This is to certify that the
dissertation entitled

Non- and semiparametric modeling of financial and macro-economic time series

presented by

Rong Liu

has been accepted towards fulfillment
of the requirements for the

Ph.D. degree in Statistics

Major Professor's Signature

8/11/09

Date

NON- AND SEMIPARAMETRIC MODELING OF FINANCIAL AND MACRO-ECONOMIC TIME SERIES

By

Rong Liu

A DISSERTATION

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

DOCTOR OF PHILOSOPHY

Statistics

2009

ABSTRACT

NON- AND SEMIPARAMETRIC MODELING OF FINANCIAL AND MACRO-ECONOMIC TIME SERIES

By

Rong Liu

Nonlinear time series analysis has gained much attention in recent years due primarily to the fact linear time series models have encountered various limitations in real applications and the development in nonparametric regression has established a solid foundation for nonlinear time series analysis. For example, the effect of technology on the economic growth, volatility of exchange returns, which follow nonlinear instead of simple linear prediction formulas. Effective tools for extracting information from such complex regression data have to be nonparametric in nature.

A smooth kernel estimator is proposed for multivariate cumulative distribution function in Chapter 2, extending the work on Yamato (1973) on univariate distribution function estimation. Under assumptions of strict stationarity and geometrically strong mixing, we establish that the proposed estimator follows the same pointwise asymptotically normal distribution of the empirical cdf, while the new estimator is a smooth instead of a step function as the empirical cdf. We also show that under stronger assumptions the smooth kernel estimator has asymptotically smaller mean integrated squared error than the empirical cdf, and converges to the true cdf uniformly almost surely at a rate of $\left(n^{-1/2}\log n\right)$. Simulated examples are provided to illustrate the theoretical properties. Using the smooth estimator, survival curves are given for real data applications.

"Curse of dimensionality" is a significant obstacle in high dimensional time series analysis, see Fan and Yao (2003). Several high dimensional data analysis techniques have been proposed to deal with this problem and Xia, Tong, Li and Zhu (2002) pointed out that there are essentially two approaches: function approximation and dimension reduction. GARCH model, Additive Coefficient Model (ACM) and Generalized Additive model (GAM) are good examples to represent these two approaches.

In Chapter 3, a cubic spline regression procedure is proposed to estimate the unknowns in the semiparametric GARCH model that is intuitively appealing due to its simplicity, and as such, can be used by non experts. The theoretical properties of the procedure is the same as the kernel procedure in Yang (2006), and simulated and real data examples show that the numerical performance is also comparable to the kernel method. The new method is computationally much more efficient and very useful for analyzing financial time series data.

In Chapter 4, a spline-backfitted kernel estimator is proposed for estimating the unknown component functions $m_{\alpha l}$ based on a geometrically strong mixing sample following model (1.3.1) under minimal smoothness assumptions. The idea is to employ one step backfitting after the spline pilot estimators, and then follow up with kernel smoothing, which combines the fast computing of polynomial spline smoothing and the good asymptotic property of kernel smoothing. Thus, the spline-backfitted kernel estimator is both computationally expedient for analyzing very high dimensional time series, and theoretically reliable to make inference on the component functions with confidence.

In Chapter 5, a spline-backfitted kernel (SBK) estimator is proposed for the Generalized Additive Model time series data with oracle efficiency. It is both computationally expedient and theoretically reliable, and simulation evidence strongly corroborates the asymptotic theory.

ACKNOWLEDGMENTS

I would like to thank many people who have helped me on the path towards this dissertation. First and foremost, I would like to express my gratitude to my advisor, Professor Lijian Yang. I could never have reached the heights or explored the depths without his generous help, unbreakable support and patient guidance.

I also wish to express my gratitude to my dissertation committee, Professor Dennis Gilliland, Professor Lyudmila Sakhanenko, Professor Emma Iglesias, Professor Yiming Xiao, Professor Richard Baillie 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. And special thanks are given to Professor James Stapleton, Professor Connie Page and Professor Raoul LePage for their numerous help, constant support and encouragement.

Thanks to the graduate school and the Department of Statistics who provided me with the Dissertation Completion Fellowship (2009), Summer Support Fellowship (2008) and Stapleton Fellowship for working on this dissertation. This dissertation is also supported in part by NSF awards DMS 0405330 and 0706518.

Last but not least, I would like to thank four of my academic sisters: Dr. Jing Wang, Dr. Li Wang, Qiongxia Song and Shujie Ma for their generous help.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

Introduction

1.1 Nonlinear Time Series Prediction Model

Nonlinear time series analysis has gained much attention in recent years due primarily to the fact linear time series models have encountered various limitations in real applications and the development in nonparametric regression has established a solid foundation for nonlinear time series analysis. For example, the effect of technology on the economic growth, volatility of exchange returns, which follow nonlinear instead of simple linear prediction formulas. Effective tools for extracting information from such complex regression data have to be nonparametric in nature. I view this line of research as developing theory that is motivated and influenced by applications.

A typical nonparametric problem in time series analysis is the classical decomposition of a realization of a time series into a slowly changing function known as a "trend component", or simply trend, a periodic function referred to as a "seasonal component", and finally a "random noise component", which in terms of the regression theory should be called the time series of residuals. In time series analysis smoothing problems occur of course in the spectral domain when we want to estimate the spectral density, e.g. for model fitting. In the time domain nonparametric prediction is one of the fields where smoothing methods are intensively used.

Two very popular forms of nonparametric regression are kernel/local polynomial type and spline type smoothing. In this work, the polynomial spline smoothing is extensively studied for nonlinear time series. The greatest advantages of spline smoothing, as pointed out in Huang and Yang (2004), Xue and Yang (2006 b) are its simplicity and fast computation. But spline smoothing also has disadvantages, such as no limiting distribution. So the combination for kernel/local polynomial and spline smoothing is studied in Chapters 4 and 5.

"Curse of dimensionality" is a significant obstacle in high dimensional time series analysis, see Fan and Yao (2003). Several high dimensional data analysis techniques have been proposed to deal with this problem and Xia, Tong, Li and Zhu (2002) pointed out that there are essentially two approaches: function approximation and dimension reduction. GARCH model and Generalized Additive model (GAM) are good examples to represent these two approaches.

1.2 Semiparametric GARCH Model

In the study of many financial time series such as foreign exchange returns, it has been a known fact that the return itself can not be predicted. It is the forecasting of the returns' volatility that is of special interests. Empirical evidences had led to the understanding that for such series, the volatility often depends on infinitely many past returns with diminishing weights. The GARCH(p, q) model of Bollerslev (1986), for example, allows the volatility function to depend on all past observations, with geometrically decaying rate.

As a special case, the GARCH$(1,1)$ model describes a process $\{Y_t\}_{t=-\infty}^{\infty}$ of the form $Y_t = \sigma_t \xi_t, t \in \mathbf{Z} = \{..., -2, -1, 0, 1, 2, ...,\}$ where the innovations $\{\xi_t\}_{t \in \mathbf{Z}}$ are i.i.d random variables satisfying $E(\xi_t) = 0, E(\xi_t^2) = 1$, and $\{\sigma_t^2\}_{t=-\infty}^{\infty}$ denotes the conditional volatility series $\sigma_t^2 = \text{var}(Y_t | Y_{t-1}, Y_{t-2}, ...)$ i.e., for some $w, \beta_0, \alpha_0 > 0, \alpha_0 + \beta_0 < 1$,

$$\sigma_t^2 = w + \beta_0 \sum\nolimits_{j=1}^{\infty} \alpha_0^{j-1} Y_{t-j}^2, t \in \mathbf{Z}. \tag{1.2.1}$$

Engle and Ng (1993) and Glosten, Jaganathan and Runkle (1993), Hentschel (1995), Duan (1997), Hafner and Herwartz (2006), Hafner (2008) had examined various useful extensions of model (1.2.1), mostly providing empirical evidence without establishing asymptotic results. For related theoretical works on GARCH model, see Peng and Yao (2003), Sun and Stengos (2006) and Chan, Deng, Peng and Xia (2007).

In recent years, there has been a surge of interests in applying nonparametric smoothing theory to volatility estimation, as in Yang, Härdle and Nielsen (1999), Dahl and Levine (2006), Levine (2006), Brown and Levine (2007). In particular, Hafner (1998) had proposed iterative algorithm for nonparametric GARCH model of the form

$$\sigma_t^2 = \sum_{j=1}^{\infty} \alpha_0^{j-1} m\left(Y_{t-j}\right), t \in \mathbf{Z}, 0 < \alpha_0 < 1$$

with unknown parameter α_0 and unknown smooth news impact function m, without asymptotic theory. A truncated version of the above nonparametric model was studied in Yang (2000), Yang (2002) with asymptotic results, yet it failed to capture the dependence of σ_t^2 on infinitely many past Y_{t-j}. In Linton and Mammen (2005), the more general model

$$\sigma_t^2 = w + \sum_{j=1}^{\infty} \psi_j(\alpha_0) m\left(Y_{t-j}\right), t \in \mathbb{Z}, \psi_j(\alpha_0) \geq 0, \sum_{j=1}^{\infty} \psi_j(\alpha_0) < \infty$$

was discussed and kernel estimator was proposed.

As an alternative, Yang (2006) formulated a class of semiparametric GARCH model, which includes the following as a special case

$$\sigma_t^2 = m\left\{\sum_{j=1}^{\infty} \alpha_0^{j-1} Y_{t-j}^2\right\}, t \in \mathbf{Z}, 0 < \alpha_0 < 1 \quad (1.2.2)$$

with unknown parameter α_0 and unknown smooth link function m, and proposed kernel estimation method for α_0 and m, with satisfactory theoretical properties and numerical accuracy in simulation and applications to real data sets. Like all the aforementioned works based on kernel smoothing, the algorithm in Yang (2006) is extremely slow due to the intensive computation of solving as many least squares problems as the sample size. The average computing time for the local linear based algorithm in Yang (2006) is contained in Table 3 for sample sizes n from 400 to 3200, and one can see that it grows at the rate of n^2. At $n = 3200$, which is a moderate sample size for financial time series, the estimation of unknown parameter α_0 takes 5 hours. The method of Yang (2006) is therefore not appealing for practical use.

In Chapter 3, cubic spline regression procedure is proposed to estimate the unknowns in the semiparametric GARCH model that is intuitively appealing due to its simplicity, and as such, can be used by non experts. The theoretical properties of the procedure is the same as the kernel procedure in Yang (2006), and simulated and real data examples show that the numerical performance is also comparable to the kernel method. The new method is computationally much more efficient and very useful for analyzing financial time series data.

1.3 Additive Coefficient Model (ACM)

Regression analysis has been widely used in econometrics studies, for instance, the estimation of production/cost function. Typical parametric regression models presume that their regression functions follow a pre-determined form with finitely many unknown parameters. Nonparametric models, on the other hand, impose less stringent assumptions on the regression functions, but pay for its flexibility the price of "curse of dimensionality". Structured models offer a sensible compromise between parametric simplicity and nonparametric flexibility, see, for example, Sperlich, Tjøstheim and Yang (2002) for additive interaction modelling for the production function of Wisconsin farms and Rodríguez-Póo, Sperlich and Vieu (2003) for a general framework of separable models. Recently Xue and Yang (2006a,b) have proposed additive coefficient model that allows a response variable Y to depend linearly on some regressors, with coefficients as smooth additive functions of other predictors, called tuning variables. Specifically

$$E\left(Y|\mathbf{X},\mathbf{T}\right) \equiv m\left(\mathbf{X},\mathbf{T}\right) \equiv \sum_{l=1}^{d_1} m_l\left(\mathbf{X}\right) T_l,\ m_l\left(\mathbf{X}\right) = m_{0l} + \sum_{\alpha=1}^{d_2} m_{\alpha l}\left(X_\alpha\right), 1 \leq l \leq d_1 \tag{1.3.1}$$

in which the predictor vector $(\mathbf{X}, \mathbf{T})$ consists of the tuning variables $\mathbf{X} = \left(X_1, ..., X_{d_2}\right)^T \in R^{d_2}$ and linear predictors $\mathbf{T} = \left(T_1, \ldots, T_{d_1}\right)^T \in R^{d_1}$. The functional coefficient model of Chen and Tsay (1993b) corresponds to the case $d_2 = 1$, the varying coefficient model of Hastie and Tibshirani (1993) corresponds to the case $d_2 = d_1$ and for each $l = 1, ..., d_1$ there is only one single significant $m_{\alpha l}$ with $\alpha = l$. Also included as special cases of model (1.3.1) are the additive model of Hastie and Tibshirani (1990), Chen and Tsay (1993a), and the multivariate linear regression model, see Xue and Yang (2006a) for detailed discussion.

In Chapter 4, a spline-backfitted kernel estimator is proposed for estimating the unknown component functions $m_{\alpha l}$ based on a geometrically strong mixing sample following model (1.3.1) under minimal smoothness assumptions. The idea is to employ one step backfitting after the spline pilot estimators, and then follow up with kernel smoothing, which combines the fast computing of polynomial spline smoothing and the good asymptotic property of kernel smoothing. Thus, the spline-backfitted kernel estimator is both computationally expedient for analyzing very high dimensional time series, and theoretically reliable to make inference on the component functions with confidence.

1.4 Generalized Additive Model(GAM)

One unavoidable issue in high dimensional time series smoothing is the "curse of dimensionality", which refers to the poor convergence rate of nonparametric estimation of general multivariate functions. One solution is autoregression in the form of additive model introduced by Hastie and Tibshirani (1990)

$$E(Y|\mathbf{X}) = g^{-1}\{m(\mathbf{X})\}, m(\mathbf{X}) = c + \sum\nolimits_{\alpha=1}^{d} m_{\alpha}(X_{\alpha}), \tag{1.4.1}$$

for the predictor vector $\mathbf{X} = (X_1, ..., X_d)^T$, and one observes a length n realization of a $(d+1)$-dimensional strictly stationary process $\left\{Y_i, \mathbf{X}_i^T\right\}_{i=1}^n = \{Y_i, X_{i1}, ..., X_{id}\}_{i=1}^n$ which follows (1.4.1). Typically, one denotes the link function g^{-1} as b' and assumes that the conditional variance function is $\sigma^2(\mathbf{X}) = \text{var}(Y|\mathbf{X}) = a(\phi)\, b''\{m(\mathbf{X})\}$, in which $a(\phi)$ is a nuisance parameter that quantifies overdispersion. One can also write the usual regression form

$$Y_i = g^{-1}\{m(\mathbf{X}_i)\} + \sigma(\mathbf{X}_i)\varepsilon_i = b'\{m(\mathbf{X}_i)\} + \sigma(\mathbf{X}_i)\varepsilon_i \tag{1.4.2}$$

for some conditional white noise ε_i that satisfy $E(\varepsilon_i|\mathbf{X}_i) = 0$, $E\left(\varepsilon_i^2|\mathbf{X}_i\right) = 1$. The regression function m takes the form in (1.4.1), and satisfies the identifibility conditions that

$$E\{m_{\alpha}(X_{\alpha})\} = 0, 1 \le l \le d_1, 1 \le \alpha \le d_2 \tag{1.4.3}$$

ensuring the unique additive representations of $m(\mathbf{x}) = m_{0l} + \sum_{\alpha=1}^{d} m_{\alpha}(x_{\alpha})$. As in most works on nonparametric smoothing, estimation of the functions $\{m_{\alpha}(x_{\alpha})\}_{\alpha=1}^{d}$ is conducted on compact sets. Without lose of generality, let the compact set be $\chi = [0,1]^d$.

In Chapter 5, we propose spline-backfitted kernel (SBK) estimator for the GAM time series data with oracle efficiency. It is both computationally expedient and theoretically reliable, thus usable for analyzing very high-dimensional time series and inference can be made on component functions with confidence. Simulation evidence strongly corroborates with the asymptotic theory.

1.5 Polynomial Spline Smoothing

Let $\{X_i, Y_i\}_{i=1}^n$ be a strictly stationary process. Assume that X_i, $i = 1, ..., n$, are supported on a compact interval $[a, b]$. Polynomial splines begin by choosing a set of knots (typically, much smaller than the number of data points n), and a set of basis functions spanning a set of piecewise polynomials satisfying continuity and smoothness constraints.

To be specific, divide $[a, b]$ into $(N+1)$ subintervals $J_j = [t_j, t_{j+1})$, $j = 0, ..., N-1, J_N = [t_N, b]$, where $T := \{t_j\}_{j=1}^N$ is a sequence of equally-spaced points, called interior knots, given as

$$t_{1-k} = ... = t_{-1} = t_0 = a < t_1 < ... < t_N < b = t_{N+1} = ... = t_{N+k},$$

in which $t_j = a + jh$, $j = 0, 1, ..., N+1, h = 1/(N+1)$ is the distance between neighboring knots. Denote by

$$C^{(k)}[a,b] = \{m|\text{the } kth \text{ order derivative of } m \text{ is continuous on } [a,b]\} \tag{1.5.1}$$

and $G^{(k-2)} = G^{(k-2)}[a,b]$ the space of all $C^{(k-2)}[a,b]$ functions that are polynomials of degree $k-1$ on each interval. 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 (2001), i.e.

$$B_{j,k}(u) = \frac{(u-t_j) B_{j,k-1}(u)}{t_{j+k-1} - t_j} - \frac{(u - t_{j+k}) B_{j+1,k-1}(u)}{t_{j+k} - t_{j+1}}, 1-k \le j \le N, \tag{1.5.2}$$

for $k > 1$, with

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

In Chapters 3, 4 and 5, spline smoothing is applied under different conditions.

CHAPTER 2

Kernel estimation of multivariate cumulative distribution

2.1 Introduction

This chapter is based on Liu and Yang (2008). The estimation of probability density functions (pdf's) and cumulative distribution functions (cdf's) occupy a central place in applied data analysis in the social sciences. While many statisticians and econometricians are familiar with various smooth nonparametric estimators of pdf's, the smooth estimation of cdf's has not been investigated as much, see Li and Racine (2007) sections 1.4 and 1.5. To properly define the problem, let $\left\{\mathbf{X}_i = (X_{i1}, ..., X_{id})^T\right\}_{i=1}^n$ be a geometrically α-mixing and strictly stationary sequence of d-dimensional variables, with a common probability density function $f \in C^{(p+1)}\left(R^d\right)$ and cumulative distribution function $F \in C^{(p+d+1)}\left(R^d\right)$, in which p is an odd integer. Traditionally, F is estimated by the empirical cumulative distribution function $\tilde{F}(\mathbf{x}) = n^{-1}\sum_{i=1}^n I\left\{\mathbf{X}_i \leq \mathbf{x}\right\}$, whose theoretical properties have been well known. One obvious drawback of $\tilde{F}$ is that it is a step function even when the true cdf F is smooth.

Yamato (1973) proposed a smooth estimator of F by integrating a kernel density estimator of the density f. To be precise, define the following kernel estimator of F

$$\hat{F}(\mathbf{x}) = \hat{F}_n(\mathbf{x}) = \int_{-\infty}^{\mathbf{x}} \hat{f}(\mathbf{u})\, d\mathbf{u} = n^{-1}\sum_{i=1}^n \int_{-\infty}^{\mathbf{x}} K_{\mathbf{h}}(\mathbf{X}_i - \mathbf{u})\, d\mathbf{u}, \forall \mathbf{x} \in R^d \tag{2.1.1}$$

where $\hat{f}(\mathbf{u})$ is the standard d-dimensional kernel density estimator (kde) of $f(\mathbf{u})$ (see Bickel and Rosenblatt, 1973)

$$\hat{f}(\mathbf{u}) = n^{-1}\sum_{i=1}^{n} K_{\mathbf{h}}(\mathbf{X}_i - \mathbf{u}), K_{\mathbf{h}}(\mathbf{u}) = \prod_{\alpha=1}^{d} \frac{1}{h_\alpha} K\left(\frac{u_\alpha}{h_\alpha}\right), \mathbf{u} = (u_1, \ldots, u_d)^T$$

in which $\mathbf{h} = (h_1, ..., h_d)^T$ are positive numbers depending on the sample size n, called bandwidths.

Theoretical properties of $\hat{F}(\mathbf{x})$ as an estimator of the unknown true distribution function $F(\mathbf{x})$ have been investigated by several authors for the case of $d = 1$ and under i.i.d assumptions, see for example Yamato (1973), Reiss (1981), Falk (1983) and more recently Cheng and Peng (2002). For feasible econometric applications of univariate kernel estimation of cumulative distribution function, such as to the testing of stochastic dominance, see Li and Racine (2007), page 23, and the references therein.

In this chapter, we examine under a strong mixing assumption and for arbitrary dimension d, the local property of $\hat{F}(\mathbf{x})$ in terms of pointwise asymptotic distribution and its global property in terms of mean integrated squared error (MISE) and maximal deviation. We have proved that the smooth estimator $\hat{F}(\mathbf{x})$ behaves asymptotically the same as the empirical cdf $\tilde{F}(\mathbf{x})$ at any point $\mathbf{x}$, have obtained its asymptotic mean integrated squared error (AMISE) and have established its uniform almost sure convergence rate.

The rest of the chapter is organized as the following. In Section 2.2, we give Theorems 2.2.1, 2.2.2 and 2.2.3, the main results on pointwise, mean integrated squared and uniform asymptotics. In Section 2.3, we describe a data-driven rule to select the asymptotically optimal bandwidths $\mathbf{h}$, which makes the MISE of $\hat{F}$ asymptotically smaller than that of the empirical cdf $\tilde{F}$ according to Theorem 2.3.2, another compelling reason that $\hat{F}$ is preferable over $\tilde{F}$ other than smoothness. In Section 2.4, we present Monte Carlo evidence that corroborates the theory and illustrates the use of $\hat{F}$ with two real data examples. The first real data example illustrates the stochastic dependence of GDP growth rate on unemployment growth rate in the US economy. Second example shows that gold and silver are substitute goods and their prices are strongly associated. All technical proofs are in the Appendix.

2.2 Asymptotic Results

Throughout this chapter, we denote

$$h_{\max} = \max(h_1, ..., h_d), \; h_{\text{prod}} = h_1 \times \cdots \times h_d$$

and for any $x \in R$, $\tilde{K}(x) = \int_{-\infty}^{x} K(u)\,du$,, where K is a kernel function in Assumption (A4). $\tilde{K}(\mathbf{x}) = \prod_{\alpha=1}^{d} \tilde{K}(x_\alpha)$ for any vector $\mathbf{x} = (x_1, \ldots, x_d)^T$. Then $\tilde{K}(\mathbf{x}) \equiv 0$ unless $\mathbf{x} \geq -\mathbf{1}$ and $\tilde{K}(\mathbf{x}) \equiv 1$ if $\mathbf{x} \geq \mathbf{1}$.where for any two vectors $\mathbf{x} = (x_1, \ldots, x_d)^T$, $\mathbf{y} = (y_1, \ldots, y_d)^T$, $\mathbf{x} \geq \mathbf{y}$ if and only if $x_\alpha \geq y_\alpha, \forall \alpha = 1, ..., d$. It is easily verified that $\int_{-1}^{1} \tilde{K}(w)\,dw = 1$, We also denote $\mu_{p+1}(K) = \int_{-1}^{1} K(u)\,u^{p+1}du$, $D(K) = 1 - \int_{-1}^{1} \tilde{K}^2(w)\,dw$. For any vector $\mathbf{x} = (x_1, \ldots, x_d)^T$ and $\forall \alpha = 1, ..., d$, we denote $\mathbf{x}_{\text{-}\alpha} = (x_1, \ldots, x_{\alpha-1}, x_{\alpha+1}, \ldots, x_d)^T$ and with slight abuse of notation, write $\mathbf{x} = (x_\alpha, \mathbf{x}_{\text{-}\alpha})^T$.

We list below some basic assumptions.

(A1) *The cumulative distribution function* $F \in C^{(p+d+1)}\left(R^d\right)$, *in which* p *is an odd integer, while all* $(p+d+1)$*-th partial derivatives of* F *belong to* $L_1\left(R^d\right)$ *and* $\max_{\mathbf{x} \in R^d} |f(\mathbf{x})| \leq C$.

(A2) *There exist positive constants* K_0 *and* λ_0 *such that* $\alpha(k) \leq K_0 \exp(-\lambda_0 k)$ *holds for all* k, *where the* k*-th order strong mixing coefficient of the strictly stationary process* $\{\mathbf{X}_s\}_{s=-\infty}^{\infty}$ *is defined as*

$$\alpha(k) = \sup_{B \in \sigma\{\mathbf{X}_s, s \leq t\}, C \in \sigma\{\mathbf{X}_s, s \geq t+k\}} |P(B \cap C) - P(B)P(C)|, k \geq 1.$$

(A3) *As* $n \to \infty$, $nh_{\text{prod}} \to \infty$, $n^{1/2}h_{\text{prod}}/(\log n)^{1/2} + n^{1/2}h_{\max}^{p+1} \to 0$.

(A4) *The univariate kernel function* $K(\cdot)$ *is of* $(p+1)$*-th order, supported on* $[-1, 1]$, *Lipschitz continuous.*

Assumptions (A1) to (A4) are all typical conditions in time series smoothing literature, see Bosq (1998) Chapter 2 for similar or even stronger assumptions. Elementary arguments show that $D(K) > 0$ under Assumption (A4).

The following theorem concerns the asymptotic distribution of $\hat{F}$ given in (2.1.1) at any $\mathbf{x} \in R^d$.

THEOREM 2.2.1. *Under Assumptions (A1)-(A4),* $\forall \mathbf{x} \in R^d$ *as* $n \to \infty$

$$\sqrt{nV^{-1}(\mathbf{x})}\left(\hat{F}(\mathbf{x}) - F(\mathbf{x})\right) \to_d N(0,1),$$

where

$$V(\mathbf{x}) = \sum_{l=-\infty}^{\infty} \gamma(l), \gamma(l) = EI\{\mathbf{X}_i \leq \mathbf{x}\} I\{\mathbf{X}_{i+l} \leq \mathbf{x}\} - F^2(\mathbf{x}).$$

Theorem 2.2.1 shows that the smooth estimator $\hat{F}(\mathbf{x})$ has asymptotically the same distribution as the empirical cdf $\tilde{F}(\mathbf{x})$. In particular, for iid process $\{\mathbf{X}_s\}, s = -\infty, ..., \infty$, the asymptotic variance function $V(\mathbf{x})$ reduces to the more familiar form of $\gamma(0) = F(\mathbf{x})\{1 - F(\mathbf{x})\}$.

The global performance of $\hat{F}(\mathbf{x})$ as an estimator of $F(\mathbf{x})$ can be measured in terms of Mean Integrated Squared Error (MISE) and maximal deviation

$$\text{MISE}\left(\hat{F}\right) = \text{MISE}\left(\hat{F};\mathbf{h}\right) = E\int\left\{\hat{F}(\mathbf{x}) - F(\mathbf{x})\right\}^2 dF(\mathbf{x}), \tag{2.2.1}$$

$$D_n\left(\hat{F}\right) = D_n\left(\hat{F};\mathbf{h}\right) = \sup_{\mathbf{x}\in R^d}\left|\hat{F}(\mathbf{x}) - F(\mathbf{x})\right|. \tag{2.2.2}$$

The next two theorems give the asymptotic formula of MISE $\left(\hat{F}\right)$ and the almost sure rate of $D_n\left(\hat{F}\right)$.

THEOREM 2.2.2. *Under Assumptions (A1)-(A4), as* $n \to \infty$,

$$\text{MISE}\left(\hat{F};\mathbf{h}\right) = \text{AMISE}\left(\hat{F};\mathbf{h}\right) + o\left(h_{\max}^{2p+2} + n^{-1}h_{\max}\right)$$

in which the Asymptotic Mean Integrated Squared Error (AMISE) is

$$\text{AMISE}\left(\hat{F};\mathbf{h}\right) = \frac{\int V(\mathbf{x})\,dF(\mathbf{x})}{n} + \frac{\mu_{p+1}^2(K)}{(p+1)!^2}\sum_{\alpha,\beta=1}^{d} h_\alpha^{p+1} h_\beta^{p+1} B_{\alpha\beta,p+1}(F)$$
$$-\frac{D(K)\sum_{\alpha=1}^{d} h_\alpha C_\alpha(F)}{n}$$

with

$$B_{\alpha\beta,p+1}(F) = \int \frac{\partial^{p+1}F(\mathbf{x})}{\partial x_\alpha^{p+1}} \frac{\partial^{p+1}F(\mathbf{x})}{\partial x_\beta^{p+1}} dF(\mathbf{x}), C_\alpha(F) = \int \frac{\partial F(\mathbf{x})}{\partial x_\alpha} dF(\mathbf{x}), \forall \alpha, \beta = 1, ..., d.$$

THEOREM 2.2.3. *Under Assumptions (A1)-(A4), as* $n \to \infty$, $D_n\left(\hat{F}\right) = O_{a.s.}\left(n^{-1/2}\log n\right)$ *while for i.i.d.* $\mathbf{X}_1, ..., \mathbf{X}_n, D_n\left(\hat{F}\right) = O_{a.s.}\left(n^{-1/2}(\log n)^{1/2}\right)$.

The first term $n^{-1}\int V(\mathbf{x})\,dF(\mathbf{x})$ in the formula of AMISE $\left(\hat{F};\mathbf{h}\right)$ is the exact MISE of the empirical cdf $\tilde{F}$. We are unaware of any published results on the MISE or the strong uniform rate of convergence for smooth estimation of multivariate distribution function based on strongly mixing data, as in Theorems 2.2.2 and 2.2.3. Since Assumptions (A1) to (A4) are mild, we believe that these strong theoretical results hold for most multiple time series data with continuous distributions.

In the next section we describe how Theorem 2.2.2 is used to compute a data-driven bandwidth for implementing the smoothed estimator $\hat{F}$.

2.3 Bandwidth Selection

To have insight into the minimization of AMISE $\left(\hat{F};\mathbf{h}\right)$ given in Theorem 2.2.2, define a function $Q : R_{+}^{d} \times M_{+}(d) \times R_{+}^{d}$ for elementwise positive vectors $\mathbf{v} = (v_1, ..., v_d)^T, \mathbf{a} = (a_1, ..., a_d)^T \in R_{+}^{d} = (0, +\infty)^d$ and $\mathbf{M} = \left(M_{\alpha\beta}\right)_{\alpha,\beta=1}^{d} \in M_{+}(d)$, the set of all positive definite $d \times d$ matrices:

$$Q(\mathbf{v},\mathbf{M},\mathbf{a}) = \sum\nolimits_{\alpha,\beta=1}^{d} v_\alpha v_\beta M_{\alpha\beta} - \sum\nolimits_{\alpha=1}^{d} a_\alpha v_\alpha^{1/(p+1)} = \mathbf{v}^T\mathbf{M}\mathbf{v} - \mathbf{a}^T\mathbf{v}^{1/(p+1)}$$

in which $\mathbf{v}^{1/(p+1)} = \left(v_1^{1/(p+1)}, ..., v_d^{1/(p+1)}\right)^T$. In the following, we denote for any d-dimensional vector $\mathbf{a} = (a_1, ..., a_d)^T$, the $d \times d$ diagonal matrix whose $(\alpha\alpha)$-th element is $a_\alpha, \alpha = 1, ..., d$ as $\operatorname{diag}(\mathbf{a})$. The following theorem is easily proved similar to Yang and Tschernig (1999).

THEOREM 2.3.1. *(i) The gradient and Hessian matrices of* $Q(\mathbf{v},\mathbf{M},\mathbf{a})$ *with respect to* $\mathbf{v}$ *are*

$$\frac{\partial}{\partial\mathbf{v}}Q(\mathbf{v},\mathbf{M},\mathbf{a}) = \left\{\operatorname{diag}\left(M_{\alpha\alpha}\right)_{\alpha=1}^{d} + \mathbf{M}\right\}\mathbf{v} - \frac{1}{p+1}\operatorname{diag}(\mathbf{a})\,\mathbf{v}^{1/(p+1)-1},$$

$$\frac{\partial^2}{\partial\mathbf{v}\partial\mathbf{v}^T}Q(\mathbf{v},\mathbf{M},\mathbf{a}) = \operatorname{diag}\left(M_{\alpha\alpha}\right)_{\alpha=1}^{d} + \mathbf{M} + \frac{p}{(p+1)^2}\operatorname{diag}\left(a_\alpha v_\alpha^{1/(p+1)-2}\right)_{\alpha=1}^{d}$$

the Hessian matrix of $Q(\mathbf{v},\mathbf{M},\mathbf{a})$ *is positive definite, hence the function* $Q(\mathbf{v},\mathbf{M},\mathbf{a})$ *is strictly convex in* $\mathbf{v}$. *(ii) For any* $\mathbf{a} \in R_{+}^{d}, \mathbf{M} \in M_{+}(d)$, *there exists a unique* $\mathbf{v} \in R_{+}^{d}$ *which minimizes* $Q(\mathbf{v},\mathbf{M},\mathbf{a})$, *denoted as* $\mathbf{v}(\mathbf{M},\mathbf{a})$, *which satisfies* $\frac{\partial}{\partial\mathbf{v}}Q(\mathbf{v},\mathbf{M},\mathbf{a}) = \mathbf{0}$. *In addition,*

$Q\{\mathbf{v}(\mathbf{M},\mathbf{a}),\mathbf{M},\mathbf{a}\} < 0$ *for any* $\mathbf{a} \in R_+^d, \mathbf{M} \in M_+(d)$. *(iii) Lastly, for any* $c_\mathbf{M}, c_\mathbf{a} > 0$

$$\begin{aligned} Q\left(c_\mathbf{a}^{(p+1)/(2p+1)} c_\mathbf{M}^{-(p+1)/(2p+1)}\mathbf{v}, c_\mathbf{M}\mathbf{M}, c_\mathbf{a}\mathbf{a}\right) &= c_\mathbf{a}^{(2p+2)/(2p+1)} c_\mathbf{M}^{-1/(2p+1)} Q(\mathbf{v},\mathbf{M},\mathbf{a}), \\ \mathbf{v}(c_\mathbf{M}\mathbf{M}, c_\mathbf{a}\mathbf{a}) &= c_\mathbf{a}^{(p+1)/(2p+1)} c_\mathbf{M}^{-(p+1)/(2p+1)}\mathbf{v}(\mathbf{M},\mathbf{a}). \end{aligned}$$

To make use of Theorem 2.3.1, we make an additional assumption on F,

(A5) *The matrices* $\mathbf{B}_{p+1}(F) = \{B_{\alpha\beta,p+1}(F)\}_{\alpha,\beta=1}^d \in M_+(d)$ *and* $\mathbf{C}(F) = \{C_\alpha(F)\}_{\alpha=1}^d \in R_+^d$.

Theorem 2.2.2, Theorem 2.3.1 (ii) and Assumption (A5) ensure the existence of a unique optimal bandwidth $\mathbf{h}_{\mathrm{opt}}$ that minimizes

$$\mathrm{AMISE}\left(\hat{F};\mathbf{h}\right) = \frac{\int V(\mathbf{x})\,dF(\mathbf{x})}{n} + Q\left(\mathbf{h}^{p+1}, \frac{\mu_{p+1}^2(K)}{(p+1)!^2}\mathbf{B}_{p+1}(F), n^{-1}D(K)\mathbf{C}(F)\right)$$

Theorem 2.3.1 (iii) then implies that

$$\begin{aligned} \mathbf{h}_{\mathrm{opt}} &= \mathbf{h}_{\mathrm{opt}}(n,K,F) = \mathbf{v}^{1/(p+1)}\left(\frac{\mu_{p+1}^2(K)}{(p+1)!^2}\mathbf{B}_{p+1}(F), n^{-1}D(K)\mathbf{C}(F)\right) \\ &= n^{-1/(2p+1)}\left\{\frac{\mu_{p+1}^2(K)}{D(K)(p+1)!^2}\right\}^{-1/(2p+1)} \mathbf{v}^{1/(p+1)}\left(\mathbf{B}_{p+1}(F),\mathbf{C}(F)\right). \end{aligned}$$

Thus to obtain the optimal bandwidth $\mathbf{h}_{\mathrm{opt}}$, one computes exactly the factors involving n and K in the above expression, and estimate the following factor

$$\boldsymbol{\theta} = \boldsymbol{\theta}(F) = (\theta_1,\ldots,\theta_d)^T = (\theta_1(F),\ldots,\theta_d(F))^T = \mathbf{v}^{1/(p+1)}\left(\mathbf{B}_{p+1}(F),\mathbf{C}(F)\right).$$

The next theorem follows from the negativity result in Theorem 2.3.1 (ii):

Theorem 2.3.2. *Under Assumptions (A1)-(A5),* $\hat{F}$ *has asymptotically smaller MISE than the empirical cdf* $\tilde{F}$. *Specifically,* $\mathrm{MISE}\left(\tilde{F}\right) = n^{-1}\int V(\mathbf{x})\,dF(\mathbf{x})$ *and as* $n \to \infty$

$$\mathrm{MISE}\left(\hat{F};\mathbf{h}_{\mathrm{opt}}\right) = \mathrm{MISE}\left(\tilde{F}\right) + n^{-(2p+2)/(2p+1)}C(K,F) + o\left(n^{-(2p+2)/(2p+1)}\right),$$

$$C(K,F) = \left\{\frac{D(K)^{2p}\mu_{p+1}^2(K)}{(p+1)!^2}\right\}^{-1/(2p+1)} Q\left(\mathbf{v}\left(\mathbf{B}_{p+1}(F),\mathbf{C}(F)\right),\mathbf{B}_{p+1}(F),\mathbf{C}(F)\right) < 0.$$

Following Yang and Tschernig (1999), we define a plug-in asymptotic optimal bandwidth

$$\hat{\mathbf{h}}_{\rm opt} = \left\{ \frac{n\mu_{p+1}^2(K)}{C(K)(p+1)!^2} \right\}^{-1/(2p+1)} \mathbf{v}^{1/(p+1)} \left(\hat{\mathbf{B}}_{p+1}(F), \hat{\mathbf{C}}(F) \right)$$

in which the plug-in estimator of the unknown parameter $\boldsymbol{\theta}$, $\hat{\boldsymbol{\theta}} = \mathbf{v}^{1/(p+1)} \left(\hat{\mathbf{B}}_{p+1}(F), \hat{\mathbf{C}}(F) \right)$, is computed by Newton-Raphson method using the gradient and Hessian formulae of Theorem 2.3.1 and where the plug-in estimators of the unknown matrices $\mathbf{B}_{p+1}(F) = \{B_{\alpha\beta,p+1}(F)\}_{\alpha,\beta=1}^{d}, \mathbf{C}(F)$ are

$$\hat{\mathbf{B}}_{p+1}(F) = \left\{ \hat{B}_{\alpha\beta,p+1}(F) \right\}_{\alpha,\beta=1}^{d}, \hat{\mathbf{C}}(F) = \left\{ \hat{C}_{\alpha}(F) \right\}_{\alpha=1}^{d},$$

$$\begin{aligned} \hat{B}_{\alpha\beta,p+1}(F) &= n^{-1} \sum_{j=1}^{n} \left\{ n^{-1} \sum_{i=1}^{n} K_{g\alpha}^{(p)}(X_{j\alpha} - X_{i\alpha}) \prod_{\gamma=1,\gamma\neq\alpha}^{d} \int_{-\infty}^{X_{j\gamma}} K_{g_\gamma}(x_\gamma - X_{i\gamma})\, dx_\gamma \right\} \\ &\times \left\{ n^{-1} \sum_{i=1}^{n} K_{g\beta}^{(p)}(X_{j\beta} - X_{i\beta}) \prod_{\gamma=1,\gamma\neq\beta}^{d} \int_{-\infty}^{X_{i\gamma}} K_{g_\gamma}(x_\gamma - X_{i\gamma})\, dx_\gamma \right\}, \end{aligned}$$

$$\hat{C}_{\alpha}(F) = n^{-1} \sum_{j=1}^{n} \left\{ n^{-1} \sum_{i=1}^{n} K_{g\alpha}(X_{j\alpha} - X_{i\alpha}) \prod_{\gamma=1,\gamma\neq\alpha} \int_{-\infty}^{X_{j\gamma}} K_{g_\gamma}(x_\gamma - X_{i\gamma})\, dx_\gamma \right\}.$$

The pilot bandwidth vector $\mathbf{g} = (g_1, ..., g_d)^T$ is the simple rule-of-thumb bandwidth for multivariate density estimation in Scott (1992).

In the next section, we present Monte Carlo evidence for Theorems 2.2.2 and 2.2.3, and illustrate the use of the smooth estimator $\hat{F}(\mathbf{x})$ with real data examples. In all computing, we use the quartic kernel $K(u) = 15/16 \times (1-u^2)^2 I(|u| \leq 1)$ with $p = 1$ and plug-in bandwidth $\hat{\mathbf{h}}_{\rm opt}$ described above. We have not experimented with other choices of K and p due to limit of space and as these choices are in general not as crucial as that of the bandwidth, see Fan and Yao (2003).

2.4 Examples

2.4.1 A simulated example

In the section, we examine the asymptotic results of Theorems 2.2.2 and 2.2.3 via simulation. The data are generated from the following vector autoregression (VAR) equation

$$\mathbf{X}_t = a\mathbf{X}_{t-1} + \varepsilon_i, \varepsilon_i \sim N(0, \Sigma), 2 \le t \le n, \Sigma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}, 0 \le a, \rho < 1$$

with stationary distribution $\mathbf{X}_t = (X_{t1}, X_{t2})^T \sim N\left(0, \left(1-a^2\right)^{-1}\Sigma\right)$. Clearly, higher values of a correspond to stronger dependence among the observations, and in particular, if $a = 0$, the data is i.i.d. The parameter ρ controls the orientation of the bivariate cdf F, and in particular, if $a = \rho = 0$, then F is a bivariate standard normal distribution. In this study, we have experimented with three cases: $\rho = 0$, $a = 0$; $\rho = 0.5$, $a = 0.2$; $\rho = 0.9$, $a = 0.2$ to cover various scenarios.

A total of 100 samples $\{\mathbf{X}_t\}_{t=1}^n$ of sizes $n = 50, 100, 200, 500$ are generated, and $\hat{F}$ is computed using the optimal bandwidths $\hat{\mathbf{h}}_{\text{opt}}$ described in section 2.3. Of interest are the mean over the 100 replications of the global maximal deviation $D_n\left(\hat{F}\right)$ defined in (2.2.2), denoted as $\bar{D}_n\left(\hat{F}\right)$, and the mean integrated squared error $\text{MISE}\left(\hat{F}; \hat{\mathbf{h}}_{\text{opt}}\right)$ defined in (2.2.1). Both measures are listed in Table 1. As one examines Table 1, both $\bar{D}_n\left(\hat{F}\right)$ and $\text{MISE}\left(\hat{F}; \hat{\mathbf{h}}_{\text{opt}}\right)$ values decrease as sample size increases in all cases, corroborating with Theorems 2.2.2 and 2.2.3. Also listed in Table 1 are the differences of the same measures for the empirical cdf $\tilde{F}$ against those of $\hat{F}$, which are always positive regardless of the data generating process (i.e., for different combinations of a, ρ) and measures of deviation (i.e., $\bar{D}_n$ or MISE). This corroborates with Theorem 2.3.2 that $\hat{F}$ has asymptotically smaller MISE than $\tilde{F}$.

Based on the above observations, we believe our kernel estimator of multivariate cdf is a convenient and reliable tool, which is also superior to the empirical cdf in terms of accuracy.

2.4.2 GDP growth and unemployment

In this section, we discuss in detail the dependence of US GDP quarterly growth rate on unemployment rate. There are three types of unemployment: frictional, structural, and cyclical. Economists regard frictional and structural unemployment as essentially unavoidable in dynamic economy, so full employment is something less than 100% employment. The full-employment rate of unemployment is also referred to as the natural rate of unemployment. It does not mean the economy will always operate at the natural rate. The economy sometimes operates at an unemployment rate higher than the natural rate due to cyclical unemployment. In contrast, the economy may on some occasions achieve an unemployment rate below the natural rate. For example, during World War II, when the natural rate was about 4% and actual rate below 2% during 1943-1945. It is caused by the pressure of wartime production resulted in an almost unlimited demand for labor. The natural rate is not forever fixed. It was about 4% in the 1960s, and economists generally agreed that the natural rate was about 6%. Today, the consensus is that the rate is about 5.5%.

GDP gap denotes the amount by which actual GDP falls short of the theoretical GDP under the natural rate. Okun's law, based on recent estimate, indicates that for every 1% which the actual unemployment rate exceeds the natural rate, a GDP gap of about 2% occurs. See Samuelson (1995), p.559 or McConnell and Brue (1999), p.214 for more details. In other words, if unemployment rate falls, then GDP growth rate increases. But unemployment rate can not keep falling because it moves around the natural rate. So it is useful to find the relationship between the GDP growth rate and unemployment growth rate.

Let X_{t1} = the seasonally adjusted quarterly unemployment growth rate in quarter t, X_{t2} = the quarterly GDP growth rate in quarter t, all data taken from the 1-st quarter of 1948 ($t = 1$) to the 2-nd quarter of 2006 ($t = 234$) . Since both data have been seasonally adjusted, it is reasonable to treat $\mathbf{X}_t = (X_{t1}, X_{t2})^T, t = 1, ..., 234$ as a strictly stationary time series, which is shown in the time plots. ACF plots also indicate that the assumption of α-mixing is satisfied. The plots are shown in Figures 1−4.

Given any interval $I = [a, b]$, the survival function of X_{t2} conditional on $X_{t1} \in I$ is defined as

$$S_I(x_2) = P(X_{t2} > x_2 | X_{t1} \in I) = 1 - \frac{F(b, x_2) - F(a, x_2)}{F(b, +\infty) - F(a, +\infty)} \tag{2.4.1}$$

in which F is the joint distribution function of X_{t1} and X_{t2}.

The function $S_I(x_2)$ can be approximated by the following plug-in estimator

$$\hat{S}_I(x_2) = 1 - \frac{\hat{F}(b, x_2) - \hat{F}(a, x_2)}{\hat{F}(b, +\infty) - \hat{F}(a, +\infty)} \tag{2.4.2}$$

in which $\hat{F}$ is the kernel estimator of F defined in (2.1.1). According to Theorems 2.2.1 and 2.2.3, for any fixed x_2, $\left|\hat{S}_I(x_2) - S_I(x_2)\right| = O_p\left(n^{-1/2}\right)$ while

$$\sup_{x_2 \in R} \left|\hat{S}_I(x_2) - S_I(x_2)\right| = O_{a.s.}\left(n^{-1/2} \log n\right),$$

so the estimator $\hat{S}_I(x_2)$ is theoretically very reliable. We therefore draw probabilistic conclusions based on the smooth estimate $\hat{S}_I(x_2)$ instead of the true $S_I(x_2)$.

In Figure 5, the estimated conditional survival curve $\hat{S}_I(x_2)$ is plotted for intervals $I = [-0.08, -0.04]$, $I = [-0.02, 0.02]$, $I = [0.04, 0.08]$. Clearly, when the unemployment growth rate is between -0.08 and -0.04, the chance to have the GDP growth rate higher than 1.5% is the greatest, which is about 0.2. This is in accordance with the Okun's law that the growth in GDP is the associated with the unemployment rate. So if policymakers want to achieve high GDP growth rate, they may find better ways to lower the unemployment rate. One can even estimate the probabilities of GDP growth rates given the policy of unemployment, which is the interval I. If current unemployment rate is close to the natural rate, then the I is an interval close to 0, such as $[-0.02, 0.02]$; if the current unemployment rate is much higher than the natural rate, then the I is an negative interval, i.e., trying to lower the unemployment rate.

On the other hand, the survival function of X_{t1} conditional on X_{t2} can be computed similarly. If certain level of GDP growth rate is planned to be achieved, one can estimate the conditional probabilities of different unemployment growth rates.

2.5 Appendix

2.5.1 Preliminaries

In this appendix, we denote by C (or c) any positive constants, by U (or u) sequences of random variables that are uniformly O (or o) of certain order and by $O_{a.s.}$ almost surely O, etc.

LEMMA 2.5.1. *[Berry-Esseen inequality, Sunklodas (1984), Theorem 1] Let* $\{\xi_i\}_{i=1}^n$ *be an* α*-mixing sequence with* $E\xi_n = 0$. *Denote* $d_\delta := \max_{1\le i\le n}\left\{E|\xi_i|^{2+\delta}\right\}, 0 < \delta \le 1$, $S_n = \sum_{i=1}^n \xi_i$, $\sigma_n^2 := ES_n^2 \ge c_0 n$ *for some* $c_0 \in (0, +\infty)$. *If* $\alpha(n) \le K_0 \exp(-\lambda_0 n)$, $\lambda_0 > 0$, $K_0 > 0$, *then there exist* $c_1 = c_1(K_0, \delta)$, $c_2 = c_2(K_0, \delta)$, *such that*

$$\Delta_n = \sup_z \left|P\left\{\sigma_n^{-1} S_n < z\right\} - \Phi(z)\right| \le c_1 \frac{d_\delta}{c_0 \sigma_n^\delta}\left\{\log\left(\sigma_n / c_0^{1/2}\right)/\lambda\right\}^{1+\delta} \tag{2.5.1}$$

for any λ *with* $\lambda_1 \le \lambda \le \lambda_2$, *where*

$$\lambda_1 = c_2\left\{\log\left(\sigma_n/c_0^{1/2}\right)\right\}^b/n, b > 2(1+\delta)/\delta; \lambda_2 = 4(2+\delta)\delta^{-1}\log\left(\sigma_n/c_0^{1/2}\right).$$

LEMMA 2.5.2. *(Bernstein's inequality, Bosq (1998), Theorem 1.4). Let* $\{\xi_t\}$ *be a zero mean real valued process,* $S_n = \sum_{i=1}^n \xi_i$. *Suppose that there exists* $c > 0$ *such that for* $i = 1, \cdots, n$, $k \ge 3, E|\xi_i|^k \le c^{k-2} k! E\xi_i^2 < +\infty, m_r = \max_{1\le i\le N} \|\xi_i\|_r, r \ge 2$. *Then for each* $n > 1$, *integer* $q \in [1, n/2]$, *each* $\varepsilon > 0$ *and* $k \ge 3$

$$P\left\{\left|\sum_{i=1}^n \xi_i\right| > n\varepsilon_n\right\} \le a_1 \exp\left(-\frac{q\varepsilon_n^2}{25m_2^2 + 5c\varepsilon_n}\right) + a_2(k)\,\alpha\left(\left[\frac{n}{q+1}\right]\right)^{\frac{2k}{2k+1}}$$

where

$$a_1 = 2\frac{n}{q} + 2\left(1 + \frac{\varepsilon_n^2}{25m_2^2 + 5c\varepsilon_n}\right), a_2(k) = 11n\left(1 + \frac{5m_k^{2k/(2k+1)}}{\varepsilon_n}\right).$$

2.5.2 Proofs of Theorems 2.2.1 and 2.2.2

LEMMA 2.5.3. *Under Assumptions (A1),(A3) and (A4), as* $n \to \infty$

$$E\left\{\hat{F}(\mathbf{x})\right\} = F(\mathbf{x}) + \frac{\mu_{p+1}(K)}{(p+1)!}\sum_{\alpha=1}^d h_\alpha^{p+1}\frac{\partial^{p+1}F(\mathbf{x})}{\partial x_\alpha^{p+1}} + u\left(h_{\max}^{p+1}\right).$$

Proof. Using the integral form of Taylor expansion and denoting $\mathbf{hv} = (h_1 v_1, ..., h_d v_d)^T$, we write

$$f(\mathbf{u}+\mathbf{hv}) \equiv f(\mathbf{u}) + \sum_{r=1}^{p} \frac{1}{r!} \left(\sum_{\alpha=1}^{d} h_\alpha v_\alpha \frac{\partial}{\partial u_\alpha} \right)^r f(\mathbf{u}) + R_{p+1},$$

$$R_{p+1} = R_{p+1}(\mathbf{u}, \mathbf{hv}) = \int_0^1 \left\{ \frac{t^p}{p!} \left(\sum_{\alpha=1}^{d} h_\alpha v_\alpha \frac{\partial}{\partial u_\alpha} \right)^{p+1} f(\mathbf{u}+t\mathbf{hv}) \right\} dt.$$

Hence Assumption (A4), (A1) and (A3) sequentially imply that

$$E\left\{\hat{F}(\mathbf{x})\right\} = E\int_{-\infty}^{\mathbf{x}} K_{\mathbf{h}}(\mathbf{X}_i - \mathbf{u})\, d\mathbf{u} = \int_{-\infty}^{\mathbf{x}} d\mathbf{u} \int_{[-1,1]^d} f(\mathbf{u}+\mathbf{hv}) K(\mathbf{v})\, d\mathbf{v}$$

$$= \int_{-\infty}^{\mathbf{x}} f(\mathbf{u}) d\mathbf{u} + \int_{-\infty}^{\mathbf{x}} d\mathbf{u} \int_{[-1,1]^d} \left[\sum_{r=1}^{p} \frac{1}{r!} \left(\sum_{\alpha=1}^{d} h_\alpha v_\alpha \frac{\partial}{\partial u_\alpha} \right)^r f(\mathbf{u}) + R_{p+1} \right] K(\mathbf{v})\, d\mathbf{v}$$

$$= F(\mathbf{x}) + \int_{-\infty}^{\mathbf{x}} d\mathbf{u} \int_{[-1,1]^d} \left[\int_0^1 \left\{ \frac{t^p}{p!} \left(\sum_{\alpha=1}^{d} h_\alpha v_\alpha \frac{\partial}{\partial u_\alpha} \right)^{p+1} f(\mathbf{u}+t\mathbf{hv}) \right\} dt \right] K(\mathbf{v})\, d\mathbf{v}$$

$$= F(\mathbf{x}) + \frac{\mu_{p+1}(K)}{(p+1)!} \int_{-\infty}^{\mathbf{x}} \sum_{\alpha=1}^{d} h_\alpha^{p+1} \frac{\partial^{p+1} f}{\partial u_\alpha^{p+1}}(\mathbf{u}) d\mathbf{u} + u\left(h_{\max}^{p+1}\right)$$

$$= F(\mathbf{x}) + \frac{\mu_{p+1}(K)}{(p+1)!} \sum_{\alpha=1}^{d} h_\alpha^{p+1} \frac{\partial^{p+1} F(\mathbf{x})}{\partial x_\alpha^{p+1}} + u\left(h_{\max}^{p+1}\right). \quad \square$$

LEMMA 2.5.4. *Under Assumptions (A1)-(A4), as $n \to \infty$*

$$E\left\{ \int_{-\infty}^{\mathbf{x}} K_{\mathbf{h}}(\mathbf{X}_i - \mathbf{u})\, d\mathbf{u} \int_{-\infty}^{\mathbf{x}} K_{\mathbf{h}}(\mathbf{X}_j - \mathbf{u})\, d\mathbf{u} \right\}$$

$$= \begin{cases} F(\mathbf{x}) - D(K) \sum_{\alpha=1}^{d} h_\alpha \frac{\partial F(\mathbf{x})}{\partial x_\alpha} + u(h_{\max}) & i = j, \\ EI\{\mathbf{X}_i \le \mathbf{x}\} I\{\mathbf{X}_j \le \mathbf{x}\} + u(h_{\max}) & i \ne j. \end{cases}$$

Proof. We begin with the case of $i = j$,

$$E\left\{ \int_{-\infty}^{\mathbf{x}} K_{\mathbf{h}}(\mathbf{X}_i - \mathbf{u})\, d\mathbf{u} \right\}^2 = \int_{-\infty}^{\infty} f(\mathbf{v}) \tilde{K}\left(\frac{\mathbf{x}-\mathbf{v}}{\mathbf{h}}\right)^2 d\mathbf{v} = \int_{-\mathbf{1}}^{\infty} f(\mathbf{x}-\mathbf{hw}) \tilde{K}^2(\mathbf{w})\, h_{\text{prod}} d\mathbf{w}$$

$$= h_{\text{prod}} \int_{-\mathbf{1}}^{\infty} \{I(\mathbf{w} \ge -\mathbf{1}) - I(\mathbf{w} \ge \mathbf{1})\} f(\mathbf{x}-\mathbf{hw}) \tilde{K}^2(\mathbf{w}) d\mathbf{w} + \int_{\mathbf{1}}^{\infty} f(\mathbf{x}-\mathbf{hw}) h_{\text{prod}} d\mathbf{w}$$

$$= h_{\text{prod}} \int_{-\mathbf{1}}^{\infty} \{I(\mathbf{w} \ge -\mathbf{1}) - I(\mathbf{w} \ge \mathbf{1})\} f(\mathbf{x}-\mathbf{hw}) \tilde{K}^2(\mathbf{w}) d\mathbf{w} + F(\mathbf{x}-\mathbf{h})$$

$$= \sum_{\alpha=1}^{d} h_{\text{prod}} \int_{\mathbf{1}}^{\infty} d\mathbf{w}_{\_\alpha} \int_{-1}^{1} dw_\alpha f(\mathbf{x}-\mathbf{hw}) \tilde{K}^2(w_\alpha) + F(\mathbf{x}) - \sum_{\alpha=1}^{d} \frac{\partial F(\mathbf{x})}{\partial x_\alpha} h_\alpha + u(h_{\max})$$

$$= F(\mathbf{x}) - \sum_{\alpha=1}^{d} h_\alpha \frac{\partial F(\mathbf{x})}{\partial x_\alpha} D(K) + u(h_{\max}).$$

Similarly, for the case of $i \neq j$, one obtains

$$E\left\{\int_{-\infty}^{\mathbf{x}} K_{\mathbf{h}}(\mathbf{X}_i - \mathbf{u})\, du \int_{-\infty}^{\mathbf{x}} K_{\mathbf{h}}(\mathbf{X}_j - \mathbf{u})\, du\right\}$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} d\mathbf{v}_i d\mathbf{v}_j f_{i,j}(\mathbf{v}_i, \mathbf{v}_j) \tilde{K}\left(\frac{\mathbf{x}-\mathbf{v}_i}{\mathbf{h}}\right) \tilde{K}\left(\frac{\mathbf{x}-\mathbf{v}_j}{\mathbf{h}}\right)$$

$$= \int_{-1}^{\infty}\int_{-1}^{\infty} f_{i,j}(\mathbf{x}-\mathbf{h}\mathbf{w}_i, \mathbf{x}-\mathbf{h}\mathbf{w}_j) \tilde{K}(\mathbf{w}_i)\, \tilde{K}(\mathbf{w}_J)\, h_{\text{prod}}^2 d\mathbf{w}_i d\mathbf{w}_j$$

$$= h_{\text{prod}}^2 \left\{\int_{-1}^{\infty} \{I(\mathbf{w}_i \geq -1) - I(\mathbf{w}_i \geq 1)\} \tilde{K}(\mathbf{w}_i)\, d\mathbf{w}_i + \int_{1}^{\infty} d\mathbf{w}_i\right\} \times$$
$$\left\{\int_{-1}^{\infty} \{I(\mathbf{w}_j \geq -1) - I(\mathbf{w}_j \geq 1)\} \tilde{K}(\mathbf{w}_j)\, d\mathbf{w}_j + \int_{1}^{\infty} d\mathbf{w}_j\right\} f_{i,j}(\mathbf{x}-\mathbf{h}\mathbf{w}_i, \mathbf{x}-\mathbf{h}\mathbf{w}_j)$$

$$= h_{\text{prod}}^2 \int_{-1}^{\infty} \{I(\mathbf{w}_i \geq -1) - I(\mathbf{w}_i \geq 1)\} \tilde{K}(\mathbf{w}_i)\, d\mathbf{w}_i \int_{1}^{\infty} d\mathbf{w}_j f_{i,j}(\mathbf{x}-\mathbf{h}\mathbf{w}_i, \mathbf{x}-\mathbf{h}\mathbf{w}_j)$$
$$+ h_{\text{prod}}^2 \int_{-1}^{\infty} \{I(\mathbf{w}_j \geq -1) - I(\mathbf{w}_j \geq 1)\} \tilde{K}(\mathbf{w}_j)\, d\mathbf{w}_j \int_{1}^{\infty} d\mathbf{w}_i f_{i,j}(\mathbf{x}-\mathbf{h}\mathbf{w}_i, \mathbf{x}-\mathbf{h}\mathbf{w}_j)$$
$$+ EI\{\mathbf{X}_i \leq \mathbf{x}-\mathbf{h}\} I\{\mathbf{X}_j \leq \mathbf{x}-\mathbf{h}\} + u(h_{\max})$$

$$= \sum_{\alpha=1}^{d} h_\alpha \int_{\mathbf{h}}^{\infty} d\mathbf{v}_j \int_{\mathbf{h}_{i\_\alpha}}^{\infty} d\mathbf{v}_{i\_\alpha} \int_{-1}^{1} \tilde{K}(w_{i\alpha})\, dw_{i\alpha} f_{i,j}(x_\alpha - h w_{i\alpha}, \mathbf{x}_{\_\alpha} - \mathbf{v}_{i\_\alpha}, \mathbf{x}-\mathbf{v}_j)$$
$$+ \sum_{\alpha=1}^{d} h_\alpha \int_{\mathbf{h}}^{\infty} d\mathbf{v}_i \int_{\mathbf{h}_{j\_\alpha}}^{\infty} d\mathbf{v}_{j\_\alpha} \int_{-1}^{1} \tilde{K}(w_{j\alpha})\, dw_{j\alpha} f_{i,j}(\mathbf{x}-\mathbf{v}_i, x_\alpha - h w_{j\alpha}, \mathbf{x}_{\_\alpha} - \mathbf{v}_{j\_\alpha})$$
$$+ EI\{\mathbf{X}_i \leq \mathbf{x}\} I\{\mathbf{X}_j \leq \mathbf{x}\} - \sum_{\alpha=1}^{d} h_\alpha \frac{\partial EI\{\mathbf{X}_i \leq \mathbf{x}\} I\{\mathbf{X}_j \leq \mathbf{x}\}}{\partial x_\alpha} + u(h_{\max})$$

$$= EI\{\mathbf{X}_i \leq \mathbf{x}\} I\{\mathbf{X}_j \leq \mathbf{x}\} - \sum_{\alpha=1}^{d} h_\alpha \frac{\partial EI\{\mathbf{X}_i \leq \mathbf{x}\} I\{\mathbf{X}_j \leq \mathbf{x}\}}{\partial x_\alpha}$$
$$+ \sum_{\alpha=1}^{d} h_\alpha \frac{\partial EI\{\mathbf{X}_i \leq \mathbf{x}\} I\{\mathbf{X}_j \leq \mathbf{x}\}}{\partial x_\alpha} + u(h_{\max})$$
$$= EI\{\mathbf{X}_i \leq \mathbf{x}\} I\{\mathbf{X}_j \leq \mathbf{x}\} + u(h_{\max}). \square$$

Denote $S_n = S_n(\mathbf{x}) = n\left\{\hat{F}(\mathbf{x}) - E\hat{F}(\mathbf{x})\right\} = \sum_{i=1}^{n} \xi_{in}$ in which

$$\xi_{i,n} = \xi_{i,n}(\mathbf{x}) = \int_{-\infty}^{\mathbf{x}} K_{\mathbf{h}}(\mathbf{X}_i - \mathbf{u})\, d\mathbf{u} - E\left\{\int_{-\infty}^{\mathbf{x}} K_{\mathbf{h}}(\mathbf{X}_i - \mathbf{u})\, du\right\},$$

then clearly $E\xi_{i,n} = 0$. Denote by $\widetilde{\gamma}(l) = \text{cov}(\xi_{i,n}, \xi_{i+l,n})$ the autocovariance function, then

COROLLARY 2.5.1. *Under Assumptions (A1)-(A4), as* $n \to \infty$

$$\operatorname{cov}\left(\xi_{i,n},\xi_{j,n}\right)=\tilde{\gamma}(i-j)=\begin{cases} F(\mathbf{x})-F^2(\mathbf{x})-D(K)\sum_{\alpha=1}^d h_\alpha \frac{\partial F(\mathbf{x})}{\partial x_\alpha}+u(h_{\max}) & i=j,\\ EI\{\mathbf{X}_i\le \mathbf{x}\}I\{\mathbf{X}_j\le\mathbf{x}\}-F^2(\mathbf{x})+u(h_{\max}) & i\ne j.\end{cases}$$

Proof. According to Lemmas 2.5.3 and 2.5.4, $\operatorname{cov}\left(\xi_{i,n},\xi_{j,n}\right)$

$$=\left[E\left\{\int_{-\infty}^{\mathbf{x}} K_{\mathbf{h}}(\mathbf{X}_i-\mathbf{u})\,d\mathbf{u}\int_{-\infty}^{\mathbf{x}} K_{\mathbf{h}}(\mathbf{X}_j-\mathbf{u})\,d\mathbf{u}\right\}-\left(E\int_{-\infty}^{\mathbf{x}} K_{\mathbf{h}}(\mathbf{X}_i-\mathbf{u})\,d\mathbf{u}\right)^2\right]$$

$$=\begin{cases} F(\mathbf{x})-D(K)\sum_{\alpha=1}^d h_\alpha\frac{\partial F(\mathbf{x})}{\partial x_\alpha}+u(h_{\max}) & i=j\\ EI\{\mathbf{X}_i\le\mathbf{x}\}I\{\mathbf{X}_j\le\mathbf{x}\}+u(h_{\max}) & i\ne j\end{cases} -$$

$$\left[F(\mathbf{x})+\frac{\mu_{p+1}(K)}{(p+1)!}\sum_{\alpha=1}^d h_\alpha^{p+1}\frac{\partial^{p+1}F(\mathbf{x})}{\partial x_\alpha^{p+1}}+u\left(h_{\max}^{p+1}\right)\right]^2,$$

the rest of the proof is trivial. □

Proofs of Theorems 2.2.1 and 2.2.2. According to Corollary 2.5.1

$$\tilde{\gamma}(l)=\begin{cases}\gamma(0)-D(K)\sum_{\alpha=1}^d h_\alpha\frac{\partial F(\mathbf{x})}{\partial x_\alpha}+u(h_{\max}) & l=0,\\ \gamma(l)+u(h_{\max}) & l\ne 0,\end{cases}\tag{2.5.2}$$

in which $\gamma(l)=\gamma(l,\mathbf{x})=EI\{\mathbf{X}_1\le\mathbf{x}\}I\{\mathbf{X}_{1+l}\le\mathbf{x}\}-F^2(\mathbf{x})$. Lemma 2.5.3 and Assumption (A3) further imply that

$$S_n=n\left\{\hat{F}(\mathbf{x})-F(\mathbf{x})-\frac{\mu_{p+1}(K)}{(p+1)!}\sum_{\alpha=1}^d h_\alpha^{p+1}\frac{\partial^{p+1}F(\mathbf{x})}{\partial x_\alpha^{p+1}}+u\left(h_{\max}^{p+1}\right)\right\}.\tag{2.5.3}$$

Meanwhile, $\sigma_n^2=ES_n^2=\operatorname{var}(S_n)=nA_n+nB_n$ where $A_n=\sum_{|l|\le c\log n}(1-|l|/n)\tilde{\gamma}(l)$ and $B_n=\sum_{c\log n<|l|<n}(1-|l|/n)\tilde{\gamma}(l)$. Because $|\gamma(l)|$ is

$$|P(\{\omega:\mathbf{X}_1(\omega)\le\mathbf{x}\}\cap\{\omega:\mathbf{X}_{1+h}(\omega)\le\mathbf{x}\})-P(\{\omega:\mathbf{X}_1(\omega)\le\mathbf{x}\})P(\{\omega:\mathbf{X}_{1+h}(\omega)\le\mathbf{x}\})|$$

which is bounded by $\alpha(l)\le K_0e^{-\lambda_0 l}$. Then $\sum_{l=-\infty}^{\infty}|\gamma(l)|\le\gamma(0)+2\sum_{l=1}^{\infty}K_0\exp(-\lambda_0 l)<\infty$ and equation (2.5.2) imply that

$$A_n=\sum_{|l|\le c\log n}(1-|l|/n)\gamma(l)+\sum_{|l|\le c\log n}(1-|l|/n)U(h_{\max})\to\sum_{l=-\infty}^{\infty}\gamma(l)\ge c_0.$$

Next, $\left|\operatorname{cov}\left(\xi_{1,n},\xi_{(1+l),n}\right)\right|\le 4\left\|\xi_{1,n}\right\|_\infty\left\|\xi_{(1+l),n}\right\|_\infty\alpha(h)\le 4K_0\exp(-\lambda_0 l)$ gives

$$|B_n|=\sum_{c\log n<|l|<n}(1-|l|/n)|\tilde{\gamma}(l)|\le\sum_{|l|>c\log n}(1-|l|/n)4K_0K_0\exp(-\lambda_0 l)$$

For $c \geq 2/\lambda_0$, $|B_n| \leq \frac{4K_0 e^{-\lambda_0 c \log n}}{1-e^{-\lambda_0}} = \frac{K_0 n^{-c\lambda_0}}{1-e^{-\lambda_0}} \leq C_1 n^{-2}$. For n large enough, $\sigma_n^2/n = A_n + B_n \to \sum_{l=-\infty}^{\infty} \gamma(l) \geq c_0$, therefore $\sum_{|l|\leq n} \gamma(l) > 0$. Then by (2.5.1) in Lemma 2.5.1,

$$\Delta_n = \sup_z \left|P\left\{\sigma_n^{-1} S_n < z\right\} - \Phi(z)\right| \leq c_1 \frac{d}{c_0 \sigma_n^{\delta}} \left\{\log\left(\sigma_n / c_0^{1/2}\right)/\lambda\right\}^{1+\delta}.$$

Let $\delta = 1$, $\lambda = 4(2+\delta)\delta^{-1}\log\left(\sigma_n/c_0^{1/2}\right) = 12\log\left(\sigma_n/c_0^{1/2}\right)$, $d = 1$, then $\Delta_n \leq \frac{c_1}{c_0\sigma_n} 12^{-2} = \frac{c}{\sigma_n} = O\left(n^{-1/2}\right)$, i.e., $S_n/\sigma_n \to_d N(0,1)$. Theorem 2.2.1 then follows because $\sqrt{n}\sqrt{V^{-1}(\mathbf{x})}\left(\hat{F}(\mathbf{x}) - F(\mathbf{x})\right) \to_d N(0,1)$ by Slutsky's theorem. Equations (2.5.2) and (2.5.3) together with $E\xi_{i,n} = 0$ imply that

$$\begin{aligned} \left\{E\hat{F}(\mathbf{x}) - F(\mathbf{x})\right\}^2 &= \frac{\mu_{p+1}^2(K)}{(p+1)!^2} \left\{\sum_{\alpha=1}^d h_\alpha^{p+1} \frac{\partial^{p+1} F(\mathbf{x})}{\partial x_\alpha^{p+1}}\right\}^2 + u\left(h_{\max}^{2p+2}\right), \\ E\left\{\hat{F}(\mathbf{x}) - E\hat{F}(\mathbf{x})\right\}^2 &= n^{-1}V(\mathbf{x}) - D(K)n^{-1}\sum_{\alpha=1}^d h_\alpha \frac{\partial F(\mathbf{x})}{\partial x_\alpha} + u\left(n^{-1}h_{\max}\right), \end{aligned}$$

hence Theorem 2.2.2 follows by computing $\int E\left\{\hat{F}(\mathbf{x}) - E\hat{F}(\mathbf{x})\right\}^2 + \left\{E\hat{F}(\mathbf{x}) - F(\mathbf{x})\right\}^2 dF(\mathbf{x})$. □

2.5.3 Proof of Theorem 2.2.3

LEMMA 2.5.5. *Denote* $g_{m_1,\cdots,m_d} = \left(a_{1,m_1},\cdots,a_{d,m_d}\right) \in R^d$, $1 \leq m_\alpha \leq M_\alpha$ *and*

$$A_n = \max_{1\leq m_\alpha \leq M_\alpha} \left|\hat{F}\left(g_{m_1,\cdots,m_d}\right) - E\left\{\hat{F}\left(g_{m_1,\cdots,m_d}\right)\right\}\right|,$$

$$B_n = \max_{1\leq m_\alpha \leq M_\alpha} \left|\hat{F}\left(g_{m_1,\cdots,m_d}\right) - F\left(g_{m_1,\cdots,m_d}\right)\right|.$$

If $\max(M_1,\cdots,M_d) \leq Cn$, *then* $A_n + B_n = O_{a.s}\left(n^{-1/2}\log n\right)$ *while for i.i.d.* $\mathbf{X}_1,...,\mathbf{X}_n$, $A_n + B_n = O_{a.s.}\left(n^{-1/2}(\log n)^{1/2}\right)$.

Proof. Note that $\hat{F}\left(g_{m_1,\cdots,m_d}\right) - E\hat{F}\left(g_{m_1,\cdots,m_d}\right) = n^{-1}\sum_{i=1}^n \zeta_{in}$ in which

$$\begin{aligned} \zeta_{in} = \zeta_{in,m_1,\cdots,m_d} = \zeta_{i,n}\left(g_{m_1,\cdots,m_d}\right) &= \int_{-\infty}^{g_{m_1,\cdots,m_d}} K_{\mathbf{h}}(\mathbf{X}_i - \mathbf{u})\,d\mathbf{u} \\ &\quad - E\left\{\int_{-\infty}^{g_{m_1,\cdots,m_d}} K_{\mathbf{h}}(\mathbf{X}_i - \mathbf{u})\,d\mathbf{u}\right\}, \end{aligned}$$

then one has $E\zeta_{in} = 0$ while

$$E(\zeta_{in}^2) = E\left(\int_{-\infty}^{g_{m_1,\cdots,m_d}} K_{\mathbf{h}}(\mathbf{X}_i - \mathbf{u})\, d\mathbf{u} - E\left\{\int_{-\infty}^{g_{m_1,\cdots,m_d}} K_{\mathbf{h}}(\mathbf{X}_i - \mathbf{u})\, d\mathbf{u}\right\}\right)^2 \leq 1,$$

and for $k \geq 2$, $E\left(|\zeta_{in}|^k\right) = E\left(|\zeta_{in}|^{k-2}\zeta_{in}^2\right)$, which is

$$E\left[\left|\int_{-\infty}^{g_{m_1,\cdots,m_d}} K_{\mathbf{h}}(\mathbf{X}_i - \mathbf{u})\, d\mathbf{u} - E\left\{\int_{-\infty}^{g_{m_1,\cdots,m_d}} K_{\mathbf{h}}(\mathbf{X}_i - \mathbf{u})\, d\mathbf{u}\right\}\right|^{k-2} \zeta_{in}^2\right]$$
$$\leq 1^{k-2} E(\zeta_{in}^2).$$

By Lemma 2.5.2 with $k = 3$, $a_2(3) = 11n\left(1 + 5m_3^{6/7}/\varepsilon_n\right)$, $m_2^2 = E(\zeta_{in}^2) \leq 1$, $\varepsilon_n = a\log n/\sqrt{n}$,

$$P\left\{\left|\sum\nolimits_{i=1}^n \zeta_{in}\right| > n\varepsilon_n\right\} \leq a_1 \exp\left(-\frac{q\varepsilon_n^2}{25m_2^2 + 5c\varepsilon_n}\right) + a_2(3)\,\alpha\left([n/(q+1)]\right)^{\frac{6}{7}}.$$

Take q such that $[n/(q+1)] \geq c_0 \log n, q \geq \dfrac{c_1 n}{\log n}$, then $\dfrac{q\varepsilon_n^2}{25m_2^2 + 5c\varepsilon_n} \geq c_2 a^2 \log n$ and

$$a_1 = 2\frac{n}{q} + 2\left(1 + \frac{\varepsilon_n^2}{25m_2^2 + 5c\varepsilon_n}\right) = O(\log n).$$

Since $m_3 = \max_{1\leq i\leq n}\|\zeta_i\|_3 \leq \left\{E(\zeta_{in}^3)\right\}^{1/3} \leq 1$, then

$$a_2(3) = 11n\left(1 + \frac{5}{\varepsilon_n}\right) \leq 11n\left\{1 + \frac{5}{an^{-\frac{1}{2}}\log n}\right\} \leq 11n\left\{1 + \frac{5}{a\log n}\right\} = O(n),$$

$$\alpha\left([n/(q+1)]\right)^{6/7} \leq \left(K_0 \exp\left(-\lambda_0 [n/(q+1)]\right)\right)^{6/7} \leq Cn^{-6\lambda_0 c_0/7}.$$

So for c_0, c_2 large enough

$$P\left\{\left|\sum\nolimits_{i=1}^n \zeta_{in}\right| > n\varepsilon_n\right\} \leq O(\log n)\exp\left(-c_2a^2\log n\right) + Cn^{1-6\lambda_0c_0/7} \leq Cn^{-(d+2)},$$

$$P\left\{\max_{1\leq m_\alpha\leq M_\alpha} n^{-1}\left|\sum\nolimits_{i=1}^n \zeta_{in,m_1,\cdots,m_d}\right| > an^{-\frac{1}{2}}\log n\right\} \leq$$

$$\sum_{m_1=1,\ldots,m_d=1}^{M_1,\ldots,M_d} P\left\{n^{-1}\left|\sum_{i=1}^n \zeta_{in,m_1,\cdots,m_d}\right| > an^{-\frac{1}{2}}\log n\right\} \leq Cn^{-(d+2)}\prod_{\alpha=1}^d M_\alpha \leq Cn^{-2}.$$

Hence Borel-Cantelli lemma implies that $A_n = O_{a.s.}\left(n^{-1/2}\log n\right)$. Meanwhile B_n is bounded by

$$\max_{1\leq m_\alpha\leq M_\alpha}\left|\hat{F}\left(g_{m_1,\cdots,m_d}\right) - E\left\{\hat{F}\left(g_{m_1,\cdots,m_d}\right)\right\}\right|$$

$$+\max_{1\le m_\alpha\le M_\alpha}\left|E\left\{\hat{F}\left(g_{m_1,\cdots,m_d}\right)\right\}-F\left(g_{m_1,\cdots,m_d}\right)\right|$$
$$=A_n+U\left(n^{-1/2}\right)=O_{a.s.}\left(n^{-1/2}\log n\right).$$

If $\mathbf{X}_1,\cdots,\mathbf{X}_n$ are i.i.d., then $A_n+B_n=O_{a.s.}\left(n^{-1/2}(\log n)^{1/2}\right)$ by using same steps above with Bernstein's inequality of i.i.d. case. □

LEMMA 2.5.6. $\forall A\subset R^d$, $\int_A|K_{\mathbf{h}}(\mathbf{v}-\mathbf{u})|\,d\mathbf{u}\le\int_{R^d}|K_{\mathbf{h}}(\mathbf{v}-\mathbf{u})|\,d\mathbf{u}\le\|K\|_{L^1}^d$.

Proof. Applying elementary arguments, $\int_A|K_{\mathbf{h}}(\mathbf{v}-\mathbf{u})|\,d\mathbf{u}\le\int_{R^d}|K_{\mathbf{h}}(\mathbf{v}-\mathbf{u})|\,d\mathbf{u}$ is bounded by

$$\int_{R^d}\left|\prod_{\alpha=1}^d h_\alpha^{-1}K\left(\frac{v_\alpha-u_\alpha}{h_\alpha}\right)\right|d\mathbf{u}=\prod_{\alpha=1}^d\int_{-1}^1|K(w_\alpha)|\,dw_\alpha\le\|K\|_{L^1}^d. \quad □$$

LEMMA 2.5.7. *Let* $-\infty=a_{\alpha,1}<\cdots<a_{\alpha,N_\alpha}=\infty$ *be such that* $\max(N_1,\cdots,N_d)\le Cn$ *and* $P\left(a_{\alpha,k}\le X_\alpha\le a_{\alpha,k+1}\right)\le 1/n,\forall 1\le k\le N_\alpha,\forall 1\le\alpha\le d$. *Then* $E\int_{g_{n_1,\cdots,n_d}}^{\mathbf{X}}|K_{\mathbf{h}}(\mathbf{X}-\mathbf{u})|\,d\mathbf{u}=u\left(n^{-1/2}(\log n)^{1/2}\right)$ *in which* $g_{n_1,\cdots,n_d}=\left(a_{1,n_1},\cdots,a_{d,n_d}\right)\in R^d$.

Proof.

$$E\int_{g_{n_1,\cdots,n_d}}^{\mathbf{X}}|K_{\mathbf{h}}(\mathbf{X}-\mathbf{u})|\,d\mathbf{u}\le\int_{-\infty}^{\infty}\int_{g_{n_1,\cdots,n_d}}^{g_{n_1+1,\cdots,n_d+1}}|K_{\mathbf{h}}(\mathbf{v}-\mathbf{u})|\,d\mathbf{u}dF(\mathbf{v})$$
$$=\int_{g_{n_1,\cdots,n_d}-(h_1,\cdots,h_d)}^{g_{n_1+1,\cdots,n_d+1}+(h_1,\cdots,h_d)}dF(\mathbf{v})\int_{g_{n_1,\cdots,n_d}}^{g_{n_1+1,\cdots,n_d+1}}|K_{\mathbf{h}}(\mathbf{v}-\mathbf{u})|\,d\mathbf{u}$$
$$\le C\int_{g_{n_1,\cdots,n_d}-(h_1,\cdots,h_d)}^{g_{n_1+1,\cdots,n_d+1}+(h_1,\cdots,h_d)}dF(\mathbf{v})$$

according to Lemma 2.5.6. $\int_{g_{n_1,\cdots,n_d}-(h_1,\cdots,h_d)}^{g_{n_1+1,\cdots,n_d+1}+(h_1,\cdots,h_d)}dF(\mathbf{v})$ equals

$$\int_{g_{n_1,\cdots,n_d}-(h_1,\cdots,h_d)}^{g_{n_1+1,\cdots,n_d+1}+(h_1,\cdots,h_d)}dF(\mathbf{v})-\int_{g_{n_1,\cdots,n_d}}^{g_{n_1+1,\cdots,n_d+1}}dF(\mathbf{v})+\int_{g_{n_1,\cdots,n_d}}^{g_{n_1+1,\cdots,n_d+1}}dF(\mathbf{v})$$
$$=\int_{g_{n_1,\cdots,n_d}-(h_1,\cdots,h_d)}^{g_{n_1+1,\cdots,n_d+1}+(h_1,\cdots,h_d)}dF(\mathbf{v})-\int_{g_{n_1,\cdots,n_d}}^{g_{n_1+1,\cdots,n_d+1}}dF(\mathbf{v})$$
$$+P\left(g_{n_1,\cdots,n_d}\le\mathbf{X}\le g_{n_1+1,\cdots,n_d+1}\right)$$

$$\leq \left(\int_{a_{1,n_1}-h_1}^{a_{1,n_1}} + \int_{a_{1,n_1}}^{a_{1,n_1}+1} + \int_{a_{1,n_1}+1}^{a_{1,n_1}+1+h_1}\right)$$
$$\cdots \left(\int_{a_{d,n_d}-h_1}^{a_{d,n_d}} + \int_{a_{d,n_d}}^{a_{d,n_d}+1} + \int_{a_{d,n_d}+1}^{a_{d,n_d}+1+h_1}\right) dF(\mathbf{v})$$
$$-\int_{g_{n_1,\cdots,n_d}}^{g_{n_1+1,\cdots,n_d+1}} dF(\mathbf{v}) + 1/n.$$

Within the above sum, the $3^d - 2^d$ terms with $\int_{a_{\alpha,n_\alpha}}^{a_{\alpha,n_\alpha+1}}$ are $O(n^{-1})$, while each of the 2^d terms without $\int_{a_{\alpha,n_\alpha}}^{a_{\alpha,n_\alpha+1}}$ is bounded by $h_{\text{prod}} \max_{\mathbf{x}\in R^d} |f(\mathbf{x})|$. Applying Assumptions (A1) and (A3),

$$E\int_{g_{n_1,\cdots,n_d}}^{\mathbf{x}} |K_{\mathbf{h}}(\mathbf{X}-\mathbf{u})|\, d\mathbf{u} \leq Ch_{\text{prod}} \max_{\mathbf{x}\in R^d} |f(\mathbf{x})| + C\left(3^d - 2^d\right)/n = u\left(n^{-1/2}(\log n)^{1/2}\right).$$

□

LEMMA 2.5.8. *Under the same conditions of Lemma 2.5.7, for* $\forall \mathbf{x} = (x_1, \cdots, x_d) \in R^d$, $n^{-1}\sum_{i=1}^n |\zeta_{in}| = U_{a.s.}\left(n^{-1/2}\log n\right)$ *in which*

$$\zeta_{in} = \zeta_{i,n}\left(g_{n_1,\cdots,n_d}\right) = \int_{g_{n_1,\cdots,n_d}}^{\mathbf{x}} \{|K_{\mathbf{h}}(\mathbf{X}_i - \mathbf{u})|\, d\mathbf{u} - E\,|K_{\mathbf{h}}(\mathbf{X}-\mathbf{u})|\}\, d\mathbf{u}.$$

while for i.i.d. $\mathbf{X}_1, ..., \mathbf{X}_n, n^{-1}\sum_{i=1}^n |\zeta_{in}| = U_{a.s.}\left(n^{-1/2}(\log n)^{1/2}\right)$.

Proof. One can show by applying Lemma 2.5.2 as in the proof of Lemma 2.5.5. □

Proof of Theorem 2.2.3. Under the same conditions of Lemma 2.5.7. one has

$$\max_{1\leq n_\alpha \leq N_\alpha} \left|\hat{F}\left(g_{n_1,\cdots,n_d}\right) - F\left(g_{n_1,\cdots,n_d}\right)\right| = O_{a.s.}\left(n^{-1/2}\log n\right)$$

by Lemma 2.5.5. For $\forall \mathbf{x} = (x_1, \cdots, x_d) \in R^d$, there exist integers $n_1, \cdots, n_d$ such that $F(g_{n_1,\cdots,n_d}) \leq F(\mathbf{x}) \leq F\left(g_{n_1+1,\cdots,n_d+1}\right)$. Hence $\left|\hat{F}(\mathbf{x}) - \hat{F}\left(g_{n_1,\cdots,n_d}\right)\right|$ is bounded by

$$\left|\frac{1}{n}\sum_{i=1}^n \int_{g_{n_1,\cdots,n_d}}^{\mathbf{x}} K_{\mathbf{h}}(\mathbf{X}_i - \mathbf{u})\, d\mathbf{u}\right| \leq \frac{1}{n}\sum_{i=1}^n \int_{g_{n_1,\cdots,n_d}}^{\mathbf{x}} |K_{\mathbf{h}}(\mathbf{X}_i - \mathbf{u})|\, d\mathbf{u}$$
$$= \frac{1}{n}\sum_{i=1}^n \int_{g_{n_1,\cdots,n_d}}^{\mathbf{x}} \{|K_{\mathbf{h}}(\mathbf{X}_i - \mathbf{u})|\, d\mathbf{u} - E\,|K_{\mathbf{h}}(\mathbf{X}-\mathbf{u})|\}\, d\mathbf{u}$$
$$+ \int_{g_{n_1,\cdots,n_d}}^{\mathbf{x}} E\,|K_{\mathbf{h}}(\mathbf{X}-\mathbf{u})|\, d\mathbf{u} = O_{a.s.}\left(n^{-1/2}\log n\right)$$

according to Lemmas 2.5.7 and 2.5.8. Then according to Lemma 2.5.5,

$$\begin{aligned}\left|\hat{F}(\mathbf{x}) - F(\mathbf{x})\right| &\leq \left|\hat{F}(\mathbf{x}) - \hat{F}\left(g_{n_1,\cdots,n_d}\right)\right| + \left|\hat{F}\left(g_{n_1,\cdots,n_d}\right) - F\left(g_{n_1,\cdots,n_d}\right)\right| \\ &\quad + \left|F\left(g_{n_1,\cdots,n_d}\right) - F(\mathbf{x})\right| \\ &= U_{a.s.}\left(n^{-1/2}\log n\right) + U_{a.s.}\left(n^{-1/2}\log n\right) + U(1/n)\end{aligned}$$

and if $\mathbf{X}_1, \cdots, \mathbf{X}_n$ are i.i.d, one can replace $\log n$ in the above inequality by $(\log n)^{1/2}$. □

CHAPTER 3

Spline estimation of a semiparametric GARCH model

3.1 Introduction

It is widely recognized that global smoothing methods such as those by spline or wavelet are computationally much more efficient than local kernel smoothing, see for example the comparison of computing time in Xue and Yang (2006b) and Wang and Yang (2007). Recent development of regression spline smoothing in terms of local asymptotics (Huang (2003)), of high dimensional and weakly dependent data (Huang and Yang (2004), Xue and Yang (2006b) and Wang and Yang (2007)) has presented convincing incentives for applying spline smoothing to solve challenging problems in time series analysis. We have applied cubic spline smoothing to the semiparametric GARCH model (1.2.2), which resulted in a procedure that is a much faster but shares the same theoretical and numerical properties of the kernel smoothing procedure in Yang (2006). Table 3 shows the computing time comparison between the proposed cubic spline method versus the local linear method in estimating parameter α_0. Clearly, the cubic spline method is superior for large sample as its computing time is proportional to n^{-1} of the corresponding time of the local linear method. The advantage of spline method had already been recognized by Engle and Ng (1993), which proposed spline estimation for the news impact curve for extensions of model (1.2.1), without developing justifications by asymptotic theory.

The chapter is organized as follows. In Section 3.2 we discuss the assumptions of the model (1.2.2), the spline estimation of the unknown parameter α_0 and asymptotic properties including its oracle efficiency. In section 3.4 we describe the implementation of the estimator. In sections 4 and 5 we apply the method to simulated and empirical examples. All technical proofs are given in the Appendix.

3.2 Estimation Method

The statistical inference of the semiparametric GARCH model (1.2.2) consists of estimating both parameter α_0 and link function m. In this chapter we focus on estimating the parameter as once α_0 is estimated with $\sqrt{n}$-consistency, the estimation of function m is a routine application of univariate smoothing.

The following assumptions on the data generating process are used

A1: The process $\{Y_t\}_{t=-\infty}^{\infty}$ is strictly stationary, and the innovations $\{\xi_t\}_{t\in\mathbf{Z}}$ have finite r-th absolute moments $E|\xi_t|^r = m_r < \infty$, $0 < r \le 6$.

A2: The link function $m(\cdot)$ is positive everywhere on R_+ and has Lipschitz continuous 4-th derivative.

For convenience, define $X_t = \sum_{j=1}^{\infty} \alpha_0^{j-1} Y_{t-j}^2, t \in \mathbf{Z}$ which simplifies model (1.2.2) to $Y_t = m^{1/2}(X_t)\xi_t, \sigma_t^2 = m(X_t), t \in \mathbf{Z}$ while the process $\{X_t\}_{t=-\infty}^{\infty}$ satisfies the Markovian equation $X_t = \alpha_0 X_{t-1} + m(X_{t-1})\xi_{t-1}^2, t \in \mathbf{Z}$. Since α_0 is an unknown parameter in $(0,1)$, to make numerical optimization feasible, we assume that α_0 lies in the interior of $A = [\alpha_1, \alpha_2]$, where $0 < \alpha_1 < \alpha_2 < 1$, are boundary values known a priori. In practice, one takes sufficiently small α_1 and sufficiently large α_2 based on prior knowledge of the data. Define next $X_{\alpha,t}$ as a series analogous to X_t but with any candidate value of $\alpha \in A$

$$X_{\alpha,t} = \sum_{j=1}^{\infty} \alpha^{j-1} Y_{t-j}^2 = \sum_{j=1}^{\infty} \alpha^j m(X_{t-j})\xi_{t-j}^2, t \in \mathbf{Z}. \tag{3.2.1}$$

We need the following assumptions on the processes $\{X_{\alpha,t}\}_{t=-\infty}^{\infty}, \alpha \in A$.

A3: The processes $\{X_{\alpha,t}\}_{t=-\infty}^{\infty}, \alpha \in A$ are jointly strictly stationary and geometrically α-mixing, i.e., the α -mixing coefficient $\alpha(k) \le c\rho^k$, for constants $c > 0$, $0 < \rho < 1$, where

$\alpha(k) = \sup_{A\in\sigma\left(X_{\alpha,t},t\le 0,\alpha\in A\right),B\in\sigma\left(X_{\alpha,t},t\ge k,\alpha\in A\right)} |P(A)P(B) - P(A\cap B)|.$

From Assumption (A3) and the fact that the innovations $\{\xi_t\}_{t=-\infty}^{\infty}$ are iid, the joint distribution of $(Y_t, \xi_t, X_{\alpha,t}, \alpha \in A)$ is strictly stationary. For each $\alpha \in A$, define the transformed variables for the $X_{\alpha,t}$ as,

$$U_{\alpha,t} = F\left(X_{\alpha,t}\right) = \frac{F_{\alpha_1}\left(X_{\alpha,t}\right) + F_{\alpha_2}\left(X_{\alpha,t}\right)}{2}, 1 \le t \le n \qquad (3.2.2)$$

in which F_{α_1} and F_{α_2} are cdfs of $X_{\alpha_1,t}$ and $X_{\alpha_2,t}$ respectively. In particular, we denote $U_t = U_{\alpha_0,t} = F\left(X_{\alpha_0,t}\right) = F\left(X_t\right)$.

A4: The pdf associated with F is $f\left(x\right) > 0, \forall x \in (0, +\infty)$ and $U_{\alpha,t}$ has a pdf $\varphi_\alpha(\cdot)$ which is Lipschitz continuous and there exist constants c_φ, C_φ such that $\inf_{\alpha\in A,0\le u\le 1}\varphi_\alpha\left(u\right) \ge c_\varphi$ and $\sup_{\alpha\in A,0\le u\le 1}\varphi_\alpha\left(u\right) \le C_\varphi$.

For any $\alpha\in A$ define the predictor of Y_t^2 based on $U_{\alpha,t}$ as $g_\alpha(u) = E(Y_t^2|U_{\alpha,t} = u), 0 < u < 1$. In particular, denote $g(U_t) = g_{\alpha_0}(U_{\alpha_0,t}) = E(Y_t^2|U_{\alpha_0,t}) = m\left(X_t\right)$. Define the risk function of α as $R(\alpha) = E\left\{Y_t^2 - g_\alpha(U_{\alpha,t})\right\}^2$. Apparently $\{Y_t\}_{t=-\infty}^{\infty}$ have finite 4-th moment due to assumption (A1) and (A2). So $R(\alpha)$ allows the usual bias-variance decomposition $R(\alpha) = E\left\{g(U_t) - g_\alpha(U_{\alpha,t})\right\}^2 + (m_4 - 1)Eg^2(U_t)$ which, together with $g(U_t) \equiv g_{\alpha_0}(U_{\alpha_0,t})$, imply that

$$R(\alpha) = E\left\{g(U_t) - g_\alpha(U_{\alpha,t})\right\}^2 + R(\alpha_0) \ge R(\alpha_0), \forall\alpha \in A.$$

We need the following assumption on the function $R(\alpha)$,

A5: The function $R(\alpha)$ has positive second derivative at α_0, i.e., $R''(\alpha_0) > 0$ and $R(\alpha)$ is locally convex at α_0, i.e., for any $\varepsilon > 0$, there exists $\delta > 0$ such that $R(\alpha) - R(\alpha_0) < \delta$ implies $|\alpha - \alpha_0| < \varepsilon$.

Thus by minimizing the prediction error of Y_t^2 on $U_{\alpha,t}$, one should be able to locate the true parameter α consistently via polynomial spline smoothing. To introduce the space of splines, we divide $[0,1]$ into $(N+1)$ subintervals $J_j = \left[t_j, t_{j+1}\right)$, $j = 0, ..., N-1$, $J_N =$

$[t_N, 1]$, where $T := \{t_j\}_{j=1}^{N}$ is a sequence of equally-spaced points, called interior knots, given as

$$t_{1-k} = ... = t_{-1} = t_0 = 0 < t_1 < ... < t_N < 1 = t_{N+1} = ... = t_{N+k},$$

in which $t_j = jh$, $j = 0, 1, ..., N+1, h = 1/(N+1)$ is the distance between neighboring knots. The j-th B-spline of order k for the knot sequence T denoted by $B_{j,k}$ is recursively defined by [14] as

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

for $k > 1$, with

$$B_{j,1}(u) = I_{\{u \in J_j\}} = \begin{cases} 1 & t_j \leq u < t_{j+1} \\ 0 & \text{otherwise} \end{cases}.$$

Define the spaces of linear, quadratic and cubic spline functions on $[0, 1]$ as

$$\Gamma^{(k-2)} = \Gamma^{(k-2)}[0,1] = \left\{ \gamma : \gamma(u) \equiv \sum_{J=1-k}^{N+1} \lambda_J B_{J,k}(u), u \in [0,1] \right\}, k = 2, 3, 4.$$

Given a realization $\{Y_t\}_{t=1}^{n}$, define for $\forall \alpha \in A$ the cubic spline estimator of $g_\alpha(\cdot)$

$$\hat{g}_\alpha(\cdot) = \underset{\gamma \in \Gamma^{(2)}}{\text{argmin}} \frac{1}{n''} \sum_{t=n'+1}^{n} \left\{ Y_t^2 - \gamma(U_{\alpha,t}) \right\}^2$$

with n' and $n'' = n - n'$. We do not use the first n' data points for implementation reasons in Section 3. Define next the empirical risk function

$$\hat{R}(\alpha) = \frac{1}{n''} \sum_{t=n'+1}^{n} \left\{ Y_t^2 - \hat{g}_\alpha(U_{\alpha,t}) \right\}^2$$

and let $\hat{\alpha}$ be the minimizer of $\hat{R}(\alpha)$, i.e.

$$\hat{\alpha} = \underset{\alpha \in A}{\text{argmin}} \hat{R}(\alpha). \tag{3.2.3}$$

We assume the following on the number of interior knots

A6: The number of interior knots N satisfies: $n^{1/6} \ll N = N_n \ll n^{1/5} (\log n)^{-2/5}$.

The next theorem establishes the strong consistency of $\hat{\alpha}$.

THEOREM 3.2.1. *Under assumptions (A1)-(A6), as $n \to \infty$, $\hat{\alpha} \to \alpha_0, a.s.$.*

Proof. According to Proposition 3.2.1, one has $\sup_{\alpha\in A} \left|\hat{R}(\alpha) - R(\alpha)\right| \to 0, a.s.$ Thus there exists an integer $n_0(\varpi)$, such that $\hat{R}(\alpha_0,\varpi) - R(\alpha_0,\varpi) < \delta/2$ when $n > n_0(\varpi)$. Notice that $\hat{\alpha}$ is the minimizer of $\hat{R}(\alpha_0,\varpi)$, so $\hat{R}(\hat{\alpha},\varpi) - R(\alpha_0,\varpi) < \delta/2$. There also exists an integer $n_1(\varpi)$, such that $R(\hat{\alpha},\varpi) - \hat{R}(\hat{\alpha},\varpi) < \delta/2$ when $n > n_1(\varpi)$. Thus, when $n > \max(n_0(\varpi), n_1(\varpi))$,

$$R(\hat{\alpha},\varpi) - R(\alpha_0,\varpi) = R(\hat{\alpha},\varpi) - \hat{R}(\hat{\alpha},\varpi) + \hat{R}(\hat{\alpha},\varpi) - R(\alpha_0,\varpi) < \delta.$$

According to Assumption (A5), R is locally convex at α_0, so for any $\varepsilon > 0$ and any ϖ, if $R(\hat{\alpha},\varpi) - R(\alpha_0,\varpi) < \delta$, then $|\hat{\alpha} \to \alpha_0| < \varepsilon$ for n large enough, which has proved the theorem. □

Denote the asymptotic variance of $\hat{\alpha}$ by the following "sandwich" formula

$$\Sigma(\alpha_0) = R''(\alpha_0)^{-1}\Psi(\alpha_0)R''(\alpha_0)^{-1} \tag{3.2.4}$$

with

$$\Psi(\alpha_0) = \frac{1}{n''}\operatorname{var}\left(\frac{2}{n''}\sum_{t=n'+1}^{n}\xi_{\alpha_0,t}\right) = 4E\left[\left\{g_\alpha(U_{\alpha_0,t}) - Y_t^2\right\}\frac{d}{d\alpha}|_{\alpha=\alpha_0}g_\alpha(U_{\alpha,t})\right]^2 \tag{3.2.5}$$

and $R''(\alpha_0) = \left.\frac{d^2}{d\alpha^2}R(\alpha)\right|_{\alpha=\alpha_0}$

$$= 2E\left[\left\{g_\alpha(U_{\alpha_0,t}) - Y_t^2\right\}\frac{d^2}{d\alpha^2}g_\alpha(U_{\alpha_0,t}) + \left\{\frac{d}{d\alpha}g_\alpha(U_{\alpha_0,t})\right\}^2\right]. \tag{3.2.6}$$

The next theorem establishes $\hat{\alpha}$'s $\sqrt{n}$-asymptotic normality.

THEOREM 3.2.2. *Under assumptions (A1)-(A6), as $n \to \infty$*

$$\sqrt{n}(\hat{\alpha} - \alpha_0) \to_d N(\mathbf{0}, \Sigma(\alpha_0)). \tag{3.2.7}$$

Proof. Denote $\hat{S}(\alpha) = \frac{d}{d\alpha}\hat{R}(\alpha)$ and

$$\xi_{\alpha,t} = \left\{g_\alpha(U_{\alpha,t}) - Y_t^2\right\}\frac{d}{d\alpha}g_\alpha(U_{\alpha,t}) - E\left[\left\{g_\alpha(U_{\alpha,t}) - Y_t^2\right\}\frac{d}{d\alpha}g_\alpha(U_{\alpha,t})\right], \tag{3.2.8}$$

then because $\frac{d}{d\alpha}R(\alpha_0) = 0$, one has

$$\left|\hat{S}(\alpha_0) - \frac{2}{n''}\sum_{t=n'+1}^{n} \xi_{\alpha_0,t}\right| = o\left(n^{-1/2}\right), a.s.. \tag{3.2.9}$$

according to (3.6.37). Mean Value Theorem then implies that for some $t \in [0,1]$

$$\hat{S}(\hat{\alpha}) - \hat{S}(\alpha_0) = \frac{d^2}{d\alpha^2}\hat{R}(t\hat{\alpha} + (1-t)\,\alpha_0)\,(\hat{\alpha} - \alpha_0)$$

and $\hat{S}(\hat{\alpha}) = 0$ because $\hat{R}(\alpha)$ attains its minimum at $\hat{\alpha}$. Thus, one has

$$-\hat{S}(\alpha_0) = \frac{d^2}{d\alpha^2}\hat{R}(t\hat{\alpha} + (1-t)\,\alpha_0)\,(\hat{\alpha} - \alpha_0)$$

i.e.,

$$\hat{\alpha} - \alpha_0 = -\hat{S}(\alpha_0)/\frac{d^2}{d\alpha^2}\hat{R}(t\hat{\alpha} + (1-t)\,\alpha_0).$$

One has

$$\frac{d^2}{d\alpha^2}\hat{R}(t\hat{\alpha} + (1-t)\,\alpha_0) \to \frac{d^2}{d\alpha^2}R(\alpha_0) = R''(\alpha_0), a.s.$$

by Theorem 3.2.1 and Proposition 3.2.1, where $R''(\alpha_0)$ is given in (3.2.6). According to (3.2.9), one has $\sqrt{n''}\hat{S}(\alpha_0) \to_d N\{0, \Psi(\alpha_0)\}$ by the Central Limit Theorem for strongly mixing processes (Theorem 1.7, [4]), where $\Psi(\alpha_0)$ is given in (3.2.5). Then Theorem 3.2.2 is proved by formula (3.2.4) and Slutsky's Theorem. □

The proofs of Theorems 3.2.1 and 3.2.2 given above have made use of complicated arguments involving spline smoothing, summarized in the following proposition, whose proof is given in the Appendix.

PROPOSITION 3.2.1. *Under Assumptions (A1)-(A6), as $n \to \infty$*

$$\sup_{\alpha \in A}\left\|\frac{d^k}{d\alpha^k}\left\{\hat{R}(\alpha) - R(\alpha)\right\}\right\|_\infty \to 0, a.s., k = 0, 1, 2.$$

According to Theorem 3.2.2, the true parameter vector α_0 can be estimated by $\hat{\alpha}$ at $\sqrt{n}$-rate. One can then use the estimate $\hat{\alpha}$ in place of the unknown α_0 for the estimation of function m. We define next the "would-be oracle" estimator of α_0 if the link function g had been "oracally" known $\widetilde{\alpha} = \operatorname{argmin}_{\alpha \in A} \widetilde{R}(\alpha)$, where the oracle empirical risk is $\widetilde{R}(\alpha) = (n'')^{-1}\sum_{t=n'+1}^{n}\left\{Y_t^2 - g(U_{\alpha,t})\right\}^2$, so $\widetilde{\alpha}$ serves as a benchmark of oracle optimality. The next theorem states the asymptotic oracle efficiency of estimator $\hat{\alpha}$.

THEOREM 3.2.3. *Under assumptions (A1)-(A6), as $n \to \infty$, the estimator $\hat{\alpha}$ is asymptotically oracally efficient, i.e., it is asymptotically as efficient as $\tilde{\alpha}$. Specifically, $\sqrt{n}(\tilde{\alpha} - \alpha_0) \to_d N(0, \Sigma(\alpha_0))$ where the variance $\Sigma(\alpha_0)$ is the same as in (3.2.4) and (3.2.7).*

The proof of Theorem 3.2.3 consists of routine arguments in parametric inference, thus it is omitted.

3.3 Implementation

For a given realization $\{Y_t\}_{t=1}^n$, denote in the following two integers

$$n' \doteq \left[2 \log n / \log\left(\alpha_2^{-1}\right)\right] + 1, n'' = n - n'.$$

It is easily verified that

$$\sup_{\alpha \in [\alpha_1, \alpha_2]} \alpha^{n'} = \alpha_2^{n'} < n^{-2}$$

which is the magnitude of error one would incur if the infinite series in (1.2.2) were truncated at n'. In practice, one always has to replace the infinite series of $X_{\alpha,t}$ in (3.2.1) by a finite truncation $\sum_{j=1}^{n'} \alpha^{j-1} Y_{t-j}^2$ for $t \in \mathbb{Z}$, the difference between the two being

$$\sum_{j=n'+1}^{\infty} \alpha^{j-1} Y_{t-j}^2 \leq \sum_{j=n'+1}^{\infty} \alpha_2^{j-1} Y_{t-j}^2 = \alpha_2^{n'} \sum_{j=1}^{\infty} \alpha_2^{j-1} Y_{t-n'-j}^2$$

$$= \alpha_2^{n'} X_{\alpha_2, t-n'} < n^{-2} X_{\alpha_2, t-n'}$$

which is bounded by n^{-2} times of a stationary process with finite variance according to Assumption (A1). Thus instead of computing the infinite sum $\sum_{j=1}^{\infty} j\alpha^{j-1} Y_{t-j}^2$, we use the slowly growing truncation $\sum_{j=1}^{n'} \alpha^{j-1} Y_{t-j}^2$ for implementing the algorithm due to practicality. Also due to practicality, we use $\hat{F}_{\alpha_1}$ and $\hat{F}_{\alpha_2}$ the empirical cdfs of $X_{\alpha_1,t}$ and $X_{\alpha_2,t}$ in place of F_{α_1} and F_{α_2} respectively to compute the transformation function F. Lastly, the number of interior knots $N = N_n$ is computed according to the formula $N = \min\left(10\left[n^{2/11}\right] + 1, n/4 - 1\right)$, which satisfies the Assumption (A6).

We compute the value of $\hat{R}$ over a equally spaced grid of points from α_1 to α_2, and take the one with smallest $\hat{R}$ value as $\hat{\alpha}$ according to (3.2.3). Functions g and m are then estimated using $\hat{\alpha}$ as the true value of α_0. In the next two sections, we present some numerical evidence of how the proposed procedures work for both simulated and real time series.

3.4 Simulation

To investigate the finite-sample precision of the proposed estimator, we applied the procedure to time series data generated according to (1.2.2) with $\alpha_0 = 0.5, A = [\alpha_1, \alpha_2] = [0.1, 0.9]$, and function

$$m(x) = 0.1\,(2x+1)\,/(1-\alpha_0). \tag{3.4.1}$$

Notice that the data generating process actually follows the standard GARCH model, possessing all the known theoretical properties presented in Engle and Ng (1993) and Glosten, Jaganathan and Runkle (1993).

For sample sizes $n = 400, 800, 1600, 3200$, a total of 100 realizations of length $n + 400$ are generated according to model (1.2.2), with functions $m(x)$ as in (3.4.1). For each realization, the last n observations are kept as our data for inference. Truncating the first 400 observations off the series ensures that the remaining series behaves like a stationary one. Estimation of the parameter α_0 is carried out according to the setups described in section 3, using cubic spline.

Table 2 shows the average sum of squared error for $n = 400, 800, 1600, 3200$, that the estimated $\hat{\alpha}$ converges to the true function α_0 as the sample size increases, corroborating the asymptotics in Theorem 3.2.2. In Figure 6, the probability density functions of $\hat{\alpha}$ are estimated by kernel smoothing based on the 100 replications and plotted for $n = 400, 800, 1600, 3200$. Clearly the empirical distribution of $\hat{\alpha}$ quickly collapses to 0, as sample size increases, conforming to Theorem 3.2.2. Since the sample sizes we have used are common for high frequency financial time series such as the two data sets in the next section, the satisfactory numerical performance in Table 2 and Figure 6 provides the assurance we need to apply the procedure to real data.

As discussed in the introduction, Table 3 shows the computing time comparison between the proposed cubic spline method versus the local linear method of Yang (2006) in estimating parameter α_0. Since for each candidate parameter value α, the cubic spline method needs to solve one linear least squares problem in order to compute the empirical risk while the local linear has to solve n, one for each data point, the ratio of their computing times is inversely proportional to n. As a matter of fact, the computing times are of order n and n^2 respectively for the cubic spline and the local linear methods. Since the theoretical properties and numerical performance of the two are similar, the cubic spline method is the one we would recommend for the estimation of parameter α_0. Once the parameter α_0 has been efficiently estimated, the estimation of functions g and m can be done via either kernel type or spline type method, using the estimated parameter value $\hat{\alpha}$ in place of α_0. In the next section, we estimate function g by the Nadaraya-Watson method.

3.5 Applications

In this section, we compare the goodness-of-fit of three models to the daily returns of Deutsche Mark against US Dollar (DEM/USD), and Deutsche Mark against British Pound (DEM/GBP) from January 2, 1980 to October 30, 1992. Both data sets consist of $n = 3212$ observations. The four modelling methods are: the semiparametric GARCH model (1.2.2) with cubic spline estimation method; the semiparametric GARCH model (1.2.2) with kernel estimation method (Yang (2006)); the GJR model of Glosten, Jaganathan and Runkle (1993); the GARCH(1,1) model of Bollerslev (1986). In analyzing the two data sets, a process $\{X_{\alpha,t}\}_{t=1}^{3212}$ is generated for every parameter value α. To have all such processes as close to strict stationarity as possible, we use only the last half for inference. Hence all estimation of parameters and function is done using $\{X_{\alpha,t}\}_{t=1607}^{3212}$ and $\{Y_t^2\}_{t=1607}^{3212}$. The parameter estimate $\hat{\alpha}$ is first obtained according to section 3. In the second step, we use the estimate $\hat{\alpha}$ in place of the unknown α_0 for the Nadaraya-Watson estimation of function g. The volatility forecasts are $\hat{\sigma}_t^2 = \hat{g}_{\hat{\alpha}}\left(U_{\hat{\alpha},t}\right), t = 1607, ..., 3212$, while the residuals are $\hat{\xi}_t = Y_t/\hat{\sigma}_t, t = 1607, ..., 3212$. For the two parametric models, the forecasts and residuals are computed similarly.

In Tables 4 and 5, the goodness-of-fit is compared for all four modelling methods, in terms of volatility prediction error and the log-likelihood, which are calculated respectively as $\sum_{t=1607}^{3212}\left(Y_t^2-\hat{\sigma}_t^2\right)^2/1606$ and $-(1/2)\sum_{t=1607}^{3212}\left\{Y_t^2/\hat{\sigma}_t^2+\ln\left(\hat{\sigma}_t^2\right)\right\}/1606$. The semi-parametric GARCH model (1.2.2) with spline estimation method has best log-likelihood and prediction error for both DEM/GBP and DEM/USD cases. In Tables 7 and 6, the frequencies of the ACF exceeding the significance limits are shown, and they are close enough for the residual absolute powers and for independent normal random samples, and hence one is reasonably sure that there is very little if any dependence left in the residuals.

Figures 7 and 8 represent graphically the fit to DEM/USD, where Figures 7 shows the standardized residuals and Figures 8 shows the estimated functions $\hat{g}_{\hat{\alpha}}$.

3.6 Appendix

3.6.1 Preliminaries

We have collected in this subsection some useful results on strongly-mixing processes and B spline.

We denote by $Q_T(g)$ the 4-th order quasi-interpolant of g corresponding to the knots T, see equation (4.12), page 146 of [15]. According to Theorem 7.7.4, page 225 of [15], the following lemma holds.

LEMMA 3.6.1. *(de Boor 2001, p.149). There exists a constant $C_\infty > 0$ such that for any $g \in C^{(4)}[0,1]$ and $0 \le k \le 2$, $\left\|\left(Q_T(g)-g\right)^{(k)}\right\|_\infty \le C_\infty \left\|g^{(4)}\right\|_\infty H^{4-k}$.*

LEMMA 3.6.2. *(B-spline Property). (i) Partition of Unity. (de boor 2001, page 96) The sequence $\left\{B_{j,k}\right\}_{j=-k+1}^{N}$ provides a positive and local partition of unity, i.e., each $B_{j,k}$ is positive on $\left(t_j, t_{j+k}\right)$, is zero off $\left[t_j, t_{j+k}\right]$, $\sum_{j=-k+1}^{N} B_{j,k}=1$.*

(ii) Differentiation. (de boor 2001, page 116)

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

(iii) Good Condition. (DeVore and Lorentz 1993, Theorem 5.4.2, page 145) There is a constant $D_k > 0$ such that for each spline $S=\sum_{j=-k+1}^{N} c_j B_{j,k}$ of order k and each

$0 < r \leq \infty$,

$$\begin{cases} D_k \left\|c'\right\|_r \leq \|S\|_r \leq \left\|c'\right\|_r, 1 \leq r \leq \infty, \\ D_k \left\|c'\right\|_r \leq \|S\|_r \leq k^{1/r} \left\|c'\right\|_r, 0 < r < 1. \end{cases}$$

For any functions $g_1, g_2 \in L_2[0,1]$, define for $\forall \alpha \in [\alpha_1, \alpha_2]$ the theoretical inner product and norm as

$$\langle g_1, g_2 \rangle_\alpha = \int_0^1 g_1(u) g_2(u) \varphi_\alpha(u)\, du, \ \|g_1\|_{2,\alpha}^2 = \langle g_1, g_1 \rangle_\alpha.$$

LEMMA 3.6.3. *There exist constants $c > 0$ such that for any $\boldsymbol{\lambda} :=$ $(\lambda_{-1,2}, \lambda_{0,2}, ..., \lambda_{N,2}, ..., \lambda_{N,4}) \in R^{3N+9}$.*

$$\begin{cases} ch^{1/r} \|\boldsymbol{\lambda}\|_r \leq \left\|\sum_{k=2}^4 \sum_{j=-k+1}^N \lambda_{j,k} B_{j,k}\right\|_r \leq \left(3^{r-1}kh\right)^{1/r} \|\boldsymbol{\lambda}\|_r, & 1 \leq r \leq \infty, \\ ch^{1/r} \|\boldsymbol{\lambda}\|_r \leq \left\|\sum_{k=2}^4 \sum_{j=-k+1}^N \lambda_{j,k} B_{j,k}\right\|_r \leq (3kh)^{1/r} \|\boldsymbol{\lambda}\|_r, & 0 < r < 1. \end{cases}$$

In particular, under Assumption (A4), $\exists$ constants $c, C \in (0, +\infty)$ such that

$$ch^{1/2} \|\boldsymbol{\lambda}\|_2 \leq \left\|\sum_{k=2}^4 \sum_{j=-k+1}^N \lambda_{j,k} B_{j,k}\right\|_{2,\alpha} \leq Ch^{1/2} \|\boldsymbol{\lambda}\|_2, \forall \alpha \in [\alpha_1, \alpha_2].$$

Proof. It follows from Lemma 3.6.2 (i) that, $\sum_{k=2}^4 \sum_{j=-k+1}^N B_{j,k} \equiv 3$ on $[0,1]$. So the right inequality follows immediate for $r = \infty$. When $1 \leq r < \infty$, Hölder's inequality implies that

$$\left|\sum_{k=2}^4 \sum_{j=-k+1}^N \lambda_{j,k} B_{j,k}\right| \leq 3^{1-1/r} \left(\sum_{k=2}^4 \sum_{j=-k+1}^N \left|\lambda_{j,k}\right|^r B_{j,k}\right)^{1/r}.$$

Since all the knots are equally spaced, Lemma 3.6.2 (i) ensures that $\int_{-\infty}^{\infty} B_{j,k}(u)\, du \leq kh$, the right inequality follows from $\int_0^1 \left|\sum_{k=2}^4 \sum_{j=-k+1}^N \lambda_{j,k} B_{j,k}(u)\right|^r du \leq 3^{r-1} kh \|\boldsymbol{\lambda}\|_r^r$. When $r < 1$, we have $\left|\sum_{k=2}^4 \sum_{j=-k+1}^N \lambda_{j,k} B_{j,k}\right|^r$ $\leq \sum_{k=2}^4 \sum_{j=-k+1}^N \left|\lambda_{j,k}\right|^r B_{j,k}^r$. Since $\int_{-\infty}^{\infty} B_{j,k}^r(u)\, du \leq t_{j+k} - t_j = kh$ and

$$\int_0^1 \left|\sum_{k=2}^4 \sum_{j=-k+1}^N \lambda_{j,k} B_{j,k}(u)\right|^r du \leq \|\boldsymbol{\lambda}\|_r^r \int_{-\infty}^{\infty} B_{j,k}^r(u)\, du \leq 3kh \|\boldsymbol{\lambda}\|_r^r,$$

the right inequality follows in this case as well. For the left inequalities, we derive from Lemma 3.6.2 (iii), for any $0 < r \leq \infty$

$$\left|\lambda_{j,k}\right|^r \leq C_1^r h^{-1} \int_{t_j}^{t_{j+1}} \left|\sum_{j=-k+1}^N \lambda_{j,k} B_{j,k}(u)\right|^r du.$$

Since each $u \in [0,1]$ appears in at most k intervals (t_j, t_{j+k}), adding up these inequalities, we obtain that

$$\begin{aligned}\|\lambda\|_r^r &\leq C_1 h^{-1} \sum_{k=1}^{4} \int_{t_j}^{t_{j+k}} \left|\sum\nolimits_{j=-k+1}^{N} \lambda_{j,k} B_{j,k}(u)\right|^r du \\ &\leq 3Ch^{-1} \left\|\sum\nolimits_{j=-k+1}^{N} \lambda_{j,k} B_{j,k}\right\|_r^r .\end{aligned}$$

The left inequality follows. □

Given a realization $\{Y_t\}_{t=1}^n$, define for any functions $g_1, g_2 \in L_2[0,1]$ and any $\alpha \in [\alpha_1, \alpha_2]$ the empirical inner product and norm as

$$\langle g_1, g_2 \rangle_{n,\alpha} = (n'')^{-1} \sum_{t=n'+1}^{n} g_1(U_{\alpha,t}) g_2(U_{\alpha,t}), \ \|g_1\|_{2,n,\alpha}^2 = \langle g_1, g_1 \rangle_{n,\alpha} .$$

LEMMA 3.6.4. *Under Assumptions (A3), (A4) and (A6), as $n \to \infty$, with probability* 1

$$\sup_{\alpha \in A} \max_{\substack{k,k'=2,3,4 \\ 1-k \leq j \leq N, 1-k' \leq j' \leq N}} \left| \left\langle B_{j,k}, B_{j',k'} \right\rangle_{n,\alpha} - \left\langle B_{j,k}, B_{j',k'} \right\rangle_{\alpha} \right| = O\left\{(nN)^{-1/2} \log n\right\}.$$

Proof. We only prove the case $k = k' = 4$, all other cases are similar. Let

$$\zeta_{\alpha,j,j',t} = B_{j,4}(U_{\alpha,t}) B_{j',4}(U_{\alpha,t}) - EB_{j,4}(U_{\alpha,t}) B_{j',4}(U_{\alpha,t})$$

with the second moment

$$E\zeta_{\alpha,j,j',t}^2 = E\left[B_{j,4}^2(U_{\alpha,t}) B_{j',4}^2(U_{\alpha,t})\right] - \left\{EB_{j,4}(U_{\alpha,t}) B_{j',4}(U_{\alpha,t})\right\}^2$$

where $E\left[B_{j,4}^2(U_{\alpha,t}) B_{j',4}^2(U_{\alpha,t})\right] \sim N^{-1}$, $\left\{EB_{j,4}(U_{\alpha,t}) B_{j',4}(U_{\alpha,t})\right\}^2 \sim N^{-2}$ uniformly for all $-3 \leq j, j' \leq N$ by Assumption (A4). Hence, $E\zeta_{\alpha,j,j',t}^2 \sim N^{-1}$ uniformly for all $-3 \leq j, j' \leq N$. The k−th moment is

$$\begin{aligned}E\left|\zeta_{\alpha,j,j',t}\right|^k &= E\left|B_{j,4}(U_{\alpha,t}) B_{j',4}(U_{\alpha,t}) - EB_{j,4}(U_{\alpha,t}) B_{j',4}(U_{\alpha,t})\right|^k \\ &\leq 2^{k-1}\left\{E\left|B_{j,4}(U_{\alpha,t}) B_{j',4}(U_{\alpha,t})\right|^k + \left|EB_{j,4}(U_{\alpha,t}) B_{j',4}(U_{\alpha,t})\right|^k\right\}\end{aligned}$$

where $E\left|B_{j,4}(U_{\alpha,t}) B_{j',4}(U_{\alpha,t})\right|^k \sim N^{-1}$, $\left|EB_{j,4}(U_{\alpha,t}) B_{j',4}(U_{\alpha,t})\right|^k \sim N^{-k}$ uniformly for all $-3 \leq j, j' \leq N$. Thus, there exists a constant $C > 0$ such that $E\left|\zeta_{\alpha,j,j',t}\right|^k \leq$

$C2^{k-1}k!E\zeta^2_{\alpha,j,j',t}$ for all $-3 \le j, j' \le N$. So Cramér's condition in Lemma 2.5.2 is satisfied, one has for $\delta_n = \delta \log n/\sqrt{nN}$ and fixed α

$$P\left\{\frac{1}{n''}\left|\sum\nolimits_{t=n'+1}^{n}\zeta_{\alpha,j,j',t}\right| > \delta_n\right\} \le n^{-10}. \tag{3.6.1}$$

We divide interval $[\alpha_1, \alpha_2]$ into n^6 equally spaced intervals with disjoint endpoints $\alpha_1 = a_1 < \cdots < a_{M_n} = \alpha_2$ and $\sup_{\alpha\in A}\max_{-3\le j,j'\le N}\left|\zeta_{\alpha,j,j',t}\right|$ is bounded by

$$\sup_{1\le r\le M_n}\max_{-3\le j,j'\le N}\left|\zeta_{a_r,j,j',t}\right| + \max_{-3\le j,j'\le N}\sup_{1\le r\le M_n}\max_{\alpha\in[a_r,a_{r+1}]}\left|\zeta_{\alpha,j,j',t} - \zeta_{a_r,j,j',t}\right|. \tag{3.6.2}$$

While (3.6.1) implies that

$$\sup_{1\le r\le M_n}\max_{-3\le j,j'\le N}\left(n''\right)^{-1}\left|\sum\nolimits_{t=n'+1}^{n}\zeta_{a_r,j,j',t}\right|, a.s. \tag{3.6.3}$$

by Borel-Cantelli Lemma. Employing Lipschitz continuity of the cubic B-spline, one has with probability 1

$$\begin{aligned}&\max_{-3\le j,j'\le N}\sup_{1\le r\le M_n}\max_{\alpha\in[a_r,a_{r+1}]}\left|\left(n''\right)^{-1}\sum\nolimits_{t=n'+1}^{n}\left(\zeta_{\alpha,j,j',t} - \zeta_{a_r,j,j',t}\right)\right| \\ &= O\left(M_n^{-1}h^{-6}\right).\end{aligned} \tag{3.6.4}$$

Therefore Assumption A4, (3.6.2), (3.6.3) and (3.6.4) lead to the result. □

Denote by $\Gamma = \Gamma^{(0)} \cup \Gamma^{(1)} \cup \Gamma^{(2)}$ the space of all linear, quadratic and cubic spline functions on $[0,1]$. We establish the uniform rate at which the empirical inner product approximates the theoretical inner product for all B-splines $B_{j,k}$ with $k = 2,3,4$.

LEMMA 3.6.5. *Under Assumptions (A3), (A4) and (A6), as $n \to \infty$, one has*

$$A_n = \sup_{\alpha\in A}\sup_{\gamma_1,\gamma_2\in\Gamma}\left|\frac{\langle\gamma_1,\gamma_2\rangle_{n,\alpha} - \langle\gamma_1,\gamma_2\rangle_\alpha}{\|\gamma_1\|_{2,\alpha}\|\gamma_2\|_{2,\alpha}}\right| = O\left\{(nh)^{-1/2}\log n\right\}, a.s.. \tag{3.6.5}$$

Proof. Denote $\gamma_a = \sum_{k=2}^{4}\sum_{j=-k+1}^{N}\gamma_{a,jk}B_{j,k}$, $a = 1,2$, without loss of generality. Then

$$\begin{aligned}\langle\gamma_1,\gamma_2\rangle_{n,\alpha} &= \sum_{k=2}^{4}\sum_{j=-k+1}^{N}\sum_{k'=2}^{4}\sum_{j'=-k+1}^{N}\gamma_{1,j,k}\gamma_{2,j'k'}\left\langle B_{j,k}, B_{j',k'}\right\rangle_{n,\alpha}, \\ \|\gamma_1\|^2_{2,\alpha} &= \sum_{k=2}^{4}\sum_{j=-k+1}^{N}\sum_{k'=2}^{4}\sum_{j'=-k+1}^{N}\gamma_{1,j,k}\gamma_{1,j'k'}\left\langle B_{j,k}, B_{j',k'}\right\rangle_{\alpha}, \\ \|\gamma_2\|^2_{2,\alpha} &= \sum_{k=2}^{4}\sum_{j=-k+1}^{N}\sum_{k'=2}^{4}\sum_{j'=-k+1}^{N}\gamma_{2,j,k}\gamma_{2,j'k'}\left\langle B_{j,k}, B_{j',k'}\right\rangle_{\alpha}.\end{aligned}$$

Let $\gamma_1 = (\gamma_{1,-1,2}, \gamma_{1,0,2}, ..., \gamma_{1,N,2}, ..., \gamma_{1,N,4})$, $\gamma_2 = (\gamma_{2,-1,2}, \gamma_{2,0,2}, ..., \gamma_{2,N,2}, ..., \gamma_{2,N,4})$. According to Lemma 3.6.3, one has for any $\alpha \in [\alpha_1, \alpha_2]$,

$$ch \|\gamma_1\|_2^2 \le \|\gamma_1\|_{2,\alpha}^2 \le Ch \|\gamma_1\|_2^2, ch \|\gamma_2\|_2^2 \le \|\gamma_2\|_{2,\alpha}^2 \le Ch \|\gamma_2\|_2^2,$$

$$ch \|\gamma_1\|_2 \|\gamma_2\|_2 \le \|\gamma_1\|_{2,\alpha} \|\gamma_2\|_{2,\alpha} \le Ch \|\gamma_1\|_2 \|\gamma_2\|_2.$$

Hence

$$\begin{aligned} A_n &= \sup_{\alpha\in A} \sup_{\gamma_1\in\gamma,\gamma_2\in\Gamma} \left| \frac{\langle\gamma_1,\gamma_2\rangle_{n,\alpha} - \langle\gamma_1,\gamma_2\rangle_\alpha}{\|\gamma_1\|_{2,\alpha}\|\gamma_2\|_{2,\alpha}} \right| \le \frac{\|\gamma_1\|_\infty \|\gamma_2\|_\infty}{c_1 h \|\gamma_1\|_2 \|\gamma_2\|_2} \\ &\quad \times \sup_{\alpha\in A} \max_{\substack{k,k'=2,3,4 \\ 1-k\le j\le N, 1-k'\le j'\le N}} \left| \frac{1}{n}\sum_{i=1}^n \left\{ \left\langle B_{j,k}, B_{j',k'} \right\rangle_{n,\alpha} - \left\langle B_{j,k}, B_{j',k'} \right\rangle_\alpha \right\} \right| \\ &\le c_0 h^{-1} \sup_{\alpha\in A} \max_{\substack{k,k'=2,3,4 \\ 1-k\le j\le N, 1-k'\le j'\le N}} \left| \frac{1}{n}\sum\nolimits_{i=1}^n \left\{ \left\langle B_{j,k}, B_{j',k'} \right\rangle_{n,\alpha} - \left\langle B_{j,k}, B_{j',k'} \right\rangle_\alpha \right\} \right|, \end{aligned}$$

which, together with Lemma 3.6.4, imply (3.6.5). □

For any fixed α, one has $\mathbf{Y}^2 = \mathbf{g}_\alpha + \mathbf{g} - \mathbf{g}_\alpha + \mathbf{E} = \mathbf{g}_\alpha + \mathbf{E}_\alpha + \mathbf{E}$, where $\mathbf{E}^T = \{g(U_t)(\xi_t^2 - 1)\}_{t=n'+1}^n, \mathbf{E}_\alpha = \{g(U_t) - g_\alpha(U_{\alpha,t})\}_{t=n'+1}^n$. Then one can break the cubic spline estimation error as

$$\hat{g}_\alpha(u) - g_\alpha(u) = \tilde{g}_\alpha(u) - g_\alpha(u) + \tilde{\varepsilon}_\alpha(u) + \hat{\varepsilon}_\alpha(u), \tag{3.6.6}$$

where

$$\begin{aligned} \tilde{g}_\alpha(u) &= \{B_{j,4}(u)\}_{-3\le j\le N}^T \mathbf{V}_{n,\alpha}^{-1} \left\{ \langle \mathbf{g}_\alpha, B_{j,4}\rangle_{n,\alpha} \right\}_{j=-3}^N, \\ \tilde{\varepsilon}_\alpha(u) &= \{B_{j,4}(u)\}_{-3\le j\le N}^T \mathbf{V}_{n,\alpha}^{-1} \left\{ \langle \mathbf{E}_\alpha, B_{j,4}\rangle_{n,\alpha} \right\}_{j=-3}^N, \\ \hat{\varepsilon}_\alpha(u) &= \{B_{j,4}(u)\}_{-3\le j\le N}^T \mathbf{V}_{n,\alpha}^{-1} \left\{ \langle \mathbf{E}, B_{j,4}\rangle_{n,\alpha} \right\}_{j=-3}^N, \\ \mathbf{V}_{n,\alpha} &= \left\{ \left\langle B_{j,4}, B_{j',4} \right\rangle_{n,\alpha} \right\}_{j,j'=-3}^N, \mathbf{V}_\alpha = \left\{ \left\langle B_{j,4}, B_{j',4} \right\rangle_\alpha \right\}_{j,j'=-3}^N. \end{aligned} \tag{3.6.7}$$

The next proposition is used in proving Proposition 3.2.1.

PROPOSITION 3.6.1. *Under Assumptions (A1)-(A4), (A6), as* $n \to \infty$

$$\sup_{\alpha\in A} \sup_{u\in[0,1]} |\hat{g}_\alpha(u) - g_\alpha(u)| = O\left\{ (nh)^{-1/2} \log n + h^4 \right\}, a.s., \tag{3.6.8}$$

$$\sup_{\alpha\in A}\max_{n'+1\le t\le n}\left|\frac{d}{d\alpha}\left\{\hat{g}_\alpha\left(U_{\alpha,t}\right)-g_\alpha\left(U_{\alpha,t}\right)\right\}\right| = O\left\{n^{-1/2}h^{-3/2}\log n + h^3\right\}, a.s., \quad (3.6.9)$$

$$\sup_{\alpha\in A}\left|\frac{d^2}{d\alpha^2}\left\{\hat{g}_\alpha\left(U_{\alpha,t}\right)-g_\alpha\left(U_{\alpha,t}\right)\right\}\right| = O\left\{n^{-1/2}h^{-5/2}\log n + h^2\right\}, a.s.. \quad (3.6.10)$$

In order to prove the above proposition, we need several technical lemmas. The following is a special case of Theorem 13.4.3 in [15].

LEMMA 3.6.6. *If a bi-infinite matrix with bandwidth r has a bounded inverse* $\mathbf{A}^{-1}$ *on* l_2 *and* $\kappa = \kappa(\mathbf{A}) = \|\mathbf{A}\|_2 \|\mathbf{A}^{-1}\|_2$ *is the condition number of* $\mathbf{A}$, *then* $\|\mathbf{A}^{-1}\|_\infty \le 2c_0(1-v)^{-1}$, *with* $c_0 = v^{-2r}\|\mathbf{A}\|_2$, $v = \left(\kappa^2-1\right)^{1/4r}\left(\kappa^2+1\right)^{-1/4r}$.

LEMMA 3.6.7. *Under Assumptions (A3), (A4) and (A6), there exist constants* $0 < c_{\mathbf{V}} < C_{\mathbf{V}}$ *such that*

$$c_{\mathbf{V}}N^{-1}\|\mathbf{w}\|_2^2 \le \mathbf{w}^T\mathbf{V}_\alpha\mathbf{w} \le C_{\mathbf{V}}N^{-1}\|\mathbf{w}\|_2^2 \quad (3.6.11)$$

$$c_{\mathbf{V}}N^{-1}\|\mathbf{w}\|_2^2 \le \mathbf{w}^T\mathbf{V}_{n,\alpha}\mathbf{w} \le C_{\mathbf{V}}N^{-1}\|\mathbf{w}\|_2^2 \quad (3.6.12)$$

with matrices $\mathbf{V}_\alpha$ *and* $\mathbf{V}_{n,\alpha}$ *defined in (3.6.7). In addition, there exists a constant* $C > 0$ *such that*

$$\sup_{\alpha\in A}\left\|\mathbf{V}_{n,\alpha}^{-1}\right\|_\infty \le CN, a.s., \sup_{\alpha\in A}\left\|\mathbf{V}_\alpha^{-1}\right\|_\infty \le CN. \quad (3.6.13)$$

Proof. Let $\mathbf{w}$ be any $(N+4)$-vector and $\gamma_{\mathbf{w}}(u) = \sum_{j=-3}^{N} w_J B_{j,4}(u)$, then $\mathbf{B}_\alpha\mathbf{w} = \left\{\gamma_{\mathbf{w}}\left(U_{\alpha,n'}\right), ..., \gamma_{\mathbf{w}}\left(U_{\alpha,n-1}\right)\right\}$ and A_n in (3.6.5) entails that

$$\|\gamma_{\mathbf{w}}\|_{2,\alpha}^2(1-A_n) \le \mathbf{w}^T\mathbf{V}_{n,\alpha}\mathbf{w} \le \|\gamma_{\mathbf{w}}\|_{2,\alpha}^2(1+A_n). \quad (3.6.14)$$

By Theorem 5.4.2 of [15] and Assumption (A4), one has

$$c_\varphi\frac{C}{N}\|\mathbf{w}\|_2^2 \le \mathbf{w}^T\mathbf{V}_\alpha\mathbf{w} \le C_\varphi\frac{C}{N}\|\mathbf{w}\|_2^2 \quad (3.6.15)$$

which, together with (3.6.14), yield

$$c_\varphi\frac{C}{N}\|\mathbf{w}\|_2^2(1-A_n) \le \mathbf{w}^T\mathbf{V}_{n,\alpha}\mathbf{w} \le C_\varphi\frac{C}{N}\|\mathbf{w}\|_2^2(1+A_n).$$

Then one has (3.6.11) and (3.6.12) by (3.6.15), (3.6.14) and (3.6.5). Next, denote by $\lambda_{\max}\left(\mathbf{V}_{n,\alpha}\right)$ and $\lambda_{\min}\left(\mathbf{V}_{n,\alpha}\right)$ the maximum and minimum eigenvalue of $\mathbf{V}_{n,\alpha}$, then

$$c_V N^{-1} \le \|\mathbf{V}_{n,\alpha}\|_2\left\|\mathbf{V}_{n,\alpha}^{-1}\right\|_2 = \lambda_{\max}\left(\mathbf{V}_{n,\alpha}\right), \lambda_{\min}^{-1}\left(\mathbf{V}_{n,\alpha}\right) = \left\|\mathbf{V}_{n,\alpha}^{-1}\right\|_2 \le C_V N^{-1}, a.s.,$$

thus $\kappa = \|\mathbf{V}_{n,\alpha}\|_2 = \lambda_{\max}(\mathbf{V}_{n,\alpha})/\lambda_{\min}(\mathbf{V}_{n,\alpha}) = C_V/c_V < \infty, a.s.$ One can also show that $\kappa \geq C > 1, a.s.$. Combining the above and Lemma 3.6.6 with $v = (\kappa^2-1)^{1/16}(\kappa^2+1)^{-1/16}$, one gets $\|\mathbf{V}_{n,\alpha}^{-1}\|_\infty \leq 2v^{-8}N(1-v)^{-1} = CN, a.s.$, which is part one of (3.6.13). Part two of (3.6.13) can be proved similarly. □

3.6.2 Proof of Proposition 3.6.1

LEMMA 3.6.8. *Under Assumptions (A2)-(A4) and (A6), as $n \to \infty$*

$$\sup_{\alpha\in A}\left\|(\tilde{g}_\alpha - g_\alpha)^{(k)}\right\|_\infty \leq C\left\|m^{(4)}\right\|_\infty h^{4-k}, a.s., 0 \leq k \leq 2. \tag{3.6.16}$$

Proof. According to Theorem A.1 of [36], there exists an absolute constant $C > 0$, such that

$$\sup_{\alpha\in A}\|\tilde{g}_\alpha - g_\alpha\|_\infty \leq C\sup_{\alpha\in A}\inf_{\gamma\in\Gamma^{(2)}}\|\gamma - g_\alpha\|_\infty \leq C\left\|m^{(4)}\right\|_\infty h^{4-k}, a.s.,$$

which proves for the case $k = 0$. Applying Lemma 3.6.1, one has for $0 \leq k \leq 2$

$$\sup_{\alpha\in A}\left\|(Q_T(g_\alpha) - g_\alpha)^{(k)}\right\|_\infty \leq C\sup_{\alpha\in A}\left\|g_\alpha^{(4)}\right\|_\infty h^{4-k} \leq C\left\|m^{(4)}\right\|_\infty h^{4-k}, a.s.. \tag{3.6.17}$$

So $\sup_{\alpha\in A}\|Q_T(g_\alpha) - \tilde{g}_\alpha\|_\infty \leq C\left\|m^{(4)}\right\|_\infty h^4$ a.s., which entails that

$$\sup_{\alpha\in A}\left\|(Q_T(g_\alpha) - \tilde{g}_\alpha)^{(k)}\right\|_\infty \leq C\left\|m^{(4)}\right\|_\infty h^{4-k}, a.s., 0 \leq k \leq 2. \tag{3.6.18}$$

Then the lemma is proved by combining (3.6.17) and (3.6.18). □

Denote $\mathbf{B}_\alpha = \{B_{j,4}(U_{\alpha,t})\}_{t=n'+1,j=-3}^{n,N}$ and

$$\mathbf{P}_\alpha = \mathbf{B}_\alpha\left(\mathbf{B}_\alpha^T\mathbf{B}_\alpha\right)^{-1}\mathbf{B}_\alpha^T \tag{3.6.19}$$

as the projection matrix onto the cubic spline space spanned by $\Gamma^{(2)}$, and $\dot{\mathbf{B}}_\alpha = d\mathbf{B}_\alpha/d\alpha, \dot{\mathbf{P}}_\alpha = d\mathbf{P}_\alpha/d\alpha$.

LEMMA 3.6.9. *Under Assumptions (A4), one has*

$$\dot{\mathbf{B}}_\alpha = \left[\{B_{j,3}(U_{\alpha,t}) - B_{j+1,3}(U_{\alpha,t})\} f(X_{\alpha,t}) h^{-1}\sum\nolimits_{j=1}^{\infty} j\alpha^{j-1}Y_{t-j}^2\right]_{t=n'+1,j=-3}^{n,N}, \tag{3.6.20}$$

$$\dot{\mathbf{P}}_\alpha = (\mathbf{I} - \mathbf{P}_\alpha)\dot{\mathbf{B}}_\alpha\left(\mathbf{B}_\alpha^T\mathbf{B}_\alpha\right)^{-1}\mathbf{B}_\alpha^T + \mathbf{B}_\alpha\left(\mathbf{B}_\alpha^T\mathbf{B}_\alpha\right)^{-1}\dot{\mathbf{B}}_\alpha^T(\mathbf{I} - \mathbf{P}_\alpha). \tag{3.6.21}$$

Proof. Property (ii) in Lemma 3.6.2 implies that

$$\dot{\mathbf{B}}_\alpha = \left\{\frac{d}{d\alpha}B_{j,4}\left(U_{\alpha,t}\right)\right\}_{t=n'+1,j=-3}^{n,N} = \left\{\frac{d}{du}B_{j,4}\left(U_{\alpha,t}\right)\frac{d}{d\alpha}U_{\alpha,t}\right\}_{t=n'+1,j=-3}^{n,N}$$

$$= \left[3\left\{\frac{B_{j,3}\left(U_{\alpha,t}\right)}{t_{j+3}-t_j} - \frac{B_{j+1,3}\left(U_{\alpha,t}\right)}{t_{j+4}-t_{j+1}}\right\} f\left(X_{\alpha,t}\right) h^{-1} \sum\nolimits_{j=1}^{\infty} j\alpha^{j-1} Y_{t-j}^2\right]_{t=n'+1,j=-3}^{n,N}$$

$$= \left[\left\{B_{j,3}\left(U_{\alpha,t}\right) - B_{j+1,3}\left(U_{\alpha,t}\right)\right\} f\left(X_{\alpha,t}\right) h^{-1} \sum\nolimits_{j=1}^{\infty} j\alpha^{j-1} Y_{t-j}^2\right]_{t=n'+1,j=-3}^{n,N}.$$

Next, note that

$$\dot{\mathbf{P}}_\alpha = \dot{\mathbf{B}}_\alpha \left(\mathbf{B}_\alpha^T \mathbf{B}_\alpha\right)^{-1} \mathbf{B}_\alpha^T + \mathbf{B}_\alpha \frac{d}{d\alpha}\left\{\left(\mathbf{B}_\alpha^T \mathbf{B}_\alpha\right)^{-1}\right\} \mathbf{B}_\alpha^T + \mathbf{B}_\alpha \left(\mathbf{B}_\alpha^T \mathbf{B}_\alpha\right)^{-1} \dot{\mathbf{B}}_\alpha^T$$

and

$$\begin{aligned}\frac{d}{d\alpha}\left\{\left(\mathbf{B}_\alpha^T \mathbf{B}_\alpha\right)^{-1}\right\} &= -\left(\mathbf{B}_\alpha^T \mathbf{B}_\alpha\right)^{-1} \frac{d}{d\alpha}\left(\mathbf{B}_\alpha^T \mathbf{B}_\alpha\right)\left(\mathbf{B}_\alpha^T \mathbf{B}_\alpha\right)^{-1} \\ &= -\left(\mathbf{B}_\alpha^T \mathbf{B}_\alpha\right)^{-1}\left(\dot{\mathbf{B}}_\alpha^T \mathbf{B}_\alpha + \mathbf{B}_\alpha^T \dot{\mathbf{B}}_\alpha\right)\left(\mathbf{B}_\alpha^T \mathbf{B}_\alpha\right)^{-1}. \quad (3.6.22)\end{aligned}$$

Hence $\dot{\mathbf{P}}_\alpha$ is

$$\begin{aligned}&\dot{\mathbf{B}}_\alpha \left(\mathbf{B}_\alpha^T \mathbf{B}_\alpha\right)^{-1} \mathbf{B}_\alpha^T - \mathbf{B}_\alpha \left(\mathbf{B}_\alpha^T \mathbf{B}_\alpha\right)^{-1} \mathbf{B}_\alpha^T \dot{\mathbf{B}}_\alpha \left(\mathbf{B}_\alpha^T \mathbf{B}_\alpha\right)^{-1} \mathbf{B}_\alpha^T \\ &-\mathbf{B}_\alpha \left(\mathbf{B}_\alpha^T \mathbf{B}_\alpha\right)^{-1} \dot{\mathbf{B}}_\alpha^T \mathbf{B}_\alpha \left(\mathbf{B}_\alpha^T \mathbf{B}_\alpha\right)^{-1} \mathbf{B}_\alpha^T + \mathbf{B}_\alpha \left(\mathbf{B}_\alpha^T \mathbf{B}_\alpha\right)^{-1} \dot{\mathbf{B}}_\alpha^T \\ = \;& \left(\mathbf{I} - \mathbf{P}_\alpha\right) \dot{\mathbf{B}}_\alpha \left(\mathbf{B}_\alpha^T \mathbf{B}_\alpha\right)^{-1} \mathbf{B}_\alpha^T + \mathbf{B}_\alpha \left(\mathbf{B}_\alpha^T \mathbf{B}_\alpha\right)^{-1} \dot{\mathbf{B}}_\alpha^T \left(\mathbf{I} - \mathbf{P}_\alpha\right). \quad (3.6.23)\end{aligned}$$

□

LEMMA 3.6.10. *Under Assumptions (A3), (A4) and (A6), as* $n \to \infty$

$$\sup_{\alpha \in A} \left\|\left(n''\right)^{-1} \mathbf{B}_\alpha^T\right\|_\infty \leq Ch, \sup_{\alpha \in A} \left\|\left(n''\right)^{-1} \dot{\mathbf{B}}_\alpha^T\right\|_\infty \leq C, a.s. \quad (3.6.24)$$

$$\sup_{\alpha \in A} \left\|\mathbf{P}_\alpha\right\|_\infty \leq C, \sup_{\alpha \in A} \left\|\dot{\mathbf{P}}_\alpha\right\|_\infty \leq Ch, \sup_{\alpha \in A} \left\|\frac{d}{d\alpha}\left\{\left(\mathbf{B}_\alpha^T \mathbf{B}_\alpha\right)^{-1}\right\}\right\|_\infty = O\left(N\right), a.s. \quad (3.6.25)$$

Proof. For any vector $\mathbf{a} \in R^{n''}$, one has

$$\left\|\left(n''\right)^{-1} \mathbf{B}_\alpha^T \mathbf{a}\right\|_\infty \leq \left\|\mathbf{a}\right\|_\infty \max_{-3 \leq j \leq N} \left|\left(n''\right)^{-1} \sum\nolimits_{t=n'+1}^{n} B_{j,4}\left(U_{\alpha,t}\right)\right| \leq Ch \left\|\mathbf{a}\right\|_\infty, a.s.$$

and using equation (3.6.20), $\left\|\left(n''\right)^{-1} \dot{\mathbf{B}}_\alpha^T \mathbf{a}\right\|_\infty$ is bounded with probability 1 by

$$\left\|\mathbf{a}\right\|_\infty \max_{-3 \leq j \leq N} \left|\left(n''h\right)^{-1} \sum_{t=n'+1}^{n} \left\{\left(B_{j,3} - B_{j+1,3}\right)\left(U_{\alpha,t}\right)\right\} f\left(X_{\alpha,t}\right) h^{-1} \sum_{j=1}^{\infty} j\alpha^{j-1} Y_{t-j}^2\right|$$

$\leq C \|\mathbf{a}\|_\infty$. Then one has (3.6.25) by (3.6.19), (3.6.13), (3.6.24), (3.6.23) and (3.6.22). Equations (3.6.20) and (3.6.21) are needed for proving the rest of the inequalities. □

LEMMA 3.6.11. *Under Assumptions (A2)-(A4) and (A6),*

$$\sup_{\alpha \in A} \left| \frac{d^k}{d\alpha^k} \left\{ \tilde{g}_\alpha (U_{\alpha,t}) - g_\alpha (U_{\alpha,t}) \right\} \right| \leq C \left\| m^{(4)} \right\|_\infty h^{4-k}, a.s., k = 1, 2. \tag{3.6.26}$$

Proof. According to the definition of $\tilde{g}_\alpha$ in (3.2.3), one has

$$\frac{d}{d\alpha} \left[\left\{ Q_T (g_\alpha) - \tilde{g}_\alpha \right\} (U_{\alpha,t}) \right] = \frac{d}{d\alpha} \mathbf{P}_\alpha \left[\left\{ Q_T (g_\alpha) - g_\alpha \right\} (U_{\alpha,t}) \right]$$

$$= \dot{\mathbf{P}}_\alpha \left[\left\{ Q_T (g_\alpha) - g_\alpha \right\} (U_{\alpha,t}) \right] + \mathbf{P}_\alpha \frac{d}{d\alpha} \left[\left\{ Q_T (g_\alpha) - g_\alpha \right\} (U_{\alpha,t}) \right],$$

$$\frac{d}{d\alpha} \left[\left\{ Q_T (g_\alpha) - g_\alpha \right\} (U_{\alpha,t}) \right] = \left[\left\{ Q_T \left(\frac{d}{d\alpha} g_\alpha \right) - \frac{d}{d\alpha} g_\alpha \right\} (U_{\alpha,t}) \right]$$

$$+ \left[\frac{d}{du} \left\{ Q_T (g_\alpha) - g_\alpha \right\} (U_{\alpha,t}) \right] f (X_{\alpha,t}) h^{-1} \sum_{j=1}^{\infty} j \alpha^{j-1} Y_{t-j}^2,$$

which yield (3.6.26) for $k = 1$ by (3.6.17) and (3.6.25). The proof for $k = 2$ is similar. □

LEMMA 3.6.12. *Under Assumptions (A2)-(A4) and (A6), as $n \to \infty$, one has with probability* 1

$$\sup_{\alpha \in A} \left\| \frac{\mathbf{B}_\alpha^T \mathbf{E}}{n''} \right\|_\infty = O\left(\frac{\log n}{\sqrt{nN}} \right), \sup_{\alpha \in A} \left\| \frac{\mathbf{B}_\alpha^T \mathbf{E}_\alpha}{n''} \right\|_\infty = O\left(\frac{\log n}{\sqrt{nN}} \right), \tag{3.6.27}$$

$$\sup_{\alpha \in A} \left\| \frac{d}{d\alpha} \left(\frac{\mathbf{B}_\alpha^T \mathbf{E}}{n''} \right) \right\|_\infty = O\left(\frac{\log n}{\sqrt{nh}} \right), \sup_{\alpha \in A} \left\| \frac{d}{d\alpha} \left(\frac{\mathbf{B}_\alpha^T \mathbf{E}_\alpha}{n''} \right) \right\|_\infty = O\left(\frac{\log n}{\sqrt{nh}} \right), \tag{3.6.28}$$

$$\sup_{\alpha \in A} \left\| \frac{\dot{\mathbf{B}}_\alpha^T \mathbf{E}}{n''} \right\|_\infty = O\left(\frac{\log n}{\sqrt{nh}} \right), \sup_{\alpha \in A} \left\| \frac{\dot{\mathbf{B}}_\alpha^T \mathbf{E}_\alpha}{n''} \right\|_\infty = O\left(\frac{\log n}{\sqrt{nh}} \right). \tag{3.6.29}$$

Proof. we prove only the first equation in (3.6.27) and the second equation of (3.6.28), other equations can be proved similarly. One has

$$\frac{\mathbf{B}_\alpha^T \mathbf{E}}{n''} = \left[(n'')^{-1} \sum\nolimits_{t=n'}^{n} B_{j,4} (U_{\alpha,t}) g (U_t) \left(\xi_t^2 - 1 \right) \right]_{j=-3}^{N}.$$

Denote $Z_t = g (U_t) (\xi_t^2 - 1) = Z_{t,1}^{D_n} + Z_{t,2}^{D_n} + Z_{t,3}^{D_n}$, where $D_n = n^\eta \, (1/3 < \eta < 2/5)$,

$$Z_{t,1}^{D_n} = g (U_t) \left(\xi_t^2 - 1 \right) I \left\{ \left| g (U_t) \left(\xi_t^2 - 1 \right) \right| > D_n \right\},$$

$$Z_{t,2}^{D_n} = g (U_{\alpha,t}) \left(\xi_t^2 - 1 \right) I \left\{ \left| g (U_t) \left(\xi_t^2 - 1 \right) \right| \leq D_n \right\} - Z_{t,3}^{D_n},$$

$$Z_{t,3}^{D_n} = E\left[g\left(U_t\right)\left(\xi_t^2-1\right)I\left\{\left|g\left(U_t\right)\left(\xi_t^2-1\right)\right| \le D_n\right\}\right].$$

Note that the B-spline basis is bounded, so it is straightforward to verify that the mean of the truncated part is uniformly bounded by D_n^{-2}

$$\sup_{\alpha\in A}\left|\left(n''\right)^{-1}\sum\nolimits_{t=n'}^{n} B_{j,4}\left(U_{\alpha,t}\right)Z_{t,3}^{D_n}\right| = O\left(D_n^{-2}\right) = o\left(n^{-2/3}\right).$$

One has $\sum_{n=n'+1}^{\infty} P\left\{\left|g\left(U_{n-1}\right)\left(\xi_n^2-1\right)\right| > D_n\right\} \le \sum_{n=n'+1}^{\infty} D_n^{-3} < \infty$ according to the assumption that $E\left(\xi_t^6\right) = m_6 < +\infty$, and Borel-Cantelli Lemma implies that the tail part

$$\left|\left(n''\right)^{-1}\sum\nolimits_{t=n'+1}^{n} B_{j,4}\left(U_{\alpha,t}\right)Z_{t,1}^{D_n}\right| = O\left(n^{-k}\right), \text{ for any } k>0.$$

For the truncated part, using Lemma 2.5.2 and discretization, one has

$$\left|\left(n''\right)^{-1}\sum\nolimits_{t=n'+1}^{n} B_{j,4}\left(U_{\alpha,t}\right)Z_{t,2}^{D_n}\right| = O\left(\log n/\sqrt{Nn}\right).$$

Therefore the first equation in (3.6.27) is established with probability 1. To prove the second equation of (3.6.28), notice that

$$\frac{\mathbf{B}_\alpha^T\mathbf{E}_\alpha}{n''} = \left[\left(n''\right)^{-1}\sum\nolimits_{t=n'+1}^{n} B_{j,4}\left(U_{\alpha,t}\right)\left\{g\left(U_t\right)-g_\alpha\left(U_{\alpha,t}\right)\right\}\right]_{j=-3}^{N},$$

$$\frac{d}{d\alpha}\left(\frac{\mathbf{B}_\alpha^T\mathbf{E}_\alpha}{n''}\right) = \left[\left(n''\right)^{-1}\sum\nolimits_{t=n'+1}^{n} \frac{d}{d\alpha}\left[B_{j,4}\left(U_{\alpha,t}\right)\left\{g\left(U_t\right)-g_\alpha\left(U_{\alpha,t}\right)\right\}\right]\right]_{j=-3}^{N}.$$

While $E\left[B_{j,4}\left(U_{\alpha,t}\right)\left\{g\left(U_t\right)-g_\alpha\left(U_{\alpha,t}\right)\right\}\right] = 0, -3 \le j \le N$ implies that

$$E\left\{\frac{d}{d\alpha}\left[B_{j,4}\left(U_{\alpha,t}\right)\left\{g\left(U_t\right)-g_\alpha\left(U_{\alpha,t}\right)\right\}\right]\right\} = 0, -3 \le j \le N, \alpha\in A,$$

which allows one to apply Lemma 2.5.2 to obtain that with probability one

$$\sup_{\alpha\in A}\left\|\frac{d}{d\alpha}\left(\frac{\mathbf{B}_\alpha^T\mathbf{E}_\alpha}{n''}\right)\right\|_\infty = O\left(\log n/\sqrt{nh}\right).$$

□

LEMMA 3.6.13. *Under Assumptions (A2)-(A4) and (A6), as* $n \to \infty$

$$\sup_{\alpha\in A}\sup_{u\in[0,1]}\left|\hat{\varepsilon}_\alpha(u)\right| = O\left(\log n/\sqrt{nh}\right), a.s., \tag{3.6.30}$$

$$\sup_{\alpha\in A}\sup_{u\in[0,1]}\left|\tilde{\varepsilon}_\alpha(u)\right| = O\left(\log n/\sqrt{nh}\right), a.s.. \tag{3.6.31}$$

Proof. We only prove (3.6.30), the proof of (3.6.31) is similar. Denote $\hat{\mathbf{a}} = (\hat{a}_{-3}, ..., \hat{a}_N)^T = \left(\mathbf{B}_\alpha^T \mathbf{B}_\alpha\right)^{-1} \mathbf{B}_\alpha^T \mathbf{E}$

$= \mathbf{V}_{n,\alpha}^{-1} \left\{ (n'')^{-1} \mathbf{B}_\alpha^T \mathbf{E} \right\}$, then $\hat{\varepsilon}_\alpha(u) = \sum_{j=-3}^{N} \hat{a}_j B_{j,4}(u)$.

$$\sup_{\alpha \in A} \sup_{u \in [0,1]} |\hat{\varepsilon}_\alpha(u)| \le \sup_{\alpha \in A} \|\hat{\mathbf{a}}\|_\infty = \sup_{\alpha \in A} \left\| \mathbf{V}_{n,\alpha}^{-1} \left(n^{-1} \mathbf{B}_\alpha^T \mathbf{E} \right) \right\|_\infty$$

$$\le CN \sup_{\alpha \in A} \left\| (n'')^{-1} \mathbf{B}_\alpha^T \mathbf{E} \right\|_\infty, a.s.,$$

where the last inequality follows from Lemmas 3.6.7 and 3.6.12. □

LEMMA 3.6.14. *Under Assumptions (A2)-(A4) and (A6), as $n \to \infty$*

$$\sup_{\alpha \in A} \max_{n'+1 \le t \le n} \left| \frac{d}{d\alpha} \hat{\varepsilon}_\alpha(U_{\alpha,t}) \right| = O\left(n^{-1/2} N^{3/2} \log n \right), a.s., \tag{3.6.32}$$

$$\sup_{\alpha \in A} \max_{n'+1 \le t \le n} \left| \frac{d}{d\alpha} \tilde{\varepsilon}_\alpha(U_{\alpha,t}) \right| = O\left(n^{-1/2} N^{3/2} \log n \right), a.s., \tag{3.6.33}$$

$$\sup_{\alpha \in A} \max_{n'+1 \le t \le n} \left| \frac{d^2}{d\alpha^2} \hat{\varepsilon}_\alpha(U_{\alpha,t}) \right| = O\left(n^{-1/2} N^{5/2} \log n \right), a.s., \tag{3.6.34}$$

$$\sup_{\alpha \in A} \max_{n'+1 \le t \le n} \left| \frac{d^2}{d\alpha^2} \tilde{\varepsilon}_\alpha(U_{\alpha,t}) \right| = O\left(n^{-1/2} N^{5/2} \log n \right), a.s.. \tag{3.6.35}$$

Proof. We only prove (3.6.32) and (3.6.33), the proofs of (3.6.34) and (3.6.35) are similar. One has $\left\{ \frac{d}{d\alpha} \hat{\varepsilon}_\alpha(U_{\alpha,t}) \right\}_{t=n'+1}^{n}$

$$= (\mathbf{I} - \mathbf{P}_\alpha) \dot{\mathbf{B}}_\alpha \left(\mathbf{B}_\alpha^T \mathbf{B}_\alpha \right)^{-1} \mathbf{B}_\alpha^T \mathbf{E} + \mathbf{B}_\alpha \left(\mathbf{B}_\alpha^T \mathbf{B}_\alpha \right)^{-1} \dot{\mathbf{B}}_\alpha^T (\mathbf{I} - \mathbf{P}_\alpha) \mathbf{E}$$

$$= (\mathbf{I} - \mathbf{P}_\alpha) \dot{\mathbf{B}}_\alpha \left(\frac{\mathbf{B}_\alpha^T \mathbf{B}_\alpha}{n} \right)^{-1} \frac{\mathbf{B}_\alpha^T \mathbf{E}}{n} + \mathbf{B}_\alpha \left(\frac{\mathbf{B}_\alpha^T \mathbf{B}_\alpha}{n} \right)^{-1} \frac{\dot{\mathbf{B}}_\alpha^T (\mathbf{I} - \mathbf{P}_\alpha) \mathbf{E}}{n}.$$

According to (3.6.13), (3.6.24), (3.6.25) and (3.6.27), one has (3.6.32). To prove (3.6.33), note that $\left\{ \frac{d}{d\alpha} \tilde{\varepsilon}_\alpha(U_{\alpha,t}) \right\}_{t=n'+1}^{n}$

$$\begin{aligned}
&= (\mathbf{I} - \mathbf{P}_\alpha) \dot{\mathbf{B}}_\alpha \left(\mathbf{B}_\alpha^T \mathbf{B}_\alpha \right)^{-1} \mathbf{B}_\alpha^T \mathbf{E}_\alpha + \mathbf{B}_\alpha^T \left(\mathbf{B}_\alpha^T \mathbf{B}_\alpha \right)^{-1} \dot{\mathbf{B}}_\alpha (\mathbf{I} - \mathbf{P}_\alpha) \mathbf{E}_\alpha \\
&\quad + \mathbf{B}_\alpha \left(\mathbf{B}_\alpha^T \mathbf{B}_\alpha \right)^{-1} \mathbf{B}_\alpha^T \frac{d}{d\alpha} \mathbf{E}_\alpha \\
&= (\mathbf{I} - \mathbf{P}_\alpha) \dot{\mathbf{B}}_\alpha \left(\mathbf{B}_\alpha^T \mathbf{B}_\alpha \right)^{-1} \mathbf{B}_\alpha^T \mathbf{E}_\alpha - \mathbf{B}_\alpha^T \left(\mathbf{B}_\alpha^T \mathbf{B}_\alpha \right)^{-1} \dot{\mathbf{B}}_\alpha \mathbf{P}_\alpha \mathbf{E}_\alpha \\
&\quad + \mathbf{B}_\alpha^T \left(\mathbf{B}_\alpha^T \mathbf{B}_\alpha \right)^{-1} \dot{\mathbf{B}}_\alpha \mathbf{E}_\alpha + \mathbf{B}_\alpha \left(\mathbf{B}_\alpha^T \mathbf{B}_\alpha \right)^{-1} \mathbf{B}_\alpha^T \frac{d}{d\alpha} \mathbf{E}_\alpha \\
&= T_1 + T_2
\end{aligned}$$

where

$$T_1 = \left\{(\mathbf{I}-\mathbf{P}_\alpha)\dot{\mathbf{B}}_\alpha - \mathbf{B}_\alpha^T\left(\mathbf{B}_\alpha^T\mathbf{B}_\alpha\right)^{-1}\dot{\mathbf{B}}_\alpha\mathbf{B}_\alpha^T\right\}\left(\mathbf{B}_\alpha^T\mathbf{B}_\alpha\right)^{-1}\mathbf{B}_\alpha^T\mathbf{E}_\alpha$$
$$= \left\{(\mathbf{I}-\mathbf{P}_\alpha)\dot{\mathbf{B}}_\alpha - \mathbf{B}_\alpha^T\left(\frac{\mathbf{B}_\alpha^T\mathbf{B}_\alpha}{n}\right)^{-1}\frac{\dot{\mathbf{B}}_\alpha\mathbf{B}_\alpha^T}{n}\right\}\left(\frac{\mathbf{B}_\alpha^T\mathbf{B}_\alpha}{n}\right)^{-1}\frac{\mathbf{B}_\alpha^T\mathbf{E}_\alpha}{n}$$
$$T_2 = \mathbf{B}_\alpha\left(\frac{\mathbf{B}_\alpha^T\mathbf{B}_\alpha}{n}\right)^{-1}\frac{d}{d\alpha}\left(\frac{\mathbf{B}_\alpha^T\mathbf{E}_\alpha}{n}\right).$$

By (3.6.13), (3.6.24), (3.6.25) (3.6.27) and (3.6.28), one has $\sup_{\alpha\in A}\|T_1\|_\infty = O\left(n^{-1/2}N^{3/2}\log n\right)$ and $\sup_{\alpha\in A}\|T_2\|_\infty$
$= O\left(n^{-1/2}N^{3/2}\log n\right)$, *a.s.* which leads to (3.6.33). □

Proof of Proposition 3.6.1. According to (3.6.6), one has (3.6.8) by (3.6.16), (3.6.30) and (3.6.31). Similarly, one has

$$\frac{d}{d\alpha}\left\{\hat{g}_\alpha\left(U_{\alpha,t}\right) - g_\alpha\left(U_{\alpha,t}\right)\right\} = \frac{d}{d\alpha}\left\{\tilde{g}_\alpha\left(U_{\alpha,t}\right) - g_\alpha\left(U_{\alpha,t}\right)\right\} + \frac{d}{d\alpha}\tilde{\varepsilon}_\alpha(U_{\alpha,t}) + \frac{d}{d\alpha}\hat{\varepsilon}_\alpha(U_{\alpha,t}).$$

Thus one has (3.6.9) by (3.6.16), (3.6.32) and (3.6.33). The proof of (3.6.10) is similar. □

3.6.3 Proof of Proposition 3.2.1

LEMMA 3.6.15. *Under Assumptions (A1)-(A6), as* $n \to \infty$, $\sup_{\alpha\in A}\left|\hat{R}(\alpha) - R(\alpha)\right| = o(1)$, *a.s.*.

Proof.

$$\hat{R}(\alpha) = \frac{1}{n-n'}\sum_{t=n'+1}^{n}\left\{Y_t^2 - \hat{g}_\alpha(U_{\alpha,t})\right\}^2$$
$$= \frac{1}{n-n'}\sum_{t=n'+1}^{n}\left\{g(U_t) + g(U_t)\left(\xi_t^2 - 1\right) - g_\alpha(U_{\alpha,t}) + g_\alpha(U_{\alpha,t}) - \hat{g}_\alpha(U_{\alpha,t})\right\}^2$$

$$
\begin{aligned}
&= \frac{1}{n-n'}\sum_{t=n'+1}^{n}\{g_\alpha(U_{\alpha,t})-\hat g_\alpha(U_{\alpha,t})\}^2+\frac{1}{n-n'}\sum_{t=n'+1}^{n}\{g(U_t)-g_\alpha(U_{\alpha,t})\}^2\\
&+\frac{2}{n-n'}\sum_{t=n'+1}^{n}\{g(U_t)-g_\alpha(U_{\alpha,t})\}\left\{g(U_t)\left(\xi_t^2-1\right)\right\}\\
&+\frac{1}{n-n'}\sum_{t=n'+1}^{n}\left\{g(U_t)\left(\xi_t^2-1\right)\right\}^2\\
&+\frac{2}{n-n'}\sum_{t=n'+1}^{n}\{g_\alpha(U_{\alpha,t})-\hat g_\alpha(U_{\alpha,t})\}\left\{g(U_t)-g_\alpha(U_{\alpha,t})+g(U_t)\left(\xi_t^2-1\right)\right\},
\end{aligned}
$$

$$
\begin{aligned}
R(\alpha) &= E\left\{Y_t^2-g_\alpha(U_{\alpha,t})\right\}^2\\
&= E\left\{g(U_t)+g(U_t)\left(\xi_t^2-1\right)-g_\alpha(U_{\alpha,t})\right\}^2\\
&= E\{g(U_t)-g_\alpha(U_{\alpha,t})\}^2+E\left\{g(U_t)\left(\xi_t^2-1\right)\right\}^2.
\end{aligned}
$$

Hence

$$
\sup_{\alpha\in A}\left|\hat R(\alpha)-R(\alpha)\right|\le I_1+I_2+I_3+I_4
$$

where

$$
I_1=\sup_{\alpha\in A}\left|\frac{1}{n-n'}\sum_{t=n'+1}^{n}\{g_\alpha(U_{\alpha,t})-\hat g_\alpha(U_{\alpha,t})\}^2\right|,
$$

$$
I_2=\sup_{\alpha\in A}\left|\frac{2}{n-n'}\sum_{t=n'+1}^{n}\{g_\alpha(U_{\alpha,t})-\hat g_\alpha(U_{\alpha,t})\}\left\{g(U_t)-g_\alpha(U_{\alpha,t})+g(U_t)\left(\xi_t^2-1\right)\right\}\right|,
$$

$$
I_3=\sup_{\alpha\in A}\left|\frac{1}{n-n'}\sum_{t=n'+1}^{n}\{g(U_t)-g_\alpha(U_{\alpha,t})\}^2-E\{g(U_t)-g_\alpha(U_{\alpha,t})\}^2\right|,
$$

$$
\begin{aligned}
I_4 &= \sup_{\alpha\in A}\left\{\left|\frac{1}{n-n'}\sum_{t=n'+1}^{n}\left\{g(U_t)\left(\xi_t^2-1\right)\right\}^2-(m_4-1)\,Eg^2(U_t)\right|\right.\\
&\left.+\left|\frac{2}{n-n'}\sum_{t=n'+1}^{n}\{g(U_t)-g_\alpha(U_{\alpha,t})\}\left\{g(U_t)\left(\xi_t^2-1\right)\right\}\right|\right\}.
\end{aligned}
$$

According to Lemma 2.5.2, one has $I_3+I_4=o(1)\,,a.s$, and (3.6.8) entails that $I_1=O\left\{\left(n^{-1/2}\log n\right)^2+\left(H^4\right)^2\right\},a.s..$ One also has

$$
I_2\le O\left(n^{-1/2}\log n+H^4\right)\sup_{\alpha\in A}\frac{2}{n''}\sum_{t=n'+1}^{n}\left|\left\{g(U_t)-g_\alpha(U_{\alpha,t})+g(U_t)\left(\xi_t^2-1\right)\right\}\right|,
$$

which is $O\left(n^{-1/2}\log n + H^4\right)$, *a.s.*. The lemma is proved by combining I_1, I_2, I_3, I_4. □

LEMMA 3.6.16. *Under Assumptions (A1)-(A6), as $n \to \infty$, one has for $k = 1, 2$*

$$\sup_{\alpha\in A}\left|\frac{d^k}{d\alpha^k}\left(\hat{R}(\alpha) - R(\alpha)\right)\right| = O\left(n^{-1/2}h^{-1/2-k}\log n + h^{4-k}\right), a.s.. \tag{3.6.36}$$

Proof. Note that

$$\begin{aligned}\frac{1}{2}\frac{d}{d\alpha}\hat{R}(\alpha) &= \frac{1}{n''}\sum_{t=n'+1}^{n}\left\{\hat{g}_\alpha(U_{\alpha,t}) - Y_t^2\right\}\frac{d}{d\alpha}\hat{g}_\alpha(U_{\alpha,t}),\\ \frac{1}{2}\frac{d}{d\alpha}R(\alpha) &= E\left[\left\{g_\alpha(U_{\alpha,t}) - Y_t^2\right\}\frac{d}{d\alpha}g_\alpha(U_{\alpha,t})\right],\end{aligned}$$

then

$$\frac{1}{2}\frac{d}{d\alpha}\left(\hat{R}(\alpha) - R(\alpha)\right) = \frac{1}{n''}\sum_{t=n'+1}^{n}\xi_{\alpha,t} + J_{\alpha,1} + J_{\alpha,2} + J_{\alpha,3}$$

where $\xi_{\alpha,t}$ is defined in (3.2.8) and $E\xi_{\alpha,t} = 0$, and where

$$\begin{aligned}J_{\alpha,1} &= \frac{1}{n''}\sum_{t=n'+1}^{n}\left\{\hat{g}_\alpha(U_{\alpha,t}) - g_\alpha(U_{\alpha,t})\right\}\frac{d}{d\alpha}\left(\hat{g}_\alpha - g_\alpha\right)(U_{\alpha,t}),\\ J_{\alpha,2} &= \frac{1}{n''}\sum_{t=n'+1}^{n}\left\{g_\alpha(U_{\alpha,t}) - Y_t^2\right\}\frac{d}{d\alpha}\left(\hat{g}_\alpha - g_\alpha\right)(U_{\alpha,t}),\\ J_{\alpha,1} &= \frac{1}{n''}\sum_{t=n'+1}^{n}\left\{\hat{g}_\alpha(U_{\alpha,t}) - g_\alpha(U_{\alpha,t})\right\}\frac{d}{d\alpha}g_\alpha(U_{\alpha,t}).\end{aligned}$$

By Lemma 2.5.2, $\sup_{\alpha\in A}\left|(n'')^{-1}\sum_{t=n'+1}^{n}\xi_{\alpha,t}\right| = O\left(n^{-1/2}\log n\right)$ *a.s.*. Meanwhile, (3.6.8) and (3.6.9) imply that $\sup_{\alpha\in A}\left|J_{\alpha,1}\right| = O\left(n^{-1}h^{-2}\log^2 n + h^7\right)$ *a.s.*. Note that

$$\begin{aligned}J_{\alpha,2} &= \frac{1}{n''}\sum_{t=n'+1}^{n}\left\{g_\alpha(U_{\alpha,t}) - Y_t^2\right\}\frac{d}{d\alpha}\left(\tilde{g}_\alpha - g_\alpha\right)(U_{\alpha,t})\\ &\quad - \frac{1}{n''}\left(\mathbf{E}_\alpha + \mathbf{E}\right)^T\frac{d}{d\alpha}\left\{\mathbf{P}_\alpha\left(\mathbf{E}_\alpha + \mathbf{E}\right)\right\}.\end{aligned}$$

One has

$$\sup_{\alpha\in A}\left|J_{\alpha,2} + \frac{1}{n''}\left(\mathbf{E}_\alpha + \mathbf{E}\right)^T\frac{d}{d\alpha}\left\{\mathbf{P}_\alpha\left(\mathbf{E}_\alpha + \mathbf{E}\right)\right\}\right| = O\left(h^3\right)\ a.s.$$

according to (3.6.16). Next

$$\left|\frac{1}{n''}\left(\mathbf{E}_\alpha + \mathbf{E}\right)^T\frac{d}{d\alpha}\left\{\mathbf{P}_\alpha\left(\mathbf{E}_\alpha + \mathbf{E}\right)\right\}\right|$$

$$= \left| \frac{1}{n''} (\mathbf{E}_\alpha + \mathbf{E})^T \frac{d}{d\alpha} \left\{ \mathbf{B}_\alpha \left(\frac{\mathbf{B}_\alpha^T \mathbf{B}_\alpha}{n''} \right)^{-1} \frac{\mathbf{B}_\alpha^T}{n''} (\mathbf{E}_\alpha + \mathbf{E}) \right\} \right|$$

$$\leq \left| \frac{1}{n''} (\mathbf{E}_\alpha + \mathbf{E})^T \dot{\mathbf{B}}_\alpha \left(\frac{\mathbf{B}_\alpha^T \mathbf{B}_\alpha}{n''} \right)^{-1} \frac{\mathbf{B}_\alpha^T}{n''} (\mathbf{E}_\alpha + \mathbf{E}) \right|$$

$$+ \left| \frac{1}{n''} (\mathbf{E}_\alpha + \mathbf{E})^T \mathbf{B}_\alpha \left(\frac{\mathbf{B}_\alpha^T \mathbf{B}_\alpha}{n''} \right)^{-1} \frac{d}{d\alpha} \frac{\mathbf{B}_\alpha^T}{n''} (\mathbf{E}_\alpha + \mathbf{E}) \right|$$

$$+ \left| \frac{1}{n''} (\mathbf{E}_\alpha + \mathbf{E})^T \mathbf{B}_\alpha \frac{d}{d\alpha} \left\{ \left(\frac{\mathbf{B}_\alpha^T \mathbf{B}_\alpha}{n''} \right)^{-1} \right\} \frac{\mathbf{B}_\alpha^T}{n''} (\mathbf{E}_\alpha + \mathbf{E}) \right|.$$

Thus

$$\sup_{\alpha \in A} \left| \frac{1}{n''} (\mathbf{E}_\alpha + \mathbf{E})^T \frac{d}{d\alpha} \{ \mathbf{P}_\alpha (\mathbf{E}_\alpha + \mathbf{E}) \} \right|$$

$$\leq O(N) \times \sup_{\alpha \in A} \left\{ \left\| \frac{1}{n''} (\mathbf{E}_\alpha + \mathbf{E})^T \dot{\mathbf{B}}_\alpha \right\|_\infty \left\| \left(\frac{\mathbf{B}_\alpha^T \mathbf{B}_\alpha}{n''} \right)^{-1} \right\|_\infty \left\| \frac{\mathbf{B}_\alpha^T}{n''} (\mathbf{E}_\alpha + \mathbf{E}) \right\|_\infty \right\}$$

$$+ O(N) \times \sup_{\alpha \in A} \left\{ \left\| \frac{1}{n''} (\mathbf{E}_\alpha + \mathbf{E})^T \mathbf{B}_\alpha \right\|_\infty \left\| \left(\frac{\mathbf{B}_\alpha^T \mathbf{B}_\alpha}{n''} \right)^{-1} \right\|_\infty \left\| \frac{d}{d\alpha} \frac{\mathbf{B}_\alpha^T}{n''} (\mathbf{E}_\alpha + \mathbf{E}) \right\|_\infty \right\}$$

$$+ O(N) \times \sup_{\alpha \in A} \left\{ \left\| \frac{1}{n''} (\mathbf{E}_\alpha + \mathbf{E})^T \mathbf{B}_\alpha \right\|_\infty \left\| \frac{d}{d\alpha} \left\{ \left(\frac{\mathbf{B}_\alpha^T \mathbf{B}_\alpha}{n''} \right)^{-1} \right\} \right\|_\infty \left\| \frac{\mathbf{B}_\alpha^T}{n''} (\mathbf{E}_\alpha + \mathbf{E}) \right\|_\infty \right\}$$

$$= O(N) \times O\left(\log n / \sqrt{nh} \right) \times O(N) \times O\left(\log n / \sqrt{nN} \right) = O\left(n^{-1} N^2 \log^2 n \right) \ a.s.$$

according to (3.6.27), (3.6.28), (3.6.29), (3.6.13) and (3.6.25). So $\sup_{\alpha \in A} |J_{\alpha,2}| = O\left(n^{-1} N^2 \log^2 n + h^3 \right)$, *a.s.*. Similarly, one can write

$$J_{\alpha,3} = \frac{1}{n''} \sum_{t=n'+1}^{n} \{ \tilde{g}_\alpha(U_{\alpha,t}) - g_\alpha(U_{\alpha,t}) \} \frac{d}{d\alpha} g_\alpha(U_{\alpha,t})$$
$$+ \frac{1}{n''} (\mathbf{E}_\alpha + \mathbf{E})^T \mathbf{B}_\alpha \left(\frac{\mathbf{B}_\alpha^T \mathbf{B}_\alpha}{n} \right)^{-1} \frac{\mathbf{B}_\alpha^T}{n} \frac{d}{d\alpha} g_\alpha.$$

and has

$$\sup_{\alpha \in A} \left| \frac{1}{n''} \sum_{t=n'+1}^{n} \{ \tilde{g}_\alpha(U_{\alpha,t}) - g_\alpha(U_{\alpha,t}) \} \frac{d}{d\alpha} g_\alpha(U_{\alpha,t}) \right| = O\left(h^4 \right) \ a.s.,$$

$$\sup_{\alpha\in A}\left|\frac{1}{n''}(\mathbf{E}_\alpha+\mathbf{E})^T\mathbf{B}_\alpha\left(\frac{\mathbf{B}_\alpha^T\mathbf{B}_\alpha}{n}\right)^{-1}\frac{\mathbf{B}_\alpha^T}{n}\frac{d}{d\alpha}g_\alpha\right|$$
$$= O\left\{\frac{\log n}{\sqrt{nN}}\times N\times h\right\}=O\left\{\frac{\log n}{\sqrt{nN}}\right\}\ a.s..$$

Thus (3.6.36) is proved for $k=1$. One can prove that for the term $\xi_{\alpha,t}$ defined in (3.2.8), with probability 1

$$\sup_{\alpha\in A}\left|\frac{1}{2}\frac{d}{d\alpha}\left\{\hat{R}(\alpha)-R(\alpha)\right\}-\frac{1}{n''}\sum_{t=n'+1}^{n}\xi_{\alpha,t}\right|=o\left(n^{-1/2}\right). \tag{3.6.37}$$

The proof of (3.6.36) for $k=2$ follows from (3.6.8), (3.6.9) and (3.6.10), since

$$\frac{1}{2}\frac{d^2}{d\alpha^2}\hat{R}(\alpha) = \frac{1}{n''}\sum_{t=n'+1}^{n}\left[\left\{\hat{g}_\alpha(U_{\alpha,t})-Y_t^2\right\}\frac{d^2}{d\alpha^2}\hat{g}_\alpha(U_{\alpha,t})+\frac{d}{d\alpha}\hat{g}_\alpha(U_{\alpha,t})\frac{d}{d\alpha}\hat{g}_\alpha(U_{\alpha,t})\right],$$
$$\frac{1}{2}\frac{d^2}{d\alpha^2}R(\alpha) = E\left[\left\{g_\alpha(U_{\alpha,t})-Y_t^2\right\}\frac{d^2}{d\alpha^2}g_\alpha(U_{\alpha,t})+\frac{d}{d\alpha}g_\alpha(U_{\alpha,t})\frac{d}{d\alpha}g_\alpha(U_{\alpha,t})\right].$$

□

Proof of Proposition 3.2.1. It follows from Lemma 3.6.15 and Lemma 3.6.16. □

CHAPTER 4

Spline-backfitted kernel smoothing of additive coefficient model

4.1 Introduction

This chapter is based on Liu and Yang (2009). Model (1.3.1)'s versatility for econometric applications is illustrated by the following example. Consider the forecasting of US GDP annual growth rate, which is modelled as the Total Factor Productivity (TFP) growth rate plus a linear function of capital growth rate and labor growth rate, according to the classic Cobb-Douglas model (Cobb and Douglas, 1928). As pointed out in Li and Racine (2007), p. 302, it is unrealistic to ignore the non neutral effect of R&D spending on the TFP growth rate and on the complementary slopes of capital and labor growth rates. Thus a smooth coefficient model should fit the production function better than the parametric Cobb-Douglas model. Indeed, Figure 9 shows that a smooth coefficient model has much smaller rolling forecast errors than the parametric Cobb-Douglas model, based on data from 1959 to 2002. In addition, Figure 10 shows that the TFP growth rate is a function of R&D spending, not a constant.

Many methods exist for the estimation of functional/varying coefficient models, see Cai, Fan and Yao (2000), Yang, Park, Xue and Härdle (2006) for kernel type estimators, Huang, Wu and Zhou (2002), Huang and Shen (2004) for spline estimators. These published works have partial success in addressing the inaccuracy of estimating multivariate nonparametric functions, commonly known as the "curse of dimensionality". Typically, optimal convergence rates of the coefficient function estimators are established, locally for kernel estimators, or globally for spline estimators.

Our view is that a satisfactory procedure for estimating the functions $\{m_{\alpha l}(x_\alpha)\}_{l=1,\alpha=1}^{d_1,d_2}$ and constants $\{m_{0l}\}_{l=1}^{d_1}$ in model (1.3.1) should meet three broad criteria. Specifically, the procedure should be (i) computationally expedient; (ii) theoretically reliable and (iii) intuitively appealing. As model (1.3.1) is a natural extension of additive model, we extend the "spline-backfitted kernel smoothing" of Wang and Yang (2007) to additive coefficient model, combining the best features of both kernel and spline methods. Kernel procedures for additive model, such as Yang, Härdle and Nielsen (1999), Sperlich, Tjøstheim and Yang (2002), Yang, Sperlich and Härdle (2003), Rodríguez-Póo, Sperlich and Vieu (2003), Hengartner and Sperlich (2005) satisfy criterion (iii) and partly (ii) as they are asymptotically normal at any given point, but not (i) since they are extremely computationally intensive when either the dimension is high or sample size is large, as illustrated in the Monte-Carlo results of Wang and Yang (2007). Spline approaches of Stone (1985), Huang (1998a,b), Huang and Yang (2004) to additive model, on the other hand, do not satisfy criterion (ii) as they lack limiting distribution, but are fast to compute, thus satisfying (i). In addition, none of the published works had established "uniform convergence rate", thus lacking in regard to (ii). The spline-backfitted kernel (SBK) and spline-backfitted local linear (SBLL) estimators we propose are essentially as fast and accurate as an univariate kernel and local linear smoothing, thus completely satisfying all three criteria (i)-(iii). Other alternatives for estimating model (1.3.1) that may satisfy criteria (i)-(iii) are possible extensions of the smoothed backfitting of Mammen, Linton & Nielsen (1999) and Nielsen & Sperlich (2005), and the two-stage estimator of Horowitz and Mammen (2004). It is important to note that although Horowitz and Mammen (2004) had used B spline in simulation, their theoretical proof was for what should be called "orthogonal series-backfitted local linear" estimator in our parlance.

We now describe the oracle smoothing idea of Linton (1997) in the context of model (1.3.1). If all the nonparametric functions of the last $d_2 - 1$ variables, $\{m_{\alpha l}(x_\alpha)\}_{l=1,\alpha=2}^{d_1,d_2}$ and all the constants $\{m_{0l}\}_{l=1}^{d_1}$ were known by "oracle", one could define a new variable

$Y_{,1} = \sum_{l=1}^{d_1} m_{1l}(X_1) T_l + \sigma(\mathbf{X}, \mathbf{T}) \varepsilon = Y - \sum_{l=1}^{d_1} \left\{ m_{0l} + \sum_{\alpha=2}^{d_2} m_{\alpha l}(X_\alpha) \right\} T_l$ and estimate all functions $\{m_{1l}(x_1)\}_{l=1}^{d_1}$ by linear regression of $Y_{,1}$ on $T_1, .., T_{d_1}$ with kernel weights computed from variable X_1. These would-be estimators do not suffer from the "curse of dimensionality" and are called "oracle smoothers". We propose to pre-estimate the functions $\{m_{\alpha l}(x_\alpha)\}_{l=1,\alpha=2}^{d_1,d_2}$ and constants $\{m_{0l}\}_{l=1}^{d_1}$ by linear spline and then use these estimates as substitutes to obtain an approximation $\hat{Y}_{,1}$ to the variable $Y_{,1}$, and construct "oracle" estimators based on $\hat{Y}_{,1}$. The theoretical contribution of this chapter is proving that the error caused by this "cheating" is negligible. Consequently, the SBK/SBLL estimators are uniformly (over the data range) equivalent to univariate kernel/local linear "oracle smoothers", automatically inheriting all their oracle efficiency properties. Our proof relies on the general principles of "reducing bias by undersmoothing" and "averaging out the variance", accomplished with the joint asymptotics of kernel and spline functions. Another innovation in this chapter is the $\sqrt{n}$-consistent oracle estimation of constants $\{m_{0l}\}_{l=1}^{d_1}$ under conditions no more than second order smoothness of $\{m_{\alpha l}(x_\alpha)\}_{l=1,\alpha=1}^{d_1,d_2}$. Xue & Yang (2006a) had provided $\sqrt{n}$-consistent estimation of constants $\{m_{0l}\}_{l=1}^{d_1}$ only under higher order smoothness Assumptions, while Xue and Yang (2006b) had failed to obtain $\sqrt{n}$-consistency for estimating $\{m_{0l}\}_{l=1}^{d_1}$.

This chapter is organized as follows. In Section 4.2 we discuss the assumptions of the model (1.3.1). In Section 4.3, we introduce the oracle smoothers and discuss its asymptotic properties. In Section 4.4 we introduce the SBK and SBLL estimators, their L_2 consistency and asymptotic normal distribution. The ideas behind our proofs of the main theoretical results are given by decomposing the estimator's "cheating" error into a bias and a variance part. In Section 4.5 we discuss the implementation of the estimators. In Section 4.6 we apply the methods to simulated and empirical examples. All technical proofs are given in the Appendix.

4.2 Assumptions

Let $\{(Y_i, \mathbf{X}_i, \mathbf{T}_i)\}_{i=1}^n$ be a sequence of strictly stationary observations, with identical distribution as $(Y, \mathbf{X}, \mathbf{T})$ in model (1.3.1). Denote the unknown conditional mean and variance functions as $m(\mathbf{X}, \mathbf{T}) = E(Y|\mathbf{X}, \mathbf{T}), \sigma^2(\mathbf{X}, \mathbf{T}) = \text{var}(Y|\mathbf{X}, \mathbf{T})$, then one has

$$Y_i = m(\mathbf{X}_i, \mathbf{T}_i) + \sigma(\mathbf{X}_i, \mathbf{T}_i)\varepsilon_i \tag{4.2.1}$$

for some conditional white noises $\{\varepsilon_i\}_{i=1}^n$ that satisfy $E(\varepsilon_i|\mathbf{X}_i, \mathbf{T}_i) = 0$, $E\left(\varepsilon_i^2|\mathbf{X}_i, \mathbf{T}_i\right) = 1$. The variables $(\mathbf{X}_i, \mathbf{T}_i)$ can consist of either exogenous variables or lagged values of Y_i. For the additive coefficient model, the regression function m takes the form in (1.3.1), and satisfies the identification conditions that

$$E\{m_{\alpha l}(X_\alpha)\} = 0, 1 \le l \le d_1, 1 \le \alpha \le d_2 \tag{4.2.2}$$

ensuring the unique additive representations of $m_l(\mathbf{x}) = m_{0l} + \sum_{\alpha=1}^{d_2} m_{\alpha l}(x_\alpha)$. As in most works on nonparametric smoothing, estimation of the functions $\{m_{\alpha l}(x_\alpha)\}_{l=1,\alpha=1}^{d_1,d_2}$ is conducted on compact sets. Without lose of generality, let the compact set be $\chi = [0,1]^{d_2}$.

Following Stone (1985), p. 693, the space of α-centered square integrable functions on $[0,1]$ is

$$\mathcal{H}_\alpha^0 = \left\{g : E\{g(X_\alpha)\} = 0, E\left\{g^2(X_\alpha)\right\} < +\infty\right\}, 1 \le \alpha \le d_2.$$

Next define the model space $\mathcal{M}$, a collection of functions on $\chi \times R^{d_1}$ as

$$\mathcal{M} = \left\{ g(\mathbf{x}, \mathbf{t}) = \sum_{l=1}^{d_1} g_l(\mathbf{x}) t_l; \quad g_l(\mathbf{x}) = g_{0l} + \sum_{\alpha=1}^{d_2} g_{\alpha l}(x_\alpha); g_{\alpha l} \in \mathcal{H}_\alpha^0 \right\},$$

in which $\{g_{0l}\}_{l=1}^{d_1}$ are finite constants. The constraints that $E\{g_{\alpha l}(X_\alpha)\} = 0$, $1 \le \alpha \le d_2$ ensure unique additive representation of m_l as expressed in (4.2.2), but are not necessary for the definition of space $\mathcal{M}$. In what follows, denote by E_n the empirical expectation, $E_n\varphi = \sum_{i=1}^n \varphi(\mathbf{X}_i, \mathbf{T}_i)/n$. We introduce two inner products on $\mathcal{M}$. For functions $g_1, g_2 \in \mathcal{M}$, the theoretical and empirical inner products are defined respectively as $\langle g_1, g_2 \rangle = E\{g_1(\mathbf{X}, \mathbf{T}) g_2(\mathbf{X}, \mathbf{T})\}$, $\langle g_1, g_2 \rangle_n = E_n\{g_1(\mathbf{X}, \mathbf{T}) g_2(\mathbf{X}, \mathbf{T})\}$. The corresponding induced norms are $\|g_1\|_2^2 = Eg_1^2(\mathbf{X}, \mathbf{T})$, $\|g_1\|_{2,n}^2 = E_n g_1^2(\mathbf{X}, \mathbf{T})$. The model space $\mathcal{M}$

is called *theoretically (empirically) identifiable*, if for any $g \in \mathcal{M}$, $\|g\|_2 = 0$ ($\|g\|_{2,n} = 0$) implies that $g = 0$ a.s.

In this chapter, for any compact interval $[a, b]$, we denote the space of p-th order smooth function as $C^{(p)}[a,b] = \left\{g|g^{(p)} \in C[a,b]\right\}$, and the class of Lipschitz continuous functions for constant $C > 0$ as $\mathrm{Lip}([a,b], C) = \{g|\,|g(x) - g(x')| \le C|x - x'|, \forall x, x' \in [a,b]\}$. We mean by "$\sim$" both sides having the same order as $n \to \infty$. We denote by $\mathbf{I}_{d_1 \times d_1}$ the $d_1 \times d_1$ identity matrix, and $\mathbf{0}_{d_1 \times d_1}$ the $d_1 \times d_1$ zero matrix. For any vector $\mathbf{x} = \left(x_1, x_2, \cdots, x_{d_2}\right)$, we denote the supremum and Euclidean norms as $|\mathbf{x}| = \max_{1 \le \alpha \le d_2} |x_\alpha|$ and $\|\mathbf{x}\| = \left(\sum_{\alpha=1}^{d_2} x_\alpha^2\right)^{1/2}$.

We need the following Assumptions on the data generating process.

(A1) *The tuning variable* $\mathbf{X} = (X_1, \ldots, X_{d_2})$ *has a continuous probability density function* $f(\mathbf{x})$ *that satisfies* $0 < c_f \le \min_{\mathbf{x} \in \chi} f(\mathbf{x}) \le \max_{\mathbf{x} \in \chi} f(\mathbf{x}) \le C_f < \infty$ *for some constants* c_f *and* C_f *and* $f(\mathbf{x}) = 0, x \notin \chi = [0,1]^{d_2}$.

(A2) *There exist constants* $0 < c_{\mathbf{Q}} \le C_{\mathbf{Q}} < +\infty$ *and* $0 < c_\delta \le C_\delta < +\infty$ *and some* $\delta > 1/2$, *such that* $c_{\mathbf{Q}} \mathbf{I}_{d_1 \times d_1} \le \mathbf{Q}(\mathbf{x}) = \{q(\mathbf{x})\}_{l,l'=1}^{d_1} = E\left(\mathbf{T}\mathbf{T}^T | \mathbf{X} = \mathbf{x}\right) \le C_{\mathbf{Q}} \mathbf{I}_{d_1 \times d_1}$ *and* $c_\delta \le E\left\{\left(T_l T_{l'}\right)^{2+\delta} |\mathbf{X} = \mathbf{x}\right\} \le C_\delta$ *for all* $\mathbf{x} \in \chi$ *and* $l, l' = 1, \ldots, d_1$.

(A3) *The vector process* $\{\varsigma_t\}_{t=-\infty}^{\infty} = \{(Y_t, \mathbf{X}_t, \mathbf{T}_t)\}_{t=-\infty}^{\infty}$ *is strictly stationary and geometrically strongly mixing, that is, its* α *-mixing coefficient* $\alpha(k) \le c\rho^k$, *for constants* $c > 0$, $0 < \rho < 1$, *where* $\alpha(k) = \sup_{A \in \sigma(\varsigma_t, t \le 0), B \in \sigma(\varsigma_t, t \ge k)} |P(A)P(B) - P(A \cap B)|$.

(A4) *The coefficient components,* $m_{\alpha l} \in C^1[0,1]$, $m'_{\alpha l} \in \mathrm{Lip}([0,1], C_\infty), \forall 1 \le \alpha \le d_2, 1 \le l \le d_1$ *with* $m_{1l} \in C^2[0,1], \forall 1 \le l \le d_1$.

(A5) *The conditional variance function* $\sigma^2(\mathbf{x}, \mathbf{t})$ *is measurable and bounded. The errors* $\{\varepsilon_i\}_{i=1}^n$ *satisfy* $E(\varepsilon_i|\mathcal{F}_i) = 0$, $E(\varepsilon_i^2|\mathcal{F}_i) = 1$, $E\left(|\varepsilon_i|^{2+\eta}|\mathcal{F}_i\right) \le C_\eta$ *for some* $\eta \in (1/2, 1]$ *and the sequence of* σ*-fields* $F_i = \sigma\left\{(\mathbf{X}_j, \mathbf{T}_j), j \le i; \varepsilon_j, j \le i-1\right\}$ *for* $i = 1, \ldots, n$.

(A6) *The marginal density* $f_1(x_1)$ *of* X_1 *and the conditional second moment matrix function* $\mathbf{Q}_1(x_1)$ *defined in (4.2.3) both have continuous derivatives on* $[0,1]$.

Assumptions (A1)-(A5) are common in the literature, see for instance, Huang & Yang (2004), Huang & Shen (2004) and especially Xue & Yang (2006b). Assumption (A6) is needed only for the asymptotic theory of oracle "kernel smoother", but not for the oracle "local linear smoother". Assumption (A2) implies also that for all $x_\alpha \in [0,1], 1 \leq \alpha \leq d_2$ and $l, l' = 1, \ldots, d_1$

$$\begin{aligned} c_{\mathbf{Q}}\mathbf{I}_{d_1 \times d_1} &\leq \mathbf{Q}_\alpha(x_\alpha) = \{q_\alpha(x_\alpha)\}_{l,l'=1}^{d_1} = E(\mathbf{T}\mathbf{T}^T | X_\alpha = x_\alpha) \leq C_{\mathbf{Q}}\mathbf{I}_{d_1 \times d_1} \quad (4.2.3) \\ c_\delta &\leq E\left\{\left(T_l T_{l'}\right)^{2+\delta} | X_\alpha = x_\alpha\right\} \leq C_\delta. \end{aligned}$$

Furthermore, Assumptions (A2) and (A5) imply that for some constant $C > 0$

$$\max_{1 \leq l \leq d_1} E|T_l|^{2+\eta} < C \max_{1 \leq l \leq d_1} E|T_l T_l|^{2+\delta} = C \max_{1 \leq l \leq d_1} E|T_l|^{4+2\delta} \leq CC_\delta < +\infty. \quad (4.2.4)$$

At one referee's request, we provide here insight into the relationship allowed between the vectors $\mathbf{T}$ and $\mathbf{X}$ under Assumption (A2). It is instructive to first understand what $\mathbf{T}$ and $\mathbf{X}$ can not be in the context of identifiability for functions $\{m_{\alpha l}(x_\alpha)\}_{l=1,\alpha=1}^{d_1,d_2}$. Suppose that the vector $\mathbf{X}$ is centered so that $E\mathbf{X} = \mathbf{0}$. Then model (1.3.1) is unidentifiable when $(T_1, T_2) = (X_1, X_2)$ since $-3X_2T_1 + 3X_1T_2 = 0$, $E(-3X_2) = E(3X_1) = 0$ and the function $m(\mathbf{x}, \mathbf{t})$ in (1.3.1) is expressed as

$$\begin{aligned} &\sum_{l=3}^{d_1}\left\{m_{0l} + \sum_{\alpha=1}^{d_2} m_{\alpha l}(x_\alpha)\right\} t_l + \left\{m_{01} + m_{21}(x_2) + \sum_{\alpha=1,\alpha\neq 2}^{d_2} m_{\alpha 1}(x_\alpha)\right\} t_1 \\ &+ \left\{m_{02} + m_{12}(x_1) + \sum_{\alpha=2}^{d_2} m_{\alpha 1}(x_\alpha)\right\} t_2 \end{aligned}$$

$$\begin{aligned} \equiv &\sum_{l=3}^{d_1}\left\{m_{0l} + \sum_{\alpha=1}^{d_2} m_{\alpha l}(x_\alpha)\right\} t_l + \left\{m_{01} + m_{21}(x_2) - 3x_2 + \sum_{\alpha=1,\alpha\neq 2}^{d_2} m_{\alpha 1}(x_\alpha)\right\} t_1 \\ &+ \left\{m_{02} + m_{12}(x_1) + 3x_1 + \sum_{\alpha=2}^{d_2} m_{\alpha 1}(x_\alpha)\right\} t_2, \end{aligned}$$

so one can use $m_{21}^*(x_2) = m_{21}(x_2) - 3x_2$ and $m_{12}^*(x_1) = m_{12}(x_1) + 3x_1$ to replace $m_{21}(x_2)$ and $m_{12}(x_1)$ without changing the data generating process (1.3.1). In other words, the functions $m_{21}(x_2)$ and $m_{12}(x_1)$ are unidentifiable. Xue and Yang (2006a), p.2523 gave a similar counterexample, and discussed why an unidentifiable model may perform better for prediction.

More generally, it is revealing to note that Assumption (A2) not only rules out the above anomaly, but it also does not allow the possibility that there exist two T_l's $(1 \leq l \leq d_1)$ almost surely equal to two Borel functions of $\mathbf{X}$. To see this, suppose that $(T_1, T_2) = \{\varphi_1(\mathbf{X}), \varphi_2(\mathbf{X})\}, a.s$ for some Borel functions φ_1 and φ_2. Assumption (A2) implies that

$$c_{\mathbf{Q}} \mathbf{I}_{2\times 2} \leq E\left\{ \left(\begin{array}{cc} T_1^2 & T_1 T_2 \\ T_1 T_2 & T_2^2 \end{array} \right) \middle| \mathbf{X} = \mathbf{x} \right\} \leq C_{\mathbf{Q}} \mathbf{I}_{2\times 2}, \forall \mathbf{x} \in \chi$$

leading to

$$c_{\mathbf{Q}} \mathbf{I}_{2\times 2} \leq \left(\begin{array}{cc} \varphi_1^2(\mathbf{x}) & \varphi_1(\mathbf{x}) \varphi_2(\mathbf{x}) \\ \varphi_1(\mathbf{x}) \varphi_2(\mathbf{x}) & \varphi_2^2(\mathbf{x}) \end{array} \right) \leq C_{\mathbf{Q}} \mathbf{I}_{2\times 2}, a.s., \forall \mathbf{x} \in \chi$$

which can not be true as for any $\mathbf{x} \in \chi$, the 2×2 matrix in the above is singular, thus can not be $\geq c_{\mathbf{Q}} \mathbf{I}_{2\times 2}$. That Assumption (A2) guarantees the identifiability of model (1.3.1) has been established in Lemma 1 of Xue and Yang (2006b). It is important to observe, however, that Assumption (A2) does not exclude the case of one $T_l, 1 \leq l \leq d_1$ almost surely equal to a Borel function of $\mathbf{X}$.

4.3 Oracle Smoothers

We now introduce what is known as the oracle smoother in Wang & Yang (2007) as a benchmark for evaluating the estimators. Denote for any vector $\mathbf{x} = \left(x_1, x_2, \cdots, x_{d_2}\right)$ the deleted vector $\mathbf{x}_{\text{-}1} = \left(x_2, \cdots, x_{d_2}\right)$ and for the random vector $\mathbf{X}_i = \left(X_{i1}, X_{i2}, \cdots, X_{id_2}\right)$ the deleted vector $\mathbf{X}_{i,\text{-}1} = \left(X_{i2}, \cdots, X_{id_2}\right)$, $1 \leq i \leq n$. For any $1 \leq l \leq d_1$, write $m_{\text{-}1,l}(\mathbf{x}_{\text{-}1}) = m_{0l} + \sum_{\alpha=2}^{d_2} m_{\alpha l}(x_\alpha)$. Denote the vector of pseudo-responses $\mathbf{Y}_1 = (Y_{1,1}, \cdots, Y_{n,1})^T$ in which

$$Y_{i,1} = Y_i - \sum_{l=1}^{d_1} \left\{ m_{0l} + m_{\text{-}1,l}(\mathbf{X}_{i,\text{-}1}) \right\} T_{il} = \sum_{l=1}^{d_1} m_{1l}(X_{i1}) T_{il} + \sigma(\mathbf{X}_i, \mathbf{T}_i) \varepsilon_i.$$

These would have been the "responses" had the unknown functions $\left\{ m_{\text{-}1,l}(\mathbf{x}_{\text{-}1}) \right\}_{1 \leq l \leq d_1}$ been given. In that case, one could "estimate" all the coefficient functions in x_1, the vector function $m_{1,\cdot}(x_1) = \left\{m_{11}(x_1), \cdots, m_{1d_1}(x_1)\right\}^T$ by solving a kernel weighted least squares problem

$$\tilde{m}_{\mathrm{K},1,\cdot}(x_1) = \left\{\tilde{m}_{\mathrm{K},11}(x_1), \cdots, \tilde{m}_{\mathrm{K},1d_1}(x_1)\right\}^T = \underset{\boldsymbol{\lambda}=(\lambda_{1l})_{1\le l\le d_1}}{\operatorname{argmin}} L(\boldsymbol{\lambda}, m_{\_1,\cdot}, x_1)$$

in which

$$L(\lambda, m_{\_1,\cdot}, x_1) = \sum_{i=1}^{n}\left(Y_{i,1} - \sum_{l=1}^{d_1}\lambda_l T_{il}\right)^2 K_h(X_{i1} - x_1).$$

Alternatively, one could rewrite the above kernel oracle smoother in matrix form

$$\tilde{m}_{\mathrm{K},1,\cdot}(x_1) = \left(\mathbf{C}_{\mathrm{K}}^T\mathbf{W}_1\mathbf{C}_{\mathrm{K}}\right)^{-1}\mathbf{C}_{\mathrm{K}}^T\mathbf{W}_1\mathbf{Y}_1 = \left(\frac{1}{n}\mathbf{C}_{\mathrm{K}}^T\mathbf{W}_1\mathbf{C}_{\mathrm{K}}\right)^{-1}\frac{1}{n}\mathbf{C}_{\mathrm{K}}^T\mathbf{W}_1\mathbf{Y}_1 \qquad (4.3.1)$$

in which

$$\mathbf{T}_i = \left(T_{i1}, \cdots, T_{id_1}\right)^T, \mathbf{C}_{\mathrm{K}} = \{\mathbf{T}_1, ..., \mathbf{T}_n\}^T,$$

$$\mathbf{W}_1 = \operatorname{diag}\{K_h(X_{11} - x_1), ..., K_h(X_{n1} - x_1)\},$$

$K_h(u) = K(u/h)/h$ for a kernel function K and bandwidth h that satisfy

(A7) *The function K is a symmetric probability density function supported on $[-1, 1]$, and $K \in \operatorname{Lip}([-1,1], C_K)$ for some $C_K > 0$, while the bandwidth $h = h_{1,n} > 0, h \sim n^{-1/5}$.*

Likewise, one can define the local linear oracle smoother of $m_{1,\cdot}(x_1)$ as

$$\tilde{m}_{\mathrm{LL},1,\cdot}(x_1) = \left(\mathbf{I}_{d_1\times d_1}, \mathbf{0}_{d_1\times d_1}\right)\left(\frac{1}{n}\mathbf{C}_{\mathrm{LL},1}^T\mathbf{W}_1\mathbf{C}_{\mathrm{LL},1}\right)^{-1}\frac{1}{n}\mathbf{C}_{\mathrm{LL},1}^T\mathbf{W}_1\mathbf{Y}_1, \qquad (4.3.2)$$

in which

$$\mathbf{C}_{\mathrm{LL},1} = \left\{\begin{array}{ccc} \mathbf{T}_1 & , \cdots , & \mathbf{T}_n \\ \mathbf{T}_1(X_{11} - x_1) & , \cdots , & \mathbf{T}_n(X_{n1} - x_1) \end{array}\right\}^T.$$

In this chapter denote $\mu_2(K) = \int u^2 K(u)\,du$, $\|K\|_2^2 = \int K(u)^2\,du$, $\mathbf{Q}_1(x_1)$ as in (4.2.3) and define the following bias and variance coefficients

$$b_{\mathrm{LL},l,l',1}(x_1) = \frac{1}{2}\mu_2(K)\, m''_{1l}(x_1) f_1(x_1) q_{ll',1}(x_1),$$

$$b_{\mathrm{K},l,l',1}(x_1) = \frac{1}{2}\mu_2(K)\left[2m'_{1l}(x_1)\frac{\partial}{\partial x_1}\left\{f_1(x_1) q_{ll',1}(x_1)\right\} + m''_{1l}(x_1) f_1(x_1) q_{ll',1}(x_1)\right],$$

$$\Sigma_1(x_1) = \|K\|_2^2 f_1(x_1) E\left\{\mathbf{T}\mathbf{T}^T\sigma^2(\mathbf{X},\mathbf{T})\,|X_1 = x_1\right\},$$

$$\left\{v_{l,l',1}(x_1)\right\}_{l,l'=1}^{d_1} = \mathbf{Q}_1(x_1)^{-1}\Sigma_1(x_1)\mathbf{Q}_1(x_1)^{-1}. \qquad (4.3.3)$$

THEOREM 4.3.1. *Under Assumptions (A1) to (A5) and (A7), for any* $x_1 \in [h, 1-h]$, *as* $n \to \infty$, *the oracle local linear smoother* $\tilde{m}_{\mathrm{LL},1,\cdot}(x_1)$ *given in (4.3.2) satisfies*

$$\sqrt{nh}\left[\tilde{m}_{\mathrm{LL},1,\cdot}(x_1) - m_{1,\cdot}(x_1) - \left\{\sum_{l=1}^{d_1} b_{\mathrm{LL},l,l',1}(x_1)\right\}_{l'=1}^{d_1} h^2\right] \to N\left(0, \left\{v_{l,l',1}(x_1)\right\}_{l,l'=1}^{d_1}\right).$$

With Assumption (A6) in addition, the oracle kernel smoother $\tilde{m}_{\mathrm{K},1,\cdot}(x_1)$ *in (4.3.1) satisfies*

$$\sqrt{nh}\left[\tilde{m}_{\mathrm{K},1,\cdot}(x_1) - m_{1,\cdot}(x_1) - \left\{\sum_{l=1}^{d_1} b_{\mathrm{K},l,l',1}(x_1)\right\}_{l'=1}^{d_1} h^2\right] \to N\left(0, \left\{v_{l,l',1}(x_1)\right\}_{l,l'=1}^{d_1}\right).$$

THEOREM 4.3.2. *Under Assumptions (A1) to (A5) and (A7), as* $n \to \infty$, *the oracle local linear smoother* $\tilde{m}_{\mathrm{LL},1,\cdot}(x_1)$ *given in (4.3.2) satisfies*

$$\sup_{x_1 \in [h,1-h]} \left|\tilde{m}_{\mathrm{LL},1,\cdot}(x_1) - m_{1,\cdot}(x_1)\right| = O_p\left(\log n / \sqrt{nh}\right).$$

With Assumption (A6) in addition, the oracle kernel smoother $\tilde{m}_{\mathrm{K},1,\cdot}(x_1)$ *in (4.3.1) satisfies*

$$\sup_{x_1 \in [h,1-h]} \left|\tilde{m}_{\mathrm{K},1,\cdot}(x_1) - m_{1,\cdot}(x_1)\right| = O_p\left(\log n / \sqrt{nh}\right).$$

Remark 1. *The above theorems hold for* $\tilde{m}_{\mathrm{LL},\alpha,\cdot}(x_\alpha)$ *and* $\tilde{m}_{\mathrm{K},\alpha,\cdot}(x_\alpha)$ *similarly constructed as* $\tilde{m}_{\mathrm{LL},1,\cdot}(x_1)$*and* $\tilde{m}_{\mathrm{K},1,\cdot}(x_1)$, *for any* $\alpha = 2, \ldots, d_2$, *i.e.,*

$$\begin{aligned}
\tilde{m}_{\mathrm{LL},\alpha,\cdot}(x_\alpha) &= \left(\mathbf{I}_{d_1 \times d_1}, \mathbf{0}_{d_1 \times d_1}\right)\left(\frac{1}{n}\mathbf{C}_{\mathrm{LL},\alpha}^T \mathbf{W}_\alpha \mathbf{C}_{\mathrm{LL},\alpha}\right)^{-1} \frac{1}{n}\mathbf{C}_{\mathrm{LL},\alpha}^T \mathbf{W}_\alpha \mathbf{Y}_\alpha, \\
\tilde{m}_{\mathrm{K},\alpha,\cdot}(x_\alpha) &= \left(\frac{1}{n}\mathbf{C}_{\mathrm{K}}^T \mathbf{W}_\alpha \mathbf{C}_{\mathrm{K}}\right)^{-1} \frac{1}{n}\mathbf{C}_{\mathrm{K}}^T \mathbf{W}_\alpha \mathbf{Y}_\alpha,
\end{aligned}$$

except that in Assumption (A4) one has to replace "$m_{1l} \in C^2[0,1], \forall 1 \le l \le d_1$*" with "*$m_{\alpha l} \in C^2[0,1], \forall 1 \le l \le d_1$*" and in Assumption (A6),* $f_1(x_1)$ *and* $Q_1(x_1)$ *have to be replaced with* $f_\alpha(x_\alpha)$ *and* $Q_\alpha(x_\alpha)$.

The same oracle idea applies to the constants as well. Define the would-be "estimators" of constants $(m_{0l})_{1 \le l \le d_1}^T$ as the following least squares solution

$$\tilde{m}_0 = (\tilde{m}_{0l})_{1 \le l \le d_1}^T = \arg\min \sum_{i=1}^{n} \left\{Y_{ic} - \sum_{l=1}^{d_1} m_{0l} T_{il}\right\}^2, \tag{4.3.4}$$

in which the oracle responses are

$$Y_{ic} = Y_i - \sum_{l=1}^{d_1}\sum_{\alpha=1}^{d_2} m_{\alpha l}(X_{i\alpha}) T_{il} = \sum_{l=1}^{d_1} m_{0l} T_{il} + \sigma(\mathbf{X}_i, \mathbf{T}_i)\,\varepsilon_i. \tag{4.3.5}$$

The following result provides optimal convergence rate of $\tilde{m}_0$ to m_0, which are needed for removing the effects of m_0 for estimating the functions $\{m_{1l}(x_1)\}_{l=1}^{d_1}$

PROPOSITION 4.3.1. *Under Assumptions (A1)-(A5) and (A8), as* $n \to \infty$, $\sup_{1\le l\le d_1} |\tilde{m}_{0l} - m_{0l}| = O_p\left(n^{-1/2}\right)$.

Although the oracle smoothers $\tilde{m}_{\mathrm{LL},\alpha,\cdot}(x_\alpha)$, $\tilde{m}_{\mathrm{K},\alpha,\cdot}(x_\alpha)$ possess the desirable theoretical properties in Theorems 4.3.1 and 4.3.2, they are not useful statistics as they are computed based on the knowledge of unavailable functions $\{m_{\alpha l}(x_\alpha)\}_{l=1,\alpha=2}^{d_1,d_2}$ and constants $\{m_{0l}\}_{l=1}^{d_1}$. They do, however, motivate the spline-backfitted estimators that we introduce in the next section.

4.4 Spline-backfitted Kernel Estimators

In this section we describe how the unknown functions $\{m_{\alpha l}(x_\alpha)\}_{l=1,\alpha=2}^{d_1,d_2}$ and constants $\{m_{0l}\}_{l=1}^{d_1}$ can be pre-estimated by linear spline and how the estimates are used to construct the "oracle estimators". To this end, we first introduce the space of linear splines. Let $0 = \xi_0 < \xi_1 < \cdots < \xi_N < \xi_{N+1} = 1$ denote a sequence of equally spaced points, called interior knots, on interval $[0, 1]$. Denote by $H = (N+1)^{-1}$ the width of each subinterval $[\xi_J, \xi_{J+1}]$, $0 \le J \le N$ and denote the degenerate knots $\xi_{-1} = 0, \xi_{N+2} = 1$. We assume that

(A8) *The number of interior knots* $N = N_n \sim n^{1/4}\log n$ *and hence* $H \sim n^{-1/4}(\log n)^{-1}$.

For $J = 0, \ldots, N+1$, define the linear B spline basis as

$$b_J(x) = (1 - |x - \xi_J|/H)_+ = \begin{cases} (N+1)x - J + 1 & , \quad \xi_{J-1} \le x \le \xi_J \\ J + 1 - (N+1)x & , \quad \xi_J \le x \le \xi_{J+1} \\ 0 & , \quad \text{otherwise} \end{cases},$$

the space of α-empirically centered linear spline functions on $[0,1]$ as

$$G_{n,\alpha}^{0}=\left\{g_{\alpha}:g_{\alpha}\left(x_{\alpha}\right)\equiv\sum_{J=0}^{N+1}\lambda_{J}b_{J}\left(x_{\alpha}\right),E_{n}\left\{g_{\alpha}\left(X_{\alpha}\right)\right\}=0\right\},1\leq\alpha\leq d_{2},$$

and the space of additive spline coefficient functions on $\chi\times R^{d_1}$ as

$$G_{n}^{0}=\left\{g\left(\mathbf{x},\mathbf{t}\right)=\sum_{l=1}^{d_1}g_{l}\left(\mathbf{x}\right)t_{l};\quad g_{l}\left(\mathbf{x}\right)=g_{0l}+\sum_{\alpha=1}^{d_2}g_{\alpha l}\left(x_{\alpha}\right);g_{0l}\in R,g_{\alpha l}\in G_{n,\alpha}^{0}\right\},$$

which is equipped with the empirical inner product $\langle\cdot,\cdot\rangle_{2,n}$.

The multivariate function $m\left(\mathbf{x},\mathbf{t}\right)$ is estimated by an additive spline coefficient function

$$\hat{m}\left(\mathbf{x},\mathbf{t}\right)=\sum_{l=1}^{d_1}\hat{m}_{l}\left(\mathbf{x}\right)t_{l}=\underset{g\in G_{n}^{0}}{\operatorname{argmin}}\sum_{i=1}^{n}\left\{Y_{i}-g\left(\mathbf{X}_{i},\mathbf{T}_{i}\right)\right\}^{2}. \qquad (4.4.1)$$

Since $\hat{m}\left(\mathbf{x},\mathbf{t}\right)\in G_{n}^{0}$, one can write $\hat{m}_{l}\left(\mathbf{x}\right)=\hat{m}_{0l}+\sum_{\alpha=1}^{d_2}\hat{m}_{\alpha l}\left(x_{\alpha}\right)$; for $\hat{m}_{0l}\in R$ and $\hat{m}_{\alpha l}\left(x_{\alpha}\right)\in G_{n,\alpha}^{0}$. Simple algebra shows that the following oracle estimators of the constants m_{0l} are exactly equal to $\hat{m}_{0l}$, in which the oracle pseudo-responses $\hat{Y}_{ic}=Y_{i}-\sum_{l=1}^{d_1}\sum_{\alpha=1}^{d_2}\hat{m}_{\alpha l}\left(X_{i\alpha}\right)T_{il}$ which mimick the oracle responses Y_{ic} in (4.3.5)

$$\hat{m}_{0}=\left(\hat{m}_{0l}\right)_{1\leq l\leq d_1}^{T}=\arg\min_{\left(\lambda_{01},\ldots,\lambda_{0d_1}\right)}\sum_{i=1}^{n}\left\{\hat{Y}_{ic}-\sum_{l=1}^{d_1}\lambda_{0l}T_{il}\right\}^{2}. \qquad (4.4.2)$$

Proposition 4.4.1. *Under Assumptions (A1) to (A5) and (A8), as* $n\to\infty$, $\sup_{1\leq l\leq d_1}\left|\hat{m}_{0l}-\tilde{m}_{0l}\right|=O_{p}\left(n^{-1/2}\right)$, *hence* $\sup_{1\leq l\leq d_1}\left|\hat{m}_{0l}-m_{0l}\right|=O_{p}\left(n^{-1/2}\right)$ *following Proposition 4.3.1.*

Define next the oracle pseudo-responses $\hat{Y}_{i1}=Y_{i}-\sum_{l=1}^{d_1}\left(\hat{m}_{0l}+\sum_{\alpha=2}^{d_2}\hat{m}_{\alpha l}\left(X_{i\alpha}\right)\right)T_{il}$ and $\hat{\mathbf{Y}}_{1}=\left(\hat{Y}_{11},\cdots,\hat{Y}_{n1}\right)^{T}$, with $\hat{m}_{0l}$ and $\hat{m}_{\alpha l}$ defined in (4.4.2) and (4.4.1) respectively. The spline-backfitted kernel (SBK) and spline-backfitted local linear (SBLL) estimators are

$$\hat{m}_{\mathrm{SBK},1,\cdot}\left(x_{1}\right)=\left(\mathbf{C}_{\mathrm{K}}^{T}\mathbf{W}_{1}\mathbf{C}_{\mathrm{K}}\right)^{-1}\mathbf{C}^{T}\mathbf{W}_{1}\hat{\mathbf{Y}}_{1}=\left(\frac{1}{n}\mathbf{C}_{\mathrm{K}}^{T}\mathbf{W}_{1}\mathbf{C}_{\mathrm{K}}\right)^{-1}\frac{1}{n}\mathbf{C}^{T}\mathbf{W}_{1}\hat{\mathbf{Y}}_{1}, \qquad (4.4.3)$$

$$\hat{m}_{\mathrm{SBLL},1,\cdot}\left(x_{1}\right)=\left(\mathbf{I}_{d_1\times d_1},\mathbf{0}_{d_1\times d_1}\right)\left(\frac{1}{n}\mathbf{C}_{\mathrm{LL},1}^{T}\mathbf{W}_{1}\mathbf{C}_{\mathrm{LL},1}\right)^{-1}\frac{1}{n}\mathbf{C}_{\mathrm{LL},1}^{T}\mathbf{W}_{1}\hat{\mathbf{Y}}_{1}. \qquad (4.4.4)$$

The following theorem states that the asymptotic uniform magnitude of difference between $\hat{m}_{\mathrm{SBK},1,\cdot}\left(x_{1}\right)$ and $\tilde{m}_{\mathrm{K},1,\cdot}\left(x_{1}\right)$ is of order $o_{p}\left(n^{-2/5}\right)$, which is dominated by the asymptotic size of $\tilde{m}_{K,1,\cdot}(x_1) - m_{1,\cdot}(x_1)$. As a result, $\hat{m}_{SBK,1,\cdot}(x_1)$ will have the same asymptotic distribution as $\tilde{m}_{K,1,\cdot}(x_1)$. The same is true for $\hat{m}_{SBLL,1,\cdot}(x_1)$ and $\tilde{m}_{LL,1,\cdot}(x_1)$.

THEOREM 4.4.1. *Under Assumptions (A1) to (A5), (A7) and (A8), as $n \to \infty$, the SBK estimator $\hat{m}_{SBK,1,\cdot}(x_1)$ in (4.4.3) and the SBLL estimator $\hat{m}_{SBLL,1,\cdot}(x_1)$ in (4.4.4) satisfy*

$$\sup_{x_1 \in [0,1]} \left| \hat{m}_{SBK,1,\cdot}(x_1) - \tilde{m}_{K,1,\cdot}(x_1) \right| + \sup_{x_1 \in [0,1]} \left| \hat{m}_{SBLL,1,\cdot}(x_1) - \tilde{m}_{LL,1,\cdot}(x_1) \right| = o_p\left(n^{-2/5}\right).$$

Theorem 4.4.1 follows from (4.4.13) and Propositions 4.4.1, 4.4.2 and 4.4.3, and remains true if the number of knots is of the more general form $N \sim n^{1/4}N'$ where $N' \to \infty, N'/n^r \to 0, \forall r > 0$ as $n \to \infty$. The following corollary provides the asymptotic distributions of $\hat{m}_{SBLL,1,\cdot}(x_1)$ and $\tilde{m}_{K,1,\cdot}(x_1)$. The proof of this corollary is straightforward from Theorems 4.3.1 and 4.4.1.

COROLLARY 4.4.1. *Under Assumptions (A1) to (A5), (A7) and (A8), for any $x_1 \in [h, 1-h]$, as $n \to \infty$, the SBLL estimator $\hat{m}_{SBLL,1,\cdot}(x_1)$ in (4.4.4) satisfies*

$$\sqrt{nh}\left[\hat{m}_{SBLL,1,\cdot}(x_1) - m_{1,\cdot}(x_1) - \left\{\sum_{l=1}^{d_1} b_{LL,l,l',1}(x_1)\right\}_{l'=1}^{d_1} h^2\right] \to N\left(0, \left\{v_{l,l',1}(x_1)\right\}_{l,l'=1}^{d_1}\right)$$

and with the additional Assumption (A6), the SBK estimator $\hat{m}_{SBK,1,\cdot}(x_1)$ in (4.4.3) satisfies

$$\sqrt{nh}\left[\tilde{m}_{K,1,\cdot}(x_1) - m_{1,\cdot}(x_1) - \left\{\sum_{l=1}^{d_1} b_{K,l,l',1}(x_1)\right\}_{l'=1}^{d_1} h^2\right] \to N\left(0, \left\{v_{l,l',1}(x_1)\right\}_{l,l'=1}^{d_1}\right)$$

where $b_{LL,l,l',1}(x_1)$, $b_{K,l,l',1}(x_1)$ and $v_{l,l',1}(x_1)$ are defined as (4.3.3).

Remark 2. *The above theorem and corollary hold for $\hat{m}_{SBK,\alpha,\cdot}(x_\alpha)$ and $\hat{m}_{SBLL,\alpha,\cdot}(x_\alpha)$ similarly constructed for any $\alpha = 2, ..., d$, i.e.,*

$$\hat{m}_{SBK,\alpha,\cdot}(x_\alpha) = \left(\frac{1}{n}\mathbf{C}_K^T \mathbf{W}_\alpha \mathbf{C}_K\right)^{-1} \frac{1}{n}\mathbf{C}_K^T \mathbf{W}_\alpha \hat{\mathbf{Y}}_\alpha, \tag{4.4.5}$$

where $\hat{Y}_{i\alpha} = Y_i - \sum_{l=1}^{d_1}\left\{\hat{m}_{0l} + \sum_{1 \le \alpha' \le d_2, \alpha' \ne \alpha} \hat{m}_{\alpha l}(X_{i\alpha})\right\}$.

4.4.1 Decomposition

In this section, we introduce the ideas of the proof of Theorem 4.4.1. Our main objective is to study the difference between the smoothed backfitted estimator $\hat{m}_{\mathrm{SBK},1l'}(x_1)$ and the smoothed "oracle" estimator $\tilde{m}_{\mathrm{K},1l'}(x_1)$. First, define the theoretical inner product of b_J and 1 with respect to the α-th marginal density $f_\alpha(x_\alpha)$ as $c_{J,\alpha} = \langle b_J(X_\alpha), 1\rangle = \int b_J(x_\alpha) f_\alpha(x_\alpha) dx_\alpha$ and define the centered B spline basis $b_{J,\alpha}(x_\alpha)$ and the standardized B spline basis $B_{J,\alpha}(x_\alpha)$ as

$$b_{J,\alpha}(x_\alpha) = b_J(x_\alpha) - \frac{c_{J,\alpha}}{c_{J-1,\alpha}} b_{J-1}(x_\alpha), B_{J,\alpha}(x_\alpha) = \frac{b_{J,\alpha}(x_\alpha)}{\|b_{J,\alpha}\|_2}, 1 \le J \le N+1, \quad (4.4.6)$$

so that $EB_{J,\alpha}(X_\alpha) \equiv 0$, $EB^2_{J,\alpha}(X_\alpha) \equiv 1$.

For any n-dimensional vector $\mathbf{\Gamma} = \{\Gamma_1, ..., \Gamma_n\}^T$, we define the additive spline coefficient function constructed from the projection of $\mathbf{\Gamma}$ on the inner product space $\left(G_n^0, \langle \cdot, \cdot \rangle_{2,n}\right)$ as $(\mathbf{P}_n\mathbf{\Gamma})(\mathbf{x}, \mathbf{t}) = \sum_{l=1}^{d_1} \left\{\hat{\gamma}_{0,l} + \sum_{\alpha=1}^{d_2} \sum_{J=1}^{N+1} \hat{\gamma}_{J,\alpha,l} B_{J,\alpha}(x_\alpha)\right\} t_l$, in which $\left\{\hat{\gamma}_{0,l}, \hat{\gamma}_{J,\alpha,l}\right\}^T_{1\le J\le N+1, 1\le\alpha\le d_2, 1\le l\le d_1}$ minimizes

$$\sum_{i=1}^n \left[\Gamma_i - \sum_{l=1}^{d_1} \left\{\gamma_{0,l} + \sum_{\alpha=1}^{d_2} \sum_{J=1}^{N+1} \gamma_{J,\alpha,l} B_{J,\alpha}(X_{i\alpha})\right\} T_{il}\right]^2, \quad (4.4.7)$$

so one can rewrite the linear spline estimator in (4.4.1) as $\hat{m}(\mathbf{x}, \mathbf{t}) = (\mathbf{P}_n\mathbf{Y})(\mathbf{x}, \mathbf{t})$, where we denote by $\mathbf{Y} = (Y_i)^T_{1\le i\le n}$ the response vector. The coefficients of the linear regressors $t_l, 1 \le l \le d_1$ are denoted as the multivariate additive spline functions

$$\left(\mathbf{P}_{n,l}\mathbf{\Gamma}\right)(\mathbf{x}) = \hat{\gamma}_{0,l} + \sum_{\alpha=1}^{d_2} \sum_{J=1}^{N+1} \hat{\gamma}_{J,\alpha,l} B_{J,\alpha}(x_\alpha), l = 1, ..., d_1.$$

Note that $\left(\mathbf{P}_{n,l}\mathbf{\Gamma}\right)(x_\alpha) = \hat{\gamma}_{0,l} + \sum_{\alpha=1}^{d_2} \left(\mathbf{P}^*_{n,\alpha,l}\mathbf{\Gamma}\right)(x_\alpha)$ where $\left(\mathbf{P}^*_{n,\alpha,l}\mathbf{\Gamma}\right)(x_\alpha) = \sum_{J=1}^{N+1} \hat{\gamma}_{J,\alpha,l} B_{J,\alpha}(x_\alpha)$, we define the empirically centered additive components $\left(\mathbf{P}_{n,\alpha,l}\mathbf{\Gamma}\right)(x_\alpha)$, $\alpha = 1, ..., d_2$

$$\left(\mathbf{P}_{n,\alpha,l}\mathbf{\Gamma}\right)(x_\alpha) = \left(\mathbf{P}^*_{n,\alpha,l}\mathbf{\Gamma}\right)(x_\alpha) - n^{-1} \sum_{i=1}^n \left(\mathbf{P}^*_{n,\alpha,l}\mathbf{\Gamma}\right)(X_{i\alpha}). \quad (4.4.8)$$

Using these notations, spline estimators of $m_l(\mathbf{x})$ and $m_{\alpha l}(x_\alpha)$ are $\hat{m}_l(\mathbf{x}) = \left(\mathbf{P}_{n,l}\mathbf{Y}\right)(\mathbf{x}), \hat{m}_{\alpha l}(x_\alpha) = \left(\mathbf{P}_{n,\alpha,l}\mathbf{Y}\right)(x_\alpha)$, while noiseless spline smoothers and variance spline components are

$$\tilde{m}_l(\mathbf{x}) = (\mathbf{P}_{n,l}\mathbf{m})(\mathbf{x}), \tilde{m}_{\alpha l}(x_\alpha) = (\mathbf{P}_{n,\alpha,l}\mathbf{m})(x_\alpha),$$

$$\tilde{\varepsilon}_l(\mathbf{x}) = (\mathbf{P}_{n,l}\mathbf{E})(\mathbf{x}), \tilde{\varepsilon}_{\alpha l}(x_\alpha) = (\mathbf{P}_{n,\alpha,l}\mathbf{E})(x_\alpha), \tag{4.4.9}$$

where $\mathbf{m} = \{m(\mathbf{X}_i, \mathbf{T}_i)\}_{1\le i\le n}^T$ is the true function vector and $\mathbf{E} = \{\sigma(\mathbf{X}_i, \mathbf{T}_i)\varepsilon_i\}_{1\le i\le n}^T$ the error vector. Due to the linearity of operators $\mathbf{P}_{n,\alpha}$ and $\mathbf{P}_{n,\alpha,l}$, $1 \le l \le d_1, 1 \le \alpha \le d_2$ and $\mathbf{Y} = \mathbf{m}+\mathbf{E}$ due to (4.2.1), one has the following crucial decomposition for proving Theorem 4.4.1,

$$\hat{m}_l(\mathbf{x}) = \tilde{m}_l(\mathbf{x}) + \tilde{\varepsilon}_l(\mathbf{x}),\ \hat{m}_{\alpha l}(x_\alpha) = \tilde{m}_{\alpha l}(x_\alpha) + \tilde{\varepsilon}_{\alpha l}(x_\alpha), 1 \le l \le d_1, 1 \le \alpha \le d_2. \tag{4.4.10}$$

We define additionally an auxiliary entity

$$\tilde{\varepsilon}^*_{\alpha l}(x_\alpha) = \left(\mathbf{P}^*_{n,\alpha,l}\mathbf{E}\right)(x_\alpha), 1 \le l \le d_1, 1 \le \alpha \le d_2. \tag{4.4.11}$$

Definition (4.4.8) implies that $\tilde{\varepsilon}_{\alpha l}(x_\alpha)$ is simply the empirical centering of $\tilde{\varepsilon}^*_{\alpha l}(x_\alpha)$, i.e.

$$\tilde{\varepsilon}_{\alpha l}(x_\alpha) \equiv \tilde{\varepsilon}^*_{\alpha l}(x_\alpha) - n^{-1}\sum_{i=1}^n \tilde{\varepsilon}^*_{\alpha l}(X_{i\alpha}). \tag{4.4.12}$$

According to (4.3.1) and (4.4.3),

$$\hat{m}_{\mathrm{SBK},1,\cdot}(x_1) - \tilde{m}_{\mathrm{K},1,\cdot}(x_1) = \left(\frac{1}{n}\mathbf{C}_{\mathrm{K}}^T\mathbf{W}\mathbf{C}_{\mathrm{K}}\right)^{-1}\frac{1}{n}\mathbf{C}_{\mathrm{K}}^T\mathbf{W}_1\left(\hat{\mathbf{Y}}_1 - \mathbf{Y}_1\right),$$

$$\begin{aligned}\hat{\mathbf{Y}}_1 - \mathbf{Y}_1 &= \left(\hat{Y}_{1,1}, \cdots, \hat{Y}_{n,1}\right)^T - (Y_{1,1}, \cdots, Y_{n,1})^T \\ &= \left[\sum_{l=1}^{d_1}\{m_{0l} - \hat{m}_{0l} + m_{\text{-}1,l}(\mathbf{X}_{i,\text{-}1}) - \hat{m}_{\text{-}1,l}(\mathbf{X}_{i,\text{-}1})\}T_{il}\right]_{1\le i\le n} \\ &= \mathbf{C}_{\mathrm{K}}(m_{0l} - \hat{m}_{0l})_{1\le l\le d_1} + \left[\sum_{l=1}^{d_1}\{m_{\text{-}1,l}(\mathbf{X}_{i,\text{-}1}) - \hat{m}_{\text{-}1,l}(\mathbf{X}_{i,\text{-}1})\}T_{il}\right]_{1\le i\le n}\end{aligned}$$

where making use of the definition of $\hat{m}_{0l}$ and the signal noise decomposition (4.4.10), the difference $\tilde{m}_{\mathrm{K},1,\cdot}(x_1) - \hat{m}_{\mathrm{SBK},1,\cdot}(x_1) - \hat{m}_{0,\cdot} + m_{0,\cdot}$ can be treated as the sum of two terms

$$\left(\frac{1}{n}\mathbf{C}_{\mathrm{K}}^T\mathbf{W}_1\mathbf{C}_{\mathrm{K}}\right)^{-1}\frac{1}{n}\mathbf{C}_{\mathrm{K}}^T\mathbf{W}\left[\sum_{l=1}^{d_1}\{m_{\text{-}1,l}(\mathbf{X}_{i,\text{-}1}) - \hat{m}_{\text{-}1,l}(\mathbf{X}_{i,\text{-}1})\}T_{il}\right]_{i=1}^n$$

$$= \left(\frac{1}{n}\mathbf{C}_{\mathrm{K}}^{T}\mathbf{W}_1\mathbf{C}_{\mathrm{K}}\right)^{-1}\left\{\Psi_b(x_1)+\Psi_v(x_1)\right\}_{l'=1}^{d_1} \tag{4.4.13}$$

where

$$\Psi_b(x_1) = \frac{1}{n}\mathbf{C}_{\mathrm{K}}^{T}\mathbf{W}_1\left[\sum_{l=1}^{d_1}\left\{m_{\text{-}1,l}\left(\mathbf{X}_{i,\text{-}1}\right)-\tilde{m}_{\text{-}1,l}\left(\mathbf{X}_{i,\text{-}1}\right)\right\}T_{il}\right]_{i=1}^{n} = \left\{\Psi_{b,l'}(x_1)\right\}_{l'=1}^{d_1}, \tag{4.4.14}$$

$$\Psi_v(x_1) = \frac{1}{n}\mathbf{C}_{\mathrm{K}}^{T}\mathbf{W}_1\left[\sum_{l=1}^{d_1}\tilde{\varepsilon}_{\text{-}1,l}\left(\mathbf{X}_{i,\text{-}1}\right)T_{il}\right]_{i=1}^{n} = \left\{\Psi_{v,l'}(x_1)\right\}_{l'=1}^{d_1}, \tilde{\varepsilon}_{\text{-}1,l}\left(\mathbf{X}_{i,\text{-}1}\right) = \sum_{\alpha=2}^{d_2}\tilde{\varepsilon}_{\alpha l}\left(X_{i\alpha}\right) \tag{4.4.15}$$

and

$$\begin{aligned}
\Psi_{b,l'}(x_1) &= \frac{1}{n}\sum_{i=1}^{n}K_h\left(X_{i1}-x_1\right)T_{il'}\sum_{l=1}^{d_1}\left\{m_{\text{-}1,l}\left(\mathbf{X}_{i,\text{-}1}\right)-\tilde{m}_{\text{-}1,l}\left(\mathbf{X}_{i,\text{-}1}\right)\right\}T_{il} \\
\Psi_{v,l'}(x_1) &= \frac{1}{n}\sum_{i=1}^{n}K_h\left(X_{i1}-x_1\right)T_{il'}\sum_{l=1}^{d_1}\tilde{\varepsilon}_{\text{-}1,l}\left(\mathbf{X}_{i,\text{-}1}\right)T_{il}.
\end{aligned}$$

The term $\Psi_b(x_1)$ is induced by the bias term $\hat{m}_{\text{-}1,l}\left(\mathbf{X}_{i,\text{-}1}\right)-m_{\text{-}1,l}\left(\mathbf{X}_{i,\text{-}1}\right)$, while $\Psi_v(x_1)$ relates to the noise terms $\tilde{\varepsilon}_{\text{-}1,l}\left(\mathbf{X}_{i,\text{-}1}\right)$. Both of these have order $o_p(n^{-2/5})$ by Propositions 4.4.2 and 4.4.3 below.

PROPOSITION 4.4.2. *Under Assumptions (A1)-(A4), (A7) and (A8), as* $n \to \infty$,

$$\sup_{1\le l'\le d_1}\sup_{x_1\in[0,1]}\left|\Psi_{b,l'}(x_1)\right| = O_p\left(n^{-1/2}+H^2\right) = o_p\left(n^{-2/5}\right).$$

PROPOSITION 4.4.3. *Under Assumptions (A1) to (A5), (A7) to (A8), as* $n \to \infty$, $\sup_{1\le l'\le d_1}\sup_{x_1\in[0,1]}\left|\Psi_{v,l'}(x_1)\right| = O_p\left(N(\log n)^2/n+H^2\right) = o_p\left(n^{-2/5}\right)$.

According to (4.4.12) and (4.4.15), we can write $\Psi_{v,l'}(x_1) = \Psi_{v,l'}^{(2)}(x_1)-\Psi_{v,l'}^{(1)}(x_1)$, where

$$\begin{aligned}
\Psi_{v,l'}^{(1)}(x_1) &= n^{-1}\sum_{i=1}^{n}\sum_{l=1}^{d_1}K_h\left(X_{i1}-x_1\right)T_{il}T_{il'}\cdot n^{-1}\sum_{i=1}^{n}\tilde{\varepsilon}_{\text{-}1,l}^{*}\left(\mathbf{X}_{i,\text{-}1}\right), && (4.4.16)\\
\Psi_{v,l'}^{(2)}(x_1) &= n^{-1}\sum_{i=1}^{n}\sum_{l=1}^{d_1}K_h\left(X_{i1}-x_1\right)T_{il}T_{il'}\tilde{\varepsilon}_{\text{-}1,l}^{*}\left(\mathbf{X}_{i,\text{-}1}\right), && (4.4.17)
\end{aligned}$$

in which $\tilde{\varepsilon}^*_{-1,l}(\mathbf{X}_{i,-1}) = \sum_{\alpha=2}^{d_2} \tilde{\varepsilon}^*_{\alpha l}(X_{i\alpha})$ and $\tilde{\varepsilon}^*_{\alpha l}(X_{i\alpha})$ is given in (4.4.11). If further one denotes

$$\omega_{J,\alpha,l,l'}(\mathbf{X}_i, x_1) = T_{il}T_{il'}K_h(X_{i1} - x_1) B_{J,\alpha}(X_{i\alpha}), \; \mu_{\omega_{J,\alpha,l,l'}}(x_1) = E\omega_{J,\alpha,l,l'}(\mathbf{X}, x_1) \tag{4.4.18}$$

then by (4.4.17), (4.7.9) and (4.4.11), $\Psi^{(2)}_{v,l'}(x_1)$ can be rewritten as

$$\Psi^{(2)}_{v,l'}(x_1) = n^{-1} \sum_{i=1}^{n} \sum_{l=1}^{d_1} \sum_{J=1}^{N+1} \sum_{\alpha=2}^{d_2} \tilde{a}_{J,\alpha,l} \omega_{J,\alpha,l,l'}(\mathbf{X}_i, x_1). \tag{4.4.19}$$

Lemma 4.4.1. *Under Assumptions (A1) to (A5), (A7) to (A8), as $n \to \infty$, $\Psi^{(1)}_{v,l'}(x_1)$ defined in (4.4.16) satisfies* $\sup_{1 \le l' \le d_1} \sup_{x_1 \in [0,1]} \left|\Psi^{(1)}_{v,l'}(x_1)\right| = O_p\left(N(\log n)^2/n\right)$.

Lemma 4.4.2. *Under Assumptions (A1) to (A5), (A7) to (A8), , as $n \to \infty$, $\Psi^{(2)}_{v}(x_1)$ defined in (4.4.17) satisfies* $\sup_{1 \le l' \le d_1} \sup_{x_1 \in [0,1]} \left|\Psi^{(2)}_{v,l'}(x_1)\right| = O_p(H^2)$.

Proof of Proposition 4.4.2 is given in the Appendix, while Proposition 4.4.3 follows from Lemmas 4.4.1 and 4.4.2. Lemma 4.4.2 follows from Lemmas 4.7.13 and 4.7.14, both proved in the Appendix, while the proof of Lemma 4.4.1 is given in the Appendix. Similar result can be proved for $\hat{m}_{\text{SBLL},1l'}(x_1)$ by extending $\Psi_{b,l'}(x_1)$ and $\Psi_{v,l'}(x_1)$ to terms such as

$$\frac{1}{n}\sum_{i=1}^{n} K_h(X_{i1} - x_1)\left(\frac{X_{i1} - x_1}{h}\right) T_{il'} \sum_{l=1}^{d_1} \left\{m_{-1,l}(\mathbf{X}_{i,-1}) - \tilde{m}_{-1,l}(\mathbf{X}_{i,-1})\right\} T_{il},$$

$$\frac{1}{n}\sum_{i=1}^{n} K_h(X_{i1} - x_1)\left(\frac{X_{i1} - x_1}{h}\right) T_{il'} \sum_{l=1}^{d_1} \tilde{\varepsilon}_{-1,l}(\mathbf{X}_{i,-1}) T_{il},$$

which do not add much difficulty as $\left|\frac{X_{i1}-x_1}{h}\right| \le 1$ whenever $K_h(X_{i1} - x_1) \ne 0$.

4.5 Implementation

We implement our procedures with the following rule-of-thumb number of interior knots

$$N = N_n = \min\left(\left[n^{1/4}\log n\right] + 1, [n/4d_1d_2 - 1/d_2] - 1\right)$$

which satisfies Assumption (A8), i.e.$N = N_n \sim n^{1/4}\log n$, and ensures that the number of parameters in the linear least squares problem (4.4.7) is no more than $n/4$, i.e., $d_1\{1 + d_2(N_n + 1)\} \le n/4$.

By Corollary 4.4.1, the asymptotic distributions of the estimators $\hat{m}_{\mathrm{SBLL},\alpha,\cdot}(x_\alpha)$ depend not only on the functions $b_{\mathrm{LL},l,l',\alpha}(x_\alpha)$ and $v_{l'l',\alpha}(x_\alpha)$ but also crucially on the choice of bandwidths h_α. So we define the optimal bandwidth of h_α, denoted by $h_{\alpha,\mathrm{opt}}$, as the minimizer of the total asymptotic mean integrated squared errors (AMISE) of $\{\hat{m}_{\alpha l}(x_a), l = 1, \ldots, d_1\}$, which is defined as

$$\mathrm{AMISE}\{\hat{m}_{\alpha,\cdot}\} = \int \sum_{l'=1}^{d_1} \left[\left\{ \sum_{l=1}^{d_1} b_{\mathrm{LL},l,l',\alpha}(x_\alpha) h_\alpha^2 \right\}^2 + v_{l'l',\alpha}(x_\alpha) / (nh_\alpha) \right] f_\alpha(x_\alpha) dx_\alpha.$$

By letting $d\,\mathrm{AMISE}\{\hat{m}_{\alpha,\cdot}\} / dh_\alpha = 0$, one gets the optimal bandwidth $h_{\alpha,\mathrm{opt}}$ as

$$h_{\alpha,\mathrm{opt}} = \left\{ \frac{n^{-1} \int \sum_{l'=1}^{d_1} v_{l'l',\alpha}(x_\alpha) f_\alpha(x_\alpha) dx_\alpha}{4 \int \sum_{l'=1}^{d_1} \left\{ \sum_{l=1}^{d_1} b_{\mathrm{LL},l,l',\alpha}(x_\alpha) \right\}^2 f_\alpha(x_\alpha) dx_\alpha} \right\}^{1/5},$$

where $4 \int \sum_{l'=1}^{d_1} \left\{ \sum_{l=1}^{d_1} b_{\mathrm{LL},l,l',\alpha}(x_\alpha) \right\}^2 f_\alpha(x_\alpha) dx_\alpha$ is approximated by

$$n^{-1} \sum_{i=1}^{n} \mu_2^2(K) \sum_{l'=1}^{d_1} \left[\sum_{l=1}^{d_1} m''_{\alpha l}(X_{i\alpha}) f_\alpha(X_{i\alpha}) q_{ll',\alpha}(X_{i\alpha}) \right]^2.$$

To implement this, we propose the following simple estimation methods for terms $m''_{\alpha l}(x_1)$, $q_{ll',\alpha}(x_\alpha)$, $v_{l'l',\alpha}(x_\alpha)$ and $f_\alpha(x_\alpha)$. The resulting bandwidth is denoted as $\hat{h}_{1,\mathrm{opt}}$.

- The derivative function $m''_{\alpha l}(X_{i\alpha})$ is estimated as $\sum_{k=2}^{3} k(k-1) \hat{a}_{\alpha,l,k} X_{i\alpha}^{k-2} + 6 \sum_{k=4}^{N+3} \hat{a}_{\alpha,l,k} (X_{i1} - t_{\alpha,k-3})$ where $\{\hat{a}_{\alpha,l,k}\}_{k=0}^{N+3}$ minimize the following least squares

$$\sum_{i=1}^{n} \left[Y_i - \sum_{l=1}^{d_1} \sum_{\alpha=1}^{d_2} \left\{ \sum_{k=0}^{3} a_{\alpha,l,k} X_{i\alpha}^k + \sum_{k=4}^{N+3} a_{\alpha,l,k} (X_{i\alpha} - t_{\alpha,k-3})^3 \right\} T_{il} \right]^2$$

  where $\min_i X_{il} = t_0 < \cdots < t_{N+1} = \max_i X_{il}$.

- $q_{ll',\alpha}(x_\alpha)$ is estimated as $\sum_{k=0}^{3} \hat{a}_{\alpha,l,k} x_\alpha^k + \sum_{k=4}^{N+3} \hat{a}_{\alpha,l,k} (x_\alpha - t_{\alpha,k-3})^3$ by minimizing

$$\sum_{i=1}^{n} \left\{ T_{il} T_{il'} - \left\{ \sum_{k=0}^{3} a_{\alpha,l,k} X_{i\alpha}^k + \sum_{k=4}^{N+3} a_{\alpha,l,k} (X_{i\alpha} - t_{\alpha,k-3})^3 \right\} \right\}^2.$$

  $E\left\{\mathbf{T}\mathbf{T}^T \sigma^2(\mathbf{X}, \mathbf{T}) | X_\alpha = x_\alpha\right\}$ is estimated as $\sum_{k=0}^{3} \hat{a}_{\alpha,l,k}^k x_\alpha + \sum_{k=4}^{N+3} \hat{a}_{\alpha,l,k} (x_\alpha - t_{\alpha,k-3})^3$ by minimizing

$$\sum_{i=1}^{n} \left[T_{il} T_{il'} \{Y_i - \hat{m}(\mathbf{X}_i, \mathbf{T}_i)\}^2 - \left\{ \sum_{k=0}^{3} a_{\alpha,l,k} X_\alpha^k + \sum_{k=4}^{N+3} a_{\alpha,l,k} (X_\alpha - t_{k-3})^3 \right\} \right]^2.$$

- Density function $f_\alpha(x_\alpha)$ is estimated by $\frac{1}{n}\sum_{i=1}^{n} K_{h_\alpha}(X_{i\alpha} - x_\alpha)$ and $f'_\alpha(x_\alpha)$ by $-\left(nh_\alpha^2\right)^{-1}\sum_{i=1}^{n} K'\left(\frac{X_{i\alpha}-x_\alpha}{h_\alpha}\right)$ with the rule-of-the-thumb bandwidth h_α.

4.6 Examples

4.6.1 Simulated example

The data are generated from the model

$$Y = \{m_{01} + m_{11}(X_1) + m_{21}(X_2)\} T_1 + \{m_{02} + m_{12}(X_1) + m_{22}(X_2)\} T_2 + \varepsilon,$$

with $m_{01} = 2$, $m_{02} = 1$ and $m_{11}(x) = \sin(4x-2) + 2\exp\left\{-2(x-0.5)^2\right\} - 1/\sqrt{2\pi}$, $m_{21}(x) = x$, $m_{12}(x) = \sin(x)$, and $m_{22}(x) = 0$. The vector $\mathbf{X} = (X_1, X_2)^T$ is uniformly distributed on $[-\pi,\pi]^2$ while $\mathbf{T} = (T_1, T_2)^T$ has distribution conditional on $\mathbf{X}$ as bivariate normal with mean $(0,0)^T$ and covariance matrix $\operatorname{diag}\left(X_1^2/\pi^2 + 1, X_2^2/\pi^2 + 1\right)$. The error ε is standard normal independent of $(\mathbf{X}, \mathbf{T})$. The functions are estimated by SBLL method. For $\alpha = 1, 2$, let $x^i_{\alpha,\min}$, $x^i_{\alpha,\max}$ denote the smallest and largest observations of the variable x_α in the i -th replication. The functions $\{m_{\alpha l}\}_{\alpha=1,l=1}^{2,2}$ are estimated on a grid of equally-spaced points $x_{\alpha,r}, r = 1, ..., n_{\alpha,\text{grid}}$ with $x_{\alpha,1} = -0.90\pi, x_{\alpha,n_{\alpha,\text{grid}}} = 0.90\pi, n_{\alpha,\text{grid}} = 51$, $\alpha = 1, 2$.

Denoting the estimator of $m_{\alpha l}$ in the k-th replication as $\hat{m}_{\text{SBLL},\alpha l,k}$ and $\{x_{\alpha,r}\}_{r=1}^{n_{\alpha,\text{grid}}}$ the grid points where the functions are evaluated, we define the (averaged) integrated squared error (ISE and AISE) as

$$\begin{aligned} \text{ISE}(\hat{m}_{\text{SBLL},\alpha l,k}) &= \frac{1}{n_{\alpha,\text{grid}}}\sum_{r=1}^{n_{\text{grid}}}\left\{\hat{m}_{\text{SBLL},\alpha l,k}(x_{\alpha,r}) - m_{\alpha l}(x_{\alpha,r})\right\}^2, \\ \text{AISE}(\hat{m}_{\text{SBLL},\alpha l}) &= \frac{1}{100}\sum_{k=1}^{100}\text{ISE}(\hat{m}_{\text{SBLL},\alpha l,k}). \end{aligned}$$

Table 8 reports the means and standard errors (in the parentheses) of $\{\hat{m}_{0l}\}_{l=1,2}$ and the AISEs of $\{\hat{m}_{\alpha l}\}_{\alpha=1,l=1}^{2,2}$ for all the two fits. Both fits are generally comparable, with the SBLL fit better than the spline fit ($p = 1$). The standard errors of the constant estimators and the AISEs of the function estimators decrease as samples size increases, confirming Corollary 4.4.1. Figure 11 gives the plot of one SBLL fit for sample size $n = 500$.

4.6.2 Real data example

In this section we illustrate how the additive coefficient model is used to extend the Cobb-Douglas model for annual US GDP growth. Denoted by Q_t the US GDP at year t, K_t the US capital at year t, L_t the US labor at year t, X_t the growth rate of ratio of R&D expenditure to GDP at year t, all data have been downloaded from the Bureau of Economic Analysis (BEA) website for years, $t = 1959, ..., 2002$ ($n = 44$). The standard Cobb-Douglas production function (Cobb & Douglas, 1928) is $Q_t = A_t K_t^{\beta_1} L_t^{1-\beta_1}$ where A_t is the Total Factor Productivity (TFP) of year t, β_1 is a parameter determined by technology. Define the following stationary time series variables

$$Y_t = \log Q_t - \log Q_{t-1}, T_{1t} = \log K_t - \log K_{t-1}, T_{2t} = \log L_t - \log L_{t-1},$$

then the Cobb-Douglas equation implies the following simple regression model

$$Y_t = (\log A_t - \log A_{t-1}) + \beta_1 T_{1t} + (1 - \beta_1) T_{2t}.$$

According to Solow (1957), the total factor productivity A_t has an almost constant rate of change, thus one might replace $\log A_t - \log A_{t-1}$ with an unknown constant and arrive at the following model

$$Y_t - T_{2t} = \beta_0 + \beta_1 (T_{1t} - T_{2t}). \tag{4.6.1}$$

The constant change rate of A_t in the period 1909-49 was mainly due to the relative low impact of technology in that era.

As technology growth is one of the biggest sub-sections of TFP, it is reasonable to examine the dependence of both β_0 and β_1 on technology rather than treating them as fixed constants. We use exogenous variables X_t (Growth rate of ratio of R&D expenditure to GDP at year t) to represent technology level and model $Y_t - T_{2t} = m_1(\mathbf{X}_t) + m_2(\mathbf{X}_t)(T_{1t} - T_{2t})$ where $m_l(\mathbf{X}_t) = m_{0l} + \sum_{\alpha=1}^{d_2} m_{\alpha l}(X_{t-\alpha+1})$, $l = 1, 2$, $\mathbf{X}_t = \left(X_t, ..., X_{t-d_2+1}\right)$. Using the BIC of Xue & Yang (2006b) for additive coefficient model with $d_2 = 5$, the following reduced model is considered optimal

$$Y_t - T_{2t} = c_1 + m_{41}(X_{t-3}) + \{c_2 + m_{52}(X_{t-4})\}(T_{1t} - T_{2t}). \tag{4.6.2}$$

The rolling forecast errors of GDP by SBLL fitting of model (4.6.2) and linear fitting of (4.6.1) are show in Figure 9. The averaged squared prediction error (ASPE)

$$\frac{1}{9}\sum_{t=1994}^{2002}\left[Y_t - T_{2t} - \hat{c}_1 - \hat{m}_{\mathrm{SBLL},41}(X_{t-3}) - \left\{\hat{c}_2 + \hat{m}_{\mathrm{SBLL},52}(X_{t-4})\right\}(T_{1t} - T_{2t})\right]^2,$$

for model (4.6.2) is 0.001818, which is about 60% of the corresponding ASPE (0.003097) for model (4.6.1). The in sample averaged squared estimation error (ASE) for model (4.6.2) is 5.2399×10^{-5}, which is about 68% of the in sample ASE (7.6959×10^{-5}) for model (4.6.1).

In model (4.6.2), $\hat{c}_1 + \hat{m}_{\mathrm{SBLL},41}(X_{t-3})$ estimates the TFP growth rate, which is shown as a function of X_{t-3} in Figure 10. It is obvious that the effect of X_{t-3} is positive when $X_{t-3} \leq 0.02$, but negative when $X_{t-3} > 0.02$. On average, the higher R&D investment spending causes faster GDP growing. However, overspending on R&D often leads to high losses (Culpepper, 2004 and Tokic, 2003).

We have also computed the average contribution of R&D to GDP growth for 1964-2001, which is about 40%. The GDP and estimated TFP growth rates is shown in Figure 12, it is obvious that TFP growth is highly correlated to the GDP growth. For more details, see Arnold (2005).

4.7 Appendix

4.7.1 Preliminaries

In the proofs that follow, we use U and u to denote sequences of random variables that are uniformly O and o of certain order.

LEMMA 4.7.1. *(Xue and Yang, 2006b, Lemma A.2, Lemma A.5) There exists a constant* $c_0 > 0$ *such that for any sets of coefficients* $\left\{a_{0l}, a_{J,\alpha,l}, 1 \leq J \leq N+1, 1 \leq l \leq d_1, 1 \leq \alpha \leq d_2\right\}$,

$$\left\|\sum_{l=1}^{d_1}\left(a_{0l} + \sum_{\alpha=1}^{d_2}\sum_{J=1}^{N+1} a_{J,\alpha,l}B_{J,\alpha}\right)t_l\right\|_2^2 \geq c_0\sum_{l=1}^{d_1}\left(a_{0l}^2 + \sum_{\alpha=1}^{d_2}\sum_{J=1}^{N+1} a_{J,\alpha,l}^2\right)$$

and that as $n \to \infty$, with probability approaching 1,

$$\left\| \sum_{l=1}^{d_1} \left(a_{0l} + \sum_{\alpha=1}^{d_2} \sum_{J=1}^{N+1} a_{J,\alpha,l} B_{J,\alpha} \right) t_l \right\|_{2,n}^2 \geq c_0 \sum_{l=1}^{d_1} \left(a_{0l}^2 + \sum_{\alpha=1}^{d_2} \sum_{J=1}^{N+1} a_{J,\alpha,l}^2 \right).$$

LEMMA 4.7.2. *Under Assumptions (A1) and (A8), one has: (i) there exist constants $c_f, C_f, c_0(f)$ and $C_0(f)$ depending on the marginal densities $f_\alpha(x_\alpha)$, $1 \leq \alpha \leq d_2$, such that $c_f H \leq c_{J,\alpha} \leq C_f H$ and $c_0(f) H \leq \|b_{J,\alpha}\|_2^2 \leq C_0(f) H$. (ii) uniformly for $J, J' = 1, ..., N+1$*

$$E\left\{B_{J,\alpha}(X_{i\alpha}) B_{J',\alpha}(X_{i\alpha})\right\} \sim \begin{cases} 1 & J' = J \\ -1/3 & |J' - J| = 1 \\ 1/6 & |J' - J| = 2 \end{cases}$$

$$E\left|B_{J,\alpha}(X_{i\alpha}) B_{J',\alpha}(X_{i\alpha})\right|^k \sim \begin{cases} H^{1-k} & |J' - J| \leq 2 \\ 0 & |J' - J| > 2 \end{cases}, k \geq 1.$$

LEMMA 4.7.3. *Under Assumption (A2), for $\mathbf{V_T}$ defined in (4.7.15) and $\mathbf{S_T} = \mathbf{V_T}^{-1}$*

$$c_\mathbf{Q} c_\mathbf{V} I_{d_1\{d_2(N+1)+1\}} \leq \mathbf{V_T} \leq C_\mathbf{Q} C_\mathbf{V} I_{d_1\{d_2(N+1)+1\}},$$

$$c_\mathbf{Q} c_\mathbf{S} I_{d_1\{d_2(N+1)+1\}} \leq \mathbf{S_T} \leq C_\mathbf{Q} C_\mathbf{S} I_{d_1\{d_2(N+1)+1\}}.$$

Proof. By definition, $\mathbf{V_T} = E\left[E\left(\mathbf{T}\mathbf{T}^T \middle| \mathbf{X}\right) \otimes \left\{\mathbf{B}(\mathbf{X}) \mathbf{B}(\mathbf{X})^T\right\}\right]$. According to Assumption (A2) and Theorem 20, p. 192 of Zhang (1999),

$$\mathbf{V_T} \leq C_\mathbf{Q} I_{d_1} \otimes E\left\{\mathbf{B}(\mathbf{X}) \mathbf{B}(\mathbf{X})^T\right\} \leq C_\mathbf{Q} C_\mathbf{V} I_{d_1\{d_2(N+1)+1\}}.$$

One can prove similarly the result for $\mathbf{S_T}$. □

Lemma 3.6.1 and Assumption (A3) ensure the existence of functions $g_{\alpha l} \in G^{(0)}[0,1]$ such that

$$\|g_{\alpha l} - m_{\alpha l}\|_\infty \leq C_\infty \|m'_{\alpha l}\|_\infty H^2, \alpha = 1, ..., d_2, l = 1, ..., d_1. \quad (4.7.1)$$

4.7.2 Oracle smoothers

In this section, we prove Theorems 4.3.1 and 4.3.2 for $\tilde{m}_{K,1,\cdot}(x_1)$. Corresponding proof for $\tilde{m}_{LL,1,\cdot}(x_1)$ would require replacing $K_h(X_{i1} - x_1)$ by $K_h(X_{i1} - x_1)\left(\frac{X_{i1} - x_1}{h}\right)$ in the proof, which does not add a great deal of difficulty. According to (4.3.1),

$$\tilde{m}_{K,1,\cdot}(x_1) - m_{1,\cdot}(x_1) = \left(\frac{1}{n} \mathbf{C}_K^T \mathbf{W}_1 \mathbf{C}_K\right)^{-1} \frac{1}{n} \mathbf{C}_K^T \mathbf{W}_1 \left\{\mathbf{Y}_1 - \mathbf{C}_K m_{1,\cdot}(x_1)\right\},$$

$$\mathbf{Y}_1 - \mathbf{C}_{\mathrm{K}} m_{1,\cdot}(x_1) = \left[\sum_{l=1}^{d_1} \{m_{1l}(X_{i1}) - m_{1l}(x_1)\} T_{il} + \sigma(\mathbf{X}_i, \mathbf{T}_i)\varepsilon_i \right]_{i=1}^{n},$$

then $\frac{1}{n}\mathbf{C}_{\mathrm{K}}^T \mathbf{W}_1 (\mathbf{Y}_1 - \mathbf{C}_{\mathrm{K}} m_{1,\cdot}(x_1))$ is

$$\left[\frac{1}{n}\sum_{i=1}^{n} K_h(X_{i1} - x_1) T_{il'} \left[\sum_{l=1}^{d_1} \{m_{1l}(X_{i1}) - m_{1l}(x_1)\} T_{il} + \sigma(\mathbf{X}_i, \mathbf{T}_i)\varepsilon_i \right] \right]_{l'=1}^{d_1}$$

$$= \{B_{l'}(x_1) + V_{l'}(x_1)\}_{l'=1}^{d_1}$$

where

$$B_{l'}(x_1) = \frac{1}{n}\sum_{i=1}^{n} K_h(X_{i1} - x_1) T_{il'} \sum_{l=1}^{d_1} \{m_{1l}(X_{i1}) - m_{1l}(x_1)\} T_{il} = \sum_{l=1}^{d_1} B_{l,l'}(x_1),$$

$$B_{l,l'}(x_1) = \frac{1}{n}\sum_{i=1}^{n} K_h(X_{i1} - x_1)\{m_{1l}(X_{i1}) - m_{1l}(x_1)\} T_{il} T_{il'}, \tag{4.7.2}$$

$$V_{l'}(x_1) = \frac{1}{n}\sum_{i=1}^{n} K_h(X_{i1} - x_1) T_{il'} \sigma(\mathbf{X}_i, \mathbf{T}_i)\varepsilon_i. \tag{4.7.3}$$

Denoting $D_{l,l'}(x_1) = \frac{1}{n}\sum_{i=1}^{n} K_h(X_{i1} - x_1) T_{il} T_{il'}$, the dispersion matrix is

$$\frac{1}{n}\mathbf{C}_{\mathrm{K}}^T \mathbf{W}_1 \mathbf{C}_{\mathrm{K}} = \left(\frac{1}{n}\sum_{i=1}^{n} K_h(X_{i1} - x_1) T_{il} T_{il'} \right)_{l,l'=1}^{d_1} = \left(D_{l,l'}(x_1) \right)_{l,l'=1}^{d_1}.$$

LEMMA 4.7.4. *Under Assumptions (A1) to (A4), (A6) to (A7), as $n \to \infty$,*

$$\sup_{1 \le l,l' \le d_1} \sup_{x_1 \in [h, 1-h]} \left| B_{l,l'}(x_1) - b_{\mathrm{K},l,l',1}(x_1) h^2 \right| = O_p\left(h^{1/2} \log n / \sqrt{n} \right)$$

where for any $x_1 \in [h, 1-h]$,

$$b_{\mathrm{K},l,l',1}(x_1) = \frac{1}{2}\mu_2(K)\left\{ 2 m_{1l}'(x_1) \frac{\partial f_1(x_1) q_{ll',1}(x_1)}{\partial x_1} + m_{1l}''(x_1) f_1(x_1) q_{ll',1}(x_1) \right\}.$$

Proof. We write the bias term $B_{l,l'}(x_1)$ in (4.7.2) as

$$h^{1/2} n^{-1} \sum_{i=1}^{n} \zeta_{i,n} + E K_h(X_1 - x_1)\{m_{1l}(X_1) - m_{1l}(x_1)\} T_l T_{l'}$$

where $\zeta_{i,n} = \zeta_{i,n}(x_1, X_{i1}, T_{il} T_{il'})$ is

$$h^{-1/2}\left[K_h(X_{i1} - x_1)\{m_{1l}(X_{i1}) - m_{1l}(x_1)\} T_{il} T_{il'} \right.$$

$$-EK_h(X_1 - x_1)\{m_{1l}(X_1) - m_{1l}(x_1)\} T_l T_{l'}] .$$

The deterministic part of $B_{l,l'}(x_1)$ is

$$EK_h(X_1 - x_1)\{m_{1l}(X_1) - m_{1l}(x_1)\} T_l T_{l'}$$

$$= \int_{[0,1]} \int t_l t_{l'} \{m_{1l}(u) - m_{1l}(x_1)\} \frac{1}{h} K\left(\frac{u - x_1}{h}\right) f(u, t_l, t_{l'}) \, dt_l dt_{l'} du$$

$$= \int_{[-1,1]} \int t_l t_{l'} \{m_{1l}(x_1 + hv) - m_{1l}(x_1)\} K(v) f(x_1 + hv, t_l, t_{l'}) \, dt_l dt_{l'} dv$$

$$= \int_{[-1,1]} \int t_l t_{l'} K(v) \left\{ m'_{1l}(x_1) hv + \frac{m''_{1l}(x_1)}{2}(hv)^2 + u\left(h^2\right) \right\}$$
$$\left\{ f(x_1, t_l, t_{l'}) + \frac{\partial f(x_1, t_l, t_{l'})}{\partial x_1} hv + u(h) \right\} dt_l dt_{l'} dv$$

$$= h^2 \mu_2(K) \left\{ m'_{1l}(x_1) \frac{\partial f_1(x_1) E(T_l T_{l'} | X_1 = x_1)}{\partial x_1} + \frac{m''_{1l}(x_1) f_1(x_1)}{2} E(T_l T_{l'} | X_1 = x_1) \right\}$$
$$+ u\left(h^3\right).$$

According to Assumption (A4),

$$EK_h(X_1 - x_1)\{m_{1l}(X_1) - m_{1l}(x_1)\} T_l T_{l'}$$

$$= h^2 \mu_2(K) \left\{ m'_{1l}(x_1) \frac{\partial f_1(x_1) q_{ll',1}(x_1)}{\partial x_1} + \frac{m''_{1l}(x_1) f(x_1)}{2} q_{ll',1}(x_1) \right\} + u\left(h^3\right). \quad (4.7.4)$$

To bound the stochastic part of $B_{l,l'}(x_1)$, define a sequence $D_n = n^\alpha$ with $0 < \alpha < \frac{2}{5}$, $\alpha(2+\delta) > 1, \alpha(1+\delta) > 2/5$, which requires $\delta > 1/2$ provided by Assumption (A2). We make use of the following truncation and tail decomposition

$$\mathrm{T}_{ill'} = T_{il} T_{il'} = \mathrm{T}_{ill',1}^{D_n} + \mathrm{T}_{ill',2}^{D_n} \quad (4.7.5)$$

where $\mathrm{T}_{ill',1}^{D_n} = T_{il} T_{il'} \{|T_{il} T_{il'}| > D_n\}, \mathrm{T}_{ill',2}^{D_n} = T_{il} T_{il'} \{|T_{il} T_{il'}| \le D_n\}$. Define correspondingly the truncated and tail parts of $\zeta_{i,n}$ as

$$\zeta_{i,n,1} = \zeta_{i,n}\left(x_1, X_i, \mathrm{T}_{ill',1}^{D_n}\right), \zeta_{i,n,2} = \zeta_{i,n}\left(x_1, X_i, \mathrm{T}_{ill',2}^{D_n}\right).$$

According to Assumption (A2),

$$\sum_{n=1}^{\infty} P\left(\left|T_{il}T_{il'}\right| > D_n\right) \leq \sum_{n=1}^{\infty} \frac{E\left|T_{il}T_{il'}\right|^{(2+\delta)}}{D_n^{(2+\delta)}} = \sum_{n=1}^{\infty}\left[\frac{E\left(\left|T_{il}T_{il'}\right|^{2+\delta}|\mathbf{X}_i\right)}{D_n^{2+\delta}}\right]$$

$$\leq \sum_{n=1}^{\infty} \frac{C_\delta}{D_n^{2+\delta}} = C_\delta^k \sum_{n=1}^{\infty} n^{-\alpha(2+\delta)} < \infty.$$

By Borel-Cantelli Lemma, one has with probability 1,

$$n^{-1}\sum_{i=1}^{n}\left\{T_{ill',1}^{D_n}\left\{m_{1l}\left(X_{i1}\right) - m_{1l}\left(x_1\right)\right\} K_h\left(X_{i1} - x_1\right)\right\} = 0$$

for large n. Therefore, one has

$$\sup_{x_1 \in [0,1]}\left|n^{-1}\sum_{i=1}^{n}\zeta_{i,n,1}\left(x_1, X_i, \mathrm{T}_{ill',1}^{D_n}\right)\right| = U\left(n^{-k}\right), k = 1,2,3.... \tag{4.7.6}$$

Next,

$$E\zeta_{i,n}^2 = h^{-1}\left[E\left\{\mathrm{T}_{ill'}\left\{m_{11}\left(X_{i1}\right) - m_{11}\left(x_1\right)\right\} K_h\left(X_{i1} - x_1\right)\right\}^2 - \left\{E\mathrm{T}_{ill'}\left\{m_{11}\left(X_{i1}\right) - m_{11}\left(x_1\right)\right\} K_h\left(X_{i1} - x_1\right)\right\}^2\right]$$

$$= h^{-1}\int_{[0,1]}\int t_l^2 t_{l'}^2 \left\{m_{11}(u) - m_{11}(x_1)\right\}^2 \frac{1}{h^2} K\left(\frac{u - x_1}{h}\right)^2 f\left(u, t_l, t_{l'}\right) dt_l dt_{l'} du + U\left(h^3\right)$$

$$= h^{-1}\int_{[-1,1]}\int t_l^2 t_{l'}^2 \frac{1}{h} K(v)^2 \left[\left\{m_{11}'(x_1) hv\right\}^2 f\left(x_1, t_l, t_{l'}\right) + U\left(h^3\right)\right] dt_l dt_{l'} dv + U\left(h^3\right)$$

$$= h^{-2}\int_{[-1,1]}\int t_l^2 t_{l'}^2 K(v)^2 \left\{m_{11}'(x_1)\right\}^2 f\left(x_1, t_l, t_{l'}\right)(hv)^2 dt_l dt_{l'} dv + u\left(h^2\right) + U\left(h^3\right)$$

$$= \left\{m_{11}'(x_1)\right\}^2 \int_{[-1,1]} K(v)^2 v^2 dv \int t_l^2 t_{l'}^2 f\left(x_1, t_l, t_{l'}\right) dt_l dt_{l'} + u\left(h^2\right)$$

$$= \left\{m_{11}'(x_1)\right\}^2 \int_{[-1,1]} K(v)^2 v^2 dv f_1(x_1) E\left(T_{il}^2 T_{il'}^2 | X_1\right) + u\left(h^2\right)$$

Then

$$E\zeta_{i,n,2}^2 = E\zeta_{i,n}^2 - E\zeta_{i,n,1}^2 = \left\{m_{11}'(x_1)\right\}^2 \int_{[-1,1]} K(v)^2 v^2 dv f_1(x_1) E\left(T_l^2 T_{l'}^2 | X_1\right) + u\left(n^{-1}\right).$$

For $k \geq 3$, $E\left|\zeta_{i,n,2}\right|^k$

$$\leq \left[\sup_{x_1 \in [0,1]}\left|\mathrm{T}_{ill',1}^{D_n}\left\{m_{11}\left(X_{i1}\right) - m_{11}\left(x_1\right)\right\} K_h\left(X_{i1} - x_1\right)\right| + U\left(h^2\right)\right]^{k-2} E(\zeta_{i,n,2}^2)$$

$$\leq \left|\mathrm{T}_{ill',1}^{D_n}\right|^{k-2} \sup_{x_1\in[0,1]} [\{m_{11}(X_{i1}) - m_{11}(x_1)\} K_h(X_{i1}-x_1)]^{k-2} E(\zeta_{i,n,2}^2)$$
$$\leq D_n^{k-2} c_0 h h^{-1} E(\zeta_{i,n,2}^2) = c_0 D_n^{k-2} E(\zeta_{i,n,2}^2)$$

by Assumptions (A4) and (A7). So there exist a constant $c_1 = c_0 D_n$ such that $E\left(|\zeta_{i,n,2}|^k\right) \leq c_1^{k-2} k! E(\zeta_{i,n,2}^2)$, $k \geq 2$. According to Lemma 2.5.2 (Bernstein's inequality),

$$P\left\{\left|\sum_{i=1}^n \zeta_{i,n,2}\right| > n\varepsilon_n\right\} \leq a_1 \exp\left(-\frac{q\varepsilon_n^2}{25m_2^2 + 5c_1\varepsilon_n}\right) + a_2(k)\,\alpha\left(\left[\frac{n}{q+1}\right]\right)^{\frac{2k}{2k+1}}$$

$$a_1 = 2\frac{n}{q} + 2\left(1 + \frac{\varepsilon_n^2}{25m_2^2 + 5c_1\varepsilon_n}\right), a_2(k) = 11n\left(1 + \frac{5m_k^{2k/(2k+1)}}{\varepsilon_n}\right).$$

Let $k = 3, a_2(3) = 11n\left(1 + \dfrac{5m_3^{6/7}}{\varepsilon_n}\right), m_2^2 = E(\zeta_{i,n,2}^2) = O(1), \varepsilon_n = a\dfrac{\log n}{\sqrt{n}}$

$$P\left\{\left|\sum_{i=1}^n \zeta_{i,n,2}\right| > n\varepsilon_n\right\} \leq a_1 \exp\left(-\frac{q\varepsilon_n^2}{25m_2^2 + 5c_1\varepsilon_n}\right) + a_2(3)\,\alpha\left(\left[\frac{n}{q+1}\right]\right)^{\frac{6}{7}}$$

and take q such that $\left[\frac{n}{q+1}\right] \geq c_2 \log n, q \geq \dfrac{c_3 n}{\log n}$ for some constants c_2, c_3.

$$\frac{q\varepsilon_n^2}{25m_2^2 + 5c\varepsilon_n} = \frac{q\dfrac{(a\log n)^2}{n}}{25m_2^2 + 5c_1\varepsilon_n} \geq \frac{\dfrac{c_3 n}{\log n} a^2 \dfrac{(\log n)^2}{n}}{25m_2^2 + 5c_1 a\dfrac{\log n}{\sqrt{nh}}}$$
$$\geq \frac{c_3 a^2 \log n}{25m_2^2 + 5c_0 D_n a\dfrac{\log n}{\sqrt{nh}}} = \frac{c_3 a^2 \log n}{25m_2^2 + 5ac_0 n^{\alpha} n^{-2/5}\log n} \sim a^2 \log n,$$

$$a_1 = 2\frac{n}{q} + 2\left(1 + \frac{\varepsilon_n^2}{25m_2^2 + 5c_1\varepsilon_n}\right) = O(\log n),$$

$$a_2(3) = 11n\left(1 + \frac{5m_3^{6/7}}{\varepsilon_n}\right), \text{ with } m_3 = \max_{1\leq i\leq N} \|\xi_{i,n,2}\|_3 \leq c_6 D_n,$$

$$a_2(3) \leq 11n\left\{1 + \frac{(c_6 D_n)^{6/7}}{an^{-\frac{1}{2}}\log n}\right\} = o\left(n^2\right),$$

$$\alpha\left(\left[\frac{n}{q+1}\right]\right)^{6/7} \leq \left(K_0 e^{-\lambda_0\left[\frac{n}{q+1}\right]}\right)^{6/7} \leq Cn^{-6\lambda_0 c_2/7},$$

$$P\left\{n^{-1}\left|\sum_{i=1}^{n}\zeta_{i,n,2}\right|>a\log n/\sqrt{n}\right\} \leq O(\log n)\exp\left(-c_5a^2\log n\right)+Cn^{2-6\lambda_0c_2/7}$$
$$= n^{-c_5a^2}O(\log n)+Cn^{2-6\lambda_0c_2/7},$$

for c_0, c_2, a large enough. For all $x_1 \in [h, 1-h]$, we discretize by equally spaced $h = x_{1,0} < x_{1,1} < \cdots < x_{1,M_n} = 1-h$, $M_n = n^4$,

$$P\left\{\max_{0\leq j\leq M_n} n^{-1}\left|\sum_{i=1}^{n}\zeta_{i,n,2}\left(x_{1,j}\right)\right|>a\log n/\sqrt{n}\right\}$$
$$\leq \sum_{0=1}^{M_n}P\left\{n^{-1}\left|\sum_{i=1}^{n}\zeta_{i,n,2}\left(x_{1,j}\right)\right|>a\log n/\sqrt{n}\right\}\leq Cn^{-8}M_n\leq Cn^{-2}$$

for a and c_2 large enough. Borel-Cantelli lemma implies that $\max_{1\leq j\leq M_n}\left|n^{-1}\sum_{i=1}^{n}\zeta_{i,n}\left(x_{1,j}\right)\right| = O_p\left(a\log n/\sqrt{n}\right)$ *a.s.*. Taking supremum over the whole interval $[h, 1-h]$, one has

$$\sup_{x_1\in[h,1-h]}\left|n^{-1}\sum_{i=1}^{n}\zeta_{i,n,2}\right| \leq \max_{0\leq j\leq M_n}\left|n^{-1}\sum_{i=1}^{n}\zeta_{i,n,2}\right|+$$
$$n^{-1}\max_{0\leq j\leq M_n-1}\sup_{x_1\in[x_{1,j},x_{1,j+1}]}\left|\sum_{i=1}^{n}\zeta_{i,n,2}-\sum_{i=1}^{n}\zeta_{i,n,2}\right| \leq Cn^{-1/2}\log n+CM_n^{-1}h^{-2}.$$

by Lipschitz continuity of kernel K. This last equation, plus (4.7.4), (4.7.5) and (4.7.6) complete the proof of lemma. □

LEMMA 4.7.5. *Under Assumptions (A1) to (A3), (A6) to (A7), as $n\to\infty$,*

$$\sup_{1\leq l,l'\leq d_1}\sup_{x_1\in[h,1-h]}\left|D_{l,l'}\left(x_1\right)-f\left(x_1\right)q_{ll',1}\left(x_1\right)+\frac{1}{2}h^2\mu_2\left(K\right)\frac{\partial^2 f\left(x_1\right)q_{ll',1}\left(x_1\right)}{\partial x_1^2}\right|$$
$$= O_p\left(n^{-1/2}\log n\right).$$

Proof. For any $x_1 \in [h, 1-h]$

$$D_{l,l'}\left(x_1\right)=\frac{1}{n}\sum_{i=1}^{n}K_h\left(X_{i1}-x_1\right)T_{il}T_{il'}$$

$$=\frac{1}{n}\sum_{i=n}\left\{T_{il}T_{il'}K_h\left(X_{i1}-x_1\right)-ET_lT_{l'}K_h\left(X_1-x_1\right)\right\}+ET_lT_{l'}K_h\left(X_1-x_1\right).$$

The deterministic part is

$$ET_lT_{l'}K_h\left(X_1-x_1\right)=\int_{[0,1]}\int t_lt_{l'}\frac{1}{h}K\left(\frac{u-x_1}{h}\right)f\left(u,t_l,t_{l'}\right)dt_ldt_{l'}du$$

$$= \int_{[-1,1]} \int t_l t_{l'} K(v) f(x_1 + hv, t_l, t_{l'})\, dt_l dt_{l'} dv$$

$$= \int_{[-1,1]} \int t_l t_{l'} K(v) \left\{ f(x_1, t_l, t_{l'}) + \frac{\partial f(x_1, t_l, t_{l'})}{\partial x_1} hv + 2^{-1} \frac{\partial^2 f(x_1, t_l, t_{l'})}{\partial x_1^2} (hv)^2 \right.$$
$$\left. + u\left(h^2\right) \right\} dt_l dt_{l'} dv$$

$$= \int t_l t_{l'} f(x_1, t_l, t_{l'})\, dt + \frac{1}{2}\mu_2(K) h^2 \int t_l t_{l'} \frac{\partial^2 f(x_1, t_l, t_{l'})}{\partial x_1^2} dt_l dt_{l'} + u\left(h^3\right)$$

$$= f(x_1) E\left(T_l T_{l'} | X_1 = x_1\right) + \frac{1}{2}\mu_2(K) h^2 \frac{\partial^2 f_1(x_1) E\left(T_l T_{l'} | X_1 = x_1\right)}{\partial x_1^2} + u\left(h^3\right).$$

Applying similar techniques as in Lemma 4.7.4, one can bound the stochastic part as

$$\sup_{x_1 \in [h, 1-h]} \left| \frac{1}{n} \sum_{i=n} \left\{ T_{il} T_{il'} K_h(X_{i1} - x_1) - E T_l T_{l'} K_h(X_1 - x_1) \right\} \right| = O_p\left(n^{-1/2} \log n\right).$$

□

Define next

$$\xi_{i,n,l'} = \xi_{i,n,l'}(x_1, \mathbf{X}_i, \mathbf{T}_i) = T_{il'} \sigma(\mathbf{X}_i, \mathbf{T}_i) \varepsilon_i K_h(X_{i1} - x_1). \tag{4.7.7}$$

Then the noise term $V_{l'}(x_1)$ in (4.7.3) equals to

$$\frac{1}{n} \sum_{i=1}^{n} K_h(X_{i1} - x_1) T_{il'} \sigma(\mathbf{X}_i, \mathbf{T}_i) \varepsilon_i = \frac{1}{n} \sum_{i=1}^{n} \xi_{i,n,l'}.$$

LEMMA 4.7.6. *Under Assumptions (A1) to (A3), (A5) and (A7), as $n \to \infty$,*

$$\sup_{x_1 \in [h, 1-h]} \left| E \xi_{i,n,l'} \xi_{i,n,l''} - h^{-1} f_1(x_1) E\left(T_{l'} T_{l''} \sigma^2(\mathbf{X}, \mathbf{T}) | X_1 = x_1\right) \|K\|_2^2 \right| = O(h),$$

$$\sup_{x_1 \in [h, 1-h]} \left| E \xi_{i,n,l'}^2 - h^{-1} f_1(x_1) E\left(T_{l'}^2 \sigma^2(\mathbf{X}, \mathbf{T}) | X_1 = x_1\right) \|K\|_2^2 \right| = O(h).$$

Proof. According to (4.7.7),

$$E \xi_{i,n,1} \xi_{i,n,2} = E T_1 T_2 \sigma^2(\mathbf{X}, \mathbf{T}) K_h^2(X_1 - x_1)$$

$$= \int_{[0,1]^{d_2}} \int_{\mathbf{R}^{d_1}} t_1 t_2 \sigma^2(\mathbf{u}, \mathbf{t}) \frac{1}{h^2} K\left(\frac{u_1 - x_1}{h}\right)^2 f(\mathbf{u}, \mathbf{t})\, d\mathbf{t} d\mathbf{u}$$

$$= \int_{\mathbf{R}^{d_1}} \int_{[0,1]^{d_2}} t_1 t_2 \sigma^2(\mathbf{u}, \mathbf{t}) \frac{1}{h^2} K\left(\frac{u_1 - x_1}{h}\right)^2 f(u_1, \mathbf{u}_{\text{-}1}, \mathbf{t})\, du_1 d\mathbf{u}_{\text{-}1} d\mathbf{t}$$

$$= h^{-1} \int_{\mathbf{R}^{d_1}} \int_{[0,1]^{d_2-1}} \int_{[-1,1]} t_1 t_2 \sigma^2 (x_1 + hv_1, \mathbf{u}_{\text{-}1}, \mathbf{t}) K(v_1)^2 f(x_1 + hv_1, \mathbf{u}_{\text{-}1}, \mathbf{t}) dv_1 d\mathbf{u}_{\text{-}1} d\mathbf{t}$$

$$= h^{-1} \int_{\mathbf{R}^{d_1}} \int_{[0,1]^{d_2-1}} \int_{[-1,1]} t_1 t_2$$
$$\left\{ \sigma^2 (x_1, \mathbf{u}_{\text{-}1}, \mathbf{t}) + \frac{\partial \sigma^2 (x_1, \mathbf{u}_{\text{-}1}, \mathbf{t})}{\partial x_1} hv_1 + \frac{\partial^2 \sigma^2 (x_1, \mathbf{u}_{\text{-}1}, \mathbf{t})}{2\partial x_1^2} (hv_1)^2 + u\left(h^2\right) \right\}$$
$$K(v_1)^2 \left\{ f(x_1, \mathbf{u}_{\text{-}1}, \mathbf{t}) + \frac{\partial f (x_1, \mathbf{u}_{\text{-}1}, \mathbf{t})}{\partial x_1} hv_1 + \frac{\partial^2 f (x_1, \mathbf{u}_{\text{-}1}, \mathbf{t})}{2\partial x_1^2} (hv_1)^2 + u\left(h^2\right) \right\}$$
$$dv_1 d\mathbf{u}_{\text{-}1} d\mathbf{t}$$

$$= h^{-1} \int_{[-1,1]} K(v_1)^2 dv_1 \int_{\mathbf{R}^{d_1}} \int_{[0,1]^{d_2-1}} t_1 t_2 \sigma^2 (x_1, \mathbf{u}_{\text{-}1}, \mathbf{t}) f(x_1, \mathbf{u}_{\text{-}1}, \mathbf{t}) d\mathbf{u}_{\text{-}1} d\mathbf{t} + U(h)$$
$$= h^{-1} f_1(x_1) E\left(T_1 T_2 \sigma^2 (\mathbf{X}, \mathbf{T}) | X_1 = x_1\right) \|K\|_2^2 + U(h).$$

Similarly, one has for any l', l''

$$E\xi_{i,n,l'}\xi_{i,n,l''} = h^{-1} f_1(x_1) E\left(T_{l'} T_{l''} \sigma^2 (\mathbf{X}, \mathbf{T}) | X_1 = x_1\right) \|K\|_2^2 + U(h),$$
$$E\xi_{i,n,l'}^2 = h^{-1} f_1(x_1) E\left(T_{l'}^2 \sigma^2 (\mathbf{X}, \mathbf{T}) | X_1 = x_1\right) \|K\|_2^2 + U(h). \square$$

LEMMA 4.7.7. *Under Assumptions (A1) to (A3), (A5) and (A7), as $n \to \infty$, there exists a constant C such that*

$$\sup_{1 \le l' \le d_1} \sup_{x_1 \in [h, 1-h]} \left| \operatorname{cov}\left(\xi_{i,n,l'}, \xi_{j,n,l''}\right) \right| \le C h^{-\frac{1+\eta}{2+\eta}} \alpha (j-i)^{\frac{\eta}{2+\eta}} \text{ for } i \ne j$$

Proof. According to Davydov's Inequality [Bosq 1998, p. 21. equation (1.10)], for $\frac{1}{p} + \frac{1}{q} + \frac{1}{r} = 1$, $\operatorname{cov}\left(\xi_{i,n,l_1'}, \xi_{j,n,l_2'}\right)$ is bounded by

$$C_2 \{2\alpha(j-i)\}^{1/p} \left\|T_{il'} \sigma(\mathbf{X}_i, \mathbf{T}_i) \varepsilon_i K_h (X_{i1} - x_1)\right\|_q \left\|T_{jl''} \sigma(\mathbf{X}_j, \mathbf{T}_j) \varepsilon_i K_h (X_{j1} - x_1)\right\|_r$$

Let $q = r = 2 + \eta, p = 1 + 2/\eta$, where η takes value in the Assumption (A5), then one has $\operatorname{cov}\left(\xi_{i,n,l'}, \xi_{j,n,l''}\right) \le C h^{-\frac{1+\eta}{2+\eta}} \alpha (j-i)^{\frac{\eta}{2+\eta}}$ for some constant C. $\square$

Proof of Theorem 4.3.1. For any $\boldsymbol{\lambda} = \left(\lambda_1, ..., \lambda_{d_1}\right)^T \in \mathbf{R}^{d_1}$,

$$\boldsymbol{\lambda}^T \left\{V_{l'}(x_1)\right\}_{l'=1}^{d_1} = \sum_{l'=1}^{d_1} \lambda_{l'} V_{l'}(x_1) = \sum_{l'=1}^{d_1} \lambda_{l'} \frac{1}{n} \sum_{i=1}^{n} \xi_{i,n,l'} = \frac{1}{n} \sum_{i=1}^{n} \sum_{l'=1}^{d_1} \lambda_{l'} \xi_{i,n,l'}.$$

Define $\xi_{i,n} = \sum_{l'=1}^{d_1} \lambda_{l'} \xi_{i,n,l'}$ and $S_n = S_n(x_1) = \sum_{i=1}^{n} \xi_{i,n} = n\boldsymbol{\lambda}^T \{V_{l'}(x_1)\}_{l'=1}^{d_1}$, then one has $ES_n = 0$. Let

$$\gamma(k) = \gamma(k, x_1) = \operatorname{cov}(\xi_{i,n}, \xi_{i+k,n}) = \operatorname{cov}\left(\sum_{l'=1}^{d_1} \lambda_{l'} \xi_{i,n,l'}, \sum_{l'=1}^{d_1} \lambda_{l'} \xi_{i+k,n,l'}\right)$$

$$\sigma_n^2 = ES_n^2 = \operatorname{var}(S_n) = \operatorname{var}\left(\sum_{i=1}^{n} \xi_{i,n}\right) = \sum_{i=1}^{n} \operatorname{var}(\xi_{i,n}) + \sum_{i \neq j} \operatorname{cov}(\xi_{i,n}, \xi_{j,n})$$

$$= n \operatorname{var}(\xi_{i,n}) + n \sum_{1 \leq |k| \leq n-1} \left(1 - \frac{|k|}{n}\right) \gamma(k) = n \operatorname{var}(\xi_{i,n}) + nA_n.$$

In the above, $\operatorname{var}(\xi_{i,n}) = \operatorname{var}\left(\sum_{l'=1}^{d_1} \lambda_{l'} \xi_{i,n,l'}\right) = h^{-1} \boldsymbol{\lambda}^T \Sigma \boldsymbol{\lambda}$ where

$$\begin{aligned} \Sigma &= h \left\{\operatorname{cov}\left(\xi_{i,n,l'}, \xi_{i,n,l''}\right)\right\}_{l',l''=1}^{d_1} = hE\left\{\xi_{i,n,l'} \xi_{i,n,l''}\right\}_{l',l''=1}^{d_1} \\ &= f(x_1) \|K\|_2^2 E\left\{\mathbf{T}\mathbf{T}^T \sigma^2(\mathbf{X}, \mathbf{T}) | X_1 = x_1\right\} \end{aligned}$$

by Lemma 4.7.6. While according to Lemma 4.7.7, one has

$$|\gamma(k)| \leq d_1^2 \max_{1 \leq l' \leq d_1} \left|\operatorname{cov}\left(\xi_{i,n,l'}, \xi_{i+k,n,l''}\right)\right| \leq Ch^{-\frac{1+\eta}{2+\delta}} \alpha(k)^{\frac{\eta}{2+\eta}}.$$

Hence

$$\begin{aligned} |A_n| &= \left|\sum_{1 \leq |l| \leq n-1} \gamma(k)\right| \leq \sum_{1 \leq |l| \leq n-1} \left(1 - \frac{|k|}{n}\right) h^{-\frac{1+\eta}{2+\eta}} \{K_0 \exp(-\lambda_0 k)\}^{\frac{\eta}{2+\eta}} \\ &\leq K_0 h^{-\frac{1+\eta}{2+\eta}} \sum_{1 \leq |l| \leq n-1} \exp\{-\lambda_0 k\eta / (2+\eta)\}, \end{aligned}$$

so there exists a constant C_1 such that $A_n \leq C_1 h^{-\frac{1+\eta}{2+\eta}}$. So $A_n / \operatorname{var}(\xi_{i,n}) \to 0$ as $n \to \infty$. Then $\sigma_n^2 \sim n \operatorname{var}(\xi_{i,n}) \geq c_0 n$ when n is large, so according to (2.5.1) in Lemma 2.5.1, there exist constants c_1 and c_2 such that for some $0 < \eta \leq 1$

$$\Delta_n = \sup_z \left|P\left\{\sigma_n^{-1} S_n < z\right\} - \Phi(z)\right| \leq c_1 \frac{d_\eta}{c_0 \sigma_n^\eta} \left\{\log\left(\sigma_n / c_0^{1/2}\right) / \lambda\right\}^{1+\eta}$$

for any λ with $\lambda_1 \leq \lambda \leq \lambda_2$, where

$$\lambda_1 = c_2 \left\{\log\left(\sigma_n / c_0^{1/2}\right)\right\}^b / n, b > 2(1+\eta)/\eta; \lambda_2 = 4(2+\eta)\eta^{-1} \log\left(\sigma_n / c_0^{1/2}\right).$$

For the η in Assumption (A5), set $\lambda = 4(2+\eta)\eta^{-1}\log\left(\sigma_n/c_0^{1/2}\right)$, then by (4.2.4) one has

$$d_\eta = \max_{1\le i\le n}\left\{E|\sum_{l'=1}^{d_1}\lambda_{l'}\xi_{i,n,l'}|^{2+\eta}\right\} = \max_{1\le i\le n}\left\{E|\sum_{l'=1}^{d_1}\lambda_{l'}T_{il'}\sigma(\mathbf{X}_i,\mathbf{T}_i)\varepsilon_i K_h(X_{i1}-x_1)|^{2+\eta}\right\}$$

$$\le CC_\delta C_\eta\left\{E|\sum_{l'=1}^{d_1}K_h(X_1-x_1)|^{2+\eta}\right\} = O\left\{h^{-(1+\eta)}\right\},$$

i.e., $\Delta_n = O\left\{h^{-(1+\eta)}/\sigma_n^\eta\right\} = O\left\{n^{(1+\eta/2)/5-\eta/2}\right\} = O\left(n^{1/5-2\eta/5}\right) \to 0$ when $1/2<\eta\le 1$. So $S_n/\sigma_n \to N(0,1)$, then $\sqrt{nh}\boldsymbol{\lambda}^T\{V_{l'}(x_1)\}_{l'=1}^{d_1} \to N\left(0,\boldsymbol{\lambda}^T\Sigma\boldsymbol{\lambda}\right)$. By Cramér-Wold device, one has $\sqrt{n}\{V_{l'}(x_1)\}_{l'=1}^{d_1} \to N(0,\Sigma)$. Then according to Slutsky's theorem, one has

$$\sqrt{nh}E\left(\mathbf{T}\mathbf{T}^T|X_1=x_1\right)\left\{\tilde{m}_{K,1,\cdot}(x_1)-m_{1,l'}(x_1)-\sum_{l=1}^{d_1}b_{l,l'}(x_1)h^2\right\}_{l'=1}^{d_1} \to N(0,\Sigma)$$

i.e., $\sqrt{nh}\left\{\tilde{m}_{K,1,\cdot}(x_1)-m_{1,l'}(x_1)-\sum_{l=1}^{d_1}b_{l,l'}(x_1)h^2\right\}_{l'=1}^{d_1} \to N\left(0,\mathbf{Q}_1(x_1)^{-1}\Sigma\mathbf{Q}_1(x_1)^{-1}\right)$, where $\mathbf{Q}_1(x_1)$ is defined in (4.2.3). □

Proof of Theorem 4.3.2. Let $D_n = n^\alpha$ with $\alpha < \frac{2}{5}$, $\alpha(2+\eta)>1, \alpha(1+\eta)>2/5$, which requires $\eta>1/2$. Rewrite $Z_i = T_{il'}\varepsilon_i = Z_{i,1}^{D_n}+Z_{i,2}^{D_n}+Z_{i,3}^{D_n}$ where $Z_{i,1}^{D_n} = Z_i\{|Z_i|>D_n\}, Z_{i,2}^{D_n} = Z_i\{|Z_i|\le D_n\}-Z_{i,3}^{D_n}, Z_{i,3}^{D_n} = EZ_i\{|Z_i|\le D_n\}$. Define

$$\xi_{i,n,l',j} = K_h(X_{i1}-x_1)\sigma(\mathbf{X}_i,\mathbf{T}_i)Z_{i,j}^{D_n},\ j=1,2,3.$$

According to Assumption (A5) and (4.2.4), one has

$$\begin{aligned}\sum_{n=1}^\infty P(|Z_i|\ge D_n) &\le \sum_{n=1}^\infty \frac{C_\sigma E|Z_i|^{2+\eta}}{D_n^{2+\eta}} = C_\sigma\sum_{n=1}^\infty\frac{E\left\{|T_{l'}|^{2+\eta}E\left(|\varepsilon|^{2+\eta}|\mathbf{X},\mathbf{T}\right)\right\}}{D_n^{2+\eta}}\\ &\le C_\sigma C_\delta E|T_{l'}|^{2+\eta}\sum_{n=1}^\infty\frac{1}{D_n^{2+\eta}} = C_\sigma C_\delta E|T_{l'}|^{2+\eta}\sum_{n=1}^\infty n^{-\alpha(2+\eta)}<\infty.\end{aligned}$$

By Borel-Cantelli Lemma, one has with probability 1, $n^{-1}\sum_{i=1}^n\xi_{i,n,l',1}=0$ for large n. Therefore, one has $\sup_{x_1\in[0,1]}\left|n^{-1}\sum_{i=1}^n\xi_{i,n,l',1}\right| = U\left(n^{-k}\right)$ for any $k>0$. Using Assumption (A5) and (4.2.4)

$$\left|Z_{i,3}^{D_n}\right| = |-EZ_i\{|Z_i|>D_n\}| \le \frac{E|Z_i|^{2+\eta}}{D_n^{1+\eta}}$$

$$= E\left\{E\left|T_{l'}\right|^{2+\eta} E\left(|\varepsilon|^{2+\eta}|\mathbf{X},\mathbf{T}\right)\right\}/D_n^{1+\eta} = O\left(n^{-2/5}\right).$$

Hence

$$\begin{aligned} n^{-1}\sum_{i=n}\xi_{i,n,3} &= n^{-1}\sum_{i=n} K_h\left(X_{i1}-x_1\right)\sigma\left(\mathbf{X}_i,\mathbf{T}_i\right) Z_{i,3}^{D_n} \\ &= n^{-1}\sum_{i=n} K_h\left(X_{i1}-x_1\right) O\left(n^{-2/5}\right) = O_p\left(n^{-2/5}\right). \end{aligned}$$

Meanwhile

$$\left(Z_{i,2}^{D_n}\right)^2 = EZ_i^2\left\{|Z_i|\le D_n\right\} - \left(Z_{i,3}^{D_n}\right)^2 = EZ_i^2 - EZ_i^2\left\{|Z_i| > D_n\right\} - \left(Z_{i,3}^{D_n}\right)^2$$

$$\le E\left\{T_{il'}^2 E\left(\varepsilon_i^2|\mathbf{X}_i,\mathbf{T}_i\right)\right\} - EZ_i^{2+\eta}\left\{|Z_i| > D_n\right\}/D_n^V - \left(Z_{i,3}^{D_n}\right)^2 = ET_{il'}^2 + U_p\left(D_n^{-\eta} + n^{-4/5}\right),$$

$$E\xi_{i,n,l',2}^2 = E\left\{K_h\left(X_{i1}-x_1\right)\sigma\left(\mathbf{X}_i,\mathbf{T}_i\right) Z_{i,2}^{D_n}\right\}^2$$

$$= h^{-1} f\left(x_1\right) E\left(T_{l'}^2\sigma^2\left(\mathbf{X},\mathbf{T}\right)|X_1 = x_1\right)\|K\|_2^2\left\{1+u\left(1\right)\right\}$$

$$E\left|\xi_{i,n,l',2}\right|^k = E\left(\left|\xi_{i,n,l',2}\right|^{k-2}\left|\xi_{i,n,l',2}\right|^2\right)$$

$$\le \sup_{x_1\in[0,1]}\left|\xi_{i,n,l',2}\right|^{k-2} E\left|\xi_{i,n,l',2}\right|^2 \le C_\sigma 2^{k-2} D_n^{k-2}/h^{k-2} E\left|\xi_{i,n,l',2}\right|^2$$

according to Assumption (A3) and the truncation of $Z_{i,2}^{D_n}$, then there exist a constant $c_1 = C_\sigma D_n/h$ such that $E\left(\left|\xi_{i,n,l',2}\right|^k\right) \le c_1^{k-2} k! E(\xi_{i,n,l',2}^2)$, $k\ge 2$.

Similar to the proof of Lemma 4.7.4, by using Lemma 2.5.2 (Bernstein's inequality), we let $k = 3, a_2\left(3\right) = 11n\left(1+\dfrac{5m_3^{6/7}}{\varepsilon_n}\right)$, $m_2^2 = E(\xi_{i,n,2}^2) = O\left(h^{-1}\right)$, $\varepsilon_n = a\dfrac{\log n}{\sqrt{nh}}$

$$P\left\{\left|\sum_{i=1}^n \xi_{i,n,2}\right| > n\varepsilon_n\right\} \le a_1\exp\left(-\frac{q\varepsilon_n^2}{25m_2^2 + 5c_1\varepsilon_n}\right) + a_2\left(3\right)\alpha\left(\left[\frac{n}{q+1}\right]\right)^{\frac{6}{7}}$$

take q such that $\left[\frac{n}{q+1}\right] \ge c_2\log n, q \ge \dfrac{c_3 n}{\log n}$ for some constants c_2, c_3.

$$\frac{q\varepsilon_n^2}{25m_2^2 + 5c\varepsilon_n} = \frac{q\dfrac{(a\log n)^2}{nh}}{25m_2^2 + 5c_1\varepsilon_n} \ge \frac{\dfrac{c_3 n}{\log n}a^2\dfrac{(\log n)^2}{nh}}{25m_2^2 + 5c_1 a\dfrac{\log n}{\sqrt{nh}}}$$

$$\ge \frac{c_3 a^2\log n}{25m_2^2 h + 5c_0 D_n/ha\dfrac{\log n}{\sqrt{nh}}h} = \frac{c_3 a^2\log n}{25m_2^2 h + 5ac_0 n^\alpha n^{-1/2}h^{-1/2}\log n} \sim a^2\log n,$$

$$a_1 = 2\frac{n}{q} + 2\left(1 + \frac{\varepsilon_n^2}{25m_2^2 + 5c_1\varepsilon_n}\right) = O(\log n),$$

$$a_2(3) = 11n\left(1 + \frac{5m_3^{6/7}}{\varepsilon_n}\right), \text{ with } m_3 = \max_{1\le i\le N}\left\|\xi_{i,n,l',2}\right\|_3 \le c_6 D_n,$$

$$a_2(3) \le 11n\left\{1 + \frac{c_6 D_n}{an^{-1/2}h^{-1/2}\log n}\right\} = o\left(n^2\right),$$

$$\alpha\left(\left[\frac{n}{q+1}\right]\right)^{6/7} \le \left(K_0 e^{-\lambda_0\left[\frac{n}{q+1}\right]}\right)^{6/7} \le Cn^{-6\lambda_0 c_2/7},$$

therefore for large n

$$\begin{aligned} P\left\{n^{-1}\left|\sum_{i=1}^{n}\xi_{i,n,l',2}\right| > a\log n/\sqrt{n}\right\} &\le O(\log n)\exp\left(-c_5 a^2\log n\right) + Cn^{2-6\lambda_0 c_2/7} \\ &= n^{-c_5 a^2}O(\log n) + Cn^{2-6\lambda_0 c_2/7} \end{aligned}$$

which implies that

$$\sup_{x_1\in[0,1]}\left|n^{-1}\sum_{i=1}^{n}\xi_{i,n,l',2}\right| = O_p\left\{(nh)^{-1/2}\log n\right\}.$$

Then $\sup_{x_1\in[0,1]}\left|n^{-1}\sum_{i=1}^{n}\xi_{i,n,l'}\right| = O_p\left\{(nh)^{-1/2}\log n\right\}$ i.e.,

$$\sup_{x_1\in[0,1]}\left|V_{l'}(x_1)\right| = O_p\left\{(nh)^{-1/2}\log n\right\} \tag{4.7.8}$$

for the term $V_{l'}(x_1)$ in (4.7.3). According to Lemma 4.7.5,

$$\tilde{m}_{K,1,\cdot}(x_1) - m_{1,\cdot}(x_1) = \left(\frac{1}{n}\mathbf{C}^T\mathbf{W}_1\mathbf{C}\right)^{-1}\left\{\sum_{l=1}^{d_1}B_{l,l'}(x_1) + V_{l'}(x_1)\right\}_{l'=1}^{d_1}$$

$$= \left\{\left\{ET_lT_{l'}K_h(X_1 - x_1)\right\}_{l,l'=1}^{d_1} + O_p\left(n^{-1/2}\log n\right)\right\}^{-1}\left\{\sum_{l=1}^{d_1}B_{l,l'}(x_1) + V_{l'}(x_1)\right\}_{l'=1}^{d_1}$$

$$= \left[\left\{ET_lT_{l'}K_h(X_1 - x_1)\right\}_{l,l'=1}^{d_1}\right]^{-1}\left\{\sum_{l=1}^{d_1}B_{l,l'}(x_1) + V_{l'}(x_1)\right\}_{l'=1}^{d_1} + O_p\left(n^{-1/2}\log n\right),$$

$$\begin{aligned} \left[\left\{ET_lT_{l'}K_h(X_1 - x_1)\right\}_{l,l'=1}^{d_1}\right]^{-1} &= \left[f(x_1)\mathbf{Q}(x_1,\mathbf{X}_{\text{-}1}) + u\left(h^2\right)\right]^{-1} \\ &= f^{-1}(x_1)\mathbf{Q}^{-1}(x_1,\mathbf{X}_{\text{-}1}) + u\left(h^2\right)^{-1}. \end{aligned}$$

Meanwhile, according to Lemma 4.7.4 and (4.7.8),

$$\sum_{l=1}^{d_1} B_{l,l'}(x_1) + V_{l'}(x_1) = U_p\left(h^{1/2}\log n/\sqrt{n} + h^2\right) + U_p\left\{(nh)^{-1/2}\log n\right\}.$$

According to Assumptions (A1) and (A2), $f^{-1}(x_1) \le c_1^{-1}$, $C_{\mathbf{Q}}^{-1} I_{d_1} \le \mathbf{Q}^{-1}(x_1, \mathbf{X}_{\text{-}1}) \le c_{\mathbf{Q}}^{-1} I_{d_1}$, so $\sup_{x_1 \in [h,1-h]} \left|\tilde{m}_{\text{K},1,\cdot}(x_1) - m_{1l}(x_1)\right| = O_p\left\{(nh)^{-1/2}\log n\right\}$. □

4.7.3 Estimation of constants

To closely examine terms $\tilde{\varepsilon}_l(\mathbf{x})$ and $\tilde{\varepsilon}_{\alpha l}(x_\alpha)$, we denote the following vector of coefficients

$$\tilde{\mathbf{a}} = \left\{\tilde{a}_{01}, \tilde{a}_{1,1,1}, \dots, \tilde{a}_{N+1,d_2,1}, \tilde{a}_{02}, \tilde{a}_{1,1,2}, \dots, \tilde{a}_{N+1,d_2,2}, \dots, \tilde{a}_{0d_1}, \tilde{a}_{1,1,d_1}, \dots, \tilde{a}_{N+1,d_2,d_1}\right\}^T \tag{4.7.9}$$

such that the noise term $\tilde{\varepsilon}_l(\mathbf{x})$ in (4.4.9) is expressed as

$$(\mathbf{P}_{n,l}\mathbf{E})(\mathbf{x}) = \tilde{\varepsilon}_l(\mathbf{x}) = \tilde{a}_{0l} + \sum_{\alpha=1}^{d_2}\sum_{J=1}^{N+1} \tilde{a}_{J,\alpha,l} B_{J,\alpha}(x_\alpha). \tag{4.7.10}$$

Equation (4.7.10) implies that $\tilde{\mathbf{a}} = \left(\mathbf{D}^T\mathbf{D}\right)^{-1}\mathbf{D}^T\mathbf{E}$, where

$$\mathbf{D} = \left\{\mathbf{D}(\mathbf{X}_1, \mathbf{T}_1), \dots, \mathbf{D}(\mathbf{X}_n, \mathbf{T}_n)\right\}^T = \left\{\mathbf{T}_1 \otimes \mathbf{B}(\mathbf{X}_1), \dots, \mathbf{T}_n \otimes \mathbf{B}(\mathbf{X}_n)\right\}^T, \tag{4.7.11}$$

$$\mathbf{B}(\mathbf{x}) = \left\{1, B_{1,1}(x_1), \dots, B_{N+1,d_2}\left(x_{d_2}\right)\right\}^T, \mathbf{t} = \left\{t_1, \dots, t_{d_1}\right\}^T. \tag{4.7.12}$$

Note that $\tilde{\mathbf{a}}$ given in (4.7.9) can be rewritten as

$$\tilde{\mathbf{a}} = \left(\frac{1}{n}\mathbf{D}^T\mathbf{D}\right)^{-1}\left(\frac{1}{n}\mathbf{D}^T\mathbf{E}\right) = \left(\mathbf{V_T} + \mathbf{V_T^*}\right)^{-1}\left(\frac{1}{n}\mathbf{D}^T\mathbf{E}\right), \tag{4.7.13}$$

where by (4.7.11)

$$\mathbf{D}^T\mathbf{D} = \sum_{i=1}^{n}\left[\left(\mathbf{T}_i\mathbf{T}_i^T\right) \otimes \left\{\mathbf{B}(\mathbf{X}_i)\mathbf{B}(\mathbf{X}_i)^T\right\}\right], \mathbf{D}^T\mathbf{E} = \sum_{i=1}^{n}\left[\left\{\mathbf{T}_i \otimes \mathbf{B}(\mathbf{X}_i)\right\}\sigma(\mathbf{X}_i, \mathbf{T}_i)\varepsilon_i\right], \tag{4.7.14}$$

and $\mathbf{V_T^*}$ is the difference between empirical and theoretical inner product matrices, i.e.

$$\mathbf{V_T} = E\left[\left(\mathbf{T}\mathbf{T}^T\right) \otimes \left\{\mathbf{B}(\mathbf{X})\mathbf{B}(\mathbf{X})^T\right\}\right] = E\left[\mathbf{Q}(\mathbf{X}) \otimes \left\{\mathbf{B}(\mathbf{X})\mathbf{B}(\mathbf{X})^T\right\}\right], \tag{4.7.15}$$

$$\mathbf{V_T^*} = \frac{1}{n}\sum_{i=1}^{n}\left[\left(\mathbf{T}_i\mathbf{T}_i^T\right) \otimes \left\{\mathbf{B}(\mathbf{X}_i)\mathbf{B}(\mathbf{X}_i)^T\right\}\right] - E\left[\mathbf{Q}(\mathbf{X}) \otimes \left\{\mathbf{B}(\mathbf{X})\mathbf{B}(\mathbf{X})^T\right\}\right].$$

Now define $\hat{\mathbf{a}} = \left\{\hat{a}_{01}, \hat{a}_{1,1,1}, ..., \hat{a}_{N,d_2,1}, \hat{a}_{02}, \hat{a}_{1,1,2}, ..., \hat{a}_{N,d_2,2}, ..., \hat{a}_{0d_1}, \hat{a}_{1,1,d_1}, ..., \hat{a}_{N,d_2,d_1}\right\}^T$ by replacing $\left(\mathbf{V_T}+\mathbf{V_T^*}\right)^{-1}$ with $\mathbf{V_T^{-1}} = \mathbf{S_T}$ in the above formula, that is

$$\hat{\mathbf{a}} = \mathbf{V_T^{-1}}\left(n^{-1}\mathbf{D}^T\mathbf{E}\right) = \mathbf{S_T}\left(n^{-1}\mathbf{D}^T\mathbf{E}\right). \tag{4.7.16}$$

LEMMA 4.7.8. *Under Assumptions (A1) to (A3), (A5) and (A8), as* $n \to \infty$

$$\|\tilde{\mathbf{a}}\| = O_p\left(n^{-1/2}N^{1/2}\log n\right), \tag{4.7.17}$$

$$\|\tilde{\mathbf{a}} - \hat{\mathbf{a}}\| = O_p\left(n^{-1}N^{3/2}\log^2 n\right), \tag{4.7.18}$$

$$\|\hat{\mathbf{a}}\| = O_p\left(n^{-1/2}N^{1/2}\log n\right). \tag{4.7.19}$$

Proof. By definition, $\tilde{\mathbf{a}}^T\mathbf{D}^T\mathbf{D}\tilde{\mathbf{a}} = \tilde{\mathbf{a}}^T\mathbf{D}^T\mathbf{D}\left(\mathbf{D}^T\mathbf{D}\right)^{-1}\mathbf{D}^T\mathbf{E} = \tilde{\mathbf{a}}\mathbf{D}^T\mathbf{E}$. Using (4.7.13), one has

$$\|\mathbf{D}\tilde{\mathbf{a}}\|_{2,n}^2 = n^{-1}\tilde{\mathbf{a}}^T\mathbf{D}^T\mathbf{D}\tilde{\mathbf{a}} = n^{-1}\tilde{\mathbf{a}}^T\mathbf{D}^{\mathbf{T}}\mathbf{E} \leq \|\tilde{\mathbf{a}}\|\left\|n^{-1}\mathbf{D}^T\mathbf{E}\right\|. \tag{4.7.20}$$

According to Lemma 4.7.1,

$$c_0\|\tilde{\mathbf{a}}\|^2 = c_0\sum_l\left(a_{0l}^2 + \sum_{J,\alpha,l}a_{J,\alpha,l}^2\right) \leq \left\|\sum_l\left(a_{0l} + \sum_{J,\alpha,l}a_{J,\alpha,l}B_{J,\alpha}\right)t_l\right\|_{2,n}^2 = \|\mathbf{D}\tilde{\mathbf{a}}\|_{2,n}^2. \tag{4.7.21}$$

So $\|\tilde{\mathbf{a}}\|$ is bounded by $c_0^{-1}\left\|n^{-1}\mathbf{D}^T\mathbf{E}\right\|$. Bernstein's inequality and truncation entail that $\left\|n^{-1}\mathbf{D}^T\mathbf{E}\right\|^2 = O_p\left\{(\log n)^2 N/n\right\}$, so (4.7.17) follows from (4.7.20) and (4.7.21). According to (4.7.13) and (4.7.16), one has $\mathbf{V_T}\ \hat{\mathbf{a}} = \left(\mathbf{V_T}+\mathbf{V_T^*}\right)\tilde{\mathbf{a}}$, which implies that $\mathbf{V_T^*}\tilde{\mathbf{a}} = \mathbf{V_T}\left(\hat{\mathbf{a}} - \tilde{\mathbf{a}}\right)$.

One obtains from (4.7.29) and (4.7.30) $\|\mathbf{V_T}(\hat{\mathbf{a}} - \tilde{\mathbf{a}})\| = \left\|\mathbf{V_T^*}\tilde{\mathbf{a}}\right\| \leq O_p\left(n^{-1/2}H^{-1}\log n\right)\|\tilde{\mathbf{a}}\|$. By (4.7.17) one has $\|\mathbf{V_T}(\hat{\mathbf{a}} - \tilde{\mathbf{a}})\| \leq O_p\left\{(\log n)^2 n^{-1}N^{3/2}\right\}$. Thus according to Lemma 4.7.3, one has $\|\hat{\mathbf{a}} - \tilde{\mathbf{a}}\| = O_p\left(n^{-1}N^{3/2}\log^2 n\right)$, which is (4.7.18). Then (4.7.19) follows (4.7.17) and (4.7.18). □

LEMMA 4.7.9. *Under Assumptions (A1) to (A3), (A5) and (A8), , as* $n \to \infty$

$$\sup_{1\leq l'\leq d}\left|\frac{1}{n}\sum_{i=1}^n T_{il'}\sum_{l=1}^{d_1}\sum_{\alpha=1}^{d_2}\tilde{\varepsilon}_{\alpha l}T_{il}\right| = O_p\left(n^{-1/2}\right). \tag{4.7.22}$$

Proof. According to (4.4.9) and (4.7.10), one has

$$\frac{1}{n}\sum_{i=1}^{n}T_{il'}\sum_{l=1}^{d_1}\sum_{\alpha=1}^{d_2}\tilde{\varepsilon}_{\alpha l}T_{il}=\frac{1}{n}\sum_{i=1}^{n}T_{il'}\sum_{l=1}^{d_1}\sum_{\alpha=1}^{d_2}\sum_{J=1}^{N+1}\tilde{a}_{J,\alpha,l}B_{J,\alpha}\left(X_{i\alpha}\right)T_{il}$$

$$=\sum_{J,\alpha,l}\tilde{a}_{J,\alpha,l}\frac{1}{n}\sum_{i=1}^{n}T_{il'}B_{J,\alpha}\left(X_{i\alpha}\right)T_{il}=I_{l'}+II_{l'}$$

where

$$I_{l'}=\sum_{J,\alpha,l}\hat{a}_{J,\alpha,l}\frac{1}{n}\sum_{i=1}^{n}T_{il'}B_{J,\alpha}\left(X_{i\alpha}\right)T_{il},$$

$$II_{l'}=\sum_{J,\alpha,l}\left(\tilde{a}_{J,\alpha,\alpha}-\hat{a}_{J,\alpha,l}\right)\frac{1}{n}\sum_{i=1}^{n}T_{il'}B_{J,\alpha}\left(X_{i\alpha}\right)T_{il}.$$

Let $I_{l'}=I_{l',1}+I_{l',2}$ where

$$I_{l',1}=\sum_{J,\alpha,l}\hat{a}_{J,\alpha,l}\left\{\frac{1}{n}\sum_{i=1}^{n}T_{il'}B_{J,\alpha}\left(X_{i\alpha}\right)T_{il}-ET_{l'}B_{J,\alpha}\left(X_{\alpha}\right)T_{l}\right\},$$

$$\left|I_{l',1}\right|\le\|\hat{\mathbf{a}}\|\sqrt{(N+1)\,d_1d_2}\sup_{J,\alpha,l}\left|\frac{1}{n}\sum_{i=1}^{n}T_{il'}B_{J,\alpha}\left(X_{i\alpha}\right)T_{il}-ET_{l'}B_{J,\alpha}\left(X_{\alpha}\right)T_{l}\right|$$

$$=O_p\left(n^{-1}N\log^2 n\right),$$

$$I_{l',2}=\sum_{J,\alpha,l}\hat{a}_{J,\alpha,l}ET_{il'}B_{J,\alpha}\left(X_{i\alpha}\right)T_{il}=\left(ET_{l'}B_{J,\alpha}\left(X_{\alpha}\right)T_{l}\right)_{J,\alpha,l}^{T}\mathbf{V}_{\mathbf{T}}^{-1}\left(n^{-1}\mathbf{D}^{T}\mathbf{E}\right).$$

Let $\mathbf{v}_{l'}=\left(ET_{l'}B_{J,\alpha}\left(X_{\alpha}\right)T_{l}\right)_{J,\alpha,l}^{T}\mathbf{V}_{\mathbf{T}}^{-1}=\left(v_{J,\alpha,l}\right)^{T}$. According to (4.7.14)

$$I_{l',2}=n^{-1}\sum_{i=1}^{n}\sum_{J,\alpha,l}v_{J,\alpha,l}B_{J,\alpha}\left(X_{i\alpha}\right)T_{il}\sigma\left(\mathbf{X}_i,\mathbf{T}_i\right)\varepsilon_i.$$

Since ε_i is martingale difference according to Assumption (A5)

$$\mathrm{var}\left(I_{l',2}\right)=n^{-2}\sum_{i=1}^{n}\mathrm{var}\left\{\sum_{J,\alpha,l}v_{J,\alpha,l}B_{J,\alpha}\left(X_{i\alpha}\right)T_{il}\sigma\left(\mathbf{X}_i,\mathbf{T}_i\right)\varepsilon_i\right\}$$

$$\le n^{-2}\sum_{i=1}^{n}C_{\sigma}E\left\{\sum_{J,\alpha,l}v_{J,\alpha,l}B_{J,\alpha}\left(X_{i\alpha}\right)T_{il}\right\}^2$$

where

$$E\left\{\sum_{J,\alpha,l}v_{J,\alpha,l}B_{J,\alpha}\left(X_{i\alpha}\right)T_{il}\right\}^2=\sum_{J,\alpha,l}\sum_{J',\alpha',l'}v_{J,\alpha,l}v_{J,\alpha,l}E\left\{B_{J,\alpha}\left(X_{\alpha}\right)T_{l}B_{J',\alpha'}\left(X_{\alpha'}\right)T_{l'}\right\}$$

$$= \mathbf{v}_{l'}\mathbf{V}_{\mathbf{T}}\mathbf{v}_{l'}^{T} = \left\{ET_{l'}B_{J,\alpha}(X_\alpha)T_l\right\}_{J,\alpha,l}^{T}\mathbf{V}_{\mathbf{T}}^{-1}\mathbf{V}_{\mathbf{T}}\mathbf{V}_{\mathbf{T}}^{-1}\left\{ET_{l'}B_{J,\alpha}(X_\alpha)T_l\right\}_{J,\alpha,l}$$

$$= \left\{ET_{l'}B_{J,\alpha}(X_\alpha)T_l\right\}_{J,\alpha,l}^{T}\mathbf{V}_{\mathbf{T}}^{-1}\left\{ET_{l'}B_{J,\alpha}(X_\alpha)T_l\right\}_{J,\alpha,l}$$

$$\leq C_{\mathbf{V}}\left\|\left\{ET_{l'}B_{J,\alpha}(X_\alpha)T_l\right\}_{J,\alpha,l}\right\|_2^2 = O(1)$$

because clearly $\left|ET_{l'}B_{J,\alpha}(X_\alpha)T_l\right| = U\left(H^{1/2}\right)$. So $\operatorname{var}\left(I_{l',2}\right) = O\left(n^{-1}\right)$, and therefore $I_{l',2} = O_p\left(n^{-1/2}\right)$. So

$$\left|I_{l'}\right| \leq \left|I_{l',1}\right| + \left|I_{l',2}\right| = O_p\left(n^{-1/2}\right). \tag{4.7.23}$$

Next, by applying Bernstein's inequality with truncation technique,

$$\sup_{J,\alpha,l}\left|\frac{1}{n}\sum_{i=1}^{n}T_{il'}B_{J,\alpha}(X_{i\alpha})T_{il} - ET_{l'}B_{J,\alpha}(X_\alpha)T_l\right| = O_p\left(n^{-1/2}\log n\right).$$

Thus $\sup_{J,\alpha,l}\left|\frac{1}{n}\sum_{i=1}^{n}T_{il'}B_{J,\alpha}(X_{i\alpha})T_{il}\right|$ is bounded by

$$\sup_{J,\alpha,l}\left|\frac{1}{n}\sum_{i=1}^{n}T_{il'}B_{J,\alpha}(X_{i\alpha})T_{il} - ET_{l'}B_{J,\alpha}(X_i)T_l\right| + \left|ET_{l'}B_{J,\alpha}(X_\alpha)T_l\right| = O\left(H^{1/2}\right).$$

Then

$$\left|II_{l'}\right| \leq \|\hat{\mathbf{a}} - \tilde{\mathbf{a}}\|\sqrt{(N+1)d_1d_2}\sup\left|\frac{1}{n}\sum_{i=1}^{n}T_{il'}B_{J,\alpha}(X_{i\alpha})T_{il}\right|$$

$$= O_p\left(n^{-1}N^{3/2}\log^2 n\right)\sqrt{(N+1)d_1d_2}\,O_p\left(H^{1/2}\right) = O_p\left(n^{-1}N^{3/2}\log^2 n\right). \tag{4.7.24}$$

Now (4.7.22) follows from (4.7.23) and (4.7.24). The lemma is proved. □

LEMMA 4.7.10. *Under Assumptions (A1) to (A5), and (A8), , as $n \to \infty$*

$$n^{-1}\sum_{i=1}^{n}\left[\sum_{l=1}^{d_1}\sum_{\alpha=1}^{d_2}\left\{\tilde{m}_{\alpha l}(X_{i\alpha}) - m_{\alpha l}(X_{i\alpha})\right\}T_{il}\right]^2 = O_p\left(n^{-1}\right) \tag{4.7.25}$$

Proof. According to (4.7.1), there exists $g_{\alpha l} \in G^{(0)}[0,1]$ such that $\|g_{\alpha l} - m_{\alpha l}\|_\infty = O\left(H^2\right) = O\left(n^{-1/2}\right)$.According to Theorem 1.7 of Bosq (1998) p. 36, $n^{-1/2}\sum_{i=1}^{n}\left(T_{il}^2 - ET_l^2\right) \Longrightarrow N\left(0,\sigma^2\right)$ where $\sigma^2 = \sum_{i=-\infty}^{\infty}\operatorname{Cov}\left(T_{0l}^2, T_{il}^2\right) < \infty$ by applying Davydov's Inequality [Bosq 1998, p. 21, equation (1.10)], then $n^{-1}\sum_{i=1}^{n}T_{il}^2 = ET_l^2 + O_p\left(n^{-1/2}\right) = O_p(1)$. So

$$n^{-1}\sum_{i=1}^{n}\left[\sum_{l=1}^{d_1}\sum_{\alpha=1}^{d_2}\left\{\tilde{m}_{\alpha l}(X_{i\alpha}) - m_{\alpha l}(X_{i\alpha})\right\}T_{il}\right]^2 \leq n^{-1}\sum_{i=1}^{n}\left[\sum_{l=1}^{d_1}\sum_{\alpha=1}^{d_2}\|\tilde{m}_{\alpha l} - m_{\alpha l}\|_\infty T_{il}\right]^2$$

$$\leq n^{-1}\sum_{i=1}^{n}\left[\sum_{l=1}^{d_1}\sum_{\alpha=1}^{d_2}\|g_{\alpha l}-m_{\alpha l}\|_\infty T_{il}\right]^2 = O\left(n^{-1}\right)\left(n^{-1}\sum_{i=1}^{n}T_{il}^2\right) = O_p\left(n^{-1}\right).\square$$

Proof of Propositions 4.3.1 and 4.4.1. According to (4.3.4), $\tilde{m}_0 - m_0 = (\mathbf{C}_K^T\mathbf{C}_K)^{-1}\mathbf{C}_K^T(\mathbf{Y}_c - m_0\mathbf{T})$
$= (\frac{1}{n}\mathbf{C}_K^T\mathbf{C}_K)^{-1}\frac{1}{n}\mathbf{C}_K^T\sigma(\mathbf{X}_i,\mathbf{T}_i)\varepsilon_i$. We know $\frac{1}{n}\mathbf{C}_K^T\mathbf{C}_K = \left(\frac{1}{n}\sum_{i=1}^n T_{il}T_{il'}\right)_{l,l'=1}^{d_1}$. Then according to Theorem 1.7 of Bosq (1998) p. 36, one has $n^{-1/2}\sum_{i=1}^n\{T_{il}T_{il'} - ET_lT_{l'}\} \Longrightarrow N(0,\sigma^2)$ where $\sigma^2 = \sum_{i=-\infty}^{\infty}\mathrm{Cov}(T_{0l}T_{0l'}, T_{il}T_{il'}) < \infty$. Therefore $\frac{1}{n}\mathbf{C}_K^T\mathbf{C}_K = (ET_lT_{l'})_{l,l'=1}^{d_1} + O_p\left(n^{-1/2}\right)$. Similarly, $\frac{1}{n}\mathbf{C}_K^T\sigma(\mathbf{X}_i,\mathbf{T}_i)\varepsilon_i = O_p\left(n^{-1/2}\right)$, implying $\sup_{1\leq l\leq d_1}|\tilde{m}_{0l} - m_{0l}| = O_p\left(n^{-1/2}\right)$, which has completed the proof of Proposition 4.3.1.

Next, According to (4.3.4) and (4.4.2),

$$\hat{m}_0 - \tilde{m}_0 = (\mathbf{C}_K^T\mathbf{C}_K)^{-1}\mathbf{C}_K^T\left(\hat{\mathbf{Y}}_c - \mathbf{Y}_c\right)$$

$$= (\mathbf{C}_K^T\mathbf{C}_K)^{-1}\mathbf{C}_K^T\left[\sum_{\alpha=1}^{d_1}\sum_{\alpha=1}^{d_2}\{\hat{m}_{\alpha l}(X_{i\alpha}) - m_{\alpha l}(X_{i\alpha})\}T_{il}\right]_{i=1}^n$$

$$= (\mathbf{C}_K^T\mathbf{C}_K)^{-1}\mathbf{C}_K^T\left[\sum_{l=1}^{d_1}\sum_{\alpha=1}^{d_2}\{\hat{m}_{\alpha l}(X_{i\alpha}) - \tilde{m}_{\alpha l}(X_{i\alpha}) + \tilde{m}_{\alpha l}(X_{i\alpha}) - m_{\alpha l}(X_{i\alpha})\}T_{il}\right]_{i=1}^n$$

$$= (\frac{1}{n}\mathbf{C}_K^T\mathbf{C}_K)^{-1}\frac{1}{n}\mathbf{C}_K^T\left[\left(\sum_{l=1}^{d_1}\sum_{\alpha=1}^{d_2}\tilde{\varepsilon}_{\alpha l}T_{il}\right)_{i=1}^n\right.$$
$$\left.+\left[\left\{\sum_{l=1}^{d_1}\sum_{\alpha=1}^{d_2}\{\tilde{m}_{\alpha l}(X_{i\alpha}) - m_{\alpha l}(X_{i\alpha})\}T_{il}\right\}\right]_{i=1}^n\right].$$

One has

$$\frac{1}{n}\sum_{i=1}^n T_{il'}\left[\sum_{l=1}^{d_1}\sum_{\alpha=1}^{d_2}(\tilde{m}_{\alpha l}(X_{i\alpha}) - m_{\alpha l}(X_{i\alpha}))T_{il}\right]$$

$$\leq \left(\frac{1}{n}\sum_{i=1}^n T^2\right)_{il'}^{1/2}\left\{\frac{1}{n}\sum_{i=1}^n\left[\sum_{l=1}^{d_1}\sum_{\alpha=1}^{d_2}(\tilde{m}_{\alpha l}(X_{i\alpha}) - m_{\alpha l}(X_{i\alpha}))T_{il}\right]^2\right\}^{1/2}$$

$$\leq O_p(1)\,O_p\left(n^{-1/2}\right) = O_p\left(n^{-1/2}\right). \tag{4.7.26}$$

by Lemma 4.7.10. Then the Proposition 4.4.1 follows (4.7.22) and (4.7.26). $\square$

4.7.4 Estimation of function components

Define

$$
\begin{aligned}
A_{n,1} &= \sup_{J,\alpha} \left| \langle 1, B_{J,\alpha} \rangle_{2,n} - \langle 1, B_{J,\alpha} \rangle_2 \right| = \sup_{J,\alpha} \left| n^{-1} \sum_{i=1}^{n} B_{J,\alpha} (X_{i\alpha}) \right|, \\
A_{n,2} &= \sup_{J,J',\alpha,l,l'} \left| \left\langle B_{J,\alpha} T_{il}, B_{J',\alpha} T_{il'} \right\rangle_{2,n} - \left\langle B_{J,\alpha} T_{il}, B_{J',\alpha} T_{il'} \right\rangle_2 \right|, \\
A_{n,3} &= \sup_{J,J',\alpha \neq \alpha',l,l'} \left| \left\langle B_{J,\alpha} T_{il}, B_{J',\alpha'} T_{il'} \right\rangle_{2,n} - \left\langle B_{J,\alpha} T_{il}, B_{J',\alpha'} T_{il'} \right\rangle_2 \right|. \quad (4.7.27)
\end{aligned}
$$

LEMMA 4.7.11. *Under Assumptions (A1) to (A3), and (A8), , as $n \to \infty$*

$$
\begin{aligned}
A_{n,1} &= O_p \left(n^{-1/2} \log n \right), \quad (4.7.28) \\
A_{n,2} &= O_p \left(n^{-1/2} H^{-1/2} \log n \right), \quad (4.7.29) \\
A_{n,3} &= O_p \left(n^{-1/2} \log n \right). \quad (4.7.30)
\end{aligned}
$$

Proof. The proof of (4.7.28) follows from Bernstein's inequality immediately, thus is omitted. Here we only prove (4.7.29) and (4.7.30). We will discuss case by case with various $l, l', \alpha, \alpha', J$ and J', via Bernstein's inequality. For brevity, we set

$$
\begin{aligned}
\xi_i &= \xi_{i,n,J,J',\alpha,\alpha',l,l'} = \xi_{i,1} + \xi_{i,2} = \xi_{i,1,n,J,J',\alpha,\alpha',l,l'} + \xi_{i,2,n,J,J',\alpha,\alpha',l,l'} \\
&= B_{J,\alpha} (X_{i\alpha}) B_{J',\alpha'} (X_{i\alpha'}) T_{il} T_{il'} - E B_{J,\alpha} (X_{i\alpha}) B_{J',\alpha'} (X_{i\alpha'}) T_{il} T_{il'}
\end{aligned}
$$

where $\xi_{i,j} = B_{J,\alpha} (X_{i\alpha}) B_{J',\alpha'} (X_{i\alpha'}) \mathrm{T}_{ill',j}^{D_n} - E B_{J,\alpha} (X_{i\alpha}) B_{J',\alpha'} (X_{i\alpha'}) \mathrm{T}_{ill',j}^{D_n}$, $j = 1, 2$ by the same truncation (4.7.5) in Lemma 4.7.4. Then $A_{n,2} = \sup_{J,J',\alpha,l,l'} n^{-1} \left| \sum_{i=1}^{n} \xi_{i,n,J,J',\alpha,\alpha,l,l'} \right|$ and $A_{n,3} = \sup_{J,J',\alpha \neq \alpha',l,l'} n^{-1} \left| \sum_{i=1}^{n} \xi_{i,n,J,J',\alpha,\alpha',l,l'} \right|$. One has with probability 1,

$$
\sup_{J,J',\alpha,\alpha',l,l'} \left| E \xi_{i,1}^2 \right| = U \left(n^{-1} \right), \quad (4.7.31)
$$

$$
\sup_{J,J',\alpha,\alpha,l,l'} \left| n^{-1} \sum_{i=1}^{n} \xi_{i,1} \right| = U \left(n^{-k} \right), k > 0. \quad (4.7.32)
$$

We will consider $\alpha = \alpha' = 1$ in the Case 1.1 to Case 1.4.

Case 1.1 when $|J - J'| > 2$. The definition of $B_{J,1}$ in (4.4.6) will guarantee that $B_{J,1} (X_{i1}) B_{J',1} (X_{i1}) = 0$ if $|J - J'| > 2$.

Case 1.2 when $J = J'$. According to Lemma 4.7.2,

$$EB_{J,\alpha}^{2}(X_{i\alpha})T_{il}T_{il'} = E\left\{B_{J,\alpha}^{2}(X_{i\alpha})E\left(T_{il}T_{il'}|X_{i\alpha}\right)\right\} = O(1)$$

$$E\left\{B_{J,\alpha}^{2}(X_{i\alpha})T_{il}T_{il'}\right\}^{2} = E\left\{B_{J,\alpha}^{4}(X_{i\alpha})E\left(T_{il}^{2}T_{il'}^{2}|X_{i\alpha}\right)\right\} \sim H^{-1}.$$

So $E\xi_i^2 = \left[E\left\{B_{J,\alpha}^{2}(X_{i\alpha})T_{il}T_{il'}\right\}^{2} - \left\{EB_{J,\alpha}^{2}(X_{i\alpha})T_{il}T_{il'}\right\}^{2}\right] \sim H^{-1}$. According to (4.7.31), one has $E\xi_{i,2}^2 = E\xi_i^2 - E\xi_{i,1}^2 \sim H^{-1}$. Lemma 4.7.2 provides a constant $C_\xi > 0$ such that

$$\begin{aligned} E\left|\xi_{i,2}\right|^{k} &= E\left|B_{J,\alpha}^{2}(X_{i\alpha})\mathrm{T}_{ill',2}^{D_n} - EB_{J,\alpha}^{2}(X_{i\alpha})\mathrm{T}_{ill',2}^{D_n}\right|^{k} \\ &\leq \sup_{J,\alpha,l,l'}\left|B_{J,\alpha}^{2}(X_{i\alpha})\mathrm{T}_{ill',2}^{D_n} - EB_{J,\alpha}^{2}(X_{i\alpha})\mathrm{T}_{ill',2}^{D_n}\right|^{k-2} E\xi_{i,2}^{2} \\ &\leq C_\xi^{k-2}H^{2-k}D_n^{k-2}E\xi_{i,2}^{2} \leq (C_\xi D_n/H)^{k-2}E\xi_{i,2}^{2}. \end{aligned}$$

Using the same technique in Lemma 4.7.4 by applying Lemma 2.5.2 and Borel -Cantelli Lemma, when $J = J', \alpha = \alpha' = 1$, one has

$$\sup_{J,\alpha,l,l'}\left|n^{-1}\sum_{i=1}^{n}\xi_{i,2}\right| = O_p\left(n^{-1/2}H^{-1/2}\log n\right)$$

and then we can get (4.7.29) combining with (4.7.32).

Case 1.3 when $\left|J - J'\right| = 1$. Without loss of generality we assume that $J' = J + 1$.

$$EB_{J,\alpha}(X_{i\alpha})B_{J+1,\alpha}(X_{i\alpha})T_{il}T_{il'} = E\left\{B_{J,\alpha}(X_{i\alpha})B_{J+1,\alpha}(X_{i\alpha})E\left(T_{il}T_{il'}|X_{i\alpha}\right)\right\} = O(1)$$

$$E\left\{B_{J,\alpha}(X_{i\alpha})B_{J+1,\alpha}(X_{i\alpha})T_{il}T_{il'}\right\}^{2} = E\left\{\left|B_{J,\alpha}(X_{i\alpha})B_{J+1,\alpha}(X_{i\alpha})\right|^{2}E\left(T_{il}^{2}T_{il'}^{2}|X_{i\alpha}\right)\right\}$$

$\sim H^{-1}$, i.e., $E\xi_i^2 \sim H^{-1}$. Similar to Case 1.2, (4.7.29) follows by using Bernstein's inequality.

Case 1.4 when $\left|J - J'\right| = 2$, all the above discussion in case 1.3 applies with replacing $J' = J + 1$ with $J' = J + 2$.

Case 2 when $\alpha = \alpha' > 1$, all the above discussion applies without modifications.

Case 3 when $\alpha \neq \alpha'$. Without loss of generality, suppose $\alpha = 1, \alpha' = 2$. First, we still need to calculate the order of second moment $E\xi_{i,2}^2$, which is $E\xi_i^2 - E\xi_{i,1}^2$. The boundedness of the density function $f(x_1, x_2)$ implies that

$$\left|EB_{J,1}(X_{i1})B_{J',2}(X_{i2})\right| \leq \left\|b_{J,1}\right\|_2^{-1}\left\|b_{J,2}\right\|_2^{-1}\int\int\left|b_{J,1}(x_1)b_{J',2}(x_2)\right|f(x_1,x_2)\,dx_1dx_2$$

$$\le \left\|b_{J,1}\right\|_2^{-1}\left\|b_{J,2}\right\|_2^{-1}\times C_f H^2 \le C_{B,1}H,$$

for some constant $C_{B,1} > 0$, where the last step is derived by Lemma 4.7.2. According to Assumption (A1) and Lemma 4.7.2,

$$E\left\{B_{J,1}(X_{i1})\,B_{J',2}(X_{i2})\right\}^2 = \left\|b_{J,1}\right\|_2^{-2}\left\|b_{J,2}\right\|_2^{-2}\int\int b_{J,1}^2(x_{i1})\,b_{J',2}^2(x_{i2})\,f(x_1,x_2)\,dx_1dx_2$$

$$\ge c_f\left\{\left\|b_{J,1}\right\|_2^{-2}\int b_{J,1}^2(x_{i1})\,dx_1\right\}\left\{\left\|b_{J,2}\right\|_2^{-2}\int b_{J',2}^2(x_{i2})\,dx_2\right\}$$

$$= c_f\frac{1}{3}\left\{2+2c_{J,1}^2/c_{J-1,1}^2 - c_{J,1}/c_{J-1,1}\right\}\left\{\left\|b_{J,1}\right\|_2^{-2}H\right\}\times.$$

$$\frac{1}{3}\left\{2+2c_{J',2}^2/c_{J'-1,2}^2 - c_{J',2}/c_{J'-1,2}\right\}\left\{\left\|b_{J,2}\right\|_2^{-2}H\right\} \ge c_{B,2}.$$

$$E\xi_i^2 = E\left\{B_{J,1}(X_{i1})\,B_{J',2}(X_{i2})\,T_{il}T_{il'}\right\}^2 - \left\{EB_{J,1}(X_{i1})\,B_{J',2}(X_{i2})\,T_{il}T_{il'}\right\}^2$$

$$= E\left\{B_{J,1}^2(X_{i1})\,B_{J',2}^2(X_{i2})\,E\left(T_{il}^2T_{il'}^2|X_{i1},X_{i2}\right)\right\}$$

$$-\left[E\left\{B_{J,1}(X_{i1})\,B_{J',2}(X_{i2})\,E\left(T_{il}T_{il'}|X_{i1},X_{i2}\right)\right\}\right]^2.$$

According to Assumption (A2), there exist constants c_ξ, such that $c_\xi \le E\xi_i^2$. Similarly, we can get $C'_\xi > 0$ such that $E\xi_i^2 \le C'_\xi$, i.e., $E\xi_i^2 \sim 1$, then $E\xi_{i,2}^2 \sim 1$ by (4.7.31). Second, the k-th moment of $|\xi_{i,2}|$ is given by

$$E\left|\xi_{i,2}\right|^k = E\left|B_{J,1}(X_{i1})\,B_{J',2}(X_{i2})\,\mathrm{T}_{ill',2}^{D_n} - E\left\{B_{J,1}(X_{i1})\,B_{J',2}(X_{i2})\,\mathrm{T}_{ill',2}^{D_n}\right\}\right|^k$$

and there is a constant C_ξ such that $E\left|\xi_{i,2}\right|^k$ is bounded by

$$\sup_{J,J',l,l'}\left|B_{J,1}(X_{i1})\,B_{J',2}(X_{i2})\,\mathrm{T}_{ill',2}^{D_n} - EB_{J,1}(X_{i1})\,B_{J',2}(X_{i2})\,\mathrm{T}_{ill',2}^{D_n}\right|^{k-2}E\xi_{i,2}^2$$

$$\le\; C_\xi^{k-2}H^{2-k}D_n^{k-2}E\xi_{i,2}^2 \le (C_\xi D_n/H)^{k-2}E\xi_{i,2}^2.$$

Similar to the proof of (4.7.29), the proof of (4.7.30) is completed. □

LEMMA 4.7.12. *Under Assumptions (A1) to (A3), (A5), and (A7) to (A8), , as* $n\to\infty$

$$\sup_{x_1\in[0,1]}\sup_{1\le J\le N+1}\sup_{2\le\alpha\le d_2}\sup_{1\le l,l'\le d_1}\left|\mu_{\omega_{J,\alpha,l,l'}}(x_1)\right| = O\left(H^{1/2}\right), \tag{4.7.33}$$

$$\sup_{x_1\in[0,1]}\sup_{1\le J\le N+1}\sup_{2\le\alpha\le d_2}\sup_{1\le l,l'\le d_1}\left|n^{-1}\sum_{i=1}^n\left\{\omega_{J,\alpha,l,l'}(\mathrm{X}_i,x_1)-\mu_{\omega_{J,\alpha,l,l'}}(x_1)\right\}\right| \tag{4.7.34}$$

$$= O_p\left(\log n/\sqrt{nh}\right), \tag{4.7.35}$$

where $\omega_{J,\alpha,l,l'}(\mathbf{X}_i, x_1)$ *and* $\mu_{\omega_{J,\alpha,l,l'}}(x_1)$ *defined in (4.4.18), hence*

$$\sup_{x_1\in[0,1]}\sup_{1\le J\le N+1}\sup_{2\le\alpha\le d_2}\sup_{1\le l,l'\le d_1}\left|n^{-1}\sum_{i=1}^{n}\omega_{J,\alpha,l,l'}(\mathbf{X}_i, x_1)\right| = O_p\left(H^{1/2}\right). \tag{4.7.36}$$

Proof. According to the definitions of $\omega_{J,\alpha,l,l'}(\mathbf{X}_i, x_1)$ and $\mu_{\omega_{J,\alpha,l,l'}}(x_1)$ in (4.4.18),

$$\left|\mu_{\omega_{J,\alpha,l,l'}}(x_1)\right| \le E\left|\omega_{J,\alpha,l,l'}(\mathbf{X}_i, x_1)\right|$$

$$= \|b_{J,\alpha}\|_2^{-1}\int\left|t_l t_{l'}\frac{1}{h}K\left(\frac{u_1-x_1}{h}\right)b_{J,\alpha}(u_\alpha)\right| f\left(u_1,u_\alpha,t_l,t_{l'}\right)du_1du_\alpha dt_l dt_{l'}$$

$$\le \|b_{J,\alpha}\|_2^{-1}\left\{\int\left|t_l t_{l'}K(v_1)b_J(u_\alpha)\right| f\left(x_1+hv_1,u_\alpha,t_l,t_{l'}\right)dv_1du_\alpha dt_l dt_{l'}\right.$$
$$\left.+\frac{c_{J,\alpha}}{c_{J-1,\alpha}}\int\left|t_l t_{l'}K(v_1)b_{J-1}(u_\alpha)\right| f\left(x_1+hv_1,u_\alpha,t_l,t_{l'}\right)dv_1du_\alpha dt_l dt_{l'}\right\}.$$

The boundedness of the joint density f and the Lipschitz continuity of the kernel K will imply that there exist constant c_2 such that

$$\int\left|t_l t_{l'}K(v_1)b_J(u_\alpha)\right| f\left(x_1+hv_1,u_\alpha,t_l,t_{l'}\right)dv_1du_\alpha dt_l dt_{l'} \le C_{\mathbf{Q}}C_K c_2 H,$$

$$\int\left|t_l t_{l'}K(v_1)b_{J-1}(u_\alpha)\right| f\left(x_1+hv_1,u_\alpha,t_l,t_{l'}\right)dv_1du_\alpha dt_l dt_{l'} \le C_{\mathbf{Q}}C_K c_2 H,$$

and therefore $E\left|\omega_{J,\alpha,l,l'}(\mathbf{X}_i, x_1)\right| = O\left(H^{1/2}\right)$. Meanwhile

$$E\left|\omega_{J,\alpha,l,l'}(\mathbf{X}_i, x_1)\right|^r = E\left|T_{il}T_{il'}K_h(X_{i1}-x_1)B_{J,\alpha}(X_{i\alpha})\right|^r$$

$$= \|b_{J,\alpha}\|_2^{-r}\int\left|\left(t_l t_{l'}\right)^r\frac{1}{h^r}K^r\left(\frac{u_1-x_1}{h}\right)b_{J,\alpha}^r(u_\alpha)\right| f\left(u_1,u_\alpha,t_l,t_{l'}\right)du_1du_\alpha dt_l dt_{l'}$$

$$= \|b_{J,\alpha}\|_2^{-r}h^{1-r}\left[\int\left|\left(t_l t_{l'}\right)^r K^r(v_1)\left\{\sum_{\alpha=0}^{r}\binom{r}{\alpha}c_{J,\alpha}^{r-\alpha}c_{J-1,\alpha}^{\alpha-r}b_J^\alpha(u_\alpha)b_{J-1}^{r-\alpha}(u_\alpha)\right\}\right|\right.$$
$$\left. f\left(x_1+hv_1,u_\alpha,t_l,t_{l'}\right)dv_1du_\alpha dt_l dt_{l'}\right].$$

The boundedness of the joint density f and the Lipschitz continuity of the kernel K will imply that there exist constant c_2 such that

$$\int\left|\left(t_l t_{l'}\right)^r K^r(v_1)b_J^\alpha(u_\alpha)b_{J-1}^{r-\alpha}(u_\alpha)\right| f\left(x_1+hv_1,u_\alpha,t_l,t_{l'}\right)dv_1du_\alpha dt_l dt_{l'} \le C_K c_2 H,$$

which implies that $E\left|\omega_{J,\alpha,l,l'}(\mathbf{X}_i,x_1)\right|^r \sim h^{1-r}H^{1-r/2}$, hence $E\omega^2_{J,\alpha,l,l'}(\mathbf{X}_i,x_1) \sim h^{-1}$. Define $\omega_{J,\alpha,l,l'}(\mathbf{X}_i,x_1) = \omega_{J,\alpha,l,l',1}(\mathbf{X}_i,x_1) + \omega_{J,\alpha,l,l',2}(\mathbf{X}_i,x_1)$ where

$$\omega_{J,\alpha,l,l',j}(\mathbf{X}_i,x_1) = K_h(X_{i1}-x_1)\,B_{J,\alpha}(X_{i\alpha})\,\mathrm{T}^{D_n}_{ill',j}, j=1,2$$

by the same truncation (4.7.5) in Lemma 4.7.4. One has with probability 1, $\sup_{x_1\in[0,1]}\left|E\left\{\omega_{J,\alpha,l,l',1}(\mathbf{X}_i,x_1)\right\}^2\right| = U(n^{-1})$ and $\sup_{x_1\in[0,1]}\left|n^{-1}\sum_{i=1}^n \omega_{J,\alpha,l,l',1}(\mathbf{X}_i,x_1)\right| = U\left(n^{-k}\right)$ for $k>0$. Define

$$\omega^*_{J,\alpha,l,l'}(\mathbf{X}_i,x_1) = \omega_{J,\alpha,l,l'}(\mathbf{X}_i,x_1) - E\omega_{J,\alpha,l,l'}(\mathbf{X}_i,x_1),$$

$$\omega^*_{J,\alpha,l,l',j}(\mathbf{X}_i,x_1) = \omega_{J,\alpha,l,l',j}(\mathbf{X}_i,x_1) - E\omega_{J,\alpha,l,l',j}(\mathbf{X}_i,x_1).$$

Then $E\left\{\omega^*_{J,\alpha,l,l',2}(\mathbf{X}_i,x_1)\right\}^2 = E\left\{\omega^*_{J,\alpha,l,l'}(\mathbf{X}_i,x_1)\right\}^2 - E\left\{\omega^*_{J,\alpha,l,l',1}(\mathbf{X}_i,x_1)\right\}^2 \sim h^{-1}$, and

$E\left|\omega^*_{J,\alpha,l,l',2}(\mathbf{X}_i,x_1)\right|^k$ is bounded by

$$\begin{aligned}&\sup_{1\le J\le N+1}\left|\omega_{J,\alpha,l,l',2}(\mathbf{X}_i,x_1) - E\omega_{J,\alpha,l,l',2}(\mathbf{X}_i,x_1)\right|^{k-2} E\left|\omega^*_{J,\alpha,l,l',2}(\mathbf{X}_i,x_1)\right|^2\\ \le\;& 2^{r-1}D_n h^{2-k}H^{1-k/2}E\left|\omega^*_{J,\alpha,l,l',2}(\mathbf{X}_i,x_1)\right|^2.\end{aligned}$$

Thud there exists a constant $c_* = cD_nh^{2-k}H^{1-k/2}$ such that $E\left|\omega^*_{J,\alpha,l,l'}(\mathbf{X}_i,x_1)\right|^r \le c_*^{r-2}r!$ $E\left|\omega^*_{J,\alpha,l,l'}(\mathbf{X}_i,x_1)\right|^2$. That means the sequence of random variables $\left\{\omega^*_{J,\alpha,l,l',2}(\mathbf{X}_l,x_1)\right\}_{l=1}^n$ satisfies the Cramér's condition, hence by the Bernstein's inequality and similar proof in the case 3 of Lemma 4.7.11, one has (4.7.34). □

In the following, we define a noise term analogous to the formula for $\Psi^{(2)}_{v,l'}(x_1)$ in (4.4.19) by replacing $\tilde{\mathbf{a}}$ in (4.7.9) with $\hat{\mathbf{a}}$ in (4.7.16)

$$\hat{\Psi}^{(2)}_{v,l'}(x_1) = n^{-1}\sum_{i=1}^n\sum_{l=1}^{d_1}\sum_{J=1}^{N+1}\sum_{\alpha=2}^{d_2}\hat{a}_{J,\alpha,l}\omega_{J,\alpha,l,l'}(\mathbf{X}_i,x_1). \tag{4.7.37}$$

LEMMA 4.7.13. *Under Assumptions (A1) to (A3), (A5) and (A8), , as $n\to\infty$,* $\sup_{1\le l'\le d_1}\sup_{x_1\in[0,1]}\left|\Psi^{(2)}_{v,l'}(x_1) - \hat{\Psi}^{(2)}_{v,l'}(x_1)\right| = O_p\left(H^2\right)$.

Proof. According to (4.4.17) and (4.7.37), one has

$$\left|\Psi^{(2)}_{v,l'}(x_1) - \hat{\Psi}^{(2)}_{v,l'}(x_1)\right| = \left|\sum_{l=1}^{d_1}\sum_{J=1}^{N+1}\sum_{\alpha=2}^{d_2}\left(\tilde{a}_{J,\alpha,l} - \hat{a}_{J,\alpha,l}\right)\frac{1}{n}\sum_{i=1}^n\omega_{J,\alpha,l}(\mathbf{X}_i,x_1)T_{i1}\right|.$$

According to (4.7.36) and (4.7.18), Cauchy-Schwartz inequality implies that

$$\sup_{1\le l'\le d_1}\sup_{x\in[0,1]}\left|\Psi^{(2)}_{v,l'}(x_1)-\hat{\Psi}^{(2)}_{v,l'}(x_1)\right| \le \sqrt{N+1}O_p\left(\frac{(\log n)^2}{nH^{3/2}}\right)O_p\left(H^{1/2}\right)$$
$$= O_p\left(\frac{(\log n)^2}{nH^{3/2}}\right).$$

Therefore the lemma follows. □

LEMMA 4.7.14. *Under Assumptions (A1) to (A3), (A5) and (A8), , as $n\to\infty$*

$$\sup_{1\le l'\le d_1}\sup_{x_1\in[0,1]}\left|\hat{\Psi}^{(2)}_{v,l'}(x_1)\right| = \sup_{1\le l'\le d_1}\sup_{x_1\in[0,1]}\left|n^{-1}\sum_{i=1}^{n}\sum_{l=1}^{d_1}\sum_{J=1}^{N+1}\sum_{\alpha=2}^{d_2}\hat{a}_{J,\alpha,l}\omega_{J,\alpha,l,l'}(\mathbf{X}_i,x_1)\right|$$
$$= O_p\left(n^{-2/5}\right).$$

Proof. Note that by definition (4.7.37)

$$\left|\hat{\Psi}^{(2)}_{v,l'}(x_1)\right| \le \left|\sum_{l=1}^{d_1}\sum_{J=1}^{N+1}\sum_{\alpha=2}^{d_2}\hat{a}_{J,\alpha,l}\mu_{\omega_{J,\alpha,l,l'}}(x_1)\right| +$$

$$\left|\sum_{l=1}^{d_1}\sum_{J=1}^{N+1}\sum_{\alpha=2}^{d_2}\hat{a}_{J,\alpha,l}n^{-1}\sum_{i=1}^{n}\left\{\omega_{J,\alpha,l,l'}(\mathbf{X}_i,x_1)-\mu_{\omega_{J,\alpha,l,l'}}(x_1)\right\}\right| = R_1(x_1)+R_2(x_1). \tag{4.7.38}$$

By Cauchy-Schwartz inequality, $R_2(x_1)$ is bounded by $\left(\sum_{l=1}^{d_1}\sum_{J=1}^{N+1}\sum_{\alpha=2}^{d_2}\hat{a}^2_{J,\alpha,l}\right)^{1/2}\times$

$$\left(\sum_{l=1}^{d_1}\sum_{J=1}^{N+1}\sum_{\alpha=2}^{d_2}\left\{\sup_{x_1\in[0,1]}\left|n^{-1}\sum_{i=1}^{n}\left\{\omega_{J,\alpha,l,l'}(\mathbf{X}_i,x_1)-\mu_{\omega_{J,\alpha,l,l'}}(x_1)\right\}\right|\right\}^2\right)^{1/2}.$$

Observe that $\|\hat{\mathbf{a}}\| = O_p\left(\log n\sqrt{N/n}\right)$ as given in (4.7.19) and

$$\sup_{x_1\in[0,1]}\left|n^{-1}\sum_{i=1}^{n}\left\{\omega_{J,\alpha,l,l'}(\mathbf{X}_i,x_1)-\mu_{\omega_{J,\alpha,l,l'}}(x_1)\right\}\right| = O_p\left(\log n/\sqrt{nh}\right),$$

which is given in (4.7.34), so the order of $\sup_{x_1\in[0,1]}R_2(x_1)$ by Assumptions (A7), (A8) is

$$O_p\left(\log n\sqrt{N/n}\right)\sqrt{(N+1)d_1(d_2-1)}O_p\left(\frac{\log n}{\sqrt{nh}}\right) = O_p\left(\frac{N(\log n)^2}{n\sqrt{h}}\right) \tag{4.7.39}$$
$$= O_p\left((\log n)^3/\sqrt{n}\right). \tag{4.7.40}$$

Using again the discretization idea, we divide the interval $[0,1]$ into $M_n \sim n$ equally spaced intervals with endpoints $0 = x_{1,0} < x_{1,1} < ... < x_{1,M_n} = 1$. Then $sup_{x_1 \in [0,1]} R_1(x_1)$

$$\leq \max_{1 \leq k \leq M_n} \left| \sum_{l=1}^{d_1} \sum_{J=1}^{N+1} \sum_{\alpha=2}^{d_2} \hat{a}_{J,\alpha,l} \mu_{\omega_{J,\alpha,l,l'}}(x_{1,k}) \right| + \max_{1 \leq k \leq M_n} \sup_{x_1 \in [x_{1,k-1}, x_{1,k}]}$$
$$\left| \sum_{l=1}^{d_1} \sum_{J=1}^{N+1} \sum_{\alpha=2}^{d_2} \hat{a}_{J,\alpha,l} \mu_{\omega_{J,\alpha,l,l'}}(x_1) - \sum_{l=1}^{d_1} \sum_{J=1}^{N+1} \sum_{\alpha=2}^{d_2} \hat{a}_{J,\alpha,l} \mu_{\omega_{J,\alpha,l,l'}}(x_{1,k}) \right|$$
$$= T_1 + T_2.$$

Noting that $\hat{a}_{J,\alpha,l}$ is

$$\sum_{l''=1}^{d_1} \sum_{J'=1}^{N+1} \sum_{\alpha'=1}^{d_2} S_{J+(\alpha-1)(N+1)+(l-1)d_2(N+1), J'+(\alpha'-1)(N+1)+(l''-1)d_2(N+1)} n^{-1} \times$$
$$\sum_{i=1}^{n} B_{J',\alpha'}(X_{i\alpha}) T_{il''} \sigma(\mathbf{X}_i, \mathbf{T}_i) \varepsilon_i,$$

according to (4.7.16), where $S_{J+(\alpha-1)(N+1)+(l-1)d_2(N+1), J'+(\alpha'-1)(N+1)+(l''-1)d_2(N+1)}$ is the corresponding element in $\mathbf{S_T} = \mathbf{V_T^{-1}}$. We define $W_{\alpha,l,\alpha',l''}$ equals

$$\max_{1 \leq k \leq M_n} \left| n^{-1} \sum_{i=1}^{n} \sum_{J=1}^{N+1} \sum_{J'=1}^{N+1} S_{J+(\alpha-1)(N+1)+(l-1)d_2(N+1), J'+(\alpha'-1)(N+1)+(l''-1)d_2(N+1)} \right.$$
$$\left. B_{J',\alpha'}(X_{i\alpha}) T_{il''} \sigma(\mathbf{X}_i, \mathbf{T}_i) \varepsilon_i \mu_{\omega_{J,\alpha,l,l'}}(x_{1,k}) \right|$$

then it is clear that $T_1 \leq \sum_{l=1}^{d_1} \sum_{\alpha=2}^{d_2} \sum_{l''=1}^{d_1} \sum_{\alpha'=1}^{d_2} W_{\alpha,l,\alpha',l''}$. To show that each term $W_{\alpha,l,\alpha',l''}$ has order $O_p\left(n^{-2/5}\right)$, we truncate the $T_{il''}\varepsilon_i$ by the same way in the proof of Theorem 4.3.2,

$$D_n = n^{\theta_0} \left(\frac{1}{2+\eta} < \theta_0 < \frac{2}{5} \right). \tag{4.7.41}$$

where η is the same as in Assumption (A5). Let $Z_i = T_{il''}\varepsilon_i = Z_{i,1}^{D_n} + Z_{i,2}^{D_n} + Z_{i,3}^{D_n}$, where

$$Z_{i,1}^{D_n} = Z_i \{|Z_i| > D_n\}, Z_{i,2}^{D_n} = Z_i \{|Z_i| \leq D_n\} - Z_{i,3}^{D_n}, Z_{i,3}^{D_n} = EZ_i \{|Z_i| \leq D_n\}$$

For fixed J, α, l, l',

$$U_{i,k} = \sum_{1 \leq J, J' \leq N} \mu_{\omega_{J,\alpha,l,l'}}(x_{1,k}) B_{J',\alpha'}(X_{i\alpha}) \sigma(\mathbf{X}_i, \mathbf{T}_i) Z_{i,2}^{D_n}$$
$$S_{J+(\alpha-1)(N+1)+(l-1)d_2(N+1), J'+(\alpha'-1)(N+1)+(l''-1)d_2(N+1)},$$

and denote $W^D_{\alpha,l,\alpha',l''}$ as the truncated centered version of $W_{\alpha,l,\alpha',l''}$, i.e.,

$$W^D_{\alpha,l,\alpha',l''} = \max_{1\le k\le M_n} \left| n^{-1} \sum_{i=1}^n U_{i,k} \right|. \qquad (4.7.42)$$

In the following, we will prove that $\left|W_{\alpha,l,\alpha',l''} - W^D_{\alpha,l,\alpha',l''}\right| = O_p(H)$.

It is clear that $\left|W_{\alpha,l,\alpha',l''} - W^D_{\alpha,l,\alpha',l''}\right| \le \Lambda_1 + \Lambda$, where

$$\Lambda_1 = \max_{1\le k\le M_n} \left| n^{-1} \sum_{i=1}^n \sum_{J=1}^{N+1} \sum_{J'=1}^{N+1} \mu_{\omega_{J,\alpha,l,l'}}(x_{1,k}) B_{J',\alpha'}(X_{i\alpha}) \sigma(\mathbf{X}_i, \mathbf{T}_i) Z^{D_n}_{i,1} \right.$$

$$\left. S_{J+(\alpha-1)(N+1)+(l-1)d_2(N+1),J'+(\alpha'-1)(N+1)+(l''-1)d_2(N+1)} \right|,$$

$$\Lambda_2 = \max_{1\le k\le M_n} \left| n^{-1} \sum_{i=1}^n \sum_{J=1}^{N+1} \sum_{J'=1}^{N+1} \mu_{\omega_{J,\alpha,l,l'}}(x_{1,k}) B_{J',\alpha'}(X_{i\alpha}) \sigma(\mathbf{X}_i, \mathbf{T}_i) Z^{D_n}_{i,3} \right.$$

$$\left. S_{J+(\alpha-1)(N+1)+(l-1)d_2(N+1),J'+(\alpha'-1)(N+1)+(l''-1)d_2(N+1)} \right|.$$

Let $\boldsymbol{\mu}_{\omega_{\alpha,l,l'}}(x_{1,k}) = \left\{\mu_{\omega_{1,\alpha,l,l'}}(x_{1,k}), \cdots, \mu_{\omega_{N,\alpha,l,l'}}(x_{1,k})\right\}^T$, then $\Lambda_2 \le C_{\mathbf{Q}} C_{\mathbf{S}} \times$

$$\max_{1\le k\le M_n} \left\{ \sum_{J=1}^{N+1} \mu^2_{\omega_{1,\alpha,l,l'}}(x_{1,k}) \sum_{J=1}^{N+1} \left\{ n^{-1} \sum_{i=1}^n B_{J',\alpha'}(X_{i\alpha}) T_{il''} \sigma(\mathbf{X}_i, \mathbf{T}_i) Z^{D_n}_{i,3} \right\}^2 \right\}^{1/2},$$

according to Lemma 4.7.3. By the proof of Theorem 4.3.2, $\left|Z^{D_n}_{i,3}\right| = O\left(D_n^{-(1+\eta)}\right)$ and

$$\sup_{J,\alpha} \left| n^{-1} \sum_{i=1}^n B_{J',\alpha'}(X_{i\alpha}) T_{il''} \sigma(\mathbf{X}_i, \mathbf{T}_i) \right| = O_p\left(\log n/\sqrt{n}\right)$$

by Bernstein's inequality given in Lemma 2.5.2. Therefore $\Lambda_2 \le C D_n^{-(1+\eta)} \times$

$$\max_{1\le k\le M_n} \left\{ \sum_{J=1}^{N+1} \mu^2_{\omega_{1,\alpha,l,l'}}(x_{1,k}) \sum_{J=1}^{N+1} \left\{ n^{-1} \sum_{i=1}^n B_{J',\alpha'}(X_{i\alpha}) \sigma(\mathbf{X}_i, \mathbf{T}_i) \right\}^2 \right\}^{1/2}$$

$$= O_p\left\{ D_n^{-(1+\eta)} \left(NHN \log^2 n/n\right)^{1/2} \right\} = O_p\left(N^{1/2} n^{-9/10} \log n\right)$$

where the last step follows from the choice of D_n in (4.7.41). One has with probability 1,

$$n^{-1} \sum_{i=1}^n \sum_{J=1}^{N+1} \sum_{J'=1}^{N+1} \mu_{\omega_{J,\alpha,l,l'}}(x_{1,k}) B_{J',\alpha'}(X_{i\alpha}) \sigma(\mathbf{X}_i, \mathbf{T}_i) Z^{D_n}_{i,3} \times$$

$$S_{J+(\alpha-1)(N+1)+(l-1)d_2(N+1),J'+(\alpha'-1)(N+1)+(l''-1)d_2(N+1)} = 0$$

for large n. Therefore, one has $\left|W_{\alpha,l,\alpha',l''} - W^D_{\alpha,l,\alpha',l''}\right| \le \Lambda_1 + \Lambda_2 = O_p\left(n^{-2/5}\right)$. Next we want to show that $W^D_{\alpha,l,\alpha',l''} = O_p\left(n^{-2/5}\right)$, with $W^D_{\alpha,l,\alpha',l''}$ defined in (4.7.42). Since

$$U_{i,k} = \mu_{\omega_{\alpha,l,l'}}\left(x_{1,k}\right)^T \mathbf{S}_{\mathbf{T},\alpha l\alpha' l''} \left\{B_{J',\alpha'}\left(X_{i\alpha}\right)\sigma\left(\mathbf{X}_i,\mathbf{T}_i\right) Z^{D_n}_{i,2}\right\}_{J'=1}^{N+1}$$

so the variance of $U_{i,k}$ is $\mu_{\omega_{\alpha,l,l'}}\left(x_{1,k}\right)^T \mathbf{S}_{\mathbf{T},\alpha l\alpha' l''}\times$

$$\operatorname{var}\left(\left\{B_{J',\alpha'}\left(X_{i\alpha}\right)\sigma\left(\mathbf{X}_i,\mathbf{T}_i\right) Z^{D_n}_{i,2}\right\}_{J'=1}^{N}\right) \mathbf{S}_{\mathbf{T},\alpha l\alpha' l''}\mu_{\omega_{\alpha,l,l'}}\left(x_{1,k}\right)^T.$$

According to Assumption (A5), $\sigma^2\left(\mathbf{x},\mathbf{t}\right)$ is measurable and bounded, so it is easy to see that

$$c^2_\sigma \mathbf{V}_{\mathbf{T},\alpha l\alpha' l''} \le \operatorname{var}\left(\left\{B_{J',\alpha'}\left(X_{i\alpha}\right)\right\}_{J'=1}^{N+1}\sigma\left(\mathbf{X}_i\right)\right) \le C^2_\sigma \mathbf{V}_{\mathbf{T},\alpha l\alpha' l''}.$$

Thus, one has

$$\operatorname{var}\left(U_{i,k}\right) \sim \mu_{\omega_{\alpha,l,l'}}\left(x_{1,k}\right)^T \mathbf{S}_{\mathbf{T},\alpha l\alpha' l''}\mathbf{V}_{\mathbf{T},\alpha l\alpha' l''}\mathbf{S}_{\mathbf{T},\alpha l\alpha' l''}\mu_{\omega_{\alpha,l,l'}}\left(x_{1,k}\right) V_{Z,D}$$

where $V_{Z,D} = \operatorname{var}\left(Z^{D_n}_{i,2}\right)$. Let $\kappa\left(x_{1,k}\right) = \left\{\mu_{\omega_{\alpha,l,l'}}\left(x_{1,k}\right)^T \mu_{\omega_{\alpha,l,l'}}\left(x_{1,k}\right)\right\}^{1/2}$

$$c_\alpha c_\mathbf{V} c_\alpha c^2_\sigma \left\{\kappa\left(x_{1,k}\right)\right\}^2 V_{Z,D} \le \operatorname{var}\left(U_{i,k}\right) \le C_\alpha C_\mathbf{V} C_\alpha C^2_\sigma \left\{\kappa\left(x_{1,k}\right)\right\}^2 V_{Z,D}.$$

When $r \ge 3$, the r-th absolute moment $E\left|U_{i,k}\right|^r$ is

$$E\left\{\left|\sum_{1\le J,J'\le N+1} \mu_{\omega_{J,\alpha,l,l'}}\left(x_{1,k}\right) B_{J',\alpha'}\left(X_{i\alpha}\right)\sigma\left(\mathbf{X}_i,\mathbf{T}_i\right)\right.\right.$$
$$\left.\left. S_{J+(\alpha-1)(N+1)+(l-1)d_2(N+1),J'+(\alpha'-1)(N+1)+(l''-1)d_2(N+1)}\right|^r E\left(\left|Z^{D_n}_{i,2}\right|^r \left|\mathbf{X}_i,\mathbf{T}_i\right.\right)\right\}$$
$$\le C^r_\alpha C^r_\sigma \left\{\kappa\left(x_{1,k}\right)\right\}^r O\left(H^{1-r/2}\right) D^{r-2}_n V_{Z,D} \le \left\{c_0\kappa\left(x_{1,k}\right) D_n H^{-1/2}\right\}^{r-2} r! E\left|U_{i,k}\right|^2,$$

which means the sequence $\left\{U_{i,k}\right\}_{i=1}^n$ satisfies the Cramér's condition with Cramér's constant equal to $c_* = c_0\kappa\left(x_{1,k}\right) D_n H^{-1/2}$, applying Bernstein's inequality for $r = 3$

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

where

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

$$m_2^2 \sim \{\kappa(x_{1,k})\}^2 V_{Z,D},\ m_3 \leq \left\{c\{\kappa(x_{1,k})\}^3 H^{-1/2} D_n V_{Z,D}\right\}^{1/3}.$$

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

$$\alpha\left(\left[\frac{n}{q+1}\right]\right)^{6/7} \leq \left\{K_0 \exp\left(-\lambda_0\left[\frac{n}{q+1}\right]\right)\right\}^{6/7} \leq Cn^{-6\lambda_0 c_0/7},$$

and as $n \to \infty$, one has

$$\frac{q\rho_n^2}{25m_2^2 + 5c_*\rho_n} \geqslant \frac{c_1\rho^2 n^{-1/5} H^{-1} \log^2 n}{25m_2^2 + 5D_n H^{-1/2}\rho n^{-3/5} H^{-1/2}\log n} \sim \frac{\rho^2 n^{-1/5} H^{-1}\log n}{D_n n^{-2/5} n^{-1/5} H^{-1}} \sim \rho \log n.$$

Thus, for n large enough,

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

Taking c_0, ρ large enough, one has for large n, $P\left\{\left|\frac{1}{n}\sum_{i=1}^{n} U_{i,k}\right| > \rho n^{-2/5}\right\} \leq n^{-3}$. Hence

$$\sum_{n=1}^{\infty} P\left(\left|W^D_{\alpha,l,\alpha',l''}\right| \geq \rho H\right) = \sum_{n=1}^{\infty}\sum_{k=1}^{M_n} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} U_{i,k}\right| \geq \rho H\right) \leq \sum_{n=1}^{\infty} M_n n^{-3} < \infty.$$

Thus, Borel-Cantelli Lemma entails that $W^D_{\alpha,l,\alpha',l''} = O_p\left(n^{-2/5}\right)$. Therefore, one has $W_{\alpha,l,\alpha',l''} = O_p\left(n^{-2/5}\right)$ since $\left|W_{\alpha,l,\alpha',l''} - W^D_{\alpha,l,\alpha',l''}\right| = O_p\left(n^{-2/5}\right)$. Hence

$$T_1 \leq \sum_{l=1}^{d_1}\sum_{\alpha=2}^{d_2}\sum_{l''=1}^{d_1}\sum_{\alpha'=1}^{d_2} W_{\alpha,l,\alpha',l''} = O_p\left(n^{-2/5}\right). \tag{4.7.43}$$

Employing Lipschitz continuity of kernel K, the term T_2 equals $\max_{1\leq k\leq M_n} \sup_{x_1\in[x_{1,k-1},x_{1,k}]}$

$$\left|\sum_{l=1}^{d_1}\sum_{J=1}^{N+1}\sum_{\alpha=2}^{d_2} \hat{a}_{J,\alpha,l}\mu_{\omega_{J,\alpha,l,l'}}(x_1) - \sum_{l=1}^{d_1}\sum_{J=1}^{N+1}\sum_{\alpha=2}^{d_2} \hat{a}_{J,\alpha,l}\mu_{\omega_{J,\alpha,l,l'}}(x_{1,k})\right|$$

is bounded by $\|\hat{\mathrm{a}}\| \times$

$$\max_{1\leq k\leq M_n}\sup_{x_1\in[x_{1,k-1},x_{1,k}]}\sum_{J=1}^{N+1} E\left[\{K_h(X_1 - x_1) - K_h(X_1 - x_{1,k})\}^2\{T_{il}T_{il'}B_{J,\alpha}(X_\alpha)\}^2\right].$$

Therefore, according to Assumption (A8), Lemma 4.7.2 (ii), and (4.7.19),

$$\begin{aligned} T_2 &\leq C_{\mathbf{Q}} O_p\left(n^{-1/2}N^{1/2}\log n\right)\frac{1}{h^4 M_n^2}\sum_{J=1}^{N+1} EB^2_{J,\alpha}(X_\alpha)T_{il}T^2_{il'} \qquad (4.7.44)\\ &= O\left(\log n\sqrt{Nn^{-1}}h^{-4}M_n^{-2}\right) = o_p\left(n^{-1/2}\right). \end{aligned}$$

Combining (4.7.43) and (4.7.44), one has $\sup_{x_1 \in [0,1]} R_1(x_1) = O_p\left(n^{-2/5}\right)$. The desired result follows from (4.7.38) and (4.7.40). □

Proof of Proposition 4.4.2. (4.7.1) implies that

$$\begin{aligned} \left|E_n g_{\text{-}1,l}(\mathbf{X}_{i,\text{-}1})\right| &\leq \left|E_n g_{\text{-}1,l}(\mathbf{X}_{i,\text{-}1}) - E_n m_{\text{-}1,l}(\mathbf{X}_{i,\text{-}1})\right| + \left|E_n m_{\text{-}1,l}(\mathbf{X}_{i,\text{-}1})\right| \quad (4.7.45) \\ &\leq C_\infty (d_2 - 1) \sup_{2 \leq \alpha \leq d_2} \left\|m'_{\alpha l}\right\|_\infty H^2 + O_p\left(n^{-1/2}\right). \end{aligned}$$

By definition (4.4.14), $\sup_{x_1 \in [0,1]} \left|\Psi_{b,l'}(x_1)\right| \leq R_1 + R_2 + R_3$ where

$$\begin{aligned} R_1 &= \sup_{x_1 \in [0,1]} \left|\frac{1}{n} \sum_{i=1}^n K_h(X_{i1} - x_1) \sum_{l=1}^{d_1} \left\{m_{\text{-}1,l}(\mathbf{X}_{i,\text{-}1}) - g_{\text{-}1,l}(\mathbf{X}_{i,\text{-}1})\right\} T_{il} T_{il'}\right|, \\ R_2 &= \sup_{x_1 \in [0,1]} \left|\frac{1}{n} \sum_{i=1}^n K_h(X_{i1} - x_1) \right. \\ &\quad \left. \sum_{l=1}^{d_1} \left\{g_{\text{-}1,l}(\mathbf{X}_{i,\text{-}1}) - E_n g_{\text{-}1,l}(\mathbf{X}_{i,\text{-}1}) - \tilde{m}_{\text{-}1,l}(\mathbf{X}_{i,\text{-}1})\right\} T_{il} T_{il'}\right|, \\ R_3 &= \sup_{x_1 \in [0,1]} \left|\frac{1}{n} \sum_{i=1}^n K_h(X_{i1} - x_1) \sum_{l=1}^{d_1} E_n g_{\text{-}1,l}(\mathbf{X}_{i,\text{-}1}) T_{il} T_{il'}\right|. \end{aligned}$$

For R_1, using (4.7.1), one has

$$\begin{aligned} R_1 &\leq C_\infty (d_2 - 1) \sup_{2 \leq \alpha \leq d_2} \left\|m'_{\alpha l}\right\|_\infty H^2 \sum_{l=1}^{d_1} \frac{1}{n} \sum_{i=1}^n \left|T_{il} T_{il'}\right| \\ &= O_p\left(H^2\right) \left\{\sum_{l=1}^{d_1} E\left|T_{il} T_{il'}\right| + O_p\left(n^{-1/2}\right)\right\} = O_p\left(H^2\right). \qquad (4.7.46) \end{aligned}$$

To bound R_2, denote the empirically centered spline basis as $B^*_{J,\alpha}(X_{i\alpha}) = B_{J,\alpha}(X_{i\alpha}) - E_n B_{J,\alpha}(X_{i\alpha})$, $1 \leq J \leq N+1$, $1 \leq \alpha \leq d_2$. Then one can write for some $\left(\tilde{a}^*_{\alpha,l}, \tilde{a}^*_{J,\alpha,l}\right)_{J,\alpha,l}$,

$$\tilde{m}_l(\mathbf{x}) - m_{0l} - \sum_{\alpha=1}^{d_2} g_{\alpha l}(\mathbf{x}) + \sum_{\alpha=1}^{d_2} E_n g_{\alpha l}(\mathbf{x}) = \tilde{a}^*_{\alpha,l} + \sum_{\alpha=1}^{d_2} \sum_{J=1}^{N+1} \tilde{a}^*_{J,\alpha,l} B^*_{J,\alpha}(x_\alpha).$$

Thus

$$R_2 = \sup_{x_1 \in [0,1]} \left|n^{-1} \sum_{i=1}^n K_h(X_{i1} - x_1) \sum_{l=1}^{d_1} \sum_{\alpha=2}^{d_2} \sum_{J=1}^{N+1} \tilde{a}^*_{J,\alpha,l} B^*_{J,\alpha}(X_{i\alpha}) T_{il} T_{il'}\right|$$

$$\leq \sum_{l=1}^{d_1}\sum_{\alpha=2}^{d_2}\sum_{J=1}^{N+1}\left|\tilde{a}^*_{J,\alpha,l}\right| \sup_{1\leq J\leq N+1,1\leq l\leq d_1,2\leq\alpha\leq d_2}\left|n^{-1}\sum_{i=1}^{n}K_h\left(X_{i1}-x_1\right)T_{il}T_{il'}B^*_{J,\alpha}\left(X_{i\alpha}\right)\right|$$

$$=\sum_{l=1}^{d_1}\sum_{\alpha=2}^{d_2}\sum_{J=1}^{N+1}\left|\tilde{a}^*_{J,\alpha,l}\right|\times$$

$$\left[\sup_{\substack{1\leq J\leq N+1,\\ 1\leq l\leq d_1,2\leq\alpha\leq d_2}}\left|n^{-1}\sum_{i=1}^{n}K_h\left(X_{i1}-x_1\right)T_{il}T_{il'}\left\{B_{J,\alpha}\left(X_{i\alpha}\right)-E_nB_{J,\alpha}\left(X_{i\alpha}\right)\right\}\right|\right].$$

Equation (4.7.33) in Lemma 4.7.12 states that

$$\sup_{x_1\in[0,1]}\sup_{1\leq J\leq N+1,1\leq l,l'\leq d_1,2\leq\alpha\leq d_2}\left|n^{-1}\sum_{i=1}^{n}K_h\left(X_{i1}-x_1\right)T_{il}T_{il'}B_{J,\alpha}\left(X_{i\alpha}\right)\right|=O_p\left(H^{1/2}\right),$$

while equation (4.7.28) of Lemma 4.7.11 states that $\sup_{1\leq J\leq N+1,}\left|E_nB_{J,\alpha}\left(X_{i\alpha}\right)\right| = O_p\left(\log n/\sqrt{n}\right)$ and standard kernel argument shows that

$$\sup_{x_1\in[0,1]}\sup_{1\leq l,l'\leq d_1}\left|n^{-1}\sum_{i=1}^{n}K_h\left(X_{i1}-x_1\right)T_{il}T_{il'}\right|=O_p\left(1\right).$$

Therefore, one has

$$R_2\leq\left\{\left(N+1\right)d_1\left(d_2-1\right)\sum_{l=1}^{d_1}\sum_{\alpha=2}^{d_2}\sum_{J=1}^{N+1}\left(\tilde{a}^*_{J,\alpha,1}\right)^2\right\}^{1/2}\left\{O_p\left(H^{1/2}\right)+O_p\left(\frac{\log n}{\sqrt{n}}\right)\right\}$$

$$\begin{aligned}
&= O_p\left(\sum_{l=1}^{d_1}\left\|\tilde{m}_l\left(\mathbf{x}\right)-m_{0l}-\sum_{\alpha=1}^{d_2}g_{\alpha l}\left(\mathbf{x}\right)+\sum_{\alpha=1}^{d_2}E_ng_{\alpha l}\left(\mathbf{x}\right)\right\|_2\right)\\
&= O_p\left(n^{-1/2}+H^2\right). \qquad (4.7.47)
\end{aligned}$$

The last step follows from

$$\left\|\tilde{m}_l\left(\mathbf{x}\right)-m_{0l}-\sum_{\alpha=1}^{d_2}g_{\alpha l}\left(\mathbf{x}\right)+\sum_{\alpha=1}^{d_2}E_ng_{\alpha l}\left(\mathbf{x}\right)\right\|_2$$

$$\leq\left\|\tilde{m}_l\left(\mathbf{x}\right)-m_l\left(\mathbf{x}\right)\right\|_2+\left\|m_l\left(\mathbf{x}\right)-m_{0l}-\sum_{\alpha=1}^{d_2}g_{\alpha l}\left(\mathbf{x}\right)\right\|_2+\left\|\sum_{\alpha=1}^{d_2}E_ng_{\alpha l}\left(\mathbf{x}\right)\right\|_2$$

$$\leq 3C_\infty\sum_{\alpha=1}^{d_2}\left\|m'_{\alpha 1}\right\|_\infty H^2+O_p\left(n^{-1/2}\right).$$

Thus $R_2 = O_p\left(n^{-1/2} + H^2\right)$. Similarly,

$$R_3 = \sup_{x_1\in[0,1]}\left|\frac{1}{n}\sum_{i=1}^{n} K_h\left(X_{i1}-x_1\right)\sum_{l=1}^{d_1} E_n g_{\text{-}1,l}\left(\mathbf{X}_{i,\text{-}1}\right) T_{il}T_{il'}\right|$$

$$\leq \left\{\sum_{l=1}^{d_1}\left|E_n g_{\text{-}1,l}\left(\mathbf{X}_{i,\text{-}1}\right)\right|\right\} \sup_{x_1\in[0,1]}\left|\frac{1}{n}\sum_{i=1}^{n} K_h\left(X_{i1}-x_1\right) T_{il}T_{il'}\right|$$

$$\leq \left\{\sum_{l=1}^{d_1}\left|E_n g_{\text{-}1,l}\left(\mathbf{X}_{i,\text{-}1}\right)\right|\right\} \sup_{x_1\in[0,1]}\left|\frac{1}{n}\sum_{i=1}^{n} K_h\left(X_{i1}-x_1\right) T_{il}T_{il'}\right| = O_p\left(n^{-1/2}+H^2\right). \tag{4.7.48}$$

by (4.7.45). Combining (4.7.46), (4.7.47) and (4.7.48), one establishes Proposition 4.4.2. □

Proof of Lemma 4.4.1. Based on formula (4.4.11), $n^{-1}\sum_{i=1}^{n}\tilde{\varepsilon}^*_{\text{-}1,l}\left(\mathbf{X}_{i,\text{-}1}\right)$ is

$$n^{-1}\sum_{i=1}^{n}\sum_{\alpha=2}^{d_2}\sum_{J=1}^{N+1}\tilde{a}_{J,\alpha,l}B_{J,\alpha}\left(X_{i\alpha}\right) = \sum_{\alpha=2}^{d_2}\sum_{J=1}^{N+1}\tilde{a}_{J,\alpha,l}\left\{n^{-1}\sum_{i=1}^{n}B_{J,\alpha}\left(X_{i\alpha}\right)\right\}.$$

Lemma 4.7.8 implies that

$$\left|\sum_{\alpha=2}^{d_2}\sum_{J=1}^{N+1}\tilde{a}_{J,\alpha,l}\right| \leq \left\{(N+1)\left(d_2-1\right)\cdot\sum_{\alpha=2}^{d_2}\sum_{J=1}^{N+1}\tilde{a}^2_{J,\alpha,l}\right\}^{1/2}$$
$$\leq \left\{(N+1)\left(d_2-1\right)\cdot\tilde{\mathbf{a}}^T\tilde{\mathbf{a}}\right\}^{1/2} = O_p\left(Nn^{-1/2}\log n\right).$$

Now it is clear from (4.7.27) and (4.7.28) that $\sup_{1\leq J\leq N+1}\left|n^{-1}\sum_{i=1}^{n}B_{J,\alpha}\left(X_{i\alpha}\right)\right| \leq A_{n,1} = O_p\left(n^{-1/2}\log n\right)$, hence

$$\frac{1}{n}\sum_{i=1}^{n}\tilde{\varepsilon}^*_{\text{-}1,l}\left(\mathbf{X}_{i,\text{-}1}\right) \leq \left|\sum_{\alpha=2}^{d_2}\sum_{J=1}^{N+1}\tilde{a}_{J,\alpha,l}\right|\cdot\sup_{J,\alpha,l}\left|n^{-1}\sum_{i=1}^{n}B_{J,\alpha}\left(X_{i\alpha}\right)\right| = O_p\left(\frac{N\left(\log n\right)^2}{n}\right). \tag{4.7.49}$$

While standard kernel theory implies that $\sup_{x_1\in[0,1]}\left|n^{-1}\sum_{i=1}^{n}\sum_{l=1}^{d_1}K_h\left(X_{i1}-x_1\right)T_{il}T_{il'}\right| = O_p(1)$.Thus the lemma follows immediately from (4.7.49) and (4.4.16). □

CHAPTER 5

Spline-backfitted kernel smoothing of generalized additive model

5.1 Introduction

Following Stone (1985), p. 693, the space of α-centered square integrable functions on $[0,1]$ is

$$\mathcal{M}=\left\{g: E\left\{g\left(X_{\alpha}\right)\right\}=0, E\left\{g^{2}\left(X_{\alpha}\right)\right\}<+\infty\right\}, 1 \leq \alpha \leq d.$$

in which g are finite constants. The constraints that $E\left\{g_{\alpha}\left(X_{\alpha}\right)\right\}=0,\ 1 \leq \alpha \leq d$ ensure unique additive representation of m_{α} as expressed in (1.4.3), but are not necessary for the definition of space $\mathcal{M}$. In what follows, denote by E_{n} the empirical expectation, $E_{n} \varphi=\sum_{i=1}^{n} \varphi\left(\mathbf{X}_{i}\right) / n$. We introduce two inner products on $\mathcal{M}$. For functions $g_{1}, g_{2} \in \mathcal{M}$, the theoretical and empirical inner products are defined respectively as $\left\langle g_{1}, g_{2}\right\rangle=E\left\{g_{1}(\mathbf{X}) g_{2}(\mathbf{X})\right\}$, $\left\langle g_{1}, g_{2}\right\rangle_{n}=E_{n}\left\{g_{1}(\mathbf{X}) g_{2}(\mathbf{X})\right\}$. The corresponding induced norms are $\left\|g_{1}\right\|_{2}^{2}=E g_{1}^{2}(\mathbf{X})$, $\left\|g_{1}\right\|_{2, n}^{2}=E_{n} g_{1}^{2}(\mathbf{X})$. The model space $\mathcal{M}$ is called *theoretically (empirically) identifiable*, if for any $g \in \mathcal{M}$, $\|g\|_{2}=0$ ($\|g\|_{2, n}=0$) implies that $g=0$ a.s.

In this chapter, for any compact interval $[a, b]$, we denote the space of p-th order smooth function as $C^{(p)}[a, b]=\left\{g \mid g^{(p)} \in C[a, b]\right\}$, and the class of Lipschitz continuous functions for constant $C>0$ as $\operatorname{Lip}([a, b], C)=\left\{g \mid\left|g(x)-g\left(x^{\prime}\right)\right| \leq C\left|x-x^{\prime}\right|, \forall x, x^{\prime} \in[a, b]\right\}$. We mean by "$\sim$" both sides having the same order as $n \rightarrow \infty$. We denote by $\mathbf{I}_{d \times d}$ the $d \times d$ identity matrix, and $\mathbf{0}_{d \times d}$ the $d \times d$ zero matrix. For any vector $\mathbf{x}=\left(x_{1}, x_{2}, \cdots, x_{d}\right)$, we denote the supremum and Euclidean norms as $|\mathbf{x}|=\max _{1 \leq \alpha \leq d}\left|x_{\alpha}\right|$ and $\|\mathbf{x}\|=\left(\sum_{\alpha=1}^{d} x_{\alpha}^{2}\right)^{1 / 2}$.

We need the following Assumptions on the data generating process.

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

(A2) *The inverse link function* b' *satisfies the following:* $b' \in C^2(\Theta)$ *where* Θ *is a compact interval such that* $m\left([0,1]^d\right)$ *is in the interior of* Θ *and* $C_b > \max_{\theta\in\Theta} b''(\theta) \geq \min_{\theta\in\Theta} b''(\theta) > c_b$ *for some constants* $C_b > c_b > 0$. *There exists a compact interval* A *such that* $m_1([0,1]) \subset A$ *and that* $A + m_{\text{-}1}\left([0,1]^{d-1}\right) \subset \Theta$ *where* $m_{\text{-}1}(\mathbf{x}_{\text{-}1}) = c + \sum_{\alpha=2}^{d} m_\alpha(x_\alpha)$ *with* $\mathbf{x}_{\text{-}1} = (x_2, ..., x_d)$.

(A3) *The conditional variance function* $\sigma^2(\mathbf{x})$ *is measurable and bounded. The errors* $\{\varepsilon_i\}_{i=1}^n$ *satisfy* $E(\varepsilon_i|\mathcal{F}_i) = 0$, $E(\varepsilon_i^2|\mathcal{F}_i) = 1$, $E\left(|\varepsilon_i|^{2+\eta}|\mathcal{F}_i\right) \leq C_\eta$ *for some* $\eta \in (1/2, 1]$ *and the sequence of* σ*-fields*
$F_i = \sigma\left\{(\mathbf{X}_j), j \leq i; \varepsilon_j, j \leq i-1\right\}$ *for* $i = 1, \ldots, n$.

(A4) *The density function* $f(\mathbf{x})$ *of* $(X_1, ..., X_d)$ *is continuous and*

$$0 < c_f \leq \inf_{\mathbf{x}\in[0,1]^d} f(\mathbf{x}) \leq \sup_{\mathbf{x}\in[0,1]^d} f(\mathbf{x}) \leq C_f < \infty.$$

The marginal densities $f_\alpha(x_\alpha)$ *of* X_α *have continuous derivatives on* $[0,1]$ *as well as the uniform upper bound* C_f *and lower bound* c_f.

(A5) *There exist positive constants* K_0 *and* λ_0 *such that* $\alpha(n) \leq K_0 e^{-\lambda_0 n}$ *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*

$$\alpha(k) = \sup_{B\in\sigma\{\mathbf{Z}_s, s\leq t\}, C\in\sigma\{\mathbf{Z}_s, s\geq t+k\}} |P(B\cap C) - P(B)P(C)|, k \geq 1.$$

5.2 Oracle Smoothers

We need following Assumption for kernel function.

(A6) *The kernel function* $K \in \mathrm{Lip}([-1,1], C_K)$ *for some positive constant* $C_K > 0$, *and is bounded, nonnegative, symmetric, and supported on* $[-1,1]$. *The bandwidth* h *of the kernel* K *is assumed to be of order* $n^{-1/5}$, *i.e.,* $c_h n^{-1/5} \leq h \leq C_h n^{-1/5}$ *for some positive constants* c_h, C_h.

If the last $d-1$ components $\{m_\alpha(x_\alpha)\}_{\alpha=2}^d$ were known by "oracle", then the only unknown component $m_1(x_1)$ could be estimated by the following procedure. Define for each $x_1 \in [h, 1-h]$ an local quasi-likelihood function $\tilde{l}(a) = \tilde{l}(a, x_1)$ as

$$n^{-1}\sum\nolimits_{i=1}^{n} \left[Y_i\{a + m_{\_1}(\mathbf{X}_{i\_1})\} - b\{a + m_{\_1}(\mathbf{X}_{i\_1})\}\right] K_h(X_{i1} - x_1) \tag{5.2.1}$$

and define the oracle smoother of $m_1(x_1)$ as

$$\tilde{m}_{\mathrm{K},1}(x_1) = \underset{a\in A}{\operatorname{argmax}}\, \tilde{l}(a). \tag{5.2.2}$$

THEOREM 5.2.1. *Under Assumptions (A1)-(A6), as $n \to \infty$*

$$\sup_{x_1\in[h,1-h]} \left|\tilde{m}_{\mathrm{K},1}(x_1) - m_1(x_1)\right| = O_{a.s.}\left(\log n/\sqrt{nh}\right) = O_{a.s.}\left(n^{-2/5}\log n\right).$$

THEOREM 5.2.2. *Under Assumptions (A1)-(A6), for any $x_1 \in [h, 1-h]$, as $n \to \infty$, the oracle kernel smoother $\tilde{m}_{\mathrm{K},1}(x_1)$ given in (5.2.2) satisfies*

$$\sqrt{nh}\left\{\tilde{m}_{\mathrm{K},1}(x_1) - m_1(x_1) - \mathrm{bias}_1(x_1)\, h^2/D_1(x_1)\right\}_{l'=1}^{d_1}$$
$$\to N\left(0, D_1(x_1)^{-1} v_1^2(x_1)\, D_1(x_1)^{-1}\right)$$

where

$$D_1(x_1) = f_1(x_1)\, E\left[b''\{m(\mathbf{X})\}\,|X_1 = x_1\right] \tag{5.2.3}$$

and

$$\begin{aligned} v_1^2(x_1) &= f_1(x_1)\, E\left\{\sigma^2(\mathbf{X})\,|X_1 = x_1\right\} \|K\|_2^2, \\ \mathrm{bias}_1(x_1) &= \mu_2(K)\left\{m_1''(x_1)\, f(x_1)\, E\left[b''\{m(\mathbf{X})\}\,|X_1 = x_1\right]\right. \\ &\quad + m_1'(x_1)\, f(x_1)\frac{\partial}{\partial x_1} E\left[b''\{m(\mathbf{X})\}\,|X_1 = x_1\right] \\ &\quad \left. - \{m_1'(x_1)\}^2 f(x_1)\, E\left[b'''\{m(\mathbf{X})\}\,|X_1 = x_1\right]\right\}. \end{aligned} \tag{5.2.4}$$

The same oracle idea applies to the constant as well. Define the the quasi-likelihood function

$$\tilde{l}_c(a) = n^{-1}\sum\nolimits_{i=1}^{n} \left[Y_i\{a + m_{\_c}(\mathbf{X}_i)\} - b\{a + m_{\_c}(\mathbf{X}_i)\}\right],$$

where $m_{\_c}(\mathbf{X}) = \sum_{\alpha=1}^{d} m_\alpha(X_\alpha)$ and then the infeasible estimator is $\tilde{c} = \operatorname{argmax}_{a\in A} \tilde{l}_c(a)$. Clearly, $\tilde{l}_c'(\tilde{c}) = 0$.

THEOREM 5.2.3. *Under Assumptions (A1)-(A5), as $n \to \infty$,*
$\tilde{c} \to_{a.s.} c$ *and* $|\tilde{c} - c| = O_p\left(n^{-1/2}\right)$.

Although the oracle smoother $\tilde{m}_{K,1}(x_1)$ possess the desirable theoretical properties in Theorems 5.2.2 and 5.2.1, it not useful statistics as it is computed based on the knowledge of unavailable functions $\{m_\alpha(x_\alpha)\}_{\alpha=2}^d$ and constants c. They do, however, motivate the spline-backfitted estimators that we introduce in the next section.

5.3 Spline-backfitted Kernel Estimators

We need following Assumption for kernel function.

(A7) *The number of interior knots $N \sim n^{1/4}\log n$, i.e., $c_N n^{1/4}\log n \le N \le C_N n^{1/4}\log n$ for some positive constants c_N,C_N, and the interval width $H = (N+1)^{-1}$.*

In what follows, we denote $\|K\|_2^2 = \int K(u)^2 du, \mu_2(K) = \int K(u) u^2 du$.

For $J = 0, \ldots, N+1$, define the linear B spline basis as

$$b_J(x) = (1 - |x - \xi_J|/H)_+ = \begin{cases} (N+1)x - J + 1 & , \quad \xi_{J-1} \le x \le \xi_J \\ J + 1 - (N+1)x & , \quad \xi_J \le x \le \xi_{J+1} \\ 0 & , \quad \text{otherwise} \end{cases},$$

the space of α-empirically centered linear spline functions on $[0,1]$ as

$$G_n^0 = \left\{g_\alpha : g_\alpha(x_\alpha) \equiv \sum\nolimits_{J=0}^{N+1} \lambda_J b_J(x_\alpha), E_n\{g_\alpha(X_\alpha)\} = 0\right\}, 1 \le \alpha \le d,$$

which is equipped with the empirical inner product $\langle\cdot,\cdot\rangle_{2,n}$. Define $\hat{L}(g) = \frac{1}{n}\sum_{i=1}^n [Y_i g(\mathbf{X}_i) - b\{g(\mathbf{X}_i)\}], g \in G_n^0$. The multivariate function $m(\mathbf{x})$ is estimated by an additive spline function

$$\hat{m}(\mathbf{x}) = \underset{g \in G_n^0}{\operatorname{argmax}}\, \hat{L}(g). \tag{5.3.1}$$

Next define the quasi-likelihood function

$$\hat{l}(a) = \frac{1}{n}\sum\nolimits_{i=1}^n [Y_i\{a + \hat{m}_{\text{-}1}(\mathbf{X}_{i,\text{-}1})\} - b\{a + \hat{m}_{\text{-}1}(\mathbf{X}_{i,\text{-}1})\}] K_h(X_{i1} - x_1) \tag{5.3.2}$$

$$\hat{m}_{\text{SBK},1}(x_1) = \underset{a \in A}{\operatorname{argmax}}\, \hat{l}(a). \tag{5.3.3}$$

THEOREM 5.3.1. *Under Assumptions (A1)-(A7),*

$$\sup_{x_1 \in [0,1]} \left| \hat{m}_{\mathrm{SBK},1}(x_1) - \tilde{m}_{\mathrm{K},1}(x_1) \right| = o_{a.s.}\left(n^{-2/5}\right).$$

Theorem 5.3.1 follows (5.6.11), Lemmas 5.6.11 and 5.6.12. The following theorems are straightforward from Theorems 5.2.2, 5.2.1 and 5.3.1.

THEOREM 5.3.2. *Under Assumptions (A1)-(A7), as $n \to \infty$*

$$\sup_{x_1 \in [h,1-h]} \left| \hat{m}_{\mathrm{SBK},1}(x_1) - m_1(x_1) \right| = O_{a.s.}\left(\log n/\sqrt{nh}\right) = O_{a.s.}\left(n^{-2/5}\log n\right).$$

THEOREM 5.3.3. *Under Assumptions (A1)-(A7), for any $x_1 \in [h, 1-h]$, as $n \to \infty$, $\hat{m}_{\mathrm{SBK},1}(x_1)$ given in (5.3.3) satisfies*

$$\sqrt{nh}\left\{\hat{m}_{\mathrm{SBK},1}(x_1) - m_1(x_1) - \mathrm{bias}_1(x_1)\, h^2/D_1(x_1)\right\}_{l'=1}^{d_1}$$
$$\to N\left(0, D_1(x_1)^{-1} v_1^2(x_1)\, D_1(x_1)^{-1}\right)$$

where $\mathrm{bias}_1(x_1)$ *and* $D_1(x_1)$ *are defined as (5.2.4) and (5.2.3).*

Then define $\hat{l}_c(a) = n^{-1}\sum_{i=1}^{n}\left[Y_i\left\{a + \hat{m}_{\_c}(\mathbf{X}_i)\right\} - b\left\{a + \hat{m}_{\_c}(\mathbf{X}_i)\right\}\right]$, where $\hat{m}_{\_c}(\mathbf{X}) = \sum_{\alpha=1}^{d} \hat{m}_\alpha(X_\alpha)$. Define next the spline-backfitted estimator $\hat{c} = \mathrm{argmax}_{a \in A}\, \hat{l}_c(a)$.

THEOREM 5.3.4. *Under Assumptions (A1)-(A5) and (A7), as $n \to \infty$, $|\hat{c} - \tilde{c}| = O_p\left(n^{-1/2}\right)$, hence $|\hat{c} - c| = O_p\left(n^{-1/2}\right)$.*

5.4 Implementation

We implement our procedures with the following rule-of-thumb number of interior knots

$$N = N_n = \min\left(\left[n^{1/4}\log n\right] + 1,\right)$$

which satisfies Assumption (A8), i.e.$N = N_n \sim n^{1/4}\log n$, and ensures that the number of parameters in the linear least squares problem.

According to Theorem 5.3.3, the asymptotic distributions of the estimators $\hat{m}_{\mathrm{SBK},\alpha}(x_\alpha)$ depend not only on the functions $\mathrm{bias}_\alpha(x_\alpha)/D_\alpha(x_\alpha)$ and $D_\alpha(x_\alpha)^{-1} v_\alpha^2(x_\alpha)\, D_\alpha(x_\alpha)^{-1}$, but also crucially on the choice of bandwidths h_α. So we define the optimal bandwidth of h_α, denoted by $h_{\alpha,\text{opt}}$, as the minimizer of the asymptotic mean integrated squared errors (AMISE) of $\{\hat{m}_\alpha(x_\alpha), l = 1, \ldots, d\}$, which is defined as

$$\begin{aligned}\text{AMISE}\{\hat{m}_\alpha\} &= \int \Big[\left\{\text{bias}_\alpha(x_\alpha) h_\alpha^2 / D_\alpha(x_\alpha)\right\}^2 \\ &\quad + D_\alpha(x_\alpha)^{-1} v_\alpha^2(x_\alpha) D_\alpha(x_\alpha)^{-1} / (n h_\alpha) \Big] f_\alpha(x_\alpha)\, dx_\alpha.\end{aligned}$$

By letting $d\,\text{AMISE}\{\hat{m}_{\alpha,\cdot}\}/dh_\alpha = 0$, one gets the optimal bandwidth $h_{\alpha,\text{opt}}$ as

$$h_{\alpha,\text{opt}} = \left\{ \frac{n^{-1} \int D_\alpha(x_\alpha)^{-1} v_\alpha^2(x_\alpha) D_\alpha(x_\alpha)^{-1} f_\alpha(x_\alpha)\, dx_\alpha}{4 \int \{\text{bias}_\alpha(x_\alpha)/D_\alpha(x_\alpha)\}^2 f_\alpha(x_\alpha)\, dx_\alpha} \right\}^{1/5},$$

which is approximated by

$$\hat{h}_{\alpha,\text{opt}} = \left\{ \frac{n^{-1} \sum_{i=1}^n D_\alpha(X_{i\alpha})^{-1} v_\alpha^2(X_{i\alpha}) D_\alpha(X_{i\alpha})^{-1}}{4 \sum_{i=1}^n \{\text{bias}_\alpha(X_{i\alpha})/D_\alpha(X_{i\alpha})\}^2} \right\}^{1/5},$$

where

$$D_\alpha(x_\alpha) = f_\alpha(x_\alpha) E\left[b''\{m(\mathbf{X})\} | X_\alpha = x_\alpha\right]$$

and

$$\begin{aligned} v_\alpha^2(x_\alpha) &= f_\alpha(x_\alpha) E\left\{\sigma^2(\mathbf{X}) | X_\alpha = x_\alpha\right\} \|K\|_2^2, \\ \text{bias}_\alpha(x_\alpha) &= \mu_2(K) \Big\{ m''_\alpha(x_\alpha) f(x_\alpha) E\left[b''\{m(\mathbf{X})\} | X_\alpha = x_\alpha\right] \\ &\quad + m'_\alpha(x_\alpha) f(x_\alpha) \frac{\partial}{\partial x_\alpha} E\left[b''\{m(\mathbf{X})\} | X_\alpha = x_\alpha\right] \\ &\quad - \{m'_\alpha(x_\alpha)\}^2 f(x_\alpha) E\left[b'''\{m(\mathbf{X})\} | X_\alpha = x_\alpha\right] \Big\}. \end{aligned}$$

To implement this, we propose following estimation methods for the terms $m'_\alpha(x_\alpha)$, $m''_\alpha(x_\alpha)$, $f_\alpha(x_\alpha)$, $E\{\sigma^2(\mathbf{X}) | X_\alpha = x_\alpha\}$, $E\left[b''\{m(\mathbf{X})\} | X_\alpha = x_\alpha\right]$, $E\left[b'''\{m(\mathbf{X})\} | X_\alpha = x_\alpha\right]$ and $\frac{\partial}{\partial x_\alpha} E\left[b''\{m(\mathbf{X})\} | X_\alpha = x_\alpha\right]$. The resulting bandwidth is denoted as $\hat{h}_{\alpha,\text{opt}}$.

- The derivative functions $m'_\alpha(X_{i\alpha})$ and $m''_\alpha(X_{i\alpha})$ are estimated as $\sum_{k=1}^3 k \hat{a}_{\alpha,l,k} X_{i\alpha}^{k-1} + 3 \sum_{k=4}^{N+3} \hat{a}_{\alpha,l,k} (X_{i1} - t_{\alpha,k-3})^2$ and

$\sum_{k=2}^{3} k(k-1)\hat{a}_{\alpha,l,k}X_{i\alpha}^{k-2} + 6\sum_{k=4}^{N+3}\hat{a}_{\alpha,l,k}\left(X_{i1} - t_{\alpha,k-3}\right)$ where $\left\{\hat{a}_{\alpha,l,k}\right\}_{k=0}^{N+3}$ maximize the following

$$\sum_{i=1}^{n}\left\{Y_i\left(\sum_{k=0}^{3} a_{\alpha,l,k}X_{i\alpha}^{k} + \sum_{k=4}^{N+3} a_{\alpha,l,k}\left(X_{i\alpha} - t_{\alpha,k-3}\right)^3\right)\right.$$
$$\left. -b\left(\sum_{k=0}^{3} a_{\alpha,l,k}X_{i\alpha}^{k} + \sum_{k=4}^{N+3} a_{\alpha,l,k}\left(X_{i\alpha} - t_{\alpha,k-3}\right)^3\right)\right\}$$

where $\min_i X_{i\alpha} = t_{\alpha,0} < \cdots < t_{\alpha,N+1} = \max_i X_{i\alpha}$.

- $E\left[b''\left\{m\left(\mathbf{X}\right)\right\}|X_\alpha = x_\alpha\right]$ is estimated as $\sum_{k=0}^{3}\hat{a}_{\alpha,l,k}^{k}x_\alpha^{k} + \sum_{k=4}^{N+3}\hat{a}_{\alpha,l,k}\left(x_\alpha - t_{\alpha,k-3}\right)^3$ by minimizing

$$\sum_{i=1}^{n}\left[b''\left\{\hat{m}\left(\mathbf{X}_i\right)\right\} - \left\{\sum_{k=0}^{3} a_{\alpha,l,k}X_\alpha^{k} + \sum_{k=4}^{N+3} a_{\alpha,l,k}\left(X_\alpha - t_{k-3}\right)^3\right\}\right]^2,$$

$\frac{\partial}{\partial x_\alpha}E\left[b''\left\{m\left(\mathbf{X}\right)\right\}|X_\alpha = x_\alpha\right]$ and $E\left[b'''\left\{m\left(\mathbf{X}\right)\right\}|X_\alpha = x_\alpha\right]$ are estimated by $\sum_{k=1}^{3} k\hat{a}_{\alpha,l,k}^{k}x_\alpha^{k-1} + 3\sum_{k=4}^{N+3}\hat{a}_{\alpha,l,k}\left(x_\alpha - t_{\alpha,k-3}\right)^2$ and $\sum_{k=0}^{3}\hat{a}_{\alpha,l,k}^{k}x_\alpha + \sum_{k=4}^{N+3}\hat{a}_{\alpha,l,k}\left(x_\alpha - t_{\alpha,k-3}\right)^3$ by minimizing

$$\sum_{i=1}^{n}\left[b'''\left\{\hat{m}\left(\mathbf{X}_i\right)\right\} - \left\{\sum_{k=0}^{3} a_{\alpha,l,k}X_\alpha^{k} + \sum_{k=4}^{N+3} a_{\alpha,l,k}\left(X_\alpha - t_{k-3}\right)^3\right\}\right]^2.$$

- $E\left\{\sigma^2\left(\mathbf{X}\right)|X_\alpha = x_\alpha\right\}$ is estimated by $\sum_{k=0}^{3}\hat{a}_{\alpha,l,k}^{k}x_\alpha + \sum_{k=4}^{N+3}\hat{a}_{\alpha,l,k}\left(x_\alpha - t_{\alpha,k-3}\right)^3$ by minimizing

$$\sum_{i=1}^{n}\left[\left[Y_i - b'\left\{\hat{m}\left(\mathbf{X}_i\right)\right\}\right]^2 - \left\{\sum_{k=0}^{3} a_{\alpha,l,k}X_\alpha^{k} + \sum_{k=4}^{N+3} a_{\alpha,l,k}\left(X_\alpha - t_{k-3}\right)^3\right\}\right]^2.$$

- Density function $f_\alpha\left(x_\alpha\right)$ is estimated by $\frac{1}{n}\sum_{i=1}^{n} K_{h_\alpha}\left(X_{i\alpha} - x_\alpha\right)$ with the rule-of-the-thumb bandwidth h_α.

5.5 Examples

5.5.1 Simulation 1

The data are generated from the model

$$Y = g^{-1}\left\{\sum_{j=1}^{d} m_j\left(X_j\right)\right\}, g^{-1}\left(x\right) = \frac{e^x}{1+e^x}$$

with $m_1 = \sin(\pi x)$, $m_2 = \Phi(3x)$ and $m_3(x) = m_4(x) = m_5(x) = x$, where Φ is the standard normal distribution function. The data are generated from the following vector autoregression (VAR) equation for $0 \le a, r < 1$,

$$\mathbf{X}_t = a\mathbf{X}_{t-1} + \varepsilon_i, \varepsilon_i \sim N(0, \Sigma), 2 \le t \le n, \Sigma = \begin{bmatrix} 1 & r & \cdots & r \\ r & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & r \\ r & \cdots & r & 1 \end{bmatrix},$$

with stationary distribution $\mathbf{X}_t = (X_{t1}, ..., X_{td})^T \sim N\left(0, \left(1 - a^2\right)^{-1} \Sigma\right)$. Clearly, Higher values of a correspond to stronger dependence among the observations, and in particular, if $a = 0$, the data is i.i.d. The parameter r controls the correlation of the bivariate X_{t1} and X_{t2}. In this study, we have experimented with two cases: $r = 0$, $a = 0$; $r = 0.5$, $a = 0.5$ to cover various scenarios. For $\alpha = 1, ..., d$, let $x^i_{\alpha,\min}$, $x^i_{\alpha,\max}$ denote the smallest and largest observations of the variable x_α in the i -th replication. The functions $\{m_\alpha\}_{\alpha=1}^d$ are estimated on sample values.

Denoting the estimator of m_l in the k-th replication as $\hat{m}_{\text{SBK},\alpha,k}$ and $X_{t\alpha}$ are the points where the functions are evaluated, we define the (mean) integrated squared error (ISE and MISE) as

$$\begin{aligned} \text{ISE}(\hat{m}_{\text{SBK},\alpha,k}) &= \frac{1}{n}\sum\nolimits_{t=1}^{n} \left\{\hat{m}_{\text{SBK},\alpha,k}(X_{t\alpha,k}) - m_\alpha(X_{t\alpha,k})\right\}^2, \\ \text{MISE}(\hat{m}_{\text{SBK},\alpha}) &= \frac{1}{100}\sum\nolimits_{k=1}^{100} \text{ISE}(\hat{m}_{\text{SBK},\alpha,k}). \end{aligned}$$

Then to see that the SBK estimator is as efficient as the "oracle smoother" $\tilde{m}_{\text{K},\alpha}(x_\alpha)$, we define the empirical relative efficiency of $\hat{m}_{\text{SBK},\alpha}(x_\alpha)$ with respect to $\tilde{m}_{\text{K},\alpha}(x_\alpha)$ as

$$\underset{\alpha}{\text{EFF}} = \left[\frac{\sum_{t=1}^{n} \left\{\tilde{m}_{\text{K},\alpha}(x_\alpha) - m_\alpha(X_{t\alpha})\right\}^2}{\sum_{t=1}^{n} \left\{\hat{m}_{\text{SBK},\alpha}(X_{t\alpha}) - m_\alpha(X_{t\alpha})\right\}^2}\right]^{1/2}.$$

Tables 9 and 10 show the MISEs of Effs of $\hat{m}_{\text{SBK},\alpha}$ and $\tilde{m}_{\text{K},\alpha}$ for $\alpha = 1, 2$. It is obvious that the SBK estimator has as good as performance of oracle estimator, and it corroborates with Theorem 5.3.1.

5.5.2 Simulation 2

Using the same model in Simulation 1 but with high dimension $d = 10$, where $m_\alpha(x_\alpha) = \sin(\pi x)$, $\alpha = 1, ..., 10$ and data are generated the same way. We have run 100 replications for sample size $n = 500$, 1000, 1500, 2000. The MISEs of Effs of $\hat{m}_{\mathrm{SBK},1}$ and $\tilde{m}_{\mathrm{K},1}$ are shown in Table 11. As expected, increases in sample size reduce MISE for both estimators and across all combinations of r and α values.

To see the convergence, Figure 13 plots the kernel density estimation of the 100 empirical efficiencies for $\alpha = 1$ and sample sizes $n = 500$, 1000, 1500, 2000 for $r = 0$, $a = 0$. The vertical line at efficiency = 1 is the standard line for the comparison of $\hat{m}_{\mathrm{SBK},1}$ and $\tilde{m}_{\mathrm{K},1}$. One can clearly see that the center of the density plots is going toward the standard line 1.0 with narrower spread when sample size n is increasing, which is confirmative to the result of Theorem 5.3.1. The basic graphic pattern of Figure 16 with $r = 0.5$, $a = 0.5$ is similar to that for the i.i.d case, though with slower convergence rate and relatively poorer efficiency.

To have some impression of the actual function estimates, for $r = 0$, $a = 0$ and $r = 0.5$, $a = 0.5$ with sample size $n = 500$, 1000, 1500, 2000, we have plotted the SBK estimators and their 95% pointwise confidence intervals (three dotted lines), oracle estimators (dashed lines) for the true functions m_1 (solid lines) in Figures 17–24. The visual impression of the SBK estimators is rather satisfactory and their performance improves with increasing sample size.

Lastly, we provide the computing time of Example 2 from 100 replications on an ordinary PC with Intel Pentium IV 1.86 GHz processor and 1.0 GB RAM. The average time run by XploRe to generate one sample of size n and compute the SBK estimator is reported in Table 12.

5.6 Appendix

5.6.1 Preliminaries

In the proofs that follow, we use U and u to denote sequences of random variables that are uniformly O and o of certain order.

LEMMA 5.6.1. *([70], Lemma A.2) Tthere exist constants* $c_0 > 0$ *such that for any* $\boldsymbol{\lambda} = (\lambda_0, \lambda_{J,\alpha})^T_{1\le J\le N+1, 1\le\alpha\le d} \in R^{1+d(N+1)}$,

$$c_0\left(\lambda_0^2 + \sum_{J,\alpha}^{2} \lambda_{J,\alpha}^2\right) \le \left\|\lambda_0 + \sum_{J,\alpha}\lambda_{J,\alpha}B_{J,\alpha}\right\|_2^2. \tag{5.6.1}$$

LEMMA 5.6.2. *([70], Lemma A.4)Under Assumptions (A2), (A4) and (A6), the uniform supremum of the rescaled difference between* $\langle g_1, g_2\rangle_{2,n}$ *and* $\langle g_1, g_2\rangle_2$ *is*

$$A_n = \sup_{g_1,g_2\in G_n^{(0)}[0,1]} \frac{\left|\langle g_1, g_2\rangle_{2,n} - \langle g_1, g_2\rangle_2\right|}{\|g_1\|_2\,\|g_2\|_2} = O_{a.s.}\left(\frac{\log n}{n^{1/2}H^{1/2}}\right). \tag{5.6.2}$$

5.6.2 Oracle smoothers

LEMMA 5.6.3. *Under Assumptions (A1)-(A6),*

$$\sup_{x_1\in[h,1-h]}\left|\tilde{l}'(m_1(x_1)) - \text{bias}_1(x_1)h^2\right| = O_{a.s.}\left(\log n/\sqrt{nh}\right)$$

where $\text{bias}_1(x_1)$ *is defined as (5.2.4).*

Proof. According to (5.2.1), $\tilde{l}'(m_1(x_1))$ equals

$$1/n\sum\nolimits_{i=1}^{n}\left[Y_i - b'\{m_1(x_1) + m_{\text{-}1}(\mathbf{X}_{i\text{-}1})\}\right]K_h(X_{i1} - x_1) \tag{5.6.3}$$

$$= 1/n\sum\nolimits_{i=1}^{n}\left[b'\{m(\mathbf{X}_i)\} - b'\{m_1(x_1) + m_{\text{-}1}(\mathbf{X}_{i\text{-}1})\} + \sigma(\mathbf{X}_i)\varepsilon_i\right]K_h(X_{i1} - x_1)$$

Let $\xi_{i,n} = \xi_{i,n}(x_1) = \xi_{i,n,1} + \xi_{i,n,2}$ is

$$\left[b'\{m(\mathbf{X}_i)\} - b'\{m_1(x_1) + m_{\text{-}1}(\mathbf{X}_{i\text{-}1})\} + \sigma(\mathbf{X}_i)\varepsilon_i\right]K_h(X_{i1} - x_1) \tag{5.6.4}$$

$$-E\left[\left[b'\{m(\mathbf{X}_i)\} - b'\{m_1(x_1) + m_{\text{-}1}(\mathbf{X}_{i\text{-}1})\} + \sigma(\mathbf{X}_i)\varepsilon_i\right]K_h(X_{i1} - x_1)\right]$$

where

$$\xi_{i,n,1} = \xi_{i,n,1}(x_1) = \sigma(\mathbf{X}_i)\varepsilon_i K_h(X_{i1} - x_1).$$

$$\begin{aligned}\xi_{i,n,2} &= \xi_{i,n,2}(x_1) = \left[b'\{m(\mathbf{X}_i)\} - b'\{m_1(x_1) + m_{\text{-}1}(\mathbf{X}_{i\text{-}1})\}\right]K_h(X_{i1} - x_1)\\ &\quad -E\left[\left[b'\{m(\mathbf{X}_i)\} - b'\{m_1(x_1) + m_{\text{-}1}(\mathbf{X}_{i\text{-}1})\}\right]K_h(X_{i1} - x_1)\right].\end{aligned}$$

Then according to (5.6.3), one can rewrite $l^{*\prime}(m_1(x_1))$ as

$$1/n \sum_{i=1}^{n} \xi_{i,n} + E\left[b'\{m(\mathbf{X}_i)\} - b'\{m_1(x_1) + m_{\text{-}1}(\mathbf{X}_{i\text{-}1})\}\right] K_h(X_{i1} - x_1).$$

While

$$\begin{aligned}
& E\left[b'\{m(\mathbf{X}_i)\} - b'\{m_1(x_1) + m_{\text{-}1}(\mathbf{X}_{i\text{-}1})\}\right] K_h(X_{i1} - x_1) \\
&= \int_{[0,1]^d} \left[b'\{m(\mathbf{u})\} - b'\{m_1(x_1) + m_{\text{-}1}(\mathbf{u}_{\text{-}1})\}\right] \frac{1}{h} K\left(\frac{u_1 - x_1}{h}\right) f(\mathbf{u})\, d\mathbf{u} \\
&= \int_{[0,1]^d} \Big[b''\{m(x_1, \mathbf{u}_{\text{-}1})\}\{m_1(u_1) - m_1(x_1)\} \\
&\quad + \frac{1}{2} b'''\{m(x_1, \mathbf{u}_{\text{-}1})\}\{m_1(u_1) - m_1(x_1)\}^2 + u\left(h^2\right)\Big] \\
&\quad \frac{1}{h} K\left(\frac{u_1 - x_1}{h}\right) f(u_1, \mathbf{u}_{\text{-}1})\, du_1 d\mathbf{u}_{\text{-}1} + u\left(h^2\right)
\end{aligned}$$

$$\begin{aligned}
&= \int_{[0,1]^{d-1}} \int_{[-1,1]} \left[b''\{m(x_1, \mathbf{u}_{\text{-}1})\}\left\{hv_1 m_1'(x_1) + \frac{(hv_1)^2}{2} m_1''(x_1) + u\left(h^2\right)\right\}\right. \\
&\quad \left. + \frac{1}{2} b'''\{m(x_1, \mathbf{u}_{\text{-}1})\}\left\{hv_1 m_1'(x_1) + (hv_1)^2 m_1''(x_1) + u\left(h^2\right)\right\}^2\right] \\
&\quad K(v_1)\left\{f(x_1, \mathbf{u}_{\text{-}1}) + hv_1 \frac{\partial f(x_1, \mathbf{u}_{\text{-}1})}{\partial x_1} + U\left(h^2\right)\right\} dv_1 d\mathbf{u}_{\text{-}1} + u\left(h^2\right)
\end{aligned}$$

$$\begin{aligned}
&= h^2 \int_{[-1,1]} v_1^2 K(v_1)\, dv_1 \left\{\frac{m_1''(x_1) f_1(x_1)}{2} \int_{[0,1]^{d-1}} b''\{m(x_1, \mathbf{u}_{\text{-}1})\} f(\mathbf{u}|x_1)\, d\mathbf{u}_{\text{-}1}\right. \\
&\quad \left. + m_1'(x_1) \int_{[0,1]^{d-1}} b''\{m(x_1, \mathbf{u}_{\text{-}1})\} \frac{\partial f(x_1, \mathbf{u}_{\text{-}1})}{\partial x_1} d\mathbf{u}_{\text{-}1}\right\} + u\left(h^2\right). \\
&= h^2 \mu_2(K)\left\{m_1''(x_1) f(x_1) E\left[b''\{m(\mathbf{X})\}|X_1 = x_1\right]\right. \\
&\quad + m_1'(x_1) \frac{\partial}{\partial x_1}\left[f(x_1) E\left[b''\{m(\mathbf{X})\}|X_1 = x_1\right]\right] \\
&\quad \left. - \{m_1'(x_1)\}^2 f(x_1) E\left[b'''\{m(\mathbf{X})\}|X_1 = x_1\right]\right\} + u\left(h^2\right).
\end{aligned}$$

Let $D_n = n^\alpha$ with $\alpha < \frac{2}{5}$, $\alpha(2+\eta) > 1, \alpha(1+\eta) > 2/5$, which requires $\eta > 1/2$. Rewrite $\varepsilon_i = \varepsilon_{i,1}^{D_n} + \varepsilon_{i,2}^{D_n} + \varepsilon_{i,3}^{D_n}$ where $\varepsilon_{i,1}^{D_n} = \varepsilon_i\{|\varepsilon_i| > D_n\}, \varepsilon_{i,2}^{D_n} = \varepsilon_i\{|\varepsilon_i| \le D_n\} - \varepsilon_{i,3}^{D_n}, \varepsilon_{i,3}^{D_n} = E\varepsilon_i\{|\varepsilon_i| \le D_n\}$. Define for $j = 1, 2, 3$, $\xi_{i,n,1,j} = \xi_{i,n,1,j}(x_1)$ is

$$\left[b'\{m(\mathbf{X}_i)\} - b'\{m_1(x_1) + m_{\text{-}1}(\mathbf{X}_{i\text{-}1})\} + \sigma(\mathbf{X}_i)\varepsilon_{i,j}^{D_n}\right] K_h(X_{i1} - x_1).$$

According to Assumption (A5), one has

$$\sum_{n=1}^{\infty} P\left(|\varepsilon_i| \geq D_n\right) \leq \sum_{n=1}^{\infty} \frac{C_\sigma E\left|\varepsilon_i\right|^{2+\eta}}{D_n^{2+\eta}}$$

$$\leq C_\sigma C_\delta \sum_{n=1}^{\infty} \frac{1}{D_n^{2+\eta}} = C_\sigma C_\delta \sum_{n=1}^{\infty} n^{-\alpha(2+\eta)} < \infty.$$

By Borel-Cantelli Lemma, one has with probability 1, $n^{-1}\sum_{i=1}^{n} \xi_{i,n,1} = 0$ for large n. Therefore, one has $\sup_{x_1 \in [0,1]} \left|n^{-1}\sum_{i=1}^{n} \xi_{i,n,1,1}\right| = U\left(n^{-k}\right)$ for any $k > 0$. Using Assumption (A5),

$$\left|\varepsilon_{i,3}^{D_n}\right| = \left|-E\varepsilon_i\left\{|\varepsilon_i| > D_n\right\}\right| \leq \frac{E\left|\varepsilon_i\right|^{2+\eta}}{D_n^{1+\eta}} = O\left(n^{-2/5}\right).$$

Hence

$$\begin{aligned} n^{-1}\sum_{i=1}^{n} \xi_{i,n,1,3} &= n^{-1}\sum_{i=1}^{n} K_h\left(X_{i1} - x_1\right)\sigma\left(\mathbf{X}_i\right)\varepsilon_{i,3}^{D_n} \\ &= n^{-1}\sum_{i=1}^{n} K_h\left(X_{i1} - x_1\right) O\left(n^{-2/5}\right) = O_{a.s.}\left(n^{-2/5}\right). \end{aligned}$$

Meanwhile

$$E\xi_{i,n,1,2}^2 = E\left[\sigma\left(\mathbf{X}_i\right)\varepsilon_i K_h\left(X_{i1} - x_1\right)\right]^2$$

$$= h^{-1} f_1\left(x_1\right) E\left\{\sigma^2\left(\mathbf{X}\right)|X_1 = x_1\right\} \|K\|_2^2 \left\{1 + u\left(1\right)\right\}.$$

$$E\left|\xi_{i,n,1,2}\right|^k = E\left(\left|\xi_{i,n,1,2}\right|^{k-2}\left|\xi_{i,n,1,2}\right|^2\right)$$

$$\leq \sup_{x_1 \in [0,1]} \left|\xi_{i,n,1,2}\right|^{k-2} E\left|\xi_{i,n,1,2}\right|^2 \leq C_\sigma 2^{k-2} D_n^{k-2}/h^{k-2} E\left|\xi_{i,n,1,2}\right|^2,$$

then there exist a constant $c_1 = C_\sigma D_n/h$ such that

$E\left(\left|\xi_{i,n,1,2}\right|^k\right) \leq c_1^{k-2} k! E(\xi_{i,n,1,2}^2)$, $k \geq 2$. By using Lemma 2.5.2, we let $k = 3, a_2\left(3\right) = 11n\left(1 + \frac{5m_3^{6/7}}{\varepsilon_n}\right)$, $m_2^2 = E(\xi_{i,n,1,2}^2) = O\left(h^{-1}\right)$, $\varepsilon_n = a\frac{\log n}{\sqrt{nh}}$,

$$P\left\{\left|\sum_{i=1}^{n} \xi_{i,n,1,2}\right| > n\varepsilon_n\right\} \leq a_1 \exp\left(-\frac{q\varepsilon_n^2}{25m_2^2 + 5c_1\varepsilon_n}\right) + a_2\left(3\right)\alpha\left(\left[\frac{n}{q+1}\right]\right)^{\frac{6}{7}}$$

take q such that $\left[\frac{n}{q+1}\right] \geq c_2 \log n, q \geq \frac{c_3 n}{\log n}$ for some constants c_2, c_3.

$$\frac{q\varepsilon_n^2}{25m_2^2 + 5c\varepsilon_n} = \frac{q\frac{\left(a\log n\right)^2}{nh}}{25m_2^2 + 5c_1\varepsilon_n} \geq \frac{\frac{c_3 n}{\log n} a^2 \frac{\left(\log n\right)^2}{nh}}{25m_2^2 + 5c_1 a\frac{\log n}{\sqrt{nh}}}$$

$$\geq \frac{c_3 a^2 \log n}{25m_2^2 h + 5c_0 D_n / ha\frac{\log n}{\sqrt{nh}}h} = \frac{c_3 a^2 \log n}{25m_2^2 h + 5ac_0 n^{\alpha} n^{-1/2} h^{-1/2} \log n} \sim a^2 \log n,$$

$$a_1 = 2\frac{n}{q} + 2\left(1 + \frac{\varepsilon_n^2}{25m_2^2 + 5c_1\varepsilon_n}\right) = O(\log n),$$

$$a_2(3) = 11n\left(1 + \frac{5m_3^{6/7}}{\varepsilon_n}\right), \text{ with } m_3 = \max_{1\leq i\leq N} \|\xi_{i,n,1,2}\|_3 \leq c_6 D_n,$$

$$a_2(3) \leq 11n\left\{1 + \frac{c_6 D_n}{an^{-1/2}h^{-1/2}\log n}\right\} = o\left(n^2\right),$$

$$\alpha\left(\left[\frac{n}{q+1}\right]\right)^{6/7} \leq \left(K_0 e^{-\lambda_0\left[\frac{n}{q+1}\right]}\right)^{6/7} \leq Cn^{-6\lambda_0 c_2/7},$$

therefore for large n

$$\begin{aligned} & P\left\{n^{-1}\left|\sum_{i=1}^{n} \xi_{i,n,1,2}\right| > a\log n/\sqrt{nh}\right\} \\ \leq\ & O(\log n)\exp\left(-c_5 a^2 \log n\right) + Cn^{2-6\lambda_0 c_2/7} \\ =\ & n^{-c_5 a^2} O(\log n) + Cn^{2-6\lambda_0 c_2/7} \end{aligned}$$

for c_2, c_5, a large enough. For all $x_1 \in [h, 1-h]$, we discrete by equally spaced $h = x_{1,0} < x_{1,1} < \cdots < x_{1,M_n} = 1-h$, $M_n = n^4$,

$$P\left\{\max_{0\leq j\leq M_n} n^{-1}\left|\sum_{i=1}^{n} \xi_{i,n,1,2}(x_{1,j})\right| > a\log n/\sqrt{nh}\right\}$$

$$\leq \sum_{0=1}^{M_n} P\left\{n^{-1}\left|\sum_{i=1}^{n} \xi_{i,n,1,2}(x_{1,j})\right| > a\log n/\sqrt{nh}\right\} \leq Cn^{-8}M_n \leq Cn^{-2}$$

for a and c_2 large enough. Borel-Cantelli lemma implies that

$$\max_{1\leq j\leq M_n}\left|n^{-1}\sum_{i=1}^{n} \xi_{i,n,1,2}(x_{1,j})\right| = O_{a.s.}\left(a\log n/\sqrt{nh}\right).$$

Taking supremum over the whole interval $[h, 1-h]$, one has

$$\sup_{x_1\in[h,1-h]}\left|n^{-1}\sum_{i=1}^{n} \xi_{i,n,1,2}(x_1)\right| \leq \max_{0\leq j\leq M_n}\left|n^{-1}\sum_{i=1}^{n} \xi_{i,n,1,2}(x_{1,j})\right| +$$

$$n^{-1}\max_{0\leq j\leq M_n-1}\ \sup_{x_1\in[x_{1,j},x_{1,j+1}]}\left|\sum_{i=1}^{n} \xi_{i,n,1,2}(x_1) - \sum_{i=1}^{n} \xi_{i,n,1,2}(x_{1,j})\right|$$

$$\leq Cn^{-1/2}\log n + CM_n^{-1}h^{-2}.$$

by Lipschitz continuity of kernel K. So one has

$$\sup_{x_1\in[0,1]}\left|n^{-1}\sum_{i=1}^{n}\xi_{i,n,1,2}\right| = O_{a.s.}\left\{(nh)^{-1/2}\log n\right\}.$$

Then putting $\xi_{i,n,1,1}, \xi_{i,n,1,2}, \xi_{i,n,1,3}$ together, one has $\sup_{x_1\in[0,1]}\left|n^{-1}\sum_{i=1}^{n}\xi_{i,n,1}\right| = O_{a.s.}\left\{(nh)^{-1/2}\log n\right\}$.

Because $E\left[\left[b'\left\{m\left(\mathbf{X}_i\right)\right\} - b'\left\{m_1\left(x_1\right) + m_{\text{-}1}\left(\mathbf{X}_{i\text{-}1}\right)\right\}\right]K_h\left(X_{i1}-x_1\right)\right] = U\left(h^2\right)$, so $E\xi_{i,n,2}^2$ is

$$h^{-2}\int_{[0,1]^d}\left[b'\left\{m\left(\mathbf{u}\right)\right\} - b'\left\{m_1\left(x_1\right) + m_{\text{-}1}\left(\mathbf{u}_{\text{-}1}\right)\right\}\right]^2 K\left(\frac{u_1-x_1}{h}\right)^2 f\left(\mathbf{u}\right)d\mathbf{u} + U\left(h^4\right)$$

$$= h^{-1}\int_{[0,1]^{d-1}}\int_{[-1,1]}\left[b'\left\{m\left(x_1+hv_1,\mathbf{u}_{\text{-}1}\right)\right\} - b'\left\{m_1\left(x_1\right) + m_{\text{-}1}\left(\mathbf{u}_{\text{-}1}\right)\right\}\right]^2$$
$$K\left(v_1\right)^2 f\left(x_1+hv_1,\mathbf{u}_{\text{-}1}\right)dv_1 d\mathbf{u}_{\text{-}1} + U\left(h^4\right)$$

$$= h^{-1}\int_{[0,1]^{d-1}}\int_{[-1,1]}\left[b''\left\{m\left(x_1,\mathbf{u}_{-1}\right)\right\}\left\{hv_1 m_1'\left(x_1\right) + U\left(h^2\right)\right\}\right]^2$$
$$K\left(v_1\right)^2\left\{f\left(x_1,\mathbf{u}_{\text{-}1}\right) + U\left(h\right)\right\}dv_1 d\mathbf{u}_{\text{-}1} + U\left(h^4\right) = U\left(h\right).$$

Note that $\sup_{x_1}\left|b'\left\{m\left(\mathbf{X}_i\right)\right\} - b'\left\{m_1\left(x_1\right) + m_{\text{-}1}\left(\mathbf{X}_{i\text{-}1}\right)\right\}\right| \leq C_b h$ when $K_h\left(X_{i1}-x_1\right) \neq 0$. Similar to the proof for $\xi_{i,n,2}$, one has

$$E\left|\xi_{i,n,2}\right|^k \leq \left(2C_b\right)^{k-2}E\xi_{i,n,1}^2$$

and then $\sup_{x_1\in[h,1-h]}\left|n^{-1}\sum\limits_{i=1}^{n}\xi_{i,n,2}\right| = O_{a.s.}\left\{(nh)^{-1/2}\log n\right\}$.
Putting $\xi_{i,n,1}, \xi_{i,n,2}$ together, the lemma is proved. □

LEMMA 5.6.4. *Under Assumptions (A2), (A4)-(A6),*

$$\sup_{x_1\in[h,1-h]}\left|\tilde{l}''\left(m_1\left(x_1\right)\right) + D_1\left(x_1\right)\right| = O_{a.s.}\left(\log n/\sqrt{nh}\right),$$

where $D_1\left(x_1\right)$ is defined as (5.2.3).

Proof. According to (5.6.3), one has $l^{*''}\left(m_1\left(x_1\right)\right)$ is

$$-1/n\sum_{i=1}^{n}\left[b''\left\{m_1\left(x_1\right) + m_{\text{-}1}\left(\mathbf{X}_{i\text{-}1}\right)\right\}\right]K_h\left(X_{i1}-x_1\right) \qquad (5.6.5)$$

Let $\zeta_{i,n} = \left[b''\left\{m_1(x_1) + m_{\text{-}1}(\mathbf{X}_{i\text{-}1})\right\}\right] K_h(X_{i1} - x_1)$, then

$$E\zeta_{i,n} = E\left[\left[b''\left\{m_1(x_1) + m_{\text{-}1}(\mathbf{X}_{i\text{-}1})\right\}\right] K_h(X_{i1} - x_1)\right]$$

$$= \int_{[0,1]^{d_2}} b''\left\{m_1(x_1) + m_{\text{-}1}(\mathbf{u}_{\text{-}1})\right\} \frac{1}{h} K\left(\frac{u_1 - x_1}{h}\right) f(\mathbf{u})\, d\mathbf{u}$$

$$= \int_{[0,1]^{d_2}} \int_{[-1,1]} b''\left\{m_1(x_1) + m_{\text{-}1}(\mathbf{u}_{\text{-}1})\right\} K(v_1) f(x_1 + hv_1, \mathbf{u}_{\text{-}1})\, dv_1 d\mathbf{u}_{\text{-}1}$$

$$= \int_{[0,1]^{d_2}} \int_{[-1,1]} b''\left\{m_1(x_1) + m_{\text{-}1}(\mathbf{u}_{\text{-}1})\right\} K(v_1) \left\{f(x_1, \mathbf{u}_{\text{-}1}) + hv_1 \frac{\partial f(x_1, \mathbf{u}_{\text{-}1})}{\partial x_1} + U\left(h^2\right)\right\} dv_1 d\mathbf{u}_{\text{-}1}$$

$$= \int_{[0,1]^{d_2}} \int_{[-1,1]} b''\left\{m_1(x_1) + m_{\text{-}1}(\mathbf{u}_{\text{-}1})\right\} K(v_1) f(x_1, \mathbf{u}_{\text{-}1})\, dv_1 d\mathbf{u}_{\text{-}1} + U\left(h^2\right)$$

$$= f_1(x_1) E\left[b''\left\{m(\mathbf{X})\right\} | X_1 = x_1\right] + U\left(h^2\right).$$

$$E\zeta_{i,n}^2 = E\left[\left[b''\left\{m_1(x_1) + m_{\text{-}1}(\mathbf{X}_{i\text{-}1})\right\}\right] K_h(X_{i1} - x_1)\right]^2$$

$$= \int_{[0,1]^{d_2}} \left[b''\left\{m_1(x_1) + m_{\text{-}1}(\mathbf{u}_{\text{-}1})\right\}\right]^2 \frac{1}{h^2} K^2\left(\frac{u_1 - x_1}{h}\right) f(\mathbf{u})\, d\mathbf{u}$$

$$= h^{-1} \int_{[0,1]^{d_2}} \int_{[-1,1]} \left[b''\left\{m_1(x_1) + m_{\text{-}1}(\mathbf{u}_{\text{-}1})\right\}\right]^2 K^2(v_1) f(x_1 + hv_1, \mathbf{u}_{\text{-}1})\, dv_1 d\mathbf{u}_{\text{-}1}$$

$$= h^{-1} \int_{[0,1]^{d_2}} \int_{[-1,1]} \left[b''\left\{m_1(x_1) + m_{\text{-}1}(\mathbf{u}_{\text{-}1})\right\}\right]^2 K^2(v_1) \left\{f(x_1, \mathbf{u}_{\text{-}1}) + hv_1 \frac{\partial f(x_1, \mathbf{u}_{\text{-}1})}{\partial x_1} + U\left(h^2\right)\right\} dv_1 d\mathbf{u}_{\text{-}1}$$

$$= h^{-1} f_1(x_1) \|K\|_2^2 E\left[\left[b''\left\{m(\mathbf{X})\right\}\right]^2 | X_1 = x_1\right] + U\left(h^2\right).$$

Similar to the proof of Lemma 5.6.3, the result follows the Lemma 2.5.2. □

LEMMA 5.6.5. *Under Assumptions (A1) to (A3), (A5) and (A7), as $n \to \infty$, there exists a constant C such that*

$$\sup_{x_1 \in [h, 1-h]} \left|\text{cov}\left(\xi_{i,n}, \xi_{j,n}\right)\right| \le Ch^{-\frac{1+\eta}{2+\eta}} \alpha(j - i)^{\frac{\eta}{2+\eta}} \text{ for } i \neq j$$

Proof. According to Davydov's Inequality, for $\frac{1}{p}+\frac{1}{q}+\frac{1}{r}=1$, $\operatorname{cov}\left(\xi_{i,n},\xi_{j,n}\right)$ is bounded by

$$C_2\left\{2\alpha\left(j-i\right)\right\}^{1/p}\left\|\xi_{i,n,1}+\xi_{i,n,2}\right\|_q\left\|\xi_{j,n,1}+\xi_{j,n,2}\right\|_r$$

$$\leq C_2\left\{2\alpha\left(j-i\right)\right\}^{1/p}\left(\left\|\xi_{i,n,1}\right\|_q+\left\|\xi_{i,n,2}\right\|_q\right)\left(\left\|\xi_{j,n,1}\right\|_r+\left\|\xi_{j,n,2}\right\|_r\right)$$

Let $q=r=2+\eta, p=1+2/\eta$, where η takes value in the Assumption (A5), then one has $\left\|\xi_{i,n,1}\right\|_q=U\left(h^{-\frac{1}{2+\eta}}\right)$ and $\left\|\xi_{i,n,1}\right\|_q=U\left(h^{-\frac{1+\eta}{2+\eta}}\right)$. $\operatorname{cov}\left(\xi_{i,n,l'},\xi_{j,n,l''}\right)\leq Ch^{-\frac{1+\eta}{2+\eta}}\alpha\left(j-i\right)^{\frac{\eta}{2+\eta}}$ for some constant C. □

Proof of Theorem 5.2.1. Existing a $\bar{m}_1\left(x_1\right)$ between $\tilde{m}_{\mathrm{K},1}\left(x_1\right)$ and $m_1\left(x_1\right)$ such that

$$\tilde{l}'\left(\tilde{m}_{\mathrm{K},1}\left(x_1\right)\right)-\tilde{l}'\left(m_1\left(x_1\right)\right)=\tilde{l}''\left(\bar{m}_1\left(x_1\right)\right)\left\{\tilde{m}_{\mathrm{K},1}\left(x_1\right)-m_1\left(x_1\right)\right\}$$

Note that $\tilde{l}'\left(\tilde{m}_{\mathrm{K},1}\left(x_1\right)\right)=0$, then

$$\tilde{m}_{\mathrm{K},1}\left(x_1\right)-m_1\left(x_1\right)=-\frac{\tilde{l}'\left(m_1\left(x_1\right)\right)}{\tilde{l}''\left(\bar{m}_1\left(x_1\right)\right)}, \tag{5.6.6}$$

Lemma 5.6.4 implies that $c\leq\sup_{x_1\in[h,1-h]}\left|-\tilde{l}''\left(\bar{m}_1\left(x_1\right)\right)\right|\leq C$ *a.s.* for some constants $0<c<C$. Then the theorem follows Lemma 5.6.3 and (5.6.6).

Proof of Theorem 5.2.2. Let $S_n=S_n\left(x_1\right)=\sum_{i=1}^n\xi_{i,n}$, where $\xi_{i,n}$ is defined as (5.6.4), then one has $ES_n=0$ and $\tilde{l}'\left(m_1\left(x_1\right)\right)=S_n/n+b\left(x_1\right)h^2+u\left(h^2\right)$.

$$\gamma\left(k\right)=\gamma\left(k,x_1\right)=\operatorname{cov}\left(\xi_{i,n},\xi_{i+k,n}\right)$$

$$\sigma_n^2=ES_n^2=\operatorname{var}\left(S_n\right)=\operatorname{var}\left(\sum_{i=1}^n\xi_{i,n}\right)=\sum_{i=1}^n\operatorname{var}\left(\xi_{i,n}\right)+\sum_{i\neq j}^n\operatorname{cov}\left(\xi_{i,n},\xi_{j,n}\right)$$
$$=n\operatorname{var}\left(\xi_{i,n}\right)+n\sum_{1\leq|k|\leq n-1}\left(1-\frac{|k|}{n}\right)\gamma\left(k\right)=n\operatorname{var}\left(\xi_{i,n}\right)+nA_n,$$

where

$$\operatorname{var}\left(\xi_{i,n}\right)=h^{-1}f_1\left(x_1\right)E\left\{\sigma^2\left(\mathrm{X}\right)|X_1=x_1\right\}\|K\|_2^2+U\left(h^4\right).$$

While according to Lemma 5.6.5, one has

$$\left|\gamma\left(k\right)\right|=\left|\operatorname{cov}\left(\xi_{i,n},\xi_{i+k,n}\right)\right|\leq Ch^{-\frac{1+\eta}{2+\delta}}\alpha\left(k\right)^{\frac{\eta}{2+\eta}}.$$

Hence

$$
\begin{aligned}
|A_n| &= \left|\sum\nolimits_{1\le|l|\le n-1}\gamma(k)\right| \le \sum\nolimits_{1\le|l|\le n-1}\left(1-\frac{|k|}{n}\right)h^{-\frac{1+\eta}{2+\eta}}\left\{K_0\exp\left(-\lambda_0 k\right)\right\}^{\frac{\eta}{2+\eta}} \\
&\le K_0 h^{-\frac{1+\eta}{2+\eta}}\sum\nolimits_{1\le|l|\le n-1}\exp\left\{-\lambda_0 k\eta/\left(2+\eta\right)\right\},
\end{aligned}
$$

so there exists a constant C_1 such that $A_n \le C_1 h^{-\frac{1+\eta}{2+\eta}}$. So $A_n/\operatorname{var}\left(\xi_{i,n}\right) \to 0$ as $n \to \infty$. . Then $\sigma_n^2 \sim n\operatorname{var}\left(\xi_{i,n}\right) \ge c_0 n$ when n is large, so according to (2.5.1) in Lemma 2.5.1, there exist constants c_1 and c_2 such that for some $0 < \eta \le 1$

$$
\Delta_n = \sup_z\left|P\left\{\sigma_n^{-1}S_n < z\right\} - \Phi(z)\right| \le c_1\frac{d_\eta}{c_0\sigma_n^\eta}\left\{\log\left(\sigma_n/c_0^{1/2}\right)/\lambda\right\}^{1+\eta}
$$

for any λ with $\lambda_1 \le \lambda \le \lambda_2$, where

$$
\lambda_1 = c_2\left\{\log\left(\sigma_n/c_0^{1/2}\right)\right\}^b/n, b > 2\left(1+\eta\right)/\eta; \lambda_2 = 4\left(2+\eta\right)\eta^{-1}\log\left(\sigma_n/c_0^{1/2}\right).
$$

For the η in Assumption (A5), set $\lambda = 4\left(2+\eta\right)\eta^{-1}\log\left(\sigma_n/c_0^{1/2}\right)$, then by Assumption (A6) d_η is

$$
\begin{aligned}
&\max_{1\le i\le n}\left\{E\left|\left[b'\left\{m\left(\mathbf{X}_i\right)\right\} - b'\left\{m_1\left(x_1\right) + m_{\text{-}1}\left(\mathbf{X}_{i\text{-}1}\right)\right\} + \sigma\left(\mathbf{X}_i\right)\varepsilon_i\right]K_h\left(X_{i1}-x_1\right)\right|^{2+\eta}\right\} \\
&\qquad = \max_{1\le i\le n}\left\{E\left|C_b h + \sigma\left(\mathbf{X}_i\right)\varepsilon_i\right|^{2+\eta}\left|K_h\left(X_{i1}-x_1\right)\right|^{2+\eta}\right\} \\
&\qquad \le CC_\delta C_\eta\left\{E\left|K_h\left(X_1-x_1\right)\right|^{2+\eta}\right\} = O\left\{h^{-(1+\eta)}\right\},
\end{aligned}
$$

i.e., $\Delta_n = O\left\{h^{-(1+\eta)}/\sigma_n^\eta\right\} = O\left\{n^{(1+\eta/2)/5-\eta/2}\right\} = O\left(n^{1/5-2\eta/5}\right) \to 0$ when $1/2 < \eta \le 1$. So $S_n/\sigma_n \to N(0,1)$, then

$$
n\left\{l^{*\prime}\left(m_1\left(x_1\right)\right) - \operatorname{bias}_1\left(x_1\right)h^2\right\}/\sqrt{nh^{-1}v_1^2\left(x_1\right)} \to N(0,1),
$$

where $v_1^2\left(x_1\right)$ defined as (5.2.4). Meanwhile, according to Theorem 5.2.1, one has as $n \to \infty$, $\sup_{x_1\in[h,1-h]}\left|\tilde{l}''\left(m_1\left(x_1\right)\right) - \tilde{l}''\left(\bar{m}_1\left(x_1\right)\right)\right| \to 0$ because $\sup_{x_1\in[h,1-h]}\left|m_1\left(x_1\right) - \bar{m}_1\left(x_1\right)\right| \to 0$. Then according to Slutsky's theorem, one has

$$
\sqrt{nh}\left\{\left\{\tilde{m}_{\mathrm{K},1}\left(x_1\right) - m_1\left(x_1\right)\right\}D_1\left(x_1\right) - \operatorname{bias}_1\left(x_1\right)h^2\right\} \to N\left(0, v_1^2\left(x_1\right)\right).
$$

where $D_1\left(x_1\right)$ is defined in (5.6.7). Then the theorem is proved.

PROOF OF THEOREM 5.2.3. According to Mean Value Theorem, there exists a $\bar{c}$ between c and $\tilde{c}$ such that $(\tilde{c}-c)\tilde{l}''_c(\bar{c}) = \tilde{l}'_c(\tilde{c}) - \tilde{l}'_c(c) = -\tilde{l}'_c(c)$, where $-\tilde{l}''_c(\bar{c}) = n^{-1}\sum_{i=1}^n b''\{\bar{c}+m_{\_c}(\mathbf{X}_i)\} > c_b > 0$ according to Assumption (A2) and where $m_{\_c}(\mathbf{X}) = \sum_{\alpha=1}^d m_\alpha(X_\alpha)$ and then the infeasible estimator is $\tilde{c} = \mathrm{argmax}_{a\in A}\tilde{l}_c(a)$. Clearly, $\tilde{l}'_c(\tilde{c}) = 0$. Using Bernstein's Inequality, one has

$$\left|\tilde{l}'_c(c)\right| = \left|n^{-1}\sum\nolimits_{i=1}^n \left[Y_i - b'\{c+m_{\_c}(\mathbf{X}_i)\}\right]\right| = \left|n^{-1}\sum\nolimits_{i=1}^n \sigma(\mathbf{X}_i)\varepsilon_i\right| \rightarrow_{a.s.} 0,$$

which implies $|\tilde{c}-c| = U_{a.s.}\left(n^{-1/2}\right)$. So $\left|\tilde{l}''_c(\bar{c}) - \tilde{l}''_c(c)\right| \rightarrow_{a.s.} 0$, in which $\tilde{l}''_c(c) = -n^{-1}\sum_{i=1}^n b''\{c+m_{\_c}(\mathbf{X}_i)\}$ and it convergents to $Eb''\{m(\mathbf{X})\}$ almost sure. Then according to central limit theorem,

$$\sqrt{n}(\tilde{c}-c) \rightarrow_d N\left(0, \left[Eb''\{m(\mathbf{X})\}\right]^{-2} E\sigma^2(\mathbf{X}_i)\right).$$

5.6.3 Spline backfitted kernel estimators

In this section, we give the proof of Theorem 5.3.1. First, define the theoretical inner product of b_J and 1 with respect to the α-th marginal density $f_\alpha(x_\alpha)$ as $c_{J,\alpha} = \langle b_J(X_\alpha), 1\rangle = \int b_J(x_\alpha) f_\alpha(x_\alpha)\,dx_\alpha$ and define the centered B spline basis $b_{J,\alpha}(x_\alpha)$ and the standardized B spline basis $B_{J,\alpha}(x_\alpha)$ as

$$\begin{aligned} b_{J,\alpha}(x_\alpha) &= b_J(x_\alpha) - \frac{c_{J,\alpha}}{c_{J-1,\alpha}} b_{J-1}(x_\alpha), \\ B_{J,\alpha}(x_\alpha) &= \frac{b_{J,\alpha}(x_\alpha)}{\left\|b_{J,\alpha}\right\|_2}, 1 \le J \le N+1, \end{aligned} \tag{5.6.7}$$

so that $EB_{J,\alpha}(X_\alpha) \equiv 0$, $EB^2_{J,\alpha}(X_\alpha) \equiv 1$. For $\forall g \in G_n^0$, one can write $g = \boldsymbol{\lambda}^T\mathbf{B}(\mathbf{X}_i)$ for a vector $\boldsymbol{\lambda} = \left(\lambda_0, \lambda_{J,\alpha}\right)^T_{1\le J\le N+1, 1\le\alpha\le d} \in R^{1+d(N+1)}$ and

$$\mathbf{B}(\mathbf{x}) = \left\{1, B_{1,1}(x_1), ..., B_{N+1,d}(x_d)\right\}^T, \tag{5.6.8}$$

Then with a slight abuse of notation, we denote

$\hat{L}(g) = \hat{L}(\boldsymbol{\lambda}) = n^{-1}\sum_{i=1}^n \left[Y_i\boldsymbol{\lambda}^T\mathbf{B}(\mathbf{X}_i) - b\left\{\boldsymbol{\lambda}^T\mathbf{B}(\mathbf{X}_i)\right\}\right]$ and then

$$\frac{\partial\hat{L}(\boldsymbol{\lambda})}{\partial\boldsymbol{\lambda}} = n^{-1}\sum\nolimits_{i=1}^n \left[Y_i\mathbf{B}(\mathbf{X}_i) - b'\left\{\boldsymbol{\lambda}^T\mathbf{B}(\mathbf{X}_i)\right\}\mathbf{B}(\mathbf{X}_i)\right]. \tag{5.6.9}$$

The multivariate function $m(\mathbf{x})$ is estimated by an additive spline function

$$\begin{aligned} \hat{m}(\mathbf{x}) &= \hat{m}_0 + \sum_{\alpha=1}^{d} \hat{m}_\alpha(x_\alpha) = \hat{\boldsymbol{\lambda}}^T \mathbf{b}(\mathbf{x}), \\ \hat{\boldsymbol{\lambda}} &= \left(\hat{\lambda}_0, \hat{\lambda}_{J,\alpha}\right)^T_{\substack{1\le\alpha\le d \\ 1\le J\le N+1}} = \underset{\boldsymbol{\lambda}}{\text{argmax}}\, \hat{L}(\boldsymbol{\lambda}). \end{aligned} \tag{5.6.10}$$

According to (5.3.2), existing a $\bar{m}_{\text{K},1}(x_1)$ between $\hat{m}_{\text{SBK},1}(x_1)$ and $\tilde{m}_{\text{K},1}(x_1)$ such that

$$\hat{l}'\left(\hat{m}_{\text{SBK},1}(x_1)\right) - \hat{l}'\left(\tilde{m}_{\text{K},1}(x_1)\right) = \hat{l}''\left(\bar{m}_{\text{K},1}(x_1)\right)\left\{\hat{m}_{\text{SBK},1}(x_1) - \tilde{m}_{\text{K},1}(x_1)\right\},$$

Then according to $\hat{l}'\left(\hat{m}_{\text{SBK},1}(x_1)\right) = 0$, one has

$$\hat{m}_{\text{SBK},1}(x_1) - \tilde{m}_{\text{K},1}(x_1) = -\frac{\hat{l}'\left(\tilde{m}_{\text{K},1}(x_1)\right)}{\hat{l}''\left(\bar{m}_{\text{K},1}(x_1)\right)} \tag{5.6.11}$$

Let $\bar{m}$ be an additive spline function such that $\|\bar{m} - m\|_\infty \le C_\infty H^2$ in the Lemma 3.6.1 and $\bar{\boldsymbol{\lambda}}$ such that

$$\bar{m}(\mathbf{x}) = \bar{\boldsymbol{\lambda}}^T \mathbf{B}(\mathbf{x}). \tag{5.6.12}$$

In what follows, we denote the dimension of vector $\bar{\boldsymbol{\lambda}}$ as $N_d = (N+1)d + 1$.

PROOF OF THEOREM 5.3.4. Existing $\bar{c}'$ between $\hat{c}$ and $\tilde{c}$ such that $\hat{c} - \tilde{c} = -\hat{l}_c'(\tilde{c})/l_c''(\bar{c}')$,where $-\hat{l}_c''(\bar{c}') = n^{-1}\sum_{i=1}^n b''\{\bar{c}' + \hat{m}_{\text{-}c}(\mathbf{X}_i)\} > c_b > 0$ according to Assumption (A6), then

$$\begin{aligned} \hat{l}_c'(\tilde{c}) = \hat{l}_c'(\tilde{c}) - \tilde{l}_c'(\tilde{c}) &= n^{-1}\sum_{i=1}^n \left[b'\{\tilde{c} + m_{\text{-}c}(\mathbf{X}_i)\} - b'\{\tilde{c} + \hat{m}_{\text{-}c}(\mathbf{X}_i)\}\right] \\ &= 1/n\sum_{i=1}^n b''\{\tilde{c} + m_{\text{-}c}(\mathbf{X}_i)\}\{m_{\text{-}c}(\mathbf{X}_i) - \hat{m}_{\text{-}c}(\mathbf{X}_i)\} \\ &\quad + O\left[1/n\sum_{i=1}^n \{m_{\text{-}c}(\mathbf{X}_i) - \hat{m}_{\text{-}c}(\mathbf{X}_i)\}^2\right] \\ &= I + O_{a.s.}\left(N_d H^4 + N_d n^{-1}\log n\right), \end{aligned}$$

by Lemma 5.6.9, where $I = I_1 + I_2$,

$$I_1 = 1/n\sum_{i=1}^n b''\{\tilde{c} + m_{\text{-}c}(\mathbf{X}_i)\}\{m_{\text{-}c}(\mathbf{X}_i) - \bar{m}_{\text{-}c}(\mathbf{X}_i)\},$$

$$I_2 = 1/n\sum_{i=1}^n b''\{\tilde{c} + m_{\text{-}c}(\mathbf{X}_i)\}\{\bar{m}_{\text{-}c}(\mathbf{X}_i) - \hat{m}_{\text{-}c}(\mathbf{X}_i)\}.$$

According to (3.6.1), $I_1 = O_{a.s.}(H^2)$, while

$$\begin{aligned} I_2 &= n^{-1}\sum_{i=1}^n b''\{\tilde{c} + m_{\text{-}c}(\mathbf{X}_i)\} \times \\ &\quad \left\{\sum_{1\le J\le N+1, 1\le\alpha\le d}\left(\hat{\lambda}_{J,\alpha} - \bar{\lambda}_{J,\alpha}\right)B_{J,\alpha}(X_{i\alpha})\right\} \end{aligned}$$

$$= I_{2,b} + I_{2,v} + I_{2,r}$$

where

$$I_{2,b} = n^{-1} \sum_{i=1}^{n} b'' \{\tilde{c} + m_{\_c}(\mathbf{X}_i)\} \left\{ \sum_{1 \le J \le N+1, 1 \le \alpha \le d} \Phi_{b,J,\alpha} B_{J,\alpha}(X_{i\alpha}) \right\},$$

$$I_{2,v} = n^{-1} \sum_{i=1}^{n} b'' \{\tilde{c} + m_{\_c}(\mathbf{X}_i)\} \left\{ \sum_{1 \le J \le N+1, 1 \le \alpha \le d} \Phi_{v,J,\alpha} B_{J,\alpha}(X_{i\alpha}) \right\},$$

$$I_{2,r} = n^{-1} \sum_{i=1}^{n} b'' \{\tilde{c} + m_{\_c}(\mathbf{X}_i)\} \left\{ \sum_{1 \le J \le N+1, 1 \le \alpha \le d} \Phi_{r,J,\alpha} B_{J,\alpha}(X_{i\alpha}) \right\}.$$

$$|I_{2,b}| \le C_b n^{-1} \sum_{i=1}^{n} \left\{ \sum_{1 \le J \le N+1, 1 \le \alpha \le d} |\Phi_{b,J,\alpha}| \, |B_{J,\alpha}(X_{i\alpha})| \right\}$$

$$\le C_b \left\{ \sum_{1 \le J \le N+1, 1 \le \alpha \le d} \Phi_{b,J,\alpha}^2 \right\}^{1/2} \times$$

$$\left[1 + \sum_{1 \le J \le N+1, 2 \le \alpha \le d} \left\{ n^{-1} \sum_{i=1}^{n} |B_{J,\alpha}(X_{i\alpha})| \right\}^2 \right]^{1/2}$$

$$= C_b \times O_{a.s.}\left(N_d^{1/2} H^{5/2}\right) \times [O_{a.s.}(1) + (N+1) \times (d-1) \times O_{a.s.}(H)]$$

$$= O_{a.s.}\left(N_d^{1/2} H^{5/2}\right)$$

according to (5.6.19) and (5.6.21), similarly

$$|I_{2,r}| = O_{a.s.}\left(N_d H^{7/2} + N_d H^{-1/2} n^{-1} \log n\right).$$

One has $I_{2,v} = \tilde{I}_{2,v} + O_{a.s.}\left(n^{-1/2}\right) \times O_{a.s.}\left(N_d^{1/2} n^{-1/2} \log n\right) \times O(N)$, where

$$\tilde{I}_{2,v} = n^{-1} \sum_{i=1}^{n} b'' \{m(\mathbf{X}_i)\} \left\{ \sum_{1 \le J \le N+1, 1 \le \alpha \le d} \Phi_{v,J,\alpha} B_{J,\alpha}(X_{i\alpha}) \right\}$$

$$= \sum_{1 \le J \le N+1, 1 \le \alpha \le d} \Phi_{v,J,\alpha} n^{-1} \sum_{i=1}^{n} b'' \{m(\mathbf{X}_i)\} B_{J,\alpha}(X_{i\alpha})$$

$$= \sum_{1 \le J \le N+1, 1 \le \alpha \le d} \Phi_{v,J,\alpha} E b'' \{m(\mathbf{X})\} B_{J,\alpha}(X_\alpha)$$

$$+ O_{a.s.}\left(N_d^{1/2} n^{-1/2} \log n\right) \times N_d^{1/2} \times O_{a.s.}\left(n^{-2/5} \log n\right)$$

$$= \bar{I}_{2,v} + O_{a.s.}\left(N_d n^{-9/10} \log^2 n\right)$$

where

$$\bar{I}_{2,v} = \sum_{1 \le J \le N+1, 1 \le \alpha \le d} \Phi_{v,J,\alpha} E b'' \{m(\mathbf{X})\} B_{J,\alpha}(X_\alpha)$$

$$= \sum\nolimits_{1 \le J \le N+1, 1 \le \alpha \le d} Eb'' \{m(\mathbf{X})\} B_{J,\alpha}(X_\alpha)$$
$$\frac{1}{n}\sum_{i=1}^{n} \sigma(\mathbf{X}_i)\varepsilon_i \left\{ S_{J,\alpha} + \sum\nolimits_{1 \le J' \le N+1, 1 \le \alpha' \le d} S_{J,\alpha,J',\alpha'} B_{J',\alpha'}(X_{i\alpha'}) \right\}$$

$$= \frac{1}{n}\sum_{i=1}^{n} \sigma(\mathbf{X}_i)\varepsilon_i \sum\nolimits_{1 \le J \le N+1, 1 \le \alpha \le d} \mu_{b,J,\alpha}(x_1)$$
$$\left\{ S_{J,\alpha} + \sum\nolimits_{1 \le J' \le N+1, 1 \le \alpha' \le d} S_{J,\alpha,J',\alpha'} B_{J',\alpha'}(X_{i\alpha'}) \right\}$$

where $S_{0,0}, S_{J,\alpha}, S_{J,\alpha,J',\alpha'}$ are the corresponding element in the matrix $\mathbf{S}_b$ defined in (5.6.13) has the form

$$\begin{bmatrix} S_{0,0} & S_{1,1} & \cdots & S_{J,\alpha} & \cdots & S_{N+1,d} \\ S_{1,1} & S_{1,1,1,1} & \cdots & S_{1,1,J,\alpha} & \cdots & S_{1,1,N+1,d} \\ \vdots & \vdots & \ddots & & & \vdots \\ S_{J,\alpha} & S_{J,\alpha,1,1} & & S_{J,\alpha,J',\alpha'} & & S_{J,\alpha,N+1,d} \\ \vdots & \vdots & & & \ddots & \vdots \\ S_{N+1,d} & S_{N+1,d,1,1} & \cdots & S_{N+1,d,J',\alpha'} & & S_{N+1,d,N+1,d} \end{bmatrix}$$

and $\mu_{b,J,\alpha} = Eb''\{m(\mathbf{X})\} B_{J,\alpha}(X_\alpha)$, which has the order $U\left(H^{1/2}\right)$. Denote $\bar{I}_{2,v} = \frac{1}{n}\sum_{i=1}^{n} \xi_i$, where

$$\xi_i = \sigma(\mathbf{X}_i)\varepsilon_i \sum\nolimits_{1 \le J \le N+1, 1 \le \alpha \le d} \mu_{b,J,\alpha}(x_1)$$
$$\left\{ S_{J,\alpha} + \sum\nolimits_{1 \le J' \le N+1, 1 \le \alpha' \le d} S_{J,\alpha,J',\alpha'} B_{J',\alpha'}(X_{i\alpha'}) \right\}$$
$$\operatorname{var}\left(\bar{I}_{2,v}\right) = n^{-2} \sum\nolimits_{i=1}^{n} \operatorname{var}(\xi_i) + n^{-2} \sum\nolimits_{i \ne j} \operatorname{cov}(\xi_i, \xi_j)$$

while

$$\operatorname{var}(\xi_i) = \boldsymbol{\mu}_{b\_c}^T \mathbf{S}_{b\_c} \operatorname{var}\{\sigma(\mathbf{X}_i)\varepsilon_i \mathbf{B}(\mathbf{X}_i)\} \mathbf{S}_{b\_c}^T \boldsymbol{\mu}_{b\_c}$$
$$\le C_\sigma C_{\mathbf{S}}^2 C_{\mathbf{V}} \boldsymbol{\mu}_{b\_c}^T \boldsymbol{\mu}_{b\_c}^T = O(1),$$

where $\boldsymbol{\mu}_{b\_c}^T = \left(\mu_{b,J,\alpha}\right)_{J=1,\alpha=1}^{N+1,d}$, $\boldsymbol{\mu}_b^T = \left[Eb''\{m(\mathbf{X})\}, \mu_{b,J,\alpha}\right]_{J=1,\alpha=1}^{N+1,d}$ and $\mathbf{S}_{b\_c}$ is a matrix with rows 2 to $1 + (N+1) \times d$ of $\mathbf{S}_b$. Then $\operatorname{var}(\xi_i) \le C_\sigma C_{\mathbf{S}} \boldsymbol{\mu}_b^T \boldsymbol{\mu}_b = U(1)$, $\operatorname{cov}(\xi_i, \xi_j) = 0$ for $i \ne j$ according to Assumption (A30), then $\operatorname{var}\left(\bar{I}_{2,v}\right) = \left(n^{-1}\right)$. So $\bar{I}_{2,v} = O_p\left(n^{-1/2}\right)$. Then it follows Theorem 5.2.3.

LEMMA 5.6.6. *Under Assumptions (A1)-(A5) and (A7),*

$$\left|\left.\frac{\partial \hat{L}(\boldsymbol{\lambda})}{\partial \boldsymbol{\lambda}}\right|_{\boldsymbol{\lambda}=\bar{\boldsymbol{\lambda}}}\right| = O_{a.s.}\left(H^2 + n^{-1/2}\log n\right),$$
$$\left\|\left.\frac{\partial \hat{L}(\boldsymbol{\lambda})}{\partial \boldsymbol{\lambda}}\right|_{\boldsymbol{\lambda}=\bar{\boldsymbol{\lambda}}}\right\| = O_{a.s.}\left(N_d^{1/2}H^2 + N_d^{1/2}n^{-1/2}\log n\right).$$

Proof.

$$\left.\frac{\partial \hat{L}(\boldsymbol{\lambda})}{\partial \boldsymbol{\lambda}}\right|_{\boldsymbol{\lambda}=\bar{\boldsymbol{\lambda}}} = \frac{1}{n}\sum_{i=1}^{n}\left[Y_i\mathbf{B}(\mathbf{X}_i) - b'\left\{\bar{\boldsymbol{\lambda}}^T\mathbf{B}(\mathbf{X}_i)\right\}\mathbf{B}(\mathbf{X}_i)\right]$$
$$= \frac{1}{n}\sum_{i=1}^{n}\left[b'\{m(\mathbf{X}_i)\} - b'\{\bar{m}(\mathbf{X}_i)\} + \sigma(\mathbf{X}_i)\varepsilon_i\right]\mathbf{B}(\mathbf{X}_i)$$

The first element of the above vector is $\frac{1}{n}\sum_{i=1}^{n}\left[\left[b'\{m(\mathbf{X}_i)\} - b'\{\bar{m}(\mathbf{X}_i)\}\right] + \sigma(\mathbf{X}_i)\varepsilon_i\right]$, which is $O_{a.s.}\left(H^2 + n^{-1/2}\log n\right)$ according to Lemmas 3.6.1 and 2.5.2. The other elements can be written as

$$\frac{1}{n}\sum_{i=1}^{n}\left[\xi_{i,J,\alpha,n} + E\left[\left[b'\{m(X_{i\alpha})\} - b'\{\bar{m}(X_{i\alpha})\}\right]B_{J,\alpha}(X_{i\alpha})\right] + \sigma(\mathbf{X}_i)\varepsilon_i B_{J,\alpha}(X_{i\alpha})\right].$$

where $\xi_{i,J,\alpha,n}$ equals

$$\left[b'\{m(X_{i\alpha})\} - b'\{\bar{m}(X_{i\alpha})\}\right]B_{J,\alpha}(X_{i\alpha}) - E\left[\left[b'\{m(X_{i\alpha})\} - b'\{\bar{m}(X_{i\alpha})\}\right]B_{J,\alpha}(X_{i\alpha})\right].$$

One has

$$E\left[\left[b'\{m(X_{i\alpha})\} - b'\{\bar{m}(X_{i\alpha})\}\right]B_{J,\alpha}(X_{i\alpha})\right] = O\left(H^{5/2}\right),$$
$$E\left[\left[b'\{m(X_{i\alpha})\} - b'\{\bar{m}(X_{i\alpha})\}\right]^2 B_{J,\alpha}^2(X_{i\alpha})\right] = O\left(H^4\right).$$

According to Lemma 2.5.2, one has $\left|\frac{1}{n}\sum_{i=1}^{n}\xi_{i,J,\alpha,n}\right| = O_{a.s.}\left(H^{3/2}n^{-1/2}\log n\right)$ and

$$\frac{1}{n}\left|\sum_{i=1}^{n}\sigma(\mathbf{X}_i)\varepsilon_i B_{J,\alpha}(X_{i\alpha})\right| = O_{a.s.}\left(n^{-1/2}\log n\right).$$

Then lemma is proved. □

Denote

$$\mathbf{V} = E\mathbf{B}(\mathbf{X})\mathbf{B}(\mathbf{X})^T, \mathbf{S} = \mathbf{V}^{-1},$$
$$\mathbf{V}_n = n^{-1}\sum_{i=1}^{n}\mathbf{B}(\mathbf{X}_i)\mathbf{B}(\mathbf{X}_i)^T, \mathbf{S}_n = \mathbf{V}_n^{-1},$$
$$\mathbf{V}_b = Eb''\{\bar{m}(\mathbf{X})\}\mathbf{B}(\mathbf{X})\mathbf{B}(\mathbf{X})^T, \mathbf{S}_b = \mathbf{V}_b^{-1}, \tag{5.6.13}$$
$$\mathbf{V}_{n,b} = n^{-1}\sum_{i=1}^{n}b''\{\bar{m}(\mathbf{X}_i)\}\mathbf{B}(\mathbf{X}_i)\mathbf{B}(\mathbf{X}_i)^T, \mathbf{S}_{n,b} = \mathbf{V}_{n,b}^{-1}.$$

LEMMA 5.6.7. *Under Assumption (A2),*

$$c_{\mathbf{V}}\mathbf{I}_{N_d} \leq \mathbf{V}_n \leq C_{\mathbf{V}}\mathbf{I}_{N_d}, c_{\mathbf{S}}\mathbf{I}_{N_d} \leq \mathbf{S}_n \leq C_{\mathbf{S}}\mathbf{I}_{N_d} \ a.s., \tag{5.6.14}$$

$$c_{\mathbf{V},b}\mathbf{I}_{N_d} \leq \mathbf{V}_{n,b} \leq C_{\mathbf{V},b}\mathbf{I}_{N_d}, c_{\mathbf{S},b}\mathbf{I}_{N_d} \leq \mathbf{S}_{n,b} \leq C_{\mathbf{S},b}\mathbf{I}_{N_d} \ a.s.. \tag{5.6.15}$$

Proof. *Take a real vector* $\boldsymbol{\lambda} \in R^{N_d}$*, one has*

$$\left\|\boldsymbol{\lambda}^T \mathbf{B}(\mathbf{X})\right\|_2^2 = \boldsymbol{\lambda}^T \begin{pmatrix} 1 & \mathbf{0}_{N_d-1}^T \\ \mathbf{0}_{N_d-1} & \left\langle B_{J,\alpha}, B_{J',\alpha'} \right\rangle_2 \end{pmatrix} \boldsymbol{\lambda} = \boldsymbol{\lambda}^{\mathbf{T}}\mathbf{V}\boldsymbol{\lambda},$$

where $\mathbf{V} = E\mathbf{B}(\mathbf{X})\mathbf{B}(\mathbf{X})^T$. *According to (5.6.1), there exist constants* $0 < c_{\mathbf{V}} < \infty$ *such that*

$$\left\|\boldsymbol{\lambda}^T \mathbf{B}(\mathbf{X})\right\|_2^2 = \lambda_0^2 + \left\|\sum_{J,\alpha} \lambda_{J,\alpha} B_{J,\alpha}(X_\alpha)\right\|_2^2 \geq c_{\mathbf{V}}\left(\lambda_0^2 + \sum_{J,\alpha} \lambda_{J,\alpha}^2\right),$$

thus one concludes that

$$\boldsymbol{\lambda}^{\mathbf{T}}\mathbf{V}\boldsymbol{\lambda} \geq c_{\mathbf{V}}\left(\lambda_0^2 + \sum_{J,\alpha} \lambda_{J,\alpha}^2\right) = c_{\mathbf{V}}\boldsymbol{\lambda}^{\mathbf{T}}\boldsymbol{\lambda},$$

which implies that $c_{\mathbf{V}}\mathbf{I}_{N_d} \leq \mathbf{V}$. *On the other hand, according to Lemma 4.7.2 and* C_r*−inequality*

$$\left\|\boldsymbol{\lambda}^T \mathbf{B}(\mathbf{X}_i)\right\|_2^2 = \lambda_0^2 + \left\|\sum_{J,\alpha} \lambda_{J,\alpha} B_{J,\alpha}(X_\alpha)\right\|_2^2 \leq C_{\mathbf{V}}\left(\lambda_0^2 + \sum_{J,\alpha} \lambda_{J,\alpha}^2\right),$$

for a constant $c_{\mathbf{V}} < C_{\mathbf{V}} < \infty$*, which implies that* $\mathbf{V} \leq C_{\mathbf{V}}\mathbf{I}_{N_d}$. *Then* $c_{\mathbf{S}}\mathbf{I}_{N_d} \leq \mathbf{S} = \mathbf{V}^{-1} \leq C_{\mathbf{S}}\mathbf{I}_{N_d}$ *follows by changing* $\boldsymbol{\lambda}$ *by* $\mathbf{V}^{-1/2}\boldsymbol{\lambda}$. *Then (5.6.14) follows immediately from Lemma 5.6.2 and (5.6.15) follows Assumption (A6).* □

Define

$$\boldsymbol{\Phi}_b = -\mathbf{S}_b \frac{1}{n} \sum_{i=1}^{n} \left[b'\{m(\mathbf{X}_i)\} - b'\{\bar{m}(\mathbf{X}_i)\}\right] \mathbf{B}(\mathbf{X}_i), \tag{5.6.16}$$

$$\boldsymbol{\Phi}_v = -\mathbf{S}_b \frac{1}{n} \sum_{i=1}^{n} \left[\sigma(\mathbf{X}_i)\varepsilon_i\right] \mathbf{B}(\mathbf{X}_i), \tag{5.6.17}$$

and

$$\boldsymbol{\Phi}_r = \hat{\boldsymbol{\lambda}} - \bar{\boldsymbol{\lambda}} - \boldsymbol{\Phi}_b - \boldsymbol{\Phi}_v. \tag{5.6.18}$$

Lemma 5.6.8. *Under Assumptions (A1)-(A5) and (A7),*

$$\begin{aligned} \left|\hat{\boldsymbol\lambda}-\bar{\boldsymbol\lambda}\right| &= O_{a.s.}\left(H^2+n^{-1/2}\log n\right), \\ \left\|\hat{\boldsymbol\lambda}-\bar{\boldsymbol\lambda}\right\| &= O_{a.s.}\left(N_d^{1/2}H^2+N_d^{1/2}n^{-1/2}\log n\right). \end{aligned} \quad (5.6.19)$$

$$\|\boldsymbol\Phi_b\| = O_{a.s.}\left(H^2N_d^{1/2}n^{-1/2}\log^2 n\right), \|\boldsymbol\Phi_v\| = O_{a.s.}\left(N_d^{1/2}n^{-1/2}\log n\right),$$

$$\|\boldsymbol\Phi_r\| = O_{a.s.}\left(N_dH^{7/2}+N_dH^{-1/2}n^{-1}\log n\right).$$

Proof. Mean Value Theorem implies that there exist an $N_d \times N_d$ diagonal matrix $\mathbf{t}$ whose diagonal elements are in $[0,1]$, such that for $\hat{\boldsymbol\lambda}^* = \mathbf{t}\hat{\boldsymbol\lambda}+\left(\mathbf{I}_{N_d}-\mathbf{t}\right)\bar{\boldsymbol\lambda}$

$$\left.\frac{\partial\hat L(\boldsymbol\lambda)}{\partial\boldsymbol\lambda}\right|_{\boldsymbol\lambda=\hat{\boldsymbol\lambda}} - \left.\frac{\partial\hat L(\boldsymbol\lambda)}{\partial\boldsymbol\lambda}\right|_{\boldsymbol\lambda=\bar{\boldsymbol\lambda}} = \left.\frac{\partial^2\hat L(\boldsymbol\lambda)}{\partial\boldsymbol\lambda\partial\boldsymbol\lambda^T}\right|_{\boldsymbol\lambda=\hat{\boldsymbol\lambda}^*}\left(\hat{\boldsymbol\lambda}-\bar{\boldsymbol\lambda}\right),$$

$$\hat{\boldsymbol\lambda}-\bar{\boldsymbol\lambda} = -\left(\left.\frac{\partial^2\hat L(\boldsymbol\lambda)}{\partial\boldsymbol\lambda\partial\boldsymbol\lambda^T}\right|_{\boldsymbol\lambda=\hat{\boldsymbol\lambda}^*}\right)^{-1}\left.\frac{\partial\hat L(\boldsymbol\lambda)}{\partial\boldsymbol\lambda}\right|_{\boldsymbol\lambda=\bar{\boldsymbol\lambda}}.$$

According to (5.6.9),

$$-\frac{\partial^2\hat L(\boldsymbol\lambda)}{\partial\boldsymbol\lambda\partial\boldsymbol\lambda^T} = \frac{1}{n}\sum_{i=1}^n b''\left\{\boldsymbol\lambda^T\mathbf{B}(\mathbf{X}_i)\right\}\mathbf{B}(\mathbf{X}_i)\mathbf{B}(\mathbf{X}_i)^T$$

So

$$\begin{aligned} c_bn^{-1}\textstyle\sum_{i=1}^n\mathbf{B}(\mathbf{X}_i)\mathbf{B}(\mathbf{X}_i)^T &\le n^{-1}\textstyle\sum_{i=1}^n\left[b''\left\{\boldsymbol\lambda^T\mathbf{B}(\mathbf{X}_i)\right\}\mathbf{B}(\mathbf{X}_i)\mathbf{B}(\mathbf{X}_i)^T\right] \\ &\le C_bn^{-1}\textstyle\sum_{i=1}^n\mathbf{B}(\mathbf{X}_i)\mathbf{B}(\mathbf{X}_i)^T \end{aligned}$$

because $\mathbf{B}(\mathbf{X}_i)\mathbf{B}(\mathbf{X}_i)^T \ge 0$ and Assumption (A7). Lemma 5.6.7 imply that

$$\begin{aligned} 0 &< c_bc_{\mathbf{V}}\mathbf{I}_{N_d} \le -\frac{\partial^2\hat L(\boldsymbol\lambda)}{\partial\boldsymbol\lambda\partial\boldsymbol\lambda^T} = \frac{1}{n}\textstyle\sum_{i=1}^n\left[b''\left\{\boldsymbol\lambda^T\mathbf{B}(\mathbf{X}_i)\right\}\mathbf{B}(\mathbf{X}_i)\mathbf{B}(\mathbf{X}_i)^T\right] \\ &\le C_bC_{\mathbf{V}}\mathbf{I}_{N_d} < \infty \ a.s.. \end{aligned}$$

Then (5.6.19) follows Lemma 5.6.6. Next,

$$-\left.\frac{\partial\hat L(\boldsymbol\lambda)}{\partial\boldsymbol\lambda}\right|_{\boldsymbol\lambda=\bar{\boldsymbol\lambda}} - \left.\frac{\partial^2\hat L(\boldsymbol\lambda)}{\partial\boldsymbol\lambda\partial\boldsymbol\lambda^T}\right|_{\boldsymbol\lambda=\bar{\boldsymbol\lambda}}\left(\hat{\boldsymbol\lambda}-\bar{\boldsymbol\lambda}\right)$$

$$= -\frac{1}{2n}\sum_{i=1}^n b'''\left\{\hat{\boldsymbol\lambda}^{*T}\mathbf{B}(\mathbf{X}_i)\right\}\left\{\left(\hat{\boldsymbol\lambda}-\bar{\boldsymbol\lambda}\right)^T\mathbf{B}(\mathbf{X}_i)\right\}^2\mathbf{B}(\mathbf{X}_i).$$

$$\hat{\boldsymbol{\lambda}} - \bar{\boldsymbol{\lambda}} = -\left(\left.\frac{\partial^2 \hat{L}(\boldsymbol{\lambda})}{\partial \boldsymbol{\lambda} \partial \boldsymbol{\lambda}^T}\right|_{\boldsymbol{\lambda}=\bar{\boldsymbol{\lambda}}}\right)^{-1} \left.\frac{\partial \hat{L}(\boldsymbol{\lambda})}{\partial \boldsymbol{\lambda}}\right|_{\boldsymbol{\lambda}=\bar{\boldsymbol{\lambda}}} +$$

$$\left(\left.\frac{\partial^2 \hat{L}(\boldsymbol{\lambda})}{\partial \boldsymbol{\lambda} \partial \boldsymbol{\lambda}^T}\right|_{\boldsymbol{\lambda}=\bar{\boldsymbol{\lambda}}}\right)^{-1} \frac{1}{2n}\sum_{i=1}^{n} b''' \left\{\hat{\boldsymbol{\lambda}}^{*T}\mathbf{B}(\mathbf{X}_i)\right\} \left\{\left(\hat{\boldsymbol{\lambda}} - \bar{\boldsymbol{\lambda}}\right)^T \mathbf{B}(\mathbf{X}_i)\right\}^2 \mathbf{B}(\mathbf{X}_i),$$

while

$$\begin{aligned}
& -\left(\left.\frac{\partial^2 \hat{L}(\boldsymbol{\lambda})}{\partial \boldsymbol{\lambda} \partial \boldsymbol{\lambda}^T}\right|_{\boldsymbol{\lambda}=\bar{\boldsymbol{\lambda}}}\right)^{-1} \left.\frac{\partial \hat{L}(\boldsymbol{\lambda})}{\partial \boldsymbol{\lambda}}\right|_{\boldsymbol{\lambda}=\bar{\boldsymbol{\lambda}}} \\
= & -\mathrm{S}_b \frac{1}{n}\sum_{i=1}^{n} \left[Y_i \mathbf{B}(\mathbf{X}_i) - b'\left\{\bar{\boldsymbol{\lambda}}^T \mathbf{B}(\mathbf{X}_i)\right\}\mathbf{B}(\mathbf{X}_i)\right] \\
= & \ \boldsymbol{\Phi}_b + \boldsymbol{\Phi}_v
\end{aligned}$$

and

$$\boldsymbol{\Phi}_r = \left(\left.\frac{\partial^2 \hat{L}(\boldsymbol{\lambda})}{\partial \boldsymbol{\lambda} \partial \boldsymbol{\lambda}^T}\right|_{\boldsymbol{\lambda}=\bar{\boldsymbol{\lambda}}}\right)^{-1} \frac{1}{2n}\sum_{i=1}^{n} b''' \left\{\hat{\boldsymbol{\lambda}}^{*T}\mathbf{B}(\mathbf{X}_i)\right\} \left\{\left(\hat{\boldsymbol{\lambda}} - \bar{\boldsymbol{\lambda}}\right)^T \mathbf{B}(\mathbf{X}_i)\right\}^2 \mathbf{B}(\mathbf{X}_i).$$

Note that

$$\begin{aligned}
& \left\|\frac{1}{2n}\sum_{i=1}^{n} b''' \left\{\hat{\boldsymbol{\lambda}}^{*T}\mathbf{B}(\mathbf{X}_i)\right\} \left\{\left(\hat{\boldsymbol{\lambda}} - \bar{\boldsymbol{\lambda}}\right)^T \mathbf{B}(\mathbf{X}_i)\right\}^2 \mathbf{B}(\mathbf{X}_i)\right\| \\
\leq & \ C\frac{1}{2n}\sum_{i=1}^{n} \left\{\left(\hat{\boldsymbol{\lambda}} - \bar{\boldsymbol{\lambda}}\right)^T \mathbf{B}(\mathbf{X}_i)\right\}^2 \|\mathbf{B}(\mathbf{X}_i)\| \leq \frac{C}{2H^{1/2}n}\sum_{i=1}^{n} \left\{\left(\hat{\boldsymbol{\lambda}} - \bar{\boldsymbol{\lambda}}\right)^T \mathbf{B}(\mathbf{X}_i)\right\}^2 \\
\leq & \ \frac{C}{2H^{1/2}}\left(\hat{\boldsymbol{\lambda}} - \bar{\boldsymbol{\lambda}}\right)^T \left\{\frac{1}{n}\sum_{i=1}^{n} \mathbf{B}(\mathbf{X}_i)\mathbf{B}(\mathbf{X}_i)^T\right\}\left(\hat{\boldsymbol{\lambda}} - \bar{\boldsymbol{\lambda}}\right) \\
\leq & \ \frac{C}{2H^{1/2}}\left\|\hat{\boldsymbol{\lambda}} - \bar{\boldsymbol{\lambda}}\right\|^2 = O_{a.s.}\left(N_d H^4 + N_d n^{-1}\log^2 n\right) \times \frac{1}{H^{1/2}} \\
= & \ O_{a.s.}\left(N_d H^{7/2} + N_d H^{-1/2} n^{-1}\log^2 n\right).
\end{aligned}$$

So $\|\boldsymbol{\Phi}_r\| = O_{a.s.}\left(N_d H^{7/2} + N_d H^{-1/2} n^{-1}\log n\right)$. Next,

$$\|\boldsymbol{\Phi}_b\|^2 = \left\|\mathrm{S}_b \frac{1}{n}\sum_{i=1}^{n} \left[b'\{m(\mathbf{X}_i)\} - b'\{\bar{m}(\mathbf{X}_i)\}\right]\mathbf{B}(\mathbf{X}_i)\right\|^2$$

$$\leq C C_b^2 C_{\mathbf{S}}^2 H^4 \left\|\frac{1}{n}\sum_{i=1}^{n} \mathbf{B}(\mathbf{X}_i)\right\|^2 = O_{a.s.}\left(H^4 N_d n^{-1}\log^2 n\right).$$

$$\begin{aligned}
\|\boldsymbol{\Phi}_v\|^2 &= \left\|\mathrm{S}_b \frac{1}{n}\sum_{i=1}^{n} \left[\sigma(\mathbf{X}_i)\varepsilon_i\right]\mathbf{B}(\mathbf{X}_i)\right\|^2 \\
&\leq C C_{\mathbf{S}}^2 \left\|\frac{1}{n}\sum_{i=1}^{n} \left[\sigma(\mathbf{X}_i)\varepsilon_i\right]\mathbf{B}(\mathbf{X}_i)\right\|^2 \\
&= O_{a.s.}\left(N_d n^{-1}\log^2 n\right)
\end{aligned}$$

□

LEMMA 5.6.9. *Under Assumptions (A1)-(A5) and (A7),*

$$\begin{aligned}\|\hat{m}-\bar{m}\|_2 &= O_{a.s.}\left(N_d^{1/2}H^2+N_d^{1/2}n^{-1/2}\log n\right),\\ \|\hat{m}-\bar{m}\|_{2,n} &= O_{a.s.}\left(N_d^{1/2}H^2+N_d^{1/2}n^{-1/2}\log n\right),\end{aligned}$$

$$\|\hat{m}-m\|_{2,n} = O_{a.s.}\left(N_d^{1/2}H^2+N_d^{1/2}n^{-1/2}\log n\right). \tag{5.6.20}$$

Proof. According to Lemma 5.6.7,

$$\|\hat{m}-\bar{m}\|_2^2 = \left\|\left(\hat{\boldsymbol{\lambda}}-\bar{\boldsymbol{\lambda}}\right)^T\mathbf{B}\left(\mathbf{X}_i\right)\right\|_2^2 = \left(\hat{\boldsymbol{\lambda}}-\bar{\boldsymbol{\lambda}}\right)^T\mathbf{V}\left(\hat{\boldsymbol{\lambda}}-\bar{\boldsymbol{\lambda}}\right)$$

$$\leq C_{\mathbf{V}}\left\|\hat{\boldsymbol{\lambda}}-\bar{\boldsymbol{\lambda}}\right\|^2 = O_{a.s.}\left(N_dH^4+N_dn^{-1}\log^2 n\right).$$

Then $\|\hat{m}-\bar{m}\|_{2,n} = O_{a.s.}\left(N_d^{1/2}H^2+N_d^{1/2}n^{-1/2}\log n\right)$ by Lemma 5.6.2. Next,

$$\begin{aligned}\|\hat{m}-m\|_{2,n} &\leq \|\hat{m}-\bar{m}\|_{2,n}+\|\bar{m}-m\|_{2,n}\\ &= O_{a.s.}\left(N_d^{1/2}H^2+N_d^{1/2}n^{-1/2}\log n\right)+O_{a.s.}\left(H^2\right)\\ &= O_{a.s.}\left(N_d^{1/2}H^2+N_d^{1/2}n^{-1/2}\log n\right).\end{aligned}$$

□

In the following denote

$$\boldsymbol{\omega}(x_1)=\left\{\omega_{J,\alpha}(x_1)\right\}_{J=1,\alpha=2}^{N+1,d},\omega_{J,\alpha}(x_1)=n^{-1}\sum\nolimits_{i=1}^n\left|B_{J,\alpha}(X_{i\alpha})\right|K_h(X_{i1}-x_1).$$

LEMMA 5.6.10. *Under Assumptions (A1) to (A3), (A5), and (A7) to (A8), , as* $n\to\infty$

$$\sup_{x_1\in[0,1]}\left|\boldsymbol{\omega}(x_1)\right| = O_{a.s.}\left(H^{1/2}\right). \tag{5.6.21}$$

Proof. First, one computes

$$\begin{aligned}E\omega_{J,\alpha}(x_1) &= \int\int K_h(u_1-x_1)\left|B_{J,\alpha}(u_\alpha)\right|f\left(u_1,u_\alpha\right)du_1du_\alpha\\ &= \int\int K(v_1)\frac{\left|b_{J,\alpha}(u_2)\right|}{\left\|b_{J,\alpha}\right\|_2}f(hv_1+x_1,u_\alpha)\,dv_1du_\alpha\\ &= \left(\left\|b_{J,\alpha}\right\|_2\right)^{-1}\left\{\int\int K(v_1)I_{J+1,2}(u_2)f(hv_1+x_1,u_2)\,dv_1du_2\right.\\ &\quad\left.+\left(\frac{c_{J+1,2}}{c_{J,2}}\right)^{1/2}\int\int K(v_1)I_{J,2}(u_2)f(hv_1+x_1,u_2)\,dv_1du_2\right\}.\end{aligned}$$

$$\leq \left\|b_{J,\alpha}\right\|_2^{-1}\left\{\int\int\left|K\left(v_1\right)b_J\left(u_\alpha\right)\right|f\left(x_1+hv_1,u_\alpha\right)dv_1du_\alpha\right.$$
$$\left.+\frac{c_{J,\alpha}}{c_{J-1,\alpha}}\int\int\left|K\left(v_1\right)b_{J-1}\left(u_\alpha\right)\right|f\left(x_1+hv_1,u_\alpha\right)dv_1du_\alpha\right\}.$$

The boundedness of the joint density f and the Lipschitz continuity of the kernel K will imply that there exist constant c_2 such that

$$\int\int\left|K\left(v_1\right)b_J\left(u_\alpha\right)\right|f\left(x_1+hv_1,u_\alpha\right)dv_1du_\alpha\leq C_Kc_2H,$$
$$\int\int\left|K\left(v_1\right)b_{J-1}\left(u_\alpha\right)\right|f\left(x_1+hv_1,u_\alpha\right)dv_1du_\alpha\leq C_Kc_2H.$$

and therefore

$$\sup_{x_1\in[0,1]}\left|E\omega\left(x_1\right)\right|=O\left(H^{1/2}\right) \tag{5.6.22}$$

by Lemma 4.7.2. Similarly, $E\omega_{J,\alpha}\left(x_1\right)^r\sim h^{1-r}H^{1-r/2}$, hence $E\omega_{J,\alpha}\left(x_1\right)^2\sim h^{-1}$. According to Lemma 2.5.2 and similar proof of Lemma A.5 in [68], one has

$$\sup_{x_1\in[0,1]}\sup_{1\leq J\leq N+1}\sup_{2\leq\alpha\leq d}\left|\omega_{J,\alpha}\left(x_1\right)-E\omega_{J,\alpha}\left(x_1\right)\right|=O_{a.s.}\left(\log n/\sqrt{nh}\right).$$

Combining with (5.6.22), the lemma is proved. □

LEMMA 5.6.11. *Under Assumptions (A1)-(A7),*

$$\sup_{x_1\in[h,1-h]}\left|\hat{l}'\left(\tilde{m}_{K,1}\left(x_1\right)\right)\right|\rightarrow o_{a.s.}\left(n^{-2/5}\right).$$

Proof. Note $\tilde{l}'\left(\tilde{m}_{K,1}\left(x_1\right)\right)=0$, one has

$$\hat{l}'\left(\tilde{m}_{K,1}\left(x_1\right)\right)=\hat{l}'\left(\tilde{m}_{K,1}\left(x_1\right)\right)-\tilde{l}'\left(\tilde{m}_{K,1}\left(x_1\right)\right)$$

$$=\ 1/n\sum\nolimits_{i=1}^n\left[b'\left\{\tilde{m}_{K,1}\left(x_1\right)+m_{\_1}\left(\mathbf{X}_{i\_1}\right)\right\}-b'\left\{\tilde{m}_{K,1}\left(x_1\right)+\hat{m}_{\_1}\left(\mathbf{X}_{i\_1}\right)\right\}\right]$$
$$K_h\left(X_{i1}-x_1\right)$$

$$=1/n\sum\nolimits_{i=1}^n b''\left\{\tilde{m}_{K,1}\left(x_1\right)+m_{\_1}\left(\mathbf{X}_{i\_1}\right)\right\}\left\{m_{\_1}\left(\mathbf{X}_{i\_1}\right)-\hat{m}_{\_1}\left(\mathbf{X}_{i\_1}\right)\right\}\times$$
$$K_h\left(X_{i1}-x_1\right)+O\left[1/n\sum\nolimits_{i=1}^n\left\{m_{\_1}\left(\mathbf{X}_{i\_1}\right)-\hat{m}_{\_1}\left(\mathbf{X}_{i\_1}\right)\right\}^2\right]=I+O_{a.s.}\left(H^3\right)$$

where $I=I_1+I_2$,

$$I_1=1/n\sum\nolimits_{i=1}^n b''\left\{\tilde{m}_{K,1}\left(x_1\right)+m_{\_1}\left(\mathbf{X}_{i\_1}\right)\right\}\left\{m_{\_1}\left(\mathbf{X}_{i\_1}\right)-\bar{m}_{\_1}\left(\mathbf{X}_{i\_1}\right)\right\}K_h\left(X_{i1}-x_1\right),$$

$$I_2 = 1/n \sum_{i=1}^{n} b'' \{\tilde{m}_{K,1}(x_1) + m_{\_1}(\mathbf{X}_{i\_1})\} \{\bar{m}_{\_1}(\mathbf{X}_{i\_1}) - \hat{m}_{\_1}(\mathbf{X}_{i\_1})\} K_h(X_{i1} - x_1).$$

According to (3.6.1), $I_1 = O_{a.s.}(H^2)$, while

$$\begin{aligned} I_2 &= n^{-1} \sum_{i=1}^{n} b'' \{\tilde{m}_{K,1}(x_1) + m_{\_1}(\mathbf{X}_{i\_1})\} \times \\ &\quad \left\{ \left(\hat{\lambda}_0 - \bar{\lambda}_0\right) + \sum_{1 \le J \le N+1, 2 \le \alpha \le d} \left(\hat{\lambda}_{J,\alpha} - \bar{\lambda}_{J,\alpha}\right) B_{J,\alpha}(X_{i\alpha}) \right\} K_h(X_{i1} - x_1) \\ &= I_{2,b} + I_{2,v} + I_{2,r} \end{aligned}$$

where

$$\begin{aligned} I_{2,b} &= n^{-1} \sum_{i=1}^{n} b'' \{\tilde{m}_{K,1}(x_1) + m_{\_1}(\mathbf{X}_{i\_1})\} \times \\ &\quad \left\{ \Phi_{b,0} + \sum_{1 \le J \le N+1, 2 \le \alpha \le d} \Phi_{b,J,\alpha} B_{J,\alpha}(X_{i\alpha}) \right\} K_h(X_{i1} - x_1), \end{aligned}$$

$$\begin{aligned} I_{2,v} &= n^{-1} \sum_{i=1}^{n} b'' \{\tilde{m}_{K,1}(x_1) + m_{\_1}(\mathbf{X}_{i\_1})\} \times \\ &\quad \left\{ \Phi_{v,0} + \sum_{1 \le J \le N+1, 2 \le \alpha \le d} \Phi_{v,J,\alpha} B_{J,\alpha}(X_{i\alpha}) \right\} K_h(X_{i1} - x_1), \end{aligned}$$

$$\begin{aligned} I_{2,r} &= n^{-1} \sum_{i=1}^{n} b'' \{\tilde{m}_{K,1}(x_1) + m_{\_1}(\mathbf{X}_{i\_1})\} \times \\ &\quad \left\{ \Phi_{r,0} + \sum_{1 \le J \le N+1, 2 \le \alpha \le d} \Phi_{r,J,\alpha} B_{J,\alpha}(X_{i\alpha}) \right\} K_h(X_{i1} - x_1) \end{aligned}$$

where $\Phi_{b,0}, \Phi_{b,0}, \Phi_{b,0}, \Phi_{b,J,\alpha}, \Phi_{b,J,\alpha}, \Phi_{b,J,\alpha}$ are the corresponding elements in the vectors Φ_b, Φ_v and Φ_r defined as (5.6.16), (5.6.17) and (5.6.18).

$$\begin{aligned} |I_{2,b}| &\le C_b n^{-1} \sum_{i=1}^{n} \left\{ |\Phi_{b,0}| + \sum_{1 \le J \le N+1, 2 \le \alpha \le d} |\Phi_{b,J,\alpha}| \, |B_{J,\alpha}(X_{i\alpha})| \right\} K_h(X_{i1} - x_1) \\ &\le C_b \left[\left\{ \Phi_{b,0}^2 + \sum_{1 \le J \le N+1, 2 \le \alpha \le d} \Phi_{b,J,\alpha}^2 \right\} \right]^{1/2} \times \left[\left\{ n^{-1} \sum_{i=1}^{n} K_h(X_{i1} - x_1) \right\}^2 \right. \\ &\quad \left. + \sum_{1 \le J \le N+1, 2 \le \alpha \le d} \left\{ n^{-1} \sum_{i=1}^{n} |B_{J,\alpha}(X_{i\alpha})| \, K_h(X_{i1} - x_1) \right\}^2 \right]^{1/2} \\ &= C_b \times O_{a.s.}\left(N_d^{1/2} H^{5/2}\right) \times [O_{a.s.}(1) + (N+1) \times (d-1) \times O_{a.s.}(H)] \\ &= O_{a.s.}\left(N_d^{1/2} H^{5/2}\right). \end{aligned}$$

according to (5.6.19) and (5.6.21), similarly

$$|I_{2,r}| = O_{a.s.}\left(N_d H^{7/2} + N_d H^{-1/2} n^{-1} \log n\right).$$

$$I_{2,v} = \widetilde{I}_{2,v} + O_{a.s.}\left(n^{-2/5}\log n\right) \times O_{a.s.}\left(N_d^{1/2} n^{-1/2}\log n\right).$$

where

$$\begin{aligned}
\widetilde{I}_{2,v} &= n^{-1}\sum_{i=1}^{n} b''\{m(\mathbf{X}_i)\}\left\{\Phi_{v,0} + \sum_{1\le J\le N+1, 2\le\alpha\le d}\Phi_{v,J,\alpha}B_{J,\alpha}(X_{i\alpha})\right\} \\
&\quad K_h(X_{i1}-x_1) \\
&= \Phi_{v,0}n^{-1}\sum_{i=1}^{n} b''\{m(\mathbf{X}_i)\}K_h(X_{i1}-x_1) + \\
&\quad \sum_{1\le J\le N+1, 2\le\alpha\le d}\Phi_{v,J,\alpha}n^{-1}\sum_{i=1}^{n} b''\{m(\mathbf{X}_i)\}B_{J,\alpha}(X_{i\alpha})K_h(X_{i1}-x_1) \\
&= \Phi_{v,0}Eb''\{m(\mathbf{X})\}K_h(X_1-x_1) \\
&\quad + \sum_{1\le J\le N+1, 2\le\alpha\le d}\Phi_{v,J,\alpha}Eb''\{m(\mathbf{X})\}B_{J,\alpha}(X_\alpha)K_h(X_1-x_1) \\
&\quad + O_{a.s.}\left(N_d^{1/2}n^{-1/2}\log n\right) \times N_d^{1/2} \times O_{a.s.}\left(n^{-2/5}\log n\right) \\
&= \widetilde{I}_{2,v,1} + \widetilde{I}_{2,v,2} + O_{a.s.}\left(N_d n^{-9/10}\log^2 n\right)
\end{aligned}$$

Where

$$\begin{aligned}
\widetilde{I}_{2,v,1} &= \Phi_{v,0}Eb''\{m(\mathbf{X})\}K_h(X_1-x_1) \\
&= Eb''\{m(\mathbf{X})\}K_h(X_1-x_1)\frac{1}{n}\sum_{i=1}^{n}\sigma(\mathbf{X}_i)\varepsilon_i \\
&\quad \left\{S_{0,0} + \sum_{1\le J\le N+1, 1\le\alpha\le d}S_{J,\alpha}B_{J,\alpha}(X_{i\alpha})\right\},
\end{aligned}$$

$$\widetilde{I}_{2,v,2} = \sum_{1\le J\le N+1, 2\le\alpha\le d}\Phi_{v,J,\alpha}Eb''\{m(\mathbf{X})\}B_{J,\alpha}(X_\alpha)K_h(X_1-x_1)$$

$$= \left\{Eb''\{m(\mathbf{X})\}B_{J,\alpha}(X_\alpha)K_h(X_1-x_1)\right\}_{1\le J\le N+1, 2\le\alpha\le d} \times \left(\Phi_{v,J,\alpha}\right)^T_{1\le J\le N+1, 2\le\alpha\le d}$$

$$\begin{aligned}
&= \sum_{1\le J\le N+1, 2\le\alpha\le d}Eb''\{m(\mathbf{X})\}B_{J,\alpha}(X_\alpha)K_h(X_1-x_1) \\
&\quad \frac{1}{n}\sum_{i=1}^{n}\sigma(\mathbf{X}_i)\varepsilon_i\left\{S_{J,\alpha} + \sum_{1\le J'\le N+1, 1\le\alpha'\le d}S_{J,\alpha,J',\alpha'}B_{J',\alpha'}(X_{i\alpha'})\right\} \\
&= \frac{1}{n}\sum_{i=1}^{n}\sigma(\mathbf{X}_i)\varepsilon_i\sum_{1\le J\le N+1, 2\le\alpha\le d}\mu_{b,k,J,\alpha}(x_1) \\
&\quad \left\{S_{J,\alpha} + \sum_{1\le J'\le N+1, 1\le\alpha'\le d}S_{J,\alpha,J',\alpha'}B_{J',\alpha'}(X_{i\alpha'})\right\}
\end{aligned}$$

where $S_{0,0}, S_{J,\alpha}, S_{J,\alpha,J',\alpha'}$ are the corresponding element in the matrix $\mathbf{S}_b$ defined in (5.6.13) has the form shown in the proof of Theorem 5.3.4 and $\mu_{b,k,J,\alpha}(x_1) = Eb''\{m(\mathbf{X})\} B_{J,\alpha}(X_\alpha) K_h(X_1 - x_1)$, which has the order $O_{a.s.}\left(H^{1/2}\right)$. Denote $D_n = n^{\theta_0}\left(\frac{1}{2+\eta} < \theta_0 < \frac{2}{5}\right)$, $\varepsilon_{i,1}^{D_n} = \varepsilon_i I\{|\varepsilon_i| > D_n\}$, $\varepsilon_{i,3}^{D_n} = E\varepsilon_i I\{|\varepsilon_i| \leq D_n\}$, $\varepsilon_{i,2}^{D_n} = \varepsilon_i I\{|\varepsilon_i| \leq D_n\} - \varepsilon_{i,3}^{D_n}$. Then $\widetilde{I}_{2,v,2} = \Lambda_1 + \Lambda_2 + \Lambda_3$ where

$$\Lambda_k = \sum\nolimits_{1 \leq J \leq N+1, 2 \leq \alpha \leq d} \mu_{b,k,J,\alpha}(x_1)(x_1) \frac{1}{n} \sum_{i=1}^{n} \sigma(\mathbf{X}_i) \varepsilon_{i,k}^{D_n}$$

$$\left\{S_{J,\alpha} + \sum\nolimits_{1 \leq J' \leq N+1, 1 \leq \alpha' \leq d} S_{J,\alpha,J',\alpha'} B_{J',\alpha'}(X_{i\alpha'})\right\}, k = 1, 2, 3.$$

Then one has with probability 1, $\Lambda_1 = 0$ for large n. Next,

$$\left|\varepsilon_{i,3}^{D_n}\right| = \left|-E\varepsilon_i I\{|\varepsilon_i| > D_n\}\right| \leq \frac{E|\varepsilon_i|^{2+\eta}}{D_n^{1+\eta}} = O\left(D_n^{-(1+\eta)}\right),$$

$$\begin{aligned}
\Lambda_3 &\leq C_{\mathbf{S}_b}\left[\sum\nolimits_{1 \leq J \leq N+1, 2 \leq \alpha \leq d} \mu_{b,k,J,\alpha}^2(x_1)\right. \\
&\quad \left.\sum\nolimits_{1 \leq J' \leq N+1, 1 \leq \alpha' \leq d}\left\{n^{-1}\sum\nolimits_{i=1}^{n} B_{J',\alpha'}(X_{i\alpha'}) \sigma(\mathbf{X}_i) \varepsilon_{i,3}^{D_n}\right\}^2\right]^{1/2} \\
&\leq CD_n^{-(1+\eta)}\left[\sum\nolimits_{1 \leq J \leq N+1, 2 \leq \alpha \leq d} \mu_{b,k,J,\alpha}^2(x_1)\right. \\
&\quad \left.\sum\nolimits_{1 \leq J' \leq N+1, 1 \leq \alpha' \leq d}\left\{n^{-1}\sum\nolimits_{i=1}^{n} B_{J',\alpha'}(X_{i\alpha'}) \sigma(\mathbf{X}_i)\right\}^2\right]^{1/2} \\
&= D_n^{-(1+\eta)} O_{a.s.}\left\{\left(NHN\log^2 n/n\right)^{1/2}\right\} = D_n^{-(1+\eta)} O_{a.s.}\left\{\left(N\log^2 n/n\right)^{1/2}\right\} \\
&= o_{a.s.}\left(n^{-2/5}\right).
\end{aligned}$$

Lastly, $\Lambda_2 = O_{a.s.}\left(n^{-3/5}H^{-1/2}\log n\right) = o_{a.s.}\left(n^{-2/5}\right)$ according to Bernstein's Inequality. Then $\widetilde{I}_{2,v,2} = o_{a.s.}\left(n^{-2/5}\right)$ according to the orders of Λ_1, Λ_2 and Λ_3. With similar proof, we can show $\widetilde{I}_{2,v,1} = o_{a.s.}\left(n^{-2/5}\right)$.

Lastly, denote $\Lambda_2 = n^{-1}\sum_{i=1}^{n} \xi_i$, where

$$\begin{aligned}
\xi_i &= \sum\nolimits_{1 \leq J \leq N+1, 2 \leq \alpha \leq d} \mu_{b,k,J,\alpha}(x_1) \sigma(\mathbf{X}_i) \varepsilon_{i,2}^{D_n} \\
&\quad \left\{S_{J,\alpha} + \sum\nolimits_{1 \leq J' \leq N+1, 1 \leq \alpha' \leq d} S_{J,\alpha,J',\alpha'} B_{J',\alpha'}(X_{i\alpha'})\right\}.
\end{aligned}$$

Then $E\xi_i = 0$, and

$$\begin{aligned}\operatorname{var}(\xi_i) &= \boldsymbol{\mu}_{b\_c}^T \mathbf{S}_{b\_1} \operatorname{var}\left(\left\{\begin{array}{c}\sigma(\mathbf{X}_i)\,\varepsilon_{i,k}^{D_n} \\ B_{J',\alpha'}(X_{i\alpha'})\,\sigma(\mathbf{X}_i)\,\varepsilon_{i,k}^{D_n}\end{array}\right\}_{J'=1,\alpha'=1}^{N+1,d}\right) \mathbf{S}_{b\_1}^T \boldsymbol{\mu}_{b\_c} \\ &\le C_\sigma C_{\mathbf{S}}^2 C_{\mathbf{V}} \boldsymbol{\mu}_{b,k\_1}^T \boldsymbol{\mu}_{b,k\_c}^T = O(1).\end{aligned}$$

Then $\Lambda_2 = O_{a.s.}\left(n^{-1/2}\log n\right) = o_{a.s.}\left(n^{-2/5}\right)$ according to Bernstein's Inequality. Then $\widetilde{I}_{2,v,2} = o_{a.s.}\left(n^{-2/5}\right)$ according to the orders of Λ_1, Λ_2 and Λ_3. With similar proof, we can show $\widetilde{I}_{2,v,1} = o_{a.s.}\left(n^{-2/5}\right)$. Then the lemma is proved. □

Lemma 5.6.12. *Under Assumptions (A1)-(A7),* $\forall a$, $c \le \sup_{x_1\in[h,1-h]}\left|-\hat{l}''(a)\right| \le C$ *a.s. for some constants* $0 < c < C$.

Proof. According to (5.3.2), one has

$$\hat{l}''(a) = -1/n\sum_{i=1}^{n}\left[b''\{a + \hat{m}_{\_1}(\mathbf{X}_{i\_1})\}\right] K_h(X_{i1} - x_1).$$

$c_b \le b''\{a + \hat{m}_{\_1}(\mathbf{X}_{i\_1})\} \le C_b$ and $\sup_{x_1\in[h,1-h]}\left|1/n\sum_{i=1}^{n} K_h(X_{i1} - x_1) - f(x_1)\right| = O_{a.s.}\left\{(nh)^{-1/2}\log n\right\}$ imply the lemma. □

Table 1. Simulated example 2.4.1

	n	$\bar{D}_n\left(\hat{F}\right)$	$\bar{D}_n\left(\tilde{F}\right) - \bar{D}_n\left(\hat{F}\right)$	MISE $\left(\hat{F}\right)$	MISE $\left(\tilde{F}\right)$ − MISE $\left(\hat{F}\right)$
	50	0.101	0.055	0.157	0.021
$\rho = 0$,	100	0.073	0.035	0.072	0.010
$a = 0$.	200	0.051	0.022	0.033	0.004
	500	0.034	0.012	0.014	0.001
	50	0.107	0.051	0.201	0.032
$\rho = 0.5$,	100	0.075	0.034	0.088	0.015
$a = 0.2$.	200	0.052	0.022	0.041	0.004
	500	0.037	0.011	0.019	0.002
	50	0.106	0.035	0.202	0.035
$\rho = 0.9$,	100	0.073	0.024	0.086	0.014
$a = 0.2$.	200	0.050	0.015	0.040	0.006
	500	0.036	0.008	0.020	0.002

Note: $\bar{D}_n$ and MISE of $\hat{F}$ and $\tilde{F}$.

Table 2. Simulated example 3.4.1

Estimation	$n = 400$	$n = 800$	$n = 1600$	$n = 3200$
$\hat{\alpha}$	0.036325	0.023289	0.013743	0.008098

Note: The mean of squared errors for 100 replications.

Table 3. Simulated example 3.4.1

n	400	800	1600	3200
Spline estimation	4	11	31	92
Local linear estimation	102	630	3200	18000
Time ratio	1 : 25	1 : 57	1 : 103	1 : 196

Note: Computing time (in seconds) of cubic spline estimation and local linear estimation of parameter α_0 for one replication with $n = 400$, 800, 1600, 3200. PC with Intel Pentium IV 1.86 GHz processor and 1.0 GB RAM.

Table 4. Fitting DEM/GBP returns

Fitted Model	Log-Likelihood	Volatility Prediction Error
GARCH(1,1)	0.5231	0.1045
GJR	0.5233	0.1039
Semi. GARCH(Kernel)	0.5306	0.0994
Semi. GARCH(Spline)	0.5786	0.0987

Table 5. Fitting DEM/USD returns

Fitted Model	Log-Likelihood	Volatility Prediction Error
GARCH(1,1)	-0.1567	0.6667
GJR	-0.1566	0.6661
Semi. GARCH(Kernel)	-0.1508	0.6529
Semi. GARCH(Spline)	-0.1485	0.6476

Table 6. Residual check for fitting DEM/GBP returns

ACF up to lag	$\|\hat{\xi}_t\|, Z_t$	$\|\hat{\xi}_t\|^2, Z_t^2$	$\|\hat{\xi}_t\|^3, Z_t^3$	$\|\hat{\xi}_t\|^4, Z_t^4$
100	0.07, 0.09	0.02, 0.06	0.02, 0.05	0.01, 0.05
200	0.045, 0.065	0.01, 0.04	0.01, 0.035	0.005, 0.035
300	0.04, 0.06	0.007, 0.037	0.007, 0.033	0.003, 0.047

Table 7. Residual check for fitting DEM/USD returns

ACF up to lag	$\|\hat{\xi}_t\|, Z_t$	$\|\hat{\xi}_t\|^2, Z_t^2$	$\|\hat{\xi}_t\|^3, Z_t^3$	$\|\hat{\xi}_t\|^4, Z_t^4$
100	0.04, 0.09	0.04, 0.06	0.06, 0.05	0.05, 0.05
200	0.025, 0.065	0.025, 0.04	0.04, 0.035	0.04, 0.035
300	0.0167, 0.06	0.0167, 0.037	0.03, 0.033	0.033, 0.047

Table 8. Simulated example 4.6.1

SBLL fit	$m_{01} = 2$	$m_{02} = 1$	m_{11}	m_{21}	m_{12}	m_{22}
$n = 200$	1.9813(0.1636)	0.9909(0.0539)	0.0255	0.0276	0.0113	0.0097
$n = 500$	1.9964(0.0980)	0.9989(0.0343)	0.0096	0.0089	0.0041	0.0030
Spline fit $p = 1$						
$n = 200$	1.9813(0.1636)	0.9909(0.0539)	0.0561	0.0125	0.0089	0.0085
$n = 500$	1.9964(0.0980)	0.9989(0.0343)	0.0185	0.0063	0.0063	0.0065

Note: the means and standard errors (in parentheses) of $\hat{m}_{01}$, $\hat{m}_{02}$ and the AISEs of $\hat{m}_{\text{SBLL},11}$, $\hat{m}_{\text{SBLL},12}$, $\hat{m}_{\text{SBLL},21}$, $\hat{m}_{\text{SBLL},22}$ by two methods: SBLL and polynomial spline.

Table 9. Simulated example 5.5.1

$d = 5$	n	MISE $(\hat{m}_{\text{SBK},1})$	MISE $(\tilde{m}_{\text{SBK},1})$	$\overline{\text{EFF}}$ $(\hat{m}_{\text{SBK},1})$	std $\{\text{EFF}\ (\hat{m}_{\text{SBK},1})\}$
$\rho = 0$, $a = 0$.	500	0.054	0.060	1.112	0.274
$r = 0.5$, $a = 0.5$.	500	0.101	0.094	1.023	0.279

Note: The MISEs and EFFs of $\hat{m}_{\text{SBK},1}$, $\tilde{m}_{\text{SBK},1}$.

Table 10. Simulated example 5.5.1

$d = 5$	n	MISE $(\hat{m}_{\text{SBK},2})$	MISE $(\tilde{m}_{\text{SBK},2})$	$\overline{\text{EFF}}$ $(\hat{m}_{\text{SBK},2})$	std $\{\text{EFF}\ (\hat{m}_{\text{SBK},2})\}$
$r = 0$, $a = 0$.	500	0.017	0.027	1.503	0.896
$r = 0.5$, $a = 0.5$.	500	0.036	0.417	0.997	0.400

Note: The MISEs and EFFs of $\hat{m}_{\text{SBK},2}$, $\tilde{m}_{\text{SBK},2}$.

Table 11. Simulated example 5.5.2

$d = 10$	n	MISE $(\hat{m}_{\text{SBK},1})$	MISE $(\tilde{m}_{\text{K},1})$	$\overline{\text{EFF}}\ (\hat{m}_{\text{SBK},1})$	std $\{\text{EFF}\ (\hat{m}_{\text{SBK},1})\}$
$r = 0,$ $a = 0.$	500	0.0965	0.0701	0.9868	0.3813
	1000	0.0491	0.0453	1.0228	0.2324
	1500	0.0298	0.0331	1.1021	0.3123
	2000	0.0246	0.0280	1.1014	0.2161
$r = 0,$ $a = 0.5.$	500	0.0992	0.0735	0.9515	0.3154
	1000	0.0453	0.0440	1.0489	0.2741
	1500	0.0285	0.0327	1.0957	0.2306
	2000	0.0259	0.0282	1.0801	0.1823
$r = 0.5,$ $a = 0.$	500	0.2318	0.1373	0.8732	0.3122
	1000	0.1343	0.0885	0.9186	0.4027
	1500	0.0756	0.0605	0.9294	0.2493
	2000	0.0567	0.0474	0.9811	0.2877
$r = 0.5,$ $a = 0.5.$	500	0.2757	0.1386	0.8509	0.3356
	1000	0.1389	0.0899	0.8950	0.2731
	1500	0.0776	0.0601	0.9686	0.2715
	2000	0.0593	0.0485	0.9885	0.3050

Note: The MISEs and EFFs of $\hat{m}_{\text{SBK},1}$, $\tilde{m}_{\text{K},1}$.

Table 12. Simulated example 5.5.2

n	500	1000	1500	2000
$r = 0, a = 0.$	5.6	22	49	86
$r = 0.5, a = 0.5.$	7.2	27	57	102

Note: Computing time of $\hat{m}_{\text{SBK},1}$.

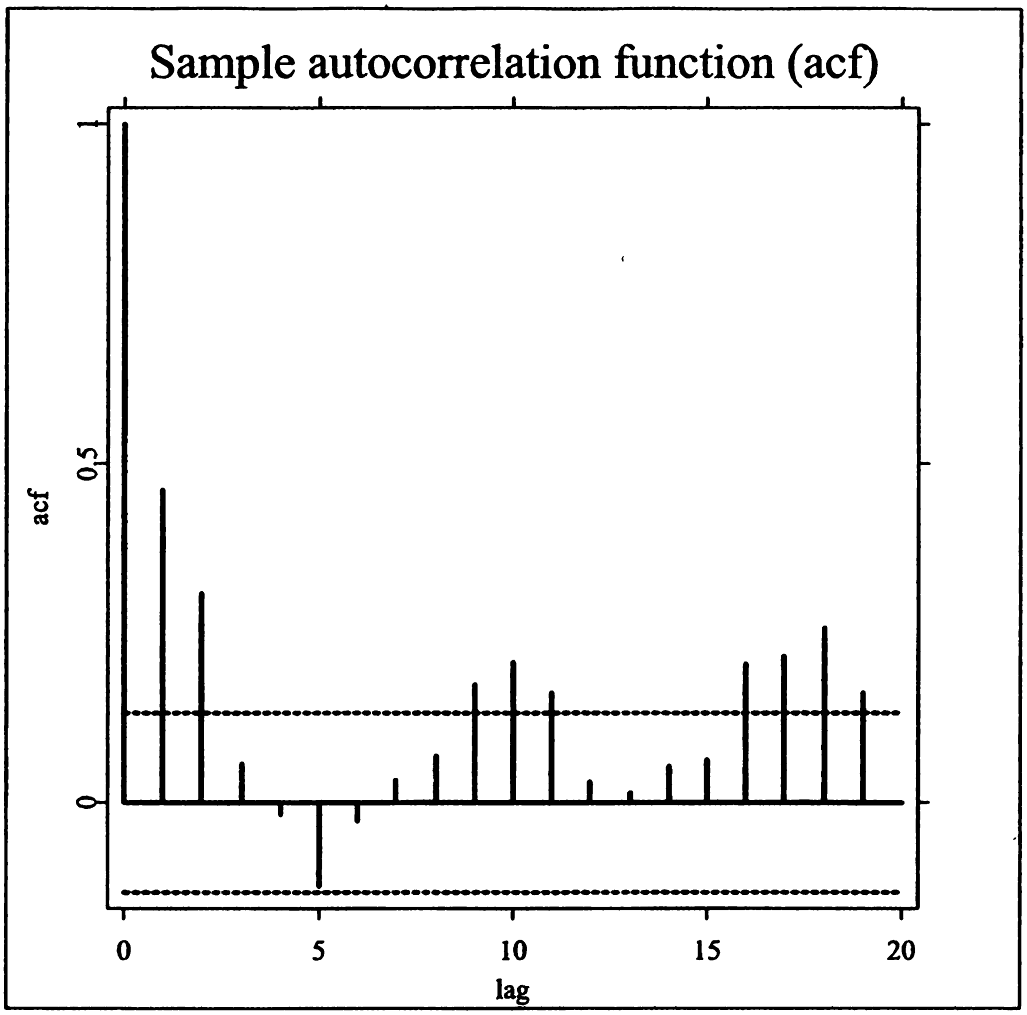

Figure 1. ACF plot of GDP quarterly growth rate.

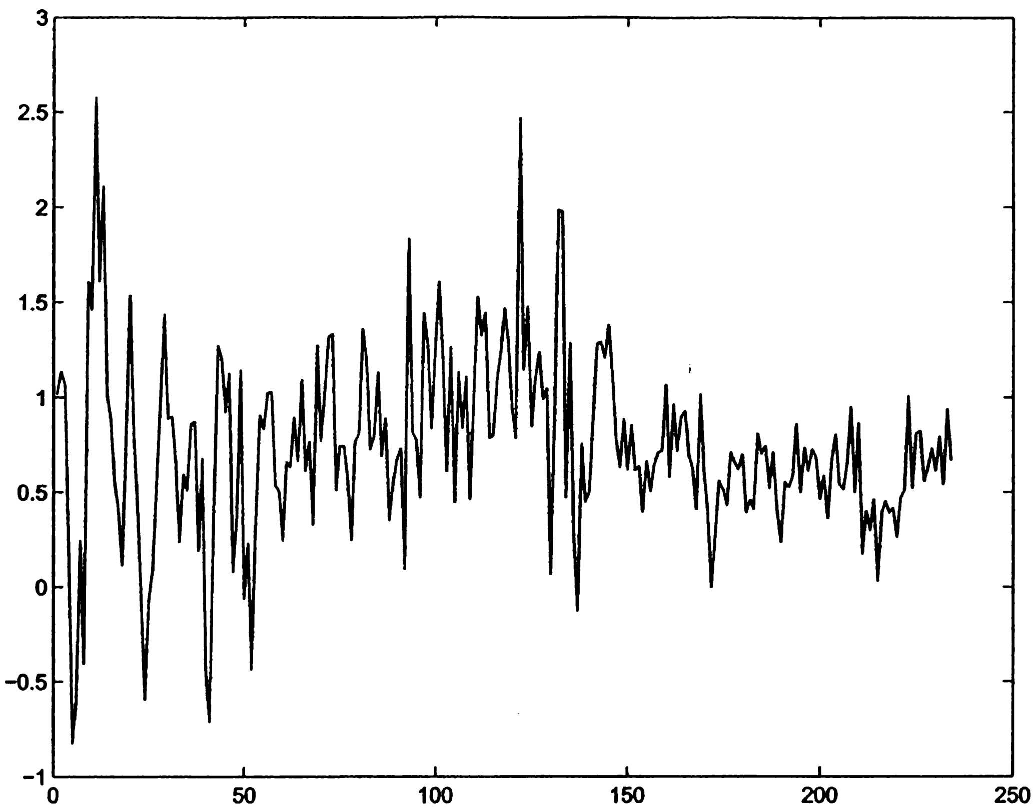

Figure 2. Timeplot of GDP quarterly growth rate.

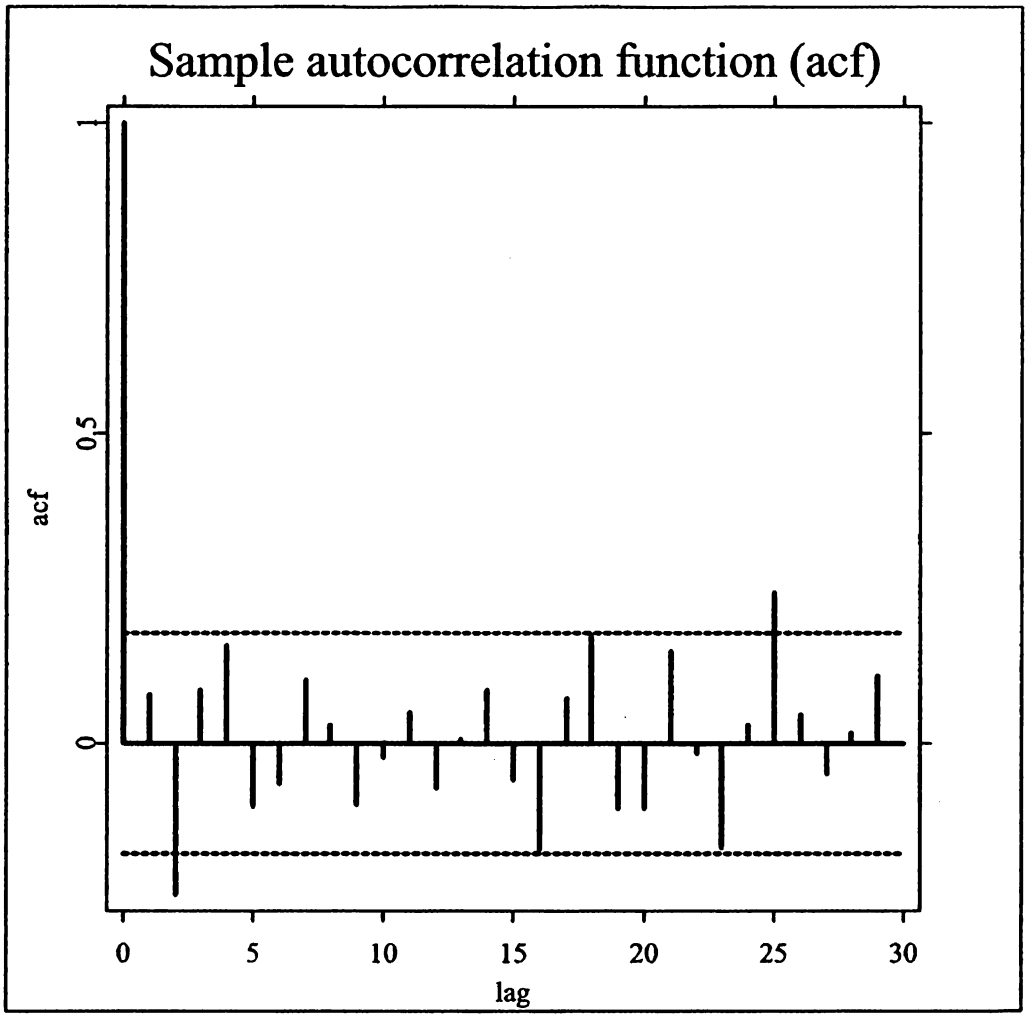

Figure 3. ACF plot of unemployment quarterly growth rate.

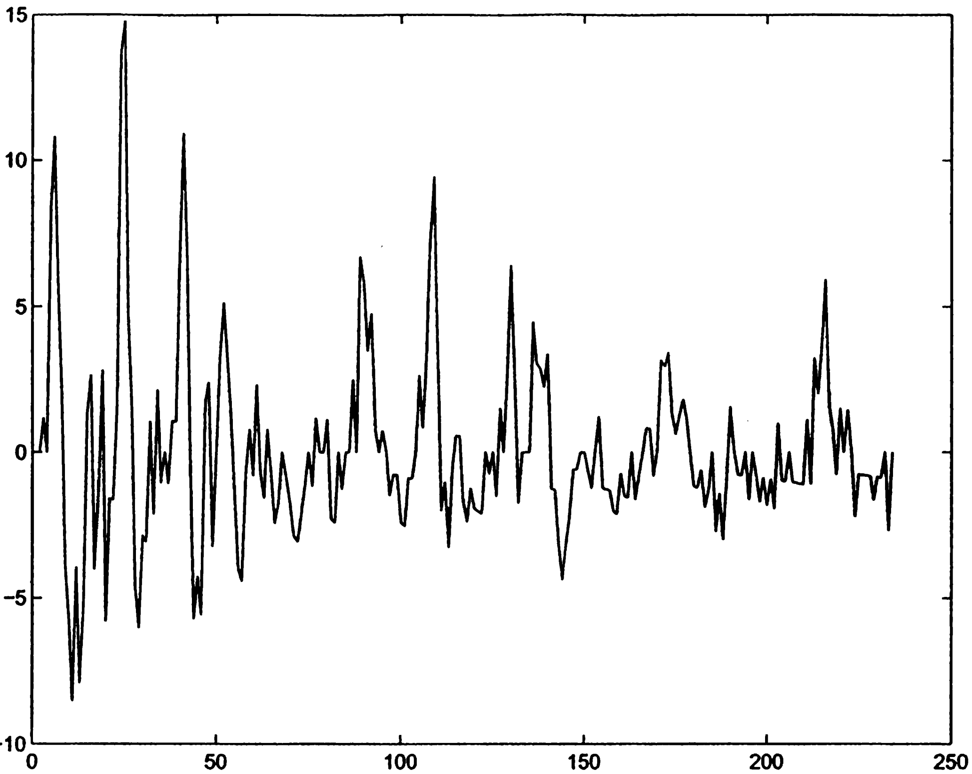

Figure 4. Timeplot of unemployment quarterly growth rate.

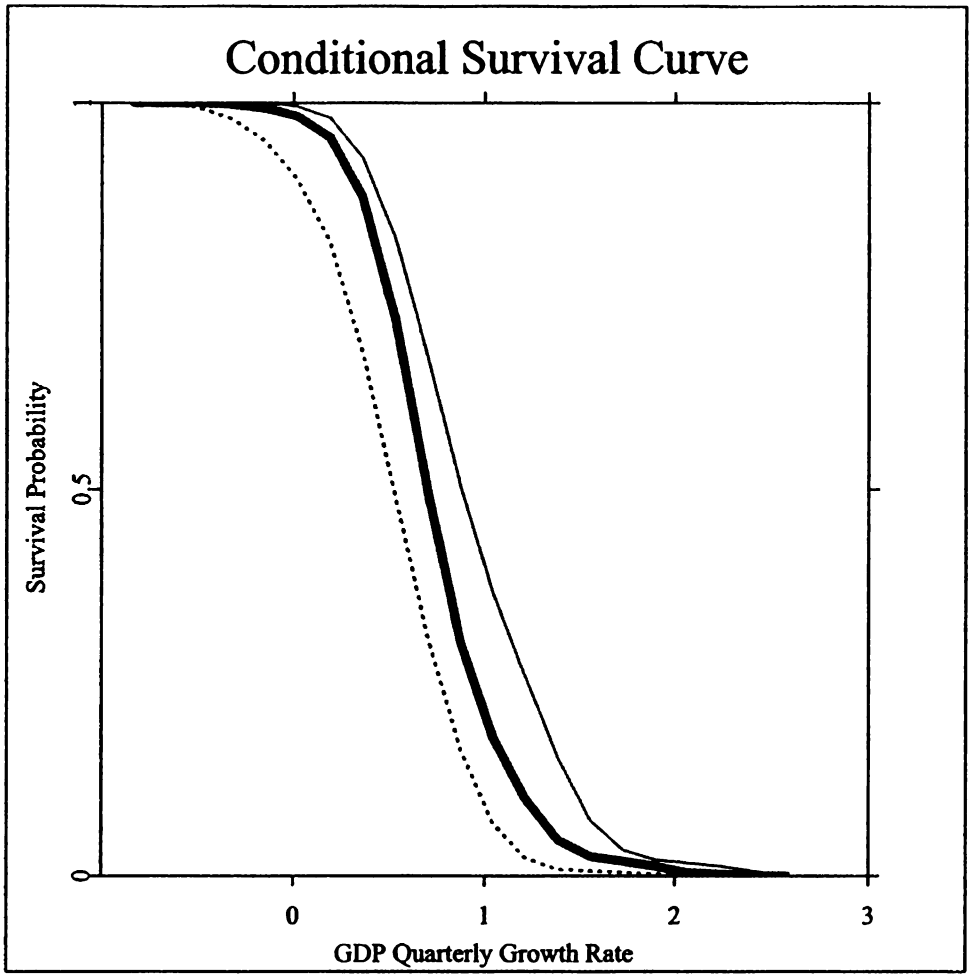

Figure 5. Survival curves of GDP growth rate conditional on unemployment growth rate. Note: $X_{t1} \in [-0.08, -0.04]$, thin solid; $X_{t1} \in [-0.02, 0.02]$, thick solid; $X_{t1} \in [0.04, 0.08]$, dotted.

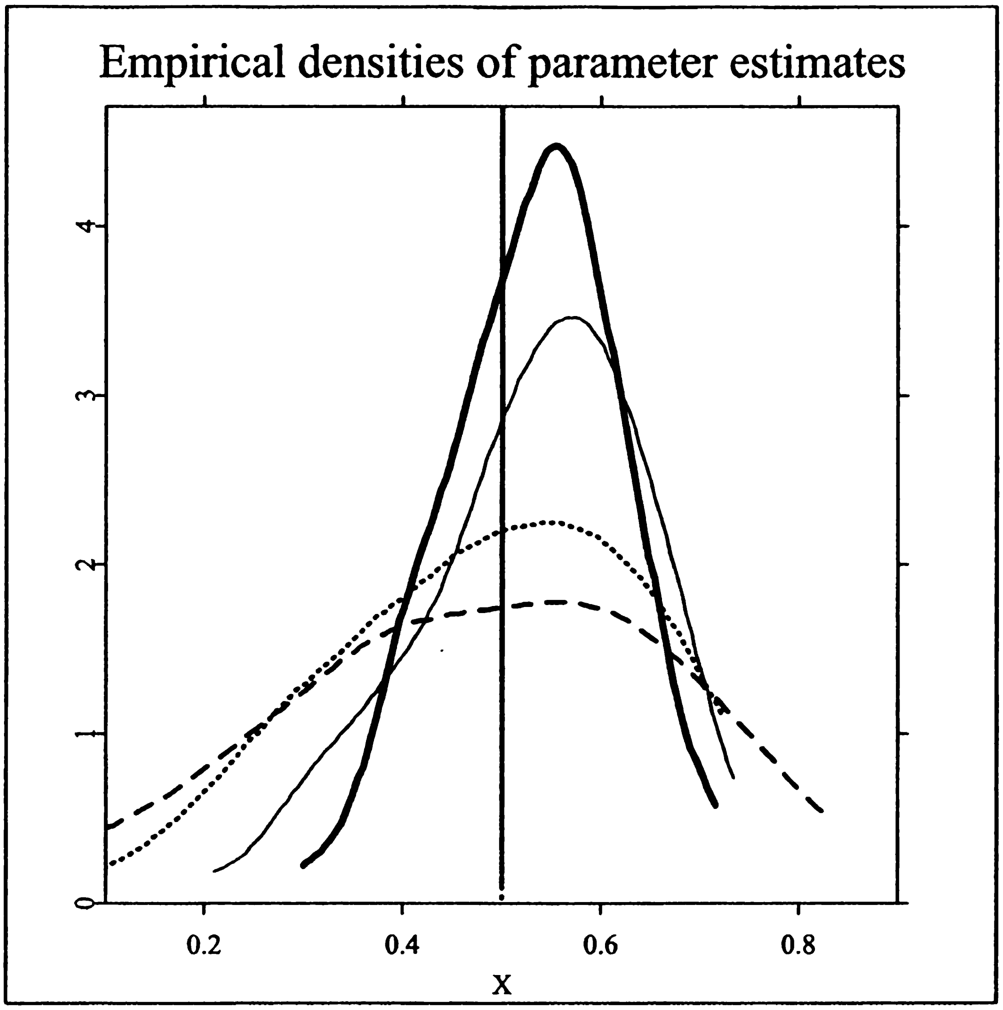

Figure 6. Plot of densities of $\hat{\alpha}$.

Note: $n = 400$ - dashed line, $n = 800$ - dotted line, $n = 1600$ - thin solid line,$n = 3200$ - thick solid line

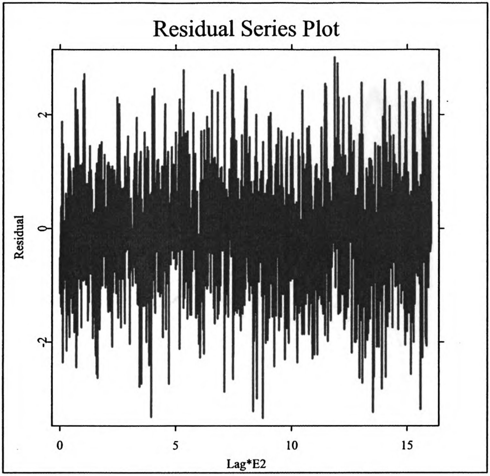

Figure 7. Residuals of DEM/USD daily returns

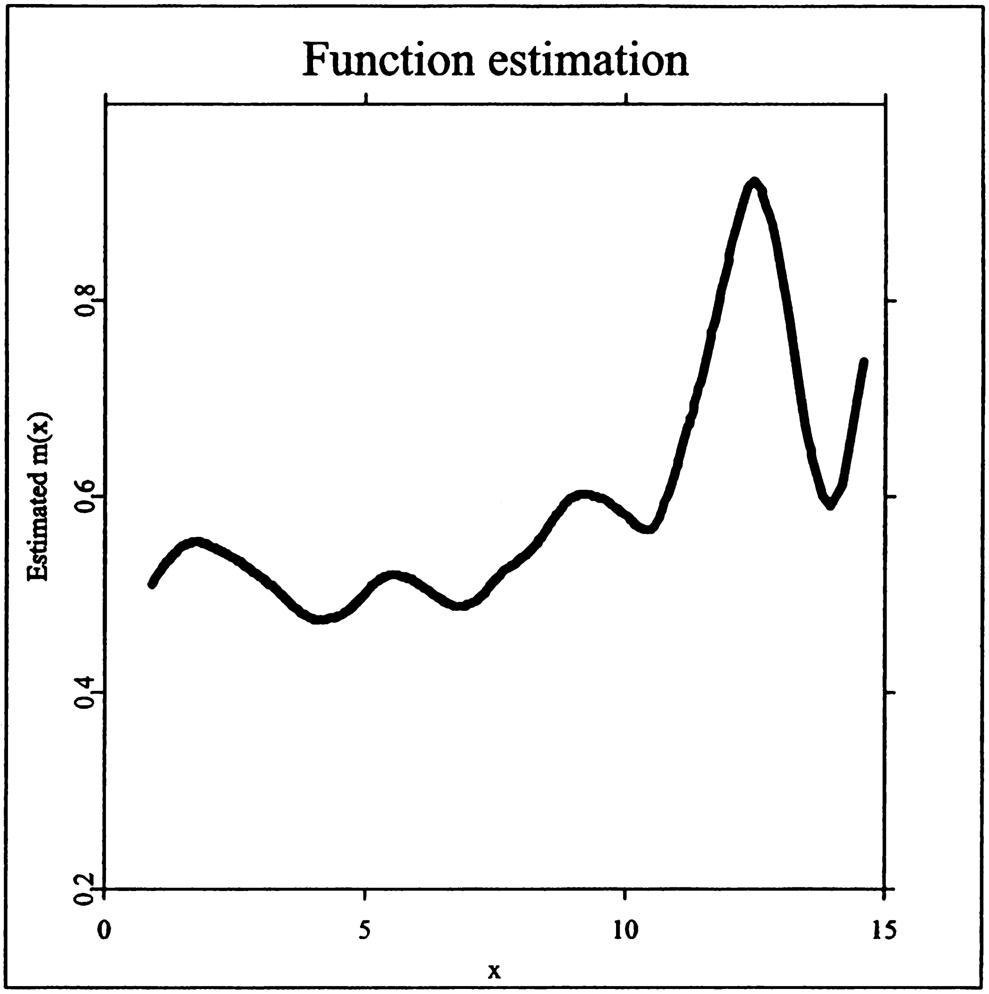

Figure 8. Estimated function m for the semiparametric GARCH model.

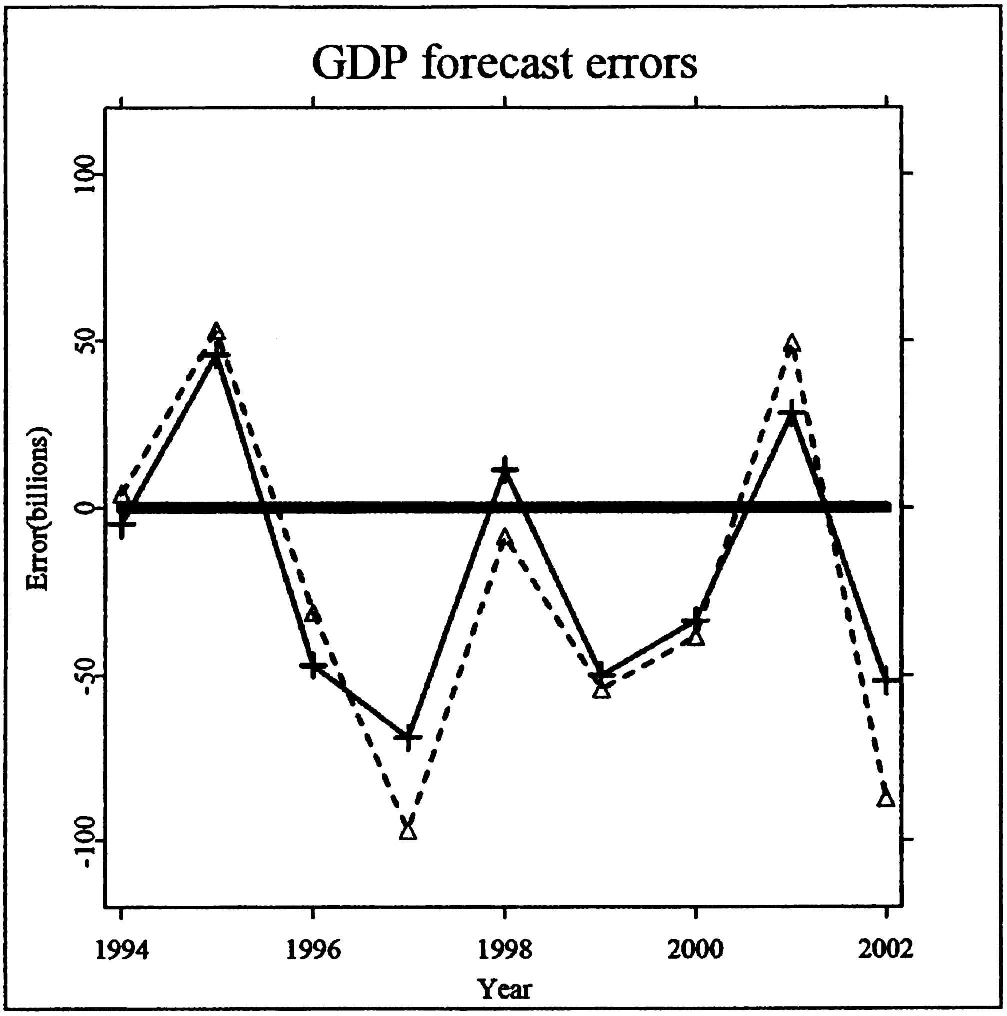

Figure 9. Errors of GDP forecasts.

Note: model (4.6.2)–solid line; model (4.6.1)–dotted line.

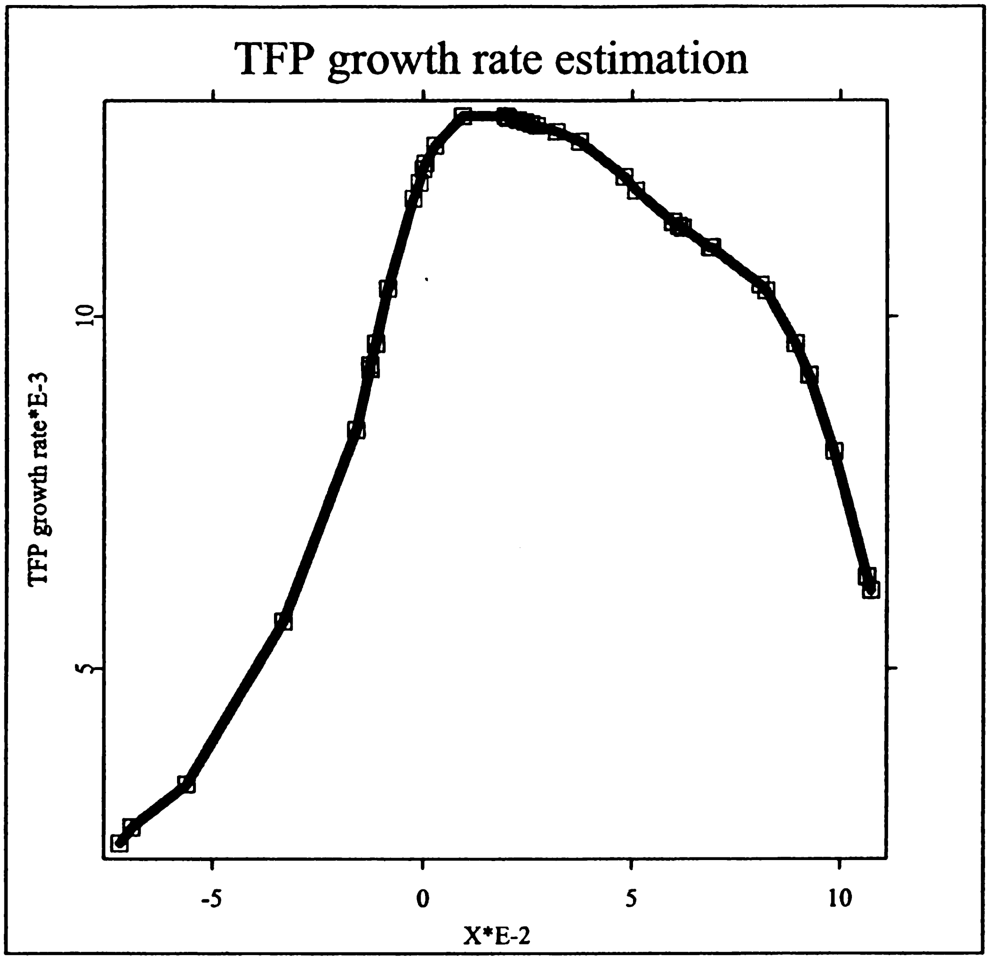

Figure 10. Estimation of function $c_1 + m_{\mathrm{SBLL},41}\,(x_{t-3})$.

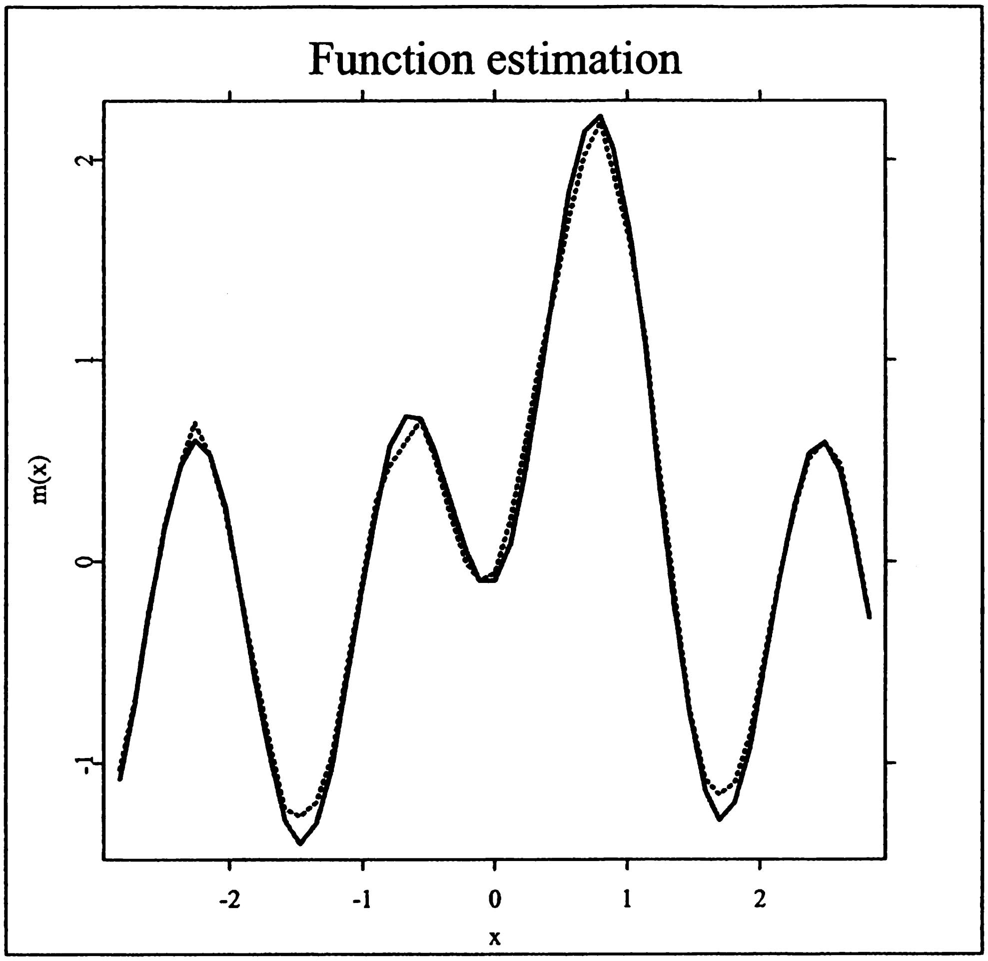

Figure 11. A typical estimator of m_{11} based on $n = 500$ observations.

Note: true function m_{11}–solid line; $\hat{m}_{\text{SBLL},11}$–dotted line.

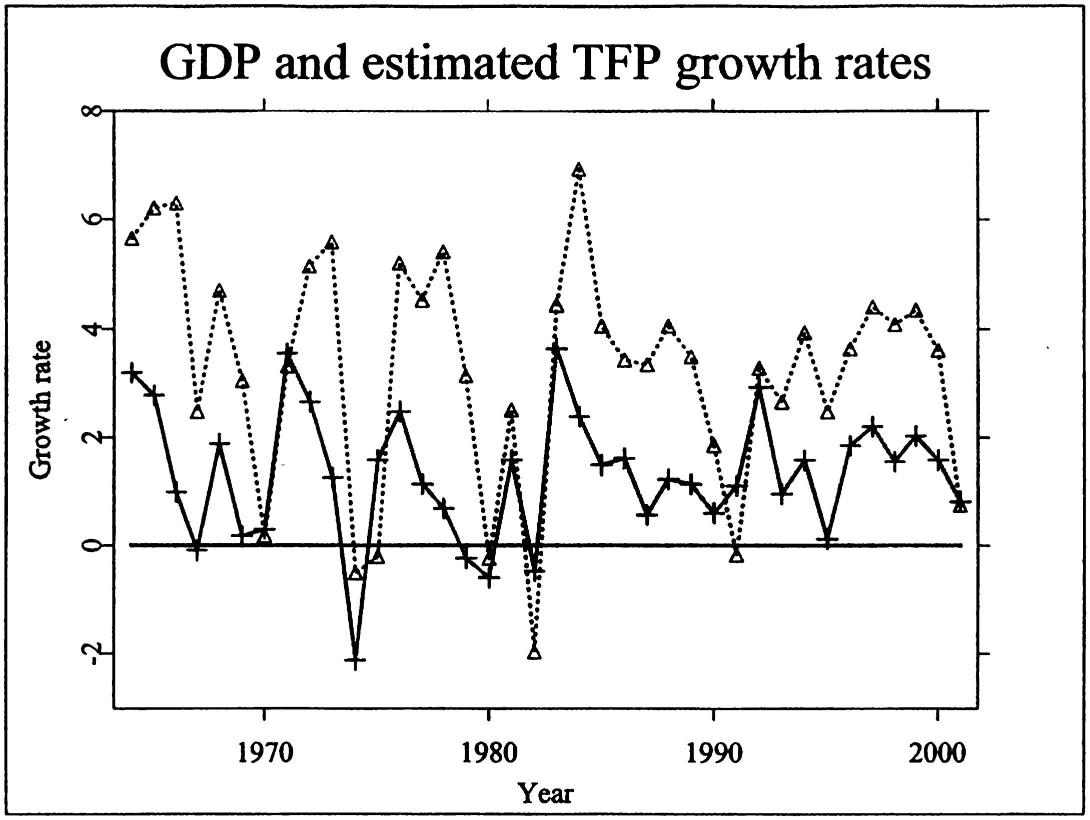

Figure 12. GDP growth rate–dotted line; estimated TFP growth rate–solid line.

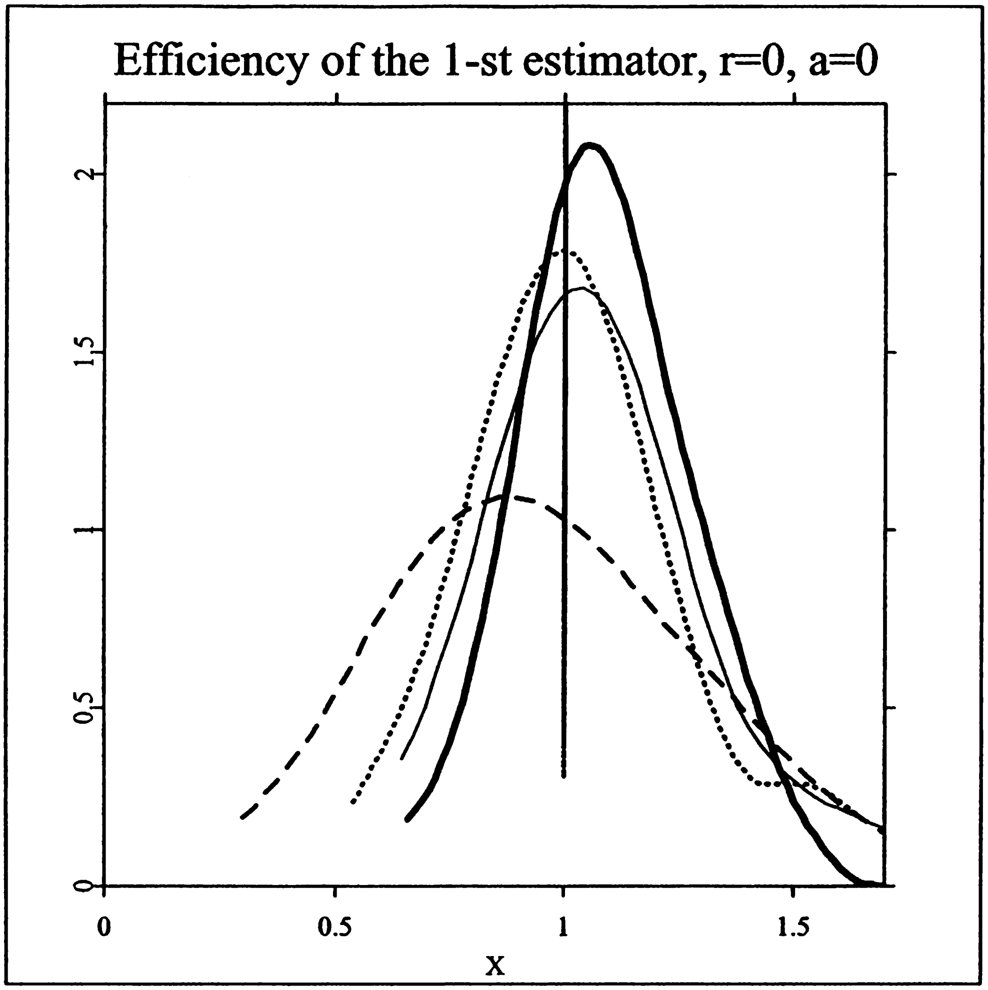

Figure 13. Plot of empirical distribution of relative efficiency: $r = 0, a = 0$.
Note: $n = 500$ - dashed line, $n = 1000$ - dotted line, $n = 1500$ - thin solid line,$n = 2000$ - thick solid line.

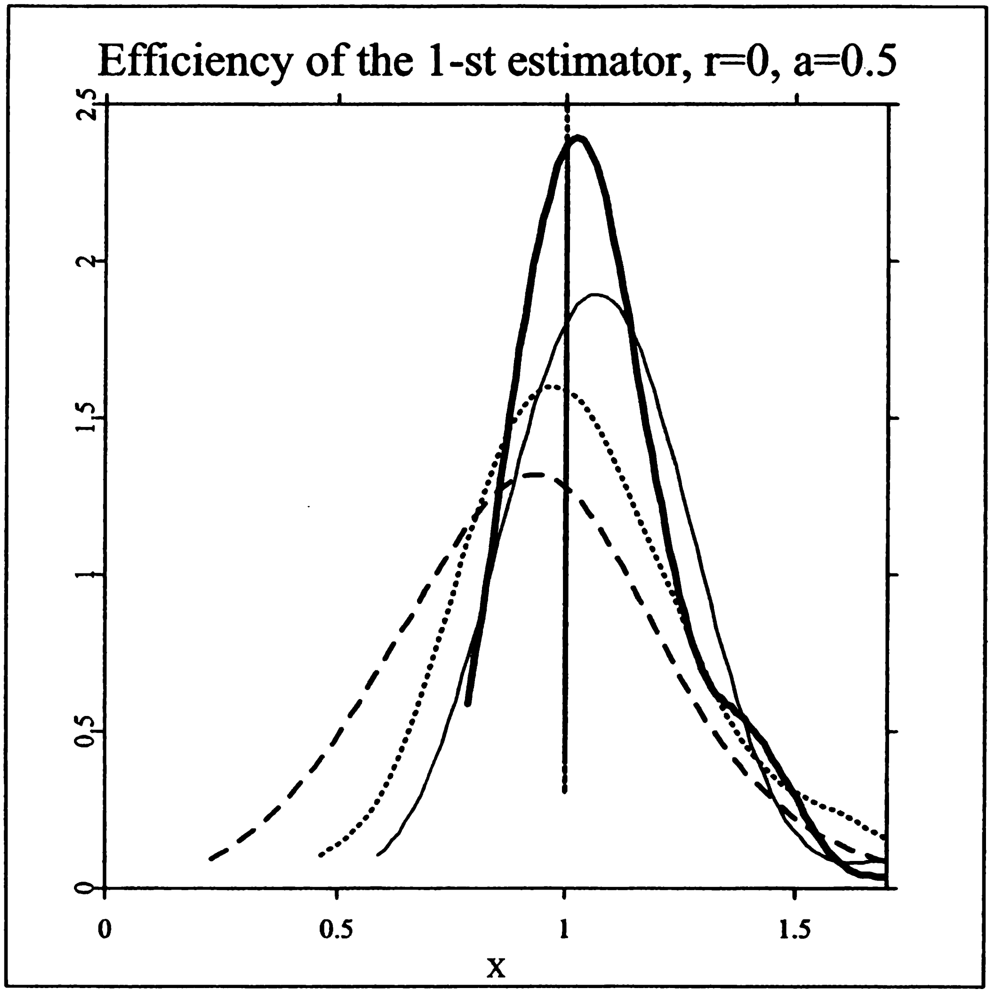

Figure 14. Plot of empirical distribution of relative efficiency: $r = 0, a = 0.5$.
Note: $n = 500$ - dashed line, $n = 1000$ - dotted line, $n = 1500$ - thin solid line,$n = 2000$ - thick solid line.

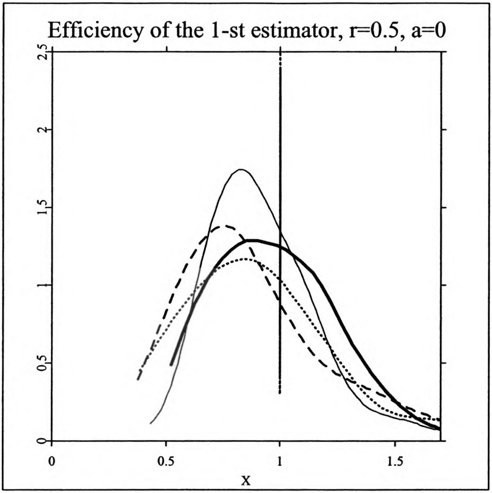

Figure 15. Plot of empirical distribution of relative efficiency: $r = 0.5, a = 0$.
Note: $n = 500$ - dashed line, $n = 1000$ - dotted line, $n = 1500$ - thin solid line,$n = 2000$ - thick solid line.

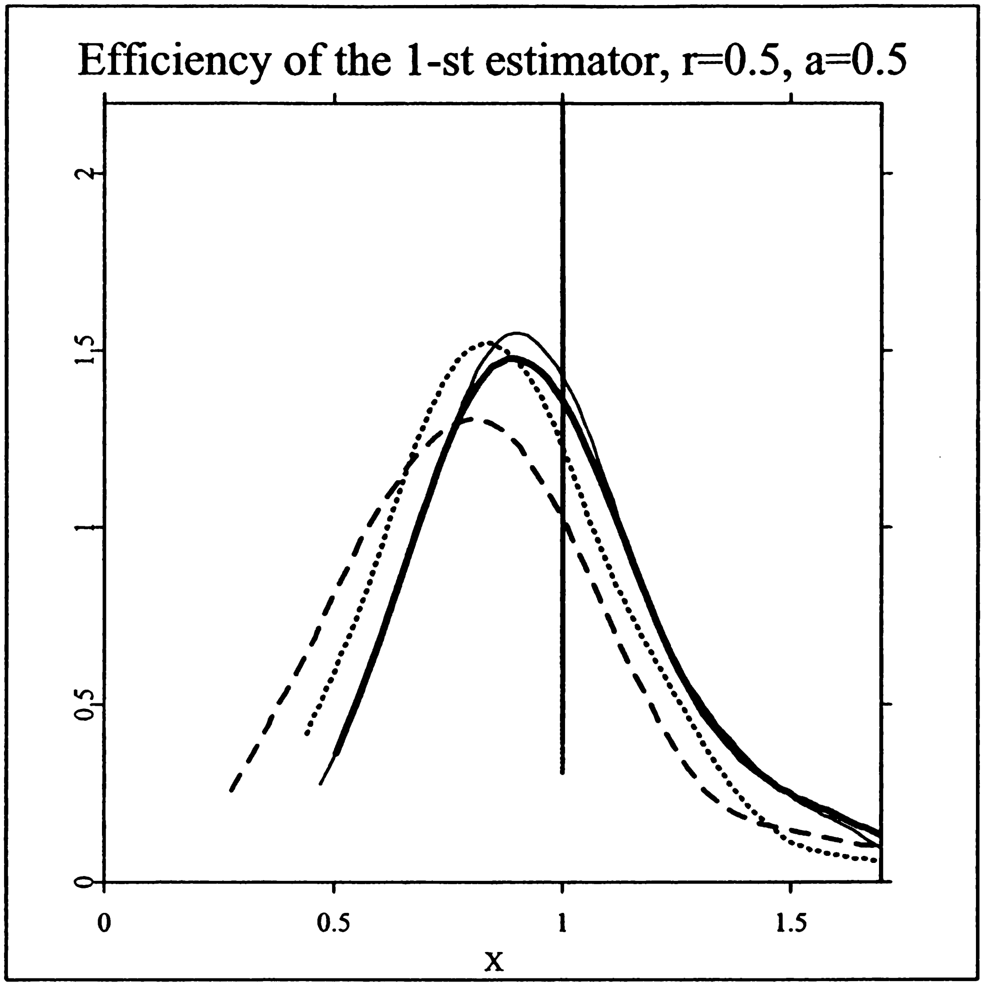

Figure 16. Plot of empirical distribution of relative efficiency: $r = 0.5, a = 0.5$.
Note: $n = 500$ - dashed line, $n = 1000$ - dotted line, $n = 1500$ - thin solid line,$n = 2000$ - thick solid line.

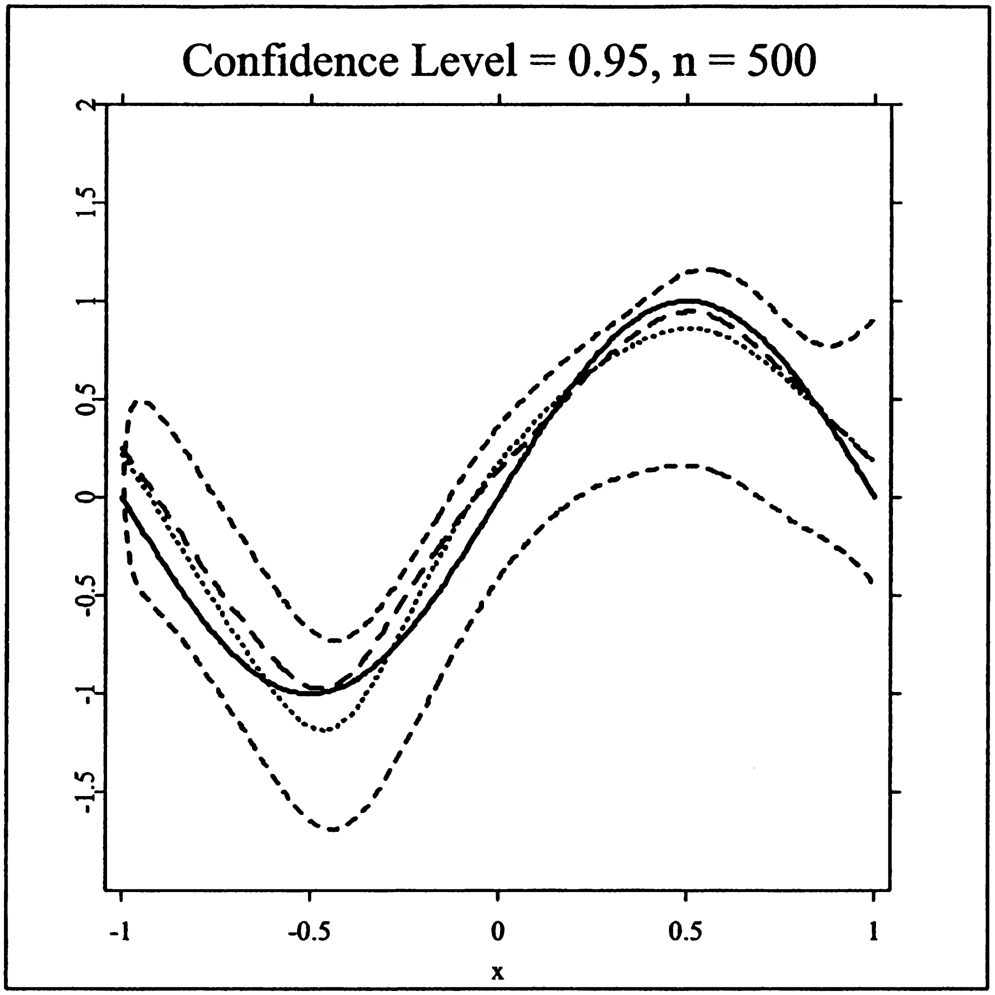

Figure 17. Plot of function estimation for $r = 0, a = 0$: $n = 500$.

Note: $m_1(x_1)$ - solid line, $\tilde{m}_{K,1}(x_1)$ - dashed line, confidence bands and $\hat{m}_{SBK,1}(x_1)$ - three dotted lines.

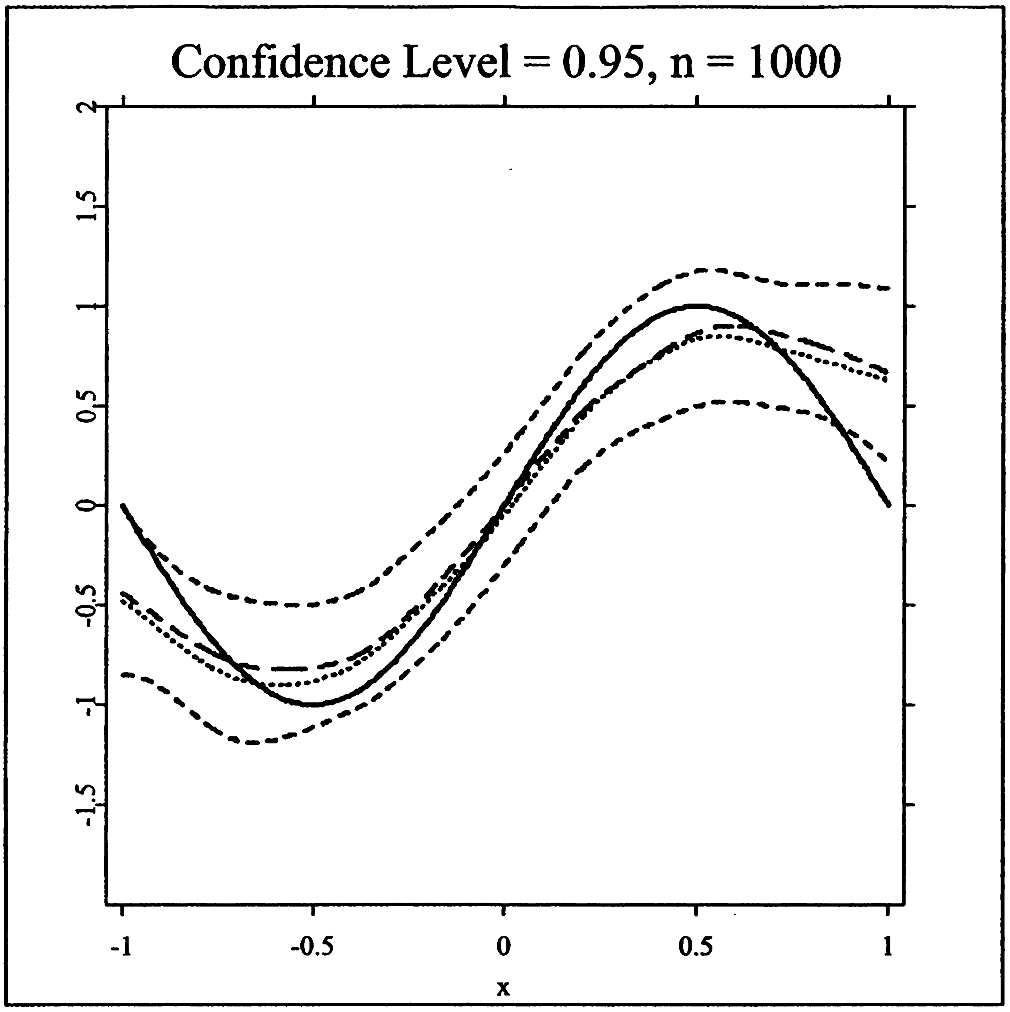

Figure 18. Plot of function estimation for $r = 0, a = 0$: $n = 1000$.
Note: $m_1(x_1)$ - solid line, $\tilde{m}_{K,1}(x_1)$ - dashed line, confidence bands and $\hat{m}_{SBK,1}(x_1)$ - three dotted lines.

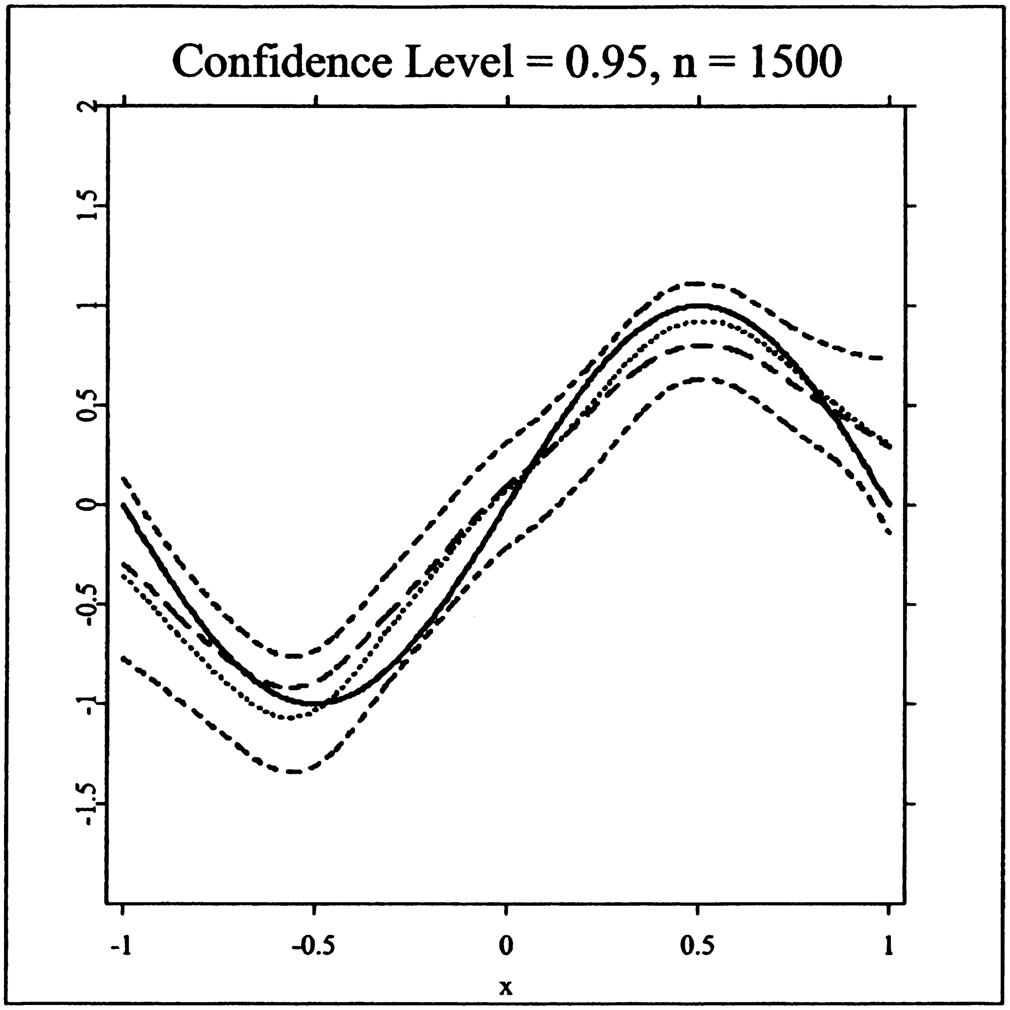

Figure 19. Plot of function estimation for $r = 0, a = 0$: $n = 1500$.

Note: $m_1(x_1)$ - solid line, $\tilde{m}_{\mathrm{K},1}(x_1)$ - dashed line, confidence bands and $\hat{m}_{\mathrm{SBK},1}(x_1)$ - three dotted lines.

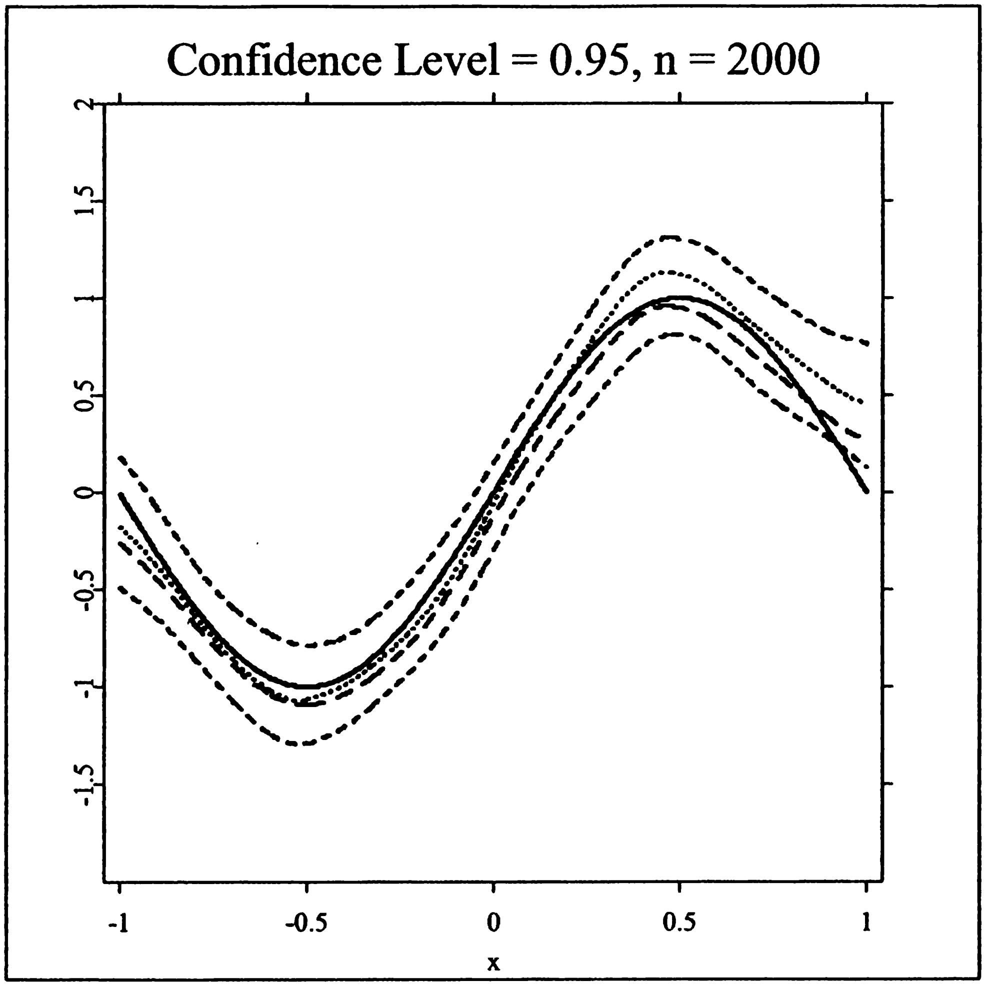

Figure 20. Plot of function estimation for $r = 0, a = 0$: $n = 2000$.

Note: $m_1(x_1)$ - solid line, $\tilde{m}_{K,1}(x_1)$ - dashed line, confidence bands and $\hat{m}_{SBK,1}(x_1)$ - three dotted lines.

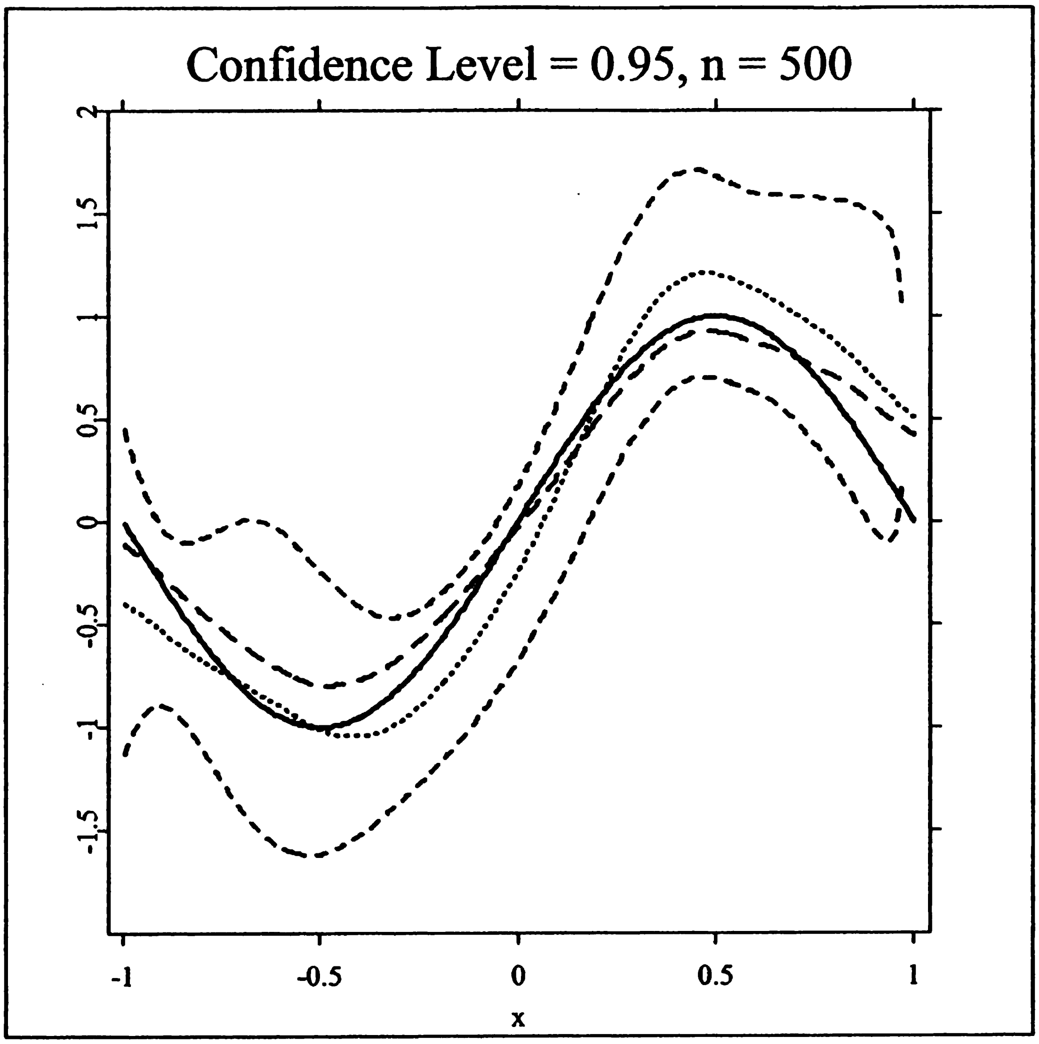

Figure 21. Plot of function estimation for $r = 0.5, a = 0.5$: $n = 500$.
Note: $m_1(x_1)$ - solid line, $\tilde{m}_{K,1}(x_1)$ - dashed line, confidence bands and $\hat{m}_{SBK,1}(x_1)$ - three dotted lines.

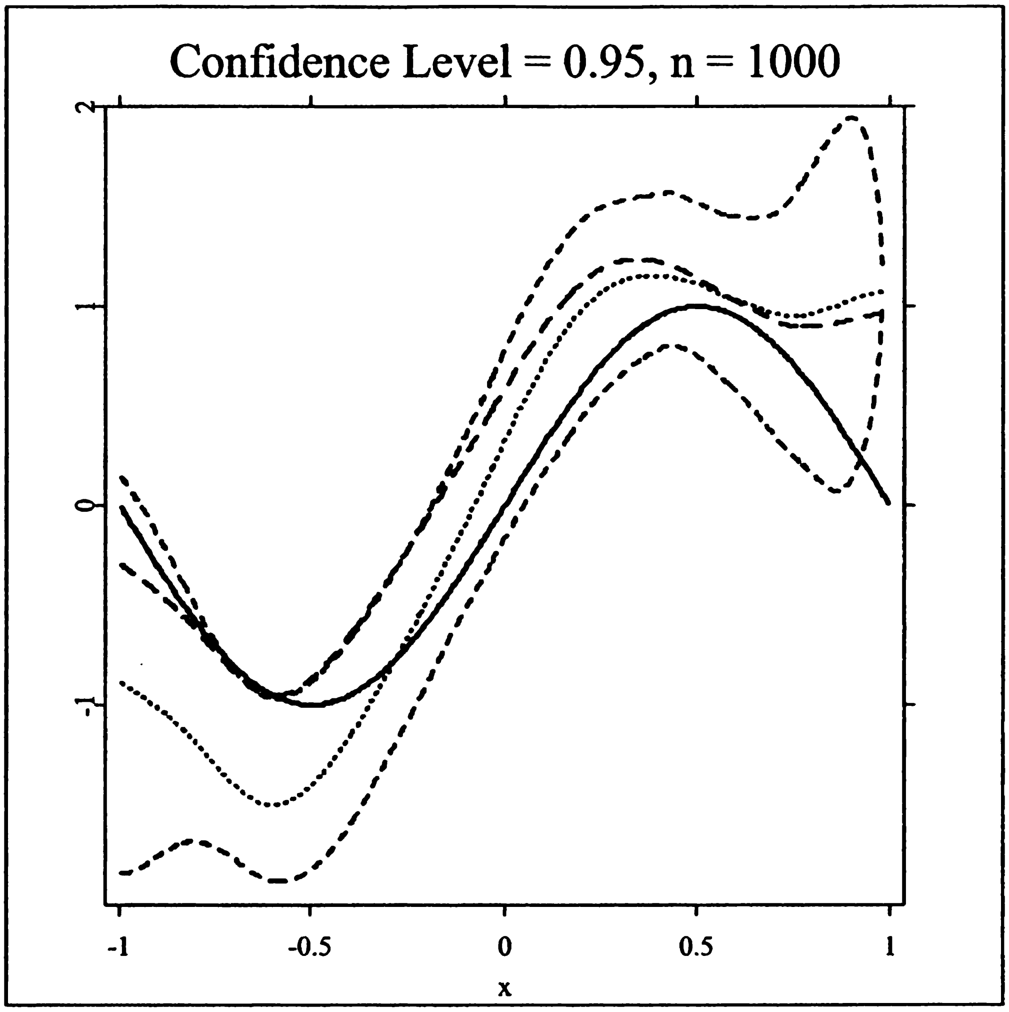

Figure 22. Plot of function estimation for $r = 0.5, a = 0.5$: $n = 1000$.
Note: $m_1(x_1)$ - solid line, $\tilde{m}_{K,1}(x_1)$ - dashed line, confidence bands and $\hat{m}_{SBK,1}(x_1)$ - three dotted lines.

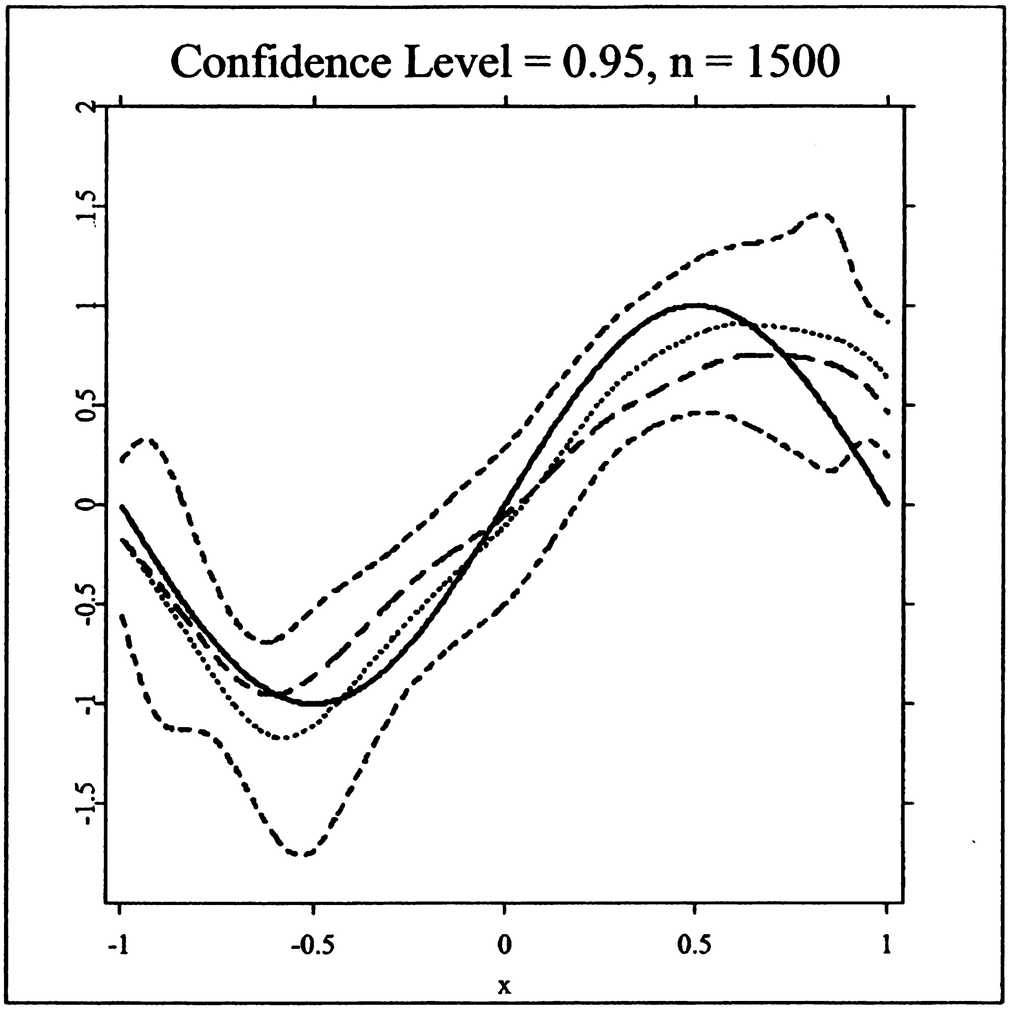

Figure 23. Plot of function estimation for $r = 0.5, a = 0.5$: $n = 1500$.
Note: $m_1(x_1)$ - solid line, $\tilde{m}_{K,1}(x_1)$ - dashed line, confidence bands and $\hat{m}_{SBK,1}(x_1)$ - three dotted lines.

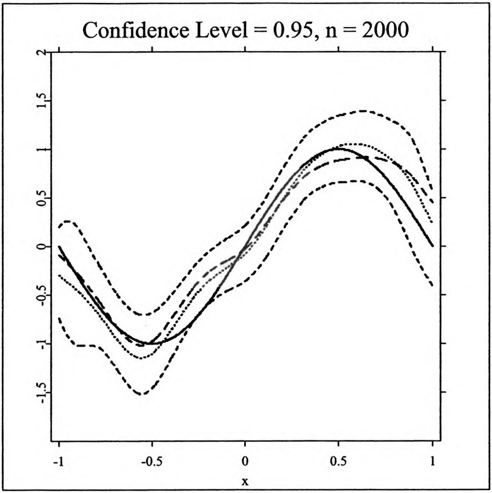

Figure 24. Plot of function estimation for $r = 0.5, a = 0.5$: $n = 2000$.
Note: $m_1(x_1)$ - solid line, $\tilde{m}_{K,1}(x_1)$ - dashed line, confidence bands and $\hat{m}_{SBK,1}(x_1)$ - three dotted lines.

BIBLIOGRAPHY

[1] Arnold, R. (2005). *R&D and Productivity Growth: A Background Paper.* Congressional Budget Office.

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

[3] Bollerslev, T. P. (1986). Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics*, **31**, 307–327.

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

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

[6] Cai, Z., Fan, J. and Yao, Q. (2000). Functional-coefficient regression models for nonlinear time series. *Journal of the American Statistical Association*, **95**, 941–956.

[7] Chan, N., Deng, S., Peng, L. and Xia, Z. (2007). Interval estimation of value-at-risk based on GARCH models with heavy-tailed innovations. *Journal of Econometrics*, **137**, 556–576.

[8] Chen, R. and Tsay, R. S. (1993a). Nonlinear additive ARX models. *Journal of the American Statistical Association*, **88**, 956–967.

[9] Chen, R. and Tsay, R. S. (1993b). Functional-coefficient autoregressive models. *Journal of the American Statistical Association*, **88**, 298–308.

[10] Cheng, M.Y. and Peng L. (2002). Regression modeling for nonparametric estimation of distribution and quantile functions. *Statistica Sinica*, **12**, 1043–1060.

[11] Cobb, C. W. and Douglas, P. H. (1928). A theory of production. *American Economic Review*, **18**, 139–165.

[12] Culpepper, W. L. (2004). *High R&D Spending Fuels Revenue Growth not Profits.* Available for downloading at *http://www.culpepper.com/eBulletin/2004/AugustRatiosArticle.asp.*

[13] Dahl, C. M. and Levine, M. (2006). Nonparametric estimation of volatility models with serially dependent innovations. *Statistics and Probability Letters*, **76**, 2007–2016.

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

[15] DeVore, R. A. and Lorentz, G. G. (1993). *Constructive Approximation: Polynomials and Splines Approximation.* Springer-Verlag, Berlin.

[16] Doukhan, P. (1994). *Mixing: Properties and Examples.* Springer-Verlag, New York.

[17] Duan, J. C. (1997). Augmented GARCH(p, q) process and its diffusion limit. *Journal of Econometrics*, **79**, 97–127.

[18] Engle, R. F. and Ng, V. (1993). Measuring and testing the impact of news on volatility. *Journal of Finance*, **48**, 1749–1778.

[19] Falk, M. (1985). Asymptotic normality of kernel type estimators of quantiles. *Annals of Statistics*, **13**, 428–433.

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

[21] Fan, J., Härdle, W. and Mammen, E. (1998). Direct estimation of low-dimensional components in additive models. *Annals of Statistics*, **26**, 943–971.

[22] Fan, J. and Jiang, J. (2005). Nonparametric inference for additive models. *Journal of the American Statistical Association*, **100**, 890–907.

[23] Glosten, L. R., Jaganathan, R. and Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance*, **48**, 1779–1801.

[24] Hafner, C. M. (1998). *Nonlinear Time Series Analysis with Applications to Foreign Exchange Rate Volatility.* Physica-Verlag, Heidelberg.

[25] Hafner, C. M. (2008). Temporal aggregation of multivariate GARCH processes. *Journal of Econometrics*, **142**, 467–483.

[26] Hafner, C. M. and Herwartz, H. (2006). Volatility impulse responses for multivariate GARCH models: An exchange rate illustration. *Journal of International Money and Finance*, **25**, 719–740.

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

[28] Hastie, T. J. and Tibshirani, R. J. (1990). *Generalized Additive Models.* Chapman and Hall, London.

[29] Hastie, T. J. and Tibshirani, R. J. (1993). Varying-coefficient models. *Journal of the Royal Statistical Society Series B*, **55**, 757–796.

[30] Hentschel, L. (1995). All in the family: nesting symmetric and asymmetric GARCH models. *Journal of Financial Economics*, **39**, 71–104.

[31] Hengartner, N. W. and Sperlich, S. (2005). Rate optimal estimation with the integration method in the presence of many covariates. *Journal of Multivariate Analysis*, **95**, 246–272.

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

[33] Horowitz, J. Klemelä, J. and Mammen, E. (2006). Optimal estimation in additive regression. *Bernoulli*, **12**, 271–298.

[34] Huang, J. Z. (1998a). Projection estimation in multiple regression with application to functional ANOVA models. *Annals of Statistics*, **26**, 242–272.

[35] Huang, J. Z. (1998b). Functional ANOVA models for generalized regression. *Journal of Multivariate Analysis*, **67**, 49–71.

[36] Huang, J. Z. (2003). Local asymptotics for polynomial spline regression. *Annals of Statistics*, **31**, 1600–1635.

[37] Huang, J. Z. and Shen, H. (2004). Functional coefficient regression models for non-linear time series: a polynomial spline approach. *Scandinavian Journal of Statistics*, **31**, 515–534.

[38] Huang, J. Z., Wu, C. O. and Zhou, L. (2002). Varying-coefficient models and basis function approximations for the analysis of repeated measurements. *Biometrika*, **89**, 111–128.

[39] Huang, J. Z. and Yang, L. (2004). Identification of nonlinear additive autoregression models. *Journal of the Royal Statistical Society Series B*, **66**, 463–477.

[40] Levine, M. (2006). Bandwidth selection for a class of difference-based variance estimators in the nonparametric regression: a possible approach. *Computational Statistics and Data Analysis*, **50**, 3405–3431.

[41] Li, R. and Liang, H. (2008). Variable selection in semiparametric regression modeling. *Annals of Statistics*, **36**, 261–286.

[42] Li, Q. and Racine, J. S. (2007). *Nonparametric Econometrics: Theory and Practice.* Princeton University Press, Princeton.

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

[44] Linton, O. B. and Härdle, W. (1996). Estimation of additive regression models with known links. *Biometrika*, **83**, 529–540.

[45] Linton, O. B. and Mammen, E. (2005). Estimating Semiparametric Arch (∞) Models by Kernel Smoothing Methods. *Econometrica*, **73**, 771-836.

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

[47] Liu, R and Yang, L. (2008). Kernel estimation of multivariate cumulative distribution function. *Journal of Nonparametric Statistics*, **20**, 661–677.

[48] Liu, R and Yang, L. (2009). Spline-backfitted kernel smoothing of additive coefficient model. **In Press**. *Econometric Theory*,

[49] Mammen, E., Linton, O. and Nielsen, J. (1999). The existence and asymptotic properties of a backfitting projection algorithm under weak conditions. *Annals of Statistics*, **27**, 1443–1490.

[50] McConnell, C. and Brue, S. (1999). *Economics: Principles, Problems, and Policies.* Irwin/McGraw-Hill, Boston.

[51] Nielsen, J. P. and Sperlich, S. (2005). Smooth backfitting in practice. *Journal of the Royal Statistical Society Series B*, **67**, 43–61.

[52] Opsomer, J. D. and Ruppert, D. (1997). Fitting a bivariate additive model by local polynomial regression. *Annals of Statistics*, **25**, 186–211.

[53] Peng, L. and Yao, Q. (2003). Least absolute deviations estimation for ARCH and GARCH models. *Biometrika*, **90**, 967–975.

[54] Pham, D. T. (1986). The mixing properties of bilinear and generalized random coefficient autoregressive models. *Stochastic Analysis and Applications*, **23**, 291–300.

[55] Reiss, R. D. (1981). Nonparametric estimation of smooth distribution functions. *Scandinavian Journal of Statistics*, **8**, 116–119.

[56] Robinson, P. M. (1983). Nonparametric estimators for time series. *Journal of Time Series Analysis*, **4**, 185–207.

[57] Rodríguez-Póo, J. M., Sperlich, S. and Vieu, P. (2003). Semiparametric estimation of separable models with possibly limited dependent variables. *Econometric Theory*, **19**, 1008–1039.

[58] Samuelson, P. (1995). *Economics.* Irwin/McGraw-Hill, New York.

[59] Scott, D. W. (1992). *Multivariate Density Estimation: Theory, Practice, and Visualization.* John Wiley & Sons, New York.

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

[61] Solow, R. M. (1957). Technical change and the aggregate production function. The *Review of Economics and Statistics*, **39**, 312–320.

[62] Stone, C. J. (1985). Additive regression and other nonparametric models. *Annals of Statistics*, **13**, 689–705.

[63] Stone, C. J. (1986). The dimensionality reduction principle for generalized additive models. *Annals of Statistics*, **14**, 590–606.

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

[65] Sun, Y. and Stengos, T. (2006). Semiparametric efficient adaptive estimation of asymmetric GARCH models. *Journal of Econometrics*, **133**, 373–386.

[66] Sunklodas, J. (1984). On the rate of convergence in the central limit theorem for strongly mixing random variables. *Lithuanian Mathematical Journal*, **24**, 182–190.

[67] Tjøstheim, D. and Auestad, B. (1994). Nonparametric identification of nonlinear time series: projections. *Journal of the American Statistical Association*, **89**, 1398–1409.

[68] Wang, L. and Yang, L. (2007). Spline-backfitted kernel smoothing of nonlinear additive autoregression model. *Annals of Statistics*, **35**, 2474–2503.

[69] Xue, L. and Yang, L. (2006a). Estimation of semiparametric additive coefficient model. *Journal of Statistical Planning and Inference*, **136**, 2506–2534.

[70] Xue, L. and Yang, L. (2006b). Additive coefficient modeling via polynomial spline. *Statistica Sinica*, **16**, 1423–1446.

[71] Yamato, H. (1973). Uniform convergence of an estimator of a distribution function. *Bulletin of Mathematical Statistics*, **15**, 69–78.

[72] Yang, L. (2000). Finite nonparametric GARCH model for foreign exchange volatility. *Communications in Statistics-Theory and Methods*, **5** & **6**, 1347–1365.

[73] Yang, L. (2002). Direct estimation in an additive model when the components are proportional. *Statistica Sinica*, **12**, 801–821.

[74] Yang, L. (2006). A semiparametric GARCH model for foreign exchange volatility. *Journal of Econometrics*, **130**, 365–384.

[75] Yang, L., Härdle, W. and Nielsen, J. P. (1999). Nonparametric autoregression with multiplicative volatility and additive mean. *Journal of Time Series Analysis*, **20**, 579–604.

[76] Yang, L., Sperlich, S. and Härdle, W. (2003). Derivative estimation and testing in generalized additive models. *Journal of Statistical Planning and Inference*, **115**, 521–542.

[77] Yang, L. and Tschernig, R. (1999). Multivariate bandwidth selection for local linear regression. *Journal of the Royal Statistical Society Series B*, **61**, 793-815.

[78] Zhang, F. (1999). *Matrix Theory*. Springer-Verlag, New York.