Additive coefficient model (Xue & Yang 2006a,b) 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 is both (i) computationally expedient so it is usable for analyzing high dimensional data, and (ii) theoretically reliable so inference can be made on the component functions with confidence. In addition, it is (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 US GDP that allows non neutral effects of the R&D on capital and labor as well as in the Total Factor Productivity (TFP).

1. INTRODUCTION

Regression analysis has been widely used in econometrics studies, for instance, the estimation of production/cost function. Typical parametric regression models presume that their regression functions follow a pre-determined form with finitely many unknown parameters. Nonparametric models, on the other hand, impose less stringent assumptions on the regression functions, but pay for their flexibility the price of “curse of dimensionality”. Structured model offers a sensible compromise between parametric simplicity and nonparametric flexibility, see for example Sperlich, Tjostheim & Yang (2002) for additive interaction modelling for the production function of Wisconsin farms and Rodríguez-Póo, Sperlich & Vieu (2003) for a general framework of separable models. Recently Xue & Yang (2006a,b) have proposed additive coefficient model that allows a response variable

---

1The 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 Statistics and Probability, Michigan State University, East Lansing, MI 48824, USA; e-mail: liurong@stt.msu.edu; and Lijian Yang, Department of Statistics and Probability, Michigan State University, East Lansing, MI 48824, USA; e-mail: yang@stt.msu.edu.
Y to depend linearly on some regressors, with coefficients as smooth additive functions of other predictors, called tuning variables. Specifically

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

in which the predictor vector \((X, T)\) consists of the tuning variables \(X = (X_1, ..., X_{d_2})^T \in R^{d_2}\) and linear predictors \(T = (T_1, ..., T_{d_1})^T \in R^{d_1}\). The functional coefficient model of Chen & Tsay (1993b) corresponds to the case \(d_2 = 1\), the varying coefficient model of Hastie & Tibshirani (1993) corresponds to the case \(d_2 = d_1\) and for each \(l = 1, ..., d_1\) only one single significant \(m_{\alpha l}\) with \(\alpha = l\).

Also included as special cases of model (1) are the additive model of Hastie & Tibshirani (1990), Chen & Tsay (1993a), and the multivariate linear regression model, see Xue & Yang (2006a) for detailed discussion. Model (1)’s versatility for econometric applications is illustrated by the following example. Consider the forecasting of US GDP annual growth rate, which is modelled as the Total Factor Productivity (TFP) growth rate plus a linear function of capital growth rate and labor growth rate, according to the classic Cobb-Douglas model (Cobb & Douglas, 1928).

As pointed out in Li & Racine (2007), p.302, it is unrealistic to ignore the non neutral effect of R&D spending on the TFP growth rate and on the complementary slopes of capital and labor growth rates. Thus a smooth coefficient model should fit the production function better than the parametric Cobb-Douglas model. Indeed, Figure 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.

(Insert Figure 1 about here)

(Insert Figure 2 about here)

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

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

We extend the oracle smoothing idea of Linton (1997) and Wang & 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 \( Y_{l1} = \sum_{l=1}^{d_1} m_{1l}(X_l) T_l + \sigma(X, T) \varepsilon = Y - \sum_{l=1}^{d_1} \left\{ m_{0l} + \sum_{\alpha=2}^{d_2} m_{\alpha l}(X_\alpha) \right\} T_l \) and estimate all functions \( \{m_{1l}(x_1)\}_{l=1}^{d_1} \) by linear regression of \( Y_{l1} \) 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 and then use these estimates as substitutes to obtain an approximation \( \hat{Y}_{l1} \) to the variable \( Y_{l1} \), and construct “oracle” estimators based on \( \hat{Y}_{l1} \). As in Wang & 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 & Yang (2007), a major theoretical complication is the dependence structure of \( T \) on \( X \), necessitating Assumption (A2) on the second moment matrix \( Q(x) = E(\mathbf{T}\mathbf{T}^T | \mathbf{X} = \mathbf{x}) \), see the detailed discussion on Assumption (A2) at the end of Section 2 and the extra step to estimate \( Q(x) \) in Section 5. In contrast, for the additive model of Wang & Yang (2007), there is no need of Assumption (A2) and of estimating \( Q(x) \equiv 1 \). 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 & Yang (2006a) had provided \( \sqrt{n} \)-consistent estimation of
constants \( \{m_{0l}\}_{l=1}^{d_1} \) only under higher order smoothness Assumptions, while Xue & Yang (2006b) had failed to obtain \( \sqrt{n} \)-consistency for estimating \( \{m_{0l}\}_{l=1}^{d_1} \). For the additive model of Wang & 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 & Yang (2007) used existing theory from kernel smoothing literature.

The paper is organized as follows. In Section 2 we discuss the assumptions of the model (1). In Section 3, we introduce the oracle smoothers and discuss its asymptotic properties. In Section 4 we introduce the SBK and SBLL estimators, their uniform onsistency 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 \( \{(Y_i, X_i, T_i)\}_{i=1}^{n} \) be a sequence of strictly stationary observations, with identical distribution as \((Y, X, T)\) in model (1). Denote the unknown conditional mean and variance functions as \( m(X, T) = E(Y|X, T), \sigma^2(X, T) = \text{var}(Y|X, T) \), then one has

\[
Y_i = m(X_i, T_i) + \sigma(X_i, T_i) \epsilon_i
\]

for some conditional white noises \( \{\epsilon_i\}_{i=1}^{n} \) that satisfy \( E(\epsilon_i|X_i, T_i) = 0, E(\epsilon_i^2|X_i, T_i) = 1 \). The variables \((X_i, T_i)\) 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

\[
E\{m_{0l}(X_\alpha)\} = 0, 1 \leq l \leq d_1, 1 \leq \alpha \leq d_2
\]

ensuring the unique additive representations of \( m_l(x) = m_{0l} + \sum_{\alpha=1}^{d_2} m_{al}(x_\alpha) \). As in most works on nonparametric smoothing, estimation of the functions \( \{m_{al}(x_\alpha)\}_{l=1,a=1}^{d_1,d_2} \) 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 \( \alpha \)-centered square integrable functions on \([0,1]\) is

\[
\mathcal{H}_\alpha = \{g : E\{g(X_\alpha)\} = 0, E\{g^2(X_\alpha)\} < +\infty\}, 1 \leq \alpha \leq d_2.
\]

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

\[
\mathcal{M} = \left\{ g(x, t) = \sum_{l=1}^{d_1} g_l(x) t_l : g_l(x) = g_{0l} + \sum_{\alpha=1}^{d_2} g_{al}(x_\alpha) : g_{al} \in \mathcal{H}_\alpha \right\},
\]

in which \( \{g_{0l}\}_{l=1}^{d_1} \) are finite constants. The constraints that \( E\{g_{al}(X_\alpha)\} = 0, 1 \leq \alpha \leq d_2 \) 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, $E_n \varphi = \sum_{i=1}^n \varphi(X_i,T_i)/n$. We introduce two inner products on $\mathcal{M}$. For functions $g_1, g_2 \in \mathcal{M}$, the theoretical and empirical inner products are defined respectively as $\langle g_1, g_2 \rangle = E\{g_1(X,T)g_2(X,T)\}$, $(g_1, g_2)_n = E_n\{g_1(X,T)g_2(X,T)\}$. The corresponding induced norms are $\|g_1\|_2 = E g_1^2(X,T)$, $\|g_1\|_{2,n} = E_n g_1^2(X,T)$. 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 $p$-th order smooth function as $C^p[a, b] = \{g | g^{(p)} \in C[a, b]\}$, and the class of Lipschitz continuous functions for constant $C > 0$ as $\text{Lip}([a, b], C) = \{g | |g(x) - g(x')| \leq C|x - x'|, \forall x, x' \in [a, b]\}$. We mean by “~” both sides having the same order as $n \rightarrow \infty$. We denote by $I_{d_1 \times d_1}$ the $d_1 \times d_1$ identity matrix, and $0_{d_1 \times d_1}$ the $d_1 \times d_1$ zero matrix. For any vector $x = (x_1, x_2, \cdots, x_{d_2})$, we denote the supremum and Euclidean norms as $|x| = \max_{1 \leq \alpha \leq d_2} |x_\alpha|$ and $\|x\| = \left(\sum_{\alpha=1}^{d_2} x_\alpha^2\right)^{1/2}$.

We need the following Assumptions on the data generating process.

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

(A2) There exist constants $0 < c_Q \leq C_Q < +\infty$ and $0 < c_\delta \leq C_\delta < +\infty$ and some $\delta > 1/2$, such that $c_Q I_{d_1 \times d_1} \leq Q(x) = \{q(x)\}_{i,j=1}^{d_1} = E(TT^T|X = x) \leq C_Q I_{d_1 \times d_1}$ and $c_\delta \leq E\left\{(T_iT_j)^{2+\delta} | X = x\right\} \leq C_\delta$ for all $x \in \chi$ and $l, l' = 1, \ldots, d_1$.

(A3) The vector process $\{\mathbf{q}_i\}_{t=-\infty}^{\infty} = \{(Y_t, X_t, T_t)\}_{t=-\infty}^{\infty}$ is strictly stationary and geometrically strongly mixing, that is, its $\alpha$-mixing coefficient $\alpha(k) \leq c_p^k$, for constants $c > 0$, $0 < \rho < 1$, where $\alpha(k) = \sup_{A \in \sigma(\mathbf{q}_s), s \leq 0, B \in \sigma(\mathbf{q}_t), t \geq k} |P(A)P(B) - P(A \cap B)|$.

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

(A5) The conditional variance function $\sigma^2(x, t)$ is measurable and bounded. The errors $\{e_i\}_{i=1}^n$ satisfy $E(e_i|F_i) = 0$, $E(e_i^2|F_i) = 1$, $E(|e_i|^{2+\eta}|F_i) \leq C_\eta$ for some $\eta \in (1/2, 1]$ and the sequence of $\sigma$-fields $F_i = \sigma\{\{X_j, T_j\}, j \leq i; e_j, j \leq i - 1\}$ for $i = 1, \ldots, n$.

(A6) The marginal density $f_1(x_1)$ of $X_1$ and the conditional second moment matrix function $Q_1(x_1)$ defined in (4) both have continuous derivatives on $[0, 1]$.

Assumptions (A1)-(A5) are common in the literature, see for instance, Huang & Yang (2004), Huang & Shen (2004) and especially Xue & Yang (2006b). Assumption (A6) is needed only for the asymptotic theory of oracle “kernel smoother”, but not for the oracle “local linear smoother”.

5
Assumption (A2) implies also that for all \( x_\alpha \in [0, 1], 1 \leq \alpha \leq d_2 \) and \( l, l' = 1, \ldots, d_1 \)
\[
 c_Q I_{d_1 \times d_1} \leq Q_\alpha (x_\alpha) = \{ q_\alpha (x_\alpha) \}_{l,l'=1}^{d_1} = E(TT^T|X_\alpha = x_\alpha) \leq C Q I_{d_1 \times d_1} \tag{4}
\]
\[
c_\delta \leq E \left\{ (T_l T_{l'})^{2+\delta}|X_\alpha = x_\alpha \right\} \leq C_\delta.
\]
Furthermore, Assumptions (A2) and (A5) imply that for some constant \( C > 0 \)
\[
 \max_{1 \leq l \leq d_1} E |T_l|^{2+\eta} < C \max_{1 \leq l \leq d_1} E |T_l|^{2+\delta} = C \max_{1 \leq l \leq d_1} E |T_l|^{4+2\delta} \leq CC_\delta < +\infty. \tag{5}
\]
At one referee’s request, we provide here insight into the relationship allowed between the vectors \( T \) and \( X \) under Assumption (A2). It is instructive to first understand what \( T \) and \( X \) can not be in the context of identifiability for functions \( \{ m_{\alpha l}(x_\alpha) \}_{l=1, \alpha=1}^{d_1, d_2} \). Suppose that the vector \( X \) is centered so that \( EX = \mathbf{0} \). Then model (1) is unidentifiable when \((T_1, T_2) = (X_1, X_2)\) since \(-3X_2 T_1 + 3X_1 T_2 = 0, E(-3X_2) = E(3X_1) = 0 \) and the function \( m(x, t) \) in (1) is expressed as
\[
\sum_{l=3}^{d_1} \left\{ m_{0l} + \sum_{\alpha=1}^{d_2} m_{\alpha l}(x_\alpha) \right\} t_l + \left\{ m_{01} + m_{21}(x_2) + \sum_{\alpha=1, \alpha \neq 2}^{d_2} m_{\alpha 1}(x_\alpha) \right\} t_1 \\
+ \left\{ m_{02} + m_{12}(x_1) + \sum_{\alpha=2}^{d_2} m_{\alpha 1}(x_\alpha) \right\} t_2
\]
\[
\equiv \sum_{l=3}^{d_1} \left\{ m_{0l} + \sum_{\alpha=1}^{d_2} m_{\alpha l}(x_\alpha) \right\} t_l + \left\{ m_{01} + m_{21}(x_2) - 3x_2 + \sum_{\alpha=1, \alpha \neq 2}^{d_2} m_{\alpha 1}(x_\alpha) \right\} t_1 \\
+ \left\{ m_{02} + m_{12}(x_1) + 3x_1 + \sum_{\alpha=2}^{d_2} m_{\alpha 1}(x_\alpha) \right\} t_2,
\]
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 & Yang (2006a), p.2523 gave a similar counterexample, and discussed why an unidentifiable model may perform better for prediction.

More generally, it is revealing to note that Assumption (A2) not only rules out the above anomaly, but it also does not allow the possibility that there exist two \( T_l \)'s (\( 1 \leq l \leq d_1 \)) almost surely equal to two Borel functions of \( X \). To see this, suppose that \((T_1, T_2) = \{ \varphi_1(X), \varphi_2(X) \}, a.s \) for some Borel functions \( \varphi_1 \) and \( \varphi_2 \). Assumption (A2) implies that
\[
c_Q I_{2 \times 2} \leq E \left\{ \begin{pmatrix} T_1^2 & T_1 T_2 \\ T_1 T_2 & T_2^2 \end{pmatrix} \right| X = x \leq C Q I_{2 \times 2}, \forall x \in \chi
\]
leading to
\[
c_Q I_{2 \times 2} \leq \begin{pmatrix} \varphi_1^2(x) & \varphi_1(x) \varphi_2(x) \\ \varphi_1(x) \varphi_2(x) & \varphi_2^2(x) \end{pmatrix} \leq C Q I_{2 \times 2}, a.s., \forall x \in \chi
\]
which can not be true as for any \( x \in \chi \), the \( 2 \times 2 \) matrix in the above is singular, thus can not be \( \geq cQ_{2 \times 2} \). That Assumption (A2) guarantees the identifiability of model (1) has been established in Lemma 1 of Xue & Yang (2006b). It is important to observe, however, that Assumption (A2) does allow the case of exactly one \( T_l, 1 \leq l \leq d_1 \) almost surely equal to a Borel function of \( X \).

3. THE ORACLE SMOOTHERS

We now introduce what is known as the oracle smoother in Wang & Yang (2007) as a benchmark for evaluating the estimators. Denote for any vector \( x = (x_1, x_2, \ldots, x_{d_2}) \) the deleted vector \( x_{\cdot,1} = (x_2, \ldots, x_{d_2}) \) and for the random vector \( X_i = (X_{i1}, X_{i2}, \ldots, X_{id_2}) \) the deleted vector \( X_{i,1} = (X_{i2}, \ldots, X_{id_2}) \), \( 1 \leq i \leq n \). For any \( 1 \leq l \leq d_1 \), write \( m_{l,1,l} (x_{\cdot,1}) = m_{0l} + \sum_{\alpha=2}^{d_2} m_{\alpha l} (x_{\alpha}) \). Denote the vector of pseudo-responses \( Y_1 = (Y_{1,1,1}, \ldots, Y_{n,1,1})^T \) in which

\[
Y_{i,1} = Y_i - \sum_{l=1}^{d_1} \{ m_{0l} + m_{l,1,l} (X_{i,1}) \} T_{il} = \sum_{l=1}^{d_1} m_{il} (X_{i1}) T_{il} + \sigma (X_{i1}, T_{i1}) \varepsilon_i.
\]

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

\[
m_{K,1,1} (x_1) = \{ m_{K,1,1} (x_1), \ldots, m_{K,1,1d_1} (x_1) \}^T = \arg\min_{\lambda=(\lambda_i)_{1 \leq i \leq d_1}} L (\lambda, m_{l,1}, x_1)
\]

in which

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

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

\[
m_{K,1,1} (x_1) = \left( C_K^T W_1 C_K \right)^{-1} C_K^T W_1 Y_1 = \left( \frac{1}{n} C_K^T W_1 C_K \right)^{-1} \frac{1}{n} C_K^T W_1 Y_1
\]

in which

\[
T_i = (T_{i1}, \ldots, T_{id_1})^T, C_K = \{ T_1, \ldots, T_n \}^T,
\]

\[
W_1 = \text{diag} \{ K_h (X_{11} - x_1), \ldots, K_h (X_{n1} - x_1) \},
\]

\[
K_h (u) = K (u/h) / h \text{ for a kernel function } K \text{ and bandwidth } h \text{ that satisfy}
\]

\[(A7) \text{ The function } K \text{ is a symmetric probability density function supported on } [-1,1], \text{ and } K \in \text{Lip } (-1,1, C_K) \text{ for some } C_K > 0, \text{ while the bandwidth } h = h_{1,n} > 0, h \sim n^{-1/3}.
\]

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

\[
m_{LL,1,1} (x_1) = (I_{d_1 \times d_1}, 0_{d_1 \times d_1}) \left( \frac{1}{n} C_{LL,1}^T W_1 C_{LL,1} \right)^{-1} \frac{1}{n} C_{LL,1}^T W_1 Y_1,
\]

\[(7)
\]

\[
7
\]
in which
\[
C_{LL,1} = \left\{ \begin{array}{cccc}
T_1 & \ldots & T_n \\
T_1(X_{i1} - x_1) & \ldots & T_n(X_{ni} - x_1)
\end{array} \right\}^T.
\]

In this paper denote \( \mu_2(K) = \int u^2 K(u) du, \|K\|_2^2 = \int K(u)^2 du, \) \( Q_1(x_1) \) as in (4) and define the following bias and variance coefficients
\[
b_{LL,l',l,1}(x_1) = \frac{1}{2} \mu_2(K) m_{ll}''(x_1) f_1(x_1) q_{ll',1}(x_1),
\]
\[
b_{K,l',l,1}(x_1) = \frac{1}{2} \mu_2(K) \left[ 2 m_{ll}'(x_1) \frac{\partial}{\partial x_1} \left\{ f_1(x_1) q_{ll',1}(x_1) \right\} + m_{ll}''(x_1) f_1(x_1) q_{ll',1}(x_1) \right],
\]
\[
\Sigma_1(x_1) = \|K\|_2^2 f_1(x_1) E\left\{ TT^T \sigma^2(X, T) | X_1 = x_1 \right\},
\]
\[
\{ v_{l,l',1}(x_1) \}_{l,l'=1}^{d_1} = Q_1(x_1)^{-1} \Sigma_1(x_1) Q_1(x_1)^{-1}.
\]

**THEOREM 1** Under Assumptions (A1) to (A5) and (A7), for any \( x_1 \in [h, 1 - h], \) as \( n \to \infty, \) the oracle local linear smoother \( \tilde{m}_{LL,1}(x_1) \) given in (7) satisfies
\[
\sqrt{n} h \left[ \tilde{m}_{LL,1}(x_1) - m_1(x_1) - \left\{ \sum_{l=1}^{d_1} b_{LL,l,l',1}(x_1) \right\}_{l'=1}^{d_1} h^2 \right] \to N \left( 0, \{ v_{l,l',1}(x_1) \}_{l,l'=1}^{d_1} \right).
\]

With Assumption (A6) in addition, the oracle kernel smoother \( \tilde{m}_{K,1}(x_1) \) in (6) satisfies
\[
\sqrt{n} h \left[ \tilde{m}_{K,1}(x_1) - m_1(x_1) - \left\{ \sum_{l=1}^{d_1} b_{K,l,l',1}(x_1) \right\}_{l'=1}^{d_1} h^2 \right] \to N \left( 0, \{ v_{l,l',1}(x_1) \}_{l,l'=1}^{d_1} \right).
\]

**THEOREM 2** Under Assumptions (A1) to (A5) and (A7), as \( n \to \infty, \) the oracle local linear smoother \( \tilde{m}_{LL,1}(x_1) \) given in (7) satisfies
\[
\sup_{x_1 \in [h, 1-h]} | \tilde{m}_{LL,1}(x_1) - m_1(x_1) | = O_p \left( \log n / \sqrt{n} h \right).
\]

With Assumption (A6) in addition, the oracle kernel smoother \( \tilde{m}_{K,1}(x_1) \) in (6) satisfies
\[
\sup_{x_1 \in [h, 1-h]} | \tilde{m}_{K,1}(x_1) - m_1(x_1) | = O_p \left( \log n / \sqrt{n} h \right).
\]

**Remark 1** The above theorems hold for \( \tilde{m}_{LL,\alpha}(x_\alpha) \) and \( \tilde{m}_{K,\alpha}(x_\alpha) \) similarly constructed as \( \tilde{m}_{LL,1}(x_1) \) and \( \tilde{m}_{K,1}(x_1) \), for any \( \alpha = 2, \ldots, d_2, i.e., \)
\[
\tilde{m}_{LL,\alpha}(x_\alpha) = (I_{d_1 \times d_1}, 0_{d_1 \times d_1}) \left( \frac{1}{n} C_{LL,\alpha}^T W_\alpha C_{LL,\alpha} \right)^{-1} \frac{1}{n} C_{LL,\alpha}^T W_\alpha Y_\alpha,
\]
\[
\tilde{m}_{K,\alpha}(x_\alpha) = \left( \frac{1}{n} C_K^T W_\alpha C_K \right)^{-1} \frac{1}{n} C_K^T W_\alpha Y_\alpha,
\]
except that in Assumption (A4) one has to replace “\( m_{ll} \in C^2 [0, 1], \forall 1 \leq l \leq d_1 \)” with “\( m_{al} \in C^2 [0, 1], \forall 1 \leq l \leq d_1 \)” and in Assumption (A6), \( f_1(x_1) \) and \( Q_1(x_1) \) have to be replaced with \( f_\alpha(x_\alpha) \) and \( Q_\alpha(x_\alpha) \).
The proofs of Theorems 1 and 2 can be found in Liu & Yang (2008), subsection A.4. The same oracle idea applies to the constants as well. Define the would-be “estimators” of constants $(m_0 l)_{l=1}^{T}$ as the following least squares solution

$$
\tilde{m}_0 = (\tilde{m}_0 l)_{l=1}^{T} = \arg \min \sum_{i=1}^{n} \left\{ Y_{ic} - \sum_{l=1}^{d_1} m_0 l T_{il} \right\}^2,
$$

in which the oracle responses are

$$
Y_{ic} = Y_i - \sum_{l=1}^{d_1} \sum_{\alpha=1}^{d_2} m_{ol} (X_{i\alpha}) T_{il} = \sum_{l=1}^{d_1} m_0 l T_{il} + \sigma (X_i, T_i) \varepsilon_i.
$$

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

**PROPOSITION 1** Under Assumptions (A1)-(A5) and (A8), as $n \to \infty$,

$$
\sup_{1 \leq d_1} |\tilde{m}_0 l - m_0 l| = O_p (n^{-1/2}).
$$

Although the oracle smoothers $\tilde{m}_{lll} (x_i)$, $\tilde{m}_{lk} (x_i)$ 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 $(m_{ol} (x_i))_{l=1, \alpha=2}^{d_1, d_2}$ and constants $(m_0 l)_{l=1}^{d_1}$. 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 $(m_{ol} (x_i))_{l=1, \alpha=2}^{d_1, d_2}$ and constants $(m_0 l)_{l=1}^{d_1}$ can be pre-estimated by linear spline and how the estimates are used to construct the “oracle estimators”. To this end, we first introduce the space of linear splines. Let $0 = \xi_0 < \xi_1 < \cdots < \xi_N < \xi_{N+1} = 1$ denote a sequence of equally spaced points, called interior knots, on interval $[0, 1]$. Denote by $H = (N + 1)^{-1}$ the width of each subinterval $[\xi_j, \xi_{j+1}]$, $0 \leq J \leq N$ and denote the degenerate knots $\xi_{-1} = 0, \xi_{N+2} = 1$. We assume that

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

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

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

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

$$
G_{n, \alpha}^{0} = \left\{ g_\alpha : g_\alpha (x) \equiv \sum_{J=0}^{N+1} \lambda_J b_J (x) \cdot E_n \{ g_\alpha (X_\alpha) \} = 0 \right\}, 1 \leq \alpha \leq d_2,
$$
and the space of additive spline coefficient functions on $\chi \times R^{d_1}$ as

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

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

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

$$\hat{m} \left( x, t \right) = \sum_{l=1}^{d_1} \hat{m}_l \left( x \right) t_l = \arg\min_{g \in G_n^0} \sum_{i=1}^{n} \left\{ Y_i - g \left( X_i, T_i \right) \right\}^2.$$

Since $\hat{m} \left( x, t \right) \in G_n^0$, one can write $\hat{m}_l \left( x \right) = \hat{m}_{0l} + \sum_{a=1}^{d_2} \hat{m}_{al} \left( x_a \right)$; for $\hat{m}_{0l} \in R$ and $\hat{m}_{al} \left( x_a \right) \in G_{n,a}^0$.

Simple algebra shows that the following oracle estimators of the constants $m_{0l}$ are exactly equal to $\hat{m}_{0l}$, in which the oracle pseudo-responses $\hat{Y}_{ic} = Y_i - \sum_{l=1}^{d_1} \sum_{a=1}^{d_2} \hat{m}_{al} \left( X_{ia} \right) T_{il}$ which mimick the oracle responses $Y_{ic}$ in (10)

$$\hat{m}_0 = (\hat{m}_{0l})_{1 \leq l \leq d_1}^{T} = \arg\min_{(\lambda_0, \ldots, \lambda_{d_1})} \sum_{i=1}^{n} \left\{ \hat{Y}_{ic} - \sum_{l=1}^{d_1} \lambda_{il} T_{il} \right\}^2.$$

**PROPOSITION 2** Under Assumptions (A1) to (A5) and (A8), as $n \to \infty$,

$$\sup_{1 \leq l \leq d_1} |\hat{m}_{0l} - m_{0l}| = O_p \left( n^{-1/2} \right), \text{ hence } \sup_{1 \leq l \leq d_1} |\hat{m}_{0l} - m_{0l}| = O_p \left( n^{-1/2} \right)$$

following Proposition 1.

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

$$\hat{m}_{SBK,1.} \left( x_1 \right) = \left( C_{K}^T W_{1} C_{K} \right)^{-1} C_{K}^T W_{1} \hat{Y}_1 = \left( \frac{1}{n} C_{K}^T W_{1} C_{K} \right)^{-1} \frac{1}{n} C_{K}^T W_{1} \hat{Y}_1,$$

$$\hat{m}_{SBLL,1.} \left( x_1 \right) = \left( I_{d_1 \times d_1}, 0_{d_1 \times d_1} \right) \left( \frac{1}{n} C_{LL,1}^T W_{1} C_{LL,1} \right)^{-1} \frac{1}{n} C_{LL,1}^T W_{1} \hat{Y}_1.$$  

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

**THEOREM 3** Under Assumptions (A1) to (A5), (A7) and (A8), as $n \to \infty$, the SBK estimator $\hat{m}_{SBK,1.} \left( x_1 \right)$ in (13) and the SBLL estimator $\hat{m}_{SBLL,1.} \left( x_1 \right)$ in (14) satisfy

$$\sup_{x_1 \in [0,1]} |\hat{m}_{SBK,1.} \left( x_1 \right) - \hat{m}_{K,1.} \left( x_1 \right)| + \sup_{x_1 \in [0,1]} |\hat{m}_{SBLL,1.} \left( x_1 \right) - \hat{m}_{LL,1.} \left( x_1 \right)| = o_p \left( n^{-2/5} \right).$$
Theorem 3 follows from (23) and Propositions 2, 3 and 4, and remains true if Assumption (A8) on the number of knots is of the more general form $N \sim n^{1/2}N'$ where $N' \to \infty, N'/n' \to 0, \forall r > 0$ 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 (A7). The following corollary provides the asymptotic distributions of $\hat{m}_{SBLL,1.}(x_1)$ and $\hat{m}_{K,1.}(x_1)$. The proof of this corollary is straightforward from Theorems 1 and 3.

**COROLLARY 1** Under Assumptions (A1) to (A5), (A7) and (A8), for any $x_1 \in [h, 1-h]$, as $n \to \infty$, the SBLL estimator $\hat{m}_{SBLL,1.}(x_1)$ in (14) satisfies

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

and with the additional Assumption (A6), the SBK estimator $\hat{m}_{SBK,1.}(x_1)$ in (13) satisfies

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

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

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

$$
\hat{m}_{SBK,\alpha.}(x_\alpha) = \left( \frac{1}{n} C_K^T W_{\alpha} C_K \right)^{-1} \frac{1}{n} C_K^T W_{\alpha} \hat{Y}_{\alpha},
$$

where $\hat{Y}_{\alpha} = Y_{i} - \sum_{l=1}^{d_1} \left\{ \hat{m}_{il} + \sum_{1 \leq \alpha' \leq d_2, \alpha' \neq \alpha} \hat{m}_{i\alpha'} \left( X_{i\alpha'} \right) \right\}$.

To understand the proof of Theorem 3, we study the difference between the smoothed backfitted estimator $\hat{m}_{SBK,1.}(x_1)$ and the smoothed “oracle” estimator $\hat{m}_{K,1.}(x_1)$. First, define the theoretical inner product of $b_J$ and 1 with respect to the $\alpha$-th marginal density $f_{\alpha}(x_\alpha)$ as $c_{J,\alpha} = \langle b_J (X_{\alpha}), 1 \rangle = \int b_J (x_\alpha) f_{\alpha}(x_\alpha) dx_\alpha$ and define the centered B spline basis $b_{J,\alpha}(x_\alpha)$ and the standardized B spline basis $B_{J,\alpha}(x_\alpha)$ as

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

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

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

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

(17)
so one can rewrite the linear spline estimator in (11) as \( \hat{m}_l(x, t) = (P_nY)(x, t) \), where we denote by \( Y = (Y_l)_{1 \leq l \leq n} \) the response vector. The coefficients of the linear regressors \( t_l, 1 \leq l \leq d_1 \) are denoted as the multivariate additive spline functions

\[
(P_n, \Gamma)(x) = \hat{\gamma}_{0,l} + \sum_{\alpha=1}^{d_2} \sum_{J=1}^{N+1} \hat{\gamma}_{J,\alpha,l}B_{J,\alpha}(x), \ l = 1, \ldots, d_1.
\]

Note that \((P_n, \Gamma)(x) = \hat{\gamma}_{0,l} + \sum_{\alpha=1}^{d_2} \left( P_{n,\alpha,l}^* \Gamma \right)(x) \) where \((P_{n,\alpha,l}^* \Gamma)(x) = \sum_{j=1}^{N+1} \hat{\gamma}_{J,\alpha,l}B_{J,\alpha}(x) \), we define the empirically centered additive components \( (P_{n,\alpha,l}^* \Gamma)(x), \ \alpha = 1, \ldots, d_2 \)

\[
(P_{n,\alpha,l}^* \Gamma)(x) = (P_{n,\alpha,l}^* \Gamma)(x) - n^{-1} \sum_{i=1}^{n} (P_{n,\alpha,l}^* \Gamma)(X_{i\alpha}). \tag{18}
\]

Using these notations, spline estimators of \( m_l(x) \) and \( m_{\alpha,l}(x) \) are \( \hat{m}_l(x) = (P_nY)(x) \), \( \hat{m}_{\alpha,l}(x) = (P_{n,\alpha,l}Y)(x) \), while noiseless splines and variance spline components are

\[
\hat{m}_l(x) = (P_n, m)(x), \quad \hat{m}_{\alpha,l}(x) = (P_{n,\alpha,l}m)(x),
\]

\[
\tilde{e}_l(x) = (P_n, E)(x), \quad \tilde{e}_{\alpha,l}(x) = (P_{n,\alpha,l}E)(x), \tag{19}
\]

where \( m = \{m(X_i, T_i)\}_{1 \leq i \leq n}^T \) is the true function vector and \( E = \{\sigma(X_i, T_i) \varepsilon_i\}_{1 \leq i \leq n}^T \) the error vector. Due to the linearity of operators \( P_{n,\alpha} \) and \( P_{n,\alpha,l} \), \( 1 \leq l \leq d_1, 1 \leq \alpha \leq d_2 \) and \( Y = m + E \) due to (2), one has the following crucial decomposition for proving Theorem 3,

\[
\hat{m}_l(x) = \hat{m}_l(x) + \tilde{e}_l(x), \quad \hat{m}_{\alpha,l}(x) = \hat{m}_{\alpha,l}(x) + \tilde{e}_{\alpha,l}(x), \quad 1 \leq l \leq d_1, 1 \leq \alpha \leq d_2. \tag{20}
\]

We define additionally an auxiliary entity

\[
\tilde{e}_{\alpha,l}^*(x) = (P_{n,\alpha,l}^* E)(x), \quad 1 \leq l \leq d_1, 1 \leq \alpha \leq d_2. \tag{21}
\]

Definition (18) implies that \( \tilde{e}_{\alpha,l}(x) \) is simply the empirical centering of \( \tilde{e}_{\alpha,l}^*(x) \), i.e.

\[
\tilde{e}_{\alpha,l}(x) \equiv \tilde{e}_{\alpha,l}^*(x) - n^{-1} \sum_{i=1}^{n} \tilde{e}_{\alpha,l}^*(X_{i\alpha}). \tag{22}
\]

According to (6) and (13),

\[
\hat{m}_{SBK,1,l}(x) - \hat{m}_{K,1,l}(x) = \left( \frac{1}{n} C_K^T W C_K \right)^{-1} \frac{1}{n} C_K^T W_1 \left( \hat{Y}_1 - Y_1 \right),
\]

\[
\hat{Y}_1 - Y_1 = \left( \hat{Y}_{1,1}, \cdots, \hat{Y}_{1,n} \right)^T - (Y_{1,1}, \cdots, Y_{1,n})^T = \left[ \sum_{l=1}^{d_1} \{m_{0l} - \hat{m}_{0l} + m_{1,l}(X_{i,1}) - \hat{m}_{1,l}(X_{i,1})\} T_{il} \right]_{1 \leq i \leq n}
\]

12
\[ = C_K (m_{01} - \hat{m}_{01})_{1 \leq l \leq d_1} + \left[ \sum_{l=1}^{d_1} \{ m_{1,l}(X_{i,l}) - \hat{m}_{1,l}(X_{i,l}) \} T_{il} \right]_{1 \leq i \leq n} \]

where making use of the definition of \( \hat{m}_{01} \) and the signal noise decomposition (20), the difference \( \hat{m}_{K,1} \cdot (x_1) - \hat{m}_{SBK,1} \cdot (x_1) - \hat{m}_{0} + m_{0} \cdot \) can be treated as the sum of two terms

\[
\left( \frac{1}{n} C_K^T W_1 C_K \right)^{-1} \frac{1}{n} C_K^T W \left[ \sum_{l=1}^{d_1} \{ m_{1,l}(X_{i,l}) - \hat{m}_{1,l}(X_{i,l}) \} T_{il} \right]_{i=1}^{n} = \left( \frac{1}{n} C_K^T W_1 C_K \right)^{-1} \left\{ \Psi_b(x_1) + \Psi_v(x_1) \right\}_{d_1, l'=1}^{d_1} (23)
\]

where

\[
\Psi_b(x_1) = \frac{1}{n} C_K^T W_1 \left[ \sum_{l=1}^{d_1} \{ m_{1,l}(X_{i,l}) - \hat{m}_{1,l}(X_{i,l}) \} T_{il} \right]_{i=1}^{n} = \left\{ \Psi_{b,l'}(x_1) \right\}_{d_1, l'=1}^{d_1}, (24)
\]

\[
\Psi_v(x_1) = \frac{1}{n} C_K^T W_1 \left[ \sum_{l=1}^{d_1} \hat{\epsilon}_{i,l}(X_{i,l}) T_{il} \right]_{i=1}^{n} = \left\{ \Psi_{v,l'}(x_1) \right\}_{d_1, l'=1}^{d_1}, \hat{\epsilon}_{i,l}(X_{i,l}) = \sum_{a=1}^{d_2} \hat{\epsilon}_{al}(X_{i,a}) (25)
\]

and

\[
\Psi_{b,l'}(x_1) = \frac{1}{n} \sum_{i=1}^{n} K_h (X_{i,l} - x_1) T_{il'} \sum_{l=1}^{d_1} \{ m_{1,l}(X_{i,l}) - \hat{m}_{1,l}(X_{i,l}) \} T_{il}
\]

\[
\Psi_{v,l'}(x_1) = \frac{1}{n} \sum_{i=1}^{n} K_h (X_{i,l} - x_1) T_{il'} \sum_{l=1}^{d_1} \hat{\epsilon}_{i,l}(X_{i,l}) T_{il}.
\]

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

**PROPOSITION 3** Under Assumptions (A1)-(A4), (A7) and (A8), as \( n \to \infty \),

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

**PROPOSITION 4** Under Assumptions (A1) to (A5), (A7) to (A8), as \( n \to \infty \),

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

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

\[
\Psi_{v,l'}^{(1)}(x_1) = n^{-1} \sum_{i=1}^{n} \sum_{l=1}^{d_1} K_h (X_{i,l} - x_1) T_{il} T_{il'} \sum_{i=1}^{n} \hat{\epsilon}_{i,l}(X_{i,l}), (26)
\]

\[
\Psi_{v,l'}^{(2)}(x_1) = n^{-1} \sum_{i=1}^{n} \sum_{l=1}^{d_1} K_h (X_{i,l} - x_1) T_{il} T_{il'} \sum_{i=1}^{n} \hat{\epsilon}_{i,l}(X_{i,l}), (27)
\]

13
in which \( \hat{\varepsilon}^*_{1,l}(X_{i,l}) = \sum_{n=2}^{d_2} \hat{\varepsilon}^*_{al}(X_{ia}) \) and \( \hat{\varepsilon}^*_{al}(X_{ia}) \) is given in (21). If further one denotes
\[ \omega_{l,a,l',l''}(X_i, x_1) = T_{il}T_{il'}K_h(x_{il} - x_{1l})B_{l,a}(X_{ia}), \; \mu_{l,a,l',l''}(X_i, x_1) = E\omega_{l,a,l',l''}(X_i, x_1) \] (28)
then by (27), (A.2) and (21), \( \Psi_{v,l'}^2(x_1) \) can be rewritten as
\[ \Psi_{v,l'}^2(x_1) = n^{-1} \sum_{i=1}^{n} \sum_{l=1}^{d_1} \sum_{l'=1}^{d_2} \tilde{a}_{l,a,l'} \omega_{l,a,l',l''}(X_i, x_1). \] (29)

**LEMMA 1** Under Assumptions (A1) to (A5), (A7) to (A8), as \( n \to \infty \), \( \Psi_{v,l'}^1(x_1) \) defined in (26) satisfies \( \sup_{1 \leq l' \leq d_1} \sup_{x_1 \in [0,1]} \left| \Psi_{v,l'}^1(x_1) \right| = O_p \left( N \log n \right) \).

**LEMMA 2** Under Assumptions (A1) to (A5), (A7) to (A8), as \( n \to \infty \), \( \Psi_{v,l'}^2(x_1) \) defined in (27) satisfies \( \sup_{1 \leq l' \leq d_1} \sup_{x_1 \in [0,1]} \left| \Psi_{v,l'}^2(x_1) \right| = O_p \left( H^2 \right) \).

Proof of Proposition 3 is given in the Appendix, while Proposition 4 follows from Lemmas 1 and 2. Lemma 2 follows from Lemmas A.11 and A.12, both proved in the Appendix, while the proof of Lemma 1 is given in the Appendix. Similar result can be proved for \( \hat{m}_{SBLL,l'}(X_i, x_1) \) by extending \( \Psi_{v,l'}(x_1) \) and \( \Psi_{v,l'}(x_1) \) to terms such as
\[ \frac{1}{n^2} \sum_{i=1}^{n} K_h(x_{i1} - x_1) \left( \frac{x_{i1} - x_1}{h} \right) T_{il'} \sum_{l=1}^{d_1} \left\{ m_{-1,l}(X_{i1}) - \hat{m}_{-1,l}(X_{i1}) \right\} T_{il}, \]
\[ \frac{1}{n^2} \sum_{i=1}^{n} K_h(x_{i1} - x_1) \left( \frac{x_{i1} - x_1}{h} \right) T_{il'} \sum_{l=1}^{d_2} \hat{\varepsilon}_{-1,l}(X_{i1}) T_{il}, \]
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
\[ N = n = \min \left( \left[ n^{1/4} \log n \right] + 1, \left[ n/4d_1d_2 - 1/d_2 \right] - 1 \right) \]
which satisfies Assumption (A8), i.e.\( N = n \sim n^{1/4} \log n \), 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 \( \hat{m}_{SBLL,\alpha}(x_\alpha) \) depend not only on the functions \( b_{LL,l',l'',\alpha}(x_\alpha) \) and \( v_{l'',\alpha}(x_\alpha) \) but also crucially on the choice of bandwidths \( h_\alpha \). So we define for the second step of kernel smoothing, the optimal bandwidth of \( h_\alpha \), denoted
by \( h_{\alpha, \text{opt}} \), as the minimizer of the total asymptotic mean integrated squared errors (AMISE) of \( \{ \hat{m}_\alpha(x) \} \), which is defined as

\[
\text{AMISE} \{ \hat{m}_\alpha \} = \int \sum_{l' = 1}^{d_1} \left\{ \sum_{l = 1}^{d_1} b_{ll',l',\alpha}(x) h_{\alpha}^2 + v_{ll',\alpha}(x) \right\} f_{\alpha}(x) \, dx.
\]

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

\[
h_{\alpha, \text{opt}} = \left\{ \frac{n^{-1} \int \sum_{l' = 1}^{d_1} v_{ll',\alpha}(x) f_{\alpha}(x) \, dx}{4 \int \sum_{l' = 1}^{d_1} \left\{ \sum_{l = 1}^{d_1} b_{ll',l',\alpha}(x) \right\}^2 f_{\alpha}(x) \, dx} \right\}^{1/5},
\]

where \( 4 \int \sum_{l' = 1}^{d_1} \left\{ \sum_{l = 1}^{d_1} b_{ll',l',\alpha}(x) \right\}^2 f_{\alpha}(x) \, dx \) is approximated by

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

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

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

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

  where \( \min_{i} X_{il} = t_0 < \cdots < t_{N + 1} = \max_{i} X_{il} \).

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

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

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

  \[
  \sum_{i = 1}^{n} \left[ T_{il}T_{il'} \{ Y_i - \hat{m}(X_i, T_i) \}^2 - \left\{ \sum_{k = 0}^{3} a_{\alpha,l,k}X_{i\alpha}^{k} + \sum_{k = 4}^{N + 3} a_{\alpha,l,k} (X_{i\alpha} - t_{k-3}) \right\}^2 \right]^2.
  \]

- Density function \( f_{\alpha}(x) \) is estimated by \( \frac{1}{n} \sum_{i = 1}^{n} \sum_{k = 1}^{d_1} K_{h_{\alpha}}(X_{i\alpha} - x_{i\alpha}) \) and \( f'_{\alpha}(x) \) by

  \[
  - (nh_{\alpha}^2)^{-1} \sum_{i = 1}^{n} K' \left( \frac{X_{i\alpha} - x_{i\alpha}}{h_{\alpha}} \right) \text{ with the rule-of-the-thumb bandwidth } h_{\alpha}.
  \]
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 & Yang (2008). In this section we illustrate how the additive coefficient model is used to extend the Cobb-Douglas model for annual US GDP growth. Denoted by \( Q_t \) the US GDP at year \( t \), \( K_t \) the US capital at year \( t \), \( L_t \) the US labor at year \( t \), \( X_t \) the growth rate of ratio of R&D expenditure to GDP at year \( t \), all data have been downloaded from the Bureau of Economic Analysis (BEA) website for years, \( t = 1959, ..., 2002 \) (\( n = 44 \)). The standard Cobb-Douglas production function (Cobb & Douglas, 1928) is \( Q_t = A_t K_t^{\beta_1} L_t^{1-\beta_1} \) where \( A_t \) is the Total Factor Productivity (TFP) of year \( t \), \( \beta_1 \) is a parameter determined by technology. Define the following stationary time series variables

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

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

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

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

\[
Y_t - T_{2t} = \beta_0 + \beta_1 (T_{1t} - T_{2t}). \tag{30}
\]

As technology growth is one of the biggest sub-sections of TFP, it is reasonable to examine the dependence of both \( \beta_0 \) and \( \beta_1 \) 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 \( Y_t - T_{2t} = m_1 (X_t) + m_2 (X_t) \) \( (T_{1t} - T_{2t}) \) where \( m_l (X_t) = m_{0l} + \sum_{a=1}^{d_2} m_{al} (X_{t-a+1}) \), \( l = 1, 2 \), \( X_t = (X_1, ..., X_{t-d_2+1}) \). Using the BIC of Xue & Yang (2006b) for additive coefficient model with \( d_2 = 5 \), the following reduced model is considered optimal

\[
Y_t - T_{2t} = c_1 + m_{41} (X_{t-3}) + \{c_2 + m_{52} (X_{t-4})\} (T_{1t} - T_{2t}). \tag{31}
\]

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

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

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 \( 5.2399 \times 10^{-5} \), which is about 68% of the in sample ASE \( (7.6959 \times 10^{-5}) \) for model (30).
In model (31), $\hat{c}_1 + \hat{m}_{SBLL,41} (X_{t-3})$ 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 $X_{t-3} \leq 0.02$, but negative when $X_{t-3} > 0.02$. On average, the higher R&D investment spending causes faster GDP growing. However, overspending on R&D often leads to high losses (Culpepper, 2004 and Tokic, 2003).

We have also computed the average contribution of R&D to GDP growth for 1964-2001, which is about 40%. The GDP and estimated TFP growth rates is shown in Figure 3, it is obvious that TFP growth is highly correlated to the GDP growth. For more details, see Arnold (2005).

(Insert Figure 3 about here)

**REFERENCES**


**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 & Yang (2008).

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

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

where

\[
a_1 = \frac{2n}{q} + 2 \left( 1 + \frac{\varepsilon_n^2}{25m_2^2 + 5\varepsilon_n} \right), \quad a_2(k) = 11n \left( 1 + \frac{5m_k^{2k/(2k+1)}}{\varepsilon_n} \right).
\]
LEMMA A.2 (Xue & Yang, 2006b, Lemma A.2, Lemma A.5) There exists a constant $c_0 > 0$ such that for any sets of coefficients \( \{a_{0l}, a_{J,l}, 1 \leq J \leq N + 1, 1 \leq l \leq d_1, 1 \leq \alpha \leq d_2\} \),
\[
\left\| \sum_{l=1}^{d_1} \left( a_{0l} + \sum_{\alpha=1}^{d_2} \sum_{J=1}^{N+1} a_{J,\alpha} B_{J,\alpha} \right) t_l \right\|_2^2 \geq c_0 \sum_{l=1}^{d_1} \left( a_{0l}^2 + \sum_{\alpha=1}^{d_2} \sum_{J=1}^{N+1} a_{J,\alpha}^2 \right)
\]
and that as \( n \to \infty \), with probability approaching 1,
\[
\left\| \sum_{l=1}^{d_1} \left( a_{0l} + \sum_{\alpha=1}^{d_2} \sum_{J=1}^{N+1} a_{J,\alpha} B_{J,\alpha} \right) t_l \right\|_{2,n}^2 \geq c_0 \sum_{l=1}^{d_1} \left( a_{0l}^2 + \sum_{\alpha=1}^{d_2} \sum_{J=1}^{N+1} a_{J,\alpha}^2 \right).
\]

LEMMA A.3 Under Assumptions (A1) and (A8), one has: (i) there exist constants \( c_f, C_f, c_0(f) \) \( \) and \( C_0(f) \) depending on the marginal densities \( f_\alpha(x_\alpha) \), \( 1 \leq \alpha \leq d_2 \), such that \( c_f H \leq c_{J,\alpha} \leq C_f H \) and \( c_0(f) H \leq \|b_{J,\alpha}\|^2 \leq C_0(f) H \). (ii) uniformly for \( J, J' = 1, \ldots, N + 1 \)
\[
E \{ B_{J,\alpha}(X_\alpha) B_{J',\alpha}(X_\alpha) \} \sim \begin{cases} 1 & J' = J \\ -1/3 & |J' - J| = 1 \\ 1/6 & |J' - J| = 2 \end{cases}
\]
\[
E |B_{J,\alpha}(X_\alpha) B_{J',\alpha}(X_\alpha)|^k \sim \begin{cases} H^{1-k} & |J' - J| \leq 2 \\ 0 & |J' - J| > 2 \end{cases}, k \geq 1.
\]

LEMMA A.4 Under Assumption (A2), for \( V_T \) defined in (A.8) and \( S_T = V_T^{-1} \)
\[
c_Q c_n I_{d_1(d_2(N+1)+1)} \leq V_T \leq c_Q C I_{d_1(d_2(N+1)+1)} \leq S_T \leq c_Q C_s I_{d_1(d_2(N+1)+1)}.
\]

Proof. see Liu & Yang (2008). \( \square \)

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 \( m' \in \text{Lip}(0,1,C_\infty) \), there is a function \( g \in G^{(0)}[0,1] \) such that \( \|g - m\|_\infty \leq C_\infty \|m'\|_\infty H^2 \).

Lemma A.5 and Assumption (A3) ensure the existence of functions \( g_{\alpha l} \in G^{(0)}[0,1] \) such that
\[
\|g_{\alpha l} - m_{\alpha l}\|_\infty \leq C_\infty \|m'_\alpha\|_\infty H^2, \alpha = 1, \ldots, d_2, l = 1, \ldots, d_1.
\] (A.1)

A.2 Estimation of Constants

To closely examine terms \( \tilde{\varepsilon}_l(x) \) and \( \tilde{\varepsilon}_{\alpha l}(x_\alpha) \), we denote the following vector of coefficients
\[
\tilde{a} = \{\tilde{a}_{01}, \tilde{a}_{1,1,1}, \ldots, \tilde{a}_{N+1,d_2,1}, \tilde{a}_{02}, \tilde{a}_{1,1,2}, \ldots, \tilde{a}_{N+1,d_2,2}, \ldots, \tilde{a}_{0d_1}, \tilde{a}_{1,1,d_1}, \ldots, \tilde{a}_{N+1,d_2,d_1}\}^T
\]
such that the noise term \( \tilde{\varepsilon}_l(x) \) in (19) is expressed as
\[
(P_{n,l}E)(x) = \tilde{\varepsilon}_l(x) = \tilde{a}_{0l} + \sum_{\alpha=1}^{d_2} \sum_{J=1}^{N+1} \tilde{a}_{J,\alpha,l} B_{J,\alpha}(x_\alpha).
\] (A.3)
Equation (A.3) implies that 
\[ \hat{a} = (D^TD)^{-1} D^TE, \] 
where
\[ D = \{ D(X_1, T_1), \ldots, D(X_n, T_n) \}^T = \{ T_1 \otimes B(X_1), \ldots, T_n \otimes B(X_n) \}^T, \tag{A.4} \]
\[ B(x) = \{ 1, B_{1,1}(x_1), \ldots, B_{N+1,d_2}(x_{d_2}) \}^T, t = \{ t_1, \ldots, t_{d_1} \}^T. \tag{A.5} \]

Note that \( \hat{a} \) given in (A.2) can be rewritten as
\[ \hat{a} = \left( \frac{1}{n} D^TD \right)^{-1} \left( \frac{1}{n} D^T E \right) = (V_T + V_T^*)^{-1} \left( \frac{1}{n} D^T E \right), \tag{A.6} \]
where by (A.4)
\[ D^TD = \sum_{i=1}^{n} \left[ (T_i T_i^T) \otimes \{ B(X_i) B(X_i)^T \} \right], D^T E = \sum_{i=1}^{n} \left[ \{ T_i \otimes B(X_i) \} \sigma(X_i, T_i) \varepsilon_i \right], \tag{A.7} \]
and \( V_T^* \) is the difference between empirical and theoretical inner product matrices, i.e.
\[ V_T = E \left[ (TT)^T \otimes \{ B(X) B(X)^T \} \right] = E \left[ Q(X) \otimes \{ B(X) B(X)^T \} \right], \tag{A.8} \]
\[ V_T^* = \frac{1}{n} \sum_{i=1}^{n} \left[ (T_i T_i^T) \otimes \{ B(X_i) B(X_i)^T \} \right] - E \left[ Q(X) \otimes \{ B(X) B(X)^T \} \right]. \]

Now define \( \hat{a} = \{ \hat{a}_{01}, \hat{a}_{1,1,1}, \ldots, \hat{a}_{N,d_2,1}, \hat{a}_{02}, \hat{a}_{1,1,2}, \ldots, \hat{a}_{N,d_2,2}, \ldots, \hat{a}_{0d_1}, \hat{a}_{1,1,d_1}, \ldots, \hat{a}_{N,d_2,d_1} \}^T \) by replacing \((V_T+V_T^*)^{-1} \) with \( V_T^{-1} = S_T \) in the above formula, that is
\[ \hat{a} = V_T^{-1} (n^{-1} D^T E) = S_T (n^{-1} D^T E). \tag{A.9} \]

**LEMMA A.6** Under Assumptions (A1) to (A3), (A5) and (A8), as \( n \to \infty \)
\[ ||\hat{a}|| = O_p\left(n^{-1/2} N^{1/2} \log n \right), \tag{A.10} \]
\[ ||\hat{a} - \hat{a}|| = O_p\left(n^{-1} N^{3/2} \log^2 n \right), \tag{A.11} \]
\[ ||\hat{a}|| = O_p\left(n^{-1/2} N^{1/2} \log n \right). \tag{A.12} \]

**Proof.** By definition, \( \hat{a}^TD^TD\hat{a} = \hat{a}^TD^TD (D^TD)^{-1} D^TE = \hat{a}D^TE. \) Using (A.6), one has
\[ ||D\hat{a}||^2_{2,n} = n^{-1} \hat{a}^TD^TD\hat{a} = n^{-1} \hat{a}^TD^TE \leq ||\hat{a}|| ||n^{-1}D^TE||. \tag{A.13} \]

According to Lemma A.2,
\[ c_0 ||\hat{a}||^2 = c_0 \sum_{l} \left( a_{0l}^2 + \sum_{j,\alpha} a_{j,\alpha}^2 \right) \leq \sum_{l} \left( a_{0l} + \sum_{j,\alpha} a_{j,\alpha} B_{j,\alpha} \right) t_l \leq ||D\hat{a}||^2_{2,n}. \tag{A.14} \]

So \( ||\hat{a}|| \) is bounded by \( c_0^{-1} ||n^{-1}D^TE||. \) Bernstein’s inequality and truncation entail that \( ||n^{-1}D^TE||^2 = O_p \left( (\log n)^2 N/n \right) \), so (A.10) follows from (A.13) and (A.14). According to (A.6) and (A.9), one has \( V_T \hat{a} = (V_T + V_T^*) \hat{a} \), which implies that \( V_T \hat{a} = V_T (\hat{a} - \hat{a}). \)
One obtains from (A.22) and (A.23) \( \| \mathbf{V}_T (\hat{a} - \bar{a}) \| = \| \mathbf{V}_T \bar{a} \| \leq O_p \left( n^{-1/2} H^{-1} \log n \right) \| \bar{a} \| \). By (A.10) one has \( \| \mathbf{V}_T (\hat{a} - \bar{a}) \| \leq O_p \left( (\log n)^2 n^{3/2} \right) \). Thus according to Lemma A.4, one has \( \| \hat{a} - \bar{a} \| = O_p \left( n^{-1} N^{3/2} \log^2 n \right) \), which is (A.11). Then (A.12) follows (A.10) and (A.11).

**Lemma A.7** Under Assumptions (A1) to (A3), (A5) and (A8), as \( n \to \infty \)

\[
\sup_{1 \leq \ell \leq d} \left| \frac{1}{n} \sum_{i=1}^{n} T_{il} \sum_{a=1}^{d_1} \sum_{d=1}^{d_2} \hat{a}_a T_{il} \right| = O_p \left( n^{-1/2} \right). \tag{A.15}
\]

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

\[
\frac{1}{n} \sum_{i=1}^{n} T_{il} \sum_{a=1}^{d_1} \sum_{d=1}^{d_2} \hat{a}_a T_{il} = \frac{1}{n} \sum_{i=1}^{n} T_{il} \sum_{a=1}^{d_1} \sum_{d=1}^{d_2} \sum_{l=1}^{N+1} \hat{a}_{ia,l} B_{ja} (X_{ia}) T_{il} = \sum_{J,a,l} \hat{a}_{ia,l} \frac{1}{n} \sum_{i=1}^{n} T_{il} B_{ja} (X_{ia}) T_{il} = I_{V,1} + I_{V,2} + I_{V}
\]

where

\[
I_{V} = \sum_{J,a,l} \hat{a}_{ia,l} \left( \frac{1}{n} \sum_{i=1}^{n} T_{il} B_{ja} (X_{ia}) T_{il} - ET_{V} B_{ja} (X_{a}) T_{il} \right)
\]

Thus \( \sup_{J,a,l} \left| \frac{1}{n} \sum_{i=1}^{n} T_{il} B_{ja} (X_{ia}) T_{il} - ET_{V} B_{ja} (X_{a}) T_{il} \right| = O_p \left( n^{-1} N \log^2 n \right) \),

\[
I_{V,2} = \sum_{J,a,l} \hat{a}_{ia,l} ET_{V} B_{ja} (X_{ia}) T_{il} = (ET_{V} B_{ja} (X_{a}) T_{il})_{J,a,l} V^{-1} (n^{-1} D^T E).
\]

Direct computation (see Liu & Yang 2008 for details) yields that \( \text{var} (I_{V,2}) = O \left( n^{-1} \right) \), and therefore \( I_{V,2} = O_p \left( n^{-1/2} \right) \). So \( \left| I_{V,1} \right| + \left| I_{V,2} \right| = O_p \left( n^{-1/2} \right) . \tag{A.16} \)

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

\[
\sup_{J,a,l} \left| \frac{1}{n} \sum_{i=1}^{n} T_{il} B_{ja} (X_{ia}) T_{il} - ET_{V} B_{ja} (X_{a}) T_{il} \right| = O_p \left( n^{-1/2} \log n \right).
\]

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

\[
\sup_{J,a,l} \left| \frac{1}{n} \sum_{i=1}^{n} T_{il} B_{ja} (X_{ia}) T_{il} - ET_{V} B_{ja} (X_{a}) T_{il} \right| + \left| ET_{V} B_{ja} (X_{a}) T_{il} \right| = O \left( H^{1/2} \right).
\]

Then

\[
\left| I_{V} \right| \leq \| \hat{a} - \bar{a} \| \sqrt{(N+1) d_1 d_2} \sup_{J,a,l} \left| \frac{1}{n} \sum_{i=1}^{n} T_{il} B_{ja} (X_{ia}) T_{il} \right| = O_p \left( n^{-1} N^{3/2} \log^2 n \right). \tag{A.17}
\]

Now (A.15) follows from (A.16) and (A.17). The lemma is proved. \qed
LEMMA A.8 Under Assumptions (A1) to (A5), and (A8), as n → ∞

\[ n^{-1} \sum_{i=1}^{n} \left[ \sum_{l=1}^{n} \sum_{a=1}^{d_l} \{ \hat{m}_{al}(X_{ia}) - m_{al}(X_{ia}) \} T_{il} \right]^2 = O_p \left( n^{-1} \right) \tag{A18} \]

Proof. According to (A1), there exists \( g_{al} \in C(0) [0, 1] \) such that \( \|g_{al} - m_{al}\|_{\infty} = O \left( H^2 \right) = O \left( n^{-1/2} \right) \). According to Theorem 1.7 of Bosq (1998) p.36, \( n^{-1/2} \sum_{i=1}^{n} \{ T_{il}^2 - ET_{il}^2 \} \Rightarrow N \left( 0, \sigma^2 \right) \)

where \( \sigma^2 = \sum_{i=-\infty}^{\infty} \text{Cov} \left( T_{il}^2, T_{il}^2 \right) < \infty \) by applying Davydov’s Inequality [Bosq 1998, p.21, equation (1.10)]. Then \( n^{-1} \sum_{i=1}^{n} T_{il}^2 = ET_{il}^2 + O_p \left( n^{-1/2} \right) \). So

\[ n^{-1} \sum_{i=1}^{n} \left[ \sum_{l=1}^{n} \sum_{a=1}^{d_l} \{ \hat{m}_{al}(X_{ia}) - m_{al}(X_{ia}) \} T_{il} \right]^2 \leq n^{-1} \sum_{i=1}^{n} \left[ \sum_{l=1}^{n} \sum_{a=1}^{d_l} \|g_{al} - m_{al}\|_{\infty} T_{il} \right]^2 = O \left( n^{-1} \right) \left( n^{-1} \sum_{i=1}^{n} T_{il}^2 \right) = O_p \left( n^{-1} \right). \]

Proof of Propositions 1 and 2. According to (9), \( \hat{m}_0 - m_0 = \left( C_K^TC_K \right)^{-1} C_K^T \left( \hat{Y}_c - Y_c \right) \)

\( = \left( \frac{1}{n} C_K^T C_K \right)^{-1} \frac{1}{n} C_K^T \sigma (X_i, T_i) \varepsilon_i \). We know \( \frac{1}{n} C_K^T C_K = \left( \frac{1}{n} \sum_{i=1}^{n} T_{il}T_{il}' \right)_{l,l'=1}^{d_l} \). Then according to Theorem 1.7 of Bosq (1998) p.36, one has \( n^{-1/2} \sum_{i=1}^{n} \{ T_{il}T_{il'} - ET_{il}T_{il'} \} \Rightarrow N \left( 0, \sigma^2 \right) \)

where \( \sigma^2 = \sum_{i=-\infty}^{\infty} \text{Cov} \left( T_{il}T_{il'}, T_{il}T_{il'} \right) < \infty \). Therefore \( \frac{1}{n} C_K^T C_K = \left( ET_{il}T_{il'} \right)_{l,l'=1}^{d_l} + O_p \left( n^{-1/2} \right) \). Similarly, \( \frac{1}{n} C_K^T \sigma (X_i, T_i) \varepsilon_i = O_p \left( n^{-1/2} \right) \), implying \( \sup_{1 \leq i \leq d_1} \| \hat{m}_{al} - m_{al} \| = O_p \left( n^{-1/2} \right) \), which has completed the proof of Proposition 1.

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

\[ \hat{m}_0 - m_0 = \left( C_K^T C_K \right)^{-1} C_K^T \left( \hat{Y}_c - Y_c \right) \]

\[ = \left( C_K^T C_K \right)^{-1} C_K^T \left[ \sum_{a=1}^{d} \sum_{l=1}^{n} \{ \hat{m}_{al}(X_{ia}) - m_{al}(X_{ia}) \} T_{il} \right]^n \]

\[ = \left( C_K^T C_K \right)^{-1} C_K^T \left[ \sum_{a=1}^{d} \sum_{l=1}^{n} \{ \hat{m}_{al}(X_{ia}) - \hat{m}_{al}(X_{ia}) + \hat{m}_{al}(X_{ia}) - m_{al}(X_{ia}) \} T_{il} \right]^n \]

\[ = \frac{1}{n} \left( C_K^T C_K \right)^{-1} \frac{1}{n} C_K^T \left[ \sum_{a=1}^{d} \sum_{l=1}^{n} \{ \hat{m}_{al}(X_{ia}) - m_{al}(X_{ia}) \} T_{il} \right]^n \]

One has

\[ \left( \frac{1}{n} \sum_{i=1}^{n} T_{il} \right)_{l,l'=1}^{d_l} \leq O_p \left( n^{-1/2} \right) . \]

By Lemma A.8. Then the Proposition 2 follows (A15) and (A19). \( \square \)
A.3 Estimation of Function Components

Define

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

\[ A_{n,2} = \sup_{j',j,\alpha,l,l'} \left| \langle B_{j,\alpha} T_{i,l}, B_{j',\alpha} T_{i,l'} \rangle_{2,n} - \langle B_{j,\alpha} T_{i,l}, B_{j',\alpha} T_{i,l'} \rangle_2 \right|, \]

\[ A_{n,3} = \sup_{j',j,\alpha \neq \alpha',l,l'} \left| \langle B_{j,\alpha} T_{i,l}, B_{j',\alpha'} T_{i,l'} \rangle_{2,n} - \langle B_{j,\alpha} T_{i,l}, B_{j',\alpha'} T_{i,l'} \rangle_2 \right|. \tag{A.20} \]

**Lemma A.9** Under Assumptions (A1) to (A3), and (A8), as \( n \to \infty \)

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

\[ A_{n,2} = O_p \left( n^{-1/2} H^{-1/2} \log n \right), \tag{A.22} \]

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

**Proof.** See Liu & Yang (2008). \( \square \)

**Lemma A.10** Under Assumptions (A1) to (A3), (A5), and (A7) to (A8), as \( n \to \infty \)

\[ \sup_{x_1 \in [0,1]} \left| \mu_{\omega_{j,\alpha,l,l'}}(x_1) \right| = O \left( H^{1/2} \right), \tag{A.24} \]

\[ \sup_{x_1 \in [0,1]} \left| \left\{ \omega_{j,\alpha,l,l'}(X_{i,1}) - \mu_{\omega_{j,\alpha,l,l'}}(x_1) \right\} \right| = O_p \left( \log n/\sqrt{n} \right), \tag{A.25} \]

where \( \omega_{j,\alpha,l,l'}(X_{i,1}) \) and \( \mu_{\omega_{j,\alpha,l,l'}}(x_1) \) defined in (28), hence

\[ \sup_{x_1 \in [0,1]} \left| \left\{ \omega_{j,\alpha,l,l'}(X_{i,1}) - \mu_{\omega_{j,\alpha,l,l'}}(x_1) \right\} \right| = O_p \left( H^{1/2} \right). \tag{A.26} \]

**Proof.** See Liu & Yang (2008). \( \square \)

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

\[ \hat{\Psi}_{v,v'}^{(2)}(x_1) = n^{-1} \sum_{i=1}^{n} \sum_{l=1}^{d_1} \sum_{J=1}^{N+1} \sum_{\alpha=2}^{d_2} \hat{a}_{j,\alpha,l} \omega_{j,\alpha,l,l'}(X_{i,1}). \tag{A.27} \]

**Lemma A.11** Under Assumptions (A1) to (A3), (A5) and (A8), as \( n \to \infty \),

\[ \sup_{1 \leq l' \leq d_1} \sup_{x_1 \in [0,1]} \left| \Psi_{v,v'}^{(2)}(x_1) - \hat{\Psi}_{v,v'}^{(2)}(x_1) \right| = O_p \left( H^2 \right). \]

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

\[ \left| \Psi_{v,v'}^{(2)}(x_1) - \hat{\Psi}_{v,v'}^{(2)}(x_1) \right| = \sum_{l=1}^{d_1} \sum_{J=1}^{N+1} \sum_{\alpha=2}^{d_2} \left| \hat{a}_{j,\alpha,l} - \hat{a}_{j,\alpha,l} \right| \frac{1}{n} \sum_{i=1}^{n} \omega_{j,\alpha,l}(X_{i,1}) T_{i,l}. \]
According to (A.26) and (A.11), Cauchy-Schwartz inequality implies that
\[
\sup_{1 \leq l' \leq d_1} \sup_{x \in [0,1]} \left| \Psi_{e,l'}(x_1) - \hat{\Psi}_{e,l'}(x_1) \right| \leq \sqrt{N} + O_p \left( \frac{(\log n)^2}{nH^{3/2}} \right) O_p \left( H^{1/2} \right) = O_p \left( \frac{(\log n)^2}{nH^{3/2}} \right).
\]
Therefore the lemma follows. \hfill \square

**Lemma A.12** Under Assumptions (A1) to (A3), (A5) and (A8), as \( n \to \infty \)
\[
\sup_{1 \leq l' \leq d_1} \sup_{x \in [0,1]} \left| \Psi_{e,l'}(x_1) \right| = \sup_{1 \leq l' \leq d_1} \sup_{x \in [0,1]} \left| \Psi_{e,l'}(x_1) \right| = O_p \left( n^{-2/5} \right).
\]

**Proof.** See Liu & Yang (2008). \hfill \square

**Proof of Proposition 3.** (A.1) implies that
\[
|E_n g_{j,l}(X_{i,l})| \leq |E_n g_{j,l}(X_{i,l}) - E_n m_{j,l}(X_{i,l})| + |E_n m_{j,l}(X_{i,l})| \leq C_\infty (d_2 - 1) \sup_{2 \leq \alpha \leq d_2} \|m'_{\alpha,l}\|_\infty H^2 + O_p \left( n^{-1/2} \right).
\]

By definition (24), \( \sup_{x \in [0,1]} |\Psi_{b,l}(x_1)| \leq R_1 + R_2 + R_3 \) where
\[
R_1 = \sup_{x \in [0,1]} \left| \frac{1}{n} \sum_{i=1}^n K_h(X_{i,l} - x_1) \sum_{l=1}^{d_1} \left\{ m_{j,l}(X_{i,l}) - g_{j,l}(X_{i,l}) \right\} T_{il} T_{il'} \right|,
\]
\[
R_2 = \sup_{x \in [0,1]} \left| \frac{1}{n} \sum_{i=1}^n K_h(X_{i,l} - x_1) \sum_{l=1}^{d_1} \left\{ g_{j,l}(X_{i,l}) - E_n g_{j,l}(X_{i,l}) - \hat{m}_{j,l}(X_{i,l}) \right\} T_{il} T_{il'} \right|,
\]
\[
R_3 = \sup_{x \in [0,1]} \left| \frac{1}{n} \sum_{i=1}^n K_h(X_{i,l} - x_1) \sum_{l=1}^{d_1} E_n g_{j,l}(X_{i,l}) T_{il} T_{il'} \right|.
\]

For \( R_1 \), using (A.1), one has
\[
R_1 \leq C_\infty (d_2 - 1) \sup_{2 \leq \alpha \leq d_2} \|m'_{\alpha,l}\|_\infty H^2 \sum_{l=1}^{d_1} \frac{1}{n} \sum_{i=1}^n |T_{il} T_{il'}| = O_p \left( H^2 \right) \left\{ \sum_{l=1}^{d_1} E \left| T_{il} T_{il'} \right| + O_p \left( n^{-1/2} \right) \right\} = O_p \left( H^2 \right). \tag{A.29}
\]

To bound \( R_2 \), denote the empirically centered spline basis as \( B_{j,l}^* (x_{i,l}) = B_{j,l} (x_{i,l}) - E_n B_{j,l} (x_{i,l}) \), \( 1 \leq J \leq N + 1, 1 \leq \alpha \leq d_2 \). Then one can write for some \( (\hat{\alpha}_{i,l,j}, \hat{\alpha}_{i,l,j}) \) \( J,l = 1 \)
\[
\hat{m}_{l}(x) - m_{0,l} - \sum_{\alpha=1}^{d_2} g_{\alpha,l}(x) + \sum_{\alpha=1}^{d_2} E_n g_{\alpha,l}(x) = \hat{\alpha}_{i,l} + \sum_{\alpha=1}^{d_2} \sum_{l=1}^{N+1} \hat{\alpha}_{i,l,j} B_{j,l}^* (x_{i,l}).
\]
Thus
\[
R_2 = \sup_{x \in [0,1]} \left| \frac{1}{n} \sum_{i=1}^n K_h(X_{i,l} - x_1) \sum_{l=1}^{d_1} \sum_{\alpha=1}^{d_2} \sum_{l=1}^{N+1} \hat{\alpha}_{i,l,j} B_{j,l}^* (x_{i,l}) T_{il} T_{il'} \right|.
\]
Thus equation (A.21) of Lemma A.9 states that
\[
\sup_{1 \leq J \leq N+1, 1 \leq d_1, 2 \leq d_2} \left| \sum_{i=1}^{n} K_h (X_{i1} - x) T_i T_i' B_{J,\alpha}(X_{i\alpha}) \right| \leq \sum_{l=1}^{d_1} \sum_{\alpha=2}^{d_2} \sum_{J=1}^{N+1} |\tilde{a}_{J,\alpha,l}| \left[ \sup_{1 \leq J \leq N+1, 1 \leq d_1, 2 \leq d_2} \left| n^{-1} \sum_{i=1}^{n} K_h (X_{i1} - x) T_i T_i' B_{J,\alpha}(X_{i\alpha}) \right| \right].
\]

Similarly, equation (A.21) of Lemma A.9 states that \( \sup_{1 \leq J \leq N+1} |E_n B_{J,\alpha}(X_{i\alpha})| = O_p (\log n / \sqrt{n}) \) and standard kernel argument shows that
\[
\sup_{x_1 \in [0,1]} \sup_{1 \leq J \leq N+1, 1 \leq l', \leq d_1, 2 \leq d_2} \left| n^{-1} \sum_{i=1}^{n} K_h (X_{i1} - x_1) T_i T_i' \{ B_{J,\alpha}(X_{i\alpha}) - E_n B_{J,\alpha}(X_{i\alpha}) \} \right| = O_p \left( H^{1/2} \right),
\]

while equation (A.24) in Lemma A.10 states that
\[
\sup_{x_1 \in [0,1]} \sup_{1 \leq J \leq N+1, 1 \leq l', \leq d_1, 2 \leq d_2} \left| n^{-1} \sum_{i=1}^{n} K_h (X_{i1} - x_1) T_i T_i' \sum_{\alpha=1}^{d_2} g_{\alpha l}(x) + \sum_{\alpha=1}^{d_2} E_n g_{\alpha l}(x) \right| = O_p (1).
\]

Therefore, one has
\[
R_2 \leq \left\{ (N + 1) d_1 (d_2 - 1) \sum_{l=1}^{d_1} \sum_{\alpha=2}^{d_2} \sum_{J=1}^{N+1} (\tilde{a}_{J,\alpha,l})^2 \right\}^{1/2} \left\{ O_p \left( H^{1/2} \right) + O_p \left( \log n / \sqrt{n} \right) \right\} = O_p \left( n^{-1/2} + H^2 \right).
\]

The last step follows from
\[
\left\| \tilde{m}_l(x) - m_{0l} - \sum_{\alpha=1}^{d_2} g_{\alpha l}(x) + \sum_{\alpha=1}^{d_2} E_n g_{\alpha l}(x) \right\|_2 \leq \left\| \tilde{m}_l(x) - m_l(x) \right\|_2 + \left\| m_l(x) - m_{0l} - \sum_{\alpha=1}^{d_2} g_{\alpha l}(x) \right\|_2 + \left\| \sum_{\alpha=1}^{d_2} E_n g_{\alpha l}(x) \right\|_2 \leq 3C_\infty \sum_{\alpha=1}^{d_2} \left\| m_{\alpha l}' \right\|_{\infty} H^2 + O_p \left( n^{-1/2} \right).
\]

Thus \( R_2 = O_p (n^{-1/2} + H^2) \). Similarly,
\[
R_3 = \sup_{x_1 \in [0,1]} \left| \frac{1}{n} \sum_{i=1}^{n} K_h (X_{i1} - x_1) \sum_{l=1}^{d_1} E_n g_{l,1,l}(X_{i,1}) T_i T_i' \right| \leq \left\{ \sum_{l=1}^{d_1} \left| E_n g_{l,1,l}(X_{i,1}) \right| \right\} \sup_{x_1 \in [0,1]} \left| \frac{1}{n} \sum_{i=1}^{n} K_h (X_{i1} - x_1) T_i T_i' \right|.
\]
\[
\leq \left\{ \sum_{l=1}^{d_1} \left| E_{n} g_{j,l} (X_{i,1}) \right| \right\} \sup_{x_1 \in [0,1]} \left| \frac{1}{n} \sum_{i=1}^{n} K_h (X_{i1} - x_1) T_{il} T_{il}' \right| = O_p \left( \frac{n^{-1/2} + H^2}{\sqrt{2}} \right). \tag{A.31}
\]

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

**Proof of Lemma 1.** Based on formula (21), \( n^{-1} \sum_{i=1}^{n} \tilde{\varepsilon}_{i,l} (X_{i,1}) \) is

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

Lemma A.6 implies that

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

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

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

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

Thus the lemma follows immediately from (A.32) and (26). \hfill \Box
Figure 1: Errors of GDP forecasts: model (31)–solid line; model (30)–dotted line.
Figure 2: Estimation of function $\hat{c}_1 + \hat{m}_{SBLL,41}(X_{t-3})$. 
GDP and estimated TFP growth rates

Figure 3: Estimation of function $\hat{c}_1 + \hat{m}_{SBLL,41}(X_{t-3})$: GDP growth rate—dotted line; $\hat{c}_1 + \hat{m}_{SBLL,41}(X_{t-3})$—solid line.