ABSTRACT

PARK, SO YOUNG. Longitudinal Functional Data Analysis with Biomedical Applications. (Under the direction of Ana-Maria Staicu.)

Longitudinal functional data consist of functional observations, e.g. profiles or images, collected from each of many subjects at multiple sequential instances (often visits). A major challenge in longitudinal functional data analysis is the presence of very complex dependence structure due to both within- and between-curves correlations, along with high dimensionality. This dissertation proposes novel statistical modeling and inference methods to address some fundamental scientific questions commonly arising from studies collecting longitudinal functional data, in a computationally efficient way.

In the first part of the dissertation we introduce simple inferential approaches for the population-level effects in a general functional mixed model framework. The fixed effects of scalar covariates are modeled nonparametrically while error covariance is left unspecified to avoid model complexity. We study confidence intervals and a $L^2$ norm based testing procedure for the fixed effects parameters. We estimate the fixed effects under the independence of functional residuals assumption and then use the bootstrap method over independent units (e.g. subject) to account for complex error dependence. Simulations show excellent coverage probability of the confidence intervals and size of tests. Methods are motivated by and applied to the Baltimore Longitudinal Study of Aging, though they are applicable to other studies that collect repeated functional data.

In the second part of the dissertation, we introduce a novel parsimonious modeling framework for longitudinal functional observations. The proposed method models both the dependence within the subject as well as the longitudinal design and extracts low-dimensional features. This way of modeling allows to study the dynamic behavior of
the underlying process over time, allows prediction of full trajectories at unobserved visit times, and is computationally efficient. Theoretical properties of this framework are studied and numerical investigations confirm excellent behavior in the finite sample setting. The proposed method is motivated by and applied to the Diffusion Tensor Imaging study of multiple sclerosis. Furthermore we provide interactive graphic tools in R to help explore longitudinal functional data and visualize the various model components and prediction obtained using the proposed method.

In the last part of the dissertation, we propose a novel inferential method for longitudinal functional data about the time-varying mean function. Specifically we consider the null hypothesis that the mean function does not vary over visit time. Both null and alternative hypotheses make no parametric assumptions on the structure of the mean function. We propose to represent the mean function using orthonormal basis functions, re-formulate the null hypothesis into a set of simpler null hypotheses in terms of the basis coefficients, and conduct pseudo likelihood ratio based tests. Simulation results confirm that the proposed testing procedure maintains nominal significance levels and has excellent power. The proposed testing procedure is applied to the DTI data.
Longitudinal Functional Data Analysis with Biomedical Applications

by

So Young Park

A dissertation submitted to the Graduate Faculty of North Carolina State University in partial fulfillment of the requirements for the Degree of Doctor of Philosophy

Statistics

Raleigh, North Carolina

2017

APPROVED BY:

________________________  _________________________
Marie Davidian            Soumendra Lahiri

________________________  _________________________
Luo Xiao                  Ana-Maria Staicu
Chair of Advisory Committee
DEDICATION

To my beloved family.
BIOGRAPHY

So Young Park was born and grew up in Seoul, Republic of Korea. She moved to the United States in 2001 and went to Wayland Academy, a high school in Beaver Dam, Wisconsin. After graduating from Wayland in 2005, she went to Carnegie Mellon University in Pittsburgh, Pennsylvania, where she earned her Bachelor of Science degree in Applied Mathematics with a concentration in Operations Research in May of 2009 and Masters in Statistical Practice in May of 2010. She joined the Department of Statistics at North Carolina State University in 2011 to pursue a Ph.D. degree in Statistics.
ACKNOWLEDGEMENTS

I would like to take this moment to express my deepest gratitude to my advisor, Dr. Ana-Maria Staicu for her endless support and guidance throughout my graduate studies. Her passion to research and professionalism have truly motivated me to constantly improve my academic and personal life.

I would also like to extend my appreciation to my committee members, Drs. Marie Davidian, Soumendra Lahiri, and Luo Xiao, for sharing their time and insights into this research, and to Dr. Anne McLaughlin for serving as the graduate school representative. I would like to thank Dr. Luo Xiao for providing many professional opportunities as well as financial support. I am also very grateful for all of the faculty members, staffs, and fellow graduate students in Statistics department at North Carolina State University. In particular, I want to thank my dear friends, Marcela Alfaro Cordoba, Matthew Austin, Janet Kim, Cai Li, Md Nazmul Islam, Saebitna Oh for always being supportive and making my graduate school life fun and memorable.

Last but not least, I thank my grandparents (Jae Min Kim, Jin Sook Seok, Gwang Chae Kim), my parents (Jong Duk Park, Chung Hee Kim), and my brother (Gwangyeon James Park) for their unconditional love and support throughout my life. I am really grateful to have such a wonderful, loving, and caring family; I could not have done this without them.
# TABLE OF CONTENTS

## LIST OF TABLES
- vii

## LIST OF FIGURES
- ix

### Chapter 1 Introduction
- 1
  1.1 Overview ........................................... 1
  1.2 Contributions and Outline .......................... 5
  1.3 Smoothing ........................................... 6
    1.3.1 Smoothing using Penalized Basis Expansions .... 7
  1.4 Functional Principal Component Analysis .......... 9

### Chapter 2 Simple Fixed Effects Inference for Complex Functional Models
- 12
  2.1 Introduction ........................................ 12
  2.2 Modeling framework .................................. 15
  2.3 Confidence bands for $\mu(s, x)$ .................... 20
  2.4 Hypothesis testing for $\mu(s, x)$ ................... 25
  2.5 Application to physical activity data ............. 27
  2.6 Simulation Study .................................... 32

### Chapter 3 Longitudinal Functional Data Analysis
- 38
  3.1 Introduction ........................................ 38
  3.2 Modeling longitudinal functional data ............. 41
  3.3 Estimation of model components .................... 44
    3.3.1 Step 1: Mean function .......................... 44
    3.3.2 Step 2: Marginal covariance. Data-based orthogonal basis 45
    3.3.3 Step 3: Covariance of the time-varying coefficients 46
    3.3.4 Step 4: Trajectories reconstruction ........... 48
  3.4 Implementation using available softwares .......... 48
  3.5 Theoretical properties ................................ 49
    3.5.1 Response curves measured without error ........ 50
    3.5.2 Response curves measured with smooth error .... 53
    3.5.3 Extensions ....................................... 54
  3.6 Simulation study ..................................... 55
  3.7 DTI application ....................................... 57
  3.8 Visualization using Interactive Graphics ........... 62

### Chapter 4 Significance Testing in Longitudinal Functional Data
- 66
  4.1 Introduction ........................................ 66
  4.2 Preliminaries ........................................ 69
4.3 Testing Procedure .................................................. 71
  4.3.1 Implementation ................................................. 74
  4.3.2 Pseudo Likelihood Ratio Test ................................ 75
4.4 Simulation Study .................................................. 78
  4.4.1 Data Generation ................................................ 78
  4.4.2 Details on Estimation ......................................... 79
  4.4.3 Competitive Methods ......................................... 80
  4.4.4 Numerical Results ............................................. 81
4.5 Diffusion Tensor Imaging Study ................................. 84

Chapter 5 Conclusion ................................................. 87

References .............................................................. 90

Appendices ............................................................. 104
  Appendix A  Additional Details for Chapter 2  ................... 105
    A.1 Validating the BLSA Testing Results via Simulation Study ... 105
    A.2 Additional results for the performance of the confidence bands ... 107
    A.3 Results obtained by bootstrapping observations by subject ... 109
    A.4 Result for the case of nonparametric modeling ................ 113
    A.5 Results for the case of having non-normal errors ............... 115
  Appendix B  Additional Details for Chapter 3  ................... 120
    B.1 Proofs of theoretical results given in Section 3.5 ............ 120
      B.1.1 Case when response curves are measured without error (Sec-
            tion 3.5.1) .................................................... 120
      B.1.2 Case when response curves are measured with smooth error
            (Section 3.5.2) ................................................ 154
    B.2 Additional details for the simulation experiment ............. 157
      B.2.1 Description of the study .................................. 157
      B.2.2 Additional results ......................................... 160
    B.3 Additional figures for the DTI data analysis .................. 163
    B.4 Additional Figure for Interactive Graphics .................... 166
  Appendix C  Additional Details for Chapter 4  ................... 167
    C.1 Additional Simulation Results ................................. 167
    C.2 Validating the DTI Testing Results via Simulation Study ....... 169
LIST OF TABLES

Table 2.1 Simulation results using bootstrap of subject-level residuals and 95% nominal level; results are based on 500 MC samples. ................. 36
Table 2.2 Empirical Type I error of the test statistic $T$ based on the $N_{sim} = 1000$ MC samples. ......................................................... 37
Table 3.1 Estimation and prediction accuracy results based on $N_{sim} = 1000$ simulations ................................................................. 58
Table 3.2 Comparison between the proposed method and Chen & Müller (2012) in the presence of correlated errors. Results based on $N_{sim} = 1000$ simulations ................................................................. 59
Table 4.1 The empirical type I error rates of the proposed method (pLRT-MT) based on 5000 simulations. Standard errors are presented in parentheses. 83
Table 4.2 Computation time (in seconds) based on 200 simulations when $n = 200$. 83
Table A.1 Empirical type I error of the test statistic $T$ based on the $N_{sim} = 1000$ MC samples; Mean function is $\mu(s, x) = \cos(2\pi s)$, $\tau = 0$ ............... 106
Table A.2 Simulation results using bootstrap of subject-level residuals and 85% nominal level; results are based on 500 MC samples. ................. 107
Table A.3 Simulation results using bootstrap of subject-level residuals and 90% nominal level; results are based on 500 MC samples. ................. 108
Table A.4 Simulation results using bootstrap of subject-level observations and 85% nominal level; results are based on 500 MC samples. ................. 110
Table A.5 Simulation results using bootstrap of subject-level observations and 90% nominal level; results are based on 500 MC samples. ................. 111
Table A.6 Simulation results using bootstrap of subject-level observations and 95% nominal level; results are based on 500 MC samples. ................. 112
Table A.7 Simulation results for the case of time-varying covariate $X_{ij}$ using bootstrap of subject-level observations; results are based on 500 MC samples.112
Table A.8 Simulation results fitting bivariate mean function using bootstrap of subject-level residuals and 85% nominal level; results are based on 500 MC samples. ......................................................... 113
Table A.9 Simulation results fitting bivariate mean function using bootstrap of subject-level residuals and 90% nominal level; results are based on 500 MC samples. ......................................................... 114
Table A.10 Simulation results fitting bivariate mean function using bootstrap of subject-level residuals and 95% nominal level; results are based on 500 MC samples. ......................................................... 114
Table A.11 Simulation results for the case of having non-Gaussian errors using bootstrap of subject-level residuals and 85% nominal level; results are based on 500 MC samples. ................................. 116
Table A.12 Simulation results for the case of having non-Gaussian errors using bootstrap of subject-level residuals and 90% nominal level; results are based on 500 MC samples. ................................. 117
Table A.13 Simulation results for the case of having non-Gaussian errors using bootstrap of subject-level residuals and 95% nominal level; results are based on 500 MC samples. ................................. 118
Table A.14 Empirical Type I error of the test statistic $T$ based on the $N_{\text{sim}} = 1000$ MC samples for the case of having non-Gaussian errors. ................. 119

Table B.1 Simulation results for estimation and prediction accuracy based on $N_{\text{sim}} = 1000$ simulations (white noise only) ......................... 161
Table B.2 Computational time (seconds) corresponding to table 3.1 .............. 162
Table B.3 Computational time (seconds) corresponding to table B.1 .............. 162

Table C.1 The empirical type I error rates of the proposed method based on 5000 simulations. Standard errors are presented in parentheses. .............. 168
Table C.2 The empirical type I error rates of the ZC-MT and Bootstrap-MT methods based on 5000 simulations. Standard errors are presented in parentheses. ......................... 169
Table C.3 The empirical type I error rates of the proposed test and other competitive methods based on 5000 simulations, using nominal level $\alpha = 0.05$. Standard errors are presented in parentheses. ......................... 170
LIST OF FIGURES

Figure 2.1 The null distribution of the test statistic in (2.4) for the null hypothesis that there is no effect of age on activity. The red dashed line is the 95 percent quantile of the null distribution of the test statistic.

Figure 2.2 Heat map of average of bootstrap estimates of log counts as a bivariate function of time of day and age (left panel) and average of bootstrap estimates of log counts for five different age groups (right panel).

Figure 2.3 Heat maps of joint confidence bands for the estimate $\hat{\mu}(t, Age)$ given in the left panel of Figure 2.2. The legend on the right applies to both plots.

Figure 2.4 Average of bootstrap estimates of log counts as a function of time of day at age 60 and the associated joint confidence bands.

Figure 2.5 Association of body mass index with mean log counts as a function of time of day and the associated joint confidence bands.

Figure 2.6 Estimated power curves for testing $H_0: \mu(s, x) = \eta(s)$ using level of significance $\alpha = 0.05$, when the true mean function $\mu(s, x) = 2 \cos(2\pi s) + \delta(x/4 - s)^3$ for $\delta = 0.01, 2, 4, 6$. The results are based on $N_{sim} = 500$ MC samples.

Figure 3.1 Left panel: 95% pointwise and joint confidence bands of the slope function $\beta_t(s)$ of $\mu(s, t)$ using bootstrap; Right: final mean estimate, $\hat{\mu}(s, t) = \hat{\mu}_0(s)$.

Figure 3.2 Top: First three eigenfunctions of the estimated marginal covariance; Bottom: estimated mean function $\hat{\mu}_0(s)$ (gray line) $\pm 2\sqrt{\hat{\lambda}_k \hat{\phi}_k(s)}$ (+ and $-$ signs, respectively).

Figure 3.3 Estimated time-varying coefficients $\hat{\xi}_{ik}(t)$ for $k = 1, 2$ and 3 using REM.

Figure 3.4 Predicted values of FA for the last visits of three randomly selected subjects; actual observations (gray); predictions using our model (black solid) and using the naive approach (black dashed).

Figure 3.5 Screenshot showing tab 1 of the interactive graphic. The plots show observed data of all subjects (top) and of the selected subject (bottom).

Figure 4.1 Estimated size and power curves for testing $H_0: \mu(s, t) = \mu_0(s)$ using $\alpha = 0.05$, when the true mean function $\mu(s, t) = \cos(\pi s/2) + 5\delta(t/4 - s)^3$ for $\delta = 0, 0.2, 0.4, 0.6, 0.8, 1$. Results are based on 5000 simulations for size ($\delta = 0$) and 1000 simulations for power ($\delta > 0$).
Figure A.1 Estimated power curves for testing $H_0 : \mu(s, x) = \eta(s)$ using $\alpha = 0.05$, when the true mean function $\mu(s, x) = \cos(2\pi s) + \delta(\hat{\mu}(s, x) - \cos(2\pi s))$ for $\delta = 0.01, 2, 4, 6, 8$. Results are based on $N_{sim} = 500$ MC samples.

Figure A.2 Normal, skew-normal, and t distributions used to generate white noises.

Figure B.1 Fitted varying coefficient model, $\hat{\mu}(s, t) = \hat{\mu}_0(s) + \hat{\beta}_t(s) t$ (left), and estimated slope function, $\hat{\beta}_t(s)$ (right).

Figure B.2 Scree plot of the marginal FPCA.

Figure B.3 Estimated eigenfunctions of the marginal FPCA for FA with PVE = 95%.

Figure B.4 Estimated basis coefficient functions, $\hat{\xi}_{ik}(t)$ using a random coefficients linear model.

Figure B.5 Screenshot showing tab 2 of the interactive graphic. The plots show the third estimated eigenfunction $\hat{\phi}_3(s)$ (top) and the predicted trajectory of the selected subject at (scaled) visit time $t = 0.3$ (bottom).

Figure C.1 The original DTI data (left) and one simulated data that mimic the DTI error structure and have a mean function equal to $\hat{\mu}(s, t)$. For illustration the observed profiles from the first subject are highlighted with colors.
Chapter 1

Introduction

1.1 Overview

Functional Data Analysis (FDA) is an actively growing research area in Statistics that involves data with each observation being a function instead of a single value. For illustration consider marathon data, where elapsed time was recorded at every mile of the race for each of 105 male participants who completed the 2016 U.S. Olympic Team Trial Marathon. In this example, FDA views a set of elapsed times from each participant as a function of miles and treat it as a single datum. Enabled by rapid advancement of technology in recent years, such data have become commonly collected in many scientific applications across different disciplines. Intensive statistical researches have naturally followed to address challenges posed by functional data and have made great progress on both theoretical and methodological developments in FDA. For comprehensive reviews of FDA we direct the reader to monographs, Ramsay & Silverman (2005); Ferraty & Vieu (2006); Ramsay et al. (2009); and Horváth & Kokoszka (2012) among others.

Suppose we observe the functional data \((s_{ir}, Y_{ir}) : i = 1, \ldots, n\) and \(r = 1, \ldots, R_i\),
where \( Y_{ir} \) is the \( r \)th repeated measurement for subject \( i \) corresponding to \( s_{ir} \in S \) for some compact interval \( S \). In the marathon data example, \( Y_{ir} \) corresponds to the elapsed time recorded at mile \( s_{ir} \) for participant \( i \). The key idea of FDA is to view the observed profile \( \{Y_{ir} : r = 1, \ldots, R_i\} \) as a realization \( X_i(\cdot) \) of a smooth unknown random process \( X(\cdot) \) that is observed at finite grid of points \( \{s_{ir} : r = 1, \ldots, R_i\} \) and is corrupted with measurement errors \( \{\epsilon_{ir} : r = 1, \ldots, R_i\} \). That is, the observed data are modeled as

\[
Y_{ir} = X_i(s_{ir}) + \epsilon_{ir},
\]

where \( X_i(\cdot) \)'s are independent and identically distributed (iid) square integrable random elements in \( L^2(S) \) with unknown smooth mean and covariance functions, and measurement errors \( \epsilon_{ir} \)'s are iid with mean zero and variance \( \sigma^2 \). FDA mainly considers two sampling designs: (i) \textit{dense design} where the set of sampling points for each subject, \( \{s_{ir} : r = 1, \ldots, R_i\} \), is dense in its domain \( S \) and (ii) \textit{sparse design} where the number of repeated measures per subject, \( R_i \), is relatively small and the set of sampling points for each subject, \( \{s_{ir} : r = 1, \ldots, R_i\} \), is random and irregular. In the sparse design setting, it is assumed that the set of pooled sampling points across all the subjects, \( \{s_{ir} : i = 1, \ldots, n \text{ and } r = 1, \ldots, R_i\} \), is dense in its domain \( S \).

Functional principal component analysis (FPCA), first introduced by Rao (1958), is recognized as a core technique in FDA for several reasons. First, it addresses some of the fundamental problems in FDA such as studying the variability of latent process \( X(\cdot) \) and identifying the main modes of variation. Second, it allows us to recover the subject-specific true underlying random process \( X_i(\cdot) \) from the noisy and discretely sampled observations, \( Y_{ir} \)'s; in this sense we can view FPCA as a smoothing technique. Third, it provides a parsimonious representation of the curves and reduces the dimension of
the data. We detail the FPCA method in Section 1.4 for both dense and sparse design settings. Due to a variety of its uses FPCA has been served as a building block of many methodologies in FDA: functional linear regression (Cardot et al., 2003b; Yao et al., 2005b; Kim et al., 2015), conditional FPCA (Cardot, 2007), multilevel FPCA (Di et al., 2009), and FPCA-based optimal design (Park et al., 2016b) to name a few. In Chapter 3 we also borrow the idea of FPCA and develop a novel parsimonious modeling framework for longitudinal functional data.

While tremendous progress has been made in FDA, most of existing methodologies are developed for independent functional data, where only one curve is observed from each of the subjects, as described in (1.1). It is relatively recent that repeated functional data are considered in the literature, though an increasing number of studies now routinely collect such complex-correlated functional data. A major characteristic of these data is the presence of a strong between-curves correlation induced by multiple functions observed on the same observational unit, often subject.

The existing studies on repeated functional data can be separated into two categories, based on whether or not the data consist of multiple independent realizations. Recent studies, which involve more than one realizations, have considered functional structures including multilevel (Morris et al., 2003; Morris & Carroll, 2006; Di et al., 2009; Crainiceanu et al., 2009); crossed (Aston et al., 2010; Shou et al., 2014); longitudinal (Greven et al., 2010; Chen & Müller, 2012; Scheipl et al., 2014; Chen et al., 2016); or spatially aligned (Baladandayuthapani et al., 2008; Staicu et al., 2010; Serban et al., 2013). Whereas in the case of having only one realization, i.e. having only one independent unit in the sample, a spatially correlated functional structure has been considered in Giraldo et al. (2007), Delicado et al. (2010), Nerini et al. (2010) Gromenko et al. (2012), Gromenko & Kokoszka (2013); whereas a temporally correlated functional structure has
been studied in Bosq (2000), Hyndman & Shang (2009), Horváth & Kokoszka (2012), Hormann & Kokoszka (2012).

In this dissertation we focus our study on longitudinal functional data, which consist of functional observations, e.g. profiles or images, collected from each of many subjects at multiple sequential instances (often visits). Some examples of longitudinal functional data applications, which motivated our work, include: (i) Baltimore Longitudinal Study of Aging (BLSA), where physical activity counts were recorded every minute using an accelerometer from each of participants over multiple consecutive days (Stone & Norris, 1966b; Xiao et al., 2015a; Goldsmith et al., 2015); (ii) Longitudinal diffusion tensor imaging (DTI) study, where various modality profiles along well-identified tracts were collected from each multiple sclerosis patient at several hospital visits (Greven et al., 2010; Goldsmith et al., 2012; Scheipl et al., 2014).

Scientific questions arising from these applications include: (i) to assess the effect of scalar covariates such as age and BMI on daily physical activity profiles, (ii) to understand the dynamics of modality profiles over multiple hospital visits, (iii) to predict the modality profile at a future visit based on the observed profiles in the past, and (iv) to examine whether the mean modality profile vary over hospital visits. These types of questions are not specific to these applications and often come up in any longitudinal studies collecting functional data. Yet there are very limited statistical methodologies available to address these questions. Motivated by such applications, we develop flexible and computationally efficient statistical methods for longitudinal functional data analysis.
1.2 Contributions and Outline

In Chapter 2 we propose simple inferential approaches for the fixed effects in complex functional models. We consider a modeling framework that is a direct generalization of the mixed model framework from longitudinal data analysis, where scalar responses are replaced with functional ones. We model the fixed effect of a scalar covariate non-parametrically while error covariance is left unspecified to avoid model complexity. We estimate the fixed effects under the independence of functional residuals assumption and then bootstrap independent units (e.g. subject) to estimate the variability of and conduct inference in the form of hypothesis testing on the fixed effects parameters. Although the estimation is based on the working independence assumption the proposed inferential methods rely on the bootstrap of subjects that accounts for the complex dependence. Simulations show excellent coverage probability of the confidence intervals and size of tests. Methods are motivated by and applied to the Baltimore Longitudinal Study of Aging, though they are applicable to other studies that collect correlated functional data.

In Chapter 3 we propose a novel parsimonious modeling framework for longitudinal functional observations that allows to extract low-dimensional features. The proposed methodology accounts for the longitudinal dependence, is designed to study the dynamic behavior of the underlying process, allows prediction of full future trajectory at unobserved visit time, and is computationally feasible. Theoretical properties of this framework are studied, and numerical investigations confirm excellent behavior in the finite sample setting. The proposed method is motivated by and applied to the Diffusion Tensor Imaging study of multiple sclerosis. In addition we provide a interactive graphic tool in R for exploring any longitudinal functional data and visualize the model components and prediction obtained using the proposed method.
In Chapter 4 we develop a novel pseudo likelihood ratio based inferential method for longitudinal functional data about the mean function. Specifically we consider testing the null hypothesis that the mean function does not vary over the actual time of visit. In both null and alternative hypotheses, no parametric assumptions are made on the structure of the mean function. We first model the mean function using orthonormal basis functions and use the basis coefficients to re-formulate the null hypothesis into a set of simpler null hypotheses that can be assessed with common existing testing procedures. Simulation results confirm that the proposed testing procedure maintains nominal significance levels and has excellent power. The proposed test is applied to the Diffusion Tensor Imaging study of multiple sclerosis.

The rest of this chapter provides a brief overview of smoothing and FPCA that are commonly used in FDA and are recurrently employed in subsequent chapters.

1.3 Smoothing

Consider model (1.1). Conceptually FDA deals with a smooth random function $X_i(\cdot)$ and studies its properties such as trend, variation, and derivatives. The smoothness assumption on $X_i(\cdot)$ is what differentiates FDA from other statistical frameworks. However in practice $X_i(\cdot)$ is latent and needs to be recovered from its noisy and discrete realization, $\{Y_{ir} : r = 1, \ldots, R_i\}$. In this regard smoothing plays an important role in FDA for two main reasons: (i) to remove measurement errors, $\epsilon_{ir}$’s, and (ii) to interpolate between discretely sampled observations, i.e. to evaluate a function at an unobserved sampling point, $s \notin \{s_{ir} : r = 1, \ldots, R_i\}$. Various smoothing techniques have been developed and are available in nonparametric and semi-parametric regression literatures; see for example Hastie & Tibshirani (1990), Eubank (1999), Ruppert et al. (2003), Wood (2006a),
1.3.1 Smoothing using Penalized Basis Expansions

Here we provide a brief overview of univariate smoothing using penalized basis expansions, which is used throughout this dissertation; see Wahba (1990), Ruppert et al. (2003) and Wood (2006b) for more details. Consider model (1.1). Suppose we want to smooth one observed profile of the $i$th subject, \{Y_{ir} : r = 1, \ldots, R_i\}, that is densely sampled with noises. The main idea is to represent the underlying smooth process $X_i(s)$ using a linear combination of a set of known basis functions defined on $S$, say \{$B_1(x), \ldots, B_K(x)$\}, for sufficiently large $K$ to capture the complexity of the data, and then control the smoothness of fit by imposing a roughness penalty. For simplicity we drop the index $i$.

Specifically we estimate the smooth function $X(\cdot)$ using the following penalized estimation criterion: $PSSE_\lambda(X) = \sum_{r=1}^{R_i} \{Y_r - X(s_r)\}^2 + \lambda \mathbf{P}\{X(\cdot)\}$, where the first term is the sum of squared residuals which measures goodness of fit; $\mathbf{P}\{X(\cdot)\}$ is a penalty term which measures roughness of the function $X(\cdot)$; and $\lambda \geq 0$ is the smoothing parameter which controls trade-off between bias (i.e. goodness of fit) and variance (i.e. smoothness of fit). The most common choice of the penalty term is $\mathbf{P}\{X(\cdot)\} = \int \{X''(s)\}^2 ds$, where $X''(s)$ is the second derivative of $X_i(s)$; this penalty term quantifies the curvature of $X(\cdot)$, i.e. measures any deviation from linearity.

Now we represent $X(s)$ using preset basis functions as $X(s) = \sum_{k=1}^{K} B_k(s)\beta_k$ for sufficiently large $K$, where $\beta_k$ is a unknown fixed basis coefficient corresponding to $B_k(s)$. It follows that the penalty term is equal to $\mathbf{P}\{X(\cdot)\} = \sum_{k=1}^{K} \sum_{k'=1}^{K} \beta_k \beta_{k'} \int B_k''(s)B_{k'}''(s)ds$, where $B_k''(s)$ is the second derivative of $B_k(s)$. Then the penalized criterion can be re-written as $PSSE_\lambda(\beta) = \sum_{r=1}^{R} \{Y_r - \sum_{k=1}^{K} B_k(s_r)\beta_k\}^2 + \lambda \beta^T \mathbf{P} \beta$, where $\beta = (\beta_1, \ldots, \beta_K)^T$.
and $P$ is the $K \times K$ penalty matrix with the $(k, k')$ element equal to $\int B''_k(s) B''_{k'}(s) ds$. Notice that when $\lambda$ is equal to zero, the penalized estimation criterion $PSSE_\lambda(\beta)$ reduces to the sum of squared residuals where no regularization is imposed; whereas with very large $\lambda$ the resulting fit $\hat{X}(s)$ becomes closer to a straight line due to heavy penalization on deviation from linearity. Several selection methods are available for selecting the optimal value of the smoothing parameter $\lambda$; for example, generalized cross-validation (GCV), maximum likelihood (ML), and restricted maximum likelihood (REML) to name a few. These selection methods have been thoroughly surveyed and compared in the literature; see for example Wahba (1990), Ruppert et al. (2003), Wood (2006b), Reiss & Ogden (2007) and Reiss & Todd Ogden (2009).

A type of basis function, $B_k(s)$, is usually chosen based on a feature of data. For example, Fourier basis is often used for periodic data; spline basis for non-periodic data; and wavelet basis for locally spiky data. In this dissertation we use a spline basis function, namely truncated power polynomials and B-splines. In general, a spline with order $d + 1$ is defined as a piecewise polynomial function of degree $d$ on each of the subintervals determined by knots at which its derivatives up to $d - 1$ order are continuous. The number of knots often depends on the number of basis functions and the degree of polynomials used. We place knots at equally spaced quantiles of the observed sampling points, as recommended in Ruppert et al. (2003) and Wood (2006a).

B-splines, which we use throughout this dissertation, have been the most popular choice in FDA (Ullah & Finch, 2013). They are computationally faster and more stable than other choices of spline basis functions because B-splines are defined such that they are locally non-zero for at most $d + 1$ adjacent subintervals (De Boor, 1978). In addition, we use truncated power polynomials in Chapter 4 for their intuitive and simple form that is often appreciated when developing a testing procedure for nested hypotheses (see
Crainiceanu & Ruppert (2004) and Staicu et al. (2014a) for example).

Detailed reviews on smoothing spline can be found in De Boor (1978), Wahba (1990), Ruppert et al. (2003), and Wood (2006a) among others.

### 1.4 Functional Principal Component Analysis

This section details the FPCA method for noisy functional data.

Consider the model (1.1). Let \( \mu(s) = E\{X(s)\} \) and \( \Sigma(s, s') = Cov\{X(s), X(s')\} \) be the mean and covariance functions of \( X(s) \). Recall that \( X(s) \) is assumed to be a square integrable random function in \( L^2(S) \), i.e. \( \int E\{X(s)\}^2 ds < \infty \); it implies that the covariance function \( \Sigma(s, s') \) is symmetric and positive definite. Then by Mercer’s theorem, the covariance function \( \Sigma(s, s') \) admits spectral decomposition \( \Sigma(s, s') = \sum_{\ell=1}^{\infty} \lambda_{\ell} \phi_{\ell}(s) \phi_{\ell}(s') \) for non-negative positive eigenvalues \( \lambda_1 \geq \lambda_2 \geq \ldots \geq 0 \) and the corresponding orthonormal eigenfunctions \( \phi_{\ell}(s) \)'s (Bosq, 2000). Using the eigenfunctions \( \phi_{\ell}(s) \)'s, we can represent \( X(s) \) as \( X(s) = \mu(s) + \sum_{\ell=1}^{\infty} \xi_{\ell} \phi_{\ell}(s) \) via Karhunen-Loève (KL) expansion, where \( \xi_{\ell} = \int \{X(s) - \mu(s)\} \phi_{\ell}(s) ds \). Here \( \xi_{\ell} \)'s are uncorrelated random variable with mean zero and variance equal to \( E(\xi_{\ell}^2) = \lambda_{\ell} \) and are referred to as FPC scores.

We can use a truncated version of the KL expansion \( X^L(s) = \mu(s) + \sum_{\ell=1}^{L} \xi_{\ell} \phi_{\ell}(s) \) to approximate \( X(s) \). The truncation value \( K \) can be selected based on various criteria, including Akaike Information Criterion (AIC) and cross-validation; see Yao et al. (2005a) and Li et al. (2013) for reference. In this dissertation we select \( K \) based on pre-specified percentage of variance (PVE), say \( \kappa \); that is, \( L \) is chosen as the smallest integer that satisfies \( \sum_{k=1}^{K} \lambda_{\ell} / \sum_{k=1}^{\infty} \lambda_{\ell} \geq \kappa \) (Di et al., 2009; Staicu et al., 2010).

Estimation of model components depends on the sampling designs. In the case of dense design, we first de-noise the data by smoothing each of the observed profiles, \( \{Y_{ir} : r = \)
1, \ldots, R_i\}, using a univariate smoother and denote by \( \hat{X}_i(s) \) the reconstructed smooth profile. Then the estimated mean and covariance functions are obtained by calculating
\[
\hat{\mu}(s) = \frac{1}{n} \sum_{i=1}^{n} \hat{X}_i(s) \quad \text{and} \quad \hat{\Sigma}(s, s') = \frac{1}{n} \sum_{i=1}^{n} \hat{X}_i(s) \hat{X}_i(s')/(n-1),
\]
respectively. We can also obtain the estimated eigen-components \( \{\hat{\lambda}_\ell, \hat{\phi}_\ell(s)\}_\ell \) from the spectral decomposition of \( \hat{\Sigma}(s, s') \) and the estimated FPC scores \( \hat{\xi}_\ell = \sum_{r=1}^{R_i} \{\hat{X}_i(s_{ir}) - \hat{\mu}(s_{ir})\} \hat{\phi}_\ell(s_{ir})(s_{ir} - s_{i(r-1)}) \) using numerical integration.

In the case of sparse design, the number of observations for each subject, \( R_i \), is small and neither smoothing each of the observed profiles nor using numerical integration for FPC score is suitable. We follow the method proposed by Yao et al. (2005a). The mean function \( \mu(s) \) is estimated by smoothing the pooled observations, \( \{Y_{ir} : i = 1, \ldots, n \text{ and } r = 1, \ldots, R_i\} \), using a univariate smoother. Let \( \tilde{Y}_{ir} = Y_{ir} - \hat{\mu}(s_{ir}) \), where \( \hat{\mu}(s) \) is the estimated mean function. For the covariance estimation, we first calculate the raw sample covariance \( \hat{\Sigma}(s_r, s_{r'}) = \sum_{i=1}^{n} \tilde{Y}_{ir} \tilde{Y}_{ir'}/n \) and then obtain the estimated covariance function \( \hat{\Sigma}(s, s') \) by smoothing \( \{\hat{\Sigma}(s_r, s_{r'}) : r \neq r'\} \) using a bivariate smoother; here the diagonal terms \( \{\hat{\Sigma}(s_r, s_{r'}) : r = r'\} \) are removed because they are inflated by the error variance \( \sigma^2 \). Again, the estimated eigen-components \( \{\hat{\lambda}_\ell, \hat{\phi}_\ell(s)\}_\ell \) are obtained from the spectral decomposition of \( \hat{\Sigma}(s, s') \). Lastly the FPC scores are predicted using conditional estimation by assuming \( \xi_\ell \) and \( \epsilon_{ij} \) are jointly Gaussian; that is, \( \hat{\xi}_\ell = \hat{E}(\xi_\ell|\mathbf{Y}_i) = \hat{\lambda}_\ell \hat{\phi}_\ell^T \hat{\Sigma}^{-1}_Y (\mathbf{Y}_i - \bar{\mu}_i) \), where \( \mathbf{Y}_i = (Y_{i1}, \ldots, Y_{iR_i})^T, \bar{\mu}_i = (\hat{\mu}(s_{i1}), \ldots, \hat{\mu}(s_{iR_i}))^T, \hat{\phi} = (\hat{\phi}_\ell(s_{i1}), \ldots, \hat{\phi}_\ell(s_{iR_i}))^T \). Here \( \hat{\Sigma}_Y \) is the \( R_i \times R_i \) dimensional matrix with \( (r, r') \)th element equal to \( \hat{\Sigma}(s_r, s_{r'}) + \hat{\sigma}^2 \mathbf{1}(r = r') \), where \( \hat{\sigma}^2 \) is the estimated variance obtained by calculating \( \int \{\hat{\Sigma}(s, s) - \hat{\Sigma}(s, s)\} ds \).

Hastie et al. (2009), and Peng & Paul (2009) for the sparse design.
Chapter 2

Simple Fixed Effects Inference for Complex Functional Models

2.1 Introduction

While repeated functional data can have highly complex dependence structures, one is often interested in simple, population-level, questions for which the multi-layered structure of the correlation is just an infinite-dimensional nuisance parameter. For example, in the Baltimore Study of Aging (BLSA) activity data are collected for each participant at the minute level for multiple days. Thus, data exhibit complex within-day (the circadian rhythm of daily activity) and between-day (the circadian rhythm of activity across days for the same subject) correlations. However, the most important questions in the BLSA tend to be simple; in particular, one may be interested in how age affects the circadian rhythm of activity or whether the effect is different by gender. In this context the high complexity and size of the data are just technical inconveniences.

Such simple questions are typically answered by estimating fixed effects in complex
functional mixed effects models. Our alternative proposal avoids the heavy associated computations by: 1) estimating the fixed (population-level) effects under the assumption of independence of functional residuals; and 2) using a nonparametric bootstrap of independent units (e.g. subjects) to construct confidence intervals and conduct tests. A natural question is whether efficiency is lost by ignoring the correlation. While the loss of efficiency is well documented in longitudinal studies with few observations per subject and small dimensional within-subject correlation, little is known about inference when there are many observations per subject with an unknown large dimensional within-subject correlation matrix. Our own view is that estimating large dimensional covariance matrices of functional data to estimate fixed effects may actually waste degrees of freedom. Indeed, a covariance matrix for an \( n \times p \) matrix of functional data (\( n = \) number of subjects and \( p = \) number of subject-specific observations) would require estimation of \( p(p + 1)/2 \) matrix covariance entries. When \( p \) is moderate or large and the covariance matrix is unstructured this is a difficult problem. Moreover, the resulting matrix has an unknown low rank and is not invertible.

We will consider cases when multiple functional observations are observed for the same subject. This structure is inspired by many current observational studies, but we will focus on the BLSA, where activity data are recorded at the minute level over multiple days, resulting in daily activity profiles observed over multiple days. Consider that the observed data is of the form \( \{Y_{ij}(\cdot), X_{ij}\} \), where \( Y_{ij}(\cdot) \) is the \( j \)th unit functional response (e.g. \( j \)th visit) for the \( i \)th subject, and \( X_{ij} \) is the corresponding vector of covariates. This general form applies to all types of functional data discussed above: multilevel, longitudinal, spatially-correlated, crossed, etc. The main objective is to make statistical inference for the population-level effects of interest.

One naïve approach to analyze data with such complex structure is to ignore the de-
pendence over the functional argument \( s \), but to account for the dependence across the repeated visits; specifically by assuming that responses \( Y_{ij}(s) \) are correlated over \( j \) and independent over \( s \). Longitudinal data analysis literature offers a wide variety of models and methods for estimating the fixed effects and their uncertainty, and for conducting tests (see for example Laird & Ware (1982); Liang & Zeger (1986); Fitzmaurice et al. (2012)). These methods allow to account for within-subject correlation, incorporate additional covariates, and make inference about the fixed effects. Nevertheless, extending these estimation procedures to functional data is difficult because specifying the dependence for functional data is not obvious while implementation may be very computationally expensive.

Another possible approach is to completely ignore the dependence across the repeated visits \( j \), but account for the functional dependence; specifically by assuming \( Y_{ij}(s) \) are dependent over \( s \), but independent over \( j \). Function on scalar/vector regression models can be used to estimate the fixed effects of interest; see for example Faraway (1997); Jiang et al. (2011); Ivanescu et al. (2014). In this context, testing procedures for hypotheses on fixed effects are available. For example, Shen & Faraway (2004) proposed the functional F statistic for testing hypotheses related to nested functional linear models. Zhang & Chen (2007) proposed \( L_2 \) norm based test for testing the effect of a linear combination of time-varying coefficients, and approximate the null sampling distribution using resampling methods. However failing to account for all sources of dependence results in tests with inflated type I error.

In contrast, development of statistical inference methods for correlated functional data has received less attention. For example, Morris & Carroll (2006) discussed Bayesian inference in the functional mixed model framework; however, their main focus was on modeling, and hypothesis testing was not studied. Crainiceanu et al. (2012) discussed
bootstrap-based inferential methods for the difference in the mean profiles. Staicu et al. (2014a) proposed likelihood-ratio type testing procedure, while Staicu et al. (2014b) considered $L_2$ norm-based testing procedures for testing the null hypothesis that multiple group mean functions are equal. Horváth et al. (2013) developed inference for the mean function of a functional time series. Nevertheless, none of these papers handle inference on fixed effects in full generality. Here we consider a modeling framework that is a direct generalization of the linear mixed model framework from longitudinal data analysis, where scalar responses are replaced with functional ones. We study confidence intervals and testing procedures for the fixed effects using bootstrap methods over subjects to account for all known sources of data dependence.

The rest of this chapter is organized as follows. Section 2.2 introduces the modeling and estimation framework and discusses several important examples. Section 2.3 describes an approach to quantifying the variability of the estimators using bootstrap. Section 2.4 proposes formal test procedure for the null hypothesis that the mean function does not depend on a covariate of interest. Applications and simulation results are presented in Sections 2.5 and 2.6, respectively.

### 2.2 Modeling framework

Consider the case when each subject is observed at $m_i$ visit times, and data at each visit consist of a functional outcome $\{Y_{ijr} = Y_{ij}(s_{ijr}) : r = 1, \ldots, R_{ij}\}$ and a vector of covariates including a scalar covariate of interest, $X_{ij}$, and additional $p$-dimensional vector of covariates, $Z_{ij}$. We assume that $s_{ijr} \in S$ for compact and closed domain $S$; take $S = [0, 1]$ for simplicity. For convenience, we assume a balanced regular sampling design, i.e. $s_{ijr} = s_r$ and $R_{ij} = R$, though all methods apply to general sampling designs.
Furthermore, we assume that \{X_{ij} : \forall i,j\} is a dense set in the closed domain \(\mathcal{X}\); this assumption is needed for the case when the fixed effect of \(X_{ij}\) is modeled nonparametrically (Ruppert et al., 2003; Fitzmaurice et al., 2012). A common approach for the study of the effect of the covariates on the functional outcome \(Y_{ij}(\cdot)\) is to posit a model of the type

\[
Y_{ij}(s) = \mu(s, X_{ij}) + Z_{ij}^T \tau + \epsilon_{ij}(s),
\]

(2.1)

where \(\mu(s, X_{ij})\) is a time-varying smooth effect of the covariate of interest, \(X_{ij}\), and \(\tau\) is a \(p\)-dimensional parameter quantifying the linear additive effect of the covariate vector, \(Z_{ij}\). Here \(\epsilon_{ij}(s)\) is a zero-mean random deviation that incorporates both the within- and between-subject variability. Below we present several examples of models for \(\mu(s, X_{ij})\) that are relevant to our problem:

2.2(a) \(\mu(s, X_{ij}) = \beta_0 + \beta_s s + \beta_x X_{ij}\)

2.2(b) \(\mu(s, X_{ij}) = \beta_0 + \beta_s s + \beta_x X_{ij} + \beta_{sx} s X_{ij}\)

2.2(c) \(\mu(s, X_{ij}) = f(s) + \beta_x X_{ij}\), where \(f(\cdot)\) is an unknown smooth function

2.2(d) \(\mu(s, X_{ij}) = h(s, X_{ij})\), where \(h(\cdot, \cdot)\) is an unknown bivariate smooth function

Models 2.2(a) and 2.2(b) assume a linear effect of both the functional argument, \(s\), and the covariate of interest, \(X_{ij}\), with or without interaction effects. In particular, model 2.2(a) assumes that the rate of change of the mean response with respect to \(s\) is constant and does not depend on \(X_{ij}\), while model 2.2(b) assumes that the rate of change depends on the covariate of interest. Model 2.2(c) describes an additive effect of the functional argument and covariate of interest, with the additional assumption that the effect of \(X_{ij}\) is linear. The mean model 2.2(d) describes a completely nonparametric structure. While this model is useful when there is no a priori information on the mean structure, fitting
a nonparametric bivariate function is computationally expensive. We considered the case when \( X_{ij} \) is univariate mainly to keep the number of indices under control. All methods can be applied in more generality.

Fitting model (2.2) with either of the mean structures 2.2(a)-2.2(d) is not new. Morris \& Carroll (2006), and Scheipl et al. (2014) discuss estimation of the mean parameters in a variety of cases using an independence working assumption across observations. Also, when \( X_{ij} \) is the actual visit time, and there are no other covariates is the study, then the approach in Chen \& Müller (2012) can be used to estimate a bivariate smooth mean under the working independence assumption. However, none of these papers discusses inference on the population level effects that accounts for the complex correlation structure of the data. The novelty of this chapter consists precisely in filling this gap in the literature. To be specific, we consider an estimation approach based on the independence working assumption, introduce pointwise and joint confidence bands for the fixed effects, and propose a hypothesis testing procedure for the null hypothesis that the covariate of interest, \( X_{ij} \), has no effect on the outcome.

When the data are modeled as in model (2.2) and \( \mu(s, X) \) has the structure 2.2(a), then the mean parameter estimates are \( \beta_0, \beta_s, \beta_x, \) and \( \tau \). They are estimated by minimizing \( SSE = \sum_{i,j,r} \left[ Y_{ijr} - \{ \beta_0 + \beta_s s_{ijr} + \beta_x X_{ij} + Z^T_{ij} \tau \} \right]^2 \). Estimators can be represented in matrix form as \([\hat{\beta}^T, \hat{\tau}^T]^T = (M^T M)^{-1} M^T Y\), where \( \beta = (\beta_0, \beta_s, \beta_x)^T \), \( M = [M_1 \vdots M_2] \), with \( M_1 \) the matrix with rows \((1, s_{ijr}, X_{ij})\) and \( M_2 \) the matrix obtained by row-stacking of \( Z^T_{ij} \). Here \( Y \) is the \( L \sum_{i=1}^n m_i \)-dimensional vector of all \( Y_{ijr} \)’s.

The estimation criterion becomes progressively more complicated as the mean structure \( \mu(s, X) \) becomes more involved. For example, for nonparametric modeling (2.2(c) and 2.2(d)) we follow standard smoothing practices using penalized splines. Of course, other types of smoothers are also acceptable and may be equally or more appropriate for
different types of data structures. The methods we discuss apply to all types of smoothers.

To be specific, consider the most complex example, 2.2(d), where $\mu(s, X)$ is an unspecified bivariate smooth function. Construct a bivariate basis by the tensor product of two univariate B-spline bases, \{${B_s^1(s), \ldots, B_s^{d_s}(s)}$\}, and \{${B_x^1(x), \ldots, B_x^{d_x}(x)}$\}, defined on $S$ and $X$ respectively. Then $\mu(s,x) = \sum_{i=1}^{d_s} \sum_{k=1}^{d_x} B_s^i(s) B_x^k(x) \beta_{ik} = B(s, x)^T \beta$; where $B(s, x)$ is the $d_sd_x$-dimensional vector of $B_s^i(s)B_x^k(x)$’s and $\beta$ is the vector of parameters $\beta_{ik}$. Typically, the number of basis functions is chosen sufficiently large to capture the maximum complexity of the mean function and smoothness is induced by a quadratic penalty on the coefficients. There are several penalties for bivariate smoothing with the most popular being the ones proposed by Marx & Eilers (2005) and Wood (2006a,b). More recently, Xiao et al. (2013, 2014) proposed a scalable sandwich penalty estimator that leads to a computationally efficient algorithm for high dimensional data. In this chapter we used the following estimation criterion

$$\arg\min_{\beta, \tau, \lambda} \sum_{i,j,r} \left| Y_{ijr} - \left\{ B(s_r, X_{ij})^T \beta + Z_{ij}^T \tau \right\} \right|^2 + \beta^T P_\lambda \beta,$$  \hspace{1cm} (2.2)$$

with a penalty matrix $P_\lambda$ described in Wood (2006a) and a vector of smoothing parameters, $\lambda$. Specifically, we used $P_\lambda = \lambda_s P_s \otimes I_{d_x} + \lambda_x I_{d_s} \otimes P_x$ and $\lambda = (\lambda_s, \lambda_x)^T$, where $\otimes$ denotes the tensor product, and $P_s$ and $\lambda_s$ are the marginal second order difference matrix and the smoothing parameter for the $s$ direction, respectively; $P_x$ and $\lambda_x$ are defined similarly for the $x$ direction. Here $I_{d_s}$ and $I_{d_x}$ are the identity matrices of dimensions $d_s$ and $d_x$. For a fixed smoothing parameter, $\lambda$, the minimizer of (2.2) has the form $[\hat{\beta}_\lambda^T, \hat{\tau}_\lambda^T]^T = (M^T M + P_\lambda)^{-1} M^T Y$, while the estimated mean is $\hat{\mu}(s, x) + Z_{ij}^T \hat{\tau} = B(s, x)^T \hat{\beta}_\lambda + Z_{ij}^T \hat{\tau}_\lambda$. 

18
Selecting the optimal value of the smoothing parameter has been discussed extensively in the literature. Two widely used criteria are the generalized cross validation (GCV) and the restricted maximum likelihood (REML). GCV is based on prediction error, whereas REML is based on likelihood estimation where the smoothing parameter is a variance parameter. Empirical evidence (Ruppert et al., 2003) suggests that REML and GCV tend to have different behaviors because of the different way they trade bias for variance. REML tends to be more biased with lower variance (Ruppert et al. (2003); Reiss & Ogden (2007); Reiss & Todd Ogden (2009); Wood (2006a)), while GCV tends to be less biased with higher variance (Ruppert et al. (2003); Wahba (1990)). New evidence (Xiao et al. (2015b)) suggests that covariance smoothing can be improved by using leave-one-subject-out generalized cross validation for functional data. However, here we investigate only estimation under independence both for the mean function and its smoothing parameters; in our numerical investigation we select the optimal smoothing parameters by GCV.

The `gam` function in the `R` (R Core Team, 2014) package `mgcv` (Wood, 2006a) is used to implement model (2.2) with various mean structures as discussed above and using row and column penalties. The `fbps` function (Xiao et al., 2013) in `R` (R Core Team, 2014) package `refund` (Crainiceanu et al., 2013) can be used to fit the smooth bivariate effects by employing fast bivariate penalized spline smoothing with a modified penalty (Xiao et al., 2014). This function requires only simple modifications to account for linear effects of additional covariates: the linear effects are fit first and the bivariate smooth effects are fitted second conditional on the linear effects estimates. The reason for considering a range of models from very simple parametric to complex nonparametric models is to show the generality of the approaches. While we will keep a close eye on the activity application, the basic principle remains simple: estimate parameters under independence and bootstrap independent units.
In the following section we discuss inference for $\mu(s,x)$ in the form of confidence bands and hypothesis testing.

### 2.3 Confidence bands for $\mu(s,x)$

Without loss of generality, assume that the mean structure is $\mu(s,x) = B(s,x)^T \beta$, where $B(s,x)^T$ can be as simple as $(1,s,x)$ or as complex as a vector of pre-specified basis functions. The mean estimator of interest is $\hat{\mu}(s,x) = B(s,x)^T \hat{\beta}$. One could study point-wise variability for every pair $(s,x)$, that is $\text{var}\{\hat{\mu}(s,x)\}$, or the joint variability for the entire domain $\mathcal{S} \times \mathcal{X}$, that is $\text{cov}\{\hat{\mu}(s,x) : s \in \mathcal{S}, x \in \mathcal{X}\}$. Irrespective of the choice, the variability is fully described by the variability of the parameter estimator $\hat{\beta}$.

In this chapter we consider a flexible dependence structure for $\epsilon_{ij}(s)$ that describes both within- and between-subject variability. We make minimal assumptions on the errors that $\epsilon_{ij}(s)$ is independent over $i$ but is correlated over $j$ and $s$. Deriving the analytical expression for the sampling variability of the estimator $\hat{\beta}$ in such contexts is challenging. We use bootstrap to study the sampling properties of the parameter estimator. Two bootstrap algorithms are discussed: bootstrap of subject-level data and bootstrap of subject-level residuals. These approaches have already been studied and popularly used under the nonparametric regression setting for independent measurements; see, for example, Härdle & Bowman (1988), Efron & Tibshirani (1994), and Hall et al. (2013) among many others. Bootstrap of functional data for fixed effects has also been considered in several literatures, including Politis & Romano (1994) for weakly dependent processes in Hilbert space, Cuevas et al. (2006) for independent functional data, and Crainiceanu et al. (2011) for two paired samples of functional data. Nonetheless, performance of the proposed bootstrap algorithms for dependent functional data with such complex error
structures that we consider in this chapter is unknown and needs to be assessed.

The first method is more generally applicable, while the second relies on two important assumptions: i) the covariates do not depend on visit, that is $X_{ij} = X_i$ and $Z_{ij} = Z_i$; and ii) both the correlation and the variance of errors are independent of the covariates. These assumptions ensure that sets of subject-level errors, i.e. $\{\epsilon_{ij}(s) : j = 1, \ldots, m_i\}$ for $i = 1, \ldots, n$, can be re-sampled over subjects without affecting the sampling distribution.

Both bootstrap methods rely on specification of $B(s, x)$. In models that require smoothing parameters, their selection is considered to be part of the estimation procedure and is repeated at each bootstrap step.

Algorithm 1 Bootstrap of the subject-level data [uncertainty estimation]

1: for $b \in \{1, \ldots, B\}$ do
2: Re-sample the subject indexes from the index set $\{1, \ldots, n\}$ with replacement. Let $I^{(b)}$ be the resulting sample of $n$ subjects.
3: Define the $b$th bootstrap data by:
   \[
data^{(b)} = \{Y_{i^*j}(s_r), X_{i^*j}, Z_{i^*j}, s_r : i^* \in I^{(b)}, j = 1, \ldots, m_{i^*}, \text{ and } r = 1, \ldots, R\}.
\]
4: Using data$^{(b)}$ fit the model (2.2) with the mean structure of interest modeled by $\mu(s, x) = B(s, x)^T\beta$, by employing criterion (2.2). Let $\hat{\beta}_{\lambda}^{(b)}$ be the corresponding estimate of the parameter of interest; similarly define $\hat{\mu}^{(b)}(s, x) = B(s, x)^T\hat{\beta}_{\lambda}^{(b)}$.
5: end for
6: Calculate the sample covariance of $\{\hat{\beta}_{\lambda}^{(1)}, \ldots, \hat{\beta}_{\lambda}^{(B)}\}$; denote it by $V_{\hat{\beta}}$.

In many applications covariates do not depend on the visit (e.g. gender, age), that is $X_{ij} = X_i$ and $Z_{ij} = Z_i$; in particular, this is the case in the BLSA data. To account for this information we propose another version of the bootstrap of the data, which relies on the assumption that the error covariance is independent of the covariates. The bootstrap of subject-level residuals shows excellent numerical results, as illustrated in the simulation.
Fit the model (2.2) with the mean structure of interest modeled by \( \mu(s, x) = \mathbf{B}(s, x)^T \beta \), by employing the estimation criterion described in (2.2). Calculate residuals by 
\[
e_{ij}(s_r) = Y_{ij}(s_r) - \mathbf{B}(s_r, X_i)^T \hat{\beta}_\lambda - \mathbf{Z}_i^T \hat{\tau}_\lambda.
\]

**Algorithm 2** Bootstrap of the subject-level residuals [uncertainty estimation]

1: for \( b \in \{1, \ldots, B\} \) do
2: Re-sample the subject indexes from the index set \( \{1, \ldots, n\} \) with replacement.
   Let \( I^{(b)} \) be the resulting sample of subjects. For each \( i = 1, \ldots, n \) denote by \( m_i^* \) the number of repeated time-visits for the \( i \)th subject selected in \( I^{(b)} \).
3: Define the \( b \)th bootstrap sample of residuals 
\[
\{ e^*_ij(s_r) : i = 1, \ldots, n, \ j = 1, \ldots, m_i^*, \ \text{and} \ \ r = 1, \ldots, R \}.
\]
4: Define the \( b \)th bootstrap data by:
\[
data^{(b)} = \{ Y_{ij}^*(s_r), X_i, Z_i, s_r \} : i = 1, \ldots, n, j = 1, \ldots, m_i^*, r = 1, \ldots, R \},\]
   where 
\[
Y_{ij}^*(s_r) = \mathbf{B}(s_r, X_i)^T \hat{\beta}_\lambda + \mathbf{Z}_i^T \hat{\tau}_\lambda + e^*_ij(s_r) : i = 1, \ldots, n, j = 1, \ldots, m_i^*, r = 1, \ldots, R \}.
\]
5: Using \( data^{(b)} \) fit the model (2.2) with the mean structure of interest modeled by 
\( \mu(s, x) = \mathbf{B}(s, x)^T \beta \), by employing criterion (2.2). Let \( \hat{\beta}^{(b)} \) be the corresponding estimate of the parameter of interest; similarly define \( \hat{\mu}^{(b)}(s, x) = \mathbf{B}(s, x)^T \hat{\beta}^{(b)}_{\lambda(b)} \).
6: end for
7: Calculate the sample covariance of \( \{ \hat{\beta}^{(1)}_{\lambda(1)}, \ldots, \hat{\beta}^{(B)}_{\lambda(B)} \} \); denote it by \( V_{\hat{\beta}} \).

For fixed \((s, x)\), the variance of the estimator \( \hat{\mu}(s, x) = \mathbf{B}(s, x)^T \hat{\beta} \) can be estimated as 
\[
\text{var}\{ \hat{\mu}(s, x) \} = \mathbf{B}(s, x)^T V_{\hat{\beta}} \mathbf{B}(s, x),
\]
by using the bootstrap-based estimate of the covariance of \( \hat{\beta} \). A 100\((1-\alpha)\)% pointwise confidence interval for \( \mu(s, x) \) can be calculated as 
\[
\hat{\mu}(s, x) \pm z_{\alpha/2} \sqrt{\text{var}\{ \hat{\mu}(s, x) \}},
\]
using normal distributional assumption for the estimator \( \hat{\mu}(s, x) \), where \( z_{\alpha/2} \) is the 100\((1 - \alpha/2)\) percentile of the standard normal. A robust
alternative is obtained by using pointwise $100(\alpha/2)\%$ and $100(1-\alpha/2)\%$ quantiles of the bootstrap estimates $\{\hat{\mu}^{(b)}(s,x) : b = 1, \ldots, B\}$.

In most cases, it makes more sense to study the variability of $\hat{\mu}(s,x)$, and draw inference about the entire true mean function $\{\mu(s,x) : (s,x) \in D_s \times D_x\}$. Thus, we focus our study on constructing a joint (or simultaneous) confidence band for $\mu(s,x)$. Constructing simultaneous confidence bands for univariate smooths has already been discussed in the nonparametric literature. For example, Degras (2009), Ma et al. (2012), and Cao et al. (2012) proposed an asymptotically correct simultaneous confidence bands using different estimators, for independently sampled curves; Crainiceanu et al. (2012) proposed bootstrap-based joint confidence bands for univariate smooths in the case of functional data with complex error processes by using ideas of Ruppert et al. (2003). Here, we present an extension of the approach considered by Crainiceanu et al. (2012) to bivariate smooth function.

Let $S^* = \{s_{g_s} : g_s = 1, \ldots, G_s\}$ and $X^* = \{x_{g_x} : g_x = 1, \ldots, G_x\}$ be evaluation points that are equally spaced in the domains $D_s$ and $D_x$, respectively. Then, we evaluate the bootstrap estimate $\hat{\mu}^{(b)}(s,x)$ of one bootstrap sample at all pairs $(s,x) \in S^* \times X^*$, and denote by $\hat{\mu}^{(b)}$ the $G_sG_x$-dimensional vector with components $\hat{\mu}^{(b)}(s,x)$. Let $B$ be the $\text{dim}(\beta) \times G_sG_x$-dimensional matrix obtained by column-stacking $B(s_{g_s},x_{g_x})$ for all $g_s$ and $g_x$. Let $\varphi(s_{g_s},x_{g_x}) = \sqrt{\text{var}\{\hat{\mu}(s_{g_s},x_{g_x})\}}$ as defined above. After adjusting for the bivariate structure of the problem the main steps of the construction of the joint confidence bands for $\mu(s,x)$ follow similarly to the ones used in (Crainiceanu et al., 2012) for univariate smooth parameter functions. For completeness we describe the steps below.

**Step 1.** Generate a random variable $u$ from the multivariate normal distribution with mean $0_{\text{dim}(\beta)}$ and variance-covariance matrix $V_\beta$; let $q(s_{g_s},x_{g_x}) = B(s_{g_s},x_{g_x})^T u$ for $g_s =
1, \ldots, G_s and g_x = 1, \ldots, G_x.

Step 2. Calculate 
\[ q^*_{max} = \max_{(s_{g_x}, x_{g_x})} \left\{ \left| q(s_{g_x}, x_{g_x}) \right| / \sqrt{\varphi(s_{g_x}, x_{g_x})} : (s_{g_x}, x_{g_x}) \in S^* \times X^* \right\}. \]

Step 3. Repeat Step 1. and Step 2. for \( \ell = 1, \ldots, L \), and obtain \( \{ q^*_{max,\ell} : \ell = 1, \ldots, L \} \).
Determine the 100(1 - \alpha)\% empirical quantile of \( \{ q^*_{max,\ell} : \ell = 1, \ldots, L \} \), say \( \hat{q}_{1-\alpha} \).

Step 4. Construct the 100(1 - \alpha)\% joint confidence band:
\[ \{ \bar{\mu}(s_{g_s}, x_{g_x}) \pm \hat{q}_{1-\alpha} \sqrt{\varphi(s_{g_s}, x_{g_x})} : (s_{g_s}, x_{g_x}) \in S^* \times X^* \} \]
Here \( \bar{\mu}(s, x) = B^{-1} \sum_{b=1}^{B} \hat{\mu}^{(b)}(s, x) \) is the sample mean of the bootstrap estimates \( \hat{\mu}^{(b)}(s, x) \)'s.

The performance of the joint confidence bands is evaluated via simulation study in Section 2.6. The joint confidence band provides a information about the entire true mean function. Moreover, the joint confidence band, in contrast to the pointwise confidence band, can be used as an inferential tool for formal global tests about the mean function, \( \mu(s, x) \).

For example, one can use the joint confidence band for testing the null hypothesis, \( H_0 : \mu(s, x) = 0 \) for all pairs \((s, x) \in D_s \times D_x\), by checking whether the confidence band
\[ \{ \bar{\mu}(s_{g_s}, x_{g_x}) \pm \hat{q}_{1-\alpha} \sqrt{\varphi(s_{g_s}, x_{g_x})} : (s_{g_s}, x_{g_x}) \in S^* \times X^* \} \] contains the vector \( 0_{G_s G_x} \). If the confidence band does not contain \( 0_{G_s G_x} \), then we conclude that there is significant evidence that the true mean function is nonzero. Furthermore, one can use this approach to study hypothesis testing that the mean function \( \mu(s, x) \) is equal to some specified bivariate smooth function, say \( f_0(s, x) \), by simply investigating whether the specified function is contained in the joint confidence band.
2.4 Hypothesis testing for $\mu(s, x)$

Next, we focus on assessing the effect of the covariate of interest $X$ on the mean function. Consider the general case when the model is (2.2) and the average effect is an unspecified bivariate smooth function, $\mu(s, x)$. One of the goals is to test if the true mean function depends on $x$, that is testing the following null hypothesis:

$$H_0 : \mu(s, x) = \eta(s) \text{ for all } s, x,$$

(2.3)

for some unknown smooth function $\eta : \mathcal{D}_s \to \mathbb{R}$ against the alternative $H_A : \mu(s, x)$ varies over $x$ for some $s$.

To the best of our knowledge, this type of hypothesis has not been studied in functional data analysis. The problem was extensively studied in nonparametric smoothing, where the primary interest centered on significance testing of a subset of covariates in a nonparametric regression model. For example, Fan & Li (1996) and Lavergne & Vuong (2000) proposed consistent, kernel-based test statistics. Delgado & Manteiga (2001) and Gu et al. (2007) also considered similar test statistics, but proposed bootstrap methods to approximate the null distribution of the test statistic. Hall et al. (2007) proposed a cross-validation based method. However, all these methods are based on the assumption that observations are independent across sampling units; in our context requiring independence of $Y_{ij}(s_{ijr})$ over $j$ and $r$ is unrealistic. Failing to account for this dependence leads to inflated type I error rates.

To test hypothesis (2.4), we propose a test statistic based on the $L^2$ distance between the mean estimators under the null and alternative hypotheses. Specifically we define the
test statistic as:

\[ T = \int_X \int_S \{ \hat{\mu}_A(s, x) - \hat{\mu}_0(s) \}^2 ds dx, \quad (2.4) \]

where \( \hat{\mu}_0(s) \) and \( \hat{\mu}_A(s, x) \) are the estimates of \( \mu(s, x) \) under the null and alternative hypothesis, respectively. In particular, \( \hat{\mu}_A(s, x) \) is estimated as in Section 2.2. The estimator \( \hat{\mu}_0(s) \) is obtained by modeling \( \mu(s) = \sum_{l=1}^{d_s} B_{sl}(s) \beta_l = B(s)^T \beta \) for the \( d_s \)-dimensional vector \( \beta \) and by estimating the mean parameters \( \beta \) based on a criterion similar to (2.2). Specifically, we use the penalized criterion \( \sum_{i,j,r} \{ Y_{ij}(s_r) - B(s_r)^T \beta - Z_{ij} \tau \}^2 + \lambda_s \beta^T P_s \beta, \)

where \( \lambda_s \) is the smoothing parameter and \( P_s \) is the \( d_s \times d_s \) penalty matrix described in Section 2.2. In practice, the two estimated effects \( \hat{\mu}_0(s) \) and \( \hat{\mu}_A(s, x) \) can be obtained using the \texttt{gam} function in the \texttt{R} (R Core Team, 2014) package \texttt{mgcv} (Wood, 2006a).

Deriving the null distribution of the test statistic \( T \) is challenging. We propose to approximate the null distribution of \( T \) using either of subjects or of subject-level residuals. Below we provide the details.

Below we provide the details of the algorithm.

\begin{algorithm}
\caption{Bootstrap approximation of the null distribution of the testing procedure} \label{alg:bootstrap}
\begin{algorithmic}
\State \textbf{for} \( b \in \{1, \ldots, B\} \) \textbf{do}
\State \text{Re-sample the subject indexes from the index set} \( \{1, \ldots, n\} \) \text{with replacement.}
\State \text{Let} \( I^{(b)} \) \text{be the obtained sample of subjects. For each} \( i = 1, \ldots, n \) \text{denote by} \( m_i^* \) \text{the number of repeated time-visits for the} \( i \text{th subject selected in} \ I^{(b)}. \)
\State \text{Define the} \( b \text{th bootstrap sample of pseudo-residuals} \)
\State \( \{ e_{ij}^*(s_r) : i = 1, \ldots, n, \ j = 1, \ldots, m_i^*, \ \text{and} \ r = 1, \ldots, R \}. \) \text{For each} \( i = 1, \ldots, n \) \text{let}
\State \( \{ Z_{ij}^* : j = 1, \ldots, m_i^* \} \) \text{the corresponding sample of the nuisance covariates for the} \( i \text{th subject selected in} \ I^{(b)}. \) \text{Similarly define} \( X_{ij}^*. \)
\State \text{Define the} \( b \text{th bootstrap data by:} \)
\State \text{data}^{(b)} = \{ [Y_{ij}^*(s_r), X_{ij}^*, Z_{ij}^*] : i = 1, \ldots, n, \ j = 1, \ldots, m_i^*, \ r = 1, \ldots, R \}, \text{where}
\end{algorithmic}
\end{algorithm}
\[ Y_{ij}^*(s_r) = \hat{\mu}_0(s_r) + Z_{ij}^* \hat{\tau}_A + e_{ij}^*(s_r) \]

5: Using data\(^{(b)}\) fit two models. First, fit model (2.2) with the mean structure modeled by \( \mu(s, x) = B(s, x)^T \beta \) and estimate \( \hat{\mu}_A^{(b)}(s, x) \). Second, fit model (2.2) with the mean structure modeled by \( \mu(s) = B(s)^T \beta \) and estimate \( \hat{\mu}_0^{(b)}(s) \). Calculate the value of the test statistic \( T^{(b)} \) using formula (2.4).

6: end for

7: Approximate the tail probability \( P(T > T_{obs}) \) by the p-value = \( B^{-1} \sum_{b=1}^{B} I(T^{(b)} > T_{obs}) \), where \( T_{obs} \) is obtained using the original data and \( I \) is the indicator function.

When the covariates \( X_{ij} \) and \( Z_{ij} \) do not depend on visit, i.e. \( X_{ij} = X_i \) and \( Z_{ij} = Z_i \), the algorithm can be modified along the lines of the ‘bootstrap of the subject-level residuals’ algorithm.

### 2.5 Application to physical activity data

Physical activity measured by wearable devices such as accelerometers provides new insights into the association between activity and health outcomes (Schrack et al., 2014); The complexity of the data also poses serious challenges to current statistical analysis. For example, accelerometers can record activity at minute level resolution for many days and for hundreds of individuals. Here we consider the physical activity data from the Baltimore Longitudinal Study on Aging (Stone & Norris, 1966a). Participants in the study wore the Actiheart portable physical activity monitor (Brage et al. 2006) 24 hours a day for a number of days. Activity counts were measured in 1-min epochs and each daily activity profile has 1440 minute-by-minute measurements of activity counts. Activity counts are proxies of activity intensity. Activity counts were log-transformed (more precisely, \( x \rightarrow \log(1 + x) \)) because they are highly skewed and then averaged in 5-min
intervals. For simplicity, hereafter we refer to the log-transformed counts as log counts. For this analysis, we focus on 1779 daily activity profiles from a single visit of 378 female participants who have at least two days of data. Women in the study are aged between 31 and 93 years old. Further details on the BLSA activity data can be found in Schrack et al. (2014) and Xiao et al. (2015b).

Our objective is to conduct inference on the marginal effect of age on women’s daily activity after adjusting for body mass index. We model the mean log counts as $\mu(s, X_i) + Z_i\beta(s)$, where $X_i$ and $Z_i$ are the age and body mass index of the $i$th woman during the visit, $\mu(s, x)$ is the baseline mean log counts for time $s$ within the day for a woman aged $x$ years old, and $\beta(s)$ is the association of body mass index with mean log counts for time $s$ within the day. We test whether $\mu(s, x)$ varies solely with $s$. We use the proposed testing statistic, $T = \int \int \left\{ \hat{\mu}_A(s, x) - \hat{\mu}_0(s) \right\}^2 dt dx$ as detailed in Section 2.4. The estimate $\hat{\mu}_A(s, x)$ is based on the tensor product of 15 cubic basis functions in $s$ and 5 cubic basis functions in $x$ and the estimate $\hat{\mu}_0(s)$ is based on 15 cubic basis functions. Figure 2.1 shows the null distribution. The observed test statistic is $T = 0.041$ and the corresponding p-value is less than 0.001 based on 1000 MC samples. This indicates that there is strong evidence that daily activity profiles in women vary with age.

Figure 2.2 shows the estimated baseline activity profile as a function of age, $\hat{\mu}(s, x)$, average of all bootstrap estimates. The plot indicates that the average log counts is a decreasing function of age for most time during the day. Furthermore, it depicts two activity peaks, one around 12pm and the other around 6pm. In particular, the peak in the evening seems to decrease faster with age, indicating that afternoon activity is more affected by age. The joint lower and upper 95% confidence limits based on methods described in Section 2.3 are displayed in Figure 2.3. Figure 2.4 displays the estimated mean activity profile for 60 years old women along with the corresponding joint 95% confidence band.
Figure 2.5 displays the estimated association of body mass index with mean log counts as a function of time of day; it suggests that women with higher body mass index have less activity during the day and evening, albeit more activity at late night and early morning.

Figure 2.1: The null distribution of the test statistic in (2.4) for the null hypothesis that there is no effect of age on activity. The red dashed line is the 95 percent quantile of the null distribution of the test statistic.
Figure 2.2: Heat map of average of bootstrap estimates of log counts as a bivariate function of time of day and age (left panel) and average of bootstrap estimates of log counts for five different age groups (right panel).

Figure 2.3: Heat maps of joint confidence bands for the estimate $\hat{\mu}(t, Age)$ given in the left panel of Figure 2.2. The legend on the right applies to both plots.
Figure 2.4: Average of bootstrap estimates of log counts as a function of time of day at age 60 and the associated joint confidence bands.

Figure 2.5: Association of body mass index with mean log counts as a function of time of day and the associated joint confidence bands.
2.6 Simulation Study

We conducted a simulation study to investigate the performance of the inferential methods introduced in this chapter. First, we evaluate the accuracy of the pointwise and joint confidence bands in terms of average coverage probability and average confidence interval length. Second, we evaluate the testing procedure with respect to Type I error and power.

Data are simulated using the model (2.2) where \( X_{ij} = X_i, \ Z_{ij} = Z_i \). Errors \( \epsilon_{ij}(s) \) are generated from \( \epsilon_{ij}(s) = \sum_{l=1}^{3} \xi_{ijl}(s) + w_{ij}(s) \), where \( \xi_{ijl} \) are random variables with zero mean, variance \( \lambda_l \) that are independent over \( i \) and \( l \), and exponential autocorrelation with a correlation parameter \( \rho \). The residuals \( w_{ij}(\cdot) \) are mutually independent with zero mean and variance \( \sigma^2 \). The number of repeated measures is fixed at \( m_i = 5 \), \( (\lambda_1, \lambda_2, \lambda_3) = (3, 2, 1/3) \), and the functions \( [\phi_1(s), \phi_2(s), \phi_3(s)] = [\sqrt{2}\cos(2\pi s), \sqrt{2}\sin(2\pi s), \sqrt{2}\cos(4\pi s)] \).

The subject-specific covariates \( X_i \) and \( Z_i \) are generated from a Uniform\([0, 1]\). The grid of points \( \{t_r : r = 1, \ldots, R\} \) is set as 101 equally spaced discrete points in \([0, 1]\). The variance of the white noise process \( \sigma^2 \) is set to 5.33, which is equivalent to ensuring a signal to noise ratio equal to 1. Here the signal to noise ratio is calculated as

\[
\text{SNR} = \frac{\int \text{var}[Y_{ij}(s)]dt}{\sigma^2} - 1 = \sum_{l=1}^{3} \lambda_l / \sigma^2.
\]

We consider different combinations of the following factors:

**F1** number of subjects: (a) \( n = 100 \), (b) \( n = 200 \), and (c) \( n = 300 \);

**F2** mean function:

**F2 i. bivariate function \( \mu(s, x) \)**

(a) \( \mu_1(s, x) = \beta_0 + \beta_s s + \beta_x x \) for \( (\beta_0, \beta_s, \beta_x) = (5, 2, 3) \), (Ex 2.2(a))

(b) \( \mu_2(s, x) = \beta_0 + \beta_s s + \beta_x x + \beta_{sx} sx \) for \( (\beta_0, \beta_s, \beta_x, \beta_{sx}) = (5, 2, 3, 7) \), (Ex 2.2(b))

(c) \( \mu_3(s, x) = \cos(2\pi s) + \beta_x x \) for \( \beta_x = 3 \), (Ex 2.2(c))

(d) \( \mu_4(s, x) = \cos(2\pi s) + \delta((x/4) - s)^3 \) for \( \delta = 0, 2, 4, \text{ and } 6 \), (Ex 2.2(d))
F2 ii. linear effect of covariate $Z_i$

(a) $\tau = 0$ (no effects), (b) $\tau = 8$;

F3 correlation: (a) $\rho = 0.2$ (weak correlation) and (b) $\rho = 0.9$ (strong correlation).

Our methodology is evaluated on these models in two ways. First, we model the data by assuming the correct model and by evaluating the accuracy of the inferential procedures; the results are detailed next. Second, we model the data using a bivariate mean, $\mu(s, x)$, and evaluate the performance of the confidence bands of $\mu(s, x)$ for covering the true mean even when the true mean has a simpler structure; results are described in Appendix section A.4.

We show now results for fitting the correct model; estimation is done as detailed in Section 2.2. When the assumed model for the mean structure of interest involves univariate or bivariate smooths, we use $d_s = 7$ and/or $d_x = 7$ cubic B-spline basis functions, and select the smoothing parameter/s via GCV; specifically for the bivariate smooth, $d_s d_x = 49$ basis functions are used. Compared to the data analysis, we use a relatively small number of basis functions because there are only 101 grid points along $s$. Estimation accuracy is measured using the bias and variance of the estimators; for univariate and bivariate smooths, single number summaries of these measures are used. Specifically, when the mean of interest is $\mu(s) = \cos(2\pi s)$, as in scenario F2 i.(c), the integrated bias defined by $\int_0^1 \{\bar{\mu}(s) - \mu(s)\}dt$ is used as a summary measure of bias, and the integrated variance, defined by $\int_0^1 \{\sum_{i=1}^{N_{sim}} \{\hat{\mu}_{i\sim}(s) - \bar{\mu}(s)\}^2/(N_{sim} - 1)\}dt$ is used as a summary measure of variance. Here $\hat{\mu}_{i\sim}(s)$ denotes the mean estimator from one simulation, $\bar{\mu}(s) = \sum_{i=1}^{N_{sim}} \hat{\mu}_{i\sim}(s)/N_{sim}$ is the sample mean of the estimator $\hat{\mu}(s)$. Inference for the parameter/s of interest is done using methods described in Sections 2.3 and 2.4. The performance of the pointwise and joint confidence bands for both univariate,
and bivariate cases is evaluated in terms of average coverage probability (ACP), and average length (AL). Specifically, let \((\hat{\mu}_i^{sim}, l(s, x), \hat{\mu}_i^{sim}, u(s, x))\) be the 100\((1 - \alpha)\)% pointwise confidence interval of \(\mu(s, x)\) obtained at the \(i_{sim}\) Monte Carlo generation of the data, then

\[
ACP_{\text{point}} = \frac{1}{N_{sim}G_sG_x} \sum_{i_{sim}=1}^{N_{sim}} \sum_{g_s=1}^{G_s} \sum_{g_x=1}^{G_x} \mathbb{1}\{\mu(s_{g_s}, x_{g_x}) \in (\hat{\mu}_i^{sim}, l(s_{g_s}, x_{g_x}), \hat{\mu}_i^{sim}, u(s_{g_s}, x_{g_x}))\}
\]

\[
AL_{\text{point}} = \frac{1}{N_{sim}G_sG_x} \sum_{i_{sim}=1}^{N_{sim}} \sum_{g_s=1}^{G_s} \sum_{g_x=1}^{G_x} |\hat{\mu}_i^{sim}, l(s_{g_s}, x_{g_x}) - \hat{\mu}_i^{sim}, u(s_{g_s}, x_{g_x})|,
\]

where \(\{s_{g_s} : g_s = 1, \ldots, G_s\}\) and \(\{x_{g_x} : g_x = 1, \ldots, G_x\}\) are equi-distanced grid points in the domains \(D_s\), and \(D_x\), respectively. Next, let \((\hat{\mu}_i^{sim}, l(s, x), \hat{\mu}_i^{sim}, u(s, x))\) be 100\((1 - \alpha)\)% joint confidence interval. The average length is calculated as above, while the average coverage probability is calculated as:

\[
ACP_{\text{joint}} = \frac{1}{N_{sim}} \sum_{i_{sim}=1}^{N_{sim}} \mathbb{1}\{\mu(s_{g_s}, x_{g_x}) \in (\hat{\mu}_i^{sim}, l(s_{g_s}, x_{g_x}), \hat{\mu}_i^{sim}, u(s_{g_s}, x_{g_x})) : \text{for all } g_s, g_x\}
\]

The performance of the test statistic \(T\) is evaluated in terms of its empirical type I error (size) for the nominal levels of 0.05, 0.10, and 0.15, and power for the nominal level of 0.05. The results for the nominal coverage of 95% are presented in Table 2.1; the results for other nominal coverages (85% and 90%) are included in Appendix A.

The results for the empirical size of the testing procedure are based on \(N_{sim} = 1000\) MC samples, while the results for the coverage probability, expected length, and power of the test are based on \(N_{sim} = 500\) MC samples. For each MC simulation we use \(B = 300\) bootstrap samples; they are obtained by bootstrapping residuals by subject.

Table 2.1 shows the bias, variance, ACP and AL for the mean structure of interest
and using the nominal level 95% when the sample size is \( n = 100 \). When the mean structure includes smooth terms, both pointwise and joint confidence intervals/bands are provided. Overall, both pointwise or/and joint confidence intervals/bands perform well. The confidence interval/bands tend to be wider when the correlation within the repeated observations is strong (\( \rho = 0.9 \)) than when is weaker (\( \rho = 0.2 \)).

The joint confidence bands based on bootstrap of subjects perform equally well when the effect of the covariate \( X \) is linear (cases F2 i.(a)-(c)). For the case of the nonlinear effect of \( X \) on the outcome (case F2 i.(d)), we consider both a covariate that is constant over visit (i.e. \( X_i \)) and a covariate that varies with visit (i.e. \( X_{ij} \)). The results show the good coverage of the joint confidence bands with the visit-varying covariate \( X_{ij} \). Additional results based on bootstrapping subject-level observations are included in Appendix section A.3. The results suggest that for a time-invariant covariate (i.e. \( X_i \)) the bootstrap of subject-level residuals gives a narrower joint confidence band with better coverage than the bootstrap of subject-level observations.

Table 2.2 shows the empirical type I error of the proposed testing procedure for testing \( H_0 : \mu(s,x) = \eta(s) \), where \( \eta(\cdot) \) is a smooth effect depending on \( s \) only. Rejection probabilities are given for various nominal levels, different correlation strength, and increasing sample sizes. Results indicate that, as sample size increases, the size of the test gets closer to the corresponding nominal levels. Including an additional covariate in the model seems to have no effect on the performance of the testing procedure. Figure 2.6 illustrates the power curves, when the true mean structure deviates from the null hypothesis. It presents the power as a function of the deviation from the null that involves both \( s \) and \( x \), \( \mu(s,x) = 2 \cos(2\pi s) + \delta(x/4 - s)^3 \). Here \( \delta \) quantifies the departure from the null hypothesis. As expected, rejection probabilities increase as the departure from the null hypothesis increases, irrespective of the direction in which it deviates. As expected,
Table 2.1: Simulation results using bootstrap of subject-level residuals and 95% nominal level; results are based on 500 MC samples.

<table>
<thead>
<tr>
<th>Case</th>
<th>True Mean Function</th>
<th>Parameter</th>
<th>$\rho$</th>
<th>Bias</th>
<th>$\sqrt{\text{var}}$</th>
<th>ACP</th>
<th>AL</th>
<th>ACP</th>
<th>AL</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a)</td>
<td>$\beta_0 + \beta_s s + \beta_x X + \tau Z$</td>
<td>$\beta_0 = 5$</td>
<td>0.20</td>
<td>0.20</td>
<td>0.20</td>
<td>0.94 (0.01)</td>
<td>0.54 ($&lt; 0.01$)</td>
<td>0.54 ($&lt; 0.01$)</td>
<td>0.90</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\beta_0 = 2$</td>
<td>0.20</td>
<td>-0.01</td>
<td>0.27</td>
<td>0.94 (0.01)</td>
<td>0.70 ($&lt; 0.01$)</td>
<td>0.70 ($&lt; 0.01$)</td>
<td>0.54 ($&lt; 0.01$)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\beta_0 = 3$</td>
<td>0.20</td>
<td>0.01</td>
<td>0.52</td>
<td>0.94 (0.01)</td>
<td>1.06 ($&lt; 0.01$)</td>
<td>1.06 ($&lt; 0.01$)</td>
<td>0.90 ($&lt; 0.01$)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\beta_0 = 4$</td>
<td>0.20</td>
<td>0.00</td>
<td>0.05</td>
<td>0.95 (0.01)</td>
<td>0.14 ($&lt; 0.01$)</td>
<td>0.14 ($&lt; 0.01$)</td>
<td>0.90 ($&lt; 0.01$)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\beta_0 = 5$</td>
<td>0.20</td>
<td>0.00</td>
<td>0.05</td>
<td>0.95 (0.01)</td>
<td>0.14 ($&lt; 0.01$)</td>
<td>0.14 ($&lt; 0.01$)</td>
<td>0.90 ($&lt; 0.01$)</td>
</tr>
</tbody>
</table>

Standard errors are presented in parentheses.

rejection probabilities increase with the sample size. Our investigation indicates that the strength of the correlation between the functional observations corresponding to the same subject affect the rejection probability: the weaker the correlation, the larger the power. There is no competitive testing method available for this null hypothesis.

The above discussion focuses on the performance of confidence intervals/bands when the correct mean structure is assumed in the estimation procedure. In Appendix section A.4 we present the corresponding results when the fitted model is completely nonparametric; of course this choice is more computationally intensive. Lastly we conducted an additional simulation study to evaluate robustness of the proposed methods to the non-Gaussian error distributions and obtained the similar results as the Gaussian case; the results are presented in Appendix section A.5.
Table 2.2: Empirical Type I error of the test statistic $T$ based on the $N_{\text{sim}} = 1000$ MC samples.

<table>
<thead>
<tr>
<th>$n$</th>
<th>$\rho$</th>
<th>$\alpha$</th>
<th>$\mu(s, x) = \cos(2\pi s)$, $\tau = 0$</th>
<th>$\mu(s, x) = \cos(2\pi s)$, $\tau = 8$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>$\alpha = 0.05$</td>
<td>$\alpha = 0.10$</td>
</tr>
<tr>
<td>100</td>
<td>0.2</td>
<td></td>
<td>0.08 (0.01)</td>
<td>0.07 (0.01)</td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td></td>
<td>0.09 (0.01)</td>
<td>0.08 (0.01)</td>
</tr>
<tr>
<td>200</td>
<td>0.2</td>
<td></td>
<td>0.07 (0.01)</td>
<td>0.08 (0.01)</td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td></td>
<td>0.08 (0.01)</td>
<td>0.12 (0.01)</td>
</tr>
<tr>
<td>300</td>
<td>0.2</td>
<td></td>
<td>0.06 (0.01)</td>
<td>0.11 (0.01)</td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td></td>
<td>0.06 (0.01)</td>
<td>0.12 (0.01)</td>
</tr>
</tbody>
</table>

Standard errors are presented in parentheses.

Figure 2.6: Estimated power curves for testing $H_0 : \mu(s, x) = \eta(s)$ using level of significance $\alpha = 0.05$, when the true mean function $\mu(s, x) = 2\cos(2\pi s) + \delta(x/4 - s)^3$ for $\delta = 0.01, 2, 4, 6$. The results are based on $N_{\text{sim}} = 500$ MC samples.
Chapter 3

Longitudinal Functional Data Analysis

3.1 Introduction

Our motivation of this chapter is the longitudinal Diffusion Tensor Imaging (DTI) study, where the objective is to investigate the evolution of the MS disease as measured by the dynamics of a common DTI modality profile - fractional anisotropy (FA) - along the corpus callosum (CCA) of the brain. Every MS subject in the study is observed over possibly multiple hospital visits and at each visit the subject’s brain is imaged using DTI. In this chapter we consider summaries of FA at 93 equally spaced locations along the brain’s CCA, which we refer to as CCA-FA profile. The change over time in the CCA-FA profiles is informative of the progression of the MS disease, and thus a model that accounts for all the dependence sources in the data has the potential to be a very useful tool in practice. We propose a modeling framework that captures the process dynamics over time and provides prediction of a full CCA-FA trajectory at a future visit.
Existing literature in longitudinal functional data can be separated into two categories, based on whether or not it accounts for the actual time \( t_{ij} \) at which the profile \( Y_{ij}(\cdot) \) is observed; here \( i \) indexes the subjects and \( j \) indexes the repeated measures of the subject. Moreover, most methods that incorporate the time \( t_{ij} \) focus on modeling the process dynamics (Greven et al., 2010) and only few can do prediction of a future full trajectory. Chen & Müller (2012) considered the latter issue and introduced an interesting perspective, but their method is very computationally expensive and its application in practice is limited as a result. We propose a novel parsimonious modeling framework to study the process dynamics and prediction of future full trajectory in a computationally feasible manner.

In this chapter we focus on the case where the sampling design of \( t_{ij} \)‘s is sparse (hence sparse longitudinal design) and the subject profiles are observed at fine grids (hence dense functional design). We propose to model \( Y_{ij}(\cdot) \) as:

\[
Y_{ij}(s) = \mu(s, t_{ij}) + X_i(s, t_{ij}) + \epsilon_{ij}(s); \quad X_i(s, t_{ij}) = \sum_{k \geq 1} \xi_{ik}(t_{ij}) \phi_k(s) \quad \text{for } s \in S \text{ and } t_{ij} \in T
\]

(3.1)

where \( S \) and \( T \) are closed compact sets, \( \mu(\cdot, t_{ij}) \) is an unknown smooth mean response corresponding to \( t_{ij} \), \( X_i(\cdot, t_{ij}) \) is a smooth random deviation from the mean at \( t_{ij} \), and \( \epsilon_{ij}(\cdot) \) is a residual process with zero-mean and unknown covariance function to be described later. The bivariate processes \( X_i(\cdot, \cdot) \)'s are independent and identically distributed (iid), the error processes \( \epsilon_{ij} \)'s are iid and furthermore are independent of \( X_i \)'s. For identifiability we require that \( X_i \) comprises solely the random deviation that is specific to the subject; the repeated time-specific deviation is included in \( \epsilon_{ij} \). Here \( \{\phi_k(\cdot)\}_k \) is an orthogonal basis in \( L^2(S) \) and \( \xi_{ik}(t_{ij}) \)'s are the corresponding basis coefficients that have zero-mean, are uncorrelated over \( i \), but correlated over \( j \). We assume that the set of
visit times of all subjects, \( \{t_{ij} : i, j\} \), is dense in \( T \). Full model assumptions are given in Section 3.2.

The class of model (3.1) is rich and includes many existent models, as we illustrate now. (i) If \( \xi_{ik}(t_{ij}) = \zeta_{0,ik} + t_{ij}\zeta_{1,ik} \) for appropriately defined random terms \( \zeta_{0,ik} \) and \( \zeta_{1,ik} \), model (3.1) can be represented as in Greven et al. (2010). (ii) If \( \text{cov}(\xi_{ik}(t), \xi_{ik}(t')) = \lambda_k \rho_k(|t - t'|; \nu) \) for some unknown variance \( \lambda_k \), known correlation function \( \rho_k(\cdot; \nu) \) with unknown parameter \( \nu \), and \( n = 1 \), model (3.1) resembles to Gromenko et al. (2012) and Gromenko & Kokoszka (2013) for spatially indexed functional data. (iii) If \( \xi_{ik}(t_{ij}) = \sum_{l \geq 1} \zeta_{ikl} \psi_{ikl}(t_{ij}) \) with orthogonal basis functions \( \psi_{ikl}(t) \)’s and the corresponding coefficients \( \zeta_{ikl} \)’s, then model (3.1) is similar to Chen & Müller (2012) who used time-varying basis functions \( \phi_k(\cdot|t) \) instead of our proposed \( \phi_k(\cdot) \) in model (3.1) and assumed a white noise residual process \( \epsilon_{ij} \).

The use of time-invariant orthogonal basis functions is one key difference between the proposed framework and Chen & Müller (2012); another important difference is the flexible error structure that our approach accommodates. The key difference leads to several major advantages of the proposed method. First, by using a time-invariant basis functions, the basis coefficients, \( \xi_{ik}(t_{ij}) \)’s extract the low dimensional features of these massive data. The longitudinal dynamics is emphasized only through the time-varying coefficients \( \xi_{ik}(t_{ij}) \)’s of (3.1) and, thus, this perspective makes the study of the process dynamics easier to understand. Second, our approach involves at most two dimensional smoothing and as a result is computationally very fast; in contrast, the time-varying basis functions \( \{\phi_k(\cdot|t)\}_k \) at each \( t \), require three dimensional smoothing which is not only complex but also computationally intensive and slow.

Nevertheless, selecting the time-invariant basis is nontrivial. One option is to use a pre-specified basis; Zhou et al. (2008) considered this approach in modeling paired
of sparse functional data. Another option is to use data-driven basis functions, such as eigenbasis of some covariance. The challenge is: what covariance to use? We take the latter direction and propose to determine \( \{ \phi_k(\cdot) \}_k \) using an appropriate marginal covariance. In this regard, let \( c((s, t), (s', t')) \) be the covariance function of \( X_i(s, t) \) and \( g(t) \) be the density of \( t_{ij} \)'s. Define \( \Sigma(s, s') = \int_T c((s, t), (s', t))g(t)dt \) for \( s, s' \in S \); we show that this bivariate function is a proper covariance function (Horváth & Kokoszka, 2012). Section 3.2 shows that the proposed basis \( \{ \phi_k(\cdot) \}_k \) has optimal properties with respect to some appropriately defined criterion. From this viewpoint, the model representation (3.1) is optimal. The idea of using the eigenbasis of the pooled covariance can be related to Jiang & Wang (2010) and Pomann et al. (2013), who considered independent functional data.

The rest of the chapter is organized as follows. Section 3.2 introduces the proposed modeling framework. Section 3.3 describes the estimation methods and implementation. The methods are studied theoretically in Section 3.5 and then numerically in Section 3.6. Section 3.7 discusses the application to the tractography DTI data.

### 3.2 Modeling longitudinal functional data

Let \([\{t_{ij}, Y_{ij}(s_r) : r = 1, \ldots, R \} : j = 1, \ldots, m_i,] \) be the observed data for the \( i \)th subject, where \( Y_{ij}(\cdot) \) is the \( j \)th profile at random time \( t_{ij} \) for subject \( i \), and each profile is observed at the fine grid of points \( \{s_1, \ldots, s_R\} \). For convenience we use the generic index \( s \) instead of \( s_r \). The number of ‘profiles’ per subject, \( m_i \) is relatively small to moderate and the set of time points of all subjects, \( \{t_{ij} : \text{for all } i, j\} \), is dense in \( T \). Without loss of generality, we set \( S = T = [0, 1] \). We model the response \( Y_{ij}(\cdot) \) using (3.1), where we assume that \( \epsilon_{ij}(s) \) is the sum of independent components \( \epsilon_{ij}(s) = \epsilon_{1,ij}(s) + \epsilon_{2,ij}(s) \). Here \( \epsilon_{1,ij}(\cdot) \) is a random square integrable function which has smooth covariance function \( \Gamma(s, s') = \)

\[ \int_T c((s, t), (s', t))g(t)dt \]
cov\{\epsilon_{1,ij}(s), \epsilon_{1,ij}(s')\} and \epsilon_{2,ij}(s) is white noise with covariance cov\{\epsilon_{2,ij}(s), \epsilon_{2,ij}(s')\} = \sigma^2 if s = s' and 0 otherwise.

Let \(c((s, t), (s', t')) = E[X_i(s, t)X_i(s', t')]\) be the covariance function of the process \(X_i(\cdot, \cdot)\) and let \(\Sigma(s, s') = \int c((s, t), (s', t))g(t)dt\), where \(g(\cdot)\) is the sampling density of \(t_{ij}\). In Section 3.5 we show that \(\Sigma(s, s')\) is a proper covariance function (Horváth & Kokoszka, 2012); due to its definition we call \(\Sigma\) as the marginal covariance function induced by \(X_i\). Chen et al. (2016) independently considered a similar marginal covariance in a related setting. Denote by \(W_i(s, t_{ij}) = X_i(s, t_{ij}) + \epsilon_{1,ij}(s)\); \(W_i\) is a bivariate process defined on \([0, 1]^2\) and its induced marginal covariance is \(\Xi(s, s') = \Sigma(s, s') + \Gamma(s, s')\). Let \(\{\phi_k(s), \lambda_k\}_k\) be the eigencomponents of \(\Xi(s, s')\), where \(\{\phi_k(\cdot) : k\}\) forms an orthogonal basis in \(L^2[0, 1]\) and \(\lambda_1 \geq \lambda_2 \geq \ldots \geq 0\). Using arguments similar to the standard functional principal component analysis (FPCA), the eigenbasis functions \(\{\phi_k(\cdot) : k = 1, \ldots, K\}\) are optimal in the sense that they minimize the following weighted mean square error: \(\text{MSE}(\theta_1(\cdot), \ldots, \theta_K(\cdot)) = \int_0^1 E\|W_i(\cdot, t) - \sum_{k=1}^K \lambda_k \phi_k(\cdot)\|^2 g(t)dt\), where \(E(f_1(\cdot), f_2(\cdot))(\cdot) = \int_0^1 f_1(s)f_2(s)ds\) is the usual inner product in \(L^2[0, 1]\).

Using the orthogonal basis in \(L^2[0, 1]\) \(\{\phi_k(\cdot)\}_k\), we can represent the square integrable smooth process \(W_i(\cdot, t)\) as \(W_i(s, t_{ij}) = \sum_{k=1}^\infty \xi_{W,ijk}\phi_k(s)\), where \(\xi_{W,ijk} = \int W_i(s, t_{ij})\phi_k(s)ds = \xi_{ik}(t_{ij}) + e_{ijk}\), and \(\xi_{W,ijk}\) are not necessarily uncorrelated over \(k\). Here \(\xi_{ik}(t_{ij}) = \int X_i(s, t_{ij})\phi_k(s)ds\) and \(e_{ijk} = \int \epsilon_{1,ij}(s)\phi_k(s)ds\) are specified by the definition of \(W_i\); for fixed \(k\) these terms are mutually independent due to the independence of the processes \(X_i\) and \(\epsilon_{1,ij}\). For each \(k\), one can easily show that, \(\xi_{ik}(\cdot)\) is a smooth zero-mean random process in \(L^2[0, 1]\) and is iid over \(i\). Furthermore \(e_{ijk}\) are zero-mean iid random variables over \(i, j\); denote by \(\sigma_{e,ik}^2\) their finite variance.

One way to model the dependence of the coefficients, \(\xi_{ik}(t_{ij})'s\), is by using common techniques in longitudinal data analysis; for example by assuming a parametric covari-
ance structure. As we discussed in Section 3.1, this leads to models similar to Greven et al. (2010); Gromenko et al. (2012); Gromenko & Kokoszka (2013). We consider this approach in the analysis of the DTI data, Section 3.7. Another approach is to assume a nonparametric covariance structure and employ a common functional data analysis technique. We detail the latter approach in this section.

For each \( k \geq 1 \) denote by \( G_k(t, t') = \text{cov}\{\xi_{ik}(t), \xi_{ik}(t')\} \) the smooth covariance function in \([0, 1] \times [0, 1]\). Mercer’s theorem provides the following convenient spectral decomposition \( G_k(t, t') = \sum_{l \geq 1} \eta_{kl} \psi_{kl}(t)\psi_{kl}(t') \), where \( \eta_{k1} \geq \eta_{k2} \geq \ldots \geq 0 \) and \( \{\psi_{kl}(\cdot)\}_{l \geq 1} \) is an orthogonal basis in \( L^2[0,1] \). Using the Karhunen-Loève (KL) expansion, we represent \( \xi_{ik}(\cdot) \) as: \( \xi_{ik}(t_{ij}) = \sum_{l=1}^{\infty} \zeta_{ikl} \psi_{kl}(t_{ij}) \), where \( \zeta_{ikl} = \int \xi_{ik}(t)\psi_{kl}(t)dt \), have zero-mean, variance equal to \( \eta_{kl} \), and are uncorrelated over \( l \). By collecting all the components, we represent the model (3.1) as \( Y_{ij}(s) = \mu(s, t_{ij}) + \sum_{k=1}^{K} \sum_{l=1}^{L_k} \zeta_{ikl} \phi_k(s) + \epsilon_{ij}(s), \) for \( \epsilon_{ij}(s) = \sum_{k=1}^{K} e_{ijk} \phi_k(s) + \epsilon_{2,ij}(s) \). In practice we truncate this expansion. Let \( K \) and \( L_1, \ldots, L_K \) such that \( Y_{ij}(s) \) is well approximated by the following truncated model based on the leading \( K \) and \( \sum_k L_k \) respective basis functions

\[
Y_i(s, t_{ij}) = \mu(s, t_{ij}) + \sum_{k=1}^{K} \sum_{l=1}^{L_k} \zeta_{ikl} \phi_k(s) + \epsilon_{ij}(s), \tag{3.2}
\]

where \( \epsilon_{ij}(s) \approx \sum_{k=1}^{K} e_{ijk} \phi_k(s) + \epsilon_{2,ij}(s) \). The truncated model (3.2) gives a parsimonious representation of the longitudinal functional data. It allows to study its dependence through two sets of eigenfunctions: one dependent solely on \( s \) and one solely on \( t_{ij} \). This approach involves two main challenges: first, determining consistent estimator of the marginal covariance and second determining consistent estimators of the time-varying coefficients \( \xi_{ik}(\cdot) \).
3.3 Estimation of model components

We discuss estimation of all model components. The mean estimation is carried out using existing methods (Chen & Müller, 2012; Scheipl et al., 2014); here we briefly describe it for completeness. Our focus and novelty is the estimation of the marginal covariance function and of the eigenfunctions $\phi_k(\cdot)$’s (see Section 3.3.2), as well as the estimation of the time-varying basis coefficients $\xi_{ik}(\cdot)$’s (see Section 3.3.3). Prediction of $Y_i(s, t)$ is detailed in Section 3.3.4.

3.3.1 Step 1: Mean function

As in Scheipl et al. (2014) we estimate the mean function, $\mu(s, t)$, using bivariate smoothing via bivariate tensor product splines (Wood, 2006b) of the pooled data $Y_{ijr} = Y_{ij}(s_r)$’s. Consider two univariate B-spline bases, \{\(B_{s,1}(s), \ldots, B_{s,d_s}(s)\)\} and \{\(B_{t,1}(t), \ldots, B_{t,d_t}(t)\)\}, where $d_s$ and $d_t$ are their respective dimensions. The mean surface is represented as a linear combination of a tensor product of the two univariate B-spline bases $\mu(s, t) = \sum_{q_1=1}^{d_s} \sum_{q_2=1}^{d_t} B_{s,q_1}(s)B_{t,q_2}(t)\beta_{q_1,q_2} = B(s, t)^T \beta$, where $B(s, t)$ is the known $d_s d_t$-dimensional vector of $B_{s,q_1}(s)B_{t,q_2}(t)$’s, and $\beta$ is the vector of unknown parameters, $\beta_{q_1,q_2}$’s. The bases dimensions, $d_s$ and $d_t$, are set to be sufficiently large to accommodate the complexity of the true mean function, and the roughness of the function is controlled through the size of the curvature in each direction separately, i.e. $\int\int \left\{\frac{\partial^2 \mu(s, t)}{\partial s^2}\right\}^2 dt ds = \beta^T (P_s \otimes I_{d_t}) \beta$ in direction $s$, and $\int\int \left\{\frac{\partial^2 \mu(s, t)}{\partial t^2}\right\}^2 dt ds = \beta^T (I_{d_s} \otimes P_t) \beta$ in $t$. The penalized criterion to be minimized is $\sum_{i,j,r} \left[ Y_{ijr} - B(s_r, t_{ij})^T \beta \right]^2 + \beta^T (\lambda_s P_s \otimes I_{d_t} + \lambda_t I_{d_s} \otimes P_t) \beta$, where $\lambda_s$ and $\lambda_t$ are smoothing parameters that control the trade-off between the smoothness of the fit and the goodness of fit. The smoothing parameters can be selected by the restricted maximum likelihood (REML) or generalized cross-validation (GCV). The es-
Estimated mean function is \( \hat{\mu}(s, t) = B(s, t)^T \hat{\beta} \). This method is a very popular smoothing technique of bivariate data.

Other available bivariate smoothers can be used to estimate the mean \( \mu(s, t) \): for example, kernel-based local linear smoother (Hastie et al., 2009), bivariate penalized spline smoother (Marx & Eilers, 2005) and the sandwich smoother (Xiao et al., 2013). The sandwich smoother (Xiao et al., 2013) is especially useful in the case of very high dimensional data for its appealing computational efficiency, in addition to its estimation accuracy.

### 3.3.2 Step 2: Marginal covariance. Data-based orthogonal basis

Once the mean function is estimated, let \( \tilde{Y}_{ijr} = Y_{ijr} - \hat{\mu}(s_r, t_{ij}) \) be the demeaned data. We use the demeaned data to estimate the marginal covariance function induced by \( W_i(s, t_{ij}) \), \( \Xi(s, s') = \Sigma(s, s') + \Gamma(s, s') \). The estimation of \( \Xi(s, s') \) consists of two steps. In the first step, a raw covariance estimator \( \tilde{\Xi}(s_r, s_{r'}) \) is obtained; the pooled sample covariance is a suitable choice, if all the curves are observed on the same grid of points:

\[
\tilde{\Xi}(s_r, s_{r'}) = \sum_{i=1}^{n} \sum_{j=1}^{m_i} \tilde{Y}_{ijr} \tilde{Y}_{ijr'}/(\sum_{i=1}^{n} m_i). \tag{3.3}
\]

As data \( Y_{ijr} \)'s are observed with white noise, \( \epsilon_{2,ij}(s_r) \), the ‘diagonal’ elements of the sample covariance, \( \tilde{\Xi}(s_r, s_r) \), are inflated by the variance of the noise, \( \sigma^2 \). In the second step, the preliminary covariance estimator is smoothed by ignoring the ‘diagonal’ terms; see also Staniswalis & Lee (1998) and Yao et al. (2005a) who used similar technique for the case of independent functional data. In our simulation and data application we use the sandwich smoother (Xiao et al., 2013). To ensure the positive semi-definiteness of the estimator the negative eigenvalues are zero-ed. The resulting smoothed covariance function, \( \hat{\Xi}(s, s') \), is
used as an estimator of $\Xi(s, s')$. In Section 3.5, we show that $\hat{\Xi}(s, s')$ is an unbiased and consistent estimator of $\Xi(s, s')$ in two settings: 1) the data are observed fully and without noise, i.e. $\epsilon_{ij}(s) \equiv 0$ and 2) the data are observed fully and with measurement error of type $\epsilon_{1,ij}(s)$, i.e $\epsilon_{ij}(s) \equiv \epsilon_{1,ij}(s)$.

Let $\{\hat{\phi}_k(s), \hat{\lambda}_k\}_k$ be the pairs of eigenvalues/eigenfunctions obtained from the spectral decomposition of the estimated covariance function, $\hat{\Xi}(s, s')$. The truncation value $K$ is determined based on pre-specified percentage of variance explained (PVE); specifically, $K$ can be chosen as the smallest integer such that

$$\frac{\sum_{k=1}^{K} \hat{\lambda}_k}{\sum_{k=1}^{\infty} \hat{\lambda}_k} > \text{PVE}$$

Di et al., 2009; Staicu et al., 2010).

### 3.3.3 Step 3: Covariance of the time-varying coefficients

Let $\tilde{\xi}_{W,ijk} = \int \tilde{Y}_{ij}(s)\hat{\phi}_k(s)ds$ be the projection of the $j$th repeated demeaned curve of the $i$th subject onto the direction $\hat{\phi}_k(\cdot)$ for $k = 1, \ldots, K$. Since $\tilde{Y}_{ij}(\cdot)$ is observed at dense grids of points $\{s_r : r = 1, \ldots, R\}$ in $[0, 1]$ for all $i$ and $j$, $\tilde{\xi}_{W,ijk}$ is approximated accurately through numerical integration. It is easy to see that the version of $\tilde{\xi}_{W,ijk}$ that uses $\mu(s, t_{ij})$ in place of $\hat{\mu}(s, t_{ij})$ and $\phi_k(s)$ in place of $\hat{\phi}_k(s)$ converges to $\xi_{W,ijk}$ with probability one, as $R$ diverges. The time-varying terms $\tilde{\xi}_{W,ijk}$ are proxy measurements of $\xi_{ik}(t_{ij})$; they will be used to study the temporal dependence along the direction $\phi_k(\cdot)$, $G_k(t, t') = \text{cov}\{\xi_{ik}(t), \xi_{ik}(t')\}$, and furthermore to obtain prediction for all times $t \in [0, 1]$.

Consider now $\{(t_{ij}, \tilde{\xi}_{W,ijk}) : j = 1, \ldots, m_i\}_i$ as the ‘observed data’. One viable approach is to assume a parametric structure for $G_k(\cdot, \cdot)$ such as AR(1) or a random effects model framework; this is typically preferable when $m_i$ is very small and the longitudinal design is balanced. We discuss random effects model for estimating the longitudinal covariance in the data application. Here we consider a more flexible approach and esti-
mate the covariance $G_k(\cdot, \cdot)$ nonparametrically, by employing FPCA techniques for sparse functional data (Yao et al., 2005a).

Let $\{\psi_{kl}(\cdot), \eta_{kl}\}_{l}$ be the pairs of eigenfunctions and eigenvalues of the covariance $G_k$; we model the proxy observations as $\tilde{\xi}_{W,ijk} = \sum_{l=1}^{\infty} \zeta_{kl} \psi_{kl}(t_{ij}) + \tilde{e}_{ijk}$ where $\zeta_{kl}$’s are random variables with zero mean and variances equal to $\eta_{kl}$, $\tilde{e}_{ijk}$’s are iid with zero-mean and variance equal to $\tilde{\sigma}_{e,k}^2$ and independent of $\zeta_{kl}$. Following Yao et al. (2005a), we first obtain the raw sample covariance, $\tilde{G}_{ik}(t_{ij}, t_{ij}') = \tilde{\xi}_{W,ijk} \tilde{\xi}_{W,ij'k}$. Then the estimated smooth covariance surface, $\hat{G}_k(t, t')$, is obtained by using bivariate smoothing of $\{(t_{ij}, t_{ij'}), \tilde{G}_k(t_{ij}, t_{ij'}) : i, j \neq j'\}$. Kernel-based local linear smoothing (Yao et al., 2005a) or penalized tensor product spline smoothing (Wood, 2006b) can be used at this step. The diagonal terms $\{\tilde{G}_k(t_{ij}, t_{ij'}) : i, j = j'\}$ are removed because the noise $\tilde{e}_{ijk}$ leads to inflated variance function. Let $\{\hat{\psi}_{kl}(\cdot), \hat{\eta}_{kl}\}_{l}$ be the pairs of eigenvalues/eigenfunctions of the estimated covariance surface, $\hat{G}_k(t, t')$. The truncation value, $L_k$, is determined based on pre-specified PVE; using similar ideas as in Section 3.3.2. The variance $\tilde{\sigma}_{e,k}^2$ is estimated as the average of the difference between a smooth estimate of the variance based on $\{t_{ij}, \tilde{\xi}_{W,ijk}^2\}$ and $\hat{G}_k(t, t)$; Yao et al. (2005a) discusses an alternative that dismisses the terms at the boundary when estimating the error variance.

Once the eigenbasis functions $\{\psi_{kl}(\cdot)\}_{l=1}^{L_k}$, eigenvalues $\eta_{kl}$’s, and error variance $\tilde{\sigma}_{e,k}^2$ are estimated, the above model framework can be viewed as a mixed effects model and the random components $\zeta_{ijk}$ can be predicted using conditional expectation and a jointly Gaussian assumption for $\zeta_{ijjk}$’s and $e_{ijk}$’s. In particular, $\hat{\zeta}_{ijkl} = \mathbb{E}[\zeta_{ijkl} | \tilde{\xi}_{W,ijk}] = \hat{\eta}_{kl} \hat{\psi}_{ikl}^T \hat{\Sigma}_{W,ik}^{-1} \tilde{\xi}_{W,ijk}$, where $\hat{\psi}_{ikl} = \{\hat{\psi}_{kl}(t_{i1}), \ldots, \hat{\psi}_{kl}(t_{im})\}^T$ is the $m_i$-dimensional column vector of the evaluations of $\hat{\psi}_{kl}(\cdot)$ at $\{t_{ij} : j = 1, \ldots, m_i\}$, $\hat{\Sigma}_{W,ik}$ is a $m_i \times m_i$ - matrix with $(j, j')$th element equal to $\hat{G}_k(t_{ij}, t_{ij'}) + \tilde{\sigma}_{e,k}^2$, for $j = j'$ and $\hat{G}_k(t_{ij}, t_{ij'})$ otherwise, and $\tilde{\xi}_{W,ik}$ is the $m_i$ - dimensional column vector of $\tilde{\xi}_{W,ijk}$’s. The predicted time-varying
coefficients corresponding to a generic time \( t \) are obtained as
\[
\hat{\xi}_{ik}(t) = \sum_{l=1}^{L_k} \hat{\zeta}_{ikl} \hat{\psi}_{kl}(t).
\]
Yao et al. (2005a) proved the consistency of the eigenfunctions and predicted trajectories when \( \xi_{W,ijk} \)’s are observed. In Section 3.5 we extend these results to the case when the proxy \( \tilde{\xi}_{W,ijk} \)’s are used instead and when the profiles \( Y_{ij}(\cdot) \) are fully observed and the noise is of the type \( \epsilon_{ij}(s) = \epsilon_{1,ij}(s) \); i.e. the data \( Y_{ij}(\cdot) \) are observed with smooth error.

### 3.3.4 Step 4: Trajectories reconstruction

We are now able to predict the full response curve at any time point \( t \in [0, 1] \) by:
\[
\hat{Y}_i(s, t) = \hat{\mu}(s, t) + \sum_{k=1}^{K} \hat{\xi}_{ik}(t) \hat{\phi}_k(s),
\]
where \( s \in [0, 1] \). In Section 3.5 we show the consistency of \( \hat{Y}_i(s, t) \).

### 3.4 Implementation using available softwares

An important advantage of the proposed approach is that its implementation can be carried using available software.

**Step 1.** Estimate the smooth mean function \( \hat{\mu}(s, t) \) using the sandwich smoother (Xiao et al., 2013) (the \texttt{fbps} function in \texttt{R} (R Core Team, 2014) package \texttt{refund} (Ciprian Crainiceanu et al., 2014)) or using the penalized tensor product spline smoothing (the \texttt{gam} and \texttt{te} functions in \texttt{R} (R Core Team, 2014) package \texttt{mgcv} (Wood, 2011)).

**Step 2.** Estimate the smooth covariance function \( \hat{\Xi}(s, s') \) with the demeaned data (dense design) using the sandwich smoother (Xiao et al., 2013) and get the eigenfunctions \( \hat{\phi}_k(s) \) using the \texttt{fpca.face} function in the \texttt{refund} package (Ciprian Crainiceanu et al., 2014). The default option of this function also provides \( \tilde{\xi}_{W,ijk} \)’s.
Step 3. For each $k$, carry out FPCA of $\{t_{ij}, \tilde{\xi}_{W,ijk} : i,j\}$. There are several available options for implementation: `fpca.sc` function in the refund package (Ciprian Crainiceanu et al., 2014) and `fpca.mle` and `fpca.pred` functions in the FPCA package (Peng & Paul, 2009; James et al., 2000) in R. Alternatively one can use the FPCA function (Yao et al., 2005a) in the MATLAB (MATLAB, 2014) package PACE (Yao et al., 2005a) available at http://www.stat.ucdavis.edu/PACE/.

Step 4. Determine the predicted trajectories, $\hat{Y}_i(s,t) = \hat{\mu}(s,t) + \sum_{k=1}^{K} \sum_{l=1}^{L_k} \hat{\xi}_{ikl} \hat{\psi}_{kl}(t) \hat{\phi}_k(s)$. We also provide a wrapper function for the proposed method, `fpca.lfda` available in the refund package (Huang et al., 2016).

3.5 Theoretical properties

Next we discuss the asymptotic properties of the estimators and the predicted trajectories. Our setting - sparse longitudinal design and dense functional design - requires new techniques than the ones commonly used for theoretical investigation of repeated functional data such as Chen & Müller (2012). Since the mean estimation has been studied previously, we assume that the response trajectories, $Y_{ij}(\cdot)$’s, have zero-mean and focus on the estimation of the model covariance. Throughout this section we assume that $Y_{ij}(\cdot)$ is observed fully as a function over the domain, $\mathcal{S} = [0, 1]$. Section 3.5.1 discusses the main theoretical results when data are observed without error, i.e. $\epsilon_{ij}(s) \equiv 0$ for $s \in [0, 1]$. Section 3.5.2 extends the results to the case when the data are corrupted with smooth error process $\epsilon_{ij}(s) \equiv \epsilon_{1,ij}(s)$. Section 3.5.3 includes a discussion on how to relax some of the assumptions. The proofs are detailed in Appendix B.1.
use $\mathcal{S}$ and $\mathcal{T}$ to distinguish between the domains.

We assume that the bivariate process $X_i(s,t)$ is a realization of a true random process, $X(s,t)$, with zero-mean and smooth covariance function, $c((s,t),(s',t'))$, which satisfies some regularity conditions:

(A1.) $X = \{X(s,t) : (s,t) \in \mathcal{S} \times \mathcal{T}\}$ is a square integrable element of the $L^2(\mathcal{S} \times \mathcal{T})$, i.e. $E[\iint X^2(s,t)dsdt] < \infty$, where $\mathcal{S}$ and $\mathcal{T}$ are compact sets.

(A2.) The sampling density $g(t)$ is continuous and $\sup_{t \in \mathcal{T}} |g(t)| < \infty$.

Under (A1.) and (A2.), the function $\Sigma(s,s')$ defined above (i) is symmetric, (ii) is positive definite, and (iii) has eigenvalues $\lambda_k$’s with $\sum_{k=1}^{\infty} \lambda_k < \infty$. Thus $\Sigma(\cdot,\cdot)$ is a proper covariance function (Horváth & Kokoszka, 2012, p.24).

3.5.1 Response curves measured without error

Assume $\epsilon_{ij}(s) \equiv 0$ and thus $Y_{ij}(s) = X_i(s,t_{ij})$ for $s \in \mathcal{S}$. The sample covariance of $Y_{ij}(s)$ is $\hat{\Sigma}(s,s') = \sum_{i=1}^{n} \sum_{j=1}^{m_i} Y_{ij}(s)Y_{ij}(s')/ (\sum_{i=1}^{n} m_i)$. In Appendix B.1.1 we first show that $\hat{\Sigma}(s,s')$ is an unbiased estimator of the true marginal covariance, $\Sigma(s,s')$ and then we prove that it is a consistent estimator. The following assumptions regard the moment behavior of $X$ and are commonly used in functional data analysis (Yao et al., 2005a; Chen & Müller, 2012); we require them in our study.

(A3.) $E[X(s,t)X(s',t)X(s',t')] < \infty$ for arbitrary $s,s' \in \mathcal{S}$ and $t,t' \in \mathcal{T}$.

(A4.) $E[\|X(\cdot,t)\|^4] < \infty$ for each $t \in \mathcal{T}$.

Theorem 1 gives the asymptotic properties of the marginal covariance estimator $\hat{\Sigma}(s,s')$.
Theorem 1. Assume (A1.) - (A3.) hold. Then $|\hat{\Sigma}(s,s') - \Sigma(s,s')| \xrightarrow{p} 0$ as $n$ diverges. If in addition (A4.) holds, then

$$\|\hat{\Sigma}(\cdot,\cdot) - \Sigma(\cdot,\cdot)\|_s \xrightarrow{p} 0 \text{ as } n \rightarrow \infty,$$  \hspace{1cm} (3.4)

where $\|k(\cdot,\cdot)\|_s = \{ \iint k^2(s,s')dsds' \}^{1/2}$ is the Hilbert-Schmidt norm of $k(\cdot,\cdot)$.

(A5.) Let $a_1 = (\lambda_1 - \lambda_2)$ and $a_k = \max[(\lambda_k - 1 - \lambda_k), (\lambda_k - \lambda_{k+1})]$ for $k \geq 2$, where $\lambda_k$ is the $k$th largest eigenvalues of $\Sigma(s,s')$. Assume that $0 < a_k < \infty$ and $\lambda_k > 0$ for all $k$ (no crossing or ties among eigenvalues).

Using Theorem 4.4 and Lemma 4.3 of Bosq (2000, p.104), the consistency result (3.4) implies that, if furthermore (A5.) holds, the eigen-elements of $\hat{\Sigma}(s,s')$ are consistent estimators of the corresponding eigen-elements of $\Sigma(s,s')$.

Corollary 1. Under the assumptions (A1.-A5.), for each $k$ we have $|\hat{\lambda}_k - \lambda_k| \xrightarrow{p} 0$, and $\|\hat{\phi}_k(\cdot) - \phi_k(\cdot)\|_s \xrightarrow{p} 0 \text{ as } n \text{ diverges.}$

Next, we focus on the estimation of the covariance $G_k(t,t') = \text{cov}\{\xi_{ik}(t_{ij}), \xi_{ik}(t_{ij}')\}$, which describes the longitudinal dynamics. If $\xi_{W,ijk}$’s were available to estimate the covariance function, $G_k(t,t')$, then the consistency of the model components follows trivially from the FPCA properties developed in Yao et al. (2005a). However, $\xi_{W,ijk}$’s are not directly observed; instead available are the proxy time-varying coefficients $\tilde{\xi}_{W,ijk} = \int Y_i(s,t_{ij})\hat{\phi}(s)ds$. We first show the uniform consistency of $\tilde{\xi}_{W,ijk}$; the result follows if $\sup_{j,s}|Y_i(s,t_{ij})|$ is bounded almost surely, which is ensured if (A6.) holds. Then, we use this result to show that the estimator of $G_k(t,t')$ based on $\tilde{\xi}_{W,ijk}$’s is asymptotically identical to that based on $\xi_{W,ijk}$. Consistency results of the remaining model components
follow directly from Yao et al. (2005a). The Gaussian assumption \((A8.)\) is needed to show the consistency of \(\hat{\zeta}_{ikl}\).

\((A6.)\) \(E[\sup_{s,t}|X(s,t)|^a] \leq M^a\) for a constant, \(M > 0\), and an arbitrary integer, \(a \geq 1\); this is equivalent to assume that \(X(s,t)\) is absolutely bounded almost surely.

\((A7.)\) Let \(b_{kl} = (\eta_{k1} - \eta_{k2})\) and \(b_{kl} = \max[(\eta_{k(l-1)} - \eta_{kl}), (\eta_{kl} - \eta_{k(l+1)})]\) for \(l \geq 2\), where \(\eta_{kl}\) is the \(l\)th largest eigenvalues of \(G_k(t,t')\). Assume that \(0 < b_{kl} < \infty\) and \(\eta_{kl} > 0\) for all \(k\) and \(l\).

\((A8.)\) \(\zeta_{ikl}\) and \(e_{ijk}\) are jointly Gaussian.

**Theorem 2.** Under the assumptions \((A1.) - (A6.)\), for each \(k\) \(\sup_j|\tilde{\xi}_{W,ijk} - \xi_{W,ijk}| \xrightarrow{p} 0\) and \(\|\hat{G}_k(\cdot,\cdot) - G_k(\cdot,\cdot)\| \xrightarrow{p} 0\) as \(n\) diverges. In fact a stronger result also holds, namely \(\sup_{t,t'}|\hat{G}_k(t,t') - G_k(t,t')| \xrightarrow{p} 0\) as \(n\) diverges.

**Corollary 2.** Assume \((A1.) - (A8.)\) hold for each \(k\) and \(l\). Then the eigenvalues \(\hat{\eta}_{kl}\) and eigenfunctions \(\hat{\psi}_{kl}(\cdot)\) of \(\hat{G}_k(\cdot,\cdot)\) satisfy \(|\hat{\eta}_{kl} - \eta_{kl}| \xrightarrow{p} 0\), and \(\|\hat{\psi}_{kl}(\cdot) - \psi_{kl}(\cdot)\| \xrightarrow{p} 0\) as \(n\) diverges. Uniform convergence of \(\hat{\psi}_{kl}(\cdot)\) also holds: \(\sup_{t}|\hat{\psi}_{kl}(t) - \psi_{kl}(t)| \xrightarrow{p} 0\). Furthermore, as \(n\) diverges, we have \(|\hat{\sigma}_{e,k}^2 - \sigma_{e,k}^2| \xrightarrow{p} 0\) and \(|\hat{\zeta}_{ikl} - \zeta_{ikl}| \xrightarrow{p} 0\), where \(\tilde{\zeta}_{ikl} = E[\zeta_{ikl}|\xi_{W,ik}]\) and \(\xi_{W,ik}\) is the \(m_i\)-dimensional column vector of \(\xi_{W,ijk}\)'s.

The consistency results for all model components imply consistency of the predicted trajectories \(\hat{Y}_i(s,t)\).

**Theorem 3.** Assume \((A1.) - (A8.)\), for each \((s,t) \in S \times T\). Then \(\hat{Y}_i(s,t)\) converges to \(\sum_{k=1}^{\infty} \sum_{l=1}^{\infty} \tilde{\zeta}_{ikl} \psi_{kl}(t) \phi_k(s)\) in probability as \(n, K\) and \(L_k\)'s \(\to \infty\).

The detailed proofs are given in Appendix B.1.1.
3.5.2 Response curves measured with smooth error

Assume next that $Y_{ij}(s)$ are observed with smooth error $\epsilon_{ij}(s) \equiv \epsilon_{1,ij}(s)$ and thus $Y_{ij}(s) = X_i(s,t_{ij}) + \epsilon_{1,ij}(s)$ for $s \in \mathcal{S}$ and $\epsilon_{1,ij}(\cdot) \in L^2(\mathcal{S})$.

The main difference from Section 3.5.1 is that the sample covariance of $Y_{ij}(s)$ is an estimator of $\Xi(s,s') = \Sigma(s,s') + \Gamma(s,s')$, not of $\Sigma(s,s')$; we denote the sample covariance of $Y_{ij}(s)$ by $\hat{\Xi}(s,s') = \sum_{i=1}^{n} \sum_{j=1}^{m} Y_{ij}(s)Y_{ij}(s') / (\sum_{i=1}^{n} m_i)$. Using similar arguments as earlier, we show that $\hat{\Xi}(s,s')$ is an unbiased estimator of $\Xi(s,s')$. Moreover similar arguments can be used to show the pointwise consistency as well as the Hilbert-Schmidt norm consistency of $\hat{\Xi}(s,s')$. Additional assumptions are required.

(A9.) Assume $\epsilon_{ij}(\cdot)$ is realization of $\epsilon = \{\epsilon(s) : s \in \mathcal{S}\}$, which is square integrable process in $L^2(\mathcal{S})$.

(A10.) $E[\|\epsilon(\cdot)\|^4] < \infty$.

(A11.) $E[\sup_s |\epsilon(s)|^{a}] \leq M^a$ for a constant, $M > 0$, and an arbitrary integer, $a \geq 1$.

**Corollary 3.** Under the assumptions (A1.) - (A3.), and (A9.), for each $(s,s')$, $|\hat{\Xi}(s,s') - \Xi(s,s')| \overset{p}{\to} 0$ as $n$ diverges. And under the assumptions (A1.)-(A4.), (A9.)-(A11.), $\|\hat{\Xi}(\cdot,\cdot) - \Xi(\cdot,\cdot)\|_s \overset{p}{\to} 0$ and $\sup_j |\widehat{\xi}_{W,ijk} - \xi_{W,ijk}| \overset{p}{\to} 0$ as $n \to \infty$.

The proofs of these results are detailed in Appendix B.1.2. As the smooth error process $\epsilon_{1,ij}(s)$ is correlated only along the functional argument, $s$, and $\epsilon_{1,ij}(s)$ are iid over $i,j$, it follows that the theoretical properties of the predictions - of the time-varying coefficients and the response curve - hold without any modification.

The theoretical results are based on the assumptions that data $Y_{ij}(s)$’s are observed fully, without white noise, $\epsilon_{2,ij}(s) \equiv 0$ for all $s$, and have mean zero. Some of these
assumptions are quite common in theoretical study involving functional data; Cardot et al. (2003a, 2004); Chen & Müller (2012). They are discussed in Appendix B.

3.5.3 Extensions

The theoretical results presented in Section 3.5 are based on the assumption that data $Y_{ij}(s)$’s are observed (i) fully, (ii) without white noise, $\epsilon_{2,ij}(s) \equiv 0$ for all $s$, and (iii) have mean zero. These assumptions are made for convenience, and they can be relaxed as we now explain.

(i) The assumption that $Y_{ij}(s)$’s are observed fully, as a continuous function, is quite common in theoretical study involving functional data; see for example Cardot et al. (2003a, 2004); Chen & Müller (2012) among many others. One possibility to bypass this assumption is to use the corresponding smooth trajectories instead. (ii) Suppose that the profiles $Y_{ij}(\cdot)$’s are observed on dense grids of points and the measurements are additionally corrupted with white noise. Zhang & Chen (2007) showed that by smoothing each profile using local linear smoother, the true de-noised curves are recovered with asymptotically negligible error, in the case of independent curves. Another possibility to handle white noise is to use ideas similar to Yao et al. (2005a). In Section 3.6 we illustrate numerically the effect of white noise on the performance accuracy. (iii) Finally, the theoretical properties of the model components estimators remain valid, when the mean function is non-zero and a consistent mean estimator is available; Chen & Müller (2012) had considered this problem and showed that under suitable assumptions such consistent mean estimator can be obtained using bivariate smoothing.
3.6 Simulation study

We study our approach in finite samples and compare its performance with Chen & Müller (2012) denoted by CM. We generate $N_{\text{sim}} = 1000$ samples from model (3.1) with $K = 2$, $Y_{ij}(s) = \mu(s, t_{ij}) + \xi_{i1}(t_{ij})\phi_1(s) + \xi_{i2}(t_{ij})\phi_2(s) + \epsilon_{ij}(s)$, where $\mu(s, t) = 1 + 2s + 3t + 4st$, and $\phi_1(s) = 1$ and $\phi_2(s) = \sqrt{2}\sin(2\pi s)$. The grid of points for $s$ is the set of 101 equispaced points in $[0, 1]$. For each $i$, there are $m_i$ profiles associated with visit times, $\{t_{ij} : j = 1, \ldots, m_i\}$; $t_{ij}$’s are randomly sampled from 41 equally spaced points in $[0, 1]$. $\xi_{ik}(t)$ are generated from various covariance structures: (a) non-parametric covariance (NP) where $\xi_{ik}(t) = \zeta_{ik1}\psi_{k1}(t) + \zeta_{ik2}\psi_{k2}(t)$; (b) random effects model (REM) $\xi_{ik}(t) = b_{ik0} + b_{ik1}t$, and (c) exponential autocorrelation (Exp) $\text{cov}\{\xi_{ik}(t), \xi_{ik}(t')\} = \lambda_{k}\rho^{\left|t-t'\right|}$. Errors are generated from $\epsilon_{ij}(s) = e_{ij1}\phi_1(s) + e_{ij2}\phi_2(s) + \epsilon_{2,ij}(s)$, where $e_{ij1}$, $e_{ij2}$ and $\epsilon_{2,ij}(s)$ are mutually independent with zero-mean and variances equal to $\sigma_{\epsilon,1}^2$, $\sigma_{\epsilon,2}^2$ and $\sigma^2$, respectively; the white noise variance, $\sigma^2$, is set based on the signal to noise ratio (SNR). The details of the models are specified in Appendix B. For each sample of size $n$ we form a training set and a test set. The test set contains 10 profiles and is obtained as follows: randomly select 10 subjects from the sample and collect the subjects’ last profile. The remaining profiles for the 10 subjects and the data corresponding to the rest ($n - 10$) of the subjects form the training set. Our model is fitted using the training set and the methods of Section 3.3. The mean function, $\mu(s, t)$, is modeled using 50 cubic spline basis functions obtained from the tensor product of $d_s = 10$ basis functions in direction $s$ and $d_t = 5$ in $t$. The smoothing parameters are selected via REML. The finite truncations $K$ and $L_k$’s are all estimated using the pre-specified level PVE = 0.95.

Estimation accuracy for the model components is evaluated using integrated mean squared errors (IMSE), while prediction performance is assessed through in-sample inte-
grated prediction errors (IN-IPE) and out-of-sample IPE (OUT-IPE). Table 3.1 shows the results for different covariance models for $\xi_{ik}(t)$, different number of repeated curve measurements per subject, different SNRs, complex error process, and varying sample sizes. The performance of the proposed estimation (see columns for $\mu$, $\phi_1$, and $\phi_2$ of this table) is slightly affected by the covariance structure of $\xi_{ik}(t)$’s and $m_i$, but in general is quite robust to the factors we investigated. As expected the estimation accuracy improves with larger sample size; see the $3 \times 3$ top left block of IMSE results corresponding to $n = 100$, $n = 300$, and $n = 500$. Moreover both the prediction of $\xi_{ik}(t)$’s and that of $Y_{ij}(\cdot)$ are considered; see columns labeled $\xi_1$, $\xi_2$, IN-IPE and OUT-IPE of Table 3.1. The underlying covariance structure of $\xi_{ik}(t)$’s affects the prediction accuracy. Furthermore increasing the number of repeated curve measurements $m_i$ improves the accuracy more than increasing the sample size $n$. This observation should not be surprising, as with larger number of repeated measurements the estimation of the covariance of the longitudinal process $\xi_{ik}(t)$’s improves and as a result it yields superior prediction. We compared our results with another, rather naïve approach: predict a subject’s profile by the average of all previously observed profiles for that subject. The naïve approach (see columns IN-IPE$_{naïve}$ and OUT-IPE$_{naïve}$) is very sensitive to the covariance structure of $\xi_{ik}(t)$. In all the cases studied the prediction accuracy is inferior to the proposed method.

Table 3.2 shows the comparison with CM, when the kernel bandwidth is fixed to $h = 0.1$ for both mean and covariance smoothing. The prediction using CM is more sensitive to the covariance structure of the underlying time-varying coefficients $\xi_{ik}(t)$ and its accuracy can be improved by up to 50% using our proposed approach. Computation-wise, there is an order of magnitude difference in the computational cost between the methods: when $n = 100$ CM takes over 16 minutes, while our approach takes about 7 seconds. The overall conclusion is that the proposed approach provides an improved
prediction performance over the existing methods in a computationally efficient manner.

### 3.7 DTI application

DTI is a magnetic resonance imaging technique, which provides different measures of water diffusivity along brain white matter tracts; its use is instrumental especially in diseases that affect the brain white matter tissue, such as MS (see Alexander et al. (2007), Basser et al. (1994), Basser et al. (2000), Basser & Pierpaoli (2011)). In this chapter we consider the DTI measure called FA along CCA; specifically we consider one-dimensional summaries of FA along CCA (CCA-FA). The DTI study involves 162 MS patients, which are observed at between one and eight hospital visits, with a total of 421 visits and a median of two visits per subject. At each visit, FA profile is recorded at 93 locations along the CCA. The measurements are registered within and between subjects using standard biological landmarks identified by an experienced neuroradiologist (Scheipl et al., 2014).

Our main objective is twofold: (i) to understand the dynamic behavior of the CCA-FA profile in MS patients over time and (ii) to make accurate predictions of the CCA-FA profile of a patient at their next visit. Various aspects of the DTI study have been also considered in Goldsmith et al. (2011), Staicu et al. (2012), Pomann et al. (2013), and Scheipl et al. (2014). Greven et al. (2010) used an earlier version of the DTI study consisting of data from fewer and possibly different patients and obtained through a different registration technique. They studied the dynamic behavior of CCA-FA over time in MS; however, their method cannot provide prediction of the entire CCA-FA profile at the subject’s next visit. By being able to predict the full CCA-FA profile at the subject’s future visit, our approach has the potential to shed lights on the understanding of the MS progression over time as well as its response to treatment.
Table 3.1: Estimation and prediction accuracy results based on $N_{\text{sim}} = 1000$ simulations

$$m_i \sim \{8, \ldots, 12\} \text{ and } SNR = 1$$

<table>
<thead>
<tr>
<th>$n$</th>
<th>$\mu$</th>
<th>$\phi_1$</th>
<th>$\phi_2$</th>
<th>$\xi_1$</th>
<th>$\xi_2$</th>
<th>IN-PE</th>
<th>IN-PE$_{\text{naive}}$</th>
<th>OUT-PE</th>
<th>OUT-PE$_{\text{naive}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>0.092</td>
<td>0.003</td>
<td>0.011</td>
<td>0.338</td>
<td>0.224</td>
<td>0.406</td>
<td>7.790</td>
<td>0.988</td>
<td>11.478</td>
</tr>
<tr>
<td>300</td>
<td>0.031</td>
<td>0.001</td>
<td>0.009</td>
<td>0.226</td>
<td>0.138</td>
<td>0.313</td>
<td>7.773</td>
<td>0.550</td>
<td>11.349</td>
</tr>
<tr>
<td>500</td>
<td>0.019</td>
<td>0.001</td>
<td>0.009</td>
<td>0.199</td>
<td>0.117</td>
<td>0.288</td>
<td>7.779</td>
<td>0.455</td>
<td>11.262</td>
</tr>
<tr>
<td>100</td>
<td>0.114</td>
<td>0.027</td>
<td>0.033</td>
<td>0.376</td>
<td>0.314</td>
<td>0.328</td>
<td>1.199</td>
<td>1.011</td>
<td>2.160</td>
</tr>
<tr>
<td>300</td>
<td>0.040</td>
<td>0.008</td>
<td>0.013</td>
<td>0.216</td>
<td>0.162</td>
<td>0.265</td>
<td>1.197</td>
<td>0.675</td>
<td>2.160</td>
</tr>
<tr>
<td>500</td>
<td>0.024</td>
<td>0.005</td>
<td>0.010</td>
<td>0.181</td>
<td>0.133</td>
<td>0.247</td>
<td>1.197</td>
<td>0.571</td>
<td>2.150</td>
</tr>
<tr>
<td>100</td>
<td>0.095</td>
<td>0.022</td>
<td>0.030</td>
<td>0.399</td>
<td>0.540</td>
<td>0.554</td>
<td>1.528</td>
<td>1.426</td>
<td>2.520</td>
</tr>
<tr>
<td>300</td>
<td>0.031</td>
<td>0.007</td>
<td>0.015</td>
<td>0.289</td>
<td>0.412</td>
<td>0.508</td>
<td>1.531</td>
<td>1.143</td>
<td>2.498</td>
</tr>
<tr>
<td>500</td>
<td>0.019</td>
<td>0.004</td>
<td>0.013</td>
<td>0.266</td>
<td>0.383</td>
<td>0.494</td>
<td>1.530</td>
<td>1.074</td>
<td>2.492</td>
</tr>
</tbody>
</table>

$$m_i \sim \{15, \ldots, 20\} \text{ and } SNR = 1$$

$$m_i \sim \{8, \ldots, 12\} \text{ and } SNR = 5$$

$$m_i \sim \{15, \ldots, 20\} \text{ and } SNR = 5$$
Table 3.2: Comparison between the proposed method and Chen & Müller (2012) in the presence of correlated errors. Results based on $N_{sim} = 1000$ simulations

$m_i \sim \{8, \ldots, 12\}$ and $SNR = 1$

<table>
<thead>
<tr>
<th></th>
<th>Chen &amp; Müller (2012)</th>
<th>Proposed method (from Tables 3.1 and B.2)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>IN-IPE</td>
<td>OUT-IPE</td>
</tr>
<tr>
<td>NP (a)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$n = 100$</td>
<td>0.880</td>
<td>2.221</td>
</tr>
<tr>
<td>$n = 300$</td>
<td>0.622</td>
<td>1.468</td>
</tr>
<tr>
<td>$n = 500$</td>
<td>0.556</td>
<td>1.298</td>
</tr>
<tr>
<td>REM (b)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$n = 100$</td>
<td>0.424</td>
<td>1.359</td>
</tr>
<tr>
<td>$n = 300$</td>
<td>0.289</td>
<td>0.729</td>
</tr>
<tr>
<td>$n = 500$</td>
<td>0.257</td>
<td>0.614</td>
</tr>
<tr>
<td>Exp (c)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$n = 100$</td>
<td>0.634</td>
<td>1.642</td>
</tr>
<tr>
<td>$n = 300$</td>
<td>0.549</td>
<td>1.251</td>
</tr>
<tr>
<td>$n = 500$</td>
<td>0.531</td>
<td>1.155</td>
</tr>
</tbody>
</table>

To start with, for each subject we define the hospital visit time $t_{ij}$ by the difference between the reported visit time and the subject’s baseline visit time; thus $t_{i1} = 0$ for all subjects $i$. Then the resulting values are scaled by the maximum value in the study so that $t_{ij} \in [0, 1]$ for all $i$ and $j$. The sampling distribution of the visit times is right-skewed with rather strong skewness; for example there are only few observations $t_{ij}$’s close to 1. The strong skewness of the sampling distribution of $t_{ij}$’s has serious implications on the estimation of the bivariate mean $\mu(s, t)$; a completely nonparametric bivariate smoothing would results in unstable and highly variable estimation. This is probably why Greven et al. (2010) first centered the times for each patient $i$, $\{t_{ij} : j = 1, \ldots, m_i\}$, and then standardized the overall set $\{t_{ij} : i,j\}$ to have unit variance. However, such subject-specific transformation of $t_{ij}$’s loses interpretability and it is not suited for prediction at unobserved times - which is crucial in our analysis. One way to bypass this issue is to assume a simpler parametric structure along the longitudinal direction, $t$, for the mean function; based on exploratory analysis we assume linearity in $t$. Specifically we consider $\mu(s, t_{ij}) = \mu_0(s) + \beta_t(s)t_{ij}$, where $\mu_0(\cdot)$ and $\beta_t(\cdot)$ are unknown, smooth functions of $s$. We estimate $\mu_0(\cdot)$ and $\beta_t(\cdot)$ using a penalized univariate cubic spline regression with 10
basis functions; the smoothing parameters are estimated using REML. The estimates \( \hat{\mu}(s, t) \) and \( \hat{\beta}_t(s) \) are displayed in Figure B.1 of Appendix B. Using the bootstrap of subjects-based methods (Park et al., 2016a) proposed in Chapter 2 of this dissertation, and \( B = 1000 \) bootstrap samples we construct 95\% joint confidence bands for \( \hat{\beta}_t(s) \); see Figure 3.1. The confidence band contains zero for all \( s \), indicating evidence that a mean model \( \mu(s, t_{ij}) = \mu_0(s) \) is more appropriate.

Next we demean the data and estimate the marginal covariance; using a preset level \( PVE = 0.95 \) we obtain \( K = 10 \) eigenfunctions. Figure 3.2 shows the leading 3 eigenfunctions that explain in turn 62.69\%, 8.37\% and 6.77\% of the total variance; the rest of the estimated eigenfunctions are given in Figure B.3 of Appendix B. Preliminary investigation (not shown here) indicates a simpler model for the longitudinal covariance: a random effects model \( \xi_{ik}(t_{ij}) = b_{0ik} + b_{1ik}t_{ij} \), where \( \text{var}(b_{lik}) = \sigma^2_{lk} \) for \( l = 0, 1 \) and \( \text{cov}(b_{0ik}, b_{1ik}) = \sigma_{01k} \). This resulting model is similar to Greven et al. (2010). The fitted time-varying coefficient functions, \( \hat{\xi}_{ik}(t) \), for \( k = 1, 2 \) and 3 are shown in Figure 3.3, and
Figure 3.2: Top: First three eigenfunctions of the estimated marginal covariance; Bottom: estimated mean function $\hat{\mu}_0(s)$ (gray line) ± $2\sqrt{\hat{\lambda}_k}\hat{\phi}_k(s)$ (+ and − signs, respectively)

the rest are shown in Figure B.4 of Appendix B. The estimated $\hat{\xi}_{i1}(t)$ suggest some longitudinal changes, but the signs generally remain constant across time. The results imply that a subject mean profile tends to stay lower than the population mean, if the first eigenfunction corresponding to that individual is positively loaded at baseline, and vice versa. In contrast, $\hat{\xi}_{i2}(t)$, are mostly constant across visit times and imply little changes over time.

Finally, we assess the goodness-of-fit and prediction accuracy of our final model. For the goodness-of-fit we use the in-sample integrated prediction error (IN-IPE): IN-IPE = $\sum_{i=1}^{162} \sum_{j=1}^{m_i} \int \{Y_{ij}(s) - \hat{Y}_{ij}(s)\}^2 ds / \{\sum_{i=1}^{162} m_i\}$, where $\hat{Y}_{ij}(s) = \hat{\mu}_0(s) + \sum_{k=1}^{K} (\hat{b}_{0ik} + \hat{b}_{1ik}t_{ij})\hat{\phi}_k(s)$, and $Y_{ij}(\cdot)$’s are the observed curve data. The square root of the IN-IPE is $2.31 \times 10^{-2}$ for our model; for comparison Greven et al. (2010) yields $2.66 \times 10^{-2}$
Figure 3.3: Estimated time-varying coefficients $\hat{\xi}_{ik}(t)$ for $k = 1, 2$ and 3 using REM and Chen & Müller (2012) gives $3.76 \times 10^{-2}$. For prediction accuracy we use leave-the last-curve-out integrated prediction error (OUT-IPE) calculated for the 106 subjects observed at two hospital visits or more: 

$$\text{OUT-IPE} = \sum_{i=1}^{106} \int \{ Y_{imi}(s) - \hat{Y}_{imi}[-imi](s) \}^2 ds / 106,$$

where $\hat{Y}_{imi}[-imi](s)$ is the predicted curve at time $t_{imi}$ for the $i$th subject using the fitted model based on all the data less the $m_i$th curve of the $i$th subject. Figure 3.4 shows such predicted curves $\hat{Y}_{imi}[-imi](s)$ obtained using our model and the naive model for three randomly selected subjects at their last visit. The square root of OUT-IPE is $3.48 \times 10^{-2}$ for our model; for comparison Chen & Müller (2012) gives $8.71 \times 10^{-2}$ and the naïve approach gives $3.52 \times 10^{-2}$. These results suggest that, in this short term study of MS, there is a small variation of CCA-FA profiles over time.

### 3.8 Visualization using Interactive Graphics

As discussed in Section 3.4 a function `fpca.lfda` is provided in the `refund` package (Huang et al., 2016) for implementation of the proposed method. We also provide an interactive graphics tool in the package `refund.shiny` (Wrobel & Goldsmith, 2016) to
help explore any longitudinal functional data and visualize estimation results obtained using the proposed method Wrobel et al. (2016).

Interactive graphics for the proposed method are implemented by the code below:

```r
> MS <- subset(DTI, case ==1)
> index.na <- which(is.na(MS$cca))
> Y <- MS$cca; Y[index.na] <- fpca.sc(Y)$Yhat[index.na]
> id <- MS$ID
> visit.index <- MS$visit
> visit.time <- MS$visit.time/max(MS$visit.time)
> fit.tfpca <- fpca.lfda(Y = Y, subject.index = id,
+   visit.index = visit.index, obsT = visit.time,
+   LongiModel.method = ‘lme’)
> plot_shiny(fit.tfpca)
```

The code produces an interface with two tabs. Tab 1 given in Figure 3.5 shows exploratory plots and includes three inset sub-tabs. The first sub-tab plots the observed curves for any user-selected subject, and includes options to display the observed curves of all subjects in the background and to display the estimated pointwise mean curve,
denoted by \( m(t) \). The second sub-tab allows the user to see the longitudinal changes of the observed curves for a user-selected subject \( i \); a slider bar animates the subject’s visit times and highlights the corresponding observed curve in the plot. The last sub-tab shows two plots of the actual visit times \( t_{ij} \): the bottom plot presents static histogram of visit times of all subjects, while the top plot presents all of observed visit times on a horizontal line to help visualize the sparsity of the longitudinal sampling.

Tab 2 given in Figure B.5 shows estimated model components and predictions, and includes 8 inset sub-tabs. Sub-tabs 1 and 2 present static images of the estimated mean surface \( \hat{\mu}(s,t) \) and estimated marginal covariance \( \hat{\Sigma}(s,s') \). Sub-tabs 3, 4, and 5 illustrate the first step of estimation, and plot estimates of eigenfunctions \( \hat{\psi}_k(t), m(t) \pm 2\sqrt{\hat{\lambda}_k}\hat{\psi}_k(t) \), and static scree plots of the estimated eigenvalues \( \hat{\lambda}_k \), respectively. Sub-tab 6 shows the estimated covariance of the time-varying FPC loadings \( c_{ik}(\cdot) \) for user specified \( k \). Sub-tab 7 shows the prediction of the time-varying basis coefficient \( c_{ik}(t) \) for any user-selected subject \( i \) and component \( k \); it also has an option of displaying predicted values of \( c_{ik}(t) \) for all subjects in the background. Lastly, sub-tab 8 shows the prediction of a full response trajectory \( Y_i(\cdot,t) \) for user-selected subject \( i \) in animation with change of values across 21 equi-spaced grid of points of \( t \) in the range of observed visit times of all subjects.
Figure 3.5: Screenshot showing tab 1 of the interactive graphic. The plots show observed data of all subjects (top) and of the selected subject (bottom).
Chapter 4

Significance Testing in Longitudinal Functional Data

4.1 Introduction

In the previous chapter our primary focus was to model the complex variation of longitudinal functional response. Now we turn our attention to the mean. This work is also motivated by the longitudinal diffusion tensor imaging (DTI) study. Recall from Chapter 3 that the longitudinal DTI study involves 162 multiple sclerosis (MS) patients, where each patient was observed at several hospital visits and at each visit a fractional anisotropy (FA) profile was measured along the corpus callosum (CCA) tract. One of the fundamental scientific questions in the DTI study is: does the mean FA trajectory vary over visit time? To formally address such question we develop a novel inferential procedure for testing about the time-varying mean for longitudinal functional data.

Let \( \{ Y_{ij}(s) : s \in S \} \) be the \( j \)-th profile of the \( i \)-th subject observed at actual time of visit \( t_{ij} \). Consider the model \( Y_{ij}(s) = \mu(s, t_{ij}) + \epsilon_i(s, t_{ij}) \), where \( \mu(\cdot, t_{ij}) \) is an unknown smooth mean function evaluated at \( t_{ij} \) and \( \epsilon_i(\cdot, t_{ij}) \) is the random deviation from the mean function. It is assumed that \( \epsilon_i \)'s are independent and identically distributed.
(i.i.d) bivariate processes with mean zero and an unknown covariance function with very complex structure induced by both functional and longitudinal designs; full model assumptions are deferred to Section 4.3. Our primary interest in this chapter is to test the null hypothesis that the mean function $\mu(s,t)$ does not vary over visit time $t$.

One naïve approach is to completely ignore the dependence over the functional argument $s$ but account for the dependence across repeated visits, i.e. to assume that $Y_{ij}(s)$ are independent over $s$ but correlated over $j$. This approach views the observed responses $Y_{ij}(s)$ as longitudinal data and an exhaustive list of modeling, estimation, and inferential methods for the mean is available in longitudinal data analysis literature; see for example Laird & Ware (1982), Liang & Zeger (1986), and Fitzmaurice et al. (2012).

Another possible approach is to ignore the dependence across the repeated visits but account for the functional dependence, i.e. to assume that $Y_{ij}(s)$ are independent over $j$ but correlated over $s$. Under this assumption the observed responses $Y_{ij}(s)$ are viewed as independent functional data and the underlying covariance is no longer dependent on visit time $t_{ij}$. In this context function on scalar regression models, such as Faraway (1997) and Jiang et al. (2011), can be used to estimate the fixed effect of $t_{ij}$, and several testing procedures are available to test fixed effects under such models; see for example Shen & Faraway (2004) and Zhang & Chen (2007).

Nevertheless, these approaches fail to account for all sources of dependence that are present in longitudinal functional data and lead to inflated type I error. Furthermore extension of the existing methods to dependent functional data in a computationally feasible way is not straightforward. On the other hand, statistical inference about the mean function for dependent functional data has received less attention. For example, Morris & Carroll (2006) discussed Bayesian inference in the functional mixed model framework; however, their main focus was on modeling and not on hypothesis testing. Crainiceanu
et al. (2011) discussed bootstrap-based inferential methods for testing the mean difference between paired functional processes. Staicu et al. (2014a) proposed likelihood ratio based testing for the mean difference between two independent samples of correlated functional processes. Staicu et al. (2014b) considered testing that mean functions are the same among multiple groups and introduced $L^2$ norm based test statistic with its asymptotic null distribution. Horváth et al. (2013) developed inference for the mean function of a functional time series. Lastly in Chapter 2 we proposed $L^2$ norm based test statistic for testing the fixed effect of a continuous covariate.

Here we propose a novel inferential procedure for testing that the mean response does not vary over actual time of visit $t$. Because visit time $t$ is closely intertwined with serial correlation of repeated responses $Y_{ij}(\cdot)$, a testing problem we consider in this chapter is quite different from one studied in Chapter 2, in which covariates of interest were either subject-specific or time-independent. The main idea of the testing procedure we propose in this chapter is: (i) model the mean function using orthonormal basis functions and (ii) represent the null hypothesis in a simpler way using the basis coefficients and reduce the original null hypothesis into a set of simpler null hypotheses that can be assessed with common existing techniques. Numerical investigation shows that the proposed test has excellent size and power performance and is computationally efficient.

The rest of the chapter is organized as follows. In Section 4.2 we formally define the null hypothesis that we aim to test and discuss its equivalent formulation with associated challenges. Section 4.3 provides details on statistical modeling framework and introduces a likelihood ratio based testing procedure. Section 4.3.1 discusses practical implementation of the proposed test. In Section 4.4 we investigate its finite sample properties via simulation study. Section 4.5 illustrates the proposed testing procedure using the longitudinal DTI study.
4.2 Preliminaries

We first introduce the notation. Let $i$ index for the subject, $j$ index for the repeated measure (or visit), and $n$ be the total number of subjects. Denote the observed data for subject $i$ by \( \{t_{ij}, (Y_{ijr}, s_r)_{r=1}^{R} : j = 1, \ldots, m_i \} \), where $t_{ij}$ is the actual time of $j$th visit for subject $i$ and $Y_{ijr} = Y_{ij}(s_r)$. We refer to $Y_{ij}(\cdot)$ as the functional response observed at time $t_{ij}$. We assume that $Y_{ij}(\cdot)$ is observed at regular grid of points, \( \{s_r : r = 1, \ldots, R \} \in S \), for sufficiently large $R$. It is assumed that for each subject $i$ number of repeated measures, denoted by $m_i$, is small and the set of pooled visit times for all subjects, \( \{t_{ij} : i = 1, \ldots, n, j = 1, \ldots, m_i \} \), is dense in $T$. The domains, $S$ and $T$, can be any compact intervals; without loss of generality, we use $S = T = [0, 1]$.

Recall the model $Y_{ijr} = \mu(s_r, t_{ij}) + \mathcal{E}_i(s_r, t_{ij})$, where $\mu(\cdot, t_{ij})$ is an unknown smooth mean function evaluated at $t_{ij}$ and $\mathcal{E}_i(\cdot, t_{ij})$ is the random deviation from the mean function. The null hypothesis we are interested to test is

$$H_0 : \mu(s, t) = \mu_0(s) \text{ for all } s \text{ and } t,$$

for some unknown smooth univariate function $\mu_0(\cdot)$ against the alternative $H_A : \mu(s, t)$ varies over $t$ for some $s$. Common types of null hypotheses about the mean function that are considered in literature include: the mean function is equal to zero (e.g. Eubank & Spiegelman (1990), Guo (2002), Antoniadis & Sapatinas (2007), Abramovich & Angelini (2006)); or, is equal to some pre-specified mean function (Zhang & Chen (2007)); or, has some specified parametric structure such as polynomial (e.g. Hardle & Mammen (1993), Crainiceanu & Ruppert (2004), Scheipl et al. (2008a), Wang & Chen (2012), Staicu et al. (2014a)); or, is additive (e.g. Eubank et al. (1995)).

In contrast the null hypothesis we consider in this chapter is completely nonparametric
that a two dimensional function is a function of only one argument. Such null hypotheses
were extensively studied in the context of nonparametric regression; see for example Fan
& Li (1996), Lavergne & Vuong (2000), Delgado & Manteiga (2001), Gu et al. (2007),
and Hall et al. (2007). However, all these methods are based on the assumption that
observations are independent across sampling units, i.e. $Y_{ijr}$ are independent over $i$, $j$
and $r$. To the best of our knowledge, other than Chapter 2 of this dissertation, such type
of hypothesis has not been considered in functional data analysis, where the independence
assumption is no longer appropriate.

Our testing procedure relies on the following observation. Consider a set of orthonor-
mal basis functions, say $\{\theta_k(s) : k \geq 1\}$. Suppose we use the orthonormal basis functions
to represent the mean function as $\mu(s, t) = \sum_{k \geq 1} \eta_k(t)\theta_k(s)$, where $\eta_k(t) = \int \mu(s, t)\theta_k(s)ds$.
Now notice that the null hypothesis (4.1) is equivalent to $\partial \mu(s, t)/\partial t = 0$ for all $s \in S$ and
t $\in T$, and furthermore $\int \{\partial \mu(s, t)/\partial t\}^2ds = 0$ for all $t \in T$. However, $\int \{\partial \mu(s, t)/\partial t\}^2ds = \int \sum_{k, \ell \geq 1} \{\eta'_k(t)\eta'_\ell(t)\}\theta_k(s)\theta_\ell(s)ds = \sum_{k \geq 1} \{\eta'_k(t)\}^2$, where $\eta'_k(t)$ is the first derivative of
$\eta_k(t)$; the last equality holds because of orthonormality of $\theta_k(s)$, i.e. $\int \theta_k(s)\theta_\ell(s)ds = 1(k = \ell)$. Then it follows that the null hypothesis (4.1) is equivalent to $\eta'_k(t) = 0$, or
$\eta_k(t) = C_k$ for some real-valued constants $C_k$ for all $k \geq 1$. Thus based on this ob-
servation we re-formulate our null hypothesis (4.1) as a set of simpler null hypotheses,
$\{H_{0,k} : \eta_k(t) = C_k, k \geq 1\}$ for some constants $C_k$’s.

The new formulation achieves some dimension reduction as it only involves univariate
functions of $t$, instead of a bivariate function of $s$ and $t$. However it still brings along a
few challenges: (i) the new formulation still has infinite number of hypothesis testings
($k \geq 1$) and requires some truncation for number of $H_{0,k}$’s; (ii) it is not obvious what
orthonormal basis functions, $\theta_k(\cdot)$’s, to use; (iii) it is not trivial to test each of $H_{0,k}$’s,
despite its simpler form; (iv) the type I error of the testing procedure becomes the family-
wise error rate that needs to be controlled appropriately. We address these challenges in the following section.

4.3 Testing Procedure

In this section we describe the full model assumptions and introduce a likelihood ratio based hypothesis testing procedure.

We assume that the random deviation \( E_i(\cdot, t_{ij}) \) consists of two independent components: a subject-specific random deviation from the population mean, denoted by \( U_i(\cdot, t_{ij}) \), and a visit-specific random deviation from the subject mean, denoted by \( \epsilon_{ij}(\cdot) \), that is also referred to as the residual process. Specifically we model the observed data as

\[
Y_{ijr} = \mu(s_r, t_{ij}) + U_i(s_r, t_{ij}) + \epsilon_{ij}(s_r)
\]

for \( i = 1, \ldots, n, j = 1, \ldots, m_i, \) and \( r = 1, \ldots, R \), where \( U_i's \) are iid square integrable random elements in \( L^2(S \times T) \) with mean zero and smooth unknown covariance

\[
\gamma((s, t), (s', t')) = E\{U_i(s, t)U_i(s', t')\},
\]

and \( \epsilon_{ij}'s \) are iid with zero mean and are independent of \( U_i's \). It is further assumed that the residual process \( \epsilon_{ij}(s) \) is the sum of two independent processes, \( \epsilon_{1,ij}(s) \) and \( \epsilon_{2,ij}(s) \), where \( \epsilon_{1,ij}'s \) are iid square integrable random elements in \( L^2(S) \) with mean zero and smooth unknown covariance \( \Gamma(s, s') = E\{\epsilon_{1,ij}(s)\epsilon_{1,ij}(s')\} \) and \( \epsilon_{2,ij}'s \) are white noise processes with mean zero and variance \( \sigma^2 \).

Suppose that we use a finite set of orthonormal basis functions, \( \{\theta_k(s) : k = 1, \ldots, K\} \), defined on \( S \), where \( K \) is much smaller than \( R \). Let \( W_{ijk} = \int Y_{ij}(s) \theta_k(s) ds \) be the projection of the observed data onto the space spanned by \( \theta_k(\cdot) \); in practice the integral can be approximated by the numerical integration, \( W_{ijk} \approx R^{-1} \sum_{r=1}^{R} Y_{ij}(s_r) \theta_k(s_r) \). As in Section 4.2 let \( \eta_k(t_{ij}) = \int \mu(s, t_{ij}) \theta_k(s) ds \); notice that \( \eta_k(t_{ij}) = \int E\{Y_{ij}(s)\} \theta_k(s) ds = E(W_{ijk}) \). Let \( w_{ik}(t_{ij}) = \int U_i(s, t_{ij}) \theta_k(s) ds \) be the projection of the subject specific random
process \( U_i(\cdot, t_{ij}) \) onto \( \theta_k(\cdot) \); it follows that \( \omega_{ik}(\cdot) \) is a zero-mean square integrable function in \( L^2(\mathcal{T}) \) with unknown smooth covariance, which we denote by \( \Pi_k(t, t') \). Lastly let \( e_{ijk} = \int \epsilon_{ij}(s)\theta_k(s)ds \) be the projection of the residual process \( \epsilon_{ij}(\cdot) \) onto \( \theta_k(\cdot) \); hence \( e_{ijk} \)'s are iid with zero mean and constant variance, denoted by \( \sigma_{e,k}^2 \), and are independent of \( \omega_{ik} \)'s. Note that the term \( e_{ijk} \) can also be written as \( e_{ijk} = e_{1,ijk} + e_{2,ijk} \), where \( e_{1,ijk} = \int \epsilon_{1,ij}(s)\theta_k(s)ds \), and \( e_{2,ijk} = \int \epsilon_{2,ij}(s)\theta_k(s)ds \). Theoretically \( e_{2,ijk} = 0 \) with probability 1, but in practice we assume that \( e_{2,ijk} \) is iid with zero mean and small variance.

It follows that for \( k = 1, \ldots, K \), the projected data \( \{(t_{ij}, W_{ijk}) : i = 1, \ldots, n \text{ and } j = 1, \ldots, m_i\} \) are modeled as

\[
W_{ijk} = \eta_k(t_{ij}) + \omega_{ik}(t_{ij}) + e_{ijk}.
\]

(4.2)

This is a common modeling framework in functional data analysis, where response \( W_{ijk} \) is viewed as a noisy independent realization of the true underlying random process, \( \eta_k(t) + \omega_{ik}(t_{ij}) \), that are sparsely sampled from its domain \( \mathcal{T} \); see Yao et al. (2005a).

A sketch of our proposed testing procedure is as follows. Based on the preliminary observation in Section 4.2, testing the null hypothesis \( H_0 \) in (4.1) can be reduced to testing simultaneously a set of null hypotheses, \( \{H_{0,k} : \eta_k(t) = C_k, \; k = 1, \ldots, K\} \) for some constants \( C_k \)'s. We propose to test each null hypothesis \( H_{0,k} \) separately in model (4.2), based on a significance level \( \alpha \) adjusted by the Bonferroni correction. Specifically we reject our original null hypothesis \( H_0 \) in (4.1) if \( \min\{p_k : k = 1, \ldots, K\} < \alpha/K \), where \( p_k \) is p-value obtained by testing the corresponding null hypothesis \( H_{0,k} \).

The null hypothesis \( H_{0,k} \) has been already considered before; see Crainiceanu & Ruppert (2004), Crainiceanu et al. (2005), Greven et al. (2012), Greven & Crainiceanu (2013), Wiencierz et al. (2011), Staicu et al. (2014a). The main idea of their approaches is to
represent $\eta_k(t)$ using penalized splines and test appropriate fixed coefficients and/or variance components in the corresponding mixed model representation. The seminal work on testing a variance component in a mixed model is Crainiceanu & Ruppert (2004), where they derived the finite sample and asymptotic distributions of the likelihood ratio test (LRT); however, their test cannot be directly applied to our problem because their results rely on the independence assumption, i.e. $w_{ik}(t_{ij})$ are independent over $i$ and $j$.

Both Wiencierz et al. (2011) and Staicu et al. (2014a) extended to the case of having a non-trivial covariance structure of $w_{ik}(t)$ and proposed to replace the true covariance by its appropriate estimator. However, Wiencierz et al. (2011) assume that the covariance structure is known with unknown parameters, whereas Staicu et al. (2014a) assume that the covariance structure is unknown. In our testing procedure, we follow Staicu et al. (2014a) and represent the random process $w_{ik}(t)$ using the truncated FPCA model, i.e.

$$w_{ik}(t) = \sum_{\ell=1}^{L_k} \xi_{k\ell} \psi_{k\ell}(t),$$

and test each $H_{0,k}$ by using the pseudo LRT (pLRT) proposed by Staicu et al. (2014a). Here $\psi_{k\ell}(t)$ is the $\ell$th eigenfunction of the true covariance $\Pi_k(t,t')$, $\xi_{k\ell}$ is the corresponding basis coefficient, and $L_k$ is a truncation value.

We conclude this section with providing several possible choices of orthonormal basis functions, $\theta_k$’s, and selection of the truncation value, $K$. One possible option for $\theta_k$ is to use pre-specified basis functions; for example, we can partition the domain $\mathcal{S}$ into $K$ equally spaced intervals and use the indicator function of the $k$th partitioned interval as $\theta_k$; or, we can use $K$ dimensional spline basis, such as Fourier basis. Another possible option is to use some data-driven orthonormal bases, such as eigenfunctions of some covariance function; here the challenge is to determine which covariance function to use.

In this chapter we take the latter option as it leads to parsimonious representation by identifying dominant modes of variation in data. We propose to use the marginal covariance function defined in Chapter 3 (Park & Staicu, 2015): $\Xi(s,s') = \int c\{(s,t),(s',t)\}g(t)dt$, 73
where $g(t)$ is the sampling density function of visit times, $t$. Now let $\tau_k$ be $k$th eigenvalue from the spectral decomposition of $\Xi(s, s')$ and let $\phi_k(s)$ be the corresponding eigenfunction. We use $\theta_k = \phi_k$ for $k = 1, \ldots, K$.

There are several different criteria for selecting $K$, such as cross-validation or AIC; see for example Yao et al. (2005a) and Li et al. (2013). Here we choose $K$ as the smallest integer that satisfies $\sum_{k=1}^{K} \tau_k / \sum_{k=1}^{\infty} \tau_k$ to be greater than the pre-specified percentage of variance explained (PVE).

### 4.3.1 Implementation

We provide a brief discussion on implementation for the proposed testing procedure, which we will refer to as *pseudo LRT with Multiple Testing* (pLRT-MT). To test the null hypothesis $H_0$ in (4.1) we first estimate the orthonormal basis functions, $\{\phi_k(s) : k = 1, \ldots, K\}$, by employing the estimation method proposed in Chapter 3. Recall that $K$ is selected based on pre-specified PVE. Once we obtain the estimated eigenfunctions, denoted by $\hat{\phi}_k(s)$, we follow Algorithm 4.

**Algorithm 4** Multiple pLRTs in the testing procedure

1. **for** $k \in \{1, \ldots, K\}$ **do**

2. Construct the projected data $\{(t_{ij}, \tilde{W}_{ijk}) : i = 1, \ldots, n \text{ and } j = 1, \ldots, m_i\}$ by calculating $\tilde{W}_{ijk} = R^{-1} \sum_{r=1}^{R} Y_{ij}(s_r) \hat{\phi}_k(s_r)$.

3. Test $H_{0,k} : \eta_k(t) = C_k$ for some constant $C_k$ by employing the pLRT (Staicu et al., 2014a) using $\tilde{W}_{ijk}$ in replacement of $W_{ijk}$.

4. Calculate p-value, $p_k$.

5. **end for**

6. Reject $H_0$ in (4.1) if $\min\{p_k : k = 1, \ldots, K\} < \alpha/K$ for a significance level $\alpha$. 

74
Because in practice responses $W_{ijk}$ in the projected data are latent and unobserved, we propose to replace $W_{ijk}$ with its estimator $\tilde{W}_{ijk}$ in Steps 1 and 2 of Algorithm 4. The pLRT can be implemented using the function `LRTsim` in the R package `RLRsim` (Scheipl et al., 2008b), which also provides p-value, $p_k$. The null distribution of the pLRT statistic is approximated using simulation; see Crainiceanu & Ruppert (2004) for details.

4.3.2 Pseudo Likelihood Ratio Test

The proposed testing procedure is developed based on a conjecture that the pLRT statistic obtained with the estimator $\tilde{W}_{ijk}$’s has the same asymptotic null distribution as the pLRT proposed in Staicu et al. (2014a), where they use the true $W_{ijk}$’s. This conjecture is suggested by the consistency result of the eigenfunctions, $\phi_k(s)$, from Chapter (3) and is strongly supported by numerical investigation (see Table C.1). However, further theoretical investigation is needed. The conjecture is formally stated in the following.

Following Staicu et al. (2014a) we use a truncated polynomial basis to represent $\eta_k(t)$ as $\eta_k(t) = \beta_{k0} + \beta_{k1}t + \ldots + \beta_{kp}t^p + \sum_{q=1}^{Q} b_{kq}(t - \kappa_q)^{p}_+$, where $(x)_+ = \max(0, x)$, $\beta_k = (\beta_{k0}, \ldots, \beta_{kp})^T$ is a vector of polynomial parameters, and $b_k = (b_{k1}, \ldots, b_{kQ})^T$ is a vector of spline coefficients. Here $\kappa_q$’s are knots placed at $Q$ equally spaced quantiles of the pooled visit times, $\{t_{ij} : \forall \ i \text{ and } j\}$. Number of knots $Q$ needs to be sufficiently large to capture the complexity of $\eta_k(\cdot)$ (Ruppert (2002) and Ruppert et al. (2003)); and smoothness is controlled by introducing a quadratic penalty term with the smoothing parameter, $\theta_k$. This can be written as mixed model representation; specifically, $\eta_k(t_{ij}) = X_{ij}\beta_k + Z_{ij}b_k$, where $X_{ij} = (1, t_{ij}, \ldots, t_{ij}^p)$, $Z_{ij} = \{(t_{ij} - \kappa_1)^{p}_+, \ldots, (t_{ij} - \kappa_Q)^{p}_+\}$, and the coefficients, $b_{kq}$’s, are assumed to be iid normal with mean zero and variance $\sigma_{b,k}^2$, where $\sigma_{b,k}^2 = \theta_k \sigma_{e,k}^2$.

Let $X_i$ be the $m_i \times (p+1)$ dimensional matrix with the $j$th row equal to $X_{ij}$. Similarly
let $Z_i$ be the $m_i \times Q$ dimensional matrix with the $j$th row equal to $Z_{ij}$. Let $\omega_{ijk} = w_{ik}(t_{ij}) + e_{ijk}$ be the random deviation from the mean in (4.2) and let $e_{ik} = (\omega_{i1k}, \ldots, \omega_{im,k})^T$. Then we can re-write model (4.2) as

$$W_{ik} = X_i\beta_k + Z_i b_k + e_{ik} \tag{4.3}$$

where $\beta_k$ is a $p$ dimensional vector of fixed effects, $b_k$ is a $Q$ dimensional vector of random effects, and $e_{ik}$ is a $m_i$ dimensional vector of errors. It is assumed that $b_k$ is distributed as multivariate Normal with mean zero and covariance $\sigma^2_{b,k} I$ and $e_k$ is distributed as multivariate Normal with mean zero and covariance $\Sigma_{W,ik}$ and is independent of $b_k$.

Note that $\Sigma_{W,ik}$ is the covariance of $\{w_{ik}(t_{ij}) + e_{ijk} : j = 1, \ldots, m_i\}$, which captures the within-subject variability. Specifically the $(j, j')$ element of $\Sigma_{W,ik}$ is equal to $\Pi_k(t_{ij}, t_{ij'}) + \sigma^2_{k,e} 1(j = j')$, where $1(\cdot)$ is the indicator function.

Let $N = \sum_{i=1}^n m_i$ be the total number of observed curves. Let $W_k$ and $e_k$ be the $N$ dimensional, stacked vectors of $W_{ik}$’s and $e_{ik}$’s, respectively. Let $X$ be the stacked matrix of $X_i$’s with dimension $N \times (p + 1)$ and let $Z$ be the stacked matrix of $Z_i$’s with dimension $N \times Q$. Let $\Sigma_{W,k}$ be $N \times N$ block diagonal matrix with the $i$th block equal to $\Sigma_{W,ik}$. When $W_k$ is observed and $\Sigma_{W,k}$ is known, twice the log-likelihood of $W_k$ is equal to

$$2 \log L_{W_k}(\beta_k, \sigma^2_{b,k}) = - \log(|\Sigma_{W,k} + \sigma^2_{b,k} ZZ^T|) - (W_k - X\beta_k)^T(\Sigma_{W,k} + \sigma^2_{b,k} ZZ^T)^{-1}(W_k - X\beta_k),$$

up to an additive constant that does not depend on the parameters; here $|\cdot|$ is the determinant of a square matrix. Then the LRT statistic is

$$LRT_{N,k} = \sup_{H_{0,k} \cup H_{A,k}} 2 \log L_{W_k}(\beta_k, \sigma^2_{b,k}) - \sup_{H_{0,k}} 2 \log L_{W_k}(\beta_k, \sigma^2_{b,k}).$$

In our problem setting, $W_k$ is unobserved and $\Sigma_{W,k}$ is unknown. We propose to use a new pseudo LRT statistic $pLRT_{N,k}$ obtained by replacing $W_{ik}$ in LRT with its estimator $\hat{W}_{ik}$ and $\Sigma_{W,ik}$ in LRT with its estimator $\hat{\Sigma}_{W,ik}$ obtained using $\hat{W}_{ijk}$’s; here $\hat{W}_{ik} = \ldots$
\((\hat{W}_{11k}, \ldots, \hat{W}_{im,k})^T\) is an estimator of \(W_{ik}\), where \(\hat{W}_{ijk} = \int Y_{ij}(s)\hat{\phi}_k(s)ds\), and \(\hat{\Sigma}_{W,k}\) an estimator of \(\Sigma_{W,k}\) obtained using \(\hat{W}_{ijk}\)'s. For a positive definite matrix \(A\), we denote by \(A^{-1/2}\) a square root of \(A^{-1}\) such that \(A^{-1} = A^{-1/2}A^{-1/2}\). Let \(\hat{V}_k = \hat{\Sigma}_{W,k}^{-1/2}\hat{W}_k\), \(\hat{X}_k = \hat{\Sigma}_{W,k}^{-1/2}X\), \(\hat{Z}_k = \hat{\Sigma}_{W,k}^{-1/2}Z\) and \(\hat{u}_k = \hat{\Sigma}_{W,k}^{-1/2}e_k\). Then twice the pseudo log-likelihood is, up to an additive constant independent of the parameters, \(2\log L_{\hat{V}_k}(\beta_k, \sigma^2_{b,k}) = -\log(|\hat{H}_{\sigma^2_{b,k}}|) - (\hat{V}_k - \hat{X}_k\beta_k)^T\hat{H}_{\sigma^2_{b,k}}^{-1}(\hat{V}_k - \hat{X}_k\beta_k)\), where \(\hat{H}_{\sigma^2_{b,k}} = \hat{\Sigma}_{W,k}^{-1} + \sigma^2_{b,k}\hat{Z}_k\hat{Z}_k^T\). And the pseudo LRT statistic for testing \(H'_{0,k}\) is

\[
\text{pLRT}_{N,k} = \sup_{H'_{0,k}} 2\log L_{\hat{V}_k}(\beta_k, \sigma^2_{b,k}) - \sup_{H_{0,k}} 2\log L_{\hat{V}_k}(\beta_k, \sigma^2_{b,k}).
\] (4.4)

Then the following conjecture is suggested by the theoretical results of Chapter 3 and Staicu et al. (2014a).

**Conjecture 1.** Assume that \(b\) and \(u_k\) are jointly Gaussian. Also assume that the following hold:

(A1) The minimum eigenvalue of \(\Sigma_{W,k}\) is bounded away from zero as \(n\) diverges. Let \(\hat{\Sigma}_{W,k}\) be an estimator of \(\Sigma_{W,k}\) satisfying \(a^T\hat{\Sigma}_{W,k}^{-1}a - a^T\Sigma_{W,k}^{-1}a = o_p(1)\), and \(a^T\hat{\Sigma}_{W,k}^{-1}e_k - a^T\Sigma_{W,k}^{-1}e_k = o_p(1)\), where \(a\) is any non random normalized vector of length \(N\).

(A2) There exists positive constants, \(\varsigma\) and \(\varsigma'\), such that \(N^{-\varsigma}Z^TZ\) and \(N^{-\varsigma'}X^TX\) converge to non-zero matrices. For every eigenvalues \(\xi_{N,kq}\) and \(\zeta_{N,kq}\) of the matrices \(N^{-\varsigma}Z^T\Sigma_{W,k}^{-1}Z\) and \(N^{-\varsigma'}\{Z^T\Sigma_{W,k}^{-1}Z - Z^T\Sigma_{W,k}^{-1}X(X^T\Sigma_{W,k}^{-1}X)^{-1}X^T\Sigma_{W,k}^{-1}Z\}\) respectively, \(\xi_{N,kq} \overset{p}{\to} \xi_{kq}\) and \(\zeta_{N,kq} \overset{p}{\to} \zeta_{kq}\) for some \(\xi_{k1}, \ldots, \xi_{kQ}, \zeta_{k1}, \ldots, \zeta_{kQ}\) that are not all zero.

Then under the null hypothesis \(H'_{0,k}\) and the regularity conditions in Chapter 3, we have

\[
\text{pLRT}_{N,k} \overset{d}{\to} \sup_{\lambda \geq 0} \text{LRT}_{\infty}(\lambda) + \nu^2
\] (4.5)

where \(\text{LRT}_{\infty}(\lambda) = \sum_{q=1}^{Q} \frac{\lambda}{1 + \lambda \zeta_{kq}} \vartheta^2_{kq} - \sum_{q=1}^{Q} \log(1 + \lambda \xi_{kq})\), with \(\vartheta_{kq} \sim N(0, \zeta_{kq})\) for \(q = 1, \ldots, Q\), \(\nu \sim N(0, 1)\), and \(\vartheta_{kq}'s\) and \(\nu\) are mutually independent.
The right hand side of (4.5) is equal to the asymptotic distribution of \( LRT_{N,k} \), where \( LRT_{N,k} \) is the LRT statistic obtained using response \( W_k \) and true covariance \( \Sigma_{W,k} \) (Crainiceanu & Ruppert, 2004). The asymptotic null distribution is not of a standard form and needs to be approximated; for the approximation, Crainiceanu & Ruppert (2004) provided a simulation-based algorithm that is computationally fast. The algorithm is implemented in the \texttt{RLRsim} package (Scheipl et al., 2008b).

### 4.4 Simulation Study

We conducted a simulation study to evaluate the finite sample performance of the proposed test with respect to type I error, power, and computational time.

#### 4.4.1 Data Generation

Each sample is generated from the following model: 
\[
Y_{ij}(s) = \mu(s, t_{ij}) + w_{i,1}(t_{ij})\phi_1(s) + w_{i,2}(t_{ij})\phi_2(s) + E_{ij}(s) \quad \text{with} \quad \phi_1(s) = \sqrt{2}\sin(2\pi s) \quad \text{and} \quad \phi_2(s) = \sqrt{2}\cos(2\pi s).
\]

Each profile \( Y_{ij}(s) \) is observed at \( R = 101 \) equally spaced sampling points, \( s_r = (r - 1)/R \) for \( r = 1, \ldots, R \). And for each \( i \) visit times \( \{t_{ij} : j = 1, \ldots, m_i\} \) are randomly sampled from 41 equally spaced points in \([0, 1]\). We consider a different number of subjects: (i) \( n = 100 \), (ii) \( n = 200 \), (iii) \( n = 300 \), and (iv) \( n = 400 \); and a different level of sparsity: (i) extreme case, where \( m_i \) is independently sampled from \( \{8, \ldots, 12\} \) and (ii) moderate case, where \( m_i \) is independently sampled from \( \{15, \ldots, 20\} \). The mean function is \( \mu(s, t) = \cos(\pi s/2) + 5\delta(t/4 - s)^3 \). If \( \delta = 0 \), \( \mu(s, t) = \cos(\pi s/2) \) does not vary with \( t \); whereas if \( |\delta| > 0 \), \( \mu(s, t) \) varies with \( t \). We use \( \delta = 0 \) for studying Type I error and \( \delta = 0.2, 0.4, 0.6, 0.8 \), and 1 for studying power.

Subject-specific time-varying coefficients \( w_{i,k}(t) \)'s are independently sampled from
Gaussian process with mean zero and smooth covariance function, \( \text{cov}\{w_{i,k}(t), w_{i,k}(t')\} = \Pi_k(t, t') \). We consider a nonparametric covariance (NP) where \( w_{i,k}(t) = \zeta_{i,k1}\sqrt{2}\sin(2\pi t) + \zeta_{i,k2}\sqrt{2}\cos(2\pi t) \), where the random coefficients are sampled independently from normal with mean zero and respective variances: \( \zeta_{i,11} \sim \text{iid } N(0, 4) \), \( \zeta_{i,12} \sim \text{iid } N(0, 2) \), \( \zeta_{i,21} \sim \text{iid } N(0, 3) \), \( \zeta_{i,22} \sim \text{iid } N(0, 1) \). Errors \( E_{ij}(s) \) are generated from \( E_{ij}(s) = e_{ij1}\phi_1(s) + e_{ij2}\phi_2(s) + \epsilon_{ij}(s) \), where \( e_{ij1}, e_{ij2}, \) and \( \epsilon_{ij}(s) \) are mutually independent with mean zero and variances equal to \( \sigma_{e,1}^2 \), \( \sigma_{e,2}^2 \), and \( \sigma^2 \), respectively. The variances are determined based on the signal-to-noise ratio (SNR); specifically, \( \sigma^2 \) is set such that \( \int \text{var}\{w_{i,1}(t) + w_{i,2}(t)\}dt/\sigma^2 \) is equal to 1, and \( \sigma_{e,k}^2 \) is set such that \( \int \text{var}\{w_{i,k}(t)\}dt/\sigma_{e,k}^2 \) is equal to 3 for \( k = 1, 2 \).

### 4.4.2 Details on Estimation

The proposed procedure, pLRT-MT, is implemented as follows. We first estimate the eigenfunctions \( \phi_k(s) \) by following the estimation methods given in Chapter 3, and then test the null hypothesis \( H_0 \) by carrying out Algorithm 4. Specifically we obtain the estimated eigenfunctions \( \hat{\phi}_k(s) \) by employing the FPCA method on the demeaned data \( \{Y_{ij}(s) - \hat{\mu}(s, t_{ij}) : \forall i \text{ and } j\} \). Here the estimated mean and marginal covariance functions, \( \hat{\mu}(s, t) \) and \( \hat{\Xi}(s, s') \), are obtained using the sandwich smoother (Xiao et al., 2013), implemented in the \texttt{fbps} function of the \texttt{refund} package (Huang et al., 2016).

To carry out the pLRT in Step 3 of Algorithm 4, we model the mean function \( \eta_k(t) \) using linear truncated power splines with \( Q \) knots placed at equally spaced quantiles of the observed sampling points \( \{t_{ij} : \forall i \text{ and } j\} \). We choose the number of knots \( Q = \max\{20, \min(0.25 \times \text{number of unique } t_{ij}, 40)\} \), following suggestion in Ruppert et al. (2003). In addition, the covariance function \( \Pi_k(s) \) is estimated by representing \( w_{ik}(t) \) using the truncated FPCA model, i.e. \( w_i(t) = \sum_{\ell=1}^{L_k} \xi_{k\ell}\psi_{k\ell}(t) \), where \( \psi_{k\ell}(t) \) is the \( \ell \)th eigenfunctions of \( \Pi_k(t, t') \) and \( \xi_{k\ell} \) is the corresponding basis coefficient. In our simulation
the \texttt{fpca.sc} function of the \texttt{refund} is used to estimate the eigenfunctions $\psi_{k\ell}(t)$ and the variance components are estimated using restricted maximum likelihood (REML). Lastly, p-value $p_k$ obtained from the \texttt{exactLRT} function of the \texttt{refund}, based on $10^5$ replications.

Throughout the simulation study a smoothing parameter is selected based on generalized cross validation (GCV) and finite truncation values, $K$ and $L_k$, are selected such that at least 90% of variance is explained.

### 4.4.3 Competitive Methods

There is no competitive inferential methods in the literature that can be directly applied to longitudinal functional data and to test the null hypothesis (4.1). Thus we compare the proposed method with variants of available testing procedures. We consider two approaches. (A1.) testing the null hypothesis $H_0$ in (4.1) directly using a $L^2$ norm based test statistic $T_1 = \int \int \{\hat{\mu}_A(s,t) - \hat{\mu}_0(s)\}^2 ds dt$ proposed in Chapter 2, where $\hat{\mu}_0(s)$ and $\hat{\mu}_A(s,t)$ are smooth estimates of the mean functions under the null hypothesis $H_0$ and the alternative, respectively; and (A2.) following a general step of our proposed testing procedure but testing a set of the null hypotheses, $H_{0,k}$’s, using test statistic $T_{2,k} = \int \{\hat{\eta}_k(t) - \hat{C}_k\}^2 dt$, instead of using the pLRT in Step 3 of Algorithm 4. Here $\hat{C}_k$ and $\hat{\eta}_k(t)$ are smooth estimates of the mean functions under the null hypothesis $H_{0,k}$ and the alternative, respectively.

The null distribution of $T_1$ in (A1.) is approximated based on the bootstrap of subjects, hence we will refer this approach as \textit{Bootstrap}; see Chapter 2 for more details. For the approach (A2.), we consider two approximation methods for the null distribution of $T_{2,k}$: (i) the bootstrap of subjects and (ii) $\chi^2$ approximation, inspired from Zhang & Chen (2007). As the approach (A2.) involves multiple testing we refer the first approximation method as \textit{Bootstrap with Multiple Testing (Bootstrap-MT)}; similarly the second one is
referred to as Zhang & Chen with Multiple Testing (ZC-MT). In the ZC-MT, the null distribution is approximated based on $10^5$ replications from the mixture of $\chi^2$ distributions, i.e. $\sum_{\ell=1}^{L_k} \hat{\gamma}_{k\ell} A_\ell$ with $A_\ell$'s being iid from the $\chi^2$ distribution with degrees of freedom equal to one. Here $\hat{\gamma}_{k\ell}$ and $L_k$ are obtained as in our testing procedure; see Section 4.4.2.

In our simulation, $\hat{\mu}_0(s)$ is obtained by fitting 10 cubic splines to the original data $Y_{ijr}$'s; and $\hat{\mu}_A(s,t)$ is obtained with 50 cubic splines from the tensor product of 10 univariate basis functions in $s$ and 5 in $t$. For the approach (A2.), $\hat{\eta}_k(t)$ is obtained by fitting 10 cubic basis functions to the projected data $\tilde{W}_{ijk}$'s, and $\hat{C}_k$ is given by the average of the projected responses, $\{\tilde{W}_{ijk} : \forall \ i \ and \ j\}$. The smoothing parameters are selected based on GCV.

### 4.4.4 Numerical Results

**Type I Error Performance.** Table 4.1 presents the type I error rates of the pLRT-MT corresponding to nominal levels $\alpha = 0.01, 0.05, 0.10,$ and $0.15$ for different sample sizes $n$ and number of profiles per subject $m_i$. When sample size is large ($n \geq 200$), the observed type I error rates of the pLRT-MT are within 2 standard errors of the nominal levels. However, when sample size is small ($n = 100$), the observed type I error rates of the pLRT-MT are inflated for both extreme and moderate levels of sparsity. When both true eigenfunctions $\phi_k(s)$ and covariance $\Pi_k(t, t')$ are used in the pLRT-MT procedure, its empirical sizes maintain its nominal levels even when the sample size $n$ and number of repeated measures per subject $m_i$ are small (see Table C.1). This suggests that the performance of the pLRT-MT heavily depends on the accuracy of both estimators, $\hat{\phi}_k(s)$ and $\hat{\Pi}_k(s, s')$. The empirical sizes of the ZC-MT and Bootstrap-MT methods are also presented in Table C.2 of Appendix C. While the ZC-MT method maintains the nominal levels with large number of repeated measures per subject $m_i$, it has poor size...
performance in the case of having small $m_i$ even when sample size is very large ($n = 400$). The Bootstrap-MT method maintains the nominal levels for all settings we considered, however it tends to be very conservative.

**Power Performance.** The empirical power functions of the pLRT-MT based on the significance level $\alpha = 0.05$ are shown in Figure 4.1 with solid lines. Because the pLRT-MT does not exhibit good size performance when $n = 100$, we only study its power for $n = 200, 300, \text{ and } 400$. Overall the pLRT-MT has excellent power properties. Both sample size $n$ and repeated measures per subject $m_i$ have a positive effect on the power performance of the pLRT-MT. Moreover, the results suggest that increasing the number of repeated measures $m_i$ improves the power performance more than increasing the sample size $n$. We compare the power performance of the pLRT-MT with the ZC-MT and Bootstrap-MT methods; their power functions are also shown in Figure 4.1 with dashed and dotted lines, respectively. Because the ZC-MT has poor size performance in the case of having small $m_i$, it is only compared when $m_i \sim \{15, \ldots, 20\}$. In comparison with the proposed method, the power of the ZC-MT and Bootstrap-MT methods are very weak and improve much slower than the ZC-MT with increasing $\delta$ and $n$.

In summary the proposed method maintains the nominal levels when $n \geq 200$ and has desirable power properties. Furthermore the proposed method offers much faster computation than the bootstrap-based tests; see Table 4.2. As the Bootstrap method involves fitting a bivariate smoother multiple times, it especially suffers from the computation burden and it was infeasible to study its power in our simulation study.
Table 4.1: The empirical type I error rates of the proposed method (pLRT-MT) based on 5000 simulations. Standard errors are presented in parentheses.

<table>
<thead>
<tr>
<th>$m_1 \sim {8, \ldots, 12}$</th>
<th>$\alpha = 0.01$</th>
<th>$\alpha = 0.05$</th>
<th>$\alpha = 0.10$</th>
<th>$\alpha = 0.15$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n = 100$</td>
<td>0.017 (0.002)</td>
<td>0.063 (0.003)</td>
<td>0.112 (0.004)</td>
<td>0.156 (0.005)</td>
</tr>
<tr>
<td>$n = 200$</td>
<td>0.012 (0.002)</td>
<td>0.052 (0.003)</td>
<td>0.103 (0.004)</td>
<td>0.151 (0.005)</td>
</tr>
<tr>
<td>$n = 300$</td>
<td>0.008 (0.001)</td>
<td>0.050 (0.003)</td>
<td>0.097 (0.004)</td>
<td>0.146 (0.005)</td>
</tr>
<tr>
<td>$n = 400$</td>
<td>0.013 (0.002)</td>
<td>0.055 (0.003)</td>
<td>0.098 (0.004)</td>
<td>0.142 (0.005)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$m_1 \sim {15, \ldots, 20}$</th>
<th>$\alpha = 0.01$</th>
<th>$\alpha = 0.05$</th>
<th>$\alpha = 0.10$</th>
<th>$\alpha = 0.15$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n = 100$</td>
<td>0.018 (0.002)</td>
<td>0.066 (0.004)</td>
<td>0.122 (0.005)</td>
<td>0.171 (0.005)</td>
</tr>
<tr>
<td>$n = 200$</td>
<td>0.012 (0.002)</td>
<td>0.053 (0.003)</td>
<td>0.106 (0.004)</td>
<td>0.156 (0.005)</td>
</tr>
<tr>
<td>$n = 300$</td>
<td>0.011 (0.001)</td>
<td>0.052 (0.003)</td>
<td>0.109 (0.004)</td>
<td>0.155 (0.005)</td>
</tr>
<tr>
<td>$n = 400$</td>
<td>0.012 (0.002)</td>
<td>0.056 (0.003)</td>
<td>0.102 (0.004)</td>
<td>0.153 (0.005)</td>
</tr>
</tbody>
</table>

Table 4.2: Computation time (in seconds) based on 200 simulations when $n = 200$.

<table>
<thead>
<tr>
<th>$m_i \sim {8, \ldots, 12}$</th>
<th>pLRT-MT</th>
<th>ZC-MT</th>
<th>Bootstrap-MT</th>
<th>Bootstrap</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>7.546</td>
<td>2.317</td>
<td>24.412</td>
<td>1126.880</td>
</tr>
<tr>
<td>$m_i \sim {15, \ldots, 20}$</td>
<td>8.010</td>
<td>2.376</td>
<td>30.081</td>
<td>1841.730</td>
</tr>
</tbody>
</table>

83
Figure 4.1: Estimated size and power curves for testing $H_0 : \mu(s, t) = \mu_0(s)$ using $\alpha = 0.05$, when the true mean function $\mu(s, t) = \cos(\pi s/2) + 5\delta(t/4 - s)^3$ for $\delta = 0, 0.2, 0.4, 0.6, 0.8, 1$. Results are based on 5000 simulations for size ($\delta = 0$) and 1000 simulations for power ($\delta > 0$).

### 4.5 Diffusion Tensor Imaging Study

Recall the diffusion tensor imaging (DTI) study, involving 162 multiple sclerosis (MS) patients from each of whom fractional anisotropy profiles along the corpus callosum tract (CCA-FA) were observed over many hospital visits. The aim of our previous analysis in Section 3.7 was to study the longitudinal dynamics of CCA-FA trajectories and to predict the subject-specific CCA-FA profile at a future visit. In this section, we study the mean trajectory of CCA-FA. Specifically we want to use the proposed testing procedure
(pLRT-MT) to formally assess if the mean CCA-FA trajectory changes over visit time.

The hospital visit time \( t_{ij} \) is defined as in Section 3.7; that is, for each subject we take the difference between the reported visit time and the subject’s baseline visit time such that \( t_{i1} = 0 \) for all subject \( i \), and then the resulting values are scaled such that \( t_{ij} \)’s for all \( i \) and \( j \) are in \([0, 1]\). We estimate the mean function using 50 basis functions from the tensor product of 10 univariate cubic B-splines in \( s \) and 5 in \( t \). The smoothing parameter is selected based on REML. Next, we demean the data and estimate the marginal covariance; we obtain \( K = 10 \) estimated eigenfunctions from the spectral decomposition of the estimated marginal covariance based on a pre-specified percentage of variance explained equal to 95%.

Based on our previous analysis in Section 3.7, for each \( k \) we model the longitudinal covariance function \( \Pi_k(t, t') = \text{cov}\{ w_{i,k}(t), w_{i,k}(t') \} \) using a random effects model \( w_{i,k}(t_{ij}) = b_{0ik} + b_{1ik}t_{ij} \), where \( \text{var}(b_{lik}) = \sigma_{l}^2 \) for \( l = 0, 1 \) and \( \text{cov}(b_{0ik}, b_{1ik}) = \sigma_{01k} \). Then we implement the proposed test by following Algorithm 4 given in Section 4.3.1. We obtain 10 p-values corresponding to 10 marginal eigenfunctions, where the minimum of the p-values, \( \min\{p_k\}_{k=1}^{10} \), is less than \( 10^{-5} \). The test result suggests that there is a strong evidence of the mean CCA-FA profile varying with visit time \( t \).

For comparison we also applied the Bootstrap and Bootstrap-MT methods described in Section 4.4.3; estimation is done exactly the same as in the simulation study of Section 4.4. In addition, we consider testing the mean function under the assumption of linearity in visit time \( t \), i.e. \( \mu(s, t) = \mu_0(s) + t\beta(s) \). We use a \( L^2 \) norm based test statistic, \( T_3 = \int \{\hat{\beta}(t)\}^2 dt \) and approximate the null distribution using the bootstrap, hence will be referred to as \textit{Bootstrap-L}. Here the estimated functions \( \hat{\mu}_0(s) \) and \( \hat{\beta}(s) \) are obtained using 10 cubic splines. For the bootstrap-based tests, the smoothing parameters are selected based on GCV to improve some computation cost. P-values obtained with the
Bootstrap and Bootstrap-L methods are $p = 0.34$ and $p = 0.227$, respectively. As the Bootstrap-MT method involves multiple testing, we obtain 10 p-values where the minimum of the p-values, $\min\{p_k\}_{k=1}^{10} = 0.25$ is greater than Bonferroni adjusted significance level, $\alpha_{adj} = 0.05/10$. The bootstrap-based test results yield to the conclusion that the mean CCA-FA profile is constant over visit time $t$, which is consistent with our preliminary analysis in Section 3.7 based on joint confidence band but contradicts the result we obtain using the proposed method.

We further investigate the performance of the proposed test and the alternative methods via simulation study which mimics the DTI data structure; data generation is described in Appendix C.2. The simulation results show that the Bootstrap-MT and Bootstrap-L methods maintain the nominal levels, whereas the Bootstrap and pLRT-MT methods, both of which requires bivariate smoothing, yield the inflated type I errors. This seems to be due to the heavily right-skewed sampling distribution of the visit times in the DTI data, negatively affecting estimation accuracy of the bivariate mean functions and the longitudinal covariance functions.
Chapter 5

Conclusion

Statistical modeling and inference methods are developed to address some fundamental scientific questions that commonly appear in longitudinal studies collecting functional data. The proposed methods are designed to be computationally feasible, readily applicable, and easy-to-interpret.

In Chapter 2 we develop bootstrap-based inferential methods to assess the population level effects of a scalar covariate on the functional response with highly correlated dependence structure. The key idea of the proposed methods is to use the bootstrap of independent sampling units (e.g. subject) to account all known sources of variability without explicitly modeling the complex error covariance, while allowing the mean structure to be flexible. Confidence bands and a $L^2$ norm-based test procedure are developed and their performance is thoroughly studied via simulation study. The numerical results confirm excellent coverage probability of the confidence intervals and size of tests. The proposed methods are applied to the Baltimore Longitudinal Study of Aging, where the objective is to formally assess the fixed effect of age on daily activity count profiles. Our proposed methods indicate that there is strong evidence that daily activity profiles in women vary with age. An R code implementing the proposed approaches is available at

In Chapter 3 we develop a novel parsimonious modeling framework for longitudinal functional responses. The main idea is to use time-invariant orthogonal basis functions and extract low-dimensional features of the data only through the basis coefficients. In this way the longitudinal dynamics of the underlying process can be easily assessed by studying the time-varying basis coefficients. Furthermore, it allows to make prediction of a full trajectory at unobserved visit time. The consistency results of the model components estimators and the predicted trajectories are obtained. Based on numerical investigation, the proposed method has excellent performance in the finite sample setting and outperforms the existing methods in terms of prediction accuracy as well as computation efficiency. We apply the proposed method to the DTI study and assess the dynamic behavior of the CCA-FA profiles in MS patients. The analysis results suggest that there is only small changes of the CCA-FA trajectories over the short duration of the DTI study. R code for illustrating the estimation procedure is available at http://www4.stat.ncsu.edu/~staicu/software/MLFD_Rcode.zip. A wrapper function fpca.lfda is available in the package refund (Huang et al., 2016), accompanied with interactive graphics in the refund.shiny package (Wrobel & Goldsmith, 2016).

In Chapter 4 we develop a pseudo likelihood ratio based testing procedure to formally assess the change of the mean function over time. The main idea is to use orthonormal basis functions to model the mean function and use the basis coefficients to re-formulate the null hypothesis into a set of simpler null hypotheses that can be tested using the pLRT developed by Staicu et al. (2014a). We propose to use the time-invariant orthonormal basis functions from Chapter 3 to achieve a parsimonious set of multiple testing. Family-wise error rate is controlled by Bonferroni correction. Based on the simulation results, the proposed test has excellent size and power performance for most cases, ex-
cept when sample size is small \( (n = 100) \). In comparison with the bootstrap-based test (Bootstrap-MT), the proposed method has much higher power and requires much less computing effort. We use the proposed method to the DTI data to formally assess if the FA mean trajectory varies over visit time. Though the proposed method indicates strong evidence of the mean function varying with visit time, our further investigation suggests that the proposed method is inappropriate for the DTI data with such small number of repeated measures per subject and unevenly distributed visit times. On the other hands, the bootstrap-based test under linearity assumption (Bootstrap-L) method indicates no evidence that the mean function varies with visit time and is validated via simulation study.
REFERENCES


Chen, K & Müller, HG (2012), ‘Modeling repeated functional observations,’ *Journal of

Ciprian Crainiceanu, Philip Reiss, Jeff Goldsmith, Lei Huang, Lan Huo & Fabian Scheipl (2014), refund: Regression with Functional Data, r package version 0.1-11.

Crainiceanu, Philip Reiss, Jeff Goldsmith, Lei Huang, Lan Huo & Fabian Scheipl (2013), refund: Regression with Functional Data, r package version 0.1-9.


Hastie, TJ & Tibshirani, RJ (1990), *Generalized additive models*, vol. 43, CRC Press.


Hyndman, RJ & Shang, H (2009), ‘Functional time series forecasting,’ *with discussion*), *Journal*.


Lavergne, P & Vuong, Q (2000), ‘Nonparametric significance testing,’ *Econometric Theory*, 16(04), pp. 576–601.


MATLAB (2014), version 8.4.0 (R2014b), The MathWorks Inc., Natick, Massachusetts.


Ruppert, D (2002), ‘Selecting the number of knots for penalized splines,’ *Journal of computational and graphical statistics*.


100


Wahba, G (1990), *Spline models for observational data*, vol. 59, Siam.


A.1 Validating the BLSA Testing Results via Simulation Study

We conducted a simulation study designed to closely mimic the BLSA data structure. Specifically, we generate data from model (2.2) with $\mu(s, x) = \cos(2\pi s) + \delta(\hat{\mu}(s, x) - \cos(2\pi s))$, where $\hat{\mu}(s, x)$ is the estimated mean log counts, $\delta$ is some parameter quantifying the distance from the null and alternative hypotheses, $\tau = 0$ (i.e. there is no additional covariate vector), and the errors $\epsilon_{ij}(s)$ are generated to have a covariance structure that mimics that of residuals from the BLSA data (Xiao et al., 2015b). Notice that when $\delta = 0$
the true mean profile $\mu(s,x) = \cos(2\pi s)$, whereas when $\delta = 1$ then $\mu(s,x) = \tilde{\mu}(s,x)$.

The covariate $X_i$ and the number of visits per subject, $m_i$, are generated uniformly from $\{30, \ldots, 90\}$ and $\{5, \ldots, 9\}$ respectively. We use $n = 378$, the number of female participants in the BLSA.

Table A.1 shows the rejection probabilities in 1000 simulations, when $\delta = 0$ and indicates that the empirical Type I error of the proposed testing procedure is close to the nominal level. Figure A.1 displays the rejection probabilities in 500 simulations, when $\delta > 0$. For all cases, we use $B = 300$ bootstrap samples to approximate the null distribution of the test statistic $T$.

Table A.1: Empirical type I error of the test statistic $T$ based on the $N_{sim} = 1000$ MC samples; Mean function is $\mu(s,x) = \cos(2\pi s)$, $\tau = 0$

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>0.05</th>
<th>0.10</th>
<th>0.15</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.06</td>
<td>0.11</td>
<td>0.16</td>
<td></td>
</tr>
<tr>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td></td>
</tr>
</tbody>
</table>

Standard errors are presented in parentheses.

Figure A.1: Estimated power curves for testing $H_0 : \mu(s,x) = \eta(s)$ using $\alpha = 0.05$, when the true mean function $\mu(s,x) = \cos(2\pi s) + \delta(\tilde{\mu}(s,x) - \cos(2\pi s))$ for $\delta = 0.01, 2, 4, 6, 8$. Results are based on $N_{sim} = 500$ MC samples.
A.2 Additional results for the performance of the confidence bands

Here we present additional results for the performance of the pointwise and joint confidence bands for different nominal coverages. Table A.2 shows results for nominal coverage of 85%, and Table A.3 for 95%. Both 85% and 90% pointwise confidence interval/bands perform fairly well while there are few cases where they fail to achieve their nominal coverages. For the joint confidence band, the 90% ones perform very well for all cases considered, whereas the 85% one fails to achieve the nominal coverage for some cases. In terms of average lengths, the confidence intervals/bands tend to be wider when there is stronger correlation.

Table A.2: Simulation results using bootstrap of subject-level residuals and 85% nominal level; results are based on 500 MC samples.

<table>
<thead>
<tr>
<th>Case</th>
<th>True Mean Function</th>
<th>Parameter</th>
<th>$\beta_0$</th>
<th>Bias</th>
<th>$\sqrt{\text{Var}}$</th>
<th>ACP$^{\text{point}}$</th>
<th>AL$^{\text{point}}$</th>
<th>ACP$^{\text{joint}}$</th>
<th>AL$^{\text{joint}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a)</td>
<td>$\beta_0 + \beta_s s + \beta_x X + \beta_S S + \tau Z$</td>
<td>$\beta_0 = 5$</td>
<td>0.20</td>
<td>0.00</td>
<td>0.20</td>
<td>0.83 (0.02)</td>
<td>0.40 (&lt; 0.01)</td>
<td>0.90</td>
<td>0.52 (&lt; 0.01)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\beta_s = 2$</td>
<td>0.20</td>
<td>0.01</td>
<td>0.40</td>
<td>0.83 (0.02)</td>
<td>0.78 (&lt; 0.01)</td>
<td>0.90</td>
<td>1.03 (&lt; 0.01)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\beta_x = 3$</td>
<td>0.20</td>
<td>0.05</td>
<td>0.40</td>
<td>0.83 (0.02)</td>
<td>0.78 (&lt; 0.01)</td>
<td>0.90</td>
<td>1.03 (&lt; 0.01)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\tau = 8$</td>
<td>0.20</td>
<td>0.05</td>
<td>0.40</td>
<td>0.83 (0.02)</td>
<td>0.78 (&lt; 0.01)</td>
<td>0.90</td>
<td>1.03 (&lt; 0.01)</td>
</tr>
<tr>
<td>(b)</td>
<td>$\beta_0 + \beta_s s + \beta_x X + \beta_S S + \tau Z$</td>
<td>$\beta_0 = 5$</td>
<td>0.20</td>
<td>-0.01</td>
<td>0.39</td>
<td>0.84 (0.02)</td>
<td>0.77 (&lt; 0.01)</td>
<td>0.90</td>
<td>0.52 (&lt; 0.01)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\beta_s = 2$</td>
<td>0.20</td>
<td>0.04</td>
<td>1.02</td>
<td>0.84 (0.02)</td>
<td>2.01 (&lt; 0.01)</td>
<td>0.90</td>
<td>1.54 (0.01)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\beta_x = 3$</td>
<td>0.20</td>
<td>0.02</td>
<td>0.67</td>
<td>0.84 (0.02)</td>
<td>1.33 (&lt; 0.01)</td>
<td>0.90</td>
<td>1.74 (0.01)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\tau = 8$</td>
<td>0.20</td>
<td>0.06</td>
<td>1.75</td>
<td>0.83 (0.02)</td>
<td>3.47 (0.01)</td>
<td>0.90</td>
<td>2.01 (0.01)</td>
</tr>
<tr>
<td>(c)</td>
<td>$\cos(2\pi s) + \beta_x X + \tau Z$</td>
<td>$f(s) = \cos(2\pi s)$</td>
<td>0.20</td>
<td>0.00</td>
<td>0.25</td>
<td>0.83 (0.01)</td>
<td>0.49 (&lt; 0.01)</td>
<td>0.90</td>
<td>1.00 (0.01)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\beta_x = 3$</td>
<td>0.20</td>
<td>0.00</td>
<td>0.50</td>
<td>0.84 (0.02)</td>
<td>0.64 (&lt; 0.01)</td>
<td>0.90</td>
<td>1.00 (0.01)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\tau = 8$</td>
<td>0.20</td>
<td>0.00</td>
<td>0.50</td>
<td>0.84 (0.02)</td>
<td>0.64 (&lt; 0.01)</td>
<td>0.90</td>
<td>1.00 (0.01)</td>
</tr>
<tr>
<td>(d)</td>
<td>$\cos(2\pi s) + 4((X/4) - s)^2 + \tau Z$</td>
<td>$\mu_4(s, X)$</td>
<td>0.20</td>
<td>0.00</td>
<td>0.61</td>
<td>0.84 (&lt; 0.01)</td>
<td>1.22 (0.01)</td>
<td>0.90</td>
<td>3.28 (0.01)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\tau = 8$</td>
<td>0.20</td>
<td>0.00</td>
<td>0.61</td>
<td>0.84 (&lt; 0.01)</td>
<td>1.22 (0.01)</td>
<td>0.90</td>
<td>3.28 (0.01)</td>
</tr>
</tbody>
</table>

Standard errors are presented in parentheses.
Table A.3: Simulation results using bootstrap of subject-level residuals and 90% nominal level; results are based on 500 MC samples.

<table>
<thead>
<tr>
<th>Case</th>
<th>True Mean Function</th>
<th>Parameter</th>
<th>$\rho$</th>
<th>Bias</th>
<th>$\sqrt{\text{var}}$</th>
<th>ACP$^{\text{point}}$</th>
<th>AL$^{\text{point}}$</th>
<th>ACP$^{\text{joint}}$</th>
<th>AL$^{\text{joint}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a)</td>
<td>$\beta_0 + \beta_s s + \beta_x X + \tau Z$</td>
<td>$\beta_0 = 5$</td>
<td>0.20</td>
<td>0.00</td>
<td>0.20</td>
<td>0.89 (0.01)</td>
<td>0.46 (&lt; 0.01)</td>
<td>0.00</td>
<td>0.20</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\beta_s = 2$</td>
<td>0.90</td>
<td>-0.01</td>
<td>0.27</td>
<td>0.89 (0.01)</td>
<td>0.59 (&lt; 0.01)</td>
<td>0.01</td>
<td>0.40</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\beta_s = 3$</td>
<td>0.90</td>
<td>0.01</td>
<td>0.52</td>
<td>0.89 (0.01)</td>
<td>1.17 (&lt; 0.01)</td>
<td>0.00</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>$\tau = 8$</td>
<td>0.90</td>
<td>0.00</td>
<td>0.05</td>
<td>0.89 (0.01)</td>
<td>0.12 (&lt; 0.01)</td>
<td>0.00</td>
<td>0.05</td>
<td>0.90 (0.01)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.90</td>
<td>0.00</td>
<td>0.05</td>
<td>0.89 (0.01)</td>
<td>0.12 (&lt; 0.01)</td>
<td>0.00</td>
<td>0.05</td>
<td>0.90 (0.01)</td>
</tr>
<tr>
<td>(b)</td>
<td>$\beta_0 + \beta_s s + \beta_x sX + \tau Z$</td>
<td>$\beta_0 = 5$</td>
<td>0.20</td>
<td>-0.01</td>
<td>0.39</td>
<td>0.89 (0.01)</td>
<td>0.88 (&lt; 0.01)</td>
<td>0.00</td>
<td>0.20</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\beta_s = 2$</td>
<td>0.90</td>
<td>-0.02</td>
<td>0.51</td>
<td>0.87 (0.01)</td>
<td>1.15 (0.01)</td>
<td>0.04</td>
<td>1.02</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\beta_s = 3$</td>
<td>0.20</td>
<td>0.03</td>
<td>0.78</td>
<td>0.89 (0.01)</td>
<td>1.75 (0.01)</td>
<td>0.02</td>
<td>0.67</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\beta_s = 7$</td>
<td>0.90</td>
<td>0.04</td>
<td>1.02</td>
<td>0.88 (0.01)</td>
<td>1.98 (0.01)</td>
<td>0.02</td>
<td>0.88</td>
</tr>
<tr>
<td></td>
<td>$\tau = 8$</td>
<td>0.20</td>
<td>-0.04</td>
<td>1.33</td>
<td>0.88 (0.01)</td>
<td>3.05 (0.01)</td>
<td>0.00</td>
<td>0.05</td>
<td>0.90 (0.01)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.00</td>
<td>0.05</td>
<td>1.75</td>
<td>0.87 (0.01)</td>
<td>3.96 (0.02)</td>
<td>0.00</td>
<td>0.12</td>
<td>0.87 (0.01)</td>
</tr>
<tr>
<td>(c)</td>
<td>$\cos(2\pi s) + \beta_s sX + \tau Z$</td>
<td>$f(s) = \cos(2\pi s)$</td>
<td>0.20</td>
<td>0.00</td>
<td>0.25</td>
<td>0.88 (0.01)</td>
<td>0.56 (&lt; 0.01)</td>
<td>0.00</td>
<td>0.25</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\beta_s = 2$</td>
<td>0.90</td>
<td>0.00</td>
<td>0.32</td>
<td>0.88 (0.01)</td>
<td>0.73 (&lt; 0.01)</td>
<td>0.00</td>
<td>0.32</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\beta_s = 3$</td>
<td>0.20</td>
<td>0.00</td>
<td>0.05</td>
<td>0.89 (0.01)</td>
<td>0.11 (&lt; 0.01)</td>
<td>0.00</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\tau = 8$</td>
<td>0.90</td>
<td>0.00</td>
<td>0.05</td>
<td>0.89 (0.01)</td>
<td>0.12 (&lt; 0.01)</td>
<td>0.00</td>
<td>0.05</td>
</tr>
<tr>
<td>(d)</td>
<td>$\cos(2\pi s) + 4((X/4) - s)^3 + \tau Z$</td>
<td>$\mu(s, X)$</td>
<td>0.20</td>
<td>0.00</td>
<td>0.01</td>
<td>0.89 (&lt; 0.01)</td>
<td>1.40 (0.01)</td>
<td>0.00</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\tau = 8$</td>
<td>0.90</td>
<td>0.00</td>
<td>0.81</td>
<td>0.88 (&lt; 0.01)</td>
<td>1.84 (0.02)</td>
<td>0.00</td>
<td>0.81</td>
</tr>
</tbody>
</table>

Standard errors are presented in parentheses.
A.3 Results obtained by bootstrapping observations by subject

Here we investigate the performance of pointwise and joint confidence interval/bands when they are obtained by bootstrapping subject-level observations (Algorithm 1) instead of bootstrapping subject-level residuals (Algorithm 2). Tables A.4 - A.6 present the relevant results for nominal coverages of 85%, 90% and 95% respectively.

First focus the case when the effect of covariate $X$ is linear on the mean response (cases F2 i. (a)-(c)). Consider the setting described in Section 2.6 of the main manuscript, and recall that for each subject a single scalar covariate of interest $X_i$ is observed. The simulation results confirm that bootstrapping subject-level observations instead of residuals has no effect on the performance of pointwise and joint confidence interval/bands, and overall they perform well. However, for the case of smooth effect of $X$, the result for the joint confidence band obtained by bootstrapping subject-level observations show substantial under-coverage. We suspect that it is because our covariate $X$ is subject-specific, i.e. $X_i$ is time-invariant; With $n = 100$ subjects in our MC samples, there are only 100 $X_i$-values to resample from. To confirm our speculation, we consider time-varying covariate (i.e. $X_{ij}$ instead of $X_i$). The results are presented in Table A.7, and the empirical coverages are similar to ones obtained with the residual-based bootstrap.
Table A.4: Simulation results using bootstrap of subject-level observations and 85% nominal level; results are based on 500 MC samples.

<table>
<thead>
<tr>
<th>Case</th>
<th>True Mean Function</th>
<th>Parameter</th>
<th>( \rho )</th>
<th>Bias</th>
<th>( \sqrt{\text{var}} )</th>
<th>ACP\textsuperscript{point}</th>
<th>AL\textsuperscript{point}</th>
<th>ACP\textsuperscript{joint}</th>
<th>AL\textsuperscript{joint}</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a)</td>
<td>( \beta_0 + \beta_s s + \beta_x X + \tau Z )</td>
<td>( \beta_0 = 5 )</td>
<td>0.20</td>
<td>0.00</td>
<td>0.20</td>
<td>0.83 (0.02)</td>
<td>0.40 (&lt; 0.01)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \beta_0 = 5 )</td>
<td>0.90</td>
<td>-0.01</td>
<td>0.27</td>
<td>0.84 (0.02)</td>
<td>0.52 (&lt; 0.01)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \beta_s = 2 )</td>
<td>0.20</td>
<td>0.01</td>
<td>0.40</td>
<td>0.83 (0.02)</td>
<td>0.78 (&lt; 0.01)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \beta_s = 2 )</td>
<td>0.90</td>
<td>0.01</td>
<td>0.52</td>
<td>0.83 (0.02)</td>
<td>1.03 (&lt; 0.01)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \beta_x = 3 )</td>
<td>0.20</td>
<td>0.00</td>
<td>0.05</td>
<td>0.85 (0.02)</td>
<td>0.10 (&lt; 0.01)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \beta_x = 3 )</td>
<td>0.90</td>
<td>0.00</td>
<td>0.05</td>
<td>0.84 (0.02)</td>
<td>0.10 (&lt; 0.01)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \tau = 8 )</td>
<td>0.20</td>
<td>0.00</td>
<td>0.05</td>
<td>0.83 (0.02)</td>
<td>0.10 (&lt; 0.01)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \tau = 8 )</td>
<td>0.90</td>
<td>0.00</td>
<td>0.05</td>
<td>0.83 (0.02)</td>
<td>0.11 (&lt; 0.01)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(b)</td>
<td>( \beta_0 + \beta_s s + \beta_x X + \beta_{sx} sX + \tau Z )</td>
<td>( \beta_0 = 5 )</td>
<td>0.20</td>
<td>-0.01</td>
<td>0.39</td>
<td>0.83 (0.02)</td>
<td>0.77 (&lt; 0.01)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \beta_0 = 5 )</td>
<td>0.90</td>
<td>-0.02</td>
<td>0.51</td>
<td>0.84 (0.02)</td>
<td>1.01 (0.01)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \beta_s = 2 )</td>
<td>0.20</td>
<td>0.03</td>
<td>0.78</td>
<td>0.83 (0.02)</td>
<td>1.54 (0.01)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \beta_s = 2 )</td>
<td>0.90</td>
<td>0.04</td>
<td>1.02</td>
<td>0.84 (0.02)</td>
<td>2.02 (0.01)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \beta_x = 3 )</td>
<td>0.20</td>
<td>0.02</td>
<td>0.67</td>
<td>0.84 (0.02)</td>
<td>1.33 (0.01)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \beta_x = 3 )</td>
<td>0.90</td>
<td>0.03</td>
<td>0.88</td>
<td>0.83 (0.02)</td>
<td>1.75 (0.01)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \beta_{sx} = 7 )</td>
<td>0.20</td>
<td>-0.04</td>
<td>1.34</td>
<td>0.84 (0.02)</td>
<td>2.66 (0.01)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \beta_{sx} = 7 )</td>
<td>0.90</td>
<td>-0.06</td>
<td>1.75</td>
<td>0.84 (0.02)</td>
<td>3.48 (0.02)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \tau = 8 )</td>
<td>0.20</td>
<td>0.00</td>
<td>0.05</td>
<td>0.83 (0.02)</td>
<td>0.10 (&lt; 0.01)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \tau = 8 )</td>
<td>0.90</td>
<td>0.00</td>
<td>0.05</td>
<td>0.83 (0.02)</td>
<td>0.11 (&lt; 0.01)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(c)</td>
<td>( \cos(2\pi s) + \beta_x X + \tau Z )</td>
<td>( f(s) = \cos(2\pi s) )</td>
<td>0.20</td>
<td>0.00</td>
<td>0.25</td>
<td>0.83 (0.01)</td>
<td>0.49 (&lt; 0.01)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>( f(s) = \cos(2\pi s) )</td>
<td>0.90</td>
<td>0.00</td>
<td>0.32</td>
<td>0.83 (0.01)</td>
<td>0.81 (0.02)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \beta_s = 3 )</td>
<td>0.20</td>
<td>0.00</td>
<td>0.05</td>
<td>0.85 (0.02)</td>
<td>0.10 (&lt; 0.01)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \beta_s = 3 )</td>
<td>0.90</td>
<td>0.00</td>
<td>0.05</td>
<td>0.84 (0.02)</td>
<td>0.10 (&lt; 0.01)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \tau = 8 )</td>
<td>0.20</td>
<td>0.00</td>
<td>0.05</td>
<td>0.83 (0.02)</td>
<td>0.10 (&lt; 0.01)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \tau = 8 )</td>
<td>0.90</td>
<td>0.00</td>
<td>0.05</td>
<td>0.83 (0.02)</td>
<td>0.11 (&lt; 0.01)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(d)</td>
<td>( \cos(2\pi s) + 4(4/4 - s)^3 + \tau Z )</td>
<td>( \mu_4(s,X) )</td>
<td>0.20</td>
<td>0.00</td>
<td>0.64</td>
<td>0.82 (&lt; 0.01)</td>
<td>1.25 (0.04)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \mu_4(s,X) )</td>
<td>0.90</td>
<td>0.00</td>
<td>0.84</td>
<td>0.82 (&lt; 0.01)</td>
<td>1.25 (0.04)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \tau = 8 )</td>
<td>0.20</td>
<td>0.00</td>
<td>0.05</td>
<td>0.85 (0.02)</td>
<td>0.11 (&lt; 0.01)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \tau = 8 )</td>
<td>0.90</td>
<td>0.00</td>
<td>0.05</td>
<td>0.83 (0.02)</td>
<td>0.11 (&lt; 0.01)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Standard errors are presented in parentheses.
Table A.5: Simulation results using bootstrap of subject-level observations and 90% nominal level; results are based on 500 MC samples.

<table>
<thead>
<tr>
<th>Case</th>
<th>True Mean Function</th>
<th>Parameter</th>
<th>( \rho )</th>
<th>Bias</th>
<th>( \sqrt{\text{var}} )</th>
<th>ACP\text{joint}</th>
<th>AL\text{joint}</th>
<th>ACP\text{joint}</th>
<th>AL\text{joint}</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a)</td>
<td>( \beta_0 + \beta_s s + \beta_x X + \tau Z )</td>
<td>( \beta_0 = 5 )</td>
<td>0.20</td>
<td>0.00</td>
<td>0.20</td>
<td>0.88 (0.01)</td>
<td>0.46 (&lt; 0.01)</td>
<td>0.40 (0.01)</td>
<td>0.50 (&lt; 0.01)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \beta_s = 2 )</td>
<td>0.20</td>
<td>0.01</td>
<td>0.27</td>
<td>0.88 (0.01)</td>
<td>0.50 (&lt; 0.01)</td>
<td>0.40 (0.01)</td>
<td>0.60 (&lt; 0.01)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \beta_x = 3 )</td>
<td>0.20</td>
<td>0.00</td>
<td>0.05</td>
<td>0.89 (0.01)</td>
<td>0.89 (&lt; 0.01)</td>
<td>0.59 (&lt; 0.01)</td>
<td>0.59 (&lt; 0.01)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \tau = 8 )</td>
<td>0.20</td>
<td>0.00</td>
<td>0.05</td>
<td>0.89 (0.01)</td>
<td>0.90 (&lt; 0.01)</td>
<td>0.46 (&lt; 0.01)</td>
<td>0.46 (&lt; 0.01)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \beta_{sx} = 7 )</td>
<td>0.20</td>
<td>-0.02</td>
<td>0.39</td>
<td>0.88 (0.01)</td>
<td>0.88 (&lt; 0.01)</td>
<td>0.88 (&lt; 0.01)</td>
<td>0.88 (&lt; 0.01)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \tau = 8 )</td>
<td>0.20</td>
<td>0.00</td>
<td>0.05</td>
<td>0.88 (0.01)</td>
<td>0.90 (&lt; 0.01)</td>
<td>0.90 (&lt; 0.01)</td>
<td>0.90 (&lt; 0.01)</td>
</tr>
<tr>
<td>(b)</td>
<td>( \beta_0 + \beta_s s + \beta_x X + \beta_{sx}sX + \tau Z )</td>
<td>( \beta_0 = 5 )</td>
<td>0.20</td>
<td>0.00</td>
<td>0.20</td>
<td>0.88 (0.01)</td>
<td>0.46 (&lt; 0.01)</td>
<td>0.40 (0.01)</td>
<td>0.50 (&lt; 0.01)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \beta_s = 2 )</td>
<td>0.20</td>
<td>0.01</td>
<td>0.40</td>
<td>0.89 (0.01)</td>
<td>0.89 (&lt; 0.01)</td>
<td>0.59 (&lt; 0.01)</td>
<td>0.59 (&lt; 0.01)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \beta_x = 3 )</td>
<td>0.20</td>
<td>0.00</td>
<td>0.05</td>
<td>0.89 (0.01)</td>
<td>0.89 (&lt; 0.01)</td>
<td>0.59 (&lt; 0.01)</td>
<td>0.59 (&lt; 0.01)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \tau = 8 )</td>
<td>0.20</td>
<td>0.00</td>
<td>0.05</td>
<td>0.89 (0.01)</td>
<td>0.90 (&lt; 0.01)</td>
<td>0.90 (&lt; 0.01)</td>
<td>0.90 (&lt; 0.01)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \beta_{sx} = 7 )</td>
<td>0.20</td>
<td>-0.02</td>
<td>0.39</td>
<td>0.88 (0.01)</td>
<td>0.88 (&lt; 0.01)</td>
<td>0.88 (&lt; 0.01)</td>
<td>0.88 (&lt; 0.01)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \tau = 8 )</td>
<td>0.20</td>
<td>0.00</td>
<td>0.05</td>
<td>0.88 (0.01)</td>
<td>0.90 (&lt; 0.01)</td>
<td>0.90 (&lt; 0.01)</td>
<td>0.90 (&lt; 0.01)</td>
</tr>
<tr>
<td>(c)</td>
<td>( \cos(2\pi s) + \beta_s s + \beta_x X + \tau Z )</td>
<td>( f(s) = \cos(2\pi s) )</td>
<td>0.20</td>
<td>0.00</td>
<td>0.25</td>
<td>0.88 (0.01)</td>
<td>0.56 (&lt; 0.01)</td>
<td>0.56 (&lt; 0.01)</td>
<td>0.56 (&lt; 0.01)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \beta_s = 2 )</td>
<td>0.20</td>
<td>0.00</td>
<td>0.32</td>
<td>0.88 (0.01)</td>
<td>0.73 (&lt; 0.01)</td>
<td>0.73 (&lt; 0.01)</td>
<td>0.73 (&lt; 0.01)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \beta_x = 3 )</td>
<td>0.20</td>
<td>0.00</td>
<td>0.05</td>
<td>0.89 (0.01)</td>
<td>0.12 (&lt; 0.01)</td>
<td>0.12 (&lt; 0.01)</td>
<td>0.12 (&lt; 0.01)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \tau = 8 )</td>
<td>0.20</td>
<td>0.00</td>
<td>0.05</td>
<td>0.88 (0.01)</td>
<td>0.12 (&lt; 0.01)</td>
<td>0.12 (&lt; 0.01)</td>
<td>0.12 (&lt; 0.01)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \beta_{sx} = 7 )</td>
<td>0.20</td>
<td>-0.02</td>
<td>0.39</td>
<td>0.88 (0.01)</td>
<td>0.88 (&lt; 0.01)</td>
<td>0.88 (&lt; 0.01)</td>
<td>0.88 (&lt; 0.01)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \tau = 8 )</td>
<td>0.20</td>
<td>0.00</td>
<td>0.05</td>
<td>0.88 (0.01)</td>
<td>0.90 (&lt; 0.01)</td>
<td>0.90 (&lt; 0.01)</td>
<td>0.90 (&lt; 0.01)</td>
</tr>
<tr>
<td>(d)</td>
<td>( \cos(2\pi s) + 4\left(\frac{X}{4} - s\right)^3 + \tau Z )</td>
<td>( \mu_4(s,X) )</td>
<td>0.20</td>
<td>0.00</td>
<td>0.64</td>
<td>0.87 (&lt; 0.01)</td>
<td>1.44 (0.01)</td>
<td>1.44 (0.01)</td>
<td>1.44 (0.01)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \beta_s = 2 )</td>
<td>0.20</td>
<td>0.00</td>
<td>0.84</td>
<td>0.87 (&lt; 0.01)</td>
<td>1.89 (0.02)</td>
<td>1.89 (0.02)</td>
<td>1.89 (0.02)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \beta_x = 3 )</td>
<td>0.20</td>
<td>0.00</td>
<td>0.05</td>
<td>0.89 (0.01)</td>
<td>1.20 (&lt; 0.01)</td>
<td>1.20 (&lt; 0.01)</td>
<td>1.20 (&lt; 0.01)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \tau = 8 )</td>
<td>0.20</td>
<td>0.00</td>
<td>0.05</td>
<td>0.88 (0.01)</td>
<td>0.12 (&lt; 0.01)</td>
<td>0.12 (&lt; 0.01)</td>
<td>0.12 (&lt; 0.01)</td>
</tr>
</tbody>
</table>

Standard errors are presented in parentheses.
Table A.6: Simulation results using bootstrap of subject-level observations and 95% nominal level; results are based on 500 MC samples.

<table>
<thead>
<tr>
<th>Case</th>
<th>True Mean Function</th>
<th>Parameter</th>
<th>$\rho$</th>
<th>Bias</th>
<th>$\sqrt{\text{var}}$</th>
<th>ACP$^{\text{joint}}$</th>
<th>AL$^{\text{joint}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a)</td>
<td>$\beta_0 + \beta_s + \beta_x + \tau Z$</td>
<td>$\beta_0 = 5$</td>
<td>0.20</td>
<td>0.00</td>
<td>0.20</td>
<td>0.94 (0.01)</td>
<td>0.54 (&lt;0.01)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\beta_s = 2$</td>
<td>0.20</td>
<td>-0.01</td>
<td>0.27</td>
<td>0.94 (0.01)</td>
<td>0.70 (&lt;0.01)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\beta_x = 3$</td>
<td>0.20</td>
<td>0.01</td>
<td>0.40</td>
<td>0.94 (0.01)</td>
<td>1.06 (&lt;0.01)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\beta_s = 3$</td>
<td>0.20</td>
<td>0.01</td>
<td>0.52</td>
<td>0.94 (0.01)</td>
<td>1.39 (0.01)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\beta_x = 3$</td>
<td>0.20</td>
<td>0.00</td>
<td>0.05</td>
<td>0.95 (0.01)</td>
<td>0.14 (&lt;0.01)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\tau = 8$</td>
<td>0.20</td>
<td>0.00</td>
<td>0.05</td>
<td>0.94 (0.01)</td>
<td>0.14 (&lt;0.01)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\tau = 8$</td>
<td>0.90</td>
<td>0.00</td>
<td>0.05</td>
<td>0.93 (0.01)</td>
<td>0.14 (&lt;0.01)</td>
</tr>
</tbody>
</table>

Standard errors are presented in parentheses.

Table A.7: Simulation results for the case of time-varying covariate $X_{ij}$ using bootstrap of subject-level observations; results are based on 500 MC samples.

<table>
<thead>
<tr>
<th>Nominal Coverage</th>
<th>Case</th>
<th>True Mean Function</th>
<th>Parameter</th>
<th>$\rho$</th>
<th>ACP$^{\text{joint}}$</th>
<th>AL$^{\text{joint}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>85%</td>
<td>(d)</td>
<td>$\cos(2\pi s) + 4\left(\frac{X}{4}\right)^3 + \tau Z$</td>
<td>$\mu_4(s,X)$</td>
<td>0.20</td>
<td>0.82 (0.02)</td>
<td>0.56 (&lt;0.01)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.90</td>
<td>0.80 (0.02)</td>
<td>2.13 (0.01)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.20</td>
<td>2.07 (0.01)</td>
<td>2.25 (0.01)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.90</td>
<td>0.80 (0.02)</td>
<td>0.89 (0.01)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.20</td>
<td>0.80 (0.02)</td>
<td>2.23 (0.01)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.90</td>
<td>0.80 (0.02)</td>
<td>2.42 (0.01)</td>
</tr>
</tbody>
</table>

Standard errors are presented in parentheses.
A.4 Result for the case of nonparametric modeling

In practice one does not know the true model for the mean structure and thus nonparametric modeling is often preferable. Thus instead of assuming a correct parametric model for $\mu(s, x)$, we consider a completely nonparametric dependence of the mean response on $s$ and $X$. For example, in the case of F2 i. (a) the generating mean model is $\mu(s, x) = 5 + 2s + 3X$, but the assumed fitting model is an unknown, smooth bivariate function of both $t$ and $X$. The relevant results are presented in Tables A.8-A.10 for nominal coverages of 85%, 90% and 95% respectively. The simulation data are generated as described in Section 2.6. We use the residual-based bootstrap. Both pointwise and joint confidence bands perform excellently for all three cases of true mean functions, and for all three nominal coverages we consider here.

Table A.8: Simulation results fitting bivariate mean function using bootstrap of subject-level residuals and 85% nominal level; results are based on 500 MC samples.

<table>
<thead>
<tr>
<th>Case</th>
<th>True Mean Function</th>
<th>Parameter</th>
<th>$\rho$</th>
<th>$\text{ACP}^{\text{point}}$</th>
<th>$\text{AL}^{\text{point}}$</th>
<th>$\text{ACP}^{\text{joint}}$</th>
<th>$\text{AL}^{\text{joint}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a)</td>
<td>$\beta_0 + \beta_s s + \beta_x X + \tau Z$</td>
<td>$\mu(s, X) = 5 + 2s + 3X$</td>
<td>0.20</td>
<td>0.84 (&lt; 0.01)</td>
<td>1.22 (0.01)</td>
<td>0.82 (0.02)</td>
<td>2.85 (0.01)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\tau = 8$</td>
<td>0.90</td>
<td>0.83 (&lt; 0.01)</td>
<td>1.61 (0.01)</td>
<td>0.81 (0.02)</td>
<td>3.76 (0.01)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.90</td>
<td>0.84 (0.02)</td>
<td>0.10 (&lt; 0.01)</td>
<td>0.84 (0.02)</td>
<td>0.10 (&lt; 0.01)</td>
</tr>
<tr>
<td>(b)</td>
<td>$\beta_0 + \beta_s s + \beta_x X + \beta_t t + \tau Z$</td>
<td>$\mu(s, X) = 5 + 2s + 3X + 7sX$</td>
<td>0.20</td>
<td>0.84 (&lt; 0.01)</td>
<td>1.22 (0.01)</td>
<td>0.82 (0.02)</td>
<td>2.85 (0.01)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\tau = 8$</td>
<td>0.90</td>
<td>0.83 (&lt; 0.01)</td>
<td>1.61 (0.01)</td>
<td>0.81 (0.02)</td>
<td>3.76 (0.01)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.90</td>
<td>0.84 (0.02)</td>
<td>0.10 (&lt; 0.01)</td>
<td>0.84 (0.02)</td>
<td>0.10 (&lt; 0.01)</td>
</tr>
<tr>
<td>(c)</td>
<td>$\cos(2\pi s) + \beta_x X + \tau Z$</td>
<td>$\mu(s, X) = \cos(2\pi s) + 3X$</td>
<td>0.20</td>
<td>0.84 (&lt; 0.01)</td>
<td>1.22 (0.01)</td>
<td>0.82 (0.02)</td>
<td>2.86 (0.01)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\tau = 8$</td>
<td>0.90</td>
<td>0.83 (&lt; 0.01)</td>
<td>1.61 (0.01)</td>
<td>0.81 (0.02)</td>
<td>3.77 (0.01)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.90</td>
<td>0.84 (0.02)</td>
<td>0.10 (&lt; 0.01)</td>
<td>0.84 (0.02)</td>
<td>0.10 (&lt; 0.01)</td>
</tr>
</tbody>
</table>

Standard errors are presented in parentheses.
Table A.9: Simulation results fitting bivariate mean function using bootstrap of subject-level residuals and 90% nominal level; results are based on 500 MC samples.

<table>
<thead>
<tr>
<th>Case</th>
<th>True Mean Function</th>
<th>Parameter</th>
<th>ρ</th>
<th>ACPmean</th>
<th>ALCmean</th>
<th>ACPmin</th>
<th>ALmin</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a)</td>
<td>( \beta_0 + \beta_s s + \beta_x X + \tau Z )</td>
<td>( \mu(s, X) = 5 + 2s + 3X )</td>
<td>0.20</td>
<td>0.89 (&lt;0.01)</td>
<td>1.39 (0.01)</td>
<td>0.87 (0.01)</td>
<td>2.99 (0.01)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \tau = 8 )</td>
<td>0.90</td>
<td>0.88 (0.01)</td>
<td>1.83 (0.02)</td>
<td>0.87 (0.02)</td>
<td>3.95 (0.01)</td>
</tr>
<tr>
<td>(b)</td>
<td>( \beta_0 + \beta_s s + \beta_x X + \beta_s x s + \tau Z )</td>
<td>( \mu(s, X) = 5 + 2s + 3X + 7sX )</td>
<td>0.20</td>
<td>0.89 (&lt;0.01)</td>
<td>1.39 (0.01)</td>
<td>0.87 (0.01)</td>
<td>2.99 (0.01)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \tau = 8 )</td>
<td>0.90</td>
<td>0.88 (0.01)</td>
<td>1.83 (0.02)</td>
<td>0.87 (0.02)</td>
<td>3.95 (0.01)</td>
</tr>
<tr>
<td>(c)</td>
<td>( \cos(2\pi s) + \beta_x X + \tau Z )</td>
<td>( \mu(s, X) = \cos(2\pi s) + 3X )</td>
<td>0.20</td>
<td>0.89 (&lt;0.01)</td>
<td>1.39 (0.01)</td>
<td>0.88 (0.01)</td>
<td>3.00 (0.01)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \tau = 8 )</td>
<td>0.90</td>
<td>0.88 (0.01)</td>
<td>1.84 (0.02)</td>
<td>0.87 (0.02)</td>
<td>3.96 (0.01)</td>
</tr>
</tbody>
</table>

Standard errors are presented in parentheses.

Table A.10: Simulation results fitting bivariate mean function using bootstrap of subject-level residuals and 95% nominal level; results are based on 500 MC samples.

<table>
<thead>
<tr>
<th>Case</th>
<th>True Mean Function</th>
<th>Parameter</th>
<th>ρ</th>
<th>ACPmean</th>
<th>ALCmean</th>
<th>ACPmin</th>
<th>ALmin</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a)</td>
<td>( \beta_0 + \beta_s s + \beta_x X + \tau Z )</td>
<td>( \mu(s, X) = 5 + 2s + 3X )</td>
<td>0.20</td>
<td>0.94 (&lt;0.01)</td>
<td>1.65 (0.01)</td>
<td>0.94 (0.01)</td>
<td>3.22 (0.01)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \tau = 8 )</td>
<td>0.90</td>
<td>0.94 (0.01)</td>
<td>2.17 (0.02)</td>
<td>0.93 (0.01)</td>
<td>4.24 (0.01)</td>
</tr>
<tr>
<td>(b)</td>
<td>( \beta_0 + \beta_s s + \beta_x X + \beta_s x s + \tau Z )</td>
<td>( \mu(s, X) = 5 + 2s + 3X + 7sX )</td>
<td>0.20</td>
<td>0.94 (&lt;0.01)</td>
<td>1.66 (0.01)</td>
<td>0.94 (0.01)</td>
<td>3.22 (0.01)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \tau = 8 )</td>
<td>0.90</td>
<td>0.93 (0.01)</td>
<td>2.17 (0.02)</td>
<td>0.93 (0.01)</td>
<td>4.24 (0.01)</td>
</tr>
<tr>
<td>(c)</td>
<td>( \cos(2\pi s) + \beta_x X + \tau Z )</td>
<td>( \mu(s, X) = \cos(2\pi s) + 3X )</td>
<td>0.20</td>
<td>0.94 (&lt;0.01)</td>
<td>1.66 (0.01)</td>
<td>0.93 (0.01)</td>
<td>3.24 (0.01)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \tau = 8 )</td>
<td>0.90</td>
<td>0.94 (0.01)</td>
<td>2.18 (0.02)</td>
<td>0.93 (0.01)</td>
<td>4.25 (0.01)</td>
</tr>
</tbody>
</table>

Standard errors are presented in parentheses.
A.5 Results for the case of having non-normal errors

We conduct additional simulation study to assess robustness of the proposed inferential methods to non-Gaussianity. Recall that $w_{ij}(s)$ is white noises in the error term, $\epsilon_{ij}(s) = \sum_{l=1}^{3} \xi_{ijl}\phi_{l}(s) + w_{ij}(s)$, defined in Section 2.6. Here we consider two non-Gaussian distributions to generate $w_{ij}(t)$: namely, (1) t-distribution with degrees of freedom, $df \approx 2.46$, and (2) skew normal distribution with mean zero, variance $\sigma^2 = 5.33$, and skewness parameter $\zeta = 50$ (Fernández & Steel, 1998). Skew normal random variables were generated using the fGarch package (Wuertz et al., 2009) in R (R Core Team, 2014). True values of the parameters, $df$ and $\sigma^2$, are determined such that signal to noise remains to be 1. Nomal, skew normal, and t distributions used to generate white noises are shown in A.2. The results are similar to the Gaussian case and suggest that the proposed methods are robust to non-Gaussian error distribution.

![Figure A.2: Normal, skew-normal, and t distributions used to generate white noises.](image-url)
Table A.11: Simulation results for the case of having non-Gaussian errors using bootstrap of subject-level residuals and 85% nominal level; results are based on 500 MC samples.

### t-distribution

<table>
<thead>
<tr>
<th>Case</th>
<th>True Mean Function</th>
<th>Parameter</th>
<th>$\rho$</th>
<th>$\text{ACP}^{\text{point}}$</th>
<th>$\text{AL}^{\text{point}}$</th>
<th>$\text{ACP}^{\text{joint}}$</th>
<th>$\text{AL}^{\text{joint}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(c)</td>
<td>$f(s) = \cos(2\pi s)$</td>
<td>$\beta_s$ = 3</td>
<td>0.20</td>
<td>0.82 (0.01)</td>
<td>0.74 (0.01)</td>
<td>0.81 (0.02)</td>
<td>1.00 (0.01)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\tau$ = 8</td>
<td>0.20</td>
<td>0.83 (0.02)</td>
<td>0.64 (0.01)</td>
<td>0.81 (0.02)</td>
<td>1.00 (0.01)</td>
</tr>
<tr>
<td>(d)</td>
<td>$f(s) = \cos(2\pi s) + 4((X/4) - s)^3 + \tau Z$</td>
<td>$\mu(s, X)$</td>
<td>0.20</td>
<td>0.83 (0.01)</td>
<td>1.22 (0.01)</td>
<td>0.90 (0.01)</td>
<td>3.10 (0.01)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\tau$ = 8</td>
<td>0.20</td>
<td>0.83 (0.08)</td>
<td>1.61 (0.01)</td>
<td>0.88 (0.01)</td>
<td>4.10 (0.01)</td>
</tr>
</tbody>
</table>

### skew-normal distribution

<table>
<thead>
<tr>
<th>Case</th>
<th>True Mean Function</th>
<th>Parameter</th>
<th>$\rho$</th>
<th>$\text{ACP}^{\text{point}}$</th>
<th>$\text{AL}^{\text{point}}$</th>
<th>$\text{ACP}^{\text{joint}}$</th>
<th>$\text{AL}^{\text{joint}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(c)</td>
<td>$f(s) = \cos(2\pi s)$</td>
<td>$\beta_s$ = 3</td>
<td>0.20</td>
<td>0.84 (0.01)</td>
<td>0.83 (0.01)</td>
<td>0.83 (0.02)</td>
<td>1.00 (0.01)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\tau$ = 8</td>
<td>0.20</td>
<td>0.84 (0.02)</td>
<td>0.64 (0.01)</td>
<td>0.83 (0.02)</td>
<td>1.00 (0.01)</td>
</tr>
<tr>
<td>(d)</td>
<td>$f(s) = \cos(2\pi s) + 4((X/4) - s)^3 + \tau Z$</td>
<td>$\mu(s, X)$</td>
<td>0.20</td>
<td>0.84 (0.01)</td>
<td>1.23 (0.01)</td>
<td>0.89 (0.01)</td>
<td>3.11 (0.01)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\tau$ = 8</td>
<td>0.20</td>
<td>0.84 (0.08)</td>
<td>1.61 (0.01)</td>
<td>0.89 (0.01)</td>
<td>4.09 (0.01)</td>
</tr>
</tbody>
</table>

Standard errors are presented in parentheses.
Table A.12: Simulation results for the case of having non-Gaussian errors using bootstrap of subject-level residuals and 90% nominal level; results are based on 500 MC samples.

| Case | True Mean Function | Parameter | \( \rho \) | DCpe\text{joint} & DCp\text{point} & DC\text{joint} & DC\text{point} |
|------|--------------------|-----------|-----|----------------|----------------|----------------|----------------|
| (c)  | \( \cos(2\pi s) + \beta_s X + \tau Z \) | \( f(s) = \cos(2\pi s) \) | 0.20 | 0.87 (0.01) | 0.55 (<0.01) | 0.87 (0.01) | 0.84 (<0.01) |
|      |                    | \( \beta_s = 3 \) | 0.20 | 0.89 (0.01) | 0.11 (<0.01) | 0.87 (0.02) | 1.09 (<0.01) |
|      |                    | \( \tau = 8 \) | 0.20 | 0.90 (0.01) | 0.11 (<0.01) | 0.89 (0.01) | 0.12 (<0.01) |
| (d)  | \( \cos(2\pi s) + 4((X/4) - s)^3 + \tau Z \) | \( \mu_4(s, X) \) | 0.20 | 0.88 (<0.01) | 1.40 (<0.01) | 0.93 (0.01) | 3.23 (0.01) |
|      |                    | \( \beta_s = 3 \) | 0.20 | 0.89 (0.01) | 0.12 (<0.01) | 0.90 (0.01) | 0.12 (<0.01) |
|      |                    | \( \tau = 8 \) | 0.20 | 0.90 (0.06) | 0.11 (<0.01) | 0.90 (0.06) | 0.11 (<0.01) |

| Case | True Mean Function | Parameter | \( \rho \) | DCpe\text{joint} & DCp\text{point} & DC\text{joint} & DC\text{point} |
|------|--------------------|-----------|-----|----------------|----------------|----------------|----------------|
| (c)  | \( \cos(2\pi s) + \beta_s X + \tau Z \) | \( f(s) = \cos(2\pi s) \) | 0.20 | 0.89 (0.01) | 0.56 (<0.01) | 0.88 (0.01) | 0.84 (<0.01) |
|      |                    | \( \beta_s = 3 \) | 0.20 | 0.89 (0.01) | 0.12 (<0.01) | 0.89 (0.01) | 0.12 (<0.01) |
|      |                    | \( \tau = 8 \) | 0.20 | 0.90 (0.01) | 0.12 (<0.01) | 0.91 (0.01) | 0.12 (<0.01) |
| (d)  | \( \cos(2\pi s) + 4((X/4) - s)^3 + \tau Z \) | \( \mu_4(s, X) \) | 0.20 | 0.88 (<0.01) | 1.40 (<0.01) | 0.92 (0.01) | 4.27 (0.01) |
|      |                    | \( \beta_s = 3 \) | 0.20 | 0.89 (0.01) | 0.12 (<0.01) | 0.91 (0.01) | 0.12 (<0.01) |
|      |                    | \( \tau = 8 \) | 0.20 | 0.89 (0.07) | 0.12 (<0.01) | 0.89 (0.07) | 0.12 (<0.01) |

Standard errors are presented in parentheses.
Table A.13: Simulation results for the case of having non-Gaussian errors using bootstrap of subject-level residuals and 95% nominal level; results are based on 500 MC samples.

<table>
<thead>
<tr>
<th>t-distribution</th>
<th>Case</th>
<th>True Mean Function</th>
<th>Parameter</th>
<th>$\rho$</th>
<th>ACPoint</th>
<th>ALPoint</th>
<th>ACJoint</th>
<th>ALJoint</th>
</tr>
</thead>
<tbody>
<tr>
<td>(c) $\cos(2\pi s) + \beta_s X + \tau Z$</td>
<td>$f(s) = \cos(2\pi s)$</td>
<td>$\beta_s = 3$</td>
<td>0.20</td>
<td>0.93 (0.01)</td>
<td>0.66 (&lt;0.01)</td>
<td>0.93 (0.01)</td>
<td>0.94 (&lt;0.01)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\tau = 8$</td>
<td>0.20</td>
<td>0.94 (0.01)</td>
<td>0.86 (&lt;0.01)</td>
<td>0.93 (0.01)</td>
<td>1.22 (&lt;0.01)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\beta_s = 3$</td>
<td>0.20</td>
<td>0.94 (0.01)</td>
<td>0.86 (&lt;0.01)</td>
<td>0.93 (0.01)</td>
<td>1.22 (&lt;0.01)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\tau = 8$</td>
<td>0.20</td>
<td>0.94 (0.01)</td>
<td>0.86 (&lt;0.01)</td>
<td>0.93 (0.01)</td>
<td>1.22 (&lt;0.01)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\tau = 8$</td>
<td>0.20</td>
<td>0.94 (0.01)</td>
<td>0.86 (&lt;0.01)</td>
<td>0.93 (0.01)</td>
<td>1.22 (&lt;0.01)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>skew-normal distribution</th>
<th>Case</th>
<th>True Mean Function</th>
<th>Parameter</th>
<th>$\rho$</th>
<th>ACPoint</th>
<th>ALPoint</th>
<th>ACJoint</th>
<th>ALJoint</th>
</tr>
</thead>
<tbody>
<tr>
<td>(c) $\cos(2\pi s) + \beta_s X + \tau Z$</td>
<td>$f(s) = \cos(2\pi s)$</td>
<td>$\beta_s = 3$</td>
<td>0.20</td>
<td>0.94 (0.01)</td>
<td>1.66 (&lt;0.01)</td>
<td>0.96 (0.01)</td>
<td>3.44 (0.01)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\tau = 8$</td>
<td>0.20</td>
<td>0.93 (&lt;0.01)</td>
<td>2.18 (&lt;0.01)</td>
<td>0.96 (0.01)</td>
<td>4.53 (0.01)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\beta_s = 3$</td>
<td>0.20</td>
<td>0.94 (0.01)</td>
<td>1.44 (&lt;0.01)</td>
<td>0.93 (0.01)</td>
<td>2.18 (0.01)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\tau = 8$</td>
<td>0.20</td>
<td>0.95 (0.01)</td>
<td>1.44 (&lt;0.01)</td>
<td>0.94 (0.01)</td>
<td>2.18 (0.01)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\tau = 8$</td>
<td>0.20</td>
<td>0.95 (0.01)</td>
<td>1.44 (&lt;0.01)</td>
<td>0.94 (0.01)</td>
<td>2.18 (0.01)</td>
<td></td>
</tr>
</tbody>
</table>

Standard errors are presented in parentheses.
Table A.14: Empirical Type I error of the test statistic $T$ based on the $N_{sim} = 1000$ MC samples for the case of having non-Gaussian errors.

<table>
<thead>
<tr>
<th></th>
<th>$\mu(s, x) = \cos(2\pi s)$, $\tau = 8$</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\alpha = 0.05$</td>
<td>$\alpha = 0.10$</td>
<td>$\alpha = 0.15$</td>
</tr>
<tr>
<td>$n = 100$ $\rho = 0.2$</td>
<td>0.08 (0.01)</td>
<td>0.15 (0.01)</td>
<td>0.21 (0.01)</td>
</tr>
<tr>
<td>$n = 200$ $\rho = 0.2$</td>
<td>0.07 (0.01)</td>
<td>0.13 (0.01)</td>
<td>0.19 (0.01)</td>
</tr>
<tr>
<td>$n = 300$ $\rho = 0.2$</td>
<td>0.07 (0.01)</td>
<td>0.12 (0.01)</td>
<td>0.17 (0.01)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>$\mu(s, x) = \cos(2\pi s)$, $\tau = 8$</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\alpha = 0.05$</td>
<td>$\alpha = 0.10$</td>
<td>$\alpha = 0.15$</td>
</tr>
<tr>
<td>$n = 100$ $\rho = 0.2$</td>
<td>0.09 (0.01)</td>
<td>0.15 (0.01)</td>
<td>0.22 (0.01)</td>
</tr>
<tr>
<td>$n = 200$ $\rho = 0.2$</td>
<td>0.05 (0.01)</td>
<td>0.09 (0.01)</td>
<td>0.14 (0.01)</td>
</tr>
<tr>
<td>$n = 300$ $\rho = 0.2$</td>
<td>0.06 (0.01)</td>
<td>0.11 (0.01)</td>
<td>0.16 (0.01)</td>
</tr>
</tbody>
</table>

Standard errors are presented in parentheses.
Appendix B

Additional Details for Chapter 3

Section B.1 provides detailed proofs of the theoretical results given in Section 3.5. Specifically Section B.1.1 corresponds to Section 3.5.1 and includes proofs of the theoretical results for the case when $Y_{ij}(s) = X_{ij}(s)$. Whereas Section B.1.2 corresponds to Section 3.5.2 and includes proofs for the case when $Y_{ij}(s) = X_{ij}(s) + \epsilon_{1,ij}(s)$. Section B.2 includes additional simulation results for the case when data are corrupted only with white noise error, and results obtained using Chen & Müller (2012)'s method. Section B.3 includes additional figures for the DTI data analysis.

B.1 Proofs of theoretical results given in Section 3.5

B.1.1 Case when response curves are measured without error (Section 3.5.1)

Here we consider the case when $Y_{ij}(s) = X_{ij}(s)$; thus in this section, $Y(s,t)$ satisfies all of the assumptions made on $X(s,t)$. We first show the following corollaries:

**Corollary 4.** Under the assumptions (A1.) and (A2.), the marginal covariance function, $\Sigma(s,s') = \int c\{(s,t),(s',t)\}g(t)dt$, (i) is symmetric, (ii) is positive definite, and (iii) has eigenvalues satisfying that $\sum_{k=1}^{\infty} \lambda_k$ is finite.
Proof. We assume that $X_i(s, t)$ is a realization of a random process, $X$, that satisfies the assumption (A1.). By definition, the covariance operator of $X$ is $C(y) = E[<X, y>]X$, $y \in L^2(S \times T)$, and it follows that

$$C(y)(s, t) = \int \int c((s, t), (s', t'))y(s', t')ds'dt',$$

where $c((s, t), (s', t')) = E[X(s, t)X(s', t')]$. Since the operator, $C$, is a properly defined covariance operator, it (i) is symmetric, i.e. $c((s, t), (s', t')) = c((s', t'), (s, t))$, (ii) is positive definite, i.e. $\int\int\int\int c((s, t), (s', t'))y(s, t)y(s', t')ds'ds'dtdt' < \infty$ (Horváth & Kokoszka, 2012). In the following, we show that $\Sigma(s, s') = \int c((s, t), (s', t))g(t)dt$ has the same properties as $c((s, t), (s', t'))$ has, and thus is also a proper covariance function.

(i) $\Sigma(s, s')$ is symmetric.

$$\Sigma(s, s') = \int c((s, t), (s', t))g(t)dt = \int E\{X_i(s, t)X_i(s', t)\}g(t)dt$$

$$= \int E\{X_i(s', t)X_i(s, t)\}g(t)dt = \int c((s', t), (s, t))g(t)dt = \Sigma(s', s)$$

(ii) $\Sigma(s, s')$ is positive definite.

To show $\Sigma(s, s')$ is positive definite we need to prove the following holds:

$$\int\int \Sigma(s, s')z(s)z(s')dsds' = \int\left[ \int c((s, t), (s', t))g(t)dt \right]z(s)z(s')dsds'$$

$$\overset{\text{Fubini's th}}{=} \int g(t)\left[ \int\int c((s, t), (s', t))z(s)z(s')dsds' \right]dt \geq 0.$$
And because a density function, \( g(t) \), is non-negative by definition, it is sufficient to show that \( \iint c((s, t), (s', t)) z(s) z'(s') dsds' \geq 0 \) for any \( z(\cdot) \in L^2([0, 1]) \) and it is equivalent to show that \( X(s, t^*) \) for arbitrary fixed \( t^* \in \mathcal{T} \) is square integrable.

We prove this by contradiction. Let \( Z(t) = \int X^2(s, t) ds \). Assume that \( E[Z(t)] = \infty \) for \( t \in \mathcal{T}_0 \subset \mathcal{T} \) such that \( \int_{\mathcal{T}_0} g(t) dt > 0 \). It follows that \( \int_{\mathcal{T}} Z(t) dt > \int_{\mathcal{T}_0} Z(t) dt \), because \( \mathcal{T}_0 \) is a subset of \( \mathcal{T} \). As expectation is a linear operator, \( E \int_{\mathcal{T}} Z(t) dt > E \int_{\mathcal{T}_0} Z(t) dt \) still holds. It implies that \( E \int_{\mathcal{T}} Z(t) dt \) is infinite as we assume that \( E[Z(t)] = \infty \) for \( t \in \mathcal{T}_0 \). However because \( X(s, t) \) is square integrable and \( E\{ \int_{\mathcal{T}} Z(t) dt \} = E\{ \iint X^2(s, t) dsdt \} \), \( E\{ \int_{\mathcal{T}} Z(t) dt \} \) is finite. Thus by contradiction, we show that \( E[Z(t)] \) is finite and \( X(s, t^*) \) is square integrable for fixed \( t^* \in \mathcal{T} \). And it follows that \( \Sigma(s, s') \) is positive definite.

(iii) Eigenvalues of \( \Sigma(s, s') \) satisfy \( \sum_{l=1}^{\infty} \lambda_l < \infty \), which is equivalent to \( \int \Sigma(s, s) ds < \infty \).

Because \( c((s, t), (s', t')) \) is a continuous function in \( S \times \mathcal{T} \) and the intervals, \( S \) and \( \mathcal{T} \), are compact, \( \iint c((s, t), (s, t)) dsdt \) is finite. Hence under the assumptions (A1.) and (A2.),

\[
\int \Sigma(s, s) ds = \iint c((s, t), (s, t)) g(t) dt ds \\
\overset{\text{Fubini's th}}{=} \int g(t) \left[ \int c((s, t), (s, t)) ds \right] dt \\
\leq \int \sup_{t \in \mathcal{T}} g(t) \left[ \int c((s, t), (s, t)) ds \right] dt \\
= \sup_{t \in \mathcal{T}} g(t) \int c((s, t), (s, t)) ds dt < \infty,
\]

and equivalently, the sum of eigenvalues of \( \Sigma(s, s') \) is finite.

Thus we obtain that \( \Sigma(s, s') = \int c((s, t), (s', t)) g(t) dt \) is a proper covariance function. The reason why we refer \( \Sigma(s, s') \) as the marginal covariance function of \( X_i(s, t) \) is as follows. Assume that there is a true latent process, \( U_i(\cdot) \), such that \( X_i(\cdot, t_{ij}) \) is \( U_i(\cdot) \) given
that \( t = t_{ij} \). Then the marginal covariance function, \( \Sigma(s, s') \), is the covariance function of the true latent process, \( U_i(s) \); specifically,

\[
\text{cov}\{U_i(s), U_i(s')\} = E_t[\text{cov}\{U_i(s), U_i(s')|t = t_{ij}\}] + E_t[\text{cov}\{U_i(s)|t = t_{ij}\}E\{U_i(s')|t = t_{ij}\}]
\]

\[
= E_t[\text{cov}\{X_i(s, t_{ij}), X_i(s', t_{ij})\}] + \text{cov}[E_t\{X_i(s, t_{ij})\}, E\{X_i(s', t_{ij})\}],
\]

where the first term is

\[
E_t[\text{cov}\{Y_i(s, t_{ij}), Y_i(s', t_{ij})\}] = \int c\{(s, t), (s', t)\}g(t)dt
\]

\[
= \Sigma(s, s'),
\]

and the second term is equal to 0.

Let \( \{t_d : d = 1, ..., D\} \) be a set of unique \( t_{ij} \)'s for all \( i \) and \( j \) in an increasing order, such that \( t_1 < t_2 < \ldots < t_D \) and assume \( t_0 = 0 \) and \( t_D = 1 \).

**Corollary 5.** \( \hat{\Sigma}(s, s') \) is an unbiased estimator of the marginal covariance function, \( \Sigma(s, s') \).

**Proof.** Here we show that that the sample covariance,

\[
\hat{\Sigma}(s, s') = \sum_{i=1}^{n} \frac{1}{\sum_{i=1}^{m_i} \sum_{j=1}^{m_i} Y_{ij}(s)Y_{ij}(s')/(\sum_{i=1}^{n} m_i), \text{ is an unbiased estimator of } \Sigma(s, s'), \text{ i.e. } E\{\hat{\Sigma}(s, s')\} = \Sigma(s, s'). \text{ Recall that here we are assuming } Y_{ij}(s) = X_{ij}(s). \text{ The proof is as follows:}
\]

\[
E\{\hat{\Sigma}(s, s')\} = E\left\{ \frac{1}{\sum_{i=1}^{m_i} \sum_{j=1}^{m_i} Y_{ij}(s)Y_{ij}(s')} \right\}
\]

\[
= E\left\{ \frac{1}{\sum_{i=1}^{m_i} \sum_{d=1}^{D} \sum_{i=1}^{n} \sum_{j=1}^{m_i} Y_{ij}(s)Y_{ij}(s')I(t_{ij} \in (t_{d-1}, t_d))} \right\}
\]

123
\[\begin{align*}
&\frac{1}{\sum_{i=1}^{n} m_i} \sum_{d=1}^{D} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \mathbb{E}\{Y_{ij}(s)Y_{ij}(s')I(t_{ij} \in (t_{d-1}, t_d)]\} \\
&= \frac{1}{\sum_{i=1}^{n} m_i} \sum_{d=1}^{D} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \mathbb{E}\{Y_i(s, t_d)Y_i(s', t_d)\} \mathbb{E}\{I(t_{ij} \in (t_{d-1}, t_d)]\} \\
&= \frac{1}{\sum_{i=1}^{n} m_i} \sum_{d=1}^{D} \sum_{i=1}^{n} \sum_{j=1}^{m_i} c((s, t_d), (s', t_d)) P(t \in (t_{d-1}, t_d]) \quad \text{(because } Y_i(s, t_d) = X_i(s, t_d)) \\
&= \sum_{d=1}^{D} c((s, t_d), (s', t_d)) P(t \in (t_{d-1}, t_d]) = \int c((s, t), (s', t)) dF(t) = \Sigma(s, s'),
\end{align*}\]

where \( F \) is a true sampling distribution function of \( t \).

**Theorem 4.1.1** (Theorem 1 in Section 3.5.1). Assume (A1.) - (A3.) hold. Then \(|\hat{\Sigma}(s, s') - \Sigma(s, s')| \xrightarrow{p} 0\) as \( n \) diverges. If in addition (A4.) holds, then

\[\|\hat{\Sigma}(\cdot, \cdot) - \Sigma(\cdot, \cdot)\|_s \xrightarrow{p} 0 \quad \text{as} \quad n \rightarrow \infty, \quad \text{(B.2)}\]

where \( \|k(\cdot, \cdot)\|_s = \left\{ \int\int k^2(s, s') dsds' \right\}^{1/2} \) is the Hilbert-Schmidt norm of \( k(\cdot, \cdot) \).

**Proof.** We prove this theorem in two steps: first we show (i) the pointwise consistency and then we show (ii) the Hilbert-Schmidt norm consistency.

(i) The pointwise consistency

Let \( M_1 = \sum_{i=1}^{n} m_i \), \( M_2 = \sum_{i=1}^{n} m_i^2 \), and \( P_d = P(t \in (t_{d-1}, t_d]) \). Let \( A_i(t) \) be a realization of \( A(t) = Y(s, t)Y(s', t) \) for fixed \( s \) and \( s' \), and let \( V_d = (D/M_1) \sum_{i=1}^{n} \sum_{j=1}^{m_i} A_i(t_d) I(t_{ij} \in (t_{d-1}, t_d)] \). Then the estimator, \( \hat{\Sigma}(s, s') \), can be written as \( \hat{\Sigma} = D^{-1} \sum_{d=1}^{D} V_d; \)

\[\begin{align*}
\hat{\Sigma}(s, s') &= \frac{1}{\sum_{i=1}^{n} m_i} \sum_{i=1}^{n} \sum_{j=1}^{m_i} Y_{ij}(s)Y_{ij}(s') \\
&= \frac{1}{D} \sum_{d=1}^{D} \sum_{i=1}^{n} \sum_{j=1}^{m_i} Y_i(s, t_d)Y_i(s', t_d) I(t_{ij} \in (t_{d-1}, t_d])
\end{align*}\]

124
\[
= \frac{1}{D} \sum_{d=1}^{D} \frac{D}{M_1} \sum_{i=1}^{n} \sum_{j=1}^{m_i} A_i(t_d)I(t_{ij} \in (t_{d-1}, t_d)) = \frac{1}{D} \sum_{d=1}^{D} V_d.
\]

Notice that \(V_d\)'s are correlated over \(d\). To show the consistency of the average of the correlated random variables, we first obtain the covariance of \(V_d\) and \(V_{d'}\),

\[
\text{cov}(V_d, V_{d'}) = E[V_d V_{d'}] - E[V_d]E[V_{d'}]
\]

and then use Theorem 5.3 of Boos & Stefanski (2013, p.208). For completeness the theorem we used is given below:

**Theorem 5.3 (Boos & Stefanski, 2013, p.208)** If \(X_1, ..., X_n\) are random variable with finite means \(\mu_i = E(X_i)\), variance \(E(X_i - \mu_i)^2 = \sigma_i^2\), and covariances \(E(X_i - \mu_i)(X_j - \mu_j) = \sigma_{ij}\) such that

\[
\text{var}(\bar{X}) = \frac{1}{n^2} \left[ \sum_{i=1}^{n} \sigma_i^2 + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \sigma_{ij} \right] \to 0 \text{ as } n \to \infty,
\]

then \(\bar{X} - \bar{\mu} \xrightarrow{p} 0 \text{ as } n \to \infty\), where \(\bar{\mu} = n^{-1} \sum_{i=1}^{n} \mu_i\)

In the following we obtain \(\text{cov}(V_d, V_{d'})\) by getting \(E[V_d]\) and \(E[V_d V_{d'}]\):

\[
\text{E}[V_d] = E \left[ \frac{D}{M_1} \sum_{i=1}^{n} \sum_{j=1}^{m_i} Y_i(s, t_d)Y_i(s', t_d)I(t_{ij} \in (t_{d-1}, t_d)) \right]
\]

\[
= D \cdot E[Y_i(s, t_d)Y_i(s', t_d)] \cdot E[I(t_{ij} \in (t_{d-1}, t_d)]
\]

\[
= D \cdot c((s, t_d), (s', t_d)) \cdot P_d \quad \text{(because } Y_i(s, t) = X_i(s, t))
\]

\[
\text{E}[V_d V_{d'}]
\]

\[
= E \left\{ \frac{D}{M_1} \sum_{i=1}^{n} \sum_{j=1}^{m_i} A_i(t_d)I(t_{ij} \in (t_{d-1}, t_d)) \right\} \left\{ \frac{D}{M_1} \sum_{i' = 1}^{n} \sum_{j' = 1}^{m_{i'}} A_{i'}(t_{d'})I(t_{i'j'} \in (t_{d'-1}, t_{d'})) \right\}
\]

\[
= \left( \frac{D}{M_1} \right)^2 E \left[ \sum_{i=1}^{n} \sum_{j=1}^{m_i} \sum_{i' = 1}^{n} \sum_{j' = 1}^{m_{i'}} A_i(t_d)I(t_{ij} \in (t_{d-1}, t_d))A_{i'}(t_{d'})I(t_{i'j'} \in (t_{d'-1}, t_{d'})) \right]
\]

125
$$= \left( \frac{D}{M_1} \right)^2 E \left[ \sum_{i=1}^{n} \sum_{i'=1}^{n} \sum_{j=1}^{m_i} \sum_{j'=1}^{m_i'} A_i(t_d)A_i'(t_{d'})I(t_{ij} \in (t_{d-1}, t_d))I(t_{ij'} \in (t_{d'-1}, t_{d'})) \right]$$

$$= \left( \frac{D}{M_1} \right)^2 \sum_{i,i'} \sum_{j,j'} E \left[ A_i(t_d)A_i'(t_{d'})I(t_{ij} \in (t_{d-1}, t_d))I(t_{ij'} \in (t_{d'-1}, t_{d'})) \right]$$

$$= \left( \frac{D}{M_1} \right)^2 \sum_{i,i'} \sum_{j,j'} E \left[ A_i(t_d)A_i'(t_{d'}) \right] E \left[ I(t_{ij} \in (t_{d-1}, t_d))I(t_{ij'} \in (t_{d'-1}, t_{d'})) \right] \quad \text{(B.3)}$$

Here we consider Equation (B.3) for two cases; when \(d = d'\) and when \(d \neq d'\).

**Case 1: \(d = d'\)**

This is equivalent to say that \(t_d = t_{d'}\) as \(\{t_d : d = 1, \ldots, D\}\) is a set of unique values. Let \(t^* = t_d = t_{d'}\).

$$E \left[ I(t_{ij} \in (t_{d-1}, t_d))I(t_{ij'} \in (t_{d'-1}, t_{d'})) \right] = P(t_{ij} \in (t_{d-1}, t_d) \text{ and } t_{ij'} \in (t_{d-1}, t_d))$$

$$= \begin{cases} 
P(t_{ij} \in (t_{d-1}, t_d))P(t_{ij'} \in (t_{d-1}, t_d)) = P_d^2 & \text{for } i \neq i', j, j' \\
P(t_{ij} \in (t_{d-1}, t_d)) = P_d & \text{for } i = i', j = j' \\
0 & \text{for } i = i', j \neq j' 
\end{cases}$$

$$E \left[ A_i(t_d)A_i'(t_{d'}) \right] = E \left[ Y_i(s, t_d)Y_i(s', t_d)Y_i'(s, t_{d'})Y_i'(s', t_{d'}) \right]$$

$$= E \left[ Y_i(s, t^*)Y_i(s', t^*)Y_i'(s, t^*)Y_i'(s', t^*) \right]$$

126
\[
\begin{align*}
\mathbb{E}[Y_i(s, t^*)Y_i(s', t^*)] & \mathbb{E}[Y_{i'}(s, t^*)Y_{i'}(s', t^*)] \\
& = c((s, t^*), (s', t^*))^2 \quad \text{for } i \neq i', j, j' \\
& \quad \text{(because } Y_i(s, t_d) = X_i(s, t_d)) \\
= & \left\{ \begin{array}{ll}
\mathbb{E}\left[ \left\{ Y_i(s, t^*)Y_i(s', t^*) \right\}^2 \right] \\
& = \text{var}\left\{ Y(s, t^*)Y(s', t^*) \right\} + \left\{ c((s, t^*), (s', t^*)) \right\}^2 \quad \text{for } i = i', j, j' \\
& \quad \text{(because } Y_i(s, t_d) = X_i(s, t_d))
\end{array} \right.
\end{align*}
\]

By summing over all \(i, i', j, j'\), Equation (B.3) for the case of \(d = d'\) equals to

\[
\begin{align*}
\mathbb{E}[V_d V_{d'}] = & \left( \frac{D}{M_1} \right)^2 \left\{ \sum_{i=1}^n m_i(M_1 - m_i) \right\} \left\{ P_d \cdot c((s, t^*), (s', t^*)) \right\}^2 \\
& + \left( \frac{D}{M_1} \right)^2 M_1 \cdot P_d \cdot \left[ \text{var}\left\{ Y(s, t^*)Y(s', t^*) \right\} + \left\{ c((s, t^*), (s', t^*)) \right\}^2 \right] \\
= & \left( D^2 - \frac{M_2 D^2}{M_1^2} \right) \left\{ P_d \cdot c((s, t^*), (s', t^*)) \right\}^2 \\
& + \left( \frac{D^2}{M_1} \right) \cdot P_d \cdot \left[ \text{var}\left\{ Y(s, t^*)Y(s', t^*) \right\} + \left\{ c((s, t^*), (s', t^*)) \right\}^2 \right] \quad \text{for } d = d'
\end{align*}
\]
Case 2: $d \neq d'$

$$E\left[I(t_{ij} \in (t_{d-1}, t_d])I(t_{i'j'} \in (t_{d'-1}, t_{d'}])\right] = P(t_{ij} \in (t_{d-1}, t_d]\text{ and } t_{i'j'} \in (t_{d'-1}, t_{d'}])$$

$$= \begin{cases} 
P(t_{ij} \in (t_{d-1}, t_d])P(t_{i'j'} \in (t_{d'-1}, t_{d'}]) = P_dP_{d'} & \text{for } i \neq i', j, j' \\
P(t_{ij} \in (t_{d-1}, t_d]\text{ and } t_{i'j'} \in (t_{d'-1}, t_{d'}]) = P_dP_{d'} & \text{for } i = i', j \neq j' \ (t_{ij} \perp t_{i'j'}) \\
0 & \text{for } i = i' \text{ and } j = j'
\end{cases}$$

For the last case ($i = i'$ and $j = j'$), $P(t_{ij} \in (t_{d-1}, t_d] \text{ and } t_{i'j'} \in (t_{d'-1}, t_{d'}])$ is equal to 0 because it is not possible that one subject is observed at two different visit times, $t_d$ and $t_{d'}$, with the same index for visit, $j = j'$.

$$E\left[A_i(t_d)A_{i'}(t_{d'})\right] = E\left[Y_i(s, t_d)Y_{i'}(s', t_d)Y_i(s, t_{d'})Y_{i'}(s', t_{d'})\right]$$

$$= \begin{cases} 
E\left[Y_i(s, t_d)Y_{i'}(s, t_d)\right]E\left[Y_{i'}(s, t_{d'})Y_{i'}(s', t_{d'})\right] \\
= c((s, t_d), (s', t_d)) \cdot c((s, t_{d'}), (s', t_{d'})) & \text{for } i \neq i', j, j' \\
& (\text{because } Y_i(s, t_d) = X_i(s, t_d)) \\
E\left[Y_i(s, t_d)Y_{i'}(s', t_d)Y_i(s, t_{d'})Y_{i'}(s', t_{d'})\right] & \text{for } i = i', j \neq j' \\
0 & \text{for } i = i', j = j'
\end{cases}$$
By summing over \( i, i', j, \) and \( j' \), Equation (B.3) for the case of \( d \neq d' \) equals to

\[
E[V_dV_{d'}] = \left( \frac{D}{M_1} \right)^2 \left\{ \sum_{i=1}^{n} m_i(M_1 - m_i) \right\} \left\{ P_d \cdot P_{d'} \cdot c((s, t_d), (s', t_d')) \cdot c((s, t_{d'}), (s', t_{d'})) \right\}
\]
\[
+ \left( \frac{D}{M_1} \right)^2 \left\{ \sum_{i=1}^{n} m_i(m_i - 1) \right\} \left\{ P_d \cdot P_{d'} \cdot E \left[ Y_i(s, t_d)Y_i(s', t_d)Y_i(s, t_{d'})Y_i(s', t_{d'}) \right] \right\}
\]
\[
= \left( D^2 - \frac{M_2D^2}{M_1^2} \right) \left\{ P_d \cdot P_{d'} \cdot c((s, t_d), (s', t_d)) \cdot c((s, t_{d'}), (s', t_{d'})) \right\}
\]
\[
+ \left( \frac{M_2D^2}{M_1^2} - \frac{D^2}{M_1} \right) \left\{ P_d \cdot P_{d'} \cdot E \left[ Y_i(s, t_d)Y_i(s', t_d)Y_i(s, t_{d'})Y_i(s', t_{d'}) \right] \right\} \text{ for } d \neq d'
\]

In summary, we have

\[
E[V_d] = D \cdot c((s, t_d), (s', t_d)) \cdot P_d \quad \text{(because } Y_i(s, t) = X_i(s, t))
\]

and

\[
E[V_dV_{d'}] =
\begin{cases}
\left\{ D^2 - \frac{M_2D^2}{M_1^2} \right\} \left\{ P_d \cdot c((s, t_d), (s', t_d)) \right\}^2 \\
+ \left\{ D^2 \right\} \left\{ P_d \cdot E \left[ Y_i(s, t_d)Y_i(s', t_d)Y_i(s, t_{d'})Y_i(s', t_{d'}) \right] \right\} & \text{for } d = d'
\end{cases}
\]

\[
E[V_dV_{d'}] =
\begin{cases}
\left\{ D^2 - \frac{M_2D^2}{M_1^2} \right\} \left\{ P_dP_{d'} \cdot c((s, t_d), (s', t_d)) \cdot c((s, t_{d'}), (s', t_{d'})) \right\} \\
+ \left\{ \frac{M_2D^2}{M_1^2} - \frac{D^2}{M_1} \right\} \left\{ P_dP_{d'} \cdot E \left[ Y_i(s, t_d)Y_i(s', t_d)Y_i(s, t_{d'})Y_i(s', t_{d'}) \right] \right\} & \text{for } d \neq d'
\end{cases}
\]

129
Finally we obtain the covariance of $V_d$ and $V_{d'}$; that is,

$$
cov(V_d, V_{d'}) = E[V_d V_{d'}] - E[V_d]E[V_{d'}]
$$

$$
= \begin{cases} 
\left\{ \frac{D^2}{M_1} \right\} \left\{ P_d \cdot E\left[ Y_i(s, t_d)Y_i(s', t_d)Y_i(s, t_{d'})Y_i(s', t_{d'}) \right] \right\} \\
\quad \quad - \left\{ \frac{M_2 D^2}{M_1^2} \right\} \left\{ P_d \cdot c((s, t_d), (s', t_d)) \right\}^2 & \text{for } d = d' \\
\left\{ \frac{M_2 D^2}{M_1^2} - \frac{D^2}{M_1} \right\} \left\{ P_{d'P_{d'}} \cdot E\left[ Y_i(s, t_d)Y_i(s', t_d)Y_i(s, t_{d'})Y_i(s', t_{d'}) \right] \right\} \\
\quad \quad - \left\{ \frac{M_2 D^2}{M_1^2} \right\} \left\{ P_{d'P_{d'}} \cdot c((s, t_d), (s', t_d)) \cdot c((s, t_{d'}), (s', t_{d'})) \right\}
\end{cases}
$$

for $d \neq d'$.

For simplicity in notation, denote the variance of $V_d$ with $\sigma_d^2$, and the covariance of $V_d$ and $V_{d'}$ with $\sigma_{d,d'}$. Under the assumptions (A1.) - (A3.), $E[V_d]$, $\sigma_d^2$ and $\sigma_{d,d'}$ are finite. Following theorem 5.3 of Boos & Stefanski (2013, p.208), we show the pointwise consistency of $\hat{\Sigma}(s, s') = \nabla$ by proving that the following holds;

$$
\frac{1}{D^2} \left[ \sum_{d=1}^D \sigma_d^2 + 2 \sum_{d=1}^{D-1} \sum_{d'=d+1}^D \sigma_{d,d'} \right] = \frac{1}{D^2} \left[ \sum_{d=1}^D \sigma_d^2 + \sum_{d=1}^{D} \sum_{d'=d+1}^D \sigma_{d,d'} \right] \overset{p}{\to} 0 \quad \text{as } D \to \infty.
$$

By plugging in $cov(V_d, V_{d'})$ that we obtained earlier, we get the first term, $\frac{1}{D^2} \sum_{d=1}^D \sigma_d^2$, equal to

$$
\frac{1}{D^2} \sum_{d=1}^D \left\{ \frac{D^2}{M_1} \right\} \left\{ P_d \cdot E\left[ Y_i(s, t_d)Y_i(s', t_d)Y_i(s, t_{d'})Y_i(s', t_{d'}) \right] \right\} \\
- \left\{ \frac{D^2 M_2}{M_1^2} \right\} \left\{ P_d \cdot c((s, t_d), (s', t_d)) \right\}^2
$$
\[
\sum_{d=1}^{D} \left\{ P_d \cdot E \left[ Y_i(s, t_d) Y_i(s', t_d') Y_i(s, t_d) Y_i(s', t_d') \right] \right\}
\]

and the second term, \(\frac{1}{D^2} \sum_{d=1}^{D} \sum_{d' \neq d} \sigma_{dd'}\), equal to

\[
\frac{1}{D^2} \sum_{d=1}^{D} \sum_{d' \neq d} \left\{ \frac{M_2 D^2}{M_1^2} - \frac{D^2}{M_1} \right\} \left\{ P_d P_{d'} \cdot E \left[ Y_i(s, t_d) Y_i(s', t_d) Y_i(s, t_{d'}) Y_i(s', t_{d'}) \right] \right\}
\]

\[
- \left\{ \frac{M_2 D^2}{M_1^2} \right\} \left\{ P_d P_{d'} \cdot c((s, t_d), (s', t_{d'})) \cdot c((s, t_{d'}), (s', t_d)) \right\}
\]

\[
= \frac{1}{D^2} \sum_{d=1}^{D} \sum_{d' \neq d} \left\{ \frac{M_2 D^2}{M_1^2} - \frac{D^2}{M_1} \right\} \left\{ P_d P_{d'} \cdot E \left[ Y_i(s, t_d) Y_i(s', t_d) Y_i(s, t_{d'}) Y_i(s', t_{d'}) \right] \right\}
\]

\[
- \frac{1}{D^2} \sum_{d=1}^{D} \sum_{d' \neq d} \left\{ \frac{M_2 D^2}{M_1^2} \right\} \left\{ P_d P_{d'} \cdot c((s, t_d), (s', t_{d'})) \cdot c((s, t_{d'}), (s', t_d)) \right\}, \quad (B.5)
\]

where

\[
\text{Term (B.5)} = \left\{ \frac{M_2}{M_1^2} - \frac{1}{M_2} \right\} \sum_{d=1}^{D} \sum_{d' \neq d} P_d P_{d'} \cdot E \left[ Y_i(s, t_d) Y_i(s', t_d) Y_i(s, t_{d'}) Y_i(s', t_{d'}) \right]
\]

\[
= \left\{ \frac{M_2}{M_1^2} - \frac{1}{M_1} \right\} \sum_{d=1}^{D} P_d \left[ \sum_{d'=1}^{D} P_{d'} E \left[ Y_i(s, t_d) Y_i(s', t_d) Y_i(s, t_{d'}) Y_i(s', t_{d'}) \right] \right]
\]

\[
- P_d E \left[ Y_i(s, t_d) Y_i(s', t_d) Y_i(s, t_{d'}) Y_i(s', t_{d'}) \right]
\]
\[
\begin{align*}
&= \left\{ \frac{M_2}{M_1^2} - \frac{1}{M_1} \right\} \sum_{d=1}^{D} P_d \cdot E_t[E[A(t)A(t')|t = t_d, t' = t_{d'}]] \\
&\quad - \left\{ \frac{M_2}{M_1^2} - \frac{1}{M_1} \right\} \sum_{d=1}^{D} P_d^2 E[A(t)^2|t = t_d] \\
&= \left\{ \frac{M_2}{M_1^2} - \frac{1}{M_1} \right\} E_t[E[A(t)A(t')|t, t']] \\
&\quad - \left\{ \frac{M_2}{M_1^2} - \frac{1}{M_1} \right\} \sum_{d=1}^{D} P_d^2 \cdot E[A(t)^2|t = t_d],
\end{align*}
\]

and

\[
\text{Term (B.6)} = - \frac{M_2}{M_1^2} \sum_{d=1}^{D} \sum_{d' \neq d} \left\{ P_d P_{d'} \cdot c((s, t_d), (s', t_d)) \cdot c((s, t_{d'}), (s', t_{d'})) \right\}
\]
\[
= - \frac{M_2}{M_1^2} \sum_{d=1}^{D} P_d \cdot c((s, t_d), (s', t_d)) \sum_{d' \neq d} P_{d'} \cdot c((s, t_{d'}), (s', t_{d'}))
\]
\[
= - \frac{M_2}{M_1^2} \sum_{d=1}^{D} P_d \cdot c((s, t_d), (s', t_d)) \times \left[ \sum_{d'=1}^{D} P_{d'} \cdot c((s, t_{d'}), (s', t_{d'})) - P_d \cdot c((s, t_d), (s', t_d)) \right]
\]
\[
= - \frac{M_2}{M_1^2} \{\Sigma(s, s')\}^2 + \frac{M_2}{M_1^2} \sum_{d=1}^{D} \{ P_d \cdot c((s, t_d), (s', t_d)) \}^2
\]

By combining everything together, we have

\[
\frac{1}{D^2} \left[ \sum_{d=1}^{D} \sigma_d^2 + \sum_{d=1}^{D} \sum_{d' \neq d} \sigma_{dd'} \right] =
\]
\[
\left[ \frac{1}{M_1} E_t[E[A(t)^2|t]] \right] + \left\{ \frac{M_2}{M_1^2} - \frac{1}{M_1} \right\} E_t[E_{t'}[E[A(t)A(t')|t, t']]]
\]

132
\[
- \left[ \left\{ \frac{M_2}{M_1^2} - \frac{1}{M_1} \right\} \sum_{d=1}^{D} P_d^2 \cdot \mathbb{E}[A(t)^2|t = t_d] \right] - \frac{M_2}{M_1^2} \Sigma(s, s')^2. \tag{B.7}
\]

As discussed before, to prove the pointwise consistency of \(\hat{\Sigma}(s, s')\) it is sufficient to show that Equation (B.7) converges to 0 as \(D\) diverges; we show this by showing each of the terms in Equation (B.7) converges to zero as \(D\) diverges.

First notice that \(M_1\) diverges with \(D\) because \(D \leq M_1\) and \(M_1/M_2 = O(1)\). And recall that here we are considering the case when \(Y(s, t) = X(s, t)\), and thus \(Y(s, t)\) satisfies all of the assumptions made on \(X(s, t)\). Because integration of a continuous function in a compact interval is finite, under the assumptions (A2.) and (A3.) it is easy to show that \(\mathbb{E}_t[\mathbb{E}[A(t)^2|t]]\), \(\mathbb{E}_t[\mathbb{E}[A(t)A'(t)|t, t']]\), and \(\Sigma(s, s')\) are finite. For example \(\mathbb{E}_t[\mathbb{E}[A(t)^2|t]] = \int_\mathcal{T} g(t) \mathbb{E}[A(t)^2|t]dt \leq \sup_t g(t) \int_\mathcal{T} \mathbb{E}[A(t)^2|t]dt\) is finite because (i) by the assumption (A2.) \(\sup_t g(t)\) is finite and (ii) \(\int_\mathcal{T} \mathbb{E}[A(t)^2|t]dt\) is finite because by the assumption (A3.) \(\mathbb{E}[A(t)^2|t]\) is continuous and finite in the compact interval \(\mathcal{T} = [0, 1]\).

And with the same reasoning we can also show that \(\sum_{d=1}^{D} P_d^2 \cdot \mathbb{E}[A(t)^2|t = t_d]\) is finite because \(\sum_{d=1}^{D} P_d^2 \cdot \mathbb{E}[A(t)^2|t = t_d] < \sup_d P_d \mathbb{E}_t[\mathbb{E}[A(t)^2|t]]\). Thus we show that each of the terms in Equation (B.7) converges to zero as \(D\) diverges.

It follows that

\[
\frac{1}{D^2} \left[ \sum_{d=1}^{D} \sigma_d^2 + \sum_{d=1}^{D} \sum_{d' \neq d} \sigma_{dd'} \right] \overset{p}{\to} 0 \text{ as } D \to \infty,
\]

and by theorem 5.3 (Boos & Stefanski, 2013), \(\hat{\Sigma}(s, s') = \overline{V} = D^{-1} \sum_{d=1}^{D} V_d\) converges in probability to \(D^{-1}\sum_{d=1}^{D} \mathbb{E}[V_d] = \sum_{d=1}^{D} P_d \cdot c((s, t_d), (s', t_d)) = \Sigma(s, s')\) as \(n\) diverges; recall that we assume \(D\) diverges with the sample size \(n\).

---

(ii) The Hilbert-Schmidt norm consistency
Because the Hilbert-Schmidt norm consistency is implied by the convergence of $E\left[ \| \hat{\Sigma}(s, s') - \Sigma(s, s') \|_s^2 \right]$ using Markov inequality, it is sufficient to show that

$$E\left[ \| \hat{\Sigma}(s, s') - E[\hat{\Sigma}(s, s')] \|_s^2 \right] = E\left[ \| \hat{\Sigma}(s, s') - \Sigma(s, s') \|_s^2 \right]$$

converges to 0 as $n$ diverges and we prove it by showing that the upper bound given in Equation (B.10) converges to 0 as $n$ diverges.

The sample covariance operator associated with $\hat{\Sigma}(s, s')$ is

$$H(x) = D^{-1} \sum_{d=1}^{D} \frac{D}{M_1} \sum_{i=1}^{n} \sum_{j=1}^{m_i} < Y_{ij}, x > Y_{ij} I(t_{ij} \in (t_{d-1}, t_d))$$

$$= D^{-1} \sum_{d=1}^{D} \frac{D}{M_1} \sum_{i=1}^{n} \sum_{j=1}^{m_i} < Y_i(\cdot, t_d), x > Y_i(\cdot, t_d) I(t_{ij} \in (t_{d-1}, t_d)), \quad x \in L^2$$

and using the covariance operator, $H(\cdot)$, $E\left[ \| \hat{\Sigma}(s, s') - E[\hat{\Sigma}(s, s')] \|_s^2 \right]$ can be written as

$$E\left[ \| \hat{\Sigma}(s, s') - E[\hat{\Sigma}(s, s')] \|_s^2 \right] = E\left[ \| H - E[H] \|_s^2 \right] = E\left[ \sum_{v=1}^{\infty} \| H(e_v) - E[H(e_v)] \|_s^2 \right]$$

$$= E\left[ \sum_{v=1}^{\infty} \left\| D^{-1} \sum_{d=1}^{D} \frac{D}{M_1} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \{ < Y_{ij}, e_v > Y_{ij} I(t_{ij} \in (t_{d-1}, t_d)) \} \right\|_s^2 \right],$$

where $\{e_v : v \geq 1\}$ is any orthonormal basis (Horváth & Kokoszka, 2012, p.22).

Let $R_{ijdv} = < Y_{ij}, e_v > Y_{ij} I(t_{ij} \in (t_{d-1}, t_d)) = < Y_i(\cdot, t_d), e_v > Y_i(\cdot, t_d) I(t_{ij} \in (t_{d-1}, t_d))$. Then it follows that

$$E\left[ \| \hat{\Sigma}(s, s') - E[\hat{\Sigma}(s, s')] \|_s^2 \right] = E\left[ \| H - E[H] \|_s^2 \right]$$

$$= E\left[ \sum_{v=1}^{\infty} \left\| D^{-1} \sum_{d=1}^{D} \frac{D}{M_1} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \{ R_{ijdv} - E[R_{ijdv}] \} \right\|_s^2 \right]$$

$$= E\left[ \sum_{v=1}^{\infty} D^{-2} \sum_{d=1}^{D} \sum_{d'=1}^{D} \frac{D^2}{M_1^2} < \sum_{i}^{n} \sum_{j}^{m_i} \{ R_{ijdv} - E[R_{ijdv}] \}, \sum_{i'}^{n} \sum_{j'}^{m_i'} \{ R_{i'j'd'v} - E[R_{i'j'd'v}] \} > \right].$$
\[ D^{-2} \sum_{d=1}^{D} \sum_{d' = 1}^{D} \frac{D^2}{M_1^2} \left[ \sum_{i}^{n} \sum_{j}^{m_\epsilon} \sum_{j'}^{m_\epsilon} \sum_{v=1}^{\infty} E \left( R_{ijdv} - E[R_{ijdv}], R_{i'j'd'v} - E[R_{i'j'd'v}] \right) \right] \]

\[ = D^{-2} \sum_{d=1}^{D} \sum_{d' = 1}^{D} Q_{dd'} \]

where we define

\[ Q_{dd'} = \sum_{i}^{n} \sum_{j}^{m_\epsilon} \sum_{j'}^{m_\epsilon} \sum_{v=1}^{\infty} E \left( R_{ijdv} - E[R_{ijdv}], R_{i'j'd'v} - E[R_{i'j'd'v}] \right) \]

\[ = \sum_{i}^{n} \sum_{j}^{m_\epsilon} \sum_{j'}^{m_\epsilon} \sum_{v=1}^{\infty} \left( E[R_{ijdv}, R_{i'j'd'v}] - E[R_{ijdv}, E[R_{i'j'd'v}]] \right). \]

Furthermore because \( Q_{dd'} \leq \left| Q_{dd'} \right| \) and \( |a + b| \leq |a| + |b| \) for any \( a, b \in \mathbb{R}, \)

\[ 0 \leq E \left[ \| H - E[H] \|_2^2 \right] = D^{-2} \sum_{d=1}^{D} \sum_{d' = 1}^{D} Q_{dd'} \leq D^{-2} \sum_{d=1}^{D} \sum_{d' = 1}^{D} |Q_{dd'}| \]

\[ = M_1^{-2} \sum_{d=1}^{D} \sum_{d' = 1}^{D} \left| \sum_{i}^{n} \sum_{j}^{m_\epsilon} \sum_{j'}^{m_\epsilon} \sum_{v=1}^{\infty} \left( E[R_{ijdv}, R_{i'j'd'v}] - E[R_{ijdv}, E[R_{i'j'd'v}]] \right) \right| \]

\[ \leq M_1^{-2} \sum_{d=1}^{D} \sum_{d' = 1}^{D} \sum_{i}^{n} \sum_{j}^{m_\epsilon} \sum_{j'}^{m_\epsilon} \sum_{v=1}^{\infty} \left( E[R_{ijdv}, R_{i'j'd'v}] - E[R_{ijdv}, E[R_{i'j'd'v}]] \right). \]

As discussed, we show the convergence of \( E \left[ \| \hat{\Sigma}(s, s') - E[\hat{\Sigma}(s, s')] \|_2^2 \right] = M_1^{-2} \sum_{d=1}^{D} \sum_{d' = 1}^{D} Q_{dd'} \) by showing that the upper bound given in Equation (B.10) converges to 0. In the following we first simplify the summand given in Equation (B.10) and then study the sum.

Let \( G_{id}(x) = \langle Y_i(\cdot, t_d), x \rangle > Y_i(\cdot, t_d) \) for \( x \in L_2 \). Notice that the summand in Equation (B.9), \( E[R_{ijdv}, R_{i'j'd'v}] - E[R_{ijdv}, E[R_{i'j'd'v}]] \), can be rewritten as

\[ E[R_{ijdv}, R_{i'j'd'v}] - E[R_{ijdv}, E[R_{i'j'd'v}]] \]

\[ = E \left[ I(t_{ij} \in (t_{d-1}, t_d]) \cdot I(t_{i'j'} \in (t_{d'-1}, t_{d'}]) \right] \times E \left[ \langle G_{id}(e_v), G_{i'd'}(e_v) \rangle \right] \]
\[ - E \left[ I(t_{ij} \in (t_{d-1}, t_d)) \right] \cdot E \left[ I(t_{i'j'} \in (t_{d'-1}, t_{d'})) \right] \times \left[ E \left[ G_{id}(e_v) \right] , \ E \left[ G_{d'}(e_v) \right] \right] \]

(B.11)

because of the following:

\[
E[< R_{ijdv}, R_{i'j'd'v} >] = E \left[ \langle < Y_i(\cdot, t_d), e_v > Y_i(\cdot, t_d)I(t_{ij} \in (t_{d-1}, t_d)), < Y_{i'}(\cdot, t_{d'}), e_v > Y_{i'}(\cdot, t_{d'})I(t_{i'j'} \in (t_{d'-1}, t_{d'})) \rangle \right] \\
\quad = E \left[ I(t_{ij} \in (t_{d-1}, t_d)) \cdot I(t_{i'j'} \in (t_{d'-1}, t_{d'})) \right] \\
\quad \quad \times E \left[ \langle < Y_i(\cdot, t_d), e_v > Y_i(\cdot, t_d), < Y_{i'}(\cdot, t_{d'}), e_v > Y_{i'}(\cdot, t_{d'}) \rangle \right] \\
\]

and

\[
<E[R_{ijdv}], E[R_{i'j'd'v}] > = E \left[ \langle < Y(\cdot, t_d), e_v > Y(\cdot, t_d)I(t_{ij} = t_d) \rangle , \ E \left[ < Y(\cdot, t_{d'}), e_v > Y(\cdot, t_{d'})I(t_{i'j'} \in (t_{d'-1}, t_{d'})) \right] \right] \\
\quad = E \left[ < Y(\cdot, t_d), e_v > Y(\cdot, t_d)I(t_{ij} \in (t_{d-1}, t_d)) \right] \cdot E \left[ I(t_{ij} \in (t_{d-1}, t_d)) \right] , \ E \left[ < Y(\cdot, t_{d'}), e_v > Y(\cdot, t_{d'})I(t_{i'j'} \in (t_{d'-1}, t_{d'})) \right] \times E \left[ I(t_{i'j'} \in (t_{d'-1}, t_{d'})) \right] \\
\quad = E \left[ I(t_{ij} \in (t_{d-1}, t_d)) \right] \cdot E \left[ I(t_{i'j'} \in (t_{d'-1}, t_{d'})) \right] \times E \left[ < Y(\cdot, t_d), e_v > Y(\cdot, t_d) \right] , \ E \left[ < Y(\cdot, t_{d'}), e_v > Y(\cdot, t_{d'}) \right] \right] \right]
\]

In the following we consider Equation (B.10) for two cases, when \(d = d'\) and when
$d \neq d'$.

**Case 1: $d = d'$**

This case is equivalent to the case of $t_d = t_d'$ as $t_d$’s are unique values. Let $t^* = t_d = t_d'$.

First observe the following:

$$
E\left[ I(t_{ij} \in (t_{d-1}, t_d)) \cdot I(t'_{ij'} \in (t'_{d-1}, t_d')) \right] = \begin{cases} 
P_d^2 & \text{for } i \neq i', j, j' \\
0 & \text{for } i = i', j \neq j' \\
P_d & \text{for } i = i', j = j' 
\end{cases}
$$

$$
E\left[ I(t_{ij} \in (t_{d-1}, t_d)) \right] \cdot E\left[ I(t'_{ij'} \in (t'_{d-1}, t_d')) \right] = P_d^2 \text{ for all } i, i', j, j'
$$

**Case 1(a): $d = d', i \neq i', j, j'$ (Using independence)**

$$
E[R_{ijd'}, R_{i'j'd'}] - E[R_{ijdv}] \cdot E[R_{i'j'd'}] \quad \text{from Equation (B.11)}
$$

$$
= P_d^2 \cdot \left\{ E\left[ \langle G_{id}(e_v), G_{i'd'}(e_v) \rangle \right] - \left\langle E\left[ G_{id}(e_v) \right], E\left[ G_{i'd'}(e_v) \right] \right\rangle \right\}
$$

$$
= P_d^2 \cdot \left\{ E\left[ \langle G_{id}(e_v), G_{i'd}(e_v) \rangle \right] - \left\langle E\left[ G_{id}(e_v) \right], E\left[ G_{i'd}(e_v) \right] \right\rangle \right\}
$$

$$
= P_d^2 \cdot \left\{ E\left[ \int G_{id}(e_v)(s) \cdot G_{i'd}(e_v)(s) \, ds \right]
\quad - \langle E\left[ G_{id}(e_v) \right], E\left[ G_{i'd}(e_v) \right] \rangle \right\}
$$

137
\[
= P_d^2 \cdot \left\{ \mathbb{E} \left[ \int < Y_i(\cdot, t_d), e_v > Y_i(s, t_d) < Y_{i'}(\cdot, t_d), e_v > Y_{i'}(s, t_d) \, ds \right] - \right.
\]
\[
\left. \mathbb{E} \left[ G_{id}(e_v) \right], \mathbb{E} \left[ G_{i'd}(e_v) \right] \right\}
\]
\[
= P_d^2 \cdot \left\{ \int \mathbb{E} \left[ < Y_i(\cdot, t_d), e_v > Y_i(s, t_d) \right] \mathbb{E} \left[ < Y_{i'}(\cdot, t_d), e_v > Y_{i'}(s, t_d) \right] \, ds - \right.
\]
\[
\left. \mathbb{E} \left[ G_{id}(e_v) \right], \mathbb{E} \left[ G_{i'd}(e_v) \right] \right\}
\]

(by commutative property of expectation with bounded operators)

(Horváth & Kokoszka, 2012, p.23)

\[
= P_d^2 \cdot \left\{ \left\langle \mathbb{E} \left[ G_{id}(e_v) \right], \mathbb{E} \left[ G_{i'd}(e_v) \right] \right\rangle - \right. \]
\[
\left. \left\langle \mathbb{E} \left[ G_{id}(e_v) \right], \mathbb{E} \left[ G_{i'd}(e_v) \right] \right\rangle \right\} = 0
\]

(by independence)

Case 1(b): \( d = d', i = i', j \neq j' \)

\[
\mathbb{E}[ < R_{ijd'v}, R_{i'j'd'v} > ] - \mathbb{E}[R_{ijd'v}, E[R_{i'j'd'v}]] >
\]
\[
= 0 \cdot \mathbb{E} \left[ \left( G_{id}(e_v), G_{i'd}(e_v) \right) \right] - P_d^2 \cdot \left\langle \mathbb{E} \left[ G_{id}(e_v) \right], \mathbb{E} \left[ G_{i'd}(e_v) \right] \right\rangle
\]
\[
= -P_d^2 \cdot \left\langle \mathbb{E} \left[ G_{id}(e_v) \right], \mathbb{E} \left[ G_{id}(e_v) \right] \right\rangle
\]
\[
= -P_d^2 \cdot \left\| \mathbb{E} \left[ G_{id}(e_v) \right] \right\|^2
\]
\[
= -P_d^2 \cdot \left\| \mathbb{E} \left[ < Y(\cdot, t_d), e_v > Y(\cdot, t_d) \right] \right\|^2
\]

138
Because $P_d^2$ and $\left\| E[<Y(\cdot, t_d), e_v > Y(\cdot, t_d)]\right\|^2$ are nonnegative, it is implied that

$$
\left| E[<R_{ijdv}, R_{i'j'd'v}>] - E[R_{ijdv}], E[R_{i'j'd'v}] \right| = P_d^2 \cdot \left\| E[<Y(\cdot, t_d), e_v > Y(\cdot, t_d)]\right\|^2 \\
\leq P_d^2 \cdot E[\left\| <Y(\cdot, t_d), e_v > Y(\cdot, t_d)\right\|^2]
$$

(by Jensen’s inequality; $\| \cdot \|^2$ is a convex function)

$$
= P_d^2 \cdot E[|<Y(\cdot, t_d), e_v |>^2 \cdot \|Y(\cdot, t_d)\|^2]
$$

It follows that Equation (B.10) for the case of $d = d', i \neq i', j, j'$ is equal to

$$
M_1^{-2} \sum_{d,d' = d} \sum_{i,i' = i} \sum_{j,j' \neq j} \sum_{v=1}^\infty \left\{ E[<R_{ijdv}, R_{i'j'd'v}>] - E[R_{ijdv}], E[R_{i'j'd'v}] \right\} \\
= M_1^{-2} \sum_{d,d' = d} \sum_{i,i' = i} \sum_{j,j' \neq j} \sum_{v=1}^\infty P_d^2 \cdot E[|<Y(\cdot, t_d), e_v |>^2 \cdot \|Y(\cdot, t_d)\|^2]
$$

$$
= M_1^{-2} \sum_{d,d' = d} \sum_{i,i' = i} \sum_{j,j' \neq j} P_d^2 \cdot E[\|Y(\cdot, t_d)\|^2 \cdot \|Y(\cdot, t_d)\|^2], \text{ by Parseval's identity}
$$

$$
= M_1^{-2} \sum_{d,d' = d} \sum_{i,i' = i} \sum_{j,j' \neq j} P_d^2 \cdot E[\|Y(\cdot, t_d)\|^4]
$$

$$
= \frac{M_2 - M_1}{M_1^2} \sum_{d=1}^P P_d^2 \cdot E[\|Y(\cdot, t_d)\|^4]
$$

$$
\leq \frac{(M_2 - M_1) \cdot \sup_d P_d}{M_1^2} \cdot E_t \left[ \|Y(\cdot, t_d)\|^4 \right]
$$

139
Case 1(c): $d = d', i = i', j = j'$

\[
E[< R_{ijdv}, R_{i'j'd'v} >] - E[R_{ijdv}], E[R_{i'j'd'v}] >
\]

\[
= P_d \cdot E\left(\left\langle G_{id}(e_v), G_{i'd'}(e_v)\right\rangle\right) - P_d^2 \cdot \left\langle E\left[ G_{id}(e_v) \right], E\left[ G_{i'd'}(e_v) \right] \right\rangle
\]

\[
= P_d \cdot E\left[\|G_{id}(e_v)\|^2\right] - P_d^2 \cdot \left\langle E\left[ G_{id}(e_v) \right] \right\rangle^2
\]

\[
\leq P_d \cdot E\left[\|< Y(\cdot, t_d), e_v > Y(\cdot, t_d)\|^2\right]
\]

(because $P_d^2$ and $\left\langle E\left[ G_{id}(e_v) \right] \right\rangle^2$ are nonnegative)

\[
= P_d \cdot E\left[\|< Y(\cdot, t_d), e_v > Y(\cdot, t_d)\|^2\right],
\]

Furthermore in this case,

\[
E[< R_{ijdv}, R_{i'j'd'v} >] - E[R_{ijdv}], E[R_{i'j'd'v}] >
\]

\[
= E[< R_{ijdv}, R_{ijdv} >] - E[R_{ijdv}], E[R_{ijdv}] >
\]

\[
= E[\|R_{ijdv}\|^2] - \left\langle E[R_{ijdv}] \right\rangle^2 = E\left[\|R_{ijdv} - E[R_{ijdv}]\|^2\right] \geq 0
\]

Thus it follows that Equation (B.10) for the case of $d = d', i = i', j = j'$ is equal to

\[
M_1^{-2} \sum_{d, d' = d} \sum_{i, i' = i} \sum_{j, j' = j} \sum_{v = 1}^{\infty} \left\{ E[< R_{ijdv}, R_{i'j'd'v} >] - E[R_{ijdv}], E[R_{i'j'd'v}] > \right\}
\]

\[
= M_1^{-2} \sum_{d, d' = d} \sum_{i, i' = i} \sum_{j, j' = j} \sum_{v = 1}^{\infty} \left\{ E[< R_{ijdv}, R_{i'j'd'v} >] - E[R_{ijdv}], E[R_{i'j'd'v}] > \right\}
\]

(because it is shown that the summand in this case is always nonnegative)

\[
\leq M_1^{-2} \sum_{d, d' = d} \sum_{i, i' = i} \sum_{j, j' = j} \sum_{v = 1}^{\infty} P_d \cdot E\left[\|< Y(\cdot, t_d), e_v > Y(\cdot, t_d)\|^2\right]
\]

\[
= M_1^{-2} \sum_{d, d' = d} \sum_{i, i' = i} \sum_{j, j' = j} \sum_{v = 1}^{\infty} P_d \cdot E\left[\|Y(\cdot, t_d)\|^2 \cdot \sum_{v = 1}^{\infty} |< Y(\cdot, t_d), e_v >|^2\right]
\]
\[
= M_1^{-2} \sum_{d, d' = 1}^D \sum_{i, i' = i, j, j' = j} P_d \cdot E\left[\|Y(\cdot, t_d)\|^4\right] \quad \text{(by Parseval’s identity)}
\]

\[
= M_1^{-1} \sum_{d=1}^D P_d \cdot E\left[\|Y(\cdot, t_d)\|^4\right] = \frac{1}{M_1} \cdot E_t \left[\|Y(\cdot, t_d)\|^4\right]
\]

**Case 2: \(d \neq d'\)**

First observe the following:

\[
E\left[I(t_{ij} \in (t_{d-1}, t_d)) \cdot I(t_{i'j'} \in (t_{d'-1}, t_{d'}))\right] = \begin{cases} P_d P_{d'} & \text{for } i \neq i', j, j' \\ P_d P_{d'} & \text{for } i = i', j \neq j' \\ 0 & \text{for } i = i', j = j' \end{cases}
\]

\[
E\left[I(t_{ij} \in (t_{d-1}, t_d))\right] \cdot E\left[I(t_{i'j'} \in (t_{d'-1}, t_{d'}))\right] = P_d P_{d'} \quad \text{for all } i, i', j, j'
\]

**Case 2(a): \(d \neq d', i \neq i', j, j'\) (Using independence)**

\[
E[< R_{ijdv}, R_{i'j'd'v} >] - E[R_{ijdv} \cdot E[R_{i'j'd'v}]]
\]

\[
= P_d P_{d'} \cdot \left\{ E\left[\langle G_{id}(e_v), G_{i'd'}(e_v) \rangle\right] - \left\langle E\left[G_{id}(e_v)\right], E\left[G_{i'd'}(e_v)\right]\right\rangle \right\}
\]

(by the same reasoning that we use for Case 1(a))
\[= P_d P_{d'} \cdot \left\{ \left\langle E\left[ G_{id}(e_v) \right], E\left[ G_{id'}(e_v) \right] \right\rangle - \left\langle E\left[ G_{id}(e_v) \right], E\left[ G_{id'}(e_v) \right] \right\rangle \right\} \]

\[= 0 \]

Case 2(b): \( d \neq d', i = i', j \neq j' \)

\[E[< R_{ijdv}, R_{i'd'j'} ] - E[R_{ijdv}], E[R_{i'd'j'} ] > = P_d P_{d'} \cdot \left\{ \left\langle E\left[ G_{id}(e_v) \right], E\left[ G_{id'}(e_v) \right] \right\rangle - \left\langle E\left[ G_{id}(e_v) \right], E\left[ G_{id'}(e_v) \right] \right\rangle \right\} \] (B.12)

Here we consider lower and upper bounds for each term in Equation (B.12). By Cauchy-Schwartz inequality,

\[L^{(1)}_{dd'v} \leq P_d P_{d'} \cdot \left\langle E\left[ G_{id}(e_v) \right], E\left[ G_{id'}(e_v) \right] \right\rangle \leq U^{(1)}_{dd'v}, \] (B.13)

where \( U^{(1)}_{dd'v} = P_d P_{d'} \cdot \left\lVert E\left[ < Y(\cdot, t_d), e_v > Y(\cdot, t_d) \right] \right\rVert \cdot \left\lVert E\left[ < Y(\cdot, t_{d'}), e_v > Y(\cdot, t_{d'}) \right] \right\rVert \), and

\[L^{(1)}_{dd'v} = -U^{(1)}_{dd'v}. \] Similarly, by Cauchy-Schwartz inequality, the second term in Equation (B.12) is bounded below and above:

\[L^{(2)}_{dd'v} \leq P_d P_{d'} \cdot \left\langle E\left[ G_{id}(e_v) \right], E\left[ G_{id'}(e_v) \right] \right\rangle \leq U^{(2)}_{dd'v}, \] (B.14)

where \( U^{(2)}_{dd'v} = P_d P_{d'} \cdot \left\lVert E\left[ < Y(\cdot, t_d), e_v > Y(\cdot, t_d) \right] \right\rVert \cdot \left\lVert E\left[ < Y(\cdot, t_{d'}), e_v > Y(\cdot, t_{d'}) \right] \right\rVert \), and

\[L^{(2)}_{dd'v} = -U^{(2)}_{dd'v}. \]

It follows that, for the case of \( d \neq d', i = i', j \neq j' \),

\[L_{dd'v} \leq P_d P_{d'} \cdot \left\{ \left\langle E\left[ G_{id}(e_v) \right], E\left[ G_{id'}(e_v) \right] \right\rangle - \left\langle E\left[ G_{id}(e_v) \right], E\left[ G_{id'}(e_v) \right] \right\rangle \right\} \leq U_{dd'v}, \]
where
\[
U_{dd'v} = U_{dd'v}^{(1)} - L_{dd'v}^{(2)} = U_{dd'v}^{(1)} + U_{dd'v}^{(2)}, \text{ and} \\
L_{dd'v} = L_{dd'v}^{(1)} - U_{dd'v}^{(2)} = -U_{dd'v}^{(1)} - U_{dd'v}^{(2)} = -U_{dd'v}.
\]

As the lower and upper bounds has the same magnitude, i.e. \(U_{dd'v} = |L_{dd'v}|\), it follows that
\[
\left| E\left[<R_{ijdv}, R_{i'j'd'v}>\right] - E[R_{ijdv}], E[R_{i'j'd'v}] > \right| \leq U_{dd'v}.
\]

Furthermore using the fact that \((a^2 + b^2)/2 \geq ab\) for any \(a, b \in \mathbb{R}\),

\[
\sum_{v=1}^{\infty} U_{dd'v}^{(1)} = \sum_{v=1}^{\infty} P_d P_{d'} \cdot E\left[||G_{id}(e_v)|| \cdot ||G_{id'}(e_v)||\right] \\
= \sum_{v=1}^{\infty} P_d P_{d'} \cdot E\left[\left< Y_i(\cdot, t_d), e_v > Y_i(\cdot, t_{d'}) \right> \right] \cdot \left< Y_i(\cdot, t_d), e_v > Y_i(\cdot, t_{d'}) \right> \right] \\
\leq \frac{1}{2} \sum_{v=1}^{\infty} P_d P_{d'} \cdot E\left[|| Y_i(\cdot, t_d), e_v > Y_i(\cdot, t_d) ||^2 + || Y_i(\cdot, t_d'), e_v > Y_i(\cdot, t_d) ||^2 \right] \\
= \frac{1}{2} \sum_{v=1}^{\infty} P_d P_{d'} \left\{ E\left[|| Y_i(\cdot, t_d), e_v > Y_i(\cdot, t_d) ||^2 \right] + \\
E\left[|| Y_i(\cdot, t_d'), e_v > Y_i(\cdot, t_d) ||^2 \right] \right\} \\
= \frac{1}{2} P_d P_{d'} \left\{ E\left[|| Y(\cdot, t_d) ||^4 \right] + E\left[|| Y(\cdot, t_{d'}) ||^4 \right] \right\} \tag{B.15}
\]
(by Parseval’s identity)

\[
\sum_{v=1}^{\infty} U_{dd'v}^{(2)} = \sum_{v=1}^{\infty} P_d P_{d'} \cdot E\left[||G_{id}(e_v)|| \cdot ||G_{id'}(e_v)||\right] \\
\leq \frac{1}{2} \sum_{v=1}^{\infty} P_d P_{d'} \left\{ E\left[< Y_i(\cdot, t_d), e_v > Y_i(\cdot, t_{d}) \right> \right]^2 + \\
\sum_{v=1}^{\infty} P_d P_{d'} \cdot E\left[\left< Y_i(\cdot, t_d), e_v > Y_i(\cdot, t_{d'}) \right> \right] \right] \\
\leq \frac{1}{2} \sum_{v=1}^{\infty} P_d P_{d'} \left\{ E\left[|| Y(\cdot, t_d) ||^4 \right] + E\left[|| Y(\cdot, t_{d'}) ||^4 \right] \right\}
\]

143
Then Equation (B.10) for the case of \(d \neq d', i = i', j \neq j'\) is equal to

\[
M_1^{-2} \sum_{d,d' \neq d,i,i'=i,j,j' \neq j} \sum_{v=1}^{\infty} \left\{ E[<R_{ijdv},R_{i'j'd'v}>] - E[R_{ijdv}], E[R_{i'j'd'v}] \right\}
\]

\[
\leq M_1^{-2} \sum_{d,d' \neq d,i,i'=i,j,j' \neq j} \sum_{v=1}^{\infty} U_{dd'v} = M_1^{-2} \sum_{d,d' \neq d,i,i'=i,j,j' \neq j} \sum_{v=1}^{\infty} \left\{ U_{1}^{(1)} + U_{2}^{(2)} \right\}
\]

\[
\leq M_1^{-2} \sum_{d,d' \neq d,i,i'=i,j,j' \neq j} \sum_{d,i,i'=i,j,j' \neq j} \sum_{v=1}^{\infty} \left\{ U_{dd'v} \right\}
\]

\[
\leq M_2 - M_1 \left\{ \sum_{d,d' \neq d} P_d P_{d'} \cdot E[\|Y(\cdot, t_d)\|^4] + \sum_{d,d' \neq d} P_d P_{d'} \cdot E[\|Y(\cdot, t_{d'})\|^4] \right\}
\]

\[
= \frac{M_2 - M_1}{M_1^2} \cdot \sum_{d=1}^{D} P_d \cdot E[\|Y(\cdot, t_d)\|^4] \cdot \sum_{d' \neq d} P_{d'}
\]

\[
+ \frac{M_2 - M_1}{M_1^2} \cdot \sum_{d=1}^{D} P_d \cdot \left\{ \sum_{d'=1}^{D} P_{d'} \cdot E[\|Y(\cdot, t_{d'})\|^4] - P_d \cdot E[\|Y(\cdot, t_d)\|^4] \right\}
\]

\[
= \frac{M_2 - M_1}{M_1^2} \cdot E_d \left[ E[\|Y(\cdot, t_d)\|^4] \right] - \frac{M_2 - M_1}{M_1^2} \cdot \sum_{d=1}^{D} P_d^2 \cdot E[\|Y(\cdot, t_d)\|^4]
\]
\[ + \frac{M_2 - M_1}{M_1^2} \cdot E_t \left[ E \left[ \|Y(\cdot, t_d)\|^4 \right] \right] - \frac{M_2 - M_1}{M_1^2} \cdot \sum_{d=1}^{D} P_d^2 \cdot E \left[ \|Y(\cdot, t_d)\|^4 \right] \leq \frac{2(M_2 - M_1)}{M_1^2} \cdot E_t \left[ E \left[ \|Y(\cdot, t_d)\|^4 \right] \right] \]

Case 2(c): \( d \neq d', i = i', j = j' \)

\[
\begin{align*}
E[<R_{ijdv}, R_{i'j'd'v}>] - <E[R_{ijdv}], E[R_{i'j'd'v}]> &= 0 \cdot E \left[ \langle G_{id}(e_v), G_{id'}(e_v) \rangle \right] - P_d P_{d'} \cdot \left< E \left[ G_{id}(e_v) \right], E \left[ G_{id'}(e_v) \right] \right> \\
&= -P_d P_{d'} \cdot \left< E \left[ G_{id}(e_v) \right], E \left[ G_{id'}(e_v) \right] \right>
\end{align*}
\]

and it follows that

\[
\left| E[<R_{ijdv}, R_{i'j'd'v}>] - <E[R_{ijdv}], E[R_{i'j'd'v}]> \right| = P_d P_{d'} \cdot \left< E \left[ G_{id}(e_v) \right], E \left[ G_{id'}(e_v) \right] \right> \leq U_{dd'}^{(2)}
\]

(by Cauchy-Schwartz inequality)

then Equation (B.10) for the case of \( d \neq d', i = i', j = j' \) is equal to

\[
M_1^{-2} \sum_{d,d' \neq d} \sum_{i,i'=i} \sum_{j,j'=j} \sum_{v=1}^{\infty} \left| E[<R_{ijdv}, R_{i'j'd'v}>] - <E[R_{ijdv}], E[R_{i'j'd'v}]> \right| \leq M_1^{-2} \sum_{d,d' \neq d} \sum_{i,i'=i} \sum_{j,j'=j} \sum_{v=1}^{\infty} U_{dd'}^{(2)}
\]

\[
\leq M_1^{-2} \sum_{d,d' \neq d} \sum_{i,i'=i} \sum_{j,j'=j} \sum_{v=1}^{\infty} \frac{1}{2} P_d P_{d'} \cdot \left\{ E \left[ \|Y(\cdot, t_d)\|^4 \right] + E \left[ \|Y(\cdot, t_{d'})\|^4 \right] \right\}
\]

\[
= \frac{1}{2M_1} \left\{ \sum_{d,d' \neq d} P_d P_{d'} \cdot E \left[ \|Y(\cdot, t_d)\|^4 \right] + \sum_{d,d' \neq d} P_d P_{d'} E \left[ \|Y(\cdot, t_{d'})\|^4 \right] \right\}
\]
Finally by combining all cases, we have

$$0 \leq \mathbb{E} \left[ \left\| H - \mathbb{E}[H] \right\|^2 \right] = M_1^{-2} \sum_{d=1}^{D} \sum_{d'=1}^{D} Q_{dd'} \leq M_1^{-2} \sum_{d=1}^{D} \sum_{d'=1}^{D} |Q_{dd'}| \leq \left( \frac{M_2 - M_1}{M_1^2} \cdot \sup_d P_d + \frac{1}{M_1} + \frac{2(M_2 - M_1)}{M_1^2} + \frac{1}{M_1} \right) \cdot \mathbb{E}_t \left[ \left\| Y(\cdot, t_d) \right\|^4 \right]. \quad (B.17)$$

In the case of \( Y(s, t) = X(s, t) \) it is implied by the assumptions (A1.), (A2.), and (A4.) that \( \mathbb{E}_t \left[ \left\| Y(\cdot, t_d) \right\|^4 \right] \) is finite. Thus Equation (B.17) converges to 0 as \( n \) diverges. It follows the convergence of \( \mathbb{E} \left[ \left\| \hat{\Sigma}(s, s') - \Sigma(s, s') \right\|^2 \right] \) as well as the Hilbert-Schmidt norm consistency.

**Corollary 4.1.1** (Corollary 1 in Section 3.5.1). *Under the assumptions (A1.)-(A5.), for each \( k \) we have \( |\hat{\lambda}_k - \lambda_k| \xrightarrow{p} 0 \), and \( \| \hat{\phi}_k(\cdot) - \phi_k(\cdot) \|_s \xrightarrow{p} 0 \) as \( n \) diverges.*

**Proof.** We prove this corollary using the Hilbert-Schmidt norm consistency result obtained in theorem 1 of Section 3.5.1 and theorem 4.4 and Lemma 4.3 of Bosq (2000, p.104). For completeness the theorem we used is given below:

**Theorem 4.4 and Lemma 4.3** (Bosq, 2000, p.104) Let \( \lambda_j \) and \( v_j \) be eigenvalues and eigenfunctions of the operator \( C \). Let \( \lambda_{jn} \) and \( v_{jn} \) be the corresponding empirical eigen-elements, such that the operator \( C_n(v_{jn}) = \lambda_{jn} v_{jn} \). Let \( v_{jn}' = \text{sgn} < v_{jn}, v_j > v_j \), \( j \geq 1 \).

\[
\sup_{j \geq 1} |\lambda_{jn} - \lambda_j| \leq \| C_n - C \|_L \leq \| C_n - C \|_s \quad \text{eq. (4.43)}
\]

\[
\| v_{jn} - v_{jn}' \| \leq a_j \| C_n - C \|_L \quad \text{eq. (4.44)}
\]
where \( a_j = 2\sqrt{2} \max[(\lambda_{j-1} - \lambda_j)^{-1}, (\lambda_j - \lambda_{j+1})^{-1}] \) if \( j \geq 2 \), and \( a_1 = 2\sqrt{2}(\lambda_1 - \lambda_2)^{-1} \).

By theorem 4.4 of Bosq (2000, p.104),

\[
0 \leq \sup_{k \geq 1} |\hat{\lambda}_k - \lambda_k| \leq \|\hat{\Sigma}(s, s') - \Sigma(s, s')\|_L \leq \|\hat{\Sigma}(s, s') - \Sigma(s, s')\|_s, \quad \text{and} \quad (B.18)
\]

\[
0 \leq \|\hat{\phi}_k(s) - \phi_k(s)\|_s \leq 2\sqrt{2}a_k^{-1}\|\hat{\Sigma}(s, s') - \Sigma(s, s')\|_s, \quad (B.19)
\]

where \( a_k = (\lambda_1 - \lambda_2) \) for \( k = 1 \) and \( \max[(\lambda_{k-1} - \lambda_k), (\lambda_k - \lambda_{k+1})] \) otherwise. Thus under the assumption (A5.) the following consistency results are implied by the Hilbert-Schmidt norm consistency:

\[
|\hat{\lambda}_k - \lambda_k| \xrightarrow{p} 0 \quad \text{and} \quad \|\hat{\phi}_k(\cdot) - \phi_k(\cdot)\|_s \xrightarrow{p} 0, \quad \text{as} \quad n \to \infty \quad (B.20)
\]

---

**Theorem 4.1.2** (theorem 2 in Section 3.5.1). *Under the assumptions (A1.) - (A6.), for each \( k \) \( \sup_j |\tilde{\xi}_{W,ijk} - \xi_{W,ijk}| \xrightarrow{p} 0 \) and \( \|\hat{G}_k(\cdot, \cdot) - G_k(\cdot, \cdot)\|_s \xrightarrow{p} 0 \) as \( n \) diverges. In fact a stronger result also holds, namely \( \sup_t |\hat{G}_k(t, t') - G_k(t, t')| \xrightarrow{p} 0 \) as \( n \) diverges.*

**Proof.** We prove this theorem in two steps: we first show that (i) \( \sup_j |\tilde{\xi}_{W,ijk} - \xi_{W,ijk}| \xrightarrow{p} 0 \) and then we use the result to show that (ii) \( \sup_t |\hat{G}_k(t, t') - G_k(t, t')| \xrightarrow{p} 0 \), which implies the Hilbert-Schmidt norm consistency of \( \hat{G}_k(\cdot, \cdot) \).

(i) *The consistency of \( \tilde{\xi}_{W,ijk} \)

For each \( k \)

\[
\sup_j |\tilde{\xi}_{W,ijk} - \xi_{W,ijk}| \xrightarrow{p} 0, \quad (B.21)
\]

as \( n \) diverges, because

\[
\sup_j |\tilde{\xi}_{W,ijk} - \xi_{W,ijk}| = \sup_j \left| \int Y_i(s, t_{ij})\hat{\phi}_k(s)ds - \int Y_i(s, t_{ij})\phi_k(s)ds \right|
\]

\[
= \sup_j \left| \int Y_i(s, t_{ij})(\hat{\phi}_k(s) - \phi_k(s))ds \right|
\]

147
\[
\leq \sup_j \int |Y_i(s, t_{ij})| \{\hat{\phi}_k(s) - \phi_k(s)\} \, ds \\
\leq \sup_j \sup_{s \in [0,1]} |Y_i(s, t_{ij})| \int \{\hat{\phi}_k(s) - \phi_k(s)\} \, ds \\
\leq \sup_{j, s} |Y_i(s, t_{ij})| \times \left\{ \int \{\hat{\phi}_k(s) - \phi_k(s)\}^2 \, ds \right\}^{1/2} \\
\text{(by Cauchy-Schwartz ineq.)} \\
= \sup_{j, s} |Y_i(s, t_{ij})| \times \|\hat{\phi}_k(\cdot) - \phi_k(\cdot)\|_s, \\
\tag{B.22}
\]

where \(\sup_{j, s} |Y_i(s, t_{ij})|\) is absolutely bounded almost surely under the assumption \((A6.)\) and \(\|\hat{\phi}_k(\cdot) - \phi_k(\cdot)\|_s\) converges to 0 as obtained in Corollary 1.

\begin{enumerate}
\item \textit{The consistency of } \(\hat{G}_k(t, t')\)
\end{enumerate}

Recall that \(G_k(t, t') = \text{cov}(\xi_{il}(t), \xi_{ik}(t'))\) is the true covariance. It is already shown in Yao et al. (2005a) the uniform consistency of its local linear estimator, denoted by \(\hat{G}_{W,k}(t, t')\), when \(\hat{G}_{W,k}(t, t')\) is obtained with \(\xi_{W,ijk}\)'s; specifically, \(\hat{G}_{W,k}(t, t')\) is obtained by the local linear smoothing of \(\{\tilde{G}_{W,ik}(t_{ij}, t_{ij}') = \xi_{W,ijk} : j \neq j'\}\). And Yao et al. (2005a) showed that \(\hat{G}_{W,k}(t, t')\) is uniformly consistent to the true covariance function, \(G_k(t, t')\). Here we show a similar result when the local linear estimator is obtained with \(\tilde{\xi}_{W,ijk}\) instead of \(\xi_{W,ijk}\):

the sample covariance of \(\tilde{\xi}_{W,ijk}\)'s is as follows:

\[
\tilde{G}_{ik}(t_{ij}, t_{ij}') = \tilde{\xi}_{W,ijk} \tilde{\xi}_{W,ij'k} \\
= \{ (\tilde{\xi}_{W,ijk} - \xi_{W,ijk}) + \xi_{W,ijk} \} \times \{ (\tilde{\xi}_{W,ij'k} - \xi_{W,ij'k}) + \xi_{W,ij'k} \} \\
= \tilde{G}_{W,ik}(t_{ij}, t_{ij}') + (\tilde{\xi}_{W,ij'k} - \xi_{W,ij'k})(\tilde{\xi}_{W,ijk} - \xi_{W,ijk}) \\
+ \xi_{W,ij'k}(\tilde{\xi}_{W,ijk} - \xi_{W,ijk}) + \xi_{W,ijk}(\tilde{\xi}_{W,ij'k} - \xi_{W,ij'k})
\]

By the uniform consistency of \(\tilde{\xi}_{W,ijk}\) given in Equation (B.21), the local linear estimator,
$\hat{G}_k(t,t')$, obtained by smoothing $\hat{G}_{ik}(t_{ij},t_{ij}')$, is asymptotically equivalent to the local linear estimator, $\hat{G}_{W,k}(t,t')$, obtained by smoothing $\tilde{G}_{W,ik}(t_{ij},t_{ij}')$. Thus the uniform consistency of $\hat{G}_k(t,t')$ is implied by the uniform consistency of $\hat{G}_{W,k}(t,t')$ shown in Yao et al. (2005a). Finally the uniform consistency implies the Hilbert-Schmidt norm consistency.

**Corollary 4.1.2** (Corollary 2 in Section 3.5.1). Assume (A1.) - (A8.) hold for each $k$ and $l$. Then the eigenvalues $\hat{\eta}_{kl}$ and eigenfunctions $\hat{\psi}_{kl}(\cdot)$ of $\hat{G}_k(\cdot,\cdot)$ satisfy $|\hat{\eta}_{kl} - \eta_{kl}| \xrightarrow{p} 0$, and $\|\hat{\psi}_{kl}(\cdot) - \psi_{kl}(\cdot)\|_s \xrightarrow{p} 0$ as $n$ diverges. Uniform convergence of $\hat{\psi}_{kl}(\cdot)$ also holds: $\sup_t |\hat{\psi}_{kl}(t) - \psi_{kl}(t)| \xrightarrow{p} 0$. Furthermore, as $n$ diverges, we have $|\hat{\sigma}_{e,k}^2 - \sigma_{e,k}^2| \xrightarrow{p} 0$ and $|\hat{\zeta}_{ikl} - \zeta_{ikl}| \xrightarrow{p} 0$, where $\zeta_{ikl} = E[\zeta_{ikl}|\xi_{W,ik}]$ and $\xi_{W,ik}$ is the $m_i$-dimensional column vector of $\xi_{W,ijk}$’s.

**Proof.** With the same reasoning that we use to show Corollary 2, we can show the consistency of the estimators of eigen-elements from the second FPCA. By theorem 4.4 of Bosq (2000, p.104),

$$0 \leq \sup_{t \geq 1} |\hat{\eta}_{kl} - \eta_{kl}| \leq \|\hat{G}_k(t,t') - G_k(t,t')\|_L \leq \|\hat{G}_k(t,t') - G_k(t,t')\|_s,$$

and

$$0 \leq \|\hat{\psi}_{kl}(t) - \psi_{kl}(t)\|_s \leq 2\sqrt{2}b_{kl}^{-1}\|\hat{G}_k(t,t') - G(t,t')\|_s,$$

where $b_{kl} = (\eta_{k1} - \eta_{k2})$ for $l = 1$ and $\max[(\eta_{k,l-1} - \eta_{k,l}), (\eta_{k,l} - \lambda_{k,l+1})]$ otherwise. Under the assumption (A7.), the consistency of $\|\hat{G}_k(t,t') - G_k(t,t')\|_s$ implies that the following holds:

$$|\hat{\eta}_{kl} - \eta_{kl}| \xrightarrow{p} 0, \text{ and } \|\hat{\psi}_{kl}(\cdot) - \psi_{kl}(\cdot)\|_s \xrightarrow{p} 0$$

as $n$ diverges.

Furthermore using Corollary 1, theorems 2 and 3 of Yao et al. (2005a), the consistency of the estimators of $\sigma_{e,k}^2$ and $\zeta_{ikl}$ as well as the uniform consistency of $\psi_{kl}(\cdot)$ can be shown with the uniform consistency result of $\hat{G}_k(t,t')$ obtained in theorem 2 of this paper. The
following proofs are similar or identical to the corresponding ones given in Yao et al. (2005a).

First we show the uniform consistency of $\psi_{kl}(\cdot)$; the proof is the same as one given in Yao et al. (2005a, p. 589). For fixed $l$,

$$
|\hat{\eta}_{kl}\hat{\psi}_{kl}(t) - \eta_{kl}\psi_{kl}(t)| = \left| \int \hat{G}_k(t, t')\hat{\psi}_{kl}(t')dt' - \int G_k(t, t')\psi_{kl}(t')dt' \right|
$$

$$
\leq \int |\hat{G}_k(t, t') - G_k(t, t')||\hat{\psi}_{kl}(t')|dt' + \int |G_k(t, t')||\hat{\psi}_{kl}(t') - \psi_{kl}(t')|dt'
$$

$$
\leq \sqrt{\int (\hat{G}_k(t, t') - G_k(t, t'))^2dt'} + \sqrt{\int (G_k(t, t'))^2dt'}\|\hat{\psi}_{kl} - \psi_{kl}\|_s,
$$

and the consistency results of $\hat{G}_k(t, t')$ and $\hat{\psi}_{kl}(t)$ that we obtained previously imply that $\sup_t|\hat{\eta}_{kl}\hat{\psi}_{kl}(t) - \eta_{kl}\psi_{kl}(t)|$ converges to 0 and so does $\sup_t|\hat{\eta}_{kl}\hat{\psi}_{kl}(t)/\eta_{kl} - \psi_{kl}(t)|$ as $n$ diverges. By Slutsky’s theorem and the consistency of $\hat{\eta}_{kl}$, it is easy to show that $\sup_t|\hat{\eta}_{kl}\hat{\psi}_{kl}(t)/\eta_{kl} - \psi_{kl}(t)|$ is asymptotically equivalent to $\sup_t|\hat{\psi}_{kl}(t) - \psi_{kl}(t)|$ and thus the uniform consistency of $\hat{\psi}_{kl}(t)$ holds.

Secondly the consistency of the estimator, $\hat{\sigma}_e^2$, follows from the uniform consistency of $\hat{G}(t, t')$ and that of $\hat{V}_k(t)$; it is shown in Yao et al. (2005a) that the local linear smoother, $\hat{V}_k(t)$, is a uniform consistent estimator of $G_k(t, t) + \sigma_e^2$. This result is analogous to Corollary 1 of Yao et al. (2005a).

Lastly by Slutsky’s theorem and the consistency results of the estimators, $\hat{\eta}_{kl}$, $\hat{\psi}_{ikk}$, $\hat{\Sigma}_{\xi_{ik}}$ and $\hat{\xi}_{W,ik}$, we have

$$
|\hat{\zeta}_{ikl} - \tilde{\zeta}_{ikl}| = \left| E[\hat{\zeta}_{ikl}\tilde{\xi}_{W,ik}] - E[\zeta_{ikl}\xi_{W,ik}] \right| \overset{p}{\to} 0 \quad (B.26)
$$

as $n$ diverges.
Theorem 4.1.3 (Theorem 3 in Section 3.5.1). Assume (A1.) - (A8.), for each \((s,t) \in S \times T\). Then \(\tilde{Y}_i(s,t) \overset{p}{\to} \sum_{k=1}^{\infty} \sum_{t=1}^{\infty} \tilde{\zeta}_{ikl} \psi_{kl}(t) \phi_k(s)\) as \(n, K\) and \(L_k\)'s \(\to \infty\).

Proof. Let \(\tilde{Y}_i(s,t) = \sum_{k=1}^{\infty} \sum_{t=1}^{\infty} \tilde{\zeta}_{ikl} \psi_{kl}(t) \phi_k(s)\) be the full (not-truncated) true trajectory. And let \(\tilde{Y}_i^K(s,t) = \sum_{k=1}^{K} \sum_{t=1}^{\infty} \tilde{\zeta}_{ikl} \psi_{kl}(t) \phi_k(s)\) and \(\tilde{Y}_i^{KL}(s,t) = \sum_{k=1}^{K} \sum_{l=1}^{L_k} \tilde{\zeta}_{ikl} \psi_{kl}(t) \phi_k(s)\) be the truncated versions of \(\tilde{Y}_i(s,t)\). Because \(|\tilde{Y}_i(s,t) - \tilde{Y}_i(s,t)| \leq |\tilde{Y}_i(s,t) - \tilde{Y}_i^K(s,t)| + |\tilde{Y}_i^K(s,t) - \tilde{Y}_i^{KL}(s,t)|\), we prove the consistency of \(\tilde{Y}_i(s,t)\) by showing that each of \(|\tilde{Y}_i(s,t) - \tilde{Y}_i^K(s,t)|\) and \(|\tilde{Y}_i^{KL}(s,t) - \tilde{Y}_i(s,t)|\) converges to 0 in probability. Using Slutsky’s theorem and the consistency results obtained for the model components estimators, it is easy to show that \(|\tilde{Y}_i(s,t) - \tilde{Y}_i^{KL}(s,t)|\) converges to 0 in probability as \(n\) diverges.

In the following we show that the second term, \(|\tilde{Y}_i(s,t) - \tilde{Y}_i^{KL}(s,t)|\), converges to 0 in probability by proving that \(E[||\tilde{Y}_i(\cdot,\cdot) - \tilde{Y}_i^{KL}(\cdot,\cdot)||^2]\) converges to 0 as \(K\) and \(L_k\)'s diverge. Because

\[
\|\tilde{Y}_i(\cdot,\cdot) - \tilde{Y}_i^{KL}(\cdot,\cdot)\|^2 = \|\{\tilde{Y}_i(\cdot,\cdot) - \tilde{Y}_i^K(\cdot,\cdot)\} + \{\tilde{Y}_i^K(\cdot,\cdot) - \tilde{Y}_i^{KL}(\cdot,\cdot)\}\|^2
\]

\[
\leq 2\{\|\tilde{Y}_i(\cdot,\cdot) - \tilde{Y}_i^K(\cdot,\cdot)\|^2 + \|\tilde{Y}_i^K(\cdot,\cdot) - \tilde{Y}_i^{KL}(\cdot,\cdot)\|^2\},
\]

we prove the convergence of \(E[||\tilde{Y}_i(\cdot,\cdot) - \tilde{Y}_i^{KL}(\cdot,\cdot)||^2]\) to 0 by proving that each of \(E[||\tilde{Y}_i(\cdot,\cdot) - \tilde{Y}_i^K(\cdot,\cdot)||^2]\) and \(E[||\tilde{Y}_i^K(\cdot,\cdot) - \tilde{Y}_i^{KL}(\cdot,\cdot)||^2]\) converges to 0 as \(K\) and \(L_k\)'s diverge.

Let \(\psi_{ikl} = \{\hat{\psi}_{ikl}(t_{i_1}), \ldots, \hat{\psi}_{ikl}(t_{im_i})\}\) be the \(m_i\)-dimensional column vector of the evaluations of \(\psi_{kl}(\cdot)\) at \(\{t_{ij} : j = 1, \ldots, m_i\}\), \(\Sigma_{\xi_{W,ik}}\) be a \(m_i \times m_i\) - matrix with \((j,j')\)th element equal to \(G_k(t_{ij}, t_{ij'}) + \sigma^2_{\psi_{kl}}\), for \(j = j'\) and \(G_k(t_{ij}, t_{ij'})\) otherwise, and \(\xi_{W,ik}\) be the \(m_i\) - dimensional column vector of \(\xi_{W,ijk}\)'s; note that these are the true parameters corresponding to the estimators, \(\hat{\psi}_{ikl}, \hat{\Sigma}_{\xi_{W,ik}}\) and \(\hat{\xi}_{W,ik}\), defined in Section 3.3.3. Now
consider the first term, $E[\|\tilde{Y}_i(\cdot, \cdot) - \tilde{Y}_i^K(\cdot, \cdot)\|^2]$

$E[\|\tilde{Y}_i(\cdot, \cdot) - \tilde{Y}_i^K(\cdot, \cdot)\|^2]$

$$= E\left[ \int \int \left\{ \sum_{k=1}^{\infty} \sum_{l=1}^{\infty} \tilde{\zeta}_{ikl} \psi_{ikl}(t) \phi_k(s) - \sum_{k=1}^{K} \sum_{l=1}^{\infty} \tilde{\zeta}_{ikl} \psi_{ikl}(t) \phi_k(s) \right\}^2 dsdt \right]$$

$$= E\left[ \int \int \left\{ \sum_{k=K+1}^{\infty} \sum_{l=1}^{\infty} \tilde{\zeta}_{ikl} \psi_{ikl}(t) \phi_k(s) \right\}^2 dsdt \right]$$

$$= E\left[ \sum_{k=K+1}^{\infty} \int \left\{ \sum_{l=1}^{\infty} \tilde{\zeta}_{ikl} \psi_{ikl}(t) \right\}^2 dt \right] \quad \text{(using the orthonormal property of } \phi_k(\cdot))$$

$$= E\left[ \sum_{k=K+1}^{\infty} \sum_{l=1}^{\infty} (\tilde{\zeta}_{ikl})^2 \right] \quad \text{(using the orthonormal property of } \psi_{ik}(\cdot))$$

$$= E\left[ \sum_{k=K+1}^{\infty} \sum_{l=1}^{\infty} \{E[\zeta_{ikl}|\xi_{W,ik}]\}^2 \right]$$

$$= \sum_{k=K+1}^{\infty} \sum_{l=1}^{\infty} \text{var} \left\{ E[\zeta_{ikl}|\xi_{W,ik}] \right\} \quad \text{(using the Monotone Convergence theorem)}$$

$$= \sum_{k=K+1}^{\infty} \sum_{l=1}^{\infty} \eta_{kl}^2 \psi_{ikl}^t \Sigma_{\xi_{W,ik}}^{-1} \psi_{ikl} \quad \text{(under Gaussian assumption)}$$

$$= \sum_{k=K+1}^{\infty} \sum_{l=1}^{\infty} \eta_{kl}^2 \cdot tr \left[ \psi_{ikl}^t \Sigma_{\xi_{W,ik}}^{-1} \psi_{ikl} \right] \quad \text{(because trace } \text{tr}(a) = a, a \in \mathbb{R})$$

$$= \sum_{k=K+1}^{\infty} \sum_{l=1}^{\infty} \eta_{kl}^2 \cdot tr \left[ \psi_{ikl}^t \left\{ \sum_{l'=1}^{\infty} \eta_{kl'} \psi_{ikl'} \psi_{ikl'}^t + \sigma_{e,k}^2 I_{m_i} \right\}^{-1} \psi_{ikl} \right]$$

(\text{using the KL expansion of } G_k(t, t'))

$$= \sum_{k=K+1}^{\infty} \text{tr} \left[ \left\{ \sum_{l'=1}^{\infty} \eta_{kl'} \psi_{ikl'} \psi_{ikl'}^t + \sigma_{e,k}^2 I_{m_i} \right\}^{-1} \sum_{l=1}^{\infty} \eta_{kl}^2 \psi_{ikl} \psi_{ikl}^t \right]$$

(by the linear and cyclic properties of trace)

$$\leq \sum_{k=K+1}^{\infty} \eta_{kl} \cdot \text{tr} \left[ \left\{ \sum_{l'=1}^{\infty} \eta_{kl'} \psi_{ikl'} \psi_{ikl'}^t + \sigma_{e,k}^2 I_{m_i} \right\}^{-1} \sum_{l=1}^{\infty} \eta_{kl} \psi_{ikl} \psi_{ikl}^t \right]$$

(because $\{\eta_{kl}\}_t$ is a strictly decreasing sequence for each $k$)
\[
\begin{align*}
&= \sum_{k=K+1}^{\infty} \eta_{k1} \cdot t_r I_{m_i} - \left\{ \sum_{t'=1}^{\infty} \eta_{kt'} \psi_{ikl'} \psi_{kll'}^t + \sigma_{e,k}^2 I_{m_i} \right\}^{-1} \sigma_{e,k}^2 I_{m_i} \\
&\leq \sum_{k=K+1}^{\infty} \eta_{k1} t_r (I_{m_i}) \\
&\quad \text{(because trace of a positive definite matrix is positive)} \\
&= \sum_{k=K+1}^{\infty} \eta_{k1} m_i \leq m_i \sum_{k=K+1}^{\infty} \sum_{l=1}^{\infty} \eta_{kl} \rightarrow 0,
\end{align*}
\]

as \( K \) diverges because \( m_i \) is finite and the square integrable property of \( X_i(s, t) \) in the assumption (A1.) ensures that \( \sum_{k=1}^{\infty} \sum_{l=1}^{\infty} \eta_{kl} \) is finite. Thus we show that \( E[\|\tilde{Y}_i(\cdot, \cdot) - \tilde{Y}_i^{KL}(\cdot, \cdot)\|^2] \) converges to 0 as \( K \) diverges.

We use the similar reasoning to show that \( E[\|\tilde{Y}_i^K(\cdot, \cdot) - \tilde{Y}_i^{KL}(\cdot, \cdot)\|^2] \) converges to 0 as \( K \) and \( L_k \)'s diverges. Specifically from the previous part, we use the fact that for each \( k \), \( \sum_{l=1}^{\infty} \eta_{kl}^2 \cdot t_r \left\{ \left( \sum_{t'=1}^{\infty} \eta_{kt'} \psi_{ikl'} \psi_{kll'}^t + \sigma_{e,k}^2 I_{m_i} \right) \right\}^{-1} \psi_{ikl'} \psi_{kll'}^t \left\| \right\| \leq \tilde{\eta}_{k1} m_i \) is finite. Most steps are identical thus omitted.

\[
E[\|\tilde{Y}_i^K(\cdot, \cdot) - \tilde{Y}_i^{KL}(\cdot, \cdot)\|^2] = E\left[ \int \int \left\{ \sum_{k=1}^{K} \sum_{l=1}^{\infty} \tilde{\zeta}_{ikl} \psi_{kl}(t) \phi_k(s) - \sum_{k=1}^{K} \sum_{l=1}^{L_k} \tilde{\zeta}_{ikl} \psi_{kl}(t) \phi_k(s) \right\}^2 ds dt \right] \\
= E\left[ \int \int \left\{ \sum_{k=1}^{K} \sum_{l=L_{k}+1}^{\infty} \tilde{\zeta}_{ikl} \psi_{kl}(t) \phi_k(s) \right\}^2 ds dt \right] \\
\leq \sum_{k=1}^{K} \sum_{l=L_{k}+1}^{\infty} \eta_{kl}^2 \cdot t_r \left\{ \left( \sum_{t'=1}^{\infty} \eta_{kt'} \psi_{ikl'} \psi_{kll'}^t + \sigma_{e,k}^2 I_{m_i} \right) \right\}^{-1} \psi_{ikl'} \psi_{kll'}^t \rightarrow 0,
\]

as \( K \) and \( L_k \)'s diverge because for each \( k \), \( \sum_{l=L_{k}+1}^{\infty} \eta_{kl}^2 \cdot t_r \left\{ \left( \sum_{t'=1}^{\infty} \eta_{kt'} \psi_{ikl'} \psi_{kll'}^t + \sigma_{e,k}^2 I_{m_i} \right) \right\}^{-1} \psi_{ikl'} \psi_{kll'}^t \) converges to 0 as \( L_k \) diverges.

The convergence of \( E[\|\tilde{Y}_i(\cdot, \cdot) - \tilde{Y}_i^{KL}(\cdot, \cdot)\|^2] \) to 0 follows from the fact that \( E[\|\tilde{Y}_i^K(\cdot, \cdot) - \tilde{Y}_i^{KL}(\cdot, \cdot)\|^2] \) and \( E[\|\tilde{Y}_i^K(\cdot, \cdot) - \tilde{Y}_i^{KL}(\cdot, \cdot)\|^2] \) converge to 0. Furthermore by Markov inequality it is implied that \( |\tilde{Y}_i(s, t) - \tilde{Y}_i^{KL}(s, t)| \) converges to 0 in probability as \( K \) and \( L_k \)'s diverge.
B.1.2 Case when response curves are measured with smooth error (Section 3.5.2)

Now we consider the case when data is corrupted with smooth error process, $\epsilon_{1,ij}(s)$; in other words, $Y_{ij}(s) = X_i(s, t_{ij}) + \epsilon_{1,ij}(s)$. The main difference from the case of having $Y_{ij}(s) = X_i(s, t_{ij})$ is that the sample covariance of $Y_{ij}(s)$ is no longer an estimator of the marginal covariance function, $\Sigma(s, s')$, but is an estimator of $\Xi(s, s') = \Sigma(s, s') + \Gamma(s, s')$, where $\Gamma(s, s')$ is a smooth covariance function of the error process, $\epsilon_{1,ij}(s)$.

As the additional error process, $\epsilon_{1,ij}(s)$, are correlated over $s$, but independent over $i$ and $j$, the theoretical results associated with the longitudinal dynamics (the second FPCA step) and their proofs remain the same. In the following we provide proofs of the theoretical results given in Section 3.5.2 that correspond to the marginal FPCA step.

We start our proofs with showing that $\widehat{\Xi}(s, s')$ is an unbiased estimator of $\Xi(s, s')$.

**Corollary 6.** The sample covariance, $\widehat{\Xi}(s, s')$, is an unbiased estimator of the covariance function, $\Xi(s, s') = \Sigma(s, s') + \Gamma(s, s')$.

**Proof.** Proof is identical to that of Corollary 5 except that $E\{Y_i(s, t_d)Y_i(s', t_d)\} = E\{X_i(s, t_d) + \epsilon_{1,ij}(s)\}{X_i(s', t_d) + \epsilon_{1,ij}(s')} = c((s, t_d), (s', t_d)) + \Gamma(s, s')$, instead of $c((s, t_d), (s', t_d))$ only. Specifically,

$$E(\widehat{\Xi}(s, s')) = E\left\{\frac{1}{\sum_{i=1}^{m_i} \sum_{d=1}^{m_d} \sum_{j=1}^{m_j} Y_{ij}(s)Y_{ij}(s')1(t_{ij} \in (t_{d-1}, t_d))}\right\}$$

$$= E\left\{\frac{1}{\sum_{i=1}^{m_i} \sum_{d=1}^{m_d} \sum_{j=1}^{m_j} Y_{ij}(s)Y_{ij}(s')1(t_{ij} \in (t_{d-1}, t_d))}\right\} = E\left\{\frac{1}{\sum_{i=1}^{m_i} \sum_{d=1}^{m_d} \sum_{j=1}^{m_j} (X_i(s, t_d) + \epsilon_{1,ij}(s))1(t_{ij} \in (t_{d-1}, t_d))}\right\}$$

$$= E\left\{\frac{1}{\sum_{i=1}^{m_i} \sum_{d=1}^{m_d} \sum_{j=1}^{m_j} (X_i(s, t_d) + \epsilon_{1,ij}(s))1(t_{ij} \in (t_{d-1}, t_d))}\right\} = E\left\{\frac{1}{\sum_{i=1}^{m_i} \sum_{d=1}^{m_d} \sum_{j=1}^{m_j} c((s, t_d), (s', t_d))1(t_{ij} \in (t_{d-1}, t_d))}\right\}$$

$$= E\left\{\frac{1}{\sum_{i=1}^{m_i} \sum_{d=1}^{m_d} \sum_{j=1}^{m_j} c((s, t_d), (s', t_d))1(t_{ij} \in (t_{d-1}, t_d))}\right\} = E\left\{\frac{1}{\sum_{i=1}^{m_i} \sum_{d=1}^{m_d} \sum_{j=1}^{m_j} \Gamma(s, s')1(t_{ij} \in (t_{d-1}, t_d))}\right\}$$

$$= E\left\{\frac{1}{\sum_{i=1}^{m_i} \sum_{d=1}^{m_d} \sum_{j=1}^{m_j} \Gamma(s, s')1(t_{ij} \in (t_{d-1}, t_d))}\right\} = E\left\{\frac{1}{\sum_{i=1}^{m_i} \sum_{d=1}^{m_d} \sum_{j=1}^{m_j} \Gamma(s, s')1(t_{ij} \in (t_{d-1}, t_d))}\right\}$$

154
\[
= \sum_{d=1}^{D} \{c((s, t_d), (s', t_d)) + \Gamma(s, s')\} \mathbb{P}(t \in (t_{d-1}, t_d]) = \int \{c((s, t_d), (s', t_d)) + \Gamma(s, s')\} dF(t)
\]
\[
= \int c((s, t_d), (s', t_d)) dF(t) + \Gamma(s, s') \quad \text{(because } \int dF(t) = 1)\]
\[
= \Sigma(s, s') + \Gamma(s, s') = \Xi(s, s'),
\]
where \( F \) is a true sampling distribution function of \( t \).

**Corollary 4.2.1** (Corollary 3 in Section 3.5.2). Under the assumptions (A1.) - (A3.), and (A9.), for each \((s, s')\), \(|\mathbb{E}(s, s') - \Xi(s, s')| \overset{p}{\to} 0 \text{ as } n \text{ diverges. And under the assumptions (A1.), (A2.), (A4.), (A9.)-(A11.),} ||\mathbb{E}(\cdot, \cdot) - \Xi(\cdot, \cdot)||_s \overset{p}{\to} 0 \text{ and sup}_j |\tilde{\xi}_{W,ijk} - \xi_{W,ijk}| \overset{p}{\to} 0 \text{ as } n \to \infty.\]

**Proof.** This corollary includes three consistency results that are corresponding to the ones given in theorems 1 and 2. As proofs are very similar to the corresponding ones detailed earlier many steps are omitted. In the following we give a short proofs for (i) the pointwise consistency of \( \hat{\Xi}(s, s') \), (ii) the Hilbert-Schmidt norm consistency of \( \hat{\Xi}(s, s') \), and (iii) the uniform consistency of \( \tilde{\xi}_{W,ijk} \).

(i) The pointwise consistency of \( \hat{\Xi}(s, s') \)

To show the pointwise consistency of \( \hat{\Xi}(s, s') \), we need to show that the following converges to 0 as \( n \) diverges:

\[
\frac{1}{D^2} \left[ \sum_{d=1}^{D} \sigma_d^2 + \sum_{d=1}^{D} \sum_{d' \neq d} \sigma_{d'd'} \right] = \frac{1}{M_1} \mathbb{E}_d [\mathbb{E}[A(t)^2 | d]] + \left[ \left\{ \frac{M_2}{M_1^2} - \frac{1}{M_1} \right\} \mathbb{E}_d [\mathbb{E}[A(t)A(t') | t, t']] \right] - \left[ \left\{ \frac{M_2}{M_1^2} - \frac{1}{M_1} \right\} \sum_{d=1}^{D} P_d^2 \cdot \mathbb{E}[A(t)^2 | t = t_d] \right] - \frac{M_2}{M_1^2} \{\Sigma(s, s') + \Gamma(s, s')\}^2, \quad (B.27)
\]

which is analogous to Equation (B.7).

Because \( X_i(s, t_{ij})'s \) and \( \epsilon_{ij}(s)'s \) are mutually independent and \( \epsilon_{ij}(s)'s \) are independent
over \(i\) and \(j\),

\[
E[A(t)A(t')|t, t'] = E[Y_{ij}(s)Y_{ij}(s')Y_{ij}(s)Y_{ij}(s')]
\]

\[
= E[X_i(s, t_{ij})X_i(s', t_{ij})X_i(s, t'_{ij})X_i(s', t'_{ij})] + \{\Gamma(s, s')\}^2
\]

\[
+ \Gamma(s, s')(c((s, t_{ij}), (s, t'_{ij}))) + \Gamma(s, s')c((s, t'_{ij}), (s', t'_{ij})).
\]

Thus under the assumptions (A3.) and (A9.), \(E[A(t)A(t')|t, t']\) is finite. Then using the same argument that we used to show the convergence of Equation (B.7), we can show that Equation (B.27) converges to 0 as \(n\) diverges. It implies that for each \((s, s')\), \(\hat{\Xi}(s, s') - \Xi(s, s')|_{P} \to 0\) as \(n\) diverges.

(ii) The Hilbert-Schmidt norm consistency of \(\hat{\Xi}(s, s')\)

To show the Hilbert-Schmidt norm consistency, we need to show that the following converges to 0 as \(n\) diverges:

\[
\left\{ \frac{(M_2 - M_1) \cdot \sup_d P_d}{M_1^2} + \frac{1}{M_1} + \frac{2(M_2 - M_1)}{M_1^2} + \frac{1}{M_1} \right\} \cdot E_t\left[ E[\|Y(\cdot, t_d)\|^4] \right]. \tag{B.28}
\]

this is the same equation as Equation (B.17). Because \(X_i(s, t_{ij})\)’s and \(\epsilon_{ij}(s)\)’s are mutually independent and \(\epsilon_{ij}(s)\)’s are independent over \(i\) and \(j\),

\[
E[\|Y_{ij}(\cdot)\|^4 | t = t_{ij}] = E[\|X_i(\cdot, t)\|^4 | t = t_{ij}] + E[\|\epsilon_{1,ij}(\cdot)\|^4]
\]

\[
+ E[\|X_i(\cdot, t)\|^2 | t = t_{ij}] \cdot E[\|\epsilon_{1,ij}(\cdot)\|^2] + 4 \int \int c((s, t_{ij}), (s', t_{ij}))\Gamma(s, s')dsds',
\]

where the first three terms are finite under the assumptions (A4.) and (A10.) and the last term is finite because by Cauchy-Schwarz inequality

\[
\int \int c((s, t_{ij}), (s', t_{ij}))\Gamma(s, s')dsds' < \sqrt{\int \int c((s, t_{ij}), (s', t_{ij}))^2dsds'} \sqrt{\int \int \Gamma(s, s')^2dsds'}.
\]
thus $E[\|Y(\cdot, t)\|^4|t]$ is also finite and under the assumptions (A1.), (A2.), (A4.) and (A10.) Equation (B.28) converges to 0 as $n$ diverges. Then it follows that $\|\hat{\Xi}(\cdot, \cdot) - \Xi(\cdot, \cdot)\|_s$ converges to 0 in probability as $n$ diverges.

(iii) The uniform consistency of $\tilde{\xi}_{W,ijk}$

Lastly the uniform consistency of $\tilde{\xi}_{W,ijk}$ holds because from Equation (B.22) we have

\[ \sup_j |\tilde{\xi}_{W,ijk} - \xi_{W,ijk}| \leq \sup_j |Y_i(s, t_{ij})| \times \|\hat{\phi}_k(\cdot) - \phi_k(\cdot)\|_s, \]

which is less than or equal to

\[ \{\sup_j |X_i(s, t_{ij})| + \sup_s |\epsilon_1_{ij}(s)|\} \times \|\hat{\phi}_k(\cdot) - \phi_k(\cdot)\|_s, \]

where $\sup_j |X_i(s, t_{ij})|$ and $\sup_s |\epsilon_1_{ij}(s)|$ are asymptotically bounded under the assumptions (A6.) and (A11.).

B.2 Additional details for the simulation experiment

B.2.1 Description of the study

Errors are generated from $\epsilon_{ij}(s) = e_{ij1}\phi_1(s) + e_{ij2}\phi_2(s) + \epsilon_{2,ij}(s)$, where $e_{ij1}$, $e_{ij2}$ and $\epsilon_{2,ij}(s)$ are mutually independent with zero-mean and variances equal to $\sigma^2_{e,1}, \sigma^2_{e,2}$ and $\sigma^2$, respectively. The white noise variance, $\sigma^2$, is set based on the signal to noise ratio (SNR),

\[ SNR = \frac{\iint \text{Var}\{Y_i(s, t)\} dsdt}{\sigma^2_1 + \sigma^2_2 + \sigma^2} - 1. \quad (B.29) \]

We consider the following experimental factors:

Case 1. covariance structure of the time-varying components:

(a) non-parametric covariance (NP): $\xi_{ik}(t) = \zeta_1 k_1(t) + \zeta_2 k_2(t)$, where

(i) $\psi_{11}(t) = \sqrt{2}\cos(2\pi t), \psi_{12}(t) = \sqrt{2}\sin(2\pi t)$, $\zeta_{11} \overset{\text{iid}}{\sim} N(0, 3), \zeta_{12} \overset{\text{iid}}{\sim} N(0, 1.5);$  

(ii) $\psi_{21}(t) = \sqrt{2}\cos(4\pi t), \psi_{22}(t) = \sqrt{2}\sin(4\pi t)$, $\zeta_{21} \overset{\text{iid}}{\sim} N(0, 2), \zeta_{22} \overset{\text{iid}}{\sim} N(0, 1).$

(b) random effects model (REM): $\xi_{ik}(t) = b_{i0} + b_{ik} t$ with

\[ \left( \begin{array}{c} b_{i10} \\ b_{i11} \end{array} \right) \overset{\text{iid}}{\sim} N \left( \begin{array}{c} 0 \\ 0 \end{array} \right), \left( \begin{array}{cc} 0 & 2.5 \\ 0 & 2 \end{array} \right), \left( \begin{array}{c} b_{20} \\ b_{21} \end{array} \right) \overset{\text{iid}}{\sim} N \left( \begin{array}{c} 0 \\ 0 \end{array} \right), \left( \begin{array}{cc} 2 & 1 \\ 1 & 1.5 \end{array} \right) \right), \]

157
(c) exponential autocorrelation model (Exp): \( \xi_{ik}(t) \) is a Gaussian process with mean zero, variance \( \lambda_k \) and auto-correlation function \( \text{corr}\{\xi_{ik}(t), \xi_{ik}(t')\} = \rho_k^{\left|t-t'\right|} \) denoted by \( GP(\lambda_k, \rho_k) \). We set \( \xi_{i1}(t) \overset{iid}{\sim} GP(4.5, 0.9) \) and \( \xi_{i2}(t) \overset{iid}{\sim} GP(3, 0.5) \).

Note that regardless of the generating models for \( \xi_{ik}(t) \), we have that \( \int \text{var}\{\xi_{ik}(t)\}dt \) is equal to 4.5 and 3 for \( k = 1, 2 \) respectively.

Case 2. number of repeated measurements per subject:

(a) \( m_i \overset{iid}{\sim} \text{Uniform}\{8, 9, \ldots, 12\} \) (about 75% missing)

(b) \( m_i \overset{iid}{\sim} \text{Uniform}\{15, 16, \ldots, 20\} \) (about 55% missing)

Case 3. variance of \( e_{ijk} \):

(a) \( (\sigma_{e1}^2, \sigma_{e2}^2) = (0, 0) \) (white noise only, i.e. \( \epsilon_{ij}(s) \overset{iid}{\sim} N(0, \sigma^2) \))

(b) \( (\sigma_{e1}^2, \sigma_{e2}^2) = (0.7, 0.3) \).

The simulation results for the Case 3.(a), i.e. no smooth error, are included in the Supplementary Material.

Case 4. signal to noise ratio: (a) \( \text{SNR} = 1 \ (\sigma^2 = 6.5) \), and (b) \( \text{SNR} = 5 \ (\sigma^2 = 0.5) \)

Case 5. number of subjects: (a) \( n = 100 \), (b) \( n = 300 \), and (c) \( n = 500 \)

For each generated sample of size \( n \) we form a training set and a test set. To determine the test set we randomly select 10 subjects from the sample. The test set is formed by collecting these subjects’ last functional observation; hence the test set contains 10 curves. The remaining functional observations for the 10 subjects and the data corresponding to the remaining subjects in the sample form the training set. Our model is fitted using the training set and the methods outlined in Section 3.3. To be more specific, the bivariate mean function, \( \mu(s, t) \), is modeled using 50 cubic spline basis functions obtained from the
tensor product of \( d_s = 10 \) basis functions in direction \( s \) and \( d_t = 5 \) in \( t \). The smoothing parameters are selected via REML. The finite truncations \( K \) and \( L_k \)'s are all estimated using the pre-specified level \( PVE = 0.95 \).

Estimation accuracy for the model components is evaluated using integrated mean squared errors (IMSE): specifically, for the bivariate mean function \( \text{IMSE}(\hat{\mu}) = \sum_{i=1}^{N_{sim}} \int \int \{ \hat{\mu}_{isim}(s,t) - \mu(s,t) \}^2 dsdt/N_{sim} \), and for the univariate eigenfunctions \( \text{IMSE}(\hat{\phi}_k) = \sum_{i=1}^{N_{sim}} \int \{ \hat{\phi}_{isim}(s) - \phi_k(s) \}^2 ds/N_{sim}, k = 1, 2 \). The prediction performance is assessed through the accuracy in predicting the time-varying model components, \( \xi_{ik}(t) \), and in predicting the response curve, \( Y_i(s,t) \). For the former assessment we use the in-sample integrated prediction errors (IPE) defined as \( \text{IPE}(\xi_k) = \sum_{i=1}^{N_{sim}} \sum_{i=1}^{n} \int \{ \hat{\xi}_{isim}(t) - \xi_{isim}(t) \}^2 dt/(nN_{sim}), k = 1, 2 \). For the later assessment we use the in-sample IPE (IN-IPE) defined as \( \text{IN-IPE}(Y) = \sum_{i=1}^{N_{sim}} \sum_{i=1}^{n} \sum_{j=1}^{n} \int \{ \hat{Y}_{ij}(s) - Y_{*ij}(s) \}^2 ds/(N_{sim} \sum_{i=1}^{n} m_i) \), where \( Y_{*ij}(s) \) is the true signal, i.e. without measurement error \( \epsilon_{ij}(s) \). Also we use the out-of-sample IPE (OOut-IPE) defined as \( \text{OOut-IPE}(Y) = \sum_{i=1}^{N_{sim}} \sum_{i \in \text{test set}} \int \{ \hat{Y}_{im}(s) - Y_{*im}(s) \}^2 ds/(10N_{sim}) \) and \( Y_{*ij}(s) \) is the true signal in the test set. The results are based on \( N_{sim} = 1000 \) simulations.

In terms of estimation performance and prediction of \( \xi_{ik}(t) \) there is no alternative approach. On the other hand, in terms of model prediction error and prediction of a subject’s future curve there are two possible alternatives. One is the CM model of Chen & Müller (2012). However due to the high computational expense required by their method, we have to restrict our comparison to few scenarios only: \( m_i \sim \{8, \ldots, 12\} \) number of repeated curves per subject, Case 3(b), and SNR = 1. The approach of Chen & Müller (2012) requires specification of several kernel bandwidths; due to the increased computation burden we use the pre-specified bandwidth \( h = 0.1 \) in smoothing both the mean and covariance functions. Even with these adjustments there is an order of
magnitude difference in the computational cost (when \( n = 100 \) the method of Chen & Müller (2012) takes approximately 984 seconds, while our approach takes about 7 seconds). As well, we also used the pre-specified level PVE = 0.95 to be consistent with our approach. A second alternative approach for prediction of a subject’s future visit trajectory is a rather naïve approach: let the future prediction equal the average of all previously observed profiles for that subject. For example, the naïve predictor of a profile of some subject in the test set is equal to the average of all profiles available in the training set for the corresponding subject.

**B.2.2 Additional results**

Simulation results for the case when the error process has trivial covariance structure are presented in table B.1. The results show that the prediction accuracy is greatly improved when the error process is just white noise. The other factors we investigated in the study seem to have the similar effects on the estimation and prediction accuracies as they did for the case of having smooth error process (table 3.1 in Section 3.6). Based on the results presented in both tables 3.1 and B.1, it seems that the exponential covariance structure (Exp) is most challenging; this most likely is due to the (very) large correlation coefficients used \( \rho_1 = 0.9 \) and \( \rho_2 = 0.5 \), which result in high temporal dependence even for the observations that are furthest apart \( t = 0 \) and \( t = 1 \), as \( t \in [0, 1] \). For completeness the average computational times of the proposed method are studied for all the cases and they are presented in tables B.2 and B.3.
Table B.1: Simulation results for estimation and prediction accuracy based on $N_{sim} = 1000$ simulations (white noise only)

$m_i \sim \{8, \ldots, 12\}$ and $SNR = 1$

<table>
<thead>
<tr>
<th>$\mu$</th>
<th>$\phi_1$</th>
<th>$\phi_2$</th>
<th>$\xi_1$</th>
<th>$\xi_2$</th>
<th>IN-P</th>
<th>IN-PE naive</th>
<th>OUT-P</th>
<th>OUT-PE naive</th>
</tr>
</thead>
<tbody>
<tr>
<td>NP (a)</td>
<td>$n = 100$</td>
<td>0.088</td>
<td>0.005</td>
<td>0.014</td>
<td>0.169</td>
<td>0.150</td>
<td>0.173</td>
<td>7.779</td>
</tr>
<tr>
<td></td>
<td>$n = 300$</td>
<td>0.029</td>
<td>0.001</td>
<td>0.010</td>
<td>0.067</td>
<td>0.066</td>
<td>0.109</td>
<td>7.788</td>
</tr>
<tr>
<td></td>
<td>$n = 500$</td>
<td>0.017</td>
<td>0.001</td>
<td>0.010</td>
<td>0.048</td>
<td>0.047</td>
<td>0.092</td>
<td>7.778</td>
</tr>
<tr>
<td>REM (b)</td>
<td>$n = 100$</td>
<td>0.110</td>
<td>0.035</td>
<td>0.043</td>
<td>0.299</td>
<td>0.303</td>
<td>0.193</td>
<td>1.197</td>
</tr>
<tr>
<td></td>
<td>$n = 300$</td>
<td>0.037</td>
<td>0.009</td>
<td>0.018</td>
<td>0.137</td>
<td>0.135</td>
<td>0.162</td>
<td>1.198</td>
</tr>
<tr>
<td></td>
<td>$n = 500$</td>
<td>0.022</td>
<td>0.005</td>
<td>0.014</td>
<td>0.108</td>
<td>0.102</td>
<td>0.156</td>
<td>1.198</td>
</tr>
<tr>
<td>Exp (c)</td>
<td>$n = 100$</td>
<td>0.091</td>
<td>0.032</td>
<td>0.041</td>
<td>0.330</td>
<td>0.526</td>
<td>0.386</td>
<td>1.529</td>
</tr>
<tr>
<td></td>
<td>$n = 300$</td>
<td>0.030</td>
<td>0.009</td>
<td>0.018</td>
<td>0.209</td>
<td>0.358</td>
<td>0.353</td>
<td>1.529</td>
</tr>
<tr>
<td></td>
<td>$n = 500$</td>
<td>0.017</td>
<td>0.005</td>
<td>0.014</td>
<td>0.189</td>
<td>0.323</td>
<td>0.344</td>
<td>1.529</td>
</tr>
</tbody>
</table>

$m_i \sim \{15, \ldots, 20\}$ and $SNR = 1$

$m_i \sim \{8, \ldots, 12\}$ and $SNR = 5$

$m_i \sim \{15, \ldots, 20\}$ and $SNR = 5$
Table B.2: Computational time (seconds) corresponding to table 3.1

<table>
<thead>
<tr>
<th>$m_i \sim {8, \ldots, 12}$ and $SNR = 1$</th>
<th>$n = 100$</th>
<th>$n = 300$</th>
<th>$n = 500$</th>
</tr>
</thead>
<tbody>
<tr>
<td>NP (a)</td>
<td>7.369</td>
<td>15.892</td>
<td>21.418</td>
</tr>
<tr>
<td>REM (b)</td>
<td>9.282</td>
<td>11.347</td>
<td>22.559</td>
</tr>
<tr>
<td>EXP (c)</td>
<td>7.514</td>
<td>16.229</td>
<td>17.109</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$m_i \sim {15, \ldots, 20}$ and $SNR = 1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>NP (a)</td>
</tr>
<tr>
<td>REM (b)</td>
</tr>
<tr>
<td>EXP (c)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$m_i \sim {8, \ldots, 12}$ and $SNR = 5$</th>
<th>$n = 100$</th>
<th>$n = 300$</th>
<th>$n = 500$</th>
</tr>
</thead>
<tbody>
<tr>
<td>NP (a)</td>
<td>7.405</td>
<td>16.162</td>
<td>21.507</td>
</tr>
<tr>
<td>REM (b)</td>
<td>9.609</td>
<td>13.283</td>
<td>21.166</td>
</tr>
<tr>
<td>EXP (c)</td>
<td>8.412</td>
<td>15.446</td>
<td>17.653</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$m_i \sim {15, \ldots, 20}$ and $SNR = 5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>NP (a)</td>
</tr>
<tr>
<td>REM (b)</td>
</tr>
<tr>
<td>EXP (c)</td>
</tr>
</tbody>
</table>

Table B.3: Computational time (seconds) corresponding to table B.1

<table>
<thead>
<tr>
<th>$m_i \sim {8, \ldots, 12}$ and $SNR = 1$</th>
<th>$n = 100$</th>
<th>$n = 300$</th>
<th>$n = 500$</th>
</tr>
</thead>
<tbody>
<tr>
<td>NP (a)</td>
<td>13.760</td>
<td>12.655</td>
<td>42.346</td>
</tr>
<tr>
<td>REM (b)</td>
<td>16.776</td>
<td>24.044</td>
<td>39.088</td>
</tr>
<tr>
<td>EXP (c)</td>
<td>10.541</td>
<td>20.195</td>
<td>30.071</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$m_i \sim {15, \ldots, 20}$ and $SNR = 1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>NP (a)</td>
</tr>
<tr>
<td>REM (b)</td>
</tr>
<tr>
<td>EXP (c)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$m_i \sim {8, \ldots, 12}$ and $SNR = 5$</th>
<th>$n = 100$</th>
<th>$n = 300$</th>
<th>$n = 500$</th>
</tr>
</thead>
<tbody>
<tr>
<td>NP (a)</td>
<td>8.222</td>
<td>27.563</td>
<td>50.990</td>
</tr>
<tr>
<td>REM (b)</td>
<td>11.318</td>
<td>17.683</td>
<td>24.464</td>
</tr>
<tr>
<td>EXP (c)</td>
<td>10.888</td>
<td>18.627</td>
<td>24.939</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$m_i \sim {15, \ldots, 20}$ and $SNR = 5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>NP (a)</td>
</tr>
<tr>
<td>REM (b)</td>
</tr>
<tr>
<td>EXP (c)</td>
</tr>
</tbody>
</table>
B.3 Additional figures for the DTI data analysis

This section includes figures that are discussed in Section 3.7.

Figure B.1: Fitted varying coefficient model, \( \hat{\mu}(s, t) = \hat{\mu}_0(s) + \hat{\beta}_t(s)t \) (left), and estimated slope function, \( \hat{\beta}_t(s) \) (right)

Figure B.2: Scree plot of the marginal FPCA
Figure B.3: Estimated eigenfunctions of the marginal FPCA for FA with PVE = 95%
Figure B.4: Estimated basis coefficient functions, $\hat{\xi}_{ik}(t)$ using a random coefficients linear model
B.4 Additional Figure for Interactive Graphics

This section includes additional figure discussed in Section 3.8

Figure B.5: Screenshot showing tab 2 of the interactive graphic. The plots show the third estimated eigenfunction $\hat{\phi}_3(s)$ (top) and the predicted trajectory of the selected subject at (scaled) visit time $t = 0.3$ (bottom)
Appendix C

Additional Details for Chapter 4

Section C.1 includes additional results of the simulation study discussed in Section 4.4.

C.1 Additional Simulation Results

Table 4.1 presents the type I error rates of the pLRT-MT corresponding to nominal levels $\alpha = 0.01, 0.05, 0.10,$ and 0.15 for different sample sizes $n$ and number of profiles per subject $m_i$.

Table C.1 presents the empirical sizes of the ZC-MT and Bootstrap-MT methods. The results indicate that the ZC-MT method has good size performance when there is relatively large number of repeated measures per subjects, $m_i \sim \{15, \ldots, 20\}$, for all sample sizes considered. However the ZC-MT method performs poorly with small number of repeated measures per subjects $m_i$ even when the sample size is large, $n = 300$. On the other hand, the Bootstrap-MT method is much more conservative and its empirical sizes are much smaller that the corresponding nominal levels for all sample sizes $n$ and number of repeated measures per subjects $m_i$, considered in our simulation.
Table C.1: The empirical type I error rates of the proposed method based on 5000 simulations. Standard errors are presented in parentheses.

<table>
<thead>
<tr>
<th>$m_1$</th>
<th>$\phi_k(s)$-choice</th>
<th>$\Pi_k(t,t')$-choice</th>
<th>$\alpha = 0.01$</th>
<th>$\alpha = 0.05$</th>
<th>$\alpha = 0.10$</th>
<th>$\alpha = 0.15$</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>True</td>
<td>True</td>
<td>0.010 (0.001)</td>
<td>0.047 (0.003)</td>
<td>0.093 (0.004)</td>
<td>0.141 (0.005)</td>
</tr>
<tr>
<td></td>
<td>True</td>
<td>Est.</td>
<td>0.013 (0.002)</td>
<td>0.058 (0.003)</td>
<td>0.105 (0.004)</td>
<td>0.155 (0.005)</td>
</tr>
<tr>
<td></td>
<td>Est.</td>
<td>Est.</td>
<td>0.017 (0.002)</td>
<td>0.063 (0.003)</td>
<td>0.112 (0.004)</td>
<td>0.156 (0.005)</td>
</tr>
<tr>
<td>200</td>
<td>True</td>
<td>True</td>
<td>0.010 (0.001)</td>
<td>0.048 (0.003)</td>
<td>0.099 (0.004)</td>
<td>0.147 (0.005)</td>
</tr>
<tr>
<td></td>
<td>True</td>
<td>Est.</td>
<td>0.012 (0.002)</td>
<td>0.051 (0.003)</td>
<td>0.102 (0.004)</td>
<td>0.154 (0.005)</td>
</tr>
<tr>
<td></td>
<td>Est.</td>
<td>Est.</td>
<td>0.012 (0.002)</td>
<td>0.052 (0.003)</td>
<td>0.103 (0.004)</td>
<td>0.151 (0.005)</td>
</tr>
<tr>
<td>300</td>
<td>True</td>
<td>True</td>
<td>0.008 (0.001)</td>
<td>0.049 (0.003)</td>
<td>0.098 (0.004)</td>
<td>0.145 (0.005)</td>
</tr>
<tr>
<td></td>
<td>True</td>
<td>Est.</td>
<td>0.008 (0.001)</td>
<td>0.051 (0.003)</td>
<td>0.099 (0.004)</td>
<td>0.147 (0.005)</td>
</tr>
<tr>
<td></td>
<td>Est.</td>
<td>Est.</td>
<td>0.008 (0.001)</td>
<td>0.050 (0.003)</td>
<td>0.097 (0.004)</td>
<td>0.146 (0.005)</td>
</tr>
<tr>
<td>400</td>
<td>True</td>
<td>True</td>
<td>0.012 (0.002)</td>
<td>0.051 (0.003)</td>
<td>0.100 (0.004)</td>
<td>0.147 (0.005)</td>
</tr>
<tr>
<td></td>
<td>True</td>
<td>Est.</td>
<td>0.013 (0.002)</td>
<td>0.054 (0.003)</td>
<td>0.098 (0.004)</td>
<td>0.144 (0.005)</td>
</tr>
<tr>
<td></td>
<td>Est.</td>
<td>Est.</td>
<td>0.013 (0.002)</td>
<td>0.055 (0.003)</td>
<td>0.098 (0.004)</td>
<td>0.142 (0.005)</td>
</tr>
<tr>
<td>$m_1$</td>
<td>$\phi_k(s)$-choice</td>
<td>$\Pi_k(t,t')$-choice</td>
<td>$\alpha = 0.01$</td>
<td>$\alpha = 0.05$</td>
<td>$\alpha = 0.10$</td>
<td>$\alpha = 0.15$</td>
</tr>
<tr>
<td>-------</td>
<td>---------------------</td>
<td>---------------------</td>
<td>-----------------</td>
<td>-----------------</td>
<td>-----------------</td>
<td>-----------------</td>
</tr>
<tr>
<td>100</td>
<td>True</td>
<td>True</td>
<td>0.011 (0.001)</td>
<td>0.055 (0.003)</td>
<td>0.103 (0.004)</td>
<td>0.156 (0.005)</td>
</tr>
<tr>
<td></td>
<td>True</td>
<td>Est.</td>
<td>0.011 (0.002)</td>
<td>0.058 (0.003)</td>
<td>0.110 (0.004)</td>
<td>0.160 (0.005)</td>
</tr>
<tr>
<td></td>
<td>Est.</td>
<td>Est.</td>
<td>0.018 (0.002)</td>
<td>0.066 (0.004)</td>
<td>0.122 (0.005)</td>
<td>0.171 (0.005)</td>
</tr>
<tr>
<td>200</td>
<td>True</td>
<td>True</td>
<td>0.010 (0.001)</td>
<td>0.048 (0.003)</td>
<td>0.100 (0.004)</td>
<td>0.148 (0.005)</td>
</tr>
<tr>
<td></td>
<td>True</td>
<td>Est.</td>
<td>0.011 (0.001)</td>
<td>0.051 (0.003)</td>
<td>0.104 (0.004)</td>
<td>0.156 (0.005)</td>
</tr>
<tr>
<td></td>
<td>Est.</td>
<td>Est.</td>
<td>0.012 (0.002)</td>
<td>0.053 (0.003)</td>
<td>0.106 (0.004)</td>
<td>0.156 (0.005)</td>
</tr>
<tr>
<td>300</td>
<td>True</td>
<td>True</td>
<td>0.010 (0.001)</td>
<td>0.050 (0.003)</td>
<td>0.103 (0.004)</td>
<td>0.154 (0.005)</td>
</tr>
<tr>
<td></td>
<td>True</td>
<td>Est.</td>
<td>0.011 (0.001)</td>
<td>0.050 (0.003)</td>
<td>0.107 (0.004)</td>
<td>0.155 (0.005)</td>
</tr>
<tr>
<td></td>
<td>Est.</td>
<td>Est.</td>
<td>0.011 (0.001)</td>
<td>0.052 (0.003)</td>
<td>0.109 (0.004)</td>
<td>0.155 (0.005)</td>
</tr>
<tr>
<td>400</td>
<td>True</td>
<td>True</td>
<td>0.012 (0.002)</td>
<td>0.054 (0.003)</td>
<td>0.100 (0.004)</td>
<td>0.150 (0.005)</td>
</tr>
<tr>
<td></td>
<td>True</td>
<td>Est.</td>
<td>0.012 (0.002)</td>
<td>0.053 (0.003)</td>
<td>0.102 (0.004)</td>
<td>0.151 (0.005)</td>
</tr>
<tr>
<td></td>
<td>Est.</td>
<td>Est.</td>
<td>0.012 (0.002)</td>
<td>0.056 (0.003)</td>
<td>0.102 (0.004)</td>
<td>0.153 (0.005)</td>
</tr>
</tbody>
</table>
Table C.2: The empirical type I error rates of the ZC-MT and Bootstrap-MT methods based on 5000 simulations. Standard errors are presented in parentheses.

<table>
<thead>
<tr>
<th>Method</th>
<th>$m_i \sim {8, \ldots, 12}$</th>
<th>$m_i \sim {15, \ldots, 20}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$n = 100$</td>
<td>$n = 100$</td>
</tr>
<tr>
<td>ZC-MT</td>
<td>$\alpha = 0.01$</td>
<td>$\alpha = 0.01$</td>
</tr>
<tr>
<td></td>
<td>0.017 (0.002)</td>
<td>0.005 (0.001)</td>
</tr>
<tr>
<td></td>
<td>$\alpha = 0.05$</td>
<td>$\alpha = 0.017$</td>
</tr>
<tr>
<td></td>
<td>0.066 (0.004)</td>
<td>0.017 (0.002)</td>
</tr>
<tr>
<td></td>
<td>$\alpha = 0.10$</td>
<td>$\alpha = 0.066$</td>
</tr>
<tr>
<td></td>
<td>0.120 (0.005)</td>
<td>0.033 (0.002)</td>
</tr>
<tr>
<td></td>
<td>$\alpha = 0.15$</td>
<td>$\alpha = 0.120$</td>
</tr>
<tr>
<td></td>
<td>0.169 (0.005)</td>
<td>0.048 (0.003)</td>
</tr>
<tr>
<td></td>
<td>$n = 200$</td>
<td>$n = 200$</td>
</tr>
<tr>
<td>ZC-MT</td>
<td>$\alpha = 0.01$</td>
<td>$\alpha = 0.01$</td>
</tr>
<tr>
<td></td>
<td>0.017 (0.002)</td>
<td>0.005 (0.001)</td>
</tr>
<tr>
<td></td>
<td>$\alpha = 0.05$</td>
<td>$\alpha = 0.017$</td>
</tr>
<tr>
<td></td>
<td>0.063 (0.003)</td>
<td>0.033 (0.002)</td>
</tr>
<tr>
<td></td>
<td>$\alpha = 0.10$</td>
<td>$\alpha = 0.063$</td>
</tr>
<tr>
<td></td>
<td>0.117 (0.005)</td>
<td>0.035 (0.003)</td>
</tr>
<tr>
<td></td>
<td>$\alpha = 0.15$</td>
<td>$\alpha = 0.117$</td>
</tr>
<tr>
<td></td>
<td>0.168 (0.005)</td>
<td>0.051 (0.003)</td>
</tr>
<tr>
<td></td>
<td>$n = 300$</td>
<td>$n = 300$</td>
</tr>
<tr>
<td>ZC-MT</td>
<td>$\alpha = 0.01$</td>
<td>$\alpha = 0.01$</td>
</tr>
<tr>
<td></td>
<td>0.014 (0.002)</td>
<td>0.004 (0.001)</td>
</tr>
<tr>
<td></td>
<td>$\alpha = 0.05$</td>
<td>$\alpha = 0.014$</td>
</tr>
<tr>
<td></td>
<td>0.065 (0.003)</td>
<td>0.019 (0.002)</td>
</tr>
<tr>
<td></td>
<td>$\alpha = 0.10$</td>
<td>$\alpha = 0.065$</td>
</tr>
<tr>
<td></td>
<td>0.114 (0.004)</td>
<td>0.035 (0.003)</td>
</tr>
<tr>
<td></td>
<td>$\alpha = 0.15$</td>
<td>$\alpha = 0.114$</td>
</tr>
<tr>
<td></td>
<td>0.164 (0.005)</td>
<td>0.051 (0.003)</td>
</tr>
<tr>
<td></td>
<td>$n = 400$</td>
<td>$n = 400$</td>
</tr>
<tr>
<td>ZC-MT</td>
<td>$\alpha = 0.01$</td>
<td>$\alpha = 0.01$</td>
</tr>
<tr>
<td></td>
<td>0.012 (0.002)</td>
<td>0.003 (0.001)</td>
</tr>
<tr>
<td></td>
<td>$\alpha = 0.05$</td>
<td>$\alpha = 0.012$</td>
</tr>
<tr>
<td></td>
<td>0.062 (0.003)</td>
<td>0.017 (0.002)</td>
</tr>
<tr>
<td></td>
<td>$\alpha = 0.10$</td>
<td>$\alpha = 0.062$</td>
</tr>
<tr>
<td></td>
<td>0.111 (0.004)</td>
<td>0.032 (0.002)</td>
</tr>
<tr>
<td></td>
<td>$\alpha = 0.15$</td>
<td>$\alpha = 0.111$</td>
</tr>
<tr>
<td></td>
<td>0.157 (0.005)</td>
<td>0.048 (0.003)</td>
</tr>
<tr>
<td></td>
<td>$m_i \sim {15, \ldots, 20}$</td>
<td>$m_i \sim {15, \ldots, 20}$</td>
</tr>
<tr>
<td>Bootstrap-MT</td>
<td>$\alpha = 0.01$</td>
<td>$\alpha = 0.01$</td>
</tr>
<tr>
<td></td>
<td>0.009 (0.001)</td>
<td>0.007 (0.001)</td>
</tr>
<tr>
<td></td>
<td>$\alpha = 0.05$</td>
<td>$\alpha = 0.009$</td>
</tr>
<tr>
<td></td>
<td>0.049 (0.003)</td>
<td>0.038 (0.003)</td>
</tr>
<tr>
<td></td>
<td>$\alpha = 0.10$</td>
<td>$\alpha = 0.049$</td>
</tr>
<tr>
<td></td>
<td>0.091 (0.004)</td>
<td>0.075 (0.004)</td>
</tr>
<tr>
<td></td>
<td>$\alpha = 0.15$</td>
<td>$\alpha = 0.091$</td>
</tr>
<tr>
<td></td>
<td>0.134 (0.005)</td>
<td>0.117 (0.005)</td>
</tr>
<tr>
<td></td>
<td>$n = 200$</td>
<td>$n = 200$</td>
</tr>
<tr>
<td>Bootstrap-MT</td>
<td>$\alpha = 0.01$</td>
<td>$\alpha = 0.01$</td>
</tr>
<tr>
<td></td>
<td>0.007 (0.001)</td>
<td>0.006 (0.001)</td>
</tr>
<tr>
<td></td>
<td>$\alpha = 0.05$</td>
<td>$\alpha = 0.007$</td>
</tr>
<tr>
<td></td>
<td>0.038 (0.003)</td>
<td>0.028 (0.002)</td>
</tr>
<tr>
<td></td>
<td>$\alpha = 0.10$</td>
<td>$\alpha = 0.038$</td>
</tr>
<tr>
<td></td>
<td>0.078 (0.004)</td>
<td>0.055 (0.003)</td>
</tr>
<tr>
<td></td>
<td>$\alpha = 0.15$</td>
<td>$\alpha = 0.078$</td>
</tr>
<tr>
<td></td>
<td>0.121 (0.005)</td>
<td>0.084 (0.004)</td>
</tr>
<tr>
<td></td>
<td>$n = 300$</td>
<td>$n = 300$</td>
</tr>
<tr>
<td>Bootstrap-MT</td>
<td>$\alpha = 0.01$</td>
<td>$\alpha = 0.01$</td>
</tr>
<tr>
<td></td>
<td>0.008 (0.001)</td>
<td>0.007 (0.001)</td>
</tr>
<tr>
<td></td>
<td>$\alpha = 0.05$</td>
<td>$\alpha = 0.008$</td>
</tr>
<tr>
<td></td>
<td>0.039 (0.003)</td>
<td>0.034 (0.003)</td>
</tr>
<tr>
<td></td>
<td>$\alpha = 0.10$</td>
<td>$\alpha = 0.039$</td>
</tr>
<tr>
<td></td>
<td>0.081 (0.004)</td>
<td>0.078 (0.004)</td>
</tr>
<tr>
<td></td>
<td>$\alpha = 0.15$</td>
<td>$\alpha = 0.081$</td>
</tr>
<tr>
<td></td>
<td>0.122 (0.005)</td>
<td>0.104 (0.004)</td>
</tr>
<tr>
<td></td>
<td>$n = 400$</td>
<td>$n = 400$</td>
</tr>
<tr>
<td>Bootstrap-MT</td>
<td>$\alpha = 0.01$</td>
<td>$\alpha = 0.01$</td>
</tr>
<tr>
<td></td>
<td>0.007 (0.001)</td>
<td>0.005 (0.001)</td>
</tr>
<tr>
<td></td>
<td>$\alpha = 0.05$</td>
<td>$\alpha = 0.007$</td>
</tr>
<tr>
<td></td>
<td>0.038 (0.003)</td>
<td>0.027 (0.002)</td>
</tr>
<tr>
<td></td>
<td>$\alpha = 0.10$</td>
<td>$\alpha = 0.038$</td>
</tr>
<tr>
<td></td>
<td>0.078 (0.004)</td>
<td>0.055 (0.003)</td>
</tr>
<tr>
<td></td>
<td>$\alpha = 0.15$</td>
<td>$\alpha = 0.078$</td>
</tr>
<tr>
<td></td>
<td>0.121 (0.005)</td>
<td>0.087 (0.004)</td>
</tr>
</tbody>
</table>

C.2 Validating the DTI Testing Results via Simulation Study

We generate data from $Y_{ij}(s) = \mu(s, t_{ij}) + \sum_{k=1}^{10} w_{i,k}(t_{ij})\hat{\phi}_k(s) + E_{ij}(s)$, where $\hat{\phi}_k(s)$’s are the estimated marginal eigenfunctions from the DTI data analysis. The time-varying coefficients are generated from the random effects model: $w_{i,k}(t) = b_{ik0} + b_{ik1}t$, where the random coefficients ($b_{ik0}, b_{ik1}$) are sampled from multivariate normal with mean zero and covariance equal to the corresponding estimated covariance from the DTI data analysis. Similarly errors $E_{ij}(s)$ are generated from the model $E_{ij}(s) = \sum_{k=1}^{10} e_{ijk}\hat{\phi}_k(s) + \epsilon_{ij}(s)$ where $e_{ijk}$ and $\epsilon_{ij}(s)$ are independently sampled from normal with mean zero and vari-
ances equal to the respective estimated variances from the DTI data analysis. We use $\mu(s, t) = \hat{\mu}_0(s)$ to study for size, where $\hat{\mu}_0(s)$ is obtained by fitting 20 univariate cubic B-splines to the DTI data. Figure C.1 shows one simulated sample that mimics the DTI error structure and has a mean function equal to $\hat{\mu}(s, t)$, i.e. $\delta = 1$.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{fig_c1.png}
\caption{The original DTI data (left) and one simulated data that mimic the DTI error structure and have a mean function equal to $\hat{\mu}(s, t)$. For illustration the observed profiles from the first subject are highlighted with colors.}
\end{figure}

Table C.3 presents the type 1 error rates corresponding to the nominal level $\alpha = 0.05$; the results are based on 5000 simulations.

Table C.3: The empirical type I error rates of the proposed test and other competitive methods based on 5000 simulations, using nominal level $\alpha = 0.05$. Standard errors are presented in parentheses.

<table>
<thead>
<tr>
<th></th>
<th>pLRT-MT</th>
<th>Bootstrap</th>
<th>Bootstrap-MT</th>
<th>Bootstrap-L</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.064</td>
<td>0.059</td>
<td>0.014</td>
<td>0.050</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.002)</td>
<td>(0.003)</td>
</tr>
</tbody>
</table>