An Adaptive Semiparametric Estimation for Partially Linear Models[1]

Kaiping Wang[1,2] and Lu Lin[1]

[1]School of Mathematics and System Sciences, Shandong University
Jinan 250100, P.R. China

[2]School of Management, Shandong University, Jinan 250100, P.R. China

Abstract

In this paper, we propose an adaptive semiparametric estimation for the nonparametric component of partially linear models. The new estimator is better than the usual nonparametric method in the sense that the convergence rate of its MSE adapts to the underlying models. The convergence rate is of optimal nonparametric rate $O(n^{-4/5})$ generally and can achieve the parametric rate $O(n^{-1})$ on some conditions. Simulation studies show that the new estimator outperforms the traditional nonparametric estimator.

Key words and phrases. Partially linear model, adaptability, adjustment, kernel estimation, local fitting.

1. Introduction

In recent years, there has been increasing interest and activity in the general area of partially linear regression smoothing. For a detailed account of this model, see the monograph by Härdle, Liang and Gao(2000).

In this paper, we mainly consider the following fixed design partially linear model

$$Y_i = x_i^\tau \beta + g(t_i) + \varepsilon_i, \quad i = 1, \cdots, n, \tag{1.1}$$

where $\{x_i\}$ and $\{t_i\}$ are respectively p-dimensional and scalar explanatory variables, data $\{Y_i\}$ are observed at (x_i, t_i), β is an unknown p-dimensional parameter, $g(t)$ is an unknown smooth function for $t \in [0, 1]$, and $\varepsilon_1, \cdots, \varepsilon_n$ are independent and identically distributed

[1]E-mail address: wkp@sdu.edu.cn. This paper is supported by RFDP project (20070422034) of China.

with mean zero and variance σ^2. Without loss of generality, assume that the design points $t_1, \cdots, t_n$ satisfy $0 \le t_1 < t_2 \cdots < t_n \le 1$.

Usually, the estimation procedure is designed as follows. For every given β, define an estimator of $g(t)$ by

$$g_n(t;\beta) = \sum_{i=1}^{n} W_{ni}(t)(Y_i - x_i^\tau \beta),$$

where $W_{ni}(t)$ is positive weight function depending on t and the design points $t_1, \cdots, t_n$. Replacing $g(t)$ by $g_n(t;\beta)$ in the model (1.1) and using the least squares (LS) criterion, we obtain the LS estimator of β as

$$\tilde{\beta}_{LS} = (\widetilde{X}^\tau \widetilde{X})^{-1} \widetilde{X}^\tau \widetilde{Y}, \tag{1.2}$$

where $\widetilde{X}^\tau = (\widetilde{x}_1, \cdots, \widetilde{x}_n)$ with $\widetilde{x}_j = x_j - \sum_{i=1}^n W_{ni}(t_j)x_i$ and $\widetilde{Y}^\tau = (\widetilde{Y}_1, \cdots, \widetilde{Y}_n)$ with $\widetilde{Y}_j = Y_j - \sum_{i=1}^n W_{ni}(t_j)Y_i$. Then, the nonparametric estimator of $g(t)$ is defined as

$$\tilde{g}_n(t) = \sum_{i=1}^{n} W_{ni}(t)(Y_i - x_i^\tau \tilde{\beta}_{LS}). \tag{1.3}$$

Under some regularity conditions, the mean squared error (MSE) of $\tilde{g}_n(t)$ satisfies

$$MSE(\tilde{g}_n(t)) = \frac{\sigma^2}{nh} + O(h^4) + o(n^{-1}h^{-1}) + o(h^4). \tag{1.4}$$

This suggests that the optimal choice for h is proportional to $n^{-1/5}$ and then $MSE(\tilde{g}_n(t))$ is proportional to $n^{-4/5}$.

Note that $n^{-4/5}$ is the standard convergence rate of nonparametric estimation. However, although we do not know which form $g(t)$ has, we are very interested in finding an estimation procedure to guarantee that the estimator of $g(t)$ has the following properties: if $g(t)$ is in fact a parametric function, then the estimator has parametric convergence rate; otherwise, the estimator has the nonparametric convergence rate. In other words, the estimation procedure should adapt to the model function $g(t)$.

In this paper, we will utilize a nonparametric adjustment to solve the above problem. Nonparametric adjustment technique has been used to improve the density estimation (Naito (2004), Hjort and Glad (1995) and Hjort and Jones (1996)). The proposed estimator can be viewed as semiparametric in such a case that it combines parametric and nonparametric methods. In the proposed approach, a parametric plug-in estimator is used as a crude guess of $g(t)$. This initial parametric approximation is adjusted via multiplication by a nonparametric factor. The resulting estimator can achieve a satisfactory convergence rate, even the parametric convergence rate, when regression function space has "good" properties in the sense that the underlying functions are sufficiently smooth or are already parametric.

It is worth pointing out that, in our estimation procedure, we need not to have any knowledge about the function $g(t)$. The convergence rate of the MSE of the new estimator is of optimal nonparametric rate $O(n^{-4/5})$ generally and can achieve the parametric rate $O(n^{-1})$ on some conditions. However, for the usual nonparametric estimator, its convergence rate is at most $O(n^{-4/5})$, and in this sense, the new estimator is better than the traditional nonparametric method.

This paper is organized as follows. In Section 2, a nonlinear least squares is used to obtain a series approximation to the nonparametric component and then this initial estimator is adjusted based on a local L_2-fitting criteria to establish the adaptive semiparametric estimator. The asymptotic behavior of the proposed estimator is investigated in Section 3. In Section 4, random design regression model as an extension is discussed. Simulation studies are reported in Section 5 to illustrate the theoretical conclusions. The technical proofs are relegated to Section 6.

2. The Estimation Procedure

We now introduce the semiparametric estimation procedure to construct our adaptive estimator. For every given β, the initial estimator of $g(t)$, denoted by $\tilde{g}_M(t,\beta) = \sum_{k=1}^M \tilde{\theta}_k p_k(t)$, can be obtained by minimizing

$$\frac{1}{n}\sum_{i=1}^n \Big\{(Y_i - x_i^\tau \beta) - \sum_{k=1}^M \theta_k p_k(t_i)\Big\}^2$$

for $\theta_k \in \Theta$, where Θ is parameter space, $p_k(t)$ are basis functions satisfying

$$\int_{-1}^1 p_k(t)dt = 0, \ \int_{-1}^1 p_j(t)p_k(t)dt = \begin{cases} 1, & \text{for } j = k \\ 0, & \text{otherwise,} \end{cases}$$

and M depends on n and tends to infinity as n tends to infinity. Let

$$\begin{aligned} P_M(t) &= (p_1(t), \cdots, p_M(t))^\tau \\ \boldsymbol{\theta}_M &= (\theta_1, \cdots, \theta_M)^\tau \\ \tilde{\boldsymbol{\theta}}_M(\beta) &= (\tilde{\theta}_1(\beta), \cdots, \tilde{\theta}_M(\beta))^\tau \\ &= \Big(n^{-1}\textstyle\sum_{i=1}^n P_M(t_i)P_M^\tau(t_i)\Big)^{-1} n^{-1}\textstyle\sum_{i=1}^n P_M(t_i)(Y_i - x_i^\tau\beta). \end{aligned}$$

Then the initial estimator can be expressed as

$$\tilde{g}_M(t,\beta) = P_M^\tau(t)\tilde{\boldsymbol{\theta}}_M(\beta).$$

To the above notations be well defined, we assume that the design points are well designed such that the matrix $\sum_{i=1}^n P_M(t_i)P_M^\tau(t_i)$ is positive definite. For the fixed design regression model, this assumption condition is somewhat difficult to be realized, but for the random design model it is common.

Next, we aim to multiple the initial estimator by an adjustment factor $\xi = \xi(t,\beta)$. We will give general criterion in Section 4 to determine the adjustment for random design models. We here define the following local L_2-fitting criterion for fixed design model as

$$q(t,\beta,\xi) = \frac{1}{h}\int_0^1 K\Big(\frac{u-t}{h}\Big)\{g(u) - (\tilde{g}_M(u,\beta)\xi)\}^2 du \tag{2.1}$$

for the solution of ξ. Let $s_0 = 0$, $s_i = (t_i + t_{i+1})/2, i = 1,\cdots,n-1$, and $s_n = 1$. By minimizing the criterion (2.1) with respect to ξ, we get the estimator of $\xi(t,\beta)$

$$\begin{aligned}\hat{\xi}(t,\beta) &= \frac{h^{-1}\int_0^1 K\big(\frac{u-t}{h}\big)g(u)\tilde{g}_M(u,\beta)du}{h^{-1}\int_0^1 K\big(\frac{u-t}{h}\big)\tilde{g}_M^2(u,\beta)du} \\ &= \frac{h^{-1}\sum_{i=1}^n\int_{s_{i-1}}^{s_i} K\big(\frac{u-t}{h}\big)g(u)\tilde{g}_M(u,\beta)du}{h^{-1}\int_0^1 K\big(\frac{u-t}{h}\big)\tilde{g}_M^2(u,\beta)du} \\ &\approx \frac{h^{-1}\sum_{i=1}^n\int_{s_{i-1}}^{s_i} K\big(\frac{u-t}{h}\big)\tilde{g}_M(u,\beta)du(Y_i - x_i^\tau\beta)}{h^{-1}\int_0^1 K\big(\frac{u-t}{h}\big)\tilde{g}_M^2(u,\beta)du} \\ &= \sum_{i=1}^n \tilde{w}_{ni}(t)(Y_i - x_i^\tau\beta),\end{aligned}$$

where

$$\tilde{w}_{ni}(t) = \frac{h^{-1}\int_{s_{i-1}}^{s_i} K\big(\frac{u-t}{h}\big)\tilde{g}_M(u,\beta)du}{h^{-1}\int_0^1 K\big(\frac{u-t}{h}\big)\tilde{g}_M^2(u,\beta)du}. \tag{2.2}$$

Consequently, the semiparametric estimator of $g(t)$ can be expressed as

$$\hat{g}_n(t,\beta) = \tilde{g}_M(t,\beta)\hat{\xi}(t,\beta) = \sum_{i=1}^n\sum_{j=1}^n w_{nij}(t)(Y_i - x_i^\tau\beta)(Y_j - x_j^\tau\beta),$$

where

$$w_{nij}(t) = P_M^\tau(t)\Big(n^{-1}\sum_{i=1}^n P_M(t_i)P_M^\tau(t_i)\Big)^{-1} n^{-1}P_M(t_i)\tilde{w}_{nj}(t). \tag{2.3}$$

By substituting $\hat{g}_n(t,\beta)$ for $g(t)$ in the model (1.1) we can obtain an LS estimator of β. However, such an estimator has no explicit representation. For convenience, we now use a $\sqrt{n}$-consistent estimator $\hat{\beta}$, e.g., $\tilde{\beta}_{LS}$ defined by (1.2), as a replacer of the estimator of β.

From the proof of Theorem 3.1, we can see that the $\sqrt{n}$-consistent condition is necessary and sufficient for the convergence of nonparametric component. Finally, the nonparametric estimator of $g(t)$ is defined as

$$\hat{g}_n(t) = \sum_{i=1}^{n}\sum_{j=1}^{n} w_{nij}(t)(Y_i - x_i^\tau \hat{\beta})(Y_j - x_j^\tau \hat{\beta}). \tag{2.4}$$

3. Asymptotic Behavior

In this section, we assume that the design points are generated by a positive, Lipschitz continuous function $\varphi(\cdot)$ in the sense that $t_i = Q\Big(\frac{i-1/2}{n}\Big)$, $i = 1, \cdots, n$, in which $Q(u) = F^{-1}(u)$ and $F(t) = \int_0^t \varphi(u)du$, $1 \le u \le 1$. We also assume that $K(\cdot)$ has support $[-1, 1]$, is Lipschitz continuous and $\int_{-1}^{1} K(u)du = 1$ and $\int_{-1}^{1} uK(u)du = 0$. From now on, we denote $J_K = \int_{-1}^{1} K^2(u)du$ and $\sigma_K^2 = \int_{-1}^{1} u^2K(u)du$.

Suppose that there exist $\theta_k \in \Theta$ such that $g(t) = \sum_{k=1}^{\infty} \theta_k p_k(t)$ and let remainder term $e_M(t) = \sum_{k=1}^{\infty} \theta_k p_k(t) - \sum_{k=1}^{M} \theta_k p_k(t)$. We need the following conditions.

(i) For $t \in (0, 1)$, there are functions $\epsilon_0(t)$, $\epsilon_1(t)$ and $\epsilon_2(t)$ such that $\epsilon_k(t) \not\equiv 0, k = 0, 1, 2,$ and

$$\lim_{M\to\infty} M^{\gamma_0} e_M(t) = \epsilon_0(t), \lim_{M\to\infty} M^{\gamma_1} (e_M(t))' = \epsilon_1(t), \lim_{M\to\infty} M^{\gamma_2} (e_M(t))'' = \epsilon_2(t)$$

for constants $\gamma_2 \le \gamma_1 \le \gamma_0$ and $\gamma_0 > 0$.

(ii) $E(\hat{\beta} - \beta)^2 = O(n^{-1})$.

Condition (i) presents the convergence rates of the remainder term and its derivatives, respectively. It also implies that the nonparametric component $g(t)$ should be smooth to some suitable extent such that the remainder term has a suitable rate to tend to zero. The decreasing relationship between the rates described by $\gamma_2 \le \gamma_1 \le \gamma_0$ is also common. For example, if $\{p_k\}$ is trigonometric function basis or polynomial basis, the remainder term has this property. Condition (ii) presents the standard convergence rate of parametric estimator and $\tilde{\beta}_{LS}$ defined by (1.2) satisfies it.

Let $h = O(n^{-\nu})$ and $M = O(n^{\delta})$ for $0 < \nu < 1$ and $0 < \delta < 1$. We have the following theorem.

Theorem 3.1 Assume that conditions (i) and (ii) hold, $g(t)$ and $p(t)$ have two continuous derivatives on (0,1). Then, for $t \in (0, 1)$, the bias and variance of $\hat{g}_n(t)$ can be expressed respectively as

$$Bias(\hat{g}_n(t)) = \tfrac{1}{2}h^2\sigma_K^2 M^{-\gamma_2}\epsilon_2(t) + o(h^2 M^{-\gamma_2}) + O(M^{-\gamma_0+1}) + O(n^{-1}M),$$
$$\text{Var}(\hat{g}_n(t)) = \tfrac{\sigma^2 J_K}{nh\varphi(t)} + O(n^{-1}) + O(n^{-2}h^{-2}).$$

The theorem above shows that, when $\delta < 2/(1+\gamma_0)$, the MSE of $\hat{g}_n(t)$ has the representation

$$MSE_h^* = \frac{\sigma^2 J_K}{nh\varphi(t)} + \frac{1}{4}h^4\sigma_K^4 M^{-2\gamma_2}(\epsilon_2(t))^2 + O(M^{-2(\gamma_0-1)}).$$

To determine the convergence rate of MSE_h^*, we need the following conditions:

(iii) $\gamma_0 - \gamma_2 > 1$.

(iv) $\gamma_0 - \gamma_2 \le 1$.

Then by minimizing MSE_h^*, we get the following corollary.

Corollary 3.1 (I) Under condition (iii), suppose the conditions of Theorem 3.1 hold, if ν and δ are chosen satisfying

$$\nu < 1/2,\ 2\nu/(\gamma_0 - \gamma_2 - 1) < \delta < \min\{2/(1+\gamma_0), (1-2\nu)/(1+\gamma_2)\},$$

then, for estimating $g(t)$, the optimal choice of h satisfies

$$h = \Big(\frac{\sigma^2 J_K}{\sigma_K^4\varphi(t)(\epsilon_2(t))^2}\Big)^{1/5}\Big(\frac{1}{nM^{-2\gamma_2}}\Big)^{1/5}.$$

In this case the the mean squared errors of $\hat{g}_n(t)$ satisfy

$$MSE(\hat{g}_n(t)) \sim \frac{\sigma^2 J_K}{\varphi(t)}\Big\{\frac{1}{4}\Big(\frac{\sigma^2 J_K}{\sigma_K^4\varphi(t)(\epsilon_2(t))^2}\Big)^{-1/5} + \Big(\frac{\sigma^2 J_K}{\sigma_K^4\varphi(t)(\epsilon_2(t))^2}\Big)^{1/5}\Big\}\Big(\frac{1}{M^{\gamma_2/4}n}\Big)^{4/5}$$

(II) Under condition (iv), suppose the conditions of Theorem 3.1 hold, if δ is chosen satisfying $\delta < 2/(1+\gamma_0)$, then, for any h,

$$MSE(\hat{g}_n(t)) \sim \frac{\sigma^2 J_K}{nh\varphi(t)} + O(M^{-2(\gamma_0-1)})$$

According to Corollary 3.1, we have the following surprising results.

(A) Conditions (iii) and (iv) are essential for the adaptability of our semiparametric estimator. Under condition (iii), the MSE has a standard asymptotic structure as in the classical nonparametric estimation. Under condition (iv), the MSE can achieve the convergence rate of parametric estimation. For example, if $\gamma_0 \geq 5/3$, δ is chosen to satisfy that $1/[2(\gamma_0 - 1)] \leq \delta < 2/(1+\gamma_0)$, and h is independent of n, then

$$MSE_h^* \sim O(n^{-1}).$$

This means that in a "good" regression function space $\mathcal{G}$, the convergence rate of the MSE of the new estimator can achieve the order of $O(n^{-1})$, the optimal rate in the parametric case.

(B) Under condition (iii), the convergence rate of the new estimator also depends on the property of the function space $\mathcal{G}$. If $\gamma_2 = 0$, the asymptotic order of bandwidth h is of order $O(n^{-1/5})$, the standard optimal convergence rate. If $\gamma_2 < 0$, the resulting estimator is under-smoothing. This means that, in a larger function space, the estimator has to be under-smoothing to guarantee optimal convergence rate. On the other hand, if $\gamma_2 > 0$, the resulting estimator is over-smoothing. This means that any function in a smaller function space can get a better first-stage estimator and, consequently, the second-stage estimator can be over-smoothing.

4. Extension

We now extend the previous method to random design regression model, i.e., $t_i = T_i$ are randomly designed in model (2.1). In this case, the regression function is defined as the conditional mean of Y given explanatory variable (X, T), and (X, T) and ε are unrelated.

Let $T_{(1)} \leq T_{(2)} \leq \cdots \leq T_{(n)}$ be the order statistics of $T_1, T_2, \cdots, T_n$ and $s_0 = 0, s_i = (T_{(i)} + T_{(i+1)})/2, i = 1, \cdots, n-1, s_n = 1$. In this case the semiparametric estimator of $g(T)$ has the same representation as in (2.4). Furthermore, if the random design points $T_1, T_2, \cdots, T_n$ are quasi-uniform, we can get the same properties as in the above theorem and corollary. Here we need the quasi-uniform design points to guarantee good approximations to the integrals as given in Section 2. The conclusions and the proofs are similar and thus the details are omitted.

However, the quasi-uniform condition is not a necessary condition if the local L_2-fitting criteria (2.1) is replaced by the following locally weighted least squares,

$$\sum_{i=1}^{n} \left\{ Y_i - [X_i^T \beta + \tilde{g}_M(T_i)\xi] \right\}^2 W_{ni}(t) \qquad (4.1)$$

where the weight functions $W_{ni}(t)$ depend on the distances $|T_i - t|$. Minimizing the criteria

(4.1), we get the estimators of $\xi(t)$ as

$$\hat{\xi}(t) = \frac{\sum_{i=1}^{n}\left\{Y_i - X_i^T\beta\right\}\tilde{g}_M(T_i)W_{ni}(t)}{\sum_{i=1}^{n}(\tilde{g}_M(T_i))^2 W_{ni}(t)}$$

Finally, the second-stage estimators of g can be expressed as

$$\hat{g}_n(t) = \tilde{g}_M(t)\frac{\sum_{i=1}^{n}\left\{Y_i - X_i^\tau\beta\right\}\tilde{g}_M(T_i,\beta)W_{ni}(t)}{\sum_{i=1}^{n}(\tilde{g}_M(T_i,\beta))^2 W_{ni}(t)} \tag{4.2}$$

The estimator (4.2) is similar to (2.4), and then has the similar properties as given before. The details are omitted.

5. Simulation Studies

Finite sample performance of the semiparametric estimator is investigated by simulations. In the following examples, the sample size is chosen as $n = 500$, the basis functions are chosen as sin-cosine functions, and the nonparametric estimator of $g(t)$ is obtained by local linear method.

Example 1. Consider the following model

$$Y_i = 0.75X_i + \sin(\pi T_i) + \varepsilon_i,\ i = 1, \cdots, 500, \tag{5.1}$$

where X_i are independently distributed as $U(0,1)$, $T_i = i/n$, and ε is distributed as $N(0, 0.125^2)$. The kernel function is chosen as $K(u) = 0.75(1-u^2)I(|u| < 1)$; the bandwidth is derived from the cross validation; the number of the basis functions is chosen to be $M = 4$.

Note that, in this model, the nonparametric component can be completely parameterized by finite basis functions. Then this example is designed for checking that, in a "good" regression space, the new semiparametric estimator can achieve a satisfactory convergence rate. In fact, a number of simulations show that the new estimator is better than the nonparametric method. Figure 1 reports the detail of one of the simulations we conducted.

(Figure 1 is about here)

Example 2. Consider the following model

$$Y_i = 0.75X_i + T_i^3 + \varepsilon_i,\ i = 1, \cdots, 500, \tag{5.2}$$

where X_i, T_i and ε are distributed as those in Example 1. The simulations we conducted again indicate that the new estimator of the nonparametric component g has the similar behavior as that of the nonparametric estimator. Figure 2 presents part of the simulations for model (5.2).

(Figure 2 is about here)

6. Proofs

Denote $g_M(t) = \sum_{k=1}^{M} \theta_k p_k(t)$ and

$$\hat{g}_M(t) = g_M(t)\frac{h^{-1}\sum_{i=1}^{n}(Y_i - x_i^\tau \hat{\beta})\int_{s_{i-1}}^{s_i} K\left(\frac{u-t}{h}\right) g_M(u)du}{h^{-1}\int_0^1 K\left(\frac{u-t}{h}\right) g_M^2(u)du}.$$

Lemma 6.1 *If the conditions of Theorem 3.1 hold, then*

$$\begin{aligned} Bias(\hat{g}_M(t)) &= \frac{\sigma_K^2 h^2}{2} M^{-\gamma_2}\epsilon_2(t) + o(h^2) + O(n^{-1}) + O(M^{-\gamma_0}), \\ Var(\tilde{g}_M(t)) &= \frac{\sigma^2 J_K}{nh\varphi(t)} + O\Big(n^{-1}\Big) + O\Big(n^{-2}h^{-2}\Big). \end{aligned}$$

Proof.

$$\begin{aligned} &E(\hat{g}_M(t)) \\ &= g_M(t)\frac{h^{-1}\sum_{i=1}^{n}(Y_i - x_i^\tau\beta + x_i^\tau\beta - x_i^\tau\hat{\beta})\int_{s_{i-1}}^{s_i} K\left(\frac{u-t}{h}\right)g_M(u)du}{h^{-1}\int_0^1 K\left(\frac{u-t}{h}\right)g_M^2(u)du} \\ &= g(t) + \frac{g_M(t)h^{-1}\sum_{i=1}^{n} g(t_i)\int_{s_{i-1}}^{s_i} K\left(\frac{u-t}{h}\right)g_M(u)du - g(t)h^{-1}\int_0^1 K\left(\frac{u-t}{h}\right)g_M^2(u)du}{h^{-1}\int_0^1 K\left(\frac{u-t}{h}\right)g_M^2(u)du} \\ &\quad + \frac{g_M(t)h^{-1}\sum_{i=1}^{n} x_i^\tau E(\beta - \hat{\beta})\int_{s_{i-1}}^{s_i} K\left(\frac{u-t}{h}\right)g_M(u)du}{h^{-1}\int_0^1 K\left(\frac{u-t}{h}\right)g_M^2(u)du} \\ &= g(t) + H_1 + H_2. \end{aligned}$$

The numerator of H_1 can be expressed as

$$\begin{aligned} &g(t)h^{-1}\int_0^1 K\left(\frac{u-t}{h}\right)g(u)g_M(u)du - g(t)h^{-1}\int_0^1 K\left(\frac{u-t}{h}\right)g_M^2(u)du \\ &\quad + O(n^{-1}) + O(M^{-\gamma_0}) \\ &= g(t)\Big(g(t)g_M(t) + \frac{\sigma_K^2 h^2}{2}(g(t)g_M(t))'' - g_M^2(t) - \frac{\sigma_K^2 h^2}{2}(g_M^2(t))''\Big) \\ &\quad + o(h^2) + O(n^{-1}) + O(M^{-\gamma_0}) \\ &= \frac{\sigma_K^2 h^2}{2} g^2(t) M^{-\gamma_2}\epsilon_2(t) + o(h^2) + O(n^{-1}) + O(M^{-\gamma_0}). \end{aligned}$$

Then $H_1 = \frac{\sigma_K^2 h^2}{2} M^{-\gamma_2}\epsilon_2(t) + o(h^2) + O(n^{-1}) + O(M^{-\gamma_0})$.

It can be easily verified that $H_2 = o(n^{-1})$. Consequently,

$$Bias(\hat{g}_M(t)) = \frac{\sigma_K^2 h^2}{2} M^{-\gamma_2}\epsilon_2(t) + o(h^2) + O(n^{-1}) + O(M^{-\gamma_0})$$

as required.

The remainder can be easily proved by the typical method, e. g. see the proof of Theorem 3.1 of Hart (1997), and then the detail is omitted here.

Proof of Theorem 3.1 Write

$$\begin{aligned} \boldsymbol{\theta}_M &= (\theta_1, \cdots, \theta_M)^\tau, \\ R_M(t) &= g(t) - \sum_{k=1}^{M} \theta_k p_k(t). \end{aligned}$$

Then the first-stage estimator of nonparametric component $g(t)$ can be expressed as

$$\begin{aligned} & P_M^\tau(t)\tilde{\boldsymbol{\theta}}_M \\ = \;& P_M^\tau(t)\Big(n^{-1}\sum_{i=1}^n P_M(t_i)\mathcal{P}_M^\tau(t_i)\Big)^{-1} n^{-1}\sum_{i=1}^n P_M(t_i)\Big(P_M^\tau(t_i)\boldsymbol{\theta}_M + R_M(t_i) + \varepsilon_i + x_i^\tau(\beta-\hat{\beta})\Big) \\ = \;& g_M(t) + P_M^\tau(t)\Big(n^{-1}\sum_{i=1}^n P_M(t_i)P_M^\tau(t_i)\Big)^{-1} n^{-1}\sum_{i=1}^n P_M(t_i)\Big(\varepsilon_i + R_M(t_i) + x_i^\tau(\beta-\hat{\beta})\Big), \end{aligned} \qquad (6.1).$$

Using Taylor expansion, we can expend $\hat{g}_n(t) - \hat{g}_M(t)$ at $g_M(t)$ and then get

$$\begin{aligned} & \hat{g}_n(t) - \hat{g}_M(t) \\ &= \tilde{g}_M(t,\hat{\beta})\frac{\sum_{i=1}^n (Y_i - x_i^\tau\hat{\beta})\int_{s_{i-1}}^{s_i} K\big(\frac{u-t}{h}\big)\tilde{g}_M(u,\hat{\beta})du}{\int_0^1 K\big(\frac{u-t}{h}\big)\tilde{g}_M^2(u,\hat{\beta})du} - g_M(t)\frac{\sum_{i=1}^n (Y_i - x_i^\tau\beta)\int_{s_{i-1}}^{s_i} K\big(\frac{u-t}{h}\big)g_M(u)du}{\int_0^1 K\big(\frac{u-t}{h}\big)g_M^2(u)du} \\ &= B(t)\Big(\tilde{g}_M(t,\hat{\beta}) - g_M(t)\Big) + o_p(n^{-1}M), \end{aligned} \qquad (6.2)$$

where

$$B(t) = \frac{\eta_1}{\eta_2} + g_M(T)\frac{\eta_3}{\eta_2} - 2g_M(T)\frac{\eta_1}{\eta_4^2}$$

with

$$\begin{aligned} \eta_1 &= \textstyle\sum_{i=1}^n \Big(Y_i - x_i^\tau\hat{\beta}\Big)\int_{s_{i-1}}^{s_i} K\Big(\frac{u-t}{h}\Big)g_M(u)du, \\ \eta_2 &= \textstyle\int_0^1 K\Big(\frac{u-u}{h}\Big)(g_M(u))^2du, \\ \eta_3 &= \textstyle\sum_{i=1}^n \Big(Y_i - x_i^\tau\hat{\beta}\Big)\int_{s_{i-1}}^{s_i} K\Big(\frac{u-t}{h}\Big)du, \\ \eta_4 &= \textstyle\int_0^1 K\Big(\frac{u-t}{h}\Big)g_M(u)du. \end{aligned}$$

From the results above and (6.1) it follows that

$$
\begin{aligned}
& E\Big(\eta_1\Big(\tilde g_M(t,\hat\beta)-g_M(t)\Big)\Big)\\
= \;& E\Big[\textstyle\sum_{i=1}^n \Big(g(t_i)+\varepsilon_i+x_i^\tau(\beta-\hat\beta)\Big)\int_{s_{i-1}}^{s_i}K\Big(\frac{u-t}{h}\Big)g_M(u)du\\
& \qquad \times P_M^\tau(t)\Big(n^{-1}\sum_{i=1}^n P_M(t_i)P_M^\tau(t_i)\Big)^{-1}n^{-1}\sum_{i=1}^n P_M(t_i)\Big(\varepsilon_i+R_M(t_i)+x_i^\tau(\beta-\hat\beta)\Big)\\
= \;& \textstyle\sum_{i=1}^n g(t_i)\int_{s_{i-1}}^{s_i}K\Big(\frac{u-t}{h}\Big)g_M(u)du\\
& \qquad \times P_M^\tau(t)\Big(n^{-1}\sum_{i=1}^n P_M(t_i)P_M^\tau(t_i)\Big)^{-1}n^{-1}\sum_{i=1}^n P_M(t_i)R_M(t_i)\\
& +\textstyle\sum_{i=1}^n g(t_i)\int_{s_{i-1}}^{s_i}K\Big(\frac{u-t}{h}\Big)g_M(u)du\\
& \qquad \times P_M^\tau(t)\Big(n^{-1}\sum_{i=1}^n P_M(t_i)P_M^\tau(t_i)\Big)^{-1}n^{-1}\sum_{i=1}^n P_M(t_i)x_i^\tau E(\beta-\hat\beta)\\
& +\frac{\sigma^2}{n}\textstyle\sum_{i=1}^n \int_{s_{i-1}}^{s_i}K\Big(\frac{u-t}{h}\Big)g_M(u)du P_M^\tau(t)\Big(n^{-1}\sum_{i=1}^n P_M(t_i)P_M^\tau(t_i)\Big)^{-1}n^{-1}\sum_{i=1}^n P_M(t_i)\\
& +E(\sum_{i=1}^n x_i^\tau(\beta-\hat\beta)\int_{s_{i-1}}^{s_i}K\Big(\frac{u-t}{h}\Big)g_M(u)du\\
& \qquad \times P_M^\tau(t)\Big(n^{-1}\sum_{i=1}^n P_M(t_i)P_M^\tau(t_i)\Big)^{-1}n^{-1}\sum_{i=1}^n P_M(t_i)\Big(\varepsilon_i+R_M(t_i)+x_i^\tau(\beta-\hat\beta)\Big)\\
= \;& O(hM^{-\gamma_0+1})+O(hM^{-2\gamma_0+1})+O(hn^{-1}M)\\
= \;& O(hM^{-\gamma_0+1})+O(hn^{-1}M).
\end{aligned}
$$

Similarly,

$$E\Big[\eta_3\Big(\tilde g_M(t,\hat\beta)-g_M(t)\Big)\Big]=O(hM^{-\gamma_0+1})+O(hn^{-1}M).$$

Consequently,

$$E\Big[B(t)\Big(\tilde g_M(t,\hat\beta)-g_M(t)\Big)\Big]=O(M^{-\gamma_0+1})+O(n^{-1}M).$$

Combining the results above and Lemma 6.1 leads to

$$Bias(\hat g_n(t))=\frac{1}{2}h^2\sigma_K^2M^{-\gamma_2}\epsilon_2(t)+o(h^2M^{-\gamma_2})+O(M^{-\gamma_0+1})+O(n^{-1}M),$$

as required.

Using the similar method and Lemma 6.1, we can prove

$$Var(\hat g_n(t)) \;=\; \frac{\sigma^2 J_K}{nh\varphi(t)}+O(n^{-1})+O(n^{-2}h^{-2}),$$

as required.

References

Härdle, W., Liang, H., Gao, J. (2000), *Partially Linear Models* (Physica-Verlag, Heidelberg).

Hart, J. D. (1997), *Nonparametric smoothing and lack-of-fit tests* (Springer, New York).

Hjort, N. L. and Glad, I. K. (1995), Nonparametric density estimation with a parametric start, *Ann. Statist*, **23**, 882-904.

Hjort, N. L. and Jones, M. C. (1996), Locally parametric nonparametric density estimation, *Ann. Statist.*, **24**, 1619-1647.

Naito, K. (2004), Semiparametric density estimation by local L_2-fitting, *Ann. Statist.*, **32**, 1162-1191.

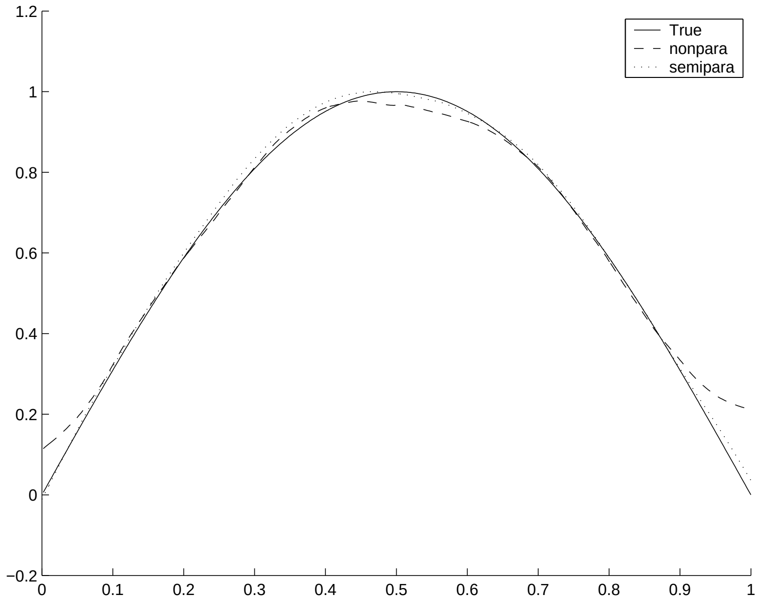

Figure 1. The estimation curve of the nonparametric component g in model (5.1).

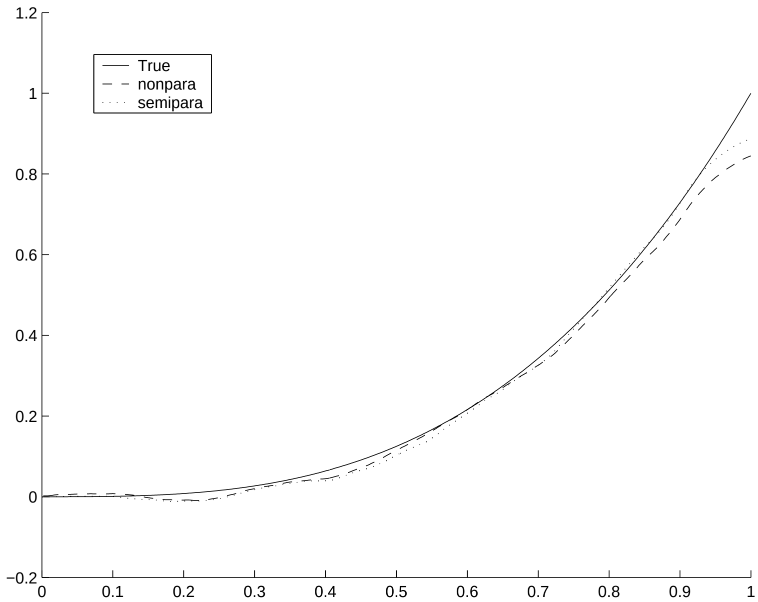

Figure 2. The estimation curve of the nonparametric component g in model (5.2).