Journal of Multivariate Analysis 101 (2010) 2008–2025

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Oracally efficient spline smoothing of nonlinear additive autoregression models with simultaneous confidence band

Qiongxia Song, Lijian Yang *

Michigan State University, East Lansing, MI 48824, USA

ARTICLE INFO

Article history:
Received 21 May 2009
Available online 27 April 2010

AMS 2000 subject classifications:
primary 62G08
62G15
secondary 62G05
62G20

Keywords:
B spline
Confidence band
Knots
Mixing
Oracle efficiency

ABSTRACT

Under weak conditions of smoothness and mixing, we propose spline-backfitted spline (SBS) estimators of the component functions for a nonlinear additive autoregression model that is both computationally expedient for analyzing high dimensional large time series data, and theoretically reliable as the estimator is oracally efficient and comes with asymptotically simultaneous confidence band. Simulation evidence strongly corroborates with the asymptotic theory.

1. Introduction

Non- and semiparametric smoothing has been proven to be useful for analyzing complex time series data due to the flexibility to "let the data speak for themselves". One unavoidable issue in high dimensional smoothing is the "curse of dimensionality", i.e., the poor convergence rate of nonparametric estimation of multivariate functions. The additive regression model of Hastie and Tibshirani [7] has been adapted by Chen and Tsay [2] to autoregression and found wide use in recent years to reduce dimension in nonparametric smoothing of time series. A nonlinear additive autoregressive model (NAAR) is of the form

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

where the sequence $\left\{Y_i, \mathbf{X}_i^T\right\}_{i=1}^n$ is a length n realization of a $(d+1)$-dimensional strictly stationary process, the d-variate functions $m(\cdot)$ and $\sigma(\cdot)$ are the mean and standard deviation of the response Y_i conditional on the predictor vector $\mathbf{X}_i = \{X_{i1}, \ldots, X_{id}\}^T$, and $E\left(\varepsilon_i|\mathbf{X}_i\right) = 0$, $E\left(\varepsilon_i^2|\mathbf{X}_i\right) = \sigma^2\left(\mathbf{X}_i\right)$. In the context of NAAR, each predictor $X_{i\gamma}$, $1 \le \gamma \le d$ can be observed lagged values of Y_i, such as $X_{i\gamma} = Y_{i-\gamma}$, or of a different times series. The additive component functions $\left\{m_\gamma\left(\cdot\right)\right\}_{\gamma=1}^d$ are subjected to the identifiability condition in (5).

* Corresponding author.
E-mail address: yang@stt.msu.edu (L. Yang).

0047-259X/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmva.2010.04.004

The inference of model (1) centers on the estimation and testing of $\left\{m_\gamma(\cdot)\right\}_{\gamma=1}^d$. The marginal integration method of Tjøstheim and Auestad [24] and Linton and Nielsen [15] came with asymptotic distribution, which was extended in [22] to include second order interactions. Other related works include Fan and Li [6], Yang, Park, Xue and Härdle [29] and Lu, Lundervold, Tjøstheim and Yao [16]. The backfitting idea promoted by [7] was made rigorous in a more complicated form of smooth backfitting by Mammen, Linton and Nielsen [17] and popularized by Nielsen and Sperlich [19]. These kernel based methods are extremely computational intensive, limiting their use for high dimension d, see [18] for numerical comparison of these methods. Spline method of Stone [23] had been extended in parallel to NAAR models in [10], which are fast and easy to implement but lack of limiting distribution. For applications of additive model in medical and environmental research, see [13,20,21].

The two-step estimators of Linton [14] for model (1) possess oracle efficiency and are theoretically superior to the aforementioned estimators of $\left\{m_\gamma(\cdot)\right\}_{\gamma=1}^d$. If all components $\left\{m_\beta(\cdot)\right\}_{\beta=1,\beta\neq\gamma}^d$ and the constant c were known and removed from the responses, one could estimate $m_\gamma(\cdot)$ from the univariate data $\left\{Y_{i\gamma}, X_{i\gamma}\right\}_{i=1}^n$ in which $\left\{Y_{i\gamma}\right\}_{i=1}^n$ are latent oracle responses to the γth covariate $\left\{X_{i\gamma}\right\}_{i=1}^n$,

$$Y_{i\gamma} = m_\gamma\left(X_{i\gamma}\right) + \varepsilon_i = Y_i - c - \sum_{\beta=1,\beta\neq\gamma}^{d} m_\beta\left(X_{i\beta}\right), \quad 1 \le i \le n, 1 \le \gamma \le d. \tag{2}$$

The key idea of [14] is to replace the true $\left\{m_\beta(\cdot)\right\}_{\beta=1,\beta\neq\gamma}^d$ and c above by some initial kernel estimates, create a pseudo-univariate data $\left\{\hat{Y}_{i\gamma}, X_{i\gamma}\right\}_{i=1}^n$, and establish the asymptotic equivalence of kernel/local polynomial estimators of $m_\gamma(\cdot)$ using either unobservable $\left\{Y_{i\gamma}, X_{i\gamma}\right\}_{i=1}^n$ or $\left\{\hat{Y}_{i\gamma}, X_{i\gamma}\right\}_{i=1}^n$. Recently, faster oracally efficient estimators have been developed for NAAR time series data by Horowitz and Mammen [8], Wang and Yang [25], making use of orthogonal series/spline initial estimates. The second step estimation is done by kernel method, with pointwise asymptotic distribution. For the sake of discussion, we call the two-step estimator of [14] kernel+kernel, of [8] orthogonal series+kernel and of [25] spline+kernel.

For the NAAR time series models, however, none of the existing methods provide any simultaneous confidence band for $m_\gamma(\cdot)$. To address this need, we propose an all new spline+spline oracally efficient estimator that is theoretically superior as it comes with an asymptotically simultaneous confidence band for $m_\gamma(\cdot)$, and also computationally more expedient than any existing estimators due to the use of spline instead of kernel in all steps. The asymptotically simultaneous confidence band is that of an univariate regression function in Wang and Yang [26], and is most convenient for inference in the global shape of function $m_\gamma(\cdot)$. Such confidence band methodology has been applied to compare the dependence of corn, soybean and wheat crop yields on wetness index under various conditions, see [9]. The spline+spline method is asymptotically oracally efficient as the spline+kernel method of [25], but can be hundreds of times faster in terms of computing, see the comparison in Table 2. We see little hope of further reducing the computing burden for model (1) over the proposed spline+spline method and still retaining the simultaneous confidence band and oracle efficiency. It seems that the only alternative worth exploring is to use penalized spline instead of B spline smoothing in the second step. For theoretical properties of penalized spline smoothing, see [11,12].

The rest of the paper is organized as follows. Section 2 describes the spline-backfitted spline (SBS) estimators and presents the main theoretical results. Section 3 illustrates the idea of proof via decomposition of error. Simulation results are showed in Section 4. Most of the technical proofs are in the Appendix.

2. The SBS estimator

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

While X_γ may be distributed on $(-\infty, \infty)$, estimation of m_γ is carried out only on compact intervals, and without loss of generality, we take all intervals to be $[0,1]$, $1 \le \gamma \le d$. Let $0 = t_0 < t_1 < \cdots < t_{N+1} = 1$ be a sequence of equally spaced knots, dividing $[0,1]$ into $(N+1)$ subintervals of length $h = h_n = 1/(N+1)$ with a preselected integer $N \sim n^{1/5}$ given in Assumption (A5), and let $0 = t_0^* < t_1^* < \cdots < t_{N^*+1}^* = 1$ be another sequence of equally spaced knots, dividing $[0,1]$ into (N^*+1) subintervals of length $H = H_n = (N^*+1)^{-1}$ where $N^* \sim n^{2/5}\log n$ is another preselected integer, see Assumption (A5). Next, we define the constant spline basis I_{J^*} for step one and the linear spline basis b_J for step two de Boor ([4], page 89) as follows,

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

$$I_{J^*}(x) = \begin{cases} 1 & J^*H \le x < \left(J^*+1\right)H, \\ 0 & \text{otherwise}, \end{cases} \quad, 1 \le J^* \le N^*,$$

$$b_J(x) = K\left(\frac{x-t_J}{h}\right), \quad 1 \le J \le N+1, \qquad K(u) = (1-|u|)_+ .$$

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

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

$$W_{i\gamma} = 1\left(0 \le X_{i\gamma} \le 1\right), \quad 1 \le i \le n, 1 \le \gamma \le d, \tag{3}$$
$$\mathbf{W}_\gamma = \mathrm{diag}(W_{1\gamma}, \ldots, W_{n\gamma}), \quad 1 \le \gamma \le d,$$

$$W_i^* = 1\left(0 \le \mathbf{X}_i \le 1\right) = \prod_{\gamma=1}^{d} W_{i\gamma}, \quad 1 \le i \le n, \tag{4}$$
$$\mathbf{W}^* = \mathrm{diag}(W_1^*, \ldots, W_n^*),$$

and impose on additive component functions the identifiability condition

$$Em_\gamma\left(X_{i\gamma}\right) W_i^* \equiv 0, \quad 1 \le \gamma \le d. \tag{5}$$

Define next a weighted spline estimator of m as

$$\hat{m}(\mathbf{x}) = \underset{g\in G^*}{\mathrm{argmin}} \sum_{i=1}^{n} \{Y_i - g(\mathbf{X}_i)\}^2 W_i^* = \hat{\lambda}'_0 + \sum_{\gamma=1}^{d}\sum_{J^*=1}^{N^*} \hat{\lambda}'_{J^*,\gamma} I_{J^*}\left(x_\gamma\right), \tag{6}$$

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

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

Pilot estimator of each component function is

$$\hat{m}_\gamma\left(x_\gamma\right) = \sum_{J^*=1}^{N^*} \hat{\lambda}'_{J^*,\gamma} \left\{I_{J^*}\left(x_\gamma\right) - n^{-1}\sum_{i=1}^{n} I_{J^*}\left(X_{i\gamma}\right) W_i^*\right\}, \quad 1 \le \gamma \le d \tag{7}$$

which satisfies the empirical analog of (5): $n^{-1}\sum_{i=1}^{n} \hat{m}_\gamma\left(X_{i\gamma}\right) W_i^* = 0, 1 \le \gamma \le d$. These pilot estimators are used to define pseudo-responses $\hat{Y}_{i\gamma}, \forall 1 \le \gamma \le d$, which approximate the "oracle" responses $Y_{i\gamma}$ in (2). Specifically, we define $\hat{Y}_{i\gamma} = Y_i - \hat{c} - \sum_{\beta=1,\beta\neq\gamma}^{d} \hat{m}_\beta\left(X_{i\beta}\right)$, where $\hat{c} = \overline{Y}_n = n^{-1}\sum_{i=1}^{n} Y_i$, which is a $\sqrt{n}$-consistent estimator of c by central limit theorem for strongly mixing sequences. Correspondingly, we denote vectors

$$\hat{\mathbf{Y}}_\gamma = \left\{\hat{Y}_{1\gamma}, \ldots, \hat{Y}_{n\gamma}\right\}^T, \qquad \mathbf{Y}_\gamma = \left\{Y_{1\gamma}, \ldots, Y_{n\gamma}\right\}^T. \tag{8}$$

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

$$\hat{m}_{\gamma,\mathrm{SBS}}\left(x_\gamma\right) = \underset{g_\gamma\in G_\gamma}{\mathrm{argmin}} \sum_{i=1}^{n} \left\{\hat{Y}_{i\gamma} - g_\gamma\left(X_{i\gamma}\right)\right\}^2 W_{i\gamma},$$

$$\tilde{m}_{\gamma,\mathrm{S}}\left(x_\gamma\right) = \underset{g_\gamma\in G_\gamma}{\mathrm{argmin}} \sum_{i=1}^{n} \left\{Y_{i\gamma} - g_\gamma\left(X_{i\gamma}\right)\right\}^2 W_{i\gamma}. \tag{9}$$

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

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

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

$$\alpha(k) = \sup_{B\in\sigma\{\mathbf{Z}_s, s\le t\}, C\in\sigma\{\mathbf{Z}_s, s\ge t+k\}} |P(B\cap C) - P(B)P(C)|, \quad k \ge 1. \tag{10}$$

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

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

(A5) *The number of interior knots in estimation step one* $N^* \sim n^{2/5}\log n$, *i.e.,* $c_{N^*}n^{2/5}\log n \le N^* \le C_{N^*}n^{2/5}\log n$ *for some positive constants* c_{N^*}, C_{N^*}. *The number of interior knots in estimation step two* $N \sim n^{1/5}$.

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

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

Theorem 1. *Under Assumptions* (A1)–(A5), *as* $n \to \infty$, *the SBS estimator* $\hat{m}_{\gamma,\mathrm{SBS}}\left(x_\gamma\right)$ *and the oracle smoother* $\tilde{m}_{\gamma,\mathrm{S}}\left(x_\gamma\right)$ *given in (9) satisfy*

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

Theorem 1 provides that the maximal deviation of $\hat{m}_{\gamma,\mathrm{SBS}}\left(x_\gamma\right)$ from $\tilde{m}_{\gamma,\mathrm{S}}\left(x_\gamma\right)$ over [0, 1] is of the order $O_p(n^{-2/5}(\log n)^{-1}) = o_p\left(n^{-2/5}(\log n)^{1/2}\right)$, which is needed for the maximal deviation of $\hat{m}_{\gamma,\mathrm{SBS}}\left(x_\gamma\right)$ from $m_\gamma\left(x_\gamma\right)$ over [0, 1] and the maximal deviation of $\tilde{m}_{\gamma,\mathrm{S}}\left(x_\gamma\right)$ from $m_\gamma\left(x_\gamma\right)$ to have the same asymptotic distribution, of order $n^{-2/5}(\log n)^{1/2}$. The estimator $\hat{m}_{\gamma,\mathrm{SBS}}\left(x_\gamma\right)$ is therefore asymptotically oracally efficient, i.e., it is asymptotically equivalent to the oracle smoother $\tilde{m}_{\gamma,\mathrm{S}}\left(x_\gamma\right)$ and in particular, the next theorem follows. The simultaneous confidence band given in (11) has width of order $n^{-2/5}(\log n)^{1/2}$ at any point $x_\gamma \in [0,1]$, consistent with published works on nonparametric simultaneous confidence bands such as [28,3].

Theorem 2. *Under Assumptions* (A1)–(A5), *for any* $p \in (0,1)$, *as* $n \to \infty$, *an asymptotic* $100(1-p)\%$ *simultaneous confidence band for* $m_\gamma\left(x_\gamma\right)$ *is*

$$\hat{m}_{\gamma,\mathrm{SBS}}\left(x_\gamma\right) \pm 2\hat{\sigma}_\gamma\left(x_\gamma\right)\left\{3\mathbf{\Delta}^T\left(x_\gamma\right)\Xi_{j(x_\gamma)}\mathbf{\Delta}\left(x_\gamma\right)\log(N+1)/2\hat{f}_\gamma\left(x_\gamma\right)nh\right\}^{1/2} \times\left[1-\{2\log(N+1)\}^{-1}\left[\log(p/4)+\frac{1}{2}\{\log\log(N+1)+\log 4\pi\}\right]\right], \tag{11}$$

where $\hat{\sigma}_\gamma\left(x_\gamma\right)$ *and* $\hat{f}_\gamma\left(x_\gamma\right)$ *are some consistent estimators of* $\sigma_\gamma\left(x_\gamma\right)$ *and* $f_\gamma\left(x_\gamma\right)$, $j\left(x_\gamma\right) = \min\left\{\left[x_\gamma/h\right], N\right\}$, $\delta\left(x_\gamma\right) = \left\{x_\gamma - t_{j(x_\gamma)}\right\}/h$, *and*

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

$$\mathbf{\Xi}_j = \begin{pmatrix} l_{j+1,j+1} & l_{j+1,j+2} \\ l_{j+2,j+1} & l_{j+2,j+2}\end{pmatrix}, \quad 0 \le j \le N,$$

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

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

We refer the proof of the theorem to [26].

3. Decomposition

In this section, we provide insight on the proof of Theorem 1. Recalling the notation of W_i^* and $W_{i\gamma}$ defined in (4) and (3), for any functions ϕ,φ on $[0,1]^d$, define the empirical inner product, empirical norm and empirical mean restricted on $[0,1]^d$ as $\langle\phi,\varphi\rangle_{2,n}^* = n^{-1}\sum_{i=1}^n \phi(\mathbf{X}_i)\varphi(\mathbf{X}_i)W_i^*$, $\|\phi\|_{2,n}^{*2} = n^{-1}\sum_{i=1}^n \phi^2(\mathbf{X}_i)W_i^*$, $E_n^*\phi = n^{-1}\sum_{i=1}^n \phi(\mathbf{X}_i)W_i^* = \langle 1,\phi\rangle_{2,n}^*$ respectively. In addition, if functions ϕ,φ are $L^2[0,1]^d$-integrable, define the theoretical inner product and its corresponding theoretical L^2 norm as $\langle\phi,\varphi\rangle_2^* = E\left\{\phi(\mathbf{X}_i)\varphi(\mathbf{X}_i)W_i^*\right\}$, $\|\phi\|_2^{*2} = E\left\{\phi^2(\mathbf{X}_i)W_i^*\right\}$. A function ϕ is called theoretically centered (empirically centered) if $E\phi W_i^* = 0$ ($E_n^*\phi = 0$). The additive component function m_γ and its pilot estimator $\hat{m}_\gamma$ defined in (7) are therefore theoretically centered (empirically centered). In the second step, for any functions ϕ,φ on $[0,1]$, for any $1\le\gamma\le d$, similarly define $\langle\phi,\varphi\rangle_{2,n,\gamma} = n^{-1}\sum_{i=1}^n \phi(X_{i\gamma})\varphi(X_{i\gamma})W_{i\gamma}$, $\|\phi\|_{2,n,\gamma}^2 = n^{-1}\sum_{i=1}^n \phi^2(X_{i\gamma})W_{i\gamma}$, $E_{n,\gamma}\phi = n^{-1}\sum_{i=1}^n \phi(X_{i\gamma})W_{i\gamma} = \langle 1,\phi\rangle_{2,n,\gamma}$ respectively. In addition, if functions ϕ,φ are $L^2[0,1]$-integrable, define the theoretical inner product and its corresponding theoretical L^2 norm as $\langle\phi,\varphi\rangle_{2,\gamma} = E\left\{\phi(X_{i\gamma})\varphi(X_{i\gamma})W_{i\gamma}\right\}$, $\|\phi\|_{2,\gamma}^2 = E\left\{\phi^2(X_{i\gamma})W_{i\gamma}\right\}$.

The function space G_γ introduced in Section 2 is expressed more conveniently for asymptotic analysis via the following standardized B spline basis

$$B_{J,\gamma}(x_\gamma) = \frac{b_J(x_\gamma)}{\|b_J\|_{2,\gamma}},\quad 0\le J\le N+1. \tag{12}$$

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

$$B_{J^*,\gamma}^*(x_\gamma) = \frac{b_{J^*,\gamma}^*(x_\gamma)}{\left\|b_{J^*,\gamma}^*\right\|_2},\quad 1\le\gamma\le d,\ 1\le J^*\le N^*, \tag{13}$$

in which

$$b_{J^*,\gamma}^*(x_\gamma) = I_{J^*+1,\gamma}(x_\gamma) - \frac{c_{J^*+1,\gamma}}{c_{J^*,\gamma}} I_{J^*,\gamma}(x_\gamma),\ c_{J^*,\gamma} = \left\langle 1, I_{J^*,\gamma}\right\rangle_2. \tag{14}$$

Simple linear algebra shows that

$$\hat{m}(\mathbf{x}) = \hat{\lambda}_0 + \sum_{\gamma=1}^d\sum_{J^*=1}^{N^*}\hat{\lambda}_{J^*,\gamma}B_{J^*,\gamma}^*(x_\gamma),\ \mathbf{x}\in[0,1]^d \tag{15}$$

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

$$\left\{\hat{\lambda}_0,\hat{\lambda}_{1,1},\ldots,\hat{\lambda}_{N^*,d}\right\}^T = \underset{R^{d(N^*)+1}}{\operatorname{argmin}}\sum_{i=1}^n\left\{Y_i - \lambda_0 - \sum_{\gamma=1}^d\sum_{J^*=1}^{N^*}\lambda_{J^*,\gamma}B_{J^*,\gamma}^*(X_{i\gamma})\right\}^2 W_i^*. \tag{16}$$

Define for any n-dimensional vector $\mathbf{\Lambda} = \{\Lambda_i\}_{i=1}^n$, the spline function constructed from the projection of $\mathbf{\Lambda}$ on the inner product space $\left(G_n,\langle\cdot,\cdot\rangle_{2,n}\right)$ as $\mathbf{P}_n\mathbf{\Lambda}(\mathbf{x}) = \hat{\lambda}_0 + \sum_{\gamma=1}^d\sum_{J^*=1}^{N^*}\hat{\lambda}_{J^*,\gamma}B_{J^*,\gamma}^*(x_\gamma)$, with coefficients $\left(\hat{\lambda}_0,\hat{\lambda}_{1,1},\ldots,\hat{\lambda}_{N^*,d}\right)$ given in (16) with Y_i's replaced by Λ_i's. The multivariate function $\mathbf{P}_n\mathbf{\Lambda}(\mathbf{x})$ has empirically centered components $\mathbf{P}_{n,\gamma}\mathbf{\Lambda}(x_\gamma)$, $\gamma = 1,\ldots,d$

$$\mathbf{P}_{n,\gamma}\mathbf{\Lambda}(x_\gamma) = \sum_{J^*=1}^{N^*}\hat{\lambda}_{J^*,\gamma}\left\{B_{J^*,\gamma}^*(x_\gamma) - n^{-1}\sum_{i=1}^n B_{J^*,\gamma}^*(X_{i\gamma})W_i^*\right\}. \tag{17}$$

The estimators $\hat{m}(\mathbf{x})$, $\hat{m}_\gamma(x_\gamma)$ in (15) and (7) are rewritten as $\hat{m}(\mathbf{x}) = \mathbf{P}_n\mathbf{Y}(\mathbf{x})$, $\hat{m}_\gamma(x_\gamma) = \mathbf{P}_{n,\gamma}\mathbf{Y}(x_\gamma)$. For linear operators $\mathbf{P}_n, \mathbf{P}_{n,\gamma}, \gamma = 1, \ldots, d$, using the relation $\mathbf{Y} = \mathbf{m} + \mathbf{E}$, where the signal and noise vectors are $\mathbf{m} = \{m(\mathbf{X}_i)\}_{i=1}^n$, $\mathbf{E} = \{\varepsilon_i\}_{i=1}^n$, one has the following decomposition for $\gamma = 1, \ldots, d$

$$\hat{m}(\mathbf{x}) = \tilde{m}(\mathbf{x}) + \tilde{\varepsilon}(\mathbf{x}), \qquad \hat{m}_\gamma(x_\gamma) = \tilde{m}_\gamma(x_\gamma) + \tilde{\varepsilon}_\gamma(x_\gamma), \tag{18}$$

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

$$\begin{aligned} \tilde{m}(\mathbf{x}) &= \mathbf{P}_n\mathbf{m}(\mathbf{x}), \qquad \tilde{m}_\gamma(x_\gamma) = \mathbf{P}_{n,\gamma}\mathbf{m}(x_\gamma), \\ \tilde{\varepsilon}(\mathbf{x}) &= \mathbf{P}_n\mathbf{E}(\mathbf{x}), \qquad \tilde{\varepsilon}_\gamma(x_\gamma) = \mathbf{P}_{n,\gamma}\mathbf{E}(x_\gamma). \end{aligned} \tag{19}$$

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

$$\mathbf{B}^*(\mathbf{x}) = \left\{1, B_{1,1}^*(x_1), \ldots, B_{N^*,d}^*(x_d)\right\}, \qquad \mathbf{B}^* = \left\{\mathbf{B}^*(\mathbf{X}_1), \ldots, \mathbf{B}^*(\mathbf{X}_n)\right\}^T. \tag{20}$$

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

$$\left\{\begin{matrix} 1 & \mathbf{0}_{dN^*}^T \\ \mathbf{0}_{dN^*} & \left\langle B_{J^*,\gamma}^*, B_{J^{*\prime},\gamma'}^*\right\rangle_{2,n}^* \end{matrix}\right\}^{-1}_{\substack{1\le\gamma,\gamma'\le d,\\ 1\le J^*,J^{*\prime}\le N^*}} \left\{\begin{matrix} \dfrac{1}{n}\displaystyle\sum_{i=1}^n W_i^*\varepsilon_i \\ \dfrac{1}{n}\displaystyle\sum_{i=1}^n B_{J^*,\gamma}^*(X_{i\gamma})W_i^*\varepsilon_i \end{matrix}\right\}_{\substack{1\le J^*\le N^*\\ 1\le\gamma\le d}}, \tag{21}$$

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

The second step spline smoothing is interpreted similarly. For notational simplicity, take $\gamma = 1$ and denote $\mathbf{X}_{i,\_1} = (X_{i2}, \ldots, X_{id})^T$ for $1 \le i \le n$, and $\mathbf{x}_{\_1} = (x_2, \ldots, x_d)^T$. Denote $B_{J^*,\_1}^*(\mathbf{x}_{\_1}) = (B_{J^*,2}^*(x_2), \ldots, B_{J^*,d}^*(x_d))^T$, and $m_{\_1}(\mathbf{x}_{\_1})$, $\hat{m}_{\_1}(\mathbf{x}_{\_1})$, $\tilde{m}_{\_1}(\mathbf{x}_{\_1})$, $\tilde{\varepsilon}_{\_1}(\mathbf{x}_{\_1})$ similarly. Define $\mathbf{B}(x_1) = \left\{B_{0,1}(x_1), \ldots, B_{N+1,1}(x_1)\right\}$, $\mathbf{B} = \{\mathbf{B}(X_{11}), \ldots, \mathbf{B}(X_{1n})\}^T$, then $\hat{m}_{1,\mathrm{SBS}}(x_1) = \mathbf{B}(x_1)\left(\frac{\mathbf{B}^T\mathbf{W}\mathbf{B}}{n}\right)^{-1}\frac{\mathbf{B}^T}{n}\mathbf{W}\hat{\mathbf{Y}}_1$, $\tilde{m}_{1,\mathrm{S}}(x_1) = \mathbf{B}(x_1)\left(\frac{\mathbf{B}^T\mathbf{W}\mathbf{B}}{n}\right)^{-1}\frac{\mathbf{B}^T}{n}\mathbf{W}\mathbf{Y}_1$, where $\hat{\mathbf{Y}}_1$ and $\mathbf{Y}_1$ are defined in (8).

Making use of the definition of $\hat{c}$ and the decomposition (18), the difference between the smoothed backfitted estimator $\hat{m}_{1,\mathrm{SBS}}(x_1)$ and the smoothed "oracle" estimator $\tilde{m}_{1,\mathrm{S}}(x_1)$, both given above, is

$$\begin{aligned} \tilde{m}_{1,\mathrm{S}}(x_1) - \hat{m}_{1,\mathrm{SBS}}(x_1) &= \mathbf{B}(x_1)\left(\frac{\mathbf{B}^T\mathbf{W}\mathbf{B}}{n}\right)^{-1}\frac{\mathbf{B}^T}{n}\mathbf{W}\left(\mathbf{Y}_1 - \hat{\mathbf{Y}}_1\right) \\ &= \mathbf{B}(x_1)\left(\frac{\mathbf{B}^T\mathbf{W}\mathbf{B}}{n}\right)^{-1}\left(\frac{\mathbf{B}^T}{n}\mathbf{W}(-\hat{c} + c) + \mathbf{\Psi}_b + \mathbf{\Psi}_v\right), \end{aligned}$$

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

$$\mathbf{\Psi}_b = \left\{n^{-1}\sum_{i=1}^n B_{J,1}(X_{i1})W_i^*\left\{\tilde{m}_{\_1}(\mathbf{X}_{i,\_1})^T - m_{\_1}(\mathbf{X}_{i,\_1})^T\right\}\mathbf{1}_{d-1}\right\}_{J=1}^{N+1}, \tag{22}$$

$$\mathbf{\Psi}_v = \left\{n^{-1}\sum_{i=1}^n B_{J,1}(X_{i1})W_i^*\tilde{\varepsilon}_{\_1}(\mathbf{X}_{i,\_1})^T\mathbf{1}_{d-1}\right\}_{J=1}^{N+1}, \tag{23}$$

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

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

4. Simulation example

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

The number of interior knots N^* and N for the spline estimation are calculated as $N^* = \min([c_{11}n^{2/5}\log n] + c_{12} + 1, [(n/2 - 1)d^{-1}])$, and $N = \left[c_{21}n^{1/5}\right] + c_{22} + 1$, in which $[a]$ denotes the integer part of a. Tuning constants $c_{11} = 5$, $c_{21} = 3$, $c_{12} = c_{22} = 1$ worked well, and we used them by default. The additional constraint that $N^* \le (n/2 - 1)d^{-1}$ ensures that the number of terms in the linear least squares problem (16), $1 + dN^*$, is no greater than $n/2$.

Table 1
Coverage frequencies from 500 replications.

	r	$n = 100$	$n = 500$	$n = 1000$
	0	0.86	0.972	0.966
$d = 4$	0.3	0.876	0.956	0.964
	0	0.0848	0.974	0.97
$d = 10$	0.3	0.842	0.962	0.966

Table 2
Comparison of computing time of model (24).

Method	$n = 100$	$n = 500$	$n = 1000$
SPBK	0.09	7.8	54
SBS	0.007	0.064	0.32
Ratio	12.88	121.88	168.75

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

Consider the following nonlinear additive heteroscedastic model

$$Y_t = \sum_{\gamma=1}^{d} \sin\left(2\pi X_{t\gamma}\right) + \varepsilon_t, \qquad \varepsilon_t \overset{i.i.d}{\sim} N\left(0, \sigma^2\left(\mathbf{X}_t\right)\right), \tag{24}$$

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

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

with stationary distribution $\mathbf{Z}_t = (Z_{t1}, \ldots, Z_{td})^T \sim N\left(0_d, \left(1-a^2\right)^{-1}\boldsymbol{\Sigma}\right)$. Hence $\{\mathbf{X}_t\}_{t=1}^n$ is a sequence of geometrically strong mixing random variables with marginal distribution $U[-0.5, 0.5]$. The standard deviation function is $\sigma\left(\mathbf{X}_t\right) = \sigma_0 \frac{1}{2} \cdot \frac{5-\exp\left(\sum_{\gamma=1}^{d}|X_{t\gamma}|/d\right)}{5+\exp\left(\sum_{\gamma=1}^{d}|X_{t\gamma}|/d\right)}$, $\sigma_0 = 0.5$, which ensures that our design is heteroscedastic.

The SBS estimator $\hat{m}_{\gamma,\mathrm{SBS}}\left(x_\gamma\right)$ and the oracle smoother $\tilde{m}_{\gamma,\mathrm{S}}\left(x_\gamma\right)$ are compared in terms of coverage probabilities of confidence bands for sample sizes $n = 100, 500, 1000$, with confidence level $1 - p = 0.95$. Table 1 contains the coverage probabilities as the percentage of complete coverage of the first true curve $\sin(2\pi x)$ at all data points $\{X_{t1}\}_{t=1}^n$ by the confidence bands in (11), over 500 replications of sample size n, for $d = 4$, 10 and $\rho = 0, 0.3$. The results are satisfactory as the empirical probabilities rapidly become greater than the nominal probability of 0.95 as n becomes large. To show that the SBS estimator $\hat{m}_{\gamma,\mathrm{SBS}}\left(x_\gamma\right)$ is as efficient as the oracle smoother $\tilde{m}_{\gamma,\mathrm{S}}\left(x_\gamma\right)$, we define the empirical relative efficiency of $\hat{m}_{\gamma,\mathrm{SBS}}\left(x_\gamma\right)$ with respect to $\tilde{m}_{\gamma,\mathrm{S}}\left(x_\gamma\right)$ as

$$\mathrm{eff}_\gamma = \left[\frac{\sum_{t=1}^{n}\left\{\tilde{m}_{\gamma,\mathrm{S}}\left(X_{t\gamma}\right) - m_\gamma\left(X_{t\gamma}\right)\right\}^2 1_{\left(0\le X_{t\gamma}\le 1\right)}}{\sum_{t=1}^{n}\left\{\hat{m}_{\gamma,\mathrm{SBS}}\left(X_{t\gamma}\right) - m_\gamma\left(X_{t\gamma}\right)\right\}^2 1_{\left(0\le X_{t\gamma}\le 1\right)}}\right]^{1/2}. \tag{25}$$

Theorem 1 indicates that the eff_γ should be close to 1 for all $\gamma = 1, \ldots, d$. Fig. 1 provides the kernel density estimators of the above empirical efficiencies computed over the 500 replications. Again, these plots show that the empirical distribution of eff_γ does rapidly converge to the point mass at 1 as n becomes larger. Finally, Fig. 2 shows typical examples of the SBS estimator with the confidence bands in (11) and the corresponding empirical relative efficiencies. The plots in these two figures illustrate graphically the summarized results on confidence band coverage and on the empirical relative efficiency.

Lastly, we provide the computing time of model (24) with dimension $d = 10$ from 100 replications on an ordinary PC with Intel(R) Quad CPU 2.4 GHz processor and 3.0 GB RAM. The average time run by R in seconds to generate one sample of size n and compute the SBS estimator and spline-backfitted spline (SPBK) estimator of [25] has been reported in Table 2. As

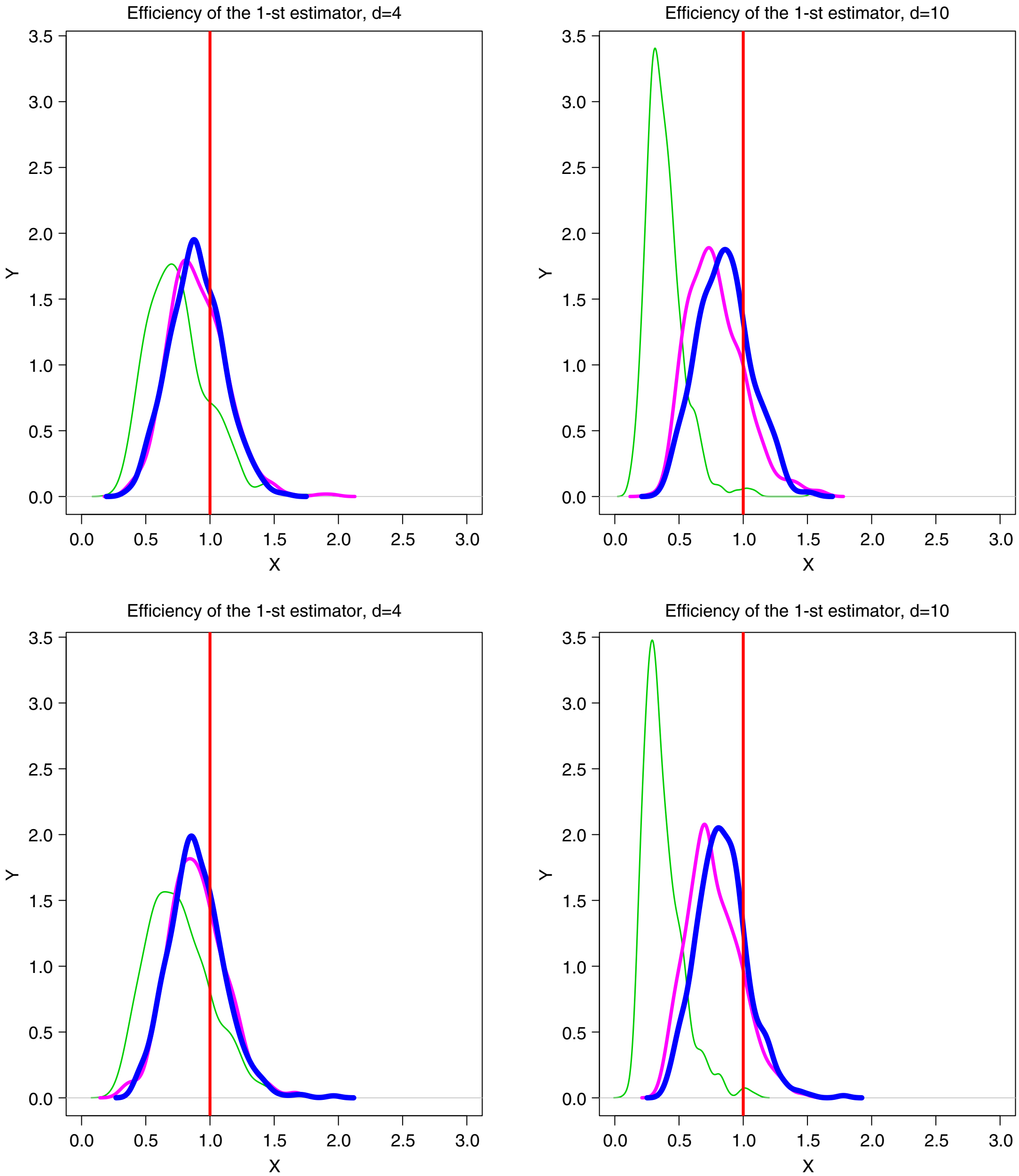

Fig. 1. Plots of the efficiency of SBS estimator $\hat{m}_{\alpha,\mathrm{SBS}}$ corresponding to oracle smoother $\tilde{m}_{\alpha,\mathrm{S}}$ for $\rho = 0$ (upper two panels), $\rho = 0.3$ (lower two panels) and $d = 4$, $d = 10$ of $m_\alpha\,(x_\alpha)$ in (25), for $\alpha = 1$ (thick curve for $n = 1000$, thin curve for $n = 500$, and solid curve for $n = 100$).

expected, the computing time of SBS is hundreds time faster than SPBK and this advantage widens with increasing sample size.

Acknowledgments

This work is part of the first author's dissertation under the supervision of the second author, and has been supported in part by NSF award DMS 0706518. The authors appreciate the helpful comments made by two referees, which have led to significant improvements of the paper.

Appendix

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

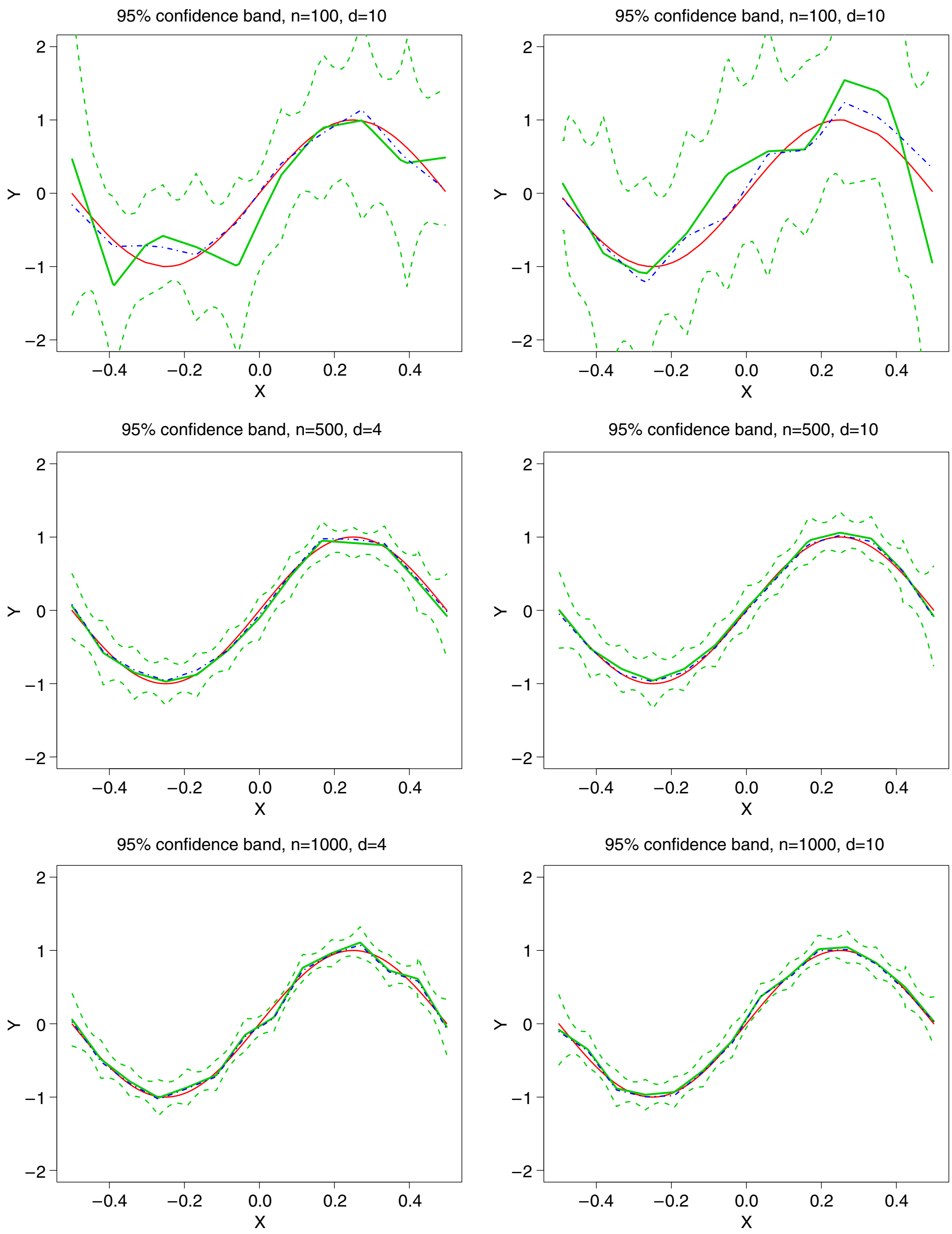

Fig. 2. For $\rho = 0$ (left plots) and $\rho = 0.3$ (right plots), plots of the oracle smoother $\tilde{m}_{\alpha,S}$ (dotted curve), SBS estimator $\hat{m}_{\alpha,SBS}$ (solid curve) and the 95% confidence bands (upper and lower dashed curves) of the function components $m_\alpha(x_\alpha)$ in (9) with $\alpha = 1$ (thin solid curve).

A.1. Propositions

Recall from Section 2 that $\|\boldsymbol{\Psi}_b\|_\infty = \sup_{0\le J\le N+1} |\{\Psi_b\}_{J=0}^{N+1}|$. In this section, we show that the bias term $\|\boldsymbol{\Psi}_b\|_\infty$ of (22) and the noise term $\boldsymbol{\Psi}_v$ given in (23) are uniformly of order $O_p\left(h^{1/2}n^{-2/5}(\log n)^{-1}\right)$.

Proposition A.1. *Under Assumptions* (A1) *to* (A2), *and* (A4) *to* (A5)

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

Lemma A.1. *Under Assumption* (A1), *there exists a function* $g(\mathbf{x}) = c + \sum_{\gamma=1}^{d} g_\gamma(x_\gamma) \in G^*$, *such that for* $\tilde{m}$ *defined in* (19),

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

Proof. By the result on page 149 of [4], there exists a constant $C_\infty > 0$ and spline functions $g_\gamma \in G^*$, such that $\left\|g_\gamma - m_\gamma\right\|_\infty \le C_\infty \left\|m'_\gamma\right\|_\infty H$, $\gamma = 1, 2, \dots, d$. Thus $\|g - m\|_\infty \le \sum_{\gamma=1}^{d} \left\|g_\gamma - m_\gamma\right\|_\infty \le C_\infty \sum_{\gamma=1}^{d} \left\|m'_\gamma\right\|_\infty H$ and $\left\|\tilde{m} - m\right\|_{2,n}^{*} \le \|g - m\|_{2,n}^{*} \le C_\infty \sum_{\gamma=1}^{d} \left\|m'_\gamma\right\|_\infty H$. Noting that $\left\|\tilde{m} - g\right\|_{2,n}^{*} \le \left\|\tilde{m} - m\right\|_{2,n}^{*} + \|g - m\|_{2,n}^{*} \le 2C_\infty \sum_{\gamma=1}^{d} \left\|m'_\gamma\right\|_\infty H$, one has

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

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

Proof of Proposition A.1. Clearly that $\|\boldsymbol{\Psi}_b\|_\infty \le R_1 + R_2 + R_3$, where

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

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

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

According to (A.1) $R_1 = O_p\left\{h^{1/2}\left(H + n^{-1/2}\right)\right\}$. For R_2, using the result on page 149 of [4], one has $R_2 \le C_\infty h^{1/2} H$. To deal with R_3, let $B_{J^*,\gamma}^{**}(\mathbf{x}_\gamma) = B_{J^*,\gamma}^{*}(\mathbf{x}_\gamma) - \left\langle 1, B_{J^*,\gamma}^{*}(\mathbf{X}_\gamma)\right\rangle_{2,n}^{*}$, for $1 \le J^* \le N^*$, $1 \le \gamma \le d$, then $\tilde{m}(\mathbf{x}) - g(\mathbf{x}) + \sum_{\gamma=1}^{d} \left\langle 1, g_\gamma(\mathbf{X}_\gamma)\right\rangle_{2,n}^{*} = \tilde{a}^* + \sum_{\gamma=1}^{d} \sum_{J^*=1}^{N^*} \tilde{a}_{J^*,\gamma}^{*} B_{J^*,\gamma}^{**}(\mathbf{x}_\gamma)$. Denote next $\omega_{J,J^*,-1}(\mathbf{X}_l) = \left\{\omega_{J,J^*,\gamma}(\mathbf{X}_l)\right\}_{\gamma=2}^{d}$, $\mu_{\omega_{J,J^*,-1}} = \left\{\mu_{\omega_{J,J^*,\gamma}}\right\}_{\gamma=2}^{d}$, where

$$\omega_{J,J^*,\gamma}(\mathbf{X}_l) = B_{J,1}(X_{l1}) B_{J^*,\gamma}^{*}(X_{l\gamma}) W_l^*, \qquad \mu_{\omega_{J,J^*,\gamma}} = E\omega_{J,J^*,\gamma}(\mathbf{X}_l). \tag{A.2}$$

Thus, $n^{-1} \sum_{i=1}^{n} B_{J,1}(X_{i1}) W_i^* \left\{\tilde{m}_{\_1}(X_{i\_1})^T - g_{\_1}(X_{i\_1})^T + E_n g_{\_1}\left(X_{i,\_1}\right)^T\right\} \mathbf{1}_{d-1}$ equals $n^{-1} \sum_{i=1}^{n} B_{J,1}(X_{i1}) W_i^* (\sum_{J^*=1}^{N^*} \tilde{a}_{J^*,\_1}^{*T} B_{J^*,\_1}^{**}(X_{i\_1}))$, bounded by

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

where $A_{n,1}$ is in (A.11). By Lemma A.11, $\sup_{1\le J\le N} \sup_{1\le J^*\le N^*} \left| n^{-1} \sum_{i=1}^{n} \omega_{J,J^*,\gamma}(\mathbf{X}_i) \right|$ is bounded by

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

Therefore, one has

$$\sup_{J=0}^{N+1} \left| \frac{1}{n} \sum_{i=1}^{n} B_{J,1}(X_{i1}) W_i^* \left\{ \tilde{m}_{\_1}(\mathbf{X}_{i\_1})^T - g_{\_1}(\mathbf{X}_{i\_1})^T + E_n g_{\_1}(\mathbf{X}_{\_1})^T \right\} \mathbf{1}_{d-1} \right|$$

$$\leq (d-1)\sup_{2\leq\gamma\leq d}\left\{N^*\sum_{J^*=1}^{N^*}\left(\tilde{a}_{J^*,\gamma}\right)^2\right\}^{1/2}\left\{O_p\left((Hh)^{1/2}\right)+O_p\left(\frac{\log n}{\sqrt{n}}\right)\right\}$$

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

Thus, by Lemma A.1

$$R_3 = O_p\left(h^{1/2}\left(n^{-1/2}+H\right)\right). \tag{A.3}$$

Combining (A.1) and (A.3), one establishes Proposition A.1. □

Define an auxiliary entity

$$\tilde{\varepsilon}^*_{\_1} = \sum_{J^*=1}^{N^*}\tilde{a}^T_{J^*,\_1}B^*_{J^*,\_1}\left(x_{\_1}\right), \tag{A.4}$$

where $\tilde{a}_{J^*,\_1} = \left\{\tilde{a}_{J^*,\gamma}\right\}_{\gamma=2}^{d}$ and $\tilde{a}_{J^*,\gamma}$ is given in (21). Definitions (17) implies that $\tilde{\varepsilon}_{\_1}\left(x_{\_1}\right)$ defined in (19) is the empirical centering of $\tilde{\varepsilon}^*_{\_1}\left(\mathbf{x}_{\_1}\right)$, i.e.

$$\tilde{\varepsilon}_{\_1}\left(\mathbf{x}_{\_1}\right) \equiv \tilde{\varepsilon}^*_{\_1}\left(x_{\_1}\right) - n^{-1}\sum_{i=1}^{n}\tilde{\varepsilon}^*_{\_1}\left(\mathbf{X}_{i\_1}\right)W_i^*. \tag{A.5}$$

Proposition A.2. *Under Assumptions* (A2) *to* (A5), *one has*

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

According to (A.5), we can write $\Psi_v = \Psi_v^{(2)} - \Psi_v^{(1)}$, in which

$$\left\{\Psi_v^{(1)}\right\}_{J=0}^{N+1} = \left\{n^{-2}\sum_{i,i'=1}^{n}B_{J,1}(X_{i1})W_{i1}W^*_{i'}\tilde{\varepsilon}^*_{\_1}\left(\mathbf{X}_{i'\_1}\right)^T\mathbf{1}_{d-1}\right\}_{J=0}^{N+1}, \tag{A.6}$$

$$\left\{\Psi_v^{(2)}\right\}_{J=0}^{N+1} = \frac{\mathbf{B}^T}{n}\mathbf{W}^*\tilde{\boldsymbol{\varepsilon}}^*_{\_1}\left(\mathbf{X}_{\_1}\right)^T\mathbf{1}_{d-1} \tag{A.7}$$

where $\tilde{\boldsymbol{\varepsilon}}^*_{\_1}\left(\mathbf{X}_{\_1}\right)$ is given in (A.4). By (A.2), (21) and (A.4), we have

$$\left\|\mathbf{\Psi}_v^{(2)}\right\|_\infty = \sup_{0\leq J\leq N+1}\left|n^{-1}\sum_{l=1}^{n}\sum_{J^*=1}^{N^*}\tilde{a}^T_{J^*,\_1}\omega_{J,J^*,\_1}\left(\mathbf{X}_{\_1}\right)\right|. \tag{A.8}$$

Proposition A.2 follows from Lemmas A.2 and A.3.

Lemma A.2. *Under Assumptions* (A2) *to* (A5), $\Psi_v^{(1)}$ *in* (A.6) *satisfies*

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

Proof. Based on (A.4), $\left\|n^{-1}\sum_{i=1}^{n}\tilde{\varepsilon}^*_{\_1}\left(\mathbf{X}_{i\_1}\right)^T\mathbf{1}_{d-1}W_i^*\right\|_\infty$ is bounded by

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

Lemma A.13 implies that $\sum_{J^*=1}^{N^*}\left|\tilde{a}^*_{J^*,\gamma}\right| \leq \left\{N^*\left(\tilde{\mathbf{a}}^{*T}\tilde{\mathbf{a}}^*\right)\right\}^{1/2} = O_p\left(N^*n^{-1/2}\log n\right)$. Further, by (A.11), $\sup_{2\leq\gamma\leq d}\sup_{1\leq J^*\leq N^*}|n^{-1}\sum_{i=1}^{n}B^*_{J^*,\gamma}\left(X_{i\gamma}\right)W_i^*| \leq A_{n,1} = O_p\left(n^{-1/2}\log n\right)$, so

$$\left\|\frac{1}{n}\sum_{i=1}^{n}\tilde{\varepsilon}^*_{\_1}\left(X_{i\_1}\right)^T\mathbf{1}_{d-1}W_i^*\right\|_\infty = O_p\left\{N^*\left(\log n\right)^2/n\right\}. \tag{A.9}$$

By $\sup_{0\leq J\leq N+1}\left|n^{-1}\sum_{i=1}^{n}B_{J,1}(X_{i1})W_i^*\right| = \sup_{0\leq J\leq N+1}\left(\left\langle 1, B_{J,1}\right\rangle_{2,n} - \left\langle 1, B_{J,1}\right\rangle_2\right) + \sup_{0\leq J\leq N+1}\left\langle 1, B_{J,1}\right\rangle_2 = O_p(\log n/\sqrt{n}) + O_p\left(h^{1/2}\right) = O_p\left(h^{1/2}\right)$. Thus with (A.9) the lemma follows immediately. □

Lemma A.3. *Under Assumptions* (A2) *to* (A5), $\Psi_v^{(2)}$ *satisfies* $\left\|\boldsymbol{\Psi}_v^{(2)}\right\|_\infty = O_p\left(Hh^{1/2}\right)$.

Lemma A.3 follows from Lemmas A.14 and A.15.

A.2. Preliminaries

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

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

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

where $\alpha(\cdot)$ *is the* α*-mixing coefficient defined in* (10) *and* $a_1 = 2\frac{n}{q} + 2\left(1 + \frac{\varepsilon^2}{25m_2^2+5c\varepsilon}\right)$, $a_2(k) = 11n\left(1 + \frac{5m_k^{2k/(2k+1)}}{\varepsilon}\right)$, *with* $m_r = \max_{1\le i\le n}\|\xi_i\|_r$, $r \ge 2$.

Lemma A.5. *Under Assumptions* (A4) *and* (A5), *one has:*

(i) $\left\|b_{J^*,\gamma}^*\right\|_2^2 \sim H$, *where* $b_{J^*,\gamma}^*$ *is given in* (14).

(ii) *for any* $\gamma = 1, 2, \ldots, d$, $E\left\{B_{J^*,\gamma}^*\left(X_{i\gamma}\right) B_{J^{*\prime},\gamma}^*\left(X_{i\gamma}\right) W_i^*\right\} \sim 1$, *for* $\left|J^{*\prime} - J^*\right| \le 1$, *and* $E\left\{B_{J,\gamma}\left(X_{i\gamma}\right) B_{J',\gamma}\left(X_{i\gamma}\right) W_{i\gamma}\right\} \sim 1$, *for* $\left|J' - J\right| \le 1$. *In addition,* $E\left|B_{J^*,\gamma}^*\left(X_{i\gamma}\right) B_{J^{*\prime},\gamma}^*\left(X_{i\gamma}\right) W_i^*\right|^k \sim H^{1-k}$, $E\left|B_{J,\gamma}\left(X_{i\gamma}\right) B_{J',\gamma}\left(X_{i\gamma}\right) W_{i\gamma}\right|^k \sim h^{1-k}$, $k \ge 1$, *where* $B_{J^*,\gamma}^*$ *and* $B_{J,\gamma}$ *are defined in* (13) *and* (12).

Lemma A.6. *Under Assumptions* (A4) *and* (A5), *there exist constants* $C_0 > c_0 > 0$ *such that for any* $\mathbf{a}^* = \left(a_0^*, a_{1,1}^*, \ldots, a_{N^*,1}^*, a_{1,2}^*, \ldots, a_{N^*,2}^*, \ldots, a_{1,d}^*, \ldots, a_{N^*,d}^*\right)$,

$$c_0\left(a_0^{*2} + \sum_{J^*,\gamma} a_{J^*,\gamma}^{*2}\right) \le \left\|a_0^* + \sum_{J^*,\gamma} a_{J^*,\gamma}^* B_{J^*,\gamma}^*\right\|_2^{*2} \le C_0\left(a_0^{*2} + \sum_{J^*,\gamma} a_{J^*,\gamma}^{*2}\right). \tag{A.10}$$

Lemma A.7. *Under Assumptions* (A2), (A4) *and* (A6), *one has*

$$A_{n,1} = \sup_{1\le J^*\le N^*,\gamma}\left|\left\langle 1, B_{J^*,\gamma}^*\right\rangle_{2,n}^* - \left\langle 1, B_{J^*,\gamma}^*\right\rangle_2^*\right| = O_p\left(n^{-1/2}\log n\right), \tag{A.11}$$

$$A_{n,2} = \sup_{1\le J^*,J'^*\le N^*,\gamma}\left|\left\langle B_{J^*,\gamma}^*, B_{J'^*,\gamma}^*\right\rangle_{2,n}^* - \left\langle B_{J^*,\gamma}^*, B_{J'^*,\gamma}^*\right\rangle_2^*\right| = O_p\left(n^{-1/2}H^{-1/2}\log n\right), \tag{A.12}$$

$$A_{n,3} = \sup_{1\le J^*,J'^*\le N^*,\gamma\ne\gamma'}\left|\left\langle B_{J^*,\gamma}^*, B_{J'^*,\gamma'}^*\right\rangle_{2,n}^* - \left\langle B_{J^*,\gamma}^*, B_{J'^*,\gamma'}^*\right\rangle_2^*\right| = O_p\left(n^{-1/2}\log n\right). \tag{A.13}$$

Lemma A.8. *Under Assumptions* (A2), (A4) *and* (A6), *one has*

$$A_n = \sup_{g_1,g_2\in G^*}\frac{\left|\langle g_1, g_2\rangle_{2,n}^* - \langle g_1, g_2\rangle_2^*\right|}{\|g_1\|_2^*\,\|g_2\|_2^*} = O_p\left(\frac{\log n}{n^{1/2}H^{1/2}}\right) = o_p(1). \tag{A.14}$$

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

$$\mathbf{V} = \begin{pmatrix} 1 & \mathbf{0}_{dN^*}^T \\ \mathbf{0}_{dN^*} & \left\langle B_{J^*,\gamma}^*, B_{J^{*\prime},\gamma'}^*\right\rangle_2^* \end{pmatrix}_{\substack{1\le\gamma,\gamma'\le d,\\ 1\le J,J'\le N^*}}. \tag{A.15}$$

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

$$\mathbf{S}=\mathbf{V}^{-1}=\begin{pmatrix}1 & \mathbf{0}_N^T & \mathbf{0}_N^T\\ \mathbf{0}_N & \mathbf{V}_{11} & \mathbf{V}_{12}\\ \mathbf{0}_N & \mathbf{V}_{21} & \mathbf{V}_{22}\end{pmatrix}^{-1}=\begin{pmatrix}1 & \mathbf{0}_N^T & \mathbf{0}_N^T\\ \mathbf{0}_N & \mathbf{S}_{11} & \mathbf{S}_{12}\\ \mathbf{0}_N & \mathbf{S}_{21} & \mathbf{S}_{22}\end{pmatrix}. \quad (A.16)$$

Lemma A.9. *Under Assumptions* (A4) *and* (A5), *for* $\mathbf{V}$, $\mathbf{S}$ *defined in* (A.15), (A.16), *there exist constants* $C_V > c_V > 0$ *and* $C_S > c_S > 0$ *such that* $c_V\mathbf{I}_{dN^*+1} \le \mathbf{V} \le C_V\mathbf{I}_{dN^*+1}$, $c_S\mathbf{I}_{dN^*+1} \le \mathbf{S} \le C_S\mathbf{I}_{dN^*+1}$.

We refer the proofs of Lemma A.5 to A.9 to Lemmas A.2, A.4, A.7, A.8 and A.9 in [25].

Lemma A.10. *Under Assumptions* (A2) *and* (A3), *there exist constants* $c(f), C(f) > 0$ *independent of* n, *such that as* $n \to \infty$, *with probability approaching* 1,

$$c(f)\,|\zeta| \le \left|\left(\frac{1}{n}\mathbf{B}^T\mathbf{W}\mathbf{B}\right)^{-1}\zeta\right| \le C(f)\,|\zeta|, \quad (A.17)$$

$$c(f)\,\|\zeta\|^2 \le \zeta^T\left(\frac{1}{n}\mathbf{B}^T\mathbf{W}\mathbf{B}\right)^{-1}\zeta \le C(f)\,\|\zeta\|^2, \quad \forall\zeta\in R^{N+2}. \quad (A.18)$$

The lemma and its proof is based on Lemma B.2 of Wang and Yang [27].

Lemma A.11. *Under Assumptions* (A4) *to* (A5), *for* $\mu_{\omega_{J,J^*,\_1}}$ *given in* (A.2)

$$\sup_{0\le J\le N+1}\sup_{1\le J^*\le N^*}\left|\mu_{\omega_{J,J^*,\_1}}\right| = O\left\{(hH)^{1/2}\right\}.$$

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

$$\begin{aligned}\left|E\left\{B_{J,1}(X_{l1})W_l^*B_{J^*,\gamma}^*\left(X_{l\gamma}\right)\right\}\right| &\le \int_0^1\cdots\int_0^1\left|B_{J,1}(u_1)B_{J^*,\gamma}^*\left(u_\gamma\right)\right|f\left(u_1,\ldots,u_d\right)\mathrm{d}u_1\cdots\mathrm{d}u_d\\ &\le C_f\int_0^1\cdots\int_0^1\left|B_{J,1}(u_1)B_{J^*,\gamma}^*\left(u_\gamma\right)\right|\mathrm{d}u_1\ldots\mathrm{d}u_d\\ &= C_f\left(\left\|b_{J,1}\right\|_2\left\|b_{J^*,\gamma}^*\right\|_2\right)^{-1}\int_0^1\int_0^1\left|b_{J,1}(u1)\left(b_{J^*,\gamma}^*\left(u_\gamma\right)\right)\right|\mathrm{d}u_1\mathrm{d}u_\gamma\\ &= \left(\left\|b_{J,1}\right\|_2\left\|b_{J^*,\gamma}^*\right\|_2\right)^{-1}\left\{\int_0^1\int_0^1 b_{J,1}(u_1)I_{J^*+1,\gamma}\left(u_\gamma\right)\mathrm{d}u_1\mathrm{d}u_\gamma\right.\\ &\quad\left.+\left(\left\|b_{J,1}\right\|_2\left\|b_{J^*,\gamma}^*\right\|_2\right)^{-1}\frac{c_{J^*+1,\gamma}}{c_{J^*,\gamma}}\int_0^1\int_0^1 b_{J,1}(u_1)I_{J^*,\gamma}\left(u_\gamma\right)\mathrm{d}u_1\mathrm{d}u_\gamma\right\},\end{aligned}$$

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

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

and the proof of the lemma is then completed by (i) of Lemma A.5. □

Lemma A.12. *Under Assumptions* (A2), (A4) *and* (A5), *one has*

$$\sup_{0\le J\le N+1}\sup_{1\le J^*\le N^*}\left\|n^{-1}\sum_{l=1}^n\left\{\omega_{J,J^*\_1}\left(\mathbf{X}_l\right)-\mu_{\omega_{J,J^*\_1}}\right\}\right\|_\infty = O_p\left(\log n/\sqrt{n}\right), \quad (A.19)$$

$$\sup_{0\le J\le N+1}\sup_{1\le J^*\le N^*}\left\|n^{-1}\sum_{l=1}^n\omega_{J,J^*\_1}\left(\mathbf{X}_l\right)\right\|_\infty = O_p\left((hH)^{1/2}\right), \quad (A.20)$$

where $\omega_{J,J^*,\_1}\left(\mathbf{X}_l\right)$ *and* $\mu_{\omega_{J,J^*,\_1}}$ *are given in* (A.2).

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

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

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

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

It is clear that $E\left|B_{J,1}(X_{l1})W_l^* B^*_{J^*,\gamma}\left(X_{l\gamma}\right)\right|^r \sim h^{(1-r/2)}H^{1-r/2}$. According to Lemma A.11, one has $\left|E\omega_{J,J^*,\gamma}(\mathbf{X}_l)\right|^r = \left|EB_{J,1}(X_{l1})W_l^* B^*_{J^*,\gamma}\left(X_{l\gamma}\right)\right|^r \sim (hH)^{r/2}$, thus $E\left|\omega_{J,J^*,\gamma}(\mathbf{X}_l)\right|^r \gg \left|\mu_{\omega_{J,J^*,\gamma}}\right|^r$. In addition, for any J and J^*,

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

so there exists $c_* = ch^{-1/2}H^{-1/2}$ such that $E\left|\omega^*_{J,J^*,\gamma}(\mathbf{X}_l)\right|^r \leq c_*^{r-2} r!E\left|\omega^*_{J,J^*,\gamma}(\mathbf{X}_l)\right|^2$, which implies that $\left\{\omega^*_{J,J^*,\gamma}(\mathbf{X}_l)\right\}_{l=1}^n$ satisfies the Cramér's condition. By the Bernstein's inequality, for $r = 3$

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

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

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

Since $5c_*\rho_n = o(1)$, by taking q such that $\left[\frac{n}{q+1}\right] \geqslant c_0 \log n$, $q \geqslant c_1 n/\log n$ for constants c_0, c_1, one has $a_1 = O(n/q) = O(\log n)$, $a_2(3) = o\left(n^2\right)$. Assumption (A2) yields that $\alpha\left(\left[\frac{n}{q+1}\right]\right)^{6/7} \leq Cn^{-6\lambda_0 c_0/7}$. Thus, for n large enough,

$$P\left\{\frac{1}{n}\left|\sum_{l=1}^n \omega^*_{J,J^*,\gamma}(\mathbf{X}_l)\right| > \frac{\rho\log n}{\sqrt{nh}}\right\} \leq cn^{-c_2\rho^2}\log n + Cn^{2-6\lambda_0 c_0/7}. \tag{A.21}$$

By (A.21), there exists large enough value $\rho > 0$ such that for any J^*,

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

which implies that

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

Thus, Borel–Cantelli Lemma entails that

$$\sup_{0\le J\le N+1}\sup_{1\le J^*\le N^*}\left|n^{-1}\sum_{l=1}^{n}\omega^*_{J,J^*,\gamma}(\mathbf{X}_l)\right|=O_p\left(\log n/\sqrt{nh}\right).\tag{A.22}$$

Then, $\sup_{0\le J\le N+1}\sup_{1\le J^*\le N^*}\left\|n^{-1}\sum_{l=1}^{n}\omega^*_{J,J^*,\_1}(\mathbf{X}_l)\right\|_\infty=O_p\left(\log n/\sqrt{nh}\right)$. As a result of Lemma A.11 and (A.19), (A.20) holds. □

The next lemma provides the size of $\tilde{\mathbf{a}}^{*T}\tilde{\mathbf{a}}^*$, where $\tilde{\mathbf{a}}^*$ is defined by (21).

Lemma A.13. *Under Assumptions* (A2) *to* (A5), $\tilde{\mathbf{a}}^*$ *satisfies*

$$\tilde{\mathbf{a}}^{*T}\tilde{\mathbf{a}}^*=\tilde{a}_0^{*2}+\sum_{J^*=1}^{N^*}\sum_{\gamma=1}^{d}\tilde{a}_{J,\gamma}^{*2}=O_p\left\{N^*(\log n)^2/n\right\}.\tag{A.23}$$

Proof. According to (20) and (21), $\tilde{\mathbf{a}}^{*T}\mathbf{B}^{*T}\mathbf{W}^*\mathbf{B}^*\tilde{\mathbf{a}}^*=\tilde{\mathbf{a}}^{*T}\left(\mathbf{B}^{*T}\mathbf{W}^*\mathbf{E}\right)$. Thus

$$\left\|\mathbf{W}^*\mathbf{B}^*\tilde{\mathbf{a}}^*\right\|_{2,n}^{*2}=\tilde{\mathbf{a}}^{*T}\begin{pmatrix}\mathbf{1} & \\ & \left\langle B^*_{J^*,\gamma},B_{J^{*\prime},\gamma'}\right\rangle^*_{2,n}\end{pmatrix}\tilde{\mathbf{a}}^*=\tilde{\mathbf{a}}^{*T}\left(n^{-1}\mathbf{B}^{*T}\mathbf{W}^*\mathbf{E}\right).\tag{A.24}$$

By (A.14), $\left\|\mathbf{B}^*\tilde{\mathbf{a}}^*\right\|_{2,n}^{*2}$ is bounded below in probability by $(1-A_n)\left\|\mathbf{B}^*\tilde{\mathbf{a}}^*\right\|_2^{*2}$. According to (A.10), one has

$$\left\|\mathbf{W}^*\mathbf{B}^*\tilde{\mathbf{a}}^*\right\|_2^{*2}=\left\|a_0^*+\sum_{J^*,\gamma}a^*_{J^*,\gamma}B^*_{J^*,\gamma}\right\|_2^{*2}\ge c_0\left(\tilde{a}_0^{*2}+\sum_{J^*,\gamma}\tilde{a}^{*2}_{J^*,\gamma}\right).\tag{A.25}$$

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

$$\sqrt{\tilde{a}_0^{*2}+\sum_{J^*,\gamma}\tilde{a}^{*2}_{J,\gamma}}\left[\left\{\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i\right\}^2+\sum_{J^*,\gamma}\left\{\frac{1}{n}\sum_{i=1}^{n}B^*_{J^*,\gamma}\left(X_{i\gamma}\right)W_i^*\varepsilon_i\right\}^2\right]^{1/2}.\tag{A.26}$$

Combining (A.24)–(A.26)), the squared norm $\tilde{\mathbf{a}}^{*T}\tilde{\mathbf{a}}^*$ is bounded by $c_0^{-2}(1-A_n)^{-2}[\{\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i\}^2+\sum_{J^*,\gamma}\{\frac{1}{n}\sum_{i=1}^{n}B^*_{J^*,\gamma}(X_{i\gamma})W_i^*\varepsilon_i\}^2]$.

Truncating ε as in Lemma A.15, Bernstein inequality entails that $\left|n^{-1}\sum_{i=1}^{n}\varepsilon_i\right|+\max_{1\le J^*\le N^*,\gamma=1,\ldots,d}|n^{-1}\sum_{i=1}^{n}B_{J^*,\gamma}(X_{i\gamma})W_i^*\varepsilon_i|=O_p\left(\log n/\sqrt{n}\right)$. Thus (A.23) holds since A_n is of order $o_p(1)$ according to A.8. □

A.3. Proof of Lemma A.3

We denote

$$\mathbf{V}^*=\begin{pmatrix}0 & \mathbf{0}^T_{dN^*}\\ \mathbf{0}_{dN^*} & \left\langle B^*_{J^*,\gamma},B^*_{J^{*\prime},\gamma'}\right\rangle^*_{2,n}-\left\langle B^*_{J^*,\gamma},B^*_{J^{*\prime},\gamma'}\right\rangle^*_{2}\end{pmatrix}_{\substack{1\le\gamma,\gamma'\le d,\\ 1\le J^*,J^{*\prime}\le N^*}},$$

then $\tilde{\mathbf{a}}^*$ in (21) can be rewritten as

$$\tilde{\mathbf{a}}^*=\left(\frac{1}{n}\mathbf{B}^{*T}\mathbf{W}^*\mathbf{B}^*\right)^{-1}\left(\frac{1}{n}\mathbf{B}^{*T}\mathbf{W}^*\mathbf{E}\right)=\left(\mathbf{V}+\mathbf{V}^*\right)^{-1}\left(\frac{1}{n}\mathbf{B}^{*T}\mathbf{W}^*\mathbf{E}\right).\tag{A.27}$$

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

$$\hat{\mathbf{a}} = \mathbf{V}^{-1}\left(n^{-1}\mathbf{B}^{*T}\mathbf{W}^{*}\mathbf{E}\right) = \mathbf{S}\left(n^{-1}\mathbf{B}^{*T}\mathbf{W}^{*}\mathbf{E}\right), \tag{A.28}$$

and define a theoretical version of $\Psi_v^{(2)}$ in (A.8) as

$$\hat{\Psi}_v^{(2)} = n^{-1}\sum_{i=1}^{n}\sum_{J^*=1}^{N^*}\hat{a}_{J^*,\_1}^{*T}\omega_{J,J^*,\_1}(\mathbf{X}_i). \tag{A.29}$$

Lemma A.14. *Under Assumptions* (A2) *to* (A5),

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

Proof. By (A.27) and (A.28), one has $\mathbf{V}\hat{\mathbf{a}}^* = (\mathbf{V}+\mathbf{V}^*)\tilde{\mathbf{a}}^*$, which implies that $\mathbf{V}^*\tilde{\mathbf{a}}^* = \mathbf{V}\left(\hat{\mathbf{a}}^* - \tilde{\mathbf{a}}^*\right)$. Using (A.12) and (A.13), one obtains that

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

According to Lemma A.13, $\left\|\tilde{\mathbf{a}}^*\right\|_2^* = O_p\left(n^{-1/2}N^{*1/2}\log n\right)$, so one has $\left\|\mathbf{V}\left(\hat{\mathbf{a}}^* - \tilde{\mathbf{a}}^*\right)\right\|_2^* \le O_p\left\{(\log n)^2 n^{-1}N^{*3/2}\right\}$. By Lemma A.9, $\left\|\left(\hat{\mathbf{a}}^* - \tilde{\mathbf{a}}^*\right)\right\|_2^* = O_p\left\{(\log n)^2 n^{-1}N^{*3/2}\right\}$. Lemma A.13 implies

$$\left\|\hat{\mathbf{a}}^*\right\|_2^* \le \left\|\left(\hat{\mathbf{a}}^* - \tilde{\mathbf{a}}^*\right)\right\|_2^* + \left\|\tilde{\mathbf{a}}^*\right\|_2^* = O_p\left(\log n\sqrt{N^*/n}\right). \tag{A.30}$$

Additionally, $\left\|\mathbf{\Psi}_v^{(2)} - \hat{\mathbf{\Psi}}_v^{(2)}\right\|_\infty = \sup_{0\le J\le N+1}\left|\sum_{J^*=1}^{N^*}\left(\tilde{a}_{J^*,\_1}^* - \hat{a}_{J^*,\_1}^*\right)\frac{1}{n}\sum_{l=1}^{n}\omega_{J,J^*,\_1}(\mathbf{X}_l)\right|$. So $\left\|\mathbf{\Psi}_v^{(2)} - \hat{\mathbf{\Psi}}_v^{(2)}\right\|_\infty \le \sqrt{N^*}O_p\left\{\frac{(\log n)^2}{nH}\right\}O_p\left((hH)^{1/2}\right) = O_p\left\{\frac{h^{1/2}(\log n)^2}{nH}\right\}$. □

Lemma A.15. *Under Assumptions* (A2) *to* (A5), *for* $\hat{\Psi}_v^{(2)}$ *in* (A.29), *one has*

$$\left\|\hat{\mathbf{\Psi}}_v^{(2)}\right\|_\infty = \sup_{0\le J\le N+1}\left|n^{-1}\sum_{i=1}^{n}\left(B_{J,1}(X_{i1})\sum_{J^*=1}^{N^*}\hat{a}_{J^*,\_1}^{*T}B_{J^*,\_1}^{*}(X_{i\_1})W_i^*\right)\right| = O_p\left(h^{1/2}H\right).$$

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

$$Q_1 = \sup_{0\le J\le N+1}\left|\sum_{J^*=1}^{N^*}\hat{a}_{J^*,\_1}^{*T}\mu_{\omega_{J,J^*,\_1}}\right|,$$
$$Q_2 = \sup_{0\le J\le N+1}\left|\sum_{J^*=1}^{N^*}\hat{a}_{J^*,\_1}^{*T}n^{-1}\sum_{i=1}^{n}\left\{\omega_{J,J^*,\_1}(\mathbf{X}_i) - \mu_{\omega_{J,J^*,\_1}}\right\}\right|. \tag{A.31}$$

By Cauchy–Schwarz inequality, (A.30), Lemma A.12, and Assumptions (A5),

$$Q_2 = O_p\left(\log n\sqrt{N^*/n}\right)\sqrt{N^*}O_p\left(\frac{\log n}{\sqrt{n}}\right) = O_p\left\{\frac{(\log n)^3}{\sqrt{n}}\right\}. \tag{A.32}$$

Define next

$$F_{1,\gamma} = \sup_{0\le J\le N+1}\left|n^{-1}\sum_{1\le i\le n}\sum_{1\le J^*,J^{*\prime}\le N^*}\mu_{\omega_{J,J^*,\gamma}}s_{J^*+N^*,J^{*\prime}+1}B_{J^{*\prime},1}(X_{i1})W_{i1}\varepsilon_i\right|,$$
$$F_{2,\gamma} = \sup_{0\le J\le N+1}\left|n^{-1}\sum_{1\le i\le n}\sum_{1\le J^*,J^{*\prime}\le N^*}\mu_{\omega_{J,J^*,\gamma}}s_{J^*+N^*,J^{*\prime}+N^*}B_{J^{*\prime},\gamma}(X_{i\gamma})W_{i\gamma}\varepsilon_i\right|,$$

then it is clear that $Q_1 \le (d-1)\left(\sup_{2\le\gamma\le d}F_{1,\gamma} + \sup_{2\le\gamma\le d}F_{2,\gamma}\right)$. Next we will show that $F_{1,\gamma} = O_p\left(h^{1/2}H\right)$. Let $D_n = n^{\theta_0}\left(\frac{1}{2+\delta} < \theta_0 < \frac{2}{5}\right)$, where δ is the same as in Assumption (A3). Define

$$\varepsilon_{i,D}^- = \varepsilon_i I(|\varepsilon_i| \le D_n), \qquad \varepsilon_{i,D}^+ = \varepsilon_i I(|\varepsilon_i| > D_n), \qquad \varepsilon_{i,D}^* = \varepsilon_{i,D}^- - E\left(\varepsilon_{i,D}^-|\mathbf{X}_i\right),$$
$$U_{i,\gamma} = \boldsymbol{\mu}_{\omega_{J,\gamma}}^T\mathbf{S}_{21}\left\{B_{1,1}^*(X_{i1}), \ldots, B_{1,N^*}^*(X_{i1})\right\}^T W_{i1}\varepsilon_{i,D}^*.$$

Denote the truncation of $F_{1,\gamma}$ as $F^D_{1,\gamma} = \left|n^{-1}\sum_{i=1}^n U_{i,\gamma}\right|$. Next we show that $\left|F_{1,\gamma} - F^D_{1,\gamma}\right| = O_p\left(h^{1/2}H\right)$. Note that $\left|F_{1,\gamma} - F^D_{1,\gamma}\right| \le \Lambda_{1,\gamma} + \Lambda_{2,\gamma}$, where

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

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

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

$$\Lambda_{1,\gamma} = \left|\boldsymbol{\mu}^T_{\omega_{J,\gamma}}\mathbf{S}_{21}\left\{n^{-1}\sum_{i=1}^n B^*_{J^{*'},1}(X_{i1})W_{i1}E\left(\varepsilon^-_{i,D}|\mathbf{X}_i\right)\right\}^{N^*}_{J^{*'}=1}\right|$$
$$\le C_S\left\{\sum_{J=1}^{N^*}\mu^2_{\omega_{JJ^*,\gamma}}\sum_{J=1}^{N^*}\left\{\frac{1}{n}\sum_{i=1}^n B^*_{J^*,1}(X_{i1})W_{i1}E\left(\varepsilon^-_{i,D}|\mathbf{X}_i\right)\right\}^2\right\}^{1/2}.$$

By Assumptions (A3), $\left|E\left(\varepsilon^-_{i,D}|\mathbf{X}_i\right)\right| = \left|E\left(\varepsilon^+_{i,D}|\mathbf{X}_i\right)\right| \le M_\delta D_n^{-(1+\delta)}$ and Lemma A.4 entails that $\sup_{J,\gamma}\left|\frac{1}{n}\sum_{i=1}^n B_{J,1}(X_{i1})W_{i1}\right| = O_p\left(\log n/\sqrt{n}\right)$. Therefore

$$\Lambda_{1,\gamma} \le M_\delta D_n^{-(1+\delta)}\sup_{0\le J\le N+1}\left[\sum_{J^*=1}^{N^*}\mu^2_{\omega_{J,J^*,\gamma}}\sum_{J^*=1}^{N^*}\left\{\frac{1}{n}\sum_{i=1}^n B^*_{J^*,1}(X_{i1})W_{i1}\right\}^2\right]^{1/2}$$
$$= O_p\left\{N^*D_n^{-(1+\delta)}h^{1/2}\log^2 n/n\right\} = O_p\left(h^{1/2}H\right),$$

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

$$\sum_{n=1}^\infty P\left(|\varepsilon_n|\ge D_n\right) \le \sum_{n=1}^\infty \frac{E|\varepsilon_n|^{2+\delta}}{D_n^{2+\delta}} = \sum_{n=1}^\infty \frac{E\left(E|\varepsilon_n|^{2+\delta}|\mathbf{X}_n\right)}{D_n^{2+\delta}} \le \sum_{n=1}^\infty \frac{M_\delta}{D_n^{2+\delta}} < \infty,$$

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

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

for large n. Therefore, one has $\left|F_{1,\gamma} - F^D_{1,\gamma}\right| \le \Lambda_{1,\gamma} + \Lambda_{2,\gamma} = O_p\left(h^{1/2}H\right)$. Next we will show that $F^D_{1,\gamma} = O_p\left(h^{1/2}H\right)$. Note that the variance of $U_{i,\gamma}$ is

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

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

Simple calculation leads to that $E\left|U_{i,\gamma}\right|^r \le \left\{c_0\kappa_\gamma D_n H^{-1/2}\right\}^{r-2} r!E\left|U_{i,\gamma}\right|^2 < +\infty$, for $r\ge 3$, so $\left\{U_{i,\gamma}\right\}_{i=1}^n$ satisfies the Cramér's condition with Cramér's constant $c_* = c_0\kappa_\gamma D_n H^{-1/2}$. Hence by the Bernstein's inequality,

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

where $m_2^2 \sim \{\kappa_\alpha\}^2 V_{\varepsilon,D}$, $m_3 \le \left\{c\{\kappa_\alpha\}^3 H^{-1/2}D_n V_{\varepsilon,D}\right\}^{1/3}$, $\rho_n = \rho h^{1/2}H$, $a_1 = 2\frac{n}{q} + 2\left(1+\frac{\rho_n^2}{25m_2^2+5c_*\rho_n}\right)$, $a_2(3) = 11n\left(1+\frac{5m_3^{6/7}}{\rho_n}\right)$. Similar arguments as in Lemma A.12 yield that as $n\to\infty$, $\frac{q\rho_n^2}{25m_2^2+5c_*\rho_n} \sim \frac{q\rho_n}{c_*} = \frac{\rho n^{2/5}}{c_0(\log n)^{5/2}D_n} \to +\infty$.

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

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

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

$$Q_1 \leq (d-1)\left(\sup_{2\leq\gamma\leq d} F_{1,\gamma} + \sup_{2\leq\gamma\leq d} F_{2,\gamma}\right) = O_p\left(h^{1/2}H\right), \tag{A.33}$$

and one has $Q_1 = O_p\left(h^{1/2}H\right)$. The result follows from (A.31) and (A.32). □

References

[1] D. Bosq, Nonparametric Statistics for Stochastic Processes, Springer-Verlag, New York, 1998.
[2] R. Chen, R.S. Tsay, Nonlinear additive ARX models, J. Amer. Statist. Assoc. 88 (1993) 956–967.
[3] G. Claeskens, I. Van Keilegom, Bootstrap confidence bands for regression curves and their derivatives, Ann. Statist. 31 (2003) 1852–1884.
[4] C. de Boor, A Practical Guide to Splines, Springer-Verlag, New York, 2001.
[5] J. Fan, I. Gijbels, Local Polynomial Modelling and its Applications, Chapman and Hall, London, 1996.
[6] Y. Fan, Q. Li, A kernel-based method for estimating additive partially linear models, Statist. Sinica 13 (2003) 739–762.
[7] T.J. Hastie, R.J. Tibshirani, Generalized Additive Models, Chapman and Hall, London, 1990.
[8] J. Horowitz, E. Mammen, Nonparametric estimation of an additive model with a link function, Ann. Statist. 32 (2004) 2412–2443.
[9] X. Huang, L. Wang, L. Yang, A.N. Kravchenko, Management practice effects on relationships of grain yields with topography and precipitation, Agronomy J. 100 (2008) 1463–1471.
[10] J.Z. Huang, L. Yang, Identification of nonlinear additive autoregression models, J. R. Stat. Soc. Ser. B Stat. Methodol. 66 (2004) 463–477.
[11] G. Kauermann, T. Krivobokova, L. Fahrmeir, Some asymptotic results on generalized penalized spline smoothing, J. R. Stat. Soc. Ser. B 71 (2009) 487–503.
[12] T. Krivobokova, G. Kauermann, A note on penalized spline smoothing with correlated errors, J. Amer. Statist. Assoc. 102 (2007) 328–1337.
[13] H. Liang, S. Thurston, D. Ruppert, T. Apanasovich, R. Hauser, Additive partial linear models with measurement errors, Biometrika 95 (2008) 667–678.
[14] O.B. Linton, Efficient estimation of additive nonparametric regression models, Biometrika 84 (1997) 469–473.
[15] O.B. Linton, J.P. Nielsen, A kernel method of estimating structured nonparametric regression based on marginal integration, Biometrika 82 (1995) 93–101.
[16] Z. Lu, A. Lundervold, D. Tjøstheim, Q. Yao, Exploring spatial nonlinearity using additive approximation, Bernoulli 13 (2007) 447–472.
[17] E. Mammen, O.B. Linton, J. Nielsen, The existence and asymptotic properties of a backfitting projection algorithm under weak conditions, Ann. Statist. 27 (1999) 1443–1490.
[18] C. Martins-Filho, K. Yang, Finite sample performance of kernel-based regression methods for non-parametric additive models under common bandwidth selection criterion, J. Nonparametr. Stat. 19 (2007) 23–62.
[19] J.P. Nielsen, S. Sperlich, Smooth backfitting in practice, J. R. Stat. Soc. Ser. B 67 (2005) 43–61.
[20] J. Roca-Pardiñas, C. Cadarso-Suárez, W. Gonzalez-Manteiga, Testing for interactions in generalized additive models: application to *SO*2 pollution data, Stat. Comput. 15 (2005) 289–299.
[21] J. Roca-Pardiñas, C. Cadarso-Suárez, P.G. Tahoces, M.J. Lado, Assessing continuous bivariate effects among different groups through nonparametric regression models: an application to breast cancer detection, Comput. Statist. Data Anal. 52 (2008) 1958–1970.
[22] S. Sperlich, D. Tjøstheim, L. Yang, Nonparametric estimation and testing of interaction in additive models, Econometric Theory 18 (2002) 197–251.
[23] C.J. Stone, Additive regression and other nonparametric models, Ann. Statist. 13 (1985) 689–705.
[24] D. Tjøstheim, B. Auestad, Nonparametric identification of nonlinear time series: projections, J. Amer. Statist. Assoc. 89 (1994) 1398–1409.
[25] L. Wang, L. Yang, Spline-backfitted kernel smoothing of nonlinear additive autoregression model, Ann. Statist. 35 (2007) 2474–2503.
[26] L. Wang, L. Yang, Simultaneous spline confidence bands for time series prediction function, J. Nonparametric Statistics (2010) (in press).
[27] J. Wang, L. Yang, Polynomial spline confidence bands for regression curves, Statist. Sinica 19 (2009) 325–342.
[28] Y. Xia, Bias-corrected confidence bands in nonparametric regression, J. R. Stat. Soc. Ser. B 60 (1998) 797–811.
[29] L. Yang, B.U. Park, L. Xue, W. Härdle, Estimation and testing of varying coefficients in additive models with marginal integration, J. Amer. Statist. Assoc. 101 (2006) 1212–1227.