Scandinavian Journal of Statistics Theory and Applications

Scandinavian Journal of Statistics, Vol. 44: 1–20, 2017
doi: 10.1111/sjos.12238

Analysis of Double Single Index Models

KUN CHEN
Department of Statistics, University of Connecticut
YANYUAN MA
Department of Statistics, University of South Carolina

ABSTRACT. Motivated from problems in canonical correlation analysis, reduced rank regression and sufficient dimension reduction, we introduce a double dimension reduction model where a single index of the multivariate response is linked to the multivariate covariate through a single index of these covariates, hence the name double single index model. Because nonlinear association between two sets of multivariate variables can be arbitrarily complex and even intractable in general, we aim at seeking a principal one-dimensional association structure where a response index is fully characterized by a single predictor index. The functional relation between the two single-indices is left unspecified, allowing flexible exploration of any potential nonlinear association. We argue that such double single index association is meaningful and easy to interpret, and the rest of the multi-dimensional dependence structure can be treated as nuisance in model estimation. We investigate the estimation and inference of both indices and the regression function, and derive the asymptotic properties of our procedure. We illustrate the numerical performance in finite samples and demonstrate the usefulness of the modelling and estimation procedure in a multi-covariate multi-response problem concerning concrete.

Key words: canonical correlation analysis, reduced rank regression, semiparametric efficiency, single index models, sufficient dimension reduction

1. Introduction

In scientific research and engineering, many statistical problems share a common goal of deciphering the associations between certain features and outcomes/responses from noisy data. When both the feature and response variables are multivariate, several different strategies exist to model their relations. Among the popular approaches are the canonical correlation analysis (CCA) (Hotelling, 1936) and the reduced rank regression (Anderson, 1951; Reinsel & Velu, 1998; Mukherjee & Zhu, 2011), both are designed to examine possible linear association between the two sets of random variables.

Specifically, write the covariate vector $\mathbf{X} \in \mathbb{R}^p$ and the response variable $\mathbf{Y} \in \mathbb{R}^q$, where $p > 1$, $q > 1$. CCA seeks linear combinations $\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{Y}$ and $\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}$ that have maximum correlation with each other. In other words, CCA searches for unit length vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ so that $\mathrm{corr}(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{Y}, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X})$ is maximized. Because correlation is chosen as the sole criterion to evaluate the closeness between $\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{Y}$ and $\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}$, CCA implicitly assumes a linear relation between these two quantities, or, at the very least, CCA is only interested in the linear relation between them. Similar to CCA, in the multivariate linear regression framework, the RRR model assumes a linear relation $\mathbf{Y} = \mathbf{C}^{\mathrm{T}}\mathbf{X} + \boldsymbol{\epsilon}^*$ between the responses and covariates, where the coefficient matrix $\mathbf{C} \in \mathbb{R}^{p\times q}$ is possibly of low rank, say, $\mathrm{rank}(\mathbf{C}) = r \leq \min(p, q)$, and $\boldsymbol{\epsilon}^*$ is usually assumed to follow a multivariate normal distribution with mean zero. The main idea of RRR amounts to seek the best low-rank approximation of $\mathbf{Y}$ supervised by the covariate information in $\mathbf{X}$, that is, minimizing $E\left\{(\mathbf{Y} - \mathbf{C}^{\mathrm{T}}\mathbf{X})^{\mathrm{T}}(\mathbf{Y} - \mathbf{C}^{\mathrm{T}}\mathbf{X})\right\}$ subject to $\mathrm{rank}(\mathbf{C}) \leq r$. When we consider the unit-rank RRR model, it becomes $\mathbf{Y} = c\boldsymbol{\alpha}\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X} + \boldsymbol{\epsilon}^*$. Here, c is the first singular value of $\mathbf{C}$ and $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ are the first left and right singular vectors of $\mathbf{C}$, respectively. This can be further written as $\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{Y} = c\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X} + \epsilon$, where ϵ is a mean zero error term. Obviously, the linear relation

between $\mathbf{Y}$ and $\mathbf{X}$ in RRR implies a linear relation between $\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{Y}$ and $\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}$. In fact, many commonly used multivariate techniques, including CCA, RRR and principal component analysis are all intrinsically related and all rely on certain linear assumption (Hotelling, 1936; Reinsel & Velu, 1998). Although in practice, multiple linearly dependent pairs of directions can be retained from CCA or RRR either sequentially or simultaneously, to focus on the main idea, we restrict our attention to the extraction of a single pair of directions in this paper, following the spirit of the single index model.

In real world applications, linearity is often too strong an assumption when characterizing variable association, and nonlinearity inevitably arises, especially in multivariate settings. However, extension of the available nonparametric techniques designed for univariate response to multivariate response is not quite straightforward, not only because of the curse of dimensionality, but also because of the difficulty in efficiently modelling the dependence structure among the response variables to fully embrace the multivariate nature of the problem. Many existing nonlinear methods were originated from classical CCA and RRR (Gifi, 1990; Hsieh, 2000; He *et al.*, 2003; Yuan *et al.*, 2007; Mukherjee & Zhu, 2011). (Xia, 2008) proposed a semiparametric approach of CCA, in which the estimation was based on minimizing $E\left\{\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{Y} - E\left(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{Y} \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}\right)\right\}^2$, where the conditional expectation $E(\cdot \mid \cdot)$ was estimated nonparametrically. SCA thus extends the classical CCA, as the latter simply assumes the conditional expectation to be linear. A similar approach is the generalized CCA proposed by (Iaci *et al.*, 2010); the method searches the pair of indices by minimizing $E\left\{\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{Y} - E\left(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{Y} \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}\right)\right\}^2 + E\left\{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X} - E\left(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X} \mid \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{Y}\right)\right\}^2$, treating the two sets of variables symmetrically. There have been several approaches that find $\left(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}, \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y}\right)$ by maximizing certain divergence measure between the joint distribution of $\left(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}, \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y}\right)$ and the product of their marginal distributions. (Iaci & Sriram, 2013) proposed two families of multivariate association measures based on power divergence and alpha divergence, and (Mandal & Cichocki, 2013) proposed a generalized method of CCA called AB-canonical analysis using Alpha-Beta divergence. For extensions more related to RRR, (Chan *et al.*, 2004) studied the properties of a general semiparametric partial linear reduced-rank regression model, and (Yuan *et al.*, 2007) proposed a nonparametric low-rank factor model using regression splines. For other methods concerning the use of dimension reduction techniques to facilitate the exploration of multivariate nonlinear association, (Li *et al.*, 2008) and the references therein.

Clearly, as soon as we venture into the territory beyond linearity, the possible multivariate association structures quickly become so rich and complex that it can even be infeasible to fully retrieve the true association structure. In this paper, to relax the assumption on linear association while still keep the model tractable, motivated by the sufficient dimension reduction literature, we introduce a flexible and yet manageable modelling strategy, where we assume there exists $\boldsymbol{\alpha} \in \mathbb{R}^q$ and $\boldsymbol{\beta} \in \mathbb{R}^p$ so that $\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{Y}$ relies on $\mathbf{X}$ through $\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}$, but we do not impose a linear relation or any specific functional link between $\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{Y}$ and $\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}$. Specifically, we only assume

$$f_{\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{Y}|\mathbf{X}}(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y}, \mathbf{x}) = f_{\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{Y}|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}}\left(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y}, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}\right). \tag{1}$$

Here, $f_{\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{Y}|\mathbf{X}}$ stands for the probability density function of $\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{Y}$ conditional on $\mathbf{X}$, and $f_{\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{Y}|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}}$ is similarly defined. The model described in (1) is what we name the double single index model (DSI), for the obvious reason that there are two single indices described by $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ respectively. Our proposal has several key ingredients. First, a variable index is often of practical interest and admits meaningful interpretation, and thus this desirable feature is retained in our model, the same as in single index models (Ichimura, 1993). Moreover, searching for a pair of associated indices is the essential objective in many real-world multivariate problems, see for example, (Witten *et al.*, 2009; Zhu *et al.*, 2014) and (Chen *et al.*, 2014). Second, to allow flexibility in

the process of pursuing nonlinearity, we do not intend to characterize the association between $\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{Y}$ and $\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}$. Instead, we aim to extract a relatively simple yet meaningful one-dimensional association between the response variables and the predictors. The DSI model is directly built on the conditional distribution of the variables, in contrast to many methods that only model the mean association structure. Third, in our approach, the desired simple DSI structure is perceived as lurking beneath other parts of the multivariate association of no direct interest. As such, these other parts are treated as nuisance and left unspecified. In estimation, only a working model is needed and the estimation is not sensitive to its misspecification; see Sections 2 and 3 for details. This is a rather important feature both practically and conceptually, especially given that modern data are obtained with ever increasing complexity, and yet often, only a few summary features of the data contribute to the actual knowledge discovery.

The DSI model has connections to several familiar multivariate models. For example, in the special case when $q = 1$, the DSI model in (1) reduces to the familiar single index model (Ichimura, 1993; Härdle *et al.*, 1993). In addition, DSI is an extension and generalization of CCA and RRR, in that it allows the association between $\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{Y}$ and $\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}$ to be nonlinear. As DSI is specified from conditional distribution, it is more comprehensive than the SCA model which only concerns the association in the mean. We also avoid specifying and modelling other possibly intractable dependence structures between $\mathbf{Y}$ and $\mathbf{X}$. The DSI model is also related to the multivariate response sufficient dimension reduction model (Li *et al.*, 2008). In this context, when the structural dimension is one, the model assumes that $\mathbf{Y}$ depends on $\mathbf{X}$ through $\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}$, that is, $f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y},\mathbf{x}) = f_{\mathbf{Y}|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}}(\mathbf{y},\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})$. This automatically leads to $f_{\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{Y}|\mathbf{X}}(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y},\mathbf{x}) = f_{\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{Y}|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}}(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y},\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})$ for any $\boldsymbol{\alpha}$. Now, if we relax the requirement so that this relation only holds for some specific $\boldsymbol{\alpha}$ instead of all $\boldsymbol{\alpha}$, then we obtain the DSI model (1).

The DSI model described in (1) arises naturally in practice. In civil engineering, it is an important topic to study the association between the quality of concrete and its composition (Yeh, 2006). Concrete is a highly complex material and consists of a mixture of several ingredients, including cement, fly ash, blast-furnace slag, water, superplasticizer and aggregate, etc. To summarize the composition of concrete is to study the proportion of these different ingredients, hence a natural way to summarize them is via their linear combination. On the other hand, in terms of quality, concrete is also measured in different aspects. Generally speaking, concrete which has high consistency at its fresh state while also has high strength at its hardened state, indicates that it has the properties of stability and durability, and is thus considered to be of high quality. The various aspects of the concrete quality include strength, stability, durability, etc, and can be summarized into a linear combination of these individual properties. It is thus natural to apply DSI to explore whether a quality index of the concrete ($\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{Y}$) exhibits some interesting linear/nonlinear relationship with certain composition of the concrete ($\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}$). In Section 4, we analyse a data example concerning concrete to further demonstrate the application of DSI.

Given that we can estimate the linear combination coefficients in $\boldsymbol{\alpha}, \boldsymbol{\beta}$, we can subsequently perform a classical univariate-covariate univariate-response nonparametric regression to identify the functional relationship between the two indices. Our proposed method thus provides a useful exploratory tool for examining potential nonlinear associations between two sets of variables.

2. Methodology

To ensure identifiability of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, we fix the last component of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ to be 1 and require (1) to hold at unique $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ locally. We write $\boldsymbol{\alpha} = \left(\boldsymbol{\alpha}_u^{\mathrm{T}}, 1\right)^{\mathrm{T}}$ and $\boldsymbol{\beta} = \left(\boldsymbol{\beta}_u^{\mathrm{T}}, 1\right)^{\mathrm{T}}$. The requirement can be easily satisfied by reordering the components in $\mathbf{Y}$ and $\mathbf{X}$ if necessary. We point out that the parameterization of requiring unit length of an index with positive first component and that

of requiring a fixed component to be one are both commonly used in the literature (Newey & Stoker, 1993; Klein & Shen, 2007; Klein & Vella, 2009); here we choose the latter to enable the semi-parametric analysis and computation to be carried out in a more straightforward way. Under this parameterization, our interest is then exclusively in the $q-1$ dimensional vector $\boldsymbol{\alpha}_u$ and the $p-1$ dimensional vector $\boldsymbol{\beta}_u$. Here the subindex $_u$ stands for unknown. Let $\boldsymbol{\gamma} = \left(\boldsymbol{\alpha}_u^{\mathrm{T}}, \boldsymbol{\beta}_u^{\mathrm{T}}\right)^{\mathrm{T}}$ be the unknown parameter of interest.

To provide a more direct and intuitive example of the model and its identifiability in (1), we consider the case when $\boldsymbol{\alpha}_u$ is zero. This occurs when the last component Y_q depends on $\mathbf{X}$ through a single index $\boldsymbol{\beta}$, while all other components in $\mathbf{Y}$, i.e. $Y_1, \ldots, Y_{q-1}$ depend on $\mathbf{X}$ through structures more complex than the single index model. For example, $Y_k = m_k(X_k X_{k+1}) + \epsilon_k$ for $k = 1, \ldots, q-1$, where ϵ_k is a mean zero random variable independent of $\mathbf{X}$ and m_k is a non-constant function. Having understood the component-wise model corresponding to the special $\boldsymbol{\alpha}$, we can then generalize the situation to the case when the response variable is further rotated and stretched by incorporating a general $\boldsymbol{\alpha}$. Further, if we restrict our interest on $\boldsymbol{\alpha}$ in a local neighbourhood, we can allow more components of $\mathbf{Y}$ to depend on $\mathbf{X}$ through single indices, as long as different components of $\mathbf{Y}$ correspond to different single indices. In this case, in a local neighbourhood, only one of these single index structures, corresponding to one particular component of $\mathbf{Y}$, will be of interest. With the additional rotation and stretching, only one linear combination of $\mathbf{Y}$ will be captured by $\boldsymbol{\alpha}$; hence, the problem is locally identifiable.

Following model (1), we write out the likelihood at one typical observation as

$$f_{\mathbf{X},\mathbf{Y}}(\mathbf{x}, \mathbf{y}) = \eta_1(\mathbf{x})\eta_2\left(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y}, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}\right)\eta_3\left(\mathbf{y}_r, \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y}, \mathbf{x}\right).$$

Here, $\mathbf{y}_r$ is the vector of the first $q-1$ components of $\mathbf{y}$, η_1 represents the probability density function (pdf) of $\mathbf{X}$ and η_3 represents the pdf of $\mathbf{Y}_r$ conditional on $\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{Y}$ and $\mathbf{X}$. We use η_2 to represent the pdf of $\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{Y}$ conditional on $\mathbf{X}$, which by the model assumption in (1) is a function of $\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y}$ and $\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}$ only. Note that η_1, η_2, η_3 are all unknown. It is now clear that (1) can be viewed as a semiparametric model where the parameter of interest is $\boldsymbol{\gamma}$ and η_1, η_2, η_3 are three nuisance parameters. We thus use the semiparametric analysis tools to derive nuisance tangent space Λ and its orthogonal complement $\Lambda^{\perp}$. The details of the derivation are in the supporting information, where we obtain the conclusion that

$$\Lambda^{\perp} = \left\{\mathbf{b}\left(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{Y}, \mathbf{X}\right) - E\left(\mathbf{b} \mid \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y}, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}\right) : \forall \mathbf{b}\left(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{Y}, \mathbf{X}\right) \in \mathbb{R}^{p+q-2}\ s.t.\ E\left(\mathbf{b} \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}\right) = E(\mathbf{b} \mid \mathbf{x})\right\}.$$

This result somewhat resembles the results in (Ma & Zhu, 2012), where their univariate Y is replaced by $\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{Y}$ here. Thus, the constructions there can be applied here as well by replacing all the instances of Y with $\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{Y}$. Let $\mathbf{a}$ and $\mathbf{a}_i$'s be arbitrary functions of $\mathbf{x}$, while $\mathbf{g}$ and $\mathbf{g}_i$'s be arbitrary functions of $\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y}$ and $\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}$. Here, $\mathbf{a}, \mathbf{a}_i, \mathbf{g}, \mathbf{g}_i$ can be scalar, vector or matrix functions as long as their dimensions conform, and the dimension of their product is $p+q-2$, that is, $\mathbf{g}\mathbf{a} \in \mathbb{R}^{p+q-2}$ and $\mathbf{g}_i\mathbf{a}_i \in \mathbb{R}^{p+q-2}$ for $i = 1, \ldots, k$. Since

$$E\left[\left\{\mathbf{g}\left(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{Y}, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}\right) - E\left(\mathbf{g} \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}\right)\right\}\left\{\mathbf{a}(\mathbf{X}) - E\left(\mathbf{a} \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}\right)\right\}\right] = \mathbf{0} \tag{2}$$

and

$$E\left[\sum_{i=1}^{k}\left\{\mathbf{g}_i\left(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{Y}, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}\right) - E\left(\mathbf{g}_i \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}\right)\right\}\left\{\mathbf{a}_i(\mathbf{X}) - E\left(\mathbf{a}_i \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}\right)\right\}\right] = \mathbf{0}, \tag{3}$$

we can use the functions inside the earlier expectations to construct root-n consistent estimators. The construction contained in (2) and (3) possesses a nice double robustness property,

in that between the two expectations $E\left\{\mathbf{g}\left(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y}, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}\right) \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}\right\}$ (or $E\left\{\mathbf{g}_i\left(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y}, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}\right) \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}\right\}$) and $E\left\{\mathbf{a}(\mathbf{x}) \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}\right\}$ $\left(\text{or} E\left\{\mathbf{a}_i(\mathbf{x}) \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}\right\}\right)$, as long as we calculate one of them correctly, we are free to mis-specify the other, and the consistency of the estimating function will still be retained. That is, for instance in (2), we have

$$E\left[\left\{\mathbf{g}\left(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y}, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}\right) - E\left(\mathbf{g} \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}\right)\right\}\left\{\mathbf{a}(\mathbf{x}) - \mathbf{h}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})\right\}\right] = \mathbf{0},$$

and

$$E\left[\left\{\mathbf{g}\left(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y}, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}\right) - \mathbf{h}\left(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}\right)\right\}\left\{\mathbf{a}(\mathbf{x}) - E\left(\mathbf{a} \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}\right)\right\}\right] = \mathbf{0}$$

for any function $\mathbf{h}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})$. However, different from the practice in (Ma & Zhu, 2012), we summarize the theoretical results of estimating $\boldsymbol{\gamma}$ based on (2) in Theorems 1, where the matrix $\mathbf{A}$ in Theorem 1 is required to have rank $p + q - 2$.

Theorem 1. *Under the regularity conditions C1-C6 listed in the supplement A.2, the estimator* $\widehat{\boldsymbol{\gamma}}$ *from the estimating equation*

$$\sum_{i=1}^{n}\left\{\mathbf{g}\left(\widehat{\boldsymbol{\alpha}}^{\mathrm{T}}\mathbf{y}_i, \widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{x}_i\right) - \widehat{E}\left(\mathbf{g} \mid \widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{x}_i\right)\right\}\left\{\mathbf{a}(\mathbf{x}_i) - \widehat{E}\left(\mathbf{a} \mid \widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{x}_i\right)\right\} = \mathbf{0}$$

is consistent, that is,.

$$\widehat{\boldsymbol{\gamma}} \to \boldsymbol{\gamma}$$

in probability when $n \to \infty$. *In addition, the estimator satisfies*

$$\sqrt{n}\mathbf{A}\left(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}\right) \to N(\mathbf{0}, \boldsymbol{B})$$

in distribution when $n \to \infty$. *Here,*

$$\mathbf{A} = E\left(\partial \operatorname{vec}\left[\left\{\mathbf{g}\left(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y}, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}\right) - E\left(\mathbf{g} \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}\right)\right\}\left\{\mathbf{a}(\mathbf{x}) - E\left(\mathbf{a} \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}\right)\right\}\right] / \partial \boldsymbol{\gamma}^{\mathrm{T}}\right),$$
$$\mathbf{B} = \operatorname{cov}\left(\operatorname{vec}\left[\left\{\mathbf{g}\left(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y}, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}\right) - E\left(\mathbf{g} \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}\right)\right\}\left\{\mathbf{a}(\mathbf{x}) - E\left(\mathbf{a} \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}\right)\right\}\right]\right).$$

Theorem 1 implies an interesting phenomenon, in that although we estimated the two expectations conditional on $\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{x}$ nonparametrically, the corresponding estimation causes no effect on the final asymptotic properties of $\widehat{\boldsymbol{\gamma}}$. In other words, if we had known how to obtain $E\left(\mathbf{a} \mid \widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{x}\right)$ and $E\left(\mathbf{g} \mid \widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{x}\right)$ exactly, the estimation of $\boldsymbol{\gamma}$ would not have been improved further. This nice property is a direct result of the double centring form of the estimating equation in Theorem 1, where we centred both $\mathbf{g}$ and $\mathbf{a}$ through subtracting their respective mean conditional on $\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{x}$, before multiplication. Similar practice has been used in other models in the partially linear model related literature (Ma *et al.*, 2006; Ma & Zhu, 2013a) and sufficient dimensional reduction literature (Ma & Zhu, 2012). How the double centring operation leads to this property is clearly shown in the proof of Theorem 1, especially through Lemma 1.2, given in the supplement A.3. It is also clear from the derivation in the supplement A.3 that if we had taken advantage of the double robustness property mentioned before and had estimated only one expectation faithfully while using an arbitrary $\mathbf{h}\left(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}\right)$ to replace the other expectation, then the fact that we had to estimate the conditional expectation would have led to an alteration of the variability in estimating $\widehat{\boldsymbol{\gamma}}$.

We now further investigate the efficient estimation issue through calculating the score and the efficient score. First, straightforward calculation yields

$$\mathbf{S}_\beta\left(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y},\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}\right)=\frac{\partial\log\eta_2\left(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y},\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}\right)}{\partial(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})}\mathbf{x}_r,$$

where we use $\mathbf{x}_r$ to denote the vector of the first $p-1$ components of $\mathbf{x}$. Now projecting $\mathbf{S}_\beta$ onto $\Lambda^\perp$, we obtain

$$\mathbf{S}_{\mathrm{eff}\,\beta}\left(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y},\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}\right)=\frac{\partial\log\eta_2\left(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y},\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}\right)}{\partial\left(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}\right)}\left\{\mathbf{x}_r-E\left(\mathbf{X}_r\mid\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}\right)\right\}.$$

This is because we can easily verify that $\mathbf{S}_{\mathrm{eff}\,\beta}\left(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y},\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}\right)\in\Lambda^\perp$ and $\left\{\partial\log\eta_2\left(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y},\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}\right)/\partial(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})\right\}E\left(\mathbf{X}_r\mid\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}\right)\in\Lambda$, based on the description of $\Lambda^\perp$ and Λ in supplement A.1. We now further calculate

$$\mathbf{S}_\alpha\left(\mathbf{y}_r,\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y},\mathbf{x}\right)=\frac{\partial\log\eta_2\left(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y},\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}\right)}{\partial\left(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y}\right)}\mathbf{y}_r+\frac{\partial\log\eta_3\left(\mathbf{y}_r,\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y},\mathbf{x}\right)}{\partial(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y})}\mathbf{y}_r.$$

Projecting $\mathbf{S}_\alpha$ onto $\Lambda^\perp$, we obtain $\mathbf{S}_{\mathrm{eff}\alpha}\left(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y},\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}\right)=E\left(\mathbf{S}_\alpha\mid\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y},\mathbf{x}\right)-E\left(\mathbf{S}_\alpha\mid\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y},\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}\right)$. This is because

$$E\left\{E\left(\mathbf{S}_\alpha\mid\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y},\mathbf{x}\right)\mid\mathbf{x}\right\}=E(\mathbf{S}_\alpha\mid\mathbf{x})=\int\frac{\partial\left\{\eta_2\left(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y},\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}\right)\eta_3\left(\mathbf{y}_r,\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y},\mathbf{x}\right)\right\}}{\partial\left(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y}\right)}d(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y})\mathbf{y}_r d\mathbf{y}_r=\mathbf{0},$$

hence, $E\left\{E\left(\mathbf{S}_\alpha\mid\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y},\mathbf{x}\right)\mid\mathbf{x}\right\}=E\left\{E\left(\mathbf{S}_\alpha\mid\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y},\mathbf{x}\right)\mid\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}\right\}$, which implies $\mathbf{S}_{\mathrm{eff}\alpha}\in\Lambda^\perp$. On the other hand, $\mathbf{S}_\alpha-E\left(\mathbf{S}_\alpha\mid\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y},\mathbf{x}\right)\in\Lambda_3$ and $E\left\{E\left(\mathbf{S}_\alpha\mid\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y},\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}\right)\mid\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}\right\}=E\left(\mathbf{S}_\alpha\mid\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}\right)=\mathbf{0}$, hence $\mathbf{S}_\alpha-\mathbf{S}_{\mathrm{eff}\alpha}\in\Lambda$ indeed. Hence, the projection of $\mathbf{S}_\alpha$ onto $\Lambda^\perp$ is indeed given by $\mathbf{S}_{\mathrm{eff}\alpha}$. Specifically, we obtain

$$\begin{aligned}&\mathbf{S}_{\mathrm{eff}\alpha}(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y},\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})\\&=\frac{\partial\log\eta_2\left(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y},\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}\right)}{\partial(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y})}\left\{E\left(\mathbf{y}_r\mid\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y},\mathbf{x}\right)-E\left(\mathbf{y}_r\mid\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y},\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}\right)\right\}\\&\quad+E\left\{\frac{\partial\log\eta_3\left(\mathbf{y}_r,\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y},\mathbf{x}\right)}{\partial(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y})}\mathbf{y}_r\mid\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y},\mathbf{x}\right\}-E\left\{\frac{\partial\log\eta_3\left(\mathbf{y}_r,\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y},\mathbf{x}\right)}{\partial\left(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y}\right)}\mathbf{y}_r\mid\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y},\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}\right\}.\end{aligned}$$

Combining the two calculations, we have $\mathbf{S}_{\mathrm{eff}}=\left(\mathbf{S}_{\mathrm{eff}\alpha}^{\mathrm{T}},\mathbf{S}_{\mathrm{eff}\beta}^{\mathrm{T}}\right)^{\mathrm{T}}$.

Unfortunately, the estimation of $E\left(\mathbf{y}_r\mid\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y},\mathbf{x}\right)$ and $\eta_3\left(\mathbf{y}_r,\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y},\mathbf{x}\right)$ is subject to curse of dimensionality because of the presence of $\mathbf{x}$ (as well as $\mathbf{y}_r$ for $\eta_3\left(\mathbf{y}_r,\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y},\mathbf{x}\right)$). Hence, the efficient estimator is unreachable in practice. This is in contrast to (Ma & Zhu, 2013b), where only a univariate Y is concerned. However, we can use the form of $\mathbf{S}_{\mathrm{eff}}$ to construct locally efficient estimators using a working model of η_3. Although we can estimate η_2, considering that in any case we cannot guarantee efficiency, we will use a working model of η_2 as well. To this end, we propose to posit the working models $\eta_2^*\left(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y},\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}\right)$ and $\eta_3^*\left(\mathbf{y}_r,\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y},\mathbf{x}\right)$. We then construct the locally efficient estimators from $\mathbf{S}_{\mathrm{eff}}^*=\left(\mathbf{S}_{\mathrm{eff}\alpha}^{*\,\mathrm{T}},\mathbf{S}_{\mathrm{eff}\beta}^{*\,\mathrm{T}}\right)^{\mathrm{T}}$, where

$$\mathbf{S}^*_{\text{eff}\alpha} = \frac{\partial \log \eta_2^* \left(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y}, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}\right)}{\partial(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y})} \left[E^* \left(\mathbf{y}_r \mid \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y}, \mathbf{x}\right) - E\left\{E^* \left(\mathbf{y}_r \mid \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y}, \mathbf{x}\right) \mid \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y}, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}\right)\right\}\right]$$
$$+ E^* \left\{ \frac{\partial \log \eta_3^* \left(\mathbf{y}_r, \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y}, \mathbf{x}\right)}{\partial \left(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y}\right)} \mathbf{y}_r \mid \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y}, \mathbf{x} \right\}$$
$$- E\left[E^* \left\{ \frac{\partial \log \eta_3^* \left(\mathbf{y}_r, \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y}, \mathbf{x}\right)}{\partial(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y})} \mathbf{y}_r \mid \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y}, \mathbf{x} \right\} \mid \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y}, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x} \right],$$

and

$$\mathbf{S}^*_{\text{eff}\beta} = \left[\frac{\partial \log \eta_2^* \left(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y}, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}\right)}{\partial(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})} - E\left\{ \frac{\partial \log \eta_2^* \left(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y}, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}\right)}{\partial(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})} \middle| \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x} \right\} \right] \left\{\mathbf{x}_r - E\left(\mathbf{X}_r \middle| \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}\right)\right\}.$$

We can obtain the locally efficient estimator through using $\mathbf{S}^*_{\text{eff}}$. Specifically, use $\mathbf{O}_i$ to denote the ith observation, and use $\mathbf{S}^*_{\text{eff}}(\mathbf{O}_i; \boldsymbol{\gamma}, \widehat{E})$ to denote the efficient score evaluated at $\mathbf{O}_i$, with E replaced by its kernel estimator $\widehat{E}$. The estimator $\widehat{\boldsymbol{\gamma}} = \left(\widehat{\boldsymbol{\alpha}}^{\mathrm{T}}, \widehat{\boldsymbol{\beta}}^{\mathrm{T}}\right)^{\mathrm{T}}$ satisfies

$$\sum_{i=1}^{n} \mathbf{S}^*_{\text{eff}} \left(\mathbf{O}_i; \widehat{\boldsymbol{\gamma}}, \widehat{E}\right) = \mathbf{0}. \tag{4}$$

We show that $\widehat{\boldsymbol{\gamma}}$ is locally efficient, that is, it is efficient when η_2 and η_3 are correctly specified; otherwise it is still consistent and asymptotically normal.

Theorem 2. *Under the regularity conditions B1-B5 listed in the supplement A.4, the estimator* $\widehat{\boldsymbol{\gamma}}$ *from the estimating equation (4) is locally efficient. Specifically, when* $n \to \infty$,

$$\sqrt{n}(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}) \to N\left(\mathbf{0}, \mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1\mathrm{T}}\right),$$

where $\mathbf{A} = -E\left\{\partial \mathbf{S}^*_{\text{eff}}(\mathbf{O}_i; \boldsymbol{\gamma})/\partial \boldsymbol{\gamma}^{\mathrm{T}}\right\}$ *and* $\mathbf{B} = E\left[\{\mathbf{S}^*_{\text{eff}}(\mathbf{O}_i; \boldsymbol{\gamma}) - \mathbf{u}^*(\mathbf{x}_i; \boldsymbol{\gamma})\}^{\otimes 2}\right]$. *Here,*

$$\mathbf{u}^*(\mathbf{x}_i; \boldsymbol{\gamma}) \equiv \int \mathbf{b}^*(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y}, \mathbf{x}_i) \eta_2 \left(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y}, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i\right) d(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y})$$
$$- \int \frac{\int_{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x} = \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i} \mathbf{b}^* \left(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y}, \mathbf{x}\right) f_{\mathbf{x}}(\mathbf{x}) d\mathbf{x} \eta_2 \left(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y}, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i\right)}{f(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i)} d\left(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y}\right).$$

In addition, when $\eta_2^* \left(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y}, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}\right) = \eta_2 \left(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y}, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}\right)$ *and* $\eta_3^* \left(\mathbf{y}_r, \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y}, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}\right) = \eta_3 \left(\mathbf{y}_r, \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y}, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}\right)$, *then* $\mathbf{A} = \mathbf{B} = E\left\{\mathbf{S}^{\otimes 2}_{\text{eff}}(\mathbf{O}_i; \boldsymbol{\gamma})\right\}$, *and the estimator is efficient. Here* $\mathbf{a}^{\otimes 2} \equiv \mathbf{a}\mathbf{a}^{\mathrm{T}}$ *for any vector or matrix* $\mathbf{a}$.

The details of the implementation of the locally efficient estimator is the following. To simplify the description of the implementation of the locally efficient estimator, we first define functions

$$\mathbf{m}_1 \left(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}\right) \equiv E\left(\mathbf{x}_r | \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}\right),$$
$$m_2 \left(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}\right) \equiv E\left\{ \frac{\partial \log \eta_2^* \left(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y}, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}\right)}{\partial \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}} \middle| \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x} \right\},$$
$$\mathbf{m}_3 \left(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y}, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}\right) \equiv E\left\{\mathbf{b}^* \left(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y}, \mathbf{x}\right) \middle| \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y}, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}\right\},$$

where

$$\mathbf{b}^*\left(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y},\mathbf{x}\right)\equiv\frac{\partial\log\eta_2^*\left(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y},\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}\right)}{\partial\left(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y}\right)}E^*\left(\mathbf{y}_r\,|\,\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y},\mathbf{x}\right)+E^*\left\{\frac{\partial\log\eta_3^*\left(\mathbf{y}_r,\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y},\mathbf{x}\right)}{\partial\left(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y}\right)}\mathbf{y}_r\,\middle|\,\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y},\mathbf{x}\right\}.$$

The Nadaraya–Watson kernel estimators of $\mathbf{m}_1, m_2$ and $\mathbf{m}_3$ are respectively

$$\begin{aligned}
\widehat{\mathbf{m}}_1(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})&=\frac{\sum_{i=1}^n K_h\left\{\boldsymbol{\beta}^{\mathrm{T}}(\mathbf{x}-\mathbf{x}_i)\right\}\mathbf{x}_{ri}}{\sum_{i=1}^n K_h\left\{\boldsymbol{\beta}^{\mathrm{T}}(\mathbf{x}-\mathbf{x}_i)\right\}},\\
\widehat{m}_2\left(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}\right)&=\frac{\sum_{i=1}^n K_h\left\{\boldsymbol{\beta}^{\mathrm{T}}(\mathbf{x}-\mathbf{x}_i)\right\}\partial\log\eta_2^*\left(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y}_i,\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i\right)/\partial\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i}{\sum_{i=1}^n K_h\left\{\boldsymbol{\beta}^{\mathrm{T}}(\mathbf{x}-\mathbf{x}_i)\right\}},\\
\widehat{\mathbf{m}}_3\left(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y},\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}\right)&=\frac{\sum_{i=1}^n K_h\left\{\boldsymbol{\beta}^{\mathrm{T}}(\mathbf{x}-\mathbf{x}_i)\right\}\mathbf{b}^*\left(t\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y},\mathbf{x}_i\right)}{\sum_{i=1}^n K_h\left\{\boldsymbol{\beta}^{\mathrm{T}}(\mathbf{x}-\mathbf{x}_i)\right\}},
\end{aligned}$$

where h is a bandwidth. To emphasize the dependence on $\mathbf{m}_1, m_2, \mathbf{m}_3$, we can write the locally efficient score function

$$\mathbf{S}_{\mathrm{eff}}^*\left\{\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y},\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x},\mathbf{m}_1\left(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}\right),\mathbf{m}_2\left(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}\right),\mathbf{m}_3\left(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y},\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}\right)\right\}.$$

The locally efficient estimator can then be obtained in practice through solving the estimating equation

$$\sum_{i=1}^n\mathbf{S}_{\mathrm{eff}}^*\left\{\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y}_i,\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i,\widehat{\mathbf{m}}_1\left(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i\right),\widehat{m}_2\left(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i\right),\widehat{\mathbf{m}}_3\left(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y}_i,\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i\right)\right\}=\mathbf{0}.$$

As long as the bandwidth h is fixed, the only unknown quantity in the estimating equation is $\boldsymbol{\gamma}$. The estimating equation can be solved by standard optimization methods such as the Newton–Raphson algorithm or the trust region method. Because a wide range of bandwidths all lead to the same asymptotic result; hence, even in finite samples, the estimator is quite insensitive to the bandwidth. Thus, we can simply use $h=n^{-1/5}$ in the implementation. One can certainly perform cross validation and use a unique bandwidth to associate with each specific nonparametric regression, at the cost of selecting more bandwidths. We have implemented our method in `MATLAB`, where Newton–Raphson algorithm is applied and numerical difference is used to approximate the local derivative functions. Based on our limited experience, the computation is stable and fast.

Having estimated $\boldsymbol{\gamma}$, we can perform nonparametric regression of $\widehat{\boldsymbol{\alpha}}^{\mathrm{T}}\mathbf{Y}$ on $\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{X}$ to further estimate η_2, following, for example, (Fan *et al.*, 2003). Because $\boldsymbol{\gamma}$ is estimated at the parametric rate of root-n, the estimation of η_2 will have the usual nonparametric estimation rate, and its first order asymptotic properties are the same as that of the estimation of η_2 using the true parameter $\boldsymbol{\gamma}$. Because the derivation and the results of the nonparametric procedure are standard, we omit the details.

Trimming (Ichimura, 1993) is often needed in nonparametric estimation to handle the potential issue of dividing by zero. However, trimming is avoided here because we only need the nonparametric evaluations at $\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i$, which is always positive because we include the ith observation in the estimator. Further, condition C3 guarantees the density of $\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}$ to be bounded away from zero. Thus, when sample size is sufficiently large, the estimated density is also bounded away from zero.

In practice, to specify η_2^* and η_3^*, we suggest the following. First, use simpler methods such as CCA or SCA to obtain starting values $(\tilde{\boldsymbol{\alpha}},\tilde{\boldsymbol{\beta}})$ for DSI. Then, η_2^* can be specified based on the empirical conditional distribution between the leading canonical pairs $\tilde{\boldsymbol{\alpha}}^{\mathrm{T}}\mathbf{y}$ and $\tilde{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{x}$,

for example, $[, \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y} \sim \mathrm{N}\left(a\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x} + b, \sigma^2\right)$, where a, b and σ^2 are estimated from a regression analysis between $\tilde{\boldsymbol{\alpha}}^{\mathrm{T}}\mathbf{y}$ and $\tilde{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{x}$. Our numerical results suggest that η_3^* can be specified in a more crude way, Section 3. For example, we can specify η_3^* by assuming the components of $\mathbf{y}_r$ are independent conditional on $\tilde{\boldsymbol{\alpha}}^{\mathrm{T}}\mathbf{y}$ and $\mathbf{x}$ and conducting regression analysis between $\mathbf{y}_r$ and some linear/nonlinear functions of $\tilde{\boldsymbol{\alpha}}^{\mathrm{T}}\mathbf{y}$ and $\mathbf{x}$.

3. Simulation

3.1. Setups

We conduct simulation studies to evaluate the finite sample performance of the proposed methods. For comparison, the most relevant nonlinear approach to our method is the semi-parametric CCA (SCA) proposed by (Xia, 2008). SCA searches the pair of indices by minimizing $E\left\{\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{Y} - E\left(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{Y}|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}\right)\right\}^2$, and the procedure involves the estimation of $E(Y_i \mid \mathbf{X})$ and its derivatives using d-th order local polynomial smoothing, where $d > p/2 + 1$ in order to achieve $\sqrt{n}$-consistency. Several classical multivariate tools based on certain linearity assumption can also be applied for such two-way search, with CCA and RRR as the popular prototypes of those. We thus compare the proposed DSI approach with CCA, RRR and SCA.

We set $p = 5$, $q = 4$, $\boldsymbol{\alpha} = (1, 1, 1, 1)^{\mathrm{T}}$ and $\boldsymbol{\beta} = (1, -1, 1, -1, 1)^{\mathrm{T}}$ in all the simulation examples. The process of generating a typical observation $(\mathbf{x}, \mathbf{y})$ is as follows.

(1) Generate $\mathbf{x}$ from η_1, the marginal distribution of $\mathbf{X}$.
(2) Compute $\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}$ and generate $\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y}$ from η_2, the conditional distribution of $\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{Y}$ given $\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}$.
(3) Generate $\mathbf{y}_r$ from η_3, the conditional distribution of $\mathbf{Y}_r$ given $\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y}$ and $\mathbf{x}$.
(4) Compute y_q from the generated values $\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y}$ and $\mathbf{y}_r$, i.e., $y_q = \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y} - \boldsymbol{\alpha}_u^{\mathrm{T}}\mathbf{y}_r$. Let $\mathbf{y} = \left(\mathbf{y}_r^{\mathrm{T}}, y_q\right)^{\mathrm{T}}$.

We set η_1 as the standard multivariate normal distribution; in practice, with the components of $\mathbf{X}$ correlated, one may orthogonalize the variables before pursuing sufficient dimension reduction. We consider three models with different choices of η_2:

Model I: η_2 is the normal distribution with mean $\mu_\eta = \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}$ and variance $\sigma_\eta^2 = 4$.
Model II: η_2 is the normal distribution with mean $\mu_\eta = (\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})^2$ and variance $\sigma_\eta^2 = 6$.
Model III: η_2 is the normal distribution with mean $\mu_\eta = (\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})^2$ and variance $\sigma_\eta^2 = \sigma^2 \exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}/3)$ where $\sigma^2 = 6$.

In each of the aforementioned models, η_3 is set as the multivariate normal distribution with mean vector $\boldsymbol{\mu}_r = (\mu_{r,1}, \ldots, \mu_{r,q-1})^{\mathrm{T}}$ with

$$\mu_{r,i} = \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y}/q + 2\sin(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y}) + a\left(\mathbf{h}_i^{\mathrm{T}}\mathbf{x}\right) + b\left(\mathbf{h}_i^{\mathrm{T}}\mathbf{x}\right)^2, \qquad i = 1, \ldots, q-1,$$

and covariance matrix $4\mathbf{I}$, where the $\mathbf{h}_i$s are orthonormal vectors that are also orthogonal to $\boldsymbol{\beta}$. The constants a and b are chosen to control the marginal correlation structure of $\mathbf{Y}$. Specifically, we set $a = 3, b = 3$ in Model 1, and $a = 3, b = 9$ in both Models II and III, so that the correlations among the $Y_i, i = 1, \ldots, q$ are roughly at or below 0.6 in magnitude. These setups ensure that $\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{Y}$ is not dominated by any particular coordinate in $\mathbf{Y}$, and it is indeed the desired simple direction, that is, a direction in $\mathbf{Y}$ that is associated with a one-dimensional sufficient dimension reduction subspace in $\mathbf{X}$. For the aforementioned models, it can be conveniently shown that

$$\frac{\partial \log \eta_2\left(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y}, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}\right)}{\partial\left(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y}\right)} = \frac{\mu_\eta - \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y}}{\sigma_\eta^2},$$

$$\frac{\partial \log \eta_2\left(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y}, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}\right)}{\partial\left(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}\right)} = -\frac{1}{2}\frac{\partial \log \sigma_\eta^2}{\partial\left(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}\right)} - \frac{2\left(\mu_\eta - \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y}\right)\frac{\partial \mu_\eta}{\partial\left(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}\right)}\sigma_\eta^2 - \left(\mu_\eta - \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y}\right)^2\frac{\partial \sigma_\eta^2}{\partial\left(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}\right)}}{2\sigma_\eta^4},$$

and

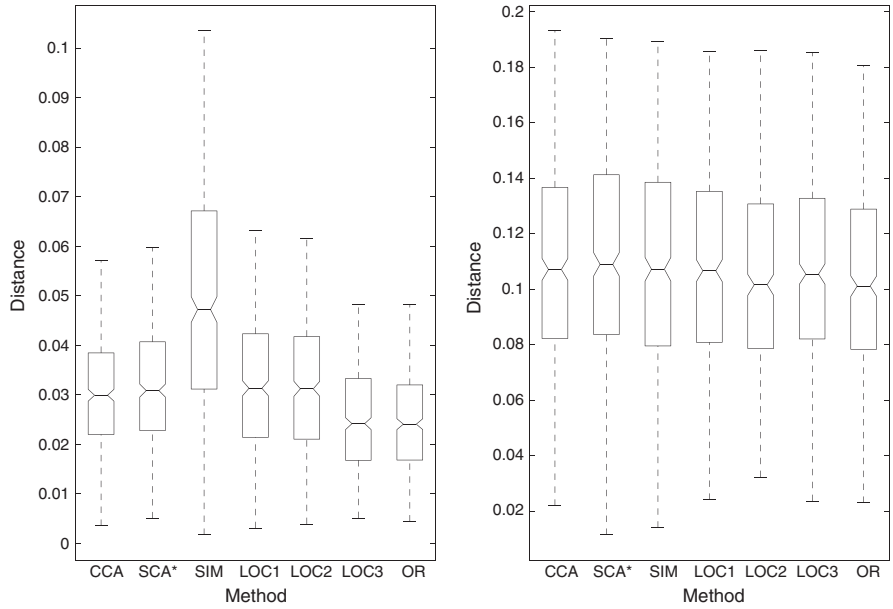
$$E\left\{\frac{\partial \log \eta_3\left(\mathbf{y}_r, \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y}, \mathbf{x}\right)}{\partial\left(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y}\right)}\mathbf{y}_r | \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y}, \mathbf{x}\right\} = \frac{\partial \boldsymbol{\mu}_r}{\partial\left(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y}\right)}.$$

The proposed locally efficient estimation is from solving the estimating equations (4) with potentially misspecified η_2 and η_3. For all three simulation examples, we set η_3^*, the working model for η_3, as normal with mean vector $\left(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y}/q + x_1^2, \ldots, \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y}/q + x_q^2\right)^{\mathrm{T}}$ and variance-covariance matrix identity. We set η_2^*, the working model of η_2, as $N\left(\mu_\eta = 2\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}, \sigma_\eta^2 = 9\right)$ in Model I and $N\left(\mu_\eta = |\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}|, \sigma_\eta^2 = 9\right)$ in Models II and III. Three locally efficient estimators are constructed: the first one is based on η_2^* and η_3^* (LOC1), the second one is based on η_2 and η_3^* (LOC2), and the third one is based on η_2^* and η_3 (LOC3). When both η_2 and η_3 are correctly specified, we obtain an efficient oracle estimator (OR). Because η_2 and η_3 are usually unknown in real problems, LOC2, LOC3 and OR are not feasible in practice, but here they may serve as benchmarks to examine the effects of model misidentification. Based on Theorem 1, we also construct a simple consistent estimator (SIM), in which we choose $\mathbf{g}\left(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y}, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}\right) = E\left(\mathbf{x} \mid \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y}\right)$ and $\mathbf{a}(\mathbf{x}) = \mathbf{x}^{\mathrm{T}}$. As we focus on the single index setup, the first leading pair of canonical variables are extracted from CCA, and a unit-rank estimator is obtained from RRR; the resulting estimators are denoted as CCA1 and RRR1, respectively. For CCA and RRR, we also extracted $\min(p, q)$ pairs of directions and recorded the one that is the closest to the true pair measured by $\left\|\widehat{\boldsymbol{\alpha}}\left(\widehat{\boldsymbol{\alpha}}^{\mathrm{T}}\widehat{\boldsymbol{\alpha}}\right)^{-1}\widehat{\boldsymbol{\alpha}}^{\mathrm{T}} - \boldsymbol{\alpha}\left(\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\alpha}\right)^{-1}\boldsymbol{\alpha}^{\mathrm{T}}\right\|_F + \left\|\widehat{\boldsymbol{\beta}}\left(\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}\right)^{-1}\widehat{\boldsymbol{\beta}}^{\mathrm{T}} - \boldsymbol{\beta}\left(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\beta}\right)^{-1}\boldsymbol{\beta}^{\mathrm{T}}\right\|_F$; the resulting estimators are denoted CCA* and RRR*, respectively. Similarly, for the semiparametric method SCA, we computed two estimators SCA1 and SCA*.

3.2. Results

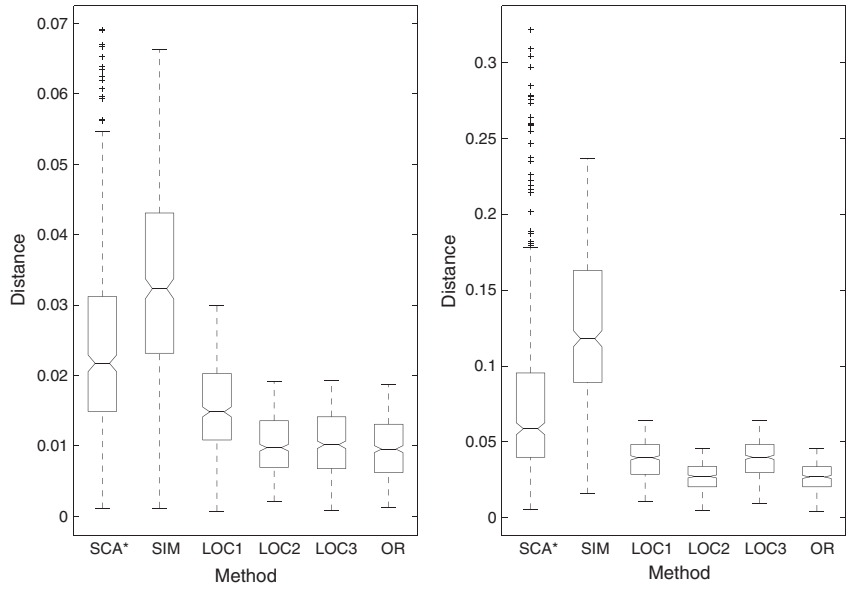
We have considered various sample sizes, that is, $n = 500, 200$ and 100, while for brevity, we mainly focus our discussion for the case $n = 500$ in the sequel, unless otherwise noted. The experiment is replicated 500 times under each setting. The obtained estimates $(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}})$ are standardized in the same way as the true $(\boldsymbol{\alpha}, \boldsymbol{\beta})$, as described in Section 2. Figures 1–3 show the boxplots of the Euclidean distances between the true parameters and their estimated counterparts from all simulation runs, that is, $d(\widehat{\boldsymbol{\alpha}}, \boldsymbol{\alpha}) = \left\|\widehat{\boldsymbol{\alpha}}\left(\widehat{\boldsymbol{\alpha}}^{\mathrm{T}}\widehat{\boldsymbol{\alpha}}\right)^{-1}\widehat{\boldsymbol{\alpha}}^{\mathrm{T}} - \boldsymbol{\alpha}\left(\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\alpha}\right)^{-1}\boldsymbol{\alpha}^{\mathrm{T}}\right\|_F$ for measuring the distance from $\widehat{\boldsymbol{\alpha}}$ to $\boldsymbol{\alpha}$, and $d(\widehat{\boldsymbol{\beta}}, \boldsymbol{\beta}) = \left\|\widehat{\boldsymbol{\beta}}\left(\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}\right)^{-1}\widehat{\boldsymbol{\beta}}^{\mathrm{T}} - \boldsymbol{\beta}\left(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\beta}\right)^{-1}\boldsymbol{\beta}^{\mathrm{T}}\right\|_F$ for measuring the distance from $\widehat{\boldsymbol{\beta}}$ to $\boldsymbol{\beta}$, where $\|\cdot\|_F$ denotes the Frobenius norm. Tables 1–3 report the average parameter estimates (ave) and their associated standard errors (std), for Models I–III, respectively. For the proposed semiparametric estimators, we also report the average of the estimated standard deviations ($\widehat{\text{std}}$) and the coverage of the estimated 95% confidence interval (95%), based on the asymptotic results.

In Model 1, the association between $\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{Y}$ and $\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}$ is linear, which should benefit the linear methods. From Table 1, CCA performs very well in estimation, but RRR performs much worse. The discrepancy in performance between these two methods is due to their different objectives: while CCA focuses on maximizing the correlation between a pair of directions in $\mathbf{Y}$ and $\mathbf{X}$, RRR focuses on explaining the variation in $\mathbf{Y}$ by $\mathbf{X}$. In our model setup, $\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{Y}$ and $\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}$ indeed



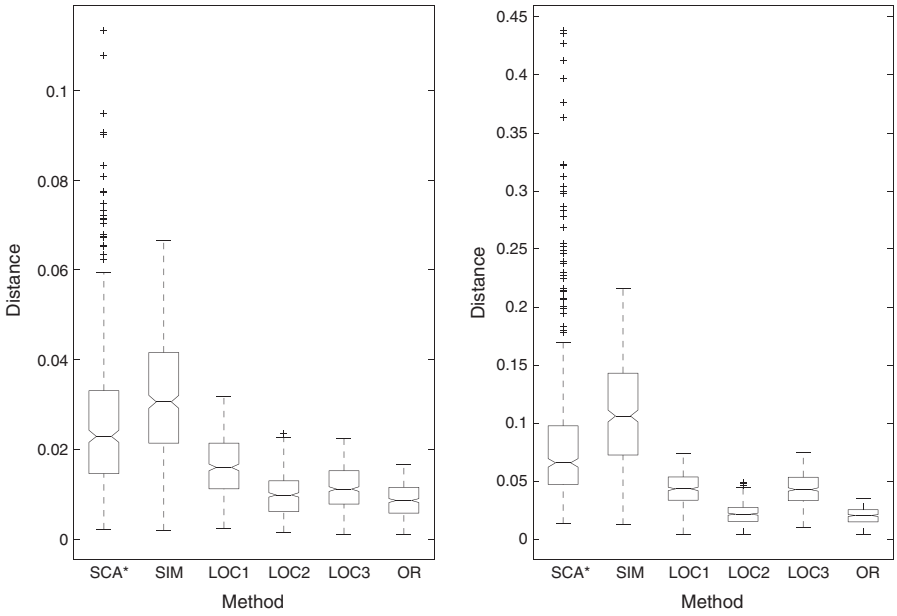
(a) $d(\widehat{\boldsymbol{\alpha}}, \boldsymbol{\alpha})$ (b) $d(\widehat{\boldsymbol{\beta}}, \boldsymbol{\beta})$

Fig. 1. Boxplots of $d(\widehat{\boldsymbol{\alpha}}, \boldsymbol{\alpha})$ and $d(\widehat{\boldsymbol{\beta}}, \boldsymbol{\beta})$ for Model I ($n = 500$). CCA, canonical correlation analysis; OR, oracle estimator; SCA, semiparametric approach of CCA; SIM, simple consistent estimator.



(a) $d(\widehat{\boldsymbol{\alpha}}, \boldsymbol{\alpha})$ (b) $d(\widehat{\boldsymbol{\beta}}, \boldsymbol{\beta})$

Fig. 2. Boxplots of $d(\widehat{\boldsymbol{\alpha}}, \boldsymbol{\alpha})$ and $d(\widehat{\boldsymbol{\beta}}, \boldsymbol{\beta})$ for Model II ($n = 500$). CCA, canonical correlation analysis; OR, oracle estimator; SCA, semiparametric approach of CCA; SIM, simple consistent estimator.



(a) $d(\widehat{\boldsymbol{\alpha}}, \boldsymbol{\alpha})$ (b) $d(\widehat{\boldsymbol{\beta}}, \boldsymbol{\beta})$

Fig. 3. Boxplots of $d(\widehat{\boldsymbol{\alpha}}, \boldsymbol{\alpha})$ and $d(\widehat{\boldsymbol{\beta}}, \boldsymbol{\beta})$ for Model III ($n = 500$). CCA, canonical correlation analysis; OR, oracle estimator; SCA, semiparametric approach of CCA; SIM, simple consistent estimator.

has the strongest linear association among all possible directions, which makes CCA suitable. However, $\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}$ does not necessarily coincide with the targeted direction of RRR, along which most of the variation in $\mathbf{Y}$ can be explained in the least squares sense. As a consequence, RRR is unsuitable here for the discovery of the desired single indices, and even RRR* performs poorly. The performance of SCA* is comparable with CCA; however, the extracted leading pair by the SCA method does not necessarily correspond to the desired pair, as seen from the performance of SCA1. If we knew the underlying model is linear, a parsimonious method like CCA would be preferable. Our results show that the proposed semiparametric approaches, which do not rely on the knowledge of linear model, work almost as well as CCA, with only a slight loss in efficiency. We plotted the results in Figure 1 to show the relative performance of the different methods. For better illustration, we omitted the estimators that perform much worse than the rest of methods.

In Models 2 and 3, the association between $\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Y}$ and $\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{X}$ is nonlinear, and any other direction in $\mathbf{Y}$ may not be adequately characterized by a single direction in $\mathbf{X}$. Not surprisingly, CCA and RRR both perform poorly. The bias in CCA* or RRR* is much smaller than CCA1 or RRR as expected, but the variance of either estimator is very high. Again, SCA1 may pick up other spurious directions to approximate a single index model. Nevertheless, it appears that the desirable pair is most likely among the ones obtained from SCA, albeit a much larger estimation error comparing with the proposed DSI methods. Also, the performance of SCA* in Model II is relatively better than that in Model III, because in Model II the two indices are related only in their mean structure, while in Model III the two indices are also related in their second moments. In all occasions, the DSI estimators continue to perform very well, clearly demonstrating the effectiveness of the proposed methods in detecting nonlinear association. LOC1 performs better than SIM in general as expected. Comparing the three locally efficient estimators and the oracle estimator, the misspecification of η_2 has a bigger impact on

Table 1. *Simulation results for Model I* ($n = 500$).

		α_1	α_2	α_3	β_1	β_2	β_3	β_4
CCA1	ave	1.0019	1.0017	1.0025	**−0.9954**	0.9978	**−0.9863**	0.9964
	std	0.0353	0.0333	0.0344	0.1415	0.1396	0.1310	0.1336
CCA*	ave	1.0019	1.0017	1.0025	**−0.9954**	0.9978	**−0.9863**	0.9964
	std	0.0353	0.0333	0.0344	0.1415	0.1396	0.1310	0.1336
RRR1	ave	**−0.4651**	**−0.5834**	**−0.4676**	0.9328	1.1671	**−0.3009**	0.3320
	std	0.6076	0.8502	0.6741	1.4070	2.0669	0.7229	0.8102
RRR*	ave	0.8305	1.0038	1.0469	**−0.7006**	0.8643	**−0.8071**	0.7799
	std	1.6763	1.4649	2.1281	3.6369	4.5548	2.5557	2.8155
SCA1	ave	**−1.2397**	**−0.1936**	0.0127	0.7962	**−0.7020**	**−0.2470**	0.2418
	std	1.7829	1.7914	1.7786	1.4098	1.4535	0.5063	0.5015
SCA*	ave	1.0027	1.0014	1.0022	**−0.9966**	0.9972	**−0.9860**	0.9976
	std	0.0367	0.0347	0.0356	0.1455	0.1414	0.1337	0.1360
SIM	ave	1.0000	0.9929	0.9968	**−1.0256**	1.0153	**−1.0300**	1.0233
	std	0.0553	0.0566	0.0570	0.1496	0.1340	0.1383	0.1331
	$\widehat{\text{std}}$	0.0548	0.0553	0.0568	0.1446	0.1437	0.1359	0.1366
	95%	0.9100	0.9140	0.9000	0.9380	0.9560	0.9520	0.9700
LOC1	ave	1.0013	0.9984	1.0001	**−1.0037**	1.0034	**−1.0092**	1.0031
	std	0.0356	0.0344	0.0327	0.1385	0.1425	0.1396	0.1322
	$\widehat{\text{std}}$	0.0322	0.0322	0.0324	0.1406	0.1448	0.1375	0.1393
	95%	0.9340	0.9280	0.9500	0.9380	0.9460	0.9380	0.9540
LOC2	ave	1.0015	0.9986	1.0002	**−1.0033**	1.0044	**−1.0083**	1.0035
	std	0.0348	0.0341	0.0326	0.1304	0.1357	0.1309	0.1239
	$\widehat{\text{std}}$	0.0332	0.0326	0.0330	0.1329	0.1365	0.1296	0.1311
	95%	0.9600	0.9300	0.9560	0.9480	0.9480	0.9380	0.9520
LOC3	ave	0.9998	0.9994	1.0004	**−1.0039**	1.0052	**−1.0113**	1.0045
	std	0.0279	0.0287	0.0267	0.1371	0.1359	0.1368	0.1294
	$\widehat{\text{std}}$	0.0280	0.0278	0.0278	0.1270	0.1297	0.1233	0.1262
	95%	0.9480	0.9440	0.9540	0.9380	0.9340	0.9220	0.9500
OR	ave	0.9996	0.9996	1.0007	**−1.0040**	1.0069	**−1.0107**	1.0052
	std	0.0272	0.0284	0.0264	0.1287	0.1292	0.1283	0.1213
	$\widehat{\text{std}}$	0.0271	0.0266	0.0267	0.1312	0.1332	0.1272	0.1293
	95%	0.9500	0.9320	0.9480	0.9520	0.9500	0.9420	0.9600

CCA, canonical correlation analysis; OR, oracle estimator; RRR, reduced rank regression; SCA, semiparametric approach of CCA; SIM, simple consistent estimator.

estimation than η_3 does. In both models, OR performs the best among all the methods, because of the fact that the search of the directions becomes more trackable when the underlying model structure is correctly chosen. On the other hand, even when both η_2 and η_3 are misspecified, LOC1 still achieves small bias and remarkable estimation accuracy, with only slightly increased standard errors.

Furthermore, we can see that the inference results based on the asymptotic analysis are accurate in general. The estimated standard errors match well with their counterparts based on Monte Carlo simulation, and the coverage probabilities are mostly close to the nominal level 95%. We notice that in Models II and III, LOC1 tends to be slightly biassed for the estimation of $\boldsymbol{\alpha}$, and the standard errors also tend to be slightly overestimated. Nevertheless, in our experiment the inference results improve when we increase the sample size.

We have also experimented with smaller sample sizes. The estimation performance of the semiparametric estimators in Model II for $n = 200$ and $n = 100$ are shown in the supporting information. While the estimation accuracy of the DSI methods is still satisfactory, it appears that the performance of SCA* deteriorates more severely. Probably this is because the SCA method requires the estimation of $E(Y_i|\mathbf{X})$ and its derivatives, which needs strong sample size

Table 2. *Simulation results for Model II* ($n = 500$).

		α_1	α_2	α_3	β_1	β_2	β_3	β_4
CCA1	ave	0.6085	0.0534	**−0.1249**	0.4247	**−1.0752**	**−0.1941**	0.1703
	std	5.3256	3.7520	4.8869	3.2969	3.3596	1.1036	1.0924
CCA*	ave	1.0022	1.0121	1.0252	**−0.9870**	0.8972	**−1.2320**	1.0596
	std	0.2406	0.2551	0.2683	2.5884	2.2488	5.2078	5.8762
RRR1	ave	**−0.3430**	**−0.4070**	**−0.3075**	0.7447	0.0802	0.0928	**−0.0913**
	std	0.5085	0.2223	0.4834	2.6632	2.7881	1.0480	0.9500
RRR*	ave	0.9439	1.0289	0.9611	**−1.1452**	0.8949	**−1.1906**	0.7960
	std	1.2258	0.6755	0.5541	3.7731	2.0927	5.3769	4.9484
SCA1	ave	**−0.3326**	**−0.1935**	**−0.1858**	0.9299	**−0.8993**	**−0.5943**	0.5953
	std	1.9992	1.8853	1.9391	2.1112	2.1041	0.5956	0.5970
SCA*	ave	1.0006	1.0005	0.9996	**−0.9979**	0.9922	**−0.9995**	0.9965
	std	0.0574	0.0383	0.0457	0.1929	0.1960	0.1233	0.1168
SIM	ave	1.0047	1.0026	1.0063	−1.0826	1.0878	**−1.0949**	1.0776
	std	0.0350	0.0368	0.0365	0.1751	0.1675	0.1787	0.1777
	$\widehat{\text{std}}$	0.0359	0.0351	0.0355	0.1711	0.1656	0.1626	0.1681
	95%	0.9260	0.9140	0.9220	0.9280	0.9320	0.9360	0.9300
LOC1	ave	1.0145	1.0133	1.0129	**−0.9999**	1.0015	**−1.0005**	1.0015
	std	0.0144	0.0160	0.0145	0.0406	0.0409	0.0401	0.0392
	$\widehat{\text{std}}$	0.0186	0.0198	0.0203	0.0394	0.0393	0.0382	0.0389
	95%	0.9540	0.9540	0.9820	0.9580	0.9420	0.9480	0.9640
LOC2	ave	1.0010	1.0012	1.0001	**−1.0022**	1.0031	**−1.0021**	1.0030
	std	0.0117	0.0115	0.0109	0.0323	0.0318	0.0323	0.0308
	$\widehat{\text{std}}$	0.0113	0.0110	0.0109	0.0322	0.0322	0.0318	0.0322
	95%	0.9580	0.9340	0.9660	0.9580	0.9560	0.9640	0.9580
LOC3	ave	1.0003	1.0011	1.0000	**−0.9991**	1.0009	**−1.0002**	1.0015
	std	0.0120	0.0118	0.0114	0.0386	0.0393	0.0390	0.0382
	$\widehat{\text{std}}$	0.0110	0.0109	0.0109	0.0386	0.0387	0.0377	0.0385
	95%	0.9380	0.9340	0.9340	0.9480	0.9480	0.9500	0.9640
OR	ave	1.0006	1.0011	1.0000	**−1.0021**	1.0029	**−1.0019**	1.0028
	std	0.0110	0.0110	0.0103	0.0319	0.0314	0.0320	0.0304
	$\widehat{\text{std}}$	0.0106	0.0105	0.0105	0.0318	0.0320	0.0315	0.0319
	95%	0.9460	0.9420	0.9640	0.9560	0.9560	0.9620	0.9600

CCA, canonical correlation analysis; OR, oracle estimator; RRR, reduced rank regression; SCA, semi-parametric approach of CCA; SIM, simple consistent estimator.

requirement depending on the predictor dimension p. We have experimented with models in which the two indices are related in the second moments but not the first, and as expected SCA* fails while the proposed method continues to perform well. We note that the inference results of DSI may become less accurate for small sample sizes. In particular, the coverage probabilities for SIM and LOC1 tend to be slightly lower than the nominal level. This is expected as the inference procedure involves numerical approximations in several places, and for complex models a larger sample size may be required to allow the asymptotic theory to take effect. Following the request of a referee, we also increased the dimensions p and q and investigated the scalability of the method. The results are very encouraging. We provide the details of the computational performance in the supporting information.

4. Concrete slump test data

As a mixture of several ingredients, concrete is a highly complex material. Understanding the relationship between the quality and composition of concrete is an important topic in the field of Civil Engineering. Generally speaking, concrete with high consistency at its fresh state and

Table 3. *Simulation results for Model III* ($n = 500$).

		α_1	α_2	α_3	β_1	β_2	β_3	β_4
CCA1	ave	0.4335	**−0.1846**	0.6086	0.5321	**−0.8804**	**−0.2799**	0.2006
	std	5.5592	4.2589	5.6687	3.9372	3.7894	1.3377	1.0205
CCA*	ave	1.0234	0.9830	1.0141	**−1.0164**	1.1505	**−0.9157**	1.1674
	std	0.2575	0.2588	0.2505	4.6014	4.5963	7.1897	9.5704
RRR1	ave	**−0.3377**	**−0.3885**	**−0.3376**	0.7026	0.0307	**−0.0410**	**−0.1217**
	std	0.5326	0.3455	0.5272	2.6470	3.0442	1.1687	1.0144
RRR*	ave	0.9656	0.9601	1.0030	**−0.7602**	0.6793	**−0.6031**	0.8798
	std	0.4133	0.4283	0.6232	4.1492	3.2080	5.3231	5.5661
SCA1	ave	−0.2611	**−0.2539**	**−0.4031**	0.9776	**−1.1616**	**−0.6185**	0.6212
	std	1.9660	1.9287	1.9992	2.1519	2.1893	0.5933	0.5946
SCA*	ave	1.0071	1.0027	1.0069	**−0.9820**	0.9957	**−0.9919**	0.9922
	std	0.1836	0.1037	0.1223	0.3488	0.3186	0.2126	0.2215
SIM	ave	1.0051	1.0024	1.0059	**−1.0837**	1.0779	**−1.0932**	1.0816
	std	0.0372	0.0344	0.0368	0.1679	0.1794	0.1671	0.1740
	$\widehat{\text{std}}$	0.0379	0.0386	0.0378	0.1870	0.1868	0.1885	0.1868
	95%	0.9300	0.9440	0.9280	0.9700	0.9540	0.9780	0.9700
LOC1	ave	1.0140	1.0129	1.0136	**−1.0034**	0.9977	**−0.9994**	0.9954
	std	0.0164	0.0157	0.0163	0.0506	0.0523	0.0537	0.0521
	$\widehat{\text{std}}$	0.0194	0.0216	0.0216	0.0501	0.0496	0.0489	0.0497
	95%	0.9420	0.9740	0.9760	0.9440	0.9480	0.9460	0.9580
LOC2	ave	1.0010	0.9999	1.0004	**−1.0017**	1.0003	**−1.0012**	1.0014
	std	0.0117	0.0099	0.0101	0.0239	0.0223	0.0236	0.0231
	$\widehat{\text{std}}$	0.0112	0.0099	0.0100	0.0239	0.0230	0.0232	0.0230
	95%	0.9480	0.9620	0.9560	0.9540	0.9740	0.9520	0.9500
LOC3	ave	1.0003	0.9999	1.0006	**−1.0037**	0.9975	**−0.9999**	0.9958
	std	0.0133	0.0127	0.0130	0.0485	0.0503	0.0523	0.0508
	$\widehat{\text{std}}$	0.0122	0.0122	0.0123	0.0495	0.0494	0.0486	0.0493
	95%	0.9340	0.9580	0.9400	0.9560	0.9580	0.9480	0.9660
OR	ave	1.0004	1.0000	1.0004	**−1.0016**	1.0000	**−1.0009**	1.0013
	std	0.0096	0.0098	0.0098	0.0220	0.0211	0.0223	0.0218
	$\widehat{\text{std}}$	0.0095	0.0096	0.0096	0.0225	0.0220	0.0223	0.0220
	95%	0.9500	0.9560	0.9440	0.9580	0.9760	0.9520	0.9500

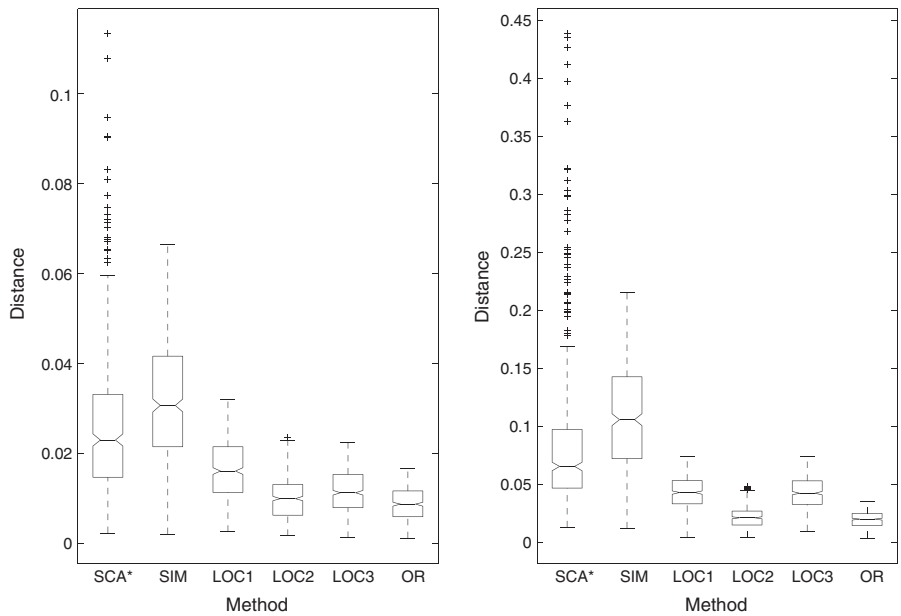
CCA, canonical correlation analysis; OR, oracle estimator; RRR, reduced rank regression; SCA, semi-parametric approach of CCA; SIM, simple consistent estimator.

with high strength at its hardened state exhibits desirable properties of stability and durability. The consistency of fresh concrete is commonly measured through a slump-cone test, by examining the behaviours of a compacted inverted cone of fresh concrete under the action of gravity: the slump is measured by the length of the drop from the top of the slumped concrete, and the slump flow is measured by its diameter. Here, we consider a slump test dataset, consisting of 103 sets of slump test measurements (Yeh, 2006; 2007). Three variables regarding the quality of concrete were recorded including slump (cm), slump flow (cm) and 28-day compressive strength (mpa). The ingredients composing the concrete were also recorded (kg/m^3), including cement, fly ash, blast furnace slag, water, superplasticizer and aggregate. Here, we apply the DSI approach to explore the association between the three quality variables ($q = 3$) and three ingredient variables ($p = 3$), the fly ash, water and superplasticizer, which are known to be important factors related to the slump and concrete quality (Yeh, 2006). All the variables are standardized prior to the analysis.

We apply CCA, RRR and SCA to identify possible linear/nonlinear relationships between the two sets of variables. We then conduct the DSI estimation, starting from 100 sets of initial

values of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, randomly generated by adding Gaussian noise $N(0, 3)$ to their CCA/RRR estimates. As the estimation problem is local in nature, this ensures that the starting points are fairly spread out in the vicinity of some initial linear estimates, enabling us to explore whether interesting directions of sufficient dimension reduction can be found when deviating away from the linear analysis. Because of the nonconvexity of the problem, multiple roots of the estimating equations may exist. In this problem, predominately, we find two roots from the 100 model fitting attempts. Upper plots of Figure 4 depict the observed data points along the estimated linear directions from CCA and RRR, together with the fitted linear regression curves. In the middle panel of Figure 4, we plotted the two sets of solutions from the DSI method, and the fitted nonparametric regression curves are also shown. The SCI methods also extracted two pairs of directions, as shown in the bottom panel of Figure 4. The parameter estimates are given in Table 4. We have used the single-indexing (leave-one-out) cross-validation method (Xia, 2008) to assess the goodness of fit of the extracted pairs, to test whether $\widehat{\boldsymbol{\alpha}}^{\mathrm{T}}\mathbf{Y}$ can be adequately predicted by a single index model of $\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{X}$, and all the six pairs mentioned earlier passed the test.

The first pair of directions found by either DSI or SCI mostly coincides with those from CCA. From the similarity of the results, as well as the fitted nonparametric curves, we can see a strong linear association along this pair of directions. It is worth pointing out that the coefficients for the slump flow has opposite sign from that of the strength or slump. This can be explained easily because in the slump test, the slump and the slump flow are in fact strongly positively correlated. Generally speaking, a lower slump implies a lower slump flow and higher compressive strength.



(a) $d(\widehat{\boldsymbol{\alpha}}, \boldsymbol{\alpha})$ (b) $d(\widehat{\boldsymbol{\beta}}, \boldsymbol{\beta})$

Fig. 4. Scatter plots along the estimated single-index directions for the slump test data analysis. CCA, canonical correlation analysis; DSI, double single index; RRR, reduced rank regression; SCA, semiparametric approach of CCA.

Table 4. *Coefficient estimation in the slump test data analysis.*

	Slump flow α_1	Strength α_2	Slump α_3	Fly ash β_1	Superplasticizer β_2	Water β_3	Lack of fit
CCA	**−1.6433**	0.2853	1.0000	0.1032	0.0715	1.0000	No
RRR	1.2951	−0.5765	1.0000	−0.1115	−0.1691	1.0000	No
SCA(1)	**−1.6917**	0.2938	1.0000	0.1380	0.0469	1.0000	No
SCA(2)	**−0.6435**	0.0612	1.0000	−0.1487	−0.5702	1.0000	No
DSI(1)	**−1.5573**	0.2700	1.0000	0.1108	0.0713	1.0000	No
DSI(2)	3.1315	**−1.5634**	1.0000	−0.1583	−0.1508	1.0000	No

CCA, canonical correlation analysis; DSI, double single index; RRR, reduced rank regression; SCA, semiparametric approach of CCA.

The second set of DSI solution reveals another interesting relation between the concrete quality and its character. In this case, the estimated $\widehat{\boldsymbol{\alpha}}$ from DSI agrees in sign with that from the RRR, although their coefficient values are quite different. The identified $\widehat{\boldsymbol{\beta}}$ directions in $\mathbf{X}$ from the two methods are similar and mainly dominated by the water content variable. This is not surprising as the water content is the most important factor influencing the property of concrete, and the fly ash and the superplasticizer are both supplemental admixtures that are expected to have some secondary impact. Up to a few outliers, the association between the identified single indices by DSI can be well characterized by the fitted robust nonlinear nonparametric regression line, as shown in Figure 4(d). The coefficient of determination (R^2) for the DSI fitted line is 0.503, while that for the RRR fitted line is 0.420. (We have removed two potential outliers, and the R^2 values before outlier removal are 0.426 and 0.364, respectively). As the water content increases, the quality index seems to increase sharply at the beginning, then flats out and eventfully decreases slightly. These findings are consistent with the results in (Yeh, 2006), in which a similar nonlinear relationship between slump and water content was detected via neural network models. From Figure (4), the second pair found by SCI appears to be spurious, and does not offer much insight to the problem. This example demonstrates that the DSI approach can be a useful and flexible tool for conveniently exploring simple nonlinear structures in complex multivariate association.

5. Discussion

Although the DSI method is illustrated in an engineering problem, it has potential in other application areas. For example, in marine ecology, DSI can be used to study the dependence between the yearly adult fish abundance, summarized from the observed fish abundances in spatial regions ($\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{Y}$) and the yearly larval abundance, summarized from observed daily spawning biomass ($\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}$) (Chen *et al.*, 2014). In portfolio construction, DSI can be used to study the relation between the asset return, summarized from the allocation of the available assets ($\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{Y}$) to the market return, summarized from market indices and macroeconomic variables ($\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}$). In genomic research, DSI can be used to study the relation between the summary of gene expression profiles ($\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{Y}$) and the summary of single-nucleotide polymorphism ($\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}$) (Witten *et al.*, 2009). More broadly, DSI can also be applied in many time series problems, where several random variables evolve together over time. In particular, the reduced-rank linear vector autoregressive (VAR) model is an important tool in modelling the vector time series (Reinsel & Velu, 1998). It can be readily seen that the DSI model extends and renovates a unit-rank VAR model, that is, the present value of an index of the vector time series has nonlinear relationship with the past value of another index.

We have developed a flexible DSI model for exploring unspecified and possibly nonlinear function relations between multivariate response and predictors. There are many directions for

future research. For example, our method can serve as a building block to study multi-index models, analogous to the multi-factor CCA or the RRR methods. To go beyond these linear methods, one challenge is how to exhaustively extract pairs of indices for sufficient dimension reduction without imposing any specific form or restrictive assumption on their functional relations. To this end, multi-index modelling and estimation strategies similar to the sufficient dimension reduction literature is one possibility. Sequentially extracting the single index pairs from both the covariate and response variables is also worth careful investigation. To further facilitate variable selection and model interpretation, we can also consider regularized estimation in the DSI model, for example, imposing sparsity assumption on $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ so that the constructed pair of indices only involves a subset of the responses and the predictors (Chen *et al.*, 2012; Chen & Huang, 2012; Bunea *et al.*, 2012).

An alternative model related to the one considered here can be constructed by further assuming that the dependence of the response variable $\mathbf{Y}$ on the covariates $\mathbf{X}$ is completely captured by the dependence of a linear combination of $\mathbf{Y}$ on $\mathbf{X}$. In other words, $\mathbf{Y}_r$ is independent of $\mathbf{X}$ conditional on $\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{Y}$. Although the assumption is stronger than the DSI model, it offers an interesting modelling approach and may have important applications. The estimation, efficiency and application of such model is certainly worth exploring.

Several possibilities exist for model checking. The general idea is that because our estimation method enables the estimation of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, one can construct both indices. This enables us to reduce the multi-covariate multi-response problem to an effective uni-covariate uni-response problem and facilitates the application of several existing methods. For example, to check whether $\widehat{\boldsymbol{\alpha}}^{\mathrm{T}}\mathbf{Y}$ can be adequately modelled by a single index model using $\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{X}$, many existing goodness-of-fit methods developed in the single index model framework can be applied (Stute & Zhu, 2005; Xia, 2008; Liang *et al.*, 2010; Ma *et al.*, 2014). In addition, a graphical tool is also possible as an exploratory tool, where one only needs to plot the data cloud formed by the two indices and inspect if the data cloud is compact along the response index. This exploratory tool is often used in the dimension reduction literature.

A potentially more fundamental problem is how to parsimoniously and flexibly approximate the multivariate conditional distribution of $\mathbf{Y}$ given $\mathbf{X}$ (Hall & Yao, 2005). Given the curse of dimensionality issue due to nonparametric estimation with multiple indices, a sequential estimation procedure, which extracts DSI model structures sequentially to improve the current approximation of the conditional distribution, can be particularly useful. Built upon the proposed DSI model, such strategy has great potential in advancing nonlinear modelling and scalable dimension reduction and is certainly on our research agenda.

Acknowledgements

This work was partially supported by the U.S. National Science Foundation (DMS-1608540, DMS-1613295) and the U.S. National Institutes of Health (U01-HL114494). The authors are grateful to the referees and the editors for their valuable comments and suggestions.

Supporting information

Additional supporting information may be found in the online version of this article at the publisher's web site.

References

Anderson, T. W. (1951). Estimating linear restrictions on regression coefficients for multivariate normal distributions. *Ann. Math. Stat.* **22**, 327–351.

Bunea, F., She, Y. & Wegkamp, M. (2012). Joint variable and rank selection for parsimonious estimation of high dimensional matrices. *Ann. Statist.* **40**, 2359–2388.
Chan, K. -S., Li, M.-C. & Tong, H. (2004). Partially linear reduced-rank regression, Technical Report, Department of Statistics, University of Iowa.
Chen, K., Chan, K.-S. & Stenseth, N. C. (2012). Reduced rank stochastic regression with a sparse singular value decomposition. *J. R. Stat. Soc. Series: B* **74**, 203–221.
Chen, K., Chan, K.-S. & Stenseth, N. C. (2014). Source-sink reconstruction through regularized multi-component regression analysis–with application to assessing whether north sea cod larvae contributed to local fjord cod in skagerrak. *J. Am. Stat. Assoc.* **109**, 560–573.
Chen, L. & Huang, J. Z. (2012). Sparse reduced-rank regression for simultaneous dimension reduction and variable selection. *J. Am. Stat. Assoc.* **107**, 1533–1545.
Fan, J., Yao, Q. & Cai, Z. (2003). Adaptive varying-coefficient linear models. *J. R. Stat. Soc. : Series B (Statistical Methodology)* **65**, 57–80.
Gifi, A. (1990). *Nonlinear Multivariate Analysis*, John and Wiley & Sons, New York.
Hall, P. & Yao, Q. (2005). Approximating conditional distribution functions using dimension reduction. *Ann. Statist.* **33**, 1404–1421.
Härdle, W., Hall, P. & Ichimura, H. (1993). Optimal smoothing in single-index models. *Ann. Statist.* **21**, 157–178.
He, G., Mller, H.-G. & Wang, J.-L. (2003). Functional canonical analysis for square integrable stochastic processes. *J. Multivariate Anal.* **85**, 54–77.
Hotelling, H. (1936). Relations between two sets of variates. *Biometrika* **28**, 321–377.
Hsieh, W. (2000). Nonlinear canonical correlation analysis by neural networks. *Neural Networks* **13**, 1095–1105.
Iaci, R. & Sriram, T. (2013). Robust multivariate association and dimension reduction using density divergences. *J. Multivar. Anal.* **117**, 281–295.
Iaci, R., Sriram, T. & Yin, X. (2010). Multivariate association and dimension reduction: A generalization of canonical correlation analysis. *Biometrics* **66**, 1107–1118.
Ichimura, H. (1993). Semiparametric least squares (sls) and weighted {SLS} estimation of single-index models. *J. Econometrics* **58**, 71–120.
Klein, R. & Shen, C. (2007). *Bias corrections in testing and estimating semiparametric, single index models*.
Klein, R. & Vella, F. (2009). A semiparametric model for binary response and continuous outcomes under index heteroscedasticity. *J. Appl. Econ.* **24**, (5), 735–762.
Li, B., Wen, S. & Zhu, L. (2008). On a projective resampling method for dimension reduction with multivariate responses. *J. Am. Stat. Assoc.* **103**, 1177–1186.
Liang, H., Liu, X., Li, R. & Tsai, C.-L. (2010). Estimation and testing for partially linear single-index models. *Ann. Statist.* **38**, 3811–3836.
Ma, S., Zhang, J., Sun, Z. & Liang, H. (2014). Integrated conditional moment test for partially linear single index models incorporating dimension-reduction. *Electron. J. Statist.* **8**, 523–542.
Ma, Y., Chiou, J.-M. & Wang, N. (2006). Efficient semiparametric estimator for heteroscedastic partially linear models. *Biometrika* **93**, 75–84.
Ma, Y. & Zhu, L. (2012). A semiparametric approach to dimension reduction. *J. Am. Stat. Assoc.* **107**, 168–179.
Ma, Y. & Zhu, L. (2013a). Doubly robust and efficient estimators for heteroscedastic partially linear single-index models allowing high dimensional covariates. *J. R. Stat. Soc. : Series B* **75**, 305–322.
Ma, Y. & Zhu, L. (2013b). Efficient estimation in sufficient dimension reduction. *Ann. Statist.* **41**, 250–268.
Mandal, A. & Cichocki, A. (2013). Non-linear canonical correlation analysis using Alpha-Beta divergences. *Entropy* **15**, 2788–2804.
MATLAB. (2010). *version 7.10.0 (R2010a)*, The MathWorks Inc., Natick, Massachusetts.
Mukherjee, A. & Zhu, J. (2011). Reduced rank ridge regression and its kernel extensions. *Stat. Anal. Data Min.* **4**, 612–622.
Newey, W. K. & Stoker, T. M. (1993). Efficiency of weighted average derivative estimators and index models. *Econometrica* **61**, (5), 1199–1223.
Reinsel, G. C. & Velu, P. (1998). *Multivariate Reduced-Rank Regression: Theory and Applications*, Springer, New York.
Stute, W. & Zhu, L.-X. (2005). Nonparametric checks for single-index models. *Ann. Statist.* **33**, 1048–1083.
Witten, D. M., Tibshirani, R. J. & Hastie, T. J. (2009). A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis. *Biostatistics* **10**, 515–534.
Xia, Y. (2008). A semiparametric approach to canonical analysis. *J. R. Stat. Soc. : Series B* **70**, 519–543.

Yeh, I.-C. (2006). Exploring concrete slump model using artificial neural networks. *J. Comput. Civ. Eng.* **20**, 217–221.
Yeh, I.-C. (2007). Modeling slump flow of concrete using second-order regressions and artificial neural networks. *Cem. Concr. Compos.* **29**, 474–480.
Yuan, M., Ekici, A., Lu, Z. & Monteiro, R. (2007). Dimension reduction and coefficient estimation in multivariate linear regression. *J. R. Stat. Soc. : Series B* **69**, 329–346.
Zhu, H., Khondker, Z., Lu, Z. & Ibrahim, J. G. (2014). Bayesian generalized low rank regression models for neuroimaging phenotypes and genetic markers. *J. Am. Stat. Assoc.* **109**, 977–990.

Received August 2015, in final form March 2016

Kun Chen, Department of Statistics, University of Connecticut, 215 Glenbrook Road U-4120, Storrs, CT 06269, USA.
E-mail: kun.chen@uconn.edu

SPi	Journal Code:				Article ID					Dispatch 27.03.17	CE: Nicol John Luzano
	S	J	O	S	1	2	2	7	6	No. of Pages: 2	ME:

Scandinavian Journal of Statistics
doi: 10.1111/sjos.12276

Correction Note to 'Analysis of Double Single Index Models'

KUN CHEN AND YANYUAN MA

In Chen & Ma (2017), a duplicate of Figure 3 was incorrectly shown as Figure 4. The correct Figure 4 is given below. We apologize for this error.

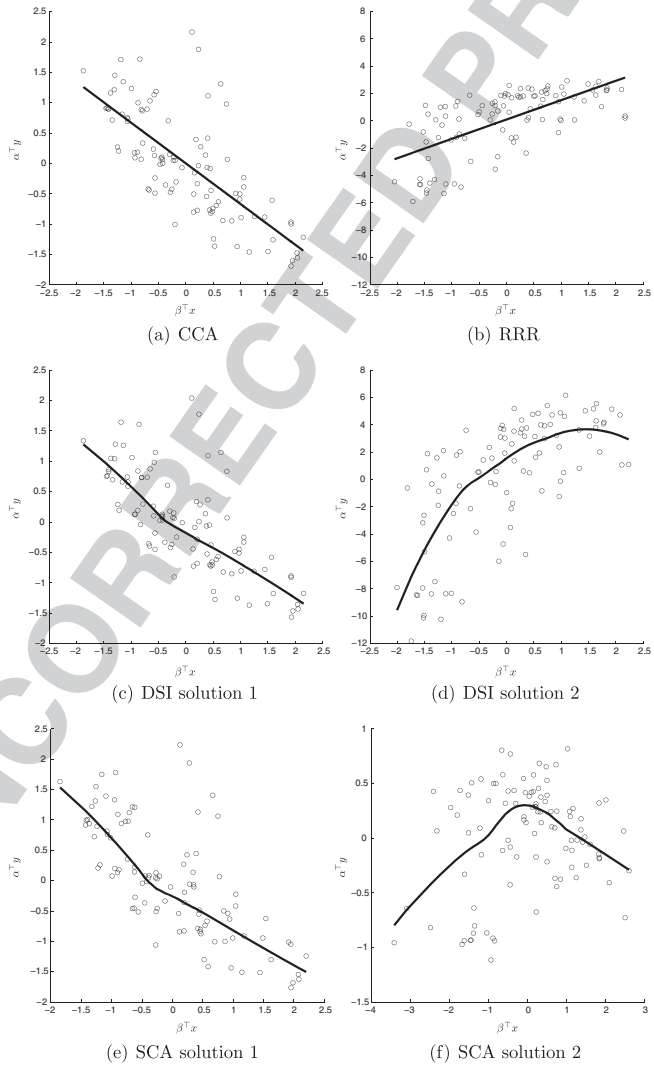


Fig. 4. Scatter plots along the estimated single-index directions for the slump test data analysis.

References

Chen, K & Ma, Y. (2017). Analysis of double single index models. *Scand. J. Stat.* **44**, 1–20.