应用概率统计 第三十卷
第四期 2014年8月

Chinese Journal of Applied Probability
and Statistics Vol.30 No.4 Aug. 2014

Bernstein Polynomial Estimation in the Partially Linear Model under Monotonicity Constraints *

Ding Jianhua[1,2] Zhang Zhongzhan[1]

([1] *College of Applied Sciences, Beijing University of Technology, Beijing, 100124*)

([2] *Department of Mathematics, Shanxi Datong University, Datong, 037009*)

Abstract

In this paper, a Bernstein-polynomial-based likelihood method is proposed for the partially linear model under monotonicity constraints. Monotone Bernstein polynomials are employed to approximate the monotone nonparametric function in the model. The estimator of the regression parameter is shown to be asymptotically normal and efficient, and the rate of convergence of the estimator of the nonparametric component is established, which could be the optimal under the smooth assumptions. A simulation study and a real data analysis are conducted to evaluate the finite sample performance of the proposed method.

Keywords: Empirical process, maximum likelihood method, monotone Bernstein polynomial, Monte Carlo.

AMS Subject Classification: 62G08.

§1. Introduction

In this paper, we consider Bernstein-polynomial-based maximum likelihood estimation for a partially linear model under monotonicity constraints. A general partially linear model takes the form

$$Y = \psi(Z) + X^T\beta + \varepsilon, \tag{1.1}$$

where $X = (x_1, \ldots, x_d)^T$ and Z are $d \times 1$ and 1 dimensional explanatory variables respectively, β is a $d \times 1$ vector of the unknown regression parameters, ψ is an unknown function, the error term ε is normally distributed with mean 0 and finite variance σ^2, and (X, Z) and ε are independent. The partially linear model is an extension of a standard linear model without having to specify the functional forms of some predictor variables. It can be an appropriate choice when the response variable Y is assumed to be linearly associated with covariate X, but the relationship between Y and Z may be nonlinear.

*The project was supported by the National Natural Science Foundation of China (11271039).
Received February 20, 2014. Revised April 8, 2014.
doi: 10.3969/j.issn.1001-4268.2014.04.005

The partially linear model (1.1) has been extensively studied by many authors, see for example, Bianco and Boente (2004), Engle et al. (1986), Green et al. (1985), Green and Silverman (1994), Robinson (1988), and Schimek (2009) among many others. A lot of methods for estimation of the parameters and the smooth nonparametric component have been suggested. Heckman (1986) explored the asymptotic properties of the estimator of β using the penalized likelihood estimation method. Chen (1988) used piecewise polynomials to approximate $\psi(\cdot)$ and showed that the estimator of β can achieve a convergence rate of $n^{-1/2}$ with smallest possible asymptotic variance. Chen and Shiau (1994) studied the asymptotic behaviors of two data-driven efficient estimators of β using the spline estimation method. Mammen and van der Geer (1997) applied empirical process theory to study the asymptotic properties of the penalized quasi-likelihood estimator of β. Speckman (1988) investigated the theoretical properties of the kernel smoothing approach for the partially linear model. Hamilton and Truong (1997) used the local linear smoother method to derive the asymptotic distributions of the estimates of β and ψ, which generalized the results of Robertson et al. (1988).

In many studies, there is a monotonic relationship between one or more of covariates and the response variable, such as the dose-response relationship in some clinical trials. In this case, one would like to give an estimator of the regression function which satisfies the monotone constraint. Earlier works for nonparametric isotonic regression models refer to, for example, Brunk (1970), Robertson and Wright (1975), and Wright (1981). Huang (2002) considered model (1.1) assuming that ψ is a smooth monotone function, and proposed a restricted least squares estimation. Under the assumption that the error ε is normally distributed, the estimator of β was shown to be asymptotically efficient among all regular estimators. The limiting distribution of the isotonic estimator of the monotone nonparametric function ψ at a fixed point was also established. Sun et al. (2012) proposed an another estimation method for model (1.1) with error-in-variable data. Although the above estimators of β performs well, the convergence rate of the nonparametric component may be improved through smoothing technics. Bernstein polynomial approximation is such a technic. It has been found for other models that Bernstein polynomials can be used to construct an isotonic estimator of a monotone function, see, for example, Chak et al. (2005), Chang et al. (2005), Chang et al. (2007), Curtis and Ghosh (2011), Petrone (1999) and Stadtmuller (1986) among others. To the best of our knowledge, however, there is no systematic study for model (1.1) based on the Bernstein polynomial approximation to ψ when it is subject to be monotone. Therefore, it would be preferable to develop a practical Bernstein polynomial procedure for the partially linear model under monotonicity constraints on ψ and study the asymptotic properties of the estimates.

Without loss of generality, assume that $\psi(\cdot)$ is nondecreasing. We employ the following monotone Bernstein polynomials to approximate $\psi(Z)$, i.e.

$$\psi(Z) \approx \Sigma_{j=0}^{N} \alpha_j b_j(Z, N), \tag{1.2}$$

subject to the constraints $\alpha_0 \le \alpha_1 \le \cdots \le \alpha_N$. The nondecreasing constraints on the coefficients α_j, $j = 0, 1, \ldots, N$, guarantee the isotonicity of the resulting Bernstein polynomial estimator (Wang and Ghosh, 2012). This approach follows the idea of the sieve method for the estimation of the infinite-dimensional parameter ψ. In sieve estimation a sequence of subspaces (sieves) that depend on the sample size n are used to approximate the original space such that the resulting estimation problem over sieves becomes less complicated. In the model presented here, the sieves are the collections of monotone Bernstein polynomials and the original space is the set of bounded nondecreasing smooth functions. By using monotone Bernstein polynomials to approximate ψ, we can estimate the Bernstein coefficients $\alpha = (\alpha_0, \ldots, \alpha_N)$ and the regression parameter β simultaneously. We show that the estimator of β is asymptotically normal and efficient and the estimator of ψ achieves the possible optimal rate of convergence under the smooth condition.

The rest of the paper is organized as follows: The Bernstein maximum likelihood estimator $(\widehat{\beta}_n, \widehat{\psi}_n)$ and the numerical algorithm are presented in Section 2. Asymptotic results are given in Section 3. A Monte Carlo simulation study and a real data analysis are displayed in Section 4. Finally, the proofs of asymptotic results are sketched in Appendix.

§2. Method and Algorithm

Let (β_0, ψ_0) be the true value of (β, ψ). Assume the parametric space Θ for β is a convex and compact subset of $\mathbb{R}^d$ and ψ_0 is a smooth monotone function. Since the change of variables specified by $t = (z - a)/(b - a)$ maps $z \in [a, b]$ to $t \in [0, 1]$ without changing the max norm of any function, we can restrict our attention to continuous functions $\psi_0(z)$ on $z \in [0, 1]$ without loss of generality. Let $(Y_1, Z_1, X_1), \ldots, (Y_n, Z_n, X_n)$ be a random sample of (Y, Z, X). The log-likelihood for this random sample is

$$l_n(\beta, \psi) = -\ln(\sqrt{2\pi}\sigma) - \sum_{i=1}^{n} (Y_i - X_i^T \beta - \psi(Z_i))^2 / (2\sigma^2), \tag{2.1}$$

subject to $\beta \in \Theta$ and ψ being monotone. Denoting by $Z_{(1)} \le \cdots \le Z_{(n)}$ the ordered values of Z_i's and letting $\psi_i = \psi(Z_{(i)})$, Huang (2002) defined the semiparametric maximum likelihood estimator of (β_0, ψ_0) as the maximizer of $l_n(\beta, \psi)$ subject to $\beta \in \Theta$ and $\psi_1 \le$

$\cdots \le \psi_n$. The obtained estimator of β is highly efficient, and the estimator of ψ is a nondecreasing step function with jumps only occurring at the observed points Z_i. In this section, we propose to estimate ψ using monotone Bernstein polynomial instead of the step function in order to achieve faster rate of convergence and better finite sample performance of the estimate of ψ.

In this paper, we construct the base function $b_j(Z, N)$ using Bernstein polynomials. For a continuous function such as $\psi(Z)$ on $[0, 1]$, the approximating Bernstein polynomial of order N is given by

$$B(Z; N, \psi) = \sum_{j=0}^{N} \psi\Big(\frac{j}{N}\Big) C_N^j (Z)^j (1-Z)^{N-j}.$$

By Weierstrass theorem, $B(\cdot; N, \psi) \to \psi(\cdot)$ uniformly over $[0, 1]$ as $N \to \infty$ (Lorentz, 1953). Denote $b_j(Z, N) = C_N^j (Z)^j (1-Z)^{N-j}$, $j = 0, \ldots, N$, where $C_N^j = N!/(j!(N-j)!)$. According to Wang and Ghosh (2012), $B(Z; N, \psi)$ is monotone nondecreasing on $[0, 1]$ if nondecreasing constraints are imposed on the coefficients $\boldsymbol{\alpha} = (\alpha_0, \ldots, \alpha_N)$. Thus,

$$\mathscr{M}_N = \Big\{ \sum_{j=0}^{N} \alpha_j \cdot b_j(Z, N) : \alpha_0 \le \cdots \le \alpha_N, \sum_{j=0}^{N} |\alpha_j| \le L_N \Big\}$$

is the collection of monotone nondecreasing polynomials on $[0, 1]$.

Replacing $\psi(Z)$ by $\sum\limits_{j=0}^{N} \alpha_j \cdot b_j(Z, N)$ in the log-likelihood function (2.1), we obtain the Bernstein-polynomial-based log-likelihood function,

$$l_n(\boldsymbol{\alpha}, \beta) = -\ln(\sqrt{2\pi}\sigma) - \sum_{i=1}^{n} \Big(Y_i - X_i^T\beta - \sum_{j=0}^{N} \alpha_j b_j(Z_i, N)\Big)^2 / (2\sigma^2), \tag{2.2}$$

subject to $\alpha_0 \le \cdots \le \alpha_N$. The advantage of this reparametrization is that we can estimate the regression parameters β and coefficients $\boldsymbol{\alpha} = (\alpha_0, \ldots, \alpha_N)$ simultaneously through maximizing the Bernstein likelihood function subject to nondecreasing constraints. The computational burden can be greatly alleviated by such fully parametric representation of Bernstein polynomial likelihood function.

Let $\widehat{\boldsymbol{\alpha}}_n = (\widehat{\alpha}_0, \ldots, \widehat{\alpha}_N)$ and $\widehat{\beta}_n$ be the values that maximize Bernstein likelihood function (2.2), we denote the Bernstein estimation of $\psi(z)$ by $\widehat{\psi}_n(z) = \sum\limits_{i=0}^{N} \widehat{\alpha}_i b_i(z, N)$.

The Bernstein likelihood estimation problem (2.2) can be formulated as an optimization problem subject to linear inequality constraints

$$\max_{\theta \in \Theta_\alpha \times \mathbb{R}^d} l_n(\theta | \text{data}),$$

where $\theta = (\boldsymbol{\alpha}, \beta)$ with $\boldsymbol{\alpha} \in \Theta_{\boldsymbol{\alpha}} = \{\boldsymbol{\alpha} : \alpha_0 \le \cdots \le \alpha_N\}$.

The above optimization problem can be effectively solved by the general quadratic programming (Rosen, 1960; Goldfarb and Idnani, 1983). Quadratic programming has also been used to solve the linear inequality constraints based on B-spline basis under linear constraints methods. In this study, we use the available R package *quadprog* by Turlach and Weingessel (2010) to solve quadratic programming problem. It is to be noted that in practice the L_N can be chosen to be a reasonably large number and hence we do not need to choose its value. The order N of the Bernstein polynomial can be chosen by the cross-validation method which is introduced in numerical examples in Section 4.1.

§3. Asymptotic Results

In this section we present asymptotic results for the Bernstein polynomial maximal likelihood estimator $\widehat{\beta}_n, \widehat{\psi}_n$. Denote $\vartheta = (\beta, \psi)$. Assume the regression parameter space Θ to be a convex and compact subset of $\mathbb{R}^d$ and the parameter space for the nonparametric function ψ is taken to be

$$\mathscr{F} = \{\psi : \psi \text{ is monotone nondecreasing on } [0,1]\}.$$

Let $\|\cdot\|$ be the Euclidean norm of $\mathbb{R}^d$. For any probability measure P, define L_2-norm $\|f\|_2 = \Big(\displaystyle\int f^2 \mathrm{d}P\Big)^{1/2}$. We study the asymptotic properties of $(\widehat{\beta}_n, \widehat{\psi}_n)$ with L_2 metric

$$\begin{aligned} d_2^2(\vartheta_1, \vartheta_2) &= \|\beta_2 - \beta_1\|^2 + \|\psi_2 - \psi_1\|_2^2 \\ &= \|\beta_2 - \beta_1\|^2 + \int |\psi_2(z) - \psi_1(z)|^2 \mathrm{d}F_Z(z), \end{aligned}$$

for any $\vartheta_i = (\beta_i, \psi_i) \in \Theta \times \mathscr{F}$, $i = 1, 2$, where $F_Z(z)$ is the marginal probability measure of the variable Z.

The following regularity conditions with respect to the order of Bernstein polynomial, the smoothness and monotonicity of ψ_0, and the underlying distributions of covariates (X, Z) are needed to derive the asymptotic results of the Bernstein polynomial maximum likelihood estimator $(\widehat{\beta}_n, \widehat{\psi}_n)$.

C1. The true function ψ_0 is strictly increasing and its first derivative is Holder continuous with the exponent a_0, i.e. there exist constant $a_0 \in [0,1]$ and constant M such that $|\psi_0^{(1)}(Z_1) - \psi_0^{(1)}(Z_2)| \le M|Z_1 - Z_2|^{a_0}$ for all $Z_1, Z_2 \in [0,1]$.

C2. The order N of Bernstein polynomials satisfies $N = O(n^\kappa)$ with $\kappa = 1/(3 + 2a_0)$.

C3. The true parameter β_0 is in the interior of Θ.

C4. There exists x_0 such that $\mathsf{P}(\|X\| \le x_0) = 1$. That is, the covariate X has a bounded support.

C5. The density function of Z is continuous and positive on $[0,1]$.

C6. For any $\beta \neq \beta_0$, $\mathsf{P}(X^T\beta \neq X^T\beta_0) > 0$.

C7. $\mathsf{E}(X - \mathsf{E}(X|Z))^{\otimes 2}$ is positive definite, where $x^{\otimes 2} = xx^T$.

C8. The function $h^*(z) = \mathsf{E}(X|Z=z)$ satisfies the Lipschitz condition on $[0,1]$.

Remark 1 (C1) and (C2) are two mild assumptions needed to derive consistency and the rate of convergence of $(\widehat{\beta}_n, \widehat{\psi}_n)$. The compactness and convexity of Θ and (C3) are common in the literature of semiparametric estimation. Assumptions that are related to observation scheme of (X,Z), (C4)-(C5), are needed for the entropy calculation in the proofs of Theorems 3.1-3.2. (C6) is required to establish the identifiability of the semiparametric model. (C7) and (C8) are useful in the proof of the asymptotic normality.

To obtain the asymptotic distribution of $\widehat{\beta}_n$, we first need to adjust for the dependence of Z and X, which is a common complication in semiparametric models. For a single observation (Y,Z,X), its log density given by

$$l(\beta,\psi) = -\ln(\sqrt{2\pi}\sigma) - (Y - \psi(Z) - X^T\beta)^2/(2\sigma^2).$$

The score function for β is

$$\dot{l}_\beta(\beta,\psi) = (Y - \psi(Z) - X^T\beta)X/\sigma^2.$$

Consider a parametric smooth submodel (β,ψ_t), where $\psi_0 = \psi$ and $\partial\psi_t/\partial t|_{t=0} = h$, more details of the submodel see Chapter 25 of van der Vaart (1996). Let $\mathscr{H}$ be the class of such h functions with bounded variation on $[0,1]$. The score function for ψ takes the form of

$$\dot{l}_\psi(\beta,\psi)h = (Y - \psi(Z) - X^T\beta)h/\sigma^2.$$

The efficient score for β at the true parameter (β_0,ψ_0) is given by

$$l_\beta^* = \dot{l}_\beta(\beta_0,\psi_0) - \dot{l}_\psi(\beta_0,\psi_0)h^*,$$

where $h^* \in \mathscr{H}^d$ satisfies

$$\mathsf{E}[\dot{l}_\beta(\beta_0,\psi_0) - \dot{l}_\psi(\beta_0,\psi_0)h^*]\dot{l}_\psi^T(\beta_0,\psi_0)h = 0,$$

for all $h \in \mathscr{H}^d$. This simplifies to

$$\mathsf{E}(Y - \psi_0(Z) - X^T\beta_0)^2(X - h^*(Z))h^T(Z) = 0,$$

for all $h \in \mathscr{H}^d$. Thus, $h^*(z) = \mathsf{E}(X|Z=z)$. So the efficient score function for β at (β_0, ψ_0) is

$$l_\beta^* = (Y - \psi_0(Z) - X^T\beta_0)(X - \mathsf{E}(X|Z))/\sigma^2.$$

The efficient information takes the form of

$$I(\beta_0) = \mathsf{E} l_\beta^{*\otimes 2} = \mathsf{E}(X - \mathsf{E}(X|Z))^{\otimes 2}/\sigma^2,$$

where $x^{\otimes 2} = xx^T$, for $x \in \mathbb{R}^d$.

Theorem 3.1 Suppose conditions (C1)-(C7) hold. Then

(i) (Consistency)

$$d_2((\widehat{\beta}_n, \widehat{\psi}_n), (\beta_0, \psi_0)) \to 0$$

in probability, as $n \to \infty$.

(ii) (Rate of convergence)

$$d_2((\widehat{\beta}_n, \widehat{\psi}_n), (\beta_0, \psi_0)) = O_P(n^{-(1+a_0)/(3+2a_0)}).$$

Thus, if $a_0 = 1$, then the rate of convergence is $O_P(n^{-2/5})$, and if $a_0 = 0$, then the rate of convergence is $O_P(n^{-1/3})$, which is the optimal rate of convergence under the smooth condition.

Theorem 3.2 (Asymptotic normality) Suppose conditions (C1)-(C8) hold. Then

$$\sqrt{n}(\widehat{\beta}_n - \beta_0) = n^{-1/2} I^{-1}(\beta_0) \sum_{i=1}^{n} l_\beta^* + o_P(1) \longrightarrow N(0, I^{-1}(\beta_0))$$

in distribution, as $n \to \infty$.

§4. Numerical Examples

4.1 Simulation Study

In this section a Monte Carlo simulation study is performed to evaluate the finite sample performance of the proposed estimation method. We generate n independently and identically distributed observations $\{(Y_i, X_i, Z_i) : i = 1, \ldots, n\}$ as follows: $Z_i \sim$ Uniform$[0,1]$; $X_i \sim N(0,1)$ and model is given by

$$Y_i = \psi(Z_i) + X_i\beta + \varepsilon_i,$$

where $\beta = 0.2$, $\psi(Z) = \sin(\pi Z/2)$ and $\varepsilon_i \sim N(0,1)$. In order to choose a proper order of Bernstein polynomial N, we use the popular V-fold cross-validation method. Given V, for each N, the cross validation term $CV(N)$ for our estimator takes the following form:

$$CV(N)=\frac{1}{V}\sum_{v=1}^{V}\sum_{i\in I_{-v}}(Y_i-X_i\widetilde{\beta}-\widetilde{\psi}(Z_i))^2, \tag{4.1}$$

where $\widetilde{\beta}$ and $\widetilde{\psi}(\cdot)$ are obtained from the v-th training data set consisting of $\lfloor n(V-1)/V\rfloor$ observation points and I_{-v} denotes the corresponding validation set consisting of $[n/V]$ points. We compute the cross validation function defined in (4.1) for a series of N values starting with $N=2$ to a relatively large integer $N_{\max}(<\lfloor n(V-1)/V-1\rfloor)$. The optimal value $\widehat{N}$ is chosen to minimize (4.1), i.e., $\widehat{N}=\mathop{\arg\min}\limits_{N\in[2,N_{\max}]} CV(N)$. The samples of $n=50,100$ and 200 observations are generated respectively, and the data generation and subsequent estimation are repeated 500 times. The order of Bernstein polynomial is estimated using 7-fold cross validation method. We also compare the constrained estimation with the unconstrained estimation.

The Monte Carlo sample bias, standard deviation (SD), and mean squared error (MSE) for the constrained estimation and the the unconstrained estimation of β are summarized in Table 1 Table 2, based on 500 repeated samples, $n=50,100$ and 200, respectively. For the current simulation setting, we can directly compute the efficient information $I(\beta_0)=1$ and the asymptotic variance $\mathrm{AVar}(\widehat{\beta}_n)=1/n$.

Table 1 The constrained estimation

n	Bias($\widehat{\beta}_n$)	SD($\widehat{\beta}_n$)	MSE($\widehat{\beta}_n$)	AVar($\widehat{\beta}_n$)
50	-0.0099	0.1562	0.0245	0.0200
100	0.0052	0.1082	0.0117	0.0100
200	-0.0080	0.0672	0.0050	0.0050

Table 2 The unconstrained estimation

n	Bias($\widehat{\beta}_n$)	SD($\widehat{\beta}_n$)	MSE($\widehat{\beta}_n$)	AVar($\widehat{\beta}_n$)
50	-0.0098	0.1645	0.0272	0.0200
100	-0.0013	0.1075	0.0116	0.0100
200	0.0009	0.0714	0.0051	0.0050

The simulation results show that the sample biases are small and the standard deviations decrease when the sample size n increases for the proposed estimation method. Moreover, the variances of $\widehat{\beta}_n$ derived from the asymptotic theory are close to the corresponding mean squared errors based on the Monte Carlo simulations, which provides a numerical justification for the asymptotic normality result in Theorem 3.2.

For nonparametric function $\psi(Z)$, we investigate both the unconstrained and constrained estimation for $\psi(Z)$. From Figure 1, we can see that one of the key distinguishing features of the proposed estimator is that the monotonicity constraint is maintained for any finite sample size and satisfied over the entire domain of the nonparametric function. Hence, the appropriate background information about the monotone constraint of nonparametric function can provide better estimator than those without such subject-matter information.

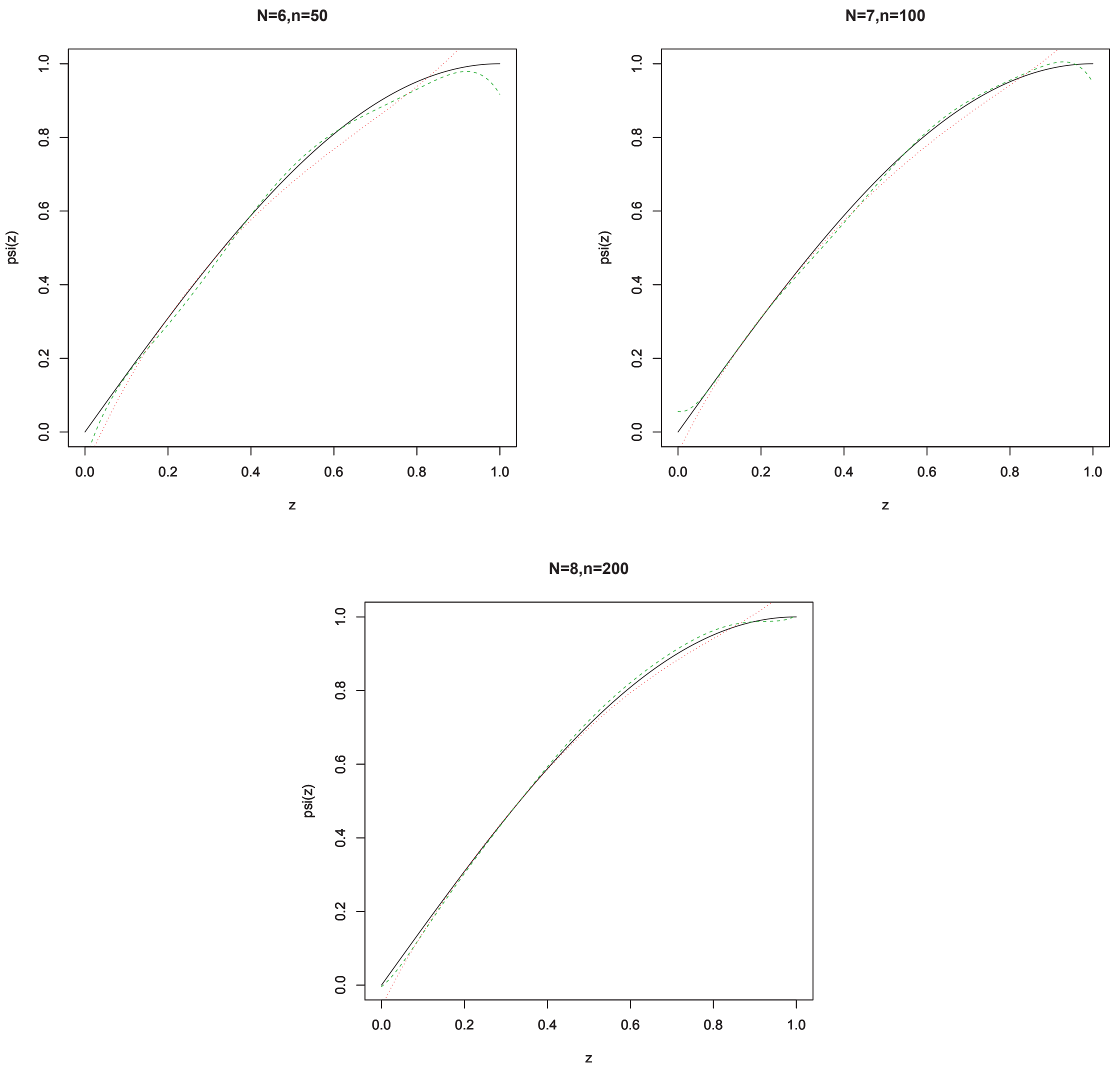

Figure 1　The real line denotes the true function curve, the dotted line denotes the constrained estimated curve, the dashed line denotes the unconstrained estimated curve, based on 500 repeated samples, sizes of sample $n = 50$, $100, 200$, and corresponding order 6,7,8, respectively.

4.2 A Real Data Example

We now illustrate the proposed procedures via an analysis of a subset of data from the Multi Center AIDS Cohort study. The dataset contains the human immunodeficiency virus (HIV) status of 283 homosexual men who were infected with HIV during the follow-up period between 1984 and 1991. Details of the study design, methods, and medical implications have been given by Kaslow et al. (1987). Fan and Zhang (2000) and Huang et al. (2002) analyzed the same dataset using varying coefficient models. Their analysis aimed to describe the trend of the mean CD4 percentage depletion over time and to evaluate the effects of cigarette smoking, pre-HIV infection CD4 percentage, and age at infection on the mean CD4 percentage after the infection. Therefore, they took the CD4 cell percentage of a subject at distinct time points after HIV infection and considered the three covariates: Smoking, Age, and PreCD4. Huang et al. (2002) fitted the data by a varying-coefficient model, and found that at significance level 0.05, only the covariate PreCD4 is linearly related to the viral load. Furthermore, neither smoking nor age had a significant impact on PerCD4. In our analysis, we believe that it is reasonable to assume that the response PerCD4 is decreasingly functionally related to the treatment time, since HIV had a devastating effect on CD4 cells after the subject was infected. All of the above motivated us to model the relation between the response CD4 and the two covariates: PreCD4 and the treatment time by model (1.1)

$$Y_i = X_i\beta + \psi(Z_i) + \varepsilon_i,$$

where Y_i represent the individual's response CD4 percentage for subject i at time Z_i, X_i the observed variable for PreCD4, and Z_i the treatment time. For a clear interpretation, we standardized PreCD4.

We consider both the unconstrained estimation and monotone constrained estimation. The order of the Bernstein polynomial is chosen $N = 8$. The constrained estimator and unconstrained estimator of β are 3.0463 and 3.0635, respectively. This shows that a somewhat stronger positive association between the CD4 percentage and PreCD4 is detected. As is seen from Figure 2, the unconstraint estimation of nonparametric function $\psi(Z)$ does not preserve decreasing property especially in the boundary points. For constrained estimation, the monotonicity property is satisfied over the entire domain of the nonparametric function. Figure 2 shows that the mean baseline CD4 percentage for the population decreased rather quickly after infection. During the three years after infection, the mean percentage of CD4 decreased from 36% to 21% approximately, which indicated a decline of 40% compared with the level before infection. The above results are similar with the results of Fan and Zhang (2000), although we ignored the correlation structure of the data in our analysis.

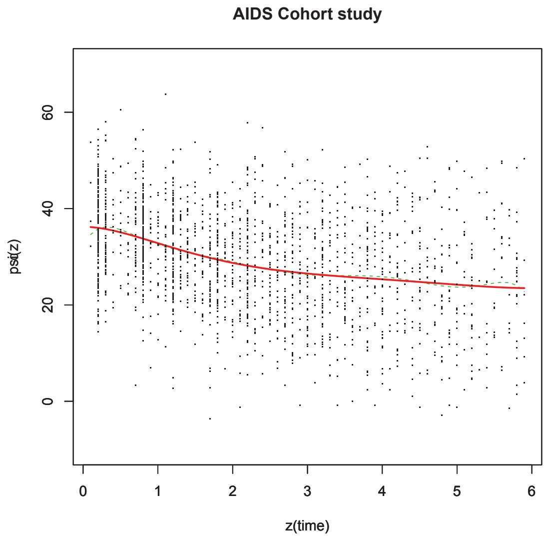

Figure 2　The real red line is the constrained estimate of $\psi(Z)$, the dashed green line is the unconstrained estimate. The dots are the residual on parametric part $r_i = y_i - \widehat{\beta} X_i$, where $\widehat{\beta}$ is the constrained estimate.

§5.　Proof of Theorems

For simplicity we assume that $X \in \mathbb{R}$. The general case can be proved similarly. Given a random sample $X_1, \ldots, X_n$ with probability measure P on a measurable space $(\mathscr{X}, \mathscr{A})$, for a measurable function $f : \mathscr{X} \mapsto \mathbb{R}$, define $Pf = \int f \mathrm{d}P$ as the expectation of f under P and $\mathbb{P}_n f = n^{-1} \sum_{i=1}^{n} f(X_i)$ as the expectation of f under the empirical measure P_n. We write $\mathbb{G}_n f = \sqrt{n}(\mathbb{P}_n - P_0) f$ for the empirical process evaluated at f and $\|\mathbb{G}_n\|_{\mathscr{F}} = \sup_{f \in \mathscr{F}} |\mathbb{G}_n f|$ for any measurable class of functions $\mathscr{F}$.

Proof of Theorem 3.1　We first prove part (i). Let $\mathbb{M}(\vartheta) = Pl(\vartheta)$ and $\mathbb{M}_n(\vartheta) = \mathbb{P}_n l(\vartheta)$. Recall that $\mathscr{F}$ is the class of monotone nondecreasing function on $[0, 1]$. Define $\mathscr{L}_1 = \{l(\beta, \psi) : (\beta, \psi) \in \Theta \times \mathscr{F}\}$. According to Example 19.11 of van der Vaart (1998), for any $\varepsilon > 0$, $\log N_{[\,]}(\varepsilon, \mathscr{F}, L_2(P)) \le K(1/\varepsilon)$. Hence, $\mathscr{F}$ is a P-Donsker class. Furthermore, X has a bounded support and Θ is compact. Therefore, we can show that $\mathscr{L}_1$ is P-Donsker class. It yields $\sup_{(\beta,\psi) \in \Theta \times \mathscr{F}} |\mathbb{M}_n(\beta, \psi) - \mathbb{M}(\beta, \psi)| = o_P(1)$. Thus, we have uniform convergence of $\mathbb{M}_n$ to $\mathbb{M}$ on $\Theta \times \mathscr{F}$.

A straightforward algebra yields $\mathbb{M}(\vartheta_0) - \mathbb{M}(\vartheta) = P((g + h)^2/(2\sigma^2))$, where $g = X\beta - X\beta_0$ and $h = \psi - \psi_0$. Note that $(Pgh)^2 = \sigma^4 (\beta - \beta_0)^2 [P\dot{l}_\psi(\beta_0, \psi_0) h \dot{l}_\beta(\beta_0, \psi_0)]^2$.

Since $Pl_\psi(\beta_0,\psi_0)hl^*_\beta(\beta_0,\psi_0)=0$, for any h, we have

$$[P\dot{l}_\psi(\beta_0,\psi_0)h\dot{l}_\beta(\beta_0,\psi_0)]^2=[P\dot{l}_\psi(\beta_0,\psi_0)h(\dot{l}_\beta(\beta_0,\psi_0)-l^*_\beta(\beta_0,\psi_0))]^2.$$

By Cauchy-Schwarz inequality and the fact that

$$P(\dot{l}_\beta(\beta_0,\psi_0)-l^*_\beta(\beta_0,\psi_0))^2=CP(\dot{l}_\beta(\beta_0,\psi_0))^2,$$

for $0<C<1$, we obtain

$$[P\dot{l}_\psi(\beta_0,\psi_0)h\dot{l}_\beta(\beta_0,\psi_0)]^2\le CP(\dot{l}_\beta(\beta_0,\psi_0))^2P(\dot{l}_\psi(\beta_0,\psi_0)h)^2.$$

Therefore, $(Pgh)^2\le CPg^2Ph^2$, for $0<C<1$. According to Lemma A.6 of Murphy and van der Vaart (1997), there exists some $C>0$ such that $P(g+h)^2\ge Cd_2^2(\vartheta,\vartheta_0)$. Hence, $\mathbb{M}(\vartheta_0)-\mathbb{M}(\vartheta)\ge Cd_2^2(\vartheta,\vartheta_0)$, for $C>0$. Then, it implies

$$\sup_{d_2(\vartheta,\vartheta_0)\ge\varepsilon}\mathbb{M}(\vartheta)\le\sup_{d_2(\vartheta,\vartheta_0)\ge\varepsilon}(\mathbb{M}(\vartheta_0)-Cd_2^2(\vartheta,\vartheta_0))\le\mathbb{M}(\vartheta_0)-C\varepsilon^2<\mathbb{M}(\vartheta_0).$$

It follows from the Property 3.2 of Wang and Ghosh (2012) and the condition (C1), we can show there exists a $\psi_{0,N}\in\mathscr{M}_N$ of order N such that the approximation error

$$|\psi_0-\psi_{0,N}|\le\frac{3}{4}(N-1)^{-(1+a_0)/2}\approx\frac{3}{4}N^{-(1+a_0)/2}$$

when N is large enough. Assume $N=O(n^\kappa)$, we have

$$\sup_{0\le t\le 1}|\psi_0-\psi_{0,N}|\le\frac{3}{4}n^{-(1+a_0)\kappa/2}.$$

Denote $\widehat{\vartheta}_n=(\widehat{\beta}_n,\widehat{\psi}_n)$ and $\vartheta_{0,N}=(\beta_0,\psi_{0,N})$. We have

$$\mathbb{M}_n(\widehat{\vartheta}_n)-\mathbb{M}_n(\vartheta_0)\ge\mathbb{M}_n(\vartheta_{0,N})-\mathbb{M}_n(\vartheta_0)=I_{n_1}+I_{n_2},$$

where $I_{n_1}=(\mathbb{P}_n-P)\{l(\vartheta_{0,N})-l(\vartheta_0)\}$ and $I_{n_2}=\mathbb{M}(\vartheta_{0,N})-\mathbb{M}(\vartheta_0)$.

Define class

$$\mathscr{L}_2=\{l(\beta_0,\psi)-l(\beta_0,\psi_0):\psi\in\mathscr{F},\|\psi-\psi_0\|_\infty\le\eta\},$$

for $\eta=O(n^{-(1+a_0)\kappa/2})$. The fact that $\mathscr{F}$ is P-Donsker and conditions (C1) and (C4) yield $\mathscr{L}_2$ is P-Donsker. Thus, we have

$$I_{n_1}=(\mathbb{P}_n-P)\{l(\beta_0,\psi_{0,N})-l(\beta_0,\psi_0)\}=o_P(1).$$

Furthermore,

$$I_{n_2}=-\frac{1}{2\sigma^2}\|\psi_{0,N}-\psi_0\|_2^2=-O(n^{-(1+a_0)/(3+2a_0)})=-o(1).$$

We conclude that

$$\mathbb{M}_n(\widehat{\vartheta}_n) - \mathbb{M}_n(\vartheta_0) > -o_P(1).$$

The uniform convergence of $\mathbb{M}_n$ to $\mathbb{M}$ on $\Theta \times \mathscr{F}$ implies $\mathbb{M}_n(\vartheta_0) \to \mathbb{M}(\vartheta_0)$ in probability, it follows that $\mathbb{M}_n(\widehat{\vartheta}_n) \geq \mathbb{M}(\vartheta_0) - o_P(1)$. Therefore,

$$\mathbb{M}(\vartheta_0) - \mathbb{M}(\widehat{\vartheta}_n) \leq \mathbb{M}_n(\widehat{\vartheta}_n) - \mathbb{M}(\widehat{\vartheta}_n) + o_P(1) \leq \sup_{\vartheta \in \Theta \times \mathscr{F}} |\mathbb{M}_n - \mathbb{M}|(\vartheta) + o_P(1) \to 0$$

in probability. The last inequality holds because of the uniform convergence of $\mathbb{M}_n$ to $\mathbb{M}$ on $\Theta \times \mathscr{F}$.

For every $\varepsilon > 0$, by $\sup\limits_{d_2(\vartheta,\vartheta_0)\geq\varepsilon} \mathbb{M}(\vartheta) < \mathbb{M}(\vartheta_0)$, there exists a number $\eta > 0$, such that $\mathbb{M}(\vartheta) < \mathbb{M}(\vartheta_0) - \eta$, for every ϑ with $d_2(\vartheta, \vartheta_0) \geq \varepsilon$. Thus, the event $d_2(\widehat{\vartheta}_n, \vartheta_0) \geq \varepsilon$ is contained in the event $\{\mathbb{M}(\widehat{\vartheta}_n) < \mathbb{M}(\vartheta_0) - \eta\}$. The probability of latter event converges to 0 by the preceding display. This completes the proof of $d_2(\widehat{\vartheta}_n, \vartheta_0) = o_P(1)$.

Next we prove rate of convergence of part (ii). We apply Theorem 3.4.1 of van der Vaart and Wellner (1996) to prove the rate of convergence. Denote the regression function by $g(z) = X\beta + \psi(z)$. Denote $g_0(z) = X\beta_0 + \psi_0(z)$. In the proof of consistency, we show that there exists a $\psi_{0,N} \in \mathscr{M}_N$ of order N such that

$$\sup_{0\leq t\leq 1} |\psi_0 - \psi_{0,N}| \leq \frac{3}{4} N^{-(1+a_0)/2} \approx \frac{3}{4} n^{-(1+a_0)\kappa/2}.$$

Let $g_n(z) = X\beta_0 + \psi_{0,N}(z)$ and the estimate of $g_0(z)$ by $\widehat{g}_n(z) = X\widehat{\beta}_n + \widehat{\psi}_{0,N}(z)$. Define $l(g) = -1/(2\sigma^2)(Y - g)^2$ and $\mathbb{M}(g) = Pl(g)$. First we need to find $\phi_n(\eta)/\eta$ is decreasing in η and

$$\mathsf{E} \sup_{\eta/2\leq\|g-g_n\|_2\leq\eta} |\mathbb{G}_n l(g) - \mathbb{G}_n l(g_n)| \leq C\phi_n(\eta).$$

Define class

$$\mathscr{L}_3 = \{l(g) - l(g_n), \psi \in \mathscr{M}_N \text{ and } \|g - g_n\|_2 \leq \eta\}.$$

For any $0 < \varepsilon \leq \eta$, by the calculation of Shen and Wong (1994), the logarithm of the bracketing number of $\mathscr{M}_N$ computed with $L_2(P)$ can be bounded by $(N+1)\log(\eta/\varepsilon)$, up to a constant. Furthermore, by conditions (C1) and (C4), we can show that, for some $C > 0$, $J_{[\,]}(\eta, \mathscr{L}_3, \|\cdot\|_{P.B}) \leq CN^{1/2}\eta$, where $\|\cdot\|_{P.B}$ is the Bernstein norm defined as $\|f\|_{P.B} = \{2P(\mathrm{e}^{|f|} - 1 - |f|)\}^{1/2}$ in van der Vaart and Wellner (1996). Moreover, some algebra leads to $\|l(g) - l(g_n)\|_{P.B}^2 \leq C\eta^2$, for some $C > 0$ and any $l(g) - l(g_n) \in \mathscr{L}_3$. According to Lemma 3.4.3 of van der Vaart and Wellner (1996), we obtain

$$\mathsf{E}_P\|\mathbb{G}_n\|_{\mathscr{L}_3} \leq J_{[\,]}(\eta, \mathscr{L}_3, \|\cdot\|_{P.B})\Big(1 + \frac{J_{[\,]}(\eta, \mathscr{L}_3, \|\cdot\|_{P.B})}{n^{1/2}\eta^2}\Big) \leq C(N^{1/2}\eta + N/n^{1/2}).$$

Hence, we choose $\phi_n(\eta) = N^{1/2}\eta + N/n^{1/2}$. Clearly $\phi_n(\eta)/\eta$ is decreasing in η. Therefore, by Theorem 3.4.1 of van der Vaart and Wellner (1996), choosing the distance d_n defined in the theorem to be $d_n^2(\widehat{g}_n, g_n) = \mathbb{M}(g_n) - \mathbb{M}(\widehat{g}_n)$, when r_n satisfies $r_n^2\phi_n(1/r_n) = r_n^2(N^{1/2}r_n^{-1} + N/n^{1/2}) = O(n^{1/2})$, we have $r_n^2(\mathbb{M}(g_n) - \mathbb{M}(\widehat{g}_n)) = O_P(1)$. It follows that $r_n = n^{(1+a_0)/(3+2a_0)}$. Note that

$$\begin{aligned}\mathbb{M}(g_n) - \mathbb{M}(\widehat{g}_n) &= \mathbb{M}(g_n) - \mathbb{M}(g_0) + \mathbb{M}(g_0) - \mathbb{M}(\widehat{g}_n)\\ &= -\|\psi_n - \psi_{0,N}\|_2^2/(2\sigma^2) + \|\widehat{g}_n - g_0\|_2^2/(2\sigma^2).\end{aligned}$$

Hence,

$$\|\widehat{g}_n - g_0\|_2^2 = (\mathbb{M}(g_n) - \mathbb{M}(\widehat{g}_n))(2\sigma^2) + \|\psi_n - \psi_{0,N}\|_2^2 = O_P(r_n^{-2}).$$

In the proof of consistency we have already shown that $\|\widehat{g}_n - g_0\|_2^2 = 2\sigma^2(\mathbb{M}(\vartheta_0) - \mathbb{M}(\widehat{\vartheta}_n)) \ge Cd_2^2(\widehat{\vartheta}_n, \vartheta_0)$, for $C > 0$. Hence, $r_n^2 d_2^2(\widehat{\vartheta}_n, \vartheta_0) = O_P(1)$. $\square$

Proof of Theorem 3.2 In this section we apply the above theorem to show that the estimator $\widehat{\beta}_n$ for β_0 is asymptotically normality. Since $\widehat{\beta}_n$ is consistent and β_0 belongs to the interior of Θ by assumption, thus for sufficiently large n, we may assume that $\widehat{\beta}_n$ belongs to the interior of Θ with large probability. Because $(\widehat{\beta}_n, \widehat{\psi}_n)$ maximizes the log-likelihood (2.2), it must satisfy the stationary equation for $\widehat{\beta}_n$, that is,

$$\sum_{i=1}^n (Y_i - \widehat{\psi}_n(Z_i) - X_i\widehat{\beta}_n)X_i = 0. \tag{5.1}$$

For h^* defined in assumption (C8), let $\Phi = h^* \circ \psi_0^{-1}$ be the composite function of h^* and ψ_0^{-1}, where ψ_0^{-1} is the inverse of ψ_0. By the Theorem 1.3.6 of Robertson et al. (1988), we have

$$\sum_{i=1}^n (Y_i - \widehat{\psi}_n(Z_i) - X_i\widehat{\beta}_n)\Phi(\widehat{\psi}_n(Z_i)) = 0. \tag{5.2}$$

Let $h_0 = y - \psi_0(z) - x\beta_0$ and $\widehat{h}_n = y - \widehat{\psi}_n(z) - x\widehat{\beta}_n$. Combine (5.1) and (5.2) to get

$$\mathbb{P}_n\widehat{h}_n(x - \Phi(\widehat{\psi}_n(z)) = 0.$$

This equation can be rewritten as

$$P\widehat{h}_n(x - \Phi(\widehat{\psi}_n(z)) = -(\mathbb{P}_n - P)\widehat{h}_n(x - \Phi(\widehat{\psi}_n(z))). \tag{5.3}$$

Note that the integration is only with respect to (y, x, z). Because ε and (X, Z) independent,

$$P\varepsilon(x - \Phi(\widehat{\psi}_n(z)) = P(x - \Phi(\widehat{\psi}_n(z))P\varepsilon = 0. \tag{5.4}$$

Define class

$$\mathscr{L}_4 = \{(y - x\beta - \psi(z))(x - \Phi(\psi(z))) : \beta \in \Theta, \psi \in \mathscr{F}, \|\psi - \psi_0\|_\infty \le \eta\},$$

where η is defined in Theorem 3.1. As in the proof of Theorem 3.1, by considering the bracketing entropy of the class of $\mathscr{L}_4$ and using maximal inequality in Lemma 3.4.3 of van der Vaart and Wellner (1996), it can be shown that

$$(\mathbb{P}_n - P)\widehat{h}_n(x - \Phi(\widehat{\psi}_n(z))) = (\mathbb{P}_n - P)h_0(x - \Phi(\psi_0(z)) + o_P(n^{-1/2}). \tag{5.5}$$

Combine (5.3)-(5.5) to get

$$P\{(x(\widehat{\beta}_n - \beta_0) - \widehat{\psi}_n(z) + \psi_0(z))(x - \Phi(\widehat{\psi}_n(z)))\} = (\mathbb{P}_n - P)h_0(x - \Phi(\widehat{\psi}_n(z))) + o_P(n^{-1/2}).$$

However, by Theorem 3.1 and the Lipschitz condition on h^* given in condition (C8), we have

$$\begin{aligned} |P(\widehat{\psi}_n(z) - \psi_0(z))(x - \Phi(\widehat{\psi}_n(z)))| &= |P(\widehat{\psi}_n(z) - \psi_0(z))(\mathsf{E}(X|z) - \Phi(\widehat{\psi}_n(z)))| \\ &= |P(\widehat{\psi}_n(z) - \psi_0(z))(\Phi(\psi_0(z)) - \Phi(\widehat{\psi}_n(z)))| \\ &\le CP(\widehat{\psi}_n(z) - \psi_0(z))^2 \\ &= o_P(n^{-1/2}), \end{aligned}$$

it follows that

$$P(x(x - \Phi(\widehat{\psi}_n(z))))(\widehat{\beta}_n - \beta_0) = (\mathbb{P}_n - P)((y - x\beta_0 - \psi_0(z))(x - \mathsf{E}(X|z))) + o_P(n^{-1/n}).$$

Since

$$P(x(x - \Phi(\widehat{\psi}_n(z)))) \to P(x(x - \Phi(\psi_0(z)))) = P(x(x - \mathsf{E}(X|z))) = P(x - \mathsf{E}(X|x))^2$$

in probability, we have

$$P(x - \mathsf{E}(X|z))^2\sqrt{n}(\widehat{\beta}_n - \beta_0) = \sqrt{n}(\mathbb{P}_n - P)((y - x\beta_0 - \psi_0(z))(x - \mathsf{E}(X|z))) + o_P(1).$$

Thus, the result follows from condition (C7), the Central Limit Theorem, and Slutsky's lemma. This completes the proof of Theorem 3.2. □

References

[1] Bianco, A. and Boente, G., Robust estimators in semiparametric partly linear regression models, *Journal of Statistical Planning and Inference*, **122(1-2)**(2004), 229–252.

[2] Brunk, H.D., Estimation of isotonic regression (with discussion), In *Nonparametric Techniques in Statistical Inference* (Editor: Puri, M.L.), Cambridge University Press, Cambridge, 1970.

[3] Chak, P.M., Madras, N. and Smith, B., Semi-nonparametric estimation with Bernstein polynomials, *Economics Letters*, **89(2)**(2005), 153–156.

[4] Chang, I.S., Hsiung, C.A., Wu, Y.J. and Yang, C.C., Bayesian survival analysis using Bernstein polynomials, *Scandinavian Journal of Statistics*, **32(3)**(2005), 447–466.

[5] Chang, I.S., Chien, L.C., Hsiung, C.A., Wen, C.C. and Wu, Y.J., Shape restricted regression with random Bernstein polynomials, *IMS Lecture Notes - Monograph Series*, **54**(2007), 187–202.

[6] Chen, H., Convergence rates for parametric components in a partly linear model, *The Annals of Statistics*, **16(1)**(1988), 136–146.

[7] Chen, H. and Shiau, J.H., Data-driven efficient estimators for a partially linear model, *The Annals of Statistics*, **22(1)**(1994), 211–237.

[8] Curtis, S.M. and Ghosh, S.K., A variable selection approach to monotonic regression with Bernstein polynomials, *Journal of Applied Statistics*, **38(5)**(2011), 961–976.

[9] Engle, R.F., Granger, C.W.J., Rice, J. and Weiss, A., Semiparametric estimates of the relation between weather and electricity sales, *Journal of the American Statistical Association*, **81(394)**(1986), 310–320.

[10] Fan, J. and Zhang, J., Two-step estimation of functional linear models with applications to longitudinal data, *Journal of the Royal Statistical Society: Series B* (*Statistical Methodology*), **62(2)**(2000), 303–322.

[11] Green, P., Jennison, C. and Seheult, A., Analysis of field experiments by least squares smoothing, *Journal of the Royal Statistical Society: Series B* (*Methodological*), **47(2)**(1985), 299–315.

[12] Green, P.J. and Silverman, B.W., *Nonparametric Regression and Generalized Linear Models*, CRC Press, 1994.

[13] Hamilton, S.A. and Truong, Y.K., Local linear estimation in partly linear models, *Journal of Multivariate Analysis*, **60(1)**(1997), 1–19.

[14] Heckman, N.E., Spline smoothing in a partly linear model. *Journal of the Royal Statistical Society: Series B* (*Methodological*), **48(2)**(1986), 244–248.

[15] Huang, J., A note on estimating a partly linear model under monotonicity constraints, *Journal of Statistical Planning and Inference*, **107(1-2)**(2002), 343–351.

[16] Huang, J.Z., Wu, C.O. and Zhou, L., Varying-coefficient models and basis function approximations for the analysis of repeated measurements, *Biometrika*, **89(1)**(2002), 111–128.

[17] Goldfarb, D. and Idnani, A., A numerically stable dual method for solving strictly convex quadratic programs, *Mathematical Programming*, **27(1)**(1983), 1–33.

[18] Kaslow, R.A., Ostrow, D.G., Detels, R., Phair, J.P., Polk, B.F. and Rinaldo, C.R., The Multicenter AIDS Cohort Study: rationale, organization, and selected characteristics of the participants, *American Journal of Epidemiology*, **126(2)**(1987), 310–318.

[19] Lorentz, G.G., *Bernstein Polynomials*, Chelsea Publishing Company, New York, 1953.

[20] Mammen, E. and van der Geer, S., Penalized quasi-likelihood estimation in partial linear models, *The Annals of Statistics*, **25(3)**(1997), 1014–1035.

[21] Murphy, S.A. and van der Vaart, A.W., Semiparametric likelihood ratio inference, *The Annals of Statistics*, **25(4)**(1997), 1471–1509.

[22] Petrone, S., Random Bernstein polynomials, *Scandinavian Journal of Statistics*, **26(3)**(1999), 373–393.

[23] Robertson, T. and Wright, F.T., Consistency in generalized isotonic regression, *The Annals of Statistics*, **3(2)**(1975), 350–362.

[24] Robertson, T., Wright, F.T. and Dykstra, R.L., *Order Restricted Statistical Inference*, Wiley, New York, 1988.

[25] Robinson, P.M., Root-n-consistent semiparametric regression, *Econometrica*, **56(4)**(1988), 931–954.

[26] Rosen, J.B., The gradient projection method for nonlinear programming. part I. linear constraints, *Journal of the Society for Industrial and Applied Mathematics*, **8(1)**(1960), 181–217.

[27] Schimek, M.G., Semiparametric penalized generalized additive models for environmental research and epidemiology, *Environmetrics*, **20(6)**(2009), 699–717.

[28] Shen, X. and Wong, W.H., Convergence rate of sieve estimates, *The Annals of Statistics*, **22(2)**(1994), 580–615.

[29] Speckman, P., Kernel smoothing in partial linear models, *Journal of the Royal Statistical Society: Series B* (*Methodological*), **50(3)**(1988), 413–436.

[30] Stadtmuller, U., Asymptotic properties of nonparametric curve estimates, *Periodica Methematrica Hungarica*, **17(2)**(1986), 83–108.

[31] Sun, Z.M., Zhang, Z.Z. and Du, J., Semiparametric analysis of isotonic errors-in-variables regression models with missing response, *Communications in Statistics - Theory and Methods*, **41(11)**(2012), 2034–2060.

[32] Turlach, B.A. and Weingessel, A., *Quadprog* (*Version 1.5-3*), R package, 2010.

[33] van der Vaart, A.W., *Asymptotic Statistics*, Cambridge University Press, Cambridge, 1998.

[34] van der Vaart, A.W. and Wellner, J.A., *Weak Convergence and Empirical Processes*, Springer, New York, 1996.

[35] Wang, J. and Ghosh, S.K., Shape restricted nonparametric regression with Bernstein polynomials, *Computational Statistics & Data Analysis*, **56(9)**(2012), 2729–2741.

[36] Wright, F.T., The asymptotic behavior of monotone regression estimates, *The Annals of Statistics*, **9(2)**(1981), 443–448.

单调约束条件下部分线性模型的Bernstein多项式估计

丁建华[1,2] 张忠占[1]

([1]北京工业大学应用数理学院, 北京, 100124; [2]山西大同大学数学系, 大同, 037009)

本文提出单调约束条件下部分线性模型基于Bernstein多项式的最大似然估计. 我们利用单调Bernstein多项式逼近单调非参数函数. 在相当宽松的条件下给出估计的渐近性质和最优收敛率. 最后通过理论模拟和实例分析来评价提出的方法.

关键词: 经验过程, 最大似然估计, 单调Bernstein多项式, Monte Carlo.

学科分类号: O212.7.