Communications for Statistical Applications and Methods
2015, Vol. 22, No. 3, 305–312

DOI: http://dx.doi.org/10.5351/CSAM.2015.22.3.305
Print ISSN 2287-7843 / Online ISSN 2383-4757

Generalized Partially Double-Index Model: Bootstrapping and Distinguishing Values

Jae Keun Yoo[1,a]

[a]Department of Statistics, Ewha Womans University, Korea

Abstract

We extend a generalized partially linear single-index model and newly define a generalized partially double-index model (GPDIM). The philosophy of sufficient dimension reduction is adopted in GPDIM to estimate unknown coefficient vectors in the model. Subsequently, various combinations of popular sufficient dimension reduction methods are constructed with the best combination among many candidates determined through a bootstrapping procedure that measures distances between subspaces. Distinguishing values are newly defined to match the estimates to the corresponding population coefficient vectors. One of the strengths of the proposed model is that it can investigate the appropriateness of GPDIM over a single-index model. Various numerical studies confirm the proposed approach, and real data application are presented for illustration purposes.

Keywords: bootstrapping, central subspace, distinguishing value, generalized partially linear single-index model, kernel matrix, regression, sufficient dimension reduction

1. Introduction

A regression is a study of the conditional distribution of a response Y given a set of predictors $\mathbf{X} = (X_1, \ldots, X_p)^{\mathrm{T}}$. Predictors are separated by two sets in regression modeling and the two sets have different relationships with the response. For example, one set of predictors is continuous and the other is categorical. Suppose that the set of continuous variables has non-linear relationship to the response, while the other is used for offset and is assumed to have a linear relation. Then, usual ordinary least square is not an ideal way to estimate unknown regression coefficients, because the non-linear relationship is known. This issue that two sets of predictors have different relationships with the response was addressed in Carroll *et al.* (1997). In the work, the following model was assumed for the two sets of predictors $\mathbf{X} = (\mathbf{U} \in \mathbb{R}^{p_u} = (X_1, \ldots, X_{p_u}), \mathbf{W} \in \mathbb{R}^{p_w} = (X_{p_u+1}, \ldots, X_p))^{\mathrm{T}}$ with $p = p_u + p_w$:

$$Y = f\left(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{U}\right) + \boldsymbol{\beta}^{\mathrm{T}}\mathbf{W} + \varepsilon, \tag{1.1}$$

where $f(\cdot)$ is an unknown function, $\boldsymbol{\alpha} \in \mathbb{R}^{p_u}, \boldsymbol{\beta} \in \mathbb{R}^{p_w}$, and ε is assumed to be a random error with zero mean and variance σ^2, which is independent of $(\mathbf{U}, \mathbf{W})$. The model in (1.1) is called textitgeneralized partially linear single-index model (GPLSIM).

This work was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Korean Ministry of Education (NRF-2014R1A2A1A11049389).

[1] Department of Statistics, Ewha Womans University, 11-1 Daehyun-dong Seodaemun-gu, Seoul 120-750, Korea.
E-mail: peter.yoo@ewha.ac.kr

Published 31 May 2015 / journal homepage: http://csam.or.kr

If f is the identity function, GPLSIM is reduced to linear regression. GPLSIM is equivalent to a single-index model if the predictors are not separable and f should be considered in the study. The single-index model generalizes linear regression and GPLSIM generalizes the single-index model; subsequently, GPLSIM forms a large class of regression models.

One limit in practice of GPLSIM is placed onto partial linear relation of one set of predictors. GPLSIM is not directly applicable if the relationship is expected to be non-linear.

The purpose of the article is placed on defining a class of regression models to generalize GPLSIM, dropping "partially linear", and on estimating $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ without prior knowledge on the true function relationships between two sets of predictors and the response. The generalized partially single-index model has been successfully used in many science fields; in addition, it is believed that GPDIM is beneficiary to progress in relevant fields of sciences because GPDIM generalizes. Another strength of the proposed approach is placed on investigating the appropriateness of GPDIM over a single-index model. Additional functional complexity can be avoided in the modeling procedure if the single-index model is chosen over GPDIM.

The organization of this article is as follows. Section 2 is devoted to defining a generalized class of GPLSIM and developing a methodology to estimate unknown coefficient vectors. Section 3 presents numerical studies and real data analysis. We summarize our work in Section 4.

2. Generalized Partially Double-Index Model

2.1. Generalized partially double-indexed model

To generalize GPLSIM in (1.1), we assume the following regression model with $\mathbf{X} = (\mathbf{U} \in \mathbb{R}^{p_u}, \mathbf{W} \in \mathbb{R}^{p_w})^{\mathrm{T}}$:

$$Y = g\left(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{U}, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{W}, \varepsilon\right). \tag{2.1}$$

The model in (2.1) is equivalent to GPLSIM, if it is assumed that

$$g\left(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{U}, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{W}, \varepsilon\right) = f\left(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{U}\right) + \boldsymbol{\beta}^{\mathrm{T}}\mathbf{W} + \varepsilon.$$

Therefore, the model in (2.1) generalizes GPLSIM, and it will be called a *generalized partially double-index model* (GPDIM). In GPDIM, the word of "partially" is used, because the regression coefficients are related with partial sets of predictors.

In GPDIM, the response of Y depends on $\mathbf{X}$ only through two linear combinations of $\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{U}$ and $\boldsymbol{\beta}^{\mathrm{T}}\mathbf{W}$. That is, the conditional relationship of Y given $\mathbf{X}$ is fully characterized by $\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{U}$ and $\boldsymbol{\beta}^{\mathrm{T}}\mathbf{W}$. Therefore, the primary interest in the GPDIM is placed on the estimation of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ without information on $g(\cdot)$. This notion is closely related with sufficient dimension reduction, which will be briefly discussed in the next subsection. Combining sufficient dimension reduction methods, a bootstrapping methodology to estimate $\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{U}$ and $\boldsymbol{\beta}^{\mathrm{T}}\mathbf{W}$ is proposed in the later section.

2.2. Sufficient dimension reduction

Sufficient dimension reduction (SDR) in regression of $Y|\mathbf{X} \in \mathbb{R}^p$ pursues to replace the original p-dimensional predictors $\mathbf{X}$ by a lower-dimensional linear projection predictor $\boldsymbol{\eta}^{\mathrm{T}}\mathbf{X}$ without a loss of information about the conditional distribution of $Y|\mathbf{X}$ such that

$$Y \perp\!\!\!\perp \mathbf{X}|\boldsymbol{\eta}^{\mathrm{T}}\mathbf{X}, \tag{2.2}$$

where $\boldsymbol{\eta} \in \mathbb{R}^{p\times q}$ with $q \leq p$ and $\perp\!\!\!\perp$ stands for independence.

Statement (2.2) directly implies that the conditional distributions of $Y|\mathbf{X}$ and $Y|\boldsymbol{\eta}^{\mathrm{T}}\mathbf{X}$ are equal, and the dimension reduction of $\mathbf{X}$ through $\boldsymbol{\eta}^{\mathrm{T}}\mathbf{X}$ is achieved. The minimal subspace spanned by the columns of $\boldsymbol{\eta}$ to satisfy (2.2) is called the *central subspace*, $\mathcal{S}_{Y|\mathbf{X}}$. The existence of $\mathcal{S}_{Y|\mathbf{X}}$ is guaranteed under various conditions such as the open and convex support of $\mathbf{X}$. Hereafter an true orthonormal basis matrix of $\mathcal{S}_{Y|\mathbf{X}}$ will be denoted as $\boldsymbol{\eta}$.

The four methods to estimate $\mathcal{S}_{Y|\mathbf{X}}$, which have been popularly used up to date among others, are: (i) Ordinary least squares (OLS; Cook, 1998; Yin and Cook, 2002); (ii) sliced inverse regression (SIR; Li, 1991); (iii) sliced average variance estimation (SAVE; Cook and Weisberg, 1991); (iv) principal Hessian directions (pHd; Li, 1992). Each method produce its own kernel matrix $\mathbf{M}_{\bullet}$, whose column spans $\mathcal{S}_{Y|\mathbf{X}}$.

2.3. Estimation: bootstrapping and distinguishing values

Considering interpretation of GPDIM in SDR context, $\mathcal{S}_{Y|\mathbf{X}}$ is spanned by the columns of

$$\boldsymbol{\eta} \in \mathbb{R}^{p\times 2} = \left\{\boldsymbol{\alpha}_0 = (\boldsymbol{\alpha}, \mathbf{0})^{\mathrm{T}}, \boldsymbol{\beta}_0 = (\mathbf{0}, \boldsymbol{\beta})^{\mathrm{T}}\right\}.$$

Therefore, the true dimension of $\mathcal{S}_{Y|\mathbf{X}}$ under GPDIM is equal to two.

Popular SDR methods introduced in the previous subsection are supposed to estimate $\boldsymbol{\alpha}_0$ and $\boldsymbol{\beta}_0$, and this induces the estimation of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. According to Ye and Weiss (2003), the estimation of $\mathcal{S}_{Y|\mathbf{X}}$ can be improved by combining two SDR methods, which will be followed hereafter.

Let $\mathbf{M}_{\mathrm{name}}$ represent the kernel matrix constructed from the method called "name" in the subscript for a regression of $Y|\mathbf{X} = (\mathbf{U}, \mathbf{W})^{\mathrm{T}}$. For example, $\mathbf{M}_{\mathrm{OLS}}$ represent the kernel matrix constructed from OLS. Then, we will consider the following four population kernel matrices:

$$\begin{aligned}
\mathbf{M}_{\mathrm{OLS}} &= \boldsymbol{\Sigma}^{-1}\operatorname{cov}(Y, \mathbf{X})\operatorname{cov}(Y, \mathbf{X})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1};\\
\mathbf{M}_{\mathrm{OLS2}} &= \boldsymbol{\Sigma}^{-1}\operatorname{cov}\left(Y^2, \mathbf{X}\right)\operatorname{cov}\left(Y^2, \mathbf{X}\right)^{\mathrm{T}}\boldsymbol{\Sigma}^{-1};\\
\mathbf{M}_{\mathrm{SAVE}} &= \boldsymbol{\Sigma}^{-\frac{1}{2}}E(\mathbf{I} - \operatorname{cov}(\mathbf{Z}|Y))^2;\\
\mathbf{M}_{\mathrm{pHd}} &= \boldsymbol{\Sigma}^{-\frac{1}{2}}E\left((Y - E(Y))\mathbf{Z}\mathbf{Z}^{\mathrm{T}}\right),
\end{aligned}$$

where $\boldsymbol{\Sigma} = \operatorname{cov}(\mathbf{X})$ and $\mathbf{Z} = \boldsymbol{\Sigma}^{-1/2}(\mathbf{X} - E(\mathbf{X}))$. Then, we consider the following weighted sums of the four kernel matrices:

$$\mathbf{M}_{\mathrm{C}} = \omega\mathbf{M}_{\square} + (1-\omega)\mathbf{M}_{\diamond},$$

where $0 \le \omega \le 1$ and $\mathbf{M}_{\square}$ and $\mathbf{M}_{\diamond}$ are two different kernel matrices among the four. If ω is equal to 0 or 1, $\mathbf{M}_{\mathrm{C}}$ is reduced to one kernel matrix among the four. According to Ye and Weiss (2003), the columns of $\mathbf{M}_{\mathrm{C}}$ span $\mathcal{S}_{Y|\mathbf{X}}$.

Once $\mathbf{M}_{\mathrm{C}}$ is constructed, it is spectral-decomposed such that $\mathbf{M}_{\mathrm{C}} = \sum_{i=1}^{p}\lambda_i\boldsymbol{\gamma}_i\boldsymbol{\gamma}_i^{\mathrm{T}}$ with $\lambda_1 \ge \lambda_2 > 0 = \cdots = 0$. Then the eigenvectors of $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2)$, corresponding to the two largest and non-zero eigenvalues of λ_1 and λ_2, forms a basis of $\mathcal{S}_{Y|\mathbf{X}}$, and can be considered as $(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0)$. Here the last $p-2$ eigenvalues of $\lambda_3, \lambda_4, \ldots, \lambda_p$ are all equal to zero, because the true dimension in model (2.1) is assumed to be two.

In practice, the four $\mathbf{M}_{\bullet}$ are replaced by its own sample quantities, denoting them as $\hat{\mathbf{M}}_{\mathrm{OLS}}$, $\hat{\mathbf{M}}_{\mathrm{OLS2}}$, $\hat{\mathbf{M}}_{\mathrm{SAVE}}$ and $\hat{\mathbf{M}}_{\mathrm{pHd}}$. Then, from the sample version of $\hat{\mathbf{M}}_{\mathrm{C}} = \omega\hat{\mathbf{M}}_{\square} + (1-\omega)\hat{\mathbf{M}}_{\diamond}$, the eigenvectors of $\hat{\boldsymbol{\gamma}} = (\hat{\boldsymbol{\gamma}}_1, \hat{\boldsymbol{\gamma}}_2)$ corresponding to its two largest eigenvalues is obtained. Finally, $\mathcal{S}(\hat{\boldsymbol{\gamma}})$ becomes the estimate of $\mathcal{S}_{Y|\mathbf{X}}$.

Table 1: Candidates of $\hat{\mathbf{M}}_{\mathrm{C}}$

Case 1:	$\hat{\mathbf{M}}_{\mathrm{OLS}}$	Case 2:	$\omega\hat{\mathbf{M}}_{\mathrm{OLS}} + (1-\omega)\hat{\mathbf{M}}_{\mathrm{OLS2}}$
Case 3:	$\omega\hat{\mathbf{M}}_{\mathrm{OLS}} + (1-\omega)\hat{\mathbf{M}}_{\mathrm{pHd}}$	Case 4:	$\omega\hat{\mathbf{M}}_{\mathrm{OLS2}} + (1-\omega)\hat{\mathbf{M}}_{\mathrm{pHd}}$
Case 5:	$\omega\hat{\mathbf{M}}_{\mathrm{OLS}} + (1-\omega)\hat{\mathbf{M}}_{\mathrm{SAVE}}$	Case 6:	$\omega\hat{\mathbf{M}}_{\mathrm{OLS2}} + (1-\omega)\hat{\mathbf{M}}_{\mathrm{SAVE}}$

As the candidates of $\hat{\mathbf{M}}_{\mathrm{C}}$, the six cases are considered with varying $\omega = 0.25, 0.5, 0.75$ for Case 2 and $\omega = 0.0, 0.25, 0.5, 0.75$ for Cases 3–6, which are summarized in Table 1. In the candidates, $\hat{\mathbf{M}}_{\mathrm{SIR}}$ is not considered, because Case 2 is similar to SIR and often shows better performances in the basis estimation of $\mathcal{S}_{Y|\mathbf{X}}$ according to Yoo (2009), if its dimension is equal to two. For Cases 3–6 with $\omega = 0$ and Case 2, $\hat{\boldsymbol{\eta}}$ is constructed from one of $\hat{\mathbf{M}}_{\mathrm{SAVE}}$ or $\hat{\mathbf{M}}_{\mathrm{pHd}}$. For Case 1, since only one eigenvalue is non-zero, it is one-dimensional. According to Shao *et al.* (2006), it is equivalent to the estimation of the coefficient vector in single-index model. It must be noted that the proposed approach can investigate the appropriateness of GPDIM over a single-index model, and this is one of strengths of the proposed approach. If Case 1 of $\hat{\mathbf{M}}_{\mathrm{C}}$=$\hat{\mathbf{M}}_{\mathrm{OLS}}$ is chosen as the best, then a single-index model had better be used over the GPDIM. However, it should be done carefully (although the single-index model is recommended) because various numerical studies show that $\hat{\mathbf{M}}_{\mathrm{OLS}}$ can be selected as the best with smaller sample sizes. So it should be suggested that one needs to do further investigation to test that $d = 1$ through various SDR methods.

Next, a best case with proper ω in Table 1 should be chosen. A bootstrapping determination procedure in Ye and Weiss (2003) is applied. First, obtain $\hat{\boldsymbol{\eta}}_{\mathrm{ref}}$ from the original sample for each case. Then construct a bootstrap sample of Y and $\mathbf{X}$, and obtain $\hat{\boldsymbol{\eta}}_b$ from the bootstrap sample for each case. Then compute a trace correlation r in Hooper (1959) to measure how close the column subspaces of $\hat{\boldsymbol{\eta}}_{\mathrm{ref}}$ and $\hat{\boldsymbol{\eta}}_b$ such that

$$r_b = \sqrt{\frac{1}{2}\mathrm{trace}\left(\left\{\hat{\boldsymbol{\eta}}_{\mathrm{ref}}\left(\hat{\boldsymbol{\eta}}_{\mathrm{ref}}^{\mathrm{T}}\hat{\boldsymbol{\eta}}_{\mathrm{ref}}\right)^{-1}\hat{\boldsymbol{\eta}}_{\mathrm{ref}}^{\mathrm{T}}\right\}\left\{\hat{\boldsymbol{\eta}}_b\left(\hat{\boldsymbol{\eta}}_b^{\mathrm{T}}\hat{\boldsymbol{\eta}}_b\right)^{-1}\hat{\boldsymbol{\eta}}_b^{\mathrm{T}}\right\}\right)}.$$

To convert the correlation (larger r, closer to each other) to distance (smaller value, closer to each other), a trace correlation distance D_b^r is considered:

$$D_b^r = 1 - r_b.$$

Iterate the bootstrap sampling k times for each case and compute

$$\bar{D}_b^r = \frac{1}{k}\sum_{b=1}^{k} D_b^r.$$

Finally, choose the combinations of ω, $\hat{\mathbf{M}}_{\square}$ and $\hat{\mathbf{M}}_{\diamond}$ to have the smallest $\bar{D}_b^r$, and set $\hat{\boldsymbol{\eta}}_{\mathrm{ref}}$ constructed from the suggested combinations as the estimate of $\boldsymbol{\eta}$. From various simulation studies and from Yoo (2011, 2013), 500 numbers of bootstrapping are normally recommended in most cases.

Although $\hat{\boldsymbol{\eta}}$ can be obtained via the bootstrapping SDR approach, it is not known which of $\hat{\boldsymbol{\eta}}_1$ or $\hat{\boldsymbol{\eta}}_2$ is an estimate of $\boldsymbol{\alpha}_0$. Denote $\hat{\boldsymbol{\alpha}}_0$ and $\hat{\boldsymbol{\beta}}_0$ as estimates of $\boldsymbol{\alpha}_0$ and $\hat{\boldsymbol{\beta}}_0$, respectively. The determination is critical, because a wrong decision automatically results in an incorrect estimation. To do this correctly, we will consider the following rationale. For $\boldsymbol{\alpha}_0 = (\boldsymbol{\alpha} \in \mathbb{R}^{p_u}, 0)$ and $\boldsymbol{\beta}_0 = (0, \boldsymbol{\beta} \in \mathbb{R}^{p_w})$, respectively, the last $p - p_u$ elements and the first $p - p_w$ elements are zeros. Suppose that $\hat{\boldsymbol{\eta}}_1$ is $\hat{\boldsymbol{\alpha}}_0$. Then, the last $p - p_u$ elements in $\hat{\boldsymbol{\eta}}_1$ should be close to zeros. Or, provided that $\hat{\boldsymbol{\eta}}_1$ is $\hat{\boldsymbol{\beta}}_0$, the first $p - p_w$ elements of $\hat{\boldsymbol{\eta}}_1$ should be close to zeros. Therefore, it is reasonable to compare the following four squared

sums under assuming each $\hat{\boldsymbol{\eta}}_i$ is the estimate of either $\boldsymbol{\alpha}_0$ or $\boldsymbol{\beta}_0$, and the squared sum will be called *distinguishing values* (DV):

$$\begin{aligned}
\mathrm{DV}_{\hat{\boldsymbol{\eta}}_1,\boldsymbol{\alpha}_0} &= \left| \sum_{\ell=1}^{p_u} \hat{\boldsymbol{\eta}}_{1_\ell}^2 - \sum_{\ell=(p_u+1)}^{p} \hat{\boldsymbol{\eta}}_{1_\ell}^2 \right|, \\
\mathrm{DV}_{\hat{\boldsymbol{\eta}}_1,\boldsymbol{\beta}_0} &= \left| \sum_{\ell=1}^{p-p_w} \hat{\boldsymbol{\eta}}_{1_\ell}^2 - \sum_{\ell=(p-p_w+1)}^{p} \hat{\boldsymbol{\eta}}_{1_\ell}^2 \right|, \\
\mathrm{DV}_{\hat{\boldsymbol{\eta}}_2,\boldsymbol{\alpha}_0} &= \left| \sum_{\ell=1}^{p_u} \hat{\boldsymbol{\eta}}_{2_\ell}^2 - \sum_{\ell=(p_u+1)}^{p} \hat{\boldsymbol{\eta}}_{2_\ell}^2 \right|, \\
\mathrm{DV}_{\hat{\boldsymbol{\eta}}_2,\boldsymbol{\beta}_0} &= \left| \sum_{\ell=1}^{p-p_w} \hat{\boldsymbol{\eta}}_{2_\ell}^2 - \sum_{\ell=(p-p_w+1)}^{p} \hat{\boldsymbol{\eta}}_{2_\ell}^2 \right|.
\end{aligned}$$

Larger DVs indicate more clear separation between the two sets. So, choose the largest one among the four candidate DVs, and set the first letter as the estimate of the second letter in the subscript of the DV. If any of $\hat{\boldsymbol{\eta}}_1$ or $\hat{\boldsymbol{\eta}}_2$ is determined as $\hat{\boldsymbol{\alpha}}_0$, the other one is automatically set to be $\hat{\boldsymbol{\beta}}_0$. For example, supposing that $DV_{\hat{\boldsymbol{\eta}}_1,\boldsymbol{\beta}_0}$ is the largest, $\hat{\boldsymbol{\eta}}_1$ becomes $\hat{\boldsymbol{\beta}}_0$, and $\hat{\boldsymbol{\eta}}_2$ does $\hat{\boldsymbol{\alpha}}_0$. After completing this determination procedure, $\boldsymbol{\alpha}$ and $\hat{\boldsymbol{\beta}}$ are estimated by the first p_u elements of $\hat{\boldsymbol{\alpha}}_0$ and the last p_w elements of $\hat{\boldsymbol{\beta}}_0$, respectively.

3. Numerical Studies and Data Analysis

3.1. Numerical studies

We considered the five simulated examples under the following variable configurations: $\mathbf{U} = (U_1, \ldots, U_5)^{\mathrm{T}}$ and $\mathbf{W} = (W_1, \ldots, W_5)^{\mathrm{T}} \overset{i.i.d}{\sim} N(0,1)$, and $\varepsilon \overset{i.i.d}{\sim} N(0,1) \perp\!\!\!\perp (\mathbf{U}, \mathbf{W})$.

Model 1: $Y = U_1 + W_1 + \varepsilon$.

Model 2: $Y = U_1 + W_1^2 + \varepsilon$.

Model 3: $Y = U_1 + \exp(W_1)\varepsilon$.

Model 4: $Y = U_1^2 + \exp(W_1)\varepsilon$.

Model 5: $Y = U_1^2 + W_1^2 + \varepsilon$.

Model 1 is a classical multiple linear regression, which is a special case of single-index models. So a vector $\boldsymbol{\eta} = (1,0,0,0,0,1,0,0,0,0)^{\mathrm{T}}$ is one-dimensional. For Models 2–5, we have $\boldsymbol{\eta} = \{(\boldsymbol{\alpha}, \mathbf{0}), (\mathbf{0}, \boldsymbol{\beta})\}^{\mathrm{T}}$, where $\boldsymbol{\alpha} = \boldsymbol{\beta} = (1,0,0,0,0)^{\mathrm{T}}$.

For each model, 50, 100, 200 sample sizes were considered with 100 iterations and 500 numbers of bootstrapping. We also computed the percentages that case1 was selected the best in all examples. To measure how well $\boldsymbol{\eta}$ is estimated, the averages $\bar{D}_\eta^r$ of the following trace correlation distance D_η^r between $\boldsymbol{\eta}$ and $\hat{\boldsymbol{\eta}}$ defined in Section 2.3 were computed:

$$D_\eta^r = 1 - \sqrt{\frac{1}{2}\mathrm{trace}\left(\left\{\hat{\boldsymbol{\eta}}\left(\hat{\boldsymbol{\eta}}^{\mathrm{T}}\hat{\boldsymbol{\eta}}\right)^{-1}\hat{\boldsymbol{\eta}}^{\mathrm{T}}\right\}\left\{\boldsymbol{\eta}\left(\boldsymbol{\eta}^{\mathrm{T}}\hat{\boldsymbol{\eta}}\right)^{-1}\boldsymbol{\eta}^{\mathrm{T}}\right\}\right)}.$$

Table 2: Simulation results for Examples 1–5: $\bar{D}^r_\eta$, the averages of the trace correlation distance between $\boldsymbol{\eta}$ and $\hat{\boldsymbol{\eta}}$

n		Model 1	Model 2	Model 3	Model 4	Model 5
50	$\bar{D}^r_\eta$	0.050	0.17	0.23	0.31	0.32
	Percent of case1	100	36.0	20.0	15.0	17.0
100	$\bar{D}^r_\eta$	0.025	0.06	0.11	0.14	0.10
	Percent of case1	100	3.0	6.0	2.0	1.0
200	$\bar{D}^r_\eta$	0.015	0.02	0.07	0.04	0.03
	Percent of case1	100	0.0	0.0	0.0	0.0

When case1 was recommended in Examples 2–5, one dimensional estimate of $\hat{\boldsymbol{\eta}} = (\hat{\eta}_1, \ldots, \hat{\eta}_{10})^{\mathrm{T}}$ was replaced by $\hat{\boldsymbol{\eta}}_{\mathrm{s}} = \{(\hat{\eta}_1, \ldots, \hat{\eta}_5, \mathbf{0}), (\mathbf{0}, \hat{\eta}_6, \ldots, \hat{\eta}_{10})\}^{\mathrm{T}}$, and D^r_η between $\boldsymbol{\eta}$ and $\hat{\boldsymbol{\eta}}_{\mathrm{s}}$ was computed. For all the artificial models, the average $\bar{D}^r_\eta$ of the trace correlation distances and the selection percentages of case1 are reported in Table 2.

Table 2 shows that, with smaller sample size $n = 50$, case1 was selected upto 36%, but the percentages dramatically decrease with $n = 100$. The computed $\bar{D}^r_\eta$ is quite reliable with $n = 100$ in all the models; however, they are relatively large with $n = 50$. Therefore, numerical studies show that, with moderate sample sizes, the proposed method is not problematic in practice.

Under CPU of AMD Phenom(tm)II and 4GB sizes of RAM, the computing times for one iteration of Example 1 with $n = 50, 100, 200$ and 500 numbers of bootstrapping were 14.33, 15.58 and 17.69 seconds, respectively. The system described above is below recent standard; therefore, the time will not matter in practice under fairly-equipped computers.

3.2. Real data application: beta-carotene plasma

For illustration purpose, we considered a data regarding plasma retinol and beta-carotene levels, which can be obtained from the following web address at http://lib.stat.cmu.edu/datasets/Plasma_Retinol. In the data, we consider a regression of beta-carotene plasma concentration levels given four continuous predictors of calories consumed per day (Calorie), grams of fiber consumed per day (Fiber), weight/height^2 (Quetelet), dietary retinol consumed per day (mcg, Ret.diet), and two categorical predictors of Gender (0 = male; 1 = female) and Smoke (0 = non-smoker; 1 = current smoker) as two sets of predictors.

The data was obtained from StatLib website. The 257th case was indicated as an outlier; therefore, it was removed from the data set, and the sample size was 314. To make the four continuous predictors satisfy requirements of SDR methods, Calorie, Fiber and Ret.diet were transformed to log-scale and Quetelet to the inverse-scale.

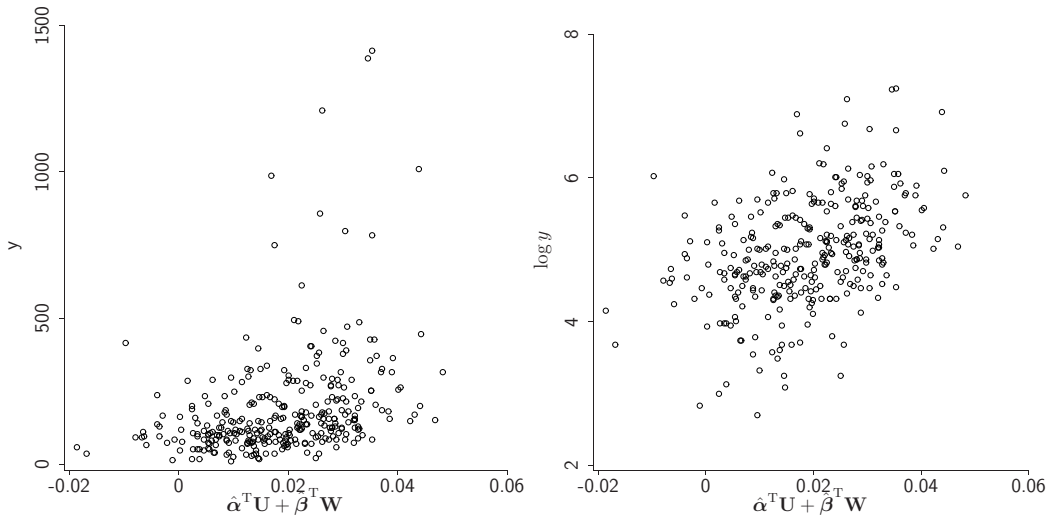
The predictors in the study naturally can be separated into two sets based on their types. Let $\mathbf{U} \in \mathbb{R}^4$ and $\mathbf{W} \in \mathbb{R}^2$ represent the four-dimensional continuous predictors and the two-dimensional categorical ones, respectively.

First, from the candidates of $\hat{\mathbf{M}}_{\mathrm{C}}$, $\bar{D}^r_b$ were computed as reported in Table 3. According to the table, all $\bar{D}^r_b$ are observed to be pretty low, although $\bar{D}^r_b$ from OLS is the smallest. Therefore, the single-index model should be a good choice for the data analysis despite considering the possibility of GPDIM. This result is similar to the founding of Yoo (2008, 2010). Then, marginally standardizing each of the remaining predictors to have a sample standard deviation of 1, the analysis might now be continued by a scatter plot of each of the response against the following estimated single-index predictor given in Figure 1(a) and (b).

$$\begin{aligned}\hat{\boldsymbol{\alpha}}^{\mathrm{T}}\mathbf{U} + \hat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{W} = &-0.008 \log \text{Calorie} + 0.018 \log \text{Fiber} - 0.001 \log \text{Ret.diet} + .999\ \text{Quetelet}^{-1} \\ &+ 0.006\ \text{Gender} - 0.012\ \text{Smoke}.\end{aligned}$$

Table 3: Determination of the best case via $\bar{D}_b^r$

Cases	ω	$\bar{D}_b^r$
Case 1, OLS		0.018
Case 2	$\omega = 0.25$	0.066
$\omega\hat{\mathbf{M}}_{\text{OLS}} + (1-\omega)\hat{\mathbf{M}}_{\text{ols2}}$	$\omega = 0.50$	0.062
	$\omega = 0.75$	0.079
Case 3	$\omega = 0.25$	0.074
$\omega\hat{\mathbf{M}}_{\text{OLS}} + (1-\omega)\hat{\mathbf{M}}_{\text{pHd}}$	$\omega = 0.50$	0.073
	$\omega = 0.75$	0.075
Case 4	$\omega = 0.25$	0.053
$\omega\hat{\mathbf{M}}_{\text{OLS2}} + (1-\omega)\hat{\mathbf{M}}_{\text{pHd}}$	$\omega = 0.50$	0.053
	$\omega = 0.75$	0.053
$\hat{\mathbf{M}}_{\text{pHd}}$ (Case 3 and Case 4 with $\omega = 0$)		0.086
Case 5	$\omega = 0.25$	0.111
$\omega\hat{\mathbf{M}}_{\text{OLS}} + (1-\omega)\hat{\mathbf{M}}_{\text{SAVE}}$	$\omega = 0.50$	0.112
	$\omega = 0.75$	0.111
Case 6	$\omega = 0.25$	0.106
$\omega\hat{\mathbf{M}}_{\text{OLS2}} + (1-\omega)\hat{\mathbf{M}}_{\text{SAVE}}$	$\omega = 0.50$	0.106
	$\omega = 0.75$	0.106
$\hat{\mathbf{M}}_{\text{SAVE}}$ (Case 5 and Case 6 with $\omega = 0$)		0.097



(a) Original scale response (b) log-transformed response

Figure 1: *Scatter plots between the response and the estimated single-index predictor in Section 3.2*

4. Discussion

This paper defines generalized partially double-index model and unknown coefficient vectors in the model are estimated via sufficient dimension reduction (SDR). For this, a combination of popular SDR methods is considered, and the best one is determined through a bootstrapping procedure. Distinguishing values are newly proposed to match the estimates to the coefficient vectors in the model properly. One of the strengths of the proposed method is that it can investigate the appropriateness of GPDIM over a single-index model. Various numerical studies support the proposed methodologies.

Since a generalized partially single-index model has been successfully used in many science fields GPDIM generalizes it; therefore, it is believed that GPDIM is beneficiary to progress in relevant

sciences fields.

The proposed method is not applicable under $n < p$, because the SDR methods used in the paper are limited in such case. As future work, a method to estimate the coefficients in the generalized partially double-index model under $n < p$ is needed to be developed with and work along these lines in progress.

Acknowledgment

The author is grateful to the associate editor and the two referees for many insightful and helpful comments.

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**, 477–489.

Cook, R. D. (1998). *Regression Graphics : Ideas for Studying Regressions through Graphics*, John Wiley & Sons, New York.

Cook, R. D. and Weisberg, S. (1991). Comment: Sliced inverse regression for dimension reduction by Ker-Chau Li, *Journal of the American Statistical Association*, **86**, 328–332.

Hooper, J. (1959). Simultaneous equations and canonical correlation theory, *Econometrica*, **27**, 245–256.

Li, K. C. (1991). Sliced inverse regression for dimension reduction, *Journal of the American Statistical Association*, **86**, 316-327.

Li, K. C. (1992). On principal Hessian directions for data visualization and dimension reduction: Another application of Stein's lemma, *Journal of the American Statistical Association*, **87**, 1025–1039.

Shao, Y., Cook, R. D. and Weisberg, S. (2006). The linearity condition and adaptive estimation in single-index regressions, regressions, Available from: http://arxiv.org/pdf/1001.4802.pdf

Ye, Z. and Weiss, R. E. (2003). Using the bootstrap to select one of a new class of dimension reduction methods, *Journal of the American Statistical Association*, **98**, 968–979.

Yin, X. and Cook, R. D. (2002). Dimension reduction for the conditional kth moment in regression, *Journal of Royal Statistical Society Series B*, **64**, 159–175.

Yoo, J. K. (2008). A novel moment-based dimension reduction approach in multivariate regression, *Computational Statistics and Data Analysis*, **52**, 3843–3851.

Yoo, J. K. (2009). Partial moment-based sufficient dimension reduction, *Statistics and Probability Letters*, **79**, 450–456.

Yoo, J. K. (2010). Integrated partial sufficient dimension reduction with heavily unbalanced categorical predictors, *The Korean Journal of Applied Statistics*, **23**, 977–985.

Yoo, J. K. (2011). Unified predictor hypothesis tests in sufficient dimension reduction: A bootstrap approach, *Computational Statistics and Data Analysis*, **40**, 217–225.

Yoo, J. K. (2013). Advances in seeded dimension reduction: Bootstrap criteria and extensions, *Computational Statistics and Data Analysis*, **60**, 70–79.

Received March 23, 2015; Revised May 8, 2015; Accepted May 8, 2015