Statistica Sinica Preprint No: SS-2017-0206.R2	
Title	Nonlinear interaction detection through model-based sufficient dimension reduction
Manuscript ID	SS-2017-0206.R2
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202017.0206
Complete List of Authors	Guoliang Fan Liping Zhu and Shujie Ma
Corresponding Author	Liping Zhu
E-mail	zhulp1@hotmail.com
Notice: Accepted version subject to English editing.	

NONLINEAR INTERACTION DETECTION THROUGH MODEL-BASED SUFFICIENT DIMENSION REDUCTION

GUOLIANG FAN, LIPING ZHU and SHUJIE MA

October 8, 2017

Abstract

In this paper we propose an efficient model-based sufficient dimension reduction method to detect interactions. We introduce a new class of multivariate adaptive varying index models (MAVIM) to investigate nonlinear interaction effects of the grouped covariates on multivariate response variables. Grouping the covariates through linear combinations in the MAVIM accommodates weak individual interaction effects as long as their joint interaction effects are strong enough to be detectable. We estimate the joint interaction effects by a weighted profile least squares method, which is numerically stable and computationally fast. The resultant profile least squares estimate is root-n consistent and asymptotically normal. We discuss how to choose an optimal weight to improve the estimation efficiency. We determine the structural dimension with a BIC-type

*Guoliang Fan (Email: hfutfgl@163.com) is a postdoctoral fellow and Liping Zhu (Email: zhu.liping@ruc.edu.cn) is the corresponding author and Professor, Institute of Statistics and Big Data and Center for Applied Statistics, Renmin University of China, 59 Zhongguancun Avenue, Haidian District, Beijing 100872, P. R. China. Guoliang Fan is also Professor, Anhui Polytechnic University, 8 Beijing Road, Wuhu, Anhui 241000, P. R. China. Shujie Ma (Email: shujie.ma@ucr.edu) is Associate Professor, Department of Statistics, University of California Riverside, CA 92521, USA.

†Fan's work is supported by National Natural Science Foundation of China (11401006), Chinese Postdoctoral Science Foundation (2017M611083) and a Visiting Program for Young Scholar in Universities (gxfx2017042), Anhui Provincial Education Department, China. Zhu's work is supported by National Natural Science Foundation of China (11371236, 11422107 and 11731011), Ministry of Eduction Project of Key Research Institute of Humanities and Social Sciences at Universities (16JJD910002) and National Youth Top-notch Talent Support Program, China.

‡The authors thank the Editor, an Associated Editor and two reviewers for their constructive comments, which have led to a dramatic improvement of the earlier version of this article.

criterion, and establish its consistency. The effectiveness of our proposal is illustrated through simulation studies and an analysis of Framingham heart study.

KEY WORDS: central mean subspace; dimension determination; high dimensionality; interaction detection; sufficient dimension reduction.

1. INTRODUCTION

With the advance of information technology, high dimensional data are effectively collected at a low cost in many scientific fields. Regression analysis is perhaps one of the most popular tools that help us gain insight into the relationship between two sets of high dimensional variables. Suppose $\mathbf{x} = (X_1, \ldots, X_p)^{\mathrm{T}} \in \mathbb{R}^p$ is the covariate vector and $\mathbf{y} = (Y_1, \ldots, Y_r)^{\mathrm{T}} \in \mathbb{R}^r$ is the response vector. In general, the goal of regression analysis is to study how the conditional mean function of $\mathbf{y}$, denoted by $E(\mathbf{y} \mid \mathbf{x})$, varies with $\mathbf{x}$. Nonparametric regression is a flexible and effective approach to estimating the conditional mean. However, it suffers from the "curse of dimensionality" when the dimension of $\mathbf{x}$ increases. To reduce the covariate dimension, sufficient dimension reduction (Cook, 1998) is an effective tool that combines the concept of sufficiency with the idea of dimension reduction. It achieves the goal of dimension reduction through replacing the high dimensional $\mathbf{x}$ with d linear combinations, denoted as $(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x})$ for $\boldsymbol{\alpha} \in \mathbb{R}^{p\times d}$, without losing information of $(\mathbf{y} \mid \mathbf{x})$. In other words, replacing $\mathbf{x}$ with $(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x})$ is "sufficient" in the sense that $E(\mathbf{y} \mid \mathbf{x}) = E(\mathbf{y} \mid \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x})$. By the very purpose of dimension reduction, d is assumed to be small and hence estimating $E(\mathbf{y} \mid \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x})$ via nonparametric smoothing is straightforward. In practice, d is unknown and needs to be estimated from the data. This differentiates the dimension reduction model $E(\mathbf{y} \mid \mathbf{x}) = E(\mathbf{y} \mid \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x})$ from the single- or multiple-index models in the literature (Ichimura, 1993; Carroll et al., 1997; Ma and Zhu, 2014). Seeking for an appropriate $\boldsymbol{\alpha}$ such that $E(\mathbf{y} \mid \mathbf{x}) = E(\mathbf{y} \mid \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x})$, is the central goal of sufficient dimension reduction when estimating $E(\mathbf{y} \mid \mathbf{x})$ is concerned. Because $\boldsymbol{\alpha}$ is not identifiable, the space spanned by $\boldsymbol{\alpha}$, denoted as $\text{span}(\boldsymbol{\alpha})$, with the minimal column dimension, is the parameter of primary interest and referred to as the central mean space (Cook and Li, 2002) in sufficient dimension reduction.

In the present article we consider the problem of sufficient dimension reduction in the presence of high dimensional controlling variables $\mathbf{z} = (Z_1, \ldots, Z_q)^{\mathrm{T}}$. This falls into the framework of partial mean dimension reduction (Li et al., 2003), which seeks for a $p \times d_0$ matrix $\boldsymbol{\beta}$ such that $E(\mathbf{y} \mid \mathbf{x}, \mathbf{z}) = E(\mathbf{y} \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}, \mathbf{z})$. Similar to $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ is not identifiable either. Therefore, our primary goal is to seek for the minimal column space of $\boldsymbol{\beta}$, denoted by $\text{span}(\boldsymbol{\beta})$, such that $E(\mathbf{y} \mid \mathbf{x}, \mathbf{z}) = E(\mathbf{y} \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}, \mathbf{z})$. Following the convention of sufficient dimension reduction (Li et al., 2003), we refer to the column space of $\boldsymbol{\beta}$, denoted by $\text{span}(\boldsymbol{\beta})$, such that $E(\mathbf{y} \mid \mathbf{x}, \mathbf{z}) = E(\mathbf{y} \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}, \mathbf{z})$ as the partial central mean dimension reduction subspace. Its column dimension, denoted by d_0, is referred to as the structural dimension of $\text{span}(\boldsymbol{\beta})$. Existing methods require $r = q = 1$. Specifically, if $\mathbf{z}$ and $\mathbf{y}$ are univariate and $\mathbf{z}$ is categorical taking a small number of values, Li et al. (2003) suggested to apply an existing sufficient dimension reduction method to $E(\mathbf{y} \mid \mathbf{x}, \mathbf{z} = \mathbf{z}_0)$ within each category of $\mathbf{z}$, say, $\mathbf{z} = \mathbf{z}_0$, and sum up all estimates to form an estimate of $\text{span}(\boldsymbol{\beta})$; if $\mathbf{y}$ is univariate and $\mathbf{z}$ is continuous and low dimensional, Feng et al. (2013) suggested to discretize $\mathbf{z}$ into a series of binary variables and then apply Li et al. (2003)'s method to form an estimate of $\text{span}(\boldsymbol{\beta})$. Hilafu and Wu (2017) suggest to recover $\text{span}(\boldsymbol{\beta})$ through regressing $\widetilde{\mathbf{y}} \stackrel{\text{def}}{=} (\mathbf{y}^{\mathrm{T}}, \mathbf{z}^{\mathrm{T}})^{\mathrm{T}}$ onto $\mathbf{x}$. In this article we consider a more general situation that both $\mathbf{z}$ and $\mathbf{y}$ are allowed to be multivariate and components in $\mathbf{z}$ can be either categorical or continuous. Such considerations are motivated by the Framingham Heart Study, where 304 subjects are collected to evaluate the effects of physical exercises on the blood pressures. The Framingham Heart Study

Data were downloaded from NCBI dbGaP with an IRB number HS-11-159. Both the systolic (Y_1) and the diastolic blood pressures (Y_2), and several measurements of physical exercises, such as hours for heavy, moderate and light activities per day, denoted by Z_1, Z_2 and Z_3, respectively, are recorded for each subject. There is a public health concern regarding the developmental effects resulting from the lack of physical exercises. It is believed that a moderate amount of physical exercises help to relieve stress, and hence are beneficial to control for the blood pressures. Therefore, our goal is to investigate whether or not the physical exercises, Z_1, Z_2 and Z_3, affect the blood pressures. The blood pressures are also relevant to the degree of obesity, which are measured with weight (X_1), height (X_2), bi-deltoid girth (X_3), right arm girth-upper third (X_4), waist girth (X_5), hip girth (X_6) and thigh girth (X_7).

In general, to understand how the conditional mean functions of $\mathbf{y} = (Y_1, \ldots, Y_r)^{\mathrm{T}}$ vary with $\mathbf{x} = (X_1, \ldots, X_p)^{\mathrm{T}}$ in the presence of $\mathbf{z} = (Z_1, \ldots, Z_q)^{\mathrm{T}}$, and to simultaneously accommodate the interaction effects between $\mathbf{x}$ and $\mathbf{z}$, we consider the following multivariate adaptive varying index models (MAVIM for short),

$$E(\mathbf{y} \mid \mathbf{x}, \mathbf{z}) = \sum_{k=1}^{q} \mathbf{m}_k(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}) Z_k, \text{or equivalently,} \tag{1.1}$$

$$E(Y_l \mid \mathbf{x}, \mathbf{z}) = \sum_{k=1}^{q} m_{kl}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}) Z_k, \quad \text{for} \quad l = 1, \cdots, r, \tag{1.2}$$

where $\mathbf{m}_k(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}) = \{m_{k1}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}), \cdots, m_{kr}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})\}^{\mathrm{T}}$, $k = 1, \cdots, q$, $\boldsymbol{\beta}$ is a $p \times d_0$ matrix with an unknown d_0. All $\mathbf{m}_k$s, $\boldsymbol{\beta}$ and d_0 have to be estimated from data. All $\mathbf{m}_k$s share an identical $\boldsymbol{\beta}$ to ensure that span($\boldsymbol{\beta}$) is identifiable (Zhu and Zhong, 2015). We group $\mathbf{x}$ through ($\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}$) to augment the interaction effects between $\mathbf{x}$ and Z_k, which is a very useful strategy if the joint interaction effects between $\mathbf{x}$ and Z_k are strong enough but the individual interaction effects between X_i and Z_k are too weak to be detectable.

The functions $\mathbf{m}_k$s accommodate nonlinear interaction effects between $\mathbf{x}$ and Z_k (Ma et al., 2011). To be precise, if $\mathbf{m}_k$ is a constant, $\mathbf{x}$ does not interact with Z_k. In addition, if the i-th row of $\boldsymbol{\beta}$ is zero, X_i does not interact with $\mathbf{z}$. In other words, the i-th row of $\boldsymbol{\beta}$ describes the joint interaction effects between X_i and $\mathbf{z}$. Model (1.1), or equivalently, (1.2), characterizes the interaction effects between $\mathbf{x}$ and $\mathbf{z}$. We achieve the goal of dimension reduction through replacing $\mathbf{x}$ with $(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})$ in the presence of $\mathbf{z}$ and such a reduction is sufficient in the sense that (1.1), or equivalently, (1.2), holds almost surely. Because we allow for a general d_0, that all $\mathbf{m}_k$s share a common $\boldsymbol{\beta}$ in model (1.1) is an imperative assumption for the purposes of identifiability.

Our goal is to estimate and make inference on span($\boldsymbol{\beta}$). Towards this goal, we recast the problem of estimating span($\boldsymbol{\beta}$) to the problem of estimating an identifiable basis matrix $\boldsymbol{\beta}$. We propose a weighted profile least squares estimation procedure for $\boldsymbol{\beta}$ in which each $\mathbf{m}_k$ is approximated by the local linear regression (Fan and Gijbels, 1996). An important methodological merit of our approach is the ease of simultaneously approximating multiple nonparametric functions to create a single objective function for $\boldsymbol{\beta}$, so that the profile least squares estimation can be established in a straightforward manner. The resultant estimate of $\boldsymbol{\beta}$ is root-n consistent and asymptotically normal. We devise a Wald chi-square testing procedure for $\boldsymbol{\beta}$ based on the asymptotic distribution of the profile least squares estimate. In the Framingham Heart Study, the systolic (Y_1) and the diastolic blood pressures (Y_2) are highly correlated and hence are considered jointly to improve the efficiency of the profile least squares estimate. We also discuss how to choose an optimal weight to improve efficiency of estimating $\boldsymbol{\beta}$, a particular basis matrix of span($\boldsymbol{\beta}$). Because the model structure is assumed in (1.1), we refer to our proposal as a model-based sufficient dimension reduction method.

This article is organized as follows. Section 2 introduces the weighted profile least squares estimation and presents asymptotic properties of the proposed estimate. We also discuss how to choose an optimal weight matrix. In Section 3 we evaluate finite sample properties of the proposed estimation and inference procedures via comprehensive simulation studies. We also illustrate the usefulness of our proposals through an analysis of the Framingham Heart Study. Some concluding remarks are given in Section 4. All technical details and additional discussions are given in an online supplement.

2. METHODOLOGY DEVELOPMENT

We seek for $\boldsymbol{\beta}$ with the minimal column dimension d_0 such that (1.1) holds, in other words, replacing $\mathbf{x}$ with $(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})$ is sufficient to describe how $E(\mathbf{y} \mid \mathbf{x}, \mathbf{z})$ varies with $\mathbf{x}$ and $\mathbf{z}$. We further assume that the upper $d_0 \times d_0$ submatrix of $\boldsymbol{\beta}$ is an identity matrix. In other words, $\boldsymbol{\beta} = (\mathbf{I}_{d_0\times d_0}, \boldsymbol{\beta}^{\mathrm{T}}_{-d_0})^{\mathrm{T}}$, where $\mathbf{I}_{d_0\times d_0}$ is a $d_0 \times d_0$ identity matrix and $\boldsymbol{\beta}_{-d_0}$ is a $(p - d_0) \times d_0$ matrix composed of the lower $(p - d_0)$ rows of $\boldsymbol{\beta}$. Recall that in single index models (Ichimura, 1993) where $d_0 = 1$, there are two options to ensure that $\boldsymbol{\beta}$ is identifiable. The first is to restrict that $\boldsymbol{\beta}$ is of unit-length and the first entry of $\boldsymbol{\beta}$ is strictly positive. The second is to simply set the first entry of $\boldsymbol{\beta}$ to be 1 and thus all other entries are free parameters. These two options are in spirit equivalent. Requiring the upper $d_0 \times d_0$ submatrix of $\boldsymbol{\beta}$ to be an identity matrix is an extension of the second option. Such a parameterization is also used by Ma and Zhu (2013) and implies that the first d_0 covariates of $\mathbf{x}$ contribute to model (1.1). If this is not the case, one can always rotate the order of the entries in $\mathbf{x}$ to guarantee that the first d_0 components of $\mathbf{x}$ are useful. Through parameterizing span$(\boldsymbol{\beta})$ with a particular basis matrix $\boldsymbol{\beta} = (\mathbf{I}_{d_0\times d_0}, \boldsymbol{\beta}^{\mathrm{T}}_{-d_0})^{\mathrm{T}}$, we convert the problem of estimating $\boldsymbol{\beta}$ into a problem of estimating the $(p - d_0) \times d_0$ matrix $\boldsymbol{\beta}_{-d_0}$, the free parameters in $\boldsymbol{\beta}$.

Because the structural dimension d_0 of span$(\boldsymbol{\beta})$ is unknown a priori, we illustrate our proposed estimation procedure with a working dimension d. In this section we propose a profile least squares method to estimate $\boldsymbol{\beta}$, or equivalently, $\boldsymbol{\beta}_{-d}$. Let $\mathbf{x}_d = (X_1, \cdots, X_d)^{\mathrm{T}}$ and $\mathbf{x}_{-d} = (X_{d+1}, \cdots, X_p)^{\mathrm{T}}$. Hence, $\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x} = \mathbf{x}_d + \boldsymbol{\beta}_{-d}^{\mathrm{T}}\mathbf{x}_{-d}$ and $\mathbf{m}_k(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}) = \mathbf{m}_k(\mathbf{x}_d + \boldsymbol{\beta}_{-d}^{\mathrm{T}}\mathbf{x}_{-d})$. Suppose that $\{(\mathbf{x}_i, \mathbf{z}_i, \mathbf{y}_i), i = 1, \cdots, n\}$ is a random sample of $(\mathbf{x}, \mathbf{z}, \mathbf{y})$, which follows model (1.1). For a given $\boldsymbol{\beta}$, we estimate $\mathbf{m}_k$s, $k = 1, \cdots, q$, using the local linear approximation (Fan and Gijbels, 1996). Specifically, for $\mathbf{U} = (\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})$ in a small neighborhood of $\mathbf{u}$, one can approximate $\mathbf{m}_k(\mathbf{U}) \approx \mathbf{m}_k(\mathbf{u}) + \mathbf{m}_k^{(1)}(\mathbf{u})(\mathbf{U} - \mathbf{u}) \stackrel{\text{def}}{=} \mathbf{a}_k + \mathbf{B}_k(\mathbf{U} - \mathbf{u})$, for $k = 1, \cdots, q$, where $\mathbf{m}_k^{(1)}(\mathbf{u})$, for $k = 1, \cdots, q$, denotes the first derivative of $\mathbf{m}_k(\mathbf{u})$ with respect to $\mathbf{u}$ and hence all of them are $r \times d$ matrices. The local linear estimators for $\mathbf{m}_k(\mathbf{u})$ and $\mathbf{m}_k^{(1)}(\mathbf{u})$ are defined as $\widehat{\mathbf{m}}_k(\mathbf{u}, \boldsymbol{\beta}) = \widehat{\mathbf{a}}_k$ and $\widehat{\mathbf{m}}_k^{(1)}(\mathbf{u}, \boldsymbol{\beta}) = \widehat{\mathbf{B}}_k$ at the fixed point $\boldsymbol{\beta}$, where $\{(\widehat{\mathbf{a}}_k, \widehat{\mathbf{B}}_k), k = 1, \cdots, q\}$ minimize the sum of the weighted least squares

$$\sum_{i=1}^{n} \left[\mathbf{y}_i - \sum_{k=1}^{q} \{\mathbf{a}_k + \mathbf{B}_k(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i - \mathbf{u})\} Z_{ik} \right]^2 K_h(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i - \mathbf{u}),$$

where $K_h(\cdot) = K(\cdot/h)/h^d$ is a product of d univariate kernel functions and h is a bandwidth. By some straightforward algebraic calculations, we derive that

$$\left\{\widehat{\mathbf{m}}_1(\mathbf{u}, \boldsymbol{\beta}), \cdots, \widehat{\mathbf{m}}_q(\mathbf{u}, \boldsymbol{\beta}), h\widehat{\mathbf{m}}_1^{(1)}(\mathbf{u}, \boldsymbol{\beta}), \cdots, h\widehat{\mathbf{m}}_q^{(1)}(\mathbf{u}, \boldsymbol{\beta})\right\}^{\mathrm{T}} = \mathbf{S}_n^{-1}(\mathbf{u}, \boldsymbol{\beta})\boldsymbol{\xi}_n(\mathbf{u}, \boldsymbol{\beta}), \quad (2.1)$$

where $\mathbf{S}_n(\mathbf{u}, \boldsymbol{\beta}) \stackrel{\text{def}}{=} \begin{pmatrix} \mathbf{S}_{n0}(\mathbf{u}, \boldsymbol{\beta}) & \mathbf{S}_{n1}^{\mathrm{T}}(\mathbf{u}, \boldsymbol{\beta}) \\ \mathbf{S}_{n1}(\mathbf{u}, \boldsymbol{\beta}) & \mathbf{S}_{n2}(\mathbf{u}, \boldsymbol{\beta}) \end{pmatrix}$ and $\boldsymbol{\xi}_n(\mathbf{u}, \boldsymbol{\beta}) \stackrel{\text{def}}{=} \begin{pmatrix} \boldsymbol{\xi}_{n0}(\mathbf{u}, \boldsymbol{\beta}) \\ \boldsymbol{\xi}_{n1}(\mathbf{u}, \boldsymbol{\beta}) \end{pmatrix}$, with

$$\mathbf{S}_{nj}(\mathbf{u}, \boldsymbol{\beta}) \stackrel{\text{def}}{=} n^{-1} \sum_{i=1}^{n} \mathbf{z}_i \mathbf{z}_i^{\mathrm{T}} \otimes \left(\frac{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i - \mathbf{u}}{h} \right)^j K_h(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i - \mathbf{u}) \quad \text{and}$$

$$\boldsymbol{\xi}_{nj}(\mathbf{u}, \boldsymbol{\beta}) \stackrel{\text{def}}{=} n^{-1} \sum_{i=1}^{n} \mathbf{z}_i \otimes \left\{ \left(\frac{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i - \mathbf{u}}{h} \right)^j \mathbf{y}_i^{\mathrm{T}} \right\} K_h(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i - \mathbf{u}).$$

Here $\mathbf{A} \otimes \mathbf{B} = (a_{ij}\mathbf{B})$ for $\mathbf{A} = (a_{ij})$, and $\mathbf{A}^0 = 1$, $\mathbf{A}^1 = \mathbf{A}$ and $\mathbf{A}^2 = \mathbf{A}\mathbf{A}^{\mathrm{T}}$. For a fixed

$\boldsymbol{\beta}$, $\mathbf{m}_k$ is now profiled out. Subsequently, we estimate $\boldsymbol{\beta}_{-d}$ through minimizing

$$\sum_{i=1}^{n}\left\{\mathbf{y}_i-\sum_{k=1}^{q}\widehat{\mathbf{m}}_k(\mathbf{x}_{d,i}+\boldsymbol{\beta}_{-d}^{\mathrm{T}}\mathbf{x}_{-d,i},\boldsymbol{\beta})Z_{ik}\right\}^{\mathrm{T}}\mathbf{W}\left\{\mathbf{y}_i-\sum_{k=1}^{q}\widehat{\mathbf{m}}_k(\mathbf{x}_{d,i}+\boldsymbol{\beta}_{-d}^{\mathrm{T}}\mathbf{x}_{-d,i},\boldsymbol{\beta})Z_{ik}\right\}, \tag{2.2}$$

where $\mathbf{W}$ is a user-specified $r\times r$ positive-definite weight matrix. Denote $\widehat{\boldsymbol{\beta}}_{-d,\mathbf{w}}$ is the profile least squares estimate of $\widehat{\boldsymbol{\beta}}_{-d,\mathbf{w}}$ if the working weight matrix $\mathbf{W}$ is used.

We assume the following regularity conditions to establish the asymptotic normality property for $\widehat{\boldsymbol{\beta}}_{-d,\mathbf{w}}$. For notational clarity, let $f(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})$ be the density function of $(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})$, $\mathbf{m}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})=\{\mathbf{m}_1(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}),\cdots,\mathbf{m}_r(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})\}^{\mathrm{T}}$, $\mathbf{m}_k^{(1)}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})=\{\mathbf{m}_{k1}^{(1)}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}),\cdots,\mathbf{m}_{kr}^{(1)}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})\}^{\mathrm{T}}$ be the first derivative of $\mathbf{m}_k(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})$ with respect to $(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})$ for $k=1,\cdots,q$.

(C1) (*The Lipschitz Continuity*) The density function $f(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})$ of $(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})$ is locally Lipschitz continuous, and bounded away from zero and infinity. In addition, $\mathbf{m}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})$, $E(\mathbf{x}\mid\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})$ and $\boldsymbol{\Omega}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})=E(\mathbf{z}\mathbf{z}^{\mathrm{T}}\mid\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})$ are locally Lipschitz continuous.

(C2) (*The Kernel Function*) The univariate kernel function $K(\cdot)$ is symmetric, has a compact support and is Lipschitz continuous. In addition,

$$\int K(u)du=1,\int u^kK(u)du=0,\ \text{for } k=1,\cdots,s-1,\ \text{and } 0\neq\int u^sK(u)du<\infty.$$

The d-dimensional kernel is a product of d univariate kernels. We abuse the notation of K here when it is sufficiently clear from the context.

(C3) (*The Bandwidth*) The bandwidth $h=O(n^{-\delta})$ for $(4s)^{-1}<\delta<(2d)^{-1}$.

(C4) (*The Moment Condition*) All the involved moments, $E\big[\{\mathbf{m}_k(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})\}^{\mathrm{T}}\{\mathbf{m}_k(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})\}\big]$, $E(\mathbf{x}^{\mathrm{T}}\mathbf{x})$, $E\{(\mathbf{y}^{\mathrm{T}}\mathbf{y})^{\kappa_1}\}$ and $E\big[\{\mathbf{m}_k^{(1)}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})\}^{\mathrm{T}}\{\mathbf{m}_k^{(1)}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})\}\big]$, for some $\kappa_1\geq 3/2$ and

$k = 1, \ldots, q$, exist.

These conditions are generally regarded as mild. In particular, condition (C1) imposes smoothness conditions on the mean and density functions, which allows us to implement local smoothers such as kernel and local polynomial regressions (Fan and Gijbels, 1996). Condition (C2) states that an s-th order kernel function is used. Condition (C3) specifies the order of the bandwidth, whose range is fairly wide and more importantly, contains an optimal order. We assume moment conditions in condition (C4) to establish the asymptotic normality. Similar conditions are also assumed in Ma and Zhu (2012, 2013).

Define

$$\mathbf{A}_{\mathbf{w}} \stackrel{\text{def}}{=} E\Big[\Big\{\sum_{k=1}^{q} \mathbf{m}_k^{(1),\mathrm{T}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}) Z_k \otimes \widetilde{\mathbf{x}}_{-d}\Big\} \mathbf{W} \Big\{\sum_{k=1}^{q} \mathbf{m}_k^{(1)}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}) Z_k \otimes \widetilde{\mathbf{x}}_{-d}^{\mathrm{T}}\Big\}\Big], \text{ and}$$

$$\mathbf{B}_{\mathbf{w}} \stackrel{\text{def}}{=} E\Big[\Big\{\sum_{k=1}^{q} \mathbf{m}_k^{(1),\mathrm{T}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}) Z_k \otimes \widetilde{\mathbf{x}}_{-d}\Big\} \mathbf{W}\boldsymbol{\Sigma}\mathbf{W} \Big\{\sum_{k=1}^{q} \mathbf{m}_k^{(1)}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}) Z_k \otimes \widetilde{\mathbf{x}}_{-d}^{\mathrm{T}}\Big\}\Big].$$

Theorem 1 suggests that $\widehat{\boldsymbol{\beta}}_{-d,\mathbf{w}}$ is root-n consistent and asymptotically normal.

Theorem 1. *Suppose that conditions (C1)-(C4) in the Appendix hold. Then*

$$n^{1/2}\big\{\mathrm{vec}(\widehat{\boldsymbol{\beta}}_{-d,\mathbf{w}}) - \mathrm{vec}(\boldsymbol{\beta}_{-d})\big\} \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbf{A}_{\mathbf{w}}^{-1}\mathbf{B}_{\mathbf{w}}\mathbf{A}_{\mathbf{w}}^{-1}),$$

where $\widetilde{\mathbf{x}}_{-d} = \mathbf{x}_{-d} - E(\mathbf{x}_{-d} \mid \boldsymbol{\beta}^{T}\mathbf{x})$ *and "* $\xrightarrow{d}$ *" stands for "convergence in distribution".*

How to specify the working weight matrix $\mathbf{W}$ is another interesting issue. As long as $\mathbf{W}$ is positive definite, $\widehat{\boldsymbol{\beta}}_{-d,\mathbf{w}}$ is root-n consistent and asymptotically normal. However, choosing an appropriate $\mathbf{W}$ properly may improve the efficiency of estimating $\boldsymbol{\beta}_{-d,\mathbf{w}}$.

We compare two options: $\mathbf{W} = \mathbf{I}_{r\times r}$ and $\mathbf{W} = \widehat{\boldsymbol{\Sigma}}^{-1}$, where

$$\widehat{\boldsymbol{\Sigma}} \overset{\text{def}}{=} n^{-1}\sum_{i=1}^{n} \widehat{\boldsymbol{\varepsilon}}_i \widehat{\boldsymbol{\varepsilon}}_i^{\mathrm{T}} \text{ and } \widehat{\boldsymbol{\varepsilon}}_i \overset{\text{def}}{=} \mathbf{y}_i - \sum_{k=1}^{q} \widehat{\mathbf{m}}_k(\widehat{\boldsymbol{\beta}}_{\mathbf{I}}^{\mathrm{T}}\mathbf{x}_i, \widehat{\boldsymbol{\beta}}_{\mathbf{I}}) Z_{ik}.$$

Theorem 2 indicates that using $\mathbf{W} = \widehat{\boldsymbol{\Sigma}}^{-1}$ yields a more efficient estimate of $\boldsymbol{\beta}_{-d,\mathbf{w}}$ than using $\mathbf{W} = \mathbf{I}_{r\times r}$.

Theorem 2. $\mathbf{A}_{\mathbf{I}}^{-1}\mathbf{B}_{\mathbf{I}}\mathbf{A}_{\mathbf{I}}^{-1} \geq \mathbf{A}_{\boldsymbol{\Sigma}^{-1}}^{-1}\mathbf{B}_{\boldsymbol{\Sigma}^{-1}}\mathbf{A}_{\boldsymbol{\Sigma}^{-1}}^{-1} = \mathbf{A}_{\boldsymbol{\Sigma}^{-1}}^{-1}$.

For the asymptotic normality of $\widehat{\boldsymbol{\beta}}_{-d,\mathbf{w}}$ stated in Theorem 1 to be useful, we provide a consistent estimate for the asymptotic covariance matrix. Let $\widehat{\boldsymbol{\beta}}_{\mathbf{w}} \overset{\text{def}}{=} (\mathbf{I}_{d\times d}, \widehat{\boldsymbol{\beta}}_{-d,\mathbf{w}}^{\mathrm{T}})^{\mathrm{T}}$, where $\mathbf{W}$ can be $\widehat{\boldsymbol{\Sigma}}^{-1}$ or $\mathbf{I}$. We further define

$$\widehat{\widetilde{\mathbf{x}}}_{-d,i} \overset{\text{def}}{=} \mathbf{x}_{-d,i} - \sum_{j=1,j\neq i}^{n} K_h(\widehat{\boldsymbol{\beta}}_{\mathbf{w}}^{\mathrm{T}}\mathbf{x}_j - \widehat{\boldsymbol{\beta}}_{\mathbf{w}}^{\mathrm{T}}\mathbf{x}_i)\mathbf{x}_{-d,i} \Big/ \sum_{j=1,j\neq i}^{n} K_h(\widehat{\boldsymbol{\beta}}_{\mathbf{w}}^{\mathrm{T}}\mathbf{x}_j - \widehat{\boldsymbol{\beta}}_{\mathbf{w}}^{\mathrm{T}}\mathbf{x}_i),$$

$$\widehat{\mathbf{A}}_{\mathbf{w}} \overset{\text{def}}{=} n^{-1}\sum_{i=1}^{n}\left\{\sum_{k=1}^{q}\widehat{\mathbf{m}}_k^{(1),\mathrm{T}}(\widehat{\boldsymbol{\beta}}_{\mathbf{w}}^{\mathrm{T}}\mathbf{x}_i, \widehat{\boldsymbol{\beta}}_{\mathbf{w}})Z_{ik}\otimes\widehat{\widetilde{\mathbf{x}}}_{-d,i}\right\}\mathbf{W}\left\{\sum_{k=1}^{q}\widehat{\mathbf{m}}_k^{(1)}(\widehat{\boldsymbol{\beta}}_{\mathbf{w}}^{\mathrm{T}}\mathbf{x}_i, \widehat{\boldsymbol{\beta}}_{\mathbf{w}})Z_{ik}\otimes\widehat{\widetilde{\mathbf{x}}}_{-d,i}^{\mathrm{T}}\right\},$$

$$\widehat{\mathbf{B}}_{\mathbf{w}} \overset{\text{def}}{=} n^{-1}\sum_{i=1}^{n}\left\{\sum_{k=1}^{q}\widehat{\mathbf{m}}_k^{(1),\mathrm{T}}(\widehat{\boldsymbol{\beta}}_{\mathbf{w}}^{\mathrm{T}}\mathbf{x}_i, \widehat{\boldsymbol{\beta}}_{\mathbf{w}})Z_{ik}\otimes\widehat{\widetilde{\mathbf{x}}}_{-d,i}\right\}\mathbf{W}\widehat{\boldsymbol{\Sigma}}\mathbf{W}\left\{\sum_{k=1}^{q}\widehat{\mathbf{m}}_k^{(1)}(\widehat{\boldsymbol{\beta}}_{\mathbf{w}}^{\mathrm{T}}\mathbf{x}_i, \widehat{\boldsymbol{\beta}}_{\mathbf{w}})Z_{ik}\otimes\widehat{\widetilde{\mathbf{x}}}_{-d,i}^{\mathrm{T}}\right\}.$$

Theorem 3. *Suppose that conditions (C1)-(C4) hold. Then* $\widehat{\mathbf{A}}_{\mathbf{w}} \xrightarrow{p} \mathbf{A}_{\mathbf{w}}$, $\widehat{\mathbf{B}}_{\mathbf{w}} \xrightarrow{p} \mathbf{B}_{\mathbf{w}}$ *and hence* $\widehat{\mathbf{A}}_{\mathbf{w}}^{-1}\widehat{\mathbf{B}}_{\mathbf{w}}\widehat{\mathbf{A}}_{\mathbf{w}}^{-1} \xrightarrow{p} \mathbf{A}_{\mathbf{w}}^{-1}\mathbf{B}_{\mathbf{w}}\mathbf{A}_{\mathbf{w}}^{-1}$, *where "*$\xrightarrow{p}$*" stands for "convergence in probability".*

Testing whether there exist interaction effects between X_i and $\mathbf{z}$ amounts to testing whether all components of the i-th row of $\boldsymbol{\beta}$ in model (1.1) are simultaneously zero. In a general context, we consider the following hypothesis testing problem:

$$H_0 : \mathbf{Q}\boldsymbol{\beta}_{-d} = \mathbf{q}_0 \quad \text{versus} \quad H_1 : \mathbf{Q}\boldsymbol{\beta}_{-d} \neq \mathbf{q}_0,$$

where $\mathbf{Q}$ is a user-specified $q_0 \times (p-d)$ matrix and $\mathbf{q}_0$ is another user-specified $q_0 \times d$ matrix. This hypothesis testing problem is general enough to include a variety of hypothesis of interest. For example, we are free to choose $\mathbf{Q} = (1, 0, \cdots, 0)_{1\times(p-d)}$ and $\mathbf{q}_0 = \mathbf{0}_{1\times d}$, which aims to test whether there exist interaction effects between X_{d_0+1} and $\mathbf{z}$. In general, we devise a Wald chi-square test:

$$\begin{aligned}T_{\mathbf{w}} =& n \left\{(\mathbf{I}_{d\times d} \otimes \mathbf{Q})\text{vec}(\widehat{\boldsymbol{\beta}}_{-d,\mathbf{w}}) - \text{vec}(\mathbf{q}_0)\right\}^{\mathrm{T}} \left\{(\mathbf{I}_{d\times d} \otimes \mathbf{Q})\widehat{\mathbf{A}}_{\mathbf{w}}^{-1}\widehat{\mathbf{B}}_{\mathbf{w}}\widehat{\mathbf{A}}_{\mathbf{w}}^{-1}(\mathbf{I}_{d\times d} \otimes \mathbf{Q}^{\mathrm{T}})\right\}^{-1}\\ &\left\{(\mathbf{I}_{d\times d} \otimes \mathbf{Q})\text{vec}(\widehat{\boldsymbol{\beta}}_{-d,\mathbf{w}}) - \text{vec}(\mathbf{q}_0)\right\}.\end{aligned}$$

A direct application of Theorem 1 yields the following corollary. Its proof is omitted.

Corollary 1. *Suppose that conditions (C1)-(C4) hold. Then under H_0, $T_{\mathbf{w}} \xrightarrow{d} \chi^2(q_0 d)$, where $\chi^2(q_0 d)$ stands for the central chi-square distribution with $(q_0 d)$ degrees of freedom.*

It remains to estimate the structural dimension of span($\boldsymbol{\beta}$), the minimum column dimension of $\boldsymbol{\beta}$, such that (1.1) holds. Following Zhu et al. (2006) and Xu et al. (2016), we suggest a BIC-type criterion. Specifically, for a working dimension d, we define

$$\begin{aligned}\mathcal{L}(d) \stackrel{\text{def}}{=} &\sum_{i=1}^{n}\left\{\mathbf{y}_i - \sum_{k=1}^{q}\widehat{\mathbf{m}}_k(\widehat{\boldsymbol{\beta}}_{d,\mathbf{w}}^{\mathrm{T}}\mathbf{x}_i, \widehat{\boldsymbol{\beta}}_{d,\mathbf{w}})Z_{ik}\right\}^{\mathrm{T}}\left\{\mathbf{y}_i - \sum_{k=1}^{q}\widehat{\mathbf{m}}_k(\widehat{\boldsymbol{\beta}}_{d,\mathbf{w}}^{\mathrm{T}}\mathbf{x}_i, \widehat{\boldsymbol{\beta}}_{d,\mathbf{w}})Z_{ik}\right\}\\ &\Bigg/\left\{\sum_{i=1}^{n}(\mathbf{y}_i - \overline{\mathbf{y}})^{\mathrm{T}}(\mathbf{y}_i - \overline{\mathbf{y}})\right\}^{1/2} \text{ and } \mathcal{L}^*(d) \stackrel{\text{def}}{=} \mathcal{L}(d) + (pd)\lambda_n,\end{aligned}$$

where $\widehat{\boldsymbol{\beta}}_{d,\mathbf{w}} = (\mathbf{I}_{d\times d}, \widehat{\boldsymbol{\beta}}_{-d,\mathbf{w}})^{\mathrm{T}}$. The estimated structural dimension is then given by

$$\widehat{d} \stackrel{\text{def}}{=} \underset{1\le d\le p}{\text{argmin}}\, \mathcal{L}^*(d). \tag{2.3}$$

Theorem 4. *Under the conditions of Theorem 1, and further assume that $\lambda_n/\log n \to$*

∞ *and* $\lambda_n n^{-1/2} \to 0$*, then* $pr(\widehat{d} = d_0) \to 1$.

Theorem 4 demonstrates that the BIC-type criterion enables us to select the true structural dimensional of span($\boldsymbol{\beta}$) consistently. The penalty term λ_n is allowed to vary over a wide range for $\widehat{d}$ to be consistent. Yet how to chose an optimal λ_n is a challenging work. Our limit simulations show that $\lambda_n = n^{2/5}$ works well. We shall use this choice of λ_n throughout our numerical studies.

We outline our algorithm for estimating $\boldsymbol{\beta}$ as follows, starting with a working dimension d and a user-specified initial value of $\boldsymbol{\beta}$.

1. Estimate $\mathbf{m}_k$ and $\mathbf{m}_k^{(1)}$ with (2.1) for a given $\boldsymbol{\beta}$.
2. Set $\mathbf{W} = \mathbf{I}_{r\times r}$. Estimate $\boldsymbol{\beta}$ with (2.2) for given $\mathbf{m}_k$ and $\mathbf{m}_k^{(1)}$.
3. Repeat the above two steps until convergence. The resultant estimate, denoted by $\widehat{\boldsymbol{\beta}}_{\mathbf{I}} = (\mathbf{I}_{d\times d}, \widehat{\boldsymbol{\beta}}^{\mathrm{T}}_{-d,\mathbf{I}})^{\mathrm{T}}$, is referred to as the unweighted profile least squares estimate.
4. We vary the working dimension d from 1 through p and repeat the above three steps. The estimated dimension $\widehat{d}$ is given in (2.3).
5. Set $\mathbf{W} = \widehat{\boldsymbol{\Sigma}}^{-1}$ and $d = \widehat{d}$ in the second step. Repeat the first two steps until convergence. The final estimate, denoted by $\widehat{\boldsymbol{\beta}}_{\widehat{\boldsymbol{\Sigma}}^{-1}} = (\mathbf{I}_{\widehat{d}\times\widehat{d}}, \widehat{\boldsymbol{\beta}}_{-\widehat{d},\widehat{\boldsymbol{\Sigma}}^{-1}})$, is referred to as the weighted profile least squares estimate.

3. NUMERICAL STUDIES

In this section we demonstrate the performance of our proposals through comprehensive simulations and an application to the Framingham Heart Study. Because existing methods cannot be used directly if $\mathbf{y}$ is multivariate, we only report the simulation results of our proposal in Section 3.1 when $\mathbf{y}$ is multivariate. In Section 3.2, we compare our proposal with existing methods proposed by Li et al. (2003), Ma and Song (2015) and Liu et al. (2016) when both $\mathbf{y}$ and $\mathbf{z}$ are univariate.

3.1. Simulation Experiments for Multivariate Response Data

We conduct simulation studies to evaluate the performance of our proposed methodology when the response is multivariate. Throughout our simulations we draw $\mathbf{x}$ and $\mathbf{z}$ independently from multivariate normal distribution with zero mean and covariance matrix $(0.5^{|k-l|})$. We fix $r = 3$, and generate $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \varepsilon_3)^{\mathrm{T}}$ from $\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$, where

$$\boldsymbol{\Sigma} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 4 \end{pmatrix} \begin{pmatrix} 1 & \rho & \rho^2 \\ \rho & 1 & \rho \\ \rho^2 & \rho & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 4 \end{pmatrix}.$$

We consider the following four simulated models.

Model I: A single-index model structure with a linear link function:

$$\begin{cases} Y_1 = 2(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})Z_1 + (\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})Z_2 + \varepsilon_1; \\ Y_2 = (\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})Z_1 + 3(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})Z_2 + \varepsilon_2; \\ Y_3 = \varepsilon_3, \end{cases}$$

Model II: A single-index model structure with a nonlinear link function:

$$\begin{cases} Y_1 = \sin(4\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})Z_1 + 2(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})Z_2 + \varepsilon_1; \\ Y_2 = \cos(2\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})Z_2 + \varepsilon_2; \\ Y_3 = 2(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})Z_1 + \sin(2\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})Z_2 + \varepsilon_3. \end{cases}$$

Model III: A multiple-index model structure with a linear link function:

$$\begin{cases} Y_1 = \{(\boldsymbol{\beta}_1^{\mathrm{T}}\mathbf{x}) + (\boldsymbol{\beta}_2^{\mathrm{T}}\mathbf{x})\}\, Z_1 + (\boldsymbol{\beta}_1^{\mathrm{T}}\mathbf{x})Z_2 + \varepsilon_1; \\ Y_2 = (\boldsymbol{\beta}_2^{\mathrm{T}}\mathbf{x})Z_1 + \{(2\boldsymbol{\beta}_1^{\mathrm{T}}\mathbf{x}) - 3(\boldsymbol{\beta}_2^{\mathrm{T}}\mathbf{x})\}\, Z_2 + \varepsilon_2; \\ Y_3 = 2(\boldsymbol{\beta}_1^{\mathrm{T}}\mathbf{x})Z_1 + 4(\boldsymbol{\beta}_2^{\mathrm{T}}\mathbf{x})Z_2 + \varepsilon_3. \end{cases}$$

Model IV: A multiple-index model structure with a nonlinear link function:

$$\begin{cases} Y_1 = (\boldsymbol{\beta}_1^{\mathrm{T}}\mathbf{x})Z_1/\{0.5 + (\boldsymbol{\beta}_2^{\mathrm{T}}\mathbf{x} + 1.5)^2\} + (\boldsymbol{\beta}_2^{\mathrm{T}}\mathbf{x})Z_2 + \varepsilon_1; \\ Y_2 = \sin^2(\boldsymbol{\beta}_1^{\mathrm{T}}\mathbf{x})Z_1 + \cos^2(\boldsymbol{\beta}_2^{\mathrm{T}}\mathbf{x})Z_2 + \varepsilon_2; \\ Y_3 = \{(2\boldsymbol{\beta}_1^{\mathrm{T}}\mathbf{x}) - (\boldsymbol{\beta}_2^{\mathrm{T}}\mathbf{x})\}^2\, Z_1 + \varepsilon_3. \end{cases}$$

We set $p = 10$, $q = 2$ and $\boldsymbol{\beta} = (1, 0.8, 0.6, 0.4, 0.2, -0.2, -0.4, -0.6, -0.8, 0)^{\mathrm{T}}$ in Models (I)-(II), and let $p = 7$, $q = 2$, $\boldsymbol{\beta}_1 = (1, 0, 0.8, -0.6, 0.4, -0.2, 0)^{\mathrm{T}}$ and $\boldsymbol{\beta}_2 = (0, 1, -0.8, 0.6, -0.4, 0.2, 0)^{\mathrm{T}}$ in Models (III)-(IV). We choose the sample size $n = 200$ and 500 and repeat each simulation 1000 times. We use Gaussian kernel and choose the bandwidth $h = (4/3n)^{1/(d+4)}s$, where s is the median of the robust estimators of the standard deviation of $(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})$.

The average of estimation bias ("bias"), the Monte Carlo standard deviation ("std"), the average of estimated standard deviation ("$\widehat{\text{std}}$"), and the empirical coverage probability ("cvp") at the nominal 95% confidence level for all free parameter are summarized in Tables 1-4 for models (I)-(IV), respectively. It can be clearly seen that all estimates have very small biases, and the biases become smaller as the sample size increases. This phenomenon shows that both the weighted and the unweighted estimates are consistent, which confirms the theoretical result of Theorem 1.

In terms of the Monte Carlo standard deviation and the average of estimated standard deviation, the weighted estimate performs competitively in comparison with the unweighted one. On the other hand, the empirical coverage probabilities for the weighted and unweighted estimators are very close to the nominal level, which implies that our inferential results are fairly reliable. In addition, the Monte Carlo standard deviations are very close to the average of the estimated standard deviations especially for large n. This finding means that the standard deviations have been estimated precisely, which verifies the consistency results stated in Theorem 3.

To demonstrate the performance of our proposed Wald test statistic $T_{\mathbf{w}}$, we test whether X_7 interacts with $\mathbf{z}$ in Model (IV). Towards this end we simply choose $\mathbf{Q} = (0, \cdots, 0, 1)_{1\times 5}$, $\mathbf{q}_0 = \mathbf{0}_{1\times 2}$ in our testing problem. To investigate the size and the power performance of our proposed Wald test, we change the last row of $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$ to (a, a) and reestimate all parameters, where $a = -0.10 : 0.02 : 0.10$. Apparently, $a = 0.00$ corresponds to the case that X_7 does not interact with $\mathbf{z}$. The power curves at the significance level 0.05 are reported in Figure 1 based on 1000 replications. All the power curves increase quickly as $|a|$ increases, indicating that our proposed test approach can detect the interaction effects very effectively.

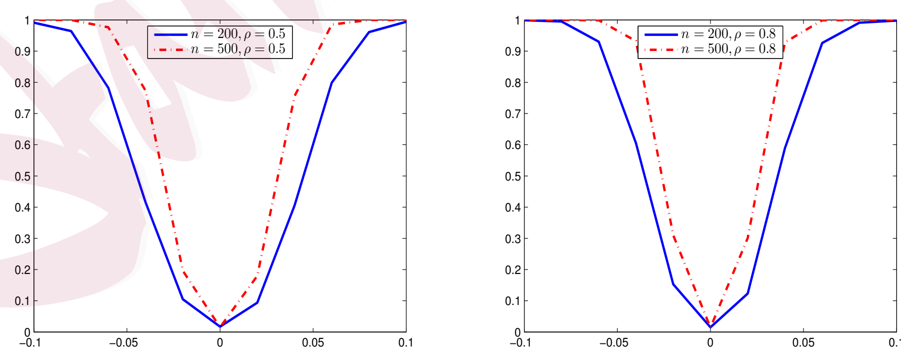

Figure 1: The power curves of $T_{\mathbf{w}}$ for $\rho = 0.5$ (left panel) and $\rho = 0.8$ (right panel) with $n = 200$ (solid line) and 500 (dot dash line).

Next we evaluate the performance of the BIC-type criterion, defined in (2.3), in estimating the structural dimension of span($\boldsymbol{\beta}$). Recall that the structure dimension $d_0 = 1$ in models (I)-(II) and $d_0 = 2$ in models (III) and (IV). For Models (I)-(IV), the percentages for each estimated dimension are charted in Table 5. It can be clearly seen that our proposed BIC-type criterion works pretty well: with high probability the estimated and the true structural dimension are equal in all models. The performance of our proposed BIC-type criterion also improves gradually as the sample size increases.

3.2. Comparison with Existing Methods for Univariate Response Data

In this section we compare our proposal (NEW for short) with existing methods proposed by Li et al. (2003) (LCC for short), Ma and Song (2015) (MS for short) and Liu et al. (2016) (LCL for short) when both $\mathbf{y}$ and $\mathbf{z}$ are univariate. When the response is univariate, the weighted and the unweighted estimates of our proposal are identical, thus we only report the unweighted estimate. The MS method yields two estimates of $\boldsymbol{\beta}$. We report the MS estimate with smaller bias and standard deviation. We compare their performance through the following single index model.

Model V: A single-index model structure with a nonlinear link function:

$$\mathbf{y} = 2(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}) + \sin(4\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})\mathbf{z} + \varepsilon,$$

where $\mathbf{x}$ is drawn independently from multivariate normal distribution with zero mean and covariance matrix $(0.5^{|k-l|})$ and ε follows standard normal distribution. We set $p = 10$ and $\boldsymbol{\beta} = (1, 0.8, 0.6, 0.4, 0.2, -0.2, -0.4, -0.6, -0.8, 0)^{\mathrm{T}}$. We consider two scenarios for $\mathbf{z}$: (i) $\mathbf{z} \sim$Bernoulli(0.5), and (ii) $\mathbf{z} \sim \mathcal{N}(0, 0.5)$. To implement the LCC method when $\mathbf{z}$ is continuous, we discretize $\mathbf{z}$ into a series of binary variables $I(\mathbf{z} \le \widetilde{\mathbf{z}})$, where $\widetilde{\mathbf{z}}$ is an independent copy of $\mathbf{z}$ and $I(A)$ is an indicator function. For each given $\widetilde{\mathbf{z}}$, we can have an estimate of span($\boldsymbol{\beta}$). We then pool all estimates together to yield an integrated estimate of span($\boldsymbol{\beta}$). For fair comparison, we rescale the resulting estimate,

Table 1: Simulation results for Model I: the average bias of the estimators ("bias"), the Monte Carlo standard deviation ("std"), the average of the estimated standard deviation ("$\widehat{\text{std}}$") based on the theoretical calculation, and the empirical coverage probability ("cvp") at the nominal 95% confidence level. All simulation results reported below are multiplied by 100.

		$\widehat{\beta}_2$	$\widehat{\beta}_3$	$\widehat{\beta}_4$	$\widehat{\beta}_5$	$\widehat{\beta}_6$	$\widehat{\beta}_7$	$\widehat{\beta}_8$	$\widehat{\beta}_9$	$\widehat{\beta}_{10}$
True value		0.8	0.6	0.4	0.2	-0.2	-0.4	-0.6	-0.8	0
$\mathbf{W}$		$\rho = 0.5, n = 200$								
$\mathbf{I}$	bias	0.83	0.44	0.16	-0.00	-0.04	-0.47	-0.17	-0.55	-0.03
	std	5.58	4.59	4.45	4.17	4.38	4.47	4.51	4.88	3.17
	$\widehat{\text{std}}$	6.17	4.83	4.54	4.37	4.37	4.56	4.85	5.22	3.85
	cvp	97.20	95.80	95.90	96.40	94.40	95.80	96.50	95.30	95.30
$\widehat{\mathbf{\Sigma}}^{-1}$	bias	0.33	0.17	0.02	-0.08	-0.01	-0.13	0.03	-0.13	-0.11
	std	4.43	3.74	3.56	3.31	3.36	3.55	3.71	3.76	2.84
	$\widehat{\text{std}}$	4.64	3.65	3.45	3.32	3.31	3.45	3.67	3.95	2.93
	cvp	95.80	94.10	94.80	94.70	94.20	93.70	95.00	96.40	95.30
$\mathbf{W}$		$\rho = 0.5, n = 500$								
$\mathbf{I}$	bias	0.39	0.19	0.05	0.11	-0.03	-0.25	-0.10	-0.22	-0.05
	std	3.66	2.97	2.87	2.75	2.77	2.79	2.87	3.06	2.29
	$\widehat{\text{std}}$	3.84	3.00	2.81	2.70	2.70	2.82	3.01	3.23	2.38
	cvp	96.30	95.10	94.20	94.60	94.80	94.70	95.90	96.50	95.90
$\widehat{\mathbf{\Sigma}}^{-1}$	bias	0.07	0.08	-0.04	0.03	-0.02	-0.10	-0.06	-0.01	-0.04
	std	2.89	2.20	2.18	2.04	2.15	2.11	2.19	2.42	1.74
	$\widehat{\text{std}}$	2.89	2.27	2.13	2.04	2.05	2.13	2.27	2.44	1.80
	cvp	94.90	96.60	94.50	94.80	93.50	95.60	95.70	94.90	95.60
$\mathbf{W}$		$\rho = 0.8, n = 200$								
$\mathbf{I}$	bias	0.38	0.21	0.18	0.12	-0.10	-0.14	-0.19	-0.38	-0.03
	std	6.28	4.95	4.52	4.60	4.41	4.48	4.92	5.29	3.95
	$\widehat{\text{std}}$	6.59	5.18	4.86	4.69	4.69	4.89	5.18	5.60	4.13
	cvp	96.40	96.90	96.20	95.50	96.10	96.90	96.80	96.80	96.50
$\widehat{\mathbf{\Sigma}}^{-1}$	bias	-0.10	0.00	0.10	-0.09	-0.01	0.05	0.06	-0.06	0.06
	std	3.51	2.90	2.61	2.66	2.66	2.67	2.89	3.04	2.31
	$\widehat{\text{std}}$	3.58	2.83	2.67	2.58	2.58	2.68	2.83	3.07	2.27
	cvp	95.70	95.20	96.40	93.90	94.20	94.40	95.10	95.20	94.30
$\mathbf{W}$		$\rho = 0.8, n = 500$								
$\mathbf{I}$	bias	0.42	0.19	0.03	0.06	-0.03	-0.22	-0.12	-0.16	-0.07
	std	3.96	3.18	3.05	2.96	2.93	2.95	3.07	3.24	2.39
	$\widehat{\text{std}}$	4.11	3.20	3.00	2.89	2.89	3.01	3.22	3.46	2.54
	cvp	96.20	95.50	94.70	94.50	95.50	95.30	95.80	96.50	96.20
$\widehat{\mathbf{\Sigma}}^{-1}$	bias	-0.01	0.05	-0.05	0.02	-0.03	-0.01	-0.07	0.01	-0.01
	std	2.23	1.67	1.60	1.55	1.64	1.62	1.73	1.88	1.38
	$\widehat{\text{std}}$	2.24	1.76	1.65	1.59	1.59	1.65	1.76	1.90	1.40
	cvp	95.00	95.90	96.10	95.60	94.10	95.20	95.50	95.00	94.70

Table 2: Simulation results for Model II: the average bias of the estimators ("bias"), the Monte Carlo standard deviation ("std"), the average of the estimated standard deviation ("$\widehat{\text{std}}$") based on the theoretical calculation, and the empirical coverage probability ("cvp") at the nominal 95% confidence level. All simulation results reported below are multiplied by 100.

		$\widehat{\beta}_2$	$\widehat{\beta}_3$	$\widehat{\beta}_4$	$\widehat{\beta}_5$	$\widehat{\beta}_6$	$\widehat{\beta}_7$	$\widehat{\beta}_8$	$\widehat{\beta}_9$	$\widehat{\beta}_{10}$
True value		0.8	0.6	0.4	0.2	-0.2	-0.4	-0.6	-0.8	0
$\mathbf{W}$		$\rho = 0.5, n = 200$								
$\mathbf{I}$	bias	2.18	0.87	0.49	0.32	-0.31	-0.84	-1.14	-0.81	-0.12
	std	11.61	9.69	8.81	8.66	8.85	9.10	9.90	10.30	7.84
	$\widehat{\text{std}}$	12.62	9.89	9.34	8.96	8.94	9.34	9.94	10.66	7.90
	cvp	95.90	95.20	95.90	96.10	94.60	95.00	94.60	95.50	95.50
$\widehat{\mathbf{\Sigma}}^{-1}$	bias	1.59	0.50	0.06	0.19	-0.15	-0.44	-0.83	-0.65	0.01
	std	6.79	5.67	5.26	4.98	5.21	5.33	5.65	5.97	4.52
	$\widehat{\text{std}}$	7.10	5.56	5.24	5.04	5.04	5.26	5.59	6.00	4.44
	cvp	95.30	95.30	94.70	95.90	94.30	93.80	95.50	95.30	95.60
$\mathbf{W}$		$\rho = 0.5, n = 500$								
$\mathbf{I}$	bias	1.58	0.47	0.50	0.11	-0.08	-0.69	-0.24	-0.89	-0.21
	std	7.54	6.23	5.71	5.57	5.52	5.84	6.18	6.52	4.96
	$\widehat{\text{std}}$	8.00	6.24	5.87	5.62	5.63	5.87	6.24	6.74	4.96
	cvp	96.40	95.50	95.90	95.30	94.70	94.30	95.20	95.90	94.70
$\widehat{\mathbf{\Sigma}}^{-1}$	bias	1.40	0.57	0.34	0.30	-0.20	-0.48	-0.58	-0.70	-0.05
	std	3.77	3.26	3.08	3.02	3.03	3.09	3.27	3.43	2.68
	$\widehat{\text{std}}$	4.29	3.34	3.14	3.01	3.02	3.14	3.35	3.61	2.66
	cvp	96.50	95.20	94.80	93.70	94.40	94.40	94.50	95.90	94.90
$\mathbf{W}$		$\rho = 0.8, n = 200$								
$\mathbf{I}$	bias	2.19	0.80	0.44	0.15	-0.40	-0.86	-0.91	-0.76	-0.25
	std	12.47	10.18	9.42	9.17	9.31	9.57	10.46	10.48	8.08
	$\widehat{\text{std}}$	13.25	10.38	9.80	9.39	9.37	9.80	10.43	11.19	8.29
	cvp	96.10	95.00	96.20	95.80	95.30	95.90	95.10	96.20	95.40
$\widehat{\mathbf{\Sigma}}^{-1}$	bias	1.53	0.45	0.21	0.22	-0.21	-0.37	-0.84	-0.61	-0.07
	std	5.69	4.80	4.34	4.20	4.22	4.44	4.67	5.01	3.77
	$\widehat{\text{std}}$	5.88	4.61	4.35	4.18	4.18	4.35	4.64	4.97	3.68
	cvp	95.00	95.90	95.00	95.20	94.90	93.60	94.20	95.60	94.10
$\mathbf{W}$		$\rho = 0.8, n = 500$								
$\mathbf{I}$	bias	1.62	0.49	0.43	0.04	-0.11	-0.68	-0.24	-0.74	-0.27
	std	7.82	6.49	6.00	6.04	5.91	6.13	6.41	6.63	5.06
	$\widehat{\text{std}}$	8.38	6.53	6.13	5.88	5.89	6.14	6.53	7.05	5.18
	cvp	96.00	94.90	96.00	94.40	94.80	95.10	95.10	96.30	95.40
$\widehat{\mathbf{\Sigma}}^{-1}$	bias	1.36	0.53	0.35	0.25	-0.17	-0.43	-0.57	-0.68	-0.03
	std	3.01	2.56	2.40	2.38	2.39	2.42	2.56	2.70	2.18
	$\widehat{\text{std}}$	3.48	2.71	2.54	2.44	2.44	2.55	2.71	2.93	2.15
	cvp	95.70	94.90	95.70	94.40	94.50	95.10	93.90	95.80	93.80

Table 3: Simulation results for Model III: the average bias of the estimators ("bias"), the Monte Carlo standard deviation ("std"), the average of the estimated standard deviation ("$\widehat{\text{std}}$") based on the theoretical calculation, and the empirical coverage probability ("cvp") at the nominal 95% confidence level. All simulation results reported below are multiplied by 100.

		$\widehat{\beta}_{13}$	$\widehat{\beta}_{14}$	$\widehat{\beta}_{15}$	$\widehat{\beta}_{16}$	$\widehat{\beta}_{17}$	$\widehat{\beta}_{23}$	$\widehat{\beta}_{24}$	$\widehat{\beta}_{25}$	$\widehat{\beta}_{26}$	$\widehat{\beta}_{27}$
True value		0.8	-0.6	0.4	-0.2	0	-0.8	0.6	-0.4	0.2	0
$\mathbf{W}$		$\rho = 0.5, n = 200$									
	bias	2.11	-1.33	0.80	-0.33	-0.12	-0.85	0.93	-0.69	0.46	-0.19
$\mathbf{I}$	std	5.96	5.86	5.30	4.95	4.14	5.07	4.92	4.41	3.93	3.34
	$\widehat{\text{std}}$	7.38	7.25	6.28	5.58	4.79	5.99	5.92	5.11	4.54	3.91
	cvp	98.40	98.20	98.20	97.40	97.60	97.50	97.50	97.00	97.20	97.10
	bias	1.21	-0.92	0.56	-0.22	-0.09	-0.86	0.81	-0.52	0.30	-0.12
$\widehat{\mathbf{\Sigma}}^{-1}$	std	3.91	3.77	3.28	3.03	2.50	3.22	3.23	2.81	2.46	1.99
	$\widehat{\text{std}}$	4.31	4.24	3.68	3.26	2.80	3.52	3.48	3.01	2.67	2.30
	cvp	97.30	97.10	96.70	96.40	96.60	95.70	95.90	96.00	96.50	97.90
$\mathbf{W}$		$\rho = 0.5, n = 500$									
	bias	1.62	-1.23	0.90	-0.37	-0.04	-0.73	0.89	-0.43	0.25	0.00
$\mathbf{I}$	std	4.27	4.27	3.68	3.22	2.93	3.59	3.47	3.09	2.76	2.40
	$\widehat{\text{std}}$	5.41	5.31	4.56	4.04	3.46	4.45	4.38	3.76	3.33	2.86
	cvp	98.80	98.00	97.90	98.60	97.60	98.40	98.50	97.80	98.30	97.80
	bias	0.88	-0.69	0.45	-0.24	0.02	-0.65	0.71	-0.47	0.23	0.02
$\widehat{\mathbf{\Sigma}}^{-1}$	std	2.72	2.77	2.43	2.05	1.91	2.32	2.17	1.88	1.71	1.39
	$\widehat{\text{std}}$	3.11	3.05	2.62	2.33	1.99	2.57	2.52	2.17	1.92	1.65
	cvp	96.90	96.90	96.40	96.90	96.10	96.80	97.00	97.50	97.90	97.90
$\mathbf{W}$		$\rho = 0.8, n = 200$									
	bias	2.13	-1.42	0.82	-0.38	-0.01	-0.73	0.93	-0.79	0.44	-0.23
$\mathbf{I}$	std	6.51	6.56	5.80	5.23	4.52	5.78	5.61	4.93	4.27	3.78
	$\widehat{\text{std}}$	8.07	7.93	6.89	6.11	5.24	6.56	6.46	5.57	4.96	4.27
	cvp	98.30	98.00	98.60	98.30	98.20	96.80	97.30	96.60	96.50	96.30
	bias	0.89	-0.64	0.39	-0.13	-0.06	-0.52	0.46	-0.30	0.17	-0.08
$\widehat{\mathbf{\Sigma}}^{-1}$	std	3.27	3.21	2.74	2.57	2.11	2.41	2.42	2.11	1.85	1.51
	$\widehat{\text{std}}$	3.60	3.55	3.07	2.73	2.34	2.61	2.59	2.23	1.98	1.71
	cvp	97.60	96.50	97.10	95.80	96.80	95.90	95.90	96.00	96.90	97.60
$\mathbf{W}$		$\rho = 0.8, n = 500$									
	bias	1.66	-1.25	0.88	-0.32	0.04	-0.79	0.91	-0.42	0.31	-0.07
$\mathbf{I}$	std	4.56	4.65	4.07	3.62	3.22	3.89	3.88	3.40	3.01	2.62
	$\widehat{\text{std}}$	5.95	5.83	5.02	4.44	3.80	4.87	4.79	4.12	3.66	3.13
	cvp	98.70	97.90	98.30	97.60	97.20	99.20	98.00	98.10	98.60	97.90
	bias	0.57	-0.39	0.26	-0.17	0.02	-0.35	0.39	-0.24	0.12	0.00
$\widehat{\mathbf{\Sigma}}^{-1}$	std	2.33	2.31	2.01	1.73	1.57	1.69	1.57	1.39	1.27	1.03
	$\widehat{\text{std}}$	2.55	2.51	2.15	1.91	1.63	1.85	1.82	1.56	1.39	1.19
	cvp	97.10	96.60	96.30	96.10	95.70	97.10	98.30	96.90	97.10	97.60

Table 4: Simulation results for Model IV: the average bias of the estimators ("bias"), the Monte Carlo standard deviation ("std"), the average of the estimated standard deviation ("$\widehat{\text{std}}$") based on the theoretical calculation, and the empirical coverage probability ("cvp") at the nominal 95% confidence level. All simulation results reported below are multiplied by 100.

		$\widehat{\beta}_{13}$	$\widehat{\beta}_{14}$	$\widehat{\beta}_{15}$	$\widehat{\beta}_{16}$	$\widehat{\beta}_{17}$	$\widehat{\beta}_{23}$	$\widehat{\beta}_{24}$	$\widehat{\beta}_{25}$	$\widehat{\beta}_{26}$	$\widehat{\beta}_{27}$
True value		0.8	-0.6	0.4	-0.2	0	-0.8	0.6	-0.4	0.2	0
W		$\rho = 0.5, n = 200$									
I	bias	-0.33	0.47	-0.29	0.22	0.00	-0.21	0.54	-0.30	0.15	0.06
	std	4.32	4.12	3.64	3.29	2.89	5.73	5.63	4.86	4.34	3.68
	$\widehat{\text{std}}$	5.10	5.04	4.34	3.90	3.34	6.54	6.46	5.58	4.98	4.27
	cvp	97.80	98.00	97.30	97.70	98.00	97.20	97.70	96.90	96.40	96.30
$\widehat{\mathbf{\Sigma}}^{-1}$	bias	0.02	0.11	-0.02	0.11	-0.06	-0.24	0.43	-0.22	0.08	0.07
	std	3.06	2.98	2.64	2.28	2.04	3.33	3.31	2.86	2.63	2.18
	$\widehat{\text{std}}$	3.66	3.61	3.12	2.79	2.39	3.89	3.84	3.31	2.94	2.53
	cvp	97.90	97.40	97.30	97.90	97.90	97.30	97.20	97.00	97.50	98.00
W		$\rho = 0.5, n = 500$									
I	bias	-0.52	0.62	-0.28	0.14	-0.08	-0.31	0.57	-0.16	0.12	-0.02
	std	2.91	2.90	2.55	2.16	1.96	4.25	4.15	3.53	3.03	2.67
	$\widehat{\text{std}}$	3.46	3.41	2.95	2.63	2.24	4.88	4.82	4.15	3.70	3.16
	cvp	97.50	97.20	97.60	98.30	96.50	97.00	97.60	98.10	98.10	97.60
$\widehat{\mathbf{\Sigma}}^{-1}$	bias	-0.28	0.37	-0.18	0.09	-0.05	-0.31	0.49	-0.31	0.11	0.05
	std	2.06	1.99	1.69	1.51	1.36	2.44	2.30	1.98	1.83	1.60
	$\widehat{\text{std}}$	2.44	2.40	2.07	1.84	1.57	2.79	2.74	2.36	2.10	1.79
	cvp	97.80	97.80	98.60	98.30	97.00	97.40	97.30	98.80	97.50	97.60
W		$\rho = 0.8, n = 200$									
I	bias	-0.33	0.45	-0.23	0.19	-0.05	-0.20	0.62	-0.22	0.20	-0.03
	std	4.54	4.43	3.92	3.58	3.02	6.51	6.05	5.48	4.85	4.27
	$\widehat{\text{std}}$	5.28	5.22	4.54	4.09	3.50	7.11	7.02	6.09	5.47	4.71
	cvp	97.40	97.60	97.60	97.70	97.40	96.50	97.50	96.00	96.90	96.70
$\widehat{\mathbf{\Sigma}}^{-1}$	bias	-0.31	0.31	-0.18	0.15	-0.04	0.05	0.10	-0.02	-0.02	0.08
	std	2.40	2.37	2.11	1.90	1.59	2.91	2.89	2.55	2.21	1.81
	$\widehat{\text{std}}$	2.84	2.81	2.45	2.20	1.88	3.22	3.18	2.77	2.48	2.13
	cvp	97.80	97.60	96.80	98.30	98.00	97.20	96.60	97.00	97.80	97.90
W		$\rho = 0.8, n = 500$									
I	bias	-0.40	0.50	-0.30	0.11	0.02	-0.40	0.50	-0.30	0.11	0.02
	std	2.98	2.99	2.70	2.34	2.03	4.44	4.50	3.88	3.41	2.79
	$\widehat{\text{std}}$	3.64	3.58	3.10	2.76	2.36	5.36	5.28	4.55	4.07	3.48
	cvp	98.20	98.00	97.60	98.20	97.20	97.80	98.10	98.20	98.40	98.10
$\widehat{\mathbf{\Sigma}}^{-1}$	bias	-0.35	0.34	-0.19	0.13	-0.03	-0.35	0.34	-0.19	0.13	-0.03
	std	1.60	1.53	1.32	1.17	1.03	2.00	1.87	1.61	1.53	1.31
	$\widehat{\text{std}}$	1.86	1.83	1.58	1.41	1.20	2.27	2.23	1.92	1.71	1.46
	cvp	97.10	97.20	97.80	98.00	97.00	97.30	97.40	98.20	96.90	97.70

Table 5: The frequency (%) of the estimated structural dimension $\widehat{d}$.

Model	$\widehat{d}=1$	$\widehat{d}=2$	$\widehat{d}\geq 3$	$\widehat{d}=1$	$\widehat{d}=2$	$\widehat{d}\geq 3$
	$\rho=0.5, n=200$			$\rho=0.8, n=200$		
I	95.90	4.10	0.00	95.30	4.70	0.00
II	87.20	12.80	0.00	84.90	15.10	0.00
III	0.90	99.10	0.00	0.50	99.50	0.00
IV	0.00	100.00	0.00	0.00	100.00	0.00
	$\rho=0.5, n=500$			$\rho=0.8, n=500$		
I	98.80	1.20	0.00	98.80	1.20	0.00
II	94.90	5.10	0.00	95.30	4.70	0.00
III	0.20	99.80	0.00	0.40	99.60	0.00
IV	0.00	100.00	0.00	0.00	100.00	0.00

denoted by $\widehat{\boldsymbol{\beta}}$, obtained through existing methods so that the first entry of $\widehat{\boldsymbol{\beta}}$, denoted by $\widehat{\beta}_1$, is one. We repeat each scenario 1000 times and report the biases and the Monte Carlo standard deviations of $(\widehat{\boldsymbol{\beta}}/\widehat{\beta}_1)$ in Table 6. It can be seen that the performance of all methods improves when the sample size n is increased from 200 to 500. In both scenarios, all these methods perform comparatively, although our proposed NEW estimate has slightly smaller biases and standard deviations.

3.3. Application to Framingham Heart Study

In this section we revisit the Framingham Heart Study described in Section 1. Let $\mathbf{y}=(Y_1,Y_2)$, $\mathbf{z}=(1,Z_1,Z_2,Z_3)^{\mathrm{T}}$, $\mathbf{x}=(X_1,\cdots,X_7)^{\mathrm{T}}$ in model (1.1). We add a column of ones in $\mathbf{z}$ to include an intercept in model (1.1). The BIC-type criterion yields that $\widehat{d}=2$. Both the unweighted and the weighted profile least squares estimates, along with their standard deviations and the p-values are given in Table 7. It can be clearly seen that the weighted profile least squares estimates have smaller standard deviations than those of the unweighted ones. In effect, $\mathrm{corr}(Y_1,Y_2)=0.4159$ and the p-value is less than 10^{-4} in the test for significance of Pearson's correlation coefficient. That is, the systolic and diastolic blood pressures are highly correlated. It is thus not surprising to see that the weighted profile least squares estimates are significantly more efficient

Table 6: Simulated results for Model V when $\mathbf{z} \sim$Bernoulli(0.5) and $\mathbf{z} \sim \mathcal{N}(0, 0.5)$, respectively: the average bias ("bias") and the Monte Carlo standard deviation ("std") of $(\widehat{\boldsymbol{\beta}}/\widehat{\beta}_1)$. All simulation results reported below are multiplied by 100.

				$\widehat{\beta}_2$	$\widehat{\beta}_3$	$\widehat{\beta}_4$	$\widehat{\beta}_5$	$\widehat{\beta}_6$	$\widehat{\beta}_7$	$\widehat{\beta}_8$	$\widehat{\beta}_9$	$\widehat{\beta}_{10}$
method	n	$\mathbf{z}$		0.8	0.6	0.4	0.2	-0.2	-0.4	-0.6	-0.8	0
NEW	200	Bernoulli	bias	0.05	0.04	0.16	0.09	-0.20	0.08	-0.21	-0.12	-0.05
			std	6.97	5.75	5.08	4.94	4.93	5.19	5.61	5.88	4.32
		Normal	bias	0.01	0.30	-0.15	0.09	-0.06	-0.01	-0.23	-0.02	-0.07
			std	6.69	5.10	5.10	4.95	4.95	5.09	5.37	5.59	4.17
	500	Bernoulli	bias	0.44	0.10	0.25	0.07	-0.08	-0.21	-0.08	-0.39	-0.02
			std	4.37	3.41	3.27	3.15	3.17	3.28	3.48	3.72	2.80
		Normal	bias	0.30	0.32	-0.07	0.18	0.05	-0.20	-0.12	-0.21	0
			std	4.11	3.25	3.19	2.95	3.04	3.18	3.33	3.57	2.71
LCC	200	Bernoulli	bias	0.36	0.17	-0.36	0.02	0.17	-0.36	-0.16	-0.39	-0.03
			std	8.18	6.15	5.85	5.72	5.60	5.94	6.39	6.76	4.79
		Normal	bias	0.57	0.33	0.06	0.01	0.08	-0.16	-0.05	-0.48	0.15
			std	7.96	6.04	5.62	5.24	5.37	5.49	5.99	6.33	4.53
	500	Bernoulli	bias	0.15	-0.11	0.24	0.05	-0.17	0.00	-0.03	0.03	0.07
			std	4.95	3.66	3.45	3.24	3.49	3.44	3.77	3.89	2.94
		Normal	bias	-0.03	0.05	-0.05	-0.06	0.05	-0.00	0.11	-0.13	0.12
			std	4.64	3.51	3.20	3.11	3.21	3.37	3.50	3.90	2.78
MS	200	Bernoulli	bias	0.26	0.32	0.11	0.15	0.06	-0.02	-0.11	-0.24	-0.12
			std	9.35	6.69	6.53	6.01	5.98	6.22	6.75	7.54	5.34
		Normal	bias	0.38	0.24	0.18	0.06	0.07	-0.17	-0.33	0.04	-0.13
			std	8.27	6.11	5.92	5.57	5.61	5.74	6.35	6.83	5.06
	500	Bernoulli	bias	0.21	0.09	0.05	0.14	-0.10	-0.21	0.15	-0.15	-0.09
			std	5.06	3.85	3.60	3.37	3.48	3.70	4.02	4.27	3.14
		Normal	bias	0.02	-0.04	0.03	-0.03	0.04	-0.08	-0.05	0.05	0.04
			std	4.61	3.63	3.45	3.35	3.32	3.53	3.55	4.03	2.88
LCL	200	Bernoulli	bias	0.65	0.34	-0.02	0.17	0.09	-0.28	-0.51	-0.45	6.43
			std	8.37	6.10	5.91	5.58	5.40	5.89	6.37	7.05	12.22
		Normal	bias	0.65	0.63	0.11	0.06	-0.08	-0.29	-0.13	-0.57	3.69
			std	7.89	6.16	5.84	5.61	5.31	5.65	6.13	6.73	9.41
	500	Bernoulli	bias	0.36	0.05	0.08	0.11	0.06	-0.23	-0.26	0.19	0.39
			std	4.69	3.57	3.17	3.05	3.22	3.26	3.69	3.75	2.42
		Normal	bias	0.22	0.18	-0.14	-0.09	0.12	-0.12	-0.13	-0.07	0.32
			std	4.67	3.50	3.21	3.20	3.26	3.38	3.65	3.94	2.21

than the unweighted ones. For $k = 3, \cdots, 7$, at least one p-value of X_k is significant at the significance level 0.05, indicating that the interactions between $\mathbf{x}$ and $\mathbf{z}$ are all significant. Therefore, we can conclude that healthy daily life styles including a moderate amount of physical exercises helps to control for the blood pressures. To show the interactions between $\mathbf{x}$ and $\mathbf{z}$ graphically, the estimated surfaces $\widehat{m}_{ij}(\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{x})$ of

$m_{ij}(\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{x})$ with $\widehat{\boldsymbol{\beta}} = (\widehat{\boldsymbol{\beta}}_1, \widehat{\boldsymbol{\beta}}_2)^{\mathrm{T}}$, for $i = 1, 2$ and $j = 2, 3, 4$ are shown in Figure 2, which clearly reveals the nonlinear modulating effect of the degree of obesity on physical exercises. Such dynamic effects are helpful to design a moderate amount of physical exercises to control for the blood pressures.

Table 7: Both the unweighted and the weighted profile least squares estimates, along with the standard errors and the p-values.

W		X_3		X_4		X_5		X_6		X_7	
		$\boldsymbol{\beta}_1$	$\boldsymbol{\beta}_2$	$\boldsymbol{\beta}_1$	$\boldsymbol{\beta}_2$	$\boldsymbol{\beta}_1$	$\boldsymbol{\beta}_2$	$\boldsymbol{\beta}_1$	$\boldsymbol{\beta}_2$	$\boldsymbol{\beta}_1$	$\boldsymbol{\beta}_2$
I	coef	0.4285	0.5149	1.0917	0.4939	0.9340	0.5008	0.5007	0.5024	0.2376	0.4919
	std	0.3805	0.0542	0.6388	0.0925	0.2754	0.0374	0.3115	0.0470	0.3343	0.0485
	p-value	0.2602	0.0000	0.0875	0.0000	0.0007	0.0000	0.1079	0.0000	0.4773	0.0000
$\widehat{\boldsymbol{\Sigma}}^{-1}$	coef	0.3811	0.5215	0.8881	0.5140	1.0031	0.5090	0.4575	0.4732	0.4355	0.4988
	std	0.3119	0.0440	0.5384	0.0741	0.2431	0.0306	0.2615	0.0382	0.2819	0.0402
	p-value	0.2217	0.0000	0.0990	0.0000	0.0000	0.0000	0.0802	0.0000	0.1224	0.0000

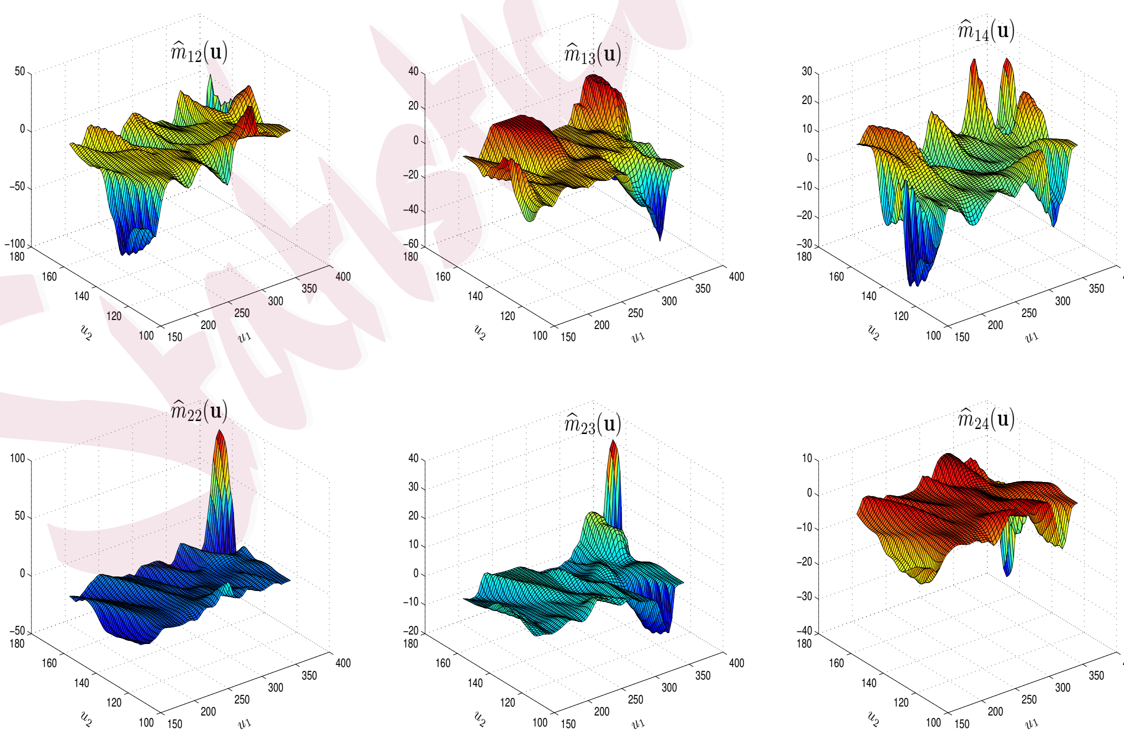

Figure 2: The estimated surfaces $\widehat{m}_{ij}(\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{x})$ of $m_{ij}(\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{x})$ with $\widehat{\boldsymbol{\beta}} = (\widehat{\boldsymbol{\beta}}_1, \widehat{\boldsymbol{\beta}}_2)$, for $i = 1, 2$ and $j = 2, 3, 4$.

4. CONCLUDING REMARKS

In this paper we introduce a new class of multivariate adaptive varying index models which encompasses many existing semiparametric models and accommodate nonlinear interactions between the covariates. We develop a profile least squares estimation procedure to estimate the dimension reduction space which is computationally efficient and conceptually simple. The resultant estimate is root-n consistent and asymptotically normal. The asymptotic properties of the dimension reduction space are useful for making inference on the interaction effects. We also suggest a BIC-type criterion to decide the structural dimension. Its consistency is also established.

There are three ancillary covariates $\mathbf{z} = (Z_1, Z_2, Z_3)^{\mathrm{T}}$ in our motivating example. These ancillary covariates are weakly correlated in that $\mathrm{corr}(Z_1, Z_2) = -0.017$, $\mathrm{corr}(Z_1, Z_3) = -0.060$ and $\mathrm{corr}(Z_2, Z_3) = 0.101$, with p-values being 0.770, 0.297 and 0.079. Their correlations are not significant, thus we do not consider the interactions among these ancillary covariates in the present work. In addition, in model (1.1) we assume implicitly that the effects of the ancillary covariates are additive. We remark here that both our proposed methodology and the theoretical results are applicable when the ancillary covariates $\mathbf{z}$ are moderately correlated. If the ancillary covariates $\mathbf{z}$ are highly correlated, it is recommended to consider the interactions among $\mathbf{z}$ as well. This however leads to quite different model structures. Accordingly, new algorithms and estimation procedures are needed. Future research along this line is warranted.

REFERENCES

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(438)**, 477–489.

Cook, R.D. (1998). *Regression Graphics.* New York: Wiley.

Cook, R.D. and Li, B. (2002). "Dimension reduction for conditional mean in regression." *Annals of Statistics*, **30(2)**, 455–474.

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

Feng, Z., Wen, X.M., Yu, Z., and Zhu, L. (2013). "On partial sufficient dimension reduction with applications to partially linear multi-index models." *Journal of the American Statistical Association*, **108(501)**, 237–246.

Hilafu, H. and Wu, W. (2017). "Partial projective resampling method for dimension reduction: With applications to partially linear models." *Computational Statistics & Data Analysis*, **109**, 1–14.

Ichimura, H. (1993). "Semiparametric least squares (sls) and weighted sls estimation of single-index models." *Journal of Econometrics*, **58(1-2)**, 71–120.

Li, B., Cook, R.D., and Chiaromonte, F. (2003). "Dimension reduction for the conditional mean in regressions with categorical predictors." *Annals of Statistics*, **31(5)**, 1636–1668.

Liu, X., Cui, Y., and Li, R. (2016). "Partial linear varying multi-index coefficient model for integrative gene-environment interactions." *Statistica Sinica*, **26**, 1037–1060.

Ma, S. and Song, P.X.K. (2015). "Varying index coefficient models." *Journal of the American Statistical Association*, **110(509)**, 341–356.

Ma, S., Yang, L., Romero, R., and Cui, Y. (2011). "Varying coefficient model for gene–environment interaction: a non-linear look." *Bioinformatics*, **27(15)**, 2119–2126.

Ma, Y. and Zhu, L. (2012). "A semiparametric approach to dimension reduction." *Journal of the American Statistical Association*, **107(497)**, 168–179.

Ma, Y. and Zhu, L. (2014). "On estimation efficiency of the central mean subspace." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **76(5)**, 885–901.

Ma, Y. and Zhu, L.P. (2013). "Efficiency loss caused by linearity condition in dimension reduction." *Biometrika*, **100(2)**, 371.

Xu, K., Guo, W., Xiong, M., Zhu, L., and Jin, L. (2016). "An estimating equation approach to dimension reduction for longitudinal data." *Biometrika*, **103(1)**, 189–203.

Zhu, L., Miao, B., and Peng, H. (2006). "On sliced inverse regression with high-dimensional covariates." *Journal of the American Statistical Association*, **101(474)**, 630–643.

Zhu, L. and Zhong, W. (2015). "Estimation and inference on central mean subspace for multivariate response data." *Computational Statistics & Data Analysis*, **92**, 68–83.