

Efficient estimation and variable selection for partially linear single-index-coefficient regression models

Young-Ju Kim^{1,a}

^aDepartment of Statistics, Kangwon National University, Korea

Abstract

A structured model with both single-index and varying coefficients is a powerful tool in modeling high dimensional data. It has been widely used because the single-index can overcome the curse of dimensionality and varying coefficients can allow nonlinear interaction effects in the model. For high dimensional index vectors, variable selection becomes an important question in the model building process. In this paper, we propose an efficient estimation and a variable selection method based on a smoothing spline approach in a partially linear single-index-coefficient regression model. We also propose an efficient algorithm for simultaneously estimating the coefficient functions in a data-adaptive lower-dimensional approximation space and selecting significant variables in the index with the adaptive LASSO penalty. The empirical performance of the proposed method is illustrated with simulated and real data examples.

Keywords: partially linear, penalized likelihood, smoothing splines, variable selection, single-index, varying coefficient model

1. Introduction

In a generalized linear model, the regression function $\mu(x, z) = E(Y|X = x, Z = z)$ is modeled linearly through a link function. Various structured models have been proposed in the literature for modeling high dimensional data. A structured model combined with both single-index and varying coefficients has been recently proposed for Gaussian error with the identity link,

$$Y = \eta^T (\alpha^T X) Z + \epsilon,$$

where Y is a response variable, $X(\in R^p)$ and $Z(\in R^d)$ are covariate vectors, $\eta(\cdot) = (\eta_1(\cdot), \dots, \eta_d(\cdot))^T$ is a vector of unknown functions, $\alpha = (\alpha_1, \dots, \alpha_p)^T$ is a vector of unknown parameters, and ϵ is a random error with mean zero. It is assumed that $\|\alpha\| = 1$ and $\text{sign}(\alpha_1) = 1$ for identifiability of the model. This model is called single-index-coefficient model (SICM). The SICM has advantages of avoiding the curse of dimensionality that multivariate nonparametric models can suffer because it uses a univariate nonparametric function with a single-index. It also has nonlinear interaction effects between index covariates and other covariates. These advantages are the reason for its popularity in modeling for many scientific fields such as medical areas, biostatistics, economics, and environmental studies.

¹ Department of Statistics, Kangwon National University, 1 Gangwondaehak-gil, Chuncheon-si, Gangwon-do, 24341, Korea. E-mail: ykim7stat@kangwon.ac.kr

Allowing a linear association between the response and covariates in SICM yields a partially linear single-index-coefficient model (PLSICM),

$$Y = \eta^T (\alpha^T X) Z_1 + \beta^T Z_2 + \epsilon. \quad (1.1)$$

The semiparametric model (1.1) includes the partially linear varying coefficient model (PLVCM) and the standard varying coefficient model without a single-index. When $Z_1 = 1$, the PLSICM becomes the classical partially linear single-index model (PLSIM).

For the estimation of index parameters and unknown coefficient function in PLSICM, various estimation methods have been proposed such as local linear method, kernel method, B-splines, empirical likelihood method, and penalized splines (Xia and Li, 1999; Xue and Wang, 2012; Huang, 2012; Yang *et al.*, 2014). In the nonparametric or semiparametric single-index models, variable selection methods have been suggested using the least absolute shrinkage and selection operator (LASSO) (Tibshirani, 1996), the smoothly clipped absolute deviation (SCAD) (Fan and Li, 2001), and the adaptive LASSO (Zou, 2006) (Peng and Huang, 2011; Foster *et al.*, 2013; Huang *et al.*, 2013; Huang *et al.*, 2014; Yang and Yang, 2014; Zhu *et al.*, 2015). However, limited studies have been done on the selection and simultaneous estimation for both varying coefficients and single-index in a smoothing splines framework.

In this paper, we propose a simple estimation method based on a smoothing splines approach for selecting variables in the index and simultaneously estimating unknown nonparametric functions, regression parameters in the partial terms, and index parameters. Smoothing splines have advantages over other nonparametric estimation methods because they can avoid the problem of choice and the placements of knots.

The paper is organized as follows. Section 2 presents a smoothing splines technique of a data-adaptive lower-dimensional approximation in a penalized likelihood method in an ordinary nonparametric setting so as to speed up the computation of function estimates without any loss of performance and extend it to PLSICM. We propose a simple and efficient method for estimating and selecting index parameters based on the penalization approach. Simulated and real data examples are illustrated to evaluate the performance of the proposed method in Section 3 and Section 4 respectively. Performance comparisons are made with different penalties and other estimation methods. The paper is concluded with a discussion in Section 5.

2. The model

2.1. Smoothing splines in partially linear single-index-coefficient model

Suppose that the data $(X_i, Z_{1i}, Z_{2i}, Y_i)$, $i = 1, \dots, n$, are independent and identically distributed samples from the PLSICM (1.1). For given α with $\|\alpha\| = 1$ and positive first element, the function estimates in (1.1) can be obtained by iteratively minimizing the following penalized least squares functional

$$\sum_{i=1}^n \left\{ Y_i - \eta^T (\alpha^T X_i) Z_{1,i} - \beta^T Z_{2,i} \right\}^2 + n \sum_{j=1}^d \lambda_j J(\eta_j), \quad (2.1)$$

where $J(\eta)$ is the penalty function of the roughness of η , and the smoothing parameter λ controls the trade-off between the lack of fit and the roughness of η .

For given α , let $u = \alpha^T X$. The minimizer of (2.1) is in infinite-dimensional space $\mathcal{H} \subseteq \{f : J(f) < \infty\}$; it lies in a Hilbert space $\mathcal{H} = \mathcal{N}_J \oplus \mathcal{H}_J$, where $\mathcal{N}_J = \{f : J(f) = 0\}$ is the null space of $J(f)$, and space \mathcal{H}_J is a reproducing kernel Hilbert space (RKHS) with $J(f)$ as the square norm. Letting

$J(f) = \int_0^1 \dot{f}^2 du$ on $\mathcal{U} = [0, 1]$ yields the popular cubic splines with $\mathcal{N}_J = \text{span}\{1, k_1(u)\}$, where $k_1(u) = u - 0.5$. $\mathcal{H}_J = \{f : \int_0^1 f du = \int_0^1 \dot{f} du = 0, J(f) < \infty\}$ with $J(f)$ as the square norm provides the reproducing kernel $R_J(u_1, u_2) = k_2(u_1)k_2(u_2) - k_4(u_1 - u_2)$, where $k_\nu = B_\nu/\nu!$ are scaled Bernoulli polynomials. Gu (2013) provides details of the RKHS and its properties.

For estimation of η and β in (1.1), a data-adaptive lower-dimensional approximation in penalized likelihood methods, as originally proposed by Gu and Kim (2002) is used to speed up the computation of function estimates without any loss of performance. It has been shown that the convergence rate of the minimizer of the penalized likelihood functional in $\mathcal{H}_n = \mathcal{N}_J \oplus \mathcal{H}_J$ is the same as that in the lower-dimensional function space $\mathcal{H}_q = \mathcal{N}_J \oplus \text{span}\{R_J(w_j, \cdot), j = 1, \dots, q\}$, where $\{w_j\}$ are random subsets of $\{u_i, i = 1, \dots, n\}$, as long as $q \asymp n^{2/(mr+1)+\delta}$, where for some $m \in [1, 2]$, $r > 1$, $\delta > 0$ is arbitrary. The smoothness of the true function is represented by m . We let $m = 2$ under the assumption that the true function is sufficiently smooth. The constant r characterizes the smoothness of the model, and $r = 4$ is used for cubic splines.

For fixed λ , the minimizer of (2.1) in \mathcal{H}_q can be written as

$$\eta_\lambda(u_i) = \sum_{\nu=1}^m d_\nu \phi_\nu(u_i) + \sum_{j=1}^q c_j R_J(w_j, u_i), \quad (2.2)$$

where $\{\phi_\nu\}$ is a basis of null space \mathcal{N}_J .

Inserting (2.2) to (2.1) becomes a minimization problem of the penalized least squares functional to find the vectors $(c_1, \dots, c_q)^T$ and $(d_1, \dots, d_m)^T$. The estimation of β can be obtained as a byproduct of partial splines by adding $\beta^T Z_2$ into the unpenalized term in (2.2). Details can be found in Kim and Gu (2004) and Gu (2013).

Selecting appropriate smoothing parameters in nonparametric function estimation is important because they determine the performance of the function estimates. Kim and Gu (2004) suggested the following modification to the generalized cross-validation (GCV) score,

$$\text{GCV}_\gamma(\lambda) = \frac{n^{-1} Y^T (I - A(\lambda))^2 Y}{[n^{-1} \text{tr}(I - \gamma A(\lambda))]^2}, \quad (2.3)$$

where $A(\lambda)$ is the smoothing matrix with the fitted values $\hat{Y} = A(\lambda)Y$. We let $\gamma = 1.4$ as suggested in Kim and Gu (2004).

2.2. Estimation and selection of single-index parameters

For given the current estimates of η and β , the estimation and selection of α is obtained by minimizing the penalized least squares functional with a penalty on α . For given α_0 , we employ a first-order approximation

$$\eta(\alpha^T X) Z_1 \approx \eta(\alpha_0^T X) Z_1 + \eta'(\alpha_0^T X) Z_1 (\alpha - \alpha_0)^T X.$$

Then we have

$$Y - \eta(\alpha^T X) Z_1 - \beta^T Z_2 \approx Y - \eta(\alpha_0^T X) Z_1 - \eta'(\alpha_0^T X) Z_1 [\alpha^T X - \alpha_0^T X] - \beta^T Z_2.$$

Therefore, we derive the following penalized least squares functional for α for given η and β ,

$$\sum_{i=1}^n (\hat{Y}_i - \alpha^T \tilde{X}_i)^2 + nP_\lambda(\alpha), \quad (2.4)$$

which is to be minimized, where $\tilde{Y} = Y - \eta(\alpha_0^T X)Z_1 + \eta'(\alpha_0^T X)Z_1(\alpha_0^T X) - \beta^T Z_2$ and $\tilde{X} = \eta'(\alpha_0^T X)Z_1 X$. Note that we assume that η is smooth enough.

For the penalty in (2.4), we use the adaptive LASSO penalty $P_\lambda(\alpha) = \lambda \sum_{j=1}^d w_j |\alpha_j|$, where λ is another smoothing parameter. We choose the weight $\hat{w}_j = 1/|\hat{\alpha}_j|$. It is well-known that the adaptive LASSO enjoys the oracle properties. It uses adaptive weights for penalizing different coefficients in the l_1 penalty (Zou, 2006). The performance of variable selection of the adaptive LASSO was compared with other penalty functions in our simulations.

2.3. Computational algorithm

Given α , the asymptotic efficiency for the function estimator of η in the lower-dimensional function space is obtained for $q \asymp n^{2/(mr+1)+\delta}$, $\forall \delta > 0$ (Gu and Kim, 2002). We take their suggestion of $q = kn^{2/(4m+1)}$ for cubic splines, $m = 2$ under the assumption that the true function η is sufficiently smooth, and $k = 10$ for the computation.

The estimator of α is calculated by a minimizer of the penalized least squares functional (2.4) for given η and β . In order to improve the selection performance, we first consider to take the estimation of α by minimizing (2.4) without penalization, which gives nonzero estimates to all index parameters. Then we take this estimate as a starting value for the selection of index parameters with the adaptive LASSO penalty. The penalized estimates of index parameters with the adaptive LASSO are computed using two nested loops: (1) for fixed smoothing parameter for the index parameters, the inner loop computes the optimal minimizers of the penalized least squares functional of α by Newton iteration. (2) In the outer loop, the optimal smoothing parameter λ for the index parameters is obtained by minimizing the ordinary GCV score by the grid search with a grid (0.0005, 0.1) by 0.002.

Algorithm

- Step 1. Start with an initial estimator of α by minimizing (2.4) with no penalty. For example, set an initial vector for α by randomly selecting from a uniform distribution on $[0, 1]$. Calculate the estimates of η and β by the smoothing splines for given $\hat{\alpha}$. Then given the current estimates of η and β , update the estimate of α by minimizing (2.4) with no penalty. Iterate these two steps until the estimate of α converges.
- Step 2. Given $\hat{\alpha}$, let $u = \hat{\alpha}^T X$. Calculate the estimates of η and β by the smoothing splines.
- Step 3. Given $\hat{\eta}$ and $\hat{\beta}$, the adaptive LASSO estimate of α is calculated by minimizing the penalized least squares functional (2.4). For the initial value for $\hat{w}_j = 1/|\hat{\alpha}_j|$ in the adaptive LASSO penalty, the minimizer of (2.4) with no penalty was used. At each iteration, \hat{w}_j was updated by the minimizer of (2.4) at the previous step.
- Step 4. Repeat step 2 and 3 until convergence. The final estimates of η and β are obtained at convergence of α . Also, a GCV score is calculated for a fixed smoothing parameter.

If there are more than one unknown functions to estimate, a Gauss-Seidel type algorithm (backfitting algorithm) estimates each of the coefficient functions iteratively. Note that the classical smoothing splines on the product domain are calculated via smoothing spline ANOVA decomposition. However, a similar decomposition cannot be used to obtain varying-coefficient function estimates in our model due to the association between predictors Z and varying-coefficient functions η (Leng, 2009).

2.4. Interval inference

Bayesian confidence intervals for a minimizer of the penalized likelihood functional were first derived by Wahba (1983) from the Bayes model of a penalized likelihood estimator. Consider $f = f_0 + f_1$, where f_0 has a diffuse prior in \mathcal{N}_J and f_1 has a zero-mean Gaussian process prior with covariance function

$$E[f_1(u_1)f_1(u_2)] = bR_J(u_1, w^T)Q^+R_J(w^T, u_2),$$

where Q^+ is the Moore-Penrose inverse of Q . Setting $b = 1/n\lambda$, the minimizer of (1.1) in \mathcal{H}_q is seen to be the posterior mode under this prior. The Bayesian confidence intervals for the coefficient function estimates are then obtained. The detailed derivations are described in Kim and Gu (2004).

3. Simulations

A simulation study was conducted to evaluate the performance of the proposed estimators. The following criteria are considered to assess the performance of the selection of index covariates; **IZ** represents the average number of nonzero index parameters that are incorrectly selected as zero; **CZ** is the average number of zero index parameters that are correctly selected. The biases and standard deviations of the estimates of α and β are calculated respectively. The performance of the estimation of η is evaluated by the square root of the average squared error (RASE) defined as

$$\text{RASE} = \left\{ \frac{1}{n_{\text{grid}}} \sum_{i=1}^{n_{\text{grid}}} \{\hat{\eta}(u_i) - \eta(u_i)\}^2 \right\}^{\frac{1}{2}}.$$

In each example, we carried out 200 simulations and the sample size in each simulation is set to $n = 100, 200, \text{ and } 300$. The results are summarized in Table 1 and Table 2. For comparison of performance of the adaptive LASSO penalty, we also used the LASSO penalty $P_\lambda(\alpha) = \lambda \sum_{j=1}^d |\alpha_j|$ and the SCAD penalty $P_\lambda(\alpha) = \lambda \sum_{j=1}^d \theta_\lambda(\alpha_j)$, where

$$\theta_\lambda(t) = \lambda \left\{ I_{(t \leq \lambda)} + \frac{(a\lambda - t)_+}{(a-1)\lambda} I_{(t > \lambda)} \right\},$$

and $a = 3.7$.

Example 1. A simple example of the SICM is the partially linear single-index model, which can be written as

$$Y = \eta(\alpha^T X) + \beta^T Z + \epsilon,$$

where $\eta(u) = \sin(\pi(u - a)/(c - a))$, $a = \sqrt{3}/2 - 1.645/\sqrt{12}$ and $c = \sqrt{3}/2 + 1.645/\sqrt{12}$, $\alpha = (3, 1.5, 0, 0, 2, 0, 0, 0)^T / \sqrt{12.25}$, and $\beta = (2, 1.6, 0.8)$. The covariates $X = (X_1, X_2, \dots, X_8)^T$ are generated independently from uniform distribution $U(0, 1)$ and $Z = (z_1, z_2, z_3)^T$ are generated from multivariate normal distribution $N(0, \Sigma)$ with $\Sigma = (\sigma_{ij})$ having entries $\sigma_{ij} = 1$ for $i = j$ and $\sigma_{ij} = 0.6$ for $i \neq j$ and ϵ is generated independently from $N(0, \sigma^2)$ with $\sigma = 0.1$.

Example 2. In this example, we consider the following SICM

$$Y = \eta_1(\alpha^T X)Z_1 + \eta_2(\alpha^T X)Z_2 + \beta^T Z_3 + \epsilon,$$

Table 1: Bias, SD, and model selection results (IZ, CZ) for the estimates of α in Examples 1–4

n	Method	α_1		α_2		α_5		IZ	CZ	
		Bias	SD	Bias	SD	Bias	SD			
Example 1	100	SS-LASSO	0.0272	0.1257	0.0221	0.0932	0.0431	0.2005	0.000	1.495
		SS-SCAD	0.0289	0.1226	0.0158	0.0894	0.0407	0.1991	0.030	1.840
		SS-ALASSO	0.0044	0.0903	0.0228	0.0753	0.0258	0.1430	0.060	4.900
	200	SS-LASSO	0.0023	0.0501	0.0079	0.0355	0.0241	0.1689	0.000	2.015
		SS-SCAD	0.0049	0.0471	0.0039	0.0307	0.0216	0.1640	0.000	2.740
		SS-ALASSO	0.0017	0.0551	0.0159	0.0620	0.0310	0.1920	0.035	4.935
	300	SS-LASSO	-0.0014	0.0080	0.0028	0.0073	0.0004	0.0101	0.000	2.375
		SS-SCAD	0.0005	0.0076	0.0003	0.0100	-0.0005	0.0098	0.000	2.850
		SS-ALASSO	-0.0029	0.0104	0.0058	0.0146	0.0005	0.0107	0.000	5.000
Example 2	100	SS-LASSO	0.0097	0.0880	0.0172	0.0586	0.0143	0.0934	0.005	2.430
		SS-SCAD	0.0236	0.0924	0.0286	0.0952	0.0163	0.1213	0.035	3.095
		SS-ALASSO	0.0225	0.1396	0.0439	0.1136	0.0187	0.1061	0.135	4.805
	200	SS-LASSO	0.0012	0.0452	0.0026	0.0086	0.0021	0.0121	0.000	3.480
		SS-SCAD	0.0069	0.0542	-0.0014	0.0095	0.0018	0.0181	0.000	3.885
		SS-ALASSO	-0.0029	0.0062	0.0044	0.0094	0.0013	0.0047	0.000	5.000
	300	SS-LASSO	-0.0015	0.0028	0.0024	0.0043	0.0005	0.0030	0.000	3.355
		SS-SCAD	-0.0001	0.0024	0.0005	0.0039	-0.0001	0.0029	0.000	3.920
		SS-ALASSO	-0.0020	0.0031	0.0037	0.0053	0.0003	0.0030	0.000	5.000
Example 3	100	SS-LASSO	0.0834	0.1481	0.0635	0.1317	0.0342	0.1920	0.035	0.965
		SS-SCAD	0.0962	0.1526	0.0515	0.1352	0.0365	0.1908	0.045	1.335
		SS-ALASSO	0.0645	0.1617	0.0653	0.1389	0.0225	0.1665	0.195	3.850
	200	SS-LASSO	0.0213	0.1112	0.0301	0.0916	0.0070	0.0912	0.000	0.780
		SS-SCAD	0.0255	0.1113	0.0261	0.0906	0.0025	0.0795	0.010	1.135
		SS-ALASSO	0.0100	0.0913	0.0228	0.0749	0.0021	0.0734	0.031	4.740
	300	SS-LASSO	-0.0010	0.0025	0.0087	0.0413	0.0047	0.0327	0.000	0.750
		SS-SCAD	0.0012	0.0246	0.0063	0.0403	0.0038	0.0324	0.000	1.060
		SS-ALASSO	-0.0048	0.0271	0.0120	0.0463	0.0024	0.0323	0.000	4.920
Example 4	100	SS-LASSO	-0.0020	0.0110	0.0081	0.0166	-0.0005	0.0102	0.000	2.225
		SS-SCAD	0.0009	0.0080	0.0032	0.0145	-0.0023	0.0094	0.000	3.420
		SS-ALASSO	-0.0079	0.0164	0.0159	0.0272	0.0013	0.0119	0.000	5.000
	200	SS-LASSO	-0.0027	0.0054	0.0040	0.0080	0.0015	0.0059	0.000	2.890
		SS-SCAD	-0.0003	0.0042	0.0009	0.0063	0.0002	0.0055	0.000	3.552
		SS-ALASSO	-0.0009	0.0552	0.0093	0.0242	0.0050	0.0495	0.005	5.000
	300	SS-LASSO	-0.0015	0.0034	0.0025	0.0047	0.0005	0.0035	0.000	3.453
		SS-SCAD	0.0003	0.0028	0.00005	0.0041	-0.0004	0.0033	0.000	3.760
		SS-ALASSO	-0.0037	0.0068	0.0068	0.0109	0.0008	0.0042	0.015	5.000

SD = standard deviation; IZ = the average number of nonzero index parameters that are incorrectly selected as zero; CZ = the average number of zero index parameters that are correctly selected; SS = smoothing spline; LASSO = least absolute shrinkage and selection operator; SCAD = smoothly clipped absolute deviation; ALASSO = adaptive LASSO.

where $\eta_1(u) = \sin(\pi(u - a)/(c - a))$, $a = \sqrt{3}/2 - 1.645/\sqrt{12}$ and $c = \sqrt{3}/2 + 1.645/\sqrt{12}$ and $\eta_2(u) = 1 + 3u^2$. The covariates $X = (X_1, X_2, \dots, X_8)^T$ are generated independently from uniform distribution $U(0, 1)$ and $Z_1 = 1$, $(Z_2, Z_3^T)^T$ are generated from the multivariate normal distribution $N(0, \Sigma)$ with $\Sigma = (\sigma_{ij})$ having entries $\sigma_{ij} = 1$ for $i = j$ and $\sigma_{ij} = 0.6$ for $i \neq j$ and α , β , and ϵ are the same as those in Example 1.

Example 3. This example is the same as the Example 2, but with $\sigma = 1$.

Example 4. This example is the same as the Example 2, but with an additional 20 noise covariates, so that $\alpha = (3, 1.5, 0, 0, 2, 0, 0, 0, \mathbf{0}_{1 \times 20})^T / \sqrt{12.25}$.

Table 1 showed that the adaptive LASSO performed well in estimating and selecting the significant

Table 2: Bias and SD of β and RASEs of η in Examples 1–4

n	Method	β_1		β_2		β_3		RASE ₁	RASE ₂	
		Bias	SD	Bias	SD	Bias	SD			
Example 1	100	SS-LASSO	-0.0018	0.0263	-0.0037	0.0177	0.0017	0.0163	0.0360	-
		SS-SCAD	-0.0020	0.0225	-0.0029	0.0162	0.0004	0.0150	0.0376	-
		SS-ALASSO	-0.0011	0.0200	-0.0030	0.0156	0.0018	0.0137	0.0331	-
	200	SS-LASSO	0.00006	0.0099	-0.0021	0.1017	0.0001	0.0090	0.0279	-
		SS-SCAD	0.0003	0.0099	-0.0022	0.0103	0.00004	0.0090	0.0239	-
		SS-ALASSO	0.0001	0.0099	-0.0030	0.0115	-0.0003	0.0090	0.0295	-
	300	SS-LASSO	-0.0003	0.0075	0.0013	0.0073	-0.00002	0.0077	0.0213	-
		SS-SCAD	-0.0003	0.0075	0.0013	0.0074	-0.0002	0.0077	0.0192	-
		SS-ALASSO	-0.0005	0.0093	0.0016	0.0073	-0.00006	0.0077	0.0192	-
Example 2	100	SS-LASSO	0.0041	0.0342	0.0055	0.0260	-0.0015	0.0306	0.0341	0.0520
		SS-SCAD	0.0009	0.0539	0.0129	0.0484	0.0011	0.0457	0.0329	0.0474
		SS-ALASSO	0.0022	0.0479	0.0082	0.0375	0.0054	0.0411	0.0310	0.0424
	200	SS-LASSO	-0.0016	0.0110	-0.0009	0.0105	-0.0004	0.0120	0.0003	0.0005
		SS-SCAD	-0.5110	0.0145	-0.0004	0.0110	-0.0001	0.0118	0.0003	0.0005
		SS-ALASSO	-0.0013	0.0100	-0.0003	0.0099	-0.0008	0.0098	0.0003	0.0004
	300	SS-LASSO	-0.0001	0.0086	-0.00004	0.0087	-0.0005	0.0083	0.0152	0.0241
		SS-SCAD	-0.0001	0.0086	-0.0009	0.0086	-0.0004	0.0083	0.0151	0.0227
		SS-ALASSO	-0.0001	0.0085	0.0004	0.0086	-0.0003	0.0084	0.0151	0.0233
Example 3	100	SS-LASSO	0.0139	0.1885	0.0488	0.1700	0.0062	0.1965	0.3651	0.6765
		SS-SCAD	0.0146	0.1808	0.0486	0.1822	0.0079	0.2018	0.3770	0.7419
		SS-ALASSO	0.0223	0.1850	0.0435	0.1695	-0.0058	0.1861	0.2818	0.4043
	200	SS-LASSO	-0.0027	0.1008	0.0133	0.1030	-0.0141	0.1033	0.1966	0.3609
		SS-SCAD	-0.0033	0.1013	0.0122	0.1030	-0.0122	0.1025	0.1962	0.3695
		SS-ALASSO	-0.0019	0.0977	0.0163	0.1026	-0.0192	0.1031	0.1387	0.2197
	300	SS-LASSO	0.0061	0.0880	0.0037	0.0877	-0.0048	0.0835	0.1375	0.2291
		SS-SCAD	0.0063	0.0881	0.0038	0.0877	-0.0046	0.0837	0.1347	0.2307
		SS-ALASSO	0.0038	0.0858	0.0073	0.0864	-0.0011	0.0835	0.1101	0.1472
Example 4	100	SS-LASSO	-0.0012	0.0168	0.0025	0.0182	-0.0025	0.0202	0.0537	0.1086
		SS-SCAD	-0.0004	0.0170	0.0043	0.0169	-0.0020	0.0196	0.0366	0.0575
		SS-ALASSO	-0.0008	0.0158	0.0057	0.0156	0.0023	0.0197	0.0313	0.0428
	200	SS-LASSO	-0.0008	0.0105	-0.0010	0.0103	-0.0027	0.0105	0.0286	0.0617
		SS-SCAD	-0.0008	0.0107	-0.0008	0.0104	-0.0025	0.0107	0.0222	0.0408
		SS-ALASSO	-0.0001	0.0107	-0.00006	0.0102	-0.0003	0.0122	0.0180	0.0322
	300	SS-LASSO	-0.0005	0.0092	0.0009	0.0091	0.0004	0.0080	0.0183	0.0325
		SS-SCAD	-0.0004	0.0091	0.0007	0.0090	0.0001	0.0080	0.0173	0.0272
		SS-ALASSO	-0.0006	0.0091	-0.0010	0.0087	-0.00005	0.0079	0.0155	0.0246

SD = standard deviation; RASE = square root of the average squared error; SS = smoothing spline; LASSO = least absolute shrinkage and selection operator; SCAD = smoothly clipped absolute deviation; ALASSO = adaptive LASSO.

variables of α . The proposed method with the adaptive LASSO has similar performance in estimating α as that of SCAD in terms of bias and standard deviation; however, it showed better performance in correct model identification. Even in the high dimensional case, the estimation and selection performance of the proposed method with the adaptive LASSO showed superior to other methods. Table 2 also confirmed that the proposed method with the adaptive LASSO showed superior performance than others in terms of bias and SD of $\hat{\beta}$ and RASE, especially in the high dimensional case. The RASEs of the coefficient functions and the biases and SDs of $\hat{\alpha}$ and $\hat{\beta}$ decrease as n increased.

4. Real data analysis

We demonstrated the proposed method to the body fat dataset. The data contain 252 observations with 14 variables, in which the response variable is the percentage of body fat determined by the

Table 3: Results for the body fat data

Method	SS-ALASSO				CP	SIM-SCAD	LM-SCAD
	PLSIM	SIM	PLSICM	PLSICM2			
Age	0.0124	0.0000	0.0199	0.0095	0.0099	0.0149	0.0489
Weight	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1457
Height	0.0000	0.0000	0.0000	-0.0366	0.0000	0.0000	-0.0395
Neck	-0.1197	-0.1828	-0.1120	-0.1504	-0.0968	-0.1691	-0.1408
Chest	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	-0.0943
Abdomen	0.9663	0.9614	0.9804	0.9659	0.9689	0.9606	0.7663
Hip	-0.2128	-0.1516	0.1276	-0.2075	0.0000	0.0000	-0.3638
Thigh	0.0000	0.0924	0.0000	0.0000	0.0000	0.0000	0.1461
Knee	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Ankle	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Biceps	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Forearm	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0413
Wrist	-0.0807	-0.1032	-0.0900	-0.0346	-0.2278	-0.2202	-0.1186
R^2	0.6797	0.6818	0.6874	0.6815	0.6691	0.6738	0.6148

SS = smoothing spline; ALASSO = adaptive least absolute shrinkage and selection operator; SIM = single-index model; SCAD = smoothly clipped absolute deviation; LM-SCAD = linear model with SCAD; PLSIM = partially linear SIM; PLSICM = partially linear single-index-coefficient model.

underwater weighting technique. The covariates include age, weight, height, and 10 body circumference measurements (neck, chest, abdomen, hip, thigh, knee, ankle, biceps, forearm, and wrist). The dataset is available from the website (<http://lib.stat.cmu.edu/datasets/bodyfat>). After excluding 6 outliers similar to Peng and Huang (2011), we adopt several structured models, including PLSICM to identify the association between the percentage of body fat and other covariates, by selecting the index parameters. We considered the PLSIM, SIM, and PLSICM as,

$$\text{PLSIM} : Y = \eta_1 (\alpha^T X_{10}) + \beta_1 Z_4 + \epsilon,$$

$$\text{SIM} : Y = \eta_1 (\alpha^T X_{13}) + \epsilon,$$

$$\text{PLSICM} : Y = \eta_1 (\alpha^T X_{10}) Z_1 + \eta_2 (\alpha^T X_{10}) Z_3 + \beta_1 Z_4 + \epsilon,$$

where $Y = \log$ (percent body fat), X_{10} is a covariate matrix of the 10 body circumference measurements, and $X_{13}^T = (X_{10}^T, Z_2^T, Z_3^T, Z_4^T)$ with $Z_1 = 1$, $Z_2 = \text{weight}$, $Z_3 = \text{height}$, $Z_4 = \text{age}$. The PLSIM was considered in Feng and Xue (2015) with their combined penalization method and the SIM was fitted in Peng and Huang (2011) by using the Kernel method with SCAD penalty.

Table 3 showed the estimation results of the body fat data for each model. For comparison, the results of Feng and Xue (2015) (CP), Peng and Huang (2011) (SIM-SCAD), and a linear model with SCAD penalty (LM-SCAD) were also presented. Age was found to have a nonzero constant effect on the percentage of body fat. Among ten body circumference variables, neck, abdomen, hip, and wrist were selected by the proposed method. All models found that abdomen was the most important measurement for the prediction of the percentage of body fat. Previous results showed that the wrist was more important than hip and thigh circumferences; however, our models showed that hip was more important than the wrist to predict the percentage of body fat. Figure 1 showed the coefficient function estimates and its 95% Bayesian confidence intervals of the PLSICM. The estimated coefficient function of index of circumferences η_1 was nonlinear, which are consistent to the results of other literatures. The function estimate of η_2 was almost flat, which leads us to consider

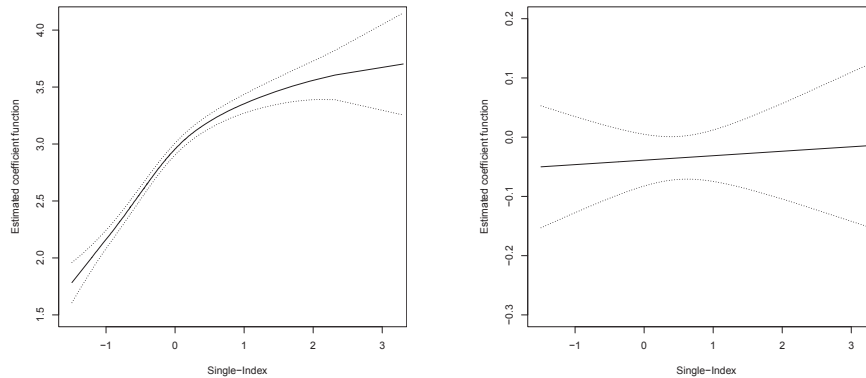


Figure 1: Varying coefficient function estimates in partially linear single-index-coefficient model of body fat data.

the following model (PLSICM2),

$$Y = \eta_1(\alpha^T X_{10})Z_1 + \beta_1 Z_4 + \beta_2 Z_3 + \epsilon,$$

and the fitted model showed consistent results to PLSICM with an additional result that the height was reversely related to the percentage of body fat at the cost of the reduced multiple R^2 .

5. Discussion

In this paper we proposed a simple nonparametric estimation method that employs smoothing splines to estimate varying-coefficient functions and select index parameters by shrinkage methods in PLSICM. This was based on the availability of reliable information that some predictors are linearly associated with the response; however, a single-index, a possible linear combination of predictors, is related to the response to different degrees according to the other predictors. The application of the splines method overcomes drawbacks suffered by high dimensional kernels. We therefore suggest a simple method based on a lower dimensional approximation in smoothing splines computation to estimate varying-coefficient functions as well as to select index parameters simultaneously using existing techniques. Simulation results have shown that the proposed method outperformed previously proposed methods. Future studies should investigate the parallel nonparametric estimation methods in partially linear single-index varying coefficient mixed effect models in the framework of smoothing spline regression.

Acknowledgements

This research was supported by 2017 Research Grant from Kangwon National University (No.52017 0503) and Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2015R1D1A3A01019998).

References

Fan J and Li R (2001). Variable selection via nonconcave penalized likelihood and its oracle properties, *Journal of the American Statistical Association*, **96**, 1348–1360.

- Feng S and Xue L (2015). Model detection and estimation for single-index varying coefficient model, *Journal of Multivariate Analysis*, **139**, 227–244.
- Foster JC, Taylor JMG, and Nan B (2013). Variable selection in monotone single-index models via the adaptive LASSO, *Statistical Medicine*, **32**, 3944–3954.
- Gu C (2013). *Smoothing Spline ANOVA Models* (2nd ed), Springer-Verlag.
- Gu C and Kim YJ (2002). Penalized likelihood regression: General formulation and efficient approximation, *Canadian Journal of Statistics*, **30**, 619–628.
- Huang Z (2012). Efficient inferences on the varying-coefficient single-index model with empirical likelihood, *Computational Statistics and Data Analysis*, **56**, 4413–4420.
- Huang Z, Lin B, Feng F, and Pang Z (2013). Efficient penalized estimating method in the partially varying-coefficient single-index model, *Journal of Multivariate Analysis*, **114**, 189–200.
- Huang Z, Pang Z, Lin B, and Shao Q (2014). Model structure selection in single-index-coefficient regression models, *Journal of Multivariate Analysis*, **125**, 159–175.
- Kim YJ and Gu C (2004). Smoothing spline Gaussian regression: more scalable computation via efficient approximation, *Journal of the Royal Statistical Society Series B*, **66**, 337–356.
- Leng C (2009). A simple approach for varying-coefficient model selection, *Journal of Statistical Planning and Inference*, **139**, 2138–2146.
- Peng H and Huang T (2011). Penalized least squares for single index models, *Journal of Statistical Planning and Inference*, **141**, 1362–1379.
- Tibshirani R (1996). Regression shrinkage and selection via the Lasso, *Journal of the Royal Statistical Society Series B*, **58**, 267–288.
- Wahba G (1983). Bayesian confidence interval for the cross-validated smoothing spline, *Journal of the Royal Statistical Society Series B*, **45**, 133–150.
- Xia Y and Li WK (1999). On single-index coefficient regression models, *Journal of the American Statistical Association*, **94**, 1275–1285.
- Xue LG and Wang QH (2012). Empirical likelihood for single-index varying-coefficient models, *Bernoulli*, **18**, 836–856.
- Yang H, Guo C, and Lv J (2014). A robust and efficient estimation method for single-index varying-coefficient models, *Statistics and Probability Letters*, **94**, 119–127.
- Yang H and Yang J (2014). The adaptive L1-penalized LAD regression for partially linear single-index models, *Journal of Statistical Planning and Inference*, **151**, 73–89.
- Zhu H, Lv Z, Yu K, and Deng C (2015). Robust variable selection in partially varying coefficient single-index model, *Journal of the Korean Statistical Society*, **44**, 45–57.
- Zou H (2006). The adaptive LASSO and its oracle properties, *Journal of the American Statistical Association*, **101**, 1418–1429.