Scientific
Research
Publishing

Efficient Shrinkage Estimation about the Partially Linear Varying Coefficient Model with Random Effect for Longitudinal Data

Wanbin Li

School of Mathematics and Statistics, Yancheng Teachers University, Yancheng, China
Email: liwanbinyc@163.com

Abstract

In this paper, an efficient shrinkage estimation procedure for the partially linear varying coefficient model (PLVC) with random effect is considered. By selecting the significant variable and estimating the nonzero coefficient, the model structure specification is accomplished by introducing a novel penalized estimating equation. Under some mild conditions, the asymptotic properties for the proposed model selection and estimation results, such as the sparsity and oracle property, are established. Some numerical simulation studies and a real data analysis are presented to examine the finite sample performance of the procedure.

1. Introduction

With the effort to reduce the risk of model misspecification, more flexible nonlinear and non/semiparametric models have been proposed for independent and subject-dependent data. See, for example, [1] introduced the varying coefficient model. Reference [2] and [3] studied the partially linear varying coefficient (PLVC) model and single index model, respectively, for longitudinal data. For longitudinal data, see [4] for an intensive review.

As a natural extension of [5], which used marginal model for longitudinal data analysis, a random effect method is developed when considering within-subject correlation and further shrinkage estimation. In the literatures in longitudinal data study, random effect method has received relatively enough attention. See, for example, [6]-[9] and so

on. Some advantages of random effect method were mentioned in [3] including its computation efficiency.

For the special case of PLVC model with random effect, an important problem is to choose the significant covariant variable. Shrinkage estimation based on regularization has attracted lots of interest. See, for example, [10]-[13]. Also, some extensions of the variable selection under the regularization framework to varying coefficient models include, [5] [14]-[16]. In this article, the variable selection problem for PLVC model with random effect is investigated. Because of the obvious simplicity and wide usage, a penalized estimating equation based shrinkage estimation procedure is introduced, following the idea of [17].

The rest of this article is organized as follows. In Section 2, the model, estimation procedure and statistical properties of the estimators are introduced. In Section 3, the practical computational issues are discussed and some numeric simulations and a real data analysis for the finite sample performance are illustrated in Section 4.

2. Estimation and Asymptotic Property

2.1. Penalized Estimating Equation

Let $(X_i, Z_i, W_i, U_i, Y_i)$ be the observed data associated with the ith subject in a longitudinal study. We consider the partially linear varying coefficient mixed effect model (PLVCMeM)

$$Y_{ij} = X_{ij}^{\mathrm{T}} \beta + Z_{ij}^{\mathrm{T}} \theta(U_{ij}) + W_{ij}^{\mathrm{T}} v_i + \varepsilon_{ij}, i = 1, \cdots, n; j = 1, \cdots, m, \tag{1}$$

where β is a $q \times 1$ coefficient vector, v_i is a $l \times 1$ vector of random effect with mean 0 and covariance matrix D, $\theta(\cdot)$ is an unknown smoothing $p \times 1$ function vector and ε_{ij} is a random variable with mean 0 and variance $\sigma^2 > 0$.

Assume that $b(u) = \left(b_1(u), \cdots, b_{k_n+h+1}(u)\right)^{\mathrm{T}}$ is a set of B-spline basis functions of order $h+1$ with k_n quasi-uniform internal knots. Then, each $\theta_i(u)$ can be expressed with a linear combination of normalized B-spline basis function $\theta_l(u) = \sum_{s=1}^{k_n+h+1} b_s(u) \gamma_{l,s} = b^{\mathrm{T}}(u) \gamma_l$, where $\gamma_l = \left(\gamma_{l,1}, \cdots \gamma_{l,k_n+h+1}\right)^{\mathrm{T}}$ is the spline coefficient vector. Therefore, with the given spline basis $b(u)$, model (1) can be approximated as

$$Y_{ij} \approx \sum_{l=1}^{p} \sum_{s=1}^{k_n+h+1} Z_{ij,l} b_s(U_{ij}) \gamma_{l,s} + \sum_{s=1}^{q} X_{ij,s}^{\mathrm{T}} \beta_s + W_{ij}^{\mathrm{T}} v_i + \varepsilon_{ij} \approx B_{ij}^{\mathrm{T}} \gamma + X_{ij}^{\mathrm{T}} \beta + W_{ij}^{\mathrm{T}} v_i + \varepsilon_{ij}, \tag{2}$$

where $B_{ij}^{\mathrm{T}} = Z_{ij}^{\mathrm{T}} \cdot I_p \otimes b^{\mathrm{T}}(u)$, $B_i = \left(B_{i1}, \cdots, B_{im}\right)^{\mathrm{T}}$, $\gamma = \left(\gamma_{l,s}\right) \in R^{p_{k_n}}$, $p_{k_n} = p(k_n + h + 1)$. Hence, the estimators for β and γ can be derived with the following weighted least square equation

$$Q(\gamma, \beta) = \frac{1}{n} \sum_{l=1}^{p} \left\{Y_i - B_i\gamma - X_i\beta\right\}^{\mathrm{T}} \Omega_i^{-1} \left\{Y_i - B_i\gamma - X_i\beta\right\}, \tag{3}$$

where $\Omega_i = W_i^{\mathrm{T}} D W_i + \sigma_\varepsilon^2 I_m$ is the variance-covariance matrix.

A primary goal for (1) is to explore useful information for Z and X, it is important to select and estimate the nonzero coefficients in β and nonzero functions in $\theta(u)$ to

enhance model prediction and interpretability. To achieve the simultaneous estimation and variable selection, a shrinkage penalty function is developed based on estimating equation with a regularization based shrinkage estimation being defined with the following penalized estimating equation, which is similarly introduced in [17] as

$$Q(\gamma, \beta) = \frac{1}{n} \sum_{i=1}^{n} \tilde{Z}_i^{\mathrm{T}} \Omega_i^{-1} \{Y_i - B_i \gamma - X_i \beta\} + n \sum_{k=1}^{p} q_{\lambda_{1k}} \left(\left\| B(\cdot)^{\mathrm{T}} \gamma_k \right\| \right) + n \sum_{l=1}^{q} q_{\lambda_{2l}} \left(\left\| \beta_l \right\| \right), \quad (4)$$

where $\tilde{Z}_i = (B_i, X_i)$. Alternatively, when denoting $\alpha = (\gamma^{\mathrm{T}}, \beta^{\mathrm{T}})^{\mathrm{T}}$, it follows that

$$U^P(\alpha) = \sum_{i=1}^{n} \tilde{Z}_i^{\mathrm{T}} \Omega_i^{-1} \{Y_i - \tilde{Z}_i \alpha\} + n q_\lambda \|\alpha\|, \quad (5)$$

where q_λ is a penalty function used to penalize the coefficients of nonsignificant variables, λ is a tuning parameter.

Among many choices of the penalty function q_λ, the smoothly clipped and deviations (SCAD) penalty function is adopted as

$$q_\lambda(x) = p'_\lambda(x) = \lambda I(x \le \lambda) + \frac{(a\lambda - x)_+}{(a-1)} I(x > \lambda), \quad (6)$$

where $x > 0$ and $a = 3.7$, which is suggested by [10]. Since the SCAD penalty is a spline function on an interval near zero and constant outside, it can shrink small value of an estimation to zero while having no impact on the large one.

2.2. Estimators of the Variance Components

An efficient estimation for the parameters of interest in model (5) depends on estimators for the variance component, therefore a consistent estimators for them is required. Suppose that the variance covariance matrix for model (1) is

$$\Omega_i = \sigma_v^2 1_m 1_m^{\mathrm{T}} + \sigma_\varepsilon^2 I_m, \quad (7)$$

where 1_m represents a $m \times 1$ vector of ones and I_m is the $m \times m$ identity matrix. Model (7) is called a variance component model and is very popular in longitudinal analysis. See, for example, [3] [6] [9] and so on. Furthermore, let $\varepsilon_i = (\varepsilon_{i1}, \cdots, \varepsilon_{im})$ $(i = 1, \cdots, n)$ with $\varepsilon_{ij} = v_i + \varepsilon_{ij}$. Hence, it follows that a moment estimator for variance component $(\sigma_v^2, \sigma_\varepsilon^2)$ of Ω_i is derived as

$$\hat{\sigma}_v^2 = \frac{1}{nm(nm-1)} \sum_{i=1}^{n} \sum_{j_1=1}^{m} \sum_{j_1 \ne j_2} \hat{\varepsilon}_{ij_1} \hat{\varepsilon}_{ij_2}, \quad \hat{\sigma}_v^2 = \frac{1}{nm} \sum_{i=1}^{n} \sum_{j_1=1}^{m} \hat{\varepsilon}_{ij_1}^2 - \hat{\sigma}_v^2. \quad (8)$$

where by the estimator $\tilde{\beta}$ and $\tilde{\theta}(u)$ obtained from (3) without the weight Ω_i, the residual $\hat{\varepsilon}_{ij}$ is defined as

$$\hat{\varepsilon}_{ij} = Y_{ij} - X_{ij}^{\mathrm{T}} \tilde{\beta} - Z_{ij}^{\mathrm{T}} \tilde{\theta}(U_{ij}), \quad (9)$$

Therefore, the estimator for Ω_i is defined as

$$\hat{\Omega}_i = \hat{\sigma}_v^2 1_m 1_m^{\mathrm{T}} + \hat{\sigma}_\varepsilon^2 I_m. \quad (10)$$

2.3. Asymptotic Properties about the Estimators

In this section, we investigate the asymptotic behavior of the estimators for the para-

metric, nonparametric and variance component as well. Throughout the article, the following assumptions are needed to facilitate the technical details, although they may not be the weakest conditions. Let β_0 and $\theta_0(u)$ be the true value of the corresponding parameters β and $\theta(u)$.

(C1) For some $r \geq 2$, $\theta_l(u), l = 1, 2, \cdots, p$ are rth continuously differentiable.

(C2) The density function $f(u)$, which gernerates the sequence of design points U_{ij}, is bounded away from 0 and infinity on [0, 1]. And assume that $f(u)$ is continuously differentiable on (0, 1).

(C3) The number of measurements m is bounded.

(C4) For an $s > 2$, assume that $E\|X\|^{2s} < \infty,$, and $E\|Z\|^{2s} < \infty.$

(C5) Let $G_1(u) = E\{Z^{\mathrm{T}}\Omega^{-1}Z|U = u\}, G_2(u) = E\{X^{\mathrm{T}}\Omega^{-1}X|U = u\}$. The eigenvalues of $G_1(u)$ and $G_2(u)$ are bounded away from infinity and zero for sufficiently large n.

Firstly, we present that the estimators given by (8) are asymptotic normal.

Theorem 1 Suppose that conditions (C1)–(C5) hold, then

$$\sqrt{n}\left(\hat{\sigma}_v^2 - \sigma_v^2\right) \xrightarrow{D} N\left(0, V_v\right), \text{ and } \sqrt{nm}\left(\hat{\sigma}_\varepsilon^2 - \sigma_\varepsilon^2\right) \xrightarrow{D} N\left(0, V_\varepsilon\right), \quad (11)$$

where

$$V_v = Var\left(v_1^2\right) + \frac{4\sigma_v^2\sigma_\varepsilon^2}{m} + \frac{2\sigma_\varepsilon^2}{m(m-1)} \text{ and } V_v = Var\left(\varepsilon_{11}^2\right) + \frac{2\sigma_\varepsilon^4}{m-1}.$$

To obtain the consistency and oracle property about the estimators, additional conditions are required as follows, which are similar to that used in [10] and [15].

(C6) Let $b_n = \max_{k,l}\left\{p''_{\lambda_{1k}}\left(\|\gamma_{k0}\|_H\right), p''_{\lambda_k}\left(|\beta_{l0}|\right): \beta_{l0} \neq 0, \gamma_{k0} \neq 0\right\}$, then $b_n \to 0$, as $n \to \infty$.

(C7) $\liminf_{n \to \infty} \liminf_{\beta_l \to 0^+} \lambda_{2l}^{-1} p'_{\lambda_{2l}}\left(|\beta_l|\right) > 0,$ and $\liminf_{n \to \infty} \liminf_{\|\gamma_k\|_H \to 0^+} \lambda_{1k}^{-1} p'_{\lambda_{1k}}\left(\|\beta_k\|_H\right) > 0,$ where $l = s+1, \cdots, q, \quad k = d+1, \cdots, p.$

Theorem 2 Under the conditions (C1)-(C7) and the number of knots $k_n = O_p\left(n^{1/(2r+1)}\right)$. Then,

i) $$\left\|\hat{\beta} - \beta_0\right\| = O\left(n^{\frac{-r}{2r+1}} + a_n\right), \quad (12)$$

ii) $$\left\|\hat{\theta}_l(\cdot) - \theta_{l0}(\cdot)\right\| = O\left(n^{\frac{-r}{2r+1}} + a_n\right), \quad (13)$$

where $a_n = \max_{k,l}\left\{p'_{\lambda_{1k}}\left(\|\gamma_{k0}\|_H\right), p'_{\lambda_k}\left(|\beta_{l0}|\right): \beta_{l0} \neq 0, \gamma_{k0} \neq 0\right\}$, and r is defined in (C1).

Theorem 2 ensures the convergence rate of the weighted estimators for not only the parametric component, but also the nonparametric component. Furthermore, the following two theorems provide us with the oracle property of the consistent estimators.

Theorem 3 Under the conditions (C1)-(C7) and the number of knots $k_n = O_p\left(n^{1/(2r+1)}\right)$.

Let $\lambda_{\max} = \max_{k,l}\left\{\lambda_{1,k}, \lambda_{2,l}\right\}$, and $\lambda_{\min} = \max_{k,l}\left\{\lambda_{1k}, \lambda_{2l}\right\}$ .If $\lambda_{\max} \to 0$, and

$$n^{\frac{r}{2r+1}}\lambda_{\min} \to \infty,$$

as $n \to \infty$. Then, with probability tending to 1, $\hat{\beta}$ and $\hat{\theta}(\cdot)$ satisfy

i)
$$\hat{\beta}_l = 0, l = s+1, \cdots, q, \tag{14}$$

ii)
$$\hat{\theta}_l(\cdot) = 0, k = d+1, \cdots, p, \tag{15}$$

According to Remark 1 in [10], the variable selection procedure can specify the model correctly and efficiently. Next theorem further shows us that under some conditions, the nonzero coefficient estimators of the parametric component have the same asymptotic distribution as that based on true model.

Let $\beta^* = (\beta_1, \cdots, \beta_s)^{\mathrm{T}}, \theta^* = (\theta_1, \cdots, \theta_s)^{\mathrm{T}}$, and β_0^* and θ_0^* be the true values of β^* and θ^*, respectively. Therefore, the following theorem about the asymptotic distribution of the estimator $\hat{\beta}^*$ is established.

Theorem 4 Under the conditions (C1)-(C7) and the number of knots $k_n = O_p\left(n^{1/(2r+1)}\right)$. Then

$$\sqrt{n}\left(\hat{\beta}^* - \beta^*\right) \to N\left(0, \Sigma^{*-1}\right), \tag{16}$$

where Σ^* is defined by (A.10) in Appendix.

3. Practical Computational Issues

Denote that Ω_i and $\hat{\Omega}_i$ are the true and estimated covariance matrix within the ith subject. With a given initial value $\hat{\alpha}^{(0)} = \left(\hat{\gamma}^{\mathrm{T}(0)}, \hat{\beta}^{\mathrm{T}(0)}\right)^{\mathrm{T}}$, we can refine the regularized estimators $\hat{\alpha}$ according to the following iterative algorithms.

Step 1. Calculate the estimator Ω_i by the formula (10) with the given value $\hat{\alpha}_0$.

Step 2. Solve the penalized estimators $\hat{\alpha}$ for the parametric and nonparametric component by the Equation (5) with the within-subject covariance matrix interpolated by the covariance matrix estimated in Step 1.

Step 3. Replace the estimator $\hat{\alpha}_0$ in Step 1 with the one $\hat{\alpha}$ in Step 2 and iterate between Step 1 and Step 2 until the convergence criterion is met. Consequently, the required estimator can be solved with this iterative procedure.

Remark 1. This modified penalized estimation procedure inherits the computational efficiency and sparsity of Lasso type solutions. And the computational details can be referred to [13].

Although our theoretical results give technical conditions on k_n and λ, in practice, these parameters is always chosen based on data to enhance the adaptive property. It follows that the cross validation method is used to choose the optimal k_n and λ. However, in the real application, the parameter k_n and $\lambda_{1k}, \lambda_{2l}$ is chosen by minimizing the following cross-validation score

$$CV\left(k_n, \lambda_{11}, \cdots, \lambda_{1p}, \lambda_{21}, \cdots, \lambda_{2q}\right) = \sum_{i=1}^{n}\left\{Y_i - \tilde{Z}_i^T \hat{\alpha}^{(-i)}\right\}^{\mathrm{T}} \Omega_i^{-1}\left\{Y_i - \tilde{Z}_i^{\mathrm{T}} \hat{\alpha}^{(-i)}\right\}, \tag{17}$$

where $\hat{\alpha}^{(-i)}$ is the solution to the function (3) with the ith subject being deleted. Considering the character of the turning parameter, in practice, the turning parameter λ is used as $\lambda_{1k} = \lambda / \left\|\hat{\gamma}_k^{(0)}\right\|_H$ and $\lambda_{2l} = \lambda / \left|\hat{\beta}_l^{(0)}\right|$, where $\hat{\gamma}_k^{(0)}$ and $\hat{\beta}_l^{(0)}$ are the estimators of (3).

4. Empirical Study

We now use two examples to illustrate the superiority of the proposed weighted shrinkage estimation to that one without considering within-subject correlation.

Example 1. Consider a partially linear varying coefficient mixed effect model

$$Y_{ij} = X_{ij}^{\mathrm{T}}\beta + Z_{ij}^{\mathrm{T}}\theta(U_{ij}) + v_i + \varepsilon_{ij}, i = 1,\cdots,n; j = 1,\cdots,m, \tag{18}$$

where $\beta = (\beta_1,\cdots,\beta_{10})$ with $\beta_4 = 3$, $\beta_5 = 1.5$, and $\beta_6 = 2$, $\theta(u) = (\theta_1(u),\cdots,\theta_{10}(u))$ with $\theta_1(u) = 8 + 0.3\exp(2u-1)$ and $\theta_2(u) = 2\sin(\pi u)$ and other parametric and nonparametric component coefficients are zero. Moreover, it is further assumed that $X_{ijk} \sim N(0,4)$, $k = 1,2,\cdots,10$ and $Z_{ijk} \sim N(0,1.5), k = 1,2,\cdots,10$. Meanwhile, denote that $v_i \sim N(0,\sigma_v^2)$ and $\varepsilon_{ij} \sim N(0,\sigma_\varepsilon^2)$, where $(\sigma_v^2,\sigma_\varepsilon^2)$ is set to be (2,0.5) and (1,0.5). It is assumed that the number of observed subject and repeated measurement within each subject are $n = 50$ and $m = 3, 4$, respectively.

To illustrate the estimation accuracy of the proposed method, we define generalized mean squared errors (GMSE) and the square root of average square error (RASE) to be

$$\text{GMSE} = (\hat{\beta} - \beta_0)^{\mathrm{T}} E(XX^{\mathrm{T}})(\hat{\beta} - \beta_0), \text{ and } \text{RASE} = \left\{\frac{1}{N}\sum_{s=1}^{N}\sum_{k=1}^{p}\left[\hat{\theta}_k(u_s) - \theta_{k0}(u_s)\right]\right\}. \tag{19}$$

And for the purpose of a intensive comparison, in addition to the proposed method in this article, two other estimation methods are also required, that are the "naive" approach based on the working independence method, and the "ideal" one based on the true within-subject covariance. And the estimator, obtained by the "naive" approach, the proposed method in this article and the "ideal" approach, are denoted to be $\hat{\beta}$, $\hat{\beta}^W$ and $\tilde{\beta}^W$, respectively. During the simulations, it is assumed that "C" and "I" means the average number of the zero coefficients being correctly set to zero and the average number of the nonzero coefficients being incorrectly set to zero, respectively.

The results about variable selection, based on 100 replications, are included in **Table 1**. **Table 1** shows that the proposed method can select the true model quite well and leads to smaller GMSE and RASE values. **Table 2** reports a satisfactory estimation for the variance component. As the sample increase, the performance becomes better.

Secondly, the variance, bias and mean square error of the estimators for the nonzero parameters, denoted to be "V", "Bias" and "MSE", are listed in **Table 3**. From **Table 3**, all the three methods can obtain consistent estimators with small bias. Moreover, the values of "V" and "MSE" also argue that the newly proposed method and the "ideal" method can derive a more efficient estimator than the "naive" method does. What's more, the asymptotic normality of the estimators for the parametric component is shown with Quantile-Quantile plot in **Figure 1**.

Finally, in **Figure 2**, all of the curves, estimated by all the three methods, fit the true nonparametric curve well. However, their 95% confidence intervals have different interval length, with the proposed method in this article and the "ideal" method showing a similar and smaller length. All these scenarios indicate a great improvement of the estimation with the proposed method in this article.

Table 1. Simulation results for the variable selection.

$\left(\sigma_v^2,\sigma_\varepsilon^2\right)$		$n = 50$, $m = 3$			$n = 50$, $m = 4$		
β		C	I	GMSE	C	I	GMSE
(2, 0.5)	$\hat{\beta}$	7	0	0.0428	7	0	0.0313
	$\hat{\beta}^w$	7	0	0.0138	7	0	0.0085
	$\tilde{\beta}^w$	7	0	0.0138	7	0	0.0083
(1, 0.5)	$\hat{\beta}$	7	0	0.0266	7	0	0.0188
	$\hat{\beta}^w$	7	0	0.1863	7	0	0.0085
	$\tilde{\beta}^w$	7	0	0.1289	7	0	0.008
θ		C	I	RASE	C	I	RASE
(2, 0.5)	$\hat{\theta}$	7.92	0	0.4471	8	0	0.3601
	$\hat{\theta}^w$	8	0	0.2466	8	0	0.1916
	$\tilde{\theta}^w$	8	0	0.2419	8	0	0.1888
(1, 0.5)	$\hat{\theta}$	7.92	0	0.3461	8	0	0.2792
	$\hat{\theta}^w$	7.95	0	0.2477	8	0	0.1916
	$\tilde{\theta}^w$	7.98	0	0.2402	8	0	0.1867

Table 2. Simulation results for the estimators of the variance components $\sigma_v^2, \sigma_\varepsilon^2$.

$\left(\sigma_v^2,\sigma_\varepsilon^2\right)$	$n = 50$, $m = 3$			$n = 50$, $m = 4$		
σ_v^2	M	V	MSE	M	V	MSE
(2, 0.5)	1.6	0.27	0.418	1.64	0.42	0.54
(1, 0.5)	0.7	0.12	0.206	0.64	0.13	0.25
σ_ε^2	M	V	MSE	M	V	MSE
(2, 0.5)	0.5	0.02	0.025	0.53	0.02	0.022
(1, 0.5)	0.4	0.08	0.008	0.50	0.01	0.01

Table 3. Simulation results ×100 for $\beta_4, \beta_5, \beta_6$ with $n = 50$ and $m = 3$.

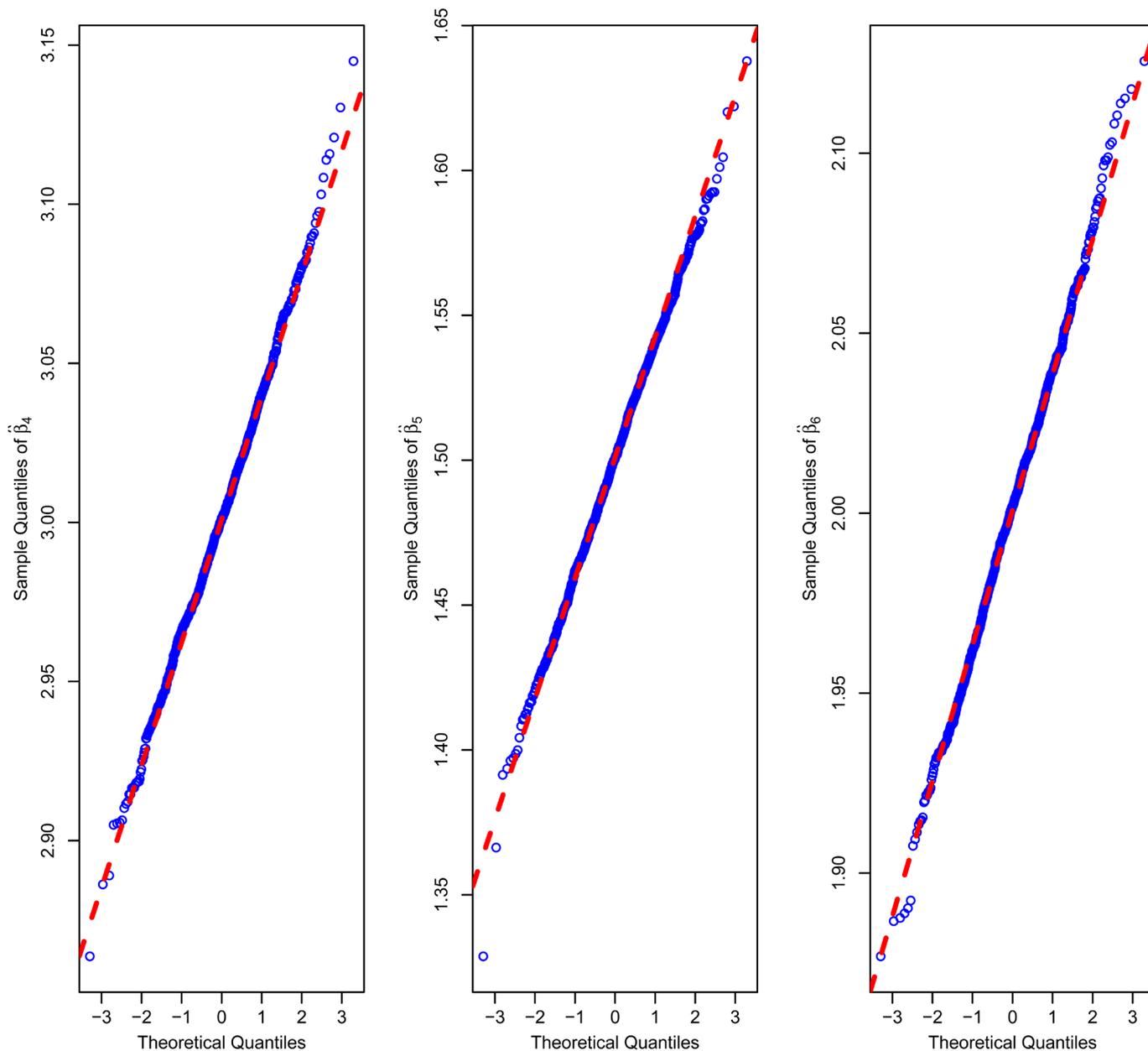| $\left(\sigma_v^2,\sigma_\varepsilon^2\right)$		β_4			β_5			β_6		
		Var	Bias	MSE	Var	Bias	MSE	Var	Bias	MSE
(2, 0.5)	$\hat{\beta}$	0.498	7.063	0.997	0.494	6.456	0.911	0.484	6.140	0.861
	$\hat{\beta}^w$	0.159	5.191	0.428	0.168	6.519	0.593	0.149	0.502	0.152
	$\tilde{\beta}^w$	0.150	5.090	0.409	0.157	6.457	0.576	0.138	0.231	0.139
(1, 0.5)	$\hat{\beta}$	0.286	−4.931	0.529	0.304	−11.76	1.688	0.291	−5.826	0.630
	$\hat{\beta}^w$	0.153	−4.128	0.323	0.152	−4.408	0.346	0.161	−5.328	0.445
	$\tilde{\beta}^w$	0.148	−4.739	0.373	0.138	−2.798	0.216	0.145	−5.386	0.435

Figure 1. Q-Q plot of the estimator $\hat{\beta}_4^W, \hat{\beta}_5^W, \hat{\beta}_6^W$.

Example 2. To illustrate the effectiveness of the proposed estimation procedure, we shall apply it to the analysis of a longitudinal AIDS data set, which is reported by [18] and comprises HIV status of 283 homosexual males who were infected with HIV during a follow-up period between 1984 and 1991. The focus in this application is to probe into the trend of the mean CD4 percentage depletion over time and evaluate the effects of cigarette smoking, preHIV infection CD4 percentage and age at infection on the mean CD4 percentage after infection.

For the jth measurement of the ith subject, let Y_{ij} be the mean CD4 percentage, t_{ij} be the time in years after HIV infection, $X_{i,1}$ be the centered preCD4 percentage and
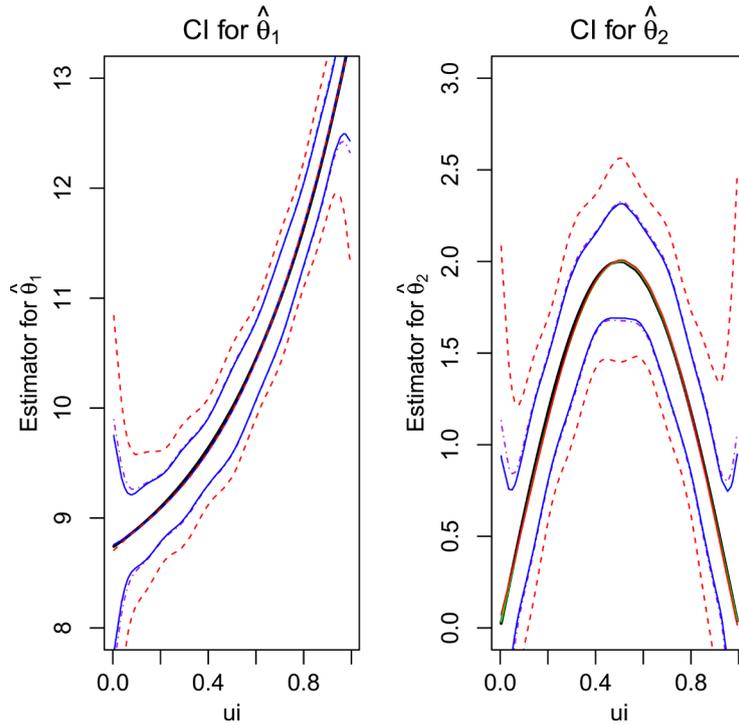
Figure 2. The estimated average curve for the nonparametric component $\theta_1(\cdot), \theta_2(\cdot)$ and their 95% pointwise confidence interval with three coincident curves being the estimated ones for the nonparametric component, and the red dotted, the coincident solid and dotted-dashed ones being the estimated confidence intervals by the "naive", the "ideal" and the proposed method, respectively.

$X_{i,2} = X_{i,1}^2$. $Z_{i,1}$ be the centered age at HIV infection and $Z_{i,2} = Z_{i,1}^2$. Let $Z_{i,3}$ be the smoking status taking a value of 1 or 0. Hence, assume the following model

$$Y_{ij} = \theta_0\left(t_{ij}\right) + Z_{ij,1}\theta_1\left(t_{ij}\right) + Z_{ij,2}\theta_2\left(t_{ij}\right) + Z_{ij,3}\theta_3\left(t_{ij}\right) + X_{ij,1}^{\mathrm{T}}\beta_1 + X_{ij,2}^{\mathrm{T}}\beta_2 + \nu_i + \varepsilon_{ij}, \quad (20)$$

where the baseline of CD4 percentage $\theta_0(t)$ is used to represent the mean CD4 percentage of t years after the infection. The random effect ν_i, representing the within-subject correlation, is also included in the assumed model.

By the analysis, there are two variables $\theta_0(t)$ and $X_{i,1}$ with nonzero estimators, which indicate the significant effect on the response variable Y_{ij}. Other variables' coefficients are zero and have no significant effect. Moreover, the estimated curve in **Figure 3** shows the trend of the mean CD4 percentage depletion over time.

5. Conclusion and Discussion

This article considered an efficient shrinkage estimation for the partially linear varying coefficient models with random effect. Variance component model was employed to take within subject correlation into consideration. Some asymptotic properties, such as convergence rate, consistency and oracle property, were established. Moreover, the effectiveness was further illustrated by a real data analysis. As a more ambitious goal, we would try to investigate the performance of variable selection issue for mixed effect
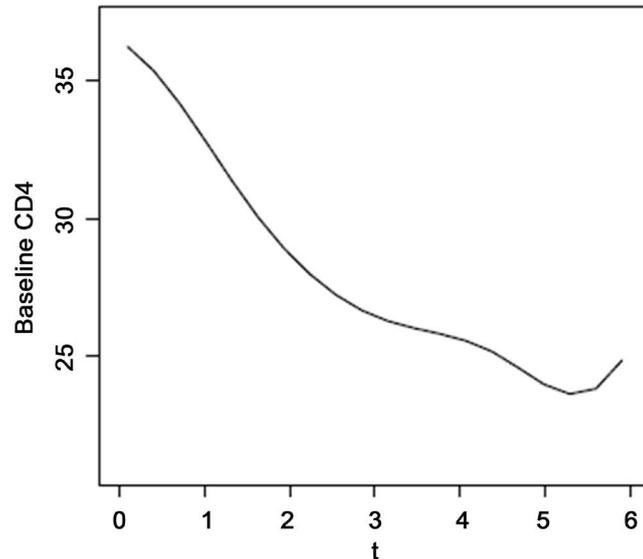
Figure 3. The estimators for the mean CD4 percentage $\theta_0(t)$, indicating the changing tendency with t.

model under a more general within-subject covariance matrix.

Acknowledgements

References

[1] Hastie, T. and Tibshirani, R. (1993) Varying Coefficient Models. *Journal of the Royal Statistical Society: Series B*, **55**, 757-796.

[2] Ahmad, I., Leelahanon, S. and Li, Q. (2005) Efficient Estimation of a Semiparametric Partially Linear Varying Coefficient Model. *Ann. Statist.*, **33**, 258-283. http://dx.doi.org/10.1214/009053604000000931

[3] 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

[4] Xue, L.G. and Zhu, L.X. (2007) Empirical Likelihood for a Varying Coefficient Model with Longitudinal Data. *Journal of the American Statistical Association*, **102**, 642-652. http://dx.doi.org/10.1198/016214507000000293

[5] Zhao, P.X. and Xue, L.G. (2010) Variable Selection for Semiparametric Varying Coefficient Partially Linear Errors-in-Variable Models. *Journal of Multivariate Analysis*, **101**, 1872-1883. http://dx.doi.org/10.1016/j.jmva.2010.03.005

[6] Ruckstuhl, A.F., Welsh, A.H. and Carroll, R.J. (2000) Nonparametric Function Estimation of Estimation of the Relationship between Two Repeatedly Measurement Variables. *Statistica Sinica*, **10**, 51-71.

[7] Wang, N. (2003) Margingal Nonparametric Kernel Regression Accounting for within Subject Correlation. *Biometrika*, **90**, 43-52. http://dx.doi.org/10.1093/biomet/90.1.43

[8] Su, L. and Ullah, A. (2007) More Efficient Estimation of Nonparametric Panel Data Models with Random Effects. *Economics Letters*, **96**, 375-380. http://dx.doi.org/10.1016/j.econlet.2007.02.018

[9] You, J. and Zhou, X. (2009) Partially Linear Models and Polynomial Spline Approximations for the Analysis of Unbalanced Panel Data. *Journal of Statistical Planning and Inference*, **139**, 679-695. http://dx.doi.org/10.1016/j.jspi.2007.04.037

[10] Fan, J. and Li, R. (2001) Variable Selection via Nonconcave Penalized Likelihood and Its Oracle Properties. *Journal of the American Statistical Association*, **96**, 710-723. http://dx.doi.org/10.1198/016214501753382273

[11] 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

[12] Yuan, M. and Lin, Y. (2006) Model Selection and Estimation in Regression with Grouped Variables. *Journal of the Royal Statistical Society: Series B*, **68**, 49-67. http://dx.doi.org/10.1111/j.1467-9868.2005.00532.x

[13] Zou, H. and Li, R.Z. (2008) One Steps Sparse Estimates in Noncave Penalized Likelihood Models. *Annals of Statistics*, **36**, 1509-1533. http://dx.doi.org/10.1214/009053607000000802

[14] Liang, H. and Li, R.Z. (2009) Variable Selection for Partially Linear Models with Measurement Errors. *Journal of the American Statistical Association*, **104**, 234-248. http://dx.doi.org/10.1198/jasa.2009.0127

[15] Zhao, P.X. and Xue, L.G. (2009) Variable Selection for Semiparametric Varying Coefficient Partially Linear Models. *Statistics and Probability Letters*, **79**, 2148-2157. http://dx.doi.org/10.1016/j.spl.2009.07.004

[16] Li, G., Lai, P. and Lian, H. (2014) Variable Selection and Estimation for Partially Linear Single-Index Models with Longitudinal Data. *Statistics & Computing*, **25**, 579-593. http://dx.doi.org/10.1007/s11222-013-9447-8

[17] Johnson, B., Lin, D. and Zeng, D. (2008) Penalized Estimating Functions and Variable Selection in Semiparametric Regression Models. *Journal of the American Statistical Association*, **103**, 672-680. http://dx.doi.org/10.1198/016214508000000184

[18] Kaslow, R.A., Ostrow, D.G., Detels, R., Phair, J.P., Polk, B.F. and Rinaldo, C.R. (1987) The Multicenter AIDS Cohort Study: Rationale, Organization and Selected Characteristics of the Participants. *American Journal of Epidemiology*, **126**, 310-318. http://dx.doi.org/10.1093/aje/126.2.310

Appendix

Supplementary material related to this article can be asked for by email.