应用概率统计 第二十六卷
第二期 2010年4月

Chinese Journal of Applied Probability
and Statistics Vol.26 No.2 Apr. 2010

The Multivariate Partially Linear Model with B-Spline *

Chen Guanglei
(*Department of Basic Sciences, the Chinese People's Armed Police Forces Academy, Langfang, 065000*)
Wang Zhaojun⋆
(*LPMC and School of Mathematical Sciences, Nankai University, Tianjin, 300071*)

Abstract

A multivariate partially linear model is considered in this paper. The B-spline least squares estimator for both the parametric and the noparametric components is proposed. Moreover, we investigate the the asymptotic normality of the estimator of the parametric component and the convergence rate of the estimator of nonparametric function.

Keywords: Multivariate regression, partially linear model, B-spline, asymptotic normality.
AMS Subject Classification: 62G05, 62G08.

§1. Introduction

The general partially linear model is given by

$$Y = X^T\beta + g(T) + e, \tag{1.1}$$

where X and T are explanatory variables, β is a vector of parameters, $g(\cdot)$ is an smooth function of T, e is random error. The model was introduced by Engle et al. (1986) and further studied by Heckman (1986), Chen (1988), Speckman (1988) and You et al. (2004). Some applications of the partially linear models have been described in the literature (Härdle et al. (2000)).

All the models under consideration are univariate models. In some applications, it may be of interest to work with a multidimensional response variable. For example, in finance, it is now widely accepted that, working with series, such as assert returns, in a multidimensional framework leads to better results than work with separate univariate model. In this paper, we consider the multivariate partially linear model

$$Y_i = X_i^T\mathbf{H} + \mathbf{g}(T_i) + e_i, \qquad i = 1, 2, \cdots, n, \tag{1.2}$$

*The project supported by the Natural Sciences Foundations of China (10771107) and Tianjin (07JCYBJC04 300).

⋆Corresponding author, E-mail: zjwang@nankai.edu.cn.

Received December 24, 2007.

where $Y_i = (y_{i1}, \cdots, y_{id})$, $X_i = (x_{i1}, \cdots, x_{ip})^T$ and T_i be ranges over a nondegenerate compact 1-dimensional interval D, the error $e_i = (e_{i1}, \cdots, e_{id})$ are assumed to be independent and identically distributed with mean $\mathbf{0}$ and variance-covariance Σ, (X_i, T_i) and e_i are independent, $\mathbf{H} = (\beta_1, \cdots, \beta_d)$ is the $p \times d$ matrix of unknown parameters and $\mathbf{g}(T) = (g_1(T), \cdots, g_d(T))$ is the $1 \times d$ vector of unknown functions. For simplicity, let

$$
\begin{aligned}
\mathbf{Y} &= (\mathbf{y}_1, \cdots, \mathbf{y}_d) = (Y_1^T, \cdots, Y_n^T)^T,\\
\mathbf{X} &= (\mathbf{x}_1, \cdots, \mathbf{x}_p) = (X_1, \cdots, X_n)^T,\\
\mathbf{G} &= (\mathbf{g}_1, \cdots, \mathbf{g}_d) = (\mathbf{g}(T_1)^T, \cdots, \mathbf{g}(T_n)^T)^T,\\
\mathbf{e} &= (\mathbf{e}_1, \cdots, \mathbf{e}_d) = (e_1^T, \cdots, e_n^T)^T,
\end{aligned}
$$

matrix form of the model (1.2) is

$$\mathbf{Y} = \mathbf{XH} + \mathbf{G} + \mathbf{e}. \tag{1.3}$$

Using Vec operator and the Kroneker products (see Christenson (1996)), the model can be rewritten as

$$\mathrm{Vec}(\mathbf{Y}) = [I_d \otimes \mathbf{X}]\mathrm{Vec}(\mathbf{H}) + \mathrm{Vec}(\mathbf{G}) + \mathrm{Vec}(\mathbf{e}). \tag{1.4}$$

Beatriz et al. (2006) considered a kernel estimation for the model (1.3). It is well known that kernel estimation, which is a local smoothing method, is a popular nonparametric smoothing technique. However, kernel type methods can be quite computationally expensive because they require re-fitting at every point where the fitted function needs to be evaluated. In this paper, we propose a spline estimation for the model (1.3). The attraction of the spline based global smoothing is that it is closely to parametric model and thus it reduces the computation substantially.

The rest of this paper is organized as follows. Section 2 presents the B-spline estimates for the model (1.3) based on global smoothing method. Section 3 states the main results. Section 4 gives some concluding remarks. Mathematical proofs are obtained in Appendix A.

§2. B-Spline Estimation

As in most works on nonparametric smoothing, without lose of generality, the estimation of the function vector $\mathbf{g}(\cdot)$ is conducted on compact set $D = [0, 1]$.

Let N_n denote a positive integer and $t_i = i/N_n$, $i = 0, 1, \cdots, N_n$ be the knot sequence. A polynomial spline of degree $m \geq 0$ on D with knot sequence $t_0, \cdots, t_{N_n}$ is a function that is a polynomial of degree m on each of the intervals $[t_i, t_{i+1})$, $i = 0, \cdots, N_n - 2$, and $[t_{N_n-1}, t_{N_n}]$, and globally has $m-1$ continuous derivatives for $m \geq 1$. A piecewise constant function, linear spline and quadratic spline correspond to $m = 0, 1, 2$, respectively. The collection of spline functions of a particular degree and knot sequence forms a linear function space and it is easy to construct a convenient basis for it. For example, the space of splines with degree 3 and knots sequence $t_0, \cdots, t_{N_n}$ form a linear space of dimension $N_n + 3$. The truncated power basis for this space is $1, t, t^2, t^3, (t-t_1)_+^3, \cdots, (t - t_{N_n-1})_+^3$, where $(t-t_i)_+ = \max\{0, t-t_i\}$, $i = 0, \cdots, N_n - 1$. A basis with better numerical properties is the B-spline basis. See de Boor (2001) and Györfi et al.(2002) for the detail of B-spline.

Let S_{m,N_n} be a space of polynomial splines on D with a fixed degree m and knots $t_0, \cdots, t_{N_n}$, $K = N_n + m$ denote the dimension of S_{m,N_n}. Under some smooth conditions, g_s can be approximated well by a spline function g_s^* in the sense that $\sup\limits_{t\in D} |g_s(t) - g_s^*(t)| \to 0$ as the number of knots of the spline tends to infinity, see de Boor (2001). Let $B_k(\cdot)$ $(k = 1, \cdots, K)$ be a set of basis functions of S_{m,N_n}. Then, exist a set of constants θ_{ks}, $k = 1, \cdots, K$, $s = 1, \cdots, d$, such that

$$g_s(t) \approx g_s^*(t) = \sum_{k=1}^{K} \theta_{ks} B_k(t).$$

For the multivariate partially linear model (1.2), we estimate the parameters β_s and $\theta_s = (\theta_{1s}, \cdots, \theta_{Ks})^T \in R^K$, $s = 1, \cdots, d$, through minimizing

$$L(\mathbf{H}, \mathbf{\Theta}) = \sum_{s=1}^{d} \sum_{i=1}^{n} \Big\{ Y_{is} - X_i^T \beta_s - \sum_{k=1}^{K} \theta_{ks} B_k(T_i) \Big\}^2 \tag{2.1}$$

with respect to β_s and θ_s. The obtained estimates of β_s and θ_s are denoted by $\widehat{\beta}_s = (\widehat{\beta}_{1s}, \cdots, \widehat{\beta}_{ps})^T$ and $\widehat{\theta}_s = (\widehat{\theta}_{1s}, \cdots, \widehat{\theta}_{Ks})^T$, respectively. Then, $g_s(t)$ is estimated by $\widehat{g}_s(t) = \sum\limits_{k=1}^{K} \widehat{\theta}_{ks} B_k(t)$. Therefore, $\mathbf{H}$ and $\mathbf{g}(\cdot)$ can be estimated by $\widehat{\mathbf{H}} = (\widehat{\beta}_1, \cdots, \widehat{\beta}_d)$, $\widehat{\mathbf{g}}(t) = (\widehat{g}_1(t), \cdots, \widehat{g}_d(t))$. We refer to $\widehat{\beta}_s$ and $\widehat{g}_s(\cdot)$, $s = 1, \cdots, d$ as the least squares spline estimates in this paper.

For convenience and simplicity, some notations are introduced.

$$\mathbf{B} = \begin{bmatrix} B_1(T_1) & B_2(T_1) & \cdots & B_K(T_1) \\ B_1(T_2) & B_2(T_2) & \cdots & B_K(T_2) \\ \vdots & \vdots & \ddots & \vdots \\ B_1(T_n) & B_2(T_n) & \cdots & B_K(T_n) \end{bmatrix}, \qquad P_{\mathbf{B}} = \mathbf{B}(\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T.$$

Then,

$$L(\mathbf{H},\mathbf{\Theta}) = [\mathrm{Vec}(\mathbf{Y}) - (I_d\otimes\mathbf{X})\mathrm{Vec}(\mathbf{H}) - (I_d\otimes\mathbf{B})\mathrm{Vec}(\mathbf{\Theta})]^T \cdot [\mathrm{Vec}(\mathbf{Y}) - (I_d\otimes\mathbf{X})\mathrm{Vec}(\mathbf{H}) - (I_d\otimes\mathbf{B})\mathrm{Vec}(\mathbf{\Theta})], \tag{2.2}$$

where I_d is the $d\times d$ identify matrix. Recalling the properties of Kronecker products

(I) $[\mathbf{A}\otimes\mathbf{B}][\mathbf{C}\otimes\mathbf{D}] = [\mathbf{AC}\otimes\mathbf{BD}]$,

(II) $[\mathbf{A}\otimes\mathbf{B}]^T = [\mathbf{A}^T\otimes\mathbf{B}^T]$,

(III) $([I_d\otimes\mathbf{A}])^{-1} = [I_d\otimes\mathbf{A}^{-1}]$,

we get the estimators $\widehat{\mathbf{H}}$ and $\widehat{\mathbf{\Theta}}$ as follows:

$$\begin{aligned}\mathrm{Vec}(\widehat{\mathbf{H}}) &= [I_d\otimes([\mathbf{X}^T(I_n-P_{\mathbf{B}})\mathbf{X}]^{-1}\mathbf{X}^T(I_n-P_{\mathbf{B}}))]\mathrm{Vec}(\mathbf{Y}),\\ \mathrm{Vec}(\widehat{\mathbf{\Theta}}) &= [I_d\otimes((\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T)](\mathrm{Vec}(\mathbf{Y}) - [I_d\otimes\mathbf{X}]\mathrm{Vec}(\widehat{\mathbf{H}})),\end{aligned} \tag{2.3}$$

which matrix forms are given by

$$\begin{aligned}\widehat{\mathbf{H}} &= [\mathbf{X}^T(I_n-P_{\mathbf{B}})\mathbf{X}]^{-1}\mathbf{X}^T(I_n-P_{\mathbf{B}})\mathbf{Y},\\ \widehat{\mathbf{\Theta}} &= (\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T(\mathbf{Y}-\mathbf{X}\widehat{\mathbf{H}}),\end{aligned} \tag{2.4}$$

which also are equivalent to, for $s=1,\cdots,d$,

$$\begin{aligned}\widehat{\beta}_s &= [\mathbf{X}^T(I_n-P_{\mathbf{B}})\mathbf{X}]^{-1}\mathbf{X}^T(I_n-P_{\mathbf{B}})\mathbf{y}_s,\\ \widehat{\theta}_s &= (\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T(\mathbf{y}_s-\mathbf{X}\widehat{\beta}_s).\end{aligned} \tag{2.5}$$

§3. Main Results

In this section, we focus on the asymptotic theories of our proposed estimators on interval $D=[0,1]$. For any real-valued function g on D, define

$$\|g\|_\infty = \sup_{t\in D}|g(t)| \qquad \text{and} \qquad \|g\|_2 = \Big\{\int_D g^2(t)\mathrm{d}t\Big\}^{1/2}.$$

For two sequences of positive number a_n and b_n, we write $a_n\preceq b_n$ if a_n/b_n is uniformly bounded and $a_n\asymp b_n$ if $a_n\preceq b_n$ and $b_n\preceq a_n$. Let $\xrightarrow{\mathrm{P}}$ and $\xrightarrow{L}$ denote the convergence in probability and distribution, respectively. Define $h=1/N_n$, and $|\cdot|$ denote either the Euclidian norm of a vector or the absolute value of a real number according to the context.

The following conditions are needed for the statement of the main results.

(C1) The distribution of T is absolutely continuous and its density is bounded away from 0 and ∞ on D.

(C2) Let l, γ and M denote real constants such that $0 < \gamma \le 1$ and $0 < M$; g_s is an l-times continuously differentiable function such that

$$|g_s^{(l)}(t') - g_s^{(l)}(t)| \le M|t - t'|^{\gamma}, \qquad \text{for } \ 0 \le t, \ t' \le 1.$$

Think of $q = l + \gamma$ as a measure of the smoothness of the function g_s, for $s = 1, \cdots, d$.

(C3) $\mathbf{V}_t = \mathsf{Cov}\,(X|T = t)$, there exist positive definite matrices $\mathbf{V}_{00}$ and $\mathbf{V}_{01}$ such that both $\mathbf{V}_t - \mathbf{V}_{00}$ and $\mathbf{V}_{01} - \mathbf{V}_t$ are nonnegative definite for all $t \in D$.

(C4) The condition mean functions $\mu_i(t) = \mathsf{E}(x_i|T = t)$, $(i = 1, ..., p)$, are all bounded and continuous on D.

(C5) The number of knots $N_n \asymp n^r$, with $1/(2q + 1) \le r < 1/2$.

Remark 3.1 Condition (C1), (C2), (C4), (C5) are common in nonparametric regression literature, see for example of Chen (1988), Huang (2003). The Condition (C3) is standard for asymptotic normality.

Set $v_{ij} = \mathsf{Cov}\,(x_i - \mu_i(T), x_j - \mu_j(T)) = \mathsf{Cov}\,(x_i, x_j) - \mathsf{Cov}\,(\mu_i(T), \mu_j(T))$, for $1 \le i$, $j \le p$, and $\mathbf{V} = (v_{ij})_{p\times p}$.

Under the above conditions, we can state our main results.

Theorem 3.1 Suppose that conditions (C1)–(C5) hold and that $m \ge q$. Then

$$\sqrt{n}(\mathrm{Vec}(\widehat{\mathbf{H}} - \mathbf{H})) \xrightarrow{L} N_{p\times d}(0, \Sigma \otimes \mathrm{V}^{-1}).$$

In particular, for each $s \in \{1, \cdots, d\}$,

$$\sqrt{n}(\widehat{\beta}_s - \beta_s) \xrightarrow{L} N_p(0, \sigma_{ss}\mathrm{V}^{-1}).$$

Theorem 3.2 Suppose that conditions (C1)–(C5) hold and that $m \ge q$. Then

$$\|\widehat{g}_s - g_s\|_2^2 = O_p((nh)^{-1} + h^{2q}), \qquad \text{for } \ s = 1, \cdots, d.$$

Remark 3.2 Theorem 3.1 not only constructs the asymptotically normal estimator of $\mathbf{H}$ but also provides a test statistics for testing $H_0 : \mathbf{H} = \mathbf{H}_0$. Theorem 3.2 shows that $\widehat{g}_s$ is consistent in estimating g_s, that is, $\|\widehat{g}_s - g_s\|_2 = o_p(1)$. If $q > 1/2$, Theorem 3.2 has the optimal order of h as $n^{-1/(2q+1)}$, in which case $\|\widehat{g}_s - g_s\|_2 = O_p(n^{-q/(2q+1)})$, which is the same optimal global convergence rates of estimators for nonparametric regression (Stone (1985)).

§4. Concluding Remarks

The results have been obtained assuming that the design points are random. When the design points are fixed, all the results also hold. In this paper we used B-splines with equally spaced knots. It is interesting to consider free-knot splines in the current context. We have focused on the case where T is one-dimensional. When T is multi-dimensional, the proposed method is conceptually applicable, where tensor product splines can be used. However, actual implementation may require substantial further development.

Appendix: Proofs of the Main Results

The spline estimates $\widehat{\mathbf{H}}$ and $\widehat{\mathbf{g}}(\cdot)$ are uniquely determined by the function space S_{m,N_n} and different sets of basis functions can give the same estimates $\widehat{\mathbf{H}}$ and $\widehat{\mathbf{g}}(\cdot)$, so we employ the B-spline basis in our proofs for convenience. However, the results do not depend on the choice of basis.

Set $B_k = K^{1/2}\varphi_k$, $k = 1, \cdots, K$, where φ_k are B-splines as defined in Chapter 5 of [6]. There are positive constants M_1 and M_2 such that

$$M_1|\theta|^2 \le \int \Big\{ \sum_{k=1}^{K} \theta_k B_k(t) \Big\}^2 \mathrm{d}z \le M_2|\theta|^2, \tag{A.1}$$

where $\theta = (\theta_1, \cdots, \theta_K)^T$ (see Theorem 4.2 of Chapter 5 of [6]).

We need the following lemmas for proving the main results.

Lemma A.1 Under the Conditions (C1) and (C5), there is an interval $[M_3, M_4]$ with $0 < M_3 < M_4 < \infty$, such that

$$\mathsf{P}\Big\{\text{all the eigenvalues of } \frac{1}{n}\mathbf{B}^T\mathbf{B} \text{ fall in } [M_3, M_4]\Big\} \to 1, \qquad \text{as } n \to \infty.$$

Proof The proofs see Burman (1999). □

Lemma A.2 If g satisfies condition (C2), then exists a constant $C > 0$ and a spline function $g^* \in S_{m,N_n}$ with $m \ge q$, such that

$$\|g - g^*\|_\infty \le Ch^q. \tag{A.2}$$

Proof This result is due to Theorem XII.1 in de Boor (2001). □

Set $\varepsilon_{ij} = x_{ij} - \mu_j(T_i)$, $\varepsilon = (\varepsilon_1, \cdots, \varepsilon_p)$, $\varepsilon_j = (\varepsilon_{1j}, \cdots, \varepsilon_{nj})^T$, $\mu(T) = (\mu_1(T), \cdots, \mu_p(T))$ and $\mu_j(T) = (\mu_j(T_1), \cdots, \mu_j(T_n))^T$ for $1 \le i \le n$, $1 \le j \le p$.

Lemma A.3 Under the conditions (C1), (C3)–(C5), then as $n \to \infty$,

$$\frac{1}{n}\mathbf{X}^T(I_n - P_{\mathbf{B}})\mathbf{X} \xrightarrow{\text{P}} \mathbf{V}. \tag{A.3}$$

Proof Observing that the (i,j)th matrix element, we obtain

$$\begin{aligned}(\mathbf{X}^T(I_n - P_{\mathbf{B}})\mathbf{X})_{ij} &= (\varepsilon_i + \mu_i(T))^T(I_n - P_{\mathbf{B}})(\varepsilon_j + \mu_j(T)) \\ &= \varepsilon_i^T(I_n - P_{\mathbf{B}})\varepsilon_j + \varepsilon_i^T(I_n - P_{\mathbf{B}})\mu_j(T) \\ &\quad + \mu_i(T)^T(I_n - P_{\mathbf{B}})\varepsilon_j + \mu_i(T)^T(I_n - P_{\mathbf{B}})\mu_j(T).\end{aligned} \tag{A.4}$$

It follows from Condition (C1) and (A.1) that $\mathsf{E}[B_k(t)]^2 \asymp 1$ and Lemma A.1, we have

$$\begin{aligned}\mathsf{E}(|\varepsilon_i^T P_{\mathbf{B}}\varepsilon_i||P_{\mathbf{B}}) &= \mathsf{E}(\varepsilon_i^T \mathbf{B}(\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T\varepsilon_i|P_{\mathbf{B}}) \\ &\asymp \frac{1}{n}\mathsf{E}(\varepsilon_i^T\mathbf{B}\mathbf{B}^T\varepsilon_i|P_{\mathbf{B}}) \\ &= \frac{1}{n}\mathsf{E}\Big\{\sum_{k=1}^{K}\Big(\sum_{l=1}^{n}B_k(T_l)\varepsilon_{li}\Big)^2\Big|P_{\mathbf{B}}\Big\} \\ &= \frac{1}{n}\sum_{k=1}^{K}\sum_{l=1}^{n}\mathsf{E}[B_k^2(T_l)\varepsilon_{li}^2|P_{\mathbf{B}}] \\ &\preceq K = o_p(n),\end{aligned}$$

and

$$\mathsf{E}[|\varepsilon_i^T P_{\mathbf{B}}\varepsilon_j|\,|P_{\mathbf{B}}] \le \{\mathsf{E}[|\varepsilon_i^T P_{\mathbf{B}}\varepsilon_i|\,|P_{\mathbf{B}}] + \mathsf{E}[|\varepsilon_i^T P_{\mathbf{B}}\varepsilon_j|\,|P_{\mathbf{B}}]\}/2 = o_p(n).$$

It follows from the law of large numbers that $\varepsilon_i^T\varepsilon_j/n$ converges to v_{ij} in probability. Consequently,

$$\varepsilon_i^T(I_n - P_{\mathbf{B}})\varepsilon_j/n \xrightarrow{\text{P}} v_{ij}, \qquad \text{as } n \to \infty. \tag{A.5}$$

If we can prove

$$n^{-1}\mu_i(T)^T(I_n - P_{\mathbf{B}})\mu_i(T) \xrightarrow{\text{P}} 0, \qquad \text{for } 1 \le i \le p, \tag{A.6}$$

then the conclusion of Lemma A.3 will hold. Since $I_n - P_{\mathbf{B}}$ is an idempotent matrix, it follows from (A.5) and (A.6), the Markov inequality and the Cauchy-Schwarz inequality that

$$n^{-1}\varepsilon_i^T(I_n - P_{\mathbf{B}})\mu_j(T) \xrightarrow{\text{P}} 0, \qquad \text{for } 1 \le i,\ j \le p,$$

except on an event whose probability tends to 0 as n tends to ∞.

It follows from condition (C4) and Lemma A.2, there exists a constant $C > 0$ and spline functions $\mu_i^*(t) \in S_{m,N_n}$ with $m \ge 1$, such that

$$\|\mu_i - \mu_i^*\|_\infty \le Ch, \qquad \text{for } 1 \le i \le p. \tag{A.7}$$

Write $\gamma_i^* = (\gamma_{1i}^*, \cdots, \gamma_{Ki}^*)^T$, $\mu_i^*(t) = \sum_{k=1}^{K} \gamma_{ki}^* B_k(t)$, and $\mu_i^*(T) = (\mu_i^*(T_1), \cdots, \mu_i^*(T_n))^T$. Then $\mu_i^*(T) = \mathbf{B}\gamma_i^*$. Observe that $|\mu_i - \mu_i^*| \le \|\mu_i - \mu_i^*\|_\infty \le Ch$, and

$$(I_n - P_{\mathbf{B}})\mu_i(T) = (I_n - P_{\mathbf{B}})(\mu_i(T) + \mathbf{B}\gamma_i^* - \mathbf{B}\gamma_i^*) = (I_n - P_{\mathbf{B}})(\mu_i(T) - \mu_i^*(T)).$$

Then,

$$\begin{aligned} |[(I_n - P_{\mathbf{B}})\mu_i(T)]_j| &= |[(I_n - P_{\mathbf{B}})(\mu_i(T) - \mu_i^*(T))]_j| \\ &\le |(\mu_i(T) - \mu_i^*(T))_j| + |(P_{\mathbf{B}}(\mu_i(T) - \mu_i^*(T))_j| \\ &\le Ch + \max_{1\le j\le n} |(\mu_i(T) - \mu_i^*(T))_j| \sum_{l=1}^{n} |a_{jl}| \\ &\le Ch + Ch \sum_{l=1}^{n} |a_{jl}|, \end{aligned} \tag{A.8}$$

where $P_{\mathbf{B}} = (a_{jl})_{n\times n}$. Let $\overline{\mathbf{B}}_j = (B_1(T_j), \cdots, B_K(T_j))^T$. By Lemma A.1 and the fact that $\sum_{k=1}^{K} \varphi_k^2(t) \le 1$ for all $t \in D$, we have

$$a_{jj} = \overline{\mathbf{B}}_j^T (\mathbf{B}^T\mathbf{B})^{-1} \overline{\mathbf{B}}_j \asymp \frac{1}{n} \overline{\mathbf{B}}_j^T \overline{\mathbf{B}}_j = \frac{1}{n} \sum_{k=1}^{K} [B_k(T_j)]^2 \le K/n,$$

except on an event whose probability goes to 0 as $n \to \infty$. On the other hand,

$$|a_{jl}| = |\overline{\mathbf{B}}_j^T (\mathbf{B}^T\mathbf{B})^{-1} \overline{\mathbf{B}}_l| \le (a_{jj} + a_{ll})/2 \le K/n,$$

except on an event whose probability goes to 0 as $n \to \infty$. Note that there are at most $O(n/K)$ nonzero elements in each row of $P_{\mathbf{B}}$ based on the local properties of the B-spline basis. Thus, by (A.8), there is a constant $C_1 > 0$ such that

$$|[(I_n - P_{\mathbf{B}})\mu_i(T)]_j| \le C_1 h, \qquad \text{for } 1 \le j \le n,\ 1 \le i \le p, \tag{A.9}$$

except on an event whose probability goes to 0 as $n \to \infty$. We have

$$n^{-1}\mu_i(T)^T (I_n - P_{\mathbf{B}})\mu_i(T) = n^{-1} \sum_{j=1}^{n} [(I_n - P_{\mathbf{B}})\mu_i(T)]_j^2 \le C_1 h^2 \xrightarrow{\text{P}} 0.$$

Hence (A.6) is true. $\square$

Lemma A.4 Let $m \ge q$. Under the Condition (C1)–(C5), one has

$$\mathbf{X}^T (I_n - P_{\mathbf{B}})\mathbf{G} = O_p(n^{1/2}h^q + nh^{q+1}). \tag{A.10}$$

In particular, Let $h = O(n^{-1/(2q+1)})$, then

$$\mathbf{X}^T (I_n - P_{\mathbf{B}})\mathbf{G} = O_p(n^{q/(2q+1)}) = o_p(n^{1/2}).$$

Proof Observe that for $1 \le s \le d$,

$$[\mathbf{X}^T(I_n - P_{\mathbf{B}})\mathbf{g}_s]_i = \mu_i(T)^T(I_n - P_{\mathbf{B}})\mathbf{g}_s + \varepsilon_i^T(I_n - P_{\mathbf{B}})\mathbf{g}_s.$$

It follows from condition (C2) and Lemma A.2, the argument used to derive (A.9) that there exists a constant $C_1 > 0$ such that

$$|[(I_n - P_{\mathbf{B}})\mathbf{g}_s]_j| \le C_1 h^q, \qquad \text{for } \ 1 \le j \le n, \ 1 \le s \le d, \tag{A.11}$$

except for an event whose probability goes to 0 with n. It follows from (A.9) and (A.11), there is a constant $C > 0$ such that

$$|\mu_i(T)^T(I_n - P_{\mathbf{B}})\mathbf{g}_s| = |\mu_i(T)^T(I_n - P_{\mathbf{B}})(I_n - P_{\mathbf{B}})\mathbf{g}_s| \le nCh^{q+1},$$

except for an event whose probability goes to 0 as $n \to \infty$. It follows from (A.5) and (A.11), the Markov inequality and the Cauchy-Schwarz inequality that

$$\varepsilon_i^T(I_n - P_{\mathbf{B}})\mathbf{g}_s = O_p(n^{1/2}h^q), \qquad \text{for } \ 1 \le i \le p. \tag{A.12}$$

The proof of Lemma A.4 is complete. $\square$

Proof of Theorem 3.1 Observe that

$$\begin{aligned}
\mathrm{Vec}(\widehat{\mathbf{H}} - \mathbf{H}) \ &= \ (I_d \otimes [(\mathbf{X}^T(I_n - P_{\mathbf{B}})\mathbf{X})^{-1}\mathbf{X}^T(I_n - P_{\mathbf{B}})])\mathrm{Vec}(\mathbf{G}) \\
&\quad + (I_d \otimes [(\mathbf{X}^T(I_n - P_{\mathbf{B}})\mathbf{X})^{-1}\mu(T)^T(I_n - P_{\mathbf{B}})])\mathrm{Vec}(\mathbf{e}) \\
&\quad + (I_d \otimes [(\mathbf{X}^T(I_n - P_{\mathbf{B}})\mathbf{X})^{-1}\varepsilon^T(I_n - P_{\mathbf{B}})])\mathrm{Vec}(\mathbf{e}).
\end{aligned} \tag{A.13}$$

Thus, Theorem 3.1 will be proved if we can show that as $n \to \infty$,

$$\sqrt{n}(I_d \otimes [(\mathbf{X}^T(I_n - P_{\mathbf{B}})\mathbf{X})^{-1}\varepsilon^T(I_n - P_{\mathbf{B}})])\mathrm{Vec}(\mathbf{e}) \xrightarrow{L} N_{p\times d}(\mathbf{0}, \Sigma \otimes \mathbf{V}^{-1}), \tag{A.14}$$

$$\sqrt{n}(\mathbf{X}^T(I_n - P_{\mathbf{B}})\mathbf{X})^{-1}\mu(T)^T(I_n - P_{\mathbf{B}})\mathbf{e} \xrightarrow{\mathrm{P}} \mathbf{0} \tag{A.15}$$

and

$$\sqrt{n}(\mathbf{X}^T(I_n - P_{\mathbf{B}})\mathbf{X})^{-1}\mathbf{X}^T(I_n - P_{\mathbf{B}})\mathbf{G} \xrightarrow{\mathrm{P}} \mathbf{0}. \tag{A.16}$$

Note that equation (A.16) is equivalent to

$$\sqrt{n}[\mathbf{X}^T(I_n - P_{\mathbf{B}})\mathbf{X}]^{-1}\mathbf{X}^T(I_n - P_{\mathbf{B}})\mathbf{g}_s \xrightarrow{\mathrm{P}} 0, \qquad s = 1, \cdots, d.$$

It follows from the Lemma A.3 and A.4, the argument used to derive (A.9) that

$$\begin{aligned}\mathsf{E}|\mu(T)^T(I_n-P_{\mathbf{B}})\mathbf{e}_s|^2 &= \mathsf{E}\Big\{\sum_{i=1}^p\Big(\sum_{j=1}^n \mathbf{e}_{js}[(I_n-P_{\mathbf{B}})\mu_i(T)]_j\Big)^2\Big\}\\ &= \sum_{i=1}^p\sum_{j=1}^n \mathsf{E}(\mathbf{e}_{js}^2[(I_n-P_{\mathbf{B}})\mu_i(T)]_j^2)\\ &= \sum_{i=1}^p\sum_{j=1}^n \sigma_{ss}^2\mathsf{E}([(I_n-P_{\mathbf{B}})\mu_i(T)]_j^2)\\ &\le nC_1\sigma_{ss}^2h^2.\end{aligned}$$

Therefore, $\mu(T)^T(I_n-P_{\mathbf{B}})\mathbf{e}_s=O_p(n^{1/2}h)=o_p(n^{1/2})$. we have

$$\begin{aligned}&\sqrt{n}(\mathbf{X}^T(I_n-P_{\mathbf{B}})\mathbf{X})^{-1}\mu(T)^T(I_n-P_{\mathbf{B}})\mathbf{e}\\ =\ &(n^{-1}\mathbf{X}^T(I_n-P_{\mathbf{B}})\mathbf{X})^{-1}n^{-1/2}\mu(T)^T(I_n-P_{\mathbf{B}})\mathbf{e}\\ =\ &O_p(1)n^{-1/2}o_p(n^{1/2})=o_p(1).\end{aligned}$$

Hence (A.15) is true.

It follows from the condition (C5), the Lemma A.3 and A.4 that

$$\begin{aligned}&\sqrt{n}[\mathbf{X}^T(I_n-P_{\mathbf{B}})\mathbf{X}]^{-1}\mathbf{X}^T(I_n-P_{\mathbf{B}})\mathbf{g}_s\\ =\ &[n^{-1}\mathbf{X}^T(I_n-P_{\mathbf{B}})\mathbf{X}]^{-1}n^{-1/2}\mathbf{X}^T(I_n-P_{\mathbf{B}})\mathbf{g}_s\\ =\ &O_p(1)n^{-1/2}O_p(n^{1/2}h^q+nh^{q+1})=o_p(1).\end{aligned}$$

Hence (A.16) is true.

Let $P_{\mathbf{B}}\varepsilon=(r_1,\cdots,r_n)^T$, then

$$\begin{aligned}&\mathsf{E}\{\{[I_d\otimes(\varepsilon^TP_{\mathbf{B}})]\mathrm{Vec}(\mathbf{e})\}^T[I_d\otimes(\varepsilon^TP_{\mathbf{B}})]\mathrm{Vec}(\mathbf{e})\}\\ =\ &\mathsf{E}\{[\mathrm{Vec}(\mathbf{e})]^T[I_d\otimes(P_{\mathbf{B}}\varepsilon\varepsilon^TP_{\mathbf{B}})]\mathrm{Vec}(\mathbf{e})]\}\\ =\ &\sum_{s=1}^d\mathsf{E}[\mathbf{e}_s^TP_{\mathbf{B}}\varepsilon\varepsilon^TP_{\mathbf{B}}\mathbf{e}_s]=\sum_{s=1}^d\mathsf{E}\Big|\sum_{j=1}^n e_{js}r_j\Big|^2\\ =\ &\sum_{s=1}^d\sum_{j=1}^n\mathsf{E}|e_{js}r_j|^2=\sum_{s=1}^d\sum_{j=1}^n\sigma_{ss}^2\mathsf{E}|r_j|^2\\ =\ &\sum_{s=1}^d\sigma_{ss}^2\mathsf{E}(\mathrm{tr}(\varepsilon^TP_{\mathbf{B}}\varepsilon))=O_p(K)\sum_{s=1}^d\sigma_{ss}^2.\end{aligned}$$

Thus,

$$\{I_d\otimes(\varepsilon^TP_{\mathbf{B}})\}\mathrm{Vec}(\mathbf{e})=o_p(n^{1/2}).\tag{A.17}$$

Obviously, it follows from the central limit theorem that

$$n^{-1/2}\{I_d\otimes(\varepsilon^T)\}\mathrm{Vec}(\mathbf{e})\xrightarrow{L}N_{p\times d}(\mathbf{0},\Sigma\otimes\mathbf{V}).\tag{A.18}$$

By (A.17), (A.18) and Lemma A.3, we obtain (A.14). This establishes Theorem 3.1 by (A.14)-(A.16). □

Proof of Theorem 3.2 It follows from (2.4) that

$$\widehat{\theta} = (\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T(\mathbf{Y} - \mathbf{X}\widehat{\mathbf{H}}).$$

Note that $\widehat{g}_s(t) = \sum\limits_{k=1}^{K} \widehat{\theta}_{ks} B_k(t)$. Set $\widetilde{\mathbf{Y}} = (\widetilde{Y}_1^T, \cdots, \widetilde{Y}_n^T)^T$ with $\widetilde{Y}_i = X_i^T(\mathbf{H} - \widehat{\mathbf{H}}) + \mathbf{g}(T_i)$. Then $\widetilde{\mathbf{Y}} = \mathbf{X}(\mathbf{H} - \widehat{\mathbf{H}}) + \mathbf{G}$. Define $\widetilde{\theta} = (\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T\widetilde{\mathbf{Y}}$. Write $\widetilde{\theta} = (\widetilde{\theta}_1, \cdots, \widetilde{\theta}_d)$, where $\widetilde{\theta}_s = (\widetilde{\theta}_{1s}, \cdots, \widetilde{\theta}_{Ks})^T$. Let $\widetilde{g}_s(t) = \sum\limits_{k=1}^{K} \widetilde{\theta}_{ks} B_k(t)$. In the following we evaluate the first magnitude of $\|\widehat{g}_s - \widetilde{g}_s\|_2$ and then that of $\|\widetilde{g}_s - g_s\|_2$.

Clearly $\widehat{\theta}_s - \widetilde{\theta}_s = (\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T\mathbf{e}_s$. Since

$$\mathsf{E}\{B_k(T_i)B_k(T_j)e_{is}e_{js}\} = 0, \qquad (i \neq j),$$

we obtain that

$$\begin{aligned}\mathsf{E}\{\mathbf{e}_s^T\mathbf{B}\mathbf{B}^T\mathbf{e}_s\} &= \mathsf{E}\Big\{\sum_{k=1}^{K}\Big[\sum_{j=1}^{n} B_k(T_j)e_{js}\Big]^2\Big\} \\ &= \sum_{k=1}^{K}\sum_{j=1}^{n} \mathsf{E}[B_k(T_j)e_{js}]^2 \preceq nK.\end{aligned}$$

Hence, $\mathbf{e}_s^T\mathbf{B}\mathbf{B}^T\mathbf{e}_s = O_p(nK)$. Using Lemma A.1, we have that

$$|\widehat{\theta}_s - \widetilde{\theta}_s|^2 = \mathbf{e}_s^T\mathbf{B}(\mathbf{B}^T\mathbf{B})^{-1}(\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T\mathbf{e}_s \preceq \frac{1}{n^2}\mathbf{e}_s^T\mathbf{B}\mathbf{B}^T\mathbf{e}_s = O_p\Big(\frac{K}{n}\Big),$$

which together with (A.1) yields

$$\|\widehat{g}_s - \widetilde{g}_s\|_2^2 \asymp |\widehat{\theta}_s - \widetilde{\theta}_s|^2 = O_p\Big(\frac{K}{n}\Big).$$

Let $g_s^* \in S_{m,N_n}$ be such that (A.11) holds. Then

$$\|g_s^* - g_s\|_2 \preceq \|g_s^* - g_s\|_\infty \preceq h^q.$$

Write $g_s^*(t) = \sum\limits_{k=1}^{K} \theta_{ks}^* B_k(t)$, $\theta_s^* = (\theta_{1s}^*, \cdots, \theta_{Ks}^*)^T$ and $\mathbf{g}_s^* = (g_s^*(T_1), \cdots, g_s^*(T_n))^T$. Obviously, $\mathbf{g}_s^* = \mathbf{B}\theta_s^*$. By (A.1) and Lemma A.1, we have

$$\|\widetilde{g}_s - g_s^*\|_2^2 \asymp |\widetilde{\theta}_s - \theta_s^*|^2 \asymp \frac{1}{n}(\widetilde{\theta}_s - \theta_s^*)^T\mathbf{B}^T\mathbf{B}(\widetilde{\theta}_s - \theta_s^*)$$

with probability tending to 1. Since $\mathbf{B}\widetilde{\theta}_s = \mathbf{B}(\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T\widetilde{\mathbf{y}}_s$ is an orthogonal projection, where $\widetilde{\mathbf{y}}_s$ is the s'th column of $\widetilde{\mathbf{Y}}$, then

$$\begin{aligned}\frac{1}{n}(\widetilde{\theta}_s-\theta_s^*)^T\mathbf{B}^T\mathbf{B}(\widetilde{\theta}_s-\theta_s^*) &\le \frac{1}{n}|\widetilde{\mathbf{y}}_s-\mathbf{B}\theta_s^*|^2=\frac{1}{n}|\mathbf{X}(\beta_s-\widehat{\beta}_s)+\mathbf{g}_s-\mathbf{g}_s^*|^2\\ &\le \frac{2}{n}\big[|\mathbf{X}(\beta_s-\widehat{\beta}_s)|^2+|\mathbf{g}_s-\mathbf{g}_s^*|^2\big]\\ &= 2(\beta_s-\widehat{\beta}_s)^T\Big(\frac{1}{n}\mathbf{X}^T\mathbf{X}\Big)(\beta_s-\widehat{\beta}_s)+\frac{2}{n}\sum_{j=1}^n[g_s(T_j)-g_s^*(T_j)]^2\\ &\le 2\lambda_1|\beta_s-\widehat{\beta}_s|^2+\frac{2}{n}\sum_{j=1}^n[g_s(T_j)-g_s^*(T_j)]^2,\end{aligned}$$

where λ_1, the largest eigenvalue of X^TX/n, is bounded. By condition (C2), $\mathsf{E}[g_s(T)-g_s^*(T)]^2 \asymp \|g_s-g_s^*\|_2^2$ and thus

$$\frac{1}{n}\sum_{j=1}^n[g_s(T_j)-g^*(T_j)]^2=O_p(\|g_s-g_s^*\|_2^2).$$

On the other hand, by condition (C3), and Theorem 3.1, we have

$$(\beta_s-\widehat{\beta}_s)^T\Big(\frac{1}{n}\mathbf{X}^T\mathbf{X}\Big)(\beta_s-\widehat{\beta}_s)=O_p(n^{-1}).$$

Consequently, $\|\widetilde{g}_s-g_s^*\|_2^2=O_p(n^{-1}+h^{2q})$. Furthermore,

$$\|\widehat{g}_s-g_s\|_2^2 \preceq \|\widehat{g}_s-\widetilde{g}_s\|_2^2+\|\widetilde{g}_s-g_s^*\|_2^2+\|g_s^*-g_s\|_2^2.$$

Therefore, it follows from the above arguments that $\|\widehat{g}_s-g_s\|_2^2=O_p((nh)^{-1}+h^{2q})$. The proof of Theorem 3.2 is complete. □

References

[1] Burman, P., Regression function estimation for dependent observations, *J. Multivariate Anal.*, **36**(1999), 263–279.

[2] Beatriz Pateiro-López, M. and González-Mantziga, W., Multivariate partially linear models, *Statistics & Probability Letters*, **76**(2006), 1543–1549.

[3] Chen, H., Covergence rates for parametric components in a partly linear model, *Ann. Statist.*, **16**(1988), 136–146.

[4] De Boor, C., *A Practical Guide to Splines*, Springer, New York, 2001.

[5] Christensen, R., *Plane Answers to Complex Questions: the Theory of Linear Models*, second edition, Springer-Verlag, New York, 1996.

[6] DeVore, R.A. and Lorentz, G.G., *Constructive Approximation*, Springer-Verlag, Berlin Heidelberg, 1993.

[7] Engle, R.F., Granger, C.W.J., Rice, J. and Weiss, A., Semiparametric estimates of the relation between weather and electricity sales, *J.A.S.A.*, **81**(1986), 310–320.

[8] Györfi, L., Kohler, M. andżak, A., *A Distribution-Free Theory of Nonparametric Regression*, Springer, New York, 2002.

[9] Hädle, W., Liang, H. and Gao, J., *Partially Linear Model*, Physica-Verlag, New York, 2000.

[10] Hamilton, S.A. and Truong, Y.K., Local linear estimation in partly linear model, *J. Multivariate Anal.*, **60**(1997), 1–19.

[11] Heckman, N.E., Spline smoothing in a partly linear model, *J. Roy. Statist. Soc. Ser. B.*, **48**(1986), 244–248.

[12] Huang, J.Z., Local asymptotics for polynomial spline regression, *Ann. Statist.*, **31**(2003), 1600–1635.

[13] Speckman, P., Kernel smoothing in partial linear model, *J. Roy. Statist. Soc. Ser. B.*, **50**(1988), 413–436.

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

[15] You, J. and Chen, G., B-spline estimation for partially linear regression model with heteroscedasticity, *Chinese Annals of Mathematics, Series A*, **25A**(2004), 661–676.

多元部分线性模型的B-样条估计

陈广雷 王兆军

(中国人民武装警察部队学院基础部, 廊坊, 065000) (南开大学数学科学学院, 天津, 300071)

本文考虑多元部分线性回归模型的估计问题, 得到了该模型参数的最小二乘估计和非参数函数的B-样条估计, 并证明了参数估计的渐近正态性, 给出了非参数函数估计的最优收敛速度.

关 键 词: 多元回归, 部分线性模型, B-样条, 渐近正态.

学 科 分 类 号: O212.7.