SFB 649 Discussion Paper 2011-016

Oracally Efficient Two-Step Estimation of Generalized Additive Model

Rong Liu*
Lijian Yang**
Wolfgang Karl Härdle***

* University of Toledo, USA
** Michigan State University, USA
*** Humboldt-Universität zu Berlin, Germany

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".

http://sfb649.wiwi.hu-berlin.de
ISSN 1860-5664

SFB 649, Humboldt-Universität zu Berlin
Spandauer Straße 1, D-10178 Berlin

Submitted to the Annals of Statistics
arXiv: math.PR/0000000

ORACALLY EFFICIENT TWO-STEP ESTIMATION OF GENERALIZED ADDITIVE MODEL *

By Rong Liu[1], Lijian Yang[2,3] and Wolfgang K. Härdle[4]

[1] *University of Toledo,* [2] *Soochow University,* [3] *Michigan State University,*
[4] *Humboldt-Universität zu Berlin*

Generalized additive models (GAM) are multivariate nonparametric regressions for non-Gaussian responses including binary and count data. We propose a spline-backfitted kernel (SBK) estimator for the component functions. Our results are for weakly dependent data and we prove oracle efficiency. The SBK techniques is both computational expedient and theoretically reliable, thus usable for analyzing high-dimensional time series. Inference can be made on component functions based on asymptotic normality. Simulation evidence strongly corroborates with the asymptotic theory.

1. Introduction. An effective semiparametric regression tool for high dimensional data is the additive model introduced by Hastie and Tibshirani (1990), which stipulates that

$$\mathsf{E}\left(Y|\mathbf{X}\right)=m\left(\mathbf{X}\right), m\left(\mathbf{X}\right)=c+\textstyle\sum_{\alpha=1}^{d}m_{\alpha}\left(X_{\alpha}\right) \tag{1.1}$$

for a response Y and a predictor vector $\mathbf{X}=(X_1,...,X_d)^{\mathsf{T}}$. When a data set $\left\{Y_i,\mathbf{X}_i^{\mathsf{T}}\right\}_{i=1}^{n}=\left\{Y_i,X_{i1},...,X_{id}\right\}_{i=1}^{n}$ of size n is observed which follows model (1.1), unknown component functions $\left\{m_\alpha\left(x_\alpha\right)\right\}_{\alpha=1}^{d}$ can be estimated via kernel, B spline and smoothing spline with a univariate convergence rate. This fact together with the interpretability of the functions has not only led to a remedy of the "curse of dimensionality", but also led to increased practical applications of additive models. A list of articles on additive models and related works include, among others, Stone (1985), Stone (1994), Huang (2004) and Xue and Yang (2006a) for B spline methods; Tjøstheim and Auestad (1994), Linton and Nielsen (1995), Linton (1997), Fan et al. (1998), Yang et al. (1999), Xue and Yang (2006b) and Yang et al.

*Supported in part by the Deutsche Forschungsgemeinschaft through the CRC 649 "Economic Risk", by NSF awards DMS 0706518, 1007594 and by a Credit Rating Grant from the National University of Singapore Risk Management Institute.

AMS 2000 subject classifications: Primary 62M10; secondary 62G08

JEL classification: C00, C14, J01, J31

Keywords and phrases: Bandwidths, B spline, knots, link function, mixing, Nadaraya-Watson estimator

(2006) for kernel methods; and more recently, spline-backfitted kernel (SBK) smoothing methods of Wang and Yang (2007), Wang and Yang (2009), Liu and Yang (2010) and Ma and Yang (2011), the spline-backfitted spline (SBS) smoothing method of Song and Yang (2010).

Certain types of responses Y, however, such as binary or Poisson responses, are much more appropriately described by GAMs. In the GAM framework, the data $\left\{Y_i, \mathbf{X}_i^{\mathsf{T}}\right\}_{i=1}^n$ are generated according to

$$\mathsf{E}\left(Y|\mathbf{X}=\mathbf{x}\right) = b'\left\{m\left(\mathbf{x}\right)\right\}, \tag{1.2}$$

with $m\left(\mathbf{x}\right)$ of additive structure as in (1.1), and a given function b' which relates $m\left(\mathbf{x}\right)$ to the conditional variance function $\sigma^2\left(\mathbf{x}\right) = \mathsf{Var}\left(Y|\mathbf{X}=\mathbf{x}\right)$ via the equation $\sigma^2\left(\mathbf{x}\right) = a\left(\phi\right)b''\left\{m\left(\mathbf{x}\right)\right\}$, in which $a\left(\phi\right)$ is a nuisance parameter that quantifies overdispersion. The inverse of b' is called the link function. For binary responses, one commonly takes $\left(b'\right)^{-1}\left(x\right) = \log\left\{x/\left(1-x\right)\right\}$, the logistic link to conduct logistic regression, while for Poisson regression, $\left(b'\right)^{-1}\left(x\right) = \log x$, the log link. If one takes $\left(b'\right)^{-1}\left(x\right) = x$, the identity link, model (1.2) becomes model (1.1).

Model (1.2) has its origin in the special case where the probability density function of Y_i conditional on $\mathbf{X}_i$ with respect to a fixed σ-finite measure forms an exponential family

$$f\left(Y_i\left|\mathbf{X}_i,\phi\right.\right) = \exp\left[\left\{Y_i m\left(\mathbf{X}_i\right) - b\left\{m\left(\mathbf{X}_i\right)\right\}\right\}/a\left(\phi\right) + h\left(Y_i,\phi\right)\right].$$

For the theoretical development in this paper, however, it is not necessary to assume that the data $\left\{Y_i, \mathbf{X}_i^{\mathsf{T}}\right\}_{i=1}^n$ comes from such exponential family, but only that the conditional variance and conditional mean are linked by the following equation

$$\mathsf{Var}\left(Y|\mathbf{X}=\mathbf{x}\right) = a\left(\phi\right)b''\left[\left(b'\right)^{-1}\left\{\mathsf{E}\left(Y|\mathbf{X}=\mathbf{x}\right)\right\}\right].$$

We can also write model (1.2) in the usual regression form

$$Y_i = b'\left\{m\left(\mathbf{X}_i\right)\right\} + \sigma\left(\mathbf{X}_i\right)\varepsilon_i \tag{1.3}$$

for conditional white noise ε_i that satisfies $\mathsf{E}\left(\varepsilon_i|\mathbf{X}_i\right) = 0$, $\mathsf{E}\left(\varepsilon_i^2|\mathbf{X}_i\right) = 1$. For identifiability, one requires that

$$\mathsf{E}\left\{m_\alpha\left(X_\alpha\right)\right\} = 0, 1 \le \alpha \le d \tag{1.4}$$

for unique additive representations of $m\left(\mathbf{x}\right) = c + \sum_{\alpha=1}^d m_\alpha\left(x_\alpha\right)$. As in most works on nonparametric smoothing, estimation of the functions $\left\{m_\alpha\left(x_\alpha\right)\right\}_{\alpha=1}^d$

is conducted on compact sets. Without lose of generality, let the compact set be $\boldsymbol{\chi} = [0,1]^d$.

Methods for the generalized additive model (1.2) are much less developed in comparison to the additive model (1.1), see for instance, the B spline method of Stone (1986) and Xue and Liang (2010), the kernel method of Linton and Härdle (1996) and Yang et al. (2003), and the two-stage methods of Horowitz and Mammen (2004) and Horowitz et al. (2006). Generally speaking, the proposed kernel methods are too computationally intensive for high dimension d, thus limiting their applicability to a small number of predictors. On the other hand, B spline methods provide only convergence rates but no asymptotic distributions, so no measures of confidence can be assigned to the estimators. In the case of the additive model (1.1), the SBK method of Wang and Yang (2007) combines the advantages of both kernel and spline methods and the result is balanced in terms of theory, computation, and interpretation. The basic idea of the SBK method for the additive model (1.1) is to first project the data with B-splines into a space of functions with additive structure and then to apply kernel smoothing to the projected objects.

In this paper we extend the SBK method to model (1.2). The desired aim is to achieve orcale efficiency. If all the nonparametric functions of the last $d-1$ variables, $\{m_\alpha(x_\alpha)\}_{\alpha=2}^d$ and the constant c were known by an "oracle", one could simply plug these in and estimate the only unknown functions $m_1(x_1)$ by maximizing the log-likelihood function with kernel weights computed from variable X_1. This estimator of $m_1(x_1)$ is called "oracle smoother" or "infeasible estimator", and it does not suffer from the "curse of dimensionality" since the smoothing operation involves w.l.o.g. only X_1. The proposed SBK method pre-estimates functions $\{m_\alpha(x_\alpha)\}_{\alpha=2}^d$ and constant c by linear splines and then use these estimates as proxies for the unknown functions $\{m_\alpha(x_\alpha)\}_{\alpha=2}^d$ and constant c. The main contribution is proving that the error caused by this approximation is uniformly negligible of order $\mathcal{O}_{a.s.}\left(n^{-1/2}\log n\right)$. Consequently, the SBK estimator is uniformly (over the data range) asymptotically equivalent to the "oracle smoother", automatically inheriting all oracle efficiency properties of the latter. Our proof relies on "reducing bias by undersmoothing" and "averaging out the variance", accomplished with the joint asymptotics of kernel and spline functions for realizations of geometrically strongly mixing time series. These results are established under substantially greater technical difficulty than existng works on additive model such as Wang and Yang (2007), Wang and Yang (2009), Liu and Yang (2010), Ma and Yang (2011), and Song and Yang (2010). The additional complication is due to the lack of decomposition of spline estimation error into the sum of a bias and a noise term when the link function $(b')^{-1}$ is nonlinear.

A similar result was proved in Horowitz and Mammen (2004) for i.i.d. rather than dependent data and only pointwise rates instead of uniform rates were derived. It is also worth emphasizing that although Horowitz and Mammen (2004) had used the B-spline estimator for the first stage in simulation, their proof is valid only for using the orthogonal series estimator in stage one. Another major contribution of this paper is establishing that the spline-backfitted estimator of the additive constant c is within an negligible error of order $\mathcal{O}_p\left(n^{-1/2}\right)$ of the infeasible estimator and thus also oracally efficient. As far as we know, our estimator of the additive constant c is the only one which has an asymptotic distribution with $n^{-1/2}$ rate.

The paper is organized as follows. In Section 2 we discuss the assumptions of the model (1.2). In Section 3, we introduce the oracle smoother or infeasible estimator for $m_1(x_1)$ and for c, and state their asymptotics. In Section 4 we introduce the SBK estimator for $m_1(x_1)$ and spline-backfitted estimator for c and present their asymptotic oracle efficiencies by showing that they differ from their infeasible counterparts only negligibly. In Section 5 we describe implementation steps of the estimators. In Section 6 we apply the methods to simulated and real examples. All technical proofs are given in the Appendix.

2. Model assumptions.

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

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

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

$$\mathcal{M}=\left\{g\left(\mathbf{x}\right)=c+\sum_{\alpha=1}^d g_\alpha\left(\mathbf{x}\right);g_\alpha\in\mathcal{H}^0\right\},$$

in which c is finite constant. The constraints that $\mathsf{E}\left\{g_\alpha\left(X_\alpha\right)\right\}=0, 1\leq\alpha\leq d$ ensure unique additive representation of m_α as expressed in (1.4), but are not necessary for the definition of space $\mathcal{M}$. In what follows, denote by E_n the empirical expectation, $\mathsf{E}_n\varphi=\sum_{i=1}^n\varphi\left(\mathbf{X}_i\right)/n$. 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 $\langle g_1,g_2\rangle=\mathrm{E}\left\{g_1\left(\mathbf{X}\right)g_2\left(\mathbf{X}\right)\right\}$, $\langle g_1,g_2\rangle_n=\mathsf{E}_n\left\{g_1\left(\mathbf{X}\right)g_2\left(\mathbf{X}\right)\right\}$. The corresponding induced norms are $\|g_1\|_2^2=\mathsf{E}\,g_1^2\left(\mathbf{X}\right)$, $\|g_1\|_{2,n}^2=\mathsf{E}_n\,g_1^2\left(\mathbf{X}\right)$. More generally, we define $\|g\|_r^r=\mathsf{E}\,g^r\left(\mathbf{X}\right)$.

Throughout the paper, for any compact interval $[a,b]$, we denote the space of p-th order smooth function as $C^{(p)}[a,b]=\left\{g|g^{(p)}\in C\left[a,b\right]\right\}$, and the class of Lipschitz continuous functions for constant $C > 0$ as $\mathrm{Lip}\left([a,b], C\right) = \{g|\, |g(x) - g(x')| \leq C|x - x'|,\ \forall x, x' \in [a,b]\}$. We mean by "$\sim$" both sides having the same order as $n \to \infty$. For any vector $\mathbf{x} = (x_1, x_2, \cdots, x_d)^{\mathsf{T}}$, we denote the supremum and p norms as $|\mathbf{x}| = \max_{1\leq\alpha\leq d}|x_\alpha|$ and $\|\mathbf{x}\|_p = \left(\sum_{\alpha=1}^{d} x_\alpha^p\right)^{1/p}$. In particular, we use $\|\mathbf{x}\|$ to denote the Euclidean norm.

We need the following Assumptions on the data generating process.

(A1) *The additive component functions* $m_\alpha \in C^{(1)}[0,1], 1 \leq \alpha \leq d$ *with* $m_1 \in C^{(2)}[0,1]$, $m'_\alpha \in \mathrm{Lip}\left([0,1], C_m\right) = 2 \leq \alpha \leq d$ *for some constant* $C_m > 0$.

(A2) *The inverse link function* b' *satisfies:* $b' \in C^2(\mathbb{R}), b''(\theta) > 0, \theta \in \mathbb{R}$ *while for a compact interval* Θ *whose interior contains* $m\left([0,1]^d\right)$, $C_b > \max_{\theta\in\Theta} b''(\theta) \geq \min_{\theta\in\Theta} b''(\theta) > c_b$ *for constants* $C_b > c_b > 0$.

(A3) *The conditional variance function* $\sigma^2(\mathbf{x})$ *is measurable and bounded. The errors* $\{\varepsilon_i\}_{i=1}^n$ *satisfy* $\mathsf{E}(\varepsilon_i|\mathcal{F}_i) = 0$, $\mathsf{E}\left(|\varepsilon_i|^{2+\eta}\right) \leq C_\eta$ *for some* $\eta \in (1/2, 1]$ *and the sequence of* σ*-fields* $\mathcal{F}_i = \sigma\{(\mathbf{X}_j), j \leq i; \varepsilon_j, j \leq i-1\}$ *for* $i = 1, \ldots, n$.

(A4) *The density function* $f(\mathbf{x})$ *of* $(X_1, ..., X_d)$ *is continuous and*

$$0 < c_f \leq \inf_{\mathbf{x}\in\boldsymbol{\chi}} f(\mathbf{x}) \leq \sup_{\mathbf{x}\in\boldsymbol{\chi}} f(\mathbf{x}) \leq C_f < \infty.$$

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) *Constants* $K_0, \lambda_0 \in (0, +\infty)$ *exist such that* $\alpha(n) \leq K_0 e^{-\lambda_0 n}$ *holds for all* n, *with the* α*-mixing coefficients for* $\left\{\mathbf{Z}_i = \left(\mathbf{X}_i^{\mathsf{T}}, \varepsilon_i\right)\right\}_{i=1}^n$ *defined as*

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

Assumptions (A1), (A2) and (A4) are standard in the GAM literature, see Stone (1986), Xue and Liang (2010), while Assumptions (A3) and (A5) are the same for weakly dependent data as in Wang and Yang (2007), Liu and Yang (2010). Assumption (A2) implies that a compact interval A exists whose interior contains $m_1([0,1])$ and that Θ's interior contains $A + m_{\text{-}1}\left([0,1]^{d-1}\right)$ where $m_{\text{-}1}(\mathbf{x}_{\text{-}1}) = c + \sum_{\alpha=2}^{d} m_\alpha(x_\alpha)$ with $x_{\text{-}1} = (x_2, ..., x_d)$.

3. Oracle smoothers.

We now introduce what is known as the oracle smoother in Wang and Yang (2007) as a benchmark for evaluating the estimators. If the last $d-1$ components $\{m_\alpha(x_\alpha)\}_{\alpha=2}^d$ were w.l.o.g. known by an "oracle", then the only unknown component $m_1(x_1)$ may be estimated by the following procedure. Define for each $x_1 \in [h, 1-h]$ a local log-likelihood function $\tilde{l}(a) = \tilde{l}(a, x_1), a \in A$ as

$$n^{-1} \sum_{i=1}^{n} \left[Y_i \left\{a + m_{\text{-}1}(\mathbf{X}_{i\text{-}1})\right\} - b\left\{a + m_{\text{-}1}(\mathbf{X}_{i\text{-}1})\right\}\right] K_h(X_{i1} - x_1) \tag{3.1}$$

with $m_{\text{-}1}(\mathbf{X}_{i\text{-}1}) = c + \sum_{\alpha=2}^{d} m_\alpha(\mathbf{X}_{i\alpha})$ and define the oracle smoother of $m_1(x_1)$ as

$$\tilde{m}_{\mathrm{K},1}(x_1) = \operatorname{argmax}_{a \in A} \tilde{l}(a, x_1). \tag{3.2}$$

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

(A6) *The kernel function K is a symmetric probability density, supported on $[-1, 1]$ and $K \in \operatorname{Lip}([-1, 1], C_K)$ for some positive constant $C_K > 0$. A constant $c_h > 0$ exists such that the bandwidth $h = h_n$ satisfies $h = \mathcal{O}\left(n^{-1/5}\right), h^{-1} = \mathcal{O}\left(n^{1/5}(\log n)^{c_h}\right)$.*

In what follows, we denote $\|K\|_2^2 = \int K^2(u)\,du$, $\mu_2(K) = \int K(u) u^2 du$. Denote the higher order error of $\tilde{m}_{\mathrm{K},1}(x_1)$ as

$$\begin{aligned} r_{\mathrm{K},1}(x_1) &= \tilde{m}_{\mathrm{K},1}(x_1) - m_1(x_1) - \operatorname{bias}_1(x_1) h^2 / D_1(x_1) \\ &\quad - n^{-1} \sum_{i=1}^{n} K_h(X_{i1} - x_1) \sigma(\mathbf{X}_i) \varepsilon_i / D_1(x_1), \end{aligned}$$

with the scale function $D_1(x_1)$ and bias function $\operatorname{bias}_1(x_1)$ defined as

$$D_1(x_1) = f_1(x_1) \mathsf{E}\left\{b''\left\{m(\mathbf{X})\right\} | X_1 = x_1\right\}, \tag{3.3}$$

$$\begin{aligned} \operatorname{bias}_1(x_1) &= \mu_2(K) \Big[m_1''(x_1) f(x_1) \mathsf{E}\left[b''\left\{m(\mathbf{X})\right\} | X_1 = x_1\right] \\ &\quad + m_1'(x_1) f(x_1) \frac{\partial}{\partial x_1} \mathsf{E}\left[b''\left\{m(\mathbf{X})\right\} | X_1 = x_1\right] \\ &\quad - \left\{m_1'(x_1)\right\}^2 f(x_1) \mathsf{E}\left[b'''\left\{m(\mathbf{X})\right\} | X_1 = x_1\right]\Big]. \end{aligned} \tag{3.4}$$

THEOREM 1. *Under Assumptions (A1)-(A6), as $n \to \infty$*

$$\sup_{x_1 \in [h, 1-h]} |r_{\mathrm{K},1}(x_1)| = \mathcal{O}_{a.s.}\left(n^{-1/2} h^{1/2} \log n\right).$$

In particular, $\sup_{x_1 \in [h, 1-h]} |\tilde{m}_{\mathrm{K},1}(x_1) - m_1(x_1)| = \mathcal{O}_{a.s.}\left(\log n / \sqrt{nh}\right)$.

THEOREM 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 (3.2) satisfies*

$$\sqrt{nh}\left\{\tilde{m}_{\mathrm{K},1}(x_1) - m_1(x_1) - \mathrm{bias}_1(x_1)\, h^2/D_1(x_1)\right\}$$

$$\xrightarrow{\mathcal{L}} N\left(0, D_1(x_1)^{-1} v_1^2(x_1)\, D_1(x_1)^{-1}\right)$$

in which

$$v_1^2(x_1) = f_1(x_1)\,\mathsf{E}\left\{\sigma^2(\mathbf{X})\,|X_1 = x_1\right\}\|K\|_2^2. \tag{3.5}$$

The same oracle idea applies to the constant as well. Define the log-likelihood function

$$\tilde{l}_c(a) = n^{-1}\textstyle\sum_{i=1}^n \left[Y_i\left\{a + m_{\text{-}c}(\mathbf{X}_i)\right\} - b\left\{a + m_{\text{-}c}(\mathbf{X}_i)\right\}\right],$$

where $m_{\text{-}c}(\mathbf{X}) = \sum_{\alpha=1}^d m_\alpha(X_\alpha)$. The infeasible estimator of c is defined as $\tilde{c} = \mathrm{argmax}_{a\in A}\,\tilde{l}_c(a)$. Clearly, $\tilde{l}_c'(\tilde{c}) = 0$.

THEOREM 3. *Under Assumptions (A1)-(A5), as $n \to \infty$*

$$\tilde{c} - c = \left[\mathsf{E}\, b''\left\{m(\mathbf{X})\right\}\right]^{-1} n^{-1}\textstyle\sum_{i=1}^n \sigma(\mathbf{X}_i)\,\varepsilon_i + \mathcal{O}_{a.s}\left(n^{-1}(\log n)^2\right).$$

Although the oracle smoother $\tilde{m}_{\mathrm{K},1}(x_1)$ enjoys the desirable theoretical properties in Theorems 1 and 2, it not useful statistics as its computation is based on the knowledge of unavailable functions $\{m_\alpha(x_\alpha)\}_{\alpha=2}^d$ and the unknown constant c, the same can be said of $\tilde{c}$. These benchmarks, however, motivate the spline-backfitted estimators that we will introduce in the next section.

4. Spline-backfitted kernel estimators.

In this section we describe how the unknown functions $\{m_\alpha(x_\alpha)\}_{\alpha=2}^d$ and constants c can be pre-estimated by linear splines and how the estimates are used to construct the SBK estimator. First, we introduce the space of linear splines defined in Liu and Yang (2010). 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 $\left[\xi_J, \xi_{J+1}\right], 0 \le J \le N$ and denote the degenerate knots $\xi_{-1} = 0, \xi_{N+2} = 1$. Assume that

(A7) *The number of interior knots* satisfies: $N \sim n^{1/4}\log n$, *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.

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

$$G_{n,\alpha}^0=\left\{g_\alpha : g_\alpha(x_\alpha)=\textstyle\sum_{J=0}^{N+1}\lambda_J b_J(x_\alpha), \mathsf{E}_n\{g_\alpha(X_\alpha)\}=0\right\}, 1\le\alpha\le d,$$

and the space of additive spline functions on $\boldsymbol{\chi}$ as

$$G_n^0=\left\{g(\mathbf{x})=c+\textstyle\sum_{\alpha=1}^d g_\alpha(x_\alpha);\quad c\in R, g_\alpha\in G_{n,\alpha}^0\right\},$$

which is equipped with the empirical inner product $\langle\cdot,\cdot\rangle_{2,n}$. Define the log-likelihood function $\hat{L}(g)=n^{-1}\sum_{i=1}^n[Y_i g(\mathbf{X}_i)-b\{g(\mathbf{X}_i)\}], g\in G_n^0$, which according to Lemma 14 of Stone (1986), has a unique maximizer with probability approaching 1. The multivariate function $m(\mathbf{x})$ is then estimated by the additive spline function

$$\hat{m}(\mathbf{x})=\operatorname{argmax}_{g\in G_n^0}\hat{L}(g).$$

Since $\hat{m}(\mathbf{x})\in G_n^0$, one can write $\hat{m}(\mathbf{x})=\hat{c}+\sum_{\alpha=1}^d\hat{m}_\alpha(x_\alpha)$ for $\hat{c}\in\mathbb{R}$ and $\hat{m}_\alpha(x_\alpha)\in G_{n,\alpha}^0$. Next define the log-likelihood function

$$\hat{l}(a)=\frac{1}{n}\sum_{i=1}^n[Y_i\{a+\hat{m}_{\text{-}1}(\mathbf{X}_{i\text{-}1})\}-b\{a+\hat{m}_{\text{-}1}(\mathbf{X}_{i\text{-}1})\}]K_h(X_{i1}-x_1) \tag{4.1}$$

where $\hat{m}_{\text{-}1}(\mathbf{X}_{i\text{-}1})=\hat{c}+\sum_{\alpha=2}^d\hat{m}_\alpha(X_{i\alpha})$. Define the SBK estimator as:

$$\hat{m}_{\text{SBK},1}(x_1)=\operatorname{argmax}_{a\in A}\hat{l}(a). \tag{4.2}$$

THEOREM 4. *Under Assumptions (A1)-(A7), as $n\to\infty$, $\hat{m}_{\text{SBK},1}(x_1)$ is oracally efficient,*

$$\sup_{x_1\in[0,1]}|\hat{m}_{\text{SBK},1}(x_1)-\tilde{m}_{\text{K},1}(x_1)|=\mathcal{O}_{a.s.}\left(n^{-1/2}\log n\right).$$

Theorem 4 follows from (A.29), Lemmas A.15 and A.16. The following corollary is a consequence of Theorems 1, 2 and 4.

COROLLARY 1. *Under Assumptions (A1)-(A7), as $n \to \infty$, the SBK estimator $\hat{m}_{\mathrm{SBK},1}(x_1)$ given in (4.2) satisfies*

$$\sup_{x_1 \in [h,1-h]} \left|\hat{m}_{\mathrm{SBK},1}(x_1) - m_1(x_1)\right| = \mathcal{O}_{a.s.}\left(\log n/\sqrt{nh}\right)$$

and for any $x_1 \in [h, 1-h]$, with $\mathrm{bias}_1(x_1)$ *as in (3.4) and $D_1(x_1)$ in (3.3)*

$$\sqrt{nh}\left\{\hat{m}_{\mathrm{SBK},1}(x_1) - m_1(x_1) - \mathrm{bias}_1(x_1)\, h^2/D_1(x_1)\right\}$$
$$\xrightarrow{\mathcal{L}} N\left(0, D_1(x_1)^{-1} v_1^2(x_1)\, D_1(x_1)^{-1}\right).$$

Define next the spline-backfitted estimator $\hat{c} = \mathrm{argmax}_{a\in A}\, \hat{l}_c(a)$ with $\hat{l}_c(a) = n^{-1}\sum_{i=1}^{n}\left[Y_i\left\{a + \hat{m}_{\text{-}c}(\mathbf{X}_i)\right\} - b\left\{a + \hat{m}_{\text{-}c}(\mathbf{X}_i)\right\}\right]$ in which $\hat{m}_{\text{-}c}(\mathbf{X}_i) = \sum_{\alpha=1}^{d}\hat{m}_\alpha(X_{i\alpha})$. Similar to Theorem 4, the main result shows that the difference between $\hat{c}$ and its infeasible counterpart $\tilde{c}$ is asymptotically negligible.

THEOREM 5. *Under Assumptions (A1)-(A5) and (A7), as $n \to \infty$, $\hat{c}$ is oracally efficient, i.e., $\sqrt{n}\,(\hat{c} - \tilde{c}) \xrightarrow{p} 0$ and hence*

$$\sqrt{n}\,(\hat{c} - c) \xrightarrow{\mathcal{L}} N\left(0, a(\phi)^{1/2}\left[\mathsf{E}\, b''\left\{m(\mathbf{X})\right\}\right]^{-1/2}\right).$$

5. Implementation.

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

$$N = N_n = \min\left(\left\lfloor n^{1/4}\log n\right\rfloor + 1, \lfloor n/4d - 1/d\rfloor - 1\right)$$

which satisfies (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 is less than $n/4$, i.e., $1 + d(N+1) \le n/4$. For more discussion, see Portnoy (1997).

According to Corollary 1, the asymptotic distribution of the estimator $\hat{m}_{\mathrm{SBK},\alpha}(x_\alpha)$ depends 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. Define the optimal bandwidth of h_α, denoted by $h_{\alpha,\mathrm{opt}}$, as the minimizer of the asymptotic mean integrated squared errors (AMISE) of $\{\hat{m}_\alpha(x_a), \alpha = 1, \ldots, d\}$:

$$\begin{aligned}\mathrm{AMISE}(\hat{m}_\alpha) \quad = \quad & \int \left[\left\{\mathrm{bias}_\alpha(x_\alpha)\, h_\alpha^2/D_\alpha(x_\alpha)\right\}^2\right. \\ & \left. + D_\alpha(x_\alpha)^{-1} v_\alpha^2(x_\alpha)\, D_\alpha(x_\alpha)^{-1}/(nh_\alpha)\right] f_\alpha(x_\alpha)\, dx_\alpha.\end{aligned}$$

By letting $d\,\mathrm{AMISE}\left(\hat{m}_\alpha\right)/dh_\alpha = 0$, one obtains an optimal bandwidth $h_{\alpha,\mathrm{opt}}$:

$$h_{\alpha,\mathrm{opt}} = \left\{ \frac{n^{-1}\int D_\alpha\left(x_\alpha\right)^{-1} v_\alpha^2\left(x_\alpha\right) D_\alpha\left(x_\alpha\right)^{-1} f_\alpha\left(x_\alpha\right)dx_\alpha}{4\int\left\{\mathrm{bias}_\alpha\left(x_\alpha\right)/D_\alpha\left(x_\alpha\right)\right\}^2 f_\alpha\left(x_\alpha\right)dx_\alpha} \right\}^{1/5},$$

which is approximated by

$$\hat{h}_{\alpha,\mathrm{opt}} = \left\{ \frac{n^{-1}\sum_{i=1}^n D_\alpha\left(X_{i\alpha}\right)^{-1} v_\alpha^2\left(X_{i\alpha}\right) D_\alpha\left(X_{i\alpha}\right)^{-1}}{4\sum_{i=1}^n\left\{\mathrm{bias}_\alpha\left(X_{i\alpha}\right)/D_\alpha\left(X_{i\alpha}\right)\right\}^2} \right\}^{1/5},$$

where

$$D_\alpha\left(x_\alpha\right) = f_\alpha\left(x_\alpha\right)\mathsf{E}\left[b''\left\{m\left(\mathbf{X}\right)\right\}|X_\alpha = x_\alpha\right]$$

and

$$\begin{aligned} v_\alpha^2\left(x_\alpha\right) &= f_\alpha\left(x_\alpha\right)\mathsf{E}\left\{\sigma^2\left(\mathbf{X}\right)|X_\alpha = x_\alpha\right\}\left\|K\right\|_2^2, \\ \mathrm{bias}_\alpha\left(x_\alpha\right) &= \mu_2\left(K\right)\Big\{m''_\alpha\left(x_\alpha\right)f\left(x_\alpha\right)\mathsf{E}\left[b''\left\{m\left(\mathbf{X}\right)\right\}|X_\alpha = x_\alpha\right] \\ &\quad + m'_\alpha\left(x_\alpha\right)f\left(x_\alpha\right)\frac{\partial}{\partial x_\alpha}\mathsf{E}\left[b''\left\{m\left(\mathbf{X}\right)\right\}|X_\alpha = x_\alpha\right] \\ &\quad - \left\{m'_\alpha\left(x_\alpha\right)\right\}^2 f\left(x_\alpha\right)\mathsf{E}\left[b'''\left\{m\left(\mathbf{X}\right)\right\}|X_\alpha = x_\alpha\right]\Big\}. \end{aligned}$$

The following estimation methods for the terms $m'_\alpha\left(x_\alpha\right)$, $m''_\alpha\left(x_\alpha\right)$, $f_\alpha\left(x_\alpha\right)$, $\mathsf{E}\left\{\sigma^2\left(\mathbf{X}\right)|X_\alpha = x_\alpha\right\}$, $\mathsf{E}\left[b''\left\{m\left(\mathbf{X}\right)\right\}|X_\alpha = x_\alpha\right]$, $\mathsf{E}\left[b'''\left\{m\left(\mathbf{X}\right)\right\}|X_\alpha = x_\alpha\right]$ and $\frac{\partial}{\partial x_\alpha}\mathsf{E}\left[b''\left\{m\left(\mathbf{X}\right)\right\}|X_\alpha = x_\alpha\right]$ are proposed. The final bandwidth is denoted as $\hat{h}_{\alpha,\mathrm{opt}}$.

1). The derivative functions $m'_\alpha\left(X_{i\alpha}\right)$ and $m''_\alpha\left(X_{i\alpha}\right)$ are estimated as $\sum_{k=1}^3 k\hat{a}_{\alpha,l,k}X_{i\alpha}^{k-1} + 3\sum_{k=4}^{N+3}\hat{a}_{\alpha,l,k}\left(X_{i1} - t_{\alpha,k-3}\right)^2$ and $\sum_{k=2}^3 k\left(k-1\right)\hat{a}_{\alpha,l,k}X_{i\alpha}^{k-2} + 6\sum_{k=4}^{N+3}\hat{a}_{\alpha,l,k}\left(X_{i1} - t_{\alpha,k-3}\right)$ where $\left\{\hat{a}_{\alpha,l,k}\right\}_{k=0}^{N+3}$ maximize:

$$\begin{aligned} &\textstyle\sum_{i=1}^n\Big[Y_i\left\{\sum_{k=0}^3 a_{\alpha,l,k}X_{i\alpha}^k + \sum_{k=4}^{N+3}a_{\alpha,l,k}\left(X_{i\alpha} - t_{\alpha,k-3}\right)^3\right\} \\ &\quad -b\left\{\textstyle\sum_{k=0}^3 a_{\alpha,l,k}X_{i\alpha}^k + \displaystyle\sum_{k=4}^{N+3}a_{\alpha,l,k}\left(X_{i\alpha} - t_{\alpha,k-3}\right)^3\right\}\Big] \end{aligned}$$

where $\min_i X_{i\alpha} = t_{\alpha,0} < \cdots < t_{\alpha,N+1} = \max_i X_{i\alpha}$.

2). $\mathsf{E}\left[b''\left\{m\left(\mathbf{X}\right)\right\}|X_\alpha = x_\alpha\right]$ 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

$$\textstyle\sum_{i=1}^n\left[b''\left\{\hat{m}\left(\mathbf{X}_i\right)\right\} - \left\{\sum_{k=0}^3 a_{\alpha,l,k}X_\alpha^k + \sum_{k=4}^{N+3}a_{\alpha,l,k}\left(X_\alpha - t_{k-3}\right)^3\right\}\right]^2,$$

$\frac{\partial}{\partial x_\alpha}\mathsf{E}\left[b''\left\{m\left(\mathbf{X}\right)\right\}|X_\alpha=x_\alpha\right]$ and $\mathsf{E}\left[b'''\left\{m\left(\mathbf{X}\right)\right\}|X_\alpha=x_\alpha\right]$ 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
$\sum_{i=1}^{n}\left[b'''\left\{\hat{m}\left(\mathbf{X}_i\right)\right\}-\left\{\sum_{k=0}^{3}a_{\alpha,l,k}X_\alpha^k+\sum_{k=4}^{N+3}a_{\alpha,l,k}\left(X_\alpha-t_{k-3}\right)^3\right\}\right]^2$.

3). $\mathsf{E}\left\{\sigma^2\left(\mathbf{X}\right)|X_\alpha=x_\alpha\right\}$ 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

$$\sum_{i=1}^{n}\left(\left[Y_i-b'\left\{\hat{m}\left(\mathbf{X}_i\right)\right\}\right]^2-\left\{\sum_{k=0}^{3}a_{\alpha,l,k}X_\alpha^k+\sum_{k=4}^{N+3}a_{\alpha,l,k}\left(X_\alpha-t_{k-3}\right)^3\right\}\right)^2.$$

4). The density function $f_\alpha\left(x_\alpha\right)$ is estimated by $n^{-1}\sum_{i=1}^{n}K_{h_\alpha}\left(X_{i\alpha}-x_\alpha\right)$ with the rule-of-the-thumb bandwidth h_α.

6. Examples. We have applied the estimation procedure described in the previous section to both simulated (Example 1 and 2) and real (Example 3) data.

6.1. *Example 1.* The data are generated from the model

$$\mathrm{P}(Y=1|\mathbf{X}=\mathbf{x})=b'\left\{c+\sum_{\alpha=1}^{d}m_\alpha\left(X_\alpha\right)\right\},b'\left(x\right)=\frac{e^x}{1+e^x}$$

with $d=5,c=0,m_1\left(x\right)=\sin\left(\pi x\right)$, $m_2\left(x\right)=\Phi\left(3x\right)$ and $m_3\left(x\right)=m_4\left(x\right)=m_5\left(x\right)=x$, where Φ is the standard normal distribution function. The predictors are generated by transforming the following vector autoregression (VAR) equation for $0\leq a,r<1$,

$$\begin{aligned}X_{t\alpha}&=\Phi\left(\sqrt{1-a^2}Z_{t\alpha}\right),2\leq t\leq n,1\leq\alpha\leq d\\ \mathbf{Z}_t&=a\mathbf{Z}_{t-1}+\boldsymbol{\varepsilon}_i,\boldsymbol{\varepsilon}_i\sim N\left(0,\Sigma\right),2\leq t\leq n,\Sigma=\left(1-r\right)\mathbf{I}_{d\times d}+r\mathbf{1}_d\mathbf{1}_d^{\mathsf{T}},\end{aligned}$$

with stationary $\mathbf{Z}_t=\left(Z_{t1},...,Z_{td}\right)^{\mathsf{T}}\sim N\left\{0,\left(1-a^2\right)^{-1}\Sigma\right\}$, $\mathbf{1}_d=\left(1,...,1\right)^{\mathsf{T}}$ and $\mathbf{I}_{d\times d}$ is the $d\times d$ identity matrix. Higher values of a correspond to stronger dependence among the observations, and in particular, if $a=0$, the data are i.i.d. The r controls the correlation of the 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 $x_{\alpha,\min}^i$, $x_{\alpha,\max}^i$ denote the smallest and largest observations of the variable x_α in the i -th replication. The component functions $\left\{m_\alpha\right\}_{\alpha=1}^{d}$ are estimated on sample values.

Denoting the estimator of m_α in the k-th sample as $\hat{m}_{\mathrm{SBK},\alpha,k}$ and $X_{t\alpha,k}$ accordingly. We define the (mean) integrated squared error (ISE and MISE):

$$\begin{aligned} \mathrm{ISE}(\hat{m}_{\mathrm{SBK},\alpha,k}) &= n^{-1}\textstyle\sum_{t=1}^{n}\left\{\hat{m}_{\mathrm{SBK},\alpha,k}(X_{t\alpha,k}) - m_\alpha(X_{t\alpha,k})\right\}^2, \\ \mathrm{MISE}(\hat{m}_{\mathrm{SBK},\alpha}) &= \frac{1}{100}\textstyle\sum_{k=1}^{100}\mathrm{ISE}(\hat{m}_{\mathrm{SBK},\alpha,k}). \end{aligned}$$

In order to show the SBK estimator's efficiency relative to the "oracle smoother" $\tilde{m}_{\mathrm{K},\alpha}(x_\alpha)$, define the empirical relative efficiency of $\hat{m}_{\mathrm{SBK},\alpha}(x_\alpha)$ with respect to $\tilde{m}_{\mathrm{K},\alpha}(x_\alpha)$ as

$$\mathrm{EFF}_\alpha = \left[\frac{\sum_{t=1}^{n}\left\{\tilde{m}_{\mathrm{K},\alpha}(x_\alpha) - m_\alpha(X_{t\alpha})\right\}^2}{\sum_{t=1}^{n}\left\{\hat{m}_{\mathrm{SBK},\alpha}(X_{t\alpha}) - m_\alpha(X_{t\alpha})\right\}^2}\right]^{1/2}.$$

Tables 1 and 2 show $\overline{\mathrm{EFF}}(\cdot)$ and $\mathrm{std}\{\mathrm{EFF}(\cdot)\}$, which are the means and standard deviations of the MISEs and EFFs of $\hat{m}_{\mathrm{SBK},\alpha}$ and $\tilde{m}_{\mathrm{K},\alpha}$ for $\alpha = 1, 2$. It is apparent that the SBK estimator performs as good as the oracle estimator, see Theorem 4.

(Insert Table 1 about here)
(Insert Table 2 about here)

6.2. *Example 2.* Using the same model in Example 1 but with a higher dimension $d = 10$, where $m_\alpha(x) = \sin(\pi x)$, $\alpha = 1, ..., 10$ 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 3. As expected, increases in sample size reduce MISE for both estimators and across all combinations of r and α values.

(Insert Table 3 about here)

The convergence properties are displayed in Figure 1 (a) showing the kernel density estimator of the simulated 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 moving towards the standard line 1.0 with a narrower spread when sample size increases, which confirms the result of Theorem 4. The basic graphic pattern of Figure 1 (b) with $r = 0.5$, $a = 0.5$ is similar to that for the i.i.d case, though with slightly slower convergent and slightly poorer efficient.

(Insert Figure 1 about here)

To have an 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 2 and 3. The results are satisfactory and show that the theory works in practice, and that performance improves with increasing sample size.

(Insert Figure 2 about here)
(Insert Figure 3 about here)

6.3. *Example 3.* We have applied the estimation to the dataset comes from the credit reform database provided by the Research Data Center (RDC) of the Humboldt Universität zu Berlin. After we exclude the missing values, it contains financial information from 18610 solvent ($y = 0$) and 1000 insolvent ($y = 1$) German companies. The time period ranges from 1997 to 2002 and in the case of the insolvent companies the information was gathered 2 years before the insolvency took place. For more details, see Härdle et al. (2010). The financial ratios we use are showed in Table 4.

(Insert Table 4 about here)

In order to satisfy (A4), we make following transformation: $X_{i\alpha} = F_{n\alpha}(Z_{i\alpha})$, $\alpha = 1, ..., 8$, where $F_{n\alpha}$ is the empirical cdf for the data $\{X_{i\alpha}\}_{i=1}^{n}$. We measure the quality of the estimation by Accuracy Ratio (AR), which is the ratio of two areas. The first one is the area between the Cumulative Accuracy Profile (CAP) curve and the diagonal line, and the second one is the area between the perfect model CAP curve and the diagonal. The second area is close to 1/2 in this example, so we have $\text{AR} \approx 2\int_0^1 \text{CAP}(x)\,dx - 1$.

As a result, our model has the AR value 62.66%. We can also estimate the functions $m_\alpha(x)$ for X_α. For example, if we are interested in the effects of Ebit/Total_Assets and log (Total_Assets), we can obtain the estimations for $m_3(x)$ and $m_8(x)$, which are showed in Figure 4.

(Insert Figure 4 about here)

It is not a surprise that the estimation for $m_8(x)$ decreases as x value increases. It means that a company with more Total_Assets has smaller probability of insolvent. While as x value increases, the estimation for $m_3(x)$ increases for most part but decreases at the end. So generally, those companies with higher Ebit/Total_Assets ratio have more probability of insolvent. But it looks like that those companies with extremely high Ebit/Total_Assets ratio have less probability of insolvent. It is an interesting topic to figure out the reason.

APPENDIX A: APPENDIX SECTION

A.1. Preliminaries. In the proofs that follow, we use "$\mathcal{U}$" and "$\mathcal{u}$" to denote sequences of random variables that are uniformly "$\mathcal{O}$" and "$\mathcal{o}$ " of certain order.

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

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

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

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

where

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

Denote 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

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

so that $\mathsf{E}\,B_{J,\alpha}(X_\alpha) = 0$, $\mathsf{E}\,B_{J,\alpha}^2(X_\alpha) = 1$.

LEMMA A.3. *(Wang and Yang (2007), Theorem A.2) Under Assumptions (A1)-(A5) and (A7), one has:*

(i) Constants $c_0\left(f\right)$, $C_0(f)$, $c_1\left(f\right)$ and $C_1(f)$ exist depending on the marginal densities $f_\alpha\left(x_\alpha\right), 1\leq\alpha\leq d$, such that $c_0\left(f\right)H\leq c_{J,\alpha}\leq C_0\left(f\right)H$ and

$$c_1\left(f\right)H\leq\left\|b_{J,\alpha}\right\|_2^2\leq C_1(f)H. \tag{A.2}$$

(ii) uniformly for $J, J'=1,...,N+1$

$$\mathsf{E}\left\{B_{J,\alpha}\left(X_{i\alpha}\right)B_{J',\alpha}\left(X_{i\alpha}\right)\right\}\sim\left\{\begin{array}{ll} 1 & J'=J \\ -1/3 & \left|J'-J\right|=1 \\ 1/6 & \left|J'-J\right|=2 \\ 0 & \left|J'-J\right|>2 \end{array}\right.$$

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

LEMMA A.4. *(De Boor (2001), p.149) A constant $C_\infty>0$ exists such that for any $m\in C^1\left[0,1\right]$ with $m'\in\mathrm{Lip}\left(\left[0,1\right],C_\infty\right)$, there is a function $g\in G_n^{(0)}\left[0,1\right]$ such that $\left\|g-m\right\|_\infty\leq C_\infty H^2$.*

LEMMA A.5. *(Wang and Yang (2007), Lemma A.2) Constants $c_0, C_0>0$ exist such that for any $\boldsymbol{\lambda}=\left(\lambda_0,\lambda_{J,\alpha}\right)^{\mathsf{T}}_{1\leq J\leq N+1,1\leq\alpha\leq d}\in\mathbb{R}^{1+d(N+1)}$,*

$$c_0\left(\lambda_0^2+\sum\nolimits_{J,\alpha}^2\lambda_{J,\alpha}^2\right)\leq\left\|\lambda_0+\sum\nolimits_{J,\alpha}\lambda_{J,\alpha}B_{J,\alpha}\right\|_2^2\leq C_0\left(\lambda_0^2+\sum\nolimits_{J,\alpha}^2\lambda_{J,\alpha}^2\right).$$

LEMMA A.6. *(Xue and Yang (2006a), Lemma A.4) Under Assumptions (A2), (A4) and (A6), as $n\to\infty$, the uniform supremum of the rescaled difference between $\left\langle g_1,g_2\right\rangle_{2,n}$ and $\left\langle g_1,g_2\right\rangle_2$ is*

$$A_n=\sup_{g_1,g_2\in G_n^{(0)}[0,1]}\frac{\left|\left\langle g_1,g_2\right\rangle_{2,n}-\left\langle g_1,g_2\right\rangle_2\right|}{\left\|g_1\right\|_2\left\|g_2\right\|_2}=\mathcal{O}_{a.s.}\left(\frac{\log n}{n^{1/2}H^{1/2}}\right).$$

A.2. Oracle smoothers.

LEMMA A.7. *Under Assumptions (A1)-(A6), as $n\to\infty$,*

$$\sup_{x_1\in[h,1-h]}\left|\tilde{l}'\left\{m_1\left(x_1\right)\right\}-\mathrm{bias}_1\left(x_1\right)h^2-n^{-1}\sum\nolimits_{i=1}^n K_h\left(X_{i1}-x_1\right)\sigma\left(\mathbf{X}_i\right)\varepsilon_i\right|$$

$$=\mathcal{O}_{a.s.}\left(n^{-1/2}h^{1/2}\log n\right)$$

where $\mathrm{bias}_1\left(x_1\right)$ is defined in (3.4).

Proof. According to (3.1) and (1.3), $\tilde{l}'\left\{m_1\left(x_1\right)\right\}$ is

$$\begin{aligned}
\text{(A.3)} \quad & n^{-1}\textstyle\sum_{i=1}^{n}\left[Y_i - b'\left\{m_1\left(x_1\right)+m_{-1}\left(\mathbf{X}_{i,-1}\right)\right\}\right]K_h\left(X_{i1}-x_1\right) \\
= \quad & n^{-1}\textstyle\sum_{i=1}^{n}\left[b'\left\{m\left(\mathbf{X}_i\right)\right\}-b'\left\{m_1\left(x_1\right)+m_{-1}\left(\mathbf{X}_{i,-1}\right)\right\}+\sigma\left(\mathbf{X}_i\right)\varepsilon_i\right] \\
& K_h\left(X_{i1}-x_1\right)
\end{aligned}$$

Let $\xi_{i,n}=\xi_{i,n}\left(x_1\right)=\xi_{i,n,1}+\xi_{i,n,2}$ in which

$$\begin{aligned}
\xi_{i,n,1}\left(x_1\right) \quad = \quad & \left[b'\left\{m\left(\mathbf{X}_i\right)\right\}-b'\left\{m_1\left(x_1\right)+m_{-1}\left(\mathbf{X}_{i,-1}\right)\right\}\right]K_h\left(X_{i1}-x_1\right) \\
& -\mathsf{E}\left[\left[b'\left\{m\left(\mathbf{X}_i\right)\right\}-b'\left\{m_1\left(x_1\right)+m_{-1}\left(\mathbf{X}_{i,-1}\right)\right\}\right]K_h\left(X_{i1}-x_1\right)\right],
\end{aligned}$$

$$\text{(A.4)} \qquad \xi_{i,n,2}=\xi_{i,n,2}\left(x_1\right)=\sigma\left(\mathbf{X}_i\right)\varepsilon_i K_h\left(X_{i1}-x_1\right).$$

Then according to (A.3), one can rewrite $l^{*\prime}\left\{m_1\left(x_1\right)\right\}$ as

$$n^{-1}\textstyle\sum_{i=1}^{n}\xi_{i,n}+\mathsf{E}\left[b'\left\{m\left(\mathbf{X}_i\right)\right\}-b'\left\{m_1\left(x_1\right)+m_{-1}\left(\mathbf{X}_{i,-1}\right)\right\}\right]K_h\left(X_{i1}-x_1\right).$$

The deterministic term is

$$\begin{aligned}
& \mathsf{E}\left[b'\left\{m\left(\mathbf{X}_i\right)\right\}-b'\left\{m_1\left(x_1\right)+m_{-1}\left(\mathbf{X}_{i,-1}\right)\right\}\right]K_h\left(X_{i1}-x_1\right) \\
= \quad & \int_{\boldsymbol{\chi}}\left[b'\left\{m\left(\mathbf{u}\right)\right\}-b'\left\{m_1\left(x_1\right)+m_{-1}\left(\mathbf{u}_{-1}\right)\right\}\right]h^{-1}K\left(\frac{u_1-x_1}{h}\right)f\left(\mathbf{u}\right)d\mathbf{u} \\
= \quad & \int_{\boldsymbol{\chi}}\left[b''\left\{m\left(x_1,\mathbf{u}_{-1}\right)\right\}\left\{m_1\left(u_1\right)-m_1\left(x_1\right)\right\}\right. \\
& \left.+\frac{1}{2}b'''\left\{m\left(x_1,\mathbf{u}_{-1}\right)\right\}\left\{m_1\left(u_1\right)-m_1\left(x_1\right)\right\}^2+u\left(h^2\right)\right] \\
& h^{-1}K\left(\frac{u_1-x_1}{h}\right)f\left(u_1,\mathbf{u}_{-1}\right)du_1 d\mathbf{u}_{-1}+u\left(h^2\right) \\
= \quad & \int_{[0,1]^{d-1}}\int_{[-1,1]}\left[b''\left\{m\left(x_1,\mathbf{u}_{-1}\right)\right\}\left\{hv_1 m_1'\left(x_1\right)+\frac{\left(hv_1\right)^2}{2}m_1''\left(x_1\right)+u\left(h^2\right)\right\}\right. \\
& \left.+\frac{1}{2}b'''\left\{m\left(x_1,\mathbf{u}_{-1}\right)\right\}\left\{hv_1 m_1'\left(x_1\right)+\left(hv_1\right)^2 m_1''\left(x_1\right)+u\left(h^2\right)\right\}^2\right] \\
& K\left(v_1\right)\left\{f\left(x_1,\mathbf{u}_{-1}\right)+hv_1\frac{\partial f\left(x_1,\mathbf{u}_{-1}\right)}{\partial x_1}+\mathcal{U}\left(h^2\right)\right\}dv_1 d\mathbf{u}_{-1}+u\left(h^2\right)
\end{aligned}$$

which equals

$$\begin{aligned}
& h^2\int_{[-1,1]}v_1^2K\left(v_1\right)dv_1\left\{\frac{m_1''\left(x_1\right)f_1\left(x_1\right)}{2}\int_{[0,1]^{d-1}}b''\left\{m\left(x_1,\mathbf{u}_{-1}\right)\right\}f\left(\mathbf{u}|x_1\right)d\mathbf{u}_{-1}\right. \\
& \left.+m_1'\left(x_1\right)\int_{[0,1]^{d-1}}b''\left\{m\left(x_1,\mathbf{u}_{-1}\right)\right\}\frac{\partial f\left(x_1,\mathbf{u}_{-1}\right)}{\partial x_1}d\mathbf{u}_{-1}\right\}+u\left(h^2\right).
\end{aligned}$$

$$\begin{aligned}
&= h^2\mu_2(K)\left\{m_1''(x_1)\,f(x_1)\,\mathsf{E}\left[b''\{m(\mathbf{X})\}\,|X_1=x_1\right]\right.\\
&\quad +m_1'(x_1)\frac{\partial}{\partial x_1}\left[f(x_1)\,\mathsf{E}\left[b''\{m(\mathbf{X})\}\,|X_1=x_1\right]\right]\\
&\quad \left.-\{m_1'(x_1)\}^2 f(x_1)\,\mathsf{E}\left[b'''\{m(\mathbf{X})\}\,|X_1=x_1\right]\right\}+u\left(h^2\right)\\
&= \mathrm{bias}_1(x_1)\,h^2+u\left(h^2\right).
\end{aligned}$$

Using the above

$$\begin{aligned}
\mathsf{E}\,\xi_{i,n,1}^2 &= h^{-2}\int_{[0,1]^d}\left[b'\{m(\mathbf{u})\}-b'\{m_1(x_1)+m_{-1}(\mathbf{u}_{-1})\}\right]^2\\
&\quad K\left(\frac{u_1-x_1}{h}\right)^2 f(\mathbf{u})\,d\mathbf{u}+\mathcal{U}\left(h^4\right)\\
&= h^{-1}\int_{[0,1]^{d-1}}\int_{[-1,1]}\left[b'\{m(x_1+hv_1,\mathbf{u}_{-1})\}-b'\{m_1(x_1)+m_{-1}(\mathbf{u}_{-1})\}\right]^2\\
&\quad K(v_1)^2 f(x_1+hv_1,\mathbf{u}_{-1})\,dv_1d\mathbf{u}_{-1}+\mathcal{U}\left(h^4\right)\\
&= h^{-1}\int_{[0,1]^{d-1}}\int_{[-1,1]}\left[b''\{m(x_1,\mathbf{u}_{-1})\}\left\{hv_1m_1'(x_1)+\mathcal{U}\left(h^2\right)\right\}\right]^2\\
&\quad K(v_1)^2\{f(x_1,\mathbf{u}_{-1})+\mathcal{U}(h)\}\,dv_1d\mathbf{u}_{-1}+\mathcal{U}\left(h^4\right)=\mathcal{U}(h).
\end{aligned}$$

Note that $\sup_{x_1}|b'\{m(\mathbf{X}_i)\}-b'\{m_1(x_1)+m_{-1}(\mathbf{X}_{i,-1})\}|\le C_bh$ whenever $K_h(X_{i1}-x_1)\neq 0$, hence $\mathsf{E}\left|\xi_{i,n,1}\right|^k\le(2C_bh)^{k-2}\,\mathsf{E}\,\xi_{i,n,1}^2$ so applying Lemma A.2 implies that $\sup_{x_1\in[h,1-h]}\left|n^{-1}\sum_{i=1}^n\xi_{i,n,1}\right|=\mathcal{O}_{a.s.}\left\{h^{1/2}n^{-1/2}\log n\right\}$. □

LEMMA A.8. *Under Assumptions (A2), (A4)-(A6), as $n\to\infty$*

$$\sup_{x_1\in[h,1-h]}\left|\tilde{l}''(m_1(x_1))+D_1(x_1)\right|=\mathcal{O}_{a.s.}\left(\log n/\sqrt{nh}\right),$$

where $D_1(x_1)$ is defined in (3.3).

PROOF. See Liu et al. (2011). □

LEMMA A.9. *Under Assumptions (A1) to (A3), (A5) and (A7), a constant C exists such that, as $n\to\infty$*

$$\sup_{x_1\in[h,1-h]}\left|\mathsf{Cov}\left(\xi_{i,n},\xi_{j,n}\right)\right|\le Ch^{-\frac{1+\eta}{2+\eta}}\alpha(j-i)^{\frac{\eta}{2+\eta}}\ \ \textit{for } i\neq j$$

Proof. According to Davydov's inequality, for $\frac{1}{p}+\frac{1}{q}+\frac{1}{r}=1$, $\mathsf{Cov}\left(\xi_{i,n},\xi_{j,n}\right)$ is bounded by

$$\begin{aligned} & C_2\left\{2\alpha\left(j-i\right)\right\}^{1/p}\left\|\xi_{i,n,1}+\xi_{i,n,2}\right\|_q\left\|\xi_{j,n,1}+\xi_{j,n,2}\right\|_r \\ \leq \quad & C_2\left\{2\alpha\left(j-i\right)\right\}^{1/p}\left(\left\|\xi_{i,n,1}\right\|_q+\left\|\xi_{i,n,2}\right\|_q\right)\left(\left\|\xi_{j,n,1}\right\|_r+\left\|\xi_{j,n,2}\right\|_r\right) \end{aligned}$$

Let $q=r=2+\eta, p=1+2/\eta$, where η takes value in the (A3), then one has $\left\|\xi_{i,n,1}\right\|_q=\mathcal{U}\left(h^{-\frac{1}{2+\eta}}\right)$ and $\left\|\xi_{i,n,1}\right\|_q=\mathcal{U}\left(h^{-\frac{1+\eta}{2+\eta}}\right)$. $\mathsf{Cov}\left(\xi_{i,n,l'},\xi_{j,n,l''}\right)\leq Ch^{-\frac{1+\eta}{2+\eta}}\alpha\left(j-i\right)^{\frac{\eta}{2+\eta}}$ for some constant C. □

Proof of Theorem 1 and Theorem 2. The Mean Value Theorem ensures the existence of 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 $\tilde{l}'\left\{\tilde{m}_{\mathrm{K},1}\left(x_1\right)\right\}=0$ yielding

$$\tilde{m}_{\mathrm{K},1}\left(x_1\right)-m_1\left(x_1\right)=-\frac{\tilde{l}'\left(m_1\left(x_1\right)\right)}{\tilde{l}''\left(\bar{m}_1\left(x_1\right)\right)}. \tag{A.5}$$

Lemma A.8, Lemma A.7 and (A.5) then imply Theorem 1.

Let $S_n=S_n\left(x_1\right)=\sum_{i=1}^n\xi_{i,n}$, where $\xi_{i,n}$ is defined as (A.4). Note that $\mathsf{E}\,S_n=0$ and $\tilde{l}'\left\{m_1\left(x_1\right)\right\}=S_n/n+b\left(x_1\right)h^2+u\left(h^2\right)$.

$$\gamma\left(k\right)=\gamma\left(k,x_1\right)=\mathsf{Cov}\left(\xi_{i,n},\xi_{i+k,n}\right)$$

$$\begin{aligned} \sigma_n^2 &= \mathsf{E}\,S_n^2=\mathsf{Var}\left(S_n\right)=\mathsf{Var}\left(\textstyle\sum_{i=1}^n\xi_{i,n}\right) \\ &= \textstyle\sum_{i=1}^n\mathsf{Var}\left(\xi_{i,n}\right)+\sum_{i\neq j}^n\mathsf{Cov}\left(\xi_{i,n},\xi_{j,n}\right) \\ &= n\,\mathsf{Var}\left(\xi_{i,n}\right)+n\textstyle\sum_{1\leq|k|\leq n-1}\left(1-\dfrac{|k|}{n}\right)\gamma\left(k\right) \\ &= n\,\mathsf{Var}\left(\xi_{i,n}\right)+nA_n, \end{aligned}$$

where

$$\mathsf{Var}\left(\xi_{i,n}\right)=h^{-1}f_1\left(x_1\right)\mathsf{E}\left\{\sigma^2\left(\mathbf{X}\right)|X_1=x_1\right\}\|K\|_2^2+\mathcal{U}\left(h^4\right).$$

According to Lemma A.9, one has

$$|\gamma\left(k\right)|=\left|\mathsf{Cov}\left(\xi_{i,n},\xi_{i+k,n}\right)\right|\leq Ch^{-\frac{1+\eta}{2+\delta}}\alpha\left(k\right)^{\frac{\eta}{2+\eta}}.$$

Hence

$$\begin{aligned}
|A_n| &= \left|\sum\nolimits_{1\leq|l|\leq n-1}\gamma(k)\right| \\
&\leq \sum\nolimits_{1\leq|l|\leq n-1}\left(1-\frac{|k|}{n}\right)h^{-\frac{1+\eta}{2+\eta}}\left\{K_0\exp(-\lambda_0 k)\right\}^{\frac{\eta}{2+\eta}} \\
&\leq K_0 h^{-\frac{1+\eta}{2+\eta}}\sum\nolimits_{1\leq|l|\leq n-1}\exp\left\{-\lambda_0 k\eta/(2+\eta)\right\},
\end{aligned}$$

so a constant C_1 exists such that $A_n \leq C_1 h^{-\frac{1+\eta}{2+\eta}}$, and therefore $A_n/\mathsf{Var}\left(\xi_{i,n}\right) \to 0$ as $n\to\infty$. Since $\sigma_n^2 \sim n\,\mathsf{Var}\left(\xi_{i,n}\right) \geq c_0 n$ when n is large, according to (A.1) in Lemma A.1, constants c_1 and c_2 exist such that for some $0<\eta\leq 1$

$$\text{(A.6)}\quad \Delta_n = \sup_z\left|\mathrm{P}\left\{\sigma_n^{-1}S_n < z\right\}-\Phi(z)\right| \leq c_1\frac{d_\eta}{c_0\sigma_n^\eta}\left\{\log\left(\sigma_n/c_0^{1/2}\right)/\lambda\right\}^{1+\eta}$$

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 η in (A3), set $\lambda = 4(2+\eta)\eta^{-1}\log\left(\sigma_n/c_0^{1/2}\right)$, then by (A6), the d_η in (A.6) is

$$\begin{aligned}
d_\eta &= \max_{1\leq i\leq n}\Big(\mathsf{E}\Big|\left[b'\left\{m(\mathbf{X}_i)\right\}-b'\left\{m_1(x_1)+m_{\text{-}1}(\mathbf{X}_{i\text{-}1})\right\}+\sigma(\mathbf{X}_i)\varepsilon_i\right] \\
&\qquad K_h(X_{i1}-x_1)\Big|^{2+\eta}\Big) \\
&= \max_{1\leq i\leq n}\left\{\mathsf{E}\left|C_b h+\sigma(\mathbf{X}_i)\varepsilon_i\right|^{2+\eta}\left|K_h(X_{i1}-x_1)\right|^{2+\eta}\right\} \\
&\leq CC_\delta C_\eta\left\{\mathsf{E}\left|K_h(X_1-x_1)\right|^{2+\eta}\right\} = \mathcal{O}\left\{h^{-(1+\eta)}\right\},
\end{aligned}$$

i.e., $\Delta_n = \mathcal{O}\left\{h^{-(1+\eta)}/\sigma_n^\eta\right\} = \mathcal{O}\left\{n^{(1+\eta/2)/5-\eta/2}\right\} = \mathcal{O}\left(n^{1/5-2\eta/5}\right)\to 0$ when $1/2<\eta\leq 1$. So $S_n/\sigma_n \xrightarrow{\mathcal{L}} N(0,1)$, then

$$n\left[l^{*\prime}\left\{m_1(x_1)\right\}-\text{bias}_1(x_1)h^2\right]/\sqrt{nh^{-1}v_1^2(x_1)} \xrightarrow{\mathcal{L}} N(0,1),$$

where $v_1^2(x_1)$ is defined in (3.5). According to Theorem 1, one has as $n\to\infty$, $\sup_{x_1\in[h,1-h]}\left|\tilde{l}''\left\{m_1(x_1)\right\}-\tilde{l}''\left\{\bar{m}_1(x_1)\right\}\right|\to 0$ because $\sup_{x_1\in[h,1-h]}\left|m_1(x_1)-\bar{m}_1(x_1)\right|\to 0$. Then according to Slutsky's theorem:

$$\sqrt{nh}\left[\left\{\tilde{m}_{\mathrm{K},1}(x_1)-m_1(x_1)\right\}D_1(x_1)-\text{bias}_1(x_1)h^2\right]\to N\left(0,v_1^2(x_1)\right).$$

where $D_1(x_1)$ is defined in (3.3). □

Proof of Theorem 3. According to the Mean Value Theorem, a constant $\bar{c}$ between c and $\tilde{c}$ exists 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 (A2) and 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, $\tilde{l}'_c(\tilde{c}) = 0$ and

$$\begin{aligned}\tilde{l}'_c(c) &= n^{-1}\textstyle\sum_{i=1}^n \left[Y_i - b'\{c + m_{\_c}(\mathbf{X}_i)\}\right] \\ &= n^{-1}\textstyle\sum_{i=1}^n \sigma(\mathbf{X}_i)\,\varepsilon_i = \mathcal{O}_{a.s}\left(n^{-1/2}\log n\right)\end{aligned}$$

by Bernstein's Inequality. Similarly, $\tilde{l}''_c(c) = -n^{-1}\sum_{i=1}^n b''\{c + m_{\_c}(\mathbf{X}_i)\}$ converges to $-\mathsf{E}\,b''\{m(\mathbf{X})\}$ almost surely at the rate of $n^{-1/2}\log n$. These imply that $|\tilde{c} - c| = \mathcal{O}_{a.s.}\left(n^{-1/2}\log n\right)$ and plugging it into
$(\tilde{c} - c) = -\tilde{l}'_c(c)/\tilde{l}''_c(\bar{c})$, Theorem 3 is proved. □

A.3. Spline backfitted kernel estimators. In this section, we present the proof of Theorem 4. We write any $g \in G_n^0$ as $g = \boldsymbol{\lambda}^\mathsf{T}\mathbf{B}(\mathbf{X}_i)$ with vector $\boldsymbol{\lambda} = (\lambda_0, \lambda_{J,\alpha})^\mathsf{T}_{1 \le J \le N+1, 1 \le \alpha \le d} \in R^{N_d}$ where $N_d = (N+1)\,d + 1$ is the dimension of the additive spline space G_n^0, and

$$\mathbf{B}(\mathbf{x}) = \{1, B_{1,1}(x_1), ..., B_{N+1,d}(x_d)\}^\mathsf{T},$$

its standardized basis. We denote with a slight abuse of notation
$\hat{L}(g) = \hat{L}(\boldsymbol{\lambda}) = n^{-1}\sum_{i=1}^n \left[Y_i\boldsymbol{\lambda}^\mathsf{T}\mathbf{B}(\mathbf{X}_i) - b\left\{\boldsymbol{\lambda}^\mathsf{T}\mathbf{B}(\mathbf{X}_i)\right\}\right]$, which yields the gradient and Hessian formulae

$$\begin{aligned}\nabla\hat{L}(\boldsymbol{\lambda}) &= n^{-1}\textstyle\sum_{i=1}^n \left[Y_i\mathbf{B}(\mathbf{X}_i) - b'\left\{\boldsymbol{\lambda}^\mathsf{T}\mathbf{B}(\mathbf{X}_i)\right\}\mathbf{B}(\mathbf{X}_i)\right], \\ \nabla^2\hat{L}(\boldsymbol{\lambda}) &= -n^{-1}\textstyle\sum_{i=1}^n b''\left\{\boldsymbol{\lambda}^\mathsf{T}\mathbf{B}(\mathbf{X}_i)\right\}\mathbf{B}(\mathbf{X}_i)\,\mathbf{B}(\mathbf{X}_i)^\mathsf{T}.\end{aligned}$$

The multivariate function $m(\mathbf{x})$ is estimated by an additive spline function

$$\begin{aligned}\hat{m}(\mathbf{x}) &= \hat{m}_0 + \textstyle\sum_{\alpha=1}^d \hat{m}_\alpha(x_\alpha) = \hat{\boldsymbol{\lambda}}^\mathsf{T}\mathbf{B}(\mathbf{x}), \\ \hat{\boldsymbol{\lambda}} &= \left(\hat{\lambda}_0, \hat{\lambda}_{J,\alpha}\right)^\mathsf{T}_{\substack{1 \le \alpha \le d \\ 1 \le J \le N+1}} = \operatorname{argmax}_{\boldsymbol{\lambda}} \hat{L}(\boldsymbol{\lambda}).\end{aligned}$$

Lemma 14 of Stone (1986) ensures that with probability approaching 1, $\hat{\boldsymbol{\lambda}}$ exists uniquely and that $\nabla\hat{L}\left(\hat{\boldsymbol{\lambda}}\right) = \mathbf{0}$. In addition, Lemma A.4 and (A1) provide a vector $\bar{\boldsymbol{\lambda}}$ and an additive spline function $\bar{m}$ such that

$$\bar{m}(\mathbf{x}) = \bar{\boldsymbol{\lambda}}^\mathsf{T}\mathbf{B}(\mathbf{x}),\, \|\bar{m} - m\|_\infty \le C_\infty H^2. \tag{A.7}$$

We first establish technical lemmas before proving Theorems 4 and 5.

LEMMA A.10. *Under Assumptions (A1)-(A5) and (A7), as $n \to \infty$*

$$\begin{aligned}\left|\nabla \hat{L}\left(\bar{\boldsymbol{\lambda}}\right)\right| &= \mathcal{O}_{a.s.}\left(H^2 + n^{-1/2}\log n\right),\\ \left\|\nabla \hat{L}\left(\bar{\boldsymbol{\lambda}}\right)\right\| &= \mathcal{O}_{a.s.}\left(H^{3/2} + H^{-1/2}n^{-1/2}\log n\right).\end{aligned}$$

PROOF.

$$\begin{aligned}\nabla \hat{L}\left(\bar{\boldsymbol{\lambda}}\right) &= n^{-1}\textstyle\sum_{i=1}^n \left[Y_i\mathbf{B}\left(\mathbf{X}_i\right) - b'\left\{\bar{\boldsymbol{\lambda}}^{\mathsf{T}}\mathbf{B}\left(\mathbf{X}_i\right)\right\}\mathbf{B}\left(\mathbf{X}_i\right)\right]\\ &= n^{-1}\textstyle\sum_{i=1}^n \left[b'\left\{m\left(\mathbf{X}_i\right)\right\} - b'\left\{\bar{m}\left(\mathbf{X}_i\right)\right\} + \sigma\left(\mathbf{X}_i\right)\varepsilon_i\right]\mathbf{B}\left(\mathbf{X}_i\right)\end{aligned}$$

The first element of the above vector is $\frac{1}{n}\sum_{i=1}^n \left[\left[b'\left\{m\left(\mathbf{X}_i\right)\right\} - b'\left\{\bar{m}\left(\mathbf{X}_i\right)\right\}\right] + \sigma\left(\mathbf{X}_i\right)\varepsilon_i\right]$, which is $\mathcal{O}_{a.s.}\left(H^2 + n^{-1/2}\log n\right)$ according to Lemmas A.4 and A.2. The other elements can be written as

$$\begin{aligned}&n^{-1}\textstyle\sum_{i=1}^n \left[\xi_{i,J,\alpha,n} + \mathsf{E}\left[b'\left\{m\left(X_{i\alpha}\right)\right\} - b'\left\{\bar{m}\left(X_{i\alpha}\right)\right\}\right]\right.\\ &\left.B_{J,\alpha}\left(X_{i\alpha}\right) + \sigma\left(\mathbf{X}_i\right)\varepsilon_i B_{J,\alpha}\left(X_{i\alpha}\right)\right],\end{aligned}$$

where

$$\begin{aligned}\xi_{i,J,\alpha,n} &= \left[b'\left\{m\left(X_{i\alpha}\right)\right\} - b'\left\{\bar{m}\left(X_{i\alpha}\right)\right\}\right]B_{J,\alpha}\left(X_{i\alpha}\right)\\ &\quad - \mathsf{E}\left[\left[b'\left\{m\left(X_{i\alpha}\right)\right\} - b'\left\{\bar{m}\left(X_{i\alpha}\right)\right\}\right]B_{J,\alpha}\left(X_{i\alpha}\right)\right].\end{aligned}$$

According to (A.2) and (A.7), one has

$$\begin{aligned}&\left|\mathsf{E}\left[b'\left\{m\left(X_{i\alpha}\right)\right\} - b'\left\{\bar{m}\left(X_{i\alpha}\right)\right\}\right]B_{J,\alpha}\left(X_{i\alpha}\right)\right|\\ \leq\ &\mathsf{E}\left|b'\left\{m\left(X_{i\alpha}\right)\right\} - b'\left\{\bar{m}\left(X_{i\alpha}\right)\right\}\right|\frac{\left|b_{J,\alpha}\left(X_{i\alpha}\right)\right|}{\left\|b_{J,\alpha}\right\|_2}\\ \leq\ &c\left\|m-\bar{m}\right\|_\infty \max_{\substack{1\leq J\leq N+1\\ 1\leq\alpha\leq d}}\left\|b_{J,\alpha}\right\|_2^{-1}\max_{\substack{1\leq J\leq N+1\\ 1\leq\alpha\leq d}}\mathsf{E}\left|b_{J,\alpha}\left(X_{i\alpha}\right)\right|\\ =\ &\mathcal{O}\left(H^2\times H^{-1/2}\times H\right) = \mathcal{O}\left(H^{5/2}\right),\end{aligned}$$

for some constant c and likewise for any $k \geq 2$

$$\begin{aligned}&\mathsf{E}\left|b'\left\{m\left(X_{i\alpha}\right)\right\} - b'\left\{\bar{m}\left(X_{i\alpha}\right)\right\}\right|^k\left|B_{J,\alpha}\left(X_{i\alpha}\right)\right|^k\\ \leq\ &c^{k-2}\left\|m-\bar{m}\right\|_\infty^{k-2}\max_{\substack{1\leq J\leq N+1\\ 1\leq\alpha\leq d}}\left\|b_{J,\alpha}\right\|_2^{-(k-2)}\max_{\substack{1\leq J\leq N+1\\ 1\leq\alpha\leq d}}\mathsf{E}\left|b_{J,\alpha}^{k-2}\left(X_{i\alpha}\right)\right|\\ &\times\mathsf{E}\left|b'\left\{m\left(X_{i\alpha}\right)\right\} - b'\left\{\bar{m}\left(X_{i\alpha}\right)\right\}\right|^2\frac{b_{J,\alpha}^2\left(X_{i\alpha}\right)}{\left\|b_{J,\alpha}\right\|_2^2}\\ \leq\ &\left(cH^{5/2}\right)^{k-2}\mathsf{E}\left|b'\left\{m\left(X_{i\alpha}\right)\right\} - b'\left\{\bar{m}\left(X_{i\alpha}\right)\right\}\right|^2\frac{b_{J,\alpha}^2\left(X_{i\alpha}\right)}{\left\|b_{J,\alpha}\right\|_2^2}\end{aligned}$$

and

$$\begin{aligned} & \mathsf{E}\left[b'\left\{m\left(X_{i\alpha}\right)\right\}-b'\left\{\bar{m}\left(X_{i\alpha}\right)\right\}\right]^{2} B_{J,\alpha}^{2}\left(X_{i\alpha}\right) \\ \leq \quad & c\left\|m-\bar{m}\right\|_{\infty}^{2} \max_{\substack{1\leq J\leq N+1 \\ 1\leq\alpha\leq d}}\left\|b_{J,\alpha}\right\|_{2}^{-2} \max_{\substack{1\leq J\leq N+1 \\ 1\leq\alpha\leq d}} \mathsf{E}\left|b_{J,\alpha}^{2}\left(X_{i\alpha}\right)\right| = \mathcal{O}\left(H^{4}\right). \end{aligned}$$

Using these bounds and applying Lemma A.2, one has $\left|n^{-1}\sum_{i=1}^{n}\xi_{i,J,\alpha,n}\right| = \mathcal{O}_{a.s.}\left(H^{2}n^{-1/2}\log n\right)$ and

$$n^{-1}\left|\sum_{i=1}^{n}\sigma\left(\mathbf{X}_{i}\right)\varepsilon_{i}B_{J,\alpha}\left(X_{i\alpha}\right)\right| = \mathcal{O}_{a.s.}\left(n^{-1/2}\log n\right).$$

The lemma is then proved. □

Define the following matrices:

$$\begin{aligned} \mathbf{V} &= \mathsf{E}\,\mathbf{B}\left(\mathbf{X}\right)\mathbf{B}\left(\mathbf{X}\right)^{\mathsf{T}}, \mathbf{S}=\mathbf{V}^{-1}, \\ \mathbf{V}_{n} &= n^{-1}\sum_{i=1}^{n}\mathbf{B}\left(\mathbf{X}_{i}\right)\mathbf{B}\left(\mathbf{X}_{i}\right)^{\mathsf{T}}, \mathbf{S}_{n}=\mathbf{V}_{n}^{-1} \end{aligned}$$

and similar matrices

$$\mathbf{V}_{b} = \mathsf{E}\,b''\left\{m\left(\mathbf{X}\right)\right\}\mathbf{B}\left(\mathbf{X}\right)\mathbf{B}\left(\mathbf{X}\right)^{\mathsf{T}} = \left[\begin{array}{cc} v_{b,00} & v_{b,0,J,\alpha} \\ v_{b,0,J',\alpha'} & v_{b,J,\alpha,J',\alpha'} \end{array}\right]_{N_d\times N_d}$$

$$\mathbf{S}_{b} = \mathbf{V}_{b}^{-1} = \left[\begin{array}{cc} s_{b,00} & s_{b,0,J,\alpha} \\ s_{b,0,J',\alpha'} & s_{b,J,\alpha,J',\alpha'} \end{array}\right]_{N_d\times N_d}, \tag{A.8}$$

For any vector $\boldsymbol{\lambda}\in\mathbb{R}^{N_d}$, denote

$$\begin{aligned} \mathbf{V}_{b}\left(\boldsymbol{\lambda}\right) &= \mathsf{E}\,b''\left\{\boldsymbol{\lambda}^{\mathsf{T}}\mathbf{B}\left(\mathbf{X}\right)\right\}\mathbf{B}\left(\mathbf{X}\right)\mathbf{B}\left(\mathbf{X}\right)^{\mathsf{T}} \\ &= \left[\begin{array}{cc} v_{b,00}\left(\boldsymbol{\lambda}\right) & v_{b,0,J,\alpha}\left(\boldsymbol{\lambda}\right) \\ v_{b,0,J',\alpha'}\left(\boldsymbol{\lambda}\right) & v_{b,J,\alpha,J',\alpha'}\left(\boldsymbol{\lambda}\right) \end{array}\right]_{N_d\times N_d}, \\ \mathbf{S}_{b}\left(\boldsymbol{\lambda}\right) &= \mathbf{V}_{b}^{-1}\left(\boldsymbol{\lambda}\right) = \left[\begin{array}{cc} s_{b,00}\left(\boldsymbol{\lambda}\right) & s_{b,0,J,\alpha}\left(\boldsymbol{\lambda}\right) \\ s_{b,0,J',\alpha'}\left(\boldsymbol{\lambda}\right) & s_{b,J,\alpha,J',\alpha'}\left(\boldsymbol{\lambda}\right) \end{array}\right]_{N_d\times N_d} \end{aligned}$$

$$\mathbf{V}_{n,b}\left(\boldsymbol{\lambda}\right) = -\nabla^{2}\hat{L}\left(\boldsymbol{\lambda}\right), \mathbf{S}_{n,b}\left(\boldsymbol{\lambda}\right) = \mathbf{V}_{n,b}^{-1}\left(\boldsymbol{\lambda}\right). \tag{A.9}$$

LEMMA A.11. *Under Assumptions (A2) and (A4)*

$$c_{\mathbf{V}}\mathbf{I}_{N_d} \leq \mathbf{V} \leq C_{\mathbf{V}}\mathbf{I}_{N_d}, c_{\mathbf{S}}\mathbf{I}_{N_d} \leq \mathbf{S} \leq C_{\mathbf{S}}\mathbf{I}_{N_d}, \tag{A.10}$$

$$c_{\mathbf{V},b}\mathbf{I}_{N_d} \leq \mathbf{V}_{b} \leq C_{\mathbf{V},b}\mathbf{I}_{N_d}, c_{\mathbf{S},b}\mathbf{I}_{N_d} \leq \mathbf{S}_{b} \leq C_{\mathbf{S},b}\mathbf{I}_{N_d}. \tag{A.11}$$

Under Assumption (A2), (A4), (A5) and (A7), as $n \to \infty$ *with probability increasing to* 1

$$c_{\mathbf{V}} \mathbf{I}_{N_d} \leq \mathbf{V}_n(\boldsymbol{\lambda}) \leq C_{\mathbf{V}} \mathbf{I}_{N_d}, c_{\mathbf{S}} \mathbf{I}_{N_d} \leq \mathbf{S}_n(\boldsymbol{\lambda}) \leq C_{\mathbf{S}} \mathbf{I}_{N_d} \tag{A.12}$$

$$c_{\mathbf{V},b} \mathbf{I}_{N_d} \leq \mathbf{V}_{n,b}(\boldsymbol{\lambda}) \leq C_{\mathbf{V},b} \mathbf{I}_{N_d}, c_{\mathbf{S},b} \mathbf{I}_{N_d} \leq \mathbf{S}_{n,b}(\boldsymbol{\lambda}) \leq C_{\mathbf{S},b} \mathbf{I}_{N_d}. \tag{A.13}$$

PROOF. For (A.10), see Lemma A.9 in Wang and Yang (2007), while (A.12) follows from Lemma A.6. The statements (A.12) and (A.13) follow from (A.10) and (A.12), together with the boundedness of b'' in (A2). □

Define three vectors $\boldsymbol{\Phi}_b, \boldsymbol{\Phi}_v, \boldsymbol{\Phi}_r$ as

$$\begin{aligned}
\boldsymbol{\Phi}_b &= \left(\Phi_{b,0}, \Phi_{b,J,\alpha}\right)^{\mathsf{T}}_{\substack{1 \leq J \leq N+1 \\ 1 \leq \alpha \leq d}} \qquad \text{(A.14)} \\
&= -\mathbf{S}_b n^{-1} \textstyle\sum_{i=1}^{n} \left[b'\left\{m\left(\mathbf{X}_i\right)\right\} - b'\left\{\bar{m}\left(\mathbf{X}_i\right)\right\}\right] \mathbf{B}\left(\mathbf{X}_i\right), \\
\boldsymbol{\Phi}_v &= \left(\Phi_{v,0}, \Phi_{v,J,\alpha}\right)^{\mathsf{T}}_{\substack{1 \leq J \leq N+1 \\ 1 \leq \alpha \leq d}} \qquad \text{(A.15)} \\
&= -\mathbf{S}_b n^{-1} \textstyle\sum_{i=1}^{n} \left[\sigma\left(\mathbf{X}_i\right) \varepsilon_i\right] \mathbf{B}\left(\mathbf{X}_i\right), \\
\boldsymbol{\Phi}_r &= \left(\Phi_{r,0}, \Phi_{r,J,\alpha}\right)^{\mathsf{T}}_{\substack{1 \leq J \leq N+1 \\ 1 \leq \alpha \leq d}} \qquad \text{(A.16)} \\
&= \hat{\boldsymbol{\lambda}} - \bar{\boldsymbol{\lambda}} - \boldsymbol{\Phi}_b - \boldsymbol{\Phi}_v.
\end{aligned}$$

LEMMA A.12. *Under Assumptions (A1)-(A5) and (A7), as* $n \to \infty$

$$\begin{aligned}
\left\|\hat{\boldsymbol{\lambda}} - \bar{\boldsymbol{\lambda}}\right\| &= \mathcal{O}_{a.s.}\left(H^2 + H^{-1/2} n^{-1/2} \log n\right), \qquad \text{(A.17)} \\
\left\|\boldsymbol{\Phi}_r\right\| &= \mathcal{O}_{a.s.}\left(H^{-3/2} n^{-1} \log n\right), \qquad \text{(A.18)} \\
\left\|\boldsymbol{\Phi}_b\right\| &= \mathcal{O}_{a.s.}\left(H^2\right), \left\|\boldsymbol{\Phi}_v\right\| = \mathcal{O}_{a.s.}\left(H^{-1/2} n^{-1/2} \log n\right). \qquad \text{(A.19)}
\end{aligned}$$

PROOF. The Mean Value Theorem implies that an $N_d \times N_d$ diagonal matrix $\mathbf{t}$ exists whose diagonal elements are in $[0,1]$, such that for $\hat{\boldsymbol{\lambda}}^* = \mathbf{t}\hat{\boldsymbol{\lambda}} + \left(\mathbf{I}_{N_d} - \mathbf{t}\right) \bar{\boldsymbol{\lambda}}$

$$\nabla \hat{L}\left(\hat{\boldsymbol{\lambda}}\right) - \nabla \hat{L}\left(\bar{\boldsymbol{\lambda}}\right) = \nabla^2 \hat{L}\left(\hat{\boldsymbol{\lambda}}^*\right)\left(\hat{\boldsymbol{\lambda}} - \bar{\boldsymbol{\lambda}}\right).$$

Since, as noted before, that $\nabla \hat{L}\left(\hat{\boldsymbol{\lambda}}\right) = \mathbf{0}$, the above equation becomes

$$\hat{\boldsymbol{\lambda}} - \bar{\boldsymbol{\lambda}} = -\left(\nabla^2 \hat{L}\left(\hat{\boldsymbol{\lambda}}^*\right)\right)^{-1} \nabla \hat{L}\left(\bar{\boldsymbol{\lambda}}\right).$$

According to (A.9),

$$-\nabla^2 \hat{L}\left(\boldsymbol{\lambda}\right) = n^{-1} \textstyle\sum_{i=1}^{n} b'' \left\{\boldsymbol{\lambda}^{\mathsf{T}} \mathbf{B}\left(\mathbf{X}_i\right)\right\} \mathbf{B}\left(\mathbf{X}_i\right) \mathbf{B}\left(\mathbf{X}_i\right)^{\mathsf{T}} = \mathbf{V}_{n,b}\left(\boldsymbol{\lambda}\right),$$

Lemma A.11 implies that with probability approaching 1

$$c_{\mathbf{V},b} \mathbf{I}_{N_d} \le -\nabla^2 \hat{L}\left(\hat{\boldsymbol{\lambda}}^*\right) \le C_{\mathbf{V},b} \mathbf{I}_{N_d}.$$

Then (A.17) follows Lemma A.10. Furthermore, $\left\|\hat{\boldsymbol{\lambda}}^* - \bar{\boldsymbol{\lambda}}\right\| = \mathcal{O}_{a.s}\left(H^{3/2} + H^{-1/2} n^{-1/2} \log n\right)$ as well according to $\hat{\boldsymbol{\lambda}}^*$'s definition. Note that Taylor expansion ensures that for any vector $\mathbf{a} \in \mathbb{R}^{N_d}$

$$\begin{aligned} & \mathbf{a}^{\mathsf{T}} \left\{\nabla^2 \hat{L}\left(\hat{\boldsymbol{\lambda}}^*\right) - \nabla^2 \hat{L}\left(\bar{\boldsymbol{\lambda}}\right)\right\} \mathbf{a} \\ \le \; & \left\|b'''\right\|_\infty \max_{1\le i\le n} \left|\hat{\boldsymbol{\lambda}}^{*T} \mathbf{B}\left(\mathbf{X}_i\right) - \bar{\boldsymbol{\lambda}}^{\mathsf{T}} \mathbf{B}\left(\mathbf{X}_i\right)\right| \mathbf{a}^{\mathsf{T}} \mathbf{V}_n \mathbf{a} \end{aligned}$$

while by Cauchy Schwartz inequality

$$\begin{aligned} & \max_{1\le i\le n} \left|\hat{\boldsymbol{\lambda}}^{*T} \mathbf{B}\left(\mathbf{X}_i\right) - \bar{\boldsymbol{\lambda}}^{\mathsf{T}} \mathbf{B}\left(\mathbf{X}_i\right)\right| \\ \le \; & \left\|\hat{\boldsymbol{\lambda}}^* - \bar{\boldsymbol{\lambda}}\right\| \sup_{\mathbf{x}\in[0,1]^d} \left\|\mathbf{B}\left(\mathbf{x}\right)\right\| \\ = \; & \mathcal{O}_{a.s}\left(H^{3/2} + H^{-1/2} n^{-1/2} \log n\right) \times \mathcal{O}\left(H^{-1/2}\right) \\ = \; & \mathcal{O}_{a.s}\left(H + H^{-1} n^{-1/2} \log n\right). \end{aligned}$$

Consequently, one has the following bound on the difference of two Hessian matrices

$$\sup_{\mathbf{a}\in R^{N_d}} \left\|\left(\nabla^2 \hat{L}\left(\hat{\boldsymbol{\lambda}}^*\right) - \nabla^2 \hat{L}\left(\bar{\boldsymbol{\lambda}}\right)\right) \mathbf{a}\right\| \left\|\mathbf{a}\right\|^{-1} = \mathcal{O}_{a.s}\left(H + H^{-1} n^{-1/2} \log n\right).$$

Denote next

$$\begin{aligned} \hat{\mathbf{a}} &= -\left\{\nabla^2 \hat{L}\left(\hat{\boldsymbol{\lambda}}^*\right)\right\}^{-1} \nabla \hat{L}\left(\bar{\boldsymbol{\lambda}}\right) = \hat{\boldsymbol{\lambda}} - \bar{\boldsymbol{\lambda}} \\ \bar{\mathbf{a}} &= -\left\{\nabla^2 \hat{L}\left(\bar{\boldsymbol{\lambda}}\right)\right\}^{-1} \nabla \hat{L}\left(\bar{\boldsymbol{\lambda}}\right) \end{aligned}$$

then $\left\|\hat{\mathbf{a}}\right\| = \mathcal{O}_{a.s.}\left(H^{3/2} + H^{-1/2} n^{-1/2} \log n\right)$ and so is $\left\|\bar{\mathbf{a}}\right\|$ by similar arguments. Furthermore,

$$\nabla^2 \hat{L}\left(\hat{\boldsymbol{\lambda}}^*\right)\left(\hat{\mathbf{a}} - \bar{\mathbf{a}}\right) = \left\{\nabla^2 \hat{L}\left(\bar{\boldsymbol{\lambda}}\right) - \nabla^2 \hat{L}\left(\hat{\boldsymbol{\lambda}}^*\right)\right\} \bar{\mathbf{a}}$$

entails that

$$\begin{aligned}\|\hat{\mathbf{a}}-\bar{\mathbf{a}}\| &= \mathcal{O}_{a.s.}\left(H^{3/2}+H^{-1/2}n^{-1/2}\log n\right)\times \mathrm{O}_{a.s}\left(H+H^{-1}n^{-1/2}\log n\right)\\ &= \mathcal{O}_{a.s.}\left(H^{5/2}+H^{-3/2}n^{-1}\log^2 n\right).\end{aligned}$$

Using similar tricks, one can show that

$$\begin{aligned}\|\tilde{\mathbf{a}}-\bar{\mathbf{a}}\| &= \mathcal{O}_{a.s.}\left(H^{3/2}+H^{-1/2}n^{-1/2}\log n\right)\times \mathrm{O}_{a.s}\left(H^2\right)\\ &= \mathcal{O}_{a.s.}\left(H^{7/2}+H^{3/2}n^{-1/2}\log n\right),\end{aligned}$$

$$\begin{aligned}\|\tilde{\mathbf{a}}-\mathbf{\Phi}_b-\mathbf{\Phi}_v\| &= \mathcal{O}_{a.s.}\left(H^{3/2}+H^{-1/2}n^{-1/2}\log n\right)\times\mathcal{O}_{a.s}\left(H^{-1/2}n^{-1/2}\log n\right)\\ &= \mathcal{O}_{a.s.}\left(Hn^{-1/2}\log n+H^{-1}n^{-1}\log^2 n\right),\end{aligned}$$

in which

$$\tilde{\mathbf{a}}=\left[n^{-1}\textstyle\sum_{i=1}^n b''\left\{m\left(\mathbf{X}_i\right)\right\}\mathbf{B}\left(\mathbf{X}_i\right)\mathbf{B}\left(\mathbf{X}_i\right)^{\mathsf{T}}\right]^{-1}\nabla\hat{L}\left(\bar{\boldsymbol{\lambda}}\right).$$

Putting together the above proves (A.18). Lastly, almost surely

$$\begin{aligned}\|\mathbf{\Phi}_b\| &= \left\|\mathbf{S}_b n^{-1}\textstyle\sum_{i=1}^n\left[b'\left\{m\left(\mathbf{X}_i\right)\right\}-b'\left\{\bar{m}\left(\mathbf{X}_i\right)\right\}\right]\mathbf{B}\left(\mathbf{X}_i\right)\right\|\\ &\leq C_{\mathbf{S},b}\left\|n^{-1}\textstyle\sum_{i=1}^n\left[b'\left\{m\left(\mathbf{X}_i\right)\right\}-b'\left\{\bar{m}\left(\mathbf{X}_i\right)\right\}\right]\mathbf{B}\left(\mathbf{X}_i\right)\right\|=\mathcal{O}_{a.s.}\left(H^2\right)\end{aligned}$$

and

$$\begin{aligned}\|\mathbf{\Phi}_v\| &= \left\|\mathbf{S}_b n^{-1}\textstyle\sum_{i=1}^n\left[\sigma\left(\mathbf{X}_i\right)\varepsilon_i\right]\mathbf{B}\left(\mathbf{X}_i\right)\right\|\\ &\leq C_{\mathbf{S},b}\left\|n^{-1}\textstyle\sum_{i=1}^n\left[\sigma\left(\mathbf{X}_i\right)\varepsilon_i\right]\mathbf{B}\left(\mathbf{X}_i\right)\right\|=\mathcal{O}_{a.s.}\left(H^{-1/2}n^{-1/2}\log^2 n\right),\end{aligned}$$

which completes the proof of the lemma. □

LEMMA A.13. *Under Assumptions (A1)-(A5) and (A7), as $n\to\infty$*

$$\begin{aligned}\|\hat{m}-\bar{m}\|_\infty+\sum\nolimits_{\alpha=1}^d\|\hat{m}_\alpha-\bar{m}_\alpha\|_\infty &= \mathcal{O}_{a.s.}\left(H^{3/2}+H^{-1}n^{-1/2}\log n\right),\\ \|\hat{m}-\bar{m}\|_{2,n}+\|\hat{m}-\bar{m}\|_2 &= \mathcal{O}_{a.s.}\left(H^2+H^{-1/2}n^{-1/2}\log n\right),\end{aligned}$$

$$\begin{aligned}\|\hat{m}-m\|_\infty+\sum\nolimits_{\alpha=1}^d\|\hat{m}_\alpha-m_\alpha\|_\infty &= \mathcal{O}_{a.s.}\left(H^{3/2}+H^{-1}n^{-1/2}\log n\right),\\ \|\hat{m}-m\|_{2,n}+\|\hat{m}-m\|_2 &= \mathcal{O}_{a.s.}\left(H^2+H^{-1/2}n^{-1/2}\log n\right).\end{aligned}$$

PROOF. According to (A.17) and the definition of $\hat{m}$ and $\bar{m}$

$$\begin{aligned}\|\hat{m}-\bar{m}\|_\infty &= \sup_{\mathbf{x}\in[0,1]^d}\left|\hat{\boldsymbol{\lambda}}^{\mathsf{T}}\mathbf{B}(\mathbf{x})-\bar{\boldsymbol{\lambda}}^{\mathsf{T}}\mathbf{B}(\mathbf{x})\right| \le \left\|\hat{\boldsymbol{\lambda}}-\bar{\boldsymbol{\lambda}}\right\| \sup_{\mathbf{x}\in[0,1]^d}\|\mathbf{B}(\mathbf{x})\| \\ &\le \mathcal{O}_{a.s.}\left(H^2+H^{-1/2}n^{-1/2}\log n\right)\times\mathcal{O}\left(H^{-1/2}\right) \\ &= \mathcal{O}_{a.s.}\left(H^{3/2}+H^{-1}n^{-1/2}\log n\right).\end{aligned}$$

The bound on $\|\hat{m}_\alpha-\bar{m}_\alpha\|_\infty$ is similarly obtained. Next, Lemma A.11 implies

$$\begin{aligned}\|\hat{m}-\bar{m}\|_{2,n}+\|\hat{m}-\bar{m}\|_2 &\le 2C_{\mathbf{V}}\left\|\hat{\boldsymbol{\lambda}}-\bar{\boldsymbol{\lambda}}\right\| \\ &= \mathcal{O}_{a.s.}\left(H^2+H^{-1/2}n^{-1/2}\log n\right).\end{aligned}$$

Since $\|\bar{m}-m\|_\infty+\|\bar{m}-m\|_2+\|\bar{m}-m\|_{2,n}=\mathcal{O}\left(H^2\right)$ by the definition in (A.7), the Lemma follows. □

In the following denote

$$\boldsymbol{\omega}(x_1)=\{\omega_{J,\alpha}(x_1)\}_{J=1,\alpha=2}^{N+1,d}, \omega_{J,\alpha}(x_1)=n^{-1}\sum\nolimits_{i=1}^n|B_{J,\alpha}(X_{i\alpha})|K_h(X_{i1}-x_1).$$

LEMMA A.14. *Under Assumptions (A1)-(A7), as $n\to\infty$,*

$$\sup_{\substack{x_1\in[0,1],2\le\alpha\le d\\1\le J\le N+1}}\left|\omega_{J,\alpha}(x_1)-\mathsf{E}\,\omega_{J,\alpha}(x_1)\right|=\mathcal{O}_{a.s.}\left(\log n/\sqrt{nh}\right) \tag{A.20}$$

$$\sup_{x_1\in[0,1]}|\boldsymbol{\omega}(x_1)|=\sup_{\substack{x_1\in[0,1],2\le\alpha\le d\\1\le J\le N+1,}}|\omega_{J,\alpha}(x_1)|=\mathcal{O}_{a.s.}\left(H^{1/2}\right). \tag{A.21}$$

PROOF. First, one computes

$$\begin{aligned}\mathsf{E}\,\omega_{J,\alpha}(x_1) &= \int\int K_h(u_1-x_1)|B_{J,\alpha}(u_\alpha)|f(u_1,u_\alpha)du_1du_\alpha \\ &= \int\int K(v_1)\frac{|b_{J,\alpha}(u_2)|}{\|b_{J,\alpha}\|_2}f(hv_1+x_1,u_\alpha)dv_1du_\alpha \\ &= \left(\|b_{J,\alpha}\|_2\right)^{-1}\left\{\int\int K(v_1)I_{J+1,2}(u_2)f(hv_1+x_1,u_2)dv_1du_2\right. \\ &\quad\left.+\left(\frac{c_{J+1,2}}{c_{J,2}}\right)^{1/2}\int\int K(v_1)I_{J,2}(u_2)f(hv_1+x_1,u_2)dv_1du_2\right\}.\end{aligned}$$

$$\begin{aligned} \leq\ & \|b_{J,\alpha}\|_2^{-1}\left\{\int\int\left|K(v_1)\,b_J(u_\alpha)\right|f(x_1+hv_1,u_\alpha)\,dv_1du_\alpha\right. \\ & \left.+\frac{c_{J,\alpha}}{c_{J-1,\alpha}}\int\int\left|K(v_1)\,b_{J-1}(u_\alpha)\right|f(x_1+hv_1,u_\alpha)\,dv_1du_\alpha\right\}. \end{aligned}$$

The boundedness of the joint density f and the Lipschitz continuity of the kernel K imply that a constant $c_2>0$ exists such that

$$\int\int\left|K(v_1)\,b_{J-1}(u_\alpha)\right|f(x_1+hv_1,u_\alpha)\,dv_1du_\alpha\leq C_Kc_2H.$$

Therefore

$$\sup_{x_1\in[0,1]}\left|\mathsf{E}\,\boldsymbol{\omega}(x_1)\right|=\mathcal{O}\left(H^{1/2}\right) \tag{A.22}$$

by Lemma A.3. Similarly, $\mathsf{E}\,\omega_{J,\alpha}(x_1)^r\sim h^{1-r}H^{1-r/2}$, hence $\mathsf{E}\,\omega_{J,\alpha}(x_1)^2\sim h^{-1}$. According to Lemma A.2 and similar proof of Lemma A.5 in Wang and Yang (2007), one proves (A.20). Combining with (A.22), the lemma is proved. □

LEMMA A.15. *Under Assumptions (A1)-(A7), as $n\to\infty$*

$$\sup_{x_1\in[0,1]}\left|\hat{l}'\left\{\tilde{m}_{\mathrm{K},1}(x_1)\right\}\right|=\mathcal{O}_{a.s.}\left(n^{-1/2}\log n\right).$$

PROOF. Note that $\tilde{l}'\left\{\tilde{m}_{\mathrm{K},1}(x_1)\right\}=0$, thus $\hat{l}'\left\{\tilde{m}_{\mathrm{K},1}(x_1)\right\}=\hat{l}'\left\{\tilde{m}_{\mathrm{K},1}(x_1)\right\}-\tilde{l}'\left\{\tilde{m}_{\mathrm{K},1}(x_1)\right\}$ equals

$$\begin{aligned} & n^{-1}\textstyle\sum_{i=1}^n\left[b'\left\{\tilde{m}_{\mathrm{K},1}(x_1)+m_{-1}(\mathbf{X}_{i-1})\right\}-b'\left\{\tilde{m}_{\mathrm{K},1}(x_1)+\hat{m}_{-1}(\mathbf{X}_{i-1})\right\}\right] \\ & K_h(X_{i1}-x_1) \\ = & \ n^{-1}\textstyle\sum_{i=1}^n b''\left\{\tilde{m}_{\mathrm{K},1}(x_1)+m_{-1}(\mathbf{X}_{i-1})\right\}\left\{m_{-1}(\mathbf{X}_{i-1})-\hat{m}_{-1}(\mathbf{X}_{i-1})\right\} \\ & K_h(X_{i1}-x_1)+\mathcal{O}\left[1/n\textstyle\sum_{i=1}^n\left\{m_{-1}(\mathbf{X}_{i-1})-\hat{m}_{-1}(\mathbf{X}_{i-1})\right\}^2\right]. \end{aligned}$$

Now Lemma A.13 together with (A7) imply:

$$\begin{aligned} 1/n\textstyle\sum_{i=1}^n\left\{m_{-1}(\mathbf{X}_{i-1})-\hat{m}_{-1}(\mathbf{X}_{i-1})\right\}^2 &= \|m_{-1}-\hat{m}_{-1}\|_{2,n}^2\leq\|m-\hat{m}\|_{2,n}^2 \\ =\mathcal{O}_{a.s.}\left(H^4+H^{-1}n^{-1}\log^2 n\right) &= \mathcal{O}_{a.s.}\left(n^{-1/2}\log n\right). \end{aligned}$$

This yields

$$\hat{l}'\left\{\tilde{m}_{\mathrm{K},1}(x_1)\right\}=I_1+I_1+\mathcal{O}_{a.s.}\left(n^{-1/2}\log^2 n\right), \tag{A.23}$$

$$
\begin{aligned}
I_1 &= n^{-1}\textstyle\sum_{i=1}^{n} b''\left\{\tilde{m}_{\mathrm{K},1}\left(x_1\right)+m_{\_1}\left(\mathbf{X}_{i\_1}\right)\right\} \\
&\quad \left\{m_{\_1}\left(\mathbf{X}_{i\_1}\right)-\bar{m}_{\_1}\left(\mathbf{X}_{i\_1}\right)\right\} K_h\left(X_{i1}-x_1\right), \\
I_2 &= n^{-1}\textstyle\sum_{i=1}^{n} b''\left\{\tilde{m}_{\mathrm{K},1}\left(x_1\right)+m_{\_1}\left(\mathbf{X}_{i\_1}\right)\right\} \\
&\quad \left\{\bar{m}_{\_1}\left(\mathbf{X}_{i\_1}\right)-\hat{m}_{\_1}\left(\mathbf{X}_{i\_1}\right)\right\} K_h\left(X_{i1}-x_1\right).
\end{aligned}
$$

Applying standard kernel theory, the boundedness of b'' and (A.4), one obtains:

$$
\begin{aligned}
|I_1| &\le C_b \textstyle\sum_{\alpha=2}^{d}\left\|m_\alpha-\bar{m}_\alpha\right\|_\infty n^{-1}\textstyle\sum_{i=1}^{n} K_h\left(X_{i1}-x_1\right) \\
&= \mathcal{O}_{a.s.}\left(H^2\right) \ a.s.
\end{aligned} \tag{A.24}
$$

again by (A7) on H, while $I_2 = I_{2,b}+I_{2,v}+I_{2,r}$ with

$$
\begin{aligned}
I_{2,b} &= n^{-1}\textstyle\sum_{i=1}^{n} b''\left\{\tilde{m}_{\mathrm{K},1}\left(x_1\right)+m_{\_1}\left(\mathbf{X}_{i\_1}\right)\right\} \times \\
&\quad \left\{\Phi_{b,0}+\textstyle\sum_{1\le J\le N+1,2\le\alpha\le d}\Phi_{b,J,\alpha}B_{J,\alpha}\left(X_{i\alpha}\right)\right\} K_h\left(X_{i1}-x_1\right),
\end{aligned}
$$

$$
\begin{aligned}
I_{2,v} &= n^{-1}\textstyle\sum_{i=1}^{n} b''\left\{\tilde{m}_{\mathrm{K},1}\left(x_1\right)+m_{\_1}\left(\mathbf{X}_{i\_1}\right)\right\} \times \\
&\quad \left\{\Phi_{v,0}+\textstyle\sum_{1\le J\le N+1,2\le\alpha\le d}\Phi_{v,J,\alpha}B_{J,\alpha}\left(X_{i\alpha}\right)\right\} K_h\left(X_{i1}-x_1\right),
\end{aligned}
$$

$$
\begin{aligned}
I_{2,r} &= n^{-1}\textstyle\sum_{i=1}^{n} b''\left\{\tilde{m}_{\mathrm{K},1}\left(x_1\right)+m_{\_1}\left(\mathbf{X}_{i\_1}\right)\right\} \times \\
&\quad \left\{\Phi_{r,0}+\textstyle\sum_{1\le J\le N+1,2\le\alpha\le d}\Phi_{r,J,\alpha}B_{J,\alpha}\left(X_{i\alpha}\right)\right\} K_h\left(X_{i1}-x_1\right)
\end{aligned}
$$

where $\Phi_{b,0}, \Phi_{b,0}, \Phi_{b,0}, \Phi_{b,J,\alpha}, \Phi_{b,J,\alpha}, \Phi_{b,J,\alpha}$ are defined in (A.14), (A.15) and (A.16). $|I_{2,b}|$ is bounded by

$$
\begin{aligned}
& C_b n^{-1}\textstyle\sum_{i=1}^{n}\left\{\left|\Phi_{b,0}\right|+\textstyle\sum_{1\le J\le N+1,2\le\alpha\le d}\left|\Phi_{b,J,\alpha}\right|\left|B_{J,\alpha}\left(X_{i\alpha}\right)\right|\right\} K_h\left(X_{i1}-x_1\right) \\
\le\ & C_b\left[\left\{\Phi_{b,0}^2+\textstyle\sum_{1\le J\le N+1,2\le\alpha\le d}\Phi_{b,J,\alpha}^2\right\}\right]^{1/2} \times \left[\left\{n^{-1}\textstyle\sum_{i=1}^{n} K_h\left(X_{i1}-x_1\right)\right\}^2\right. \\
& \left.+\textstyle\sum_{1\le J\le N+1,2\le\alpha\le d}\left\{n^{-1}\textstyle\sum_{i=1}^{n}\left|B_{J,\alpha}\left(X_{i\alpha}\right)\right| K_h\left(X_{i1}-x_1\right)\right\}^2\right]^{1/2} \\
=\ & C_b \times \left\|\mathbf{\Phi}_b\right\| \times \left[\mathcal{O}_{a.s.}(1)+(N+1)\times(d-1)\times\mathcal{O}_{a.s.}(H)\right],
\end{aligned}
$$

so

$$
\left|I_{2,b}\right| = \mathcal{O}_{a.s.}\left(H^2\right) \tag{A.25}
$$

according to (A.17) and (A.21). Similarly

$$
\begin{aligned}
\left|I_{2,r}\right| &= \mathcal{O}_{a.s}\left(\left\|\mathbf{\Phi}_r\right\|\right) = \mathcal{O}_{a.s.}\left(H^{-3/2}n^{-1}\log n\right) \\
&= \mathcal{O}_{a.s.}\left(n^{-1/2}\log n\right)
\end{aligned} \tag{A.26}
$$

$$I_{2,v}-\widetilde{I}_{2,v} = \mathcal{O}_{a.s.}\left(\log n/\sqrt{nh}\right)\times\mathcal{O}_{a.s.}\left(H^{-1/2}n^{-1/2}\log n\right) = \mathcal{O}_{a.s.}\left(n^{-1/2}\log n\right) \tag{A.27}$$

by making use of (A7) on H and (A6) on h, where

$$\widetilde{I}_{2,v} = \Phi_{v,0}n^{-1}\textstyle\sum_{i=1}^{n} b''\left\{m\left(\mathbf{X}_i\right)\right\}K_h\left(X_{i1}-x_1\right)+\sum_{1\le J\le N+1,2\le\alpha\le d} \Phi_{v,J,\alpha}n^{-1}\sum_{i=1}^{n} b''\left\{m\left(\mathbf{X}_i\right)\right\}B_{J,\alpha}\left(X_{i\alpha}\right)K_h\left(X_{i1}-x_1\right).$$

Applying Lemma A.2, we have $\widetilde{I}_{2,v}$ equals

$$\begin{aligned} &\Phi_{v,0}\,\mathsf{E}\,b''\left\{m\left(\mathbf{X}\right)\right\}K_h\left(X_1-x_1\right)\\ &+\textstyle\sum_{1\le J\le N+1,2\le\alpha\le d}\Phi_{v,J,\alpha}\,\mathsf{E}\,b''\left\{m\left(\mathbf{X}\right)\right\}B_{J,\alpha}\left(X_\alpha\right)K_h\left(X_1-x_1\right)\\ &+\mathcal{O}_{a.s.}\left(H^{-1/2}n^{-1/2}\log n\right)\times N_d^{1/2}\times\mathcal{O}_{a.s.}\left(\log n/\sqrt{nh}\right)\\ =\;&\widetilde{I}_{2,v,1}+\widetilde{I}_{2,v,2}+\mathcal{O}_{a.s.}\left(n^{-1/2}\log^2 n\right),\end{aligned}\tag{A.28}$$

in which

$$\begin{aligned}\widetilde{I}_{2,v,1} &= \Phi_{v,0}\,\mathsf{E}\,b''\left\{m\left(\mathbf{X}\right)\right\}K_h\left(X_1-x_1\right)\\ &= \mathsf{E}\,b''\left\{m\left(\mathbf{X}\right)\right\}K_h\left(X_1-x_1\right)n^{-1}\textstyle\sum_{i=1}^{n}\sigma\left(\mathbf{X}_i\right)\varepsilon_i\\ &\quad\left\{S_{0,0}+\textstyle\sum_{1\le J\le N+1,1\le\alpha\le d}S_{J,\alpha}B_{J,\alpha}\left(X_{i\alpha}\right)\right\},\end{aligned}$$

$$\begin{aligned}\widetilde{I}_{2,v,2} &= \textstyle\sum_{1\le J\le N+1,2\le\alpha\le d}\Phi_{v,J,\alpha}\,\mathsf{E}\,b''\left\{m\left(\mathbf{X}\right)\right\}B_{J,\alpha}\left(X_\alpha\right)K_h\left(X_1-x_1\right)\\ &= \left\{\mathsf{E}\,b''\left\{m\left(\mathbf{X}\right)\right\}B_{J,\alpha}\left(X_\alpha\right)K_h\left(X_1-x_1\right)\right\}_{1\le J\le N+1,2\le\alpha\le d}\\ &\quad\left(\Phi_{v,J,\alpha}\right)^{\mathsf{T}}_{1\le J\le N+1,2\le\alpha\le d}\\ &= \textstyle\sum_{1\le J\le N+1,2\le\alpha\le d}\mathsf{E}\,b''\left\{m\left(\mathbf{X}\right)\right\}B_{J,\alpha}\left(X_\alpha\right)K_h\left(X_1-x_1\right)\\ &\quad n^{-1}\textstyle\sum_{i=1}^{n}\sigma\left(\mathbf{X}_i\right)\varepsilon_i\left\{S_{J,\alpha}+\sum_{1\le J'\le N+1,1\le\alpha'\le d}S_{J,\alpha,J',\alpha'}B_{J',\alpha'}\left(X_{i\alpha'}\right)\right\}\\ &= n^{-1}\textstyle\sum_{i=1}^{n}\sigma\left(\mathbf{X}_i\right)\varepsilon_i\sum_{1\le J\le N+1,2\le\alpha\le d}\mu_{b,k,J,\alpha}\left(x_1\right)\\ &\quad\left\{S_{J,\alpha}+\textstyle\sum_{1\le J'\le N+1,1\le\alpha'\le d}S_{J,\alpha,J',\alpha'}B_{J',\alpha'}\left(X_{i\alpha'}\right)\right\}\end{aligned}$$

where $S_{0,0}, S_{J,\alpha}$, $S_{J,\alpha,J',\alpha'}$ are the corresponding elements in the matrix $\mathbf{S}_b$ defined in (A.8), and $\mu_{b,k,J,\alpha}\left(x_1\right) = \mathsf{E}\,b''\left\{m\left(\mathbf{X}\right)\right\}B_{J,\alpha}\left(X_\alpha\right)K_h\left(X_1-x_1\right)$, the supremum of which has the order $\mathcal{O}\left(H^{1/2}\right)$. Denote

$D_n = n^{\theta_0}\left(\frac{1}{2+\eta} < \theta_0 < \frac{2}{5}\right)$, $\varepsilon_{i,1}^{D_n} = \varepsilon_i I\left\{|\varepsilon_i| > D_n\right\}$, $\varepsilon_{i,3}^{D_n} = \mathsf{E}\,\varepsilon_i I\left\{|\varepsilon_i| \leq D_n\right\}$, $\varepsilon_{i,2}^{D_n} = \varepsilon_i I\left\{|\varepsilon_i| \leq D_n\right\} - \varepsilon_{i,3}^{D_n}$. Then $\widetilde{I}_{2,v,2} = \Lambda_1 + \Lambda_2 + \Lambda_3$ where

$$\Lambda_k = \textstyle\sum_{1\leq J\leq N+1, 2\leq\alpha\leq d} \mu_{b,k,J,\alpha}\left(x_1\right)\left(x_1\right) n^{-1} \sum_{i=1}^n \sigma\left(\mathbf{X}_i\right) \varepsilon_{i,k}^{D_n}$$
$$\left\{S_{J,\alpha} + \textstyle\sum_{1\leq J'\leq N+1, 1\leq\alpha'\leq d} S_{J,\alpha,J',\alpha'} B_{J',\alpha'}\left(X_{i\alpha'}\right)\right\}, k = 1, 2, 3.$$

With probability 1, $\Lambda_1 = 0$ for large n. Next

$$\left|\varepsilon_{i,3}^{D_n}\right| = \left|-\mathsf{E}\,\varepsilon_i I\left\{|\varepsilon_i| > D_n\right\}\right| \leq \frac{\mathsf{E}\left|\varepsilon_i\right|^{2+\eta}}{D_n^{1+\eta}} = \mathcal{O}\left(D_n^{-(1+\eta)}\right),$$

$$\begin{aligned}
\Lambda_3 \quad &\leq \quad C_{\mathbf{S}_b}\Big[\textstyle\sum_{1\leq J\leq N+1, 2\leq\alpha\leq d} \mu_{b,k,J,\alpha}^2\left(x_1\right) \\
&\qquad \textstyle\sum_{1\leq J'\leq N+1, 1\leq\alpha'\leq d}\left\{n^{-1}\sum_{i=1}^n B_{J',\alpha'}\left(X_{i\alpha'}\right)\sigma\left(\mathbf{X}_i\right)\varepsilon_{i,3}^{D_n}\right\}^2\Big]^{1/2} \\
&\leq \quad C D_n^{-(1+\eta)}\Big[\textstyle\sum_{1\leq J\leq N+1, 2\leq\alpha\leq d} \mu_{b,k,J,\alpha}^2\left(x_1\right) \\
&\qquad \textstyle\sum_{1\leq J'\leq N+1, 1\leq\alpha'\leq d}\left\{n^{-1}\sum_{i=1}^n B_{J',\alpha'}\left(X_{i\alpha'}\right)\sigma\left(\mathbf{X}_i\right)\right\}^2\Big]^{1/2} \\
= \quad & D_n^{-(1+\eta)}\mathcal{O}_{a.s.}\left\{\left(NHN\log^2 n/n\right)^{1/2}\right\} = D_n^{-(1+\eta)}\mathcal{O}_{a.s.}\left\{\left(N\log^2 n/n\right)^{1/2}\right\} \\
= \quad & \scriptstyle{\mathcal{O}}_{a.s.}\left(n^{-1/2}\right) = \mathcal{O}_{a.s.}\left(n^{-1/2}\log n\right).
\end{aligned}$$

Lastly, denote $\Lambda_2 = n^{-1}\sum_{i=1}^n \xi_i$, where

$$\begin{aligned}
\xi_i \quad = \quad & \textstyle\sum_{1\leq J\leq N+1, 2\leq\alpha\leq d} \mu_{b,k,J,\alpha}\left(x_1\right)\sigma\left(\mathbf{X}_i\right)\varepsilon_{i,2}^{D_n} \\
& \left\{S_{J,\alpha} + \textstyle\sum_{1\leq J'\leq N+1, 1\leq\alpha'\leq d} S_{J,\alpha,J',\alpha'} B_{J',\alpha'}\left(X_{i\alpha'}\right)\right\}.
\end{aligned}$$

Then $\mathsf{E}\,\xi_i = 0$, and

$$\begin{aligned}
\mathsf{Var}\left(\xi_i\right) \quad &= \quad \boldsymbol{\mu}_{b\_c}^{\mathsf{T}}\mathbf{S}_{b\_1}\,\mathsf{Var}\left(\left\{\begin{array}{c}\sigma\left(\mathbf{X}_i\right)\varepsilon_{i,k}^{D_n} \\ B_{J',\alpha'}\left(X_{i\alpha'}\right)\sigma\left(\mathbf{X}_i\right)\varepsilon_{i,k}^{D_n}\end{array}\right\}_{J'=1,\alpha'=1}^{N+1,d}\right)\mathbf{S}_{b\_1}^{\mathsf{T}}\boldsymbol{\mu}_{b\_c} \\
&\leq \quad C_\sigma C_{\mathbf{S}}^2 C_{\mathbf{V}}\boldsymbol{\mu}_{b,k\_1}^{\mathsf{T}}\boldsymbol{\mu}_{b,k\_c}^{\mathsf{T}} = \mathcal{O}\left(1\right).
\end{aligned}$$

Then $\Lambda_2 = \mathcal{O}_{a.s.}\left(n^{-1/2}\log n\right)$ according to Bernstein's Inequality. Then $\widetilde{I}_{2,v,2} = \mathcal{O}_{a.s.}\left(n^{-1/2}\log n\right)$ according to the orders of Λ_1, Λ_2 and Λ_3. With similar proof, we can show $\widetilde{I}_{2,v,1} = \mathcal{O}_{a.s.}\left(n^{-1/2}\log n\right)$. The lemma is proved by putting together (A.23), (A.24), (A.25), (A.26), (A.27), (A.28) and the above bound on $\widetilde{I}_{2,v,2}$ and $\widetilde{I}_{2,v,1}$. □

LEMMA A.16. *Under Assumptions (A1)-(A7), constants c, C exist such that $0 < c \leq \left|-\hat{l}''(a, x_1)\right| \leq C < \infty$ a.s. for $a \in A, x_1 \in [0,1]$.*

PROOF. According to (4.1), one has

$$\hat{l}''(a) = -1/n \sum_{i=1}^{n} \left[b''\left\{a + \hat{m}_{\_1}(\mathbf{X}_{i\_1})\right\}\right] K_h(X_{i1} - x_1).$$

$c_b \leq b''\left\{a + \hat{m}_{\_1}(\mathbf{X}_{i\_1})\right\} \leq C_b$ and $\sup_{x_1 \in [h, 1-h]} \left|1/n \sum_{i=1}^{n} K_h(X_{i1} - x_1) - f(x_1)\right| = \mathcal{O}_{a.s.}(1)$ imply the lemma. □

PROOF OF THEOREM 4. According to (4.1) and the Mean Value Theorem, a $\bar{m}_{\mathrm{K},1}(x_1)$ between $\hat{m}_{\mathrm{SBK},1}(x_1)$ and $\tilde{m}_{\mathrm{K},1}(x_1)$ exists such that

$$\hat{l}'\left\{\hat{m}_{\mathrm{SBK},1}(x_1)\right\} - \hat{l}'\left\{\tilde{m}_{\mathrm{K},1}(x_1)\right\} = \hat{l}''\left(\bar{m}_{\mathrm{K},1}(x_1)\right)\left\{\hat{m}_{\mathrm{SBK},1}(x_1) - \tilde{m}_{\mathrm{K},1}(x_1)\right\},$$

Then according to $\hat{l}'\left\{\hat{m}_{\mathrm{SBK},1}(x_1)\right\} = 0$, one has

$$\hat{m}_{\mathrm{SBK},1}(x_1) - \tilde{m}_{\mathrm{K},1}(x_1) = -\frac{\hat{l}'\left\{\tilde{m}_{\mathrm{K},1}(x_1)\right\}}{\hat{l}''\left\{\bar{m}_{\mathrm{K},1}(x_1)\right\}}. \tag{A.29}$$

The theorem then follows from Lemmas A.15 and A.16. □

PROOF OF THEOREM 5. The Mean Value Theorem implies the existence of $\bar{c}'$ between $\hat{c}$ and $\tilde{c}$ such that $\hat{c} - \tilde{c} = -\hat{l}'_c(\tilde{c})/l''_c(\bar{c}')$,where $-\hat{l}''_c(\bar{c}') = n^{-1} \sum_{i=1}^{n} b''\left\{\bar{c}' + \hat{m}_{\_c}(\mathbf{X}_i)\right\} > c_b > 0$ according to (A6), then

$$\hat{l}'_c(\tilde{c}) = \hat{l}'_c(\tilde{c}) - \tilde{l}'_c(\tilde{c}) = n^{-1} \sum_{i=1}^{n} \left[b'\left\{\tilde{c} + m_{\_c}(\mathbf{X}_i)\right\} - b'\left\{\tilde{c} + \hat{m}_{\_c}(\mathbf{X}_i)\right\}\right]$$

$$\begin{aligned} &= n^{-1} \sum_{i=1}^{n} b''\left\{\tilde{c} + m_{\_c}(\mathbf{X}_i)\right\}\left\{m_{\_c}(\mathbf{X}_i) - \hat{m}_{\_c}(\mathbf{X}_i)\right\} \\ &\quad + \mathcal{O}\left[1/n \sum_{i=1}^{n} \left\{m_{\_c}(\mathbf{X}_i) - \hat{m}_{\_c}(\mathbf{X}_i)\right\}^2\right] \\ &= I + \mathcal{O}_{a.s.}\left(N_d H^4 + N_d n^{-1} \log n\right), \end{aligned}$$

by Lemma A.13, where $I = I_1 + I_2$,

$$I_1 = n^{-1} \sum_{i=1}^{n} b''\left\{\tilde{c} + m_{\_c}(\mathbf{X}_i)\right\}\left\{m_{\_c}(\mathbf{X}_i) - \bar{m}_{\_c}(\mathbf{X}_i)\right\},$$

$$I_2 = n^{-1} \sum_{i=1}^{n} b''\left\{\tilde{c} + m_{\_c}(\mathbf{X}_i)\right\}\left\{\bar{m}_{\_c}(\mathbf{X}_i) - \hat{m}_{\_c}(\mathbf{X}_i)\right\}.$$

According to Lemma A.4, $I_1 = \mathcal{O}_{a.s.}\left(H^2\right)$, while

$$\begin{aligned} I_2 &= n^{-1} \sum_{i=1}^{n} b''\left\{\tilde{c} + m_{\_c}(\mathbf{X}_i)\right\} \times \\ &\quad \left\{\sum_{1 \leq J \leq N+1, 1 \leq \alpha \leq d} \left(\hat{\lambda}_{J,\alpha} - \bar{\lambda}_{J,\alpha}\right) B_{J,\alpha}(X_{i\alpha})\right\} \end{aligned}$$

$$= I_{2,b} + I_{2,v} + I_{2,r}$$

where

$$\begin{aligned} I_{2,b} &= n^{-1}\textstyle\sum_{i=1}^{n} b''\left\{\tilde{c} + m_{\_c}\left(\mathbf{X}_i\right)\right\}\left\{\sum_{1\le J\le N+1, 1\le\alpha\le d}\Phi_{b,J,\alpha}B_{J,\alpha}\left(X_{i\alpha}\right)\right\}, \\ I_{2,v} &= n^{-1}\textstyle\sum_{i=1}^{n} b''\left\{\tilde{c} + m_{\_c}\left(\mathbf{X}_i\right)\right\}\left\{\sum_{1\le J\le N+1, 1\le\alpha\le d}\Phi_{v,J,\alpha}B_{J,\alpha}\left(X_{i\alpha}\right)\right\}, \\ I_{2,r} &= n^{-1}\textstyle\sum_{i=1}^{n} b''\left\{\tilde{c} + m_{\_c}\left(\mathbf{X}_i\right)\right\}\left\{\sum_{1\le J\le N+1, 1\le\alpha\le d}\Phi_{r,J,\alpha}B_{J,\alpha}\left(X_{i\alpha}\right)\right\}. \end{aligned}$$

We have $\left|I_{2,b}\right| = \mathcal{o}_{a.s.}\left(n^{-1/2}\right)$ according to (A.17) and (A.21), see Liu et al. (2011). Similarly

$$\left|I_{2,r}\right| = \mathcal{O}_{a.s.}\left(N_d H^{7/2} + N_d H^{-1/2} n^{-1}\log n\right) = \mathcal{o}_{a.s.}\left(n^{-1/2}\right).$$

We have $I_{2,v} = \widetilde{I}_{2,v} + \mathcal{O}_{a.s.}\left(n^{-1/2}\right) \times \mathcal{O}_{a.s.}\left(N_d^{1/2} n^{-1/2}\log n\right) \times \mathcal{O}\left(N\right)$, where

$$\begin{aligned} \widetilde{I}_{2,v} &= n^{-1}\textstyle\sum_{i=1}^{n} b''\left\{m\left(\mathbf{X}_i\right)\right\}\left\{\sum_{1\le J\le N+1, 1\le\alpha\le d}\Phi_{v,J,\alpha}B_{J,\alpha}\left(X_{i\alpha}\right)\right\} \\ &= -n^{-1}\textstyle\sum_{i=1}^{n} b''\left\{m\left(\mathbf{X}_i\right)\right\} n^{-1}\sum_{i'=1}^{n}\sigma\left(\mathbf{X}_{i'}\right)\varepsilon_{i'}\mathbf{B}^{\mathsf{T}}\left(\mathbf{X}_{i'}\right)\mathbf{S}_{b\_c}\mathbf{B}_{\_c}\left(\mathbf{X}_i\right) \end{aligned}$$

where $\mathbf{S}_{b\_c} = \mathbf{S}_b\left(\mathbf{0}_{N_d-1}, \mathbf{I}_{N_d-1}\right)^{\mathsf{T}}$ consists of columns 2 to N_d of $\mathbf{S}_b$ defined in (A.8) and $\mathbf{B}_{\_c}\left(\mathbf{x}\right) = \left\{B_{1,1}\left(x_1\right), ..., B_{N+1,d}\left(x_d\right)\right\}^{\mathsf{T}}$. So

$$\begin{aligned} \widetilde{I}_{2,v} &= -n^{-1}\textstyle\sum_{i'=1}^{n}\sigma\left(\mathbf{X}_{i'}\right)\varepsilon_{i'} + n^{-1}\sum_{i'=1}^{n}\sigma\left(\mathbf{X}_{i'}\right)\varepsilon_{i'}v_{b,00}\left(s_{b,00}, s_{b,0,J,\alpha}\right)\mathbf{B}\left(\mathbf{X}_{i'}\right) \\ &\quad + \mathcal{o}_{a.s.}\left(n^{-1/2}\right) \end{aligned}$$

by Liu et al. (2011). Putting the above together, and noticing that $v_{b,00} = \mathsf{E}\, b''\left\{m\left(\mathbf{X}\right)\right\}$, one has

$$\hat{c} - \tilde{c} = T_n + \mathcal{o}_{a.s.}\left(n^{-1/2}\right), \tag{A.30}$$

$$T_n = n^{-1}\textstyle\sum_{i'=1}^{n}\sigma\left(\mathbf{X}_{i'}\right)\varepsilon_{i'}\left(s_{b,00} - v_{b,00}^{-1}, s_{b,0,J,\alpha}\right)\mathbf{B}\left(\mathbf{X}_{i'}\right). \tag{A.31}$$

According to (A.8) and matrix theory

$$\begin{aligned} \mathbf{S}_b &= \begin{pmatrix} v_{b,00}^{-1} + v_{b,00}^{-2} B F^{-1} B^{\mathsf{T}} & -v_{b,00}^{-1} B F^{-1} \\ -v_{b,00}^{-1} F^{-1} B^{\mathsf{T}} & F^{-1} \end{pmatrix} \\ B &= \left(v_{b,0,J,\alpha}\right), F = \left(v_{b,J,\alpha,J',\alpha'}\right) - B^{\mathsf{T}} B v_{b,00}^{-1}. \end{aligned}$$

According to (A.11)

$$0 < c_{V,b} \le v_{b,00} \le C_{V,b} < +\infty$$

$$c_{V,b}\mathbf{I}_{N_d-1} \le \left(v_{b,J,\alpha,J',\alpha'}\right) \le C_{V,b}\mathbf{I}_{N_d-1}$$

while the definition of $\mathbf{B}\left(\mathbf{X}_i\right)$ implies that $\left\|B\right\|_\infty = \mathcal{O}\left(H^{3/2}\right)$ and hence $\left\|B\right\| = \mathcal{O}\left(H\right)$. Thus a constant $c_F > 0$ exists such that for sufficiently large n, $F \ge c_F\mathbf{I}_{N_d-1}$ and hence $F^{-1} \le c_F^{-1}\mathbf{I}_{N_d-1}$. Putting the above together leads to

$$s_{b,00} = v_{b,00}^{-1} + v_{b,00}^{-2}BF^{-1}B^{\top} = v_{b,00}^{-1} + \mathcal{O}\left(H^2\right)$$

$$s_{b,0,J,\alpha} = -v_{b,00}^{-1}BF^{-1}, \left\|s_{b,0,J,\alpha}\right\|_2 = \mathcal{O}\left(H\right).$$

Applying Cauchy-Schwartz inequality, $\left|T_n\right|$ is bounded by

$$\left\{\left(s_{b,00} - v_{b,00}^{-1}\right)^2 + \left\|s_{b,0,J,\alpha}\right\|_2^2\right\}^{1/2} \times$$

$$\left[\left\{n^{-1}\textstyle\sum_{i=1}^n \sigma\left(\mathbf{X}_i\right)\varepsilon_i\right\}^2 + \displaystyle\sum_{J,\alpha}\left\{n^{-1}\textstyle\sum_{i=1}^n \sigma\left(\mathbf{X}_i\right)\varepsilon_i B_{J,\alpha}\left(X_{i\alpha}\right)\right\}^2\right]^{1/2}$$

$$= \mathcal{O}\left(H\right) \times \mathcal{O}_{a.s}\left(N^{1/2}n^{-1/2}\log n\right) = \mathcal{O}_{a.s}\left(n^{-1/2}\right).$$

This, together with (A.30) and (A.31) prove the Theorem. □

REFERENCES

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

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

Fan, J. Härdle, W. and Mammen, E. (1998). Direct estimation of low-dimensional components in additive models. *Ann. Statist.* **26**, 943–971.

Härdle, W., Hoffmann, L. and Moro, R. (2011). *Learning Machines Supporting Bankruptcy prediction. Statistical Tools in Finance and Insurance* (2nd ed.), Cizek, Härdle, Weron, Springer Verlag.

Hastie, T. J. and Tibshirani, R. J. (1990). *Generalized Additive Models.* Chapman and Hall, London.

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

Horowitz, J. Klemelä, J. and Mammen, E. (2006). Optimal estimation in additive regression. *Bernoulli* **12** 271–298.

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

Linton, O. B. and Nielsen, J. P. (1995). A kernel method of estimating structured nonparametric regression based on marginal integration. *Biometrika* **82** 93–100.

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

Linton, O. B. and Härdle, W. (1996). Estimation of additive regression models with known links. *Biometrika* **83** 529–540.

Liu, R. and Yang, L. (2010). Spline-backfitted kernel smoothing of additive coefficient model. *Econometric Theory* **26** 29–59.

Liu, R., Yang, L. and Härdle, W. (2011) Supplement to "Oracally efficient two-step estimation of generalized addtive model". Manuscript.

Ma, S. and Yang, L. (2011). Spline-backfitted kernel smoothing of partially linear additive model. *Journal of Statistical Planning and Inference* **141**, 204–219.

Portnoy, S. (2011). Local asymptotics for quantile smoothing splines. *Ann. Statist.* **25**, 414–434.

Song, Q. and Yang, L. (2010). Oracally efficient spline smoothing of nonlinear additive autoregression model with simultaneous confidence band. *Journal of Multivariate Analysis* **101**, 2008–2025.

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

Stone, C. J. (1986). The dimensionality reduction principle for generalized additive models. *Ann. Statist.* **14** 590–606.

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

Sunklodas, J. (1984). On the rate of convergence in the central limit theorem for strongly mixing random variables. *Lithuanian Mathematical Journal* **24** 182–190.

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

Wang, L. and Yang, L. (2007). Spline-backfitted kernel smoothing of nonlinear additive autoregression model. *Ann. Statist.* **35** 2474–2503.

Wang, J. and Yang, L. (2009). Efficient and fast spline-backfitted kernel smoothing of additive regression model. *Annals of the Institute of Statistical Mathematics* **61** 663–690.

Xue, L. and Liang, H. (2010). Polynomial spline estimation for a generalized additive coefficient model. *Scandinavian Journal of Statistics* **37** 26–46.

Xue, L. and Yang, L. (2006a). Additive coefficient modeling via polynomial spline. *Statistica Sinica* **16** 1423–1446.

Xue, L. and Yang, L. (2006b). Estimation of semiparametric additive coefficient model. *Journal of Statistical Planning and Inference* **136** 2506–2534.

Yang, L., Härdle, W. and Nielsen, J. P. (1999). Nonparametric autoregression with multiplicative volatility and additive mean. *J. Time Ser. Anal.* **20** 579–604.

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.

Yang, L., Park, B. U., Xue, L. and Härdle, W. (2006). Estimation and testing for varying coefficients in additive models with marginal integration. *J. Amer. Statist. Assoc.* **101** 1212–1227.

$d = 5$	n	$\mathrm{MISE}\,(\hat{m}_{\mathrm{SBK},1})$	$\mathrm{MISE}\,(\tilde{m}_{\mathrm{SBK},1})$	$\overline{\mathrm{EFF}}\,(\hat{m}_{\mathrm{SBK},1})$	$\mathrm{std}\,\{\mathrm{EFF}\,(\hat{m}_{\mathrm{SBK},1})\}$
$\rho = 0$, $a = 0$.	500	0.0548	0.0600	1.1120	0.2741
$r = 0.5$, $a = 0.5$.	500	0.1017	0.0944	1.0233	0.2796

Table 1

Example 1: the means and standard deviations of MISEs and EFFs of $\hat{m}_{\mathrm{SBK},1}$, $\tilde{m}_{\mathrm{SBK},1}$ *for* $d = 5$, $n = 500$.

$d = 5$	n	$\mathrm{MISE}\,(\hat{m}_{\mathrm{SBK},2})$	$\mathrm{MISE}\,(\tilde{m}_{\mathrm{SBK},2})$	$\overline{\mathrm{EFF}}\,(\hat{m}_{\mathrm{SBK},2})$	$\mathrm{std}\,\{\mathrm{EFF}\,(\hat{m}_{\mathrm{SBK},2})\}$
$r = 0$, $a = 0$.	500	0.0179	0.0271	1.5032	0.8965
$r = 0.5$, $a = 0.5$.	500	0.0365	0.4178	0.9977	0.4006

Table 2

Example 1: the means and standard deviations of MISEs and EFFs of $\hat{m}_{\mathrm{SBK},2}$, $\tilde{m}_{\mathrm{SBK},2}$ *for* $d = 5$, $n = 500$.

$d = 10$	n	$\mathrm{MISE}\,(\hat{m}_{\mathrm{SBK},1})$	$\mathrm{MISE}\,(\tilde{m}_{\mathrm{K},1})$	$\overline{\mathrm{EFF}}\,(\hat{m}_{\mathrm{SBK},1})$	$\mathrm{std}\,\{\mathrm{EFF}\,(\hat{m}_{\mathrm{SBK},1})\}$
$r = 0$, $a = 0$.	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

Table 3

Simulated example 2: the MISEs and EFFs of $\hat{m}_{\mathrm{SBK},1}$, $\tilde{m}_{\mathrm{K},1}$ *for* $d = 10$, $n = 500$, 1000, 1500, 2000.

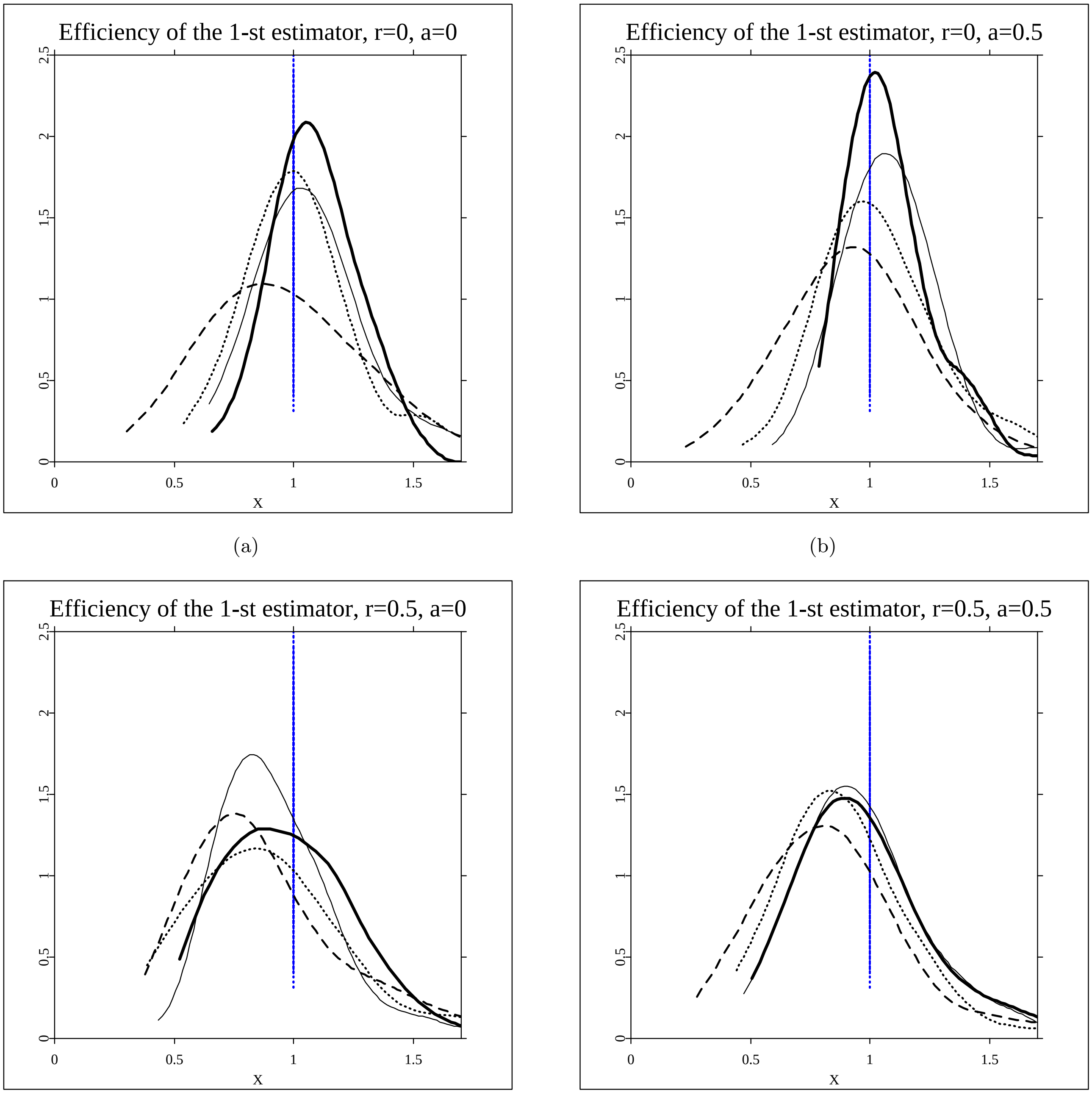
Ratio No.	Definition	Ratio No.	Definition
Z_1	Net_Income/Sales	Z_5	Cash/Total_Assets
Z_2	Operating_Income/Total_Assets	Z_6	Inventories/Sales
Z_3	Ebit/Total_Assets	Z_7	Accounts_Payable/Sales
Z_4	Total_Liabilities/Total_Assets	Z_8	log(Total_Assets)

TABLE 4
Real data example 3: Definitions of financial ratios.

Rong Liu
Department of Mathematics,
University of Toledo,
Toledo, OH 43606, USA
E-mail: rong.liu@utoledo.edu

Lijian Yang
Center for Advanced Statistics and Econometrics Research,
Soochow University,
Suzhou 215006,
People's Republic of China
and
Department of Statistics and Probability,
Michigan State University,
East Lansing, MI 48824, USA
E-mail: yang@stt.msu.edu

Wolfgang K. Härdle
Center for Applied Statistics and Economics,
Humboldt-Universität zu Berlin,
Unter den Linden 6,
10099 Berlin, Germany
E-mail: stat@wiwi.hu-berlin.de



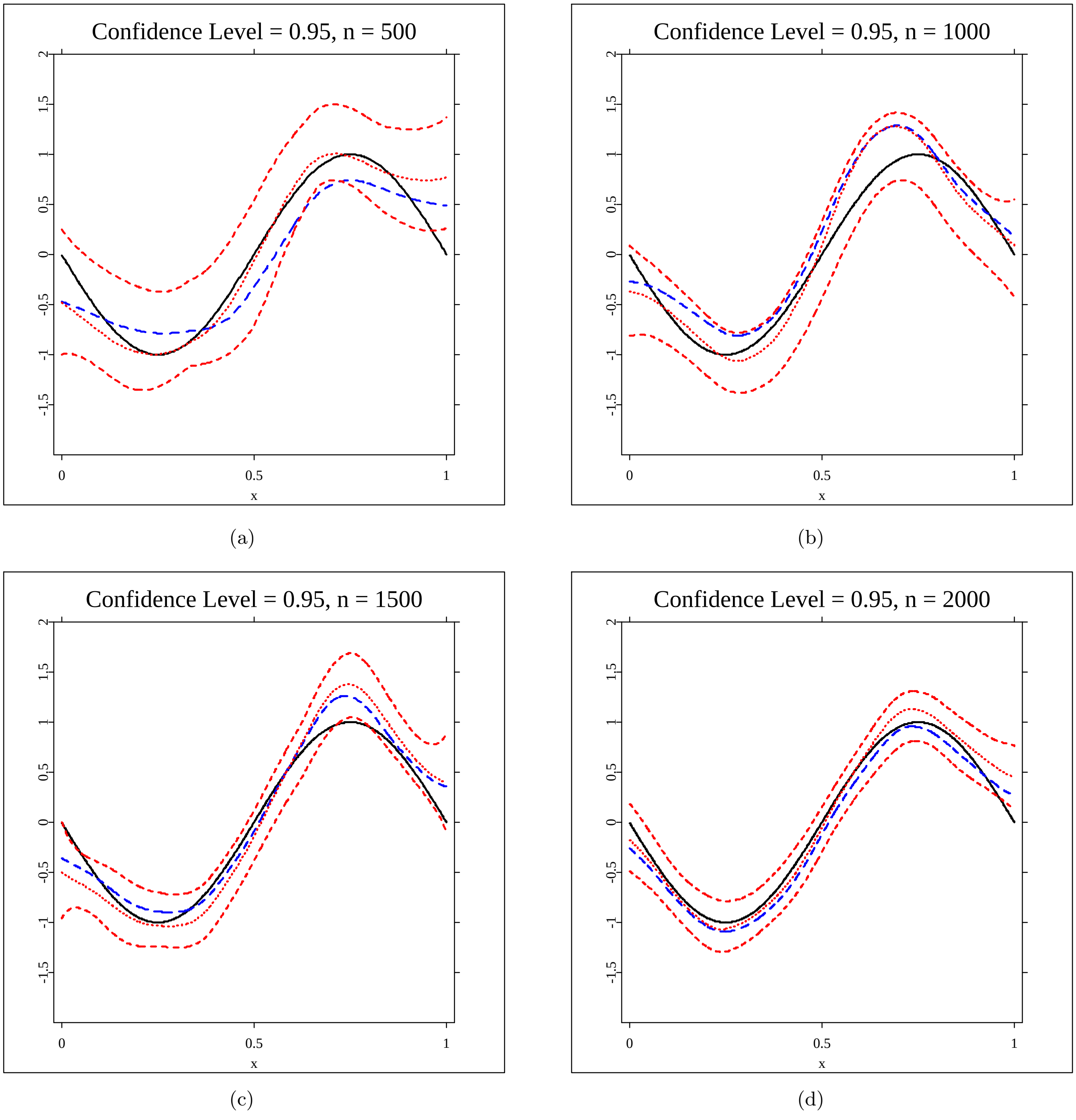
(a)

(b)

(c)

(d)

FIG 1. *Plots of empirical distribution of relative efficiency of* $n = 500$ *- dashed line,* $n = 1000$ *- dotted line,* $n = 1500$ *- thin solid line,*$n = 2000$ *- thick solid line for (a)* $r = 0, a = 0$*, (b)* $r = 0, a = 0.5$*, (c)* $r = 0.5, a = 0$*, (d)* $r = 0.5, a = 0.5$*.*



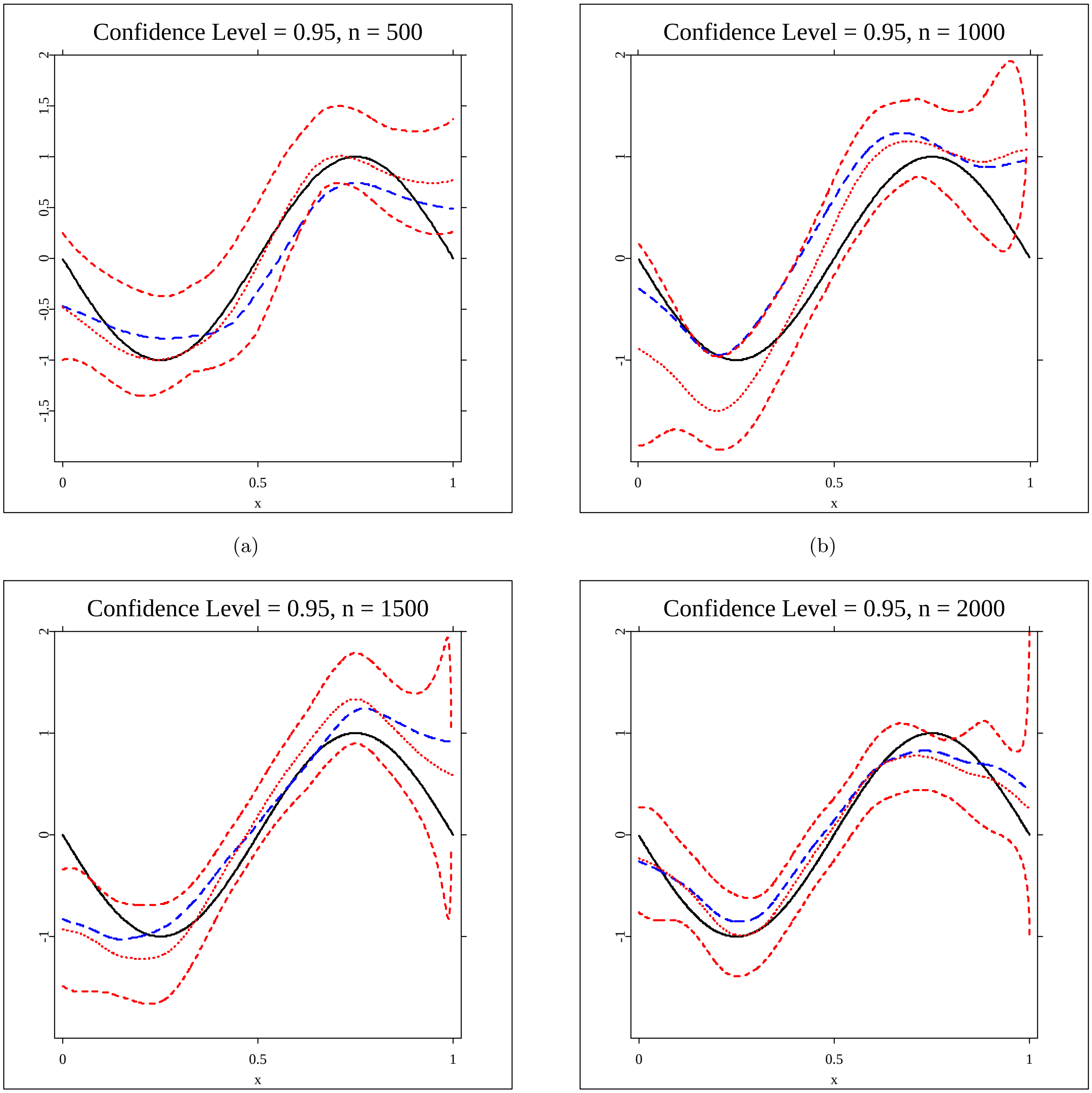
(a)

(b)

(c)

(d)

FIG 2. *Plots of* $m_1(x_1)$ *- solid line,* $\tilde{m}_{\mathrm{K},1}(x_1)$ *- dashed line, confidence bands and* $\hat{m}_{\mathrm{SBK},1}(x_1)$ *- three dotted lines for* $r = 0, a = 0$ *and (a)* $n = 500$*, (b)* $n = 1000$*, (c)* $n = 1500$*, (d)* $n = 2000$*.*



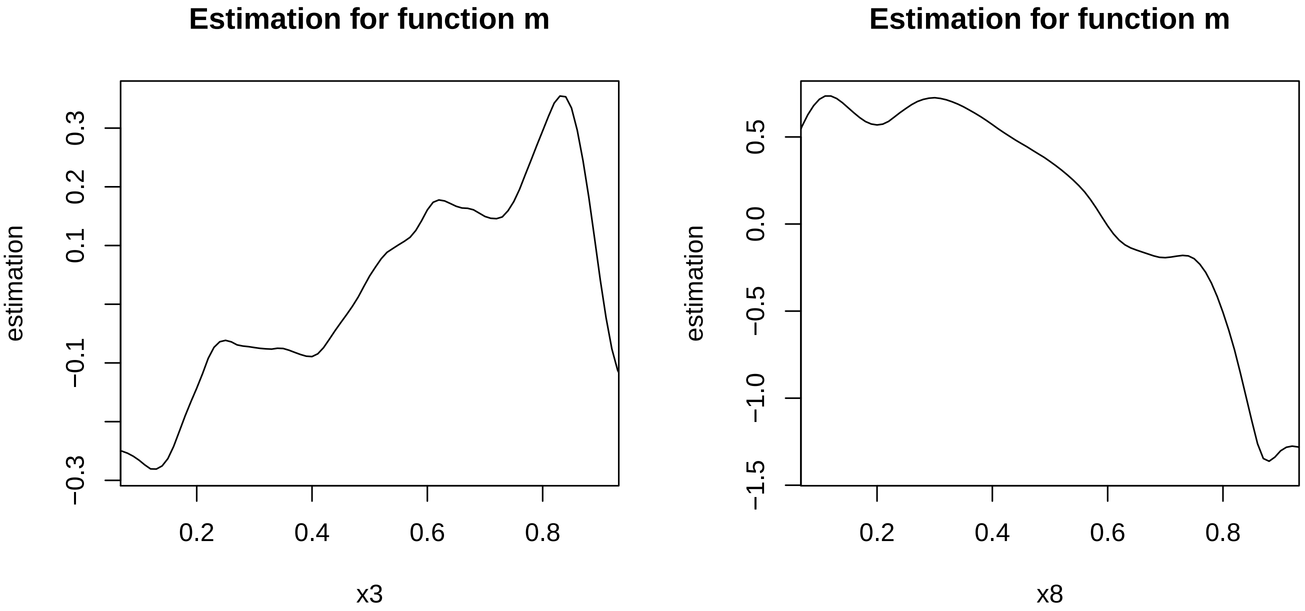
(a)

(b)

(c)

(d)

Fig 3. *Plots of* $m_1(x_1)$ *- solid line,* $\tilde{m}_{\mathrm{K},1}(x_1)$ *- dashed line, confidence bands and* $\hat{m}_{\mathrm{SBK},1}(x_1)$ *- three dotted lines for* $r = 0.5, a = 0.5$ *and (a)* $n = 500$*, (b)* $n = 1000$*, (c)* $n = 1500$*, (d)* $n = 2000$*.*



(a)

(b)

Fig 4. *Estimations for (a) $m_3(x)$ and (b) $m_8(x)$.*

SFB 649 Discussion Paper Series 2011

For a complete list of Discussion Papers published by the SFB 649, please visit http://sfb649.wiwi.hu-berlin.de.

001 "Localising temperature risk" by Wolfgang Karl Härdle, Brenda López Cabrera, Ostap Okhrin and Weining Wang, January 2011.

002 "A Confidence Corridor for Sparse Longitudinal Data Curves" by Shuzhuan Zheng, Lijian Yang and Wolfgang Karl Härdle, January 2011.

003 "Mean Volatility Regressions" by Lu Lin, Feng Li, Lixing Zhu and Wolfgang Karl Härdle, January 2011.

004 "A Confidence Corridor for Expectile Functions" by Esra Akdeniz Duran, Mengmeng Guo and Wolfgang Karl Härdle, January 2011.

005 "Local Quantile Regression" by Wolfgang Karl Härdle, Vladimir Spokoiny and Weining Wang, January 2011.

006 "Sticky Information and Determinacy" by Alexander Meyer-Gohde, January 2011.

007 "Mean-Variance Cointegration and the Expectations Hypothesis" by Till Strohsal and Enzo Weber, February 2011.

008 "Monetary Policy, Trend Inflation and Inflation Persistence" by Fang Yao, February 2011.

009 "Exclusion in the All-Pay Auction: An Experimental Investigation" by Dietmar Fehr and Julia Schmid, February 2011.

010 "Unwillingness to Pay for Privacy: A Field Experiment" by Alastair R. Beresford, Dorothea Kübler and Sören Preibusch, February 2011.

011 "Human Capital Formation on Skill-Specific Labor Markets" by Runli Xie, February 2011.

012 "A strategic mediator who is biased into the same direction as the expert can improve information transmission" by Lydia Mechtenberg and Johannes Münster, March 2011.

013 "Spatial Risk Premium on Weather Derivatives and Hedging Weather Exposure in Electricity" by Wolfgang Karl Härdle and Maria Osipenko, March 2011.

014 "Difference based Ridge and Liu type Estimators in Semiparametric Regression Models" by Esra Akdeniz Duran, Wolfgang Karl Härdle and Maria Osipenko, March 2011.

015 "Short-Term Herding of Institutional Traders: New Evidence from the German Stock Market" by Stephanie Kremer and Dieter Nautz, March 2011.

016 "Oracally Efficient Two-Step Estimation of Generalized Additive Model" by Rong Liu, Lijian Yang and Wolfgang Karl Härdle, March 2011.

SFB 649, Ziegelstraße 13a, D-10117 Berlin
http://sfb649.wiwi.hu-berlin.de

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".