Econometric Theory, **26**, 2010, 29–59.
doi:10.1017/S0266466609090604

SPLINE-BACKFITTED KERNEL SMOOTHING OF ADDITIVE COEFFICIENT MODEL

Rong Liu
University of Toledo

Lijian Yang
Michigan State University

Additive coefficient model (Xue and Yang, 2006a, 2006b) is a flexible regression and autoregression tool that circumvents the "curse of dimensionality." We propose spline-backfitted kernel (SBK) and spline-backfitted local linear (SBLL) estimators for the component functions in the additive coefficient model that are both (i) computationally expedient so they are usable for analyzing high dimensional data, and (ii) theoretically reliable so inference can be made on the component functions with confidence. In addition, they are (iii) intuitively appealing and easy to use for practitioners. The SBLL procedure is applied to a varying coefficient extension of the Cobb-Douglas model for the U.S. GDP that allows nonneutral effects of the R&D on capital and labor as well as in total factor productivity (TFP).

1. INTRODUCTION

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

The comments from two anonymous referees and co-editor Oliver Linton have resulted in substantial improvement of the work. This research is part of the first author's dissertation work under the supervision of the second author, and has been supported in part by NSF awards DMS 0405330 and DMS 0706518. Address correspondence to Rong Liu, Department of Mathematics, University of Toledo, Toledo, OH 43606, USA; e-mail: rong.liu@utoledo.edu; or to Lijian Yang, Department of Statistics and Probability, Michigan State University, East Lansing, MI 48824, USA; e-mail: yang@stt.msu.edu.

 0266-4666/10 $15.00

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

in which the predictor vector $(\mathbf{X},\mathbf{T})$ consists of the tuning variables $\mathbf{X} = (X_1,\ldots,X_{d_2})^T \in R^{d_2}$ and linear predictors $\mathbf{T} = (T_1,\ldots,T_{d_1})^T \in R^{d_1}$. The functional coefficient model of Chen and Tsay (1993b) corresponds to the case $d_2 = 1$, the varying coefficient model of Hastie and Tibshirani (1993) corresponds to the case $d_2 = d_1$ and for each $l = 1,\ldots,d_1$ only one single significant $m_{\alpha l}$ with $\alpha = l$. Also included as special cases of model (1) are the additive model of Hastie and Tibshirani (1990) and Chen and Tsay (1993a), and the multivariate linear regression model (see Xue and Yang, 2006a, for detailed discussion). Model (1)'s versatility for econometric applications is illustrated by the following example: Consider the forecasting of the U.S. GDP annual growth rate, which is modeled as the total factor productivity (TFP) growth rate plus a linear function of the capital growth rate and the labor growth rate, according to the classic Cobb-Douglas model (Cobb and Douglas, 1928). As pointed out in Li and Racine (2007, p. 302), it is unrealistic to ignore the nonneutral effect of R&D spending on the TFP growth rate and on the complementary slopes of capital and labor growth rates. Thus, a smooth coefficient model should fit the production function better than the parametric Cobb-Douglas model. Indeed, Figure 1 shows that a smooth coefficient model has much smaller rolling forecast errors than the parametric Cobb-Douglas model, based on data from 1959 to 2002. In addition, Figure 2 shows that the TFP growth rate is a function of R&D spending, not a constant.

Many methods exist for the estimation of functional/varying coefficient models; see Cai, Fan, and Yao (2000) and Yang, Park, Xue, and Härdle (2006) for kernel type estimators; see Huang, Wu, and Zhou (2002) and Huang and Shen (2004) for spline estimators. These published works have partial success in addressing the inaccuracy of estimating multivariate nonparametric functions,

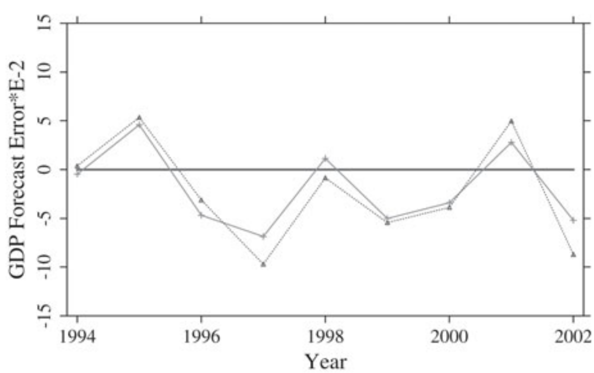

FIGURE 1. Errors of GDP forecasts: solid line = model (31); dotted line = model (30).

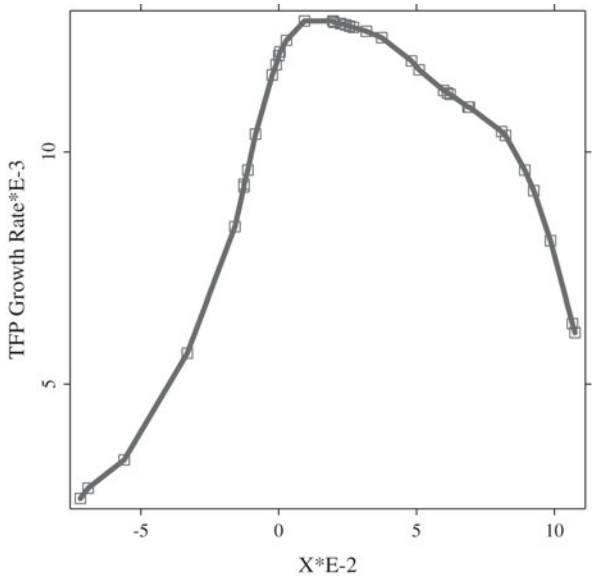

FIGURE 2. Estimation of TFP growth rate function $c_1 + m_{\text{SBLL},41}\left(X_{t-3}\right)$.

commonly known as the "curse of dimensionality." Typically, optimal convergence rates of the coefficient function estimators are established, locally for kernel estimators or globally for spline estimators.

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

We extend the oracle smoothing idea of Linton (1997) and Wang and Yang (2007) to model (1). If all the nonparametric functions of the last $d_2 - 1$ variables, $\{m_{\alpha l}(x_\alpha)\}_{l=1,\alpha=2}^{d_1,d_2}$, and all the constants $\{m_{0l}\}_{l=1}^{d_1}$ were known by "oracle", one could define a new variable $Y_{,1} = \sum_{l=1}^{d_1} m_{1l}(X_1)T_l + \sigma(\mathbf{X},\mathbf{T})\varepsilon = Y - \sum_{l=1}^{d_1}\left\{m_{0l} + \sum_{\alpha=2}^{d_2} m_{\alpha l}(X_\alpha)\right\}T_l$ and estimate all functions $\{m_{1l}(x_1)\}_{l=1}^{d_1}$ by linear regression of $Y_{,1}$ on $T_1,\ldots,T_{d_1}$ with kernel weights computed from variable X_1. These would-be estimators do not suffer from the "curse of dimensionality" and are called "oracle smoothers." We propose to pre-estimate the functions $\{m_{\alpha l}(x_\alpha)\}_{l=1,\alpha=2}^{d_1,d_2}$ and constants $\{m_{0l}\}_{l=1}^{d_1}$ by linear spline, then use these estimates as substitutes to obtain an approximation $\hat{Y}_{,1}$ to the variable $Y_{,1}$, and construct "oracle" estimators based on $\hat{Y}_{,1}$. As in Wang and Yang (2007), the theoretical contribution of this paper is proving that the error caused by this "cheating" is negligible. Consequently, the SBK/SBLL estimators are uniformly (over the data range) equivalent to univariate kernel/local linear "oracle smoothers," automatically inheriting all their oracle efficiency properties. Our proof relies on the same principles of "reducing bias by undersmoothing in step one" and "averaging out the variance in step two," accomplished with the joint asymptotics of kernel and spline functions. Compared to Wang and Yang (2007), a major theoretical complication is the dependence structure of $\mathbf{T}$ on $\mathbf{X}$, necessitating Assumption 2 on the second moment matrix $\mathbf{Q}(\mathbf{x}) = \mathrm{E}\left(\mathbf{T}\mathbf{T}^T\,|\mathbf{X}=\mathbf{x}\right)$; see the detailed discussion on Assumption 2 at the end of Section 2 and the extra step to estimate $\mathbf{Q}(\mathbf{x})$ in Section 5. In contrast, for the additive model of Wang and Yang (2007), there is no need for Assumption 2 and of estimating $\mathbf{Q}(\mathbf{x}) \equiv 1$. Another innovation in this paper is the $\sqrt{n}$-consistent oracle estimation of constants $\{m_{0l}\}_{l=1}^{d_1}$ under conditions no more than second-order smoothness of $\{m_{\alpha l}(x_\alpha)\}_{l=1,\alpha=1}^{d_1,d_2}$. Xue and Yang (2006a) have provided $\sqrt{n}$-consistent estimation of constants $\{m_{0l}\}_{l=1}^{d_1}$ only under higher order smoothness assumptions, and Xue and Yang (2006b) failed to obtain $\sqrt{n}$-consistency for estimating $\{m_{0l}\}_{l=1}^{d_1}$. For the additive model of Wang and Yang (2007), there is only one such unknown constant, and it is $\sqrt{n}$-consistently estimated by the sample mean $\overline{Y}$. Lastly, asymptotic theory for the oracle smoothers is developed in Section 3 separately, whereas Wang and Yang (2007) used existing theory from kernel smoothing literature.

The paper is organized as follows. In Section 2 we discuss the assumptions of model (1). In Section 3, we introduce the oracle smoothers and discuss their asymptotic properties. In Section 4, we introduce the SBK and SBLL estimators, their uniform consistency, and asymptotic normal distributions. The ideas behind our proofs of the main theoretical results are given by decomposing the estimator's "cheating" error into a bias and a variance part. In Section 5, we discuss the implementation of the estimators. In Section 6, we apply the methods to an empirical example. All technical proofs are given in the Appendix.

2. ASSUMPTIONS ON THE MODEL

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

$$Y_i = m(\mathbf{X}_i, \mathbf{T}_i) + \sigma(\mathbf{X}_i, \mathbf{T}_i)\varepsilon_i \tag{2}$$

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

$$\mathrm{E}\{m_{\alpha l}(X_\alpha)\} = 0, 1 \le l \le d_1, 1 \le \alpha \le d_2, \tag{3}$$

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

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

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

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

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

in which $\{g_{0l}\}_{l=1}^{d_1}$ are finite constants. The constraints that $\mathrm{E}\{g_{\alpha l}(X_\alpha)\} = 0$, $1 \le \alpha \le d_2$ ensure unique additive representation of m_l as expressed in (3), but are not necessary for the definition of space $\mathcal{M}$. In what follows, denote by E_n the empirical expectation, $\mathrm{E}_n\varphi = \sum_{i=1}^n \varphi(\mathbf{X}_i, \mathbf{T}_i)/n$. We introduce two inner products on $\mathcal{M}$. For functions $g_1, g_2 \in \mathcal{M}$, the theoretical and empirical inner products are defined respectively as $\langle g_1, g_2\rangle = \mathrm{E}\{g_1(\mathbf{X}, \mathbf{T})g_2(\mathbf{X}, \mathbf{T})\}$, $\langle g_1, g_2\rangle_n = \mathrm{E}_n\{g_1(\mathbf{X}, \mathbf{T})g_2(\mathbf{X}, \mathbf{T})\}$. The corresponding induced norms are $\|g_1\|_2^2 = \mathrm{E}g_1^2(\mathbf{X}, \mathbf{T})$,

$\|g_1\|_{2,n}^2 = \mathrm{E}_n g_1^2(\mathbf{X},\mathbf{T})$. The model space $\mathcal{M}$ is called *theoretically (empirically) identifiable*, if for any $g \in \mathcal{M}$, $\|g\|_2 = 0$ ($\|g\|_{2,n} = 0$) implies that $g = 0$, a.s.

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

We need the following assumptions on the data-generating process.

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

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

Assumption 3. The vector process $\{\varsigma_t\}_{t=-\infty}^{\infty} = \{(Y_t, \mathbf{X}_t, \mathbf{T}_t)\}_{t=-\infty}^{\infty}$ is strictly stationary and geometrically strongly mixing, that is, its α-mixing coefficient $\alpha(k) \le c\rho^k$, for constants $c > 0$, $0 < \rho < 1$, where $\alpha(k) = \sup_{A \in \sigma(\varsigma_t, t \le 0), B \in \sigma(\varsigma_t, t \ge k)} |P(A)P(B) - P(A \cap B)|$.

Assumption 4. The coefficient components, $m_{\alpha l} \in C^1[0,1]$, $m'_{\alpha l} \in \mathrm{Lip}([0,1], C_\infty), \forall 1 \le \alpha \le d_2, 1 \le l \le d_1$ with $m_{1l} \in C^2[0,1], \forall 1 \le l \le d_1$.

Assumption 5. The conditional variance function $\sigma^2(\mathbf{x},\mathbf{t})$ is measurable and bounded. The errors $\{\varepsilon_i\}_{i=1}^n$ satisfy $\mathrm{E}(\varepsilon_i | \mathcal{F}_i) = 0$, $\mathrm{E}\left(\varepsilon_i^2 | \mathcal{F}_i\right) = 1$, $\mathrm{E}\left(|\varepsilon_i|^{2+\eta} | \mathcal{F}_i\right) \le C_\eta$ for some $\eta \in (1/2, 1]$ and the sequence of σ-fields $F_i = \sigma\left\{\left(\mathbf{X}_j, \mathbf{T}_j\right), j \le i; \varepsilon_j, j \le i-1\right\}$ for $i = 1, \ldots, n$.

Assumption 6. The marginal density $f_1(x_1)$ of X_1 and the conditional second-moment matrix function $\mathbf{Q}_1(x_1)$ defined in (4) both have continuous derivatives on $[0,1]$.

Assumptions 1–5 are common in the literature; see, for instance, Huang and Yang (2004), Huang and Shen (2004), and especially Xue and Yang (2006b). Assumption 6 is needed only for the asymptotic theory of oracle "kernel smoother," but not for the oracle "local linear smoother." Assumption 2 implies also that for all $x_\alpha \in [0,1], 1 \le \alpha \le d_2$ and $l, l' = 1, \ldots, d_1$,

$$c_{\mathbf{Q}} \mathbf{I}_{d_1 \times d_1} \le \mathbf{Q}_\alpha(x_\alpha) = \{q_\alpha(x_\alpha)\}_{l,l'=1}^{d_1} = \mathrm{E}(\mathbf{T}\mathbf{T}^T | X_\alpha = x_\alpha) \le C_{\mathbf{Q}} \mathbf{I}_{d_1 \times d_1} \quad (4)$$

$$c_\delta \le \mathrm{E}\left\{\left(T_l T_{l'}\right)^{2+\delta} | X_\alpha = x_\alpha\right\} \le C_\delta.$$

Furthermore, Assumptions 2 and 5 imply that for some constant $C > 0$,

$$\max_{1\le l\le d_1} \mathrm{E}\left|T_l\right|^{2+\eta} < C \max_{1\le l\le d_1} \mathrm{E}\left|T_l T_l\right|^{2+\delta} = C \max_{1\le l\le d_1} \mathrm{E}\left|T_l\right|^{4+2\delta} \le C C_\delta < +\infty. \qquad \textbf{(5)}$$

At one referee's request, we provide here insight into the relationship allowed between the vectors $\mathbf{T}$ and $\mathbf{X}$ under Assumption 2. It is instructive to first understand what $\mathbf{T}$ and $\mathbf{X}$ cannot be in the context of identifiability for functions $\{m_{\alpha l}(x_\alpha)\}_{l=1,\alpha=1}^{d_1,d_2}$. Suppose that the vector $\mathbf{X}$ is centered so that $\mathrm{E}\mathbf{X} = \mathbf{0}$. Then model (1) is unidentifiable when $(T_1, T_2) = (X_1, X_2)$ since $-3X_2T_1 + 3X_1T_2 = 0$, $\mathrm{E}(-3X_2) = \mathrm{E}(3X_1) = 0$, and the function $m(\mathbf{x}, \mathbf{t})$ in (1) is expressed as

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

so one can use $m_{21}^*(x_2) = m_{21}(x_2) - 3x_2$ and $m_{12}^*(x_1) = m_{12}(x_1) + 3x_1$ to replace $m_{21}(x_2)$ and $m_{12}(x_1)$ without changing the data generating process (1). In other words, the functions $m_{21}(x_2)$ and $m_{12}(x_1)$ are unidentifiable. Xue and Yang (2006a, p. 2523) gave a similar counterexample and discussed why an unidentifiable model may perform better for prediction.

More generally, it is revealing to note that Assumption 2 not only rules out the above anomaly, but it also does not allow the possibility that there exist two T_l's $(1 \le l \le d_1)$ almost surely equal to two Borel functions of $\mathbf{X}$. To see this, suppose that $(T_1, T_2) = \{\varphi_1(\mathbf{X}), \varphi_2(\mathbf{X})\}$, a.s., for some Borel functions φ_1 and φ_2. Assumption 2 implies that

$$c_{\mathbf{Q}}\mathbf{I}_{2\times 2} \le \mathrm{E}\left\{\left.\begin{pmatrix} T_1^2 & T_1T_2 \\ T_1T_2 & T_2^2\end{pmatrix}\right| \mathbf{X} = \mathbf{x}\right\} \le C_{\mathbf{Q}}\mathbf{I}_{2\times 2}, \forall \mathbf{x} \in \chi,$$

leading to

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

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

3. THE ORACLE SMOOTHERS

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

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

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

$$\tilde{m}_{\mathrm{K},1,\cdot}\left(x_1\right) = \left\{\tilde{m}_{\mathrm{K},11}\left(x_1\right), \ldots, \tilde{m}_{\mathrm{K},1d_1}\left(x_1\right)\right\}^T = \underset{\lambda=(\lambda_l)_{1\leq l \leq d_1}}{\operatorname{argmin}} L\left(\lambda, m_{\text{-}1,\cdot}, x_1\right),$$

in which

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

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

$$\tilde{m}_{\mathrm{K},1,\cdot}\left(x_1\right) = \left(\mathbf{C}_{\mathrm{K}}^T \mathbf{W}_1 \mathbf{C}_{\mathrm{K}}\right)^{-1} \mathbf{C}_{\mathrm{K}}^T \mathbf{W}_1 \mathbf{Y}_1 = \left(\frac{1}{n} \mathbf{C}_{\mathrm{K}}^T \mathbf{W}_1 \mathbf{C}_{\mathrm{K}}\right)^{-1} \frac{1}{n} \mathbf{C}_{\mathrm{K}}^T \mathbf{W}_1 \mathbf{Y}_1, \qquad \textbf{(6)}$$

in which

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

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

$K_h(u) = K(u/h)/h$ for a kernel function K and bandwidth h that satisfy Assumption 7 below.

Assumption 7. The function K is a symmetric probability density function supported on $[-1,1]$, and $K \in \mathrm{Lip}([-1,1], C_K)$ for some $C_K > 0$, while the bandwidth $h = h_{1,n} > 0, h \sim n^{-1/5}$.

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

$$\tilde{m}_{\mathrm{LL},1,\cdot}(x_1) = \left(\mathbf{I}_{d_1 \times d_1}, \mathbf{0}_{d_1 \times d_1}\right) \left(\frac{1}{n}\mathbf{C}_{\mathrm{LL},1}^T \mathbf{W}_1 \mathbf{C}_{\mathrm{LL},1}\right)^{-1} \frac{1}{n}\mathbf{C}_{\mathrm{LL},1}^T \mathbf{W}_1 \mathbf{Y}_1, \tag{7}$$

in which

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

In this paper, denote $\mu_2(K) = \int u^2 K(u)\,du$, $\|K\|_2^2 = \int K(u)^2\,du$, $\mathbf{Q}_1(x_1)$ as in (4) and define the following bias and variance coefficients:

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

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

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

$$\left\{ v_{l,l',1}(x_1) \right\}_{l,l'=1}^{d_1} = \mathbf{Q}_1(x_1)^{-1} \Sigma_1(x_1) \mathbf{Q}_1(x_1)^{-1}. \tag{8}$$

THEOREM 1. *Under Assumptions 1–5 and 7, for any $x_1 \in [h, 1-h]$, as $n \to \infty$ the oracle local linear smoother $\tilde{m}_{\mathrm{LL},1,\cdot}(x_1)$ given in (7) satisfies*

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

With Assumption 6 in addition, the oracle kernel smoother $\tilde{m}_{\mathrm{K},1,\cdot}(x_1)$ in (6) satisfies

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

THEOREM 2. *Under Assumptions 1–5 and 7, as $n \to \infty$, the oracle local linear smoother $\tilde{m}_{\mathrm{LL},1,\cdot}(x_1)$ given in (7) satisfies*

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

With Assumption 6 in addition, the oracle kernel smoother $\tilde{m}_{\mathrm{K},1,\cdot}(x_1)$ in (6) satisfies

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

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

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

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

except that in Assumption 4 one has to replace "$m_{1l} \in C^2[0,1], \forall 1 \le l \le d_1$" with "$m_{\alpha l} \in C^2[0,1], \forall 1 \le l \le d_1$" and in Assumption 6, $f_1(x_1)$ and $Q_1(x_1)$ have to be replaced with $f_\alpha(x_\alpha)$ and $Q_\alpha(x_\alpha)$.

The proofs of Theorems 1 and 2 can be found in Liu and Yang (2008, Sect. A.4). The same oracle idea applies to the constants as well. Define the would-be estimators of constants $(m_{0l})_{1 \le l \le d_1}^T$ as the following least squares solution:

$$\tilde{m}_0 = (\tilde{m}_{0l})_{1<l<d_1}^T = \arg\min \sum_{i=1}^n \left\{ Y_{ic} - \sum_{l=1}^{d_1} m_{0l} T_{il} \right\}^2, \tag{9}$$

in which the oracle responses are

$$Y_{ic} = Y_i - \sum_{l=1}^{d_1} \sum_{\alpha=1}^{d_2} m_{\alpha l}(X_{i\alpha}) T_{il} = \sum_{l=1}^{d_1} m_{0l} T_{il} + \sigma(\mathbf{X}_i, \mathbf{T}_i) \varepsilon_i. \tag{10}$$

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

PROPOSITION 1. *Under Assumptions 1–5 and 8, as $n \to \infty$,*
$\sup_{1 \le l \le d_1} |\tilde{m}_{0l} - m_{0l}| = O_p\left(n^{-1/2}\right).$

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

4. SPLINE-BACKFITTED KERNEL ESTIMATORS

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

Assumption 8. The number of interior knots $N = N_n \sim n^{1/4}\log n$, and hence $H \sim n^{-1/4}(\log n)^{-1}$.

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

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

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

$$G_{n,\alpha}^0 = \left\{ g_\alpha : g_\alpha(x_\alpha) \equiv \sum_{J=0}^{N+1} \lambda_J b_J(x_\alpha), \mathrm{E}_n\{g_\alpha(X_\alpha)\} = 0 \right\}, 1 \le \alpha \le d_2,$$

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

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

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

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

$$\hat{m}(\mathbf{x},\mathbf{t}) = \sum_{l=1}^{d_1} \hat{m}_l(\mathbf{x}) t_l = \operatorname*{argmin}_{g \in G_n^0} \sum_{i=1}^{n} \{Y_i - g(\mathbf{X}_i,\mathbf{T}_i)\}^2. \tag{11}$$

Since $\hat{m}(\mathbf{x},\mathbf{t}) \in G_n^0$, one can write $\hat{m}_l(\mathbf{x}) = \widehat{m}_{0l} + \sum_{\alpha=1}^{d_2} \hat{m}_{\alpha l}(x_\alpha)$; for $\widehat{m}_{0l} \in R$ and $\hat{m}_{\alpha l}(x_\alpha) \in G_{n,\alpha}^0$. Simple algebra shows that the following oracle estimators of the constants m_{0l} are exactly equal to $\widehat{m}_{0l}$, in which the oracle pseudo-responses $\hat{Y}_{ic} = Y_i - \sum_{l=1}^{d_1}\sum_{\alpha=1}^{d_2} \hat{m}_{\alpha l}(X_{i\alpha}) T_{il}$, which mimick the oracle responses Y_{ic} in (10)

$$\hat{m}_0 = \left(\hat{m}_{0l}\right)_{1\le l\le d_1}^T = \arg\min_{\left(\lambda_{01},\ldots,\lambda_{0d_1}\right)} \sum_{i=1}^{n}\left\{\hat{Y}_{ic} - \sum_{l=1}^{d_1}\lambda_{0l}T_{il}\right\}^2. \tag{12}$$

PROPOSITION 2. *Under Assumptions 1–5 and 8, as $n \to \infty$,* $\sup_{1\le l\le d_1}\left|\hat{m}_{0l} - \tilde{m}_{0l}\right| = O_p\left(n^{-1/2}\right)$, *hence* $\sup_{1\le l\le d_1}\left|\hat{m}_{0l} - m_{0l}\right| = O_p\left(n^{-1/2}\right)$, *following Proposition 1.*

Define next the oracle pseudo-responses $\hat{Y}_{i1} = Y_i - \sum_{l=1}^{d_1}\left(\hat{m}_{0l} + \sum_{\alpha=2}^{d_2}\hat{m}_{\alpha l}(X_{i\alpha})\right)T_{il}$ and $\hat{\mathbf{Y}}_1 = \left(\hat{Y}_{11},\cdots,\hat{Y}_{n1}\right)^T$, with $\hat{m}_{0l}$ and $\hat{m}_{\alpha l}$ defined in (12) and (11), respectively. The spline-backfitted kernel (SBK) and spline-backfitted local linear (SBLL) estimators are

$$\hat{m}_{\mathrm{SBK},1,\cdot}(x_1) = \left(\mathbf{C}_{\mathrm{K}}^T\mathbf{W}_1\mathbf{C}_{\mathrm{K}}\right)^{-1}\mathbf{C}^T\mathbf{W}_1\hat{\mathbf{Y}}_1 = \left(\frac{1}{n}\mathbf{C}_{\mathrm{K}}^T\mathbf{W}_1\mathbf{C}_{\mathrm{K}}\right)^{-1}\frac{1}{n}\mathbf{C}^T\mathbf{W}_1\hat{\mathbf{Y}}_1, \tag{13}$$

$$\hat{m}_{\mathrm{SBLL},1,\cdot}(x_1) = \left(\mathbf{I}_{d_1\times d_1}, \mathbf{0}_{d_1\times d_1}\right)\left(\frac{1}{n}\mathbf{C}_{\mathrm{LL},1}^T\mathbf{W}_1\mathbf{C}_{\mathrm{LL},1}\right)^{-1}\frac{1}{n}\mathbf{C}_{\mathrm{LL},1}^T\mathbf{W}_1\hat{\mathbf{Y}}_1. \tag{14}$$

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

THEOREM 3. *Under Assumptions 1–5, 7, and 8, as $n \to \infty$, the SBK estimator $\hat{m}_{\mathrm{SBK},1,\cdot}(x_1)$ in (13) and the SBLL estimator $\hat{m}_{\mathrm{SBLL},1,\cdot}(x_1)$ in (14) satisfy*

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

Theorem 3 follows from Propositions 2, 3, and 4, and remains true if Assumption 8 on the number of knots is of the more general form $N \sim n^{1/4}N'$, where $N' \to \infty, N'/n^r \to 0, \forall r > 0$ as $n \to \infty$. In other words, one slightly undersmoothes in the first step of linear spline regression to reduce the bias. The larger variance is reduced in the second step of kernel smoothing, where a bandwidth h of optimal order is used, see Assumption 7. The following corollary provides the asymptotic distributions of $\hat{m}_{\mathrm{SBLL},1,\cdot}(x_1)$ and $\tilde{m}_{\mathrm{K},1,\cdot}(x_1)$. The proof of this corollary is straightforward from Theorems 1 and 3.

COROLLARY 1. *Under Assumptions 1–5, 7, and 8, for any* $x_1 \in [h, 1-h]$, *as* $n \to \infty$, *the SBLL estimator* $\hat{m}_{\mathrm{SBLL},1,\cdot}(x_1)$ *in (14) satisfies*

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

$$\to N\left(0, \left\{v_{l,l',1}(x_1)\right\}_{l,l'=1}^{d_1}\right),$$

and with the additional Assumption 6, the SBK estimator $\hat{m}_{\mathrm{SBK},1,\cdot}(x_1)$ *in (13) satisfies*

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

$$\to N\left(0, \left\{v_{l,l',1}(x_1)\right\}_{l,l'=1}^{d_1}\right),$$

where $b_{\mathrm{LL},l,l',1}(x_1)$, $b_{\mathrm{K},l,l',1}(x_1)$, *and* $v_{l,l',1}(x_1)$ *are defined as (8).*

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

$$\hat{m}_{\mathrm{SBK},\alpha,\cdot}(x_\alpha) = \left(\frac{1}{n}\mathbf{C}_{\mathrm{K}}^T \mathbf{W}_\alpha \mathbf{C}_{\mathrm{K}}\right)^{-1} \frac{1}{n}\mathbf{C}_{\mathrm{K}}^T \mathbf{W}_\alpha \hat{\mathbf{Y}}_\alpha, \tag{15}$$

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

To understand the proof of Theorem 3, we study the difference between the smoothed backfitted estimator $\hat{m}_{\mathrm{SBK},1l'}(x_1)$ and the smoothed "oracle" estimator $\tilde{m}_{\mathrm{K},1l'}(x_1)$. First, define the theoretical inner product of b_J and 1 with respect to the αth 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

$$b_{J,\alpha}(x_\alpha) = b_J(x_\alpha) - \frac{c_{J,\alpha}}{c_{J-1,\alpha}} b_{J-1}(x_\alpha), \qquad B_{J,\alpha}(x_\alpha) = \frac{b_{J,\alpha}(x_\alpha)}{\left\|b_{J,\alpha}\right\|_2}, 1 \le J \le N+1, \tag{16}$$

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

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

$$\sum_{i=1}^{n}\left[\Gamma_i - \sum_{l=1}^{d_1}\left\{\gamma_{0,l} + \sum_{\alpha=1}^{d_2}\sum_{J=1}^{N+1}\gamma_{J,\alpha,l}B_{J,\alpha}\left(X_{i\alpha}\right)\right\}T_{il}\right]^2, \tag{17}$$

so one can rewrite the linear spline estimator in (11) as $\hat{m}\left(\mathbf{x},\mathbf{t}\right) = \left(\mathbf{P}_n\mathbf{Y}\right)\left(\mathbf{x},\mathbf{t}\right)$, where we denote by $\mathbf{Y} = \left(Y_i\right)^T_{1\le i\le n}$ the response vector. The coefficients of the linear regressors $t_l, 1 \le l \le d_1$ are denoted as the multivariate additive spline functions

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

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

$$\left(\mathbf{P}_{n,\alpha,l}\Gamma\right)\left(x_\alpha\right) = \left(\mathbf{P}^*_{n,\alpha,l}\Gamma\right)\left(x_\alpha\right) - n^{-1}\sum_{i=1}^{n}\left(\mathbf{P}^*_{n,\alpha,l}\Gamma\right)\left(X_{i\alpha}\right). \tag{18}$$

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

$$\tilde{m}_l\left(\mathbf{x}\right) = \left(\mathbf{P}_{n,l}\mathbf{m}\right)\left(\mathbf{x}\right), \tilde{m}_{\alpha l}\left(x_\alpha\right) = \left(\mathbf{P}_{n,\alpha,l}\mathbf{m}\right)\left(x_\alpha\right),$$
$$\tilde{\varepsilon}_l\left(\mathbf{x}\right) = \left(\mathbf{P}_{n,l}\mathbf{E}\right)\left(\mathbf{x}\right), \tilde{\varepsilon}_{\alpha l}\left(x_\alpha\right) = \left(\mathbf{P}_{n,\alpha,l}\mathbf{E}\right)\left(x_\alpha\right), \tag{19}$$

where $\mathbf{m} = \left\{m\left(\mathbf{X}_i,\mathbf{T}_i\right)\right\}^T_{1\le i\le n}$ is the true function vector and $\mathbf{E} = \left\{\sigma\left(\mathbf{X}_i, \mathbf{T}_i\right)\varepsilon_i\right\}^T_{1\le i\le n}$ the error vector. Due to the linearity of operators $\mathbf{P}_{n,\alpha}$ and $\mathbf{P}_{n,\alpha,l}$, $1 \le l \le d_1, 1 \le \alpha \le d_2$ and $\mathbf{Y} = \mathbf{m} + \mathbf{E}$ due to (2), one has the following crucial decomposition for proving Theorem 3:

$$\hat{m}_l\left(\mathbf{x}\right) = \tilde{m}_l\left(\mathbf{x}\right) + \tilde{\varepsilon}_l\left(\mathbf{x}\right),$$
$$\hat{m}_{\alpha l}\left(x_\alpha\right) = \tilde{m}_{\alpha l}\left(x_\alpha\right) + \tilde{\varepsilon}_{\alpha l}\left(x_\alpha\right), \qquad 1 \le l \le d_1, 1 \le \alpha \le d_2. \tag{20}$$

We define additionally an auxiliary entity

$$\tilde{\varepsilon}^*_{\alpha l}\left(x_\alpha\right) = \left(\mathbf{P}^*_{n,\alpha,l}\mathbf{E}\right)\left(x_\alpha\right), \qquad 1 \le l \le d_1, 1 \le \alpha \le d_2. \tag{21}$$

Definition (18) implies that $\tilde{\varepsilon}_{\alpha l}\left(x_\alpha\right)$ is simply the empirical centering of $\tilde{\varepsilon}^*_{\alpha l}\left(x_\alpha\right)$; i.e.,

$$\tilde{\varepsilon}_{\alpha l}\left(x_\alpha\right) \equiv \tilde{\varepsilon}^*_{\alpha l}\left(x_\alpha\right) - n^{-1}\sum_{i=1}^{n}\tilde{\varepsilon}^*_{\alpha l}\left(X_{i\alpha}\right). \tag{22}$$

According to (6) and (13),

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

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

where making use of the definition of $\hat{m}_{0l}$ and the signal noise decomposition (20), the difference $\tilde{m}_{\mathrm{K},1,\cdot}(x_1) - \hat{m}_{\mathrm{SBK},1,\cdot}(x_1) - \hat{m}_{0,\cdot} + m_{0,\cdot}$ can be treated as the sum of two terms

$$\begin{aligned}&\left(\frac{1}{n}\mathbf{C}_{\mathrm{K}}^T\mathbf{W}_1\mathbf{C}_{\mathrm{K}}\right)^{-1}\frac{1}{n}\mathbf{C}_{\mathrm{K}}^T\mathbf{W}\left[\sum_{l=1}^{d_1}\left\{m_{\text{-}1,l}\left(\mathbf{X}_{i,\text{-}1}\right) - \hat{m}_{\text{-}1,l}\left(\mathbf{X}_{i,\text{-}1}\right)\right\}T_{il}\right]_{i=1}^{n} \\ &\quad = \left(\frac{1}{n}\mathbf{C}_{\mathrm{K}}^T\mathbf{W}_1\mathbf{C}_{\mathrm{K}}\right)^{-1}\left\{\Psi_b(x_1) + \Psi_v(x_1)\right\}_{l'=1}^{d_1}, \end{aligned} \tag{23}$$

where

$$\Psi_b(x_1) = \frac{1}{n}\mathbf{C}_{\mathrm{K}}^T\mathbf{W}_1\left[\sum_{l=1}^{d_1}\left\{m_{\text{-}1,l}\left(\mathbf{X}_{i,\text{-}1}\right) - \tilde{m}_{\text{-}1,l}\left(\mathbf{X}_{i,\text{-}1}\right)\right\}T_{il}\right]_{i=1}^{n} = \left\{\Psi_{b,l'}(x_1)\right\}_{l'=1}^{d_1}, \tag{24}$$

$$\begin{aligned}\Psi_v(x_1) &= \frac{1}{n}\mathbf{C}_{\mathrm{K}}^T\mathbf{W}_1\left[\sum_{l=1}^{d_1}\tilde{\varepsilon}_{\text{-}1,l}\left(\mathbf{X}_{i,\text{-}1}\right)T_{il}\right]_{i=1}^{n} \\ &= \left\{\Psi_{v,l'}(x_1)\right\}_{l'=1}^{d_1},\ \tilde{\varepsilon}_{\text{-}1,l}\left(\mathbf{X}_{i,\text{-}1}\right) = \sum_{\alpha=2}^{d_2}\tilde{\varepsilon}_{\alpha l}\left(X_{i\alpha}\right),\end{aligned} \tag{25}$$

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

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

The term $\Psi_b(x_1)$ is induced by the bias term $\hat{m}_{\text{-}1,l}\left(\mathbf{X}_{i,\text{-}1}\right) - m_{\text{-}1,l}\left(\mathbf{X}_{i,\text{-}1}\right)$, and $\Psi_v(x_1)$ relates to the noise terms $\tilde{\varepsilon}_{\text{-}1,l}\left(\mathbf{X}_{i,\text{-}1}\right)$. Both of these have order $o_p(n^{-2/5})$ by Propositions 3 and 4 below.

PROPOSITION 3. *Under Assumptions 1–4, 7, and 8, as $n \to \infty$,*

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

PROPOSITION 4. *Under Assumptions 1–5 and 7–8, as $n \to \infty$,* $\sup_{1\le l'\le d_1}\sup_{x_1\in[0,1]}\left|\Psi_{v,l'}(x_1)\right| = O_p\left(N(\log n)^2/n + H^2\right) = o_p\left(n^{-2/5}\right)$.

According to (22) and (25), we can write $\Psi_{v,l'}(x_1) = \Psi^{(2)}_{v,l'}(x_1) - \Psi^{(1)}_{v,l'}(x_1)$, where

$$\Psi^{(1)}_{v,l'}(x_1) = n^{-1}\sum_{i=1}^{n}\sum_{l=1}^{d_1} K_h(X_{i1}-x_1)\,T_{il}T_{il'}\cdot n^{-1}\sum_{i=1}^{n}\tilde{\varepsilon}^*_{\text{-}1,l}\left(\mathbf{X}_{i,\text{-}1}\right) \quad \text{and} \tag{26}$$

$$\Psi^{(2)}_{v,l'}(x_1) = n^{-1}\sum_{i=1}^{n}\sum_{l=1}^{d_1} K_h(X_{i1}-x_1)\,T_{il}T_{il'}\tilde{\varepsilon}^*_{\text{-}1,l}\left(\mathbf{X}_{i,\text{-}1}\right), \tag{27}$$

in which $\tilde{\varepsilon}^*_{\text{-}1,l}\left(\mathbf{X}_{i,\text{-}1}\right) = \sum_{\alpha=2}^{d_2}\tilde{\varepsilon}^*_{\alpha l}(X_{i\alpha})$ and $\tilde{\varepsilon}^*_{\alpha l}(X_{i\alpha})$ is given in (21). Furthermore, if one denotes

$$\omega_{J,\alpha,l,l'}(\mathbf{X}_i,x_1) = T_{il}T_{il'}K_h(X_{i1}-x_1)\,B_{J,\alpha}(X_{i\alpha}), \quad \mu_{\omega_{J,\alpha,l,l'}}(x_1) = \mathrm{E}\omega_{J,\alpha,l,l'}(\mathbf{X},x_1), \tag{28}$$

then by (27), (A.2), and (21), $\Psi^{(2)}_{v,l'}(x_1)$ can be rewritten as

$$\Psi^{(2)}_{v,l'}(x_1) = n^{-1}\sum_{i=1}^{n}\sum_{l=1}^{d_1}\sum_{J=1}^{N+1}\sum_{\alpha=2}^{d_2}\tilde{a}_{J,\alpha,l}\,\omega_{J,\alpha,l,l'}(\mathbf{X}_i,x_1). \tag{29}$$

LEMMA 1. *Under Assumptions 1–5 and 7–8, as $n \to \infty$, $\Psi^{(1)}_{v,l'}(x_1)$ defined in (26) satisfies* $\sup_{1\le l'\le d_1}\sup_{x_1\in[0,1]}\left|\Psi^{(1)}_{v,l'}(x_1)\right| = O_p\left(N(\log n)^2/n\right)$.

LEMMA 2. *Under Assumptions 1–5 and 7–8, as $n \to \infty$, $\Psi^{(2)}_{v}(x_1)$ defined in (27) satisfies* $\sup_{1\le l'\le d_1}\sup_{x_1\in[0,1]}\left|\Psi^{(2)}_{v,l'}(x_1)\right| = O_p\left(H^2\right)$.

Proof of Proposition 3 is given in the Appendix; Proposition 4 follows from Lemmas 1 and 2. Lemma 2 follows from Lemmas A.11 and A.12, both proved in the Appendix, and the proof of Lemma 1 is given in the Appendix. A similar result can be proved for $\hat{m}_{\mathrm{SBLL},1l'}(x_1)$ by extending $\Psi_{b,l'}(x_1)$ and $\Psi_{v,l'}(x_1)$ to terms such as

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

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

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

5. IMPLEMENTATION

We implement our procedures with the following undersmoothing rule of thumb for the number of interior knots in the first step of linear spline smoothing:

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

which satisfies Assumption 8; i.e., $N = N_n \sim n^{1/4}\log n$, and ensures that the number of parameters in the linear least squares problem (17) is no more than $n/4$, i.e., $d_1\{1+d_2(N_n+1)\} \leq n/4$.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

6. EXAMPLE

We have applied the estimation procedure described in the previous section to both simulated and real data. Simulation results provide strong evidence in support of the asymptotic theory (for details, see Liu and Yang, 2008). In Section 6, we illustrate how the additive coefficient model is used to extend the Cobb-Douglas model for annual U.S. GDP growth. Denoted by Q_t the U.S. GDP at year t, K_t the U.S. capital at year t, L_t the U.S. labor at year t, and X_t the growth rate of the ratio of R&D expenditure to GDP at year t. All data have been downloaded from the Bureau of Economic Analysis (BEA) website for years $t = 1959, \ldots, 2002$ ($n = 44$). The standard Cobb-Douglas production function

(Cobb and Douglas, 1928) is $Q_t = A_t K_t^{\beta_1} L_t^{1-\beta_1}$, where A_t is the total factor productivity (TFP) of year t and β_1 is a parameter determined by technology. Define the following stationary time series variables:

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

Then the Cobb-Douglas equation implies the following simple regression model:

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

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

$$Y_t - T_{2t} = \beta_0 + \beta_1 (T_{1t} - T_{2t}). \quad \textbf{(30)}$$

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

$$Y_t - T_{2t} = c_1 + m_{41}(X_{t-3}) + \{c_2 + m_{52}(X_{t-4})\}(T_{1t} - T_{2t}). \quad \textbf{(31)}$$

The rolling forecast errors of GDP by SBLL fitting of model (31) and linear fitting of (30) are shown in Figure 1. The averaged squared prediction error (ASPE) for model (31) is

$$\frac{1}{9}\sum_{t=1994}^{2002}\left[Y_t - T_{2t} - \hat{c}_1 - \hat{m}_{\mathrm{SBLL},41}(X_{t-3}) \right.$$
$$\left. - \left\{\hat{c}_2 + \hat{m}_{\mathrm{SBLL},52}(X_{t-4})\right\}(T_{1t} - T_{2t})\right]^2 = 0.001818,$$

which is about 60% of the corresponding ASPE (0.003097) for model (30). The in-sample averaged squared estimation error (ASE) for model (31) is 5.2399×10^{-5}, which is about 68% of the in-sample ASE (7.6959×10^{-5}) for model (30).

In model (31), $\hat{c}_1 + \hat{m}_{\mathrm{SBLL},41}(X_{t-3})$ estimates the TFP growth rate, which is shown as a function of X_{t-3} in Figure 2. It is obvious that the effect of X_{t-3} is positive when $X_{t-3} \le 0.02$, but negative when $X_{t-3} > 0.02$. On average, the higher R&D investment spending causes faster GDP growth. However, overspending on R&D often leads to high losses (Culpepper, 2004; Tokic, 2003).

We have also computed the average contribution of R&D to GDP growth for 1964–2001, which is about 40%. From the GDP and estimated TFP growth rates,

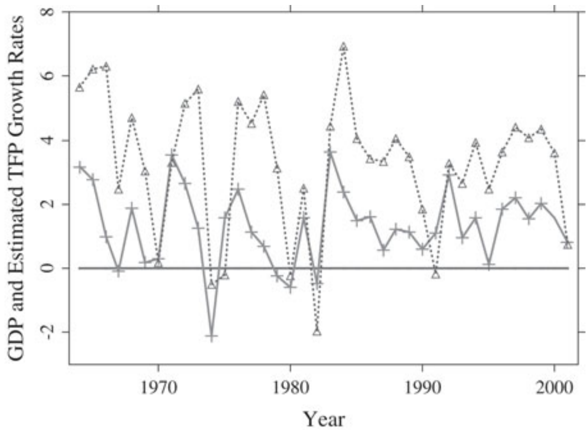

FIGURE 3. Estimation of function $c_1 + m_{\text{SBLL},41}\left(X_{t-3}\right)$: dotted line = GDP growth rate; solid line = $\hat{c}_1 + \hat{m}_{\text{SBLL},41}\left(X_{t-3}\right)$.

shown in Figure 3, it is obvious that TFP growth is highly correlated to GDP growth. For more details, see Arnold (2005).

REFERENCES

Arnold, R. (2005) *R&D and Productivity Growth: A Background Paper*. Congressional Budget Office.

Bosq, D. (1998) *Nonparametric Statistics for Stochastic Processes: Estimation and Prediction*. Springer-Verlag.

Cai, Z., J. Fan, & Q.W. Yao (2000) Functional-coefficient regression models for nonlinear time series. *Journal of the American Statistical Association* 95, 941–956.

Chen, R. & R.S. Tsay (1993a) Nonlinear additive ARX models. *Journal of the American Statistical Association* 88, 955–967.

Chen, R. & R.S. Tsay (1993b) Functional-coefficient autoregressive models. *Journal of the American Statistical Association* 88, 298–308.

Cobb C.W. & P.H. Douglas (1928) A theory of production. *American Economic Review* 18, 139–165.

Culpepper, W.L. (2004) *High R&D Spending Fuels Revenue Growth Not Profits*. Available for downloading at *http://www.culpepper.com/eBulletin/2004/ AugustRatiosArticle.asp.*

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

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

Hastie, T.J. & R.J. Tibshirani (1993) Varying-coefficient models. *Journal of the Royal Statistical Society Series B* 55, 757–796.

Hengartner, N.W. & S. Sperlich (2005) Rate optimal estimation with the integration method in the presence of many covariates. *Journal of Multivariate Analysis* 95, 246–272.

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

Huang, J.Z. (1998a) Projection estimation in multiple regression with application to functional ANOVA models. *Annals of Statistics* 26, 242–272.

Huang, J.Z. (1998b) Functional ANOVA models for generalized regression. *Journal of Multivariate Analysis* 67, 49–71.

Huang, J.Z. & H. Shen (2004) Functional coefficient regression models for non-linear time series: A polynomial spline approach. *Scandinavian Journal of Statistics* 31, 515–534.

Huang, J.Z., C.O. Wu, & L. Zhou (2002) Varying-coefficient models and basis function approximations for the analysis of repeated measurements. *Biometrika* 89, 111–128.
Huang, J.Z. & L. Yang (2004) Identification of nonlinear additive autoregressive models. *Journal of the Royal Statistical Society Series B* 66, 463–477.
Li, Q. & J.S. Racine (2007) *Nonparametric Econometrics: Theory and Practice*. Princeton University Press.
Linton, O.B. (1997) Efficient estimation of additive nonparametric regression models. *Biometrika* 84, 469–473.
Liu, R. & L. Yang (2008) Spline-backfitted kernel smoothing of additive coefficient model. Available at http://www.msu.edu/˜yangli/sbkaddcoefffull.pdf.
Mammen, E., O.B. Linton, & J.P. Nielsen (1999) The existence and asymptotic properties of a backfitting projection algorithm under weak conditions. *Annals of Statistics* 27, 1443–1490.
Nielsen, J.P. & S. Sperlich (2005) Smooth backfitting in practice. *Journal of the Royal Statistical Society Series B* 67, 43–61.
Rodríguez-Póo, J.M., S. Sperlich, & P. Vieu (2003) Semiparametric estimation of separable models with possibly limited dependent variables. *Econometric Theory* 19, 1008–1039.
Solow, R.M. (1957) Technical change and the aggregate production function. The *Review of Economics and Statistics* 39, 312–320.
Sperlich, S., D. Tjøstheim, & L. Yang (2002) Nonparametric estimation and testing of interaction in additive models. *Econometric Theory* 18, 197–251.
Stone, C. J. (1985) Additive regression and other nonparametric models. *Annals of Statistics* 13, 689–705.
Tokic, D. (2003) How efficient were R&D and advertising investments for internet firms before the bubble burst? A DEA approach. *Credit and Financial Management Review* 9, 39–51.
Wang, L. & L. Yang (2007) Spline-backfitted kernel smoothing of nonlinear additive autoregression model. *Annals of Statistics* 35, 2474–2503.
Xue, L. & L. Yang (2006a) Estimation of semiparametric additive coefficient model. *Journal of Statistical Planning and Inference* 136, 2506–2534.
Xue, L. & L. Yang (2006b) Additive coefficient modelling via polynomial spline. *Statistica Sinica* 16, 1423–1446.
Yang, L., W. Härdle, & J.P. Nielsen (1999) Nonparametric autoregression with multiplicative volatility and additive mean. *Journal of Time Series Analysis* 20, 579–604.
Yang, L., B.U. Park, L. Xue, & W. Härdle (2006) Estimation and testing of varying coefficients in additive models with marginal integration. *Journal of the American Statistical Association* 101, 1212–1227.
Yang, L., S. Sperlich, & W. Härdle (2003) Derivative estimation and testing in generalized additive models. *Journal of Statistical Planning and Inference* 115, 521–542.

APPENDIX

A. 1. Preliminaries. In the proofs that follow, we use U and u to denote sequences of random variables that are uniformly O and o of certain order. In several places, we have omitted details and referred to Liu and Yang (2008).

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

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

where

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

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

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

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

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

LEMMA A.3. *Under Assumptions 1 and 8, (i) there exist constants* $c_f, C_f, c_0(f)$ *and* $C_0(f)$ *depending on the marginal densities* $f_\alpha(x_\alpha), 1 \le \alpha \le d_2$, *such that* $c_f H \le c_{J,\alpha} \le C_f H$ *and* $c_0(f) H \le \|b_{J,\alpha}\|_2^2 \le C_0(f) H$; *and (ii) uniformly for* $J, J' = 1, \ldots, N+1$,

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

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

LEMMA A.4. *Under Assumption 2, for* $\mathbf{V_T}$ *defined in (A.8) and* $\mathbf{S_T} = \mathbf{V_T}^{-1}$,

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

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

Proof. See Liu and Yang (2008). ∎

LEMMA A.5 (de Boor, 2001, p. 149). *There exists a constant* $C_\infty > 0$ *such that for any* $m \in C^1[0,1]$ *with* $m' \in \mathrm{Lip}([0,1], C_\infty)$, *there is a function* $g \in G^{(0)}[0,1]$ *such that* $\|g - m\|_\infty \le C_\infty \|m'\|_\infty H^2$.

Lemma A.5 and Assumption 3 ensure the existence of functions $g_{\alpha l} \in G^{(0)}[0,1]$ such that

$$\|g_{\alpha l} - m_{\alpha l}\|_\infty \le C_\infty \|m'_{\alpha l}\|_\infty H^2, \alpha = 1, \ldots, d_2, l = 1, \ldots, d_1. \tag{A.1}$$

A. 2. Estimation of Constants. To closely examine terms $\tilde{\varepsilon}_l(\mathbf{x})$ and $\tilde{\varepsilon}_{\alpha l}(x_\alpha)$, we denote the following vector of coefficients:

$$\tilde{\mathbf{a}} = \left\{\tilde{a}_{01}, \tilde{a}_{1,1,1}, \ldots, \tilde{a}_{N+1,d_2,1}, \tilde{a}_{02}, \tilde{a}_{1,1,2}, \ldots, \tilde{a}_{N+1,d_2,2}, \ldots, \tilde{a}_{0d_1}, \right.$$
$$\left. \times \tilde{a}_{1,1,d_1}, \ldots, \tilde{a}_{N+1,d_2,d_1}\right\}^T, \tag{A.2}$$

such that the noise term $\tilde{\varepsilon}_l(\mathbf{x})$ in (19) is expressed as

$$\left(\mathbf{P}_{n,l}\mathbf{E}\right)(\mathbf{x}) = \tilde{\varepsilon}_l(\mathbf{x}) = \tilde{a}_{0l} + \sum_{\alpha=1}^{d_2}\sum_{J=1}^{N+1} \tilde{a}_{J,\alpha,l} B_{J,\alpha}(x_\alpha). \tag{A.3}$$

Equation (A.3) implies that $\tilde{\mathbf{a}} = \left(\mathbf{D}^T\mathbf{D}\right)^{-1}\mathbf{D}^T\mathbf{E}$, where

$$\mathbf{D} = \{\mathbf{D}(\mathbf{X}_1, \mathbf{T}_1), \ldots, \mathbf{D}(\mathbf{X}_n, \mathbf{T}_n)\}^T = \{\mathbf{T}_1 \otimes \mathbf{B}(\mathbf{X}_1), \ldots, \mathbf{T}_n \otimes \mathbf{B}(\mathbf{X}_n)\}^T, \tag{A.4}$$

$$\mathbf{B}(\mathbf{x}) = \left\{1, B_{1,1}(x_1), \ldots, B_{N+1,d_2}\left(x_{d_2}\right)\right\}^T, \mathbf{t} = \left\{t_1, \ldots, t_{d_1}\right\}^T. \tag{A.5}$$

Note that $\tilde{\mathbf{a}}$ given in (A.2) can be rewritten as

$$\tilde{\mathbf{a}} = \left(\frac{1}{n}\mathbf{D}^T\mathbf{D}\right)^{-1}\left(\frac{1}{n}\mathbf{D}^T\mathbf{E}\right) = \left(\mathbf{V}_\mathbf{T} + \mathbf{V}_\mathbf{T}^*\right)^{-1}\left(\frac{1}{n}\mathbf{D}^T\mathbf{E}\right), \tag{A.6}$$

where, by (A.4),

$$\mathbf{D}^T\mathbf{D} = \sum_{i=1}^n \left[\left(\mathbf{T}_i\mathbf{T}_i^T\right) \otimes \left\{\mathbf{B}(\mathbf{X}_i)\mathbf{B}(\mathbf{X}_i)^T\right\}\right], \qquad \mathbf{D}^T\mathbf{E} = \sum_{i=1}^n \left[\{\mathbf{T}_i \otimes \mathbf{B}(\mathbf{X}_i)\}\sigma(\mathbf{X}_i, \mathbf{T}_i)\varepsilon_i\right], \tag{A.7}$$

and $\mathbf{V}_\mathbf{T}^*$ is the difference between empirical and theoretical inner product matrices; i.e.,

$$\mathbf{V}_\mathbf{T} = \mathrm{E}\left[\left(\mathbf{T}\mathbf{T}^T\right) \otimes \left\{\mathbf{B}(\mathbf{X})\mathbf{B}(\mathbf{X})^T\right\}\right] = \mathrm{E}\left[\mathbf{Q}(\mathbf{X}) \otimes \left\{\mathbf{B}(\mathbf{X})\mathbf{B}(\mathbf{X})^T\right\}\right], \tag{A.8}$$

$$\mathbf{V}_\mathbf{T}^* = \frac{1}{n}\sum_{i=1}^n \left[\left(\mathbf{T}_i\mathbf{T}_i^T\right) \otimes \left\{\mathbf{B}(\mathbf{X}_i)\mathbf{B}(\mathbf{X}_i)^T\right\}\right] - \mathrm{E}\left[\mathbf{Q}(\mathbf{X}) \otimes \left\{\mathbf{B}(\mathbf{X})\mathbf{B}(\mathbf{X})^T\right\}\right].$$

Now define $\hat{\mathbf{a}} = \left\{\hat{a}_{01}, \hat{a}_{1,1,1}, \ldots, \hat{a}_{N,d_2,1}, \hat{a}_{02}, \hat{a}_{1,1,2}, \ldots, \hat{a}_{N,d_2,2}, \ldots, \hat{a}_{0d_1}, \hat{a}_{1,1,d_1}, \ldots, \hat{a}_{N,d_2,d_1}\right\}^T$ by replacing $\left(\mathbf{V}_\mathbf{T} + \mathbf{V}_\mathbf{T}^*\right)^{-1}$ with $\mathbf{V}_\mathbf{T}^{-1} = \mathbf{S}_\mathbf{T}$ in the above formula; that is,

$$\hat{\mathbf{a}} = V_\mathbf{T}^{-1}\left(n^{-1}\mathbf{D}^T\mathbf{E}\right) = \mathbf{S}_\mathbf{T}\left(n^{-1}\mathbf{D}^T\mathbf{E}\right). \tag{A.9}$$

LEMMA A.6. *Under Assumptions 1–3 and 5–8, as $n \to \infty$,*

$$\|\tilde{\mathbf{a}}\| = O_p\left(n^{-1/2}N^{1/2}\log n\right), \tag{A.10}$$

$$\|\tilde{\mathbf{a}} - \hat{a}\| = O_p\left(n^{-1}N^{3/2}\log^2 n\right), \quad \text{and} \tag{A.11}$$

$$\|\hat{\mathbf{a}}\| = O_p\left(n^{-1/2}N^{1/2}\log n\right). \tag{A.12}$$

Proof. By definition, $\tilde{\mathbf{a}}^T\mathbf{D}^T\mathbf{D}\tilde{\mathbf{a}}=\tilde{\mathbf{a}}^T\mathbf{D}^T\mathbf{D}\left(\mathbf{D}^T\mathbf{D}\right)^{-1}\mathbf{D}^T\mathbf{E}=\tilde{\mathbf{a}}\mathbf{D}^T\mathbf{E}$. Using (A.6), one has

$$\left\|\mathbf{D}\tilde{\mathbf{a}}\right\|_{2,n}^2=n^{-1}\tilde{\mathbf{a}}^T\mathbf{D}^T\mathbf{D}\tilde{\mathbf{a}}=n^{-1}\tilde{\mathbf{a}}^T\mathbf{D}^{\mathbf{T}}\mathbf{E}\le\left\|\tilde{\mathbf{a}}\right\|\left\|n^{-1}\mathbf{D}^T\mathbf{E}\right\|. \tag{A.13}$$

According to Lemma A.2,

$$c_0\left\|\tilde{\mathbf{a}}\right\|^2=c_0\sum_l\left(a_{0l}^2+\sum_{J,\alpha,l}a_{J,\alpha,l}^2\right)\le\left\|\sum_l\left(a_{0l}+\sum_{J,\alpha,l}a_{J,\alpha,l}B_{J,\alpha}\right)t_l\right\|_{2,n}^2=\left\|\mathbf{D}\tilde{\mathbf{a}}\right\|_{2,n}^2. \tag{A.14}$$

So $\left\|\tilde{\mathbf{a}}\right\|$ is bounded by $c_0^{-1}\left\|n^{-1}\mathbf{D}^T\mathbf{E}\right\|$. Bernstein's inequality and truncation entail that $\left\|n^{-1}\mathbf{D}^T\mathbf{E}\right\|^2=O_p\left\{(\log n)^2N/n\right\}$, so (A.10) follows from (A.13) and (A.14). According to (A.6) and (A.9), one has $\mathbf{V_T}\,\hat{\mathbf{a}}=\left(\mathbf{V_T}+\mathbf{V_T^*}\right)\tilde{\mathbf{a}}$, which implies that $\mathbf{V_T^*}\tilde{\mathbf{a}}=\mathbf{V_T}\left(\hat{\mathbf{a}}-\tilde{a}\right)$.

One obtains from (A.22) and (A.23) $\left\|\mathbf{V_T}\left(\hat{\mathbf{a}}-\tilde{a}\right)\right\|=\left\|\mathbf{V_T^*}\tilde{\mathbf{a}}\right\|\le O_p\left(n^{-1/2}H^{-1}\log n\right)$ $\left\|\tilde{\mathbf{a}}\right\|$. By (A.10), one has $\left\|\mathbf{V_T}\left(\hat{\mathbf{a}}-\tilde{a}\right)\right\|\le O_p\left\{(\log n)^2n^{-1}N^{3/2}\right\}$. Thus, according to Lemma A.4, one has $\left\|\hat{\mathbf{a}}-\tilde{a}\right\|=O_p\left(n^{-1}N^{3/2}\log^2 n\right)$, which is (A.11). Then (A.12) follows (A.10) and (A.11). ■

LEMMA A.7. *Under Assumptions 1–3 and 5–8, as n → ∞,*

$$\sup_{1\le l'\le d}\left|\frac{1}{n}\sum_{i=1}^n T_{il'}\sum_{l=1}^{d_1}\sum_{\alpha=1}^{d_2}\tilde{\varepsilon}_{\alpha l}T_{il}\right|=O_p\left(n^{-1/2}\right). \tag{A.15}$$

Proof. According to (19) and (A.3), one has

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^n T_{il'}\sum_{l=1}^{d_1}\sum_{\alpha=1}^{d_2}\tilde{\varepsilon}_{\alpha l}T_{il}&=\frac{1}{n}\sum_{i=1}^n T_{il'}\sum_{l=1}^{d_1}\sum_{\alpha=1}^{d_2}\sum_{J=1}^{N+1}\tilde{a}_{J,\alpha,l}B_{J,\alpha}\left(X_{i\alpha}\right)T_{il}\\&=\sum_{J,\alpha,l}\tilde{a}_{J,\alpha,l}\frac{1}{n}\sum_{i=1}^n T_{il'}B_{J,\alpha}\left(X_{i\alpha}\right)T_{il}=I_{l',1}+I_{l',2}+II_{l'},\end{aligned}$$

where

$$II_{l'}=\sum_{J,\alpha,l}\left(\tilde{a}_{J,\alpha,\alpha}-\hat{a}_{J,\alpha,l}\right)\frac{1}{n}\sum_{i=1}^n T_{il'}B_{J,\alpha}\left(X_{i\alpha}\right)T_{il},$$

$$I_{l',1}=\sum_{J,\alpha,l}\hat{a}_{J,\alpha,l}\left\{\frac{1}{n}\sum_{i=1}^n T_{il'}B_{J,\alpha}\left(X_{i\alpha}\right)T_{il}-\mathrm{E}T_{l'}B_{J,\alpha}\left(X_\alpha\right)T_l\right\},$$

$$\begin{aligned}\left|I_{l',1}\right|&\le\left\|\hat{\mathbf{a}}\right\|\sqrt{(N+1)d_1d_2}\sup_{J,\alpha,l}\left|\frac{1}{n}\sum_{i=1}^n T_{il'}B_{J,\alpha}\left(X_{i\alpha}\right)T_{il}-\mathrm{E}T_{l'}B_{J,\alpha}\left(X_\alpha\right)T_l\right|\\&=O_p\left(n^{-1}N\log^2 n\right),\quad\text{and}\end{aligned}$$

$$I_{l',2}=\sum_{J,\alpha,l}\hat{a}_{J,\alpha,l}\mathrm{E}T_{il'}B_{J,\alpha}\left(X_{i\alpha}\right)T_{il}=\left(\mathrm{E}T_{l'}B_{J,\alpha}\left(X_\alpha\right)T_l\right)_{J,\alpha,l}^T\mathbf{V_T^{-1}}\left(n^{-1}\mathbf{D}^T\mathbf{E}\right).$$

Direct computation (see Liu and Yang, 2008, for details) yields that $\operatorname{var}\left(I_{l',2}\right) = O\left(n^{-1}\right)$, and therefore, $I_{l',2} = O_p\left(n^{-1/2}\right)$. So

$$\left|I_{l',1}\right| + \left|I_{l',2}\right| = O_p\left(n^{-1/2}\right). \tag{A.16}$$

Next, by applying Bernstein's inequality with truncation technique,

$$\sup_{J,\alpha,l}\left|\frac{1}{n}\sum_{i=1}^{n} T_{il'}B_{J,\alpha}\left(X_{i\alpha}\right)T_{il} - \mathrm{E}T_{l'}B_{J,\alpha}\left(X_{\alpha}\right)T_{l}\right| = O_p\left(n^{-1/2}\log n\right).$$

Thus $\sup_{J,\alpha,l}\left|\frac{1}{n}\sum_{i=1}^{n} T_{il'}B_{J,\alpha}\left(X_{i\alpha}\right)T_{il}\right|$ is bounded by

$$\sup_{J,\alpha,l}\left|\frac{1}{n}\sum_{i=1}^{n} T_{il'}B_{J,\alpha}\left(X_{i\alpha}\right)T_{il} - \mathrm{E}T_{l'}B_{J,\alpha}\left(X_{i}\right)T_{l}\right| + \left|\mathrm{E}T_{l'}B_{J,\alpha}\left(X_{\alpha}\right)T_{l}\right| = O\left(H^{1/2}\right).$$

Then

$$\left|II_{l'}\right| \le \left\|\hat{\mathbf{a}} - \tilde{a}\right\|\sqrt{(N+1)d_1d_2}\sup\left|\frac{1}{n}\sum_{i=1}^{n} T_{il'}B_{J,\alpha}\left(X_{i\alpha}\right)T_{il}\right| = O_p\left(n^{-1}N^{3/2}\log^2 n\right). \tag{A.17}$$

Now (A.15) follows from (A.16) and (A.17). The lemma is proved. ■

LEMMA A.8. *Under Assumptions 1–5 and 8, as $n \to \infty$,*

$$n^{-1}\sum_{i=1}^{n}\left[\sum_{l=1}^{d_1}\sum_{\alpha=1}^{d_2}\left\{\tilde{m}_{\alpha l}\left(X_{i\alpha}\right) - m_{\alpha l}\left(X_{i\alpha}\right)\right\}T_{il}\right]^2 = O_p\left(n^{-1}\right). \tag{A.18}$$

Proof. According to (A.1), there exists $g_{\alpha l} \in G^{(0)}[0,1]$ such that $\|g_{\alpha l} - m_{\alpha l}\|_\infty = O\left(H^2\right) = O\left(n^{-1/2}\right)$. According to Theorem 1.7 of Bosq (1998, p. 36), $n^{-1/2}\sum_{i=1}^{n}\left(T_{il}^2 - \mathrm{E}T_l^2\right) \Longrightarrow N\left(0,\sigma^2\right)$, where $\sigma^2 = \sum_{i=-\infty}^{\infty}\operatorname{Cov}\left(T_{0l}^2, T_{il}^2\right) < \infty$ by applying Davydov's Inequality (Bosq, 1998, p. 21, eqn. (1.10)), then $n^{-1}\sum_{i=1}^{n} T_{il}^2 = \mathrm{E}T_l^2 + O_p\left(n^{-1/2}\right) = O_p(1)$. So

$$n^{-1}\sum_{i=1}^{n}\left[\sum_{l=1}^{d_1}\sum_{\alpha=1}^{d_2}\left\{\tilde{m}_{\alpha l}\left(X_{i\alpha}\right) - m_{\alpha l}\left(X_{i\alpha}\right)\right\}T_{il}\right]^2$$

$$\le n^{-1}\sum_{i=1}^{n}\left[\sum_{l=1}^{d_1}\sum_{\alpha=1}^{d_2}\|\tilde{m}_{\alpha l} - m_{\alpha l}\|_\infty T_{il}\right]^2 \le n^{-1}\sum_{i=1}^{n}\left[\sum_{l=1}^{d_1}\sum_{\alpha=1}^{d_2}\|g_{\alpha l} - m_{\alpha l}\|_\infty T_{il}\right]^2$$

$$= O\left(n^{-1}\right)\left(n^{-1}\sum_{i=1}^{n} T_{il}^2\right) = O_p\left(n^{-1}\right).$$ ■

Proof of Propositions 1 and 2. According to (9), $\tilde{m}_0 - m_0 = (\mathbf{C}_K^T\mathbf{C}_K)^{-1}\mathbf{C}_K^T(\mathbf{Y}_c - m_0\mathbf{T}) = (\frac{1}{n}\mathbf{C}_K^T\mathbf{C}_K)^{-1}\frac{1}{n}\mathbf{C}_K^T\sigma(\mathbf{X}_i,\mathbf{T}_i)\varepsilon_i$. We know that $\frac{1}{n}\mathbf{C}_K^T\mathbf{C}_K = \left(\frac{1}{n}\sum_{i=1}^n T_{il}T_{il'}\right)_{l,l'=1}^{d_1}$. Then according to Theorem 1.7 of Bosq (1998, p. 36), one has $n^{-1/2}\sum_{i=1}^n\{T_{il}T_{il'} - \mathrm{E}T_lT_{l'}\} \Longrightarrow N\left(0,\sigma^2\right)$, where $\sigma^2 = \sum_{i=-\infty}^{\infty}\mathrm{Cov}\left(T_{0l}T_{0l'}, T_{il}T_{il'}\right) < \infty$. Therefore $\frac{1}{n}\mathbf{C}_K^T\mathbf{C}_K = \left(\mathrm{E}T_lT_{l'}\right)_{l,l'=1}^{d_1} + O_p\left(n^{-1/2}\right)$. Similarly, $\frac{1}{n}\mathbf{C}_K^T\sigma(\mathbf{X}_i,\mathbf{T}_i)\varepsilon_i = O_p\left(n^{-1/2}\right)$, implying $\sup_{1\le l\le d_1}|\tilde{m}_{0l} - m_{0l}| = O_p\left(n^{-1/2}\right)$, which completes the proof of Proposition 1.

Next, according to (9) and (12),

$$\hat{m}_0 - \tilde{m}_0 = (\mathbf{C}_K^T\mathbf{C}_K)^{-1}\mathbf{C}_K^T\left(\hat{\mathbf{Y}}_c - \mathbf{Y}_c\right)$$

$$= (\mathbf{C}_K^T\mathbf{C}_K)^{-1}\mathbf{C}_K^T\left[\sum_{\alpha=1}^{d_1}\sum_{\alpha=1}^{d_2}\left\{\hat{m}_{\alpha l}(X_{i\alpha}) - m_{\alpha l}(X_{i\alpha})\right\}T_{il}\right]_{i=1}^n$$

$$= (\mathbf{C}_K^T\mathbf{C}_K)^{-1}\mathbf{C}_K^T\left[\sum_{l=1}^{d_1}\sum_{\alpha=1}^{d_2}\left\{\hat{m}_{\alpha l}(X_{i\alpha}) - \tilde{m}_{\alpha l}(X_{i\alpha})\right.\right.$$

$$\left.\left.+\tilde{m}_{\alpha l}(X_{i\alpha}) - m_{\alpha l}(X_{i\alpha})\right\}T_{il}\right]_{i=1}^n$$

$$= \left(\frac{1}{n}\mathbf{C}_K^T\mathbf{C}_K\right)^{-1}\frac{1}{n}\mathbf{C}_K^T\left[\left(\sum_{l=1}^{d_1}\sum_{\alpha=1}^{d_2}\tilde{\varepsilon}_{\alpha l}T_{il}\right)_{i=1}^n\right.$$

$$\left.+\left[\left\{\sum_{l=1}^{d_1}\sum_{\alpha=1}^{d_2}\{\tilde{m}_{\alpha l}(X_{i\alpha}) - m_{\alpha l}(X_{i\alpha})\}T_{il}\right\}\right]_{i=1}^n\right].$$

One has

$$\frac{1}{n}\sum_{i=1}^n T_{il'}\left[\sum_{l=1}^{d_1}\sum_{\alpha=1}^{d_2}(\tilde{m}_{\alpha l}(X_{i\alpha}) - m_{\alpha l}(X_{i\alpha}))T_{il}\right]$$

$$\le \left(\frac{1}{n}\sum_{i=1}^n T^2\right)_{il'}^{1/2}\left\{\frac{1}{n}\sum_{i=1}^n\left[\sum_{l=1}^{d_1}\sum_{\alpha=1}^{d_2}(\tilde{m}_{\alpha l}(X_{i\alpha}) - m_{\alpha l}(X_{i\alpha}))T_{il}\right]^2\right\}^{1/2}$$

$$\le O_p(1)\,O_p\left(n^{-1/2}\right) = O_p\left(n^{-1/2}\right) \tag{A.19}$$

by Lemma A.8. Then Proposition 2 follows from (A.15) and (A.19). ■

A. 3. Estimation of Function Components. Define

$$A_{n,1} = \sup_{J,\alpha} \left| \left\langle 1, B_{J,\alpha} \right\rangle_{2,n} - \left\langle 1, B_{J,\alpha} \right\rangle_2 \right| = \sup_{J,\alpha} \left| n^{-1} \sum_{i=1}^{n} B_{J,\alpha}\left(X_{i\alpha}\right) \right|,$$

$$A_{n,2} = \sup_{J,J',\alpha,l,l'} \left| \left\langle B_{J,\alpha} T_{il}, B_{J',\alpha} T_{il'} \right\rangle_{2,n} - \left\langle B_{J,\alpha} T_{il}, B_{J',\alpha} T_{il'} \right\rangle_2 \right|, \quad \text{and}$$

$$A_{n,3} = \sup_{J,J',\alpha \neq \alpha',l,l'} \left| \left\langle B_{J,\alpha} T_{il}, B_{J',\alpha'} T_{il'} \right\rangle_{2,n} - \left\langle B_{J,\alpha} T_{il}, B_{J',\alpha'} T_{il'} \right\rangle_2 \right|. \tag{A.20}$$

LEMMA A.9. *Under Assumptions 1–3 and 8, as $n \to \infty$,*

$$A_{n,1} = O_p\left(n^{-1/2} \log n\right), \tag{A.21}$$

$$A_{n,2} = O_p\left(n^{-1/2} H^{-1/2} \log n\right), \quad \text{and} \tag{A.22}$$

$$A_{n,3} = O_p\left(n^{-1/2} \log n\right). \tag{A.23}$$

Proof. See Liu and Yang (2008). ■

LEMMA A.10. *Under Assumptions 1–3, 5, and 7–8, as $n \to \infty$,*

$$\sup_{x_1 \in [0,1]} \sup_{1 \le J \le N+1} \sup_{2 \le \alpha \le d_2} \sup_{1 \le l,l' \le d_1} \left| \mu_{\omega_{J,\alpha,l,l'}}\left(x_1\right) \right| = O\left(H^{1/2}\right), \tag{A.24}$$

$$\sup_{x_1 \in [0,1]} \sup_{1 \le J \le N+1} \sup_{2 \le \alpha \le d_2} \sup_{1 \le l,l' \le d_1} \left| n^{-1} \sum_{i=1}^{n} \left\{ \omega_{J,\alpha,l,l'}\left(\mathbf{X}_i, x_1\right) - \mu_{\omega_{J,\alpha,l,l'}}\left(x_1\right) \right\} \right|$$

$$= O_p\left(\log n / \sqrt{nh}\right), \tag{A.25}$$

where $\omega_{J,\alpha,l,l'}\left(\mathbf{X}_i, x_1\right)$ and $\mu_{\omega_{J,\alpha,l,l'}}\left(x_1\right)$ defined in (28), hence

$$\sup_{x_1 \in [0,1]} \sup_{1 \le J \le N+1} \sup_{2 \le \alpha \le d_2} \sup_{1 \le l,l' \le d_1} \left| n^{-1} \sum_{i=1}^{n} \omega_{J,\alpha,l,l'}\left(\mathbf{X}_i, x_1\right) \right| = O_p\left(H^{1/2}\right). \tag{A.26}$$

Proof. See Liu and Yang (2008). ■

In the following, we define a noise term analogous to the formula for $\Psi^{(2)}_{v,l'}\left(x_1\right)$ in (29) by replacing $\tilde{\mathbf{a}}$ in (A.2) with $\hat{\mathbf{a}}$ in (A.9)

$$\hat{\Psi}^{(2)}_{v,l'}\left(x_1\right) = n^{-1} \sum_{i=1}^{n} \sum_{l=1}^{d_1} \sum_{J=1}^{N+1} \sum_{\alpha=2}^{d_2} \hat{a}_{J,\alpha,l} \omega_{J,\alpha,l,l'}\left(\mathbf{X}_i, x_1\right). \tag{A.27}$$

LEMMA A.11. *Under Assumptions 1–3, 5, and 8, as $n \to \infty$,* $\sup_{1\le l'\le d_1}\sup_{x_1\in[0,1]}\left|\Psi^{(2)}_{\upsilon,l'}(x_1)-\hat{\Psi}^{(2)}_{\upsilon,l'}(x_1)\right| = O_p\left(H^2\right)$.

Proof. According to (27) and (A.27), one has

$$\left|\Psi^{(2)}_{\upsilon,l'}(x_1)-\hat{\Psi}^{(2)}_{\upsilon,l'}(x_1)\right| = \left|\sum_{l=1}^{d_1}\sum_{J=1}^{N+1}\sum_{\alpha=2}^{d_2}\left(\tilde{a}_{J,\alpha,l}-\hat{a}_{J,\alpha,l}\right)\frac{1}{n}\sum_{i=1}^{n}\omega_{J,\alpha,l}(\mathbf{X}_i,x_1)\,T_{i1}\right|.$$

According to (A.26) and (A.11), Cauchy-Schwartz inequality implies that

$$\sup_{1\le l'\le d_1}\sup_{x\in[0,1]}\left|\Psi^{(2)}_{\upsilon,l'}(x_1)-\hat{\Psi}^{(2)}_{\upsilon,l'}(x_1)\right|$$
$$\le \sqrt{N+1}\,O_p\left(\frac{(\log n)^2}{nH^{3/2}}\right)O_p\left(H^{1/2}\right) = O_p\left(\frac{(\log n)^2}{nH^{3/2}}\right).$$

Therefore, the lemma follows. ■

LEMMA A.12. *Under Assumptions 1–3, 5, and 8, as $n \to \infty$,*

$$\sup_{1\le l'\le d_1}\sup_{x_1\in[0,1]}\left|\hat{\Psi}^{(2)}_{\upsilon,l'}(x_1)\right|$$
$$= \sup_{1\le l'\le d_1}\sup_{x_1\in[0,1]}\left|n^{-1}\sum_{i=1}^{n}\sum_{l=1}^{d_1}\sum_{J=1}^{N+1}\sum_{\alpha=2}^{d_2}\hat{a}_{J,\alpha,l}\,\omega_{J,\alpha,l,l'}(\mathbf{X}_i,x_1)\right| = O_p\left(n^{-2/5}\right).$$

Proof. See Liu and Yang (2008). ■

Proof of Proposition 3. (A.1) implies that

$$\begin{aligned}\left|\mathrm{E}_n g_{\text{-}1,l}\left(\mathbf{X}_{i,\text{-}1}\right)\right| &\le \left|\mathrm{E}_n g_{\text{-}1,l}\left(\mathbf{X}_{i,\text{-}1}\right)-\mathrm{E}_n m_{\text{-}1,l}\left(\mathbf{X}_{i,\text{-}1}\right)\right| + \left|\mathrm{E}_n m_{\text{-}1,l}\left(\mathbf{X}_{i,\text{-}1}\right)\right| \\ &\le C_\infty\left(d_2-1\right)\sup_{2\le\alpha\le d_2}\left\|m'_{\alpha l}\right\|_\infty H^2 + O_p\left(n^{-1/2}\right).\end{aligned} \tag{A.28}$$

By definition (24), $\sup_{x_1\in[0,1]}\left|\Psi_{b,l'}(x_1)\right| \le R_1+R_2+R_3$ where

$$R_1 = \sup_{x_1\in[0,1]}\left|\frac{1}{n}\sum_{i=1}^{n}K_h\left(X_{i1}-x_1\right)\sum_{l=1}^{d_1}\left\{m_{\text{-}1,l}\left(\mathbf{X}_{i,\text{-}1}\right)-g_{\text{-}1,l}\left(\mathbf{X}_{i,\text{-}1}\right)\right\}T_{il}T_{il'}\right|,$$

$$R_2 = \sup_{x_1\in[0,1]}\left|\frac{1}{n}\sum_{i=1}^{n}K_h\left(X_{i1}-x_1\right)\sum_{l=1}^{d_1}\left\{g_{\text{-}1,l}\left(\mathbf{X}_{i,\text{-}1}\right)\right.\right.$$
$$\left.\left.-\mathrm{E}_n g_{\text{-}1,l}\left(\mathbf{X}_{i,\text{-}1}\right)-\tilde{m}_{\text{-}1,l}\left(\mathbf{X}_{i,\text{-}1}\right)\right\}T_{il}T_{il'}\right|, \quad \text{and}$$

$$R_3 = \sup_{x_1\in[0,1]}\left|\frac{1}{n}\sum_{i=1}^{n}K_h\left(X_{i1}-x_1\right)\sum_{l=1}^{d_1}\mathrm{E}_n g_{\text{-}1,l}\left(\mathbf{X}_{i,\text{-}1}\right)T_{il}T_{il'}\right|.$$

For R_1, using (A.1), one has

$$R_1 \leq C_\infty (d_2 - 1) \sup_{2 \leq \alpha \leq d_2} \left\| m'_{\alpha l} \right\|_\infty H^2 \sum_{l=1}^{d_1} \frac{1}{n} \sum_{i=1}^{n} \left| T_{il} T_{il'} \right|$$

$$= O_p\left(H^2\right) \left\{ \sum_{l=1}^{d_1} \mathrm{E} \left| T_{il} T_{il'} \right| + O_p\left(n^{-1/2}\right) \right\} = O_p\left(H^2\right). \tag{A.29}$$

To bound R_2, denote the empirically centered spline basis as $B^*_{J,\alpha}(X_{i\alpha}) = B_{J,\alpha}(X_{i\alpha}) - \mathrm{E}_n B_{J,\alpha}(X_{i\alpha})$, $1 \leq J \leq N+1$, $1 \leq \alpha \leq d_2$. Then one can write for some $\left(\tilde{a}^*_{\alpha,l}, \tilde{a}^*_{J,\alpha,l}\right)_{J,\alpha,l}$,

$$\tilde{m}_l(\mathbf{x}) - m_{0l} - \sum_{\alpha=1}^{d_2} g_{\alpha l}(\mathbf{x}) + \sum_{\alpha=1}^{d_2} \mathrm{E}_n g_{\alpha l}(\mathbf{x}) = \tilde{a}^*_{\alpha,l} + \sum_{\alpha=1}^{d_2} \sum_{J=1}^{N+1} \tilde{a}^*_{J,\alpha,l} B^*_{J,\alpha}(x_\alpha).$$

Thus,

$$R_2 = \sup_{x_1 \in [0,1]} \left| n^{-1} \sum_{i=1}^{n} K_h(X_{i1} - x_1) \sum_{l=1}^{d_1} \sum_{\alpha=2}^{d_2} \sum_{J=1}^{N+1} \tilde{a}^*_{J,\alpha,l} B^*_{J,\alpha}(X_{i\alpha}) T_{il} T_{il'} \right|$$

$$\leq \sum_{l=1}^{d_1} \sum_{\alpha=2}^{d_2} \sum_{J=1}^{N+1} \left| \tilde{a}^*_{J,\alpha,l} \right| \sup_{1 \leq J \leq N+1, 1 \leq l \leq d_1, 2 \leq \alpha \leq d_2}$$

$$\times \left| n^{-1} \sum_{i=1}^{n} K_h(X_{i1} - x_1) T_{il} T_{il'} B^*_{J,\alpha}(X_{i\alpha}) \right|$$

$$= \sum_{l=1}^{d_1} \sum_{\alpha=2}^{d_2} \sum_{J=1}^{N+1} \left| \tilde{a}^*_{J,\alpha,l} \right|$$

$$\times \left[\sup_{\substack{1 \leq J \leq N+1, \\ 1 \leq l \leq d_1, 2 \leq \alpha \leq d_2}} \left| n^{-1} \sum_{i=1}^{n} K_h(X_{i1} - x_1) T_{il} T_{il'} \left\{ B_{J,\alpha}(X_{i\alpha}) - \mathrm{E}_n B_{J,\alpha}(X_{i\alpha}) \right\} \right| \right].$$

Equation (A.24) in Lemma A.10 states that

$$\sup_{x_1 \in [0,1]} \sup_{1 \leq J \leq N+1, 1 \leq l, l' \leq d_1, 2 \leq \alpha \leq d_2} \left| n^{-1} \sum_{i=1}^{n} K_h(X_{i1} - x_1) T_{il} T_{il'} B_{J,\alpha}(X_{i\alpha}) \right|$$
$$= O_p\left(H^{1/2}\right),$$

while equation (A.21) of Lemma A.9 states that $\sup_{1 \leq J \leq N+1,} \left| \mathrm{E}_n B_{J,\alpha}(X_{i\alpha}) \right| = O_p\left(\log n / \sqrt{n}\right)$, and standard kernel argument shows that

$$\sup_{x_1 \in [0,1]} \sup_{1 \leq l, l' \leq d_1} \left| n^{-1} \sum_{i=1}^{n} K_h(X_{i1} - x_1) T_{il} T_{il'} \right| = O_p(1).$$

Therefore, one has

$$R_2 \le \left\{ (N+1)\, d_1\, (d_2-1) \sum_{l=1}^{d_1} \sum_{\alpha=2}^{d_2} \sum_{J=1}^{N+1} \left(\tilde{a}^*_{J,\alpha,1} \right)^2 \right\}^{1/2} \left\{ O_p\left(H^{1/2}\right) + O_p\left(\frac{\log n}{\sqrt{n}}\right) \right\}$$
$$= O_p\left(\left\{ \sum_{l=1}^{d_1} \sum_{\alpha=2}^{d_2} \sum_{J=1}^{N+1} \left(\tilde{a}^*_{J,\alpha,1} \right)^2 \right\}^{1/2} \right)$$
$$= O_p\left(\sum_{l=1}^{d_1} \left\| \tilde{m}_l(\mathbf{x}) - m_{0l} - \sum_{\alpha=1}^{d_2} g_{\alpha l}(\mathbf{x}) + \sum_{\alpha=1}^{d_2} \mathrm{E}_n g_{\alpha l}(\mathbf{x}) \right\|_2 \right) = O_p\left(n^{-1/2} + H^2\right). \tag{A.30}$$

The last step follows from

$$\left\| \tilde{m}_l(\mathbf{x}) - m_{0l} - \sum_{\alpha=1}^{d_2} g_{\alpha l}(\mathbf{x}) + \sum_{\alpha=1}^{d_2} \mathrm{E}_n g_{\alpha l}(\mathbf{x}) \right\|_2$$
$$\le \|\tilde{m}_l(\mathbf{x}) - m_l(\mathbf{x})\|_2 + \left\| m_l(\mathbf{x}) - m_{0l} - \sum_{\alpha=1}^{d_2} g_{\alpha l}(\mathbf{x}) \right\|_2 + \left\| \sum_{\alpha=1}^{d_2} \mathrm{E}_n g_{\alpha l}(\mathbf{x}) \right\|_2$$
$$\le 3C_\infty \sum_{\alpha=1}^{d_2} \left\| m'_{\alpha 1} \right\|_\infty H^2 + O_p\left(n^{-1/2}\right).$$

Thus $R_2 = O_p\left(n^{-1/2} + H^2\right)$. Similarly,

$$R_3 = \sup_{x_1 \in [0,1]} \left| \frac{1}{n} \sum_{i=1}^{n} K_h(X_{i1} - x_1) \sum_{l=1}^{d_1} \mathrm{E}_n g_{\text{-}1,l}\left(\mathbf{X}_{i,\text{-}1}\right) T_{il} T_{il'} \right|$$
$$\le \left\{ \sum_{l=1}^{d_1} \left| \mathrm{E}_n g_{\text{-}1,l}\left(\mathbf{X}_{i,\text{-}1}\right) \right| \right\} \sup_{x_1 \in [0,1]} \left| \frac{1}{n} \sum_{i=1}^{n} K_h(X_{i1} - x_1) T_{il} T_{il'} \right|$$
$$\le \left\{ \sum_{l=1}^{d_1} \left| \mathrm{E}_n g_{\text{-}1,l}\left(\mathbf{X}_{i,\text{-}1}\right) \right| \right\} \sup_{x_1 \in [0,1]} \left| \frac{1}{n} \sum_{i=1}^{n} K_h(X_{i1} - x_1) T_{il} T_{il'} \right| = O_p\left(n^{-1/2} + H^2\right), \tag{A.31}$$

by (A.28). Combining (A.29), (A.30), and (A.31), one establishes Proposition 3. ■

Proof of Lemma 1. Based on formula (21), $n^{-1} \sum_{i=1}^{n} \tilde{\varepsilon}^*_{\text{-}1,l}\left(\mathbf{X}_{i,\text{-}1}\right)$ is

$$n^{-1} \sum_{i=1}^{n} \sum_{\alpha=2}^{d_2} \sum_{J=1}^{N+1} \tilde{a}_{J,\alpha,l} B_{J,\alpha}(X_{i\alpha}) = \sum_{\alpha=2}^{d_2} \sum_{J=1}^{N+1} \tilde{a}_{J,\alpha,l} \left\{ n^{-1} \sum_{i=1}^{n} B_{J,\alpha}(X_{i\alpha}) \right\}.$$

Lemma A.6 implies that

$$\left| \sum_{\alpha=2}^{d_2} \sum_{J=1}^{N+1} \tilde{a}_{J,\alpha,l} \right| \le \left\{ (N+1)(d_2-1) \cdot \sum_{\alpha=2}^{d_2} \sum_{J=1}^{N+1} \tilde{a}^2_{J,\alpha,l} \right\}^{1/2}$$
$$\le \left\{ (N+1)(d_2-1) \cdot \tilde{\mathbf{a}}^T \tilde{\mathbf{a}} \right\}^{1/2} = O_p\left(N n^{-1/2} \log n\right).$$

Clearly, (A.20) and (A.21) imply $\sup_{1\le J\le N+1}\left|n^{-1}\sum_{i=1}^{n} B_{J,\alpha}\left(X_{i\alpha}\right)\right| \le A_{n,1} = O_p\left(n^{-1/2}\log n\right)$, hence

$$\frac{1}{n}\sum_{i=1}^{n}\tilde{\varepsilon}^{*}_{\text{-}1,l}\left(\mathbf{X}_{i,\text{-}1}\right) \le \left|\sum_{\alpha=2}^{d_2}\sum_{J=1}^{N+1}\tilde{a}_{J,\alpha,l}\right| \cdot \sup_{J,\alpha,l}\left|n^{-1}\sum_{i=1}^{n} B_{J,\alpha}\left(X_{i\alpha}\right)\right| = O_p\left(\frac{N\left(\log n\right)^2}{n}\right), \tag{A.32}$$

while standard kernel theory implies that $\sup_{x_1\in[0,1]}\left|n^{-1}\sum_{i=1}^{n}\sum_{l=1}^{d_1} K_h\left(X_{i1}-x_1\right)T_{il}\, T_{il'}\right| = O_p(1)$. Thus the lemma follows immediately from (A.32) and (26). ■