Statistics and Its Interface Volume 10 (2017) 495–504

Varying-coefficient single-index model for longitudinal data*

Hongmei Lin, Riquan Zhang[†], Jianhong Shi, and Yuedong Wang

In this paper we consider a general class of varying-coefficient single-index models for longitudinal data. This class of models provides a tool for simultaneous dimension reduction and the exploration of dynamic patterns. We develop an estimation procedure using Cholesky decomposition, local linear and backfitting technique. Asymptotic normality for the proposed estimators of varying-coefficient functions, link function and parameters will be established. Monte Carlo simulation studies show excellent finite-sample performance. We illustrate our methods with a real data example.

AMS 2000 subject classifications: Primary 62G05; secondary 62E20.
Keywords and phrases: Varying-coefficient single-index models, Cholesky decomposition, Local linear regression, Longitudinal data.

1. INTRODUCTION

Regression analysis is commonly used to study the relationship between a response variable and a vector of covariates. In recent years nonparametric and semiparametric regression models have attracted a great deal of attention due to their flexibility and power to uncover hidden relationship between the response and covariates. See [13], [11], [7] and [27] for an introduction to nonparametric and semiparametric models.

Varying-coefficient models are useful for exploring dynamic patterns [14, 3, 15, 23], and single-index models are useful for dimension reduction [12, 30, 1, 21]. In practice it may be desirable to explore dynamic patterns for some covariates while reduce dimensionality for others. This had led [28] to propose the following varying-coefficient single-index model (VCSIM) for independent data:

$$y_i = g_0(\boldsymbol{\alpha}^T \boldsymbol{x}_i) + \boldsymbol{g}^T(u_i)\boldsymbol{z}_i + \varepsilon_i, \quad i = 1, \cdots, n, \tag{1}$$

where y_i's are observations of the response variable, $\boldsymbol{x}_i \in \mathbb{R}^q$, $\boldsymbol{z}_i \in \mathbb{R}^p$, and $u_i \in \mathbb{R}$ are observations of the associated covariates, ε_i's are independent random errors with $\mathrm{E}(\varepsilon_i) = 0$ and $\mathrm{Var}(\varepsilon_i) = \sigma^2$, $g_0(\cdot)$ is an unknown univariate smooth index function, $\boldsymbol{g}(\cdot) = (g_1(\cdot), \cdots, g_p(\cdot))^T$ is an unknown vector of coefficient functions, and $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_q)^T$ is a vector of unknown parameters. For identifiability, it was assumed that $\alpha_1 \geq 0$ and $\boldsymbol{\alpha}^T\boldsymbol{\alpha} = 1$.

The VCSIM (1) is a very inclusive model. It can be seen as a generation of many important models such as varying-coefficient models, single index models, partially linear single index models [4, 29] and varying-coefficient partially linear models[31, 8, 34].

The varying-coefficient single-index model is proposed first by [28], when they analyse the environmental data of Hong Kong. They develop the theory and method to estimate the unknown parameter α_0 and the unknown functions $g(\cdot)$ and $\mathbf{g}(\cdot)$, based on the local linear method, average method and backfitting technique. Recently, [27] also study the VCSIM by the stepwise approaches. The distinct difference between the two methods is in the way they obtained initial estimators of unknown functions. [28] obtained the so called quasi-initial estimators by combining the bi-local linear smoother and the average method. However, [27] developed stepwise approaches to estimate the unknown index parameter α and functions g_0, $\mathbf{g}$. They first rewrote the model as a varying-coefficient model by pretending some unknown link function g_0 and the single-index parameter vector $\boldsymbol{\alpha}$ to be known. Then, they use the local linear regression technique to obtain an initial estimator of varying -coefficient function $\mathbf{g}$. Again, they estimated the link function g_0 by pretending the unknown single-index parameter vector $\boldsymbol{\alpha}$ to be known. However, none of these works addresses longitudinal data, which is the focus of our paper.

Longitudinal studies are common in many research areas including social, economic, health and medical sciences. In this paper, we consider a VCSIM for longitudinal data. Consisting of outcome measurements repeatedly taken on each subject, longitudinal data are usually correlated. Many methods have been developed to model and estimate within-subject correlation. In particular, the Cholesky decomposition proposed by [25] is a powerful approach since it automatically leads to positive definite covariance matrices. In addition, the decomposition is appealing since the parameters have meaningful interpretations as autoregressive coefficients and these parameters can be modeled via regression

*The research was supported in part by NSF DMS-1507620, National Natural Science Foundation of China (11571112, 11501372, 11571148), Doctoral Fund of Ministry of Education of China (20130076110004), Program of Shanghai Subject Chief Scientist (14XD1401600), the 111 Project of China (B14019) and the Fund of ECNU for Overseas Studies, International Conferences and Domestic Scholar Visits(40600-511232-14204/003).
[†]Corresponding author.

techniques. This method has received considerable attention in the statistics literature (see for example [22, 33, 17, 32]). We will develop an estimation procedure using the Cholesky decomposition, local linear and backfitting technique that estimates all parameters and nonparametric functions simultaneously. One of the advantages of our method is that there is no need to specify the correlation structure. The proposed estimation procedure is easy to implement and computationally fast. We derive the asymptotic normality of the resulting estimators and evaluate finite sample performance using Monte Carlo simulations. The usefulness of the new model is illustrated through a real data set.

The article is organized as follows. Section 2 introduces the VCSIM for longitudinal data and the estimation procedure. Sections 3 and 4 present asymptotic properties and finite sample evaluations of the proposed estimators. Section 5 illustrates the proposed methods using a real data example. Technical proofs are given in an Appendix.

2. MODEL AND ESTIMATION

Let y_{ij} be the jth observation of the response variable from subject i, and $\boldsymbol{x}_{ij} \in \mathbb{R}^q$, $\boldsymbol{z}_{ij} \in \mathbb{R}^p$ and $u_{ij} \in \mathbb{R}$ be the jth observation of covariates from subject i. A VCSIM for longitudinal data assumes that

$$y_{ij} = g_0(\boldsymbol{\alpha}^T \boldsymbol{x}_{ij}) + \boldsymbol{g}^T(u_{ij})\boldsymbol{z}_{ij} + \varepsilon_{ij}, \quad i = 1, 2, \cdots, n, \quad j = 1, 2, \cdots, J_i, \tag{2}$$

where ε_{ij}'s are random errors with $\mathrm{E}(\varepsilon_{ij}) = 0$, $g_0(\cdot)$ is an unknown univariate smooth index function, $\boldsymbol{g}(\cdot) = (g_1(\cdot), \cdots, g_p(\cdot))^T$ is an unknown vector of p smooth coefficient functions, and $\boldsymbol{\alpha} = (\alpha_1, \cdots, \alpha_q)^T$ is q-dimensional vector of unknown parameters. For identifiability, we assume that $\alpha_1 \geq 0$ and $\boldsymbol{\alpha}^T\boldsymbol{\alpha} = 1$.

The observations within subjects are usually correlated. Many parametric correlation structures have been proposed in the literature [24]. Herein, we will extend the work of [32], model and estimate the correlation structure using the Cholesky decomposition. For easy of presentation, we will first consider the balanced case with $J_i = J$. Denote $\varepsilon_i = (\varepsilon_{i1}, \varepsilon_{i2}, \ldots, \varepsilon_{iJ})^T$ and $\mathrm{cov}(\varepsilon_i | \boldsymbol{x}_i, \mathbf{z}_i, \mathbf{u}_i) = \boldsymbol{\Sigma}$. Based on the Cholesky decomposition, there exists a lower triangle matrix $\boldsymbol{\Phi}$ with 1's on the main diagonal such that $\mathrm{cov}(\boldsymbol{\Phi}\varepsilon_i) = \boldsymbol{\Phi}\boldsymbol{\Sigma}\boldsymbol{\Phi}^T = \mathbf{D}$ where $\mathbf{D} = \mathrm{diag}(d_1^2, d_2^2, \ldots, d_J^2)$ is a diagonal matrix. Let $\boldsymbol{e}_i = (e_{i1}, e_{i2}, \ldots, e_{iJ})^T = \boldsymbol{\Phi}\varepsilon_i$. Then we have

$$\varepsilon_{ij} = \sum_{k=1}^{j-1} \phi_{jk}\varepsilon_{ik} + e_{ij}, \quad i = 1, 2, \ldots, n, \quad j = 1, \ldots, J, \tag{3}$$

where ϕ_{jk} is the negative of the (j, k)-element of $\boldsymbol{\Phi}$ and e_{ij}'s are uncorrelated with $\mathrm{var}(e_{ij}) = d_j^2$. By convention, $\sum_{k=1}^{0} \phi_{1k}\varepsilon_{ik} = 0$ when $j = 1$. Substituting equation (3) into model (2), we obtain the following model with uncorrelated error terms e_{ij}:

$$y_{ij} = g_0(\boldsymbol{\alpha}^T \boldsymbol{x}_{ij}) + \boldsymbol{g}^T(u_{ij})\boldsymbol{z}_{ij} + \sum_{k=1}^{j-1} \phi_{jk}\varepsilon_{ik} + e_{ij}, \quad i = 1, 2, \ldots, n, \quad j = 1, \ldots, J. \tag{4}$$

The terms ε_{ij} in (4) are not observable. We will estimate them by $\check{\varepsilon}_{ij} = y_{ij} - \check{g}_0(\check{\boldsymbol{\alpha}}^T \boldsymbol{x}_{ij}) - \check{\boldsymbol{g}}^T(u_{ij})\boldsymbol{z}_{ij}$, where $\check{\boldsymbol{\alpha}}$, $\check{g}_0$ and $\check{\boldsymbol{g}}$ are local linear estimators for model (2) with independent random errors [28]. Let $\boldsymbol{\phi} = (\phi_{21}, \phi_{31}, \ldots, \phi_{JJ-1})^T$ and $\mathbf{F}_{ij} = (\mathbf{0}^T_{(j-2)(j-1)/2}, \check{\varepsilon}_{i1}, \ldots, \check{\varepsilon}_{ij-1}, \mathbf{0}^T_{J(J-1)/2-j(j-1)/2})^T$, where $\mathbf{0}_k$ is a k-dimension column vector with all entries 0. Replacing the ε_{ij}'s in (4) with $\check{\varepsilon}_{ij}$'s, we have

$$y_{ij} = g_0(\boldsymbol{\alpha}^T \boldsymbol{x}_{ij}) + \boldsymbol{g}^T(u_{ij})\boldsymbol{z}_{ij} + \mathbf{F}_{ij}^T\boldsymbol{\phi} + e_{ij}, \quad i = 1, 2, \ldots, n, \quad j = 1, \ldots, J. \tag{5}$$

Under model (5), our primary interest is to estimate $\boldsymbol{\alpha}$, $\boldsymbol{\phi}$, $g_0(\cdot)$ and $\boldsymbol{g}(\cdot)$. Suppose that $g_r(\cdot)$ has a continuous second derivative for every $r = 0, 1, \ldots, p$. Then we approximate $g_r(\cdot)$ locally by a linear function

$$g_r(v) \approx g_r(u) + g_r'(u)(v-u) \equiv a_r + b_r(v-u), \quad r = 0, 1, \cdots, p,$$

for v in a neighborhood of u. Throughout the paper, let $K_{i,h_i}(\cdot) = K_i(\cdot/h_i)/h_i$ where $K_i(\cdot)$ is a kernel function and $h_i > 0$ is a bandwidth, and $\mu_k(K_i) = \int t^k K_i(t)\mathrm{d}t$, $\nu_k(K_i) = \int t^k K_i^2(t)\mathrm{d}t$.

We first describe the derivation of initial estimates for $\boldsymbol{\alpha}$, $\boldsymbol{\phi}$, $g_0(\cdot)$ and $\boldsymbol{g}(\cdot)$. We apply the minimum average variance estimation (MAVE) method to obtain initial estimates $\hat{\boldsymbol{\alpha}}^0$ and $\hat{\boldsymbol{\phi}}^0$ of $\boldsymbol{\alpha}$ and $\boldsymbol{\phi}$ by fitting a partially linear single index model in (5) [29]. We then consider the following weighted sum of squares:

$$\sum_{i=1}^{n}\sum_{j=1}^{J}(y_{ij}^* - a_0 - b_0(\hat{\boldsymbol{\alpha}}^{0T}\boldsymbol{x}_{ij} - t) - \mathbf{a}^T\boldsymbol{z}_{ij} - \mathbf{b}^T(u_{ij} - u)\boldsymbol{z}_{ij})^2 K_{1,h_1}(\hat{\boldsymbol{\alpha}}^{0T}\boldsymbol{x}_{ij} - t)K_{2,h_2}(u_{ij} - u),$$

where $y_{ij}^* = y_{ij} - \mathbf{F}_{ij}^T\hat{\boldsymbol{\phi}}^0$, and a_0, $\mathbf{a}^T = (a_1, \cdots, a_p)$, b_0 and $\mathbf{b}^T = (b_1, \cdots, b_p)$ are parameters. Setting estimates of g_0 and $\boldsymbol{g}$ as the estimates of a_0 and $\mathbf{a}$ respectively, we have

$$\tilde{g}_0(t, u; \hat{\boldsymbol{\alpha}}^0, \hat{\boldsymbol{\phi}}^0) = \sum_{i=1}^{n}\sum_{j=1}^{J} \boldsymbol{c}_{1,2p+2}^T(\boldsymbol{V}^T\boldsymbol{W}\boldsymbol{V})^{-1}\mathbf{v}_{ij} K_{1,h_1}(\hat{\boldsymbol{\alpha}}^{0T}\boldsymbol{x}_{ij} - t)K_{2,h_2}(u_{ij} - u)y_{ij}^*,$$

$$\tilde{\boldsymbol{g}}(t, u; \hat{\boldsymbol{\alpha}}^0, \hat{\boldsymbol{\phi}}^0) = \sum_{i=1}^{n}\sum_{j=1}^{J} \boldsymbol{E}_{p,2p+2}(\boldsymbol{V}^T\boldsymbol{W}\boldsymbol{V})^{-1}\mathbf{v}_{ij} K_{1,h_1}(\hat{\boldsymbol{\alpha}}^{0T}\boldsymbol{x}_{ij} - t)K_{2,h_2}(u_{ij} - u)y_{ij}^*,$$

where $\boldsymbol{c}_{j,2p+2}$ is a $(2p+2)$ unit vector with one at the jth position, $\boldsymbol{E}_{p,2p+2}$ is a $p \times (2p+2)$ matrix with $\boldsymbol{c}_{j+1,2p+2}$ as the jth row, $\boldsymbol{V}$ is an $nJ \times (2p+2)$ matrix with $\mathbf{v}_{ij}^T = (1, \boldsymbol{z}_{ij}^T, (\hat{\boldsymbol{\alpha}}^{0T}\boldsymbol{x}_{ij} - t)/h_1, \boldsymbol{z}_{ij}^T(u_{ij} - u)/h_2)$ as the $((i-1)J + j)$th row, and $\boldsymbol{W} = \mathrm{diag}(\{K_{1,h_1}(\hat{\boldsymbol{\alpha}}^{0T}\boldsymbol{x}_{ij} - t)K_{2,h_2}(u_{ij} - u)\}_{i=1,j=1}^{n,\ \ J})$.

The initial estimates of functions $g_r(t)$ for $r = 0, \ldots, p$ are defined as

$$\breve{g}_0(t; \hat{\boldsymbol{\alpha}}^0, \hat{\boldsymbol{\phi}}^0) = \frac{1}{nJ}\sum_{i=1}^{n}\sum_{j=1}^{J}\tilde{g}_0(t, u_{ij}; \hat{\boldsymbol{\alpha}}^0, \hat{\boldsymbol{\phi}}^0),$$

$$\breve{\boldsymbol{g}}(u; \hat{\boldsymbol{\alpha}}^0, \hat{\boldsymbol{\phi}}^0) = \frac{1}{nJ}\sum_{i=1}^{n}\sum_{j=1}^{J}\tilde{\boldsymbol{g}}(\hat{\boldsymbol{\alpha}}^{0T}\boldsymbol{x}_{ij}, u; \hat{\boldsymbol{\alpha}}^0, \hat{\boldsymbol{\phi}}^0). \tag{6}$$

We now describe our estimation procedure. We apply the backfitting technique to compute estimates of parameters and nonparametric functions iteratively. For fixed $\boldsymbol{\alpha}$ and $\boldsymbol{\phi}$ at the current estimates denoted as $\hat{\boldsymbol{\alpha}}_{\_}$ and $\hat{\boldsymbol{\phi}}_{\_}$, let $\hat{\beta}_0$ and $\hat{\gamma}_0$ be the minimizers of the following weighted sum of squares

$$\sum_{i=1}^{n}\sum_{j=1}^{J}\hat{d}_{ij}^{-2}(\tilde{y}_{ij} - \beta_0 - \gamma_0(\hat{\boldsymbol{\alpha}}_{\_}^T\boldsymbol{x}_{ij} - t))^2 K_{3,h_3}(\hat{\boldsymbol{\alpha}}_{\_}^T\boldsymbol{x}_{ij} - t),$$

where $\tilde{y}_{ij} = y_{ij} - \mathbf{F}_{ij}^T\hat{\boldsymbol{\phi}}_{\_} - \breve{\boldsymbol{g}}(u_{ij}; \hat{\boldsymbol{\alpha}}_{\_}, \hat{\boldsymbol{\phi}}_{\_})^T\boldsymbol{z}_{ij}$. Then the updated estimate of g_0 is

$$\hat{g}_0(t) = \hat{\beta}_0 = \sum_{i=1}^{n}\sum_{j=1}^{J}\hat{d}_{ij}^{-2}\boldsymbol{c}_{1,2}(\tilde{\boldsymbol{X}}^T\boldsymbol{W}_1\tilde{\boldsymbol{X}})^{-1}\tilde{\boldsymbol{x}}_{ij} K_{3,h_3}(\hat{\boldsymbol{\alpha}}_{\_}^T\boldsymbol{x}_{ij} - t)\tilde{y}_{ij}, \tag{7}$$

where $\boldsymbol{c}_{1,2} = (1,0)^T$, $\tilde{\boldsymbol{X}}$ denotes an $nJ \times 2$ matrix with $\tilde{\boldsymbol{x}}_{ij}^T = (1, (\hat{\boldsymbol{\alpha}}_{\_}^T\boldsymbol{x}_{ij} - t)/h_3)$ as its rows, $\boldsymbol{W}_1 = \mathrm{diag}(\{\hat{d}_{ij}^{-2}K_{3,h_3}(\hat{\boldsymbol{\alpha}}_{\_}^T\boldsymbol{x}_{ij} - t)\}_{i=1,j=1}^{n,\ \ J})$, and $\hat{d}_{ij}$ is any consistent estimate of d_{ij}. In our simulations and real data example, d_{ij}'s are estimated by sample standard deviations of residuals from model (3) [32].

Let $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\gamma}}$ be the minimizers of the following weighted sum of squares

$$\sum_{i=1}^{n}\sum_{j=1}^{J}\hat{d}_{ij}^{-2}(\tilde{\tilde{y}}_{ij} - \boldsymbol{\beta}^T\boldsymbol{z}_{ij} - \boldsymbol{\gamma}^T\boldsymbol{z}_{ij}(u_{ij} - u))^2 K_{4,h_4}(u_{ij} - u),$$

where $\tilde{\tilde{y}}_{ij} = y_{ij} - \mathbf{F}_{ij}^T\hat{\boldsymbol{\phi}}_{\_} - \breve{g}_0(\hat{\boldsymbol{\alpha}}_{\_}^T\boldsymbol{x}_{ij}; \hat{\boldsymbol{\alpha}}_{\_}, \hat{\boldsymbol{\phi}}_{\_})$. Then the updated estimate of $\boldsymbol{g}$ is

$$\hat{\boldsymbol{g}}(u) = \hat{\boldsymbol{\beta}} = \sum_{i=1}^{n}\sum_{j=1}^{J}\hat{d}_{ij}^{-2}\boldsymbol{E}_{p,2p}(\tilde{\mathbf{U}}^T\boldsymbol{W}_2\tilde{\mathbf{U}})^{-1}\tilde{\mathbf{u}}_{ij} K_{4,h_4}(u_{ij} - u)\tilde{\tilde{y}}_{ij}, \tag{8}$$

where $\boldsymbol{E}_{p,2p}$ is a $p \times 2p$ matrix with $\boldsymbol{c}_{j,2p}$ as its jth row, $\tilde{\mathbf{U}}$ denotes an $nJ \times 2p$ matrix with $\tilde{\mathbf{u}}_{ij}^T = (\boldsymbol{z}_{ij}^T, \boldsymbol{z}_{ij}^T(u_{ij} - u)/h_4)$ as its rows, and $\boldsymbol{W}_2 = \mathrm{diag}(\{\hat{d}_{ij}^{-2}K_{4,h_4}(u_{ij} - u)\}_{i=1,j=1}^{n,\ \ J})$.

For fixed g_0 and $\boldsymbol{g}$ at the current estimates denoted as $\hat{g}_{0\_}$ and $\boldsymbol{g}_{\_}$, to update $\boldsymbol{\alpha}$ and $\boldsymbol{\phi}$, we consider the following weighted sum of squares

$$\sum_{i=1}^{n}\sum_{j=1}^{J}\hat{d}_{ij}^{-2}(y_{ij} - \hat{g}_{0\_}(\boldsymbol{\alpha}^T\boldsymbol{x}_{ij}) - \hat{\boldsymbol{g}}_{\_}^T(u_{ij})\boldsymbol{z}_{ij} - \mathbf{F}_{ij}^T\boldsymbol{\phi})^2, \quad \text{subject to } \boldsymbol{\alpha}^T\boldsymbol{\alpha} = 1. \tag{9}$$

Approximating $\hat{g}_{0\_}(\boldsymbol{\alpha}^T\boldsymbol{x}_{ij})$ by its first order Taylor expansion, we update the estimates of $\boldsymbol{\alpha}$ and $\boldsymbol{\phi}$ by minimizing the following approximated weighted sum of squares

$$\sum_{i=1}^{n}\sum_{j=1}^{J}\hat{d}_{ij}^{-2}(\tilde{\tilde{\tilde{y}}}_{ij} - \hat{g}_{0\_}'(\hat{\boldsymbol{\alpha}}_{\_}^T\boldsymbol{x}_{ij})\boldsymbol{x}_{ij}^T\boldsymbol{\alpha} - \mathbf{F}_{ij}^T\boldsymbol{\phi})^2, \quad \text{subject to } \boldsymbol{\alpha}^T\boldsymbol{\alpha} = 1, \tag{10}$$

where $\tilde{\tilde{\tilde{y}}}_{ij} = y_{ij} - \hat{g}_{0\_}(\hat{\boldsymbol{\alpha}}_{\_}^T\boldsymbol{x}_{ij}) + \hat{g}_{0\_}'(\hat{\boldsymbol{\alpha}}_{\_}^T\boldsymbol{x}_{ij})\hat{\boldsymbol{\alpha}}_{\_}^T\boldsymbol{x}_{ij} - \hat{\boldsymbol{g}}_{\_}^T(u_{ij})\boldsymbol{z}_{ij}$. [4] showed that estimates based on (9) and (10) are asymptotically equivalent.

One of the key factors in our estimation procedure is the selection of bandwidths. Theoretical conditions will be discussed in the next section. In our simulations and real data example, we will use leave-one-subject-out cross-validation as in [7], [6] and [16]. When the number of subjects is large, one may consider a K-fold cross-validation to reduce the computational cost [18, 5].

The computation procedure is summarized in the following algorithm.

Algorithm for fitting a VCSIM

1. *Initialize*: Derive initial estimates of $\boldsymbol{\alpha}$ and $\boldsymbol{\phi}$ and $g_r(t)$ for $r = 0, \ldots, p$.
2. *Cycle*: Alternate between (a) and (b) until convergence.
   (a) Conditional on current estimates of $\boldsymbol{\alpha}$ and $\boldsymbol{\phi}$, update $g_r(t)$ for $r = 0, \ldots, p$ using equations (6), (7) and (8).
   (b) Conditional on current estimates of $g_r(t)$ for $r = 0, \ldots, p$, update $\boldsymbol{\alpha}$ and $\boldsymbol{\phi}$ using equation (10).

We now describe our estimation procedure when the design is unbalanced. Let $\varepsilon_i = (\varepsilon_{i1}, \varepsilon_{i2}, \ldots, \varepsilon_{iJ_i})^T$ and $\mathrm{cov}(\varepsilon_i) = \boldsymbol{\Sigma}_i$. There exists a lower triangle matrix $\boldsymbol{\Phi}_i$ with 1's on the main diagonal such that $\mathrm{cov}(\boldsymbol{\Phi}_i\varepsilon_i) = \boldsymbol{\Phi}_i\boldsymbol{\Sigma}_i\boldsymbol{\Phi}_i^T = \mathbf{D}_i$, where $\mathbf{D}_i$ is a diagonal matrix. Letting $\phi_{jl}^{(i)}$ be the negative of the (j,l)-element of $\boldsymbol{\Phi}_i$, we have

$$y_{ij} = g_0(\boldsymbol{\alpha}^T\boldsymbol{x}_{ij}) + \boldsymbol{g}^T(u_{ij})\boldsymbol{z}_{ij} + \sum_{k=1}^{j-1}\phi_{jk}^{(i)}\breve{\varepsilon}_{ik} + e_{ij},$$
$$i = 1, \ldots, n, \quad j = 1, \ldots, J_i,$$

where e_{ij}'s are uncorrelated. We estimate $\boldsymbol{\Sigma}_i$ by a working covariance matrix using the same method proposed in

[9]. The working covariance matrix provides estimates of $\mathbf{\Phi}_i$ and $\mathbf{D}_i$. With $\mathbf{\Phi}_i$ and $\mathbf{D}_i$ being fixed at these estimates, the parameters $\boldsymbol{\alpha}$ and functions g_r for $r = 0, \ldots, p$ can be estimated by similar steps described in this section with straightforward modifications.

3. ASYMPTOTIC PROPERTIES

In this section we study asymptotic properties of the estimators proposed in Section 2 for the balanced case. Covariates as random variables are denoted as upper case letters such as $\mathbf{X}$, $\mathbf{Z}$ and U, while their realizations are denoted as lower case letters.

The following conditions are needed:

(C1) For every $r = 0, 1, \ldots, p$, $g_r(\cdot)$ has a Lipschitz continuous second derivative.

(C2) The kernels $K_i(\cdot)$, $i = 1, \ldots, 4$ are bounded symmetric density functions with bounded support and satisfy the Lipschitz condition.

(C3) The marginal density of $\boldsymbol{\alpha}^T\mathbf{X}_{ij}$, $f_{1j}(\cdot)$, is Lipschitz continuous, bounded away from 0 and has a continuous second derivative on its support.

(C4) The marginal density of U_{ij}, $f_{2j}(\cdot)$, has a bounded support Ω, and is Lipschitz continuous and bounded away from 0 on Ω. The U_{ij}'s within each subject are allowed to be correlated.

(C5) There exists an $s > 2$ such that $\mathrm{E}\|\mathbf{F}_{1j}\|^s < \infty$, $\forall j$, and for some $\xi > 0$ such that $n^{1-2s^{-1}-2\xi}h \longrightarrow \infty$.

(C6) $\sup_{s\in\Omega}|\breve{g}_r(s) - g_r(s)| = o_p(n^{-1/4})$, where $\breve{g}_r(s)$ is obtained by local linear regression under the assumption that the random errors are independent and identically distributed.

(C7) Every entry of $\mathbf{Z}_{ij}$ will not degenerate to a constant. $\mathrm{E}(\mathbf{Z_{ij}}^T|\boldsymbol{\alpha}^T\mathbf{X}_{ij} = t,\ U_{ij} = u)$, $\mathrm{E}(\mathbf{Z}_{ij}\mathbf{Z}_{ij}^T|\boldsymbol{\alpha}^T\mathbf{X}_{ij} = t,\ U_{ij} = u)$ and $\mathrm{E}(\mathbf{H}_{ij}\tau_2(U_{ij})\tau_1^{-1}(U_{ij})\mathbf{Z}_{ij}^T|U_{ij})$ are all Lipschitz continuous.

Theorem 3.1. *When $\boldsymbol{\alpha}$ and $\boldsymbol{\phi}$ are known constants or estimated with the order $O_p(N^{-1/2})$, under the regularity conditions (C1)–(C7), $h_1/h_3 \to 0$, $h_2/h_3 \to 0$, $Nh_1h_2 \to \infty$ and $h_3 = Cn^{-1/5}$, as $N \to \infty$ we have*

$$\sqrt{Nh_3}\left(\hat{g}_0(t) - g_0(t) - \frac{1}{2}h_3^2\mu_2(K_3)g_0''(t)\right) \xrightarrow{\text{d}} N\left(0, \frac{\nu_0(K_3)}{\tau_1(t)}\right), \tag{11}$$

where $\tau_1(t) = \frac{1}{J}\sum_{j=1}^J f_{1j}(t)/d_{ij}^2$.

Remark 1. The asymptotic distribution of $\hat{g}_0(t)$ in (11) is the same as the estimator for the nonparametric regression model. See, Theorem 1(b) of [32].

Theorem 3.2. *When $\boldsymbol{\alpha}$ and $\boldsymbol{\phi}$ are known constants or estimated with the order $O_p(N^{-1/2})$, under the regularity conditions (C1)–(C7), $h_1/h_4 \to 0$, $h_2/h_4 \to 0$, $Nh_1h_2 \to \infty$ and $h_4 = Cn^{-1/5}$, as $N \to \infty$ we have*

$$\sqrt{Nh_4}\left(\hat{\mathbf{g}}(u) - \mathbf{g}(u) - \frac{1}{2}h_4^2\mu_2(K_4)\mathbf{g}''(u)\right) \xrightarrow{\text{d}} N\left(0, \nu_0(K_4)\tau_2^{-1}(u)\right), \tag{12}$$

where $\tau_2(u) = \frac{1}{J}\sum_{j=1}^J E(\mathbf{Z}_{ij}\mathbf{Z}_{ij}^T \mid U_{ij} = u)f_{2j}(u)/d_{ij}^2$.

Remark 2. The asymptotic distribution of $\hat{\mathbf{g}}(u)$ in (12) is same as the estimator for the varying-coefficient model. Note that $\tau_2(u)$ is independent of i since $\mathbf{Z}_{ij}$ and U_{ij} with different i and fixed j are iid random samples.

Theorem 3.3. *Suppose that the conditions (C1)–(C7) hold. If $h_i/h_j \to 0$, $Nh_j^2 \to \infty$, $Nh_j^4 \to 0$, $i = 1, 2$; $j = 3, 4$, as $N \to \infty$, we have*

$$\sqrt{N}\begin{pmatrix} \hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha} \\ \hat{\boldsymbol{\phi}} - \boldsymbol{\phi} \end{pmatrix} \xrightarrow{\text{d}} N(\mathbf{0}, \mathbf{B}^{-1}\mathbf{\Sigma}\mathbf{B}^{-1}), \tag{13}$$

where

$$\mathbf{\Sigma} = E\Big\{\frac{1}{J}\sum_{j=1}^J d_{ij}^{-2}\Big(E(\mathbf{H}_{ij}\tau_3(U_{ij})\tau_2^{-1}(U_{ij})\mathbf{Z}_{ij} \mid U_{ij})\Big)^{\otimes 2}\Big\},$$

$$\mathbf{B} = \mathrm{E}\left\{\frac{1}{J}\sum_{j=1}^J d_{ij}^{-2}\mathbf{H}_{ij}^{\otimes 2}\right\}, \quad \mathbf{H_{ij}} = \begin{pmatrix} g_0'(\boldsymbol{\alpha}^T\mathbf{X_{ij}})\mathbf{X_{ij}} \\ \mathbf{F_{ij}} \end{pmatrix},$$

$$\tau_3(u) = \frac{1}{J}\sum_{j=1}^J E(\mathbf{Z}_{ij}^T \mid U = u)f_{2j}(u)/d_{ij}^2.$$

Remark 3. From (13) we see that the estimator $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\phi}}$ follow the usual asymptotic distribution, and their variances are associated with the nonparametric components g_0 and $\boldsymbol{g}$. Moreover, Theorem 3 has an important restriction on the bandwidths h_3 and h_4. To estimate $\boldsymbol{\alpha}$ and $\boldsymbol{\phi}$ at the rate $N^{-1/2}$, one must undersmooth the nonparametric functions g_0 and $\boldsymbol{g}$. The need for undersmoothing so as to obtain the usual rates of convergence is standard in the kernel literature and has analogs in the spline literature.

4. SIMULATION STUDIES

In this section we evaluate the finite sample performance of the proposed estimation methods. We consider the following VCSIM:

$$y_{ij} = g_0(\alpha_1 x_{ij1} + \alpha_2 x_{ij2}) + g_1(u_{ij})z_{ij1} + g_2(u_{ij})z_{ij2} + \varepsilon_{ij},$$
$$i = 1, 2, \cdots, n, \quad j = 1, 2, \cdots, J_i,$$

where $g_0(t) = t + 8\exp(-16t^2)$, $g_1(u) = 5\sin(2\pi u)$, $g_2(u) = 5\cos(2\pi u)$, x_{ij1} and x_{ij2} are iid realizations of $\mathrm{U}(-0.5, 0.5)$, z_{ij1} and z_{ij2} are iid realizations of $\mathrm{N}(0, 1)$, u_{ij} are iid realizations of $\mathrm{U}(0, 1)$. For the sake of comparison, we consider the following two cases:

Table 1. Averages, standard errors (SE), biases and MSEs of estimates from the proposed method $\hat{\alpha}_1$ and $\hat{\alpha}_2$ and [28]'s independent estimates $\breve{\alpha}_1$ and $\breve{\alpha}_2$ for Case I

n	Estimation	$\breve{\alpha}_1$	$\hat{\alpha}_1$	$\breve{\alpha}_2$	$\hat{\alpha}_2$
50	Average	0.6194	0.7142	1.0614	1.1154
	SE	0.3215	0.5300	0.3117	0.4151
	Bias	0.1722	0.2670	0.1670	0.2210
	MSE	0.1330	0.3522	0.1250	0.2211
100	Average	0.3531	0.5605	0.7983	1.0137
	SE	0.2152	0.3194	0.2761	0.3026
	Bias	-0.0941	0.1133	-0.0961	0.1193
	MSE	0.0552	0.1149	0.0855	0.1058
200	Average	0.5206	0.3522	0.9746	0.9914
	SE	0.1511	0.2085	0.1473	0.2123
	Bias	0.0734	-0.0950	0.0802	0.0970
	MSE	0.0282	0.0525	0.0281	0.0545

Table 2. Averages, standard errors (SE), biases and MSEs of estimates from the proposed method $\hat{\alpha}_1$ and $\hat{\alpha}_2$, oracle estimates $\hat{\alpha}_1^*$ and $\hat{\alpha}_2^*$ and [28]'s independent estimates $\breve{\alpha}_1$ and $\breve{\alpha}_2$ for Case II

n	Estimation	$\breve{\alpha}_1$	$\hat{\alpha}_1$	$\hat{\alpha}_1^*$	$\breve{\alpha}_2$	$\hat{\alpha}_2$	$\hat{\alpha}_2^*$
50	Average	0.3821	0.4138	0.4175	0.8182	0.9340	0.8628
	SE	0.7199	0.2167	0.1252	0.7102	0.2124	0.1216
	Bias	-0.0651	-0.0334	-0.0297	-0.0762	0.0396	-0.0316
	MSE	0.5225	0.0481	0.0166	0.5102	0.0467	0.0158
100	Average	0.4929	0.4695	0.4592	0.9403	0.8671	0.9102
	SE	0.4145	0.1097	0.0945	0.4065	0.1064	0.0996
	Bias	0.0457	0.0223	0.0120	0.0459	-0.0273	0.0158
	MSE	0.1739	0.0125	0.0091	0.1673	0.0121	0.0102
200	Average	0.4143	0.4617	0.4564	0.9248	0.8827	0.8847
	SE	0.2523	0.0740	0.0659	0.2606	0.0691	0.0713
	Bias	-0.0329	0.0145	0.0092	0.0304	-0.0117	-0.0097
	MSE	0.0647	0.0057	0.0044	0.0688	0.0049	0.0052

(a) case I, the ε_{ij}'s are independent and $\sigma^2 = 4$;

(b) case II, the random error ε_{ij} are generated from an AR(1) model with variance $\sigma^2 = 4$ and autocorrelation $\rho = 0.9$.

We set $\alpha_1 = 1/\sqrt{5}$, $\alpha_2 = 2/\sqrt{5}$. Three sample sizes for the number of subjects will be considered: $n = 50$, $n = 100$ and $n = 200$. We conduct simulations under two situations: balanced and unbalanced designs. For the balanced design, $J_i = 5$. For the unbalanced design, J_i's are iid realizations of a random variables with distribution $p(j) = 0.1$ for $j = 1, \ldots, 10$. To save space we present results from balanced design only. Results from the unbalanced design are similar. All simulations are replicated for 500 times. The good results presented in this section are typical of the performances we observed over several runs with different choices of function, sample sizes.

To investigate the effect of correlated errors on the estimation, we compute oracle estimates $\hat{\alpha}_1^*$ and $\hat{\alpha}_2^*$ using the true covariance matrix. These oracle estimates serve as a benchmark for the comparison. Moreover, to examine the efficiency of our proposed estimators, we also compute the estimates $\breve{\alpha}_1$ and $\breve{\alpha}_2$ via the method of [28] pretending that random errors are independent.

Performance of estimates of α_1 and α_2 are assessed using bias and mean squared error (MSE). Performance of estimates of nonparametric functions g_k are assessed using the root of average squared errors (RASE) defined as

$$\mathrm{RASE}\left(\hat{g}_k\right) = \left[\frac{1}{100}\sum_{j=1}^{100}\left(\hat{g}_k(t_j) - g_k(t_j)\right)^2\right]^{\frac{1}{2}}, \quad k = 0, 1, 2,$$

where $\{t_j, j = 1, \ldots, 100\}$ are the grid points at which the functions $\hat{g}_k(\cdot)$ are evaluated.

Table 3. Summary of the RASE (Standard Deviation) for Case I

Coef. curve	n	Wong (2008)	New
$g_0(t)$	50	0.5883(0.1935)	0.6488(0.2107)
	100	0.3829(0.1594)	0.4188(0.1700)
	200	0.3105(0.0601)	0.3401(0.0744)
$g_1(u)$	50	0.4204(0.1036)	0.4755(0.1172)
	100	0.3096(0.0710)	0.3355(0.0776)
	200	0.2100(0.0417)	0.2414(0.0473)
$g_2(u)$	50	0.4152(0.1083)	0.4809(0.1271)
	100	0.3011(0.0684)	0.3394(0.0762)
	200	0.2083(0.0400)	0.2520(0.0577)

Table 4. Summary of the RASE (Standard Deviation) for Case II

Coef. curve	n	Wong (2008)	New	Oracle
$g_0(t)$	50	0.8757(0.2607)	0.4526(0.1491)	0.3853(0.1405)
	100	0.6421(0.1900)	0.3326(0.1117)	0.2619(0.1058)
	200	0.4128(0.1162)	0.2448(0.0683)	0.1809(0.0643)
$g_1(u)$	50	0.6098(0.1624)	0.3959(0.1090)	0.2943(0.0873)
	100	0.4803(0.1282)	0.2515(0.0582)	0.1893(0.0438)
	200	0.2173(0.0418)	0.1786(0.0378)	0.1499(0.0299)
$g_2(u)$	50	0.6212(0.1619)	0.4050(0.1227)	0.3015(0.0983)
	100	0.4952(0.1065)	0.2596(0.0625)	0.1844(0.0418)
	200	0.2150(0.1017)	0.1296(0.0391)	0.1283(0.0351)

Tables 1 and 2 list averages, standard errors (SE), biases and MSEs of estimates of α_1 and α_2 for the case I and case II designs respectively. The sample mean and standard deviation of the RASEs for the case I and case II designs are summarized in Tables 3–4 respectively. We can see that the new and oracle methods have smaller MSE, RASE than the independence model when the data are correlated and the gain in efficiency can be achieved even for moderate sample size. Moreover, our proposed estimation methods performed very well and as well as the oracle method even for moderate sample sizes. As expected, the performance improves as the number of subjects increases. In particular, for independent data (case I), our proposed method does not lose much

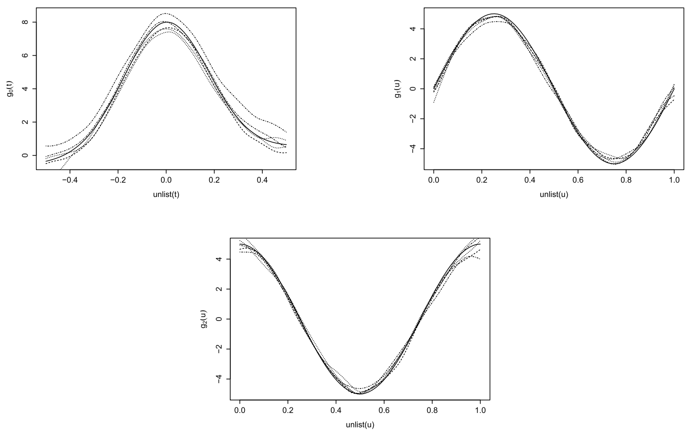

Figure 1. Balanced design, plots of the true functions (solid lines) together with their estimates correoponding to the minimum, lower quartile, median, upper quartile and maximum RASEs (dotted and dashed lines) of $g_0(\cdot)$, $g_1(\cdot)$ and $g_2(\cdot)$.

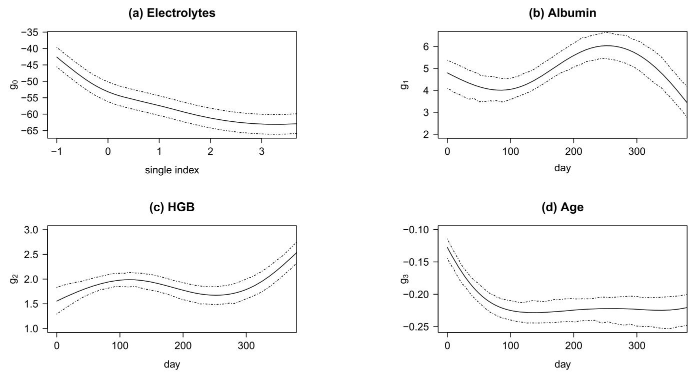

Figure 2. Estimates of nonparametric functions (solid lines) and their 95% pointwise confidence intervals (dotted lines).

Table 5. Estimates of index coefficient estimates

	α_1	α_2	α_3
Average	0.5429	-0.4377	0.2644
SE	0.1925	0.2184	0.1718

efficiency for estimating the correlation structure when compared with the independence model.

For the sample size $n = 100$, Figures 1 shows the true functions together with their estimates corresponding to the minimum, lower quartile, median, upper quartile and maximum RASEs for the balanced designs. The estimates are good even for the ones with the largest RASEs.

5. LONGITUDINAL PROGRESSION OF GFR AMONG CKD PATIENTS

Chronic kidney disease (CKD) is a worldwide public health problem. More than 23 million people in the United States have CKD. CKD patients are at increased risk for kidney failure, cardiovascular disease, hospitalization and mortality. CKD is typically characterized by progressive loss of renal function as evidenced by declining glomerular filtration rate (GFR) over time. The trajectory of GFR has been well-studied in the medical literature [2], and it has been noted recently that the trajectory can be quite complex and there is a large heterogeneity among patients [19]. The pattern of GFR trajectory may be related to demographic and clinical characteristics. In spite of its importance, there is only one study by [20] to investigate time-varying risk factors. [20] divided GFR profile for each patient into two periods, stable and rapidly declining, and then compared risk factors between these two periods. We note that not all patients follow the two periods stable and rapidly declining pattern. In addition, the approach in [20] does not investigate the relationship between GFR and time-varying risk factors directly.

We consider a data set consisting of 727 observations over time from 85 patients. All patients started at stage 2 CKD. At each time point we have measurement of GFR ($\mathrm{mL/min/1.73\,m^2}$), albumin (g/dL), hemoglobin (HGB) (g/dL), as well as electrolytes calcium (mg/dL), chloride (mg/dL) and phosphate (mg/dL). GFR was estimated using the CKD-EPI equation. Age of each patient at the start of the study is also available. We use this data set to show how the VCSIM may be used to investigate potential effects of time-varying risk factors as well as time-varying effects. This example is intended as an illustration of the usefulness of the new methodology rather than a formal data analysis.

We are interested in the relationship between GFR and the time-independent covariate age as well as the time-varying covariates albumin, HGB, calcium, chloride and phosphate. We consider a single index model for electrolytes to reduce dimension. We will allow dynamic relationships between GFR and age, albumin and HGB. Specifically, we consider the following VCSIM

$$
\begin{aligned}
y_{ij} &= g_0(\alpha_1 x_{ij1} + \alpha_2 x_{ij2} + \alpha_3 x_{ij3}) + g_1(t_{ij}) z_{ij1} \\
&\quad + g_2(t_{ij}) z_{ij2} + g_3(t_{ij}) z_{i3} + \varepsilon_{ij}, \\
i &= 1, 2, \cdots, n, \quad j = 1, 2, \cdots, J_i,
\end{aligned}
$$

where y_{ij} is the GFR at time t_{ij} from the ith patient, x_{ij1}, x_{ij2}, x_{ij3}, z_{ij1} and z_{ij2} are measurements of calcium, chloride, phosphate, albumin and HGB respectively at time t_{ij}, and z_{i3} is the age of the ith patient. Variables z_{ij1}, z_{ij2} and z_{i3} are centered to facilitate the interpretation. The design is unbalanced and the proposed method for the unbalanced design was used to fit the data.

Table 5 lists the estimates and standard errors of α_1, α_2 and α_3. The estimates and 95% pointwise confidence intervals of nonparametric functions g_0, g_1, g_2 and g_3 are shown in Figure 2. The bootstrap method with 200 replications was used to compute the standard errors in Table 5 and 95% confidence intervals in Figure 2.

Figure 2(a) indicates that GFR decreases as the index of electrolytes increases. The index is positively associated with calcium and phosphate and negatively associated with chloride. Figure 2(b) and Figure 2(c) indicate that GFR increases as albumin or HGB increase. The positive association between GFR and albumin become stronger during 200 and 300 days. The positive association between GFR and HGB grows stronger over time. Figure 2(d) indicates that GFR decreases as age increases and this negatively association grows stronger over time.

APPENDIX A. PROOFS

We show proofs of Theorems 1 and 2 under the assumption that $\boldsymbol{\alpha}$ and $\boldsymbol{\phi}$ are known constants. When $\boldsymbol{\alpha}$ and $\boldsymbol{\phi}$ are estimated to the order of $N^{-1/2}$, the proofs can be completed by noting that $\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha} = O_p(N^{-1/2})$ and $\hat{\boldsymbol{\phi}} - \boldsymbol{\phi} = O_p(N^{-1/2})$. The complete proof is omitted to save space. In addition, since d_{ij} can be estimated by a parametric rate, we shall assume that d_{ij} is known in our proof, without loss of generality.

Proof of Theorem 1. Note that $1/N(\tilde{\mathbf{X}}^T \boldsymbol{W}_1 \tilde{\mathbf{X}}) = \tau_1(t)\text{diag}(1, \mu_2(K_3))(1 + o_p(1))$, then

$$\begin{aligned}
\hat{g}_0(t) &= \hat{g}_{01}(t) + \hat{g}_{02}(t),\\
g_0(t) &= \frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{J}\frac{1}{\tau_1(t)}K_{3,h_3}(\boldsymbol{\alpha}^T\boldsymbol{x}_{ij} - t)/d_{ij}^2(g_0(t)\\
&\quad + g_0^{'}(t)(\boldsymbol{\alpha}^T\boldsymbol{x}_{ij} - t))(1 + o_p(1)),
\end{aligned}$$

where

$$\begin{aligned}
\hat{g}_{01}(t) &= \frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{J}\frac{1}{\tau_1(t)}K_{3,h_3}(\boldsymbol{\alpha}^T\boldsymbol{x}_{ij} - t)/d_{ij}^2\\
&\quad (y_{ij}^* - \mathbf{g}^T(u_{ij})\boldsymbol{z}_{ij})(1 + o_p(1)),\\
\hat{g}_{02}(t) &= \frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{J}\frac{1}{\tau_1(t)}K_{3,h_3}(\boldsymbol{\alpha}^T\boldsymbol{x}_{ij} - t)/d_{ij}^2\\
&\quad (\mathbf{g}(u_{ij}) - \breve{\mathbf{g}}(u_{ij}))^T\boldsymbol{z}_{ij}(1 + o_p(1)).
\end{aligned}$$

Furthermore,

$$\hat{g}_0(t) - g_0(t) = (\Delta_1 + \Delta_2)(1 + o_p(1)) + \hat{g}_{02}(t),$$

where

$$\begin{aligned}
\Delta_1 &= \frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{J}\frac{1}{\tau_1(t)}K_{3,h_3}(\boldsymbol{\alpha}^T\boldsymbol{x}_{ij} - t)/d_{ij}^2\\
&\quad (g_0(\boldsymbol{\alpha}^T\boldsymbol{x}_{ij}) - g_0(t) - g_0^{'}(t)(\boldsymbol{\alpha}^T\boldsymbol{x}_{ij} - t)),\\
\Delta_2 &= \frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{J}\frac{1}{\tau_1(t)}K_{3,h_3}(\boldsymbol{\alpha}^T\boldsymbol{x}_{ij} - t)/d_{ij}^2 e_{ij}.
\end{aligned}$$

Then the proof of Theorem 1 is completed by noting the fact that $\Delta_1 = \frac{1}{2}g_0^{''}(t)\mu_2(K_3)h_3^2(1 + o_p(1))$, $\sqrt{Nh_3}\Delta_2 \xrightarrow{d} \text{N}(0, \nu_0(K_3)/\tau_1(t))$, and $\sqrt{Nh_3}\hat{g}_{02}(t) = o_p(1)$. □

Proof of Theorem 2. Note that $1/N(\tilde{\mathbf{U}}^T \boldsymbol{W}_2 \tilde{\mathbf{U}}) = \text{diag}(\tau_2(u), \tau_2(u)\mu_2(K_4))(1 + o_p(1))$, then

$$\begin{aligned}
\hat{\mathbf{g}}(u) &= \hat{\mathbf{g}}_1(u) + \hat{\mathbf{g}}_2(u),\\
\mathbf{g}(u) &= \frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{J}\tau_2^{-1}(u)\boldsymbol{z}_{ij}K_{4,h_4}(u_{ij} - u)/d_{ij}^2\\
&\quad (\mathbf{g}(u) + \mathbf{g}^{'}(u)(u_{ij} - u))^T\boldsymbol{z}_{ij}(1 + o_p(1)),
\end{aligned}$$

where

$$\begin{aligned}
\hat{\mathbf{g}}_1(u) &= \frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{J}\tau_2^{-1}(u)\boldsymbol{z}_{ij}K_{4,h_4}(u_{ij} - u)/d_{ij}^2\\
&\quad (y_{ij}^* - g_0(\boldsymbol{\alpha}^T\boldsymbol{x}_{ij}))(1 + o_p(1)),\\
\hat{\mathbf{g}}_2(u) &= \frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{J}\tau_2^{-1}(u)\boldsymbol{z}_{ij}K_{4,h_4}(u_{ij} - u)/d_{ij}^2\\
&\quad (g_0(\boldsymbol{\alpha}^T\boldsymbol{x}_{ij}) - \breve{g}_0(\boldsymbol{\alpha}^T\boldsymbol{x}_{ij}))(1 + o_p(1)).
\end{aligned}$$

Furthermore,

$$\hat{\mathbf{g}}(u) - \mathbf{g}(u) = (Q_1 + Q_2)(1 + o_p(1)) + \hat{\mathbf{g}}_2(u),$$

where

$$\begin{aligned}
Q_1 &= \frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{J}\tau_2^{-1}(u)\boldsymbol{z}_{ij}K_{4,h_4}(u_{ij} - u)/d_{ij}^2\\
&\quad (\mathbf{g}(u_{ij}) - \mathbf{g}(u) - \mathbf{g}^{'}(u)(u_{ij} - u))^T\boldsymbol{z}_{ij},\\
Q_2 &= \frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{J}\tau_2^{-1}(u)\boldsymbol{z}_{ij}K_{4,h_4}(u_{ij} - u)/d_{ij}^2 e_{ij}.
\end{aligned}$$

Then the proof of Theorem 2 is completed by noting the fact that $Q_1 = \frac{1}{2}\mathbf{g}^{''}(u)\mu_2(K_4)h_4^2(1 + o_p(1))$, $\sqrt{Nh_4}Q_2 \xrightarrow{d} \text{N}(0, \nu_0(K_4)\tau_2^{-1}(u))$, and $\sqrt{Nh_3}\hat{\mathbf{g}}_2(u) = o_p(1)$. □

Proof of Theorem 3. With λ as the Lagrange multiplier, it follows from (9) that $(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\phi}})$ is the solution to

$$\begin{aligned}
&\frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{J}d_{ij}^{-2}(y_{ij} - \hat{g}_0(\hat{\boldsymbol{\alpha}}^T\mathbf{x}_{ij}; \hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\phi}}) - \hat{\mathbf{g}}^T(u_{ij}; \hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\phi}})\mathbf{z}_{ij}\\
&\quad - \hat{\boldsymbol{\phi}}^T\mathbf{F}_{ij})\begin{pmatrix}\hat{g}_0'(\hat{\boldsymbol{\alpha}}^T\mathbf{x}_{ij}; \hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\phi}})\mathbf{x}_{ij}\\ \mathbf{F}_{ij}\end{pmatrix} + \lambda\begin{pmatrix}\hat{\boldsymbol{\alpha}}\\ 0\end{pmatrix} = 0.
\end{aligned}$$

Using Taylor expansion, we have

$$0 \;=\; \lambda\begin{pmatrix}\hat{\boldsymbol{\alpha}}\\ 0\end{pmatrix} + \frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{J}d_{ij}^{-2}\mathbf{H}_{ij}e_{ij}$$

$$-\frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{J} d_{ij}^{-2}\mathbf{H}_{ij}(\hat{\boldsymbol{\phi}}-\boldsymbol{\phi})^T\mathbf{F}_{ij}$$
$$-\frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{J} d_{ij}^{-2}\mathbf{H}_{ij}[\hat{g}_0(\hat{\boldsymbol{\alpha}}^T\mathbf{x}_{ij};\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\phi}})-g_0(\boldsymbol{\alpha}^T\mathbf{x}_{ij})]$$
$$-\frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{J} d_{ij}^{-2}\mathbf{H}_{ij}[\hat{\mathbf{g}}(u_{ij};\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\phi}})-\mathbf{g}(u_{ij})]^T\mathbf{z}_{ij}$$
$$+o_P(1), \tag{14}$$

where $\mathbf{H}_{ij}=\begin{pmatrix} g_0'(\boldsymbol{\alpha}^T\mathbf{x}_{ij})\mathbf{x}_{ij} \\ \mathbf{F}ij \end{pmatrix}$. Note that

$$\begin{aligned}
&\hat{g}_0(\hat{\boldsymbol{\alpha}}^T\mathbf{x}_{ij};\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\phi}})-g_0(\boldsymbol{\alpha}^T\mathbf{x}_{ij})\\
=\ &\hat{g}_0(\hat{\boldsymbol{\alpha}}^T\mathbf{x}_{ij};\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\phi}})-\hat{g}_0(\boldsymbol{\alpha}^T\mathbf{x}_{ij};\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\phi}})\\
&+\hat{g}_0(\boldsymbol{\alpha}^T\mathbf{x}_{ij};\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\phi}})-g_0(\boldsymbol{\alpha}^T\mathbf{x}_{ij})\\
=\ &\hat{g}_0'(\boldsymbol{\alpha}^T\mathbf{x}_{ij};\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\phi}})(\hat{\boldsymbol{\alpha}}-\boldsymbol{\alpha})^T\mathbf{x}_{ij}\\
&+\hat{g}_0(\boldsymbol{\alpha}^T\mathbf{x}_{ij};\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\phi}})-g_0(\boldsymbol{\alpha}^T\mathbf{x}_{ij})+o_P(N^{-1/2})\\
=\ &g_0'(\boldsymbol{\alpha}^T\mathbf{x}_{ij})(\hat{\boldsymbol{\alpha}}-\boldsymbol{\alpha})^T\mathbf{x}_{ij}+\hat{g}_0(\boldsymbol{\alpha}^T\mathbf{x}_{ij};\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\phi}})\\
&-g_0(\boldsymbol{\alpha}^T\mathbf{x}_{ij})+o_P(N^{-1/2}).
\end{aligned}\tag{15}$$

Substituting (15) into (14), we have

$$\begin{aligned}
&\frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{J} d_{ij}^{-2}\mathbf{H}_{ij}\mathbf{H}_{ij}^T\begin{pmatrix}\hat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}\\ \hat{\boldsymbol{\phi}}-\boldsymbol{\phi}\end{pmatrix}\\
=\ &\lambda\begin{pmatrix}\hat{\boldsymbol{\alpha}}\\ 0\end{pmatrix}+\frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{J} d_{ij}^{-2}\mathbf{H}_{ij}e_{ij}\\
&-\frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{J} d_{ij}^{-2}\mathbf{H}_{ij}[\hat{g}_0(\boldsymbol{\alpha}^T\mathbf{x}_{ij};\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\phi}})-g_0(\boldsymbol{\alpha}^T\mathbf{x}_{ij})]\\
&-\frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{J} d_{ij}^{-2}\mathbf{H}_{ij}[\hat{\mathbf{g}}(u_{ij};\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\phi}})-\mathbf{g}(u_{ij})]^T\boldsymbol{z}_{ij}\\
&+o_P(N^{-1/2}).
\end{aligned}\tag{16}$$

By the proofs of Theorem 1 and 2, we have

$$\hat{g}_0(t;\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\phi}})-g_0(t)=\Big\{\frac{1}{2}g_0^{''}(t)\mu_2(K_3)h_3^2$$
$$+\frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{J}\frac{1}{\tau_1(t)}K_{3,h_3}(\boldsymbol{\alpha}^T\boldsymbol{x}_{ij}-t)/d_{ij}^2e_{ij}\Big\}(1+o_p(1)),$$
$$\hat{\mathbf{g}}(u;\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\phi}})-\mathbf{g}(u)=\Big\{\frac{1}{2}\mathbf{g}^{''}(u)h_4^2\mu_2(K_4)$$
$$+\frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{J}\tau_2^{-1}(u)\boldsymbol{z}_{ij}K_{4,h_4}(u_{ij}-u)/d_{ij}^2e_{ij}\Big\}(1+o_p(1)).$$

Consequently,

$$\begin{aligned}
&\hat{g}_0(\boldsymbol{\alpha}^T\mathbf{x}_{ij};\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\phi}})-g_0(\boldsymbol{\alpha}^T\mathbf{x}_{ij})\\
=\ &\Big\{\frac{1}{2}g_0^{''}(\boldsymbol{\alpha}^T\mathbf{x}_{ij})\mu_2(K_3)h_3^2+\frac{1}{N}\sum_{k=1}^{n}\sum_{l=1}^{J}\frac{1}{\tau_1(\boldsymbol{\alpha}^T\mathbf{x}_{ij})}\\
&K_{3,h_3}(\boldsymbol{\alpha}^T\boldsymbol{x}_{kl}-\boldsymbol{\alpha}^T\mathbf{x}_{ij})/d_{kl}^2e_{kl}\Big\}(1+o_p(1)),\\
&\hat{\mathbf{g}}(u_{ij};\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\phi}})-\mathbf{g}(u_{ij})\\
=\ &\Big\{\frac{1}{2}\mathbf{g}^{''}(u_{ij})h_4^2\mu_2(K_4)+\frac{1}{N}\sum_{k=1}^{n}\sum_{l=1}^{J}\tau_2^{-1}(u_{ij})\boldsymbol{z}_{kl}\\
&K_{4,h_4}(u_{kl}-u_{ij})/d_{kl}^2e_{kl}\Big\}(1+o_p(1)),
\end{aligned}$$

which implies that, by $Nh_3^4\longrightarrow 0$ and $Nh_4^4\longrightarrow 0$, (16) becomes

$$\begin{aligned}
&\frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{J}\mathbf{H}_{ij}\mathbf{H}_{ij}^T\begin{pmatrix}\hat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}\\ \hat{\boldsymbol{\phi}}-\boldsymbol{\phi}\end{pmatrix}\\
=&\lambda\begin{pmatrix}\hat{\boldsymbol{\alpha}}\\ 0\end{pmatrix}+\frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{J} d_{ij}^{-2}\mathbf{H}_{ij}e_{ij}\\
&-\frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{J}\Big\{\frac{1}{N}\sum_{k=1}^{n}\sum_{l=1}^{J}\frac{\mathbf{H}_{ij}}{\tau_1(\boldsymbol{\alpha}^T\mathbf{x}_{ij})d_{ij}^2}\\
&K_{3,h_3}(\boldsymbol{\alpha}^T\boldsymbol{x}_{kl}-\boldsymbol{\alpha}^T\mathbf{x}_{ij})/d_{kl}^2\\
&+\frac{1}{N}\sum_{k=1}^{n}\sum_{l=1}^{J}\frac{\mathbf{H}_{ij}}{d_{ij}^2}\boldsymbol{z}_{kl}^T\tau_2^{-1}(u_{ij})\boldsymbol{z}_{ij}\\
&K_{4,h_4}(u_{kl}-u_{ij})/d_{kl}^2\Big\}e_{kl}+o_P(N^{-1/2}).
\end{aligned}\tag{17}$$

Since the term in the big brackets converges to $d_{ij}^{-2}\mathbf{H}_{ij}+d_{ij}^{-2}\mathrm{E}(\mathbf{H}_{ij}\tau_3(U_{ij})\tau_2^{-1}(U_{ij})\mathbf{Z}_{ij}\mid U_{ij})$ uniformly in i and j, then (17) reduces to

$$\begin{aligned}
&\frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{J} d_{ij}^{-2}\mathbf{H}_{ij}\mathbf{H}_{ij}^T\begin{pmatrix}\hat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}\\ \hat{\boldsymbol{\phi}}-\boldsymbol{\phi}\end{pmatrix}\\
=\ &-\frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{J} d_{ij}^{-2}\mathrm{E}(\mathbf{H}_{ij}\tau_3(U_{ij})\tau_2^{-1}(U_{ij})\mathbf{Z}_{ij}\mid U_{ij})e_{ij}\\
&+\lambda\begin{pmatrix}\hat{\boldsymbol{\alpha}}\\ 0\end{pmatrix}+o_P(N^{-1/2}).
\end{aligned}$$

Multiplying the above equation by $P_{\boldsymbol{\alpha}}=\mathrm{diag}(\mathbf{I}-\boldsymbol{\alpha}\boldsymbol{\alpha}^T,\mathbf{I})$ where $\mathbf{I}$ is a $q\times q$ identity matrix, we obtain

$$\begin{aligned}
&P_{\boldsymbol{\alpha}}\frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{J} d_{ij}^{-2}\mathbf{H}_{ij}\mathbf{H}_{ij}^T\begin{pmatrix}\hat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}\\ \hat{\boldsymbol{\phi}}-\boldsymbol{\phi}\end{pmatrix}\\
=\ &-P_{\boldsymbol{\alpha}}\frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{J} d_{ij}^{-2}\mathrm{E}(\mathbf{H}_{ij}\tau_3(U_{ij})\tau_2^{-1}(U_{ij})\mathbf{Z}_{ij}\mid U_{ij})e_{ij}\\
&+o_P(N^{-1/2}).
\end{aligned}\tag{18}$$

The proof of Theorem 3 is completed by applying the central limit theorem to (18). □

ACKNOWLEDGEMENTS

We sincerely thank the associate editor and two referees for their valuable comments that have led to a greatly improved presentation of our work. We also thank Fresenius Medical Care North America for providing deidentified CKD registry data and their collaboration in this analysis.

Received 3 September 2015

REFERENCES

[1] 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. MR2460484

[2] Boucquemont, J., Heinze, G., Jager, K., Oberbauer, R. and Leffondre, K. (2014). Regression methods for investigating risk factors of chronic kidney disease outcomes: the state of the art. *BMC Nephrology* **15**.

[3] Cai, Z., Fan, J. and Li, R. (2000). Efficient estimation and inferences for varying-coefficient models. *Journal of the American Statistical Association* **95** 888–902. MR1804446

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

[5] Chen, D., Hall, P. and Müller, H. G. (2011). Single and multiple index functional regression models with nonparametric link. *The Annals of Statistics* **39** 1720–1747. MR2850218

[6] Diggle, P., Heagerty, P., Liang, K. Y. and Zeger, S. (2002). *Analysis of Longitudinal Data*, Oxford University Press. MR2049007

[7] Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications*, Chapman and Hall, London. MR1383587

[8] Fan, J. and Huang, T. (2005). Profile likelihood inferences on semiparametric varying-coefficient partially linear models. *Bernoulli* **11** 1031–1057. MR2189080

[9] Fan, J., Huang, T. and Li, R. (2007). Analysis of longitudinal data with semiparametric estimation of covariance function. *Journal of the American Statistical Association* **102** 632–641. MR2370857

[10] Gasser, T., Hall, P. and Presnell, B. (1998). Nonparametric estimation of the mode of a distribution of random curves. *Journal of the Royal Statistical Society, Series B* **60** 681–691. MR1649539

[11] Green, P. J. and Silverman, B. W. (1994). *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach*, Chapman and Hall, London. MR1270012

[12] Härdle, W., Hall, P. and Ichimura, H. (1993). Optimal smoothing in single-index models. *The Annals of Statistics* **21** 157–178. MR1212171

[13] Hastie, T. and Tibshirani, R. (1990). *Generalized Additive Models*, Chapman and Hall, London. MR1082147

[14] Hastie, T. and Tibshirani, R. (1993). Varying-coefficient models. *Journal of the Royal Statistical Society, Series B* **55** 757–796. MR1229881

[15] Huang, J. Z., Wu, C. O. and Zhou, L. (2002). Varying-coefficient models and basis function approximations for the analysis of repeated measurements. *Biometrika* **89** 111–128. MR1888349

[16] Huang, J. Z., Wu, C. O. and Zhou, L. (2004). Polynomial spline estimation and inference for varying coefficient models with longitudinal data. *Statistica Sinica* **14** 763–788. MR2087972

[17] Leng, C., Zhang, W. and Pan, J. (2010). Semiparametric mean-covariance regression analysis for longitudinal data. *Journal of the American Statistical Association* **105** 181–193. MR2656048

[18] Jiang, C. R. and Wang, J. L. (2011). Functional single index models for longitudinal data. *The Annals of Statistics* **39** 362–388. MR2797850

[19] Li, L., Astor, B. C., Lewis, J., Hu, B., Appel, L. J., Lipkowitz, M. S., Toto, R. D., Wang, X., Wright, J. T. and Greene, T. (2012). Longitudinal progression trajectory of GFR among patients with CKD. *Am J Kidney Dis* **59** 504–512.

[20] Li, L., Chang, A., Rostand, S., Hebert, L., Appel, L. J., Astor, B. C., Lipkowitz, M. S., Wright, J. T., Wang, X. and Greene, T. (2014). A within-patient analysis for time-varying risk factors of CKD progression. *J Am Soc Nephrol* **25** 606–613.

[21] Liu, J. C., Zhang, R. Q. and Zhao, W. H. (2013). A robust and efficient estimation method for single index models. *Journal of Multivariate Analysis* **122** 226–238. MR3189320

[22] Pan, J. and Mackenzie, G. (2003). On modelling mean-covariance structures in longitudinal studies. *Bimetrika* **90** 239–244. MR1966564

[23] Park, B. U., Mammen, E., Lee, Y. K. and Lee, E. R. (2013). Varying coefficient regression models: a review and new developments. *International Statistical Review*.

[24] Pinheiro, J. C. and Bates, D. M. (2000). *Mixed-Effects Models in S and S-PLUS*, Springer.

[25] Pourahmadi, M. (1999). Joint mean-covariance models with applications to longitudinal data: unconstrained parameterisation. *Bimetrika* **86** 677–900. MR1723786

[26] Wang, Y. (2011). *Smoothing Splines: Methods and Applications*, Chapman and Hall. MR2814838

[27] Wang, Q. and Xue, L. (2011). Statistical inference in partially-varying-coefficient single-index model. *Journal of Multivariate Analysis* **102** 1–19. MR2729416

[28] Wong, H., Ip, W. and Zhang, R. Q. (2008). Varying-coefficient single-index model. *Computational Statistics* & *Data Analysis* **52** 1458–1476. MR2422748

[29] Xia, Y. and Härdle, W. (2006). Semi-parametric estimation of partially linear single index models. *Journal of Multivariate Analysis* **97** 1162–1184. MR2276153

[30] Xia, Y. and Li, W. (1999). On single-index coefficient regression models. *Journal of the American Statistical Association* **94** 1275–1285. MR1731489

[31] Xia, Y., Li, W. and Tong, H. (2004). Efficient estimation for semivarying-coefficient models. *Bimetrika* **91** 661–681. MR2090629

[32] Yao, W. and Li, R. (2013). New local estimation procedure for a nonparametric regression function for longitudinal data. *Journal of the Royal Statistical Society, Series B* **75** 123–138. MR3008274

[33] Ye, H. and Pan, J. (2006). Modelling of covariates structures in generalized estimating equations for longitudinal data. *Bimetrika* **93** 927–941. MR2285080

[34] Zhang, R., Zhao, W. and Liu, J. (2013). Robust estimation and variable selection for semiparametric partially linear varying coefficient model based on modal regression. *Journal of Nonparametric Statistics* **25** 523–544. MR3056100

Hongmei Lin
School of Statistics
East China Normal University
Shanghai 200241
P. R. China
E-mail address: xuemeijiayi@163.com

Riquan Zhang
School of Statistics
East China Normal University
Shanghai 200241
P. R. China
E-mail address: zhangriquan@163.com

Jianhong Shi
School of Mathematics and Computer Science
Shanxi Normal University
Linfen 041004
P. R. China
E-mail address: shijh70@163.com

Yuedong Wang
Department of Statistics and Applied Probability
University of California
Santa Barbara, CA 93106
USA
E-mail address: yuedong@pstat.ucsb.edu