobserved points.

An effective nonparametric estimate of the covariance function is of prime interest. Both Diggle and Verbyla (1998) and Staniswalis and Lee (1998) used two-dimensional smoothing techniques, Fan and Zhang (2002) proposed a two-step procedure containing a raw estimation by standard linear model followed by a refinement by smoothing. This idea was borrowed by Lee et al. (2002) and Wu and Pourahmad (2003). The former implemented a usual principal component analysis based on data of Format 1 to get the raw

estimates of eigenvalues and eigenfunctions, then polished the estimates by smoothing and finally estimated the covariance by spectral decomposition. The latter used autoregression techniques to guarantee that the estimated covariance matrix is positive definite.

Below we consider smoothing issues in inverse regression. The basic idea is to plug in the smoothed estimation of covariance into the FSIR and FKIR procedures.

In the following, we use $smooth1d$ and $smooth2d$ to denote the one-dimensional and two-dimensional local linear smoothing procedures (Bowman and Azzalini 1997), respectively. Let $X = \{x_{ij} = x_i(t_{ij}), j = 1,\ldots,m_i, i = 1,\ldots,n\}$, where t_{ij}'s are in an interval T. The smoothing operations for mean and covariance functions are described as follows:

Operation $SM(X)$ —- Smoothed estimation of the mean

$$\hat{\mu} = smooth1d\left(\{(t_{ij}, x_{ij}), j = 1,\ldots,m_i, i = 1,\ldots,n\}\right).$$

This operation includes cross-validation (CV) procedure in which one minimizes the CV score with respect to the bandwidth b given by

$$\mathrm{CV}(b) = \sum_{i=1}^{n}\sum_{j=1}^{m_i}\{X_{ij} - \hat{\mu}^{(-i)}(t_{ij}; b)\}^2/N,$$

where $\hat{\mu}^{(-i)}$ is the estimate after removing the ith curve.

Operation $DM(X)$ —- Centering X after SM(X), that is, performing the calculation,

$$\{\hat{x}_{ij} = \hat{x}_i(t_{ij}) = x_{ij} - \hat{\mu}(t_{ij}), j = 1,\ldots,m_i, i = 1,\ldots,n\}$$

Operation $SC(X)$ —- Smoothed estimation of the covariance, which includes two parts.

Variance:

$$\hat{R}(t,t) = smooth1d\left(\{(t_{ij}, \hat{x}_{ij}^2), j = 1,\ldots,m_i, i = 1,\ldots,n\}\right);$$

Covariance: Let

$$\hat{u}(t,s) = smooth2d\left(\{(t_{ij}, t_{ik}, \frac{1}{2}(\hat{x}_{ij} - \hat{x}_{ik})^2), 1 \le j < k \le m_i, i = 1, \ldots, n\}\right),$$

then

$$\hat{R}(t,s) = \frac{1}{2}\{\hat{R}(t,t) + \hat{R}(s,s)\} - \hat{u}(t,s),$$

or directly

$$\hat{R}(t,s) = smooth2d\left(\{(t_{ij}, t_{ik}, \hat{x}_{ij}\hat{x}_{ik}), 1 \le j < k \le m_i, i = 1, \ldots, n\}\right),$$

The CV procedures similar to that in $SM(X)$ are also included in $SC(X)$.

Then we can design the smoothing versions of FSIR and FKIR algorithms. Let $T_q = \{t_1, \ldots, t_q\}$.

Smoothed FSIR

Step 0. Apply $SM(X)$, $\hat{X} = DM(X)$ and $\hat{R}_{q,n} = SC(X)$ at $T_q \times T_q$.

Step 1. S-partition the range of Y to form $\{I_s, s = 1, \ldots, S\}$.

Step 2. $n_s = \sum_{i=1}^{n} I(Y_i \in I_s)$, $\hat{p}_{s,n} = \frac{n_s}{n}$, $\hat{X}_s = \{\hat{X}_{ij} | Y_i \in I_s\}$ and $\hat{\mu}_{s,q,n} = SM(\hat{X}_s)$ at T_q, $s = 1, \ldots, S$.

Step 3. $\hat{\tilde{L}}_{q,n} = \sum_{s=1}^{S} \hat{p}_{s,n}\left(\hat{\mu}_{s,q,n} \otimes_{\hat{R}_{q,n}} \hat{\mu}_{s,q,n}\right)$.

Step 4. Implement the eigen-decomposition of $\hat{\tilde{L}}_{q,n}$.

Smoothed FKIR

Step 0. Apply $SM(X)$, $\hat{X} = DM(X)$, $\hat{R}_{q,n} = SC(X)$ at $T_q \times T_q$, and $\hat{X}_{i,q} =$ smooth each $\hat{X}_i$ based on individual or all curves and output at T_q.

Step 1. Choose a kernel function and calculate

$$\hat{g}_{q,n}(y) = \frac{1}{nh}\sum_{i=1}^{n}\hat{X}_{i,q}K\left(\frac{Y_i - y}{h}\right)$$

and

$$\hat{f}(y) = \frac{1}{nh}\sum_{i=1}^{n}K\left(\frac{Y_i - y}{h}\right).$$

Step 2. Choose a small positive number b and compute

$$\hat{f}_b(y) = max(b, \hat{f}(y)).$$

Step 3. Implement the Waston-Nadaraya type kernel estimation

$$\hat{m}_{b,q,n}(y) = \frac{\hat{g}_{q,n}(y)}{\hat{f}_b(y)},$$

where the bandwidth could be selected by cross-validation procedure.

Step 4. Let $\hat{M}_{q,n} = SC(\{\hat{m}_{b,q,n}(Y_i), i = 1,\ldots,n\})$, $\hat{L}_{q,n} = \hat{M}_{q,n}\hat{R}_{q,n}$.

Step 5. Implement the eigen-decomposition of $\hat{L}_{q,n}$.

The smoothed estimation of covariance function could also be plugged in the stage of performing the back-transformation stage to provide additional smoothing to the procedure.

6.2 Inverse of Covariance Operator

Let X be an H-valued random variable, where H is a Hilbert space with inner product $\langle\cdot,\cdot\rangle$ and the induced norm $\|\cdot\|$. If $\mathrm{E}\|X\|^2 < \infty$, we can define the covariance operator $\Gamma = \mathrm{E}(X \otimes X)$ which belongs to the trace-class operators (Zhang and Guo 1990), this implies its compactness, and hence if $dim(H) = \infty$ then the inverse of Γ either does not exist or is unbounded. As a result, the behavior of the estimated inverse covariance matrix would be unstable with the size getting large.

The projection method proposed by Bosq (1991) and described in Chapter II is a nice solution to handle the inverse. Let $\{(\lambda_i, v_i), i \in \mathbf{N}\}$ be the eigenvalues and corresponding eigenfunctions of Γ, where $\lambda_1 \geq \lambda_2 \geq \ldots \geq 0$ (since Γ is a positive operator) and v_i's comprise an orthonormal basis of H. Suppose $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_K > 0$, then define $\Pi_K = \sum_{k=1}^{K} v_k \otimes v_k$, the projector operator on $V_K = \text{span}\{v_1, \ldots, v_K\}$.

Consider the operator $\Gamma_K^+ = (\Pi_K \Gamma \Pi_K)^{-1}$, which has the following two properties:

(1) $\Gamma_K^+ = \sum_{k=1}^{K} \frac{1}{\lambda_k} v_k \otimes v_k$.

Proof. By the spectral decomposition of Γ and properties of tensor product,

$$\begin{aligned} \Pi_K \Gamma \Pi_K &= \Pi_K \sum_{i=1}^{\infty} \lambda_i (v_i \otimes v_i) \Pi_K \\ &= \sum_{i=1}^{\infty} \lambda_i (\Pi_K v_i) \otimes (\Pi_K v_i) \\ &= \sum_{i=1}^{K} \lambda_i (v_i \otimes v_i). \end{aligned}$$

Hence $(\Pi_K \Gamma \Pi_K)^{-1} = \sum_{k=1}^{K} \frac{1}{\lambda_k} v_k \otimes v_k$. ♢

Where $\sum_{i=1}^{K} \lambda_i (v_i \otimes v_i)$ is called a partial spectral decomposition.

(2) $\Gamma_K^+ \Gamma \Gamma_K^+ = \Gamma_K^+$.

Proof. Plugging in the result of (1),

$$\begin{aligned} \Gamma_K^+ \Gamma \Gamma_K^+ &= \Gamma_K^+ \Gamma \left(\sum_{k=1}^{K} \frac{1}{\lambda_k} v_k \otimes v_k \right) \\ &= \Gamma_K^+ \left(\sum_{k=1}^{K} \frac{1}{\lambda_k} v_k \otimes \Gamma v_k \right) \\ &= \left(\sum_{k=1}^{K} \frac{1}{\lambda_k} v_k \otimes v_k \right) \left(\sum_{k=1}^{K} v_k \otimes v_k \right) \\ &= \sum_{k=1}^{K} \frac{1}{\lambda_k} v_k \otimes v_k. \end{aligned}$$

$\diamondsuit$

This implies that Γ_K^+ is a generalized inverse of Γ.

The counterpart of the above is easier to understand for covariance matrix. Let Σ be a $p \times p$ non-negative definite matrix. By the Cholesky's decomposition, $\Sigma = V\Lambda V'$, where $\Lambda = diag(\lambda_1, \ldots, \lambda_p)$, $V = (\nu_1, \ldots, \nu_p)$, $\lambda_1 \geq \lambda_2 \geq \ldots \geq 0$ are the eigenvalues, and $\nu_1, \ldots, \nu_p$ are corresponding orthonormal eigenvectors. For some $k < p$, a partial spectral decomposition is defined as $V_{(k)}\Lambda_{(k)}V'_{(k)}$, where $\Lambda_{(k)} = diag(\lambda_1, \ldots, \lambda_k)$, $V_{(k)} = (\nu_1, \ldots, \nu_k)$, and $\lambda_1 \geq \lambda_2 \geq \ldots \lambda_k > 0$. And define

$$\Sigma_k^+ = V_{(k)}(V_{(k)}\Lambda_{(k)}V'_{(k)})^{-1}V'_{(k)},$$

which implies $\Sigma_k^+ = \sum_{i=1}^k \frac{1}{\lambda_i}\nu_i\nu_i'$ and $\Sigma_k^+\Sigma\Sigma_k^+ = \Sigma_k^+$.

The partial spectral decomposition and the projected inverse are used extensively in the practice of our methods, both in RKHS and the back-transformation.

CHAPTER VII

EMPIRICAL STUDIES

Empirical studies of the FSIR, the FKIR, the back-transformation, and their smoothed versions will be reported in this chapter. Section 7.1 lists some simulation results, two examples of data analysis are described in Section 7.2. IR for hybrid data is developed in Section 7.3, which also includes some simulations and an application.

7.1 Monte Carlo Experiments

We will consider a specific type of the stochastic multiple-index model (3.1) based on a L^2 process $\{X_t, t \in [a,b]\}$, the functional additive model

$$y = \sum_{i=1}^{p} f_i(\xi_i) + \varepsilon,$$

where $\xi_i = \int_a^b \beta_i(t)X(t)\,dt, i = 1,\ldots,p,$ which are L^2-integrable (Ash and Gardner 1975). We intend to estimate span$(\beta_1,\ldots\beta_p)$.

Three processes are included in the simulation studies: a simple three-component process, the standard Brownian motion and the fractional Brownian motion. The three-component process is defined as

$$S(t) = \sqrt{2}[2\sin(3\pi t)U_1 + 1.5\sin(5\pi t)U_2 + 0.8\cos(7\pi t)U_3],$$

where U_1, U_2 and U_3 are i.i.d. standard normal variables. Its covariance function is

$$\text{cov}(S(t),S(s)) = 8\sin(3\pi s)\sin(3\pi t) + 4.5\sin(5\pi s)\sin(5\pi t) + 1.28\sin(7\pi s)\sin(7\pi t).$$

Brownian motion is one of the most important stochastic processes. The definition of a Brownian motion is:

Definition VII.1. (Brownian Motion)

A Gaussian process $\{B(t), t \in \mathbf{R}\}$ is called a Brownian motion (Bm), if it satisfies that

(1)$B(0) = 0$ a.s. and $B(t)$ is a continuous function of t;

(2)$B(t+h) - B(t)$ is distributed as $N(0, \sigma^2|h|)$,

where σ is a positive constant. It is called a standard Bm when $\sigma = 1$.

Properties of Bm could be referred to Ash and Gardner (1975). One of the generalizations of Bm is the following.

Definition VII.2. (Fractional Brownian Motion)

A Gaussian process $\{B_H(t), t \in \mathbf{R}\}$ is called a fractional Brownian motion (fBm) of Hurst index $H \in (0,1)$ such that

(1)$B_H(0) = 0$ a.s. and $B_H(t)$ is a continuous function of t;

(2)$B_H(t+h) - B_H(t)$ is distributed as $N(0, \sigma^2|h|^{2H})$,

Nuzman and Poor (2001) applied the RKHS methods to fBm. The case $H = 1/2$ reduces to Bm, and a fBm can be presented as an integral of Bm with the following version from Mandelbrot and Van Ness (1968)

$$B_H(t) = \begin{cases} \frac{1}{\Gamma(H+\frac{1}{2})} \int_{-\infty}^{t} \left(|t-s|^{H-1/2} - |s|^{H-1/2} \right) \mathrm{d}B(s) & t < 0 \\ 0 & t = 0 \\ \frac{1}{\Gamma(H+\frac{1}{2})} \left(\int_{-\infty}^{0} \left(|t-s|^{H-1/2} - |s|^{H-1/2} \right) \mathrm{d}B(s) + \int_{0}^{t} |t-s|^{H-1/2} \mathrm{d}B(s) \right) & t > 0 \end{cases} \tag{7.1}$$

where $B(s)$ is a standard BM and $\Gamma(r) = \int_0^\infty x^{r-1} e^{-x} \mathrm{d}x$, $r > 0$.

The covariance function of a fBm is derived from the definition

$$\mathrm{cov}(B_H(t), B_H(s)) = \frac{\sigma^2}{2} \left(|t|^{2H} + |s|^{2H} - |t-s|^{2H} \right).$$

As a direct consequence, the covariance of BM is $\text{cov}(B(t),B(s)) = \frac{\sigma^2}{2}\min(t,s)$.

Without loss of generality, the index set T will be restricted to $[0,1]$. Two trajectories of S and Bm are illustrated in Figure 1, and two trajectories of fBm ($H = 0.25$) and fBm ($H = 0.75$) in Figure 2 , respectively.

We firstly consider the single-index model

$$y = f(\xi) + \varepsilon,$$

where $\xi = \int_0^1 \beta(t)x(t)\,dt$, $\beta(t)$ is the index coefficient function and $\varepsilon \sim N(0,\sigma^2)$.

Experiment 1. Let $\{x(t), t \in [0,1]\}$ be a standard Bm, $\varepsilon \sim N(0,\sigma^2 = 0.1)$, $f(t) = \exp(t)$ and $\beta(t) = t$. Let $h(t) = \Psi(\xi)(t)$, which can be calculated as

$$h(t) = \int_0^1 t(s)\min(s,t)\,ds = \frac{t}{2} - \frac{t^3}{6}.$$

We generate $n = 500$ data points and each trajectory has been sampled at $q = 101$ equally spaced time points in $[0,1]$. By FSIR in the RKHS, the first eigenvalue takes 96% of the total which detects a single-index model significantly. The normalized estimation in RKHS is shown in the upper panel of Figure 3. We then transform it to the L^2 space and equivalently get the normalized estimate of β, which is shown in the lower panel of Figure 3. Both estimates are compared to the target functions normalized in the corresponding spaces, respectively.

Experiment 2. Let $\{x(t), t \in [0,1]\}$ be a standard Bm, $\varepsilon \sim N(0,\sigma^2 = 0.01)$, $f(t) = \tan^{-1}(t)$ and $\beta(t) = t$.

We intend to see the performance of smoothed FSIR. We generate a relatively small sample with $n = 10$ curves. For each curve, we calculate y throughout $q = 51$ equally spaced time points. Then we follow Hsing (2004) to generate a sampling scheme of Format 3 introduced in Section 6.1. Suppose that we get a dataset following sampling

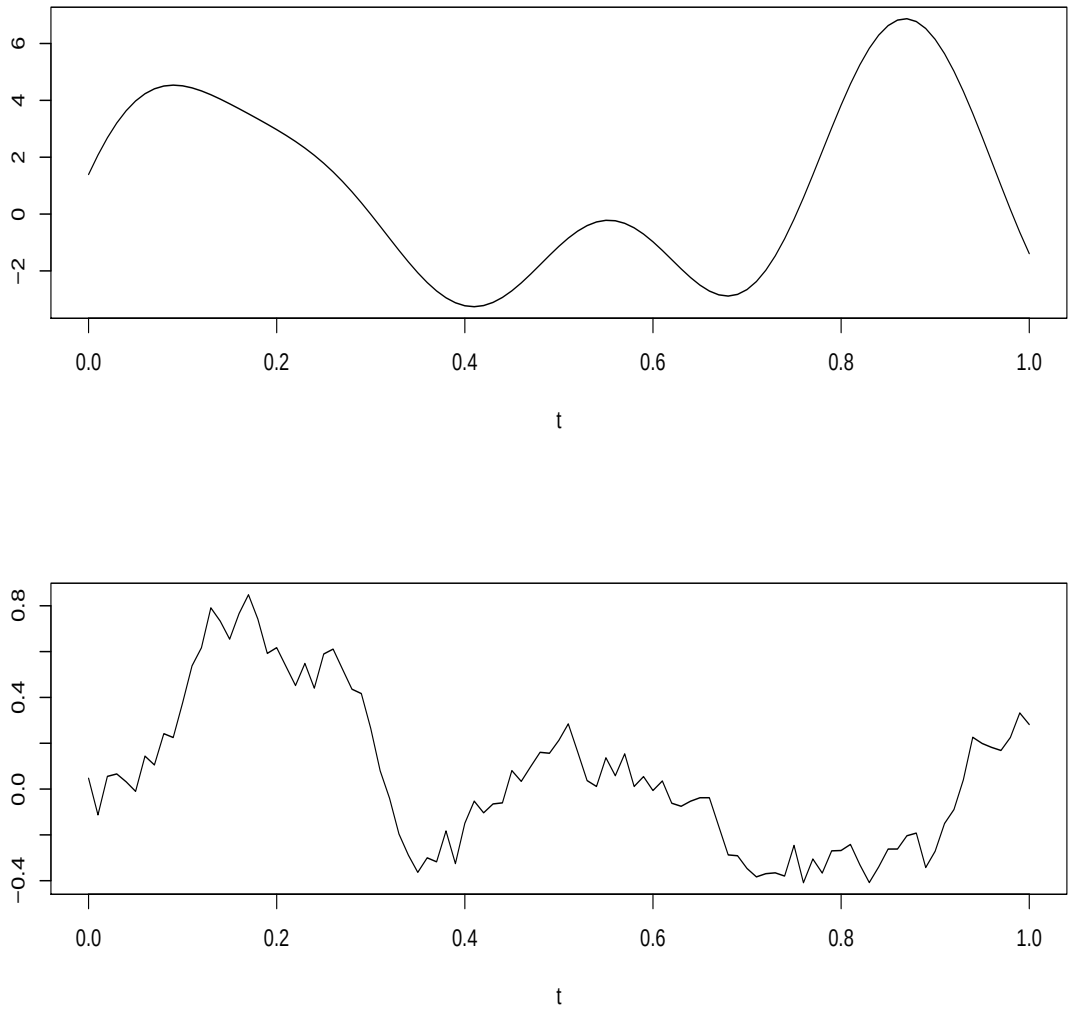

Figure 1. Illustrations of the Three-Component Process and Brownian Motion. The upper and lower panels correspond to trajectories of two-component process and Bm, respectively.

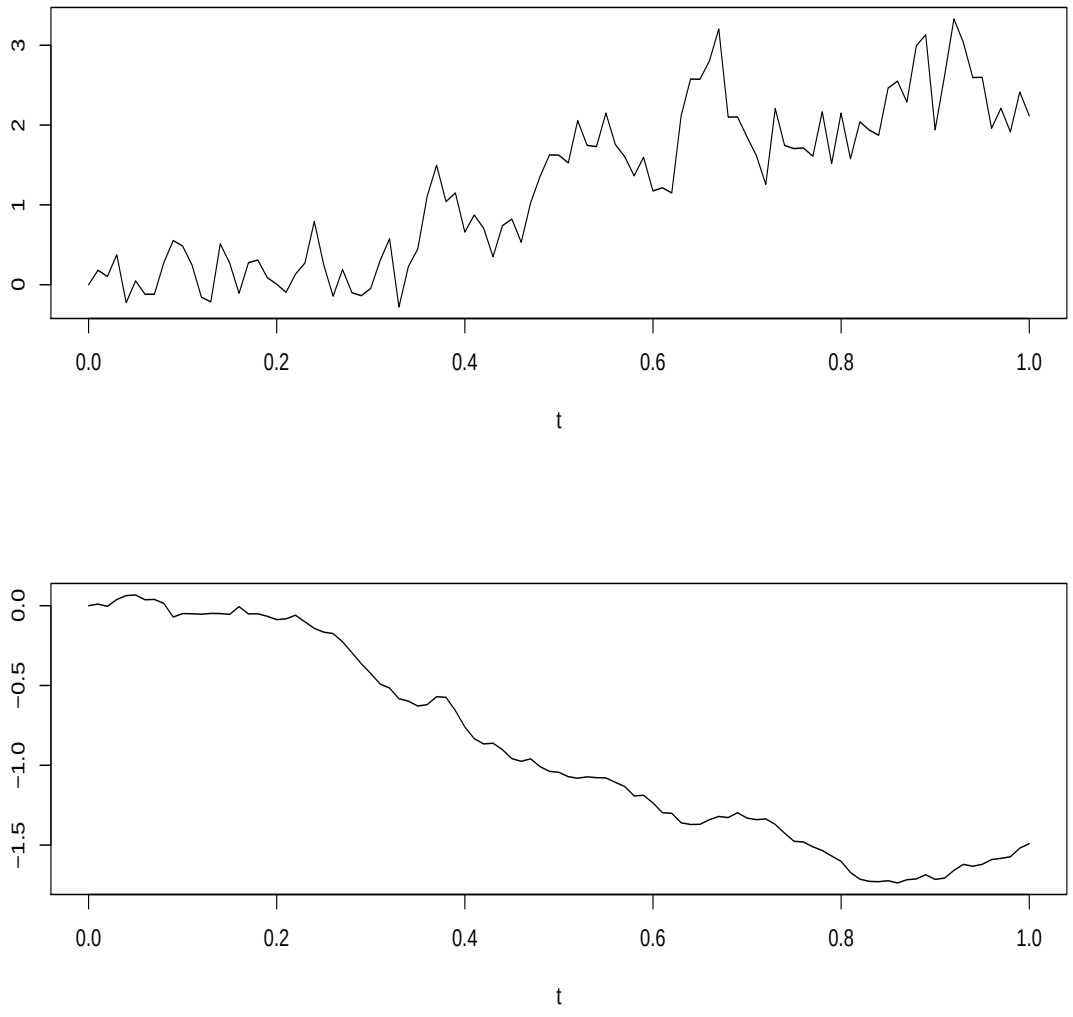

Figure 2. Illustrations of Fractional Brownian Motions. The upper and lower panels correspond to trajectories of fBm ($H = 0.25$) and fBm ($H = 0.75$), respectively.

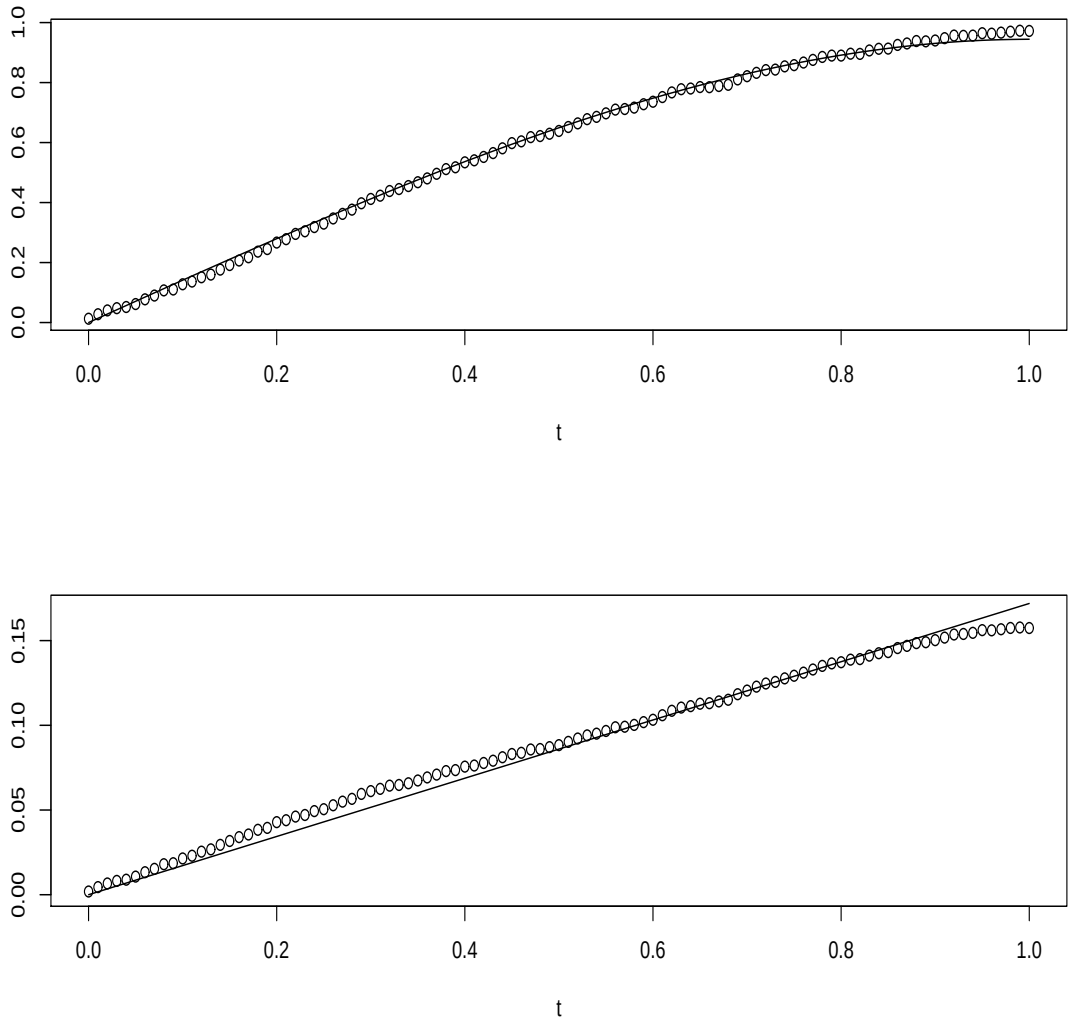

Figure 3. FSIR Estimation for Experiment 1. The upper and lower panels correspond to the Functions in RKHS and L^2 , respectively. In each panel, the solid line stands for the target function and the circled line is the estimation.

Format 2 in Section 6.1, that is, $\{x_i(t_j), j = 1,\ldots,M, i = 1,\ldots,n\}$. Further assume that each $x_i(t_j)$ can be observed according to whether a random variable b_{ij} is 1 or 0, where b_{ij} are i.i.d. Bernoulli(p) distributed. We choose $M = 25$ and $p = 0.8$. The sampled data is shown in Figure 4.

Smoothed FSIR is applied for this data. The normalized estimation in RKHS is shown in the upper panel of Figure 5. We then transform it to the L^2 space and equivalently get the normalized estimate of β, which is shown in the lower panel of Figure 5. Both estimates are compared to the target functions normalized in the corresponding spaces, respectively.

Experiment 3. Let $\{x(t), t \in [0,1]\}$ be the three-component process $S(t)$, $\varepsilon \sim N(0, \sigma^2 = 0.01)$, $f(t) = \exp(t)$ and $\beta(t) = t$, hence

$$h(t) = 0.848826\sin(3\pi t) + 0.286479\sin(5\pi t) + 0.0582052\sin(7\pi t).$$

We generate $n = 100$ data points and each trajectory has been sampled at $q = 51$ equally spaced time points in $[0,1]$. By FSIR the normalized estimation in RKHS is compared with the normalized target function and shown in Figure 6. However, all three transformation methods mentioned in Chapter 5 fail for this case.

Experiment 4. Let $\{x(t), t \in [0,1]\}$ be a fBm ($H = 0.75$), $\varepsilon \sim N(0, \sigma^2 = 0.01)$, $f(t) = \tan^{-1}(t)$ and $\beta(t) = \sin(3\pi t/2)$.

We generate $n = 30$ paths and observing at $q = 25$ locations randomly by the same way as in Experiment 2. Figure 7 shows the result from smoothed FKIR.

Throughout the above numerical experiments, the methods of partial spectral decomposition and generalized inverse of covariance matrix mentioned in Chapter VI are extensively applied to the principal component analysis (see formula (4.1)) and the transformation stages.

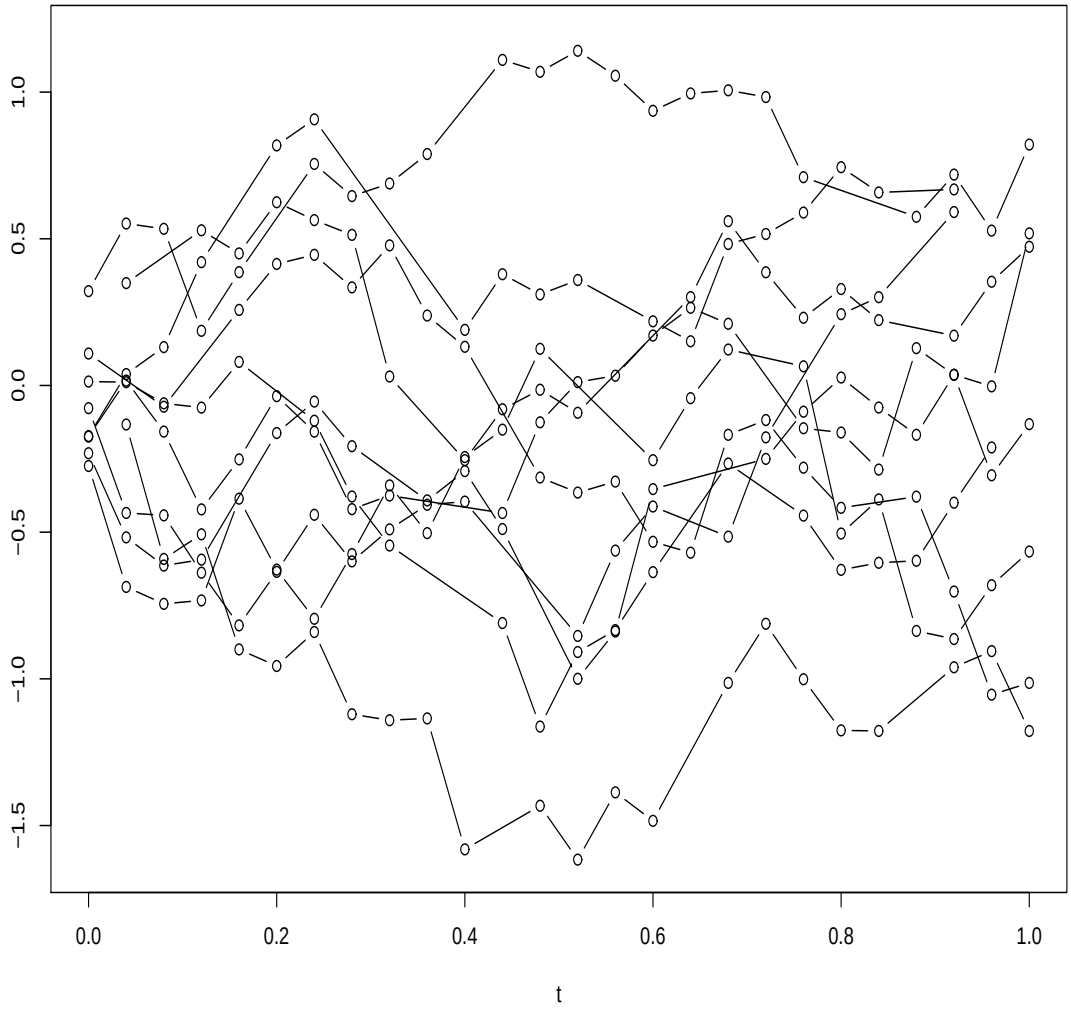

Figure 4. The Randomly Sampled Bm Trajectories in Experiment 2. The circles represent the sampled points.

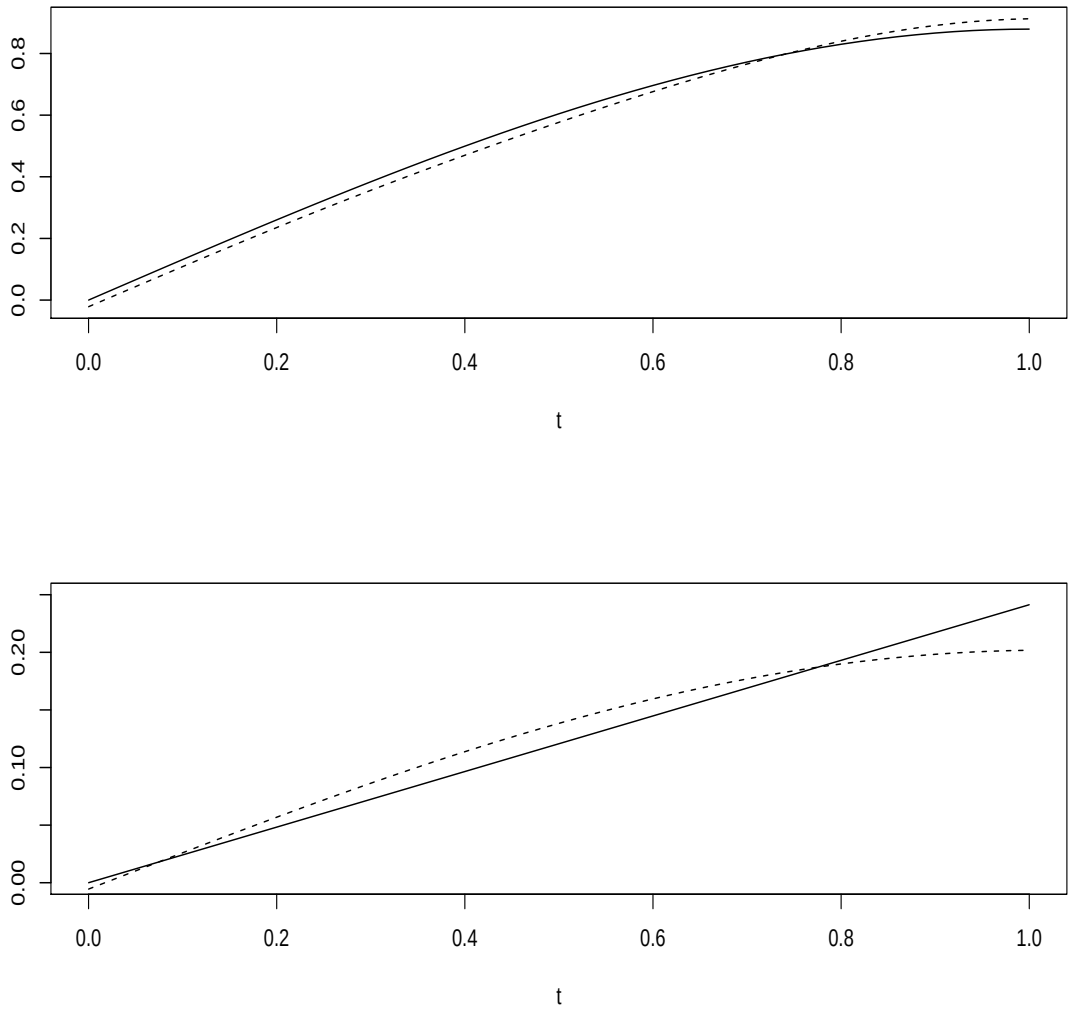

Figure 5. Smoothed FSIR Estimation for Experiment 2. The upper and lower panels correspond to the Functions in RKHS and L^2 , respectively. In each panel, the solid line stands for the target function and the dotted line is the estimation.

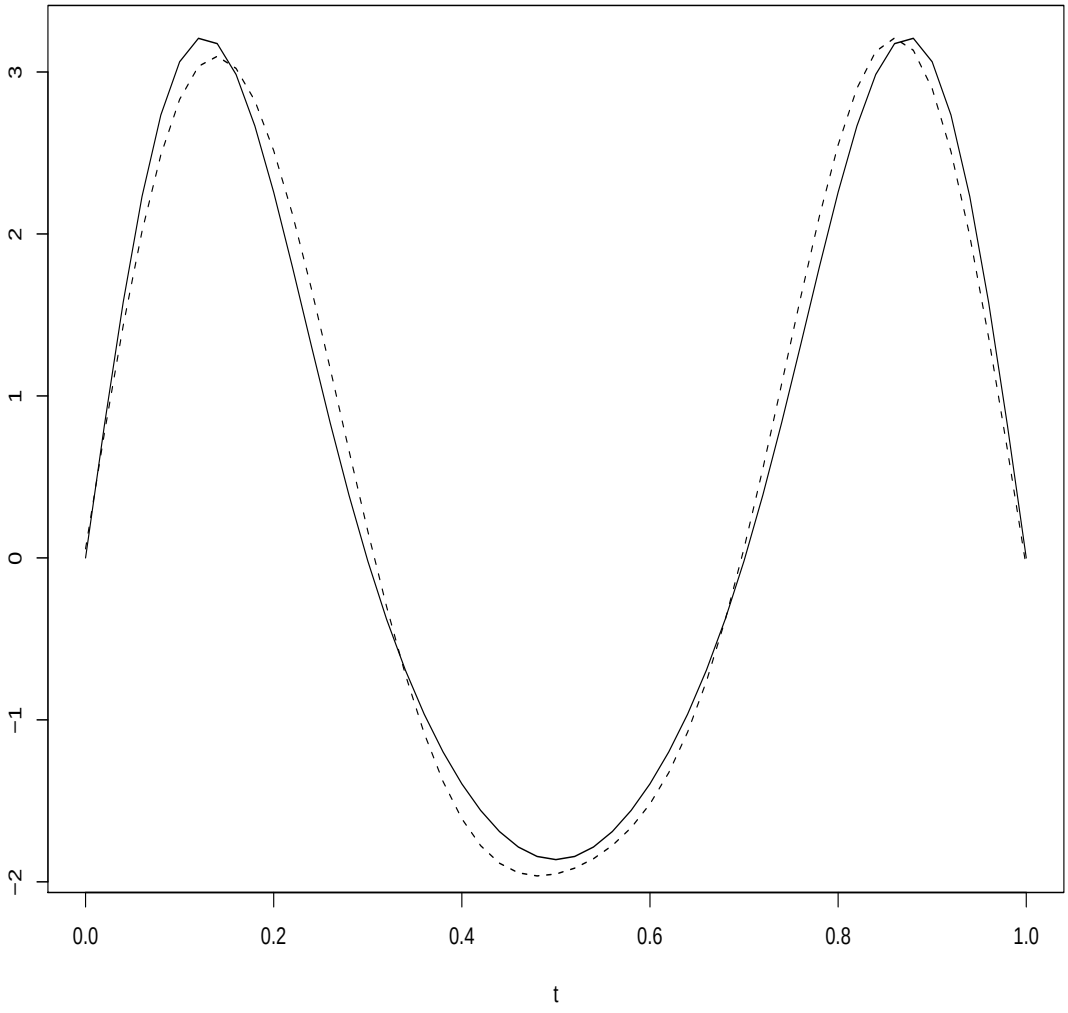

Figure 6. FSIR Estimation for Experiment 3. The solid line stands for the target function and the dotted line is the estimation.

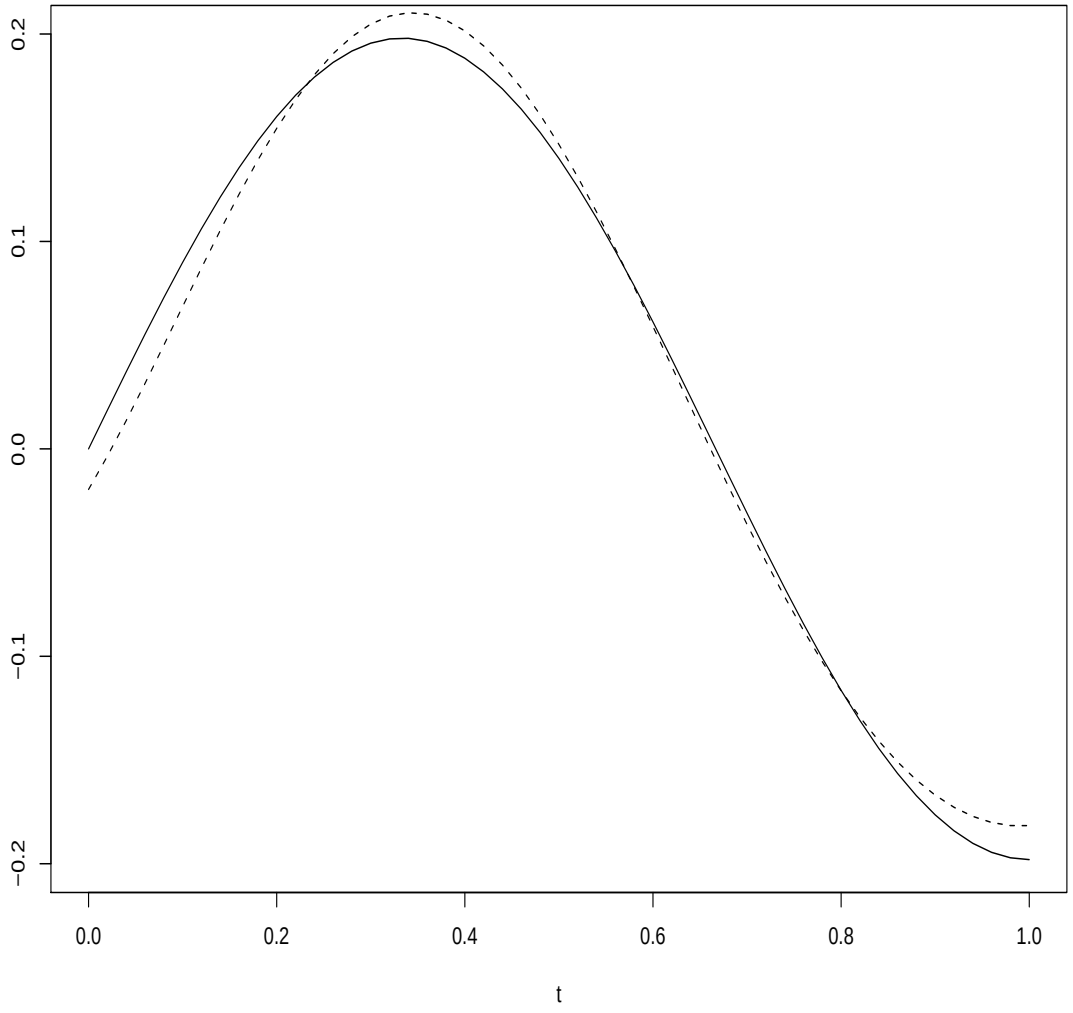

Figure 7. Smoothed FKIR Estimation for Experiment 4. The solid line stands for the target function and the dotted line is the estimation.

For two-index model, we consider

$$Y = \exp(\int_0^1 \sin(3\pi t/2)x(t)dt) + \exp(|\int_0^1 \sin(5\pi t/2)x(t)dt|) + \varepsilon,$$

where $\{x(t), t \in [0,1]\}$ is a standard Bm and $\varepsilon \sim N(0, \sigma^2 = 0.1)$. Generating $n = 500$ paths and sampling at $q = 100$ same locations for each path. The FSIR detects two indices since the first two eigenvalues from the principal component analysis take 98% of the sum of all eigenvalues. Figure 8 shows The estimated functions with the target functions, which are all normalized.

7.2 Data Analysis

The stochastic multiple-index model with IR could be extensively applied to many areas. Here we illustrate two examples, one is in biological science, another is from spectroscopic chemistry.

7.2.1 Application in Biology — Analysis of Medfly Data

A series of papers including Chiou et al. (2003) and Müller and Stadtmüller (2005) were analyzing the medfly data via a various of FDA approaches. One of the purposes of the medfly studies is to investigate the relationship of longevity and reproduction. There were hundreds of female medflies in the experiment. For each medfly, the reproduction trajectory was collected by counting the daily eggs it laid, and the lifetime was also recorded.

In the first analysis, we want to model the relationship between the lifetime in days (y) and the reproduction trajectory (x) by a functional multiple-index model based on the sample $\{(x_i, y_i), i = 1, \ldots, 534\}$, where $x_i(t)$=number of eggs laid on the tth day by the ith medfly, $t = [0, 30]$. To illustrate the data curves, we randomly selected 30 paths from 534 medflies and show them in Figure 9, where the circled line indicates one typical path.

FKIR is implemented and two indices are detected. The first two eigenvalues take

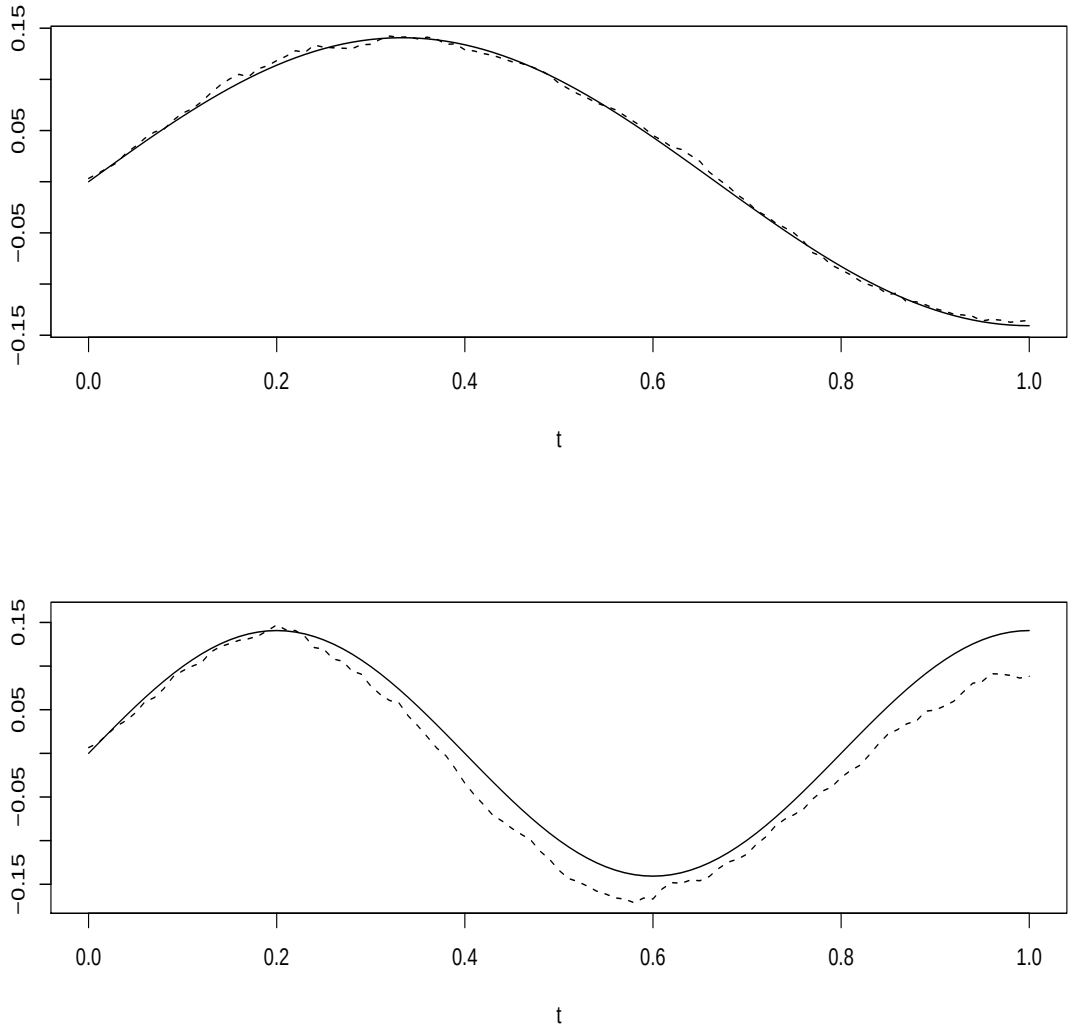

Figure 8. Estimated Index Coefficient Functions for a Two-Index Model. The upper and lower panels correspond to two indices, respectively. In each panel, the solid line is the true function and the dotted line is the estimation.

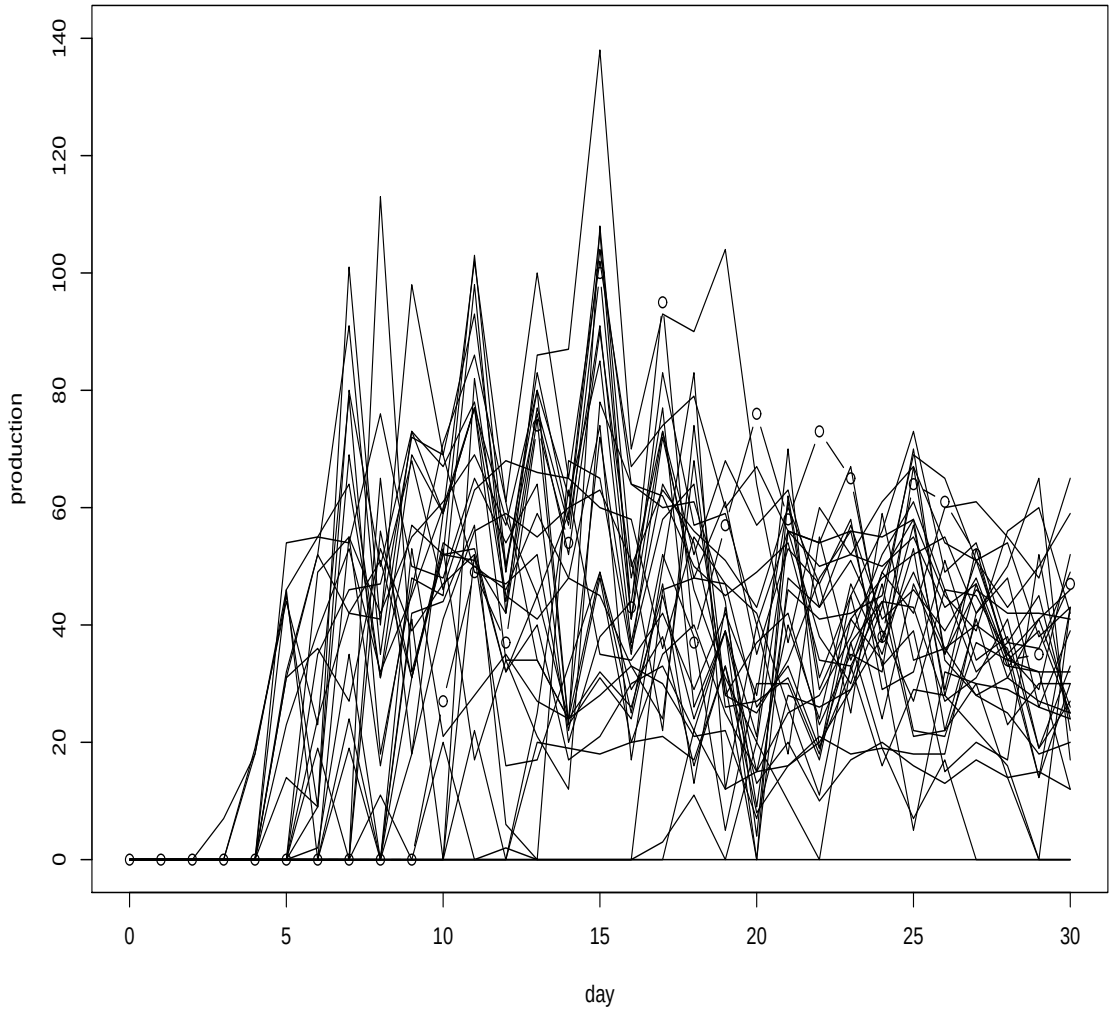

Figure 9. The Reproductive Trajectories of 30 *Medflies. The circled line indicates one path.*

85% of the total. Two coefficient functions are illustrated in Figure 10. We then estimate the link function by two-dimensional local linear regression and it is shown in Figure 11.

In the second analysis, following the approach of Müller and Stadtmüller (2005), we define the following long- or short- lived indicator

$$z = I(y \geq 44 \text{ days }) = \begin{cases} 1, & \text{long} - \text{lived} \\ 0, & \text{short} - \text{lived} \end{cases}.$$

Since z is binary, FSIR with two slices would be a direct approach. The principal component analysis shows that the first eigenvalue takes 98% of the sum of all eigenvalues, which indicates a significant single-index model. The index coefficient function is drawn in Figure 12.

7.2.2 Application in Chemometrics—Analysis of Tecator Data

Goutist (1998) and Ferraty and Vieu (2002) introduced FDA to study Spectrometric data. Spectrometrics is an important branch in Chemometrics, which uses infrared spectroscopy for the structural analysis of organic and inorganic compounds by assigning absorption bands to fundamental vibrations of the investigated molecule.

The Tecator data are recorded by the Tecator Infratec food and feed analyzer which is a near-infrared spectrometer. Each food sample contains finely minced pure pork meat with different contents of fat, protein and moisture. During the experiment, the spectrometer measured the spectrum of light transmitted through the sample in the region $850 - 1050$ nanometers (nm). For each meat sample, the data consists of a 100 channel spectrum of absorbances and the contents of fat, protein and moisture. The absorbance is transformed to $-log_{10}$ of its original value. Since the contents of fat, protein and moisture are recorded in percentages, we take the normalized transformation $log(\frac{a}{1-a})$ for all of these contents, and still call them the contents in the following context. The sample size in this analysis is 172.

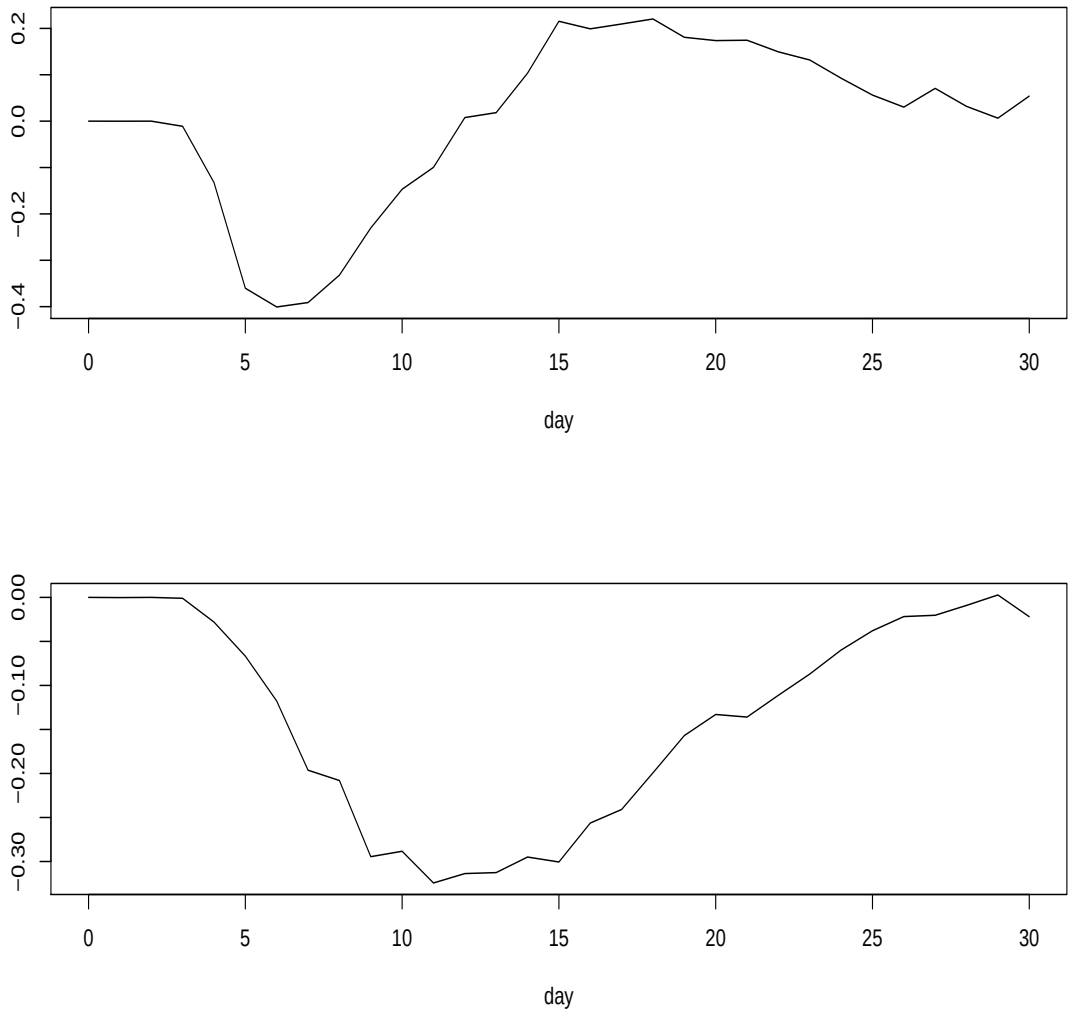

Figure 10. The Index Coefficient Functions in Modeling the Lifetime of Medflies. The upper and lower panels correspond to the first and second indices, respectively.

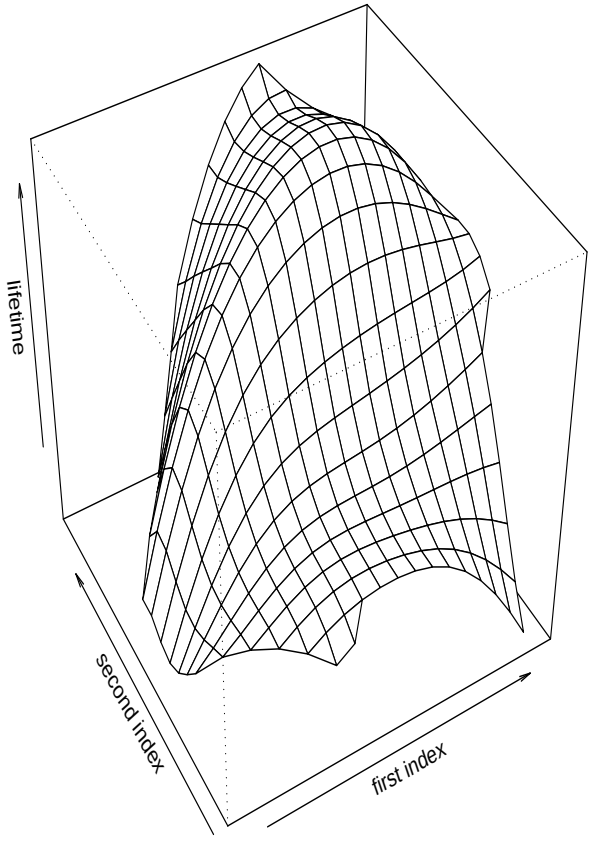

Figure 11. The Estimated Link Function in Modeling the Lifetime of Medflies.

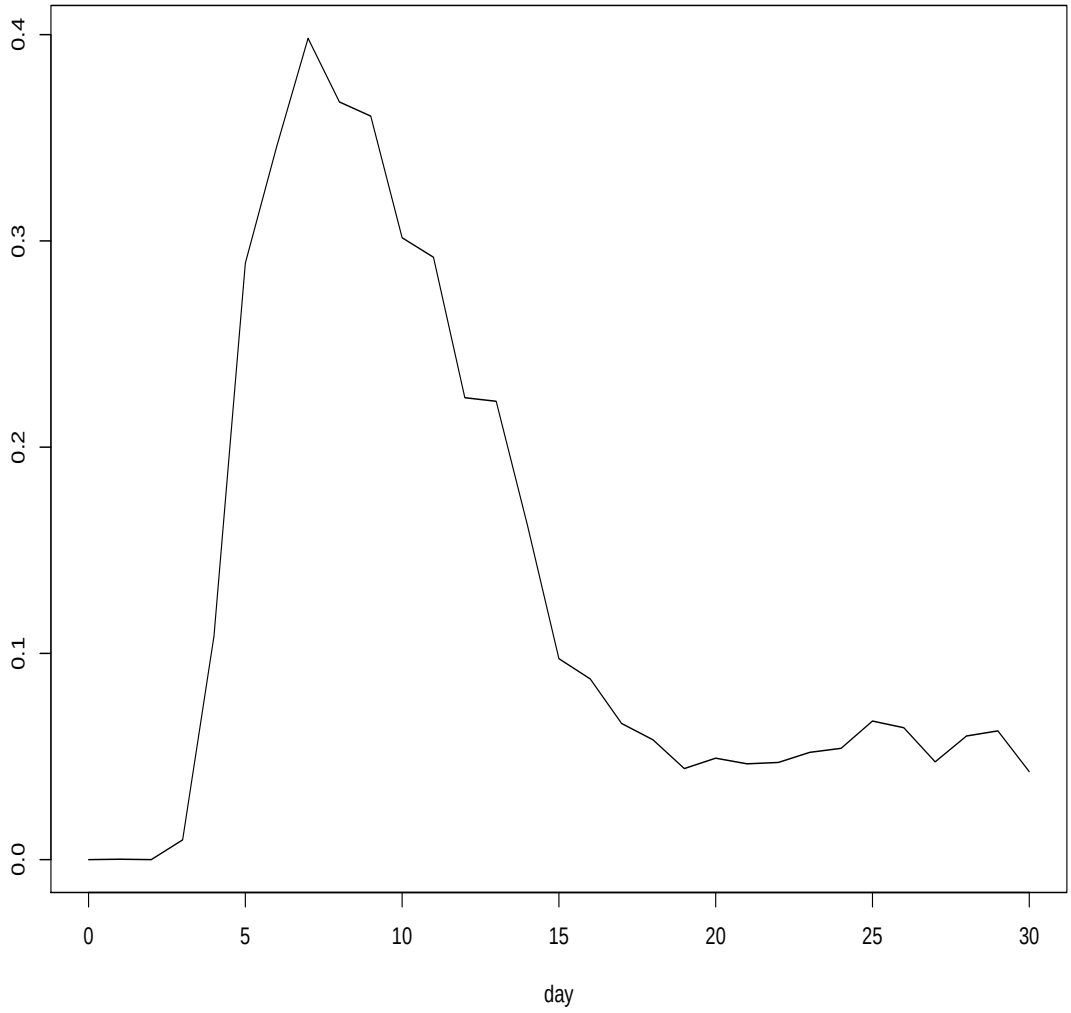

Figure 12. The Single Index Coefficient Function in Modeling Lifetime Indicator of Medflies.

Let x = the trajectory of channel spectrum of absorbance, we randomly selected 30 paths from all 172 paths and show them in Figure 13. In this subsection we intend to model y = content of moisture with x by the functional multiple-index model

$$y = f(\int \beta_1(t)x(t)dt, \cdots, \int \beta_k(t)x(t)dt, \varepsilon).$$

The principal component analysis in FSIR detects a two-index model with the eigenvalues take 69.5% and 26.4% of sum of all eigenvalues, respectively. The first and second index coefficient function are displayed on Figure 14. We estimate the response by fitting a two-dimensional local linear regression. The comparison of the estimated and the observed response values is presented in the left panel of Figure 15 with correlation coefficient 0.83. The boxplot of the error is displayed in the right panel, which passes the Shapiro-Wilk normality test with p-value 0.14.

7.3 IR of Hybrid Data

In Chapter 8 of Ramsay and Silverman (1997), the principal component analysis was developed to access the mixed data or hybrid data which includes both a vector and a curve observed on each individual. In functional regression-type studies, this hybrid data analysis problem is motivated by the situation where the exploratory variables contain both functional curves and multivariate covariates. Hence the following hybrid multiple-index model is proposed:

$$y = f(\mathbf{a_1}'\mathbf{z} + \int_0^1 \beta_1(t)x(t)\,\mathrm{d}t, \ldots, \mathbf{a_p}'\mathbf{z} + \int_0^1 \beta_p(t)x(t)\,\mathrm{d}t, \varepsilon), \tag{7.2}$$

where x is a L^2 Gaussian process on $[0,1]$, $\mathbf{z}$ is d-dim normally distributed variable, $\{\mathbf{a_1}, \ldots, \mathbf{a_p}\} \subset \mathbf{R}^d$ and $\{\beta_1, \ldots, \beta_p\}$ are functions on $[0,1]$.

It is interesting that by re-defining the index set, $T = [0,1] \cup \{2, \ldots, d+1\}$, associated with the extended process, $x(i) = z_i, i = 2, \ldots, d+1$, model (7.2) is not new, but a special

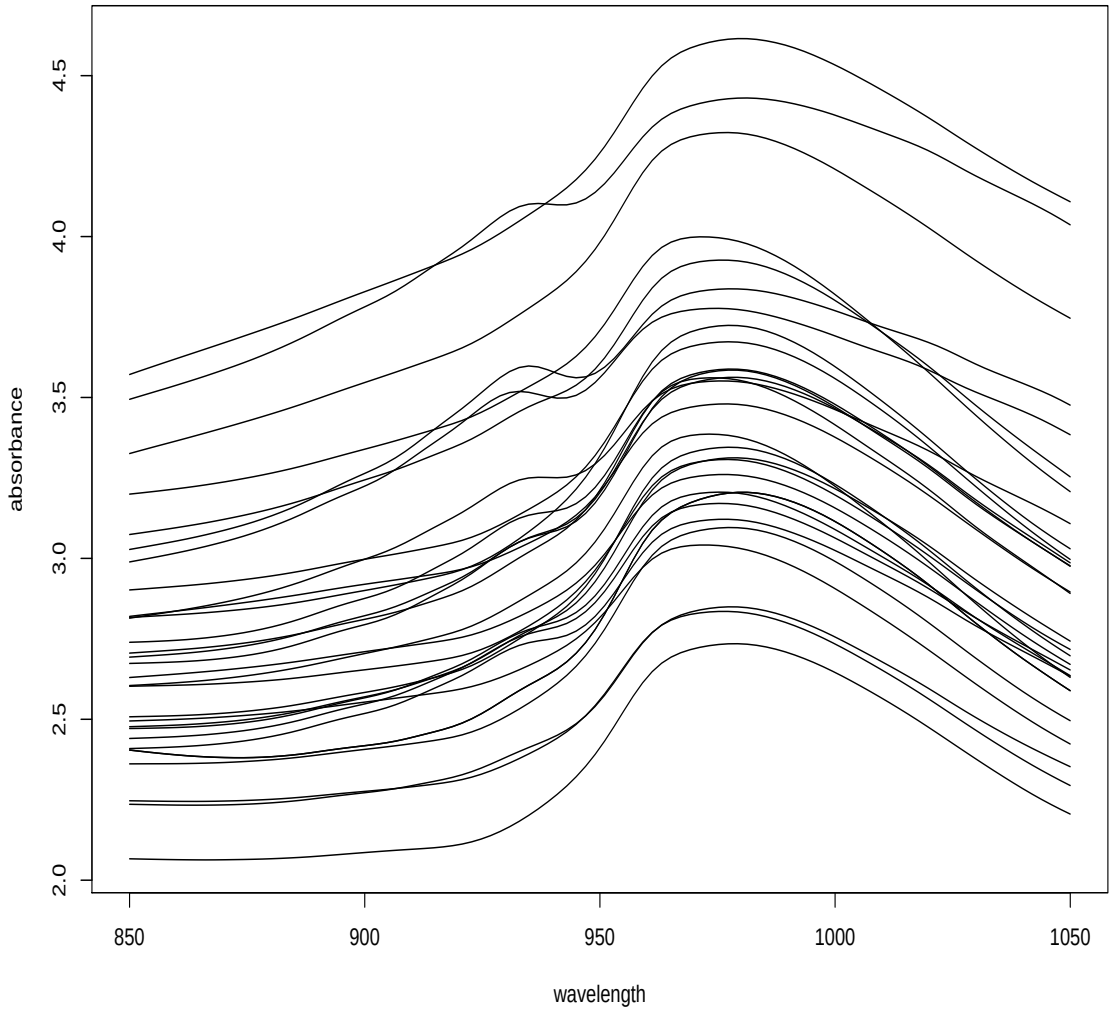

Figure 13. The Spectra of 30 *Selected Samples.*

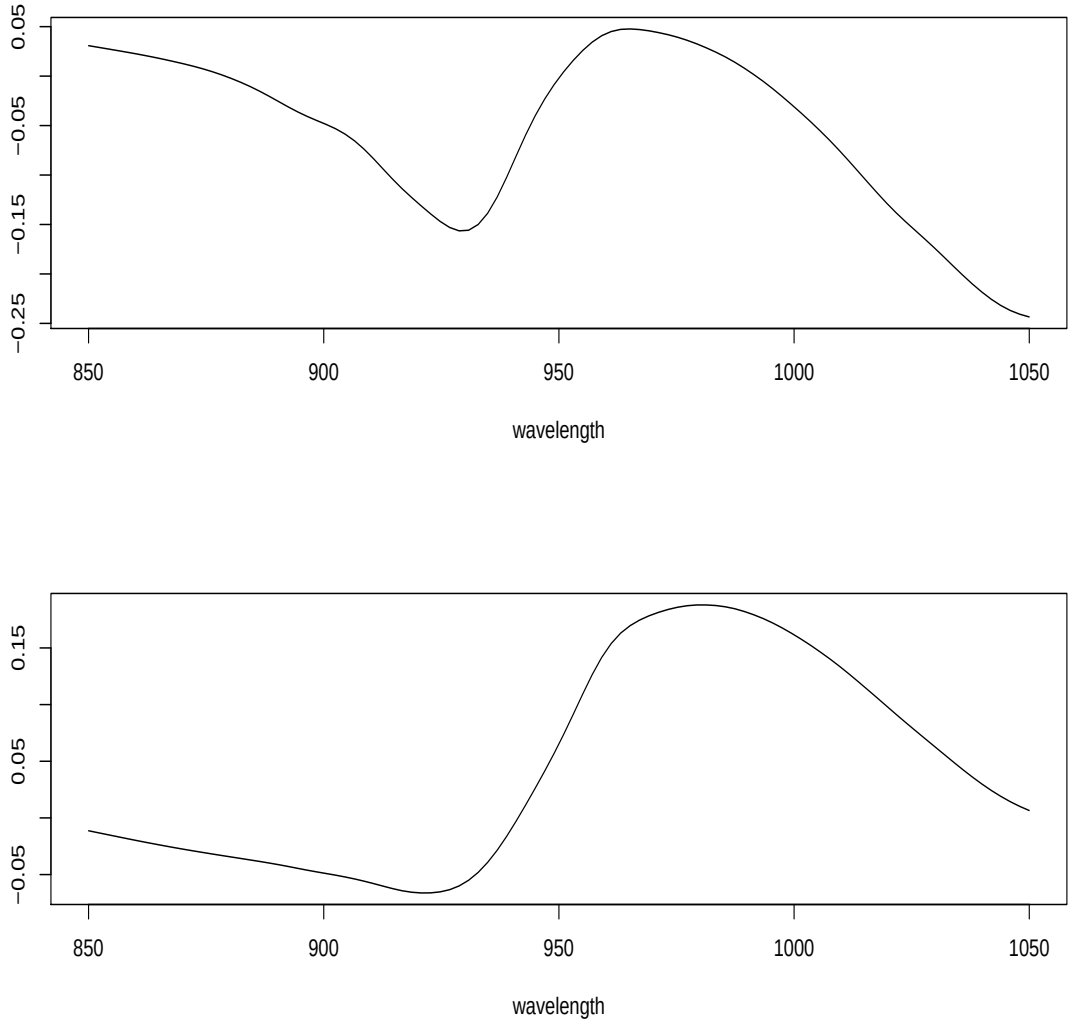

Figure 14. The Index Coefficient Functions in Modeling the Tecator Data. The upper and lower panels correspond to the first and second indices, respectively.

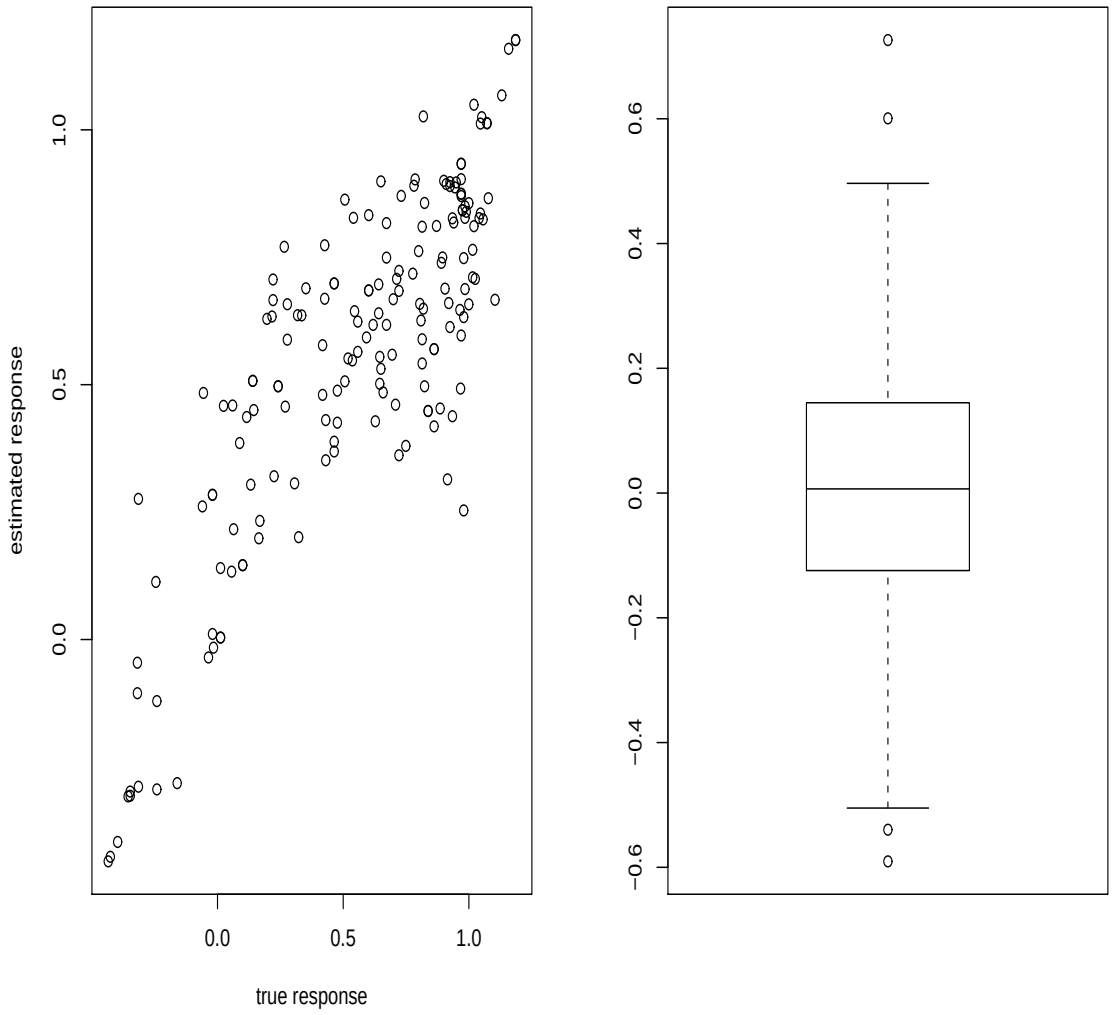

Figure 15. The Estimation in Modeling the Tecator Data. The left and right panels correspond to the comparison of the estimated and observed response values and the boxplot of the error, respectively.

case of the general stochastic multiple-index model (3.1). This makes IR work directly for hybrid data.

Assuming that the hybrid data are collected as

$$D = \left\{ \left(\{z_{i1}, \ldots, z_{id}\}, \{x_{ij}, j = 1, \ldots, m_i\}, y_i \right), i = 1, \ldots, n \right\},$$

we can propose the following IR algorithm:

Step 1. Smooth each curve based on $\{x_{ij}, j = 1, \ldots, m, i = 1, \ldots, n\}$ to get

$$D' = \left\{ \left(\{z_{i1}, \ldots, z_{id}\}, \{\hat{x}_{ij}, j = 1, \ldots, m\}, y_i \right), i = 1, \ldots, n \right\},$$

and denote it as

$$D' = \{ (\{s_{ik}, k = 1, \ldots, m+d\}, y_i), i = 1, \ldots, n \},$$

where

$$s_{ik} = \begin{cases} z_{ik}, & \text{if } k = 1, \ldots, d \\ \hat{x}_{i,k-d}, & \text{if } k = d+1, \ldots, d+m. \end{cases}$$

Step 2. Implement FSIR or FKIR with the back-transformation on D', get the estimated number of indices $\hat{p}$ and the estimated eigenvectors

$$\hat{\mathbf{b}}_i = (b_{i,1}, \ldots, b_{i,m+d})', \text{ for } i = 1, \ldots, \hat{p}.$$

Step 3. Normalize $(b_{i,1}, \ldots, b_{i,d})'$ and $(b_{i,d+1}, \ldots, b_{i,m+d})'$ to estimate the coefficients vector and the coefficient functions, respectively, for $i = 1, \ldots, \hat{p}$.

To demonstrate how IR works for hybrid data, we consider the single-index model

$$y = f(\mathbf{a}'\mathbf{z} + \int_0^1 \beta(t) x(t)\, \mathrm{d}t) + \varepsilon,$$

where $\mathbf{a}' = (2,1)$, $\mathbf{z}$ is a standard two dimensional normal vector, x is a standard Bm, $\beta(t) = t$, $f = \tan^{-1}$ and $\varepsilon \sim N(0, 0.1)$.

We generate $n = 1000$ data points and each trajectory has been sampled at $q = 101$ equally spaced time points in $[0,1]$. We randomly selected half of the data ($n = 500$) as the training set to estimate $\mathbf{a}$ and β, $\hat{\mathbf{a}}$ and $\hat{\beta}$, respectively, then we use the rest half of the data to validate the estimation by comparing the true index $\xi = \mathbf{a}'\mathbf{z} + \int_0^1 \beta_(t)x(t)$ and the predicted index $\hat{\xi} = \hat{\mathbf{a}}'\mathbf{z} + \int_0^1 \hat{\beta}_(t)x(t)$. Among 100 replicates, there are 36 times that $\text{corr}(\xi, \hat{\xi}) > 0.95$, we illustrate one plot of ξ versus $\hat{\xi}$ in Figure 16, in this case, $\text{corr}(\xi, \hat{\xi}) = 0.9989850$ and $\hat{\mathbf{a}}' = (0.9140922, 0.4055063)$, which is close to the normalized target $(0.8944272, 0.4472136)$.

As an application of this hybrid IR, we continue analyze the Tecator data. Denote z_1 = content of fat and z_2 = content of protein, we expect to find the relationship between y=content of moisture and x = the trajectory of channel spectrum of absorbance with covariates z_1 and z_2 by the functional multiple-index model

$$y = f(a_1^{(1)}z_1 + a_2^{(1)}z_2 + \int \beta_1(t)x(t), \cdots, a_1^{(k)}z_1 + a_2^{(k)}z_2 + \int \beta_k(t)x(t), \varepsilon).$$

We find a significant single-index model by FSIR, as the first eigenvalue takes 93% of the summation. The estimation of (a_1, a_2) is $(0.9788277 - 0.2046861)$, and the estimated β is shown in Figure 17. We estimate the link function and response by fitting a one-dimensional local linear regression. We show the estimated link function in Figure 18 which is a nonlinear function. The comparison of the estimated and the observed response values is presented in the left panel of Figure 19 with correlation coefficient 0.99. The boxplot of the error is displayed in the right panel.

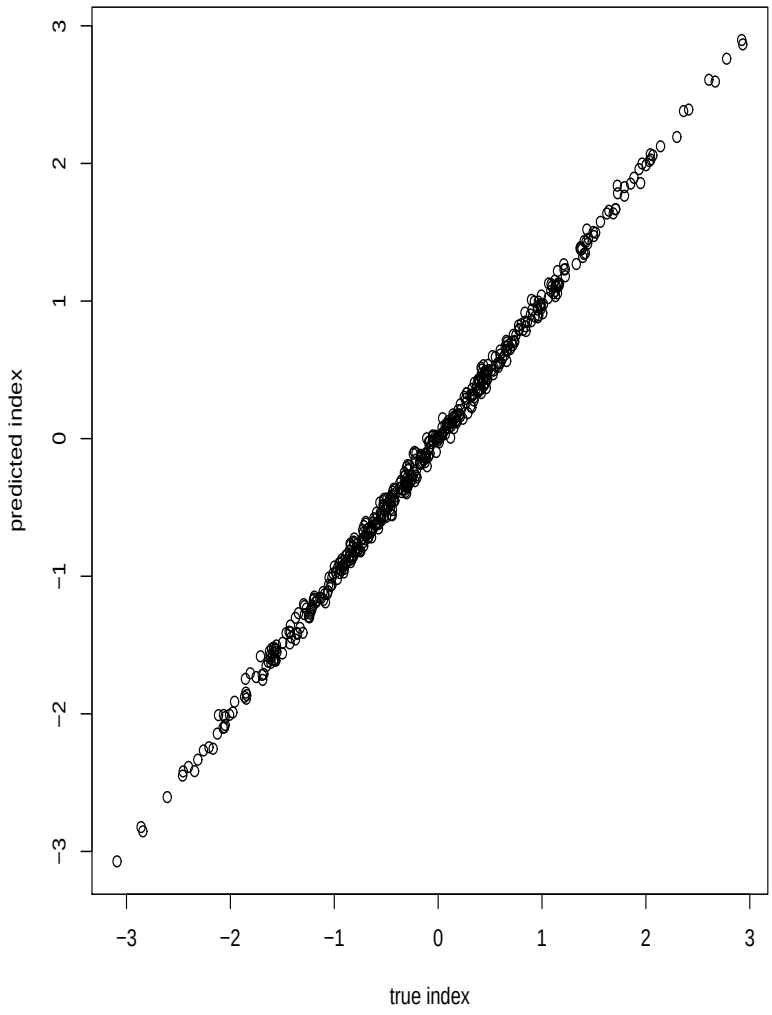

Figure 16. The True Indices Versus the Predicted Indices for the Hybrid Data. The $x-$axis stands for ξ and $y-$axis stands for $\hat{\xi}$.

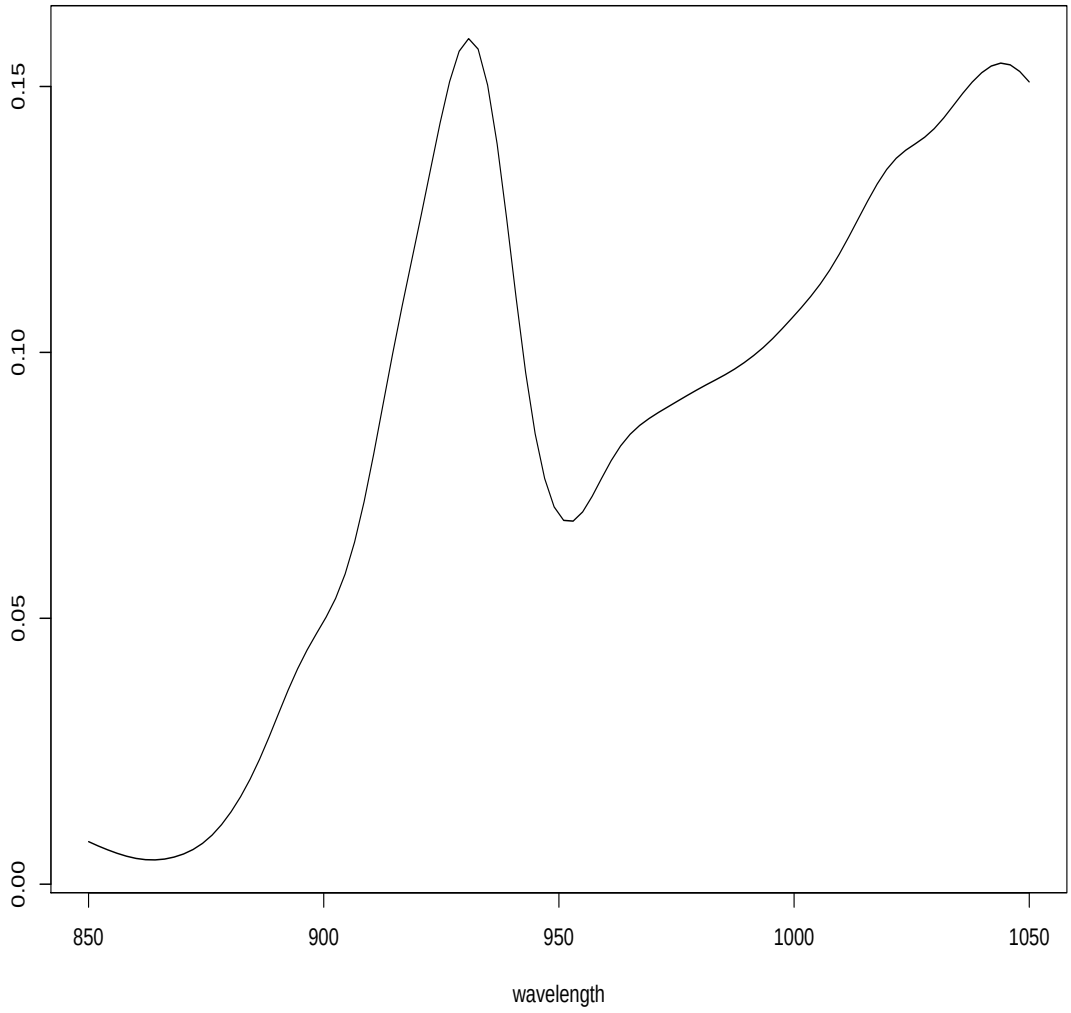

Figure 17. The Single-Index Coefficient Function in Modeling the Hybrid Tecator Data.

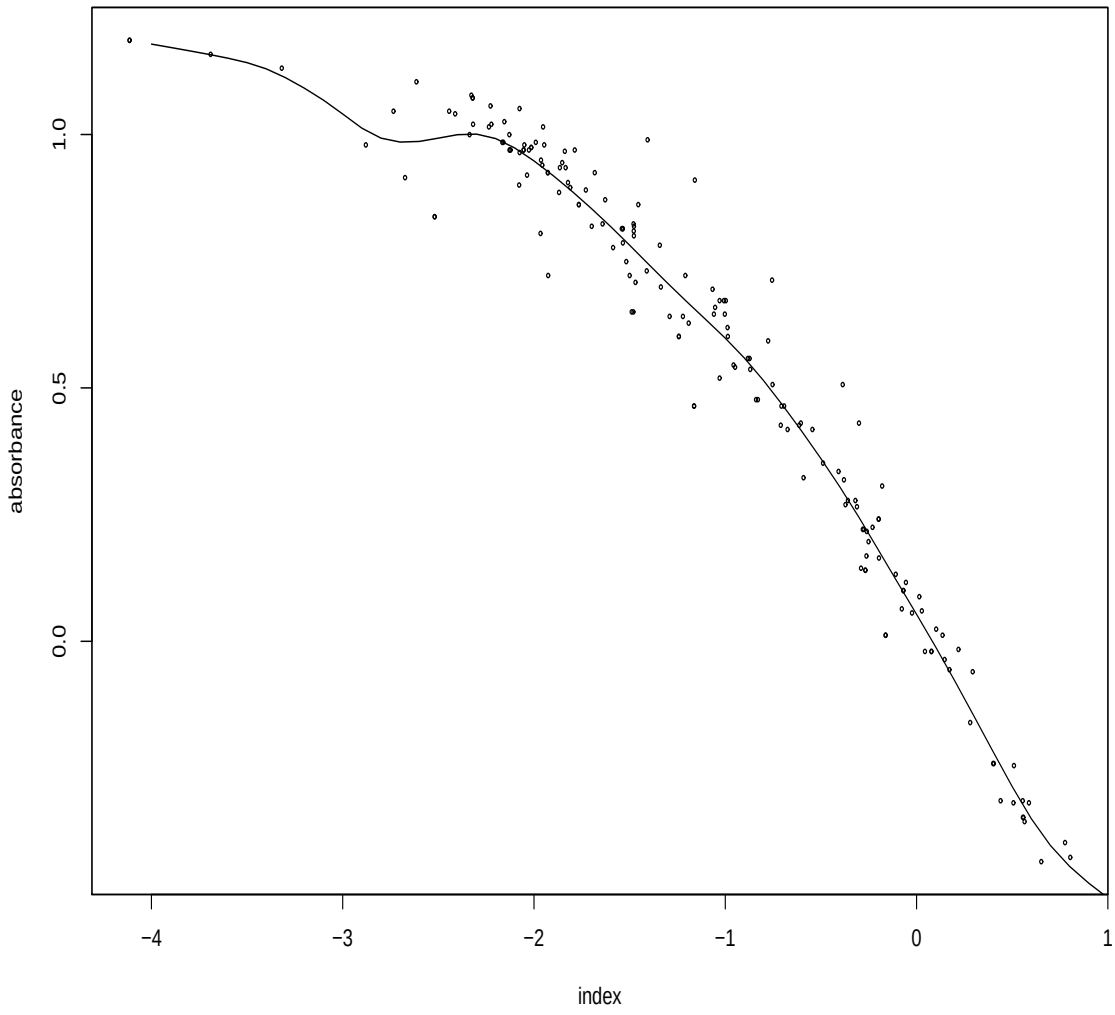

Figure 18. The Estimated Link Function in Modeling the Hybrid Tecator Data. The points are the observations and the solid is the estimated function.

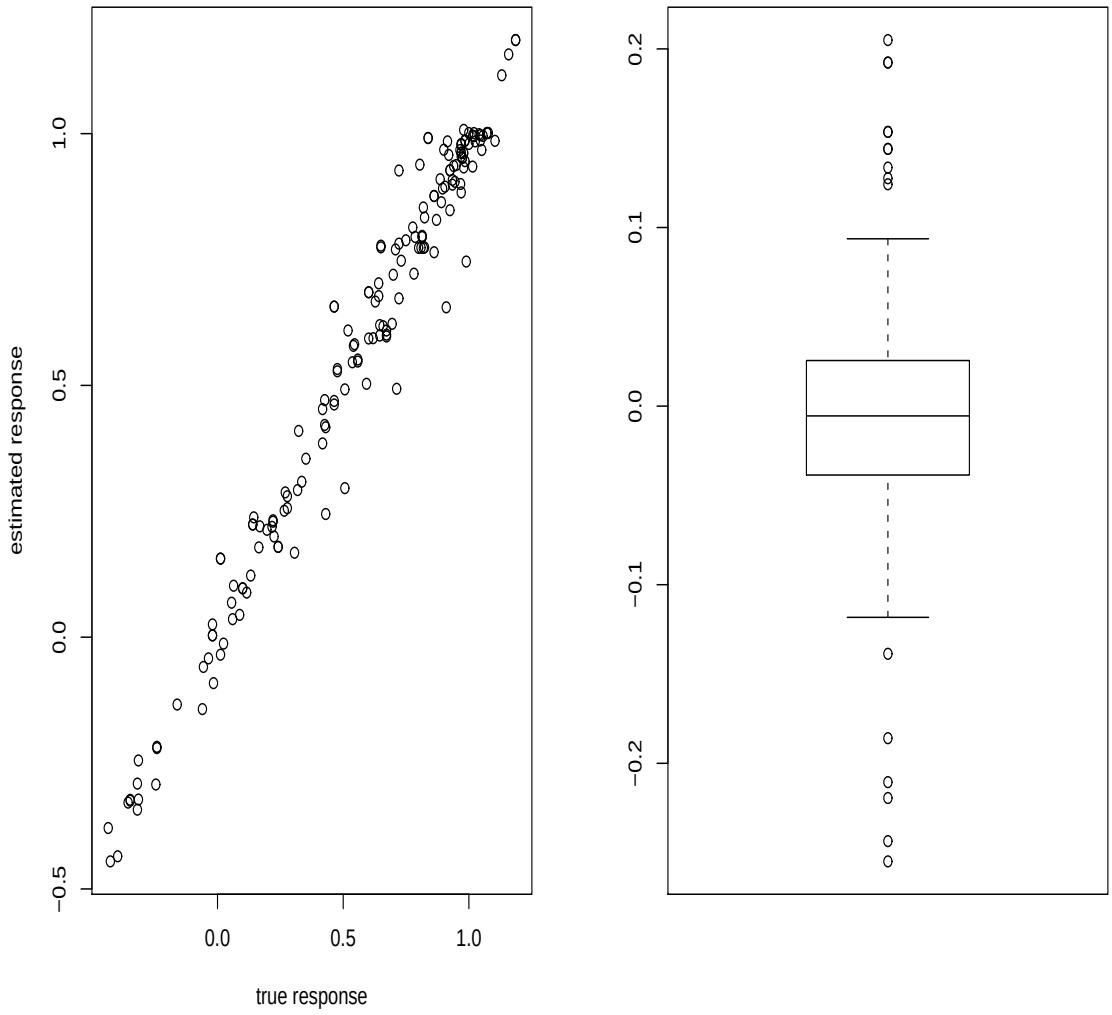

Figure 19. The Estimation in Modeling the Hybrid Tecator Data. The left and right panels correspond to the comparison of the estimated and observed response values and the boxplot of the error, respectively.

CHAPTER VIII

CONCLUSIONS

FDA is still a relatively unexplored and hence exciting area. It brings the traditional statistics to a new era. Theoretically, FDA makes statistics work under infinite-dimensional spaces and unifies stochastic statistics and longitudinal data analysis.

Within the trend of implementation of more statistical techniques to FDA, our research generalizes the seminal inverse regression to a functional setting. We proposed a semiparametric model with multiple indices via an unknown link function related to a second order stochastic process, and a RKHS approach is developed to estimate the EDRS.

Due to the isometric isomorphism between the Hilbert space spanned by the process and the RKHS generated by the covariance function of the process, we discovered and proved Theorem III.1, which reveals the probabilistic or geometrical structure of the multiple-index model. This key result decomposes the estimation of the EDRS into two stages. During the first stage, the EDRS in RKHS is estimated by IR. We proposed FSIR and FKIR procedures implemented in the empirical RKHS. The asymptotic theory associated with these approaches will provide more possible applications of RKHS methods in FDA. The second stage is the back-transformation where we transform the estimated EDRS in RKHS back to the Hilbert space of the space, that is, the original EDRS. Hence, we complete the estimation.

The following comparisons with Li (1991) and Ferré and Yao (2003) are given:

(a) Suppose that X is finite dimensional, recalling the proof of Theorem III.1, the result $E(X|Y) \in \mathcal{H}_X$ a.s. naturally holds without conditions (P1)-(P2). Hence under (IR1) and (IR2), we have $E(X|Y) \in H_{X,e}$ a.s.. Further with the discussion of the back-transformation in Section 5.3, our approach therefore coincides with Li's approach

in that setting, which implies that a generalization from multivariate IR to general IR.

(b) Ferré and Yao (2000) redefines Li's problem in the FDA setting. They considered a special case of ours by assuming that the sample paths of X_t are elements of $L^2[a,b]$ with inner product $\langle f,g\rangle = \int_a^b f(t)g(t)dt$, and that the elements ξ_i are of the form

$$\xi_i = \langle \beta_i, X\rangle,\ 1 \leq i \leq p.$$

(c) The separability of index set, T, is much more general than the settings in both works. In the data analysis, we showed a case study with hybrid data. Since in a metric space compactness implies separability, we can consider the cases where T has more complicated structures.

The determination of number of indices or EDRS dimensions is not included in this research, which remains for future studies. Ferré (1998) is a possible approach, but the computational cost is high. Other issues including the model specification and diagnostics are also valuable topics. With the RKHS approach proposed in this thesis, the multiple-index model can be effectively used to further the study of FDA.

REFERENCES

Amato, U., Antoniadis, A., and Feis, I. D. (2004), "Dimension Reduction in Functional Regression with Applications," *Computational Statistics and Data Analysis*, To Appear.

Aronszajn, N. (1950), "Theory of Reproducing Kernel," *Transactions of the American Mathematical Society*, 68, 337–404.

Ash, R. B., and Gardner, M. N. (1975), *Topics in Stochastic Processes*, New York: Academic Press, INC.

Barton, R. J., and Poor, H. V. (1990), "An RKHS Approach to Robust L^2 Estimation and Signal Detection," *IEEE Transactions on Information Theory*, 36, 485–501.

Becker, C., and Fried, R. (2002), "Sliced Inverse Regression for High-Dimensional Time Series," *Exploratory Data Analysis in Empirical Research*, Eds. Opitz, O. and Schwaiger, M., Berlin: Springer, 3–11.

Berlinet, A., and Thomas–Agnan, C. (2004), *Reproducing Kernel Hilbert Spaces in Probability and Statistics*, Boston: Kluwer Academic Publishers.

Besse, P., Cardot, H., and Ferraty F. (1997), "Simultaneous Nonparametric Regression of Unbalanced Longitudinal Data," *Computational Statistics and Data Analysis*, 24, 255–270.

Bosq, D. (1991), *Linear Process in Function Spaces*, Lecture Notes in Statistics, Vol. 149, Berlin: Springer.

——— (2000), "Modelization, Nonparametric Estimation and Prediction for Continuous Time Processes," *Nata Asi Series C*, 335, 509–529.

Bowman, A. W. and Azzalini, A. (1997), *Applied Smoothing Techniques for Data Analysis: The Kernel Approach with S-Plus Illustrations*, New York: Oxford Science Publications.

Bura, E., and Cook, R. D. (2001), "Estimating the Structural Dimension of Regression via Parametric Inverse Regression," *Journal of the Royal Statistical Society*, Ser. B, 63, 393–410.

Bura, E., and Pfeiffer, R. (2003), "Graphical Methods for Class Prediction Using Dimension Reduction Techniques on DNA Microarray Data," *Bioinformatics*, 19, 1252–1258.

Cai, T. T., and Hall, P. (2005), "Prediction in Functional Linear Regression," *Manuscript*, Department of Statistics, Univeristy of Pennsylvania, PA.

Cardot, H. (2000), "Nonparametric Estimation of Smoothed Principal Components Analysis of Sampled Noisy Functions," *Journal of Nonparametric Statistics*, 12, 502–538.

Cardot, H., Crambes, C., and Sarda, P. (2004), "Conditional Quantiles for Functional Covariates: An Application to Ozone Pollution Forecasting," *COMPSTAT 2004 Proceedings*, Prague, Czech, Ed. Antoch, J., Heidelberg, Germany: Physica-Verlag/Springer, 769–776.

Cardot, H., Ferraty F., Mas, A., and Sarda, P. (2003), "Testing Hypotheses in the Functional Linear Model," *Scandinavian Journal of Statistics*, 30, 241–255.

Cardot, H., Ferraty F., and Sarda, P. (1999), "Functional Linear Model," *Statistics and Probability Letters*, 45, 11–22.

——— (2003), "Spline Estimators for the Functional Linear Model," *Statistica Sinica*, 13, 571–591.

Cardot, H., Goia, A., and Sarda, P. (2004), "Testing for No Effect in Functional Linear Regression Models, Some Computational Approaches," *Communications in Statistics – Simulation and Computation*, 33, 179–199.

Cardot, H., and Sarda, P. (2005), "Estimation in Generalized Linear Models for Functional Data via Penalized Likelihood," *Journal of Multivariate Analysis*, 92, 24–41.

Carroll, R. J., and Li, K. C. (1992), "Measurement Error Regression with Unknown Link:

Dimension Reduction and Data Visualization," *Journal of the American Statistical Association*, 87, 1040–1050.

Chen, C.–H., and Li, K. C. (1998), "Can SIR Be as Popular as Multiple Linear Regression?" *Statistica Sinica*, 8, 289–316.

Chiou, J.–M., Müller, H.–G., and Wang, J.–L. (2003), "Functional Quasi-Likelihood Regression Models with Smooth Random Effects," *Journal of the Royal Statistical Society*, Ser. B, 65, 405–423.

Chiou, J.–M., Müller, H.–G., Wang, J.–L., and Carey, J. R. (2003), "A Functional Multiplicative Effects Model for Longitudinal Data, with Application to Reproductive Histories of Female Medflies," *Statistica Sinica*, 13, 1119–1133.

Cook, R. D. (1998), *Regression Graphics*, New York: Wiley.

Cook, R. D., and Weisberg, S. (1991), Comment on "Sliced Inverse Regression," by K. C. Li, *Journal of the American Statistical Association*, 86, 328–332.

Crambes, C. (2005), "Total Least Squares for Functional Data," *ASMDA Brest 2005 Proceedings*, France, 619–626.

Cueva, A., Febrero M., and Fraiman, R. (2002), "Linear Functional Regression: The Case of Fixed Design and Functional Response," *Scandinavian Journal of Statistics*, 30, 285–300.

Dauxois, J., Pousse, A., and Romain, Y. (1982), "Asymptotic Theory for the Principal Component Analysis of a Vector Random Function: Some Applications to Statistical Inference," *Journal of Multivariate Analysis*, 12, 136–154.

Diggle, P. J., and Verbyla, A. P. (1998), "Nonparametric Estimation of Covariance Structure in Longitudinal Data," *Biometrics*, 54, 401–415.

Donkers, B., and Schafgans, M. (2003), "A Derivative Based Estimator for Semiparametric Index Models," *Discussion Paper*, No. 22, Tilburg, Netherlands: Tilburg University.

Driscoll, M. F. (1973), "The Reproducing Kernel Hilbert Space Structure of the Sam-

ple Paths of a Gaussian Process," *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 26, 309–316.

Duan, N., and Li, K. C. (1991), "Slicing Regression: A Link-Free Regression Method," *The Annals of Statistics*, 19, 505–530.

Escabias, M., Aguilear, A. M., and Valderrama, M. J. (2004), "Principal Component Estimation of Functional Logistic Regression: Discussion of Two Different Approaches," *Nonparametric Statistics*, 16, 365–384.

Eubank, R. (1999), *Spline Smoothing and Nonparametric Regression, 2nd Edition*, New York: Dekker.

Eubank, R., and Hsing, T. (2005), "Canonical Correlation for Stochastic Processes," *Manuscript*, Department of Statistics, Ohio State University, Columbus, OH.

Fan, J., and Zhang J.–T. (2002), "Two-Step Estimation of Functional Linear Models with Applications to Longitudinal Data," *Journal of the Royal Statistical Society*, Ser. B, 62, 303–322.

Ferraty, F., and Vieu, P. (2002), "The Functional Nonparametric Model and Application to Spectrometric Data," *Computational Statistics*, 17, 545–564.

——— (2004), "Nonparametric Models for Functional Data, with Application in Regression, Time-series Prediction and Curve Discrimination," *Nonparametric Statistics*, 16, 111–125.

Ferré, L. (1997), "Dimension Choice for Sliced Inverse Regression Based on Ranks," *Student*, 2, 95–108.

Ferré, L. (1998), "Determining the Dimension in Sliced Inverse Regression and Related Methods," *Journal of the American Statistical Association*, 93, 132–140.

Ferré, L., and Yao, A. F. (2003), "Functional Sliced Inverse Regression Analysis," *Statistics*, 37, 475–488.

Fine, J. (2003), "Asymptotic Study of Canonical Correlation Analysis: From Matrix and

Analytic Approach to Operator and Tensor Approach," *SORT*, 27, 165–174.

Fung, W. K., He, X., Liu, L., and Shi, P. (2002), "Dimension Reduction Based on Canonical Correlation," *Statistica Sinica*, 12, 1093–1113.

Gannoun, A., Girard, S., Guinot, C., and Saracco, J. (2004), "Reference Curve Estimation via Alternating Sliced Inverse Regression," *Environmetrics*, 15, 81–99.

Gannoun, A., and Saracco, J. (2003), "An Asymptotic Theory for SIR_{α} Method," *Statistica Sinica*, 13, 297–310.

Gasser, T., and Kneip, A. (1995), "Searching for Structure in Curve Samples," *Journal of the American Statistical Association*, 90, 1179–1188.

Gather, U., Hilker, T., and Becker, C. (2001), "A Robustified Version of Sliced Inverse Regression Procedures," *Statistics in Genetics and in the Environmental Sciences*, Eds. Fernholz, L. T. Morgenthaler, S. and Stahel, W., Berlin: Birkhäuser, 147–157.

——— (2002), "A Note on Outlier Sensitivity of Sliced Inverse Regression," *Statistics*, 36, 271–281.

Goia, A. (2003), "Selection Model in Functional Linear Regression Models for Scalar Response," *Developments in Applied Statistics*, Eds. Feligoj, A and Mrvar, A., Ljubliana, Slovenia: FDV, 97–110.

Goutis, C. (1998), "Second-Derivative Functional Regression with Applications to Near Infra-Red Spectroscopy," *Journal of the Royal Statistical Society*, Ser. B, 60, 103–114.

Green, P. J., and Silverman, B. W. (1994), *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach*, London: Chapman and Hall.

Gu, C. (2002), *Smoothing Spline ANOVA Models*, New York: Springer.

Guo, W. (2002), "Functional Mixed Effects Models," *Biometrics*, 89, 145–158.

Hall, P., and Horowitz, J. L. (2004), "Methodology and Convergence Rates for Functional Linear Regression," *Manuscript*, Department of Economics, Northwestern University,

Chicago, IL.

Hall, P., and Li, K. C. (1993), "On Almost Linearity of Low Dimensional Projections from High Dimensional Data," *The Annals of Statistics*, 21, 867–889.

Hastie, T., Buja, A., and Tibshirani, R. (1995), "Penalized Discriminant Analysis," *The Annals of Statistics*, 23, 73–102.

Härdle, W., and Stocker, T. M. (1989), "Investigating Smooth Multiple Regression by Method of Average Derivatives," *Journal of the American Statistical Association*, 84, 986–995.

He, G., Müller, H.–G., and Wang, J.–L. (2002), "Functional Canonical Analysis for Square Integrable Stochastic Processes," *Journal of Multivariate Analysis*, 85, 54–77.

——— (2004), "Methods of Canonical Analysis for Functional Data," *Journal of Statistical Planning and Inference*, 122, 141–159.

He, P., Fang, K.–T., and Xu, C.–J. (2003), "The Classification Tree Combined with SIR and Its Application to Classification of Mass Spectra," *Journal of Data Science*, 1, 425–445.

Hristache, M., Juditsky, A., Polzehl, J., and Spokoiny, V. (2001), "Structure Adaptive Approach for Dimension Reduction," *The Annals of Statistics*, 29, 1537–1566.

Hristache, M., Juditsky, A., and Spokoiny, V. (2001), "Direct Estimation of the Index Coefficient in a Single-Index Model," *The Annals of Statistics*, 29, 595–623.

Hsing, T. (1999), "Nearest Neighbor Inverse Regression," *The Annals of Statistics*, 27, 697–731.

Hsing, T. (2004), *Lecture Notes for Functional Data Analysis*, Department of Statistics, Texas A&M University, College Station, TX.

Hsing, T., and Carroll, R. J. (1992), "An Asymptotic Theory of Sliced Inverse Regression," *The Annals of Statistics*, 20, 1040–1061.

Huang, S. P., Quek, S. T., and Phoon, K. K. (2001), "Convergence of the Truncated

Karhunen–Loéve Expansion for Simulation of Stochastic Processes," *International Journal for Numerical Methods in Engineering*, 52, 1029–1043.

James, G. M. (2002), "Generalized Linear Models with Functional Predictors," *Journal of the Royal Statistical Society*, Ser. B, 64, 411–432.

James, G. M., Hastie, T. J., and Sugar, C. A. (2000), "Principal Component Models for Sparse Functional Data," *Biometrica*, 87, 587–602.

James, G. M., and Silverman, B. W. (2005), "Functional Adaptive Model Estimation," *Journal of the American Statistical Association*, 100, 565–576.

Kailath, T. (1971), "RKHS Approach to Detection and Estimation Problems–Part I: Deterministic Signals in Gaussian Noise," *IEEE Trans. Inform. Theory*, IT-17, 530–549.

Kneip, A., and Gasser, T. (1992), "Statistical Tools to Analyze Data Representing a Sample of Curves," *The Annals of Statistics*, 20, 1266–1305.

Laukaitis, A., and Račkauskas, A. (2002), "Functional Data Analysis of Payment Systems," *Nonlinear Analysis: Modeling and Control*, 7, 53–68.

Lee, S.–Y., Zhang, W., and Song X.–Y. (2002), "Estimating the Covariance Functions with Functional Data," *British Journal of Mathematical and Statistical Psychology*, 55, 247–261.

Leurgans, S. E., Moyeed, R. A., and Silverman, B. W. (1993), "Canonical Corrlation Analysis When the Data Are Curves," *Journal of the Royal Statistical Society*, Ser. B, 55, 725–740.

Li, K. C. (1991), "Sliced Inverse Regression for Dimension Reduction," *Journal of the American Statistical Association*, 86, 319–327.

——— (1997), "Nonlinear Confounding in High-Dimensional Regression," *The Annals of Statistics*, 25, 577–612.

Li, K. C., Aragon, Y., Shedden, K., and Agnan, C. T. (2003), "Dimension Reduction for Multivariate Response Data," *Journal of the American Statistical Association*, 98, 99–

109.

Liu, X., and Müller, H.–G. (2004), "Functional Convex Averaging and Synchronization for Time-Warping Random Curves," *Journal of the American Statistical Association*, 99, 687–699.

Loéve, M. (1948), "Fonctions Aléatoires du Second Ordre, Supplement to P. Lévy," *Processus Stochastiques et Mouvement Brownien*, Gauthier–Villars, Paris.

Luckić, M. N., and Beder, J. H. (2001), "Stochastic Process with Sample Paths in Producing Kernel Hilbert Spaces," *Transactions of the American Mathematical Society*, 353, 3945–3969.

Malfait, N., and Ramsay, J. O. (2003), "The Historical Functional Linear Model," *The Canadian Journal of Statistics*, 31, 115–128.

Mandelbrot, B. B., and Van Ness, W. N. (1968), "Fractional Brownian Motions, Fractional Noises, and Applications," *SIAM*, 10, 423–437.

Müller, H.–G., and Stadtmüller, U. (2005), "Generalized Functional Linear Models," *The Annals of Statistics*, 33, 774–805.

Naik, P., and Tsai C.–L. (2000), "Partial Least Squares Estimator for Single-Index Models," *Journal of the Royal Statistical Society*, Ser. B, 62, 763–771.

——— (2005), "Constrained Inverse Regression for Incorporating Prior Information," *Journal of the American Statistical Association*, 100, 204–211.

Nuzman, C. J., and Poor, H. V. (2001), "RKHS Methods for Wide-Sense Self-Similar Processes," *The Annals of Applied Probability*, 11, 1199–1219.

Parzen, E. (1959), "Statistical Inference on Time Series by Hilbert Space Methods I," *Technical Report*, No. 23, Department of Statistics, Stanford University, CA.

——— (1961a), "An Approach to Time Series Analysis," *The Annal of Mathematical Statistics*, 32, 951–989.

——— (1961b), "Regression Analysis of Continuous Parameter Time Series," *Proceedings*

of the Fourth Berkerly Symposium on Mathematical Statistics and Probability, Berkeley, CA, 1, 469–489.

——— (1963), "Probability Density Functionals and Reproducing Kernel Hilbert Spaces," *Time Series Analysis,* 155–169.

Preda, C., and Saporta, G. (2005a), "PLS Regression on an Stochastic Process," *Computational Statistics and Data Analysis*, 48, 149–158.

——— (2005b), "Clusterwise PLS Regression on an Stochastic Process," *Computational Statistics and Data Analysis*, 49, 99–108.

Ramosay, J. O., and Dalzell, C. J. (1991), "Some Tools for Functional Data Analysis (with Discussion)," *Journal of the Royal Statistical Society,* Ser. B, 53, 539–572.

Ramosay, J. O., and Li, X. (1996), "Curve Registration," *Journal of the Royal Statistical Society*, Ser. B, 60, 351–363.

Ramosay, J. O., and Silverman, B. W. (1997), *Functional Data Analysis*, New York: Springer.

——— (2002), *Applied Functional Data Analysis*, New York: Springer.

Rao, M. M. (2000), *Stochastic Process: Inference Theory*, Boston: Kluwer Academic Publishers.

Rice, J. A. (2004), "Functional and Longitudinal Data Analysis: Perspecptives on Smoothing," *Statistica Sinica*, 14, 613–629.

Rice, J. A., and Silverman, B. W. (1991), "Estimating the Mean and Covariance Structure Nonparametrically When the Data Are Curves," *Journal of the Royal Statistical Society*, Ser. B, 53, 233–243.

Rice, J. A., and Wu, C. (2001), "Nonparametric Mixed Effects Models for Unequally Sampled Noisy Curves," *Biometrics*, 57, 253–259.

Rossi, F., Delannay, N., Conan–Guez, B., and Verleysen, M. (2005), "Representation of Functional Data in Neural Networks," *Neurocomputing*, 64, 183–210.

Schott, J. R. (1994), "Determining the Dimensionality in Sliced Inverse Regression," *Journal of the American Statistical Association*, 89, 141–148.

Setodji, C. M., and Cook, R. D. (2004), "K-Means Inverse Regression," *Technometrics*, 46, 421–429.

Silverman, B. W. (1996), "Smoothed Functional Principal Components Analysis by Choice of Norm," *The Annals of Statistics*, 24, 1–24.

Simonoff, J. S. (1996), *Smoothing Methods in Statistics*, New York: Springer.

Staniswallis, J. G., and Lee, J. J. (1998), "Nonparametric Regression Analysis of Longitudinal Data," *Journal of the American Statistical Association*, 93, 1403–1418.

Wahba, G. (1990), *Spline Models for Observational Data*, CBMS 59, Philadelphia: SIAM.

Wand, M. P., and Jones, M. C. (1995), *Kernel Smoothing*, London: Chapman and Hall.

Weinert, H. L. (1982), *Reproducing Kernel Hilbert Spaces: Applications in Statistical Signal Processing*, New York: Hutchinson Ross Publishing Company.

Wu, W. B., and Pourahmadi, M. (2003), "Nonparametric Estimation of Large Covariance Matrices of Longitudinal Data," *Biometrika*, 90, 831–844.

Xia, Y., Tong, H., Li, W. K., and Zhu, L.–X. (2002), "An Adaptive Estimation of Dimension Reduction Space," *Journal of the Royal Statistical Society*, Ser. B, 64, 363–410.

Yao, F., Müller, H.–G., Clifford, A. J., Dueker, S. R., Follett, J., Lin, Y., Buchholz, B. A., and Vogel, J. S. (2003), "Shiringkage Estimation for Functional Principal Component Scores with Application to the Population Kinetics of Plasma Folate," *Biometrics*, 59, 676–685.

Yao, F., Müller, H.–G., and Wang, J.–L. (2005), "Functional Data Analysis for Sparse Longitudinal Data," *Journal of the American Statistical Association*, 100, 577–590.

Yu, Y., and Ruppert, D. (2002), "Penalized Spline Estimation for Partially Linear Single-Index Models," *Journal of the American Statistical Association*, 97, 1042–1054.

Zhang, G., and Guo, M. (1990), *Lecture Notes of Functional Analysis, II*, Department of

Mathematics, Peking University.

Zhang, G., and Lin, Y. (1987), *Lecture Notes of Functional Analysis, I*, Department of Mathematics, Peking University.

Zhu, L.–X., and Fang, K.–K. (1996), "Asymptotics for Kernel Estimate of Sliced Inverse Regression," *The Annals of Statistics*, 24, 1053–1068.

Zhu, L.–X., and Ng, K. W. (1995), "Asymptotics of Sliced Inverse Regression," *Statistica Sinica*, 5, 727–736.

VITA

Haobo Ren was born in Beijing, China. He received a Bachelor of Science degree in Statistics in June 1994 and a Master of Science degree in Statistics in June 1997 from Peking University, Beijing, China. He came to the United States to pursue his Ph.D. in August of 1999 in the Department of Statistics, Texas A&M University. He received his Ph.D. under the direction of Dr. Tailen Hsing in August 2005. His permanent address is

17 Zhixing Garden #0101, Haidian District

Beijing, China