01 Jan 2023

Variable-Dependent Partial Dimension Reduction

Lu Li

Kai Tan

Xuerong Meggie Wen
Missouri University of Science and Technology, wenx@mst.edu

Zhou Yu

Recommended Citation

L. Li et al., "Variable-Dependent Partial Dimension Reduction," *Test*, Springer, Jan 2023.
The definitive version is available at https://doi.org/10.1007/s11749-022-00841-y

TEST
https://doi.org/10.1007/s11749-022-00841-y

ORIGINAL PAPER

Variable-dependent partial dimension reduction

Lu Li[1] · **Kai Tan**[2] · **Xuerong Meggie Wen**[3] · **Zhou Yu**[4,5]

Received: 17 February 2022 / Accepted: 29 November 2022

Abstract

Sufficient dimension reduction reduces the dimension of a regression model without loss of information by replacing the original predictor with its lower-dimensional linear combinations. Partial (sufficient) dimension reduction arises when the predictors naturally fall into two sets $\mathbf{X}$ and $\mathbf{W}$, and pursues a partial dimension reduction of $\mathbf{X}$. Though partial dimension reduction is a very general problem, only very few research results are available when $\mathbf{W}$ is continuous. To the best of our knowledge, none can deal with the situation where the reduced lower-dimensional subspace of $\mathbf{X}$ varies with $\mathbf{W}$. To address such issue, we in this paper propose a novel variable-dependent partial dimension reduction framework and adapt classical sufficient dimension reduction methods into this general paradigm. The asymptotic consistency of our method is investigated. Extensive numerical studies and real data analysis show that our variable-dependent partial dimension reduction method has superior performance compared to the existing methods.

Keywords Directional regression · Sliced average variance estimation · Sliced inverse regression · Sufficient dimension reduction · Order determination

Mathematics Subject Classification 62G08 · 62G20

✉ Zhou Yu
zyu@stat.ecnu.edu.cn

1 School of Mathematical Sciences, Shanghai Jiao Tong University, 800 Dongchuan Rd., Shanghai 200240, China

2 Department of Statistics, Rutgers, The State University of New Jersey, 110 Frelinghuysen Rd., Piscataway, NJ 08854, USA

3 Department of Mathematics and Statistics, Missouri University of Science and Technology, Rolla, MO 65409, USA

4 School of Statistics, East China Normal University, 3663 North Zhongshan Rd., Shanghai, China

5 Key Laboratory of Advanced Theory and Application in Statistics and Data Science, MOE, Shanghai, China

Published online: 10 January 2023

1 Introduction

The rapid developments of brain imaging, microarray data analysis, computer vision, network analysis, econometrics, and many other applications call for the analysis of high-dimensional data. Sufficient dimension reduction (SDR) (Li 1991; Cook 1998) is arguably one of the most important tools in analyzing high-dimensional data. Let Y be a univariate response, $\mathbf{X} = (X_1, \ldots, X_p)^{\mathrm{T}} \in \mathbb{R}^p$ be a p-dimensional predictors. Sufficient dimension reduction methods aim to find a lower-dimensional subspace of $\mathbf{X}$ without loss of information on the conditional distribution of $Y|\mathbf{X}$, and without pre-specifying a model for the regression. This subspace is then called a dimension reduction subspace for the regression. The goal of SDR is to search for the smallest dimension reduction subspace, the central subspace (CS, $\mathcal{S}_{Y|\mathbf{X}}$) and its dimension d, which is called the structural dimension of the regression. We refer readers to Cook (1998) for more details. Many methods have been developed in the past two decades due to the ubiquity of large high-dimension data sets which are now more readily available than in the past. To name a few: sliced inverse regression [SIR; Li (1991)], sliced average variance estimation [SAVE; Cook and Weisberg (1991)], minimum average variance estimation [MAVE; Xia et al. (2002)], the kth moment estimation (Yin and Cook 2002, 2003), inverse regression (Cook and Ni 2005), directional regression [DR; Li and Wang (2007)], sliced regression [SR; Wang and Xia (2008)], likelihood acquired directions [LAD; Cook and Forzani (2009)], and semiparametric approaches of Ma and Zhu (2012, 2013a, b, 2014). More detailed discussion can be found in Xue et al. (2018).

Partial dimension reduction (PDR) (Chiaromonte et al. 2002; Wen and Cook 2007; Feng et al. 2013) arises when the predictors naturally fall into two groups, $\mathbf{X} = (X_1, \ldots, X_p)^{\mathrm{T}}$ and $\mathbf{W} = (W_1, \ldots, W_q)^{\mathrm{T}}$, and a partial dimension reduction of $\mathbf{X}$ is pursued. This might happen when $\mathbf{W}$ plays a particular role in the regression and must, therefore, be shielded from the reduction process. Considering the Boston Housing dataset (Feng et al. 2013), which was collected by the U.S. Census Service concerning housing in the 18 areas of Boston, where the goal was to study how the house prices are affected by certain given attributes regarding those houses. Among all those features, it is well known that *Crime rate* ($\mathbf{W}$) plays an important role in the housing price, hence it should be treated discriminately and the dimension reduction should focus on the remaining features ($\mathbf{X}$).

To be specific, PDR performs regression of Y on $(\mathbf{X}, \mathbf{W})$ by seeking a projection $\mathbf{P}_{\mathcal{S}}\mathbf{X}$ of $\mathbf{X}$ that preserves information on $Y|(\mathbf{X}, \mathbf{W})$, where $\mathbf{P}_{\mathcal{S}}$ indicates the projection onto the subspace $\mathcal{S}$ in the usual inner product. If the intersection of all subspaces $\mathcal{S} \subseteq \mathbb{R}^p$ such that

$$Y \perp\!\!\!\perp \mathbf{X}|(\mathbf{P}_{\mathcal{S}}\mathbf{X}, \mathbf{W}), \tag{1}$$

also satisfies condition (1), we call it the partial central subspace, and denote it by $\mathcal{S}^{(\mathbf{W})}_{Y|\mathbf{X}}$. $\dim\{\mathcal{S}^{(\mathbf{W})}_{Y|\mathbf{X}}\} = d$ is called the structural dimension of the partial central subspace. The concept of partial central subspace was first proposed by Chiaromonte et al. (2002) to deal with regressions with a mixture of continuous ($\mathbf{X}$) and categorical predictors ($\mathbf{W}$).

Feng et al. (2013) developed a method called PDEE to incorporate the continuous $\mathbf{W}$ scenario via a dichotomization transformation. Though PDEE widens the application of PDR, it could not deal with the situation where $\mathbf{W}$ is continuous, and $Y|(\mathbf{X}, \mathbf{W} = \mathbf{w})$ is a function of $\mathbf{w}$, which is often the case in real-world applications.

To overcome the limitations mentioned above, we propose the concept of variable-dependent partial dimension reduction, where the variable-dependent partial CS, $\mathcal{S}_{Y|\mathbf{X}}^{(\mathbf{W}=\mathbf{w})}$ is allowed to vary with $\mathbf{w}$. Hence the aim of variable-dependent partial dimension reduction is to find a matrix of smooth functions of $\mathbf{W} = \mathbf{w}$ with minimum rank, $\mathbf{B}(\mathbf{w}) \in \mathbb{R}^{p\times d(\mathbf{w})}$, such that

$$Y \perp\!\!\!\perp \mathbf{X}|(\mathbf{B}^{\mathrm{T}}(\mathbf{w})\mathbf{X}, \mathbf{W} = \mathbf{w}). \tag{2}$$

Then $d(\mathbf{w}) \triangleq \mathrm{rank}\{\mathbf{B}(\mathbf{w})\}$ is the structural dimension function of the variable-dependent partial CS. It is worth noting that the covariance matrix $\mathrm{Cov}(\mathbf{X}|\mathbf{W} = \mathbf{w})$, the column space of $\mathbf{B}(\mathbf{w})$ and the structural dimension $d(\mathbf{w})$ may all vary as $\mathbf{w}$ changes, which poses a great challenge for the estimation procedure. Furthermore, it is easy for us to see that $\mathcal{S}_{Y|\mathbf{X}}^{(\mathbf{W}=\mathbf{w})} \subseteq \mathcal{S}_{Y|\mathbf{X}}^{(\mathbf{W})}$. Therefore, variable-dependent partial dimension reduction can be regarded as an extension of the original partial dimension reduction.

The contribution of this paper is twofold. First, we adapt three classical SDR methods into the new framework to develop variable-dependent partial dimension reduction methods and establish the corresponding asymptotic normality and consistency properties rigorously. Second, we propose to determine the structural dimension, $d(\mathbf{w})$, by a nonparametric version of the ladle estimator (Luo and Li 2016), and also derive the consistency property for our nonparametric ladle estimator as well. More importantly, our proposed variable-dependent partial dimension reduction method has outstanding advantages in dealing with the case that $d(\mathbf{w})$ vary with different $\mathbf{w}$. To the best of our knowledge, this is first way to solve this kind of sufficient dimension reduction problem. Moreover, our proposed method works well to estimate $\mathcal{S}_{Y|\mathbf{X}}^{(\mathbf{W}=\mathbf{w})}$ and $d(\mathbf{w})$.

The rest of this paper is organized as follows. In Sect. 2, we first introduce the principles of variable-dependent partial dimension reduction, then develop variable-dependent partial SIR, and also propose its estimation schemes along with the large sample theories. In Sect. 3, we develop the nonparametric ladle estimator to determine the structural dimension of variable-dependent partial dimension reduction subspace. Section 4 focuses on how to conduct the bandwidth selection involved in the kernel estimation. Section 5 presents the finite sample performance of our proposed methods via extensive simulation studies. To illustrate the efficiency of our proposed methods, four real data analyses are conducted in Sect. 6. Results extended to variable-dependent partial SAVE and DR are included in Appendix. For the ease of presentation, we defer all proofs to Appendix.

2 The principle of variable-dependent partial dimension reduction

Let $(\mathbf{X}_{\mathbf{w}}, Y_{\mathbf{w}})$ denote a generic pair distributed like $(\mathbf{X}, Y)|(\mathbf{W} = \mathbf{w})$, and $\mathcal{S}_{Y|(\mathbf{X},\mathbf{W}=\mathbf{w})} = \mathcal{S}_{Y_{\mathbf{w}}|\mathbf{X}_{\mathbf{w}}}$. As we discussed in Sect. 1, the aim of variable-dependent

PDR is to find a subspace spanned by the columns of matrix $\mathbf{B}(\mathbf{w}) \in \mathbb{R}^{p\times d(\mathbf{w})}$ such that

$$Y_{\mathbf{w}} \perp\!\!\!\perp \mathbf{X}_{\mathbf{w}} | \mathbf{B}^{\mathrm{T}}(\mathbf{w})\mathbf{X}_{\mathbf{w}}, \tag{3}$$

where $\mathbf{B}(\mathbf{w})$ is a matrix of smooth functions of $\mathbf{w}$. The meaning of $\mathbf{X}_{\mathbf{w}}$ and $Y_{\mathbf{w}}$ may not be intuitive. Hence we present a simple example here. Suppose W is gender, predictor $\mathbf{X}$ is a vector consisting of several other characteristics of a person, such as weight, height, age, etc., response variable Y is the blood pressure. If w is male, then $\mathbf{X}_w$ represents the male's characteristics and Y_w is the male's blood pressure.

We employ nonparametric covariance models to analyze this variable-dependent scenario. Let $\mathbf{m}(\mathbf{w}) = \left(m_1(\mathbf{w}), \ldots, m_p(\mathbf{w})\right)^{\mathrm{T}}$ and $\boldsymbol{\Sigma}_{\boldsymbol{w}} = \{\sigma_{ij}(\mathbf{w})\}_{p\times p}$ denote the mean and covariance of $\mathbf{X}_{\mathbf{w}}$, where both $\mathbf{m}(\mathbf{w})$ and $\boldsymbol{\Sigma}_{\boldsymbol{w}}$ are smooth functions of $\mathbf{w}$. Equation (3) implies the reduction of the predictor from $\mathbf{X}_{\mathbf{w}}$ to $\mathbf{B}^{\mathrm{T}}(\mathbf{w})\mathbf{X}_{\mathbf{w}}$. If $\boldsymbol{\Sigma}_{\boldsymbol{w}}$ is invertible, one can also work with the standardized data $\mathbf{Z}_{\mathbf{w}} = \boldsymbol{\Sigma}_{\boldsymbol{w}}{}^{-1/2}\{\mathbf{X} - \mathbf{m}(\mathbf{w})\}$ to obtain $\mathcal{S}_{Y_{\mathbf{w}}|\mathbf{Z}_{\mathbf{w}}}$ and then recover the variable-dependent partial CS, $\mathcal{S}_{Y_{\mathbf{w}}|\mathbf{X}_{\mathbf{w}}}$ by the well-known invariance property $\mathcal{S}_{Y_{\mathbf{w}}|\mathbf{X}_{\mathbf{w}}} = \boldsymbol{\Sigma}_{\boldsymbol{w}}{}^{-1/2}\mathcal{S}_{Y_{\mathbf{w}}|\mathbf{Z}_{\mathbf{w}}}$ [see Cook (1998), Proposition 6.1]. Note that the estimation of $\mathcal{S}_{Y_{\mathbf{w}}|\mathbf{X}_{\mathbf{w}}}$ consists of two parts, the order determination for $d(\mathbf{w})$ and the basis estimation for $\mathcal{S}_{Y_{\mathbf{w}}|\mathbf{X}_{\mathbf{w}}}$. We first consider the basis estimation assuming $d(\mathbf{w})$ is known, then propose an order determination method for $d(\mathbf{w})$.

2.1 Variable-dependent partial SIR

We adopt three popular sufficient dimension reduction approaches, SIR (Li 1991), SAVE (Cook and Weisberg 1991), DR (Li and Wang 2007), to perform variable-dependent PDR. We now illustrate the variable-dependent partial SIR in full details here. The development of variable-dependent partial SAVE and DR is put in the supplementary file.

Let $\mathbf{B}(\mathbf{w}) \in \mathbb{R}^{p\times d(\mathbf{w})}$ be a matrix such that $\mathrm{Span}(\mathbf{B}(\mathbf{w})) = \mathcal{S}_{\mathbf{X}_{\mathbf{w}}|Y_{\mathbf{w}}}$. Prior to the main development of our variable-dependent partial dimension reduction methods, we first present the following two assumptions,

(A1) (Linear conditional mean) $E\{\mathbf{X}_{\mathbf{w}}|\mathbf{B}^{\mathrm{T}}(\mathbf{w})\mathbf{X}_{\mathbf{w}}\}$ is a linear function of $\mathbf{B}^{\mathrm{T}}(\mathbf{w})\mathbf{X}_{\mathbf{w}}$.
(A2) (Constant conditional variance) $\mathrm{Cov}\{\mathbf{X}_{\mathbf{w}}|\mathbf{B}^{\mathrm{T}}(\mathbf{w})\mathbf{X}_{\mathbf{w}}\}$ is a nonrandom matrix.

Condition (A1) has been commonly assumed in sufficient dimension reduction literature and is indispensable for almost all inverse-regression-based methods. Condition (A2) is similar to (A1) in nature and is important for all the second-order sufficient dimension reduction methods. Both conditions are guaranteed when $\mathbf{X}_{\mathbf{w}}$ is normally distributed. Since conditions (A1) and (A2) have to hold for all $\mathbf{w}$, they seem stronger than the standard assumptions used within the general sufficient dimension reduction framework. However, it is not the case. Hall and Li (1993) claimed that as long as the dimension of $\mathbf{X}$ is large, standard conditions in the sufficient dimension literature approximately hold. Therefore, even if $\mathbf{X}_{\mathbf{w}}$ violates the normal distribution, as long as the dimension of $\mathbf{X}$ is sufficiently large, we can also ensure conditions (A1) and (A2). More discussions about conditions (A1) and (A2) can be found in Chiaromonte et al. (2002), Li et al. (2003), Li and Dong (2009), Dong and Li (2010) and Li (2018).

We now briefly review the development of SIR (Li 1991). The main idea of SIR is to work with the inverse regression, the conditional distribution of $\mathbf{X}|Y$, and in particular by examining the kernel matrix $\mathrm{Cov}\{E(\mathbf{X}|Y)\}$. SIR procedure starts with the partition of the response Y. Let $\{J_1, J_2, \ldots, J_H\}$ be a measurable partition of the sample space of Y, consider the discretized version $\widetilde{Y} = \sum_{l=1}^{H} l \cdot \mathbf{1}(Y \in J_l)$, where $\mathbf{1}(.)$ is the indicator function. If Y is categorical or H is sufficiently large ($H \geq d+1$), Bura and Cook (2001) and Cook and Forzani (2009) verified that there is no loss of information for identifying $\mathcal{S}_{Y|\mathbf{X}}$ when Y is replaced with $\widetilde{Y}$.

Let $\mathbf{P}_{\mathbf{B}(\mathbf{w})}$ be the projection onto $\mathcal{S}_{Y_\mathbf{w}|\mathbf{X}_\mathbf{w}}$ with respect to the inner product $\langle a, b\rangle := a^{\mathrm{T}}\boldsymbol{\Sigma}_{\boldsymbol{w}} b$, assuming (A1), the following proposition states that the random vector $\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\{\mathrm{E}(\mathbf{X}_\mathbf{w}|Y_\mathbf{w}) - \mathbf{m}(\mathbf{w})\}$ belongs to $\mathcal{S}_{Y_\mathbf{w}|\mathbf{X}_\mathbf{w}}$ almost surely.

Proposition 1 *Given* $\mathbf{W} = \mathbf{w}$*, suppose that the linear conditional mean condition* (A1) *holds, then*

$$\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}(E\{\mathbf{X}|(Y, \mathbf{W} = \mathbf{w})\} - \mathbf{m}(\mathbf{w})) = \mathbf{P}_{\mathbf{B}(\mathbf{w})}\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}(E\{\mathbf{X}|(Y, \mathbf{W} = \mathbf{w})\} - \mathbf{m}(\mathbf{w})).$$

Proposition 1 indicates that the random vector $\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}(E(\mathbf{X}_\mathbf{w}|Y_\mathbf{w}) - \mathbf{m}(\mathbf{w}))$ belongs to the range of the projection operator $\mathbf{P}_{\mathbf{B}(\mathbf{w})}$, which is actually $\mathcal{S}_{Y_\mathbf{w}|\mathbf{X}_\mathbf{w}}$. Consequently, the column space of the matrix $\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\mathrm{Cov}\{E(\mathbf{X}_\mathbf{w}|Y_\mathbf{w})\}$ is a subspace of $\mathcal{S}_{Y_\mathbf{w}|\mathbf{X}_\mathbf{w}}$. The proof of Proposition 1 relies on Li and Dong (2009), and is provided in Appendix.

Motivated by this finding, we now construct the following kernel matrix for variable-dependent partial SIR,

$$\mathbf{M}_{\mathrm{SIR}}(\mathbf{w}) \triangleq \mathrm{Cov}\{E(\mathbf{X}|\widetilde{Y}, \mathbf{w})\} = \sum_{l=1}^{H} P_{l,\mathbf{w}}\mathbf{V}_{l,\mathbf{w}}\mathbf{V}_{l,\mathbf{w}}^{\mathrm{T}} - \mathbf{m}(\mathbf{w})\mathbf{m}(\mathbf{w})^{\mathrm{T}}, \tag{4}$$

where $P_{l,\mathbf{w}} = \mathrm{pr}(\widetilde{Y} = l|\mathbf{w})$, $\mathbf{V}_{l,\mathbf{w}} = \mathrm{E}\{\mathbf{X}|\widetilde{Y} = l, \mathbf{w}\}$ and $\mathbf{m}(\mathbf{w}) = \mathrm{E}(\mathbf{X}_\mathbf{w})$. Note that the term $\mathbf{V}_{l,\mathbf{w}} = \mathrm{E}\{\mathbf{X}|\widetilde{Y} = l, \mathbf{w}\}$ contains two conditional variables, and thus makes it hard to deal with. However, this difficulty can be overcome by using the following proposition, whose proof is provided in Appendix.

Proposition 2 *Given* $\mathbf{W} = \mathbf{w}$*, for each* $l = 1, 2, \ldots, H$*, we have*

$$E(\mathbf{X}|\widetilde{Y} = l, \mathbf{w}) = \frac{E\{\mathbf{X}\mathbf{1}(\widetilde{Y} = l)|\mathbf{w}\}}{E\{\mathbf{1}(\widetilde{Y} = l)|\mathbf{w}\}}. \tag{5}$$

Proposition 2 shows that the term $\mathbf{V}_{l,\mathbf{w}}$ in (4) can be written as a fraction of simple conditional expectations. We can rewrite the kernel matrix $\mathbf{M}_{\mathrm{SIR}}(\mathbf{w})$ in the following form:

$$\mathbf{M}_{\mathrm{SIR}}(\mathbf{w}) = \sum_{l=1}^{H} \frac{\mathbf{U}_{l,\mathbf{w}}\mathbf{U}_{l,\mathbf{w}}^{\mathrm{T}}}{P_{l,\mathbf{w}}} - \mathbf{m}(\mathbf{w})\mathbf{m}(\mathbf{w})^{\mathrm{T}}, \tag{6}$$

where $\mathbf{U}_{l,\mathbf{w}} = E\{\mathbf{X}\mathbf{1}(\widetilde{Y} = l)|\mathbf{w}\}$.

Let $\{(Y_i, \mathbf{X}_i, \mathbf{W}_i), i = 1, \ldots, n\}$ be random samples from $(Y, \mathbf{X}, \mathbf{W})$. Since the structure of $\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\mathbf{M}_{\mathrm{SIR}}(\mathbf{w})$ is variable dependent, we employ the nonparametric covariance model Yin et al. (2010) for estimation. Specifically, we adopt the following Nadaraya–Watson estimator Nadaraya (1964), Watson (1964) of $\mathbf{m}(\mathbf{w})$

$$\widehat{\mathbf{m}}(\mathbf{w}) = \frac{\sum_{i=1}^{n} \mathbf{X}_i K_h(\mathbf{W}_i - \mathbf{w})}{\sum_{i=1}^{n} K_h(\mathbf{W}_i - \mathbf{w})}.$$

Similarly, the Nadaraya–Watson estimators of $\mathbf{U}_{l,\mathbf{w}}$ and $P_{l,\mathbf{w}}$ are given by

$$\widehat{\mathbf{U}}_{l,\mathbf{w}} = \frac{\sum_{i=1}^{n} \mathbf{X}_i \mathbf{1}(\widetilde{Y}_i = l) K_h(\mathbf{W}_i - \mathbf{w})}{\sum_{i=1}^{n} K_h(\mathbf{W}_i - \mathbf{w})}, \quad \widehat{P}_{l,\mathbf{w}} = \frac{\sum_{i=1}^{n} \mathbf{1}(\widetilde{Y}_i = l) K_h(\mathbf{W}_i - \mathbf{w})}{\sum_{i=1}^{n} K_h(\mathbf{w}_i - \mathbf{w})}.$$

Then it's straightforward to obtain the sample estimator of $\mathbf{M}_{\mathrm{SIR}}(\mathbf{w})$ by substituting $\widehat{\mathbf{m}}(\mathbf{w})$, $\widehat{\mathbf{U}}_{l,\mathbf{w}}$, and $\widehat{P}_{l,\mathbf{w}}$ into Eq. (6), that is,

$$\widehat{\mathbf{M}}_{\mathrm{SIR}}(\mathbf{w}) = \sum_{l=1}^{H} \frac{\widehat{\mathbf{U}}_{l,\mathbf{w}}\widehat{\mathbf{U}}_{l,\mathbf{w}}^{\mathrm{T}}}{\widehat{P}_{l,\mathbf{w}}} - \widehat{\mathbf{m}}(\mathbf{w})\widehat{\mathbf{m}}(\mathbf{w})^{\mathrm{T}}. \tag{7}$$

Remark 1 As for the estimation of the conditional covariance matrix, one may use different bandwidths for different elements of $\boldsymbol{\Sigma}_{\boldsymbol{w}}$. However, the resulting estimate with different bandwidth is not guaranteed to be positive definite (Li and Zhu 2007), which is the desired property in practice. Thus, we suggest using the same bandwidth for all elements. And the selection of bandwidth will be discussed in Sect. 4.

Recall that $\mathcal{S}_{Y_{\mathbf{w}}|\mathbf{X}_{\mathbf{w}}}$ is a $d(\mathbf{w})$-dimensional subspace of $\mathbb{R}^p$. Proposition 1 leads us to consider the singular value decomposition of $\widehat{\boldsymbol{\Sigma}}_{\mathbf{w}}^{-1}\widehat{\mathbf{M}}_{\mathrm{SIR}}(\mathbf{w})$. Let

$$\begin{aligned}
\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\mathbf{M}_{\mathrm{SIR}}(\mathbf{w}) &= \sum_{k=1}^{p} \boldsymbol{\lambda}_k^{\mathrm{SIR}}(\mathbf{w})\boldsymbol{\beta}_k^{\mathrm{SIR}}(\mathbf{w})(\boldsymbol{\eta}_k^{\mathrm{SIR}}(\mathbf{w}))^{\mathrm{T}}, \\
\boldsymbol{\lambda}_1^{\mathrm{SIR}}(\mathbf{w}) \geq \cdots \geq \boldsymbol{\lambda}_{d(\mathbf{w})}^{\mathrm{SIR}}(\mathbf{w}) &= 0 = \cdots = \boldsymbol{\lambda}_p^{\mathrm{SIR}}(\mathbf{w}), \\
\widehat{\boldsymbol{\Sigma}}_{\mathbf{w}}^{-1}\widehat{\mathbf{M}}_{\mathrm{SIR}}(\mathbf{w}) &= \sum_{k=1}^{p} \widehat{\boldsymbol{\lambda}}_k^{\mathrm{SIR}}(\mathbf{w})\widehat{\boldsymbol{\beta}}_k^{\mathrm{SIR}}(\mathbf{w})(\widehat{\boldsymbol{\eta}}_k^{\mathrm{SIR}}(\mathbf{w}))^{\mathrm{T}}, \\
\widehat{\boldsymbol{\lambda}}_1^{\mathrm{SIR}}(\mathbf{w}) \geq \cdots \geq \widehat{\boldsymbol{\lambda}}_{d(\mathbf{w})}^{\mathrm{SIR}}(\mathbf{w}) &\geq \cdots \geq \widehat{\boldsymbol{\lambda}}_p^{\mathrm{SIR}}(\mathbf{w}),
\end{aligned}$$

be the singular value decomposition of $\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\mathbf{M}_{\mathrm{SIR}}(\mathbf{w})$ and $\widehat{\boldsymbol{\Sigma}}_{\mathbf{w}}^{-1}\widehat{\mathbf{M}}_{\mathrm{SIR}}(\mathbf{w})$, respectively.

Then we can use $\mathrm{Span}\{\widehat{\boldsymbol{\beta}}_1^{\mathrm{SIR}}(\mathbf{w}), \ldots, \widehat{\boldsymbol{\beta}}_{d(\mathbf{w})}^{\mathrm{SIR}}(\mathbf{w})\}$ to estimate $\mathcal{S}_{Y_{\mathbf{w}}|\mathbf{X}_{\mathbf{w}}}$. By large sample theory and singular value decomposition, the asymptotic normality of $\widehat{\mathbf{M}}_{\mathrm{SIR}}(\mathbf{w})$ and the asymptotic expansion of $\widehat{\boldsymbol{\beta}}_k^{\mathrm{SIR}}(\mathbf{w})$ with $k = 1, \ldots, d(\mathbf{w})$ are presented in Theorem 3.

Theorem 3 *Let $G \subset \{\mathbf{w} : f(\mathbf{w}) > 0\}$ be a compact subset on the support of $\mathbf{W}$, where $f(\mathbf{w})$ is the density of $\mathbf{W}$. Under the condition* (A1) *and assumptions* (C1)–(C8) *listed in Appendix, we have*

$$\sqrt{nh}\left(\mathbf{vech}\{\widehat{\mathbf{M}}_{\mathrm{SIR}}(\mathbf{w})\} - \mathbf{vech}\{\mathbf{M}_{\mathrm{SIR}}(\mathbf{w})\} - \mathbf{vech}\{\mathbf{B}_{\mathrm{SIR}}(\mathbf{w})\}\right) \xrightarrow{d} \mathbf{N}\left(\mathbf{0}, f^{-1}(\mathbf{w})\omega_0\mathbf{C}^{\mathrm{SIR}}(\mathbf{w})\right).$$

Assume that $\mathbf{X}_{\mathbf{w}}$ has finite fourth moment and all the nonzero eigenvalues of $\mathbf{M}_{\mathrm{SIR}}(\mathbf{w})$ are distinct, then for $k = 1, \ldots, d(\mathbf{w})$, we have

$$\sqrt{nh}\left(\widehat{\boldsymbol{\beta}}_k^{\mathrm{SIR}}(\mathbf{w}) - \boldsymbol{\beta}_k^{\mathrm{SIR}}(\mathbf{w}) - \mathbf{B}_k^{\mathrm{SIR}}(\mathbf{w})\right) \xrightarrow{d} \mathbf{N}(\mathbf{0}, \boldsymbol{\Sigma}_k^{\mathrm{SIR}}(\mathbf{w})),$$

where $\mathbf{vech}(\cdot)$ is the vectorization of the upper triangular part of a matrix, and the closed forms of $\mathbf{B}_{\mathrm{SIR}}(\mathbf{w})$, ω_0, $\mathbf{C}^{\mathrm{SIR}}(\mathbf{w})$, $\mathbf{B}_k^{\mathrm{SIR}}(\mathbf{w})$ and $\boldsymbol{\Sigma}_k^{\mathrm{SIR}}(\mathbf{w})$ are provided by (9), (11), (12), (22) and (23) in Appendix, respectively.

We now present the uniform convergence result in Theorem 4. Denote

$$\widehat{T}_l(\mathbf{w}) = \frac{1}{nh}\sum_{j=1}^{n}\mathbf{1}(\widetilde{Y}_j = l)K_h(\mathbf{w} - \mathbf{w}_j),$$
$$\widehat{f}(\mathbf{w}) = \frac{1}{nh}\sum_{j=1}^{n}K_h(\mathbf{w} - \mathbf{w}_j),$$
$$\widehat{g}_i(\mathbf{w}) = \frac{1}{nh}\sum_{j=1}^{n}X_{ij}K_h(\mathbf{w} - \mathbf{w}_j),$$
$$\widehat{R}_{il}(\mathbf{w}) = \frac{1}{nh}\sum_{j=1}^{n}X_{ij}\mathbf{1}(\widetilde{Y}_j = l)K_h(\mathbf{w} - \mathbf{w}_j).$$

Let $\widehat{\mathbf{R}}_l(\mathbf{w}) = (\widehat{R}_{1l}(\mathbf{w}), \ldots, \widehat{R}_{pl}(\mathbf{w}))^{\mathrm{T}}$ and $\widehat{\mathbf{g}}(\mathbf{w}) = (\widehat{g}_1(\mathbf{w}), \ldots, \widehat{g}_p(\mathbf{w}))^{\mathrm{T}}$, then $\widehat{\mathbf{U}}_{l,\mathbf{w}} = \frac{\widehat{\mathbf{R}}_l(\mathbf{w})}{\widehat{f}(\mathbf{w})}$, $\widehat{P}_{l,\mathbf{w}} = \frac{\widehat{T}_l(\mathbf{w})}{\widehat{f}(\mathbf{w})}$, $\widehat{\mathbf{m}}(\mathbf{w}) = \frac{\widehat{\mathbf{g}}(\mathbf{w})}{\widehat{f}(\mathbf{w})}$ and

$$\widehat{\mathbf{M}}_{\mathrm{SIR}}(\mathbf{w}) = \sum_{l=1}^{H}\frac{\widehat{\mathbf{R}}_l(\mathbf{w})\widehat{\mathbf{R}}_l(\mathbf{w})^{\mathrm{T}}}{\widehat{T}_l(\mathbf{w})\widehat{f}(\mathbf{w})} - \frac{\widehat{\mathbf{g}}(\mathbf{w})\widehat{\mathbf{g}}(\mathbf{w})^{\mathrm{T}}}{\widehat{f}^2(\mathbf{w})}.$$

Easy to see $\mathbf{R}_l(\mathbf{w}) = \mathbf{U}_{l,\mathbf{w}}f(\mathbf{w})$, $T_l(\mathbf{w}) = P_{l,\mathbf{w}}f(\mathbf{w})$ and $\mathbf{g}(\mathbf{w}) = \mathbf{m}(\mathbf{w})f(\mathbf{w})$. To prove the uniform convergence result, we need the following conditions.

(A3) $R_{jl}(\mathbf{w})$, $g_j(\mathbf{w})$, $T_l(\mathbf{w})$, $j = 1, \ldots, p$, $l = 1, \ldots, H$, and $f(\mathbf{w})$ are thrice differentiable and their third derivatives satisfy the following conditions: there exists a neighborhood of the origin, say $I^{\mathbf{U}}$, and a constant $c > 0$ such that, for any

$I^{\mathbf{u}} \in I^{\mathbf{U}}$,

$$|f^{(3)}(\mathbf{w}+I^{\mathbf{u}}) - f^{(3)}(\mathbf{w})| \le c|I^{\mathbf{u}}|, \quad |R_{jl}^{(3)}(\mathbf{w}+I^{\mathbf{u}}) - R_{jl}^{(3)}(\mathbf{w})| \le c|I^{\mathbf{u}}|,$$
$$|g_j^{(3)}(\mathbf{w}+I^{\mathbf{u}}) - g_j^{(3)}(\mathbf{w})| \le c|I^{\mathbf{u}}|, \quad |T_l^{(3)}(\mathbf{w}+I^{\mathbf{u}}) - T_l^{(3)}(\mathbf{w})| \le c|I^{\mathbf{u}}|,$$

where $j = 1, \ldots, p, l = 1, \ldots, H$.

(A4) For each pair $1 \le s, t \le p$ and for any $I^{\mathbf{u}} \in I^{\mathbf{U}}$,

$$|R_{sl}(\mathbf{w}+I^{\mathbf{u}})R_{tl}(\mathbf{w}+I^{\mathbf{u}}) - R_{sl}(\mathbf{w})R_{tl}(\mathbf{w})| \le c|I^{\mathbf{u}}|,$$
$$|g_s(\mathbf{w}+I^{\mathbf{u}})g_t(\mathbf{w}+I^{\mathbf{u}}) - g_s(\mathbf{w})g_t(\mathbf{w})| \le c|I^{\mathbf{u}}|.$$

(A5) The continuous kernel function $K(\cdot)$ has the following properties:

(a) the support of $K(\cdot)$ is the interval $[-1, 1]$;
(b) $K(\cdot)$ is symmetric about 0;
(c) $\int_{-1}^{1} K(z)dz = 1$, and $\int_{-1}^{1} z^i K(z)dz = 0, i = 1, 2, 3$.

(A6) As $n \to \infty$, $h \sim n^{-c}$ and the notation "$\sim$" means that two quantities have the same convergence order.

It is worthwhile to note that conditions (A3) and (A4) are concerned with the smoothness of the density function of $\mathbf{w}$ and kernel matrix $\widehat{\mathbf{M}}_{\mathrm{SIR}}(\mathbf{w})$. Furthermore, condition (A5) and (A6) are traditional conditions for nonparametric estimation in the literature (Zhu and Fang 1996; Yin et al. 2010).

Theorem 4 *Suppose that conditions* (A3)–(A6) *hold. Then assume* $\mathbf{X}_{\mathbf{w}}$ *has finite fourth moment and all the nonzero eigenvalues of* $\mathbf{M}_{\mathrm{SIR}}(\mathbf{w})$ *are distinct, as* $n \to \infty$*, we have*

$$\sup_{\mathbf{w}} \|\widehat{\mathbf{M}}_{\mathrm{SIR}}(\mathbf{w}) - \mathbf{M}_{\mathrm{SIR}}(\mathbf{w})\|_F = O(h^4 + n^{-1/2}h^{-1}\log n) \quad a.s.$$
$$\sup_{\mathbf{w}} \|\widehat{\boldsymbol{\beta}}_k^{\mathrm{SIR}}(\mathbf{w}) - \boldsymbol{\beta}_k^{\mathrm{SIR}}(\mathbf{w})\|_F = O(h^4 + n^{-1/2}h^{-1}\log n) \quad a.s.$$
$$\sup_{\mathbf{w}} |\widehat{\boldsymbol{\lambda}}_k^{\mathrm{SIR}} - \boldsymbol{\lambda}_k^{\mathrm{SIR}}(\mathbf{w})| = O(h^4 + n^{-1/2}h^{-1}\log n) \quad a.s. \tag{8}$$

where F denotes Frobenius norm.

As we know, the convergence rate of the dimension reduction matrix of the existing classic inverse regression methods is $\sqrt{n}$. However, $\mathcal{S}_{Y|\mathbf{X}}^{(\mathbf{W}=\mathbf{w})}$ is a completely free function of $\mathbf{W}$ in this paper, since we use $\mathbf{M}(\mathbf{W})$ to estimate it. As shown in Theorem 2, we know that the convergence rate of our proposed variable-dependent partial dimension reduction methods is $h^4 + n^{-1/2}h^{-1}\log n$, which is much slower than classic inverse regression methods. However, if $\mathbf{M}(\mathbf{W})$ is a nonconstant function of $\mathbf{W}$ and $\mathcal{S}_{Y|\mathbf{X}}^{(\mathbf{W}=\mathbf{w})}$ is constant over $\mathbf{W}$, the convergence rate of our proposed variable-dependent method will also be $\sqrt{n}$, which is the same as classic inverse regression methods. This phenomenon is not surprising, since we use $E(\mathbf{M}(\mathbf{W}))$ to estimate $\mathcal{S}_{Y|\mathbf{X}}^{(\mathbf{W}=\mathbf{w})}$. Zhu and Fang (1996) and Zhu and Zhu (2007) proposed kernel method to estimate SIR and

SAVE matrix and used $E(\mathbf{M}(Y))$ to estimate $\mathcal{S}_{Y|\mathbf{X}}$, which is equivalent to our case. They proved that the convergence rate of their methods is $\sqrt{n}$.

3 Determination of dimensionality $d(\mathbf{w})$

Recall that when we estimate the $\mathcal{S}_{Y|\mathbf{X}}^{(\mathbf{W}=\mathbf{w})}$ in Sect. 2, we assume that the variable-dependent structural dimension $d(\mathbf{w})$ is known. However, $d(\mathbf{w})$ is usually unknown in practice, and its estimation is of independent interest. In this section, we extend the state-of-the-art ladle estimator in Luo and Li (2016) into a nonparametric version and establish its consistency property as well.

Let F be the distribution function of $(\mathbf{X}_{\mathbf{w}}, Y_{\mathbf{w}})$, and let F_n be the empirical distribution based on $(\mathbf{X}_1, Y_1), \ldots, (\mathbf{X}_n, Y_n)$ conditional on $\mathbf{W} = \mathbf{w}$. Let $\{(\mathbf{X}_{1,i}^*, Y_{1,i}^*), \ldots, (\mathbf{X}_{n,i}^*, Y_{n,i}^*)\}_{i=1}^n$ be n independent and identically distributed bootstrap sample from F_n, and let F_n^* be the empirical distribution based on the bootstrap sample.

Let $\mathbf{M}(\mathbf{w})$ denote the kernel matrix of a specific dimension reduction approach, let $\mathbf{G}(\mathbf{w}) = \boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\mathbf{M}(\mathbf{w})\mathbf{M}(\mathbf{w})\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}$, and $d(\mathbf{w})$ be the rank of $\mathbf{G}(\mathbf{w})$. Rearrange the eigenvalues of $\mathbf{G}(\mathbf{w})$ as $\boldsymbol{\lambda}_1^2(\mathbf{w}) \geq \ldots \geq \boldsymbol{\lambda}_{d(\mathbf{w})}^2(\mathbf{w}) > 0 = \boldsymbol{\lambda}_{d(\mathbf{w})+1}^2(\mathbf{w}) = \ldots = \boldsymbol{\lambda}_p^2(\mathbf{w})$, and denote the corresponding eigenvectors by $\boldsymbol{\beta}_1(\mathbf{w}), \ldots, \boldsymbol{\beta}_p(\mathbf{w})$, where $\boldsymbol{\lambda}_k(\mathbf{w})$ is the singular value of $\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\mathbf{M}(\mathbf{w})$ and $\boldsymbol{\beta}_k(\mathbf{w})$ is the corresponding left singular vector of $\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\mathbf{M}(\mathbf{w})$ for $k = 1, \ldots, p$. Let $\widehat{\mathbf{G}}(\mathbf{w})$ be the sample kernel matrix based on the sample $\{Y_i, \mathbf{X}_i, \mathbf{W}_i\}_{i=1}^n$, and $\mathbf{G}^*(\mathbf{w})$ be the sample kernel matrix based on the bootstrap sample. In parallel, we can define $\{\widehat{\boldsymbol{\lambda}}_1^2(\mathbf{w}), \ldots, \widehat{\boldsymbol{\lambda}}_p^2(\mathbf{w}), \widehat{\boldsymbol{\beta}}_1(\mathbf{w}), \ldots, \widehat{\boldsymbol{\beta}}_p(\mathbf{w})\}$ and $\{\boldsymbol{\lambda}_1^{*2}(\mathbf{w}), \ldots, \boldsymbol{\lambda}_p^{*2}(\mathbf{w}), \boldsymbol{\beta}_1^*(\mathbf{w}), \ldots, \boldsymbol{\beta}_p^*(\mathbf{w})\}$ for $\widehat{\mathbf{G}}(\mathbf{w})$ and $\mathbf{G}^*(\mathbf{w})$. For each $k < p$, let

$$\mathbf{T}_k(\mathbf{w}) = \Big(\boldsymbol{\beta}_1(\mathbf{w}), \ldots, \boldsymbol{\beta}_k(\mathbf{w})\Big),$$
$$\widehat{\mathbf{T}}_k(\mathbf{w}) = \Big(\widehat{\boldsymbol{\beta}}_1(\mathbf{w}), \ldots, \widehat{\boldsymbol{\beta}}_k(\mathbf{w})\Big),$$
$$\mathbf{T}_k^*(\mathbf{w}) = \Big(\boldsymbol{\beta}_1^*(\mathbf{w}), \ldots, \boldsymbol{\beta}_k^*(\mathbf{w})\Big).$$

Since $\mathbf{T}_k^*(\mathbf{w})$ is repeatedly calculated for n bootstrap samples, we denote its realization at the ith bootstrap sample by $\mathbf{T}_{k,i}^*(\mathbf{w})$.

Conditional on $\mathbf{W} = \mathbf{w}$, define a function from $\{0, 1, \ldots, p-1\}$ to $\mathbb{R}$ as

$$f_n^0(\mathbf{w}, k) = \begin{cases} 0, & k = 0; \\ n^{-1}\sum_{i=1}^n [1 - |\det\{\widehat{\mathbf{T}}_k^{\top}(\mathbf{w})\mathbf{T}_{k,i}^*(\mathbf{w})\}|], & k = 1, \ldots, p-1. \end{cases}$$

As in Ye and Weiss (2003), $1 - |\det\{\widehat{\mathbf{T}}_k^{\top}(\mathbf{w})\mathbf{T}_{k,i}^*(\mathbf{w})\}|$ is a number between 0 and 1 that measures the discrepancy between column spaces of $\widehat{\mathbf{T}}_k(\mathbf{w})$ and $\mathbf{T}_{k,i}^*(\mathbf{w})$, with 1 representing the largest discrepancy. Therefore, $f_n^0(\mathbf{w}, k)$ measures the variability of the bootstrap estimates $\mathbf{T}_{k,1}^*(\mathbf{w}), \ldots, \mathbf{T}_{k,n}^*(\mathbf{w})$ around the full sample estimate $\widehat{\mathbf{T}}_k(\mathbf{w})$.

We then normalize $f_n^0(\mathbf{w}, k)$ to be $f_n(\mathbf{w}, k) = f_n^0(\mathbf{w}, k)/\{1 + \sum_{i=0}^{p-1} f_n^0(\mathbf{w}, i)\}$. The asymptotic behavior of $f_n(\mathbf{w}, \cdot)$ is presented in Lemma 4 in Appendix.

Similarly, we normalize the sample eigenvalues and define the function $\phi_n(\mathbf{w}, \cdot) : \{0, \ldots, p\} \to \mathbb{R}$ as

$$\phi_n(\mathbf{w}, k) = \widehat{\boldsymbol{\lambda}}_{k+1}^2(\mathbf{w}) \Big/ \left\{ 1 + \sum_{i=0}^{p-1} \widehat{\boldsymbol{\lambda}}_{i+1}^2(\mathbf{w}) \right\},$$

where the constant 1 in the denominator is introduced to stabilize the performance of the criterion when $d(\mathbf{w}) = 0$. The technique here is to shift the eigenvalues so that $\phi_n(\mathbf{w}, \cdot)$ takes a small value at $k = d(\mathbf{w})$ instead of at $k = d(\mathbf{w}) + 1$. Lemma 5 gives the asymptotic property of $\phi_n(\mathbf{w}, \cdot)$ in Appendix. Now, we can define the objective function of our estimator as

$$\begin{aligned} g_n(\mathbf{w}, \cdot) \quad &: \{0, \ldots, p-1\} \to \mathbb{R}, \\ g_n(\mathbf{w}, k) &= f_n(\mathbf{w}, k) + \phi_n(\mathbf{w}, k) \\ &= \frac{f_n^0(\mathbf{w}, k)}{1 + \sum_{i=0}^{p-1} f_n^0(\mathbf{w}, i)} + \frac{\widehat{\boldsymbol{\lambda}}_{k+1}^2(\mathbf{w})}{1 + \sum_{i=0}^{p-1} \widehat{\boldsymbol{\lambda}}_{i+1}^2(\mathbf{w})}, \end{aligned} \tag{9}$$

which collects information from both the eigenvectors and the eigenvalues. The reason for using this objective function is that the eigenvalue term $\phi_n(\mathbf{w}, \cdot)$ is large when $k < d(\mathbf{w})$, while the eigenvector term $f_n(\mathbf{w}, \cdot)$ is large when $k > d(\mathbf{w})$, and they are both small when $k = d(\mathbf{w})$.

In most applications it is reasonable to assume $d(\mathbf{w}) \leq \lfloor p/\log(p) \rfloor$, where $\lfloor a \rfloor$ stands for the greatest integer less than or equal to a, and thus objective function (9) becomes

$$\begin{aligned} g_n(\mathbf{w}, k) &= f_n(\mathbf{w}, k) + \phi_n(\mathbf{w}, k) \\ &= \frac{f_n^0(\mathbf{w}, k)}{1 + \sum_{i=0}^{\lfloor p/\log(p) \rfloor} f_n^0(\mathbf{w}, i)} + \frac{\widehat{\boldsymbol{\lambda}}_{k+1}^2(\mathbf{w})}{1 + \sum_{i=0}^{\lfloor p/\log(p) \rfloor} \widehat{\boldsymbol{\lambda}}_{i+1}^2(\mathbf{w})}. \end{aligned} \tag{10}$$

Let $\mathcal{D}(f)$ denote the domain of a function f. Following Luo and Li (2016), we define the ladle estimator for $d(\mathbf{w})$ by

$$\widehat{d}(\mathbf{w}) = \arg\min\{g_n(\mathbf{w}, k) : k \in \mathcal{D}[g_n(\mathbf{w}, \cdot)]\}, \tag{11}$$

where $g_n(\mathbf{w}, \cdot)$ is defined by (9) if $p \leq 10$ or by (10) if $p > 10$.

Theorem 5 establishes the consistency of the nonparametric ladle estimator for variable-dependent partial dimension reduction.

Theorem 5 *Under assumptions* (C9)–(C12), *for positive semi-definite matrix* $\mathbf{G}(\mathbf{w}) \in \mathbb{R}^{p \times p}$ *of rank* $d(\mathbf{w}) \in \{0, \ldots, p-1\}$, *the nonparametric ladle estimator* (11) *enjoys* *the following property:*

$$P\left\{\lim_{n\to\infty} P(\widehat{d}(\mathbf{w}) = d(\mathbf{w})|\mathcal{S}) = 1\right\} = 1.$$

4 The bandwidth selection for variable-dependent partial SIR

Bandwidth selection for both kernel regression estimator and local estimator has been well studied. Since Cook and Yin (2001) showed that SIR can be viewed as linear discriminant analysis, we can choose the bandwidth in a way similar to the tuning parameter selection based on linear discriminant analysis.

Assuming that $\mathbf{X}|(\widetilde{Y} = l, \mathbf{W} = \mathbf{w}) \sim N(\mathbf{m}_l(\mathbf{w}), \mathbf{\Sigma}_{l\mathbf{w}})$, with density function $f_{\mathbf{X}|\widetilde{Y}=l,\mathbf{W}=\mathbf{w}}(x)$, thus, $\mathbf{X}|\mathbf{w}$ follows a mixture multivariate normal distribution, and its likelihood function is given by

$$\begin{aligned} L(\mathbf{m}_l(\mathbf{w}), \mathbf{\Sigma}_{l\mathbf{w}}|\mathbf{x}) &= \prod_{i=1}^{n}\sum_{l=1}^{H} P_{l,\mathbf{w}} f_{\mathbf{X}|\widetilde{Y}=l,\mathbf{W}=\mathbf{w}}(\mathbf{x}_i) \\ &= \prod_{i=1}^{n}\sum_{l=1}^{H} P_{l,\mathbf{w}} \left[\frac{e^{-\frac{1}{2}\{\mathbf{x}_i-\mathbf{m}_l(\mathbf{w})\}^{\mathrm{T}}\mathbf{\Sigma}_{l,\mathbf{w}}^{-1}\{\mathbf{x}_i-\mathbf{m}_l(\mathbf{w})\}}}{\sqrt{(2\pi)^p|\mathbf{\Sigma}_{l,\mathbf{w}}|}}\right], \end{aligned} \tag{12}$$

where $P_{l,\mathbf{w}}$ is defined as before. Recall that $\mathbf{m}_l(\mathbf{w})$ and $\mathbf{\Sigma}_{l\mathbf{w}}$ are both of conditional structure, so we propose to estimate them by the following Nadaraya–Watson (NW) kernel estimators:

$$\begin{aligned} \widehat{\mathbf{m}}_l(\mathbf{w}) &= \frac{\sum_{i=1}^{n}\mathbf{x}_i\mathbf{1}(\widetilde{Y}_i = l)K_h(\mathbf{w}_i - \mathbf{w})}{\sum_{i=1}^{n}\mathbf{1}(\widetilde{Y}_i = l)K_h(\mathbf{w}_i - \mathbf{w})}, \\ \widehat{\mathbf{\Sigma}}_{l\mathbf{w}} &= \frac{\sum_{i=1}^{n}\mathbf{x}_i\mathbf{x}_i^{\mathrm{T}}\mathbf{1}(\widetilde{Y}_i = l)K_h(\mathbf{w}_i - \mathbf{w})}{\sum_{i=1}^{n}\mathbf{1}(\widetilde{Y}_i = l)K_h(\mathbf{w}_i - \mathbf{w})} - \widehat{\mathbf{m}}_l(\mathbf{w})\widehat{\mathbf{m}}_l^{\mathrm{T}}(\mathbf{w}), \\ \widehat{\mathbf{\Sigma}}_{\mathbf{w}} &= \sum_{l=1}^{H}\widehat{P}_{l,\mathbf{w}}\widehat{\mathbf{\Sigma}}_{l\mathbf{w}}. \end{aligned}$$

Since it's too hard to calculate the log-likelihood type of (12) directly, we consider finding out the optimal bandwidth in each slice instead of targeting at the overall bandwidth, which is much more reasonable and computationally feasible. Following the argument of leave-one-out cross validation in Jiang et al. (2017), we propose to find the h_l which is the bandwidth for l-th slice such that

$$\begin{aligned} CV(h_l) = \frac{1}{n_l}\sum_{i=1}^{n_l}\Big[&\{\mathbf{x}_i - \widehat{\mathbf{m}}_l^{(-i)}(\mathbf{w})\}^{\mathrm{T}}\widehat{\mathbf{\Sigma}}_{\mathbf{w}(-i)}^{-1}(\mathbf{w})\{\mathbf{x}_i \\ &- \widehat{\mathbf{m}}_l^{(-i)}(\mathbf{w})\} + \log|\widehat{\mathbf{\Sigma}}_{\mathbf{w}(-i)}(\mathbf{w})|\Big] \end{aligned} \tag{13}$$

is minimized, where $\widehat{\mathbf{m}}_l^{(-i)}(\mathbf{w})$ and $\widehat{\boldsymbol{\Sigma}}_{\mathbf{w}(-i)}$ are estimators of the mean and covariance matrix of $\mathbf{X}|\widetilde{Y} = l, \mathbf{W} = \mathbf{w}$, computed without the i-th observation, n_l is the total number of observations in the l-th slice. We choose the value of h_l which maximizes (12) as h_{opt}, which is the bandwidth selected for variable-dependent partial SIR.

5 Simulation studies

In this section, we conduct simulation studies to evaluate our variable-dependent PDR methods. We consider the following five models:

1. Model I: $Y = X_1|W| + 3X_2 \cos W + 0.2\varepsilon$,
2. Model II: $Y = 2\exp\{X_1 \exp(W) - X_2 \cos W + 1\} \cdot \text{sign}\{0.01X_1 \cos W + 2(W + 1)^2 X_2\} + 0.2\varepsilon$,
3. Model III: $Y = \{X_1 \sin(W) + 5X_2 \cos(W)\}^2 + 0.2\varepsilon$,
   vModel IV: $Y = \exp\{(X_1|W| + X_2)^2\} \log\{(X_3 \cos W)^2\} + 0.2\varepsilon$,
4. Model V: $Y = 10\frac{\exp\{X_1 \sin W + 5X_2|W|\}}{X_1 \exp(W) - X_2 \cos W} + 0.2\varepsilon$,where sign$(\cdot)$ is the sign function, and $\varepsilon \sim N(0, 1)$. For Models I-V, W is univariate and has a uniform distribution $U(-1, 1)$, $\mathbf{X}|W \sim N_p(\mathbf{m}(W), \boldsymbol{\Sigma}_W)$, where $\mathbf{m}(W) = \frac{\sin(W)}{2}\mathbf{1}_p$, $\boldsymbol{\Sigma}_W = (\sigma_{ij})_{p\times p}$ with $\sigma_{ij} = 1$ for $i = j$, $\sigma_{ij} = \frac{1}{2}\sin(W)$ for $i \neq j$. Hence, $d(\mathbf{w}) = 1$ for Models I and III, and $d(\mathbf{w}) = 2$ for Models II, IV. For Model V, $d(\mathbf{w}) = 2$ when $\mathbf{w} \neq 0$, and 1 when $\mathbf{w} = 0$.

Our simulation studies are twofold. We first use the nonparametric ladle estimator in (11) to determine $d(\mathbf{w})$ for each model and then estimate the basis for $\mathcal{S}_{Y_{\mathbf{w}}|\mathbf{X}_{\mathbf{w}}}$ on the estimated structural dimension. Since the inverse conditional mean is symmetric about 0 for Models III and IV, we expect that variable-dependent partial SIR may provide a poor estimate for these two models. However, partial variable-dependent SIR might have advantages over variable-dependent partial SAVE and DR for the rest of simulation models when the sample size is small since SIR is based on first inverse moments.

5.1 Estimation of structural dimension

Based on the nonparametric ladle estimator, we use variable-dependent partial SIR (VDPSIR), variable-dependent partial SAVE (VDPSAVE) and variable-dependent partial DR (VDPDR) to estimate $d(\mathbf{w})$ for different w. The percentages of correct order estimates in 500 replications for Model I to V are presented in Table 1. It shows that our proposed nonparametric ladle estimator works pretty well, with the percentage of correct order estimation approaches 100%. Also, as we expected, variable-dependent partial SIR fails for Models III and V and outperforms the other two variable-dependent PDR methods for the remaining models due to SIR is based on the estimation of the conditional mean. Figure 1shows the percentage of correct order determination among 500 replications at different values of w for Models I to IV.

Table 2shows that, for Model V, under the scenario $(n, p) = (150, 5)$, if w is away from 0, the structural dimension can be accurately estimated by variable-dependent

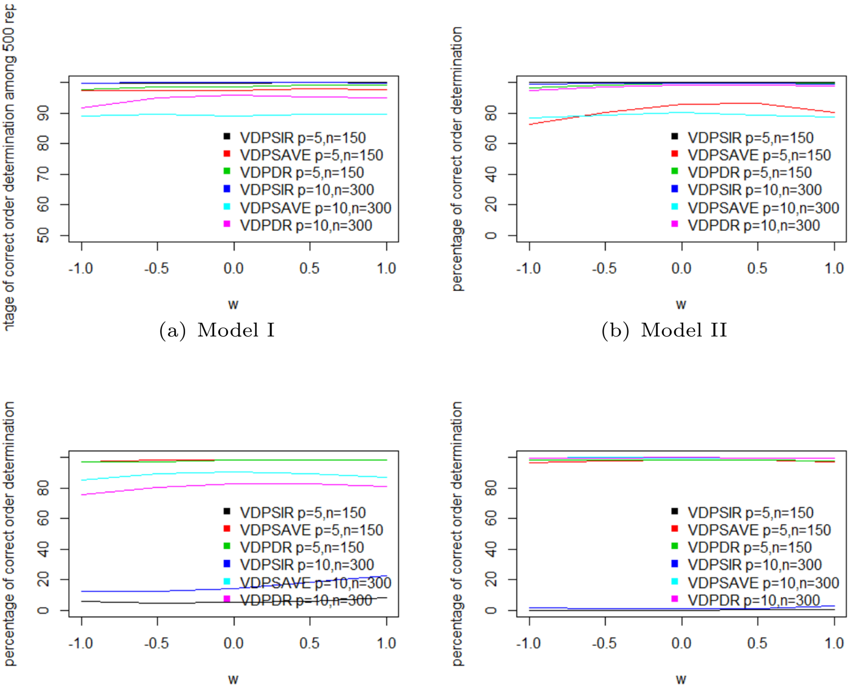
Table 1 Percentage of correct order determinations among 500 replications for Models I–V (%)

Model	w	$(n, p) = (150, 5)$			$(n, p) = (300, 10)$		
		VDPSIR	VDPSAVE	VDPDR	VDPSIR	VDPSAVE	VDPDR
I	0	99.800	97.200	98.600	100.000	89.000	95.800
	− 1	99.800	97.400	97.600	99.800	89.000	91.600
	1	100.000	97.600	99.200	99.600	89.600	93.400
	− 0.5	99.800	97.400	98.400	100.000	89.600	94.800
	0.5	100.000	98.000	99.200	100.000	89.600	95.200
II	0	100.000	85.800	99.600	99.400	80.200	98.200
	− 1	100.000	72.600	96.600	99.000	76.600	94.800
	1	100.000	80.200	99.200	99.000	77.200	97.400
	− 0.5	100.000	80.600	98.000	99.200	78.600	97.200
	0.5	100.000	86.400	99.400	99.200	78.400	98.200
III	0	5.400	98.000	98.200	14.400	90.400	82.600
	− 1	5.800	97.000	97.200	12.400	84.800	75.600
	1	8.200	98.000	98.200	22.600	86.800	81.000
	− 0.5	4.600	98.000	97.200	12.400	89.400	80.200
	0.5	5.800	98.200	98.200	18.600	89.200	82.800
IV	0	0.200	98.400	98.400	1.400	99.600	99.800
	− 1	0	96.400	98.400	2.000	99.400	99.400
	1	0.600	96.800	97.600	3.000	99.600	99.600
	− 0.5	0.200	97.800	98.400	1.200	99.800	99.600
	0.5	0.400	98.000	98.400	1.400	99.600	99.600
V	− 1	98.200	70.600	81.400	99.400	70.600	72.600
	1	99.800	70.000	83.800	98.800	70.000	74.400
	− 0.5	99.600	72.800	88.400	99.400	73.400	86.000
	0.5	99.400	72.800	83.400	99.200	72.000	78.800

partial SIR approach all the times. When $(n, p) = (300, 10)$, at least 96% of the times, variable-dependent partial SIR provides with the correct estimates of $d = 2$.

Figure 2 shows the percentage of correct order determination varies when w lies within a small neighborhood of zero, whose shape is different from Fig. 1, since $w = 0$ is a singular point.

We also compared our simulation results from 500 runs with 100 runs for order determination for Model I–IV, and did not see significant differences. Hence, due to computational cost, all the following simulation results are from 100 runs. Table 3 suggests that the percentage of correct estimation for Model V when $w = 0$ is somewhat unsatisfactory when the sample size n is small. However, as n increases, the correct estimation percentage improves steadily, which is consistent with large sample theory.



(a) Model I (b) Model II (c) Model III (d) Model IV

Fig. 1 Percentage of correct order determination among 500 replications for Models I–IV

Table 2 Percentage of correct order determinations among 500 replications for Model V (%)

Model	w	$(n, p) = (150, 5)$			$(n, p) = (300, 10)$		
		VDPSIR	VDPSAVE	VDPDR	VDPSIR	VDPSAVE	VDPDR
V	-1	98.200	70.600	81.400	99.400	70.600	72.600
	1	99.800	70.000	83.800	98.800	70.000	74.400
	-0.5	99.600	72.800	88.400	99.400	73.400	86.000
	0.5	99.400	72.800	83.400	99.200	72.000	78.800
	0	60.800	6.000	14.000	54.800	9.200	15.400
	0.001	80.400	59.400	67.400	78.000	56.600	64.200
	0.002	85.200	61.000	69.600	83.200	59.400	67.400
	0.003	88.600	61.200	70.200	88.000	60.200	68.800
	0.004	90.000	63.200	73.600	88.800	61.000	70.400
	0.005	92.800	63.800	74.200	91.800	61.400	71.600
	0.006	93.000	64.600	74.000	93.000	62.600	72.800
	0.007	94.200	65.000	75.400	93.400	63.400	73.200
	0.008	96.000	65.200	76.600	95.000	64.200	75.800
	0.009	96.400	69.400	76.800	95.400	67.600	75.800

Table 2 continued

Model	w	$(n, p) = (150, 5)$			$(n, p) = (300, 10)$		
		VDPSIR	VDPSAVE	VDPDR	VDPSIR	VDPSAVE	VDPDR
	0.01	98.200	70.800	78.600	96.200	68.000	76.200
	0.0125	98.600	70.600	79.400	97.600	68.800	77.400
	0.015	98.600	71.800	79.800	97.200	70.200	77.600
	0.0175	98.400	72.000	79.400	97.800	70.600	77.600
	0.02	98.600	72.000	80.600	97.800	70.600	78.000
	0.025	98.600	72.200	80.600	98.200	71.000	78.200
	0.03	99.000	72.000	82.400	98.000	71.800	78.600
	0.035	98.800	72.400	82.200	98.400	71.600	79.400
	0.04	99.000	72.200	84.000	99.200	71.400	81.200
	0.045	99.200	72.400	84.000	99.000	71.800	82.400
	0.05	99.000	72.800	84.400	99.200	71.800	82.600
	− 0.001	79.600	58.400	65.200	76.400	54.200	63.400
	− 0.002	82.800	60.400	67.600	80.600	57.600	66.400
	− 0.003	86.600	61.200	67.800	86.200	59.400	66.800
	− 0.004	88.400	63.600	70.200	87.800	60.000	67.600
	− 0.005	90.200	63.200	72.800	90.400	60.600	68.400
	− 0.006	91.600	66.000	73.600	91.000	62.400	69.800
	− 0.007	92.800	68.400	75.400	93.400	63.200	72.600
	− 0.008	95.600	68.800	76.200	95.200	63.600	73.800
	− 0.009	96.800	70.000	76.000	95.400	64.200	74.000
	− 0.01	97.600	70.800	76.200	96.000	64.800	75.800
	− 0.0125	97.600	70.800	79.800	96.400	65.400	76.200
	− 0.015	98.000	71.200	80.200	97.000	66.400	77.400
	− 0.0175	98.000	71.600	80.000	96.600	66.800	77.600
	− 0.02	98.400	72.000	80.600	96.400	68.000	78.800
	− 0.025	98.200	71.800	82.200	96.600	68.800	78.600
	− 0.03	98.400	72.800	82.000	98.000	68.200	79.200
	− 0.035	98.200	72.600	82.400	97.800	69.400	79.000
	− 0.04	99.000	72.400	82.800	98.600	69.800	80.200
	− 0.045	98.800	72.400	82.800	99.000	70.400	80.000
	− 0.05	99.200	72.600	82.600	98.800	70.400	81.200

Table 3 Percentage of correct order determination among 100 replications for Model V with $\mathbf{w} = 0$ (%)

(n, p)	VDPSIR	VDPSAVE	VDPDR	(n, p)	VDPSIR	VDPSAVE	VDPDR
(150, 5)	62.000	21.000	25.000	(300, 10)	59.000	19.000	27.000
(500, 5)	73.000	51.000	60.000	(500, 10)	65.000	42.000	53.000
(800, 5)	86.000	74.000	82.000	(800, 10)	75.000	57.000	68.000
(1000, 5)	93.000	88.000	90.000	(1000, 10)	81.000	70.000	78.000

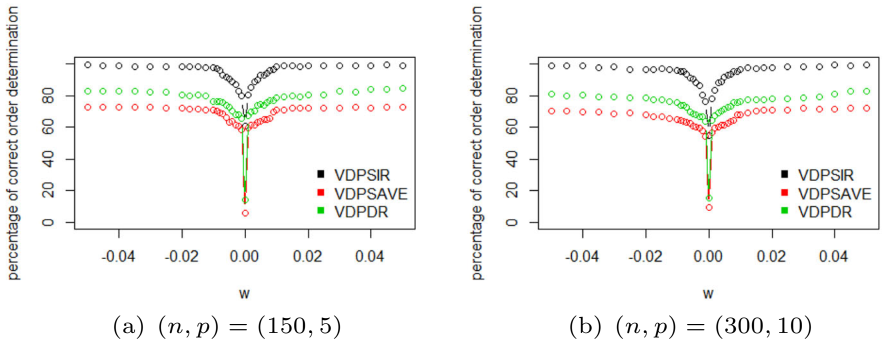


Fig. 2 Percentage of correct order determination among 500 replications for Model V

5.2 Estimation of $\mathcal{S}_{Y_\mathbf{w}|\mathbf{X}_\mathbf{w}}$

To assess the accuracy of our variable-dependent PDR methods, we adopt the trace correlation $r^2_{d(\mathbf{w})}$ proposed by Ferré (1998). Let $\mathcal{S}_1$ and $\mathcal{S}_2$ be two $d(\mathbf{w})$-dimensional subspaces of $\mathbb{R}^p$, the distance between subspace $\mathcal{S}_1$ and $\mathcal{S}_2$ can be measured by the following trace correlation coefficient,

$$r^2_{d(\mathbf{w})} = \text{Tr}(P_{\mathcal{S}_1} P_{\mathcal{S}_2})/d(\mathbf{w}),$$

where $P_{\mathcal{S}_1}$ and $P_{\mathcal{S}_2}$ are orthogonal projections onto $\mathcal{S}_1$ and $\mathcal{S}_2$, and $\text{Tr}(\cdot)$ is the trace of a square matrix. It can be justified that $r^2_{d(\mathbf{w})} \in [0, 1]$, $r^2_{d(\mathbf{w})} = 1$ if $\mathcal{S}_1 = \mathcal{S}_2$, and $r^2_{d(\mathbf{w})} = 0$ if $\mathcal{S}_1 \perp \mathcal{S}_2$ (the two subspaces are perpendicular). Note that a larger value of $r^2_{d(\mathbf{w})}$ implies that $\mathcal{S}_1$ and $\mathcal{S}_2$ are closer. Li and Dong (2009) and Dong and Li (2010) applied a similar criterion to assess the performance of sufficient dimension reduction estimator with non-elliptically distributed predictors.

We first compare the performance among the three variable-dependent PDR methods under two configurations $(n, p) = (150, 5)$ and $(n, p) = (300, 10)$. To be fair, we set the number of slices H to be 5 for all three methods. Due to space limitations, we only present part of the results. We present in Table 4the mean of trace correlations between the true and estimated variable-dependent partial CS at different values of $\mathbf{w}$ for the first five models based on 100 repetitions.

Table 4 shows that variable-dependent partial SIR works pretty well in most cases except for Models III and IV just as expected, while both variable-dependent partial SAVE and DR perform stably for all models, providing with trace correlations greater than 0.9 most of the time.

6 Real data analysis

In this section, we consider analyzing four real-world datasets: Body Fat data, Wage data, Hongkong environmental data, and Boston Housing data. For each dataset, we

Table 4 Trace correlation for the Models I–V

Model	**w**	$(n, p) = (150, 5)$			$(n, p) = (300, 10)$		
		VDPSIR	VDPSAVE	VDPDR	VDPSIR	VDPSAVE	VDPDR
I	0	0.978	0.962	0.945	0.980	0.958	0.933
	−1	0.892	0.920	0.918	0.889	0.918	0.919
	1	0.894	0.920	0.922	0.888	0.919	0.918
	−0.5	0.991	0.994	0.985	0.989	0.992	0.978
	0.5	0.990	0.995	0.986	0.988	0.992	0.977
II	0	0.973	0.909	0.914	0.958	0.835	0.856
	−1	0.940	0.910	0.911	0.927	0.831	0.854
	1	0.941	0.898	0.913	0.932	0.817	0.851
	−0.5	0.961	0.909	0.913	0.947	0.835	0.856
	0.5	0.969	0.905	0.914	0.959	0.830	0.854
III	0	0.297	0.967	0.967	0.313	0.955	0.952
	−1	0.218	0.880	0.858	0.203	0.847	0.838
	1	0.492	0.941	0.945	0.536	0.939	0.945
	−0.5	0.230	0.956	0.950	0.242	0.938	0.935
	0.5	0.389	0.965	0.969	0.418	0.957	0.955
IV	0	0.361	0.906	0.908	0.221	0.878	0.879
	−1	0.433	0.938	0.940	0.321	0.919	0.917
	1	0.418	0.941	0.941	0.320	0.916	0.918
	−0.5	0.379	0.965	0.966	0.256	0.940	0.940
	0.5	0.375	0.966	0.966	0.249	0.940	0.941
V	0	0.895	0.850	0.883	0.833	0.813	0.814
	−1	0.833	0.827	0.857	0.813	0.815	0.821
	1	0.904	0.847	0.884	0.951	0.840	0.898
	−0.5	0.912	0.826	0.872	0.905	0.818	0.839
	0.5	0.946	0.871	0.890	0.948	0.873	0.903

compare our proposals, variable-dependent partial SIR, variable-dependent partial SAVE, and variable-dependent partial DR with the PDEE methods which include PDEE-SIR, PDEE-SAVE, and PDEE-DR.

6.1 Description of the datasets

We illustrate the application of variable-dependent PDR to the following four data sets in the literature.

Body Fat dataset we first consider the Body Fat data, which has been analyzed in Penrose et al. (1985), Hoeting et al. (1999) and Leng (2010). The Body Fat data contains 252 observations and 14 attributes. Following the analysis of Leng (2010), we treat *brozek* as the response Y, *age* as W, and the other 12 predictors (i.e., *weight, height, neck, chest, abdomen, hip, thigh, knee, ankle, biceps, forearm, wrist*) as $\mathbf{X}$. For the structural determination, we estimate $\hat{d}(w)$ for $W = w_i$ and get that $\hat{d}(w_i) = 1$ for $i = 1, \ldots, 252$. This is consistent with the previous studies in Leng (2010) and Zhang et al. (2013), both estimated the structural dimension as 1.

Wage dataset Wage dataset contains the wage information of 534 workers and their education, living region, gender, union membership, race, occupation, sector, marriage status information and their years of experience. This data set has been investigated in Berndt (1991), Xie and Huang (2009) and Zhang et al. (2013). We take *wage* as the response Y, *years of experience* as W, and the remaining 8 predictors as $\mathbf{X}$. The order determination procedure yields that $\hat{d}(w_i) = 1$ for $i = 1, \ldots, 534$.

Hongkong environmental dataset Hongkong environmental dataset has been analyzed in Li et al. (2015). This data set was collected between January 1 of 1994 and December 31 of 1995. To be more specific, it is a collection of numbers of daily total hospital admissions for circulatory and respiratory problems, measurements of pollutants and many other environmental factors in Hong Kong. We take *the number of daily total hospital admissions for circulatory and respiratory problems* as the response Y, *time* as W, and the remaining predictors as $\mathbf{X}$. The order determination procedure yields that $\hat{d}(w_i) = 1$ for $i = 1, \ldots, 730$.

Boston Housing dataset. Boston housing data (Harrison and Rubinfeld 1978) has been widely used as a classical dataset in regression study. For example, it has been studied in Fan and Huang (2005) and Chen et al. (2010). It contains information collected by the U.S. Census Service concerning housing in the area of Boston. The original data consist of 14 variables (features) and 506 data points. Following Chen et al. (2010), we only keep 374 observations with *per capital crime rate by town* smaller than 3.2 in the subsequent analysis. We take *the median value of the owner-occupied homes in $1000's* as the response Y, *crime rate* as W, and the remaining predictors as $\mathbf{X}$. The order determination procedure yields that $\hat{d}(w_i) = 2$ for $i = 1, \ldots, 374$.

To implement each of the methods, we first need to estimate the structural dimension. For our proposed methods, we use the nonparametric ladle estimator in (11) to estimate $d(w_i)$, while using the ladle estimator in Luo and Li (2016) to estimate d for the PDEE methods. Then based on the estimated structural dimension, we obtain the variable-dependent partial dimension reduction directions $\widehat{\mathbf{B}}(w_i)$ and partial dimension reduction directions $\widehat{\mathbf{B}}$. We use the mean distance correlations (Székely et al. 2007) between Y and $\widehat{\mathbf{B}}^{\mathrm{T}}(w_i)\mathbf{X}$ and distance correlation between Y and $\widehat{\mathbf{B}}^{\mathrm{T}}\mathbf{X}$ to evaluate the performance of the estimates $\widehat{\mathbf{B}}(w)$ and $\widehat{\mathbf{B}}$.

We derive distance correlations between Y and $\widehat{\mathbf{B}}(w_i)^{\mathrm{T}}\mathbf{X}$ for each i, then calculate the mean distance correlation based on all w. The mean distance correlations between Y and $\widehat{\mathbf{B}}(w_i)^{\mathrm{T}}\mathbf{X}$ are reported in columns 2-4 in Table 5, where $\widehat{\mathbf{B}}(w_i)$ is the estimate by our proposals (i.e., VDPSIR, VDPSAVE, VDPDR), and the distance correlations between Y and $\widehat{\mathbf{B}}^{\mathrm{T}}\mathbf{X}$ are reported in columns 5–7, where $\widehat{\mathbf{B}}$ is the estimate by PDEE methods (i.e., PDEE-SIR, PDEE-SAVE, PDEE-DR).

It's obvious that our variable-dependent methods consistently beat the PDEE approaches for each variant of SIR, SAVE, DR procedure. To be specific, for the Body Fat dataset, the variable-dependent partial SIR estimation gives a distance correlation of 0.83, compared with 0.69 resulted from PDEE-SIR. For the Boston Housing

Table 5 Distance correlation for each estimation approaches

Dataset	VDPSIR	VDPSAVE	VDPDR	PDEE-SIR	PDEE-SAVE	PDEE-DR
Body Fat	0.830	0.215	0.531	0.691	0.167	0.267
Wage	0.520	0.222	0.504	0.517	0.151	0.150
Hongkong environmental	0.459	0.206	0.194	0.216	0.138	0.119
Boston Housing	0.911	0.662	0.869	0.803	0.342	0.541

dataset, the distance correlation from variable-dependent partial DR is 0.87 compared with 0.54 from PDEE-DR. The more accurate dimension reduction estimates we obtained could greatly facilitate further modeling and analysis.

7 Discussions

In this paper, we propose a variable-dependent approach to better deal with partial sufficient dimension reduction. The dimension $d(\mathbf{w})$ is a function of $\mathbf{w}$ and can vary with different $\mathbf{w}$. As we show in Model V, which is a special and important model in our paper, our proposed variable-dependent partial dimension reduction method works well to estimate $d(\mathbf{w})$, since we can correctly specify $d(w) = 1$ when $w = 0$ and $d(w) = 2$ when $w \neq 0$. For all the other simulation models, our method can also correctly estimate $d(w)$. For the purpose of statistical estimation, a kernel matrix is developed and its asymptotic properties are thoroughly investigated. We also develop a nonparametric ladle estimator to determine the structural dimension. For future work, we plan to investigate how to apply variable-dependent PDR to conduct variable selections, which is of special practical importance since different $\mathbf{w}$ may lead to different variable selection results.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11749-022-00841-y.

Acknowledgements Zhou Yu is supported by The Basic Research Project of Shanghai Science and Technology Commission “22JC1400800" and the National Natural Science Foundation of China Grants “11971170" and “11831008”.

References

Berndt ER (1991) The practice of econometrics: classic and contemporary. Addison Wesley Publishing Company

Bura E, Cook RD (2001) Estimating the structural dimension of regressions via parametric inverse regression. J R Stat Soc: Ser B (Stat Methodol) 63(2):393–410

Chen X, Zou C, Cook RD (2010) Coordinate-independent sparse sufficient dimension reduction and variable selection. Ann Stat 38(6):3696–3723

Chiaromonte F, Cook RD, Li B (2002) Sufficient dimension reduction in regressions with categorical predictors. Ann Stat 30(2):475–497

Cook RD (1998) Regression graphics. Wiley, New York

Cook RD, Forzani L (2009) Likelihood-based sufficient dimension reduction. J Am Stat Assoc 104(485):197–208
Cook RD, Ni L (2005) Sufficient dimension reduction via inverse regression: a minimum discrepancy approach. J Am Stat Assoc 100(470):410–428
Cook RD, Weisberg S (1991) Discussion of "Sliced inverse regression for dimension reduction". J Am Stat Assoc 86(414):328–332
Cook RD, Yin X (2001) Theory and methods: Special invited paper: dimension reduction and visualization in discriminant analysis (with discussion). Aust N Z J Stat 43(2):147–199
Dong Y, Li B (2010) Dimension reduction for non-elliptically distributed predictors: second-order methods. Biometrika 97(2):279–294
Fan J, Huang T (2005) Profile likelihood inferences on semiparametric varying-coefficient partially linear models. Bernoulli 11(6):1031–1057
Feng Z, Wen XM, Yu Z, Zhu LX (2013) On partial sufficient dimension reduction with applications to partially linear multi-index models. J Am Stat Assoc 108(501):237–246
Ferré L (1998) Determining the dimension in sliced inverse regression and related methods. J Am Stat Assoc 93(441):132–140
Hall P, Li KC (1993) On almost linearity of low dimensional projections from high dimensional data. Ann Stat 21(2):867–889
Hoeting JA, Madigan D, Raftery AE, Volinsky CT (1999) Bayesian model averaging: a tutorial. Stat Sci 14(4):382–401
Jiang B, Chen Z, Leng C (2017) Dynamic linear discriminant analysis in high dimensional space. Bernoulli 26(2):1234–1268
Leng C (2010) Variable selection and coefficient estimation via regularized rank regression. Stat Sin 20(1):167–181
Li B (2018) Sufficient dimension reduction: methods and applications with R. CRC Press
Li KC (1991) Sliced inverse regression for dimension reduction. J Am Stat Assoc 86(414):316–327
Li B, Dong Y (2009) Dimension reduction for nonelliptically distributed predictors. Ann Stat 37(3):1272–1298
Li B, Wang S (2007) On directional regression for dimension reduction. J Am Stat Assoc 102(479):997–1008
Li Y, Zhu LX (2007) Asymptotics for sliced average variance estimation. Ann Stat 35(1):41–69
Li B, Cook RD, Chiaromonte F (2003) Dimension reduction for the conditional mean in regressions with categorical predictors. Ann Stat 31(5):1636–1668
Li D, Ke Y, Zhang W (2015) Model selection and structure specification in ultra-high dimensional generalised semi-varying coefficient models. Ann Stat 43(6):2676–2705
Luo W, Li B (2016) Combining eigenvalues and variation of eigenvectors for order determination. Biometrika 103(4):875–887
Ma Y, Zhu L (2012) A semiparametric approach to dimension reduction. J Am Stat Assoc 107(497):168–179
Ma Y, Zhu L (2013) A review on dimension reduction. Int Stat Rev 81(1):134–150
Ma Y, Zhu L (2013) Efficient estimation in sufficient dimension reduction. Ann Stat 41(1):250
Ma Y, Zhu L (2014) On estimation efficiency of the central mean subspace. J R Stat Soc: Ser B (Stat Methodol) 76(5):885–901
Nadaraya EA (1964) On estimating regression. Theory Probab Appl 9(1):141–142
Penrose KW, Nelson AG, Fisher AG (1985) Generalized body composition prediction equation for men using simple measurement techniques. Med Sci Sports Exerc 17(2):189
Székely GJ, Rizzo ML, Bakirov NK (2007) Measuring and testing dependence by correlation of distances. Ann Stat 35(6):2769–2794
Wang H, Xia Y (2008) Sliced regression for dimension reduction. J Am Stat Assoc 103(482):811–821
Watson GS (1964) Smooth regression analysis, Sankhyi: the Indian journal of statistics. Series A 26(4):359–372
Wen X, Cook RD (2007) Optimal sufficient dimension reduction in regressions with categorical predictors. J Stat Plan Inference 137(6):1961–1978
Xia Y, Tong H, Li WK, Zhu LX (2002) An adaptive estimation of dimension reduction space. J R Stat Soc: Ser B (Stat Methodol) 64(3):363–410
Xie H, Huang J (2009) SCAD-penalized regression in high-dimensional partially linear models. Ann Stat 37(2):673–696
Xue Y, Wang Q, Yin X (2018) A unified approach to sufficient dimension reduction. J Stat Plan Inference 197:168–179

Ye Z, Weiss RE (2003) Using the bootstrap to select one of a new class of dimension reduction methods. J Am Stat Assoc 98(464):968–979
Yin X, Cook RD (2002) Dimension reduction for the conditional kth moment in regression. J R Stat Soc: Ser B (Stat Methodol) 64(2):159–175
Yin X, Cook RD (2003) Estimating central subspaces via inverse third moments. Biometrika 90(1):113–125
Yin J, Geng Z, Li R, Wang H (2010) Nonparametric covariance model. Stat Sin 20(1):469–479
Zhang R, Zhao W, Liu J (2013) Robust estimation and variable selection for semiparametric partially linear varying coefficient model based on modal regression. J Nonparametr Stat 25(2):523–544
Zhu LX, Fang KT (1996) Asymptotics for kernel estimate of sliced inverse regression. Ann Stat 24(3):1053–1068
Zhu LP, Zhu LX (2007) On kernel method for sliced average variance estimation. J Multivar Anal 98(5):970–991