

# Automatic Variable Selection for Single-Index Random Effects Models with Longitudinal Data

Suigen Yang<sup>1,2\*</sup>, Liugen Xue<sup>1</sup>

<sup>1</sup>College of Applied Sciences, Beijing University of Technology, Beijing, China

<sup>2</sup>College of Sciences, Tianjin University of Commerce, Tianjin, China

Email: [yangsuigen@163.com](mailto:yangsuigen@163.com)

Received 19 March 2014; revised 19 April 2014; accepted 26 April 2014

Copyright © 2014 by authors and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

---

## Abstract

We consider the problem of variable selection for the single-index random effects models with longitudinal data. An automatic variable selection procedure is developed using smooth-threshold. The proposed method shares some of the desired features of existing variable selection methods: the resulting estimator enjoys the oracle property; the proposed procedure avoids the convex optimization problem and is flexible and easy to implement. Moreover, we use the penalized weighted deviance criterion for a data-driven choice of the tuning parameters. Simulation studies are carried out to assess the performance of our method, and a real dataset is analyzed for further illustration.

## Keywords

Variable Selection; Single-Index Model; Random Effects; Longitudinal Data

---

## 1. Introduction

With the increasing availability of longitudinal data, both theoretical and applied works in longitudinal data analysis have become more popular in recent years. Diggle *et al.* [1] provided an excellent overview of the longitudinal data analysis. To avoid the so-called “curse of dimensionality” in the multivariate nonparametric regression with longitudinal data and to generate an association correlation structure between the repeated measurements, we consider the following single-index random effects models with longitudinal data,

$$Y_{ij} = g\left(X_{ij}^T \beta_0\right) + Z_{ij}^T b_i + \varepsilon_{ij}, i = 1, \dots, n, j = 1, \dots, m, \quad (1)$$

---

\*Corresponding author.

where  $\beta_0$  is a  $p \times 1$  index coefficients vector associated with the covariates  $X_{ij}$ ;  $b_i$  are independent  $q \times 1$  vectors of random effects with mean zero and covariance matrix  $\Phi$ ,  $g(\cdot)$  is an unknown link function;  $\varepsilon_{ij}$  are independent mean zero random variables with variance  $\sigma_\varepsilon^2 > 0$ . Here  $\Phi$  is a positive definite matrix depending on a parameter vector  $\phi$ ;  $X_{ij}$  and  $Y_{ij}$  are the observable random variables, and  $Z_{ij}$  are  $q \times 1$  known fixed design vectors. We suppose that  $b_i$  and  $\varepsilon_{ij}$  are mutually independent and follow gaussian distribution, and  $\|\beta_0\| = 1$  with the first nonzero element of  $\beta_0$  being positive to ensure identifiability. Pang and Xue [2] considered estimators of parameters and non-parameter for model (1). Yang *et al.* [3] considered simultaneous confidence band for the model (1).

Since the single-index models are popular and efficient modeling tools in multivariate nonparametric regression, the single-index models have recently received much attention, including those from Carroll *et al.* [4], Xia *et al.* [5], Zhu and Xue [6], Wang *et al.* [7], and among others. Pang and Xue [2], Yang *et al.* [3] and Chen *et al.* [8] considered the single-index models for longitudinal/panel data. Further, random effects models have become very popular for the analysis of longitudinal or panel data, because they are flexible and widely applicable. Given the importance of the random effects models, it is not surprising that methodologies for random effects models have emerged in the extensive literatures, such as Zeger and Diggle [9], Ke and Wang [10], Wu and Zhang [11] and Field *et al.* [12], and among others. However, it has a lot of challenges for the studies and the applications of single-index models with longitudinal data when the random effects in the models exist. Pang and Xue [2] proposed an iterative estimation procedure to estimate the index parameter vector and the link function, and they proved the asymptotic properties of the resulting estimators. However, in practical application, we do not know which covariates  $X$  have significant effects on the corresponding variable  $Y$ . In this paper, we consider the problem of variable selection for the single-index random effects models with longitudinal data.

Various penalty functions have been used in the variable selection literature for linear regression models. Frank and Friedman [13] considered the  $L_q$  penalty, which yields a ‘‘Bridge Regression’’. Tibshirani [14] proposed the Lasso, which can be viewed as a solution to the penalized least squares with the  $L_1$  penalty. Zou [15] further developed the adaptive lasso. Through combining both ridge ( $L_2$ ) and lasso ( $L_1$ ) penalty together, Zou and Hastie [16] proposed the Elastic-Net, which also has the sparsity property, to solve the collinearity problems. Fan and Li [17] proposed the SCAD penalty method and proved that the SCAD estimators enjoy the Oracle properties. All these variable selection procedures are based on penalized estimation using penalty functions, which have a singularity at zero. Consequently, these estimation procedures require convex optimization, which incurs a computational burden. To overcome this problem, Ueki [18] developed a new variable selection procedure called the smooth-threshold estimating equations that can automatically eliminate irrelevant parameters by setting them as zero. In addition, the resulting estimator enjoys the oracle property in the sense that Fan and Li [17] suggested. Li *et al.* [19] focus on marginal longitudinal generalized linear models and develop a variable selection technique.

Motivated by the idea of Ueki [18] and Li *et al.* [19], an automatic variable selection procedure is developed for the single-index random effects models. There are two difficulties. One notable difficulty in our setting is that we have to treat the nuisance parameters  $\Phi$  and  $\sigma_\varepsilon^2$  involved in the working covariance matrix, which affect the final estimator of  $\beta$ . Computationally, we need to update the values of these nuisance parameters together with the main parameter of interest. We propose an iterative algorithm to implement the procedures in Section 2 and obtain the efficient estimator of  $\beta$ . The proposed method shares some of the desired features of existing variable selection methods: the resulting estimator enjoys the oracle property; the proposed procedure avoids the convex optimization problem and is flexible and easy to implement. Moreover, we use the penalized weighted deviance criterion for a data-driven choice of the tuning parameters, see, Li *et al.* [19]. Simulation studies are carried out to assess the performance of our method, and a real dataset is analyzed for further illustration.

The paper is organized as follows. In Section 2, the iterative estimation procedure is given for model (1) and the asymptotic properties of the proposed estimator are established in Section 3. In Section 4 simulation studies are conducted to evaluate the performance of the proposed method, and a real data set is analyzed to illustrate the proposed method.

## 2. Estimation Procedure

Throughout this paper, let  $\beta_0$  be the fixed true value of  $\beta$  and let  $n \rightarrow \infty$ , while the  $m$  is uniformly bounded.

We partition  $\beta_0$  into active (nonzero) and inactive (zero) coefficients as follows: let  $A_0 = \{j: \beta_{0j} \neq 0\}$  and  $A_0^c = \{j: \beta_{0j} = 0\}$  be the complement of  $A_0$ . Denote by  $s = |A_0|$  the number of true zero parameters.

Suppose that the sample comes from model (1). Let  $Y_i = (Y_{i1}, \dots, Y_{im})^T$ ,  $X_i = (X_{i1}, \dots, X_{im})^T$ ,  $Z_i = (Z_{i1}, \dots, Z_{im})^T$ ,  $\varepsilon_i = (\varepsilon_{i1}, \dots, \varepsilon_{im})^T$  and  $G(X_i\beta_0) = (g(X_{i1}^T\beta_0), \dots, g(X_{im}^T\beta_0))^T$ . Model (1) can be rewritten as

$$Y_i = G(X_i\beta_0) + Z_i b_i + \varepsilon_i, i = 1, \dots, n. \tag{2}$$

It is easy to see that  $E(Y_i|X_i) = G(X_i\beta_0)$  and  $\text{Cov}(Y_i|X_i) = Z_i\Phi Z_i^T + \sigma_\varepsilon^2 I_m \equiv \Lambda_i$ , where  $I_m$  is the  $m \times m$  identity matrix and  $Z_i$  is  $m \times q$  known fixed design matrix. A naive idea to estimate  $\beta_0$  is to minimize

$$R_n(\beta) \equiv \frac{1}{n} \sum_{i=1}^n \{Y_i - G(X_i\beta)\}^T \Lambda_i^{-1} \{Y_i - G(X_i\beta)\} \tag{3}$$

Since  $\|\beta_0\| = 1$  means that the true value of  $\beta_0$  is the boundary point on the unit sphere,  $g(X_{ij}^T\beta_0)$  does not have derivative at the point  $\beta_0$ . However, we must use the derivative of  $g(X_{ij}^T\beta_0)$  on  $\beta_0$ , when constructing the estimating equation for  $\beta_0$ . The “delete-one-component” method (see Zhu and Xue [6], Wang et al. [7]) is used to solve this problem. The detail is as follows. Let  $\beta_0 = (\beta_1, \dots, \beta_p)^T$  and  $\beta^{(r)} = (\beta_1, \dots, \beta_{r-1}, \beta_{r+1}, \dots, \beta_p)^T$  be a  $p-1$  dimensional parameters vector deleting the  $r$ th component  $\beta_r$ . Without loss of generality, we may assume that the true vector  $\beta_0$  has a positive component  $\beta_r$ . Then, the true parameter  $\beta^{(r)}$  satisfies the constraint  $\|\beta^{(r)}\| < 1$ . Thus,  $\beta_0$  is infinitely differentiable in a neighborhood of the true parameter  $\beta^{(r)}$ , the Jacobian matrix is

$$J_{\beta^{(r)}} = (\gamma_1, \dots, \gamma_p)^T,$$

where  $\gamma_s (1 \leq s \leq p, s \neq r)$  is a  $(p-1)$  dimensional unit vector with  $s$  component 1, and

$$\gamma_r = -(1 - \|\beta^{(r)}\|)^{-1/2} \beta^{(r)}.$$

Based on the estimation procedure in Pang and Xue [2] and Yang et al. [3], we outline the iterative steps for estimating procedures for  $\beta_0$ ,  $g(\cdot)$  and its derivative  $g'(\cdot)$ .

**Step 0:** We first give a consistent estimator of  $\beta_0$ , which is denoted by  $\tilde{\beta}$ .

**Step 1:** Estimation of the link function  $g(\cdot)$  and its derivative  $g'(\cdot)$ . Given the initial estimator  $\tilde{\beta}$ , we apply the local linear regression technique in Fan and Gijbels [20] to estimate the link function and its derivative. The estimators of  $g(\cdot)$  and  $g'(\cdot)$  are obtained by minimizing the weighted sum of squares

$$\sum_{i=1}^n \sum_{j=1}^m \{Y_{ij} - a - b(X_{ij}^T \tilde{\beta} - u)\}^2 K_h(X_{ij}^T \tilde{\beta} - u)$$

with respect to  $a$  and  $b$  where  $K_h(\cdot) = h^{-1}K(\cdot/h)$ ,  $K(\cdot)$  be a kernel function, and  $h = h_n$  is the bandwidth. Specifically, the local linear estimators of  $g(\cdot)$  and  $g'(\cdot)$  are defined as  $\hat{g}(u; \tilde{\beta}) = \hat{a}$  and  $\hat{g}'(u; \tilde{\beta}) = \hat{b}$  for the initial estimator  $\tilde{\beta}$ . By some simple calculations, we have

$$\left(\hat{g}(u; \tilde{\beta}), h\hat{g}'(u; \tilde{\beta})\right)^T = S_n(u; \tilde{\beta})^{-1} V_n(u; \tilde{\beta}),$$

where  $S_n(u; \tilde{\beta}) = \begin{pmatrix} S_{n,0}(u; \tilde{\beta}) & S_{n,1}(u; \tilde{\beta}) \\ S_{n,1}(u; \tilde{\beta}) & S_{n,2}(u; \tilde{\beta}) \end{pmatrix}$  and  $V_n(u; \tilde{\beta}) = (V_{n,0}(u; \tilde{\beta}), V_{n,1}(u; \tilde{\beta}))^T$  with

$$S_{n,l}(u; \tilde{\beta}) = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^m \left(\frac{X_{ij}^T \tilde{\beta} - u}{h}\right)^l K_h(X_{ij}^T \tilde{\beta} - u), l = 0, 1, 2,$$

$$V_{n,l}(u; \tilde{\beta}) = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^m \left(\frac{X_{ij}^T \tilde{\beta} - u}{h}\right)^l K_h(X_{ij}^T \tilde{\beta} - u) Y_{ij}, l = 0, 1.$$

**Step 2:** Estimation of the variance components. To obtain the estimator of index parameter, we need to get the consistent estimators of the variance components. Suppose that the variance-covariance matrix for model (2)

is

$$\Lambda = \sigma_b^2 J_m J_m^T + \sigma_\varepsilon^2 I_m,$$

where  $J_m = (1, \dots, 1)^T$  is the  $m$ -vector of ones. Assume that the random effects  $b_i$  and the error  $\varepsilon_{ij}$  are Gaussian distributed, then the observation  $Y_i$  have independent  $N(G(X_i \beta_0), \Lambda)$  distributions. Based on the estimator  $\tilde{\beta}$  and the estimator  $\hat{g}(u, \tilde{\beta})$ , the log-likelihood function for  $\sigma_\varepsilon^2$  and  $\sigma_b^2$  can be written as

$$-n(m-1)\log(\sigma_\varepsilon^2) - n\log(\sigma_\varepsilon^2 + m\sigma_b^2) - \frac{m}{\sigma_\varepsilon^2 + m\sigma_b^2} \sum_{i=1}^n (\bar{Y}_i - \bar{\hat{g}}_i)^2 - \frac{1}{\sigma_\varepsilon^2} \sum_{i=1}^n \sum_{j=1}^m \left\{ Y_{ij} - \hat{g}(X_{ij}^T \tilde{\beta}; \tilde{\beta}) - (\bar{Y}_i - \bar{\hat{g}}_i) \right\}^2,$$

where  $\bar{Y}_i = m^{-1} \sum_{j=1}^m Y_{ij}$  and  $\bar{\hat{g}}_i = m^{-1} \sum_{j=1}^m \hat{g}(X_{ij}^T \tilde{\beta}; \tilde{\beta})$ . The maximum likelihood estimators of  $\sigma_\varepsilon^2$  and  $\sigma_b^2$  are defined by

$$\hat{\sigma}_\varepsilon^2(\tilde{\beta}) = \frac{1}{n(m-1)} \sum_{i=1}^n \sum_{j=1}^m \left\{ Y_{ij} - \hat{g}(X_{ij}^T \tilde{\beta}; \tilde{\beta}) - (\bar{Y}_i - \bar{\hat{g}}_i) \right\}^2,$$

$$\hat{\sigma}_b^2(\tilde{\beta}) = \frac{1}{n} \sum_{i=1}^n (\bar{Y}_i - \bar{\hat{g}}_i)^2 - \frac{1}{m} \hat{\sigma}_\varepsilon^2(\tilde{\beta}).$$

Then we can obtain the  $\Lambda$ 's estimator  $\hat{\Lambda}(\tilde{\beta}) = \hat{\sigma}_b^2(\tilde{\beta}) J_m J_m^T + \hat{\sigma}_\varepsilon^2(\tilde{\beta}) I_m$ .

**Step 3:** Estimation of index parameter. Based on the initial estimator  $\tilde{\beta}$  and the estimator of  $\Lambda_i$ , the estimator  $\tilde{\beta}^{(r)}$  of  $\beta^{(r)}$  can be obtained by solving the following estimating equation

$$Q_n(\hat{G}, \beta^{(r)}) = \frac{1}{n} \sum_{i=1}^n J_{\beta^{(r)}}^T X_i^T \hat{G}'_\Delta(X_i \beta_0) \hat{\Lambda}_i^{-1}(\tilde{\beta}) \{Y_i - \hat{G}(X_i \beta_0)\} = 0,$$

where  $\hat{G}'_\Delta(X_i \beta_0) = \text{diag}\{\hat{g}'(X_{i1}^T \beta_0; \beta_0), \dots, \hat{g}'(X_{im}^T \beta_0; \beta_0)\}$ .

Motivated by the idea of Ueki [18] and Li *et al.* [19], we can use the following smooth-threshold estimating equations to estimate  $\beta^{(r)}$ ,

$$Q_n^*(\hat{G}, \beta^{(r)}) = (I_{(p-1)} - \Delta) Q_n(\hat{G}, \beta^{(r)}) + \Delta \beta = 0 \quad (3)$$

where  $\Delta$  is the diagonal matrix whose diagonal elements are  $(\delta_i)_{i=1, \dots, (p-1)}$ , and  $I_{(p-1)}$  is the  $p-1$  dimensional identity matrix. Note that  $\delta_i = 1$  reduces to  $\beta_i = 0$ . Therefore equation (3) can yield a sparse solution. Unfortunately, we cannot direct obtain the estimator of  $\beta^{(r)}$  by solving (3), because (3) involves  $\delta_i$ , which need be chosen using some data driven criteria. For the choice of  $\delta_i$ , Ueki [18] suggested that  $\delta_i$  can be chosen by  $\hat{\delta}_i = \min\left(1, \lambda / |\tilde{\beta}_i|^{1+\gamma}\right)$ , where  $(\lambda, \gamma)$  are two tuning parameters, which can be computed by a penalized weighted deviance criterion, see Li *et al.* [19]. Similarly, we can define the active set  $A = \{i: \hat{\delta}_i \neq 1\}$  which is the set of indices of nonzero parameters. Replacing  $\Delta$  in Equation (3) by  $\hat{\Delta}$  with diagonal elements  $\hat{\delta}_i, i = 1, \dots, p-1$ , we propose the following modified iterative procedure for  $\beta_A$ ,

$$\hat{\beta}_A^* = \tilde{\beta}_A + J_{\tilde{\beta}_A} B_n^{*-1}(\hat{G}, \tilde{\beta}_A^{(r)}) Q_n^*(\hat{G}, \tilde{\beta}_A^{(r)}), \quad (4)$$

where  $B_n^*(\hat{G}, \tilde{\beta}_A^{(r)}) = (I_{(p-1)} - \hat{\Delta}) B_n(\hat{G}, \tilde{\beta}_A^{(r)}) (I_{(p-1)} - \hat{\Delta}) + \hat{\Delta}$ . Reset  $\tilde{\beta}_A = \hat{\beta}_A^* / \|\hat{\beta}_A^*\|$ . Repeat (4) until convergence. We denote the final estimator of  $\beta_A$  by  $\hat{\beta}_A$ .

**Step 4:** Repeat step 1 to step 3 until convergence. Finally, instead  $\tilde{\beta}$  with  $\hat{\beta}_A$  in  $\hat{g}(u; \tilde{\beta})$ , we obtain the final estimators of  $g(\cdot)$  and  $g'(\cdot)$ , which is denoted by  $\hat{g}(u) = \hat{g}(u; \hat{\beta})$  and  $\hat{g}'(u) = \hat{g}'(u; \hat{\beta})$ , respectively.

**Remark 1:** In Step 0, we need to choose a suitable initial estimator of  $\beta_0$ . For the numerical studies and real data analysis in Section 4, the initial estimator can be obtained using two steps. In the first step, we use independent data  $\{(Y_{ik}, X_{ik}, Z_{ik}), i = 1, \dots, n\}$  to get estimators  $\tilde{\beta}_{(k)}$  for  $k = 1, \dots, m$ . In the second step, we average  $\tilde{\beta}_{(k)}$  over  $k = 1, \dots, m$ , then  $\tilde{\beta} = m^{-1} \sum_{k=1}^m \tilde{\beta}_{(k)}$  is taken as the initial estimator.

**Remark 2:** It is well-known that the convergence rate of the estimator  $\hat{g}'(u)$  is slower than that of the es-

timator  $\hat{g}(u)$  if the same bandwidth is used. This leads to a slower convergence rate than root- $n$  for the estimator  $\hat{\beta}$  of  $\beta_0$ . This motivates us to introduce another bandwidth  $h_1$  to control the variability of the estimator of  $g'(u)$ .

**Remark 3:** We use following penalized weighted deviance criterion (see Li *et al.* [19]) to select tuning parameters  $(\lambda, \gamma)$ :

$$PWD_{(\lambda, \gamma)} = WSSE + DF_{(\lambda, \gamma)} \log(n)$$

where  $DF_{(\lambda, \gamma)} = \sum_{i=1}^p 1(\hat{\delta}_i \neq 1)$  denotes the number of nonzero parameters with  $1(\cdot)$  the indicator function,

$WSSE = \sum_{i=1}^n r_i^T \Lambda_i^{-1} r_i$  with the deviance residual  $r_i = Y_i - g(X_i \tilde{\beta})$ . We can choose  $(\lambda, \gamma)$  by minimizing the  $PWD_{(\lambda, \gamma)}$ .

### 3. Asymptotic Properties

In this section, we assume, under the regularity conditions, the initial estimator using the full model is consistent and asymptotically normally distributed by solving the GEE (see Liang and Zeger [21]). Following Fan and Li [17], it is possible to prove the oracle properties for the estimators, including  $\sqrt{n}$ -consistency, variable selection consistency, and asymptotic normality.

**Theorem 1.** Under mild regularity conditions, for any positive  $\lambda$  and  $\gamma$  such that  $n^{1/2}\lambda \rightarrow 0$  and  $n^{(1+\gamma)/2}\lambda \rightarrow 0$ , there exists a sequence  $\hat{\beta}_{\lambda, \gamma}$  of the solutions of (3) such that

$$\|\hat{\beta}_{\lambda, \gamma} - \beta_0 = O_p(n^{-1/2})\|.$$

**Note 1:** The mild regularity conditions in Theorem 1 are same with the conditions in Yang *et al.* [3] Theorem 1.

**Theorem 2.** Suppose that the conditions of Theorem 1 hold, as  $n \rightarrow \infty$ , we have

- 1) variable selection consistency, *i.e.*  $P(A = A_0) \rightarrow 1$
- 2) asymptotic normality, *i.e.*

$$\sqrt{n}(\hat{\beta}_{\lambda, \gamma, A_0} - \beta_{A_0}) \xrightarrow{D} N\left(0, J_{\beta_{A_0}^{(r)}} \Omega^{-1} J_{\beta_{A_0}^{(r)}}^T\right)$$

where  $\Omega$  is the limit in probability of

$$\frac{1}{n} \left[ J_{\beta_{A_0}^{(r)}}^T \left\{ X_i - G_1(X_i, \beta_{A_0}) \right\}^T G'^2 \Lambda_i^{-2} \left\{ X_i - G_1(X_i, \beta_{A_0}) \right\} J_{\beta_{A_0}^{(r)}} \right]$$

as  $n \rightarrow \infty$ .

The proof of Theorem 1 and Theorem 2 can be obtained similarly to the proof of Theorem 1 and Theorem 2 in Li *et al.* [19].

## 4. Numerical Studies and Application

### 4.1. Numerical Studies

In this subsection, we conduct simulation studies to illustrate the finite sample properties of proposed procedure. Throughout the simulation studies, we take Epanechnikov kern  $K(t) = 0.75(1-t^2)_+$  for estimating the link function, and the bandwidth  $h$  is chosen by the cross validation (CV) method.

For each case we repeat the experiment 100 times and applied the penalized weighted deviance criterion to select the tuning parameters. We consider the following example.

For a single-index random effects model,

$$Y_{ij} = 5(X_{ij}^T \beta_0)^2 + b_i + \sigma_\varepsilon \varepsilon_{ij}, i = 1, \dots, n, j = 1, \dots, m, \tag{5}$$

where  $\beta_0 = (\sqrt{3}/2, 0, 0.5, 0, 0)^T$ ,  $X_{ij}$  is a five-dimensional vector with independent uniform  $[0,1]$ ,  $b_i$  is a normal variable with mean zero and variance  $\sigma_b^2$ ,  $\varepsilon_{ij}$  is a standard normal variable.

For the simulations, we consider the number of subjects  $n = 50, 100$  subjects and  $m = 3$ . For comparison, we consider  $\sigma_\varepsilon = 0.5, 1$ , and  $\sigma_b = 1, 1.5$ , respectively. Based on the experiment time  $M = 100$ , the simulation results are reported in **Table 1**. In the tables, values in the column labeled “Correct” denote the average number of coefficients of the true zeros, correctly set to zero, and those in the column labeled “Incorrect” denote the average number of the true nonzeros incorrectly set to zero.

**Table 1** and **Figure 1** indicate the following simulation results:

- 1) From **Table 1**, it is easy to see that “correct” increases to 3 (true number) as  $n$  increases. Therefore, the proposed method is able to correctly identify the true submodel.
- 2) From **Table 1**, we find that “correct” increases to 3 as  $\sigma_b$  and the  $\sigma_\varepsilon$  decrease, respectively.
- 3) **Figure 1** shows that the estimators of  $\beta$  have asymptotic normality.

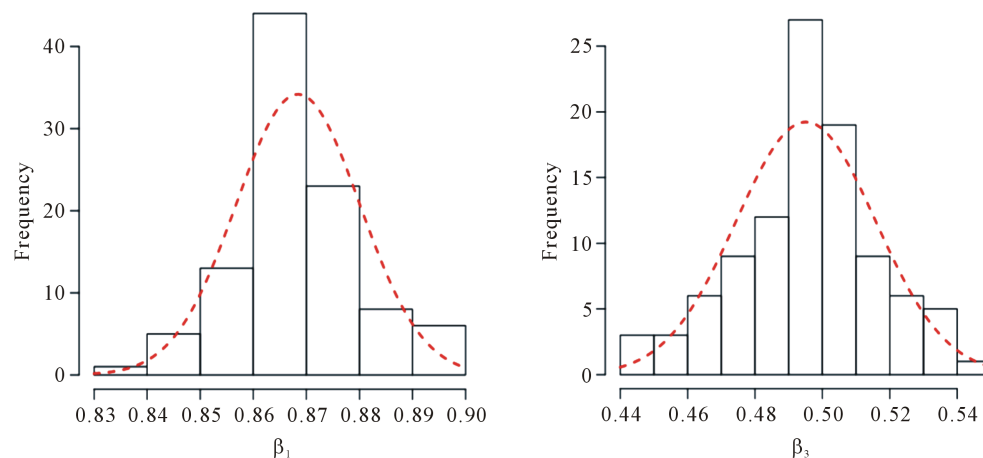
## 4.2. Application to Real Data

The data set comes from an epileptic study (Thall and Vail [22], Bai *et al.* [23], and Pang and Xue [2]). Two different treatments (placebo and antiepileptic drug progabide) were administered to 59 epileptics during the experimental period. Patients were randomized to receive either of the two treatments. The patients attended clinic visits every two weeks for four consecutive times and the number of seizures occurring over the previous two weeks was reported. For this dataset, the number of seizures in a two-week period (NS) is taken as the response variable, the logarithm of age in year (LA), and the baseline seizure count (which is divided by 4 and then log-transformed, let BSC) are considered as the covariates. A scientific question here is whether the drug helps to reduce the rate of epileptic seizures. To illustrate the proposed method, we consider the following single-index model,

$$NS_{ij} = g(\beta_1 LA_{ij} + \beta_2 BSC_{ij} + \beta_3 X_{1ij} + \beta_4 X_{2ij} + \beta_5 X_{3ij}) + \varepsilon_{ij}, i = 1, \dots, 59, j = 1, \dots, 4. \quad (6)$$

**Table 1.** Variable selections for model (5) using our method.

$\sigma_b$	$\sigma_\varepsilon$	$n = 50$		$n = 100$	
		Correct	Incorrect	Correct	Incorrect
1	0.5	2.96	0	2.99	0
	1	2.88	0	2.94	0
1.5	0.5	2.82	0	2.99	0
	1	2.71	0	2.95	0



**Figure 1.** The hist plot about  $\beta_1$  (left) and  $\beta_3$  (right).

where  $\beta_3 = \beta_4 = \beta_5 = 0$ , and  $X_{kij}$ ,  $k=1,2,3$  are random number with independent uniform  $[0,1]$ . The estimation procedure proposed in Section 2 is used to estimate the single-index model (6), the non zero estimated of the index coefficients (standard error of estimated) are  $\hat{\beta}_1 = 0.8342$  (0.0563),  $\hat{\beta}_2 = -0.5515$  (0.1287), and  $\hat{\beta}_3 = \hat{\beta}_4 = \hat{\beta}_5 = 0$ . Therefore, the proposed method is feasible in practical application.

## 5. Concluding Remarks

In this paper, we have done automatic variable select to parameters of index  $\beta$  for single-index random effects model with longitudinal data. We further derive the asymptotic distributions for estimator of  $\beta$  for single-index random effects model. The proposed estimator has good asymptotic behavior and select number of zero parameters very close to the nominal level in our simulation study. A real data analysis illustrates the practical use of the variable select. The methodology in this paper is general and widely applicable, and therefore, we expect further research along these lines to yield deep theoretical results with interesting applications for other nonparametric or semiparametric models with random effects.

## Acknowledgements

Liugen Xue's research was supported by the National Natural Science Foundation of China (11171012), the Science and Technology Project for the Supervisor of Excellent Doctoral Dissertation of Beijing (20111000503), the Specialized Research Fund for the Doctoral Program of Higher Education of China (20121103110004) and National Natural Science Foundation of China (11331011).

Suigen Yang's research was supported by the NNSF (11101014, 11002005) of China, the Specialized Research Fund for the Doctoral Program of Higher Education of China (20101103120016), PHR (IHLB, PHR20110822), the Training Programme Foundation for the Beijing Municipal Excellent Talents (2010D005015000002) and the Fundamental Research Foundation of Beijing University of Technology (X4006013201101).

## References

- [1] Diggle, P.J., Heagerty, P., Liang, K. and Zeger, S.L. (2002) Analysis of Longitudinal Data. 2nd Edition, Oxford University, Oxford.
- [2] Pang, Z. and Xue, L.G. (2012) Estimation for the Single-Index Models with Random Effects. *Computational Statistics and Data Analysis*, **56**, 1837-1853. <http://dx.doi.org/10.1016/j.csda.2011.11.007>
- [3] Yang, S.G., Xue, L.G. and Li, G.R. (2014) Simultaneous Confidence Band for Single-Index Random Effects Models with Longitudinal Data. *Statistics and Probability Letters*, **85**, 6-14. <http://dx.doi.org/10.1016/j.spl.2013.10.014>
- [4] Carroll, R.J., Fan, J., Gijbels, I. and Wand, M.P. (1998) Generalized Partially Linear Single-Index Models. *Journal of the American Statistical Association*, **92**, 477-489. <http://dx.doi.org/10.1080/01621459.1997.10474001>
- [5] Xia, Y., Li, W.K., Tong, H. and Zhang, D. (2004) A Goodness-Of-Fit Test for Single-Index Models. *Statistica Sinica*, **14**, 1-39.
- [6] Zhu, L.X. and Xue, L.G. (2006) Empirical Likelihood Confidence Regions in a Partially Linear Single-Index Model. *Journal of the Royal Statistical Society: Series B*, **68**, 549-570. <http://dx.doi.org/10.1111/j.1467-9868.2006.00556.x>
- [7] Wang, J.L., Xue, L.G., Zhu, L.X. and Chong, Y. (2010) Estimation for a Partial-Linear Single-Index Model. *Annals of Statistics*, **38**, 246-274.
- [8] Chen, J., Gao, J. and Li, D. (2013) Estimation in a Single-Index Panel Data Models with Heterogeneous Link Functions. *Econometric Reviews*, **33**, 928-955. <http://dx.doi.org/10.1080/07474938.2012.690687>
- [9] Zeger, S.L. and Diggle, P.J. (1994) Semiparametric Models for Longitudinal Data with Application to CD4 Cell Numbers in HIV Seroconverters. *Biometrics*, **50**, 689-699. <http://dx.doi.org/10.2307/2532783>
- [10] Ke, C.L. and Wang, Y.D. (2001) Semiparametric Nonlinear Mixed-Effects Models and Their Applications (with Discussion). *Journal of the American Statistical Association*, **96**, 1272-1298. <http://dx.doi.org/10.1198/016214501753381913>
- [11] Wu, H.L. and Zhang, J.T. (2002). Local Polynomial Mixed-Effects Models for Longitudinal Data. *Journal of the American Statistical Association*, **97**, 883-897. <http://dx.doi.org/10.1198/016214502388618672>
- [12] Field, C.A., Pang, Z. and Welsh, A.H. (2008) Bootstrapping Data with Multiple Levels of Variation. *Canadian Journal of Statistics*, **36**, 521-539. <http://dx.doi.org/10.1002/cjs.5550360403>



- [13] Frank, I.E. and Friedman, J.H. (1993) A Statistical View of Some Chemometrics Regression Tools. *Technometrics*, **35**, 109-148. <http://dx.doi.org/10.1080/00401706.1993.10485033>
- [14] Tibshirani, R. (1996) Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society: Series B*, **58**, 267-288.
- [15] Zou, H. and Hastie, T. (2005) Regularization and Variable Selection via the Elastic Net. *Journal of the Royal Statistical Society: Series B*, **67**, 301-320. <http://dx.doi.org/10.1111/j.1467-9868.2005.00503.x>
- [16] Zou, H. (2006) The Adaptive Lasso and Its Oracle Properties. *Journal of the American Statistical Association*, **101**, 1418-1429. <http://dx.doi.org/10.1198/016214506000000735>
- [17] Fan, J.Q. and Li, R.Z. (2001) Variable Selection via Nonconcave Penalized Likelihood and Its Oracle Properties. *Journal of the American Statistical Association*, **96**, 1348-1360. <http://dx.doi.org/10.1198/016214501753382273>
- [18] Ueki, M. (2009) A Note on Automatic Variable Selection Using Smooth-Threshold Estimating Equations. *Biometrika*, **96**, 1005-1011. <http://dx.doi.org/10.1093/biomet/asp060>
- [19] Li, G.R., Lian, H., Feng, S.Y. and Zhu, L.X. (2013) Automatic Variable Selection for Longitudinal Generalized Linear Models. *Computational Statistics and Data Analysis*, **61**, 174-186. <http://dx.doi.org/10.1016/j.csda.2012.12.015>
- [20] Fan, J. and Gijbels, I. (1996) Local Polynomial Modeling and Its Applications. Chapman & Hall, London.
- [21] Liang, K.Y. and Zeger, S.L. (1986) Longitudinal Data Analysis Using Generalized Linear Models. *Biometrika*, **73**, 13-22. <http://dx.doi.org/10.1093/biomet/73.1.13>
- [22] Thall, P. and Vail, S.C. (1990) Some Covariance Models for Longitudinal Count Data with Over Dispersion. *Biometrics*, **46**, 657-671. <http://dx.doi.org/10.2307/2532086>
- [23] Bai, Y., Fung, W.K. and Zhu, Z.Y. (2009) Penalized Quadratic Inference Functions for Single-Index Models with Longitudinal Data. *Journal of Multivariate Analysis*, **100**, 152-161. <http://dx.doi.org/10.1016/j.jmva.2008.04.004>