University of Michigan School of Public Health

The University of Michigan Department of Biostatistics Working Paper Series

Year 2013 *Paper* 100

VARYING INDEX COEFFICIENT MODELS

Shujie Ma* Peter Xuekun Song†

*University of California - Riverside, shujie.ma@ucr.edu

†University of Michigan, Ann Arbor, pxsong@umich.edu

VARYING INDEX COEFFICIENT MODELS

Shujie Ma and Peter Xuekun Song

Abstract

It has been a long history of utilizing interactions in regression analysis to investigate interactive effects of covariates on response variables. In this paper we aim to address two kinds of new challenges resulted from the inclusion of such high-order effects in the regression model for complex data. The first kind arises from a situation where interaction effects of individual covariates are weak but those of combined covariates are strong, and the other kind pertains to the presence of nonlinear interactive effects. Generalizing the single index coefficient regression model (Xia and Li, 1999), we propose a new class of semiparametric models with varying index coefficients, which enables us to model and assess nonlinear interaction effects between grouped covariates on the response variable. As a result, most of the existing semiparametric regression models are special cases of our proposed models. We develop a numerically stable and computationally fast estimation procedure utilizing both profile least squares method and local fitting. We establish both estimation consistency and asymptotic normality for the proposed estimators of index coefficients as well as the oracle property for the nonparametric function estimator. In addition, a generalized likelihood ratio test is provided to test for the existence of interaction effects or the existence of nonlinear interaction effects. Our models and estimation methods are illustrated by both simulation studies and an analysis of body fat dataset.

Varying Index Coefficient Models [1]

Shujie Ma and Peter X.-K. Song

Abstract

It has been a long history of utilizing interactions in regression analysis to investigate interactive effects of covariates on response variables. In this paper we aim to address two kinds of new challenges resulted from the inclusion of such high-order effects in the regression model for complex data. The first kind arises from a situation where interaction effects of individual covariates are weak but those of combined covariates are strong, and the other kind pertains to the presence of nonlinear interactive effects. Generalizing the single index coefficient regression model (Xia and Li, 1999), we propose a new class of semiparametric models with varying index coefficients, which enables us to model and assess nonlinear interaction effects between grouped covariates on the response variable. As a result, most of the existing semiparametric regression models are special cases of our proposed models. We develop a numerically stable and computationally fast estimation procedure utilizing both profile least squares method and local fitting. We establish both estimation consistency and asymptotic normality for the proposed estimators of index coefficients as well as the oracle property for the nonparametric function estimator. In addition, a generalized likelihood ratio test is provided to test for the existence of interaction effects or the existence of nonlinear interaction effects. Our models and estimation methods are illustrated by both simulation studies and an analysis of body fat dataset.

Some Key Words: Semiparametric regression, interactions, B splines, profile estimation, two-step estimation, oracle property

Short title: Varying Index Coefficient Models

[1]Shujie Ma is Assistant Professor, Department of Statistics, University of California-Riverside, Riverside, CA 92521 (Email: shujie.ma@ucr.edu). Peter X.-K. Song is Professor, Department of Biostatistics, University of Michigan, Ann Arbor , MI 48109-2029. Song's research was partially supported by NSF grant DMS 1208939. Part of the research was conducted by the second author during his visit at the Department of Statistics and Applied Probability, National University of Singapore. He is thankful to the department for computational and other logistic support. The correspondence should be addressed to Shujie Ma.

1 Introduction

Regression analysis has played a central role in studying relationships between variables in the statistical literature. In a linear regression model, the dependent variable is typically assumed to be a linear function of one or more independent variables plus an error given as follows:

$$Y = \mathbf{Z}^{\mathrm{T}}\boldsymbol{\beta} + \varepsilon, \tag{1}$$

where Y is the response variable, $\mathbf{Z} = (Z_1, \ldots, Z_p)^{\mathrm{T}}$ is the p-dimensional vector of covariates of interest, ε is the error term with mean 0, and $\boldsymbol{\beta} = \left(\beta_1, \ldots, \beta_p\right)^{\mathrm{T}}$ is the p-dimensional vector of regression coefficients. As one of the most widely used regression methods, the linear model and properties of parameter estimators have been extensively studied. One challenge arising from applications of the linear model in practical studies is the violation of the linearity assumption on the relationship between Y and $\mathbf{Z}$. Such misspecification may give rise to large bias in estimation and incorrect inference, and hence misleading conclusions. For example, it is pointed out in popular public health monographs (e.g. Behnke and Wilmore (1974) pp. 66-67, Wilmore (1976) pp. 247 and Katch and McArdle (1977) pp. 120-132), body fat can be predicted by body circumference measurements such as abdominal, chest and hip circumferences. In a dataset available online (http://lib.stat.cmu.edu/datasets/bodyfat), it contains measured percentages of body fat determined by underwater weighing, and 12 circumference measurements from 252 men aged from 22 to 81 years old. By a routine analysis, one may fit the data by a linear model (1), where the response Y is the log-transformed body fat percentage, and six covariates $Z_1, \ldots, Z_6$ are, respectively, measured circumferences of chest, abdomen, hip, thigh, forearm and wrist. Denote the obtained least squares estimate of $\boldsymbol{\beta}$ by $\widehat{\boldsymbol{\beta}}$. To check the validity of linearity assumption, one approach would be based on the following nonparametric regression of Y and $\mathbf{Z}$: regress Y on linear predictor $U = \mathbf{Z}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}$ by the means of local linear fitting (Fan and Gijbels (1996)). The left panel of Figure 1 shows the fitted curve in U obtained by the local linear fitting (solid line), as well as the routine linear fitted line (dotted line). The contrast between these two fitted curves unveils a possible violation of linear relationship between Y and $\mathbf{Z}$.

The above approach of capturing nonlinear relationships is referred to the semiparametric single-index model, in which the response variable depends on an unknown but smooth nonlinear function of an index that takes a form of a linear combination of some covariates, given as follows:

$$Y = m\left(\mathbf{Z}^{\mathrm{T}}\boldsymbol{\beta}\right) + \varepsilon, \tag{2}$$

where both the smooth function $m(\cdot)$ and the index coefficients $\boldsymbol{\beta}$ are unknown. Studied intensively in the literature, this class of semiparametric single-index models is widely used in practice, because they enable us to deal with the curse of dimensionality (Bellman (1961)) and to achieve dimension reduction in nonparametric regression; also see Härdle et al. (1993), Stute and Zhu (2005), Cui et al. (2011), among others, for an overview and more references therein. It is worth noting that those covariates used in the index variables (e.g. circumference measurements) are often correlated variables of same or similar types.

Insert Figure 1 here

In the research of obesity or nutritional health science, age has been repeatedly reported as an important factor with a positive effect on body fat percentage (see Zamboni et al. (1997) and Jackson et al. (2002)). The right panel of Figure 1 displays two fitted curves over age obtained by nonparametric local linear fitting (solid line) and linear regression (dotted line), respectively. It is interesting to notice that body fat percent shows a nonlinear pattern for middle-aged men (40-60 years old), during which men's hormones are deemed to change significantly. We can observe that between ages 22 and 39, the fitted solid line indicates a linear increasing pattern. After age 39 the line rises up faster, followed by a phase of decreasing pattern after age 45. It returns the stable increasing mode after age 60. To better understand how the relationship between body fat percent and circumference measurements interact with covariate age, we ran nonparametric regression of Y on $U = \mathbf{Z}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}$, respectively, for three different age groups of 22-39, 40-60, and 61-81. Figure 2 shows the resulting three fitted curves over the estimated circumference index $u = z^{\mathrm{T}}\widehat{\boldsymbol{\beta}}$, each for one age group. Clearly, the three curves have demonstrated different patterns, which implies that there exist strong interaction effects between age and circumference index; in other words, the profiles of circumferences may have modified the rate of change regarding body fat percentage over age.

Insert Figure 2 here

To evaluate both nonlinear main effects of $\mathbf{Z}$ (e.g. circumferences) and interaction effects with X (e.g. age), we propose to extend model (2) as follows:

$$Y = m_0\left(\mathbf{Z}^{\mathrm{T}}\boldsymbol{\beta}_0\right) + m_1\left(\mathbf{Z}^{\mathrm{T}}\boldsymbol{\beta}_1\right)X + \varepsilon, \tag{3}$$

where $m_0(\cdot)$ and $m_1(\cdot)$ are unknown nonlinear functions, and $\boldsymbol{\beta}_l = \left(\beta_{l1}, \ldots, \beta_{lp}\right)^{\mathrm{T}}$ are coefficient vectors for $l = 1, 2$. It is interesting to note that $m_1(\cdot)$ behaves as a varying index coefficient of age X that explains the mechanism how covariates (e.g. suspected endocrine disrupting compounds) modify collectively the rate of growth. To generalize model (3) to include multiple covariates, we let $\mathbf{X}$ be a d-dimensional vector $\mathbf{X} = (X_1, \ldots, X_d)^{\mathrm{T}}$. A class of varying index coefficient models (VICM) is specified as follows:

$$Y = m(\mathbf{Z}, \mathbf{X}, \boldsymbol{\beta}) = \sum\nolimits_{l=1}^{d} m_l\left(\mathbf{Z}^{\mathrm{T}}\boldsymbol{\beta}_l\right)X_l + \varepsilon, \tag{4}$$

where $m_l(\cdot)$ are unknown smooth functions, and $\boldsymbol{\beta}_l = \left(\beta_{l1}, \ldots, \beta_{lp}\right)^{\mathrm{T}}$ are coefficient vectors for $1 \leq l \leq d$.

Note that in our model (4), the index coefficient vectors $\boldsymbol{\beta}_l$ are assumed to be different, unlike the setup of the single-index coefficient regression model proposed by Xia and Li (1999) in which a common coefficient vector $\boldsymbol{\beta}$ is assumed. Such difference, when the $\boldsymbol{\beta}$ vectors are given, gives rise to different nonparametric models, namely their varying-coefficient model and our additive model. Technically, the former involves one nonparametric function and the latter contains multiple nonparametric functions in estimation and inference. As a matter of fact, the former may be regarded as a special case of the latter, so the proposed profile estimation procedure in this paper for model (4) may also be applied to Xia and Li's model with minor modifications. It is worth noting that in real data analysis,with little knowledge about the nonparametric model structures it seems more natural to start with a model that include a full set of $\mathbf{Z}$ covariates in each coefficient function $m_l(\cdot)$, and then identify any important ones interacting with X_l through, for example, a hypothesis testing procedure. Denote $U_l(\boldsymbol{\beta}_l) = \mathbf{Z}^{\mathrm{T}}\boldsymbol{\beta}_l$. We assume that $U_l(\boldsymbol{\beta}_l)$ is confined in a compact set $[a, b]$. Without loss of generality, let $[a, b] = [0, 1]$. In this paper, we do not assume any distributions for error term ε, instead only requiring $E(\varepsilon\,|\mathbf{Z}, \mathbf{X}) = 0$ and $\mathrm{Var}(\varepsilon\,|\mathbf{Z}, \mathbf{X}) = \sigma^2(\mathbf{Z}, \mathbf{X})$.

For the sake of identifiability, let $\boldsymbol{\beta} = \left(\boldsymbol{\beta}_1^{\mathrm{T}}, \ldots, \boldsymbol{\beta}_d^{\mathrm{T}}\right)^{\mathrm{T}}$ belong to the parameter space:

$$\Theta = \left\{\boldsymbol{\beta} = \left(\boldsymbol{\beta}_l^{\mathrm{T}} : 1 \leq l \leq d\right)^{\mathrm{T}} : \|\boldsymbol{\beta}_l\|_2 = 1, \beta_{l1} > 0, \boldsymbol{\beta}_l \in R^p\right\},$$

where $\|\cdot\|_2$ denotes the L_2 norm of a vector such that $\|\boldsymbol{\zeta}\|_2 = (|\zeta_1|^2 + \cdots + |\zeta_s|^2)^{1/2}$ for any vector $\boldsymbol{\zeta} = (\zeta_1, \ldots, \zeta_s)^{\mathrm{T}} \in R^s$.

The class of models specified by (4) is quite general, containing many existing models as special cases. Some of the examples include:

1. When $m_l(\cdot) \equiv m_l$, where m_l are unknown constants, model (4) is reduced to a linear regression model.

2. When $d = 1$ and $X_1 \equiv 1$, model (4) becomes a single-index model. When $X_1 \equiv 1$ and $m_l\left(\mathbf{Z}^{\mathrm{T}}\boldsymbol{\beta}_l\right) \equiv m_l$ for $l \geq 2$, it is a partially linear single-index model (PLSiM, Carroll et al. (1997), Xia et al. (1999), Lu et al. (2006) and Liang et al. (2010)).

3. When $X_l \equiv 1$ for all $1 \leq l \leq d$, model (4) is an additive index model (Yuan (2011)).

4. When $\mathbf{Z}$ is a scalar ($p = 1$), model (4) becomes a varying-coefficient model (VCM, Hastie and Tibishirani (1993), Cai et al. (2000), Lin et al. (2007) and Ma et al. (2011)).

5. By treating $\mathbf{Z}^{\mathrm{T}}\boldsymbol{\beta}_l$ as a covariate U_l and letting $X_l \equiv 1$ for all $1 \leq l \leq d$, model (4) may be regarded as an additive model (Hastie and Tibshirani (1990) and Wang and Yang (2007)), and moreover by letting $m_l(\cdot) \equiv m_l$ for some l, it is a partially linear additive model (PLAM, Wang et al. (2011) and Ma and Yang (2011)).

6. As mentioned above, when $\boldsymbol{\beta}_l$ are the same for all $1 \leq l \leq d$, model (4) is the single-index coefficient model (SiCM) studied in Xia and Li (1999).

We develop a profile least squares (LS) estimation procedure to estimate the parameter vector $\boldsymbol{\beta} = \left(\boldsymbol{\beta}_1^{\mathrm{T}}, \ldots, \boldsymbol{\beta}_d^{\mathrm{T}}\right)^{\mathrm{T}}$. Precisely, for a given $\boldsymbol{\beta}$, we apply the LS estimation to approximate each $m_l(\cdot)$ via B-spline basis functions (de Boor (2001)), and the resulting estimator of $m_l(\cdot)$ is a function of $\boldsymbol{\beta}$. By replacing $m_l(\cdot)$ with its spline estimator in the conditional mean, we obtain the estimator of $\boldsymbol{\beta}$ by the LS method. This proposed profile spline estimation is motivated by the profile kernel estimation studied in Liang et al. (2010) and Cui et al. (2011) for PLSiMs and single-index models. Other existing methods of estimating parameters in single-index models include the backfitting algorithm (Carroll et al. (1997)), the penalized spline estimation (Yu and Ruppert (2002)) and the minimum average variance estimation (MAVE, Xia and Li (1999), Xia et al. (1999) and Xia and Härdle (2006)). It is noteworthy that to deal with the estimation in model (4) the backfitting algorithm may be unstable, while the penalized spline estimation may be inefficient. Although MAVE overcomes some of these limitations, it could encounter the so-called sparseness problem as noted by Cui et al. (2011). Since the proposed LS profile estimator of $\boldsymbol{\beta}$ implicitly involves the spline estimates of the nonparametric functions with a divergent number of parameters, the existing asymptotic distribution for the estimator in parametric models cannot be directly applied. In this paper, we propose a new approach to establishing the asymptotic normality for the profile LS estimator of $\boldsymbol{\beta}$ in model (4).

Another challenge arises in the estimation of the nonparametric functions $m_l(\cdot)$ in model (4), which requires more sophisticated estimation procedures than the kernel smoothing method employed in both Xia and Li (1999) for the SiCM and Liang et al. (2010) for the PLSiM. Note that when parameters $\boldsymbol{\beta}$ are fixed by known values or by their root-n consistent estimates, the SiCM is simplified as to be a VCM in which each coefficient function is univariate, and thus some of the existing nonparametric smoothing methods proposed for the VCM may be directly applied; for example, the kernel-based method (Cai et al. (2000); Fan and Zhang (2008)) and the spline-based method (Huang et al. (2004)). Our VICM (4), however, involves multiple additive nonparametric functions due to different parametric vectors $\boldsymbol{\beta}_l$. Moreover, these nonparametric functions interact with covariates X_l to form nonlinear interaction effects of scientific interest. Obviously, the estimation methods for univariate nonparametric functions fails to directly applicable in model (4). In the literature, several methods have been proposed for estimation in multivariate additive models, summarized as follows. It is shown in Stone (1985) that the one-step LS B-spline estimators of the additive nonparametric functions have the univariate convergence rate, but no asymptotic distribution is available. Later, there are several alternative estimation methods proposed for the additive models that provide asymptotically normally distributed estimators, including the backfitting algorithm (Hastie and Tibshirani (1990), Mammen et al. (1999) and Opsomer and Ruppert (1997)), the marginal integration method (Linton and Nielsen (1995) and Fan et al. (1998)), the two-stage estimation (Linton (1997), Horowitz (2006) and Horowitz and Mammen (2004)), and the spline backfitted local linear (SBLL, Wang and Yang (2007)).

In this paper, we adopt the SBLL procedure, which theoretically enjoys the oracle property and is computationally expedient. The SBLL procedure is briefly outlined as follows: in model (4), for a given l, to estimate $m_l(\cdot)$, (i) first utilize the one-step B-spline estimation for all the other functions $m_{l'}(\cdot)$, $l' \neq l$, as the initial estimates in replace of $m_{l'}(\cdot)$, $l' \neq l$; and (ii) then estimate $m_l(\cdot)$ by the means of local linear smoothing. The resulting SBLL estimator will be shown to satisfy the oracle property; that is, it has the same asymptotic distribution as that of the univariate oracle estimator under the assumption that all the other nonparametric functions were known. Such useful properties are achieved by taking the advantage of joint asymptotics of kernel and spline functions. Furthermore, by taking advantage of the asymptotic normality for the estimators of the index parameters $\boldsymbol{\beta}$ and the nonparametric functions $m_l(\cdot)$, we construct a Wald test and a generalized likelihood ratio test (see Fan et al. (2001); Fan and Jiang (2007)) to make statistical inferences.

The rest of this paper is organized as follows. Section 2 introduces the profile LS estimation and presents asymptotic properties of the proposed estimators. Section 3 discusses the SBLL estimation and inference for parameter $\boldsymbol{\beta}$ and the nonparametric function $m_l(\cdot)$. In Section 4, we describe the procedure of selecting smoothing parameters. In Sections 5, we evaluate finite sample properties of the proposed estimation and inference procedures via simulation studies. Section 6 illustrates the proposed model and method through the analysis of body fat percentage data. Some concluding remarks are given in Section 7. All technical details including detailed proofs are provided in the Appendix.

2 Profile Least Squares Estimation

Suppose $(Y_i, \mathbf{Z}_i, \mathbf{X}_i, \mathbf{U}_i(\boldsymbol{\beta}))$, $1 \leq i \leq n$, are the i.i.d. realizations of $(Y, \mathbf{Z}, \mathbf{X}, \mathbf{U}(\boldsymbol{\beta}))$, where $\mathbf{U}(\boldsymbol{\beta}) = (U_1(\boldsymbol{\beta}_1), \ldots, U_d(\boldsymbol{\beta}_d))^{\mathrm{T}}$. We propose an estimation of parameter $\boldsymbol{\beta}$ by a profile LS procedure. Letting

$\boldsymbol{\beta}$ be fixed, we estimate nonparametric functions $m_l(u_l)$ by B splines described as follows. Let $\mathcal{G}_n$ denote the space of polynomial splines of order $q \geq 2$. Consider a knot sequence with $N \equiv N_n$ interior knots, denoted by

$$\xi_1 = \cdots = 0 = \xi_q < \xi_{q+1} < \cdots < \xi_{q+N} < 1 = \xi_{N+q+1} = \cdots = \xi_{N+2q},$$

where N increases along with the number of subjects n. Space $\mathcal{G}_n$ consists of functions, say ϖ, satisfying (i) ϖ is a polynomial of degree $q-1$ on each of subintervals $I_s = \left[\xi_s, \xi_{s+1}\right)$, $s = 0, ..., N_n - 1$, and $I_{N_n} = \left[\xi_{N_n}, 1\right]$; (ii) for $q \geq 2$, function ϖ is $q-2$ times continuously differentiable on $[0,1]$. For $0 \leq s \leq N_n$, let $H_s = \xi_{s+1} - \xi_s$ be the distance between neighboring knots and let $H = \max_{0\leq s\leq N_n} H_s$. Following Zhou et al. (1998), to study asymptotic properties of the spline estimator of $m_l(\cdot)$, we assume that $\max_{0\leq s\leq N_n-1}|H_{s+1}-H_s| = o\left(N^{-1}\right)$ and $H/\min_{0\leq s\leq N_n} H_s \leq M$, where $M > 0$ is a predetermined constant. Such an assumption assures that $M^{-1} < N_n H < M$, which is necessary for numerical implementation. Let $J_n = N_n + q$. Denote the q-th order B spline basis for $\mathcal{G}_n$ (de Boor (2001), p. 89) as $\mathbf{B}_q(u) = (B_{s,q}(u) : 1 \leq s \leq J_n)^{\mathrm{T}}$, $u \in [0,1]$, with some $q \geq 3$. Then $m_l(u_l)$, $l = 1, \ldots, d$, are estimated by the following spline functions:

$$\widehat{m}_l(u_l, \boldsymbol{\beta}) = \sum\nolimits_{s=1}^{J_n} B_{s,q}(u_l)\,\widehat{\lambda}_{s,l}(\boldsymbol{\beta}) = \mathbf{B}_q(u_l)^{\mathrm{T}}\,\widehat{\boldsymbol{\lambda}}_l(\boldsymbol{\beta}), \tag{5}$$

where $\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}) = \left(\widehat{\boldsymbol{\lambda}}_1(\boldsymbol{\beta})^{\mathrm{T}}, \ldots, \widehat{\boldsymbol{\lambda}}_d(\boldsymbol{\beta})^{\mathrm{T}}\right)^{\mathrm{T}}$, with $\widehat{\boldsymbol{\lambda}}_l(\boldsymbol{\beta}) = \left(\widehat{\lambda}_{s,l}(\boldsymbol{\beta}) : 1 \leq s \leq J_n\right)^{\mathrm{T}}$, is given by

$$\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}) = \operatorname{argmin}_{\boldsymbol{\lambda}\in R^{dJ_n}} \sum\nolimits_{i=1}^{n} \left\{Y_i - \sum\nolimits_{l=1}^{d} \sum\nolimits_{s=1}^{J_n} B_{s,q}(U_{il}(\boldsymbol{\beta}_l))\,\lambda_{s,l} X_{il}\right\}^2. \tag{6}$$

Denote $D_i(\boldsymbol{\beta}) = (D_{i,sl}(\boldsymbol{\beta}_l), 1 \leq s \leq J_n, 1 \leq l \leq d)^{\mathrm{T}}$ with $D_{i,sl}(\boldsymbol{\beta}_l) = B_{s,q}(U_{il}(\boldsymbol{\beta}_l))\,X_{il}$ and $\mathbf{D}(\boldsymbol{\beta}) = \left[(D_1(\boldsymbol{\beta}), \ldots, D_n(\boldsymbol{\beta}))^{\mathrm{T}}\right]_{n\times J_n d}$. Thus the solution to (6) is expressed as

$$\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}) = \left\{\mathbf{D}(\boldsymbol{\beta})^{\mathrm{T}}\mathbf{D}(\boldsymbol{\beta})\right\}^{-1}\mathbf{D}(\boldsymbol{\beta})^{\mathrm{T}}\mathbf{Y}, \tag{7}$$

where $\mathbf{Y} = (Y_1, \ldots, Y_n)^{\mathrm{T}}$. In the estimation of $\boldsymbol{\beta}_l$, it requires estimates of both m_l and its first order derivative $\dot{m}_l$. According to de Boor (2001, page 116), $\dot{m}_l$ can be approximated by the spline functions with one order lower than that of m_l. Thus, a spline estimator of $\dot{m}_l$ is given by

$$\widehat{\dot{m}}_l(u_l, \boldsymbol{\beta}) = \sum\nolimits_{s=1}^{J_n} \dot{B}_{s,q}(u_l)\,\widehat{\lambda}_{s,l}(\boldsymbol{\beta}) = \sum\nolimits_{s=2}^{J_n} B_{s,q-1}(u_l)\,\widehat{\omega}_{s,l}(\boldsymbol{\beta}), \tag{8}$$

where

$$\widehat{\omega}_{s,l}(\boldsymbol{\beta}) = (q-1)\left\{\widehat{\lambda}_{s,l}(\boldsymbol{\beta}) - \widehat{\lambda}_{s-1,l}(\boldsymbol{\beta})\right\} / \left(\xi_{s+q-1} - \xi_s\right),$$

for $2 \leq s \leq J_n$. In addition, $\widehat{m}_l(u_l, \boldsymbol{\beta})$ can be re-expressed as $\widehat{m}_l(u_l, \boldsymbol{\beta}) = \mathbf{B}_{q-1}(u_l)^{\mathrm{T}} \mathbf{D}_1 \widehat{\boldsymbol{\lambda}}_l(\boldsymbol{\beta})$, where $\mathbf{B}_{q-1}(u_l) = (B_{s,q-1}(u_l) : 2 \leq s \leq J_n)^{\mathrm{T}}$ and

$$\mathbf{D}_1 = (q-1)\begin{pmatrix} \frac{-1}{\xi_{q+1}-\xi_2} & \frac{1}{\xi_{q+1}-\xi_2} & 0 & \cdots & 0 \\ 0 & \frac{-1}{\xi_{q+2}-\xi_3} & \frac{1}{\xi_{q+2}-\xi_3} & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{-1}{\xi_{N+2q-1}-\xi_{N+q}} & \frac{1}{\xi_{N+2q-1}-\xi_{N+q}} \end{pmatrix}_{(J_n-1)\times J_n}. \quad (9)$$

In the estimation of $\boldsymbol{\beta}$, to ensure identifiability, we exclude the first component β_{l1} of $\boldsymbol{\beta}_l$ by setting $\beta_{l1} = \left(1 - \|\boldsymbol{\beta}_{l,-1}\|_2^2\right)^{1/2}$, where $\boldsymbol{\beta}_{l,-1} = \left(\beta_{l2}, \ldots, \beta_{lp}\right)^{\mathrm{T}}$, for all $1 \leq l \leq d$ (see Cui et al. (2011)), and reformulate the parameter space of $\boldsymbol{\beta}_l$, $l = 1, \ldots, d$, as follows:

$$\Theta_{-1} = \left[\left\{\left(1 - \|\boldsymbol{\beta}_{l,-1}\|_2^2\right)^{1/2}, \beta_{l2}, \ldots, \beta_{lp}\right\}^{\mathrm{T}} : \|\boldsymbol{\beta}_{l,-1}\|_2^2 < 1\right].$$

Let $\boldsymbol{\beta}_{l,-1} = \left(\beta_{l2}, \ldots, \beta_{lp}\right)^{\mathrm{T}}$ and $\mathbf{J}_l = \partial\boldsymbol{\beta}_l / \partial\boldsymbol{\beta}_{l,-1}^{\mathrm{T}}$ be the Jacobian matrix of size $p \times (p-1)$, which is $\mathbf{J}_l = \begin{pmatrix} -\boldsymbol{\beta}_{l,-1}^{\mathrm{T}} / \sqrt{1 - \|\boldsymbol{\beta}_{l,-1}\|_2^2} \\ \mathbf{I}_{p-1} \end{pmatrix}$. Denote the estimate of $\boldsymbol{\beta}_{-1} = \left(\boldsymbol{\beta}_{1,-1}^{\mathrm{T}}, \ldots, \boldsymbol{\beta}_{d,-1}^{\mathrm{T}}\right)^{\mathrm{T}}$ by $\widehat{\boldsymbol{\beta}}_{-1} = \left(\widehat{\boldsymbol{\beta}}_{1,-1}, \ldots, \widehat{\boldsymbol{\beta}}_{d,-1}\right)^{\mathrm{T}}$, which can be obtained by $\widehat{\boldsymbol{\beta}}_{-1} = \arg\min_{\boldsymbol{\beta}_{-1} \in \Theta_{-1}} L_n(\boldsymbol{\beta})$, where

$$L_n(\boldsymbol{\beta}) = 2^{-1} \sum_{i=1}^{n} \left\{Y_i - \sum_{l=1}^{d} \sum_{s=1}^{J_n} B_{s,l}\left(U_{il}(\boldsymbol{\beta}_l)\right) \widehat{\lambda}_{s,l}(\boldsymbol{\beta}) X_{il}\right\}^2, \boldsymbol{\beta}_{-1} \in \Theta_{-1}.$$

Moreover, one can obtain $\widehat{\boldsymbol{\beta}}_{-1}$ as the solution of the following estimation equations:

$$\begin{aligned} \partial L_n(\boldsymbol{\beta}) / \partial \boldsymbol{\beta}_{-1} &= -\sum_{i=1}^{n} \left\{Y_i - \sum_{l=1}^{d} \sum_{s=1}^{J_n} B_{s,l}\left(U_{il}(\boldsymbol{\beta}_l)\right) \widehat{\lambda}_{s,l}(\boldsymbol{\beta}) X_{il}\right\} \times \\ & \left[\begin{array}{c} \left\{\widehat{m}_1(U_{i1}(\boldsymbol{\beta}_1), \boldsymbol{\beta}) X_{i1} \mathbf{J}_1^{\mathrm{T}} \mathbf{Z}_i + \left(\partial \widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta})^{\mathrm{T}} / \partial \boldsymbol{\beta}_{1,-1}\right) D_i(\boldsymbol{\beta})\right\} \\ \vdots \\ \left\{\widehat{m}_d(U_{id}(\boldsymbol{\beta}_d), \boldsymbol{\beta}) X_{id} \mathbf{J}_d^{\mathrm{T}} \mathbf{Z}_i + \left(\partial \widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta})^{\mathrm{T}} / \partial \boldsymbol{\beta}_{d,-1}\right) D_i(\boldsymbol{\beta})\right\} \end{array}\right] \\ &= 0, \end{aligned} \quad (10)$$

where $\widehat{m}_l(, \boldsymbol{\beta})$ is given in (8). Now define the space $\mathcal{M}$ as a collection of functions with finite L_2 norm on $[0,1]^d \times R^d$ by

$$\mathcal{M} = \left\{g(\mathbf{u}, \mathbf{x}) = \sum_{l=1}^{d} g_l(u_l) x_l, E g_l(U_l)^2 < \infty\right\},$$

where $\mathbf{u} = (u_1, \ldots, u_d)^{\mathrm{T}}$ and $\mathbf{x} = (x_1, \ldots, x_d)^{\mathrm{T}}$. To study the large-sample properties of parameter estimators, let $\boldsymbol{\beta}^0 = \left\{ (\boldsymbol{\beta}_1^0)^{\mathrm{T}}, \ldots, (\boldsymbol{\beta}_d^0)^{\mathrm{T}} \right\}^{\mathrm{T}}$ with $\boldsymbol{\beta}_l^0 = \left\{ \beta_{l1}^0, (\boldsymbol{\beta}_{l,-1}^0)^{\mathrm{T}} \right\}^{\mathrm{T}}$ and $\boldsymbol{\beta}_{l,-1}^0 = \left(\beta_{l2}^0, \ldots, \beta_{lp}^0 \right)^{\mathrm{T}}$ for $1 \le l \le d$ be the true parameters in model (4). For $1 \le k \le p$, define g_k^0 as the one satisfying:

$$\mathbb{P}(Z_k) = g_k^0 \left(\mathbf{U}\left(\boldsymbol{\beta}^0\right), \mathbf{X} \right) = \sum_{l=1}^d g_{l,k}^0 \left(U_l(\boldsymbol{\beta}_l^0) \right) X_l = \arg\min_{g \in \mathcal{M}} E \left\{ Z_k - g\left(\mathbf{U}\left(\boldsymbol{\beta}^0\right), \mathbf{X} \right) \right\}^2. \quad (11)$$

Let $\mathbb{P}(\mathbf{Z}) = \{\mathbb{P}(Z_1), \ldots, \mathbb{P}(Z_p)\}^{\mathrm{T}}$, $\widetilde{\mathbf{Z}} = \mathbf{Z} - \mathbb{P}(\mathbf{Z})$ and

$$\Phi\left(\mathbf{X}, \mathbf{Z}, \boldsymbol{\beta}^0\right) = \left[\left\{ \dot{m}_l(U_l(\boldsymbol{\beta}_l^0), \boldsymbol{\beta}^0) X_l \mathbf{J}_l^{\mathrm{T}} \widetilde{\mathbf{Z}} \right\}^{\mathrm{T}}, 1 \le l \le d \right]^{\mathrm{T}}. \quad (12)$$

Here $\Phi\left(\mathbf{X}, \mathbf{Z}, \boldsymbol{\beta}^0\right)$ is a vector with $(p-1)\,d$ elements. For any matrix $\mathbf{A}$, denote $\mathbf{A}^{\otimes 2} = \mathbf{A}\mathbf{A}^{\mathrm{T}}$.

Theorem 1. *Under Conditions (C1)-(C5) in the Appendix, and* $nN^{-4} \to \infty$ *and* $nN^{-2q-2} \to 0$, *we have*

(i) (consistency) $\left\| \widehat{\boldsymbol{\beta}}_{-1} - \boldsymbol{\beta}_{-1}^0 \right\|_2 = O_p\left(n^{-1/2}\right)$;

(ii) (asymptotic normality) as $n \to \infty$,

$$\begin{aligned} \sqrt{n}\left(\widehat{\boldsymbol{\beta}}_{-1} - \boldsymbol{\beta}_{-1}^0 \right) &= \left\{ n^{-1} \sum_{i=1}^n \Phi\left(\mathbf{X}_i, \mathbf{Z}_i, \boldsymbol{\beta}^0\right)^{\otimes 2} \right\}^{-1} \times \\ &\quad \left\{ n^{-1/2} \sum_{i=1}^n \left(Y_i - m\left(\mathbf{Z}_i, \mathbf{X}_i\right)\right) \Phi\left(\mathbf{X}_i, \mathbf{Z}_i, \boldsymbol{\beta}^0\right) \right\} + o_p(1). \end{aligned}$$

Moreover $\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_{-1} - \boldsymbol{\beta}_{-1}^0 \right) \xrightarrow{d} \mathcal{N}_{d(p-1)}(\mathbf{0}, \Sigma)$, *as* $n \to \infty$, *where*

$$\Sigma = \left[E\left\{ \Phi\left(\mathbf{X}, \mathbf{Z}, \boldsymbol{\beta}^0\right)^{\otimes 2} \right\} \right]^{-1} \left[E\left\{ \sigma^2(\mathbf{Z}, \mathbf{X})\, \Phi\left(\mathbf{X}, \mathbf{Z}, \boldsymbol{\beta}^0\right)^{\otimes 2} \right\} \right] \left[E\left\{ \Phi\left(\mathbf{X}, \mathbf{Z}, \boldsymbol{\beta}^0\right)^{\otimes 2} \right\} \right]^{-1}. \quad (13)$$

Remark 1. If we assume homoscedasticity to the random noise ε in model (4), that is, $\sigma^2(\mathbf{Z}, \mathbf{X}) = \sigma^2$ for some constant $\sigma^2 > 0$, then the asymptotic variance matrix given in (13) is reduced to

$$\Sigma = \sigma^2 \left[E\left\{ \Phi\left(\mathbf{X}, \mathbf{Z}, \boldsymbol{\beta}^0\right)^{\otimes 2} \right\} \right]^{-1}. \quad (14)$$

Let $\mathbf{J}_{dp \times d(p-1)} = \oplus_{l=1}^d \mathbf{J}_l = \mathrm{diag}(\mathbf{J}_1, \ldots, \mathbf{J}_d)$ be the direct sum of Jacobian matrices $\mathbf{J}_1, \ldots, \mathbf{J}_d$. For $1 \le l \le d$, β_{l1} is estimated by $\widehat{\beta}_{l1} = \left(1 - \sum_{k=2}^p \widehat{\beta}_{lk}^2\right)^{1/2}$. Let $\widehat{\boldsymbol{\beta}}_l = \left(\widehat{\beta}_{l1}, \ldots, \widehat{\beta}_{lp}\right)^{\mathrm{T}}$. Both consistency and asymptotic normality of $\widehat{\boldsymbol{\beta}} = \left(\widehat{\boldsymbol{\beta}}_1^{\mathrm{T}}, \ldots. \widehat{\boldsymbol{\beta}}_d^{\mathrm{T}}\right)^{\mathrm{T}}$ follow directly from Theorem 1 with an application of the multivariate delta-method. Thus we obtain $\sqrt{n}\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0\right) \xrightarrow{d} \mathcal{N}_{dp}\left(0, \mathbf{J}\Sigma\mathbf{J}^{\mathrm{T}}\right)$, $n \to \infty$.

Next we consider the spline estimator of the nonparametric function $m_l(\cdot)$ given as follows:

$$\widehat{m}_l(u_l, \widehat{\boldsymbol{\beta}}) = \sum_{s=1}^{J_n} B_{s,q}(u_l)\,\widehat{\lambda}_{s,l}(\widehat{\boldsymbol{\beta}}) = \mathbf{B}_q(u_l)^{\mathrm{T}}\,\widehat{\boldsymbol{\lambda}}_l(\widehat{\boldsymbol{\beta}}), \tag{15}$$

where $\widehat{\boldsymbol{\lambda}}(\widehat{\boldsymbol{\beta}}) = \left(\widehat{\boldsymbol{\lambda}}_1(\widehat{\boldsymbol{\beta}})^{\mathrm{T}}, \ldots, \widehat{\boldsymbol{\lambda}}_d(\widehat{\boldsymbol{\beta}})^{\mathrm{T}}\right)^{\mathrm{T}}$ with $\widehat{\boldsymbol{\lambda}}_l(\widehat{\boldsymbol{\beta}}) = \left(\widehat{\lambda}_{s,l}(\widehat{\boldsymbol{\beta}}) : 1 \le s \le J_n\right)^{\mathrm{T}}$ given by (7) in which $\boldsymbol{\beta}$ is replaced with $\widehat{\boldsymbol{\beta}}$. The following theorem provides the convergence rate of $\widehat{m}_l(u_l, \widehat{\boldsymbol{\beta}})$.

Theorem 2. *Under Conditions (C1)-(C5) in the Appendix, and $nN^{-4} \to \infty$ and $nN^{-2q-2} \to 0$, we have for each $1 \le l \le d$, $\left|\widehat{m}_l(u_l, \widehat{\boldsymbol{\beta}}) - m_l(u_l)\right| = O_p\left(\sqrt{N/n} + N^{-q}\right)$ uniformly for any $u_l \in [0,1]$.*

Remark 2. The order assumptions regarding N, $nN^{-4} \to \infty$, and $nN^{-2q-2} \to 0$, in Theorem 2 implies that $N \asymp n^{1/(2q+1)}$, which is the optimal order for the number of interior knots needed to estimate the nonparametric functions. The resulting convergence rate is then $O_p\left\{n^{-q/(2q+1)}\right\}$. For instance, when $q = 4$, which is the order for cubic splines, the optimal convergence rate is $O_p\left(N^{-4/9}\right)$.

Remark 3. To estimate the asymptotic covariance matrix Σ given in (13), we need estimation of $\Phi\left(\mathbf{X}, \mathbf{Z}, \boldsymbol{\beta}^0\right)$ given by (12). There $\widetilde{\mathbf{Z}}$ can be estimated by $\widehat{\mathbf{Z}} = \mathbf{Z} - \mathbb{P}_n(\mathbf{Z})$, with $\mathbb{P}_n(\mathbf{Z}) = \{\mathbb{P}_n(Z_1), \ldots, \mathbb{P}_n(Z_p)\}^{\mathrm{T}}$ and $\mathbb{P}_n(Z_k) = \sum_{l=1}^{d} \widehat{g}^0_{l,k}\left(U_l\left(\widehat{\boldsymbol{\beta}}\right), \widehat{\boldsymbol{\beta}}\right) X_l$, where $\widehat{g}^0_{l,k}\left(\cdot, \widehat{\boldsymbol{\beta}}\right)$ is the spline estimate of $g^0_{l,k}(\cdot)$ obtained by carrying out the same procedure as for $\widehat{m}_l(\cdot, \widehat{\boldsymbol{\beta}})$ with the response Y replaced by Z_k. Thus $\Phi\left(\mathbf{X}, \mathbf{Z}, \boldsymbol{\beta}^0\right)$ is estimated by

$$\widehat{\Phi}\left(\mathbf{X}, \mathbf{Z}, \widehat{\boldsymbol{\beta}}\right) = \left[\left\{\widehat{m}_l(U_l(\widehat{\boldsymbol{\beta}}_l), \widehat{\boldsymbol{\beta}}) X_l \mathbf{J}_l^{\mathrm{T}} \widehat{\mathbf{Z}}\right\}^{\mathrm{T}}, 1 \le l \le d\right]^{\mathrm{T}},$$

and the resulting estimate of Σ defined in (13) is given by

$$\begin{aligned} \widehat{\Sigma} &= n\left\{\sum_{i=1}^{n} \widehat{\Phi}\left(\mathbf{X}_i, \mathbf{Z}_i, \widehat{\boldsymbol{\beta}}\right)^{\otimes 2}\right\}^{-1} \left\{\sum_{i=1}^{n} \widehat{e}^2(\mathbf{Z}_i, \mathbf{X}_i)\,\widehat{\Phi}\left(\mathbf{X}_i, \mathbf{Z}_i, \widehat{\boldsymbol{\beta}}\right)^{\otimes 2}\right\} \\ &\quad \times \left\{\sum_{i=1}^{n} \widehat{\Phi}\left(\mathbf{X}_i, \mathbf{Z}_i, \widehat{\boldsymbol{\beta}}\right)^{\otimes 2}\right\}^{-1}, \end{aligned} \tag{16}$$

where $\widehat{e}(\mathbf{X}_i, \mathbf{Z}_i) = Y_i - \sum_{l=1}^{d} \widehat{m}_l\left(\mathbf{Z}_i^{\mathrm{T}} \widehat{\boldsymbol{\beta}}_l, \widehat{\boldsymbol{\beta}}\right) X_{il}$. For the homoscedasticity case, Σ in (14) is estimated by

$$\widehat{\Sigma} = \widehat{\sigma}^2 n \left\{\sum_{i=1}^{n} \widehat{\Phi}\left(\mathbf{X}_i, \mathbf{Z}_i, \widehat{\boldsymbol{\beta}}\right)^{\otimes 2}\right\}^{-1}, \tag{17}$$

where $\widehat{\sigma}^2 = \sum_{i=1}^{n} \widehat{e}^2(\mathbf{X}_i, \mathbf{Z}_i) / \{n - d(J_n + p)\}$.

3 Inference

3.1 Oracle property of SBLL estimation for $m_l(\cdot)$

In Theorem 2 we show that the spline estimator $\widehat{m}_l\left(\cdot,\widehat{\boldsymbol{\beta}}\right)$ obtained from the profile estimation procedure in (15) is a consistent estimator of $m_l(\cdot)$. The asymptotic distribution of $\widehat{m}_l\left(\cdot,\widehat{\boldsymbol{\beta}}\right)$, however, is not available. Thus, no measure of confidence can be established in statistical inference. To overcome this, we consider a two-step spline backfitted local linear (SBLL) estimation for the nonparametric function $m_l(\cdot)$, for which the spline estimate $\widehat{m}_l\left(\cdot,\widehat{\boldsymbol{\beta}}\right)$ given in (15) will be used as the initial estimate. We will establish the asymptotic normality for the SBLL estimators. The SBLL estimation proceeds as follows. Without loss of generality, we focus on the estimation of the first nonparametric function $m_1(\cdot)$. The spline estimates $\widehat{m}_l(\cdot,\widehat{\boldsymbol{\beta}})$, $l \geq 2$, given in (15) are used as the initial estimates and held fixed in the estimation of $m_1(\cdot)$. When $m_l(\cdot)$ for $l \geq 2$ were known, we could define the oracle pseudo response $Y_{i,1} = Y_i - \sum_{l=2}^{d} m_l\left(\mathbf{Z}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_l\right)X_{il} = m_1\left(\mathbf{Z}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_1\right)X_{i1} + \varepsilon_i$, where $\widehat{\boldsymbol{\beta}}_l$ are the LS profile estimators given in Section 2. For each given u_1, $m_1(u_1)$ is estimated by the means of local linear fitting, namely $\widetilde{m}_{\mathrm{LL},1}\left(u_1,\widehat{\boldsymbol{\beta}}\right) = \widehat{a}\left(\widehat{\boldsymbol{\beta}}\right)$, where $\widehat{a}\left(\widehat{\boldsymbol{\beta}}\right)$ and $\widehat{b}\left(\widehat{\boldsymbol{\beta}}\right)$ minimize the following local kernel objective function:

$$\sum_{i=1}^{n}\left\{Y_{i,1} - aX_{i1} - b\left(U_{i1}(\widehat{\boldsymbol{\beta}}_1) - u_1\right)X_{i1}\right\}^2 K_{h_1}\left(U_{i1}(\widehat{\boldsymbol{\beta}}_1) - u_1\right).$$

Here $K_{h_1}(u) = K(u/h_1)/h_1$ is a symmetric kernel function and h_1 is a bandwidth. Let

$$\begin{aligned}
\mathbf{C}\left(u_1,\widehat{\boldsymbol{\beta}}_1\right) &= \begin{bmatrix} X_{11} & \cdots & X_{1n} \\ X_{11}\left\{\left(U_{11}(\widehat{\boldsymbol{\beta}}_1) - u_1\right)/h_1\right\} & \cdots & X_{1n}\left\{\left(U_{1n}(\widehat{\boldsymbol{\beta}}_1) - u_1\right)/h_1\right\} \end{bmatrix}^{\mathrm{T}}, \\
\mathbf{W}\left(u_1,\widehat{\boldsymbol{\beta}}_1\right) &= \operatorname{diag}\left\{K_{h_1}\left(U_{11}(\widehat{\boldsymbol{\beta}}_1) - u_1\right),\ldots,K_{h_1}\left(U_{1n}(\widehat{\boldsymbol{\beta}}_1) - u_1\right)\right\},
\end{aligned}$$

and $\mathbf{Y}_1 = (Y_{1,1},\ldots,Y_{n,1})^{\mathrm{T}}$. Then we have

$$\widehat{a}\left(\widehat{\boldsymbol{\beta}}\right) = (1,0)\left\{\mathbf{C}\left(u_1,\widehat{\boldsymbol{\beta}}_1\right)^{\mathrm{T}}\mathbf{W}\left(u_1,\widehat{\boldsymbol{\beta}}_1\right)\mathbf{C}\left(u_1,\widehat{\boldsymbol{\beta}}_1\right)\right\}^{-1}\mathbf{C}\left(u_1,\widehat{\boldsymbol{\beta}}_1\right)^{\mathrm{T}}\mathbf{W}\left(u_1,\widehat{\boldsymbol{\beta}}_1\right)\mathbf{Y}_1. \tag{18}$$

Because $m_l(u_l)$ for $l \geq 2$ are actually unknown, we modify (18) by replacing $m_l(u_l)$ with their spline estimators $\widehat{m}_l(u_l,\widehat{\boldsymbol{\beta}})$ given in (15), which is equivalent to replacing $\mathbf{Y}_1$ in (18) by $\widehat{\mathbf{Y}}_1$, where $\widehat{\mathbf{Y}}_1 = \left(\widehat{Y}_{1,1},\ldots,\widehat{Y}_{n,1}\right)^{\mathrm{T}}$ and $\widehat{Y}_{i,1} = Y_i - \sum_{l=2}^{d}\widehat{m}_l\left(\mathbf{Z}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_l,\widehat{\boldsymbol{\beta}}\right)X_{il}$. The resulting SBLL estimator is denoted by $\widehat{m}_{\mathrm{SBLL},1}\left(u_1,\widehat{\boldsymbol{\beta}}\right)$. Denote $\mu_2(K) = \int u^2K(u)\,du$ and $\|K\|_2^2 = \int K^2(u)\,du$.

Theorem 3. *Under Conditions (C1)-(C6) in the Appendix, and $h_1 \asymp n^{-1/5}$, $nN^{-4} \to \infty$ and $nN^{-2q-2} \to 0$, as $n \to \infty$, for any $u_1 \in [h_1, 1-h_1]$, we have*

$$\sup_{u_1 \in [h_1, 1-h_1]} \left| \widetilde{m}_{LL,1}\left(u_1, \widehat{\boldsymbol{\beta}}\right) - m_1(u_1) \right| = O_p\left(\sqrt{log(n) / (nh_1)}\right) = O_p\left(n^{-2/5}\sqrt{log(n)}\right),$$

and

$$\sqrt{nh_1}\left\{\widetilde{m}_{LL,1}\left(u_1, \widehat{\boldsymbol{\beta}}\right) - m_1(u_1) - b_1(u_1) h_1^2\right\} \xrightarrow{d} \mathcal{N}(0, v_1(u_1)),$$

where

$$\begin{aligned} b_1(u_1) &= \mu_2(K)\ddot{m}_1(u_1)/2, \\ v_1(u_1) &= \left\{E\left(X_1^2 | u_1\right)\right\}^{-2} f_1^{-1}(u_1) \|K\|_2^2 E\left\{X_1^2 \sigma^2(\mathbf{Z}, \mathbf{X}) | u_1\right\}. \end{aligned}$$

in which $\ddot{m}_1(\cdot)$ is the second order derivative of m_1 and $f_1(\cdot)$ is the density function of $\mathbf{Z}^T\boldsymbol{\beta}_1^0$.

The theorem below presents the uniform oracle efficiency of the SBLL estimator $\widehat{m}_{\mathrm{SBLL},1}\left(u_1, \widehat{\boldsymbol{\beta}}\right)$ such that the absolute difference between $\widehat{m}_{\mathrm{SBLL},1}\left(u_1, \widehat{\boldsymbol{\beta}}\right)$ and $\widetilde{m}_{\mathrm{LL},1}\left(u_1, \widehat{\boldsymbol{\beta}}\right)$ is of order $o_p\left(n^{-2/5}\right)$ uniformly. As a result, $\widehat{m}_{\mathrm{SBLL},1}\left(u_1, \widehat{\boldsymbol{\beta}}\right)$ has the same asymptotic distribution as $\widetilde{m}_{\mathrm{LL},1}\left(u_1, \widehat{\boldsymbol{\beta}}\right)$.

Theorem 4. *Under Conditions (C1)-(C6) in the Appendix, and $nN^{-4} \to \infty$, $nN^{-2q-2} \to 0$ and $nN^{-5q/2} \to 0$, we have*

$$\sup_{u_1 \in [0,1]} \left| \widehat{m}_{SBLL,1}\left(u_1, \widehat{\boldsymbol{\beta}}\right) - \widetilde{m}_{LL,1}\left(u_1, \widehat{\boldsymbol{\beta}}\right) \right| = O_p\left(n^{-1/2} + N^{-q}\right) = o_p\left(n^{-2/5}\right).$$

Corollary 1. *Under Conditions (C1)-(C6) in the Appendix, and $h_1 \asymp n^{-1/5}$, $nN^{-4} \to \infty$, $nN^{-2q-2} \to 0$ and $nN^{-5q/2} \to 0$, for any $u_1 \in [h_1, 1-h_1]$, as $n \to \infty$, we have*

$$\sqrt{nh_1}\left\{\widehat{m}_{SBLL,1}\left(u_1, \widehat{\boldsymbol{\beta}}\right) - m_1(u_1) - b_1(u_1) h_1^2\right\} \xrightarrow{d} \mathcal{N}(0, v_1(u_1)).$$

Remark 4. Since the spline order q needs to be no smaller than 3, under the assumption of N given in Corollary 1, the same order $N \asymp n^{1/(2q+1)}$ as given in Remark 2 can be applied in the first step of spline estimation. For instance, when $q = 4$ for cubic splines, $N \asymp n^{1/9}$. In the second step of SBLL estimation, the bandwidth satisfies the optimal order $h_1 \asymp n^{-1/5}$.

3.2 Inference for index parameter $\boldsymbol{\beta}$

With the availability of asymptotic normality in Theorem 1, we can easily derive a Wald chi-square testing procedure to test whether a subset of $\boldsymbol{\beta}_l = \left(\beta_{2l}, \ldots, \beta_{pl}\right)$, $l = 1, \ldots, d$, equals to zero. Let K be an integer satisfying $2 \leq K \leq p$, and let $(k_1, \ldots k_K)$ be a subset of indices in $\{2, \ldots, p\}$. The null hypothesis of interest is: $H_0 : \beta_{k_1 l} = \beta_{k_2 l} = \cdots = \beta_{k_K l} = 0$ for the i-th index coefficients. Following Theorem 1, a Wald test statistic takes the form $\chi^2_W = \left(\widehat{\boldsymbol{\beta}}_{Kl} - \mathbf{0}_K\right)^{\mathrm{T}} \left\{\widehat{\mathrm{V}}\left(\widehat{\boldsymbol{\beta}}_{Kl}\right)\right\}^{-1} \left(\widehat{\boldsymbol{\beta}}_{Kl} - \mathbf{0}_K\right)$, where $\widehat{\boldsymbol{\beta}}_{Kl} = \left(\widehat{\beta}_{k_1 l}, \widehat{\beta}_{k_2 l}, \ldots, \widehat{\beta}_{k_K l}\right)^{\mathrm{T}}$ is the profile estimate of $\boldsymbol{\beta}_{Kl} = \left(\beta_{k_1 l}, \beta_{k_2 l}, \ldots, \beta_{k_K l}\right)^{\mathrm{T}}$, and $\left\{\widehat{\mathrm{V}}\left(\widehat{\boldsymbol{\beta}}_{Kl}\right)\right\}^{-1}$ is the inverse of the estimated asymptotic variance-covariance matrix of $\widehat{\boldsymbol{\beta}}_{Kl}$. Under H_0, χ^2_W follows asymptotically the central chi-square distribution with K degrees of freedom.

3.3 Inference for nonparametric function $m_l(\cdot)$

For a given $1 \leq l \leq d$, both main and interaction effects of X_l are related to the nonparametric function $m_l(\cdot)$. To test whether $m_l(\cdot)$ has a specific parametric form, we set up the hypothesis testing as: $H_0 : m_l(\cdot) = m_{\boldsymbol{\theta},l}(\cdot)$ versus $H_a : m_l(\cdot) \neq m_{\boldsymbol{\theta},l}(\cdot)$, where $m_{\boldsymbol{\theta},l}(\cdot)$ is a certain given parametric function with the p_θ-dimensional parameter vector $\boldsymbol{\theta}$. For example, setting $m_{\boldsymbol{\theta},l}(u_l) \equiv \theta_{l0}$ (constant), we aim to test whether there exist any interaction effects, while setting $m_{\boldsymbol{\theta},l}(u_l) = \theta_{l1} + \theta_{l2} u_l$ (a linear function), we attempt to test whether there exists a linear interaction effect between U_l and X_l. Following Fan et al. (2001) and Liang et al. (2010), we construct generalized likelihood ratio (GLR) statistics based on the SBLL estimator $\widehat{m}_{\mathrm{SBLL},l}\left(u_l, \widehat{\boldsymbol{\beta}}\right)$ given in section 3.1. First we construct a GLR statistic and establish its asymptotic distribution by using the oracle estimator $\widetilde{m}_{\mathrm{LL},l}\left(u_l, \widehat{\boldsymbol{\beta}}\right)$ assuming that all the other nonparametric functions $m_{l'}(\cdot)$ for $l' \neq l$ were known. Because of Theorem 4, the same asymptotic distribution will be satisfied by the GLR statistic by plugging in the SBLL estimates.

For example, let us consider $l = 1$. Applying the same procedure as proposed in Liang et al. (2010), under H_a, we estimate $m_{\boldsymbol{\theta},1}(u_1)$ by minimizing $\sum_{i=1}^n \left\{Y_{i,1} - m_{\boldsymbol{\theta},1}\left(U_{i1}(\widehat{\boldsymbol{\beta}}_1), \widehat{\boldsymbol{\beta}}\right) X_{i1}\right\}^2$, denoted as $\widetilde{m}_{\widehat{\boldsymbol{\theta}},1}\left(u_1, \widehat{\boldsymbol{\beta}}\right)$, and the resulting residual sum of squares under the null and alternative hypotheses are given as

$$\begin{aligned}\mathrm{RSS}_{\mathrm{LL},1}\left(H_0\right) &= \sum\nolimits_{i=1}^{n}\left\{Y_{i,1}-\widetilde{m}_{\widehat{\boldsymbol{\theta}},1}\left(U_{i1}(\widehat{\boldsymbol{\beta}}_1),\widehat{\boldsymbol{\beta}}\right)X_{i1}\right\}^2,\\ \mathrm{RSS}_{\mathrm{LL},1}\left(H_1\right) &= \sum\nolimits_{i=1}^{n}\left\{Y_{i,1}-\widetilde{m}_{\mathrm{LL},1}\left(U_{i1}(\widehat{\boldsymbol{\beta}}_1),\widehat{\boldsymbol{\beta}}\right)X_{i1}\right\}^2,\end{aligned}$$

where $\widehat{\boldsymbol{\beta}}$ and $\widetilde{m}_{\mathrm{LL},1}\left(u_1,\widehat{\boldsymbol{\beta}}\right)$ are the profile and local linear estimates of $\boldsymbol{\beta}$ and $m_1\left(u_1\right)$, respectively. It follows that a GLR statistic is defined by

$$\mathcal{T}_{\mathrm{LL},1}=\frac{n\left\{\mathrm{RSS}_{\mathrm{LL},1}\left(H_0\right)-\mathrm{RSS}_{\mathrm{LL},1}\left(H_1\right)\right\}}{2\mathrm{RSS}_{\mathrm{LL},1}\left(H_1\right)}.$$

Let

$$\Gamma_1\left(u_1\right)=E\left(X_1^2\left|U_1=u_1\right.\right)f_1\left(u_1\right),\ \Gamma_1^*\left(u_1\right)=E\left\{X_1^2\sigma^2\left(\mathbf{Z},\mathbf{X}\right)\left|U_1=u_1\right.\right\}f_1\left(u_1\right).$$

Corollary 2. *Assume Conditions (C1)-(C7) in the Appendix, and* $h_1\asymp n^{-1/5}$, $nN^{-4}\to\infty$ *and* $nN^{-2q-2}\to 0$.

(i) Suppose $H_0: m_{\boldsymbol{\theta},1}\left(\cdot\right)$ *is linear such that* $m_{\boldsymbol{\theta},1}\left(u_1\right)=\theta_{11}+\theta_{12}u_1$. *Then under* H_0, $\tau_K\mathcal{T}_{LL,1}$ *has an asymptotic* χ^2 *distribution with* df_n *degrees of freedom, where*

$$\begin{aligned}\tau_K &= \left\{K\left(0\right)-0.5\int K^2\left(u\right)du\right\}/\int\left\{K\left(u\right)-0.5\int K*K\left(u\right)du\right\}^2du,\\ df_n &= \tau_K\left\{K\left(0\right)-0.5\int K^2\left(u\right)du\right\}/h,\end{aligned}$$

and $K*K\left(u\right)$ *denotes the convolution of* K*;*

(ii) Suppose $H_0: m_{\boldsymbol{\theta},1}\left(\cdot\right)$ *is a constant such that* $m_{\boldsymbol{\theta},1}\left(u_1\right)=\theta_{10}$. *Then under* H_0, $\widetilde{\tau}_K\mathcal{T}_{LL,1}$ *has an asymptotic* χ^2 *distribution with* $\widetilde{df}_n$ *degrees of freedom, where*

$$\begin{aligned}\widetilde{\tau}_K &= \tau_K E\left\{\sigma^2\left(\mathbf{Z},\mathbf{X}\right)\right\}\left\{\int\left(\Gamma_1^*\left(u_1\right)\Gamma_1^{-1}\left(u_1\right)\right)du_1\right\}\left\{\int\left(\Gamma_1^*\left(u_1\right)\Gamma_1^{-1}\left(u_1\right)\right)^2du_1\right\}^{-1},\\ \widetilde{df}_n &= \tau_K c_K h^{-1}\left\{\int\left(\Gamma_1^*\left(u_1\right)\Gamma_1^{-1}\left(u_1\right)\right)du_1\right\}^2\left\{\int\left(\Gamma_1^*\left(u_1\right)\Gamma_1^{-1}\left(u_1\right)\right)^2du_1\right\}^{-1},\end{aligned}$$

where $c_K=K\left(0\right)-0.5\left\|K\right\|_2^2$.

Results (i) and (ii) in Corollary 2 can be proved by following the same reasoning as the proofs of Theorems 5 and 9 given in Fan et al. (2001) as well as the proofs of Theorem 5 given by Liang et al. (2010). Now we construct a sample version of GLR statistic by using the SBLL estimator

$\widehat{m}_{\mathrm{SBLL},1}\left(u_1,\widehat{\boldsymbol{\beta}}\right)$. Similarly, denote by $\widehat{m}_{\widehat{\boldsymbol{\theta}},1}\left(u_1,\widehat{\boldsymbol{\beta}}\right)$ the least squares estimator that minimizes $\sum_{i=1}^{n}\left\{\widehat{Y}_{i,1}-m_{\boldsymbol{\theta},1}\left(U_{i1}(\widehat{\boldsymbol{\beta}}_1),\widehat{\boldsymbol{\beta}}\right)X_{i1}\right\}^2$. Then a GLR statistic is defined by

$$\mathcal{T}_{\mathrm{SBLL},1}=\frac{n\left\{\mathrm{RSS}_{\mathrm{SBLL},1}\left(H_0\right)-\mathrm{RSS}_{\mathrm{SBLL},1}\left(H_1\right)\right\}}{2\mathrm{RSS}_{\mathrm{SBLL},1}\left(H_1\right)},$$

where

$$\begin{aligned}\mathrm{RSS}_{\mathrm{SBLL},1}\left(H_0\right) &= \sum_{i=1}^{n}\left\{\widehat{Y}_{i,1}-\widehat{m}_{\widehat{\boldsymbol{\theta}},1}\left(U_{i1}(\widehat{\boldsymbol{\beta}}_1),\widehat{\boldsymbol{\beta}}\right)X_{i1}\right\}^2, \\ \mathrm{RSS}_{\mathrm{SBLL},1}\left(H_1\right) &= \sum_{i=1}^{n}\left\{\widehat{Y}_{i,1}-\widehat{m}_{\mathrm{SBLL},1}\left(U_{i1}(\widehat{\boldsymbol{\beta}}_1),\widehat{\boldsymbol{\beta}}\right)X_{i1}\right\}^2.\end{aligned}$$

By the oracle property given in Theorem 4, under Conditions (C1)-(C7) and the order requirements of h_1 and N given in Corollary 1, it is easy to show that the above test statistic $\mathcal{T}_{\mathrm{SBLL},1}$ has the same asymptotic distribution as that of $\mathcal{T}_{\mathrm{LL}}$ established in Corollary 2. The implementation of such GLR test is carried out by the bootstrap method as suggested by Fan and Jiang (2007).

4 Smoothing Parameter Selection

In the profile LS estimation of $\boldsymbol{\beta}$, the nonparametric functions $m_l(\cdot)$ are approximated by cubic spline ($q=4$), where the number of interior knots is set as $N=\left[2n^{1/(2q+1)}\right]+1=\left[2n^{1/9}\right]+1$, which satisfies the optimal order of N as discussed in Remark 2. Here $[a]$ denotes the closest integer to a. After we obtain the estimate of $\boldsymbol{\beta}$, each $m_l(\cdot)$ is estimated by its B spline estimate $\widehat{m}_l\left(\cdot,\widehat{\boldsymbol{\beta}}\right)$ with the number of interior knots selected by minimizing the BIC criterion on the range $\left[n^{1/9}\right]\leq N\leq\left[2n^{1/9}\right]+1$ given as

$$\mathrm{BIC}\left(N\right)=\log\left[n^{-1}\sum_{i=1}^{n}\left\{Y_i-\widehat{m}\left(\mathbf{Z},\mathbf{X}\right)\right\}^2\right]+\frac{\log n}{n}d\left(N+q\right),$$

where $\widehat{m}\left(\mathbf{Z},\mathbf{X}\right)=\sum_{l=1}^{d}\widehat{m}_l\left(\mathbf{Z}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_l,\widehat{\boldsymbol{\beta}}\right)\mathbf{X}_l$. Then one selects the optimal number of interior knots $\widehat{N}=\mathrm{argmin}_{N\in I_N}\mathrm{BIC}(N)$. In the second step, the SBLL estimation for $m_1(\cdot)$ is performed with the optimal bandwidth $h_{1,\mathrm{opt}}$, which minimizes the total asymptotic mean integrated squared errors (AMISE):

$$\mathrm{AMISE}\left(\widehat{m}_{\mathrm{SBLL},1}\right)=\int\left[\left\{b_1\left(u_1\right)h_1^2\right\}^2+v_1\left(u_1\right)/\left(nh_1\right)\right]f_1\left(u_1\right)du_1.$$

It is easy to show that the optimal bandwidth $h_{1,\mathrm{opt}}$ is

$$h_{1,\mathrm{opt}}=\left\{\frac{n^{-1}\int v_1\left(u_1\right)f_1\left(u_1\right)du_1}{4\int b_1\left(u_1\right)^2f_1\left(u_1\right)du_1}\right\}^{1/5}.$$

In this paper, we use Epanechnikov kernel function, $K(u) = \frac{3}{4}(1-u^2) I(|u| \le 1)$. The optimal bandwidth $h_{1,\mathrm{opt}}$ is estimated by

$$\widehat{h}_{1,\mathrm{opt}} = \left[(4n)^{-1} \left\{ \sum_{i=1}^{n} \widehat{v}_1 \left(\mathbf{X}_i^{\mathrm{T}} \widehat{\boldsymbol{\beta}}_1 \right) \right\} / \left\{ \sum_{i=1}^{n} \widehat{b}_1 \left(\mathbf{X}_i^{\mathrm{T}} \widehat{\boldsymbol{\beta}}_1 \right)^2 \right\} \right]^{1/5},$$

where

$$\begin{aligned} \widehat{b}_1(u_1) &= 2^{-1} \mu_2(K) \widehat{\ddot{m}}_1 \left(u_1, \widehat{\boldsymbol{\beta}} \right), \\ \widehat{v}_1(u_1) &= \left\{ \widehat{E} \left(X_1^2 | u_1 \right) \right\}^{-2} \widehat{f}_1^{-1}(u_1) \|K\|_2^2 \widehat{E} \left\{ X_1^2 \widehat{e}^2 (\mathbf{Z}, \mathbf{X}) | u_1 \right\}, \end{aligned}$$

in which $\widehat{f}_1(u_1)$ is a kernel density estimate of $f_1(u_1)$, $\widehat{e}(\mathbf{Z}, \mathbf{X}) = Y - \sum_{l=1}^{d} \widehat{m}_l \left(\mathbf{Z}^{\mathrm{T}} \widehat{\boldsymbol{\beta}}_l, \widehat{\boldsymbol{\beta}} \right) \mathbf{X}_l$, and $\widehat{\ddot{m}}_1 \left(u_1, \widehat{\boldsymbol{\beta}} \right)$, $\widehat{E} \left(X_1^2 | u_1 \right)$ and $\widehat{E} \left\{ X_1^2 \widehat{\sigma}^2 (\mathbf{Z}, \mathbf{X}) | u_1 \right\}$ are respectively the spline estimators given as follows: $\widehat{\ddot{m}}_1 \left(u_1, \widehat{\boldsymbol{\beta}} \right) = \sum_{s=1}^{J_n} \ddot{B}_{s,1}(u_1) \widehat{\lambda}_{s,1}(\widehat{\boldsymbol{\beta}})$, $\widehat{E} \left(X_1^2 | u_1 \right) = \sum_{s=1}^{J_n} B_{s,1}(u_1) \widehat{\zeta}_{s,1}$, $\widehat{E} \left\{ X_1^2 \widehat{e}^2 (\mathbf{Z}, \mathbf{X}) | u_1 \right\} = \sum_{s=1}^{J_n} B_{s,1}(u_1) \widehat{\eta}_{s,1}$, where $\widehat{\zeta}_1 = \left(\widehat{\zeta}_{s,1} \right)_{s=1}^{J_n}$ and $\widehat{\eta}_1 = \left(\widehat{\eta}_{s,1} \right)_{s=1}^{J_n}$ are obtained by minimizing, respectively,

$$\sum_{i=1}^{n} \left\{ X_{i1}^2 - \sum_{s=1}^{J_n} B_{s,1}(U_{i1}) \zeta_1 \right\}^2 \text{ and } \sum_{i=1}^{n} \left\{ X_{i1}^2 \widehat{e}_i^2 - \sum_{s=1}^{J_n} B_{s,1}(U_{i1}) \eta_1 \right\}^2,$$

with

$$\widehat{e}_i^2 = \left\{ Y_i - \sum_{l=1}^{d} \sum_{s=1}^{J_n} B_{s,l} \left(U_{il}(\widehat{\boldsymbol{\beta}}_l) \right) \widehat{\lambda}_{s,l}(\widehat{\boldsymbol{\beta}}) X_{il} \right\}^2.$$

5 Simulation Experiments

In this section, we conduct simulation studies to evaluate the performance of the proposed estimation methods. We generate $\mathbf{Z}_i = (Z_{i1}, Z_{i2}, Z_{i3})^{\mathrm{T}}$ independently from Uniform $[0,1]$, X_{i1} from Bernoulli $(0.5) - 0.5$, and $(X_{i2}, X_{i3})^{\mathrm{T}}$ from a bivariate normal distribution with mean 0, variance 1 and covariance 0.2. Set true parameters as $\boldsymbol{\beta}_1 = \frac{1}{\sqrt{14}}(2,1,3)^{\mathrm{T}}$, $\boldsymbol{\beta}_2 = \frac{1}{\sqrt{14}}(3,2,1)^{\mathrm{T}}$ and $\boldsymbol{\beta}_2 = \frac{1}{\sqrt{14}}(2,3,1)^{\mathrm{T}}$. Also set $m_1(u_1) = 5\sin(\pi u_1)$, $m_2(u_2) = 5\cos(\pi u_2) - 10/3$, and $m_3(u_3) = 5\{\sin(\pi u_3) + \cos(\pi u_3)\} - 10/3$. The random errors ε_i are generated from $N(0, \sigma^2)$ with $\sigma = 1$. Then Y_i, $1 \le i \le n$, are generated from the following VICM model:

$$Y_i = m_1 \left(\mathbf{Z}_i^{\mathrm{T}} \boldsymbol{\beta}_1 \right) X_{i1} + m_2 \left(\mathbf{Z}_i^{\mathrm{T}} \boldsymbol{\beta}_2 \right) X_{i2} + m_3 \left(\mathbf{Z}_i^{\mathrm{T}} \boldsymbol{\beta}_3 \right) X_{i3} + \varepsilon_i.$$

The sample size is set as $n = 200, 500, 1000$, respectively, and 500 simulation replications are run to draw summary statistics. Table 1 shows the empirical coverage rates of the 95% confidence intervals for $\boldsymbol{\beta}_1 = (\beta_{11}, \beta_{12}, \beta_{13})^{\mathrm{T}}$, $\boldsymbol{\beta}_2 = (\beta_{21}, \beta_{22}, \beta_{23})^{\mathrm{T}}$ and $\boldsymbol{\beta}_3 = (\beta_{31}, \beta_{32}, \beta_{33})^{\mathrm{T}}$ for $n = 200, 500, 1000$. The standard errors are calculated according to the asymptotic formula given in (17). We can observe that the coverage rates get closer to 95% as the sample size increases. This result is confirmatory to the asymptotic normal distribution of the parameter estimators established in Theorem 1.

Insert Table 1 here

Tables 2 presents the average bias. We can observe that the biases are close to 0 for all cases. This result confirms the asymptotic property that the parameter estimators are asymptotically unbiased as given in Theorem 1. It also indicates that estimation consistency is achieved even with a relatively small sample size $n = 200$. Table 3 shows the average asymptotic standard error (ASE) calculated according to Theorem 1 and the empirical standard error (ESE) among 500 replications for $n = 200, 500, 1000$. With no surprise, the standard errors become smaller as n increases, due to the fact of root-n consistency of the parameter estimators. it is more important that the ASEs are very similar to the corresponding ESEs for all cases, suggesting that the asymptotic covariance matrix is correctly derived.

Insert Tables 2 and 3 here

To evaluate the performance of the two-step SBLL estimator $\widehat{m}_{\mathrm{SBLL},l}(\cdot)$ for a given l, we define the median integrated squared error (MISE) as the median value of the $\mathrm{ISE}(\widehat{m}_{\mathrm{SBLL},l}) = n^{-1}\sum_{i=1}^{n}\left\{\widehat{m}_{\mathrm{SBLL},l}(U_{il}\left(\widehat{\boldsymbol{\beta}}_l\right), \widehat{\boldsymbol{\beta}}) - m_l(U_{il})\right\}^2$ among the 500 replications. The MISE for the oracle estimator $\widetilde{m}_{\mathrm{LL},l}(\cdot)$ is defined in the same way. Table 4 shows the MISEs for the two-step SBLL estimators $\widehat{m}_{\mathrm{SBLL},l}$ and the oracle estimators $\widetilde{m}_{\mathrm{LL},l}(\cdot)$ for$1 \leq l \leq 3$, $n = 200, 500, 1000$. We can observe that the MISE values for the SBLL estimators become closer to those values for the oracle estimators as n increases, and the MISE values decrease as n increases.

Insert Table 4 here

To visualize the estimated functions, Figure 3 displays the estimated curves by the SBLL estimator $\widehat{m}_{\mathrm{SBLL},l}(\cdot)$ (thick line), with the upper and lower 95% pointwise confidence intervals (upper and lower thick lines), and by the oracle estimator $\widetilde{m}_{\mathrm{LL},l}(\cdot)$ (thin line) and the true function $m_l(\cdot)$ (dashed line) for $l = 1, 2, 3$ in the setting of $n = 200$. It is evident that the proposed SBLL estimators perform well.

Insert Figure 3 here

The proposed estimation procedure is computationally fast. We ran the above simulation experiments on Macbook Pro with 2 GHz Intel Core. The average operation time per simulated dataset in R is 1.375 seconds, 2.429 seconds and 4.068 seconds for sample size $n = 200, 500, 1000$, respectively, including the total running time of generating one data sample and computing both the profile LS estimate of $\boldsymbol{\beta}_l$, $l = 1, 2, 3$, and the SBLL estimate of $m_l(\cdot)$, $l = 1, 2, 3$.

6 Application

In this section we illustrate our method via the analysis of body fat dataset introduced in Section 1. It is suggested in the public health science literature that percentage of body fat is an important biomarker of health status (see Bailey (1994)). Since available procedures accurately measuring body fat are all complex, expensive, and impractical on the daily use, it is desirable to develop some practical formulas that enable to calculate body fat percentage conveniently. In many studies, body circumference measurements are extensively used as surrogate variables to approximate the determination of body fat percentage; see the published work by Behnke and Wilmore (1974), Wilmore (1976), Katch and McArdle (1977), among others. However, their formulas are mostly derived by using the ordinary multiple linear regression models. As a matter of fact, in addition to circumference measurements, body fat is also potentially related to age, weight, height and other personal characteristics. As an illustration, in this analysis we consider 6 circumferences as $\mathbf{Z}$ covariates to form index coefficients and two other covariates, age and fat free weight, are treated as the $\mathbf{X}$ covariates of interest to build a varying index coefficient model. Our model will reveal how the relationships of the body fact percent with age and fat free weight are modified by the profiles of circumference indices.

To specify the model, let the response variable Y =log(percent body fat + 1). The percent body fat is measured by Brozek's equation 457/Density (gm/cmˆ3)−414.2. Also the centered and standardized versions of 6 circumferences (cm) serve as the vector of covariates $\mathbf{Z}$ in the index coefficient including Z_1 = chest circumference, Z_2 = abdomen circumference, Z_3 = hip circumference, Z_4 = thigh circumference, Z_5 = forearm circumference, and Z_6 = wrist circumference. Covariates of interest include $X_1 = 1$, X_2 =age (yrs), X_3 =fat free weight =(1 - fraction of body fat) ×weight, and both X_2 and X_3 are also centered and standardized. Thus our model takes the following form:

$$Y = \sum_{l=1}^{3} m_l \left(\mathbf{Z}^{\mathrm{T}} \boldsymbol{\beta}_l\right) X_l + \varepsilon, \tag{19}$$

where $m_l(\cdot)$ are unknown nonparametric functions and $\boldsymbol{\beta}_l = (\beta_{l1}, \ldots, \beta_{l6})^{\mathrm{T}}$ are unknown coefficient vectors for $l = 1, 2, 3$, both of which will be estimated.

To begin, we first apply a principle component analysis (PCA) on $\mathbf{Z}$, which allows us to form an index variable $U^{\mathrm{PCA}} = \mathbf{Z}^{\mathrm{T}} w$, where w is the vector of loadings for the first principle component. Then we fit a varying coefficient model $Y = \sum_{l=1}^{3} m_l \left(U^{\mathrm{PCA}}\right) X_l + \varepsilon$. This simple analysis provides us with reasonable initial estimates of the nonparametric functions $\widehat{m}_l^{\mathrm{ini}}(\cdot)$. In the meanwhile, the initial estimates of $\boldsymbol{\beta}_l$, denoted by $\widehat{\boldsymbol{\beta}}_l^{\mathrm{ini}}$, can be obtained by minimizing

$$L_n(\boldsymbol{\beta}) = 2^{-1} \sum_{i=1}^{n} \left\{ Y_i - \sum_{l=1}^{3} \widehat{m}_l^{\mathrm{ini}} \left(\mathbf{Z}_i^{\mathrm{T}} \boldsymbol{\beta}_l\right) X_{il} \right\}^2 .$$

In our analysis, the number of interior knots and the bandwidth are chosen based on the criteria discussed in Section 4. Fitting model (19) by the proposed profile LS estimation procedure, we obtain the estimates (EST) of $\boldsymbol{\beta}_l$ and their standard errors (SE) according to (16), $1 \leq l \leq 3$, as well as their lower bound (LB) and upper bound (UB) of 95% confidence intervals (CI) according to Theorem 1. Table 5 lists all the results, including the corresponding p-values.

Insert Table 5 here

Table 5 indicates that for $\boldsymbol{\beta}_1$ in the first coefficient index with X_1 =intercept, the estimated coefficients for Z_1 (chest), Z_2 (abdomen), Z_3 (hip) and Z_5 (forearm) are significantly different from zero. This means that in terms of the main effects of circumferences, chest, abdomen, hip and forearm are significant factors on body fat percentage. For $\boldsymbol{\beta}_2$ with X_2 =age, the abdomen and thigh circumferences (Z_2 and Z_4) appear to be important factors for the interactions of circumferences and age. In other words, the association between body fat percent and age is modified by a combination of abdomen and thigh circumferences. For $\boldsymbol{\beta}_3$ with X_3 =fat free weight, a combination of chest, abdomen and thigh circumferences alters the association between body fat percent and fat free weight.

We also conduct the Wald chi-square test described in Section 3.2 for each subset of $\boldsymbol{\beta}_l = (\beta_{l2}, \ldots, \beta_{l6})^{\mathrm{T}}$, $l = 1, 2, 3$, and results are summarized in Tables 6. Whenever a subset contains more than three $\mathbf{Z}$ components, it is found to be significant with p-value much smaller than 0.01. Therefore, we do not report the results of four circumferences or more in Table 6 for the sake of saving space. Tables 6 lists critical values (C-value) and p-values for each significant subset of the three $\mathbf{Z}$ variables. We observe that for $\boldsymbol{\beta}_1$, all subsets are significant at significance level 0.05. For $\boldsymbol{\beta}_2$ and $\boldsymbol{\beta}_3$, $\boldsymbol{\beta}_2$ has just one more significant subset, i.e. $(2, 5, 6)$, than $\boldsymbol{\beta}_3$, and all the other significant subsets are the same. To examine interaction effects of the circumferences with the intercept, age and fat free weight, we conduct the GLR test proposed in Section 3.3. For the intercept, age and fat free weight, we obtain the p-values of GLR test statistics all less than 0.0001 in the following hypothesis tests. First, we consider $H_0 : m_l(\cdot)$ is constant (or absence of interaction effect for covariate X_l) versus $H_a : m_l(\cdot)$ is not constant. Second, we look at $H_0 : m_l(\cdot)$ is linear (or existence of a linear interaction with X_l) versus $H_a : m_l(\cdot)$ is nonlinear. The very small p-values are not in favor of the null hypotheses, implying strong nonlinear main effects of the circumferences and more importantly the presence of strong nonlinear interaction effects of the circumferences with age and fat free weight. Such findings are consistent with the graphic evidence presented in Figure 4.

Insert Table 6 here

To further illustrate the change pattern of the estimates of $m_l(\cdot)$ along with the estimated circumference index $\mathbf{Z}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_l$, Figure 4 displays the fitted curves obtained by our two-step SBLL method (middle solid line), the one-step spline estimate given in (15) (middle dashed line), and their 95% pointwise confidence intervals (lower and upper lines) of $m_l(\cdot)$, $1 \le l \le 3$. In addition, the estimates $\widehat{m}_{\boldsymbol{\theta},l} = \widehat{\theta}_{l0}$ (horizontal dashed lines) by assuming $m_l(\cdot)$ is a constant and $\widehat{m}_{\boldsymbol{\theta},l} =$

$\widehat{a}_l + \widehat{b}_l U_l \left(\widehat{\boldsymbol{\beta}}_l\right)$ (straight thin lines) by assuming linearity of $m_l(\cdot)$ are included for comparison.

Insert Figure 4 here

The first plot for intercept of Figure 4 clearly shows that the estimated function $\widehat{m}_1(\cdot)$ is an increasing function of the circumference index U_1 and the increasing speed declines as the index value increases. The parametric models by assuming constant and linear coefficients, respectively, apparently missed the opportunity to capture this feature. This finding is clearly corroborative with the GLR test results. In Table 5, the estimated coefficients of significant $\mathbf{Z}$ are positive, so it is of scientific importance to unveil the pattern that $\widehat{m}_1(\cdot)$ increases along with the higher chest, abdomen, hip and forearm circumferences. The second plot for age of Figure 4 shows that the modification by the circumferences on the association of body fat percent and age is highly nonlinear. Note that this association with age starts from a positive value and drops quickly to around zero as the important circumference measurements increase, and then becomes stable. This pattern of change illustrates the complexity in terms of interaction effects between body circumferences and age ranged between 22 and 81 years old. The other two simple parametric models cannot provide these informative relationships of human body development. The third plot for fat free weight of Figure 4 again demonstrates that the interaction effect between the circumferences and fat free weight is not a constant or linear. Moreover, the one-step spline method and the two-step SBLL method yield similar estimated curves for m_1 and m_2, while the former method results in a curve with more waves for m_3. However, the two curves by the two methods clearly have the same change pattern in general.

Finally, we compare model (19) with the varying coefficient model (VCM) of the following form:

$$Y = \sum\nolimits_{l=1}^{3} m_l \left(U^{\mathrm{PCA}}\right) X_l + \varepsilon. \tag{20}$$

We perform the leave-one-out cross validations for the proposed model (19) and the VCM (20), as well as two linear models by assuming constant and linear coefficient functions, respectively. The cross-validation (CV) errors are 0.042, 0.059, 0.066, 0.215, respectively. The proposed model has the smallest CV error, while the linear model with only main effects of age and fat free weight has the largest CV error, consistent with what we learn above from all the figures and tables.

7 Discussion

In this paper, we propose a new class of semiparametric models with varying index coefficients, which allows us to study nonlinear interactive effects that are of scientific importance in the understanding of the response-covariate relationship. We demonstrate that regression coefficient of a covariate can be altered or directed by a nonlinear function of multiple other covariates. The proposed modeling framework gives rise to a rich class of regression models, including many popular semiparametric models as special cases. Utilizing the least squares estimation approach, we develop a profile estimation procedure that is both conceptually simple and computationally efficient, and the resulting estimators are consistent and asymptotically normal.

One of the co-authors is currently involved in multiple collaborative projects studying effects of environmental pollutants on the somatic growth of children in USA. We believe that the proposed model has a great potential to investigate the developmental effects resulting from ubiquitous environmental exposure to known or suspected endocrine disrupting components(EDCs) among children. In this kind of study, a single EDC has typically a weak effect to alter the rate of somatic growth but a bundle of EDCs (termed as a mixture of EDCs in environmental health sciences) is possibly attributive to altered hormone secretion and hence modifies the rate of growth among children in a nonlinear fashion. Similar analysis has been also called in the study of exposure to EDCs affecting the pregnancy as well as early life of children. Our VICM model provides a comprehensive way to understand interactions between environmental factors and physiological variables in the study of human growth and diseases.

Our future studies will be focused on the extension of the proposed model for longitudinal data as well as on discrete or categorical response variables along the line of quasi-likelihood estimation inference. Since the proposed model is quite general, it may involve a large number of parameters (e.g. index coefficients) to estimate given that we assume each coefficient function depends on different index parameters. In order to achieve model sparsity, variable selection procedures via regularization will be investigated as future work. Also a user friendly R package for the implementation of the VICM will be released to the public.

Appendix

A.1 Assumptions

For positive numbers a_n and b_n, let $a_n \asymp b_n$ denote that $\lim_{n\to\infty} a_n/b_n = c$, where c is some nonzero constant. For any vector $\boldsymbol{\zeta} = (\zeta_1, \ldots, \zeta_s)^{\mathrm{T}} \in R^s$, denote $\|\boldsymbol{\zeta}\|_\infty = \max_{1\le l\le s}|\zeta_l|$. For any symmetric matrix $\mathbf{A}_{s\times s}$, denote its L_r norm as $\|\mathbf{A}\|_r = \max_{\boldsymbol{\zeta}\in\mathbf{s},\boldsymbol{\zeta}\neq\mathbf{0}} \|\mathbf{A}\boldsymbol{\zeta}\|_r \|\boldsymbol{\zeta}\|_r^{-1}$. For any matrix $\mathbf{A} = (A_{ij})_{i=1,j=1}^{s,t}$, denote $\|\mathbf{A}\|_\infty = \max_{1\le i\le s}\sum_{j=1}^t |A_{ij}|$.

We denote the space of p-th order smooth function as $C^{(p)}[0,1] = \left\{\varphi \left| \varphi^{(p)} \in C[0,1]\right.\right\}$. Let $C^{0,1}(\mathcal{X}_w)$ be the space of Lipschitz continuous function on $\mathcal{X}_w$, i.e.,

$$C^{0,1}(\mathcal{X}_w) = \left\{\varphi : \|\varphi\|_{0,1} = \sup_{w\neq w', w, w'\in\mathcal{X}_w} \frac{|\varphi(w) - \varphi(w')|}{|w - w'|} < \infty\right\},$$

in which $\|\varphi\|_{0,1}$ is the $C^{0,1}$-norm of φ. To establish the consistency and asymptotic normality for the proposed estimators, we need the following regularity conditions.

(C1) For every $1 \le l \le d$, the density function $f_{U_l(\boldsymbol{\beta}_l)}(\cdot)$ of random variable $U_l(\boldsymbol{\beta}_l) = \mathbf{Z}^{\mathrm{T}}\boldsymbol{\beta}_l$ is bounded away from 0 on $\mathcal{S}_w$ and $f_{U_l(\boldsymbol{\beta}_l)}(\cdot) \in C^{0,1}(\mathcal{S}_w)$ for $\boldsymbol{\beta}_l$ in the neighborhood of $\boldsymbol{\beta}_l^0$, where $\mathcal{S}_w = \left\{\mathbf{Z}^{\mathrm{T}}\boldsymbol{\beta}_l, \mathbf{Z} \in S\right\}$ and S is a compact support set of $\mathbf{Z}$. Without loss of generality, we assume $\mathcal{S}_w = [0,1]$.

(C2) For every $1 \le l \le d$, the nonparametric function $m_l \in C^{(q)}[0,1]$.

(C3) The conditional variance function $\sigma^2(\mathbf{z},\mathbf{x})$ is measurable and bounded above from C_σ, for some constant $0 < C_\sigma < \infty$.

(C4) There exist constants $0 < c_Q \le C_Q < \infty$, such that $c_Q \le Q(\mathbf{z}) = E\left(\mathbf{X}\mathbf{X}^{\mathrm{T}} | \mathbf{Z} = \mathbf{z}\right) \le C_Q$ for all $\mathbf{z} \in S$.

(C5) For $1 \le k \le p$ and $1 \le l \le d$, $g_{l,k}^0 \in C^{(1)}[0,1]$.

(C6) The kernel function K is a symmetric probability density, supported on $[-1,1]$ and $K \in C^{0,1}[-1,1]$.

(C7) The functions $u^3K(u)$ and $u^3K'(u)$ are bounded and $\int u^4K(u)\,du < \infty$. $E|\varepsilon|^4 < \infty$.

It is noteworthy that Condition (C1) is the same as Condition (d) in Cui et al. (2011). Condition (C2) is given in Theorem 2.1 of Zhou et al. (1998). Condition (C3) is the same as Condition (C5) of Xue and Yang (2006). Condition (C4) is given in Condition (C2) of Xia and Härdle (2006) and Condition (C5) of Xue and Liang (2010). Condition (C5) gives the smoothness condition of functions $g_{l,k}^{0}$ defined in (11). Condition (C6) is a common assumption on the kernel function in the nonparametric smoothing literature. Condition (C7) is the same as Conditions (A3) and (A4) in Fan et al. (2001), which is used for obtaining the asymptotic distribution of the GLR statistic.

A.2 Proofs of Theorems 1 and 2

Denote $\mathbf{Y} = (Y_1, \ldots, Y_n)^{\mathrm{T}}$ and $\mathbf{m} = \left\{m\left(\mathbf{Z}_1, \mathbf{X}_1, \boldsymbol{\beta}^0\right), \ldots, m\left(\mathbf{Z}_n, \mathbf{X}_n, \boldsymbol{\beta}^0\right)\right\}^{\mathrm{T}}$. By (7), $\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta})$ can be decomposed into $\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}) = \widehat{\boldsymbol{\lambda}}_m(\boldsymbol{\beta}) + \widehat{\boldsymbol{\lambda}}_e(\boldsymbol{\beta})$, where

$$\widehat{\boldsymbol{\lambda}}_m(\boldsymbol{\beta}) = \left\{\mathbf{D}(\boldsymbol{\beta})^{\mathrm{T}}\mathbf{D}(\boldsymbol{\beta})\right\}^{-1}\mathbf{D}(\boldsymbol{\beta})^{\mathrm{T}}\mathbf{m}, \widehat{\boldsymbol{\lambda}}_e(\boldsymbol{\beta}) = \left\{\mathbf{D}(\boldsymbol{\beta})^{\mathrm{T}}\mathbf{D}(\boldsymbol{\beta})\right\}^{-1}\mathbf{D}(\boldsymbol{\beta})^{\mathrm{T}}(\mathbf{Y} - \mathbf{m}). \tag{A.1}$$

We first present several lemmas which will be used in the proofs of Theorems 1 and 2. Define

$$\mathbf{V}(\boldsymbol{\beta}) = E\left(D_i(\boldsymbol{\beta})D_i(\boldsymbol{\beta})^{\mathrm{T}}\right), \widehat{\mathbf{V}}(\boldsymbol{\beta}) = n^{-1}\mathbf{D}(\boldsymbol{\beta})^{\mathrm{T}}\mathbf{D}(\boldsymbol{\beta}). \tag{A.2}$$

Lemma A.1. *Under Conditions (C1) and (C4), for any vector* $\boldsymbol{\alpha} = \left\{\left(\boldsymbol{\alpha}_1^T, \ldots, \boldsymbol{\alpha}_d^T\right)^T\right\}_{dJ_n \times 1}$ *with* $\boldsymbol{\alpha}_l = (\alpha_{s,l} : 1 \le s \le J_n)^T$, *there are constants* $0 < c_V < C_V < \infty$, *such that for* $\forall \boldsymbol{\beta} \in \Theta$ *and for large enough* n,

$$c_V J_n^{-1}\boldsymbol{\alpha}^T\boldsymbol{\alpha} \le \boldsymbol{\alpha}^T\mathbf{V}(\boldsymbol{\beta})\boldsymbol{\alpha} \le C_V J_n^{-1}\boldsymbol{\alpha}^T\boldsymbol{\alpha},\ C_V^{-1}J_n\boldsymbol{\alpha}^T\boldsymbol{\alpha} \le \boldsymbol{\alpha}^T\mathbf{V}(\boldsymbol{\beta})^{-1}\boldsymbol{\alpha} \le c_V^{-1}J_n\boldsymbol{\alpha}^T\boldsymbol{\alpha}. \tag{A.3}$$

$$\begin{aligned}\sup_{1\le s,s'\le J_n, 1\le l\le d}&\left|n^{-1}\sum\nolimits_{i=1}^{n} D_{i,sl}(\boldsymbol{\beta}_l)D_{i,s'l}(\boldsymbol{\beta}_l) - E\left\{D_{i,sl}(\boldsymbol{\beta}_l)D_{i,s'l}(\boldsymbol{\beta}_l)\right\}\right| \\ &= O_{a.s.}\left(\sqrt{J_n^{-1}n^{-1}\log n}\right), \end{aligned}\tag{A.4}$$

$$\begin{aligned}\sup_{1\le s,s'\le J_n, l\ne l'}&\left|n^{-1}\sum\nolimits_{i=1}^{n} D_{i,sl}(\boldsymbol{\beta}_l)D_{i,s'l'}(\boldsymbol{\beta}_l) - E\left\{D_{i,sl}(\boldsymbol{\beta}_l)D_{i,s'l'}(\boldsymbol{\beta}_l)\right\}\right| \\ &= O_{a.s.}\left(J_n^{-1}\sqrt{n^{-1}\log n}\right). \end{aligned}\tag{A.5}$$

Proof. By Theorem 5.4.2 of DeVore and Lorentz (1993) and Condition (C1), one has for large enough n, there are constants $0 < c_l \le C_l < \infty$, for $\forall \boldsymbol{\beta} \in \Theta$, such that

$$c_l J_n^{-1} \boldsymbol{\alpha}_l^{\mathrm{T}} \boldsymbol{\alpha}_l \le \boldsymbol{\alpha}_l^{\mathrm{T}} E\left(\mathbf{B}_q\left(U_{il}(\boldsymbol{\beta}_l)\right) \mathbf{B}_q\left(U_{il}(\boldsymbol{\beta}_l)\right)^{\mathrm{T}}\right) \boldsymbol{\alpha}_l \le C_l J_n^{-1} \boldsymbol{\alpha}_l^{\mathrm{T}} \boldsymbol{\alpha}_l.$$

Let $G_{il} = \sum_s \alpha_{s,l} B_{s,q}\left(U_{il}(\boldsymbol{\beta}_l)\right)$ and $G_i = (G_{i1}, \ldots, G_{id})^{\mathrm{T}}$. By Conditions (C1) and (C4) and the above result, for large enough n,

$$\begin{aligned}
\boldsymbol{\alpha}^{\mathrm{T}} E\left(D_i(\boldsymbol{\beta}) D_i(\boldsymbol{\beta})^{\mathrm{T}}\right) \boldsymbol{\alpha} &= \sum\nolimits_{l,l'} \sum\nolimits_{s,s'} E\left\{\alpha_{s,l}\alpha_{s',l'} B_{s,q}\left(U_{il}(\boldsymbol{\beta}_l)\right) B_{s',q}\left(U_{il'}(\boldsymbol{\beta}_{l'})\right) X_{il} X_{il'}\right\} \\
&= E\left(\sum\nolimits_l G_{il} X_{il}\right)^2 = E\left(G_i^{\mathrm{T}} \mathbf{X}_i \mathbf{X}_i^{\mathrm{T}} G_i\right) \ge c_Q E\left(G_i^{\mathrm{T}} G_i\right) \\
&= C_Q \sum\nolimits_l \boldsymbol{\alpha}_l^{\mathrm{T}} E\left(\mathbf{B}_q\left(U_{il}(\boldsymbol{\beta}_l)\right) \mathbf{B}_q\left(U_{il}(\boldsymbol{\beta}_l)\right)^{\mathrm{T}}\right) \boldsymbol{\alpha}_l \ge C_Q d \min\left(c_l\right) J_n^{-1} \boldsymbol{\alpha}^{\mathrm{T}} \boldsymbol{\alpha}.
\end{aligned}$$

Similarly it can be proved that $\boldsymbol{\alpha}^{\mathrm{T}} E\left(D_i(\boldsymbol{\beta}) D_i(\boldsymbol{\beta})^{\mathrm{T}}\right) \boldsymbol{\alpha} \le C_Q d \max\left(c_l\right) J_n^{-1} \boldsymbol{\alpha}^{\mathrm{T}} \boldsymbol{\alpha}$. The second result in (A.3) follows directly from the first result. Results A.4 and A.5 can be proved by Bernstein's inequality in Bosq (1998). ∎

By Lemma A.1, one has with probability approaching 1, for large enough n, for $\forall \boldsymbol{\beta} \in \Theta$,

$$c_V J_n^{-1} \boldsymbol{\alpha}^{\mathrm{T}} \boldsymbol{\alpha} \le \boldsymbol{\alpha}^{\mathrm{T}} \widehat{\mathbf{V}}(\boldsymbol{\beta}) \boldsymbol{\alpha} \le C_V J_n^{-1} \boldsymbol{\alpha}^{\mathrm{T}} \boldsymbol{\alpha},\ C_V^{-1} J_n \boldsymbol{\alpha}^{\mathrm{T}} \boldsymbol{\alpha} \le \boldsymbol{\alpha}^{\mathrm{T}} \widehat{\mathbf{V}}(\boldsymbol{\beta})^{-1} \boldsymbol{\alpha} \le c_V^{-1} J_n \boldsymbol{\alpha}^{\mathrm{T}} \boldsymbol{\alpha} \tag{A.6}$$

for any vector $\boldsymbol{\alpha} = \left\{\left(\boldsymbol{\alpha}_1^{\mathrm{T}}, \ldots, \boldsymbol{\alpha}_d^{\mathrm{T}}\right)^{\mathrm{T}}\right\}_{dJ_n \times 1}$ with $\boldsymbol{\alpha}_l = (\alpha_{s,l} : 1 \le s \le J_n)^{\mathrm{T}}$. By (A.3) and Demko (1986), it can be proved that for $\forall \boldsymbol{\beta} \in \Theta$ and for large enough n, there is a constant $0 < C_V^* < \infty$ such that $\left\|\mathbf{V}(\boldsymbol{\beta})^{-1}\right\|_\infty \le C_V^* J_n$. Following this result, (A.4) and (A.5), it can be proved that for $\forall \boldsymbol{\beta} \in \Theta$, $\left\|\widehat{\mathbf{V}}(\boldsymbol{\beta})^{-1}\right\|_\infty = O_p\left(J_n\right)$. Let $\mathbf{E} = \mathbf{Y} - \mathbf{m} = (\varepsilon_1, \ldots, \varepsilon_n)^{\mathrm{T}}$.

Lemma A.2. *Under Conditions (C1), (C3) and (C4), for* $\forall \boldsymbol{\beta} \in \Theta$, $\left\|n^{-1} \mathbf{D}(\boldsymbol{\beta})^{\mathrm{T}} \mathbf{E}\right\|_2 = O_p\left(n^{-1/2}\right)$.

Proof. With probability approaching 1,

$$\begin{aligned}
\left\|n^{-1} \mathbf{D}(\boldsymbol{\beta})^{\mathrm{T}} \mathbf{E}\right\|_2^2 &= \sum\nolimits_{l,s} \left\{n^{-1} \sum\nolimits_{i=1}^n B_{s,q}\left(U_{il}(\boldsymbol{\beta}_l)\right) X_{il} \varepsilon_i\right\}^2 \\
&\asymp n^{-2} \sum\nolimits_{l,s} \sum\nolimits_{i=1}^n E\left\{B_{s,q}\left(U_{il}(\boldsymbol{\beta}_l)\right) X_{il} \varepsilon_i\right\}^2 = O\left(n^{-1}\right).
\end{aligned}$$

∎

The proposition below presents the convergence rate of the estimators $\widehat{m}_l(u_l, \boldsymbol{\beta}^0)$ and $\widehat{\dot{m}}_l(u_l, \boldsymbol{\beta}^0)$ for the nonparametric function $m_l\left(u_l\right)$ and its first derivative $\dot{m}_l\left(u_l\right)$, for $l = 1, \ldots, d$.

Proposition A.1. *Under Conditions (C1)-(C4), and $N \to \infty$ and $nN^{-1} \to \infty$, as $n \to \infty$ one has (i) $\left|\widehat{m}_l(u_l, \boldsymbol{\beta}^0) - m_l(u_l)\right| = O_p\left(n^{-1/2}N^{1/2} + N^{-q}\right)$ uniformly for any $u_l \in [0,1]$; and (ii) under $N \to \infty$ and $nN^{-3} \to \infty$, as $n \to \infty$, $\left|\widehat{\dot{m}}_l(u_l, \boldsymbol{\beta}^0) - \dot{m}_l(u_l)\right| = O_p\left(n^{-1/2}N^{3/2} + N^{-q+1}\right)$ uniformly for any $u_l \in [0,1]$.*

Proof. Let $\widehat{\boldsymbol{\lambda}}_e(\boldsymbol{\beta}) = \left\{\widehat{\boldsymbol{\lambda}}_{1,e}(\boldsymbol{\beta})^{\mathrm{T}}, \ldots, \widehat{\boldsymbol{\lambda}}_{d,e}(\boldsymbol{\beta})^{\mathrm{T}}\right\}^{\mathrm{T}}$, where $\widehat{\boldsymbol{\lambda}}_{l,e}(\boldsymbol{\beta}) = \left\{\widehat{\lambda}_{s,l,e}(\boldsymbol{\beta}) : 1 \le s \le J_n\right\}^{\mathrm{T}}$ and $\widehat{\boldsymbol{\lambda}}_m(\boldsymbol{\beta}) = \left\{\widehat{\boldsymbol{\lambda}}_{1,m}(\boldsymbol{\beta})^{\mathrm{T}}, \ldots, \widehat{\boldsymbol{\lambda}}_{d,m}(\boldsymbol{\beta})^{\mathrm{T}}\right\}^{\mathrm{T}}$, where $\widehat{\boldsymbol{\lambda}}_{l,m}(\boldsymbol{\beta}) = \left\{\widehat{\lambda}_{s,l,m}(\boldsymbol{\beta}) : 1 \le s \le J_n\right\}^{\mathrm{T}}$. Thus

$$\widehat{m}_l(u_l, \boldsymbol{\beta}) = \widehat{m}_{l,e}(u_l, \boldsymbol{\beta}) + \widehat{m}_{l,m}(u_l, \boldsymbol{\beta}), \tag{A.7}$$

where

$$\widehat{m}_{l,e}(u_l, \boldsymbol{\beta}) = \mathbf{B}_q(u_l)^{\mathrm{T}} \widehat{\boldsymbol{\lambda}}_{l,e}(\boldsymbol{\beta}) \text{ and } \widehat{m}_{l,m}(u_l, \boldsymbol{\beta}) = \mathbf{B}_q(u_l)^{\mathrm{T}} \widehat{\boldsymbol{\lambda}}_{l,m}(\boldsymbol{\beta}). \tag{A.8}$$

According to the result on page 149 of de Boor (2001), for m_l satisfying Condition (C2), there is a function $m_l^0(u_l) = \mathbf{B}_q(u_l)^{\mathrm{T}} \boldsymbol{\lambda}_l \in G_n$, such that

$$\sup_{u_l \in [0,1]} \left|m_l^0(u_l) - m_l(u_l)\right| = O\left(J_n^{-q}\right). \tag{A.9}$$

Let $\mathbb{B}_q(\mathbf{u}) = \begin{bmatrix} \mathbf{B}_q(u_1)^{\mathrm{T}} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \mathbf{B}_q(u_d)^{\mathrm{T}} \end{bmatrix}_{d \times J_n d}$, where $\mathbf{u} = (u_1, \ldots u_d)^{\mathrm{T}}$. Thus $\widehat{m}_{l,e}(u_l, \boldsymbol{\beta}^0) = \mathbf{1}_l^{\mathrm{T}} \mathbb{B}_q(\mathbf{u}) \widehat{\boldsymbol{\lambda}}_e(\boldsymbol{\beta}^0)$ and $\widehat{m}_{l,m}(u_l, \boldsymbol{\beta}^0) = \mathbf{1}_l^{\mathrm{T}} \mathbb{B}_q(\mathbf{u}) \widehat{\boldsymbol{\lambda}}_m(\boldsymbol{\beta}^0)$, where $\mathbf{1}_l$ is the $d \times 1$ vector with the l-th element as "1" and other elements as "0". Let $\boldsymbol{\lambda} = \left\{\boldsymbol{\lambda}_1^{\mathrm{T}}, \ldots, \boldsymbol{\lambda}_d{}^{\mathrm{T}}\right\}^{\mathrm{T}}$. By Berstein's inequality in Bosq (1998), it can be proved that $\left\|n^{-1}\mathbf{D}(\boldsymbol{\beta}^0)^{\mathrm{T}}\mathbf{1}_n\right\|_\infty = O_p\left(J_n^{-1}\right)$. Thus by (A.6) and (A.9), for every $u_l \in [0,1]$,

$$\begin{aligned}
&\left|\widehat{m}_{l,m}(u_l, \boldsymbol{\beta}^0) - m_l^0(u_l)\right| \\
=\ & \left|n^{-1}\mathbf{1}_l^{\mathrm{T}} \mathbb{B}_q(\mathbf{u}) \widehat{\mathbf{V}}(\boldsymbol{\beta}^0)^{-1}\mathbf{D}(\boldsymbol{\beta}^0)^{\mathrm{T}} \left\{\mathbf{m} - \mathbf{D}(\boldsymbol{\beta}^0)\boldsymbol{\lambda}\right\}\right| \\
\le\ & \left|\sum_{s=1}^{J_n} B_{s,q}(u_l)\right| \left\|\widehat{\mathbf{V}}(\boldsymbol{\beta})^{-1}\right\|_\infty \left\|n^{-1}\mathbf{D}(\boldsymbol{\beta}^0)^{\mathrm{T}}\mathbf{1}_n\right\|_\infty O\left(J_n^{-q}\right) \\
=\ & O_p(J_n) O_p\left(J_n^{-1}\right) O\left(J_n^{-q}\right) = O_p\left(J_n^{-q}\right).
\end{aligned} \tag{A.10}$$

Moreover, for every $u_l \in [0,1]$, by (A.1), (A.6) and Condition (C3), with probability approaching

1,

$$\begin{aligned}
& E\left\{\widehat{m}_{l,e}(u_l,\boldsymbol{\beta}^0)\,|\mathbf{X},\mathbf{Z}\right\}^2 \\
= \;& n^{-2}\mathbf{1}_l^{\mathrm{T}}\mathbb{B}_q(\mathbf{u})\widehat{\mathbf{V}}(\boldsymbol{\beta}^0)^{-1}\mathbf{D}(\boldsymbol{\beta}^0)^{\mathrm{T}}E\left(\mathbf{E}\mathbf{E}^{\mathrm{T}}\,|\mathbf{X},\mathbf{Z}\right)\mathbf{D}(\boldsymbol{\beta}^0)\widehat{\mathbf{V}}(\boldsymbol{\beta}^0)^{-1}\mathbb{B}_q(\mathbf{u})^{\mathrm{T}}\mathbf{1}_l \\
\leq \;& n^{-1}C_\sigma\mathbf{1}_l^{\mathrm{T}}\mathbb{B}_q(\mathbf{u})\widehat{\mathbf{V}}(\boldsymbol{\beta}^0)^{-1}\mathbb{B}_q(\mathbf{u})^{\mathrm{T}}\mathbf{1}_l \\
\leq \;& n^{-1}C_\sigma\left\|\mathbb{B}_q(\mathbf{u})^{\mathrm{T}}\mathbf{1}_l\right\|_2^2\left\|\widehat{\mathbf{V}}(\boldsymbol{\beta}^0)^{-1}\right\|_2 = O\left(J_n/n\right). \qquad (A.11)
\end{aligned}$$

Thus by the weak law of large numbers, for every $u_l \in [0,1]$, $\widehat{m}_{l,e}(u_l,\boldsymbol{\beta}^0) = O_p\left(J_n^{1/2}n^{-1/2}\right)$. Therefore, by (A.9), (A.11) and (A.10), $\left|\widehat{m}_l(u_l,\boldsymbol{\beta}^0) - m_l(u_l)\right| = O_p\left(J_n^{1/2}n^{-1/2} + J_n^{-q}\right)$, uniformly for every $u_l \in [0,1]$. Results in (i) of Proposition A.1 are proved. Similarly, $\widehat{\dot{m}}_l(u_l,\boldsymbol{\beta}^0)$ can be written as $\widehat{\dot{m}}_{l,e}(u_l,\boldsymbol{\beta}^0) + \widehat{\dot{m}}_{l,m}(u_l,\boldsymbol{\beta}^0)$, where $\widehat{\dot{m}}_{l,e}(u_l,\boldsymbol{\beta}^0) = \mathbf{B}_{q-1}(u_l)^{\mathrm{T}}\mathbf{D}_1\widehat{\boldsymbol{\lambda}}_{l,e}(\boldsymbol{\beta}^0)$ and $\widehat{\dot{m}}_{l,m}(u_l,\boldsymbol{\beta}^0) = \mathbf{B}_{q-1}(u_l)^{\mathrm{T}}\mathbf{D}_1\widehat{\boldsymbol{\lambda}}_{l,m}(\boldsymbol{\beta}^0)$. It is easy to prove that $\|\mathbf{D}_1\|_\infty = O(J_n)$, where $\mathbf{D}_1$ is defined in (9). Following the similar reasoning as the proof for $\widehat{m}_{l,m}(u_l,\boldsymbol{\beta}^0)$, one can proved that

$$\widehat{\dot{m}}_l(u_l,\boldsymbol{\beta}^0) - \dot{m}_l(u_l) = O_p\left(J_n^{3/2}n^{-1/2} + J_n^{-q+1}\right),$$

uniformly for every $u_l \in [0,1]$. Thus, results in (ii) of Proposition A.1 are proved. ∎

Define $\mathbb{P}_n(\mathbf{Z}_i) = D_i(\boldsymbol{\beta}^0)^{\mathrm{T}}\widehat{\delta}$, where

$$\widehat{\delta} = \arg\min_{\delta \in R^{dJ_n \times p}} \sum_{i=1}^n \left\{\mathbf{Z}_i - \delta^{\mathrm{T}}D_i(\boldsymbol{\beta}^0)\right\}^{\mathrm{T}}\left\{\mathbf{Z}_i - \delta^{\mathrm{T}}D_i(\boldsymbol{\beta}^0)\right\}.$$

Let $\mathbb{Z} = (\mathbf{Z}_1,\ldots,\mathbf{Z}_n)^{\mathrm{T}}$. Thus

$$\widehat{\delta} = \left\{\mathbf{D}(\boldsymbol{\beta}^0)^{\mathrm{T}}\mathbf{D}(\boldsymbol{\beta}^0)\right\}^{-1}\mathbf{D}(\boldsymbol{\beta}^0)^{\mathrm{T}}\mathbb{Z},$$

Lemma A.3. *Under Conditions (C1)-(C5), and $nN^{-4} \to \infty$ and $nN^{-2q-2} \to 0$, as $n \to \infty$,*
$\partial L_n\left(\boldsymbol{\beta}^0\right)/\partial\boldsymbol{\beta}_{-1} = -\sum_{i=1}^n\left\{Y_i - \sum_{l=1}^d m_l\left(\mathbf{Z}_i^T\boldsymbol{\beta}_l^0\right)X_{il}\right\}\left[\dot{m}_l(U_{il}(\boldsymbol{\beta}_l^0),\boldsymbol{\beta}^0)X_{il}\mathbf{J}_l^T\widetilde{\mathbf{Z}}_i\right]_{l=1}^d + o_p\left(n^{1/2}\right).$

Proof. For $\widehat{\boldsymbol{\lambda}}_e(\boldsymbol{\beta}^0)$ defined in (A.1), first we will show that for every $1 \leq l \leq d$,

$$\left\{\partial\widehat{\boldsymbol{\lambda}}_e(\boldsymbol{\beta}^0)^{\mathrm{T}}/\partial\boldsymbol{\beta}_{l,-1}\right\}D_i(\boldsymbol{\beta}^0) = -\dot{m}_l(U_{il}(\boldsymbol{\beta}_l^0),\boldsymbol{\beta}^0)X_{il}\mathbf{J}_l^{\mathrm{T}}\mathbb{P}_n(\mathbf{Z}_i) + O_p\left(J_n^{-q+1} + n^{-1/2}\right). \qquad (A.12)$$

Let $\Psi(\boldsymbol{\beta}^0) = \left[\widehat{\mathbf{V}}(\boldsymbol{\beta}^0)^{-1}\mathbf{D}(\boldsymbol{\beta}^0)^{\mathrm{T}}\right]_{J_nd\times n}$. Then

$$\begin{aligned}
D_i(\boldsymbol{\beta}^0)^{\mathrm{T}}\left\{\partial\widehat{\boldsymbol{\lambda}}_e(\boldsymbol{\beta}^0)/\partial\boldsymbol{\beta}_{l,-1}^{\mathrm{T}}\right\} = \;& n^{-1}D_i(\boldsymbol{\beta}^0)^{\mathrm{T}}\left\{\partial\Psi(\boldsymbol{\beta}^0)/\partial\boldsymbol{\beta}_{l,-1}^{\mathrm{T}}\right\}(\mathbf{Y}-\mathbf{m}) \\
& + D_i(\boldsymbol{\beta}^0)^{\mathrm{T}}\Psi(\boldsymbol{\beta}^0)\left\{\partial(\mathbf{Y}-\mathbf{m})/\partial\boldsymbol{\beta}_{l,-1}^{\mathrm{T}}\right\},
\end{aligned}$$

where $\partial\Psi(\boldsymbol{\beta}^0)/\partial\boldsymbol{\beta}_{l,-1}^{\mathrm{T}} = (\Psi_{i,sl,k})$ is $J_n d \times n \times (p-1)$ dimensional array and $\partial\left(\mathbf{Y}-\mathbf{m}\right)/\partial\boldsymbol{\beta}_{l,-1}^{\mathrm{T}}$is $n \times (p-1)$ dimensional matrix. By the weak law of large numbers, it can be proved that

$$\begin{aligned}&D_i(\boldsymbol{\beta}^0)^{\mathrm{T}}\left\{\partial\Psi(\boldsymbol{\beta}^0)/\partial\boldsymbol{\beta}_{l,-1}^{\mathrm{T}}\right\}(\mathbf{Y}-\mathbf{m})\\ =\;& n^{-1}\sum\nolimits_{i'=1}^{n}\sum\nolimits_{l=1}^{d}\sum\nolimits_{s=1}^{J_n} D_{i,sl}(\boldsymbol{\beta}_l^0)\left(\Psi_{i',sl,k}\right)_{k=2}^{p}\varepsilon_{i'} = O_p\left(n^{-1/2}\right).\end{aligned}$$

Let $\zeta_{il}' = \dot{m}_l\left(\mathbf{Z}_i^{\mathrm{T}}\boldsymbol{\beta}_l^0\right)X_{il}$ and $\mathbb{P}_n\left(\zeta_{il}'\right)$ be defined in the same way as $\mathbb{P}_n\left(\mathbf{Z}_i\right)$. Following similar reasoning as the proof for (A.10) and the fact that $\sup_{u_l\in[0,1]}\left|\dot{m}_l^0(u_l)-\dot{m}_l(u_l)\right| = O\left(J_n^{-q+1}\right)$, we have $\left|\zeta_{il}'-\mathbb{P}_n\left(\zeta_{il}'\right)\right| = O_p\left(J_n^{-q+1}\right)$. Thus

$$\begin{aligned}&D_i(\boldsymbol{\beta}^0)^{\mathrm{T}}\Psi(\boldsymbol{\beta}^0)\left\{\partial\left(\mathbf{Y}-\mathbf{m}\right)/\partial\boldsymbol{\beta}_{l,-1}^{\mathrm{T}}\right\} = -D_i(\boldsymbol{\beta}^0)^{\mathrm{T}}\Psi(\boldsymbol{\beta}^0)\left[\dot{m}_l(U_{il}(\boldsymbol{\beta}_l^0),\boldsymbol{\beta}^0)X_{il}\mathbf{Z}_i^{\mathrm{T}}\mathbf{J}_l\right]_{i=1}^{n}\\ =\;& -\mathbb{P}_n\left(\zeta_{li}'\right)D_i(\boldsymbol{\beta}^0)^{\mathrm{T}}\Psi(\boldsymbol{\beta}^0)\mathbb{Z}\mathbf{J}_l + O_p\left(J_n^{-q+1}\right) = -\dot{m}_l(U_{il}(\boldsymbol{\beta}_l^0),\boldsymbol{\beta}^0)X_{il}\mathbb{P}_n\left(\mathbf{Z}_i^{\mathrm{T}}\right)\mathbf{J}_l + O_p\left(J_n^{-q+1}\right).\end{aligned}$$

Therefore, (A.12) is proved by the above results. For $\widehat{\boldsymbol{\lambda}}_m(\boldsymbol{\beta}^0)$ defined in (A.1), by (A.3) and (A.10), with probability approaching 1,

$$\begin{aligned}\left\|\widehat{\boldsymbol{\lambda}}_m(\boldsymbol{\beta}^0)-\boldsymbol{\lambda}\right\|_\infty &= \left\|\left\{\mathbf{D}(\boldsymbol{\beta}^0)^{\mathrm{T}}\mathbf{D}(\boldsymbol{\beta}^0)\right\}^{-1}\mathbf{D}(\boldsymbol{\beta}^0)^{\mathrm{T}}\left\{\mathbf{m}-\mathbf{D}(\boldsymbol{\beta}^0)\boldsymbol{\lambda}\right\}\right\|_\infty \\ &\le c_V^{-1}J_n\left\|n^{-1}\mathbf{D}(\boldsymbol{\beta}^0)^{\mathrm{T}}\left\{\mathbf{m}-\mathbf{D}(\boldsymbol{\beta}^0)\boldsymbol{\lambda}\right\}\right\|_\infty \\ &\le c_V^{-1}J_n\sup\nolimits_{s,l}n^{-1}\sum\nolimits_{i=1}^{n}\left|B_{s,q}\left(U_{il}(\boldsymbol{\beta}_l^0)\right)X_{il}\right|O\left(J_n^{-q}\right) \\ &\asymp c_V^{-1}\sup\nolimits_{s,l}E\left|B_{s,q}\left(U_{il}(\boldsymbol{\beta}_l^0)\right)X_{il}\right|O\left(J_n^{-q+1}\right) \\ &\le O\left(J_n^{-1}\right)O\left(J_n^{-q+1}\right) = O\left(J_n^{-q}\right).\end{aligned} \tag{A.13}$$

By the decomposition of $\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}^0)$, (A.1), (A.12) and (A.13), one has

$$\left\{\partial\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}^0)^{\mathrm{T}}/\partial\boldsymbol{\beta}_{l,-1}\right\}D_i(\boldsymbol{\beta}^0) = -\dot{m}_l(U_{il}(\boldsymbol{\beta}_l^0),\boldsymbol{\beta}^0)X_{il}\mathbf{J}_l^{\mathrm{T}}\mathbb{P}_n\left(\mathbf{Z}_i\right) + O_p\left(J_n^{-q+1}+n^{-1/2}\right).$$

By result (ii) in Proposition A.1,

$$\widehat{\dot{m}}_l(U_{il}(\boldsymbol{\beta}_l^0),\boldsymbol{\beta}^0)X_{il}\mathbf{J}_l^{\mathrm{T}}\mathbf{Z}_i = \dot{m}_l(U_{il}(\boldsymbol{\beta}_l^0),\boldsymbol{\beta}^0)X_{il}\mathbf{J}_l^{\mathrm{T}}\mathbf{Z}_i + O_p\left(J_n^{3/2}n^{-1/2}+J_n^{-q+1}\right).$$

By Condition (C5), it can be proved that $\left\|\mathbb{P}_n\left(\mathbf{Z}_i\right)-\mathbb{P}\left(\mathbf{Z}_i\right)\right\|_\infty = O_p\left((J_n/n)^{1/2}+J_n^{-1}\right)$. For notation simplicity, we let $\widehat{\dot{m}}_{il} = \widehat{\dot{m}}_l(U_{il}(\boldsymbol{\beta}_l^0),\boldsymbol{\beta}^0)$ and $\dot{m}_{il} = \dot{m}_l(U_{il}(\boldsymbol{\beta}_l^0),\boldsymbol{\beta}^0)$. Thus

$$\widehat{\dot{m}}_{il}X_{il}\mathbf{J}_l^{\mathrm{T}}\mathbf{Z}_i + \left\{\partial\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}^0)^{\mathrm{T}}/\partial\boldsymbol{\beta}_{l,-1}\right\}D_i(\boldsymbol{\beta}^0) = \dot{m}_{il}X_{il}\mathbf{J}_l^{\mathrm{T}}\widetilde{\mathbf{Z}}_i + O_p\left(J_n^{3/2}n^{-1/2}+J_n^{-1}\right).$$

Let 1_s be the $s \times 1$-dimensional vector with "1"'s as its elements, and let $m_i = m\left(\mathbf{Z}_i, \mathbf{X}_i, \boldsymbol{\beta}^0\right)$. Hence, by (10) and the above result, one has

$$
\begin{aligned}
\partial L_n(\boldsymbol{\beta}^0)/\partial\boldsymbol{\beta}_{-1} &= -\sum\nolimits_{i=1}^n \left\{Y_i - D_i(\boldsymbol{\beta}^0)^{\mathrm{T}}\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}^0)\right\}\left[\dot{m}_{il}X_{il}\mathbf{J}_l^{\mathrm{T}}\widetilde{\mathbf{Z}}_i + O_p\left(J_n^{3/2}n^{-1/2} + J_n^{-1}\right)\times 1_{p-1}\right]_{l=1}^d \\
&= -\sum\nolimits_{i=1}^n \left\{Y_i - m_i + m_i - D_i(\boldsymbol{\beta}^0)^{\mathrm{T}}\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}^0)\right\}\left[\dot{m}_{li}X_{il}\mathbf{J}_l^{\mathrm{T}}\widetilde{\mathbf{Z}}_i + O_p\left(J_n^{3/2}n^{-1/2} + J_n^{-1}\right)\times 1_{p-1}\right]_{l=1}^d \\
&= -\sum\nolimits_{i=1}^n \left\{Y_i - m_i\right\}\left[\dot{m}_{li}X_{il}\mathbf{J}_l^{\mathrm{T}}\widetilde{\mathbf{Z}}_i\right]_{l=1}^d - \left(\mathbf{I}_1 + \mathbf{I}_2 + \mathbf{I}_3\right),
\end{aligned}
$$

where

$$
\begin{aligned}
\mathbf{I}_1 &= \sum\nolimits_{i=1}^n \left\{m_i - D_i(\boldsymbol{\beta}^0)^{\mathrm{T}}\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}^0)\right\}\left[\dot{m}_{li}X_{il}\mathbf{J}_l^{\mathrm{T}}\widetilde{\mathbf{Z}}_i\right]_{l=1}^d, \\
\mathbf{I}_2 &= \sum\nolimits_{i=1}^n \left\{Y_i - m_i\right\} O_p\left(J_n^{3/2}n^{-1/2} + J_n^{-1}\right)\times 1_{d(p-1)}, \\
\mathbf{I}_3 &= \sum\nolimits_{i=1}^n \left\{m_i - D_i(\boldsymbol{\beta}^0)^{\mathrm{T}}\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}^0)\right\} O_p\left(J_n^{3/2}n^{-1/2} + J_n^{-1}\right)\times 1_{d(p-1)}.
\end{aligned}
$$

We will prove that $\|\mathbf{I}_j\|_\infty = o_p\left(n^{1/2}\right)$ for each $j = 1, 2, 3$. By Lemmas A.1 and A.2,

$$
\sum\nolimits_{i=1}^n \left\{m_i - D_i(\boldsymbol{\beta}^0)^{\mathrm{T}}\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}^0)\right\} = \Pi_1 + \Pi_2 + \Pi_3,
$$

where $\Pi_1 = 1_n^{\mathrm{T}}\left(\mathbf{m} - \mathbf{D}(\boldsymbol{\beta}^0)\boldsymbol{\lambda}\right) = O\left(nJ_n^{-q}\right)$, $D_i(\boldsymbol{\beta}^0) = \left(D_{i,sl}(\boldsymbol{\beta}_l^0), 1 \le s \le J_n, 1 \le l \le d\right)^{\mathrm{T}}$ with $D_{i,sl}(\boldsymbol{\beta}_l^0) = B_{s,q}\left(U_{il}(\boldsymbol{\beta}_l^0)\right)X_{il}$ and $\mathbf{D}(\boldsymbol{\beta}^0) = \left[\left\{D_1(\boldsymbol{\beta}^0), \ldots, D_n(\boldsymbol{\beta}^0)\right\}^{\mathrm{T}}\right]_{n\times J_n d}$.. With probability approaching 1,

$$
\begin{aligned}
\left\|\mathbf{D}(\boldsymbol{\beta}^0)^{\mathrm{T}}1_n\right\|_2^2 &= \sum\nolimits_{l=1}^d \sum\nolimits_{s=1}^{J_n}\left\{\sum\nolimits_{i=1}^n B_{s,q}\left(U_{il}(\boldsymbol{\beta}_l^0)\right)X_{il}\right\}^2 \\
&\asymp n^2\sum\nolimits_{l=1}^d \sum\nolimits_{s=1}^{J_n}\left[E\left\{B_{s,q}\left(U_{il}(\boldsymbol{\beta}_l^0)\right)X_{il}\right\}\right]^2 \asymp n^2J_n^{-1}.
\end{aligned}
$$

By (A.6) and (A.9),

$$
\begin{aligned}
|\Pi_2| &= \left|1_n^{\mathrm{T}}\mathbf{D}(\boldsymbol{\beta}^0)\left\{\mathbf{D}(\boldsymbol{\beta}^0)^{\mathrm{T}}\mathbf{D}(\boldsymbol{\beta}^0)\right\}^{-1}\mathbf{D}(\boldsymbol{\beta}^0)^{\mathrm{T}}\left(\mathbf{m} - \mathbf{D}(\boldsymbol{\beta}^{\mathbf{0}})\boldsymbol{\lambda}\right)\right| \\
&\le \left\|\mathbf{D}(\boldsymbol{\beta}^0)^{\mathrm{T}}1_n\right\|_2^2\left\|\widehat{\mathbf{V}}_n\left(\boldsymbol{\beta}^0\right)^{-1}\right\|_2 O\left(n^{-1}J_n^{-q}\right) \\
&= O_p\left(n^2J_n^{-1}\right)O_p\left(J_n\right)O\left(n^{-1}J_n^{-q}\right) = O_p\left(nJ_n^{-q}\right),
\end{aligned}
$$

$$
\begin{aligned}
|\Pi_3| &= \left|1_n^{\mathrm{T}}\mathbf{D}(\boldsymbol{\beta}^0)\left\{\mathbf{D}(\boldsymbol{\beta}^0)^{\mathrm{T}}\mathbf{D}(\boldsymbol{\beta}^0)\right\}^{-1}\mathbf{D}(\boldsymbol{\beta}^0)^{\mathrm{T}}\mathbf{E}\right| \\
&\le \left\|\mathbf{D}(\boldsymbol{\beta}^0)^{\mathrm{T}}1_n\right\|_2\left\|\widehat{\mathbf{V}}_n\left(\boldsymbol{\beta}^0\right)^{-1}\right\|_2\left\|n^{-1}\mathbf{D}(\boldsymbol{\beta}^0)^{\mathrm{T}}\mathbf{E}\right\|_2 \\
&= O_p\left(nJ_n^{-1/2}\right)O_p\left(J_n\right)O_p\left(n^{-1/2}\right) = O_p\left(n^{1/2}J_n^{1/2}\right).
\end{aligned}
$$

Thus $\sum_{i=1}^{n}\left\{m_i - D_i(\boldsymbol{\beta}^0)^{\mathrm{T}}\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}^0)\right\} = O_p\left(n^{1/2}J_n^{1/2} + nJ_n^{-q}\right)$. By Proposition A.1, one has

$$\left|m\left(\mathbf{Z}_i, \mathbf{X}_i\right) - D_i(\boldsymbol{\beta}^0)^{\mathrm{T}}\widehat{\boldsymbol{\lambda}}(\boldsymbol{\beta}^0)\right| = O_p\left(J_n^{1/2}n^{-1/2} + J_n^{-q}\right).$$

Therefore,

$$\begin{aligned}
\|\mathbf{I}_1\|_\infty &= O_p\left(J_n^{1/2}n^{-1/2} + J_n^{-q}\right)O_p\left(n^{1/2}\right) = O_p\left(J_n^{1/2} + J_n^{-q}n^{1/2}\right) = o_p\left(n^{1/2}\right),\\
\|\mathbf{I}_2\|_\infty &= O_p\left(J_n^{3/2}n^{-1/2} + J_n^{-1}\right)O_p\left(n^{1/2}\right) = o_p\left(n^{1/2}\right),\\
\|\mathbf{I}_3\|_\infty &= O_p\left(n^{1/2}J_n^{1/2} + nJ_n^{-q}\right)O_p\left(J_n^{3/2}n^{-1/2} + J_n^{-1}\right) = o_p\left(n^{1/2}\right).
\end{aligned}$$

Thus, Lemma A.3 is proved. ∎

Proof of Theorem 1. Under the conditions of Theorem 1, we follow similar arguments as presented in Ichimura (1993) to show that $\widehat{\boldsymbol{\beta}}_{-1}$ is a root-n consistent estimator of $\boldsymbol{\beta}^0_{-1}$, and thus the proof is omitted. By Lemma A.3, it is straightforward to prove that

$$\partial L_n(\boldsymbol{\beta}^0)/\partial\boldsymbol{\beta}_{-1}\partial\boldsymbol{\beta}_{-1}^{\mathrm{T}} = \sum\nolimits_{i=1}^{n}\left[\left[\dot{m}_l(U_{il}(\boldsymbol{\beta}_l^0), \boldsymbol{\beta}^0)X_{il}\mathbf{J}_l^{\mathrm{T}}\widetilde{\mathbf{Z}}_i\right]_{l=1}^{d}\right]^{\otimes 2} + o_p\left(n\right).$$

By Taylor expansion, Lemma A.3 and the above result,

$$\begin{aligned}
\widehat{\boldsymbol{\beta}}_{-1} - \boldsymbol{\beta}^0_{-1} &= -\left\{\partial L_n(\boldsymbol{\beta}^0)/\partial\boldsymbol{\beta}_{-1}\partial\boldsymbol{\beta}_{-1}^{\mathrm{T}}\right\}^{-1}\left\{\partial L_n(\boldsymbol{\beta}^0)/\partial\boldsymbol{\beta}_{-1}\right\}\left\{1 + o_p\left(1\right)\right\}\\
&= \left[E\left[\left\{\dot{m}_l(U_{il}(\boldsymbol{\beta}_l^0), \boldsymbol{\beta}^0)X_{il}\mathbf{J}_l^{\mathrm{T}}\widetilde{\mathbf{Z}}_i\right\}_{l=1}^{d}\right]^{\otimes 2}\right]^{-1}\times\\
&\qquad n^{-1}\sum\nolimits_{i=1}^{n}\varepsilon_i\left[\dot{m}_l(U_{il}(\boldsymbol{\beta}_l^0), \boldsymbol{\beta}^0)X_{il}\mathbf{J}_l^{\mathrm{T}}\widetilde{\mathbf{Z}}_i\right]_{l=1}^{d} + o_p\left(n^{-1/2}\right).
\end{aligned}$$

Theorem 1 can be proved by Lindeberg-Feller Central Limit Theorem. ∎

Proof of Theorem 2. Since $\left\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0\right\|_2 = O_p\left(n^{-1/2}\right)$, Theorem 2 follows from this result and Proposition A.1. ∎

A.3 Proofs of Theorems 3 and 4

Following the same techniques employed in Fan and Zhang (2008), it can be proved that the oracle estimator $\widetilde{m}_{\mathrm{LL},1}\left(u_1, \boldsymbol{\beta}^0\right)$ has the asymptotic distribution and convergence rate given in Theorem 3. The detailed proof is thus omitted. Since $\left\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0\right\|_2 = O_p\left(n^{-1/2}\right)$, Theorem 3 is proved by Slutsky's theorem. We will focus on the proof of Theorem 4.

According to (18) and (A.7),

$$\begin{aligned} & \widehat{m}_{\mathrm{SBLL},1}\left(u_1, \boldsymbol{\beta}^0\right) - \widetilde{m}_{\mathrm{LL},1}\left(u_1, \boldsymbol{\beta}^0\right) \\ = & -(1,0)\left\{\mathbf{C}\left(u_1, \boldsymbol{\beta}_1^0\right)^{\mathrm{T}} \mathbf{W}\left(u_1, \boldsymbol{\beta}_1^0\right) \mathbf{C}\left(u_1, \boldsymbol{\beta}_1^0\right)\right\}^{-1} \mathbf{C}\left(u_1, \boldsymbol{\beta}_1^0\right)^{\mathrm{T}} \mathbf{W}\left(u_1, \boldsymbol{\beta}_1^0\right) \times \\ & \left[\sum_{l=2}^{d}\left\{\widehat{m}_l(U_{il}\left(\boldsymbol{\beta}^0\right), \boldsymbol{\beta}^0) - m_l\left(U_{il}\right)\right\} X_{il}\right]_{i=1}^{n} \\ = & -(1,0)\left\{n^{-1}\mathbf{C}\left(u_1, \boldsymbol{\beta}_1^0\right)^{\mathrm{T}} \mathbf{W}\left(u_1, \boldsymbol{\beta}_1^0\right) \mathbf{C}\left(u_1, \boldsymbol{\beta}_1^0\right)\right\}^{-1}\left\{\left(\begin{array}{c} \Psi_{v1}\left(u_1, \boldsymbol{\beta}^0\right) \\ \Psi_{v2}\left(u_1, \boldsymbol{\beta}^0\right) \end{array}\right) + \left(\begin{array}{c} \Psi_{b1}\left(u_1, \boldsymbol{\beta}^0\right) \\ \Psi_{b2}\left(u_1, \boldsymbol{\beta}^0\right) \end{array}\right)\right\}, \end{aligned}$$

where

$$\begin{aligned} \Psi_{v1}\left(u_1, \boldsymbol{\beta}^0\right) &= n^{-1}\sum_{i=1}^{n}\sum_{l=2}^{d} X_{i1}X_{il}K_{h_1}\left(U_{i1}(\boldsymbol{\beta}_1^0) - u_1\right)\widehat{m}_{l,\varepsilon}(U_{il}, \boldsymbol{\beta}^0), \\ \Psi_{v2}\left(u_1, \boldsymbol{\beta}^0\right) &= n^{-1}\sum_{i=1}^{n}\sum_{l=2}^{d}\left\{\left(U_{i1}(\boldsymbol{\beta}_1^0) - u_1\right)/h_1\right\} X_{i1}X_{il}K_{h_1}\left(U_{i1}(\boldsymbol{\beta}_1^0) - u_1\right)\widehat{m}_{l,\varepsilon}(U_{il}, \boldsymbol{\beta}^0), \end{aligned}$$

$$\begin{aligned} \Psi_{b1}\left(u_1, \boldsymbol{\beta}^0\right) &= n^{-1}\sum_{i=1}^{n}\sum_{l=2}^{d} X_{i1}X_{il}K_{h_1}\left(U_{i1}(\boldsymbol{\beta}_1^0) - u_1\right)\left\{\widehat{m}_{l,m}(U_{il}, \boldsymbol{\beta}^0) - m_l\left(U_{il}\right)\right\}, \\ \Psi_{b2}\left(u_1, \boldsymbol{\beta}^0\right) &= n^{-1}\sum_{i=1}^{n}\sum_{l=2}^{d}\left\{\left(U_{i1}(\boldsymbol{\beta}_1^0) - u_1\right)/h_1\right\} X_{i1}X_{il} \times \\ & \quad K_{h_1}\left(U_{i1}(\boldsymbol{\beta}_1^0) - u_1\right)\left\{\widehat{m}_{l,m}(U_{il}, \boldsymbol{\beta}^0) - m_l\left(U_{il}\right)\right\}. \end{aligned}$$

Lemma A.4. *Under Conditions (C1), (C3), (C4) and (C6), and $N \to \infty$ and $nN^{-1} \to \infty$, as $n \to \infty$, one has* $\sup_{u_1 \in [0,1]}\left|\Psi_{v1}\left(u_1, \boldsymbol{\beta}^0\right)\right| + \sup_{u_1 \in [0,1]}\left|\Psi_{v2}\left(u_1, \boldsymbol{\beta}^0\right)\right| = O_p\left(n^{-1/2}\right)$.

Proof. Let

$$\xi_{sl} = n^{-1}\sum_{i=1}^{n} X_{i1}X_{il}K_{h_1}\left(U_{i1}(\boldsymbol{\beta}_1^0) - u_1\right)B_{s,q}\left(U_{il}(\boldsymbol{\beta}_l^0)\right),$$

and $\xi\left(u_1\right) = \left\{\left(\xi_1\left(u_1\right)^{\mathrm{T}}, \ldots, \xi_d\left(u_1\right)^{\mathrm{T}}\right)^{\mathrm{T}}\right\}_{dJ_n \times 1}$ with $\xi_l\left(u_1\right) = \left\{\xi_{s,l}\left(u_1\right) : 1 \le s \le J_n\right\}^{\mathrm{T}}$. Then for every $u_1 \in [0,1]$, $E\left\{\xi_{sl}\left(u_1\right)\right\} \asymp J_n^{-1}$. It can be proved by Bernstein's inequality in de Boor (2001) that $\sup_{u_1 \in [0,1]}\sup_{1 \le l \le d, 1 \le s \le J_n}\left|\xi_{sl}\left(u_1\right) - E\left\{\xi_{sl}\left(u_1\right)\right\}\right| = O_p\left(J_n^{-1/2}n^{-1/2}\right)$, and thus for $J_n n^{-1} = o(1)$, $\sup_{u_1 \in [0,1]}\left\|\xi\left(u_1\right)\right\|_2 = O_p\left(J_n^{-1/2}\right)$. By (A.8),

$$\Psi_{v1}\left(u_1, \boldsymbol{\beta}^0\right) = \sum_{l=2}^{d}\sum_{s=1}^{J_n}\xi_{sl}\widehat{\lambda}_{s,l,e}(\boldsymbol{\beta}^0) = \xi^{\mathrm{T}}\widehat{\boldsymbol{\lambda}}_e(\boldsymbol{\beta}^0) - \xi_l^{\mathrm{T}}\widehat{\boldsymbol{\lambda}}_{l,e}(\boldsymbol{\beta}^0).$$

Thus $E\left\{\xi\left(u_1\right)^{\mathrm{T}}\widehat{\boldsymbol{\lambda}}_e(\boldsymbol{\beta}^0)\right\} = 0$ and with probability approaching 1,

$$\begin{aligned} & \sup_{u_1 \in [0,1]} E\left\{\xi\left(u_1\right)^{\mathrm{T}}\widehat{\boldsymbol{\lambda}}_e(\boldsymbol{\beta}^0)\left|\mathbf{X}, \mathbf{Z}\right.\right\}^2 \\ = & \sup_{u_1 \in [0,1]} n^{-2}\xi\left(u_1\right)^{\mathrm{T}}\widehat{\mathbf{V}}(\boldsymbol{\beta}^0)^{-1}\mathbf{D}(\boldsymbol{\beta}^0)^{\mathrm{T}}E\left(\mathbf{E}\mathbf{E}^{\mathrm{T}}\left|\mathbf{X}, \mathbf{Z}\right.\right)\mathbf{D}(\boldsymbol{\beta}^0)\widehat{\mathbf{V}}(\boldsymbol{\beta}^0)^{-1}\xi\left(u_1\right) \\ \le & \sup_{u_1 \in [0,1]} n^{-1}C_\sigma\left\|\xi\left(u_1\right)\right\|_2^2\left\|\widehat{\mathbf{V}}(\boldsymbol{\beta}^0)^{-1}\right\|_2 = O\left(n^{-1}\right). \end{aligned}$$

Therefore, by the weak law of large numbers, $\sup_{u_1\in[0,1]}\left|\xi\left(u_1\right)^{\mathrm{T}}\widehat{\boldsymbol{\lambda}}_e(\boldsymbol{\beta}^0)\right| = O_p\left(n^{-1/2}\right)$. Similarly, we can prove that $\sup_{u_1\in[0,1]}\left|\xi_l\left(u_1\right)^{\mathrm{T}}\widehat{\boldsymbol{\lambda}}_{l,e}(\boldsymbol{\beta}^0)\right| = O_p\left(n^{-1/2}\right)$. Thus $\sup_{u_1\in[0,1]}\left|\Psi_{v1}\left(u_1,\boldsymbol{\beta}^0\right)\right| = O_p\left(n^{-1/2}\right)$. Since $\left|\left(U_{i1}(\boldsymbol{\beta}_1^0)-u_1\right)/h_1\right| \leq 1$, following the same reasoning, it can be proved that $\sup_{u_1\in[0,1]}\left|\Psi_{v1}\left(u_1,\boldsymbol{\beta}^0\right)\right| = O_p\left(n^{-1/2}\right)$. ∎

Lemma A.5**.** *Under Conditions (C1), (C4) and (C6), and* $N\to\infty$*, as* $n\to\infty$*, one has* $\sup_{u_1\in[0,1]}\left|\Psi_{b1}\left(u_1,\boldsymbol{\beta}^0\right)\right| + \sup_{u_1\in[0,1]}\left|\Psi_{b2}\left(u_1,\boldsymbol{\beta}^0\right)\right| = O_p\left(J_n^{-q}\right)$.

Proof. By (A.9) and (A.10), $\left|\widehat{m}_{l,m}(U_{il},\boldsymbol{\beta}^0) - m_l\left(U_{il}\right)\right| = O_p\left(J_n^{-q}\right)$, $E\left\{X_{i1}X_{il}K_{h_1}\left(U_{i1}(\boldsymbol{\beta}_1^0)-u_1\right)\right\} \asymp 1$. It can be proved by Bernstein's inequality in de Boor (2001) that

$$\sup_{u_1\in[0,1]}\left|n^{-1}\sum\nolimits_{i=1}^{n}\sum\nolimits_{l=2}^{d}X_{i1}X_{il}K_{h_1}\left(U_{i1}(\boldsymbol{\beta}_1^0)-u_1\right)\right| = O_p\left(1\right).$$

Thus $\sup_{u_1\in[0,1]}\left|\Psi_{b1}\left(u_1,\boldsymbol{\beta}^0\right)\right| = O_p\left(J_n^{-q}\right)$. Similarly, one has $\sup_{u_1\in[0,1]}\left|\Psi_{b2}\left(u_1,\boldsymbol{\beta}^0\right)\right| = O_p\left(J_n^{-q}\right)$. ∎

Proof of Theorem 4. It is straightforward to prove that

$$\sup_{u_1\in[0,1]}\left\|\left\{n^{-1}\mathbf{C}\left(u_1,\boldsymbol{\beta}_1^0\right)^{\mathrm{T}}\mathbf{W}\left(u_1,\boldsymbol{\beta}_1^0\right)\mathbf{C}\left(u_1,\boldsymbol{\beta}_1^0\right)\right\}^{-1}\right\|_2 \leq C$$

for some constants $0<C<\infty$. Thus by Lemmas A.5 and A.4, one has

$$\sup_{u_1\in[0,1]}\left|\widehat{m}_{\mathrm{SBLL},1}\left(u_1,\boldsymbol{\beta}^0\right) - \widetilde{m}_{\mathrm{LL},1}\left(u_1,\boldsymbol{\beta}^0\right)\right| = O_p\left(n^{-1/2}+J_n^{-q}\right).$$

Since $\left\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^0\right\|_2 = O_p\left(n^{-1/2}\right)$, $\sup_{u_1\in[0,1]}\left|\widehat{m}_{\mathrm{SBLL},1}\left(u_1,\widehat{\boldsymbol{\beta}}\right) - \widetilde{m}_{\mathrm{LL},1}\left(u_1,\widehat{\boldsymbol{\beta}}\right)\right| = O_p\left(n^{-1/2}+J_n^{-q}\right)$. Therefore, under the assumption that $nN^{-5q/2}=o\left(1\right)$ and $n^{-1}N=o\left(1\right)$, Theorem 4 is proved. ∎

References

Bailey, C. (1994). *Smart Exercise: Burning Fat, Getting Fit.* Houghton-Mifflin Co., Boston.

Behnke, A. R. and Wilmore, J. H. (1974). *Evaluation and Regulation of Body Build and Composition.* Prentice-Hall, Englewood Cliffs, N. J.

Bellman, R.E. (1961). *Adaptive Control Processes.* Princeton University Press, Princeton, NJ.

Bosq, D. (1998). *Nonparametric Statistics for Stochastic Processes.* Springer-Verlag, New York.

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.

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

Cui, X., Härdle, W., and Zhu, L. (2011). The EFM approach for single-index models. *Annals of Statistics,* 39, 1658-1688.

de Boor, C. (2001). *A Practical Guide to Splines.* Springer, New York.

DeVore, R. A. and Lorentz, G. G. (1993). *Constructive Approximation.* Berlin: Springer-Verlag.

Demko, S. (1986). Spectral bounds for $\left|a^{-1}\right|_{\infty}$. *Journal of Approximation Theory*, 48, 207-212.

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and its Applications.* Chapman and Hall, London.

Fan, J. and Jiang, J. (2007). Nonparametric inference with generalized likelihood ratio tests. *Test*, 16, 409-444.

Fan, J., Härdle, W., and Mammen, E. (1998). Direct estimation of low-dimensional components in additive models. *The Annals of Statistics,* 26, 943-971.

Fan, J., Zhang, C., and Zhang, J. (2001). Generalized Likelihood Ratio Statistics and Wilks Phenomenon. *The Annals of Statistics,* 29, 153-193.

Fan, J. and Zhang, W. (2008). Statistical methods with varying coefficient models. *Statistics and its interface,* 1, 179-195.

Hastie, T. and Tibshirani, R. (1990). *Generalized additive models.* Chapman and Hall, London.

Hastie, T. J. and Tibishirani, R. J. (1993). Varying-coefficient models. *Journal of the Royal Statistical Society, Series B*, 55, 757-796.

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

Horowitz, J., Klemelä, J., and Mammen, E. (2006). Optimal estimation in additive regression. *Bernoulli,* 12, 271-298.

Horowitz, J. and Mammen, E. (2004). Nonparametric estimation of an additive model with a link function. *Annals of Statistics.* 32, 2412-2443.

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.

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

Jackson, A. S., Stanforth, P. R., Gagnon, J., Rankinen, T, Leon, A. S., Rao, D. C., Skinner, J. S., Bouchard, C. and Wilmore, J. H. (2002). The effect of sex, age and race on estimating percentage body fat from body mass index: The Heritage Family Study. *International Journal of Obesity,* 26, 789-796.

Jackson, A. S. and Pollock, M. L. (1978). Generalized equations for predicting body density. *British Journal of Nutrition*, 40, 497-504.

Johnson, R. W. (1996). Fitting percentage of body fat to simple body measurements. *Journal of Statistics Education*, v.4, n.1.

Katch, F. and McArdle, W. (1977). *Nutrition, Weight Control, and Exercise.* Houghton Mifflin Co., Boston.

Liang, H., Liu, X., Li, R. and Tsai, C. L. (2010). Estimation and testing for partially linear single-index models. *The Annals of Statistics*, 38, 3811-3836.

Lin, H., Song, X.-K. P., and Zhou, Q. M. (2007). Varying-coefficient marginal models and applications in longitudinal data analysis. *Sankhya,* 69, 581-614.

Linton, O. B. (1997). Efficient estimation of additive nonparametric regression models. *Biometrika,* 84, 469-473.

Linton, O. B. and Nielsen, J. P. (1995). A kernel method of estimating structured nonparametric regression based on marginal integration. *Biometrika,* 82, 93-100.

Lu, X., Chen, G., Singh, R. and Song, X.-K. P. (2006). A class of partially linear single-index survival models. *Canadian Journal of Statistics,* 34, 97-112.

Ma, S. and Yang, L. (2011). Spline-backfitted kernel smoothing of partially linear additive model. *Journal of Statistical Planning and Inference,* 141, 204-219.

Ma, S., Yang, L., Romero, R., and Cui, Y. (2011). Varying-coefficient models for gene-environment interaction: a non-linear look. *Bioinformatics*, 27, 2119-2126.

Mammen, E., Linton, O., and Nielsen, J. (1999). The existence and asymptotic properties of a back tting projection algorithm under weak conditions. *The Annals of Statistics*, 27 1443-1490.

Opsomer, J. D. and Ruppert, D. (1997). Fitting a bivariate additive model by local polynomial regression. *The Annals of Statistics*, 25 186-211.

Shake, C. L., Schlichting, C., Mooney, L. W., Callahan, A. B., and Cohen, M. E. (1993). Predicting percent body fat from circumference measurements. *Military Medicine*, 158, 26-31.

Stone, C. (1985). Additive regression and other nonparametric models. *The Annals of Statistics*, 13, 689-705.

Stute, W. and Zhu, L. X. (2005) Nonparametric checks for single-index models. *The Annals of Statistics,* 33, 1048-1083.

Wang, L. and Yang, L. (2007). Spline-backfitted kernel smoothing of nonlinear additive autoregression model. *The Annals of Statistics,* 35, 2474-2503.

Wang, L., Liu X., Liang, H., and Carroll, R. J. (2011). Estimation and variable selection for generalized additive partial linear models. *The Annals of Statistics*, 39, 1827-1851

Wilmore, J. (1976). *Athletic Training and Physical Fitness: Physiological Principles of the Conditioning Process.* Allyn and Bacon, Inc., Boston.

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

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

Xia, Y., Tong, H. and Li, W. K. (1999). On extended partially linear single-index models. *Biometrika,* 86, 831-842.

Xue, L. and Liang, H. (2010). Polynomial spline estimation for the generalized additive coefficient model. *Scandinavian Journal of Statistics,* 37, 26-46.

Xue, L. and Yang, L. (2006). Additive coefficient modeling via polynomial spline. *Statistica Sinica*, 16, 1423-1446.

Yuan, M. (2011). On the identifiability of additive index models. *Statistica Sinica*, 21, 1901-1911

Yu, Y. and Ruppert, D. (2002). Penalized spline estimation for partially linear single-index models. *Journal of the American Statistical Association,* 97, 1042-1054.

Zamboni, M., Armellini, F., Harris, T., Turcato, E., Micciolo, R., Bergamo-Andreis, I. A., and Bosello, O. (1997). Effects of age on body fat distribution and cardiovascular risk factors in women. *The American Journal of Clinical Nutrition,* 66, 111-5.

Zhou, S., Shen, X. and Wolfe, D. A. (1998). Local asymptotics for regression splines and confidence regions. *The Annals of Statistics*, 26, 1760-1782.

Table 1: The empirical coverage rates of the 95% confidence intervals for $\beta_1 = (\beta_{11}, \beta_{12}, \beta_{13})^{\mathrm{T}}$, $\beta_2 = (\beta_{21}, \beta_{22}, \beta_{23})^{\mathrm{T}}$ and $\beta_3 = (\beta_{31}, \beta_{32}, \beta_{33})^{\mathrm{T}}$ for $n = 200, 500, 1000$.

n	β_{11}	β_{12}	β_{13}	β_{21}	β_{22}	β_{23}	β_{31}	β_{32}	β_{33}
200	0.908	0.918	0.940	0.908	0.912	0.914	0.938	0.950	0.912
500	0.956	0.930	0.954	0.934	0.926	0.932	0.934	0.952	0.934
1000	0.950	0.946	0.946	0.956	0.956	0.950	0.940	0.946	0.942

Table 2: The average bias $(\times 10^{-2})$ of the estimators for $\beta_1 = (\beta_{11}, \beta_{12}, \beta_{13})^{\mathrm{T}}$, $\beta_2 = (\beta_{21}, \beta_{22}, \beta_{23})^{\mathrm{T}}$ and $\beta_3 = (\beta_{31}, \beta_{32}, \beta_{33})^{\mathrm{T}}$ for $n = 200, 500, 1000$.

n	β_{11}	β_{12}	β_{13}	β_{21}	β_{22}	β_{23}	β_{31}	β_{32}	β_{33}
200	−0.3932	0.0942	−0.0422	−0.0896	−0.1393	0.0587	−0.0220	−0.0023	−0.1031
500	−0.1683	0.0095	0.0248	−0.0728	0.0396	0.0024	−0.1286	0.0663	0.0137
1000	0.0439	0.0327	−0.0796	0.0226	−0.0525	−0.0205	−0.0003	−0.0140	0.0218

Table 3: The average asymptotic standard error (ASE) $\left(\times 10^{-2}\right)$ and empirical standard error (ESE) $\left(\times 10^{-2}\right)$ of the estimators for $\beta_1 = (\beta_{11}, \beta_{12}, \beta_{13})^{\mathrm{T}}$, $\beta_2 = (\beta_{21}, \beta_{22}, \beta_{23})^{\mathrm{T}}$ and $\beta_3 = (\beta_{31}, \beta_{32}, \beta_{33})^{\mathrm{T}}$ for $n = 200, 500, 1000$.

n		β_{11}	β_{12}	β_{13}	β_{21}	β_{22}	β_{23}	β_{31}	β_{32}	β_{33}
200	ASE	3.4309	4.1912	2.4164	2.0593	2.8174	2.8519	1.5215	1.0825	1.7550
	ESE	3.8157	4.7051	2.6441	2.3449	3.1650	3.2594	1.5801	1.1116	1.9398
500	ASE	2.0555	2.5270	1.4322	1.1691	1.6041	1.5767	0.8737	0.6179	1.0033
	ESE	2.0411	2.7095	1.4118	1.2757	1.7195	1.6535	0.8743	0.6254	1.0983
1000	ASE	1.4330	1.7724	1.0008	0.8053	1.1102	1.0838	0.6035	0.4280	0.6943
	ESE	1.4844	1.7461	1.0367	0.8019	1.0966	1.1122	0.6288	0.4593	0.7113

Table 4: The MISE values for the two-step SBLL estimator $\widehat{m}_{\mathrm{SBLL},l}$ and the oracle estimators $\widetilde{m}_{\mathrm{LL},l}(\cdot)$ for$1 \leq l \leq 3$.

n	$\widehat{m}_{\mathrm{SBLL},1}$	$\widetilde{m}_{\mathrm{LL},1}$	$\widehat{m}_{\mathrm{SBLL},2}$	$\widetilde{m}_{\mathrm{LL},2}$	$\widehat{m}_{\mathrm{SBLL},3}$	$\widetilde{m}_{\mathrm{LL},3}$
200	0.16189	0.13362	0.09505	0.08161	0.07961	0.07288
500	0.07745	0.07210	0.04150	0.03945	0.03604	0.03727
1000	0.04270	0.04214	0.02078	0.02070	0.01773	0.01768

Table 5: The estimates (EST), standard errors (SE) and lower bound (LB) and upper bound (UB) of 95% confidence intervals of β_l, $1 \le l \le 3$, in model 19.

		EST	LB	UB	P-value
X_1 = intercept					
β_1	β_{11}	0.469	0.318	0.621	< 0.001
	β_{12}	0.547	0.365	0.730	< 0.001
	β_{13}	0.656	0.510	0.801	< 0.001
	β_{14}	0.162	−0.013	0.337	0.070
	β_{15}	0.129	0.014	0.243	0.028
	β_{16}	0.088	−0.016	0.192	0.097
X_2 = age					
β_2	β_{21}	0.020	−0.145	0.186	0.812
	β_{22}	0.309	0.211	0.408	< 0.001
	β_{23}	0.001	−0.255	0.257	0.995
	β_{24}	0.950	0.919	0.982	< 0.001
	β_{25}	0.018	−0.044	0.080	0.581
	β_{26}	0.013	−0.036	0.061	0.616
X_3 = fat free weight					
β_3	β_{31}	0.390	0.177	0.604	< 0.001
	β_{32}	0.455	0.061	0.850	0.024
	β_{33}	0.334	−0.046	0.714	0.085
	β_{34}	0.727	0.354	1.101	< 0.001
	β_{35}	0.007	−0.169	0.182	0.940
	β_{36}	0.003	−0.205	0.210	0.979

Table 6: The indices of the components in the significant subsets of β_l with no more than 3 components, the corresponding critical values (C-value) and p-values.

β_1	C-value	P-value	β_2	C-value	P-value	β_3	C-value	P-value
(2,3)	337.92	< 0.0001	(2,3)	52.30	< 0.0001	(2,3)	8.93	0.0115
(2,4)	37.20	< 0.0001	(2,4)	228094.60	< 0.0001	(2,4)	92.53	< 0.0001
(2,5)	36.43	< 0.0001	(2,5)	41.29	< 0.0001	(2,5)	6.06	< 0.0001
(2,6)	43.18	< 0.0001	(2,6)	37.76	< 0.0001	(2,6)	6.95	0.0310
(3,4)	113.5054	< 0.0001	(3,4)	4642.92	< 0.0001	(3,4)	52.59	< 0.0001
(3,5)	79.74	< 0.0001						
(3,6)	64.53	< 0.0001						
(4,5)	7.96	0.0187	(4,5)	3662.91	< 0.0001	(4,5)	24.30	< 0.0001
(4,6)	6.14	0.0464	(4,6)	79.44	< 0.0001	(4,6)	34.79	< 0.0001
(5,6)	13.80	0.0010						
(2,3,4)	512.97	< 0.0001	(2,3,4)	980942.20	< 0.0001	(2,3,4)	360.04	< 0.0001
(2,3,5)	339.72	< 0.0001	(2,3,5)	52.39	< 0.0001	(2,3,5)	9.38	0.0247
(2,3,6)	338.85	< 0.0001	(2,3,6)	52.62	< 0.0001	(2,3,6)	13.44	0.0038
(2,4,5)	37.43	< 0.0001	(2,4,5)	476617.70	< 0.0001	(2,4,5)	92.60	< 0.0001
(2,4,6)	43.97	< 0.0001	(2,4,6)	252412.90	< 0.0001	(2,4,6)	104.57	< 0.0001
(2,5,6)	47.36	< 0.0001	(2,5,6)	43.65	< 0.0001			
(3,4,5)	142.75	< 0.0001	(3,4,5)	4809.96	< 0.0001	(3,4,5)	55.27	< 0.0001
(3,4,6)	113.97	< 0.0001	(3,4,6)	4642.93	< 0.0001	(3,4,6)	60.55	< 0.0001
(3,5,6)	85.89	< 0.0001						
(4,5,6)	18.28	0.0004	(4,5,6)	4270.81	< 0.0001	(4,5,6)	38.66	< 0.0001

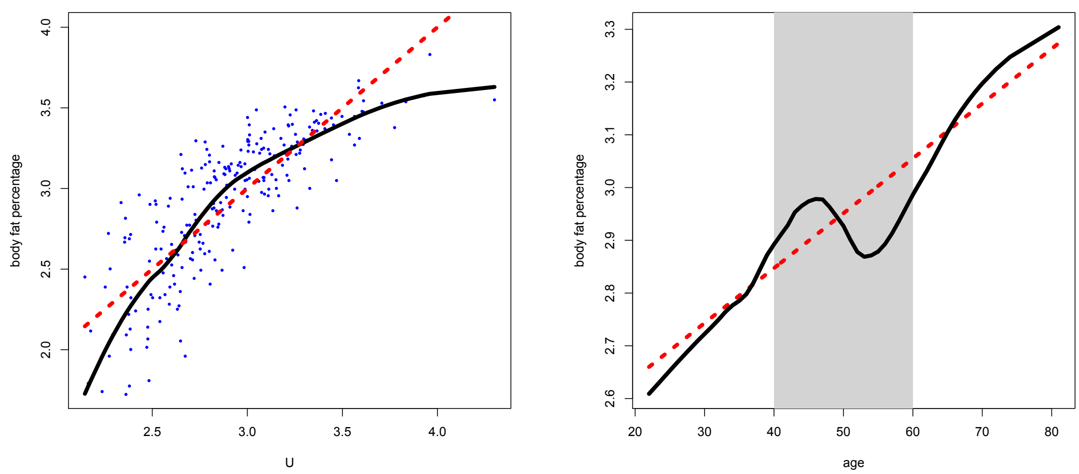

Figure 1: Plots of the fitted curves where in the left panel u is an index covariate given as a linear combination of 6 circumference measurements. The solid curve denotes the local linear fitting, the dotted curve stands for the conventional linear regression, and the dots are the observed response values.

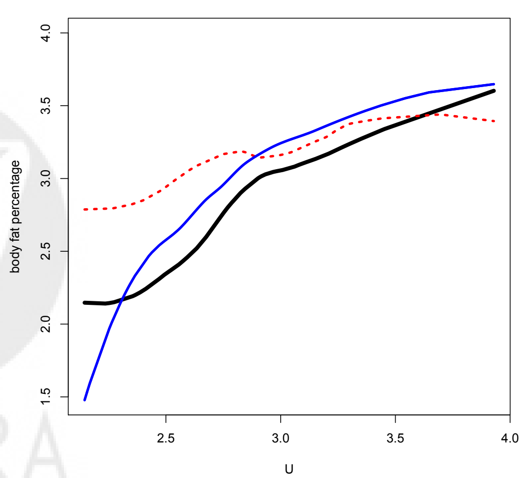

Figure 2: Three fitted curves over the circumference index $u = z^{\mathrm{T}}\widehat{\beta}$ for three age groups 22-39 (thick line), 40-60 (thin line) and 61-81 (dotted line).

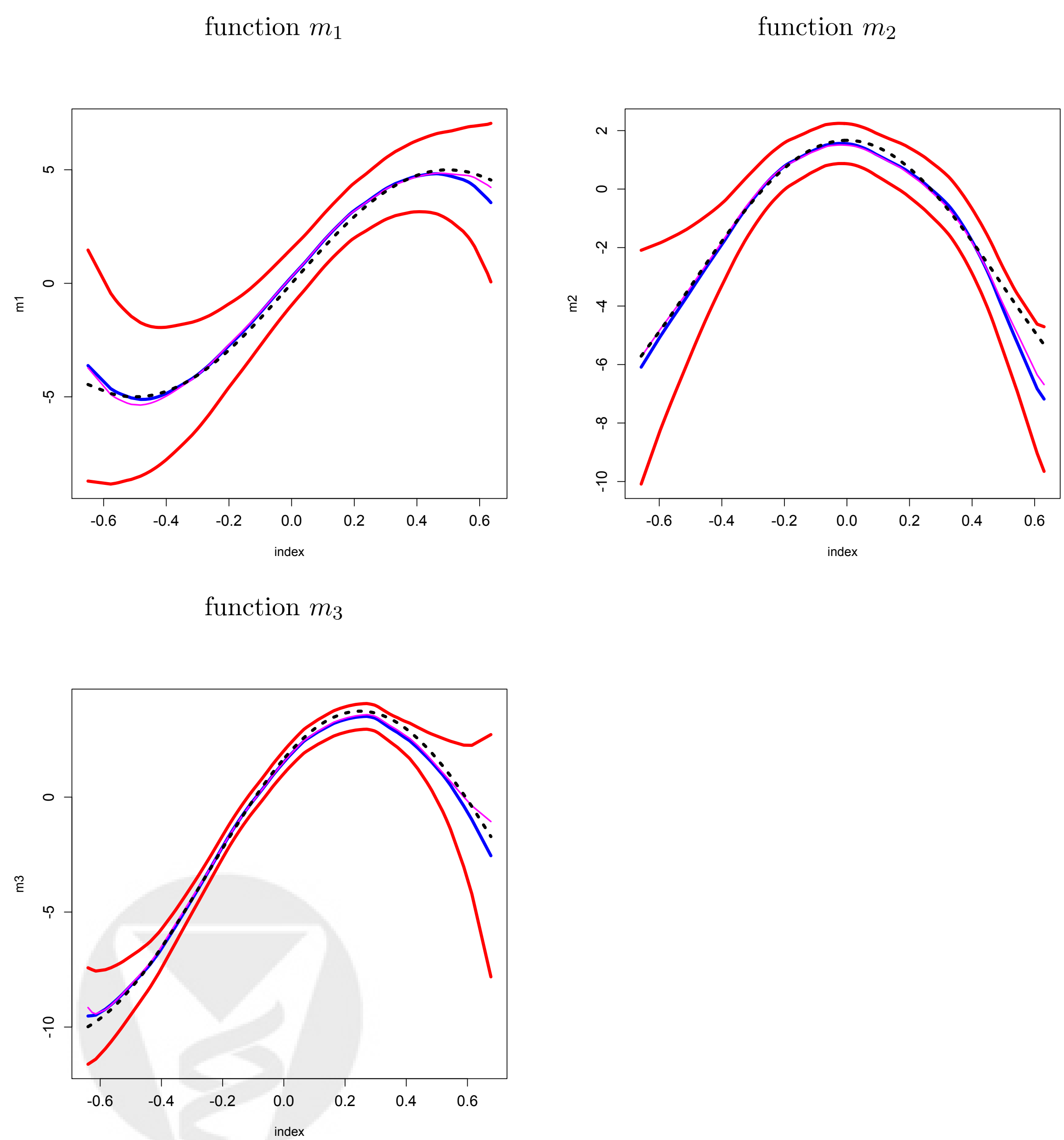

Figure 3: Plots of the two-step SBLL estimator $\widehat{m}_{\mathrm{SBLL},l}(\cdot)$ (thick line), the upper and lower 95% pointwise confidence intervals (upper and lower thick lines), the oracle estimator $\widetilde{m}_{\mathrm{LL},l}(\cdot)$ (thin line) and the true function $m_l(\cdot)$ (dashed line) for $l = 1, 2, 3$ based on one sample with $n = 200$.

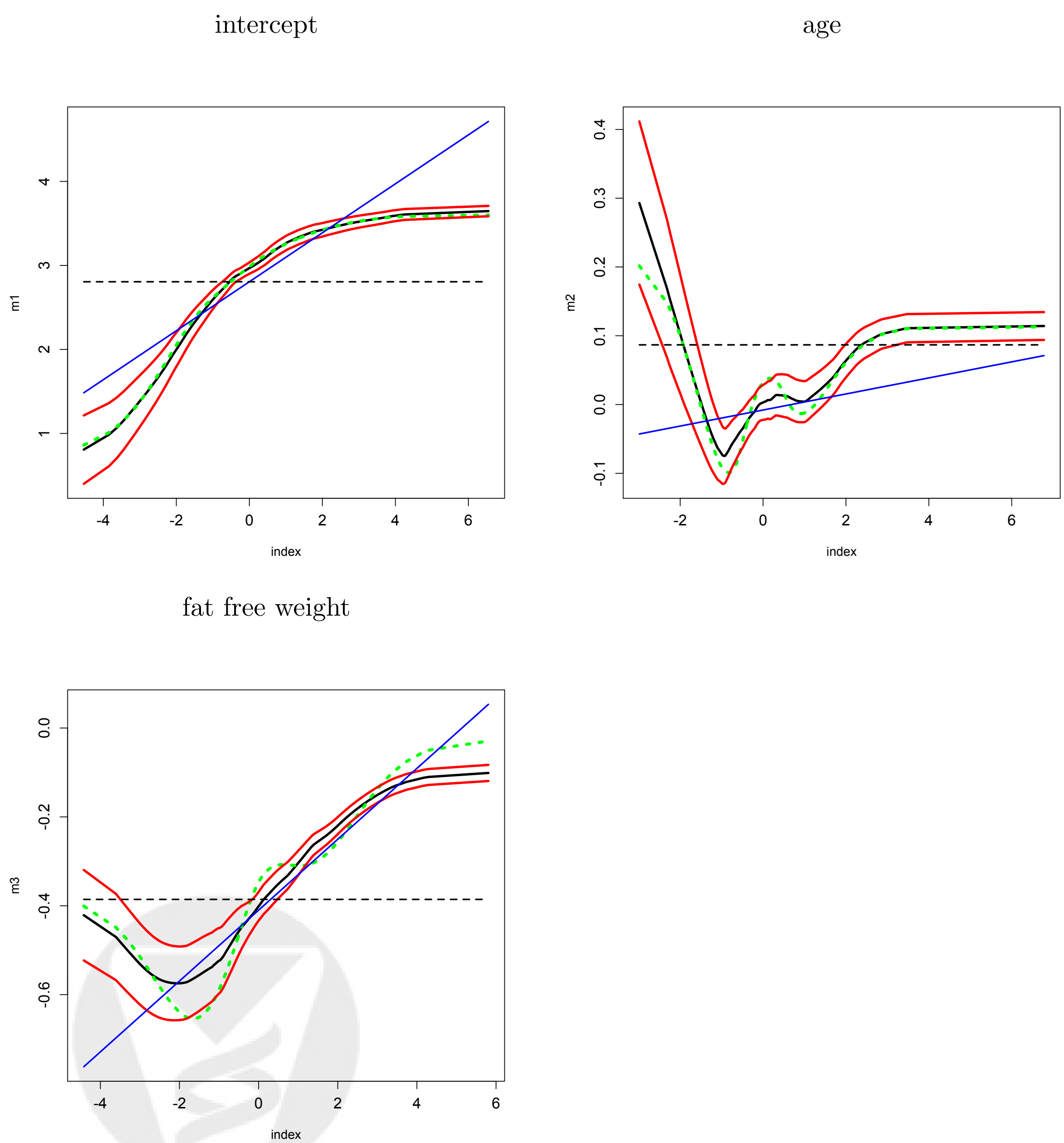

Figure 4: Plots of the SBLL estimator (middle solid line), the one-step spline estimator (middle dashed line), and the 95% pointwise confidence intervals (lower and upper lines) of $m_l(\cdot)$, $1 \le l \le 3$, as well as the estimates $\widehat{m}_{\boldsymbol{\theta},l} = \widehat{\theta}_{l0}$ (horizontal dashed lines) and $\widehat{m}_{\boldsymbol{\theta},l} = \widehat{\theta}_{l1} + \widehat{\theta}_{l2} U_l\left(\widehat{\boldsymbol{\beta}}_l\right)$ (straight thin lines).