Statistics and Its Interface Volume 11 (2018) 19–29

Monotone function estimation in partially linear models

Yi Zhang* and Shaoli Wang†

A kernel-based method is proposed for the monotone estimation of the nonparametric function component of a partially linear regression model. The estimated monotone function is constructed via a density estimate and numerical inversion. This procedure does not require constrained optimization and hence is fast to compute. Asymptotic normality is established for the proposed monotone function estimator. We apply the proposed method to analyze mammalian eye gene expression data and reveal a complex nonlinear relation within a gene network; we also analyze the German SOEP data using our method and validate the human capital theory.

AMS 2000 subject classifications: Primary 62G05.
Keywords and phrases: Asymptotic normality, Density estimation, Kernel estimation, Monotone function, Nonparametric function, Partially linear models.

1. INTRODUCTION

A major goal in biomedical research is to discover genetic mechanisms that underline complex heritable diseases. Recent advances in microarray technology have enabled researchers to use expression quantitative trait locus (eQTL) mapping to identify genetic variation that are relevant to human diseases. The Bardet-Biedl syndrome (BBS) is a genetically heterogeneous disorder that is characterized by multiple symptoms including retinopathy [1, 2, 14]. EQTL has identified TRIM32, an E3 ubiquitin ligase, as a BBS gene [9, 27]. Using a novel sure independence screening procedure, [24] discovered that other genes may be functionally related to gene TRIM32; some genes are linearly correlated with gene TRIM32, some are nonlinearly correlated with gene TRIM32. This motivates us to use a partially linear model to study the functional relation between gene TRIM32 and other genes that are potentially involved in regulation of BBS.

A partially linear model is given by

$$Y = \boldsymbol{X}^T\boldsymbol{\beta} + g(\boldsymbol{T}) + \epsilon, \quad (1)$$

where Y is the response variable, $\boldsymbol{X} = (X_1, \ldots, X_p)^T$ and $\boldsymbol{T} = (T_1, \ldots, T_d)^T$ are two vectors of explanatory variables. $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^T$ is a vector of unknown parameters, $g : \mathbb{R}^d \to \mathbb{R}^1$ is an unknown function, ϵ is a random error with mean zero and finite variance. In the above eQTL mapping application, Y is the expression level of gene TRIM32, $\boldsymbol{X}$ denotes the expression levels of genes that are linearly correlated with TRIM32, and $\boldsymbol{T}$ the expression levels of genes that are nonlinearly correlated with TRIM32. Since it contains a parametric part $\boldsymbol{X}^T\boldsymbol{\beta}$ and a nonparametric part $g(\cdot)$, model (1) is a semiparametric regression model. The partially linear model is more flexible than the fully parametric (linear regression) models because it allows a nonparametric component $g(\cdot)$, hence reduces the possibility of model misspecification. On the other hand, with a parametric part $\boldsymbol{X}^T\boldsymbol{\beta}$, model (1) has an advantage over a fully nonparametric model as it diminishes the "curse of dimensionality" of fully nonparametric models. In their pioneering work [11], Engle, Granger, Rice and Weiss (1986) modeled the electricity demand as a sum of a nonparametric function of temperature and a linear function of electricity price and income. Since then, the partially linear models have found many applications in various fields; see [32, 28, 18], among others. The monograph [18] gives a detailed account of recent advances in partially linear models.

Several methods have been proposed for the estimation of the linear coefficients $\boldsymbol{\beta}$ and function $g(\cdot)$ in model (1). Penalized least squares approach was introduced by [11, 15, 31, 33, 34], among others. [6] established a parametric convergence rate for parameter vector $\boldsymbol{\beta}$ while employing a piecewise polynomial estimator for the nonparametric function $g(\cdot)$. [32] proposed to use kernel smoothing for estimation, and obtained the asymptotic bias and variance for $\boldsymbol{\beta}$. [7] used a two-stage spline smoothing method for estimating the parametric and nonparametric components, and showed that the parametric component can be estimated at the parametric rate without undersmoothing the nonparametric component. Many other methods have also been proposed for estimation of model (1), for example, profiled likelihood [29, 4], empirical likelihood [30], local linear estimation [17], wavelet thresholding [5, 13].

Without any parametric assumptions on the function form, the nonparametric component $g(\cdot)$ in model (1) provides a flexibility in modeling. [24] found out that the nonlinear relationship between gene TRIM32 and gene Zmat1 appears to be a monotone function. In fact, it is desired in

*Ph.D. student at Shanghai University of Finance and Economics, and Lecturer at Shanghai Lixin University of Accounting and Finance.
†Corresponding author. Associate Professor at Shanghai University of Finance and Economics.

many applications that the nonparametric component $g(\cdot)$ be a monotone function. To study the monotone estimation of $g(\cdot)$, let us begin with a completely nonparametric regression model:

$$Y = m(X) + \epsilon.$$

For this model, there has been a vast amount of research devoted to the estimation of a monotone regression function. Estimation of the monotone nonparametric function can be based on maximum likelihood [3, 23, 8, 35, 12, 21], constrained spline smoothing [25, 19, 22], or kernel smoothing [16, 10]. In this paper we will adopt the estimation procedure given by [10], which is easy to carry out and has good sampling properties. First, [10] constructs a kernel estimate of m^{-1} by

$$\frac{1}{Nh_d}\int_{-\infty}^{t}\sum_{k=1}^{N} K_d\left(\frac{m(U_k)-u}{h_d}\right)du, \tag{2}$$

where the independent identically distributed random variables $U_1, \dots, U_N$ are from the uniform distribution over interval $[0,1]$, and N is a positive integer. Then, by replacing U_k with $k/N, k = 1, \dots, N$ in (2), they define the estimate of m^{-1} as

$$\hat{m}^{-1}(t) = \frac{1}{Nh_d}\int_{-\infty}^{t}\sum_{i=1}^{N} K_d\left(\frac{\hat{m}(i/N)-u}{h_d}\right)du,$$

where $\hat{m}$ is a kernel estimator of the nonparametric function, K_d is a symmetric kernel function. The monotone estimate of the regression function $m(\cdot)$ is finally defined as the inverse of the function $\hat{m}^{-1}$.

In this paper, we assume that $d = 1$ and the unknown nonparametric function $g(\cdot) : [0,1] \to \mathbb{R}$ is monotonous and twice continuously differentiable. We propose a monotone estimation procedure for the nonparametric component $g(\cdot)$ in the partially linear model (1) based on the idea of [10], and establish the asymptotic normality for the proposed estimator of the monotone function. The performance of the estimation procedure is illustrated via simulations and real data analysis.

The paper is organized as follows. In Section 2 we introduce the monotone estimation procedure for $g(\cdot)$ and discuss the bandwidth selection problem. In Section 3 we study the asymptotic properties of the proposed estimator of the monotone function. Simulations and applications to mammalian eye gene expression data and German SOEP data are presented in Section 4. Some discussions are given in Section 5. Technical conditions and proofs are given in the Appendix A.

2. ESTIMATION PROCEDURE

2.1 Estimation method

In this section, we describe the monotone estimation procedure. Assume that $\{(\boldsymbol{X}_i, Y_i, T_i), i = 1, \dots, n\}$ is a random sample from the population $(\boldsymbol{X}, Y, T)$, $\boldsymbol{X}_i = (X_{i1}, X_{i2}, \dots, X_{ip})^T$. We can obtain the Nadaraya-Watson estimate (referred to simply as the N-W estimate in this article) of $g(\cdot)$ as in [32]. If we define $Z = Y - \boldsymbol{X}^T\boldsymbol{\beta}$, then model (1) becomes a nonparametric regression model $Z = g(T) + \epsilon$. Consequently, the nonparametric function g can be estimated by kernel smoothing:

$$\hat{g}(t) = \frac{\sum_{i=1}^{n} K_r\left(\frac{t-T_i}{h_r}\right) Z_i}{\sum_{j=1}^{n} K_r\left(\frac{t-T_j}{h_r}\right)},$$

where $K_r(\cdot)$ is a probability kernel function, h_r is a bandwidth. To simplify the notations, we define

$$\omega_{ni}(t) = \frac{K_r\left(\frac{t-T_i}{h_r}\right)}{\sum_{j=1}^{n} K_r\left(\frac{t-T_j}{h_r}\right)}, \quad V = \begin{pmatrix} \omega_{n1}(T_1)\cdots\omega_{nn}(T_1) \\ \vdots \qquad \ddots \qquad \vdots \\ \omega_{n1}(T_n)\cdots\omega_{nn}(T_n) \end{pmatrix}.$$

Then we can replace $g(T)$ with $\sum_{i=1}^{n}\omega_{ni}(T)(Y_i - \boldsymbol{X}_i^T\boldsymbol{\beta})$ for every given $\boldsymbol{\beta}$. After making a change of variables $\tilde{\boldsymbol{X}} = \boldsymbol{X}^T - V\boldsymbol{X}^T, \tilde{Y} = Y - VY$, we can obtain the least squares estimate of $\boldsymbol{\beta}$, the N-W estimate of $g(\cdot)$ and the estimate of σ^2:

$$\begin{aligned} \hat{\boldsymbol{\beta}}_{LS} &= (\tilde{\boldsymbol{X}}^T\tilde{\boldsymbol{X}})^{-1}\tilde{\boldsymbol{X}}^T\tilde{Y}, \\ \hat{g}(t) &= \sum_{i=1}^{n}\omega_{ni}(t)(Y_i - \boldsymbol{X}_i^T\hat{\boldsymbol{\beta}}_{LS}), \\ \hat{\sigma}^2 &= \frac{1}{n}\sum_{i=1}^{n}(\tilde{Y}_i - \tilde{\boldsymbol{X}}_i\hat{\boldsymbol{\beta}}_{LS})^2. \end{aligned} \tag{3}$$

$\hat{g}(t)$ in (3) is a consistent estimate of $g(t)$, but not necessary a monotonous function. We use the method in [10] to guarantee the monotonicity. If $g(\cdot)$ is a strictly increasing function on the interval $[0,1]$, we define

$$\hat{g}_I^{-1}(t) = \frac{1}{Nh_d}\int_{-\infty}^{t}\sum_{k=1}^{N} K_d\left(\frac{\hat{g}(k/N)-u}{h_d}\right)du \tag{4}$$

as the estimate of $g^{-1}(t)$, where N is a large enough positive integer. In practice, if the sample size n is large enough, one may use $N = n$. As a result, the monotonous estimate $\hat{g}_I$ of g is obtained as the inverse function of $\hat{g}_I^{-1}(t)$. In order to study the consistency of $\hat{g}_I$, we replace the N-W estimate $\hat{g}(\cdot)$ of equation (4) with the true value $g(\cdot)$ and define

$$g_N^{-1}(t) = \frac{1}{Nh_d}\int_{-\infty}^{t}\sum_{k=1}^{N} K_d\left(\frac{g(\frac{k}{N})-u}{h_d}\right)du.$$

By Lemma 2.2 of [10], $g_N(t)$ is a good approximation of $g(t)$ under certain assumptions. If $g(\cdot)$ is a strictly decreasing function on the interval $[0,1]$, then we define

$$(5) \qquad \hat{g}_D^{-1}(t) = 1 - \frac{1}{Nh_d}\int_{-\infty}^{t}\sum_{k=1}^{N} K_d\left(\frac{\hat{g}(k/N)-u}{h_d}\right)du$$

as the estimate of $g^{-1}(t)$. The monotonous estimate of $g(\cdot)$ can be obtained similarly as in the strictly increasing case.

2.2 Bandwidth selection

Bandwidth selection is a fundamental issue in nonparametric smoothing. We will use the method from [18] to determine the bandwidth h_r for each sample. For any fixed h and $1 \le i \le n$, we define

$$\tilde{\omega}_{i,n}(t) = K_r\left(\frac{T_i - t}{h}\right) \Big/ \sum_{j=1, j\neq i}^{n} K_r\left(\frac{T_j - t}{h}\right),$$

$$\tilde{g}_{i,n}(t, \boldsymbol{\beta}) = \sum_{j=1, j\neq i}^{n} \tilde{\omega}_{j,n}(t)(Y_j - \boldsymbol{X}_j^T\boldsymbol{\beta}),$$

and

$$\tilde{\boldsymbol{\beta}}(h) = \arg\min_{\boldsymbol{\beta}} \sum_{i=1}^{n}\{Y_i - \boldsymbol{X}_i^T\boldsymbol{\beta} - \tilde{g}_{i,n}(T_i, \boldsymbol{\beta})\}^2.$$

The cross-validation (CV) function is

$$CV(h) = \frac{1}{n}\sum_{i=1}^{n}\{Y_i - \boldsymbol{X}_i^T\tilde{\boldsymbol{\beta}}(h) - \tilde{g}_{i,n}(T_i, \tilde{\boldsymbol{\beta}}(h))\}^2.$$

As in [18], the optimal bandwidth is determined as $h_r = \arg\min_{h\in\Theta_h} CV(h)$, with $\Theta_h = [\lambda_1 n^{-1/5-\eta}, \lambda_2 n^{-1/5+\eta}]$, where $\lambda_1, \lambda_2, \eta$ are constants which satisfy that $0 < \lambda_1 < \lambda_2 < \infty, 0 < \eta < 1/20$.

As discussed in Section 3, the asymptotic distribution of monotonous estimator of $g(\cdot)$ depends on the asymptotic behavior of the ratio h_r/h_d. In fact, there are two cases, $\lim_{n\to\infty}\frac{h_r}{h_d} = a \in [0,\infty)$ and $\lim_{n\to\infty}\frac{h_r}{h_d} = \infty$. [10] argued that bandwidths satisfying $h_d = o(h_r)$ are preferred if we want to minimize the MSE of the estimator. Hence we use $h_d = h_r^3$ as in [10].

3. PROPERTIES

It is shown in [18] that the least squares estimator of $\boldsymbol{\beta}$ has the convergence rate of $\sqrt{n}$. In addition, the N-W estimate of $g(\cdot)$ converges uniformly to the true value with certain rate. In this section, we will study the asymptotic behaviors of the proposed monotonous function estimator. Six assumptions given in the Appendix A will be used in the deduction of the sampling properties of our estimators.

Lemma 1 ([18] Theorem 2.1.1)**.** *Under assumptions 1 and 2, $\hat{\boldsymbol{\beta}}_{LS}$ is an asymptotically normal estimator of $\boldsymbol{\beta}$, i.e.,*

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_{LS} - \boldsymbol{\beta}) \xrightarrow{L} \mathcal{N}(0, \sigma^2\Sigma^{-1}).$$

Lemma 2 ([18] Theorem 3.2.2)**.** *Assume that assumptions 1, 3 and 4 hold, and that $E|\epsilon_1|^3 < \infty$. Then*

$$\sup_t |\hat{g}(t) - g(t)| = O\left(\left(\frac{\log n}{nh_r}\right)^{1/2}\right) + O(h_r^2) + O_p(n^{-1/2}).$$

Lemma 3 ([10] Lemma 2.2)**.** *If $g(\cdot)$ is strictly increasing and assumption 5 is satisfied, then for any $t \in (0,1)$ with $g'(t) > 0$*

$$g_N(t) = g(t) + \kappa_2(K_d)h_d^2\frac{g''(t)}{(g'(t))^2} + o(h_d^2) + O(\frac{1}{Nh_d}),$$

where the constant $\kappa_2(K)$ is given by

$$\kappa_2(K) = \frac{1}{2}\int_{-1}^{1} v^2K(v)dv.$$

The asymptotic biases, variances, and normality of the proposed monotonous function estimator are given in the next two theorems. The proofs are given in the Appendix A.

Theorem 3.1. *Assume that assumptions 1-6 hold. If $\lim_{n\to\infty}\frac{h_r}{h_d} = a \in [0,\infty)$, $nh_d^7 = o(1)$, $\frac{(\log n)^2}{nh_r^2 h_d} = o(1)$ and g is strictly increasing, then for all $t \in (g(0), g(1))$ with $g'(g^{-1}(t)) > 0$, we have*

$$\sqrt{nh_d}\left\{\hat{g}_I^{-1}(t) - g_n^{-1}(t) + \mathcal{B}\circ(g^{-1}(t))\right\} \xrightarrow{d} \mathcal{N}(0, r^2(t)),$$

where

$$\begin{aligned}\mathcal{B}(u) =& h_r^2\kappa_2(K_r)\frac{g''(u)f_T(u) + 2f_T'(u)g'(u)}{f_T(u)g'(u)},\\ r^2(t) =& \frac{\sigma^2}{g' f_T}\circ(g^{-1}(t))I(t),\\ I(t) =& \iiint K_d\left(p + ag'(g^{-1}(t)(q-s))\right)K_d(p)\\ &\cdot K_r(q)K_r(s)dpdqds.\end{aligned}$$

If $\lim_{n\to\infty}\frac{h_r}{h_d} = \infty$, $\frac{nh_r^9}{h_d^2} = o(1)$, $\frac{(\log n)^2}{nh_r h_d^2} = o(1)$ and g is strictly increasing, then for all $t \in (g(0), g(1))$ with $g'(g^{-1}(t)) > 0$, we have

$$\sqrt{nh_d}\left\{\hat{g}_I^{-1}(t) - g_n^{-1}(t) + \mathcal{B}\circ(g^{-1}(t))\right\} \xrightarrow{d} \mathcal{N}(0, \tilde{r}^2(t)),$$

where

$$\tilde{r}^2(t) = \frac{\sigma^2}{g'^2 f_T}\circ(g^{-1}(t))\int K_r^2(s)ds.$$

Theorem 3.2. *Assume that assumptions 1-6 hold. If $\lim_{n\to\infty}\frac{h_r}{h_d} = a \in [0,\infty)$, $nh_d^7 = o(1)$, $\frac{(\log n)^2}{nh_r^2 h_d} = o(1)$ and g is strictly increasing, then for all $t \in (0,1)$ with $g'(t) > 0$, we have*

$$\sqrt{nh_d}\left\{\hat{g}_I(t) - g_n(t) - \mathcal{D}(t)\right\} \xrightarrow{d} \mathcal{N}(0, s^2(t)),$$

where

$$\mathcal{D}(t) = h_r^2\kappa_2(K_r)\frac{g''(t)f_T(t) + 2f_T'(t)g'(t)}{f_T(t)},$$

Table 1. Mean squared errors of estimated $\boldsymbol{\beta}$ and σ^2

$\varepsilon, g(\cdot)$	n	h_r	β_1 (10^{-5})	β_2 (10^{-5})	β_3 (10^{-5})	σ^2 (10^{-6})
$N(0, 0.01)$ $g(T) = T^3$	100	0.0939	1.433	0.572	1.272	3.6644
	200	0.076	7.199	2.405	5.892	1.6554
	400	0.0636	3.431	1.246	2.517	0.68644
$N(0, 0.01)$ $g(T) = \frac{-1}{0.2+e^{10T}}$	100	0.073	1.634	0.573	1.363	13.475
	200	0.0569	7.161	2.406	5.75	6.6355
	400	0.0458	3.439	1.161	3.058	3.5857
$0.05\sqrt{2}(\chi_1^2 - 1)$ $g(T) = T^3$	100	0.0951	1.514	0.515	1.251	4.3981
	200	0.0769	7.349	2.729	5.606	1.8753
	400	0.0643	3.228	1.195	2.801	0.79659
$0.05\sqrt{2}(\chi_1^2 - 1)$ $g(T) = \frac{-1}{0.2+e^{10T}}$	100	0.0711	1.562	0.547	1.321	13.821
	200	0.0563	7.642	2.476	6.186	7.0492
	400	0.0461	3.474	1.127	2.957	3.6179

$$s^2(t) = \frac{\sigma^2 g'(t)}{f_T(t)} I(t),$$

$$I(t) = \iiint K_d\left(p + a g'(t)(q - s)\right) K_d(p) \cdot K_r(q) K_r(s)\, dp\, dq\, ds.$$

If $\lim_{n\to\infty} \frac{h_r}{h_d} = \infty$, $\frac{n h_r^9}{h_d^2} = o(1)$, $\frac{(\log n)^2}{n h_r h_d^2} = o(1)$ *and* g *is strictly increasing, then for all* $t \in (0,1)$ *with* $g'(t) > 0$*, we have*

$$\sqrt{n h_r}\left\{\hat{g}_I(t) - g_n(t) - \mathcal{D}(t)\right\} \xrightarrow{d} \mathcal{N}(0, \tilde{s}^2(t)),$$

where

$$\tilde{s}^2(t) = \frac{\sigma^2}{f_T(t)} \int K_r^2(u)\, du.$$

4. SIMULATION AND APPLICATIONS

4.1 Simulation study

In this part we will apply our method to simulated datasets. We generate data from a partially linear model

$$Y_i = \boldsymbol{X}_i^T \boldsymbol{\beta}_0 + g(T_i) + \varepsilon_i$$

for $i = 1, 2, \ldots, n$, where $\boldsymbol{\beta}_0 = (1.2, 1.3, 1.4)^T, T_i \sim U[0,1]$ and

$$\boldsymbol{X}_i \sim N(0, \Sigma_x) \text{ with } \Sigma_x = \begin{pmatrix} 0.81 & 0.1 & 0.2 \\ 0.1 & 2.25 & 0.1 \\ 0.2 & 0.1 & 1 \end{pmatrix}.$$

We take two cases for the distribution of ε_i, namely $\varepsilon_i \sim N(0, 0.01), \varepsilon_i \sim 0.05\sqrt{2}(\chi_1^2 - 1)$, where $\chi_1^2 - 1$ is a chisquare distribution with 1 degree of freedom minus 1. The nonparametric function $g(\cdot)$ takes two forms, namely $g(T) = T^3$ and $g(T) = -1/(0.2 + e^{10T})$.

In the simulation, we use sample sizes $n = 200$, 400 and 800, and for each sample size, we replicate 1000 times. Epanechnikov kernel function is used for smoothing. The mean squared errors(MSEs) for $\boldsymbol{\beta}$ and σ^2 are reported in Table 1. Table 2 compares the means and standard deviations of the square root of the average squared errors(RASEs), where

Table 2. Mean and standard deviation of RASEs for N-W estimate and monotone estimate

$\varepsilon, g(\cdot)$	n	RASE of $\hat{g}$	RASE of $\hat{g}_I$
$N(0, 0.01)$ $g(T) = T^3$	100	.0336(.0080)	.0311(.0075)
	200	.0248(.0053)	.0230(.0049)
	400	.0187(.0036)	.0174(.0034)
$N(0, 0.01)$ $g(T) = -1/(0.2 + e^{10T})$	100	.0399(.0091)	.0358(.0092)
	200	.0298(.0054)	.0256(.0053)
	400	.0216(.0037)	.0190(.0035)
$0.05\sqrt{2}(\chi_1^2 - 1)$ $g(T) = T^3$	100	.0336(.0092)	.0307(.0076)
	200	.0249(.0055)	.0231(.0050)
	400	.0185(.0037)	.0173(.0032)
$0.05\sqrt{2}(\chi_1^2 - 1)$ $g(T) = -1/(0.2 + e^{10T})$	100	.0388(.0107)	.0354(.0104)
	200	.0288(.0065)	.0260(.0064)
	400	.0216(.0041)	.0190(.0041)

$$RASE^2 = N^{-1} \sum_{j=1}^{N} \{\hat{g}(u_j) - g(u_j)\}^2.$$

We compare the unconstrained N-W estimate $\hat{g}$ (left) with the monotonous estimate $\hat{g}_I$(right) in Figure 1. From Table 1 and 2, we can see that the proposed procedure gives better results than the N-W estimate; for example, RASEs of $g(\cdot)$ and MSE of $\boldsymbol{\beta}$ are significantly reduced.

The residuals of the fitted partially linear model are $\hat{\epsilon} = Y - X^T\hat{\boldsymbol{\beta}} - \hat{g}_I(T)$. Let $\hat{F}_n$ be the centered empirical distribution function of $\hat{\epsilon}$. For any given $\boldsymbol{X}$ and T, independent samples $\epsilon_1^*, \cdots, \epsilon_n^*$ are drawn from $\hat{F}_n$, and let

$$Y^* = \mathbf{X}^T\hat{\boldsymbol{\beta}} + \hat{g}_I(T) + \epsilon^*.$$

Based on 1000 bootstrap replications, we construct 95% pointwise confidence intervals for the true nonparametric function. The confidence intervals for the monotonous estimates and N-W estimates are shown in Figure 2, together

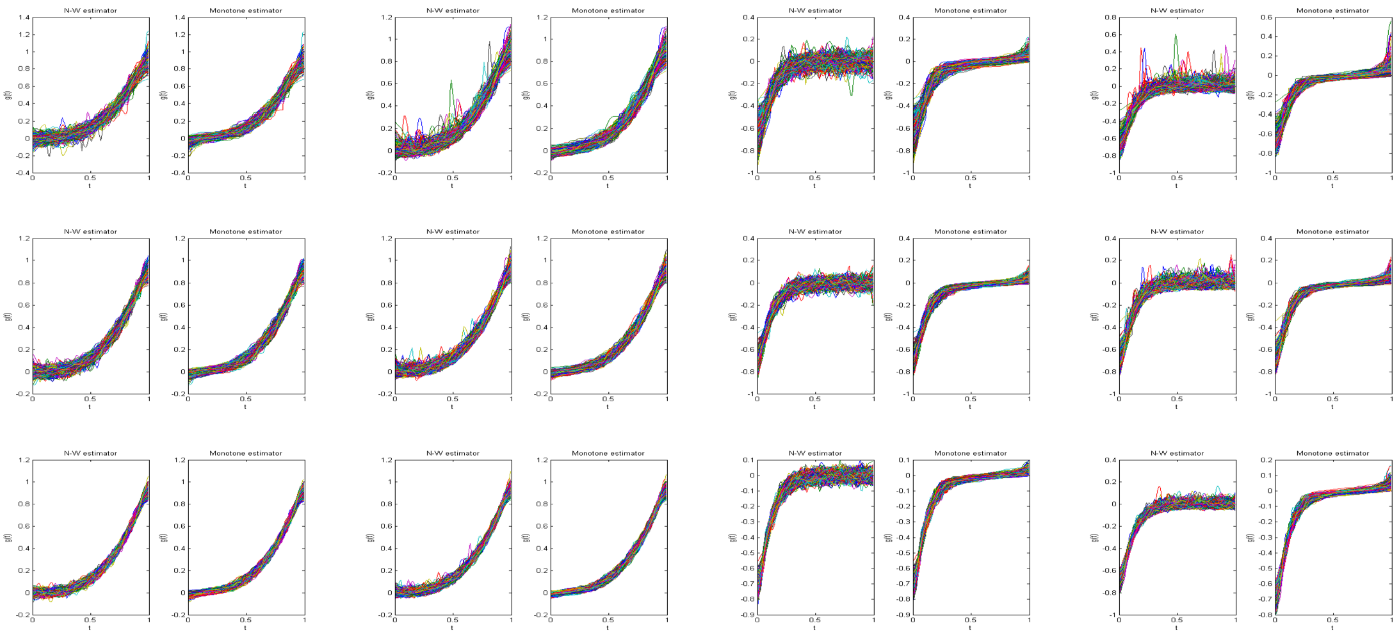

Figure 1. Comparison of estimates of $g(\cdot)$ with different function forms and error distributions. (N-W estimate: left, monotone estimate: right). $g(T) = T^3, N(0, 0.01)$ (column 1), $g(T) = T^3, 0.05\sqrt{2}(\chi_1^2 - 1)$ (column 2), $g(T) = -1/(0.2 + e^{10T}), N(0, 0.01)$ (column 3), $g(T) = -1/(0.2 + e^{10T}), 0.05\sqrt{2}(\chi_1^2 - 1)$ (column 4).

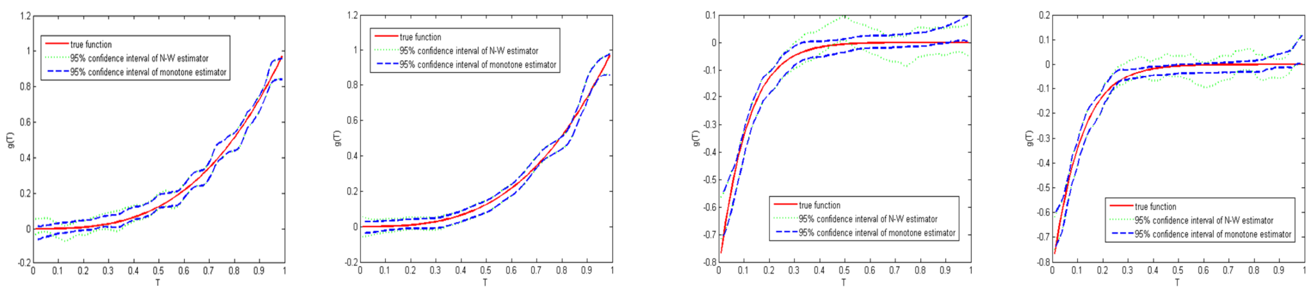

Figure 2. 95% pointwise confidence intervals of $g(\cdot)$ with different function forms and error distributions, sample size $n = 200$. (red lines: true functions; blue dashed lines: monotonous estimates; green dotted lines: N-W estiamtes). (a) $g(T) = T^3, N(0, 0.01)$, (b) $g(T) = T^3, 0.05\sqrt{2}(\chi_1^2 - 1)$, (c) $g(T) = -1/(0.2 + e^{10T}), N(0, 0.01)$, (d) $g(T) = -1/(0.2 + e^{10T}), 0.05\sqrt{2}(\chi_1^2 - 1)$.

with the true nonparametric function. From Figure 2 we can see that the true function lies more likely within the confidence intervals based on the monotonous estimation than based on the N-W estimation. The coverage probabilities (CP) and median lengths (ML) of the confidence intervals based on normal approximation with nominal confidence level 95% are reported in Table 3. These illustrations indicate the good performance of the proposed methodology in general.

4.2 Applications to real data

In this subsection, we apply the newly proposed procedure to mammalian eye gene expression data and German SOEP data.

I. Mammalian eye gene expression data

In a year 2006 study of the regulation of gene expression in the mammalian eye and its relevance to eye disease, [27] analyzed the microarrays containing more than 31,000 gene probes from 120 twelve-week-old male F_2 offspring of laboratory rats. The logarithm of the gene expression levels was used in the analysis. The dataset is obtainable from the Gene Expression Omnibus repository, www.ncbi.nlm.nih.gov/geo (accession no. GSE5680).

It has been shown that gene TRIM32, corresponding to probe 1389163_at, is associated with human disease Bardet-Biedl syndrome [9, 27]. Nonpositional analysis can help detect functionally related genes and identify new gene networks. It is also useful in deducing unknown biological

Table 3. The coverage probabilities(CP) and median lengths(ML) of confidence intervals for N-W estimate and monotone estimate

$\varepsilon, g(\cdot)$	n	$CP_{\hat{g}}$	$ML_{\hat{g}}$	$CP_{\hat{g}_I}$	$ML_{\hat{g}_I}$
$N(0, 0.01)$ $g(T) = T^3$	100	.94	.1132	.94	.0972
	200	.99	.0860	.99	.0752
	400	.94	.0615	.94	.0548
$N(0, 0.01)$ $g(T) = -1/(0.2 + e^{10T})$	100	.96	.1407	.98	.0768
	200	.91	.0921	.90	.0571
	400	1	.0745	1	.0396
$0.05\sqrt{2}(\chi_1^2 - 1)$ $g(T) = T^3$	100	.99	.0869	.99	.0771
	200	.96	.0697	.96	.0634
	400	.98	.0639	.99	.0588
$0.05\sqrt{2}(\chi_1^2 - 1)$ $g(T) = -1/(0.2 + e^{10T})$	100	.97	.1040	.98	.0644
	200	.97	.0890	.96	.0436
	400	.99	.0805	.98	.0386

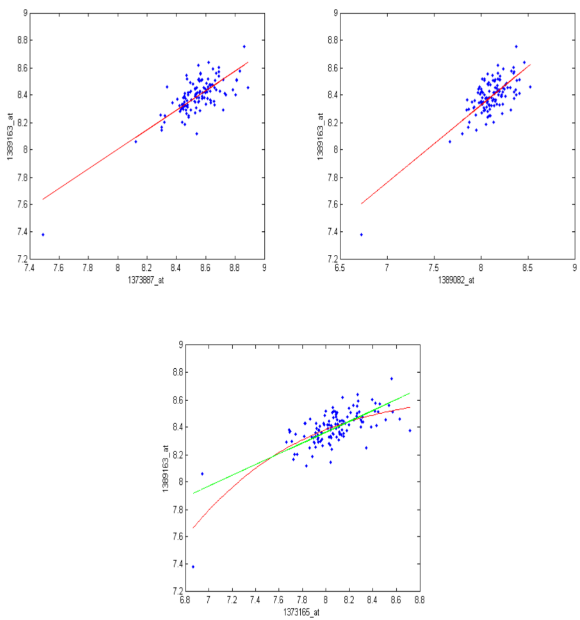

Figure 3. The scatter plots.

pathways. [24] performed an analysis on the above mammalian eye gene expression data using sure independence screen procedures and showed that the expression of gene TRIM32 is highly correlated with that of an unnamed gene (probe 1373887_at), gene Fam63b (probe 1389082_at) and gene Zmat1 (probe 1373165_at); see Figure 3. It can be seen that the unnamed gene (probe 1373887_at) and gene Fam63b are linearly correlated with gene TRIM32, while the relation between genes Zmat1 and TRIM32 appears to be a monotone nonlinear function.

In this paper we use partially linear model (1) to study the functional relation between gene TRIM32 and the unnamed gene (probe 1373887_at), Fam63b and Zmat1, with $\boldsymbol{X}$ being the 2-dimensional vector of the expression levels of the unnamed gene (probe 1373887_at) and gene Fam63b,

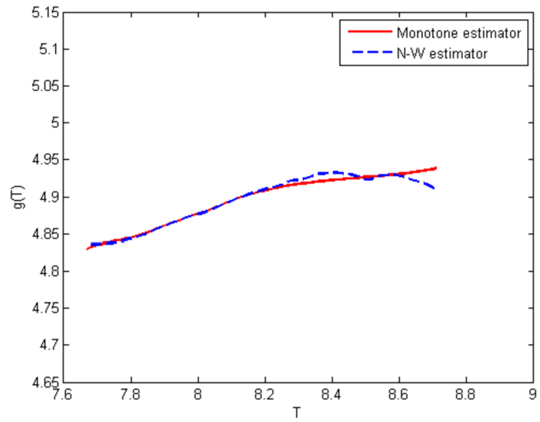

Figure 4. The monotonous estimator (red solid line) and the N-W estimator (blue dashed line).

and T the expression level of gene Zmat1. The nonparametric component $g(\cdot)$ is assumed to be a smooth monotone function. The range of variable T is $[6.8653, 8.7130]$. Two observations with the smallest T values (6.8653 and 6.9455) seem to be outliers and thus are excluded in the analysis for the purpose of kernel smoothing. The estimated monotonous curve $g(\cdot)$ is depicted in Figure 4, together with the N-W estimator.

From Figure 4 we can see that, after taking into account the expression levels of the unnamed gene and gene Fam63b, gene TRIM32 is monotonously correlated with gene Zmat1. The expression level of gene TRIM32 increases as the expression level of gene Zmat1 increases; The slope of the function depicting the relation between gene TRIM32 and gene Zmat1 is high for expression level T approximately being less than or equal to 8.2, and then diminishes a little after 8.2. This indicates that the interaction within a gene network can assume a complicated nonlinear form. Traditional linear analysis cannot detect this kind of nonlinear interaction [24] and fails in identification of potential important genes such as Zmat1 that play roles in regulation of mammalian eye disease.

II. German SOEP data

The German Socio-Economic Panel (SOEP) is a wide-ranging representative longitudinal study performed by the German Institute for Economic Research. The study collects data on household members living in Germany, including household composition, occupational biographies, employment, earnings, health and so on. The data are available at http://www.diw.de/en/soep. Based on the data of SOEP 1985, 1989 and 1993, [26] studied the relationship between the log-earnings of an individual and personal characteristics (gender, marital status) and measures of a person's human capital, such as time spent in school and labor market experience. Experience suggests that there is a nonlinear relationship between log earnings and labor market experience [26, 20]. As in [20], we fit to the data a partially linear

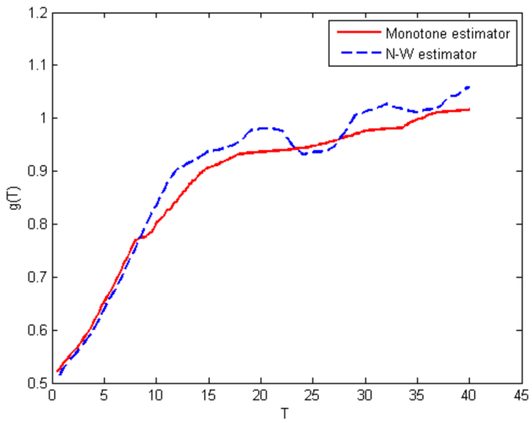

Figure 5. The monotonous estimator (red solid line) and the N-W estimator (blue dashed line).

model:

$$\ln(Y) = X\beta + g(T) + \varepsilon, \tag{6}$$

where Y is the earning of an individual, X the level of secondary school completed by the individual, and T is the number of years spent in the labor market. Human capital theory suggests that the function $g(\cdot)$ is an increasing function: human capital accumulation in the labor market is associated with rising earnings. By using the local kernel smoothing method, we first obtain the N-W estimate of function $g(\cdot)$, and then modify it to attain a monotonous estimate that is shown in Figure 5. This figure reveals that accumulation of human capital in the early stage of an individual's career results in a rapid increase of earnings, and the marginal effect of human capital on earnings peaks somewhere during mid-life and drops off thereafter. In other words, the nonparametric component $g(\cdot)$ is a concave function. This validates the law of diminishing marginal utility in the human capital theory.

5. DISCUSSION

In this article, we extended the monotone estimation method given by [10] to the partially linear models. Two kernel functions with different bandwidths are used in the estimation procedure. The asymptotic distributions of the proposed monotone function estimator vary depending on the asymptotic behavior of the ratio of the two bandwidths. The asymptotic biases are of the same order, while the asymptotic variances differ. Even when the true nonparametric component is a monotone function, traditional partial linear models do not necessarily produce a monotone estimate for the component. The proposed method can guarantee to give a nonparametric estimate for the nonparametric component, and hence is particularly useful in applications when monotonicity is desired in partial linear modeling. This procedure does not require constrained optimization and hence is fast to compute. By using kernel smoothing technique, our method requires enough data points around a point of interest. Estimation of the nonparametric component is no longer reliable in a region of sparse data points.

APPENDIX A

We outline the key steps for the proofs of Theorems 3.1 and 3.2.

A.1 Assumptions

The following six assumptions will be used.

Assumption 1 $\sup_{0\le t\le 1} E(\|\boldsymbol{X}_1\|^3|T = t) < \infty, \Sigma = cov(\boldsymbol{X}_1 - E(\boldsymbol{X}_1|T_1))$ *is a positive definite matrix. The random errors* ϵ_i *are independent of* $(\boldsymbol{X}_i, Y_i)$.

Assumption 2 *Let* $h_j(T_i) = E(X_{ij}|T_i)$ *and* $U_{ij} = X_{ij} - h_j(T_i)$ *for* $i = 1, \cdots, n,\ j = 1, \cdots, p$. *The first two derivatives of* $g(\cdot)$ *and* $h_j(\cdot)$ *are Lipschitz continuous of order one.*

Assumption 3 *the following equations hold uniformly over* $[0, 1]$ *and* $n \ge 1$:

$$(i) \sum_{i=1}^n |\omega_{ni}(t)| I(|t - T_i| > \mu_n) = O(h_r^2),$$

$$(ii) \sup_{i\le n} |\omega_{ni}(t)| = O((nh_r)^{-1}),$$

$$(iii) \sum_{i=1}^n \omega_{ni}^2(t) E\epsilon_i^2 = \frac{\sigma_0^2}{nh_r} + o\left(\frac{1}{nh_r}\right) \quad \text{for some } \sigma_0^2 > 0,$$

where bandwidth h_r *satisfying* $\lim_{n\to\infty} n^{-1/2}\log n/h_r = 0$ *and* $\limsup_{n\to\infty} n^2 h_r^4 < \infty$.

Assumption 4 *The weight functions* $\omega_{ni}(\cdot)$ *satisfy*

$$\max_{i\ge 1} |\omega_{ni}(s) - \omega_{ni}(t)| \le C_2|s - t|$$

uniformly over $n \ge 1$ *and* $s, t \in [0, 1]$, *where* C_2 *is a positive constant.*

Assumption 5 K_r *and* K_d *are symmetric kernels with compact support* $[-1, 1]$, *and finite second moments,* h_r *and* h_d *are the corresponding bandwidths converging to 0 with the sample size* n. K_d *is twice continuously differentiable on its support.*

Assumption 6 $nh_r^5 = O(1), n = O(N)$.

A.2 Proof of Theorem 3.1

Proof. Let

$$\Delta_n(t) = \hat{g}_I^{-1}(t) - g_n^{-1}(t) = \Delta_n^{(1)}(t) + \frac{1}{2}\Delta_n^{(2)}(t),$$

where

$$\Delta_n^{(1)}(t) = \frac{1}{nh_d^2} \sum_{i=1}^n \int_{-\infty}^t K_d'\left(\frac{g\left(\frac{i}{n}\right) - u}{h_d}\right) \left\{\hat{g}\left(\frac{i}{n}\right) - g\left(\frac{i}{n}\right)\right\} du$$

$$=\frac{-1}{nh_d}\sum_{i=1}^{n}K_d\left(\frac{g\left(\frac{i}{n}\right)-t}{h_d}\right)\left\{\hat{g}\left(\frac{i}{n}\right)-g\left(\frac{i}{n}\right)\right\},$$

$$\Delta_n^{(2)}(t)$$
$$=\frac{1}{nh_d^3}\sum_{i=1}^{n}\int_{-\infty}^{t}K_d''\left(\frac{\xi_i-u}{h_d}\right)\left\{\hat{g}\left(\frac{i}{n}\right)-g\left(\frac{i}{n}\right)\right\}^2du$$
$$=\frac{-1}{nh_d^2}\sum_{i=1}^{n}K_d'\left(\frac{\xi_i-t}{h_d}\right)\left\{\hat{g}\left(\frac{i}{n}\right)-g\left(\frac{i}{n}\right)\right\}^2$$

with $|\xi_i-g(i/n)|<|\hat{g}(i/n)-g(i/n)|, j=1,\ldots,n$. A straight calculation shows that

$$|\Delta_n^{(2)}(t)|$$
$$<\frac{1}{h_d^2}\left|\frac{1}{n}\sum_{i=1}^{n}K_d'\left(\frac{\xi_i-t}{h_d}\right)\left\{\hat{g}\left(\frac{i}{n}\right)-g\left(\frac{i}{n}\right)\right\}^2\right|$$
$$=\frac{1}{h_d^2}\left|\int_0^1K_d'\left(\frac{g(x)-t}{h_d}\right)\{\hat{g}(x)-g(x)\}^2dx\right|(1+o_p(1)).$$

If $\lim_{n\to\infty}\frac{h_r}{h_d}=a\in[0,\infty)$, $nh_d^7=o(1)$ and $\frac{(\log n)^2}{nh_r^2h_d}=o(1)$, then

$$\sqrt{nh_d}\Delta_n^{(2)}(t)$$
$$=O_p\left(\frac{\sqrt{nh_d}}{h_d}\left(h_r^4+\frac{\log n}{nh_r}+n^{-1}\right)\right)$$
$$=O_p\left(\sqrt{\frac{nh_r^8}{h_d}}+\sqrt{\frac{(\log n)^2}{nh_r^2h_d}}+\frac{1}{\sqrt{nh_d}}\right)=o_p(1).$$

If $\lim_{n\to\infty}\frac{h_r}{h_d}=\infty$, $\frac{nh_r^9}{h_d^2}=o(1)$ and $\frac{(\log n)^2}{nh_rh_d^2}=o(1)$, then

$$\sqrt{nh_r}\Delta_n^{(2)}(t)$$
$$=O_p\left(\frac{\sqrt{nh_r}}{h_d}\left(h_r^4+\frac{\log n}{nh_r}+n^{-1}\right)\right)$$
$$=O_p\left(\sqrt{\frac{nh_r^9}{h_d^2}}+\sqrt{\frac{(\log n)^2}{nh_rh_d^2}}+\sqrt{\frac{h_r}{nh_d^2}}\right)=o_p(1).$$

We split $\Delta_n^{(1)}(t)$ as follows.

$$\Delta_n^{(1)}(t)$$
$$=\frac{-1}{nh_d}\sum_{i=1}^{n}K_d\left(\frac{g\left(\frac{i}{n}\right)-t}{h_d}\right)\left(\hat{g}\left(\frac{i}{n}\right)-g\left(\frac{i}{n}\right)\right)$$
$$=\frac{-1}{n^2h_d}\sum_{i=1}^{n}K_d\left(\frac{g\left(\frac{i}{n}\right)-t}{h_d}\right)\left\{\sum_{j=1}^{n}\omega_{nj}\left(\frac{i}{n}\right)g(T_j)-g\left(\frac{i}{n}\right)\right\}$$
$$+\frac{-1}{n^2h_d}\sum_{i=1}^{n}K_d\left(\frac{g\left(\frac{i}{n}\right)-t}{h_d}\right)\sum_{j=1}^{n}\omega_{nj}\left(\frac{i}{n}\right)\epsilon_j$$
$$+\frac{-1}{n^2h_d}\sum_{i=1}^{n}K_d\left(\frac{g\left(\frac{i}{n}\right)-t}{h_d}\right)\sum_{j=1}^{n}\omega_{nj}\left(\frac{i}{n}\right)X_j^T(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}_{LS}).$$

Since $\hat{\boldsymbol{\beta}}_{LS}-\boldsymbol{\beta}=O_p(1)$ and $|\sum_{i=1}^n\omega_{ni}(t)x_{ij}|=O(1)$, the third term of above expression is $O_p(n^{-1})$. Denote $\hat{f}_T(t)=\frac{1}{nh_r}\sum_{i=1}^nK_r\left(\frac{T_i-t}{h_r}\right)$, then $\hat{f}_T(t)\xrightarrow{P}g_T(t)$, $\frac{1}{\hat{f}_T(t)}=\frac{1}{g_T(t)}(1+o_p(1))$. Therefore the first two terms of above expression can be rewritten as $\Delta_n^{(1,1)}(t)(1+o_p(1))$ and $\Delta_n^{(1,2)}(t)(1+o_p(1))$ respectively, where

$$\Delta_n^{(1,1)}(t)$$
$$=\frac{-1}{n^2h_dh_r}\sum_{i=1}^{n}K_d\left(\frac{g\left(\frac{i}{n}\right)-t}{h_d}\right)$$
$$\times\sum_{j=1}^{n}K_r\left(\frac{T_j-\frac{i}{n}}{h_r}\right)\frac{g(T_j)-g\left(\frac{i}{n}\right)}{f_T\left(\frac{i}{n}\right)}$$
$$\Delta_n^{(1,2)}(t)$$
$$=\frac{-1}{n^2h_dh_r}\sum_{i=1}^{n}K_d\left(\frac{g\left(\frac{i}{n}\right)-t}{h_d}\right)\sum_{j=1}^{n}K_r\left(\frac{T_j-\frac{i}{n}}{h_r}\right)\frac{\epsilon_j}{f_T\left(\frac{i}{n}\right)}.$$
$$E(\Delta_n^{(1,1)}(t))$$
$$=\frac{-1}{n^2h_dh_r}\sum_{i=1}^{n}K_d\left(\frac{g\left(\frac{i}{n}\right)-t}{h_d}\right)$$
$$\times\sum_{j=1}^{n}K_r\left(\frac{T_j-\frac{i}{n}}{h_r}\right)\frac{g(T_j)-g\left(\frac{i}{n}\right)}{f_T\left(\frac{i}{n}\right)}$$
$$=\frac{-(1+o(1))}{h_dh_r}\cdot$$
$$\iint K_d\left(\frac{g(u)-t}{h_d}\right)K_r\left(\frac{w-u}{h_r}\right)\frac{g(w)-g(u)}{f_T(u)}f_T(w)dwdu.$$

Let

$$\frac{w-u}{h_r}=p,\frac{g(u)-t}{h_d}=q.$$

Since $\frac{f_T(w)}{f_T(u)}=1+\frac{f_T'(u)}{f_T(u)}ph_r+o(ph_r)$ and $g(w)-g(u)=g'(u)ph_r+\frac{1}{2}g''(u)p^2h_r^2+o(p^2h_r^2)$,

$$E(\Delta_n^{(1,1)}(t))$$
$$=-(1+o(1))h_r^2\kappa_2(K_r)\int K_d(q)\frac{g''f_T+2f_T'g'}{f_Tg'}$$
$$\circ(g^{-1}(t+qh_d)dq$$
$$=-h_r^2\kappa_2(K_r)\frac{g''f_T+2f_T'g'}{f_Tg'}\circ(g^{-1}(t))(1+o(1)).$$
$$var(\Delta_n^{(1,1)}(t))$$
$$=\frac{1+o(1)}{nh_d^2h_r^2}var\left\{\int K_d\left(\frac{g(u)-t}{h_d}\right)K_r\left(\frac{T_1-u}{h_r}\right)\right.$$
$$\left.\times\frac{g(T_1)-g(u)}{f_T(u)}du\right\}$$
$$\leq\frac{1+o(1)}{nh_d^2h_r^2}E\left\{\int K_d\left(\frac{g(u)-t}{h_d}\right)K_r\left(\frac{T_1-u}{h_r}\right)\right.$$

$$\times \frac{g(T_1)-g(u)}{f_T(u)}du\Big\}^2$$

$$= \frac{1+o(1)}{nh_d^2h_r^2}\iiint K_d\left(\frac{g(u)-t}{h_d}\right)K_d\left(\frac{g(v)-t}{h_d}\right)$$

$$\times K_r\left(\frac{w-u}{h_r}\right)\cdot$$

$$K_r\left(\frac{w-v}{h_r}\right)\frac{(g(w)-g(u))(g(w)-g(v))f_T(w)}{f_T(u)f_T(v)}dudvdw.$$

Let

$$\frac{g(u)-t}{h_d}=p, \frac{w-u}{h_r}=q, \frac{w-v}{h_r}=s,$$

then $var(\Delta_n^{(1,1)}(t)) = o\left(\frac{1}{nh_d}\right)$. Further,

$$\Delta_n^{(1,1)}(t)+h_r^2\kappa_2(K_r)\frac{g''f_T+2f_T'g'}{f_Tg'}\circ(g^{-1}(t)) = o_p\left(\frac{1}{\sqrt{nh_d}}\right).$$

By simple calculation, $E(\Delta_n^{(1,2)}(t)) = 0$ and

$$var(\Delta_n^{(1,2)}(t))$$

$$=\frac{\sigma^2(1+o(1))}{nh_d^2h_r^2}\iiint K_d\left(\frac{g(u)-t}{h_d}\right)K_d\left(\frac{g(v)-t}{h_d}\right)\cdot$$

$$K_r\left(\frac{w-u}{h_r}\right)K_r\left(\frac{w-v}{h_r}\right)\frac{f_T(w)}{f_T(u)f_T(v)}dudvdw.$$

If $\lim_{n\to\infty}\frac{h_r}{h_d} = a\in[0,\infty)$, $nh_d^7 = o(1)$, $\frac{(\log n)^2}{nh_r^2h_d} = o(1)$ and g is strictly increasing. For all $t\in(g(0),g(1))$ with $g'(g^{-1})>0$, let

$$\frac{g(u)-t}{h_d}=p, \frac{w-u}{h_r}=q, \frac{w-v}{h_r}=s,$$

then

$$var(\Delta_n^{(1,2)}(t))$$

$$=\frac{\sigma^2(1+o(1))}{nh_dg'f_T}\circ(g^{-1}(t))\cdot$$

$$\iiint K_d\left(p+\frac{h_r}{h_d}g'(g^{-1}(t))(q-s)\right)K_d(p)K_r(q)K_r(s)dpdqds.$$

Therefore,

$$\sqrt{nh_d}\Delta_n^{(1,2)}(t)\xrightarrow{d}\mathcal{N}(0,r^2(t)).$$

If $\lim_{n\to\infty}\frac{h_r}{h_d}=\infty$, $\frac{nh_r^9}{h_d^2}=o(1)$, $\frac{(\log n)^2}{nh_rh_d^2}=o(1)$ and g is strictly increasing. For all $t\in(g(0),g(1))$ with $g'(g^{-1})>0$, let

$$\frac{g(u)-t}{h_d}=p, \frac{g(v)-t}{h_d}=q, \frac{w-u}{h_r}=s,$$

then $var(\Delta_n^{(1,2)}(t)) = \frac{\sigma^2(1+o(1))}{nh_rg'^2f_T}\circ(g^{-1}(t))\int K_r^2(s)ds$. Therefore,

$$\sqrt{nh_r}\Delta_n^{(1,2)}(t)\xrightarrow{d}\mathcal{N}(0,\tilde{r}^2(t)).$$

If $\lim_{n\to\infty}\frac{h_r}{h_d}=a\in[0,\infty)$, then

$$\sum_{j=1}^n E\Bigg\{\frac{\epsilon_j}{n^{3/2}h_d^{1/2}h_r}\sum_{i=1}^n K_d\left(\frac{g\left(\frac{i}{n}\right)-t}{h_d}\right)$$

$$\times K_r\left(\frac{T_j-\frac{i}{n}}{h_r}\right)\frac{1}{f_T\left(\frac{i}{n}\right)}\Bigg\}^4$$

$$=O\left(\frac{1}{nh_d}\right)=o(1),$$

where we make a change of variables

$$\frac{g(u_1)-t}{h_d}=\tilde{w},\quad \frac{w-u_p}{h_r}=s_p,\quad p=1,2,3,4.$$

If $\lim_{n\to\infty}\frac{h_r}{h_d}=\infty$, then

$$\sum_{j=1}^n E\Bigg\{\frac{\epsilon_j}{n^{3/2}h_dh_r^{1/2}}\sum_{i=1}^n K_d\left(\frac{g\left(\frac{i}{n}\right)-t}{h_d}\right)$$

$$\times K_r\left(\frac{T_j-\frac{i}{n}}{h_r}\right)\frac{1}{f_T\left(\frac{i}{n}\right)}\Bigg\}^4$$

$$=O\left(\frac{1}{nh_r}\right)=o(1),$$

Where we make a change of variables

$$\frac{g(u_p)-t}{h_d}=\tilde{w}_p,\ p=1,2,3,4.\quad \frac{w-u_1}{h_r}=s.$$

□

A.3 Proof of Theorem 3.2

Proof. Let

$$A_n = -\frac{\hat{g}_I^{-1}-g_n^{-1}}{(g_n^{-1})'}\circ g_n(t)$$

and

$$B_n = 2\left\{\hat{g}_I(t)-g_n(t)-A_n\right\}.$$

From Lemma A.1 of [10], B_n can be rewritten as $2B_{n1}-B_{n2}$, where

$$B_{n1} =\frac{(\hat{g}_I^{-1}-g_n^{-1})(\hat{g}_I^{-1}-g_n^{-1})'(t_n)}{\{(g_{nc}^{-1}+\lambda^*(\hat{g}_I^{-1}-g_n^{-1}))'\}^2(t_n)},$$

$$B_{n2} =\frac{(\hat{g}_I^{-1}-g_n^{-1})^2(g_n^{-1}+\lambda^*(\hat{g}_I^{-1}-g_n^{-1}))''(t_n)}{\{(\hat{g}_I^{-1}+\lambda^*(\hat{g}_I^{-1}-g_n^{-1}))'\}^3(t_n)},$$

$$t_n =(g_n^{-1}+\lambda^*(\hat{g}_I^{-1}-g_n^{-1}))^{-1}(t).$$

for some $\lambda*\in[0,1]$.

Let

$$D_n =(\hat{g}_I^{-1}-g_n^{-1})\circ g_n(t)-(\hat{g}_I^{-1}-g_n^{-1})\circ g(t)$$

$$=(\hat{g}_I^{-1}-g_n^{-1})'(\xi_n)(g_n(t)-g(t)),$$

where $|\xi_n-g(t)|\le|g_n(t)-g(t)|$.

$$(\hat{g}_I^{-1}-g_n^{-1})'(\xi_n)$$

$$=\frac{1}{nh_d}\sum_{i=1}^n\left\{K_d\left(\frac{\hat g(i/n)-\xi_n}{h_d}\right)-K_d\left(\frac{g(i/n)-\xi_n}{h_d}\right)\right\}$$
$$=-\frac{1}{nh_d^2}\sum_{i=1}^n K_d'\left(\frac{\eta_{i,n}-\xi_n}{h_d}\right)\left\{\hat g\left(\frac{i}{n}\right)-g\left(\frac{i}{n}\right)\right\},$$

where $|\eta_{i,n}-g(i/n)|\le|\hat g(i/n)-g(i/n)|=O(R_n)$ almost surely, with

$$R_n=h_r^2+(\log(n)/nh_r)^{1/2}.$$

Further, we have

$$(\hat g_I^{-1}-g_n^{-1})'(\xi_n)=O\left(\frac{R_n}{h_d}+\frac{R_n^2}{h_d^3}+\frac{1}{nh_d}\right)\quad a.s.$$

Because $\frac{h_r^4}{h_d^3}=o(1)$ and $\frac{\log n}{nh_rh_d^3}=o(1)$, we have $\frac{R_n^2}{h_d^3}=o(1)$. If $\lim_{n\to\infty}\frac{h_r}{h_d}=a\in[0,\infty)$, $nh_d^7=o(1)$ and $\frac{(\log n)^2}{nh_r^2h_d}=o(1)$, then

$$\sqrt{nh_d}D_n=O(\sqrt{nh_d^3}R_n+\sqrt{n/h_d}R_n^2+\sqrt{h_d^3/n})\quad a.s.$$
$$=o(1)\quad a.s.$$

By the fact that $(g_n^{-1})'(t)=(g^{-1})'(t)+o(1)$, we obtain

$$A_n=-\frac{\hat g_I^{-1}-g_n^{-1}}{(g^{-1})'}\circ g(t)+o_p\left(\frac{1}{\sqrt{nh_d}}\right).$$

Note that $t_n\xrightarrow{P}g(t)$, $(g_n^{-1}+\lambda^*(\hat g_I^{-1}-g_n^{-1}))\xrightarrow{P}g^{-1}$, we obtain

$$B_{n1}=O_p\left(\frac{1}{\sqrt{nh_d}}\left(\frac{R_n}{h_d}+\frac{R_n^2}{h_d^3}\right)\right)=o_p\left(\frac{1}{\sqrt{nh_d}}\right),$$
$$B_{n2}=O_p\left(\frac{1}{nh_d}\right).$$

As a consequence, we obtain

$$\hat g_I(t)-g_n(t)=-\frac{\hat g_I^{-1}-g_n^{-1}}{(g^{-1})'}\circ g(t)+o_p\left(\frac{1}{\sqrt{nh_d}}\right),$$

and

$$\sqrt{nh_d}\left\{\hat g_I(t)-g_n(t)-\mathcal{D}(t)\right\}$$
$$=-\sqrt{nh_d}\left(\frac{\hat g_I^{-1}-g_n^{-1}}{(g^{-1})'}\circ g(t)+\mathcal{D}(t)\right)+o_p(1)$$
$$=-g'(t)\sqrt{nh_d}\left\{(\hat g_I^{-1}-g_n^{-1})\circ g(t)+\mathcal{D}(t)\right\}+o_p(1)$$
$$\overset{d}{\Rightarrow}\mathcal{N}(0,s^2(t)).$$

The second assertion follows by exactly the same arguments. If $\lim_{n\to\infty}\frac{h_r}{h_d}=\infty$, $\frac{h_r^4}{h_d^3}=o(1)$ and $\frac{\log n}{nh_rh_d^3}=o(1)$, then $\frac{nh_r^9}{h_d^2}=o(1)$, $\frac{(\log n)^2}{nh_rh_d^2}=o(1)$ and

$$\sqrt{nh_r}D_n$$
$$=O\left(\sqrt{nh_rh_d^2}R_n+\sqrt{nh_r/h_d^2}R_n^2+\sqrt{h_rh_d^2/n}\right)\quad a.s.$$
$$=o(1)\quad a.s.$$

By the fact that $(g_n^{-1})'(t)=(g^{-1})'(t)+o(1)$, we obtain

$$A_n=-\frac{\hat g_I^{-1}-g_n^{-1}}{(g^{-1})'}\circ g(t)+o_p\left(\frac{1}{\sqrt{nh_r}}\right).$$

Note that $t_n\xrightarrow{P}g(t)$, $(g_n^{-1}+\lambda^*(\hat g_I^{-1}-g_n^{-1}))\xrightarrow{P}g^{-1}$, we obtain

$$B_{n1}=O_p\left(\frac{1}{\sqrt{nh_r}}\left(\frac{R_n}{h_d}+\frac{R_n^2}{h_d^3}\right)\right)=o_p\left(\frac{1}{\sqrt{nh_r}}\right),$$
$$B_{n2}=O_p\left(\frac{1}{nh_r}\right).$$

As a consequence, we obtain

$$\hat g_I(t)-g_n(t)=-\frac{\hat g_I^{-1}-g_n^{-1}}{(g^{-1})'}\circ g(t)+o_p\left(\frac{1}{\sqrt{nh_r}}\right).$$

and

$$\sqrt{nh_r}\left\{\hat g_I(t)-g_n(t)-\mathcal{D}(t)\right\}$$
$$=-\sqrt{nh_r}\left(\frac{\hat g_I^{-1}-g_n^{-1}}{(g^{-1})'}\circ g(t)+\mathcal{D}(t)\right)+o_p(1)$$
$$=-\sqrt{nh_r}\left(\frac{(\hat g_I^{-1}-g_n^{-1})\circ g(t)+\mathcal{D}(t)(g^{-1})'\circ g(t)}{(g^{-1})'\circ g(t)}\right)+o_p(1)$$
$$=-g'(t)\sqrt{nh_r}\left\{(\hat g_I^{-1}-g_n^{-1})\circ g(t)+\mathcal{D}(t)\right\}+o_p(1)$$
$$\overset{d}{\Rightarrow}\mathcal{N}(0,\tilde s^2(t)).$$

□

ACKNOWLEDGEMENTS

The authors thank the editor, associate editor and referees for their helpful comments. We are grateful to Professor Hua Liang and Xueqin Wang for providing us with the real data. Zhang's research is partially supported by Graduate Innovation Foundation of Shanghai University of Finance and Economics, China (CXJJ-2014-460). Wang's research is partially supported by NSFC grant (11371235) and by Program for Innovative Research Team of Shanghai University of Finance and Economics.

Received 17 August 2016

REFERENCES

[1] Bardet, G. (1920). *Sur un syndrome d'obésité congenitale avec polydactylie et rétinite pigmentaire: Contribution a l'etude des formes cliniques de l'obésité hypophysaire*. PhD thesis.

[2] Biedl, A. (1922). Ein geschwisterpaar mit adiposo-genitaler dystrophie. *Dtsch Med Wochenschr*, 48:1630.

[3] Brunk, H. D. (1955). Maximum likelihood estimates of monotone parameters. *The Annals of Mathematical Statistics*, pages 607–616. MR0073894

[4] Carroll, R. J., Fan, J., Gijbels, I., and Wand, M. P. (1997). Generalized partially linear single-index models. *Journal of the American Statistical Association*, 92(438):477–489. MR1467842

[5] Chang, X.-W. and Qu, L. (2004). Wavelet estimation of partially linear models. *Computational statistics & data analysis*, 47(1):31–48. MR2087929

[6] Chen, H. et al. (1988). Convergence rates for parametric components in a partly linear model. *The Annals of Statistics*, 16(1):136–146. MR0924861

[7] Chen, H. and Shiau, J.-J. H. (1991). A two-stage spline smoothing method for partially linear models. *Journal of Statistical Planning and Inference*, 27(2):187–201. MR1096678

[8] Cheng, K.-F. and Lin, P.-E. (1981). Nonparametric estimation of a regression function. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 57(2):223–233. MR0626817

[9] Chiang, A. P., Beck, J. S., Yen, H.-J., Tayeh, M. K., Scheetz, T. E., Swiderski, R. E., Nishimura, D. Y., Braun, T. A., Kim, K.-Y. A., Huang, J., et al. (2006). Homozygosity mapping with snp arrays identifies trim32, an e3 ubiquitin ligase, as a bardet–biedl syndrome gene (bbs11). *Proceedings of the National Academy of Sciences*, 103(16):6287–6292.

[10] Dette, H., Neumeyer, N., Pilz, K. F., et al. (2006). A simple nonparametric estimator of a strictly monotone regression function. *Bernoulli*, 12(3):469–490. MR2232727

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

[12] Friedman, J. and Tibshirani, R. (1984). The monotone smoothing of scatterplots. *Technometrics*, 26(3):243–250.

[13] Gannaz, I. (2007). Robust estimation and wavelet thresholding in partially linear models. *Statistics and Computing*, 17(4):293–310. MR2409795

[14] Green, J. S., Parfrey, P. S., Harnett, J. D., Farid, N. R., Cramer, B. C., Johnson, G., Heath, O., McManamon, P. J., O'Leary, E., and Pryse-Phillips, W. (1989). The cardinal manifestations of bardet–biedl syndrome, a form of laurence–moon–biedl syndrome. *New England Journal of Medicine*, 321(15):1002–1009.

[15] Green, P., Jennison, C., and Seheult, A. (1985). Analysis of field experiments by least squares smoothing. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 299–315. MR0816095

[16] Hall, P. and Huang, L.-S. (2001). Nonparametric kernel regression subject to monotonicity constraints. *Annals of Statistics*, pages 624–647. MR1865334

[17] Hamilton, S. A. and Truong, Y. K. (1997). Local linear estimation in partly linear models. *Journal of Multivariate Analysis*, 60(1):1–19. MR1441456

[18] Härdle, W. and Liang, H. (2007). *Partially linear models*. Springer. MR2376405

[19] Kelly, C. and Rice, J. (1990). Monotone smoothing with application to dose-response curves and the assessment of synergism. *Biometrics*, pages 1071–1085.

[20] Liang, H. (2006). Estimation in partially linear models and numerical comparisons. *Computational Statistics & Data Analysis*, 50(3):675–687. MR2207001

[21] Mammen, E. (1991). Estimating a smooth monotone regression function. *The Annals of Statistics*, pages 724–740. MR1105841

[22] Mammen, E. and Thomas-Agnan, C. (1999). Smoothing splines and shape restrictions. *Scandinavian Journal of Statistics*, 26(2):239–252. MR1707587

[23] Mukerjee, H. (1988). Monotone nonparametric regression. *The Annals of Statistics*, pages 741–750. MR0947574

[24] Pan, W., Wang, X., Xiao, W., and Zhu, H. (2016). A generic sure independence screening procedure. *Unpublished manuscript*.

[25] Ramsay, J. (1998). Estimating smooth monotone functions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 60(2):365–375. MR1616049

[26] Rendtel, U. and Schwarze, J. (1995). Zum zusammenhang zwischen lohnhöhe und arbeitslosigkeit: Neue befunde auf basis semi-parametrischer schätzungen und eines verallgemeinerten varianz-komponenten-modells. *in* V. Steiner & L. Bellmann (Hrsg.), *Mikroökonomik des Arbeitsmarktes, Beiträge zur Arbeitsmarkt-und Berufsforschung,* Nürnberg, 192:115–138.

[27] Scheetz, T. E., Kim, K.-Y. A., Swiderski, R. E., Philp, A. R., Braun, T. A., Knudtson, K. L., Dorrance, A. M., DiBona, G. F., Huang, J., Casavant, T. L., Sheffield, V. C., and Stone, E. M. (2006). Regulation of gene expression in the mammalian eye and its relevance to eye disease. *Proceedings of the National Academy of Sciences*, 103(39):14429–14434.

[28] Schmalensee, R. and Stoker, T. M. (1999). Household gasoline demand in the united states. *Econometrica*, 67(3):645–662.

[29] Severini, T. A. and Wong, W. H. (1992). Profile likelihood and conditionally parametric models. *The Annals of statistics*, pages 1768–1802. MR1193312

[30] Shi, J. and Lau, T.-S. (2000). Empirical likelihood for partially linear models. *Journal of Multivariate Analysis*, 72(1):132–148. MR1747427

[31] Shiau, J.-J., Wahba, G., and Johnson, D. R. (1986). Partial spline models for the inclusion of tropopause and frontal boundary information in otherwise smooth two-and three-dimensional objective analysis. *Journal of Atmospheric and Oceanic Technology*, 3(4):714–725.

[32] Speckman, P. (1988). Kernel smoothing in partial linear models. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 413–436. MR0970977

[33] Wahba, G. (1983). *Cross validated spline methods for the estimation of multivariate functions from data on functionals*. University of Wisconsin, Department of Statistics.

[34] Wahba, G. (1984). Partial spline models for the semiparametric estimation of functions of several variables. In *Statistical analysis of time series*, volume 319, page 329. Institute of Statistical Mathematics Tokyo.

[35] Wright, F. et al. (1981). The asymptotic behavior of monotone regression estimates. *The Annals of Statistics*, 9(2):443–448. MR0606630

Yi Zhang
School of Statistics and Management
Shanghai University of Finance and Economics
Shanghai 200433
P.R. China
School of Insurance
Shanghai Lixin University of Accounting and Finance
Shanghai 201209
P.R. China
E-mail address: zhangyi034@live.cn

Shaoli Wang
School of Statistics and Management
Shanghai Key Laboratory of Financial Information Technology
Shanghai University of Finance and Economics
Shanghai 200433
P.R. China
E-mail address: swang@shufe.edu.cn