

STATISTICAL INFERENCE FOR FUNCTIONAL AND LONGITUDINAL DATA

By

Guanqun Cao

A DISSERTATION

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

DOCTOR OF PHILOSOPHY

Statistics

2012

ABSTRACT

STATISTICAL INFERENCE FOR FUNCTIONAL AND LONGITUDINAL DATA

By

Guanqun Cao

Advances in modern technology have facilitated the collection of high-dimensional functional and low dimensional longitudinal data. For these data, it is often of interest to describe the key signals of the data (mean functions, covariance functions, derivative functions, etc.). Functional data analysis (FDA) and longitudinal data analysis (LDA) techniques have played a central role in the analysis of these data. The primary goal of this dissertation is to provide some novel statistical inference methods for FDA and LDA.

In Chapter 1, we describe the structure (design, notations, etc.) of functional data and describe the spline smoothing technique as a tool to analysis these data. Longitudinal data analysis with missing not at random response is also discussed.

In Chapter 2, a polynomial spline estimator is proposed for the mean function of dense functional data together with a simultaneous confidence band which is asymptotically correct. In addition, the spline estimator and its accompanying confidence band enjoy semi-parametric efficiency in the sense that they are asymptotically the same as if all random trajectories are observed entirely and without errors. The confidence band is also extended to the difference of mean functions of two populations of functional data. Simulation experiments provide strong evidence that corroborates the asymptotic theory while computing is efficient. The confidence band procedure is illustrated by analyzing the near infrared spectroscopy data.

A nonparametric estimation of the covariance function for dense functional data using tensor product B-splines is considered in Chapter 3. We develop both local and global asymptotic distributions for the proposed estimator, and show that our estimator is as efficient as an “oracle” estimator. Monte Carlo simulation experiments and two real data examples are also provided to illustrate the proposed method in this chapter.

In Chapter 4, we develop a new procedure to construct simultaneous confidence bands for derivatives of mean curves in FDA. The technique involves polynomial splines that provide an approximation to the derivatives of the mean functions, the covariance functions and the associated eigenfunctions. The confidence band procedure is illustrated through numerical simulation studies and a real life example.

In Chapter 5, we consider data generated from a longitudinal study with potentially non random missing data. For these data, a joint model for the missing data process and the outcome process, is found to be at best weakly identifiable. Due to this identifiability concerns, tests concerning the parameters of interest may not be able to use conventional theories and it may not be clear how to assess statistical significance. We extend the literature by developing a testing procedure that can be used to evaluate hypotheses under non and weakly identifiable semiparametric models. We derive the limiting distribution of this statistic and propose theoretically justified resampling approaches to approximate its asymptotic distribution. The methodology’s practical utility is illustrated in simulations and an analysis of quality-of-life outcomes from a longitudinal study on breast cancer.

To my beloved parents

ACKNOWLEDGMENTS

Foremost, I would like to express my sincere gratitude to my advisor Professor Lijian Yang for the continuous support of my Ph.D study and research, for his patience, motivation, enthusiasm, and immense knowledge. His guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better mentor for my Ph.D study.

I would like to express my gratitude to my co-advisor Professor David Todem, for offering invaluable assistance, support and guidance. He provided me a fabulous opportunity to study biostatistics topics. Moreover, I could not finish my job application promptly without his generous help.

Besides my advisors, I would like to thank the rest of my thesis committee: Professors Lyudmila Sakhanenko and Lifeng Wang for their hard questions, encourage and insightful comments. My sincere thanks also goes to Professor Jiaguo Qi in the Center for Global Change and Earth Observations at MSU, for offering me the collaboration opportunities in his group and leading me working on diverse exciting projects.

Thanks also goes to the entire faculty and staff members in the Department of Statistics and Probability who have taught me and help me during my study at MSU. My special thanks goes to Professor James Stapleton, for listening to my presentation rehearsal many times and helpful suggestions which leads to my successful presentations.

Thanks to the graduate school and the Department of Statistics and Probability who provided me with the Dissertation Continuation Fellowship and Dissertation Completion Fellowship for working on the dissertation. This dissertation is also supported in part by NSF award DMS 0706518, 1007594 and NCI/NIH K-award, 1K01 CA131259.

I also would like to thank my academic family members: Dr. Lan Xue, Dr. Jing Wang,

Dr. Li Wang, Dr. Rong Liu, Dr. Qionxia Song and Shuzhuan Zheng for the stimulating discussions, for the sleepless nights we were working together before submitting papers, and for all the fun we have had in the last five years.

Last but not the least, I would like to thank my beloved parents Jinxiang Cao and Lazhen Sun, for giving birth to me at the first place and supporting me spiritually throughout my life.

TABLE OF CONTENTS

List of Tables	ix
List of Figures	xi
Chapter 1 Introduction	1
1.1 Functional data analysis	1
1.1.1 The basics	1
1.1.2 Polynomial spline	2
1.1.3 The near infrared spectroscopy data	4
1.1.4 Speech recognition data	4
1.2 Longitudinal data analysis	5
1.2.1 Problems	5
1.2.2 The breast cancer data	6
Chapter 2 Confidence Bands for Mean Functions	7
2.1 Introduction	7
2.2 Main results	11
2.3 Error decomposition	17
2.4 Implementation	20
2.5 Simulation	21
2.6 Empirical example	31
Chapter 3 Confidence Envelopes for Covariance Functions	47
3.1 Introduction	47
3.2 Spline covariance estimation	50
3.2.1 Data structure and model assumptions	50
3.2.2 Spline covariance estimator	51
3.3 Asymptotic theory and simultaneous confidence envelopes	54
3.3.1 Assumptions and the oracle property	54
3.3.2 Asymptotic confidence envelopes	56
3.3.3 Extension to two-sample test problems	58
3.4 Implementation	59
3.5 Simulation	61

3.6	Empirical examples	65
3.6.1	Tecator near infrared spectra data	65
3.6.2	Speech recognition data	66
3.7	Summary	71
Chapter 4	Spline Confidence Bands for Functional Derivatives	96
4.1	Introduction	96
4.2	Models and spline estimators	98
4.2.1	Models	98
4.2.2	Spline estimators	99
4.2.3	Convergence rate	101
4.3	Confidence bands	103
4.3.1	Asymptotic confidence bands	103
4.3.2	Implementation	104
4.4	Numerical studies	107
4.4.1	Simulated examples	107
4.4.2	Tecator data	117
Chapter 5	Testing Hypotheses Under Weakly Identifiability	137
5.1	Introduction	137
5.2	The method	143
5.2.1	The general framework	143
5.2.2	Large sample properties of $\hat{\theta}(\beta)$	144
5.2.3	Global sensitivity testing	147
5.3	Numerical studies	153
5.3.1	Pseudo-likelihood models with missing data	153
5.3.2	Real data analysis	156
5.3.3	Simulation study	160
5.4	Discussion	165
Bibliography		173

LIST OF TABLES

Table 2.1	Coverage frequencies from 500 replications using linear spline (??) with $p = 2$, $N_m = \lceil cn^{1/(2p)\log(n)} \rceil$ and $\sigma = 0.5$	25
Table 2.2	Coverage frequencies from 500 replications using linear spline (??) with $p = 2$, $N_m = \lceil cn^{1/(2p)\log(n)} \rceil$ and $\sigma = 0.3$	25
Table 2.3	Coverage frequencies from 500 replications using cubic spline (??) with $p = 4$, $N_m = \lceil cn^{1/(2p)\log(n)} \rceil$ and $\sigma = 0.5$	26
Table 2.4	Coverage frequencies from 500 replications using cubic spline (??) with $p = 4$, $N_m = \lceil cn^{1/(2p)\log(n)} \rceil$ and $\sigma = 0.3$	27
Table 2.5	Coverage frequencies from 500 replications using least squares Bonferroni band and least squares Bootstrap band.	28
Table 2.6	Five number summary of ratios of confidence band widths.	29
Table 2.7	Empirical power and type I error of two-sample test using cubic spline.	30
Table 3.1	Simulation results: uniform coverage rates from 500 replications.	95
Table 4.1	Uniform coverage rates using cubic spline in Model I	108
Table 4.2	Uniform coverage rates using cubic spline in Model II	117
Table 5.1	P-values for evaluating the Treatment and Time effects using data from Study VI of the IBCSG trials	159
Table 5.2	Empirical type I error and power of the infimum test and Wald test	164

Table 5.3	Empirical power of the infimum test to detect the interaction effect $\tilde{\alpha}_3 = 1$ for two ranges Ξ of β with true value being $\beta_0 = -1$	166
Table 5.4	Empirical power of the infimum test to detect the interaction effect $\tilde{\alpha}_3 = 0.7$ for two sets Ξ of β with true value being $\beta_0 = -1$	166

LIST OF FIGURES

Figure 2.1	Plots of the linear spline estimator for simulated data and 95% confidence bands for $m(x)$	23
Figure 2.2	Plots of the cubic spline estimator for simulated data and 95% confidence bands for $m(x)$	24
Figure 2.3	Left: Plot of Tecator data. Right: Sample curves for the Tecator data. Each class has 10 sample curves.	32
Figure 2.4	Plots of the fitted linear and cubic spline regressions of $m_1(x) - m_2(x)$ and confidence bands	34
Figure 3.1	Plots of the true covariance functions of the simulated data and their 95% confidence envelopes: $n=200, N=100, \sigma=0.5. p_1 = p_2 = 2$. . .	63
Figure 3.2	Plots of the true covariance functions of the simulated data and their 95% confidence envelopes: $n=200, N=100, \sigma=0.5. p_1 = p_2 = 4$. . .	64
Figure 3.3	Plot of the Tecator data	67
Figure 3.4	Plots of the cubic tensor spline estimator for the Tecator data	68
Figure 3.5	Plots of hypothesis test (3.12) results	69
Figure 3.6	Plots of the speech recognition data	71
Figure 3.7	Plots of tensor spline estimators for speech recognition data sets . .	72
Figure 3.8	Plots of hypothesis test (3.13) results	73

Figure 4.1	Plots of the cubic spline estimators and 99% confidence bands of $m^{(1)}(t)$ in Model I. $n = 30, N = 30$	109
Figure 4.2	Plots of the cubic spline estimators and 99% confidence bands of $m^{(1)}(t)$ in Model I. $n = 30, N = 60$	110
Figure 4.3	Plots of the cubic spline estimators and 99% confidence bands of $m^{(1)}(t)$ in Model I. $n = 50, N = 50$	111
Figure 4.4	Plots of the cubic spline estimators and 99% confidence bands of $m^{(1)}(t)$ in Model I. $n = 50, N = 100$	112
Figure 4.5	Plots of the cubic spline estimators and 99% confidence bands of $m^{(1)}(t)$ in Model II. $n = 30, N = 30$	113
Figure 4.6	Plots of the cubic spline estimators and 99% confidence bands of $m^{(1)}(t)$ in Model II. $n = 30, N = 60$	114
Figure 4.7	Plots of the cubic spline estimators and 99% confidence bands of $m^{(1)}(t)$ in Model II. $n = 50, N = 50$	115
Figure 4.8	Plots of the cubic spline estimators and 99% confidence bands of $m^{(1)}(t)$ in Model II. $n = 50, N = 100$	116
Figure 4.9	Plots of the cubic spline estimators and 99% confidence bands . . .	119
Figure 5.1	Supremum of the pseudo-likelihood function profiled across β	140
Figure 5.2	Plot of the exact and the resampled CDF	152
Figure 5.3	Simultaneous confidence bands for time effects on PACIS and Mood	161

Chapter 1

Introduction

1.1 Functional data analysis

1.1.1 The basics

Functional data analysis (FDA) has recently become a focal area in statistical research, as recent technological progress in measuring devices now allow one to observe spatiotemporal phenomena on arbitrarily fine grids, that is, almost in a continuous manner. This area remains distinct due to its usefulness to climatology, medicine, meteorology, etc.

The functional data that we consider are a collection of trajectories $\{\eta_i(x)\}_{i=1}^n$ which are i.i.d. realizations of a smooth random function $\eta(x)$, defined on a continuous interval $\mathcal{X}$. Assume that $\{\eta(x), x \in \mathcal{X}\}$ is an L^2 process, i.e. $E \int_{\mathcal{X}} \eta^2(x) dx < +\infty$, and define the mean and covariance functions $m(x) = E\{\eta(x)\}$ and $G(x, x') = \text{cov}\{\eta(x), \eta(x')\}$. The covariance function is a symmetric nonnegative-definite function with a spectral decomposition, $G(x, x') = \sum_{k=1}^{\kappa} \lambda_k \psi_k(x) \psi_k(x')$, where $\lambda_1 \geq \lambda_2 \geq \dots \geq 0$, $\sum_{k=1}^{\kappa} \lambda_k < +\infty$, are the eigenvalues, κ is either a positive integer or infinity and $\{\psi_k(x)\}_{k=1}^{\kappa}$ are the corresponding

eigenfunctions and form a set of orthonormal functions in $L^2(\mathcal{X})$. By the Karhunen-Loève representation, $\eta_i(x) = m(x) + \sum_{k=1}^{\kappa} \xi_{ik} \phi_k(x)$, where the random coefficients ξ_{ik} are uncorrelated with mean 0 and variance 1, and the functions $\phi_k = \sqrt{\lambda_k} \psi_k$.

Characterizing nonlinear variation in FDA is a challenging problem. In particular, when random curves are observed on regular dense grids, the existing literature on FDA focuses on pointwise estimation and inference. This, however, is not sufficient to provide understanding of the variability of the estimator of the whole regression curve, its derivatives and covariance function, nor can it be used to correctly answer questions about the curve's or surface's shape.

In Chapters 2, 3 and 4, we propose oracle estimators and establish asymptotic correctness of the proposed simultaneous confidence bands/envelopes for mean, covariance and its derivative functions while the number of observations for each subject tends to infinite as sample size goes to infinite, using various properties of spline smoothing.

1.1.2 Polynomial spline

To describe the spline functions, we first introduce a sequence of equally-spaced points $\{t_J\}_{J=1}^{N_m}$, called interior knots which divide the interval $[0, 1]$ into $(N_m + 1)$ equal subintervals $I_J = [t_J, t_{J+1})$, $J = 0, \dots, N_m - 1$, $I_{N_m} = [t_{N_m}, 1]$. For any positive integer p , introduce left boundary knots $t_{1-p}, \dots, t_0$, and right boundary knots $t_{N_m+1}, \dots, t_{N_m+p}$,

$$t_{1-p} = \dots = t_0 = 0 < t_1 < \dots < t_{N_m} < 1 = t_{N_m+1} = \dots = t_{N_m+p},$$

$$t_J = Jh_m, \quad 0 \leq J \leq N_m + 1, \quad h_m = 1/(N_m + 1),$$

in which h_m is the distance between neighboring knots. Denote by $\mathcal{H}^{(p-2)}$ the space of p -th order spline space, i.e., $p - 2$ times continuously differentiable functions on $[0, 1]$ that

are polynomials of degree $p - 1$ on $[t_J, t_{J+1}]$, $J = 0, \dots, N_m$.

The J -th B-spline of order p is

$$B_{J,p}(x) = \frac{(x - t_J) B_{J,p-1}(x)}{t_{J+p-1} - t_J} - \frac{(x - t_{J+p}) B_{J+1,p-1}(x)}{t_{J+p} - t_{J+1}}, \quad 1 - p \leq J \leq N_m,$$

for $p > 1$, with

$$B_{J,1}(x) = I_J(x) = \begin{cases} 1 & t_J \leq x < t_{J+1} \\ 0 & \text{otherwise.} \end{cases}$$

Hence, the B-spline basis of $\mathcal{H}^{(-1)}$, the space of constant splines, are indicator functions of interval I_J , $B_{J,1}(x) = I_J(x)$, $J = 0, \dots, N_m$, while the B-spline basis of $\mathcal{H}^{(0)}$, the space of linear splines, are $B_{J,2}(x) = K \left\{ (x - t_{J+1}) h_m^{-1} \right\}$, $J = -1, \dots, N_m$, where $K(u) = (1 - |u|)_+$.

In this work, $\mathcal{X} = [0, 1]$ or $[0, 1]^2$. For any function ϕ on a domain $\mathcal{X}$, let $\|\phi\|_\infty = \sup_{\mathbf{x} \in \mathcal{X}} |\phi(\mathbf{x})|$. For any Lebesgue measurable L_2 -integrable functions ϕ and φ on $[0, 1]$, define their theoretical and empirical inner products as $\langle \phi, \varphi \rangle = \int_{\mathcal{X}} \phi(x) \varphi(x) dx$ and $\langle \phi, \varphi \rangle_{2,N} = N^{-1} \sum_{j=1}^N \phi(j/N) \varphi(j/N)$, with the corresponding theoretical and empirical inner product norms defined as $\|\phi\|_2^2 = \int_0^1 \phi^2(x) dx$ and $\|\phi\|_{2,N}^2 = N^{-1} \sum_{j=1}^N \phi^2(j/N)$, respectively. The theoretical and empirical inner product matrices of $\{B_{J,p}(x)\}_{J=1-p}^{N_m}$ are denoted as

$$\mathbf{V}_p = \left(\langle B_{J,p}, B_{J',p} \rangle \right)_{J,J'=1-p}^{N_m}, \hat{\mathbf{V}}_p = \left(\langle B_{J,p}, B_{J',p} \rangle_{2,N} \right)_{J,J'=1-p}^{N_m}.$$

1.1.3 The near infrared spectroscopy data

The near infrared spectroscopy data are recorded on a Tecator Infrared Food and Feed Analyzer working in the wavelength range 850 - 1050 nm by the Near Infrared Transmission (NIT) principle. Each sample contains finely chopped pure meat with different moisture, fat and protein contents. The task is to predict the fat content of a meat sample on the basis of its near infrared absorbance spectrum.

The spectral data can be naturally considered as functional data as the number of observation points for each trajectory is relatively large compared with the sample size. In Chapter 2, we will perform a two-sample test to evaluate whether absorbance from the spectrum differs significantly due to difference in fat content. In Chapter 4, we will study the behavior of the first order derivative of mean absorbance measurements for this data set.

1.1.4 Speech recognition data

Speech recognition data were extracted from the TIMIT database (TIMIT Acoustic-Phonetic Continuous Speech Corpus, NTIS, US Dept of Commerce) which is a widely used resource for research in speech recognition. The data set we use was formed by selecting five phonemes for classification based on digitized speech from this database. From continuous speech of 50 male speakers, 4509 speech frames of 32 msec duration were selected. From each speech frame, a log-periodogram was used as transformation for casting speech data in a form suitable for speech recognition. The five phonemes in this data set are transcribed as follows: “sh” as in “she”, “dcl” as in “dark”, “iy” as the vowel in “she”, “aa” as the vowel in “dark”, and “ao” as the first vowel in “water”. For illustration purpose, we focus on the “sh” and “ao” phoneme classes as representatives of consonants and vowels. There are

$n_1 = 872$ log-periodograms in the “sh” class, and $n_2 = 1022$ log-periodograms in the “ao” group. Each log-periodogram consists $N = 256$ equally spaced points.

This data set was first analyzed by Hastie et al. (1995) using penalized linear discriminant analysis. One of the basic assumptions is that the covariance functions are the same for different classes. In Chapter 3, we would like to test the equal covariance assumption formally by the proposed tensor product spline confidence envelopes.

1.2 Longitudinal data analysis

1.2.1 Problems

Identifiability issues commonly arise with non-random missing data, where the parameters in the model for the missingness may not be jointly identifiable with those in the model for the outcomes of interest using only the observed data. Analyses which assume identifiability may be unreliable, with the joint selection and outcome model yielding flat “estimation” surfaces potentially having multiple modes.

In Chapter 5, we consider these missing data issues when analyzing longitudinal data with informative dropout employing the model of Troxel, Lipsitz and Harrington (1998b). The model is semiparametric, with the parameter being estimated denoted by (θ, β) . The parameter β is the selection parameter that measures the extent of non-randomness of the missing data mechanism and θ consists of the remaining finite dimensional parameters of the selection and outcome models. The hypotheses of interest concern covariate effects on the outcome, which are contained in θ . We extend the profiling idea to arbitrary estimating functions involving θ and β but which do not require a complete parametric model specifica-

tion. The generalization of the infimum test and confidence bands to non-likelihood settings is nontrivial. We present generic conditions which establish the large sample properties of the estimating function for θ profiled on β , including the uniform consistency and weak convergence of the θ estimator as a function of β . Owing to the complexity of the asymptotic distributions of the infimum test and confidence bands, resampling is needed. A theoretically justified procedure is discussed for approximating such distributions.

1.2.2 The breast cancer data

These breast cancer data are from the International Breast Cancer Study Group-IBCSG, previously reported by Hürny et al. (1992); and Troxel et al. (1998b). This is a group of randomized breast cancer studies with primary endpoints being survival and relapse; and quality of life being a secondary endpoint. One study, Study VI, is a randomized trial of adjuvant chemotherapy following surgery for the treatment of breast cancer. In this study, 4 treatments (A, B, C and D) were randomly assigned to 431 pre-menopausal cancer patients and several domains of quality of life were assessed. In this thesis, we focus on three quality-of-life domains; 1) PACIS (perceived adjustment to chronic illness scale), 2) Mood and 3) Appetite. A full description of Study VI and other IBCSG trials may be found elsewhere (Hürny et al., 1992; Troxel et al., 1998a). In Chapter 5, we assess the treatment and time effects on the mean quality of life by apply the proposed test and resampling procedure.

Chapter 2

Confidence Bands for Mean Functions

2.1 Introduction

In functional data analysis problems, estimation of mean function is the fundamental first step; see Cardot (2000); Rice and Wu (2001); Cuevas, Febrero and Frainman (2006); Ferraty and Vieu (2006); Degras (2011) and Ma, Yang and Carroll (2012) for example. According to Ramsay and Silverman (2005), functional data consist of a collection of i.i.d. realizations $\{\eta_i(x)\}_{i=1}^n$ of a smooth random function $\eta(x)$, with unknown mean function $E\eta(x) = m(x)$ and covariance function $G(x, x') = \text{cov}\{\eta(x), \eta(x')\}$. Although the domain of $\eta(\cdot)$ is an entire interval $\mathcal{X}$, the recording of each random curve $\eta_i(x)$ is only over a finite number N_i of points in $\mathcal{X}$, and contaminated with measurement errors. Without loss of generality, we take $\mathcal{X} = [0, 1]$.

Denote by Y_{ij} the j -th observation of the random curve $\eta_i(\cdot)$ at time point X_{ij} , $1 \leq i \leq n, 1 \leq j \leq N_i$. Although we refer to variable X_{ij} as time, it could also be other numerical measures, such as wavelength in Section 3.6. In this work, we examine the equally spaced

dense design, in other words, $X_{ij} = j/N, 1 \leq i \leq n, 1 \leq j \leq N$ with N going to infinity. For the i -th subject, $i = 1, 2, \dots, n$, its sample path $\{j/N, Y_{ij}\}$ is the noisy realization of the continuous time stochastic process $\eta_i(x)$ in the sense that $Y_{ij} = \eta_i(j/N) + \sigma(j/N) \varepsilon_{ij}$, with errors ε_{ij} satisfying $E(\varepsilon_{ij}) = 0$, $E(\varepsilon_{ij}^2) = 1$, and $\{\eta_i(x), x \in [0, 1]\}$ are i.i.d. copies of the process $\{\eta(x), x \in [0, 1]\}$ satisfying $E \int_{[0,1]} \eta^2(x) dx < +\infty$.

For the standard process $\{\eta(x), x \in [0, 1]\}$, let sequences $\{\lambda_k\}_{k=1}^\infty$, $\{\psi_k(x)\}_{k=1}^\infty$ be the eigenvalues and eigenfunctions of $G(x, x')$ respectively, in which $\lambda_1 \geq \lambda_2 \geq \dots \geq 0$, $\sum_{k=1}^\infty \lambda_k < \infty$, $\{\psi_k\}_{k=1}^\infty$ form an orthonormal basis of $L^2([0, 1])$ and covariance function $G(x, x') = \sum_{k=1}^\infty \lambda_k \psi_k(x) \psi_k(x')$, which implies that $\int G(x, x') \psi_k(x') dx' = \lambda_k \psi_k(x)$. The process $\{\eta_i(x), x \in [0, 1]\}$ has the Karhunen-Loève L^2 representation $\eta_i(x) = m(x) + \sum_{k=1}^\infty \xi_{ik} \phi_k(x)$, where the random coefficients ξ_{ik} are uncorrelated with mean 0 and variance 1, and $\phi_k = \sqrt{\lambda_k} \psi_k$. In what follows, we assume that $\lambda_k = 0$, for $k > \kappa$, where κ is a positive integer or ∞ , thus $G(x, x') = \sum_{k=1}^\kappa \phi_k(x) \phi_k(x')$ and the model that we consider is

$$Y_{ij} = m(j/N) + \sum_{k=1}^\kappa \xi_{ik} \phi_k(j/N) + \sigma(j/N) \varepsilon_{ij}. \quad (2.1)$$

Although the sequences $\{\lambda_k\}_{k=1}^\kappa$, $\{\phi_k(\cdot)\}_{k=1}^\kappa$ and the random coefficients ξ_{ik} exist mathematically, they are unknown or unobservable.

The existing literature focuses on two data types. Yao, Müller and Wang (2005) studied sparse longitudinal data for which N_i , i.e. the number of observations for the i -th curve, is bounded and follows a given distribution, in which case Ma, Yang and Carroll (2012) obtained asymptotically simultaneous confidence band for the mean function of the functional data, using piecewise constant spline estimation. Li and Hsing (2010a) established uniform convergence rate for local linear estimators of mean and covariance function of dense functional

data, where $\min_{1 \leq i \leq n} N_i \gg (n/\log n)^{1/4}$, as $n \rightarrow \infty$, similar to our Assumption (A3), but did not provide asymptotic distribution of maximal deviation or simultaneous confidence bands. Degras (2011) built asymptotically correct simultaneous confidence bands for dense functional data using local linear estimators. Bunea, Ivanescu and Wegkamp (2011) proposed asymptotically conservative rather than correct confidence set for the mean function of Gaussian functional data.

In this chapter, we propose polynomial spline confidence bands for the mean function based on dense functional data. In function estimation problems, simultaneous confidence bands are an important tool to address the variability in the mean curve, see Zhao and Wu (2008); Zhou, Shen and Wolfe (1998) and Zhou and Wu (2010) for related theory and applications. The fact that simultaneous confidence bands have not been widely used for functional data analysis is certainly not due to lack of interesting applications, but due to the greater technical difficulty to formulate such bands for functional data and establish their theoretical properties. In this work, we have established asymptotic correctness of the proposed confidence bands using various properties of spline smoothing. The spline estimators and the accompanying confidence bands are asymptotically the same as if all the n random curves are recorded over the entire interval, without measurement errors. They are oracally efficient despite the use of spline smoothing, see Remark 2.2.1. This provides partial theoretical justification for treating functional data as perfectly recorded random curves over the entire data range, as in Ferraty and Vieu (2006). Theorem 3 of Hall, Müller and Wang (2006) stated mean square (rather than the stronger uniform) oracle efficiency for local linear estimation of eigenfunctions and eigenvalues (rather than the mean function), under assumptions similar to ours, but provided only an outline of proof. Among

the existing works on functional data analysis, Ma, Yang and Carroll (2012) proposed the simultaneous confidence bands for sparse functional data. However, their result does not enjoy the oracle efficiency stated in Theorem 2.2.1, since there are not enough observations for each subject to obtain a good estimate of the individual trajectories. As a result, it has the slow nonparametric convergence rate of $n^{-1/3}\log n$, instead of the parametric rate of $n^{-1/2}$ as in the present work. This essential difference completely separates dense functional data from sparse ones.

The aforementioned confidence bands are also extended to the difference of two regression functions. This is motivated by Li and Yu (2008), who applied functional segment discriminant analysis to a Tecator data set, see Figure 2.3. In this data set, each observation (meat) consists of a 100-channel absorbance spectrum in the wavelength with different fat, water and protein percent. Li and Yu (2008) used the spectra to predict whether the fat percentage is greater than 20%. On the flip side, we are interested in building a $100(1 - \alpha)\%$ confidence band for the difference between regression functions from the spectra of the less than 20% fat group and the higher than 20% fat group. If this $100(1 - \alpha)\%$ confidence band covers the zero line, one accepts the null hypothesis of no difference between the two groups, with p-value no greater than α . Test for equality between two groups of curves based on the adaptive Neyman test and wavelet thresholding techniques were proposed in Fan and Lin (1998), who did not provide an estimator of the difference of the two mean functions nor a simultaneous confidence band for such estimator. As a result, their test did not extend to testing other important hypotheses on the difference of the two mean functions while our Theorem 2.2.3 provides a benchmark for all such testing. More recently, Benko, Härdle and Kneip (2009) developed two-sample bootstrap tests for the equality of eigenfunctions, eigen-

values and mean functions by using common functional principal components and bootstrap tests.

This chapter is organized as follows. Section 2.2 states main theoretical results on confidence bands constructed from polynomial splines. Section 2.3 provides further insights into the error structure of spline estimators. The actual steps to implement the confidence bands are provided in Section 2.4. A simulation study is presented in Section 2.5, and an empirical illustration on how to use the proposed spline confidence band for inference is reported in Section 2.6. Technical proofs are collected in the Appendix.

2.2 Main results

To describe the spline functions, we first define a sequence of equally-spaced points $\{t_J\}_{J=1}^{N_m}$, called interior knots, which have been introduced in Chapter 1. Denote by $\mathcal{H}^{(p-2)}$ the space of p -th order spline space, i.e., $p - 2$ times continuously differentiable functions on $[0, 1]$ that are polynomials of degree $p - 1$ on $[t_J, t_{J+1}]$, $J = 0, \dots, N_m$. Then $\mathcal{H}^{(p-2)} = \left\{ \sum_{J=1}^{N_m} b_{J,p} B_{J,p}(x), b_{J,p} \in \mathcal{R}, x \in [0, 1] \right\}$, where $B_{J,p}$ is the J -th B-spline basis of order p as defined in de Boor (2001).

We propose to estimate the mean function $m(x)$ by

$$\hat{m}_p(x) = \underset{g(\cdot) \in \mathcal{H}^{(p-2)}}{\operatorname{argmin}} \sum_{i=1}^n \sum_{j=1}^N \left\{ Y_{ij} - g(j/N) \right\}^2. \quad (2.2)$$

The technical assumptions we need are as follows:

(A1) *The regression function $m \in C^{p-1,1}[0, 1]$, i.e., $m^{(p-1)} \in C^{0,1}[0, 1]$.*

(A2) *The standard deviation function $\sigma(x) \in C^{0,\mu}[0, 1]$ for some $\mu \in (0, 1]$.*

(A3) As $n \rightarrow \infty$, $N^{-1}n^{1/(2p)} \rightarrow 0$ and $N = O(n^\theta)$ for some $\theta > 1/(2p)$; the number of interior knots N_m satisfies $NN_m^{-1} \rightarrow \infty$, $N_m^{-p}n^{1/2} \rightarrow 0$, $N^{-1/2}N_m^{1/2}\log n \rightarrow 0$ or equivalently $Nh_m \rightarrow \infty$, $h_m^p n^{1/2} \rightarrow 0$, $N^{-1/2}h_m^{-1/2}\log n \rightarrow 0$.

(A4) There exists $C_G > 0$ such that $G(x, x) \geq C_G$, $x \in [0, 1]$; for $k \in \{1, \dots, \kappa\}$, $\phi_k(x) \in C^{0,\mu}[0, 1]$, $\sum_{k=1}^{\kappa} \|\phi_k\|_\infty < \infty$ and as $n \rightarrow \infty$, $h_m^\mu \sum_{k=1}^{\kappa n} \|\phi_k\|_{0,\mu} = o(1)$ for a sequence $\{\kappa_n\}_{n=1}^\infty$ of increasing integers, with $\lim_{n \rightarrow \infty} \kappa_n = \kappa$ and the constant $\mu \in (0, 1]$ as in Assumption (A2). In particular, $\sum_{k=\kappa_n+1}^{\kappa} \|\phi_k\|_\infty = o(1)$.

(A5) There are constants $C_1, C_2 \in (0, +\infty)$, $\gamma_1, \gamma_2 \in (1, +\infty)$, $\beta \in (0, 1/2)$ and i.i.d. $N(0, 1)$ variables $\left\{Z_{ik,\xi}\right\}_{i=1,k=1}^{n,\kappa}$, $\left\{Z_{ij,\varepsilon}\right\}_{i=1,j=1}^{n,N}$ such that

$$\max_{1 \leq k \leq \kappa} P \left[\max_{1 \leq t \leq n} \left| \sum_{i=1}^t \xi_{ik} - \sum_{i=1}^t Z_{ik,\xi} \right| > C_1 n^\beta \right] < C_2 n^{-\gamma_1}, \quad (2.3)$$

$$P \left\{ \max_{1 \leq j \leq N} \max_{1 \leq t \leq n} \left| \sum_{i=1}^t \varepsilon_{ij} - \sum_{i=1}^t Z_{ij,\varepsilon} \right| > C_1 n^\beta \right\} < C_2 n^{-\gamma_2}. \quad (2.4)$$

Assumptions (A1)-(A2) are typical for spline smoothing, see Huang and Yang (2004), Xue and Yang (2006), Wang and Yang (2009a), Liu and Yang (2010) and Ma and Yang (2011). Assumption (A3) concerns the number of observations for each subject, and the number of knots of B-splines. Assumption (A4) ensures that the principal components have collectively bounded smoothness. Assumption (A5) provides Gaussian approximation of estimation error process, and is ensured by the following elementary assumption:

(A5') There exist $\tau_1 > 4$, $\tau_2 > 4+2\theta$ such that $E|\xi_{ik}|^{\tau_1} + E|\varepsilon_{ij}|^{\tau_2} < +\infty$, for $1 \leq i < \infty$,

$1 \leq k \leq \kappa$, $1 \leq j < \infty$. The number κ of nonzero eigenvalues is finite or κ is infinite

while the variables $\{\xi_{ik}\}_{1 \leq i < \infty, 1 \leq k < \infty}$ are i.i.d..

Degras (2011) makes a restrictive assumption (A.2) on the Hölder continuity of the stochastic process $\eta(x) = m(x) + \sum_{k=1}^{\infty} \xi_k \phi_k(x)$. It is elementary to construct examples where our Assumptions (A4) and (A5) are satisfied while assumption (A.2) of Degras (2011) is not.

The part of Assumption (A4) on ϕ_k 's holds trivially if κ is finite and all $\phi_k(x) \in C^{0,\mu}[0,1]$. Note also that by definition, $\phi_k = \sqrt{\lambda_k} \psi_k$, $\|\phi_k\|_{\infty} = \sqrt{\lambda_k} \|\psi_k\|_{\infty}$, $\|\phi_k\|_{0,\mu} = \sqrt{\lambda_k} \|\psi_k\|_{0,\mu}$, in which $\{\psi_k\}_{k=1}^{\infty}$ form an orthonormal basis of $L^2([0,1])$, hence, Assumption (A4) is fulfilled for $\kappa = \infty$ as long as λ_k decreases to zero sufficiently fast. Following one Referee's suggestion, we provide the following example. One takes $\lambda_k = \rho^{2[k/2]}$, $k = 1, 2, \dots$ for any $\rho \in (0, 1)$, with $\{\psi_k\}_{k=1}^{\infty}$ the canonical orthonormal Fourier basis of $L^2([0,1])$

$$\begin{aligned} \psi_1(x) &\equiv 1, \psi_{2k+1}(x) \equiv \sqrt{2} \cos(k\pi x) \\ \psi_{2k}(x) &\equiv \sqrt{2} \sin(k\pi x), k = 1, 2, \dots, x \in [0, 1]. \end{aligned}$$

In this case, $\sum_{k=1}^{\infty} \|\phi_k\|_{\infty} = 1 + \sum_{k=1}^{\infty} \rho^k (\sqrt{2} + \sqrt{2}) = 1 + 2\sqrt{2}\rho(1 - \rho)^{-1} < \infty$, while for any $\{\kappa_n\}_{n=1}^{\infty}$ with κ_n increasing, odd and $\kappa_n \rightarrow \infty$, and Lipschitz order $\mu = 1$

$$\begin{aligned} h_m \sum_{k=1}^{\kappa_n} \|\phi_k\|_{0,1} &= h_m \sum_{k=1}^{(\kappa_n-1)/2} \rho^k (\sqrt{2}k\pi + \sqrt{2}k\pi) \\ &\leq 2\sqrt{2}\pi h_m \rho \sum_{k=1}^{\infty} \rho^{k-1} k = 2\sqrt{2}\pi h_m (1 - \rho)^{-2} \\ &= O(h_m) = o(1). \end{aligned}$$

Denote by $\zeta(x)$, $x \in [0,1]$ a standardized Gaussian process such that $E\zeta(x) \equiv 0$,

$E\zeta^2(x) \equiv 1, x \in [0, 1]$ with covariance function

$$E\zeta(x)\zeta(x') = G(x, x') \left\{ G(x, x) G(x', x') \right\}^{-1/2}, x, x' \in [0, 1]$$

and define the $100 \times (1 - \alpha)$ -th percentile of the absolute maxima distribution of $\zeta(x)$, $\forall x \in [0, 1]$, i.e., $P \left\{ \sup_{x \in [0, 1]} |\zeta(x)| \leq Q_{1-\alpha} \right\} = 1 - \alpha, \forall \alpha \in (0, 1)$. Denote by $z_{1-\alpha/2}$ the $100(1 - \alpha/2)$ -th percentile of the standard normal distribution. Define also the following “infeasible estimator” of function m

$$\bar{m}(x) = \bar{\eta}(x) = n^{-1} \sum_{i=1}^n \eta_i(x), x \in [0, 1]. \quad (2.5)$$

The term “infeasible” refers to the fact that $\bar{m}(x)$ is computed from unknown quantity $\eta_i(x)$, $x \in [0, 1]$, and it would be the natural estimator of $m(x)$ if all the i.i.d. random curves $\eta_i(x)$, $x \in [0, 1]$ were observed, a view taken in Ferraty and Vieu (2006).

We now state our main results in the following theorem.

Theorem 2.2.1. *Under Assumptions (A1)-(A5), for $\forall \alpha \in (0, 1)$, as $n \rightarrow \infty$, the “infeasible estimator” $\bar{m}(x)$ converges at the $\sqrt{n}$ rate*

$$\begin{aligned} P \left\{ \sup_{x \in [0, 1]} n^{1/2} |\bar{m}(x) - m(x)| G(x, x)^{-1/2} \leq Q_{1-\alpha} \right\} &\rightarrow 1 - \alpha, \\ P \left\{ n^{1/2} |\bar{m}(x) - m(x)| G(x, x)^{-1/2} \leq z_{1-\alpha/2} \right\} &\rightarrow 1 - \alpha, \forall x \in [0, 1], \end{aligned}$$

while the spline estimator $\hat{m}_p$ is asymptotically equivalent to $\bar{m}$ up to order $n^{1/2}$, i.e.

$$\sup_{x \in [0, 1]} n^{1/2} |\bar{m}(x) - \hat{m}_p(x)| = o_P(1).$$

Remark 2.2.1. The significance of Theorem 2.2.1 lies in the fact that one does not need to distinguish between the spline estimator $\hat{m}_p$ and the “infeasible estimator” $\bar{m}$ in (2.5), which converges with $\sqrt{n}$ rate like a parametric estimator. We therefore have established oracle efficiency of the nonparametric estimator $\hat{m}_p$.

Corollary 2.2.2. *Under Assumptions (A1)-(A5), as $n \rightarrow \infty$, an asymptotic $100(1 - \alpha)\%$ correct confidence band for $m(x), x \in [0, 1]$ is*

$$\hat{m}_p(x) \pm G(x, x)^{1/2} Q_{1-\alpha} n^{-1/2}, \forall \alpha \in (0, 1)$$

while an asymptotic $100(1 - \alpha)\%$ pointwise confidence interval for $m(x), x \in [0, 1]$, is $\hat{m}_p(x) \pm G(x, x)^{1/2} z_{1-\alpha/2} n^{-1/2}$.

We next describe a two-sample extension of Theorem 2.2.1. Denote two samples indicated by $d = 1, 2$, which satisfy

$$Y_{dij} = m_d(j/N) + \sum_{k=1}^{\kappa_d} \xi_{dik} \phi_{dk}(j/N) + \sigma_d(j/N) \varepsilon_{dij}, \quad 1 \leq i \leq n_d, \quad 1 \leq j \leq N$$

with covariance functions $G_d(x, x') = \sum_{k=1}^{\kappa_d} \phi_{dk}(x) \phi_{dk}(x')$ respectively. We denote the ratio of two sample sizes as $\hat{r} = n_1/n_2$ and assume that $\lim_{n_1 \rightarrow \infty} \hat{r} = r > 0$.

For both groups, let $\hat{m}_{1p}(x)$ and $\hat{m}_{2p}(x)$ be the order p spline estimates of mean functions $m_1(x)$ and $m_2(x)$ by (2.2). Also denote by $\zeta_{12}(x), x \in [0, 1]$ a standardized Gaussian process such that $E\zeta_{12}(x) \equiv 0, E\zeta_{12}^2(x) \equiv 1, x \in [0, 1]$ with covariance function

$$E\zeta_{12}(x) \zeta_{12}(x') = \frac{G_1(x, x') + rG_2(x, x')}{\{G_1(x, x) + rG_2(x, x)\}^{1/2} \{G_1(x, x') + rG_2(x, x')\}^{1/2}}, \quad x, x' \in [0, 1].$$

Denote by $Q_{12,1-\alpha}$ the $(1-\alpha)$ -th quantile of the absolute maxima deviation of $\zeta_{12}(x)$, $x \in [0, 1]$ as above. We mimic the two sample t-test and state the following theorem whose proof is analogous to that of Theorem 2.2.1.

Theorem 2.2.3. *If Assumptions (A1)-(A5) are modified for each group accordingly, then for any $\alpha \in (0, 1)$, as $n_1 \rightarrow \infty$, $\hat{r} \rightarrow r > 0$,*

$$P \left\{ \sup_{x \in [0,1]} \frac{n_1^{1/2} \left| \left(\hat{m}_{1p} - \hat{m}_{2p} - m_1 + m_2 \right) (x) \right|}{\{(G_1 + rG_2)(x, x)\}^{1/2}} \leq Q_{12,1-\alpha} \right\} \rightarrow 1 - \alpha.$$

Theorems 2.2.3 yields uniform asymptotic confidence band for $m_1(x) - m_2(x)$, $x \in [0, 1]$.

Corollary 2.2.4. *If Assumptions (A1)-(A5) are modified for each group accordingly, as $n_1 \rightarrow \infty$, $\hat{r} \rightarrow r > 0$, a $100 \times (1 - \alpha)\%$ asymptotically correct confidence band for $m_1(x) - m_2(x)$, $x \in [0, 1]$ is $\left(\hat{m}_{1p} - \hat{m}_{2p} \right) (x) \pm n_1^{-1/2} Q_{12,1-\alpha} \{(G_1 + rG_2)(x, x)\}^{1/2}$, $\forall \alpha \in (0, 1)$.*

If the confidence band in Corollary 2.2.2 is used to test hypothesis

$$H_0 : m(x) = m_0(x), \forall x \in [0, 1] \longleftrightarrow H_a : m(x) \neq m_0(x), \text{ for some } x \in [0, 1],$$

for some given function $m_0(x)$, as one referee pointed out, the asymptotic power of the test is α under H_0 , 1 under H_1 due to Theorem 2.2.1. The same can be said for testing hypothesis about $m_1(x) - m_2(x)$ using the confidence band in Corollary 2.2.4.

2.3 Error decomposition

In this section, we break the estimation error $\hat{m}_p(x) - m(x)$ into three terms. We begin by discussing the representation of the spline estimator $\hat{m}_p(x)$ in (2.2).

The definition of $\hat{m}_p(x)$ in (2.2) means that

$$\hat{m}_p(x) \equiv \sum_{J=1-p}^{N_m} \hat{\beta}_{J,p} B_{J,p}(x),$$

with coefficients $\{\hat{\beta}_{1-p,p}, \dots, \hat{\beta}_{N_m,p}\}^T$ solving the following least squares problem

$$\begin{aligned} & \left\{ \hat{\beta}_{1-p,p}, \dots, \hat{\beta}_{N_m,p} \right\}^T \\ = & \underset{\left\{ \beta_{1-p,p}, \dots, \beta_{N_m,p} \right\} \in R^{N_m+p}}{\operatorname{argmin}} \sum_{i=1}^n \sum_{j=1}^N \left\{ Y_{ij} - \sum_{J=1-p}^{N_m} \beta_{J,p} B_{J,p}(j/N) \right\}^2. \end{aligned} \quad (2.6)$$

Applying elementary algebra, one obtains

$$\hat{m}_p(x) = \left\{ B_{1-p,p}(x), \dots, B_{N_m,p}(x) \right\} \left(\mathbf{X}^T \mathbf{X} \right)^{-1} \mathbf{X}^T \mathbf{Y}, \quad (2.7)$$

where $\mathbf{Y} = (\bar{Y}_{.1}, \dots, \bar{Y}_{.N})^T$, $\bar{Y}_{.j} = n^{-1} \sum_{i=1}^n Y_{ij}$, $1 \leq j \leq N$, and the design matrix $\mathbf{X}$ is

$$\mathbf{X} = \begin{pmatrix} B_{1-p,p}(1/N) & \cdots & B_{N_m,p}(1/N) \\ \vdots & \ddots & \vdots \\ B_{1-p,p}(N/N) & \cdots & B_{N_m,p}(N/N) \end{pmatrix}_{N \times (N_m+p)}.$$

Projecting via (2.7) the relationship in model (2.1) onto the linear subspace of R^{Nm+p} spanned by $\left\{B_{J,p}(j/N)\right\}_{1 \leq j \leq N, 1-p \leq J \leq Nm}$, we obtain the following crucial decomposition in the space $\mathcal{H}^{(p-2)}$ of spline functions:

$$\hat{m}_p(x) = \tilde{m}_p(x) + \tilde{e}_p(x) + \tilde{\xi}_p(x), \quad (2.8)$$

where

$$\begin{aligned} \tilde{m}_p(x) &= \sum_{J=1-p}^{Nm} \tilde{\beta}_{J,p} B_{J,p}(x), \quad \tilde{e}_p(x) = \sum_{J=1-p}^{Nm} \tilde{a}_{J,p} B_{J,p}(x), \\ \tilde{\xi}_p(x) &= \sum_{k=1}^{\kappa} \tilde{\xi}_{k,p}(x), \quad \tilde{\xi}_{k,p}(x) = \sum_{J=1-p}^{Nm} \tilde{\tau}_{k,J,p} B_{J,p}(x). \end{aligned} \quad (2.9)$$

The vectors $\{\tilde{\beta}_{1-p}, \dots, \tilde{\beta}_{Nm}\}^T$, $\{\tilde{a}_{1-p}, \dots, \tilde{a}_{Nm}\}^T$ and $\{\tilde{\tau}_{k,1-p}, \dots, \tilde{\tau}_{k,Nm}\}^T$ in (2.9) are solutions to (2.6) with Y_{ij} replaced by $m(j/N)$, $\sigma(j/N)\varepsilon_{ij}$ and $\xi_{ik}\phi_k(j/N)$ respectively. Alternatively,

$$\begin{aligned} \tilde{m}_p(x) &= \left\{B_{1-p,p}(x), \dots, B_{Nm,p}(x)\right\} \left(\mathbf{X}^T \mathbf{X}\right)^{-1} \mathbf{X}^T \mathbf{m} \\ \tilde{e}_p(x) &= \left\{B_{1-p,p}(x), \dots, B_{Nm,p}(x)\right\} \left(\mathbf{X}^T \mathbf{X}\right)^{-1} \mathbf{X}^T \mathbf{e} \\ \tilde{\xi}_{k,p}(x) &= \bar{\xi}_{.k} \left\{B_{1-p,p}(x), \dots, B_{Nm,p}(x)\right\} \left(\mathbf{X}^T \mathbf{X}\right)^{-1} \mathbf{X}^T \boldsymbol{\phi}_k, \quad 1 \leq k \leq \kappa \end{aligned}$$

in which $\mathbf{m} = (m(1/N), \dots, m(N/N))^T$ is the signal vector, $\mathbf{e} = (\sigma(1/N)\bar{\varepsilon}_{.1}, \dots, \sigma(N/N)\bar{\varepsilon}_{.N})^T$, $\bar{\varepsilon}_{.j} = n^{-1} \sum_{i=1}^n \varepsilon_{ij}$, $1 \leq j \leq N$ is the noise vector and $\boldsymbol{\phi}_k = (\phi_k(1/N), \dots, \phi_k(N/N))^T$ are the eigenfunction vectors, and $\bar{\xi}_{.k} = n^{-1} \sum_{i=1}^n \xi_{ik}$, $1 \leq k \leq \kappa$.

We cite next an important result from de Boor (2001), p. 149.

Theorem 2.3.1. *There is an absolute constant $C_{p-1,\mu} > 0$ such that for every $\phi \in C^{p-1,\mu}[0,1]$ for some $\mu \in (0,1]$, there exists a function $g \in \mathcal{H}^{(p-1)}[0,1]$ for which*

$$\|g - \phi\|_{\infty} \leq C_{p-1,\mu} \left\| \phi^{(p-1)} \right\|_{0,\mu} h_m^{\mu+p-1}.$$

The next three propositions concern $\tilde{m}_p(x)$, $\tilde{e}_p(x)$ and $\tilde{\xi}_p(x)$ given in (2.8).

Proposition 2.3.1. *Under Assumptions (A1) and (A3), as $n \rightarrow \infty$*

$$\sup_{x \in [0,1]} n^{1/2} |\tilde{m}_p(x) - m(x)| = o_P(1). \quad (2.10)$$

Proposition 2.3.2. *Under Assumptions (A2)-(A4), as $n \rightarrow \infty$*

$$\sup_{x \in [0,1]} n^{1/2} |\tilde{e}_p(x)| = o_P(1). \quad (2.11)$$

Proposition 2.3.3. *Under Assumptions (A2)-(A4), as $n \rightarrow \infty$*

$$\sup_{x \in [0,1]} n^{1/2} \left| \tilde{\xi}_p(x) - (\bar{m}(x) - m(x)) \right| = o_P(1) \quad (2.12)$$

also for any $\alpha \in (0,1)$

$$P \left\{ \sup_{x \in [0,1]} n^{1/2} \left| \tilde{\xi}_p(x) \right| G(x,x)^{-1/2} \leq Q_{1-\alpha} \right\} \rightarrow 1 - \alpha. \quad (2.13)$$

Equations (2.10), (2.11) and (2.12) yield the asymptotic efficiency of the spline estimator $\hat{m}_p$, i.e. $\sup_{x \in [0,1]} n^{1/2} |\bar{m}(x) - \hat{m}_p(x)| = o_P(1)$. The Appendix contains the proofs of the above three propositions, which together with (2.8), imply Theorem 2.2.1.

2.4 Implementation

This section describes procedures to implement the confidence band of Corollary 2.2.2.

Given any data set $\left(j/N, Y_{ij}\right)_{j=1, i=1}^{N, n}$ from model (2.1), the spline estimator $\hat{m}_p(x)$ is obtained from (2.7), the number of interior knots in estimating $m(x)$ is taken to be $N_m = \lceil cn^{1/(2p)\log(n)} \rceil$, in which $\lceil a \rceil$ denotes the integer part of a . Our experience shows that the choice of constant $c = 0.2, 0.3, 0.5, 1, 2$ seems quite adequate, and that is what we recommend. When constructing the confidence bands, one needs to estimate the unknown functions $G(\cdot, \cdot)$ and the quantile $Q_{1-\alpha}$ and then plug in these estimators: the same approach is taken in Ma, Yang and Carroll (2012) and Wang and Yang (2009a).

The pilot estimator $\hat{G}_p(x, x')$ of covariance function $G(x, x')$ is

$$\hat{G}_p = \underset{g(\cdot, \cdot) \in \mathcal{H}^{(p-2), 2}}{\operatorname{argmin}} \sum_{j \neq j'}^N \left\{ C_{.jj'} - g\left(j/N, j'/N\right) \right\}^2,$$

with $C_{.jj'} = n^{-1} \sum_{i=1}^n \left\{ Y_{ij} - \hat{m}_p(j/N) \right\} \left\{ Y_{ij'} - \hat{m}_p(j'/N) \right\}$, $1 \leq j \neq j' \leq N$ and the tensor product spline space $\mathcal{H}^{(p-2), 2} = \left\{ \sum_{J, J'=1-p}^{N_G} b_{JJ'} B_{J,p}(t) B_{J',p}(s), b_{JJ'} \in \mathcal{R}, t, s \in [0, 1] \right\}$ in which $N_G = \lceil n^{1/(2p)\log(\log(n))} \rceil$. A detailed discussion of the consistent property of this plug-in estimator can be found in Chapter 3.

In order to estimate $Q_{1-\alpha}$, one first does the eigenfunction decomposition of $\hat{G}_p(x, x')$, i.e. $N^{-1} \sum_{j=1}^N \hat{G}_p(j/N, j'/N) \hat{\psi}_k(j/N) = \hat{\lambda}_k \hat{\psi}_k(j'/N)$, to obtain the estimated eigenvalues $\hat{\lambda}_k$ and eigenfunctions $\hat{\psi}_k$. Next, one chooses the number κ of eigenfunctions by using the following standard and efficient criterion, i.e.

$\kappa = \operatorname{argmin}_{1 \leq l \leq T} \left\{ \sum_{k=1}^l \hat{\lambda}_k / \sum_{k=1}^T \hat{\lambda}_k > 0.95 \right\}$, where $\{\lambda_k\}_{k=1}^T$ are the first T estimated positive eigenvalues. Finally, one simulates $\hat{\zeta}_b(x) = \hat{G}_p(x, x)^{-1/2} \sum_{k=1}^{\kappa} Z_{k,b} \hat{\phi}_k(x)$, where

$\hat{\phi}_k = \sqrt{\hat{\lambda}_k} \hat{\psi}_k$, $Z_{k,b}$ are i.i.d standard normal variables with $1 \leq k \leq \kappa$ and $b = 1, \dots, b_M$, and b_M is a preset large integer, the default of which is 1000. We take the maximal absolute value for each copy of $\hat{\zeta}_b(x)$ and estimates $Q_{1-\alpha}$ by the empirical quantile $\hat{Q}_{1-\alpha}$ of these maximum values. We then use the following confidence band

$$\hat{m}_p(x) \pm n^{-1/2} \hat{G}_p(x, x)^{1/2} \hat{Q}_{1-\alpha}, \quad x \in [0, 1], \quad (2.14)$$

for the mean function. We estimate $Q_{12,1-\alpha}$ in a similar way as $\hat{Q}_{1-\alpha}$ and compute

$$\left(\hat{m}_{1p} - \hat{m}_{2p} \right)(x) \pm n_1^{-1/2} \hat{Q}_{12,1-\alpha} \left\{ \left(\hat{G}_{1p} + \hat{r} \hat{G}_{2p} \right)(x, x) \right\}^{1/2}, \quad (2.15)$$

as confidence band for $m_1(x) - m_2(x)$. Although beyond the scope of this work, as one referee pointed out, the confidence band in (2.14) is expected to enjoy the same asymptotic coverage as if true values of $Q_{1-\alpha}$ and $G(x, x)$ were used instead, due to the consistency of $\hat{G}_p(x, x)$ estimating $G(x, x)$. The same holds for the confidence band in (2.15).

2.5 Simulation

To demonstrate the practical performance of our theoretical results, we perform a set of simulation studies. Data are generated from model

$$Y_{ij} = m(j/N) + \sum_{k=1}^2 \xi_{ik} \phi_k(j/N) + \sigma \varepsilon_{ij}, \quad 1 \leq j \leq N, 1 \leq i \leq n, \quad (2.16)$$

where $\xi_{ik} \sim N(0, 1)$, $k = 1, 2$, $\varepsilon_{ij} \sim N(0, 1)$, for $1 \leq i \leq n$, $1 \leq j \leq N$, $m(x) = 10 + \sin \{2\pi(x - 1/2)\}$, $\phi_1(x) = -2 \cos \{\pi(x - 1/2)\}$ and $\phi_2(x) = \sin \{\pi(x - 1/2)\}$. This setting

implies $\lambda_1 = 2$ and $\lambda_2 = 0.5$. The noise levels are set to be $\sigma = 0.5$ and 0.3 . The number of subjects n is taken to be $60, 100, 200, 300$ and 500 , and under each sample size the number of observations per curve is assumed to be $N = \lceil n^{0.25} \log^2(n) \rceil$. This simulated process has a similar design as one of the simulation models in Yao, Müller and Wang (2005), except that each subject is densely observed. We consider both linear and cubic spline estimators, and use confidence levels $1 - \alpha = 0.95$ and 0.99 for our simultaneous confidence bands. The constant c in the definition of N_m in Section 2.4 is taken to be $0.2, 0.3, 0.5, 1$ and 2 . Each simulation is repeated 500 times.

Figures 2.1 and 2.2 show the estimated mean functions and their 95% confidence bands for the true curve $m(\cdot)$ in model (2.16) with $\sigma = 0.3$ and $n = 100, 200, 300, 500$, respectively. As expected when n increases, the confidence band becomes narrower and the linear and cubic spline estimators are closer to the true curve.

Tables 2.2 to 2.4 show the empirical frequency that the true curve $m(\cdot)$ is covered by the linear and cubic spline confidence bands (2.14) at 100 points $\{1/100, \dots, 99/100, 1\}$ respectively. At all noise levels, the coverage percentages for the confidence bands are close to the nominal confidence levels 0.95 and 0.99 for linear splines with $c = 0.5, 1$ (Tables 2.1 and 2.2), and cubic splines with $c = 0.3, 0.5$ (Tables 2.3 and 2.4) but decline slightly for $c = 2$ and markedly for $c = 0.2$. The coverage percentages thus depend on the choice of N_m , and the dependency becomes stronger when sample sizes decrease. For large sample sizes $n = 300, 500$, the effect of the choice of N_m on the coverage percentages is negligible. Although our theory indicates no optimal choice of c , we recommend using $c = 0.5$ for data analysis as its performance in simulation for both linear and cubic splines is either optimal or near optimal.

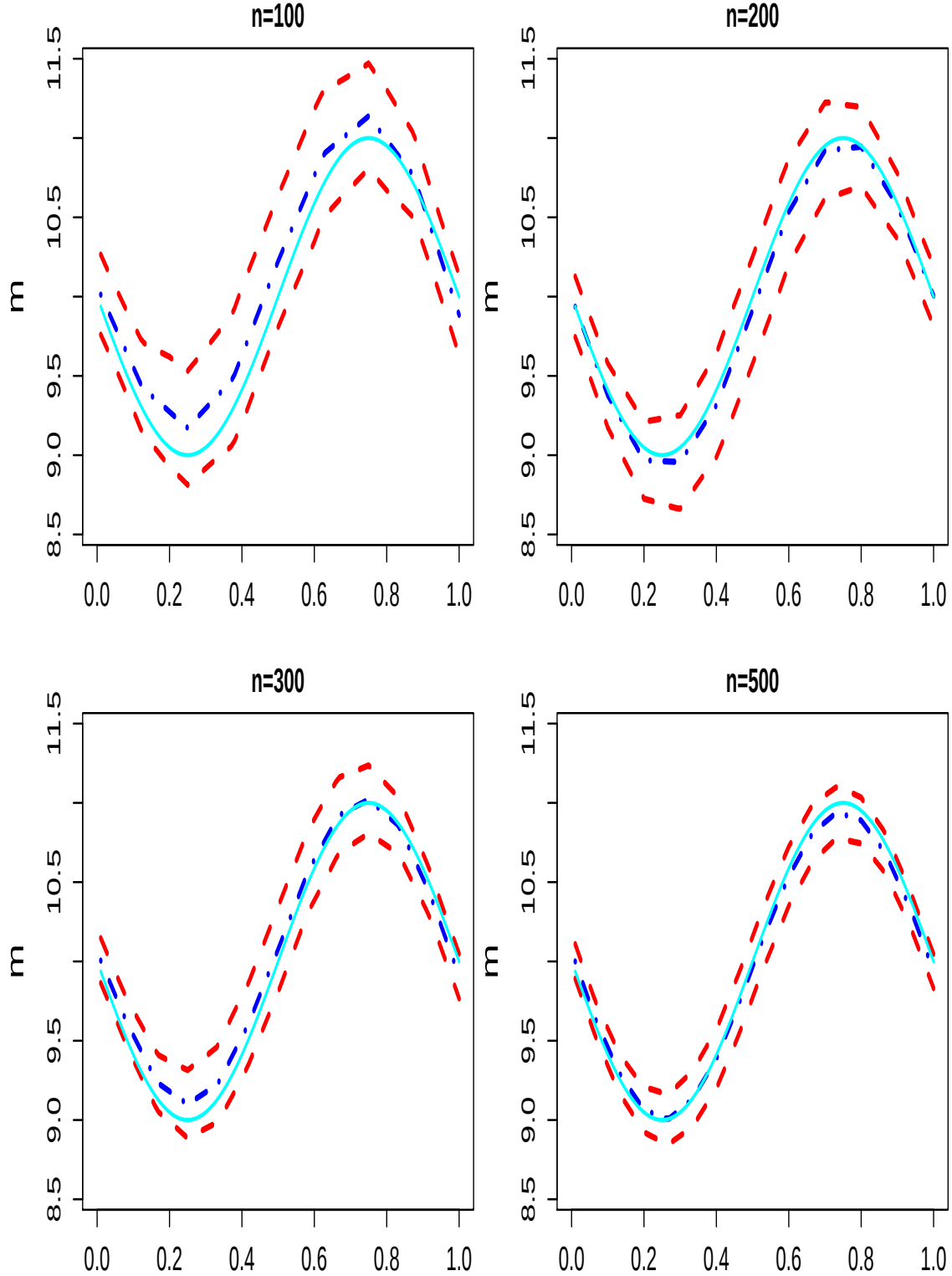


Figure 2.1: For interpretation of the references to color in this and all other figures, the reader is referred to the electronic version of this dissertation. Plots of the linear spline estimator (2.2) for simulated data (dashed-dotted line) and 95% confidence bands (2.14) (upper and lower dashed lines) (2.14) for $m(x)$ (solid lines). In all panels, $\sigma = 0.3$.

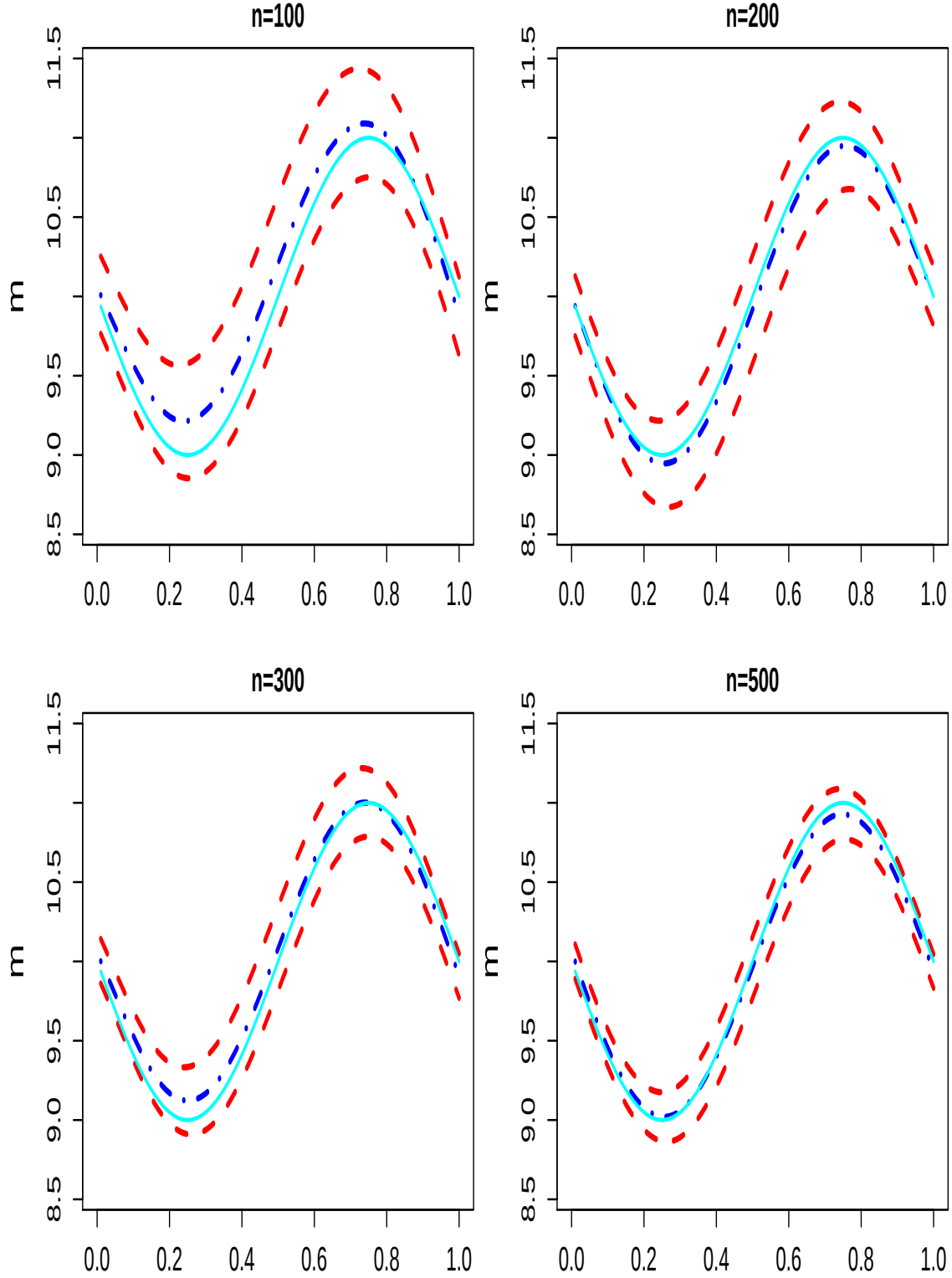


Figure 2.2: Plots of the cubic spline estimator (2.2) for simulated data (dashed-dotted line) and 95% confidence bands (2.14) (upper and lower dashed lines) (2.14) for $m(x)$ (solid lines). In all panels, $\sigma = 0.3$.

Table 2.1: Coverage frequencies from 500 replications using linear spline (2.14) with $p = 2$, $N_m = \lceil cn^{1/(2p)\log(n)} \rceil$ and $\sigma = 0.5$.

n	$1 - \alpha$	$c = 0.2$	$c = 0.3$	$c = 0.5$	$c = 1$	$c = 2$
60	0.950	0.384	0.790	0.876	0.894	0.852
	0.990	0.692	0.938	0.970	0.976	0.942
100	0.950	0.184	0.826	0.886	0.884	0.838
	0.990	0.476	0.936	0.964	0.966	0.944
200	0.950	0.418	0.856	0.914	0.922	0.862
	0.990	0.712	0.966	0.976	0.990	0.972
300	0.950	0.600	0.888	0.920	0.932	0.874
	0.990	0.834	0.978	0.976	0.980	0.972
500	0.950	0.772	0.880	0.922	0.886	0.894
	0.990	0.902	0.964	0.984	0.976	0.976

Table 2.2: Coverage frequencies from 500 replications using linear spline (2.14) with $p = 2$, $N_m = \lceil cn^{1/(2p)\log(n)} \rceil$ and $\sigma = 0.3$.

n	$1 - \alpha$	$c = 0.2$	$c = 0.3$	$c = 0.5$	$c = 1$	$c = 2$
60	0.950	0.410	0.786	0.930	0.914	0.884
	0.990	0.702	0.950	0.972	0.966	0.954
100	0.950	0.198	0.822	0.916	0.916	0.896
	0.990	0.496	0.940	0.974	0.974	0.968
200	0.950	0.414	0.862	0.946	0.942	0.926
	0.990	0.720	0.966	0.984	0.984	0.980
300	0.950	0.602	0.896	0.940	0.934	0.926
	0.990	0.840	0.982	0.984	0.986	0.980
500	0.950	0.768	0.888	0.954	0.950	0.942
	0.990	0.906	0.968	0.992	0.994	0.988

Table 2.3: Coverage frequencies from 500 replications using cubic spline (2.14) with $p = 4$, $N_m = \lceil cn^{1/(2p)} \log(n) \rceil$ and $\sigma = 0.5$.

n	$1 - \alpha$	$c = 0.2$	$c = 0.3$	$c = 0.5$	$c = 1$	$c = 2$
60	0.950	0.644	0.916	0.902	0.890	0.738
	0.990	0.866	0.980	0.958	0.964	0.888
100	0.950	0.596	0.902	0.904	0.876	0.846
	0.990	0.786	0.970	0.968	0.956	0.952
200	0.950	0.928	0.942	0.932	0.936	0.904
	0.990	0.978	0.992	0.982	0.992	0.978
300	0.950	0.920	0.948	0.926	0.948	0.898
	0.990	0.976	0.986	0.986	0.988	0.980
500	0.950	0.928	0.922	0.954	0.902	0.898
	0.990	0.980	0.982	0.990	0.976	0.978

Table 2.4: Coverage frequencies from 500 replications using cubic spline (2.14) with $p = 4$, $N_m = \lceil cn^{1/(2p)} \log(n) \rceil$ and $\sigma = 0.3$.

n	$1 - \alpha$	$c = 0.2$	$c = 0.3$	$c = 0.5$	$c = 1$	$c = 2$
60	0.950	0.672	0.922	0.940	0.940	0.916
	0.990	0.884	0.986	0.986	0.984	0.982
100	0.950	0.610	0.916	0.914	0.914	0.896
	0.990	0.798	0.980	0.974	0.970	0.964
200	0.950	0.938	0.952	0.950	0.948	0.934
	0.990	0.982	0.984	0.992	0.982	0.984
300	0.950	0.922	0.956	0.948	0.942	0.938
	0.990	0.982	0.984	0.988	0.984	0.982
500	0.950	0.928	0.928	0.936	0.932	0.916
	0.990	0.980	0.982	0.990	0.990	0.992

We compare by simulation the proposed spline confidence band to the least squares Bonferroni and least squares bootstrap bands in Bunea, Ivanescu and Wegkamp (2011) (BIW). Table 2.5 presents the empirical frequency that the true curve $m(\cdot)$ for model (2.16) is covered by these bands at $\{1/100, \dots, 99/100, 1\}$ respectively as Tables 2.1 and 2.2. The coverage frequency of the BIW Bonferroni band is much higher than the nominal level making it too conservative. The coverage frequency of the BIW bootstrap band is consistently lower than the nominal level by at least 10%, thus not recommended for practical use.

We also compare the widths of the three bands. For each replication, we calculate the ratios of widths of the two BIW bands against the spline band at $\{1/100, \dots, 99/100, 1\}$ and then average these 100 ratios. Table 2.6 shows the five number summary of these 500 averaged ratios for $\sigma = 0.3$ and $p = 4$. The BIW Bonferroni band is much wider than

Table 2.5: Coverage frequencies from 500 replications using least squares Bonferroni band and least squares Bootstrap band.

n	$1 - \alpha$	Coverage frequency		Coverage frequency	
		least squares Bonferroni		least squares bootstrap	
		$\sigma = 0.5$	$\sigma = 0.3$	$\sigma = 0.5$	$\sigma = 0.3$
60	0.950	0.990	0.988	0.742	0.744
	0.990	0.994	0.994	0.856	0.864
100	0.950	0.996	0.998	0.678	0.712
	0.990	0.998	1.000	0.860	0.870
200	0.950	0.988	0.992	0.710	0.734
	0.990	1.000	1.000	0.856	0.888
300	0.950	0.988	0.998	0.704	0.720
	0.990	1.000	1.000	0.868	0.870
500	0.950	0.996	0.998	0.718	0.732
	0.990	1.000	1.000	0.856	0.860

Table 2.6: Five number summary of ratios of confidence band widths.

n	$1 - \alpha$	least squares Bonferroni/cubic spline					least squares bootstrap/cubic spline				
		Min.	Q1	Med.	Q3	Max.	Min.	Q1	Med.	Q3	Max.
60	0.950	0.964	1.219	1.299	1.397	1.845	0.522	0.667	0.716	0.770	0.967
	0.990	0.907	1.114	1.188	1.285	1.730	0.527	0.662	0.715	0.770	1.048
100	0.950	0.995	1.263	1.331	1.415	1.684	0.565	0.675	0.714	0.754	0.888
	0.990	0.910	1.148	1.219	1.295	1.603	0.536	0.665	0.708	0.752	0.925
200	0.950	1.169	1.326	1.383	1.433	1.653	0.600	0.683	0.715	0.743	0.855
	0.990	1.045	1.197	1.250	1.300	1.507	0.557	0.668	0.702	0.740	0.888
300	0.950	1.169	1.363	1.412	1.462	1.663	0.574	0.690	0.717	0.742	0.838
	0.990	1.067	1.228	1.277	1.322	1.509	0.587	0.676	0.707	0.739	0.850
500	0.950	1.273	1.395	1.432	1.476	1.601	0.620	0.691	0.714	0.737	0.818
	0.990	1.132	1.243	1.288	1.334	1.465	0.607	0.674	0.707	0.734	0.839

cubic spline band, making it undesirable. While the BIW bootstrap band is narrower, we have mentioned previously that its coverage frequency is too low to be useful in practice. Simulation for other cases (e.g. $p = 2$, $\sigma = 0.5$) leads to the same conclusion.

To examine the performance of the two-sample test based on spline confidence band, Table 2.7 reports the empirical power and type I error for the proposed two-sample test. The data were generated from (2.16) with $\sigma = 0.5$ and $m_1(x) = 10 + \sin\{2\pi(x - 1/2)\} + \delta(x)$, $n = n_1$ for the first group, and $m_2(x) = 10 + \sin\{2\pi(x - 1/2)\}$, $n = n_2$ for the another group. The remaining parameters, ξ_{ik} , ε_{ij} , $\phi_1(x)$ and $\phi_2(x)$ were set to the same values for each group as in (2.16). In order to mimic the real data in Section 2.6, we set $N = 50$, 100 and 200 when $n_1 = 160$, 80 and 40 and $n_2 = 320$, 160 and 80 accordingly. The studied

Table 2.7: Empirical power and type I error of two-sample test using cubic spline.

$\delta(x)$	$n_1 = 160, n_2 = 320$		$n_1 = 80, n_2 = 160$		$n_1 = 40, n_2 = 80$	
	Nominal test level		Nominal test level		Nominal test level	
	0.05	0.01	0.05	0.01	0.05	0.01
$0.6t$	1.000	1.000	0.980	0.918	0.794	0.574
$0.7\sin(x)$	1.000	1.000	0.978	0.910	0.788	0.566
0	0.058	0.010	0.068	0.010	0.096	0.028
Monte Carlo SE	0.001	0.004	0.001	0.004	0.001	0.004

hypotheses are:

$$H_0 : m_1(x) = m_2(x), \forall x \in [0, 1] \longleftrightarrow H_a : m_1(x) \neq m_2(x), \text{ for some } x \in [0, 1].$$

Table 2.7 shows the empirical frequencies of rejecting H_0 in this simulation study with nominal test level equal to 0.05 and 0.01. If $\delta(x) \neq 0$, these empirical powers should be close to 1, and for $\delta(x) \equiv 0$, the nominal levels. Each set of simulations consists of 500 Monte Carlo runs. Asymptotic standard errors (as the number of Monte Carlo iterations tends to infinity) are reported in the last row of the table. Results are listed only for cubic spline confidence bands, as those of the linear spline are similar. Overall, the two-sample test performs well, even with a rather small difference ($\delta(x) = 0.7\sin(x)$), providing a reasonable empirical power. Moreover, the differences between nominal levels and empirical type I error do diminish as the sample size increases.

2.6 Empirical example

In this section, we revisit the Tecator data mentioned in Section 1, which can be downloaded at <http://lib.stat.cmu.edu/datasets/tecator>. In this data set, there are measurements on $n = 240$ meat samples, where for each sample a $N = 100$ channel near-infrared spectrum of absorbance measurements was recorded, and contents of moisture (water), fat and protein were also obtained. The Feed Analyzer worked in the wavelength range from 850 nm to 1050 nm. Figure 2.3 shows the scatter plot of this data set. The spectral data can be naturally considered as functional data, and we will perform a two-sample test to see whether absorbance from the spectrum differs significantly due to difference in fat content.

This data set has been used for comparing four classification methods (Li and Yu, 2008), building a regression model to predict the fat content from the spectrum (Li and Hsing, 2010b). Following Li and Yu (2008), we separate samples according to their fat contents being less than 20% or not. The right panel of Figure 2.3 shows 10 samples from each group. Here, hypothesis of interest is:

$$H_0 : m_1(x) = m_2(x), \forall x \in [850, 1050] \longleftrightarrow H_a : m_1(x) \neq m_2(x), \text{ for some } x \in [850, 1050],$$

where $m_1(x)$ and $m_2(x)$ are the regression functions of absorbance on spectrum, for samples with fat content less than 20% and greater than or equal to 20%, respectively. Among 240 samples, there are $n_1 = 155$ with fat content less than 20%, the rest $n_2 = 85$ no less than 20%. The numbers of interior knots in (2.2) are computed as in Section 2.4 with $c = 0.5$ and are $N_{1m} = 4$ and $N_{2m} = 3$ for cubic spline fit and $N_{1m} = 8$ and $N_{2m} = 6$ for linear spline fit. Figure 2.4 depicts the linear and cubic spline confidence bands according

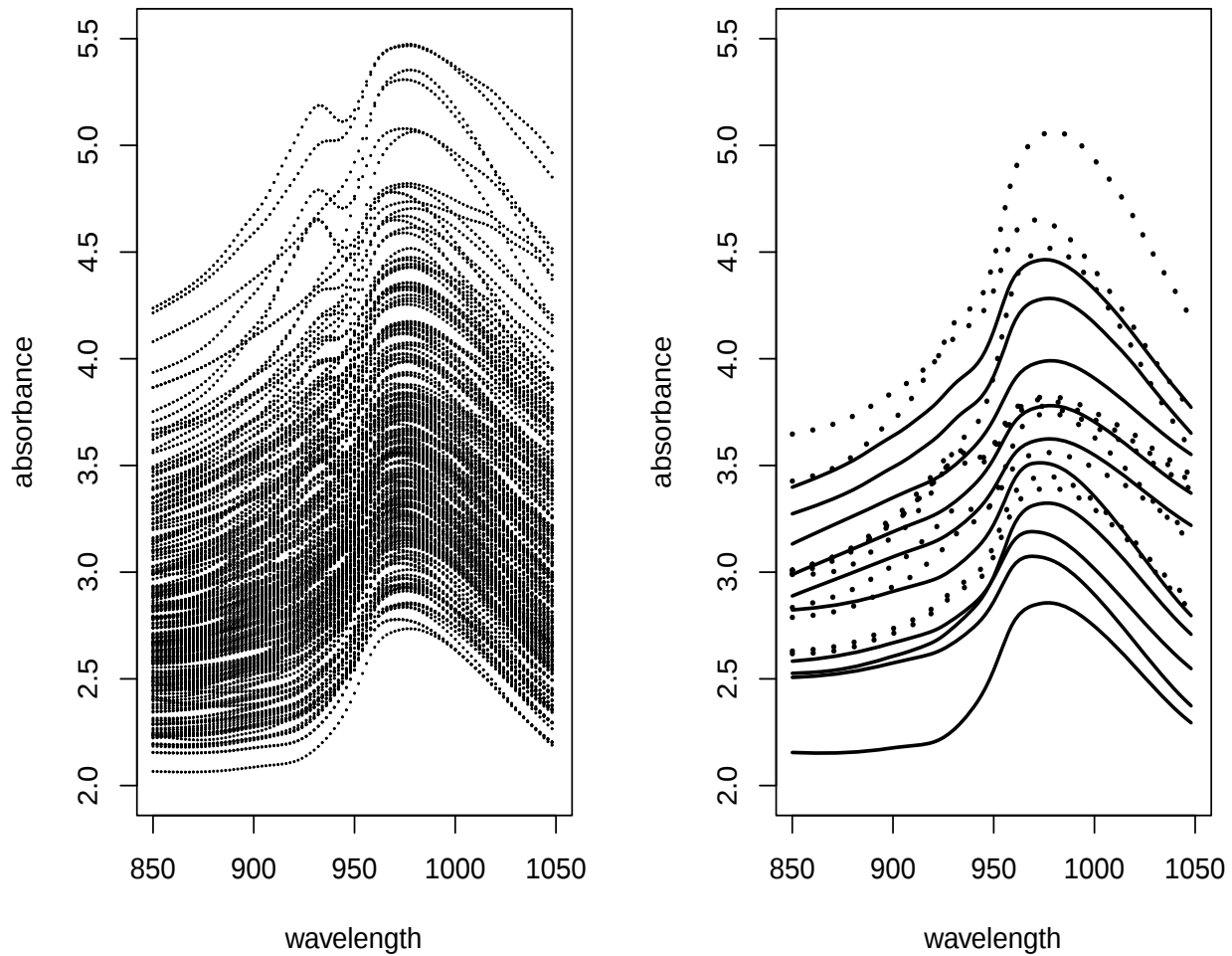


Figure 2.3: Left: Plot of Tecator data. Right: Sample curves for the Tecator data. Each class has 10 sample curves. Dashed lines represent spectra with fact $> 20\%$ and solid lines represent spectra with fact $< 20\%$.

to (2.15) at confidence levels 0.99 (upper and lower dashed lines) and 0.999995 (upper and lower dotted lines), with the center dashed-dotted line representing the spline estimator $\hat{m}_1(x) - \hat{m}_2(x)$ and a solid line representing zero. Since even the 99.9995% confidence band does not contain the zero line entirely, the difference of low fat and high fat populations' absorbance was extremely significant. In fact, Figure 2.4 clearly indicates that the less the fat contained, the higher the absorbance is.

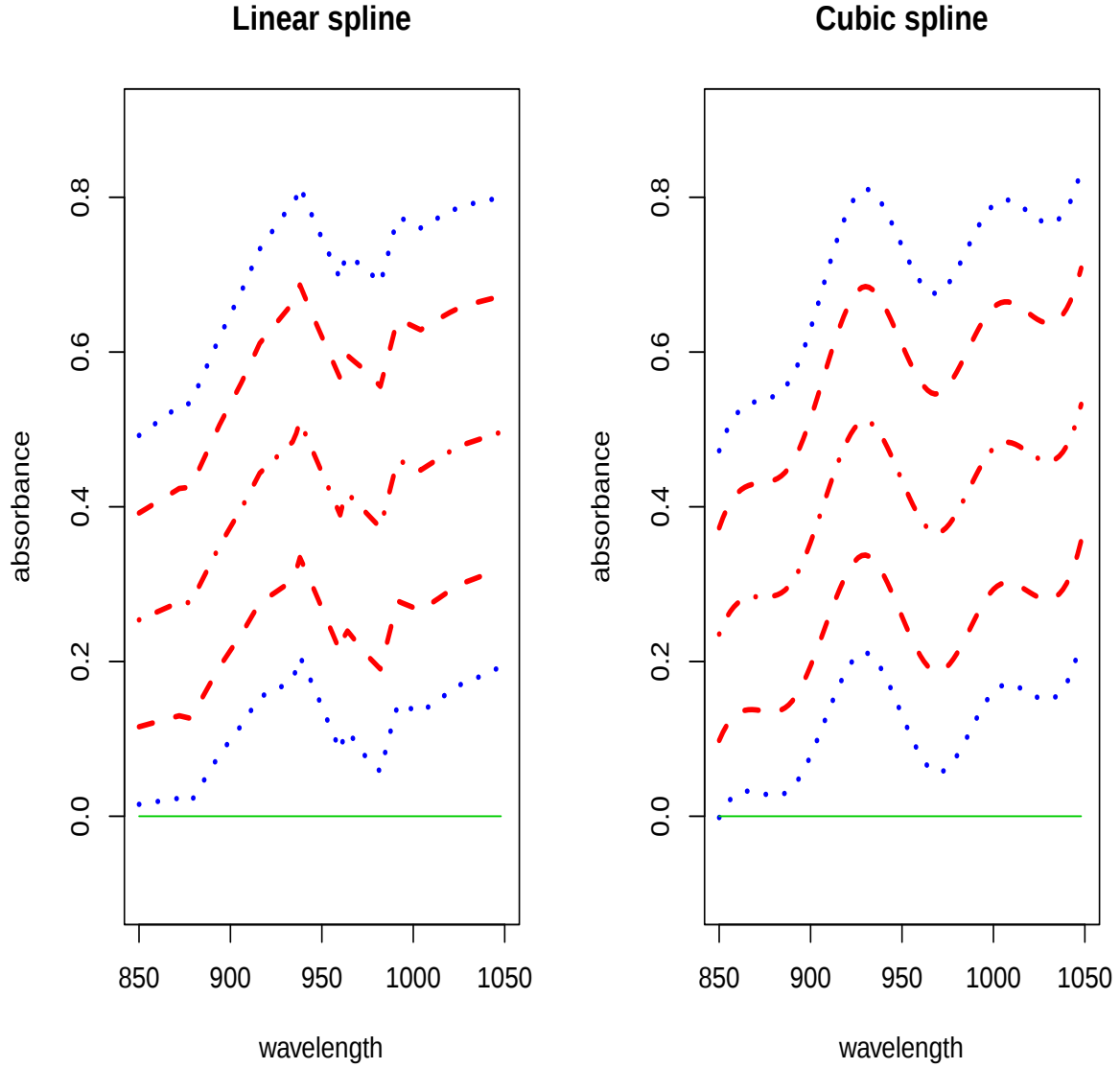


Figure 2.4: Plots of the fitted linear and cubic spline regressions of $m_1(x) - m_2(x)$ for the Tecator data (dashed-dotted line), 99% confidence bands (2.15) (upper and lower dashed lines), 99.9995% confidence bands (2.15) (upper and lower dotted lines) and the zero line (solid line).

APPENDIX

In this appendix, we use C to denote a generic positive constant unless stated otherwise.

Preliminaries

For any vector $\zeta = (\zeta_1, \dots, \zeta_s) \in R^s$, denote the norm $\|\zeta\|_r = (|\zeta_1|^r + \dots + |\zeta_s|^r)^{1/r}$, $1 \leq r < +\infty$, $\|\zeta\|_\infty = \max(|\zeta_1|, \dots, |\zeta_s|)$. For any $s \times s$ symmetric matrix $\mathbf{A}$, we define $\lambda_{\min}(\mathbf{A})$ and $\lambda_{\max}(\mathbf{A})$ as its smallest and largest eigenvalues, and its L_r norm as $\|\mathbf{A}\|_r = \max_{\zeta \in R^s, \zeta \neq \mathbf{0}} \|\zeta\|_r^{-1} \|\mathbf{A}\zeta\|_r$. In particular, $\|\mathbf{A}\|_2 = \lambda_{\max}(\mathbf{A})$, and if $\mathbf{A}$ is also nonsingular, $\|\mathbf{A}^{-1}\|_2 = \lambda_{\min}^{-1}(\mathbf{A})$.

For any functions $\phi, \varphi \in L_2[0, 1]$, define the theoretical and empirical inner products as $\langle \phi, \varphi \rangle = \int_0^1 \phi(x) \varphi(x) dx$ and $\langle \phi, \varphi \rangle_{2,N} = N^{-1} \sum_{j=1}^N \phi(j/N) \varphi(j/N)$. The corresponding norms are $\|\phi\|_2^2 = \langle \phi, \phi \rangle$, $\|\phi\|_{2,N}^2 = \langle \phi, \phi \rangle_{2,N}$.

We state a strong approximation result, which is used in the proof of Lemma 2.6.6.

Lemma 2.6.1. *[Theorem 2.6.7 of Csörgő and Révész (1981)] Suppose that ξ_i , $1 \leq i < \infty$ are i.i.d. with $E(\xi_1) = 0$, $E(\xi_1^2) = 1$ and $H(x) > 0$ ($x \geq 0$) is an increasing continuous function such that $x^{-2-\gamma}H(x)$ is increasing in x for some $\gamma > 0$ and $x^{-1}\log H(x)$ is decreasing in x with $EH(|\xi_1|) < \infty$. Then there exist a sequence of Brownian motions $\{W_n(l)\}_{n=1}^\infty$ and constants $C_1, C_2, a > 0$, depending only on the distribution of ξ_1 and such that for any $\{x_n\}_{n=1}^\infty$ satisfying $H^{-1}(n) < x_n < C_1(n \log n)^{1/2}$ and $S_l = \sum_{i=1}^l \xi_i$,*

$$P \left\{ \max_{1 \leq l \leq n} |S_l - W_n(l)| > x_n \right\} \leq C_2 n \{H(ax_n)\}^{-1}.$$

The next lemma is a special case of Theorem 13.4.3, Page 404 of DeVore and Lorentz (1993). Let p be a positive integer. A matrix $\mathbf{A} = (a_{ij})$ is said to have bandwidth p if $a_{ij} = 0$ when $|i - j| \geq p$, and p is the smallest integer with this property. Denote by

$d = \|\mathbf{A}\|_2 \|\mathbf{A}^{-1}\|_2$ and d is the condition number of $\mathbf{A}$.

Lemma 2.6.2. *If a matrix $\mathbf{A}$ with bandwidth p has an inverse $\mathbf{A}^{-1}$, then $\|\mathbf{A}^{-1}\|_\infty \leq 2c_0(1-\tau)^{-1}$, with $c_0 = \tau^{-2p} \|\mathbf{A}^{-1}\|_2$, $\tau = \left((d^2 - 1)/(d^2 + 1)\right)^{1/(4p)}$.*

One writes $\mathbf{X}^T \mathbf{X} = N \hat{\mathbf{V}}_p$, $\mathbf{X}^T \mathbf{Y} = \left\{ \sum_{j=1}^N B_{J,p}(j/N) \bar{Y}_{\cdot j} \right\}_{J=1-p}^{Nm}$, where the theoretical and empirical inner product matrices of $\left\{ B_{J,p}(x) \right\}_{J=1-p}^{Nm}$ are denoted as

$$\mathbf{V}_p = \left(\left\langle B_{J,p}, B_{J',p} \right\rangle \right)_{J,J'=1-p}^{Nm}, \hat{\mathbf{V}}_p = \left(\left\langle B_{J,p}, B_{J',p} \right\rangle_{2,N} \right)_{J,J'=1-p}^{Nm}. \quad (2.17)$$

We establish next that the theoretical inner product matrix $\mathbf{V}_p$ defined in (2.17) has an inverse with bounded L_∞ norm.

Lemma 2.6.3. *For any positive integer p , there exists a constant $M_p > 0$ depending only on p , such that $\|\mathbf{V}_p^{-1}\|_\infty \leq M_p h_m^{-1}$, where $h_m = (N_m + 1)^{-1}$.*

PROOF. According to Lemma A.1 in Wang and Yang (2009b), $\mathbf{V}_p$ is invertible since it is a symmetric matrix with all eigenvalues positive, i.e. $0 < c_p N_m^{-1} \leq \lambda_{\min}(\mathbf{V}_p) \leq \lambda_{\max}(\mathbf{V}_p) \leq C_p N_m^{-1} < \infty$, where c_p and C_p are positive real numbers. The compact support of B-spline basis makes $\mathbf{V}_p$ of bandwidth p , hence one can apply Lemma 2.6.2. Since $d_p = \lambda_{\max}(\mathbf{V}_p) / \lambda_{\min}(\mathbf{V}_p) \leq C_p / c_p$, hence

$$\tau_p = \left(d_p^2 - 1\right)^{1/4p} \left(d_p^2 + 1\right)^{-1/4p} \leq \left(C_p^2 c_p^{-2} - 1\right)^{1/4p} \left(C_p^2 c_p^{-2} + 1\right)^{-1/4p} < 1.$$

If $p = 1$, then $\mathbf{V}_p^{-1} = h_m^{-1} \mathbf{I}_{N_m+p}$, the lemma holds with $M_p = 1$. If $p > 1$, let $\mathbf{u}_{1-p} = \left(1, \mathbf{0}_{N_m+p-1}^T\right)^T$, $\mathbf{u}_0 = \left(\mathbf{0}_{p-1}^T, 1, \mathbf{0}_{N_m}^T\right)^T$, then $\|\mathbf{u}_{1-p}\|_2 = \|\mathbf{u}_0\|_2 = 1$. Also lemma A.1

in Wang and Yang (2009b) implies that

$$\begin{aligned}\lambda_{\min}(\mathbf{V}_p) &= \lambda_{\min}(\mathbf{V}_p) \|\mathbf{u}_{1-p}\|_2^2 \leq \mathbf{u}_{1-p}^T \mathbf{V}_p \mathbf{u}_{1-p} = \|B_{1-p,p}\|_2^2, \\ \mathbf{u}_0^T \mathbf{V}_p \mathbf{u}_0 &= \|B_{0,p}\|_2^2 \leq \lambda_{\max}(\mathbf{V}_p) \|\mathbf{u}_0\|_2^2 = \lambda_{\max}(\mathbf{V}_p),\end{aligned}$$

hence $d_p = \lambda_{\max}(\mathbf{V}_p) / \lambda_{\min}(\mathbf{V}_p) \geq \|B_{0,p}\|_2^2 \|B_{1-p,p}\|_2^{-2} = r_p > 1$, where r_p is an absolute constant depending only on p . Thus

$\tau_p = (d_p^2 - 1)^{\frac{1}{4p}} (d_p^2 + 1)^{-\frac{1}{4p}} \geq (r_p^2 - 1)^{\frac{1}{4p}} (r_p^2 + 1)^{-\frac{1}{4p}} > 0$. Applying Lemma 2.6.2 and putting the above bounds together, one obtains

$$\begin{aligned}\|\mathbf{V}_p^{-1}\|_{\infty} h_m &\leq 2\tau_p^{-2p} \|\mathbf{V}_p^{-1}\|_2 (1 - \tau_p)^{-1} h_m \\ &\leq 2 \left(\frac{r_p^2 + 1}{r_p^2 - 1} \right)^{1/2} \lambda_{\min}^{-1}(\mathbf{V}_p) \times \left\{ 1 - \left(\frac{C_p^2 c_p^{-2} - 1}{C_p^2 c_p^{-2} + 1} \right)^{1/4p} \right\}^{-1} h_m \\ &\leq 2 \left(\frac{r_p^2 + 1}{r_p^2 - 1} \right)^{1/2} c_p^{-1} \left\{ 1 - \left(\frac{C_p^2 c_p^{-2} - 1}{C_p^2 c_p^{-2} + 1} \right)^{1/4p} \right\}^{-1} \equiv M_p.\end{aligned}$$

The lemma is proved.

For any function $\phi \in C[0, 1]$, denote the vector $\boldsymbol{\phi} = (\phi(1/N), \dots, \phi(N/N))^T$ and function

$$\tilde{\phi}(x) \equiv \{B_{1-p,p}(x), \dots, B_{N_m,p}(x)\} \left(\mathbf{X}^T \mathbf{X} \right)^{-1} \mathbf{X}^T \boldsymbol{\phi}.$$

Lemma 2.6.4. *Under Assumption (A3), for $\mathbf{V}_p$ and $\hat{\mathbf{V}}_p$ defined in (2.17), $\|\mathbf{V}_p - \hat{\mathbf{V}}_p\|_{\infty} = O(N^{-1})$ and $\|\hat{\mathbf{V}}_p^{-1}\|_{\infty} \leq 2M_p h_m^{-1}$, for large enough n . There exists $c_{\phi,p} \in (0, \infty)$ such that when n is large enough, $\|\tilde{\phi}\|_{\infty} \leq c_{\phi,p} \|\phi\|_{\infty}$ for any $\phi \in C[0, 1]$. Furthermore, if*

$\phi \in C^{p-1,\mu}[0,1]$ for some $\mu \in (0,1]$, then for $\tilde{C}_{p-1,\mu} = (c_{\phi,p} + 1) C_{p-1,\mu}$

$$\|\tilde{\phi} - \phi\|_{\infty} \leq \tilde{C}_{p-1,\mu} \|\phi^{(p-1)}\|_{0,\mu} h_m^{\mu+p-1}. \quad (2.18)$$

PROOF. We first show that $\|\mathbf{V}_p - \hat{\mathbf{V}}_p\|_{\infty} = O(N^{-1})$. Suppose $p = 1$, define for any $0 \leq J \leq N_m$, the number of design points j/N in the J -th interval I_J as N_J , then $N_J = \#\{j : j \in [NJ/(N_m + 1), N(J+1)/(N_m + 1))\}$, for $0 \leq J < N_m$, and $N_J = \#\{j : j \in [NJ/(N_m + 1), N(J+1)/(N_m + 1)]\}$, for $J = N_m$.

Clearly $\max_{0 \leq J \leq N_m} |N_J - Nh_m| \leq 1$, and hence

$$\begin{aligned} \|\mathbf{V}_1 - \hat{\mathbf{V}}_1\|_{\infty} &= \max_{0 \leq J \leq N_m} \left| \|B_{J,1}\|_{2,N}^2 - \|B_{J,1}\|_2^2 \right| \\ &= \max_{0 \leq J \leq N_m} \left| N^{-1} \sum_{j=1}^N B_{J,1}^2(j/N) - h_m \right| \\ &= \max_{0 \leq J \leq N_m} |N^{-1} N_J - h_m| = N^{-1} \max_{0 \leq J \leq N_m} |N_J - Nh_m| \leq N^{-1}. \end{aligned}$$

For $p > 1$, de Boor (2001), Page 96, B-spline property ensures that there exists a constant $C_{1,p} > 0$ such that

$$\begin{aligned} &\max_{1-p \leq J, J' \leq N_m} \max_{1 \leq j \leq N} \sup_{x \in [(j-1)/N, j/N]} |B_{J,p}(j/N) B_{J',p}(j/N) - B_{J,p}(x) B_{J',p}(x)| \\ &\leq C_{1,p} N^{-1} h_m^{-1}, \end{aligned}$$

while there exists a constant $C_{2,p} > 0$ such that $\max_{1-p \leq J, J' \leq N_m} N_{J,J'} \leq C_{2,p} Nh_m$

where $N_{J,J'} = \#\left\{j : 1 \leq j \leq N, B_{J,p}(j/N) B_{J',p}(j/N) > 0\right\}$. Hence

$$\begin{aligned}
& \left\| \mathbf{V}_p - \hat{\mathbf{V}}_p \right\|_{\infty} \\
&= \max_{1-p \leq J, J' \leq N_m} \left| N^{-1} \sum_{j=1}^N B_{J,p}(j/N) B_{J',p}(j/N) - \int_0^1 B_{J,p}(x) B_{J',p}(x) dx \right| \\
&\leq \max_{1-p \leq J, J' \leq N_m} \sum_{j=1}^N \int_{(j-1)/N}^{j/N} \left| B_{J,p}(j/N) B_{J',p}(j/N) - B_{J,p}(x) B_{J',p}(x) \right| dx \\
&\leq C_{2,p} N h_m \times N^{-1} \times C_{1,p} N^{-1} h_m^{-1} \leq C N^{-1}.
\end{aligned}$$

According to Lemma 2.6.3, for any $(N_m + p)$ vector $\boldsymbol{\gamma}$, $\left\| \mathbf{V}_p^{-1} \boldsymbol{\gamma} \right\|_{\infty} \leq h_m^{-1} \|\boldsymbol{\gamma}\|_{\infty}$. Hence, $\left\| \mathbf{V}_p \boldsymbol{\gamma} \right\|_{\infty} \geq M_p^{-1} h_m \|\boldsymbol{\gamma}\|_{\infty}$. By Assumption (A3), $N^{-1} = o(h_m)$ so if N is large enough, for any $\boldsymbol{\gamma}$, one has

$$\begin{aligned}
& \left\| \hat{\mathbf{V}}_p \boldsymbol{\gamma} \right\|_{\infty} \geq \left\| \mathbf{V}_p \boldsymbol{\gamma} \right\|_{\infty} - \left\| \mathbf{V}_p \boldsymbol{\gamma} - \hat{\mathbf{V}}_p \boldsymbol{\gamma} \right\|_{\infty} \geq h_m M_p^{-1} \|\boldsymbol{\gamma}\|_{\infty} - O\left(N^{-1}\right) \|\boldsymbol{\gamma}\|_{\infty} \\
&= \frac{h_m}{2} M_p^{-1} \|\boldsymbol{\gamma}\|_{\infty}.
\end{aligned}$$

Hence $\left\| \hat{\mathbf{V}}_p^{-1} \right\|_{\infty} \leq 2 M_p h_m^{-1}$.

To prove the last statement of the lemma, note that for any $x \in [0, 1]$ at most $(p+1)$ of the numbers $B_{1-p,p}(x), \dots, B_{N_m,p}(x)$ are between 0 and 1, others being 0, so

$$\begin{aligned}
\left| \tilde{\phi}(x) \right| &\leq (p+1) \left| \left(\mathbf{X}^T \mathbf{X} \right)^{-1} \mathbf{X}^T \boldsymbol{\phi} \right| = (p+1) \left| \hat{\mathbf{V}}_p^{-1} \left(\mathbf{X}^T \boldsymbol{\phi} N^{-1} \right) \right| \\
&\leq (p+1) \left\| \hat{\mathbf{V}}_p^{-1} \right\|_{\infty} \left| \mathbf{X}^T \boldsymbol{\phi} N^{-1} \right| \leq 2(p+1) M_p h_m^{-1} \left| \mathbf{X}^T \mathbf{I}_N N^{-1} \right| \|\boldsymbol{\phi}\|_{\infty}
\end{aligned}$$

in which $\mathbf{I}_N = (1, \dots, 1)^T$. Clearly $\left| \mathbf{X}^T \mathbf{I}_N N^{-1} \right| \leq C h_m$ for some $C > 0$, hence $\left| \tilde{\phi}(x) \right| \leq 2 M_p (p+1) C \|\boldsymbol{\phi}\|_{\infty} = c_{\phi,p} \|\boldsymbol{\phi}\|_{\infty}$. Now if $\boldsymbol{\phi} \in C^{p-1,\mu}[0, 1]$ for some $\mu \in (0, 1]$, let

$g \in \mathcal{H}^{(p-1)}[0, 1]$ be such that $\|g - \phi\|_\infty \leq C_{p-1, \mu} \left\| \phi^{(p-1)} \right\|_{0, \mu} h_m^{\mu+p-1}$ according to Theorem 2.3.1, then $\tilde{g} \equiv g$ as $g \in \mathcal{H}^{(p-1)}[0, 1]$ hence

$$\begin{aligned} \left\| \tilde{\phi} - \phi \right\|_\infty &= \left\| \tilde{\phi} - \tilde{g} - (\phi - g) \right\|_\infty \leq \left\| \tilde{\phi} - \tilde{g} \right\|_\infty + \|\phi - g\|_\infty \\ &\leq (c_{\phi, p} + 1) \|\phi - g\|_\infty \leq (c_{\phi, p} + 1) C_{p-1, \mu} \left\| \phi^{(p-1)} \right\|_{0, \mu} h_m^{\mu+p-1} \end{aligned}$$

proving (2.18).

Lemma 2.6.5. *Under Assumption (A5), for $C_0 = C_1 \left(1 + \beta C_2 \sum_{s=1}^{\infty} s^{\beta-1-\gamma_1} \right)$ and $n \geq 1$*

$$\max_{1 \leq k \leq \kappa} E \left| \bar{\xi}_{.,k} - \bar{Z}_{.,k,\xi} \right| \leq C_0 n^{\beta-1}, \quad (2.19)$$

$$\max_{1 \leq j \leq N} \left| \bar{\varepsilon}_{.,j} - \bar{Z}_{.,j,\varepsilon} \right| = O_{a.s.} \left(n^{\beta-1} \right) \quad (2.20)$$

where $\bar{Z}_{.,k,\xi} = n^{-1} \sum_{i=1}^n Z_{ik,\xi}$, $\bar{Z}_{.,j,\varepsilon} = n^{-1} \sum_{i=1}^n Z_{ij,\varepsilon}$, $1 \leq j \leq N$, $1 \leq k \leq \kappa$. Also

$$\max_{1 \leq k \leq \kappa} E \left| \bar{\xi}_{.,k} \right| \leq n^{-1/2} (2/\pi)^{1/2} + C_0 n^{\beta-1}. \quad (2.21)$$

PROOF. The proof of (2.20) is trivial. Assumption (A5) entails that

$\bar{F}_{n+t,k} < C_2 (n+t)^{-\gamma_1}$, $k = 1, \dots, \kappa$, $t = 0, 1, \dots, \infty$, in which

$\bar{F}_{n+t,k} = P \left[\left| \sum_{i=1}^n \xi_{ik} - \sum_{i=1}^n Z_{ik,\xi} \right| > C_1 (n+t)^\beta \right]$. Taking expectation, one has

$$\begin{aligned}
& E \left| \sum_{i=1}^n \xi_{ik} - \sum_{i=1}^n Z_{ik,\xi} \right| \\
& \leq C_1 (n+0)^\beta + \sum_{t=1}^{\infty} C_1 (n+t)^\beta \left(\bar{F}_{n+t-1,k} - \bar{F}_{n+t,k} \right) \\
& \leq C_1 n^\beta + \sum_{t=0}^{\infty} C_1 C_2 (n+t)^{-\gamma_1} \beta (n+t)^{\beta-1} \\
& \leq C_1 \left\{ n^\beta + \beta C_2 \sum_{t=0}^{\infty} (n+t)^{\beta-1-\gamma_1} \right\} \\
& \leq n^\beta C_1 \left[1 + \beta C_2 n^{-1-\gamma_1} \sum_{s=1}^{\infty} \sum_{t=sn-n}^{sn-1} (1+t/n)^{\beta-1-\gamma_1} \right] \\
& \leq n^\beta C_1 \left[1 + \beta C_2 n^{-1-\gamma_1} \times n \sum_{t=1}^{\infty} t^{\beta-1-\gamma_1} \right] \leq C_0 n^\beta,
\end{aligned}$$

which proves (2.19) if one divides the above inequalities by n . The fact that $\bar{Z}_{.,k,\xi} \sim N(0, 1/n)$ entails that $E \left| \bar{Z}_{.,k,\xi} \right| = n^{-1/2} (2/\pi)^{1/2}$ and thus $\max_{1 \leq k \leq \kappa} E \left| \bar{\xi}_{.,k} \right| \leq n^{-1/2} (2/\pi)^{1/2} + C_0 n^{\beta-1}$.

Lemma 2.6.6. *Assumption (A5) holds under Assumption (A5').*

PROOF. Under Assumption (A5'), $E \left| \xi_{ik} \right|^{\tau_1} < +\infty$, $\tau_1 > 4$, $E \left| \varepsilon_{ij} \right|^{\tau_2} < +\infty$, $\tau_2 > 4 + 2\theta$, so there exists some $\beta \in (0, 1/2)$ such that $\tau_1 > 2/\beta$, $\tau_2 > (2 + \theta)/\beta$.

Now let $H(x) = x^{\tau_1}$, then Lemma 2.6.1 entails that there exists constants C_{1k}, C_{2k}, a_k which depend on the distribution of ξ_{ik} , such that for

$x_n = C_{1k} n^\beta$, $\frac{n}{H(a_k x_n)} = a_k^{-\tau_1} C_{1k}^{-\tau_1} n^{1-\tau_1\beta}$ and i.i.d. $N(0, 1)$ variables $Z_{ik,\xi}$ such that

$$P \left[\max_{1 \leq t \leq n} \left| \sum_{i=1}^t \xi_{ik} - \sum_{i=1}^t Z_{ik,\xi} \right| > C_{1k} n^\beta \right] < C_{2k} a_k^{-\tau_1} C_{1k}^{-\tau_1} n^{1-\tau_1\beta}.$$

Since $\tau_1 > 2/\beta$, $\gamma_1 = \tau_1\beta - 1 > 1$. If the number κ of k is finite, so there are common con-

stants $C_1, C_2 > 0$ such that $P \left[\max_{1 \leq t \leq n} \left| \sum_{i=1}^t \xi_{ik} - \sum_{i=1}^t Z_{ik, \xi} \right| > C_1 n^\beta \right] < C_2 n^{-\gamma_1}$ which entails (2.3) since κ is finite. If κ is infinite but all the ξ_{ik} 's are i.i.d., then C_{1k}, C_{2k}, a_k are the same for all k , so the above is again true.

Likewise, under Assumption (A5'), Lemma 2.6.1 applied to $H(x) = x^{\tau_2}$ implies that there exists constants C_1, C_2 and a which depend on the distribution of ε_{ij} and i.i.d. $N(0, 1)$ variables $Z_{ij, \varepsilon}$, such that with $x_n = C_1 n^\beta$, $\frac{n}{H(a_k x_n)} = a^{-\tau_2} C_1^{-\tau_2} n^{1-\tau_2 \beta}$ and

$$\max_{1 \leq j \leq N} P \left\{ \max_{1 \leq t \leq n} \left| \sum_{i=1}^t \varepsilon_{ij} - \sum_{i=1}^t Z_{ij, \varepsilon} \right| > C_1 n^\beta \right\} \leq C_2 a^{-\tau_2} C_1^{-\tau_2} n^{1-\tau_2 \beta},$$

Hence, $\tau_2 \beta > 2 + \theta$ which implies that there is $\gamma_2 > 1$ such that $\tau_2 \beta - 1 > \gamma_2 + \theta$ and (2.4) follows.

PROOF OF PROPOSITION 2.3.1.

Applying (2.18), $\|\tilde{m}_p - m\|_\infty \leq C_{p-1,1} h_m^p$. Since Assumption (A3) implies that $O(h_m^p n^{1/2}) = o(1)$, equation (2.10) is proved.

PROOF OF PROPOSITION 2.3.2.

Denote by $\tilde{\mathbf{Z}}_{p, \varepsilon}(x) = \left\{ B_{1-p, p}(x), \dots, B_{N_m, p}(x) \right\} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Z}$, where $\mathbf{Z} = \left(\sigma(1/N) \bar{Z}_{\cdot 1, \varepsilon}, \dots, \sigma(N/N) \bar{Z}_{\cdot N, \varepsilon} \right)^T$. By (2.20), one has $\|\mathbf{Z} - \mathbf{e}\|_\infty = O_{a.s.}(n^{\beta-1})$, while

$$\begin{aligned} \left\| N^{-1} \mathbf{X}^T (\mathbf{Z} - \mathbf{e}) \right\|_\infty &\leq \|\mathbf{Z} - \mathbf{e}\|_\infty \max_{1-p \leq J \leq N_m} \left\langle B_{J, p}, 1 \right\rangle_{2, N} \\ &\leq C \|\mathbf{Z} - \mathbf{e}\|_\infty \max_{1-p \leq J \leq N_m} \# \left\{ j : B_{J, p}(j/N) > 0 \right\} N^{-1} \leq C \|\mathbf{Z} - \mathbf{e}\|_\infty h_m. \end{aligned}$$

Also for any fixed $x \in [0, 1]$, one has

$$\begin{aligned} \left\| \tilde{\mathbf{Z}}_{p,\varepsilon}(x) - \tilde{e}_p(x) \right\|_\infty &= \left\| \left\{ B_{1-p,p}(x), \dots, B_{N_m,p}(x) \right\} \hat{\mathbf{V}}_p^{-1} N^{-1} \mathbf{X}^T (\mathbf{Z} - \mathbf{e}) \right\|_\infty \\ &\leq C \left\| \hat{\mathbf{V}}_p^{-1} \right\|_\infty \left\| \mathbf{Z} - \mathbf{e} \right\|_\infty h_m = O_{a.s.} \left(n^{\beta-1} \right). \end{aligned}$$

Note next that the random vector $\hat{\mathbf{V}}_p^{-1} N^{-1} \mathbf{X}^T \mathbf{Z}$ is $(N_m + p)$ -dimensional normal with covariance matrix $N^{-2} \hat{\mathbf{V}}_p^{-1} \mathbf{X}^T \text{var}(\mathbf{Z}) \mathbf{X} \hat{\mathbf{V}}_p^{-1}$, bounded above by

$$\max_{x \in [0,1]} \sigma^2(x) \left\| n^{-1} N^{-1} \hat{\mathbf{V}}_p^{-1} \hat{\mathbf{V}}_p \hat{\mathbf{V}}_p^{-1} \right\|_\infty \leq C N^{-1} n^{-1} \left\| \hat{\mathbf{V}}_p^{-1} \right\|_\infty \leq C N^{-1} n^{-1} h_m^{-1},$$

bounding the tail probabilities of entries of $\hat{\mathbf{V}}_p^{-1} N^{-1} \mathbf{X}^T \mathbf{Z}$ and applying Borel-Cantelli Lemma leads to

$$\begin{aligned} \left\| \hat{\mathbf{V}}_p^{-1} N^{-1} \mathbf{X}^T \mathbf{Z} \right\|_\infty &= O_{a.s.} \left(N^{-1/2} n^{-1/2} h_m^{-1/2} \log^{1/2}(N_m + p) \right) \\ &= O_{a.s.} \left(N^{-1/2} n^{-1/2} h_m^{-1/2} \log^{1/2} n \right). \end{aligned}$$

Hence, $\sup_{x \in [0,1]} \left| n^{1/2} \tilde{\mathbf{Z}}_{p,\varepsilon}(x) \right| = O_{a.s.} \left(N^{-1/2} h_m^{-1/2} \log^{1/2} n \right)$ and

$$\sup_{x \in [0,1]} \left| n^{1/2} \tilde{e}_p(x) \right| = O_{a.s.} \left(n^{\beta-1/2} + N^{-1/2} h_m^{-1/2} \log^{1/2} n \right) = o_{a.s.}(1).$$

Thus (2.11) holds according to Assumption (A3).

PROOF OF PROPOSITION 2.3.3.

We denote $\tilde{\zeta}_k(x) = \bar{Z}_{\cdot k, \xi} \phi_k(x)$, $k = 1, \dots, \kappa$ and define

$$\tilde{\zeta}(x) = n^{1/2} \left[\sum_{k=1}^{\kappa} \{\phi_k(x)\}^2 \right]^{-1/2} \sum_{k=1}^{\kappa} \tilde{\zeta}_k(x) = n^{1/2} G(x, x)^{-1/2} \sum_{k=1}^{\kappa} \tilde{\zeta}_k(x).$$

It is clear that $\tilde{\zeta}(x)$ is a Gaussian process with mean 0, variance 1 and covariance $E \tilde{\zeta}(x) \tilde{\zeta}(x') = G(x, x)^{-1/2} G(x, x')^{-1/2} G(x, x')$, for any $x, x' \in [0, 1]$. Thus $\tilde{\zeta}(x)$, $x \in [0, 1]$ has the same distribution as $\zeta(x)$, $x \in [0, 1]$.

Using Lemma 2.6.4, one obtains that

$$\|\tilde{\phi}_k\|_{\infty} \leq c_{\phi, p} \|\phi_k\|_{\infty}, \|\tilde{\phi}_k - \phi_k\|_{\infty} \leq \tilde{C}_{0, \mu} \|\phi_k\|_{0, \mu} h_m^{\mu}, 1 \leq k \leq \kappa. \quad (2.22)$$

Applying the above (2.22), (2.21) and Assumptions (A3), (A4), one has

$$\begin{aligned} & E n^{1/2} \sup_{x \in [0, 1]} G(x, x)^{-1/2} \left| \sum_{k=1}^{\kappa} \bar{\xi}_{\cdot k} \left\{ \phi_k(x) - \tilde{\phi}_k(x) \right\} \right| \\ & \leq C n^{1/2} \left\{ \sum_{k=1}^{\kappa n} E |\bar{\xi}_{\cdot k}| \|\phi_k\|_{0, \mu} h_m^{\mu} + \sum_{k=\kappa n+1}^{\kappa} E |\bar{\xi}_{\cdot k}| \|\phi_k\|_{\infty} \right\} \\ & \leq C \left\{ \sum_{k=1}^{\kappa n} \|\phi_k\|_{0, \mu} h_m^{\mu} + \sum_{k=\kappa n+1}^{\kappa} \|\phi_k\|_{\infty} \right\} = o(1), \end{aligned}$$

hence

$$n^{1/2} \sup_{x \in [0, 1]} G(x, x)^{-1/2} \left| \sum_{k=1}^{\kappa} \bar{\xi}_{\cdot k} \left\{ \phi_k(x) - \tilde{\phi}_k(x) \right\} \right| = o_P(1). \quad (2.23)$$

In addition, (2.19) and Assumptions (A3) and (A4) entail

$$\begin{aligned} & E n^{1/2} \sup_{x \in [0, 1]} G(x, x)^{-1/2} \left| \sum_{k=1}^{\kappa} \left(\bar{Z}_{\cdot k, \xi} - \bar{\xi}_{\cdot k} \right) \phi_k(x) \right| \\ & \leq C n^{\beta-1/2} \sum_{k=1}^{\kappa} \|\phi_k\|_{\infty} = o(1). \end{aligned}$$

Hence

$$n^{1/2} \sup_{x \in [0,1]} G(x, x)^{-1/2} \left| \sum_{k=1}^{\kappa} \left(\bar{Z}_{\cdot k, \xi} - \bar{\xi}_{\cdot k} \right) \phi_k(x) \right| = o_P(1). \quad (2.24)$$

Note that

$$\begin{aligned} \bar{m}(x) - m(x) - \tilde{\xi}_p(x) &= \sum_{k=1}^{\kappa} \bar{\xi}_{\cdot k} \left\{ \phi_k(x) - \tilde{\phi}_k(x) \right\}, \\ n^{-1/2} G(x, x)^{1/2} \tilde{\zeta}(x) - \{\bar{m}(x) - m(x)\} &= \sum_{k=1}^{\kappa} \left(\bar{Z}_{\cdot k, \xi} - \bar{\xi}_{\cdot k} \right) \phi_k(x) \end{aligned}$$

hence

$$\begin{aligned} n^{1/2} \sup_{x \in [0,1]} G(x, x)^{-1/2} \left| \bar{m}(x) - m(x) - \tilde{\xi}_p(x) \right| &= o_P(1), \\ \sup_{x \in [0,1]} \left| \tilde{\zeta}(x) - n^{1/2} G(x, x)^{-1/2} \{\bar{m}(x) - m(x)\} \right| &= o_P(1). \end{aligned}$$

according to (2.23) and (2.24), which leads to both (2.12) and (2.13).

Chapter 3

Confidence Envelopes for Covariance Functions

3.1 Introduction

Covariance estimation is crucial in both functional and longitudinal data analysis. For longitudinal data, a good estimation of the covariance function improves the estimation efficiency of the mean parameters (Wang, Carroll and Lin, 2005; Fan, Huang and Li, 2007). In functional data analysis (Ramsay and Silverman, 2005), covariance estimation plays a critical role in functional principal component analysis (James, Hastie and Sugar, 2000; Zhao, Marron and Wells, 2004; Yao, Müller and Wang, 2005a; Hall, Müller and Wang, 2006; Yao and Lee, 2006; Zhou, Huang and Carroll, 2008; Li and Hsing, 2010a), functional generalized linear models (Cai and Hall, 2005; Yao, Müller and Wang, 2005b; Li, Wang and Carroll, 2010), and other functional nonlinear models (James and Silverman, 2005; Li and Hsing, 2010b). Other related work on functional data analysis includes Ferraty and Vieu (2006) and Morris and Carroll (2006).

There are some important recent works on nonparametric covariance estimation in functional data, which are mostly based on kernel smoothing, for example Yao et al. (2005a), Hall et al. (2006) and Li and Hsing (2010a). More recently, Cai and Yuan (2010) also proposed a smoothing spline covariance estimator. So far, all existing work concentrated on estimation and the corresponding asymptotic convergence rate. There is no theoretical or methodological development for inference procedures on the covariance functions, such as simultaneous or uniform confidence envelopes. Nonparametric simultaneous confidence regions are powerful tools for making global inference on functions; see Härdle and Marron (1991), Claeskens and Van Keilegom (2003) and Zhao and Wu (2008) for related theory and applications.

In this chapter, we consider a typical functional data setting where the functions are recorded on a dense regular grid in an interval $\mathcal{X}$ and the measurements are contaminated with measurement errors. Some recent applications of this type of functional data include near infrared spectra (Li and Hsing, 2010a), recorded speeches for voice recognition (Hastie, Tibshirani and Buja, 1995), electroencephalogram (EEG) data (Crainiceanu, Stacu and Di, 2009). We propose to estimate the covariance function by tensor product B-splines. In contrast with the kernel methods (Yao et al., 2005a; Hall et al., 2006; Li and Hsing, 2010b), our proposed spline estimator is much more efficient in terms of computation. The reason is that the kernel smoothers are evaluated pointwisely, while for the spline estimator, we only need to solve for a small number of spline coefficients to have an explicit expression for the whole function. For smoothing a two-dimensional covariance surface with a moderate sample size, the kernel smoother might take up to half an hour, while our spline estimator only takes a few seconds. Computation efficiency is a huge advantage for the spline methods in analyzing large data sets and in performing simulation studies. The reader is referred to

Huang and Yang (2004) for more discussions on the computational merits of spline methods. Compared with the smoothing spline approach of Cai and Yuan (2010), our method uses reduced rank tensor product B-splines, and is potentially faster while analyzing large data sets.

We show that the estimation error in the mean function is asymptotically negligible in estimating the covariance function, and our covariance estimator is as efficient as an “oracle” estimator where the true mean function is known. We derive both local and global asymptotic distribution for the proposed spline covariance estimator. Especially, based on the asymptotic distribution of the maximum deviation of the estimator, we propose a new simultaneous confidence envelope for the covariance function, which can be used to visualize the variability of the covariance estimator and to make global inferences on the shape of the true covariance.

We apply the proposed confidence envelope method to a Tecator near infrared spectra data set to test the hypothesis that the covariance is stationary. In a speech recognition application, the classic functional linear discriminant analysis (Hastie et al., 1995; James and Hastie, 2001) assumes that the random curves from different classes share a common covariance function. We further extend our confidence envelope method to a two-sample problem, where one can test whether the covariance functions from two groups are different.

We organize this chapter as follows. In Section 3.2 we describe the data structure and the proposed spline covariance estimator. In Section 3.3, we study the local and global asymptotic properties of the proposed estimator. Based on the theory, we propose a new confidence envelope approach and extend the method to two-sample hypothesis testing problems. More implementation details of the proposed confidence envelopes are provided in Section 3.4. We present simulation studies in Section 3.5 and applications to the Tecator infrared spec-

troscopy and the speech recognition data set in Sections 3.6. Some concluding remarks are provided in Section 3.7. All proofs of the theorems and technical lemmas are provided in the appendix and the supplementary material.

3.2 Spline covariance estimation

3.2.1 Data structure and model assumptions

Following Ramsay and Silverman (2005), the data that we consider are a collection of trajectories $\{\eta_i(x)\}_{i=1}^n$ which are i.i.d. realizations of a smooth random function $\eta(x)$, defined on a continuous interval $\mathcal{X}$. Assume that $\{\eta(x), x \in \mathcal{X}\}$ is a L^2 process, i.e. $E \int_{\mathcal{X}} \eta^2(x) dx < +\infty$, and define the mean and covariance functions as $m(x) = E\{\eta(x)\}$ and $G(x, x') = \text{cov}\{\eta(x), \eta(x')\}$. The covariance function is a symmetric nonnegative-definite function with a spectral decomposition, $G(x, x') = \sum_{k=1}^{\kappa} \lambda_k \psi_k(x) \psi_k(x')$, where $\lambda_1 \geq \lambda_2 \geq \dots \geq 0$, $\sum_{k=1}^{\kappa} \lambda_k < +\infty$, are the eigenvalues, and $\{\psi_k(x)\}_{k=1}^{\kappa}$ are the corresponding eigenfunctions and are a set of orthonormal functions in $L^2(\mathcal{X})$. By the Karhunen-Loève representation, $\eta_i(x) = m(x) + \sum_{k=1}^{\kappa} \xi_{ik} \phi_k(x)$, where the random coefficients ξ_{ik} are uncorrelated with mean 0 and variance 1, and the functions $\phi_k = \sqrt{\lambda_k} \psi_k$. In the standard Karhunen-Loève expansion, κ can diverge to ∞ . For practical consideration, κ is always truncated at a finite number in real data analysis. Our main theoretical results are developed under the assumption that κ is a finite positive integer, but some of our theoretical results can be further generalized to the infinite dimension case, which will be discussed in Section 3.3.

Without loss of generality, we take $\mathcal{X} = [0, 1]$. Then the observed data are $Y_{ij} = \eta_i(X_{ij}) + \sigma(X_{ij}) \varepsilon_{ij}$, for $1 \leq i \leq n$, $1 \leq j \leq N$, where $X_{ij} = j/N$, ε_{ij} are i.i.d.

random errors with $E(\varepsilon_{11}) = 0$ and $E(\varepsilon_{11}^2) = 1$, and $\sigma^2(x)$ is the variance function of the measurement errors. By the Karhunen-Loève representation, the observed data can be written as

$$Y_{ij} = m(j/N) + \sum_{k=1}^{\kappa} \xi_{ik} \phi_k(j/N) + \sigma(j/N) \varepsilon_{ij}.$$

We model $m(\cdot)$ and $G(\cdot, \cdot)$ as nonparametric functions, and hence $\{\lambda_k\}_{k=1}^{\kappa}$, $\{\phi_k(\cdot)\}_{k=1}^{\kappa}$ and $\{\xi_{ik}\}_{k=1}^{\kappa}$ are unknown and need to be estimated.

3.2.2 Spline covariance estimator

To describe the tensor product spline estimator of the covariance functions, we first introduce some notation. Denote a sequence of equally-spaced points $\{t_J\}_{J=1}^{N_s}$, called interior knots which have been defined in Chapter 1. Let $h_s = 1/(N_s + 1)$ be the distance between neighboring knots. Let $\mathcal{H}^{(p-2)} = \mathcal{H}^{(p-2)}[0, 1]$ be the polynomial spline space of order p . The J^{th} B-spline of order p is denoted by $B_{J,p}$ as in de Boor (2001). Thus we define the tensor product spline space as

$$\begin{aligned} \mathcal{H}^{(p-2),2}[0,1]^2 &\equiv \mathcal{H}^{(p-2),2} = \mathcal{H}^{(p-2)} \otimes \mathcal{H}^{(p-2)} \\ &= \left\{ \sum_{J,J'=1-p}^{N_s} b_{JJ'p} B_{J,p}(x) B_{J',p}(x'), b_{JJ'p} \in R, x, x' \in [0,1] \right\}. \end{aligned}$$

If the mean function $m(x)$ was known, one could compute the errors

$$U_{ij} \equiv Y_{ij} - m(j/N) = \sum_{k=1}^{\kappa} \xi_{ik} \phi_k(j/N) + \sigma(j/N) \varepsilon_{ij}, \quad 1 \leq i \leq n, \quad 1 \leq j \leq N.$$

Denote $\bar{U}_{.jj'} = n^{-1} \sum_{i=1}^n U_{ij} U_{ij'}$, $1 \leq j \neq j' \leq N$, one can then define the “oracle”

estimator of the covariance function

$$\tilde{G}_{p_2}(\cdot, \cdot) = \underset{g(\cdot, \cdot) \in \mathcal{H}^{(p_2-2), 2}}{\operatorname{argmin}} \sum_{1 \leq j \neq j' \leq N} \left\{ \bar{U}_{\cdot jj'} - g(j/N, j'/N) \right\}^2, \quad (3.1)$$

using tensor product splines of order $p_2 \geq 2$. Since the mean function $m(x)$ is unavailable when one analyzes data, one can use instead the spline smoother of $m(x)$, i.e.,

$$\hat{m}_{p_1}(\cdot) = \underset{g(\cdot) \in \mathcal{H}^{(p_1-2)}}{\operatorname{argmin}} \sum_{i=1}^n \sum_{j=1}^N \left\{ Y_{ij} - g(j/N) \right\}^2, \quad p_1 \geq 1.$$

To mimic the above “oracle” smoother, we define

$$\hat{G}_{p_1, p_2}(\cdot, \cdot) = \underset{g(\cdot, \cdot) \in \mathcal{H}^{(p_2-2), 2}}{\operatorname{argmin}} \sum_{1 \leq j \neq j' \leq N} \left\{ \hat{\bar{U}}_{\cdot jj', p_1} - g(j/N, j'/N) \right\}^2, \quad (3.2)$$

where $\hat{\bar{U}}_{\cdot jj', p_1} = n^{-1} \sum_{i=1}^n \hat{U}_{ijp_1} \hat{U}_{ij'p_1}$ with $\hat{U}_{ijp_1} = Y_{ij} - \hat{m}_{p_1}(j/N)$. Let N_{s_1} be the number of interior knots for mean estimation, and N_{s_2} be the number of interior knots for $\hat{G}_{p_1, p_2}(x, x')$ in each coordinate. In other words, we have $N_{s_2}^2$ interior knots for the tensor product spline space $\mathcal{H}^{(p_2-2), 2}$.

We now provide detailed algorithm for the spline covariance estimator. For simplicity, denote $B_{JJ', p_2}(x, x') = B_{J, p_2}(x) B_{J', p_2}(x')$ and

$$\begin{aligned} \mathbf{B}_{p_2}(x, x') &= \left(B_{1-p_2, 1-p_2, p_2}(x, x'), \dots, B_{N_{s_2}, 1-p_2, p_2}(x, x'), \right. \\ &\quad \left. \dots, B_{1-p_2, N_{s_2}, p_2}(x, x'), \dots, B_{N_{s_2}, N_{s_2}, p_2}(x, x') \right)^T, \\ \mathbf{X} &= \left\{ \mathbf{B}_{p_2}\left(\frac{2}{N}, \frac{1}{N}\right), \dots, \mathbf{B}_{p_2}\left(1, \frac{1}{N}\right), \dots, \mathbf{B}_{p_2}\left(\frac{1}{N}, 1\right), \dots, \mathbf{B}_{p_2}\left(1 - \frac{1}{N}, 1\right) \right\}^T. \end{aligned}$$

Then $\hat{G}_{p_1,p_2}(x, x')$ defined in (3.2) can be rewritten as

$$\hat{G}_{p_1,p_2}(x, x') \equiv \hat{\boldsymbol{\beta}}_{p_1,p_2}^T \mathbf{B}_{p_2}(x, x'), \quad (3.3)$$

where $\hat{\boldsymbol{\beta}}_{p_1,p_2}$ is the collector of the estimated spline coefficients by solving the following least squares

$$\hat{\boldsymbol{\beta}}_{p_1,p_2} = \underset{\mathbf{b}_{p_2} \in R^{(N_{s_2}+p_2)^2}}{\operatorname{argmin}} \sum_{1 \leq j \neq j' \leq N} \left\{ \hat{\bar{U}}_{\cdot jj', p_1} - \mathbf{b}_{p_2}^T \mathbf{B}_{p_2}(j/N, j'/N) \right\}^2.$$

By elementary algebra, one obtains

$$\begin{aligned} \hat{G}_{p_1,p_2}(x, x') &= \mathbf{B}_{p_2}^T(x, x') \left(\mathbf{X}^T \mathbf{X} \right)^{-1} \mathbf{X}^T \hat{\bar{\mathbf{U}}}_{p_1}, \\ \tilde{G}_{p_2}(x, x') &= \mathbf{B}_{p_2}^T(x, x') \left(\mathbf{X}^T \mathbf{X} \right)^{-1} \mathbf{X}^T \bar{\mathbf{U}}, \end{aligned} \quad (3.4)$$

where

$$\begin{aligned} \hat{\bar{\mathbf{U}}}_{p_1} &= \left(\hat{\bar{U}}_{\cdot 21, p_1}, \dots, \hat{\bar{U}}_{\cdot N1, p_1}, \dots, \hat{\bar{U}}_{\cdot 1N, p_1}, \dots, \hat{\bar{U}}_{\cdot (N-1)N, p_1} \right)^T, \\ \bar{\mathbf{U}} &= \left(\bar{U}_{\cdot 21}, \dots, \bar{U}_{\cdot N1}, \dots, \bar{U}_{\cdot 1N}, \dots, \bar{U}_{\cdot (N-1)N} \right)^T. \end{aligned}$$

3.3 Asymptotic theory and simultaneous confidence envelopes

3.3.1 Assumptions and the oracle property

For any $\nu \in (0, 1]$, we denote $C^{q,\nu}[0, 1]$ as the space of ν -Hölder continuous functions on $[0, 1]$,

$$C^{q,\nu}[0, 1] = \left\{ \phi : \sup_{x \neq x', x, x' \in [0, 1]} \frac{|\phi^{(q)}(x) - \phi^{(q)}(x')|}{|x - x'|^\nu} < +\infty \right\}.$$

We need the following technical assumptions.

(B1) *The regression function $m \in C^{p_1-1,1}[0, 1]$.*

(B2) *The standard deviation function $\sigma(x) \in C^{0,\nu}[0, 1]$. For any $k = 1, 2, \dots, \kappa$, $\phi_k(x) \in C^{p_2-1,\nu}[0, 1]$. Also $\sup_{(x,x') \in [0,1]^2} G(x, x') < C$, for some positive constant C and $\min_{x \in [0,1]} G(x, x) > 0$.*

(B3) *The number of knots N_{s_1} and N_{s_2} satisfy $n^{1/(4p_1)} \ll N_{s_1} \ll N$, $n^{1/(2p_2)} \ll N_{s_2} \ll \min(N^{1/2}, n^{1/3})$ and $N_{s_2} \ll N_{s_1}^{p_1}$.*

(B4) *The number κ of nonzero eigenvalues is finite. The variables $(\xi_{ik})_{i=1, k=1}^{\infty, \kappa}$ and $(\varepsilon_{ij})_{i=1, j=1}^{\infty, \infty}$ are independent. In addition, $E\varepsilon_{11} = 0$, $E\varepsilon_{11}^2 = 1$, $E\xi_{1k} = 0$, $E\xi_{1k}^2 = 1$ and $\max_{1 \leq k \leq \kappa} E|\xi_{1k}|^{\delta_1} < +\infty$, $E|\varepsilon_{11}|^{\delta_2} < +\infty$, for some $\delta_1, \delta_2 > 4$.*

Assumptions (B1)-(B4) are standard in the spline smoothing literature; see Huang (2003), for instance. In particular, (B1) and (B2) guarantee the orders of the bias terms of the spline smoothers for $m(x)$ and $\phi_k(x)$. Assumption (B3) is a weak assumption to ensure the order

of the bias and noise terms in Propositions 3.3.1 and 3.3.2. Assumption (B4) is necessary for strong approximation. More discussion about the assumptions is in Section 3.4.

To gain a deeper understanding on the behavior of the spline covariance estimator $\hat{G}_{p_2}$ in (3.2), we first study the asymptotic property of $\tilde{G}_{p_2}$ in (3.1).

Denote by

$$\Delta(x, x') = \sum_{k, k'=1}^{\kappa} \phi_k(x) \phi_{k'}(x') \left(\bar{\xi}_{\cdot, kk'} - \delta_{kk'} \right), \quad (3.5)$$

where $\bar{\xi}_{\cdot, kk'} = n^{-1} \sum_{i=1}^n \xi_{ik} \xi_{ik'}$ and $\delta_{kk'} = 1$ for $k = k'$ and 0 otherwise.

Proposition 3.3.1. *Under Assumptions (B2)-(B4), one has*

$$\sup_{(x, x') \in [0, 1]^2} \left| \tilde{G}_{p_2}(x, x') - G(x, x') - \Delta(x, x') \right| = o_p(n^{-1/2}). \quad (3.6)$$

The proof of Proposition 3.3.1 is provided in the supplementary material. The next proposition provides that the tensor product spline estimator $\hat{G}_{p_1, p_2}$ is uniformly close to the “oracle” smoother at the rate of $o_p(n^{-1/2})$.

Proposition 3.3.2. *Under Assumptions (B1)-(B4), one has*

$$\sup_{(x, x') \in [0, 1]^2} \left| \hat{G}_{p_1, p_2}(x, x') - \tilde{G}_{p_2}(x, x') \right| = o_p(n^{-1/2}).$$

The proof of Proposition 3.3.2 is provided in the supplementary material. As a result of Propositions 3.3.1 and 3.3.2,

$$\sup_{(x, x') \in [0, 1]^2} \left| \hat{G}_{p_1, p_2}(x, x') - G(x, x') - \Delta(x, x') \right| = o_p(n^{-1/2}).$$

3.3.2 Asymptotic confidence envelopes

The next theorem provides a pointwise approximation to the mean squared error of $\hat{G}_{p_1, p_2}(x, x')$.

Theorem 3.3.1. *Under Assumptions (B1)-(B4),*

$$nE[\hat{G}_{p_1, p_2}(x, x') - G(x, x')]^2 = V(x, x') + o(1),$$

where $V(x, x') = G(x, x')^2 + G(x, x)G(x', x') + \sum_{k=1}^{\kappa} \phi_k^2(x) \phi_k^2(x') (E\xi_{1k}^4 - 3)$.

Remark 3.3.1. Although the convergence result in Theorem 3.3.1 is derived under the assumption that κ is a finite positive number, it continues to hold when $\kappa \rightarrow \infty$ as long as the sum in the definition of $V(x, x')$ still converges. The existence of $V(x, x')$ in the infinite dimension case is guaranteed by imposing an additional assumption that $E[\{\sup_{x \in \mathcal{X}} \eta(x)\}^{4+\delta}] < \infty$, for some $\delta > 0$, which is commonly assumed in functional data analysis (see Li and Hsing, 2010a). The assumption κ being finite is merely a practical consideration for the development of inference procedures. Further discussion on how to choose κ in real data will be provided in Section 3.4.

To obtain the quantile of the distribution of $n^{1/2} |\hat{G}_{p_1, p_2}(x, x') - G(x, x')| V^{-1/2}(x, x')$, one defines

$$\begin{aligned} & \zeta_Z(x, x') \\ &= \left\{ \sum_{k \neq k'}^{\kappa} Z_{kk'} \phi_k(x) \phi_{k'}(x') + \sum_{k=1}^{\kappa} \phi_k(x) \phi_k(x') Z_k (E\xi_{1k}^4 - 1)^{1/2} \right\} V^{-1/2}(x, x'), \end{aligned} \tag{3.7}$$

where $Z_{kk'} = Z_{k'k}$ and Z_k are i.i.d. standard gaussian random variables. Hence, for any $(x, x') \in [0, 1]^2$, $\zeta_Z(x, x')$ is a standardized gaussian field such that $E\zeta_Z(x, x') = 0$,

$E\zeta_Z^2(x, x') = 1$. Define $Q_{1-\alpha}$ as the $100(1-\alpha)^{th}$ percentile of the absolute maxima distribution of $\zeta_Z(x, x')$, $\forall (x, x') \in [0, 1]^2$, i.e.

$$P \left\{ \sup_{(x, x') \in [0, 1]^2} |\zeta_Z(x, x')| \leq Q_{1-\alpha} \right\} = 1 - \alpha, \quad \forall \alpha \in (0, 1).$$

The following theorem and corollary address the simultaneous envelopes for $G(x, x')$.

Theorem 3.3.2. *Under Assumptions (B1)-(B4), for any $\alpha \in (0, 1)$,*

$$\lim_{n \rightarrow \infty} P \left\{ \sup_{(x, x') \in [0, 1]^2} n^{1/2} |\hat{G}_{p_1, p_2}(x, x') - G(x, x')| V^{-1/2}(x, x') \leq Q_{1-\alpha} \right\} = 1 - \alpha,$$

$$\lim_{n \rightarrow \infty} P \left\{ n^{1/2} |\hat{G}_{p_1, p_2}(x, x') - G(x, x')| V^{-1/2}(x, x') \leq Z_{1-\alpha/2} \right\} = 1 - \alpha, \quad \forall (x, x') \in [0, 1]^2,$$

where $Z_{1-\alpha/2}$ is the $100(1-\alpha/2)^{th}$ percentile of the standard normal distribution.

Remark 3.3.2. Although this covariance function estimator cannot be guaranteed to be positive definite, it tends to the true positive definite covariance function in probability.

The next result follows directly from Theorem 3.3.2.

Corollary 3.3.3. *Under Assumptions (B1)-(B4), as $n \rightarrow \infty$, an asymptotic $100(1-\alpha)\%$ confidence envelope for $G(x, x')$, $\forall (x, x') \in [0, 1]^2$ is*

$$\hat{G}_{p_1, p_2}(x, x') \pm n^{-1/2} Q_{1-\alpha} V^{1/2}(x, x'), \quad \forall \alpha \in (0, 1), \quad (3.8)$$

while an asymptotic $100(1-\alpha)\%$ pointwise confidence envelope for $G(x, x')$, $\forall (x, x') \in [0, 1]^2$ is

$$\hat{G}_{p_1, p_2}(x, x') \pm n^{-1/2} Z_{1-\alpha/2} V^{1/2}(x, x'), \quad \forall \alpha \in (0, 1).$$

3.3.3 Extension to two-sample test problems

In functional analysis of variance and linear discriminant analysis, it is commonly assumed that the covariance functions are the same across different treatment groups. It is natural to extend our method to the two-sample problems, where we can construct confidence envelopes for the difference between the covariances functions from two independent groups. This procedure is equivalent to two-sample t-test.

Suppose we have two independent groups of curves with sample sizes n_1 and n_2 , respectively. We denote the ratio of two sample sizes as $\hat{r} = n_1/n_2$ and assume that $\lim_{n_1 \rightarrow \infty} \hat{r} = r > 0$. Let $\hat{G}_{p_1, p_2}^{(1)}(x, x')$ and $\hat{G}_{p_1, p_2}^{(2)}(x, x')$ be the spline estimates of covariance functions $G^{(1)}(x, x')$ and $G^{(2)}(x, x')$ by (3.2). Also denote by $\zeta_{12}(x, x')$, $\forall (x, x') \in [0, 1]^2$ a standardized Gaussian process such that $E\zeta_{12}(x, x') \equiv 0$, $E\zeta_{12}^2(x, x') \equiv 1$, $\forall (x, x') \in [0, 1]^2$ with covariance function, for $x, x' \in [0, 1]$,

$$E\zeta_{12}(x, x') \zeta_{12}(x, x') = \frac{V_1(x, x') + rV_2(x, x')}{\{V_1(x, x) + rV_2(x, x)\}^{1/2} \{V_1(x', x') + rV_2(x', x')\}^{1/2}}.$$

Denote by $Q_{12, 1-\alpha}$ the $(1 - \alpha)$ -th quantile of the absolute maxima deviation of $\zeta_{12}(x, x')$, $\forall (x, x') \in [0, 1]^2$ as above.

We have the following theorem, the proof of which is analogous to that of Theorem 3.3.2 and therefore omitted.

Theorem 3.3.4. *Under Assumptions (B1)-(B4) modified for each group accordingly, for*

any $\alpha \in (0, 1)$, as $n_1 \rightarrow \infty$, $\hat{r} \rightarrow r > 0$,

$$P \left\{ \sup_{(x, x') \in [0, 1]^2} n_1^{1/2} \frac{|\hat{G}_{p_1, p_2}^{(1)}(x, x') - \hat{G}_{p_1, p_2}^{(2)}(x, x') - G^{(1)}(x, x') + G^{(2)}(x, x')|}{(V_1(x, x') + rV_2(x, x'))^{1/2}} \leq Q_{12, 1-\alpha} \right\} = 1 - \alpha.$$

Remark 3.3.3. Under Assumptions (B1)-(B4), an asymptotic $100(1 - \alpha)\%$ confidence envelope for $G^{(1)}(x, x') - G^{(2)}(x, x')$, $\forall (x, x') \in [0, 1]^2$ is:

$$\hat{G}_{p_1, p_2}^{(1)}(x, x') - \hat{G}_{p_1, p_2}^{(2)}(x, x') \pm n_1^{-1/2} Q_{12, 1-\alpha} \left(V_1(x, x') + rV_2(x, x') \right)^{1/2}, \quad \forall \alpha \in (0, 1). \quad (3.9)$$

one can use this confidence envelope to test any hypothesis on $G^{(1)}(x, x') - G^{(2)}(x, x')$.

3.4 Implementation

In this section, we describe the procedure to implement the confidence envelopes. Given the data set $(j/N, Y_{ij})_{j=1, i=1}^{N, n}$, the number of interior knots N_{s_1} for $\hat{m}_{p_1}(x)$ is taken to be $[n^{1/(4p_1)} \log n]$, where $[a]$ denotes the integer part of a . Meanwhile, the spline estimator $\hat{G}_{p_1, p_2}(x, x')$ is obtained by (3.3) with the number of interior knots $N_{s_2} = [n^{1/(2p_2)} \log \log n]$. These choices of knots satisfy condition (B3) in our theory.

To construct the confidence envelopes, one needs to evaluate the percentile $Q_{1-\alpha}$ and

estimate the variance function $V(x, x')$. An estimator $\hat{V}(x, x')$ of $V(x, x')$ is

$$\begin{aligned} & \hat{V}(x, x') \\ &= \hat{G}_{p_1, p_2}(x, x')^2 + \hat{G}_{p_1, p_2}(x, x) \hat{G}_{p_1, p_2}(x', x') \\ &+ \sum_{k=1}^{\kappa} \hat{\phi}_k^2(x) \hat{\phi}_k^2(x') \left(n^{-1} \sum_{i=1}^n \hat{\xi}_{ik}^4 - 3 \right), \end{aligned}$$

where $\hat{\phi}_k$ and $\hat{\xi}_{ik}$ are the estimators of ϕ_k and ξ_{ik} , respectively. According to Yao et al. (2005b), the estimates of eigenfunctions and eigenvalues correspond to the solutions $\hat{\phi}_k$ and $\hat{\lambda}_k$ of the eigen-equations,

$$\int_0^1 \hat{G}_{p_1, p_2}(x, x') \hat{\phi}_k(x) dx = \hat{\lambda}_k \hat{\phi}_k(x'), \quad (3.10)$$

where the $\hat{\phi}_k$ are subject to $\int_0^1 \hat{\phi}_k^2(t) dt = \hat{\lambda}_k$ and $\int_0^1 \hat{\phi}_k(t) \hat{\phi}_{k'}(t) dt = 0$ for $k' < k$. Since N is sufficiently large, (3.10) can be approximated by

$$N^{-1} \sum_{j=1}^N \hat{G}_{p_1, p_2}(j/N, j'/N) \hat{\phi}_k(j/N) = \hat{\lambda}_k \hat{\phi}_k(j'/N).$$

For the same reason, the estimation of ξ_{ik} has the form of

$$\hat{\xi}_{ik} = N^{-1} \sum_{j=1}^N \hat{\lambda}_k^{-1} \left(Y_{ij} - \hat{m}_{p_1}(j/N) \right) \hat{\phi}_k(j/N).$$

To choose the number of principal components, κ , Müller (2009) described two methods. The first method is the “pseudo-AIC” criterion proposed in Yao et al. (2005a). The second is a simple “fraction of variation explained” method, i.e. select the number of eigenvalues that can explain, say, 95% of the variation in the data. From our experience in the numerical

studies, the simple “fraction of variation explained” method often works well.

Finally, to evaluate $Q_{1-\alpha}$, we need to simulate the Gaussian random field $\zeta_Z(x, x')$ in (3.7). The definition of $\zeta_Z(x, x')$ involves $\phi_k(x)$ and $V(x, x')$, which are replaced by their estimators described above. The fourth moment of ξ_{1k} is replaced by the empirical moments of $\hat{\xi}_{ik}$. We simulate a large number of independent realizations of $\zeta_Z(x, x')$, and take the maximal absolute deviation for each copy of $\zeta_Z(x, x')$. Then $Q_{1-\alpha}$ is estimated by the empirical percentiles of these maximum values.

For the two-sample hypothesis testing problem, we will center the two groups of curves separately by their own mean functions, since we do allow each group to have a different mean function. Analogous to $\hat{Q}_{1-\alpha}$, we estimate $Q_{12,1-\alpha}$ and further

$$\hat{G}_{p_1, p_2}^{(1)}(x, x') - \hat{G}_{p_1, p_2}^{(2)}(x, x') \pm n_1^{-1/2} \hat{Q}_{12,1-\alpha} \left(\hat{V}_1(x, x') + \hat{r} \hat{V}_2(x, x') \right)^{1/2}, \quad \forall \alpha \in (0, 1). \quad (3.11)$$

is applied for two samples covariance functions in the practice. The rest of the procedure follows as described in Section 3.3.3.

3.5 Simulation

To illustrate the finite-sample performance of the spline approach, we generate data from the model

$$Y_{ij} = m(j/N) + \sum_{k=1}^2 \xi_{ik} \phi_k(j/N) + \sigma \varepsilon_{ij}, \quad 1 \leq j \leq N, 1 \leq i \leq n,$$

where $\xi_{ik} \sim N(0, 1)$, $k = 1, 2$, $\varepsilon_{ij} \sim N(0, 1)$, for $1 \leq i \leq n$, $1 \leq j \leq N$, $m(x) = 10 + \sin\{2\pi(x - 1/2)\}$, $\phi_1(x) = -2 \cos\{\pi(x - 1/2)\}$ and $\phi_2(x) = \sin\{\pi(x - 1/2)\}$. This setting implies $\lambda_1 = 2$ and $\lambda_2 = 0.5$. The noise levels are set to be $\sigma = 0.5$ and 1.0 . The number of subjects n is taken to be 50, 100, 200, 300 and 500, and under each sample size the number of observations per curve is assumed to be $N = 4[n^{0.3} \log(n)]$. This simulated process has a similar design to one of the simulation models in Yao et al. (2005a), except that each subject is densely observed. We consider both linear and cubic spline estimators, and use confidence levels $1 - \alpha = 0.95$ and 0.99 for our simultaneous confidence envelopes. Each simulation is repeated 500 times.

Table 3.1 shows the empirical frequency that the true surface $G(x, x')$ is entirely covered by the confidence envelopes. At both noise levels, one observes that, as sample size increases, the true coverage probability of the confidence envelopes becomes closer to the nominal confidence level, which shows a positive confirmation of Theorem 3.3.2.

We present two estimation schemes: a) both mean and covariance functions are estimated by linear splines, i.e., $p_1 = p_2 = 2$; b) both are estimated by cubic splines, i.e. $p_1 = p_2 = 4$. Since the true covariance function is smooth in our simulation, the cubic spline estimator provides better estimate of the covariance function. However, as can be seen from Table 3.1, the two spline estimators behave rather similarly in terms of coverage probability. We also did simulation studies for the cases $p_1 = 4$, $p_2 = 2$ and $p_1 = 2$, $p_2 = 4$, the coverage rates are not shown here because they are similar to the cases presented in Table 3.1.

We show in Figures 3.1 and 3.2 the spline covariance estimator and the 95% confidence envelopes for $n = 200$ and $\sigma = 0.5$. Figures 3.1 and 3.2 correspond to linear ($p_1 = p_2 = 2$) and cubic ($p_1 = p_2 = 4$) spline estimators respectively. In each plot, the true covariance function is overlayed by the two confidence envelopes.

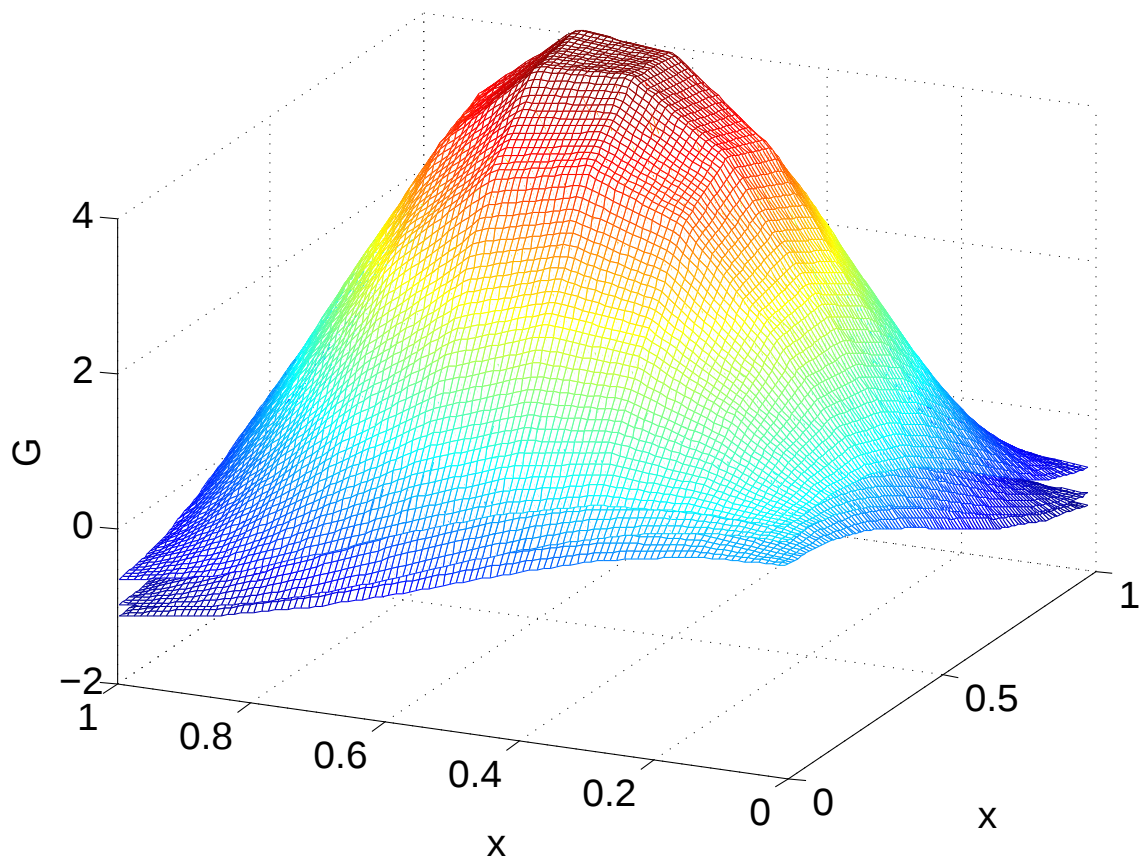


Figure 3.1: Plots of the true covariance functions (middle surfaces) of the simulated data and their 95% confidence envelopes (3.11) (upper and lower surfaces): $n=200$, $N=100$, $\sigma=0.5$. $p_1 = p_2 = 2$.

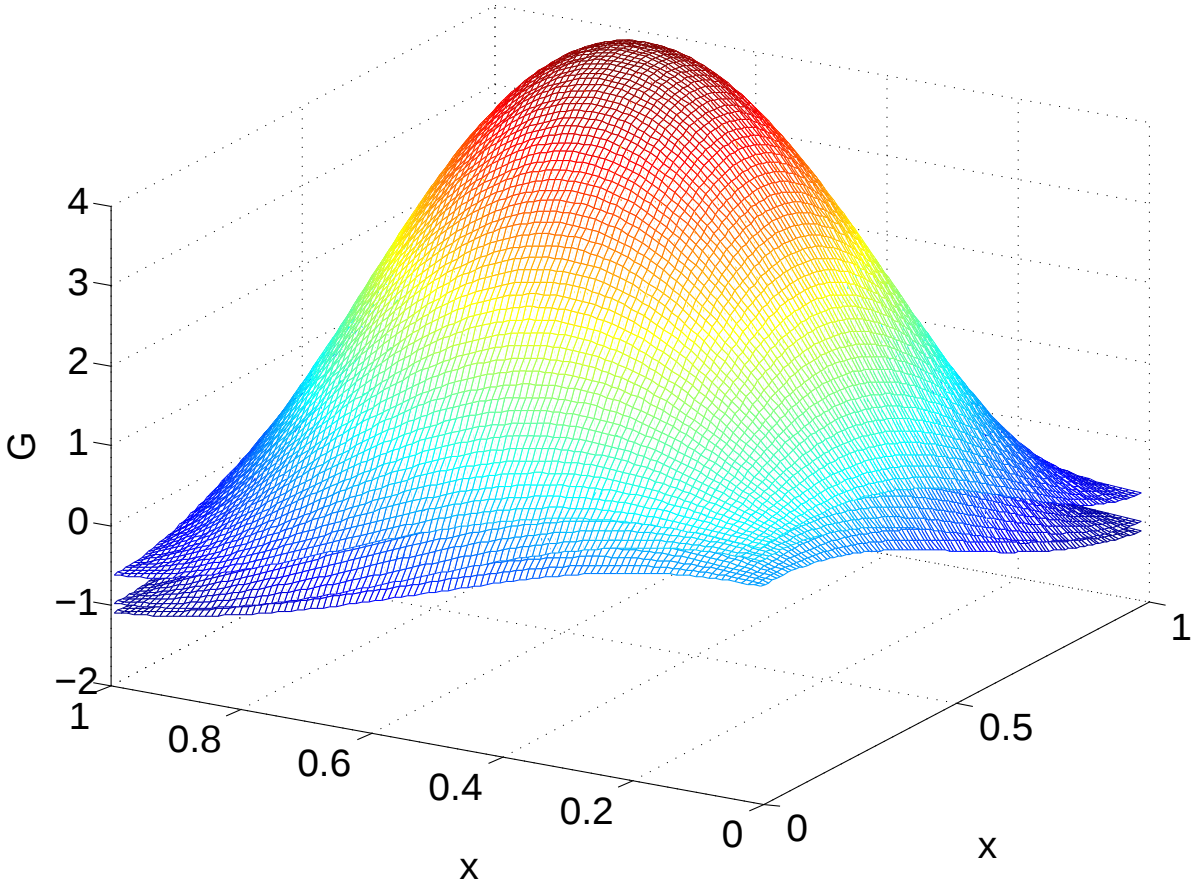


Figure 3.2: Plots of the true covariance functions (middle surfaces) of the simulated data and their 95% confidence envelopes (3.11) (upper and lower surfaces): $n=200$, $N=100$, $\sigma=0.5$. $p_1 = p_2 = 4$.

3.6 Empirical examples

3.6.1 Tecator near infrared spectra data

We first apply our methodology to the Tecator data mentioned in Section 1 and chapter 2.6. Figure 3.3 shows the scatter plot of the spectra. As we can see, the spectra can be naturally considered as functional data, since they are recorded on a dense grid of points with little measurement error. On the other hand, there is a lot of variation among different curves. We show the estimated covariance function and the 95% confidence envelope in Figures 3.4. These results are obtained by applying cubic spline smoothing to both the mean and covariance functions, with the number of knots $N_{s1} = 10$, $N_{s2} = 6$, respectively. We also tried other combinations of knots numbers and linear spline estimators. The results are very similar, and hence are not shown here. From Figure 3.4, we can see that the within curve covariance is positive and quite significant, since the zero hyperplane is far below the lower bound of the confidence envelope.

Using the simultaneous confidence envelopes, one can test other interesting hypotheses on the true covariance function, such as the true covariance being stationary. Specifically, we are interested in the following hypothesis,

$$\begin{aligned} H_0 : G(x, x') &\equiv g(|x - x'|), \forall (x, x') \in [a, b]^2 \\ v.s. H_a : G(x, x') &\neq g(|x - x'|), \exists (x, x') \in [a, b]^2, \end{aligned} \quad (3.12)$$

where $g(\cdot)$ is a stationary covariance function, and $[a, b]$ is the range of wavelength.

To test the hypothesis in (3.12), we need to generate a new estimator under the stationarity assumption and check if this estimator can be covered by the simultaneous confidence

envelope. Letting $\hat{G}(x, x')$ be the tensor product B-spline covariance estimator, we define $\hat{G}_S(u) = (b - a - u)^{-1} \int_a^{b-u} \hat{G}(x, x + u) dx$ for $0 \leq u \leq b - a$ and $\hat{G}_S(u) = \hat{G}_S(-u)$ for $a - b < u < 0$. Similar to $\hat{G}$, the new estimator $\hat{G}_S$ is not guaranteed to be positive semi-definite, but it is sufficient for our purpose. Under the stationarity assumption, $\hat{G}_S$ is a better estimator of the true covariance. We will pretend that $\hat{G}_S$ is the true covariance and reject the null hypothesis if this function is not covered by the confidence envelope.

Figure 3.5 shows cubic tensor spline envelopes with 0.9995 confidence level, and the center surface is $\hat{G}_S(x - x')$ as a two-dimensional function. As we can see, even for such a high confidence level, the estimator under the stationarity assumption is still not fully covered in the envelopes. We conclude that the covariance structure in these Tecator spectra is non-stationary. The same conclusion can be drawn using linear tensor spline method.

3.6.2 Speech recognition data

The data were extracted from the TIMIT database (TIMIT Acoustic-Phonetic Continuous Speech Corpus, NTIS, US Dept of Commerce) which is a widely used resource for research in speech recognition. The data set we use was formed by selecting five phonemes for classification based on digitized speech from this database. From continuous speech of 50 male speakers, 4509 speech frames of 32 msec duration were selected. From each speech frame, a log-periodogram was used as transformation for casting speech data in a form suitable for speech recognition. The five phonemes in this data set are transcribed as follows: “sh” as in “she”, “dcl” as in “dark”, “iy” as the vowel in “she”, “aa” as the vowel in “dark”, and “ao” as the first vowel in “water”. For illustration purpose, we focus on the “sh” and “ao” phoneme classes as representatives of consonants and vowels. There are $n_1 = 872$

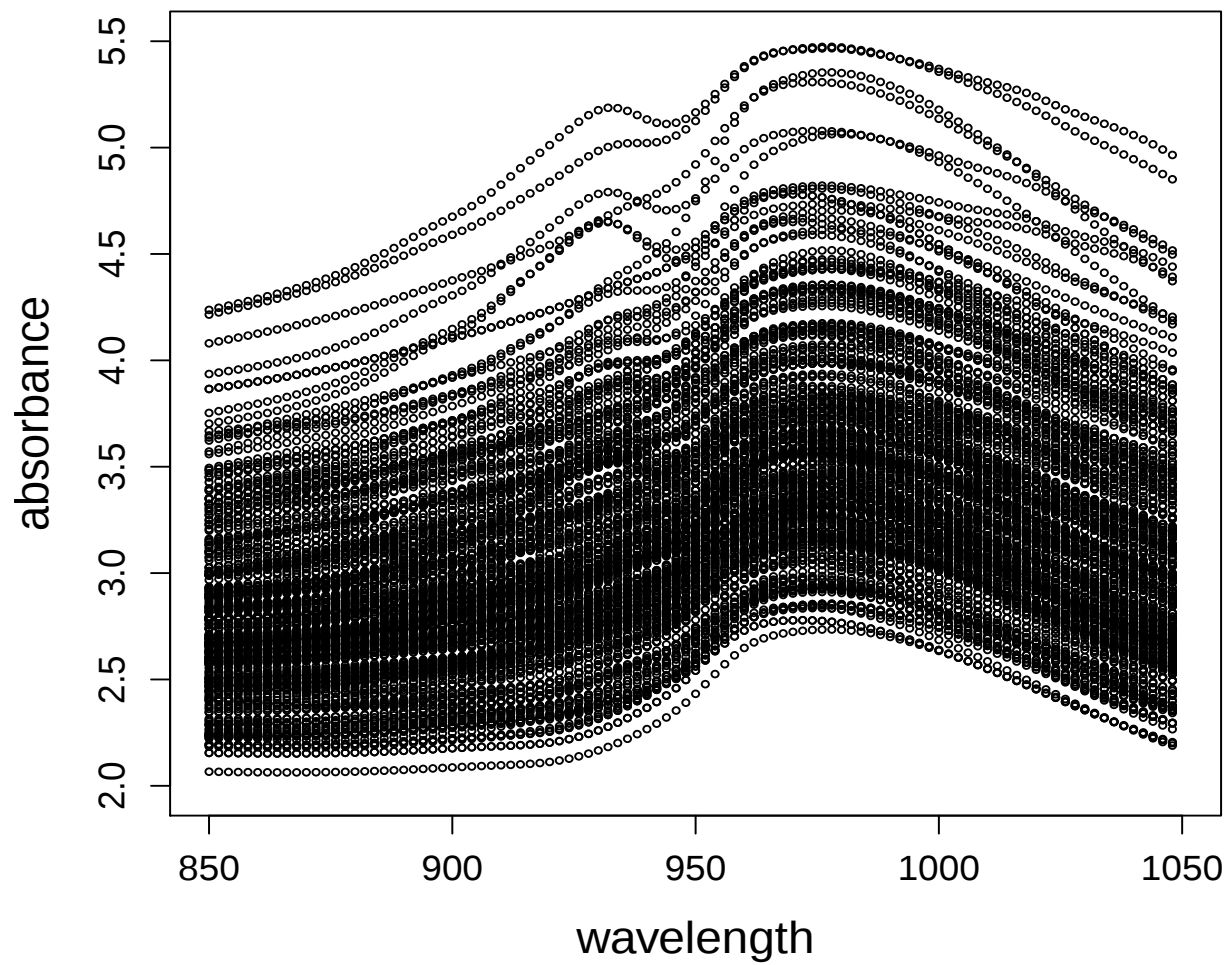


Figure 3.3: Plot of the Tecator data.

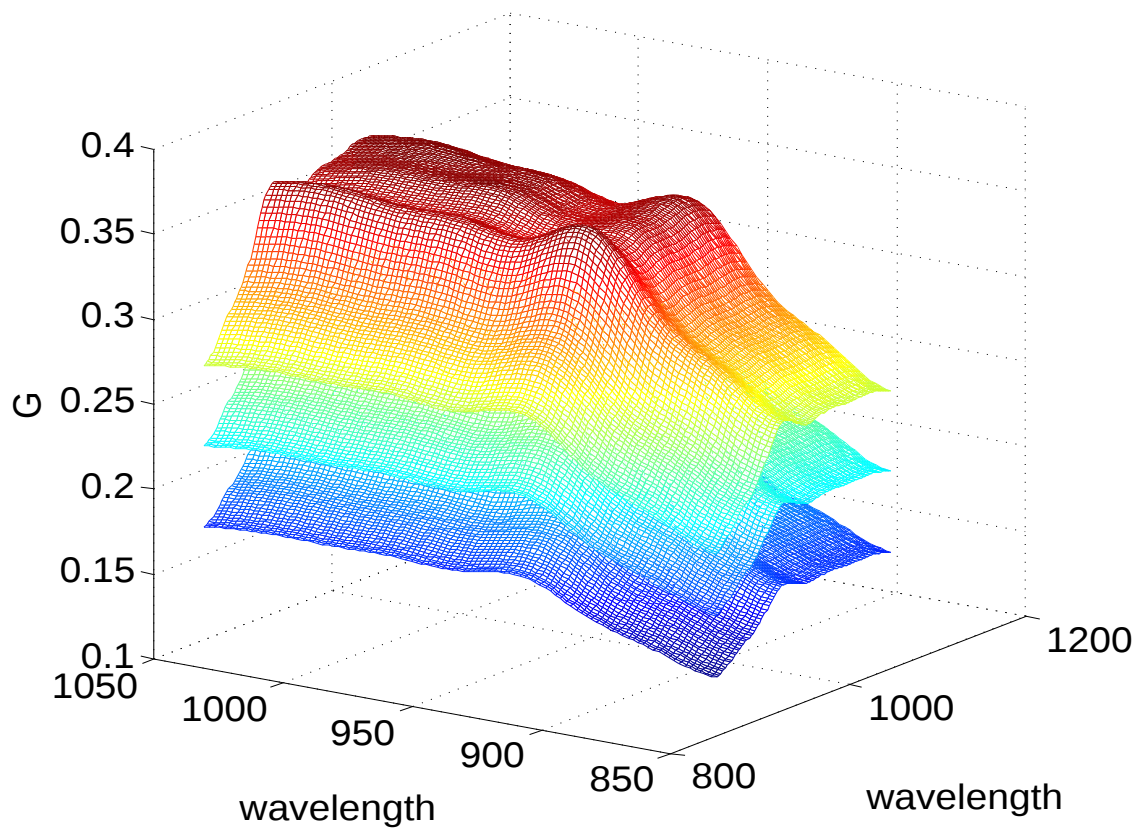


Figure 3.4: Plots of the cubic tensor product spline covariance estimator (3.3) for the Tecator data (middle surface) and the 95% simultaneous confidence envelope (3.11) (upper and lower surfaces).

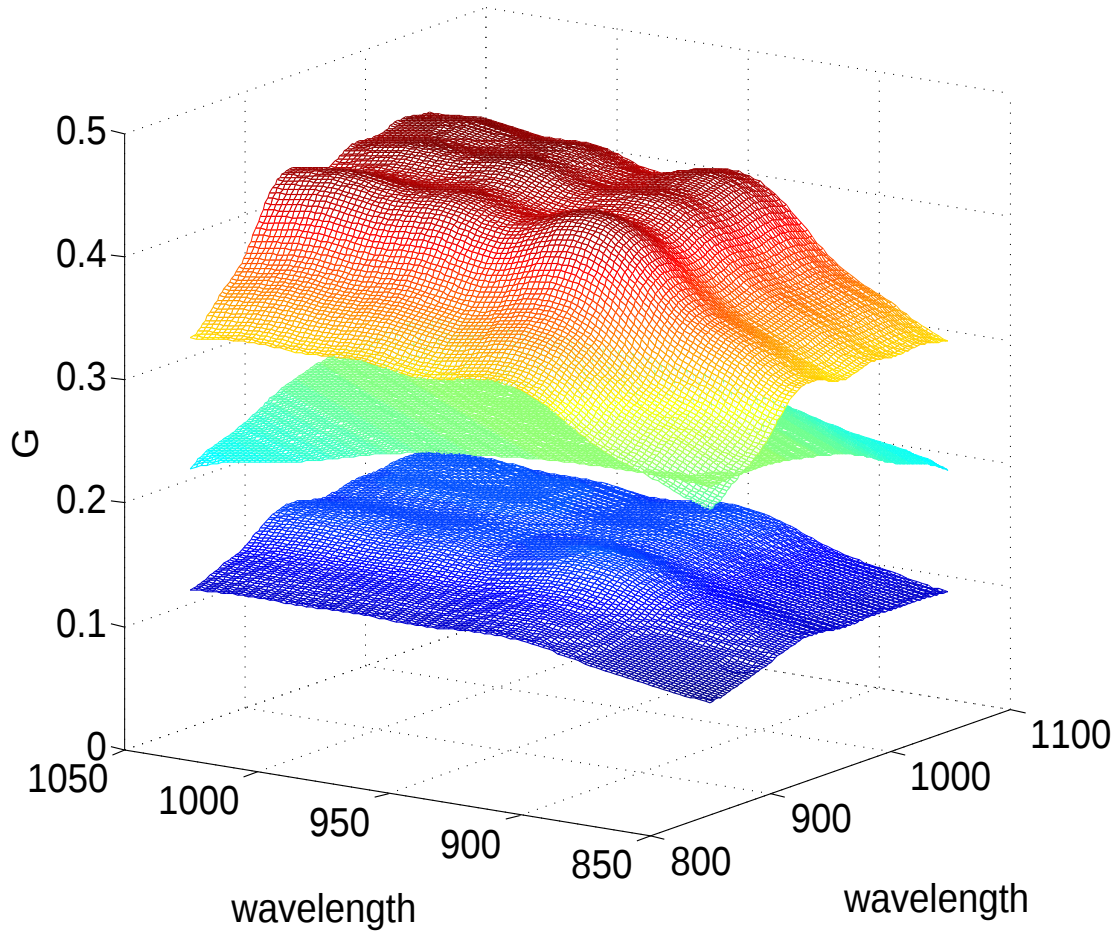


Figure 3.5: Plot for testing hypothesis (3.12) for the Tecator data. The upper and lower surfaces are the 99.95% confidence envelopes for the covariance function, and the middle surface is the covariance estimator under stationarity assumption, $\hat{G}_S(x - x')$.

log-periodograms in the “sh” class, and $n_2 = 1022$ log-periodograms in the “ao” group. Each log-periodogram consists $N = 256$ equally spaced points. Figure 3.6 shows a sample 10 log-periodograms from each of the two phoneme classes.

This data set was first analyzed by Hastie et al. (1995) using penalized linear discriminant analysis. One of the basic assumptions is that the covariance functions are the same for different classes. Judging from the scatter plot of the data in Figure 3.6, despite the clear difference between the mean functions of the two groups, there is no obvious indication of difference in covariance structures.

We first obtain the cubic tensor product spline covariance estimators for the two phoneme classes separately, which are shown in Figure 3.7. These results are obtained by using $N_{s1} = 10$, $N_{s2} = 4$ number of knots for the “sh” class, and $N_{s1} = 11$ and $N_{s2} = 5$ for the “ao” class. Different number of knots between the two groups reflects that the sample sizes are different.

By comparing the covariance estimators in Figure 3.7, there seems to be a visible difference between the two classes. We now would like to test the equal covariance assumption formally. The hypotheses of interest are

$$\begin{aligned} H_0 : G^{(1)}(x, x') &\equiv G^{(2)}(x, x'), \forall (x, x') \in [0, 1]^2 \\ v.s. \quad H_a : G^{(1)}(x, x') &\neq G^{(2)}(x, x'), \exists (x, x') \in [0, 1]^2. \end{aligned} \quad (3.13)$$

The 99.95% confidence envelopes for the difference of the two covariance functions are provided in Figure 3.8, and the zero hyperplane is used as a reference. Since the zero hyperplane is not covered by the envelopes, the equal covariance hypothesis is rejected with p-value < 0.0005 . We also tried different numbers of knots and the test result is not sensitive to this

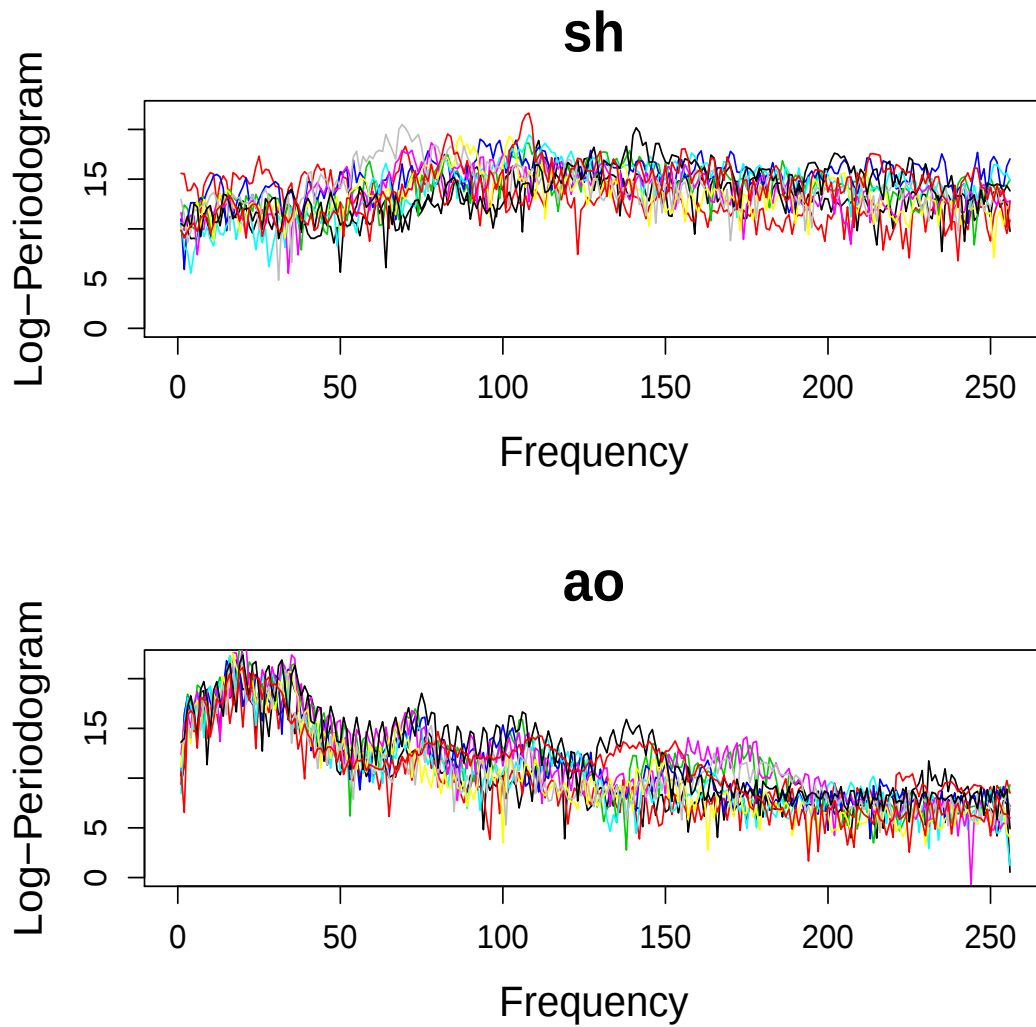


Figure 3.6: Plots of the speech recognition data.

choice. Our results suggest that a quadratic discriminant analysis, that takes into account the difference in the within-group covariance functions, might yield a better classification error rate than a linear discriminant analysis in this data set.

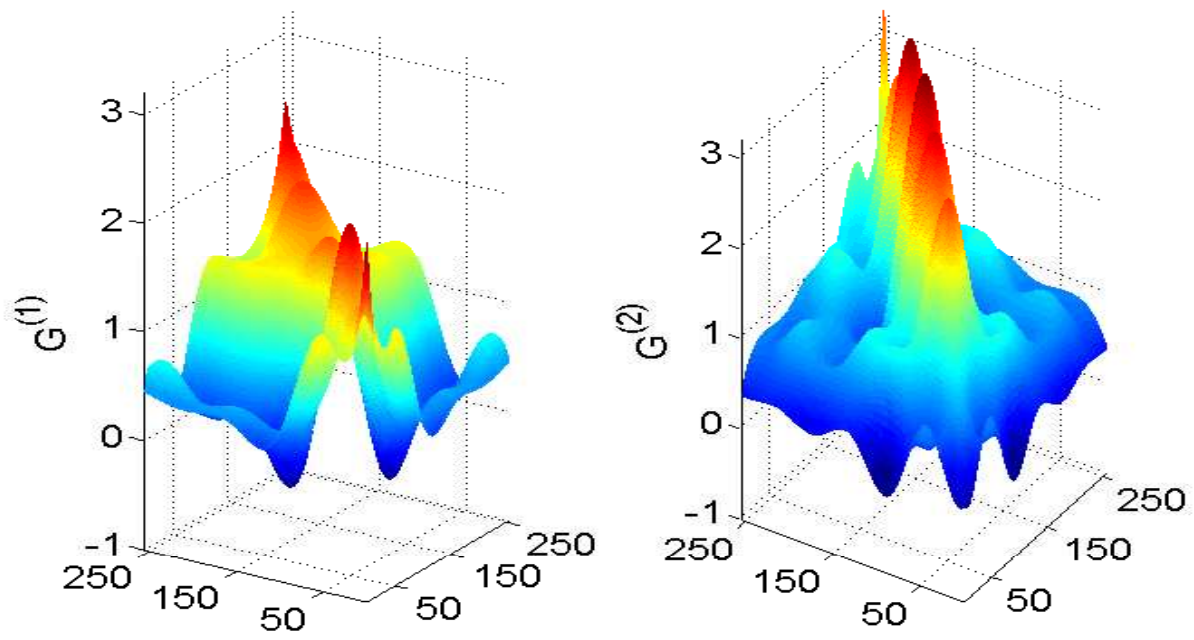


Figure 3.7: Plots of tensor spline estimators for “sh” and “ao” data sets. Right: “sh”; Left: “ao”.

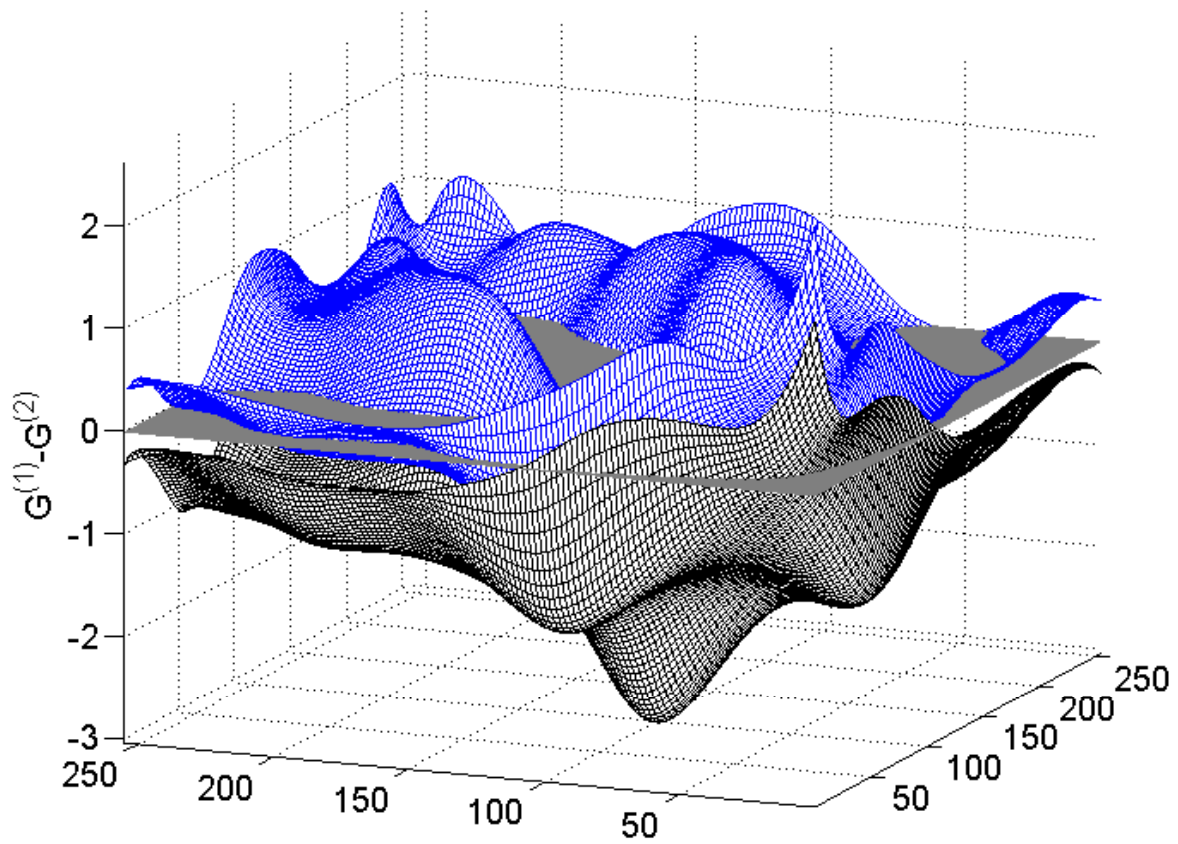


Figure 3.8: Plots of hypothesis test (3.13) results with the cubic tensor product spline 99.95% confidence envelopes (3.11) (upper and lower surfaces) for the speech recognition data and the zero surfaces (middle flat surface).

3.7 Summary

In this chapter, we consider covariance estimation in functional data and propose a new computationally efficient tensor-product B-spline estimator. The proposed estimator can be used as a building block for further data analysis, such as principal component analysis, linear discriminant analysis and analysis of variance. We study both local and global asymptotic properties of our estimator and proposed a simultaneous envelope approach to make inference on the true covariance function. The method is applied to a Tecator near-infrared spectra data to test the stationarity assumption on the covariance. In a classification problem, we further extend our method to a two-sample problem to test the equal covariance assumption between different treatment groups.

APPENDIX

Throughout this section, C means some positive constant in this whole section.

The detailed proofs of the following lemmas, Proposition 3.3.1 and Proposition 3.3.2 can be found in the supplement.

Preliminaries

For any positive integer p , denote the theoretical and empirical inner product matrices of $\left\{B_{J,p}(x)\right\}_{J=1-p}^{N_S}$ as

$$\mathbf{V}_p = \left(\left\langle B_{J,p}, B_{J',p} \right\rangle\right)_{J,J'=1-p}^{N_S}, \quad \hat{\mathbf{V}}_p = \left(\left\langle B_{J,p}, B_{J',p} \right\rangle_{2,N}\right)_{J,J'=1-p}^{N_S}.$$

The following lemma is from Chapter 1, which established the upper bound of $\left\|\mathbf{V}_p^{-1}\right\|_{\infty}$.

Lemma 3.7.1. *For any positive integer p , there exists a constant $M_p > 0$ depending only on p , such that $\left\|\mathbf{V}_p^{-1}\right\|_{\infty} \leq M_p h_S^{-1}$, for a large enough n , where $h_S = (N_S + 1)^{-1}$.*

Denote by “ $\otimes$ ” the Kronecker product of two matrices. Note that $\left\|(\mathbf{A} \otimes \mathbf{A})^{-1}\right\|_{\infty} = \left\|\mathbf{A}^{-1} \otimes \mathbf{A}^{-1}\right\|_{\infty} \leq \left\|\mathbf{A}^{-1}\right\|_{\infty}^2$, for any invertible matrix $\mathbf{A}$, which, together with Lemma 3.7.1, leads to the following result.

Lemma 3.7.2. *For any positive integer p , there exists a constant $M_p > 0$ depending only on p , such that $\left\|(\mathbf{V}_p \otimes \mathbf{V}_p)^{-1}\right\|_{\infty} \leq M_p^2 h_S^{-2}$.*

Next we define the theoretical and empirical inner product matrices of tensor product

spline basis $\left\{ B_{JJ',p_2}(x, x') \right\}_{J,J'=1-p_2}^{N_{s2}}$ as

$$\begin{aligned} \mathbf{V}_{p_2,2} &= \left(\left\langle B_{JJ',p_2}, B_{J''J''',p_2} \right\rangle \right)_{J,J',J'',J'''=1-p_2}^{N_{s2}} \\ \hat{\mathbf{V}}_{p_2,2} &= \left(\left\langle B_{JJ',p_2}, B_{J''J''',p_2} \right\rangle_{2,N} \right)_{J,J',J'',J'''=1-p_2}^{N_{s2}}. \end{aligned} \quad (3.14)$$

The following results show that the difference of $\mathbf{V}_{p_2,2}$ and $\hat{\mathbf{V}}_{p_2,2}$ is negligible. Using the results from Lemma 3.7.2, we can obtain the upper bound of the norm of $\hat{\mathbf{V}}_{p_2,2}^{-1}$.

Lemma 3.7.3. *Under Assumption (A3), for $\mathbf{V}_{p_2,2}$ and $\hat{\mathbf{V}}_{p_2,2}$ defined in (3.14),*

$$\left\| \mathbf{V}_{p_2,2} - \hat{\mathbf{V}}_{p_2,2} \right\|_{\infty} = O\left(N^{-1}\right) \text{ and } \left\| \hat{\mathbf{V}}_{p_2,2}^{-1} \right\|_{\infty} = O\left(h_{s2}^{-2}\right).$$

Lemma 3.7.4. *For $\hat{\mathbf{V}}_{p_2,2}$ defined in (3.14) and any $N(N-1)$ vector $\boldsymbol{\rho} = (\rho_{jj'})$, there exists a constant $C > 0$, such that*

$$\sup_{(x,x') \in [0,1]^2} \left\| N^{-2} \mathbf{B}_{p_2}^{\top}(x, x') \hat{\mathbf{V}}_{p_2,2}^{-1} \mathbf{X}^T \boldsymbol{\rho} \right\|_{\infty} \leq C \|\boldsymbol{\rho}\|_{\infty}.$$

Denote $\tilde{\phi}_{kk'}(x, x') = \mathbf{B}_{p_2}^{\top}(x, x') (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \phi_{kk'}$, and

$$\begin{aligned} \phi_{kk'} &= \\ &\left(\phi_k(2/N) \phi_{k'}(1/N), \dots, \phi_k(1) \phi_{k'}(1/N), \dots, \phi_k(1/N) \phi_{k'}(1), \dots, \phi_k(1-1/N) \phi_{k'}(1) \right)^T. \end{aligned}$$

The following lemma is a direct result from de Boor (2001), p. 149 and Theorem 5.1 of Huang (2003), thus the proof is omitted.

Lemma 3.7.5. *There is an absolute constant $C_g > 0$ such that for every $g \in C^{p-1,\mu}[0,1]$,*

there exists a function $g^ \in \mathcal{H}^{(p-1)}[0,1]$ and some $\mu \in (0,1]$ such that $\sup_{x \in [0,1]} |g(x) - g^*(x)| \leq C_g h_s^{p-1+\mu}$. If Assumption (B2) holds, $\sup_{(x,x') \in [0,1]^2} \left| \phi_{kk'}(x, x') - \tilde{\phi}_{kk'}(x, x') \right|_{\infty} =$*

$$O\left(h_{s_2}^{p_2}\right).$$

Proofs of theorems 3.3.1 and 3.3.2

PROOF OF THEOREM 3.3.1. By Propositions 3.3.1,

$$E[\tilde{G}_{p_1, p_2}(x, x') - G(x, x')]^2 = E\Delta^2(x, x') + o(1),$$

where $\Delta(x, x')$ is defined in (3.5). Let $\bar{\xi}_{\cdot k k'} = n^{-1} \sum_{i=1}^n \xi_{ik} \xi_{ik'}$, $1 \leq k, k' \leq \kappa$. According to (3.16) and (3.17), one has

$$\Delta(x, x') = \sum_{k \neq k'}^{\kappa} \bar{\xi}_{\cdot k k'} \phi_k(x) \phi_{k'}(x') + \sum_{k=1}^{\kappa} (\bar{\xi}_{\cdot k k} - 1) \phi_k(x) \phi_k(x').$$

Since

$$\begin{aligned} & nE \left[\Delta(x, x') \right]^2 \\ &= \sum_{k, k'=1}^{\kappa} \phi_k^2(x) \phi_{k'}^2(x') + \sum_{k, k'=1}^{\kappa} \phi_k(x') \phi_k(x) \phi_{k'}(x) \phi_{k'}(x') \\ &+ \sum_{k=1}^{\kappa} \phi_k^2(x) \phi_k^2(x') (E\xi_{1k}^4 - 3) \\ &= G(x, x')^2 + G(x, x) G(x', x') + \sum_{k=1}^{\kappa} \phi_k^2(x) \phi_k^2(x') (E\xi_{1k}^4 - 3) \equiv V(x, x'), \end{aligned}$$

and the desired result follows from Proposition 3.3.2.

Next define $\zeta(x, x') = n^{1/2} V^{-1/2}(x, x') \Delta(x, x')$.

Lemma 3.7.6. *Under Assumptions (B2)-(B4), $\sup_{(x, x') \in [0, 1]^2} \left| \zeta_Z(x, x') - \zeta(x, x') \right| =$*

$o.a.s. (1)$, where $\zeta_Z(x, x')$ is given in (3.7).

PROOF OF THEOREM 3.3.2. According to Lemma 3.7.6, Propositions 3.3.1 and 3.3.2 and Theorem 3.3.1, for $\forall \alpha \in (0, 1)$, as $n \rightarrow \infty$,

$$\begin{aligned}
& \lim_{n \rightarrow \infty} P \left\{ \sup_{(x, x') \in [0, 1]^2} n^{1/2} \left| \hat{G}_{p_1, p_2}(x, x') - G(x, x') \right| V(x, x')^{-1/2} \leq Q_{1-\alpha} \right\} \\
&= \lim_{n \rightarrow \infty} P \left\{ \sup_{(x, x') \in [0, 1]^2} n^{1/2} \left| \tilde{G}_{p_1, p_2}(x, x') - G(x, x') \right| V(x, x')^{-1/2} \leq Q_{1-\alpha} \right\} \\
&= \lim_{n \rightarrow \infty} P \left\{ \sup_{(x, x') \in [0, 1]^2} \left| \zeta(x, x') \right| \leq Q_{1-\alpha} \right\} \\
&= \lim_{n \rightarrow \infty} P \left\{ \sup_{(x, x') \in [0, 1]^2} \left| \zeta_Z(x, x') \right| \leq Q_{1-\alpha} \right\}.
\end{aligned}$$

Supplement

This supplement contains proofs for Lemmas 3.7.3, 3.7.4 and 3.7.6, Propositions 3.3.1 and 3.3.2.

PROOF OF LEMMA 3.7.3. Note that

$$\begin{aligned}
& \hat{\mathbf{V}}_{p_2,2} \\
&= \left\{ N^{-2} \sum_{1 \leq j \neq j' \leq N} B_{JJ',p_2} \left(\frac{j}{N}, \frac{j'}{N} \right) B_{J''J''',p_2} \left(\frac{j}{N}, \frac{j'}{N} \right) \right\}_{J,J',J'',J'''=1-p_2}^{N_{s_2}} \\
&= \left\{ N^{-2} \left[\sum_{j=1}^N B_{JJ'',p_2} \left(\frac{j}{N}, \frac{j}{N} \right) \right] \left[\sum_{j=1}^N B_{J'J''',p_2} \left(\frac{j}{N}, \frac{j}{N} \right) \right] \right\}_{J,J',J'',J'''=1-p_2}^{N_{s_2}} \\
&\quad - \left\{ N^{-2} \sum_{j=1}^N B_{JJ',p_2} \left(\frac{j}{N}, \frac{j}{N} \right) B_{J''J''',p_2} \left(\frac{j}{N}, \frac{j}{N} \right) \right\}_{J,J',J'',J'''=1-p_2}^{N_{s_2}} \\
&= \hat{\mathbf{V}}_{p_2} \otimes \hat{\mathbf{V}}_{p_2} - \left\{ N^{-2} \sum_{j=1}^N B_{JJ',p_2} \left(\frac{j}{N}, \frac{j}{N} \right) B_{J''J''',p_2} \left(\frac{j}{N}, \frac{j}{N} \right) \right\}_{J,J',J'',J'''=1-p_2}^{N_{s_2}}.
\end{aligned}$$

Note that the entries in the matrix $\hat{\mathbf{V}}_{p_2,2} - \hat{\mathbf{V}}_{p_2} \otimes \hat{\mathbf{V}}_{p_2}$ are zero when the maximum absolute difference between any two of the indices (J, J'') , (J', J''') is greater than p ; otherwise

$$\begin{aligned}
& N^{-2} \sum_{j=1}^N B_{JJ',p_2} \left(\frac{j}{N}, \frac{j}{N} \right) B_{J''J''',p_2} \left(\frac{j}{N}, \frac{j}{N} \right) \\
&= N^{-1} \left[\int_0^1 B_{JJ',p_2} (x, x) B_{J''J''',p_2} (x, x) dx + O(N^{-1} h_{s_2}^{-1}) \right] \\
&= O \left(N^{-1} h_{s_2} + N^{-2} h_{s_2}^{-1} \right).
\end{aligned}$$

Hence, $\left\| \hat{\mathbf{V}}_{p_2,2} - \hat{\mathbf{V}}_{p_2} \otimes \hat{\mathbf{V}}_{p_2} \right\|_{\infty} = O\left(N^{-1}h_{s_2} + N^{-2}h_{s_2}^{-1}\right)$. Since

$$\begin{aligned}
& \left\| \mathbf{V}_{p_2} \otimes \mathbf{V}_{p_2} - \hat{\mathbf{V}}_{p_2} \otimes \hat{\mathbf{V}}_{p_2} \right\|_{\infty} \\
&= \max_{1-p_2 \leq J', J'' \leq N_{s_2}} \sum_{J, J''=1-p_2}^{N_{s_2}} \left| N^{-2} \sum_{j, j'=1}^N B_{JJ', p_2} \left(\frac{j}{N}, \frac{j'}{N} \right) B_{J'' J''', p_2} \left(\frac{j}{N}, \frac{j'}{N} \right) \right. \\
&\quad \left. - \int_0^1 \int_0^1 B_{JJ', p_2} (x, x') B_{J'' J''', p_2} (x, x') dx dx' \right| \\
&\leq \max_{1-p_2 \leq J', J'' \leq N_{s_2}} \sum_{J, J''=1-p_2}^{N_{s_2}} \sum_{j, j'=1}^N \left| \int_{(j'-1)/N}^{j'/N} \int_{(j-1)/N}^{j/N} B_{JJ', p_2} (j/N, j'/N) \right. \\
&\quad \left. \times B_{J'' J''', p_2} (j/N, j'/N) - B_{JJ', p_2} (x, x') B_{J'' J''', p_2} (x, x') \right| dx dx' \\
&\leq C h_{s_2}^{-2} (N h_{s_2})^2 \times N^{-2} \times N^{-2} h_{s_2}^{-2} = C N^{-2} h_{s_2}^{-2},
\end{aligned}$$

applying Assumption (B3) one has $\left\| \mathbf{V}_{p_2} \otimes \mathbf{V}_{p_2} - \hat{\mathbf{V}}_{p_2,2} \right\|_{\infty} = o\left(N^{-1}\right)$.

According to Lemma 3.7.2, for any $(N_{s_2} + p_2)^2$ vector $\boldsymbol{\tau}$, one has $\left\| (\mathbf{V}_{p_2} \otimes \mathbf{V}_{p_2})^{-1} \boldsymbol{\tau} \right\|_{\infty} \leq h_{s_2}^{-2} \|\boldsymbol{\tau}\|_{\infty}$. Hence, $\left\| (\mathbf{V}_{p_2} \otimes \mathbf{V}_{p_2}) \boldsymbol{\tau} \right\|_{\infty} \geq h_{s_2}^2 \|\boldsymbol{\tau}\|_{\infty}$. Note that

$$\left\| \hat{\mathbf{V}}_{p_2,2} \boldsymbol{\tau} \right\|_{\infty} \geq \left\| (\mathbf{V}_{p_2} \otimes \mathbf{V}_{p_2}) \boldsymbol{\tau} \right\|_{\infty} - \left\| (\mathbf{V}_{p_2} \otimes \mathbf{V}_{p_2}) \boldsymbol{\tau} - \hat{\mathbf{V}}_{p_2,2} \boldsymbol{\tau} \right\|_{\infty} = O\left(h_{s_2}^2\right) \|\boldsymbol{\tau}\|_{\infty}.$$

If $\boldsymbol{\tau}$ satisfies that $\left\| \hat{\mathbf{V}}_{p_2,2}^{-1} \right\|_{\infty} = \left\| \hat{\mathbf{V}}_{p_2,2}^{-1} \boldsymbol{\tau} \right\|_{\infty} \leq O\left(h_{s_2}^{-2}\right) \|\boldsymbol{\tau}\|_{\infty} = O\left(h_{s_2}^{-2}\right)$, the lemma is proved.

PROOF OF LEMMA 3.7.4. Note that for any matrix $\mathbf{A} = (a_{ij})_{i=1, j=1}^{m,n}$ and any n by

1 vector $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_n)^T$, one has $\|\mathbf{A}\boldsymbol{\alpha}\|_\infty \leq \|\mathbf{A}\|_\infty \|\boldsymbol{\alpha}\|_\infty$. It is clear that

$$\begin{aligned} \|N^{-2}\mathbf{X}^T\boldsymbol{\rho}\|_\infty &= \left\| N^{-2} \left\{ \sum_{1 \leq j \neq j' \leq N} B_{JJ', p_2} \left(j/N, j'/N \right) \rho_{jj'} \right\}_{J, J'=1-p_2}^{N_{s_2}} \right\|_\infty \\ &\leq \|\boldsymbol{\rho}\|_\infty \max_{1-p_2 \leq J, J' \leq N_{s_2}} \left| N^{-2} \sum_{1 \leq j \neq j' \leq N} B_{JJ', p_2} \left(j/N, j'/N \right) \right| \\ &\leq h_{s_2}^2 \|\boldsymbol{\rho}\|_\infty. \end{aligned}$$

One also observe that

$$\left\| N^{-2} \mathbf{B}_{p_2}^\top(x, x') \hat{\mathbf{V}}_{p_2, 2}^{-1} \mathbf{X}^T \boldsymbol{\rho} \right\|_\infty \leq \left\| \mathbf{B}_{p_2}^\top(x, x') \right\|_\infty \left\| N^{-2} \mathbf{X}^T \boldsymbol{\rho} \right\|_\infty \left\| \hat{\mathbf{V}}_{p_2, 2}^{-1} \right\|_\infty,$$

which, together with the boundedness of spline functions and Lemma 3.7.3, leads to the desired result.

PROOF OF LEMMA 3.7.6. According to Assumption (B4) and multivariate Central Limit Theorem

$$\sqrt{n} \left\{ \bar{\xi}_{\cdot k k'}, \left(\bar{\xi}_{\cdot k}^2 - 1 \right) \left(E \xi_{1k}^4 - 1 \right)^{-1/2} \right\}_{1 \leq k < k' \leq \kappa} \rightarrow_d N \left(\mathbf{0}_{\kappa(\kappa+1)/2}, \mathbf{I}_{\kappa(\kappa+1)/2} \right).$$

Applying Skorohod's Theorem, there exist i.i.d. variables $Z_{kk'} = Z_{k'k} \sim N(0, 1)$, $Z_k \sim N(0, 1)$, $1 \leq k < k' \leq \kappa$, such that as $n \rightarrow \infty$,

$$\max_{1 \leq k < k' \leq \kappa} \left\{ \left| \sqrt{n} \bar{\xi}_{\cdot k k'} - Z_{kk'} \right|, \left| \sqrt{n} \left(\bar{\xi}_{\cdot k}^2 - 1 \right) - Z_k \left(E \xi_{1k}^4 - 1 \right)^{1/2} \right| \right\} = o_{\text{a.s.}}(1). \quad (3.15)$$

The desired result follows from (3.15).

Proof of Proposition 3.3.1

Recall that the error terms defined in Section 3.2 are $U_{ij} = Y_{ij} - m(j/N)$, $i = 1, \dots, n$, $j = 1, \dots, N$. Note that

$$\begin{aligned}\bar{U}_{\cdot jj'} &= n^{-1} \sum_{i=1}^n U_{ij} U_{ij'} \\ &= n^{-1} \sum_{i=1}^n \left(\sum_{k=1}^{\kappa} \xi_{ik} \phi_k(j/N) + \sigma(j/N) \varepsilon_{ij} \right) \left(\sum_{k=1}^{\kappa} \xi_{ik} \phi_k(j'/N) + \sigma(j'/N) \varepsilon_{ij'} \right) \\ &= \bar{U}_{1jj'} + \bar{U}_{2jj'} + \bar{U}_{3jj'} + \bar{U}_{4jj'},\end{aligned}$$

where

$$\begin{aligned}\bar{U}_{1jj'} &= \sum_{k \neq k'}^{\kappa} \bar{\xi}_{\cdot k k'} \phi_k(j/N) \phi_{k'}(j'/N), \\ \bar{U}_{2jj'} &= \sum_{k=1}^{\kappa} \bar{\xi}_{\cdot k k} \phi_k(j/N) \phi_k(j'/N), \\ \bar{U}_{3jj'} &= n^{-1} \sum_{i=1}^n \sigma(j/N) \sigma(j'/N) \varepsilon_{ij} \varepsilon_{ij'}, \\ \bar{U}_{4jj'} &= n^{-1} \sum_{i=1}^n \left\{ \sum_{k=1}^{\kappa} \xi_{ik} \phi_k(j/N) \sigma(j'/N) \varepsilon_{ij'} + \sum_{k=1}^{\kappa} \xi_{ik} \phi_k(j'/N) \sigma(j/N) \varepsilon_{ij} \right\}.\end{aligned}$$

Let $\tilde{\mathcal{U}}_{ip_2}(x, x') = \mathbf{B}_{p_2}^T(x, x') (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \bar{\mathbf{U}}_i$, where $\bar{\mathbf{U}}_i = \{\bar{U}_{ijj'}\}_{1 \leq j \neq j' \leq N}$, for $i = 1, 2, 3, 4$. Then we can have the following decomposition for $\tilde{G}_{p_2}(x, x')$

$$\tilde{G}_{p_2}(x, x') = \tilde{\mathcal{U}}_{1p_2}(x, x') + \tilde{\mathcal{U}}_{2p_2}(x, x') + \tilde{\mathcal{U}}_{3p_2}(x, x') + \tilde{\mathcal{U}}_{4p_2}(x, x').$$

Define

$$\mathcal{U}_1(x, x') = \sum_{k \neq k'}^{\kappa} \bar{\xi}_{\cdot k k'} \phi_k(x) \phi_{k'}(x'), \quad (3.16)$$

$$\mathcal{U}_2(x, x') = G(x, x') + \sum_{k=1}^{\kappa} \left\{ \phi_k(x) \phi_k(x') (\bar{\xi}_{\cdot k k} - 1) \right\}. \quad (3.17)$$

Next we illustrate the facts that $\tilde{\mathcal{U}}_{1p_2}(x, x')$ and $\tilde{\mathcal{U}}_{2p_2}(x, x')$ are the dominating terms in the above decomposition, which converge uniformly to $\mathcal{U}_1(x, x')$ and $\mathcal{U}_2(x, x')$ respectively, while $\tilde{\mathcal{U}}_{3p_2}(x, x')$ and $\tilde{\mathcal{U}}_{4p_2}(x, x')$ are negligible noise terms.

By the definition of $\bar{\mathbf{U}}_1$, one has that

$$\begin{aligned} \tilde{\mathcal{U}}_{1p_2}(x, x') &= N^{-2} \mathbf{B}_{p_2}^T(x, x') \hat{\mathbf{V}}_{p_2, 2}^{-1} \mathbf{X}^T \left\{ \sum_{k \neq k'}^{\kappa} \bar{\xi}_{\cdot k k'} \phi_k(j/N) \phi_{k'}(j'/N) \right\} \\ &= \sum_{k \neq k'}^{\kappa} \bar{\xi}_{\cdot k k'} \tilde{\phi}_{kk'}(x, x'). \end{aligned}$$

Lemma 3.7.5 and Assumption (B3) imply that

$$\begin{aligned} \sup_{x, x' \in [0, 1]^2} \left| (\tilde{\mathcal{U}}_{1p_2} - \mathcal{U}_1)(x, x') \right| &= \sup_{x, x' \in [0, 1]^2} \left| \left\{ \tilde{\mathcal{U}}_{1p_2} - \sum_{k \neq k'}^{\kappa} \bar{\xi}_{\cdot k k'} \tilde{\phi}_{kk'} \right\}(x, x') \right| \\ &\leq \kappa^2 \max_{1 \leq k \neq k' \leq \kappa} \left\{ \left| \bar{\xi}_{\cdot k k'} \right| \sup_{x, x' \in [0, 1]^2} \left| (\phi_{kk'} - \tilde{\phi}_{kk'})(x, x') \right| \right\} \\ &= O_p(h_{s_2}^{p_2} n^{-1/2}) = o_p(n^{-1/2}). \end{aligned}$$

Similarly,

$$\begin{aligned}
& \sup_{x, x' \in [0, 1]^2} \left| (\tilde{\mathcal{U}}_{2p_2} - \mathcal{U}_2) (x, x') \right| \\
& \leq \kappa \max_{1 \leq k \leq \kappa} \left\{ \left| \bar{\xi}_{\cdot k k} \right| \sup_{x, x' \in [0, 1]^2} \left| (\phi_{kk'} - \tilde{\phi}_{kk'}) (x, x') \right| \right\} \\
& = O_p \left(h_{s_2}^{p_2} \right) = o_p \left(n^{-1/2} \right).
\end{aligned}$$

Therefore, one has

$$\sup_{x, x' \in [0, 1]^2} \left| (\tilde{\mathcal{U}}_{1p_2} + \tilde{\mathcal{U}}_{2p_2} - \mathcal{U}_1 - \mathcal{U}_2) (x, x') \right| = o_p \left(n^{-1/2} \right).$$

Denote that

$$N^{-2} \mathbf{X}^T \bar{\mathbf{U}}_3 = \left\{ n^{-1} \sum_{i=1}^n A_{iJJ'} \right\}_{J, J'=1-p_2}^{N_{s_2}},$$

where

$$A_{iJJ'} = N^{-2} \sum_{1 \leq j \neq j' \leq N} B_{J, p_2} (j/N) \sigma (j/N) B_{J', p_2} (j'/N) \sigma (j'/N) \varepsilon_{ij} \varepsilon_{ij'}.$$

It is easy to see that $EA_{iJJ'} = 0$ and $EA_{iJJ'}^2 = O \left(h_{s_2}^2 N^{-2} \right)$. Using standard arguments in Wang and Yang (2009), one has

$$\left\| N^{-2} \mathbf{X}^T \bar{\mathbf{U}}_3 \right\|_{\infty} = o_{\text{a.s.}} \left\{ N^{-1} n^{-1/2} h_{s_2} \log^{1/2} (n) \right\}.$$

Therefore, according to the definition of $\tilde{U}_{3p_2}(x, x')$, one has

$$\begin{aligned}
& \sup_{x, x' \in [0, 1]^2} \left\| \mathbf{B}_{p_2}^T(x, x') \hat{\mathbf{V}}_{p_2, 2}^{-1} N^{-2} \mathbf{X}^T \bar{\mathbf{U}}_3 \right\|_{\infty} \\
& \leq C_{p_2} \sup_{x, x' \in [0, 1]^2} \left\| \mathbf{B}_{p_2}(x, x') \right\|_{\infty} \left\| \hat{\mathbf{V}}_{p_2, 2}^{-1} N^{-2} \mathbf{X}^T \bar{\mathbf{U}}_3 \right\|_{\infty} \\
& = o_{\text{a.s.}} \left\{ N^{-1} n^{-1/2} h_{s_2}^{-1} \log^{1/2} n \right\} = o_{\text{a.s.}} \left(n^{-1/2} \right).
\end{aligned}$$

Likewise, in order to get the upper bound of $\left| \tilde{U}_{4p_2}(x, x') \right|$, one has

$$\begin{aligned}
& \mathbf{X}^T \bar{\mathbf{U}}_4 \\
& = \left\{ \frac{2}{n} \sum_{i=1}^n \sum_{k=1}^{\kappa} \xi_{ik} \right. \\
& \quad \left. \sum_{1 \leq j \neq j' \leq N} \phi_k(j/N) \sigma(j'/N) B_{JJ', p_2} \left(\frac{j}{N}, \frac{j'}{N} \right) \varepsilon_{ij'} \right\}_{J, J'=1-p_2}^{N_{s_2}}.
\end{aligned}$$

Let

$$D_{iJJ'} = N^{-2} \sum_{k=1}^{\kappa} \left(\xi_{ik} \sum_{1 \leq j \neq j' \leq N} \phi_k(j/N) \sigma(j'/N) B_{JJ', p} \left(j/N, j'/N \right) \varepsilon_{ij'} \right),$$

then $ED_{iJJ'} = 0$,

$$\begin{aligned}
ED_{iJJ'}^2 &= N^{-4} \sum_{k=1}^{\kappa} \sum_{1 \leq j \neq j' \leq N} \phi_k^2(j/N) \sigma^2(j'/N) B_{JJ', p_2}^2(j/N, j'/N) E\xi_{ik}^2 E\varepsilon_{ij'}^2 \\
&\leq CN^{-4} \sum_{k=1}^{\kappa} \sum_{1 \leq j \neq j' \leq N} \phi_k^2(j/N) B_{J, p_2}^2(j/N) \sigma^2(j'/N) B_{J', p_2}^2(j'/N) \\
&= O(h_{s_2}^2 N^{-2}).
\end{aligned}$$

Similar arguments in Wang and Yang (2009) leads to

$$\left\| \frac{\mathbf{X}^T \bar{\mathbf{U}}_4}{N^2} \right\|_{\infty} = o_{\text{a.s.}} \left\{ N^{-1} n^{-1/2} h_{s_2} \log^{1/2}(n) \right\}.$$

Thus,

$$\left\| \mathbf{B}_{p_2}^T(x, x') \hat{\mathbf{V}}_{p_2, 2}^{-1} \frac{\mathbf{X}^T \bar{\mathbf{U}}_4}{N^2} \right\|_{\infty} = o_{\text{a.s.}} \left\{ N^{-1} n^{-1/2} h_{s_2}^{-1} \log^{1/2}(n) \right\} = o_{\text{a.s.}} \left(n^{-1/2} \right).$$

Proof of Proposition 3.3.2

For simplicity, denote

$$\mathbf{X}_1 = \begin{pmatrix} B_{1-p_1, p_1}(1/N) & \cdots & B_{N_{s_1}, p_1}(1/N) \\ \dots & \dots & \dots \\ B_{1-p_1, p_1}(N/N) & \cdots & B_{N_{s_1}, p_1}(N/N) \end{pmatrix}_{N \times (N_{s_1} + p_1)}$$

for the positive integer p_1 . We decompose $\hat{m}_{p_1}(j/N)$ into three terms $\tilde{m}_{p_1}(j/N)$, $\tilde{\xi}_{p_1}(j/N)$

and $\tilde{\varepsilon}_{p_1}(j/N)$ in the space $\mathcal{H}^{(p_1-2)}$ of spline functions:

$$\hat{m}_{p_1}(x) = \tilde{m}_{p_1}(x) + \tilde{\varepsilon}_{p_1}(x) + \tilde{\xi}_{p_1}(x),$$

where

$$\tilde{m}_p(x) = \left\{ B_{1-p_1,p_1}(x), \dots, B_{N_s,p_1}(x) \right\} \left(\mathbf{X}_1^T \mathbf{X}_1 \right)^{-1} \mathbf{X}_1^T \mathbf{m},$$

$$\tilde{\varepsilon}_{p_1}(x) = \left\{ B_{1-p_1,p_1}(x), \dots, B_{N_s,p_1}(x) \right\} \left(\mathbf{X}_1^T \mathbf{X}_1 \right)^{-1} \mathbf{X}_1^T \mathbf{e},$$

$$\tilde{\xi}_{p_1}(x) = \left\{ B_{1-p_1,p_1}(x), \dots, B_{N_s,p_1}(x) \right\} \left(\mathbf{X}_1^T \mathbf{X}_1 \right)^{-1} \mathbf{X}_1^T \sum_{k=1}^{\kappa} \bar{\xi}_{\cdot k} \phi_k,$$

where $\mathbf{m} = (m(1/N), \dots, m(N/N))^T$ is the signal vector,

$\mathbf{e} = (\sigma(1/N)\bar{\varepsilon}_{\cdot 1}, \dots, \sigma(N/N)\bar{\varepsilon}_{\cdot N})^T$, $\bar{\varepsilon}_{\cdot j} = n^{-1} \sum_{i=1}^n \varepsilon_{ij}$, $1 \leq j \leq N$, is the noise vector and $\phi_k = (\phi_k(1/N), \dots, \phi_k(N/N))^T$ are the eigenfunction vectors, and $\bar{\xi}_{\cdot k} = n^{-1} \sum_{i=1}^n \xi_{ik}$, $1 \leq k \leq \kappa$.

Thus, one can write the residuals $\hat{U}_{ij,p_1} = Y_{ij} - \hat{m}_{p_1}(j/N)$ as

$$\hat{U}_{ij,p_1} = m(j/N) - \tilde{m}_{p_1}(j/N) - \tilde{\xi}_{p_1}(j/N) - \tilde{\varepsilon}_{p_1}(j/N) + U_{ij}.$$

Let $\hat{\bar{U}}_{\cdot jj',p_1} = n^{-1} \sum_{i=1}^n \hat{U}_{ij,p_1} \hat{U}_{ij',p_1}$. We calculate the difference of $\hat{G}_{p_1,p_2}(x, x') - \tilde{G}_{p_2}(x, x')$ by checking the difference $\hat{\bar{U}}_{\cdot jj',p_1} - \bar{U}_{\cdot jj'}$ first. For any $1 \leq j \neq j' \leq N$, one

has

$$\begin{aligned}
\hat{U}_{\cdot jj', p_1} - \bar{U}_{\cdot jj'} &= n^{-1} \sum_{i=1}^n \left[U_{ij}(\tilde{m}_{p_1} - m) \left(j'/N \right) + U_{ij'}(\tilde{m}_{p_1} - m) (j/N) \right. \\
&\quad + U_{ij} \tilde{\xi}_{p_1} \left(j'/N \right) + U_{ij'} \tilde{\xi}_{p_1} (j/N) + U_{ij} \tilde{\varepsilon}_{p_1} \left(j'/N \right) + U_{ij'} \tilde{\varepsilon}_{p_1} (j/N) \Big] \\
&\quad + \left(\hat{m}_{p_1} - m \right) \left(j'/N \right) \left(\hat{m}_{p_1} - m \right) (j/N).
\end{aligned}$$

Next, we calculate the super norm of each part of $\hat{G}_{p_1, p_2}(x, x') - \tilde{G}_{p_2}(x, x')$ respectively. One can write $\tilde{\varepsilon}_{p_1} \left(j'/N \right) = \sum_{J=1-p_1}^{N_{s1}} B_{J, p_1} \left(j'/N \right) w_J$, where $\{w_J\}_{J=1-p_1}^{N_{s1}} = N^{-1} \hat{\mathbf{V}}_{p_1}^{-1} \mathbf{X}_1^T \mathbf{e}$ and

$$\begin{aligned}
&\mathbf{X}^T \left\{ \sum_{i=1}^n U_{ij} \tilde{\varepsilon}_{p_1} \left(j'/N \right) \right\}_{1 \leq j \neq j' \leq N} \\
&= \left\{ \sum_{i=1}^n \sum_{1 \leq j \neq j' \leq N} U_{ij} \tilde{\varepsilon}_{p_1} \left(j'/N \right) B_{JJ', p_2} \left(j/N, j'/N \right) \right\}_{J, J'=1-p_2}^{N_{s2}}.
\end{aligned}$$

Thus,

$$\begin{aligned}
& \left\| \mathbf{X}^T \left\{ n^{-1} \sum_{i=1}^n U_{ij} \tilde{\varepsilon}_{p_1} \left(j'/N \right) \right\}_{1 \leq j \neq j' \leq N} \right\|_2^2 \\
&= \sum_{J, J'=1-p_2}^{N_{s_2}} \left[n^{-1} \sum_{i=1}^n \sum_{1 \leq j \neq j' \leq N} U_{ij} \left(\sum_{J''=1-p_1}^{N_{s_1}} B_{J'', p_1} \left(j'/N \right) w_{J''} \right) B_{JJ', p_2} \left(j/N, j'/N \right) \right]^2 \\
&\leq \sum_{J, J'=1-p_2}^{N_{s_2}} \sum_{J''=1-p_1}^{N_{s_1}} \left[n^{-1} \sum_{i=1}^n \sum_{1 \leq j \neq j' \leq N} U_{ij} B_{J'', p_1} \left(j'/N \right) B_{JJ', p_2} \left(j/N, j'/N \right) \right]^2 \\
&\times \sum_{J''=1-p_1}^{N_{s_1}} w_{J''}^2 = I \times II,
\end{aligned}$$

where

$$I = \sum_{J, J'=1-p_2}^{N_{s_2}} \sum_{J''=1-p_1}^{N_{s_1}} \left\{ n^{-1} \sum_{i=1}^n \sum_{1 \leq j \neq j' \leq N} U_{ij} B_{JJ', p_2} \left(j/N, j'/N \right) B_{J'', p_1} \left(j'/N \right) \right\}^2$$

$$\text{and } II = \left\| N^{-1} \hat{\mathbf{V}}_{p_1}^{-1} \mathbf{X}_1^T \mathbf{e} \right\|_2^2.$$

Let $h_* = \min \{h_{s_1}, h_{s_2}\}$. The definition of spline function implies that

$$\begin{aligned}
E[I] &= n^{-1} \sum_{J, J'=1-p_2}^{N_{s_2}} \sum_{J''=1-p_1}^{N_{s_1}} \left\{ \sum_{j, j''=1}^N E \left(U_{1j} U_{1j''} \right) B_{J, p_2} (j/N) B_{J, p_2} (j''/N) \right. \\
&\quad \times \sum_{j' \neq j}^N \sum_{j''' \neq j''}^N B_{J', p_2} (j'/N) B_{J'', p_1} (j'/N) B_{J', p_2} (j'''/N) B_{J'', p_1} (j'''/N) \left. \right\} \\
&\leq C(G, \sigma^2) n^{-1} N^4 h_{s_2}^2 h_*^2 N_{s_2} \max \{N_{s_1}, N_{s_2}\} \leq C(G, \sigma^2) n^{-1} N^4 h_{s_2} h_*.
\end{aligned}$$

Hence,

$$\left\| N^{-2} \mathbf{X}^T \left\{ n^{-1} \sum_{i=1}^n U_{ij} \tilde{\varepsilon}_{p_1} (j'/N) \right\}_{1 \leq j \neq j' \leq N} \right\|_2 = O \left(n^{-1/2} h_{s_2}^{1/2} h_*^{1/2} \right).$$

Meanwhile, one has that

$$II \leq C_{p_1} \left\| N^{-1} \mathbf{X}_1^T \mathbf{e} \right\|_2^2 h_{s_1}^{-2} = O_{\text{a.s.}} \left\{ (N n h_{s_1}^2)^{-1} \right\}.$$

By Lemma 3.7.4, one has

$$\begin{aligned}
& n^{1/2} \sup_{x, x' \in [0, 1]^2} \left\| \mathbf{B}_{p_2}^T(x, x') \hat{\mathbf{V}}_{p_2, 2}^{-1} N^{-2} \mathbf{X}^T \left\{ n^{-1} \sum_{i=1}^n U_{ij} \tilde{\varepsilon}_{p_1} (j'/N) \right\}_{1 \leq j \neq j' \leq N} \right\|_\infty \\
&= O \left\{ n^{-1/2} N^{-1/2} h_{s_1}^{-1} h_{s_2}^{1/2} h_*^{1/2} h_{s_2}^{-2} \right\} = O \left\{ n^{-1/2} N^{-1/2} h_{s_1}^{-1/2} h_{s_2}^{-3/2} \right\} = o(1).
\end{aligned}$$

Similarly, $\tilde{\xi}_{p_1}(j'/N) = \sum_{J=1-p_1}^{N_{s_1}} B_{J,p_1}(j'/N) s_J$, where

$$\{s_J\}_{J=1-p_1}^{N_{s_1}} = N^{-1} \hat{\mathbf{V}}_{p_1}^{-1} \mathbf{X}_1^T \sum_{k=1}^{\kappa} \bar{\xi}_{\cdot k} \phi_k.$$

Assumption (B3) ensures that

$$\begin{aligned} & \left\| \mathbf{X}^T \left\{ n^{-1} \sum_{i=1}^n U_{ij} \tilde{\xi}_{p_1}(j'/N) \right\}_{1 \leq j \neq j' \leq N} \right\|_2^2 \\ & \leq \sum_{J, J'=1-p_2}^{N_{s_2}} \sum_{J''=1-p_1}^{N_{s_1}} \left[n^{-1} \sum_{i=1}^n \sum_{1 \leq j \neq j' \leq N} U_{ij} B_{J'', p_1}(j'/N) B_{JJ', p_2}(j/N, j'/N) \right]^2 \\ & \times \sum_{J''=1-p_1}^{N_{s_1}} s_{J''}^2 = I \times III, \end{aligned}$$

where $III = \left\| N^{-1} \hat{\mathbf{V}}_{p_1}^{-1} \mathbf{X}_1^T \sum_{k=1}^{\kappa} \bar{\xi}_{\cdot k} \phi_k \right\|_2^2$. Note that

$$N^{-1} \mathbf{X}_1^T \sum_{k=1}^{\kappa} \bar{\xi}_{\cdot k} \phi_k = \left\{ \sum_{k=1}^{\kappa} \bar{\xi}_{\cdot k} N^{-1} \sum_{j=1}^N B_{J, p_1}(j/N) \phi_k(j/N) \right\}_{J=1-p_1}^{N_{s_1}}$$

and

$$E \left[\sum_{k=1}^{\kappa} \bar{\xi}_{\cdot k} N^{-1} \sum_{j=1}^N B_{J, p_1}(j/N) \phi_k(j/N) \right]^2 = O(n^{-1} h_{s_1}^2),$$

hence

$$III \leq C_{p_1} \left\| N^{-1} \mathbf{X}_1^T \sum_{k=1}^{\kappa} \bar{\xi}_{\cdot k} \phi_k \right\|_2^2 h_{s_1}^{-2} = O_p \left\{ (n h_{s_1})^{-1} \right\}$$

and

$$\begin{aligned}
& n^{1/2} \sup_{x, x' \in [0,1]^2} \left\| \mathbf{B}_{p_2}^T(x, x') \hat{\mathbf{V}}_{p_2, 2}^{-1} N^{-2} \mathbf{X}^T \left\{ n^{-1} \sum_{i=1}^n U_{ij} \tilde{\xi}_{p_1} \left(j'/N \right) \right\}_{1 \leq j \neq j' \leq N} \right\|_{\infty} \\
&= O \left\{ n^{-1/2} h_{s_1}^{-1/2} h_{s_2}^{1/2} h_*^{1/2} h_{s_2}^{-2} \right\} = O(n^{-1/2} h_{s_2}^{-3/2}) = o(1).
\end{aligned}$$

Next one obtains that

$$\begin{aligned}
& E \left\{ n^{-1} N^{-2} \sum_{i=1}^n \sum_{1 \leq j \neq j' \leq N} U_{ij} B_{J, p_2}(j/N) B_{J', p_2}(j'/N) (m - \tilde{m}_{p_1})(j'/N) \right\}^2 \\
&= n^{-1} N^{-4} \sum_{j, j''=1}^N E \left(U_{1j} U_{1j''} \right) B_{J, p_2}(j/N) B_{J, p_2}(j''/N) \\
&\quad \times \sum_{j' \neq j}^N \sum_{j''' \neq j''}^N B_{J', p_2}(j'/N) B_{J', p_1}(j'''/N) (m - \tilde{m}_{p_1})(j'/N) (m - \tilde{m}_{p_1})(j'''/N) \\
&\leq C(G, \sigma^2) h_{s_1}^{2p_1} n^{-1} N^{-4} (N h_{s_2})^4 = C(G, \sigma^2) h_{s_1}^{2p_1} n^{-1} h_{s_2}^4.
\end{aligned}$$

Therefore,

$$\begin{aligned}
& n^{1/2} \sup_{x, x' \in [0,1]^2} \left\| \mathbf{B}_{p_2}^T(x, x') \hat{\mathbf{V}}_{p_2, 2}^{-1} N^{-2} \mathbf{X}^T \left\{ n^{-1} \sum_{i=1}^n U_{ij} \left(m \left(j'/N \right) - \tilde{m}_{p_1} \left(j'/N \right) \right) \right\}_{j \neq j'} \right\|_{\infty} \\
&= O(h_{s_2}^{-2} h_{s_1}^{p_1} h_{s_2}) = O(h_{s_2}^{-1} h_{s_1}^{p_1}) = o(1).
\end{aligned}$$

Finally, we derive the upper bound of

$\sup_{x, x' \in [0, 1]^2} \left\| \mathbf{B}_{p_2}^T(x, x') \hat{\mathbf{V}}_{p_2, 2}^{-1} N^{-2} \mathbf{X}^T (\mathbf{m} - \hat{\mathbf{m}}_{p_1})^{\otimes 2} \right\|_{\infty}$, where

$$(\mathbf{m} - \hat{\mathbf{m}}_{p_1})^{\otimes 2} = \left\{ (m - \hat{m}_{p_1})(j/N) (m - \hat{m}_{p_1})(j'/N) \right\}_{1 \leq j \neq j' \leq N}.$$

In order to apply Lemma 3.7.4, one needs to find the upper bound of $\left\| (\mathbf{m} - \hat{\mathbf{m}}_{p_1})^{\otimes 2} \right\|_{\infty}$.

Using the similar proof as Lemma A.8 in Wang and Yang (2009) and Assumption (B3), one

has $\sup_{x \in [0, 1]} |\tilde{\varepsilon}_{p_1}(x)| + \sup_{x \in [0, 1]} |\tilde{\xi}_{p_1}(x)| = o(n^{-1/2})$. Therefore,

$$\begin{aligned} & \sup_{(x, x') \in [0, 1]^2} (m(x) - \hat{m}_{p_1}(x)) (m(x') - \hat{m}_{p_1}(x')) \\ & \leq \left[\sup_{x \in [0, 1]} (m(x) - \hat{m}_{p_1}(x)) \right]^2 \\ & \leq \left(\sup_{x \in [0, 1]} |m(x) - \tilde{m}_{p_1}(x)| + \sup_{x \in [0, 1]} |\tilde{\varepsilon}_{p_1}(x)| + \sup_{x \in [0, 1]} |\tilde{\xi}_{p_1}(x)| \right)^2 \\ & \leq \left[O(h_{s_1}^{p_1} + n^{-1/2}) \right]^2 = O(h_{s_1}^{2p_1} + n^{-1} + h_{s_1}^{p_1} n^{-1/2}) = o(n^{-1/2}). \end{aligned}$$

Hence $\left\| (\mathbf{m} - \hat{\mathbf{m}}_{p_1})^{\otimes 2} \right\|_{\infty} = o(n^{-1/2})$. Hence, the proposition has been proved.

Table 3.1: Simulation results: uniform coverage rates from 500 replications.

σ	n	$1 - \alpha$	Coverage proportion ($p_1 = p_2 = 4$)	Coverage proportion ($p_1 = p_2 = 2$)
0.5	50	0.950	0.720	0.710
		0.990	0.824	0.828
	100	0.950	0.858	0.834
		0.990	0.946	0.930
	200	0.950	0.912	0.898
		0.990	0.962	0.956
	300	0.950	0.890	0.884
		0.990	0.960	0.958
	500	0.950	0.908	0.894
		0.990	0.976	0.964
1.0	50	0.950	0.626	0.690
		0.990	0.720	0.796
	100	0.950	0.752	0.796
		0.990	0.874	0.904
	200	0.950	0.798	0.852
		0.990	0.912	0.944
	300	0.950	0.822	0.828
		0.990	0.922	0.936
	500	0.950	0.864	0.858
		0.990	0.946	0.946

Chapter 4

Spline Confidence Bands for Functional Derivatives

4.1 Introduction

In exploratory FDA, it is often of interest to estimate the mean functions; see for example, Ramsay and Silverman (2005), Yao, Müller and Wang (2005a,b), Ferraty and Vieu (2006), Li and Hsing (2010) and Cao, Yang and Todem (2012). In some settings, however, estimation and inference of derivatives of the mean functions in FDA are of equal importance. For example, in economics, consistent and direct estimation of derivatives are essential for estimating elasticities, returns to scale, substitution rates and average derivatives. Often, these index (derivative) functions are as interesting as the mean functions themselves. Another example is in the fields of engineering and biomedical sciences, where the estimation of velocity and acceleration are of great importance in addition to obtaining a smooth curve of the measurements.

The problem of estimation and inference of derivatives for functional data is very chal-

lenging; see Ramsay and Silverman (2005), Liu and Müller (2009), and Hall, Müller and Yao (2009) for some discussions. Existing methodologies for derivatives of the regression function in FDA often rely on a pointwise analysis. For example, in Liu and Müller (2009) the theoretical focus was primarily on obtaining consistency and asymptotic normality of the proposed estimators, thereby providing the necessary ingredients to construct pointwise confidence intervals. This approach is important but its usefulness in conducting global inferences is limited. To our knowledge, we are not aware of any methodology that provides simultaneous confidence bands for functional derivatives in FDA. In this chapter, we develop such methodology with the primary aim to better understand the variability and shape of the mean curve.

Nonparametric simultaneous confidence bands are powerful tools for global inference of functions. Some work has been conducted to study the simultaneous confidence band of the mean curves for FDA; see Degras (2011), Ma, Yang and Carroll (2012) and Cao, Yang and Todem (2012). The research work on confidence bands for functional derivatives is actually sparse. This is partially due to the technical difficulty to formulate such bands for FDA and establish the associated theoretical properties.

Some smoothing tools are necessary to construct the confidence bands. Popular smoothing methods include kernels (Gasser and Müller, 1984; Härdle, 1989; Xia, 1998; Claeskens and Van Keilegom, 2003), local polynomials (Fan and Gijbels, 1996), splines (Wahba, 1990; Stone, 1994) and series expansion methods (Morris and Carroll, 2006). In this chapter, we use B-splines, which can be readily implemented due to their explicit expression, to construct the bands. B-spline approximation has also been employed to estimate the functional mixed-effect models in Shi, Weiss and Taylor (1996) and Rice and Wu (2001), and to study functional data via principle components in Yao and Lee (2006) and Zhou, Huang and Car-

roll (2008). Other works include Zhou, Shen and Wolfe (1998) and Wang and Yang (2009a) who have proposed B-spline confidence bands for regression functions.

The proposed confidence bands are asymptotically the same as if all the random trajectories are correctly recorded over the entire interval. As discussed in Section 4.3, the estimators are semiparametrically efficient thereby providing partial theoretical justification for treating functional data as perfectly recorded random curves over the entire data range, as in Cao, Yang and Todem (2012).

The rest of the chapter is organized as follows. In Section 4.2, we introduce the model and the spline estimators for the mean curves and their derivatives. Section 4.3 presents the simultaneous confidence bands for the derivatives of the mean curves. Specifically, in Section 4.3.1, we show that the bands have asymptotically correct coverage probabilities; and in Section 4.3.2, we discuss how to estimate the unknown components involved in the band construction and other issues of the implementation. Section 4.4 reports findings from a simulation study and a real data set. Proofs of technical results are relegated to the Appendix.

4.2 Models and spline estimators

4.2.1 Models

We consider a collection of trajectories $\{X_i(t)\}_{i=1}^n$ which are i.i.d. realizations of a smooth random function $X(t)$, defined on a continuous interval $\mathcal{T}$. Assume that $\{X(t), t \in \mathcal{T}\}$ is a $L^2(\mathcal{T})$ process, i.e. $E \int_{\mathcal{T}} X^2(t) dt < +\infty$, and define the mean and covariance functions as $m(t) = E\{X(t)\}$ and $G(t, s) = \text{cov}\{X(t), X(s)\}$, $t, s \in \mathcal{T}$. The covariance func-

tion is a symmetric nonnegative-definite function with a spectral decomposition, $G(t, s) = \sum_{k=1}^{\infty} \lambda_k \phi_k(t) \phi_k(s)$, where $\lambda_1 \geq \lambda_2 \geq \dots \geq 0$, are the eigenvalues satisfying $\sum_{k=1}^{\infty} \lambda_k < \infty$, and $\{\phi_k(t)\}_{k=1}^{\infty}$ are the corresponding eigenfunctions that form an orthonormal basis. By the standard Karhunen-Loève representation (Hall and Hosseini-Nasab, 2006), $X_i(t) = m(t) + \sum_{k=1}^{\infty} \xi_{ik} \phi_k(t)$, where the random coefficients ξ_{ik} are uncorrelated with mean 0 and variance λ_k . In what follows, we assume that $\lambda_k = 0$, for $k > \kappa$, where κ is a positive integer or ∞ .

We consider a typical functional data setting where $X_i(\cdot)$ is recorded on a regular grid in $\mathcal{T}$, and assumed to be contaminated with measurement errors. Without loss of generality, we take $\mathcal{T} = [0, 1]$. Then the observed data are $Y_{ij} = X_i(T_{ij}) + \sigma(T_{ij}) \varepsilon_{ij}$, for $1 \leq i \leq n$, $1 \leq j \leq N$, where $T_{ij} = j/N$, ε_{ij} are independent random errors with $E(\varepsilon_{ij}) = 0$ and $E(\varepsilon_{ij}^2) = 1$, and $\sigma(\cdot)$ is the standard deviation of the measurement errors. By the Karhunen-Loève representation, the observed data can be written as

$$Y_{ij} = m(j/N) + \sum_{k=1}^{\kappa} \xi_{ik} \phi_k(j/N) + \sigma(j/N) \varepsilon_{ij}, \quad (4.1)$$

where $m(\cdot)$, $\sigma(\cdot)$ and $\{\phi_k(\cdot)\}_{k=1}^{\kappa}$ are smooth but unknown functions of t . In addition, $\{\phi_k(\cdot)\}_{k=1}^{\kappa}$ are further subject to constraints $\int_0^1 \phi_k^2(t) dt = 1$, and $\int_0^1 \phi_k(t) \phi_{k'}(t) dt = 0$, for $k' \neq k$.

4.2.2 Spline estimators

We first introduce some notation of the B-spline space. Divide the interval $\mathcal{T} = [0, 1]$ into $(N_m + 1)$ subintervals $I_J = [\omega_J, \omega_{J+1})$, $J = 0, \dots, N_m - 1$, $I_{N_m} = [\omega_{N_m}, 1]$, where $\varpi_m := \{\omega_J\}_{J=1}^{N_m}$ is a sequence of equally-spaced points, called interior knots. Let $\mathcal{H}^{(p-2)}$

be the polynomial spline space of order p on $[0, 1]$. The J -th B-spline of order p is denoted by $B_{J,p}$. We augment the boundary and the number of interior knots as $\omega_{1-p} = \dots = \omega_{-1} = \omega_0 = 0 < \omega_1 < \dots < \omega_{N_m} < 1 = \omega_{N_m+1} = \dots = \omega_{N_m+p}$, in which $\omega_J = Jh_m$, $J = 0, 1, \dots, N_m + 1$ and $h_m = 1/(N_m + 1)$ is the distance between neighboring knots.

Following Cao, Yang and Todem (2012), we estimate the mean function $m(\cdot)$ in (4.1) by

$$\hat{m}(\cdot) = \arg \min_{g(\cdot) \in \mathcal{H}^{(p-2)}} \sum_{i=1}^n \sum_{j=1}^N \left\{ Y_{ij} - g(j/N) \right\}^2 = \sum_{J=1-p}^{N_m} \hat{b}_{J,p} B_{J,p}(\cdot),$$

where the coefficients

$$\hat{\mathbf{b}}_p = \left\{ \hat{b}_{1-p,p}, \dots, \hat{b}_{N_m,p} \right\}^T = \arg \min_{\mathbb{R}^{N_m+p}} \sum_{i=1}^n \sum_{j=1}^N \left\{ Y_{ij} - \sum_{J=1-p}^{N_m} b_{J,p} B_{J,p}(j/N) \right\}^2.$$

Let $\mathbf{Y} = (\bar{Y}_{\cdot 1}, \dots, \bar{Y}_{\cdot N})^T$ and $\bar{Y}_{\cdot j} = n^{-1} \sum_{i=1}^n Y_{ij}$, $1 \leq j \leq N$. Applying elementary algebra, one obtains

$$\hat{m}(t) = \mathbf{B}_p(t) \left(\mathbf{B}^T \mathbf{B} \right)^{-1} \mathbf{B}^T \mathbf{Y} \quad (4.2)$$

in which $\mathbf{B}_p(t) = \left(B_{1-p,p}(t), \dots, B_{N_m,p}(t) \right)$ and $\mathbf{B} = \left(\mathbf{B}_p^T(1/N), \dots, \mathbf{B}_p^T(N/N) \right)^T$ is the design matrix.

We denote by $m^{(\nu)}(t)$ the ν -th order derivative of $m(t)$ with respect to t . Since $\hat{m}(t)$ is an estimator of $m(t)$, it is natural to consider $\hat{m}^{(\nu)}(t)$ as the estimator of $m^{(\nu)}(t)$, for any $\nu = 1, \dots, p-2$, i.e.

$$\hat{m}^{(\nu)}(t) = \mathbf{B}_p^{(\nu)}(t) \left(\mathbf{B}^T \mathbf{B} \right)^{-1} \mathbf{B}^T \mathbf{Y}, \quad (4.3)$$

where $\mathbf{B}_p^{(\nu)}(t) = \left(B_{1-p,p}^{(\nu)}(t), \dots, B_{N_m,p}^{(\nu)}(t) \right)$. According to B-spline property in de Boor

(2001), for $p > 2$ and $2 - p \leq J \leq Nm - 1$,

$$\frac{d}{dt}B_{J,p}(t) = (p-1) \left(\frac{B_{J,p-1}(t)}{\omega_{J+p-1} - \omega_J} - \frac{B_{J+1,p-1}(t)}{\omega_{J+p} - \omega_{J+1}} \right).$$

Therefore, $\mathbf{B}_p^{(\nu)}(t) = \mathbf{B}_{p-\nu}(t) \mathbf{D}_{(\nu)}^T$, in which $\mathbf{D}_{(\nu)} = \mathbf{D}_1 \cdots \mathbf{D}_{\nu-1} \mathbf{D}_\nu$, with matrix

$$\mathbf{D}_l = (p-l) \begin{pmatrix} \frac{-1}{\omega_1 - \omega_{1-p+l}} & 0 & 0 & \cdots & 0 & 0 \\ \frac{1}{\omega_1 - \omega_{1-p+l}} & \frac{-1}{\omega_2 - \omega_{2-p+l}} & 0 & \cdots & 0 & 0 \\ 0 & \frac{1}{\omega_2 - \omega_{2-p+l}} & \frac{-1}{\omega_3 - \omega_{3-p+l}} & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & \frac{1}{\omega_{Nm+p-l} - \omega_{Nm}} \end{pmatrix}$$

for $1 \leq l \leq \nu \leq p-2$, which is the same as equation (6) in Zhou and Wolfe (2000).

4.2.3 Convergence rate

Define the following “infeasible estimator” of function $m^{(\nu)}$

$$\bar{m}^{(\nu)}(t) = \bar{X}^{(\nu)}(t) = n^{-1} \sum_{i=1}^n X_i^{(\nu)}(t), \quad t \in [0, 1]. \quad (4.4)$$

The term “infeasible”, borrowed from Cao, Yang and Todem (2012), refers to the fact that $\bar{m}^{(\nu)}(\cdot)$ would be a natural estimator of $m^{(\nu)}(\cdot)$ if all random curves $X_i^{(\nu)}(\cdot)$ were observed.

In the following, we want to show that the spline estimator $\hat{m}^{(\nu)}(\cdot)$ in (4.3) is asymptotically

equivalent to $\bar{m}^{(\nu)}(\cdot)$.

We break the error $\hat{m}^{(\nu)}(\cdot) - m^{(\nu)}(\cdot)$ into three terms. Let $\bar{\varepsilon}_{\cdot,j} = n^{-1} \sum_{i=1}^n \varepsilon_{ij}$, $1 \leq j \leq N$. Denote the signal vector, the noise vector and the eigenfunction vectors by $\mathbf{m} = \left(m\left(\frac{1}{N}\right), \dots, m\left(\frac{N}{N}\right) \right)^T$, $\mathbf{e} = \left(\sigma\left(\frac{1}{N}\right) \bar{\varepsilon}_{\cdot,1}, \dots, \sigma\left(\frac{N}{N}\right) \bar{\varepsilon}_{\cdot,N} \right)^T$ and $\boldsymbol{\phi}_k = \left(\phi_k\left(\frac{1}{N}\right), \dots, \phi_k\left(\frac{N}{N}\right) \right)^T$. Projecting the relationship in model (4.2) onto the linear subspace of $\mathbb{R}^{Nm+p}$ spanned by $\left\{ B_{J,p}(j/N) \right\}_{1 \leq j \leq N, 1-p \leq J \leq Nm}$, we obtain the following crucial decomposition:

$$\hat{m}^{(\nu)}(t) = \tilde{m}^{(\nu)}(t) + \tilde{e}^{(\nu)}(t) + \tilde{\xi}^{(\nu)}(t), \quad (4.5)$$

where $\tilde{m}^{(\nu)}(t) = \Gamma^{(\nu)}(t) \mathbf{m}$, $\tilde{e}^{(\nu)}(t) = \Gamma^{(\nu)}(t) \mathbf{e}$ and $\tilde{\xi}^{(\nu)}(t) = \sum_{k=1}^{\kappa} \bar{\xi}_{\cdot,k} \Gamma^{(\nu)}(t) \boldsymbol{\phi}_k$ with $\Gamma^{(\nu)}(t) = \mathbf{B}_p^{(\nu)}(t) \left(\mathbf{B}^T \mathbf{B} \right)^{-1} \mathbf{B}^T$ and $\bar{\xi}_{\cdot,k} = n^{-1} \sum_{i=1}^n \xi_{ik}$, $1 \leq k \leq \kappa$.

The following proposition provides asymptotic properties of the three terms.

Proposition 4.2.1. *Under Assumptions (C1)-(C6) in Appendix, one has*

$$\sup_{t \in [0,1]} \left| \tilde{m}^{(\nu)}(t) - m^{(\nu)}(t) \right| = o\left(n^{-1/2}\right), \quad (4.6)$$

$$\sup_{t \in [0,1]} \left| \tilde{\xi}^{(\nu)}(t) - (\bar{m}^{(\nu)}(t) - m^{(\nu)}(t)) \right| = o_P\left(n^{-1/2}\right), \quad (4.7)$$

$$\sup_{t \in [0,1]} \left| \tilde{e}^{(\nu)}(t) \right| = o_P\left(n^{-1/2}\right). \quad (4.8)$$

Appendix A.2 contains proofs for the above proposition, which together with (4.5), leads to the following semiparametric efficiency result.

Theorem 4.2.1. *Under Assumptions (C1)-(C6) in Appendix, the B-spline estimator $\hat{m}^{(\nu)}$*

is asymptotically equivalent to $\bar{m}^{(\nu)}$ with the $\sqrt{n}$ approximation power, i.e.

$$\sup_{t \in [0,1]} \left| \hat{m}^{(\nu)}(t) - \bar{m}^{(\nu)}(t) \right| = o_P \left(n^{-1/2} \right).$$

Remark 4.2.1. Since the “infeasible estimator” $\bar{m}^{(\nu)}(t)$ is the sample average of i.i.d. trajectories $\{X_i(t)\}_{i=1}^n$, an application of the central limit theorem gives

$\sup_{t \in [0,1]} \left| \bar{m}^{(\nu)}(t) - m^{(\nu)}(t) \right| = O_P \left(n^{-1/2} \right)$. Thus combining with the results in Theorem 4.2.1, one has

$$\sup_{t \in [0,1]} \left| \hat{m}^{(\nu)}(t) - m^{(\nu)}(t) \right| = O_P \left(n^{-1/2} \right).$$

4.3 Confidence bands

In this section, we develop the simultaneous confidence bands for the derivative function $m^{(\nu)}(\cdot)$.

4.3.1 Asymptotic confidence bands

Let $\Sigma(\cdot, \cdot)$ be a positive definite function, and defined as $\Sigma(t, s) = \sum_{k=1}^{\kappa} \lambda_k \phi_k^{(\nu)}(t) \phi_k^{(\nu)}(s)$, $t, s \in [0, 1]$. Denote by $\zeta(t)$, $t \in [0, 1]$ a standardized Gaussian process such that $E\zeta(t) \equiv 0$, $E\zeta^2(t) \equiv 1$ with covariance function $E\zeta(t)\zeta(s) = \Sigma(t, s) \{\Sigma(t, t)\Sigma(s, s)\}^{-1/2}$, $t, s \in [0, 1]$. Denote by $q_{1-\alpha}$ the $100(1-\alpha)^{th}$ percentile of the absolute maxima distribution of $\zeta(t)$, $t \in [0, 1]$, i.e. $P \left\{ \sup_{t \in [0,1]} |\zeta(t)| \leq q_{1-\alpha} \right\} = 1 - \alpha, \forall \alpha \in (0, 1)$.

Theorem 4.3.1. *Under Assumptions (C1)-(C6) in Appendix, $\forall \alpha \in (0, 1)$, as $n \rightarrow \infty$,*

$$P \left\{ \sup_{t \in [0,1]} n^{1/2} \left| \bar{m}^{(\nu)}(t) - m^{(\nu)}(t) \right| \Sigma(t, t)^{-1/2} \leq q_{1-\alpha} \right\} \rightarrow 1 - \alpha.$$

Applying Theorems 4.2.1 and 4.3.1 gives asymptotic confidence bands for $m^{(\nu)}(t)$, $t \in [0, 1]$.

Corollary 4.3.2. *Under Assumptions (C1)-(C6) in Appendix, $\forall \alpha \in (0, 1)$, as $n \rightarrow \infty$, an asymptotic $100 \times (1 - \alpha) \%$ exact confidence band for $m^{(\nu)}(t)$ is*

$$P \left\{ m^{(\nu)}(t) \in \hat{m}^{(\nu)}(t) \pm n^{-1/2} q_{1-\alpha} \Sigma(t, t)^{1/2}, \quad t \in [0, 1] \right\} \rightarrow 1 - \alpha.$$

4.3.2 Implementation

When constructing the confidence bands, one needs to estimate the unknown function $\Sigma(t, s)$. Note that $\Sigma(t, s) = G(t, s)$, when $\nu = 0$. Following Liu and Müller (2009), we estimate $\phi_k^{(\nu)}$ through the derivatives of $G(t, s)$. According to Cao, Wang, Li and Yang (2012), $G(t, s)$ is estimated by

$$\hat{G}(t, s) = \sum_{J, J'=1-p}^{N_G} \hat{b}_{JJ'} B_{J,p}(t) B_{J',p}(s) \quad (4.9)$$

where $\hat{R}_{\cdot jj'} = n^{-1} \sum_{i=1}^n \left\{ Y_{ij} - \hat{m}(j/N) \right\} \left\{ Y_{ij'} - \hat{m}(j'/N) \right\}$, $1 \leq j \neq j' \leq N$, N_G is the number of interior knots for B-spline, and the coefficients

$$\begin{aligned} & \left\{ \hat{b}_{JJ'} \right\}_{J, J'=1-p}^{N_G} \\ &= \arg \min_{\mathbb{R}^{N_G+p} \otimes \mathbb{R}^{N_G+p}} \sum_{j \neq j'}^N \left\{ \hat{R}_{\cdot jj'} - \sum_{1-p \leq J, J' \leq N_G} b_{JJ'} B_{J,p}(j/N) B_{J',p}(j'/N) \right\}^2 \end{aligned}$$

with “ $\otimes$ ” being the tensor product of two spaces. They showed that $\hat{G}$ converges to G as n goes to ∞ . In this section, we further show that $\hat{G}$ and G are asymptotically equivalent up to the ν -th partial derivative. We define the ν -th derivative with respect to s for $G(t, s)$ and $\hat{G}(t, s)$ as

$$G^{(0,\nu)}(t, s) = \frac{\partial^\nu}{\partial s^\nu} G(t, s), \quad \hat{G}^{(0,\nu)}(t, s) = \frac{\partial^\nu}{\partial s^\nu} \hat{G}(t, s) = \sum_{J, J'} \hat{b}_{JJ'} B_{J,p}(t) B_{J',p}^{(\nu)}(s). \quad (4.10)$$

Theorem 4.3.3. *Under Assumptions (C1)-(C6), one has*

$$\sup_{(t,s) \in [0,1]^2} \left| \hat{G}^{(0,\nu)}(t, s) - G^{(0,\nu)}(t, s) \right| = o_P(1), \quad 1 \leq \nu \leq p-2.$$

The proof of Theorem 4.3.3 is given in Appendix A.3.

According to Liu and Müller (2009), we estimate the ν -th derivative of eigenfunctions $\hat{\phi}_k^{(\nu)}$ using the following eigenequations,

$$\frac{d^\nu}{ds^\nu} \int_0^1 \hat{G}(t, s) \hat{\phi}_k(t) dt = \int_0^1 \frac{\partial^\nu}{\partial s^\nu} \hat{G}(t, s) \hat{\phi}_k(t) dt = \hat{\lambda}_k \hat{\phi}_k^{(\nu)}(s), \quad (4.11)$$

where $\hat{\phi}_k$ are subject to $\int_0^1 \hat{\phi}_k^2(t) dt = 1$ and $\int_0^1 \hat{\phi}_k(t) \hat{\phi}_{k'}(t) dt = 0$ for $k' < k$. If N is sufficiently large, the left hand side of (4.11) can be approximated by

$$\frac{1}{N} \sum_{j=1}^N \hat{G}^{(0,\nu)}\left(\frac{j}{N}, \frac{j'}{N}\right) \hat{\phi}_k\left(\frac{j}{N}\right). \quad \text{Then we estimate } \Sigma(t, s) \text{ by}$$

$$\hat{\Sigma}(t, s) = \sum_{k=1}^{\kappa} \hat{\lambda}_k \hat{\phi}_k^{(\nu)}(t) \hat{\phi}_k^{(\nu)}(s).$$

The following theorem shows that $\hat{\Sigma}(\cdot, \cdot)$ and $\Sigma(\cdot, \cdot)$ are asymptotically equivalent.

Theorem 4.3.4. *Under Assumptions (C1)-(C6), one has*

$$\sup_{(s,t) \in [0,1]^2} \left| \hat{\Sigma}(t,s) - \Sigma(t,s) \right| = o_P(1).$$

The proof of Theorem 4.3.4 is given in Appendix A.4.

In practice, we choose the first L positive eigenvalues $\hat{\lambda}_1 \geq \dots \geq \hat{\lambda}_L > 0$ by eigenvalue decomposition of $\hat{G}(t,s)$. Then we apply a standard criterion in Müller (2009), to choose the number of eigenfunctions, i.e. $\kappa = \arg \min_{1 \leq l \leq L} \left\{ \sum_{k=1}^l \hat{\lambda}_k / \sum_{k=1}^L \hat{\lambda}_k > 0.95 \right\}$. Müller (2009) suggests the “pseudo-AIC” and this simple method of counting the percentage of variation explained can be used to choose the number of principal components. The simple method performed well in our simulations and is used for our numerical studies.

To construct the confidence bands, we use cubic splines to estimate the mean and covariance functions and their first order derivatives. Generalized cross-validation is used to choose the number of knots N_m (from 2 to 20), to smooth out the mean function. According to Assumption (C3), the number of knots for smoothing the covariance function is taken to be $N_G = \lceil n^{1/(2p)} \log(n) \rceil$, where $\lceil a \rceil$ denotes the integer part of a .

Finally, in order to estimate $q_{1-\alpha}$, we generate i.i.d standard normal variables $Z_{k,b}$, $1 \leq k \leq \kappa$, $b = 1, \dots, 5000$. Let $\hat{\zeta}_b(t) = \hat{\Sigma}(t,t)^{-1/2} \sum_{k=1}^{\kappa} \sqrt{\hat{\lambda}_k} Z_{k,b} \hat{\phi}_k^{(\nu)}(t)$, $t \in [0,1]$, $q_{1-\alpha}$ can be estimated by 100(1- α)-th percentile of $\{\sup_{t \in [0,1]} |\hat{\zeta}_b(t)|\}_{b=1}^{5000}$. Therefore, in application we recommend the following band

$$\hat{m}^{(\nu)}(t) \pm n^{-1/2} \hat{\Sigma}(t,t)^{1/2} \hat{q}_{1-\alpha}, \quad t \in [0,1]. \quad (4.12)$$

4.4 Numerical studies

4.4.1 Simulated examples

To illustrate the finite-sample performance of the confidence band in (4.12), we generate data from the following:

$$Y_{ij} = m(j/N) + \sum_{k=1}^{\kappa} \xi_{ik} \phi_k(j/N) + \varepsilon_{ij}, \quad \varepsilon_{ij} \stackrel{i.i.d.}{\sim} N(0, 0.1^2),$$

for $1 \leq i \leq n$, $1 \leq j \leq N$. We consider two scenarios.

Model I: $m(t) = 5t + 4 \sin(2\pi(t-0.5))$, $\phi_1(t) = -\sqrt{2} \cos(2\pi(t-0.5))$, $\phi_2(t) = \sqrt{2} \sin(4\pi(t-0.5))$, $\xi_{ik} \sim N(0, \lambda_k)$, $\lambda_1 = 2$, $\lambda_2 = 1$, $\kappa = 2$;

Model II: $m(t) = 4t + \frac{1}{\sqrt{2\pi}0.1} \exp\left(-\frac{(t-0.5)^2}{2(0.1)^2}\right)$, $\phi_k(t) = \sqrt{2} \sin(\pi kt)$, $\xi_{ik} \sim N(0, \lambda_k)$, $\lambda_k = 2^{-(k-1)}$, $k = 1, 2, \dots, \kappa = 8$.

The second case has similar design as in Simulation C of Liu and Müller (2009). We use the proposed method in (4.12) and its “oracle” version with true $\Sigma(t, t)$ to construct the confidence bands for $m^{(1)}(\cdot)$ respectively in both studies. We consider two confidence levels: $1 - \alpha = 0.95, 0.99$. The number of trajectories n is taken to be 30, 50, 100, 200, and for each n , we try different numbers of observations on the trajectory. Each simulation consists of 1000 Monte Carlo samples.

We evaluate the coverage of the bands over 200 equally spaced points on $[0, 1]$ and test whether the true functions are covered by the confidence bands at these points. Tables 4.1 and 4.2 show the empirical coverage probabilities out of 1000 replications for Models I and II, respectively. From Tables 4.1 and 4.2, we observe that coverage probabilities for

Table 4.1: Coverage rates of the spline confidence bands in Model I.

n	N	95%		99%	
		Est.	Oracle	Est.	Oracle
30	30	0.756	0.831	0.901	0.936
	60	0.833	0.900	0.926	0.969
50	50	0.824	0.863	0.932	0.933
	100	0.856	0.904	0.947	0.979
100	100	0.851	0.897	0.943	0.971
	200	0.856	0.907	0.949	0.971
200	200	0.866	0.910	0.950	0.972
	400	0.869	0.944	0.959	0.987

both estimated bands and “oracle” bands approach the nominal levels, which show positive confirmation of Theorem 4.3.1. In most of the scenarios the “oracle” confidence bands outperform the estimated bands, and the “oracle” bands arrive at about the nominal coverage for large n and N . The convergence rates of estimated bands are slower than those “oracle” bands, but the convergence trend to nominal level is clearly.

Figure 4.1 to Figure 4.8 show the estimated functions and their 99% confidence bands for the first order derivative curve $m^{(1)}(\cdot)$ for Models I and II, respectively. As expected when n and N increase, the confidence band is narrower and the cubic spline estimator is closer to the true derivative curve. For Model I, the boundary effects in all four panels are almost unnoticeable. For Model II, there seems to be some boundary effects for small n and N , which are attenuated as n and N increase.

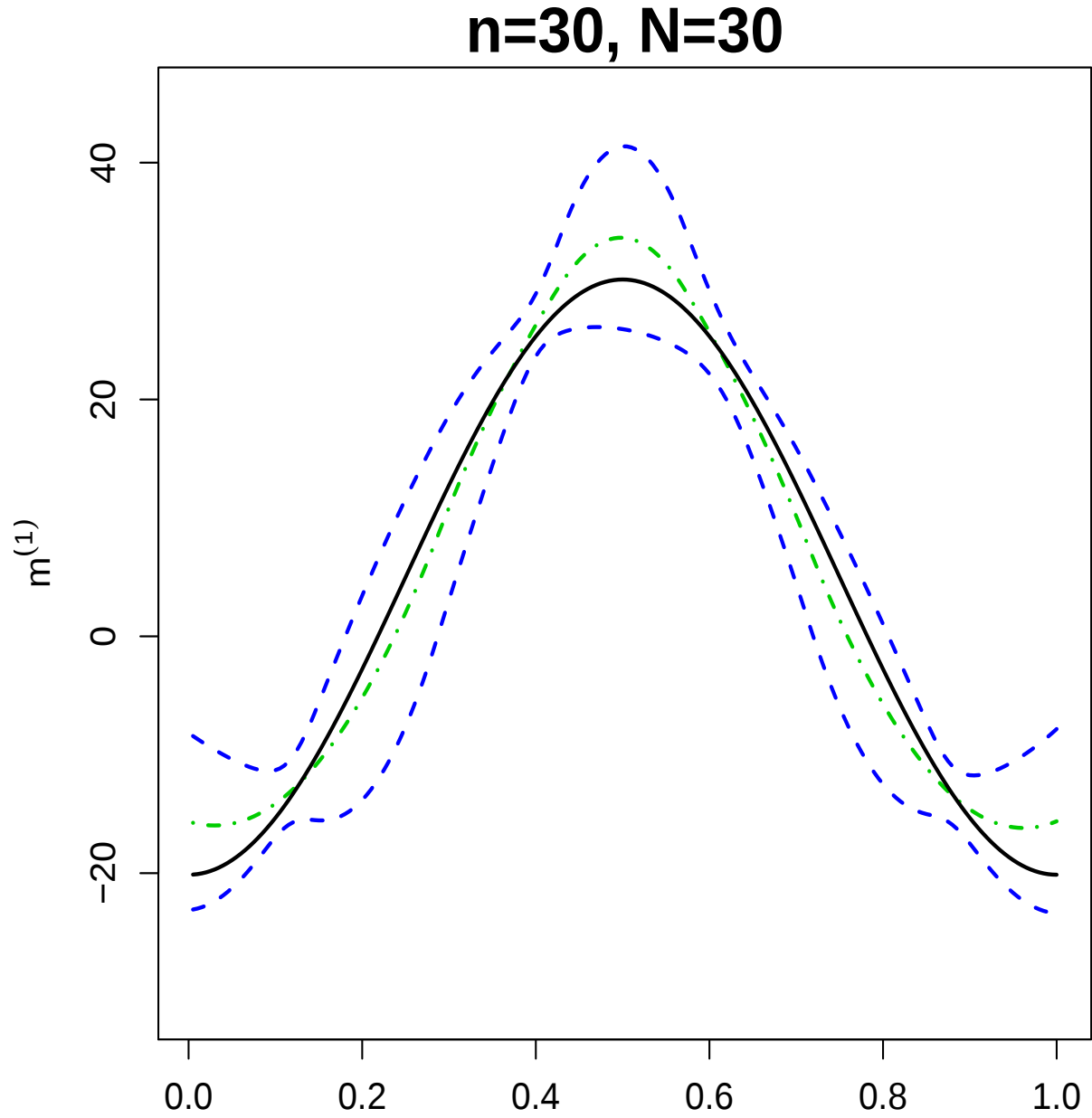


Figure 4.1: Plots of the cubic spline estimators (dotted-dashed line) and 99% confidence bands (upper and lower dashed lines) of $m^{(1)}(t)$ (solid line) in Model I. $n = 30$, $N = 30$.

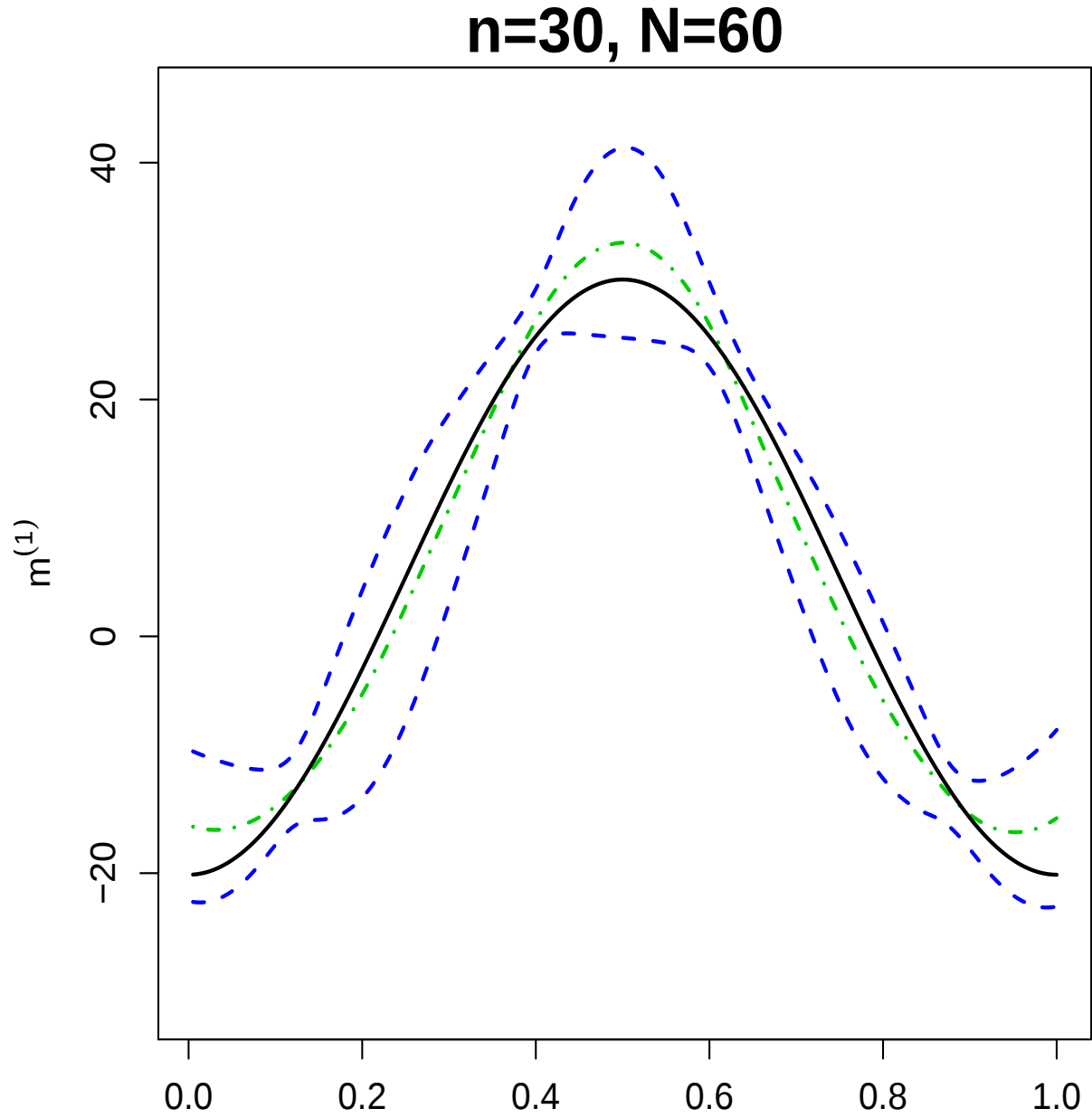


Figure 4.2: Plots of the cubic spline estimators (dotted-dashed line) and 99% confidence bands (upper and lower dashed lines) of $m^{(1)}(t)$ (solid line) in Model I. $n = 30$, $N = 60$.

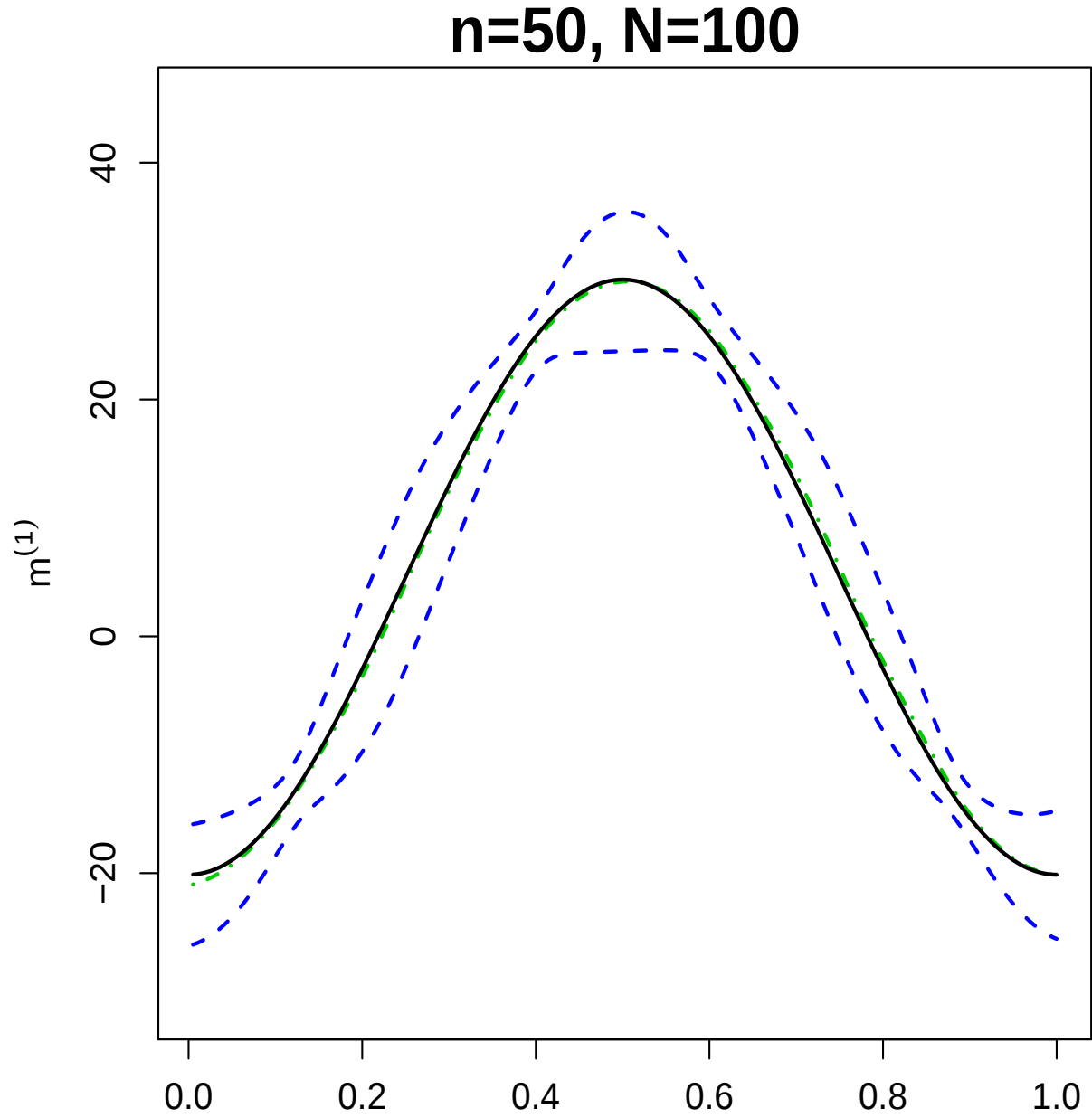


Figure 4.3: Plots of the cubic spline estimators (dotted-dashed line) and 99% confidence bands (upper and lower dashed lines) of $m^{(1)}(t)$ (solid line) in Model I. $n = 50$, $N = 50$.

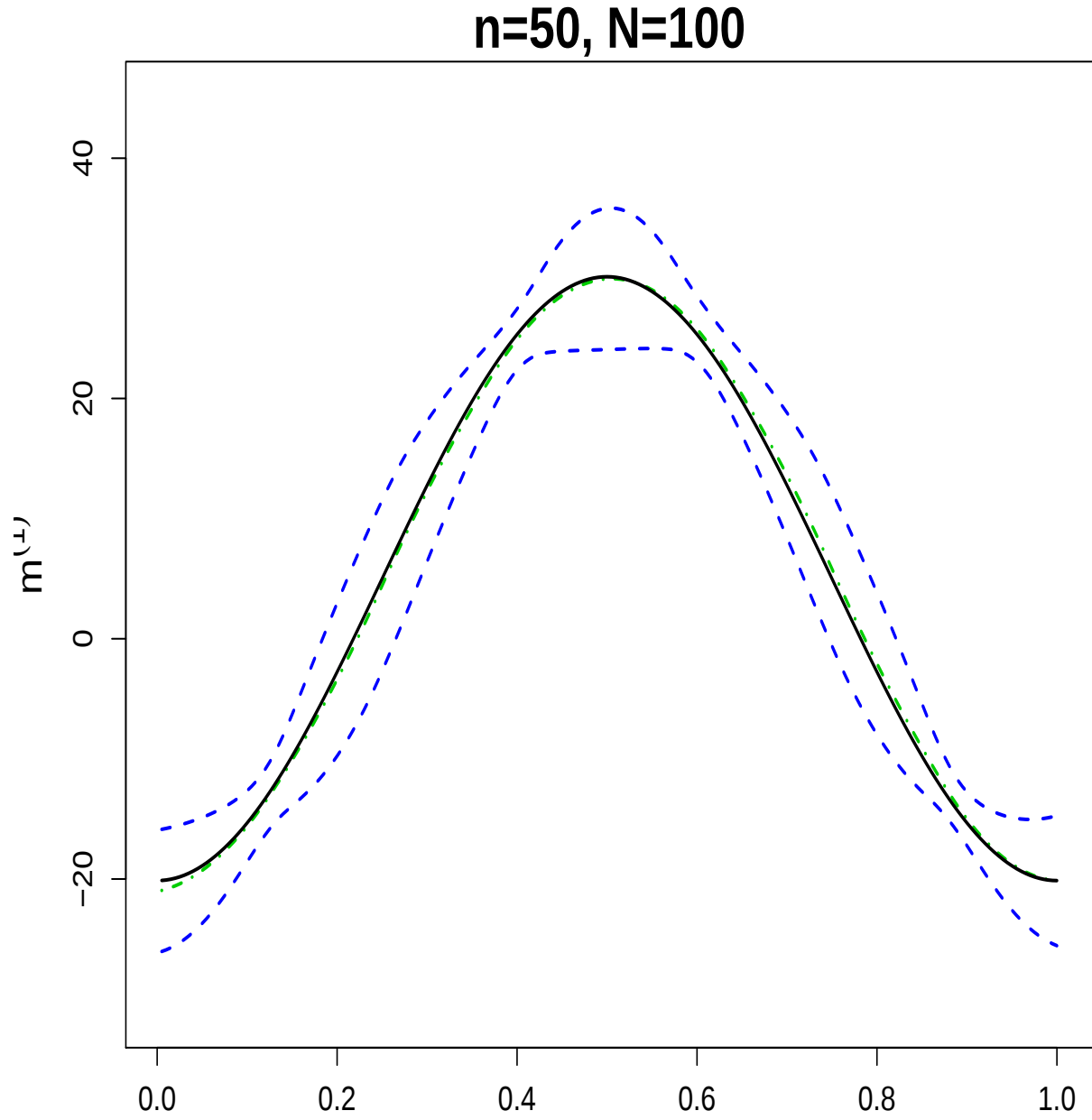


Figure 4.4: Plots of the cubic spline estimators (dotted-dashed line) and 99% confidence bands (upper and lower dashed lines) of $m^{(1)}(t)$ (solid line) in Model I. $n = 50$, $N = 100$.

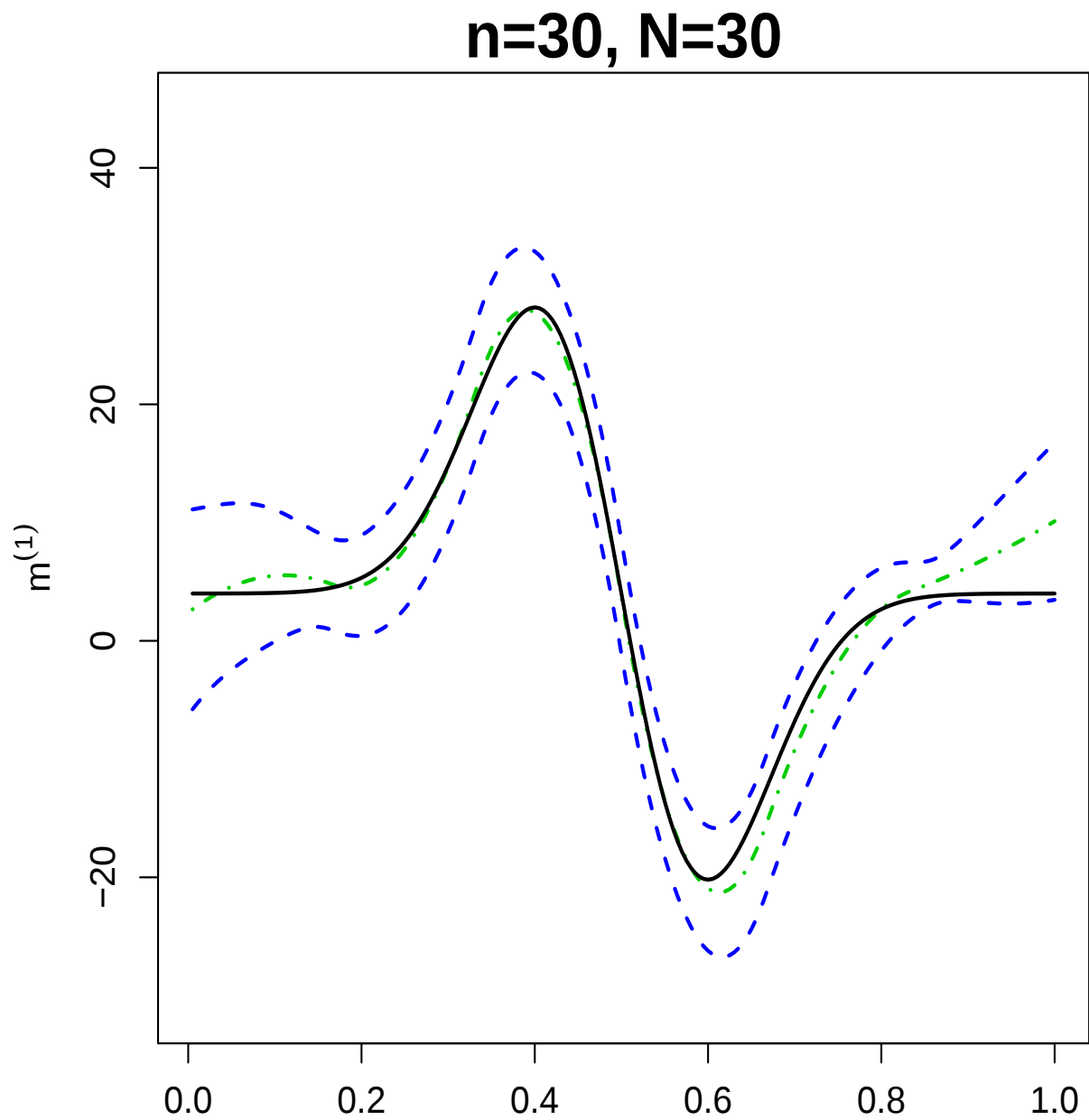


Figure 4.5: Plots of the cubic spline estimators (dotted-dashed line) and 99% confidence bands (upper and lower dashed lines) of $m^{(1)}(t)$ (solid line) in Model II. $n = 30$, $N = 30$.

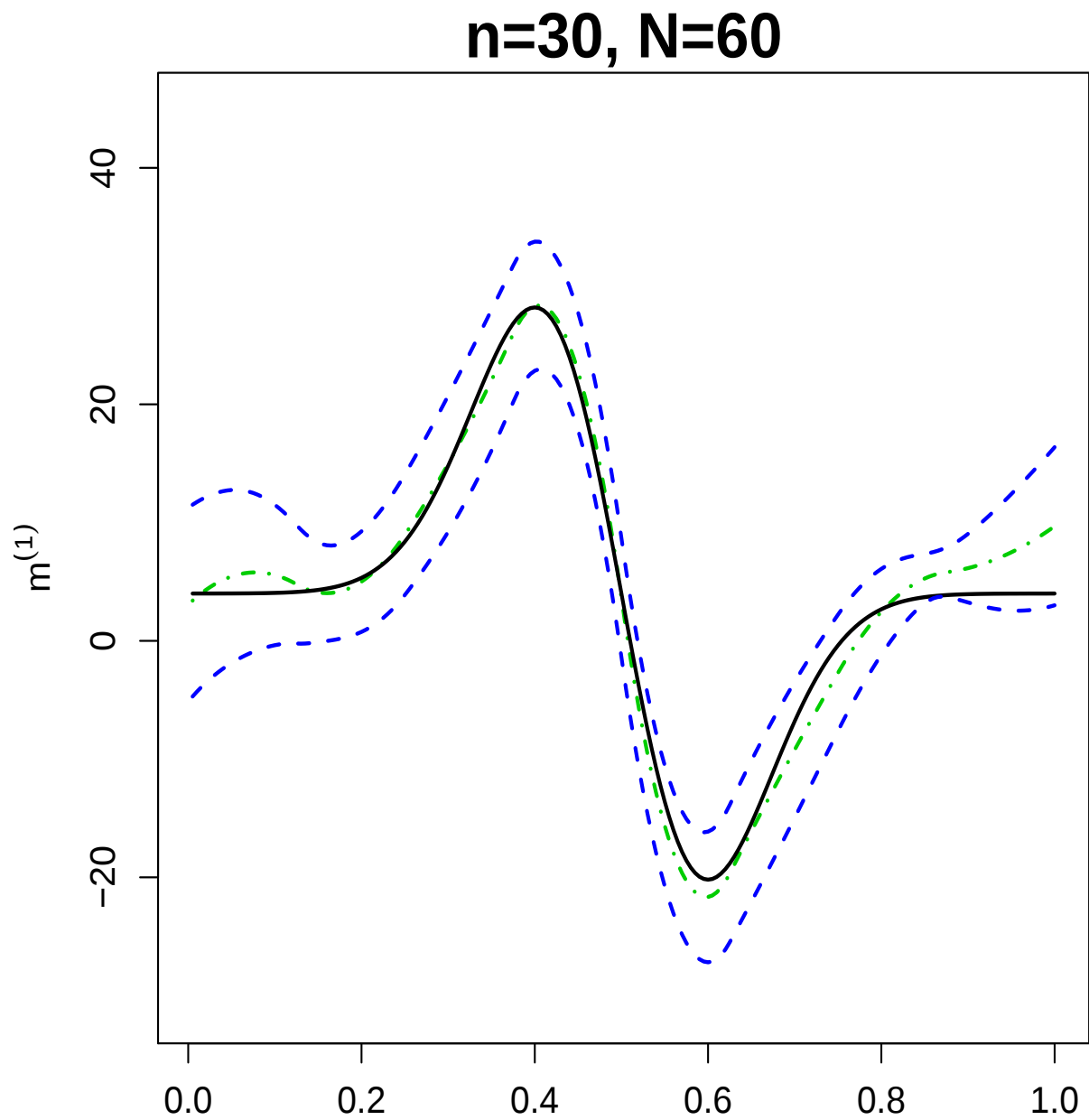


Figure 4.6: Plots of the cubic spline estimators (dotted-dashed line) and 99% confidence bands (upper and lower dashed lines) of $m^{(1)}(t)$ (solid line) in Model II. $n = 30$, $N = 60$.

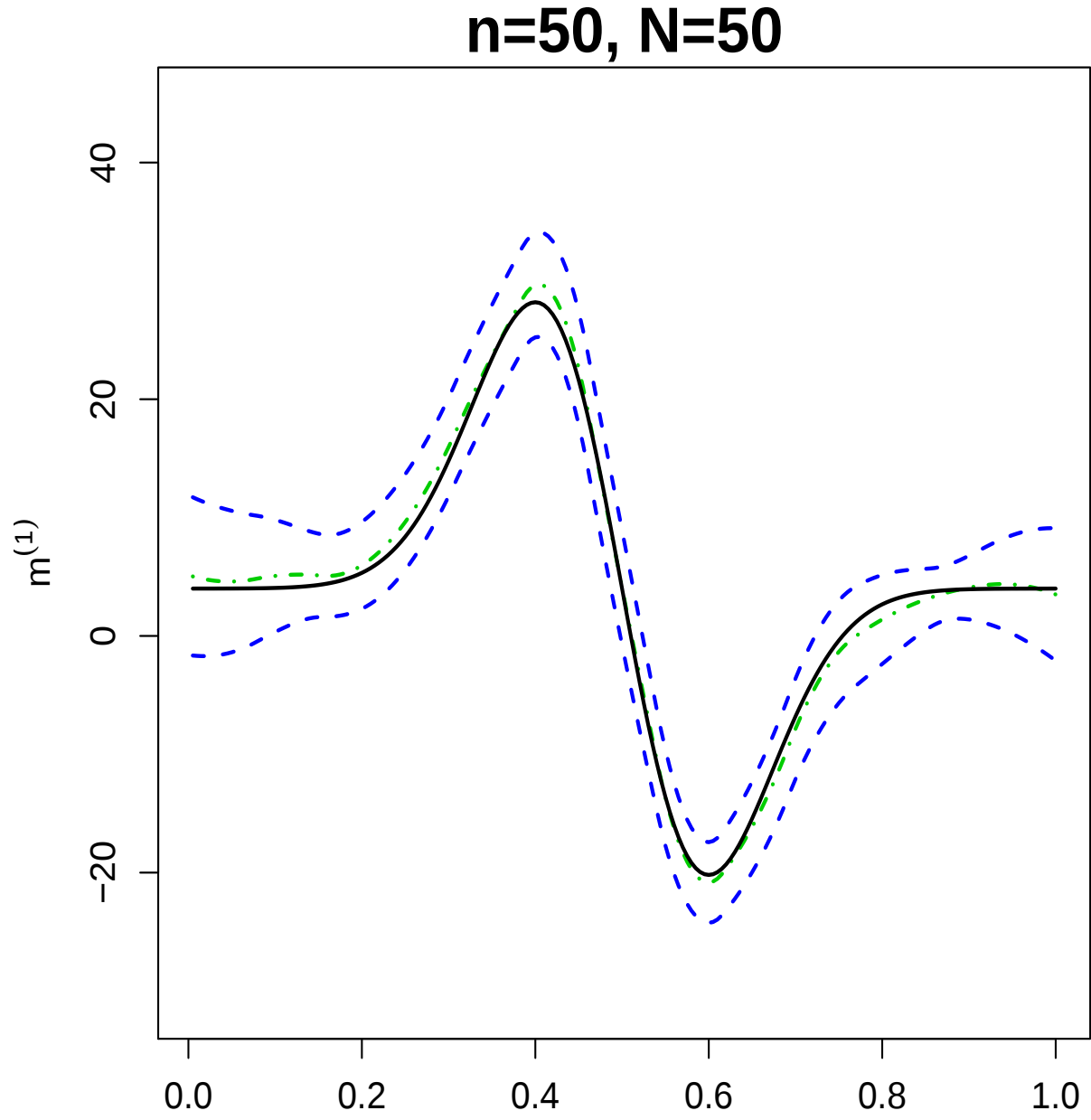


Figure 4.7: Plots of the cubic spline estimators (dotted-dashed line) and 99% confidence bands (upper and lower dashed lines) of $m^{(1)}(t)$ (solid line) in Model II. $n = 50$, $N = 50$.

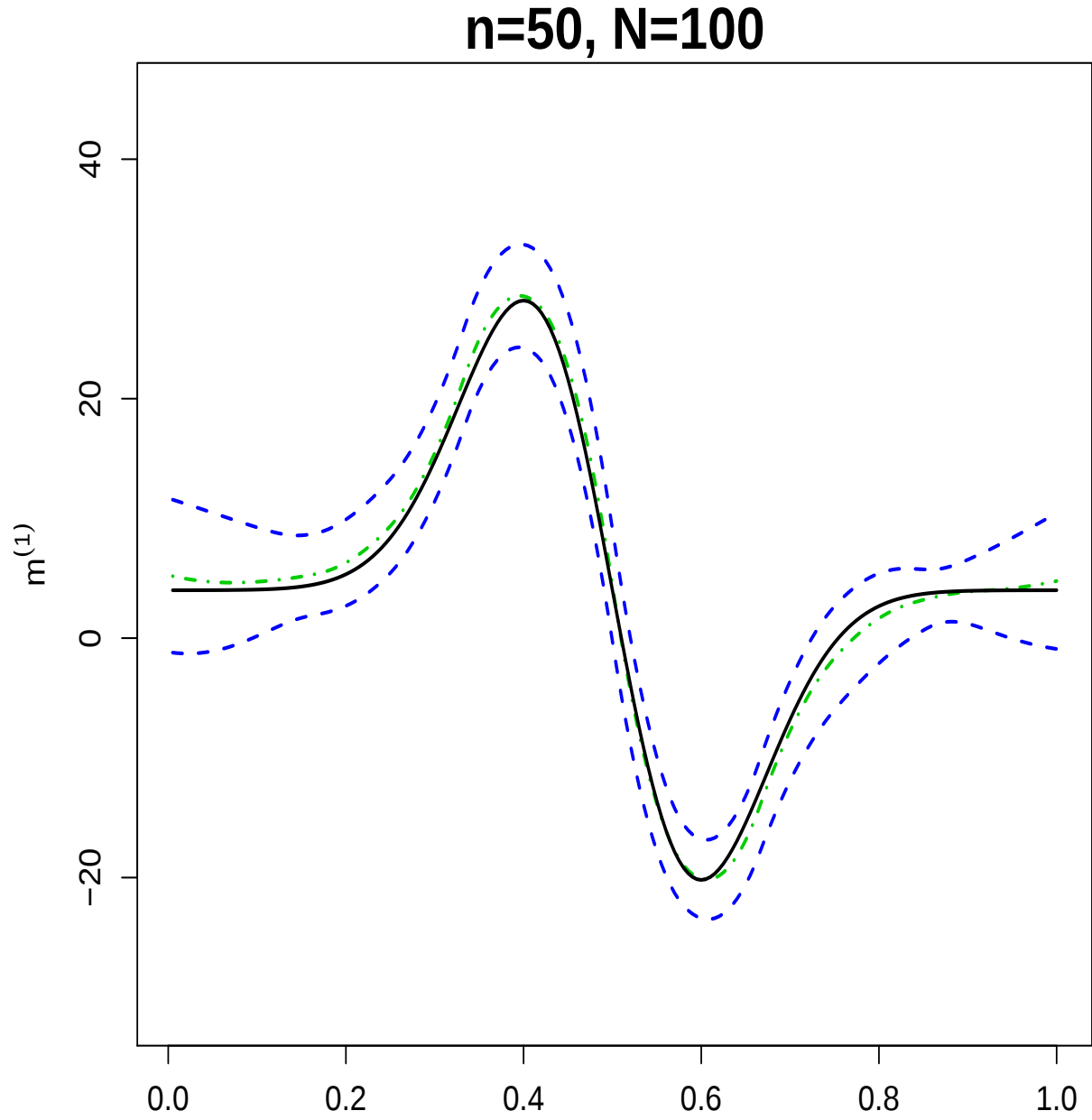


Figure 4.8: Plots of the cubic spline estimators (dotted-dashed line) and 99% confidence bands (upper and lower dashed lines) of $m^{(1)}(t)$ (solid line) in Model II. $n = 50$, $N = 100$.

Table 4.2: Coverage rates of the spline confidence bands in Model II.

n	N	95%		99%	
		Est.	Oracle	Est.	Oracle
30	30	0.650	0.730	0.830	0.854
	60	0.695	0.839	0.860	0.941
50	50	0.817	0.911	0.930	0.982
	100	0.830	0.929	0.933	0.986
100	100	0.858	0.940	0.948	0.986
	200	0.876	0.939	0.960	0.986
200	200	0.874	0.939	0.949	0.980
	400	0.889	0.946	0.963	0.991

4.4.2 Tecator data

Here we apply the proposed method to the Tecator dataset, which can be introduced in chapter 2.6. Figure 4.9 shows the estimated mean absorbance measurements $m(\cdot)$ and its estimated first order derivative $m^{(1)}(\cdot)$ in the lower panel. Their 99% confidence bands (dashed lines) are also included in the figure, both bands have similar band width around 0.1 to 0.2 even though the bands for $m^{(1)}(\cdot)$ looks much narrower in the figure.

As shown in Figure 4.9, in the region of 850 – 950 nm the derivative estimate of mean absorbance is increasing gradually above 0, which corroborates with the convex behavior of the corresponding estimated mean function. For wavelength between 950 and 970 nm, the big bump in the derivative graph is consistent with the changing pattern of the mean estimate before it reaches the turning point at around 970 nm. When wavelength is larger than 970 nm, its estimated derivative turns negative and is relatively flat after 1000 nm, and

it is in accordance with the quick dip of the mean absorbance and a linear decreasing trend for wavelength larger than 1000 nm.

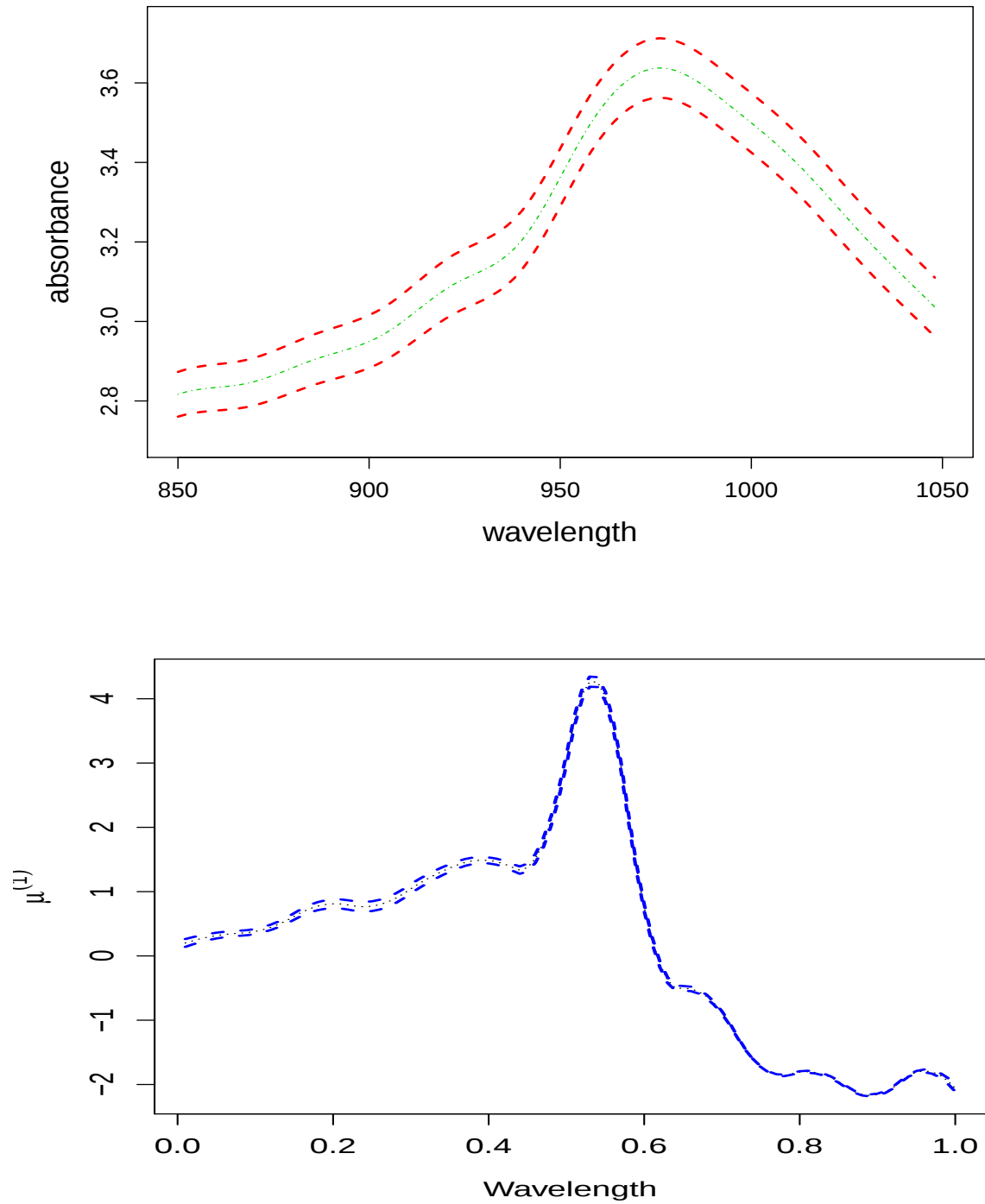


Figure 4.9: Plots of the cubic spline estimators (dotted-dashed line) and 99% confidence bands (upper and lower dashed lines) of the mean function and its first order derivative.

APPENDIX

For any $r \in (0, 1]$, we denote $C^{q,r} [0, 1]$ as the space of Hölder continuous functions on $[0, 1]$, $C^{q,r} [0, 1] = \left\{ \phi : \|\phi\|_{q,r} < +\infty \right\}$. The technical assumptions we need are as follows:

- (C1) *The regression function $m \in C^{p-1,1} [0, 1]$;*
- (C2) *The standard deviation function $\sigma(t) \in C^{0,\delta} [0, 1]$ for some $\delta \in (0, 1]$;*
- (C3) *The number of observations for each trajectory $N \gg n^\theta$ for some $\theta > \frac{1+2\nu}{2(p-\nu)}$; the number of interior knots satisfies $n^{\frac{1}{2(p-\nu)}} \ll N_m \ll (N/\log(n))^{\frac{1}{1+2\nu}}$, $n^{\frac{1}{2p}} \ll N_G \ll n^{\frac{1}{2+2\nu}}$;*
- (C4) *There exists a constant $C > 0$ such that $\Sigma(t, t) > C$, for any $t \in [0, 1]$;*
- (C5) *For $k \in \{1, \dots, \kappa\}$, $\iota = 0, 1, \dots, p-2$, $\phi_k^{(\iota)}(t) \in C^{0,\delta} [0, 1]$, for some $\delta \in (0, 1]$, $\sum_{k=1}^{\kappa} \sqrt{\lambda_k} \left\| \phi_k^{(\iota)} \right\|_{\infty} < \infty$; and for a sequence $\{\kappa_n\}_{n=1}^{\infty}$ of increasing integers with $\lim_{n \rightarrow \infty} \kappa_n = \kappa$, $N_m^{-\delta} \sum_{k=1}^{\kappa_n} \sqrt{\lambda_k} \left\| \phi_k^{(\iota)} \right\|_{0,\delta} = o(1)$;*
- (C6) *There exist $\delta_1 > 0, \delta_2 > 0$ such that $E \left| \xi_{ik} \right|^{4+\delta_1} + E \left| \varepsilon_{ij} \right|^{4+\delta_2} < +\infty$, for $1 \leq i < \infty$, $1 \leq k \leq \kappa$, $1 \leq j < \infty$. The number κ of nonzero eigenvalues is finite or κ is infinite while the variables $\{\xi_{ik}\}_{1 \leq i < \infty, 1 \leq k < \infty}$ are i.i.d..*

Assumptions (C1)-(C6) are standard in the spline smoothing literature; see Huang (2003) and Cao, Yang and Todem (2012), for instance. In particular, (C1) and (C2) guarantee the convergence rates of $\hat{m}(t)$ and its derivatives. Assumption (C3) states the requirement of the number of observations within each curve to the sample size, and the order of the number of knots of splines. Assumptions (C4) and (C5) concern that the derivatives of principal components have collectively bounded smoothness. When $\iota = 0$, Assumption (C5) is the

same as (C4) in Cao, Yang and Todem (2012). Assumption (C6) is necessary when using strong approximation result in Lemma 4.4.4.

If κ is finite and all $\phi_k^{(\nu)}(t) \in C^{0,\delta}[0,1]$, then Assumption (C5) on $\phi_k^{(\nu)}$'s holds trivially. For $\kappa = \infty$, Assumption (C5) is fulfilled as long as λ_k decreases to zero sufficiently fast. For example, considering the following canonical orthonormal Fourier basis of $L^2([0,1])$:

$$\begin{aligned}\phi_1(t) &\equiv 1, \phi_{2k+1}(t) \equiv \sqrt{2} \cos(k\pi t) \\ \phi_{2k}(t) &\equiv \sqrt{2} \sin(k\pi t), k = 1, 2, \dots, t \in [0, 1],\end{aligned}$$

we can take $\lambda_1 = 1$ and $\lambda_k = ([k/2]\pi)^{-2\nu} \rho^{2[k/2]}$, $k = 2, \dots$, for any $\rho \in (0,1)$, then $\sum_{k=1}^{\infty} \sqrt{\lambda_k} \left\| \phi_k^{(\nu)} \right\|_{\infty} = 1 + \sum_{k=1}^{\infty} \rho^k (\sqrt{2} + \sqrt{2}) = 1 + 2\sqrt{2}\rho(1-\rho)^{-1} < \infty$. While for any $\delta = 1$ and $\{\kappa_n\}_{n=1}^{\infty}$ with κ_n increasing, odd and $\kappa_n \rightarrow \infty$, one has

$$\begin{aligned}N_m^{-1} \sum_{k=1}^{\kappa_n} \sqrt{\lambda_k} \left\| \phi_k^{(\nu)} \right\|_{0,1} &= N_m^{-1} \sum_{k=1}^{(\kappa_n-1)/2} \rho^k \left(\sqrt{2}k\pi + \sqrt{2}k\pi \right) \\ &\leq 2\sqrt{2}\pi N_m^{-1} \rho \sum_{k=1}^{\infty} \rho^{k-1} k = 2\sqrt{2}\pi N_m^{-1} (1-\rho)^{-2} \\ &= O\left(N_m^{-1}\right) = o(1).\end{aligned}$$

In the following, define the theoretical and empirical inner product matrices of

$$\left\{ B_{J,p}(t) \right\}_{J=1-p}^{N_m} \text{ as}$$

$$\begin{aligned}\mathbf{V}_p &= \left(\left\langle B_{J,p}, B_{J',p} \right\rangle \right)_{J,J'=1-p}^{N_m} = \left(v_{JJ',p} \right)_{J,J'=1-p}^{N_m}, \\ \hat{\mathbf{V}}_p &= \left(\left\langle B_{J,p}, B_{J',p} \right\rangle_{2,N} \right)_{J,J'=1-p}^{N_m} = \left(\hat{v}_{JJ',p} \right)_{J,J'=1-p}^{N_m}.\end{aligned}\tag{4.13}$$

We establish next that $\hat{\mathbf{V}}_p$ has an inverse with bounded L_{∞} norm.

Lemma 4.4.1. *[Cao, Yang and Todem (2012)] Under Assumption (C3), for $\mathbf{V}_p$ and $\hat{\mathbf{V}}_p$ defined in (4.13), $\|\mathbf{V}_p - \hat{\mathbf{V}}_p\|_\infty = O(N^{-1})$ and $\|\hat{\mathbf{V}}_p^{-1}\|_\infty = O(N_m)$.*

Proof of Proposition 4.2.1

Following Wang and Yang (2009b), we introduce the p -th order quasi-interpolant of m corresponding to the knots ϖ , denoted by $Q_\varpi(m)$; see equation (4.12) on page 146 of DeVore and Lorentz (1993) for details. According to Theorem 7.7.4, DeVore and Lorentz (1993), the following lemma holds.

Lemma 4.4.2. *There exists a constant $C > 0$, such that for $0 \leq \nu \leq p - 2$ and $m \in C^{p,1}[0, 1]$,*

$$\|(m - Q_\varpi(m))^{(\nu)}\|_\infty \leq C \|m^{(p)}\|_\infty h_m^{p-\nu}.$$

Lemma 4.4.3. *Under Assumptions (C2), (C3) and (C6), one has*

$$\frac{1}{N} \sum_{j=1}^N B_{J,p}(j/N) \sigma(j/N) \bar{\varepsilon}_{\cdot j} = O_P \left(\sqrt{\frac{\log(n)}{nNN_m}} \right).$$

PROOF. We first truncate the random error ε_{ij} by $u_n = (nN)^\gamma$ ($2/9 < \gamma < 1/3$) and write $\varepsilon_{ij} = \varepsilon_{ij,1} + \varepsilon_{ij,2} + a_{ij}$, where $\varepsilon_{ij,1} = \varepsilon_{ij} I \left\{ |\varepsilon_{ij}| > u_n \right\}$, $\varepsilon_{ij,2} = \varepsilon_{ij} I \left\{ |\varepsilon_{ij}| \leq u_n \right\} - a_{ij}$ and $a_{ij} = E \left[\varepsilon_{ij} I \left\{ |\varepsilon_{ij}| \leq u_n \right\} \right]$. It is easy to see that $|a_{ij}| = \left| -E \left[\varepsilon_{ij} I \left\{ |\varepsilon_{ij}| > u_n \right\} \right] \right| \leq E \left(|\varepsilon_{ij}|^{4+\delta_2} \right) u_n^{-(3+\delta_2)}$. It is straightforward from the boundedness of spline basis that

$$\left| \frac{1}{nN} \sum_{j=1}^N B_{J,p}(j/N) \sigma(j/N) \sum_{i=1}^n a_{ij} \right| = O(u_n^{-(3+\delta_2)}).$$

Next we show that the tail part vanishes almost surely. Note that

$$\sum_{n=1}^{\infty} P \left\{ \left| \varepsilon_{ij} \right| > u_n \right\} \leq \sum_{n=1}^{\infty} \frac{E \left| \varepsilon_{ij} \right|^{4+\delta_2}}{u_n^{4+\delta_2}} \leq M_{\delta} \sum_{n=1}^{\infty} u_n^{-(4+\delta_2)} < \infty.$$

By the Borel-Cantelli Lemma, $P \left\{ \omega | \exists N(\omega), \left| \varepsilon_{ij}(\omega) \right| \leq u_n \text{ for } n > N(\omega) \right\} = 1$. Let $v_{\varepsilon} = \max \left\{ \left| \varepsilon_{ij} \right|, 1 \leq i, j \leq N(\omega) \right\}$ and there exists $N_1(\omega) > N(\omega)$, $u_{N_1(\omega)} > v_{\varepsilon}$. Since $u_n = (nN)^{\gamma}$ is an increasing function, $u_n > u_{N_1(\omega)} > v_{\varepsilon}$ and $n > N_1(\omega)$.

Therefore $P \{ \omega | \exists N(\omega), \left| \varepsilon_{ij}(\omega) \right| \leq u_n, 1 \leq i \leq n, 1 \leq j \leq N, \text{ for } \min(n, N) > N(\omega) \} = 1$, which implies $P \{ \omega | \exists N(\omega), \left| \varepsilon_{ij,1} \right| = 0, 1 \leq i \leq n, 1 \leq j \leq N \text{ for } \min(n, N) > N(\omega) \} = 1$. Thus

$$\left| \frac{1}{nN} \sum_{j=1}^N B_{J,p}(j/N) \sigma(j/N) \sum_{i=1}^n \varepsilon_{ij,1} \right| = O_{a.s.} (nN)^{-k}, \text{ for any } k > 0.$$

Next denote $D_j = (nN)^{-1} B_{J,p}(j/N) \sigma(j/N) \sum_{i=1}^n \varepsilon_{ij,2}$. Since

$$\text{Var} \left(\varepsilon_{ij,2} \right) = 1 - E \left[\varepsilon_{ij}^2 I \left\{ \left| \varepsilon_{ij} \right| > u_n \right\} \right] - a_{ij}^2 = 1 + O_P \left\{ u_n^{-\delta_2} + u_n^{-2(1+\delta_2)} \right\},$$

one has $V_n^2 = \text{Var} \left(\sum_{j=1}^N D_j \right) = c(nNN_m)^{-1}$ for a constant $c > 0$. Now Minkowski's inequality implies that

$$E \left| \varepsilon_{ij,2} \right|^k \leq 2^{k-1} E \left\{ \varepsilon_{ij}^k I \left\{ \left| \varepsilon_{ij} \right| \leq u_n \right\} + a_{ij}^k \right\} \leq 2^{k-2} u_n^{k-2} E \left| \varepsilon_{ij,2} \right|^2 k!, \quad k \geq 2.$$

Thus $E \left| D_j \right|^k \leq \left(2n^{-1} N^{-1} u_n \right)^{k-2} k! E(D_j^2) < \infty$ with the Cramer constant $c^* = \frac{2u_n}{nN}$.

For any δ , let $\delta_n = \delta \sqrt{\frac{\log(n)}{nNNm}}$. By the Bernstein inequality, for any large enough $\delta > 0$,

$$P \left\{ \left| \sum_{j=1}^N D_j \right| \geq \delta_n \right\} \leq 2 \exp \left\{ \frac{-\delta_n^2}{4V_n^2 + 2c^* \delta_n} \right\} = 2 \exp \left\{ \frac{-\delta_n^2}{\frac{4c}{nNNm} + 4\frac{u_n}{nN} \delta_n} \right\} \leq 2n^{-3}.$$

Hence $\sum_{n=1}^{\infty} P \left(\left| \frac{1}{nN} \sum_{j=1}^N B_{J,p}(j/N) \sigma(j/N) \sum_{i=1}^n \varepsilon_{ij,2} \right| \geq \delta_n \right) \leq 2 \sum_{n=1}^{\infty} n^{-3} < \infty$, for such $\delta > 0$. Thus Borel-Cantelli's lemma implies the desired result.

Lemma 4.4.4. *[Theorem 2.6.7 of Csörgő and Révész (1981)] Suppose that ξ_i , $1 \leq i \leq n$ are i.i.d. with $E(\xi_1) = 0$, $E(\xi_1^2) = 1$ and $H(x) > 0$ ($x \geq 0$) is an increasing continuous function such that $x^{-2-\gamma}H(x)$ is increasing for some $\gamma > 0$ and $x^{-1}\log H(x)$ is decreasing with $EH(|\xi_1|) < \infty$. Then there exist constants $C_1, C_2, a > 0$ which depend only on the distribution of ξ_1 and a sequence of Brownian motions $\{W_n(m)\}_{n=1}^{\infty}$, such that for any $\{x_n\}_{n=1}^{\infty}$ satisfying $H^{-1}(n) < x_n < C_1(n \log(n))^{1/2}$ and $S_m = \sum_{i=1}^m \xi_i$, then $P \left\{ \max_{1 \leq m \leq n} |S_m - W_n(m)| > x_n \right\} \leq C_2 n \{H(ax_n)\}^{-1}$.*

PROOF OF PROPOSITION 4.2.1. We first show (4.6). According to Theorem A.1 of Huang (2003), there exists an absolute constant $C > 0$, such that

$$\|\tilde{m} - m\|_{\infty} \leq C \inf_{g \in C^{p,1}} \|g - m\|_{\infty} \leq C \|m^{(p)}\|_{\infty} h_m^p. \quad (4.14)$$

Applying Lemma 4.4.2, for $0 \leq \nu \leq p - 2$,

$$\left\| \{Q_{\varpi}(m) - m\}^{(\nu)} \right\|_{\infty} \leq C \|m^{(p)}\|_{\infty} h_m^{p-\nu}. \quad (4.15)$$

As a consequence of (4.14) and (4.15) if $\nu = 0$, one has $\|Q_{\varpi}(m) - \tilde{m}\|_{\infty} \leq C \|m^{(p)}\|_{\infty} h_m^p$,

which, according to the differentiation of B-spline given in de Boor (2001), entails that

$$\left\| \{Q_{\varpi}(m) - \tilde{m}\}^{(\nu)} \right\|_{\infty} \leq C \left\| m^{(p)} \right\|_{\infty} h_m^{p-\nu}, \quad (4.16)$$

for $0 \leq \nu \leq p-2$. Combining (4.15) and (4.16) proves (4.6) for $\nu = 1, \dots, p-2$.

Next we prove (4.7). Similar to the definition of $\tilde{m}^{(\nu)}(t)$ and $\tilde{\xi}_k^{(\nu)}(t)$, in the following we denote $\tilde{\phi}_k^{(\nu)}(t) = \Gamma^{(\nu)}(t) \phi_k$, for any $k \geq 1$. Using the similar arguments as in the proof of (4.6), we can show that, for any $k \geq 1$,

$$\left\| \phi_k^{(\nu)} - \tilde{\phi}_k^{(\nu)} \right\|_{\infty} \leq C_{\phi} h_m^{p-\nu}. \quad (4.17)$$

Also, according to triangle inequality one has that $\left\| \tilde{\phi}_k^{(\nu)} \right\|_{\infty} \leq c_{\phi} \left\| \phi_k^{(\nu)} \right\|_{\infty} = O(1)$.

According to Assumption (C6), $E|\xi_{ik}|^{4+\delta_1} < +\infty$, $\delta_1 > 0$, so there exists some $\beta \in (0, 1/2)$, such that $4 + \delta_1 > 2/\beta$. Let $H(x) = x^{4+\delta_1}$, $x_n = n^{\beta}$, then $\frac{n}{H(ax_n)} = a^{-4-\delta_1} n^{1-(4+\delta_1)\beta} = O(n^{-\gamma_1})$ for some $\gamma_1 > 1$. Applying Lemma 4.4.4 and Borel-Cantelli Lemma, one finds i.i.d. variables $Z_{ik,\xi} \sim N(0, 1)$ such that

$$\max_{k \geq 1} \left| \bar{\xi}_{\cdot k} - \bar{Z}_{\cdot k, \xi} \sqrt{\lambda_k} \right| = O_{\text{a.s.}}(n^{\beta-1}), \quad (4.18)$$

where $\bar{Z}_{\cdot k, \xi} = n^{-1} \sum_{i=1}^n Z_{ik, \xi}$, $k \geq 1$.

If κ is finite, according to (4.18) note that $|\bar{\xi}_{\cdot k}| \leq \left| \bar{Z}_{\cdot k, \xi} \right| \sqrt{\lambda_k} + \left| \bar{\xi}_{\cdot k} - \bar{Z}_{\cdot k, \xi} \sqrt{\lambda_k} \right|$, $1 \leq k \leq \kappa$, so $\max_{1 \leq k \leq \kappa} |\bar{\xi}_{\cdot k}| = O_P(n^{-1/2} + n^{\beta-1})$. Then the definition of $\bar{m}(t)$ in (4.4)

and (4.17) entail that

$$\left\| \bar{m}^{(\nu)} - m^{(\nu)} - \tilde{\xi}^{(\nu)} \right\|_{\infty} \leq \kappa \max_{1 \leq k \leq \kappa} |\bar{\xi}_{\cdot k}| \max_{1 \leq k \leq \kappa} \left\| \phi_k^{(\nu)} - \tilde{\phi}_k^{(\nu)} \right\|_{\infty} = o_P \left(n^{-1/2} \right).$$

Thus (4.8) holds according to Assumption (C3).

If $\kappa = \infty$, using similar arguments in Cao, Yang and Todem (2012), by (4.18) one can show that $|\bar{\xi}_{\cdot k}| \lambda_k^{-1/2} \leq |\bar{Z}_{\cdot k, \xi}| + \left| \bar{\xi}_{\cdot k} \lambda_k^{-1/2} - \bar{Z}_{\cdot k, \xi} \right|$, for any $k \geq 1$, so $|\bar{\xi}_{\cdot k}| \lambda_k^{-1/2} = O_P(n^{-1/2} + n^{\beta-1})$. Also following Assumption (C5) one has

$$\begin{aligned} & E \sup_{t \in [0,1]} \left| \bar{m}^{(\nu)}(t) - m^{(\nu)}(t) - \tilde{\xi}^{(\nu)}(t) \right| \\ & \leq \sum_{k=1}^{\kappa} E |\bar{\xi}_{\cdot k}| \sup_{t \in [0,1]} \left| \phi_k^{(\nu)}(t) - \tilde{\phi}_k^{(\nu)}(t) \right| \leq C \sum_{k=1}^{\kappa} E |\bar{\xi}_{\cdot k}| \sup_{t \in [0,1]} \left| \phi_k^{(\nu)}(t) \right| \\ & \leq C \left\{ \sum_{k=1}^{\kappa n} \left(E |\bar{\xi}_{\cdot k}| \lambda_k^{-1/2} \right) \lambda_k^{1/2} \left\| \phi_k^{(\nu)} \right\|_{0, \delta} h_m^{\delta} \right. \\ & \quad \left. + \sum_{k=\kappa n+1}^{\kappa} \left(E |\bar{\xi}_{\cdot k}| \lambda_k^{-1/2} \right) \lambda_k^{1/2} \left\| \phi_k^{(\nu)} \right\|_{\infty} \right\} \\ & \leq C n^{-1/2} \left\{ \sum_{k=1}^{\kappa n} \lambda_k^{1/2} \left\| \phi_k^{(\nu)} \right\|_{0, \delta} h_m^{\delta} + \sum_{k=\kappa n+1}^{\kappa} \lambda_k^{1/2} \left\| \phi_k^{(\nu)} \right\|_{\infty} \right\} = o \left(n^{-1/2} \right). \end{aligned}$$

Hence, $\left\| \bar{m}^{(\nu)} - m^{(\nu)} - \tilde{\xi}^{(\nu)} \right\|_{\infty} = o_P \left(n^{-1/2} \right)$. According to Lemmas 4.4.1 and 4.4.3,

finally we have

$$\begin{aligned} \left\| \tilde{e}^{(\nu)} \right\|_{\infty} &= \left\| \mathbf{B}_p^{(\nu)}(t) \hat{\mathbf{V}}_p^{-1} N^{-1} \mathbf{B}^T \mathbf{e} \right\|_{\infty} \leq C h_m^{-\nu} \left\| \hat{\mathbf{V}}_p^{-1} \right\|_{\infty} \left\| N^{-1} \mathbf{B}^T \mathbf{e} \right\|_{\infty} \\ &= O_{\text{a.s.}} \left(n^{-1/2} N^{-1/2} h_m^{-1/2-\nu} \sqrt{\log(n)} \right). \end{aligned}$$

Thus (4.8) holds according to Assumption (C3).

Proof of Theorem 4.3.1

We denote $\tilde{\zeta}_k(t) = \sqrt{\lambda_k} \bar{Z}_{\cdot k, \xi} \phi_k^{(\nu)}(t)$, $k = 1, \dots, \kappa$ and define

$$\tilde{\zeta}(t) = n^{1/2} \left[\sum_{k=1}^{\kappa} \lambda_k \left\{ \phi_k^{(\nu)}(t) \right\}^2 \right]^{-1/2} \sum_{k=1}^{\kappa} \tilde{\zeta}_k(t) = n^{1/2} \Sigma(t, t)^{-1/2} \sum_{k=1}^{\kappa} \tilde{\zeta}_k(t).$$

It is clear that, for any $t \in [0, 1]$, $\tilde{\zeta}(t)$ is Gaussian with mean 0 and variance 1, and the covariance $E \tilde{\zeta}(t) \tilde{\zeta}(s) = \Sigma(t, t)^{-1/2} \Sigma(s, s)^{-1/2} \Sigma(t, s)$, for any $t, s \in [0, 1]$. That is, the distribution of $\tilde{\zeta}(t)$, $t \in [0, 1]$ and the distribution of $\zeta(t)$, $t \in [0, 1]$ in Section 3.1 are identical. Similar to the proof of (4.7), Note that

$$\begin{aligned} & E \sup_{t \in [0, 1]} \left| \tilde{\zeta}(t) - n^{1/2} \Sigma(t, t)^{-1/2} \tilde{\xi}^{(\nu)}(t) \right| \\ &= n^{1/2} E \sup_{t \in [0, 1]} \Sigma(t, t)^{-1/2} \left| \sum_{k=1}^{\kappa} \tilde{\zeta}_k(t) - \tilde{\xi}^{(\nu)}(t) \right| \\ &\leq n^{1/2} E \sup_{t \in [0, 1]} \Sigma(t, t)^{-1/2} \sum_{k=1}^{\kappa} \left(\left| \bar{Z}_{\cdot k, \xi} \sqrt{\lambda_k} - \bar{\xi}_{\cdot k} \right| \left| \phi_k^{(\nu)}(t) \right| + \left| \bar{\xi}_{\cdot k} \right| \left| \phi_k^{(\nu)}(t) - \tilde{\phi}_k^{(\nu)}(t) \right| \right) \end{aligned}$$

If κ is finite, then $E \sup_{t \in [0,1]} \left| \tilde{\zeta}(t) - n^{1/2} \Sigma(t, t)^{-1/2} \tilde{\xi}^{(\nu)}(t) \right| = O \left(n^{\beta-1/2} + h_m^{p-\nu} \right) = o(1)$. If $\kappa = \infty$, by Assumption (C5) one has

$$\begin{aligned}
& n^{1/2} E \sup_{t \in [0,1]} \Sigma(t, t)^{-1/2} \sum_{k=1}^{\kappa} \left(\left| \sqrt{\lambda_k} \bar{Z}_{\cdot k, \xi} - \bar{\xi}_{\cdot k} \right| \left| \phi_k^{(\nu)}(t) \right| + \left| \bar{\xi}_{\cdot k} \right| \left| \phi_k^{(\nu)}(t) - \tilde{\phi}_k^{(\nu)}(t) \right| \right) \\
& \leq n^{1/2} \sup_{t \in [0,1]} \Sigma(t, t)^{-1/2} \\
& \quad \left\{ n^{\beta-1} \sum_{k=1}^{\kappa} \sqrt{\lambda_k} \left| \phi_k^{(\nu)}(t) \right| + n^{-1/2} \sum_{k=1}^{\kappa} \sqrt{\lambda_k} \left| \phi_k^{(\nu)}(t) - \tilde{\phi}_k^{(\nu)}(t) \right| \right\} \\
& = o(1).
\end{aligned}$$

Theorem 4.3.1 follows directly.

Proof of Theorem 4.3.3

Following Cao, Wang, Li and Yang (2012), we define the tensor product spline space as

$$\begin{aligned}
\mathcal{H}^{(p-2),2}_{[0,1]^2} & \equiv \mathcal{H}^{(p-2),2} = \mathcal{H}^{(p-2)} \otimes \mathcal{H}^{(p-2)} \\
& = \left\{ \sum_{J, J'=1-p}^{N_G} b_{JJ'} B_{J,p}(t) B_{J',p}(s), \quad t, s \in [0, 1] \right\}.
\end{aligned}$$

Let $R_{ij} \equiv Y_{ij} - m(j/N)$, $1 \leq i \leq n$, $1 \leq j \leq N$ and $\bar{R}_{\cdot jj'} = n^{-1} \sum_{i=1}^n R_{ij} R_{ij'}$, $1 \leq j, j' \leq N$. Note that the spline estimator in (4.9) is

$$\hat{G}(\cdot, \cdot) = \arg \min_{g(\cdot, \cdot) \in \mathcal{H}^{(p-2),2}} \sum_{j \neq j'}^N \left\{ \hat{R}_{\cdot jj'} - g(j/N, j'/N) \right\}^2,$$

so we define the “infeasible estimator” of the covariance function

$$\tilde{G}(\cdot, \cdot) = \operatorname{argmin}_{g(\cdot, \cdot) \in \mathcal{H}^{(p-2), 2}} \sum_{1 \leq j \neq j' \leq N} \left\{ \bar{R}_{\cdot jj'} - g\left(j/N, j'/N\right) \right\}^2. \quad (4.19)$$

Denote $\mathbf{X} = \mathbf{B} \otimes \mathbf{B}$, $\mathbf{B}_p(t, s) = \mathbf{B}_p(t) \otimes \mathbf{B}_p(s)$ and $\mathbf{B}_p^{(0, \nu)}(t, s) = \mathbf{B}_p(t) \otimes \mathbf{B}_p^{(\nu)}(s)$. Let vector $\bar{\mathbf{R}} = \left\{ \bar{R}_{\cdot jj'} \right\}_{1 \leq j, j' \leq N}$, then we have

$$\tilde{G}^{(0, \nu)}(t, s) \equiv \frac{\partial^\nu}{\partial s^\nu} \hat{G}(t, s) = \mathbf{B}_p^{(0, \nu)}(t, s) \left(\mathbf{X}^T \mathbf{X} \right)^{-1} \mathbf{X}^T \bar{\mathbf{R}}. \quad (4.20)$$

In the following we write $\phi_{kk'}(t, s) = \phi_k(t) \phi_{k'}(s)$, $\phi_{kk'}^{(0, \nu)}(t, s) = \phi_k(t) \phi_{k'}^{(\nu)}(s)$. Let $\phi_{kk'} = \phi_k \otimes \phi_{k'}$, where ϕ_k is a N -dimensional vector defined in Section 2.3, $\tilde{\phi}_{kk'}(t, s) = \mathbf{B}_p(t, s) \left(\mathbf{X}^T \mathbf{X} \right)^{-1} \mathbf{X}^T \phi_{kk'}$. Further we denote $\tilde{\phi}_{kk'}^{(0, \nu)}(t, s) = \mathbf{B}_p^{(0, \nu)}(t, s) \left(\mathbf{X}^T \mathbf{X} \right)^{-1} \mathbf{X}^T \phi_{kk'}$.

Lemma 4.4.5. *Under Assumptions (C5), for any $1 \leq \nu \leq p-2$, and $\kappa = \infty$, one has*

$$\sum_{k, k' \geq 1} \sqrt{\lambda_k \lambda_{k'}} \left\| \phi_{kk'}^{(0, \nu)} - \tilde{\phi}_{kk'}^{(0, \nu)} \right\|_\infty = o(1).$$

PROOF. When $\kappa = \infty$, one has $\lambda_k > 0$ for any $k \geq 1$. Note that

$$\sum_{k, k' \geq 1} \sqrt{\lambda_k \lambda_{k'}} \left\| \phi_{kk'}^{(0, \nu)} \right\|_\infty \leq \sum_{k \geq 1} \sqrt{\lambda_k} \|\phi_k\|_\infty \sum_{k' \geq 1} \sqrt{\lambda_{k'}} \left\| \phi_{k'}^{(\nu)} \right\|_\infty = o(1).$$

Also similarly,

$$\begin{aligned}
\sum_{k,k' \geq 1} \sqrt{\lambda_k \lambda_{k'}} \left\| \tilde{\phi}_{kk'}^{(0,\nu)} \right\|_{\infty} &\leq \sum_{k \geq 1} \sqrt{\lambda_k} \left\| \tilde{\phi}_k \right\|_{\infty} \sum_{k' \geq 1} \sqrt{\lambda_{k'}} \left\| \tilde{\phi}_{k'}^{(\nu)} \right\|_{\infty} \\
&\leq C \sum_{k \geq 1} \sqrt{\lambda_k} \left\| \phi_k \right\|_{\infty} \sum_{k' \geq 1} \sqrt{\lambda_{k'}} \left\| \phi_{k'}^{(\nu)} \right\|_{\infty} = o(1).
\end{aligned}$$

Lemma 4.4.6. *Under Assumption (C5), for any $0 \leq \nu \leq p-2$ and $k, k' \geq 1$, one has*

$$\left\| \phi_{kk'}^{(0,\nu)} - \tilde{\phi}_{kk'}^{(0,\nu)} \right\|_{\infty} = O\left(h_G^{p-\nu}\right).$$

PROOF. According to Theorem 12.8 of Schumaker (2007), there exists an absolute constant $C > 0$, such that

$$\left\| \tilde{\phi}_{kk'} - \phi_{kk'} \right\|_{\infty} \leq C \left\| \phi_k^{(p)} \phi_{k'} + \phi_k \phi_{k'}^{(p)} \right\|_{\infty} h_G^p,$$

which proves (4.6) for the case $\nu = 0$. Let $Q(f)$ the p -th order quasi-interpolant of a function f ; see the definition in (12.29) of Schumaker (2007), for $0 \leq \nu \leq p-2$,

$$\left\| \left\{ Q(\phi_{kk'}) - \phi_{kk'} \right\}^{(0,\nu)} \right\|_{\infty} \leq C \left\| \phi_k^{(p)} \phi_{k'} + \phi_k \phi_{k'}^{(p)} \right\|_{\infty} h_G^{p-\nu}.$$

For the case $\nu = 0$, one has

$$\begin{aligned}
\left\| Q(\phi_{kk'}) - \tilde{\phi}_{kk'} \right\|_{\infty} &= \left\| Q(\phi_{kk'}) - Q(\tilde{\phi}_{kk'}) \right\|_{\infty} \leq C \left\| \tilde{\phi}_{kk'} - \phi_{kk'} \right\|_{\infty} \\
&\leq C \left\| \phi_k^{(p)} \phi_{k'} + \phi_k \phi_{k'}^{(p)} \right\|_{\infty} h_G^p,
\end{aligned}$$

which, according to the differentiation of B-spline given in de Boor (2001), entails that

$$\left\| \left\{ Q(\phi_{kk'}) - \tilde{\phi}_{kk'} \right\}^{(0,\nu)} \right\|_{\infty} \leq C \left\| \phi_k^{(p)} \phi_{k'} + \phi_k \phi_{k'}^{(p)} \right\|_{\infty} h_G^{p-\nu}, \text{ for } 0 \leq \nu \leq p-2.$$

Lemma 4.4.7. *Under Assumptions (C1)-(C6), for any $1 \leq \nu \leq p-2$*

$$\left\| \tilde{G}^{(0,\nu)} - G^{(0,\nu)} \right\|_{\infty} = O_P \left(n^{-1/2} + h_G^{p-\nu} \right) + o_P \left(N^{-1} n^{-1/2} h_G^{-1-\nu} \log^{1/2} n \right), \quad (4.21)$$

$$\left\| \hat{G}^{(0,\nu)} - \tilde{G}^{(0,\nu)} \right\|_{\infty} = O_P \left(n^{-1} h_G^{-3/2-\nu} + n^{-1/2} h_G^{-1-\nu} h_m^p \right), \quad (4.22)$$

where $\hat{G}^{(0,\nu)}$ and $\tilde{G}^{(0,\nu)}$ are given in (4.10) and (4.20).

PROOF. Let $\bar{\xi}_{\cdot k k'} = n^{-1} \sum_{i=1}^n \xi_{ik} \xi_{ik'}$ and $\bar{\varepsilon}_{\cdot j j'} = n^{-1} \sum_{i=1}^n \varepsilon_{ij} \varepsilon_{ij'}$. To show (4.21), we decompose $\bar{R}_{\cdot j j'}$ in (4.19) into

$$\begin{aligned} \bar{R}_{1 j j'} &= \sum_{k \neq k'}^{\kappa} \bar{\xi}_{\cdot k k'} \phi_{k k'} \left(\frac{j}{N}, \frac{j'}{N} \right), \quad \bar{R}_{2 j j'} = \sum_{k=1}^{\kappa} \bar{\xi}_{\cdot k k} \phi_{k k} \left(\frac{j}{N}, \frac{j'}{N} \right), \\ \bar{R}_{3 j j'} &= \sigma \left(\frac{j}{N} \right) \sigma \left(\frac{j'}{N} \right) \bar{\varepsilon}_{\cdot j j'}, \end{aligned}$$

$$\bar{R}_{4 j j'} = n^{-1} \sum_{i=1}^n \left\{ \sum_{k=1}^{\kappa} \xi_{ik} \phi_k \left(\frac{j}{N} \right) \sigma \left(\frac{j'}{N} \right) \varepsilon_{ij'} + \sum_{k=1}^{\kappa} \xi_{ik} \phi_k \left(\frac{j'}{N} \right) \sigma \left(\frac{j}{N} \right) \varepsilon_{ij} \right\}.$$

Denote $\bar{\mathbf{R}}_i = \left\{ \bar{R}_{i j j'} \right\}_{1 \leq j, j' \leq N}$, $\tilde{\mathcal{R}}_i^{(0,\nu)}(t, s) = \mathbf{B}_p^{(0,\nu)}(t, s) \left(\mathbf{X}^T \mathbf{X} \right)^{-1} \mathbf{X}^T \bar{\mathbf{R}}_i$, $i = 1, 2, 3, 4$.

Then $\tilde{G}^{(0,\nu)}(t, s) = \tilde{\mathcal{R}}_1^{(0,\nu)}(t, s) + \tilde{\mathcal{R}}_2^{(0,\nu)}(t, s) + \tilde{\mathcal{R}}_3^{(0,\nu)}(t, s) + \tilde{\mathcal{R}}_4^{(0,\nu)}(t, s)$. Next we define

$$\begin{aligned} \mathcal{R}_1^{(0,\nu)}(t, s) &= \sum_{k \neq k'}^{\kappa} \bar{\xi}_{\cdot k k'} \phi_{k k'}^{(0,\nu)}(t, s), \\ \mathcal{R}_2^{(0,\nu)}(t, s) &= G^{(0,\nu)}(t, s) + \sum_{k=1}^{\kappa} \left\{ \phi_{k k}^{(0,\nu)}(t, s) (\bar{\xi}_{\cdot k k} - \lambda_k) \right\}. \end{aligned}$$

Note that $\tilde{\mathcal{R}}_1^{(0,\nu)}(t, s) = \sum_{k \neq k'}^{\kappa} \bar{\xi}_{\cdot k k'} \tilde{\phi}_{k k'}^{(0,\nu)}(t, s)$, then Lemma 4.4.5 and Assumption (C5)

imply that if κ is ∞ , $\lambda_k \lambda_{k'} > 0$ and one has

$$\begin{aligned} & \sup_{(t,s) \in [0,1]^2} E \left| \tilde{\mathcal{R}}_1^{(0,\nu)}(t,s) - \mathcal{R}_1^{(0,\nu)}(t,s) \right| \\ & \leq \sum_{k \neq k'}^{\kappa} E \left| \bar{\xi}_{\cdot k k'} \left(\lambda_k \lambda_{k'} \right)^{-1/2} \right| \sqrt{\lambda_k \lambda_{k'}} \left\| \phi_{k k'}^{(0,\nu)} - \tilde{\phi}_{k k'}^{(0,\nu)} \right\|_{\infty} = o(1). \end{aligned}$$

Similarly, one has

$$\begin{aligned} & \sup_{(t,s) \in [0,1]^2} E \left| \tilde{\mathcal{R}}_2^{(0,\nu)}(t,s) - \mathcal{R}_2^{(0,\nu)}(t,s) \right| \\ & \leq \sum_{k=1}^{\kappa} E \left| \bar{\xi}_{\cdot k k} \lambda_k^{-1} - 1 \right| \lambda_k \left\| \phi_{k k}^{(0,\nu)} - \tilde{\phi}_{k k}^{(0,\nu)} \right\|_{\infty} = o(1). \end{aligned}$$

If κ is finite, then Lemma A.6 and Assumption (C3) imply that

$$\left\| \tilde{\mathcal{R}}_1^{(0,\nu)} - \mathcal{R}_1^{(0,\nu)} \right\|_{\infty} \leq \kappa^2 \max_{1 \leq k \neq k' \leq \kappa} \left| \bar{\xi}_{\cdot k k'} \right| \left\| \phi_{k k'}^{(0,\nu)} - \tilde{\phi}_{k k'}^{(0,\nu)} \right\|_{\infty} = O_P \left(h_G^{p-\nu} n^{-1/2} \right).$$

Similarly,

$$\left\| \tilde{\mathcal{R}}_2^{(0,\nu)} - \mathcal{R}_2^{(0,\nu)} \right\|_{\infty} \leq \kappa \max_{1 \leq k \leq \kappa} \left| \bar{\xi}_{\cdot k k} - \lambda_k \right| \left\| \phi_{k k}^{(0,\nu)} - \tilde{\phi}_{k k}^{(0,\nu)} \right\|_{\infty} = O_P \left(h_G^{p-\nu} \right).$$

Hence, $\left\| \tilde{\mathcal{R}}_2^{(0,\nu)} - \mathcal{R}_2^{(0,\nu)} \right\|_{\infty} + \left\| \tilde{\mathcal{R}}_1^{(0,\nu)} - \mathcal{R}_1^{(0,\nu)} \right\|_{\infty} = o(1)$. By Proposition 3.1 in Cao,

Wang, Li and Yang (2012), one has

$$\left\| N^{-2} \mathbf{X}^T \bar{\mathbf{R}}_3 \right\|_{\infty} = \left\| N^{-2} \mathbf{X}^T \bar{\mathbf{R}}_4 \right\|_{\infty} = o_P \left(N^{-1} n^{-1/2} h_G \log^{1/2}(n) \right).$$

Hence, $\left\| \tilde{\mathcal{R}}_3^{(0,\nu)} \right\|_\infty = \left\| \tilde{\mathcal{R}}_4^{(0,\nu)} \right\|_\infty = o_P \left(N^{-1} n^{-1/2} h_G^{-1-\nu} \log^{1/2} n \right)$. Therefore,

$$\begin{aligned} \left\| \tilde{G}^{(0,\nu)} - G^{(0,\nu)} \right\|_\infty &\leq \sum_{k \neq k'}^{\kappa} \bar{\xi}_{\cdot, kk'} \left\| \phi_{kk'}^{(0,\nu)} \right\|_\infty + \sum_{k=1}^{\kappa} (\bar{\xi}_{\cdot, kk} - \lambda_k) \left\| \phi_{kk}^{(0,\nu)} \right\|_\infty \\ &+ \left\| \tilde{\mathcal{R}}_1^{(0,\nu)} - \mathcal{R}_1^{(0,\nu)} \right\|_\infty + \left\| \tilde{\mathcal{R}}_2^{(0,\nu)} - \mathcal{R}_2^{(0,\nu)} \right\|_\infty + \left\| \tilde{\mathcal{R}}_3^{(0,\nu)} \right\|_\infty + \left\| \tilde{\mathcal{R}}_4^{(0,\nu)} \right\|_\infty \\ &= O_P \left(n^{-1/2} + h_G^{p-\nu} \right) + o_P \left(N^{-1} n^{-1/2} h_G^{-1-\nu} \log^{1/2} n \right). \end{aligned}$$

The proof of (4.22) is similar to Proposition 2.1 in Cao, Wang, Li and Yang (2012), thus omitted.

PROOF OF THEOREM 4.3.3: According to Lemma 4.4.7, one has

$$\left\| \hat{G}^{(0,\nu)} - G^{(0,\nu)} \right\|_\infty \leq \left\| \tilde{G}^{(0,\nu)} - G^{(0,\nu)} \right\|_\infty + \left\| \hat{G}^{(0,\nu)} - \tilde{G}^{(0,\nu)} \right\|_\infty = o_P(1).$$

Proof of Theorem 4.3.4

We first show asymptotic consistency of $\hat{\lambda}_k$ and $\hat{\phi}_k(\cdot)$, for $k \geq 1$, in the following lemma.

Lemma 4.4.8. *Under Assumptions (C1)-(C6), one has*

$$\left| \hat{\lambda}_k - \lambda_k \right| = o_P(1), \quad \left\| \hat{\phi}_k - \phi_k \right\|_\infty = o_P(1), \quad k \geq 1.$$

PROOF. We first want to show that for any $k \geq 1$, $\left\| \Delta \phi_k \right\|_\infty = o_P(1)$, in which Δ is the integral operator with kernel $\hat{G} - G$. Note that $(\Delta \phi_k)(t) = \int (\hat{G} - G)(s, t) \phi_k(s) ds$. By Theorem 4.3.3, when $\nu = 0$, $\left\| \hat{G} - G \right\|_\infty = o_P(1)$. Thus, for any $k \geq 1$, $\left\| \Delta \phi_k \right\|_\infty = o_P(1)$.

Hall and Hosseini-Nasab (2006) gives the L^2 expansion

$$\hat{\phi}_k - \phi_k = \sum_{k' \neq k} (\lambda_k - \lambda_{k'})^{-1} \langle \Delta \phi_k, \phi_{k'} \rangle \phi_{k'} + O(\|\Delta\|_2^2),$$

where $\|\Delta\|_2 = \left[\int \int \left\{ \hat{G}(s, t) - G(s, t) \right\}^2 ds dt \right]^{1/2}$. By Bessel's inequality, one has $\|\hat{\phi}_k - \phi_k\|_2 \leq C \left(\|\Delta \phi_k\|_\infty^2 + \|\Delta\|_2^2 \right) = o_P(1)$. By (4.9) in Hall, Müller and Wang (2006) and Theorem 4.3.3

$$\hat{\lambda}_k - \lambda_k = \int \int (\hat{G} - G)(s, t) \phi_k(s) \phi_k(t) ds dt + O(\|\Delta \phi_k\|_2^2) = o_P(1).$$

Thus $|\hat{\lambda}_k - \lambda_k| = o_P(1)$ for any $k \geq 1$. Next note that

$$\begin{aligned} \hat{\lambda}_k \hat{\phi}_k(t) - \lambda_k \phi_k(t) &= \int \hat{G}(s, t) \hat{\phi}_k(s) ds - \int G(s, t) \phi_k(s) ds \\ &= \int (\hat{G} - G)(s, t) (\hat{\phi}_k(s) - \phi_k(s)) ds + \int (\hat{G} - G)(s, t) \phi_k(s) ds \\ &\quad + \int G(s, t) \{ \hat{\phi}_k(s) - \phi_k(s) \} ds. \end{aligned}$$

By the Cauchy-Schwarz inequality, uniformly for all $t \in [0, 1]$

$$\int G(s, t) (\hat{\phi}_k(s) - \phi_k(s)) ds \leq \left(\int G^2(s, t) ds \right)^{1/2} \|\hat{\phi}_k - \phi_k\|_2 = o_P(1).$$

Similar arguments and Theorem 4.3.3 imply that $\int (\hat{G} - G)(s, t) (\hat{\phi}_k(s) - \phi_k(s)) ds = o_P(1)$ and $\int (\hat{G} - G)(s, t) \phi_k(s) ds = o_P(1)$. All the above together yield $\|\hat{\lambda}_k \hat{\phi}_k - \lambda_k \phi_k\|_\infty =$

$o_P(1)$. By the triangle inequality and

$$\lambda_k \left\| \hat{\phi}_k - \phi_k \right\|_{\infty} \leq \left\| \hat{\lambda}_k \hat{\phi}_k - \lambda_k \phi_k \right\|_{\infty} + \left| \hat{\lambda}_k - \lambda_k \right| \left\| \hat{\phi}_k \right\|_{\infty} = o_P(1),$$

the second result in Lemma 4.4.8 follows directly.

PROOF OF THEOREM 4.3.4. According to Lemma 4.4.8 and Theorem 4.3.3, one has

$$\begin{aligned} & \left| \hat{\phi}_k^{(\nu)}(s) - \phi_k^{(\nu)}(s) \right| = \left| \hat{\lambda}_k^{-1} \int_0^1 \hat{G}^{(0,\nu)}(t, s) \hat{\phi}_k(t) dt - \lambda_k^{-1} \int_0^1 G^{(0,\nu)}(t, s) \phi_k(t) dt \right| \\ & \leq \left| \hat{\lambda}_k^{-1} - \lambda_k^{-1} \right| \times \int_0^1 \left| \hat{G}^{(0,\nu)}(t, s) \hat{\phi}_k(t) \right| dt \\ & \quad + \lambda_k^{-1} \int_0^1 \left| \hat{G}^{(0,\nu)}(t, s) \hat{\phi}_k(t) - G^{(0,\nu)}(t, s) \phi_k(t) \right| dt. \end{aligned}$$

Hence, $\sup_{s \in [0,1]} \left\{ \int_0^1 \left| \left(\hat{G}^{(0,\nu)} - G^{(0,\nu)} \right)(t, s) \hat{\phi}_k(t) + G(t, s) \left(\hat{\phi}_k - \phi_k \right)(t) \right| dt \right\} = o_P(1)$
and $\left\| \hat{\phi}_k^{(\nu)} - \phi_k^{(\nu)} \right\|_{\infty} = o_P(1)$. Theorem 4.3.4 follows from the definition of $\Sigma(t, s)$.

Chapter 5

Testing Hypotheses Under Weakly Identifiability

5.1 Introduction

Data in statistical research are often well described by models, in which the scientific questions of interest are described by an unknown, finite-dimensional parameter vector. Such models may be either fully parametric or semiparametric, where other aspects of the model may be described by infinite dimensional parameters which are completely unspecified. In such settings, it is often of interest to use the observed data in order to draw inferences about the parameters of interest. Standard inferential techniques may be applied if the parameters of interest can be well estimated by minimizing a parametric loss function or more generally by solving a parametric estimating function which does not involve infinite dimensional nuisance parameters. In many situations, however, these parameters may be nonidentifiable or at best weakly identifiable from the estimating function so that the standard inferential theories may not be valid. The goal of this paper is to develop hypothesis tests for scenarios

in which the model parameters are weakly identifiable. Conceptually, the term weak identifiability refers to the situations where data contain some information about model parameters but not enough to identify them uniquely.

To illustrate the problem quite sharply, we consider a simple theoretical example where a fully parametric model is indexed by an unknown parameter vector (θ, β) for an observable random quantity Y . We assume that realizations $\{Y_i\}_{i=1}^n$ of Y are independent and identically distributed (i.i.d) normal $\mathcal{N}(\theta + \beta, 1)$ variates. The objective is to evaluate the hypothesis $H_0 : \theta_0 = 0$, where θ_0 is the true value of θ . Using only observed data and assuming that β_0 , the true value of β is unknown, inferences for θ_0 may not be conducted using standard techniques due to identifiability problems arising from the mean model being overparameterized.

Another interesting, more practical illustration of this problem comes from the missing data literature where weakly identifiable models are frequently encountered. Specific examples include the study of publication bias in meta-analysis (Chambers and Welsh, 1993; Copas, 1999; Copas and Li, 1997) and the analysis of longitudinal data subject to non-random nonresponses (Scharfstein et al., 1999; Kenward et al., 2001; Rotnitzky et al., 2001; Little and Rubin, 2002). Identifiability issues commonly arise with non-random missing data, where the parameters in the model for the missingness may not be jointly identifiable with those in the model for the outcomes of interest using only the observed data, particularly with semiparametric models, where some of the nuisance parameters may be infinite dimensional. Analyses which assume identifiability may be unreliable, with the joint selection and outcome model yielding flat “estimation” surfaces potentially having multiple modes. These phenomena have previously been reported by several authors in modeling potentially non-ignorable missing data models (Scharfstein et al., 1999; Todem et al., 2010) .

In section 3, we consider these missing data issues when analyzing longitudinal data with informative dropout employing the model of Troxel et al.(1998b). The model is semiparametric, with the parameter being estimated denoted by (θ, β) , where β is the selection parameter that measures the extent of non-randomness of the missing data mechanism and θ consists of the remaining finite dimensional parameters of the selection and outcome models. The hypotheses of interest concern covariate effects on the outcome, which are contained in θ . In Troxel et al. (1998b), a so-called pseudo-likelihood analysis, described in detail in Section 3, was carried out under the assumption of parameter identifiability. The resulting estimating function only involves (θ, β) , with the longitudinal dependence in the outcomes completely unspecified and not estimated. We investigated the parameter identifiability assumption in a reanalysis of the cancer data from Troxel et al. (1998b) by profiling the pseudo-likelihood analysis in β (Figure 5.1). The profile pseudolikelihood is flat in β , suggesting a model that is at best weakly identifiable. These results draw into question inferences which assume identifiability of θ and β .

Due to identifiability concerns, tests concerning the model parameters cannot use conventional theory to assess statistical significance. Essentially, the standard estimation and inference techniques may fail due to the models being overparameterized. A natural remedy is to partition the parameter indexing the estimating function into certain parameters of interest and other parameters which may be viewed as secondary parameters. For the theoretical example discussed earlier, where Y is normally distributed, the parameter of interest in light of the hypothesis under study is θ , while β is the secondary parameter. In the missing data application (Troxel et al., 1998b), the parameter β which describes the informativeness may be viewed as the secondary parameter, while the covariate effects in θ may be of primary interest in hypothesis testing. In practice, the choice of θ and β will

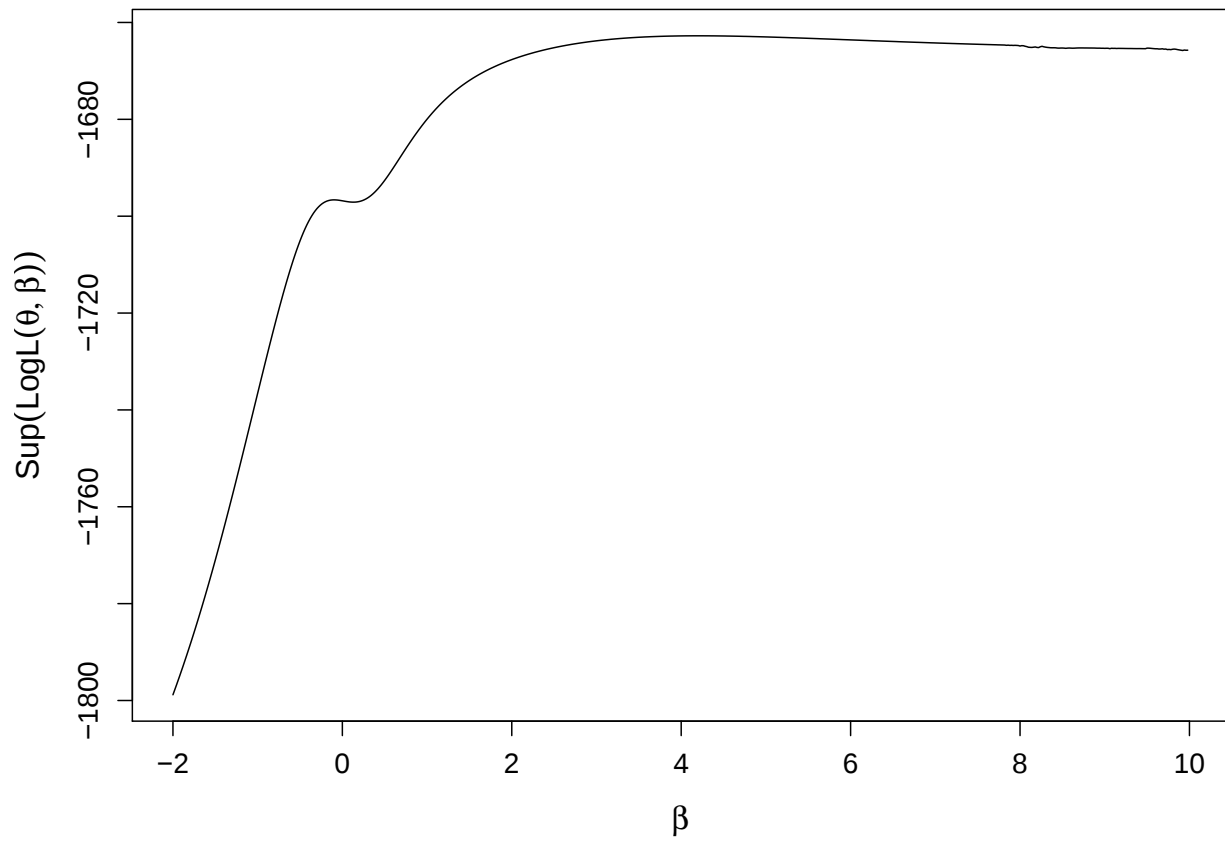


Figure 5.1: Supremum of the pseudo-likelihood function profiled across β , the parameter measuring the extent of non-randomness of the missing data mechanism in the study.

depend on the application.

Various approaches to the problem of non identifiable parameters that have appeared in the literature focused primarily on maximum likelihood based procedures. Almost all previous works in hypothesis testing deal with the case where non identifiability only occurs under the null hypothesis. Examples include Davies (1977, 1987), Hansen (1996), Ritz and Skovgaard (2005) and Song et al. (2009). Generally, this requires that the model is identifiable under the alternative hypothesis. In sensitivity analysis, the testing problem has a different formulation. The model may not be identifiable under either the null or the alternative hypothesis. Moreover, even after fixing a set of parameters, it may not be clear whether the parameters of interest can be consistently estimated under the null hypothesis. To be concrete, in the normal example, for each value of β , the maximum likelihood estimator of θ consistently estimates $\theta_0 + \beta - \beta_0$, where θ_0 and β_0 are the true values of θ and β . This only equals θ_0 when $\beta = \beta_0$. Our approach to inference about the parameters of interest is to adapt the profiling strategy from the earlier works described above. Because the testing problem is fundamentally different, the resulting developments are nonstandard, with relatively little work in the literature on this problem. Since the model may not be identifiable even after profiling, we need to consider the behavior of the profile estimator under model misspecification under the null.

This inferential strategy poses substantial technical challenges beyond those encountered with supremum tests which assume identifiability under the alternative. In missing data applications used to motivate the sensitivity analysis, rigorous results for full likelihood analyses have been established (Lu and Copas, 2004), essentially requiring model identifiability. More recently, Todem et al. (2010) demonstrated how to conduct likelihood inference via infimum tests, including a precise analysis of the behavior of the profile estimators under

model misspecification and the distribution of the corresponding infimum test. Such tests are particularly important when the quantity being tested does not increase or decrease monotonically as the nonidentified parameters are increased or decreased. Under monotonicity, it is only necessary to perform the tests at the limits of the nonidentified parameter space. They developed simultaneous confidence bands which enable identification of those values of the sensitivity parameter for which significant results are obtained. Although these likelihood-based methods are useful, they require a full distribution specification for the data. This can be a difficult task in practice, especially when observed data do not have enough information to fully identify the parameter of interest.

In this paper, we extend the profiling idea to arbitrary estimating functions involving θ and β but which do not require a complete parametric model specification. Our set-up includes the likelihood score functions as a special case. The generalization of the infimum test and confidence bands to non-likelihood settings is nontrivial. The infimum test has the advantage that it is simply defined directly in terms of contrasts whereas the supremum tests are obtained through nontrivial derivations using the log-likelihood functions (Dacunha-Castelle and Gassiat, 1999). We present generic conditions which establish the large sample properties of the estimating function for θ profiled on β , including the uniform consistency and weak convergence of the θ estimator as a function of β . To our knowledge, these theoretical results are novel, with issues related to nonidentifiable estimating functions not having been studied rigorously, previously. We accommodate misspecification and uniformity in β in a general paradigm which permits the profiling to be carried out with respect to any suitable estimating function. Owing to the complexity of the asymptotic distributions of the infimum test and confidence bands, resampling is needed. A theoretically justified procedure is discussed for approximating such distributions.

The rest of this chapter is organized as follows. In Section 5.2, we present the general framework of the problem, the proposed test and the resampling procedure, along with a proof of the key asymptotic properties. In Section 5.3, the methodology is exhibited using the cancer dataset in Troxel et al. (1998b) and in simulations, where the naive Wald test may have either inflated type I error rate or reduced power. Some remaining issues are discussed in Section 5.4.

5.2 The method

5.2.1 The general framework

We consider a model involving a finite dimensional parameter $\varpi \in \Omega$ for an observable random quantity Y . The parameter ϖ may not completely determine the distribution of Y , that is, there may be other aspects of the model which are unspecified. The interest is drawing inferences about ϖ with i.i.d realizations $\{Y_i\}_{i=1}^n$ of Y and a general estimating function $S_Y(\varpi)$. Denote by ϖ_0 the true value of ϖ . If $E\{S_Y(\varpi_0)\} = 0$, then an estimator $\hat{\varpi}$ of ϖ_0 usually can be obtained by solving the estimating equation, $S_Y(\varpi) = 0$; see Chapter 5 of van der Vaart (2000b) for an overview of Z-estimators. If S_Y identifies ϖ_0 , then under other mild regularity conditions, this estimating equation yields a consistent and asymptotically normal parameter estimator. Under such regularity conditions, inferences about ϖ_0 can be conducted using the large sample properties of $\hat{\varpi}$. Problems may occur if the model as a function of ϖ is “overparameterized”, with multiple values of ϖ satisfying $E\{S_Y(\varpi)\} = 0$. In this case, the estimator may not have the usual asymptotic properties.

Nonidentifiability can be addressed by fixing some components of ϖ , conditional upon

which the remaining parameters are uniquely defined by S_Y . One may partition $\varpi = (\theta, \beta)$, where θ , a p -dimensional vector, is assumed to be “identifiable” for a fixed q -dimensional vector β , as defined in Section 2.2 below. If the true value β_0 of the nonidentified parameter β is known, the estimator $\hat{\theta}_0$ at $\beta = \beta_0$ can be used to conduct reliable inferences about θ_0 , the true value of θ . This estimator is readily available by solving the estimating equation $S_Y(\theta, \beta_0) = 0$, for fixed and known β_0 . The approach is unfeasible, as the true value β_0 is usually unknown to the analyst in practice. A common strategy is to fix β and study the estimator of θ at various values of $\beta \in \Xi$. To highlight the dependence on β , we denote by $\hat{\theta}(\beta)$, the estimator of θ for a fixed β . The estimator of θ when $\beta = \beta_0$ is $\hat{\theta}_0 = \hat{\theta}(\beta_0)$.

For the simple normal example, $\hat{\theta}(\beta) = \bar{Y} - \beta$ and $\hat{\theta}(\beta_0) = \bar{Y} - \beta_0$, where $\bar{Y}$ is the sample mean. This estimator is normally distributed with mean $\theta_0 + \beta_0 - \beta$ and variance n^{-1} , uniformly in β , for each fixed n . Of course, in general, it is not possible to obtain clean finite sample results and large sample approximations are needed. In the subsection below, we study the uniform asymptotic properties of $\hat{\theta}(\beta)$ for $\beta \in \Xi$.

5.2.2 Large sample properties of $\hat{\theta}(\beta)$

When β is fixed at its true value β_0 , it is well established that for an estimating function $S_Y(\theta, \beta_0)$ which is smooth in θ , the estimator $\hat{\theta}$ is consistent and approximately normal under mild regularity conditions (see, for example, van der Vaart and Wellner, 2000a). That is, $n^{\frac{1}{2}}\{\hat{\theta}(\beta_0) - \theta_0\} \rightarrow_d \mathcal{N}(0, \Sigma_0)$, where $\Sigma_0 = (D(\theta_0))^{-1} \text{var}(S_Y(\theta_0, \beta_0))(D^{-1}(\theta_0))^T$, with $D(\theta_0)$ being the expected value of the first order derivative of $S_Y(\theta, \beta_0)$ with respect to θ . These properties of $\hat{\theta}(\beta_0)$ can be used to conduct large-sample inferences about θ_0 .

For a given β , the estimator $\hat{\theta}(\beta)$ will converge to a quantity $\theta^*(\beta)$, which is generally

different from θ_0 if $\beta \neq \beta_0$. For the simple normal example, $\theta^*(\beta) = \theta_0 + \beta_0 - \beta$. This contrasts with set-ups on testing with nonidentifiability under the null (Davies, 1977, 1987), where it is generally assumed that $\theta^*(\beta) = \theta_0$ for all β . Moreover, appropriately standardized, $\hat{\theta}(\beta)$ will be asymptotically normal, with variance which may be estimated using a sandwich variance approach. This is an extension of standard pointwise asymptotic theory for maximum likelihood estimation with misspecified models, originating in the seminal work of Huber (1967) and White (1982). We study below the uniform convergence of this estimator across all values of $\beta \in \Xi$.

Suppose the data consist of iid realizations $\{Y_i\}_{i=1}^n$ of Y . Let $s_{Y_i}(\theta, \beta)$ be the contribution of subject i to the estimating function $S_Y(\theta, \beta)$. Define $S_n(\theta, \beta) = n^{-1} \sum_{i=1}^n s_{Y_i}(\theta, \beta)$ and $\tilde{S}(\theta, \beta) = E\{s_{Y_1}(\theta, \beta)\}$. Let $g_{Y_i}(\theta, \beta) = \partial s_{Y_i}(\theta, \beta) / \partial \theta$, $W_Y(\theta, \beta) = n^{-1} \sum_{i=1}^n g_{Y_i}(\theta, \beta)$ and $\tilde{W}(\theta, \beta) = E\{g_{Y_1}(\theta, \beta)\}$. For any given $\beta \in \Xi$, let $\hat{\theta}(\beta)$ denote the solution to $S_Y(\theta, \beta) = 0$, that is $S_Y(\hat{\theta}(\beta), \beta) = 0$. The “least false” (White, 1982) parameter $\theta^*(\beta)$, satisfies $\tilde{S}(\theta^*(\beta), \beta) = 0$. Define $\mathcal{G}_1 = \{s_{Y_i}(\theta, \beta) : i = 1, \dots, n, \theta \in \Theta, \beta \in \Xi\}$ and $\mathcal{G}_2 = \{g_{Y_i}(\theta, \beta) : i = 1, \dots, n, \theta \in \Theta, \beta \in \Xi\}$.

We assume the following regularity conditions:

- C1. The sets $\Theta \subset \mathbb{R}^p$ and $\Xi \subset \mathbb{R}^q$ are compact and $\theta^*(\beta)$ is an interior point of Θ for any $\beta \in \Xi$.
- C2. The function classes, $\mathcal{G}_1$ and $\mathcal{G}_2$, are pointwise measurable and satisfy the uniform entropy condition (van der Vaart and Wellner, 2000a).
- C3. $\inf_{\theta \in \Theta, \beta \in \Xi} \lambda_{\min}\{-\tilde{W}(\theta, \beta)\} > 0$, where $\lambda_{\min}(\cdot)$ denotes the minimum eigenvalue of a matrix.

C4. The estimating function $S_Y(\theta, \beta)$ has continuous first order derivatives with respect to θ for any given $\beta \in \Xi$.

Condition C1 defines the parameter space for the implied parameter $\theta^*(\beta)$ for a given β . Because $\theta^*(\beta)$ may be nonconstant in β , the parameter space for $\theta^*(\beta)$ across β is contained in a suitably defined functional space. Conditions C2 and C3 give conditions under which uniform asymptotic results for $\theta^*(\beta)$ may be obtained. The entropy condition C2 ensures that the estimating function is well behaved across all β . The condition is satisfied by functions which are uniformly bounded and uniformly Lipschitz of order $> \{\dim(\theta) + \dim(\beta)\}/2$, where $\dim(\cdot)$ denotes the dimension of a vector. Condition C3 guarantees the identifiability of $\theta^*(\beta)$ for all β . The longitudinal data model presented in Section 3 meets these requirements. Note that the smoothness specified in condition C4 only applies to θ . Differentiability in β is not assumed. Non-smoothness in θ could be accommodated under stronger assumptions.

The proof of the following theorem is provided in the appendix.

Theorem 5.2.1. *Under Conditions C1-C4, $\sup_{\beta \in \Xi} \|\hat{\theta}(\beta) - \theta^*(\beta)\| \rightarrow_p 0$, where $\|\cdot\|$ represents the Euclidean norm. Furthermore, $n^{\frac{1}{2}} \left(\hat{\theta}(\beta) - \theta^*(\beta) \right)$ converge weakly to a tight Gaussian process with positive definite covariance function*

$$\begin{aligned} \Sigma^*(\beta_1, \beta_2) &= \lim_{n \rightarrow \infty} \text{cov} \left\{ n^{\frac{1}{2}} \left(\hat{\theta}(\beta_1) - \theta^*(\beta_1) \right), n^{\frac{1}{2}} \left(\hat{\theta}(\beta_2) - \theta^*(\beta_2) \right) \right\} \\ &= \left[\left\{ \tilde{W}(\theta^*(\beta_1), \beta_1) \right\}^{-1} \right]^T E \left\{ s_{Y_1}(\theta, \beta_1) s_{Y_1}^T(\theta, \beta_2) \right\} \left\{ \tilde{W}(\theta^*(\beta_2), \beta_2) \right\}^{-1}. \end{aligned}$$

For fixed β ,

$$\begin{aligned}\Sigma^*(\beta, \beta) &= \lim_{n \rightarrow \infty} \text{var} \left\{ n^{\frac{1}{2}} \left(\hat{\theta}(\beta) - \theta^*(\beta) \right) \right\} \\ &= \left[\left\{ \tilde{W}(\theta^*(\beta), \beta) \right\}^{-1} \right]^T \text{E} \left\{ s_{Y_1}(\theta, \beta) s_{Y_1}^T(\theta, \beta) \right\} \left\{ \tilde{W}(\theta^*(\beta), \beta) \right\}^{-1}.\end{aligned}$$

The covariance function may be easily estimated using a robust sandwich variance estimator along the lines of White (1982), which is valid under model misspecification. This estimator may be used to construct pointwise confidence intervals for $\theta^*(\beta)$ at fixed β using the pointwise asymptotic normality of $\hat{\theta}(\beta)$. However, for the testing and confidence band procedures described below, the complexity of the limiting distribution across β is prohibitive for conducting inference, even with variance estimation. For such scenarios, we suggest resampling to approximate the distribution of the estimator.

It can easily be shown that the regularity conditions are satisfied for the simple normal example. Interestingly, $\hat{\theta}(\beta) - \theta^*(\beta) = \bar{Y} - \theta_0 - \beta_0$, which does not depend on β . This greatly simplifies the results of Theorem 1, since the standardized estimators are identical for all β , which is not generally true. One should note that the form of the mean model is critical. If we assumed that $E(Y) = \theta\beta$, then the eigenvalue condition, C3, would be violated at $\beta = 0$ and the uniform convergence in Theorem 1 would fail to hold on intervals containing zero.

5.2.3 Global sensitivity testing

Suppose we are interested in evaluating the null hypothesis: $\mathbf{H}_0 : C\theta_0 = c$, where θ_0 is the true value of θ and C an $r \times \dim(\theta_0)$ contrast matrix for assessing single and multiple

linear combinations of model parameters. For example, when testing the j th component of θ , one takes C to be $1 \times \dim(\theta)$ vector with a one at the j th position and zeros elsewhere. Under nonidentifiability, the above hypothesis cannot be tested without imposing unverifiable restrictions. If the true sensitivity parameter β_0 is known, then $\mathbf{H}_0 : C\theta^*(\beta_0) = c$, where $\theta^*(\beta_0) = \theta_0$.

In practice, where β_0 is unknown, one may consider the process $\theta^*(\beta)$, observing that the trivial inequality,

$$0 \leq \inf_{\beta \in \Xi} \|C\theta^*(\beta) - c\| \leq \|C\theta^*(\beta_0) - c\| \leq \sup_{\beta \in \Xi} \|C\theta^*(\beta) - c\|,$$

permits a conservative assessment of $\mathbf{H}_0$. To do so, we formulate the infimum hypothesis:

$$\mathbf{H}_{inf} : \inf_{\beta \in \Xi} \|C\theta^*(\beta) - c\| = 0.$$

The infimum statistic $\mathcal{T}_{inf} = \inf_{\beta \in \Xi} \|C\hat{\theta}(\beta) - c\|$ can be used to evaluate this hypothesis. The distribution of this statistic can be derived analytically in some simple situations. As an example, we revisit the normal scenario discussed earlier where the interest is in evaluating the hypothesis, $\mathbf{H}_0 : \theta_0 = 0$, using the processes $\hat{\theta}(\beta) = \bar{Y} - \beta$. For ease of illustration, assume $\Xi = [0, 1]$, such that the infimum statistic becomes $\mathcal{T}_{inf} = \inf_{\beta \in [0, 1]} |\bar{Y} - \beta|$. This is a mixture of a point mass at 0 with probability $\Pr(\bar{Y} \in [0, 1])$ and two truncated normal

distributions. Specifically,

$$\inf_{\beta \in [0,1]} |\bar{Y} - \beta| = \begin{cases} -\bar{Y} & \text{if } \bar{Y} < 0, \\ 0 & \text{if } \bar{Y} \in [0, 1], \\ \bar{Y} - 1 & \text{if } \bar{Y} > 1. \end{cases}$$

The corresponding cumulative distribution function-CDF $F_{\inf}(x) = \Pr(\inf_{\beta \in [0,1]} |\bar{Y} - \beta| \leq x)$ is

$$F_{\inf}(x) = \Pr(\bar{Y} \leq x + 1) - \Pr(\bar{Y} \leq -x), \text{ for } x \geq 0. \quad (5.1)$$

In particular, we have $F_{\inf}(0) = \Pr(\bar{Y} \leq 1) - \Pr(\bar{Y} \leq 0) = \Pr(\bar{Y} \in [0, 1])$, reflecting the point mass at 0 for $\mathcal{T}_{\inf}$.

In general, because of the complexity of the limiting distribution of the infimum of the test process, simple general analytic results do not appear tractable. Instead, resampling may be utilized. A simple nonparametric bootstrap (Efron and Tibshirani, 1993) may be used to compute variance estimators, and to carry out the simultaneous inferences necessary for the infimum tests and the confidence bands, described below. The validity of the bootstrap follows automatically from empirical process theory under the regularity conditions given in van der Vaart and Wellner (2000b) even under model misspecification. This requires the boundedness of the estimating function for fixed $\beta \in \Xi$. A difficulty with the nonparametric bootstrap is that it requires solving the estimating function for all β in each bootstrap sample, which may be computationally demanding. An alternative resampling technique which does not require repeatedly solving the estimating function may be constructed. The basic

idea is to generate realizations directly from the limiting distribution of $\hat{\theta}(\beta)$ and to use these realizations to approximate the distribution of the infimum test and confidence bands. This resampling technique has been extensively used in the literature when the true asymptotic distribution is hard if not impossible to derive analytically (see for example, Parzen et al., 1994 and Zhu and Zhang, 2006). To do this, one fixes the estimator based on the observed data and then “perturbs” this estimator using a disturbance which conditionally on data has mean zero and variance-covariance in β equalling that of $\hat{\theta}(\beta)$ in Theorem 5.2.1. The procedure is given by the following steps:

Step 1. Generate n i.i.d random variables from a standard normal model ζ , denoted $\{\zeta_1^{(b)}, \dots, \zeta_n^{(b)}\}$, where superscript (b) represents replications.

Step 2. Given the realizations of the data, $\{Y_i\}_{i=1}^n$, and values of $\beta \in \Xi$, calculate $\tilde{\theta}^{(b)}(\beta)$ using the simulated $\{\zeta_1^{(b)}, \dots, \zeta_n^{(b)}\}$ and the equation,

$$\tilde{\theta}^{(b)}(\beta) = \hat{\theta}(\beta) + \left[n^{-1} \sum_{i=1}^n s_{Y_i}(\hat{\theta}(\beta), \beta) \zeta_i^{(b)} \right] W_Y^{-1}(\hat{\theta}(\beta), \beta), \quad (5.2)$$

where the statistic $\hat{\theta}(\beta)$ takes value $\hat{\theta}^O(\beta)$ for observed data $\{Y_i\}_{i=1}^n$.

Step 3. Calculate $\mathcal{T}_{inf}^{(b)} = \inf_{\beta \in \Xi} \|C\tilde{\theta}^{(b)}(\beta) - c\|$ using $\tilde{\theta}^{(b)}(\beta), \beta \in \Xi$.

By repeatedly generating the normal variates $\{\zeta_j\}_{j=1}^n$, B times, and repeating steps 2 and 3 for each generated sample, we obtain the empirical distribution of $\mathcal{T}_{inf}^{(b)}$ given observed data. Theorem 2 below establishes that this empirical distribution converges to the marginal asymptotic distribution of $\mathcal{T}_{inf}$ as $n \rightarrow \infty$. Let $\mathbf{1}(\mathcal{E})$ be the indicator function for event $\mathcal{E}$. The p-value of the test is then $B^{-1} \sum_{b=1}^B \mathbf{1}(\mathcal{T}_{inf}^{(b)} \geq \mathcal{T}_{inf}^O)$, the proportion of these bootstrap observations which exceed $\mathcal{T}_{inf}^O$ the observed value of the statistic.

For the simple normal example, we compare the resampling null distribution of $\mathcal{T}_{inf}^{(b)}$ to the analytical distribution $F_{inf}(\cdot)$ in (5.1) for a finite sample size. Setting $\theta_0 = 0$ under the null and $\beta_0 = 0.5$, we generate $\{Y_i\}_{i=1}^n$ from a normal distribution $\mathcal{N}(0.5, 10^2)$. Furthermore, we take $\Xi = [0, 1]$ and for each resample $b = 1, \dots, B$, we compute $\mathcal{T}_{inf}^{(b)} = \inf_{\beta \in [0, 1]} |\tilde{\theta}^{(b)}(\beta)|$, where $\tilde{\theta}^{(b)}(\beta) = \bar{Y} - \beta - n^{-1} \sum_{i=1}^n (Y_i - \bar{Y}) \zeta_i^{(b)}$. Results with $n = 100$ and $B = 10000$ resamples are plotted in Figure 5.2. The resampling distribution provides a good approximation to the analytical distribution for this simple hypothetical example.

If the infimum (null) hypothesis cannot be rejected, then a supremum test or equivalently a simultaneous confidence region may be used to check whether $\|C\theta^*(\beta) - c\| > 0$ in some regions of Ξ . The supremum hypothesis $\mathbf{H}_{sup}$ may be tested with the statistic $\mathcal{T}_{sup}^o = \sup_{\beta \in \Xi} \|C\hat{\theta}(\beta) - c\|$ using the bootstrap realizations of $\tilde{\theta}^{(b)}(\beta), \beta \in \Xi$. The p-value of the supremum test is then $B^{-1} \sum_{b=1}^B \mathbf{1}(\mathcal{T}_{sup}^{(b)} \geq \mathcal{T}_{sup}^o)$, where $\mathcal{T}_{sup}^{(b)}$ are the bootstrap realizations of the statistic. Alternatively, a simultaneous confidence region for $C\theta^*(\beta) - c$ across all values of β may be constructed. Let $0 < \varphi < 1$. A simultaneous confidence region for $C\theta^*(\beta) - c, \beta \in \Xi$ is given by $\{\vartheta(\beta) : \Xi \rightarrow R^r; \|\vartheta(\beta) - C\hat{\theta}(\beta) + c\| < \rho_\varphi\}$, where ρ_φ is the $(1 - \varphi)$ th empirical percentile of $\{\sup_{\beta \in \Xi} \|C\tilde{\theta}^{(b)}(\beta) - C\hat{\theta}^o(\beta)\|\}_{b=1}^B$, with $\hat{\theta}^o(\beta)$ being the value of the statistic $\hat{\theta}(\beta)$ for observed data $\{Y_i\}_{i=1}^n$.

The following result supports the validity of the resampling based infimum test and confidence bands.

Theorem 5.2.2. *Under Conditions C1-C4, the conditional distribution of the process $n^{1/2}\{\tilde{\theta}(\beta) - \hat{\theta}^o(\beta)\}$ given realizations $\{Y_i\}_{i=1}^n$ of Y , is asymptotically equivalent to the unconditional distribution of the process $n^{1/2}\{\hat{\theta}(\beta) - \theta^*(\beta)\}, \beta \in \Xi$.*

Theorem 5.2.2 (proof provided in the appendix) coupled with a continuous mapping

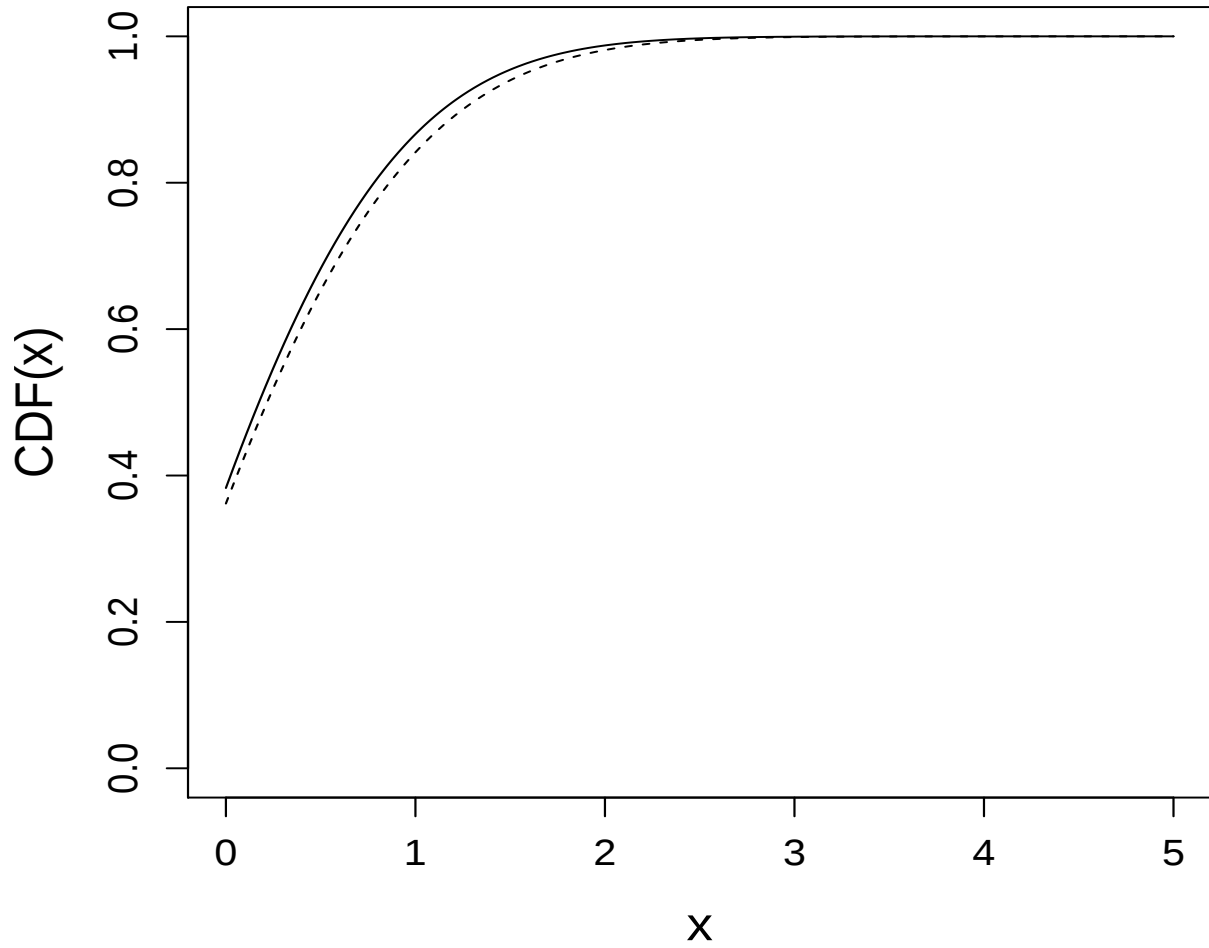


Figure 5.2: Plot of the exact (solid line) and the resampled (dashed line) CDF ($\text{CDF}(x) = \Pr(\inf_{\beta \in [0,1]} |\bar{Y} - \beta| \leq x)$) of the infimum test statistic under the null $\theta_0 = 0$ for the simple normal example, assuming the true parameter $\beta_0 = 0.5$, sample size $n = 100$ and $B = 10000$ resamples.

theorem gives that the infimum and supremum tests can be carried out using this resampling procedure. For the simple normal example, $n^{1/2}\{\hat{\theta}(\beta) - \theta^*(\beta)\} = n^{1/2}(\bar{Y} - \theta_0 - \beta_0)$ and $n^{1/2}\{\tilde{\theta}(\beta) - \hat{\theta}^o(\beta)\} = -n^{-1/2}\sum_{i=1}^n(Y_i - \bar{Y})\zeta_i$, which do not depend on β . The random quantity $n^{1/2}(\bar{Y} - \theta_0 - \beta_0)$ is normally distributed with mean 0 and variance 1, uniformly in β , for each fixed n . Given observed data $\{Y_i\}_{i=1}^n$, $-n^{-1/2}\sum_{i=1}^n(Y_i - \bar{Y})\zeta_i$ is also normally distributed with mean 0 and variance $n^{-1}\sum_{i=1}^n(Y_i - \bar{Y})^2$ which converges almost surely to 1 as $n \rightarrow \infty$.

The choice of the support Ξ of β is critically important in performing the test in practice. If values of β are selected in some data-driven fashion, the limiting distribution in Theorem 1 will be invalid. This is similar to Hansen (1996) for the case where the model is identifiable under the null after profiling on β , that is, when $\theta^*(\beta) = \theta_0, \forall \beta \in \Xi$. On the other hand, an approach which ignores sample information about Ξ may be unnecessarily conservative and potentially sacrifices power. One possible solution is to consult with subject-matter experts on the choice of Ξ . This choice ideally should be based on prior studies, as in the breast cancer analysis in Section 3, where closely related datasets were used to select the range for the sensitivity parameter. From a technical standpoint, this choice should also be computationally feasible.

5.3 Numerical studies

5.3.1 Pseudo-likelihood models with missing data

We consider the data set-up and model described in Troxel et al. (1998b) for potentially non-random missing data in longitudinal studies. The model will be referred to as the TLH model.

The data arise from a longitudinal study where each subject i ($i = 1, \dots, n$), is to be observed at K occasions. For subject i , we have a $K \times 1$ response vector, $Y_i^* = (Y_{i1}^*, \dots, Y_{iK}^*)^T$ which may not be fully observed. To accommodate missingness, subject i has a vector of missing data indicators $R_i = (R_{i1}, \dots, R_{iK})^T$, where $R_{it} = 1$ if Y_{it}^* is observed and 0 otherwise. Let $Y_{i,obs}^*$ and $Y_{i,miss}^*$ denote the observed and missing components of Y_i^* , respectively. Each individual also has a $K \times J$ covariate matrix X_i , which is assumed fully observed. The response Y_i in our general formulation is $\{Y_{i,obs}^*, R_i, X_i\}$.

The key idea of the TLH methodology is to model the time point pair (Y_{it}^*, R_{it}) , without accounting for the dependence on other time points. Let $f(u | w)$ denote the density function of random quantity u conditional on possibly non random quantity w . We assume a simple selection model given by, $f(Y_{it}^*, R_{it} | X_{it}, \varpi) = f(Y_{it}^* | X_{it}, \varpi)f(R_{it} | Y_{it}^*, X_{it}, \varpi)$, where ϖ is a finite but unknown parameter and X_{it} may contain both time dependent and independent covariates.

The TLH model assumes that density $f(Y_{it}^* | X_{it}, \varpi)$ is that of normal $\mathcal{N}(\mu_{it}, \sigma_t)$, where μ_{it} and σ_t ($t = 1, \dots, K$) are elements of ϖ . The missing data process is assumed to satisfy $R_{it} \sim \text{Bernoulli}(1 - \pi_{it})$ where the failure probability $\pi_{it} = \Pr(R_{it} = 0 | Y_{it}^*, X_{it}, \varpi)$. We assume a logistic regression model relating the missing data probability to potentially unobserved responses, that is,

$$\text{logit}(\pi_{it}) = \gamma_{0t} + \gamma_{1t}X_{it} + \beta_t Y_{it}^*, \quad (5.3)$$

where γ_{jt} and β_t ($j = 0, 1; t = 1, \dots, K$) are unknown parameters and elements of ϖ . The parameter β_t measures the extent of non-randomness of the missing data mechanism in the study at time t . Specifically, $\exp\{\beta_t\}$ represents the odds ratio for missing response at time t

for each additional unit increase of the hypothetical response Y_{it}^* . Here, π_{it} in (5.3) depends on Y_{it}^* and not on previous elements of Y_i^* . Following warnings by Troxel et al. (1998b), we emphasize that this model could suffer from misspecification if the approximation of the logistic link function to the true link function fails.

The TLH model lends itself to a pseudo-likelihood analysis (Gong and Samaniego, 1981), where the longitudinal association is naively ignored in the likelihood construction. Specifically, the independence pseudo-likelihood function based on observed data $\{Y_{i,obs}^*, R_i, X_i\}_{i=1}^n$ is

$$\begin{aligned}
\ell_{ind}(\varpi) &= \prod_{i=1}^n f(Y_{i,obs}^*, R_i | \varpi) = \prod_{i=1}^n \int \cdots \int f(Y_{i,obs}^*, Y_{i,miss}^*, R_i | \varpi) dY_{i,miss}^* \\
&= \prod_{i=1}^n \prod_{t=1}^K \{f(Y_{it}^*, R_{it} | \varpi)\}^{R_{it}} \left\{ \int f(Y_{it}^*, R_{it} | \varpi) dY_{it}^* \right\}^{1-R_{it}} \\
&= \prod_{i=1}^n \prod_{t=1}^K \{f(Y_{it}^* | \varpi) f(R_{it} | Y_{it}^*, \varpi)\}^{R_{it}} \\
&\quad \left\{ \int f(Y_{it}^* | \varpi) f(R_{it} | Y_{it}^*, \varpi) dY_{it}^* \right\}^{1-R_{it}} \\
&= \prod_{i=1}^n \prod_{t=1}^K \{f(Y_{it}^* | \varpi) (1 - \pi_{it})\}^{R_{it}} \left\{ \int f(Y_{it}^* | \varpi) \pi_{it} dY_{it}^* \right\}^{1-R_{it}}.
\end{aligned}$$

We have suppressed the dependence on covariates in $\ell_{ind}(\varpi)$. As a pseudo-likelihood model, conditions C1-C4 are easily verified and the asymptotic results hold. The densities in the TLH model are normal and Bernoulli, which are smooth functions of the unknown parameters.

5.3.2 Real data analysis

To illustrate our methodology, we consider data from the International Breast Cancer Study Group-IBCSG, previously reported by Hürny et al. (1992); and Troxel et al. (1998b). This is a group of randomized breast cancer studies with primary endpoints being survival and relapse; and quality of life being a secondary endpoint. One study, Study VI, is a randomized trial of adjuvant chemotherapy following surgery for the treatment of breast cancer. In this study, 4 treatments (A, B, C and D) were randomly assigned to 431 pre-menopausal cancer patients and several domains of quality of life were assessed. In this paper, we focus on three quality-of-life domains; 1) PACIS (perceived adjustment to chronic illness scale), 2) Mood and 3) Appetite. These variables were originally measured on a 0 – 100 scale but are normalized using a square-root transformation as recommended by Troxel et al. (1998b). Questionnaires for the quality of life assessment were administered to study patients at baseline and every three months for two years. Our analysis employs the first three time points, with rates of missing data equalling 16%, 33% and 37% for PACIS, 16%, 33% and 38% for Mood, and 15%, 33% and 38% for Appetite. A full description of Study VI and other IBCSG trials may be found elsewhere (Hürny et al., 1992; Troxel et al., 1998a).

As in earlier analyses of Study VI, we consider the following model for the measurement outcome,

$$\mu_{it} = \mu_{0t} + \alpha_1 X_{1i} + \alpha_2 X_{2i} + \alpha_3 X_{3i}, \quad (t = 1, 2, 3),$$

where μ_{0t} is a time-dependent intercept and α_j is a slope associated with $X_{ji}, j = 1, 2, 3$.

$$\text{Here } (X_{1i}, X_{2i}, X_{3i}) = \begin{cases} (1, 0, 0) & \text{if treatment A,} & (0, 0, 1) & \text{if treatment C} \\ (0, 1, 0) & \text{if treatment B,} & (0, 0, 0) & \text{if treatment D.} \end{cases}$$

The missing data model is

$$\text{logit}(\pi_{it}) = \gamma_{0t} + \beta Y_{it}^* \quad (i = 1, \dots, 431; t = 1, 2, 3),$$

where γ_{0t} is a time-dependent intercept and β is a slope associated with Y_{it}^* . As discussed previously, β quantifies the nonrandomness of the missing data process. A constant σ_t is assumed across time.

Our objective is to assess the treatment and time effects on the mean quality of life. Under the assumed model, the hypotheses of interest are $\alpha_1 = \alpha_2 = \alpha_3 = 0$ and $\mu_{01} = \mu_{02} = \mu_{03}$ for the treatment and time effects, respectively. As a preliminary analysis, we first evaluated these hypotheses under identifiability assumptions. Specifically, we fit the TLH model by simultaneously estimating both β and $\theta = (\alpha_1, \alpha_2, \alpha_3, \mu_{01}, \mu_{02}, \mu_{03}, \gamma_{01}, \gamma_{02}, \gamma_{03})$ via the independence pseudolikelihood estimating function. A Wald test based on the sandwich estimator of the covariance matrix of the regression parameter estimates was performed to evaluate the hypotheses of interest. P-values of these Wald tests for the three responses are given in Table 1. In brief, these inferences suggest that there is no treatment effect on PACIS and Appetite and no time effect on PACIS and Mood. The treatment effect on Mood and the time effect on Appetite are significant at 1% level. In addition to these analyses, we also conducted two crude analyses that do not explicitly model the missing data mechanism. The first analysis used only subjects with complete data sequences, therefore removing subjects with incomplete data profiles. The second analysis ignored the missing data and conduct the so-called ignorable (missing at random) inferences by forcing β , the non-randomness missing data parameter, to 0. Results of these analyses are also summarized in Table 1. From these additional exploratory analyses, the treatment and time effects are found to be statistically

significant for Mood at 5% level. The ignorable analysis also appears to yield a statistically significant time effects on Appetite. Of course, these crude analyses may not be reliable as they rely on assumptions that are not verifiable using observed data at hand.

The Wald tests conducted under the assumption of identifiability may not have desirable properties if identifiability is violated. As illustrated in Figure 5.1, the model was at best weakly identifiable for the outcome PACIS. Model identifiability was also a concern for β for the other two responses. We performed the infimum test to conservatively evaluate the treatment and time effects on the three quality of life domains. To conduct these tests, the set Ξ for the range of β was obtained from an independent source. We considered data on post-menopausal cancer patients from Study VII of the IBCSG trials. Objectives of this study were similar to those of Study VI, except that the menopausal status of study participants differed. The joint model appeared to be identifiable when applied to Study VII data. Based on these results, we derived 99% confidence intervals to use as ranges for β in the infimum tests for Study VI. The ranges for PACIS, Mood and Appetite were $[-4, 0]$, $[-3, 0]$ and $[-5.6, -1.6]$, respectively. Recall that in the missing data model, $\exp\{-\beta\}$ represents the odds ratio of being observed at any time point for each additional unit increase of the hypothetical response Y_{it}^* . Since Y_{it}^* takes values in the range $0 - 10$ on a square-root scale, for the selected ranges, the odds ratio may be as high as; $\exp\{4\} = 54.60$ for PACIS, $\exp\{3\} = 20.09$ for Mood, and $\exp\{5.6\} = 270.43$ for Appetite. One might criticize these upper bounds as being scientifically unreasonable. However, permitting such extreme scenarios provides for a conservative test, which is in the spirit of sensitivity analysis. For computational feasibility, the ranges were approximated on fine grids with equally spaced points of 0.02. P-values of the infimum tests are given in Table 1.

The infimum hypothesis for the treatment effect was rejected for Mood at the 5% level (p-

Table 5.1: P-values for evaluating the Treatment and Time effects using data from Study VI of the IBCSG trials

			Reponses		
			PACIS	Mood	Appetite
Treatment effect	Wald test	TLH	0.281	0.008	0.229
		COM [†]	0.170	0.041	0.370
		IGN [‡]	0.303	0.011	0.376
	Infimum test		0.231(0.522*)	0.025	0.281(0.369*)
Time effect	Wald test	TLH	0.216	0.552	<0.001
		COM [†]	0.302	0.030	0.163
		IGN [‡]	0.142	0.006	0.023
	Infimum test		0.134(<0.001*)	0.063(<0.001*)	<0.001

* P-value of supremum test

[†] Complete cases; [‡] Ignorable cases

value= 0.025), but not for PACIS (p-value= 0.281) and Appetite (p-value= 0.231). For the time variable, a strongly significant effect was detected only for Appetite (p-value< 0.001). For nonsignificant infimum test results, a supremum test was conducted to see if one could not reject the null hypothesis for all values $\beta \in \Xi$. The supremum test for the treatment effect was not rejected on PACIS (p-value= 0.522) and Appetite (p-value= 0.369), but was strongly rejected for the time effect on PACIS (p-value< 0.001) and Mood (p-value< 0.001).

When the supremum test was rejected, a sensitivity analysis was conducted using a simultaneous 95% confidence band approach to identify regions of β for which the point-wise null hypotheses are rejected. Plots of these analyses for contrasts $\mu_{01}^*(\beta) - \mu_{02}^*(\beta)$ and $\mu_{02}^*(\beta) - \mu_{03}^*(\beta)$ for PACIS and Mood are given in Figure 3. For PACIS, the 95% simultaneous confidence band for $\mu_{02}^*(\beta) - \mu_{01}^*(\beta)$, $-4 \leq \beta < -0.4$; and $\mu_{03}^*(\beta) - \mu_{02}^*(\beta)$, $-4 \leq \beta \leq -3$, did not contain 0. Similar analyses for Mood revealed that a 95% simultaneous confidence band for $\mu_{02}^*(\beta) - \mu_{01}^*(\beta)$, $-3 \leq \beta < -0.7$, did not contain 0. The confidence band for $\mu_{03}^*(\beta) - \mu_{02}^*(\beta)$ did not exclude 0 over the selected range of β ($-3 \leq \beta < 0$). The Wald tests which assume identifiability were nonsignificant for all pairwise comparisons at the 5% level.

5.3.3 Simulation study

In this section, we report results of a simulation study comparing the performance of the infimum test to that of the naive Wald test derived under identifiability assumptions. The simulations were conducted under a TLH model specified so as to roughly approximate data from Study VI of the IBCSG trials. For simplicity, only two treatments (A and B) and two time points ($K = 2$) were considered. The outcome vector (Y_{i1}^*, Y_{i2}^*) , assuming dependence on subject i , was generated from a two-dimensional normal distribution with univariate mean

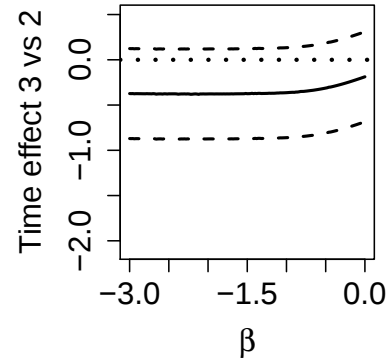
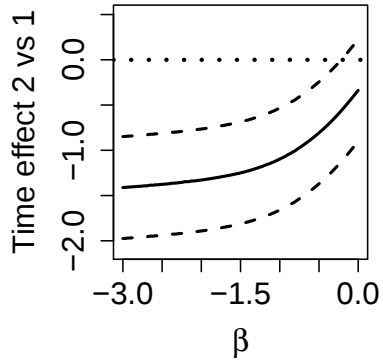
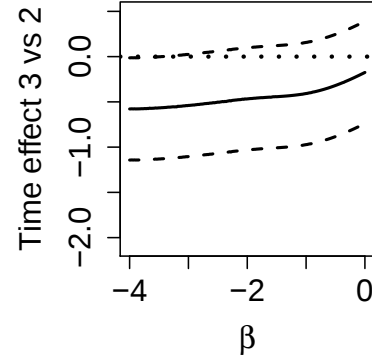
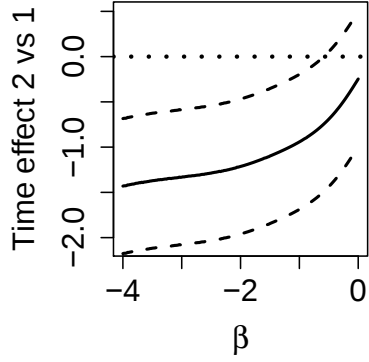


Figure 5.3: The top panel corresponds to PACIS; and the bottom panel to Mood. In each panel, the solid lines represent $\hat{\mu}_{02}(\beta) - \hat{\mu}_{01}(\beta)$ (on the left) and $\hat{\mu}_{03}(\beta) - \hat{\mu}_{02}(\beta)$ (on the right) for fixed values of the parameter β . The dashed lines are the corresponding 95% simultaneous confidence bands and the dotted lines are the null values.

models,

$$\mu_{it} = \mu_{0t} + \alpha_t X_{1i}, t = 1, 2,$$

and time-point variances $\sigma_t, t = 1, 2$, and correlation coefficient ρ . The parameters μ_{0t} and α_t are time-dependent intercepts and slopes associated with covariate X_{1i} , which equals 1 if treatment B and 0 otherwise. We reparameterized μ_{0t} and α_t as, $\mu_{0t} = \tilde{\alpha}_0 + \tilde{\alpha}_1 I(t = 2)$ and $\alpha_t = \tilde{\alpha}_2 + \tilde{\alpha}_3 I(t = 2)$, where $I(t = 2)$ is an indicator variable taking value 1 at the second time point. Throughout our simulations, we fixed the variances $\sigma_t, t = 1, 2$, to 1 and the correlation coefficient ρ to 0.4. Missing observations were generated using a logistic model relating the dropout probability π_{it} to the response Y_{it}^* as,

$$\text{logit}(\pi_{it}) = \gamma_{0t} + \gamma_{1t} X_{1i} + \beta Y_{it}^*,$$

where γ_{0t} , γ_{1t} and β are respectively the intercept and slopes associated with X_{1i} and Y_{it}^* . Time-dependent parameters γ_{0t} and γ_{1t} were reparameterized as, $\gamma_{0t} = \tilde{\gamma}_0 + \tilde{\gamma}_1 I(t = 2)$ and $\gamma_{1t} = \tilde{\gamma}_2 + \tilde{\gamma}_3 I(t = 2), t = 1, 2$.

We study the size and power of the infimum and Wald tests for $\tilde{\alpha}_3$, the parameter that captures the interaction effect of time and treatment on the mean response. We set $\tilde{\alpha}_3 = 0$ and $\tilde{\alpha}_3 = 1$ for the size and power of the test respectively. Additionally, $(\tilde{\alpha}_1, \tilde{\alpha}_2) = (0, 0)$ and $(\tilde{\gamma}_1, \tilde{\gamma}_2, \tilde{\gamma}_3) = (0.5, -2, 0.2)$ when evaluating size, and $(\tilde{\alpha}_1, \tilde{\alpha}_2) = (0.1, 1)$ and $(\tilde{\gamma}_1, \tilde{\gamma}_2, \tilde{\gamma}_3) = (1, -3, 1)$ when evaluating power.

The parameter $\tilde{\gamma}_0$ was varied throughout our simulations to produce different missing data rates. Specifically, to study the size of the test $\tilde{\gamma}_0$ was fixed to 0.5, to produce about 15% and 22% missing observations at the first and second time point, respectively and at

1.8 to produce about 33% and 43% missing observations at the first and second time point, respectively. For the power, $\tilde{\gamma}_0$ was fixed to 0.5, producing rates of missing observations roughly 14% and 26% at the first and second time point, respectively and to 2, producing rates of missing observations roughly 32% and 46% at the first and second time point, respectively. Finally, throughout our simulations, we set the true β to -1 .

One thousand datasets were generated with sample sizes 100 and 300. Equal proportions of subjects were assigned to treatment A and B. The infimum tests were performed on the interval $\Xi = [-2, 0]$. To ensure computational feasibility, a fine grid of equally spaced points of 0.02, was considered. We used 1000 resamples from the alternative resampling scheme discussed in section 2.3 to approximate the null distribution of the infimum test.

The infimum and Wald tests were performed using working regression models having the same form as those used to generate data. These models saturate the number of parameters, leading to potential nonidentifiability as a result of overparameterization. Table 2 shows the rejection rates for nominal test levels 0.01, 0.05, and 0.1. Asymptotic standard errors (as the number of Monte Carlo iterations tends to infinity) are reported in the last row of the table. Overall, the infimum tests perform well, with the resampling distribution of the test providing a reasonable approximation to the nominal level. The Wald test appears to be very liberal when compared to the infimum test. The anti-conservativeness of the Wald test does not diminish as the sample size increases. Based on these results, our recommendation is to avoid the Wald test when identifiability is of concern. Because the empirical type I error rate of the infimum test and that of the Wald test are different, comparing their empirical powers is not appropriate. Nevertheless for both methods, a larger sample size improves the power of detecting the alternatives under consideration, a finding consistent with the literature. Moreover, the power decreases with increasing missing data rates.

Table 5.2: Empirical type I error and power of the infimum test[‡] and Wald test (in parenthesis) for evaluating the interaction effect represented by the parameter $\tilde{\alpha}_3$

True value	Missing ⁺ data rate	n=100			n=300		
		Nominal test level			Nominal test level		
		0.1	0.05	0.01	0.1	0.05	0.01
$\tilde{\alpha}_3 = 0$	15%, 22%	0.099	0.049	0.011	0.103	0.057	0.010
	($\tilde{\gamma}_0 = 0.5$)	(0.147)	(0.092)	(0.034)	(0.158)	(0.117)	(0.059)
	33%, 43%	0.110	0.044	0.010	0.122	0.062	0.012
	($\tilde{\gamma}_0 = 1.8$)	(0.149)	(0.097)	(0.051)	(0.165)	(0.113)	(0.073)
$\tilde{\alpha}_3 = 1$	14%, 26%	0.982	0.959	0.870	>0.999	>0.999	>0.999
	($\tilde{\gamma}_0 = 0.5$)	(0.961)	(0.944)	(0.848)	(0.948)	(0.942)	(0.934)
	32%, 46%	0.905	0.846	0.675	0.999	0.999	0.993
	($\tilde{\gamma}_0 = 2$)	(0.889)	(0.827)	(0.659)	(0.961)	(0.949)	(0.926)
Monte carlo SE		0.003	0.007	0.009	0.003	0.007	0.009

[‡] Test performed using $\Xi = [-2, 0]$

⁺ First and second time point missing data rate

While the ability to choose an appropriate support set Ξ of β to perform the infimum tests is highly desirable in practice, our simulations (results not shown) indicate that only a minimal inflation of type I error rate is observed under a modest misspecification of the set Ξ . For example, when Ξ does not contain the true β but β_0 is not far away from the boundaries of the set, close to the nominal level is still achieved under the null hypothesis. As an example, we performed the infimum test on the interval $[0, 2]$, which does not contain $\beta_0 = -1$. For this range of β , the infimum tests nearly maintain their sizes at all significance levels. However, when $[10, 12]$ was selected for the range of β , the infimum tests were overly anti-conservative.

Another simulation study was conducted to evaluate the effects of the choice of the set Ξ on the power of the infimum tests. Specifically, we generated data as before, but performed the infimum tests on wider intervals, namely $[-3, 3]$ and $[-5, 5]$. Results of this simulation study are given in Table 5.2. As expected, the power decreases as the interval widens, which occurs regardless of the missing data rate. Following a referee's recommendation, further simulations were conducted to evaluate the loss of power when the infimum test is performed on a given support set of β compared to the ideal set $\Xi = \{\beta_0\}$. For this, we generated the data as before with the only difference that $\tilde{\alpha}_3 = 0.7$. We then performed the infimum test using $\Xi = [-2, 0]$ and $\Xi = \{-1\}$. Results revealed a minor loss of power of the infimum test on $\Xi = [-2, 0]$ compared to the ideal set $\Xi = \{-1\}$ (see Table 5.4).

5.4 Discussion

While hypothesis testing under nonidentifiability has been previously considered, the framework is often too restrictive for sensitivity analyses. In a sensitivity analysis, the model may

Table 5.3: Empirical power of the infimum test to detect the interaction effect $\tilde{\alpha}_3 = 1$ for two ranges Ξ of β with true value being $\beta_0 = -1$

Ξ	Missing ⁺ data rate	n=100			n=300		
		Nominal test level			Nominal test level		
		0.10	0.05	0.01	0.10	0.05	0.01
[-3, 3]	14%, 26%	0.912	0.865	0.720	0.999	0.997	0.987
	32%, 46%	0.783	0.706	0.529	0.981	0.960	0.905
[-5, 5]	14%, 26%	0.899	0.836	0.655	0.997	0.990	0.970
	32%, 46%	0.767	0.695	0.492	0.953	0.925	0.858
Monte carlo SE		0.003	0.007	0.009	0.003	0.007	0.009

⁺ First and second time point missing data rate

Table 5.4: Empirical power of the infimum test to detect the interaction effect $\tilde{\alpha}_3 = 0.7$ for two sets Ξ of β with true value being $\beta_0 = -1$

Ξ	Missing ⁺ data rate	n=100			n=300		
		Nominal test level			Nominal test level		
		0.10	0.05	0.01	0.10	0.05	0.01
[-2, 0]	13%, 23%	0.806	0.699	0.470	0.996	0.989	0.949
{-1}	13%, 23%	0.829	0.725	0.469	0.998	0.998	0.982
[-2, 0]	31%, 44%	0.677	0.570	0.342	0.978	0.957	0.834
{-1}	31%, 44%	0.707	0.589	0.348	0.992	0.974	0.915
Monte carlo SE		0.003	0.007	0.009	0.003	0.007	0.009

⁺ First and second time point missing data rate

not be identifiable under either the null or alternative hypothesis, and profiling may not lead to consistent estimation of the parameter of interest under the null. As a result, the supremum test may not be appropriate. As discussed in this paper, a theoretically rigorous approach to this testing problem may be based on infimum statistics, whose distribution must be carefully considered under model misspecification under the null hypothesis.

The infimum testing approach was previously studied for likelihood analyses of parametric models (Todem et al., 2010). In this paper, we have extended these results to general estimating functions for parametric models. This includes limiting results for the profile estimators and the infimum test and confidence bands, as well as the validity of the bootstrap procedure. Such results are critically important in sensitivity analyses of complex data arising in longitudinal studies, where full model specification may be difficult and partially specified models may be more easily analyzed using non-likelihood based approaches.

APPENDIX

Proof of Theorem 5.2.1

i) We show that $\sup_{\beta \in \Xi} \|\hat{\theta}(\beta) - \theta^*(\beta)\| \rightarrow_p 0$.

Condition C2 implies that $\mathcal{G}_1$ and $\mathcal{G}_2$ are Donsker and hence Glivenko-Cantelli (van der Vaart and Wellner, 2000a, 2000b). Therefore,

$$\sup_{\theta \in \Theta, \beta \in \Xi} \|S_Y(\theta, \beta) - \tilde{S}(\theta, \beta)\| \rightarrow_p 0 \quad \text{and} \quad \sup_{\theta \in \Theta, \beta \in \Xi} \|W_Y(\theta, \beta) - \tilde{W}(\theta, \beta)\| \rightarrow_p 0. \quad (5.4)$$

The definitions of $\hat{\theta}(\beta)$ and $\theta^*(\beta)$ and Condition C4 imply that

$$\begin{aligned} 0 &= S_Y(\hat{\theta}(\beta), \beta) - \tilde{S}(\theta^*(\beta), \beta) \\ &= \left(S_Y(\hat{\theta}(\beta), \beta) - S_Y(\theta^*(\beta), \beta) \right) + \left(S_Y(\theta^*(\beta), \beta) - \tilde{S}(\theta^*(\beta), \beta) \right) \\ &= \tilde{W}(\check{\theta}(\beta), \beta) \left(\hat{\theta}(\beta) - \theta^*(\beta) \right) + v_{2n}(\beta) \left(\hat{\theta}(\beta) - \theta^*(\beta) \right) + \nu_{1n}(\beta), \end{aligned} \quad (5.5)$$

where $\check{\theta}(\beta)$ is on the line segment between $\hat{\theta}(\beta)$ and $\theta^*(\beta)$. Also, $\sup_{\beta \in \Xi} |v_{2n}(\beta)| \rightarrow_p 0$ and $\nu_{1n}(\beta) = S_Y(\theta^*(\beta), \beta) - \tilde{S}(\theta^*(\beta), \beta)$. From (5.5), we have,

$$\begin{aligned} \sup_{\beta \in \Xi} \|\hat{\theta}(\beta) - \theta^*(\beta)\| &= \sup_{\beta \in \Xi} \left\| - \left(\tilde{W}^{-1}(\check{\theta}(\beta), \beta) + v_{2n}(\beta) \right) \nu_{1n}(\beta) \right\| \\ &\leq \sup_{\beta \in \Xi} \left\| -\tilde{W}^{-1}(\check{\theta}(\beta), \beta) + v_{2n}(\beta) \right\| \sup_{\beta \in \Xi} \|\nu_{1n}(\beta)\|. \end{aligned}$$

Because of Condition C3, for any $\theta \in \Theta$, for any $\beta \in \Xi$, there exists a positive number λ_1 , such that $\lambda_{\min}(\beta) > \lambda_1 > 0$. For any $s \times s$ symmetric matrix $\mathbf{A}$, denote its Euclidean norm as $\|\mathbf{A}\| = \lambda_{\max}(\mathbf{A})$, where $\lambda_{\max}(\mathbf{A})$ is the largest eigenvalue of $\mathbf{A}$ and if $\mathbf{A}$ is also nonsingular, $\|\mathbf{A}^{-1}\| = \lambda_{\min}^{-1}(\mathbf{A})$. Therefore, $\left\| -\tilde{W}^{-1}(\check{\theta}(\beta), \beta) \right\| = \lambda_{\min}^{-1}(\beta) \leq \lambda_1^{-1}$, and $\sup_{\beta \in \Xi} \|\hat{\theta}(\beta) - \theta^*(\beta)\| \leq \lambda_1^{-1} \sup_{\beta \in \Xi} \|\nu_n(\beta)\|$. The uniform consistency of

$\hat{\theta}(\beta)$ to $\theta^*(\beta)$ follows from $\sup_{\beta \in \Xi} \|\nu_n(\beta)\| = \sup_{\beta \in \Xi} \|S_Y(\theta^*(\beta), \beta) - \tilde{S}(\theta^*(\beta), \beta)\| \leq \sup_{\theta \in \Theta, \beta \in \Xi} \|S_Y(\theta, \beta) - \tilde{S}(\theta, \beta)\| \rightarrow_p 0$, according to (5.4).

ii) We show that $n^{1/2}(\hat{\theta}(\beta) - \theta^*(\beta))$ converge weakly to a tight Gaussian process.

Based on the uniform consistency of $\hat{\theta}(\beta)$, and (5.4) and (5.5), applying the Taylor expansion to $S_Y(\hat{\theta}(\beta), \beta)$ around $S_Y(\theta^*(\beta), \beta)$ gives

$$\begin{aligned} n^{1/2} \left(\hat{\theta}(\beta) - \theta^*(\beta) \right) &\approx -n^{-1/2} \tilde{W}^{-1}(\theta^*(\beta), \beta) \nu_{1n}(\beta) \\ &= -n^{-1/2} \sum_{i=1}^n \tilde{W}^{-1}(\theta^*(\beta), \beta) \left(s_{Y_i}(\theta^*(\beta), \beta) - \mathbb{E} s_{Y_1}(\theta^*(\beta), \beta) \right) \\ &= -n^{-1/2} \sum_{i=1}^n \tilde{W}^{-1}(\theta^*(\beta), \beta) s_{Y_i}(\theta^*(\beta), \beta) \equiv -n^{-1/2} \sum_{i=1}^n \eta_i(\beta), \end{aligned}$$

where $\approx$ denotes asymptotic equivalence uniformly in $\beta \in \Xi$. Because Condition C2 implies that $\mathcal{G}_1$ is Donsker and using previous results that $\tilde{W}^{-1}(\theta^*(\beta), \beta)$ is uniformly bounded for $\beta \in \Xi$, the function class $\{\tilde{W}^{-1}(\theta^*(\beta), \beta) s_{Y_i}(\theta^*(\beta), \beta), \beta \in \Xi, i = 1, \dots, n\}$ is Donsker. This permits the application of a functional central limit theory to establish the weak convergence of $\hat{\theta}(\beta)$. Therefore, $\lim_{n \rightarrow \infty} \text{cov} \left\{ n^{1/2} \left(\hat{\theta}(\beta_1) - \theta^*(\beta_1) \right), n^{1/2} \left(\hat{\theta}(\beta_2) - \theta^*(\beta_2) \right) \right\} = \mathbb{E} \left(\eta_1(\beta_1) \eta_1^T(\beta_2) \right) = \Sigma^*(\beta_1, \beta_2)$. For a given β , $\text{var} \left\{ n^{1/2} \left(\hat{\theta}(\beta) - \theta^*(\beta) \right) \right\} = \mathbb{E} \left(\eta_1(\beta) \eta_1^T(\beta) \right)$.

Proof of Theorem 5.2.2

Applying the Taylor expansion to $s_{Y_i}(\theta^*(\beta), \beta)$ around $s_{Y_i}(\hat{\theta}^O(\beta), \beta)$ gives

$$\begin{aligned} &n^{-1/2} \sum_{i=1}^n s_{Y_i}(\theta^*(\beta), \beta) \zeta_i \\ &= n^{-1/2} \sum_{i=1}^n s_{Y_i}(\hat{\theta}^O(\beta), \beta) \zeta_i + n^{-1/2} \left(\theta^*(\beta) - \hat{\theta}^O(\beta) \right) \sum_{i=1}^n g_{Y_i}(\bar{\theta}(\beta), \beta) \zeta_i, \end{aligned}$$

where $\bar{\theta}(\beta)$ is on the line segment between $\hat{\theta}^O(\beta)$ and $\theta^*(\beta)$. Given observations $\{Y_i\}_{i=1}^n$, Condition C4 and $\sup_{\beta \in \Xi} \|\theta^*(\beta) - \hat{\theta}^O(\beta)\| = o_p(1)$, one has

$$n^{-1/2} \sum_{i=1}^n s_{Y_i}(\theta^*(\beta), \beta) \zeta_i \approx n^{-1/2} \sum_{i=1}^n s_{Y_i}(\hat{\theta}^O(\beta), \beta) \zeta_i.$$

Based on the definition of $\tilde{\theta}(\beta)$ in (5.2) and (5.4), one has

$$\begin{aligned} n^{1/2} \left(\tilde{\theta}(\beta) - \hat{\theta}^O(\beta) \right) &= -n^{-1/2} W_Y^{-1}(\hat{\theta}^O(\beta), \beta) \sum_{i=1}^n s_{Y_i}(\hat{\theta}^O(\beta), \beta) \zeta_i \\ &\approx -n^{-1/2} W_Y^{-1}(\hat{\theta}^O(\beta), \beta) \sum_{i=1}^n s_{Y_i}(\theta^*(\beta), \beta) \zeta_i \\ &\approx -n^{-1/2} \tilde{W}^{-1}(\theta^*(\beta), \beta) \sum_{i=1}^n s_{Y_i}(\theta^*(\beta), \beta) \zeta_i \\ &= -n^{-1/2} \sum_{i=1}^n \eta_i(\beta) \zeta_i. \end{aligned}$$

Hence, conditional on observations $\{Y_i\}_{i=1}^n$, $n^{1/2} \left(\tilde{\theta}(\beta) - \hat{\theta}^O(\beta) \right)$ converges weakly to a Gaussian process with mean 0 and covariance function

$\Sigma^O(\beta_1, \beta_2) = \lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n \mathbb{E}(\eta_i(\beta_1) \zeta_i \zeta_i^T \eta_i^T(\beta_2) \mid Y)$. We also have

$$\begin{aligned} &\Sigma^O(\beta_1, \beta_2) - \Sigma^*(\beta_1, \beta_2) \\ &= \lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n \left[\mathbb{E}(\eta_i(\beta_1) \zeta_i \zeta_i^T \eta_i^T(\beta_2) \mid Y) \right] - \mathbb{E} \left(\eta_1(\beta_1) \eta_1^T(\beta_2) \right) \\ &= \lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n \left(\eta_i(\beta_1) \eta_i^T(\beta_2) \right) - \mathbb{E} \left(\eta_1(\beta_1) \eta_1^T(\beta_2) \right) = 0. \end{aligned}$$

Hence, $\lim_{n \rightarrow \infty} \text{cov} \left\{ n^{1/2} \left(\tilde{\theta}(\beta_1) - \hat{\theta}^O(\beta_1) \right), n^{1/2} \left(\tilde{\theta}(\beta_2) - \hat{\theta}^O(\beta_2) \right) \right\} = \Sigma^*(\beta_1, \beta_2)$, the conditional distribution of $n^{1/2} \left(\tilde{\theta}(\beta) - \hat{\theta}^O(\beta) \right)$ is asymptotically equivalent to the unconditional distribution of $n^{1/2} \left(\hat{\theta}(\beta) - \theta^*(\beta) \right)$.

BIBLIOGRAPHY

BIBLIOGRAPHY

- [1] Cai, T. and Hall, P. (2005). Prediction in functional linear regression. *Annals of Statistics*, 34, 2159-2179.
- [2] Cai, T. and Yuan, M. (2010). Nonparametric covariance function estimation for functional and longitudinal data. *Technical report*.
- [3] Cao, G., Wang, L., Li, Y. and Yang, L. (2012). Spline confidence envelopes for covariance function in dense functional/longitudinal data. *Manuscript*.
- [4] Cao, G., Yang, L. and Todem, D. (2012). Simultaneous inference for the mean function of dense functional data. *Journal of Nonparametric Statistics*, 24, 359-377
- [5] Cao, G., Wang, J., Wang, L. and Todem, D. (2012). Spline confidence bands for functional derivatives. *Journal of Statistical Planning and Inference*, 104, 1557-1570.
- [6] Cao, G., Todem, D., Yang, L. and Fine, J. (2012). Evaluating statistical hypotheses using weakly identifiable estimating functions. *Scandinavian Journal of Statistics*, accepted.
- [7] Chambers, R. L. and Welsh, A. H. (1993). Log-linear models for survey data with non-ignorable non-response. *Journal of the Royal Statistical Society, Series B: Statistical Methodology*, 55, 157-170.
- [8] Claeskens, G. and Van Keilegom, I. (2003). Bootstrap confidence bands for regression curves and their derivatives. *Annals of Statistics*, 31, 1852-1884.
- [9] Crainiceanu, C. M., Staicu, A. and Di, C. (2009). Generalized multilevel functional regression, *Journal of the American Statistical Association*, 104, 1550-1561
- [10] Copas, J. (1999). What works?: Selectivity models and meta-analysis. *Journal of the Royal Statistical Society, Series A: Statistics in Society*, 162, 95-109.

- [11] Copas, J. B. and Li, H. G. (1997). Inference for non-random samples. *Journal of the Royal Statistical Society, Series B: Methodological*, 59, 55-77.
- [12] Csörgő, M. and Révész, P. (1981). *Strong Approximations in Probability and Statistics*. Academic Press, New York-London.
- [13] Dacunha-Castelle, D. and Gassiat, E. (1999). Testing the order of a model using locally conic parametrization: population mixtures and stationary ARMA processes. *Annals of Statistics*, 27, 1178-1209.
- [14] Davies, R. B. (1977). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika*, 64, 247-254.
- [15] Davies, R. B. (1987). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika*, 74, 33-43.
- [16] de Boor, C. (2001) *A Practical Guide to Splines*. Springer-Verlag, New York.
- [17] Degras, D. A. (2011). Simultaneous confidence bands for nonparametric regression with functional data. *Statistica Sinica*, 21, 1735-1765.
- [18] DeVore, R. and Lorentz, G. (1993). *Constructive Approximation: Polynomials and Splines Approximation*. Springer-Verlag, Berlin.
- [19] Efron, B. and Tibshirani, R. (1993). *An Introduction to the Bootstrap*. Chapman & Hall Ltd.
- [20] Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and its Applications*. Chapman & Hall, London.
- [21] Ferraty, F. and Vieu, P. (1996). *Nonparametric Functional Data Analysis: Theory and Practice*. Springer, New York.
- [22] Gasser, T. and Müller, H. G. (1984). Estimating regression functions and their derivatives by the kernel method. *Scandinavian Journal of Statistics*, 11, 171-185.
- [23] Gong, G. and Samaniego, F. J. (1981). Pseudo maximum likelihood estimation: Theory and applications. *The Annals of Statistics*, 9, 861-869.

- [24] Hall, P. and Hosseini-Nasab, M. (2006). On properties of functional principal components analysis. *Journal of the Royal Statistical Society: Series B*, 68, 109-126.
- [25] Hall, P., Müller, H. G. and Wang, J. L. (2006). Properties of principal component methods for functional and longitudinal data analysis. *Annals of Statistics*, 34, 1493-1517.
- [26] Hall, P., Müller, H. G. and Yao, F. (2009). Estimation of functional derivatives. *Annals of Statistics*, 37, 3307-3329.
- [27] Hansen, B. E. (1996). Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica*, 61, 413-430.
- [28] Härdle, W. (1989). Asymptotic maximal deviation of M-smoothers. *Journal of Multivariate Analysis*, 29, 163-179.
- [29] Huang, J. (2003). Local asymptotics for polynomial spline regression. *Annals of Statistics*, 31, 1600-1635.
- [30] Huber, P. J. (1967). *The behavior of the maximum likelihood estimates under non-standard conditions* (in Proceedings of the fifth Berkeley Symposium in mathematical Statistics and Probability, ed.). Berkeley: University of California Press.
- [31] Hürny, C., Bernhard, J., Gelber, R., Coates, A., Castiglione, M., Isley, M., Dreher, D., Peterson, H., Goldhirsch, A. and Senn, H. (1992). Quality of life measures for patients receiving adjuvant therapy for breast cancer: an international trial. *Eur.J. Cancer*, 28, 118-124.
- [32] James, G. M., Hastie, T. and Sugar, C. (2000). Principal component models for sparse functional data. *Biometrika*, 87, 587-602.
- [33] James, G. and Hastie, T. (2001). Functional linear discriminant analysis for irregularly sampled curves. *Journal of the Royal Statistical Society Series B*, 63, 533-550.
- [34] James, G. M. and Silverman, B. W. (2005). Functional adaptive model estimation. *Journal of the American Statistical Association*, 100, 565-576.
- [35] Kenward, M. G., Goetghebeur, E. J. T. and Molenberghs, G. (2001). Sensitivity for incomplete categorical data. *Statistical Modelling*, 1, 31-48.

- [36] Li, Y. and Hsing, T. (2010a). Uniform convergence rates for nonparametric regression and principal component analysis in functional/longitudinal data. *Annals of Statistics*, 38, 3321-3351.
- [37] Li, Y. and Hsing, T. (2010b). Deciding the dimension of effective dimension reduction space for functional and high-dimensional data. *Annals of Statistics*, 38, 3028-3062.
- [38] Li, Y., Wang, N. and Carroll, R. J. (2010). Generalized functional linear models with semiparametric single-index interactions, *Journal of the American Statistical Association*, 105, 621-633.
- [39] Little, R. J. A. and Rubin, D. B. (2002). *Statistical Analysis with Missing Data*. John Wiley & Sons.
- [40] Liu, B. and Müller, H. G. (2009). Estimating derivatives for samples of sparsely observed functions, with application to online auction dynamics. *Journal of the American Statistical Association*, 104, 704-717.
- [41] Lu, G. and Copas, J. B. (2004). Missing at random, likelihood ignorability and model completeness. *Annals of Statistics*, 32, 754-765.
- [42] Ma, S., Yang, L. and Carroll, R. J. (2012). A simultaneous confidence band for sparse longitudinal regression. *Statistica Sinica*, 22, 95-122.
- [43] Morris, J. S. and Carroll, R. J. (2006). Wavelet-based functional mixed models. *Journal of the Royal Statistical Society, Series B*, 68, 179-199.
- [44] Müller, H. G. (2009). *Functional modeling of longitudinal data*. In: *Longitudinal Data Analysis (Handbooks of Modern Statistical Methods)*, Ed. Fitzmaurice, G., Davidian, M., Verbeke, G., Molenberghs, G., Wiley, New York, 223-252.
- [45] Parzen, M. I., Wei, L. J. and Ying, Z. (1994). A resampling method based on pivotal estimating functions. *Biometrika*, 81, 341-350.
- [46] Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis*. Springer, New York.
- [47] Rice, J. A. and Wu, C. O. (2001). Nonparametric mixed effects models for unequally sampled noisy curves. *Biometrics*, 57, 253-259.

- [48] Ritz, C. and Skovgaard, I. M. (2005). Likelihood ratio tests in curved exponential families with nuisance parameters present only under the alternative. *Biometrika*, 92, 507–517.
- [49] Rotnitzky, A., Scharfstein, D., Su, T.-L. and Robins, J. (2001). Methods for conducting sensitivity analysis of trials with potentially nonignorable competing causes of censoring. *Biometrics*, 57, 103-113.
- [50] Scharfstein, D. O., Rotnitzky, A. and Robins, J. M. (1999). Adjusting for nonignorable drop-out using semiparametric nonresponse models (C/R: P1121-1146). *Journal of the American Statistical Association*, 94, 1096-1120.
- [51] Schumaker, L. (2007). *Spline Functions: Basic Theory*. Third Edition. Cambridge University Press, Cambridge.
- [52] Shi, M., Weiss, R. E. and Taylor, J. M. G. (1996). An analysis of paediatric CD4 counts for acquired immune deficiency syndrome using flexible random curves. *Journal of the Royal Statistical Society, Series B*, 45, 151-163.
- [53] Song, R., Kosorok, R. and Fine, J. (2009). On asymptotically optimal tests under loss of identifiability in semiparametric models. *Annals of Statistics*, 37, 2409-2444.
- [54] Stone, C. J. (1994). The use of polynomial splines and their tensor products in multivariate function estimation. *Annals of Statistics*, 22, 118-184.
- [55] Todem, D., Fine, J. and Peng, L. (2010). A global sensitivity test for evaluating statistical hypotheses with nonidentifiable models. *Biometrics*, 66, 558-566.
- [56] Troxel, A. B., Harrington, D. P. and Lipsitz, S. R. (1998a). Analysis of longitudinal data with non-ignorable non-monotone missing values. *Journal of the Royal Statistical Society, Series C: Applied Statistics*, 47, 425-438.
- [57] Troxel, A. B., Lipsitz, S. R. and Harrington, D. P. (1998b). Marginal models for the analysis of longitudinal measurements with nonignorable non-monotone missing data. *Biometrika*, 85, 661-672.
- [58] van der Vaart, A. W. and Wellner, J. A. (2000a). *Preservation theorems for Glivenko-Cantelli and uniform Glivenko-Cantelli theorems. High Dimensional Probability II*. Birkhäuser, Boston.: eds: E Giné, DM Mason, JA Wellner.

- [59] van der Vaart, A. W. and Wellner, J. A. (2000b). *Weak convergence and empirical processes*. New York: Springer.
- [60] Wahba, G. (1990). *Spline Models for Observational Data.*, 59. SIAM, Philadelphia, PA.
- [61] Wang, J. and Yang, L. (2009a). Polynomial spline confidence bands for regression curves. *Statistica Sinica*, 19, 325-342.
- [62] Wang, L. and Yang, L. (2009b). Spline estimation of single index model. *Statistica Sinica*, 19, 765-783.
- [63] Wang, N., Carroll, R. J. and Lin, X. (2005). Efficient semiparametric marginal estimation for longitudinal/clustering data. *Journal of the American Statistical Association*, 100, 147-157.
- [64] White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica*, 50, 141-161.
- [65] Xia, Y. (1998). Bias-corrected confidence bands in nonparametric regression. *Journal of the Royal Statistical Society, Series B*, 60, 797-811.
- [66] Yao, F. and Lee, T. C. M. (2006). Penalized spline models for functional principal component analysis. *Journal of the Royal Statistical Society, Series B*, 88, 3-25.
- [67] Yao, F., Müller, H. G. and Wang, J. L. (2005a). Functional linear regression analysis for longitudinal data. *Annals of Statistics*, 33, 2873-2903.
- [68] Yao, F., Müller, H. G. and Wang, J. L. (2005b). Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association*, 100, 577-590.
- [69] Zhao, X., Marron, J. S. and Wells, M. T. (2004). The functional data analysis view of longitudinal data. *Statistica Sinica*, 14, 789-808.
- [70] Zhao, Z. and Wu, W. (2008). Confidence bands in nonparametric time series regression. *Annals of Statistics*, 36, 1854-1878.
- [71] Zhou, L., Huang, J. and Carroll, R. J. (2008). Joint modelling of paired sparse functional data using principal components. *Biometrika*, 95, 601-619.

- [72] Zhou, S., Shen, X. and Wolfe, D. A. (1998). Local asymptotics of regression splines and confidence regions. *Annals of Statistics*, 26, 1760-1782.
- [73] Zhou, S. and Wolfe, D. A. (2000). On derivative estimation in spline regression. *Statistica Sinica*, 10, 93-108.
- [74] Zhu, H., Zhang, H. (2006). Generalized score test of homogeneity for mixed effects models. *Annals of Statistics*, 34, 1545–1569.