NOVEL METHODS FOR FUNCTIONAL DATA ANALYSIS WITH APPLICATIONS TO NEUROIMAGING STUDIES

By

Pratim Guha Niyogi

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Statistics – Doctor of Philosophy

2022

ABSTRACT

NOVEL METHODS FOR FUNCTIONAL DATA ANALYSIS WITH APPLICATIONS TO NEUROIMAGING STUDIES

By

Pratim Guha Niyogi

In recent years, there has been explosive growth in different neuroimaging studies such as functional magnetic resonance imaging (fMRI) and diffusion tensor imaging (DTI). The data generated from such studies are often complex structured which are collected for different individuals, via various time-points and across various modalities, thus paving the way for interesting problems in statistical methodology for analysis of such data. In this dissertation, some efficient methodologies are proposed with considerable development which have nice statistical properties and can be useful not only in neuroimaging but also in other scientific domains.

A brief overview of the dissertation is provided in Chapter 1 and in particular, different kinds of data structures that are commonly used in consecutive chapters are described. Some useful mathematical results frequently used in the theoretical derivations in various chapters are also provided. Moreover, we raise some fundamental questions that arise due to some specific data structures with applications in neuroimaging and answer these questions in subsequent chapters.

In Chapter 2, we consider the problem of estimation of coefficients in constant linear effect models for semi-parametric functional regression with functional response, where each response curve is decomposed into the overall mean function indexed by a covariate function with constant regression parameters and random error process. We provide an alternative semi-parametric solution to estimate the parameters using quadratic inference approach by estimating bases functions non-parametrically. Therefore, the proposed method can be easily implemented without assuming any working correlation structure. Moreover, we achieve a parametric $\sqrt{n}$-convergence rate of the proposed estimator under the proper choice of bandwidth and establish its asymptotic normality.

A multi-step estimation procedure to simultaneously estimate the varying-coefficient functions using a local linear generalized method of moments (GMM) based on continuous moment conditions

is developed in Chapter 3 under heteroskedasticity of unknown form. To incorporate spatial dependence, the continuous moment conditions are first projected onto eigen-functions and then combined by weighted eigen-values. This approach solves the challenges of using an inverse covariance operator directly. We propose an optimal instrumental variable that minimizes the asymptotic variance function among the class of all local linear GMM estimators, and it is found to outperform the initial estimates that do not incorporate spatial dependence.

Neuroimaging data are increasingly being combined with other non-imaging modalities, such as behavioral and genetic data. The data structure of many of these modalities can be expressed as time-varying multidimensional arrays (tensors), collected at different time-points on multiple subjects. In Chapter 4, we consider a new approach to study neural correlates in the presence of tensor-valued brain images and tensor-valued predictors, where both data types are collected over the same set of time-points. We propose a time-varying tensor regression model with an inherent structural composition of responses and covariates. This development is a non-trivial extension of function-on-function concurrent linear models for complex and large structural data where the inherent structures are preserved.

Through extensive simulation studies and real-data analyses, we demonstrate the opportunities and advantages of the proposed methods.

To my mother Maya for her sacrifices.
To my father Kajal for his support and protection.
To my wife Debolina for her unconditional love and support.

ACKNOWLEDGEMENTS

My journey till the completion of this dissertation has been a long one which has taught me many things. Now that I look back into all those moments, all I can feel is gratitude towards time, place, events and some very important people. With humbleness, I thank my academic advisors Dr. Ping-Shou Zhong and Dr. Tapabrata Maiti for providing invaluable technical support during my doctoral journey without which my journey would not have reached its goal. They trained me to pursue research independently, think out of the box, and deal with multiple problems simultaneously. Their inspiration boosted my confidence and persuaded me to contribute to statistical sciences. I am especially grateful to Dr. Zhong for continuing to mentor me even after moving to University of Illinois at Chicago. This long-distance mentoring was difficult, but he carried on with it effortlessly and never left me struggling. I thank Dr. Lyudmila Sakhanenko and Dr. Alla Sikorskii for serving in my dissertation committee and for providing helpful constructive suggestions to improve the quality of this dissertation.

Reflecting upon the start of my journey, I would like to thank the Graduate Students' Selection Committee of 2016 for offering me the opportunity to pursue the doctoral program here at Michigan State University (MSU) and for providing assistantship during my education. Dr. Tapabrata Maiti, the then graduate director, considered me worthy of the opportunity and expressed his belief in me when I was just another fresh Masters from India venturing into this foreign land for a new endeavour. His encouraging words have helped me believe in myself and seek excellence. I thank Dr. Alla Sikorskii for supporting me through research assistantship for more than three years from her grants and helping me grow in collaborative research. This opportunity gifted me the beneficial experience of interdisciplinary research and exposure to statistics in health sciences. I will never forget her immense help and support. I would like to thank Drs. Shlomo Levental, Ping-Shou Zhong and Lyudmila Sakhanenko for teaching four very important courses: probability theory, linear models, and asymptotics and non-parametric curve estimation, respectively, with care and depth. The lessons learned from those courses have helped a lot in my research. I am thankful to my

seniors especially Rejaul Karim, who was my roommate for three years and I spent some of the best days at East Lansing in his company, and Atrayee Majumder for her valuable suggestions. I would like to make a special mention of Andy Hufford for being approachable and helping out with just about any technology-related issues. Also, I thank Tami Hankey for her assistance in completing numerous paper-works all along the way. A huge role on this dissertation was played by the high-performance computing technology of MSU without which running codes with voluminous data would not be possible. I am therefore thankful to computational resources and services provided by the Institute for Cyber-Enabled Research at MSU.

There are also people outside the University who have been of immense help. The person to whom I am extremely grateful for providing me with the opportunity to spend a wonderful summer at the Biostatistics department of Johns Hopkins University is Dr. Martin Lindquist. I learnt many things from him and am honoured to have been able to collaborate with him in one of my projects. I would also like to thank Dr. Xiaohong Joe Zhou of University of Illinois at Chicago for providing me with valuable data that helped me validate my research findings.

Going back to my roots, I would like to thank my high-school statistics teacher Ashoke Panda, faculty of Hare School, Kolkata who for the first time introduced me to this colorful subject and motivated me to pursue it. I thank Arindam Bhattacharyya, who has been a valuable guide all the way upto my Masters. I am thankful to my alma maters Presidency University and Indian Statistical Institute for their intellectual environment that shaped my critical thinking ability and propelled me to pursue research, and I will forever remain grateful to my roots. I am thankful to all professors who helped me, guided me, and motivated me during those times. I am especially thankful to Drs. Saurabh Ghosh and Ayanendranath Basu of Indian Statistical Institute and Dr. Subhra Sankar Dhar of Indian Institute of Technology, Kanpur who engaged me in research at various phases of my educational levels; their training and treatment were very beneficial to me. Nevertheless, I am indebted to Dr. Partha Sarathi Chakraborty for his excellent teaching, which created a strong foundation during my undergraduate days. Also, my sincere gratitude to the Ramakrishna Mission Institute of Culture for awarding me the Debesh-Kamal Scholarship for Higher Studies in 2016

which provided me the necessary funds to travel to USA.

My family has been a pillar of strength throughout my life. I thank my parents Maya and Kajal Guha Niyogi for not stopping me from trying things out and letting me fail so I could learn the right way. I am grateful to them for never comparing me to anyone else, for always believing in me, and for teaching me humility and respect. I am thankful to my wife Debolina Chatterjee for her unconditional love and support and for being the first reader of all my writings. She is my "better half" in uncountable ways and I thank her for bringing out the best in me.

Last but not least, I thank the Almighty for this life and for making things fall into place.

Pratim Guha Niyogi

June 29, 2022

East Lansing, MI

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ALGORITHMS

KEY TO ABBREVIATIONS

ADC Apparent Diffusion Coefficient

BOLD Blood Oxygen Level Dependent

DTI Diffusion Tensor Imaging

FA Fractional Anisotropy

(f)MRI (Functional) Magnetic Resonance Imaging

FDA Functional Data Analysis

FPCA Functional Principal Component Analysis

FVE Fraction of variance explained

GEE Generalized Estimating Equations

GLM Generalized Linear Model

(G)MM (Generalized) Method of Moments

i.i.d. Independent and Identically Distributed

LDA Longitudinal Data Analysis

MD Mean Diffusivity

OSA Obstructive Sleep Apnea

PCA Principal Component Analysis

PM Particulate Matter

QIF Quadratic Inference Functions

ROI Region of Interest

TR Repetition Time

VCM Varying Coefficient Model

CHAPTER 1

PROLOGUE

In this chapter, we provide a brief overview of the relevant topics and state the problems that we are going to present in the following chapters.

1.1 Big data analysis

The advances in scientific research and technological developments have led to the collection and storage of huge amounts of data which are not only voluminous but also complex in structure. These are commonly called "Big data". Big data give rise to statistical problems in natural science, engineering, social sciences, and humanities. The analysis of such data having massive volumes and complex structures for decision-making and scientific discovery is a challenge faced by statisticians and computer scientists, which requires innovative statistical and computational methods, sophisticated statistical modelling, and theoretical results. Collectively, this is known as "Data science" which has nowadays become a multi-disciplinary field involving knowledge from various disciplines for developing new methodologies for various kinds of data: low or high dimensional; structured, unstructured, or semi-structured. In recent decades, big data has become a significant part of scientific interest, where images, videos, texts, and other objects can be considered as a form of massive data. Therefore, the statistician plays an important role in proposing new methodologies for discovering information from available big data. In the following subsections, we discuss the different data structures that lead to different methodologies. In Sub-section 1.1.1 we discuss functional data analysis, and in Sub-section 1.1.2 we discuss analysis of a complex structured multidimensional array (also known as tensor data analysis).

1.1.1 Functional data analysis

This subsection is dedicated to a discussion on functional data analysis (FDA) and its relevance in the dissertation. Some relevant review articles include Morris (2015); Müller (2016); Wang

et al. (2016). Owing to the types of data generated in various scientific research in the fields of biology, audiology, environmental sciences, earth sciences, and economics (to name a few), there was a need for a statistical methodology that could analyze data which are observed as functions varying over time, space, or other continuum domains. This led to the development of functional data analysis. Although the term "functional data analysis" was coined by Jim Ramsey in the famous 1982 paper (Ramsay, 1982), its origin dates back to the late 1940s in the Ph.D. theses of Kari Karhunen (Karhunen, 1946) and Ulf Grenander (Grenander, 1950). In their seminal works, Karhunen and Grenander respectively, discussed the decomposition of square integrable continuous-time stochastic processes into series expansions to obtain representations in a Hilbert space. The idea to expand random curves appeared in Rao (1958) and Tucker (1958) around the same time. In the last three decades, FDA gained considerable momentum in statistics literature, of which some significant works are Ramsay and Silverman (2005, 2007); Ferraty and Vieu (2006); Horváth and Kokoszka (2012); Zhang (2013); Hsing and Eubank (2015) and some notable survey articles are Morris (2015); Wang et al. (2016); Greven and Scheipl (2017); Li et al. (2022). The main feature that makes functional data distinct from other types of data, especially those having a large p (number of parameters) and small n (sample size) framework, is that the functional data are infinite-dimensional in nature, since the underlying statistical quantity of the measurement is a curve or a surface over a continuum domain. Thus, the commonly used classical multivariate statistical methods (Anderson et al., 1958) do not suffice for these types of analyses. Moreover, in asymptotic analysis, the space between the function arguments is assumed to approach zero, hence making the number of arguments tend to infinity. This is essentially the large p (rather p_n, where the number of arguments of the function is p_n for the sample size n) problem in high-dimensional statistics. In fact, this dimensionality issue is a blessing in disguise because we end up with more data, with an extra cost being paid by the smoothness assumption on some standard spaces. The smoothness assumption tells us that the information from measurements at neighboring arguments can be pooled, thereby overcoming the curse of dimensionality.

Some significant research maneuvering the use of FDA include

- Functional generalized linear models (Müller and Stadtmüller, 2005)
- Functional sliced inverse regression (Ferré and Yao, 2005)
- Multi-level functional data analysis (Crainiceanu et al., 2009; Huang et al., 2014; Xu et al., 2018)
- Functional time series (Hörmann and Kokoszka, 2010; Aue et al., 2015; Kowal et al., 2017; van Delft and Eichler, 2018)
- Spatially dependent functional data (Zhu et al., 2014; Kuenzer et al., 2021)
- Spatio-temporal point process (Li and Guan, 2014; Goldsmith et al., 2015)
- Longitudinal functional data analysis (Goldsmith et al., 2012; Chen and Müller, 2012; Park and Staicu, 2015; Staicu et al., 2020)

In FDA, continuous functional data are available at every time-point or can at least be evaluated for some time-points. In practice, however, data are observed in discrete domains such as time-points, with or without measurement errors. For the demonstration of the theoretical results, without loss of generality, it suffices to assume that the functional data are observed continuously without measurement error. Note that the theory behind the estimation can be different for different measurement schedules/ sampling plans such as densely or sparsely observing data over time-points. In most cases, the analyses of sparse and dense functional data are different, although sparse and dense data are asymptotic concepts and are difficult to use in practice. While for dense functional data, one can smooth each of the curves separately and then proceed with further estimation and inference procedures based on pre-smoothed curves (Castro et al., 1986; Rice and Silverman, 1991; Zhang and Chen, 2007), for sparse functional data, the pre-smoothing step is not required (Yao et al., 2005). Various smoothing techniques are available in non-parametric literature to deal with functional data. The different types of non-parametric smoothing techniques commonly used in the literature are

- Spline smoothing (Rice and Silverman, 1991; Cai and Hall, 2006)
- B-spline (Cardot et al., 1999; James et al., 2000; Rice and Wu, 2001)
- Penalized splines (Ruppert et al., 2003; Yao and Lee, 2006)
- Local polynomial smoothing (Fan and Gijbels, 1996; Zhang and Chen, 2007; Yao and Li, 2013)

Principal component analysis (PCA) in FDA is a generalization of the classical high-dimensional statistics for finite-dimensional matrix-valued observations to the case of infinite-dimensional continuum domain, and it is termed as functional principal component analysis (FPCA). The main objective of FPCA is to express the underlying stochastic processes as a truncated sum of a countable sequence of uncorrelated random variables, thereby reducing the problem from infinite into that of finite dimension, so that the tools of multivariate data analysis can be applied to the resulting random vector of scores. FPCA based on spline smoothing was studied in James et al. (2000); Zhou et al. (2008), whereas, FPCA based on kernel were discussed in Hall et al. (2006); Müller and Yao (2010); Li and Hsing (2010). Asymptotic theories based on kernel smoothing (a.k.a. local polynomial smoothing) are more profound in the literature. For fully observed dense data, Hall and Hosseini-Nasab (2009) derived a stochastic expansion of estimators of eigen-values and eigen-functions based on the principles of operator theory, the statistical implementations of which were provided in Hall and Hosseini-Nasab (2006). For sparse functional data, FPCA approach was studied in Yao et al. (2005); Liu and Müller (2009). Hall et al. (2006) discussed the theoretical properties of FPCA based on local linear smoother. In one of the seminal works in Li and Hsing (2010), an estimation procedure was discussed for all types of sampling strategies. It was found that in some specific ranges for the rate of the number of functional points, for dense sampling strategy, pre-smoothing was found to be asymptotically negligible, and other important commonly used statistics such as mean, covariance, and eigen-components could be estimated using the parametric rate. On the other hand, for sparse functional data, those statistics could only be estimated with a non-parametric convergence rate. It was shown that the estimation of the eigen-values was not

as sensitive to the sampling design as the estimation of the eigen-functions. This was the first time where a phase transition was observed. Zhang and Wang (2016) investigated local linear estimation of mean and covariance functions with general weighting schemes, where equal weight per observation and equal weight per subject were two special cases. All works mentioned till now were based on univariate functional data. Multivariate FPCA was discussed in Viviani et al. (2005); Wang (2008); Berrendero et al. (2011); Chiou et al. (2014); Happ and Greven (2018) among many others.

Regression analysis for FDA is one of the most active research domains for the analysis of functional data wherein the modelling of the data depends on the type of variables. For example,

- when the response is functional, but the covariates are vectors, the approach is called function-on-scalar regression (Zhu et al., 2014; Chen et al., 2019).

- when the response is vector valued, while the covariate is functional, this approach is called scalar-on-function regression (Cardot et al., 1999, 2003; Müller and Stadtmüller, 2005; Cai and Hall, 2006; Hall et al., 2007; Li and Hsing, 2007; Goldsmith et al., 2011; Kato, 2012). In this approach, the covariates and the varying coefficient are expressed as the same set of orthogonal functional bases.

- when the response and covariates are both functional, the approach is called function-on-function regression. It was introduced by (Ramsay and Dalzell, 1991). In this regression set-up, a varying coefficient model (Hastie and Tibshirani, 1993) was implemented. These regression models are often referred to as concurrent linear models. Recent literature for functional concurrent linear models include Faraway (1997); Zhang and Chen (2007); Zhang et al. (2010); Wang et al. (2016); Fang et al. (2020). Other techniques to estimate the regression function can be found in Hoover et al. (1998); He et al. (2003); Yao et al. (2005); He et al. (2018).

In this dissertation, in Chapter 3, we consider the first case where the response is functional but the covariates are vectors and we consider the third case where the covariate and response both are

functional in Chapters 2 and 4.

1.1.2 Tensor data analysis

In many scientific researches, for instance, in areas of imaging studies, network sciences, economics, computer technologies, genetics, recommendation systems, etc. data appear structured. Such high-dimensional as well as multi-dimensional structures have raised various challenges to their analysis. Thus, multidimensional arrays, popularly known as "tensors", came as a savior for understanding the structure of these complex data. The tensor as a generalization of matrices appeared for the first time in the literature during 1928 (Hitchcock, 1928) and was used to represent and store data efficiently. Ever since then, its use has seen a boom in the scientific community. Some significant research surveys can be found in Ji et al. (2019); Bi et al. (2021). Sub-section 1.3.1 discusses some basic notation and properties of the tensor. We will consider such kind of data in chapter 4 in more detail.

1.2 Some applications

In December 2, 1956, eminent statistician Professor P. C. Mahalanobis emphasized that

> Statistics is the universal tool of inductive inference, research in natural and social sciences, and technological applications. Statistics, therefore, must have a clearly defined purpose, either in the pursuit of knowledge or in the promotion of human welfare.

In this dissertation, some advanced methodologies driven by their applications are proposed for two types of neuroimaging studies.

1.2.1 (Functional) magnetic resonance imaging

A remarkable research area developed in magnetic resonance imaging (MRI) for studying the structure and functioning of the human brain in the years following 1977 after the first MRI scanner

was developed. In these studies, the differences in magnetic properties of certain molecules (especially water molecules) in the brain are measured by using the fact that their density differs in different media like air, white matter, gray matter, blood vessels, and tumors. Functional MRI, also known as fMRI, has recently gained popularity as a pre-surgical procedure to map the functional architecture of a subject's brain without exciting the tissues associated with some critical skills like vision, hearing, etc. Occurrence of neural activity in a certain portion of the brain results in increased metabolic activity, causing a rush of oxygen-carrying hemoglobin in that particular area, whereas immediately following the end of neural activity, the oxygen level drops. These changes in the oxygen levels give rise to a measure called the blood oxygen level dependent (BOLD) signal, which is the ratio between oxygenated and de-oxygenated hemoglobin in blood. The objective of fMRI studies is to observe the neural activity of the brain in instantaneous time with high spatial resolution by detecting changes in the BOLD signal. Typically, the BOLD signal happens to rise well above the baseline with a peak at around 6 seconds following a neural activity, and decays back to baseline over a period of 20 seconds. Due to the observable nature of neural activities, we can use fMRI data to make various inferences if we can assess the relationship between neural activity and the BOLD response.

In fMRI data, images are collected over time; therefore, in order to maintain high temporal resolution, spatial resolution is sacrificed. High-resolution structural images are used to get back the spatial resolution from the fMRI data, and the spatial coordinates are used to identify the activation regions during the fMRI scans by examining the aligned structural coordinates. The times between two successive scans are called repetition time (TR). Subjects are aligned in the scanner which is assumed to be a three-dimensional coordinate system with coordinates (X, Y, Z) to the bore of the magnet, where the Z direction is downward to the bore (from feet to head) and the X and Y directions refer to the plane which is perpendicular to the Z axis. A schematic diagram of the MRI scanner is provided in Figure 1.1. The brain is naturally a continuous medium due to the existence of neurons in almost all coordinates, but can be made discrete by dividing the brain into a set of cubes. These cubes are commonly called voxels. A typical MRI scan of a healthy human brain is

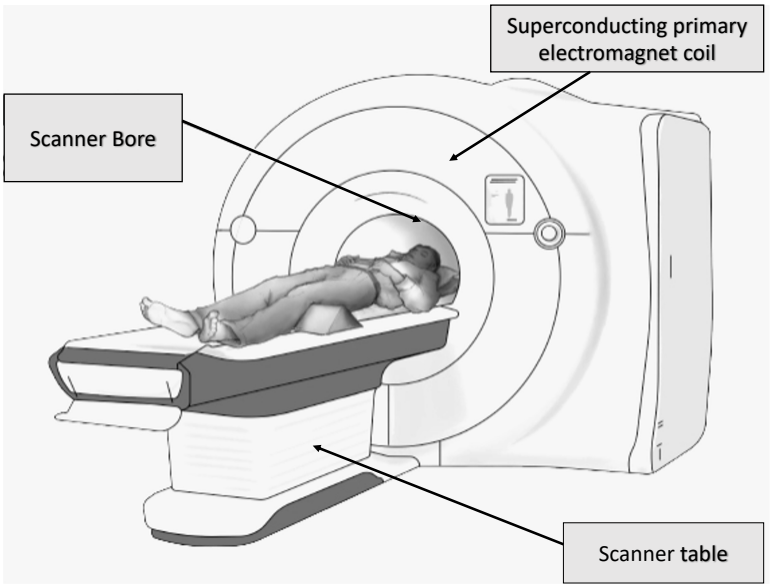

Figure 1.1 A schematic diagram of an MRI scanner.

provided in Figure 1.2.

fMRI data consist of spatial and temporal correlations; therefore, we need sophisticated tools to analyze them. For example, brain tissue in neighboring voxels is supplied by the same kind of vasculature; as a result, a large response in one voxel in a neighborhood set increases the probability that the neighboring voxels consist of a large response (spatial dependency) or, under the same set of stimuli over time, the brain activation is expected to be similar. Moreover, fMRI data can often be corrupted with noise arising due to the thermal motion of electrons inside the bore of the magnet, the brain itself, and due to other physiological reasons. In order to reduce such inherent unaccounted and uncontrolled errors due to head motion scanner drift, a series of prepossessing is performed (see Appendix A for more details). The main objective of fMRI data analysis is to identify regions of the brain that show task-related activity. For more information on fMRI data analysis, please refer to Huettel et al. (2004); Lindquist (2008); Ashby (2011); Wager and Lindquist (2015).

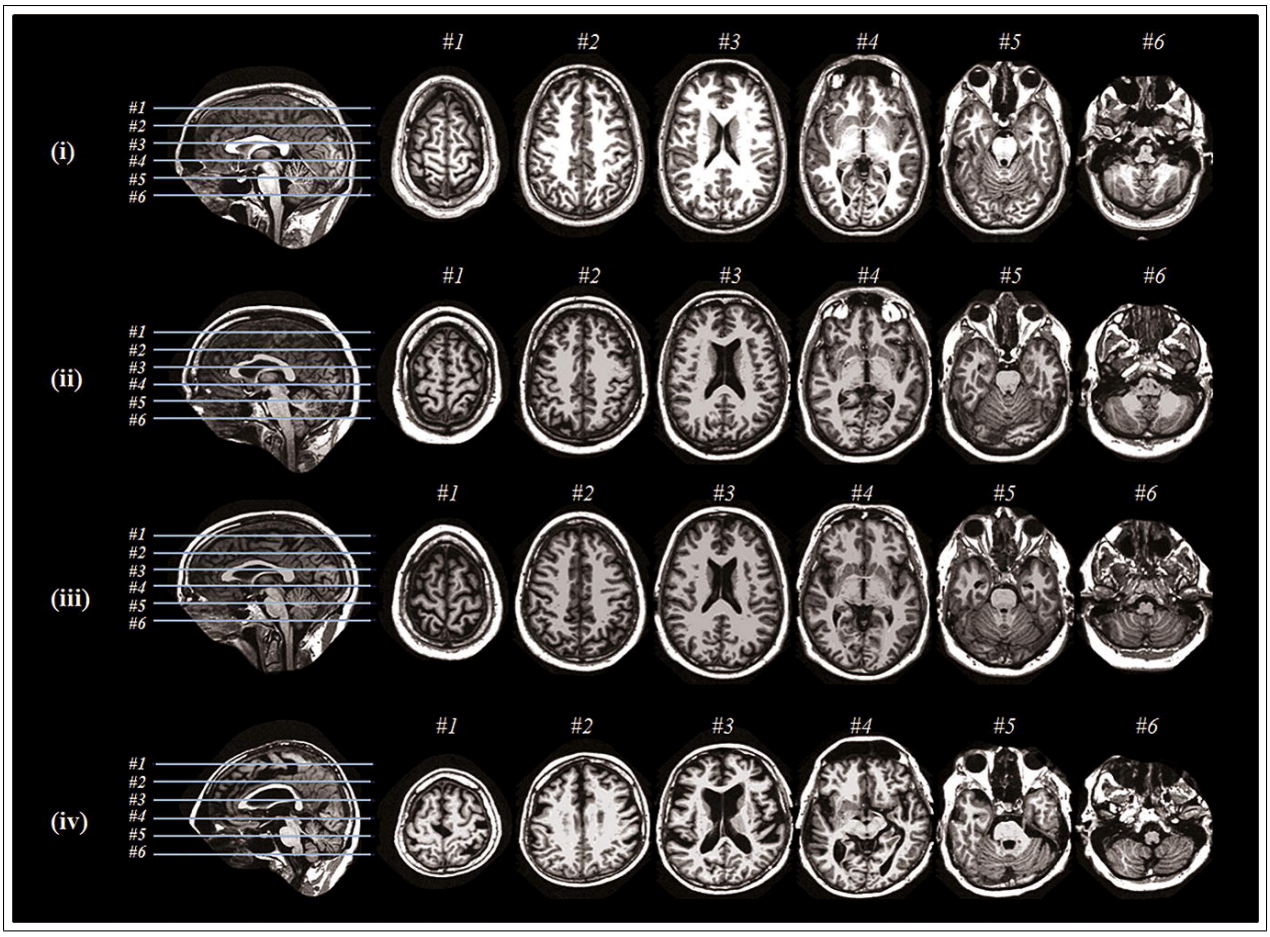

Figure 1.2 A typical example of MRI scan of healthy human brain. Source: Long et al. (2012)

1.2.2 Diffusion tensor imaging

Diffusion tensor imaging, popularly known as DTI measures the restricted diffusion of water molecules in the brain tissues in order to produce neural tract images. When water molecules are located in fiber tracts, their movement is restricted and they are more likely to be anisotropic; whereas those molecules in the rest of brain are less restricted in their movement, and are therefore isotropic. Diffusion causes water molecules to diverge from a central point and gradually reach the surface of an ellipsoid when the medium is anisotropic. In an isotropic medium, water molecules move out at the same rate in all directions. Thus, using the laws of physics (such as attenuation), the signal of an MRI voxel can be converted into numerical measures of diffusion, which are taken care of by physicians. Thus, each brain voxel has one or more pairs of parameters, such as the

rate of diffusion, direction of diffusion etc. The properties of each voxel of a single DTI image are usually calculated by vector or higher-order multi-dimensional arrays.

Consider an ellipsoid tensor in a three-dimensional Cartesian grid, where there exist three projections of the given ellipsoid into three different axes. These projections provide apparent diffusion coefficient (ADC) and are denoted as ADC_x, ADC_y and ADC_z corresponding to X, Y and Z axes respectively. Therefore, the average diffusivity in a given voxel is defied by $\text{ADC} = (\text{ADC}_x + \text{ADC}_y + \text{ADC}_z)/3$. Note that ellipsoid has three axes, one principle axis (longest) and two small axes passing through center, where the direction and length of these axes are eigen-vectors and eigenvalues, respectively, in the context of tensor algebra. The diffusion along the principle axis is termed as axial diffusivity (denoted as L_1) and the average diffusivity along two other minor axes is termed as radial diffusivity (denoted as L_{23} where $L_{23} = (L_2 + L_3)/2$, L_2 and L_3 are eigen-values corresponding to minor axes). The mean diffusivity is defined as $\text{MD} = (L_1 + L_2 + L_3)/3$. The degree of anisotropy of a diffusion process is termed as fractional anisotropy (FA), which is a scaled measure that belongs to the interval $[0, 1]$. The quantity FA takes the value zero when diffusion is isotropic (i.e., unrestricted in all directions) and takes the value one when diffusion occurs only on one and is fully restricted in other directions. FA can be calculated using the following formula.

$$\text{FA} = \sqrt{\frac{3}{2}} \frac{\left\{(L_1 - \text{MD})^2 + (L_2 - \text{MD})^2 + (L_3 - \text{MD})^2\right\}^{1/2}}{\left\{L_1^2 + L_2^2 + L_3^2\right\}^{1/2}} \tag{1.1}$$

For more information on DTI, please refer to O'Donnell and Westin (2011); Van Hecke et al. (2016).

1.3 Mathematical preliminaries

In this section, we introduce some notation and basic mathematical principles that will be used frequently throughout the dissertation.

1.3.1 Notations of tensor/matrix object

Tensor is a multidimensional array indexed by D many indices. The first-order tensor is a vector for $D = 1$, second-order tensor is a matrix for $D = 2$ and for $D > 2$, we call the set of objects higher-order tensors. In the following paragraph, we provide a brief summary of tensors and define important notation. Interested readers can refer to a survey article by Kolda and Bader (2009) for more details.

A D-dimensional tensor is denoted by Sans-serif upper-face letters $\mathsf{A} \in \mathbb{R}^{I_1 \times \cdots \times I_D}$ where the size I_d along each mode or dimension d for $d = 1, \cdots, D$. Therefore, the number of elements in the tensor A is $I = \prod_{d=1}^{D} I_d$ and the order of the tensor is the number of dimensions. Here and henceforth, matrices are denoted by bold-face capital letters (examples: $\mathbf{A}, \mathbf{B}, \cdots$), vectors are written as bold-face lower-case letters (examples: $\mathbf{a}, \mathbf{b}, \cdots$) and scalars are presented as Latin alphabets $(a, b, \cdots)$. The entry on i-th row and j-th column of a matrix $\mathbf{A}$ is denoted by $(\mathbf{A})_{i,j} = a_{ij}$ and $(i_1, \cdots, i_D)$-th entry of a D dimensional tensor is denoted as $(\mathsf{A})_{i_1,\cdots,i_D} = a_{i_1,\cdots,i_D}$. Slices are two-dimensional sections of the tensor defined by fixing all but two indices and thus become a $I_d \times I_{d'}$ dimensional matrix. For a D-way tensor $\mathsf{A} \in \mathbb{R}^{I_1 \times \cdots \times I_D}$ with the element $a_{i_1,\cdots,i_D}$ at the position with mode $i_d, d = 1, \cdots, D$, vectorization operator $\mathrm{vec}(\cdot)$ is defined as a vector with length $\prod_{d=1}^{D} I_d$ which is formed by stacking the nodes of A into a single column vector, i.e.,

$$\mathrm{vec}(\mathsf{A}) \left[i_1 + \sum_{d=2}^{D} \left(\prod_{k=1}^{d-1} I_k \right) (i_d - 1) \right] = a_{i_1,\cdots,i_D} \tag{1.2}$$

By simplifying, for a matrix $\mathbf{A}$ of order $I \times J$, $\mathrm{vec}(\mathbf{A}) = (a_{1,1}, \cdots, a_{I,1}, \cdots, a_{1,J}, \cdots, a_{I,J})^{\mathrm{T}}$. Similarly to the vectorization operator, one can unfold as d-mode matricization or unfolding, a D-array A, to form a matrix $\mathbf{A}_{(d)}$ with I_d rows and $\prod_{d':d' \neq d} I_{d'}$ columns where the element $a_{i_1,\cdots,i_D}$ is at the row i_d and column $\left\{ 1 + \sum_{\substack{d_1=1 \\ d_1 \neq d}}^{D} (i_{d_1} - 1) \prod_{\substack{d_2=1 \\ d_2 \neq d}}^{d_1-1} I_{d_2} \right\}$.

1.3.2 Different kinds of products

Analogous to the Frobenius norm in a 2D space, the norm of a tensor A is the square root of the sum of squares of its indices, denoted as

$$\|\mathsf{A}\|_{\mathscr{F}} = \sqrt{\sum_{i_1=1}^{I_1} \cdots \sum_{i_D=1}^{I_D} a_{i_1,\cdots,i_d}^2} \tag{1.3}$$

The scalar product $\langle \mathsf{A}, \mathsf{B} \rangle$ of two D-dimensional tensors with the same size is defined as

$$\langle \mathsf{A}, \mathsf{B} \rangle = \sum_{i_1,\cdots,i_D} b_{i_1,\cdots,i_D} a_{i_1,\cdots,i_D} \tag{1.4}$$

Thus, immediately, the Frobenius norm of the tensor A can be expressed as $\|\mathsf{A}\|_{\mathscr{F}} = \sqrt{\langle \mathsf{A}, \mathsf{A} \rangle}$. Two tensors A and B are said to be orthogonal if $\langle \mathsf{A}, \mathsf{B} \rangle = 0$. Furthermore, consider the contracted tensor product between two tensors with different mode dimensions. For two tensors $\mathsf{A} \in \mathbb{R}^{I_1 \times \cdots \times I_K \times P_1 \times \cdots \times P_L}$ and $\mathsf{B} \in \mathbb{R}^{P_1 \times \cdots \times P_L \times Q_1 \times \cdots \times Q_M}$, contracted tensor product (Lock, 2018; Raskutti et al., 2019) is defined as $\langle \mathsf{A}, \mathsf{B} \rangle_L$ with $(i_1, \cdots, i_K, q_1, \cdots, q_M)$-th element $\sum_{p_1,\cdots,p_L} a_{i_1,\cdots,i_K,p_1,\cdots,p_L} \times b_{p_1,\cdots,p_L,q_1,\cdots,q_K}$. As a special case, $\langle \mathbf{A}, \mathbf{B} \rangle_1 = \mathbf{AB}$ where $\mathbf{A}$ and $\mathbf{B}$ are $I \times P$ and $P \times Q$ matrices.

A D-way tensor A has a rank-1 when it can be written as the outer product of D vectors $\mathbf{u}^{(1)}, \cdots, \mathbf{u}^{(D)}$ of length $I_1, \cdots, I_K$ respectively, i.e.,

$$\mathsf{A} = \mathbf{u}^{(1)} \circ \cdots \circ \mathbf{u}^{(D)} \tag{1.5}$$

where $(i_1, \cdots, i_D)$-th element of A is $\prod_{d=1}^{D} u_{i_d}^{(d)}$. The Kronecker product of the matrices $\mathbf{A} \in \mathbb{R}^{I \times J}$ and $\mathbf{B} \in \mathbb{R}^{K \times L}$ is denoted by $\mathbf{A} \otimes \mathbf{B}$, an $(IK) \times (JL)$ matrix, defined by

$$\mathbf{A} \otimes \mathbf{B} = (a_{ij}\mathbf{B})_{i,j} = [\mathbf{a}_1 \otimes \mathbf{b}_1,\ \mathbf{a}_1 \otimes \mathbf{b}_2,\ \cdots,\ \mathbf{a}_J \otimes \mathbf{b}_{L-1},\ \mathbf{a}_J \otimes \mathbf{b}_L] \tag{1.6}$$

The Khatri-Rao product of matrices $\mathbf{A} \in \mathbb{R}^{I \times K}$, $\mathbf{B} \in \mathbb{R}^{J \times K}$, denoted as $\mathbf{A} \odot \mathbf{B}$, is defined by

$$\mathbf{A} \odot \mathbf{B} = [\mathbf{a}_1 \otimes \mathbf{b}_1,\ \cdots,\ \mathbf{a}_K \otimes \mathbf{b}_K] \tag{1.7}$$

which is an $(IJ) \times K$ matrix. The Hadamard product is the element-wise matrix product which is denoted by $\mathbf{A} * \mathbf{B}$ where $\mathbf{A}, \mathbf{B}$ are $I \times J$.

1.3.3 Tensor decomposition

Now the question is how to represent the tensor as a sum of a finite number of rank-1 tensors? The answers came from psychometrics in the form of canonical decomposition or CANDECOMP (Carroll and Chang, 1970) and parallel factors or PARAFAC (Harshman et al., 1970) decomposition and now in the literature of tensor decomposition, it is known as CANDECOMP/PARAFAC (CP) decomposition which is an extension of matrix singular value decomposition (Tucker, 1966; Kiers, 2000). Therefore, the CP decomposition factorizes a tensor into a sum of rank-1 tensors, i.e., mathematically,

$$\mathsf{A} = \sum_{r=1}^{R} \mathbf{u}_r^{(1)} \circ \cdots \circ \mathbf{u}_r^{(D)} \tag{1.8}$$

where $\mathbf{u}_r^{(d)} \in \mathbb{R}^{I_d}, d = 1, \cdots, D$ and $r = 1, \cdots, R$ are column vectors and A cannot be written as a sum of less than R outer products for a positive integer R which is the rank of the tensor. Equation (1.8) is sometimes denoted as $\mathsf{A} = [[\mathbf{U}_1, \cdots, \mathbf{U}_D]]$ where $\mathbf{U}_1, \cdots, \mathbf{U}_D$ have linearly independent columns $\mathbf{U}_d = [\mathbf{u}_1^{(d)}, \cdots, \mathbf{u}_R^{(d)}] \in \mathbb{R}^{I_d \times R}$ for each $d = 1, \cdots D$.

1.3.4 Some useful results

In this sub-section, we will present some useful well-known results without proofs. We define $\mathbf{1}_R$ as an R-dimensional vector with all elements 1.

1. $\text{vec}(\mathbf{u}^{(1)} \circ \mathbf{u}^{(2)} \circ \cdots \circ \mathbf{u}^{(D)}) = \mathbf{u}^{(D)} \otimes \cdots \otimes \mathbf{u}^{(1)}$.

2. For two vectors $\mathbf{a}$ and $\mathbf{b}$, $\mathbf{a} \otimes \mathbf{b} = \mathbf{a} \odot \mathbf{b}$, $\mathbf{a} \circ \mathbf{b} = \mathbf{a}\mathbf{b}^{\mathrm{T}}$.

3. For any matrices $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ so that the required matrix multiplications are possible, we have the following:

   a) $(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = \mathbf{AC} \otimes \mathbf{BD}$.

   b) $(\mathbf{A} \otimes \mathbf{B})^{-1} = \mathbf{A}^{-1} \otimes \mathbf{B}^{-1}$

   c) $\mathbf{A} \odot \mathbf{B} \odot \mathbf{C} = (\mathbf{A} \odot \mathbf{B}) \odot \mathbf{C} = \mathbf{A} \odot (\mathbf{B} \odot \mathbf{C})$

d) $(\mathbf{A} \odot \mathbf{B})^{\mathrm{T}}(\mathbf{A} \odot \mathbf{B}) = (\mathbf{A}^{\mathrm{T}}\mathbf{A}) * (\mathbf{B}^{\mathrm{T}}\mathbf{B})$

e) $(\mathbf{A} \odot \mathbf{B})^{-1} = \{(\mathbf{A}^{\mathrm{T}}\mathbf{A}) * (\mathbf{B}^{\mathrm{T}}\mathbf{B})\}^{-1}(\mathbf{A} \odot \mathbf{B})^{\mathrm{T}}$

f) $\mathrm{vec}(\mathbf{A} \odot \mathbf{B}) = ((\mathbf{I} \odot \mathbf{A}) \otimes \mathbf{I})\,\mathrm{vec}(\mathbf{B})$

g) $\mathrm{vec}(\mathbf{B} \odot \mathbf{A}) = \left\{\mathbf{I} \odot (\mathbf{A}(\mathbf{I} \otimes \mathbf{1}^{\mathrm{T}}))\right\}\mathrm{vec}(\mathbf{B})$

h) $\mathrm{trace}(\mathbf{AB}) = \mathrm{trace}(\mathbf{BA}) = \mathrm{vec}(\mathbf{A}^{\mathrm{T}})^{\mathrm{T}}\,\mathrm{vec}(\mathbf{B})$

i) $\mathrm{vec}(\mathbf{ABC}) = (\mathbf{C}^{\mathrm{T}} \otimes \mathbf{A})\,\mathrm{vec}(\mathbf{B})$

j) $\mathrm{rank}(\mathbf{A} \odot \mathbf{B}) \leq \mathrm{rank}(\mathbf{A} \otimes \mathbf{B}) \leq \mathrm{rank}(\mathbf{A})\mathrm{rank}(\mathbf{B})$

k) If $\mathbf{A}$ is a matrix of order $m_1 \times m_2$ with (i, j)-th element a_{ij}, then the Frobenius norm of $\mathbf{A}$ is defined as $\|\mathbf{A}\|_{\mathscr{F}} = \sqrt{\sum_{i=1}^{m_1}\sum_{j=1}^{m_2}|a_{ij}|^2} = \sqrt{\mathrm{trace}(\mathbf{A}^{\mathrm{T}}\mathbf{A})} = \sqrt{\sum_{i=1}^{\min(m_1,m_2)}\sigma_i^2(\mathbf{A})}$ where $\sigma_i(\mathbf{A})$ is the i-th order singular value of $\mathbf{A}$ and $\mathrm{trace}(\mathbf{A})$ is the trace operator of a square matrix $\mathbf{A}$.

4. If the tensor A admits a rank-R decomposition, then $\mathbf{A}_{(d)} = \mathbf{U}_d(\mathbf{U}_D \odot \cdots \odot \mathbf{U}_{d+1} \odot \mathbf{U}_{d-1} \odot \cdots \odot \mathbf{U}_1)^T$ and $\mathrm{vec}(\mathsf{A}) = (\mathbf{U}_D \odot \cdots \odot \mathbf{U}_1)\mathbf{1}_R$

5. $\|\mathsf{A}\|_{\mathscr{F}} = \|\mathbf{A}_{(d)}\|_{\mathscr{F}} = \|\,\mathrm{vec}(\mathbf{A}_{(d)})\|_2$ for $d = 1, \cdots, D$

6. $\mathrm{vec}(\mathsf{A}) = \mathbf{P}^{(d)}_{I_1,\cdots,I_D} \times \mathrm{vec}(\mathbf{A}_{(d)})$ where $\mathbf{P}^{(d)}_{I_1,\cdots,I_D}$ are permutation matrices such that $\mathbf{P}^{(d)-1}_{I_1,\cdots,I_D} = \mathbf{P}^{(d)\mathrm{T}}_{I_1,\cdots,I_D}$.

1.4 Dissertation outline

The main objective of this dissertation is to answer some fundamental questions that appear in different domains of statistics due to real-life situations. Let us dive into these questions one by one and briefly introduce them.

Question 1. How should we handle the dense functional response in *quadratic inference method*?

We consider the problem of estimation for constant linear effect models in semi-parametric functional regression with functional response, where each response curve is decomposed into

the overall mean function indexed by a covariate function with constant regression parameters and random error. In Chapter 2, we provide an alternative solution using a popular method for the analysis of correlated data, viz., the quadratic inference approach for such models. Here, we use the estimated basis functions which are being estimated non-parametrically. Therefore, the proposed method can be easily implemented without assuming any working correlation structure. Moreover, we achieve a parametric $\sqrt{n}$-convergence rate under the proper choice of bandwidth when the number of repeated measurements per trajectory is larger than n^{a_0} where n is the number of trajectories and establish the asymptotic normality of the resulting estimator. The performance of the proposed method is compared with that of existing methods through extensive simulation studies. Real data analysis is also carried out to demonstrate the proposed method.

Question 2. How should the *heteroskedastic* functional data be analyzed?

Motivated by recent work on diffusion tensor imaging, we propose a novel varying-coefficient model in Chapter 3. We develop a multi-step estimation procedure to simultaneously estimate the varying-coefficient functions using a local linear generalized method of moments (GMM) based on continuous moment conditions. To incorporate spatial dependence, the continuous moment conditions are first projected onto eigen-functions and then combined by weighted eigen-values. This approach solves the challenges of using an inverse covariance operator directly. We propose an optimal instrumental variable that minimizes the asymptotic variance function among the class of all local linear GMM estimators and it outperforms the initial estimates which do not incorporate the spatial dependence. It is shown that with our proposed method, accuracy of the estimation is significantly improved under heteroskedasticity conditions. We investigate the asymptotic properties of the initial and proposed estimators. Extensive simulation studies illustrate the finite sample performance and the analysis of real data confirms the efficacy of the proposed method.

Question 3. How should functional regression be preformed for complex structured data such as

tensor?

All neuroimaging modalities have their own strengths and limitations. A current trend is towards interdisciplinary approaches that use multiple imaging methods to overcome limitations of each method in isolation. At the same time neuroimaging data is increasingly being combined with other non-imaging modalities, such as behavioral and genetic data. The data structure of many of these modalities can be expressed as time-varying multidimensional arrays (tensors), collected at different time-points on multiple subjects. In Chapter 4, we consider a new approach for the study of neural correlates in the presence of tensor-valued brain images and tensor-valued predictors, where both data types are collected over the same set of time-points. We propose a time-varying tensor regression model with an inherent structural composition of responses and covariates. Regression coefficients are expressed using the B-spline technique, and basis function coefficients are estimated using CP-decomposition by minimizing a penalized loss function. We develop a varying-coefficient model for the tensor-valued regression model, where both predictors and responses are modeled as tensors. This development is a non-trivial extension of function-on-function concurrent linear models for complex and large structural data where the inherent structures are preserved. In addition to the methodological and theoretical development, the usefulness of the proposed method based on both simulated and real data analysis (e.g., the combination of eye-tracking data and functional magnetic resonance imaging (fMRI) data) is also discussed.

Putting it all together, in this chapter, we have introduced the concepts of functional data, its computational framework, required mathematical notations, definitions and real-life applications thereby establishing a foundation of the upcoming chapters of this dissertation.

CHAPTER 2

IMPROVING QUADRATIC INFERENCE APPROACH FOR FUNCTIONAL RESPONSES

2.1 Introduction

The key characteristic of longitudinal data analysis (LDA) is the collection of repeated measurements on the same set of individuals over multiple time-points, thus allowing study of changes in responses over time and identification of factors that influence those changes. Unlike cross-sectional studies, where one can estimate only the "between-individual" responses, since they are measured at a single time-point; in LDA, it is possible to capture the "with-in individual" changes as repeated measurements on each individual are available. Moreover, longitudinal data is always observed as clusters, where each cluster pertains to repeated measurements obtained from each individual. Although longitudinal studies are performed for data that are observed sparsely over irregular time-points, such studies do not suffice when voluminous data are observed in a continuum domain. As technologies advance, this type of data is being observed more often, so sophisticated methods are needed to handle it. Since functional data are natural generalizations to multivariate data from finite to infinite dimension, functional data analysis (FDA) has turned out to be an important methodological tool.

In the following two paragraphs, we will present a brief review of some significant research in the past decades that led to the current research. In LDA, the data are generally observed with noise for measurements at each time-point (Taris, 2000; Diggle et al., 2002; Hedeker and Gibbons, 2006; Hand and Crowder, 2017). Moreover, a few repeated measurements are required in LDA and the data are observed sparsely with noise. On the other hand, in FDA, data are densely observed as a continuous-time stochastic process without noise (Zhang and Wang, 2016). Often, the sampling plan can have an effect on the performance of the estimation procedures and inference (Hall and Hosseini-Nasab, 2006). In some situations, data are typically functions by

nature and are observed densely over time. Chiou et al. (2003) proposed a class of semi-parametric functional regression models to describe the influence of vector-valued covariates on a sample of the response curve. When data collection leads to experimental error, smoothing is performed at closely spaced time-points in order to reduce the effect of noise. The current developments of functional regression techniques have been rigorously studied in Fan et al. (1999); Hall et al. (2007); Chen et al. (2019). The applicability of FDA spans across various scientific domains such as medical imaging, speech recognition, growth curves, climatology, price index analysis, and many more. Some recent literature on applications of FDA include Ramsay and Silverman (2005); Ferraty and Vieu (2006); Ramsay and Silverman (2007); Zhang (2013); Hsing and Eubank (2015); Morris (2015); Wang et al. (2016); Kokoszka and Reimherr (2017).

Methodologically, in LDA in the past few years, the generalized estimating equation (GEE) technique proposed by Liang and Zeger (1986) has been extensively used for estimation of parameters. Although it is an efficient technique, the GEE is unable to estimate the parameters of interest efficiently when the correlation matrix of covariates is not specified correctly. Hence, without requiring the estimation of the correlation parameters, the quadratic inference function (QIF) approach proposed by Qu et al. (2000) is useful for parameter estimation in longitudinal studies (Diggle et al., 2002) and cluster randomized trials (Turner et al., 2017). By representing the inverse of the working correlation matrix in terms of linear combinations of the basis matrices and involving multiple sets of score functions, the QIF approach has improved efficiency over GEE when the working correlation matrix is not specified correctly. Although it maintains the same efficiency as in the situation where the working correlation matrix is specified correctly, the QIF method is not independent of the choice of the working correlation matrix. A QIF method-based approach to varying-coefficient models for longitudinal data was proposed by Qu and Li (2006). The related work of Bai et al. (2008) is an extension of QIF for the partial linear model. An alternative method was presented in Yu et al. (2020) where each set of score equations was solved separately and their solutions were combined afterwards; thereby providing results on inference for an optimally weighted estimator and extending those insights to the general setting with over-

identified estimating equations. Zhao et al. (2020, 2021) proposed variable selection method for the varying-coefficient model when some of the covariates were contaminated with additive errors based on bias-corrected penalized QIFs that are defined by combining the bias function approximation to the coefficient functions and bias-corrected QIF with shrinkage estimation. Zhou and Qu (2012) proposed QIF based strategy which minimizes the norm of the difference between two estimating functions based on empirical correlation information. Tian et al. (2014) focused on the selection of variables for the semi-parametric varying-coefficient model based on the combination of the approximated basis function and the QIFs. A longitudinal principal component analysis was proposed in Kinson et al. (2020) based on eigen-decomposition of random effects, while data on correlations information of multivariate observations over time were decomposed by nonparametric splines. Zheng et al. (2018) proposed a method based on a time-varying linear representation of the inverse of the correlation matrix which is projected over the span of basis matrices.

The fundamental limitations that all the above-mentioned powerful techniques suffer from are: (1) all the above methodologies require prior information on the working correlation structure; and (2) performance of the classical QIF approach is unknown for dense functional data. Our study is motivated by problems from multiple real-data applications that involve dense functional data when information on the working correlation structure is lacking. Let us discuss two motivating examples that we will use to illustrate the proposed method in this chapter (see Section 2.5 for more details).

- ***Beijing2017-data*** **example** - In different locations in China, particulate matter (PM) with diameter less than 2.5 micrometer is collected over different time-points. Scientists are interested in knowing the linear dependence of the pollution factor $PM_{2.5}$ with other atmospheric chemicals (Liang et al., 2015). Figure 2.1 pictorially demonstrates the readings of $PM_{2.5}$ for the given locations over several hourly time-points; therefore, dense functional data analysis can be implemented.

- ***Apnea-data*** **example** - In neuroimaging data analysis, scientists are interested in modelling

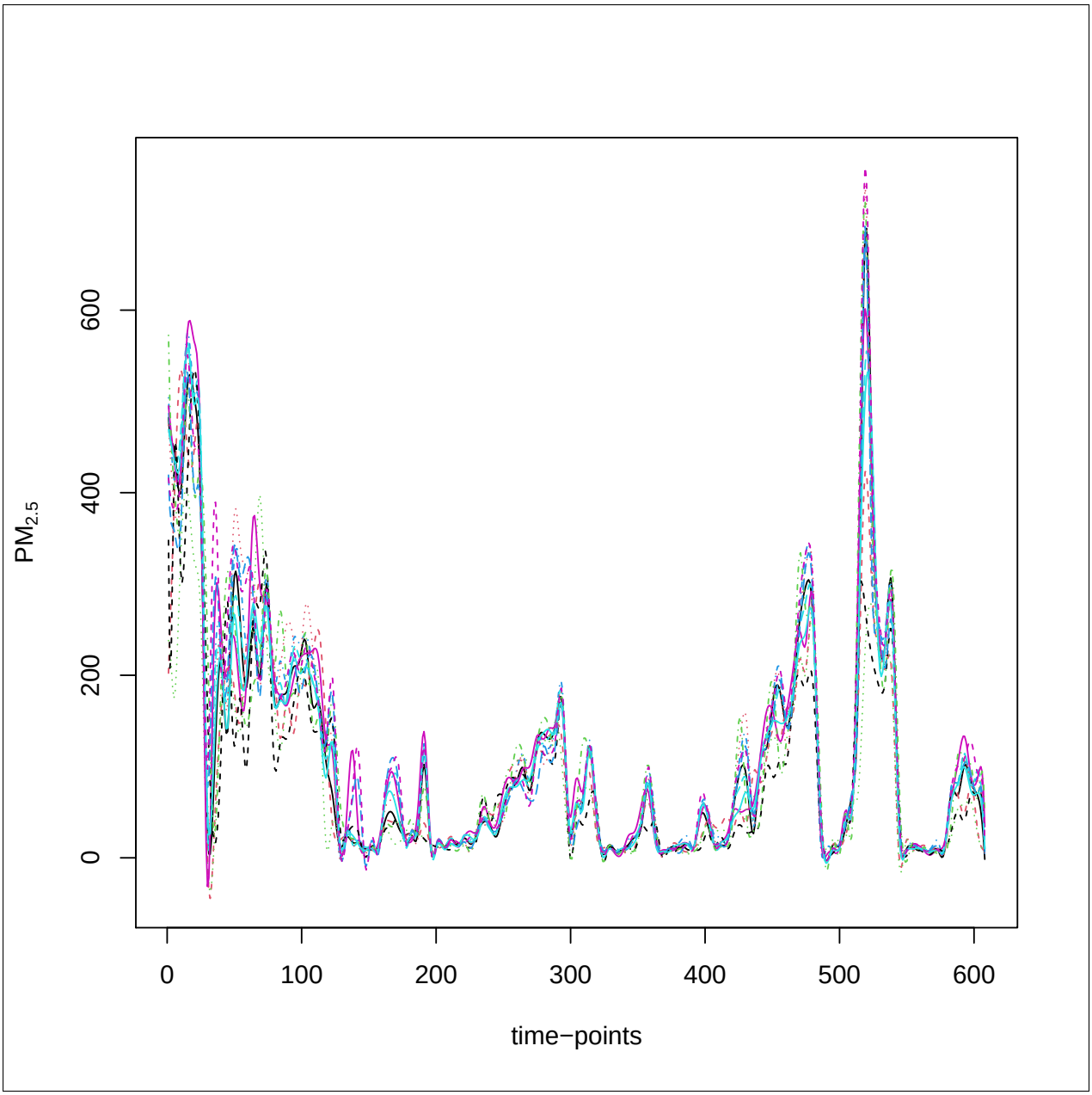

Figure 2.1 *Beijing2017-data*: Reading of hourly $PM_{2.5}$ measures for twelve different locations over 608 hourly time-points during January 2017.

the change of responses among voxels in each region of interest (ROI) of the human brain. Therefore, we can fit a linear regression model and compare the estimated coefficient across each ROIs. Needless to say, there exist a large number of voxels and the responses change smoothly across the voxels in each ROI, therefore, the data are functional and dense in nature. In recent literature, Xiong et al. (2017) investigates white matter structural alterations

using diffusion tensor imaging (DTI) in obstructive sleep apnea (OSA) patients. Here, the change of DTI parameters such as fractional anisotropy (FA) with interaction of count of lapses obtained from the Psychomotor Vigilance Task and voxel locations are investigated and compared to the results obtained in each ROIs.

We propose a data-driven way to select the working covariance matrix and express the inverse of the covariance function in terms of the empirical eigen-functions of the covariance operator. The covariance operator can be estimated as in Hsing and Eubank (2015) and other related methods based on functional principal component analysis (FPCA) as found in Dauxois et al. (1982); Yao et al. (2005); Hall and Hosseini-Nasab (2006); Hall et al. (2007); Li and Hsing (2010). Note that the estimation of the eigen-functions creates some error in the proposed estimation method. In this chapter, we try to answer the following question: *while we estimate the eigen-functions from the data, is the estimation of coefficient vectors in a semi-parametric problem* $\sqrt{n}$*- consistent in dense functional data, and can we achieve asymptotic normality?* The advantages of our proposed method are the following: First, our method preserves the good properties of the QIF method and is easier to implement since the eigen-functions can be estimated using the existing packages in statistical softwares such as R. Second, under some mild conditions, our proposed estimator can obtain the optimal convergence rate and is asymptotically normally distributed with less variance as compared to the classical QIF methods. Third, asymptotic results show the estimation accuracy of the coefficient in semi-parametric functional model, therefore, making the influence of the dimension reduction step using FPCA redundant. The error in the estimation of the eigen-functions contributes to the error in the estimation of the parameters. Under some mild bandwidth conditions, the above-mentioned error contribution is of the same order of magnitude as error in parameter estimation if eigen-functions are known in advance.

The rest of the chapter is organized as follows. In Section 2.2 we introduce the basic concept of QIF along with our proposed method. The asymptotic results for the proposed estimator are presented in Section 2.3. In Section 2.4, we demonstrate the performance for finite samples. We also apply the proposed method to real data-sets in Section 2.5. We conclude in some remarks in

Section 2.6. All technical proofs are given in Section 2.7.

2.2 Functional response model and estimation procedure

2.2.1 Basic model

To analyze longitudinal data, a straightforward application of a generalized linear model (GLM) (McCullagh and Nelder, 1989) for single response variables is not applicable due to the lack of independence between repeated measures. To account for the high correlation in the longitudinal data, some special techniques are required. A seminal work by Liang and Zeger (1986) proposed the use of GLM for the analysis of longitudinal data. The model we consider in this chapter is commonly observed in spatial modeling, where associations among variables do not change over the functional domain (see Zhang and Banerjee (2021) and references therein); this is termed a constant linear effects model. In this chapter, the variable "time" is used as a functional domain variable.

Let $y(t)$ be the response variable at time-point t and $\mathbf{x}(t)$ be p-dimensional covariates observed at time $t \in \mathcal{T}$ where $\mathcal{T} = \left[\underline{a}, \overline{a}\right], -\infty < \underline{a} < \overline{a} < \infty$ is the spectrum of the time-points. Without loss of generality, assume that $\underline{a} = 0$ and $\overline{a} = 1$ in the rest of this chapter. The stochastic process $y(t)$ is square-integrable with marginal mean $\mathbb{E}\{y(t)|\mathbf{x}(t)\}$ and finite covariance function; the regression parameter $\boldsymbol{\beta}$ is unknown and is to be efficiently estimated. Thus, linear models with longitudinal data have the following expression.

$$y(t) = \mathbf{x}(t)^{\mathrm{T}}\boldsymbol{\beta} + e(t) \tag{2.1}$$

where, the stochastic process $y(t)$ is decomposed into two parts: one is the mean function $\mu(t) = \mathbf{x}(t)^{\mathrm{T}}\boldsymbol{\beta}$ that depends on time-varying covariates and vector-valued coefficient vector $\boldsymbol{\beta}$, and other is the random error part $e(t)$ where $\mathbb{E}\{e\} = 0$ and has finite second-order covariance. Let y_i be i.i.d. copies of the stochastic process and for each individual, the measurements are taken on m_i discrete time-points T_{ij} for $j = 1, \cdots, m_i; i = 1, \cdots, n$. Therefore, at time T_{ij}, we observe a $m_i \times 1$ response vector $y_i(T_{ij})$ and corresponding covariates $\mathbf{x}_i(T_{ij})$ for the i-th subject. We assume that

m_i's are all of the same order as $m = n^a$ for some $a \geq 0$, thus m_i/m are bounded below and above by some constants. Functional data are considered to be sparse depending on the choice of a (Hall and Hosseini-Nasab, 2006). Data with bounded m or $a = 0$ are called sparse functional data, and if $a \geq a_0$ where a_0 is a transition point are called dense functional data. Moreover, the regions $(0, a_0)$ are sometimes referred to as moderately dense. Furthermore, we denote y_{ij} and $\mathbf{x}_{ij}$ as $y_i(T_{ij})$ and $\mathbf{x}_i(T_{ij})$ respectively. $(y_{i1}, \cdots, y_{im_i})^{\mathrm{T}}$ and $(\mu_{i1}, \cdots, \mu_{im_i})^{\mathrm{T}}$ are m_i component vectors, denoted as $\mathbf{y}_i$ and $\boldsymbol{\mu}_i$ respectively. The derivative of $\boldsymbol{\mu}$, denoted as $\dot{\boldsymbol{\mu}}$, is a $m_i \times p$ matrix.

In the classical problem of GEE, we estimate $\boldsymbol{\beta}$ by solving the quasi-likelihood equations (Liang and Zeger, 1986):

$$\sum_{i=1}^{n} \dot{\boldsymbol{\mu}}_i^{\mathrm{T}} \mathbf{V}_i^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_i) = 0 \tag{2.2}$$

We denote $\mathbf{V}_i = \nu \mathbf{A}_i^{1/2} \mathbf{R}_i(\rho) \mathbf{A}_i^{1/2}$ where $\mathbf{R}_i(\rho)$ is the working correlation matrix, ν is an over-dispersion parameter and $\mathbf{A}_i$ is a diagonal matrix where entries are marginal variances $\mathrm{Var}(y_{i1}), \cdots,$ $\mathrm{Var}(y_{im_i})$. In this article, we simply set $\nu = 1$ while the extension to a general ν is straightforward. The GEE approach is robust in the sense that it does not require the true knowledge of the likelihood function.

Note that, in practice, the prior knowledge of the working correlation matrix is not known, and the estimation of the coefficient is influenced by its choice. Therefore, Qu et al. (2000) suggested an expansion of the inverse of the working correlation matrix as $\mathbf{R}(\rho)^{-1} = \sum_{k=1}^{\kappa_0} a_k(\rho) \mathbf{M}_k$ where $\mathbf{M}_k$ are some basis matrices. Zhou and Qu (2012) modified linear representation by grouping the basis matrices into an identity matrix and some symmetric basis matrices. For example, if the working correlation matrix is exchangeable/ compound symmetric, $\mathbf{R}(\rho)^{-1} = c_1 \mathbf{I}_m + c_2 \mathbf{J}_m$ where $\mathbf{I}_m$ is the $m \times m$ identity matrix and $\mathbf{J}_m$ is the $m \times m$ matrix such that 0 is in diagonal and 1 is in off-diagonal positions. On the other hand, for first-order auto-regressive correlation matrix, $\mathbf{R}(\rho)^{-1} = c_1 \mathbf{I}_m + c_2 \mathbf{J}_m^{(1)} + c_3 \mathbf{J}_m^{(2)}$ where $\mathbf{J}_m^{(1)}$ is a matrix with 1 in the two main off-diagonal positions and 0 otherwise, $\mathbf{J}_m^{(2)}$ is a matrix such that 1 is in the corner positions, viz. $(1, 1)$ and (m, m) and 0 elsewhere. Here, c_ks are real constants that depend on the nuisance parameter ρ. Therefore,

Equation (2.2) reduces to the linear combination of the score vectors:

$$\bar{\mathbf{g}}(\beta) = \frac{1}{n}\sum_{i=1}^{n}\mathbf{g}_i(\beta) = \begin{bmatrix} \frac{1}{n}\sum_{i=1}^{n}\dot{\boldsymbol{\mu}}_i^{\mathrm{T}}\mathbf{A}_i^{-1/2}\mathbf{M}_1\mathbf{A}_i^{-1/2}(\mathbf{y}_i-\boldsymbol{\mu}_i) \\ \vdots \\ \frac{1}{n}\sum_{i=1}^{n}\dot{\boldsymbol{\mu}}_i^{\mathrm{T}}\mathbf{A}_i^{-1/2}\mathbf{M}_{\kappa_0}\mathbf{A}_i^{-1/2}(\mathbf{y}_i-\boldsymbol{\mu}_i) \end{bmatrix} \tag{2.3}$$

Due to the higher dimension of $\bar{\mathbf{g}}$, Qu et al. (2000) used the generalized method of moments (GMM) (Hansen, 1982) for which the method of estimation boils down to minimization of the quadratic inference function $Q(\boldsymbol{\beta}) = n\bar{\mathbf{g}}(\boldsymbol{\beta})^{\mathrm{T}}\widehat{\mathbf{C}}(\boldsymbol{\beta})^{-1}\bar{\mathbf{g}}(\boldsymbol{\beta})$ where $\widehat{\mathbf{C}}(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^{n}\bar{\mathbf{g}}_i(\boldsymbol{\beta})\bar{\mathbf{g}}_i(\boldsymbol{\beta})^{\mathrm{T}}$ is the sample covariance matrix of Equation (2.3). In order to obtain the solution of $\boldsymbol{\beta}$, Newton-Raphson method is used which iteratively updates the value of $\boldsymbol{\beta}$.

2.2.2 Incorporating eigen-functions in QIF

Now, due to standard Karhunen-Loève expansions of $e_i(t) = y_i(t) - \mu_i(t)$ (Karhunen, 1946; Loève, 1946)

$$e_i(t) = \sum_{r=1}^{\infty}\xi_{ir}\phi_r(t) \tag{2.4}$$

where independently distributed random variables $\xi_{ir} \sim (0, \lambda_r)$ for ordered eigen-values λ_r such that $\lambda_1 \geq \lambda_2 \geq \cdots \geq 0$ and ϕ_rs are orthonormal eigen-functions such that $\int \phi_r(t)\phi_l(t) = \mathbf{1}(r = l)$. We extract the main directions of the variation of the response variables using FPCA. In this situation, we take the first κ_0 terms, which provide a good approximation of the infinite sum in Equation (2.4) by considering that the majority of the variations in the data are contained in the subspace spanned by few eigen-functions (Chen et al., 2019). For finite $\kappa_0 \geq 1$, we, therefore, consider the rank-κ_0 FPCA model,

$$\mathbb{E}\{y(t)|\mathbf{x}(t)\} = \mu(t) + \sum_{r=1}^{\kappa_0}\mathbb{E}\{\xi_r|\mathbf{x}(t)\}\phi_r(t) \tag{2.5}$$

An analogue of the truncated empirical version of Equation (2.22) defined in Section 2.7 and Equation (2.4) can be provided easily and we discuss the proposed method based on this truncated version. Moreover, we discuss how to choose κ_0 in our situation in Section 2.3 in detail.

In this chapter, we propose a data-driven way to compute the basis matrices to obtain the approximate inverse of $\mathbf{V}$ as discussed earlier. In this approach, it is enough to find the eigen-functions to construct a GEE. Let us define

$$\bar{\mathbf{g}}(\boldsymbol{\beta}) = \begin{bmatrix} \frac{1}{n}\sum_{i=1}^{n} \dot{\boldsymbol{\mu}}_i^{\mathrm{T}} \widehat{\boldsymbol{\Phi}}_{i1}(\mathbf{y}_i - \boldsymbol{\mu}_i) \\ \vdots \\ \frac{1}{n}\sum_{i=1}^{n} \dot{\mu}_i^{\mathrm{T}} \widehat{\boldsymbol{\Phi}}_{i\kappa_0}(\mathbf{y}_i - \boldsymbol{\mu}_i) \end{bmatrix} \tag{2.6}$$

where for $k = 1, \cdots, \kappa_0$, we define $\widehat{\boldsymbol{\Phi}}_{ik} = \left(m_i^{-2} \widehat{\phi}_k(T_{ij}) \widehat{\phi}_k(T_{ij'}) \right)_{j,j'=1,\cdots,m_i}$. Since the dimension of $\overline{\mathbf{g}}$ in Equation (2.6) is greater than the number of parameters to estimate, instead of setting $\overline{\mathbf{g}}$ to zero, we minimize the following quadratic function.

$$\widehat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \mathcal{Q}(\boldsymbol{\beta}) \text{ where } \mathcal{Q}(\boldsymbol{\beta}) = n\bar{\mathbf{g}}(\boldsymbol{\beta})^{\mathrm{T}} \widehat{\mathbf{C}}(\boldsymbol{\beta})^{-1} \bar{\mathbf{g}}(\boldsymbol{\beta}) \tag{2.7}$$

where, $\widehat{\mathbf{C}}(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{g}_i(\boldsymbol{\beta}) \mathbf{g}_i(\boldsymbol{\beta})^{\mathrm{T}}$. For the existence of $\widehat{\mathbf{C}}^{-1}$ we need the additional restriction: $n \geq dim(\mathbf{g}_i) = p \times \kappa_0$ where κ_0 is the number of eigen-functions. Under the given set-up, by Equation (8) in Qu et al. (2000) the estimating equation for $\boldsymbol{\beta}$ will be

$$\dot{\mathcal{Q}}(\boldsymbol{\beta}) \approx 2\dot{\bar{\mathbf{g}}}(\boldsymbol{\beta})^{\mathrm{T}} \widehat{\mathbf{C}}(\boldsymbol{\beta})^{-1} \bar{\mathbf{g}}(\boldsymbol{\beta}) \tag{2.8}$$

For obtaining the solution of the above equation, we use a Newton-like method. In practice, the standard Newton method does not lead to a decrease in the objective function, that is, at each step of the iteration, there is no guarantee that $\mathcal{Q}(\boldsymbol{\beta}_{s+1}) < \mathcal{Q}(\boldsymbol{\beta}_s)$. Therefore, we use the following algorithm to estimate $\boldsymbol{\beta}$ using the Quasi-Newton method with halving (Givens and Hoeting, 2012).

2.2.3 Estimation of eigen-functions

Estimation of eigen-functions is an important step in our proposed quadratic inference technique. In general, FPCA plays an important role as a dimension reduction technique in functional data analysis. Some important theories on FPCA have been developed in recent years. In particular, Hall and Hosseini-Nasab (2006) proved various asymptotic expressions for FPCA for densely observed functional data. Later, Hall and Hosseini-Nasab (2009) showed more common theoretical

Algorithm 2.1 Estimation of $\boldsymbol{\beta}$ using the Quasi-Newton method with halving.

Data: $\widehat{\boldsymbol{\beta}}_0$ (initial estimates) and calculate $\mathcal{Q}(\widehat{\boldsymbol{\beta}_0})$, $\dot{\mathcal{Q}}(\widehat{\boldsymbol{\beta}_0})$, $\ddot{\mathcal{Q}}(\widehat{\boldsymbol{\beta}_0})$ respectively. ϵ_0 (threshold, a small number) and max.count (maximum number of repetition)

Result: Estimate $\boldsymbol{\beta}$ using proposed method

Calculate: $\widehat{\boldsymbol{\beta}}_1 \leftarrow \widehat{\boldsymbol{\beta}}_0 - \ddot{\mathcal{Q}}(\widehat{\boldsymbol{\beta}_0})^{-1}\dot{\mathcal{Q}}(\widehat{\boldsymbol{\beta}_0})$

while *Error* $> \epsilon_0$ **do**

  Calculate: $\dot{\mathcal{Q}}(\widehat{\boldsymbol{\beta}}_1)$ and $\ddot{\mathcal{Q}}(\widehat{\boldsymbol{\beta}}_1)$ based on $\widehat{\boldsymbol{\beta}}_1$

  Initialise: $r_0 = 1$

  $\widehat{\boldsymbol{\beta}}_2 \leftarrow \widehat{\boldsymbol{\beta}}_1 - r_0\ddot{\mathcal{Q}}(\widehat{\boldsymbol{\beta}}_1)^{-1}\dot{\mathcal{Q}}(\widehat{\boldsymbol{\beta}}_1)$

  Calculate $\mathcal{Q}(\widehat{\boldsymbol{\beta}}_1)$ and $\mathcal{Q}(\widehat{\boldsymbol{\beta}}_2)$ based on $\widehat{\boldsymbol{\beta}}_1$ and $\widehat{\boldsymbol{\beta}}_2$ respectively using proposed method

  while $\mathcal{Q}(\widehat{\boldsymbol{\beta}}_2) > \mathcal{Q}(\widehat{\boldsymbol{\beta}}_1)$ **do**

    $r_0 \leftarrow r_0/2$

    $\widehat{\boldsymbol{\beta}}_2 \leftarrow \widehat{\boldsymbol{\beta}}_1 - r_0\ddot{\mathcal{Q}}(\widehat{\boldsymbol{\beta}}_1)^{-1}\dot{\mathcal{Q}}(\widehat{\boldsymbol{\beta}}_1)$

    Calculate $\mathcal{Q}(\widehat{\boldsymbol{\beta}}_1)$ and $\mathcal{Q}(\widehat{\boldsymbol{\beta}}_2)$ based on $\widehat{\boldsymbol{\beta}}_1$ and $\widehat{\boldsymbol{\beta}}_2$ respectively using proposed method

  end

  Calculate: Error $= \|\widehat{\boldsymbol{\beta}}_2 - \widehat{\boldsymbol{\beta}}_1\|^2$

  $\widehat{\boldsymbol{\beta}}_0 \leftarrow \widehat{\boldsymbol{\beta}}_1$

  $\widehat{\boldsymbol{\beta}}_1 \leftarrow \widehat{\boldsymbol{\beta}}_2$

end

arguments, including the effect of gap between eigen-value (a.k.a., spacing) on the property of eigen-value estimators. In Li and Hsing (2010), uniform rates of convergence of mean and covariance functions are given, which are equipped for all possible choices/scenarios of m_is. In this section, we adopt the estimation of covariance functions mostly from Li and Hsing (2010).

Note that the error process $e(t)$ has mean zero, defined on compact set $\mathcal{T} = [0, 1]$ satisfying $\int_{\mathcal{T}} \mathbb{E}\{e^2\} < \infty$. The functional principal components can be constructed via the covariance function $R(s, t)$ defined as

$$R(s, t) = \mathbb{E}\{e(s)e(t)\} \tag{2.9}$$

which is assumed to be square-integrable. This function R induces the kernel operator $\mathscr{F}$ as defined in Sub-section 2.7.1. An empirical analogue of the spectral decomposition of R can be obtained

$$\widehat{R}(s, t) = \sum_{r=1}^{\infty} \widehat{\lambda}_r \widehat{\phi}_r(s)\widehat{\phi}_r(t) \tag{2.10}$$

where the random variables $\widehat{\lambda}_1 \geq \widehat{\lambda}_2 \geq \cdots \geq 0$ are the eigen-values of the estimated operator $\widehat{\mathscr{F}}$ and the corresponding sequence of eigen-functions are $\widehat{\phi}_1, \widehat{\phi}_2, \cdots$. Further, assume that $\int_{\mathscr{T}} \phi_r \widehat{\phi}_r \geq 0$ to avoid the issue regarding change of sign (Hall and Hosseini-Nasab, 2006) for practical comparison of eigen-functions, otherwise there is no impact on the convergence rate of eigen-functions and hence the proposed estimators. Our proposed method can be generalized for finite ties of the true eigen-values λ_r but to avoid more technicalities, we assume that the eigen-values are distinct.

Suppose that T_{ij} are observational points with a positive density function $f_T(\cdot)$. Assume $m_i \geq 2$ and define $N = \sum_{i=1}^{n} N_i$ where $N_i = m_i(m_i - 1)$. This approach is based on local linear smoother, which is popular in functional data analysis, including Fan and Gijbels (1996); Li and Hsing (2010) among many others. Let $K(\cdot)$ be a symmetric probability density function on $[-1, 1]$, which is used as kernel, and $h > 0$ be bandwidth; thus the re-scaled kernel function is defined as $K_h(\cdot) = \frac{1}{h}K(\cdot)$. Therefore, for given $s, t \in \mathscr{T}$, choose $(\widehat{a}_0, \widehat{b}_1, \widehat{b}_2)$ be the minimizer of the following equation.

$$\frac{1}{n}\sum_{i=1}^{n}\frac{1}{N_i}\sum_{\substack{j_1=1 \\ j_1 \neq j_2}}^{m_i}\sum_{j_2=1}^{m_i}\{e_i(T_{ij_1})e_i(T_{ij_2}) - a_0 - b_1(T_{ij_1} - s) - b_2(T_{ij_2} - t)\}^2 K_h\left(T_{ij_1} - s\right) K_h\left(T_{ij_2} - t\right) \quad (2.11)$$

Thus, we estimate $R(s, t) = \mathbb{E}\{e(s)e(t)\}$ using the quantity $\widehat{a}_0$, viz., $\widehat{R}(s, t) = \widehat{a}_0$. The operator $\widehat{\mathscr{F}}$ is in general positive semi-definite and the estimated eigen-values $\widehat{\lambda}_r$ are non-negative; indeed, $\widehat{R}$ is symmetric. Along with the lines in the existing literature, we define the following.

- $S_{a,b}(s, t) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{N_i}\sum_{\substack{j_1=1 \\ j_1 \neq j_2}}^{m_i}\sum_{j_2=1}^{m_i}\left(\frac{T_{ij_1} - s}{h}\right)^a\left(\frac{T_{ij_2} - t}{h}\right)^b K_h(T_{ij_1} - s)K_h(T_{ij_2} - t)$
- $\mathscr{R}_{a,b}(s, t) = \frac{1}{N_i}\sum_{\substack{j_1=1 \\ j_1 \neq j_2}}^{m_i}\sum_{j_2=1}^{m_i}\left(\frac{T_{ij_1} - s}{h}\right)^a\left(\frac{T_{ij_2} - t}{h}\right)^b e_i(T_{ij_1})e_i(T_{ij_2})K_h(T_{ij_1} - s)K_h(T_{ij_2} - t)$
- $\mathscr{A}_1 = S_{20}S_{02} - S_{11}^2$, $\mathscr{A}_2 = S_{10}S_{02} - S_{01}S_{11}$, and $\mathscr{A}_3 = S_{01}S_{20} - S_{10}S_{11}$
- $\mathscr{B} = \mathscr{A}_1 S_{00} - \mathscr{A}_2 S_{10} - \mathscr{A}_3 S_{01}$

Therefore, $\widehat{R}(s, t) = (\mathscr{A}_1\mathscr{R}_{00} - \mathscr{A}_2\mathscr{R}_{10} - \mathscr{A}_3\mathscr{R}_{01})\mathscr{B}^{-1}$.

2.3 Asymptotic properties

In this section, we study the asymptotic properties of the proposed estimator. Let us introduce some notation. Assume that m_is are all of the same order, i.e., $m \equiv m(n) = n^a$ for some $a \geq 0$. Define $d_{n1}(h) = h^2 + h\overline{m}/m$ and $d_{n2}(h) = h^4 + h^3\overline{m}/m + h^2\overline{\overline{m}}/m^2$ where $\overline{m} = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} m/m_i$ and $\overline{\overline{m}} = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n}(m/m_i)^2$. Denote $\delta_{n1}(h) = \left\{d_{n1}(h)\log n/(nh^2)\right\}^{1/2}$ and $\delta_{n2}(h) = \left\{d_{n2}(h)\log n/(nh^4)\right\}^{1/2}$. Further, $\nu_{a,b} = \int t^a K^b(t)dt$. Define $\mathbf{W} = (\boldsymbol{\phi}(t_1)^{\mathrm{T}}, \cdots, \boldsymbol{\phi}(t_m)^{\mathrm{T}})^{\mathrm{T}}$ is a matrix of order $m \times \kappa_0$ obtained after stacking all $\boldsymbol{\phi}_k$s and random components $\xi_i = (\xi_{i1}, \cdots, \xi_{i\kappa_0})^{\mathrm{T}}$. Further, $\boldsymbol{\xi}$ has mean zero and variance $\boldsymbol{\Lambda}$ which is a diagonal matrix with components $\lambda_1, \cdots, \lambda_{\kappa_0}$. The sign " $\lesssim$ " indicates that the left-hand side of the inequality is bounded by the right-hand side up to a multiplicative positive constant, i.e. for two positive variables f_1 and f_2 we define $f_1 \lesssim f_2$ as $f_1 \leq Cf_2$ where C is a positive constant not involving n. The following conditions are needed for further discussion of the asymptotic properties.

(C1) Kernel function $K(\cdot)$ is a symmetric density function defined on bounded support $[-1, 1]$.

(C2) Density function f_T of T is bounded above and away from infinity. Also the density function is bounded below away from zero. Moreover, f is differentiable and the derivative is continuous.

(C3) $R(s, t)$ is twice differentiable and all second order partial derivatives are bounded on $[0, 1]^2$.

(C4) $\mathbb{E}\{\sup_{t\in[0,1]} |e(t)|^\gamma\} < \infty$ and $\mathbb{E}\{\sup_{t\in[0,1]} |\mathbf{x}_i(t)|^{2\gamma}\} < \infty$ for some $\gamma \in (4, \infty)$.

(C5) $h \to 0$ as $n \to \infty$ such that $d_{n1}^{-1}(\log n/n)^{1-2/\gamma} \to 0$ and $d_{n2}^{-1}(\log n/n)^{1-4/\gamma} \to 0$ for $\gamma \in (4, \infty)$.

(C6) Condition for eigen-components.

a) for each $1 \leq k < r < \infty$ and for non-zero finite generic constant C_0,

$$\frac{\max\{\lambda_k, \lambda_r\}}{|\lambda_k - \lambda_r|} \leq C_0 \frac{\max\{k, r\}}{|k - r|} \tag{2.12}$$

b) For some $\alpha > 0$, with the condition $V_r \lambda_r^{-2} r^{1+\alpha} \to 0$ as $r \to \infty$ where $V_r = \mathbb{E}\{\int \dot{\mu}(t)\phi_r(t)dt\}^2$.

The above two conditions hold if $\lambda_r = r^{-\tau_1}\Lambda(r)$ and $V_r = r^{-\tau_2}\Gamma(r)$ for slowly varying functions Λ and Γ where $\tau_2 > 1 + 2\tau_1 > 3$.

c) $\int \phi_k^4(t)dt$ and $\int \dot{\mu}^2(t)\phi_k^2(t)dt$ are finite for all $k \geq 1$.

(C7) $\widehat{\mathbf{C}}(\boldsymbol{\beta})$ converges almost surely to an invertible matrix $\mathbf{C}_0 = \mathbb{E}\{\mathbf{g}(\boldsymbol{\beta}_0)\mathbf{g}(\boldsymbol{\beta}_0)^{-1}\}$.

(C8) Conditions for h and κ_0. For $\tau = \alpha + \tau_1$,

a) If $a > 1/4$, $\kappa_0 = O(n^{1/(3-\tau)})$ and $n^{-1/4} \lesssim h \lesssim n^{-(a+1)/5}$

b) If $a \leq 1/4$, $\kappa_0 = O(n^{4(1+a)/5(3-\tau)})$ and $h \lesssim n^{-1/4}$

Remark 2.3.1. *Condition (C1) is commonly used in non-parametric regression. The bound condition for the density function of time-points has the standard Condition (C2) for random design. Similar results can be obtained for fixed design where the grid-points are pre-fixed according to the design density* $\int_0^{T_j} f(t)dt = j/m$ *for* $j = 1, \cdots, m$, *for* $m \geq 1$. *Furthermore, it is important to note that this approach does not involve the requirement to obtain sample path differentiation when we invoke the estimation of eigenfunctions from Li and Hsing (2010). Therefore, the method could be applicable for Brownian motion which has a continuously non-differentiable sample path. To expand in Taylor series, Condition (C3) is required, and is also common in non-parametric regression. Condition (C4) is required for a uniform bound for certain higher-order expectations to show uniform convergence. This is a similar condition adopted from Li and Hsing (2010). Smoothness conditions in (C5) and (C8) are common in kernel smoothing and functional data to control bias and variance. The first condition for tuning the parameters mentioned in (C5) is similar to Li and Hsing (2010). The required spacing assumptions for eigen-values in Conditions (C6)a and (C6)b are similar as in Hall and Hosseini-Nasab (2009). Condition (C6)c is the trivial assumption that frequently arises in the functional data analysis literature. In most situations, this condition automatically holds. Using the weak law of large numbers, Condition (C7) holds for large n. Similar kind of conditions can be invoked, such as the convexity assumption, i.e.,*

$\lambda_r - \lambda_{r+1} \leq \lambda_{r-1} - \lambda_r$ *for all* $r \geq 2$. *Condition (C8) is determined to control the rate of the number of repeated measurements.*

Now, the following theorem provides the asymptotic expansion and consistency of the proposed estimator for $\widehat{\boldsymbol{\beta}}$.

Theorem 2.3.1. *Let* $\boldsymbol{\beta}_0$ *be the true value of* $\boldsymbol{\beta}$. *Under the Conditions (C1), (C2), (C3), (C4), (C5), (C6)a, (C6)b and (C6)c, for* $k = 1, \cdots, \kappa_0$, *we have the asymptotic mean square error for* $\overline{\mathbf{g}}_k(\boldsymbol{\beta}_0)$ *as*

$$AMSE\{\overline{\mathbf{g}}_k(\boldsymbol{\beta}_0)\} = O\left(n^{-1} + n^{-1}\kappa_0^{3-\tau} R_n(h)\right) \qquad almost\ surely \tag{2.13}$$

where $R_n(h) = \left\{h^4 + \frac{1}{n} + \frac{1}{nmh} + \frac{1}{n^2m^2h^2} + \frac{1}{n^2m^4h^4} + \frac{1}{n^2mh} + \frac{1}{n^2m^3h^3}\right\}$.
Moreover, under Condition (C8), $AMSE\{\widehat{\mathbf{g}}_k(\boldsymbol{\beta}_0)\} = O(n^{-1})$. *Therefore, if in addition, Condition (C7) holds, as* $n \to \infty$, $\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\| = O(n^{-1/2})$ *in probability.*

The following theorem states the results of asymptotic normality.

Theorem 2.3.2. *Define* $\mathscr{C}_i = \sum_{k_1=1}^{\kappa_0}\sum_{k_2=1}^{\kappa_0} \boldsymbol{\Phi}_{k_1}\mathbf{X}_i\mathbf{C}_{k_1,k_2}^{-1}\mathbf{X}_i^T\boldsymbol{\Phi}_{k_2}$, *where* $\mathbf{C}_{k_1,k_2}^{-1}$ *is a* (k_1, k_2) *block of* $\mathbf{C}_0^{-1}$ *with* $\mathbf{C}_0 = \mathbb{E}\left\{\mathbf{g}(\boldsymbol{\beta}_0)\mathbf{g}_i^T(\boldsymbol{\beta}_0)\right\}$. *Assume that the conditions for Theorem 2.3.1 hold. Then* $\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \xrightarrow{d} N(0, \boldsymbol{\Sigma})$ *where* $\boldsymbol{\Sigma} = \mathbf{B}^{-1}\mathbf{A}\mathbf{B}^{-1}$. *The quantities* $\mathbf{A}$ *and* $\mathbf{B}$ *are, respectively, limits of* $\widehat{\mathbf{A}} = \frac{1}{n}\sum_{i=1}^n \mathbf{X}_i^T\widehat{\mathscr{C}}_i\widehat{\mathbf{e}}_i\widehat{\mathbf{e}}_i^T\widehat{\mathscr{C}}_i\mathbf{X}_i$ *and* $\widehat{\mathbf{B}} = \frac{1}{n}\sum_{i=1}^n \mathbf{X}_i^T\widehat{\mathscr{C}}_i\mathbf{X}_i$, *and "* $\xrightarrow{d}$ *" denotes the convergence in distribution.*

Remark 2.3.2. *Here, the selection of the bandwidth only effects the second-order term of the MSE of* $\widehat{\boldsymbol{\beta}}$ *and has no effect on the asymptotic result of normality as long as h satisfies the Conditions (C5) and (C8) along with some restrictions on* κ_0.

All proofs with relevant technical details are available in Section 2.7.

2.4 Simulation studies

We conduct the numerical studies to compare the finite sample performance to the corresponding longitudinal approach of quadratic inference proposed in Qu et al. (2000) under different correlation structures.

2.4.1 Simulation set-up

Consider the normal response model

$$y_i(T_{ij}) = \mathbf{x}_i(T_{ij})^{\mathrm{T}}\boldsymbol{\beta} + e_i(T_{ij}) \tag{2.14}$$

For $p = 2$, we set coefficient vectors, $\boldsymbol{\beta} = (\beta_1, \beta_2)^{\mathrm{T}}$ where $\beta_1 = 1$ and $\beta_2 = 0.5$. The covariates are generated in following way.

$$x_{ik}(t) = \chi_{i1}^{(k)} + \chi_{i2}^{(k)}\sqrt{2}\sin(\pi t) + \chi_{i3}^{(k)}\sqrt{2}\cos(\pi t) \tag{2.15}$$

The coefficients $\chi_{i1}^{(k)} \sim N(0, (2^{-0.5(k-1)})^2)$, $\chi_{i2}^{(k)} \sim N(0, (0.85 \times 2^{-0.5(k-1)})^2)$, $\chi_{i3}^{(k)} \sim N(0, (0.7 \times 2^{-0.5(k-1)})^2)$ and χ_{ij}s are mutually independent for each trajectories i and each j. Consider the following simulation design.

- **Observational times-points**. In a fixed-design situation, associated observational times are fixed. Sample trajectories are observed at $m = 100$ equidistant time-points $\{t_1, \cdots, t_m\}$ on $[0, 1]$.

- **Choice of residuals**. The residual process $e_i(t)$ is a smoothed function with mean zero and unknown covariance function, where each e_i is distributed as $e_i = \sum_{k\geq 1}\xi_i\phi_i$ and ξ_ks are independent normal random variables with mean zero and respective variances λ_k. For numerical computation, we truncate the finite series at $k = 3$ in Karhunen-Loève expansions for Situations (a), (b) and (c) as described below. In Situations (d) and (e), the error process is generated from given covariance functions.

(a) Brownian motion. The covariance function for the Brownian motion is $\min(s,t)$, $\lambda_k = \frac{4}{\pi^2(2k-1)^2}$ and $\phi_k(t) = \sqrt{2}\sin(t/\sqrt{\lambda_k})$.

(b) Linear process. Consider the eigen-values be $\lambda_k = k^{-2l_0}$ and $\phi_k(t) = \sqrt{2}\cos(k\pi t)$. We fix $l_0 \in \{1,2,3\}$.

(c) Ornstein Uhlenbeck (OU) process. For positive constants μ_0 and ρ_0, we have a stochastic differential equation for $e(t)$ as $\partial e(t) = -\mu_0 e(t)\partial t + \rho_0 \partial w(t)$ for the Brownian motion $w(t)$. It can be shown that $\text{cov}\{e(t), e(s)\} = c\exp\{-\mu_0|t-s|\}$ where $c = \rho_0^2/2\mu_0$. Here we assume $c = 1$. Thus, by solving the integral equation we have $\phi_k(t) = A_k\cos(\omega_k t) + B_k\sin(\omega_k t)$ and $\lambda_k = \frac{2\mu_0}{\omega_k^2+\mu_0^2}$ where ω is solution of $\cot(\omega) = \frac{\omega^2-\mu_0^2}{2\mu_0\omega}$. The constants A_k and B_k are defined as $B_k = \mu_0 A_k/\omega_k$ where $A_k = \sqrt{\frac{2\omega_k^2}{2\mu_0+\mu_0^2+\omega_k^2}}$. Here μ_0 is chosen to be 1 or 3.

(d) Power exponential. $R(s,t) = \exp\{(-|s-t|/a_0)^{b_0}\}$ where scale parameter $a_0 = 1$ and shape parameter $b_0 \in \{1,2,5\}$.

(e) Rational quadratic. $R(s,t) = \left\{1 + \left(\frac{s-t}{a_0}\right)^2\right\}^{-b_0}$ where scale parameter $a_0 = 1$ and shape parameter $b_0 \in \{1,2,5\}$.

- **Sample size parameter.** Number of individuals, $n \in \{100, 300, 500\}$.

2.4.2 Comparison and evaluation

For each of the situations, we perform 500 simulation replicates. To execute Qu et al. (2000)'s approach, we construct the scores using basis matrices as described in Example 1 (approximation of the compound symmetric correlation structure, denoted as *ldaCS* in the tables) and Example 2 (for the first-order autoregressive correlation structure, denoted as *ldaAR* in the tables) in their paper. Ordinary least squares estimate (denoted as *init* in the tables) is taken as the initial estimate of $\boldsymbol{\beta}$ for both ours (denoted as *fda-k* for specific k in the tables) and Qu et al. (2000)'s approach. The estimation procedure in the iterative algorithm converges when the square difference between the estimated values of two consecutive steps is bounded by a small number 10^{-10} or the maximum

number of steps crosses 500, whichever happens earlier. To make theoretical results and numerical examples consistent, we use "*FPCA*" function in R which is available in *fdapace* packages (Gajardo et al., 2021) or in the MATLAB (MATLAB, 2014) package PACE available at http://www.stat.ucdavis.edu/PACE/ to estimate the eigen-functions. The key references for the PACE approach and associated works include in Yao et al. (2003, 2005); Müller and Yao (2010); Li and Hsing (2010). Bandwidths are selected using generalized cross-validation and the Epanechnikov kernel $K(x) = 0.75(1 - x^2)_+$ is used for estimation where $(a)_+ = \max(a, 0)$.

The means and standard deviations (SD) of the regression coefficients based on 500 simulations are given as summary measures. We calculate the standard deviation mentioned in the tables based on 500 estimates from 500 replications that can be viewed as the true standard error. Moreover, we also compute the following statistics to compare the performance of estimation, where for b-th replication $\widehat{\boldsymbol{\beta}}_b$ be the estimated value for $\boldsymbol{\beta}$,

- Absolute bias, AB = $\frac{1}{500}\sum_{b=1}^{500} |\widehat{\boldsymbol{\beta}}_b - \boldsymbol{\beta}|$
- Mean square error, MSE = $\frac{1}{500}\sum_{b=1}^{500} (\widehat{\boldsymbol{\beta}}_b - \boldsymbol{\beta})^2$

MSEs are reported in the order of 10^{-2}. The number of selected eigen-functions plays a critical role in our proposed method. We choose κ_0 based on a scree plot where the elbow of the graph is found and the components to the left are considered as significant.

Table 2.1 Performance of the estimation procedure where the residuals are generated from Brownian motion (a). Mean of the estimated coefficients, standard deviation, absolute bias, mean square error ($\times 100$) and FVE in percentage are summarized upto four decimal places.

Method	β_1				β_2				FVE
	Mean	SD	AB	MSE	Mean	SD	AB	MSE	%-age
					$n = 100$				
init	0.9999	0.0373	0.0297	0.1391	0.4995	0.0486	0.0384	0.2354	
ldaAR	0.9991	0.0331	0.0265	0.1095	0.5004	0.0445	0.0353	0.1972	
ldaCS	0.9987	0.0316	0.0253	0.1000	0.4997	0.0411	0.0322	0.1685	
fda-1	0.9998	0.0564	0.0447	0.3180	0.5006	0.0743	0.0587	0.5516	86.2672
fda-2	1.0001	0.0269	0.0213	0.0723	0.4971	0.0362	0.0290	0.1314	96.3746
fda-3	0.9998	0.0231	0.0181	0.0532	0.4978	0.0317	0.0251	0.1010	99.9220
fda-4	1.0004	0.0052	0.0014	0.0028	0.4994	0.0092	0.0022	0.0085	99.9979
fda-5	0.9999	0.0021	0.0008	0.0004	0.4999	0.0051	0.0012	0.0026	100.0000
fda-6	0.9999	0.0021	0.0008	0.0004	0.4999	0.0051	0.0012	0.0026	100.0000
fda-7	0.9999	0.0021	0.0008	0.0004	0.4999	0.0051	0.0012	0.0026	100.0000
					$n = 300$				
init	1.0002	0.0200	0.0162	0.0401	0.5003	0.0288	0.0226	0.0825	
ldaAR	1.0003	0.0184	0.0147	0.0336	0.5000	0.0259	0.0203	0.0670	
ldaCS	1.0007	0.0170	0.0134	0.0288	0.4995	0.0242	0.0190	0.0583	
fda-1	1.0002	0.0309	0.0251	0.0955	0.5008	0.0443	0.0350	0.1962	86.7578
fda-2	1.0002	0.0142	0.0114	0.0202	0.4998	0.0213	0.0169	0.0451	96.4747
fda-3	1.0002	0.0122	0.0098	0.0150	0.4992	0.0179	0.0144	0.0321	99.9745
fda-4	1.0002	0.0021	0.0003	0.0004	0.4998	0.0032	0.0005	0.0010	99.9993
fda-5	1.0000	0.0002	0.0001	0.0000	0.5000	0.0003	0.0002	0.0000	100.0000
fda-6	1.0000	0.0002	0.0001	0.0000	0.5000	0.0003	0.0002	0.0000	100.0000
fda-7	1.0000	0.0002	0.0001	0.0000	0.5000	0.0003	0.0002	0.0000	100.0000
					$n = 500$				
init	1.0002	0.0148	0.0117	0.0219	0.5006	0.0223	0.0177	0.0497	
ldaAR	1.0005	0.0138	0.0111	0.0189	0.5000	0.0206	0.0162	0.0422	
ldaCS	1.0000	0.0128	0.0102	0.0164	0.4992	0.0184	0.0146	0.0340	
fda-1	1.0007	0.0234	0.0185	0.0545	0.5012	0.0348	0.0277	0.1213	86.7520
fda-2	0.9996	0.0105	0.0083	0.0110	0.5002	0.0157	0.0126	0.0247	96.5174
fda-3	0.9991	0.0091	0.0074	0.0084	0.4999	0.0133	0.0107	0.0177	99.9851
fda-4	1.0000	0.0002	0.0001	0.0000	0.5000	0.0003	0.0002	0.0000	99.9996
fda-5	1.0000	0.0002	0.0001	0.0000	0.5000	0.0003	0.0002	0.0000	100.0000
fda-6	1.0000	0.0002	0.0001	0.0000	0.5000	0.0003	0.0002	0.0000	100.0000
fda-7	1.0000	0.0002	0.0001	0.0000	0.5000	0.0003	0.0002	0.0000	100.0000

Table 2.2 Performance of the estimation procedure where the residuals are generated from linear process (b) with $l_0 = 1$. Mean of the estimated coefficients, standard deviation, absolute bias, mean square error ($\times 100$) and FVE in percentage are summarized upto four decimal places.

Method	β_1				β_2				FVE
	Mean	SD	AB	MSE	Mean	SD	AB	MSE	%-age
					$n = 100$				
init	1.0019	0.0340	0.0267	0.1155	0.5001	0.0456	0.0356	0.2078	
ldaAR	1.0007	0.0230	0.0187	0.0528	0.5010	0.0361	0.0285	0.1303	
ldaCS	1.0000	0.0006	0.0005	0.0000	0.4999	0.0008	0.0006	0.0001	
fda-1	1.0093	0.1436	0.1150	2.0675	0.5020	0.2010	0.1570	4.0311	73.0607
fda-2	1.0036	0.1055	0.0828	1.1123	0.5052	0.1337	0.1070	1.7869	91.6726
fda-3	1.0038	0.1024	0.0804	1.0473	0.5056	0.1303	0.1020	1.6969	99.7657
fda-4	1.0000	0.0092	0.0021	0.0084	0.5006	0.0096	0.0024	0.0093	99.9993
fda-5	1.0000	0.0011	0.0007	0.0001	0.5000	0.0017	0.0010	0.0003	100.0000
fda-6	1.0000	0.0011	0.0007	0.0001	0.5000	0.0017	0.0010	0.0003	100.0000
fda-7	1.0000	0.0011	0.0007	0.0001	0.5000	0.0017	0.0010	0.0003	100.0000
					$n = 300$				
init	0.9991	0.0181	0.0144	0.0326	0.5006	0.0268	0.0212	0.0715	
ldaAR	0.9995	0.0133	0.0105	0.0178	0.5000	0.0212	0.0169	0.0447	
ldaCS	1.0000	0.0003	0.0002	0.0000	0.5000	0.0005	0.0004	0.0000	
fda-1	0.9958	0.0767	0.0616	0.5888	0.5037	0.1163	0.0918	1.3523	73.4907
fda-2	0.9974	0.0567	0.0458	0.3220	0.5011	0.0804	0.0648	0.6460	91.7757
fda-3	0.9974	0.0564	0.0455	0.3182	0.5007	0.0800	0.0643	0.6391	99.9225
fda-4	1.0000	0.0005	0.0001	0.0000	0.5000	0.0009	0.0002	0.0001	99.9998
fda-5	1.0000	0.0002	0.0001	0.0000	0.5000	0.0003	0.0002	0.0000	100.0000
fda-6	1.0000	0.0002	0.0001	0.0000	0.5000	0.0003	0.0002	0.0000	100.0000
fda-7	1.0000	0.0002	0.0001	0.0000	0.5000	0.0003	0.0002	0.0000	100.0000
					$n = 500$				
init	0.9999	0.0152	0.0121	0.0230	0.5027	0.0207	0.0163	0.0436	
ldaAR	1.0008	0.0100	0.0079	0.0100	0.5019	0.0161	0.0129	0.0263	
ldaCS	1.0000	0.0003	0.0002	0.0000	0.5000	0.0004	0.0003	0.0000	
fda-1	0.9990	0.0657	0.0523	0.4303	0.5113	0.0883	0.0698	0.7913	73.4098
fda-2	1.0013	0.0468	0.0371	0.2185	0.5078	0.0651	0.0520	0.4292	91.8501
fda-3	1.0014	0.0459	0.0364	0.2107	0.5074	0.0650	0.0516	0.4274	99.9490
fda-4	1.0000	0.0001	0.0001	0.0000	0.5000	0.0001	0.0001	0.0000	99.9999
fda-5	1.0000	0.0001	0.0001	0.0000	0.5000	0.0001	0.0001	0.0000	100.0000
fda-6	1.0000	0.0001	0.0001	0.0000	0.5000	0.0001	0.0001	0.0000	100.0000
fda-7	1.0000	0.0001	0.0001	0.0000	0.5000	0.0001	0.0001	0.0000	100.0000

Table 2.3 Performance of the estimation procedure where the residuals are generated from linear process (b) with $l_0 = 2$. Mean of the estimated coefficients, standard deviation, absolute bias, mean square error ($\times 100$) and FVE in percentage are summarized upto four decimal places.

Method	β_1				β_2				FVE
	Mean	SD	AB	MSE	Mean	SD	AB	MSE	%-age
					$n = 100$				
init	1.0020	0.0331	0.0261	0.1094	0.4999	0.0451	0.0349	0.2028	
ldaAR	1.0002	0.0167	0.0133	0.0278	0.5014	0.0232	0.0183	0.0538	
ldaCS	1.0000	0.0003	0.0002	0.0000	0.5000	0.0004	0.0003	0.0000	
fda-1	1.0096	0.1431	0.1144	2.0520	0.5022	0.2003	0.1561	4.0029	92.5933
fda-2	1.0014	0.0648	0.0508	0.4188	0.5037	0.0842	0.0667	0.7094	98.5532
fda-3	1.0009	0.0535	0.0406	0.2860	0.5018	0.0744	0.0570	0.5524	99.7251
fda-4	1.0000	0.0074	0.0017	0.0055	0.4999	0.0103	0.0024	0.0106	99.9991
fda-5	1.0000	0.0045	0.0009	0.0021	0.5000	0.0021	0.0009	0.0004	100.0000
fda-6	1.0000	0.0045	0.0009	0.0021	0.5000	0.0021	0.0009	0.0004	100.0000
fda-7	1.0000	0.0045	0.0009	0.0021	0.5000	0.0021	0.0009	0.0004	100.0000
					$n = 300$				
init	0.9991	0.0175	0.0140	0.0308	0.5006	0.0263	0.0208	0.0689	
ldaAR	0.9999	0.0091	0.0071	0.0083	0.4997	0.0127	0.0102	0.0161	
ldaCS	1.0000	0.0002	0.0001	0.0000	0.5000	0.0002	0.0002	0.0000	
fda-1	0.9957	0.0767	0.0616	0.5893	0.5038	0.1164	0.0919	1.3541	92.9525
fda-2	0.9989	0.0365	0.0295	0.1331	0.4998	0.0511	0.0410	0.2608	98.7539
fda-3	0.9992	0.0349	0.0282	0.1218	0.4991	0.0483	0.0385	0.2332	99.9061
fda-4	0.9999	0.0017	0.0002	0.0003	0.4999	0.0033	0.0004	0.0011	99.9997
fda-5	1.0000	0.0002	0.0001	0.0000	0.5000	0.0002	0.0001	0.0000	100.0000
fda-6	1.0000	0.0002	0.0001	0.0000	0.5000	0.0002	0.0001	0.0000	100.0000
fda-7	1.0000	0.0002	0.0001	0.0000	0.5000	0.0002	0.0001	0.0000	100.0000
					$n = 500$				
init	0.9998	0.0148	0.0118	0.0219	0.5026	0.0201	0.0158	0.0409	
ldaAR	1.0006	0.0073	0.0059	0.0054	0.5008	0.0098	0.0079	0.0097	
ldaCS	1.0000	0.0001	0.0001	0.0000	0.5000	0.0002	0.0001	0.0000	
fda-1	0.9990	0.0656	0.0523	0.4295	0.5114	0.0882	0.0698	0.7897	92.9459
fda-2	1.0013	0.0296	0.0233	0.0874	0.5039	0.0415	0.0329	0.1735	98.7957
fda-3	1.0012	0.0285	0.0224	0.0812	0.5036	0.0407	0.0322	0.1670	99.9389
fda-4	1.0000	0.0001	0.0001	0.0000	0.5000	0.0001	0.0001	0.0000	99.9998
fda-5	1.0000	0.0001	0.0001	0.0000	0.5000	0.0001	0.0001	0.0000	100.0000
fda-6	1.0000	0.0001	0.0001	0.0000	0.5000	0.0001	0.0001	0.0000	100.0000
fda-7	1.0000	0.0001	0.0001	0.0000	0.5000	0.0001	0.0001	0.0000	100.0000

Table 2.4 Performance of the estimation procedure where the residuals are generated from linear process (b) with $l_0 = 3$. Mean of the estimated coefficients, standard deviation, absolute bias, mean square error ($\times 100$) and FVE in percentage are summarized upto four decimal places.

Method	β_1				β_2				FVE
	Mean	SD	AB	MSE	Mean	SD	AB	MSE	%-age
					$n = 100$				
init	1.0020	0.0328	0.0260	0.1077	0.4998	0.0451	0.0349	0.2027	
ldaAR	1.0000	0.0087	0.0069	0.0076	0.5009	0.0122	0.0096	0.0149	
ldaCS	1.0000	0.0001	0.0001	0.0000	0.5000	0.0002	0.0002	0.0000	
fda-1	1.0096	0.1430	0.1143	2.0496	0.5023	0.2001	0.1559	3.9978	97.9586
fda-2	1.0000	0.0326	0.0250	0.1058	0.5023	0.0440	0.0344	0.1941	99.5699
fda-3	0.9998	0.0144	0.0072	0.0208	0.4986	0.0209	0.0103	0.0436	99.8941
fda-4	1.0002	0.0053	0.0013	0.0028	0.5002	0.0067	0.0018	0.0044	99.9991
fda-5	1.0003	0.0037	0.0008	0.0014	0.4998	0.0042	0.0011	0.0018	100.0000
fda-6	1.0003	0.0037	0.0008	0.0014	0.4998	0.0042	0.0011	0.0018	100.0000
fda-7	1.0003	0.0037	0.0008	0.0014	0.4998	0.0042	0.0011	0.0018	100.0000
					$n = 300$				
init	0.9991	0.0174	0.0139	0.0303	0.5006	0.0262	0.0207	0.0684	
ldaAR	1.0000	0.0047	0.0037	0.0022	0.4997	0.0066	0.0052	0.0044	
ldaCS	1.0000	0.0001	0.0001	0.0000	0.5000	0.0001	0.0001	0.0000	
fda-1	0.9957	0.0767	0.0616	0.5896	0.5038	0.1164	0.0919	1.3543	98.2262
fda-2	0.9996	0.0197	0.0158	0.0386	0.4996	0.0274	0.0219	0.0750	99.7663
fda-3	1.0002	0.0151	0.0101	0.0227	0.4990	0.0186	0.0125	0.0347	99.9255
fda-4	1.0001	0.0022	0.0003	0.0005	0.5001	0.0017	0.0003	0.0003	99.9997
fda-5	1.0000	0.0002	0.0001	0.0000	0.5000	0.0002	0.0001	0.0000	100.0000
fda-6	1.0000	0.0002	0.0001	0.0000	0.5000	0.0002	0.0001	0.0000	100.0000
fda-7	1.0000	0.0002	0.0001	0.0000	0.5000	0.0002	0.0001	0.0000	100.0000
					$n = 500$				
init	0.9998	0.0147	0.0117	0.0216	0.5025	0.0199	0.0157	0.0400	
ldaAR	1.0003	0.0038	0.0031	0.0015	0.5003	0.0052	0.0041	0.0027	
ldaCS	1.0000	0.0001	0.0001	0.0000	0.5000	0.0001	0.0001	0.0000	
fda-1	0.9990	0.0656	0.0523	0.4293	0.5114	0.0882	0.0698	0.7895	98.2518
fda-2	1.0008	0.0159	0.0126	0.0253	0.5015	0.0222	0.0175	0.0495	99.8021
fda-3	1.0001	0.0126	0.0091	0.0158	0.5002	0.0179	0.0130	0.0320	99.9449
fda-4	1.0000	0.0001	0.0001	0.0000	0.5000	0.0001	0.0001	0.0000	99.9998
fda-5	1.0000	0.0001	0.0001	0.0000	0.5000	0.0001	0.0001	0.0000	100.0000
fda-6	1.0000	0.0001	0.0001	0.0000	0.5000	0.0001	0.0001	0.0000	100.0000
fda-7	1.0000	0.0001	0.0001	0.0000	0.5000	0.0001	0.0001	0.0000	100.0000

Table 2.5 Performance of the estimation procedure where the residuals are generated from Ornstein-Uhlenbeck process (c) with $\mu_0 = 1$. Mean of the estimated coefficients, standard deviation, absolute bias, mean square error ($\times 100$) and FVE in percentage are summarized upto four decimal places.

Method	β_1				β_2				FVE
	Mean	SD	AB	MSE	Mean	SD	AB	MSE	%-age
					$n = 100$				
init	1.0003	0.0541	0.0434	0.2922	0.4994	0.0711	0.0563	0.5048	
ldaAR	1.0001	0.0476	0.0383	0.2261	0.4984	0.0650	0.0513	0.4214	
ldaCS	0.9994	0.0398	0.0316	0.1581	0.4978	0.0534	0.0421	0.2849	
fda-1	1.0009	0.0705	0.0563	0.4964	0.5006	0.0947	0.0747	0.8954	79.4156
fda-2	1.0001	0.0453	0.0358	0.2044	0.4982	0.0608	0.0482	0.3690	94.9669
fda-3	0.9993	0.0386	0.0307	0.1491	0.4978	0.0511	0.0405	0.2613	99.9949
fda-4	1.0003	0.0127	0.0081	0.0162	0.4991	0.0242	0.0146	0.0583	99.9992
fda-5	0.9997	0.0084	0.0071	0.0071	0.4991	0.0159	0.0124	0.0254	100.0000
fda-6	0.9997	0.0084	0.0071	0.0071	0.4991	0.0159	0.0124	0.0254	100.0000
fda-7	0.9997	0.0084	0.0071	0.0071	0.4991	0.0159	0.0124	0.0254	100.0000
					$n = 300$				
init	1.0000	0.0288	0.0233	0.0829	0.5003	0.0416	0.0329	0.1728	
ldaAR	1.0000	0.0258	0.0206	0.0662	0.4996	0.0368	0.0294	0.1350	
ldaCS	1.0002	0.0212	0.0170	0.0449	0.4987	0.0314	0.0254	0.0983	
fda-1	0.9997	0.0388	0.0316	0.1503	0.5014	0.0560	0.0440	0.3130	79.9233
fda-2	1.0005	0.0240	0.0193	0.0576	0.4983	0.0352	0.0284	0.1242	95.0283
fda-3	0.9999	0.0202	0.0161	0.0409	0.4994	0.0298	0.0240	0.0884	99.9984
fda-4	0.9999	0.0072	0.0064	0.0051	0.4997	0.0129	0.0111	0.0166	99.9998
fda-5	1.0000	0.0072	0.0064	0.0051	0.4997	0.0128	0.0111	0.0165	100.0000
fda-6	1.0000	0.0072	0.0064	0.0051	0.4997	0.0128	0.0111	0.0165	100.0000
fda-7	1.0000	0.0072	0.0064	0.0051	0.4997	0.0128	0.0111	0.0165	100.0000
					$n = 500$				
init	1.0000	0.0288	0.0233	0.0829	0.5003	0.0416	0.0329	0.1728	
ldaAR	1.0000	0.0258	0.0206	0.0662	0.4996	0.0368	0.0294	0.1350	
ldaCS	1.0002	0.0212	0.0170	0.0449	0.4987	0.0314	0.0254	0.0983	
fda-1	0.9997	0.0388	0.0316	0.1503	0.5014	0.0560	0.0440	0.3130	79.9233
fda-2	1.0005	0.0240	0.0193	0.0576	0.4983	0.0352	0.0284	0.1242	95.0283
fda-3	0.9999	0.0202	0.0161	0.0409	0.4994	0.0298	0.0240	0.0884	99.9984
fda-4	0.9999	0.0072	0.0064	0.0051	0.4997	0.0129	0.0111	0.0166	99.9998
fda-5	1.0000	0.0072	0.0064	0.0051	0.4997	0.0128	0.0111	0.0165	100.0000
fda-6	1.0000	0.0072	0.0064	0.0051	0.4997	0.0128	0.0111	0.0165	100.0000
fda-7	1.0000	0.0072	0.0064	0.0051	0.4997	0.0128	0.0111	0.0165	100.0000

Table 2.6 Performance of the estimation procedure where the residuals are generated from Ornstein-Uhlenbeck process (c) with $\mu_0 = 3$. Mean of the estimated coefficients, standard deviation, absolute bias, mean square error ($\times 100$) and FVE in percentage are summarized upto four decimal places.

Method	β_1				β_2				FVE
	Mean	SD	AB	MSE	Mean	SD	AB	MSE	%-age
					$n = 100$				
init	1.0001	0.0454	0.0363	0.2056	0.4990	0.0591	0.0469	0.3487	
ldaAR	0.9996	0.0390	0.0312	0.1521	0.4979	0.0521	0.0414	0.2717	
ldaCS	0.9999	0.0429	0.0341	0.1841	0.4981	0.0564	0.0448	0.3176	
fda-1	1.0005	0.0557	0.0445	0.3100	0.5003	0.0743	0.0588	0.5511	59.1692
fda-2	1.0005	0.0459	0.0362	0.2107	0.4981	0.0604	0.0480	0.3639	87.0120
fda-3	1.0000	0.0436	0.0348	0.1894	0.4975	0.0568	0.0455	0.3221	99.9908
fda-4	1.0004	0.0136	0.0058	0.0185	0.4989	0.0237	0.0106	0.0562	99.9960
fda-5	1.0001	0.0049	0.0033	0.0024	0.4997	0.0099	0.0061	0.0098	100.0000
fda-6	1.0001	0.0049	0.0033	0.0024	0.4997	0.0099	0.0061	0.0098	100.0000
fda-7	1.0001	0.0049	0.0033	0.0024	0.4997	0.0099	0.0061	0.0098	100.0000
					$n = 300$				
init	1.0002	0.0239	0.0191	0.0568	0.4998	0.0345	0.0274	0.1190	
ldaAR	1.0003	0.0209	0.0167	0.0435	0.4990	0.0303	0.0244	0.0916	
ldaCS	1.0003	0.0224	0.0178	0.0500	0.4992	0.0328	0.0265	0.1075	
fda-1	0.9998	0.0305	0.0247	0.0928	0.5011	0.0443	0.0349	0.1957	59.4418
fda-2	1.0004	0.0240	0.0192	0.0574	0.4991	0.0347	0.0279	0.1201	86.9290
fda-3	1.0003	0.0222	0.0177	0.0492	0.4992	0.0327	0.0264	0.1070	99.9972
fda-4	1.0002	0.0043	0.0039	0.0018	0.4998	0.0075	0.0065	0.0057	99.9987
fda-5	1.0001	0.0038	0.0035	0.0014	0.4998	0.0068	0.0059	0.0046	100.0000
fda-6	1.0001	0.0038	0.0035	0.0014	0.4998	0.0068	0.0059	0.0046	100.0000
fda-7	1.0001	0.0038	0.0035	0.0014	0.4998	0.0068	0.0059	0.0046	100.0000
					$n = 500$				
init	1.0003	0.0180	0.0141	0.0323	0.5017	0.0268	0.0213	0.0720	
ldaAR	1.0001	0.0155	0.0123	0.0241	0.5016	0.0232	0.0186	0.0541	
ldaCS	1.0002	0.0171	0.0134	0.0291	0.5013	0.0251	0.0202	0.0629	
fda-1	1.0005	0.0229	0.0180	0.0523	0.5021	0.0346	0.0276	0.1201	59.3104
fda-2	1.0004	0.0177	0.0138	0.0314	0.5022	0.0268	0.0217	0.0724	87.0183
fda-3	1.0001	0.0173	0.0136	0.0300	0.5010	0.0249	0.0197	0.0620	99.9983
fda-4	1.0000	0.0042	0.0038	0.0017	0.5005	0.0075	0.0065	0.0056	99.9992
fda-5	1.0000	0.0038	0.0035	0.0015	0.5004	0.0066	0.0058	0.0044	100.0000
fda-6	1.0000	0.0038	0.0035	0.0015	0.5004	0.0066	0.0058	0.0044	100.0000
fda-7	1.0000	0.0038	0.0035	0.0015	0.5004	0.0066	0.0058	0.0044	100.0000

Table 2.7 Performance of the estimation procedure where the residuals are generated with power exponential covariance function (d) where scale parameter $a_0 = 1$ and shape parameter $b_0 = 1$. Mean of the estimated coefficients, standard deviation, absolute bias, mean square error (×100) and FVE in percentage are summarized upto four decimal places.

Method	β_1				β_2				FVE
	Mean	SD	AB	MSE	Mean	SD	AB	MSE	%-age
					$n = 100$				
init	0.9968	0.0525	0.0423	0.2758	0.4961	0.0755	0.0603	0.5705	
ldaAR	0.9985	0.0486	0.0387	0.2361	0.4962	0.0702	0.0562	0.4938	
ldaCS	0.9978	0.0389	0.0309	0.1514	0.4986	0.0549	0.0438	0.3013	
fda-1	0.9960	0.0708	0.0564	0.5018	0.4951	0.1010	0.0813	1.0195	73.1399
fda-2	0.9975	0.0439	0.0347	0.1929	0.4980	0.0629	0.0496	0.3948	87.5389
fda-3	0.9975	0.0381	0.0305	0.1453	0.4987	0.0531	0.0425	0.2811	92.2253
fda-4	0.9978	0.0392	0.0311	0.1540	0.4977	0.0540	0.0434	0.2914	94.4211
fda-5	0.9978	0.0388	0.0302	0.1508	0.4975	0.0527	0.0420	0.2773	95.6881
fda-6	0.9979	0.0393	0.0306	0.1546	0.4977	0.0534	0.0424	0.2852	96.5184
fda-7	0.9990	0.0396	0.0308	0.1570	0.4986	0.0535	0.0423	0.2857	97.0882
fda-8	0.9985	0.0405	0.0317	0.1637	0.4981	0.0540	0.0429	0.2915	97.5117
					$n = 300$				
init	0.9975	0.0297	0.0236	0.0885	0.4989	0.0428	0.0352	0.1832	
ldaAR	0.9972	0.0281	0.0224	0.0793	0.4992	0.0389	0.0316	0.1509	
ldaCS	0.9988	0.0226	0.0180	0.0513	0.4995	0.0300	0.0238	0.0899	
fda-1	0.9967	0.0391	0.0310	0.1539	0.4986	0.0588	0.0482	0.3456	73.4971
fda-2	0.9986	0.0254	0.0203	0.0648	0.4990	0.0335	0.0266	0.1122	87.5190
fda-3	0.9987	0.0221	0.0173	0.0490	0.4996	0.0293	0.0233	0.0859	92.1425
fda-4	0.9986	0.0219	0.0171	0.0479	0.4997	0.0294	0.0233	0.0862	94.3152
fda-5	0.9988	0.0213	0.0167	0.0456	0.5000	0.0287	0.0232	0.0823	95.5616
fda-6	0.9989	0.0213	0.0166	0.0454	0.5001	0.0289	0.0232	0.0832	96.3760
fda-7	0.9988	0.0212	0.0166	0.0449	0.5003	0.0291	0.0234	0.0843	96.9441
fda-8	0.9988	0.0212	0.0166	0.0449	0.5001	0.0292	0.0236	0.0852	97.3641
					$n = 500$				
init	0.9996	0.0212	0.0169	0.0450	0.4989	0.0316	0.0252	0.0999	
ldaAR	0.9993	0.0189	0.0151	0.0356	0.4983	0.0294	0.0237	0.0868	
ldaCS	0.9996	0.0170	0.0135	0.0288	0.4998	0.0235	0.0193	0.0553	
fda-1	0.9996	0.0278	0.0222	0.0773	0.4984	0.0422	0.0333	0.1777	73.6172
fda-2	0.9992	0.0186	0.0150	0.0347	0.4995	0.0264	0.0216	0.0694	87.5609
fda-3	1.0002	0.0164	0.0130	0.0268	0.5001	0.0225	0.0183	0.0505	92.1497
fda-4	1.0001	0.0163	0.0129	0.0264	0.4999	0.0226	0.0184	0.0508	94.3084
fda-5	1.0000	0.0160	0.0127	0.0255	0.4996	0.0223	0.0182	0.0496	95.5492
fda-6	1.0000	0.0161	0.0128	0.0260	0.4995	0.0222	0.0182	0.0493	96.3576
fda-7	0.9999	0.0161	0.0128	0.0258	0.4994	0.0219	0.0179	0.0481	96.9249
fda-8	0.9999	0.0161	0.0128	0.0258	0.4994	0.0219	0.0177	0.0478	97.3423

Table 2.8 Performance of the estimation procedure where the residuals are generated with power exponential covariance function (d) where scale parameter $a_0 = 1$ and shape parameter $b_0 = 2$. Mean of the estimated coefficients, standard deviation, absolute bias, mean square error (×100) and FVE in percentage are summarized upto four decimal places.

Method	β_1				β_2				FVE
	Mean	SD	AB	MSE	Mean	SD	AB	MSE	%-age
					$n = 100$				
init	0.9967	0.0563	0.0454	0.3170	0.4957	0.0811	0.0649	0.6591	
ldaAR	0.9980	0.0506	0.0399	0.2562	0.4970	0.0724	0.0578	0.5234	
ldaCS	0.9982	0.0373	0.0294	0.1389	0.4983	0.0531	0.0425	0.2814	
fda-1	0.9957	0.0766	0.0611	0.5872	0.4947	0.1094	0.0881	1.1964	85.6976
fda-2	0.9979	0.0440	0.0348	0.1934	0.4975	0.0633	0.0503	0.4010	98.9964
fda-3	0.9993	0.0239	0.0187	0.0569	0.4988	0.0343	0.0273	0.1177	99.9585
fda-4	0.9984	0.0224	0.0176	0.0505	0.5004	0.0305	0.0242	0.0931	99.9987
fda-5	0.9984	0.0224	0.0176	0.0505	0.5004	0.0305	0.0242	0.0931	100.0000
fda-6	0.9984	0.0224	0.0176	0.0505	0.5004	0.0305	0.0242	0.0931	100.0000
fda-7	0.9984	0.0224	0.0176	0.0505	0.5004	0.0305	0.0242	0.0931	100.0000
fda-8	0.9984	0.0224	0.0176	0.0505	0.5004	0.0305	0.0242	0.0931	100.0000
					$n = 300$				
init	0.9973	0.0318	0.0253	0.1014	0.4987	0.0461	0.0378	0.2118	
ldaAR	0.9972	0.0297	0.0239	0.0888	0.4992	0.0400	0.0323	0.1597	
ldaCS	0.9990	0.0216	0.0172	0.0468	0.4995	0.0289	0.0229	0.0835	
fda-1	0.9964	0.0424	0.0336	0.1804	0.4984	0.0637	0.0521	0.4047	86.0991
fda-2	0.9987	0.0253	0.0201	0.0642	0.4990	0.0336	0.0266	0.1129	99.0443
fda-3	0.9992	0.0146	0.0114	0.0213	0.5003	0.0197	0.0159	0.0388	99.9593
fda-4	0.9993	0.0128	0.0100	0.0165	0.5000	0.0176	0.0139	0.0309	99.9987
fda-5	0.9993	0.0128	0.0100	0.0165	0.5000	0.0176	0.0139	0.0309	100.0000
fda-6	0.9993	0.0128	0.0100	0.0165	0.5000	0.0176	0.0139	0.0309	100.0000
fda-7	0.9993	0.0128	0.0100	0.0165	0.5000	0.0176	0.0139	0.0309	100.0000
					$n = 500$				
init	0.9994	0.0226	0.0180	0.0512	0.4987	0.0340	0.0270	0.1153	
ldaAR	0.9993	0.0201	0.0161	0.0403	0.4983	0.0309	0.0250	0.0954	
ldaCS	0.9993	0.0163	0.0130	0.0266	0.4995	0.0227	0.0186	0.0517	
fda-1	0.9995	0.0301	0.0240	0.0906	0.4983	0.0456	0.0361	0.2081	86.1915
fda-2	0.9990	0.0186	0.0150	0.0346	0.4992	0.0265	0.0217	0.0699	99.0585
fda-3	1.0004	0.0111	0.0087	0.0122	0.5002	0.0148	0.0121	0.0219	99.9596
fda-4	1.0007	0.0100	0.0080	0.0100	0.5007	0.0129	0.0106	0.0168	99.9987
fda-5	1.0007	0.0100	0.0080	0.0100	0.5007	0.0129	0.0106	0.0168	100.0000
fda-6	1.0007	0.0100	0.0080	0.0100	0.5007	0.0129	0.0106	0.0168	100.0000
fda-7	1.0007	0.0100	0.0080	0.0100	0.5007	0.0129	0.0106	0.0168	100.0000
fda-8	1.0007	0.0100	0.0080	0.0100	0.5007	0.0129	0.0106	0.0168	100.0000

Table 2.9 Performance of the estimation procedure where the residuals are generated with power exponential covariance function (d) where scale parameter $a_0 = 1$ and shape parameter $b_0 = 5$. Mean of the estimated coefficients, standard deviation, absolute bias, mean square error (×100) and FVE in percentage are summarized upto four decimal places.

Method	β_1				β_2				FVE
	Mean	SD	AB	MSE	Mean	SD	AB	MSE	%-age
					$n = 100$				
init	0.9970	0.0582	0.0468	0.3390	0.4952	0.0841	0.0676	0.7089	
ldaAR	0.9977	0.0528	0.0415	0.2789	0.4961	0.0772	0.0621	0.5956	
ldaCS	0.9996	0.0274	0.0218	0.0747	0.4979	0.0396	0.0321	0.1572	
fda-1	0.9956	0.0810	0.0646	0.6564	0.4943	0.1157	0.0933	1.3399	92.2949
fda-2	0.9995	0.0375	0.0296	0.1400	0.4966	0.0548	0.0441	0.3013	99.7252
fda-3	0.9993	0.0283	0.0203	0.0801	0.5004	0.0383	0.0274	0.1462	99.8787
fda-4	0.9991	0.0121	0.0053	0.0147	0.4995	0.0165	0.0073	0.0273	99.9461
fda-5	1.0001	0.0116	0.0022	0.0134	0.5005	0.0146	0.0030	0.0212	99.9842
fda-6	1.0000	0.0133	0.0029	0.0177	0.5004	0.0184	0.0041	0.0336	99.9979
fda-7	1.0000	0.0133	0.0028	0.0176	0.5004	0.0183	0.0040	0.0334	99.9990
fda-8	1.0000	0.0133	0.0028	0.0176	0.5004	0.0183	0.0040	0.0334	99.9998
					$n = 300$				
init	0.9972	0.0328	0.0260	0.1082	0.4986	0.0482	0.0395	0.2317	
ldaAR	0.9971	0.0310	0.0251	0.0968	0.4994	0.0426	0.0344	0.1810	
ldaCS	0.9995	0.0156	0.0123	0.0244	0.4996	0.0216	0.0171	0.0467	
fda-1	0.9962	0.0449	0.0356	0.2027	0.4982	0.0673	0.0552	0.4524	92.6741
fda-2	0.9990	0.0213	0.0168	0.0454	0.4994	0.0291	0.0229	0.0846	99.8076
fda-3	0.9995	0.0196	0.0153	0.0384	0.4992	0.0271	0.0213	0.0736	99.9391
fda-4	1.0005	0.0075	0.0044	0.0056	0.4994	0.0105	0.0064	0.0111	99.9716
fda-5	0.9999	0.0023	0.0006	0.0005	0.5000	0.0037	0.0010	0.0014	99.9889
fda-6	1.0000	0.0025	0.0006	0.0006	0.4999	0.0049	0.0011	0.0024	99.9979
fda-7	1.0000	0.0012	0.0003	0.0001	0.5002	0.0039	0.0006	0.0015	99.9989
fda-8	1.0000	0.0012	0.0003	0.0001	0.5002	0.0039	0.0006	0.0015	99.9997
					$n = 500$				
init	0.9994	0.0232	0.0185	0.0539	0.4985	0.0352	0.0280	0.1242	
ldaAR	0.9989	0.0209	0.0169	0.0437	0.4981	0.0327	0.0264	0.1073	
ldaCS	0.9991	0.0119	0.0095	0.0142	0.4992	0.0168	0.0134	0.0281	
fda-1	0.9995	0.0318	0.0254	0.1011	0.4981	0.0483	0.0382	0.2328	92.7586
fda-2	0.9988	0.0158	0.0127	0.0252	0.4988	0.0227	0.0183	0.0517	99.8272
fda-3	0.9989	0.0153	0.0123	0.0235	0.4989	0.0218	0.0176	0.0478	99.9574
fda-4	0.9997	0.0070	0.0050	0.0049	0.5002	0.0105	0.0072	0.0110	99.9811
fda-5	1.0000	0.0024	0.0007	0.0006	0.5001	0.0035	0.0011	0.0012	99.9923
fda-6	1.0000	0.0026	0.0006	0.0007	0.5003	0.0045	0.0011	0.0021	99.9979
fda-7	1.0001	0.0017	0.0004	0.0003	0.5000	0.0038	0.0007	0.0014	99.9989
fda-8	1.0001	0.0015	0.0003	0.0002	0.5001	0.0035	0.0007	0.0012	99.9997

Table 2.10 Performance of the estimation procedure where the residuals are generated with rational quadratic covariance function (e) where scale parameter $a_0 = 1$ and shape parameter $b_0 = 1$. Mean of the estimated coefficients, standard deviation, absolute bias, mean square error (×100) and FVE in percentage are summarized upto four decimal places.

Method	β_1				β_2				FVE
	Mean	SD	AB	MSE	Mean	SD	AB	MSE	%-age
					$n = 100$				
init	0.9967	0.0565	0.0455	0.3193	0.4957	0.0814	0.0651	0.6624	
ldaAR	0.9982	0.0507	0.0399	0.2567	0.4968	0.0725	0.0580	0.5257	
ldaCS	0.9983	0.0349	0.0275	0.1215	0.4986	0.0495	0.0396	0.2444	
fda-1	0.9957	0.0774	0.0617	0.5994	0.4946	0.1105	0.0890	1.2213	87.2296
fda-2	0.9980	0.0413	0.0326	0.1706	0.4979	0.0594	0.0471	0.3521	98.4366
fda-3	0.9989	0.0265	0.0210	0.0704	0.4988	0.0379	0.0302	0.1434	99.8285
fda-4	0.9985	0.0271	0.0212	0.0733	0.4991	0.0372	0.0296	0.1383	99.9800
fda-5	0.9987	0.0259	0.0203	0.0671	0.4987	0.0350	0.0276	0.1224	99.9977
fda-6	0.9987	0.0259	0.0203	0.0671	0.4987	0.0350	0.0276	0.1224	99.9997
fda-7	0.9987	0.0259	0.0203	0.0671	0.4987	0.0350	0.0276	0.1224	100.0000
fda-8	0.9987	0.0259	0.0203	0.0671	0.4987	0.0350	0.0276	0.1224	100.0000
					$n = 300$				
init	0.9973	0.0319	0.0253	0.1021	0.4987	0.0463	0.0381	0.2144	
ldaAR	0.9971	0.0297	0.0239	0.0889	0.4992	0.0404	0.0327	0.1630	
ldaCS	0.9991	0.0203	0.0161	0.0412	0.4995	0.0271	0.0215	0.0736	
fda-1	0.9964	0.0429	0.0339	0.1846	0.4984	0.0643	0.0527	0.4132	87.6273
fda-2	0.9988	0.0238	0.0190	0.0568	0.4991	0.0317	0.0251	0.1002	98.4851
fda-3	0.9991	0.0159	0.0125	0.0255	0.5002	0.0215	0.0173	0.0462	99.8287
fda-4	0.9991	0.0157	0.0121	0.0248	0.5001	0.0214	0.0171	0.0457	99.9800
fda-5	0.9993	0.0144	0.0113	0.0209	0.5005	0.0202	0.0164	0.0408	99.9977
fda-6	0.9993	0.0144	0.0113	0.0209	0.5005	0.0202	0.0164	0.0408	99.9997
fda-7	0.9993	0.0144	0.0113	0.0209	0.5005	0.0202	0.0164	0.0408	100.0000
fda-8	0.9993	0.0144	0.0113	0.0209	0.5005	0.0202	0.0164	0.0408	100.0000
					$n = 500$				
init	0.9995	0.0227	0.0181	0.0514	0.4987	0.0341	0.0271	0.1160	
ldaAR	0.9993	0.0200	0.0160	0.0399	0.4982	0.0310	0.0250	0.0962	
ldaCS	0.9994	0.0153	0.0122	0.0235	0.4996	0.0213	0.0174	0.0454	
fda-1	0.9995	0.0304	0.0243	0.0924	0.4982	0.0461	0.0365	0.2125	87.7225
fda-2	0.9991	0.0176	0.0142	0.0310	0.4993	0.0249	0.0204	0.0620	98.5022
fda-3	1.0003	0.0121	0.0096	0.0146	0.5002	0.0163	0.0133	0.0265	99.8293
fda-4	1.0005	0.0120	0.0096	0.0145	0.5004	0.0160	0.0131	0.0256	99.9801
fda-5	1.0002	0.0114	0.0090	0.0129	0.5000	0.0151	0.0121	0.0227	99.9977
fda-6	1.0002	0.0114	0.0090	0.0129	0.5000	0.0151	0.0121	0.0227	99.9997
fda-7	1.0002	0.0114	0.0090	0.0129	0.5000	0.0151	0.0121	0.0227	100.0000
fda-8	1.0002	0.0114	0.0090	0.0129	0.5000	0.0151	0.0121	0.0227	100.0000

Table 2.11 Performance of the estimation procedure where the residuals are generated with rational quadratic covariance function (e) where scale parameter $a_0 = 1$ and shape parameter $b_0 = 2$. Mean of the estimated coefficients, standard deviation, absolute bias, mean square error (×100) and FVE in percentage are summarized upto four decimal places.

Method	β_1				β_2				FVE
	Mean	SD	AB	MSE	Mean	SD	AB	MSE	%-age
					$n = 100$				
init	0.9967	0.0547	0.0441	0.2993	0.4960	0.0788	0.0628	0.6206	
ldaAR	0.9983	0.0498	0.0393	0.2473	0.4968	0.0714	0.0571	0.5091	
ldaCS	0.9977	0.0425	0.0337	0.1805	0.4981	0.0602	0.0481	0.3618	
fda-1	0.9957	0.0730	0.0583	0.5343	0.4951	0.1042	0.0839	1.0868	78.3365
fda-2	0.9972	0.0479	0.0381	0.2297	0.4975	0.0687	0.0544	0.4721	96.4030
fda-3	0.9981	0.0368	0.0293	0.1358	0.4982	0.0520	0.0418	0.2699	99.4758
fda-4	0.9982	0.0377	0.0299	0.1424	0.4979	0.0528	0.0421	0.2791	99.9264
fda-5	0.9986	0.0356	0.0280	0.1266	0.4975	0.0483	0.0384	0.2335	99.9902
fda-6	0.9986	0.0356	0.0280	0.1266	0.4975	0.0483	0.0384	0.2335	99.9987
fda-7	0.9986	0.0356	0.0280	0.1266	0.4975	0.0483	0.0384	0.2335	99.9998
fda-8	0.9986	0.0356	0.0280	0.1266	0.4975	0.0483	0.0384	0.2335	100.0000
					$n = 300$				
init	0.9974	0.0309	0.0246	0.0959	0.4988	0.0445	0.0365	0.1977	
ldaAR	0.9972	0.0290	0.0233	0.0848	0.4991	0.0393	0.0317	0.1542	
ldaCS	0.9987	0.0246	0.0195	0.0606	0.4994	0.0327	0.0259	0.1066	
fda-1	0.9966	0.0403	0.0320	0.1633	0.4986	0.0607	0.0497	0.3682	78.7132
fda-2	0.9985	0.0276	0.0220	0.0764	0.4989	0.0364	0.0290	0.1327	96.4278
fda-3	0.9985	0.0218	0.0171	0.0477	0.4999	0.0290	0.0232	0.0842	99.4738
fda-4	0.9985	0.0218	0.0171	0.0478	0.4998	0.0293	0.0235	0.0858	99.9259
fda-5	0.9988	0.0198	0.0157	0.0394	0.5004	0.0271	0.0224	0.0733	99.9900
fda-6	0.9988	0.0198	0.0157	0.0394	0.5004	0.0271	0.0224	0.0733	99.9987
fda-7	0.9988	0.0198	0.0157	0.0394	0.5004	0.0271	0.0224	0.0733	99.9998
fda-8	0.9988	0.0198	0.0157	0.0394	0.5004	0.0271	0.0224	0.0733	100.0000
					$n = 500$				
init	0.9995	0.0221	0.0176	0.0490	0.4989	0.0330	0.0263	0.1086	
ldaAR	0.9993	0.0197	0.0158	0.0389	0.4983	0.0302	0.0245	0.0910	
ldaCS	0.9994	0.0184	0.0146	0.0339	0.4996	0.0257	0.0211	0.0658	
fda-1	0.9996	0.0288	0.0229	0.0826	0.4984	0.0435	0.0344	0.1892	78.8101
fda-2	0.9991	0.0201	0.0162	0.0405	0.4993	0.0286	0.0235	0.0819	96.4486
fda-3	1.0003	0.0162	0.0128	0.0262	0.5000	0.0222	0.0181	0.0491	99.4752
fda-4	1.0003	0.0162	0.0129	0.0263	0.4999	0.0224	0.0182	0.0501	99.9261
fda-5	0.9999	0.0151	0.0120	0.0227	0.4994	0.0208	0.0170	0.0434	99.9900
fda-6	0.9999	0.0151	0.0120	0.0227	0.4994	0.0208	0.0170	0.0434	99.9987
fda-7	0.9999	0.0151	0.0120	0.0227	0.4994	0.0208	0.0170	0.0434	99.9998
fda-8	0.9999	0.0151	0.0120	0.0227	0.4994	0.0208	0.0170	0.0434	100.0000

Table 2.12 Performance of the estimation procedure where the residuals are generated with rational quadratic covariance function (e) where scale parameter $a_0 = 1$ and shape parameter $b_0 = 5$. Mean of the estimated coefficients, standard deviation, absolute bias, mean square error (×100) and FVE in percentage are summarized upto four decimal places.

Method	β_1				β_2				FVE
	Mean	SD	AB	MSE	Mean	SD	AB	MSE	%-age
					$n = 100$				
init	0.9967	0.0504	0.0407	0.2545	0.4965	0.0725	0.0577	0.5251	
ldaAR	0.9986	0.0466	0.0369	0.2173	0.4965	0.0671	0.0539	0.4508	
ldaCS	0.9970	0.0473	0.0377	0.2238	0.4978	0.0668	0.0531	0.4458	
fda-1	0.9959	0.0648	0.0517	0.4205	0.4960	0.0922	0.0742	0.8505	62.1253
fda-2	0.9964	0.0505	0.0405	0.2561	0.4976	0.0721	0.0572	0.5199	89.2976
fda-3	0.9970	0.0462	0.0368	0.2136	0.4974	0.0644	0.0513	0.4142	97.4903
fda-4	0.9981	0.0464	0.0363	0.2149	0.4957	0.0647	0.0520	0.4201	99.4795
fda-5	0.9986	0.0441	0.0345	0.1941	0.4952	0.0621	0.0497	0.3872	99.9012
fda-6	0.9987	0.0446	0.0350	0.1986	0.4958	0.0622	0.0500	0.3885	99.9827
fda-7	0.9986	0.0440	0.0343	0.1934	0.4959	0.0625	0.0504	0.3921	99.9971
fda-8	0.9986	0.0440	0.0343	0.1934	0.4959	0.0625	0.0504	0.3921	99.9995
					$n = 300$				
init	0.9977	0.0286	0.0229	0.0820	0.4991	0.0406	0.0332	0.1646	
ldaAR	0.9975	0.0271	0.0215	0.0737	0.4991	0.0367	0.0296	0.1346	
ldaCS	0.9983	0.0272	0.0216	0.0739	0.4994	0.0362	0.0288	0.1309	
fda-1	0.9972	0.0355	0.0283	0.1264	0.4991	0.0539	0.0441	0.2896	62.2661
fda-2	0.9983	0.0289	0.0229	0.0835	0.4990	0.0384	0.0308	0.1473	89.2161
fda-3	0.9981	0.0266	0.0209	0.0712	0.4992	0.0355	0.0282	0.1257	97.4664
fda-4	0.9981	0.0260	0.0205	0.0677	0.4992	0.0351	0.0278	0.1227	99.4727
fda-5	0.9984	0.0244	0.0192	0.0594	0.4996	0.0332	0.0267	0.1097	99.8990
fda-6	0.9985	0.0243	0.0191	0.0589	0.4997	0.0334	0.0269	0.1115	99.9820
fda-7	0.9985	0.0237	0.0189	0.0562	0.4998	0.0334	0.0270	0.1112	99.9969
fda-8	0.9985	0.0237	0.0189	0.0562	0.4998	0.0334	0.0270	0.1112	99.9995
					$n = 500$				
init	0.9996	0.0207	0.0165	0.0430	0.4992	0.0304	0.0245	0.0921	
ldaAR	0.9993	0.0187	0.0151	0.0351	0.4985	0.0281	0.0229	0.0792	
ldaCS	0.9996	0.0201	0.0160	0.0404	0.4997	0.0282	0.0230	0.0792	
fda-1	0.9997	0.0256	0.0204	0.0654	0.4988	0.0385	0.0304	0.1484	62.3377
fda-2	0.9992	0.0210	0.0168	0.0440	0.4995	0.0299	0.0244	0.0892	89.2460
fda-3	1.0001	0.0193	0.0153	0.0372	0.4999	0.0270	0.0219	0.0728	97.4696
fda-4	0.9997	0.0188	0.0150	0.0355	0.4993	0.0269	0.0219	0.0724	99.4726
fda-5	0.9994	0.0180	0.0142	0.0322	0.4989	0.0260	0.0211	0.0674	99.8988
fda-6	0.9993	0.0180	0.0142	0.0323	0.4988	0.0258	0.0210	0.0668	99.9818
fda-7	0.9992	0.0176	0.0140	0.0309	0.4988	0.0258	0.0209	0.0663	99.9969
fda-8	0.9992	0.0176	0.0140	0.0309	0.4988	0.0258	0.0209	0.0663	99.9995

2.4.3 Simulation results

Simulation results associated with the Brownian motion are shown in Table 2.1. In this situation, we observe that our approach produces better results in terms of the dispersion measures. Tables 2.2, 2.3 and 2.4 show results for linear processes, here our proposed method performs better in situations with working correlation matrix as AR, but is comparable for an exchangeable structure for $l_0 = 1, 2, 3$. Moreover, in our proposed method, as l increases, all dispersion measures, such as MSE, decrease. The results based on the OU process are documented in Tables 2.5 and 2.6. Our method outperforms the existing methods for both situations and, as μ_0 increases, the MSE decreases. For three different parameter choices of the power exponential and rational quadratic covariance structure, numerical results are presented in Tables 2.7, 2.8, 2.9 and 2.10, 2.11, 2.12, respectively. As before, we observe that our proposed method is finer than the existing ones in all sub-cases; but interestingly, when b_0 increases, MSE decreases for the power exponential, whereas it increases for the rational quadratic covariance structure, as expected due to the covariance structure. Overall, we observe that for all the above situations, as sample size increases, the dispersion measures, for example SD and MSE decrease. It establishes that as the sample size increases, the parameter estimates get closer and closer to the true parameters. From empirical studies, it was observed that if the estimated number of principal components $\widehat{\kappa} > \kappa_0$ then $\mathcal{Q}(\widehat{\boldsymbol{\beta}})$ may not be continuous at $\widehat{\boldsymbol{\beta}}$. In each of the above situations, the SDs of the proposed methods decrease as we increase κ_0 and stabilize after some value of κ_0 where the fraction of variance (FVE) is approximately 100%.

2.5 Real data analysis

In this section, we apply our proposed method to motivating examples in two different data-sets.

2.5.1 Beijing's $PM_{2.5}$ pollution study

In the atmosphere, suspended microscopic particles of solid and liquid matter are commonly known as particulates or particulate matter (PM). Such particulates often have a strong noxious

impact on human health, climate, and visibility. One such common and fine type of atmospheric particle is $PM_{2.5}$ with a diameter less than 2.5 micrometers. Many developed and developing cities across the world are experiencing chronic air pollution, with major pollutant being $PM_{2.5}$; Beijing and a substantial part of China are among such places. Some studies show that there are many non-ignorable sources of variability in the distribution and transmission pattern of $PM_{2.5}$, which are confounded with secondary chemical generation. The atmospheric $PM_{2.5}$ data used in our analysis were collected from the UCI machine learning repository https://archive.ics.uci.edu/ml/datasets/Beijing+Multi-Site+Air-Quality+Data (Liang et al., 2015). The data-set includes daily measurements of $PM_{2.5}$ and associated covariates at twelve different locations in China, viz., Aotizhongxin, Changping, Dingling, Dongsi, Guanyuan, Gucheng, Huairou, Nongzhanguan, Shunyi, Tiantan, Wanliu, and Wanshouxigong during January 2017. After excluding missing data, there were 608 hourly data points in *Beijing2017-data*. We assume that the atmospheric measurements are independent since they are located quite apart. The objective of our analysis is to describe the trend of the functional response $PM_{2.5}$ (as shown in Figure 2.1) and to evaluate the effect of covariates including chemical compounds such as sulfur dioxide (SO_2), nitrogen dioxide (NO_2), carbon monoxide (CO) and ozone (O_3) over time. We smoothed the covariates and responses to reduce variability and center them. Subsequently, we consider the following model

$$Y_i(t) = \beta_0 + \mathrm{SO}_2(t)\beta_1 + \mathrm{NO}_2(t)\beta_2 + \mathrm{CO}(t)\beta_3 + \mathrm{O}_3(t)\beta_4 + e_i(t) \tag{2.16}$$

We use Algorithm 2.1 to estimate the coefficients of the regression model mentioned above. Through the simulation results, we observe that if the values of κ_0 increase, the standard deviation of the coefficients decreases. For small FVEs such as 50%, the corresponding $\kappa_0 = 1$ and the estimation procedure performs poorly; whereas for large FVE percentages, the estimation procedure has adequately improved in terms of standard error. The estimated values for $\beta_0, \beta_1, \beta_2, \beta_3$, and β_4 produce similar results across different choices of κ_0. From the estimated standard error, using the scree plot, we conclude that the suitable choice of κ_0 is approximately 10. The estimated scaled eigen-values are provided in Figure 2.2 which clearly shows their decay rate. The estimated

coefficients with standard error are 0.0009 (1.1644), 0.0829 (0.2584), 0.9503 (0.1586), 0.0196 (0.0037) and 1.1523 (0.1198) respectively.

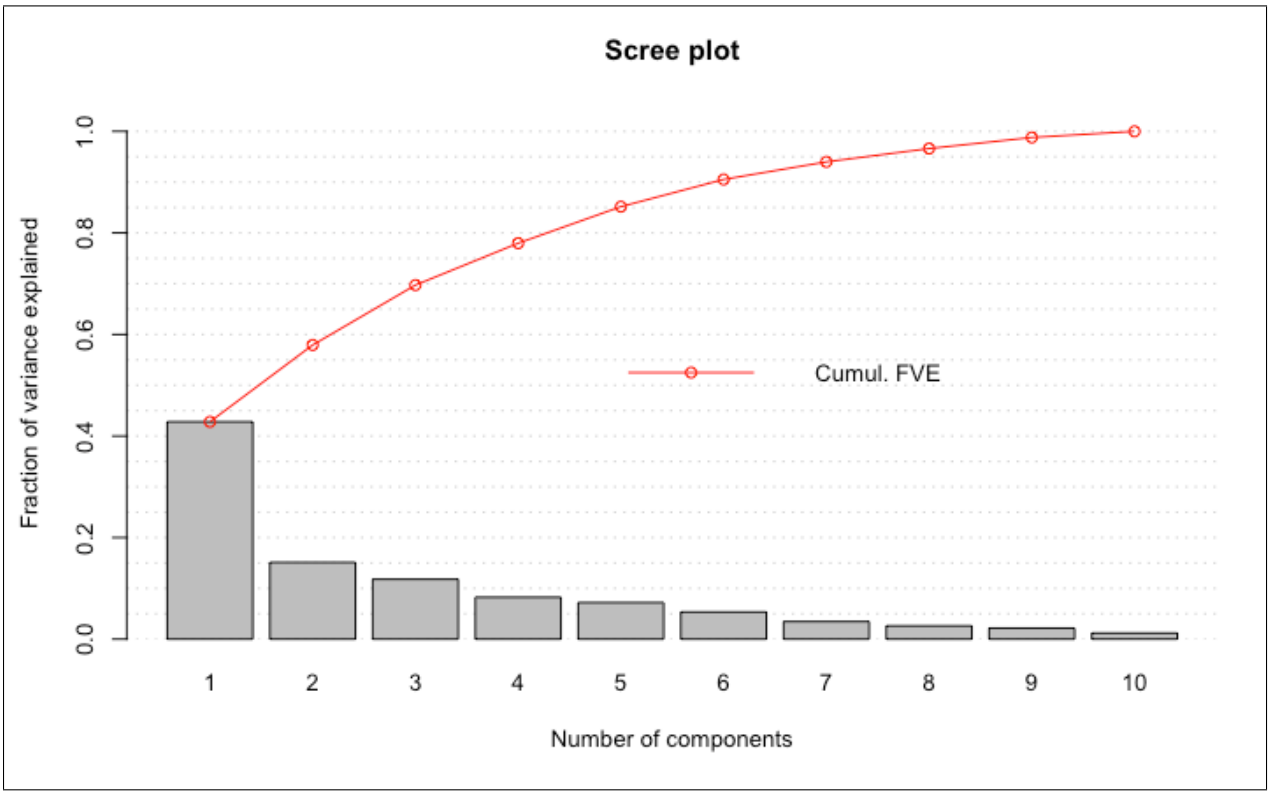

Figure 2.2 *Beijing2017-data* results: Scree plots of fraction of variance explained (FVE).

2.5.2 DTI study for sleep apnea patients

MRI is a powerful technique for investigating the structural and functional changes in the brain during pathological and neuro-psychological processes. Due to the advancement in diffusion tensor imaging (DTI), several studies on white matter alterations associated with clinical variables can be found in the recent literature. For our analysis, we use *Apnea-data* obtained from one such study on obstructive sleep apnea (OSA) patients (Xiong et al., 2017). The data consist of 29 male patients between the ages of 30-55 years who underwent a study for the diagnosis of continuous positive airway pressure (CPAP) therapy. Among those who have sleep disorder other than OSA, night-shift workers, patients with psychiatric disorders, hypertension, diabetes, and other neurological disorders were excluded. In this study, the psychomotor vigilance task (PVT) was performed in

which a light was randomly switched-on on a screen for several seconds in a certain interval of time and subjects were asked to press a button as soon as they saw the light appear on screen; such an experiment provides a numerical measure of sleepiness by counting the number of "lapses" for each individual.

DTI was performed on a 3T MRI scanner using a commercial 32-channel head coil, followed by analysis using tract-based spatial statistics to investigate the difference in fractional anisotropy (FA) and other DTI parameters. Image acquisition is as follows. An axial T1-weighted image of the brain (3D-BRAVO) is collected with repetition time (TR) = 12ms, echo time (TE) = 5,2ms, flip angle = 13°, inversion time = 450 ms, matrix = 384×256, voxel size = $1.2 \times 0.57 \times 0.69$mm and scan time = 2 min 54 s. DTI are obtained in the axial plane using a spin-echo echo planner imaging sequence with TR = 4500ms, TE = 89.4ms, field of view = $20 \times 20\text{cm}^2$, matrix size = 160×132, slice thickness = 3mm, slice spacing = 1mm, b-values = 0, 1000 s/mm^2.

Our objective is to investigate the structural alteration of white matter using DTI in the patients with OSA over each voxel at various regions of the brain (called ROIs). Thus, our response variable is one of the DTI parameters, viz., fractional anisotropy (FA) and we are interested in studying the effect of the changes of FA over continuous domain such as voxels with the interaction of the lapses and the voxel locations in each ROIs. We consider the following model for each ROI.

$$\text{FA}_i(s) = \beta_0 + \beta_1 \text{lapses}_i \times s + e_i(s) \tag{2.17}$$

where $s \in \mathcal{S}$, a set of voxels in the considered ROIs. Using the Algorithm 2.1, we estimate the coefficients β_1 and β_2 as mentioned in the model 2.17 and the results are presented in Table 2.13. We find that the coefficient estimates are close enough to their initial estimates and the estimated standard error is smaller for the coefficients based on the proposed method. Here κ_0 (i.e., the number of eigen-functions) are determined by FVE for simplicity.

2.6 Discussion

In this chapter, we propose an estimation procedure for the constant linear effects model, which is commonly used in statistics (Zhang and Banerjee, 2021) especially in spatial modelling. One of

Table 2.13 *Apnea-data* results: Estimated values and associated standard errors for the regression coefficients are provided upto four decimal places based on the existing and proposed methods. First line corresponding to each ROI shows results based on initial estimates and the second line corresponds to that of proposed estimates.

		β_0		β_1	
	# functional points	Estimate	Std. Error (×100)	Estimate (×100)	Std. Error (×100)
ROI.6	659	0.4512	0.1343	-0.0606	0.0130
		0.4512	0.0983	-0.0605	0.0028
ROI.7	1362	0.5048	0.0628	0.0309	0.0061
		0.5050	0.0681	0.0342	0.0007
ROI.8	1370	0.5256	0.0586	-0.0667	0.0057
		0.5271	0.0346	-0.0733	0.0006
ROI.9	690	0.4951	0.0910	0.2904	0.0088
		0.5443	0.0874	0.1660	0.0014
ROI.10	699	0.4951	0.0892	0.3314	0.0086
		0.5262	0.1398	0.4231	0.0014
ROI.11	968	0.4372	0.0979	0.1323	0.0095
		0.4380	0.0637	0.1311	0.0009
ROI.12	968	0.4529	0.0948	0.0965	0.0092
		0.4664	0.0750	0.0504	0.0013
ROI.13	992	0.5448	0.1060	0.3453	0.0103
		0.5449	0.0856	0.3559	0.0011
ROI.14	992	0.5435	0.1068	0.3432	0.0104
		0.5436	0.0754	0.3437	0.0003
ROI.37	1236	0.3695	0.0779	-0.1126	0.0076
		0.3713	0.0669	-0.1175	0.0017
ROI.38	1155	0.3564	0.0819	-0.1356	0.0079
		0.3578	0.0420	-0.1356	0.0009
ROI.39	1124	0.4618	0.0760	0.1972	0.0074
		0.4621	0.0615	0.1996	0.0007
ROI.40	1125	0.4786	0.0658	0.0953	0.0064
		0.4780	0.0369	0.1016	0.0005
ROI.45	380	0.4189	0.1071	0.1647	0.0104
		0.4190	0.0175	0.1648	0.0001
ROI.46	376	0.4074	0.1033	0.1988	0.0100
		0.4074	0.0159	0.1994	0.0002
ROI.47	596	0.4596	0.0932	0.1304	0.0090
		0.4594	0.0191	0.1349	0.0001
ROI.48	600	0.4045	0.0868	0.1100	0.0084
		0.4036	0.0644	0.1067	0.0006

the key factors of this estimation procedure is the fact that it is based on the quadratic inference methodology, which has played a huge role in the analysis of correlated data since it was discovered by Qu et al. (2000). In contrast to the existing method, our approach allows the number of repeated measurements to grow with sample size; therefore, the trajectories of individuals can be observed on a dense grid of a continuum domain. Instead of assuming a working correlation structure, we propose a data-driven way by estimating the eigen-functions that are obtained by functional principal component analysis. Here, we achieve $\sqrt{n}-$consistency of the parametric estimates in the regression model, even though the eigen-functions are estimated non-parametrically.

Additionally, our method is easy to implement in a wide range of applications. The applicability of the proposed method is illustrated by extensive simulation studies. Moreover, two real-data applications in different scientific domains confirm the efficacy of the proposed method.

2.7 Technical details

2.7.1 Some preliminary definitions and concepts of operators

In this section, we discuss some basic concepts of operators and discuss some useful properties on it. This can be found in Dunford and Schwartz (1988); Riss and Sz-Nagy (1990) along many more textbooks in functional analysis. Since FDA deals with continuous-time stochastic process, we need to be equipped with the dealing of random function and hence an overview of functional spaces is required. Perturbation theory of compact operators is required to discuss functional principal component analysis and these are demonstrated here with some useful results in our context. In the next subsection, we discuss the functional principal component analysis in brief and show the estimation techniques of eigen-values and eigen-functions based on the data at hand. This plays a fundamental role in our proposed method.

Consider the standard $\mathscr{L}^2[0,1]$ space that defines the set of square-integrable functions defined on the closed set $[0,1]$ that takes values on the real line. The space $\mathscr{L}^2[0,1]$ is equipped with an inner product and is defined as

$$\langle f, g\rangle = \int_0^1 f(t)g(t)dt \tag{2.18}$$

for f and g in that space, and forms a Hilbert space. Moreover, we denote the norm $\|\cdot\|_2$ in $\mathscr{L}^2$ which is defined as $\|f\|_2 = \left\{\int f^2(u)du\right\}^{1/2}$. Define $\mathscr{F}$ be an operator that assigns an element f in $\mathscr{L}^2[0,1]$ to a new element $\mathscr{F}f$ in $\mathscr{L}^2[0,1]$. The operator is linear if $\mathscr{F}(\alpha f + \beta g) = \alpha\mathscr{F}f + \beta\mathscr{F}g$ for any scalar α and β. It is said to be bounded if for any positive constant M (which may depend on f) we have $\|\mathscr{F}f\| \leq M\|f\|$ for all $f \in \mathscr{L}^2[0,1]$ where the largest bound for all M is called the norm of the operator $\mathscr{F}$, denoted by $\|\mathscr{F}\|$, and it is defined as $\|\mathscr{F}\| = \sup_{\|f\|\leq 1}\|\mathscr{F}f\|$. The operator is bounded if and only if it is continuous. $\mathscr{F}$ is said to be self-adjoint if $\langle\mathscr{F}f, g\rangle = \langle f, \mathscr{F}g\rangle$ and becomes non-negetive definite if $\langle\mathscr{F}f, f\rangle \geq 0$ for all $f \in \mathscr{L}^2[0,1]$.

A linear mapping $\mathscr{F}f(\cdot) = \int R(\cdot,u)f(u)du$ for any function $f \in \mathscr{L}^2[0,1]$ and for some integrable function $R(\cdot,\cdot)$ on $[0,1]\times[0,1]$. This function is preferably known as integral operator and the bivariate function R is known as a kernel in statistics and functional analysis literature. Note that the above linear mapping is bounded since

$$\begin{aligned} |\mathscr{F}f(t)|^2 &\leq \int R^2(t,u)du \int f^2(u)du \qquad \text{using Cauchy-Schwarz inequality} \\ &= \|f\|_2^2 \int R^2(t,u)du \end{aligned} \tag{2.19}$$

Furthermore, $\|\mathscr{F}f\|_2^2 \leq \|f\|_2^2 \int\int R^2(s,t)dsdt$. under the assumption that $\int\int R^2(u,v)dudv < \infty$. It is easy to see that $\mathscr{F}f(\cdot)$ is uniformly continuous and compact for a non-negative definite symmetric kernel R.

For some λ, in Fredholm integral equation , $\mathscr{F}\phi = \lambda\phi$ has non-zero solution ϕ then we call λ as eigen-value of $\mathscr{F}$ and the solution of the eigen-equation is called eigen-functions, altogether, the pair of eigen-values and eigen-function, viz., (λ,ϕ) are called eigen-elements. Let f_1 and f_2 be the elements in Hilbert space $\mathscr{H}$ then the tensor product operator $(f_1 \bigotimes f_2) : \mathscr{H} \to \mathscr{H}$ and defined by $(f_1 \bigotimes f_2)(u) = \langle f_1, g\rangle f_2$ for $g \in \mathscr{H}$. For a compact self-adjoint operator in $\mathscr{L}^2[0,1]$, let $\{(\lambda_k,\phi_k) : k \geq 1\}$ be as set of eigen-components such that ϕ_ks are orthogonal. Then for any function $f \in \mathscr{L}^2[0,1]$, it can be represented as

$$f = f_0 + \sum_{r=1}^{\infty} \mathscr{P}_r \tag{2.20}$$

where $\mathscr{P}_r$ is the projection operator for eigen-spaces λ_r. In our particular situation $\mathscr{P}_r = \phi_r \bigotimes \phi_r$. Moreover, for a suitable f_0 such that $\mathscr{F} f_0 = 0$, it can be shown that

$$\mathscr{F} = \sum_{r=1}^{\infty} \lambda_r \mathscr{P}_r \tag{2.21}$$

where λ_r is repeated as its multiplicity. Due to non-negative definiteness of $\mathscr{F}$, the eigen-values are ordered as $\lambda_1 \geq \lambda_2 \geq \cdots \geq 0$.

Now we discuss Mercer's theorem (J Mercer, 1909) for a symmetric continuous non-negative definite kernel function R. It states that, for $\{(\lambda_r, \phi_r)\}_{r\geq 1}$ be the set of eigen-components, such kernel R has the following representation

$$R(s,t) = \sum_{r=1}^{\infty} \lambda_r \phi_r(s)\phi_r(t) \tag{2.22}$$

where the sum is absolutely and uniformly convergence.

In this paper, we assume that the eigen-values are distinct for mathematical simplicity. We conclude this subsection perturbation theory of compact operators in the sense that every subsequences of $\{\mathscr{F} f_n\}$ is a Cauchy sequence. In the statistical literature, this can be found in Hall and Hosseini-Nasab (2006, 2009). This is useful to find the bound of eigen-components in different applications.

Suppose for self-adjoint compact operator on Hilbert space $\mathscr{H}$ consider two operators $\mathscr{F}$ and $\mathscr{G}$, define perturbation operator $\Delta = \mathscr{G} - \mathscr{F}$ such that $\mathscr{G} = \mathscr{F} + \Delta$ where $\mathscr{G}$ is an approximation to $\mathscr{F}$ where Δ amount of error is occurred. Let $\mathscr{F}$ and $\mathscr{G}$ have kernels F and G respectively with eigen-elements (θ_r, ψ_r) and (λ_r, ϕ_r). For simplicity, we assume that the eigen-values are distinct. Then the following Lemma provides perturbation of the eigen-functions.

Lemma 2.7.1 (Theorem 5.1.8 in Hsing and Eubank (2015))**.** *Let (λ, ϕ) be the eigen-components of $\mathscr{F}$ and (θ, ψ) be that of $\mathscr{G}$ with multiplicity of all eigen-values are restricted to be 1. Define $\eta_k = \min_{r\neq k} |\lambda_r - \lambda_k|$. Assume $\langle \phi_r, \psi_r \rangle \geq 0$ and $\eta_k > 0$. Then*

$$\psi_k - \phi_k = \sum_{\substack{r=1 \\ r\neq k}}^{\infty} (\theta_k - \lambda_r)^{-1} \mathscr{P}_r \Delta \psi_k + \mathscr{P}_k(\psi_k - \phi_k) \tag{2.23}$$

The above equation follows

$$\psi_k - \phi_k = \sum_{\substack{r=1 \\ r\neq k}}^{\infty} (\theta_k - \lambda_r)^{-1} \mathscr{P}_r \Delta \psi_k + O(\|\Delta\|^2) \tag{2.24}$$

Remark 2.7.1. *Equation (2.24) plays an important role in finding the bound of the proposed estimator introduced in Section 2.2.2. Note that* $\sup_{r\geq 1} |\theta_r - \lambda_r| \leq \|\Delta\| \leq \inf_{r\neq k} |\lambda_k - \lambda_r|$ *(see Theorem 4.2.8 in Hsing and Eubank (2015) for proof). Thus, it is easy to see,* $|\theta_r - \lambda_r| \leq |\lambda_k - \lambda_r|$ *which implies from Equation (2.23)*

$$\begin{aligned}
\psi_k - \phi_k &= \sum_{\substack{r=1 \\ r\neq k}}^{\infty} (\lambda_k - \lambda_r)^{-1} \sum_{s=0}^{\infty} \left(\frac{\lambda_k - \theta_r}{\lambda_k - \lambda_r}\right)^s \mathscr{P}_r \Delta\{\phi_k + (\psi_k - \phi_k)\} + \mathscr{P}_k(\psi_k - \phi_k) \\
&= \sum_{\substack{r=1 \\ r\neq k}}^{\infty} (\lambda_k - \lambda_r)^{-1} \mathscr{P}_r \Delta \phi_k + \sum_{\substack{r=1 \\ r\neq k}}^{\infty} (\lambda_k - \lambda_r)^{-1} \mathscr{P}_r \Delta(\psi_k - \phi_k) \\
&\quad + \sum_{\substack{r=1 \\ r\neq k}}^{\infty} \sum_{s=1}^{\infty} \frac{(\lambda_k - \lambda_s)^s}{(\lambda_k - \lambda_r)^{s+1}} \mathscr{P}_r \Delta \psi_k + \mathscr{P}_k(\psi_k - \phi_k)
\end{aligned} \tag{2.25}$$

Moreover, using Bessel's inequality [1] *, we can bound last three terms in the above equation by* $\|\Delta\|^2$.

Another useful tool in functional data analysis is Karhunen-Loève expansions Karhunen (1946); Loève (1946) for random function $e(t)$ which is mean zero, second order process with kernel R is defined in Sub-section 2.2.3. It states that, with probability 1, the random function can be expressed as

$$e_i = \sum_{r=1}^{\infty} \xi_{ir} \phi_r, \qquad \text{where } \xi_{ir} := \langle e_i, \phi_r \rangle \tag{2.26}$$

The random variables ξ_{ir} are uncorrelated with mean zero and variance λ_r. This provides a sufficient and necessary condition for the decomposition of a random process.

2.7.2 Some useful lemmas

In this section, we represent some useful lemmas. For convenience, let us recall the notation. Assume that m_is are all of the same order, viz, $m \equiv m(n)$. Define, $d_{n1}(h) = h^2 + h\overline{m}/m$ and $d_{n2}(h) =$

[1] Bessel's inequality: for any $f \in \mathscr{H}$, $\sum_{k=1}^{\infty} |\langle f, \phi_k \rangle|^2 \leq \|f\|^2$

$h^4 + h^3\overline{m}/m + h^2\overline{\overline{m}}/m^2$ where $\overline{m} = \lim_{n\to\infty} n^{-1}\sum_{i=1}^{n} m/m_i$ and $\overline{\overline{m}} = \lim_{n\to\infty} n^{-1}\sum_{i=1}^{n}(m/m_i)^2$. Denote $\delta_{n1}(h) = \left\{d_{n1}(h)\log n/(nh^2)\right\}^{1/2}$, $\delta_{n2}(h) = \left\{d_{n2}(h)\log n/(nh^4)\right\}^{1/2}$ and $\overline{\delta}_n(h) = h^2 + \delta_{n1}(h) + \delta_{n2}^2(h)$. Further, $\nu_{a,b} = \int t^a K^b(t)dt$. Define, $\mathbf{W} = (\boldsymbol{\phi}(t_1)^{\mathrm{T}}, \cdots, \boldsymbol{\phi}(t_m)^{\mathrm{T}})^{\mathrm{T}}$ be matrix of order $m \times \kappa_0$ obtained after stacking all $\boldsymbol{\phi}_k$s and random components $\boldsymbol{\xi}_i = (\xi_{i1}, \cdots, \xi_{i\kappa_0})^{\mathrm{T}}$. Furthermore, $\boldsymbol{\xi}$ has mean zero and variance $\boldsymbol{\Lambda}$ which is a diagonal matrix with components $\lambda_1, \cdots, \lambda_{k_0}$. Define the indicator function $\mathbf{1}(a = b) = 1$ if $a = b$ and zero otherwise.

Lemma 2.7.2. *Consider $Z_1, \cdots, Z_n$ be independent and identically distributed random variables with mean zero and finite variance. Suppose that there exists an M such that $P(|Z_i| \leq M) = 1$ for all $i = 1, \cdots, n$. Let $T_n = \frac{1}{n}\sum_{i=1}^{n} Z_i$. then, $\frac{1}{n}\sum_{i=1}^{n} Z_i = O((\log n/n)^{1/2}$ almost surely. If $\sqrt{Var(T_n)} = O((\log n/n)^{1/2})$ then $T_n = O(\log n/n)$ almost surely.*

Proof. Bernstein's inequality states that if $Z_1, \cdots, Z_n$ be centered independent bounded random variables with probability 1. Let $T_n = \frac{1}{n}\sum_{i=1}^{n} Z_i$, then let $\mathrm{Var}\{T_n\} = \sigma_n^2$. Then for any positive real number u, we have $P(|T_n| \geq u) \leq \exp\{-\frac{nu^2}{2\sigma_n^2 + 2Mu/3}\}$ where M is such that $P(|Z_i| \leq M) = 1$. Moreover, if T_n converges to its limit in probability fast enough, then it converge almost surely in the limit, i.e., if for any $u > 0$, $\sum_{n=1}^{\infty} P(|T_n| \geq u) < \infty$ them T_n converges to zero almost surely. Now, choose $u = \sqrt{\frac{4\sigma_n^2 \log n}{n}} + \frac{4M\log n}{3n}$. Thus, $\sum_{n=1}^{\infty} P(|T_n| \geq u) < \sum_{n=1}^{\infty} 1/n^2$ which is finite. Therefore, $T_n = O(u)$ almost surely. Now let $\sigma_n \leq \sqrt{4M^2 \log n/9n}$, we have, $T_n = O(\log n/n)$ and if $\sigma_n = O(1)$ then $T_n = O((\log n/n)^{1/2})$ almost surely. □

Lemma 2.7.3. *Suppose T_{ij} are i.i.d. with density f_T. Then for fixed $i = 1, \cdots, n$, any k and $l \geq 1$, under assumptions (C2) and (C6)c, the following holds.*

$$\frac{1}{m_i}\sum_{j=1}^{m_i} \phi_k(T_{ij})\phi_l(T_{ij}) = \boldsymbol{1}(k = l) + O((\log m_i/m_i)^{1/2}) \qquad \textit{almost surely} \tag{2.27}$$

Proof. Note that,

$$\mathbb{E}\left\{\frac{1}{m_i}\sum_{j=1}^{m_i} \phi_k(T_{ij})\phi_l(T_{ij})\right\} = \int \phi_k(t)\phi_l(t)dt = \mathbf{1}(k = l) \tag{2.28}$$

and,

$$\begin{aligned}
&\mathrm{Var}\left\{\frac{1}{m_i}\sum_{j=1}^{m_i}\phi_k(T_{ij})\phi_l(T_{ij})\right\} = \mathbb{E}\left\{\frac{1}{m_i}\sum_{j=1}^{m_i}\phi_k(T_{ij})\phi_l(T_{ij})\right\}^2 - \mathbf{1}(k=1)\\
&= \frac{1}{m_i^2}\sum_{j=1}^{m_i}\mathbb{E}\{\phi_k^2(T_{ij})\phi_l^2(T_{ij})\} + \frac{1}{m_i^2}\sum_{\substack{j_1=1\\ j_1\neq j_2}}^{m_i}\sum_{j_2=1}^{m_i}\mathbb{E}\{\phi_k(T_{ij_1})\phi_k(T_{ij_2})\phi_l(T_{ij_1})\phi_l(T_{ij_2})\} - \mathbf{1}(k=l)\\
&= \frac{1}{m_i^2}\sum_{j=1}^{m_i}\mathbb{E}\{\phi_k^2(T_{ij})\phi_l^2(T_{ij})\} + \frac{1}{m_i^2}\sum_{\substack{j_1=1\\ j_1\neq j_2}}^{m_i}\sum_{j_2=1}^{m_i}\mathbb{E}\{\phi_k(T_{ij_1})\phi_l(T_{ij_1})\}\mathbb{E}\{\phi_k(T_{ij_2})\phi_l(T_{ij_2})\} - \mathbf{1}(k=l)\\
&= \begin{cases}
\frac{1}{m_i}\int\phi_k^4(t)dt + \frac{m_i-1}{m_i}(\int\phi_k^2(t)dt)^2 - 1 & \text{if } k=l\\
\frac{1}{m_i}\int\phi_k^2(t)\phi_l^2(t)dt + \frac{m_i-1}{m_i}(\int\int\phi_k(t)\phi_l(t)\phi_k(t')\phi_l(t')dtdt') & \text{if } k\neq l
\end{cases}\\
&= O(1/m_i)
\end{aligned} \tag{2.29}$$

Therefore, by applying Lemma 2.7.2, we get Equation (2.27). □

Lemma 2.7.4. *Suppose T_{ij} are i.i.d with density f_T. Then for fixed $i = 1, \cdots, n$, for any $k \geq 1$, under assumptions (C2), (C6)c, the following holds.*

$$\frac{1}{m_i}\sum_{j=1}^{m_i}\dot{\mu}_i(T_{ij})\phi_k(T_{ij}) = \int\dot{\mu}_i(t)\phi_k(t)dt + O\left((\log m_i/m_i)^{1/2}\right) \qquad \textit{almost surely} \tag{2.30}$$

Proof. Note that,

$$\mathbb{E}\left\{\frac{1}{m_i}\sum_{j=1}^{m_i}\dot{\mu}_i(T_{ij})\phi_k(T_{ij})\right\} = \int\dot{\mu}_i(t)\phi_k(t)dt \tag{2.31}$$

and,

$$\begin{aligned}
&\mathrm{Var}\left\{\frac{1}{m_i}\sum_{j=1}^{m_i}\dot{\mu}_i(T_{ij})\phi_k(T_{ij})\right\} \leq \mathbb{E}\left\{\frac{1}{m_i}\sum_{j=1}^{m_i}\dot{\mu}_i(T_{ij})\phi_k(T_{ij})\right\}^2\\
&= \frac{1}{m_i^2}\sum_{j=1}^{m_i}\mathbb{E}\left\{\dot{\mu}_i^2(T_{ij})\phi_k(T_{ij})^2\right\} + \frac{1}{m_i^2}\sum_{\substack{j_1=1\\ j_1\neq j_2}}^{m_i}\sum_{j_2=1}^{m_i}\mathbb{E}\{\dot{\mu}_i(T_{ij_1})\dot{\mu}_i(T_{ij_2})\phi_k(T_{ij_1})\phi_k(T_{ij_2})\}\\
&= O(1/m_i) \qquad \text{since } \int\dot{\mu}^2(t)\phi_k^2(t)dt < \infty
\end{aligned} \tag{2.32}$$

Therefore, applying the Lemma 2.7.2, we get Equation (2.30). □

Lemma 2.7.5. *Define,* $\mathscr{M}_{ir} = \int \dot{\mu}_i(t)\phi_r(t)dt$ *and* $V_r = \mathbb{E}\{\int \dot{\mu}(t)\phi_r(t)dt\}^2$ *for* $r \geq 2$. *Then under Conditions (C6)a, (C6)b, for some* $\alpha > 0$ *such that* $V_r\lambda_r^{-2}r^{1+\alpha} \to 0$ *as* $r \to \infty$ *(due to Condition (C6)b)*

$$\sum_{\substack{r=1\\r\neq k}}^{\infty}(\lambda_k - \lambda_r)^{-1}\frac{1}{n}\sum_{i=1}^{n}\mathscr{M}_{ir}\xi_{ik} = O\left((\log n/n)^{1/2}\lambda_k^{1/2}k^{(1-\alpha)/2}\right) \qquad \textit{almost surely} \tag{2.33}$$

Proof. It is easy to see that,

$$\mathbb{E}\left\{\sum_{\substack{r=1\\r\neq k}}^{\infty}(\lambda_k - \lambda_r)^{-1}\frac{1}{n}\sum_{i=1}^{n}\mathscr{M}_{ir}\xi_{ik}\right\} = 0 \tag{2.34}$$

Using the spacing condition among the eigen-values in (C6)a, for each $1 \leq k < r < \infty$ and for non-zero finite generic constant C_0,

$$\frac{\max\{\lambda_k, \lambda_r\}}{|\lambda_k - \lambda_r|} \leq C_0\frac{\max\{k, r\}}{|k - r|} \tag{2.35}$$

Similar kind of conditions can be invoked such as the convexity assumption, i.e. $\lambda_r - \lambda_{r+1} \leq \lambda_{r-1} - \lambda_r$ for all $r \geq 2$. Thus, using Inequality (2.35), for some $\alpha > 0$, with condition $V_r\lambda_r^{-2}r^{1+\alpha} \to 0$ as $r \to \infty$ we can write

$$\begin{aligned}
&\sum_{\substack{r=1\\r\neq k}}^{\infty}V_r(\lambda_k - \lambda_r)^{-2} \lesssim \sum_{\substack{r=1\\r\neq k}}^{\infty}V_r\left\{\frac{\max(k, r)}{|k - r|\max(\lambda_k, \lambda_r)}\right\}^2\\
&= \sum_{r\leq k/2}V_r\lambda_r^{-2}\frac{k^2}{(k-r)^2} + \sum_{r>2k}V_r\lambda_k^{-2}\frac{r^2}{(k-r)^2} + \sum_{k/2<r<k}V_r\lambda_r^{-2}\frac{k^2}{(k-r)^2} + \sum_{k<r<2k}V_r\lambda_k^{-2}\frac{r^2}{(k-r)^2}\\
&\lesssim \sum_{r\leq k/2, r>2k}V_r\lambda_r^{-2} + k^2\sum_{k/2<r<2k}V_r\lambda_r^{-2}(k-r)^{-2}\\
&\lesssim 1 + k^{1-\alpha}\sum_{k/2<r<2k}(k-r)^{-2} \lesssim k^{1-\alpha}.
\end{aligned} \tag{2.36}$$

This follows the line of proofs in Hall and Hosseini-Nasab (2009) in different contexts. Thus, using the Inequality (2.36), it follows that

$$\mathrm{Var}\left\{\sum_{\substack{r=1\\r\neq k}}^{\infty}(\lambda_k - \lambda_r)^{-1}\frac{1}{n}\sum_{i=1}^{n}\mathscr{M}_{ir}\xi_{ik}\right\} = \frac{1}{n}\lambda_k\sum_{\substack{r=1\\r\neq k}}^{\infty}V_r(\lambda_k - \lambda_r)^{-2} = O\left(n^{-1}\lambda_k k^{(1-\alpha)}\right) \tag{2.37}$$

Therefore, applying Lemma 2.7.2, we get Equation (2.33). □

Lemma 2.7.6. *For, $\mathscr{M}_{ir} = \int \dot{\mu}(t)\phi_r(t)dt$ and $\eta_k = \min_{r\neq k} |\lambda_k - \lambda_r| > 0$, under conditions (C6)a and (C6)b, we almost surely have the following.*

$$\sum_{\substack{r_1\neq 1\\ r_1\neq k}}^{\infty}\sum_{\substack{r_2\neq 1\\ r_2\neq k}}^{\kappa_0}(\lambda_k-\lambda_{r_1})^{-1}(\lambda_k-\lambda_{r_2})^{-1}\frac{1}{n}\sum_{i=1}^{n}\mathscr{M}_{ir_1}\xi_{ir_2} = O\left((\log n/n)^{1/2}\kappa_0^{(3-\alpha)/2}\lambda_{\kappa_0}^{-1}\left\{\sum_{r=1}^{\kappa_0}\lambda_r\right\}^{1/2}\right) \tag{2.38}$$

Proof. It is easy to see that,

$$\mathbb{E}\left\{\sum_{\substack{r_1=1\\ r_1\neq k}}^{\infty}\sum_{\substack{r_2=1\\ r_2\neq k}}^{\kappa_0}(\lambda_k-\lambda_{r_1})^{-1}(\lambda_k-\lambda_{r_2})^{-1}\frac{1}{n}\sum_{i=1}^{n}\mathscr{M}_{ir_1}\xi_{ir_2}\right\} = 0 \tag{2.39}$$

Moreover, using the spacing condition mentioned in (C6)a, one can derive the upper bound of η_k^{-1} by

$$\begin{aligned}\eta_k^{-1} &= \left\{\min_{r\neq k}|\lambda_k-\lambda_r|\right\}^{-1} = \max_{r\neq k}|\lambda_k-\lambda_r|^{-1}\\ &\lesssim \max_{r\neq k}\left\{\frac{\max(k,r)}{|k-r|\max(\lambda_k,\lambda_r)}\right\} \leq \lambda_k^{-1}k\end{aligned} \tag{2.40}$$

Due to the monotonic decreasing property of eigen-values, for fixed $k = 1, \cdots, \kappa_0$, we have

$$\sum_{\substack{r=1\\ r\neq k}}^{\kappa_0}\lambda_r(\lambda_k-\lambda_r)^{-2} \lesssim \eta_k^{-2}\sum_{r=1}^{\kappa_0}\lambda_r \lesssim \lambda_k^{-2}k^2\sum_{r=1}^{\kappa_0}\lambda_r \lesssim \lambda_{\kappa_0}^{-2}\kappa_0^2\sum_{r=1}^{\kappa_0}\lambda_r \tag{2.41}$$

Therefore, the following holds under similar conditions to obtain the Inequality (2.36),

$$\begin{aligned}&\mathbb{E}\left\{\sum_{\substack{r_1=1\\ r_1\neq k}}^{\infty}\sum_{\substack{r_2=1\\ r_2\neq k}}^{\kappa_0}(\lambda_k-\lambda_{r_1})^{-2}(\lambda_k-\lambda_{r_2})^{-2}\left(\frac{1}{n}\sum_{i=1}^{n}\mathscr{M}_{ir_1}\xi_{ir_2}\right)^2\right\}\\ &= \frac{1}{n}\sum_{\substack{r_1=1\\ r_1\neq k}}^{\infty}\sum_{\substack{r_1=1\\ r_1\neq k}}^{\kappa_0}(\lambda_k-\lambda_{r_1})^{-2}(\lambda_k-\lambda_{r_2})^{-2}V_{r_1}\lambda_{r_2}\\ &\lesssim \frac{1}{n}k^{1-\alpha}\sum_{\substack{r=1\\ r\neq k}}^{\kappa_0}\lambda_r(\lambda_k-\lambda_r)^{-2} \lesssim \frac{1}{n}\kappa_0^{3-\alpha}\lambda_{\kappa_0}^{-2}\sum_{r=1}^{\kappa_0}\lambda_r\end{aligned} \tag{2.42}$$

Therefore, applying the Lemma 2.7.2, we get Equation (2.38). □

2.7.3 Proof of Theorem 2.3.1

For the $k-$th element of $\overline{\mathbf{g}}_n(\boldsymbol{\beta}_0)$ $(1 \leq k \leq \kappa_0)$,

$$\begin{aligned}
\overline{\mathbf{g}}_{n,k}(\boldsymbol{\beta}_0) &= \frac{1}{n}\sum_{i=1}^{n}\frac{1}{m_i^2}\dot{\boldsymbol{\mu}}_i^{\mathrm{T}}\widehat{\boldsymbol{\Phi}}_k(\mathbf{y}_i - \boldsymbol{\mu}_i) \\
&= \frac{1}{n}\sum_{i=1}^{n}\frac{1}{m_i^2}\dot{\boldsymbol{\mu}}_i^{\mathrm{T}}\widehat{\boldsymbol{\Phi}}_k\mathbf{W}\boldsymbol{\xi}_i \\
&= \frac{1}{n}\sum_{i=1}^{n}\frac{1}{m_i^2}\dot{\boldsymbol{\mu}}_i^{\mathrm{T}}\boldsymbol{\Phi}_k\mathbf{W}\boldsymbol{\xi}_i + \frac{1}{n}\sum_{i=1}^{n}\frac{1}{m_i^2}\dot{\boldsymbol{\mu}}_i^{\mathrm{T}}(\widehat{\boldsymbol{\Phi}}_k - \boldsymbol{\Phi}_k)\mathbf{W}\boldsymbol{\xi}_i \\
&:= J_k^{n1} + J_k^{n2} \qquad (2.43)
\end{aligned}$$

Now, using Lemmas 2.7.3 and 2.7.4, the first part of the expression of $\overline{\mathbf{g}}_{n,k}(\boldsymbol{\beta}_0)$ becomes

$$\begin{aligned}
J_k^{n1} &= \frac{1}{n}\sum_{i=1}^{n}\frac{1}{m_i^2}\dot{\boldsymbol{\mu}}_i^{\mathrm{T}}\boldsymbol{\Phi}_k\mathbf{W}\boldsymbol{\xi}_i \\
&= \frac{1}{n}\sum_{i=1}^{n}\frac{1}{m_i^2}\sum_{j_1=1}^{m_i}\sum_{j_2=1}^{m_i}\sum_{l=1}^{\kappa_0}\dot{\mu}_i(T_{ij_1})\phi_k(T_{ij_1})\phi_k(T_{ij_2})\phi_l(T_{ij_2})\xi_{il} \\
&= \frac{1}{n}\sum_{i=1}^{n}\frac{1}{m_i}\sum_{j_1=1}^{m_i}\dot{\mu}_i(T_{ij_1})\phi_k(T_{ij_1})\left\{\sum_{l=1}^{\kappa_0}[\mathbf{1}(k=l) + O((\log m/m)^{1/2})]\right\}\xi_{il} \\
&\lesssim \frac{1}{n}\sum_{i=1}^{n}\left\{\mathscr{M}_{ik} + O((\log m/m)^{1/2})\right\}\left\{1 + O((\log m/m)^{1/2})\right\}\xi_{ik} \quad \text{where } \mathscr{M}_{ik} = \int \dot{\mu}_i(t)\phi_k(t)dt \\
&= \frac{1}{n}\sum_{i=1}^{n}\mathscr{M}_{ik}\xi_{ik}\left\{1 + O\left((\log m/m)^{1/2}\right)\right\} \\
&= O\left((\log n/n)^{1/2}\left\{1 + (\log m/m)^{1/2}\right\}\right) \qquad \text{almost surely} \qquad (2.44)
\end{aligned}$$

On the other hand, the last part of $\overline{\mathbf{g}}_{n,k}(\boldsymbol{\beta})$ can be expressed as

$$\begin{aligned}
J_k^{n2} &= \frac{1}{n}\sum_{i=1}^{n}\frac{1}{m_i^2}\dot{\boldsymbol{\mu}}_i^{\mathrm{T}}(\widehat{\boldsymbol{\Phi}}_k - \boldsymbol{\Phi}_k)\mathbf{W}\boldsymbol{\xi}_i \\
&= \frac{1}{n}\sum_{i=1}^{n}\frac{1}{m_i^2}\sum_{j_1=1}^{m_i}\sum_{j_2=1}^{m_i}\sum_{l=1}^{\kappa_0}\dot{\mu}_i(T_{ij_1})d_{ik}(T_{ij_1},T_{ij_2})\phi_l(T_{ij_2})\xi_{il} \qquad (2.45)
\end{aligned}$$

where $d_{ik}(T_{ij_1},T_{ij_2}) := \widehat{\phi}_k(T_{ij_1})\widehat{\phi}_k(T_{ij_2}) - \phi_k(T_{ij_1})\phi_k(T_{ij_2})$. Now replace $\mathscr{G}$, θ and ψ by $\widehat{\mathscr{F}}$, $\widehat{\lambda}$ and $\widehat{\phi}$ respectively since $\widehat{\mathscr{F}}$ be the approximation of $\mathscr{F}$ and Δ is the corresponding perturbation operator

in Equation (2.23). Therefore, Lemma 2.7.1 immediately implies the following expansion, which is the key fact to represent the objective function in QIF.

$$\widehat{\phi}_k - \phi_k = \sum_{\substack{r=1 \\ r\neq k}}^{\infty} (\lambda_k - \lambda_r)^{-1} \langle \phi_r, \Delta\phi_k \rangle \phi_r + O(\|\Delta\|^2) \qquad \text{almost surely} \tag{2.46}$$

where Δ be the integral operator with kernel $\widehat{R} - R$. Therefore, almost surely, we have,

$$\begin{aligned}
&d_{ik}(T_{ij_1}, T_{ij_2}) := \widehat{\phi}_k(T_{ij_1})\widehat{\phi}_k(T_{ij_2}) - \phi_k(T_{ij_1})\phi_k(T_{ij_2}) \\
&= \left\{ \phi_k(T_{ij_1}) + \sum_{\substack{r=1 \\ r\neq k}}^{\infty} (\lambda_k - \lambda_r)^{-1} \langle \phi_r, \Delta\phi_k \rangle \phi_r(T_{ij_1}) + O(\|\Delta\|^2) \right\} \\
&\qquad \times \left\{ \phi_k(T_{ij_2}) + \sum_{\substack{r=1 \\ r\neq k}}^{\infty} (\lambda_k - \lambda_r)^{-1} \langle \phi_r, \Delta\phi_k \rangle \phi_r(T_{ij_2}) + O(\|\Delta\|^2) \right\} - \phi_k(T_{ij_1})\phi_k(T_{ij_2}) \\
&= \sum_{\substack{r=1 \\ r\neq k}}^{\infty} (\lambda_k - \lambda_r)^{-1} \langle \phi_r, \Delta\phi_k \rangle \left\{ \phi_r(T_{ij_1})\phi_k(T_{ij_2}) + \phi_k(T_{ij_1})\phi_r(T_{ij_2}) \right\} \\
&\qquad + \sum_{r_1\neq k} \sum_{r_2\neq k} (\lambda_k - \lambda_{r_1})^{-1}(\lambda_k - \lambda_{r_2})^{-1} \left\langle \phi_{r_1}, \Delta\phi_k \right\rangle \left\langle \phi_{r_2}, \Delta\phi_k \right\rangle \phi_{r_1}(T_{ij_1})\phi_{r_2}(T_{ij_2}) + O(\|\Delta\|^2) \\
&:= I^{n1}_{ik}(T_{ij_1}, T_{ij_2}) + I^{n2}_{ik}(T_{ij_1}, T_{ij_2}) + O(\|\Delta\|^2) \qquad \text{almost surely}
\end{aligned} \tag{2.47}$$

Therefore, by placing the expression of $d_{ik}(T_{ij_1}, T_{ij_2})$ in Equation (2.45), we have the following.

$$\begin{aligned}
J^{n2}_k &= \frac{1}{n} \sum_{i=1}^{n} \frac{1}{m_i^2} \dot{\boldsymbol{\mu}}_i^{\mathrm{T}} (\widehat{\boldsymbol{\Phi}}_k - \boldsymbol{\Phi}_k) \mathbf{W} \boldsymbol{\xi}_i \\
&= \frac{1}{n} \sum_{i=1}^{n} \frac{1}{m_i^2} \sum_{j_1=1}^{m_i} \sum_{j_2=1}^{m_i} \sum_{l=1}^{\kappa_0} \dot{\mu}_i(T_{ij_1}) d_{ik}(T_{ij_1}, T_{ij_2}) \phi_l(T_{ij_2}) \xi_{il} \\
&= \frac{1}{n} \sum_{i=1}^{n} \frac{1}{m_i^2} \sum_{j_1=1}^{m_i} \sum_{j_2=1}^{m_i} \sum_{l=1}^{\kappa_0} \dot{\mu}_i(T_{ij_1}) \left\{ I^{n1}_{ik}(T_{ij_1}, T_{ij_2}) + I^{n2}_{ik}(T_{ij_1}, T_{ij_2}) \right\} \phi_l(T_{ij_2}) \xi_{il} + O(\|\Delta\|^2) \\
&:= J^{n2}_{k1} + J^{n2}_{k2} + O(\|\Delta\|^2) \qquad \text{almost surely}
\end{aligned} \tag{2.48}$$

Under assumptions (C1), (C2), (C3), (C4) and (C5), by using Theorem 3.3 in Li and Hsing (2010), we have the following.

$$\|\Delta\|^2 = O(h^4 + \delta^2_{n2}(h)) \qquad \text{almost surely} \tag{2.49}$$

Now observe that

$$
\begin{aligned}
J_{k1}^{n2} &= \frac{1}{n}\sum_{i=1}^{n}\frac{1}{m_i^2}\sum_{j_1=1}^{m_i}\sum_{j_2=1}^{m_i}\sum_{l=1}^{\kappa_0}\dot{\mu}_i(T_{ij_1})I_{ik}^{n1}(T_{ij_1})\phi_l(T_{ij_2})\xi_{il} \\
&= \frac{1}{n}\sum_{i=1}^{n}\frac{1}{m_i^2}\sum_{j_1=1}^{m_i}\sum_{j_2=1}^{m_i}\sum_{l=1}^{\kappa_0}\sum_{\substack{r=1\\r\neq k}}^{\infty}(\lambda_k-\lambda_r)^{-1}\dot{\mu}_i(T_{ij_1})\Big\{\phi_r(T_{ij_1})\phi_k(T_{ij_2})+\phi_k(T_{ij_1})\phi_r(T_{ij_2})\Big\}\phi_l(T_{ij_2}) \\
&\qquad \times\langle\phi_r,\Delta\phi_k\rangle\,\xi_{il} \\
&\lesssim \frac{1}{n}\sum_{i=1}^{n}\frac{1}{m_i}\sum_{j_1=1}^{m_i}\sum_{\substack{r=1\\r\neq k}}^{\infty}\sum_{l=1}^{\kappa_0}(\lambda_k-\lambda_r)^{-1}\dot{\mu}_i(T_{ij_1})\phi_r(T_{ij_1})\left\{\mathbf{1}(l=k)+O((\log m/m)^{1/2})\right\}\langle\phi_r,\Delta\phi_k\rangle\,\xi_{il} \\
&+ \frac{1}{n}\sum_{i=1}^{n}\frac{1}{m_i}\sum_{j_1=1}^{m_i}\sum_{\substack{r=1\\r\neq k}}^{\infty}\sum_{l=1}^{\kappa_0}(\lambda_k-\lambda_r)^{-1}\dot{\mu}_i(T_{ij_1})\phi_k(T_{ij_1})\left\{\mathbf{1}(r=l)+O((\log m/m)^{1/2})\right\}\langle\phi_r,\Delta\phi_k\rangle\,\xi_{il} \\
&\lesssim \frac{1}{n}\sum_{i=1}^{n}\frac{1}{m_i}\sum_{j_1=1}^{m_i}\sum_{\substack{r=1\\r\neq k}}^{\infty}(\lambda_k-\lambda_r)^{-1}\dot{\mu}_i(T_{ij_1})\phi_r(T_{ij_1})\langle\phi_r,\Delta\phi_k\rangle\,\xi_{ik}\left\{1+O((\log m/m)^{1/2})\right\} \\
&\qquad + \frac{1}{n}\sum_{i=1}^{n}\frac{1}{m_i}\sum_{j_1=1}^{m_i}\sum_{\substack{r=1\\r\neq k}}^{\kappa_0}(\lambda_k-\lambda_r)^{-1}\dot{\mu}_i(T_{ij_1})\phi_k(T_{ij_1})\langle\phi_r,\Delta\phi_k\rangle\,\xi_{ir}\left\{1+O((\log m/m)^{1/2})\right\} \\
&:= (U_{k1}^n+U_{k2}^n)\left\{1+O((\log m/m)^{1/2})\right\} \qquad \text{almost surely}
\end{aligned}
\tag{2.50}
$$

Then applying the triangle inequality, we have the following.

$$
\begin{aligned}
U_{k1}^n &= \frac{1}{n}\sum_{i=1}^{n}\frac{1}{m_i}\sum_{j=1}^{m_i}\sum_{\substack{r=1\\r\neq k}}^{\infty}(\lambda_k-\lambda_r)^{-1}\dot{\mu}_i(T_{ij})\phi_r(T_{ij})\langle\Delta\phi_k,\phi_r\rangle\,\xi_{ik} \\
&\lesssim \|\Delta\phi_k\|\sum_{\substack{r=1\\r\neq k}}^{\infty}(\lambda_k-\lambda_r)^{-1}\frac{1}{n}\sum_{i=1}^{n}\left\{\mathscr{M}_{ir}+O((\log m/m)^{1/2})\right\}\xi_{ik} \\
&= \|\Delta\phi_k\|\sum_{\substack{r=1\\r\neq k}}^{\infty}(\lambda_k-\lambda_r)^{-1}\frac{1}{n}\sum_{i=1}^{n}\mathscr{M}_{ir}\xi_{ik}\left\{1+O((\log m/m)^{1/2})\right\}
\end{aligned}
\tag{2.51}
$$

where $\mathscr{M}_{ir}=\int\dot{\mu}_i(t)\phi_r(t)dt$. By Lemma 6 of Li and Hsing (2010), under conditions (C1), (C2), (C3), (C4), (C5), for any measurable bounded function e on $[0,1]$, one can show the following.

$$
\|\Delta\phi_k\| = O(h^2+\delta_{n1}(h)+\delta_{n2}^2(h)) \equiv O(\overline{\delta}_n(h)) \qquad \text{almost surely}
\tag{2.52}
$$

where $\overline{\delta}_n(h) = h^2 + \delta_{n1}(h) + \delta^2_{n2}(h)$. Thus, combining the Inequalities (2.37) in Lemma 2.7.5 and Equation (2.52), we obtain $U^n_{k1} = O\left(\overline{\delta}_n(h)(\log n/n)^{1/2}\lambda_k^{1/2}k^{(1-\alpha)/2}\{1 + (\log m/m)^{1/2}\}\right)$ almost surely. Next, under the spacing condition mentioned earlier and in assumption (C6)a, using the Inequality (2.40) recall $\eta_k^{-1} \lesssim \lambda_k^{-1}k$. Thus, observe that

$$\begin{aligned}
U^n_{k2} &= \frac{1}{n}\sum_{i=1}^{n}\frac{1}{m_i}\sum_{j=1}^{m_i}\sum_{\substack{r=1\\ r\neq k}}^{\kappa_0}(\lambda_k - \lambda_r)^{-1}\dot{\mu}_i(T_{ij})\phi_k(T_{i_j})\left\langle \Delta\phi_k, \phi_r\right\rangle \xi_{ir} \\
&\lesssim \|\Delta\phi_k\|\sum_{\substack{r=1\\ r\neq k}}^{\kappa_0}(\lambda_k - \lambda_r)^{-1}\left\{\frac{1}{n}\sum_{i=1}^{n}\mathscr{M}_{ik}\xi_{ir}\right\}\left\{1 + O((\log m/m)^{1/2})\right\} \\
&\lesssim \|\Delta\phi_k\|\eta_k^{-1}\sum_{\substack{r=1\\ r\neq k}}^{\kappa_0}\left\{\frac{1}{n}\sum_{i=1}^{n}\mathscr{M}_{ik}\xi_{ir}\right\}\left\{1 + O((\log m/m)^{1/2})\right\} \qquad (2.53)
\end{aligned}$$

Using Condition (C6)b we also have $V_k^{1/2}\eta_k^{-1} \lesssim V_k^{1/2}\lambda_k^{-1}k = O(k^{(1-\alpha)/2})$. Finally, combining with the bounds for U^n_{k1}, U^n_{k2}, we have, almost surely,

$$\begin{aligned}
J^{n2}_{k1} &= O\left((\log n/n)^{1/2}\overline{\delta}_n(h)\left\{\lambda_k^{1/2}k^{(1-\alpha)/2} + \eta_k^{-1}V_k^{1/2}\sum_{\substack{r=1\\ r\neq k}}^{\kappa_0}\lambda_r^{1/2}\right\}\left(1 + (\log m/m)^{1/2}\right)\right) \\
&= O\left((\log n/n)^{1/2}\overline{\delta}_n(h)k^{(1-\alpha)/2}\sum_{k=1}^{\kappa_0}\lambda_r^{1/2}\left\{1 + (\log m/m)^{1/2}\right\}\right) \\
&:= O(\omega_{k1}(n,h)) \qquad (2.54)
\end{aligned}$$

where $\omega_{k1}(n,h) = (\log n/n)^{1/2}\overline{\delta}_n(h)k^{(1-\alpha)/2}\sum_{k=1}^{\kappa_0}\lambda_r^{1/2}\left\{1 + (\log m/m)^{1/2}\right\}$. It is easy to see that $\sum_{r=1}^{\kappa_0}\lambda_r^{1/2} \sim \kappa_0^{-\frac{\tau_1}{2}+1}$. Therefore, $\omega_{k1}(n,h) \sim (\log n/n)^{1/2}\overline{\delta}_n(h)\kappa_0^{(3-\tau)/2}\left\{1 + (\log m/m)^{1/2}\right\}$ where $\tau = \alpha + \tau_1$.

Similarly to the derivation of the bound for J^{n2}_{k1}, we can write

$$\begin{aligned}
J^{n2}_{k2} &= \frac{1}{n}\sum_{i=1}^{n}\frac{1}{m_i^2}\sum_{j_1=1}^{m_i}\sum_{j_2=1}^{m_i}\sum_{l=1}^{\kappa_0}\dot{\mu}_i(T_{ij_1})I^{n2}_{ik}(T_{ij_1}, T_{ij_2})\phi_l(T_{ij_2})\xi_{il} \\
&= \frac{1}{n}\sum_{i=1}^{n}\frac{1}{m_i^2}\sum_{j_1=1}^{m_i}\sum_{j_2=1}^{m_i}\sum_{l=1}^{\kappa_0}\sum_{\substack{r_1=1\\ r_1\neq k}}^{\infty}\sum_{\substack{r_2=1\\ r_2\neq k}}^{\infty}(\lambda_k - \lambda_{r_1})^{-1}(\lambda_k - \lambda_{r_2})^{-1}\left\langle \phi_{r_1}, \Delta\phi_k\right\rangle\left\langle \phi_{r_2}, \Delta\phi_k\right\rangle \\
&\quad \times \dot{\mu}_i(T_{ij_1})\phi_{r_1}(T_{ij_1})\phi_{r_2}(T_{ij_2})\phi_l(T_{ij_2})\xi_{il}
\end{aligned}$$

$$
\begin{aligned}
&\lesssim \frac{1}{n}\sum_{i=1}^{n}\frac{1}{m_i}\sum_{j_1=1}^{m_i}\sum_{l=1}^{\kappa_0}\sum_{\substack{r_1=1\\ r_1\neq k}}^{\infty}\sum_{\substack{r_2=1\\ r_2\neq k}}^{\infty}(\lambda_k-\lambda_{r_1})^{-1}(\lambda_k-\lambda_{r_2})^{-1}\left\langle\phi_{r_1},\Delta\phi_k\right\rangle\left\langle\phi_{r_2},\Delta\phi_k\right\rangle \\
&\qquad\times\dot{\mu}_i(T_{ij_1})\phi_{r_1}(T_{ij_1})\left\{\mathbf{1}(r_2=l)+O((\log m/m)^{1/2})\right\}\xi_{il} \\
&= \frac{1}{n}\sum_{i=1}^{n}\frac{1}{m_i}\sum_{j_1=1}^{m_i}\sum_{\substack{r_1=1\\ r_1\neq k}}^{\infty}\sum_{\substack{r_2=1\\ r_2\neq k}}^{\kappa_0}(\lambda_k-\lambda_{r_1})^{-1}(\lambda_k-\lambda_{r_2})^{-1}\left\langle\phi_{r_1},\Delta\phi_k\right\rangle\left\langle\phi_{r_2},\Delta\phi_k\right\rangle \\
&\qquad\times\dot{\mu}_i(T_{ij_1})\phi_{r_1}(T_{ij_1})\xi_{ir_2}\left\{1+O((\log m/m)^{1/2})\right\} \\
&\lesssim \|\Delta\phi_k\|^2\sum_{\substack{r_1=1\\ r_1\neq k}}^{\infty}\sum_{\substack{r_2=1\\ r_2\neq k}}^{\kappa_0}(\lambda_k-\lambda_{r_1})^{-1}(\lambda_k-\lambda_{r_2})^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}\left\{\mathscr{M}_{ir_1}+O((\log m/m)^{1/2})\right\}\xi_{ir_2}\right) \\
&\qquad\times\left\{1+O((\log m/m)^{1/2})\right\} \\
&= \|\Delta\phi_k\|^2\sum_{\substack{r_1=1\\ r_1\neq k}}^{\infty}\sum_{\substack{r_2=1\\ r_2\neq k}}^{\kappa_0}(\lambda_k-\lambda_{r_1})^{-1}(\lambda_k-\lambda_{r_2})^{-1}\times\left\{\frac{1}{n}\sum_{i=1}^{n}\mathscr{M}_{ir_1}\xi_{ir_2}\right\}\left\{1+O((\log m/m)^{1/2})\right\}
\end{aligned}
\tag{2.55}
$$

Therefore, Inequality (2.55) immediately follows using Lemma 2.7.6,

$$
J_{k2}^{n2}=O\left((\log n/n)^{1/2}\overline{\delta}_n^2(h)\kappa_0^{(3-\alpha)/2}\lambda_{\kappa_0}^{-1}\{\sum_{r=1}^{\kappa_0}\lambda_r\}^{1/2}\left\{1+(\log m/m)^{1/2}\right\}\right):=O(\omega_{k2}(n,h)) \tag{2.56}
$$

almost surely where $\omega_{k2}(n,h)=(\log n/n)^{1/2}\overline{\delta}_n^2(h)\kappa_0^{(3-\alpha)/2}\lambda_{\kappa_0}^{-1}\sum_{r=1}^{\kappa_0}\lambda_r\left\{1+(\log m/m)^{1/2}\right\}$. It is easy to see $\sum_{r=1}^{\kappa_0}\lambda_r\sim\kappa_0^{-\tau_1+1}$ under assumption (C6)b. Thus, $\omega_{k2}(n,h)\sim(\log n/n)\overline{\delta}^2(h)\kappa_0^{(4-\tau)/2}\{1+(\log m/m)^{1/2}\}$.

Since $\omega_{n2}=O(\omega_{n1})$ and $\overline{\delta}_n^2(h)=O(\omega_{n1})$, in summary, for each $k=1,\cdots,\kappa_0$, the following holds almost surely.

$$
\overline{\mathbf{g}}_k(\boldsymbol{\beta})=O\left((\log n/n)^{1/2}\left\{1+(\log m/m)^{1/2}\right\}+\omega_{k1}(n,h)\right) \tag{2.57}
$$

Therefore, AMSE of $\overline{\mathbf{g}}_k(\boldsymbol{\beta})$ is the following.

$$
\begin{aligned}
&\mathrm{AMSE}\{\overline{\mathbf{g}}_k(\boldsymbol{\beta}_0)\} \\
&=O\left(\frac{1}{n}\left(1+\frac{1}{m}\right)\left[1+\kappa_0^{3-\tau}\left\{h^4+\frac{1}{n}+\frac{1}{nmh}+\frac{1}{n^2m^2h^2}+\frac{1}{n^2m^4h^4}+\frac{1}{n^2mh}+\frac{1}{n^2m^3h^3}\right\}\right]\right)
\end{aligned}
$$

$$= O\left(\frac{1}{n}\left(1+\frac{1}{m}\right)\left(1+\kappa_0^{3-\tau}R_n(h)\right)\right)$$
$$= O\left(n^{-1}+n^{-1}\kappa_0^{3-\tau}R_n(h)\right) \qquad \text{since } (1/nm) = O(1/n) \tag{2.58}$$

where $R_n(h) = \left\{h^4 + \frac{1}{n} + \frac{1}{nmh} + \frac{1}{n^2m^2h^2} + \frac{1}{n^2m^4h^4} + \frac{1}{n^2mh} + \frac{1}{n^2m^3h^3}\right\}$. Combining the above conditions, we find that if $a > 1/4$, $\kappa_0 = O(n^{1/(3-\tau)})$ and $n^{-1/4} \lesssim h \lesssim n^{-(a+1)/5}$ then $\text{AMSE}\{\overline{\mathbf{g}_k}(\boldsymbol{\beta}_0)\} = O(1/n)$. On the other hand, if $a \leq 1/4$, $\kappa_0 = O(n^{4(1+a)/5(3-\tau)})$ and $h \lesssim n^{-1/4}$ $\text{AMSE}\{\overline{\mathbf{g}_k}(\boldsymbol{\beta}_0)\} = O(1/n)$.

Note that for a three-dimensional array $(\partial C/\partial\beta_1, \cdots, \partial C/\partial\beta_p)$ such that the following is a $p \times 1$ vector.

$$\overline{\mathbf{g}}(\boldsymbol{\beta}_0)^{\mathrm{T}}\mathbf{C}^{-1}(\boldsymbol{\beta}_0)\dot{\mathbf{C}}(\boldsymbol{\beta}_0)\mathbf{C}^{-1}(\boldsymbol{\beta}_0)\overline{\mathbf{g}}(\boldsymbol{\beta}_0) \tag{2.59}$$

Therefore,

$$n^{-1}\dot{\mathcal{Q}}(\boldsymbol{\beta}_0) = 2\dot{\overline{\mathbf{g}}}(\boldsymbol{\beta}_0)^{\mathrm{T}}\mathbf{C}^{-1}(\boldsymbol{\beta}_0)\overline{\mathbf{g}}(\boldsymbol{\beta}_0) - \overline{\mathbf{g}}(\boldsymbol{\beta}_0)^{\mathrm{T}}\mathbf{C}^{-1}(\boldsymbol{\beta}_0)\dot{\mathbf{C}}(\boldsymbol{\beta}_0)\mathbf{C}^{-1}(\boldsymbol{\beta}_0)\overline{\mathbf{g}}(\boldsymbol{\beta}_0) \tag{2.60}$$

And

$$n^{-1}\ddot{\mathcal{Q}}(\boldsymbol{\beta}_0) = 2\dot{\overline{\mathbf{g}}}(\boldsymbol{\beta}_0)^{\mathrm{T}}\mathbf{C}^{-1}(\boldsymbol{\beta}_0)\dot{\overline{\mathbf{g}}}(\boldsymbol{\beta}_0) + r_{n1} + r_{n2} + r_{n3} + r_{n4} \tag{2.61}$$

where

$$\begin{aligned}
r_{n1} &= 2\ddot{\overline{\mathbf{g}}}(\boldsymbol{\beta}_0)\mathbf{C}^{-1}(\boldsymbol{\beta}_0)\overline{\mathbf{g}}(\boldsymbol{\beta}_0)\\
r_{n2} &= -4\dot{\overline{\mathbf{g}}}(\boldsymbol{\beta}_0)^{\mathrm{T}}\mathbf{C}^{-1}(\boldsymbol{\beta}_0)\dot{\mathbf{C}}(\boldsymbol{\beta}_0)\mathbf{C}^{-1}(\boldsymbol{\beta}_0)\overline{\mathbf{g}}(\boldsymbol{\beta}_0)\\
r_{n3} &= 2\dot{\overline{\mathbf{g}}}(\boldsymbol{\beta}_0)\mathbf{C}^{-1}(\boldsymbol{\beta}_0)\mathbf{C}^{-1}(\boldsymbol{\beta}_0)\dot{\mathbf{C}}(\boldsymbol{\beta}_0)\mathbf{C}^{-1}(\boldsymbol{\beta}_0)\overline{\mathbf{g}}(\boldsymbol{\beta}_0)\\
r_{n4} &= -\overline{\mathbf{g}}(\boldsymbol{\beta}_0)^{\mathrm{T}}\mathbf{C}^{-1}(\boldsymbol{\beta}_0)\ddot{\mathbf{C}}(\boldsymbol{\beta}_0)\mathbf{C}^{-1}(\boldsymbol{\beta}_0)\overline{\mathbf{g}}(\boldsymbol{\beta}_0)
\end{aligned} \tag{2.62}$$

Since $\overline{\mathbf{g}}(\boldsymbol{\beta}_0) = O_P(n^{-1/2})$ and the weight matrix converges almost surely to an invertible matrix, $\overline{\mathbf{g}}(\boldsymbol{\beta}_0)^{\mathrm{T}}\mathbf{C}^{-1}(\boldsymbol{\beta}_0)\dot{\mathbf{C}}(\boldsymbol{\beta}_0)\mathbf{C}^{-1}(\boldsymbol{\beta}_0)\overline{\mathbf{g}}(\boldsymbol{\beta}_0) = o(n^{-1})$ almost surely. Furthermore, $r_{n1} = O(n^{-1/2})$, $r_{n2} = o(n^{-1/2})$, $r_{n3} = O(n^{-1/2})$, and $r_{n4} = O(n^{-1})$ almost surely. Combining these bounds, we have $r_n = o(1)$ almost surely. Therefore, $\|n^{-1}\dot{\mathcal{Q}} - 2\dot{\overline{\mathbf{g}}}(\boldsymbol{\beta}_0)^{\mathrm{T}}\mathbf{C}^{-1}(\boldsymbol{\beta}_0)\overline{\mathbf{g}}(\boldsymbol{\beta}_0)\| = o_P(n^{-1})$ and $\|n^{-1}\ddot{\mathcal{Q}} - 2\dot{\overline{\mathbf{g}}}(\boldsymbol{\beta}_0)^{\mathrm{T}}\mathbf{C}^{-1}(\boldsymbol{\beta}_0)\dot{\overline{\mathbf{g}}}(\boldsymbol{\beta}_0)\| = o_P(1)$.

The following lines are based on common steps in the GEE literature that includes McCullagh and Nelder (1989); Balan et al. (2005); Tian et al. (2014) among many others. Let $\boldsymbol{\beta}_n = \boldsymbol{\beta}_0 + \delta\mathbf{d}$

where set $\delta = n^{-1/2}$. We have to show that for any $\epsilon > 0$ there exists a large constant c such that

$$P\{\inf_{\|d\|=c} \mathcal{Q}(\boldsymbol{\beta}_n) \geq \mathcal{Q}(\boldsymbol{\beta}_0)\} > 1 - \epsilon \tag{2.63}$$

Note that the above statement is always true if $\epsilon \geq 1$. Thus, we assume that $\epsilon \in (0, 1)$. Due to Taylor series expansion,

$$\mathcal{Q}(\boldsymbol{\beta}_n) = \mathcal{Q}(\boldsymbol{\beta}_0 + \delta\mathbf{d}) = \mathcal{Q}(\boldsymbol{\beta}_0) + \delta\mathbf{d}^{\mathrm{T}}\dot{\mathcal{Q}}(\boldsymbol{\beta}_0) + \frac{1}{2}\delta\mathbf{d}^{\mathrm{T}}\ddot{\mathcal{Q}}(\boldsymbol{\beta}_0)\mathbf{d} + \|\mathbf{d}\|^2 o_P(1) \tag{2.64}$$

Now, observe that, using Equation (2.60),

$$\delta\mathbf{d}^{\mathrm{T}}\dot{\mathcal{Q}}(\boldsymbol{\beta}_0) = \|\mathbf{d}\|O_P(\sqrt{n}\delta) + \|\mathbf{d}\|O_P(\delta) \tag{2.65}$$

and

$$\frac{1}{2}\delta^2\mathbf{d}^{\mathrm{T}}\ddot{\mathcal{Q}}(\boldsymbol{\beta}^*)\mathbf{d} = n\delta^2\mathbf{d}^{\mathrm{T}}\dot{\bar{\mathbf{g}}}(\boldsymbol{\beta}_0)\mathbf{C}^{-1}(\boldsymbol{\beta}_0)\dot{\mathbf{g}}(\boldsymbol{\beta}_0)\mathbf{d} + n\delta^2\|\mathbf{d}\|^2 o_P(1) \tag{2.66}$$

Therefore, for given $\epsilon > 0$, there exists a large enough c such that the above equation (2.63) holds. This implies that there exists a $\hat{\boldsymbol{\beta}}$ that satisfies $\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\| = O_P(\delta)$. Thus, for large n, with probability 1, $\mathcal{Q}(\boldsymbol{\beta})$ attains the minimal value at $\hat{\boldsymbol{\beta}}$ and therefore, $\dot{\mathcal{Q}} = 0$.

2.7.4 Proof of Theorem 2.3.2

Recall, $\mathscr{C}_i = \sum_{k_1=1}^{\kappa_0}\sum_{k_2=1}^{\kappa_0}\boldsymbol{\Phi}_{k_1}\mathbf{X}_i\mathbf{C}^{-1}_{k_1,k_2}\mathbf{X}_i^{\mathrm{T}}\boldsymbol{\Phi}_{k_2}$, where $\mathbf{C}^{-1}_{k_1,k_2}$ is the (k_1, k_2) block of $\mathbf{C}_0^{-1}$. Similarly, we can define $\widehat{\mathscr{C}}_i = \sum_{k_1=1}^{\kappa_0}\sum_{k_2=1}^{\kappa_0}\widehat{\boldsymbol{\Phi}}_{k_1}\mathbf{X}_i\mathbf{C}^{-1}_{k_1,k_2}\mathbf{X}_i^{\mathrm{T}}\widehat{\boldsymbol{\Phi}}_{k_2}$. It is easy to observe that $\widehat{\mathscr{C}}_i = \mathscr{C}_i + \sum_{k_1=1}^{\kappa_0}\sum_{k_2=1}^{\kappa_0}(\widehat{\boldsymbol{\Phi}}_{k_1} - \boldsymbol{\Phi}_{k_1})\mathbf{X}_i\mathbf{C}^{-1}_{k_1,k_2}\mathbf{X}_i^{\mathrm{T}}(\widehat{\boldsymbol{\Phi}}_{k_2} - \boldsymbol{\Phi}_{k_2}) + 2\sum_{k_1=1}^{\kappa_0}\sum_{k_2=1}^{\kappa_0}\boldsymbol{\Phi}_{k_1}\mathbf{X}_i\mathbf{C}^{-1}_{k_1,k_2}\mathbf{X}_i^{\mathrm{T}}(\widehat{\boldsymbol{\Phi}}_{k_2} - \boldsymbol{\Phi}_{k_2})$. Therefore, $\frac{1}{n}\sum_{i=1}^n \dot{\boldsymbol{\mu}}_i^{\mathrm{T}}(\widehat{\mathscr{C}}_i - \mathscr{C})\mathbf{X}_i = \frac{1}{n}\sum_{i=1}^n\sum_{k_1=1}^{\kappa_0}\sum_{k_2=1}^{\kappa_0}\mathbf{P}_{ik_1}\mathbf{C}^{-1}_{k_1,k_2}\mathbf{P}_{ik_2}$ and $\frac{1}{n}\sum_{i=1}^n \dot{\boldsymbol{\mu}}_i^{\mathrm{T}}(\widehat{\mathscr{C}}_i - \mathscr{C})\mathbf{y}_i = \frac{1}{n}\sum_{i=1}^n\sum_{k_1=1}^{\kappa_0}\sum_{k_2=1}^{\kappa_0}\mathbf{P}_{ik_1}\mathbf{C}^{-1}_{k_1,k_2}\mathbf{Q}_{ik_2}$ where $\mathbf{P}_{i,k} = \dot{\boldsymbol{\mu}}_i^{\mathrm{T}}\mathbf{D}_{ik}\mathbf{X}_i$ and $\mathbf{Q}_{ik} = \dot{\boldsymbol{\mu}}_i^{\mathrm{T}}\mathbf{D}_{ik}\mathbf{y}_i$ with $\mathbf{D}_{ik}$ be the difference matrix with (j_1, j_2)-th element is $d_i(T_{ij_1}, T_{ij_2})$. Thus, note that, almsot surely we have the following relation,

$$\begin{aligned}
\mathbf{P}_{ik} &= \frac{1}{m_i^2}\sum_{j_1=1}^{m_i}\sum_{j_2=1}^{m_i}\dot{\mu}_i(T_{ij_1})d_i(T_{ij_1}, T_{ij_2})x_i(T_{ij_2}) \\
&= \frac{1}{m_i^2}\sum_{j_1=1}^{m_i}\sum_{j_2=1}^{m_i}\dot{\mu}_i(T_{ij_1})\left\{I_{ik}^{n1}(T_{ij_1}, T_{ij_2}) + I_{ik}^{n2}(T_{ij_1}, T_{ij_2}) + O(\|\Delta\|^2)\right\}x_i(T_{ij_2})
\end{aligned}$$

$$\lesssim \frac{1}{m_i^2}\sum_{j_1=1}^{m_i}\sum_{j_2=1}^{m_i}\dot{\mu}_i(T_{ij_1})\sum_{\substack{r=1\\ r\neq k}}^{\infty}(\lambda_k-\lambda_r)^{-1}\langle\phi_r,\Delta\phi_k\rangle\left\{\phi_r(T_{ij_1})\phi_k(T_{ij_2})+\phi_k(T_{ij_1})\phi_r(T_{ij_2})\right\}+O(\|\Delta\|^2)$$

$$\lesssim \sum_{\substack{r=1\\ r\neq k}}^{\infty}(\lambda_k-\lambda_r)^{-1}\|\Delta\phi_k\|+O(\|\Delta\|^2)$$

since $\frac{1}{m_i}\sum_{j=1}^{m_i}\dot{\mu}(T_{ij})\phi_k(T_{ij})$ and $\frac{1}{m_i}\sum_{j=1}^{m_i}x_i(T_{ij})\phi_k(T_{ij})$ are finite

$$= O(\varpi) \tag{2.67}$$

where $\varpi=\sum_{\substack{r=1\\ r\neq k}}^{\infty}(\lambda_k-\lambda_r)^{-1}\overline{\delta}_n(h)+h^2+\delta_{n2}^2(h)$. A similar result can be obtained for $\mathbf{Q}_{ik}$. Combining such results, in summary, we have $-\frac{2}{n}\sum_{i=1}^n\mathbf{X}_i^{\mathrm{T}}\widehat{\mathscr{C}}_i(\mathbf{y}_i-\mathbf{X}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}})=-\frac{2}{n}\sum_{i=1}^n\mathbf{X}_i^{\mathrm{T}}\mathscr{C}_i(\mathbf{y}_i-\mathbf{X}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}})+O(\varpi_n)$. Since, for $n\to\infty$, $\mathcal{Q}(\boldsymbol{\beta})$ attains a minimal value at $\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}$, we therefore have $\dot{\mathcal{Q}}(\widehat{\boldsymbol{\beta}})=0$. Thus,

$$\begin{aligned}\dot{\mathcal{Q}}(\widehat{\boldsymbol{\beta}}) &= -\frac{2}{n}\sum_{i=1}^n\mathbf{X}_i^{\mathrm{T}}\widehat{\mathscr{C}}_i(\mathbf{y}_i-\mathbf{X}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}})\\ &= -\frac{2}{n}\sum_{i=1}^n\mathbf{X}_i^{\mathrm{T}}(\widehat{\mathscr{C}}_i-\mathscr{C}_i)(\mathbf{y}_i-\mathbf{X}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}})-\frac{2}{n}\sum_{i=1}^n\mathbf{X}_i^{\mathrm{T}}\mathscr{C}_i(\mathbf{y}_i-\mathbf{X}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}})=0\end{aligned} \tag{2.68}$$

Therefore, almost surely, we have,

$$\begin{aligned}-\frac{2}{n}\sum_{i=1}^n\mathbf{X}_i^{\mathrm{T}}\mathscr{C}_i(\mathbf{y}_i-\mathbf{X}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}})+O(\varpi_n)&=0\\ -\frac{2}{n}\sum_{i=1}^n\mathbf{X}_i^{\mathrm{T}}\mathscr{C}_i(\mathbf{X}_i^{\mathrm{T}}\boldsymbol{\beta}_0+\mathbf{e}_i-\mathbf{X}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}})+O(\varpi_n)&=0\\ \sqrt{n}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)\left\{\frac{1}{n}\sum_{i=1}^n\mathbf{X}_i^{\mathrm{T}}\mathscr{C}_i\mathbf{X}_i\right\}&=\frac{1}{\sqrt{n}}\sum_{i=1}^n\mathbf{X}_i^{\mathrm{T}}\mathscr{C}_i\mathbf{e}_i\end{aligned} \tag{2.69}$$

Now, using the central limit theorem, we can obtain the following.

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n\mathbf{X}_i^{\mathrm{T}}\mathscr{C}_i\mathbf{e}_i\xrightarrow{d}N(0,\mathbf{A}) \tag{2.70}$$

In addition, by the law of large numbers $\frac{1}{n}\sum_{i=1}^n\mathbf{X}_i^{\mathrm{T}}\mathscr{C}_i\mathbf{X}_i\to\mathbf{B}$ in probability. Therefore, using the Slutsky theorem, we complete the proof of Theorem 2.3.2.

CHAPTER 3

ESTIMATION FOR VARYING-COEFFICIENT MODEL IN FUNCTIONAL DATA ANALYSIS UNDER UNKNOWN HETEROSKEDASTICITY: A GMM-BASED APPROACH

3.1 Introduction

Due to modern advancements of technology, varying-coefficient models in functional data have become popular to analyze data coming from several imaging technologies such as magnetic resonance imaging (MRI), diffusion tensor imaging (DTI) etc. We consider the problem of estimating non-parametric coefficient function $\boldsymbol{\beta}(s)$ which is defined on functional domain (for example, space, time) $\mathcal{S}$ to understand the relationship between the functional response $Y(s)$ and the real-valued covariates denoted by $\mathbf{X} = (X_1, \cdots, X_p)^{\mathrm{T}}$, which takes the following form.

$$Y(s) = \mathbf{X}^{\mathrm{T}}\boldsymbol{\beta}(s) + U(s) \tag{3.1}$$

where $\boldsymbol{\beta}(s) = (\beta_1(s), \cdots, \beta_p(s))^{\mathrm{T}}$ is a p-dimensional vector of unknown smooth functions, and it is assumed that $\boldsymbol{\beta}(\cdot)$ is twice-differentiable with continuous second-order derivatives. The random error $\{U_i(s) : s \in \mathcal{S}\}$ is assumed to be a stochastic process indexed by $s \in \mathcal{S}$ and it characterizes the within curve dependence with mean zero and an unknown covariance function $\Sigma(s, s') = \mathrm{cov}\{U(s), U(s')|\mathbf{X}\}$. The varying-coefficient model (VCM) in Equation (3.1) allows its regression coefficient to vary over some predictors of interest. It was introduced in the literature by Hastie and Tibshirani (1993) and systematically studied in Hoover et al. (1998); Fan et al. (1999); Wu and Chiang (2000); Huang et al. (2002); Fan et al. (2003); Huang et al. (2004); Chiou et al. (2004); Ramsay and Silverman (2005); Zhang and Chen (2007); Cardot and Sarda (2008); Fan and Zhang (2008); Wang et al. (2008); Zhu et al. (2014); Kokoszka and Reimherr (2017).

The notion of density is not well defined for functional responses (in general for any random function) (Delaigle and Hall, 2010), as a result of which it is difficult to take advantage of likelihood-based inference; therefore, we need to rely on the moment conditions. Typically, we assume that

the error term $U(s)$ satisfies the conditional mean-zero assumption, such as $\mathbb{E}\{U(s)|\mathbf{X}\} = 0$. By the iterated law of expectation, it is easy to see that, for a given point $s \in \mathcal{S}$, we can define least-square estimates as solution of the sample version of $\mathbb{E}\{\mathbf{X}[Y(s) - \mathbf{X}^{\mathrm{T}}\boldsymbol{\beta}(s)]\} = 0$. Equivalently, we can obtain these estimates by a minimizer of the sample version of $\mathbb{E}\{[Y(s) - \mathbf{X}^{\mathrm{T}}\boldsymbol{\beta}(s)]^2\}$ which is termed as non-parametric local linear estimates (Fan and Gijbels, 1996). Since the above estimates rely only on the conditional mean-zero assumption, they become inefficient in the presence of heteroskedasticity. Therefore, to analyze such data, there is need for a robust estimation procedure, which does not require distributional assumption and can accommodate heteroskedasticity of unknown form. Therefore, we introduce the functional generalized method of moments (GMM) estimation procedure for such VCM.

In classical statistics, the method of moment (MM) estimator solves the sample moment conditions corresponding to the population moment conditions to obtain solutions for the unknown parameters. For example, if the data Y_i are independently and identically distributed with mean μ, then $\mathbb{E}\{Y_i - \mu\} = 0$, and the MM estimate of μ is simply $\hat{\mu} = \overline{Y}$, the sample mean. Now consider the linear regression model $Y_i = \mathbf{X}_i^{\mathrm{T}}\boldsymbol{\beta} + U_i$ where Y_i is the response, $\mathbf{X}$ and $\boldsymbol{\beta}$ are, respectively, the covariates of dimensions p and the unknown regression coefficient with simple moment restrictions $E\{U_i|\mathbf{X}_i\} = 0$. By applying the law of iterated expectation, we get the unconditional moment condition $\mathbb{E}\{\mathbf{X}_i U_i\} = 0$ for random error U_i. It is easy to see that MM estimates coincide with the ordinary least squares (OLS) estimates for simple linear regression model. For a non-linear regression model with additive error $Y_i = \mathfrak{M}(\mathbf{X}_i, \boldsymbol{\beta}) + U_i$, the moment condition is similar to the classical linear model which is essentially $\mathbb{E}\{g(\mathbf{X}_i)U_i\} = 0$ for any function $g(\cdot)$ of $\mathbf{X}$. An obvious choice is $g(\mathbf{X}) = \frac{\partial \mathfrak{M}(\mathbf{x},\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}$ as first-order conditions in non-linear least squares estimator.

In simple linear regression, let us consider that the covariates are decomposed into p_1 and p_2 components such that $\mathbf{X}_i = (\mathbf{X}_{i1}, \mathbf{X}_{i2})^{\mathrm{T}}$ with $p_1 + p_2 = p$. It is a well-known fact that without any further assumptions, an asymptotically efficient estimator of $\boldsymbol{\beta}$ is the OLS estimator. Assume that $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)^{\mathrm{T}}$, where $\boldsymbol{\beta}_2$ is known and is $\mathbf{0}$. Then, we can rewrite the above model as $Y_i = \mathbf{X}_{i1}^{\mathrm{T}}\boldsymbol{\beta}_1 + U_i$ with a similar restriction $\mathbb{E}\{\mathbf{X}_{i1}U_i\} = 0$. For estimating $\boldsymbol{\beta}_1$ based on the data $\{Y_i, \mathbf{X}_{i1} : i = 1, \cdots, n\}$,

the MM estimates become inefficient since the number of moment conditions (p) is larger than the number of parameters to estimate (p_1). This is called "over-identified" situation whereas when $p = p_1$, it is referred as "just-identified" situation. In general, for over-identified situation, let $\mathbf{g}(Y, \mathbf{X}; \boldsymbol{\beta})$ be a d-dimensional function of $\boldsymbol{\beta} \in \mathbb{R}^p$ where $d \geq p$ such that

$$\mathbb{E}\mathbf{g}(Y_i, \mathbf{X}_i; \boldsymbol{\beta}_0) = 0 \tag{3.2}$$

where $\boldsymbol{\beta}_0$ is the vector of true parameters. The set of restrictions in Equation (3.2) is often called "estimating equations". In a seminal paper, Hansen (1982) proposed an extended version of the MM approach for over-identified model. Let $W_1, \cdots, W_n$ be a set of random variables indexed by p-dimensional parameter vector $\boldsymbol{\beta}$ with moment conditions $\mathbb{E}\{\mathbf{g}(W, \boldsymbol{\beta}_0)\} = 0$, where $\mathbf{g}(W, \boldsymbol{\beta})$ be a d-dimensional vector such that $\mathbf{g}(\cdot; \boldsymbol{\beta})$ be the function of W. The estimator is formed by choosing $\boldsymbol{\beta}$ such that the sample average of $\mathbf{g}_i$ is close to zero. For over-identification problem, it is not possible for all the moment conditions to satisfy. Therefore, GMM approach is used to define an estimator which brings the sample mean of $\mathbf{g}$, (viz., $\overline{\mathbf{g}}(W, \boldsymbol{\beta})$) close to zero. GMM estimates minimize the following objective function.

$$\mathcal{J}(\boldsymbol{\beta}) = \overline{\mathbf{g}}(W; \boldsymbol{\beta})^{\mathrm{T}} \mathbf{W}(\boldsymbol{\beta}) \overline{\mathbf{g}}(W; \boldsymbol{\beta}) \tag{3.3}$$

Note that, the above Equation (3.3) is a well-defined norm as long as the weight matrix $\mathbf{W}(\boldsymbol{\beta})$ is symmetric positive definite with dimension $d \times d$.

When the likelihood is not specified or ill-specified, GMM is an alternative estimation technique to the likelihood principle and has become quite popular in statistics and econometrics in the last few decades due to its intuitive idea and applicability; also its properties are quite well-known. In the original version of the proposed method described in Hansen (1982) a two-step GMM was described by the following algorithm: First, compute $\breve{\boldsymbol{\beta}} \in \arg\min_{\boldsymbol{\beta}} \mathbf{g}(\boldsymbol{\beta})^{\mathrm{T}} \mathbf{g}(\boldsymbol{\beta})$; then estimate the precision matrix $\mathbf{W}$ based on the first-step estimates, $\mathbf{W}(\breve{\boldsymbol{\beta}})$. Therefore, the two-step GMM estimates are $\widehat{\boldsymbol{\beta}} \in \arg\min_{\boldsymbol{\beta}} \overline{\mathbf{g}}(\boldsymbol{\beta})^{\mathrm{T}} \mathbf{W}(\breve{\boldsymbol{\beta}}) \overline{\mathbf{g}}(\boldsymbol{\beta})$. Under some sufficient conditions, it is well known that the GMM estimator is constant and asymptotically normally distributed with the asymptotic covariance matrix $n^{-1} \left\{\mathbf{G}(\boldsymbol{\beta}_0)^{\mathrm{T}} \mathbf{W} \mathbf{G}(\boldsymbol{\beta}_0)\right\}^{-1} \mathbf{G}(\boldsymbol{\beta}_0)^{\mathrm{T}} \mathbf{W} \boldsymbol{\Omega} \mathbf{W} \mathbf{G}(\boldsymbol{\beta}_0) \left\{\mathbf{G}(\boldsymbol{\beta}_0)^{\mathrm{T}} \mathbf{W} \mathbf{G}(\boldsymbol{\beta}_0)\right\}^{-1}$, where $\mathbf{G}(\boldsymbol{\beta}_0) =$

$\mathbb{E}\{\nabla_{\boldsymbol{\beta}_0}\mathbf{g}(W;\boldsymbol{\beta}_0)\}$ and $\boldsymbol{\Omega} = \mathbb{E}\{\mathbf{g}(W;\boldsymbol{\beta}_0)\mathbf{g}(W;\boldsymbol{\beta}_0)^{\mathrm{T}}\}$ for true $\boldsymbol{\beta}_0$ (Newey and McFadden, 1994). Note that the variance of the GMM estimator depends on the weight matrix and is optimal if the weight matrix is $\mathbf{W} = \boldsymbol{\Omega}^{-1}$. Therefore the asymptotic variance would be $n^{-1}\left\{\mathbf{G}(\boldsymbol{\beta}_0)^{\mathrm{T}}\mathbf{W}\mathbf{G}(\boldsymbol{\beta}_0)\right\}^{-1}$.

A crucial problem of the above estimation procedure is identifying the number of moment conditions. An increase in the number of moment conditions may lead to the problem of finite sample bias. Based on the situation, additional moment conditions may be beneficial. Adding more moment conditions leads to decrease (or at least no change) in the asymptotic variance of the estimator, since the optimal weight matrix for fewer moment conditions is not optimal for all moment conditions. In the presence of heteroskedasticity, the set of moment conditions $\mathbf{g}(W;\boldsymbol{\beta}) = (\mathbf{X}U, \mathfrak{M}(\mathbf{X})U)^{\mathrm{T}}$ produces an estimator more efficient than the least squares estimates for the simple linear regression model by efficiently choosing $\mathfrak{M}$. It is of interest to figure out how many functions should be used to get a better estimator. Moreover, some choices of $\mathbf{g}$ are better than others depending on additional assumptions. For example, in a linear regression model, if we choose $\mathbf{g} = \mathbf{X}U$ then the resulting estimator becomes OLS. On the other hand, if we choose $\mathbf{g} = \mathbf{X}U/\mathrm{Var}\{U|\mathbf{X}\}$ we land upon a generalized least squares estimator under heteroskedasticity. $\mathrm{Var}\{U|\mathbf{X}\}$ can be modelled parametrically and substituted in the above mentioned mean-zero function. If the complete form of the likelihood structure is known, one can choose $\mathbf{g} = \nabla_{\boldsymbol{\beta}} \ln\{\mathscr{P}(U|\mathbf{X})\}$ where $\mathscr{P}(\cdot)$ is the density of U. When explanatory variables are endogenous, additional moment conditions may create bias and, as a result, increase the small sample variance. However, this issue does not arise when the explanatory variables are exogenous and the presence of heteroskedasticity does not cause OLS inconsistency problems in classical linear regression. Lu and Wooldridge (2020) obtained an asymptotic efficient estimator using Cragg (1983) which showed existence of GMM estimators which are more efficient than OLS in the presence of heteroskedasticity of unknown form.

It is well known that the constant or varying-coefficient models may often be misspecified and, therefore, this may lead to inconsistent estimation. In varying-coefficient non-parametric models, mostly exogenous regression situation has been considered so far (Hastie and Tibshirani, 1993; Fan et al., 1999). In last two decades, some semi-parametric models were considered under

endogenous variables and a non-parametric or semi-parametric GMM with instrument variables approach was considered. For example, Cai and Li (2008) proposed a one-step local linear GMM estimator that corresponds to the local linear GMM discussed in Su et al. (2013) with an identity weight matrix. Tran and Tsionas (2009) provided a local constant two-step GMM estimator with specified weight matrix by minimizing the asymptotic variance. Su et al. (2013) developed a local linear GMM estimator procedure of functional coefficient instrument variable models with a general weight matrix under exogenous conditions. Cai et al. (2006) proposed a two-step local linear estimation procedure to estimate the functional coefficient which include the estimation of high-dimensional non-parametric model in first step and later estimate the functional coefficients using the first-step non-parametric estimates as generated regressor. As opposed to the classical GMM, for non-parametric local linear GMM estimator, integrated mean square error increases as the number of instrument variables increase for arbitrary choice of instrument variables (Bravo, 2021).

The current work is motivated by the problem encountered in diffusion tensor imaging (DTI) where multiple diffusion properties are measured along common major white matter fiber tracts across multiple individuals to characterize the structure and orientation of white matter in the human brain. Recently a study has been performed to understand white matter structural alteration using DTI for obstructive sleep apnea patients (Xiong et al., 2017). As an illustration, we present smoothed functional data to analyze the efficiency properties of network generated by diffusion properties of the water molecules. We plot the graphical characteristics of one of the diffusion properties called fractional anisotropy (FA) over different significant levels to obtain the graphical connectivity from 29 patients in Figures 3.1. Scientists are often interested to know the individual association of average path length (APL) of the network generated from FA with a set of covariates of interests such as age and lapses score (see Section 2.5). Moreover, in this data, there is sufficient evidence of heteroskedasticity in the covariates. Details about the data-set and associated variables are described in Section 3.6. We therefore need an estimation procedure which (1) does not need knowledge of distribution, (2) can handle heteroskedasticity of covariates, (3) can estimate non-

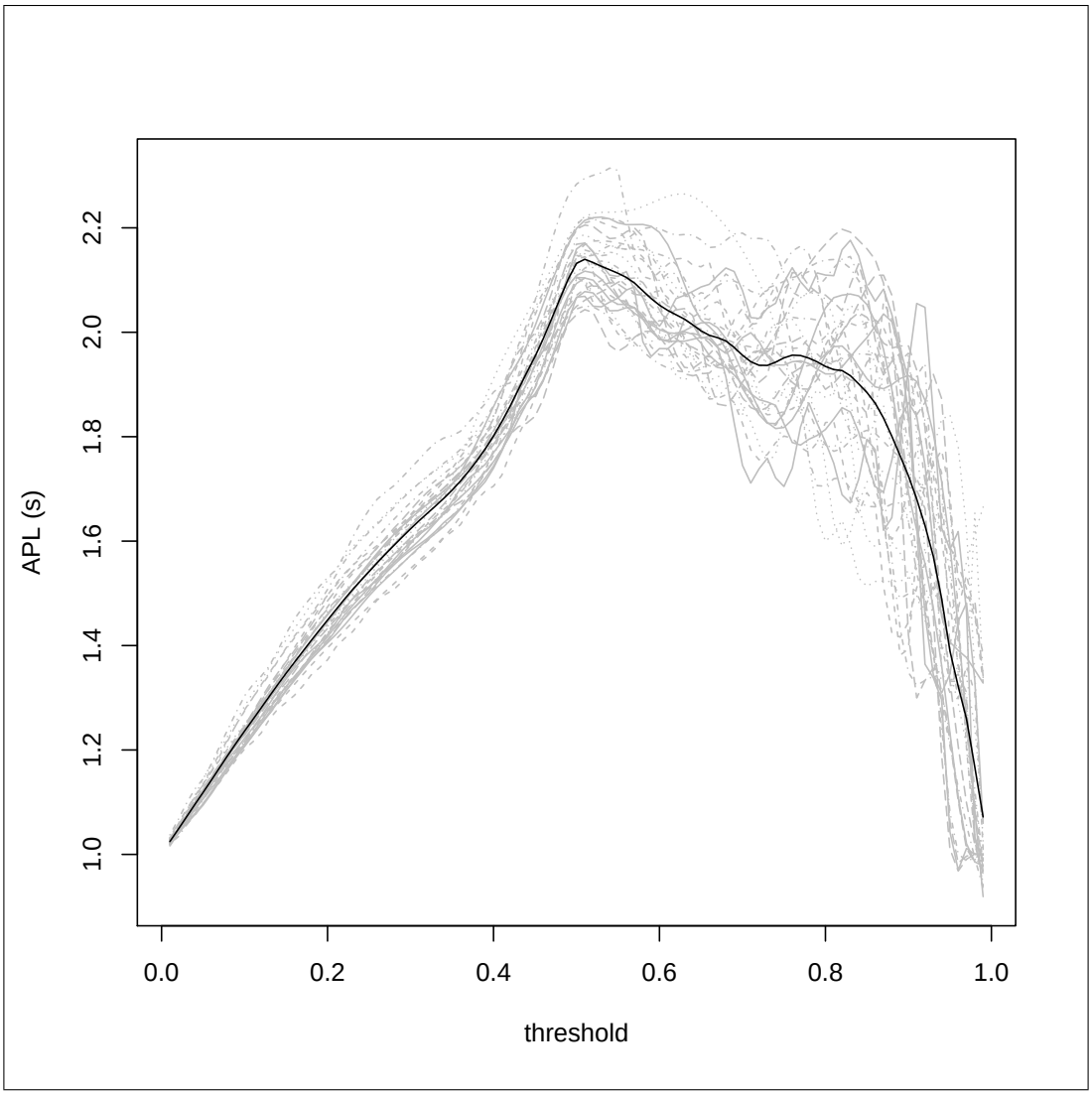

Figure 3.1 *Apnea-data*: Smoothed average path length (APL) from 29 patients over different thresholds. Black solid line indicates the mean of APL over thresholds.

parametric coefficient functions from varying-coefficient models, and (4) has a systematic technique for computing an efficient estimator. In this chapter, we develop a local linear GMM estimation procedure for varying-coefficient model. For given instrument variables, we propose an optimal local-linear GMM estimator motivated by Lu and Wooldridge (2020). However, the key difference in our approach from the later is that we model the variance of integrated squared error using a non-parametric function of covariates whereas they assume a parametric form in case of classical regression. Therefore, we can ensure that the proposed estimator is at least as efficient as local linear estimates (initial estimator) and more efficient than that under the presence of heteroskedasticity.

This chapter is organized as follows. In Section 3.2, we introduce our varying-coefficient model and propose a local linear GMM estimator. In Section 3.3, we present a multi-step estimation procedure. We establish asymptotic results in Section 3.4. We perform a set of simulations studied to understand the finite sample performance of the proposed estimator and present those in Section 3.5. In Section 3.6, we apply the proposed method in a real imaging data-set on obstructive sleep apnea (OSA). In Section 3.7, we conclude this chapter with some discussion. All technical details are provided in Section 3.8.

3.2 Varying-coefficient functional model and moment conditions

In this section, we first introduce heteroskedastic conditions for SVC model and thereafter, propose a heuristic method to construct a mean-zero function.

3.2.1 Model

Let $\{Y_i(s), \mathbf{X}_i\}$ for $i = 1, \cdots, n$ be independent copies of $\{Y(s), \mathbf{X}\}$. Instead of observing the entire functional trajectory, one can observe $Y(s)$ only on the discrete spatial grid $\{s_1, \cdots, s_r\}$ on the functional domain $\mathcal{S}$. Data can be Gaussian or non-Gaussian and homoskedastic or heteroskedastic depending on the real applications. Therefore, the observed data for the i-th individual are $\{s_j, Y_i(s_j), \mathbf{X}_i : j = 1, \cdots, r\}$. For simplifying the notation, define $Y_{ij} = Y_i(s_j)$ and $U_{ij} = U_i(s_j)$.

Considering the functional principal component analysis (FPCA) model for $U_i(s)$, we assume that $U_i(s)$ is square-integrable and admits the Karhunen-Loève expansion (Karhunen, 1946; Loève, 1946). Let $\omega_1(\mathbf{X}) \geq \omega_2(\mathbf{X}) \geq \cdots \geq 0$ be ordered eigen-values of the linear operator determined by Σ with $\sum_{k=1}^{\infty} \omega_k(\mathbf{X})$ being finite and $\psi_k(s)$'s being the corresponding orthonormal eigen-functions or principal components. Thus, the spectral decomposition (J Mercer, 1909) is given by

$$\Sigma(s, s') = \sum_{k=1}^{\infty} \omega_k(\mathbf{X}) \psi_k(s) \psi_k(s') \tag{3.4}$$

Therefore, $U_i(s)$ admits the Karhunen-Loève expansion as follows.

$$U_i(s) = \sum_{k=1}^{\infty} \xi_k(\mathbf{X}_i) \psi_k(s) \tag{3.5}$$

where $\xi_k(\mathbf{X}_i) = \int_{\mathcal{S}} U_i(s)\psi_i(s)ds$, which is termed as the k-th functional principal score for i-th individual. The $\xi_k(\mathbf{X}_i)$ are uncorrelated over k with $\mathbb{E}\{\xi_k(\mathbf{X}_i)|\mathbf{X}_i\} = 0$ and $\text{Var}\{\xi_k(\mathbf{X}_i)|\mathbf{X}_i\} = \omega_k(\mathbf{X}_i)$, $k \geq 1$. Furthermore, assume that the eigen-values vary with $\mathbf{X}_i$ such that $\omega_k(\mathbf{X}_i) = \theta_k\sigma^2(\mathbf{X}_i)$ for some unknown function $\sigma(\mathbf{X}) \geq 0$ and $\theta_1 \geq \theta_2 \geq \cdots \geq 0$. For identifiability, we need some restrictions on θ_ks, such as known or fixed θ_1. Therefore, the above assumption on eigen-values for spectral decomposition allow us to incorporate heteroskedasticity into the model. To best of our knowledge, this is the first attempt to model SVC with unknown heteroskedasticity.

3.2.2 Local-linear mean-zero function

Let us reiterate our main objective: we want to efficiently estimate the varying-coefficient functions based on GMM for the case of continuum moment conditions together with infinite-dimensional parameters. Therefore, we need to construct a mean-zero function which will be described in this sub-section.

Since $\boldsymbol{\beta}(\cdot)$ in model 3.1 is twice continuously differentiable, we can apply the Taylor series expansion to $\boldsymbol{\beta}(s_j)$ around an interior point s_0 and get

$$\boldsymbol{\beta}(s_j) = \boldsymbol{\beta}(s) + \dot{\boldsymbol{\beta}}(s_0)(s_j - s_0) + \ddot{\boldsymbol{\beta}}(s^*)(s_j - s_0)^2/2 \tag{3.6}$$

where s^* lies between s_j and s_0 for all $j = 1, \cdots, r$ and $\dot{\boldsymbol{\beta}}$ and $\ddot{\boldsymbol{\beta}}$ denote the gradients of $\boldsymbol{\beta}$ and $\dot{\boldsymbol{\beta}}$ with respect to s. Thus, $\boldsymbol{\beta}(s_j)$ can be approximated as $\beta_k(s_j) \approx \beta(s_0) + \frac{\partial\beta_k(s_0)}{\partial s}(s_j - s_0)$. So in matrix notation, the first-order Taylor series expansion of the coefficient functions becomes

$$\boldsymbol{\beta}(s_j) \approx \mathbf{A}(s_0)\mathbf{z}_h(s_j - s_0) \tag{3.7}$$

where $\mathbf{z}_h(s_j - s_0) = \left(1, \frac{s_j - s_0}{h}\right)^{\mathrm{T}}$ and $\mathbf{A}(s) = [\boldsymbol{\beta}(s_0), h\dot{\boldsymbol{\beta}}(s_0)]$ which is a $p \times 2$ matrix. Hence, applying the approximation procedure in Equation (3.7), we can rewrite the model 3.1 as

$$\begin{aligned} Y_{ij} &\approx \mathbf{X}_i^{\mathrm{T}}\left\{\mathbf{A}(s)\mathbf{z}_h(s_j - s_0)\right\} + U_{ir} \\ &= \left\{\mathbf{z}_h(s_j - s_0) \otimes \mathbf{X}_i\right\}^{\mathrm{T}} \text{vec}\{\mathbf{A}(s_0)\} + U_{ir} \\ &= \mathbf{W}_{ij}(s_0)^{\mathrm{T}}\boldsymbol{\gamma}(s_0) + U_{ir} \end{aligned} \tag{3.8}$$

such that s_j are sufficiently close to s_0, where $\mathbf{W}_{ij}(s_0) = [\mathbf{z}_h(s_j - s_0) \otimes \mathbf{X}_i]$ and $\boldsymbol{\gamma}(s_0) = (\boldsymbol{\beta}(s_0)^{\mathrm{T}}, h\dot{\boldsymbol{\beta}}(s_0)^{\mathrm{T}})^{\mathrm{T}}$, both of which are vectors of length $2p \times 1$.

Let $K(\cdot)$ be a symmetric probability density function which is used as a kernel and $h > 0$ be the bandwidth; thus a re-scaled kernel function is defined as $K_h(\cdot) = \frac{1}{h}K(\cdot)$. It is easy to see that for a given location $s_0 \in \mathcal{S}$, we can construct a least squares estimator of $\boldsymbol{\gamma}(s)$ defined in Equation (3.8) by minimizing the sample version of mean squared error $\mathbb{E}\{[Y_{ij} - \mathbf{W}_{ij}^{\mathrm{T}}(s_0)\boldsymbol{\gamma}(s_0)]^2|\mathbf{X}_i\}$. Let $\mathfrak{M}(\mathbf{X})$ be a q-dimensional instrument variable with $q \geq p$; the moment condition can be written as $\mathbb{E}\left\{\frac{1}{r}\sum_{j=1}^{r} K_h(s_j - s_0)\boldsymbol{\Delta}_{ij}(s_0)\right\} = 0$ where $\boldsymbol{\Delta}_{ij}(s_0) = \mathfrak{M}(\mathbf{X}_i)\left\{Y_{ij} - \mathbf{W}_{ij}(s_0)^{\mathrm{T}}\boldsymbol{\gamma}(s_0)\right\}$ is a zero-mean stochastic process. Two popular approaches to construct optimal instrument variables are proposed by Newey (1990); Ai and Chen (2003). Due to functional dependence and the existence of heteroskedasticity of an unknown form, these approaches can not be undertaken for the model 3.8 that we have considered. Since Taylor series expansion provides a local approximation of the function, we need to incorporate this phenomenon into the instrument variables during construction of the mean-zero function.

Motivated from the idea of local linear estimator Fan and Gijbels (1996), consider the local linear instrument variables $\mathbf{Q}_{ir}(s_0) = (\mathfrak{M}(\mathbf{X}_i), \mathfrak{M}(\mathbf{X}_i)(s_j - s_0)/h)^{\mathrm{T}}$. Therefore, consider the following non-parametrically localizing augmented orthogonal moment conditions for estimating $\boldsymbol{\beta}(s)$.

$$\begin{aligned}
\mathbf{g}_i\{\boldsymbol{\gamma}(s_0)\} &= \frac{1}{r}\sum_{j=1}^{r} K_h(s_j - s_0)\mathbf{Q}_{ir}(s_0)\left\{Y_{ij} - \mathbf{W}_{ij}(s_0)^{\mathrm{T}}\boldsymbol{\gamma}(s_0)\right\} \\
&= \begin{pmatrix} \frac{1}{r}\sum_{j=1}^{r} K_h(s_j - s_0)\boldsymbol{\Delta}_{ij}(s_0) \\ \frac{1}{r}\sum_{j=1}^{r} K_h(s_j - s_0)\frac{(s_j - s_0)}{h}\boldsymbol{\Delta}_{ij}(s_0) \end{pmatrix} \\
&= \frac{1}{r}\sum_{j=1}^{r} K_h(s_j - s_0)\mathbf{z}_h(s_j - s_0) \otimes \boldsymbol{\Delta}_{ij}(s_0) \qquad (3.9)
\end{aligned}$$

and note that $\{\mathbf{g}_i(\boldsymbol{\gamma}(s))\} : i = 1, \cdots, n\}$ are independent and $\mathbb{E}\{\mathbf{g}_i(\boldsymbol{\gamma}(s))\} = \mathbf{0}_{2q\times 1}$ for $s \in \mathcal{S}$.

Most of the varying-coefficient models that exist in the literature assume homoskedasticity in covariates and are limited to weakly dependent non-parametric models (Su et al., 2013; Sun, 2016), that differs significantly in our model assumptions. In contrast, we assume a varying-coefficient model under heteroskedasticity of an unknown form.

3.3 Multi-step estimation procedure

This section develops a multi-step estimation procedure to estimate $\boldsymbol{\beta}(s)$ simultaneously across all functional points. Essentially, the multi-step procedure can be broken down as, Step-I: an initial estimation; Step-II: estimation of the variance function, mean zero function, and eigen-components and Step-III: GMM estimation. The key ideas of each step are described below.

Step-I. Calculate the least squares estimates of $\boldsymbol{\beta}(s)$ as initial estimates, denoted by $\breve{\boldsymbol{\beta}}(s)$ across all $s \in \mathcal{S}$.

Step-II. Estimate the conditional variance of integrated square residuals non-parametrically, subsequently estimate the covariance of mean-zero function. Estimate the eigen-components using multivariate FPCA.

Step-III. Project the continuous moment conditions onto eigen-functions and then combine them by weighted eigenvalues to incorporate spatial dependence and thus obtain the updated estimate of $\boldsymbol{\beta}(s)$, denoted by $\widehat{\boldsymbol{\beta}}(s)$ across all $s \in \mathcal{S}$.

3.3.1 Step-I: Initial least squares estimates

We consider a local linear smoother (Fan and Gijbels, 1996) to obtain an initial estimator of $\boldsymbol{\beta}(\cdot)$ ignoring functional dependencies. In this case, the non-linear least squares function of the model 3.1 can be defined as an objective function given by $\mathcal{J}_{init}\{\boldsymbol{\beta}(\cdot)\} = \frac{1}{nr}\sum_{i=1}^{n}\sum_{j=1}^{r}\{Y_{ij} - \mathbf{X}_i^{\mathrm{T}}\boldsymbol{\beta}(s_j)\}^2$. By the local linear smoothing method we estimate $\boldsymbol{\gamma}$ at functional point s_0, by minimizing

$$\mathcal{J}_{init}\{\boldsymbol{\gamma}(s_0)\} = \frac{1}{nr}\sum_{i=1}^{n}\sum_{j=1}^{r} K_h(s_j - s_0)\left\{Y_{ij} - \mathbf{W}_{ij}(s_0)^{\mathrm{T}}\boldsymbol{\gamma}(s_0)\right\}^2 \tag{3.10}$$

The solution of the above least-squares problem can be expressed as

$$\breve{\boldsymbol{\gamma}}(s_0) = \left\{\frac{1}{nr}\sum_{i=1}^{n}\sum_{j=1}^{r} K_h(s_j - s_0)\mathbf{W}_{ij}(s_0)\mathbf{W}_{ij}(s_0)^{\mathrm{T}}\right\}^{-1}\left\{\frac{1}{nr}\sum_{i=1}^{n}\sum_{j=1}^{r} K_h(s_j - s_0)\mathbf{W}_{ij}(s_0)Y_{ij}\right\} \tag{3.11}$$

Consequently, the estimator of the coefficient function vector $\boldsymbol{\beta}(s)$ at s_0 is

$$\breve{\boldsymbol{\beta}}(s_0) = [(1,0) \otimes \mathbf{I}_p]\breve{\boldsymbol{\gamma}}(s_0) \tag{3.12}$$

We determine the tuning parameter h by using some data-driven techniques such as cross-validation and generalized cross-validation (Hastie and Tibshirani, 2017).

3.3.2 Step-II: Intermediate steps

Step-II consists of two important steps in determining the class of GMM estimator. First in Step-II.A, we propose a method to obtain optimal instrument variables and therefore estimate the eigen-components which are used in local linear GMM objective function in Step-III. To estimate eigen-components, we essentially need to use a multivariate version of FPCA which is quite uncommon in the literature. We borrow the method proposed by Wang (2008).

Step-II.A: Choice of instrument variables

The choice of instrument variables is critical and the required identification condition is $q \geq p$, which ensures that the dimension of $\mathbf{Q}_{ij}(s_0)$ is at least equal to the dimension of $\boldsymbol{\gamma}(s_0)$. In our model, as discussed in Section 3.2, the error term has a potential heteroskedasticity of unknown form. We define a set of independent and identically distributed random variables $R_1, R_2, \cdots, R_n$ for n individuals, where $R_i = \int U_i^2(s)ds$ for each i, termed as integrated square of residuals, and $\mathbb{E}\{R_i|\mathbf{X}_i\} = \sigma^2(\mathbf{X}_i)\sum_{k=1}^{\infty}\theta_k$. Therefore, consider the following non-parametric regression problem.

$$\log R_i = \log \sigma^2(\mathbf{X}_i) + \epsilon_i \tag{3.13}$$

where ϵ_i is the mean zero random variable with constant variance. The above model in Equation (3.13) boils down to the problem of estimation of $\log \sigma^2(\mathbf{X}_i)$ by regressing the logarithmic value of the integrated squared residual variables on the covariates $\mathbf{X}_i$. This approach is inspired by Yu and Jones (2004); Wasserman (2006) although used in a different context. Since U_is are not observable, we replace U_i by an efficient estimate that is obtained from Step-I, viz., $\breve{U}_i(s) = Y_{ij} - \mathbf{X}_i^{\mathrm{T}}\breve{\boldsymbol{\beta}}(s)$ for

all $s \in \mathcal{S}$. This step can easily be implemented using "*gam*" function available in *mgcv* package in R to get an estimate of the non-parametric mean function, denoted by $\widehat{\mu}(\mathbf{X})$ and therefore $\widehat{\sigma}^2(\mathbf{X}) = \exp\{\widehat{\mu}(\mathbf{X})\}$. Given the estimate of $\sigma(\cdot)$, we can, therefore, choose instrument variables as $\mathfrak{M}(\mathbf{X}_i) = \left(\mathbf{X}_i, \mathbf{X}_i/\widehat{\sigma}^2(\mathbf{X}_i)\right)^{\mathrm{T}}$.

Step-II.B: Estimation of eigen-components

Without loss of generality, assume that the spectrum of functional domain $\mathcal{S} = [0, 1]$ and the dimension of mean-zero function $\mathbf{g}(s) = (g_1(s), \cdots, g_d(s))^{\mathrm{T}}$ is d (in our problem, it equals $2q$) for simplicity. Note that $\mathbf{g}(s)$ is defined on an interval $[0, 1]$ such that $\sum_{l=1}^{d} \int \mathbb{E}\{g_l^2(s)\}ds$ is finite and the covariance function $\mathbf{C}(s, s') = \mathbb{E}\left\{\mathbf{g}\{\boldsymbol{\gamma}(s)\}\mathbf{g}\{\boldsymbol{\gamma}(s)\}^{\mathrm{T}}\right\}$. Under the condition (C6) mentioned in Section 3.4, using the lining-up method in Wang (2008), define a new stochastic process $e(s)$ on the interval $[0, d]$ with eigen-function ϕ_e such that,

$$e(s) = \begin{cases} g_1(s) & 0 \le s < 1 \\ g_2(s-1) & 1 \le s < 2 \\ \cdots & \\ g_l(s-(l-1)) & l-1 \le s < l \\ \cdots & \\ g_d(s-d+1) & d-1 \le s < d \end{cases} \qquad \phi_e(s) = \begin{cases} \phi_1(s) & 0 \le s < 1 \\ \phi_2(s-1) & 1 \le s < 2 \\ \cdots & \\ \phi_l(s-(l-1)) & l-1 \le s < l \\ \cdots & \\ \phi_d(s-d+1) & d-1 \le s < d \end{cases}$$

where we define the eigen-function for each g_l as ϕ_l for $l = 1, \cdots, d$. Therefore, the covariance function between $g_l(s)$ and $g_{l'}(s')$ can be expressed as $C_{l,l'}(s, s') = \mathrm{cov}\{g_l(s-(l-1)), g_{l'}(s'-(l'-1))\}$ for $l - 1 \le s < l$ and $l' - 1 \le s' < l'$; $l, l' = 1, \cdots, d$. Note that for d-dimensional processes, the Fredholm integral equation (Porter et al., 1990) is equivalent to d-simultaneous integral equations where each of them corresponds to a specific functional interval of $e(s)$. For $l - 1 \le s < l$; $l = 1, \cdots, d$, the Fredholm integral equation yields

$$\int_0^d \mathrm{cov}\{e(s), e(s')\}\phi_e(s)ds = \lambda\phi_e(s) \tag{3.14}$$

Now observe that for $(l-1) \leq s < l; l = 1, \cdots, d$, the above relation is equivalent to the following.

$$\sum_{l'=1}^{d} \int_{l'-1}^{l'} \text{cov}\{g_l(s-(l-1)), g_{l'}(s'-(l'-1))\phi_{l'}(s')ds' = \lambda\phi_l(s-(l-1))$$
$$\sum_{l'=1}^{d} \int_{0}^{1} \text{cov}\{g_l(s-(l-1)), g_{l'}(s')\}\phi_{l'}(s')ds' = \lambda\phi_l(s-(l-1))$$
$$\sum_{l'=1}^{d} \int_{0}^{1} \text{cov}\{g_l(s), g_{l'}(s')\}\phi_{l'}(s')ds' = \lambda\phi_l(s) \tag{3.15}$$

In a multivariate setting, the orthogonality condition is

$$\begin{aligned}\int_0^d \phi_{e,l}(s)\phi_{e,l'}(s)ds = \mathbf{1}(l = l') &= \sum_{k=1}^{d} \int_{k-1}^{k} \phi_{k,l}(s-(k-1))\phi_{k,l'}(s-(k-1))ds \\ &= \sum_{k=1}^{d} \int_{0}^{1} \phi_{k,l}(s)\phi_{k,l'}(s)ds \end{aligned} \tag{3.16}$$

Using the generalized Mercer's theorem (J Mercer, 1909), the results for the covariance function can be briefly shown using the lining-up method. Assume that the covariance function is continuous after the lining-up processes, so for $(l_1 - 1) \leq s < l_1$ and $(l_2 - 1) \leq s < l_2$; $l_1, l_2 = 1, \cdots, d$, the covariance function between $g_{l_1}(s)$ and $g_{l_2}(s')$ can be expressed as

$$\begin{aligned} C_{l,l'}(s,s') = \text{cov}\{g_l(s), g_{l'}(s)\} &= \sum_{k=1}^{\infty} \lambda_k \phi_{k,l}(s-(l-1))\phi_{k,l'}(s-(l'-1)) \\ &= \sum_{k=1}^{\infty} \lambda_k \phi_{k,l}(s)\phi_{k,l'}(s') \end{aligned} \tag{3.17}$$

Therefore, using the above argument, we can define the multivariate spectral decomposition

$$\mathbf{C}(s,s') = \sum_{k=1}^{\infty} \lambda_k \boldsymbol{\phi}_k(s)\boldsymbol{\phi}_k(s')^{\mathrm{T}} \tag{3.18}$$

with the orthogonality condition 3.16. Since the lining-up data are univariate, we can adopt the existing techniques of estimating functional eigen-values and eigen-function in the literature (Yao et al., 2003, 2005; Müller and Yao, 2010; Li and Hsing, 2010) to estimate λ and $\phi_e(s)$, and hence can estimate $\boldsymbol{\phi}(s)$ by stacking all components for aliened eigen-functions $\phi_e(s)$.

3.3.3 Step-III: Final estimates

Finally, we demonstrate our proposed estimator based on local-linear GMM where the proposed mean-zero function can be projected onto eigen-function and then combined by the weighted eigen-values. Then, for any positive α, the objective function is given by

$$\begin{aligned}\mathcal{J}\{\boldsymbol{\gamma}(s_0)\} &= \sum_{k=1}^{\infty} \frac{\widehat{\lambda}_k}{\widehat{\lambda}_k^2+\alpha} \left\{\overline{\mathbf{g}}(\boldsymbol{\gamma}(s_0))^{\mathrm{T}}\widehat{\boldsymbol{\phi}}_k(s_0)\right\}^2 \\ &= \sum_{k=1}^{\infty} \frac{\widehat{\lambda}_k}{\widehat{\lambda}_k^2+\alpha} \left\{\frac{1}{nr}\sum_{i=1}^{n}\sum_{j=1}^{r} K_h(s_j-s_0)\widehat{\boldsymbol{\phi}}_k(s_0)^{\mathrm{T}}\mathbf{Q}_{ij}(s_0)\left[Y_{ij}-\mathbf{W}_{ij}(s_0)^{\mathrm{T}}\boldsymbol{\gamma}(s_0)\right]\right\}^2\end{aligned} \tag{3.19}$$

By minimizing the above objective function, we obtain the following.

$$\begin{aligned}&\sum_{k=1}^{\infty} \frac{\widehat{\lambda}_k}{\widehat{\lambda}_k^2+\alpha} \left\{\frac{1}{nr}\sum_{i=1}^{n}\sum_{j=1}^{r} K_h(s_j-s_0)\widehat{\boldsymbol{\phi}}_k(s_0)^{\mathrm{T}}\mathbf{Q}_{ij}(s_0)\mathbf{W}_{ij}(s_0)\right\} \\ &\qquad \times \left\{\frac{1}{nr}\sum_{i=1}^{n}\sum_{j=1}^{r} K_h(s_j-s_0)\widehat{\boldsymbol{\phi}}_k(s_0)^{\mathrm{T}}\mathbf{Q}_{ij}(s_0)\left[Y_{ij}-\mathbf{W}_{ij}(s_0)^{\mathrm{T}}\boldsymbol{\gamma}(s_0)\right]\right\} \\ &:= \sum_{k=1}^{\infty} \frac{\widehat{\lambda}_k}{\widehat{\lambda}_k^2+\alpha} \mathcal{X}_k(s_0)\left\{\mathcal{Y}_k(s_0)-\mathcal{X}_k(s_0)^{\mathrm{T}}\boldsymbol{\gamma}(s_0)\right\}\end{aligned} \tag{3.20}$$

where $\mathcal{X}_k(s_0) = \frac{1}{nr}\sum_{i=1}^{n}\sum_{j=1}^{r} K_h(s_j-s_0)\widehat{\boldsymbol{\phi}}_k(s_0)^{\mathrm{T}}\mathbf{Q}_{ij}(s_0)\mathbf{W}_{ij}(s_0)$ and $\mathcal{Y}_k(s_0) = \frac{1}{nr}\sum_{i=1}^{n}\sum_{j=1}^{r} K_h(s_j-s_0)\widehat{\boldsymbol{\phi}}_k(s_0)^{\mathrm{T}}\mathbf{Q}_{ij}(s_0)Y_{ij}$. Therefore, the final estimate of the coherent function is $\widehat{\boldsymbol{\beta}}(s_0) = [(1,0)\otimes \mathbf{I}_p]\widehat{\boldsymbol{\gamma}}(s_0)$ where

$$\widehat{\boldsymbol{\gamma}}(s_0) = \left\{\sum_{k=1}^{\infty} \frac{\widehat{\lambda}_k}{\widehat{\lambda}_k^2+\alpha}\mathcal{X}_k(s_0)\mathcal{X}_k(s_0)^{\mathrm{T}}\right\}^{-1}\left\{\sum_{k=1}^{\infty} \frac{\widehat{\lambda}_k}{\widehat{\lambda}_k^2+\alpha}\mathcal{X}_k(s_0)\mathcal{Y}_k(s_0)\right\} \tag{3.21}$$

The algorithm 3.1 summarizes the proposed method. For demonstration purposes, we choose the tuning parameters using cross-validation as discussed in the algorithm. In the proposed algorithm, α controls the number of eigen-values, and can be chosen so that the condition (C8) defined in Section 3.4 is valid. Moreover, a continuity condition for lining-up is required for theoretical justification, by empirical studies, in the present of discontinuity in ϕ_e, the end results are still adequate to use in practice.

Algorithm 3.1 Estimation of $\boldsymbol{\beta}(s) : s \in \mathcal{S}$ for the proposed local-GMM based estimation procedure.

Data: $\{Y_i(s_j), X_i, s_j\}$, for $j = 1, \cdots, r; i = 1, \cdots, n$
Result: Estimate $\boldsymbol{\beta}(s)$ using proposed method

1: Calculate optimal h: $\widehat{h}_{init} \leftarrow \arg\min_{h \in \mathscr{H}} \frac{1}{nr} \sum_{i=1}^{n} \sum_{r=1}^{r} \left\{ Y_{ij} - \mathbf{X}_i^{\mathrm{T}} \breve{\boldsymbol{\beta}}^{-i}(s_r; h) \right\}^2$

2: Calculate $\breve{\boldsymbol{\gamma}}(s; \widehat{h}_{init})$

3: $\mathbf{g}_i\{\boldsymbol{\gamma}(s)\} = \frac{1}{r} \sum_{j=1}^{r} K_{\widehat{h}_{init}}(s_j - s) \mathbf{Q}_{ij}(s; \widehat{h}_{init}) \{ Y_{ij} - \mathbf{W}_{ij}(s; \widehat{h}_{init})^{\mathrm{T}} \breve{\boldsymbol{\gamma}}(s; \widehat{h}_{init}) \}$

4: Determine the instrument variables $\mathfrak{M}(\mathbf{X})$

5: Compute eigen-components $\widehat{\lambda}_k, \widehat{\boldsymbol{\phi}}_k(s)$ and get the value of α using the condition (C8).

6: Calculate optimal h: $\widehat{h}_{opt} \leftarrow \arg\min_{h \in \mathscr{H}} \frac{1}{nr} \sum_{i=1}^{n} \sum_{r=1}^{r} \left\{ Y_{ij} - \mathbf{X}_i^{\mathrm{T}} \widehat{\boldsymbol{\beta}}^{-i}(s_r; h) \right\}^2$

7: Calculate $\widehat{\boldsymbol{\beta}}(s; h_{opt})$

3.4 Asymptotic results

In this section, we provide some assumptions and then study the asymptotic properties of the local linear GMM estimator. Here, we allow the sample size n and the number of functional domains r to grow to infinity. Detailed technical proofs are provided in Section 3.8.

Let $\boldsymbol{\beta}_0(s_0)$ be the true value of $\boldsymbol{\beta}(s_0)$ at the location s_0. For simplicity, define $\delta_{n1}(h) = \left\{ (1 + (hr)^{-1}) \log n/n \right\}^{1/2}$ and $\delta_{n2}(h) = \left\{ (1 + (hr)^{-1} + (hr)^{-2}) \log n/n \right\}^{1/2}$. $\nu_{a,b} = \int t^a K^b(t) dt$. Consider the following conditions that will be useful in asymptotic results.

(C1) Kernel function $K(\cdot)$ is a symmetric density function defined on the bounded support $[-1, 1]$ and is Lipschitz continuous.

(C2) Density function f_T of T is bounded above and away from infinity, and also below and away from zero. Moreover, f is differentiable, and the derivative is continuous.

(C3) $\mathbb{E}\{\|X\|^a\} < \infty$ and $\mathbb{E}\{\sup_{s \in \mathcal{S}} |U(s)|^a\} < \infty$ for some positive $a > 1$. Define $\mathbb{E}\{\mathfrak{M}(\mathbf{X})\mathbf{X}\} = \boldsymbol{\Omega}$ with rank p.

(C4) The true coefficient function $\boldsymbol{\beta}_0(s)$ is three-times continuously differentiable and $\Sigma(s, s')$ are twice continuously differentiable.

(C5) $\{U(s) : s \in [0,1]\}$ and $\{\mathfrak{M}(\mathbf{X})U(s) : s \in [0,1]\}$ are Donsker class, where $\mathbf{X} \subset \mathfrak{M}(\mathbf{X})$

(C6) a) $\lim_{s \searrow 1} \mathbb{E}\{|g_l(s-1) - g_l(0)|^2\} = 0$ for $l = 1, \cdots, d$

b) $\lim_{s \nearrow 1} \mathbb{E}\{|g_{l-1}(s) - g_l(0)|^2\} = 0$ for $l = 2, \cdots, d$

(C7) All second order partial derivatives of $\mathbf{C}(s, s')$ exist and are bounded on the support of the functional domain.

(C8) For some $\kappa_0 \geq 1$ and $\alpha^{-1} = o\left(\sum_{k=1}^{\kappa_0} \lambda_k^{-1} / \sum_{k=\kappa_0+1}^{\infty} \lambda_k\right)$

(C9) The numbers of individuals and functional grid-points are growing to infinity such that $h \to 0$ and $rh \to \infty$. For some positive number $a \in (2,4)$, $|\log h|^{1-2/a}/h \leq r^{1-2/a}$. For $a > 2$, $(h^4 + h^3/r + h^2/r^2)^{-1}(\log n/n)^{1-2/a} \to 0$ as $n \to \infty$.

Remark 3.4.1. *Conditions (C1) and (C2) are commonly used in the literature of non-parametric regression. The bound condition for the density function in (C2) of the functional points is standard for random design. Similar results can be obtained for fixed design where the grid-points are pre-fixed according to the design density* $\int_0^{s_j} f(s)dt = j/r$ *for* $j = 1, \cdots, r$, *for* $r \geq 1$. *Condition (C3) is similar to that in Li and Hsing (2010) which requires the bound on the higher order moment of* $\mathbf{X}$. *Moreover, the rank condition of* $\mathbf{\Omega}$ *is required for the identification of the functional coefficient and its first-order derivatives (Su et al., 2013). Condition (C4) is also common in functional data analysis literature (Wang et al., 2016). This condition allows us to perform the Taylor series expansion. Condition in (C5) avoids the smoothness condition of the sample path (Zhu et al., 2012, 2014) which is commonly assumed in Hall and Hosseini-Nasab (2006); Zhang and Chen (2007); Cardot et al. (2013). Conditions (C6) are required to check the continuity in the mean-zero function, which is equivalent to checking the mean square continuity of the process after lining-up (Hadinejad-Mahram et al., 2002). Here, the first condition shows the limits from right and remain always right; therefore, it involves only one process. A similar but opposite phenomenon occurs in the second condition. Moreover, if the vector process* $\mathbf{g}(s)$ *is mean-square continuous, then both approaches are equivalent, as a result, the covariance function is continuous after lining-up the*

process. To obtain the asymptotic expression of $\widehat{\boldsymbol{\beta}}(s)$*, observed for a fixed sample size, there exists* κ_0 *such that* $k \leq \kappa_0$*,* λ_k^2 *is much larger than* α*, thus the ratio* $\lambda_k/(\lambda_k^2+\alpha) \approx \lambda_k^{-1}$*. On the other hand, if* $k > \kappa_0$*,* $\lambda_k^2 << \alpha$*, as a result, the fraction* $\lambda_k/(\lambda_k^2+\alpha)$ *can be approximately written as* λ_k/α*. Therefore, by the assumption that we make in (C8), we can write for* $s \in \mathcal{S}$*,*

$$\sum_{k=1}^{\kappa_0} \lambda_k^{-1} \boldsymbol{\phi}_k(s)\boldsymbol{\phi}_k(s')^T + \alpha^{-1} \sum_{k=\kappa_0+1}^{\infty} \lambda_k \boldsymbol{\phi}_k(s)\boldsymbol{\phi}_k(s')^T = \sum_{k=1}^{\kappa_0} \lambda_k^{-1} \boldsymbol{\phi}_k(s)\boldsymbol{\phi}_k(s')^T \{1+o(1)\} \tag{3.22}$$

Condition (C9) provide the range of bandwidth. Under the fixed sampling design, this condition can be weakened; see Zhu et al. (2012).

The following result provide the asymptotic properties of the initial estimates mentioned in Step-I.

Theorem 3.4.1. *Under conditions (C1), (C2), (C3), (C4), (C5), and (C9)*

$\left\{\sqrt{n}\left(\breve{\boldsymbol{\beta}}(s_0) - \boldsymbol{\beta}_0(s_0) - 0.5h^2\nu_{21}\ddot{\boldsymbol{\beta}}_0(s_0)\right) \times (1+o_{a.s.}(1)) : s_0 \in \mathcal{S}\right\}$ *weakly converges to a mean zero Gaussian process with a covariance matrix* $\Sigma(s_0,s_0)\boldsymbol{\Omega}_{\mathbf{x}}^{-1}$ *where* $\boldsymbol{\Omega}_{\mathbf{x}} = \mathbb{E}\{\mathbf{X}\mathbf{X}^T\}$.

Next, we study the convergence rates of the estimated eigen-components based on the proposed lining-up method. The following lemma is the output of the asymptotic expansion of eigen-components of an estimated covariance function developed by Li and Hsing (2010).

Lemma 3.4.2. *Under assumptions (C1), (C2), (C3), (C6), (C7), (C8), and (C9) the following convergence holds almost surely for any finite-dimensional mean-zero function* $\mathbf{g}(s)$*.*

1. $\left|\widehat{\lambda}_k - \lambda_k\right| = O(h^2 + \delta_{n1}(h) + \delta_{n2}(h))$
2. $\sup_{s_0\in\mathcal{S}} \left|\widehat{\boldsymbol{\phi}}_k(s_0) - \boldsymbol{\phi}_k(s_0)\right| = O(h^2 + \delta_{n1}(h) + \delta_{n2}^2(h))$

for all $k = 1, \cdots, \kappa_0$*.*

We skip the proof of the above lemma, as it is well developed in the literature of functional data analysis including Hall (2004); Hall and Hosseini-Nasab (2006); Li and Hsing (2010). Next, we show the asymptotic results of the proposed estimation.

Theorem 3.4.3. *Suppose the conditions (C1), (C2), (C3), (C4), (C5), (C6), (C7), (C8), and (C9) hold, then for the proposed local linear GMM estimator $\widehat{\boldsymbol{\beta}}(s)$, have the following results hold.*

$\left\{\sqrt{n}\left(\widehat{\boldsymbol{\beta}}(s) - \boldsymbol{\beta}_0(s) - 0.5h^2\nu_{21}\ddot{\boldsymbol{\beta}}(s)\right)\right)(1 + o_{a.s.}(1)) : s \in \mathcal{S}\Big\}$ *weakly converges to a mean zero Gaussian process with a covariance matrix*

$$\mathscr{A}(s_0, s_0) = \left(\boldsymbol{\Omega}\mathbf{C}^{-1}_{\kappa_0,11}(s_0, s_0)\boldsymbol{\Omega}^T\right)^{-1}\boldsymbol{\Omega}\mathbf{C}^{-1}_{\kappa_0,11}(s_0, s_0)\boldsymbol{\Sigma}(s_0, s_0)\mathbf{C}^{-1}_{\kappa_0,11}(s_0, s_0)\boldsymbol{\Omega}\left(\boldsymbol{\Omega}\mathbf{C}_{\kappa_0,11}(s_0, s_0)^{-1}\boldsymbol{\Omega}^T\right)^{-1}$$

.

The proofs of Theorems 3.4.1 and 3.4.3 are provided in Section 3.8.

3.5 Simulation studies

We conduct numerical studies to compare finite sample performance under different correlation structures and heterogeneity conditions. Data are generated from the following model.

$$Y_i(s) = X_i\boldsymbol{\beta}(s) + U_i(s) \tag{3.23}$$

where we generate trajectories observed at r spatial locations for i-th curve, $i = 1, \cdots, n$. Assume that the functional fixed effect be $\beta(s) = \cos(2\pi s)$ and corresponding fixed effect covariate is generated from the normal distribution with unit mean and variance. The error process is generated as

$$U_i(s) = \xi_1(X_i)\psi_1(s) + \xi_2(X_i)\psi_2(s) \tag{3.24}$$

where $\xi_1(X_i)$ and $\xi_2(X_i)$ are independent central normal random variables with variance $3\sigma^2(X_i)\theta_0^2$ and $1.5\sigma^2(X_i)\theta_0^2$ where θ_0 is determined by the relative importance of random error signal-to-noise ratio, denoted as SNR_θ which is interpreted as the ratio of standard deviation of the additive prediction without noise divided by the standard error of the random noise function. For example, $\text{SNR}_\theta = 2$ means that the contribution of each functional random noise to the variability in $Y(s)$ is about double that of the fixed effect (Scheipl et al., 2015). Here, we use scaled orthonormal functions $\psi_1(s) \propto (1.5 - \sin(2\pi s) - \cos(2\pi s))$ and $\psi_2(s) \propto \sin(4\pi s)$; due to orthonormality, the proportionality constant can be easily determined. Contributions to the conditional variances in $\xi_k(X)$ are specified below.

S.0 $\sigma^2(x) = 1$ (homoskedastic)

S.1 $\sigma^2(x) = (1 + x^2/2)^2$

S.2 $\sigma^2(x) = \exp(1 + x^2/2)$

S.3 $\sigma^2(x) = \exp(1 + |x| + x^2)$

S.4 $\sigma^2(x) = (1 + |x|/2)^2$

The following parameters are considered for each of the above scenarios.

1. **Observational spatial points**. We sample the trajectories at r equidistant spatial points $\{s_1, \cdots, s_r\}$ on $[0, 1]$. Let $s_i = (j - 0.5)/r$ for $j = 1, \cdots, r$ for the i-th curve. The number of spatial points is assumed to be 200 for each case.

2. **Sample size**. Number of trajectories $n \in \{50, 100, 200, 500\}$.

3. **Signal to noise ratio**. The controlling parameter θ_0 is determined using signal-to-noise ratio, SNR_θ which is assumed to be either 0.5 or 1.

Here, for each of the above situations, we perform 500 simulation replicates. To make it consistent with theoretical results and numerical examples, we use "*FPCA*" function in R which is available in *fdapace* packages (Gajardo et al., 2021) or in the MATLAB (MATLAB, 2014) package PACE available at http://www.stat.ucdavis.edu/PACE/ to estimate the eigen-functions. Bandwidths are selected using five-fold generalized cross-validation in all situations and for estimation, the Epanechnikov kernel $K(x) = 0.75(1 - x^2)_+$ is used; where $(a)_+ = \max(a, 0)$. Accuracy of the parameter estimation is assessed using integrated mean square error (IMSE) and integrated mean absolute error (IMAE) which for the $b-$th replication are defined as

$$\text{IMSE}_b = \left[\sum_{j=1}^{r} \left(\widehat{\boldsymbol{\beta}}_b(s_j) - \boldsymbol{\beta}(s_j)\right)^2 \Delta(s_r)\right] \tag{3.25}$$

and

$$\text{IMAE}_b = \left[\sum_{j=1}^{r} |\widehat{\boldsymbol{\beta}}_b(s_j) - \boldsymbol{\beta}(s_j)| \Delta(s_r)\right] \tag{3.26}$$

with $\Delta(s_j) = s_j - s_{j-1}$ and $s_0 = 0$ and $s_1 < \cdots < s_r$ are the observed points over the support set of observational points. We have noticed that the results can be improved by multiplying h^* by a constant in a certain range where h^* is the optimal bandwidth obtained from cross-validation. We use $\widehat{\boldsymbol{\beta}}$ corresponding to bandwidth $0.75h^*$ for our numerical studies. We present Tables 3.1 and 3.2 where IMSEs and IMAEs are averaged over 500 replications for each situation. We denote *LLE*, *LLGMM* and *LLGMM-opt* by local linear smoothing estimator described in Step-I, local linear GMM without incorporating weight matrix and local linear GMM with weight matrix as proposed in Step-III in Section 3.3 respectively. As expected, for all situations, IMSE and IMAE are significantly reduced if we increase the sample size and/or SNR_θ. For the homoskedastic case, the error rates of *LLE* are similar for *LLGM* but under the presence of heteroskedasticity of unknown form, our proposed method outperforms.

Table 3.1 Performance of the estimation procedure with $\text{SNR}_\theta = 0.5$

		n = 50		n = 100		n = 200		n = 500	
Case	Method	IMSE	IMAE	IMSE	IMAE	IMSE	IMAE	IMSE	IMAE
S.0	*LLE*	0.0372	0.1429	0.0189	0.1016	0.0099	0.0737	0.0041	0.0472
	LLGMM	0.0375	0.1435	0.0191	0.1022	0.0099	0.0737	0.0041	0.0471
	LLGMM-opt	0.0388	0.1460	0.0198	0.1038	0.0100	0.0740	0.0042	0.0474
S.1	*LLE*	0.0939	0.2271	0.0516	0.1679	0.0261	0.1189	0.0106	0.0766
	LLGMM	0.0816	0.2123	0.0443	0.1556	0.0227	0.1109	0.0091	0.0708
	LLGMM-opt	0.0585	0.1820	0.0292	0.1262	0.0135	0.0867	0.0050	0.0517
S.2	*LLE*	0.1381	0.2810	0.0812	0.2169	0.0468	0.1632	0.0209	0.1094
	LLGMM	0.0804	0.2105	0.0372	0.1409	0.0164	0.0902	0.0048	0.0486
	LLGMM-opt	0.0557	0.1462	0.0134	0.0817	0.0045	0.0471	0.0015	0.0262
S.3	*LLE*	0.1762	0.3330	0.1018	0.2538	0.0581	0.1913	0.0265	0.1291
	LLGMM	0.0328	0.1069	0.0126	0.0619	0.0055	0.0376	0.0023	0.0243
	LLGMM-opt	0.1067	0.0600	0.0021	0.0201	0.0004	0.0093	0.0003	0.0064
S.4	*LLE*	0.0588	0.1798	0.0309	0.1298	0.0158	0.0928	0.0065	0.0596
	LLGMM	0.0577	0.1782	0.0303	0.1285	0.0155	0.0920	0.0063	0.0589
	LLGMM-opt	0.0576	0.1792	0.0303	0.1287	0.0153	0.0914	0.0063	0.0585

Table 3.2 Performance of the estimation procedure with $\text{SNR}_\theta = 1$

		n = 50		n = 100		n = 200		n = 500	
Case	Method	IMSE	IMAE	IMSE	IMAE	IMSE	IMAE	IMSE	IMAE
S.0	*LLE*	0.0097	0.0729	0.0048	0.0515	0.0025	0.0372	0.0010	0.0237
	LLGMM	0.0099	0.0738	0.0051	0.0526	0.0025	0.0373	0.0010	0.0238
	LLGMM-opt	0.0101	0.0741	0.0052	0.0532	0.0025	0.0374	0.0010	0.0238
S.1	*LLE*	0.0248	0.1169	0.0135	0.0860	0.0068	0.0608	0.0027	0.0387
	LLGMM	0.0142	0.0887	0.0070	0.0617	0.0034	0.0430	0.0013	0.0269
	LLGMM-opt	0.0126	0.0836	0.0062	0.0573	0.0029	0.0403	0.0012	0.0253
S.2	*LLE*	0.0363	0.1440	0.0215	0.1117	0.0124	0.0842	0.0055	0.0560
	LLGMM	0.0103	0.0734	0.0046	0.0480	0.0019	0.0304	0.0006	0.0172
	LLGMM-opt	0.0069	0.0589	0.0029	0.0376	0.0012	0.0239	0.0004	0.0142
S.3	*LLE*	0.0466	0.1705	0.0274	0.1314	0.0157	0.0994	0.0071	0.0669
	LLGMM	0.0050	0.0398	0.0025	0.0273	0.0011	0.0168	0.0004	0.0101
	LLGMM-opt	0.0006	0.0133	0.0003	0.0078	0.0002	0.0069	0.0001	0.0052
S.4	*LLE*	0.0155	0.0924	0.0080	0.0662	0.0041	0.0472	0.0016	0.0300
	LLGMM	0.0141	0.0881	0.0073	0.0631	0.0036	0.0447	0.0015	0.0285
	LLGMM-opt	0.0142	0.0883	0.0074	0.0634	0.0037	0.0449	0.0015	0.0285

3.6 Real data analysis

For illustrating the application of our proposed method and the estimation procedure therein, we use *Apnea-data* to understand white matter structural alterations using diffusion tensor imaging (DTI) in obstructive sleep apnea (OSA) patients (Xiong et al., 2017). The details of this data-set have already been discussed in Chapter 2, Subsection 2.5.2.

FA varies systematically along the trajectory of each white matter fascicle. Several pre- and post-processing steps were performed by the FSL software. The brain was extracted using brain segmentation tools. After generating FA maps using the FMRIB diffusion toolbox, images from all individuals were aligned to an FA standard template through non-linear co-registration. The Johns Hopkins University (JHU) white matter tractography atlas was used as a standard template for white matter parcellation with 50 regions of interest (ROIs). All imaging parameters were calculated by averaging the voxel values in each ROI.

For each subject, we calculate the similarity matrix $\mathbf{C}$ with dimension 50×50. The (k, l)-th

element of the matrix $\mathbf{C}$ is defined as $c_{kl} = |y_k - y_l|$ where y_k is the measure of FA associated with k-th ROI. For simplicity, we scale the similarity matrix such that the range of elements of the matrix is $[0, 1]$. To create the network, we threshold each similarity matrix to build an adjacency matrix $\mathbf{G}$ with elements $\{1, 0\}$ depending on whether the similarity values exceed the threshold or not. Since this threshold controls the topology of the data, we contract the adjacent matrix over a set of threshold parameters from 0.01 to 0.99, and this set is denoted as $\mathcal{S}$ with cardinally 99. A popular measure of the connectivity is average path length (APL) which is defined as the average number of steps along the shortest path for all possible pairs of the network nodes. Therefore, it measures the efficacy of information on a network (Albert and Barabási, 2002). For a series of threshold parameters (s), we observe the APL for FA as shown in Figure 3.1. Scientists are often interested to know the association of APL of the network generated from FA with a set of covariates of interests such as age and lapses score.

We fit the model 3.1 to APLs that are collected over continuous spatial domains (viz, thresholds) from all individuals in which $\mathbf{X}_i$ included clinical variables such as lapses, age. We discard the subjects from the analysis with missing clinical variables and therefore sample size $n = 27$. Here we used three-fold cross-validation to obtain the tuning parameters and the FVE is set be at 0.99. In Figure 3.2, we present the estimated coefficient functions corresponding to age and number of lapses associated with APL where it can be observed that the coefficient of the network property is negative with age but positive with lapses counts. Moreover the effect for the APL is found to be increasing when the significant level is small to moderate and decreasing at moderate to large significance levels; whereas, the effect of APL is more-or-less similar upto the larger values of threshold, and after that it significantly decreases. Here small values of significance thresholds represent sparse connected graph where the true connection might be eliminated due to lenient thresholding; on the other hand, for large significant values, the generated graphs are densely connected.

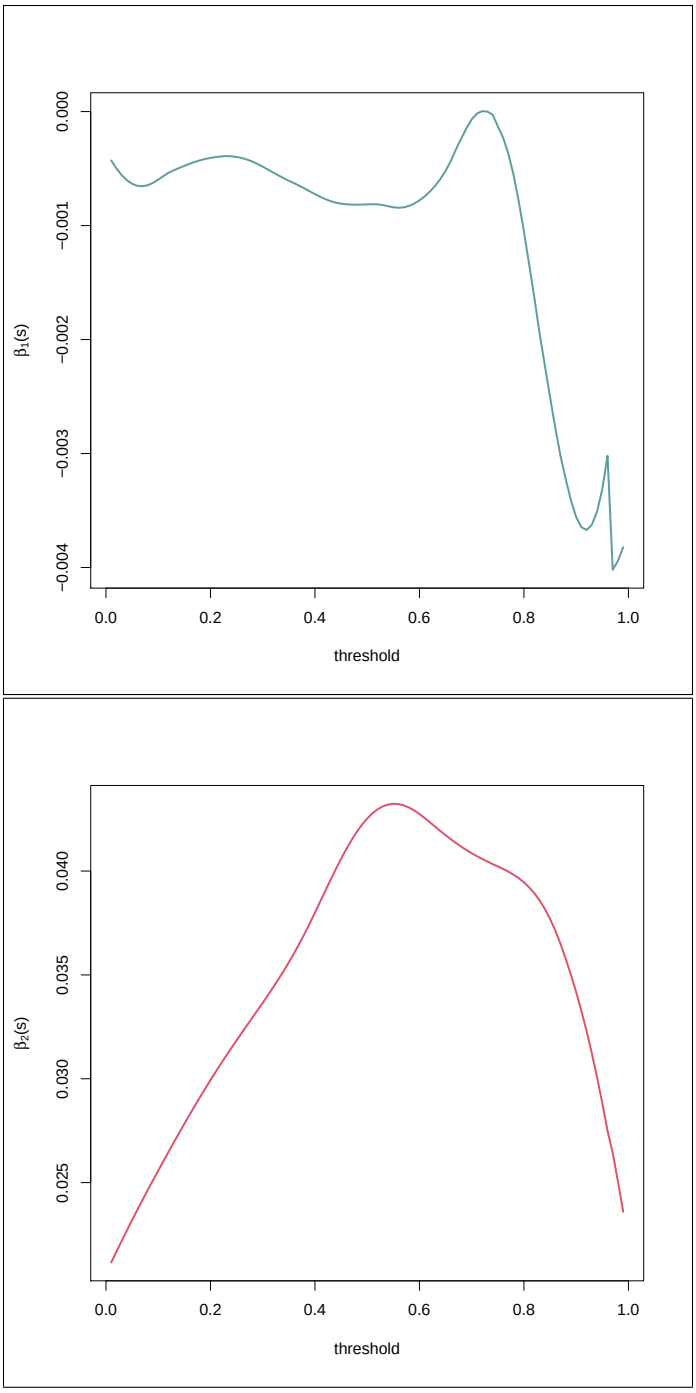

Figure 3.2 *Apnea-data* analysis: Plots of estimated coefficient functions of age (top panel) and number of lapses (bottom panel) for average path length associated with Fractional Anisotropy (FA) in DTI analysis.

3.7 Discussion

In this chapter, we propose an efficient estimation procedure for the varying-coefficient model which is commonly used in neuroimaging and econometrics. We understand that this procedure is an efficient approach to incorporate heteroskedasticity in the analysis of functional data. To best of our knowledge, this is the first initiative to incorporate such a condition into the model. This model is therefore equipped with a more complex relationship between the functional response and real-valued covariates. Additionally, our method is easy to implement in a wide range of applications due to the multi-stage structure of the algorithm. The applicability of the proposed method is illustrated by simulation studies and real data analysis. We leave the testing of hypotheses for linear constraints on $\boldsymbol{\beta}_0(\cdot)$ for future studies.

3.8 Technical details

In this section, we provide technical details of the proposed theorems in Section 3.4. We prove theorems 3.4.1 and 3.4.3 by proving the following lemmas.

3.8.1 Some useful lemmas

Lemma 3.8.1. *Under the conditions (C5)* $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} U_i(s_0)\mathfrak{M}(\mathbf{X}_i)$ *is tight.*

Proof. Consider the class of function $\mathscr{C} = \{U(s_0)\mathfrak{M}(\mathbf{X}_i) : s_0 \in [0,1]\}$. Therefore, due the assumption (C5), $\mathscr{C}$ is a P-Donsker class. Therefore, $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} U_i(s_0)\mathfrak{M}(\mathbf{X}_i)$ is tight. □

Lemma 3.8.2. *Under the assumptions (C1), (C2) and (C9), the following holds for any power* $c \geq 0$.

$$\sup_{s\in[0,1]} \left| \int K_h(t-s) \{(t-s)/h\}^c \, d\Pi(t) - \Pi(t) \right| = O(1/(rh)^{-1/2}) \tag{3.27}$$

The above bound can be replaced by $O(1/rh)$ *for fixed design case.*

Proof. This can be proved by using the empirical process techniques by observing that the class $\{K((\cdot - s/h))((\cdot - s/h))^c : s \in [0,1]\}$ is a P-Donsker class (Zhu et al., 2012). For the balanced case, the results can be shown using Tayler's series expansion. □

Lemma 3.8.3. *Define* $\mathbf{I}(s_0) = \frac{1}{nr}\sum_{i=1}^{n}\sum_{j=1}^{r} K_h(s_j - s_0)\mathbf{W}_{ij}(s_0)\mathbf{Q}_{ij}(s_0)^T$. *Under the conditions (C1), (C2), (C3) and (C9)* $\mathbf{I}(s_0) = f(s_0)diag(1, \nu_{21}) \otimes \mathbf{\Omega} + O(h + \delta_{n1}(h))$ *almost surely, where* $\mathbf{\Omega} = \mathbb{E}\{\mathfrak{M}(\mathbf{X})\mathbf{X}^T\}$.

Proof. Observe the following.

$$\begin{aligned}
\mathbf{I}(s_0) &= \frac{1}{nr}\sum_{i=1}^{n}\sum_{j=1}^{r} K_h(s_j - s_0)\mathbf{W}_{ij}(s_0)\mathbf{Q}_{ij}(s_0)^{\mathrm{T}} \\
&= \frac{1}{nr}\sum_{i=1}^{n}\sum_{j=1}^{r} K_h(s_j - s_0)\left\{\mathbf{z}_h(s_j - s_0)\otimes\mathfrak{M}(\mathbf{X}_i)\right\}\left\{\mathbf{z}_h(s_j - s_0)\otimes\mathbf{X}_i\right\}^{\mathrm{T}} \\
&= \frac{1}{nR}\sum_{i=1}^{n}\sum_{j=1}^{r} K_h(s_j - s)\left\{\mathbf{z}_h(s_j - s_0)^{\otimes^2}\otimes\mathfrak{M}(\mathbf{X}_i)\mathbf{X}_i^{\mathrm{T}}\right\} \\
&= \frac{1}{nr}\sum_{i=1}^{n}\sum_{j=1}^{r} K_h(s_j - s_0)\begin{pmatrix} \mathfrak{M}(\mathbf{X}_i)\mathbf{X}_i^{\mathrm{T}} & \mathfrak{M}(\mathbf{X}_i)\mathbf{X}_i^{\mathrm{T}}(s_j - s_0)/h \\ \mathfrak{M}(\mathbf{X}_i)\mathbf{X}_i^{\mathrm{T}}(s_j - s_0)/h & \mathfrak{M}(\mathbf{X}_i)\mathbf{X}_i^{\mathrm{T}}((s_j - s_0)/h)^2 \end{pmatrix} \\
&:= \begin{pmatrix} \mathbf{I}_{11}(s_0) & \mathbf{I}_{12}(s_0) \\ \mathbf{I}_{21}(s_0) & \mathbf{I}_{22}(s_0) \end{pmatrix}
\end{aligned} \tag{3.28}$$

Let us define $\mathbf{I}_{a,b} = \frac{1}{nr}\sum_{i=1}^{n}\sum_{j=1}^{r} K_h(s_j - s_0)(s_j - s_0)^{a+b}\mathfrak{M}(\mathbf{X}_i)\mathbf{X}_i^{\mathrm{T}}$. Assume that ν_{41} is finite and due to condition (C2),for general index c, we can derive the uniform bound of for all $s_0 \in \mathcal{S}$.

$$\begin{aligned}
\mathbb{E}\{\mathbf{I}_{a,b}(s_0)\} &= \mathbb{E}\left\{\frac{1}{nr}\sum_{i=1}^{n}\sum_{j=1}^{r} K_h(s_j - s_0)((s_j - s_0)/h)^c\mathfrak{M}(\mathbf{X}_i)\mathbf{X}_i^{\mathrm{T}}\right\} \\
&= \mathbf{\Omega}\mathbb{E}\left\{\frac{1}{r}\sum_{j=1}^{r} K_h(s_j - s_0)((s_j - s_0)/h)^c\right\} \\
&= \mathbf{\Omega}\int K_h(u - s_0)((u - s_0)/h)^c f(u)du \\
&= \mathbf{\Omega}\int K(u)u^c f(s_0 + hu)du \\
&= \mathbf{\Omega}\int K(u)u^c\left\{f(s_0) + huf'(s_0) + 0.5h^2u^2f''(s_0) + \cdots\right\}du
\end{aligned}$$

$$= \mathbf{\Omega} \begin{cases} f(s_0) + O(h^2) & c = 0, \text{ provided } \nu_{21} < \infty, f'' \text{ exists and finite} \\ O(h) & c = 1, \text{ provided } \nu_{21} < \infty, f' \text{ exists and finite} \\ f(s_0)\nu_{21} + O(h^2) & c = 2, \text{ provided } \nu_{41} < \infty, f'' \text{ exists and finite} \\ O(h) & c = 3, \text{ provided } \nu_{41} < \infty, f' \text{ exists and finite} \end{cases} \tag{3.29}$$

Moreover, under the condition (C3), we have $\mathbb{E}\|\mathbf{X}\|^a$ is finite for some $a > 2$ and can define, $b_n = h^2 + h/r$ where $h \to 0$ such that $b_n^{-1}(\log n/n)^{1-2/a} = o(1)$. Thus, $\delta_{n1}(h) = \{b_n \log n/nh^2\}^{1/2}$. Now to establish the uniform bound for $\mathbf{I}(s_0)$, by using Lemma 2 in Li and Hsing (2010) for each of $\mathbf{I}_{a,b}(s_0)$ for $a, b = 1, 2$, we have

$$\mathbf{I}(s_0) = f(s_0)(\text{diag}(1, \nu_{21})) \otimes \mathbf{\Omega} + O(h + \delta_{n1}(h)) \qquad \text{almost surely} \tag{3.30}$$

□

Lemma 3.8.4. *Define,* $\mathbf{J}(s_0) = \frac{1}{nr}\sum_{i=1}^{n}\sum_{j=1}^{r} K_h(s_j - s_0)\mathbf{Q}_{ij}(s_0)\mathbf{X}_i^T\boldsymbol{\beta}_0(s_j)$. *Thus, under the conditions (C1), (C2), (C4), (C3) and (C9),* $\mathbf{J}(s_0) - \mathbf{I}(s_0)\boldsymbol{\gamma}_0(s_0) = f(s_0)(\nu_{21}, 0) \otimes \mathbf{\Omega} + O(\delta_{n1}(h) + h)$ *almost surely, where* $\boldsymbol{\gamma}_0(s_0) = (\boldsymbol{\beta}_0(s_0)^T, h\dot{\boldsymbol{\beta}}_0(s_0)^T)^T$. *Moreover,* $\mathbf{T}(s_0) = \frac{1}{nr}\sum_{i=1}^{n}\sum_{j=1}^{r} K_h(s_j - s_0)\mathbf{Q}_{ij}(s_0)U_{ij} = O(\delta_{n1}(h))$ *almost surely.*

Proof. Observe that, because of condition (C4), using Taylor's series expansion,

$$\begin{aligned} \mathbf{J}(s_0) &= \frac{1}{nr}\sum_{i=1}^{n}\sum_{j=1}^{r} K_h(s_j - s_0)\mathbf{Q}_{ij}(s_0)\mathbf{X}_i^{\mathrm{T}}\boldsymbol{\beta}_0(s_0) \\ &= \frac{1}{nr}\sum_{i=1}^{n}\sum_{j=1}^{r} K_h(s_j - s_0)\mathbf{Q}_{ij}(s_0)\left\{\mathbf{X}_i^{\mathrm{T}}\boldsymbol{\beta}_0(s_0) + (s_j - s_0)\mathbf{X}_i^{\mathrm{T}}\dot{\boldsymbol{\beta}}_0(s_0)\right. \\ &\qquad \left. +0.5(s_j - s_0)^2\mathbf{X}_i^{\mathrm{T}}\ddot{\boldsymbol{\beta}}_0(s_0)\right\} + o(h^2) \\ &= \mathbf{I}(s_0)\boldsymbol{\gamma}_0(s_0) + 0.5h^2\mathbf{I}_{21}(s_0) + o(h^2) \end{aligned} \tag{3.31}$$

Using similar arguments, due to Lemma 2 in (Li and Hsing, 2010), under the conditions (C1), (C2)

and (C3), with ν_{41} being finite, we have

$$\begin{aligned}\mathbf{I}_{21}(s_0) &= \frac{1}{nr}\sum_{i=1}^{n}\sum_{j=1}^{r} K_h(s_j - s_0)\begin{pmatrix}((s_j - s_0)/h)^2\\ ((s_j - s_0)/h)^3\end{pmatrix}\mathfrak{M}(\mathbf{X}_i)\mathbf{X}_i^{\mathrm{T}} \\ &= f(s_0)(\nu_{21}, 0)\otimes\mathbf{\Omega} + O(\delta_{n1}(h) + h) \qquad \text{almost surely} \end{aligned} \tag{3.32}$$

and

$$\begin{aligned}\mathbf{T}(s_0) &= \begin{pmatrix}\frac{1}{nr}\sum_{i=1}^{n}\sum_{j=1}^{r} K_h(s_j - s_0)\mathfrak{M}(\mathbf{X}_i)U_{ij}\\ \frac{1}{nr}\sum_{i=1}^{n}\sum_{j=1}^{r} K_h(s_j - s_0)((s_j - s_0)/h)\mathfrak{M}(\mathbf{X}_i)U_{ir}\end{pmatrix} \\ &= O(\delta_{n1}(h)) \qquad \text{almost surely}\end{aligned} \tag{3.33}$$

□

Lemma 3.8.5. *Under conditions (C1),(C2), (C3), (C5), (C9),*

$$\sqrt{n}\mathbf{T}(s_0)(1 + o_{a.s.}(1)) \xrightarrow{d} N(0, f^2(s_0)\Sigma(s_0, s_0)\otimes\mathbf{\Omega}) \tag{3.34}$$

where $\mathbf{T}(s_0)$ *is defined in Lemma 3.8.4.*

Proof. Note that

$$\sqrt{n}\mathbf{T}(s_0) = \frac{1}{\sqrt{nr}}\sum_{i=1}^{n}\sum_{j=1}^{r} K_h(s_j - s_0)\left[\mathbf{z}_h(s_j - s_0)\otimes\mathfrak{M}(\mathbf{X}_i)\right]U_{ij} \tag{3.35}$$

Therefore, the variance of the above quantity is

$$\begin{aligned}&\mathrm{Var}\{\sqrt{n}\mathbf{T}(s_0)\} \\ &= \frac{1}{n}\mathbb{E}\left\{\sum_{i=1}^{n}\sum_{j=1}^{r}\sum_{j'=1}^{r} K_h(s_j - s_0)K_h(s_{j'} - s_0)\left[\mathbf{z}_h(s_j - s_0)\mathbf{z}_h(s_{j'} - s_0)^{\mathrm{T}}\otimes\mathfrak{M}(\mathbf{X}_i)^{\otimes^2}\right]U_{ij}U_{ij'}\right\} \\ &= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left\{\frac{1}{r^2}\sum_{j=1}^{r}\sum_{j'=1}^{r} K_h(s_j - s_0)K_h(s_{j'} - s_0)\left[\mathbf{z}_h(s_j - s_0)\mathbf{z}_h(s_{j'} - s_0)\otimes\mathfrak{M}(\mathbf{X}_i)^{\otimes^2}\right]\Sigma(s_j, s_{j'})\right\} \\ &= \mathbb{E}\left\{\frac{1}{r^2}\sum_{j=1}^{r}\sum_{j'=1}^{r} K_h(s_j - s_0)K_h(s_{j'} - s_0)\mathbf{z}_h(s_j - s_0)\mathbf{z}_h(s_j - s_0)^{\mathrm{T}}\Sigma(s_j, s_{j'})\right\}\otimes\mathbf{\Omega} \\ &= [\mathbb{E}\{\mathbf{D}_1(s_0)\} + \mathbb{E}\{\mathbf{D}_2(s_0)\}]\otimes\mathbf{\Omega}\end{aligned} \tag{3.36}$$

where $\mathbf{D}_1(s_0) = \frac{1}{r^2}\sum_{j=1}^{n} K_h^2(s_j - s_0)\mathbf{z}_h(s_j - s_0)^{\otimes^2}\Sigma(s_j, s_j)$ and $\mathbf{D}_2(s_0) = \frac{1}{r(r-1)}\sum_{j=1}^{n}\sum_{\substack{j'=1 \\ j\neq j'}}^{r} K_h(s_j - s_0)K_h(s_{j'} - s_0)\mathbf{z}_h(s_j - s_0)\mathbf{z}_h(s_{j'} - s_0)^{\mathrm{T}}\Sigma(s_j, s_{j'})$. Note that

$$\begin{aligned}
\mathbb{E}\{\mathbf{D}_2(s_0)\} &= \mathbb{E}\left\{\frac{1}{r^2}\sum_{j=1}^{r} K^2(s_j - s_0)\mathbf{z}_h(s_j - s_0)^{\otimes^2}\Sigma(s_j, s_j)\right\} \\
&= \frac{1}{r}\int K_h^2(t - s_0)\mathbf{z}_h(t - s_0)^{\otimes^2}\Sigma(t, t)f(t)dt \\
&= \frac{1}{hr}\int K^2(t)\begin{pmatrix} 1 & t \\ t & t^2 \end{pmatrix}\Sigma(s_0 + hu)f(s_0 + ht)dt \\
&= \frac{1}{hr}\{f(s_0)\mathrm{diag}(\nu_{02}, \nu_{22})\Sigma(s_0, s_0) + O(h)\} \qquad (3.37)
\end{aligned}$$

Now assume that $\Theta(s_0) = \mathbb{E}\{\mathbf{D}_2(s_0)\}$ with (l, l')-th entry $\theta_{l,l'}$ and $\mathscr{P}(t) = \int_{\mathcal{S}} K_h(t - s_0)K_h(t' - s_0)\mathbf{z}_h(t - s_0)\mathbf{z}_h(t' - s_0)\Sigma(t, t')f(t')dt'$ with (l, l')-the element $\mathscr{P}_{l,l'}$. Therefore, using Hájek projection (Vaart and Wellner, 1996), we have

$$\mathbf{D}_{1,l.l'}(s_0) = \theta_{l,l'}(s_0) + \frac{2}{r}\sum_{j=1}^{r}\left\{\mathscr{P}_{l,l'}(s_j) - \theta_{l,l'}(s_0)\right\} + \tilde{\epsilon}_{l,l'}(s_0) \qquad (3.38)$$

where $\frac{2}{r}\sum_{j=1}^{r}\left\{\mathscr{P}_{l,l'}(s_j) - \theta_{l,l'}(s_0)\right\}$ is the projection on $\mathbf{D}_{2,',l'}(s_0) - \theta_{l,l'}(s_0)$ onto the set of all statistics of the linear order form. Thus, it is easy to see $\mathrm{Var}\{\tilde{\epsilon}\} = O(1/(rh)^2)$ (Zhu et al., 2012). Since the Taylor series expansion for small $h \to 0$, we have $\theta_{l,l'}(s_0) = f(s_0)^2\nu_{l-1,}\nu_{l'-1,1}\Sigma(s_0, s_0)$. Therefore, in summery, we have $\mathrm{Var}\{\sqrt{n}\mathbf{T}(s_0)\} = f^2(s_0)\mathscr{U}\Sigma(s_0, s_0)$. where the element (l, l') of the matrix $\mathscr{U}$ is $\nu_{l-1}\nu_{l'-1}$.

To hold the above asymptotic results, we need to show that $\sqrt{n}\mathbf{T}(s_0)$ be tight asymptotically. Therefore, consider the following, for suitable choice of $\underline{l} < \bar{l}$ after change of variables,

$$\begin{aligned}
&\sqrt{n}\mathbf{T}(s_0) \\
&= \frac{1}{\sqrt{n}r}\sum_{i=1}^{n}\sum_{r=1}^{r} K_h(s_j - s_0)[\mathbf{z}_h(s_j - s_0) \otimes \mathfrak{M}(\mathbf{X}_i)]U_{ij} \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\frac{1}{r}\sum_{j=1}^{r} K_h(s_j - s_0)\mathbf{z}_h(s_j - s_0)U_{ij} - \int_0^1 K_h(t - s_0)\mathbf{z}_h(t - s_0)U_i(t)f(t)dt\right\} \otimes \mathfrak{M}(\mathbf{X}_i) \\
&\quad + \frac{1}{\sqrt{n}}\sum_{i=1}^{n} U_i(s_0)\int_{\underline{l}}^{\bar{l}} K(t)(1, t)^{\mathrm{T}} f(s_0 + ht)dt \otimes \mathfrak{M}(\mathbf{X}_i)
\end{aligned}$$

$$+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_{\underline{l}}^{\bar{l}}K(t)(1,t)^{\mathrm{T}}\left\{U_i(s_0+ht)-U_i(s_0)\right\}f(s_0+ht)dt\otimes\mathfrak{M}(\mathbf{X}_i)$$

$$:=\mathbf{T}_1(s_0)+\mathbf{T}_2(s_0)+\mathbf{T}_3(s_0) \tag{3.39}$$

Note that,

$$\begin{aligned}
&\mathbf{T}_1(s_0)\\
&=\frac{1}{r}\sum_{r=1}^{r}K_h(s_j-s_0)\mathbf{z}_h(s_j-s_0)\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}U_{ij}\otimes\mathfrak{M}(\mathbf{X}_i)-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}U_i(t)\otimes\mathfrak{M}(\mathbf{X}_i)\right\}\\
&\quad+\left\{\frac{1}{r}\sum_{j=1}^{r}K_h(s_j-s_0)\mathbf{z}_h(s_j-s_0)-\int_{\underline{l}}^{\bar{l}}K_h(t-s_0)\mathbf{z}_h(t-s_0)f(t)dt\right\}\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}U_i(t)\otimes\mathfrak{M}(\mathbf{X}_i)\right\}\\
&\quad+\int_{\underline{l}}^{\bar{l}}K_h(t-s_0)\mathbf{z}_h(t-s_0)\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{U_i(s_0)-U_i(t)\right\}f(t)dt\otimes\mathfrak{M}(\mathbf{X}_i)\\
&:=\mathbf{T}_{11}(s_0)+\mathbf{T}_{12}(s_0)+\mathbf{T}_{13}(s_0)
\end{aligned} \tag{3.40}$$

Due to the Donsker Theorem, we have $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathfrak{M}(\mathbf{X}_i)U_i(s)$ weekly converges to a centered Gaussian process and $\sup_{s\in[0,1]}|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathfrak{M}(\mathbf{X}_i)U_i(s)|=O_p(1)$ (Vaart and Wellner, 1996). Therefore,

$$\begin{aligned}
&|\mathbf{T}_{11}(s_0)|\\
&\leq\frac{1}{r}\sum_{j=1}^{r}K_h(s_j-s_0)\|\mathbf{z}_h(s_j-s_0)\|_2\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}U_{ij}\otimes\mathfrak{M}(\mathbf{X}_i)-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}U_i(s)\otimes\mathfrak{M}(\mathbf{X}_i)\right|\\
&\leq\frac{1}{r}\sum_{j=1}^{r}K_h(s_j-s_0)\|\mathbf{z}_h(s_j-s_0)\|_2\sup_{|s-s_0|\leq h}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}U_i(s)\otimes\mathfrak{M}(\mathbf{X}_i)-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}U_i(s_0)\otimes\mathfrak{M}(\mathbf{X}_i)\right|\\
&=o_P(1)
\end{aligned} \tag{3.41}$$

$$\begin{aligned}
&|\mathbf{T}_{12}(s_0)|\\
&\leq\left|\frac{1}{r}\sum_{j=1}^{r}K_h(s_j-s_0)\mathbf{z}_h(s_j-s_0)-\int_{\underline{l}}^{\bar{l}}K_h(t-s_0)\mathbf{z}_h(t-s_0)f(t)dt\right|\sup_{t\in[0,1]}\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}U_i(t)\otimes\mathfrak{M}(\mathbf{X}_i)\right\}\\
&=O_P(1/\sqrt{rh})O_P(1)=o_P(1)
\end{aligned} \tag{3.42}$$

The above bound holds for Lemma 3.8.2 and Condition (C9) so that $mh \to \infty$.

$$
\begin{aligned}
&|\mathbf{T}_{13}(s_0)| \\
&\leq \sup_{|s-s_0|\leq h} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} U_i(s) \otimes \mathfrak{M}(\mathbf{X}_i) - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} U_i(s_0) \otimes \mathfrak{M}(\mathbf{X}_i) \right| \int_{\underline{l}}^{\bar{l}} K_h(s_j - s_0) \|\mathbf{z}_h(s_j - s_0)\|_2 f(s) ds \\
&= O_P(1) \qquad (3.43)
\end{aligned}
$$

By combining the above three bounds, due to conditions (C1),(C2), (C3), (C5), (C9), we obtain $\mathbf{T}_1(s_0) = o_P(1)$. Now, rewrite $\mathbf{T}_3(s_0)$ as

$$
\begin{aligned}
\mathbf{T}_3(s) &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int_{\underline{l}}^{\bar{l}} K(t)(1,t)^{\mathrm{T}} \left\{ U_i(s_0 + ht) - U_i(s_0) \right\} f(s_0 + ht) dt \otimes \mathfrak{M}(\mathbf{X}_i) \\
&= \int_{\underline{l}}^{\bar{l}} K(t)(1,t)^{\mathrm{T}} \otimes \left\{ U_i(s_0 + ht) - U_i(s_0) \right\} \mathfrak{M}(\mathbf{X}_i) f(s_0 + ht) dt \qquad (3.44)
\end{aligned}
$$

Since, $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathfrak{M}(\mathbf{X}_i) U_i(s_0)$is asymptotically tight, for any $h \to 0$, we have the following (Vaart and Wellner, 1996).

$$
\sup_{s_0 \in [0,1]:|t|\leq 1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathfrak{M}(\mathbf{X}_i) \left\{ U_i(s_0 + ht) - U_i(s_0) \right\} = o_P(1) \qquad (3.45)
$$

Now it is enough to show that $\mathbf{T}_2(s_0)$ is tight. First, observe that

$$
\begin{aligned}
&(1,0) \int_{\underline{l}}^{\bar{l}} K(t) \mathrm{diag}(1, \nu_{21}^{-1})(1,t)^{\mathrm{T}} f(s_0 + ht) dt \\
&= \int_{\underline{l}}^{\bar{l}} K(t) f(s_0 + ht) dt \\
&= \int_{\underline{l}}^{\bar{l}} K(t) \left\{ f(s_0) + htf'(s_0) + \cdots \right\} \\
&= f(s_0) + o(h) \qquad (3.46)
\end{aligned}
$$

Therefore, $\mathbf{T}_2(s_0)(1 + o_P(h)) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} U_i(s_0) \otimes \mathfrak{M}(\mathbf{X}_i)$. By assumption (C5), $\mathbf{T}_2(s_0)$ is tight. □

3.8.2 Proof of Theorem 3.4.1

Under the initial estimates, by considering $\mathfrak{M}(\mathbf{X}) = \mathbf{X}$, $\mathbf{\Omega}$ can be replaced by $\mathbf{\Omega}_{\mathbf{x}}$ in Equation (3.47) and inverse of $\mathbf{\Omega}_{\mathbf{x}}$ exits. Therefore, by using Lemma 3.8.3, it is easy to observe that, almost surely

$$
\mathbf{I}(s_0)^{-1} = f(s_0)^{-1} (\mathrm{diag}(1, \nu_{21})^{-1}) \otimes \mathbf{\Omega}_{\mathbf{x}}^{-1} + O(h + \delta_{n1}(h)) \qquad (3.47)
$$

Similarly, for the numerator, we have the following.

$$\begin{aligned}
&\frac{1}{nr}\sum_{i=1}^{n}\sum_{j=1}^{r} K_h(s_j - s_0)\left\{\mathbf{z}_h(s_j - s_0)\otimes \mathbf{X}_i\right\} Y_{ij} \\
&= \frac{1}{nr}\sum_{i=1}^{n}\sum_{j=1}^{r} K_h(s_j - s_0)\left\{\mathbf{z}_h(s_j - s_0)\otimes \mathbf{X}_i\right\}\left\{\mathbf{X}_i^{\mathrm{T}}\boldsymbol{\beta}_0(s_j) + U_{ij}\right\} \\
&= \mathbf{I}(s_0)\boldsymbol{\gamma}_0(s_0) + 0.5h^2\mathbf{I}_{21}(s_0)\ddot{\boldsymbol{\beta}}_0(s_0) + \mathbf{T}(s_0) + o(h^2) \qquad (3.48)
\end{aligned}$$

Thus, using Equation (3.47) and (3.48), we can derive,

$$\begin{aligned}
\check{\boldsymbol{\beta}}(s_0) &= [(1,0)\otimes \mathbf{I}_p]\mathbf{I}(s_0)^{-1}\left\{\mathbf{I}(s_0)\boldsymbol{\gamma}_0(s_0) + 0.5h^2\mathbf{I}_{21}(s_0)\ddot{\boldsymbol{\beta}}_0(s_0) + \mathbf{T}(s_0) + o(h^2)\right\} \\
&= \boldsymbol{\beta}_0(s_0) + [(1,0)\otimes \mathbf{I}_p]f(s_0)^{-1}\left\{\mathrm{diag}(1,\nu_{21})^{-1}\otimes \boldsymbol{\Omega}_{\mathbf{x}}^{-1}\right\}\{f(s_0)(\nu_{21},0)\otimes \boldsymbol{\Omega}_{\mathbf{x}}\}\,0.5h^2\ddot{\boldsymbol{\beta}}_0(s_0) \\
&\quad + O(\delta_{n1}(h) + h) \\
&= \boldsymbol{\beta}_0(s_0) + 0.5h^2\nu_{21}\ddot{\boldsymbol{\beta}}_0(s_0) + O(\delta_{n1}(h) + h) \\
&= \boldsymbol{\beta}_0(s_0) + O(\delta_{n1}(h) + h) \qquad \text{almost surely} \qquad (3.49)
\end{aligned}$$

Therefore, $\sup_{s_0\in\mathcal{S}}\left|\check{\boldsymbol{\beta}}(s_0) - \boldsymbol{\beta}_0(s_0)\right| = O(\delta_{n1} + h)$ almost surely. Furthermore, observe that the bias of the initial estimator is

$$\mathbb{E}\{\check{\boldsymbol{\beta}}(s_0)\} - \boldsymbol{\beta}_0(s_0) = 0.5h^2\nu_{21}\ddot{\boldsymbol{\beta}}_0(s_0)\left\{1 + O_P(\delta_{n1}(h) + h)\right\} \qquad (3.50)$$

Now, to calculate the variance, note that

$$\begin{aligned}
&\sqrt{n}\{\check{\boldsymbol{\beta}}(s_0) - \boldsymbol{\beta}(s_0) - 0.5h^2\nu_{21}\ddot{\beta}_0(s_0)\}(1 + o_{a.s.}(1)) \\
&= [(1,0)\otimes \mathbf{I}_p]f(s_0)\left\{\mathrm{diag}(1,\nu_{21})^{-1}\otimes \boldsymbol{\Omega}_{\mathbf{x}}^{-1}\right\}\sqrt{n}\mathbf{T}(s_0) \qquad (3.51)
\end{aligned}$$

By Lemma 3.8.5, we have the variance of the above quantity $\Sigma(s_0, s_0)\boldsymbol{\Omega}_{\mathbf{x}}^{-1}$.

3.8.3 Proof of Theorem 3.4.3

Define $\mathbf{C}_{\kappa_0}(s, s') = \sum_{k=1}^{\kappa_0}\lambda_k\boldsymbol{\phi}_k(s)\boldsymbol{\phi}_k(s')^{\mathrm{T}}$ and hence, we can define $\mathbf{C}_{\kappa_0}^{-1}(s, s')$ with possible block matrix

$$\mathbf{C}_{\kappa_0}^{-1}(s, s') = \sum_{k=1}^{\kappa_0}\lambda_k^{-1}\boldsymbol{\phi}_k(s)\boldsymbol{\phi}_k(s')^{\mathrm{T}} = \begin{pmatrix} \mathbf{C}_{\kappa_0,1,1}^{-1}(s, s') & 0 \\ 0 & \mathbf{C}_{\kappa_0,2,2}^{-1}(s, s') \end{pmatrix} \qquad (3.52)$$

Also define,

$$\widehat{\boldsymbol{\gamma}}_{\kappa_0}(s_0) = \left\{ \sum_{k=1}^{\kappa_0} \frac{\lambda_k}{\lambda_k^2 + \alpha} \mathscr{X}_k(s_0; \kappa_0) \mathscr{X}_k(s_0; \kappa_0)^{\mathrm{T}} \right\}^{-1} \left\{ \sum_{k=1}^{\kappa_0} \frac{\lambda_k}{\lambda_k^2 + \alpha} \mathscr{X}_k(s_0; \kappa_0) \mathscr{Y}_k(s_0; \kappa_0) \right\} \tag{3.53}$$

where

$$\mathscr{X}_k(s_0; \kappa_0) = \frac{1}{nr} \sum_{j=1}^{n} \sum_{j=1}^{r} K_h(s_j - s_0) \boldsymbol{\phi}_k(s_0)^{\mathrm{T}} \mathbf{Q}_{ij}(s_0) \mathbf{W}_{ij}(s_0) \tag{3.54}$$

and

$$\mathscr{Y}_k(s_0; \kappa_0) = \frac{1}{nr} \sum_{j=1}^{n} \sum_{j=1}^{r} K_h(s_j - s_0) \boldsymbol{\phi}_k(s_0)^{\mathrm{T}} \mathbf{Q}_{ij}(s_0) Y_{ij} \tag{3.55}$$

Therefore, we have the following.

$$\begin{aligned}
&\sum_{k=1}^{\kappa_0} \frac{\lambda_k}{\lambda_k^2 + \alpha} \mathscr{X}_k(s_0; \kappa_0) \mathscr{X}_k(s_0; \kappa_0)^{\mathrm{T}} \\
&= \left\{ \frac{1}{nr} \sum_{i=1}^{n} \sum_{j=1}^{r} K_h(s_j - s_0) \mathbf{W}_{ij}(s_0) \mathbf{Q}_{ij}(s_0)^{\mathrm{T}} \right\} \\
&\quad \times \sum_{k=1}^{\kappa_0} \lambda_k^{-1} \boldsymbol{\phi}_k(s_0) \boldsymbol{\phi}_k(s_0)^{\mathrm{T}} \left\{ \frac{1}{nr} \sum_{i=1}^{n} \sum_{j=1}^{r} K_h(s_j - s_0) \mathbf{W}_{ij}(s_0) \mathbf{Q}_{ij}(s_0)^{\mathrm{T}} \right\}^{\mathrm{T}} \\
&= \mathbf{I}(s_0) \mathbf{C}_{\kappa_0}^{-1}(s_0, s_0) \mathbf{I}(s_0)^{\mathrm{T}} \\
&= f^2(s_0) \left[\mathrm{diag}(1, \nu_{21}) \otimes \boldsymbol{\Omega} \right] \mathbf{C}_{\kappa_0}^{-1}(s_0, s_0) \left[\mathrm{diag}(1, \nu_{21}) \otimes \boldsymbol{\Omega} \right]^{\mathrm{T}} + O(\delta(h)) \\
&= \mathscr{V}(s_0) + O(\delta(h))
\end{aligned} \tag{3.56}$$

where we define $\mathscr{V}(s_0) = f^2(s_0) \mathrm{diag} \left(\boldsymbol{\Omega} \mathbf{C}_{\kappa_0,1,1,}^{-1}(s_0, s_0) \boldsymbol{\Omega}^{\mathrm{T}}, \nu_{21}^2 \boldsymbol{\Omega} \mathbf{C}_{\kappa_0,2,2,}^{-1}(s_0, s_0) \boldsymbol{\Omega}^{\mathrm{T}} \right)$ and

$$\begin{aligned}
&\sum_{k=1}^{\kappa_0} \frac{\lambda_k}{\lambda_k^2 + \alpha} \mathscr{X}_k(s_0; \kappa_0) \mathscr{Y}_k(s_0; \kappa_0) \\
&= \left\{ \frac{1}{nr} \sum_{i=1}^{n} \sum_{j=1}^{r} K_h(s_j - s_0) \mathbf{W}_{ij}(s_0) \mathbf{Q}_{ij}(s_0)^{\mathrm{T}} \right\} \\
&\quad \sum_{k=1}^{\kappa_0} \lambda_k^{-1} \boldsymbol{\phi}_k(s_0) \boldsymbol{\phi}_k(s_0)^{\mathrm{T}} \left\{ \frac{1}{nr} \sum_{i=1}^{n} \sum_{j=1}^{r} K_h(s_j - s_0) \mathbf{Q}_{ij}(s_0) Y_{ij} \right\} \\
&= \mathbf{I}(s_0) \mathbf{C}_{\kappa_0}^{-1}(s_0, s_0) \left\{ \mathbf{I}(s_0) \boldsymbol{\gamma}_0(s_0) + 0.5 h^2 \mathbf{I}_{21}(s_0) \ddot{\boldsymbol{\beta}}(s_0) + \mathbf{T}(s_0) + o(h^2) \right\}
\end{aligned} \tag{3.57}$$

Therefore,

$$
\begin{aligned}
&\widehat{\boldsymbol{\beta}}(s_0) - \boldsymbol{\beta}_0(s_0) \\
&= 0.5h^2[\text{diag}(1,0) \otimes \mathbf{I}_p]\mathscr{V}(s_0)^{-1}\mathbf{I}(s_0)\mathbf{C}_{\kappa_0}^{-1}(s_0,s_0)\mathbf{I}_{21}(s_0)\ddot{\boldsymbol{\beta}}(s_0) \\
&\quad + [\text{diag}(1,0) \otimes \mathbf{I}_p]\mathscr{V}(s_0)^{-1}\mathbf{I}(s_0)\mathbf{C}_{\kappa_0}^{-1}(s_0,s_0)\mathbf{T}(s_0) + O(\delta(h)) \\
&= 0.5h^2 f^2(s_0)[\text{diag}(1,0) \otimes \mathbf{I}_p]\mathscr{V}(s_0)^{-1}[\text{diag}(1,\nu_{21}) \otimes \boldsymbol{\Omega}]\mathbf{C}_{\kappa_0}^{-1}(s_0,s_0)[(\nu_{21},0) \otimes \boldsymbol{\Omega}]^{\mathrm{T}}\ddot{\boldsymbol{\beta}}(s_0) \\
&\quad + f^2(s_0)[\text{diag}(1,0) \otimes \mathbf{I}_p]\mathscr{V}(s_0)^{-1}[\text{diag}(1,\nu_{21}) \otimes \boldsymbol{\Omega}]\mathbf{C}_{\kappa_0}^{-1}(s_0,s_0)\mathbf{T}(s_0) + O(\delta(h)) \\
&= 0.5h^2\nu_{21}\ddot{\boldsymbol{\beta}}(s_0) + \mathscr{T}(s_0) + O(\delta(h)) \qquad (3.58)
\end{aligned}
$$

In order to obtain the asymptotic variance, consider, using Lemmas 3.8.3, 3.8.4 and 3.8.5, we have $\sqrt{n}\{\widehat{\boldsymbol{\beta}}(s_0) - \boldsymbol{\beta}_0(s_0) - 0.5h^2\nu_{21}\ddot{\boldsymbol{\beta}}(s_0)\} \xrightarrow{d} N(0, \mathscr{A}(s_0,s_0))$ where $\mathscr{A}(s_0,s_0)$ is the asymptotic variance of $\sqrt{n}\mathscr{T}(s_0)$, where we derive, $\mathscr{A}(s_0,s_0) = [(1,0)\otimes\mathbf{I}_p]\mathscr{V}^{-1}(s_0)\tilde{\mathscr{A}}(s_0,s_0)\mathscr{V}^{-1}(s_0)[(1,0)\otimes\mathbf{I}_p]$ for $\tilde{\mathscr{A}}(s_0,s_0) = [\text{diag}(1,\nu_{21}) \otimes \boldsymbol{\Omega}]\mathbf{C}_{\kappa_0}^{-1}(s_0,s_0)\text{diag}(\Sigma(s_0,s_0), \nu_{11}^2\Sigma(s_0,s_0))\mathbf{C}_{\kappa_0}^{-1}(s_0,s_0)[\text{diag}(1,\nu_{21}) \otimes \boldsymbol{\Omega}]^{\mathrm{T}}$. By simple calculation, it can be shown that

$$
\mathscr{A}(s_0,s_0) = (\boldsymbol{\Omega}\mathbf{C}_{\kappa_0,11}^{-1}(s_0,s_0)\boldsymbol{\Omega}^{\mathrm{T}})^{-1}\boldsymbol{\Omega}\mathbf{C}_{\kappa_0,11}^{-1}(s_0,s_0)\boldsymbol{\Sigma}(s_0,s_0)\mathbf{C}_{\kappa_0,11}^{-1}(s_0,s_0)\boldsymbol{\Omega}(\boldsymbol{\Omega}\mathbf{C}_{\kappa_0,11}^{-1}(s_0,s_0)\boldsymbol{\Omega}^{\mathrm{T}})^{-1} \qquad (3.59)
$$

CHAPTER 4

TENSOR BASED SPATIO-TEMPORAL MODELS FOR ANALYSIS OF FUNCTIONAL NEUROIMAGING DATA

4.1 Introduction

Recent years have seen an explosive growth in the number of neuroimaging studies performed. Popular imaging modalities include functional magnetic resonance imaging (fMRI), electroencephalography (EEG), diffusion tensor imaging (DTI), positron emission tomography (PET), and single-photon emission computed tomography (SPECT). Each of these techniques has its own limitations and strengths. Therefore, a current trend is toward interdisciplinary approaches that use multiple imaging techniques to overcome limitations of each method in isolation. As an example, Figure 4.1 illustrates the combination of fMRI and EEG data. At the same time, neuroimaging data is increasingly being combined with non-imaging modalities, such as behavioral and genetic data.

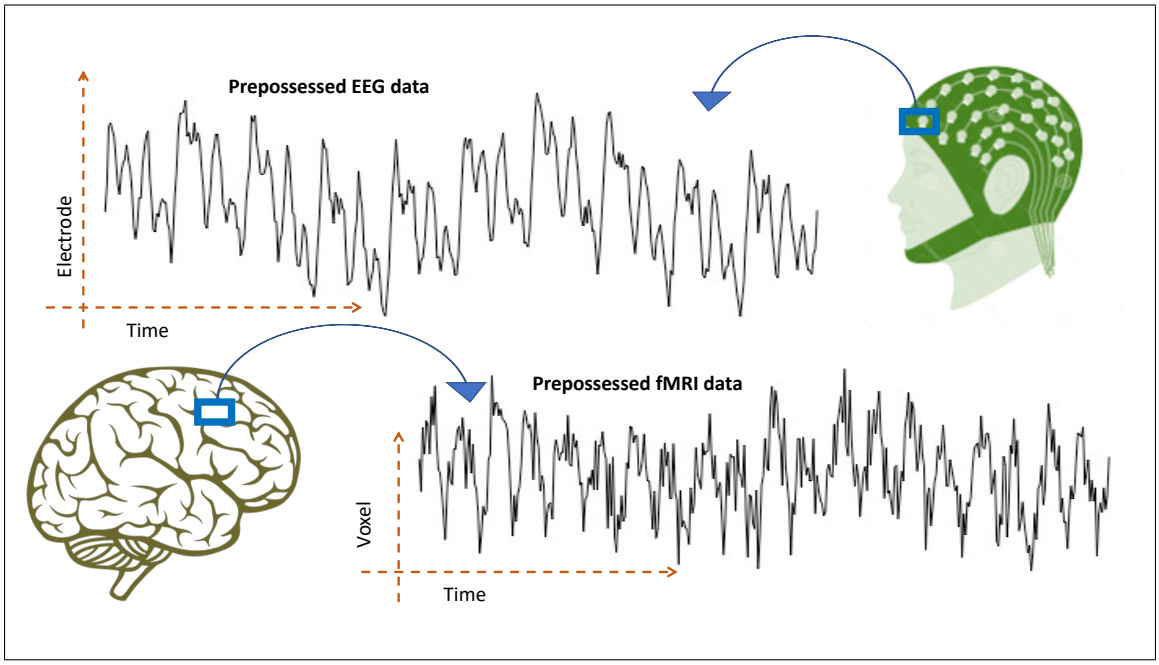

Figure 4.1 *Multi-modal-data*: An example of multi-modal data analysis which seeks to explore the relationship between EEG and fMRI data.

Multi-modal analysis is an increasingly important topic of research, and to fully realize its promise, novel statistical techniques are needed. Here, we present a new approach towards performing such analysis. It is common for the data generated from neuroimaging studies to consist of time-varying signal measured over a large three-dimensional (3D) domain (Lindquist, 2008; Ombao et al., 2016). Hence, the data are inherently spatio-temporal in nature. Due to the massive size of the data along with its complex anatomical structure, classical vector-based spatio-temporal statistical methods are often deemed unrealistic and inadequate. It is becoming increasingly clear that any new model and methodology should address three fundamental concerns. First, standard spatio-temporal covariance modelling techniques are based on many parametric assumptions, which are often hard to validate in large high-dimensional data such as fMRI. Second, modelling of spatio-temporal interactions often produces large covariance matrices containing millions of elements that are hard to estimate properly. Third, storage of these large data-sets while performing analysis is nearly impossible.

The current research is motivated by the experiment *studyforrest* (http://studyforrest.org/) which investigates high-level cognition in the human brain using complex natural stimulation, namely watching the Hollywood movie *Forrest Gump* (1994). The data consist of several hours of fMRI scans, structural brain images, eye-tracking data, and extensive annotations of the movie. Details of this experiment are presented in Section 4.6. In our motivating example, we focus on data consisting of voxel-wise fMRI images, measured over a large number of spatial locations (voxels) at 451 time-points. The goal of our analysis is to use the multivariate eye-tracking data, measured while the participants watch the movie, as covariates in a model that explains changes in the multivariate brain data. The vast size and scale of these data call for well-equipped statistical techniques to find the association between brain regions and other covariates over time-varying activities. It is useful to consider this as a regression problem with a multi-dimensional array of outcomes and predictors. These multi-dimensional arrays are popularly known as tensors. Figure 4.2 illustrates the reason for considering a time-varying multi-dimensional array for analysis. Although the signals in both modalities (in this case fMRI and eye-tracking) are measured discretely over time, we consider

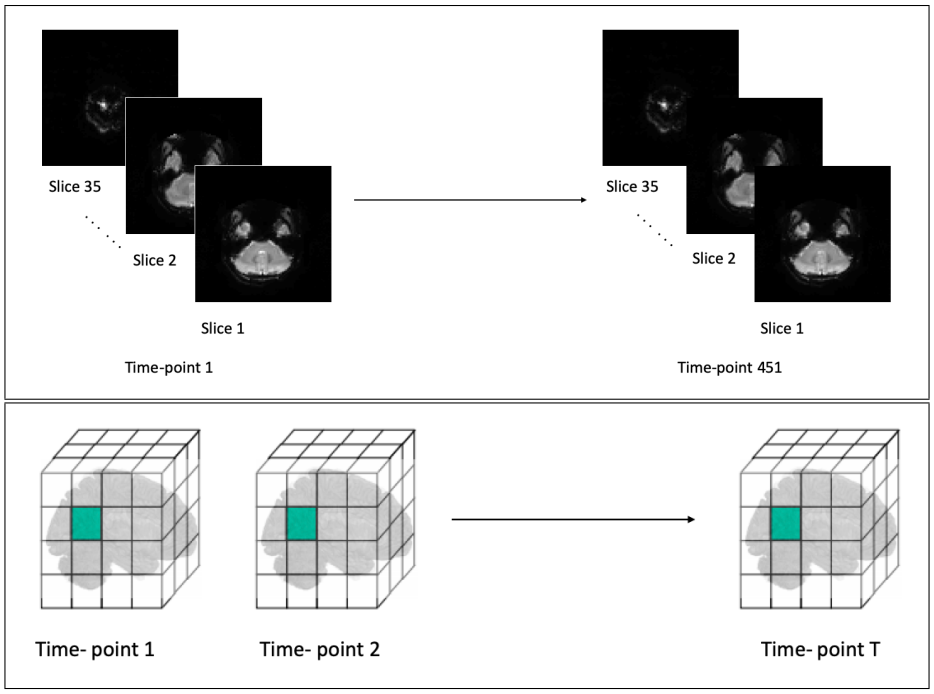

Figure 4.2 *ForrestGump-data*: (Top panel) BOLD fMRI for an example subject during their first run (see Section 4.6 for details). 35 axial slices (thickness 3.0 mm) represents the third mode of the tensor with 80×80 voxels (3.0×3.0 mm) in-plate resolution measured at every repetition time (TR) of 2 seconds. (Bottom panel) fMRI data-set consists of a time series of 3D images (tensors) at each TR (source: Wager and Lindquist (2015)).

them to be discrete measures of a smooth underlying function over time in a certain interval. This assumption is reasonable in the context of both brain activity and eye movement, as they can potentially change at any moment.

There are two main advantages to taking a tensor-based approach (Guo et al., 2012) towards modeling this data-set. First, we can represent the unknown parameters to be estimated as a linear combination of rank-1 components, where the latter are expressed as the outer product of low-dimensional vectors. This allows the estimation of fewer parameters, which is consistent with variable selection or dimension reduction problems in statistics. Second, because of the need to estimate fewer parameters, the computational complexity is significantly reduced.

In an exploratory analysis of multi-dimensional data, principal component analysis (PCA) is one of the most common tools for reducing dimensionality. Its use in tensor data has been studied in various articles; for example, Liu et al. (2017) provided a generalized classical PCA that can

deal with data matrices and tensors and can explore the spatial and temporal dependencies of data simultaneously. (Allen et al., 2014) proposed generalization of singular value decomposition (SVD) to quantify two-way regularization of PCA. This generalization involves a class of penalty functions that can be used to regularize the matrix factors. In recent years, the methodology for modelling tensor data has developed considerably with interesting applications. Hoff et al. (2011) proposed a class of multi-dimensional normal distributions by applying multi-linear transformation to an array of independently and identically distributed (i.i.d.) $N(0,1)$ items and hence studied the maximum likelihood estimator of separable complex covariance structures. Hoff (2011) discussed a model-based version of low rank decomposition.

In a previous work, Zhou et al. (2013) formulated a regression framework that considers clinical outcomes as response and images as covariates. Their method efficiently explored the spatial dependence of images in the form of a multi-dimensional array structure. By extending the generalized linear regression to a multi-way parameter corresponding to the tensor-structured predictor, they proposed a penalized likelihood approach with adaptive lasso penalties, which are imposed on the individual margins of PARAFAC decomposition. Guhaniyogi et al. (2017) later proposed a Bayesian approach using a similar setup as Zhou et al. (2013), but with a novel multi-way shrinkage prior, which can identify important cells in the tensor predictor appropriately. However, a shortcoming of both these approaches is that they unable to address the issue when responses are multi-dimensional images, and the covariates are also multi-dimensional variables (e.g., clinical data or data from another imaging modality), as in the case of multi-modal analysis and our motivating example. This necessitates the use of a tensor-on-tensor type model. The recent work of Zhang et al. (2014) presents a tensor generalized estimation equation (GEE) for longitudinal data analysis using low-rank CANDECOMP/PARAFAC (CP) decomposition on the coefficient array in GEE. This decomposition approach accommodates the longitudinal correlation of the data. Hoff (2015) proposed multi-linear regression model for longitudinal data using the least squares method. In practice, there might be the possibility that there exists an effect among the relations between the numbers of different pairs of modes. The general multi-linear regression

model can address this via separable, Kronecker-structured regression parameters along with a separable covariance structure.

A tensor-on-tensor regression approach was proposed in Lock (2018), followed by a general multiple tensor-on-tensor regression in Gahrooei et al. (2018). Furthermore, Guhaniyogi and Spencer (2018) discussed a tensor response regression where the coefficients corresponding to each vector covariate are assumed to be tensors in the Bayesian framework. Recently, Liu et al. (2020) have represented a generalized multi-linear tensor-on-tensor ridge regression model via tensor train representation. Melzer et al. (2019) proposed a joint tensor regression which is weighted at expectile levels. Their estimation technique is based on low-rank factorization combined with regularization techniques using the smooth fast iterative shrinkage-thresholding algorithm (Beck and Teboulle, 2009). An adaptive tensor-based SVD estimation is also discussed in the light of Remannian trust method in Conn et al. (2000).

A varying-coefficient model in functional data analysis (FDA) literature allows the regression coefficient to vary over some predictors of interest (say, T). In some cases, these predictors are confounded with covariates $\mathbf{X}$ or some special variables such as time. This kind of model was introduced and discussed by Hastie and Tibshirani (1993) and has since been widely studied by researchers. The non-constant relationship between functional response and predictors has been described in Fan et al. (1999); Ramsay and Silverman (2005); Ferraty and Vieu (2006); Horváth and Kokoszka (2012); Bongiorno et al. (2014); Hsing and Eubank (2015), which are some good references in FDA among many others.

The current chapter provides the following contributions to this literature. First, we propose a method of modelling image data that can efficiently process large amounts of information and identify associations while preserving the structure of the 3D images and multi-layer covariates. Second, we consider the time-varying function-on-function concurrent linear model (Hastie and Tibshirani, 1993) and generalize it to the tensor-on-tensor regression case, thus moving a step further than Lock (2018), which did not consider the time-varying coefficient. Consequently, our generalization provides an extension to classical functional concurrent regression with tensor

predictors and tensor covariates. To the best of our knowledge, such an approach has not yet been proposed in the statistics literature. Here, we express the regression coefficients using the B-spline technique, and the coefficients of the basis functions are estimated using CP-decomposition, thereby reducing computational complexity. Furthermore, our model requires minimum assumptions compared to those in the existing literature. Our approach does not require the estimation of covariance separately. Thus, our proposal offers an important addition to the literature on functional and imaging data analysis. Our methods are flexible and general; therefore, they are applicable using data from different domains such as multi-phenotype analysis and imaging genetics (Casey et al., 2010). This makes it an ideal approach for modeling multi-modal data of the type described in our motivating example.

The rest of this chapter is organized as follows. The proposed tensor-on-tensor functional regression models are described in Section 4.2. Section 4.3 provides the theoretical properties of the proposed estimator. Section 4.4 presents the algorithm and implementation of the method. The simulation results are presented in Section 4.5 and real data examples are shown in Section 4.6. Section 4.7 concludes with a discussion of future extensions. The technical proofs are presented in Section 4.8.

4.2 Tensor-on-tensor functional regression

Recall the notations and definition in matrix algebra from Section 1.3 in Chapter 1. A D-dimensional tensor is denoted by Sans-serif upper-face letters $\mathsf{A} \in \mathbb{R}^{I_1 \times \cdots \times I_D}$ where the size I_d along each mode or dimension d for $d = 1, \cdots, D$. Therefore, the number of elements in tensor A is $I = \prod_{d=1}^{D} I_d$ and the order of the tensor is the number of dimensions. Here and henceforth, matrices are denoted by bold-face capital letters (examples: $\mathbf{A}, \mathbf{B} \cdots$), vectors are written as bold-face lower-case letters (examples: $\mathbf{a}, \mathbf{b}, \cdots$) and scalars are presented as Latin alphabets $(a, b, \cdots)$.

In this section, we discuss tensor-on-tensor functional regression with time-varying coefficients. Let $\mathsf{Y}(t) \in \mathbb{R}^{Q_1 \times \cdots \times Q_M}$ with $(q_1, \cdots, q_M)$-th element $y_{q_1, \cdots, q_M}$ for all possible indices, be a set of time-varying response variables observed at time t and $\{\mathsf{Y}(t) : t \in \mathcal{T}\}$ be the underlying continuous

stochastic process defined on a compact interval $\mathscr{T}$. Without loss of generality, we assume $\mathscr{T} = [0, T], T > 0$. Suppose there are N individuals/trajectories on $\mathscr{T}$. Observations are taken at J distinct points for each individual. Collection of points for the i-th individual is denoted as $\mathscr{T}_i^\dagger = \{0 \leq t_{i1} < \cdots < t_{iJ} \leq T\}$. Therefore, for i-th individual at a set of discrete time-points $\mathscr{T}_i^\dagger$, we observe the responses $\mathsf{Y}_i(t_i) = (\mathsf{Y}_i(t_{i1}), \cdots, \mathsf{Y}_i(t_{iJ})) \in \mathbb{R}^{J \times Q_1 \times \cdots \times Q_M}$ which are distinct realizations of the corresponding stochastic process. The covariate $\mathsf{X}(t) \in \mathbb{R}^{P_1 \times \cdots \times P_L}$ with $(p_1, \cdots, p_L)$-th element $x_{p_1, \cdots, p_L}(t)$ for all indices, observed at $\mathscr{T}_i^\dagger$ is denoted as $\mathsf{X}_i(t_i) = (\mathsf{X}_i(t_{i1}), \cdots, \mathsf{X}_i(t_{iJ})) \in \mathbb{R}^{J \times P_1 \times \cdots \times P_L}$. The time-varying tensor coefficient $\boldsymbol{\beta}(t) \in \mathbb{R}^{P_1 \times \cdots \times P_L \times Q_1 \times \cdots \times Q_M}$ is assumed to vary smoothly over time. Therefore, we can apply local polynomial smoothing (Hardle, 1990; Wahba, 1990; Wand and Jones, 1995; Fan and Gijbels, 1996; Eubank, 1999), smoothing spline (Wahba, 1990; Green and Silverman, 1993; Eubank, 1999), regression spline (Eubank, 1999), P-spline Ruppert et al. (2003). In this chapter, we use B-spline bases which are very popular in mathematics, computer science and statistics (De Boor et al., 1978). Suppose, $\{\tau_h\}_{h=1}^{K_N}$ be K_N interior knots within the compact interval $[0, T]$ and the partition of the interval $[0, T]$ at these knots be denoted as $\mathscr{K} = \left\{0 = \tau_0 < \tau_1 < \cdots < \tau_{K_N} < \tau_{K_N+1} = T\right\}$. The polynomial spline of order $v+1$ is a function of polynomials with degree v on the intervals $[\tau_{h-1}, \tau_h)$ for $h = 1, \cdots, K_N$ and $\left[\tau_{K_N}, \tau_{K_N+1}\right]$ and $v-1$ continuous derivatives globally. Let $\mathcal{S}_{K_N}^v(t)$ denotes a set of such spline functions, i.e., $s(t)$ belongs to $\mathcal{S}_{K_N}^v(t)$ if and only if $s(t)$ belongs to $C^{v-1}[0, T]$ and its restriction to each intervals $[\tau_{h-1}, \tau_h)$ is a polynomial of degree atleast v. Define for $h = 1, \cdots, H_N$

$$\mathbb{B}_h(t) = (\tau_h - \tau_{h-v-1})[\tau_{h-v-1}, \cdots, \tau_h](z-t)_+^v \tag{4.1}$$

where $H := H_N = K_N + v + 1$, $[\tau_{h-v-1}, \cdots, \tau_h]f$ denotes the $(v+1)$-st order divided difference of the function f and $\tau_h = \tau_0$ for $h = -v, \cdots, -1$ and $\tau_h = \tau_{K_N+1}$ for $h = K_N + 2, \cdots, H_N$. Therefore, $\{\mathbb{B}_h\}_{h=1}^{H_N}$ forms the basis for $\mathcal{S}_{K_N}^v(t)$ (Schumaker, 2007).

Now, for $1 \leq p_l \leq P_l, 1 \leq q_m \leq Q_m, 1 \leq l \leq L, 1 \leq m \leq M$, each function $\beta_{p_1, \cdots, p_L, q_1, \cdots, q_M}(t)$ can be approximated by

$$\beta_{p_1, \cdots, p_L, q_1, \cdots, q_M}(t) = \sum_{h=1}^{H} b_{h, p_1, \cdots, p_L, q_1, \cdots, q_M} \mathbb{B}_h(t) = \mathbf{b}_{p_1, \cdots, p_L, q_1, \cdots, q_M}^{\mathrm{T}} \mathscr{B}(t) \tag{4.2}$$

where $\mathbf{b}_{p_1,\cdots,p_L,q_1,\cdots,q_M} = (b_{1,p_1,\cdots,p_L,q_1,\cdots,q_M},\cdots,b_{H,p_1,\cdots,p_L,q_1,\cdots,q_M})^{\mathrm{T}}$ is the collection of basis coefficients and $\mathscr{B}(t) = (\mathbb{B}_1(t),\cdots,\mathbb{B}_H(t))^{\mathrm{T}}$ is a vector of known B-spline bases.

In practice, we can use different basis functions in mode to approximate $\beta_{p_1,\cdots,p_L,q_1,\cdots,q_M}(t)$. However, for convenience, we use the same set of bases in this chapter. Instead of the B-spline, one can use other basis functions to approximate the coefficient functions. We use the B-spline base for its simplicity and numerical tractability. Although this method does not produce a desirable approximation for discontinuous functions, in this chapter we restrict ourselves to smooth continuous coefficients.

We propose a general time-varying tensor-on-tensor regression model,

$$\mathsf{Y}_i(t) = \langle \mathsf{X}_i(t), \boldsymbol{\beta}(t)\rangle_L + \mathsf{E}_i(t) \tag{4.3}$$

which can be reduced into the following mode-wise time-varying coefficient model.

$$y_{i,q_1,\cdots,q_M}(t) = \sum_{p_1=1}^{P_1}\cdots\sum_{p_L=1}^{P_L} x_{i,p_1,\cdots,p_L}(t)\beta_{p_1,\cdots,p_L,q_1,\cdots,q_M}(t) + \epsilon_{i,q_1,\cdots,q_M}(t) \tag{4.4}$$

where $\epsilon_{i,q_1,\cdots,q_M}(t)$ is a random error with mean zero. Errors can be correlated over time and modes, but are independent over the trajectories. After plugging-in the approximate expression of $\beta_{p_1,\cdots,p_L,q_1,\cdots,q_M}(t)$ at each mode, the model now boils down to

$$y_{i,q_1,\cdots,q_M}(t) = \sum_{p_1=1}^{P_1}\cdots\sum_{p_L=1}^{P_L}\sum_{h=1}^{H} b_{h,p_1,\cdots,p_L,q_1,\cdots,q_M} x_{i,p_1,\cdots,p_L}(t)\mathbb{B}_h(t) + \epsilon_{i,q_1,\cdots,q_M}(t) \tag{4.5}$$

The multi-dimensional basis coefficients $\mathsf{B}_0 = \left\{b_{h,p_1,\cdots,p_L,q_1,\cdots,q_M} : 1 \le h \le H,\ 1 \le p_l \le P_l,\ 1 \le q_m \le Q_m,\ 1 \le l \le L,\ 1 \le m \le M\right\}$ can be estimated by minimizing the mode-wise penalized integrated sum of square errors with respect to B_0. Let us denote the smoothness penalty by Ω_{sm} where

$$\begin{aligned}&\Omega_{sm}(\mathsf{B}_0)\\ &= \sum_{p_1=1}^{P_1}\cdots\sum_{p_L=1}^{P_L}\sum_{q_1=1}^{Q_1}\cdots\sum_{q_M=1}^{Q_M}\int \theta_{p_1,\cdots,p_L,q_1,\cdots,q_M}\left\{\beta''_{p_1,\cdots,p_L,q_1,\cdots,q_M}(t)\right\}^2 dt\end{aligned}$$

$$= \sum_{p_1=1}^{P_1} \cdots \sum_{p_L=1}^{P_L} \sum_{q_1=1}^{Q_1} \cdots \sum_{q_M=1}^{Q_M} \theta_{p_1,\cdots,p_L,q_1,\cdots,q_M} \mathbf{b}^{\mathrm{T}}_{p_1,\cdots,p_L,q_1,\cdots,q_M} \int \mathbf{B}''(t)\mathbf{B}''(t)^{\mathrm{T}} dt \mathbf{b}_{p_1,\cdots,p_L,q_1,\cdots,q_M} \tag{4.6}$$

Hence, the loss function turns out to be

$$\begin{aligned}
&\mathscr{L}(\mathsf{B}_0) \\
&= \frac{1}{N} \int_{\mathscr{T}} \sum_{i=1}^{N} \sum_{q_1=1}^{Q_1} \cdots \sum_{q_M=1}^{Q_M} \left(y_{i,q_1,\cdots,q_M}(t) - \sum_{p_1=1}^{P_1} \cdots \sum_{p_L=1}^{P_L} \sum_{h=1}^{H} b_{h,p_1,\cdots,p_L,q_1,\cdots,q_M} x_{i,p_1,\cdots,p_L}(t) B_h(t) \right)^2 dt \\
&\quad + \Omega_{sm}(\mathsf{B}_0)
\end{aligned} \tag{4.7}$$

In Equation (4.6), $\{\theta_{p_1,\cdots,p_L,q_1,\cdots,q_M}\}_{p_1,\cdots,p_L,q_1,\cdots,q_M}$ are the tuning parameters for smoothness. Penalty due to smoothness is widespread in the literature of functional data analysis (Ramsay and Silverman (2005) among many others). In practice, it is unrealistic to find these large numbers of pre-assigned tuning parameters. By considering $\theta_{p_1,\cdots,p_L,q_1,\cdots,q_M} = \theta$, for all possible $p_1, \cdots, p_L, q_1, \cdots, q_M$, the simplest version of smoothness penalty would be,

$$\Omega_{sm}(\mathsf{B}_0) = \theta \, \mathrm{vec}(\mathsf{B}_0)^{\mathrm{T}} (\mathbf{I}_Q \otimes \mathbf{I}_P \otimes \int \mathbf{B}''(t)\mathbf{B}''(t)^{\mathrm{T}} dt) \, \mathrm{vec}(\mathsf{B}_0) \tag{4.8}$$

Note that, $\mathrm{vec}(\mathsf{B}_0) = (\mathbf{b}_{11}, \cdots, \mathbf{b}_{P1}, \mathbf{b}_{12}, \cdots, \mathbf{b}_{P2}, \cdots, \mathbf{b}_{1Q}, \cdots, \mathbf{b}_{PQ})^{\mathrm{T}}$. Therefore, the penalized likelihood estimating equation for functional tensor-on-tensor regression problem is

$$\mathscr{L}(\mathsf{B}_0) = \int_{\mathscr{T}} \frac{1}{N} \sum_{i=1}^{N} \|\mathsf{Y}_i(t) - \langle \mathsf{Z}_i(t), \mathsf{B}_0 \rangle_{L+1}\|^2_{\mathscr{F}} \, dt + \Omega_{sm}(\mathsf{B}_0) \tag{4.9}$$

where $\langle \cdot, \cdot \rangle_{L+1}$ is the contracted tensor product defined in Section 1.3.1 and $\| \cdot \|_{\mathscr{F}}$ is the Frobenius norm. The first term of the Equation (4.9) is integrated sum of squares and the second term is the smoothness penalty.

Let the response tensor for time t, $\mathsf{Y}(t) \in \mathbb{R}^{N \times Q_1 \times \cdots \times Q_M}$ with its $(i, q_1, \cdots, q_M)$-th element be $y_{i,q_1,\cdots,q_M}(t)$ for all $i = 1, \cdots, N; q_m = 1, \cdots, Q_m; m = 1, \cdots, M$. Similarly, we define an updated covariate tensor contaminated with B-spline bases $\mathsf{Z}(t) \in \mathbb{R}^{N \times H \times P_1 \times \cdots \times P_L}$ where $(i, h, p_1, \cdots, p_L)$-th element of the tensor is defined as $z_{i,h,p_1,\cdots,p_L}(t) = x_{i,p_1,\cdots,p_L}(t)\mathbb{B}_h(t)$. Therefore, the corresponding penalized loss function in Equation (4.9) is equivalent to

$$\mathscr{L}(\mathsf{B}_0) = \int_{\mathscr{T}} \|\mathsf{Y}(t) - \langle \mathsf{Z}(t), \mathsf{B}_0 \rangle_{L+1}\|^2_{\mathscr{F}} \, dt + \Omega_{sm}(\mathsf{B}_0) \tag{4.10}$$

Remark 4.2.1. *For $Q = 0$, the proposed model reduces to the classical concurrent linear model Ramsay and Silverman (2005). For $Q = 1$ and $P = 1$, the time-varying network model (Xue et al., 2018) is a special case of our proposed model for a specific choice of covariates. For $Q = 2$, $y_{i,q_1,q_2}(t)$ is the observation of the quantity of interest at time t for sub-unit q_2 from unit q_1 of a treatment group i in a hierarchical model (Zhou et al., 2010).*

Let $P = \prod_{l=1}^{L} P_l$ be the total number of predictors for each observation and $Q = \prod_{m=1}^{M} Q_m$ be the total number of outcomes for each predictor over time. To minimize the penalized integrated sum of squared residuals described in Equation (4.10), the solution for B_0 could be inconsistent. Since the unknown coefficient tensor B_0 has $H \prod_{l=1}^{L} P_l \prod_{m=1}^{M} Q_m$ many parameters, we need to adopt the dimension reduction technique. Inspired by the novel idea discussed in Lock (2018), we consider rank R decomposition of B_0 as $\mathsf{B}_0 = [[\mathbf{U}_0, \mathbf{U}_1, \cdots, \mathbf{U}_L, \mathbf{V}_1, \cdots, \mathbf{V}_M]]$ where $\mathbf{U}_0$, $\mathbf{U}_l$ and $\mathbf{V}_m$ are matrices with dimensions $H \times R$, $P_l \times R$ and $Q_m \times R$ respectively for all $1 \leq l \leq L, 1 \leq m \leq M$. After dimension reduction, the number of unknown parameters reduces to $R(H + \sum_{l=1}^{L} P_l + \sum_{m=1}^{M} Q_m)$. Therefore, the estimate of the coefficient tensor is as follows.

$$\widetilde{\mathsf{B}}_0 = \arg \min_{\text{rank}(\mathsf{B}_0) \leq R} \mathscr{L}(\mathsf{B}_0) \tag{4.11}$$

However, this estimated coefficient tensor suffers from over-fitting and instability problems due to multi-collinearity of Z and/or the large number of observed outcomes. Thus, we obtain an alternative estimate of the coefficient tensor B_0

$$\widehat{\mathsf{B}}_0 = \arg \min_{\text{rank}(\mathsf{B}_0) \leq R} \mathscr{Q}(\mathsf{B}_0) \tag{4.12}$$

which is based on the modified loss function, $\mathscr{Q}$, defined by

$$\mathscr{Q}(\mathbf{B}_0) = \frac{1}{N} \int_{\mathscr{T}} \|\mathsf{Y}(t) - \langle \mathsf{Z}(t), \mathsf{B}_0 \rangle_{L+1}\|_{\mathscr{F}}^2 \, dt + \Omega(\mathsf{B}_0) \tag{4.13}$$

where

$$\Omega(\mathsf{B}_0) = \theta \, \text{vec}(\mathsf{B}_0)^{\mathrm{T}} (\mathbf{I}_Q \otimes \mathbf{I}_P \otimes \int \mathbf{B}''(t) \mathbf{B}''(t)^{\mathrm{T}} dt) \, \text{vec}(\mathsf{B}_0) + \phi \, \text{vec}(\mathsf{B}_0)^{\mathrm{T}} \, \text{vec}(\mathsf{B}_0) \tag{4.14}$$

Equation (4.14) suggests the penalization of the smoothness and sparsity of the coefficient functions simultaneously.

Remark 4.2.2. *Tuning parameter selection: The number of knots and tuning parameters θ and ϕ are unknown and we need to select it using Mallows's C_p (Mallows, 1973), generalized cross-validation (Craven and Wahba, 1978) or leave-one-out cross validation method (Stone, 1974). Also, the selection of rank is a separate problem altogether. A rank selection method can be proposed based on Chen et al. (2013).*

4.3 Asymptotic properties

In this section, we will study identifiability of the model and consistency of parameter estimates under the proposed model as the number of subjects N goes to infinity while we assume that the rank of the basis tensor coefficient is known and fixed.

4.3.1 Identifiability

Identifiability issue plays important roles in tensor regression (Lock, 2018; Zhou et al., 2013; Guhaniyogi et al., 2017). The model discussed in Section 4.2 would be identifiable for $\boldsymbol{\beta}(t)$, if for $\boldsymbol{\beta}(t) \neq \boldsymbol{\beta}^*(t)$ implies $\langle \mathsf{X}(t), \boldsymbol{\beta}(t) \rangle_L \neq \langle \mathsf{X}(t), \boldsymbol{\beta}^*(t) \rangle_L$ for some $t \in \mathscr{T}$ and some $\mathsf{X}(t) \in \mathbb{R}^{P_1 \times \cdots \times P_L}$. Using the basis expansion in Equation (4.2), we can say that B_0 is identifiable if and only if $\boldsymbol{\beta}(t)$ is identifiable for all $t \in \mathscr{T}$. Therefore, the reduced model is identifiable if for $\mathsf{B}_0 \neq \mathsf{B}_0^*$ implies $\langle \mathsf{Z}(t), \mathsf{B}_0 \rangle_{L+1} = \left\langle \mathsf{Z}(t), \mathsf{B}_0^* \right\rangle_{L+1}$ for some $t \in \mathscr{T}$ and for some $\mathsf{Z}(t) \in \mathbb{R}^{H \times P_1 \times \cdots \times P_L}$. Assume that, for $t = t_0$, $\mathsf{Z}_{h,p_{k_1},\cdots,p_{k_L}}(t_0) = 1$ at $k_1 = 1, \cdots, k_L = L$ and 0 otherwise, then the product becomes $b_{h,p_1,\cdots,p_L,q_1,\cdots,q_M}$. Furthermore, $\mathbf{U}_0, \mathbf{U}_1, \cdots, \mathbf{U}_L, \mathbf{V}_1, \cdots, \mathbf{V}_M$ in the expression of CP-decomposition are not identifiable. Therefore, The identifiability conditions can be imposed in the following way (Sidiropoulos and Bro, 2000).

1. Restrictions for scale and non-uniqueness: B_0 will remain same after replacing $\mathbf{U}_0$, $\mathbf{U}_l$ and $\mathbf{V}_m$ by $c_s\mathbf{U}_0$, $c_{u_l}\mathbf{U}_l$ and $c_{v_m}\mathbf{V}_m$ respectively, where $\{c_s, c_{u_l}, c_{v_m}\}$ is the set of constants with $c_s \prod_{l=1}^{L} c_{u_l} \prod_{l=1}^{M} c_{v_l} = 1$. This problem can be solved by introducing the condition that the norm of each of $\mathbf{u}_{r0}$, $\mathbf{u}_{rl}$ and $\mathbf{v}_{rm}$ is set to be 1, $1 \leq r \leq R, 1 \leq l \leq L, 1 \leq m \leq M$.

2. Restriction for permutation: For any permutation $\pi(\cdot)$ of $\{1, \cdots, R\}$, $\sum_{r=1}^{R} \mathbf{u}_{r0} \circ \mathbf{u}_{r1} \circ \cdots \circ \mathbf{u}_{rL} \circ \mathbf{v}_{r1} \circ \cdots \circ \mathbf{v}_{rM}$ is same as $\sum_{r=1}^{R} \mathbf{u}_{\pi(r)0} \circ \mathbf{u}_{\pi(r)1} \circ \cdots \circ \mathbf{u}_{\pi(r)L} \circ \mathbf{v}_{\pi(r)1} \circ \cdots \circ \mathbf{v}_{\pi(r)M}$. Therefore, we impose the restriction $\|\mathbf{u}_{01}\| \geq \cdots \geq \|\mathbf{u}_{0R}\|$.

These indeterminacies are enough to ensure the identifiability for $L + M \geq 2$. Therefore, we do not need the additional orthogonality condition that appeared in Lock (2018); Zhou et al. (2013); Guhaniyogi et al. (2017).

4.3.2 Convergence rate

In this subsection, we study the asymptotic properties of the estimate of the time-varying tensor regression parameter $\boldsymbol{\beta}(t)$ based on polynomial spline approximation and the CP decomposition. To proceed further, we introduce some regularity conditions which are required to establish the asymptotic properties.

(C1) Without loss of generality, assume $\mathscr{T} = [0, 1]$. The observation times t_{ij} for $i = 1, \cdots, N; j = 1, \cdots, J$ are independent and follow a distribution $f_T(t)$ over the support $\mathscr{T}$. the density function $f_T(t)$ is assumed to be absolutely continuous and bounded by a nonzero and finite constant.

(C2) $\{\tau_h\}_{h=1}^{K_n}$ be K_n interior knots within the compact interval $\mathscr{K} = [0, 1]$ and the partition of the interval $[0, T]$ with K_N knots can be denoted as $\mathbb{I} = \left\{0 = \tau_0 < \tau_1 < \cdots < \tau_{K_N} < \tau_{K_N+1} = 1\right\}$.

(C3) The polynomial spline of order $v + 1$ are the function with degree v of polynomials on the interval $[\tau_{h-1}, \tau_h)$ for $h = 1, \cdots, K$ and $\left[\tau_{K_N}, \tau_{K_N+1}\right]$ and $v - 1$ continuous derivatives globally.

(C4) For $t \in \mathscr{T}$, $\epsilon_{i,q_1,\cdots,q_M}(t)$'s are i.i.d. copies with mean zero and finite second order moment over i. Moreover, for each i and the coordinates $q_1, \cdots, q_M$, $\epsilon_{i,q_1,\cdots,q_M}(t_{ij})$ are locally stationary time series of the form given in appendix. Assume physical dependence measure $\Delta(k, a)$ is upper bounded by $k^{-\kappa_0}$ for some positive κ_0 and for all $j \geq 1$.

(C5) The covariate $x_{i,p_1,\cdots,p_L}(t)$'s are i.i.d. for index i and it is bounded almost everywhere.

Remark 4.3.1. *Conditions (C1), (C2), (C3) are standard conditions in the context of polynomial spline regression and require the consistency of spline estimation of varying-coefficient models. Condition (C3) provides the degree of smoothness on the time-varying coefficients. We assume Condition (C4) to represent a wide class of stationary, locally stationary, and non-linear processes. Similar conditions can be found in Ding and Zhou (2020); Ding et al. (2021). This is a natural assumption of temporal short-range dependent process where temporal correlation decays in polynomial order. This phenomenon can also be observed in well-known Ornstein–Uhlenbeck process and the linear process with the standard basis expansion* $\epsilon_{i,\bullet}(t) = \sum_{k=1}^{\infty} a_{ik,\bullet}\phi_k(t)$ *where* $a_{ik,\bullet}$ *is an uncorrelated mean zero, finite variance random variables over* (i,k) *and* $\sup_t \phi_k(t) \leq Ck^{-a}$ *for some positive constants* C *and* a *.*

Since the number of modes is fixed, we reduce the objective function by following the notation $\mathbf{Y} \in \mathbb{R}^{NJ\times Q}$ and $\mathsf{Z} \in \mathbb{R}^{NJ\times H_N\times P}$

$$\mathcal{Q}(\mathbf{B}_0) = \frac{1}{NJ}\|\mathbf{Y} - \langle \mathsf{Z}, \mathsf{B}_0\rangle_2\|_{\mathscr{F}}^2 + \|\mathsf{B}_0\|_{\mathscr{F},\mathbf{W}_\omega}^2 \tag{4.15}$$

where $\|\mathsf{B}_0\|_{\mathscr{F},\mathbf{W}_\omega}$ be the weighted Frobenius norm, defined as $\|\mathsf{B}_0\|_{\mathscr{F},\mathbf{W}_\omega} = \sqrt{\mathrm{vec}(\mathsf{B}_0)^{\mathrm{T}}\mathbf{W}_\omega\,\mathrm{vec}(\mathbf{B}_0)}$ where ω is a set of all tuning parameters. Moreover, assume that $\mathrm{rank}(\mathsf{B}_0) = R_0$ which is assumed to be known and fixed. Further assume that

(C6) $\lambda_{\min}\left(\mathsf{Z}_{(1)}^{\mathrm{T}}\mathbf{Z}_{(1)}\right) = \sigma_{\min}(\mathsf{Z}_{(1)})^2 \geq \lambda_{\min}(\mathbb{B}^{\mathrm{T}}\mathbb{B})\lambda_{\min}(\mathbf{X}^{\mathrm{T}}\mathbf{X}) > \lambda$

where $\lambda_i(\mathbf{A})$ and $\sigma_i(\mathbf{A})$ denotes i-th eigen-value and singular value respective for a matrix $\mathbf{A}$. Define, the constants $\mathscr{C}(\delta) = 1+2/\delta$ such that $\mathscr{C}(\delta) \leq \lambda^2/2\mu$ where $\mu = (NJ)(\theta\lambda_{\max}(\int \mathbf{B}''(t)\mathbf{B}''(t)^{\mathrm{T}}dt) + \phi)\sqrt{2R_0}$. Further, define $\xi = \sup_{1\leq h\leq H}\sup_{t\in[0,1]}|\mathbb{B}_h(t)|$ which is typically bounded. Additionally, define $\sigma_1(\mathsf{C}) = \max\{\sigma_1(\mathsf{C}_{(1)}), \sigma_1(\mathsf{C}_{(2)}), \sigma_1(\mathsf{C}_{(3)})\}$. Therefore, we propose the following theorem for the estimation and prediction performance of the coefficient tensor.

Theorem 4.3.1. *Under assumptions (C4) and (C6), when both the number of time-points and trajectories are large enough, there exists a constant* C_a*, with probability atleast* $1 - C_aN^{-a\tau}$*, such*

that we have the following.

$$\| \left\langle \mathsf{Z}, (\widehat{\mathsf{B}}_0 - \mathsf{B}_0) \right\rangle \|_{\mathscr{F}}^2 \leq \lambda^{-1} \left(\mathscr{C}(\delta)^{-1} - 2\mu\lambda^{-2} \right)^{-1} \left\{ 4\mu\sigma_1^2(\mathsf{C}) + 2R_0(1+\delta)Q^2\xi^2 N^{2\tau+2} J \right\} \quad (4.16)$$

for any $H_N \times P \times Q$ matrix C with rank$(\mathsf{C}) \leq R_0$, By choosing $\mathsf{C} = \mathsf{B}_0$, a simplified prediction error could be obtained. Under the same set of assumptions, the estimation error of the matrix B_0 is

$$\|\widehat{\mathbf{B}}_0 - \mathbf{B}_0\|_{\mathscr{F}}^2 \leq \lambda^{-1} \left(\mathscr{C}(\delta)^{-1} - 2\mu\lambda^{-2} \right)^{-1} \left\{ 4\mu\sigma_1^2(\mathsf{C}) + 2R_0(1+\delta)Q^2\xi^2 N^{2\tau+2} J \right\} \quad (4.17)$$

Additionally, we introduce the following theorem which states the consistency result for the coefficient tensor function.

Theorem 4.3.2. *Under assumptions (C1)-(C6), with probability, we have the following with probability $1 - C_a N^{-a\tau}$,*

$$\int_{\mathscr{T}} |\widehat{\beta}_\bullet(t) - \beta_\bullet(t)|^2 f_T(t) dt = O\left\{ \lambda^{-1} \left(\mathscr{C}(\delta)^{-1} - 2\mu\lambda^{-2} \right)^{-1} \left\{ 4\mu\sigma_1^2(\mathsf{C}_{(1)}) + 2R_0(1+\delta)Q\xi N^{\tau+1}\sqrt{J} \right\} + K_N^{-2(v+1)} \right\}$$

4.4 Algorithm and implementation

In this section, we propose a general algorithm to estimate the basis coefficient tensor using the objective function described in Section 4.2. For given time-points $t_1, \cdots, t_J$, define $\mathscr{X}$ and $\mathscr{Y}$ as the combined tensor after staking over all time-points. Therefore, $\mathscr{X}$ and $\mathscr{Y}$ are the tensors of order $NJ \times H \times P_1 \times \cdots P_L \times Q_1 \times \cdots Q_M$ and $NJ \times Q_1 \times \cdots Q_M$ respectively. Moreover, define $\check{\mathbf{B}}_0$ as the matrix of coefficient of order $HP \times Q$, where columns and rows of B_0 are obtained by vectorizing first $(L+1)$ and last M modes of B_0 respectively. For the alternative expression of the penalty term in Equation (4.13), observe the following.

1. $\|\mathsf{B}_0\|^2 = \|\check{\mathbf{B}}_0\|^2 = \text{vec}(\mathsf{B}_0)^{\mathrm{T}} \text{vec}(\mathsf{B}_0) = \text{trace}(\check{\mathbf{B}}_0 \check{\mathbf{B}}_0^{\mathrm{T}})$, where trace$(\mathbf{A})$ denotes the trace of a square matrix $\mathbf{A}$.
2. $\left[\mathbf{I}_Q \otimes \mathbf{I}_P \otimes \left(\theta \int \mathbf{B}''(t)\mathbf{B}''(t)^{\mathrm{T}} dt + \phi \mathbf{I}_H \right)^{1/2} \right] \text{vec}(\mathsf{B}_0)$

$$= \text{vec}\left((\mathbf{I}_P \otimes \left(\theta \int \mathbf{B}''(t)\mathbf{B}''(t)^{\mathrm{T}} dt + \phi \mathbf{I}_H\right)^{1/2}) \breve{\mathbf{B}}_0 \mathbf{I}_Q\right)$$

Therefore, equivalently, the optimization problem reduces to an unregulated least squares problem with modified predictor and outcome variables for estimating B_0.

$$\widehat{\mathsf{B}}_0 = \arg \min_{\text{rank}(\mathsf{B}_0) \leq R} \frac{1}{NJ} \int_{\mathscr{T}} \|\widetilde{\mathscr{Y}} - \langle \widetilde{\mathscr{X}}, \mathsf{B}_0 \rangle_{L+1}\|^2 dt \tag{4.18}$$

where $\widetilde{\mathscr{X}} \in \mathbb{R}^{(NJ+HP)\times H \times P_1 \times \cdots \times P_L \times Q_1 \times \cdots Q_M}$ be the contamination of $\mathsf{Z}(t)$ along with smoothing term and sparsity. $\widetilde{\mathscr{Y}} \in \mathbb{R}^{(NJ+HP)\times Q_1 \times \cdots \times Q_M}$ which is a contamination of $\mathsf{Y}(t)$ and zero tensor function. The unfolding of $\widetilde{\mathscr{X}}$ and $\widetilde{\mathscr{Y}}$ along the first dimension produce the following matrices:

$$\widetilde{\mathscr{X}}_{(1)} = \begin{bmatrix} \mathscr{X}_{(1)} \\ (\mathbf{I}_P \otimes \left(\theta \int \mathbf{B}''(t)\mathbf{B}''(t)^{\mathrm{T}} dt + \phi \mathbf{I}_H\right)^{1/2}) \end{bmatrix} \quad \text{and} \quad \widetilde{\mathscr{Y}}_{(1)} = \begin{bmatrix} \mathscr{Y}_{(1)} \\ 0_{HP\times Q} \end{bmatrix} \tag{4.19}$$

Thus, applying the following Algorithm 4.1, we obtain an estimate of coefficient tensor for the known rank of the coefficient array and hence the coefficient function $\boldsymbol{\beta}(t)$.

The above algorithm is similar to function "*rrr*" available in the package *MultiwayRegression* in R. The selection of the adjustment parameters θ, ϕ and the rank R of the coefficient tensor is crucial. It can be done using integrated predictive accuracy in a training and test set. K-fold cross-validation can be used to obtain these tuning parameters; however, it is computationally expensive. Fortunately, our estimate is robust for the selection of θ and ϕ. The rank of CP decomposition of the coefficient tensor is the number of rank-1 terms that are necessary to represent the coefficient tensor. For large R, every B_0 can be represented by the CP decomposition. Therefore, it determines the complexity of the model. We leave the optimal determination of the rank for future research.

4.5 Simulation studies

In this section, we conduct numerical studies to compare the finite sample performance to estimate four-way time-varying tensor coefficient $\boldsymbol{\beta}(t)$. Data are generated from the following model for each mode p_1, p_2, q_1, q_2

$$y_{i,q_1,q_2}(t) = \sum_{p_1=1}^{P_1} \sum_{p_2=1}^{P_2} x_{i,p_1,p_2}(t) \beta_{p_1,p_2,q_1,q_2}(t) + \epsilon_{i,q_1,q_2}(t),\ i = 1, \cdots, N; t \in [0,1] \tag{4.20}$$

Algorithm 4.1 Estimation of $\boldsymbol{\beta}(t) : t \in [0,T]$ for tensor based function-on-function regression method.

Data: $\mathsf{X}(t)$, $\mathsf{Y}(t)$ for $t \in [0,T]\,, T > 0$ observed on a grid in $[0,T]$
Result: Estimate $\boldsymbol{\beta}(s)$ using proposed method
Tuning parameters: $\{\theta, \phi\}$, rank $R \in \mathbb{N}$, number of knots K_N, a vector of known B-spline bases $\mathscr{B}(t) = (\mathbb{B}_1(t), \cdots, \mathbb{B}_H(t))^{\mathrm{T}}$
Stopping parameter: $\epsilon_0 > 0$
Create: $\mathscr{X}$ and $\mathscr{Y}$ as mentioned in Equation (4.19)
Initialize: $\mathbf{U}_0, \mathbf{U}_1, \cdots, \mathbf{U}_L, \mathbf{V}_1, \cdots, \mathbf{V}_M$ be randomly chosen matrices of specific order
while *Error* $> \epsilon_0$ **do**
  for $l \leftarrow 1$ *to* $\#\{H, P_1, \cdots, P_L\}$ **do**
    Set $d^{(l)}$ be the l-th entry of $\{H, P_1, \cdots, P_L\}$
    for $r = 1, \cdots, R$ **do**
      $\mathsf{C}_r \leftarrow \langle \widetilde{\mathscr{X}}, \mathbf{u}_{r0} \circ \cdots \circ \mathbf{u}_{r,k-1} \circ \mathbf{u}_{r,k+1} \circ \cdots \circ \mathbf{u}_{rL} \circ \mathbf{v}_{r1} \circ \cdots \circ \mathbf{v}_{rM} \rangle_L$ which is a tensor of dimension $(NJ + HP) \times d^{(l)} \times Q_1 \times \cdots \times Q_M$
      Unfolding C_r along with dimension corresponding to $d^{(l)}$
      Obtain a $(NJ + HP)Q \times d^{(l)}$ dimension matrix $\mathbf{C}_r$
    end
    $\mathbf{C} \leftarrow [\mathbf{C}_1, \cdots, \mathbf{C}_R] \in \mathbb{R}^{(NJ+HP)Q \times Rd^{(l)}}$
    $\mathrm{vec}(\mathbf{U}_l) \leftarrow (\mathbf{C}^{\mathrm{T}}\mathbf{C})^{-1}\mathbf{C}^{\mathrm{T}}\,\mathrm{vec}(\widetilde{\mathscr{Y}})$
  end
  for $m \leftarrow 1$ *to* $\#\{Q_1, \cdots, Q_M\}$ **do**
    Set $d^{(m)}$ be the m-th entry of $\{Q_1, \cdots, Q_L\}$
    $\widetilde{\mathscr{Y}}_{d^{(m)}}$ is unfolded along the mode corresponding to $d^{(m)}$ and obtain a $d^{(m)} \times (NJ + HP) \prod_{m \neq k} Q_m$
    for $r = 1, \cdots, R$ **do**
      $D_r \leftarrow \mathrm{vec}(\langle \widetilde{\mathscr{X}}, \mathbf{u}_{r0} \circ \mathbf{u}_{r1} \circ \cdots \circ \mathbf{u}_{rL} \circ \mathbf{v}_{r1} \circ \cdots \mathbf{v}_{r,k-1} \circ \mathbf{v}_{r,k+1} \circ \cdots \circ \mathbf{v}_{rM} \rangle_{L+1})$
    end
    $\mathbf{D} \leftarrow [D_1, \cdots, D_R] \in \mathbb{R}^{(NJ+HP)\prod_{m \neq k} Q_m \times R}$
    $\mathbf{V}_m \leftarrow \widetilde{\mathscr{Y}}_{d^{(m)}} \mathbf{D}(\mathbf{D}^{\mathrm{T}}\mathbf{D})^{-1}$
  end
  Compute $\mathbf{B} = [[\mathbf{U}_0, \mathbf{U}_1, \cdots, \mathbf{U}_L, \mathbf{V}_1, \cdots, \mathbf{V}_M]]$
  Calculate: Error $= \frac{\|\widetilde{\mathscr{Y}} - \left\langle \widetilde{\mathscr{X}}, \widehat{\mathbf{B}} \right\rangle_{L+1}\|_{\mathscr{F}}^2}{\|\widehat{\mathscr{Y}}\|_{\mathscr{F}}^2}$
end
Compute $\beta_{p_1,\cdots,p_L,q_1,\cdots,q_M}(t) = \mathbf{b}^{\mathrm{T}}_{p_1,\cdots,p_L,q_1,\cdots,q_M} \mathscr{B}(t)$ using Equation (4.2) for each node

Regression functions are given by

$$\beta_{p_1,p_2,q_1,q_2}(t) = p_1 \cos(2\pi t) + q_1 \sin(2\pi t) + p_2 \sin(4\pi t) + q_2 \cos(4\pi t) \tag{4.21}$$

Here, changes in one unit of the index of each mode produce a change in one unit of the coefficient when the time is fixed. The covarites are generated in the following way,

$$x_{i,p_1,p_2}(t) = \chi^{(1)}_{i,p_1,p_2} + \chi^{(2)}_{i,p_1,p_2} \sin(\pi t) + \chi^{(3)}_{i,p_1,p_2} \cos(\pi t) \tag{4.22}$$

and errors are generated as follows.

$$\epsilon_{i,q_1,q_2}(t) = \eta^{(1)}_{i,q_1,q_2} \sqrt{2} \cos(\pi t) + \eta^{(2)}_{i,q_1,q_2} \sqrt{2} \sin(\pi t)$$

for all $p_1 = 1, \cdots, P_1, p_2 = 1, \cdots, P_2, q_1 = 1, \cdots, Q_1$ and $q_2 = 1, \cdots, Q_2$. Moreover, we assume that $x_{i,p_1,p_2}(t)$ are observed with measurement error, i.e., $u_{i,p_1,p_2}(t) = x_{i,p_1,p_2} + \delta_{p_1,p_2}$ where $\delta_{p_1,p_2} \sim N(0, 0.6^2)$. Assume that the set of random variables $\{\chi^{(l)}_{i,p_1,p_2} : l = 1, 2, 3\}$ and $\{\eta^{(l)}_{i,q_1,q_2} : l = 1, 2\}$ are mutually independent. The data generation process is influenced by Kim et al. (2018) although in an entirely different situation. We observe the data at 81 equidistant time-points in $[0, 1]$ with $t_j = (j - 0.5)/J$ for all $j = 1, \cdots, J$. We also fix $P_1 \times P_2 = 5 \times 2$ and $Q_1 \times Q_2$ as either 5×2 or 15×12. Set, number of subjects, $N \in \{30, 100\}$. We consider the following scenarios.

(Situation1) We choose $\chi^{(1)}_{i,p_1,p_2} \sim N(0, 1^2)$, $\chi^{(2)}_{i,p_1,p_2} \sim N(0, 0.85^2)$, $\chi^{(3)}_{i,p_1,p_2} \sim N(0, 0.7^2)$ and they are mutually independent. $\eta^{(1)}_{i,q_1,q_2} \sim N(0, 2^2)$, $\eta^{(2)}_{i,q_1,q_2} \sim N(0, 0.75^2)$ and they are mutually independent. Here, the covariates do not depend on the modes of the data structure.

(Situation2) In addition, with the assumption of the coefficients of covariates, impose the spatial correlation structure to address the mode-wise dependencies. We consider the following two cases.

a) $\chi^{(l)}_{i,p_1,p_2}$ at mode (p_1, p_2) is $\rho_s(\mathrm{ED}_{p_1,p_2}; \theta)$, where ρ_s is the exponential correlation function, ED_{p_1,p_2} is defined as scaled Euclidean distance between two modes, having scaled by a constant θ, therefore, θ defines an isotropic covariance function. In this simulation setup, θ is taken as 8.

b) $\chi^{(l)}_{i,p_1,p_2}$ at mode (p_1, p_2) is $\rho_M(d_{p_1,p_2}; \kappa, \nu)$, where d_{q_1,q_2} denotes the Euclidean distance between two different modes and ρ_M is the correlation function, belongs to Matérn family. The Matérn isotropic auto-correlation function has a specific form

$$\rho_M(d; \kappa, \nu) = \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{2d\sqrt{\nu}}{\kappa}\right)^{\nu} K_\nu \left(\frac{2d\sqrt{\nu}}{\kappa}\right), \qquad \kappa, \nu > 0 \tag{4.23}$$

Here, $K_\nu(\cdot)$ is termed as Bessel function of order ν. The positive range parameter κ controls the decay of the correlation between the observations at a large distance d. The order ν controls the behavior of the auto-correlation function for observations that are separated by a small distance. For our numerical example, we set the scale $\kappa = 0.55$ and the smoothness parameter $\nu = 1$.

The above mentioned situations have been implemented using "*stationary.image.cov*" and "*matern.image.cov*" functions respectively available in *fields* package in R (Douglas Nychka et al., 2017).

We run the simulation 100 times for each scenario for the evaluation of our method. For each of the simulation setups, we take the number of knots as $[J/4]$, where $[a]$ denotes the integer part of a. We compare the overall performance of the models to estimate the parameter curves for different choices of ranks by studying several error rates based on different norms. We choose the smoothing parameters θ from the set $\{0, 0.001, 0.005, 0.01, 0.05, 0.1\}$, on the other hand, ϕ is chosen from a grid from $\{0, 0.5, 3, 10\}$ and allow the different values from 1 to 5 for the choice of rank R. In the following tables, we denote $FToTM_r$ as proposed functional tensor-on-tensor model with rank r. To compare with the existing literature, we apply the concurrent linear model (Ramsay and Silverman, 2005) (*CLM*) for mode-wise analysis and implement this method using "*pffr*" function available in *refund* (Goldsmith et al., 2020) package in R, with the penalized concurrent effect of functional covariates (Ivanescu et al., 2015).

Tables 4.1, 4.2 and 4.3 show the results of integrated mean square errors and absolute errors IMSE $= \int_{t\in\mathcal{T}} \|\widehat{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}(t)\|^2_{\mathcal{F}} dt$ and IMAE $= \int_{t\in\mathcal{T}} \sum_{p_1,p_2,q_1,q_2} \left|\widehat{\beta}_{p_1,p_2,q_1,q_2}(t) - \beta_{p_1,p_2,q_1,q_2}(t)\right| dt$

Table 4.1 Results of simulation situations (Situation1) where each modes are assumed to be independent for $\mathsf{X}(t)$ and $\mathsf{E}(t)$ for fixed time-points. Here we assume each of $\{\chi^{(k)}_{p_1,p_2}\}_{p_1,p_2}$ and $\{\eta_{q_1,q_2}\}^{(k)}_{q_1,q_2}$ are independent for (p_1, p_2) and (q_1, q_2) respectively.

Method	IMSE (SD)	RIMSE (SD)	IMAE (SD)	RIMAE (SD)
$N = 30, P_1 \times P_2 = 5 \times 2, Q_1 \times Q_2 = 5 \times 2$				
CLM	0.14294 (0.02046)	0.01059 (0.00152)	0.28311 (0.02027)	0.09244 (0.00662)
$FToTM_1$	1.48469 (0.05628)	0.10998 (0.00417)	0.96636 (0.01626)	0.31552 (0.00531)
$FToTM_2$	0.45773 (0.02218)	0.03391 (0.00164)	0.53786 (0.01068)	0.17561 (0.00349)
$FToTM_3$	0.15078 (0.01316)	0.01117 (0.00097)	0.29482 (0.01452)	0.09626 (0.00474)
$FToTM_4$	0.01065 (0.00383)	0.00079 (0.00028)	0.07871 (0.01367)	0.0257 (0.00446)
$FToTM_5$	0.01558 (0.00582)	0.00115 (0.00043)	0.09412 (0.01695)	0.03073 (0.00553)
$N = 30, P_1 \times P_2 = 5 \times 2, Q_1 \times Q_2 = 15 \times 12$				
CLM	0.1448 (0.01339)	0.00193 (0.00018)	0.28468 (0.0132)	0.04054 (0.00188)
$FToTM_1$	9.24824 (0.06732)	0.12304 (0.0009)	2.27313 (0.01304)	0.32372 (0.00186)
$FToTM_2$	1.79804 (0.06786)	0.02392 (0.0009)	1.02121 (0.01836)	0.14543 (0.00261)
$FToTM_3$	0.23289 (0.02089)	0.0031 (0.00028)	0.36104 (0.01293)	0.05142 (0.00184)
$FToTM_4$	0.06108 (0.06808)	0.00081 (0.00091)	0.15243 (0.13744)	0.02171 (0.01957)
$FToTM_5$	0.00195 (0.00053)	0.00003 (0.00001)	0.03348 (0.00451)	0.00477 (0.00064)
$N = 100, P_1 \times P_2 = 5 \times 2, Q_1 \times Q_2 = 5 \times 2$				
CLM	0.03087 (0.00348)	0.00229 (0.00026)	0.13236 (0.00731)	0.04322 (0.00239)
$FToTM_1$	1.46268 (0.04068)	0.10835 (0.00301)	0.95921 (0.01095)	0.31319 (0.00358)
$FToTM_2$	0.43737 (0.01418)	0.0324 (0.00105)	0.52551 (0.00725)	0.17158 (0.00237)
$FToTM_3$	0.13651 (0.00541)	0.01011 (0.0004)	0.27253 (0.01099)	0.08898 (0.00359)
$FToTM_4$	0.00303 (0.00091)	0.00022 (0.00007)	0.04222 (0.00632)	0.01379 (0.00206)
$FToTM_5$	0.0037 (0.00115)	0.00027 (0.00008)	0.04663 (0.00696)	0.01523 (0.00227)
$N = 100, P_1 \times P_2 = 5 \times 2, Q_1 \times Q_2 = 15 \times 12$				
CLM	0.03082 (0.00163)	0.00041 (0.00002)	0.1328 (0.00357)	0.01891 (0.00051)
$FToTM_1$	9.21298 (0.04487)	0.12257 (0.0006)	2.26689 (0.01132)	0.32283 (0.00161)
$FToTM_2$	1.76018 (0.04482)	0.02342 (0.0006)	1.00917 (0.01218)	0.14372 (0.00173)
$FToTM_3$	0.22276 (0.03467)	0.00296 (0.00046)	0.35168 (0.02647)	0.05008 (0.00377)
$FToTM_4$	0.05837 (0.06468)	0.00078 (0.00086)	0.14918 (0.14726)	0.02124 (0.02097)
$FToTM_5$	0.00085 (0.00033)	0.00001 (<0.00001)	0.02197 (0.00403)	0.00313 (0.00057)

respectively. Similarly, we report the relative mean and absolute errors, which are defined as RIMSE = $\frac{\int_{t\in\mathcal{T}} \|\widehat{\boldsymbol{\beta}}(t)-\boldsymbol{\beta}(t)\|^2_{\mathcal{F}} dt}{\int_{t\in\mathcal{T}} \|\boldsymbol{\beta}(t)\|^2_{\mathcal{F}} dt}$ and RIMAE = $\frac{\int_{t\in\mathcal{T}} \sum_{p_1,p_2,q_1,q_2} \left|\widehat{\beta}_{p_1,p_2,q_1,q_2}(t)-\beta_{p_1,p_2,q_1,q_2}(t)\right| dt}{\int_{t\in\mathcal{T}} \sum_{p_1,p_2,q_1,q_2} \left|\beta_{p_1,p_2,q_1,q_2}(t)\right| dt}$ respectively. The advantages of these simulation situations are that these models are not based on the reduced-rank model. Here, we observe the curves in the presence of errors. All integrals are approximated

Table 4.2 Results of simulation situations (Situation2)a where each modes are assumed to be independent for $\mathsf{E}(t)$ for fixed time-points whereas modes for $\mathsf{X}(t)$ are assumed to be dependent. Here we assume $\{\chi^{(k)}_{p_1,p_2}\}_{p_1,p_2}$ is spatially dependent with exponential covariance function.

Method	IMSE (SD)	RIMSE (SD)	IMAE (SD)	RIMAE (SD)
$N = 30, P_1 \times P_2 = 5 \times 2, Q_1 \times Q_2 = 5 \times 2$				
CLM	11.03513 (2.27364)	0.81742 (0.16842)	2.45079 (0.24082)	0.8002 (0.07863)
$FToTM_1$	1.46631 (0.0141)	0.10862 (0.00104)	0.96402 (0.00583)	0.31476 (0.0019)
$FToTM_2$	0.60273 (0.01917)	0.04465 (0.00142)	0.60152 (0.01318)	0.1964 (0.0043)
$FToTM_3$	0.32753 (0.01741)	0.02426 (0.00129)	0.42707 (0.01962)	0.13944 (0.00641)
$FToTM_4$	0.21328 (0.21078)	0.0158 (0.01561)	0.35394 (0.13306)	0.11556 (0.04344)
$FToTM_5$	0.13694 (0.02654)	0.01014 (0.00197)	0.30854 (0.0384)	0.10074 (0.01254)
$N = 30, P_1 \times P_2 = 5 \times 2, Q_1 \times Q_2 = 15 \times 12$				
CLM	11.36335 (1.34533)	0.15118 (0.0179)	2.49712 (0.14845)	0.35562 (0.02114)
$FToTM_1$	9.21977 (0.02778)	0.12266 (0.00037)	2.27091 (0.0079)	0.32341 (0.00112)
$FToTM_2$	1.76995 (0.02734)	0.02355 (0.00036)	1.01769 (0.01081)	0.14493 (0.00154)
$FToTM_3$	0.41264 (0.16057)	0.00549 (0.00214)	0.48365 (0.08798)	0.06888 (0.01253)
$FToTM_4$	0.18218 (0.21293)	0.00242 (0.00283)	0.32063 (0.13906)	0.04566 (0.0198)
$FToTM_5$	0.06936 (0.05182)	0.00092 (0.00069)	0.19811 (0.08864)	0.02821 (0.01262)
$N = 100, P_1 \times P_2 = 5 \times 2, Q_1 \times Q_2 = 5 \times 2$				
CLM	2.55232 (0.45649)	0.18906 (0.03381)	1.19172 (0.10708)	0.38911 (0.03496)
$FToTM_1$	1.45974 (0.00711)	0.10813 (0.00053)	0.96178 (0.00323)	0.31403 (0.00105)
$FToTM_2$	0.58776 (0.01049)	0.04354 (0.00078)	0.59246 (0.00766)	0.19344 (0.0025)
$FToTM_3$	0.31275 (0.00961)	0.02317 (0.00071)	0.411 (0.01063)	0.1342 (0.00347)
$FToTM_4$	0.18492 (0.20235)	0.0137 (0.01499)	0.32604 (0.13409)	0.10646 (0.04378)
$FToTM_5$	0.11149 (0.03648)	0.00826 (0.0027)	0.27665 (0.06128)	0.09033 (0.02001)
$N = 100, P_1 \times P_2 = 5 \times 2, Q_1 \times Q_2 = 15 \times 12$				
CLM	2.5259 (0.21061)	0.0336 (0.0028)	1.18808 (0.05122)	0.1692 (0.00729)
$FToTM_1$	9.26995 (0.13929)	0.12333 (0.00185)	2.28525 (0.03385)	0.32545 (0.00482)
$FToTM_2$	1.74798 (0.01575)	0.02325 (0.00021)	1.00948 (0.00691)	0.14376 (0.00098)
$FToTM_3$	0.61359 (0.30173)	0.00816 (0.00401)	0.58812 (0.16308)	0.08376 (0.02322)
$FToTM_4$	0.66733 (0.41716)	0.00888 (0.00555)	0.596 (0.24026)	0.08488 (0.03422)
$FToTM_5$	0.0914 (0.04684)	0.00122 (0.00062)	0.23906 (0.07987)	0.03405 (0.01137)

using the Riemann sum. Since our proposed method involves an iterative procedure, which depends on the initial estimates, the computational time is therefore not comparable to that of the classical CLM, which is not an iterative method. For all situations, our proposed method does a much better job in terms of low error rates in estimating the parameter $\boldsymbol{\beta}(t)$.

Table 4.3 Results of simulation situations (Situation2)b where each modes are assumed to be independent for $\mathsf{E}(t)$ for fixed time-points whereas modes for $\mathsf{X}(t)$ are assumed to be dependent. Here we assume $\{\chi^{(k)}_{p_1,p_2}\}_{p_1,p_2}$ is spatially dependent with Matérn covariance function.

$N = 30, P_1 \times P_2 = 5 \times 2, Q_1 \times Q_2 = 5 \times 2$				
Method	IMSE (SD)	RIMSE (SD)	IMAE (SD)	RIMAE (SD)
CLM	0.26393 (0.04919)	0.01955 (0.00364)	0.38374 (0.03318)	0.12529 (0.01083)
$FToTM_1$	1.45885 (0.02061)	0.10806 (0.00153)	0.9599 (0.00731)	0.31342 (0.00239)
$FToTM_2$	0.46879 (0.02445)	0.03473 (0.00181)	0.54097 (0.01118)	0.17663 (0.00365)
$FToTM_3$	0.16291 (0.01629)	0.01207 (0.00121)	0.30998 (0.01506)	0.10121 (0.00492)
$FToTM_4$	0.0087 (0.01146)	0.00064 (0.00085)	0.06782 (0.0274)	0.02214 (0.00895)
$FToTM_5$	0.0111 (0.00525)	0.00082 (0.00039)	0.07909 (0.01855)	0.02582 (0.00606)
$N = 30, P_1 \times P_2 = 5 \times 2, Q_1 \times Q_2 = 15 \times 12$				
CLM	0.26313 (0.02958)	0.0035 (0.00039)	0.3835 (0.02167)	0.05462 (0.00309)
$FToTM_1$	9.22145 (0.02791)	0.12268 (0.00037)	2.27063 (0.00894)	0.32337 (0.00127)
$FToTM_2$	1.77848 (0.02878)	0.02366 (0.00038)	1.02026 (0.01052)	0.1453 (0.0015)
$FToTM_3$	0.23293 (0.01206)	0.0031 (0.00016)	0.36047 (0.00952)	0.05134 (0.00136)
$FToTM_4$	0.05929 (0.06315)	0.00079 (0.00084)	0.15872 (0.13817)	0.0226 (0.01968)
$FToTM_5$	0.00175 (0.00133)	0.00002 (0.00002)	0.03081 (0.00931)	0.00439 (0.00133)
$N = 100, P_1 \times P_2 = 5 \times 2, Q_1 \times Q_2 = 5 \times 2$				
CLM	0.05833 (0.00912)	0.00432 (0.00068)	0.18217 (0.01463)	0.05948 (0.00478)
$FToTM_1$	1.44275 (0.00963)	0.10687 (0.00071)	0.95499 (0.00374)	0.31181 (0.00122)
$FToTM_2$	0.44676 (0.01346)	0.03309 (0.001)	0.52798 (0.00559)	0.17239 (0.00183)
$FToTM_3$	0.14999 (0.00779)	0.01111 (0.00058)	0.29657 (0.00869)	0.09683 (0.00284)
$FToTM_4$	0.00231 (0.00143)	0.00017 (0.00011)	0.03593 (0.00956)	0.01173 (0.00312)
$FToTM_5$	0.00284 (0.00125)	0.00021 (0.00009)	0.04026 (0.00816)	0.01314 (0.00266)
$N = 100, P_1 \times P_2 = 5 \times 2, Q_1 \times Q_2 = 15 \times 12$				
CLM	0.05746 (0.00427)	0.00076 (0.00006)	0.18093 (0.00695)	0.02577 (0.00099)
$FToTM_1$	9.18754 (0.00744)	0.12223 (0.0001)	2.26337 (0.00385)	0.32233 (0.00055)
$FToTM_2$	1.73663 (0.00773)	0.0231 (0.0001)	1.00481 (0.0038)	0.1431 (0.00054)
$FToTM_3$	0.2181 (0.00522)	0.0029 (0.00007)	0.34535 (0.00306)	0.04918 (0.00044)
$FToTM_4$	0.05167 (0.05987)	0.00069 (0.0008)	0.13999 (0.14339)	0.01994 (0.02042)
$FToTM_5$	0.00081 (0.00061)	0.00001 (0.00001)	0.02055 (0.00654)	0.00293 (0.00093)

4.6 Application to *ForrestGump-data*

4.6.1 Details about the data-set

The *studyforrest* (website: https://www.studyforrest.org/) describes a publicly available data-set for the study of neural language and story processing. The imaging data analyzed here are publicly available through OpenfMRI (https://openneuro.org/datasets/ds000113/versions/1.3.0) (Hanke et al., 2014; Sengupta et al., 2016). In total 15 right-handed participants (mean age 29.4 years, range 21–39, 40% females, native German speaker) volunteered for a series of studies including eye-tracking experiments using natural signal stimulation with a motion picture. Volunteers have no known hearing problem without permanent or current temporary impairments, and no neurological disorder. Participants viewed a feature film " Forrest Gump" (Robert Zemeckis, Paramount Pictures, 1994 with German audio track) in eight back-to-back 15-minute movie sessions, which were presented chronologically in two back-to-back sessions on the same day. Each session contained four segments, each approximately 15 minutes long. The eye tracking camera was fitted just outside the scanner bore, approximately centered and viewing the left eye of the participant at a distance of 100 cm through a small gap between the top of the back projection screen and the scanner bore ceiling. Participants were allowed to perform free eye movements without having to fixate or keep the eye open. The eye-gaze recording started as soon as the computer received the first fMRI trigger signal. In the audio-visual movie, the video-track of the movie was extracted and encoded as H.264 (1280×720 at 25 fps). The movie was shown on a 1280×1024 pixel screen with a 63 cm viewing distance in 720p resolution. The temporal resolution of the participants' eye gaze recording was 1000Hz.

All fMRI acquisitions had the following parameters: T2*- weighted echo-planner images with 2 second repetition time (TR), 30 ms echo time, and 90-degree flip angle were acquired during stimulation using a 3 Tesla MRI scanner. The dimension of the images for each time-point was $80 \times 80 \times 35$ (with pixel dimension $3 \times 3 \times 3.3 mm^3$). The number of volumes acquired for the selected session was 451.

All files related to data acquisition for a particular subject are available as sub-<ID>/ses-movie/ directory, where ID is the numeric subject ID. fMRI data files are available with the file name ses-movie_task-movie_<run>_bold. In the scanner, the normalized eye-gaze coordinate time series are located at sub-<ID>/ses-movie/func/sub-<ID>_recording-eyegaze_physio.tsv.gz which contain X and Y coordinates of the eye-gaze, pupil area measurements, and numerical ID of movie frame presented at the time of measurement. Since the sampling rate is uniformly 1000 Hz, we have 1000 lines per second, with the first line corresponding to the onset of the movie stimulus. Here, the coordinates $(0, 0)$ were located at the top-left corner of the movie frame, and the lower right corner is located at $(1280, 546)$. Both measurements were taken excluding the gray bar of the frame. All in-scanner recordings were temporally normalized by shifting the time series by the minimal video onset. Stimulus timing information was recorded in events.tsv files which contain the onset, duration of each movie frame. In eye-gazing data, there is huge change of loss of information due to eye blinks and those are marked as *nan* in the data-set and perform spline interpolation. We used 14 individuals and removed Subject 5 due to excessive missing data.

We use the eye position in the angular unit (i.e., polar coordinates) instead of Cartesian coordinates, where we report magnitude changes of eye position in the screen reference system. Moreover, X and Y coordinates, related polar coordinates, and pupil area were down-sampled to the match fMRI sampling frequency. Table 4.4 reports the covariance between angles and distance for each participants and quotient of the standard deviations of distance and angle, and that of vertical and horizontal direction of the frame. Fransson et al. (2014) investigates the relationship between spontaneous charges in eye position during passive fixation and intrinsic brain position in a block-related task in a resting-state fMRI and concurrent recordings of eye-gaze experiments.

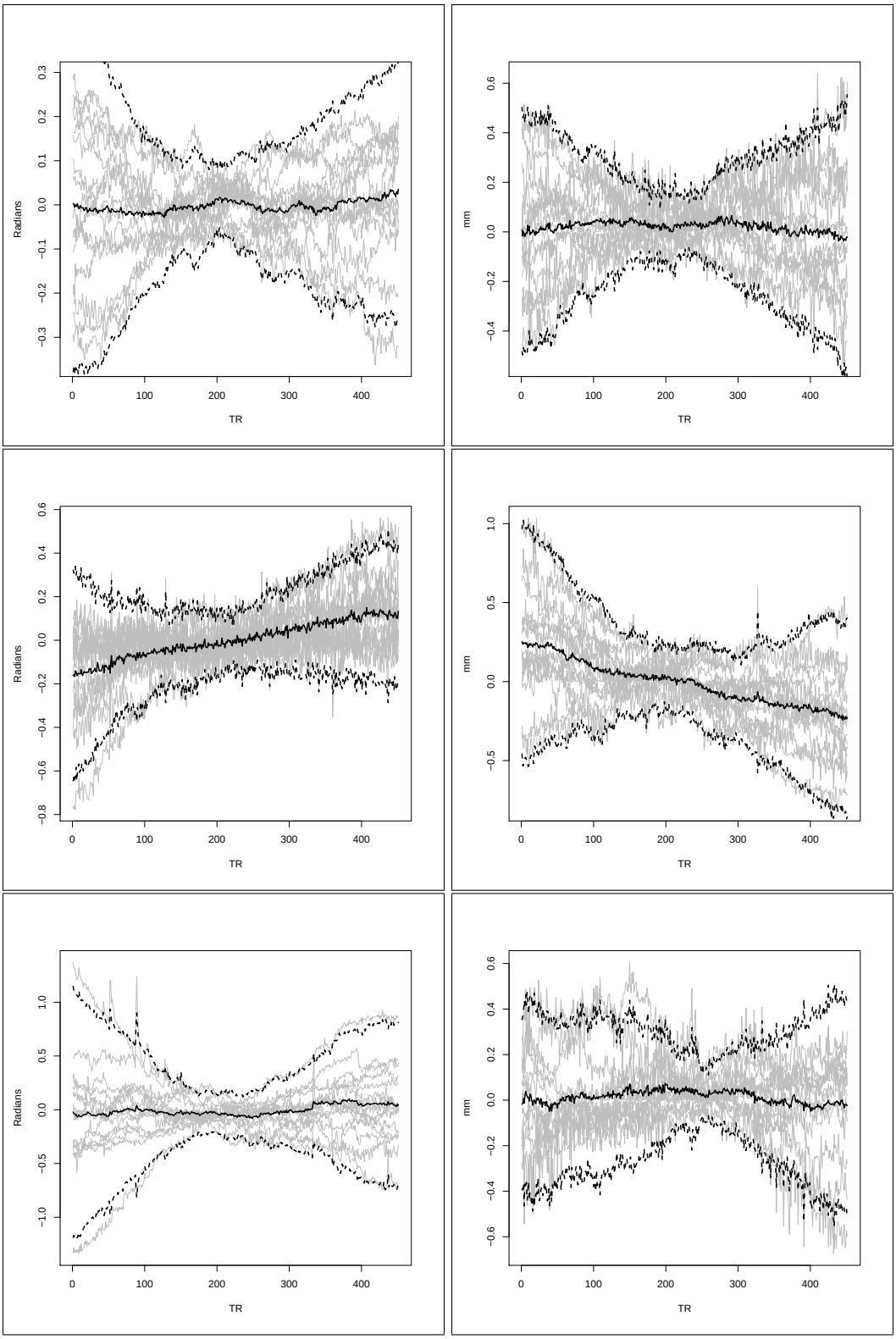

Figure 4.3 *ForrestGump-data*: Summary statistics for the parameters estimates of head motion correction across TRs and participants. (Left panel) Magnitude of three rotational parameters (in radians) and (Right panel) Magnitude of three translation parameters (in millimeters) for each individual on each of the 451 TRs. In each plot, solid black line indicates mean over the individuals through TRs and black dotted lines indicate mean±2sd over the individuals through TRs.

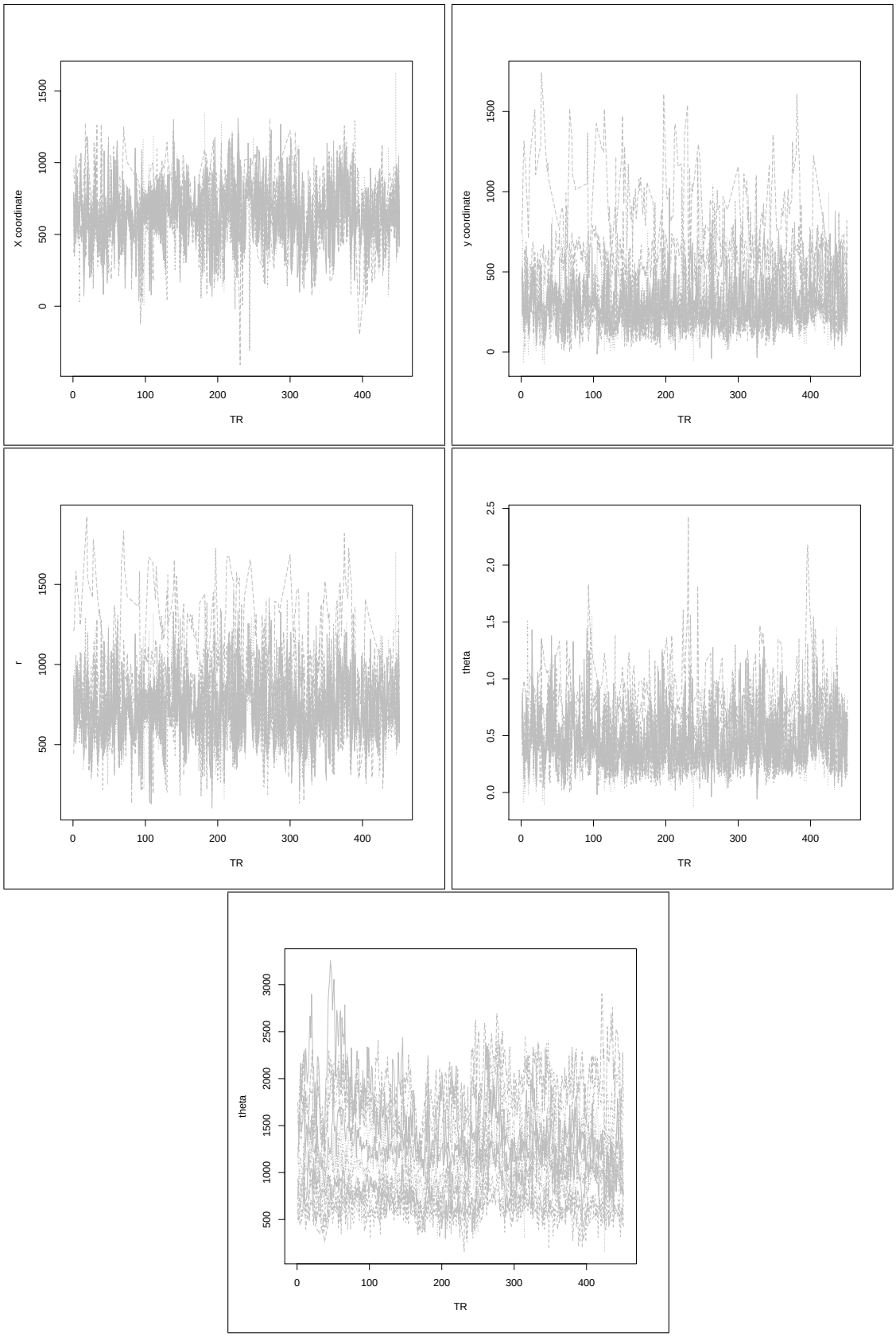

Figure 4.4 *ForrestGump-data*: Covariates of interest. Cartesian and polar coordinates and pupil area are shown across TRs and participants.

ID	sd(X)/sd(Y)	sd(dist)/sd(angle)	corr(dist, angle)
1	1.7044	922.2200	-0.3640
2	1.5247	1004.1233	-0.2552
3	1.6271	919.6505	-0.3124
4	1.6990	839.1719	-0.4135
5	1.1656	1156.9217	-0.4048
6	1.3943	772.5980	-0.2928
9	1.4403	762.4111	-0.3457
10	1.1282	821.5585	0.0177
14	1.9538	967.5292	-0.4522
15	1.1085	753.6706	-0.0512
16	1.7194	899.7864	-0.4071
17	1.9600	904.5869	-0.4612
18	1.6047	890.4765	-0.3568
19	1.3867	1069.9573	-0.2195
20	1.4631	944.8355	-0.3688

Table 4.4 *ForrestGump-data*: Summary statistics across participants.

4.6.2 Analysis

To analyze the data on a local computer, we only used the first run of the experiment for each individual and down-sampled the images to $64 \times 64 \times 64$ via nearest-neighbor interpolation using "*resize*" function in Matlab, where the number of time-points was 451. Details of the pre-processing steps performed along with further information of data acquisitions are described in Appendix A.

Our scientific question of interest was to understand the association between brain image pattern in the presence of audio-visual inputs. This is the first approach to statistically analyze such a study by exploiting the complex structure of the data. We fit a time-varying tensor regression coefficient model as described in Section 4.2. Our covariate is a 3-mode tensor representing normalized eye-gaze coordinate time-series; each mode representing scaled polar coordinates of the eye-gaze and pupil area measurements, respectively. The response of the model is pre-processed fMRI data. Response and covariates are collected simultaneously. The coefficient functions $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_2$ and $\boldsymbol{\beta}_3$ are amplitudes over the time associated with distance, angle of eye-gaze and pupil area respectively; included to detect the effect of movie in a visual form in BOLD response change. We choose the

rank for reduced-rank extraction to be 3 since it has lowest prediction error.

For interpretation purposes, we evaluate estimates $\widehat{\boldsymbol{\beta}}(t)$ by taking average values over eight different functional networks in the brain. This was achieved by first parcellating the brain into the 268 regions of the Shen atlas (Shen et al., 2013). These regions were thereafter further combined into eight functional networks (Finn et al., 2015): medial frontal, frontoparietal, default mode, subcortical-cerebellum, motor, visual I, visual II, and visual association. Figures 4.5, 4.6, 4.7 represent the average estimated coefficient function corresponding to three visual features (distance, angle of eye-gaze, and pupil area) over all the time-points for each network respectively. Vertical lines represent scene changes in the movie. The first segment, consisting of approximately 84 time-points corresponds to the opening sequence, which shows a feather floating through the sky as credits are shown. The second segment consists of the famous scene where the protagonist of the movie sits on a bench at a bus stop and begins discussing the story of his life. During this scene, there is heightened activation in several brain networks in reaction to different visual features. Throughout the time course, the changes in visual features have the greatest impact on activation in "visual I", which is depicted using purple lines and is consistent with what we should expect. Moreover, subsequent segments represent scene changes alternating between interior and exterior settings; see Häusler and Hanke (2016) for more details.

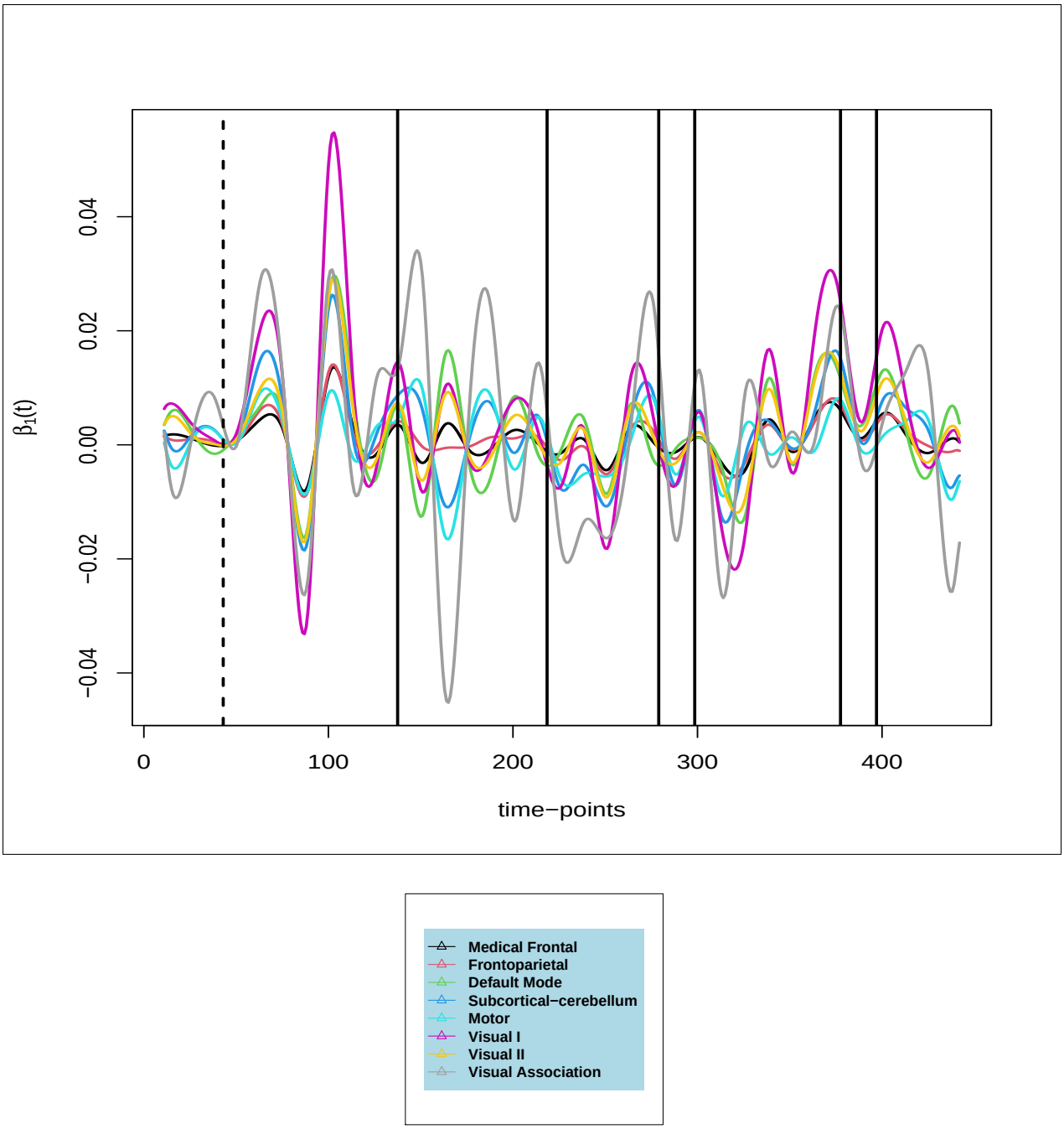

Figure 4.5 *ForrestGump-data* results: Estimate of the coefficient $\boldsymbol{\beta}_1(t)$ corresponding to visual feature *distance of eye-gaze* for different location. Legends for different parcellation as mentioned in Shen et al. (2013) are also provided.

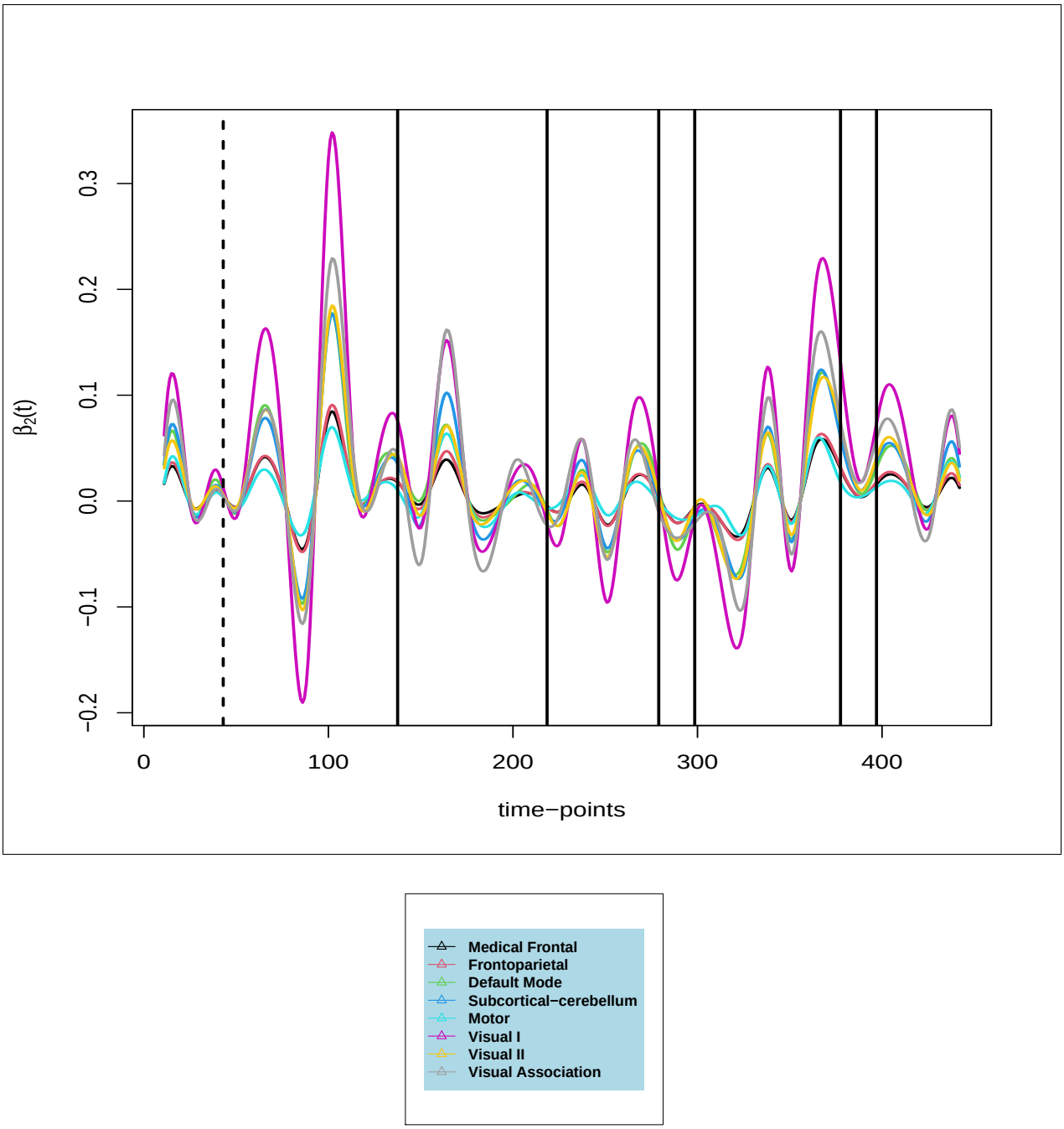

Figure 4.6 *ForrestGump-data* results: Estimate of the coefficient $\boldsymbol{\beta}_2(t)$ corresponding to visual feature *angle of eye-gaze* for different location. Legends for different parcellation as mentioned in Shen et al. (2013) are also provided.

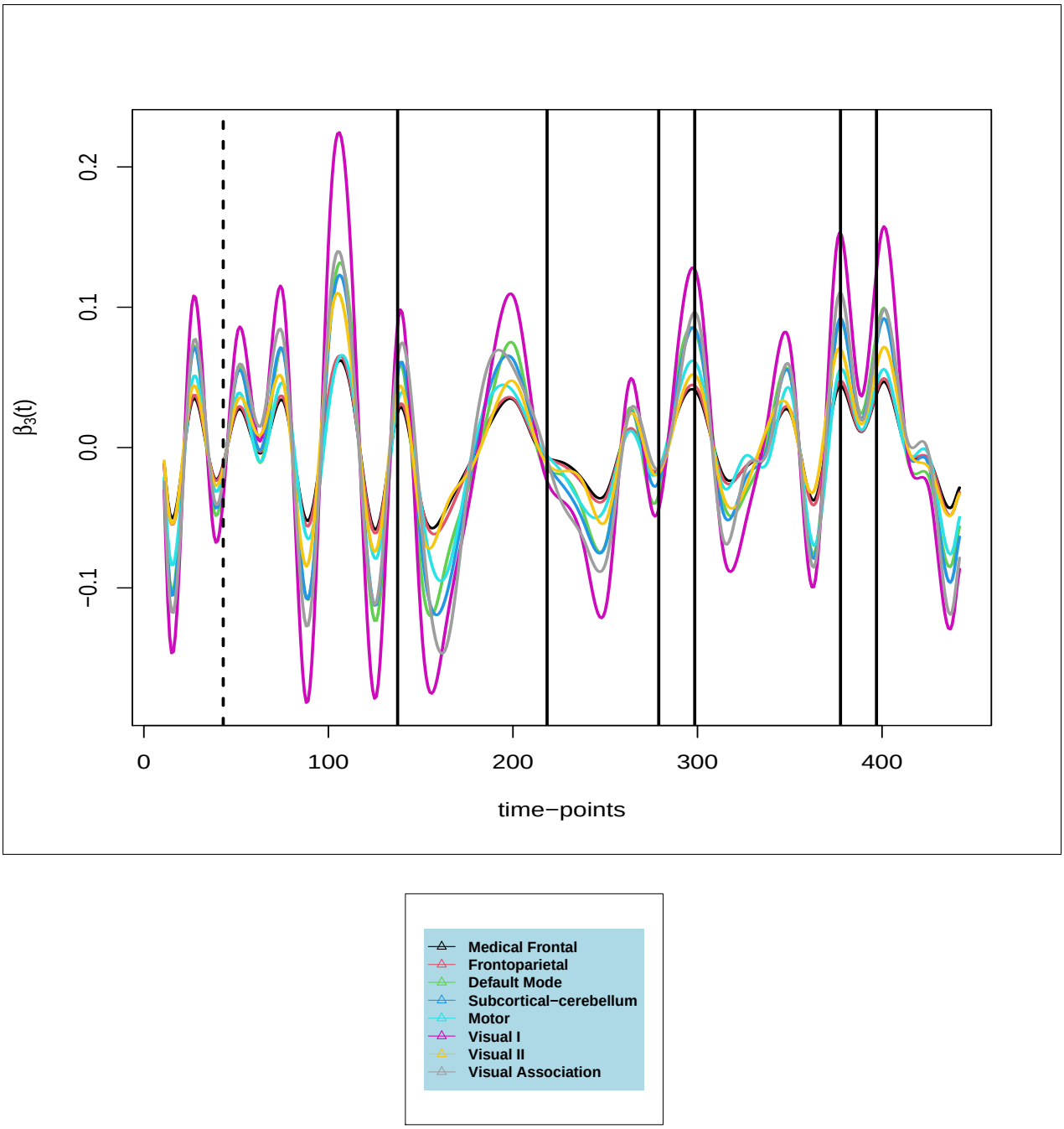

Figure 4.7 *ForrestGump-data* results: Estimate of the coefficient $\boldsymbol{\beta}_3(t)$ corresponding to visual feature *pupil area* for different location. Legends for different parcellation as mentioned in Shen et al. (2013) are also provided.

4.7 Discussion

In this chapter, we have proposed a time-varying tensor-on-tensor regression model and a method to estimate the coefficient tensors which belong to an infinite-dimensional space. We believe that the method provides an efficient approach towards performing multi-modal data analysis using neuroimaging data. Regression coefficients are expressed using the B-spline technique, and the coefficients of the B-spline bases are estimated using low-rank tensor decomposition. This method reduces the vastness of the parameters of interest and computational complexity. We have provided a meaningful simulation study as well as performed real data analysis combining fMRI and eye-tracking data. The results of our data analysis suggest that the approach has promise for identifying brain regions responding to an external stimulus, which in this case is movie-watching.

Although our tensor data can be compactly represented by a CP model, it is NP hard to determine the rank of the low-rank decomposition (Johan, 1990). To determine the tuning parameters, one can perform the cross-validation technique. However, our main objective is not to choose the optimal rank of the low-rank decomposition in the algorithm, and we leave this for future research. Furthermore, the tensor train representation (Liu et al., 2020) could be an alternative representation of the multi-dimensional array. In conclusion, our work provides an important direction for dealing with massive structured data such as time-varying tensors for analysis in multi-modal neuroimaging studies.

4.8 Technical details

4.8.1 Technical lemmas

Lemma 4.8.1. *For any positive definite matrices* $\mathbf{A}$ *and* $\mathbf{B}$ *we have*

$$\lambda_{\min}(\mathbf{A})trace\{\mathbf{B}\} \leq trace\{\mathbf{AB}\} \leq \lambda_{\max}(\mathbf{A})trace\{\mathbf{B}\} \tag{4.24}$$

where $\lambda_{\max}(\mathbf{A})$ *is the largest eigen-value of* $\mathbf{A}$ *and* $\lambda_{\min}(\mathbf{A})$ *is the smallest eigen-value of* $\mathbf{A}$.

Proof. See Fang et al. (1994) for the proof in detail. □

Before introducing the next lemma, define a P-dimensional vector $\mathbf{u} = (u_1, \cdots u_P)^{\mathrm{T}}$ which is sub-Gaussian with some parameters σ, then for all $\boldsymbol{\alpha} \in \mathbb{R}^P$,

$$\mathbb{E}\{\exp \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{u}\} \leq \exp(\|\boldsymbol{\alpha}\|^2\sigma^2/2) \tag{4.25}$$

Define the locally stationary time series $u_j = \mathscr{G}(j/J, \mathscr{F}_j)$ where $\mathscr{F}_j = (\cdots, \eta_{j-1}, \eta_j, \cdots)$, η_js are i.i.d. random variables and $\mathscr{G} : [0,1]\times\mathbb{R}^\infty \to \mathbf{R}$ is a measurable function such that $\xi_j(t) = \mathscr{G}(t, \mathscr{F}_j)$. Let $\{\eta'\}$ be i.i.d. copies of η and assume that for some $a > 0$, define the L_a-norm $\|\eta\|_a = \{\mathbb{E}|\eta|^a\}^{1/a}$. Then for $k \geq 0$ define the physical dependence measure $\Delta(k, a) = \sup_{t\in[0,1]} \max_j \|\mathscr{G}(t, \mathscr{F}_j) - \mathscr{G}(t, \mathscr{F}_{j,k})\|_a$ where $\mathscr{F}_{j,k} = (\mathscr{F}_{j-k-1}, \eta'_{j-k}, \eta_{j-k+1}, \cdots, \eta_j)$. Moreover, recall the condition (C4) where for some large $a, \kappa_0 > 0$, there exists a universal constant $C > 0$ such that $\Delta(k, a) \leq Ck^{-\kappa_0}$ for $k \geq 1$. Furthermore, let $\|\eta\|_a$ be finite for some $a > 1$.

Lemma 4.8.2. *Under condition (C4), with the above explanation, for some constant $C_a > 0$,*

$$\mathbb{P}\left\{\frac{1}{NJ}\sigma_1(\mathscr{P}\mathbf{E}) \leq \frac{Q\xi N^\tau}{\sqrt{J}}\right\} \geq 1 - C_a N^{-a\tau} \tag{4.26}$$

where τ is some small positive real number and $\xi = \sup_{1\leq h\leq H} \sup_{t\in[0,1]} |\mathbb{B}_h(t)|$

Proof. See Ding et al. (2021) and the references herein for the proof in detail. □

Lemma 4.8.3. *Define $\mathcal{S}_n$ be a collection of spline such that the function $g_\bullet(t) = \sum_{h=1}^{K_N+v+1} b_{h,\bullet}B_h(t)$, where $\{B_h, h = 1, \cdots, (K_N + v + 1)\}$ is a set of B-spline bases in S_n. Under conditions (C2) and (C3), there exists a spline function $g_\bullet(t) \in S_n$ such that*

$$\sup_{t\in\mathscr{T}} |\beta_\bullet(t) - g_\bullet(t)| = O\left(\frac{1}{K_N^{v+1}}\right) \tag{4.27}$$

Proof. This proof follows from De Boor et al. (1978). □

4.8.2 Proof of Theorem 4.3.1

For simplicity, assume $\mathbf{Y} \in \mathbb{R}^{NJ\times Q}$ and $\mathsf{Z} \in \mathbb{R}^{NJ\times H\times P}$, thus $\mathsf{B} \in \mathbb{R}^{H\times P\times Q}$. The contracted inner product in this proof is of order 2, i.e., $< \cdot, \cdot >_2$, for simplicity, we drop subscript 2 from the inner

product. By the definition of $\widehat{\mathsf{B}}_0$, for all matrices C of rank R_0 with order $H_N \times P \times Q$, we have

$$\|\mathbf{Y} - \left\langle \mathsf{Z}, \widehat{\mathsf{B}}_0 \right\rangle\|_{\mathscr{F}}^2 + (NJ)\|\widehat{\mathsf{B}}_0\|_{\mathscr{F},\mathbf{W}_\omega}^2 \leq \|\mathbf{Y} - \langle \mathsf{Z}, \mathsf{C}\rangle\|_{\mathscr{F}}^2 + (NJ)\|\mathsf{C}\|_{\mathscr{F},\mathbf{W}_\omega}^2 \tag{4.28}$$

In addition, the following two equations hold for any tensor C,

$$\begin{aligned}
\|\mathbf{Y} - \langle \mathsf{Z}, \mathsf{C}\rangle\|_{\mathscr{F}}^2 &= \|\mathbf{Y} - \langle \mathsf{Z}, \mathsf{B}_0\rangle\|_{\mathscr{F}}^2 + \|\langle \mathsf{Z}, (\mathsf{B}_0 - \mathsf{C})\rangle\|_{\mathscr{F}}^2 + 2\langle \mathbf{E}, \langle \mathsf{Z}, (\mathsf{B}_0 - \mathsf{C})\rangle\rangle_{\mathscr{F}} \\
\|\mathbf{Y} - \left\langle \mathsf{Z}, \widehat{\mathsf{B}}_0 \right\rangle\|_{\mathscr{F}}^2 &= \|\mathbf{Y} - \langle \mathsf{Z}, \mathsf{B}_0\rangle\|_{\mathscr{F}}^2 + \|\left\langle \mathsf{Z}, (\mathsf{B}_0 - \widehat{\mathsf{B}}_0)\right\rangle\|_{\mathscr{F}}^2 + 2\left\langle \mathbf{E}, \left\langle \mathsf{Z}, (\mathsf{B}_0 - \widehat{\mathsf{B}}_0)\right\rangle\right\rangle_{\mathscr{F}}
\end{aligned} \tag{4.29}$$

with $\langle \mathbf{A}, \mathbf{B}\rangle_{\mathscr{F}} = \text{trace}\{\mathbf{A}^{\mathrm{T}}\mathbf{B}\}$ for any matrices $\mathbf{A}$ and $\mathbf{B}$ such that the matrix product of $\mathbf{A}^{\mathrm{T}}\mathbf{B}$ is permissible. Define, $\mathscr{P} = \mathsf{Z}_{(1)}(\mathsf{Z}_{(1)}^{\mathrm{T}}\mathsf{Z}_{(1)})^{-1}\mathsf{Z}_{(1)}^{\mathrm{T}}$, then by the definition of Frobenius inner product, $\left\langle \mathbf{E}, \left\langle \mathsf{Z}, (\widehat{\mathsf{B}}_0 - \mathsf{B})\right\rangle\right\rangle_{\mathscr{F}} = \left\langle \mathscr{P}\mathbf{E}, \left\langle \mathsf{Z}, (\widehat{\mathsf{B}}_0 - \mathsf{C})\right\rangle\right\rangle_{\mathscr{F}}$. Moreover, the inner product norm $\langle \cdot, \cdot\rangle_{\mathscr{F}}$, operator norm $\|\cdot\|_2 = \sigma_1(\cdot)$ and nuclear norm $\|\cdot\|_* = \sum_i \sigma_i(\cdot)$ are related using the inequalities $\langle \mathbf{A}, \mathbf{B}\rangle_{\mathscr{F}} \leq \|\mathbf{A}\|_2\|\mathbf{B}\|_*$ and $\|\mathbf{B}\|_* \leq \sqrt{r}\|\mathbf{B}\|_{\mathscr{F}}$ where r be the rank of the matrix $\mathbf{B}$ and $\sigma_i(\cdot)$ represents the i^{th} largest singular value of a matrix. By subtracting the two Equations in 4.29 and exercising the properties of different norms mentioned above, we get the following inequalities.

$$\begin{aligned}
\|\left\langle \mathsf{Z}, (\widehat{\mathsf{B}}_0 - \mathsf{B}_0)\right\rangle\|_{\mathscr{F}}^2 &\leq \|\langle \mathsf{Z}, (\mathsf{C} - \mathsf{B}_0)\rangle\|_{\mathscr{F}}^2 + 2\left\langle \mathbf{E}, \left\langle \mathsf{Z}, (\widehat{\mathsf{B}}_0 - \mathsf{C})\right\rangle\right\rangle_{\mathscr{F}} + (NJ)\left\{\|\mathsf{C}\|_{\mathscr{F},\mathbf{W}_\omega}^2 - \|\widehat{\mathsf{B}}_0\|_{\mathscr{F},\mathbf{W}_\omega}^2\right\} \\
&= \|\langle \mathsf{Z}, (\mathsf{C} - \mathsf{B}_0)\rangle\|_{\mathscr{F}}^2 + 2\left\langle \mathscr{P}\mathbf{E}, \left\langle \mathsf{Z}, (\widehat{\mathsf{B}}_0 - \mathsf{C})\right\rangle\right\rangle_{\mathscr{F}} \\
&\quad + (NJ)\left\{\|\mathsf{C}\|_{\mathscr{F},\mathbf{W}_\omega}^2 - \|\widehat{\mathsf{B}}_0\|_{\mathscr{F},\mathbf{W}_\omega}^2\right\} \\
&\leq \|\langle \mathsf{Z}, (\mathsf{C} - \mathsf{B}_0)\rangle\|_{\mathscr{F}}^2 + 2\sigma_1(\mathscr{P}\mathbf{E})\sqrt{2R_0}\|\left\langle \mathsf{Z}, (\widehat{\mathsf{B}}_0 - \mathsf{C})\right\rangle\|_{\mathscr{F}} \\
&\quad + (NJ)\left\{\|\mathsf{C}\|_{\mathscr{F},\mathbf{W}_\omega}^2 - \|\widehat{\mathsf{B}}_0\|_{\mathscr{F},\mathbf{W}_\omega}^2\right\}
\end{aligned} \tag{4.30}$$

Define, $\mathbf{P} = \mathbf{I}_Q \otimes \mathbf{I}_P \otimes \int \mathbf{B}''(t)\mathbf{B}''(t)^{\mathrm{T}}dt$ and observe the fact that $\lambda_{\max}(\mathbf{P}) = \lambda_{\max}(\int \mathbf{B}''(t)\mathbf{B}''(t)^{\mathrm{T}}dt)$. Now consider, for any tensor with C, using Lemma 4.8.1,

$$\begin{aligned}
&\text{vec}(\mathsf{C})^{\mathrm{T}}\mathbf{P}\,\text{vec}(\mathsf{C}) - \text{vec}(\widehat{\mathsf{B}}_0)^{\mathrm{T}}\mathbf{P}\,\text{vec}(\widehat{\mathsf{B}}_0) \\
&= \text{trace}\{\mathbf{P}(\text{vec}(\mathsf{C})\,\text{vec}(\mathsf{C})^{\mathrm{T}} - \text{vec}(\widehat{\mathsf{B}}_0)\,\text{vec}(\widehat{\mathsf{B}}_0)^{\mathrm{T}})\} \\
&\leq \lambda_{\max}(\mathbf{P})\text{trace}\{\text{vec}(\mathsf{C})\,\text{vec}(\mathsf{C})^{\mathrm{T}} - \text{vec}(\widehat{\mathsf{B}}_0)\,\text{vec}(\widehat{\mathsf{B}}_0)^{\mathrm{T}}\}
\end{aligned}$$

$$= \lambda_{\max}\Big(\int \mathbf{B}''(t)\mathbf{B}''(t)^{\mathrm{T}}dt\Big)\{\|\mathsf{C}\|_{\mathscr{F}}^2 - \|\widehat{\mathsf{B}}_0\|_{\mathscr{F}}^2\} \tag{4.31}$$

As a consequence of the above inequality,

$$\begin{aligned}
\|\mathsf{C}\|_{\mathscr{F},\mathbf{W}_\omega}^2 - \|\widehat{\mathsf{B}}_0\|_{\mathscr{F},\mathbf{W}_\omega}^2 &= \mathrm{vec}(\mathsf{C})^{\mathrm{T}}\mathbf{W}_\omega\,\mathrm{vec}(\mathsf{C}) - \mathrm{vec}(\widehat{\mathsf{B}}_0)^{\mathrm{T}}\mathbf{W}_\omega\,\mathrm{vec}(\widehat{\mathsf{B}}_0) \\
&= \theta\left(\mathrm{vec}(\mathsf{C})^{\mathrm{T}}\mathbf{P}\,\mathrm{vec}(\mathsf{C})\} - \mathrm{vec}(\widehat{\mathsf{B}}_0)^{\mathrm{T}}\mathbf{P}\,\mathrm{vec}(\widehat{\mathsf{B}}_0)\}\right) \\
&\quad + \phi\left(\mathrm{vec}(\mathsf{C})^{\mathrm{T}}\,\mathrm{vec}(\mathsf{C})\} - \mathrm{vec}(\widehat{\mathsf{B}}_0)^{\mathrm{T}}\,\mathrm{vec}(\widehat{\mathsf{B}}_0)\}\right) \\
&\leq (\theta\lambda_{\max}(\mathbf{P}) + \phi)\left\{\|\mathsf{C}\|_{\mathscr{F}}^2 - \|\widehat{\mathsf{B}}_0\|_{\mathscr{F}}^2\right\} \\
&= (\theta\lambda_{\max}(\int \mathbf{B}''(t)\mathbf{B}''(t)^{\mathrm{T}}dt) + \phi)\left\{\|\mathsf{C}\|_{\mathscr{F}}^2 - \|\widehat{\mathsf{B}}_0\|_{\mathscr{F}}^2\right\}
\end{aligned} \tag{4.32}$$

Then for tensor C with $\mathrm{rank}(\mathsf{C}) \leq R_0$ and $I = \min(H, PQ)$, we have the following inequalities.

$$\begin{aligned}
&\|\mathsf{C}\|_{\mathscr{F}}^2 - \|\widehat{\mathsf{B}}_0\|_{\mathscr{F}}^2 \\
&= \sum_{i=1}^{I}\sigma_i^2(\mathsf{C}_{(1)}) - \sum_{i=1}^{I}\sigma_i^2(\widehat{\mathsf{B}}_{0(1)}) \\
&\leq \left\{\sigma_1(\mathsf{C}_{(1)}) + \sigma_1(\widehat{\mathsf{B}}_{0(1)})\right\}\left\{\sum_{i=1}^{I}\left(\sigma_i(\mathsf{C}_{(1)}) - \sigma_i(\widehat{\mathsf{B}}_{0(1)})\right)\right\} \\
&\overset{(i)}{\leq} \left\{2\sigma_1(\mathsf{C}_{(1)}) + \sigma_1(\widehat{\mathsf{B}}_{0(1)} - \mathsf{C}_{(1)})\right\}\left\{\sum_{i=1}^{I}\sigma_i(\widehat{\mathsf{B}}_{0(1)} - \mathsf{C}_{(1)})\right\} \\
&= \left\{2\sigma_1(\mathsf{C}_{(1)}) + \sigma_1(\widehat{\mathsf{B}}_{0(1)} - \mathsf{C}_{(1)})\right\}\left\{\sum_{i=1}^{R_0}\sigma_i(\widehat{\mathsf{B}}_{0(1)} - \mathsf{C}_{(1)})\right\} \\
&\overset{(ii)}{\leq} \left\{2\sigma_1(\mathsf{C}_{(1)}) + \|\widehat{\mathsf{B}}_0 - \mathsf{C}\|_{\mathscr{F}}\right\}\left\{\sqrt{2R_0}\|\widehat{\mathsf{B}}_0 - \mathsf{C}\|_{\mathscr{F}}\right\} \\
&\leq \sqrt{2R_0}\left\{2\sigma_1(\mathsf{C}_{(1)}) + \|\widehat{\mathsf{B}}_0 - \mathsf{C}\|_{\mathscr{F}}\right\}^2
\end{aligned} \tag{4.33}$$

where inequality (i) follows since $\sigma_{i+j-1}(\mathbf{A}+\mathbf{B}) \leq \sigma_i(\mathbf{A}) + \sigma_j(\mathbf{B})$, or in other words due to Weyl's additive perturbation theory which states that $\sigma_{i+j-1}(\mathbf{A}) \leq \sigma_i(\mathbf{B}) + \sigma_j(\mathbf{A}-\mathbf{B})$. Inequality (ii) holds since by definition $\sigma_1(\mathbf{A}) = \|\mathbf{A}\|_2$, operator norm. Moreover, $\|\mathbf{A}\|_2 \leq \|\mathbf{A}\|_{\mathscr{F}}$ and due to Cauchy-Schwarz inequality along with the fact that $\mathrm{rank}(\mathbf{A}+\mathbf{B}) \leq rank(\mathbf{A}) + rank(\mathbf{B})$. Also, $\|\left\langle \mathsf{Z}, (\widehat{\mathsf{B}}_0 - \mathsf{C})\right\rangle\|_{\mathscr{F}}^2 = \|\left\langle \mathsf{Z}_{(1)}, (\widehat{\mathsf{B}}_0 - \mathsf{C})_{(3)}\right\rangle\|_{\mathscr{F}}^2 = \|\mathsf{Z}_{(1)}^{\mathrm{T}}(\widehat{\mathsf{B}}_0 - \mathsf{C})_{(3)}\|_{\mathscr{F}}^2 \geq \|\widehat{\mathsf{B}}_0 - \mathsf{C}\|_{\mathscr{F}}^2\lambda_{\min}(\mathbf{Z}_{(1)}^{\mathrm{T}}\mathbf{Z}_{(1)})$ due to 4.8.1. Therefore, using the inequality $(x+y)^2 \leq 2(x^2+y^2)$ we have for

$\mu = (NJ)(\theta\lambda_{\max}(\int \mathbf{B}''(t)\mathbf{B}''(t)^{\mathrm{T}}dt) + \phi)\sqrt{2R_0}$

$$
\begin{aligned}
(NJ)\left\{\|\mathsf{C}\|^2_{\mathscr{F},\mathbf{W}_\omega} - \|\widehat{\mathsf{B}}_0\|^2_{\mathscr{F},\mathbf{W}_\omega}\right\} &\leq \mu\left\{2\sigma_1(\mathsf{C}_{(1)}) + \lambda^{-1}_{\min}(\mathsf{Z}^{\mathrm{T}}_{(1)}\mathbf{Z}_{(1)})\|\left\langle \mathsf{Z}, (\widehat{\mathsf{B}}_0 - \mathsf{C})\right\rangle\|_{\mathscr{F}}\right\}^2 \\
&\leq \mu\left\{4\sigma^2_1(\mathsf{C}_{(1)}) + \lambda^{-2}_{\min}(\mathsf{Z}^{\mathrm{T}}_{(1)}\mathsf{Z}_{(1)})\|\left\langle \mathsf{Z}, (\widehat{\mathsf{B}}_0 - \mathsf{C})\right\rangle\|^2_{\mathscr{F}}\right\} \\
&\leq 4\mu\sigma^2_1(\mathsf{C}_{(1)}) + 2\mu\lambda^{-2}_{\min}(\mathsf{Z}^{\mathrm{T}}_{(1)}\mathbf{Z}_{(1)})\|\left\langle \mathsf{Z}, (\widehat{\mathsf{B}}_0 - \mathsf{B}_0)\right\rangle\|^2_{\mathscr{F}} \\
&\quad + 2\mu\lambda^{-2}_{\min}(\mathsf{Z}^{\mathrm{T}}_{(1)}\mathsf{Z}_{(1)})\|\left\langle \mathsf{Z}, (\mathsf{C} - \mathsf{B}_0)\right\rangle\|^2_{\mathscr{F}} \qquad (4.34)
\end{aligned}
$$

Therefore, we obtain the bound for the prediction error as the following way using the assumption that $\lambda_{\min}(\mathsf{Z}^{\mathrm{T}}_{(1)}\mathsf{Z}_{(1)})$ is bounded below by λ with high probability and by inequality $2xy \leq x^2/a + ay^2$ in $(\star)$, consider the following from Equation (4.30),

$$
\begin{aligned}
\|\left\langle \mathsf{Z}, (\widehat{\mathsf{B}}_0 - \mathsf{B}_0)\right\rangle\|^2_{\mathscr{F}} &\leq \|\left\langle \mathsf{Z}, (\mathsf{C} - \mathsf{B}_0)\right\rangle\|^2_{\mathscr{F}} + 2\sigma_1(\mathscr{P}\mathbf{E})\sqrt{2R_0}\|\left\langle \mathsf{Z}, (\widehat{\mathsf{B}}_0 - \mathsf{C})\right\rangle\|_{\mathscr{F}} \\
&\quad + 2\mu\lambda^{-2}\|\left\langle \mathsf{Z}, (\widehat{\mathsf{B}}_0 - \mathsf{B}_0)\right\rangle\|^2_{\mathscr{F}} + 2\mu\lambda^{-2}\|\left\langle \mathsf{Z}, (\mathsf{C} - \mathsf{B}_0)\right\rangle\|^2_{\mathscr{F}} + 4\mu\sigma^2_1(\mathsf{C}_{(1)}) \\
&\leq \|\left\langle \mathsf{Z}, (\mathsf{C} - \mathsf{B}_0)\right\rangle\|^2_{\mathscr{F}} + 2\sigma_1(\mathscr{P}\mathbf{E})\sqrt{2R_0}\|\left\langle \mathsf{Z}, (\widehat{\mathsf{B}}_0 - \mathsf{B}_0)\right\rangle\|_{\mathscr{F}} \\
&\quad + 2\sigma_1(\mathscr{P}\mathbf{E})\sqrt{2R_0}\|\left\langle \mathsf{Z}, (\mathsf{C} - \mathsf{B}_0)\right\rangle\|_{\mathscr{F}} \\
&\quad + 2\mu\lambda^{-2}\|\left\langle \mathsf{Z}, (\widehat{\mathsf{B}}_0 - \mathsf{B}_0)\right\rangle\|^2_{\mathscr{F}} + 2\mu\lambda^{-2}\|\left\langle \mathsf{Z}, (\mathsf{C} - \mathsf{B}_0)\right\rangle\|^2_{\mathscr{F}} + 4\mu\sigma^2_1(\mathsf{C}_{(1)}) \\
&\overset{(\star)}{\leq} 4\mu\sigma^2_1(\mathsf{C}_{(1)}) + \|\left\langle \mathsf{Z}, (\mathsf{C} - \mathsf{B}_0)\right\rangle\|^2_{\mathscr{F}} \\
&\quad + 2R_0 a\sigma^2_1(\mathscr{P}\mathbf{E}) + \|\left\langle \mathsf{Z}, (\widehat{\mathsf{B}}_0 - \mathsf{B}_0)\right\rangle\|^2_{\mathscr{F}}/a + 2\mu\lambda^{-2}\|\left\langle \mathsf{Z}, (\widehat{\mathsf{B}}_0 - \mathsf{B}_0)\right\rangle\|^2_{\mathscr{F}} \\
&\quad + 2R_0 b\sigma^2_1(\mathscr{P}\mathbf{E}) + \|\left\langle \mathsf{Z}, (\mathsf{C} - \mathsf{B}_0)\right\rangle\|^2_{\mathscr{F}}/b + 2\mu\lambda^{-2}\|\left\langle \mathsf{Z}, (\mathsf{C} - \mathsf{B}_0)\right\rangle\|^2_{\mathscr{F}} \\
&\leq 4\mu\sigma^2_1(\mathsf{C}_{(1)}) + 2(a+b)R_0\sigma^2_1(\mathscr{P}\mathbf{E}) \\
&\quad + \left(\frac{b+1}{b} + 2\mu\lambda^{-2}\right)\|\left\langle \mathsf{Z}, (\mathsf{C} - \mathsf{B}_0)\right\rangle\|^2_{\mathscr{F}} + \left(\frac{1}{a} + 2\mu\lambda^{-2}\right)\|\left\langle \mathsf{Z}, (\widehat{\mathsf{B}}_0 - \mathsf{B}_0)\right\rangle\|^2_{\mathscr{F}}
\end{aligned}
\qquad (4.35)
$$

Therefore, by doing some algebra, we have the following.

$$
\begin{aligned}
\left(\frac{a-1}{a} - 2\mu\lambda^{-2}\right)\|\left\langle \mathsf{Z}, (\widehat{\mathsf{B}}_0 - \mathsf{B}_0\right\rangle\|^2_{\mathscr{F}} &\leq 4\mu\sigma^2_1(\mathsf{C}_{(1)}) + 2(a+b)R_0\sigma^2_1(\mathscr{P}\mathbf{E}) \\
&\quad + \left(\frac{b+1}{b} + 2\mu\lambda^{-2}\right)\|\left\langle \mathsf{Z}, (\mathsf{C} - \mathsf{B}_0)\right\rangle\|^2_{\mathscr{F}}
\end{aligned}
$$

$$\| \left\langle \mathsf{Z}, (\widehat{\mathsf{B}}_0 - \mathsf{B}_0) \right\rangle \|_{\mathscr{F}}^2 \leq \left(\mathscr{C}(\delta)^{-1} - 2\mu\lambda^{-2} \right)^{-1} \left\{ 4\mu\sigma_1^2(\mathsf{C}) + 2(1+\delta)R_0\sigma_1^2(\mathscr{P}\mathbf{E}) \right\} + \left(\frac{\mathscr{C}(\delta) + 2\mu\lambda^{-2}}{\mathscr{C}(\delta)^{-1} - 2\mu\lambda^{-2}} \right) \| \langle \mathsf{Z}, (\mathsf{C} - \mathsf{B}) \rangle \|_{\mathscr{F}}^2 \tag{4.36}$$

where $\mathscr{C}(\delta) = 1 + 2/\delta$ and $\sigma_1(\mathsf{C}) = \max\{\sigma_1(\mathsf{C}_{(1)}), \sigma_1(\mathsf{C}_{(2)}), \sigma_1(\mathsf{C}_{(3)})\}$. The last inequality holds after choosing $a = 1 + \delta/2$ and $b = \delta/2$. Now, it is enough to provide an upper bound of the largest singular value of $\mathscr{P}\mathbf{E}$. For some positive constant C_0, with high probability $1 - C_0 N^{-a\tau}$, by Lemma 4.8.2,

$$\| \left\langle \mathsf{Z}, (\widehat{\mathsf{B}}_0 - \mathsf{B}_0) \right\rangle \|_{\mathscr{F}}^2 \leq \left(\mathscr{C}(\delta)^{-1} - 2\mu\lambda^{-2} \right)^{-1} \left\{ 4\mu\sigma_1^2(\mathsf{C}) + 2R_0(1+\delta)Q^2\xi^2 N^{2\tau+2} J \right\} + \left(\frac{\mathscr{C}(\delta) + 2\mu\lambda^{-2}}{\mathscr{C}(\delta)^{-1} - 2\mu\lambda^{-2}} \right) \| \langle \mathsf{Z}, (\mathsf{C} - \mathsf{B}_0) \rangle \|_{\mathscr{F}}^2 \tag{4.37}$$

Since C is an arbitrary matrix with $\text{rank}(\mathsf{B}) \leq R_0$, the choosing $\mathsf{C} = \mathsf{B}_0$, we have,

$$\| \left\langle \mathsf{Z}, (\widehat{\mathsf{B}}_0 - \mathsf{B}_0) \right\rangle \|_{\mathscr{F}}^2 \leq \left(\mathscr{C}(\delta)^{-1} - 2\mu\lambda^{-2} \right)^{-1} \left\{ 4\mu\sigma_1^2(\mathsf{C}) + 2R_0(1+\delta)Q^2\xi^2 N^{2\tau+2} J \right\} \tag{4.38}$$

The estimation bound can be derived from the above expression under condition $\lambda_{\min}(\mathsf{Z}_{(1)}^{\mathrm{T}}\mathsf{Z}_{(1)}) \geq \lambda$, from inequality 4.38, we have

$$\|\widehat{\mathsf{B}}_0 - \mathsf{B}_0\|_{\mathscr{F}}^2 \leq \lambda^{-1} \left(\mathscr{C}(\delta)^{-1} - 2\mu\lambda^{-2} \right)^{-1} \left\{ 4\mu\sigma_1^2(\mathsf{C}) + 2R_0(1+\delta)Q^2\xi^2 N^{2\tau+2} J \right\} \tag{4.39}$$

4.8.3 Proof of Theorem 4.3.2

Observe that, due to Lemma 4.8.3 and the fact that,

$$\int_{\mathscr{T}} [\mathbb{B}(t)^{\mathrm{T}}(\widehat{\mathsf{B}}_0 - \mathsf{B}_0)]^2 f_T(t)dt = (\widehat{\mathsf{B}}_0 - \mathsf{B}_0)^{\mathrm{T}} \left(\int_{\mathscr{T}} \mathbb{B}_h(t)\mathbb{B}_h(t)^{\mathrm{T}} f_T(t)dt \right) (\widehat{\mathsf{B}}_0 - \mathsf{B}_0) \propto \|\widehat{\mathbf{B}}_0 - \mathbf{B}_0\|^2 = O_P(a_N) \tag{4.40}$$

Therefore,

$$\begin{aligned} &\int_{\mathscr{T}} (\widehat{\beta}_\bullet(t) - \beta_\bullet(t))^2 f_T(t)dt \\ &= \int_{\mathscr{T}} \left(\mathbb{B}(t)^{\mathrm{T}}(\widehat{\mathsf{B}}_0 - \mathsf{B}_0) + \mathbb{B}(t)^{\mathrm{T}}\mathsf{B}_0 - \beta(t) \right)^2 f_T(t)dt \\ &\leq 2\|\widehat{\mathsf{B}}_0 - \mathsf{B}_0\|_{\mathscr{F}}^2 + 2\int_{\mathscr{T}} [\mathbb{B}(t)^{\mathrm{T}}\mathsf{B}_0 - \beta(t)]^2 f_T(t)dt \\ &\leq O_P(a_N) + O(K_N^{-2(\nu+1)}) \end{aligned} \tag{4.41}$$

CHAPTER 5

EPILOGUE

5.1 Conclusions

In this dissertation, we develop new methods and theories for functional data analysis with dependence and complex structures that are often produced in neuroimaging studies. Classical statistical approaches either do not adequately take advantage of dependence or are not applicable to data with complex structures. The first part of this dissertation (Chapters 2 and 3) focuses on improving existing non-parametric methods via incorporating dependence in functional data. In Chapter 2 we develop an efficient and robust estimation technique based on quadratic inference for a coefficient vector in a constant linear effects model with dense functional responses. The proposed method uses a data-driven approach to construct bases, which avoids the possible mis-specification and improves estimation efficiency. Then, in Chapter 3 we develop a multi-step estimation procedure for estimating non-parametric coefficient function in a varying-coefficient linear model with heteroskedastic errors. This method incorporates the dependence via a local linear generalized method of moments based on continuous moment conditions. In the second part of the dissertation (Chapter 4) we develop a new approach for studying neural correlates in the presence of tensor-valued brain images and tensor-valued predictors. We consider a time-varying tensor regression model where the inherent structural composition of responses and covariates is preserved. Extensive simulation studies and real data analyses are conducted to justify the efficacy of the proposed methods.

5.2 Future directions

We would like to provide some possible directions with feasible applications to work on in the future, which directly follow from this dissertation.

1. Can quadratic inference approach improve the efficiency of parameter estimation for longitudinal tensor data?

- In many applications, we are convinced that brain images appear as structured data. Therefore, scientists are often interested in analyzing such complex structured longitudinal data which are observed at finitely many time-points or clusters due to the number of visits of patients in the clinic. Zhang et al. (2014) proposed a unifying regression framework with tensor-variate image predictor with tensor-GEE for longitudinal imaging analysis. The proposed GEE approach takes into account the intra-subject correlation responses where a low-rank tensor decomposition of the coefficient array becomes effective during estimation and prediction. Similarly to the existing problem of GEE in classical longitudinal analysis, the estimation of the parameter tensor becomes inconsistent when the working correlation structure is miss-specified. Therefore, a quadratic inference-based method can be developed for tensor-GEE to improve the efficiency of the tensor-variate regression model.

2. Can a time-varying tensor approach improve FDR control for fMRI data?

- In the statistics and neuroimaging literature, false discovery rate (FDR) is commonly used for inference. A straightforward application of FDR violates the complex data features, thereby producing unsatisfactory performance. Brown and Behrmann (2017) correctly observed that the overstatement of Eklund et al. (2016) regarding FDR cast doubt on fMRI technique for studying brain function and caused damage to the field of cognitive neuroscience. Subsequently, Cox et al. (2017); Kessler et al. (2017) offered several clarifications. Among other issues, these PNAS papers recognized that accounting for the spatio-temporal aspects is utterly important for fMRI and is a remarkable methodology for understanding brain function and its relationship to behavioral characteristics. Therefore, based on the model proposed in Chapter 4, one can propose a new thresholding technique for the multiple hypothesis testing problem for spatially dependent tests over continuum null hypotheses to identify activated voxels over time based on a tensor-structured "statistical parametric map", which can be an interesting development.

Apart from the above two immediate future directions, the methodologies presented in this

dissertation will provide essential tools for analysis of such complex structured data not only in neuroimaging studies, but also in other scientific areas such as network analysis, recommendation systems and statistical genetics.

APPENDIX

APPENDIX

PRE-PROCESSING OF IMAGING DATA USING R

The main goal of studying fMRI data is to identify brain areas activated by a task. Scanner drift may occur when the strength of the magnetic field inside the bore changes slowly over time during a scan session. Therefore, it is not recommended to conduct any statistical analyses using the raw data coming from the scanner. As a result, an important role is played by the pre-processing steps of fMRI data, which consist of all required transformations needed to prepare the data for analysis. In statistics, noise is the fundamental uncertainty, but sometimes it consists of systematic variability and it is possible to remove the noise from the data. Hence, the main goal of these pre-processing steps is to remove the systematic variability that can arise due to movement of the head during an experiment, size of the brain, etc. which are mainly sources of variability due to artifacts.

Pre-processing of the fMRI data is almost similar for all kinds of experiment and typically involves a number of steps such as aligning the functional and structural scans, correcting the possible head movements, skull stripping, registration to the template, and smoothing to reduce noise; although, the type of smoothing depends on the objective of the statistical analysis. In most of the cases, the fMRIs are in NIfTI (Neuroimaging Informatics Technology Initiative) format with extension of the file "*.nii" and this can be read as a multi-dimensional array. We read the data in R and perform all required pre-processing steps using some tools and packages in R such as the *fslr* (Jenkinson et al., 2012; Muschelli et al., 2015). Interested readers are encouraged to study Ashby (2011); Wager and Lindquist (2015) for further details in the pre-processing steps.

The six commonly used pre-processing steps are performed in the order as listed below.

1. **Slice timing correction:** This method corrects the variability in the BOLD responses that is due to the fact that data in different voxels are acquired at different times. This step is performed using the function "*slicetimer*" where indexing is from top, the order of acquisition is continuous, and the interpolation using this function is done using "sinc" filter. For example,

in *ForrestGump-data*, the brain image consists of 35 slices and the replication time is 2 seconds. Therefore, the time between acquisition of the first and last slices should be almost the same as 2 seconds. Moreover, the variability in the data due to differences in the times of slice acquirement can be reduced by this pre-processing step using interpolation and/or by analyzing task-related activation using the flexible *hrf* model. Later, the *bias_correct* function is used for bias field corrections.

2. **Motion correction:** This step is performed to correct for variability due to head movement. Motion correction is a special case of image registration where a series of images are aligned by considering mean image over all time-points as the target image for each individuals. This is one of the most important steps of pre-processing. When a subject moves their head, a specific brain region either moves to a different region or out of the scanning area. This correction procedure is based on the assumption that the shape and size of the brain remain intact irrespective of the subject moving their head. Therefore, the rigid body registration (Ashburner and Friston, 2007) method can be applied. It is easy to visualize that any rigid body movement can be described by six parameters. When a subject lies inside the scanner, the center of any voxel in their head occupies a point in space that can be characterized by the triplet (x, y, z). By convention, the Z-axis is parallel to the bore of the magnet, the X-axis passes through the subject's ears from left to right side and the Y-axis is a pole that enters through the back of the head and exits in the forehead. Based on this coordinate system, possible rigid body movements are translations and rotations along the X, Y and Z axes. BOLD responses at the mean over different TR is taken as the standard and then rigid body transformation is performed for all TRs until each of the data-sets agree as closely as possible with the mean data. Motion-corrected images have the same dimension, voxel spacing, origin, and direction as those of images collected from the scanner. Here we use "*antsrMotionCalculation*" function which provides an R-wrapper around the Insight Segmentation and Registration Toolkit (ITK). A rigid body transformation for the calculation of frame-wise motion parameters is illustrated in Figure 4.3 where first three parameters

contain the rotation matrix (rotation along the X, Y and Z axes, respectively) and the other three parameters are translation vectors (translation along the X, Y and Z axes, respectively) at each TR.

3. **Co-registration:** Since the functional images are collected with low spatial resolution, and the voxel size is much smaller than the structural images, in this step we align the functional and structural images of the pre-processing steps. We perform brain extraction of T1-weighted images (after skull stripping and bias-correction since brain activity is restricted to brain tissues only; therefore, brain extraction of the anatomical image and inhomogeneity correction must be performed to remove artifacts). The method of co-registration is similar to motion correction, rather simpler, since it involves only two images to be aligned, but challenge is that voxels are no longer one-to-one between two images, and the functional and structural images are run with different imaging parameters (or modalities), as a result of which their contrasts are different. In this step, only the structural image (mostly the mean image) and any one of the functional images must align, since all functional images are already in alignment after the head movement corrections. We use the function "*registration*" to register the average fMRI with spatial resolution of $3.0 \times 3.0 \times 3.0 mm^3$ to an $0.7 \times 0.67 \times 0.67 mm^3$ T1-weighted image by applying affine transformation and non-linear registration using the symmetric normalization (SyN) algorithm in Advance Normalizing Tools (ANTs). Note that the resulting images have the same dimension, voxel spacing, origin, and direction as those of the anatomical coordinate systems.

4. **Normalization:** This step is performed to wrap the structural images in the standard brain atlas. There exist huge disparities between two brains in terms of size and shape across the subjects. Therefore, it is difficult to make decisions regarding task-related activation in order to figure which voxel/region of the brain is activated in a subject. The common practice is to map the data onto a "standard brain" for which coordinates of all major brain areas have already been discovered and henceforth determine the activated region in a brain

atlas. Here we use an atlas in Section 4.6 proposed by Montreal Neurological Institute (MNI) where the MNI-atlas was created by averaging the results from high-resolution structural images taken over 152 different brains with dimension $182 \times 218 \times 182$ with pixel dimension $1 \times 1 \times 1\text{mm}^3$ and it is also provided in *FSL* as MNI152_T1_1mm_brain. Furthermore, the functional brain atlas provides information about the location of the functional brain region, obtaining knowledge on the brain functionality. The steps of the normalization process are the following.

- Co-registering the T1-weighted image into the coordinate system of the MNI space.
- Co-registering functional and T1-weighted structural images.
- Applying the calculated non-linear forward transformation from previous steps to project fMRI time-series to MNI space.

Here we use "*registration*" and "*antsApplyTransforms*" accordingly.

5. **Spatial smoothing:** This step reduces high-frequency noise that changes rapidly in small regions of the brain. As a result of local averaging due to spatial smoothing, the resulting images become blurred due to the reduction of intensity of BOLD responses. The standard choice of the kernel function in this smoothing method is the Gaussian kernel, which is centered at the smoothing location and $\sigma_\bullet^2$ is the width of the kernel in "$\bullet$" direction. In the context of fMRI data analysis, this is termed as full width at half maximum (FWHM) which is the width of the interval between the points at which the height of the kernel along "$\bullet$" direction is half of the peak height. In our data processing, we take (6, 6, 7) as kernel width (FWHM).

6. **Temporal filtering:** This is done to reduce the effect of slow fluctuations in the local magnetic field properties of the scanner. This pre-processing step smooths the data at each voxel across neighboring TRs. Like spatial smoothing, the goal here is to reduce noise and, thereafter, make it easier to identify signal. We perform de-noising by regressing BOLD time-series

on nuisance regressors including 6 principal component scores (obtained using "*compcor*" function (Behzadi et al., 2007)) and motion parameters obtained from motion correction step. In the final stage, we perform high-pass filtering using the "*frequencyFilterfMRI*" function with the lower and the higher frequency limits in band-pass filter 0.01 and 0.1 respectively.

- **Additional step - extraction of ROI:** This is an additional step that is performed as per requirement for a given data. It is well-known that functional brain atlases provide information about the location of functional brain regions. One of the recent atlases is that of Shen286 (Shen et al., 2013; Finn et al., 2015) and related website https://sites.google.com/dartmouth.edu/canlab-brainpatterns/brain-atlases-and-parcellations/ which provides a brain parcellation into 268 regions that were obtained as a result of a resting-state fMRI study. The T1-weighted image of *Shen268* atlas, with size $181 \times 217 \times 181$ and dimensions $1 \times 1 \times 1 mm^3$ is commonly used. This image is then transformed into the MNI space as described in the normalization step. Symmetric Normalization (SyN) is performed using Advance Normalising Tools (ANTs), furthermore, deformable registration using the "*registration*" function in *fslr* package of R is performed. Estimated transformation is used to wrap the atlas to the associated MNI space, and labels are created by the nearest-neighbor interpolation method. In addition, for DTI data, the JHU-ICBM-FA-1mm template with dimension $182 \times 218 \times 182$ where the pixel dimension is $1 \times 1 \times 1$ with 50 ROIs is commonly used (Mori et al., 2008).

BIBLIOGRAPHY

BIBLIOGRAPHY

Ai, C. and Chen, X. (2003). Efficient estimation of models with conditional moment restrictions containing unknown functions. *Econometrica*, 71(6):1795–1843.

Albert, R. and Barabási, A.-L. (2002). Statistical mechanics of complex networks. *Reviews of modern physics*, 74(1):47.

Allen, G. I., Grosenick, L., and Taylor, J. (2014). A generalized least-square matrix decomposition. *Journal of the American Statistical Association*, 109(505):145–159.

Anderson, T. W., Anderson, T. W., Anderson, T. W., Anderson, T. W., and Mathématicien, E.-U. (1958). *An introduction to multivariate statistical analysis*, volume 2. Wiley New York.

Ashburner, J. and Friston, K. J. (2007). Rigid body registration. *Statistical parametric mapping: The analysis of functional brain images*, pages 49–62.

Ashby, F. G. (2011). *Statistical analysis of fMRI data*. MIT press.

Aue, A., Norinho, D. D., and Hörmann, S. (2015). On the prediction of stationary functional time series. *Journal of the American Statistical Association*, 110(509):378–392.

Bai, Y., Zhu, Z., and Fung, W. K. (2008). Partial linear models for longitudinal data based on quadratic inference functions. *Scandinavian Journal of Statistics*, 35(1):104–118.

Balan, R. M., Schiopu-Kratina, I., et al. (2005). Asymptotic results with generalized estimating equations for longitudinal data. *The Annals of Statistics*, 33(2):522–541.

Beck, A. and Teboulle, M. (2009). A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM journal on imaging sciences*, 2(1):183–202.

Behzadi, Y., Restom, K., Liau, J., and Liu, T. T. (2007). A component based noise correction method (compcor) for bold and perfusion based fmri. *Neuroimage*, 37(1):90–101.

Berrendero, J. R., Justel, A., and Svarc, M. (2011). Principal components for multivariate functional data. *Computational Statistics & Data Analysis*, 55(9):2619–2634.

Bi, X., Tang, X., Yuan, Y., Zhang, Y., and Qu, A. (2021). Tensors in statistics. *Annual review of statistics and its application*, 8:345–368.

Bongiorno, E. G., Salinelli, E., Goia, A., and Vieu, P. (2014). *Contributions in infinite-dimensional statistics and related topics*. Societa Editrice Esculapio.

Bravo, F. (2021). Second order expansions of estimators in nonparametric moment conditions models with weakly dependent data. *Econometric Reviews*, pages 1–24.

Brown, E. N. and Behrmann, M. (2017). Controversy in statistical analysis of functional magnetic resonance imaging data. *Proceedings of the National Academy of Sciences*, 114(17):E3368–E3369.

Cai, T. T. and Hall, P. (2006). Prediction in functional linear regression. *The Annals of Statistics*, 34(5):2159–2179.

Cai, Z., Das, M., Xiong, H., and Wu, X. (2006). Functional coefficient instrumental variables models. *Journal of Econometrics*, 133(1):207–241.

Cai, Z. and Li, Q. (2008). Nonparametric estimation of varying coefficient dynamic panel data models. *Econometric Theory*, 24(5):1321–1342.

Cardot, H., Degras, D., and Josserand, E. (2013). Confidence bands for horvitz–thompson estimators using sampled noisy functional data. *Bernoulli*, 19(5A):2067–2097.

Cardot, H., Ferraty, F., and Sarda, P. (1999). Functional linear model. *Statistics & Probability Letters*, 45(1):11–22.

Cardot, H., Ferraty, F., and Sarda, P. (2003). Spline estimators for the functional linear model. *Statistica Sinica*, pages 571–591.

Cardot, H. and Sarda, P. (2008). Varying-coefficient functional linear regression models. *Communications in statistics—theory and methods*, 37(20):3186–3203.

Carroll, J. D. and Chang, J.-J. (1970). Analysis of individual differences in multidimensional scaling via an n-way generalization of "eckart-young" decomposition. *Psychometrika*, 35(3):283–319.

Casey, B., Soliman, F., Bath, K. G., and Glatt, C. E. (2010). Imaging genetics and development: challenges and promises. *Human Brain Mapping*, 31(6):838–851.

Castro, P. E., Lawton, W. H., and Sylvestre, E. (1986). Principal modes of variation for processes with continuous sample curves. *Technometrics*, 28(4):329–337.

Chen, K., Dong, H., and Chan, K.-S. (2013). Reduced rank regression via adaptive nuclear norm penalization. *Biometrika*, 100(4):901–920.

Chen, K. and Müller, H.-G. (2012). Modeling repeated functional observations. *Journal of the American Statistical Association*, 107(500):1599–1609.

Chen, X., Li, H., Liang, H., and Lin, H. (2019). Functional response regression analysis. *Journal of Multivariate Analysis*, 169:218–233.

Chiou, J.-M., Chen, Y.-T., and Yang, Y.-F. (2014). Multivariate functional principal component analysis: A normalization approach. *Statistica Sinica*, pages 1571–1596.

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: Series B (Statistical Methodology)*, 65(2):405–423.

Chiou, J.-M., Müller, H.-G., and Wang, J.-L. (2004). Functional response models. *Statistica Sinica*, pages 675–693.

Conn, A. R., Gould, N. I., and Toint, P. L. (2000). *Trust region methods*, volume 1. Siam.

Cox, R. W., Chen, G., Glen, D. R., Reynolds, R. C., and Taylor, P. A. (2017). fmri clustering and false-positive rates. *Proceedings of the National Academy of Sciences*, 114(17):E3370–E3371.

Cragg, J. G. (1983). More efficient estimation in the presence of heteroscedasticity of unknown form. *Econometrica: Journal of the Econometric Society*, pages 751–763.

Crainiceanu, C. M., Staicu, A.-M., and Di, C.-Z. (2009). Generalized multilevel functional regression. *Journal of the American Statistical Association*, 104(488):1550–1561.

Craven, P. and Wahba, G. (1978). Smoothing noisy data with spline functions. *Numerische mathematik*, 31(4):377–403.

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(1):136–154.

De Boor, C., De Boor, C., Mathématicien, E.-U., De Boor, C., and De Boor, C. (1978). *A practical guide to splines*, volume 27. springer-verlag New York.

Delaigle, A. and Hall, P. (2010). Defining probability density for a distribution of random functions. *The Annals of Statistics*, pages 1171–1193.

Diggle, P., Diggle, P. J., Heagerty, P., Liang, K.-Y., Heagerty, P. J., Zeger, S., et al. (2002). *Analysis of longitudinal data*. Oxford University Press.

Ding, X., Yu, D., Zhang, Z., and Kong, D. (2021). Multivariate functional response low-rank regression with an application to brain imaging data. *Canadian Journal of Statistics*, 49(1):150–181.

Ding, X. and Zhou, Z. (2020). Estimation and inference for precision matrices of nonstationary time series. *The Annals of Statistics*, 48(4):2455–2477.

Douglas Nychka, Reinhard Furrer, John Paige, and Stephan Sain (2017). fields: Tools for spatial

data. R package version 12.3.

Dunford, N. and Schwartz, J. T. (1988). *Linear operators, part 1: general theory*, volume 10. John Wiley & Sons.

Eklund, A., Nichols, T. E., and Knutsson, H. (2016). Cluster failure: Why fmri inferences for spatial extent have inflated false-positive rates. *Proceedings of the national academy of sciences*, 113(28):7900–7905.

Eubank, R. L. (1999). *Nonparametric regression and spline smoothing*. CRC press.

Fan, J. and Gijbels, I. (1996). *Local polynomial modelling and its applications*. Chapman & Hall/CRC.

Fan, J., Yao, Q., and Cai, Z. (2003). Adaptive varying-coefficient linear models. *Journal of the Royal Statistical Society: series B (statistical methodology)*, 65(1):57–80.

Fan, J. and Zhang, W. (2008). Statistical methods with varying coefficient models. *Statistics and its Interface*, 1(1):179.

Fan, J., Zhang, W., et al. (1999). Statistical estimation in varying coefficient models. *The annals of Statistics*, 27(5):1491–1518.

Fang, E. X., Ning, Y., Li, R., et al. (2020). Test of significance for high-dimensional longitudinal data. *Annals of Statistics*, 48(5):2622–2645.

Fang, Y., Loparo, K. A., and Feng, X. (1994). Inequalities for the trace of matrix product. *IEEE Transactions on Automatic Control*, 39(12):2489–2490.

Faraway, J. J. (1997). Regression analysis for a functional response. *Technometrics*, 39(3):254–261.

Ferraty, F. and Vieu, P. (2006). *Nonparametric functional data analysis: theory and practice*. Springer Science & Business Media.

Ferré, L. and Yao, A.-F. (2005). Smoothed functional inverse regression. *Statistica Sinica*, pages 665–683.

Finn, E. S., Shen, X., Scheinost, D., Rosenberg, M. D., Huang, J., Chun, M. M., Papademetris, X., and Constable, R. T. (2015). Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. *Nature neuroscience*, 18(11):1664–1671.

Fransson, P., Flodin, P., Seimyr, G. Ö., and Pansell, T. (2014). Slow fluctuations in eye position and resting-state functional magnetic resonance imaging brain activity during visual fixation. *European Journal of Neuroscience*, 40(12):3828–3835.

Gahrooei, M. R., Yan, H., Paynabar, K., and Shi, J. (2018). A novel approach for fusion of heterogeneous sources of data. *arXiv preprint arXiv:1803.00138*.

Gajardo, A., Carroll, C., Chen, Y., Dai, X., Fan, J., Hadjipantelis, P. Z., Han, K., Ji, H., Mueller, H.-G., and Wang, J.-L. (2021). *fdapace: Functional Data Analysis and Empirical Dynamics*. R package version 0.5.7.

Givens, G. H. and Hoeting, J. A. (2012). *Computational statistics*, volume 703. John Wiley & Sons.

Goldsmith, J., Bobb, J., Crainiceanu, C. M., Caffo, B., and Reich, D. (2011). Penalized functional regression. *Journal of computational and graphical statistics*, 20(4):830–851.

Goldsmith, J., Crainiceanu, C. M., Caffo, B., and Reich, D. (2012). Longitudinal penalized functional regression for cognitive outcomes on neuronal tract measurements. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 61(3):453–469.

Goldsmith, J., Scheipl, F., Huang, L., Wrobel, J., Di, C., Gellar, J., Harezlak, J., McLean, M. W., Swihart, B., Xiao, L., Crainiceanu, C., and Reiss, P. T. (2020). *refund: Regression with Functional Data*. R package version 0.1-23.

Goldsmith, J., Zipunnikov, V., and Schrack, J. (2015). Generalized multilevel function-on-scalar regression and principal component analysis. *Biometrics*, 71(2):344–353.

Green, P. J. and Silverman, B. W. (1993). *Nonparametric regression and generalized linear models: a roughness penalty approach*. CRC Press.

Grenander, U. (1950). Stochastic processes and statistical inference. *Arkiv för matematik*, 1(3):195–277.

Greven, S. and Scheipl, F. (2017). A general framework for functional regression modelling. *Statistical Modelling*, 17(1-2):1–35.

Guhaniyogi, R., Qamar, S., and Dunson, D. B. (2017). Bayesian tensor regression. *The Journal of Machine Learning Research*, 18(1):2733–2763.

Guhaniyogi, R. and Spencer, D. (2018). Bayesian tensor response regression with an application to brain activation studies. Technical report, Technical report, UCSC. 2, 13.

Guo, W., Kotsia, I., and Patras, I. (2012). Tensor learning for regression. *IEEE Transactions on Image Processing*, 21(2):816–827.

Hadinejad-Mahram, H., Dahlhaus, D., and Blomker, D. (2002). Karhunen-loéve expansion of vector random processes. Technical report, Technical Report No. IKTNT 1019, Communications Technological Laboratory.

Hall, A. R. (2004). *Generalized method of moments*. OUP Oxford.

Hall, P., Horowitz, J. L., et al. (2007). Methodology and convergence rates for functional linear regression. *The Annals of Statistics*, 35(1):70–91.

Hall, P. and Hosseini-Nasab, M. (2006). On properties of functional principal components analysis. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(1):109–126.

Hall, P. and Hosseini-Nasab, M. (2009). Theory for high-order bounds in functional principal components analysis. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 146, page 225. Cambridge University Press.

Hall, P., Müller, H.-G., and Wang, J.-L. (2006). Properties of principal component methods for functional and longitudinal data analysis. *The annals of statistics*, pages 1493–1517.

Hand, D. and Crowder, M. (2017). *Practical longitudinal data analysis*. Routledge.

Hanke, M., Baumgartner, F. J., Ibe, P., Kaule, F. R., Pollmann, S., Speck, O., Zinke, W., and Stadler, J. (2014). A high-resolution 7-tesla fmri dataset from complex natural stimulation with an audio movie. *Scientific data*, 1:140003.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica: Journal of the Econometric Society*, pages 1029–1054.

Happ, C. and Greven, S. (2018). Multivariate functional principal component analysis for data observed on different (dimensional) domains. *Journal of the American Statistical Association*, 113(522):649–659.

Hardle, W. (1990). *Applied Nonparametric Regression*. Cambridge University Press.

Harshman, R. A. et al. (1970). Foundations of the PARAFAC procedure: Models and conditions for an “explanatory" multimodal factor analysis. Technical report, University of California at Los Angeles Los Angeles, CA.

Hastie, T. and Tibshirani, R. (1993). Varying-coefficient models. *Journal of the Royal Statistical Society: Series B (Methodological)*, 55(4):757–779.

Hastie, T. J. and Tibshirani, R. J. (2017). *Generalized additive models*. Routledge.

Häusler, C. O. and Hanke, M. (2016). An annotation of cuts, depicted locations, and temporal progression in the motion picture “forrest gump”. *F1000Research*, 5.

He, G., Muller, H., and Wang, J.-L. (2018). Extending correlation and regression from multivariate to functional data. In *Asymptotics in statistics and probability*, pages 197–210. De Gruyter.

He, G., Müller, H.-G., and Wang, J.-L. (2003). Functional canonical analysis for square integrable stochastic processes. *Journal of Multivariate Analysis*, 85(1):54–77.

Hedeker, D. and Gibbons, R. D. (2006). *Longitudinal data analysis*, volume 451. John Wiley & Sons.

Hitchcock, F. L. (1928). Multiple invariants and generalized rank of a p-way matrix or tensor. *Journal of Mathematics and Physics*, 7(1-4):39–79.

Hoff, P. D. (2011). Hierarchical multilinear models for multiway data. *Computational Statistics & Data Analysis*, 55(1):530–543.

Hoff, P. D. (2015). Multilinear tensor regression for longitudinal relational data. *The annals of applied statistics*, 9(3):1169.

Hoff, P. D. et al. (2011). Separable covariance arrays via the tucker product, with applications to multivariate relational data. *Bayesian Analysis*, 6(2):179–196.

Hoover, D. R., Rice, J. A., Wu, C. O., and Yang, L.-P. (1998). Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. *Biometrika*, 85(4):809–822.

Hörmann, S. and Kokoszka, P. (2010). Weakly dependent functional data. *The Annals of Statistics*, 38(3):1845–1884.

Horváth, L. and Kokoszka, P. (2012). *Inference for functional data with applications*, volume 200. Springer Science & Business Media.

Hsing, T. and Eubank, R. (2015). *Theoretical foundations of functional data analysis, with an introduction to linear operators*, volume 997. John Wiley & Sons.

Huang, H., Li, Y., and Guan, Y. (2014). Joint modeling and clustering paired generalized longitudinal trajectories with application to cocaine abuse treatment data. *Journal of the American Statistical Association*, 109(508):1412–1424.

Huang, J. Z., Wu, C. O., and Zhou, L. (2002). Varying-coefficient models and basis function approximations for the analysis of repeated measurements. *Biometrika*, 89(1):111–128.

Huang, J. Z., Wu, C. O., and Zhou, L. (2004). Polynomial spline estimation and inference for varying coefficient models with longitudinal data. *Statistica Sinica*, pages 763–788.

Huettel, S. A., Song, A. W., McCarthy, G., et al. (2004). *Functional magnetic resonance imaging*, volume 1. Sinauer Associates Sunderland, MA.

Ivanescu, A. E., Staicu, A.-M., Scheipl, F., and Greven, S. (2015). Penalized function-on-function regression. *Computational Statistics*, 30(2):539–568.

J Mercer, B. (1909). Xvi. functions of positive and negative type, and their connection the theory of integral equations. *Phil. Trans. R. Soc. Lond. A*, 209(441-458):415–446.

James, G. M., Hastie, T. J., and Sugar, C. A. (2000). Principal component models for sparse functional data. *Biometrika*, 87(3):587–602.

Jenkinson, M., Beckmann, C. F., Behrens, T. E., Woolrich, M. W., and Smith, S. M. (2012). Fsl. *Neuroimage*, 62(2):782–790.

Ji, Y., Wang, Q., Li, X., and Liu, J. (2019). A survey on tensor techniques and applications in machine learning. *IEEE Access*, 7:162950–162990.

Johan, H. (1990). Tensor rank is np-complete. *Journal of Algorithms*, 4(11):644–654.

Karhunen, K. (1946). Zur spektraltheorie stochastischer prozesse. *Ann. Acad. Sci. Fennicae, AI*, 34.

Kato, K. (2012). Estimation in functional linear quantile regression. *The Annals of Statistics*, 40(6):3108–3136.

Kessler, D., Angstadt, M., and Sripada, C. S. (2017). Reevaluating "cluster failure" in fmri using nonparametric control of the false discovery rate. *Proceedings of the National Academy of Sciences*, 114(17):E3372–E3373.

Kiers, H. A. (2000). Towards a standardized notation and terminology in multiway analysis. *Journal of Chemometrics: A Journal of the Chemometrics Society*, 14(3):105–122.

Kim, J. S., Maity, A., and Staicu, A.-M. (2018). Additive nonlinear functional concurrent model. *Statistics and its interface*, 11(4):669–685.

Kinson, C., Tang, X., Zuo, Z., and Qu, A. (2020). Longitudinal principal component analysis with an application to marketing data. *Journal of Computational and Graphical Statistics*, 29(2):335–350.

Kokoszka, P. and Reimherr, M. (2017). *Introduction to functional data analysis*. Chapman and Hall/CRC.

Kolda, T. and Bader, B. (2009). Tensor decompositions and applications. *SIAM Review*, 51(3):455–500.

Kowal, D. R., Matteson, D. S., and Ruppert, D. (2017). A bayesian multivariate functional dynamic linear model. *Journal of the American Statistical Association*, 112(518):733–744.

Kuenzer, T., Hörmann, S., and Kokoszka, P. (2021). Principal component analysis of spatially indexed functions. *Journal of the American Statistical Association*, 116(535):1444–1456.

Li, Y. and Guan, Y. (2014). Functional principal component analysis of spatiotemporal point processes with applications in disease surveillance. *Journal of the American Statistical Association*, 109(507):1205–1215.

Li, Y. and Hsing, T. (2007). On rates of convergence in functional linear regression. *Journal of Multivariate Analysis*, 98(9):1782–1804.

Li, Y. and Hsing, T. (2010). Uniform convergence rates for nonparametric regression and principal component analysis in functional/longitudinal data. *The Annals of Statistics*, 38(6):3321–3351.

Li, Y., Qiu, Y., and Xu, Y. (2022). From multivariate to functional data analysis: Fundamentals, recent developments, and emerging areas. *Journal of Multivariate Analysis*, 188:104806.

Liang, K.-Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, 73(1):13–22.

Liang, X., Zou, T., Guo, B., Li, S., Zhang, H., Zhang, S., Huang, H., and Chen, S. X. (2015). Assessing beijing's pm2. 5 pollution: severity, weather impact, apec and winter heating. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 471(2182):20150257.

Lindquist, M. A. (2008). The statistical analysis of fmri data. *Statistical science*, 23(4):439–464.

Liu, B. and Müller, H.-G. (2009). Estimating derivatives for samples of sparsely observed functions, with application to online auction dynamics. *Journal of the American Statistical Association*, 104(486):704–717.

Liu, T., Yuan, M., and Zhao, H. (2017). Characterizing spatiotemporal transcriptome of human brain via low rank tensor decomposition. *arXiv preprint arXiv:1702.07449*.

Liu, Y., Liu, J., and Zhu, C. (2020). Low-rank tensor train coefficient array estimation for tensor-on-tensor regression. *IEEE Transactions on Neural Networks and Learning Systems*.

Lock, E. F. (2018). Tensor-on-tensor regression. *Journal of Computational and Graphical Statistics*, 27(3):638–647.

Loève, M. (1946). Functions aleatoire de second ordre. *Revue science*, 84:195–206.

Long, X., Liao, W., Jiang, C., Liang, D., Qiu, B., and Zhang, L. (2012). Healthy aging: an automatic analysis of global and regional morphological alterations of human brain. *Academic radiology*, 19(7):785–793.

Lu, C. and Wooldridge, J. M. (2020). A gmm estimator asymptotically more efficient than ols and wls in the presence of heteroskedasticity of unknown form. *Applied Economics Letters*, 27(12):997–1001.

Mallows, C. L. (1973). Some comments on cp. *Technometrics*, 15(4):661–675.

McCullagh, P. and Nelder, J. (1989). Generalized linear models ii.

Melzer, A., Härdle, W. K., and Cabrera, B. L. (2019). Joint tensor expectile regression for electricity day-ahead price curves. *Available at SSRN 3363167*.

Mori, S., Oishi, K., Jiang, H., Jiang, L., Li, X., Akhter, K., Hua, K., Faria, A. V., Mahmood, A., Woods, R., et al. (2008). Stereotaxic white matter atlas based on diffusion tensor imaging in an icbm template. *Neuroimage*, 40(2):570–582.

Morris, J. S. (2015). Functional regression. *Annual Review of Statistics and Its Application*, 2:321–359.

Müller, H.-G. (2016). Peter hall, functional data analysis and random objects. *The Annals of Statistics*, 44(5):1867–1887.

Müller, H.-G. and Stadtmüller, U. (2005). Generalized functional linear models. *the Annals of Statistics*, 33(2):774–805.

Müller, H.-G. and Yao, F. (2010). Empirical dynamics for longitudinal data. *The Annals of Statistics*, 38(6):3458–3486.

Muschelli, J., Sweeney, E., Lindquist, M., and Crainiceanu, C. (2015). fslr: Connecting the fsl software with r. *The R Journal*, 7(1):163–175.

Newey, W. K. (1990). Efficient instrumental variables estimation of nonlinear models. *Econometrica: Journal of the Econometric Society*, pages 809–837.

Newey, W. K. and McFadden, D. (1994). Large sample estimation and hypothesis testing. *Handbook of econometrics*, 4:2111–2245.

Ombao, H., Lindquist, M., Thompson, W., and Aston, J. (2016). *Handbook of neuroimaging data analysis*. Chapman and Hall/CRC.

O'Donnell, L. J. and Westin, C.-F. (2011). An introduction to diffusion tensor image analysis. *Neurosurgery Clinics*, 22(2):185–196.

Park, S. Y. and Staicu, A.-M. (2015). Longitudinal functional data analysis. *Stat*, 4(1):212–226.

Porter, D., Stirling, D. S., and David, P. (1990). *Integral equations: a practical treatment, from spectral theory to applications*, volume 5. Cambridge university press.

Qu, A. and Li, R. (2006). Quadratic inference functions for varying-coefficient models with longitudinal data. *Biometrics*, 62(2):379–391.

Qu, A., Lindsay, B. G., and Li, B. (2000). Improving generalised estimating equations using quadratic inference functions. *Biometrika*, 87(4):823–836.

Ramsay, J. O. (1982). When the data are functions. *Psychometrika*, 47(4):379–396.

Ramsay, J. O. and Dalzell, C. (1991). Some tools for functional data analysis. *Journal of the Royal Statistical Society: Series B (Methodological)*, 53(3):539–561.

Ramsay, J. O. and Silverman, B. W. (2005). *Functional data analysis*. Springer series in statistics.

Ramsay, J. O. and Silverman, B. W. (2007). *Applied functional data analysis: methods and case studies*. Springer.

Rao, C. R. (1958). Some statistical methods for comparison of growth curves. *Biometrics*, 14(1):1–17.

Raskutti, G., Yuan, M., and Chen, H. (2019). Convex regularization for high-dimensional multiresponse tensor regression. *Ann. Statist.*, 47(3):1554–1584.

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: Series B (Methodological)*, 53(1):233–243.

Rice, J. A. and Wu, C. O. (2001). Nonparametric mixed effects models for unequally sampled noisy curves. *Biometrics*, 57(1):253–259.

Riss, F. and Sz-Nagy, B. (1990). *Functional analysis*. Dover Publications Inc.

Ruppert, D., Wand, M. P., and Carroll, R. J. (2003). *Semiparametric regression*. Cambridge university press.

Scheipl, F., Staicu, A.-M., and Greven, S. (2015). Functional additive mixed models. *Journal of Computational and Graphical Statistics*, 24(2):477–501.

Schumaker, L. (2007). *Spline functions: basic theory*. Cambridge University Press.

Sengupta, A., Kaule, F. R., Guntupalli, J. S., Hoffmann, M. B., Häusler, C., Stadler, J., and Hanke, M. (2016). A studyforrest extension, retinotopic mapping and localization of higher visual areas. *Scientific data*, 3:160093.

Shen, X., Tokoglu, F., Papademetris, X., and Constable, R. T. (2013). Groupwise whole-brain parcellation from resting-state fmri data for network node identification. *Neuroimage*, 82:403–415.

Sidiropoulos, N. D. and Bro, R. (2000). On the uniqueness of multilinear decomposition of n-way

arrays. *Journal of Chemometrics: A Journal of the Chemometrics Society*, 14(3):229–239.

Staicu, A.-M., Islam, M. N., Dumitru, R., and Heugten, E. v. (2020). Longitudinal dynamic functional regression. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 69(1):25–46.

Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society: Series B (Methodological)*, 36(2):111–133.

Su, L., Murtazashvili, I., and Ullah, A. (2013). Local linear gmm estimation of functional coefficient iv models with an application to estimating the rate of return to schooling. *Journal of Business & Economic Statistics*, 31(2):184–207.

Sun, Y. (2016). Functional-coefficient spatial autoregressive models with nonparametric spatial weights. *Journal of Econometrics*, 195(1):134–153.

Taris, T. (2000). *Longitudinal data analysis*. Sage.

Tian, R., Xue, L., and Liu, C. (2014). Penalized quadratic inference functions for semiparametric varying coefficient partially linear models with longitudinal data. *Journal of Multivariate Analysis*, 132:94–110.

Tran, K. C. and Tsionas, E. G. (2009). Local gmm estimation of semiparametric panel data with smooth coefficient models. *Econometric Reviews*, 29(1):39–61.

Tucker, L. R. (1958). Determination of parameters of a functional relation by factor analysis. *Psychometrika*, 23(1):19–23.

Tucker, L. R. (1966). Some mathematical notes on three-mode factor analysis. *Psychometrika*, 31(3):279–311.

Turner, E. L., Li, F., Gallis, J. A., Prague, M., and Murray, D. M. (2017). Review of recent methodological developments in group-randomized trials: part 1—design. *American journal of public health*, 107(6):907–915.

Vaart, A. W. and Wellner, J. A. (1996). Weak convergence. In *Weak convergence and empirical processes*, pages 16–28. Springer.

van Delft, A. and Eichler, M. (2018). Locally stationary functional time series. *Electronic Journal of Statistics*, 12(1):107–170.

Van Hecke, W., Emsell, L., and Sunaert, S. (2016). *Diffusion tensor imaging: a practical handbook*. Springer.

Viviani, R., Grön, G., and Spitzer, M. (2005). Functional principal component analysis of fmri

data. *Human brain mapping*, 24(2):109–129.

Wager, T. D. and Lindquist, M. A. (2015). Principles of fmri. *New York: Leanpub*.

Wahba, G. (1990). *Spline models for observational data*, volume 59. Siam.

Wand, M. and Jones, M. (1995). Kernel smoothing chapman & hall. *Monographs on Statistics and Applied Probability, London*.

Wang, J.-L., Chiou, J.-M., and Müller, H.-G. (2016). Functional data analysis. *Annual Review of Statistics and Its Application*, 3:257–295.

Wang, L. (2008). *Karhunen-Loéve expansions and their applications*. London School of Economics and Political Science (United Kingdom).

Wang, L., Li, H., and Huang, J. Z. (2008). Variable selection in nonparametric varying-coefficient models for analysis of repeated measurements. *Journal of the American Statistical Association*, 103(484):1556–1569.

Wasserman, L. (2006). *All of nonparametric statistics*. Springer Science & Business Media.

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

Xiong, Y., Zhou, X. J., Nisi, R. A., Martin, K. R., Karaman, M. M., Cai, K., and Weaver, T. E. (2017). Brain white matter changes in cpap-treated obstructive sleep apnea patients with residual sleepiness. *Journal of Magnetic Resonance Imaging*, 45(5):1371–1378.

Xu, Y., Li, Y., and Nettleton, D. (2018). Nested hierarchical functional data modeling and inference for the analysis of functional plant phenotypes. *Journal of the American Statistical Association*, 113(522):593–606.

Xue, L., Shu, X., and Qu, A. (2018). Time-varying estimation and dynamic model selection with an application of network data. *Statistica Sinica*.

Yao, F. and Lee, T. C. (2006). Penalized spline models for functional principal component analysis. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(1):3–25.

Yao, F., Müller, H.-G., Clifford, A. J., Dueker, S. R., Follett, J., Lin, Y., Buchholz, B. A., and Vogel, J. S. (2003). Shrinkage estimation for functional principal component scores with application to the population kinetics of plasma folate. *Biometrics*, 59(3):676–685.

Yao, F., Müller, H.-G., and Wang, J.-L. (2005). Functional linear regression analysis for longitudinal data. *The Annals of Statistics*, pages 2873–2903.

Yao, W. and Li, R. (2013). New local estimation procedure for a non-parametric regression function for longitudinal data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 75(1):123–138.

Yu, H., Tong, G., and Li, F. (2020). A note on the estimation and inference with quadratic inference functions for correlated outcomes. *Communications in Statistics-Simulation and Computation*, pages 1–12.

Yu, K. and Jones, M. (2004). Likelihood-based local linear estimation of the conditional variance function. *Journal of the American Statistical Association*, 99(465):139–144.

Zhang, C., Peng, H., and Zhang, J.-T. (2010). Two samples tests for functional data. *Communications in Statistics—Theory and Methods*, 39(4):559–578.

Zhang, J.-T. (2013). *Analysis of variance for functional data*. Chapman and Hall/CRC.

Zhang, J.-T. and Chen, J. (2007). Statistical inferences for functional data. *The Annals of Statistics*, 35(3):1052–1079.

Zhang, L. and Banerjee, S. (2021). Spatial factor modeling: A bayesian matrix-normal approach for misaligned data. *Biometrics*.

Zhang, X., Li, L., Zhou, H., Shen, D., et al. (2014). Tensor generalized estimating equations for longitudinal imaging analysis. *arXiv preprint arXiv:1412.6592*.

Zhang, X. and Wang, J.-L. (2016). From sparse to dense functional data and beyond. *The Annals of Statistics*, 44(5):2281–2321.

Zhao, M., Gao, Y., and Cui, Y. (2020). Variable selection for longitudinal varying coefficient errors-in-variables models. *Communications in Statistics-Theory and Methods*, pages 1–26.

Zhao, M., Xu, X., Zhu, Y., Zhang, K., and Zhou, Y. (2021). Model estimation and selection for partial linear varying coefficient ev models with longitudinal data. *Journal of Applied Statistics*, pages 1–23.

Zheng, X., Xue, L., and Qu, A. (2018). Time-varying correlation structure estimation and local-feature detection for spatio-temporal data. *Journal of Multivariate Analysis*, 168:221–239.

Zhou, H., Li, L., and Zhu, H. (2013). Tensor regression with applications in neuroimaging data analysis. *Journal of the American Statistical Association*, 108(502):540–552. PMID: 24791032.

Zhou, J. and Qu, A. (2012). Informative estimation and selection of correlation structure for longitudinal data. *Journal of the American Statistical Association*, 107(498):701–710.

Zhou, L., Huang, J. Z., and Carroll, R. J. (2008). Joint modelling of paired sparse functional data

using principal components. *Biometrika*, 95(3):601–619.

Zhou, L., Huang, J. Z., Martinez, J. G., Maity, A., Baladandayuthapani, V., and Carroll, R. J. (2010). Reduced rank mixed effects models for spatially correlated hierarchical functional data. *Journal of the American Statistical Association*, 105(489):390–400.

Zhu, H., Fan, J., and Kong, L. (2014). Spatially varying coefficient model for neuroimaging data with jump discontinuities. *Journal of the American Statistical Association*, 109(507):1084–1098.

Zhu, H., Li, R., and Kong, L. (2012). Multivariate varying coefficient model for functional responses. *Annals of statistics*, 40(5):2634.