This is to certify that the
dissertation entitled

APPLICATION OF SIMULTANEOUS CONFIDENCE BANDS
IN STATISTICAL INFERENCE FOR HETEROSCEDASTIC,
HIGH DIMENSIONAL AND FUNCTIONAL DATA

presented by

Qiongxia Song

has been accepted towards fulfillment
of the requirements for the

Doctoral degree in Statistics

Major Professor's Signature

8/9/2010

Date

PLACE IN RETURN BOX to remove this checkout from your record.
TO AVOID FINES return on or before date due.
MAY BE RECALLED with earlier due date if requested.

DATE DUE	DATE DUE	DATE DUE

APPLICATION OF SIMULTANEOUS CONFIDENCE BANDS IN STATISTICAL INFERENCE FOR HETEROSCEDASTIC, HIGH DIMENSIONAL AND FUNCTIONAL DATA

By

Qiongxia Song

A DISSERTATION

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

DOCTOR OF PHILOSOPHY

Statistics

2010

ABSTRACT

APPLICATION OF SIMULTANEOUS CONFIDENCE BANDS IN STATISTICAL INFERENCE FOR HETEROSCEDASTIC, HIGH DIMENSIONAL AND FUNCTIONAL DATA

By

Qiongxia Song

This dissertation studies simultaneous confidence bands for heteroscedastic, high dimensional and functional data with their applications in statistical inference.

Nonparametric simultaneous confidence bands are a powerful tool of global inference for functions. Chapter 1 provides a bird's eye view of the state-of-the-art and challenges for constructing such confidence bands, a brief introduction to later chapters. An introduction to the nonlinear spline smoothing and local linear smoothing is also provided in chapter 1.

In Chapter 2, asymptotically exact and conservative confidence bands are obtained for possibly heteroscedastic variance function, using piecewise constant and piecewise linear spline estimation, respectively. The variance estimation possesses oracle efficiency and the widths of the confidence bands are of optimal order. Simulation experiments provide strong evidence that corroborates the asymptotic theory while the computing is extremely fast. Also, in simulation, the proposed confidence bands is compared with some other testing heteroscedasticity methods. As illustration of the applicability of the methods, the linear spline band has been applied to test for heteroscedasticity in a fossil data and in the motorcycle data.

Chapter 3 provides the method for constructing simultaneous confidence bands for nonlinear additive autoregressive models (NAAR), which have found wide use in recent years to reduce dimension in nonparametric smoothing of time series. Under weak conditions of smoothness and mixing, we propose spline-backfitted spline (SBS) estimators of the component functions for nonlinear additive autoregressive model

that is both computationally expedient for analyzing high dimensional large time series data, and theoretically reliable as the estimator is oracally efficient and comes with asymptotically simultaneous confidence band. Simulation evidence strongly corroborates with the asymptotic theory.

Chapter 4 focuses on constructing confidence bands for densely spaced functional data. We illustrate the use of local linear smoothing to construct simultaneous confidence bands for the mean function. Our approach works under mild conditions for the case of densely spaced observations and differs from sparse and irregular longitudinal data. Simulation experiments provide strong evidence that corroborates the asymptotic theory. The confidence band procedure is illustrated by analyzing the near infrared spectroscopy data.

ACKNOWLEDGMENT

I would like to thank many people who have helped me on the path towards this dissertation. Most importantly, I want to thank my thesis adviser, Prof. Lijian Yang, who set me on the right path after I have joined his group. His systematic guidance and constant push not only work as the major sources of the completion of this dissertation and its corresponding publications, but also cultivates me from a student to an independent researcher. His support and encouragement in every research related aspect sustain my confidence in entering a highly competitive academic area.

I also wish to express my gratitude to my dissertation committee, Professor Connie Page, Professor Yuehua Cui and Professor Timothy Vogelsang, for sparing their precious time to serve on my committee and giving valuable comments and suggestions.

I am grateful to the entire faculty and staff in the Department of Statistics and Probability who have taught me and assisted me during my study at MSU. My special thanks go to Prof. James Stapleton for his numerous help, constant support and encouragement.

My PhD degree will not be completed without the help and support from my family. My parents and my sister have been giving their support from the other side of the earth. My husband, Weihua Geng, is the one I can really rely on when I have to face difficulties and challenges. I owe them a great deal.

Michigan State University is such a great university to study, research and live. In particular, I am grateful to the graduate school and the Department of Statistics and Probability in terms of providing me Dissertation Completion Fellowship. This dissertation is also supported by Prof. Yang's NSF grant award DMS 0706518. I would like thank the group members for their generous help, and they are Dr. Lan Xue, Dr. Jing Wang, Dr. Lily Wang, Dr. Rong Liu, Mrs. Shujie Ma, Mr. Shuzhuan Zheng and Mrs. Guanqun Cao.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction to confidence bands

1.1 Status and challenges

Nonparametric regression has gained much attention since it relaxes the usual assumption of linearity and enables one to explore the data more flexibly. Many of the properties of nonparametric regression estimators have been thoroughly investigated. However, as Eubank and Speckman [13] pointed out, techniques for constructing interval estimates to accompany the regression function estimators have been slow to develop, even in the case of independent and identically distributed (IID) observations.

Consider the nonparametric regression model

$$Y_i = m\left(X_i\right) + \varepsilon_i, \quad i = 1, 2, \cdots, n. \tag{1.1}$$

A natural definition for asymptotic exact (conservative) $100(1-\alpha)\%$ confidence bands for an unknown function $m(x)$ over interval $[a, b]$ consists of an estimator $\hat{m}(x)$ of $m(x)$, lower and upper confidence limit ($l_{n,L}(x)$ and $l_{n,U}(x)$) at every $x \in [a, b]$, such that,

$$\begin{aligned}\lim_{n\to\infty} P\left\{m(x) \in \left[\hat{m}(x) - l_{n,L}(x), \hat{m}(x) + l_{n,U}(x)\right], \forall x \in [a,b]\right\} &= 1-\alpha,\\ \liminf_{n\to\infty} P\left\{m(x) \in \left[\hat{m}(x) - l_{n,L}(x), \hat{m}(x) + l_{n,U}(x)\right], \forall x \in [a,b]\right\} &\geq 1-\alpha.\end{aligned}$$

Confidence bands are closely related to confidence intervals, which represent the uncertainty in an estimate of a single numerical value. While, confidence bands arise whenever a statistical analysis focuses on estimating a function or constructing interval estimates. A confidence band is used in statistical analysis to represent the uncertainty in an estimate of a curve or function based on limited or noisy data. For instance, with the simultaneous confidence bands, we can test whether m is of certain parametric form: $H_0 : m = m_\theta$, where $\theta \in \Theta$ and Θ is a parameter space. For example, we can test whether $m = c$ with c a constant or we can test whether $m(x) = \beta_0 + \beta_1 x$ with (β_0, β_1) linear regression estimate. If so, then we accept at level $1-\alpha$ the null hypothesis that m is constant or linear. Otherwise H_0 is rejected.

Construction simultaneous confidence bands has been developed slowly since it is difficult to establish asymptotic sample distribution theory for nonparametric regression estimates. In the last two decades, many statisticians have worked on the theory and applications of nonparametric simultaneous confidence bands, see [7, 13, 16, 22, 23, 25, 74, 75, 87, 89].

All these methods are local polynomial smoothing based. Confidence bands of kernel type estimators are computationally intensive since a least square estimation has to be done at every point. In contrast, it is enough to solve only one least square to get the polynomial spline estimator. Recently, some research has been done to provide confidence bands results using polynomial spline smoothing. See, Wang and Yang [70] and Wang and Yang [72]. For the application, see Wang et. al. [30]. In this thesis, I tackle this difficult problem in many scenarios, using polynomial spline smoothing mainly. In this introductory chapter, I state, without proof, those basic facts about our target models. We construct confidence bands for all these models with statistical inference.

1.2 Nonparametric smoothing

Smoothing techniques make an important class of tools for identifying the true signal hidden in highly noisy data. They offer the art of nonlinear curve/surface estimation by relaxing the linear assumption in regression and have very broad applications in many areas. I give a brief introduction to the smoothing techniques used in our research and analysis, namely regression splines and kernel smoothers.

Regression spline smoothing is a projection method for fitting splines. Let $\{X_i, Y_i\}_{i=1}^n$ be a strictly stationary process. Assume that X_i, $1 \le i \le n$ are supported on a compact interval $[a, b]$. Polynomial splines begin by choosing a set of knots, and a set of basis functions spanning a set of piecewise polynomials satisfying continuity and smoothness constraints.

Let $a = t_{1-k} = \cdots = t_0 < t_1 < ... < t_{N+1} = \cdots = t_{N+k} = b$ be a sequence of equally spaced knots, dividing $[a, b]$ into $(N+1)$ subintervals of length $h = (b-a)/(N+1)$. The j-th B-spline of order k for the knot sequence T denoted by $B_{j,k}$ is recursively defined by the de Boor [10], i.e.

$$B_{j,k}(u) = \frac{(u-t_j)B_{j,k-1}(u)}{t_{j+k-1}-t_j} - \frac{(u-t_{j+k})B_{j+1,k-1}(u)}{t_{j+k}-t_{j+1}}, 1-k \le j \le N$$

for $k > 1$, with

$$B_{j,1}(u) = \begin{cases} 1 & t_j \le u < t_{j+1}, \\ 0 & \text{otherwise}, \end{cases}$$

We denote by $G^{(p-2)}[a, b]$ the linear space spanned by $\left\{B_{j,p}\left(x_\gamma\right)\right\}_{J=1-k}^{N}$, whose elements are $C^{(p-2)}[a,b]$ functions that are polynomials of degree $p-1$ on each subinterval. We denote by

$$C^{(p)}[a,b] = \{m|\text{the p-th order derivative of m is continuous on}[a,b]\}.$$

The polynomial spline estimator for regression model (1.1) is

$$\hat{m}_k(\cdot) = \underset{g(\cdot)\in G^{(k-2)}[a,b]}{\text{argmin}} \sum_{i=1}^{n} \{Y_i - g(X_i)\}^2, k > 0$$

Locally linear smoothing is used for the last chapter to develop the confidence bands for functional data. This smoother combines the strict local nature of the data and the smooth weights of kernel smoothers. Kernel smoothers are expensive to compute ($O(n^2)$ for the whole sequence), but are visually smooth if the kernel is smooth.

A local linear approximation is

$$\mu(x_i) \approx a + b(x_i - x)$$

The local approximation can be fitted by locally weighted least squares. A weight function and bandwidth are defined as kernel regression. In the case of local linear regression, coefficient estimates $\hat{a}$ and $\hat{b}$ are chosen to minimize

$$\left(\hat{a}, \hat{b}\right) = \arg\min \sum_{i=1}^{n} \{y_i - a - b\,(x_i - x)\}^2 K_h\,(x_i - x)$$

with $K_h\,(u) = \frac{1}{h} K\left(\frac{u}{h}\right)$, $h = h_n \to 0$, as $n \to \infty$. When $\left(\mathbf{X}^T\mathbf{W}\mathbf{X}\right)$ is invertible, one has the explicit representation

$$\hat{a} = e_0^T \left(\mathbf{X}^T\mathbf{W}\mathbf{X}\right)^{-1} \mathbf{X}^T\mathbf{W}\mathbf{Y}$$

in which $\mathbf{Y} = (Y_1, \ldots, Y_n)^T$, $e_0^T = (1, 0)$, and the design matrix $\mathbf{X}$ is

$$\mathbf{X} = \begin{pmatrix} 1 & (x_1 - x) \\ \vdots & \vdots \\ 1 & (x_n - x) \end{pmatrix}_{n \times 2},$$

and $\mathbf{W} = \text{diag}\left\{K\left(\frac{x_i - x}{h}\right)\right\}_{i=1}^{n}$.

1.3 Variance function bands

The importance of being able to detect heteroscedasticity in regression is widely recognized because of efficient inference for the regression function requires that heteroscedasticity is taken into account. In many applications of regression models the usual assumptions of homoscedastic disturbances cannot be guaranteed a priori. Although the problem of testing hypothesis regarding the regression function has been discussed by many researchers much less attention has been paid to the problem of testing hypotheses regarding the variance structure in a nonparametric regression model. By constructing confidence bands for variance function, we provide a simple consistent test for heteroscedasticity in a nonparametric regression set-up.

In the second chapter, we propose polynomial spline confidence bands for heteroscedastic variance function in a nonparametric regression model, and the result is the only existing confidence band result for variance functions. The greatest advantages of polynomial spline estimation are its simplicity of implementation and fast computation. It is desirable from a theoretical as well as a practical point of view to have confidence bands for polynomial spline estimators.

We assume that observations $\{(X_i, Y_i)\}_{i=1}^{n}$ and unobserved errors $\{\varepsilon_i\}_{i=1}^{n}$ are i.i.d. copies of (X, Y, ε) satisfying the regression model (1.1) where the error ε is conditional noise, with $E(\varepsilon|X) \equiv 0$, $E\left(\varepsilon^2|X\right) \equiv \sigma^2(X)$. We constructed a simultaneous confidence band for $\sigma^2(x)$ over $[a,b]$. In addition, the proposed variance estimator is asymptotically as efficient as the infeasible estimator, i.e., the asymptotic mean squared error is as small as if the conditional mean function $m(x)$ is given (equivalently, as if the unobservable error ε is actually observed).

We applied our result on a motorcycle data. The result shows that with a p-value as small as 0.008, one rejects the null hypotheses that the conditional variance function of the data is a constant as no horizontal line can be squeezed into the 99.2% variance function confidence band. The details of the theoretic results and applications are the content of the chapter two.

1.4 SBS estimate and NAAR models bands

Non- and semiparametric smoothing has been proven to be useful for analyzing complex time series data due to the flexibility to "let the data speak for themselves". One unavoidable issue in high dimensional smoothing is the "curse of dimensionality", i.e., the poor convergence rate of nonparametric estimation of multivariate functions. Additive regression models has been found wide use in recent years to reduce dimension in nonparametric smoothing of time series.

A nonlinear additive autoregressive model (NAAR) is of the form

$$Y_i = m(\mathbf{X}_i) + \varepsilon_i, \quad m(x_1, ..., x_d) = c + \sum_{\gamma=1}^{d} m_\gamma(x_\gamma), \tag{1.2}$$

where the sequence $\left\{Y_i, \mathbf{X}_i^T\right\}_{i=1}^n$ is a length n realization of a $(d+1)$-dimensional strictly stationary process, the d-variate functions $m(\cdot)$ and $\sigma(\cdot)$ are the mean and standard deviation of the response Y_i conditional on the predictor vector $\mathbf{X}_i = \{X_{i1}, ..., X_{id}\}^T$, and $E(\varepsilon_i|\mathbf{X}_i) = 0, E\left(\varepsilon_i^2|\mathbf{X}_i\right) = \sigma^2(\mathbf{X}_i)$. In the context of

NAAR, each predictor $X_{i\gamma}, 1 \leq \gamma \leq d$ can be observed lagged values of Y_i, such as $X_{i\gamma} = Y_{i-\gamma}$, or of a different times series.

Inference of model (1.2) centers on the estimation and testing of $\{m_\gamma(\cdot)\}_{\gamma=1}^d$. The two-step estimators for model (1.2) possess oracle efficiency. If all components $\left\{m_\beta(\cdot)\right\}_{\beta=1,\beta\neq\gamma}^d$ and the constant c were known and removed from the responses, one could estimate $m_\gamma(\cdot)$ from the univariate data $\left\{Y_{i\gamma}, X_{i\gamma}\right\}_{i=1}^n$ in which $\left\{Y_{i\gamma}\right\}_{i=1}^n$ are latent oracle responses to the γ-th covariate $\left\{X_{i\gamma}\right\}_{i=1}^n$,

$$Y_{i\gamma} = m_\gamma\left(X_{i\gamma}\right) + \varepsilon_i = Y_i - c - \sum_{\beta=1,\beta\neq\gamma}^{d} m_\beta\left(X_{i\beta}\right), 1 \leq i \leq n, 1 \leq \gamma \leq d.$$

For the NAAR time series models, however, none of the existing methods provide any simultaneous confidence band for $m_\gamma(\cdot)$. To address this need, we propose an all new spline+spline oracally efficient estimator that is theoretically superior as it comes with an asymptotically simultaneous confidence band for $m_\gamma(\cdot)$, and also computationally more expedient than any existing estimators due to the use of spline instead of kernel in all steps.

1.5 Functional data bands

Traditional statistical methods fail often as we deal with functional data. Indeed, if for instance we consider a sample of finely discretized curves, two crucial statistical problems appear. The first comes from the ratio between the size of the sample and the number of variables (each real variable corresponding to one discretized point). The second, is due to the existence of strong correlations between the variables and becomes an ill-conditioned problem in the context of multivariate linear model. So, there is a real necessity to develop statistical methods/models in order to take into account the functional structure of this kind of data.

Functional data with different design are increasingly common in modern data analysis. A simultaneous confidence band for this data set has been more and more in need. A functional data set has the form $\left\{X_{ij}, Y_{ij}\right\}, 1 \le i \le n, 1 \le j \le N$, in which N observations are taken for each subject, with X_{ij} and Y_{ij} the j^{th} predictor and response variables, respectively, for the i^{th} subject. In this paper we only deal with the equally spaced design. Without loss of generality, the predictor X_{ij} takes values $\{1/N, 2/N, ..., N/N\}$ for the i^{th} subject, $i = 1, 2, ..., n$. For the i^{th} subject, its sample path $\left\{j/N, Y_{ij}\right\}$ is the noisy realization of a continuous time stochastic process $\xi_i(x)$ in the sense that $Y_{ij} = \xi_i(j/N) + \sigma(j/N)\varepsilon_{ij}$,with errors ε_{ij} satisfying $E\left(\varepsilon_{ij}\right) = 0, E(\varepsilon_{ij}^2) = 1$, and $\{\xi_i(x), x \in \mathcal{X}\}$ are iid copies of a process $\{\xi(x), x \in \mathcal{X}\}$ which is L^2, i.e., $E\int_{\mathcal{X}} \xi^2(x)dx < +\infty$.

For the standard process $\{\xi(x), x \in \mathcal{X}\}$, one defines the mean function $m(x) = E\{\xi(x)\}$ and the covariance function $G\left(x, x'\right) = \text{cov}\left\{\xi(x), \xi(x')\right\}$. Let sequences $\{\lambda_k\}_{k=1}^{\infty}, \{\psi_k(x)\}_{k=1}^{\infty}$ be the eigenvalues and eigenfunctions of $G\left(x, x'\right)$ respectively, in which $\lambda_1 \ge \lambda_2 \ge \cdots \ge 0, \sum_{k=1}^{\infty} \lambda_k < \infty$, $\{\psi_k\}_{k=1}^{\infty}$ form an orthonormal basis of $L^2(\mathcal{X})$ and $G\left(x, x'\right) = \sum_{k=1}^{\infty} \lambda_k \psi_k(x) \psi_k\left(x'\right)$, which implies that $\int G\left(x, x'\right) \psi_k\left(x'\right) dx' = \lambda_k \psi_k(x)$.

The process $\{\xi_i(x), x \in \mathcal{X}\}$ allows the Karhunen-Loève L^2 representation

$$\xi_i(x) = m(x) + \sum_{k=1}^{\infty} \xi_{ik} \phi_k(x),$$

where the random coefficients ξ_{ik} are uncorrelated with mean 0 and variances 1, and the functions $\phi_k = \sqrt{\lambda_k}\psi_k$. In what follows, we assume that $\lambda_k = 0$, for $k > \kappa$, where κ is a positive integer or $+\infty$, thus $G(x, x') = \sum_{k=1}^{\kappa} \phi_k(x)\phi_k\left(x'\right)$ and the data generating process is now written as

$$Y_{ij} = m(j/N) + \sum_{k=1}^{\kappa} \xi_{ik}\phi_k(j/N) + \sigma(j/N)\varepsilon_{ij}. \tag{1.3}$$

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

Two distinct types of functional data have been studied: sparse longitudinal data ($1 \leq j \leq N_i$ and N_i's are iid copies of an integer valued positive random variable) and dense functional data ($N_i \to \infty$ as $n \to \infty$). For the dense functional data, strong uniform convergence rates are developed for local-linear smooth estimators, but without uniform confidence bands. The fact that simultaneous confidence band has not been established for functional data analysis is certainly not due to lack of interesting applications, but to the greater technical difficulty to formulate such bands for functional data and establish their theoretical properties. In this thesis, we present simultaneous confidence bands for $m(x)$ in dense longitudinal data given in (1.3) via local linear smoothing approach.

Chapter 2

Spline confidence bands for variance function

2.1 Introduction

Quantification of local variability of regression data is an indispensable ingredient for many scientific investigations. The most intuitive measure of such is the conditional variance function, whose estimation has been the subject of Müller and Stadtmüller [50], Hall and Carroll [20], Ruppert et. al. [61] and Fan and Yao [15], which employed kernel type smoothing methods for the nonparametric variance function. Similar smoothing methods have also been used to estimate noise-to-signal ratio in Yao and Tong [83] with applications to time series volatility estimation. These existing works estimate the conditional variance function via kernel smoothing of the squares of residuals from an initial kernel smoothing of the regression data. Such two-stage smoothing technique has also been used in estimating homoscedastic variance in Hall and Marron [21]. More recently, a new approach to variance estimation based on differencing has been proposed, which can successfully handle serially correlated errors, see Dahl and Levine [9] and Brown and Levine [4].

What has been lacking is uniform confidence band for the whole variance curve over an entire bounded range, and explicit formula for the estimated variance function. The former is useful for making inference on the shape of the variance function, such as testing of homoscedasticity, while the latter is appealing to practitioners without much statistics expertise but wish to implement nonparametric procedures. Uniform confidence bands have been constructed for conditional mean function in Hall and Titterington [26], Härdle [23], Xia [75], Claeskens and Van Keilegom [7], and for probability density function in Bickel and Rosenblatt [1]. All these and other related works such as Mack and Silverman [46], are based on kernel smoothing and make use of the "Hungarian embedding" type results such as in Rosenblatt [59] and Tusnády [69]. More recently, Zhao et. al. [89], Wang and Yang [70] constructed confidence bands for conditional mean function using polynomial spline method with explicit formulae for both the estimated conditional mean function and the confidence band. In particular, Wang and Yang [70] allows for heteroscedastic and nonnormal errors, and is useful for testing hypothesis on the shape of regression curve.

In this chapter, we propose polynomial spline confidence bands for heteroscedastic variance function in a nonparametric regression model. The greatest advantages of polynomial spline estimation are its simplicity of implementation and fast computation, see for instance, Stone [67] and Huang [28] for the basic theory of polynomial spline smoothing, and Xue and Yang [76] for computing speed comparison of spline vs. kernel smoothing. Hence, it is desirable from a theoretical as well as a practical point of view to have confidence bands for polynomial spline estimators.

We assume that observations $\{(X_i, Y_i)\}_{i=1}^n$ and unobserved errors $\{\varepsilon_i\}_{i=1}^n$ are i.i.d. copies of (X, Y, ε) satisfying the regression model

$$Y = m(X) + \varepsilon, \tag{2.1}$$

where the error ε is conditional noise, with $E(\varepsilon|X) \equiv 0$, $E\left(\varepsilon^2|X\right) \equiv \sigma^2(X)$, see Assumption (A4) in Section 2.2 for details. The conditional mean and conditional variance functions $m(x)$ and $\sigma^2(x)$, defined on interval $[a,b]$, need not be of any known form.

Our goal is to construct a simultaneous confidence band for $\sigma^2(x)$ over $[a,b]$. In addition, the proposed variance estimator is asymptotically as efficient as the infeasible estimator, i.e., the asymptotic mean squared error is as small as if the conditional mean function $m(x)$ is given (equivalently, as if the unobservable error ε is actually observed). As an example, consider the motor cycle data, Figure 2.4 shows that with a p-value as small as 0.008, one rejects the null hypotheses that the conditional variance function of the data is a constant as no horizontal line can be squeezed into the 99.2% variance function confidence band. For other methods of testing the heteroscedasticity or the lack-of-fit of regression function, see Dette and Munk [11] and Bissantz et. al. [2], and Section 2.5 for simulation comparison of our method with that of Dette and Munk [11].

The chapter is based on a published work Song and Yang [63], and the chapter is organized as follows. In Section 2.2, we state our main results on variance confidence bands using constant/linear splines. In Section 2.3 we investigate the error structure of spline variance estimators leading to insights of proof. We give the actual steps to implement the confidence band in Section 2.4, and in Section 2.5, we report simulation results and applications to a fossil data and the well known motorcycle data. Appendix contains all the technical proofs needed for the main results.

2.2 Main results

An asymptotic exact and conservative $100(1-\alpha)\%$ confidence band for the unknown $\sigma^2(x)$ over the interval $[a,b]$ consists of an estimator $\hat{\sigma}^2(x)$ of $\sigma^2(x)$, lower and upper confidence limits $\hat{\sigma}^2(x) - l_{n,L}(x)$, $\hat{\sigma}^2(x) + l_{n,U}(x)$ at every $x \in [a,b]$ such that

$$
\begin{aligned}
\lim_{n\to\infty} P\left\{\sigma^2(x) \in \left[\hat{\sigma}^2(x) - l_{n,L}(x), \hat{\sigma}^2(x) + l_{n,U}(x)\right], \forall x \in [a,b]\right\} &= 1-\alpha, \\
\liminf_{n\to\infty} P\left\{\sigma^2(x) \in \left[\hat{\sigma}^2(x) - l_{n,L}(x), \hat{\sigma}^2(x) + l_{n,U}(x)\right], \forall x \in [a,b]\right\} &\geq 1-\alpha
\end{aligned}
$$

respectively.

If the mean function $m(x)$ were known, one could compute the errors $\varepsilon_i = Y_i - m(X_i), 1 \leq i \leq n$ and make use of the fact that $E\left(\varepsilon_i^2 \,|X_i = x\right) \equiv \sigma^2(x)$ to carry out polynomial spline regression of the data $\{(X_i, Z_i)\}_{i=1}^n$, in which $Z_i = \varepsilon_i^2$ are the squared errors. Specifically, one could define the "infeasible estimator" of the variance function as $\tilde{\sigma}_{p_2}^2(x) = \operatorname{argmin}_{g \in G_{N_2}^{(p_2-2)}[a,b]} \sum_{i=1}^n \{Z_i - g(X_i)\}^2$, in which $G_{N_2}^{(p_2-2)} = G_{N_2}^{(p_2-2)}[a,b]$ is the space of functions that are piecewise polynomials of degree $(p_2 - 1)$ on interval $[a,b]$, defined precisely below, for some positive integer p_2.

To mimic the above unattainable spline smoother, we define

$$
\hat{\sigma}_{p_1,p_2}^2(x) = \operatorname*{argmin}_{g \in G_{N_2}^{(p_2-2)}[a,b]} \sum_{i=1}^n \left\{\hat{Z}_{i,p_1} - g(X_i)\right\}^2, \tag{2.2}
$$

where $\hat{Z}_{i,p_1} = \hat{\varepsilon}_{i,p_1}^2$ are the squares of residuals $\hat{\varepsilon}_{i,p_1}$ obtained from spline regression,

$$
\hat{\varepsilon}_{i,p_1} = Y_i - \hat{m}_{p_1}(X_i), \quad 1 \leq i \leq n, \tag{2.3}
$$

for some positive integer p_1, in which

$$
\hat{m}_{p_1}(x) = \operatorname*{argmin}_{g \in G_{N_1}^{(p_1-2)}[a,b]} \sum_{i=1}^n \{Y_i - g(X_i)\}^2. \tag{2.4}
$$

To introduce spline functions, for the two steps $\nu = 1, 2$, we divide the finite interval $[a, b]$ into $(N_\nu + 1)$ subintervals $J_j = \left[t_j, t_{j+1}\right)$, $j = 0, ..., N_\nu - 1$, $J_{N_\nu} = \left[t_{N_\nu}, b\right]$. A sequence of equally-spaced interior knots $\left\{t_j\right\}_{j=1}^{N_\nu}$, are given as

$$t_0 = a < t_1 < \cdots < t_{N_\nu} < b = t_{N_\nu+1},\ t_j = a + jh_\nu,\ j = 0, 1, ..., N_\nu + 1,$$

in which $h_\nu = (b - a) / (N_\nu + 1)$ is the distance between neighboring knots. We denote by $G_{N_\nu}^{(p_\nu - 2)} = G_{N_\nu}^{(p_\nu - 2)}[a, b]$ the space of functions that are polynomials of degree $(p_\nu - 1)$ on each J_j and have continuous $(p_\nu - 2)$th derivative. For example, $G_{N_\nu}^{-1}$ denotes the space of functions that are constant on each J_j, and $G_{N_\nu}^{0}$ the space of functions that are linear on each J_j and continuous on $[a, b]$.

In what follows, $\|\cdot\|_\infty$ denotes the supremum norm of a function w on $[a, b]$, i.e. $\|w\|_\infty = \sup_{x\in[a,b]} |w(x)|$, and the moduli of continuity of a continuous function w on $[a, b]$ is denoted by $\omega(w, h_\nu) = \max_{x,x'\in[a,b],|x-x'|\leq h_\nu} \left|w(x) - w\left(x'\right)\right|$. That $\lim_{h_\nu\to 0} \omega(w, h_\nu) = 0$ follows from the uniform continuity of w on compact $[a, b]$.

Our approach is to construct the error bound function $l_n(x)$ around the spline estimators $\hat\sigma^2_{p_1,p_2}(x)$. The technical assumptions we need are as follows:

(A1) *The regression function* $m(\cdot) \in C^{(p_1)}[a, b]$.

(A2) *The density function* $f(\cdot)$ *of* X *is continuous and positive on the interval* $[a, b]$.

(A3) *The subinterval length* $h_\nu \sim n^{-1/(2p_\nu+1)}$, *i.e., the number of interior knots* $N_\nu \sim n^{1/(2p_\nu+1)}$, $\nu = 1, 2$

(A4) *The joint distribution* $F(x, \varepsilon)$ *of random variables* (X, ε) *satisfies:*

(a) *There exists a positive value* $\eta > 1/p_2$ *and a finite positive* M_η *such that*

$$\sup_{x\in[a,b]} E\left(|\varepsilon|^{4+2\eta} | X = x\right) < M_\eta.$$

(b) *The error is conditional noise:* $E(\varepsilon|X=x) \equiv 0$, $E\left(\varepsilon^2|X=x\right) \equiv \sigma^2(x)$ *with* $E\left(\varepsilon^4|X=x\right) \equiv \mu_4(x)$ *which is a positive function on* $[a,b]$ *with bounded variation. The variance function* $\sigma^2(\cdot) \in C^{(p_2)}[a,b]$ *and has a positive lower bound on* $[a,b]$.

Assumptions (A1)-(A4) are adapted from [70] for sample $\{(X_i, Z_i)\}_{i=1}^n$. In particular, Assumption (A4) (a) implies that $\text{var}\left(\varepsilon^2|X=x\right) \equiv \mu_4(x) - \sigma^4(x)$ is the conditional variance of $Z = \varepsilon^2$, denoted as $v_Z^2(x)$. We denote also $p_* = \min(p_1, p_2)$, $p^* = \max(p_1, p_2)$, $N_* = \min(N_1, N_2)$, $N^* = \max(N_1, N_2)$. The idea of allowing different degrees of smoothness for m and σ comes from one referee.

To properly define the confidence bands, we denote for any $x \in [a,b]$, define its location and relative position indices $j_\nu(x), r_\nu(x)$ as

$$j_\nu(x) = j_{n\nu}(x) = \min\{[(x-a)/h_\nu], N_\nu\}, \quad r_\nu(x) = \left\{x - t_{j_\nu(x)}\right\}/h_\nu. \quad (2.5)$$

Since any x is between two consecutive knots, it is clear that $t_{j_{n\nu}(x)} \leq x < t_{j_{n\nu}(x)+1}$, $0 \leq r_\nu(x) < 1, \forall x \in [a,b)$, and $r_\nu(b) = 1$. We denote by $\|\phi\|_2$ the theoretical L^2 norm of a function ϕ on $[a,b]$, i.e. $\|\phi\|_2^2 = E\left\{\phi^2(X)\right\} = \int_a^b \phi^2(x) f(x)\,dx$, and the empirical L^2 norm as $\|\phi\|_{2,n}^2 = n^{-1}\sum_{i=1}^n \phi^2(X_i)$. Corresponding inner products are defined by

$$\langle\phi,\varphi\rangle = \int_a^b \phi(x)\varphi(x) f(x)\,dx = E\{\phi(X)\varphi(X)\},$$
$$\langle\phi,\varphi\rangle_n = n^{-1}\sum_{i=1}^n \phi(X_i)\varphi(X_i)$$

for any L^2-integrable functions ϕ, φ on $[a,b]$. Clearly $E\langle\phi,\varphi\rangle_n = \langle\phi,\varphi\rangle$.

Algebra shows that the space $G_{N_\nu}^{(p_\nu-2)}$ can be spanned linearly by the B-spline basis introduced below or the truncated power basis introduced in Section 2.4, see [10]. Hence the same estimator $\hat{m}_{p_1}(x)$ can be expressed as a linear combination of either of the two bases. While the truncated power basis is convenient for implementation, it is easier to work with the B-spline basis for theoretical analysis. The B-spline basis of $G_{N_\nu}^{(-1)}$, the space of piecewise constant splines, are indicator functions of intervals J_j, $b_{j,1}(x) = I_j(x) = I_{J_j}(x), 0 \le j \le N_\nu$. The B-spline basis of $G_{N_\nu}^{0}$, the space of piecewise linear splines, are $\left\{b_{j,2}(x)\right\}_{j=-1}^{N_\nu}$, where

$$b_{j,2}(x) = K\left(\frac{x - t_{j+1}}{h_1}\right), \; j = -1, 0, ..., N_\nu, \quad \text{for } K(u) = (1 - |u|)_+ .$$

Define the rescaled B-spline basis $\left\{B_{j,p_\nu}(x)\right\}_{j=1-p_\nu}^{N_\nu}$ for $G_{N_\nu}^{(p_\nu-2)}$

$$B_{j,p_\nu}(x) \equiv b_{j,p_\nu}(x) \left\|b_{j,p_\nu}\right\|_2^{-1}, \quad 1 - p_\nu \le j \le N_\nu.$$

Obviously all the rescaled basis functions will have theoretical norm 1.

To express the estimator $\hat{m}_{p_1}(x)$ based on the basis $\left\{B_{j,p_1}(x)\right\}_{j=1-p_1}^{N_1}$, we introduce the following vectors in R^n: $\mathbf{Y} = (Y_1, ..., Y_n)^T$,

$$\mathbf{B}_{j,p_1}(\mathbf{X}) = \left\{B_{j,p_1}(X_1), ..., B_{j,p_1}(X_n)\right\}^T, \quad j = 1 - p_1, ..., N_1,$$

and let the design matrix for spline regression be

$$\mathbf{B}_{p_1} = \mathbf{B}_{p_1}(\mathbf{X}) = \left\{\mathbf{B}_{1-p_1,p_1}(\mathbf{X}), ..., \mathbf{B}_{N_1,p_1}(\mathbf{X})\right\},$$

then the estimator $\hat{m}_{p_1}(x)$ in (2.4) is expressed as

$$\begin{aligned}\hat{m}_{p_1}(x) &= \left\{B_{1-p_1,p_1}(x), ..., B_{N_1,p_1}(x)\right\}\left(\mathbf{B}_{p_1}^T \mathbf{B}_{p_1}\right)^{-1} \mathbf{B}_{p_1}^T \mathbf{Y} \\ &= \sum_{j=1-p_1}^{N_1} \hat{\lambda}_{j,p_1} B_{j,p_1}(x),\end{aligned}$$

where the coefficients $\left\{\hat{\lambda}_{1-p_1,p_1},...,\hat{\lambda}_{N_1,p_1}\right\}^T$ are solutions of the following least squares problem

$$\left\{\hat{\lambda}_{1-p_1,p_1},...,\hat{\lambda}_{N_1,p_1}\right\}^T = \underset{R^{N_1+p_1}}{\text{argmin}} \sum_{i=1}^{n} \left\{Y_i - \sum_{j=1-p_1}^{N_1} \lambda_{j,p_1} B_{j,p_1}(X_i)\right\}^2,$$

or equivalently, of the normal equation

$$\begin{aligned} &\left(\left\langle B_{j,p_1}, B_{j',p_1}\right\rangle_n\right)_{j,j'=1-p_1}^{N_1} \left(\hat{\lambda}_{j,p_1}\right)_{j=1-p_1}^{N_1} \\ = &\left(n^{-1}\sum_{i=1}^{n} B_{j,p_1}(X_i)\, Y_i\right)_{j=1-p_1}^{N_1}. \end{aligned}$$

It is straightforward that $\left\langle B_{j,p_1}, B_{j',p_1}\right\rangle \equiv 0, \left|j-j'\right| \geq p_1$, thus the inner product matrix on the left side of the normal equation is diagonal for the constant B spline basis ($p_1 = 1$), and tridiagonal for the linear B spline basis ($p_1 = 2$). According to Lemma 2.2, it is approximated by its deterministic version, whose inverse has an explicit formula given in [70].

For $p_2 = 2$, define the inverse of inner product matrix as S with its 2×2 diagonal submatrices $\left\{S_j, 0 \leq j \leq N_2\right\}$

$$S = \left(s_{j,j'}\right)_{j,j'=-1}^{N_2} = \left(\left\langle B_{j,2}, B_{j',2}\right\rangle\right)^{-1}, \quad S_j = \begin{pmatrix} s_{j-1,j-1} & s_{j-1,j} \\ s_{j,j-1} & s_{j,j} \end{pmatrix}. \tag{2.6}$$

The widths of the confidence bands depend on the variance function:

$$v_{n,1}^2(x) = \frac{\int_{I_{j(x)}} v_Z^2(v)\, f(v)\, dv}{n\|b_{j(x),1}\|_2^2}, \quad v_{n,2}^2(x) = \sum_{j,j',l,l'=-1}^{N_2} \frac{B_{j',2}(x)\, B_{l',2}(x)\, s_{jj'} s_{ll'} v_{jl}}{n}, \tag{2.7}$$

with $j(x)$ defined in (2.5), and $s_{ll'}$ in (2.6), and

$$\left(v_{jl}\right)_{j,j'=-1}^{N_2} = \Sigma = \left\{\int v_Z^2(v)\, B_{j,2}(v)\, B_{l,2}(v)\, f(v)\, dv\right\}_{j,j'=-1}^{N_2}.$$

Under all assumptions, applying [70] to the unobserved sample $\{(X_i, Z_i)\}_{i=1}^n$, an asymptotic $100(1-\alpha)\%$ exact confidence band for $\sigma^2(x)$ over $[a,b]$ is

$$\tilde{\sigma}_1^2(x) \pm v_{n,1}(x)\{2\log(N_2+1)\}^{1/2} d_n(\alpha),$$

where $v_{n,1}(x)$ is given in (2.7) and replaceable by $\hat{v}_Z(x)\left\{\hat{f}(x)\, nh_1\right\}^{-1/2}$ and

$$d_n = 1 - \{2\log(N_2+1)\}^{-1}\left[\log\left\{-\frac{\log(1-\alpha)}{2}\right\} + \frac{\log\log(N_2+1) + \log 4\pi}{2}\right], \quad (2.8)$$

and an asymptotic $100(1-\alpha)\%$ conservative confidence band for $\sigma^2(x)$ over $[a,b]$ is

$$\tilde{\sigma}_2^2(x) \pm v_{n,2}(x)\{2\log(N_2+1) - 2\log\alpha\}^{1/2},$$

where $v_{n,2}(x)$ is as in (2.7), replaceable by $\hat{v}_{n,2}(x)$ in (2.16).

We state our main results in the next theorems.

Theorem 2.1. *Under Assumptions (A1)-(A4), as $n \to \infty$, the spline estimator $\hat{\sigma}_{p_1,p_2}^2$ of σ^2 is asymptotically as efficient as "infeasible estimator", i.e.*

$$\left\|\hat{\sigma}_{p_1,p_2}^2 - \tilde{\sigma}_{p_2}^2\right\|_\infty = \sup_{x\in[a,b]}\left|\hat{\sigma}_{p_1,p_2}^2(x) - \tilde{\sigma}_{p_2}^2(x)\right| = o_P\left(n^{-p_1/(2p_1+1)}\right).$$

Theorem 2.1 and the aforementioned properties of $\hat{\sigma}_{p_1,p_2}^2$, imply the following:

Theorem 2.2. *Under Assumptions (A1)-(A4), an asymptotic* $100(1-\alpha)\%$ *exact or* *conservative confidence band for $\sigma^2(x)$ over the interval $[a, b]$ for $p_2 = 1$ or 2 is*

$$\hat{\sigma}^2_{1,1}(x) \pm v_{n,1}(x)\{2\log(N_2+1)\}^{1/2} d_n(\alpha),$$
$$\hat{\sigma}^2_{2,2}(x) \pm v_{n,2}(x)\{2\log(N_2+1) - 2\log\alpha\}^{1/2},$$

respectively. That is,

$$\lim_{n\to\infty} P\left\{\sigma^2(x) \in \hat{\sigma}^2_{1,1}(x) \pm v_{n,1}(x)\{2\log(N_2+1)\}^{1/2} d_n(\alpha), \forall x \in [a,b]\right\}$$
$$= 1-\alpha,$$
$$\liminf_{n\to\infty} P\left\{\sigma^2(x) \in \hat{\sigma}^2_{2,2}(x) \pm v_{n,2}(x)\left\{2\log\frac{(N_2+1)}{\alpha}\right\}^{1/2}, \forall x \in [a,b]\right\}$$
$$\geq 1-\alpha.$$

The proof of Theorem 2.1 and therefore also of Theorem 2.2, depend on Propositions 2.1, 2.2 and 2.3 in the next section, and the proofs of the propositions are given in the Appendix.

2.3 Error decomposition

In this section, we break the estimation error $\hat{\sigma}^2_{p_2,p_1}(x) - \tilde{\sigma}^2_{p_2}(x)$ into three parts, so we can deal with the convergence rate for each part in the proof. To understand this decomposition, we begin by discussing the spline space $G^{(p_1-2)}$ introduced in Chapter 1 and the representation of the linear spline estimators $\hat{m}_{p_1}(x)$ in (2.4) and $\hat{\sigma}^2_{p_2,p_1}(x)$ in (2.2).

We write $\mathbf{Y}$ as the sum of a signal vector $\mathbf{m}$ and a noise vector $\mathbf{E}$

$$\mathbf{Y} = \mathbf{m} + \mathbf{E},$$
$$\mathbf{m} = \{m(X_1), ..., m(X_n)\}^T,$$

$$\mathbf{E} = \{\varepsilon_1, ..., \varepsilon_n\}^T.$$

Projecting the response $\mathbf{Y}$ onto the linear space $G_n^{(p_1-2)}$ spanned by $\left\{\mathbf{B}_{j,p_1}(\mathbf{X})\right\}_{j=1-p_1}^{N_1}$, one gets

$$\begin{aligned}
\hat{\mathbf{m}}_{p_1} &= \left\{\hat{m}_{p_1}(X_1), ..., \hat{m}_{p_1}(X_n)\right\}^T \\
&= \operatorname*{Proj}_{G_n^{(p_1-2)}} \mathbf{Y} = \operatorname*{Proj}_{G_n^{(p_1-2)}} \mathbf{m} + \operatorname*{Proj}_{G_n^{(p_1-2)}} \mathbf{E}.
\end{aligned}$$

Correspondingly in the space $G^{(p_1-2)}$, one has

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

$$\begin{aligned}
\tilde{m}_{p_1}(x) &= \left[\left\{B_{j,p_1}(x)\right\}_{j=1-p_1}^{N_1}\right]^T \left(\mathbf{B}_{p_1}^T \mathbf{B}_{p_1}\right)^{-1} \mathbf{B}_{p_1}^T \mathbf{m}, \\
\tilde{\varepsilon}_{p_1}(x) &= \left[\left\{B_{j,p_1}(x)\right\}_{j=1-p_1}^{N_1}\right]^T \left(\mathbf{B}_{p_1}^T \mathbf{B}_{p_1}\right)^{-1} \mathbf{B}_{p_1}^T \mathbf{E}. \qquad (2.9)
\end{aligned}$$

Regarding variance, we define $\mathbf{Z} = \left\{\varepsilon_1^2, ..., \varepsilon_n^2\right\}^T$, $\hat{\mathbf{Z}}_{p_1} = \left\{\hat{\varepsilon}_{1,p_1}^2, ..., \hat{\varepsilon}_{n,p_1}^2\right\}^T$, then

$$\begin{aligned}
\tilde{\sigma}_{p_2}^2(x) &= \left[\left\{B_{j,p_2}(x)\right\}_{j=1-p_2}^{N_2}\right]^T \left(\mathbf{B}_{p_2}^T \mathbf{B}_{p_2}\right)^{-1} \mathbf{B}_{p_2}^T \mathbf{Z}, \\
\hat{\sigma}_{p_2,p_1}^2(x) &= \left[\left\{B_{j,p_2}(x)\right\}_{j=1-p_2}^{N_2}\right]^T \left(\mathbf{B}_{p_2}^T \mathbf{B}_{p_2}\right)^{-1} \mathbf{B}_{p_2}^T \hat{\mathbf{Z}}_{p_1}.
\end{aligned}$$

Taking difference,

$$\begin{aligned}
&\hat{\sigma}_{p_2,p_1}^2(x) - \tilde{\sigma}_{p_2}^2(x) \\
&= \left[\left\{B_{j,p_2}(x)\right\}_{j=1-p_2}^{N_1}\right]^T \left(\mathbf{B}_{p_2}^T \mathbf{B}_{p_2}\right)^{-1} \mathbf{B}_{p_2}^T \left(\hat{\mathbf{Z}}_{p_1} - \mathbf{Z}\right)
\end{aligned}$$

$$= \left[\left\{B_{j,p_2}(x)\right\}_{j=1-p_2}^{N_1}\right]^T \left(\mathbf{B}_{p_2}^T \mathbf{B}_{p_2}\right)^{-1} \mathbf{B}_{p_2}^T \times$$
$$\left(\left\{Y_1 - \hat{m}_{p_1}(X_1)\right\}^2 - \varepsilon_1^2, ..., \left\{Y_n - \hat{m}_{p_1}(X_n)\right\}^2 - \varepsilon_n^2\right)^T$$

Then one writes

$$\hat{\sigma}_{p_2,p_1}^2(x) - \tilde{\sigma}_{p_2}^2(x) = I_{p_2,p_1}(x) + II_{p_2,p_1}(x) + III_{p_2,p_1}(x), \tag{2.10}$$

in which

$$\begin{aligned}
I_{p_2,p_1} &= I_{p_2,p_1}(x) \\
&= \left[\left\{B_{j,p_2}(x)\right\}_{j=1-p_2}^{N_1}\right]^T \left(\mathbf{B}_{p_2}^T \mathbf{B}_{p_2}\right)^{-1} \mathbf{B}_{p_2}^T \left(I_{1,p_1}, ..., I_{n,p_1}\right)^T \\
II_{p_2,p_1} &= II_{p_2,p_1}(x) \\
&= \left[\left\{B_{j,p_2}(x)\right\}_{j=1-p_2}^{N_1}\right]^T \left(\mathbf{B}_{p_2}^T \mathbf{B}_{p_2}\right)^{-1} \mathbf{B}_{p_2}^T \left(II_{1,p_1}, ..., II_{n,p_1}\right)^T \\
III_{p_2,p_1} &= III_{p_2,p_1}(x) \\
&= \left[\left\{B_{j,p_2}(x)\right\}_{j=1-p_2}^{N_1}\right]^T \left(\mathbf{B}_{p_2}^T \mathbf{B}_{p_2}\right)^{-1} \mathbf{B}_{p_2}^T \left(III_{1,p_1}, ..., III_{n,p_1}\right)^T
\end{aligned}$$

$$I_{i,p_1} = \left\{m(X_i) - \tilde{m}_{p_1}(X_i)\right\}^2 + \tilde{\varepsilon}_{p_1}^2(X_i) + 2\left\{m(X_i) - \tilde{m}_{p_1}(X_i)\right\}\tilde{\varepsilon}_{p_1}(X_i)$$

$$II_{i,p_1} = 2\tilde{\varepsilon}_{p_1}(X_i)\,\varepsilon_i$$

$$III_{i,p_1} = 2\left\{m(X_i) - \tilde{m}_{p_1}(X_i)\right\}\varepsilon_i.$$

2.4 Implementation

In this section, we describe procedures to implement the confidence bands in Theorem 2.2. Our codes are written in XploRe for convenience in order to use kernel smoothing, see Härdle et. al. [24].

Given any sample $\{(X_i, Y_i)\}_{i=1}^{n}$ from model (2.1), we use $\min(X_1, ..., X_n)$ and $\max(X_1, ..., X_n)$ respectively as the endpoints of interval $[a, b]$. Motivated by the comment of one referee, we select the number of interior knots N_ν using a BIC criteria. For knot location, we use equally space knots. According to Assumption (A3), the optimal order of N_ν is $n^{1/(2p_\nu+1)}$. Thus we propose selecting the "optimal" N_ν, denoted by $\hat{N}_\nu^{\text{opt}}$, from $[0.5N_{r\nu},\ \min(5N_{r\nu}, Tb)]$, with $N_{r\nu} = n^{1/(2p_\nu+1)}$ and $Tb = n/4 - 1$ to ensure that the total number of parameters in the least square estimation is less than $n/4$.

To be specific, let $q_n = (1 + N_n)$ be the total number of parameters. Then $\hat{N}_\nu^{\text{opt}}$ is the one minimizing the BIC value

$$\hat{N}_\nu^{\text{opt}} = \underset{N_n \in [0.5N_{r\nu},\ \min(5N_{r\nu}, Tb)]}{\text{argmin}} \text{BIC}(N_n)$$

where $\text{BIC} = \log(\text{MSE}) + q_n \log(n)/n$, with $\text{MSE} = \sum_{i=1}^{n} \{Y_i - \hat{Y}_i\}^2/n$.The least squares problem in (2.4) can be solved via the truncated power basis $\left\{1, x, ..., x^{p_1-1}, \left(x - t_j\right)_+^{p_1-1}, j = 1, ..., N_1\right\}$. In other words

$$\hat{m}_{p_1}(x) = \sum_{k=0}^{p_1-1} \hat{\gamma}_k x^k + \sum_{j=1}^{N_1} \hat{\gamma}_{j,p_1} \left(x - t_j\right)_+^{p_1-1},$$

where the coefficients $\left\{\hat{\gamma}_0, ..., \hat{\gamma}_{p_1-1}, \hat{\gamma}_{1,p_1}, ..., \hat{\gamma}_{N_1,p_1}\right\}^T$ are solutions to the following least squares problem

$$\left\{\hat{\gamma}_0, ..., \hat{\gamma}_{N_1,p_1}\right\}^T = \underset{R^{N_1+p_1}}{\text{argmin}} \sum_{i=1}^{n} \left\{Y_i - \sum_{k=0}^{p_1-1} \gamma_k X_i^k - \sum_{j=1}^{N_1} \gamma_{j,p_1} \left(X_i - t_j\right)_+^{p_1-1}\right\}^2.$$

The variance estimators $\hat{\sigma}^2_{p_1,p_2}(x)$ are computed likewise.

When constructing the confidence bands, one needs to evaluate the functions $v^2_{n,p_2}(x)$ in (2.7) differently for the exact and conservative bands, and the description is separated into two subsections. For both cases, one estimates the unknown functions $f(x)$ and $v^2_Z(x)$ and then plugs in these estimates, as in [70]. This is analogous to using $\overline{X} \pm 1.96 \times s_n/\sqrt{n}$ instead of $\overline{X} \pm 1.96 \times \sigma/\sqrt{n}$ as a large sample 95% confidence interval for a normal population mean μ, where the sample standard deviation s_n is a plugin substitute for the unknown population standard deviation σ.

Let $\widetilde{K}(u) = 15\left(1-u^2\right)^2 I\{|u| \leq 1\}/16$ be the quadric kernel, s_n =the sample standard deviation of $(X_i)_{i=1}^n$ and

$$\begin{aligned} \hat{f}(x) &= n^{-1}\sum\nolimits_{i=1}^n h^{-1}_{2\,\mathrm{rot},f}\widetilde{K}\left(\frac{X_i - x}{h_{2\,\mathrm{rot},f}}\right), \\ h_{2\,\mathrm{rot},f} &= (4\pi)^{1/10}\left(\frac{140}{3}\right)^{1/5} n^{-1/5}s_n, \end{aligned} \tag{2.11}$$

with $h_{2\,\mathrm{rot},f}$ the rule-of-thumb bandwidth in Silverman [62].

Define $\Xi_{p_2} = \left\{\Xi_{i,p_2}, 1 \leq i \leq n\right\}^T$, $\Xi_{i,p_2} = \left\{\hat{Z}_{i,p_1} - \hat{\sigma}^2_{p_1,p_2}(X_i)\right\}^2$, and

$$\mathbf{X} = \mathbf{X}(\mathbf{x}) = \begin{pmatrix} 1 & ,\ldots, & 1 \\ X_1 - x & ,\ldots, & X_n - x \end{pmatrix}^T,$$

$$\mathbf{W} = \mathbf{W}(x) = \mathrm{diag}\left\{\widetilde{K}\left(\frac{X_i - x}{h_{2\,\mathrm{rot},\sigma}}\right)\right\}_{i=1}^n,$$

where $h_{2\,\mathrm{rot},\sigma}$ is the rule-of-thumb bandwidth of Fan and Gijbels [14] based on data $\left(X_i, \Xi_{i,p_2}\right)_{i=1}^n$. Define the following estimators of $v^2_Z(x)$,

$$\hat{v}^2_{Z,p_2}(x) = \begin{pmatrix} 1, & 0 \end{pmatrix}\left(\mathbf{X}^T\mathbf{W}\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{W}\Xi_{p_2}. \tag{2.12}$$

The following uniform consistency results are provided in [1] and [14]

$$\max_{p_2} \sup_{x\in[a,b]} \left|\hat{v}^2_{Z,p_2}(x) - v^2_Z(x)\right| + \sup_{x\in[a,b]} \left|\hat{f}(x) - f(x)\right| = o_p(1). \tag{2.13}$$

2.4.1 Implementing the exact band

The function $v_{n,1}(x)$ is approximated by the following, with $\hat{f}(x)$ and $\hat{v}_{Z,1}(x)$ defined in (2.11) and (2.12), $j(x)$ defined in (2.5)

$$\hat{v}_{n,1}(x) = \hat{v}_{Z,1}(x)\hat{f}^{-1/2}(x)n^{-1/2}h_2^{-1/2}.$$

Then (2.13) and (2.8) imply that as $n \to \infty$, the band below is asymptotically exact

$$\hat{\sigma}^2_{1,1}(x) \pm \hat{v}_{n,1}(x)\{2\log(N_2+1)\}^{1/2} d_n. \tag{2.14}$$

2.4.2 Implementing the conservative band

The band below is asymptotically conservative

$$\hat{\sigma}^2_{2,2}(x) \pm \hat{v}_{n,2}(x)\{2\log(N_2+1) - 2\log\alpha\}^{1/2}, \tag{2.15}$$

where the function $v_{n,2}(x)$ in (2.7) for the linear band is estimated consistently by

$$\hat{v}_{n,2}(x) = \left\{\mathbf{\Delta}^T(x) L_{j_2(x)} \mathbf{\Delta}(x)\right\}^{1/2} \hat{v}_{Z,2}(x) \left\{\frac{2}{3}\hat{f}(x) n h_2\right\}^{-1/2}, \tag{2.16}$$

with $j_2(x)$ defined in (2.5), and $\hat{f}(x)$ and $\hat{v}^2_{Z,2}(x)$ defined in (2.11) and (2.12), $\mathbf{\Delta}(x)$ and L_j defined as follows:

$$\mathbf{\Delta}(x) = \begin{pmatrix} c_{j_2(x)-1}\{1 - r_2(x)\} \\ c_{j_2(x)} r_2(x) \end{pmatrix},$$

$$c_j = \begin{cases} \sqrt{2} & j = -1, N_2 \\ 1 & j = 0, ..., N_2 - 1 \end{cases},$$

$$L_j = \begin{pmatrix} l_{j+1,j+1} & l_{j+1,j+2} \\ l_{j+2,j+1} & l_{j+2,j+2} \end{pmatrix}, \; j = 0, 1, ..., N_2. \tag{2.17}$$

The terms $l_{ik}, |i-k| \leq 1$ are defined through the following matrix inversion

$$M_{N_2+2} = \begin{pmatrix} 1 & \sqrt{2}/4 & & & & 0 \\ \sqrt{2}/4 & 1 & 1/4 & & & \\ & 1/4 & 1 & \ddots & & \\ & & \ddots & \ddots & 1/4 & \\ & & & 1/4 & 1 & \sqrt{2}/4 \\ 0 & & & & \sqrt{2}/4 & 1 \end{pmatrix}_{(N_2+2)\times(N_2+2)}$$
$$= (l_{ik})^{-1}_{(N_2+2)\times(N_2+2)},$$

and computed via (2.18), (2.19), and (2.20) given below, which are needed for (2.17). Letting

$$z_1 = \frac{2+\sqrt{3}}{4}, \quad z_2 = \frac{2-\sqrt{3}}{4}, \quad \theta = \frac{z_2}{z_1} = \left(2-\sqrt{3}\right)^2 = 7 - 4\sqrt{3}, \tag{2.18}$$

and applying matrix theory from Gantmacher and Krein [19] and Zhang [84], we have the following

$$l_{11} = l_{N_2+2,N_2+2}$$
$$= \frac{8z_1^2\left(1-\theta^{N_2+1}\right) - z_1\left(1-\theta^{N_2}\right)}{8z_1^2\left(1-\theta^{N_2+1}\right) - 2z_1\left(1-\theta^{N_2}\right) + \left(1-\theta^{N_2-1}\right)/8},$$

$$l_{i,i}=\frac{\left\{8z_1\left(1-\theta^{N_2+2-i}\right)-\left(1-\theta^{N_2+1-i}\right)\right\}\left\{8z_1\left(1-\theta^{i-1}\right)-\left(1-\theta^{i-2}\right)\right\}}{(z_1-z_2)\left\{64z_1^2\left(1-\theta^{N_2+1}\right)-16z_1\left(1-\theta^{N_2}\right)+\left(1-\theta^{N_2-1}\right)\right\}} \tag{2.19}$$

for $2 \le i \le N_2 + 1$ and

$$\begin{aligned} l_{12} &= l_{N_2+1,N_2+2} \\ &= \frac{(-2\sqrt{2})\, z_1\left(1-\theta^{N_2}\right)-\left(1-\theta^{N_2-1}\right)/8}{8z_1^2\left(1-\theta^{N_2+1}\right)-2z_1\left(1-\theta^{N_2}\right)+8\left(1-\theta^{N_2-1}\right)/8}, \end{aligned}$$

$$l_{i,i+1}=\frac{\left\{8z_1\left(1-\theta^{N_2+1-i}\right)-\left(1-\theta^{N_2-i}\right)\right\}\left\{8z_1\left(1-\theta^{i-1}\right)-\left(1-\theta^{i-2}\right)\right\}}{4z_1(z_1-z_2)\left\{64z_1^2\left(1-\theta^{N_2+1}\right)-16z_1\left(1-\theta^{N_2}\right)+\left(1-\theta^{N_2-1}\right)\right\}} \tag{2.20}$$

for $2 \le i \le N_2$. By the symmetry of the matrix M_{N_2+2}, the lower diagonal entries are $l_{i+1,i} = l_{i,i+1}$, $\forall i = 1, ..., N_2 + 1$. See [70] for details.

2.4.3 Implementing the bootstrap band

In this subsection, we use wild bootstrap for improved performance following the suggestion of one referee. We define the residuals $\hat{\xi}_{i,p_1,p_2} = \hat{\varepsilon}^2_{i,p_1} - \hat{\sigma}^2_{p_1,p_2}(X_i)$, where $\hat{\varepsilon}_{i,p_1}$ are defined in (2.3), and denote a predetermined integer by n_B, whose default value is 500. The steps to compute bootstrap band, similar to Yang [77], are described in the following.

Step 1, Let $\left\{\delta_{i,k}\right\}_{1\le k\le n_B}$, $1 \le i \le n$ be i.i.d. samples of the following discrete distribution $\delta_{i,k} = \pm 1$ with probability 1/2, it is easily verified that $E(\delta_{i,k}) = 0$, $\mathrm{Var}\left(\delta_{i,k}\right) = 1$.

Step 2, For any $1 \le k \le n_B$, define the k-th wild bootstrap sample $\hat{\varepsilon}^2_{i,p_1,k} = \hat{\sigma}^2_{p_1,p_2}(X_i) + \hat{\xi}_{i,p_1,p_2}\delta_{i,k}$, $1 \le i \le n$.Taking $\mathbf{E}_{p_1,k} = \left\{\hat{\varepsilon}^2_{i,p_1,k}\right\}_{i=1}^{n}$, we apply linear spline on $\mathbf{E}_{p_1,k}$ to get the spline estimate

$$\hat{\sigma}^2_{p_1,p_2,k}(x) = \left[\left\{B_{j,p_1}(x)\right\}_{j=1-p_1}^{N_1}\right]^T \left(\mathbf{B}_{p_1}^T \mathbf{B}_{p_1}\right)^{-1} \mathbf{B}_{p_1}^T \mathbf{E}_{p_1,k} \tag{2.21}$$

Step 3, The wild bootstrap $(1-\alpha)$ pointwise confidence interval for function value $\sigma^2(x)$ at one point x is $\left[\widehat{\sigma}^2_{L,\alpha/2}(x), \widehat{\sigma}^2_{U,\alpha/2}(x)\right]$, where $\widehat{\sigma}^2_{L,\alpha/2}(x)$ and $\widehat{\sigma}^2_{U,\alpha/2}(x)$ are the lower and upper $100(\alpha/2)\%$ quantiles of the set $\hat{\sigma}^2_{p_1,p_2,k}(x)_{1\le k\le n_B}$ obtained from (2.21) for each of the bootstrap sample generated in Step 2.

Step 4, According to [70], the uniform confidence band is wider than the pointwise confidence interval by an inflation factor of $z^{-1}_{1-\alpha/2}\sqrt{2\left\{\log(N_2+1)-\log(\alpha/2)\right\}}$ when localized at any point x, hence we define the wild bootstrap $(1-\alpha)$ confidence band for the function $\sigma^2(x)$ over $[a,b]$ as $\left[\widehat{\sigma}^2_{L,\alpha/2}(x), \widehat{\sigma}^2_{U,\alpha/2}(x)\right]$, $x \in [a,b]$ where

$$\begin{aligned}
&\hat{\sigma}^2_{L,\alpha/2}(x) = \\
&\hat{\sigma}^2_{p_1,p_2}(x) + \left(\widehat{\sigma}^2_{L,\alpha/2}(x) - \hat{\sigma}^2_{p_1,p_2}(x)\right) z^{-1}_{1-\alpha/2}\sqrt{2\left\{\log(N_2+1)-\log(\alpha/2)\right\}}, \\
&\hat{\sigma}^2_{U,\alpha/2}(x) = \\
&\hat{\sigma}^2_{p_1,p_2}(x) + \left(\widehat{\sigma}^2_{U,\alpha/2}(x) - \hat{\sigma}^2_{p_1,p_2}(x)\right) z^{-1}_{1-\alpha/2}\sqrt{2\left\{\log(N_2+1)-\log(\alpha/2)\right\}}.
\end{aligned}$$

As one referee pointed out, instead of resampling at each point x and then inflate by a universal factor K_n, it is also possible to resample the maximal deviation distribution, as was done in Neumann and Kreiss [54], and obtain bootstrap lower and upper $100(\alpha/2)\%$ quantiles of $\sup_{x\in[a,b]}\left|\hat{\sigma}^2_{p_1,p_2}(x) - \sigma^2(x)\right| v^{-1}_{n,p_2}(x)$. Our approach, however, has the advantage of adaptivity since the confidence band is locally calibrated at each point x, without the constraint of symmetry.

2.5 Examples

2.5.1 Simulation example

To illustrate the finite-sample behavior of our confidence bands, we simulate data from model (2.1), with $X \sim U[-1/2, 1/2]$, and

$$m(x) = \sin(2\pi x), \quad \sigma(x) = \sigma_0 \frac{c - \exp(x)}{c + \exp(x)}, \quad \varepsilon | x \sim N_1\{0, \sigma^2(x)\}. \tag{2.22}$$

The noise levels are $\sigma_0 = 0.2, 0.5$, while sample sizes are taken to be $n = 100, 200, 500$. Confidence level $1 - \alpha = 0.99, 0.95$. For $c = 100$ and $c = 5$, Tables 2.1 and 2.2 contain the coverage probabilities as the percentage of coverage of the true curve $\sigma(x)$ at all data points $\{X_i\}_{i=1}^n$ by the confidence bands in (2.14), (2.15) and using bootstrap method, over 500 replications of sample size n. Following the suggestion of one referee, we have included variance functions $\sigma^2(x)$ that are strongly heteroscedastic ($c = 5$) and nearly homoscedastic ($c = 100$).

In all cases, the performance of constant band is worse than the linear band in terms of coverage, while the bootstrap band has the best coverage. In all cases the coverage improves with sample sizes increasing, showing a positive confirmation of Theorem 2.2. The bootstrap band achieves reasonable coverage rate for moderate sample size as low as 100, while for the nearly homoscedastic case of $c = 100$, the asymptotic linear band has good coverage for sample size as low as $n = 200$. For the strongly heteroscedastic case $c = 5$, it seems that the bootstrap band is the only satisfactory one. We therefore recommend using the bootstrap band for analyzing real data.

The graphs in Figures 2.1 and 2.2 are created based on two samples of size 100 and 500 respectively, for $c = 100$ and 5 respectively, each with three types of symbols: center thin solid line (true curve), center dotted line (the estimated curve), upper and lower thick solid line (bootstrap confidence band). In all figures, the confidence bands for $n = 500$ are thinner and fit better than those for $n = 100$.

We next compare by simulation the testing of heteroscedasticity based on the proposed bootstrap confidence band to the results of [11] for the following three models

$$
\begin{aligned}
m(x) = 1 + \sin(x),\ \sigma(x) = \sigma \exp(cx) \quad &\text{(monotone, model I)} \\
m(x) = 1 + x,\ \sigma(x) = \sigma \{1 + c\sin(10x)\}^2 \quad &\text{(high frequency, model II)} \\
m(x) = 1 + x,\ \sigma(x) = \sigma(1 + cx)^2 \quad &\text{(unimodal, model III)}
\end{aligned}
\tag{2.23}
$$

for $c = 0, 0.5, 1.0$ and $\sigma^2 = 0.25$ with standard normal errors. The design points X were generated uniformly from $[0, 1]$ and the sample sizes were $n = 50,\ 100,\ 200$. Table 2.3 shows the relative proportion of rejections for the various situations using both our method and the results from [11], Table 1, p. 700 (in brackets). Our method performs poorly when heteroscedasticity is weak ($c = 0.5$) for models I and III, so the type II error is larger than [11]. For strongly heteroscedastic model ($c = 1$), however, our method achieves higher rejection power for models II and III, and comparable rejection power for model I, so the type II error is either comparable to [11] or lower. For homoscedastic model ($c = 0$), our rejection rate is always lower, hence the bootstrap confidence band based test has smaller type I error than [11]. Based on the above simulation, our method is better than [11] at detecting strong heteroscedasticity and retaining homoscedasticity, while [11] is better than ours at discovering weak heteroscedasticity.

2.5.2 Fossil data and motorcycle data

In this subsection we apply the bootstrap band to two real data sets, both of which have sample size below 200.

Table 2.1: Coverage probabilities for $c = 100$ from 500 replications.

σ_0	n	$1-\alpha$	Constant Band	Linear Band	Bootstrap Band
0.2	100	0.99	0.882	0.886	0.944
		0.95	0.806	0.858	0.858
	200	0.99	0.940	0.970	0.996
		0.95	0.874	0.958	0.968
	500	0.99	0.984	0.994	1
		0.95	0.942	0.992	0.984
0.5	100	0.99	0.764	0.892	0.956
		0.95	0.690	0.870	0.886
	200	0.99	0.896	0.970	0.992
		0.95	0.830	0.962	0.960
	500	0.99	0.974	0.996	0.998
		0.95	0.926	0.994	0.984

The fossil data reflects global climate millions of years ago through ratios of strontium isotopes found in fossil shells. These were studied by Chaudhuri and Marron[5] to detect the structure via kernel smoothing. The corresponding penalized spline fit was provided in Ruppert et. al. [60]. In this section we test the heteroscedasticity of the fossil data variance. The null hypothesis is $H_0 : \sigma^2(x) = \sigma_0^2 > 0$. The response Y is the strontium isotopes ratio after linear transformation, $Y = 0.70715 + \text{ratio} * 10^{-5}$, since all the values are very close to 0.707, while the predictor X is the fossil shell age in million years.

In Figure 2.3, the center dotted line is the linear spline fit $\hat{\sigma}^2_{2,2}(x)$ for the variance function $\sigma^2(x)$. The upper/lower thick solid lines represent bootstrap confidence band. The constant horizontal line between the upper/lower thick lines represents the average of the minimum of the upper line and the maximum of the lower line, which indicates if one can fit a constant line into the confidence band. Since the variance band of high confidence level $100(1 - 0.20)\%$ contains the fitted constant line entirely, we have failed to reject the null hypothesis of homoscedasticity with p-value 0.20.

Table 2.2: Coverage probabilities for $c = 5$ from 500 replications.

σ_0	n	$1-\alpha$	Constant Band	Linear Band	Bootstrap Band
		0.99	0.824	0.858	0.944
	100	0.95	0.764	0.834	0.874
		0.99	0.912	0.896	0.986
0.2	200	0.95	0.832	0.884	0.954
		0.99	0.978	0.970	1
	500	0.95	0.916	0.964	0.992
		0.99	0.886	0.856	0.946
	100	0.95	0.648	0.828	0.878
		0.99	0.916	0.918	0.992
0.5	200	0.95	0.688	0.904	0.958
		0.99	0.958	0.966	1
	500	0.95	0.726	0.964	0.986

A second data used to illustrate our technique is the well-known motorcycle data. The X-values denote time (in milliseconds) after a simulated impact with motorcycles. The response variable Y is the head acceleration of a PTMO (post mortem human test object).

In Figure 2.4, the center dotted line is the linear spline fit $\hat{\sigma}^2_{2,2}(x)$ for $\sigma^2(x)$. The upper/lower thick solid lines represent bootstrap confidence band. The constant line between the upper/lower thick lines represents the average of the minimum of the upper line and the maximum of the lower line. Since the variance band of an extremely high confidence level $100(1-0.008)\%$ does not contain the fitted constant line entirely, we reject the null hypothesis of homoscedasticity with p-value ≤ 0.008.

In both Figures 2.3 and 2.4, there exists an exact correspondence of high $\hat{\sigma}^2_{2,2}(x)$ value in the upper plot to greater width of the confidence band for the conditional mean function in the lower plot, throughout the entire data range.

Table 2.3: Simulated rejection probabilities of test homoscedasticity from 500 replications.

	$n = 50$			$n = 100$			$n = 200$		
α	2.5%	5%	10%	2.5%	5%	10%	2.5%	5%	10%
c	model I								
0	0.004	0.004	0.012	0	0	0.002	0	0	0
	(0.038)	(0.056)	(0.101)	(0.028)	(0.057)	(0.093)	(0.037)	(0.059)	(0.105)
0.5	0.014	0.020	0.030	0.002	0.006	0.018	0	0.004	0.034
	(0.055)	(0.084)	(0.132)	(0.064)	(0.097)	(0.151)	(0.086)	(0.134)	(0.200)
1.0	0.038	0.058	0.110	0.024	0.072	0.254	0.150	0.362	0.690
	(0.095)	(0.148)	(0.223)	(0.153)	(0.215)	(0.313)	(0.249)	(0.337)	(0.458)
c	model II								
0	0.004	0.004	0.012	0	0	0.002	0	0	0
	(0.031)	(0.053)	(0.100)	(0.026)	(0.049)	(0.089)	(0.032)	(0.056)	(0.100)
0.5	0.082	0.106	0.158	0.296	0.484	0.766	0.694	0.918	0.992
	(0.197)	(0.276)	(0.390)	(0.333)	(0.433)	(0.568)	(0.527)	(0.637)	(0.761)
1.0	0.316	0.422	0.612	0.356	0.512	0.734	0.656	0.884	0.984
	(0.272)	(0.365)	(0.481)	(0.477)	(0.557)	(0.674)	(0.693)	(0.790)	(0.884)
c	model III								
0	0.004	0.004	0.012	0	0	0.002	0	0	0
	(0.034)	(0.054)	(0.097)	(0.028)	(0.053)	(0.100)	(0.031)	(0.053)	(0.094)
0.5	0.02	0.034	0.066	0.010	0.030	0.110	0.032	0.142	0.394
	(0.073)	(0.113)	(0.185)	(0.105)	(0.158)	(0.233)	(0.175)	(0.239)	(0.342)
1.0	0.078	0.112	0.216	0.122	0.312	0.642	0.668	0.984	0.978
	(0.136)	(0.198)	(0.291)	(0.221)	(0.304)	(0.412)	(0.378)	(0.476)	(0.598)

2.6 Appendix

The goals of this Appendix are to prove Propositions 2.1, 2.2 and 2.3. These clearly establish Theorem 2.1 and Theorem 2.2. In what follows, we denote by $\|\xi\|$ the Euclidean norm and by $|\xi|$ the largest absolute value of the elements of any vector ξ. We use c, C to denote positive constants in the generic sense.

The following result is based on Theorem 3.2 and Propositions 3.1, 3.2 of [70], see also [28] and Leadbetter et. al. [38].

Lemma 2.1. *Under Assumptions (A1)-(A4), there exists a constant* $C_{p_1} > 0, p_1 \geq 1$

such that for any $m \in C^{(p_1)}[a,b]$ and the function $\tilde{m}_{p_1}(x)$ given in (2.9),

$$\left\|\tilde{m}_{p_1}(x) - m(x)\right\|_\infty \le C_{p_1} \inf_{g \in G^{(p_1-2)}} \|g - m\|_\infty = O_P\left(h_1^{p_1}\right). \tag{2.24}$$

Moreover, for the function $\tilde{\varepsilon}_{p_1}(x)$ given in (2.9),

$$\left\|\tilde{\varepsilon}_{p_1}(x)\right\|_\infty = O_P\left(h_1^{p_1}\sqrt{\log n}\right). \tag{2.25}$$

According to Lemma 2.1, the bias term $\tilde{m}_{p_1}(x) - m(x)$ is uniformly of order $O_P\left(h_1^{p_1}\right) = O_P\left(n^{-p_1/(2p_1+1)}\right)$, while the noise term $\tilde{\varepsilon}_{p_1}(x)$ is uniformly of order $O_P\left(h_1^{p_1}\sqrt{\log n}\right) = O_P\left(n^{-p_1/(2p_1+1)}\sqrt{\log n}\right)$.

The following lemma on uniform convergence of the empirical inner product to the theoretical counterparts is from Lemma 3.1 of [70].

Lemma 2.2. *Under Assumptions (A2) and (A3), as $n \to \infty$,*

$$\begin{aligned} A_{n,p_1} &= \sup_{g_1, g_2 \in G^{(p_1-2)}} \left|\frac{\langle g_1, g_2\rangle_n - \langle g_1, g_2\rangle}{\|g_1\|_2 \|g_2\|_2}\right| \\ &= O_P\left(\sqrt{n^{-1}h_1^{-1}\log(n)}\right). \end{aligned} \tag{2.26}$$

The next result on the empirical inner product matrix is based on Lemma B.2 of [70] and Lemma A.5 of [76].

Lemma 2.3. *Under Assumptions (A2) and (A3), there exist constants $c(f), C(f) > 0$ independent of n but dependent on f, such that as $n \to \infty$, with probability approaching 1, for all $\xi \in R^{N_\nu + p_\nu}$, $\nu = 1, 2$*

$$c(f)|\xi| \le \left|\left(n^{-1}\mathbf{B}_{p_\nu}^T \mathbf{B}_{p_\nu}\right)^{-1}\xi\right| \le C(f)|\xi|, \tag{2.27}$$

$$c(f)\|\xi\|^2 \le \xi^T\left(n^{-1}\mathbf{B}_{p_\nu}^T \mathbf{B}_{p_\nu}\right)^{-1}\xi \le C(f)\|\xi\|^2. \tag{2.28}$$

Using the above three results, we establish two additional technical lemmas to be used in proving Propositions 2.1, 2.2 and 2.3.

Lemma 2.4. *Under Assumptions (A2) and (A3), as $n \to \infty$,*

$$\sup_{x\in[a,b]} \left\{ \left| \left\{ B_{j,p_\nu}(x) \right\}_{j=1-p_\nu}^{N_\nu} \right| + \left\| \left\{ B_{j,p_\nu}(x) \right\}_{j=1-p_\nu}^{N_\nu} \right\| \right\} = O\left(h_\nu^{-1/2}\right), \quad (2.29)$$

$$\max_{j=1-p_\nu}^{N_\nu} \left\{ \left\langle B_{j,p_\nu}, 1 \right\rangle \right\} = O\left(h_\nu^{1/2}\right), \quad \max_{j=1-p_\nu}^{N_\nu} \left\{ \left\langle B_{j,p_\nu}, 1 \right\rangle_n \right\}$$
$$= O_p\left(h_\nu^{1/2} + \sqrt{n^{-1} h_\nu^{-1} \log n} \right). \quad (2.30)$$

Proof. For each $x \in [a,b]$, at most p_ν of the $B_{j,p_\nu}(x)$'s are nonzero, (2.29) follows directly from the definition of $\left\{ B_{j,p_\nu}(x) \right\}_{j=1-p_\nu}^{N_\nu}$ and the simple fact that

$$\|b_{j,p_\nu}\|_2 \geq c h_\nu^{1/2}, 1 - p_\nu \leq j \leq N_\nu.$$

The same definition and fact also imply that

$$\max_{j=1-p_\nu}^{N_\nu} \left\{ \left\langle B_{j,p_\nu}, 1 \right\rangle \right\} \leq p_\nu \times h_\nu \times \left(c h_\nu^{1/2} \right)^{-1} = O\left(h_\nu^{1/2} \right).$$

As all $\left\{ B_{j,p_\nu}(x) \right\}_{j=1-p_\nu}^{N_\nu}$ are standardized, the definition and rate of A_{n,p_ν} in (2.26) imply the second half of (2.30).

Lemma 2.5. *Under Assumptions (A2) and (A3), as $n \to \infty$,*

$$\sum_{j=1-p_2}^{N_2} \sum_{k=1-p_1}^{N_1} \left\{ n^{-1} \sum_{i=1}^{n} B_{j,p_2}(X_i)\, \varepsilon_i B_{k,p_1}(X_i) \right\}^2$$
$$= O_p\left(n^{5/2(2p_*+1) - 1/2(2p^*+1) - 1} \right), \quad (2.31)$$

while for any continuous function r *defined on* $[a,b]$,

$$E\sum_{j=1-p_1}^{N_1}\left[n^{-1}\sum_{i=1}^{n}B_{j,p_1}(X_i)\,r(X_i)\,\varepsilon_i\right]^2$$
$$\leq \|\sigma\|_\infty^2\|r\|_\infty^2(N_1+p_1)\,n^{-1}. \tag{2.32}$$

Proof.

$$E\sum_{j=1-p_2}^{N_2}\sum_{k=1-p_1}^{N_1}\left\{n^{-1}\sum_{i=1}^{n}B_{j,p_2}(X_i)\,\varepsilon_i B_{k,p_1}(X_i)\right\}^2$$
$$= \sum_{j=1-p_2}^{N_2}\sum_{k=1-p_1}^{N_1}n^{-2}\sum_{i=1}^{n}E\left\{B_{j,p_2}(X_i)^2B_{k,p_1}(X_i)^2\sigma^2(X_i)\right\}$$
$$\leq n^{-1}\max(N_1+p_1,N_2+p_2)\,N_*^{-1}N^*$$
$$\times\max_{|k-j|\leq p_1}E\left\{B_{j,p_1}(X_1)^2B_{k,p_1}(X_1)^2\sigma^2(X_1)\right\}.$$

With the definition of $B_{j,p_1}(x)\equiv b_{j,p_1}(x)\,\|b_{j,p_1}\|_2^{-1},\ 1-p_1\leq j\leq N_1$, we have

$$\max_{|k-j|\leq p_1}E\left\{B_{j,p_2}(X_1)^2B_{k,p_1}(X_1)^2\sigma^2(X_1)\right\}$$
$$\leq c(\sigma)\frac{c(f)\sqrt{h_1h_2}}{C(f)\,h_1h_2}=\frac{C(f,\sigma)}{\sqrt{h_1h_2}}.$$

Thus (2.31) follows from

$$E\sum_{j=1-p_2}^{N_2}\sum_{k=1-p_1}^{N_1}\left\{n^{-1}\sum_{i=1}^{n}B_{j,p_2}(X_i)\,\varepsilon_iB_{k,p_1}(X_i)\right\}^2$$
$$\leq n^{-1}\max(N_1+p_1,N_2+p_2)\,N_*^{-1}N^*\times\frac{C(f,\sigma)}{\sqrt{h_1h_2}}$$
$$= O\left(n^{5/2(2p_*+1)-1/2(2p^*+1)-1}\right).$$

To prove (2.32), we argue that

$$
\begin{aligned}
& E \sum_{j=1-p_1}^{N_1} \left[n^{-1} \sum\nolimits_{i=1}^{n} B_{j,p_1}(X_i)\, r(X_i)\, \varepsilon_i\right]^2 \\
= & \sum_{j=1-p_1}^{N_1} n^{-1} E\left\{B_{j,p_1}^2(X_1)\, r^2(X_1)\, \sigma^2(X_1)\right\} \\
\leq & \|\sigma\|_\infty^2 \|r\|_\infty^2 n^{-1} \sum_{j=1-p_1}^{N_1} E\left\{B_{j,p_1}(X_1)^2\right\} \\
= & \|\sigma\|_\infty^2 \|r\|_\infty^2 (N_1 + p_1)\, n^{-1}.
\end{aligned}
$$

The next three propositions show the asymptotical property of the three terms, I_{p_2,p_1}, II_{p_2,p_1} and III_{p_2,p_1} in (2.10), decomposed from section 2.3, then establish Theorem 2.1.

Proposition 2.1. *Under Assumptions (A1)-(A4),* $\left\|I_{p_2,p_1}\right\|_\infty = \sup_{x\in[a,b]} \left|I_{p_2,p_1}(x)\right|$, *as* $n \to \infty$, *is of order*

$$
O_p\left(h_1^{2p_1} \log n\right) = O_p\left(n^{-2p_1/(2p_1+1)} \log n\right) = o_p\left(n^{-p_2/(2p_2+1)}\right).
$$

Proof. By Cauchy-Schwarz inequality,

$$
\left|I_{i,p_1}\right| \leq 2\left\{m(X_i) - \tilde{m}_{p_1}(X_i)\right\}^2 + 2\tilde{\varepsilon}_{p_1}^2(X_i),
$$

thus $\max_{i=1}^n \left|I_{i,p_1}\right|$ is bounded by

$$
\begin{aligned}
& 2\left\{\max_{i=1}^{n}\left\{m(X_i) - \tilde{m}_{p_1}(X_i)\right\}^2 + \max_{i=1}^{n} \tilde{\varepsilon}_{p_1}^2(X_i)\right\} \\
\leq & 2\left\{\left\|m - \tilde{m}_{p_1}\right\|_\infty^2 + \left\|\tilde{\varepsilon}_{p_1}\right\|_\infty^2\right\}.
\end{aligned}
$$

It follows that

$$\begin{aligned}
&\left\|I_{p_1,p_2}\right\|_\infty \\
= &\sup_{x\in[a,b]}\left|\left\{B_{j,p_2}(x)\right\}_{j=1-p_2}^{N_2}\left(\mathbf{B}_{p_2}^T\mathbf{B}_{p_2}\right)^{-1}\mathbf{B}_{p_2}^T\left(I_{i,p_1},1\le i\le n\right)^T\right|,
\end{aligned}$$

which, as for each $x\in[a,b]$, $B_{j,p_2}(x)\neq 0$ for at most p_2 values of j, is bounded by

$$p_2\max_{j=1-p_2}^{N_2}B_{j,p_2}(x)\left|\left(n^{-1}\mathbf{B}_{p_2}^T\mathbf{B}_{p_2}\right)^{-1}\times n^{-1}\mathbf{B}_{p_2}^T\left(\left|I_{i,p_1}\right|,1\le i\le n\right)^T\right|.$$

Using (2.29) in Lemma 2.4 and (3.41) in Lemma 3.10, the above is bounded by $p_2c(f)\,h_2^{-1/2}\times C(f)\times\left|n^{-1}\mathbf{B}_{p_2}^T\left(\left|I_{i,p_1}\right|,1\le i\le n\right)^T\right|$. THen, using the bound on $\max_{i=1}^n\left|I_{i,p_1}\right|$, we have

$$\begin{aligned}
&\left\|I_{p_1,p_2}\right\|_\infty \\
\le\ & C(f)\,h_2^{-1/2}\times\left\{\left\|m-\tilde{m}_{p_1}\right\|_\infty^2+\left\|\tilde{\varepsilon}_{p_1}\right\|_\infty^2\right\}\times\max_{j=1-p_2}^{N_2}\left\{\left\langle B_{j,p_2},1\right\rangle_n\right\}
\end{aligned}$$

which, applying (2.24) and (2.25) in Lemma 2.1, and (2.26), (2.30) in Lemma 2.4, is bounded by

$$\begin{aligned}
&O_p\left\{h_2^{-1/2}\times\left(h_1^{2p_1}+h_1^{2p_1}\log n\right)\times\left(h_2^{1/2}+\sqrt{n^{-1}h_2^{-1}\log n}\right)\right\} \\
=\ & O_p\left(h_1^{2p_1}\log n\right).
\end{aligned}$$

Proposition 2.2. *Under Assumptions (A1)-(A4), as* $n\to\infty$,

$$\begin{aligned}
\left\|II_{p_2,p_1}\right\|_\infty &= \sup_{x\in[a,b]}\left|II_{p_2,p_1}(x)\right| \\
&= O_p\left(n^{3/(2p_*+1)-3/2}\right)=o_p\left(n^{-p_2/(2p_2+1)}\right).
\end{aligned}$$

Proof. By definition

$$
\begin{aligned}
& II_{p_1,p_2}(x) \\
= & \left\{B_{1-p_2,p_2}(x), ..., B_{N_2,p_2}(x)\right\}\left(\mathbf{B}_{p_2}^T\mathbf{B}_{p_2}\right)^{-1}\mathbf{B}_{p_2}^T\left(II_{1,p_1}, ..., II_{n,p_1}\right)^T \\
= & 2\left\{B_{1-p_2,p_2}(x), ..., B_{N_2,p_2}(x)\right\}\left(n^{-1}\mathbf{B}_{p_2}^T\mathbf{B}_{p_2}\right)^{-1} \\
& \times\left\{n^{-1}\sum_{i=1}^n B_{j,p_2}(X_i)\tilde{\varepsilon}_{p_1}(X_i)\varepsilon_i\right\}_{j=1-p_2}^{N_2}.
\end{aligned}
$$

Applying (3.41) in Lemma 3.10, $\left|II_{p_1,p_2}(x)\right|$, with probability approaching 1, is bounded by

$$
\begin{aligned}
& C\left\|\left\{B_{1-p_2,p_2}(x), ..., B_{N_2,p_2}(x)\right\}\right\|C(f) \\
& \times\left\|\left\{n^{-1}\sum_{i=1}^n B_{j,p_2}(X_i)\tilde{\varepsilon}_{p_1}(X_i)\varepsilon_i\right\}_{j=1-p_2}^{N_2}\right\|,
\end{aligned}
$$

applying (2.29) in Lemma 2.4,

$$
\sup_{x\in[a,b]}\left|II_{p_1,p_2}(x)\right| \le C(f)\,h_2^{-1/2}\left\|\left\{n^{-1}\sum_{i=1}^n B_{j,p_2}(X_i)\tilde{\varepsilon}_{p_1}(X_i)\varepsilon_i\right\}_{j=1-p_2}^{N_2}\right\|.
$$

Next, one can write for any $1-p_2 \le j \le N_2$,

$$
\begin{aligned}
& n^{-1}\sum_{i=1}^n B_{j,p_2}(X_i)\tilde{\varepsilon}_{p_1}(X_i)\varepsilon_i \\
= & n^{-1}\sum_{i=1}^n B_{j,p_2}(X_i)\varepsilon_i\left\{B_{1-p_2,p_2}(X_i), ..., B_{N_2,p_2}(X_i)\right\} \\
& \times\left(n^{-1}\mathbf{B}_{p_1}^T\mathbf{B}_{p_1}\right)^{-1}n^{-1}\mathbf{B}_{p1}^T\mathbf{E} \\
= & \left\{n^{-1}\sum_{i=1}^n B_{j,p_2}(X_i)\varepsilon_i B_{k,p_1}(X_i)\right\}_{k=1-p_1}^{N_1} \\
& \times\left(n^{-1}\mathbf{B}_{p_1}^T\mathbf{B}_{p_1}\right)^{-1}n^{-1}\mathbf{B}_{p_1}^T\mathbf{E},
\end{aligned}
$$

hence, $\sup_{x\in[a,b]}\left|II_{p_1,p_2}(x)\right|$ is bounded by

$$\begin{aligned}
& C(f)\,h_2^{-1/2}\sqrt{\sum_{j=1-p_2}^{N_2}\left[\left\{\frac{1}{n}\sum_{i=1}^{n}B_{j,p_2}(X_i)\,\varepsilon_i B_{k,p_1}(X_i)\right\}_{k=1-p_1}^{N_1}\right.} \\
& \times\sqrt{\left.\left(\frac{1}{n}\mathbf{B}_{p_1}^T\mathbf{B}_{p_1}\right)^{-1}n^{-1}\mathbf{B}_{p_1}^T\mathbf{E}\right]^2} \\
\leq\; & C(f)\,h_2^{-1/2}\sqrt{\sum_{j=1-p_2}^{N_2}\left\|\left\{\frac{1}{n}\sum_{i=1}^{n}B_{j,p_2}(X_i)\,\varepsilon_i B_{k,p_1}(X_i)\right\}_{k=1-p_1}^{N_1}\right\|^2} \\
& \times\left\|\left(\frac{1}{n}\mathbf{B}_{p_1}^T\mathbf{B}_{p_1}\right)^{-1}\frac{1}{n}\mathbf{B}_{p_1}^T\mathbf{E}\right\| \\
=\; & C(f)\,h_2^{-1/2}\sqrt{\sum_{j=1-p_2}^{N_2}\sum_{k=1-p_1}^{N_1}\left\{\frac{1}{n}\sum_{i=1}^{n}B_{j,p_2}(X_i)\,\varepsilon_i B_{k,p_1}(X_i)\right\}^2} \\
& \times\left\|\left(\frac{1}{n}\mathbf{B}_{p_1}^T\mathbf{B}_{p_1}\right)^{-1}\frac{1}{n}\mathbf{B}_{p_1}^T\mathbf{E}\right\|
\end{aligned}$$

by Cauchy-Schwarz inequality.

Note that with probability approaching 1,

$$\begin{aligned}
& \left\|\left(n^{-1}\mathbf{B}_{p_1}^T\mathbf{B}_{p_1}\right)^{-1}n^{-1}\mathbf{B}_{p_1}^T\mathbf{E}\right\|^2 \\
\leq\; & C(f)^2\left\{\left(n^{-1}\mathbf{B}_{p_1}^T\mathbf{E}\right)^T n^{-1}\mathbf{B}_{p_1}^T\mathbf{E}\right\} \\
=\; & C(f)^2\sum_{k=1-p_1}^{N_1}\left\{n^{-1}\textstyle\sum_{i=1}^{n}B_{j,p_1}(X_i)\,\varepsilon_i\right\}^2 \\
=\; & O_p\left\{(N_1+p_1)\,n^{-1}\right\}=O_p\left(N_1 n^{-1}\right)
\end{aligned}$$

according to (2.32) of Lemma 2.5 with function $r(x)\equiv 1$. Meanwhile, according to

(2.31) in Lemma 2.5 we have,

$$
\begin{aligned}
& \sup_{x\in[a,b]} \left| II_{p_1,p_2}(x) \right| \\
= \; & O_p\left(h_2^{-1/2} \times n^{5/2(2p_*+1)-1/2(2p^*+1)-1} \times \sqrt{N_1/n} \right) \\
= \; & O_p\left(n^{1/2(2p_2+1)} \times n^{5/2(2p_*+1)-1/2(2p^*+1)-1} \times n^{1/2(2p_1+1)} n^{-1/2} \right) \\
= \; & O_p\left(n^{3/(2p_*+1)-3/2} \right).
\end{aligned}
$$

Proposition 2.3. *Under Assumptions (A1)-(A4), as $n \to \infty$,*

$$
\begin{aligned}
\left\| III_{p_2,p_1} \right\|_\infty &= \sup_{x\in[a,b]} \left| III_{p_2,p_1}(x) \right| \\
&= O_p\left(n^{3/(2p_*+1)-1} \right) = o_p\left(n^{-p_2 1/(2p_2+1)} \right).
\end{aligned}
$$

Proof.

$$
\begin{aligned}
& III_{p_1,p_2}(x) \\
= \; & \left\{ B_{1-p_2,p_2}(x), ..., B_{N_2,p_2}(x) \right\} \left(\frac{1}{n} \mathbf{B}_{p_2}^T \mathbf{B}_{p_2} \right)^{-1} \\
& \frac{1}{n} \mathbf{B}_{p_2}^T \left(III_{1,p_1}, ..., III_{n,p_1} \right)^T \\
= \; & 2 \left\{ B_{j,p_2}(x) \right\}_{j=1-p_2}^{N_2} \left(\frac{1}{n} \mathbf{B}_{p_2}^T \mathbf{B}_{p_2} \right)^{-1} \\
& \left\{ \frac{1}{n} \sum_{i=1}^n B_{j,p_2}(X_i) \left\{ m(X_i) - \tilde{m}_{p_1}(X_i) \right\} \varepsilon_i \right\}_{j=1-p_2}^{N_2} \\
= \; & 2 \left\{ B_{j,p_2}(x) \right\}_{j=1-p_2}^{N_2} \left(\frac{1}{n} \mathbf{B}_{p_2}^T \mathbf{B}_{p_2} \right)^{-1} \\
& \left\{ \frac{1}{n} \sum_{i=1}^n B_{j,p_2}(X_i) \left\{ m(X_i) - g_{p_1}(X_i) \right\} \varepsilon_i \right\}_{j=1-p_2}^{N_2}
\end{aligned}
$$

$$
\begin{aligned}
&+ 2\left\{B_{j,p_2}(x)\right\}_{j=1-p_2}^{N_2}\left(\frac{1}{n}\mathbf{B}_{p_2}^T\mathbf{B}_{p_2}\right)^{-1} \\
&\left\{\frac{1}{n}\sum_{i=1}^{n}B_{j,p_2}(X_i)\left\{g_{p_1}(X_i)-\tilde{m}_{p_1}(X_i)\right\}\varepsilon_i\right\}_{j=1-p_2}^{N_2} \\
&= 2\left\{B_{j,p_2}(x)\right\}_{j=1-p_2}^{N_2}\left(\frac{1}{n}\mathbf{B}_{p_2}^T\mathbf{B}_{p_2}\right)^{-1} \\
&\left\{\frac{1}{n}\sum_{i=1}^{n}B_{j,p_2}(X_i)\left\{m(X_i)-g_{p_1}(X_i)\right\}\varepsilon_i\right\}_{j=1-p_2}^{N_2} \\
&+ 2\left\{B_{j,p_2}(x)\right\}_{j=1-p_2}^{N_2}\left(\frac{1}{n}\mathbf{B}_{p_2}^T\mathbf{B}_{p_2}\right)^{-1} \\
&\left\{\left\{\frac{1}{n}\sum_{i=1}^{n}B_{j,p_2}(X_i)\varepsilon_i B_{k,p_1}(X_i)\right\}_{k=1-p_1}^{N_1}\right. \\
&\left.\left(\frac{1}{n}\mathbf{B}_{p_1}^T\mathbf{B}_{p_1}\right)^{-1}\frac{1}{n}\mathbf{B}_{p_1}^T\left(\mathbf{g}_{p_1}-\mathbf{m}\right)\right\}_{j=1-p_2}^{N_2},
\end{aligned}
$$

in which the spline function $g_{p_1} \in G^{(p_1-1)}$ satisfies $\| m - g_{p_1} \|_\infty \leq Ch_1^{p_1}$, and $\mathbf{g}_{p_1} = \left\{g_{p_1}(X_1), ..., g_{p_1}(X_n)\right\}^T$.

With probability approaching 1, according to (2.29) in Lemma 2.4, the first term in the above is bounded by

$$
\begin{aligned}
&\left|2\left\{B_{1-p_2,p_2}(x)\right\}_{j=1-p_2}^{N_2}\left(\frac{1}{n}\mathbf{B}_{p_2}^T\mathbf{B}_{p_2}\right)^{-1}\right| \\
&\left|\left\{\frac{1}{n}\sum_{i=1}^{n}B_{j,p_2}(X_i)\left\{m(X_i)-g_{p_1}(X_i)\right\}\varepsilon_i\right\}_{j=1-p_2}^{N_2}\right| \\
&\leq C(f)\,h_2^{-1/2}\left\|\left\{\frac{1}{n}\sum_{i=1}^{n}B_{j,p_2}(X_i)\left\{m(X_i)-g_{p_1}(X_i)\right\}\varepsilon_i\right\}_{j=1-p_2}^{N_2}\right\|.
\end{aligned}
$$

By (2.32) in Lemma 2.5 with $r(x) = m(x) - g_{p_1}(x)$, the above has order

$$O_P\left(h_2^{-1/2}\sqrt{\|\sigma\|_\infty^2 \left\|m - g_{p_1}\right\|_\infty^2 (N_1 + p_1) n^{-1}}\right) = O_P\left(N_2^{1/2} N_1/n\right).$$

For the second term

$$\begin{aligned}
&\Bigg|2\left\{B_{1-p_2,p_2}(x), \ldots, B_{N_2,p_2}(x)\right\}\left(\frac{1}{n}\mathbf{B}_{p_2}^T\mathbf{B}_{p_2}\right)^{-1} \\
&\left\{\left\{\frac{1}{n}\sum_{i=1}^n B_{j,p_2}(X_i)\,\varepsilon_i B_{k,p_1}(X_i)\right\}_{k=1-p_1}^{N_1}\right. \\
&\left.\left(\frac{1}{n}\mathbf{B}_{p_1}^T\mathbf{B}_{p_1}\right)^{-1}\frac{1}{n}\mathbf{B}_{p_1}^T\left(\mathbf{g}_{p_1} - \mathbf{m}\right)\right\}_{j=1-p_2}^{N_2}\Bigg| \\
\leq\ & \frac{C(f)}{h_2^{1/2}}\Bigg\|\left\{\left\{\frac{1}{n}\sum_{i=1}^n B_{j,p_2}(X_i)\,\varepsilon_i B_{k,p_1}(X_i)\right\}_{k=1-p_1}^{N_1}\right. \\
&\left.\left(\frac{1}{n}\mathbf{B}_{p_1}^T\mathbf{B}_{p_1}\right)^{-1}\frac{1}{n}\mathbf{B}_{p_1}^T\left(\mathbf{g}_{p_1} - \mathbf{m}\right)\right\}_{j=1-p_2}^{N_2}\Bigg\| \\
\leq\ & \frac{C(f)}{h_2^{1/2}}\sqrt{\sum_{j=1-p_2}^{N_2}\left\|\left\{\frac{1}{n}\sum_{i=1}^n B_{j,p_2}(X_i)\,\varepsilon_i B_{k,p_1}(X_i)\right\}_{k=1-p_1}^{N_1}\right\|^2} \\
&\left\|\left(\frac{1}{n}\mathbf{B}_{p_1}^T\mathbf{B}_{p_1}\right)^{-1}\frac{1}{n}\mathbf{B}_{p_1}^T\left(\mathbf{g}_{p_1} - \mathbf{m}\right)\right\|.
\end{aligned}$$

The order of

$$\sqrt{\sum_{j=1-p_2}^{N_2}\left\|\left\{n^{-1}\sum_{i=1}^n B_{j,p_2}(X_i)\,\varepsilon_i B_{k,p_1}(X_i)\right\}_{k=1-p_1}^{N_1}\right\|^2}$$

is $O_P\left(n^{5/2(2p_*+1) - 1/2(2p^*+1) - 1}\right)$ according to (2.31) in Lemma 2.5. And with probability approaching 1, (3.42) of Lemma 3.10 implies that

$$\begin{aligned}
&\left\|\left(n^{-1}\mathbf{B}_{p_1}^T\mathbf{B}_{p_1}\right)^{-1} n^{-1}\mathbf{B}_{p_1}^T\left(\mathbf{g}_{p_1}-\mathbf{m}\right)\right\| \\
\leq\ & C(f)\left\|n^{-1}\mathbf{B}_{p_1}^T\left(\mathbf{g}_{p_1}-\mathbf{m}\right)\right\|,
\end{aligned}$$

while $\left\|n^{-1}\mathbf{B}_{p_1}^T\left(\mathbf{g}_{p_1}-\mathbf{m}\right)\right\|$ is bounded by

$$\begin{aligned}
&\sqrt{\sum_{j=1-p_1}^{N_1}\left\{\frac{1}{n}\sum_{i=1}^{n}B_{j,p_1}(X_i)\left|g_{p_1}-m\right|(X_i)\right\}^2} \\
&\leq \left\|\mathbf{g}_{p_1}-\mathbf{m}\right\|_\infty \sqrt{\sum_{j=1-p_1}^{N_1}\left\{\frac{1}{n}\sum_{i=1}^{n}B_{j,p_1}(X_i)\right\}^2} \\
&= O_p\left(h_1^{p_1}\right)\sqrt{\sum_{j=1-p_1}^{N_1}\max_{j=1-p_1}^{N_1}\left\{\left\langle B_{j,p_1},1\right\rangle_n\right\}^2},
\end{aligned}$$

which is of order $O_p\left\{h_1^{p_1}\times\sqrt{N_1}\times\left(h_1^{1/2}+\sqrt{n^{-1}h_1^{-1}\log n}\right)\right\} = O_p\left(h_1^{p_1}\right)$ by (2.30) in Lemma 2.4. Combining them, the order of the second term is

$$\begin{aligned}
& O_p\left(h_1^{-1/2}\times n^{5/2(2p_*+1)-1/2(2p^*+1)-1}\times h_1^{p_1}\right) \\
=\ & O_p\left(n^{5/2(2p_*+1)-1/2(2p^*+1)-3/2}\right) \\
=\ & o_p\left(n^{3/2(2p_*+1)-1}\right).
\end{aligned}$$

Putting the first and second term together, we have established that

$$\sup_{x\in[a,b]}\left|III_{p_1}(x)\right| = O_p\left(n^{3/2(2p_*+1)-1}\right).$$

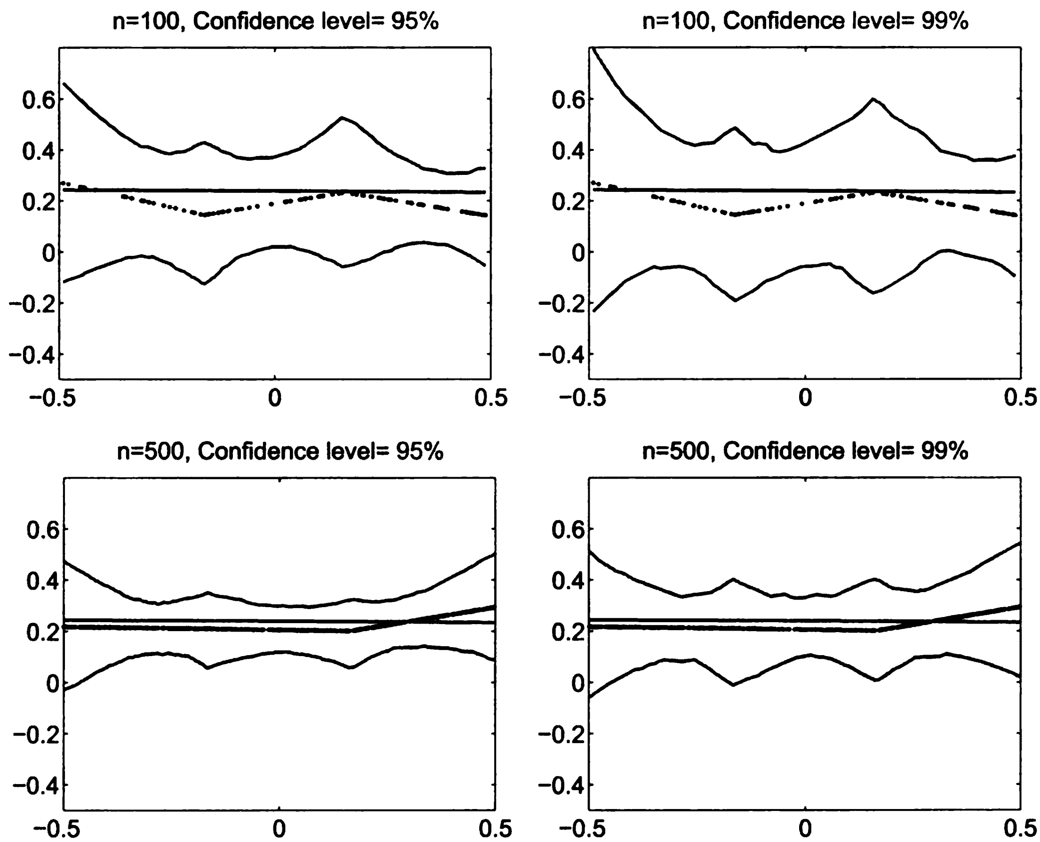


Figure 2.1: For data generated from model (2.22) (with $\sigma_0 = .5, c = 100$) of different sample size n and confidence level $1 - \alpha$, plots of confidence bands for variance (thick solid), the linear spline estimator $\hat{\sigma}^2_{2,2}(x)$ (dotted), and the true function $\sigma^2(x)$ (solid). The bands are computed from bootstrap method.

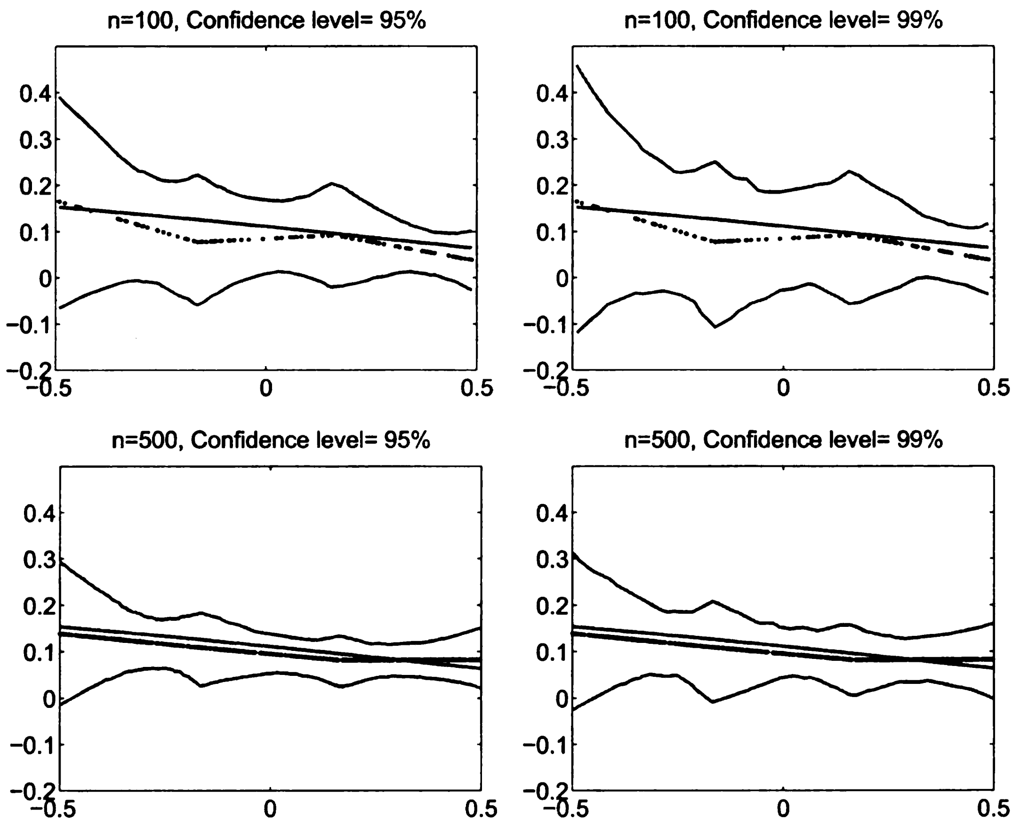


Figure 2.2: For data generated from model (2.22) (with $\sigma_0 = .5, c = 5$) of different sample size n and confidence level $1 - \alpha$, plots of confidence bands for variance (thick solid), the linear spline estimator $\hat{\sigma}^2_{2,2}(x)$ (dotted), and the true function $\sigma^2(x)$ (solid). The bands are computed from bootstrap method.

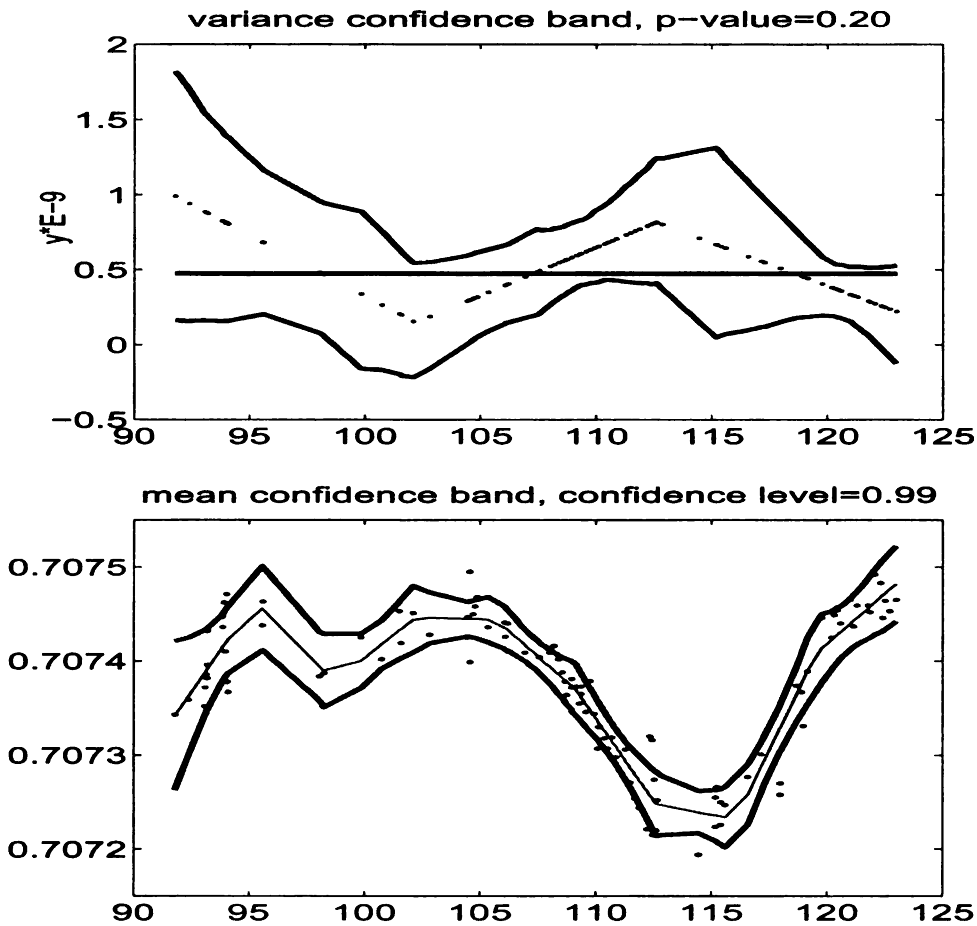


Figure 2.3: For the fossil data, plots of variance confidence bands (thick solid) computed by bootstrap method, the linear spline estimator $\hat{\sigma}^2_{2,2}(x)$ (dotted) and a constant variance function that fits in the confidence band (solid). The lower picture is the data scatter plot and the confidence band for mean (thin solid).

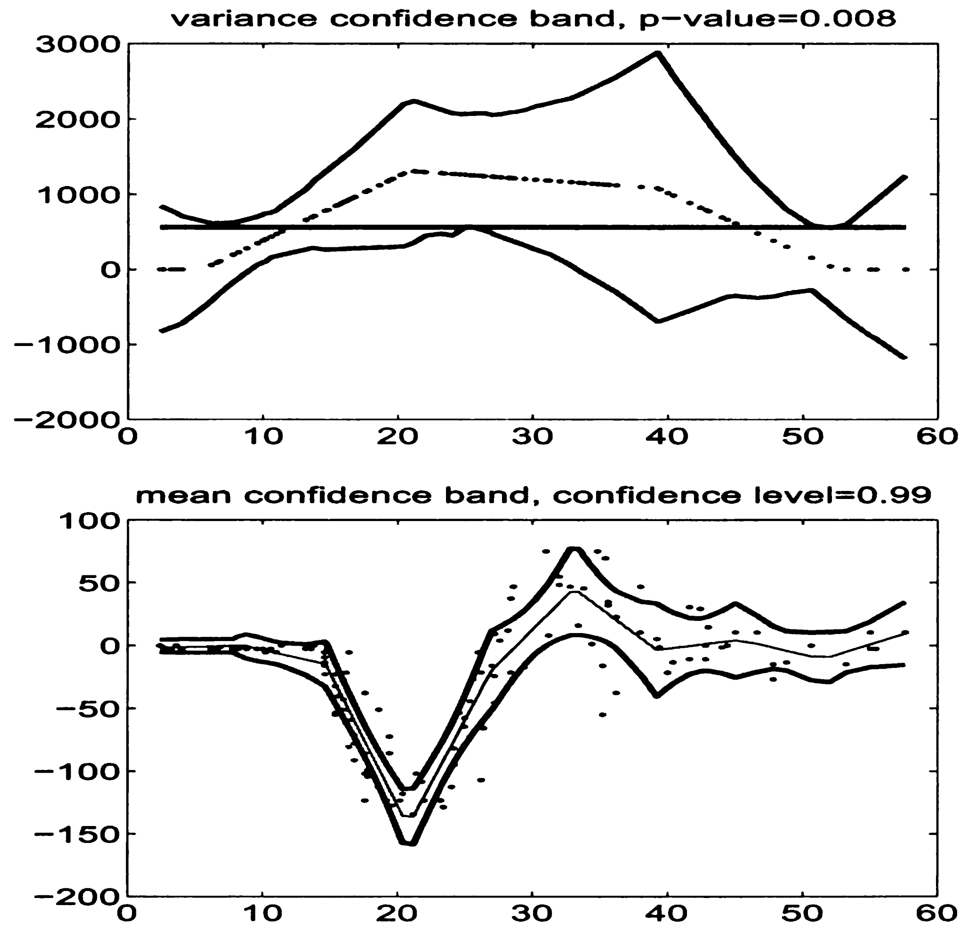


Figure 2.4: For the motorcycle data, plots of variance confidence bands (thick solid) computed by bootstrap method, the linear spline estimator $\hat{\sigma}^2_{2,2}(x)$ (dotted) and a constant variance function that fits in the confidence band (solid). The lower picture is the data scatter plot and the confidence band for mean (thin solid).

Chapter 3

Oracally efficient spline smoothing of NAAR models with simultaneous confidence bands

3.1 Introduction

Non- and semiparametric smoothing has been proven to be useful for analyzing complex time series data due to the flexibility to "let the data speak for themselves". One unavoidable issue in high dimensional smoothing is the "curse of dimensionality", i.e., the poor convergence rate of nonparametric estimation of multivariate functions. Additive regression model of Hastie and Tibshirani [26] has been adapted by Chen and Tsay [6] to autoregression and found wide use in recent years to reduce dimension in nonparametric smoothing of time series. A nonlinear additive autoregressive model (NAAR) is of the form

$$Y_i = m\left(\mathbf{X}_i\right) + \varepsilon_i, \quad m\left(x_1, ..., x_d\right) = c + \sum_{\gamma=1}^{d} m_\gamma\left(x_\gamma\right), \tag{3.1}$$

where the sequence $\left\{Y_i, \mathbf{X}_i^T\right\}_{i=1}^n$ is a length n realization of a $(d+1)$-dimensional strictly stationary process, the d-variate functions $m(\cdot)$ and $\sigma(\cdot)$ are the mean and standard deviation of the response Y_i conditional on the predictor vector $\mathbf{X}_i = \{X_{i1}, ..., X_{id}\}^T$, and $E(\varepsilon_i |\mathbf{X}_i) = 0, E\left(\varepsilon_i^2 |\mathbf{X}_i\right) = \sigma^2(\mathbf{X}_i)$. In the context of NAAR, each predictor $X_{i\gamma}, 1 \le \gamma \le d$ can be observed lagged values of Y_i, such as $X_{i\gamma} = Y_{i-\gamma}$, or of a different times series. The component functions $\{m_\gamma(\cdot)\}_{\gamma=1}^d$ are subjected to the identifiability condition $Em_\gamma\left(X_{i\gamma}\right) \equiv 0, 1 \le \gamma \le d$.

Inference of model (3.1) centers on the estimation and testing of $\{m_\gamma(\cdot)\}_{\gamma=1}^d$. The marginal integration method of Tjøstheim and Auestad [68] and Linton and Nielsen [43] came with asymptotic distribution, which was extended in Sperlich, Tjøstheim and Yang [65] to include second order interactions. Other related works include Fan and Li [17], Yang, Park, Xue and Härdle [78] and Lu, Lundervold, Tjøstheim and Yao [44]. The backfitting idea promoted by [26] was made rigorous in a more complicated form of smooth backfitting by Mammen, Linton and Nielsen [47] and popularized by Nielsen and Sperlich [55] . These kernel based methods are extremely computational intensive, limiting their use for high dimension d, see Martins-Filho and Yang [48] for numerical comparison of these methods. Spline method of Stone [66] had been extended in parallel to NAAR models in Huang and Yang [29], which are fast and easy to implement but lack of limiting distribution. For applications of additive model in medical and environmental research, see Liang et al [41], Roca-Pardinas, Cadarso-Suárez and Gonzalez-Manteiga [57] and Roca-Pardiñas, Cadarso-Suárez, Tahoces and Lado [58].

The two-step estimators of Linton [42] for model (3.1) possess oracle efficiency and are theoretically superior to the aforementioned estimators of $\{m_\gamma(\cdot)\}_{\gamma=1}^d$. If all components $\left\{m_\beta(\cdot)\right\}_{\beta=1,\beta\neq\gamma}^d$ and the constant c were known and removed from the responses, one could estimate $m_\gamma(\cdot)$ from the univariate data $\left\{Y_{i\gamma}, X_{i\gamma}\right\}_{i=1}^n$ in which $\left\{Y_{i\gamma}\right\}_{i=1}^{n}$ are latent oracle responses to the γ-th covariate $\left\{X_{i\gamma}\right\}_{i=1}^{n}$,

$$Y_{i\gamma} = m_{\gamma}\left(X_{i\gamma}\right) + \varepsilon_{i} = Y_{i} - c - \sum_{\beta=1,\beta\neq\gamma}^{d} m_{\beta}\left(X_{i\beta}\right), 1 \leq i \leq n, 1 \leq \gamma \leq d. \quad (3.2)$$

The key idea of [42] is to replace the true $\left\{m_{\beta}(\cdot)\right\}_{\beta=1,\beta\neq\gamma}^{d}$ and c above by some initial kernel estimates, create a pseudo univariate data $\left\{\hat{Y}_{i\gamma}, X_{i\gamma}\right\}_{i=1}^{n}$, and establish the asymptotic equivalence of kernel/local polynomial estimators of $m_{\gamma}(\cdot)$ using either unobservable $\left\{Y_{i\gamma}, X_{i\gamma}\right\}_{i=1}^{n}$ or $\left\{\hat{Y}_{i\gamma}, X_{i\gamma}\right\}_{i=1}^{n}$. Recently, faster oracally efficient estimators have been developed for NAAR time series data by Horowitz and Mammen [27], Wang and Yang [71], making use of orthogonal series/spline initial estimates. The second step estimation is done by kernel method, with pointwise asymptotic distribution. For the sake of discussion, we call the two-step estimator of [42] kernel+kernel, of [27] orthogonal series+kernel and of [71] spline+kernel.

For the NAAR time series models, however, none of the existing methods provide any simultaneous confidence band for $m_{\gamma}(\cdot)$. To address this need, we propose an all new spline+spline oracally efficient estimator that is theoretically superior as it comes with an asymptotically simultaneous confidence band for $m_{\gamma}(\cdot)$, and also computationally more expedient than any existing estimators due to the use of spline instead of kernel in all steps. The asymptotically simultaneous confidence band is that of an univariate regression function in Wang and Yang [72], and is most convenient for inference in the global shape of function $m_{\gamma}(\cdot)$. Such confidence band methodology has been applied to compare the dependence of corn, soybean and wheat crop yields on wetness index under various conditions, see Huang, Wang, Yang and Kravchenko [30]. The spline+spline method is asymptotically oracally efficient as the spline+kernel method of [71], but can be hundreds of times faster in terms of computing, see the comparison in Table 3.2. We see little hope of further reducing the computing burden for model (3.1) over the proposed spline+spline method and still retaining the simultaneous confidence band and oracle efficiency. It seems that the only alternative worth exploring is to use penalized spline instead of B spline smoothing in the second step. For theoretical properties of penalized spline smoothing, see Kauermann, Krivobokova and Fahrmeir [36] and Krivobokova and Kauermann [37].

The chapter is based on a published work Song and Yang [64]. The rest of the chapter is organized as follows. Section 3.2 describes the spline-backfitted spline (SBS) estimators and presents the main theoretical results. Section 3.3 illustrates the idea of proof via decomposition of error. Simulation results are showed in Section 3.4. Most of the technical proofs are in the Appendix.

3.2 The SBS estimator

In this section, we describe the spline-backfitted spline estimation procedure. For convenience, we denote vectors as $\mathbf{x} = (x_1, ..., x_d)$ and take $\|\cdot\|$ as the usual Euclidean norm on R^d, i.e., $\|\mathbf{x}\| = \sqrt{\sum_{\gamma=1}^{d} x_\gamma^2}$, and $\|\cdot\|_\infty$ the sup norm, i.e., $\|\mathbf{x}\|_\infty = \sup_{1\leq\gamma\leq d} |x_\gamma|$. In what follows, denote $\mathbf{Y} = (Y_1, ..., Y_n)^T$ the response vector and $(\mathbf{X}_1, ..., \mathbf{X}_n)^T$ the design matrix. We denote by $\mathbf{1}_k$ the k-vector with all elements 1, and $\mathbf{I}_{k\times k}$ the $k \times k$ identity matrix. Throughout this chapter, we denote the space of the second order smooth functions as $C^{(2)}[0,1] = \left\{m \left| m'' \in C[0,1]\right.\right\}$.

While X_γ may be distributed on $(-\infty, \infty)$, estimation of m_γ is carried out only on compact intervals, and without loss of generality, we take all intervals to be $[0,1], 1 \leq \gamma \leq d$. Let $0 = t_0 < t_1 < ... < t_{N+1} = 1$ be a sequence of equally spaced knots, dividing $[0,1]$ into $(N+1)$ subintervals of length $h = h_n = 1/(N+1)$ with a preselected integer $N \sim n^{1/5}$ given in Assumption (A5), and let $0 = t_0^* < t_1^* < ... < t_{N^*+1}^* = 1$ be another sequence of equally-spaced knots, dividing $[0,1]$ into (N^*+1) subintervals of length $H = H_n = (N^*+1)^{-1}$ where $N^* \sim n^{2/5} \log n$ is another preselected integer, see Assumption (A5). Next, we define the constant spline basis I_{J^*} for step one and the linear spline basis b_J for step two de Boor ([10], page 89) as follows,

$$I_0(x) \equiv 1, 0 < x < 1$$

$$\begin{aligned} I_{J^*}(x) &= \begin{cases} 1 & J^*H \le x < (J^*+1)H, \\ 0 & \text{otherwise}, \end{cases}, 1 \le J^* \le N^*, \\ b_J(x) &= K\left(\frac{x-t_J}{h}\right), 1 \le J \le N+1, K(u) = (1-|u|)_+ . \end{aligned}$$

We denote by G_γ the linear space spanned by $\{b_J(x_\gamma)\}_{J=0}^{N+1} = \{1, b_J(x_\gamma)\}_{J=1}^{N+1}$, whose elements are called linear splines, piecewise linear functions of x_γ which are continuous on $[0,1]$ and linear on each subinterval $[t_J, t_{J+1}]$, $0 \le J \le N$. We denote by $G_{n,\gamma} \subset R^n$ the corresponding subspace of R^n spanned by $\left\{1, \left\{b_J\left(X_{i\gamma}\right)\right\}_{i=1}^n\right\}_{J=1}^{N+1}$. Similarly, define the $\{1+dN^*\}$-dimensional space G^* of additive constant spline functions as the space spanned by $\{1, I_{J^*}(x_\gamma)\}_{\gamma=1,J^*=1}^{d,N^*}$, and the corresponding subspace spanned by $\left\{1, \left\{I_{J^*}\left(X_{i\gamma}\right)\right\}_{i=1}^n\right\}_{\gamma=1,J^*=1}^{d,N^*}$ as $G_n^* \subset R^n$. As $n \to \infty$, with probability approaching one, the dimension of $G_{n,\gamma}$ becomes $N+2$, and the dimension of G_n^* becomes $1+dN^*$.

The function $m(\mathbf{x})$ has a multivariate additive regression spline (MARS) estimator $\hat{m}(\mathbf{x}) = \hat{m}_n(\mathbf{x})$, the unique element of G^*, so the vector $\{\hat{m}(\mathbf{X}_1), ..., \hat{m}(\mathbf{X}_n)\}^T \in G_n^*$ best approximates the response vector $\mathbf{Y}$. For spline regression, we introduce the following weights,

$$\begin{aligned} W_{i\gamma} &= 1\left(0 \le X_{i\gamma} \le 1\right), 1 \le i \le n, 1 \le \gamma \le d, & (3.3) \\ \mathbf{W}_\gamma &= \text{diag}(W_{1\gamma}, ..., W_{n\gamma}), 1 \le \gamma \le d, \\ W_i^* &= 1\left(0 \le \mathbf{X}_i \le 1\right) = \Pi_{\gamma=1}^d W_{i\gamma}, 1 \le i \le n, & (3.4) \end{aligned}$$

$$\mathbf{W}^* = \text{diag}(W_1^*, ..., W_n^*),$$

and impose on additive component functions the identifiability condition

$$Em_\gamma\left(X_{i\gamma}\right)W_i^* \equiv 0, 1 \leq \gamma \leq d. \tag{3.5}$$

Define next a weighted spline estimator of m as

$$\hat{m}(\mathbf{x}) = \underset{g \in G^*}{\text{argmin}} \sum_{i=1}^{n} \{Y_i - g(\mathbf{X}_i)\}^2 W_i^* = \hat{\lambda}_0' + \sum_{\gamma=1}^{d} \sum_{J^*=1}^{N^*} \hat{\lambda}_{J^*,\gamma}' I_{J^*}(x_\gamma), \tag{3.6}$$

where $\left(\hat{\lambda}_0', \hat{\lambda}_{1,1}', ..., \hat{\lambda}_{N^*,d}'\right)$ is the solution of the weighted least squares problem

$$\begin{aligned} &\left\{\hat{\lambda}_0', \hat{\lambda}_{1,1}', ..., \hat{\lambda}_{N^*,d}'\right\}^T \\ = &\underset{R^{d(N^*)+1}}{\text{argmin}} \sum_{i=1}^{n} \left\{Y_i - \lambda_0 - \sum_{\gamma=1}^{d} \sum_{J^*=1}^{N^*} \lambda_{J^*,\gamma} I_{J^*}\left(X_{i\gamma}\right)\right\}^2 W_i^*. \end{aligned}$$

Pilot estimator of each component function is

$$\hat{m}_\gamma(x_\gamma) = \sum_{J^*=1}^{N^*} \hat{\lambda}_{J^*,\gamma}' \left\{I_{J^*}(x_\gamma) - n^{-1} \sum_{i=1}^{n} I_{J^*}\left(X_{i\gamma}\right) W_i^*\right\}, \quad 1 \leq \gamma \leq d \tag{3.7}$$

which satisfies the empirical analog of (3.5): $n^{-1}\sum_{i=1}^{n} \hat{m}_\gamma\left(X_{i\gamma}\right) W_i^* = 0, 1 \leq \gamma \leq d$. These pilot estimators are used to define pseudo-responses $\hat{Y}_{i\gamma}$, $\forall 1 \leq \gamma \leq d$, which approximate the "oracle" responses $Y_{i\gamma}$ in (3.2). Specifically, we define $\hat{Y}_{i\gamma} = Y_i - \hat{c} - \sum_{\beta=1,\beta\neq\gamma}^{d} \hat{m}_\beta\left(X_{i\beta}\right)$, where $\hat{c} = \overline{Y}_n = n^{-1}\sum_{i=1}^{n} Y_i$, which is a $\sqrt{n}$-consistent estimator of c by central limit theorem for strongly mixing sequences. Correspondingly, we denote vectors

$$\hat{\mathbf{Y}}_\gamma = \left\{\hat{Y}_{1\gamma}, ..., \hat{Y}_{n\gamma}\right\}^T, \mathbf{Y}_\gamma = \left\{Y_{1\gamma}, ..., Y_{n\gamma}\right\}^T. \tag{3.8}$$

We define the spline-backfitted spline (SBS) estimator of $m_\gamma\left(x_\gamma\right)$ as $\hat{m}_{\gamma,\mathrm{SBS}}\left(x_\gamma\right)$ based on $\left\{\hat{Y}_{i\gamma}, X_{i\gamma}\right\}_{i=1}^n$, which attempts to mimic the would-be spline estimator $\tilde{m}_{\gamma,\mathrm{S}}\left(x_\gamma\right)$ of $m_\gamma\left(x_\gamma\right)$ based on $\left\{Y_{i\gamma}, X_{i\gamma}\right\}_{i=1}^n$ if the unobservable "oracle" responses $\left\{Y_{i\gamma}\right\}_{i=1}^n$ were available. To be precise, for $0 \le x_\gamma \le 1$,

$$\begin{aligned} \hat{m}_{\gamma,\mathrm{SBS}}\left(x_\gamma\right) &= \underset{g_\gamma \in G_\gamma}{\operatorname{argmin}} \sum_{i=1}^n \left\{\hat{Y}_{i\gamma} - g_\gamma\left(X_{i\gamma}\right)\right\}^2 W_{i\gamma}, \\ \tilde{m}_{\gamma,\mathrm{S}}\left(x_\gamma\right) &= \underset{g_\gamma \in G_\gamma}{\operatorname{argmin}} \sum_{i=1}^n \left\{Y_{i\gamma} - g_\gamma\left(X_{i\gamma}\right)\right\}^2 W_{i\gamma}. \end{aligned} \tag{3.9}$$

Before presenting the main results, we state the following assumptions.

(A1) *The additive component functions* $m_\gamma\left(x_\gamma\right) \in C^{(2)}[0,1], \forall\gamma = 1, ..., d$

(A2) *There exist positive constants* K_0 *and* λ_0 *such that* $\alpha(n) \le K_0 e^{-\lambda_0 n}$ *holds for all* n*, with the* α*-mixing coefficients for* $\left\{\mathbf{Z}_i = \left(\mathbf{X}_i^T, \varepsilon_i\right)\right\}_{i=1}^n$ *defined as*

$$\alpha(k) = \sup_{B\in\sigma\{\mathbf{Z}_s, s\le t\}, C\in\sigma\{\mathbf{Z}_s, s\ge t+k\}} \left|P(B\cap C) - P(B)P(C)\right|, \ k \ge 1. \tag{3.10}$$

(A3) *The noise* ε_i *satisfies* $E\left(\varepsilon_i \left|\mathbf{X}_i\right.\right) = 0$, $E\left(\varepsilon_i^2 \left|\mathbf{X}_i\right.\right) = \sigma^2\left(\mathbf{X}_i\right)$, $E\left(\left|\varepsilon_i\right|^{2+\delta}\left|\mathbf{X}_i\right.\right) < M_\delta$ *for some* $\delta > 1/2$ *and a finite positive* M_δ *and* $\sigma(\mathbf{x})$ *is continuous on* $[0,1]^d$, $0 < c_\sigma \le \inf_{\mathbf{x}\in[0,1]^d} \sigma(\mathbf{x}) \le \sup_{\mathbf{x}\in[0,1]^d} \sigma(\mathbf{x}) \le C_\sigma < \infty$. *Consequently, for* $\gamma = 1, ..., d, \sigma_\gamma^2\left(x_\gamma\right) = E\left\{\sigma^2(\mathbf{X})\left|X_\gamma = x_\gamma\right.\right\}$ *satisfies also* $c_\sigma \le \inf_{x_\gamma\in[0,1]} \sigma_\gamma\left(x_\gamma\right) \le \sup_{x_\gamma\in[0,1]} \sigma_\gamma\left(x_\gamma\right) \le C_\sigma$.

(A4) *The density function* $f(\mathbf{x})$ *of* $\mathbf{X}$ *is continuous and* $0 < c_f \le \inf_{\mathbf{x}\in[0,1]^d} f(\mathbf{x}) \le \sup_{\mathbf{x}\in[0,1]^d} f(\mathbf{x}) \le C_f < \infty$.

(A5) *The number of interior knots in estimation step one* $N^* \sim n^{2/5}\log n$*, i.e.,*

$c_{N^*} n^{2/5} \log n \leq N^* \leq C_{N^*} n^{2/5} \log n$ *for some positive constants* c_{N^*}, C_{N^*}. *The number of interior knots in estimation step two* $N \sim n^{1/5}$.

Remark. 1. The smoothness Assumption (A1) is nearly minimal. (A2)-(A4) are typical in the nonparametric literature, for instance, Fan and Gijbels [14]. For (A5), the optimal order of N in the second step ensures bias and variance trade-off. Theorem 3.4 on the oracle efficiency of $\hat{m}_{\gamma,\mathrm{SBS}}(x_\gamma)$ remains true if N^* is of the more general form $n^{2/5} N'$, where the sequence N' satisfies $\log(n)/N' = O(1)$, $n^{-\theta} N' \to 0$ for any $\theta > 0$, see Proposition A.1., A.2. and Lemma A.1., A.2. for the proof of Theorem 3.4 in Appendix.

Remark. 2. Assumptions (A1)-(A4) are satisfied by many commonly used time series models, such as those in Chen and Tsay [6].

Theorem 3.4. *Under Assumptions (A1) to (A5), as* $n \to \infty$, *the SBS estimator* $\hat{m}_{\gamma,\mathrm{SBS}}(x_\gamma)$ *and the oracle smoother* $\tilde{m}_{\gamma,\mathrm{S}}(x_\gamma)$ *given in (3.9) satisfy*

$$\sup_{x_\gamma \in [0,1]} \left| \hat{m}_{\gamma,\mathrm{SBS}}(x_\gamma) - \tilde{m}_{\gamma,\mathrm{S}}(x_\gamma) \right| = O_P\left(n^{-2/5} (\log n)^{-1} \right).$$

Theorem 3.4 provides that the maximal deviation of $\hat{m}_{\gamma,\mathrm{SBS}}(x_\gamma)$ from $\tilde{m}_{\gamma,\mathrm{S}}(x_\gamma)$ over $[0,1]$ is of the order $O_P\left(n^{-2/5}(\log n)^{-1}\right) = o_P\left(n^{-2/5}(\log n)^{1/2}\right)$, which is needed for the maximal deviation of $\hat{m}_{\gamma,\mathrm{SBS}}(x_\gamma)$ from $m_\gamma(x_\gamma)$ over $[0,1]$ and the maximal deviation of $\tilde{m}_{\gamma,\mathrm{S}}(x_\gamma)$ from $m_\gamma(x_\gamma)$ to have the same asymptotic distribution, of order $n^{-2/5}(\log n)^{1/2}$. The estimator $\hat{m}_{\gamma,\mathrm{SBS}}(x_\gamma)$ is therefore asymptotically oracally efficient, i.e., it is asymptotically equivalent to the oracle smoother $\tilde{m}_{\gamma,\mathrm{S}}(x_\gamma)$ and in particular, the next theorem follows. The simultaneous confidence band given in (3.11) has width of order $n^{-2/5}(\log n)^{1/2}$ at any point $x_\gamma \in [0,1]$, consistent with published works on nonparametric simultaneous confidence bands such as Xia [75], Claeskens and Van Keilegom [7].

Theorem 3.5. *Under Assumptions (A1)-(A5), for any* $p \in (0,1)$ *, as* $n \to \infty$*, an asymptotic* $100(1-p)\%$ *simultaneous confidence band for* $m_\gamma(x_\gamma)$ *is*

$$\hat{m}_{\gamma,\mathrm{SBS}}(x_\gamma) \pm 2\hat{\sigma}_\gamma(x_\gamma)\left\{3\boldsymbol{\Delta}^T(x_\gamma)\Xi_{j(x_\gamma)}\boldsymbol{\Delta}(x_\gamma)\log\left(\frac{N+1}{2}\right)\hat{f}_\gamma(x_\gamma)\right.$$
$$\left.nh\right\}^{1/2}\left[1-\{2\log(N+1)\}^{-1}\left[\log(p/4)+\frac{1}{2}\log\{4\pi\log(N+1)\}\right]\right], \quad (3.11)$$

where $\hat{\sigma}_\gamma(x_\gamma)$ *and* $\hat{f}_\gamma(x_\gamma)$ *are some consistent estimators of* $\sigma_\gamma(x_\gamma)$ *and* $f_\gamma(x_\gamma)$,

$$j(x_\gamma) = \min\{[x_\gamma/h], N\}, \delta(x_\gamma) = \left\{x_\gamma - t_{j(x_\gamma)}\right\}/h,$$

and

$$\boldsymbol{\Delta}(x_\gamma) = \begin{pmatrix} c_{j(x_\gamma)-1}\{1-\delta(x_\gamma)\} \\ c_{j(x_\gamma)}\delta(x_\gamma) \end{pmatrix}, \ c_j = \begin{cases} \sqrt{2} & j = 0, N+1 \\ 1 & 1 \le j \le N \end{cases},$$

$$\Xi_j = \begin{pmatrix} l_{j+1,j+1} & l_{j+1,j+2} \\ l_{j+2,j+1} & l_{j+2,j+2} \end{pmatrix}, 0 \le j \le N,$$

where terms $\{l_{ik}\}_{|i-k|\le 1}$ *are the entries of the inverse of the* $(N+2)\times(N+2)$ *matrix* $\mathbf{M}_{N+2}$,

$$\mathbf{M}_{N+2} = \begin{pmatrix} 1 & \sqrt{2}/4 & & & & 0 \\ \sqrt{2}/4 & 1 & 1/4 & & & \\ & 1/4 & 1 & \ddots & & \\ & & \ddots & \ddots & 1/4 & \\ & & & 1/4 & 1 & \sqrt{2}/4 \\ 0 & & & & \sqrt{2}/4 & 1 \end{pmatrix}.$$

We refer the proof of the theorem to Wang and Yang [72].

3.3 Decomposition

In this section, we provide insight on the proof of Theorem 3.4. Recalling the notaion of W_i^* and $W_{i\gamma}$ defined in (3.4), (3.3), for any functions ϕ, φ on $[0,1]^d$, define the empirical inner product, empirical norm and empirical mean restricted on $[0,1]^d$ as $\langle \phi, \varphi \rangle_{2,n}^* = n^{-1} \sum_{i=1}^n \phi(\mathbf{X}_i) \varphi(\mathbf{X}_i) W_i^*$, $\|\phi\|_{2,n}^{*2} = n^{-1} \sum_{i=1}^n \phi^2(\mathbf{X}_i) W_i^*$, $E_n^* \phi = n^{-1} \sum_{i=1}^n \phi(\mathbf{X}_i) W_i^* = \langle 1, \phi \rangle_{2,n}^*$ respectively. In addition, if functions ϕ, φ are $L^2[0,1]^d$-integrable, define the theoretical inner product and its corresponding theoretical L^2 norm as $\langle \phi, \varphi \rangle_2^* = E\left\{\phi(\mathbf{X}_i) \varphi(\mathbf{X}_i) W_i^*\right\}$, $\|\phi\|_2^{*2} = E\left\{\phi^2(\mathbf{X}_i) W_i^*\right\}$. A function ϕ is called theoretically centered (empirically centered) if $E\phi W_i^* = 0$ ($E_n^* \phi = 0$). The additive component function m_γ and its pilot estimator $\hat{m}_\gamma$ defined in (3.7) are therefore theoretically centered (empirically centered). In the second step, for any functions ϕ, φ on $[0,1]$, for any $1 \le \gamma \le d$, similarly define $\langle \phi, \varphi \rangle_{2,n,\gamma} = n^{-1} \sum_{i=1}^n \phi\left(X_{i\gamma}\right) \varphi\left(X_{i\gamma}\right) W_{i\gamma}$, $\|\phi\|_{2,n,\gamma}^2 = n^{-1} \sum_{i=1}^n \phi^2\left(X_{i\gamma}\right) W_{i\gamma}$, $E_{n,\gamma} \phi = n^{-1} \sum_{i=1}^n \phi\left(X_{i\gamma}\right) W_{i\gamma} = \langle 1, \phi \rangle_{2,n,\gamma}$ respectively. In addition, if functions ϕ, φ are $L^2[0,1]$-integrable, define the theoretical inner product and its corresponding theoretical L^2 norm as $\langle \phi, \varphi \rangle_{2,\gamma} = E\left\{\phi\left(X_{i\gamma}\right) \varphi\left(X_{i\gamma}\right) W_{i\gamma}\right\}$, $\|\phi\|_{2,\gamma}^2 = E\left\{\phi^2\left(X_{i\gamma}\right) W_{i\gamma}\right\}$.

The function space G_γ introduced in Section 3.2 is expressed more conveniently for asymptotic analysis via the following standardized B spline basis

$$B_{J,\gamma}(x_\gamma) = \frac{b_J(x_\gamma)}{\|b_J\|_{2,\gamma}}, 0 \le J \le N+1. \tag{3.12}$$

Likewise, G^* is spanned by $\left\{1, B_{J^*,\gamma}^*(x_\gamma)\right\}_{\gamma=1,J^*=1}^{d,N^*}$, in which the new theoretically centered and standardized B spline basis are

$$B_{J^*,\gamma}^*(x_\gamma) = \frac{b_{J^*,\gamma}^*(x_\gamma)}{\left\|b_{J^*,\gamma}^*\right\|_2}, 1 \le \gamma \le d, 1 \le J^* \le N^*, \tag{3.13}$$

in which

$$b^*_{J^*,\gamma}(x_\gamma) = I_{J^*+1,\gamma}(x_\gamma) - \frac{c_{J^*+1,\gamma}}{c_{J^*,\gamma}} I_{J^*,\gamma}(x_\gamma), \tag{3.14}$$
$$c_{J^*,\gamma} = \left\langle 1, I_{J^*,\gamma} \right\rangle_2 .$$

Simple linear algebra shows that

$$\hat{m}(\mathbf{x}) = \hat{\lambda}_0 + \sum_{\gamma=1}^{d} \sum_{J^*=1}^{N^*} \hat{\lambda}_{J^*,\gamma} B^*_{J^*,\gamma}(x_\gamma), \quad \mathbf{x} \in [0,1]^d \tag{3.15}$$

where $\left(\hat{\lambda}_0, \hat{\lambda}_{1,1}, ..., \hat{\lambda}_{N^*,d}\right)$ are solutions of the following least squares problem

$$\left\{\hat{\lambda}_0, \hat{\lambda}_{1,1}, ..., \hat{\lambda}_{N^*,d}\right\}^T$$
$$= \underset{R^{d(N^*)+1}}{\operatorname{argmin}} \sum_{i=1}^{n} \left\{ Y_i - \lambda_0 - \sum_{\gamma=1}^{d} \sum_{J^*=1}^{N^*} \lambda_{J^*,\gamma} B^*_{J^*,\gamma}\left(X_{i\gamma}\right) \right\}^2 W_i^* . \tag{3.16}$$

Define for any n-dimensional vector $\mathbf{\Lambda} = \{\Lambda_i\}_{i=1}^n$, the spline function constructed from the projection of $\mathbf{\Lambda}$ on the inner product space $\left(G_n, \langle\cdot,\cdot\rangle_{2,n}\right)$ as $\mathbf{P}_n\mathbf{\Lambda}(\mathbf{x}) = \hat{\lambda}_0 + \sum_{\gamma=1}^{d} \sum_{J^*=1}^{N^*} \hat{\lambda}_{J^*,\gamma} B^*_{J^*,\gamma}(x_\gamma)$, with coefficients $\left(\hat{\lambda}_0, \hat{\lambda}_{1,1}, ..., \hat{\lambda}_{N^*,d}\right)$ given in (3.16) with Y_i's replaced by Λ_i's. The multivariate function $\mathbf{P}_n\mathbf{\Lambda}(\mathbf{x})$ has empirically centered components $\mathbf{P}_{n,\gamma}\mathbf{\Lambda}(x_\gamma)$, $\gamma = 1, ..., d$

$$\mathbf{P}_{n,\gamma}\mathbf{\Lambda}(x_\gamma) = \sum_{J^*=1}^{N^*} \hat{\lambda}_{J^*,\gamma} \left\{ B^*_{J^*,\gamma}(x_\gamma) - n^{-1} \sum_{i=1}^{n} B^*_{J^*,\gamma}\left(X_{i\gamma}\right) W_i^* \right\}. \tag{3.17}$$

The estimators $\hat{m}(\mathbf{x}), \hat{m}_\gamma(x_\gamma)$ in (3.15) and (3.7) are rewritten as $\hat{m}(\mathbf{x}) = \mathbf{P}_n\mathbf{Y}(\mathbf{x})$, $\hat{m}_\gamma(x_\gamma) = \mathbf{P}_{n,\gamma}\mathbf{Y}(x_\gamma)$. For linear operators $\mathbf{P}_n$, $\mathbf{P}_{n,\gamma}$, $\gamma = 1, ..., d$, using the relation $\mathbf{Y} = \mathbf{m} + \mathbf{E}$, where the signal and noise vectors are $\mathbf{m} = \{m(\mathbf{X}_i)\}_{i=1}^n, \mathbf{E} =$

$\{\varepsilon_i\}_{i=1}^n$, one has the following decomposition for $\gamma = 1, ..., d$

$$\hat{m}(\mathbf{x}) = \tilde{m}(\mathbf{x}) + \tilde{\varepsilon}(\mathbf{x}),\ \hat{m}_\gamma(x_\gamma) = \tilde{m}_\gamma(x_\gamma) + \tilde{\varepsilon}_\gamma(x_\gamma), \tag{3.18}$$

in which the noiseless spline smoothers and the variance spline components are

$$\begin{aligned} \tilde{m}(\mathbf{x}) &= \mathbf{P}_n \mathbf{m}(\mathbf{x}), \tilde{m}_\gamma(x_\gamma) = \mathbf{P}_{n,\gamma}\mathbf{m}(x_\gamma), \\ \tilde{\varepsilon}(\mathbf{x}) &= \mathbf{P}_n \mathbf{E}(\mathbf{x}), \tilde{\varepsilon}_\gamma(x_\gamma) = \mathbf{P}_{n,\gamma}\mathbf{E}(x_\gamma). \end{aligned} \tag{3.19}$$

Additionally, we can write $\tilde{\varepsilon}(\mathbf{x}) = \tilde{\mathbf{a}}^{*T}\mathbf{B}^*(\mathbf{x})$, $\tilde{\mathbf{a}}^* = \left\{\tilde{a}_0^*, \tilde{a}_{1,1}^*, ..., \tilde{a}_{N^*,d}^*\right\}^T = \left(\mathbf{B}^{*T}\mathbf{W}^*\mathbf{B}^*\right)^{-1}\mathbf{B}^{*T}\mathbf{W}^*\mathbf{E}$, where vector $\mathbf{B}^*(\mathbf{x})$ and matrix $\mathbf{B}^*$ are defined as

$$\mathbf{B}^*(\mathbf{x}) = \left\{1, B_{1,1}^*(x_1), ..., B_{N^*,d}^*(x_d)\right\},\ \mathbf{B}^* = \left\{\mathbf{B}^*(\mathbf{X}_1), ..., \mathbf{B}^*(\mathbf{X}_n)\right\}^T. \tag{3.20}$$

Clearly $\tilde{\mathbf{a}}^*$ equals to

$$\left\{\begin{array}{cc} 1 & \mathbf{0}_{dN^*}^T \\ \mathbf{0}_{dN^*} & \left\langle B_{J^*,\gamma}^*, B_{J^{*\prime},\gamma'}^*\right\rangle_{2,n}^* \end{array}\right\}_{\substack{1\le\gamma,\gamma'\le d, \\ 1\le J^*,J^{*\prime}\le N^*}}^{-1} \times \left\{\begin{array}{l} \frac{1}{n}\sum_{i=1}^n W_i^*\varepsilon_i \\ \frac{1}{n}\sum_{i=1}^n B_{J^*,\gamma}^*\left(X_{i\gamma}\right)W_i^*\varepsilon_i \end{array}\right\}_{\substack{1\le J^*\le N^* \\ 1\le\gamma\le d}},$$

where $\mathbf{0}_p$ is a p-vector with all elements 0.

The second step spline smoothing is interpreted similarly. For notational simplicity, take $\gamma = 1$ and denote $\mathbf{X}_{i,\text{-}1} = (X_{i2}, ..., X_{id})^T$ for $1 \le i \le n$, and $\mathbf{x}_{\text{-}1} = (x_2, ..., x_d)^T$. Denote $B_{J^*,\text{-}1}^*(\mathbf{x}_{\text{-}1}) = (B_{J^*,2}^*(x_2), ..., B_{J^*,d}^*(x_d))^T$, and so $m_{\text{-}1}(\mathbf{x}_{\text{-}1})$, $\hat{m}_{\text{-}1}(\mathbf{x}_{\text{-}1})$, $\tilde{m}_{\text{-}1}(\mathbf{x}_{\text{-}1})$ and $\tilde{\varepsilon}_{\text{-}1}(\mathbf{x}_{\text{-}1})$. Define $\mathbf{B}(x_1) = \left\{B_{0,1}(x_1), ..., B_{N+1,1}(x_1)\right\}$,

$\mathbf{B} = \{\mathbf{B}(X_{11}), ..., \mathbf{B}(X_{1n})\}^T$, then $\hat{m}_{1,\mathrm{SBS}}(x_1) = \mathbf{B}(x_1)\left(\frac{\mathbf{B}^T\mathbf{W}\mathbf{B}}{n}\right)^{-1}\frac{\mathbf{B}^T}{n}\mathbf{W}\hat{\mathbf{Y}}_1$, $\tilde{m}_{1,\mathrm{S}}(x_1) = \mathbf{B}(x_1)\left(\frac{\mathbf{B}^T\mathbf{W}\mathbf{B}}{n}\right)^{-1}\frac{\mathbf{B}^T}{n}\mathbf{W}\mathbf{Y}_1$, where $\hat{\mathbf{Y}}_1$ and $\mathbf{Y}_1$ are defined in (3.8).

Making use of the definition of $\hat{c}$ and the decomposition (3.18), the difference between the smoothed backfitted estimator $\hat{m}_{1,\mathrm{SBS}}(x_1)$ and the smoothed "oracle" estimator $\tilde{m}_{1,\mathrm{S}}(x_1)$, both given above, is

$$\tilde{m}_{1,\mathrm{S}}(x_1) - \hat{m}_{1,\mathrm{SBS}}(x_1) = \mathbf{B}(x_1)\left(\frac{\mathbf{B}^T\mathbf{W}\mathbf{B}}{n}\right)^{-1}\frac{\mathbf{B}^T}{n}\mathbf{W}\left(\mathbf{Y}_1 - \hat{\mathbf{Y}}_1\right)$$

$$= \mathbf{B}(x_1)\left(\frac{\mathbf{B}^T\mathbf{W}\mathbf{B}}{n}\right)^{-1}\left(\frac{\mathbf{B}^T}{n}\mathbf{W}(-\hat{c}+c) + \mathbf{\Psi}_b + \mathbf{\Psi}_v\right),$$

$\mathbf{\Psi}_b$ and $\mathbf{\Psi}_v$ are the following vectors

$$\mathbf{\Psi}_b = \left\{n^{-1}\sum_{i=1}^{n} B_{J,1}(X_{i1})W_i^*\left\{\tilde{m}_{\text{-}1}\left(\mathbf{X}_{i,\text{-}1}\right)^T - m_{\text{-}1}\left(\mathbf{X}_{i,\text{-}1}\right)^T\right\}\mathbf{1}_{d-1}\right\}_{J=1}^{N+1} \quad (3.21)$$

$$\mathbf{\Psi}_v = \left\{n^{-1}\sum_{i=1}^{n} B_{J,1}(X_{i1})W_i^*\tilde{\varepsilon}_{\text{-}1}\left(\mathbf{X}_{i,\text{-}1}\right)^T\mathbf{1}_{d-1}\right\}_{J=1}^{N+1}, \quad (3.22)$$

here we need the fact that $W_i^* W_{i\gamma} = W_i^*$.

According to Propositions 3.1 and 3.2 in Appendix, both of these two terms have order $O_p\left(h^{1/2}n^{-2/5}(\log n)^{-1}\right) = O_p\left(n^{-1/2}(\log n)^{-1}\right)$.

3.4 Simulation example

In this section, we carry out simulation experiments to illustrate the finite-sample behavior of SBS estimators. The programming codes are available in R, see http://www.r-project.org.

The number of interior knots N^* and N for the spline estimation are calculated as $N^* = \min\left(\left[c_{11}n^{2/5}\log n\right] + c_{12} + 1, \left[(n/2-1)\,d^{-1}\right]\right)$, and $N = \left[c_{21}n^{1/5}\right] + c_{22} + 1$, in which $[a]$ denotes the integer part of a. Tuning constants $c_{11} = 5, c_{21} = 3, c_{12} = c_{22} = 1$ worked well, and we used them by default. The additional constraint that $N^* \leq (n/2-1)\,d^{-1}$ ensures that the number of terms in the linear least squares problem (3.16), $1 + dN^*$, is no greater than $n/2$.

Alternatively, one can use BIC to choose the number of knots. To be specific, in the second step, let $q_n = (1 + N_n)$ be the total number of parameters. Then $\hat{N}^{\text{opt}}$ is the one minimizing the BIC value. $\text{BIC} = \log(\text{MSE}) + q_n \log(n)/n$, with $\text{MSE} = \sum_{i=1}^{n}\{Y_i - \hat{Y}_i\}^2/n$. For computing speed consideration, we have not experimented with this option in this chapter.

Consider the following nonlinear additive heteroscedastic model

$$Y_t = \sum_{\gamma=1}^{d} \sin\left(2\pi X_{t\gamma}\right) + \varepsilon_t,\ \varepsilon_t \overset{i.i.d}{\sim} N\left(0, \sigma^2(\mathbf{X}_t)\right), \tag{3.23}$$

in which $\mathbf{X}_t = \{X_{t1}, ..., X_{td}\}^T$ is generated as $X_{t\gamma} = \Phi\left\{\left(1-a^2\right)^{-1/2} Z_{t\gamma}\right\} - 1/2, 1 \leq \gamma \leq d$ where the $Z_{t\gamma}$'s follow a vector autoregression (VAR) equation

$$\begin{aligned}
\mathbf{Z}_1 &\sim N\left(0_d, \left(1-a^2\right)^{-1}\boldsymbol{\Sigma}\right),\ \mathbf{Z}_t = a\mathbf{Z}_{t-1} + \boldsymbol{\varepsilon}_t,\ \boldsymbol{\varepsilon}_t \sim N(0, \boldsymbol{\Sigma}),\ 2 \leq t \leq n, \\
\boldsymbol{\Sigma} &= (1-\rho)\,\mathbf{I}_{d\times d} + \rho \mathbf{1}_d \mathbf{1}_d^T,\ a = 0.3,\ 0 < \rho < 1,
\end{aligned}$$

with stationary distribution $\mathbf{Z}_t = (Z_{t1}, ..., Z_{td})^T \sim N\left(0_d, \left(1-a^2\right)^{-1}\boldsymbol{\Sigma}\right)$. Hence $\{\mathbf{X}_t\}_{t=1}^n$ is a sequence of geometrically strong mixing random variables with marginal distribution $U[-0.5, 0.5]$. The standard deviation function is

$$\sigma(\mathbf{X}_t) = \sigma_0 \frac{1}{2} \cdot \frac{5 - \exp\left(\sum_{\gamma=1}^{d}\left|X_{t\gamma}\right|/d\right)}{5 + \exp\left(\sum_{\gamma=1}^{d}\left|X_{t\gamma}\right|/d\right)}, \sigma_0 = 0.5,$$

which ensures that our design is heteroscedastic.

The SBS estimator $\hat{m}_{\gamma,\mathrm{SBS}}(x_\gamma)$ and the oracle smoother $\tilde{m}_{\gamma,\mathrm{S}}(x_\gamma)$ are compared in terms of coverage probabilities of confidence bands for sample sizes $n = 100, 500, 1000$, with confidence level $1 - p = 0.95$. Table 3.1 contains the coverage probabilities as the percentage of complete coverage of the first true curve $\sin(2\pi x)$ at all data points $\{X_{t1}\}_{t=1}^{n}$ by the confidence bands in (3.11), over 500 replications of sample size n, for $d = 4, 10$ and $\rho = 0, 0.3$. The results are satisfactory as the empirical probabilities rapidly become greater than the nominal probability of 0.95 as n becomes large. To show that the SBS estimator $\hat{m}_{\gamma,\mathrm{SBS}}(x_\gamma)$ is as efficient as the

Table 3.1: Coverage frequencies from 500 replications.

	r	$n = 100$	$n = 500$	$n = 1000$
	0	0.86	0.972	0.966
$d = 4$	.3	0.876	0.956	0.964
	0	.0848	0.974	0.97
$d = 10$	.3	0.842	0.962	0.966

oracle smoother $\tilde{m}_{\gamma,\mathrm{S}}(x_\gamma)$, we define the empirical relative efficiency of $\hat{m}_{\gamma,\mathrm{SBS}}(x_\gamma)$ with respect to $\tilde{m}_{\gamma,\mathrm{S}}(x_\gamma)$ as

$$\mathrm{eff}_\gamma = \left[\frac{\sum_{t=1}^{n}\left\{\tilde{m}_{\gamma,\mathrm{S}}\left(X_{t\gamma}\right) - m_\gamma\left(X_{t\gamma}\right)\right\}^2 1_{\left(0\le X_{t\gamma}\le 1\right)}}{\sum_{t=1}^{n}\left\{\hat{m}_{\gamma,\mathrm{SBS}}\left(X_{t\gamma}\right) - m_\gamma\left(X_{t\gamma}\right)\right\}^2 1_{\left(0\le X_{t\gamma}\le 1\right)}}\right]^{1/2}. \tag{3.24}$$

Theorem 3.4 indicates that the eff_γ should be close to 1 for all $\gamma = 1, ..., d$. Figure 3.1 and 3.2 provide the kernel density estimators of the above empirical efficiencies computed over the 500 replications. Again, these plots show that the empirical distribution of eff_γ does rapidly converge to the point mass at 1 as n becomes larger. Finally, Figure 3.3 and 3.4 show typical examples of the SBS estimator with the confidence bands in (3.11) and the corresponding empirical relative efficiencies. The plots in these two figures illustrate graphically the summarized results on confidence band coverage and on the empirical relative efficiency.

Lastly, we provide the computing time of model (3.23) with dimension $d = 10$ from 100 replications on an ordinary PC with Intel(R) Quad CPU 2.4 GHz processor and 3.0 GB RAM. The average time run by R in seconds to generate one sample of size n and compute the SBS estimator and spline backfitted spline (SPBK) estimator of [71] has been reported in Table 3.2. As expected, the computing time of SBS is hundreds time faster than SPBK and this advantage widens with increasing sample size.

Table 3.2: Comparison of computing time of Model (3.23).

Method	$n = 100$	$n = 500$	$n = 1000$
SPBK	0.09	7.8	54
SBS	0.007	.064	0.32
Ratio	12.88	121.88	168.75

3.5 Appendix

Throughout this section, $a_n \gg b_n$ means $\lim_{n\to\infty} b_n/a_n = 0$, and $a_n \sim b_n$ means $\lim_{n\to\infty} b_n/a_n = c$, where c is a nonzero constant. Whenever we write ~ 1 for some quantity that depends on $0 \leq J^* \leq N^*$ or $0 \leq J \leq N+1$ it means it holds for all possible J^* or J values as $n \to \infty$.

A.1. Propositions

Recall from section 3.2 that $\|\Psi_b\|_\infty = \sup_{0\leq J\leq N+1} |\{\Psi_b\}_{J=0}^{N+1}|$. In this section, we show that the bias term $\|\Psi_b\|_\infty$ of (3.21) and the noise term Ψ_v given in (3.22) are uniformly of order $O_p\left(h^{1/2}n^{-2/5}(\log n)^{-1}\right)$.

Proposition 3.1. *Under Assumptions (A1) to (A2), and (A4) to (A5)*

$$\|\Psi_b\|_\infty = O_p\left(h^{1/2}\left(n^{-1/2}+H\right)\right) = O_p\left(h^{1/2}n^{-2/5}(\log n)^{-1}\right).$$

Lemma 3.1. *Under Assumption (A1), there exists* $g(\mathbf{x}) = c + \sum_{\gamma=1}^{d} g_\gamma(x_\gamma) \in G^*$, *such that for* $\tilde{m}$ *defined in (3.19),*

$$\left\|\tilde{m} - g + \sum_{\gamma=1}^{d}\langle 1, g_\gamma(X_\gamma)\rangle_{2,n}\right\|_{2,n}^{*} = O_p\left(n^{-1/2}+H\right).$$

Proof. By the result on page 149 of [10], there exists a constant $C_\infty > 0$ and spline functions $g_\gamma \in G^*$, such that $\|g_\gamma - m_\gamma\|_\infty \le C_\infty \left\|m'_\gamma\right\|_\infty H$, $\gamma = 1, 2, ..., d$. Thus $\|g - m\|_\infty \le \sum_{\gamma=1}^{d}\|g_\gamma - m_\gamma\|_\infty \le C_\infty \sum_{\gamma=1}^{d}\left\|m'_\gamma\right\|_\infty H$ and $\|\tilde{m} - m\|_{2,n}^* \le \|g - m\|_{2,n}^* \le C_\infty \sum_{\gamma=1}^{d}\left\|m'_\gamma\right\|_\infty H$. Noting that $\|\tilde{m} - g\|_{2,n}^* \le \|\tilde{m} - m\|_{2,n}^* + \|g - m\|_{2,n}^* \le 2C_\infty \sum_{\gamma=1}^{d}\left\|m'_\gamma\right\|_\infty H$, one has

$$\begin{aligned}\left|\langle 1, g_\gamma(\boldsymbol{X}_\gamma)\rangle_{2,n}^*\right| &\le \left|\langle 1, g_\gamma(\boldsymbol{X}_\gamma)\rangle_{2,n}^* - \langle 1, m_\gamma(\boldsymbol{X}_\gamma)\rangle_{2,n}^*\right| + \left|\langle 1, m_\gamma(\boldsymbol{X}_\gamma)\rangle_{2,n}^*\right| \\ &\le C_\infty\left\|m'_\gamma\right\|_\infty H + O_p\left(n^{-1/2}\right). \end{aligned} \tag{3.25}$$

So

$$\begin{aligned}&\left\|\tilde{m} - g + \sum_{\gamma=1}^{d}\langle 1, g_\gamma(\boldsymbol{X}_\gamma)\rangle_{2,n}\right\|_{2,n}^{*} \\ \le\ & \|\tilde{m} - g\|_{2,n}^* + \sum_{\gamma=1}^{d}\left|\langle 1, g_\gamma(\boldsymbol{X}_\gamma)\rangle_{2,n}^*\right| \\ \le\ & 3C_\infty \sum_{\gamma=1}^{d}\left\|m'_\gamma\right\|_\infty H + O_p\left(n^{-1/2}\right) = O_p\left(n^{-1/2}+H\right).\end{aligned}$$

Proof of Proposition 3.1. Clearly that $\left\|\Psi_b\right\|_\infty \leq R_1 + R_2 + R_3$, where

$$R_1 = \sup_{J=0}^{N+1} \left| n^{-1} \sum_{i=1}^{n} \sum_{\gamma=2}^{d} B_{J,1}(X_{i1}) W_i^* \left\langle 1, g_\gamma \left(X_{i\gamma}\right)\right\rangle_{2,n}^{*} \right|,$$

$$R_2 = \sup_{J=0}^{N+1} \left| n^{-1} \sum_{i=1}^{n} \sum_{\gamma=2}^{d} B_{J,1}(X_{i1}) W_i^* \left\{ g_\gamma \left(X_{i\gamma}\right) - m_\gamma \left(X_{i\gamma}\right) \right\} \right|,$$

$$\begin{aligned} R_3 &= \sup_{J=0}^{N+1} \left| n^{-1} \sum_{i=1}^{n} \sum_{\gamma=2}^{d} B_{J,1}(X_{i1}) W_i^* \right. \\ &\quad \left. \times \left\{ \tilde{m}_\gamma \left(X_{i\gamma}\right) - g_\gamma \left(X_{i\gamma}\right) + \left\langle 1, g_\gamma \left(X_{i\gamma}\right)\right\rangle_{2,n}^{*} \right\} \right|. \end{aligned}$$

According to (3.25) $R_1 = O_p\left\{h^{1/2}\left(H + n^{-1/2}\right)\right\}$. For R_2, using the result on page 149 of [10], one has $R_2 \leq C_\infty h^{1/2} H$. To deal with R_3, let $B_{J^*,\gamma}^{**}\left(\mathbf{x}_\gamma\right) = B_{J^*,\gamma}^{*}\left(\mathbf{x}_\gamma\right) - \left\langle 1, B_{J^*,\gamma}^{*}\left(\mathbf{X}_\gamma\right)\right\rangle_{2,n}^{*}$, for $1 \leq J^* \leq N^*$, $1 \leq \gamma \leq d$, then $\tilde{m}\left(\mathbf{x}\right) - g\left(\mathbf{x}\right) + \sum_{\gamma=1}^{d} \left\langle 1, g_\gamma \left(\mathbf{X}_\gamma\right)\right\rangle_{2,n}^{*} = \tilde{a}^* + \sum_{\gamma=1}^{d} \sum_{J^*=1}^{N^*} \tilde{a}_{J^*,\gamma}^{*} B_{J^*,\gamma}^{**}\left(\mathbf{x}_\gamma\right)$. Denote next $\omega_{J,J^*,\text{-}1}\left(\mathbf{X}_l\right) = \left\{\omega_{J,J^*,\gamma}\left(\mathbf{X}_l\right)\right\}_{\gamma=2}^{d}, \mu_{\omega_{J,J^*,\text{-}1}} = \left\{\mu_{\omega_{J,J^*,\gamma}}\right\}_{\gamma=2}^{d}$, where

$$\omega_{J,J^*,\gamma}\left(\mathbf{X}_l\right) = B_{J,1}(X_{l1}) B_{J^*,\gamma}^{*}\left(X_{l\gamma}\right) W_l^*, \ \mu_{\omega_{J,J^*,\gamma}} = E\omega_{J,J^*,\gamma}\left(\mathbf{X}_l\right). \quad (3.26)$$

Thus,

$$\frac{1}{n} \sum_{i=1}^{n} B_{J,1}(X_{i1}) W_i^* \left\{ \tilde{m}_{\text{-}1}\left(X_{i,\text{-}1}\right)^T - g_{\text{-}1}\left(X_{i,\text{-}1}\right)^T + E_n g_{\text{-}1}\left(X_{i,\text{-}1}\right)^T \right\} \mathbf{1}_{d-1}$$

$$= n^{-1} \sum_{i=1}^{n} B_{J,1}(X_{i1}) W_i^* \left(\sum_{J^*=1}^{N^*} \tilde{a}_{J^*,\text{-}1}^{*T} B_{J^*,\text{-}1}^{**}\left(X_{i,\text{-}1}\right) \right),$$

bounded by

$$
\begin{aligned}
&(d-1)\sup_{2\le\gamma\le d}\left(\sum_{J^*=1}^{N^*}\left|\tilde a^*_{J^*,\gamma}\right|\right.\\
&\left.\sup_{1\le J\le N}\sup_{1\le J^*\le N^*}\left|n^{-1}\sum_{i=1}^{n}B_{J,1}(X_{i1})W_i^*B^{**}_{J^*,\gamma}\left(X_{i\gamma}\right)\right|\right)\\
\le\ &(d-1)\sup_{2\le\gamma\le d}\sum_{J^*=1}^{N^*}\left|\tilde a^*_{J^*,\gamma}\right|\sup_{1\le J\le N}\sup_{1\le J^*\le N^*}\left(\left|n^{-1}\sum_{i=1}^{n}\omega_{J,J^*,\gamma}\left(\mathbf{X}_i\right)\right|\right.\\
&\left.+A_{n,1}\left|n^{-1}\sum_{i=1}^{n}B_{J,1}(X_{i1})W_i^*\right|\right),
\end{aligned}
$$

where $A_{n,1}$ is in (3.35). By Lemma 3.11,

$$
\begin{aligned}
&\sup_{1\le J\le N}\sup\nolimits_{1\le J^*\le N^*}\left|n^{-1}\sum_{i=1}^{n}\omega_{J,J^*,\gamma}\left(\mathbf{X}_i\right)\right|\\
\le\ &\sup_{1\le J\le N}\sup_{1\le J^*\le N^*}\left|\frac{1}{n}\sum_{i=1}^{n}\omega_{J,J^*,\gamma}\left(\mathbf{X}_i\right)-\mu_{\omega_{J,J^*,\gamma}}\right|\\
&+\sup_{1\le J\le N}\sup_{1\le J^*\le N^*}\left|\mu_{\omega_{J,J^*,\gamma}}\right|\\
&=O_p\left(\log n/\sqrt{n}\right)+O_p\left((Hh)^{1/2}\right)=O_p\left((Hh)^{1/2}\right).
\end{aligned}
$$

Therefore, one has

$$
\begin{aligned}
&\sup_{J=0}^{N+1}\left|\frac{1}{n}\sum_{i=1}^{n}B_{J,1}(X_{i1})W_i^*\right.\\
&\left.\left\{\tilde m_{\_1}\left(\boldsymbol{X}_{i\_1}\right)^T-g_{\_1}\left(\boldsymbol{X}_{i\_1}\right)^T+E_ng_{\_1}\left(\boldsymbol{X}_{\_1}\right)^T\right\}\mathbf{1}_{d-1}\right|\\
\le\ &(d-1)\sup_{2\le\gamma\le d}\left\{N^*\sum_{J^*=1}^{N^*}\left(\tilde a_{J^*,\gamma}\right)^2\right\}^{1/2}\left\{O_p\left((Hh)^{1/2}\right)+O_p\left(\frac{\log n}{\sqrt{n}}\right)\right\}
\end{aligned}
$$

$$= O_p\left(h^{1/2}\left\{\sum_{J^*=1}^{N^*}\left(\tilde{a}^*_{J^*,\gamma}\right)^2\right\}^{1/2}\right)$$

$$= O_p\left(h^{1/2}\left\|\tilde{m} - g + \sum_{\gamma=1}^{d}\langle 1, g_\gamma(X_\gamma)\rangle^*_{2,n}\right\|_2^*\right).$$

Thus, by lemma 3.1

$$R_3 = O_p\left(h^{1/2}\left(n^{-1/2} + H\right)\right). \tag{3.27}$$

Combining (3.25) and (3.27), one establishes Proposition 3.1.

Define an auxiliary entity

$$\tilde{\varepsilon}^*_{\text{-}1} = \sum_{J^*=1}^{N^*} \tilde{a}^T_{J^*,\text{-}1} B^*_{J^*,\text{-}1}(x_{\text{-}1}), \tag{3.28}$$

where $\tilde{a}_{J^*,\text{-}1} = \left\{\tilde{a}_{J^*,\gamma}\right\}_{\gamma=2}^{d}$ and $\tilde{a}_{J^*,\gamma}$ is given in (3.21). Definitions (3.17) imply that $\tilde{\varepsilon}_{\text{-}1}(x_{\text{-}1})$ defined in (3.19) is the empirical centering of $\tilde{\varepsilon}^*_{\text{-}1}(\mathbf{x}_{\text{-}1})$, i.e.

$$\tilde{\varepsilon}_{\text{-}1}(\mathbf{x}_{\text{-}1}) \equiv \tilde{\varepsilon}^*_{\text{-}1}(x_{\text{-}1}) - n^{-1}\sum_{i=1}^{n}\tilde{\varepsilon}^*_{\text{-}1}(\boldsymbol{X}_{i,\text{-}1}) W^*_i. \tag{3.29}$$

Proposition 3.2. *Under Assumptions (A2) to (A5), one has*

$$\|\boldsymbol{\Psi}_v\|_\infty = O_p\left(Hh^{1/2}\right) = O_p\left(h^{1/2}n^{-2/5}(\log n)^{-1}\right).$$

According to (3.29), we can write $\Psi_v = \Psi_v^{(2)} - \Psi_v^{(1)}$, in which

$$\left\{\Psi_v^{(1)}\right\}_{J=0}^{N+1} = \left\{n^{-2}\sum_{i,i'=1}^{n} B_{J,1}(X_{i1})W_{i1}W^*_{i'}\tilde{\varepsilon}^*_{\text{-}1}\left(\boldsymbol{X}_{i',\text{-}1}\right)^T \mathbf{1}_{d-1}\right\}_{J=0}^{N+1}, \tag{3.30}$$

$$\left\{\Psi_v^{(2)}\right\}_{J=0}^{N+1} = \frac{\mathbf{B}^T}{n}\mathbf{W}^*\tilde{\varepsilon}^*_{\text{-}1}\left(\mathbf{X}_{\text{-}1}\right)^T \mathbf{1}_{d-1}. \tag{3.31}$$

where $\tilde{\varepsilon}^*_{\text{-}1}\left(\mathbf{X}_{\text{-}1}\right)$ is given in (3.28). By (3.26), (3.21) and (3.28), we have

$$\left\|\mathbf{\Psi}_v^{(2)}\right\|_\infty = \sup_{0\le J\le N+1}\left|n^{-1}\sum_{l=1}^{n}\sum_{J^*=1}^{N^*}\tilde{a}^T_{J^*,\text{-}1}\omega_{J,J^*,\text{-}1}\left(\mathbf{X}_{\text{-}1}\right)\right|. \tag{3.32}$$

Proposition 3.2 follows from Lemmas 3.2 and 3.3.

Lemma 3.2. *Under Assumptions (A2) to (A5),* $\Psi_v^{(1)}$ *in (3.30) satisfies*

$$\left\|\mathbf{\Psi}_v^{(1)}\right\|_\infty = O_p\left\{h^{1/2}N^*\left(\log n\right)^2/n\right\}.$$

Proof. Based on (3.28), $\left\|n^{-1}\sum_{i=1}^n \tilde{\varepsilon}^*_{\text{-}1}\left(\mathbf{X}_{i,\text{-}1}\right)^T\mathbf{1}_{d-1}W_i^*\right\|_\infty$ is bounded by

$$(d-1)\sup_{2\le\gamma\le d}\left\{\left(\sum_{J^*=1}^{N^*}\left|\tilde{a}^*_{J^*,\gamma}\right|\right)\cdot\sup_{1\le J^*\le N^*}\left|\frac{1}{n}\sum_{i=1}^n B^*_{J^*,\gamma}\left(X_{i\gamma}\right)W_i^*\right|\right\}.$$

Lemma 3.13 implies $\sum_{J^*=1}^{N^*}\left|\tilde{a}^*_{J^*,\gamma}\right| \le \left\{N^*\left(\tilde{\mathbf{a}}^{*T}\tilde{\mathbf{a}}^*\right)\right\}^{1/2} = O_p\left(N^*n^{-1/2}\log n\right)$. Further, by (3.35),

$$\begin{aligned}&\sup_{2\le\gamma\le d}\sup_{1\le J^*\le N^*}\left|n^{-1}\sum_{i=1}^n B^*_{J^*,\gamma}\left(X_{i\gamma}\right)W_i^*\right| \\ \le\ & A_{n,1} = O_p\left(n^{-1/2}\log n\right),\end{aligned}$$

so

$$\left\|\frac{1}{n}\sum_{i=1}^n \tilde{\varepsilon}^*_{\text{-}1}\left(X_{i,\text{-}1}\right)^T\mathbf{1}_{d-1}W_i^*\right\|_\infty = O_p\left\{N^*(\log n)^2/n\right\}. \tag{3.33}$$

By

$$
\begin{aligned}
& \sup_{0\le J\le N+1}\left|n^{-1}\sum_{i=1}^{n} B_{J,1}(X_{i1})W_i^*\right| \\
= & \sup_{0\le J\le N+1}\left(\left\langle 1, B_{J,1}\right\rangle_{2,n} - \left\langle 1, B_{J,1}\right\rangle_{2}\right) + \sup_{0\le J\le N+1}\left\langle 1, B_{J,1}\right\rangle_{2} \\
= & O_P(\log n/\sqrt{n}) + O_P\left(h^{1/2}\right) = O_P\left(h^{1/2}\right).
\end{aligned}
$$

Thus with (3.33) the lemma follows immediately.

Lemma 3.3. *Under Assumptions (A2) to (A5), we have* $\left\|\Psi_v^{(2)}\right\|_\infty = O_p\left(Hh^{1/2}\right)$.

Lemma 3.3 follows from Lemmas 3.14 and 3.15.

A.2. Preliminaries

We first give the Bernstein's inequality for geometrically γ-mixing sequence, which is used often in many of our proofs.

Lemma 3.4. *[Theorem 1.4, page 31 of Bosq [3]] Let* $\{\xi_t, t\in\mathbb{Z}\}$ *be a zero mean real valued* α*-mixing process,* $S_n = \sum_{i=1}^n \xi_i$. *Suppose that there exists* $c>0$ *such that for* $i=1,...,n, k=3,4,...,$ $E\left|\xi_i\right|^k \le c^{k-2}k!E\xi_i^2 < +\infty$, *then for each* $n>1$, *integer* $q\in[1,n/2]$, *each* $\varepsilon>0$ *and* $k\ge 3$

$$
P(|S_n|\ge n\varepsilon) \le a_1 \exp\left(-\frac{q\varepsilon^2}{25m_2^2+5c\varepsilon}\right) + a_2(k)\,\alpha\left(\left[\frac{n}{q+1}\right]\right)^{2k/(2k+1)},
$$

where $\alpha(\cdot)$ *is the* α*-mixing coefficient in (3.10) and* $a_1 = 2\frac{n}{q} + 2\left(1+\frac{\varepsilon^2}{25m_2^2+5c\varepsilon}\right)$,
$a_2(k) = 11n\left(1+\frac{5m_k^{2k/(2k+1)}}{\varepsilon}\right)$, *with* $m_r = \max_{1\le i\le n}\left\|\xi_i\right\|_r$, $r\ge 2$.

Lemma 3.5. *Under Assumptions (A4) and (A5), one has:*

(i) $\left\|b_{J^*,\gamma}^*\right\|_2^2 \sim H$, *where* $b_{J^*,\gamma}^*$ *is given in (3.14).*

(ii) for any $\gamma = 1, 2, ..., d$,

$$E\left\{B^*_{J^*,\gamma}\left(X_{i\gamma}\right) B^*_{J^{*\prime},\gamma}\left(X_{i\gamma}\right) W^*_i\right\} \sim 1,$$

for $\left|J^{*\prime} - J^*\right| \leq 1$, *and*

$$E\left\{B_{J,\gamma}\left(X_{i\gamma}\right) B_{J',\gamma}\left(X_{i\gamma}\right) W_{i\gamma}\right\} \sim 1,$$

for $\left|J' - J\right| \leq 1$. *In addition,*

$$E\left|B^*_{J^*,\gamma}\left(X_{i\gamma}\right) B^*_{J^{*\prime},\gamma}\left(X_{i\gamma}\right) W^*_i\right|^k \sim H^{1-k},$$
$$E\left|B_{J,\gamma}\left(X_{i\gamma}\right) B_{J',\gamma}\left(X_{i\gamma}\right) W_{i\gamma}\right|^k \sim h^{1-k},$$

for $k \geq 1$, *where* $B^*_{J^*,\gamma}$ *and* $B_{J,\gamma}$ *are defined in (3.13) and (3.12).*

Lemma 3.6. *Under Assumptions (A4) and (A5), there exist constants* $C_0 > c_0 > 0$ *such that for any* $\mathbf{a}^* = \left(a^*_0, a^*_{1,1}, ..., a^*_{N^*,1}, a^*_{1,2}, ..., a^*_{N^*,2}, ...a^*_{1,d}, ..., a^*_{N^*,d}\right)$,

$$c_0\left(a^{*2}_0 + \sum_{J^*,\gamma} a^{*2}_{J^*,\gamma}\right) \leq \left\|a^*_0 + \sum_{J^*,\gamma} a^*_{J^*,\gamma} B^*_{J^*,\gamma}\right\|^{*2}_2 \leq C_0\left(a^{*2}_0 + \sum_{J^*,\gamma} a^{*2}_{J^*,\gamma}\right). \tag{3.34}$$

Lemma 3.7. *Under Assumptions (A2), (A4) and (A6), one has*

$$\begin{aligned} A_{n,1} &= \sup_{1 \leq J^* \leq N^*,\gamma}\left|\left\langle 1, B^*_{J^*,\gamma}\right\rangle^*_{2,n} - \left\langle 1, B^*_{J^*,\gamma}\right\rangle^*_2\right| \\ &= O_P\left(n^{-1/2}\log n\right), \end{aligned} \tag{3.35}$$

$$A_{n,2} \tag{3.36}$$
$$= \sup_{1\le J^*,J'^*\le N^*,\gamma}\left|\left\langle B^*_{J^*,\gamma},B^*_{J'^*,\gamma}\right\rangle^*_{2,n}-\left\langle B^*_{J^*,\gamma},B^*_{J'^*,\gamma}\right\rangle^*_2\right|$$
$$= O_p\left(n^{-1/2}H^{-1/2}\log n\right),$$

$$A_{n,3} \tag{3.37}$$
$$= \sup_{1\le J^*,J'^*\le N^*,\gamma\neq\gamma'}\left|\left\langle B^*_{J^*,\gamma},B^*_{J'^*,\gamma'}\right\rangle^*_{2,n}-\left\langle B^*_{J^*,\gamma},B^*_{J'^*,\gamma'}\right\rangle^*_2\right|$$
$$= O_p\left(n^{-1/2}\log n\right).$$

Lemma 3.8. *Under Assumptions (A2), (A4) and (A6), one has*

$$A_n = \sup_{g_1,g_2\in G^*}\frac{\left|\langle g_1,g_2\rangle^*_{2,n}-\langle g_1,g_2\rangle^*_2\right|}{\|g_1\|^*_2\,\|g_2\|^*_2} = O_p\left(\frac{\log n}{n^{1/2}H^{1/2}}\right) = o_p(1). \tag{3.38}$$

Denote next by $\mathbf{V}$ as the theoretical inner product of the B spline basis $\left\{1, B^*_{J^*,\gamma}(x_\gamma), J^*=1,...,N^*,\gamma=1,...,d\right\}$, i.e.

$$\mathbf{V} = \begin{pmatrix} 1 & \mathbf{0}^T_{dN^*} \\ \mathbf{0}_{dN^*} & \left\langle B^*_{J^*,\gamma},B^*_{J^{*\prime},\gamma'}\right\rangle^*_2 \end{pmatrix}_{\substack{1\le\gamma,\gamma'\le d,\\ 1\le J,J'\le N^*}}. \tag{3.39}$$

Let $\mathbf{S}$ be the inverse matrix of $\mathbf{V}$, i.e.,

$$\mathbf{S} = \mathbf{V}^{-1} = \begin{pmatrix} 1 & \mathbf{0}^T_N & \mathbf{0}^T_N \\ \mathbf{0}_N & \mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{0}_N & \mathbf{V}_{21} & \mathbf{V}_{22} \end{pmatrix}^{-1} = \begin{pmatrix} 1 & \mathbf{0}^T_N & \mathbf{0}^T_N \\ \mathbf{0}_N & \mathbf{S}_{11} & \mathbf{S}_{12} \\ \mathbf{0}_N & \mathbf{S}_{21} & \mathbf{S}_{22} \end{pmatrix}. \tag{3.40}$$

Lemma 3.9. *Under Assumptions (A4) and (A5), for* $\mathbf{V}$, $\mathbf{S}$ *defined in (3.39), (3.40), there exist constants* $C_V > c_V > 0$ *and* $C_S > c_S > 0$ *such that* $c_V\mathbf{I}_{dN^*+1} \le \mathbf{V} \le C_V\mathbf{I}_{dN^*+1}$, $c_S\mathbf{I}_{dN^*+1} \le \mathbf{S} \le C_S\mathbf{I}_{dN^*+1}$.

We refer the proofs of Lemmas 3.5 to 3.9 to Lemmas A.2, A.4, A.7, A.8 and A.9 in [71].

Lemma 3.10. *Under Assumptions (A2) and (A3), there exist constants* $c(f)$, $C(f) > 0$ *independent of* n*, such that as* $n \to \infty$*, with probability approaching* 1,

$$c(f)|\varsigma| \leq \left|\left(\frac{1}{n}\mathbf{B}^T\mathbf{W}\mathbf{B}\right)^{-1}\varsigma\right| \leq C(f)|\varsigma|, \tag{3.41}$$

$$c(f)\|\varsigma\|^2 \leq \varsigma^T\left(\frac{1}{n}\mathbf{B}^T\mathbf{W}\mathbf{B}\right)^{-1}\varsigma \leq C(f)\|\varsigma\|^2, \forall\varsigma \in R^{N+2}. \tag{3.42}$$

The lemma and its proof is based on Lemma B.2 of Wang and Yang [70].

Lemma 3.11. *Under Assumptions (A4) to (A5), for* $\mu_{\omega_{J,J^*,\text{-}1}}$ *given in (3.26)*

$$\sup_{0\leq J\leq N+1}\sup_{1\leq J^*\leq N^*}\left|\mu_{\omega_{J,J^*,\text{-}1}}\right| = O\left\{(hH)^{1/2}\right\}.$$

Proof. For $\gamma = 2,...,d$, $J = 0,...,N+1$, $J^* = 1,...,N^*$, by the boundedness of the density f,

$$\begin{aligned}
&\left|E\left\{B_{J,1}(X_{l1})W_l^*B_{J^*,\gamma}^*\left(X_{l\gamma}\right)\right\}\right| \\
\leq\ & \int_0^1\cdots\int_0^1\left|B_{J,1}(u_1)B_{J^*,\gamma}^*(u_\gamma)\right|f(u_1,...,u_d)\,du_1...du_d \\
\leq\ & C_f\int_0^1\cdots\int_0^1\left|B_{J,1}(u_1)B_{J^*,\gamma}^*(u_\gamma)\right|du_1...du_d \\
=\ & C_f\left(\left\|b_{J,1}\right\|_2\left\|b_{J^*,\gamma}^*\right\|_2\right)^{-1}\int_0^1\int_0^1\left|b_{J,1}(ul)\left(b_{J^*,\gamma}^*(u_\gamma)\right)\right|du_1du_\gamma \\
=\ & \left(\left\|b_{J,1}\right\|_2\left\|b_{J^*,\gamma}^*\right\|_2\right)^{-1}\left\{\int_0^1\int_0^1 b_{J,1}(u_1)I_{J^*+1,\gamma}(u_\gamma)\,du_1du_\gamma+\right. \\
&\left.\left(\left\|b_{J,1}\right\|_2\left\|b_{J^*,\gamma}^*\right\|_2\right)^{-1}\frac{c_{J^*+1,\gamma}}{c_{J^*,\gamma}}\int_0^1\int_0^1 b_{J,1}(u_1)I_{J^*,\gamma}(u_\gamma)\,du_1du_\gamma\right\},
\end{aligned}$$

where $c_{J^*,\gamma} = \left\langle 1, I_{J^*,\gamma} \right\rangle_2$.

$$\sup_{0 \le J \le N+1} \sup_{1 \le J^* \le N^*} \int \int b_{J,1}(u_1) I_{J^*,\gamma}\left(u_\gamma\right) du_1 du_\gamma = O\left\{hH\right\},$$

and the proof of the lemma is then completed by (i) of Lemma 3.5.

Lemma 3.12. *Under Assumptions (A2), (A4) and (A5), one has*

$$\sup_{0 \le J \le N+1} \sup_{1 \le J^* \le N^*} \left\| n^{-1} \sum_{l=1}^{n} \left\{ \omega_{J,J^*,1}\left(\mathbf{X}_l\right) - \mu_{\omega_{J,J^*,1}} \right\} \right\|_\infty = O_p\left(\log n / \sqrt{n}\right), \tag{3.43}$$

$$\sup_{0 \le J \le N+1} \sup_{1 \le J^* \le N^*} \left\| n^{-1} \sum_{l=1}^{n} \omega_{J,J^*,1}\left(\mathbf{X}_l\right) \right\|_\infty = O_p\left((hH)^{1/2}\right), \tag{3.44}$$

where $\omega_{J,J^*,1}\left(\mathbf{X}_l\right)$ *and* $\mu_{\omega_{J,J^*,1}}$ *are given in (3.26).*

Proof. For simplicity, denote $\omega^*_{J,J^*,\gamma}\left(\mathbf{X}_l\right) = \omega_{J,J^*,\gamma}\left(\mathbf{X}_l\right) - \mu_{\omega_{J,J^*,\gamma}}$. Then $E\left\{\omega^*_{J,J^*,\gamma}\left(\mathbf{X}_l\right)\right\}^2 = E\omega^2_{J,J^*,\gamma}\left(\mathbf{X}_l\right) - \mu^2_{\omega_{J,J^*,\gamma}}$, while

$$\begin{aligned} E\omega^2_{J,J^*,\gamma}\left(\mathbf{X}_l\right) &= E\left(B_{J,1}(X_{l1}) W^*_l B^*_{J^*,\gamma}\left(X_{l\gamma}\right)\right)^2 \\ &= \left(\left\|b_{J,1}\right\|_2 \left\|b^*_{J^*,\gamma}\right\|_2\right)^{-2} \int_0^1 \cdots \int_0^1 \left(b_{J,1}(u1)\right)^2 \left(b^*_{J^*,\gamma}\left(u_\gamma\right)\right)^2 \\ &\quad f\left(u_1, ..., u_d\right) du_1 ... du_d, \end{aligned}$$

$E\omega^2_{J,J^*,\gamma}\left(\mathbf{X}_l\right) \sim 1$ and $E\omega^2_{J,J^*,\gamma}\left(\mathbf{X}_l\right) \gg \mu^2_{\omega_{J,J^*,\gamma}}$. Hence $E\left\{\omega^*_{J,J^*,\gamma}\left(\mathbf{X}_l\right)\right\}^2 = E\omega^2_{J,J^*,\gamma}\left(\mathbf{X}_l\right) - \mu^2_{\omega_{J,J^*,\gamma}} \ge c^*$ for n sufficiently large and some positive constant c^*. When $r \ge 3$, the r-th moment $E\left|\omega_{J,J^*,\gamma}\left(\mathbf{X}_l\right)\right|^r$ is

$$\frac{1}{\left(\left\|b_{J,1}\right\|_2 \left\|b^*_{J^*,\gamma}\right\|_2\right)^r} \int_0^1 \cdots \int_0^1 b_{J,1}(u_1)^r \left|b^*_{J^*,\gamma}\left(u_\gamma\right)\right|^r f\left(u_1, ..., u_d\right) du_1 ... du_d.$$

It is clear that $E\left|B_{J,1}(X_{l1})W_l^* B_{J^*,\gamma}^*\left(X_{l\gamma}\right)\right|^r \sim h^{(1-r/2)}H^{1-r/2}$. According to Lemma 3.11, one has $\left|E\omega_{J,J^*,\gamma}\left(\mathbf{X}_l\right)\right|^r = \left|EB_{J,1}(X_{l1})W_l^* B_{J^*,\gamma}^*\left(X_{l\gamma}\right)\right|^r \sim (hH)^{r/2}$, thus $E\left|\omega_{J,J^*,\gamma}\left(\mathbf{X}_l\right)\right|^r \gg \left|\mu_{\omega_{J,J^*,\gamma}}\right|^r$. In addition, for any J and J^*,

$$E\left|\omega_{J,J^*,\gamma}^*\left(\mathbf{X}_l\right)\right|^r \leq \left\{\frac{c}{(hH)^{1/2}}\right\}^{(r-2)} r!E\left|\omega_{J,J^*,\gamma}^*\left(\mathbf{X}_l\right)\right|^2,$$

so there exists $c_* = ch^{-1/2}H^{-1/2}$ such that

$$E\left|\omega_{J,J^*,\gamma}^*\left(\mathbf{X}_l\right)\right|^r \leq c_*^{r-2} r!E\left|\omega_{J,J^*,\gamma}^*\left(\mathbf{X}_l\right)\right|^2,$$

which implies that $\left\{\omega_{J,J^*,\gamma}^*\left(\mathbf{X}_l\right)\right\}_{l=1}^n$ satisfies the Cramér's condition. By the Bernstein's inequality, for $r = 3$

$$\begin{aligned} & P\left\{\left|\frac{1}{n}\sum_{l=1}^n \omega_{J,J^*,\gamma}^*\left(\mathbf{X}_l\right)\right| \geq \rho_n\right\} \\ \leq \; & a_1 \exp\left(-\frac{q\rho_n^2}{25m_2^2 + 5c_*\rho_n}\right) + a_2(3)\,\alpha\left(\left[\frac{n}{q+1}\right]\right)^{6/7} \end{aligned}$$

with $m_2^2 \sim h^{-1}$, $m_3 = \max_{1\leq i\leq n}\left\|\omega_{J,J^*,\gamma}^*\left(\mathbf{X}_l\right)\right\|_3 \leq \left\{C_0\left(2h^{-1}\right)^2\right\}^{1/3}$ and

$$\rho_n = \rho\frac{\log n}{\sqrt{nh}}, a_1 = 2\frac{n}{q} + 2\left(1 + \frac{\rho_n^2}{25m_2^2 + 5c_*\rho_n}\right), a_2(3) = 11n\left(1 + \frac{5m_3^{6/7}}{\rho_n}\right).$$

Since $5c_*\rho_n = o(1)$, by taking q such that $\left[\frac{n}{q+1}\right] \geqslant c_0\log n$, $q \geqslant c_1 n/\log n$ for constants c_0, c_1, one has $a_1 = O(n/q) = O(\log n)$, $a_2(3) = o\left(n^2\right)$. Assumption (A2) yields that

$$\alpha\left(\left[\frac{n}{q+1}\right]\right)^{6/7} \leq Cn^{-6\lambda_0 c_0/7}.$$

Thus, for n large enough,

$$P\left\{\frac{1}{n}\left|\sum_{l=1}^{n}\omega^*_{J,J^*,\gamma}(\mathbf{X}_l)\right| > \frac{\rho\log n}{\sqrt{nh}}\right\} \le cn^{-c_2\rho^2}\log n + Cn^{2-6\lambda_0 c_0/7}. \quad (3.45)$$

By (3.45), there exists large enough value $\rho > 0$ such that for any J^*,

$$P\left\{\frac{1}{n}\left|\sum_{l=1}^{n}\omega^*_{J,J^*,\gamma}(\mathbf{X}_l)\right| > \rho(nh)^{-1/2}\log n\right\} \le n^{-10}, 1 \le J^* \le N^*,$$

which implies that

$$\begin{aligned}
&\sum_{n=1}^{\infty} P\left\{\sup_{0\le J\le N+1}\sup_{1\le J^*\le N^*}\left|n^{-1}\sum_{l=1}^{n}\omega^*_{J,J^*,\gamma}(\mathbf{X}_l)\right| \ge \rho\frac{\log n}{\sqrt{nh}}\right\} \\
\le\ & \sum_{n=1}^{\infty}\sum_{J=0}^{N+1}\sum_{J^*=1}^{N^*} P\left\{\left|n^{-1}\sum_{l=1}^{n}\omega^*_{J,J^*,\gamma}(\mathbf{X}_l)\right| \ge \rho\frac{\log n}{\sqrt{nh}}\right\} \\
\le\ & \sum_{n=1}^{\infty} N(N^*)n^{-10} < \infty.
\end{aligned}$$

Thus, Borel-Cantelli Lemma entails that

$$\sup_{0\le J\le N+1}\sup_{1\le J^*\le N^*}\left|n^{-1}\sum_{l=1}^{n}\omega^*_{J,J^*,\gamma}(\mathbf{X}_l)\right| = O_p\left(\log n/\sqrt{nh}\right). \quad (3.46)$$

Then,

$$\sup_{0\le J\le N+1}\sup_{1\le J^*\le N^*}\left\|n^{-1}\sum_{l=1}^{n}\omega^*_{J,J^*,-1}(\mathbf{X}_l)\right\|_{\infty} = O_p\left(\log n/\sqrt{nh}\right).$$

As a result of Lemma 3.11 and (3.43), (3.44) holds.

The next lemma provides the size of $\tilde{\mathbf{a}}^{*T}\tilde{\mathbf{a}}^*$, where $\tilde{\mathbf{a}}^*$ is defined by (3.21).

Lemma 3.13. *Under Assumptions (A2) to (A5),* $\tilde{\mathbf{a}}^*$ *satisfies*

$$\tilde{\mathbf{a}}^{*T}\tilde{\mathbf{a}}^* = \tilde{a}_0^{*2} + \sum_{J^*=1}^{N^*}\sum_{\gamma=1}^{d} \tilde{a}_{J,\gamma}^{*2} = O_P\left\{N^*(\log n)^2/n\right\}. \tag{3.47}$$

Proof. According to (3.20) and (3.21), $\tilde{\mathbf{a}}^{*T}\mathbf{B}^{*T}\mathbf{W}^*\mathbf{B}^*\tilde{\mathbf{a}}^* = \tilde{\mathbf{a}}^{*T}\left(\mathbf{B}^{*T}\mathbf{W}^*\mathbf{E}\right)$. Thus

$$\|\mathbf{W}^*\mathbf{B}^*\tilde{\mathbf{a}}^*\|_{2,n}^{*2} = \tilde{\mathbf{a}}^{*T}\begin{pmatrix} 1 & \\ & \left\langle B_{J^*,\gamma}^*, B_{J^{*\prime},\gamma'}\right\rangle_{2,n}^* \end{pmatrix}\tilde{\mathbf{a}}^* = \tilde{\mathbf{a}}^{*T}\left(n^{-1}\mathbf{B}^{*T}\mathbf{W}^*\mathbf{E}\right). \tag{3.48}$$

By (3.38), $\|\mathbf{B}^*\tilde{\mathbf{a}}^*\|_{2,n}^{*2}$ is bounded below in probability by $(1-A_n)\|\mathbf{B}^*\tilde{\mathbf{a}}^*\|_2^{*2}$. According to (3.34), one has

$$\|\mathbf{W}^*\mathbf{B}^*\tilde{\mathbf{a}}^*\|_2^{*2} = \left\|a_0^* + \sum_{J^*,\gamma} a_{J^*,\gamma}^* B_{J^*,\gamma}^*\right\|_2^{*2} \geq c_0\left(\tilde{a}_0^{*2} + \sum_{J^*,\gamma}\tilde{a}_{J^*,\gamma}^{*2}\right). \tag{3.49}$$

Meanwhile one can show that $\tilde{\mathbf{a}}^{*T}\left(n^{-1}\mathbf{B}^{*T}\mathbf{W}^*\mathbf{E}\right)$ is bounded above by

$$\sqrt{\tilde{a}_0^{*2} + \sum_{J^*,\gamma}\tilde{a}_{J,\gamma}^{*2}}\left[\left\{\frac{1}{n}\sum_{i=1}^n \varepsilon_i\right\}^2 + \sum_{J^*,\gamma}\left\{\frac{1}{n}\sum_{i=1}^n B_{J^*,\gamma}^*\left(X_{i\gamma}\right)W_i^*\varepsilon_i\right\}^2\right]^{1/2}. \tag{3.50}$$

Combining (3.48), (3.49) and (3.50), the squared norm $\tilde{\mathbf{a}}^{*T}\tilde{\mathbf{a}}^*$ is bounded by

$$c_0^{-2}(1-A_n)^{-2}\left[\left\{\frac{1}{n}\sum_{i=1}^n \varepsilon_i\right\}^2 + \sum_{J^*,\gamma}\left\{\frac{1}{n}\sum_{i=1}^n B_{J^*,\gamma}^*\left(X_{i\gamma}\right)W_i^*\varepsilon_i\right\}^2\right].$$

Truncating ε as in Lemma 3.15, Bernstein inequality entails that

$\left|n^{-1}\sum_{i=1}^{n}\varepsilon_i\right| + \max_{1\leq J^*\leq N^*,\gamma=1,\ldots,d}\left|n^{-1}\sum_{i=1}^{n}B_{J^*,\gamma}\left(X_{i\gamma}\right)W_i^*\varepsilon_i\right| =$ $O_p\left(\log n/\sqrt{n}\right)$. Thus (3.47) holds since A_n is of order $o_p(1)$ by lemma 3.8.

A.3. Proof of Lemma 3.3 We denote

$$\mathbf{V}^* = \begin{pmatrix} 0 & \mathbf{0}_{dN^*}^T \\ \mathbf{0}_{dN^*} & \left\langle B_{J^*,\gamma}^*, B_{J^{*\prime},\gamma'}^*\right\rangle_{2,n}^* - \left\langle B_{J^*,\gamma}^*, B_{J^{*\prime},\gamma'}^*\right\rangle_2^* \end{pmatrix}_{\substack{1\leq\gamma,\gamma'\leq d,\\ 1\leq J^*,J^{*\prime}\leq N^*}},$$

then $\tilde{\mathbf{a}}^*$ in (3.21) can be rewritten as

$$\tilde{\mathbf{a}}^* = \left(\frac{1}{n}\mathbf{B}^{*T}\mathbf{W}^*\mathbf{B}^*\right)^{-1}\left(\frac{1}{n}\mathbf{B}^{*T}\mathbf{W}^*\mathbf{E}\right) = \left(\mathbf{V}+\mathbf{V}^*\right)^{-1}\left(\frac{1}{n}\mathbf{B}^{*T}\mathbf{W}^*\mathbf{E}\right). \quad (3.51)$$

Now define $\hat{\mathbf{a}} = \left\{\hat{a}_0, \hat{a}_{1,1}, \ldots, \hat{a}_{N,1}, \hat{a}_{1,2}, \ldots, \hat{a}_{N,2}\right\}^T$ as

$$\hat{\mathbf{a}} = \mathbf{V}^{-1}\left(n^{-1}\mathbf{B}^{*T}\mathbf{W}^*\mathbf{E}\right) = \mathbf{S}\left(n^{-1}\mathbf{B}^{*T}\mathbf{W}^*\mathbf{E}\right), \quad (3.52)$$

and define a theoretical version of $\Psi_v^{(2)}$ in (3.32) as

$$\hat{\Psi}_v^{(2)} = n^{-1}\sum_{i=1}^{n}\sum_{J^*=1}^{N^*}\hat{a}_{J^*,-1}^{*T}\omega_{J,J^*,-1}\left(\mathbf{X}_i\right). \quad (3.53)$$

Lemma 3.14. *Under Assumptions (A2) to (A5),*

$$\left\|\Psi_v^{(2)} - \hat{\Psi}_v^{(2)}\right\|_\infty = O_p\left\{h^{1/2}\left(\log n\right)^2/nH\right\}.$$

Proof. By (3.51) and (3.52), one has $\mathbf{V}\ \hat{\mathbf{a}}^* = \left(\mathbf{V}+\mathbf{V}^*\right)\ \tilde{\mathbf{a}}^*$, which implies that $\mathbf{V}^*\tilde{\mathbf{a}}^* = \mathbf{V}\left(\hat{\mathbf{a}}^* - \tilde{\mathbf{a}}^*\right)$. Using (3.36) and (3.37), one obtains that

$$\left\|\mathbf{V}\left(\hat{\mathbf{a}}^*-\tilde{\mathbf{a}}^*\right)\right\|_2^* = \left\|\mathbf{V}^*\tilde{\mathbf{a}}^*\right\|_2^* \leq O_p\left(n^{-1/2}H^{-1}\log n\right)\left\|\tilde{\mathbf{a}}^*\right\|_2^*.$$

According to Lemma 3.13, $\|\tilde{\mathbf{a}}^*\|_2^* = O_P\left(n^{-1/2}N^{*1/2}\log n\right)$, so one has

$$\|\mathbf{V}\left(\hat{\mathbf{a}}^* - \tilde{\mathbf{a}}^*\right)\|_2^* \leq O_P\left\{(\log n)^2 n^{-1} N^{*3/2}\right\}.$$

By Lemma 3.9, $\|\left(\hat{\mathbf{a}}^* - \tilde{\mathbf{a}}^*\right)\|_2^* = O_P\left\{(\log n)^2 n^{-1} N^{*3/2}\right\}$. Lemma 3.13 implies

$$\|\hat{\mathbf{a}}^*\|_2^* \leq \|\left(\hat{\mathbf{a}}^* - \tilde{\mathbf{a}}^*\right)\|_2^* + \|\tilde{\mathbf{a}}^*\|_2^* = O_P\left(\log n\sqrt{N^*/n}\right). \tag{3.54}$$

Additionally,

$$\left\|\Psi_v^{(2)} - \hat{\Psi}_v^{(2)}\right\|_\infty = \sup_{0\leq J\leq N+1}\left|\sum_{J^*=1}^{N^*}\left(\tilde{a}_{J^*,\text{-}1}^* - \hat{a}_{J^*,\text{-}1}^*\right)\frac{1}{n}\sum_{l=1}^{n}\omega_{J,J^*,\text{-}1}\left(\mathbf{X}_l\right)\right|.$$

So

$$\begin{aligned}\left\|\Psi_v^{(2)} - \hat{\Psi}_v^{(2)}\right\|_\infty &\leq \sqrt{N^*}O_P\left\{\frac{(\log n)^2}{nH}\right\}O_P\left((hH)^{1/2}\right) \\ &= O_P\left\{\frac{h^{1/2}(\log n)^2}{nH}\right\}.\end{aligned}$$

Lemma 3.15. *Under Assumptions (A2) to (A5), for $\hat{\Psi}_v^{(2)}$ in (3.53), one has*

$$\begin{aligned}\left\|\hat{\Psi}_v^{(2)}\right\|_\infty &= \sup_{0\leq J\leq N+1}\left|n^{-1}\sum_{i=1}^{n}\left(B_{J,1}\left(X_{i1}\right)\sum_{J^*=1}^{N^*}\hat{a}_{J^*,\text{-}1}^{*T}B_{J^*,\text{-}1}^*\left(X_{i,\text{-}1}\right)W_i^*\right)\right| \\ &= O_P\left(h^{1/2}H\right).\end{aligned}$$

Proof. Note that $\left\|\hat{\Psi}_v^{(2)}\right\|_\infty$ is bounded by $Q_1 + Q_2$, where

$$Q_1 = \sup_{0\leq J\leq N+1}\left|\sum_{J^*=1}^{N^*}\hat{a}_{J^*,\text{-}1}^{*T}\mu_{\omega_{J,J^*,\text{-}1}}\right|,$$

$$Q_2 = \sup_{0\le J\le N+1}\left|\sum_{J^*=1}^{N^*} \hat{a}^{*T}_{J^*,-1} n^{-1}\sum_{i=1}^{n}\left\{\omega_{J,J^*,-1}(\mathbf{X}_i) - \mu_{\omega_{J,J^*,-1}}\right\}\right|. \quad (3.55)$$

By Cauchy-Schwartz inequality, (3.54), Lemma 3.12, and Assumptions (A5),

$$Q_2 = O_p\left(\log n\sqrt{N^*/n}\right)\sqrt{N^*}O_p\left(\frac{\log n}{\sqrt{n}}\right) = O_p\left\{\frac{(\log n)^3}{\sqrt{n}}\right\}. \quad (3.56)$$

Define next

$$F_{1,\gamma} = \sup_{0\le J\le N+1}\left|n^{-1}\sum_{1\le i\le n}\sum_{1\le J^*,J^{*\prime}\le N^*}\right.$$

$$\left.\mu_{\omega_{J,J^*,\gamma}} s_{J^*+N^*,J^{*\prime}+1} B_{J^{*\prime},1}(X_{i1})W_{i1}\varepsilon_i\right|,$$

$$F_{2,\gamma} = \sup_{0\le J\le N+1}\left|n^{-1}\sum_{1\le i\le n}\sum_{1\le J^*,J^{*\prime}\le N^*}\right.$$

$$\left.\mu_{\omega_{J,J^*,\gamma}} s_{J^*+N^*,J^{*\prime}+N^*} B_{J^{*\prime},\gamma}(X_{i\gamma})W_{i\gamma}\varepsilon_i\right|,$$

then it is clear that

$$Q_1 \le (d-1)\left(\sup_{2\le\gamma\le d} F_{1,\gamma} + \sup_{2\le\gamma\le d} F_{2,\gamma}\right).$$

Next we will show that $F_{1,\gamma} = O_p\left(h^{1/2}H\right)$. Let $D_n = n^{\theta_0}\left(\frac{1}{2+\delta} < \theta_0 < \frac{2}{5}\right)$, where δ is the same as in Assumption (A3). Define

$$\varepsilon^-_{i,D} = \varepsilon_i I\left(|\varepsilon_i| \le D_n\right),\ \varepsilon^+_{i,D} = \varepsilon_i I\left(|\varepsilon_i| > D_n\right),\ \varepsilon^*_{i,D} = \varepsilon^-_{i,D} - E\left(\varepsilon^-_{i,D}\,|\mathbf{X}_i\right),$$

$$U_{i,\gamma} = \boldsymbol{\mu}^T_{\omega_{J,\gamma}}\mathbf{S}_{21}\left\{B^*_{1,1}(X_{i1}),\cdots,B^*_{1,N^*}(X_{i1})\right\}^T W_{i1}\varepsilon^*_{i,D}.$$

Denote the truncation of $F_{1,\gamma}$ as $F^D_{1,\gamma} = \left|n^{-1}\sum_{i=1}^n U_{i,\gamma}\right|$. Next we show that

$\left|F_{1,\gamma} - F_{1,\gamma}^D\right| = O_P\left(h^{1/2}H\right)$. Note that $\left|F_{1,\gamma} - F_{1,\gamma}^D\right| \leq \Lambda_{1,\gamma} + \Lambda_{2,\gamma}$, where

$$\Lambda_{1,\gamma} = \left|\frac{1}{n}\sum_{i=1}^{n} \sum_{1\leq J^*,J^{*\prime}\leq N^*} \mu_{\omega J,J^*,\gamma} s_{J+N+1,J'+1} B^*_{J^{*\prime},1}(X_{i1})W_{i1}E\left(\varepsilon^-_{i,D}\left|\mathbf{X}_i\right.\right)\right|,$$

$$\Lambda_{2,\gamma} = \left|\frac{1}{n}\sum_{i=1}^{n} \sum_{1\leq J^*,J^{*\prime}\leq N^*} \mu_{\omega J,J^*,\gamma} s_{J+N+1,J'+1} B^*_{J^{*\prime},1}(X_{i1})W_{i1}\varepsilon^+_{i,D}\right|.$$

Let $\boldsymbol{\mu}_{\omega J,\gamma} = \left\{\mu_{\omega J,1,\gamma}, \cdots, \mu_{\omega J,N^*,\gamma}\right\}^T$, then

$$\Lambda_{1,\gamma} = \left|\boldsymbol{\mu}^T_{\omega J,\gamma}\mathbf{S}_{21}\left\{n^{-1}\sum_{i=1}^{n} B^*_{J^{*\prime},1}(X_{i1})W_{i1}E\left(\varepsilon^-_{i,D}\left|\mathbf{X}_i\right.\right)\right\}_{J^{*\prime}=1}^{N^*}\right|$$

$$\leq C_S\left\{\sum_{J=1}^{N^*}\mu^2_{\omega JJ^*,\gamma}\sum_{J=1}^{N^*}\left\{\frac{1}{n}\sum_{i=1}^{n} B^*_{J^*,1}(X_{i1})W_{i1}E\left(\varepsilon^-_{i,D}\left|\mathbf{X}_i\right.\right)\right\}^2\right\}^{1/2}.$$

By Assumption (A3), $\left|E\left(\varepsilon^-_{i,D}\left|\mathbf{X}_i\right.\right)\right| = \left|E\left(\varepsilon^+_{i,D}\left|\mathbf{X}_i\right.\right)\right| \leq M_\delta D_n^{-(1+\delta)}$ and Lemma 3.4 entails that $\sup_{J,\gamma}\left|\frac{1}{n}\sum_{i=1}^{n} B_{J,1}(X_{i1})W_{i1}\right| = O_P\left(\log n/\sqrt{n}\right)$. Therefore

$$\begin{aligned}\Lambda_{1,\gamma} &\leq M_\delta D_n^{-(1+\delta)} \\ &\times \sup_{0\leq J\leq N+1}\left[\sum_{J^*=1}^{N^*}\mu^2_{\omega J,J^*,\gamma}\sum_{J^*=1}^{N^*}\left\{\frac{1}{n}\sum_{i=1}^{n} B^*_{J^*,1}(X_{i1})W_{i1}\right\}^2\right]^{1/2} \\ &= O_P\left\{N^*D_n^{-(1+\delta)}h^{1/2}\log^2 n/n\right\} = O_P\left(h^{1/2}H\right),\end{aligned}$$

where the last step follows from the choice of D_n. Meanwhile

$$\begin{aligned}\sum_{n=1}^{\infty} P\left(|\varepsilon_n| \geq D_n\right) &\leq \sum_{n=1}^{\infty} \frac{E\left|\varepsilon_n\right|^{2+\delta}}{D_n^{2+\delta}} = \sum_{n=1}^{\infty} \frac{E\left(E\left|\varepsilon_n\right|^{2+\delta}|\mathbf{X}_n\right)}{D_n^{2+\delta}} \\ &\leq \sum_{n=1}^{\infty} \frac{M_\delta}{D_n^{2+\delta}} < \infty,\end{aligned}$$

since $\delta > 1/2$. By Borel-Cantelli Lemma, one has with probability 1,

$$n^{-1} \sum_{i=1}^{n} \sum_{1 \leq J^*, J^{*\prime} \leq N^*} \mu_{\omega_{J,J^*,\gamma}} s_{J^*+N^*, J^{*\prime}+1} B^*_{J^{*\prime},1}(X_{i1}) W_{i1} \varepsilon^+_{i,D} = 0,$$

for large n. Therefore, one has

$$\left|F_{1,\gamma} - F^D_{1,\gamma}\right| \leq \Lambda_{1,\gamma} + \Lambda_{2,\gamma} = O_p\left(h^{1/2}H\right)$$

. Next we will show that $F^D_{1,\gamma} = O_p\left(h^{1/2}H\right)$. Note that the variance of $U_{i,\gamma}$ is

$$\boldsymbol{\mu}^T_{\omega_{J,\gamma}} \mathbf{S}_{21} \operatorname{var}\left(\left\{B^*_{1,1}(X_{i1}), \cdots, B^*_{1,N^*}(X_{i1})\right\}^T W_{i1} \varepsilon^*_{i,D}\right) \mathbf{S}_{21} \boldsymbol{\mu}_{\omega_{J,\gamma}}.$$

By Assumption (A3), $c_\sigma^2 \mathbf{V}_{11} \leq \operatorname{var}\left(\left\{B^*_{1,1}(X_{i1}), \cdots, B^*_{1,N^*}(X_{i1})\right\}^T W_{i1}\right) \leq C_\sigma^2 \mathbf{V}_{11}$, $\operatorname{var}\left(U_{i,\gamma}\right) \sim \boldsymbol{\mu}^T_{\omega_{J,\gamma}} \mathbf{S}_{21} \mathbf{V}_{11} \mathbf{S}_{21} \boldsymbol{\mu}_{\omega_{J,\gamma}} V_{\varepsilon,D} = \boldsymbol{\mu}^T_{\omega_{J,\gamma}} \mathbf{S}_{21} \boldsymbol{\mu}_{\omega_{J,\gamma}} V_{\varepsilon,D}$, where $V_{\varepsilon,D} = \operatorname{var}\left\{\varepsilon^*_{i,D} |\mathbf{X}_i\right\}$. Let $\kappa_\gamma = \left\{\boldsymbol{\mu}^T_{\omega_{J,\gamma}} \boldsymbol{\mu}_{\omega_{J,\gamma}}\right\}^{1/2}$, then

$$c_S c_\sigma^2 \left\{\kappa_\gamma\right\}^2 V_{\varepsilon,D} \leq \operatorname{var}\left(U_i\right) \leq C_S C_\sigma^2 \left\{\kappa_\gamma\right\}^2 V_{\varepsilon,D}.$$

Simple calculation leads to that

$$E\left|U_{i,\gamma}\right|^r \leq \left\{c_0 \kappa_\gamma D_n H^{-1/2}\right\}^{r-2} r! E\left|U_{i,\gamma}\right|^2 < +\infty,$$

where the last step follows from the choice of D_n. Meanwhile

$$\begin{aligned}\sum_{n=1}^{\infty} P(|\varepsilon_n| \geq D_n) &\leq \sum_{n=1}^{\infty} \frac{E|\varepsilon_n|^{2+\delta}}{D_n^{2+\delta}} = \sum_{n=1}^{\infty} \frac{E\left(E|\varepsilon_n|^{2+\delta}|\mathbf{X}_n\right)}{D_n^{2+\delta}} \\ &\leq \sum_{n=1}^{\infty} \frac{M_\delta}{D_n^{2+\delta}} < \infty,\end{aligned}$$

since $\delta > 1/2$. By Borel-Cantelli Lemma, one has with probability 1,

$$n^{-1}\sum_{i=1}^{n} \sum_{1\leq J^*,J^{*\prime}\leq N^*} \mu_{\omega J,J^*,\gamma} s_{J^*+N^*,J^{*\prime}+1} B^*_{J^{*\prime},1}(X_{i1})W_{i1}\varepsilon^+_{i,D} = 0,$$

for large n. Therefore, one has

$$\left|F_{1,\gamma} - F^D_{1,\gamma}\right| \leq \Lambda_{1,\gamma} + \Lambda_{2,\gamma} = O_p\left(h^{1/2}H\right)$$

. Next we will show that $F^D_{1,\gamma} = O_p\left(h^{1/2}H\right)$. Note that the variance of $U_{i,\gamma}$ is

$$\boldsymbol{\mu}^T_{\omega J,\gamma}\mathbf{S}_{21} \operatorname{var}\left(\left\{B^*_{1,1}(X_{i1}),\cdots,B^*_{1,N^*}(X_{i1})\right\}^T W_{i1}\varepsilon^*_{i,D}\right)\mathbf{S}_{21}\boldsymbol{\mu}_{\omega J,\gamma}.$$

By Assumption (A3), $c_\sigma^2\mathbf{V}_{11} \leq \operatorname{var}\left(\left\{B^*_{1,1}(X_{i1}),\cdots,B^*_{1,N^*}(X_{i1})\right\}^T W_{i1}\right) \leq C_\sigma^2\mathbf{V}_{11}$, $\operatorname{var}\left(U_{i,\gamma}\right) \sim \boldsymbol{\mu}^T_{\omega J,\gamma}\mathbf{S}_{21}\mathbf{V}_{11}\mathbf{S}_{21}\boldsymbol{\mu}_{\omega J,\gamma}V_{\varepsilon,D} = \boldsymbol{\mu}^T_{\omega J,\gamma}\mathbf{S}_{21}\boldsymbol{\mu}_{\omega J,\gamma}V_{\varepsilon,D}$, where $V_{\varepsilon,D} = \operatorname{var}\left\{\varepsilon^*_{i,D}|\mathbf{X}_i\right\}$. Let $\kappa_\gamma = \left\{\boldsymbol{\mu}^T_{\omega J,\gamma}\boldsymbol{\mu}_{\omega J,\gamma}\right\}^{1/2}$, then

$$c_S c_\sigma^2\left\{\kappa_\gamma\right\}^2 V_{\varepsilon,D} \leq \operatorname{var}(U_i) \leq C_S C_\sigma^2\left\{\kappa_\gamma\right\}^2 V_{\varepsilon,D}.$$

Simple calculation leads to that

$$E\left|U_{i,\gamma}\right|^r \leq \left\{c_0\kappa_\gamma D_n H^{-1/2}\right\}^{r-2} r! E\left|U_{i,\gamma}\right|^2 < +\infty,$$

for $r \geq 3$, so $\left\{U_{i,\gamma}\right\}_{i=1}^{n}$ satisfies the Cramér's condition with Cramér's constant $c_* = c_0 \kappa_\gamma D_n H^{-1/2}$. Hence by the Bernstein's inequality,

$$P\left\{\left|n^{-1}\sum_{l=1}^{n} U_{i,\gamma}\right| \geq \rho_n\right\} \leq a_1 \exp\left(-\frac{q\rho_n^2}{25m_2^2 + 5c_*\rho_n}\right) + a_2(3)\,\alpha\left(\left[\frac{n}{q+1}\right]\right)^{6/7},$$

where $m_2^2 \sim \{\kappa\alpha\}^2 V_{\varepsilon,D}$, $m_3 \leq \left\{c\{\kappa\alpha\}^3 H^{-1/2} D_n V_{\varepsilon,D}\right\}^{1/3}$, $\rho_n = \rho h^{1/2} H$, $a_1 = 2\frac{n}{q} + 2\left(1 + \frac{\rho_n^2}{25m_2^2 + 5c_*\rho_n}\right)$, $a_2(3) = 11n\left(1 + \frac{5m_3^{6/7}}{\rho_n}\right)$. Similar arguments as in Lemma 3.12 yield that as $n \to \infty$, $\frac{q\rho_n^2}{25m_2^2 + 5c_*\rho_n} \sim \frac{q\rho_n}{c_*} = \frac{\rho n^{2/5}}{c_0(\log n)^{5/2} D_n} \to +\infty$. For c_0, ρ large enough,

$$P\left\{\frac{1}{n}\left|\sum_{i=1}^{n} U_{i,\gamma}\right| > \rho h^{1/2} H\right\} \leq c\log n \exp\left\{-c_2\rho^2 \log n\right\} + Cn^{2-6\lambda_0 c_0/7} \leq n^{-3},$$

for n large enough. Hence

$$\sum_{n=1}^{\infty} P\left(\left|W_{1,\gamma}^D\right| \geq \rho h^{1/2} H\right) = \sum_{n=1}^{\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} U_i\right| \geq \rho h^{1/2} H\right) \leq \sum_{n=1}^{\infty} n^{-3} < \infty.$$

Thus, Borel-Cantelli Lemma entails that $F_{1,\gamma}^D = O_p\left(h^{1/2} H\right)$. Noting the fact that $\left|F_{1,\gamma} - F_{1,\gamma}^D\right| = O_p\left(h^{1/2} H\right)$, one has that $F_{1,\gamma} = O_p\left(h^{1/2} H\right)$. Similarly $F_{2,\gamma} = O_p\left(h^{1/2} H\right)$. Thus

$$Q_1 \leq (d-1)\left(\sup_{2\leq\gamma\leq d} F_{1,\gamma} + \sup_{2\leq\gamma\leq d} F_{2,\gamma}\right) = O_p\left(h^{1/2} H\right), \tag{3.57}$$

and one has $Q_1 = O_p\left(h^{1/2} H\right)$. The result follows from (3.55) and (3.56).

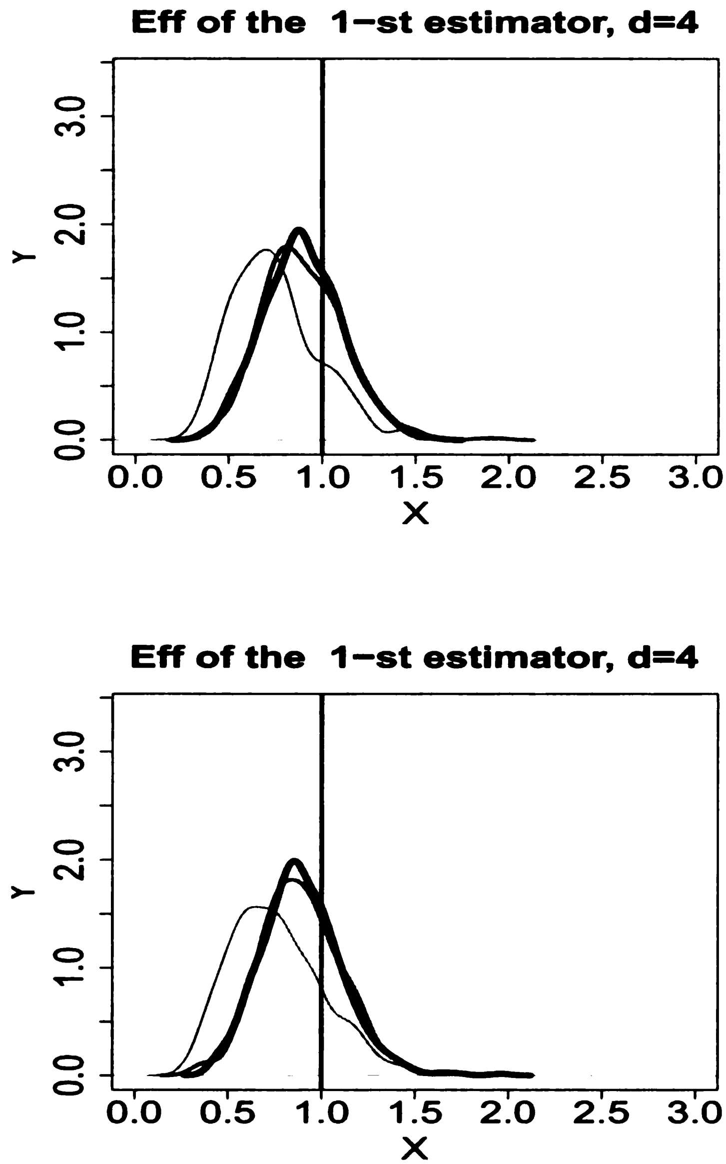


Figure 3.1: Plots of the efficiency of SBS estimator $\hat{m}_{\alpha,\mathrm{SBS}}$ corresponding to oracle smoother $\tilde{m}_{\alpha,\mathrm{S}}$ for $d = 4$ and $\rho = 0$ (upper panel), $\rho = .3$ (lower panel) of $m_\alpha(x_\alpha)$ in (3.24), for $\alpha = 1$ (thick curve for $n = 1000$, thin curve for $n = 500$, and solid curve for $n = 100$).

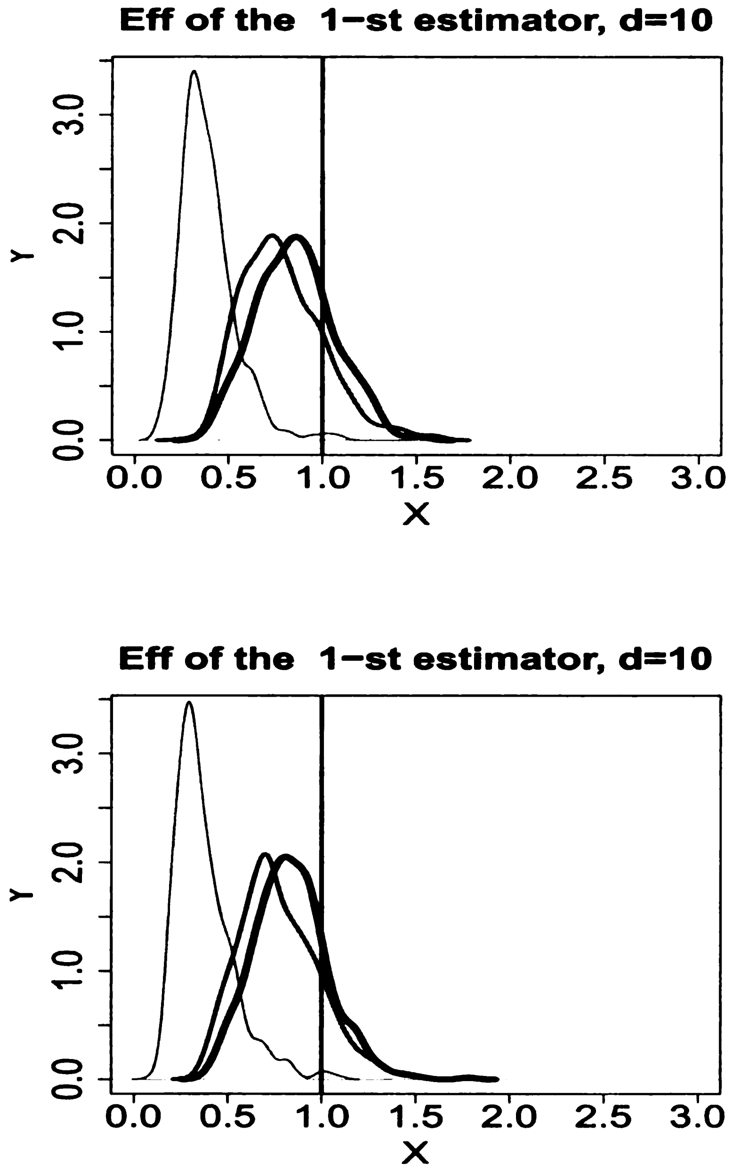


Figure 3.2: Plots of the efficiency of SBS estimator $\hat{m}_{\alpha,\mathrm{SBS}}$ corresponding to oracle smoother $\tilde{m}_{\alpha,\mathrm{S}}$ for $d = 10$ and $\rho = 0$ (upper panel), $\rho = .3$ (lower panel) of $m_\alpha(x_\alpha)$ in (3.24), for $\alpha = 1$ (thick curve for $n = 1000$, thin curve for $n = 500$, and solid curve for $n = 100$).

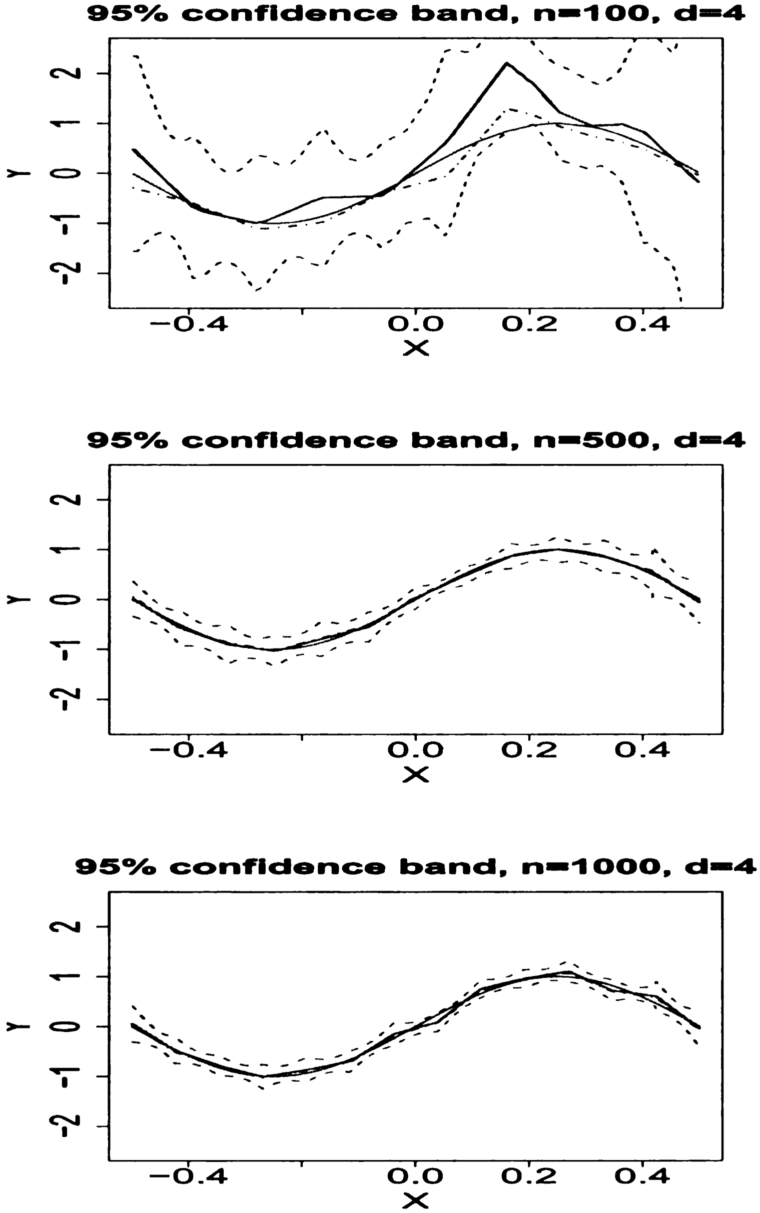


Figure 3.3: For $\rho = 0$, plots of the oracle smoother $\tilde{m}_{\alpha,\mathrm{S}}$ (dotted curve), SBS estimator $\hat{m}_{\alpha,\mathrm{SBS}}$ (solid curve) and the 95% confidence bands (upper and lower dashed curves) of the function components $m_{\alpha}(x_{\alpha})$ in (3.9) with $\alpha = 1$ (thin solid curve).

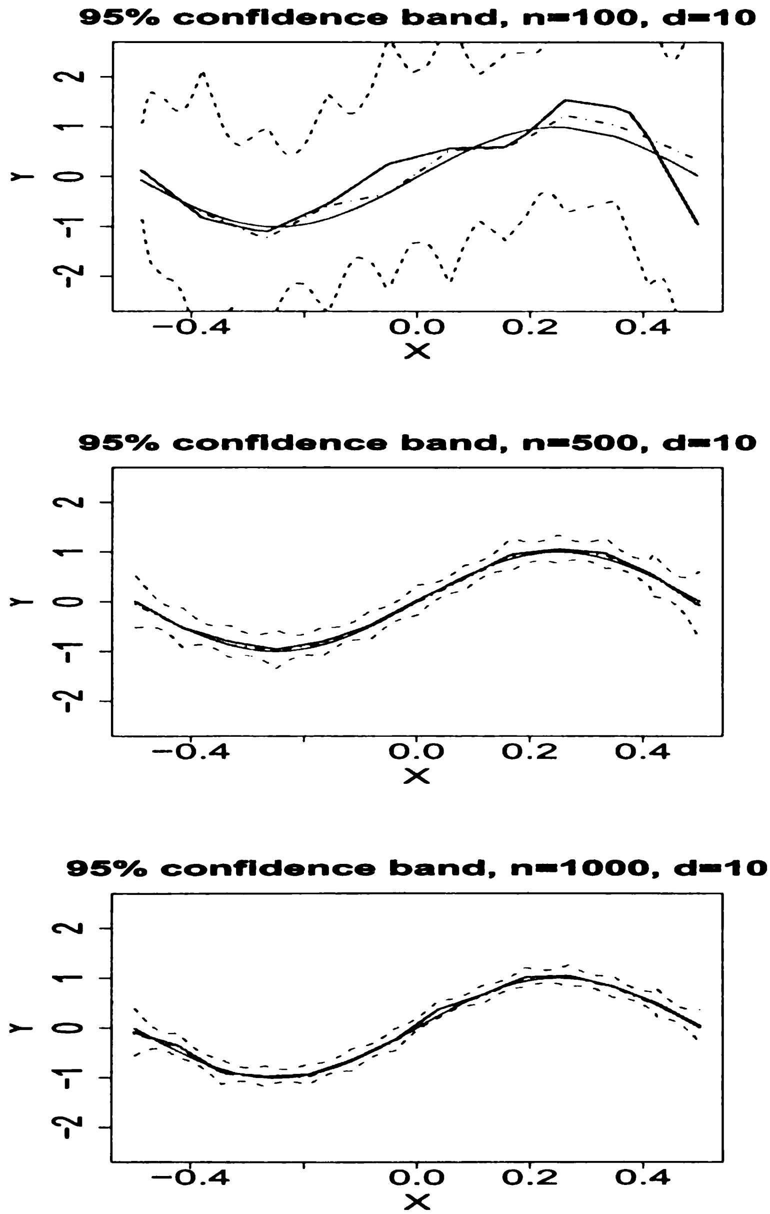


Figure 3.4: For $\rho = .3$, plots of the oracle smoother $\tilde{m}_{\alpha,\mathrm{S}}$ (dotted curve), SBS estimator $\hat{m}_{\alpha,\mathrm{SBS}}$ (solid curve) and the 95% confidence bands (upper and lower dashed curves) of the function components $m_\alpha(x_\alpha)$ in (3.9) with $\alpha = 1$ (thin solid curve).

Chapter 4

A simultaneous confidence band for dense longitudinal regression

4.1 Introduction

Traditional statistical methods fail often as we deal with functional data. Indeed, if for instance we consider a sample of finely discretized curves, two crucial statistical problems appear. The first comes from the ratio between the size of the sample and the number of variables (each real variable corresponding to one discretized point). The second, is due to the existence of strong correlations between the variables and becomes an ill-conditioned problem in the context of multivariate linear model. So, there is a real necessity to develop statistical methods/models in order to take into account the functional structure of this kind of data.

Functional data with different design are increasingly common in modern data analysis. A functional data set has the form $\left\{X_{ij}, Y_{ij}\right\}, 1 \leq i \leq n, 1 \leq j \leq N_i$, in which N_i observations are taken for i^{th} subject, with X_{ij} and Y_{ij} the j^{th} predictor and response variables, respectively, for the i^{th} subject. In this chapter we only deal with the equally spaced design. For simplicity, we only consider the case $N_1 =$

$N_2 = \cdots = N_n = N$. Without loss of generality, the predictor X_{ij} takes values $\{1/N, 2/N, ..., N/N\}$ for the i^{th} subject, $i = 1, 2, ..., n$. For the i^{th} subject, its sample path $\left\{j/N, Y_{ij}\right\}$ is the noisy realization of a continuous time stochastic process $\xi_i(x)$ in the sense that

$$Y_{ij} = \xi_i(j/N) + \sigma(j/N)\varepsilon_{ij},$$

with errors ε_{ij} satisfying $E\left(\varepsilon_{ij}\right) = 0$, $E(\varepsilon_{ij}^2) = 1$, and $\{\xi_i(x), x \in \mathcal{X}\}$ are iid copies of a process $\{\xi(x), x \in \mathcal{X}\}$ which is L^2, i.e., $E\int_{\mathcal{X}} \xi^2(x)dx < +\infty$.

For the standard process $\{\xi(x), x \in \mathcal{X}\}$, one defines the mean function $m(x) = E\{\xi(x)\}$ and the covariance function $G\left(x, x'\right) = \text{cov}\left\{\xi(x), \xi(x')\right\}$. Let sequences $\{\lambda_k\}_{k=1}^{\infty}$, $\{\psi_k(x)\}_{k=1}^{\infty}$ be the eigenvalues and eigenfunctions of $G\left(x, x'\right)$ respectively, in which $\lambda_1 \geq \lambda_2 \geq \cdots \geq 0$ with $\sum_{k=1}^{\infty} \lambda_k < \infty$, $\{\psi_k\}_{k=1}^{\infty}$ form an orthonormal basis of $L^2(\mathcal{X})$ and $G\left(x, x'\right) = \sum_{k=1}^{\infty} \lambda_k \psi_k(x) \psi_k\left(x'\right)$, which implies that $\int G\left(x, x'\right) \psi_k\left(x'\right) dx' = \lambda_k \psi_k(x)$.

The process $\{\xi_i(x), x \in \mathcal{X}\}$ allows the Karhunen-Loève L^2 representation

$$\xi_i(x) = m(x) + \sum_{k=1}^{\infty} \xi_{ik}\phi_k(x),$$

where the random coefficients ξ_{ik} are uncorrelated with mean 0 and variances 1, and the functions $\phi_k = \sqrt{\lambda_k}\psi_k$. In what follows, we assume that $\lambda_k = 0$, for $k > \kappa$, where κ is a positive integer or $+\infty$, thus $G(x, x') = \sum_{k=1}^{\kappa} \phi_k(x)\phi_k\left(x'\right)$ and the data generating process is now written as

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

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

Two distinct types of functional data have been studied. Yao, Müller and Wang

[80, 81], Yao [82] and Ma, Yang and Carroll [45] studied sparse longitudinal data for which $1 \leq j \leq N_i$ and N_i's are iid copies of an integer valued positive random variable. While Li and Hsing [39, 40] concern dense functional data. For the dense functional data, strong uniform convergence rates are developed for local-linear smooth estimators, but no uniform confidence bands have been given. The fact that simultaneous confidence band has not been established for functional data analysis is certainly not due to lack of interesting applications, but to the greater technical difficulty to formulate such bands for functional data and establish their theoretical properties.

In this chapter, we present simultaneous confidence bands for $m(x)$ in dense longitudinal data given in (4.1) via local linear smoothing approach.

The chapter is a joint work with Yang, L., Liu, R. and Shao, Q. We organize our chapter as follows. In Section 4.2 we state our main results on confidence bands constructed from local linear smoothing. In Section 4.3 we provide further insights into the error structure of local linear estimators. Section 4.4 describes the actual steps to implement the confidence bands. Section 4.5 reports findings of a simulation study. An empirical example in Section 4.6 illustrates how to use the proposed spline estimator with confidence band for inference. Proofs of technical lemmas are in the Appendix.

4.2 Main results

For any Lebesgue measurable function ϕ on $[0,1]$, denote $\|\phi\|_r = \left\{\int_0^1 |\phi(x)|^r \, dx\right\}^{1/r}$, $1 \leq r < \infty$ and $\|\phi\|_\infty = \sup_{x\in[0,1]} |\phi(x)|$, and for a continuous function ϕ on $[0,1]$ denote the modulus of continuity as

$$\omega(\phi, \delta) = \max_{x,x' \in [0,1], |x-x'| \leq \delta} \left|\phi(x) - \phi\left(x'\right)\right|.$$

For any $\beta \in (0,1]$, we denote by $C^{0,\beta}[0,1]$ the space of order β Hölder continuous function on $[0,1]$, i.e.,

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

in which $\|\phi\|_{0,\beta}$ is called the $C^{0,\beta}$-norm of ϕ. Clearly, $C^{0,\beta}[0,1] \subset C[0,1]$ and if $\phi \in C^{0,\beta}[0,1]$, then $\omega(\phi,\delta) \leq \|\phi\|_{0,\beta}\delta^{\beta}$. For any vector $\boldsymbol{\zeta} = (\zeta_1, ..., \zeta_s) \in R^s$, denote the norm $\|\boldsymbol{\zeta}\|_r = (|\zeta_1|^r + \cdots + |\zeta_s|^r)^{1/r}, 1 \leq r < +\infty$, $\|\boldsymbol{\zeta}\|_\infty = \max(|\zeta_1|, ..., |\zeta_s|)$.

We are using the local linear estimation in this paper.

The technical assumptions we need are as follows:

(A1) *The regression function* $m' \in C[0,1]$.

(A2) *The standard deviation function* $\sigma(x) \in C^{0,\beta}[0,1]$. *For any* $k = 1, 2, \ldots \kappa$, $\phi_k(x) \in C^{0,\beta}[0,1]$ for $\beta \in [0,1]$ *and* $\min_{x \in [0,1]} G(x,x) > 0$.

(A3) *As* $n \to \infty$, $Nn^{-1/4}(\log n)^{-1} \to \infty$ *and* $N = O\left(n^{\theta}\right)$ *for some* $\theta > 5/8$. *The bandwidth* h *satisfies* $Nh(\log n)^{-1} \to \infty, nh^4 \to 0$ as $n \to \infty$.

(A4) *The number* κ *of nonzero eigenvalues is finite. The variables* $(\xi_{ik})_{i=1,k=1}^{\infty,\kappa}$ and $\left(\varepsilon_{ij}\right)_{i=1,j=1}^{\infty,\infty}$ *are independent. In addition,* $\max_{1 \leq k \leq \kappa} E|\xi_{1k}|^{\eta_1} < +\infty$ *for some* $\eta_1 > 4$ *while* $E|\varepsilon_{11}|^{\eta_2} < +\infty$, *for some* $\eta_2 > 4 + 2\theta$ *with* θ *being the constant in Assumption (A3).*

(A5) *The kernel* $K_h(x)$ *is a second order smooth function and satisfies the following conditions:* $K_h(x) = \frac{1}{h}K\left(\frac{x}{h}\right)$, *where* $K(\cdot)$ *is a density function with bounded support* $[-1,1]$ *and symmetric about 0 unless special conditions are indicated. It is Lipschitz continuous.*

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 $100(1-\alpha)^{th}$ percentile of the absolute maxima distribution of $\zeta(x)$, for all $x \in [0,1]$,

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

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

$$\overline{m}(x) = \bar{\xi}(x) = n^{-1}\sum\nolimits_{i=1}^{n} \xi_i(x), x \in [0,1]. \tag{4.2}$$

The term "infeasible" refers to the fact that $\overline{m}(x)$ is computed from unknown quantity $\xi_i(x), x \in [0,1]$, while $\overline{m}(x)$ would be the natural estimator of $m(x)$ if all the iid random curves $\xi_i(x), x \in [0,1]$ were observed, a view taken in Ferraty and Vieu [18]. We propose to estimate the mean function $m(x)$ by solving the local linear least square

$$\left(\hat{a}, \hat{b}\right) = \arg\min \sum_{i=1}^{n}\sum_{j=1}^{N}\left\{Y_{ij} - a - b\left(\frac{j}{N} - x\right)\right\}^2 K_h\left(\frac{j}{N} - x\right)$$

with $K_h(u) = \frac{1}{h}K\left(\frac{u}{h}\right)$, $h = h_n \to 0$, as $n \to \infty$. For any $x \in [0,1]$,

$$\hat{m}(x) = \hat{a} = e_0^T\left(\mathbf{X}^T\mathbf{W}\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{W}\mathbf{Y} \tag{4.3}$$

in which $\mathbf{Y} = \left(\bar{Y}_{.,1}, \ldots, \bar{Y}_{.,N}\right)^T, \bar{Y}_{.,j} = n^{-1}\sum_{i=1}^{n} Y_{ij}, 1 \leq j \leq N$, $e_0^T = (1,0)$, and the design matrix $\mathbf{X}$ is

$$\mathbf{X} = \begin{pmatrix} 1 & (\frac{1}{N} - x) \\ \vdots & \vdots \\ 1 & (\frac{N}{N} - x) \end{pmatrix}_{N\times 2}, \tag{4.4}$$

and $\mathbf{W}$ =diag$\left\{\left\{K_h\left(j/N - x\right)/N\right\}_{j=1}^{N}\right\}$.

We now state our main results in the following theorems.

Theorem 4.1. *Under Assumptions (A1)-(A4), for* $\forall \alpha \in (0,1)$ *,as* $n \to \infty$*, the "infeasible estimator"* $\overline{m}(x)$ *converges at the* $\sqrt{n}$ *rate*

$$P\left\{\sup_{x\in[0,1]} n^{1/2} \left|\overline{m}(x) - m(x)\right| G\left(x,x\right)^{-1/2} \leq Q_{1-\alpha}\right\} \to 1-\alpha,$$

$$P\left\{n^{1/2} \left|\overline{m}(x) - m(x)\right| G\left(x,x\right)^{-1/2} \leq z_{1-\alpha/2}\right\} \to 1-\alpha, \forall x \in [0,1],$$

while the local linear estimator $\hat{m}$ *is asymptotically equivalent to* $\overline{m}$

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

Corollary 4.1. *Under Assumptions (A1)-(A4), as* $n \to \infty$*, an asymptotic* $100\left(1-\alpha\right)\%$ *exact confidence band for* $m(x), x \in [0,1]$ *is*

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

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

4.3 Decomposition

In this section, we break the estimation error $\hat{m}(x)-m(x)$ into a bias term and a noise term. To understand this decomposition, we begin by discussing the representation of the local linear estimator $\hat{m}(x)$ in (4.3). We obtain the following crucial decomposition

$$\hat{m}(x) = \widetilde{m}(x) + \widetilde{e}(x) + \widetilde{\xi}(x), \tag{4.5}$$

with

$$\begin{aligned} \widetilde{m}\left(x\right) &= e_0^T \left(\mathbf{X}^T \mathbf{W} \mathbf{X}\right)^{-1} \mathbf{X}^T \mathbf{W} \mathbf{m} \\ \widetilde{e}\left(x\right) &= e_0^T \left(\mathbf{X}^T \mathbf{W} \mathbf{X}\right)^{-1} \mathbf{X}^T \mathbf{W} \mathbf{e} \\ \tilde{\xi}\left(x\right) &= e_0^T \left(\mathbf{X}^T \mathbf{W} \mathbf{X}\right)^{-1} \mathbf{X}^T \mathbf{W} \boldsymbol{\xi}, \end{aligned} \tag{4.6}$$

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

The next three propositions concerns $\widetilde{m}(x), \widetilde{e}(x)$ and $\tilde{\xi}(x)$ given in (4.5).

Proposition 4.1. *Under Assumptions (A1) and (A3), as $n \to \infty$*

$$\sup_{x\in[0,1]} n^{1/2} \left|\widetilde{m}(x) - m(x)\right| G\left(x,x\right)^{-1/2} = o\left(1\right).$$

Proposition 4.2. *Under Assumptions (A2)-(A4), as $n \to \infty$*

$$\sup_{x\in[0,1]} n^{1/2} \left|\overline{m}(x) - m\left(x\right) - \widetilde{\xi}(x)\right| = o_p\left(1\right), \tag{4.7}$$

and there is a version $\widetilde{\zeta}(x)$ *of* $\zeta(x)$ *such that*

$$\sup_{x\in[0,1]} n^{1/2}\left|\widetilde{\zeta}(x)-\widetilde{\xi}(x)\right| = o_p(1), \tag{4.8}$$

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

$$\begin{aligned} P\left\{\sup_{x\in[0,1]} n^{1/2}\left|\overline{m}(x)-m(x)\right| G(x,x)^{-1/2} \le Q_{1-\alpha}\right\} &\rightarrow 1-\alpha, \\ P\left\{\sup_{x\in[0,1]} n^{1/2}\left|\widetilde{\xi}(x)\right| G(x,x)^{-1/2} \le Q_{1-\alpha}\right\} &\rightarrow 1-\alpha. \end{aligned} \tag{4.9}$$

Proposition 4.3. *Under Assumptions (A2)-(A4), as* $n \to \infty$

$$\sup_{x\in[0,1]} n^{1/2}\left|\widetilde{e}(x)\right| G(x,x)^{-1/2} = o_p(1).$$

The Appendix contains proofs for the above three propositions. Combining these propositions with the decomposition of $\hat{m}(x)$ as given in (4.5), we can easily get Theorem 4.1.

4.4 Implementation

In this section, we describe procedures to implement the confidence bands and intervals given in Corollary 4.1. Given any data set $\left(j/N, Y_{ij}\right)_{j=1,i=1}^{N,n}$ from model (4.1), the local linear estimator $\hat{m}(x)$ is obtained by (4.3). When constructing the confidence bands, one needs to evaluate $100(1-\alpha)^{th}$ percentile $Q_{1-\alpha}$ by estimating the unknown functions $G(x,x)$.

The pilot estimator of covariance function $G\left(x,x'\right)$ is $\hat{G}\left(x,x'\right)=\hat{a}\left(x,x'\right)$ such that

$$\left\{\hat{a}\left(x,x'\right), \hat{b}_1\left(x,x'\right), \hat{b}_2\left(x,x'\right)\right\} = \underset{a,b_1,b_2}{\operatorname{argmin}} \sum_{j,j'=1}^{N}$$

$$\left\{C_{\cdot jj'} - a - b_1 (j/N - x) - b_2 \left(j'/N - x'\right)\right\}^2 K_h (j/N - x) K_h \left(j'/N - x'\right),$$

where

$$C_{\cdot jj'} = n^{-1} \sum_{i=1}^{n} \left\{Y_{ij} - \hat{m}_p (j/N)\right\} \left\{Y_{ij} - \hat{m}_p \left(j'/N\right)\right\},$$

for $1 \leq j, j' \leq N$.

Therefore. as $n \to \infty$,

$$\hat{m}_p(x) \pm \hat{G}(x,x)^{1/2} Q_{1-\alpha} n^{-1/2} \tag{4.10}$$

and $\hat{m}_p(x) \pm \hat{G}(x,x)^{1/2} Z_{1-\alpha/2} n^{-1/2}$ have asymptotic confidence level $1-\alpha$.

4.5 Simulation

We carried out some simulations to illustrate the finite sample behavior of the proposed confidence bands defined in Section 2. We generated data from model

$$Y_{ij} = m (j/N) + \sum_{k=1}^{2} \xi_{ik} \phi_k (j/N) + \sigma \varepsilon_{ij}, 1 \leq j \leq N, 1 \leq i \leq n, \tag{4.11}$$

with $\xi_{ik} \sim \text{Normal}(0,1), k = 1,2$, $\varepsilon \sim \text{Normal}(0,1)$, for $1 \leq i \leq n$, and $m(x) = \sin\{2\pi (x - 1/2)\}$. We take orthonormal functions $\phi_1(x) = -2\cos\{\pi (x-1/2)\}$ and $\phi_2(x) = \sin\{\pi (x - 1/2)\}$ to be the eigenfunctions, thus $\lambda_1 = 2, \lambda_2 = 1/2$. Different noise levels $\sigma = 0.5, 1$ were used to interpret the result, and the number of subjects n was taken to be 50, 100, 200 and 500. We used $N = [n^{0.8} \log n]$ to determine the number of grid for each subject.

Table 4.1 shows the coverage frequencies from 200 replications for the confidence levels $1 - \alpha = 0.95$ and 0.99. As we expected, the coverage rates go to the nominal ones as the sample sizes increase.

Table 4.1: Coverage frequencies from 200 replications.

σ	n	$1-\alpha=.95$	$1-\alpha=.99$
.5	50	0.9	0.97
	100	0.91	0.995
	200	0.94	0.99
	500	0.94	0.99
1	50	0.855	0.95
	100	0.905	0.96
	200	0.89	0.975
	500	0.865	0.97

4.6 Empirical example

In this section, we have applied the confidence band procedure of Section 4.4 to the 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. In this study, we used 240 meat samples with each consisting of a 100 channel spectrum of absorbance and the contents of moisture (water), fat and protein. Figure 4.3 shows this data set with the confidence band for the mean. We can clearly see that there is no linear or quadratic pattern for the Tecator mean.

4.7 Appendix

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

4.7.1 Preliminaries

We first state some results are used in the proofs of Lemma 4.2.

Lemma 4.1. *[Theorem 2.6.7 of [8]] Suppose that* $\xi_i, 1 \le i \le n$ *are iid with* $E(\xi_1) = 0, E(\xi_1^2) = 1$ *and* $H(x) > 0$ $(x \ge 0)$ *is an increasing continuous function such* *that* $x^{-2-\gamma}H(x)$ *is increasing for some* $\gamma > 0$ *and* $x^{-1}logH(x)$ *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 *such that for any* $\{x_n\}_{n=1}^{\infty}$ *satisfying* $H^{-1}(n) < x_n < C_1(nlogn)^{1/2}$ *and* $S_t = \sum_{i=1}^{t} \xi_i$

$$P\left\{\max_{1\leq t\leq n} |S_t - W(t)| > x_n\right\} \leq C_2 n \{H(ax_n)\}^{-1}.$$

4.7.2 Proof of Theorem

PROOF OF PROPOSITION 4.1. $\tilde{m}(x) = e_0^T \left(\mathbf{X}^T\mathbf{W}\mathbf{X}\right)^{-1} \mathbf{X}^T\mathbf{W}\mathbf{m}$. The dispersion matrix

$$\mathbf{X}^T\mathbf{W}\mathbf{X} = \text{diag}(1,h)\,\mathbf{D}_{N,x}\,\text{diag}(1,h),$$

where

$$\mathbf{D}_{N,x} = \begin{pmatrix} s_{N,0}(x) & s_{N,1}(x) \\ s_{N,1}(x) & s_{N,2}(x) \end{pmatrix}$$

where $s_{N,l}(x) = N^{-1}\sum_{j=1}^{N} K_h(j/N - x)\{(j/N - x)/h\}^l, l = 0, 1, 2$. Denote

$$\mathbf{D}_x = \begin{pmatrix} \mu_{0,x}(K) & \mu_{1,x}(K) \\ \mu_{1,x}(K) & \mu_{2,x}(K) \end{pmatrix}$$

where

$$\mu_{l,x}(K) = \begin{cases} \int_{-x/h}^{1} v^l K(v)\,dv & x \in [0, h) \\ \int_{-1}^{1} v^l K(v)\,dv & x \in [h, 1-h] \\ \int_{-1}^{(1-x)/h} v^l K(v)\,dv & x \in (1-h, 1] \end{cases}$$

$$\mathbf{D}_{n,x} = \mathbf{D}_x + U(h)$$

$$\left(\mathbf{X}^T\mathbf{W}\mathbf{X}\right)^{-1} = \text{diag}\left(1, h^{-1}\right)\left\{\mathbf{D}_x^{-1} + U(h)\right\}\text{diag}\left(1, h^{-1}\right)$$

Without loss of generality, let $x \in [h, 1-h]$, one has

$$\begin{aligned}
\widetilde{m}(x) - m(x) &= e_0^T \left(\mathbf{X}^T \mathbf{W} \mathbf{X}\right)^{-1} \mathbf{X}^T \mathbf{W} \left\{\mathbf{m} - m(x)\mathbf{X}e_0 - m'(x)\mathbf{X}e_1\right\} \\
&= e_0^T \operatorname{diag}\left(1, h^{-1}\right) \left\{\mathbf{D}_x^{-1} + U(h)\right\} \\
&\quad \operatorname{diag}\left(1, h^{-1}\right) \mathbf{X}^T \mathbf{W} \left\{\mathbf{m} - m(x)\mathbf{X}e_0 - m'(x)\mathbf{X}e_1\right\}
\end{aligned}$$

$$\begin{aligned}
&\operatorname{diag}\left(1, h^{-1}\right) \mathbf{X}^T \mathbf{W} \left\{\mathbf{m} - m(x)\mathbf{X}e_0 - m'(x)\mathbf{X}e_1\right\} \\
= &\begin{pmatrix} N^{-1}\sum_{j=1}^N K_h(j/N - x) \\ \times\left\{m(j/N) - m(x) - m'(x)(j/N - x)\right\} \\ N^{-1}\sum_{j=1}^N K_h(j/N - x)\{(j/N - x)/h\} \\ \times\left\{m(j/N) - m(x) - m'(x)(j/N - x)\right\} \end{pmatrix} \\
= &\begin{pmatrix} N^{-1}\sum_{j=1}^N K_h(j/N - x) \\ \times\left\{\frac{1}{2}m''(x)(j/N - x)^2 + u\left(h^2\right)\right\} \\ N^{-1}\sum_{j=1}^N K_h(j/N - x)\{(j/N - x)/h\} \\ \times\left\{\frac{1}{2}m''(x)(j/N - x)^2 + u\left(h^2\right)\right\} \end{pmatrix} \\
= &\frac{1}{2}m''(x)h^2 \begin{pmatrix} N^{-1}\sum_{j=1}^N K_h(j/N - x)\{(j/N - x)/h\}^2 + u(1) \\ N^{-1}\sum_{j=1}^N K_h(j/N - x)\{(j/N - x)/h\}^3 + u(1) \end{pmatrix} \\
= &\frac{1}{2}m''(x)h^2 \begin{pmatrix} \mu_{2,x}(K) + u(1) \\ \mu_{3,x}(K) + u(1) \end{pmatrix}.
\end{aligned}$$

Combining the above two big equations, we have

$$\begin{aligned}
&m(x) - m(x) \\
= \;&e_0^T \operatorname{diag}\left(1, h^{-1}\right)\left\{\mathbf{D}_x^{-1} + U(h)\right\} \frac{1}{2}m''(x)h^2 \begin{pmatrix} \mu_2(K) + u(1) \\ \mu_3(K) + u(1) \end{pmatrix}
\end{aligned}$$

$$= \frac{1}{2}m''(x)\, h^2 e_0^T \operatorname{diag}\left(1, h^{-1}\right) \mathbf{D}_x^{-1} \begin{pmatrix} \mu_2(K) + u(1) \\ \mu_3(K) + u(1) \end{pmatrix} + U\left(h^3\right)$$

$$= U\left(h^2\right).$$

Lemma 4.2. *Under Assumptions (A2)-(A4), there exist iid standard normal random variables $Z_{ik,\xi}, Z_{ij,\varepsilon}, 1 \leq i \leq n, 1 \leq j \leq N, 1 \leq k \leq \kappa$ and some $\beta \in (0, 1/2)$ such that as $n \to \infty$*

$$\max_{1\leq k\leq \kappa}\left|\bar{\xi}_{\cdot,k} - \bar{Z}_{\cdot,k,\xi}\right| + \max_{1\leq j\leq N}\left|\bar{\varepsilon}_{\cdot,j} - \bar{Z}_{\cdot,j,\varepsilon}\right| = O_{a.s.}\left(n^{\beta-1}\right) \tag{4.12}$$

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

Proof. Aaccording to Assumption (A4), $E\left|\xi_{ik}\right|^{\eta_1} < +\infty$, $\eta_1 > 4, E\left|\varepsilon_{ij}\right|^{\eta_2} < +\infty, \eta_2 > 4 + 2\theta$, so there exists some $\beta \in (0, 1/2)$ such that $\eta_1 > 2/\beta$, $\eta_2 > (2+\theta)/\beta$. Let $H(x) = x^{\eta_1}$, $x_n = n^\beta$ in Lemma 4.1, then

$$\frac{n}{H(ax_n)} = a^{-\eta_1} n^{1-\eta_1\beta} = O\left(n^{-\gamma_1}\right)$$

for some $\gamma_1 > 1$. Applying Borel-Cantelli Lemma, one finds iid variables $Z_{ik,\xi} \sim N(0,1)$ such that

$$\max_{1\leq k\leq \kappa}\max_{1\leq t\leq n}\left|\sum_{i=1}^{t}\xi_{ik} - \sum_{i=1}^{t} Z_{ik,\xi}\right| = O_{\text{a.s.}}\left(n^{\beta}\right).$$

Likewise, if one lets $H(x) = x^{\eta_2}, x_n = n^\beta$ in Lemma 4.1, then

$$\frac{n}{H(ax_n)} = a^{-\eta_2} n^{1-\eta_2\beta} = O\left(n^{-\gamma_2-\theta}\right)$$

for some $\gamma_2 > 1$. Applying Borel-Cantelli Lemma and Lemma 4.1, one finds iid variables $Z_{ij,\varepsilon} \sim N(0,1)$ such that

$$\max_{1\le j\le N} P\left\{\max_{1\le t\le n}\left|\sum_{i=1}^{t}\varepsilon_{ij}-\sum_{i=1}^{t}Z_{ij,\varepsilon}\right|>n^{\beta}\right\}\le C\frac{n}{H(axn)}=Cn^{1-\eta_2\beta},$$

which entails that $\max_{1\le j\le N} P\left\{\left|\sum_{i=1}^{n}\varepsilon_{ij}-\sum_{i=1}^{n}Z_{ij,\varepsilon}\right|>n^{\beta}\right\}\le Cn^{1-\eta_2\beta}$ and

$$\begin{aligned}
& P\left\{\max_{1\le j\le N}\left|\sum_{i=1}^{n}\varepsilon_{ij}-\sum_{i=1}^{n}Z_{ij,\varepsilon}\right|>n^{\beta}\right\} \\
\le & \sum_{1\le j\le N} P\left\{\left|\sum_{i=1}^{n}\varepsilon_{ij}-\sum_{i=1}^{n}Z_{ij,\varepsilon}\right|>n^{\beta}\right\} \\
\le & Cn^{1-\eta_2\beta}\times N\le Cn^{1-\eta_2\beta+\theta}\le Cn^{-\gamma_2},
\end{aligned}$$

in which $\gamma_2>1$ as described before. Thus Borel-Cantelli Lemma implies that

$$\max_{1\le j\le N}\left|\sum_{i=1}^{n}\varepsilon_{ij}-\sum_{i=1}^{n}Z_{ij,\varepsilon}\right|=O_{\text{a.s.}}\left(n^{\beta}\right).$$

Putting together all the above proves (4.12).

Denote

$$\tilde{\xi}(x)=e_0^T\left(\mathbf{X}^T\mathbf{W}\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{W}\boldsymbol{\xi}=\sum_{k=1}^{\kappa}\tilde{\xi}_k(x),$$

where $\tilde{\xi}_k(x)=\bar{\xi}_{,k}e_0^T\left(\mathbf{X^TWX}\right)^{-1}\mathbf{X^TW}\boldsymbol{\phi}_k$ and $\boldsymbol{\phi}_k=(\phi_k(1/N),\ldots,\phi_k(N/N))^T$. Let $\tilde{\phi}_k(x)$ be the solution to the least square problem

$$\operatorname{argmin}\sum_{j=1}^{N}\left\{\phi_k(j/N)-a-b(j/N-x)\right\}^2K_h(j/N-x),$$

we have $\tilde{\phi}_k(x)=e_0^T\left(\mathbf{X^TWX}\right)^{-1}\mathbf{X^TW}\boldsymbol{\phi}_k$ and $\tilde{Z}_k(x)=\bar{Z}_{\cdot,k,\xi}\tilde{\phi}_k(x)$, $k=1,\ldots,\kappa$, similar to the definition of $\tilde{m}(x)$ and $\tilde{\xi}_k(x)$ in (4.6). Also denote $\tilde{\zeta}_k(x)=\bar{Z}_{\cdot,k,\xi}\phi_k(x)$,

$k = 1, \ldots, \kappa$ and define

$$\begin{aligned}
\widetilde{\zeta}(x) &= n^{1/2}\left\{\sum\nolimits_{k=1}^{\kappa}\phi_k^2(x)\right\}^{-1/2}\sum\nolimits_{k=1}^{\kappa}\widetilde{\zeta}_k(x) \\
&= n^{1/2}G(x,x)^{-1/2}\sum\nolimits_{k=1}^{\kappa}\widetilde{\zeta}_k(x) \qquad (4.13)
\end{aligned}$$

PROOF OF PROPOSITION 4.2: Note first fact that $\overline{Z}_{\cdot,k,\xi}$ are independent $N\left(0, n^{-1}\right)$ variables implies that $\max_{1\leq k\leq\kappa}\left|\overline{Z}_{\cdot,k,\xi}\right| = O_P\left(n^{-1/2}\right)$. By Assumption (A2), $\phi_k(x) \in c^{0,\beta}[0,1]$. Similar to Proposition 4.1, one has

$$\max_{1\leq k\leq\kappa}\left\|\phi_k(x) - \widetilde{\phi}_k(x)\right\|_\infty = U\left(h^2\right).$$

The definition of $\widetilde{\zeta}(x)$ in (4.13), together with definition of $\overline{m}(x)$ in (4.2), the strong approximation in (4.12), the above bound on $\max_{1\leq k\leq\kappa}\left|\overline{Z}_{\cdot,k,\xi}\right|$ entail that

$$\begin{aligned}
&\sup_{x\in[0,1]}\left|\overline{m}(x) - m(x) - \widetilde{\xi}(x)\right| \\
\leq\ & \max_{1\leq k\leq\kappa}\left|\overline{\xi}_{\cdot k}\right|\sup_{x\in[0,1]}\left\|\phi_k(x) - \widetilde{\phi}_k(x)\right\|_\infty \\
\leq\ & C\max_{1\leq k\leq\kappa}\left(\left|\overline{Z}_{\cdot,k,\xi}\right| + \left|\overline{\xi}_{\cdot,k} - \overline{Z}_{\cdot,k,\xi}\right|\right)h^2 \\
=\ & O_P\left(n^{-1/2}h^2 + n^{\beta-1}h^2\right) \\
=\ & o_P\left(n^{-1/2}\right).
\end{aligned}$$

While,

$$\begin{aligned}
&\sup_{x\in[0,1]}\left|\overline{m}(x) - m(x) - \widetilde{\zeta}(x)\right| \\
\leq\ & \max_{1\leq k\leq\kappa}\left|\overline{\xi}_{\cdot,k} - \overline{Z}_{\cdot,k,\xi}\right|\sup_{x\in[0,1]}\left\|\phi_k(x)\right\|_\infty \\
=\ & O_P\left(n^{\beta-1}\right) = o_P\left(n^{-1/2}\right).
\end{aligned}$$

Now for any $x \in [0,1]$, $\widetilde{\zeta}(x)$ is Gaussian with $E\widetilde{\zeta}(x) \equiv 0, E\widetilde{\zeta}^2(x) \equiv 1, x \in [0,1]$ and covariance $E\widetilde{\zeta}(x)\widetilde{\zeta}\left(x'\right)$ equal to

$$
\begin{aligned}
& n^{1/2} G(x,x)^{-1/2} n^{1/2} G\left(x',x'\right)^{-1/2} \\
& \times \operatorname{cov}\left[\left\{\sum_{k=1}^{\kappa} \overline{Z}_{\cdot,k,\xi}\phi_k(x)\right\}, \left\{\sum_{k=1}^{\kappa} \overline{Z}_{\cdot,k,\xi}\phi_k(x')\right\}\right] \\
& = G(x,x)^{-1/2} G\left(x',x'\right)^{-1/2} G\left(x,x'\right), \forall x, x' \in [0,1],
\end{aligned}
$$

so $\mathcal{L}\left\{\widetilde{\zeta}(x), x \in [0,1]\right\} = \mathcal{L}\left\{\zeta(x), x \in [0,1]\right\}$. Proposition 4.2 is proved.

PROOF OF PROPOSITION 4.3

Proof. We use C_i to denote a constant in the context. Since $G(x,x)$ is bounded, we only need to consider $\sup_{x\in[0,1]} |\widetilde{e}(x)|$. Notice that

$$
\begin{aligned}
\widetilde{e}(x) &= \mathbf{e}_0^T \left(\mathbf{X}^T\mathbf{W}\mathbf{X}\right)^{-1} \mathbf{X}^T\mathbf{W}\mathbf{e} \\
&= \mathbf{e}_0^T \operatorname{diag}\left(1, h^{-1}\right)\left\{\mathbf{D}_x^{-1} + U(h)\right\} \operatorname{diag}\left(1, h^{-1}\right) \mathbf{X}^T\mathbf{W}\mathbf{e} \\
&= Q_{N,h}(x)\{C_0 + U(h)\}, \qquad (4.14)
\end{aligned}
$$

where $Q_{N,h}(x) = N^{-1}\sum_{j=1}^{N} K_h(j/N - x)\,\bar{\varepsilon}_{\cdot,j}$. We discretize the interval $[0,1]$ and partition it into $N^* = \sqrt{N/h^3}$ subintervals $\{I_k\}$ of equal length. Let x_k be the center of I_k. For $x \in I_k$,

$$
\begin{aligned}
|Q_{N,h}(x)| &\leq |Q_{N,h}(x) - Q_{N,h}(x_k)| + |Q_{N,h}(x_k)| \\
&= |Q_{N,h}(x_k)| + N^{-1}\sum_{j=1}^{N} |\{K_h(j/N - x) - K_h(j/N - x_k)\}\bar{\varepsilon}_{\cdot,j}| \\
&= |Q_{N,h}(x_k)| + O_p\left\{(nNh)^{-1/2}\right\}. \qquad (4.15)
\end{aligned}
$$

The above is obtained because the kernel function $K(\cdot)$ is Lipchitz continuous. According to (4.14) and (4.15), we obtain

$$\sup_{x\in[0,1]} |\widetilde{e}(x)| \leq \max_{1\leq k\leq N^*} \left|Q_{N,h}(x_k)\right| + o_P(n^{-1/2}). \tag{4.16}$$

In the following, we will show that

$$\max_{1\leq k\leq N^*} \left|Q_{N,h}(x_k)\right| = o_P(n^{-1/2}). \tag{4.17}$$

Define $R_{j,h}(x) = N^{-1}K_h\,(j/N - x)\,\bar{Z}_{.,j,\varepsilon}$, where $\bar{Z}_{.,j,\varepsilon}$ is defined in Lemma A.2. According to Lemma A.2, $\{\bar{Z}_{.,j,\varepsilon}, 1 \leq j \leq N\}$ are independent and identically distributed as $N(0, 1/n)$. Then

$$\sum_{j=1}^{N} P\left(\left|\bar{Z}_{.,j,\varepsilon}\right| \leq j^{-1/2}\right) \;\leq\; \sum_{j=1}^{N} E|\bar{Z}_{.,j,\varepsilon}|^4 j^2 < \infty.$$

Based on the Borel-Cantelli Lemma, it is straightforward to show that with probability 1, for large enough j, $|\bar{Z}_{.,j,\varepsilon}| \leq j^{-1/2}$.

In the following, we only focus on large enough N such that $|\bar{Z}_{.,j,\varepsilon}| \leq N^{-1/2}$ and define

$$R_{j,h}(x) = N^{-1}K_h\,(j/N - x)\,\bar{Z}_{.,j,\varepsilon}I_{\{\bar{Z}_{.,j,\varepsilon}\leq B_N\}}. \tag{4.18}$$

It is straightforward to show that $\{R_{j,h}(x_k), 1 \leq j \leq N\}$ are independent bounded random variables with mean 0. Notice that $\left|R_{j,h}(x_k)\right| \leq C_1/(N^{3/2}h)$ and

$$\sum_{j=1}^{N} E\{R_{j,h}(x_k)\}^2 \leq C_2/(nNh).$$

Therefore, according to the Bernstein's inequality,

$$P\left(\sum_{j=1}^{N}\left|R_{j,h}(x_k)\right| \geq \eta\right) \leq 2\exp\left[-\frac{1}{2}\eta^2 \Big/ \left\{C_2/(nNh) + C_1\eta/(N^{3/2}h)\right\}\right].$$

In particular, if $\eta = \sqrt{\log N/(nNh)}$, $P\left(\sum_{j=1}^{N}\left|R_{j,h}(x_k)\right| \geq \eta\right) \longrightarrow 0$ under the Assumption (A3), which implies that

$$\sum_{j=1}^{N}\left|R_{j,h}(x_k)\right| = O_P\left\{\sqrt{\log N/(nNh)}\right\}$$

Hence,

$$\begin{aligned}
|Q_{N,h}(x_k)| &\leq \left|N^{-1}\sum_{j=1}^{N} K_h\left(j/N - x_k\right)\left(\bar{\varepsilon}_{.,j} - \bar{Z}_{.,j}\right)\right| + \left|\sum_{j=1}^{N} R_{N,h}(x_k)\right| \\
&= O_P(n^{\beta-1}) + O_P\left\{\sqrt{\log N/(nNh)}\right\} = o_P(n^{-1/2}).
\end{aligned}$$

This completes the proof.

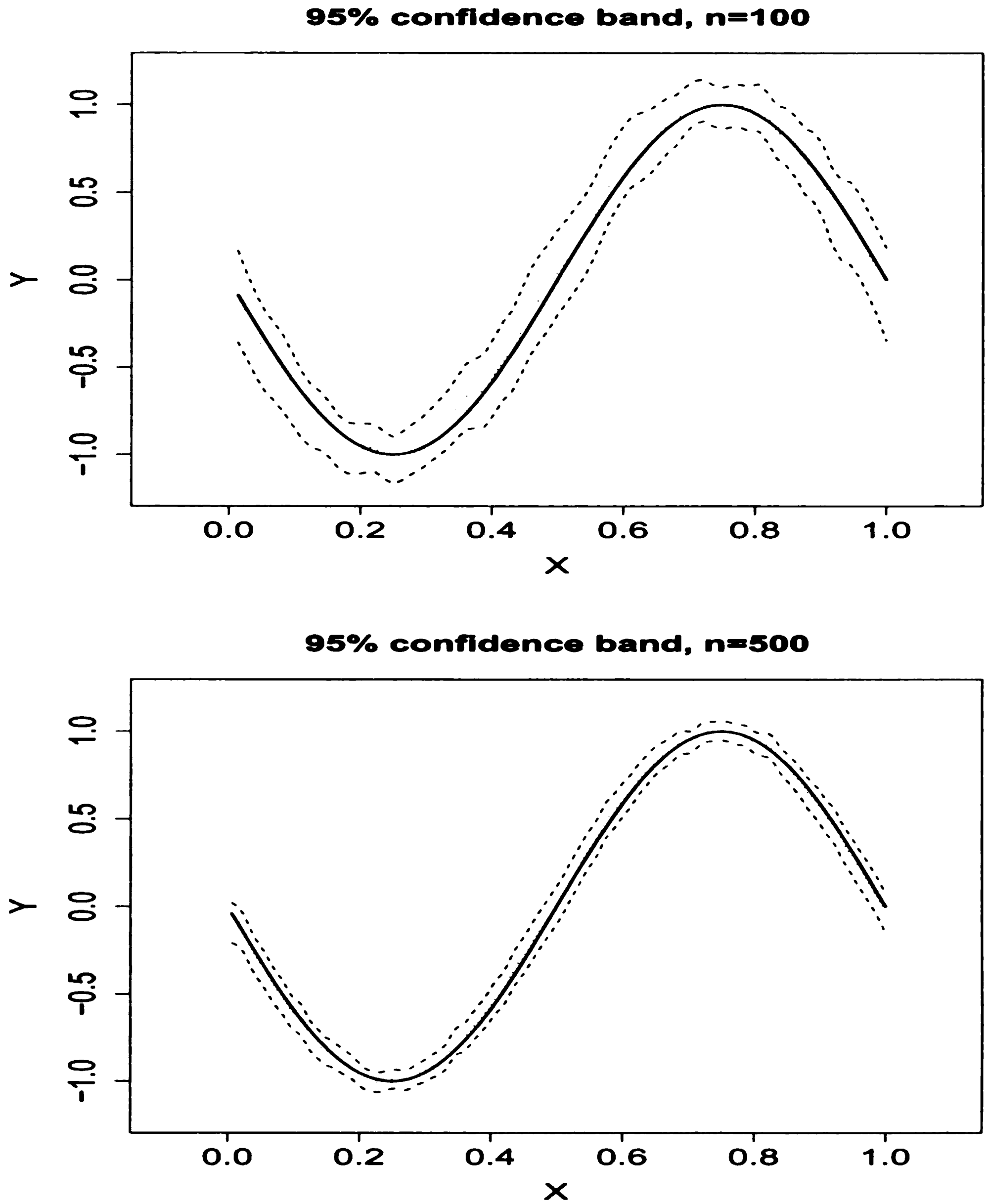


Figure 4.1: For data generated from model (4.11) (with $\sigma_0 = .5$) of different sample size n and confidence level 95%, plots of confidence bands for mean (dashed lines), the local linear estimator $\hat{m}(x)$ (dotted line), and the true function $m(x)$ (thick solid line).

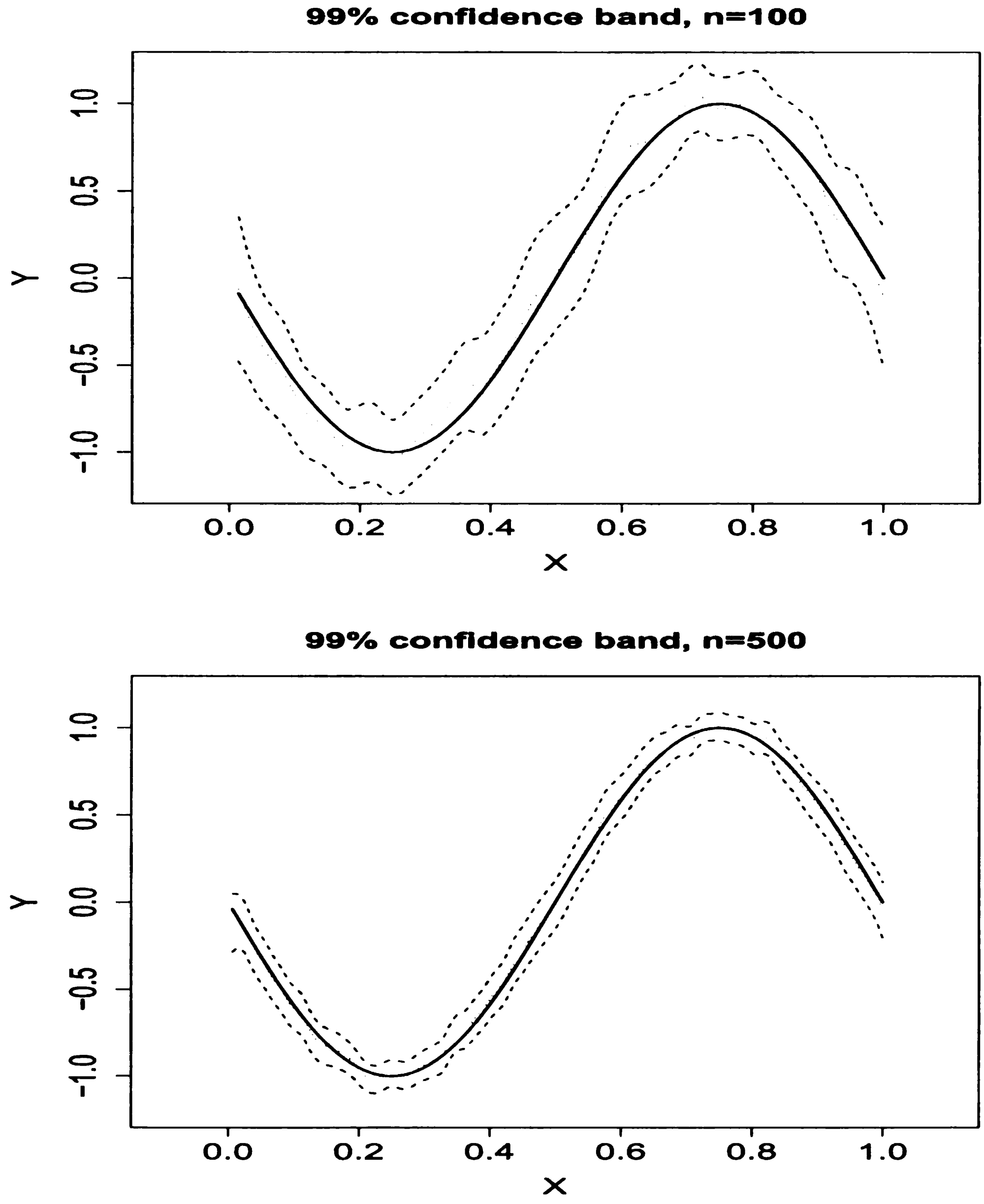
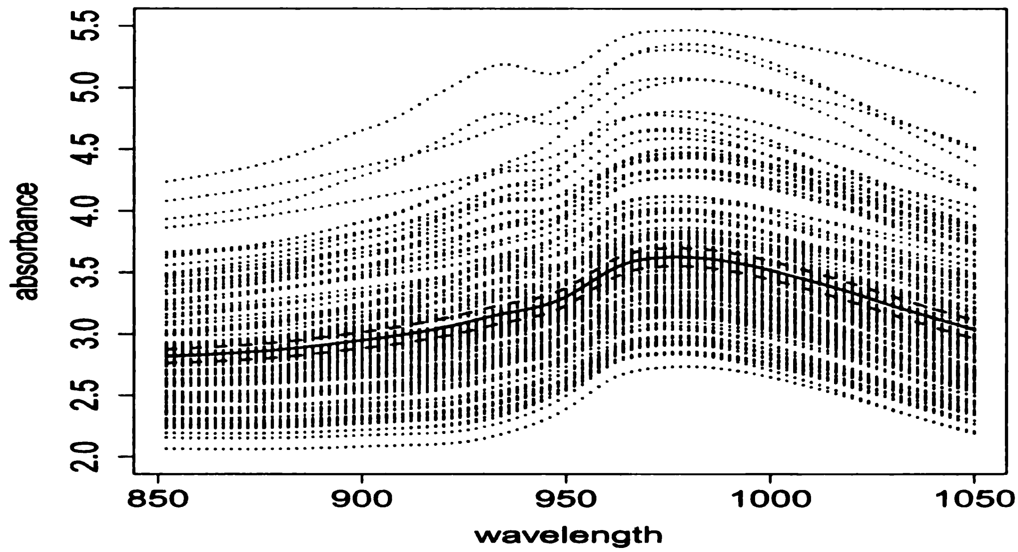


Figure 4.2: For data generated from model (4.11) (with $\sigma_0 = 1$) of different sample size n and confidence level 99%, plots of confidence bands for mean (dashed lines), the local linear estimator $\hat{m}(x)$ (dotted line), and the true function $m(x)$ (thick solid line).



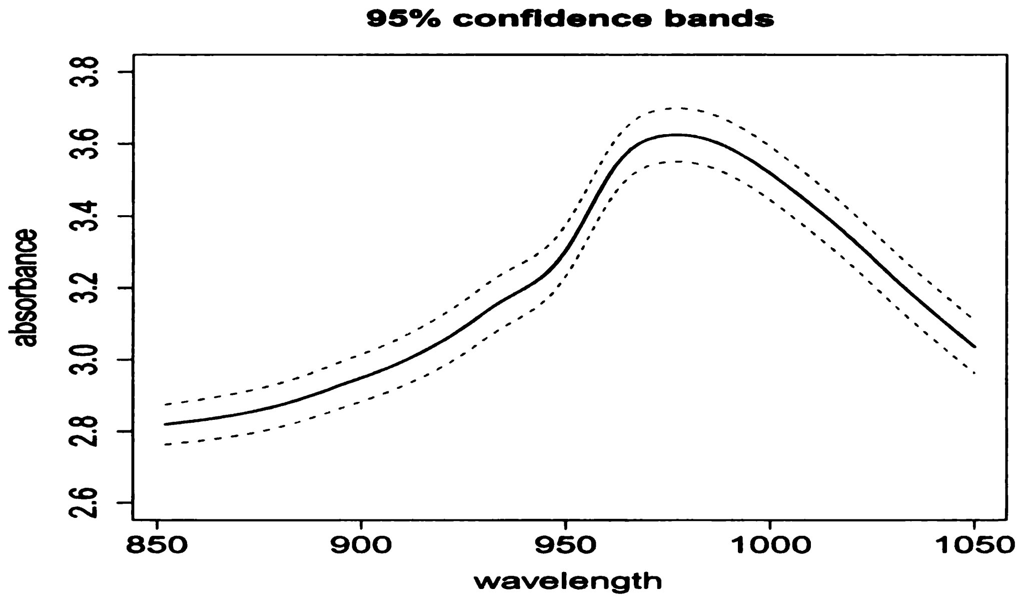


Figure 4.3: The upper plot shows the Tecator data with the 95% confidence band (dashed thick lines) for the mean estimate (thick solid line). The lower plot is the confidence band (thin dashed lines) for the mean estimate (thick solid line) in a different scale.

Chapter 5

Summary of thesis contribution

The main contributions of this thesis are the follows: the construction of simultaneous confidence bands for heteroscedastic, high dimensional and functional data.

Construction of simultaneous confidence bands has been developed slowly since it is difficult to establish asymptotic sample distribution theory for nonparametric regression estimates. In the last two decades, many statisticians have worked on the theory and applications of nonparametric simultaneous confidence bands. For the first time, we constructed confidence bands for variance function, nonlinear additive models and dense functional data.

Among all the nonparametric smoothing methods, polynomial spline smoothing has the advantage of fast computation and simple implementation, see for instance, Stone [67] and Huang [28] for the basic theory of polynomial spline smoothing, and Xue and Yang [76] for computing speed comparison of spline vs kernel smoothing. We used polynomial spline smoothing to do the nonparametric regression in the chapter 2 and chapter 3 of the thesis.

The importance of being able to detect heteroscedasticity in regression is widely recognized because of efficient inference for the regression function requires that heteroscedasticity is taken into account. In chapter 2, we proposed polynomial spline confidence bands for heteroscedastic variance function in a nonparametric regression model. It is desirable from a theoretical as well as a practical point of view to have confidence bands for polynomial spline estimators.

In chapter 3, we proposed an all new spline+spline oracally efficient estimator. For the NAAR time series models, none of any existing methods provide simultaneous confidence band for the additive components. To address this need, we proposed an all new spline+spline oracally efficient estimator that is theoretically superior as it comes with an asymptotically simultaneous confidence band for the additive component, and also computationally more expedient than any existing estimators due to the use of spline instead of kernel in all steps. The spline+spline method is asymptotically oracally efficient as the spline+kernel method of [71], but can be hundreds of times faster in terms of computing, see the comparison in Table 3.2.

Locally linear smoothing is used in chapter 4 to develop the confidence bands of mean function for dense functional data. This smoother combines the strict local nature of the data and the smooth weights of kernel smoothers. Kernel smoothers are expensive to compute ($O(n^2)$ for the whole sequence), but are visually smooth if the kernel is smooth. The confidence bands for dense functional data by locally linear smoothing is very easy to use for practitioners.

BIBLIOGRAPHY

[1] Bickel, P. J. and Rosenblatt, M. (1973). On some global measures of the deviations of density function estimates. *Ann. Statist.* **1**, 1071–1095.

[2] Bissantz, N., Claeskens, G., Holzmann, H. and Munk, A. (2009). Testing for lack of fit in inverse regression – with applications to biophotonic imaging. *J. Royal Statist. Soc. Ser. B.* **71**, 25–48.

[3] Bosq, D. (1998). Nonparametric Statistics for Stochastic Processes. Springer-Verlag, New York.

[4] Brown, L. D. and Levine, M. (2007). Variance estimation in nonparametric regression via the difference sequence method. *Ann. Statist.* **35**, 2219–2232.

[5] Chaudhuri, P. and Marron, J. S. (1999). SiZer for exploration of structures in curves. *J. Amer. Statist. Assoc.* **94**, 807–823.

[6] Chen, R. and Tsay, R. S. (1993). Nonlinear additive ARX models. *J. Amer. Statist. Assoc.* **88**, 956–967.

[7] Claeskens, G. and Van Keilegom, I. (2003). Bootstrap confidence bands for regression curves and their derivatives. *Ann. Statist.* **31**, 1852–1884.

[8] Csőrgő, M. and Révész, P. (1981). Strong Approximations in Probability and Statistics. Academic Press, New York-London.

[9] Dahl, C. M. and Levine, M. (2006). Nonparametric estimation of volatility models with serially dependent innovations. *Statist. Probab. Lett.* **76**, 2007–2016.

[10] de Boor, C. (2001). A Practical Guide to Splines, Springer-Verlag, New York.

[11] Dette, H. and Munk, A. (1998). Testing heteroscedasticity in nonparametric regression *J. Royal Statist. Society B.* **60**, 693–708.

[12] DeVore,R. and Lorentz, G. (1993). Constructive approximation : polynomials and splines approximation. Springer-Verlag, Berlin, New York.

[13] Eubank, R. L. and Speckman, P. L. (1993). Confidence bands in nonparametric regression. *J. Am. Statist. Ass.*, **88**, 1287–1301.

[14] Fan, J. and Gijbels, I. (1996). Local Polynomial Modelling and Its Applications, Chapman and Hall, London.

[15] Fan, J. and Yao, Q. (1998). Efficient estimation of conditional variance functions in stochastic regression. *Biometrika.* **85**, 645–660.

[16] Fan, J. and Zhang, W. (2000). Simultaneous confidence bands and hypothesis testing in varying-coefficient models. *Scandinavian Journal of Statistics.* **27**, 715–731.

[17] Fan, Y. and Li, Q. (2003). A kernel-based method for estimating additive partially linear models. *Statistica Sinica.* **13**, 739–762.

[18] Ferraty, F. and Vieu, P. (2006). Nonparametric Functional Data Analysis: Theory and Practice. Springer Series in Statistics, Springer: Berlin.

[19] Gantmacher, F. R. and Krein, M. G. (1960). Oszillationsmatrizen, Oszillationskerne und kleine Schwingungen mechanischer Systeme. Akademie-Verlag, Berlin.

[20] Hall, P. and Carroll, R. J. (1989). Variance function estimation in regression: the effect of estimating the mean. *J. Roy. Statist. Soc. Ser. B.* **51**, 3–14.

[21] Hall, P. and Marron, J. S. (1990). On variance estimation in nonparametric regression. *Biometrika.* **77**, 415–419.

[22] Hall, P. and Titterington, D. M. (1988). On confidence bands in nonparametric density estimation and regression. *J. Multi. Analys.* **27**, 228–254.

[23] Härdle, W. (1989). Asymptotic maximal deviation of M-smoothers. *J. Multi. Analys.* **29**, 163–179.

[24] Härdle, W., Hlávka, Z. and Klinke, S. (2000). *XploRe Application Guide.* Springer-Verlag, Berlin.

[25] Härdle, W. and Marron, J. S. (1991). Bootstrap simultaneous error bars for nonparametric regression. *Ann. Statist.* **19**, 778–796.

[26] Hastie, T. J. and Tibshirani, R. J. (1990). Generalized additive models. Chapman and Hall, London.

[27] Horowitz, J. and Mammen, E. (2004). Nonparametric estimation of an additive model with a link function. *Ann. Statist.* **32**, 2412–2443.

[28] Huang, J. Z. (2003). Local asymptotics for polynomial spline regression. *Ann. Statist.* **31**, 1600–1635.

[29] Huang, J. Z. and Yang, L. (2004). Identification of nonlinear additive autoregression models. *J. R. Stat. Soc. Ser. B.* **66**, 463–477.

[30] Huang, X., Wang, L., Yang, L. and Kravchenko, A. N. (2008). Management practice effects on relationships of grain yields with topography and precipitation. *Agronomy Journal.* **100**, 1463–1471.

[31] Izem, R. and Marron, J. S. (2007). Analysis of nonlinear modes of variation for functional data. *Elect. J. Statist.* **1**, 641–676.

[32] James, G. M., Hastie, T. and Sugar, C. (2000). Principal Component Models for Sparse Functional Data. *Biometrika.* **87**, 587–602.

[33] James, G. M. (2002). Generalized linear models with functional predictors. *J. R. Stat. Soc. Ser. B.* **64**, 411–432.

[34] James, G. M. and Silverman, B. W. (2005). Functional adaptive model estimation. *J. Amer. Statist. Assoc.* **100**, 565–576.

[35] James, G. M. and Sugar, C. A. (2003). Clustering for sparsely sampled functional data. *J. Amer. Statist. Assoc.* **98**, 397–408.

[36] Kauermann, G., Krivobokova, T. and Fahrmeir, L. (2009). Some asymptotic results on generalized penalized spline smoothing, *J. R. Stat. Soc. Ser. B.* **71** 487–503.

[37] Krivobokova, T. and Kauermann, G. (2007). A note on penalized spline smoothing with correlated errors. *J. Amer. Statist. Assoc.* **102**, 328–1337.

[38] Leadbetter, M. R., Lindgren, G. and Rootzén, H. (1983). Extremes and related properties of random sequences and processes. Springer-Verlag, New York.

[39] Li, Y. and Hsing, T. (2007). On rates of convergence in functional linear regression. *J. Multi. Analys.* **98**, 1782–1804.

[40] Li, Y. and Hsing, T. (2009). Uniform convergence rates for nonparametric regression and principal component analysis in functional/longitudinal data. *Ann. Statist.* In press.

[41] Liang, H., Thurston, S., Ruppert, D., Apanasovich, T. and Hauser, R. (2008). Additive partial linear models with measurement errors. *Biometrika.* **95**, 667–678.

[42] Linton, O. B. (1997). Efficient estimation of additive nonparametric regression models, *Biometrika.* **84**, 469–473.

[43] Linton, O. B. and Nielsen, J. P. (1995). A kernel method of estimating structured nonparametric regression based on marginal integration, *Biometrika.* **82**, 93–101.

[44] Lu, Z., Lundervold, A., Tjøstheim, D. and Yao, Q. (2007). Exploring spatial nonlinearity using additive approximation. *Bernoulli.* **13**, 447–472.

[45] Ma, S., Yang,L. and Carroll, R. J. (2010). Simultaneous confidence band for sparse longitudinal regression curve. Manuscript.

[46] Mack, Y. P. and Silverman, B. W. (1982). Weak and strong uniform consistency of kernel regression estimates. *Z. Wahrscheinlichkeitstheorie verm. Gebiete* **61**, 405–415.

[47] Mammen, E., Linton, O. B. and Nielsen, J. P. (1999). The existence and asymptotic properties of a backfitting projection algorithm under weak conditions. *Ann. Statist.* **27**, 1443–1490.

[48] Martins-Filho, C. and Yang, K. (2007). Finite sample performance of kernel-based regression methods for non-parametric additive models under common bandwidth selection criterion. *J. Nonparametr. Stat.* **19**, 23–62.

[49] Morris, J. S. and Carroll, R. J. (2006). Wavelet-based functional mixed models. *J. R. Stat. Soc. Ser. B.* **68**, 179–199.

[50] Müller, H. G and Stadtmüller, U. (1997). Variable bandwidth kernel estimators of regression curves. *Ann. Statist.* **15**, 182–201.

[51] Müller, H. G. and Stadtmüller, U. (2005). Generalized functional linear models. *Ann. Statist.* **33**, 774–805.

[52] Müller, H. G., Stadtmüller, U. and Yao, F. (2006). Functional variance processes. *J. Amer. Statist. Assoc.* **101**, 1007–1018.

[53] Müller, H. G. and Yao, F. (2008). Functional additive models. *J. Amer. Statist. Assoc.* **103**, 1534–1544.

[54] Neumann, M. H. and Kreiss, J. P. (1998). Regression-type inference in nonparametric autoregression. *Ann. Statist.* **26**, 1570–1613.

[55] Nielsen, J. P. and Sperlich, S. (2005). Smooth backfitting in practice. *J. R. Stat. Soc. Ser. B.* **67**, 43–61.

[56] Ramsay, J. O. and Silverman, B. W. (2005). Functional Data Analysis. Second Edition. Springer Series in Statistics. Springer: New York.

[57] Roca-Pardiñas, J., Cadarso-Suárez, C. and Gonzalez-Manteiga, W. (2005). Testing for interactions in generalized additive models: Application to $SO2$ pollution data. *Stat. Comput.* **15**, 289–299.

[58] Roca-Pardiñas, J., Cadarso-Suárez, C., Tahoces, PG. and Lado, MJ. (2008). Assessing continuous bivariate effects among different groups through nonparametric regression models: An application to breast cancer detection. *Comput. Statist. Data. Analysis.* **52**, 1958–1970.

[59] Rosenblatt, M. (1976). On the maximal deviation of k-dimensional density estimates. *Ann. Prob.* **4**, 1009–1015.

[60] Ruppert, D., Wand, M. P. and Carroll, R. J. (2003). Semiparametric regression. Cambridge University Press, Cambridge.

[61] Ruppert, D., Wand, M. P., Holst, U. and Hössjer, O. (1997). Local polynomial variance-function estimation. *Technometrics.* **39**, 262–273.

[62] Silverman, B. W. (1986). Density estimation for statistics and data analysis. Chapman and Hall, London.

[63] Song, Q. and Yang, L. (2009). Spline confidence bands for variance function. *J. Nonparametri. Statis.* **21**, 589–609.

[64] Song, Q. and Yang, L. (2010). Oracally efficient spline smoothing of nonlinear additive autoregression models with simultaneous confidence bands. *J. Multi. Analys.* In press.

[65] Sperlich, S., Tjøstheim, D. and Yang, L. (2002). Nonparametric estimation and testing of interaction in additive models. *Econometric Theory.* **18**, 197–251.

[66] Stone, C. J. (1985). Additive regression and other nonparametric models. *Ann. Statist.* **13**, 689–705.

[67] Stone, C. J. (1994). The use of polynomial splines and their tensor products in multivariate function estimation. *Ann. Statist.* **22**, 118–184.

[68] Tjøstheim, D. and Auestad, B. (1994). Nonparametric identification of nonlinear time series: projections. *J. Amer. Statist. Assoc.* **89**, 1398–1409.

[69] Tusnády, G. (1977). A remark on the approximation of the sample df in the multidimensional case. *Periodica Mathematica Hungarica.* **8**, 53–55.

[70] Wang, J. and Yang, L. (2009). Polynomial spline confidence bands for regression curves. *Statistica Sinica.* **19**, 325–342.

[71] Wang, L. and Yang, L. (2007). Spline-backfitted kernel smoothing of nonlinear additive autoregression model. *Ann. Statist.* **35**, 2474–2503.

[72] Wang, L. and Yang, L. (2010). Simultaneous spline confidence bands for time series prediction function. *J. Nonparametri. Stat.* in press.

[73] Wang, N., Carroll, R. J. and Lin, X. (2005). Efficient semiparametric marginal estimation for longitudinal/clustered data. *J. Amer. Statist. Assoc.* **100**, 147–157.

[74] Wu, W. and Zhao, Z. (2007). Inference of trends in time series. *J. R. Stat. Soc. Ser. B.* **69**, 391–410.

[75] Xia, Y. (1998). Bias-corrected confidence bands in nonparametric regression. *J. Roy. Statist. Soc. Ser. B.* **60**, 797–811.

[76] Xue, L. and Yang, L. (2006). Additive coefficient modeling via polynomial spline. *Statistica Sinica.* **16**, 1423–1446.

[77] Yang, L. (2008). Confidence band for additive regression model. *Journal of Data Science* **6**, 207–217.

[78] Yang, L., Park, B.U., Xue, L. and Härdle, W. (2006). Estimation and testing of varying coefficients in additive models with marginal integration. *J. Amer. Statist. Assoc¿* **101**, 1212–1227.

[79] Yao, F. and Lee, T. C. M. (2006). Penalized spline models for functional principal component analysis. *J. Roy. Statist. Soc. Ser. B.* **68**, 3–25.

[80] Yao, F., Müller, H. G. and Wang, J. L. (2005a). Functional linear regression analysis for longitudinal data. *Ann. Statist.* **33**, 2873–2903.

[81] Yao, F., Müller, H. G. and Wang, J. L. (2005b). Functional data analysis for sparse longitudinal data. *J. Amer. Statist. Assoc.* **100**, 577–590.

[82] Yao, F. (2007). Asymptotic distributions of nonparametric regression estimators for longitudinal or functional data. *J. Multi. Analys.* **98**, 40–56.

[83] Yao, Q. and Tong, H. (2000). Nonparametric estimation of ratios of noise to signal in stochastic regression. *Statistica Sinica.* **10**, 751–770.

[84] Zhang, F. (1999). Matrix theory: basic results and techniques. Springer-Verlag, New York.

[85] Zhang, J. T. and Chen, J. (2007). Statistical inferences for functional data. *Ann. Statist.* **35**, 1052–1079.

[86] Zhao, X., Marron, J. S. and Wells, M. T. (2004). The functional data analysis view of longitudinal data. *Statistica Sinica.* **14**, 789–808.

[87] Zhao, Z. and Wu, W. (2008). Confidence bands in nonparametric time series regression. *Ann. Statist.* **36**, 1854–1878.

[88] Zhou, L., Huang, J. and Carroll, R. J. (2008). Joint modelling of paired sparse functional data using principal components. *Biometrika.* **95**, 601–619.

[89] Zhou, S., Shen, X. and Wolfe, D. A. (1998). Local asymptotics of regression splines and confidence regions. *Ann. Statist.* **26**, 1760–1782.