Communications in Statistics—Simulation and Computation®, 39: 1953–1961, 2010

ISSN: 0361-0918 print/1532-4141 online
DOI: 10.1080/03610918.2010.526735

Smoothing Spline Estimation for Partially Linear Single-index Models

BARTHOLOMEW LEUNG[1], HEUNG WONG[1], RIQUAN ZHANG[2,3] AND SHENGTONG HAN[2]

[1]Department of Applied Mathematics, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong
[2]Department of Statistics and Actuarial Science East China Normal University, Shanghai, P.R. China
[3]Department of Mathematics, Shanxi Datong University, Datong, P.R. China

In this article, the partially linear single-index models are discussed based on smoothing spline and average derivative estimation method. This proposed technique consists of two stages: one is to estimate the vector parameter in the linear part using the smoothing cubic spline method, simultaneously, obtaining the estimator of unknown single-index function; the other is to estimate the single-index coefficients in the single-index part by the using average derivative estimator procedure. Some simulated and real examples are presented to illustrate the performance of this method.

Keywords Average derivative estimate; Partially linear single-index models; Smoothing spline.

Mathematics Subject Classification Primary 62F10; Secondary 62G05.

1. Introduction

The single-index model is an efficient tool in multivariate nonparametric regression, which has the form

$$Y = \eta_0(\alpha_0^T X) + \varepsilon, \tag{1}$$

where $\eta_0(\cdot)$ is an unknown univariable measurable function, $\alpha_0 \in R^q$ is an unknown parametric vector with $\|\alpha_0\| = 1$ for identifiability, $X \in R^q$, $Y \in R$ and the error ε is independent of X with $E(\varepsilon) = 0$ and $\mathrm{Var}(\varepsilon) = \sigma^2$. There are a number of articles for the model (1), such as Härdle and Stoker (1989), Härdle et al. (1993), Li (1991),

Received February 1, 2010; Accepted September 13, 2010
Address correspondence to Riquan Zhang, Department of Statistics and Actuarial Science, East China Normal University, Shanghai 200241, P.R. China; E-mail: zhangriquan@163.com

Ichimura (1993), and so on. The appeal of the model (1) is that by reducing the dimensionality from multivariate predictors to an index $\alpha_0^T X$, the so-called "curse of dimensionality" is avoided and the important features are still captured in high-dimensional data.

To capture more accurately the underlying relationship between the response variable and the covariates, Carroll et al. (1997) proposed the partially linear single-index model (PLSIM) by combining the single-index and linear regression model, i.e.,

$$Y = \eta_0(\alpha_0^T X) + \boldsymbol{\beta}_0^T Z + \varepsilon, \tag{2}$$

where $Z \in R^p$ is covariate; the unknown parameter $\boldsymbol{\beta}_0$ is in R^p; the error ε is independent of X and Z with $E(\varepsilon) = 0$ and $\mathrm{Var}(\varepsilon) = \sigma^2$. Other quantities are defined as in the model (1). Carroll et al. (1997) employed the local linear method to obtain the quasi-likelihood estimators of the unknown parameters and the unknown function in model (2) and gave their asymptotic distributions. More discussions can be found in Yu and Ruppert (2002) and Xia and Härdle (2006). These articles use local polynomial or penalized spline method.

In this article, we use smoothing spline to estimate the model (2) by combining average derivative method. We first apply smoothing cubic spline method to obtain the estimate of unknown smooth function $\eta_0(\cdot)$ and get the estimate of the vector parameter β_0 simultaneously under given index parameter, then, use average derivative estimator (ADE) technique by using the estimated function value obtained previously to establish the estimate of the parameter vector α_0 in the single-index part. The estimation procedure is an iteration procedure.

The rest of the article is organized as follows. We give detailed estimation procedure in Sec. 2, where details regarding to choosing smoothing parameters is included as well. An optimal algorithm for these estimators is presented in Sec. 3. Sec. 4 presents some simulated examples to reveal the performance of this estimation procedure and the real data analysis is also included in this section. Some concluding remarks are left in Sec. 5.

2. Estimation

2.1. *Estimation of $\boldsymbol{\beta}_0$ and $\boldsymbol{\eta}_0(\cdot)$*

For any given α_0, write $U = \alpha_0^T X$, then the model (2) becomes

$$Y = \eta_0(U) + \beta_0^T Z + \varepsilon. \tag{3}$$

Model (3) is the partially linear model. We adopt a cubic spline and the penalized least squares technique to estimate the function $\eta_0(\cdot)$ and parameter β_0. Suppose that $\eta_0(\cdot)$ is a cubic smoothing spline and has second derivative at every knot. In our setting, all available data points of the covariate U_i are taken as the knots. We can estimate the model (3) by minimizing

$$S(\eta, \beta) = \sum_{i=1}^{n}\{Y_i - \eta(U_i) - \beta^T Z_i\}^2 + \lambda \int [\eta''(u)]^2\, \mathrm{d}u. \tag{4}$$

The idea of the above representation aries naturally, because it takes the goodness of fit into account in the first term, and at the same time, the smoothness of resulting curve also is included in the second part. The solution of (4) fairly balances the two criterions and can be regarded as the compromise of the two considerations. For computational simplicity, (4) can also be expressed as

$$S(\eta, \beta) = (\mathbf{Y} - \mathbf{Z}\beta - \eta)^T(\mathbf{Y} - \mathbf{Z}\beta - \eta) + \lambda \int [\eta''(u)]^2 \, \mathrm{d}u, \tag{5}$$

where $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_n)^T$, $\mathbf{Z} = (Z_1, Z_2, \ldots, Z_n)^T$, $\eta = (\eta_1, \eta_2, \ldots, \eta_n)^T$, and $\eta_i = \eta(U_i)$. Firstly, we consider the second part of (5). As pointed out by Green and Silverman (1994), the second term can be represented as $\eta^T K \eta$. Here, $K = QR^{-1}Q^T$. Q, R are two band matrices which are defined in the following.

Suppose that these ascending U_i, $i = 1, 2, \ldots, n$, are distinct knots, define $h_i = U_{i+1} - U_i$, for $i = 1, \ldots, n-1$. Q is defined as a $n \times (n-2)$matrix with elements q_{ij} given by

$$q_{j-1,j} = h_{j-1}^{-1}, \; q_{j,j} = -h_{j-1}^{-1} - h_j^{-1}, \quad \text{and} \quad q_{j+1,j} = h_j^{-1}$$

for $j = 2, \ldots, n-1$, and $q_{ij} = 0$ for $|i - j| \geq 2$, for $i = 1, \ldots, n$ and $j = 2, \ldots, n-1$. R is defined as a $(n-2) \times (n-2)$ symmetric matrix, the columns of which start from 2 and so the top left element of it is r_{22}. Its elements have the forms of

$$r_{ii} = \frac{1}{3}(h_{i-1} + h_i) \quad \text{for } i = 2, \ldots, n-1,$$

$$r_{i,i+1} = r_{i+1,i} = \frac{1}{6}h_i \quad \text{for } i = 2, \ldots, n-2,$$

and $r_{i,j} = 0$ for $|i - j| \geq 2$.

Then we have

$$S(\eta, \beta) = (\mathbf{Y} - \mathbf{Z}\beta - \eta)^T(\mathbf{Y} - \mathbf{Z}\beta - \eta) + \lambda\eta^T K \eta. \tag{6}$$

By deriving with respect to η and β, we get

$$\begin{pmatrix} \mathbf{Z}^T \\ I \end{pmatrix} \mathbf{Y} = \begin{pmatrix} \mathbf{Z}^T\mathbf{Z} & \mathbf{Z}^T \\ \mathbf{Z} & I + \lambda K \end{pmatrix} \begin{pmatrix} \beta \\ \eta \end{pmatrix}. \tag{7}$$

Solve the linear matrix equation, then we have the estimation $\widehat{\beta}_0$ and $\widehat{\eta}_0$ of β_0 and η_0 at every knot, respectively. What we want to do next is to employ ADE technique to estimate parameter α_0.

2.2. *Estimation of* α_0

Although the estimation of function $\eta_0(\cdot)$ at every knot and parameter β_0 are obtained through the method described above under given the parameter α_0. Next, we are going to employ the average derivative estimator (ADE) approach to get the estimator of α_0. Replace β_0 by its estimator $\hat{\beta}_0$, and write $Y_{new} = Y - \widehat{\beta}_0^T Z$, then model (2) can be rewrote asymptotically

$$Y_{new} \doteq \eta_0(\alpha_0^T X) + \varepsilon. \tag{8}$$

Model (8) is considered as the single-index model and our next task is to get the estimate $\hat{\alpha}_0$ of α_0 by use of average derivative estimate method.

Average derivative estimate method (ADE) was introduced in Stoker (1986) and Powell et al. (1989). The idea of this method is to estimate the expected value of the gradient $(\partial/\partial x)\eta_0(\alpha_0^T x) = \alpha_0 \eta_0'(\alpha_0^T x)$ of the regression function $\eta_0(\cdot)$, which is proportional to α_0 (Hristache et al., 2001). So we only need to know the direction of function $\nabla\eta_0(\cdot)$ in order to estimate the parameter α_0. This method leads to the $\sqrt{n}$-consistent estimator of the index vector. The idea above naturally leads to the following representation:

$$\alpha_0^* := 1/n \sum_{j=1}^{n} \nabla\eta_0(\alpha_0^T X_j). \tag{9}$$

The main problem which arises when implementing this approach is that the gradient function is not smooth and some rather restrictive assumptions on the design and on the link function η_0 must hold to make sure the desire $\sqrt{n}$-consistency of the corresponding estimator (Härdle and Tsybakov, 1993; Samarov, 1993). So the estimator $\hat{\alpha}_0$ of α_0 can be estimated naturally using the following expression:

$$\hat{\alpha}_0^* = 1/n \sum_{j=1}^{n} \widehat{\nabla\eta_0}(\alpha_0^T X_j) \quad \text{and} \quad \hat{\alpha}_0 = \frac{\hat{\alpha}_0^*}{|\hat{\alpha}_0^*|}. \tag{10}$$

Here, the $\widehat{\nabla\eta_0}(\alpha_0^T X_j)$ is a consistent estimator of the gradient $\nabla\eta_0(\alpha_0^T X_j)$. We introduce the estimation procedure given by Hristache et al. (2001). The local linear method is employed to estimate $\nabla\eta_0(\cdot)$. Suppose for a moment that we know α_0 and estimate $\nabla\eta_0(\alpha_0^T X_j)$. By using the local least algorithm

$$\begin{pmatrix} \widehat{\eta_0}(\alpha_0^T X_j) \\ \widehat{\nabla\eta_0}(\alpha_0^T X_j) \end{pmatrix} = \arg \inf_{c\in R, \theta\in R^d} \sum_{i=1}^{n} [Y_{new,i} - c - \theta^T (X_i - X_j)]^2 K\left(\frac{|\alpha_0^T (X_i - X_j)|^2}{\rho^2} \right), \tag{11}$$

where $|\cdot|$ is the Euclidean norm, ρ is a small positive value, $K(\cdot)$ is a kernel function and supported on $[-1, 1]$. The solution to the problem above is represented as

$$\begin{pmatrix} \widehat{\eta_0}(\alpha_0^T X_j) \\ \widehat{\nabla\eta_0}(\alpha_0^T X_j) \end{pmatrix} = \left\{ \sum_{i=1}^{n} \begin{pmatrix} 1 \\ X_{ij} \end{pmatrix} \begin{pmatrix} 1 \\ X_{ij} \end{pmatrix}^T K\left(\frac{|\alpha_0^T X_{ij}|^2}{\rho^2} \right) \right\}^{-1} \sum_{i=1}^{n} Y_{new,i} \begin{pmatrix} 1 \\ X_{ij} \end{pmatrix} K\left(\frac{|\alpha_0^T X_{ij}|^2}{\rho^2} \right). \tag{12}$$

Here, $X_{ij} = X_i - X_j$. Then the estimator of $\nabla\eta_0(\alpha_0^T X_j)$ can obtained. Since the parameter α_0 is unknown, the algorithm described above involves iterative procedure. As pointed by Hristache et al. (2001), the algorithm reads as follows.

Step 1. Initialization: specify parameters $h_1 = n^{-1/(4\vee q)}$, $h_{max} = 1$, $a_h = e^{1/2(4\vee q)}$, $\rho_1 = 1$, $\rho_{min} = n^{-1/3}$, $a_\rho = e^{-1/6}$, $k = 1$, $\hat{\theta}_0^{(0)} = 0$.

Step 2. Compute $S_k = (I + \rho_k^{-2} \widehat{\theta}_0^{(k-1)} (\widehat{\theta}_0^{(k-1)})^T)^{1/2}$.

Step 3. For every $j = 1, \ldots, n$, compute $\widehat{\nabla\eta_0}((\theta_0^{(k)})^T X_j)$ from the following expression $\begin{pmatrix} \widehat{\eta_0}((\theta_0^{(k)})^T X_j) \\ \widehat{\nabla\eta_0}((\theta_0^{(k)})^T X_j) \end{pmatrix} = \left\{ \sum_{i=1}^{n} \begin{pmatrix} 1 \\ X_{ij} \end{pmatrix} \begin{pmatrix} 1 \\ X_{ij} \end{pmatrix}^T K\left(\frac{|S_k X_{ij}|^2}{h_k^2} \right) \right\}^{-1} \sum_{i=1}^{n} Y_{new,i} \begin{pmatrix} 1 \\ X_{ij} \end{pmatrix} K\left(\frac{|S_k X_{ij}|^2}{h_k^2} \right)$.

Step 4. Compute the vector $\widehat{\theta}_0^{(k)} = \frac{1}{n}\sum_{j=1}^{n} \widehat{\nabla\eta_0}((\theta_0^{(k)})^T X_j)$.

Step 5. Set $h_{k+1} = a_h h_k$, $\rho_{k+1} = a_\rho \rho_k$. If $\rho_{k+1} > \rho_{min}$, then set $k = k+1$ and continue with step 1; otherwise terminate and let $\hat{\alpha}_0^* = \hat{\theta}_0^{(k)}$, $\hat{\alpha}_0 = \hat{\alpha}_0^*/|\hat{\alpha}_0^*|$.

In all, the target estimators can be established through the two-stage estimation procedure stated previously, which employs both the cubic spline method and average derivative method together. The next section will provide how to choose the appropriate smoothing parameter λ.

2.3. *Choosing the Smoothing Parameter*

The selection of the smoothing parameter is important in modeling. There are mainly two views regarding to the choice of the smoothing parameter. One is based on people's experience that is free of the data and the other is chosen by the data or namely data driven method. The first one in reality probably is the most useful. The data-driven method includes cross-validation (CV), generalized cross-validation (GCV), AIC, BIC, and so on. In the following, we will describe the CV approach in general which is probably the oldest and widely used one in many literatures. The general idea of it is to minimize some function concerning the smoothing parameter with respect to this parameter to get the "idea" parameter value theoretically. For the model (3), we minimize the following equation

$$CV(\lambda) = n^{-1}\sum_{i=1}^{n}\{Y_i - \widehat{\eta_0}^{(-i)}(U_i, \lambda) - (\beta^{(-i)})^T Z_i\}^2 \tag{13}$$

with respect to λ. We choose the smoothing parameter λ to be the minimizer of (13). Here, $\widehat{\eta_0}^{(-i)}(U_i, \lambda)$ and $\beta^{(-i)}$ are the minimizer of

$$\sum_{j\neq i}\{Y_j - \eta(U_j) - \beta^T Z_j\}^2 + \lambda\int [\eta''(u)]^2 du.$$

It cannot be guaranteed that the function $CV(\lambda)$ has a unique minimum, so attention should be paid on the uniqueness of the solution. It is suggested that a simple grid search is probably the best approach in practice. The simulated examples in Sec. 4 adopt the simple grid search method when seeking the effective value of smoothing parameter.

3. Estimation Procedure

In this section, we will illustrate how the estimate procedure presented previously would be implemented in practice, and of all the difficulties one is the selection of the initial value of the single-index parameter α_0 for the estimation procedure. Precisely, it is an iteration procedure. We randomly select α_0 from the unit half-sphere from R^d with the first non zero element positive. Here is the detailed procedure.

Step 1. Randomly select the initial value of $\alpha_0^{(0)}$ from the unit half-sphere from R^d with first nonzero element positive.

Step 2. Apply the cubic spline algorithm which has been described in Sec. 2 to get the estimator $\widehat{\beta_0}$ of β_0.

Step 3. Make a simple transformation (8) and then employ the ADE technique in Sec. 2.2 to obtain the $\widehat{\alpha_0}^{(1)}$ for the first iteration.

Step 4. View the $\widehat{\alpha_0}^{(1)}$ as the initial value of the second iteration and continue from Step 2 until it converges.

Step 5. Obtain the estimator of α_0: let $m(n)$ be the total number of iterations. Set $\hat{\alpha}_0 = \widehat{\alpha_0}^{(m(n))}/|\widehat{\alpha_0}^{(m(n))}|$.

Step 6. Obtain the estimators of $\eta_0(\cdot)$ and β_0: replacing α_0 by $\hat{\alpha}_0$ in (7) and resulting it, the estimators of $\eta_0(\cdot)$ and β_0 can be obtained.

4. Simulation and Real Example

4.1. *Simulation Example*

The model we want to analyze is the same as the one that appears in Yu and Ruppert (2002). Data are generated from the model

$$Y_i = sin\left\{\frac{\pi(\alpha^T X_i - A)}{C - A}\right\} + \beta Z_i + \varepsilon_i. \tag{14}$$

Here, the X_i are trivariate with independent uniform (0,1) components, $Z_i = 0$ for i odd and $Z_i = 1$ for i even. A is taken to be $A = \sqrt{3}/2 - 1.645/\sqrt{12}$ and $C = \sqrt{3}/2 + 1.645/\sqrt{12}$ as did Carroll et al. (1997). We present the results for cases where $\alpha_0 = \frac{1}{\sqrt{3}}(1, 1, 1)$, $\beta = 0.3$, $\sigma = 0.1$, $n = 100$ and 200, respectively. For each case with sample size of 100 or 200, we will simulate 200, 400, and 600 times and provide corresponding figures and tables to illustrate the performance of our proposed method. During the ADE procedure we take the kernel function to be $k(t) = \frac{3}{4}(1 - t^2)_+^2$.

Table 1 presented below is the summary of estimated values of parameters for different simulated times with sample size of 100. The means (Mean), standard errors (SE), bias, and mean squared errors (MSE) of these estimators of the parameters $\alpha_0^{(i)}$ (the i-th entry of parameter α_0), $i = 1, 2, 3$ and β_0 are listed, Table 2 for sample size of 200. Figure 1 shows in a typical estimate of unknown function $\eta_0(\cdot)$ and its limits of the 95% confidence intervals: (a) for $n = 100$; (b) for $n = 200$. The typical sample is selected in such a way that its MSE is equal to the median in the 200 replications.

From Tables 1 and 2, we can see that these means of these estimators of these parameters are very close those true values, respectively, and these standard errors, bias, and mean squared errors of these estimators of the parameters are very small, no matter for the sample size of 100 or 200, and no matter in 200 replications, or in 400 replications, or in 600 replications. From Fig. 1 the estimate of the unknown function $\eta_0(\cdot)$ agrees with the true function very closely. In a word, our proposed estimation procedure works well.

Table 1
Summary of the estimated values of parameters for different simulated times with the sample size of 100

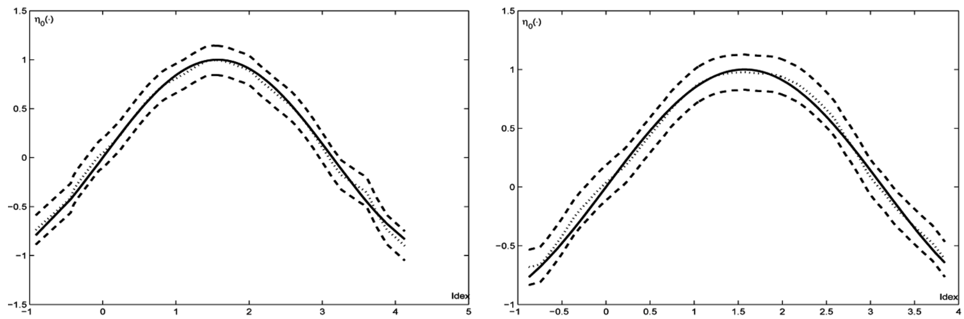
Parameter	Simulated times	Mean	SE	Bias(10^{-3})	MSE(10^{-3})
$\alpha_0^{(1)}$	200	0.5726985	0.0486753	−4.851769	2.390924
	400	0.5750721	0.05167847	−2.278120	2.675854
	600	0.575258	0.05522577	−2.092252	3.054263
$\alpha_0^{(2)}$	200	0.5788902	0.04812349	1.539934	2.318241
	400	0.5752077	0.05075389	−2.142522	2.580548
	600	0.5731454	0.01316511	−4.204869	2.215793
$\alpha_0^{(3)}$	200	0.567978	0.08217174	−9.372222	6.840033
	400	0.5734995	0.06548053	−3.850754	4.302529
	600	0.5713011	0.008129085	−6.049206	6.644795
β_0	200	0.2913312	0.035679996	−1.865432	5.234168
	400	0.2966826	0.026547232	−1.221245	4.389936
	600	0.3001698	0.006222996	2.0086643	3.312657

Table 2
Summary of the estimates of parameters for different simulated times with the sample size of 200

Parameter	Simulated times	Mean	SE	Bias(10^{-3})	MSE(10^{-3})
$\alpha_0^{(1)}$	200	0.5741909	0.02387754	−3.199375	0.5803727
	400	0.5789361	0.02540472	1.585842	0.6479149
	600	0.5779919	0.02347418	0.6416725	0.5514488
$\alpha_0^{(2)}$	200	0.5777333	0.02501009	0.3830533	0.6256514
	400	0.574365	0.02414208	−2.485251	0.5917516
	600	0.5764323	0.02413244	−0.9179995	0.5832174
$\alpha_0^{(3)}$	200	0.5786073	0.02453342	1.257038	0.6034691
	400	0.5772789	0.02161280	−0.07136874	0.4671184
	600	0.5761418	0.02407984	−1.208506	0.581299
β_0	200	0.2934423	0.03005678	−1.331233	5.112311
	400	0.2956342	0.01965788	−1.004568	4.008679
	600	0.3000896	0.00321675	2.0095467	3.012386

5. Real Data–Air Pollution Data

The air pollution data are concerned with environment study of how the concentration y of the air pollutant ozone relates with three meteorological variables $\mathbf{X}$: wind speed, x_1; temperature, x_2; and radiation, x_3. The data are daily measurements of the four variables for $n = 111$ days. Yu and Ruppert (2002) employed a partially linear single-index model using a P-splines in which temperature and wind are the two components of the index and radiation is the linear term to analyze the data. To see the performance of our proposed method, we also deal with the data and compare our results with that in Yu and Ruppert (2002). Similar to Yu and Ruppert (2002), we also let x_1, x_2 be treated as components



(a) (b)

Figure 1. Simulation results: (a) the estimate (dot-line) of the unknown function $\eta_0(\cdot)$ (solid-line) and its limits of the 95% confidence bands (dash lines) in the 200 replications for sample size of 100; (b) for sample size of 200.

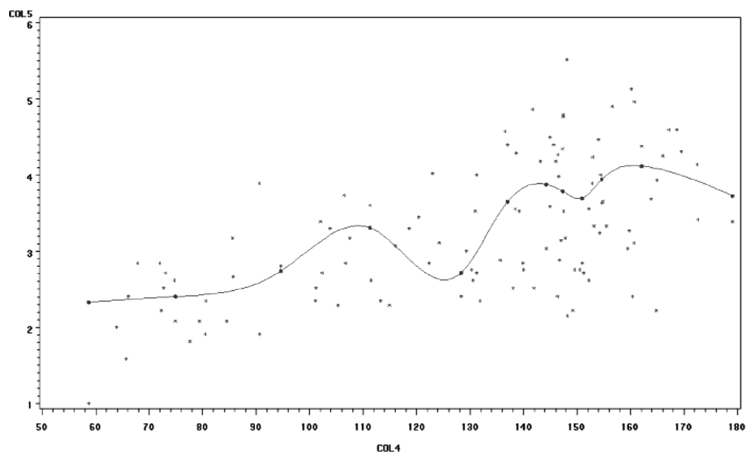


Figure 2. Curve estimates for air pollution data. The data are represented by open circles, and the solid curve corresponds to the estimate of the single index function $\eta_0(\cdot)$ of the partially linear single-index model by our method.

Table 3
Summary of two fits for air pollution data

	Parameter		
Method	Temperature	Wind	Radiation
Method 1	0.5288	−0.8569	0.0021
SE	0.0002	0.0001	0.00001
Method 2	0.5224	−0.8527	0.0024
SE	0.0005	0.0003	0.00005

of single-index part and x_3 the liner term in a partially linear single-index model. Figure 2 shows the singe index curve estimates $\hat{\eta}_0$ of the air pollution data. Table 3 gives the fitting results (parameter estimates, standard errors (SE)) by our method (Method 1) and by Yu and Ruppert (2002) (Method 2).

We conclude from Table 3 that our method is better than that by Yu and Ruppert (2002), since every SE of estimator of every parameter obtained is less than that by Yu and Ruppert (2002).

Acknowledgments

The research was supported in part by National Natural Science Foundation of China (10871072), Doctoral Fund of Ministry of Education of China (20090076110001), Shanxi's Natural Science Foundation of China (2007011014).

References

Carroll, R. J., Fan, J. Q., Gijbel, I., Wand, M. P. (1997). Generalized partially linear single-index models. *Journal of the American Statistical Association* 92:477–489.

Green, P. J., Silverman, B. W. (1994). *Nonparametric regression and generalized linear models: a roughness penalty approach.* London: Chapman and Hall.

Härdle, W., Stoker, T. M. (1989). Investing smooth multiple regression by the method of average derivatives. *Journal of the American Statistical Association* 84:986–995.

Härdle, W., Hall, P., Ichimura, H. (1993). Optimal smoothing in single-index models. *The Annals of Statistics* 21:157–178.

Härdle, W., Tsybakov, A. B. (1993). How sensitive are average derivatives. *Journal of Econometrics* 58:31–48.

Hristache, M., Juditsky, A., Spokoiny, V. (2001). Direct estimation of the index coefficient in a single-indeex model. *The Annals of Statistics* 29:595–623.

Ichimura, H. (1993). Semiparametric least squares (SLS) and weighted SLS estimation of single-index models. *Journal of Econometrics* 58:71–120.

Li, K. C. (1991). Sliced inverse regression for dimension reduction. *Journal of the American Statistical Association* 86:316–342.

Powell, J. L., Stock, J. H., Stoker, T. M. (1989). Semiparametric estimation of index model coefficients. *Econometrica* 57:1403–1430.

Samarov, A. (1993). Exploring regression structure using nonparametric functional estimation. *Journal of the American statistical Association* 88:836–847.

Stoker, T. M. (1986). Consistent estimation of scaled coefficients. *Econometrica* 54:1461–1481.

Xia, Y., Härdle, W. (2006). Semi-parametric estimation of partially linear single-index models. *Journal of Multivariate Analysis* 97:1162–1184.

Yu, Y., Ruppert, D. (2002). Penalized spline estimation for partically linear single-index models. *Journal of the American Statistical Association* 97(460):1042–2002.