Statistica Sinica

VARIABLE DEPENDENT PARTIAL DIMENSION REDUCTION

Lu Li, Kai Tan, Xuerong Meggie Wen and Zhou Yu

East China Normal University, East China Normal University,

Missouri University of Science and Technology, East China Normal University

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, these methods generally perform poorly when $\mathbf{X}$ and $\mathbf{W}$ are related, furthermore, 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 comparing to the existing methods.

Key words and phrases: Directional Regression; Sliced Average Variance Estimation; Sliced Inverse Regression; Sufficient dimension reduction; Order determination.

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, Wang and Yin (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 pursues a partial dimension reduction of $\mathbf{X}$. 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 area 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 $Y \mid (\mathbf{X}, \mathbf{W})$, 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

$$Y \perp\!\!\!\perp \mathbf{X} | (\mathbf{P}_{\mathcal{S}}\mathbf{X}, \mathbf{W}), \tag{1.1}$$

also satisfies condition (1.1), we call it the partial central subspace, and denote it by $\mathcal{S}_{Y|\mathbf{X}}^{(\mathbf{W})}$. And $\dim\{\mathcal{S}_{Y|\mathbf{X}}^{(\mathbf{W})}\} = d$ 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 $\mathcal{S}_{Y|\mathbf{X}}^{(\mathbf{W})}$ varies with continuous $\mathbf{W}$, which is often the case in real-world applications. It is also worth addressing that PDEE depends on the fact that $\mathbf{W}$ is independent of $\mathbf{X}$, which is a considerably stronger condition in the real problem in the practice.

To overcome the limitations mentioned above, we propose the concept of *Variable Dependent Partial Dimension Reduction*, where the partial CS,

$\mathcal{S}_{Y|\mathbf{X}}^{(\mathbf{W})}$ is allowed to vary with $\mathbf{W}$ which is dependent or independent of variable $\mathbf{X}$. Hence the aim of variable dependent partial dimension reduction is to find a matrix of smooth functions of $\mathbf{W}$ with minimum rank, $\mathbf{B}(\mathbf{W}) \in \mathbb{R}^{p\times d(\mathbf{W})}$, such that

$$Y \perp\!\!\!\perp \mathbf{X}|(\mathbf{B}^{\mathrm{T}}(\mathbf{W})\mathbf{X}, \mathbf{W}). \tag{1.2}$$

Then $d(\mathbf{W}) \triangleq \text{rank}\{\mathbf{B}(\mathbf{W})\}$ is the structural dimension function of the variable dependent partial CS. It is worth noting that the covariance matrix $\text{Cov}(\mathbf{X}|\mathbf{W})$, 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.

The contribution of this paper is threefold. First, to the best of our knowledge, the proposed variable dependent partial dimension reduction is the first attempt to perform partial dimension reduction for the case that $\mathcal{S}_{Y|\mathbf{X}}^{(\mathbf{W})}$ varies with $\mathbf{W}$ that is dependent or independent of variable $\mathbf{X}$. Second, 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. Last but not the least, we propose to determine the structural dimension 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.

The rest of this paper is organized as follows. In Section 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 Section 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. Results extended to variable dependent partial SAVE and DR are included in section 5. Section 6 presents the finite sample performance of our proposed methods via extensive simulation studies. To illustrate the efficiency of our proposed methods, four real data analysis are conducted in section 7. For the ease of presentation, we defer all proofs to the 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 Section 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

$$Y_{\mathbf{w}} \perp\!\!\!\perp \mathbf{X}_{\mathbf{w}} | \mathbf{B}^{\mathrm{T}}(\mathbf{w})\mathbf{X}_{\mathbf{w}}, \tag{1.3}$$

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. Then we present a simple example here. Suppose $\mathbf{W}$ is gender, predictor $\mathbf{X}$ is a vector include a person's some characters, such as weight, height, age and something else, response variable Y is blood pressure. If $\mathbf{w}$ is male, then $\mathbf{X}_{\mathbf{w}}$ represents the male's characters and $Y_{\mathbf{w}}$ is the male's blood pressure.

We employ nonparametric covariance models to analyze this variable dependent scenario. Let $\mathbf{m}(\mathbf{w}) = (\mathbf{m}_1(\mathbf{w}), \ldots, \mathbf{m}_p(\mathbf{w}))^{\mathrm{T}}$ and $\mathbf{\Sigma}_{\mathbf{w}} = \{\sigma_{ij}(\mathbf{w})\}_{p\times p}$ denote the mean and covariance of $\mathbf{X}_{\mathbf{w}}$, where both $\mathbf{m}(\mathbf{w})$ and $\mathbf{\Sigma}_{\mathbf{w}}$ are smooth functions of $\mathbf{w}$. Equation (1.3) implies the reduction of the predictor from $\mathbf{X}_{\mathbf{w}}$ to $\mathbf{B}^{\mathrm{T}}(\mathbf{w})\mathbf{X}_{\mathbf{w}}$. If $\mathbf{\Sigma}_{\mathbf{w}}$ is invertible, one can also work with the standardized data $\mathbf{Z}_{\mathbf{w}} = {\mathbf{\Sigma}_{\mathbf{w}}}^{-1/2}\{\mathbf{X} - \mathbf{m}(\mathbf{w})\}$ 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 $\mathcal{S}_{Y_{\mathbf{w}}|\mathbf{X}_{\mathbf{w}}} = {\mathbf{\Sigma}_{\mathbf{w}}}^{-1/2}\mathcal{S}_{Y_{\mathbf{w}}|\mathbf{Z}_{\mathbf{w}}}$ (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.

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) $\mathrm{E}\{\mathbf{X}_\mathbf{w}|\mathbf{B}^{\mathrm{T}}(\mathbf{w})\mathbf{X}_\mathbf{w}\}$ 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 it 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 (A1) and (A2) are guaranteed when $\mathbf{X}_\mathbf{w}$ is normally distributed. More discussions about conditions (A1) and (A2) can be found in Chiaromonte et al. (2002), Li et al. (2003), Li and Dong (2009) and Dong and Li (2010), Li (2018), etc.

Firstly, we 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 $\mathrm{Cov}\{\mathrm{E}(\mathbf{X}|\mathrm{Y})\}$. 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 $\widetilde{Y} = \sum_{l=1}^{H} l \cdot \mathbb{1}(Y \in J_l)$. 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 on to $\mathcal{S}_{Y_{\mathbf{w}}|\mathbf{X}_{\mathbf{w}}}$ with respect to the inner product $\langle a, b\rangle := a^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{w}}b$, 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*

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

Proposition 1 indicates that the random vector $\mathbf{\Sigma}_{\mathbf{w}}^{-1}(\mathrm{E}(\mathbf{X}_{\mathbf{w}}|Y_{\mathbf{w}})-\mathbf{m}(\mathbf{w}))$ 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 $\mathbf{\Sigma}_{\mathbf{w}}^{-1}\mathrm{Cov}\{\mathrm{E}(\mathbf{X}_{\mathbf{w}}|Y_{\mathbf{w}})\}$ 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 the Appendix.

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

$$\begin{aligned}\mathbf{M}_{\mathrm{SIR}}(\mathbf{w}) &\triangleq \mathrm{Cov}\{\mathrm{E}(\mathbf{X}|\widetilde{\mathrm{Y}},\mathbf{w})\}\\ &= \sum_{l=1}^{H}\mathbf{P}_{l,\mathbf{w}}(\mathbf{V}_{l,\mathbf{w}}-\mathbf{m}(\mathbf{w}))(\mathbf{V}_{l,\mathbf{w}}-\mathbf{m}(\mathbf{w}))^{\mathrm{T}} \\ &= \sum_{l=1}^{H}\mathbf{P}_{l,\mathbf{w}}\mathbf{V}_{l,\mathbf{w}}\mathbf{V}_{l,\mathbf{w}}^{\mathrm{T}}-\mathbf{m}(\mathbf{w})\mathbf{m}(\mathbf{w})^{\mathrm{T}},\end{aligned} \tag{1.4}$$

where $\mathbf{P}_{l,\mathbf{w}}=\mathrm{pr}(\widetilde{Y}=l|\mathbf{w})$, $\mathbf{V}_{l,\mathbf{w}}=\mathrm{E}\{\mathbf{X}|\widetilde{Y}=l,\mathbf{w}\}$ 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 the Appendix.

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

$$\mathrm{E}(\mathbf{X}|\widetilde{Y}=l,\mathbf{w})=\frac{\mathrm{E}\{\mathbf{X}\mathbb{1}(\widetilde{Y}=l)|\mathbf{w}\}}{\mathrm{E}\{\mathbb{1}(\widetilde{Y}=l)|\mathbf{w}\}}, \tag{1.5}$$

where $\mathbb{1}(.)$ is the indicator function. Proposition 2 shows that the term $\mathbf{V}_{l,\mathbf{w}}$ in (1.4) can be written as a fraction of simple conditional expectations. We can rewrite the kernel matrix $\mathbf{M}_{\text{SIR}}(\mathbf{w})$ in the following form:

$$\mathbf{M}_{\text{SIR}}(\mathbf{w}) = \sum_{l=1}^{H} \frac{\mathbf{U}_{l,\mathbf{w}}\mathbf{U}_{l,\mathbf{w}}^{\mathrm{T}}}{\mathbf{P}_{l,\mathbf{w}}} - \mathbf{m}(\mathbf{w})\mathbf{m}(\mathbf{w})^{\mathrm{T}}, \tag{1.6}$$

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

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 $\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\mathbf{M}_{\text{SIR}}(\mathbf{w})$ 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 $\mathbf{P}_{l,\mathbf{w}}$ are given by

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

Then it's straightforward to obtain the sample estimator of $\mathbf{M}_{\text{SIR}}(\mathbf{w})$ by substituting $\widehat{\mathbf{m}}(\mathbf{w})$, $\widehat{\mathbf{U}}_{l,\mathbf{w}}$, and $\widehat{\mathbf{P}}_{l,\mathbf{w}}$ into equation (1.6), that is,

$$\widehat{\mathbf{M}}_{\mathrm{SIR}}(\mathbf{w}) = \sum\nolimits_{l=1}^{H} \frac{\widehat{\mathbf{U}}_{l,\mathbf{w}}\widehat{\mathbf{U}}_{l,\mathbf{w}}^{\mathrm{T}}}{\widehat{\mathbf{P}}_{l,\mathbf{w}}} - \widehat{\mathbf{m}}(\mathbf{w})\widehat{\mathbf{m}}(\mathbf{w})^{\mathrm{T}}. \tag{1.7}$$

Remark 1. *As for the estimation of the conditional covariance matrix, one may use different bandwidths for different elements of $\boldsymbol{\Sigma}_{\mathbf{w}}$. 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 Section .*

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

$$\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\mathbf{M}_{\mathrm{SIR}}(\mathbf{w}) = \sum_{k=1}^{p} \lambda_k^{\mathrm{SIR}}(\mathbf{w})\boldsymbol{\beta}_k^{\mathrm{SIR}}(\mathbf{w})\boldsymbol{\eta}_k^{\mathrm{SIR}}(\mathbf{w}), \quad \lambda_1^{\mathrm{SIR}}(\mathbf{w}) \geq \cdots \geq \lambda_d^{\mathrm{SIR}}(\mathbf{w}) = 0 = \cdots = \lambda_p^{\mathrm{SIR}}(\mathbf{w}),$$
$$\widehat{\boldsymbol{\Sigma}}_{\mathbf{w}}^{-1}\widehat{\mathbf{M}}_{\mathrm{SIR}}(\mathbf{w}) = \sum_{k=1}^{p} \widehat{\lambda}_k^{\mathrm{SIR}}(\mathbf{w})\widehat{\boldsymbol{\beta}}_k^{\mathrm{SIR}}(\mathbf{w})\widehat{\boldsymbol{\eta}}_k^{\mathrm{SIR}}(\mathbf{w}), \quad \widehat{\lambda}_1^{\mathrm{SIR}}(\mathbf{w}) \geq \cdots \geq \widehat{\lambda}_d^{\mathrm{SIR}}(\mathbf{w}) \geq \cdots \geq \widehat{\lambda}_p^{\mathrm{SIR}}(\mathbf{w}),$$

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 $\widehat{\boldsymbol{\beta}}_k^{\mathrm{SIR}}(\mathbf{w})$ are presented in Theorem 1.

Theorem 1. *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 the 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 (S2.9), (S2.11), (S2.12), (S2.22) and (S2.23) in the Appendix, respectively.

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

Recall that when we estimate the $\mathcal{S}_{Y_{\mathbf{w}}|\mathbf{X}_{\mathbf{w}}}$ in Section , 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 $\mathbf{G}(\mathbf{w}) = \boldsymbol{\Sigma}_\mathbf{w}^{-1}\mathbf{M}(\mathbf{w})$, and $d(\mathbf{w})$ be the rank of $\mathbf{G}(\mathbf{w})$. Rearrange the eigenvalues of $\mathbf{G}(\mathbf{w})$ as $\lambda_1(\mathbf{w}) \geq \ldots \geq \lambda_d(\mathbf{w}) > 0 = \lambda_{d+1}(\mathbf{w}) = \ldots = \lambda_p(\mathbf{w})$, and denote the corresponding eigenvectors by $\boldsymbol{\beta}_1(\mathbf{w}), \ldots, \boldsymbol{\beta}_p(\mathbf{w})$. 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 $\{\widehat{\lambda}_1(\mathbf{w}), \ldots, \widehat{\lambda}_p(\mathbf{w}), \widehat{\boldsymbol{\beta}}_1(\mathbf{w}), \ldots, \widehat{\boldsymbol{\beta}}_p(\mathbf{w})\}$ and $\{\lambda^*_1(\mathbf{w}), \ldots, \lambda^*_p(\mathbf{w}), \boldsymbol{\beta}^*_1(\mathbf{w}), \ldots, \boldsymbol{\beta}^*_p(\mathbf{w})\}$ for $\widehat{\mathbf{G}}(\mathbf{w})$ and $\mathbf{G}^*(\mathbf{w})$. For each $k < p$, let

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

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

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

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

As in Ye and Weiss (2003), $1 - |\det\{\widehat{\mathbf{T}}_k^{\mathrm{T}}(\mathbf{w}) \mathbf{T}_{k,i}^*(\mathbf{w})\}|$ 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 1 in the Appendix.

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

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

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 2 gives the asymptotic property of $\phi_n(\mathbf{w}, \cdot)$ in the Appendix. Now, we can define the objective function of our estimator as

$$g_n(\mathbf{w}, \cdot) : \{0, \dots, p-1\} \to \mathbb{R}, \quad g_n(\mathbf{w}, k) = f_n(\mathbf{w}, k) + \phi_n(\mathbf{w}, k), \tag{1.8}$$

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 small 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 $k = d(\mathbf{w})$.

As a rule of thumb, 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, thus it suffices to minimize $g_n(\mathbf{w}, \cdot)$ over $\{0, 1, \dots, \lfloor p/\log(p) \rfloor\}$, which yields

$$\begin{aligned} g_n(\mathbf{w}, k) &= f_n(\mathbf{w}, k) + \phi_n(\mathbf{w}, k) \\ &= \frac{f_n^0(\mathbf{w}, k)}{1 + \sum_{i=0}^{\lfloor p/\log(p) \rfloor} f_n^0(\mathbf{w}, i)} + \frac{\widehat{\lambda}_{k+1}(\mathbf{w})}{1 + \sum_{i=0}^{\lfloor p/\log(p) \rfloor} \widehat{\lambda}_{i+1}(\mathbf{w})}. \end{aligned} \tag{1.9}$$

Let $\mathcal{D}(f)$ denote the domain of a function f. Since the sample estimator of

$\mathbf{G}(\mathbf{w})$ is a nonparametric estimator, which is similar to the ladle estimator in Luo and Li (2016), we define the nonparametric ladle estimator for $d(\mathbf{w})$ by

$$\widehat{d}(\mathbf{w}) = \arg\min\{g_n(\mathbf{w}, k) : k \in \mathcal{D}[g_n(\mathbf{w}, \cdot)]\}, \tag{1.10}$$

where $g_n(\mathbf{w}, \cdot)$ is defined by (1.8) if $p \leq 10$ or by (1.9) if $p > 10$.

The following theorem 2 establishes the consistency of the nonparametric ladle estimator for variable dependent partial dimension reduction.

Theorem 2. *Under assumptions (C9), (C10), (C11) and (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 (1.10) enjoys the following property:*

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

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 $f_{\mathbf{X}|\widetilde{Y}=l,\mathbf{W}=\mathbf{w}}(\mathbf{x})$, thus, $\mathbf{X}|\mathbf{w}$ follows a mixture multivariate normal distribution, and its likelihood function is given by

$$\begin{aligned} L(\mathbf{m}_l(\mathbf{w}), \mathbf{\Sigma}_{l\mathbf{w}}|\mathbf{x}) &= \prod_{i=1}^{n} \sum_{l=1}^{H} \mathbf{P}_{l,\mathbf{w}} f_{\mathbf{X}|\widetilde{Y}=l,\mathbf{W}=\mathbf{w}}(\mathbf{x_i}) \\ &= \prod_{i=1}^{n} \sum_{l=1}^{H} \mathbf{P}_{l,\mathbf{w}} \left[\frac{e^{-\frac{1}{2}\{\mathbf{x}_i - \mathbf{m}_l(\mathbf{w})\}^{\mathrm{T}} \mathbf{\Sigma}_{l,\mathbf{w}}^{-1} \{\mathbf{x}_i - \mathbf{m}_l(\mathbf{w})\}}}{\sqrt{(2\pi)^p |\mathbf{\Sigma}_{l,\mathbf{w}}|}} \right], \end{aligned} \tag{1.11}$$

where $\mathbf{P}_{l,\mathbf{w}}$ 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

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

Since it's too hard to calculate the log-likelihood type of (1.11) 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

$$CV(h_l) = \frac{1}{n_l}\sum_{i=1}^{n_l}\left[\{\mathbf{x}_i - \widehat{\mathbf{m}}_l^{(-i)}(\mathbf{w})\}^{\mathrm{T}}\widehat{\boldsymbol{\Sigma}}_{\mathbf{w}(-i)}^{-1}(\mathbf{w})\{\mathbf{x}_i - \widehat{\mathbf{m}}_l^{(-i)}(\mathbf{w})\} + \log|\widehat{\boldsymbol{\Sigma}}_{\mathbf{w}(-i)}(\mathbf{w})|\right] \tag{1.12}$$

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 (1.11) as h_{opt}, which is the bandwidth selected for variable dependent partial SIR.

5. Variable Dependent Partial SAVE and DR

In this section, we also prove the large sample properties of proposed variable dependent partial SAVE and DR. We present the estimation procedure of variable dependent partial SAVE and DR in the Appendix. The order determination for variable dependent SAVE and DR is similar to the variable dependent SIR, which replaces the SIR matrix with SAVE or DR matrix. And in the Appendix, we also introduce the bandwidth selection for variable dependent SAVE and DR in detail, which are quite different from the variable dependent partial SIR.

Using large sample theory and singular value decomposition, we get the asymptotic normality of $\widehat{\mathbf{M}}_{\mathrm{SAVE}}(\mathbf{w})$ and the asymptotic expansion of

$\widehat{\boldsymbol{\beta}}_k^{\mathrm{SAVE}}(\mathbf{w})$ as follows.

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), (A2) and assumptions (C1)-(C8) listed in Appendix, we have*

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

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

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

where the closed form of ω_0, $\mathbf{B}_{\mathrm{SAVE}}(\mathbf{w})$, $\mathbf{C}^{\mathrm{SAVE}}(\mathbf{w})$, $\mathbf{B}_k^{\mathrm{SAVE}}(\mathbf{w})$ and $\boldsymbol{\Sigma}_k^{\mathrm{SAVE}}(\mathbf{w})$ are provided by (S2.11), (S3.31), (S3.34), (S3.35) and (S3.36) in the Appendix, respectively.

Similarly, the asymptotic normality of $\widehat{\mathbf{M}}_{\mathrm{DR}}(\mathbf{w})$ and the asymptotic expansion of $\widehat{\boldsymbol{\beta}}_k^{\mathrm{DR}}(\mathbf{w})$ are presented in Theorem 4.

Theorem 4. *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 function of $\mathbf{W}$. Under the condition (A1),* *(A2) and assumptions (C1)-(C8) listed in Appendix, we have*

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

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

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

where the closed form of ω_0, $\mathbf{B}_{\mathrm{DR}}(\mathbf{w})$, $\mathbf{C}^{\mathrm{DR}}(\mathbf{w})$, $\mathbf{B}_k^{\mathrm{DR}}(\mathbf{w})$ and $\boldsymbol{\Sigma}_k^{\mathrm{DR}}(\mathbf{w})$ are provided by (S2.11), (S4.38) (S4.40), (S4.41) and (S4.42) in the Appendix, respectively.

6. Simulation studies

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

Model I: $Y=X_1|W|+3X_2\cos W+0.2\varepsilon$,

Model II: $Y=2\exp\{X_1\exp(W)-X_2\cos W+1\}\cdot\mathrm{sign}\{0.01X_1\cos W+2(W+1)^2X_2\}+0.2\varepsilon$,

Model III: $Y=\{X_1\sin(W)+5X_2\cos(W)\}^2+0.2\varepsilon$,

Model IV: $Y=\exp\{(X_1|W|+X_2)^2\}\log\{(X_3\cos W)^2\}+0.2\varepsilon$,

Model V: $Y=10\frac{\exp\{X_1\sin W+5X_2|W|\}}{X_1\exp(W)-X_2\cos W}+0.2\varepsilon$, Model VI: $Y = \frac{X_1(W_2+10)+X_2\sin W_1+7}{X_1\exp(W_1)+10X_2\cos W_2} + 0.2\varepsilon$,

where $\mathrm{sign}(\cdot)$ is the sign function, and $\varepsilon \sim N(0,1)$. For Models I-V, W is univariate and has a uniform distribution $U(-1,1)$, $\mathbf{X}|\mathbf{W} \sim N_p(\mathbf{m}(W), \mathbf{\Sigma}_W)$, where $\mathbf{m}(W) = \frac{\sin(W)}{2}\mathbf{1}_p$, $\mathbf{\Sigma}_W = (\sigma_{ij})_{p\times p}$ with $\sigma_{ij} = 1$ for $i = j$, $\sigma_{ij} = \frac{1}{2}\sin(W)$ for $i \neq j$. For Model VI, $\mathbf{W} = (W_1, W_2)^{\mathrm{T}}$ is a two-dimensional vector with $W_1, W_2 \overset{\text{iid}}{\sim} U(-1,1)$, $\mathbf{X}|\mathbf{W} \sim N_p(\mathbf{m}(\mathbf{W}), \mathbf{\Sigma}_{\mathbf{W}})$, where $\mathbf{m}(\mathbf{W}) = \frac{\sin(W_1)+\cos(W_2)}{2}\mathbf{1}_p$, and $\mathbf{\Sigma}_{\mathbf{W}} = (\sigma_{ij})_{p\times p}$ with $\sigma_{ij} = 1$ for $i = j$, $\sigma_{ij} = \frac{1}{2}(\sin(W_1)+\cos(W_2))$ for $i \neq j$. Hence, $d(\mathbf{w}) = 1$ for Models I and III, and $d(\mathbf{w}) = 2$ for Models II, IV and VI. For Model V, $d(\mathbf{w}) = 2$ when $\mathbf{w} \neq 0$, and 1 when $\mathbf{w} = 0$.

Our simulation studies are two folds. We first use the nonparametric ladle estimator in (1.10) 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.

6.1. Estimation of Structural Dimension

Table 1: Correct order determinations among 100 runs for Models I-V

Model	$\mathbf{w}$	$(n,p)=(150,5)$			$(n,p)=(300,10)$		
		VDPSIR	VDPSAVE	VDPDR	VDPSIR	VDPSAVE	VDPDR
I	0	100	97	100	100	90	100
	-1	100	96	100	99	91	99
	1	100	96	100	100	89	99
	-0.5	100	96	100	98	93	99
	0.5	100	96	100	99	92	99
II	0	100	86	100	100	87	100
	-1	100	84	100	100	86	100
	1	100	85	100	100	87	100
	-0.5	100	86	99	100	86	100
	0.5	100	86	100	100	87	100
III	0	7	99	95	17	90	84
	-1	8	96	94	17	87	84
	1	8	98	95	18	88	85
	-0.5	7	97	96	18	88	86
	0.5	9	99	96	14	88	82
IV	0	0	97	97	1	99	99
	-1	0	97	97	1	99	99
	1	0	98	98	3	99	99
	-0.5	0	99	98	1	99	99
	0.5	0	98	97	3	99	99
V	-1	100	72	73	99	70	70
	1	100	74	83	99	71	79
	-0.5	100	71	79	98	70	77
	0.5	100	76	87	99	72	82

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 $\mathbf{w}$. The percentages of correct order estimates in 100 runs 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 IV, and outperforms the other two variable dependent PDR methods for the remaining models due to the small sample sizes we have ($n = 150$

or 300) in our simulation studies.

Table 2: Order determination for Model V with $\mathbf{w}$ in a neighborhood of 0

Model	$\mathbf{w}$	$(n,p)=(150,5)$			$(n,p)=(300,10)$		
		VDPSIR	VDPSAVE	VDPDR	VDPSIR	VDPSAVE	VDPDR
V	0.01	100	72	83	99	70	80
	0.02	100	74	82	98	71	79
	0.03	100	75	85	100	72	81
	0.04	100	71	83	99	69	80
	0.05	100	72	82	99	71	79
	-0.01	100	72	83	98	69	80
	-0.02	100	75	83	99	72	81
	-0.03	100	76	80	99	73	77
	-0.04	100	71	80	98	69	78
	-0.05	100	69	84	99	67	81

Table 2 shows that, for Model V, under the scenario $(n,p)=(150,5)$, if $\mathbf{w}\neq 0$, the structural dimension can be accurately estimated by variable dependent partial SIR approach all the times. When $(n,p)=(300,10)$, at least 98% of the times, variable dependent partial SIR provides with the correct estimates of $d=2$.

Table 3 suggests that the percentage of correct estimation for Model V when $\mathbf{w}=0$ 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.

Table 3: Order determination for Model V with $\mathbf{w}=0$

(n,p)	VDPSIR	VDPSAVE	VDPDR	(n,p)	VDPSIR	VDPSAVE	VDPDR
(150, 5)	62	21	25	(300, 10)	59	19	27
(500, 5)	73	51	60	(500, 10)	65	42	53
(800, 5)	86	74	82	(800, 10)	75	57	68
(1000, 5)	93	88	90	(1000, 10)	81	70	78

For Model VI, since $\mathbf{W}=(W_1,W_2)$ is two-dimensional, we present the order determination results for different $\mathbf{w}$ values in Tables 4 and 5. The entries are the counts of correct estimates of the structural dimension out of 100 repetitions by variable dependent partial SIR, variable dependent partial SAVE, variable dependent partial DR, respectively.

Table 4: Order Determination for Model VI with $(n, p) = (150, 5)$

$\mathbf{w}_1$ \ $\mathbf{w}_2$	-1			-0.5			0			0.5			1		
-1	88	76	85	88	76	86	92	78	86	90	84	86	91	78	86
-0.5	90	77	87	90	77	84	89	80	87	89	82	86	93	77	87
0	89	77	85	89	84	85	91	75	87	89	78	87	88	77	88
0.5	91	77	87	90	75	87	91	76	87	91	80	87	88	80	89
1	92	78	83	91	79	87	89	76	84	88	79	86	89	85	86

Table 5: Order Determination for Model VI with $(n, p) = (300, 10)$

$\mathbf{w}_1$ \ $\mathbf{w}_2$	-1			-0.5			0			0.5			1		
-1	97	80	98	97	81	98	97	78	98	96	81	98	95	78	98
-0.5	96	84	98	95	78	98	95	77	99	97	81	98	97	86	98
0	97	82	99	97	78	98	95	80	98	95	77	98	97	77	98
0.5	95	79	98	96	83	98	96	86	98	95	80	99	96	82	98
1	96	79	99	98	79	98	97	79	99	96	80	99	97	80	98

Table 4 and 5 suggest that our nonparametric ladles works pretty well for estimation of structural dimension when $\mathbf{W}$ is two dimensional.

We also use PDEE methods to estimate the dimension of $\mathbf{B}$ based on the ladle estimator. Since PDEE methods cannot estimate the dimension of $\mathbf{B}(\mathbf{w})$, it could not deal with situations such as Model V, the following table gives the number of correct estimates of structural dimensions among 100 simulation runs for all models except Model V. It is easy to see that PDEE methods perform poorly regarding the order determination, especially for complex models. Further, simulation results (unreported here) also show that, when PDEE fails to select the correct structural dimension, it tends to under-select the dimension than to over-select it, which makes PDEE even less desirable in practice.

Table 6: Estimated structural dimension based on 100 replications for PDEE methods

		PDEE-SIR	PDEE-SAVE	PDEE-DR
(n,p)=(150,5)	Model I	63	5	100
	Model II	0	0	0
	Model III	0	6	39
	Model IV	0	0	8
	Model VI	0	0	1
(n,p)=(300,10)	Model I	100	11	100
	Model II	0	0	0
	Model III	0	9	76
	Model IV	0	0	18
	Model VI	0	0	0

6.2. 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 Ferre (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,

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

where $\mathbf{P}_{\mathcal{S}_1}$ and $\mathbf{P}_{\mathcal{S}_2}$ are orthogonal projections onto $\mathcal{S}_1$ and $\mathcal{S}_2$, and $\mathrm{Tr}(\cdot)$ 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 7 the 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 7 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.

Simulation results for Model VI are provided in Tables 8 and 9 based on different combinations of $\mathbf{W} = (W_1, W_2)$. These results reaffirm the

Table 7: Trace Correlation for the Models 1-5

Model	$\mathbf{w}$	$(n,p)=(150,5)$			$(n,p)=(300,10)$		
		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

good performance of our variable dependent PDR methods even when $\mathbf{W}$ is multivariate.

Table 8: Trace Correlation for Model 6 with $(n,p)=(150,5)$

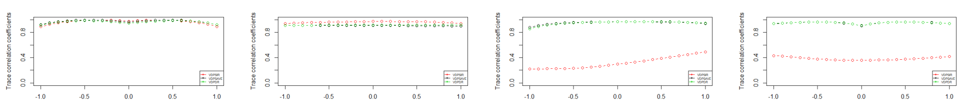
$\mathbf{w}_1$ \ $\mathbf{w}_2$	-1			-0.5			0			0.5			1		
-1	.851	.829	.842	.874	.868	.876	.880	.890	.888	.862	.892	.874	.815	.872	.825
-0.5	.884	.861	.873	.906	.896	.898	.912	.912	.915	.894	.915	.931	.864	.899	.893
0	.905	.881	.907	.925	.912	.919	.929	.924	.931	.915	.927	.92	.886	.911	.921
0.5	.912	.893	.90	.931	.919	.925	.935	.924	.931	.921	.920	.918	.894	.903	.896
1	.900	.883	.892	.925	.900	.925	.930	.908	.922	.916	.894	.910	.883	.859	.864

The following Figures give a direct visual presentation of our simulation results, which also agree with what we discussed previously. It seems that variable dependent partial SAVE and DR are very reliable and accurate

Table 9: Trace Correlation for Model VI with $(n, p) = (300, 10)$

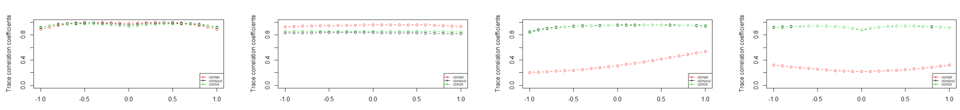
$\mathbf{w}_2$ / $\mathbf{w}_1$	-1			-0.5			0			0.5			1		
-1	.872	.830	.880	.891	.861	.875	.890	.880	.885	.885	.883	.878	.824	.869	.837
-0.5	.894	.859	.876	.921	.889	.911	.923	.910	.915	.906	.909	.903	.879	.896	.854
0	.915	.878	.908	.931	.910	.924	.945	.921	.926	.923	.922	.91	.894	.908	.8778
0.5	.927	.890	.921	.947	.918	.932	.951	.921	.945	.936	.918	.931	.903	.899	.900
1	.907	.876	.898	.932	.892	.92	.936	.903	.924	.925	.890	.916	.889	.856	.876

regarding the estimation of the variable dependent partial CS in all models, while variable dependent partial SIR might fail under certain conditions. We recommend variable dependent partial DR, and use variable dependent partial SIR as a complementary method.



(a) Model 1 (b) Model 2 (c) Model 3 (d) Model 4

Figure 1: Trace correlation coefficient for Models I-IV with $(n, p) = (150, 5)$

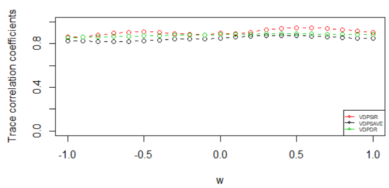


(a) Model 1 (b) Model 2 (c) Model 3 (d) Model 4

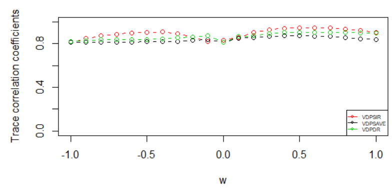
Figure 2: Trace correlation coefficient for Models I-IV with $(n, p) = (300, 10)$

The 3D plot for Model VI is omitted since it cannot clearly demonstrate the trend of trace correlation coefficient when $\mathbf{W}$ is two-dimensional.

Furthermore, to illustrate the advantages of our variable dependent approach, we compare the average trace correlation coefficient deriving from



(a) $(n, p) = (150, 5)$



(b) $(n, p) = (300, 10)$

Figure 3: Trace correlation coefficient for Model VI

those at different $\mathbf{w}$ values, with the trace correlation coefficient obtained via PDEE (Feng et al., 2013) estimation methods. Since the PDEE method can not handle variable dependent structural dimension scenario, we only investigate the estimation accuracy in Model I-IV and VI, given a fixed estimated structural dimension. The results of the mean trace correlation coefficient from 100 simulation runs are shown in Table 10.

Table 10: Mean trace correlation coefficients of 100 replications

(n,p)	Model	VDPSIR	VDPSAVE	VDPDR	PDEE-SIR	PDEE-SAVE	PDEE-DR
(150,5)	I	0.9723	0.9646	0.9665	0.9483	0.7848	0.3897
	II	0.8076	0.6940	0.7936	0.7935	0.5222	0.5387
	III	0.1911	0.9147	0.9290	0.4937	0.8638	0.8739
	IV	0.3721	0.9033	0.9278	0.5951	0.8722	0.8832
	VI	0.9090	0.9104	0.9055	0.8214	0.6437	0.7446
(300,10)	I	0.9728	0.9600	0.9633	0.9553	0.6892	0.3234
	II	0.8096	0.5844	0.7881	0.8014	0.3758	0.4826
	III	0.1520	0.9218	0.9207	0.3959	0.7600	0.8461
	IV	0.1929	0.8948	0.9104	0.4931	0.8490	0.8625
	VI	0.8929	0.8631	0.8746	0.8659	0.5161	0.6868

It is noticeable that our proposals consistently outperform PDEE approaches. Our methods can estimate $\mathbf{B}(\mathbf{W})$ at $\mathbf{W} = \mathbf{w}$, which changes with the value of $\mathbf{w}$, while the PDEE approach can only estimate $\mathbf{B}$, which is fixed no matter how $\mathbf{w}$ varies. As a result, the variable dependent PDR is a better approach to handle partial dimension reduction with continuous $\mathbf{W}$.

7. 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 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.

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 (1.10) to estimate $d(\mathbf{w})$, 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 $\widehat{\mathbf{B}}(\mathbf{w})$ and partial dimension reduction directions $\widehat{\mathbf{B}}$. We use the distance correlations (Szekely et al., 2007) between Y and $\widehat{\mathbf{B}}^{\mathrm{T}}(\mathbf{w})\mathbf{X}$ (or $\widehat{\mathbf{B}}^{\mathrm{T}}\mathbf{X}$) to evaluate the performance of the estimates $\widehat{\mathbf{B}}(\mathbf{w})$ and $\widehat{\mathbf{B}}$.

7.1. 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), 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 get $\hat{d}(\mathbf{w}) = \hat{d} = 1$. 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 $\hat{d}(\mathbf{w}) = \hat{d} = 1$.

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 circulationary and respirationary problems, measurements of pollutants and many other environmental factors in Hong Kong. We take *the number of daily total hospital admissions for circulationary and respirationary problems* as the response Y, *time* as W, and the remaining predictors as $\mathbf{X}$. The order determination procedure yields that $\hat{d}(\mathbf{w}) = \hat{d} = 1$.

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 capita crime rate by twon* 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 $\hat{d}(\mathbf{w}) = \hat{d} = 2$.

7.2. Comparison of each method for the data analysis

The distance correlations between Y and $\widehat{\mathbf{B}}(\mathbf{w})^{\mathrm{T}}\mathbf{X}$ are reported in columns

2-4 in Table 11, where $\widehat{\mathbf{B}}(\mathbf{w})$ 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

Table 11: Distance correlation for each estimation approaches

Dataset	VDPSIR	VDPSAVE	VDPDR	PDEE-SIR	PDEE-SAVE	PDEE-DR
Body Fat	0.8301	0.2149	0.5310	0.6906	0.1664	0.2665
Wage	0.5195	0.2215	0.5044	0.5171	0.1512	0.1500
Hongkong environmental	0.4592	0.2060	0.1942	0.2155	0.1379	0.1193
Boston Housing	0.9112	0.6618	0.8688	0.8030	0.3418	0.5408

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

8. Discussions

In this paper, we propose a variable dependent approach to better dealing with partial sufficient dimension reduction. 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 Materials

The proofs of Theorem 1, Theorem 2, Theorem 3 and Theorem 4 are given in the online supplementary document.

References

[1] Berndt, E. R. (1991) The practice of econometrics: classic and contemporary. *Addison Wesley Publishing Company.*

[2] Bura, E., and Cook, R. D. (2001) Estimating the structural dimension of regressions via parametric inverse regression. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **63(2)** 393-410.

[3] Chen, X., Zou, C., and Cook, R. D. (2010) Coordinate-independent sparse sufficient dimension reduction and variable selection. *The Annals of Statistics*, **38(6)** 3696-3723.

[4] Chiaromonte, F., Cook, R. D., and Li, B. (2002) Sufficient dimension reduction in regressions with categorical predictors. *The Annals of Statistics*, 475-497.

[5] Cook, R. D. (1998) Regression Graphics. *New York: Wiley*, **237**.

[6] Cook, R. D., and Forzani, L. (2009) Likelihood-based sufficient dimension reduction. *Journal of the American Statistical Association*, **104(485)** 197-208.

[7] Cook, R. D., and Ni, L. (2005) Sufficient dimension reduction via inverse regression: A minimum discrepancy approach. *Journal of the American Statistical Association*, **100(470)** 410-428.

[8] Cook, R. D., and Weisberg, S. (1991) Comment. *Journal of the American Statistical Association*, **86(414)** 328-332.

[9] Cook, R. D., and Yin, X. (2001) Theory and methods: Special invited paper: Dimension reduction and visualization in discriminant analysis (with discussion). *Australian and New Zealand Journal of Statistics*, **43(2)** 147-199.

[10] Dong and Li (2010) Dimension reduction for non-elliptically distributed predictors: second-order methods. *Biometrika*, **97(2)** 279-294.

[11] Fan, J., and Huang, T. (2005) Profile likelihood inferences on semiparametric varying-coefficient partially linear models. *Bernoulli*, **11(6)** 1031-1057.

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

[13] Ferré, L. (1998) Determining the dimension in sliced inverse regression

and related methods. *Journal of the American Statistical Association*, **93(441)** 132-140.

[14] Hoeting, J. A., Madigan, D., Raftery, A. E., and Volinsky, C. T. (1999) Bayesian model averaging: a tutorial. *Statistical Science*, 382-401.

[15] Jiang, B., Chen, Z., and Leng, C. (2017) Dynamic Linear Discriminant Analysis in High Dimensional Space.

[16] Leng, C. (2010) Variable selection and coefficient estimation via regularized rank regression. *Statistica Sinica*, 167-181.

[17] Li, K. C. (1991) Sliced inverse regression for dimension reduction. *Journal of the American Statistical Association*, **86(414)** 316-327.

[18] Li, B. (2018) Sufficient Dimension Reduction: Methods and Applications with R. *CRC Press*.

[19] Li, B., Cook, R. D., and Chiaromonte, F. (2003) Dimension reduction for the conditional mean in regressions with categorical predictors. *The Annals of Statistics*, 1636-1668.

[20] Li, B., and Dong, Y. (2009) Dimension reduction for nonelliptically distributed predictors. *The Annals of Statistics*, **37(3)** 1272-1298.

[21] Li, D., Ke, Y., and Zhang, W. (2015) Model selection and structure specification in ultra-high dimensional generalised semi-varying coefficient models. *The Annals of Statistics*, **43(6)** 2676-2705.

[22] Li, B., and Wang, S. (2007) On directional regression for dimension reduction. *Journal of the American Statistical Association*, **102(479)** 997-1008.

[23] Li, Y., and Zhu, L. X. (2007) Asymptotics for sliced average variance estimation. *The Annals of Statistics*, **35(1)** 41-69.

[24] Luo, W., and Li, B. (2016) Combining eigenvalues and variation of eigenvectors for order determination. *Biometrika*, **103(4)** 875-887.

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

[26] Ma, Y., and Zhu, L. (2013a) A review on dimension reduction. *International Statistical Review*, **81(1)** 134-150.

[27] Ma, Y., and Zhu, L. (2013b) Efficient estimation in sufficient dimension reduction. *The Annals of Statistics*, **41(1)** 250.

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

[29] Nadaraya (1964) On estimating regression. *Theory of Probability* & *Its Applications*, **9(1)** 141-142.

[30] Penrose, K. W., Nelson, A. G., and Fisher, A. G. (1985) Generalized body composition prediction equation for men using simple measurement techniques. *Medicine* & *Science in Sports* & *Exercise*, **17(2)** 189.

[31] Székely, G. J., Rizzo, M. L., and Bakirov, N. K. (2007) Measuring and testing dependence by correlation of distances. *The Annals of Statistics*, **35(6)** 2769-2794.

[32] Wang, H., and Xia, Y. (2008) Sliced regression for dimension reduction. *Journal of the American Statistical Association*, **103(482)** 811-821.

[33] Watson, G. S. (1964) Smooth regression analysis. *Sankhyi: The Indian Journal of Statistics, Series A*, 359-372.

[34] Wen, X., and Cook, R. D. (2007) Optimal sufficient dimension reduction in regressions with categorical predictors. *Journal of Statistical Planning and Inference*, 137(6) 1961-1978.

[35] Xia, Y., Tong, H., Li, W. K., & Zhu, L. X. (2002). An adaptive estimation of dimension reduction space. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **64(3)**, 363-410.

[36] Xie, H., and Huang, J. (2009) SCAD-penalized regression in high-dimensional partially linear models. *The Annals of Statistics*, **37(2)** 673-696.

[37] Xue, Y., Wang, Q., and Yin, X. (2018) A unified approach to sufficient dimension reduction. *Journal of Statistical Planning and Inference*, **197** 168-179.

[38] Ye, Z., and Weiss, R. E. (2003) Using the bootstrap to select one of a new class of dimension reduction methods. *Journal of the American Statistical Association*, **98(464)** 968-979.

[39] Yin, X., and Cook, R. D. (2002) Dimension reduction for the conditional kth moment in regression. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **64(2)** 159-175.

[40] Yin, X., and Cook, R. D. (2003) Estimating central subspaces via inverse third moments. *Biometrika*, **90(1)** 113-125.

[41] Yin, J., Geng, Z., Li, R., and Wang, H. (2010) Nonparametric covariance model. *Statistica Sinica*, **20** 469.

[42] Zhang, R., Zhao, W., and Liu, J. (2013) Robust estimation and variable selection for semiparametric partially linear varying coefficient model based on modal regression. *Journal of Nonparametric Statistics*, **25(2)** 523-544.

School of Statistics, East China Normal University, Shanghai 200241, China.

E-mail: 52174404009@stu.ecnu.edu.cn

School of Statistics, East China Normal University, Shanghai 200241, China.

E-mail: kaitan.stat.@uky.edu

Department of Mathematics and Statistics, Missouri University of Science and Technology, Missouri 65409, The United States.

E-mail: wenx@mst.edu

School of Statistics, East China Normal University, Shanghai 200241, China.

E-mail: zyu@stat.ecnu.edu.cn

SUPPLEMENT TO "VARIABLE DEPENDENT PARTIAL DIMENSION REDUCTION"

Lu Li, Kai Tan, Xuerong Meggie Wen and Zhou Yu

East China Normal University, East China Normal University,

Missouri University of Science and Technology, East China Normal University

S1. Regularity conditions

To prove the theoretical results of this paper, we need some regularity conditions. They are not the weakest possible conditions, but they are imposed to facilitate the proofs.

(C1). (The density of the index variable) We assume that W has a compact support. Then, on the support, we further assume that the probability density function of W, denoted by $f(W)$, is bounded away from 0 and has continuous derivatives up to second order.

(C2). (The moment requirement) For any $1 \leq j_1, j_2 \leq p$, there exists a constant $\delta \in [0, 1]$, such that $\sup_{\mathbf{w}} \mathrm{E}\{|\mathbf{X}_{j_1}(\mathbf{w})\mathbf{X}_{j_2}(\mathbf{w})|\}^{2+\delta} < \infty$.

(C3). (Smoothness of the conditional mean) Assume that the conditional mean $\mathbf{m}_j(\cdot)$ has continuous derivatives up to second order.

(C4). (Smoothness of the conditional variance) We assume that $\mathrm{E}\{\mathbf{X}_{j_1}^{k_1}\mathbf{X}_{j_2}^{k_2}\mathbf{X}_{j_3}^{k_3}\mathbf{X}_{j_4}^{k_4}|\mathbf{W} = \mathbf{w}\}$ has continuous derivatives up to second order in $\mathbf{w}$ for $k_1, k_2, k_3, k_4 \in \{0, 1\}$, where j_1, j_2, j_3, j_4 are not necessarily different components in the $(\mathbf{X}, \mathbf{W})$ vectors.

(C5). (Smoothness of the conditional indicator mean) Assume that the conditional indicator mean $\mathrm{E}\{\mathbf{X}\mathbb{1}(Y \in \mathbf{J}_l)|\mathbf{W} = \mathbf{w}\}$ has continuous derivatives up to second order, where $l = 1, \ldots, H$.

(C6). (Smoothness of the conditional indicator variance) We assume that $\mathrm{E}\{\mathbf{X}_{j_1}^{k_1}\mathbf{X}_{j_2}^{k_2}\mathbf{X}_{j_3}^{k_3}$
$\mathbf{X}_{j_4}^{k_4}\mathbb{1}(\widetilde{Y} = l)|\mathbf{W} = \mathbf{w}\}$ has continuous derivatives up to second order in $\mathbf{w}$ for $k_1, k_2, k_3, k_4 \in \{0, 1\}$, $l = 1, \ldots, H$, where j_1, j_2, j_3, j_4 are not necessarily different components in the $(\mathbf{X}, \mathbf{W})$ vectors.

(C7). (The bandwidth) $h \to 0$ and $nh^5 \to c > 0$ for some $c > 0$.

(C8). (The kernel function) We assume that $K(\mathbf{w})$ is a bounded probability density function symmetric about 0. Furthermore, for the δ in (C2), we assume that $\int K^{2+\delta}(v)v^j dv < \infty$, for $j = 0, 1, 2$. Lastly, for two arbitrary indices $\mathbf{w}_1$ and $\mathbf{w}_2$, we must have $|K(\mathbf{w}_1) - K(\mathbf{w}_2)| \leq K_c|\mathbf{w}_1 - \mathbf{w}_2|$ for some positive constant K_c.

(C9) The bootstrap estimator $\mathbf{M}^*$ satisfies

$$(nh)^{1/2}\{\mathbf{vech}(\mathbf{M}^*) - \mathbf{vech}(\widehat{\mathbf{M}}) - \mathbf{vech}(\mathbf{B}^*)\} \xrightarrow{d} N(0, \mathrm{Var}_F[\mathbf{vech}\{\mathbf{H}(\mathbf{X}, Y)\}]). \tag{S1.1}$$

(C10) For any sequence of nonnegative random variables $\{\mathbf{Z}_n : n = 1, 2, \ldots\}$ involved hereafter, if $\mathbf{Z}_n = \mathbf{O}(c_n)$ for some sequence $\{c_n : n \in N\}$ with $c_n > 0$, then $\mathrm{E}(c_n^{-1}\mathbf{Z}_n)$ exists for each n and $\mathrm{E}(c_n^{-1}\mathbf{Z}_n) = \mathbf{O}(1)$.

(C11) $\sum_{n=3}^{\infty}(h/\log\log n)\mathrm{E}\{S^2\mathbf{I}(|S| > a_n)\} < \infty$, where S represents response variable in nonparametric regression function and $a_n = \mathbf{o}\{(nh^{-1}\log\log n)^{1/2}/(\log n)^2\}$.

(C12) $\lim_{\epsilon\to 0}\limsup_{n\to\infty}\sup_{m\in\Gamma_{n,\epsilon}}|h(m)/h(n) - 1| = 0$, where $\Gamma_{n,\epsilon} = \{m : |m - n| \leq \epsilon n\}$.

Both conditions (C1) and (C2) are standard technical assumptions [Yin et al. (2010)], and conditions (C3), (C4), (C5) and (C6) are necessary smoothness constraints [Fan (1993); Yao and Tong (1998)]. By condition (C7) we know that the optimal convergence rate of $n^{-1/5}$ can be used. Condition (C8) is a standard requirement for the kernel function [Yao and Tong (1996)] , which is trivially satisfied by both Gaussian kernel and Epanechnikov kernel.

Condition (C9) is quite mild: it is satisfied if the statistical functional $\mathbf{M}$ is Fr*é*chet differentiable [Luo and Li (2016)], where $\mathbf{B}^*$ represents the bias term of $\mathbf{M}^* - \widehat{\mathbf{M}}$. Condition (C10) amounts to asserting that the asymptotic behaviour of $(nh)^{1/2}(\mathbf{M}^* - \widehat{\mathbf{M}} - \mathbf{B}^*)$ mimics that of $(nh)^{1/2}(\widehat{\mathbf{M}} - \mathbf{M} - \mathbf{B})$, where $\mathbf{B}$ represents the bias term of $\widehat{\mathbf{M}} - \mathbf{M}$. The validity of this self-similarity has been discussed [Bickel and Freedman (1981), Luo and Li (2016)], where $\mathbf{vech}(\cdot)$ is the vectorization of the upper triangular part of a matrix and $\mathrm{var}_F[\mathbf{vech}\{\mathbf{H}(\mathbf{X}, Y)\}]$ is positive definite. Condition (C11) and condition (C12) are widely used in law of the iterated logarithm for nonparametric regression Hardle (1984). These conditions should not be too restrictive on the applicability of our estimator.

S2. The proofs of the main results

Proof of Proposition 1. Following Li et al. (2003), for convenience, we often use the abbreviation

$$\mathrm{E}\{f(\mathbf{X}, Y)|g(\mathbf{X}), \mathbf{W} = \mathbf{w}\} \triangleq \mathrm{E}\{f(\mathbf{X}_{\mathbf{w}}, Y_{\mathbf{w}})|g(\mathbf{X}_{\mathbf{w}})\}.$$

In our case, we define

$$\mathrm{E}(\mathbf{X}|Y,\mathbf{W}=\mathbf{w}) = \mathrm{E}(\mathbf{X}_{\mathbf{w}}|Y_{\mathbf{w}}), \quad \mathrm{E}\{\mathbf{X}|Y,\mathbf{B}^{\mathrm{T}}(\mathbf{w})\mathbf{X},\mathbf{W}=\mathbf{w}\} = \mathrm{E}\{\mathbf{X}_{\mathbf{w}}|Y_{\mathbf{w}},\mathbf{B}^{\mathrm{T}}(\mathbf{w})\mathbf{X}_{\mathbf{w}}\}.$$

Since $\mathbf{X}_{\mathbf{w}}$ satisfies linear conditional mean (A1), given $\mathbf{W}=\mathbf{w}$, it's easy for us to get

$$\mathrm{E}\{\mathbf{X}_{\mathbf{w}}|\mathbf{B}^{\mathrm{T}}(\mathbf{w})\mathbf{X}_{\mathbf{w}}\} = \left[\mathbf{B}(\mathbf{w})\{\mathbf{B}^{\mathrm{T}}(\mathbf{w})\boldsymbol{\Sigma}_{\mathbf{w}}\mathbf{B}(\mathbf{w})\}^{-1}\mathbf{B}^{\mathrm{T}}(\mathbf{w})\boldsymbol{\Sigma}_{\mathbf{w}}\right]^{\mathrm{T}}\mathbf{X}_{\mathbf{w}}.$$

Hence,

$$\begin{aligned}
&\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\{\mathrm{E}(\mathbf{X}|Y,\mathbf{W}=\mathbf{w})-\mathbf{m}(\mathbf{w})\}\\
&=\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\{\mathrm{E}(\mathbf{X}_{\mathbf{w}}|Y_{\mathbf{w}})-\mathrm{E}(\mathbf{X}_{\mathbf{w}})\}\\
&=\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}[\mathrm{E}\{\mathrm{E}(\mathbf{X}_{\mathbf{w}}|Y_{\mathbf{w}},\mathbf{B}^{\mathrm{T}}(\mathbf{w})\mathbf{X}_{\mathbf{w}})|Y_{\mathbf{w}}\}-\mathrm{E}\{\mathrm{E}(\mathbf{X}_{\mathbf{w}}|\mathbf{B}^{\mathrm{T}}(\mathbf{w})\mathbf{X}_{\mathbf{w}})\}]\\
&=\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}[\mathrm{E}\{\mathrm{E}(\mathbf{X}_{\mathbf{w}}|\mathbf{B}^{\mathrm{T}}(\mathbf{w})\mathbf{X}_{\mathbf{w}})|Y_{\mathbf{w}}\}-\mathrm{E}\{\mathrm{E}(\mathbf{X}_{\mathbf{w}}|\mathbf{B}^{\mathrm{T}}(\mathbf{w})\mathbf{X}_{\mathbf{w}})\}]\\
&=\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}[\mathbf{B}(\mathbf{w})\{\mathbf{B}^{\mathrm{T}}(\mathbf{w})\boldsymbol{\Sigma}_{\mathbf{w}}\mathbf{B}(\mathbf{w})\}^{-1}\mathbf{B}^{\mathrm{T}}(\mathbf{w})\boldsymbol{\Sigma}_{\mathbf{w}}]^{\mathrm{T}}\{\mathrm{E}(\mathbf{X}_{\mathbf{w}}|Y_{\mathbf{w}})-\mathrm{E}(\mathbf{X}_{\mathbf{w}})\}\\
&=\mathbf{B}(\mathbf{w})\{\mathbf{B}^{\mathrm{T}}(\mathbf{w})\boldsymbol{\Sigma}_{\mathbf{w}}\mathbf{B}(\mathbf{w})\}^{-1}\mathbf{B}^{\mathrm{T}}(\mathbf{w})\boldsymbol{\Sigma}_{\mathbf{w}}\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\{\mathrm{E}(\mathbf{X}|Y,\mathbf{W}=\mathbf{w})-\mathbf{m}(\mathbf{w})\}.
\end{aligned}$$

The second equality follows tower property about conditional expectation. Then the third equality is based on the fact that

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

Denote $\mathbf{P}_{\mathbf{B}(\mathbf{w})} = \mathbf{B}(\mathbf{w})\{\mathbf{B}^{\mathrm{T}}(\mathbf{w})\boldsymbol{\Sigma}_{\mathbf{w}}\mathbf{B}(\mathbf{w})\}^{-1}\mathbf{B}^{\mathrm{T}}(\mathbf{w})\boldsymbol{\Sigma}_{\mathbf{w}}$. Then we have,

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

$\square$

Proof of Proposition 2. At first, we need to prove the left side of the equation (1.5),

$$\begin{aligned}
\mathrm{E}(\mathbf{X}|\widetilde{Y} = l, \mathbf{w}) &= \int \mathbf{X} \frac{f(\mathbf{X}, \widetilde{Y} = l, \mathbf{w})}{f(\widetilde{Y} = l, \mathbf{w})} d\mathbf{X} \\
&= \int \mathbf{X} \frac{f(\mathbf{X}, \mathbf{w}|\widetilde{Y} = l)\mathbf{P}(\widetilde{Y} = l)}{f(\mathbf{w}|\widetilde{Y} = l)\mathbf{P}(\widetilde{Y} = l)} d\mathbf{X} \\
&= \frac{\int \mathbf{X} f(\mathbf{X}, \mathbf{w}|\widetilde{Y} = l) d\mathbf{X}}{f(\mathbf{w}|\widetilde{Y} = l)}.
\end{aligned}$$

Then we show the right side of the equation (1.5), since

$$\mathrm{E}(\mathbb{1}(\widetilde{Y}=l)|\mathbf{w}) = \frac{\mathbb{1}(\widetilde{Y}=l)\sum_{k=1}^{H} f(\mathbf{w}|\widetilde{Y}=k)\mathbf{P}(\widetilde{Y}=k)}{f(\mathbf{w})} = \frac{f(\mathbf{w}|\widetilde{Y}=l)\mathbf{P}(\widetilde{Y}=l)}{f(\mathbf{w})}, \tag{S2.2}$$

where $f(\mathbf{w})$ is the density function of $\mathbf{w}$. Hence

$$\begin{aligned}
\mathrm{E}(\mathbf{X}\mathbb{1}(\widetilde{Y}=l)|\mathbf{w}) &= \int \mathbf{X}\mathbb{1}(\widetilde{Y}=l)\frac{f(\mathbf{X},\widetilde{Y}=l,\mathbf{w})}{f(\mathbf{w})}d\mathbf{X}d\widetilde{Y} \\
&= \int \mathbf{X}\mathbb{1}(\widetilde{Y}=l)\frac{\sum_{k=1}^{H} f(\mathbf{X},\mathbf{w}|\widetilde{Y}=k)\mathbf{P}(\widetilde{Y}=k)}{f(\mathbf{w})}d\mathbf{X} \\
&= \frac{\int \mathbf{X}f(\mathbf{X},\mathbf{w}|\widetilde{Y}=l)\mathbf{P}(\widetilde{Y}=l)}{f(\mathbf{w})}.
\end{aligned} \tag{S2.3}$$

Thus, we get

$$\begin{aligned}
\frac{\mathrm{E}\{\mathbf{X}\mathbb{1}(\widetilde{Y}=l)|\mathbf{w}\}}{\mathrm{E}\{\mathbb{1}(\widetilde{Y}=l)|\mathbf{w}\}} &= \frac{\int \mathbf{X}f(\mathbf{X},\mathbf{w}|\widetilde{Y}=l)P(\widetilde{Y}=l)d\mathbf{X}/f(\mathbf{w})}{f(\mathbf{w}|\widetilde{Y}=l)P(\widetilde{Y}=l)/f(\mathbf{w})} \\
&= \frac{\int f(\mathbf{X},\mathbf{w}|\widetilde{Y}=l)d\mathbf{X}}{f(\mathbf{w}|\widetilde{Y}=l)} = \mathrm{E}(\mathbf{X}|\widetilde{Y}=l,\mathbf{w}).
\end{aligned}$$

$\square$

Proof of Theorem 1. Denote $k_i(\mathbf{w}) = K(\frac{\mathbf{w}_i - \mathbf{w}}{h})$, we get

$$\frac{1}{nh}\sum_{i=1}^{n} k_i(\mathbf{w}) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{\mathbf{w}_i - \mathbf{w}}{h}\right) = \widehat{f}(\mathbf{w}).$$

Then under condition (C1) and (C8), by Yin et al. (2010), we conclude that $|\widehat{f}(\mathbf{w}) - f(\mathbf{w})| = \mathbf{O}_P\left\{\sqrt{\frac{\log(n)}{nh}} + h^2\right\}$. Next under condition (C1), we have uniformly for $\mathbf{w} \in G$,

$$\left\{\frac{1}{nh}\sum_{i=1}^{n} k_i(\mathbf{w})\right\}^{-1} = f^{-1}(\mathbf{w})\left\{1 + \mathbf{O}_P\left(\sqrt{\frac{\log(n)}{nh}} + h^2\right)\right\}.$$

Define for $j = 1, 2$,

$$s_j(\mathbf{w}) = \frac{1}{nh}\sum_{i=1}^{n}\left(\frac{\mathbf{w}_i - \mathbf{w}}{h}\right) K\left(\frac{\mathbf{w}_i - \mathbf{w}}{h}\right).$$

Then, by the same method as in Yin et al. (2010), we have

$$s_1(\mathbf{w}) - h\dot{\mathbf{P}}_{l,\mathbf{w}}\omega_2 = \mathbf{O}_p\left(\sqrt{\frac{\log(n)}{nh}} + \mathbf{o}(h)\right) = \mathbf{o}_P(h), \tag{S2.4}$$

where $\omega_2 = \int_{-\infty}^{\infty} \mathbf{w}^2 K(\mathbf{w}) d\mathbf{w}$. Next, by Lemma 2 of Yao and Tong (1998) we know that

$$\sup_{\mathbf{w}\in\mathbb{R}} |s_2(\mathbf{w}) - f(\mathbf{w})\omega_2| = \mathbf{o}_P(1). \tag{S2.5}$$

By Taylor's expansion, we then have

$$
\begin{aligned}
&\widehat{\mathbf{P}}_{l,\mathbf{w}} - \mathbf{P}_{l,\mathbf{w}} \\
&= \frac{1}{nhf(\mathbf{w})}\left\{1 + \mathbf{O}_P\left(\sqrt{\frac{\log(n)}{nh}} + h^2\right)\right\}\sum_{i=1}^{n} K\left(\frac{\mathbf{w}_i - \mathbf{w}}{h}\right)\left\{\mathbb{1}(\widetilde{Y}_i = l) - \mathbf{P}_{l,\mathbf{w}_i}\right\} \\
&+ \frac{1}{f(\mathbf{w})}\left\{1 + \mathbf{O}_P\left(\sqrt{\frac{\log(n)}{nh}} + h^2\right)\right\}\left\{h\dot{\mathbf{P}}_{l,\mathbf{w}}s_1(\mathbf{w}) + \frac{h^2\ddot{\mathbf{P}}_{l,\mathbf{w}}}{2}s_2(\mathbf{w}) + \mathbf{o}(h^2)\right\}
\end{aligned}
\tag{S2.6}
$$

Then it follows by (S2.4) and (S2.5) that

$$
\widehat{\mathbf{P}}_{l,\mathbf{w}} - \mathbf{P}_{l,\mathbf{w}} = \mathbf{C}_{\mathbf{P},\mathbf{w}} + \mathbf{B}_{\mathbf{P},\mathbf{w}} + \mathbf{O}_P\{\mathbf{R}_1(\mathbf{w})\}. \tag{S2.7}
$$

where

$$
\mathbf{C}_{\mathbf{P},\mathbf{w}} = \frac{1}{nhf(\mathbf{w})}\sum_{i=1}^{n} k_i(\mathbf{w})\left\{\mathbb{1}(\widetilde{Y}_i = l) - \mathbf{P}_{l,\mathbf{w}_i}\right\}, \quad \mathbf{B}_{\mathbf{P},\mathbf{w}} = \frac{h^2\omega_2}{2}\left(2\frac{\dot{f}(\mathbf{w})}{f(\mathbf{w})}\dot{\mathbf{P}}_{l,\mathbf{w}} + \ddot{\mathbf{P}}_{l,\mathbf{w}}\right),
$$

and

$$
\mathbf{R}_1(\mathbf{w}) = \frac{1}{nf(\mathbf{w})}\left[\left|\sum_{i=1}^{n} K\left(\frac{\mathbf{w}_i - \mathbf{w}}{h}\right)\{\mathbb{1}(\widetilde{Y}_i = l) - \mathbf{P}_{l,\mathbf{w}_i}\}\right|\right] + \mathbf{o}(h^2).
$$

$\widehat{\mathbf{P}}_{l,\mathbf{w}} - \mathbf{P}_{l,\mathbf{w}}$ follows by the similar argument. Then,

$$
\widehat{\mathbf{U}}_{l,\mathbf{w}} - \mathbf{U}_{l,\mathbf{w}} = \mathbf{C}_{\mathbf{U},\mathbf{w}} + \mathbf{B}_{\mathbf{U},\mathbf{w}} + \mathbf{O}_P\{\mathbf{R}_2(\mathbf{w})\}, \tag{S2.8}
$$

where

$$\mathbf{C}_{\mathbf{U},\mathbf{w}} = \frac{1}{nhf(\mathbf{w})}\sum_{i=1}^{n} k_i(\mathbf{w})\left\{\mathbf{x}_i \mathbb{1}(\widetilde{Y}_i = l) - \mathbf{U}_{l,\mathbf{w}_i}\right\}, \quad \mathbf{B}_{\mathbf{U},\mathbf{w}} = \frac{h^2\omega_2}{2}\left(2\frac{\dot{f}(\mathbf{w})}{f(\mathbf{w})}\dot{\mathbf{U}}_{l,\mathbf{w}} + \ddot{\mathbf{U}}_{l,\mathbf{w}}\right),$$

and

$$\mathbf{R}_2(\mathbf{w}) = \frac{1}{nf(\mathbf{w})}\left\{\left|\sum_{i=1}^{n} K\left(\frac{\mathbf{w}_i - \mathbf{w}}{h}\right)\left(\mathbf{x}_i \mathbb{1}(\widetilde{Y}_i = l) - \mathbf{U}_{l,\mathbf{w}_i}\right)\right|\right\} + \mathbf{o}(h^2),$$

Yin et al. (2010) has already shown us that

$$\widehat{\mathbf{m}}(\mathbf{w}) - \mathbf{m}(\mathbf{w}) = \mathbf{C}_{\mathbf{m},\mathbf{w}} + \mathbf{B}_{\mathbf{m},\mathbf{w}} + \mathbf{O}_P\{\mathbf{R}_3(\mathbf{w})\}, \quad \widehat{\boldsymbol{\Sigma}}_{\mathbf{w}} - \boldsymbol{\Sigma}_{\mathbf{w}} = \mathbf{C}_{\boldsymbol{\Sigma},\mathbf{w}} + \mathbf{B}_{\boldsymbol{\Sigma},\mathbf{w}} + \mathbf{O}_P\{\mathbf{R}_4(\mathbf{w})\},$$

where

$$\mathbf{C}_{\mathbf{m},\mathbf{w}} = \frac{1}{nhf(\mathbf{w})}\sum_{i=1}^{n} k_i(\mathbf{w})\left\{\mathbf{x}_i - \mathbf{m}(\mathbf{w}_i)\right\}, \quad \mathbf{B}_{\mathbf{m},\mathbf{w}} = \frac{h^2\omega_2}{2}\left(2\frac{\dot{f}(\mathbf{w})}{f(\mathbf{w})}\dot{\mathbf{m}}(\mathbf{w}) + \ddot{\mathbf{m}}(\mathbf{w})\right),$$

$$\mathbf{C}_{\boldsymbol{\Sigma},\mathbf{w}} = \frac{1}{nhf(\mathbf{w})}\sum_{i=1}^{n} k_i(\mathbf{w})\left[\{\mathbf{x}_i - \widehat{\mathbf{m}}(\mathbf{w}_i)\}\{\mathbf{x}_i - \widehat{\mathbf{m}}(\mathbf{w}_i)\}^{\mathrm{T}} - \boldsymbol{\Sigma}_{\mathbf{w}} - (\mathbf{w}_i - \mathbf{w})\dot{\boldsymbol{\Sigma}}_{\mathbf{w}}\right],$$

$$\mathbf{B}_{\boldsymbol{\Sigma},\mathbf{w}} = h^2\dot{\boldsymbol{\Sigma}}_{\mathbf{w}}\omega_2\frac{\dot{f}(\mathbf{w})}{f(\mathbf{w})},$$

and

$$\mathbf{R}_3(\mathbf{w}) = \frac{1}{nf(\mathbf{w})}\left\{\left|\sum_{i=1}^{n} K\left(\frac{\mathbf{w}_i - \mathbf{w}}{h}\right)(\mathbf{x}_i - \mathbf{m}(\mathbf{w}_i))\right|\right\} + \mathbf{o}(h^2),$$
$$\mathbf{R}_4(\mathbf{w}) = \frac{1}{nf(\mathbf{w})}\left\{\left|\sum_{i=1}^{n} K\left(\frac{\mathbf{w}_i - \mathbf{w}}{h}\right)\left[\{\mathbf{x}_i - \widehat{\mathbf{m}}(\mathbf{w}_i)\}\{\mathbf{x}_i - \widehat{\mathbf{m}}(\mathbf{w}_i)\}^{\mathrm{T}} - \mathbf{\Sigma}_{\mathbf{w}} - (\mathbf{w}_i - \mathbf{w})\dot{\mathbf{\Sigma}}_{\mathbf{w}}\right]\right|\right\} + \mathbf{o}(h^2),$$

Thus

$$\begin{aligned}
&\widehat{\mathbf{M}}_{\mathrm{SIR}}(\mathbf{w}) - \mathbf{M}_{\mathrm{SIR}}(\mathbf{w}) \\
&= \sum_{l=1}^{H}\left(\frac{\widehat{\mathbf{U}}_{l,\mathbf{w}}\widehat{\mathbf{U}}_{l,\mathbf{w}}^{\mathrm{T}}}{\widehat{\mathbf{P}}_{l,\mathbf{w}}} - \sum_{l=1}^{H}\frac{\mathbf{U}_{l,\mathbf{w}}\mathbf{U}_{l,\mathbf{w}}^{\mathrm{T}}}{\mathbf{P}_{l,\mathbf{w}}}\right) + \mathbf{m}(\mathbf{w})\mathbf{m}(\mathbf{w})^{\mathrm{T}} - \widehat{\mathbf{m}}(\mathbf{w})\widehat{\mathbf{m}}(\mathbf{w})^{\mathrm{T}} \\
&= \sum_{l=1}^{H}\left\{\frac{(\widehat{\mathbf{U}}_{l,\mathbf{w}} - \mathbf{U}_{l,\mathbf{w}})\mathbf{U}_{l,\mathbf{w}}^{\mathrm{T}}}{\mathbf{P}_{l,\mathbf{w}}} + \frac{\mathbf{U}_{l,\mathbf{w}}(\widehat{\mathbf{U}}_{l,\mathbf{w}} - \mathbf{U}_{l,\mathbf{w}})^{\mathrm{T}}}{\mathbf{P}_{l,\mathbf{w}}} - \frac{1}{\mathbf{P}_{l,\mathbf{w}}^2}(\widehat{\mathbf{P}}_{l,\mathbf{w}} - \mathbf{P}_{l,\mathbf{w}})\mathbf{U}_{l,\mathbf{w}}\mathbf{U}_{l,\mathbf{w}}^{\mathrm{T}}\right\} \\
&- \{\widehat{\mathbf{m}}(\mathbf{w}) - \mathbf{m}(\mathbf{w})\}\mathbf{m}(\mathbf{w})^{\mathrm{T}} - \mathbf{m}(\mathbf{w})\{\widehat{\mathbf{m}}(\mathbf{w}) - \mathbf{m}(\mathbf{w})\}^{\mathrm{T}} + \mathbf{o}_P\left(\frac{1}{\sqrt{nh}}\right) \\
&= \mathbf{B}_{\mathrm{SIR}}(\mathbf{w}) + \mathbf{C}_{\mathrm{SIR}}(\mathbf{w}) + \mathbf{o}_P\left(\frac{1}{\sqrt{nh}}\right),
\end{aligned}$$

where

$$\mathbf{B}_{\mathrm{SIR}}(\mathbf{w}) = \sum_{l=1}^{H}\left\{\frac{\mathbf{B}_{\mathbf{U},\mathbf{w}}\mathbf{U}_{l,\mathbf{w}}^{\mathrm{T}} + \mathbf{U}_{l,\mathbf{w}}\mathbf{B}_{\mathbf{U},\mathbf{w}}^{\mathrm{T}}}{\mathbf{P}_{l,\mathbf{w}}} - \frac{1}{\mathbf{P}_{l,\mathbf{w}}^2}\mathbf{B}_{\mathbf{P},\mathbf{w}}\mathbf{U}_{l,\mathbf{w}}\mathbf{U}_{l,\mathbf{w}}^{\mathrm{T}}\right\} - \mathbf{B}_{\mathbf{m},\mathbf{w}}\mathbf{m}(\mathbf{w})^{\mathrm{T}} - \mathbf{m}(\mathbf{w})\mathbf{B}_{\mathbf{m},\mathbf{w}}^{\mathrm{T}}, \tag{S2.9}$$

and

$$\mathbf{C}_{\mathrm{SIR}}(\mathbf{w})=\sum_{l=1}^{H}\left\{\frac{\mathbf{C}_{\mathbf{U},\mathbf{w}}\mathbf{U}_{l,\mathbf{w}}^{\mathrm{T}}+\mathbf{U}_{l,\mathbf{w}}\mathbf{B}_{\mathbf{U},\mathbf{w}}^{\mathrm{T}}}{\mathbf{P}_{l,\mathbf{w}}}-\frac{1}{\mathbf{P}_{l,\mathbf{w}}^{2}}\mathbf{C}_{\mathbf{P},\mathbf{w}}\mathbf{U}_{l,\mathbf{w}}\mathbf{U}_{l,\mathbf{w}}^{\mathrm{T}}\right\}-\mathbf{C}_{\mathbf{m},\mathbf{w}}\mathbf{m}(\mathbf{w})^{\mathrm{T}}-\mathbf{m}(\mathbf{w})\mathbf{C}_{\mathbf{m},\mathbf{w}}^{\mathrm{T}}.$$

Following by the similar argument of Theorem 1 in Yin et al. (2010), we have

$$\sqrt{nh}\left(\mathbf{vech}\{\widehat{\mathbf{M}}_{\mathrm{SIR}}(\mathbf{w})\}-\mathbf{vech}\{M_{\mathrm{SIR}}(\mathbf{w})\}-\mathbf{vech}\{\mathbf{B}_{\mathrm{SIR}}(\mathbf{w})\}\right)\xrightarrow{d}\mathbf{N}(\mathbf{0},f^{-1}(\mathbf{w})\omega_0\mathbf{C}^{\mathrm{SIR}}(\mathbf{w})),\tag{S2.10}$$

where

$$\omega_0=\int_{-\infty}^{\infty}K^2(\mathbf{w})d\mathbf{w}.\tag{S2.11}$$

Hence

$$\mathbf{C}^{\mathrm{SIR}}(\mathbf{w})=\mathrm{Cov}\left[\mathbf{vech}\{\mathbf{C}_{\mathrm{SIR}}(\mathbf{w})\}|\mathbf{w}\right].\tag{S2.12}$$

This completes the proof of the part one of Theorem 1, then we prove the part two. Observe that $\lambda_k(\mathbf{w})$ and $\boldsymbol{\beta}_k(\mathbf{w})$ satisfy the following singular value decomposition equation:

$$\mathbf{G}(\mathbf{w})\mathbf{G}^{\mathrm{T}}(\mathbf{w})\boldsymbol{\beta}_k(\mathbf{w})=\lambda_k^2(\mathbf{w})\boldsymbol{\beta}_k(\mathbf{w}),\quad k=1,\ldots,p;$$

where $\mathbf{G}(\mathbf{w}) = \boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\mathbf{M}(\mathbf{w})$. Hence,

$$\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\mathbf{M}(\mathbf{w})\mathbf{M}^{\mathrm{T}}(\mathbf{w})\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\boldsymbol{\beta}_k(\mathbf{w}) = \lambda_k^2(\mathbf{w})\boldsymbol{\beta}_k(\mathbf{w}), \quad k = 1, \ldots, p;$$

where $\boldsymbol{\beta}_k^{\mathrm{T}}(\mathbf{w})\boldsymbol{\beta}_k(\mathbf{w}) = 1$ and $\boldsymbol{\beta}_k^{\mathrm{T}}(\mathbf{w})\boldsymbol{\beta}_\rho(\mathbf{w}) = 0$ for $k \neq \rho$. Similarly, in the sample level, we have

$$\widehat{\boldsymbol{\Sigma}}_{\mathbf{w}}^{-1}\widehat{\mathbf{M}}(\mathbf{w})\widehat{\mathbf{M}}^{\mathrm{T}}(\mathbf{w})\widehat{\boldsymbol{\Sigma}}_{\mathbf{w}}^{-1}\widehat{\boldsymbol{\beta}}_k(\mathbf{w}) = \widehat{\lambda}_k^2(\mathbf{w})\widehat{\boldsymbol{\beta}}_k(\mathbf{w}), \quad k = 1, \ldots, p;$$

and $\widehat{\boldsymbol{\beta}}_k^{\mathrm{T}}(\mathbf{w})\widehat{\boldsymbol{\beta}}_k(\mathbf{w}) = 1$ and $\widehat{\boldsymbol{\beta}}_k^{\mathrm{T}}(\mathbf{w})\widehat{\boldsymbol{\beta}}_\rho(\mathbf{w}) = 0$ for $k \neq \rho$. The singular value decomposition form in the sample level implies that

$$\begin{aligned}
&\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\mathbf{M}(\mathbf{w})\{\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\mathbf{M}(\mathbf{w})\}^{\mathrm{T}}\{\widehat{\boldsymbol{\beta}}_k(\mathbf{w}) - \boldsymbol{\beta}_k(\mathbf{w})\} + (\widehat{\boldsymbol{\Sigma}}_{\mathbf{w}}^{-1} - \boldsymbol{\Sigma}_{\mathbf{w}}^{-1})\mathbf{M}(\mathbf{w})\mathbf{M}^{\mathrm{T}}(\mathbf{w})\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\boldsymbol{\beta}_k(\mathbf{w}) \\
&+\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\{\widehat{\mathbf{M}}(\mathbf{w}) - \mathbf{M}(\mathbf{w})\}\mathbf{M}^{\mathrm{T}}(\mathbf{w})\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\boldsymbol{\beta}_k(\mathbf{w}) + \boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\mathbf{M}(\mathbf{w})\mathbf{M}^{\mathrm{T}}(\mathbf{w})(\widehat{\boldsymbol{\Sigma}}_{\mathbf{w}}^{-1} - \boldsymbol{\Sigma}_{\mathbf{w}}^{-1})\boldsymbol{\beta}_k(\mathbf{w}) \\
&+\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\mathbf{M}(\mathbf{w})\{\widehat{\mathbf{M}}(\mathbf{w}) - \mathbf{M}(\mathbf{w})\}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\boldsymbol{\beta}_k(\mathbf{w}) = \lambda_k(\mathbf{w})\{\widehat{\lambda}_k(\mathbf{w}) - \lambda_k(\mathbf{w})\}\boldsymbol{\beta}_k(\mathbf{w}) \\
&+\{\widehat{\lambda}_k(\mathbf{w}) - \lambda_k(\mathbf{w})\}\lambda_k(\mathbf{w})\boldsymbol{\beta}_k(\mathbf{w}) + \lambda_k^2(\mathbf{w})\{\widehat{\boldsymbol{\beta}}_k(\mathbf{w}) - \boldsymbol{\beta}_k(\mathbf{w})\} + \mathbf{o}_p(\frac{1}{\sqrt{nh}}),
\end{aligned} \tag{S2.13}$$

for $k = 1, \ldots, d$. Multiply both sides of (S2.13) by $\boldsymbol{\beta}_k^{\mathrm{T}}(\mathbf{w})$ from the left, we get

$$\boldsymbol{\beta}_k^{\mathrm{T}}(\mathbf{w})\Big[(\widehat{\boldsymbol{\Sigma}}_{\mathbf{w}}^{-1} - \boldsymbol{\Sigma}_{\mathbf{w}}^{-1})\mathbf{M}(\mathbf{w})\mathbf{M}^{\mathrm{T}}(\mathbf{w})\boldsymbol{\Sigma}_{\mathbf{w}}^{-1} + \boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\{\widehat{\mathbf{M}}(\mathbf{w}) - \mathbf{M}(\mathbf{w})\}\mathbf{M}^{\mathrm{T}}(\mathbf{w})\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}$$

$$+\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\mathbf{M}(\mathbf{w})\mathbf{M}^{\mathrm{T}}(\mathbf{w})(\widehat{\boldsymbol{\Sigma}}_{\mathbf{w}}^{-1}-\boldsymbol{\Sigma}_{\mathbf{w}}^{-1})+\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\mathbf{M}(\mathbf{w})\{\widehat{\mathbf{M}}(\mathbf{w})-\mathbf{M}(\mathbf{w})\}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\Big]\boldsymbol{\beta}_k(\mathbf{w})$$
$$=\{\widehat{\lambda}_k(\mathbf{w})-\lambda_k(\mathbf{w})\}\lambda_k(\mathbf{w})+\lambda_k\{\widehat{\lambda}_k(\mathbf{w})-\lambda_k(\mathbf{w})\}+\mathbf{o}_p(\frac{1}{\sqrt{nh}}),$$

which further suggests that

$$\begin{aligned}\widehat{\lambda}_k(\mathbf{w}) =&\lambda_k(\mathbf{w})+\frac{\boldsymbol{\beta}_k^{\mathrm{T}}(\mathbf{w})}{2\lambda_k(\mathbf{w})}\Bigg[(\widehat{\boldsymbol{\Sigma}}_{\mathbf{w}}^{-1}-\boldsymbol{\Sigma}_{\mathbf{w}}^{-1})\mathbf{M}(\mathbf{w})\mathbf{M}^{\mathrm{T}}(\mathbf{w})\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\\ &+\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\{\widehat{\mathbf{M}}(\mathbf{w})-\mathbf{M}(\mathbf{w})\}\mathbf{M}^{\mathrm{T}}(\mathbf{w})\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}+\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\mathbf{M}(\mathbf{w})\mathbf{M}^{\mathrm{T}}(\mathbf{w})(\widehat{\boldsymbol{\Sigma}}_{\mathbf{w}}^{-1}-\boldsymbol{\Sigma}_{\mathbf{w}}^{-1})\\ &+\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\mathbf{M}(\mathbf{w})\{\widehat{\mathbf{M}}(\mathbf{w})-\mathbf{M}(\mathbf{w})^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\}\Bigg]\boldsymbol{\beta}_k(\mathbf{w})+\mathbf{o}_p(\frac{1}{\sqrt{nh}}).\end{aligned} \tag{S2.14}$$

By lemma A.2 of Cook and Ni (2005) we know that

$$\widehat{\boldsymbol{\Sigma}}_{\mathbf{w}}^{-1}-\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}=-\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}(\widehat{\boldsymbol{\Sigma}}_{\mathbf{w}}-\boldsymbol{\Sigma}_{\mathbf{w}})\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}+\mathbf{o}_p(\frac{1}{\sqrt{nh}}).$$

Hence, equation (S2.14) becomes

$$\begin{aligned}\widehat{\lambda}_k(\mathbf{w}) =&\lambda_k(\mathbf{w})+\frac{\boldsymbol{\beta}_k^{\mathrm{T}}(\mathbf{w})}{2\lambda_k(\mathbf{w})}\Big[-\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}(\widehat{\boldsymbol{\Sigma}}_{\mathbf{w}}-\boldsymbol{\Sigma}_{\mathbf{w}})\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\mathbf{M}(\mathbf{w})\mathbf{M}^{\mathrm{T}}(\mathbf{w})\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\\ &+\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\{\widehat{\mathbf{M}}(\mathbf{w})-\mathbf{M}(\mathbf{w})\}\mathbf{M}^{\mathrm{T}}(\mathbf{w})\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}-\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\mathbf{M}(\mathbf{w})\mathbf{M}^{\mathrm{T}}(\mathbf{w})\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}(\widehat{\boldsymbol{\Sigma}}_{\mathbf{w}}-\boldsymbol{\Sigma}_{\mathbf{w}})\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\\ &+\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\mathbf{M}(\mathbf{w})\{\widehat{\mathbf{M}}(\mathbf{w})-\mathbf{M}(\mathbf{w})\}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\Big]\boldsymbol{\beta}_k(\mathbf{w})+\mathbf{o}_p(\frac{1}{\sqrt{nh}}),\\ =&\lambda_k(\mathbf{w})+\mathbf{C}_{\lambda_k}(\mathbf{w})+\mathbf{B}_{\lambda_k}(\mathbf{w})+\mathbf{o}_p(\frac{1}{\sqrt{nh}})\end{aligned} \tag{S2.15}$$

where

$$\mathbf{C}_{\lambda_k}(\mathbf{w}) = \frac{\boldsymbol{\beta}_k^{\mathrm{T}}(\mathbf{w})}{2\lambda_k(\mathbf{w})}\Big[- \boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\mathbf{C}_{\boldsymbol{\Sigma},\mathbf{w}}\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\mathbf{M}(\mathbf{w})\mathbf{M}^{\mathrm{T}}(\mathbf{w})\boldsymbol{\Sigma}_{\mathbf{w}}^{-1} + \boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\mathbf{C}_{\mathbf{M},\mathbf{w}}\mathbf{M}^{\mathrm{T}}(\mathbf{w})\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}$$
$$-\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\mathbf{M}(\mathbf{w})\mathbf{M}^{\mathrm{T}}(\mathbf{w})\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\mathbf{C}_{\boldsymbol{\Sigma},\mathbf{w}}\boldsymbol{\Sigma}_{\mathbf{w}}^{-1} + \boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\mathbf{M}(\mathbf{w})\mathbf{C}_{\mathbf{M},\mathbf{w}}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\Big]\boldsymbol{\beta}_k(\mathbf{w}), \tag{S2.16}$$

and

$$\mathbf{B}_{\lambda_k}(\mathbf{w}) = \frac{\boldsymbol{\beta}_k^{\mathrm{T}}(\mathbf{w})}{2\lambda_k(\mathbf{w})}\Big[- \boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\mathbf{B}_{\boldsymbol{\Sigma},\mathbf{w}}\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\mathbf{M}(\mathbf{w})\mathbf{M}^{\mathrm{T}}(\mathbf{w})\boldsymbol{\Sigma}_{\mathbf{w}}^{-1} + \boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\mathbf{B}_{\mathbf{M},\mathbf{w}}\mathbf{M}^{\mathrm{T}}(\mathbf{w})\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}$$
$$-\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\mathbf{M}(\mathbf{w})\mathbf{M}^{\mathrm{T}}(\mathbf{w})\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\mathbf{B}_{\boldsymbol{\Sigma},\mathbf{w}}\boldsymbol{\Sigma}_{\mathbf{w}}^{-1} + \boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\mathbf{M}(\mathbf{w})\mathbf{B}_{\mathbf{M},\mathbf{w}}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\Big]\boldsymbol{\beta}_k(\mathbf{w}). \tag{S2.17}$$

Now we turn to the expansion of $\widehat{\boldsymbol{\beta}}_k(\mathbf{w})$. Since $(\boldsymbol{\beta}_1(\mathbf{w}), \dots, \boldsymbol{\beta}_p(\mathbf{w}))$ is a basis of $\mathbb{R}^p$, then there exists c_{kj}^* for $j = 1, \dots, p$, such that $\widehat{\boldsymbol{\beta}}_k(\mathbf{w}) - \boldsymbol{\beta}_k(\mathbf{w}) = \sum_{j=1}^p c_{kj}^* \boldsymbol{\beta}_j(\mathbf{w})$ and $c_{kj}^* = \mathbf{O}_p(\frac{1}{\sqrt{nh}} + h^2)$. We will derive the explicit form of c_{kj}^* in the next step. Note that (S2.13) can be rewritten as

$$\begin{aligned}
&\left[\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\mathbf{M}(\mathbf{w})\mathbf{M}^{\mathrm{T}}(\mathbf{w})\boldsymbol{\Sigma}_{\mathbf{w}}^{-1} - \lambda_k^2(\mathbf{w})\right] \sum\nolimits_{j=1}^p c_{kj}^* \boldsymbol{\beta}_j(\mathbf{w}) \\
&= \lambda_k(\mathbf{w}) \left\{\widehat{\lambda}_k(\mathbf{w}) - \lambda_k(\mathbf{w})\right\} \boldsymbol{\beta}_k(\mathbf{w}) + \left\{\widehat{\lambda}_k(\mathbf{w}) - \lambda_k(\mathbf{w})\right\} \lambda_k(\mathbf{w})\boldsymbol{\beta}_k(\mathbf{w}) \\
&+ \Big[\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}(\widehat{\boldsymbol{\Sigma}}_{\mathbf{w}} - \boldsymbol{\Sigma}_{\mathbf{w}})\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\mathbf{M}(\mathbf{w})\mathbf{M}^{\mathrm{T}}(\mathbf{w})\boldsymbol{\Sigma}_{\mathbf{w}}^{-1} - \boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\{\widehat{\mathbf{M}}(\mathbf{w}) - \mathbf{M}(\mathbf{w})\}\mathbf{M}^{\mathrm{T}}(\mathbf{w})\boldsymbol{\Sigma}_{\mathbf{w}}^{-1} \\
&+ \boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\mathbf{M}(\mathbf{w})\mathbf{M}^{\mathrm{T}}(\mathbf{w})\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}(\widehat{\boldsymbol{\Sigma}}_{\mathbf{w}} - \boldsymbol{\Sigma}_{\mathbf{w}})\boldsymbol{\Sigma}_{\mathbf{w}}^{-1} - \boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\mathbf{M}(\mathbf{w})\{\widehat{\mathbf{M}}(\mathbf{w}) - \mathbf{M}(\mathbf{w})\}\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\Big]\boldsymbol{\beta}_k(\mathbf{w}).
\end{aligned} \tag{S2.18}$$

Denote

$$\mathbf{A}(\mathbf{w}) = \boldsymbol{\Sigma}_{\mathbf{w}}^{-1}(\widehat{\boldsymbol{\Sigma}}_{\mathbf{w}} - \boldsymbol{\Sigma}_{\mathbf{w}})\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\mathbf{M}(\mathbf{w})\mathbf{M}^{\mathrm{T}}(\mathbf{w})\boldsymbol{\Sigma}_{\mathbf{w}}^{-1} - \boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\{\widehat{\mathbf{M}}(\mathbf{w}) - \mathbf{M}(\mathbf{w})\}\mathbf{M}^{\mathrm{T}}(\mathbf{w})\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}$$
$$+ \boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\mathbf{M}(\mathbf{w})\mathbf{M}^{\mathrm{T}}(\mathbf{w})\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}(\widehat{\boldsymbol{\Sigma}}_{\mathbf{w}} - \boldsymbol{\Sigma}_{\mathbf{w}})\boldsymbol{\Sigma}_{\mathbf{w}}^{-1} - \boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\mathbf{M}(\mathbf{w})\{\widehat{\mathbf{M}}(\mathbf{w}) - \mathbf{M}(\mathbf{w})\}\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}.$$

Multiply both sides of (S2.18) by $\boldsymbol{\beta}_j^{\mathrm{T}}(\mathbf{w})$ $(j \neq k)$ from the left, we have

$$c_{kj}^* = \frac{\boldsymbol{\beta}_j^{\mathrm{T}}(\mathbf{w})\mathbf{A}(\mathbf{w})\boldsymbol{\beta}_k(\mathbf{w})}{\lambda_j^2(\mathbf{w}) - \lambda_k^2(\mathbf{w})}, j \neq k;$$

in addition, $\boldsymbol{\beta}_k^{\mathrm{T}}(\mathbf{w})\boldsymbol{\beta}_k(\mathbf{w}) = \widehat{\boldsymbol{\beta}}_k^{\mathrm{T}}(\mathbf{w})\widehat{\boldsymbol{\beta}}_k(\mathbf{w}) = 1$ indicates that

$$0 = \{\sum_{j=1}^{p} c_{kj}^*\boldsymbol{\beta}_j(\mathbf{w})\}^{\mathrm{T}}\boldsymbol{\beta}_k(\mathbf{w}) + \boldsymbol{\beta}_k(\mathbf{w})^{\mathrm{T}}\{\sum_{j=1}^{p} c_{kj}^*\boldsymbol{\beta}_j(\mathbf{w})\},$$

which further implies that $c_{kk}^* = 0$. Let

$$\mathbf{A}_1(\mathbf{w}) = \boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\mathbf{B}_{\boldsymbol{\Sigma},\mathbf{w}}\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\mathbf{M}(\mathbf{w})\mathbf{M}^{\mathrm{T}}(\mathbf{w})\boldsymbol{\Sigma}_{\mathbf{w}}^{-1} - \boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\mathbf{B}_{\mathbf{M},\mathbf{w}}\mathbf{M}^{\mathrm{T}}(\mathbf{w})\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}$$
$$+ \boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\mathbf{M}(\mathbf{w})\mathbf{M}^{\mathrm{T}}(\mathbf{w})\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\mathbf{B}_{\boldsymbol{\Sigma},\mathbf{w}}\boldsymbol{\Sigma}_{\mathbf{w}}^{-1} - \boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\mathbf{M}(\mathbf{w})\mathbf{B}_{\mathbf{M},\mathbf{w}}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{w}}^{-1},$$

and

$$\mathbf{A}_2(\mathbf{w}) = \boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\mathbf{C}_{\boldsymbol{\Sigma},\mathbf{w}}\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\mathbf{M}(\mathbf{w})\mathbf{M}^{\mathrm{T}}(\mathbf{w})\boldsymbol{\Sigma}_{\mathbf{w}}^{-1} - \boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\mathbf{C}_{\mathbf{M},\mathbf{w}}\mathbf{M}^{\mathrm{T}}(\mathbf{w})\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}$$

$$+ \boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\mathbf{M}(\mathbf{w})\mathbf{M}^{\mathrm{T}}(\mathbf{w})\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\mathbf{C}_{\boldsymbol{\Sigma},\mathbf{w}}\boldsymbol{\Sigma}_{\mathbf{w}}^{-1} - \boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\mathbf{M}(\mathbf{w})\mathbf{C}_{\mathbf{M},\mathbf{w}}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}.$$

Hence, we have

$$\widehat{\boldsymbol{\beta}}_k(\mathbf{w}) = \boldsymbol{\beta}_k(\mathbf{w}) + \mathbf{B}_k(\mathbf{w}) + \mathbf{C}_k(\mathbf{w}) + \mathbf{o}_P(\frac{1}{\sqrt{nh}}). \tag{S2.19}$$

Denote

$$\mathbf{B}_k(\mathbf{w}) = \sum_{j\neq k} \frac{\boldsymbol{\beta}_j(\mathbf{w})\boldsymbol{\beta}_j^{\mathrm{T}}(\mathbf{w})\mathbf{A}_1(\mathbf{w})\boldsymbol{\beta}_k(\mathbf{w})}{\lambda_j^2(\mathbf{w}) - \lambda_k^2(\mathbf{w})}, \tag{S2.20}$$

and

$$\mathbf{C}_k(\mathbf{w}) = \sum_{j\neq k} \frac{\boldsymbol{\beta}_j(\mathbf{w})\boldsymbol{\beta}_j^{\mathrm{T}}(\mathbf{w})\mathbf{A}_2(\mathbf{w})\boldsymbol{\beta}_k(\mathbf{w})}{\lambda_j^2(\mathbf{w}) - \lambda_k^2(\mathbf{w})}. \tag{S2.21}$$

The asymptotic normality is then straightforward via the central limit theorem and

$$\boldsymbol{\Sigma}_k(\mathbf{w}) = \mathrm{Cov}\{\mathbf{C}_k(\mathbf{w})|\mathbf{w}\}.$$

In partial variable dependent SIR, denote $\mathbf{G}(\mathbf{w}) = \boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\mathbf{M}_{\mathrm{SIR}}(\mathbf{w})$, then substitute it into (S2.20) and (S2.21),

$$\mathbf{B}_k^{\mathrm{SIR}}(\mathbf{w}) = \sum_{j\neq k} \frac{\boldsymbol{\beta}_j^{\mathrm{SIR}}(\mathbf{w})\{\boldsymbol{\beta}_j^{\mathrm{SIR}}(\mathbf{w})\}^{\mathrm{T}}}{\{\lambda_j^{\mathrm{SIR}}(\mathbf{w})\}^2 - \{\lambda_k^{\mathrm{SIR}}(\mathbf{w})\}^2} \Big[\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\mathbf{B}_{\boldsymbol{\Sigma},\mathbf{w}}\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\mathbf{M}_{\mathrm{SIR}}(\mathbf{w})\mathbf{M}_{\mathrm{SIR}}^{\mathrm{T}}(\mathbf{w})\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}$$
$$- \boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\mathbf{B}_{\mathrm{SIR}}(\mathbf{w})\mathbf{M}_{\mathrm{SIR}}^{\mathrm{T}}(\mathbf{w})\boldsymbol{\Sigma}_{\mathbf{w}}^{-1} + \boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\mathbf{M}_{\mathrm{SIR}}(\mathbf{w})\mathbf{M}_{\mathrm{SIR}}^{\mathrm{T}}(\mathbf{w})\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\mathbf{B}_{\boldsymbol{\Sigma},\mathbf{w}}\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}$$
$$- \boldsymbol{\Sigma}_{\mathbf{w}}^{-1}M_{\mathrm{SIR}}(\mathbf{w})\mathbf{B}_{\mathrm{SIR}}^{\mathrm{T}}(\mathbf{w})\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\Big]\boldsymbol{\beta}_k^{\mathrm{SIR}}(\mathbf{w}), \tag{S2.22}$$

$$\mathbf{C}_k^{\mathrm{SIR}}(\mathbf{w}) = \sum_{j\neq k} \frac{\boldsymbol{\beta}_j^{\mathrm{SIR}}(\mathbf{w})\{\boldsymbol{\beta}_j^{\mathrm{SIR}}(\mathbf{w})\}^{\mathrm{T}}}{\{\lambda_j^{\mathrm{SIR}}(\mathbf{w})\}^2 - \{\lambda_k^{\mathrm{SIR}}(\mathbf{w})\}^2} \Big[\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\mathbf{C}_{\boldsymbol{\Sigma},\mathbf{w}}\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\mathbf{M}_{\mathrm{SIR}}(\mathbf{w})\mathbf{M}_{\mathrm{SIR}}^{\mathrm{T}}(\mathbf{w})\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}$$
$$- \boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\mathbf{C}_{\mathrm{SIR}}(\mathbf{w})\mathbf{M}_{\mathrm{SIR}}^{\mathrm{T}}(\mathbf{w})\boldsymbol{\Sigma}_{\mathbf{w}}^{-1} + \boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\mathbf{M}_{\mathrm{SIR}}(\mathbf{w})\mathbf{M}_{\mathrm{SIR}}^{\mathrm{T}}(\mathbf{w})\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\mathbf{C}_{\boldsymbol{\Sigma},\mathbf{w}}\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}$$
$$- \boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\mathbf{M}_{\mathrm{SIR}}(\mathbf{w})\mathbf{C}_{\mathrm{SIR}}^{\mathrm{T}}(\mathbf{w})\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\Big]\boldsymbol{\beta}_k^{\mathrm{SIR}}(\mathbf{w}).$$

Hence,

$$\boldsymbol{\Sigma}_k^{\mathrm{SIR}}(\mathbf{w}) = \mathrm{Cov}\{\mathbf{C}_k^{\mathrm{SIR}}(\mathbf{w})|\mathbf{w}\} \tag{S2.23}$$

□

Theorem 2 is closely related to the following two assertions:

(a) if $\lambda_k(\mathbf{w}) > \lambda_{k+1}(\mathbf{w})$, then $f_n^0(k) = \mathbf{O}_P(\frac{1}{nh})$ almost surely $\mathbf{P}_{\mathcal{S}}$ given $\mathbf{W} = \mathbf{w}$;

(b) if $\lambda_k(\mathbf{w}) = \lambda_{k+1}(\mathbf{w})$, then $f_n^0(k) = \mathbf{O}_P^{+}(c_n)$ almost surely $\mathbf{P}_{\mathcal{S}}$ given

$$\mathbf{W} = \mathbf{w},$$

where $c_n = [\log\{\log(n)\}]^{-1}$ and $\mathbf{O}_P^+$ symbols has been shown in Luo and Li (2016). We will prove these assertions, and then prove Theorem 2 based on them.

The nonparametric ladle estimator can pinpoint the rank of a matrix more precisely than the other order-determination methods when they are used for the nonparametric model. We established the consistency of the nonparametric ladle estimator. The next Lemma regulates the order between the eigenvalues and the variability of eigenvectors. In particular, it shows that distant eigenvalues are related to the small variability of eigenvectors. Ladle estimator allows asymmetric matrices, therefore we apply it to $\widehat{\mathbf{G}}(\mathbf{w})\widehat{\mathbf{G}}^{\mathrm{T}}(\mathbf{w})$, which amounts to replacing the eigenvalues and eigenvectors of $\widehat{\mathbf{G}}(\mathbf{w})$ by its squared singular values and singular vectors, since $\mathbf{G}(\mathbf{w})$ is not a symmetric and semi-positive definite matrix.

Lemma 1. *Let $c_n = \{\log(\log n)\}^{-1}$. If conditions (C9), (C10), (C11) and (C12) hold, and $\mathbf{G}(\mathbf{w})\mathbf{G}(\mathbf{w})^{\tau} \in \mathbb{R}^{p\times p}$ is a positive semi-definite matrix of rank $d(\mathbf{w}) \in \{0, \ldots, p-1\}$, then for any $k = 1, \ldots, p-1$, the following rela-*

tion holds for almost every sequence $S = \{(Y_1, \mathbf{X}_1, \mathbf{W}_1), (Y_2, \mathbf{X}_2, \mathbf{W}_2), \ldots\}$:

$$f_n(\mathbf{w}, k) = \begin{cases} \mathbf{O}_P(\frac{1}{nh}), & \lambda_k(\mathbf{w}) > \lambda_{k+1}(\mathbf{w}); \\ \mathbf{O}_P^+(c_n), & \lambda_k(\mathbf{w}) = \lambda_{k+1}(\mathbf{w}); \end{cases}$$

where $\mathbf{O}_P^+$ *is defined in Luo and Li (2016).*

Lemma 2. *Let* $c_n = \{\log(\log n)\}^{-1}$ *and* $r = \lfloor p/\log(p) \rfloor$. *For any positive semi-definite candidate matrix* $\mathbf{G}(\mathbf{w})\mathbf{G}(\mathbf{w})^T \in \mathbb{R}^{p\times p}$ *of rank* $d(\mathbf{w})$, *and for each* $k \in \{0, 1, \ldots, r\}$, *we have*

$$\phi_n(\mathbf{w}, k) = \begin{cases} \mathbf{O}_P^+(1), & \text{if} \quad k < d(\mathbf{w}); \\ \mathbf{O}_P(\frac{1}{\sqrt{c_n nh}}), & \text{if} \quad k \geq d(\mathbf{w}); \end{cases} \quad \text{almost surely } \mathbf{P}_{\mathcal{S}}.$$

The asymptotic behavior of $f_n(\mathbf{w}, \cdot)$ is presented in Lemma 1 and Lemma 2 gives the asymptotic property of $\phi_n(\mathbf{w}, \cdot)$. The proof of Lemma 1 and 2 are similar to that of Theorem 1 and Lemma F in Luo and Li (2016), thus omitted here.

Lemma 3. *For any* $i, j = 1, \ldots, p$ *with* $i \neq j$,

$$|\boldsymbol{\beta}_i^T(\mathbf{w})\widehat{\boldsymbol{\beta}}_j(\mathbf{w})|[\lambda_i(\mathbf{w}) - \lambda_j(\mathbf{w}) - \{\mathbf{B}_{\lambda_j}(\mathbf{w}) + \mathbf{B}(\mathbf{w})\}] = \mathbf{O}_p(\frac{1}{\sqrt{nh}}),$$

where the closed form of $\mathbf{B}_{\lambda_j}(\mathbf{w})$ *and* $\mathbf{B}(\mathbf{w})$ *are provided in* (S2.17) *and* (S2.26) *in the Appendix, respectively.*

Proof of Lemma 3. Luo and Li (2016) has shown that

$$\mathbf{G}(\mathbf{w})\widehat{\boldsymbol{\beta}}_j(\mathbf{w}) = \mathbf{G}(\mathbf{w})\left\{\sum_{i=1}^{p} r_{ij}\boldsymbol{\beta}_i(\mathbf{w})\right\} = \sum_{i=1}^{p} r_{ij}\lambda_i(\mathbf{w})\boldsymbol{\beta}_i(\mathbf{w}), \tag{S2.24}$$

and

$$\mathbf{G}(\mathbf{w})\widehat{\boldsymbol{\beta}}_j(\mathbf{w}) = \widehat{\mathbf{G}}(\mathbf{w})\widehat{\boldsymbol{\beta}}_j(\mathbf{w}) + \{\mathbf{G}(\mathbf{w}) - \widehat{\mathbf{G}}(\mathbf{w})\}\widehat{\boldsymbol{\beta}}_j(\mathbf{w}) = \widehat{\lambda}_j(\mathbf{w})\widehat{\boldsymbol{\beta}}_j(\mathbf{w}) + \{\mathbf{G}(\mathbf{w}) - \widehat{\mathbf{G}}(\mathbf{w})\}\widehat{\boldsymbol{\beta}}_j(\mathbf{w}), \tag{S2.25}$$

where $r_{ij} = \boldsymbol{\beta}_i^{\mathrm{T}}(\mathbf{w})\widehat{\boldsymbol{\beta}}_j(\mathbf{w})$, $\widehat{\boldsymbol{\beta}}_j(\mathbf{w}) = \sum_{i=1}^{p} r_{ij}\boldsymbol{\beta}_i(\mathbf{w})$ and $\sum_{i=1}^{p} r_{ij}^2 = 1$. $\widehat{\mathbf{G}}(\mathbf{w}) - \mathbf{G}(\mathbf{w}) = \mathbf{B}(\mathbf{w}) + \mathbf{O}_P\left(\frac{1}{\sqrt{nh}}\right)$ and $\widehat{\lambda}_j(\mathbf{w}) - \lambda_j(\mathbf{w}) = \mathbf{B}_{\lambda_j}(\mathbf{w}) + \mathbf{O}_p(\frac{1}{\sqrt{nh}})$. Here, we denote $\mathbf{G}(\mathbf{w}) = \boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\mathbf{M}(\mathbf{w})$. Since

$$\widehat{\mathbf{G}}(\mathbf{w}) - \mathbf{G}(\mathbf{w}) = \boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\{\widehat{\mathbf{M}}(\mathbf{w}) - \mathbf{M}(\mathbf{w})\} + (\widehat{\boldsymbol{\Sigma}}_{\mathbf{w}}^{-1} - \boldsymbol{\Sigma}_{\mathbf{w}}^{-1})\mathbf{M}(\mathbf{w}) + \mathbf{o}_p(\frac{1}{\sqrt{nh}}).$$

Hence

$$\mathbf{B}(\mathbf{w}) = \boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\mathbf{B}_{M,\mathbf{w}} + \boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\mathbf{B}_{\boldsymbol{\Sigma},\mathbf{w}}\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\mathbf{M}(\mathbf{w}). \tag{S2.26}$$

Combing (S2.24) and (S2.25), we have

$$\begin{aligned}\sum_{i=1}^{p} r_{ij}\lambda_i(\mathbf{w})\boldsymbol{\beta}_i(\mathbf{w}) =& \lambda_j(\mathbf{w})\widehat{\boldsymbol{\beta}}_j(\mathbf{w}) + \mathbf{B}_{\lambda_j}(\mathbf{w})\widehat{\boldsymbol{\beta}}_j(\mathbf{w}) + \mathbf{B}(\mathbf{w})\widehat{\boldsymbol{\beta}}_j(\mathbf{w}) + \mathbf{O}_p(\frac{1}{\sqrt{nh}})\\ =& \sum_{i=1}^{p} r_{ij}\lambda_j(\mathbf{w})\boldsymbol{\beta}_i(\mathbf{w}) + \{\mathbf{B}_{\lambda_j}(\mathbf{w}) + \mathbf{B}(\mathbf{w})\}\widehat{\boldsymbol{\beta}}_j(\mathbf{w}) + \mathbf{O}_p(\frac{1}{\sqrt{nh}})\\ =& \sum_{i=1}^{p} r_{ij}\lambda_j(\mathbf{w})\boldsymbol{\beta}_i(\mathbf{w}) + \sum_{i=1}^{p} r_{ij}\{\mathbf{B}_{\lambda_j}(\mathbf{w}) + \mathbf{B}(\mathbf{w})\}\boldsymbol{\beta}_i(\mathbf{w}) + \mathbf{O}_p(\frac{1}{\sqrt{nh}}).\end{aligned}$$

Hence

$$\sum_{i=1}^{p} r_{ij}[\lambda_i(\mathbf{w}) - \lambda_j(\mathbf{w}) - \{\mathbf{B}_{\lambda_j}(\mathbf{w}) + \mathbf{B}(\mathbf{w})\}]\boldsymbol{\beta}_i(\mathbf{w}) = \mathbf{O}_p(\frac{1}{\sqrt{nh}}).$$

Since $\boldsymbol{\beta}_1(\mathbf{w}), \ldots, \boldsymbol{\beta}_p(\mathbf{w})$ are orthogonal, we have, for each $i \neq j$,

$$|\boldsymbol{\beta}_i^{\mathrm{T}}(\mathbf{w})\widehat{\boldsymbol{\beta}}_j(\mathbf{w})|[\lambda_i(\mathbf{w}) - \lambda_j(\mathbf{w}) - \{\mathbf{B}_{\lambda_j}(\mathbf{w}) + \mathbf{B}(\mathbf{w})\}] = \mathbf{O}_p(\frac{1}{\sqrt{nh}}).$$

$\square$

Since the order of bias term is h^2, then under condition (C7),

$$\frac{\sqrt{nh} \times h^2}{\sqrt{\log(\log n)}} \to 0, \text{ as } n \to \infty,$$

the bias term $\{\mathbf{B}_{\lambda_j}(\mathbf{w}) + \mathbf{B}(\mathbf{w})\}$ will vanish, such that a direct application of this Lemma leads to the next lemma.

Lemma 4. *Under condition (C9) and (C13), for any positive semi-definite candidate matrix* $\mathbf{G}(\mathbf{w}) \in \mathbb{R}^{p\times p}$ *and any* $i, j \in \{1, \ldots, p\}$*, if* $\lambda_i(\mathbf{w}) > \lambda_j(\mathbf{w})$*,then*

$$\widehat{\boldsymbol{\beta}}_i(\mathbf{w})^T \boldsymbol{\beta}_j^*(\mathbf{w}) = \mathbf{O}_P(\frac{1}{\sqrt{nh}}),$$

almost surely $\mathbf{P}_{\mathcal{S}}$.

Proof of Lemma 4. Let $A_1 \in \mathcal{F}$ be the event that $\Big\{(nh)^{1/2}\{\widehat{\lambda}_i(\mathbf{w}) - \lambda_i(\mathbf{w})\}/\sqrt{\log(\log n)} : n \in \mathbb{N}\Big\}$ is a bounded sequence for each $i = 1, \ldots, p$. From Assumption (C12) it follows Hardle (1984) that the bias term of $\widehat{\lambda}_i(\mathbf{w}) - \lambda_i(\mathbf{w})$ vanishes. By the law of the iterated logarithm and Lemma 3.2 of Zhao et al. (1986), $\mathrm{pr}(A_1) = 1$. For any $s \in A_1$ and $i, j = 1, \ldots, p$, we have

$$|\widehat{\lambda}_i(\mathbf{w}) - \widehat{\lambda}_j(\mathbf{w})| = |\lambda_i(\mathbf{w}) - \lambda_j(\mathbf{w})| + \mathbf{o}(1). \tag{S2.27}$$

Let $A_2 \in \mathcal{F}$ be the event in Assumption (C10). Then $\mathrm{pr}(A_2) = 1$. Hence $\mathrm{pr}(A_1 \cap A_2) = 1$. For any fixed $s \in A_1 \cap A_2$, by Lemma 3,

$$\widehat{\boldsymbol{\beta}}_i^{\mathrm{T}}(\mathbf{w})\boldsymbol{\beta}_j^*(\mathbf{w})\{\widehat{\lambda}_i(\mathbf{w}) - \widehat{\lambda}_j(\mathbf{w})\} = \mathbf{O}_P(\frac{1}{\sqrt{nh}}).$$

By (S2.27), we have $\widehat{\boldsymbol{\beta}}_i^{\mathrm{T}}(\mathbf{w})\boldsymbol{\beta}_j^*(\mathbf{w}) = \mathbf{O}_P(\frac{1}{\sqrt{nh}})$. □

Proof of Assertion (a). By the law of iterated logarithm, similar to the proof of Assertion (a) of Luo and Li (2016), we can show that, when $\lambda_k(\mathbf{w}) > \lambda_{k+1}(\mathbf{w})$,

$$1 - |\det\{\mathbf{G}_{11}(\mathbf{w})\}| = \mathbf{O}_P(\frac{1}{nh}) \text{ almost surely } \mathbf{P}_{\mathcal{S}},$$

where $\mathbf{G}_{11}(\mathbf{w}) = \widehat{\mathbf{T}}_k^{\mathrm{T}}\mathbf{T}_k^*$. Thus $f_n^0(\mathbf{w}, k) = \mathbf{O}_P(\frac{1}{nh})$ almost surely $\mathbf{P}_{\mathcal{S}}$, as desired. □

Proof of Assertion (b) is omitted here since it's similar in Luo and Li (2016). We now prove Theorem 2 by combining lemma 1, lemma 2, lemma 3, lemma 4 and Assertion (a) and (b).

Proof of Theorem 2. It's easy to see that

$$\mathbf{O}_P(\frac{1}{nh}) = \mathbf{o}_P(c_n), \quad \mathbf{O}_P^+(1) = \mathbf{O}_P^+(c_n), \quad \mathbf{O}_P(\frac{1}{\sqrt{c_n nh}}) = \mathbf{o}_P(c_n).$$

Let $r = p - 1$ if $p \leq 10$ and $r = [p/\log(p)]$ otherwise. By assertion (a), assertion (b) and Lemma 1 and 2, for any $k \in \{0, 1, \ldots, r\}$,

$$\begin{cases} f_n(\mathbf{w}, k) \geq 0, & \phi_n(\mathbf{w}, k) = \mathbf{O}_P^+(c_n), \quad \text{if} \quad k < d(\mathbf{w}); \\ f_n(\mathbf{w}, k) = \mathbf{o}_P(c_n), & \phi_n(\mathbf{w}, k) = \mathbf{o}_P(c_n), \quad \text{if} \quad k = d(\mathbf{w}); \\ f_n(\mathbf{w}, k) = \mathbf{O}_P^+(c_n), & \phi_n(\mathbf{w}, k) > 0, \quad \text{if} \quad k > d(\mathbf{w}); \end{cases}$$

almost surely $\mathbf{P}_{\mathcal{S}}$. Since $g_n(\mathbf{w}) = f_n(\mathbf{w}) + \phi_n(\mathbf{w})$, lemma D (i) of Luo and Li (2016) implies that

$$g_n(\mathbf{w}, k) = \begin{cases} \mathbf{O}_P^+(c_n), & \text{if} \quad k \neq d(\mathbf{w}); \\ \mathbf{o}_P(c_n), & \text{if} \quad k = d(\mathbf{w}); \end{cases} \quad \text{almost surely } \mathbf{P}_{\mathcal{S}}.$$

By Lemma D (ii) of Luo and Li (2016), g_n is minimized at $d(\mathbf{w})$ in probability almost surely $\mathbf{P}_{\mathcal{S}}$. □

S3. Variable Dependent Partial SAVE

Though SIR has received much attention, it cannot recover any vector in the central subspace $\mathcal{S}_{Y|\mathbf{X}}$ if the regression function is symmetric about the origin because SIR is based on the estimation of the conditional mean. To address this, SAVE (Cook and Weisberg, 1991) was proposed to estimate the central space by utilizing the conditional variance function of the covariates when the response is given. For partial dimension reduction, Shao et al. (2009) also developed partial SAVE and showed that partial SAVE is more comprehensive than partial SIR. As an extension of partial SAVE, we now develop the variable dependent partial SAVE, which is based on the proposition S3.1. Parallel to SAVE, we define the following kernel matrix for variable dependent partial SAVE.

$$\begin{aligned}\mathbf{M}_{\mathrm{SAVE}}(\mathbf{w}) &\triangleq \mathrm{E}_{\widetilde{Y}}\{\mathrm{Cov}(\mathbf{X}_{\mathbf{w}}) - \mathrm{Cov}(\mathbf{X}|\widetilde{Y} = l, \mathbf{w})\}^2 \\ &= \sum\nolimits_{l=1}^{H} \mathbf{P}_{l,\mathbf{w}} \left(\boldsymbol{\Sigma}_{\mathbf{w}} - \mathbf{R}_{l,\mathbf{w}} + \mathbf{V}_{l,\mathbf{w}}\mathbf{V}_{l,\mathbf{w}}^{\mathrm{T}}\right)^2,\end{aligned}$$

where $\mathbf{R}_{l,\mathbf{w}} = \mathrm{E}(\mathbf{X}\mathbf{X}^{\mathrm{T}}|\widetilde{Y} = l, \mathbf{w})$ and $\mathbf{V}_{l,\mathbf{w}}$, $\mathbf{P}_{l,\mathbf{w}}$ are defined as previously. $\mathrm{E}_{\widetilde{Y}}$ represents expectation with respect to $\widetilde{Y}$.

Proposition S3.1. *Conditional on $\mathbf{W} = \mathbf{w}$, suppose that linear conditional mean condition (A1) and constant conditional variance condition (A2) hold, then*

$$\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\{\boldsymbol{\Sigma}_{\mathbf{w}} - \mathrm{Cov}(\mathbf{X}|\mathrm{Y}, \mathbf{W} = \mathbf{w})\} = \mathbf{P}_{\mathbf{B}(\mathbf{w})}\{\mathbf{I}_{\mathrm{p}} - \boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\mathrm{Cov}(\mathbf{X}|\mathrm{Y}, \mathbf{W} = \mathbf{w})\}\mathbf{P}_{\mathbf{B}(\mathbf{w})}.$$

PROOF OF PROPOSITION S3.1. The proof of Proposition S3.1 is based on the Theorem 1 of Cook and Lee (1999). Use the similar abbreviation in Proposition 1,

$$\mathrm{Cov}(\mathbf{X}|\mathrm{Y}, \mathbf{W} = \mathbf{w}) = \mathrm{Cov}(\mathbf{X}_{\mathbf{w}}|\mathrm{Y}_{\mathbf{w}}).$$

Assume that $\mathbf{X}_{\mathbf{w}}$ satisfies linear conditional mean (A1) and constant conditional variance (A2), furthermore, constant conditional variance (A2) implies $\mathrm{Cov}\{\mathbf{X}_{\mathbf{w}}|\mathbf{B}^{\mathrm{T}}(\mathbf{w})\mathbf{X}_{\mathbf{w}}\} = \mathbf{\Sigma}_{\mathbf{w}}\{\mathbf{I}_{\mathrm{p}} - \mathbf{P}_{\mathbf{B}(\mathbf{w})}\}$. Conditional on $\mathbf{W} = \mathbf{w}$, it's easy for us to get

$$\begin{aligned}
&\mathbf{\Sigma}_{\mathbf{w}}^{-1}\mathrm{Cov}(\mathbf{X}|\mathrm{Y}, \mathbf{W} = \mathbf{w}) = \mathbf{\Sigma}_{\mathbf{w}}^{-1}\mathrm{Cov}(\mathbf{X}_{\mathbf{w}}|\mathrm{Y}_{\mathbf{w}}) \\
&= \mathbf{\Sigma}_{\mathbf{w}}^{-1}\mathrm{E}[\mathrm{Cov}\{\mathbf{X}_{\mathbf{w}}|\mathbf{B}^{\mathrm{T}}(\mathbf{w})\mathbf{X}_{\mathbf{w}}, \mathrm{Y}_{\mathbf{w}}\}|\mathrm{Y}_{\mathbf{w}}] + \mathbf{\Sigma}_{\mathbf{w}}^{-1}\mathrm{Cov}[\mathrm{E}\{\mathbf{X}_{\mathbf{w}}|\mathbf{B}^{\mathrm{T}}(\mathbf{w})\mathbf{X}_{\mathbf{w}}, \mathrm{Y}_{\mathbf{w}}\}|\mathrm{Y}_{\mathbf{w}}] \\
&= \mathbf{\Sigma}_{\mathbf{w}}^{-1}\mathrm{E}[\mathrm{Cov}\{\mathbf{P}_{\mathbf{B}(\mathbf{w})}^{\mathrm{T}}\mathbf{X}_{\mathbf{w}} + \mathbf{Q}_{\mathbf{B}(\mathbf{w})}^{\mathrm{T}}\mathbf{X}_{\mathbf{w}}|\mathbf{B}^{\mathrm{T}}(\mathbf{w})\mathbf{X}_{\mathbf{w}}\}|\mathrm{Y}_{\mathbf{w}}] + \mathbf{\Sigma}_{\mathbf{w}}^{-1}\mathrm{Cov}[\mathrm{E}\{\mathbf{X}_{\mathbf{w}}|\mathbf{B}^{\mathrm{T}}(\mathbf{w})\mathbf{X}_{\mathbf{w}}\}|\mathrm{Y}_{\mathbf{w}}] \\
&= \mathbf{\Sigma}_{\mathbf{w}}^{-1}\mathbf{Q}_{\mathbf{B}(\mathbf{w})}^{\mathrm{T}}\mathrm{E}[\mathrm{Cov}\{\mathbf{X}_{\mathbf{w}}|\mathbf{B}^{\mathrm{T}}(\mathbf{w})\mathbf{X}_{\mathbf{w}}\}|\mathrm{Y}_{\mathbf{w}}]\mathbf{Q}_{\mathbf{B}(\mathbf{w})} + \mathbf{\Sigma}_{\mathbf{w}}^{-1}\mathbf{P}_{\mathbf{B}(\mathbf{w})}^{\mathrm{T}}\mathrm{Cov}(\mathbf{X}_{\mathbf{w}}|\mathrm{Y}_{\mathbf{w}})\mathbf{P}_{\mathbf{B}(\mathbf{w})} \\
&= \mathbf{Q}_{\mathbf{B}(\mathbf{w})}\mathbf{\Sigma}_{\mathbf{w}}^{-1}\mathrm{E}[\mathrm{Cov}\{\mathbf{X}_{\mathbf{w}}|\mathbf{B}^{\mathrm{T}}(\mathbf{w})\mathbf{X}_{\mathbf{w}}\}|\mathrm{Y}_{\mathbf{w}}]\mathbf{Q}_{\mathbf{B}(\mathbf{w})} + \mathbf{P}_{\mathbf{B}(\mathbf{w})}\mathbf{\Sigma}_{\mathbf{w}}^{-1}\mathrm{Cov}(\mathbf{X}_{\mathbf{w}}|\mathrm{Y}_{\mathbf{w}})\mathbf{P}_{\mathbf{B}(\mathbf{w})} \\
&= \mathbf{Q}_{\mathbf{B}(\mathbf{w})} + \mathbf{P}_{\mathbf{B}(\mathbf{w})}\mathbf{\Sigma}_{\mathbf{w}}^{-1}\mathrm{Cov}(\mathbf{X}_{\mathbf{w}}|\mathrm{Y}_{\mathbf{w}})\mathbf{P}_{\mathbf{B}(\mathbf{w})},
\end{aligned}$$

where $\mathbf{Q}_{\mathbf{B}(\mathbf{w})} = \mathbf{I}_p - \mathbf{P}_{\mathbf{B}(\mathbf{w})}$. Such that we can derive that

$$\mathbf{I}_p - \mathbf{\Sigma}_{\mathbf{w}}^{-1}\mathrm{Cov}(\mathbf{X}|\mathrm{Y}, \mathbf{W} = \mathbf{w}) = \mathbf{P}_{\mathbf{B}(\mathbf{w})}\{\mathbf{I}_{\mathrm{p}} - \mathbf{\Sigma}_{\mathbf{w}}^{-1}\mathrm{Cov}(\mathbf{X}|\mathrm{Y}, \mathbf{W} = \mathbf{w})\}\mathbf{P}_{\mathbf{B}(\mathbf{w})}.$$

The proof is completed. $\square$

Proposition S3.1 implies that $\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\mathbf{M}_{\text{SAVE}}(\mathbf{w}) \subseteq \mathcal{S}_{Y_{\mathbf{w}}|\mathbf{X}_{\mathbf{w}}}$ almost surely. Similar to Proposition 2, the proposition S3.2 shows that the term $\mathbf{R}_{l,\mathbf{w}} = \mathrm{E}(\mathbf{X}\mathbf{X}^{\mathrm{T}}|\widetilde{Y} = l, \mathbf{w})$ can be expressed as a fraction, whose numerator and denominator are easy to be estimated.

Proposition S3.2. *Conditional on* $\mathbf{W} = \mathbf{w}$*, for each* $l = 1, 2, \ldots, H$*, we have*

$$\mathrm{E}(\mathbf{X}\mathbf{X}^{T}|\widetilde{Y} = l, \mathbf{w}) = \frac{\mathrm{E}\{\mathbf{X}\mathbf{X}^{T}\mathbb{1}(\widetilde{Y} = l)|\mathbf{w}\}}{\mathrm{E}\{\mathbb{1}(\widetilde{Y} = l)|\mathbf{w}\}}. \tag{S3.28}$$

The proof of Proposition S3.2 is similar to that of Proposition 2, and is omitted.

Thus, the kernel matrix of partial variable dependent SAVE can be rewritten as

$$\mathbf{M}_{\text{SAVE}}(\mathbf{w}) = \sum\nolimits_{l=1}^{H} \mathbf{P}_{l,\mathbf{w}} \left\{ \boldsymbol{\Sigma}_{\mathbf{w}} - \frac{\mathbf{N}_{l,\mathbf{w}}}{\mathbf{P}_{l,\mathbf{w}}} + \frac{\mathbf{U}_{l,\mathbf{w}}\mathbf{U}_{l,\mathbf{w}}^{\mathrm{T}}}{\mathbf{P}_{l,\mathbf{w}}\mathbf{P}_{l,\mathbf{w}}} \right\}^{2}, \tag{S3.29}$$

where $\mathbf{N}_{l,\mathbf{w}} = \mathrm{E}\{\mathbf{X}\mathbf{X}^{\mathrm{T}}\mathbb{1}(\widetilde{Y} = l)|\mathbf{w}\}$.

Recall that the goal of variable dependent partial SAVE is to estimate the $\mathcal{S}_{Y_{\mathbf{w}}|\mathbf{X}_{\mathbf{w}}}$ by the estimates of $\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\mathbf{M}_{\text{SAVE}}(\mathbf{w})$. Similar to variable dependent partial SIR, we have the following NW kernel estimator of $\mathbf{N}_{l,\mathbf{w}}$:

$$\widehat{\mathbf{N}}_{l,\mathbf{w}} = \frac{\sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^{\mathrm{T}} \mathbb{1}(\widetilde{Y}_i = l) K_h(\mathbf{w}_i - \mathbf{w})}{\sum_{i=1}^{n} K_h(\mathbf{w}_i - \mathbf{w})}.$$

Then it is easy for us to get the sample estimator of $\mathbf{M}_{\mathrm{SAVE}}(\mathbf{w})$:

$$\widehat{\mathbf{M}}_{\mathrm{SAVE}}(\mathbf{w}) = \sum_{l=1}^{H} \widehat{\mathbf{P}}_{l,\mathbf{w}} \left\{ \widehat{\boldsymbol{\Sigma}}_{\mathbf{w}} - \frac{\widehat{\mathbf{N}}_{l,\mathbf{w}}}{\widehat{\mathbf{P}}_{l,\mathbf{w}}} + \frac{\widehat{\mathbf{U}}_{l,\mathbf{w}} \widehat{\mathbf{U}}_{l,\mathbf{w}}^{\mathrm{T}}}{\widehat{\mathbf{P}}_{l,\mathbf{w}} \widehat{\mathbf{P}}_{l,\mathbf{w}}} \right\}^2. \tag{S3.30}$$

Now we can use $\widehat{\boldsymbol{\Sigma}}_{\mathbf{w}}^{-1} \widehat{\mathbf{M}}_{\mathrm{SAVE}}(\mathbf{w})$ to estimate $\mathcal{S}_{Y_{\mathbf{w}}|\mathbf{X}_{\mathbf{w}}}$. Proposition S3.1 also leads us to consider the singular value decomposition. Let

$$\begin{aligned}
\boldsymbol{\Sigma}_{\mathbf{w}}^{-1} \mathbf{M}_{\mathrm{SAVE}}(\mathbf{w}) &= \sum_{k=1}^{p} \lambda_k^{\mathrm{SAVE}}(\mathbf{w}) \boldsymbol{\beta}_k^{\mathrm{SAVE}}(\mathbf{w}) \boldsymbol{\eta}_k^{\mathrm{SAVE}}(\mathbf{w}), \\
\lambda_1^{\mathrm{SAVE}}(\mathbf{w}) &\geq \cdots \geq \lambda_d^{\mathrm{SAVE}}(\mathbf{w}) = 0 = \cdots = \lambda_p^{\mathrm{SAVE}}(\mathbf{w}), \\
\widehat{\boldsymbol{\Sigma}}_{\mathbf{w}}^{-1} \widehat{\mathbf{M}}_{\mathrm{SAVE}}(\mathbf{w}) &= \sum_{k=1}^{p} \widehat{\lambda}_k^{\mathrm{SAVE}}(\mathbf{w}) \widehat{\boldsymbol{\beta}}_k^{\mathrm{SAVE}}(\mathbf{w}) \widehat{\boldsymbol{\eta}}_k^{\mathrm{SAVE}}(\mathbf{w}), \\
\widehat{\lambda}_1^{\mathrm{SAVE}}(\mathbf{w}) &\geq \cdots \geq \widehat{\lambda}_d^{\mathrm{SAVE}}(\mathbf{w}) \geq \cdots \geq \widehat{\lambda}_p^{\mathrm{SAVE}}(\mathbf{w}),
\end{aligned}$$

be the singular value decomposition of $\boldsymbol{\Sigma}_{\mathbf{w}}^{-1} \mathbf{M}_{\mathrm{SAVE}}(\mathbf{w})$ and $\widehat{\boldsymbol{\Sigma}}_{\mathbf{w}}^{-1} \widehat{\mathbf{M}}_{\mathrm{SAVE}}(\mathbf{w})$, respectively. Note that $\mathrm{Span}\{\boldsymbol{\beta}_1^{\mathrm{SAVE}}(\mathbf{w}), \ldots, \boldsymbol{\beta}_{d(\mathbf{w})}^{\mathrm{SAVE}}(\mathbf{w})\} = \mathcal{S}_{Y_{\mathbf{w}}|\mathbf{X}_{\mathbf{w}}}$, naturally, we propose to use $\mathrm{Span}\{\widehat{\boldsymbol{\beta}}_1^{\mathrm{SAVE}}(\mathbf{w}), \ldots, \widehat{\boldsymbol{\beta}}_{d(\mathbf{w})}^{\mathrm{SAVE}}(\mathbf{w})\}$ to estimate $\mathcal{S}_{Y_{\mathbf{w}}|\mathbf{X}_{\mathbf{w}}}$.

Proof of Theorem 3. Following by similar argument in Theorem 1, we have

$$\widehat{\mathbf{N}}_{l,\mathbf{w}} - \mathbf{N}_{l,\mathbf{w}} = \mathbf{C}_{\mathbf{N},\mathbf{w}} + \mathbf{B}_{\mathbf{N},\mathbf{w}} + \mathbf{O}_P\{\mathbf{R}_5(\mathbf{w})\},$$

where

$$\mathbf{C}_{\mathbf{N},\mathbf{w}} = \frac{1}{nhf(\mathbf{w})}\sum_{i=1}^{n} k_i(\mathbf{w})\left\{\mathbf{x}_i\mathbf{x}_i^{\mathrm{T}}\mathbb{1}(\widetilde{Y}_i = l) - \mathbf{N}_{l,\mathbf{w}_i}\right\}, \quad \mathbf{B}_{\mathbf{N},\mathbf{w}} = \frac{h^2\omega_2}{2}\left(2\frac{\dot{f}(\mathbf{w})}{f(\mathbf{w})}\dot{\mathbf{N}}_{l,\mathbf{w}} + \ddot{\mathbf{N}}_{l,\mathbf{w}}\right),$$

and

$$\mathbf{R}_5(\mathbf{w}) = \frac{1}{nf(\mathbf{w})}\left\{\left|\sum_{i=1}^{n} K(\frac{\mathbf{w}_i - \mathbf{w}}{h})\left(\mathbf{x}_i\mathbf{x}_i^{\mathrm{T}}\mathbb{1}(\widetilde{Y}_i = l) - \mathbf{N}_{l,\mathbf{w}_i}\right)\right|\right\} + \mathbf{o}(h^2).$$

Denote

$$\mathbf{E}_{l,\mathbf{w}} = \mathbf{\Sigma}_{\mathbf{w}} - \frac{\mathbf{N}_{l,\mathbf{w}}}{\mathbf{P}_{l,\mathbf{w}}} + \frac{\mathbf{U}_{l,\mathbf{w}}\mathbf{U}_{l,\mathbf{w}}^{\mathrm{T}}}{\mathbf{P}_{l,\mathbf{w}}\mathbf{P}_{l,\mathbf{w}}},$$

then the sample estimator of $\mathbf{E}_{l,\mathbf{w}}$ is

$$\widehat{\mathbf{E}}_{l,\mathbf{w}} = \widehat{\mathbf{\Sigma}}_{\mathbf{w}} - \frac{\widehat{\mathbf{N}}_{l,\mathbf{w}}}{\widehat{\mathbf{P}}_{l,\mathbf{w}}} + \frac{\widehat{\mathbf{U}}_{l,\mathbf{w}}\widehat{\mathbf{U}}_{l,\mathbf{w}}^{\mathrm{T}}}{\widehat{\mathbf{P}}_{l,\mathbf{w}}\widehat{\mathbf{P}}_{l,\mathbf{w}}},$$

thus

$$
\begin{aligned}
&\widehat{\mathbf{M}}_{\mathrm{SAVE}}(\mathbf{w})-\mathbf{M}_{\mathrm{SAVE}}(\mathbf{w})=\sum\nolimits_{l=1}^{H}\left(\widehat{\mathbf{P}}_{l,\mathbf{w}}\widehat{\mathbf{E}}_{l,\mathbf{w}}^{2}-\mathbf{P}_{l,\mathbf{w}}\mathbf{E}_{l,\mathbf{w}}^{2}\right)\\
&=\sum_{l=1}^{H}\left\{\left(\widehat{\mathbf{P}}_{l,\mathbf{w}}-\mathbf{P}_{l,\mathbf{w}}\right)\widehat{\mathbf{E}}_{l,\mathbf{w}}^{2}+\mathbf{P}_{l,\mathbf{w}}\left(\widehat{\mathbf{E}}_{l,\mathbf{w}}^{2}-\mathbf{E}_{l,\mathbf{w}}^{2}\right)\right\}\\
&=\sum\nolimits_{l=1}^{H}\left(\widehat{\mathbf{P}}_{l,\mathbf{w}}-\mathbf{P}_{l,\mathbf{w}}\right)\mathbf{E}_{l,\mathbf{w}}^{2}+\sum\nolimits_{l=1}^{H}\mathbf{P}_{l,\mathbf{w}}\left(\widehat{\mathbf{E}}_{l,\mathbf{w}}-\mathbf{E}_{l,\mathbf{w}}\right)\times\left(\widehat{\mathbf{E}}_{l,\mathbf{w}}+\mathbf{E}_{l,\mathbf{w}}\right)+\mathbf{o}_{P}\left(\frac{1}{\sqrt{nh}}\right)\\
&=\sum\nolimits_{l=1}^{H}\left(\widehat{\mathbf{P}}_{l,\mathbf{w}}-\mathbf{P}_{l,\mathbf{w}}\right)\mathbf{E}_{l,\mathbf{w}}^{2}+\sum\nolimits_{l=1}^{H}2\mathbf{P}_{l,\mathbf{w}}\Bigg\{\left(\widehat{\boldsymbol{\Sigma}}_{\mathbf{w}}-\boldsymbol{\Sigma}_{\mathbf{w}}\right)\boldsymbol{\Sigma}_{\mathbf{w}}\\
&+\frac{\mathbf{N}_{l,\mathbf{w}}\left(\widehat{\mathbf{P}}_{l,\mathbf{w}}-\mathbf{P}_{l,\mathbf{w}}\right)-\left(\widehat{\mathbf{N}}_{l,\mathbf{w}}-\mathbf{N}_{l,\mathbf{w}}\right)\mathbf{P}_{l,\mathbf{w}}}{\mathbf{P}_{l,\mathbf{w}}^{2}}\boldsymbol{\Sigma}_{\mathbf{w}}\\
&+\frac{\left(\widehat{\mathbf{U}}_{l,\mathbf{w}}-\mathbf{U}_{l,\mathbf{w}}\right)\mathbf{U}_{l,\mathbf{w}}^{\mathrm{T}}\mathbf{P}_{l,\mathbf{w}}+\mathbf{U}_{l,\mathbf{w}}\left(\widehat{\mathbf{U}}_{l,\mathbf{w}}-\mathbf{U}_{l,\mathbf{w}}\right)^{\mathrm{T}}\mathbf{P}_{l,\mathbf{w}}-2\mathbf{U}_{l,\mathbf{w}}\mathbf{U}_{l,\mathbf{w}}^{\mathrm{T}}\left(\widehat{\mathbf{P}}_{l,\mathbf{w}}-\mathbf{P}_{l,\mathbf{w}}\right)}{\mathbf{P}_{l,\mathbf{w}}^{3}}\boldsymbol{\Sigma}_{\mathbf{w}}\\
&-\left(\widehat{\boldsymbol{\Sigma}}_{\mathbf{w}}-\boldsymbol{\Sigma}_{\mathbf{w}}\right)\frac{\mathbf{N}_{l,\mathbf{w}}}{\mathbf{P}_{l,\mathbf{w}}}-\frac{\mathbf{N}_{l,\mathbf{w}}\left(\widehat{\mathbf{P}}_{l,\mathbf{w}}-\mathbf{P}_{l,\mathbf{w}}\right)-\left(\widehat{\mathbf{N}}_{l,\mathbf{w}}-\mathbf{N}_{l,\mathbf{w}}\right)\mathbf{P}_{l,\mathbf{w}}}{\mathbf{P}_{l,\mathbf{w}}^{2}}\frac{\mathbf{N}_{l,\mathbf{w}}}{\mathbf{P}_{l,\mathbf{w}}}\\
&-\frac{\left(\widehat{\mathbf{U}}_{l,\mathbf{w}}-\mathbf{U}_{l,\mathbf{w}}\right)\mathbf{U}_{l,\mathbf{w}}^{\mathrm{T}}\mathbf{P}_{l,\mathbf{w}}+\mathbf{U}_{l,\mathbf{w}}\left(\widehat{\mathbf{U}}_{l,\mathbf{w}}-\mathbf{U}_{l,\mathbf{w}}\right)^{\mathrm{T}}\mathbf{P}_{l,\mathbf{w}}-2\mathbf{U}_{l,\mathbf{w}}\mathbf{U}_{l,\mathbf{w}}^{\mathrm{T}}\left(\widehat{\mathbf{P}}_{l,\mathbf{w}}-\mathbf{P}_{l,\mathbf{w}}\right)}{\mathbf{P}_{l,\mathbf{w}}^{3}}\frac{\mathbf{N}_{l,\mathbf{w}}}{\mathbf{P}_{l,\mathbf{w}}}\\
&+\left(\widehat{\boldsymbol{\Sigma}}_{\mathbf{w}}-\boldsymbol{\Sigma}_{\mathbf{w}}\right)\frac{\mathbf{U}_{l,\mathbf{w}}\mathbf{U}_{l,\mathbf{w}}^{\mathrm{T}}}{\mathbf{P}_{l,\mathbf{w}}^{2}}+\frac{\mathbf{N}_{l,\mathbf{w}}\left(\widehat{\mathbf{P}}_{l,\mathbf{w}}-\mathbf{P}_{l,\mathbf{w}}\right)-\left(\widehat{\mathbf{N}}_{l,\mathbf{w}}-\mathbf{N}_{l,\mathbf{w}}\right)\mathbf{P}_{l,\mathbf{w}}}{\mathbf{P}_{l,\mathbf{w}}^{2}}\frac{\mathbf{U}_{l,\mathbf{w}}\mathbf{U}_{l,\mathbf{w}}^{\mathrm{T}}}{\mathbf{P}_{l,\mathbf{w}}^{2}}\\
&+\frac{\left(\widehat{\mathbf{U}}_{l,\mathbf{w}}-\mathbf{U}_{l,\mathbf{w}}\right)\mathbf{U}_{l,\mathbf{w}}^{\mathrm{T}}\mathbf{P}_{l,\mathbf{w}}+\mathbf{U}_{l,\mathbf{w}}\left(\widehat{\mathbf{U}}_{l,\mathbf{w}}-\mathbf{U}_{l,\mathbf{w}}\right)^{\mathrm{T}}\mathbf{P}_{l,\mathbf{w}}-2\mathbf{U}_{l,\mathbf{w}}\mathbf{U}_{l,\mathbf{w}}^{\mathrm{T}}\left(\widehat{\mathbf{P}}_{l,\mathbf{w}}-\mathbf{P}_{l,\mathbf{w}}\right)}{\mathbf{P}_{l,\mathbf{w}}^{3}}\frac{\mathbf{U}_{l,\mathbf{w}}\mathbf{U}_{l,\mathbf{w}}^{\mathrm{T}}}{\mathbf{P}_{l,\mathbf{w}}^{2}}\Bigg\}\\
&+\mathbf{o}_{P}\left(\frac{1}{\sqrt{nh}}\right)=\mathbf{B}_{\mathrm{SAVE}}+\mathbf{C}_{\mathrm{SAVE}}+\mathbf{o}_{P}\left(\frac{1}{\sqrt{nh}}\right),
\end{aligned}
$$

where

$$
\begin{aligned}
\mathbf{B}_{\text{SAVE}} = &\sum\nolimits_{l=1}^{H} \mathbf{B}_{\mathbf{P},\mathbf{w}}\mathbf{E}_{l,\mathbf{w}}^{2} + 2\sum\nolimits_{l=1}^{H} \mathbf{P}_{l,\mathbf{w}}\Big\{\mathbf{B}_{\boldsymbol{\Sigma},\mathbf{w}}\boldsymbol{\Sigma}_{\mathbf{w}} + \frac{\mathbf{N}_{l,\mathbf{w}}\mathbf{B}_{\mathbf{P},\mathbf{w}} - \mathbf{B}_{\mathbf{N},\mathbf{w}}\mathbf{P}_{l,\mathbf{w}}}{\mathbf{P}_{l,\mathbf{w}}^{2}}\boldsymbol{\Sigma}_{\mathbf{w}} \\
&+ \frac{\mathbf{B}_{\mathbf{U},\mathbf{w}}\mathbf{U}_{l,\mathbf{w}}^{\mathrm{T}}\mathbf{P}_{l,\mathbf{w}} + \mathbf{U}_{l,\mathbf{w}}\mathbf{B}_{\mathbf{U},\mathbf{w}}^{\mathrm{T}}\mathbf{P}_{l,\mathbf{w}} - 2\mathbf{U}_{l,\mathbf{w}}\mathbf{U}_{l,\mathbf{w}}^{\mathrm{T}}\mathbf{B}_{\mathbf{P},\mathbf{w}}}{\mathbf{P}_{l,\mathbf{w}}^{3}}\boldsymbol{\Sigma}_{\mathbf{w}} - \mathbf{B}_{\boldsymbol{\Sigma},\mathbf{w}}\frac{\mathbf{N}_{l,\mathbf{w}}}{\mathbf{P}_{l,\mathbf{w}}} \\
&- \frac{\mathbf{N}_{l,\mathbf{w}}\mathbf{B}_{\mathbf{P},\mathbf{w}} - \mathbf{B}_{\mathbf{N},\mathbf{w}}\mathbf{P}_{l,\mathbf{w}}}{\mathbf{P}_{l,\mathbf{w}}^{2}}\frac{\mathbf{N}_{l,\mathbf{w}}}{\mathbf{P}_{l,\mathbf{w}}} - \frac{\mathbf{B}_{\mathbf{U},\mathbf{w}}\mathbf{U}_{l,\mathbf{w}}^{\mathrm{T}}\mathbf{P}_{l,\mathbf{w}} + \mathbf{U}_{l,\mathbf{w}}\mathbf{B}_{\mathbf{U},\mathbf{w}}^{\mathrm{T}}\mathbf{P}_{l,\mathbf{w}} - 2\mathbf{U}_{l,\mathbf{w}}\mathbf{U}_{l,\mathbf{w}}^{\mathrm{T}}\mathbf{B}_{\mathbf{P},\mathbf{w}}}{\mathbf{P}_{l,\mathbf{w}}^{3}}\frac{\mathbf{N}_{l,\mathbf{w}}}{\mathbf{P}_{l,\mathbf{w}}} \\
&+ \mathbf{B}_{\boldsymbol{\Sigma},\mathbf{w}}\frac{\mathbf{U}_{l,\mathbf{w}}\mathbf{U}_{l,\mathbf{w}}^{\mathrm{T}}}{\mathbf{P}_{l,\mathbf{w}}^{2}} + \frac{\mathbf{N}_{l,\mathbf{w}}\mathbf{B}_{\mathbf{P},\mathbf{w}} - \mathbf{B}_{\mathbf{N},\mathbf{w}}\mathbf{P}_{l,\mathbf{w}}}{\mathbf{P}_{l,\mathbf{w}}^{2}}\frac{\mathbf{U}_{l,\mathbf{w}}\mathbf{U}_{l,\mathbf{w}}^{\mathrm{T}}}{\mathbf{P}_{l,\mathbf{w}}^{2}} \\
&+ \frac{\mathbf{B}_{\mathbf{U},\mathbf{w}}\mathbf{U}_{l,\mathbf{w}}^{\mathrm{T}}\mathbf{P}_{l,\mathbf{w}} + \mathbf{U}_{l,\mathbf{w}}\mathbf{B}_{\mathbf{U},\mathbf{w}}^{\mathrm{T}}\mathbf{P}_{l,\mathbf{w}} - 2\mathbf{U}_{l,\mathbf{w}}\mathbf{U}_{l,\mathbf{w}}^{\mathrm{T}}\mathbf{B}_{\mathbf{P},\mathbf{w}}}{\mathbf{P}_{l,\mathbf{w}}^{3}}\frac{\mathbf{U}_{l,\mathbf{w}}\mathbf{U}_{l,\mathbf{w}}^{\mathrm{T}}}{\mathbf{P}_{l,\mathbf{w}}^{2}}\Big\},
\end{aligned}
\tag{S3.31}
$$

and

$$
\begin{aligned}
\mathbf{C}_{\text{SAVE}} = &\sum\nolimits_{l=1}^{H} \mathbf{C}_{\mathbf{P},\mathbf{w}}\mathbf{E}_{l,\mathbf{w}}^{2} + 2\sum\nolimits_{l=1}^{H} \mathbf{P}_{l,\mathbf{w}}\Big\{\mathbf{C}_{\boldsymbol{\Sigma},\mathbf{w}}\boldsymbol{\Sigma}_{\mathbf{w}} + \frac{\mathbf{N}_{l,\mathbf{w}}\mathbf{C}_{\mathbf{P},\mathbf{w}} - \mathbf{C}_{\mathbf{N},\mathbf{w}}\mathbf{P}_{l,\mathbf{w}}}{\mathbf{P}_{l,\mathbf{w}}^{2}}\boldsymbol{\Sigma}_{\mathbf{w}} \\
&+ \frac{\mathbf{C}_{\mathbf{U},\mathbf{w}}\mathbf{U}_{l,\mathbf{w}}^{\mathrm{T}}\mathbf{P}_{l,\mathbf{w}} + \mathbf{U}_{l,\mathbf{w}}\mathbf{C}_{\mathbf{U},\mathbf{w}}^{\mathrm{T}}\mathbf{P}_{l,\mathbf{w}} - 2\mathbf{U}_{l,\mathbf{w}}\mathbf{U}_{l,\mathbf{w}}^{\mathrm{T}}\mathbf{C}_{\mathbf{P},\mathbf{w}}}{\mathbf{P}_{l,\mathbf{w}}^{3}}\boldsymbol{\Sigma}_{\mathbf{w}} - \mathbf{C}_{\boldsymbol{\Sigma},\mathbf{w}}\frac{\mathbf{N}_{l,\mathbf{w}}}{\mathbf{P}_{l,\mathbf{w}}} \\
&- \frac{\mathbf{N}_{l,\mathbf{w}}\mathbf{C}_{\mathbf{P},\mathbf{w}} - \mathbf{C}_{\mathbf{N},\mathbf{w}}\mathbf{P}_{l,\mathbf{w}}}{\mathbf{P}_{l,\mathbf{w}}^{2}}\frac{\mathbf{N}_{l,\mathbf{w}}}{\mathbf{P}_{l,\mathbf{w}}} - \frac{\mathbf{C}_{\mathbf{U},\mathbf{w}}\mathbf{U}_{l,\mathbf{w}}^{\mathrm{T}}\mathbf{P}_{l,\mathbf{w}} + \mathbf{U}_{l,\mathbf{w}}\mathbf{C}_{\mathbf{U},\mathbf{w}}^{\mathrm{T}}\mathbf{P}_{l,\mathbf{w}} - 2\mathbf{U}_{l,\mathbf{w}}\mathbf{U}_{l,\mathbf{w}}^{\mathrm{T}}\mathbf{C}_{\mathbf{P},\mathbf{w}}}{\mathbf{P}_{l,\mathbf{w}}^{3}}\frac{\mathbf{N}_{l,\mathbf{w}}}{\mathbf{P}_{l,\mathbf{w}}} \\
&+ \mathbf{C}_{\boldsymbol{\Sigma},\mathbf{w}}\frac{\mathbf{U}_{l,\mathbf{w}}\mathbf{U}_{l,\mathbf{w}}^{\mathrm{T}}}{\mathbf{P}_{l,\mathbf{w}}^{2}} + \frac{\mathbf{N}_{l,\mathbf{w}}\mathbf{C}_{\mathbf{P},\mathbf{w}} - \mathbf{C}_{\mathbf{N},\mathbf{w}}\mathbf{P}_{l,\mathbf{w}}}{\mathbf{P}_{l,\mathbf{w}}^{2}}\frac{\mathbf{U}_{l,\mathbf{w}}\mathbf{U}_{l,\mathbf{w}}^{\mathrm{T}}}{\mathbf{P}_{l,\mathbf{w}}^{2}} \\
&+ \frac{\mathbf{C}_{\mathbf{U},\mathbf{w}}\mathbf{U}_{l,\mathbf{w}}^{\mathrm{T}}\mathbf{P}_{l,\mathbf{w}} + \mathbf{U}_{l,\mathbf{w}}\mathbf{C}_{\mathbf{U},\mathbf{w}}^{\mathrm{T}}\mathbf{P}_{l,\mathbf{w}} - 2\mathbf{U}_{l,\mathbf{w}}\mathbf{U}_{l,\mathbf{w}}^{\mathrm{T}}\mathbf{C}_{\mathbf{P},\mathbf{w}}}{\mathbf{P}_{l,\mathbf{w}}^{3}}\frac{\mathbf{U}_{l,\mathbf{w}}\mathbf{U}_{l,\mathbf{w}}^{\mathrm{T}}}{\mathbf{P}_{l,\mathbf{w}}^{2}}\Big\}.
\end{aligned}
\tag{S3.32}
$$

Similar to the proof of Theorem 1, we have

$$\sqrt{nh}\left(\mathbf{vech}\{\widehat{\mathbf{M}}_{\text{SAVE}}(\mathbf{w})\} - \mathbf{vech}\{\mathbf{M}_{\text{SAVE}}(\mathbf{w})\} - \mathbf{vech}\{\mathbf{B}_{\text{SAVE}}(\mathbf{w})\}\right) \xrightarrow{d} \mathbf{N}(0, f^{-1}(\mathbf{w})\omega_0\mathbf{C}^{\text{SAVE}}(\mathbf{w})), \tag{S3.33}$$

where ω_0 is defined as follows. Hence

$$\mathbf{C}^{\text{SAVE}}(\mathbf{w}) = \text{Cov}[\mathbf{vech}\{\mathbf{C}_{\text{SAVE}}(\mathbf{w})\}|\mathbf{w}]. \tag{S3.34}$$

Then for singular value decomposition, in partial variable dependent SAVE, denote $\mathbf{G}(\mathbf{w}) = \mathbf{\Sigma}_{\mathbf{w}}^{-1}\mathbf{M}_{\text{SAVE}}(\mathbf{w})$, then substitute it into (S2.20) and (S2.21),

$$\begin{aligned}\mathbf{B}_k^{\text{SAVE}}(\mathbf{w}) = \sum_{j\neq k}\frac{\boldsymbol{\beta}_j^{\text{SAVE}}(\mathbf{w})\{\boldsymbol{\beta}_j^{\text{SAVE}}(\mathbf{w})\}^{\mathrm{T}}}{\{\lambda_j^{\text{SAVE}}(\mathbf{w})\}^2 - \{\lambda_k^{\text{SAVE}}(\mathbf{w})\}^2}\Big[&\mathbf{M}_{\text{SAVE}}^{\mathrm{T}}(\mathbf{w})\mathbf{\Sigma}_{\mathbf{w}}^{-1}\mathbf{B}_{\mathbf{\Sigma},\mathbf{w}}\mathbf{\Sigma}_{\mathbf{w}}^{-1}\mathbf{\Sigma}_{\mathbf{w}}^{-1}\mathbf{M}_{\text{SAVE}}(\mathbf{w})\\ &- \mathbf{B}_{\text{SAVE}}^{\mathrm{T}}(\mathbf{w})\mathbf{\Sigma}_{\mathbf{w}}^{-1}\mathbf{\Sigma}_{\mathbf{w}}^{-1}\mathbf{M}_{\text{SAVE}}(\mathbf{w}) + \{\mathbf{\Sigma}_{\mathbf{w}}^{-1}\mathbf{M}_{\text{SAVE}}(\mathbf{w})\}^{\mathrm{T}}\mathbf{\Sigma}_{\mathbf{w}}^{-1}\mathbf{B}_{\mathbf{\Sigma},\mathbf{w}}\mathbf{\Sigma}_{\mathbf{w}}^{-1}\mathbf{M}_{\text{SAVE}}(\mathbf{w})\\ &- \{\mathbf{\Sigma}_{\mathbf{w}}^{-1}\mathbf{M}_{\text{SAVE}}(\mathbf{w})\}^{\mathrm{T}}\mathbf{\Sigma}_{\mathbf{w}}^{-1}\mathbf{B}_{\text{SAVE}}(\mathbf{w})\Big]\boldsymbol{\beta}_k^{\text{SAVE}}(\mathbf{w}),\end{aligned} \tag{S3.35}$$

$$\begin{aligned}\mathbf{C}_k^{\text{SAVE}}(\mathbf{w}) = \sum_{j\neq k}\frac{\boldsymbol{\beta}_j^{\text{SAVE}}(\mathbf{w})\{\boldsymbol{\beta}_j^{\text{SAVE}}(\mathbf{w})\}^{\mathrm{T}}}{\{\lambda_j^{\text{SAVE}}(\mathbf{w})\}^2 - \{\lambda_k^{\text{SAVE}}(\mathbf{w})\}^2}\Big[&\mathbf{M}_{\text{SAVE}}^{\mathrm{T}}(\mathbf{w})\mathbf{\Sigma}_{\mathbf{w}}^{-1}\mathbf{C}_{\mathbf{\Sigma},\mathbf{w}}\mathbf{\Sigma}_{\mathbf{w}}^{-1}\mathbf{\Sigma}_{\mathbf{w}}^{-1}\mathbf{M}_{\text{SAVE}}(\mathbf{w})\\ &- \mathbf{C}_{\text{SAVE}}(\mathbf{w})^{\mathrm{T}}\mathbf{\Sigma}_{\mathbf{w}}^{-1}\mathbf{\Sigma}_{\mathbf{w}}^{-1}\mathbf{M}_{\text{SAVE}}(\mathbf{w}) + \{\mathbf{\Sigma}_{\mathbf{w}}^{-1}\mathbf{M}_{\text{SAVE}}(\mathbf{w})\}^{\mathrm{T}}\mathbf{\Sigma}_{\mathbf{w}}^{-1}\mathbf{C}_{\mathbf{\Sigma},\mathbf{w}}\mathbf{\Sigma}_{\mathbf{w}}^{-1}\mathbf{M}_{\text{SAVE}}(\mathbf{w})\\ &- \{\mathbf{\Sigma}_{\mathbf{w}}^{-1}\mathbf{M}_{\text{SAVE}}(\mathbf{w})\}^{\mathrm{T}}\mathbf{\Sigma}_{\mathbf{w}}^{-1}\mathbf{C}_{\text{SAVE}}(\mathbf{w})\Big]\boldsymbol{\beta}_k^{\text{SAVE}}(\mathbf{w}).\end{aligned}$$

Hence,

$$\boldsymbol{\Sigma}_k^{\text{SAVE}}(\mathbf{w}) = \text{Cov}\{\mathbf{C}_k^{\text{SAVE}}(\mathbf{w})|\mathbf{w}\} \tag{S3.36}$$

$\square$

S4. Variable Dependent Partial DR

Another popular method of sufficient dimension reduction is Directional Regression (DR) (Li and Wang, 2007), which implicitly synthesizes sliced inverse regression and sliced average variance estimation. DR enjoys the advantage of high accuracy and convenient computation, and has received substantial attention in the literature of sufficient dimension reduction (Yu, 2014; Yu and Dong, 2016). Parallel to variable dependent partial SIR and variable dependent partial SAVE, we now propose variable dependent partial DR approach to perform variable dependent partial dimension reduction.

Now we can define the kernel matrix for variable dependent partial DR,

$$\mathbf{M}_{\text{DR}}(\mathbf{w}) \triangleq \text{E}_{Y,\breve{Y}}\{2\boldsymbol{\Sigma}_{\mathbf{w}} - \text{E}\{(\mathbf{X}-\breve{\mathbf{X}})(\mathbf{X}-\breve{\mathbf{X}})^{\text{T}}|Y,\breve{Y},\mathbf{W}=\mathbf{w}\}\}^2,$$

Where $\text{E}_{Y,\breve{Y}}$ represents expectation with respect to Y and $\breve{Y}$.

Proposition S4.3. *Conditional on* $\mathbf{W} = \mathbf{w}$*, assume that linear conditional mean condition (A1) and constant conditional variance condition (A2) hold, then*

$$
\begin{aligned}
&\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}[2\boldsymbol{\Sigma}_{\mathbf{w}} - \mathrm{E}\{(\mathbf{X} - \breve{\mathbf{X}})(\mathbf{X} - \breve{\mathbf{X}})^{T}|Y, \breve{Y}, \mathbf{W} = \mathbf{w}\}]\\
=&\mathbf{P}_{\mathbf{B}(\mathbf{w})}[2\mathbf{I}_p - \boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\mathrm{E}\{(\mathbf{X} - \breve{\mathbf{X}})(\mathbf{X} - \breve{\mathbf{X}})^{T}|Y, \breve{Y}, \mathbf{W} = \mathbf{w}\}]\mathbf{P}_{\mathbf{B}(\mathbf{w})},
\end{aligned}
$$

where $(\breve{\mathbf{X}}, \breve{Y})$ *is an independent copy of* $(\mathbf{X}, Y)$*.*

Proof of Proposition S4.3. Assume that $\mathbf{X}_{\mathbf{w}}$ satisfies linear conditional mean (A1) and constant conditional variance (A2), conditional on $\mathbf{W} = \mathbf{w}$, it's easy for us to get

$$
\begin{aligned}
&\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\mathrm{E}\{(\mathbf{X} - \breve{\mathbf{X}})(\mathbf{X} - \breve{\mathbf{X}})^{\mathrm{T}}|Y, \breve{Y}, \mathbf{W} = \mathbf{w}\}\\
&= \boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\mathrm{E}\{(\mathbf{X}_{\mathbf{w}} - \breve{\mathbf{X}}_{\mathbf{w}})(\mathbf{X}_{\mathbf{w}} - \breve{\mathbf{X}}_{\mathbf{w}})^{\mathrm{T}}|Y_{\mathbf{w}}, \breve{Y}_{\mathbf{w}}\}\\
&= \boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\{\mathrm{E}(\mathbf{X}_{\mathbf{w}}\mathbf{X}_{\mathbf{w}}^{\mathrm{T}}|Y_{\mathbf{w}}) - \mathrm{E}(\mathbf{X}_{\mathbf{w}}|Y_{\mathbf{w}})\mathrm{E}(\breve{\mathbf{X}}_{\mathbf{w}}^{\mathrm{T}}|\breve{Y}_{\mathbf{w}}) - \mathrm{E}(\breve{\mathbf{X}}_{\mathbf{w}}|\breve{Y}_{\mathbf{w}})\mathrm{E}(\mathbf{X}_{\mathbf{w}}^{\mathrm{T}}|Y_{\mathbf{w}}) + \mathrm{E}(\breve{\mathbf{X}}_{\mathbf{w}}\breve{\mathbf{X}}_{\mathbf{w}}^{\mathrm{T}}|\breve{Y}_{\mathbf{w}})\}\\
&= \boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\Big(\mathrm{E}\left[\mathrm{E}\{\mathbf{X}_{\mathbf{w}}\mathbf{X}_{\mathbf{w}}^{\mathrm{T}}|\mathbf{B}^{\mathrm{T}}(\mathbf{w})\mathbf{X}_{\mathbf{w}}, Y_{\mathbf{w}}\}|Y_{\mathbf{w}}\right] - \mathrm{E}[\mathrm{E}\{\mathbf{X}_{\mathbf{w}}|\mathbf{B}^{\mathrm{T}}(\mathbf{w})\mathbf{X}_{\mathbf{w}}, Y_{\mathbf{w}}\}|Y_{\mathbf{w}}]\mathrm{E}[\mathrm{E}\{\breve{\mathbf{X}}_{\mathbf{w}}^{\mathrm{T}}|\mathbf{B}^{\mathrm{T}}(\mathbf{w})\breve{\mathbf{X}}_{\mathbf{w}}, \breve{Y}_{\mathbf{w}}\}|\breve{Y}_{\mathbf{w}}]\\
&- \mathrm{E}[\mathrm{E}\{\breve{\mathbf{X}}_{\mathbf{w}}|\mathbf{B}^{\mathrm{T}}(\mathbf{w})\breve{\mathbf{X}}_{\mathbf{w}}, \breve{Y}_{\mathbf{w}}\}|\breve{Y}_{\mathbf{w}}]\mathrm{E}[\mathrm{E}\{\mathbf{X}_{\mathbf{w}}^{\mathrm{T}}|\mathbf{B}^{\mathrm{T}}(\mathbf{w})\mathbf{X}_{\mathbf{w}}, Y_{\mathbf{w}}\}|Y_{\mathbf{w}}] + \mathrm{E}\left[\mathrm{E}\{\mathbf{X}_{\mathbf{w}}\mathbf{X}_{\mathbf{w}}^{\mathrm{T}}|\mathbf{B}^{\mathrm{T}}(\mathbf{w})\mathbf{X}_{\mathbf{w}}, Y_{\mathbf{w}}\}|Y_{\mathbf{w}}\right]\Big)\\
&= \boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\Big(\mathrm{E}\left[\mathrm{Cov}\{\mathbf{X}_{\mathbf{w}}|\mathbf{B}^{\mathrm{T}}(\mathbf{w})\mathbf{X}_{\mathbf{w}}\}|\mathrm{Y}_{\mathbf{w}}\right] + \mathrm{E}[\mathrm{E}\{\mathbf{X}_{\mathbf{w}}|\mathbf{B}^{\mathrm{T}}(\mathbf{w})\mathbf{X}_{\mathbf{w}}\}\mathrm{E}\{\mathbf{X}_{\mathbf{w}}^{\mathrm{T}}|\mathbf{B}^{\mathrm{T}}(\mathbf{w})\mathbf{X}_{\mathbf{w}}\}|Y_{\mathbf{w}}]\\
&- \mathrm{E}[\mathrm{E}\{\mathbf{X}_{\mathbf{w}}|\mathbf{B}^{\mathrm{T}}(\mathbf{w})\mathbf{X}_{\mathbf{w}}\}|Y_{\mathbf{w}}]\mathrm{E}[\mathrm{E}\{\breve{\mathbf{X}}_{\mathbf{w}}^{\mathrm{T}}|\mathbf{B}^{\mathrm{T}}(\mathbf{w})\breve{\mathbf{X}}_{\mathbf{w}}\}|\breve{Y}_{\mathbf{w}}] - \mathrm{E}[\mathrm{E}\{\breve{\mathbf{X}}_{\mathbf{w}}|\mathbf{B}^{\mathrm{T}}(\mathbf{w})\breve{\mathbf{X}}_{\mathbf{w}}\}|\breve{Y}_{\mathbf{w}}]\mathrm{E}[\mathrm{E}\{\mathbf{X}_{\mathbf{w}}^{\mathrm{T}}|\mathbf{B}^{\mathrm{T}}(\mathbf{w})\mathbf{X}_{\mathbf{w}}\}|Y_{\mathbf{w}}]\\
&+ \mathrm{E}\left[\mathrm{Cov}\{\breve{\mathbf{X}}_{\mathbf{w}}|\mathbf{B}^{\mathrm{T}}(\mathbf{w})\breve{\mathbf{X}}_{\mathbf{w}}\}|\breve{\mathrm{Y}}_{\mathbf{w}}\right] + \mathrm{E}\left[\mathrm{E}\{\breve{\mathbf{X}}_{\mathbf{w}}|\mathbf{B}^{\mathrm{T}}(\mathbf{w})\breve{\mathbf{X}}_{\mathbf{w}}\}\mathrm{E}\{\breve{\mathbf{X}}_{\mathbf{w}}^{\mathrm{T}}|\mathbf{B}^{\mathrm{T}}(\mathbf{w})\breve{\mathbf{X}}_{\mathbf{w}}\}|\breve{Y}_{\mathbf{w}}\right]\Big)
\end{aligned}
$$

$$= \mathbf{Q}_{\mathbf{B}(\mathbf{w})} + \mathbf{P}_{\mathbf{B}(\mathbf{w})}\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\mathrm{E}(\mathbf{X}_{\mathbf{w}}\mathbf{X}_{\mathbf{w}}^{\mathrm{T}}|Y_{\mathbf{w}})\mathbf{P}_{\mathbf{B}(\mathbf{w})} - \mathbf{P}_{\mathbf{B}(\mathbf{w})}\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\mathrm{E}(\mathbf{X}_{\mathbf{w}}|Y_{\mathbf{w}})\mathrm{E}(\breve{\mathbf{X}}_{\mathbf{w}}^{\mathrm{T}}|\breve{Y}_{\mathbf{w}})\mathbf{P}_{\mathbf{B}(\mathbf{w})}$$

$$- \mathbf{P}_{\mathbf{B}(\mathbf{w})}\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\mathrm{E}(\breve{\mathbf{X}}_{\mathbf{w}}|\breve{Y}_{\mathbf{w}})\mathrm{E}(\mathbf{X}_{\mathbf{w}}^{\mathrm{T}}|Y_{\mathbf{w}})\mathbf{P}_{\mathbf{B}(\mathbf{w})} + \mathbf{Q}_{\mathbf{B}(\mathbf{w})} + \mathbf{P}_{\mathbf{B}(\mathbf{w})}\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\mathrm{E}(\breve{\mathbf{X}}_{\mathbf{w}}\breve{\mathbf{X}}_{\mathbf{w}}^{\mathrm{T}}|\breve{Y}_{\mathbf{w}})\mathbf{P}_{\mathbf{B}(\mathbf{w})}.$$

Recall that $\mathbf{Q}_{\mathbf{B}(\mathbf{w})} = \mathbf{I}_p - \mathbf{P}_{\mathbf{B}(\mathbf{w})}$. Then we can derive that

$$\begin{aligned} &\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}[2\boldsymbol{\Sigma}_{\mathbf{w}} - \mathrm{E}\{(\mathbf{X} - \breve{\mathbf{X}})(\mathbf{X} - \breve{\mathbf{X}})^{\mathrm{T}}|Y, \breve{Y}, \mathbf{W} = \mathbf{w}\}] \\ = &\mathbf{P}_{\mathbf{B}(\mathbf{w})}[2\mathbf{I}_p - \boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\mathrm{E}\{(\mathbf{X} - \breve{\mathbf{X}})(\mathbf{X} - \breve{\mathbf{X}})^{\mathrm{T}}|Y, \breve{Y}, \mathbf{W} = \mathbf{w}\}]\mathbf{P}_{\mathbf{B}(\mathbf{w})}. \end{aligned}$$

The proof is completed. □

Proposition S4.3 implies that $\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\mathbf{M}_{\mathrm{DR}}(\mathbf{w}) \subseteq \mathcal{S}_{Y_{\mathbf{w}}|\mathbf{X}_{\mathbf{w}}}$ almost surely. According to Proposition 1 of Li and Wang (2007), the kernel matrix $\mathbf{M}_{\mathrm{DR}}(\mathbf{w})$ can be rewritten as

$$\begin{aligned} \mathbf{M}_{\mathrm{DR}}(\mathbf{w}) =&2\sum_{l=1}^{H}\mathbf{P}_{l,\mathbf{w}}\left(\mathbf{R}_{l,\mathbf{w}} - \boldsymbol{\Sigma}_{\mathbf{w}}\right)^2 + 2\left(\sum_{l=1}^{H}\mathbf{P}_{l,\mathbf{w}}\mathbf{V}_{l,\mathbf{w}}\mathbf{V}_{l,\mathbf{w}}^{\mathrm{T}}\right)^2 \\ &+ 2\left(\sum_{l=1}^{H}\mathbf{P}_{l,\mathbf{w}}\mathbf{V}_{l,\mathbf{w}}^{\mathrm{T}}\mathbf{V}_{l,\mathbf{w}}\right)\left(\sum_{l=1}^{H}\mathbf{P}_{l,\mathbf{w}}\mathbf{V}_{l,\mathbf{w}}\mathbf{V}_{l,\mathbf{w}}^{\mathrm{T}}\right), \end{aligned}$$

where $\mathbf{P}_{l,\mathbf{w}}$, $\mathbf{R}_{l,\mathbf{w}}$ and $\mathbf{V}_{l,\mathbf{w}}$ are defined as previously. And the sample estimator of $\mathbf{M}_{\mathrm{DR}}(\mathbf{w})$ can be easily found by

$$\begin{aligned}\widehat{\mathbf{M}}_{\mathrm{DR}}(\mathbf{w}) =& 2\sum_{l=1}^{H}\widehat{\mathbf{P}}_{l,\mathbf{w}}\left(\widehat{\mathbf{R}}_{l,\mathbf{w}}-\widehat{\boldsymbol{\Sigma}}_{\mathbf{w}}\right)^2+2\left(\sum_{l=1}^{H}\widehat{\mathbf{P}}_{l,\mathbf{w}}\widehat{\mathbf{V}}_{l,\mathbf{w}}\widehat{\mathbf{V}}_{l,\mathbf{w}}^{\mathrm{T}}\right)^2 \\ &+2\left(\sum_{l=1}^{H}\widehat{\mathbf{P}}_{l,\mathbf{w}}\widehat{\mathbf{V}}_{l,\mathbf{w}}^{\mathrm{T}}\widehat{\mathbf{V}}_{l,\mathbf{w}}\right)\left(\sum_{l=1}^{H}\widehat{\mathbf{P}}_{l,\mathbf{w}}\widehat{\mathbf{V}}_{l,\mathbf{w}}\widehat{\mathbf{V}}_{l,\mathbf{w}}^{\mathrm{T}}\right).\end{aligned}\tag{S4.37}$$

Then we use $\widehat{\boldsymbol{\Sigma}}_{\mathbf{w}}^{-1}\widehat{\mathbf{M}}_{\mathrm{DR}}(\mathbf{w})$ to estimate $\mathcal{S}_{Y_{\mathbf{w}}|\mathbf{X}_{\mathbf{w}}}$. Proposition S4.3 also leads us to consider the singular value decomposition. Let

$$\begin{aligned}\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\mathbf{M}_{\mathrm{DR}}(\mathbf{w}) &= \sum_{k=1}^{p}\lambda_k^{\mathrm{DR}}(\mathbf{w})\boldsymbol{\beta}_k^{\mathrm{DR}}(\mathbf{w})\boldsymbol{\eta}_k^{\mathrm{DR}}(\mathbf{w}), \quad \lambda_1^{\mathrm{DR}}(\mathbf{w})\ge\cdots\ge\lambda_d^{\mathrm{DR}}(\mathbf{w})=0=\cdots=\lambda_p^{\mathrm{DR}}(\mathbf{w}),\\ \widehat{\boldsymbol{\Sigma}}_{\mathbf{w}}^{-1}\widehat{\mathbf{M}}_{\mathrm{DR}}(\mathbf{w}) &= \sum_{k=1}^{p}\widehat{\lambda}_k^{\mathrm{DR}}(\mathbf{w})\widehat{\boldsymbol{\beta}}_k^{\mathrm{DR}}(\mathbf{w})\widehat{\boldsymbol{\eta}}_k^{\mathrm{DR}}(\mathbf{w}), \quad \widehat{\lambda}_1^{\mathrm{DR}}(\mathbf{w})\ge\cdots\ge\widehat{\lambda}_d^{\mathrm{DR}}(\mathbf{w})\ge\cdots\ge\widehat{\lambda}_p^{\mathrm{DR}}(\mathbf{w}),\end{aligned}$$

be the singular value decomposition of $\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\mathbf{M}_{\mathrm{DR}}(\mathbf{w})$ and $\widehat{\boldsymbol{\Sigma}}_{\mathbf{w}}^{-1}\widehat{\mathbf{M}}_{\mathrm{DR}}(\mathbf{w})$, respectively. Then $\mathrm{Span}\{\boldsymbol{\beta}_1^{\mathrm{DR}}(\mathbf{w}),\ldots,\boldsymbol{\beta}_{d(\mathbf{w})}^{\mathrm{DR}}(\mathbf{w})\} = \mathcal{S}_{Y_{\mathbf{w}}|\mathbf{X}_{\mathbf{w}}}$, and the final sample estimator of $\mathcal{S}_{Y_{\mathbf{w}}|\mathbf{X}_{\mathbf{w}}}$ is $\mathrm{Span}\{\widehat{\boldsymbol{\beta}}_1^{\mathrm{DR}}(\mathbf{w}),\ldots,\widehat{\boldsymbol{\beta}}_{d(\mathbf{w})}^{\mathrm{DR}}(\mathbf{w})\}$.

Proof of Theorem 4. Firstly, we have

$$\begin{aligned}&\widehat{\mathbf{M}}_{\mathrm{DR}}(\mathbf{w})-\mathbf{M}_{\mathrm{DR}}(\mathbf{w})\\ =&\,2\sum_{l=1}^{H}\left\{\widehat{\mathbf{P}}_{l,\mathbf{w}}\left(\frac{\widehat{\mathbf{N}}_{l,\mathbf{w}}}{\widehat{\mathbf{P}}_{l,\mathbf{w}}}-\widehat{\boldsymbol{\Sigma}}_{\mathbf{w}}\right)^2-\mathbf{P}_{l,\mathbf{w}}\left(\frac{\mathbf{N}_{l,\mathbf{w}}}{\mathbf{P}_{l,\mathbf{w}}}-\boldsymbol{\Sigma}_{\mathbf{w}}\right)^2\right\}+2\left\{\left(\sum_{l=1}^{H}\frac{\widehat{\mathbf{U}}_{l,\mathbf{w}}\widehat{\mathbf{U}}_{l,\mathbf{w}}^{\mathrm{T}}}{\widehat{\mathbf{P}}_{l,\mathbf{w}}}\right)^2-\left(\sum_{l=1}^{H}\frac{\mathbf{U}_{l,\mathbf{w}}\mathbf{U}_{l,\mathbf{w}}^{\mathrm{T}}}{\mathbf{P}_{l,\mathbf{w}}}\right)^2\right\}\\ &+2\left\{\left(\sum_{l=1}^{H}\frac{\widehat{\mathbf{U}}_{l,\mathbf{w}}^{\mathrm{T}}\widehat{\mathbf{U}}_{l,\mathbf{w}}}{\widehat{\mathbf{P}}_{l,\mathbf{w}}}\right)\left(\sum_{l=1}^{H}\frac{\widehat{\mathbf{U}}_{l,\mathbf{w}}\widehat{\mathbf{U}}_{l,\mathbf{w}}^{\mathrm{T}}}{\widehat{\mathbf{P}}_{l,\mathbf{w}}}\right)-\left(\sum_{l=1}^{H}\frac{\mathbf{U}_{l,\mathbf{w}}^{\mathrm{T}}\mathbf{U}_{l,\mathbf{w}}}{\mathbf{P}_{l,\mathbf{w}}}\right)\left(\sum_{l=1}^{H}\frac{\mathbf{U}_{l,\mathbf{w}}\mathbf{U}_{l,\mathbf{w}}^{\mathrm{T}}}{\mathbf{P}_{l,\mathbf{w}}}\right)\right\}\end{aligned}$$

$$
\begin{aligned}
&= 2\sum_{l=1}^{H}\Bigg\{ \left(\widehat{\mathbf{P}}_{l,\mathbf{w}}-\mathbf{P}_{l,\mathbf{w}}\right)\left(\frac{\mathbf{N}_{l,\mathbf{w}}}{\mathbf{P}_{l,\mathbf{w}}}-\boldsymbol{\Sigma}_{\mathbf{w}}\right)^{2}+\mathbf{P}_{l,\mathbf{w}}\left(\frac{\widehat{\mathbf{N}}_{l,\mathbf{w}}}{\widehat{\mathbf{P}}_{l,\mathbf{w}}}-\widehat{\boldsymbol{\Sigma}}_{\mathbf{w}}+\frac{\mathbf{N}_{l,\mathbf{w}}}{\mathbf{P}_{l,\mathbf{w}}}-\boldsymbol{\Sigma}_{\mathbf{w}}\right)\\
&\times\left(\frac{\widehat{\mathbf{N}}_{l,\mathbf{w}}}{\widehat{\mathbf{P}}_{l,\mathbf{w}}}-\widehat{\boldsymbol{\Sigma}}_{\mathbf{w}}-\frac{\mathbf{N}_{l,\mathbf{w}}}{\mathbf{P}_{l,\mathbf{w}}}+\boldsymbol{\Sigma}_{\mathbf{w}}\right)\Bigg\}+2\sum_{l=1}^{H}\left(\frac{\widehat{\mathbf{U}}_{l,\mathbf{w}}\widehat{\mathbf{U}}_{l,\mathbf{w}}^{\mathrm{T}}}{\widehat{\mathbf{P}}_{l,\mathbf{w}}}+\frac{\mathbf{U}_{l,\mathbf{w}}\mathbf{U}_{l,\mathbf{w}}^{\mathrm{T}}}{\mathbf{P}_{l,\mathbf{w}}}\right)\\
&\times\sum_{l=1}^{H}\left(\frac{\widehat{\mathbf{U}}_{l,\mathbf{w}}\widehat{\mathbf{U}}_{l,\mathbf{w}}^{\mathrm{T}}}{\widehat{\mathbf{P}}_{l,\mathbf{w}}}-\frac{\mathbf{U}_{l,\mathbf{w}}\mathbf{U}_{l,\mathbf{w}}^{\mathrm{T}}}{\mathbf{P}_{l,\mathbf{w}}}\right)+2\sum_{l=1}^{H}\left(\frac{\widehat{\mathbf{U}}_{l,\mathbf{w}}^{\mathrm{T}}\widehat{\mathbf{U}}_{l,\mathbf{w}}}{\widehat{\mathbf{P}}_{l,\mathbf{w}}}-\frac{\mathbf{U}_{l,\mathbf{w}}^{\mathrm{T}}\mathbf{U}_{l,\mathbf{w}}}{\mathbf{P}_{l,\mathbf{w}}}\right)\left(\sum_{l=1}^{H}\frac{\mathbf{U}_{l,\mathbf{w}}\mathbf{U}_{l,\mathbf{w}}^{\mathrm{T}}}{\mathbf{P}_{l,\mathbf{w}}}\right)\\
&+2\sum_{l=1}^{H}\frac{\mathbf{U}_{l,\mathbf{w}}^{\mathrm{T}}\mathbf{U}_{l,\mathbf{w}}}{\mathbf{P}_{l,\mathbf{w}}}\sum_{l=1}^{H}\left(\frac{\widehat{\mathbf{U}}_{l,\mathbf{w}}\widehat{\mathbf{U}}_{l,\mathbf{w}}^{\mathrm{T}}}{\widehat{\mathbf{P}}_{l,\mathbf{w}}}-\frac{\mathbf{U}_{l,\mathbf{w}}\mathbf{U}_{l,\mathbf{w}}^{\mathrm{T}}}{\mathbf{P}_{l,\mathbf{w}}}\right)\Bigg\}+\mathbf{o}_{P}\left(\frac{1}{\sqrt{nh}}\right)\\
&= 2\sum_{l=1}^{H}\Bigg\{ \left(\widehat{\mathbf{P}}_{l,\mathbf{w}}-\mathbf{P}_{l,\mathbf{w}}\right)\left(\frac{\mathbf{N}_{l,\mathbf{w}}}{\mathbf{P}_{l,\mathbf{w}}}-\boldsymbol{\Sigma}_{\mathbf{w}}\right)^{2}+2\mathbf{N}_{l,\mathbf{w}}\left(\frac{\widehat{\mathbf{N}}_{l,\mathbf{w}}}{\widehat{\mathbf{P}}_{l,\mathbf{w}}}-\frac{\mathbf{N}_{l,\mathbf{w}}}{\mathbf{P}_{l,\mathbf{w}}}\right)-2\mathbf{N}_{l,\mathbf{w}}\left(\widehat{\boldsymbol{\Sigma}}_{\mathbf{w}}-\boldsymbol{\Sigma}_{\mathbf{w}}\right)\\
&-2\mathbf{P}_{l,\mathbf{w}}\boldsymbol{\Sigma}_{\mathbf{w}}\left(\frac{\widehat{\mathbf{N}}_{l,\mathbf{w}}}{\widehat{\mathbf{P}}_{l,\mathbf{w}}}-\frac{\mathbf{N}_{l,\mathbf{w}}}{\mathbf{P}_{l,\mathbf{w}}}\right)+2\mathbf{P}_{l,\mathbf{w}}\boldsymbol{\Sigma}_{\mathbf{w}}\left(\widehat{\boldsymbol{\Sigma}}_{\mathbf{w}}-\boldsymbol{\Sigma}_{\mathbf{w}}\right)\Bigg\}\\
&+4\Bigg\{\sum_{l=1}^{H}\frac{\mathbf{U}_{l,\mathbf{w}}\mathbf{U}_{l,\mathbf{w}}^{\mathrm{T}}}{\mathbf{P}_{l,\mathbf{w}}}\sum_{l=1}^{H}\left(\frac{\widehat{\mathbf{U}}_{l,\mathbf{w}}\widehat{\mathbf{U}}_{l,\mathbf{w}}^{\mathrm{T}}}{\widehat{\mathbf{P}}_{l,\mathbf{w}}}-\frac{\mathbf{U}_{l,\mathbf{w}}\mathbf{U}_{l,\mathbf{w}}^{\mathrm{T}}}{\mathbf{P}_{l,\mathbf{w}}}\right)\Bigg\}+2\sum_{l=1}^{H}\left(\frac{\widehat{\mathbf{U}}_{l,\mathbf{w}}^{\mathrm{T}}\widehat{\mathbf{U}}_{l,\mathbf{w}}}{\widehat{\mathbf{P}}_{l,\mathbf{w}}}-\frac{\mathbf{U}_{l,\mathbf{w}}^{\mathrm{T}}\mathbf{U}_{l,\mathbf{w}}}{\mathbf{P}_{l,\mathbf{w}}}\right)\left(\sum_{l=1}^{H}\frac{\mathbf{U}_{l,\mathbf{w}}\mathbf{U}_{l,\mathbf{w}}^{\mathrm{T}}}{\mathbf{P}_{l,\mathbf{w}}}\right)\\
&+2\sum_{l=1}^{H}\frac{\mathbf{U}_{l,\mathbf{w}}^{\mathrm{T}}\mathbf{U}_{l,\mathbf{w}}}{\mathbf{P}_{l,\mathbf{w}}}\sum_{l=1}^{H}\left(\frac{\widehat{\mathbf{U}}_{l,\mathbf{w}}\widehat{\mathbf{U}}_{l,\mathbf{w}}^{\mathrm{T}}}{\widehat{\mathbf{P}}_{l,\mathbf{w}}}-\frac{\mathbf{U}_{l,\mathbf{w}}\mathbf{U}_{l,\mathbf{w}}^{\mathrm{T}}}{\mathbf{P}_{l,\mathbf{w}}}\right)+\mathbf{o}_{P}\left(\frac{1}{\sqrt{nh}}\right)\\
&= 2\sum_{l=1}^{H}\Bigg\{ \left(\widehat{\mathbf{P}}_{l,\mathbf{w}}-\mathbf{P}_{l,\mathbf{w}}\right)\left(\frac{\mathbf{N}_{l,\mathbf{w}}}{\mathbf{P}_{l,\mathbf{w}}}-\boldsymbol{\Sigma}_{\mathbf{w}}\right)^{2}+2\mathbf{N}_{l,\mathbf{w}}\frac{\left(\widehat{\mathbf{N}}_{l,\mathbf{w}}-\mathbf{N}_{l,\mathbf{w}}\right)\mathbf{P}_{l,\mathbf{w}}-\mathbf{N}_{l,\mathbf{w}}\left(\widehat{\mathbf{P}}_{l,\mathbf{w}}-\mathbf{P}_{l,\mathbf{w}}\right)}{\mathbf{P}_{l,\mathbf{w}}^{2}}\\
&-2\mathbf{N}_{l,\mathbf{w}}\left(\widehat{\boldsymbol{\Sigma}}_{\mathbf{w}}-\boldsymbol{\Sigma}_{\mathbf{w}}\right)-2\boldsymbol{\Sigma}_{\mathbf{w}}\frac{\left(\widehat{\mathbf{N}}_{l,\mathbf{w}}-\mathbf{N}_{l,\mathbf{w}}\right)\mathbf{P}_{l,\mathbf{w}}-\mathbf{N}_{l,\mathbf{w}}\left(\widehat{\mathbf{P}}_{l,\mathbf{w}}-\mathbf{P}_{l,\mathbf{w}}\right)}{\mathbf{P}_{l,\mathbf{w}}}+2\mathbf{P}_{l,\mathbf{w}}\boldsymbol{\Sigma}_{\mathbf{w}}\left(\widehat{\boldsymbol{\Sigma}}_{\mathbf{w}}-\boldsymbol{\Sigma}_{\mathbf{w}}\right)\Bigg\}\\
&+2\sum_{l=1}^{H}\left(2\frac{\mathbf{U}_{l,\mathbf{w}}\mathbf{U}_{l,\mathbf{w}}^{\mathrm{T}}}{\mathbf{P}_{l,\mathbf{w}}}+\frac{\mathbf{U}_{l,\mathbf{w}}^{\mathrm{T}}\mathbf{U}_{l,\mathbf{w}}}{\mathbf{P}_{l,\mathbf{w}}}\right)\\
&\times\sum_{l=1}^{H}\frac{\left(\widehat{\mathbf{U}}_{l,\mathbf{w}}-\mathbf{U}_{l,\mathbf{w}}\right)\mathbf{U}_{l,\mathbf{w}}^{\mathrm{T}}\mathbf{P}_{l,\mathbf{w}}+\mathbf{U}_{l,\mathbf{w}}\left(\widehat{\mathbf{U}}_{l,\mathbf{w}}-\mathbf{U}_{l,\mathbf{w}}\right)^{\mathrm{T}}\mathbf{P}_{l,\mathbf{w}}-\mathbf{U}_{l,\mathbf{w}}\mathbf{U}_{l,\mathbf{w}}^{\mathrm{T}}\left(\widehat{\mathbf{P}}_{l,\mathbf{w}}-\mathbf{P}_{l,\mathbf{w}}\right)}{\mathbf{P}_{l,\mathbf{w}}^{2}}\\
&+2\sum_{l=1}^{H}\frac{\left(\widehat{\mathbf{U}}_{l,\mathbf{w}}-\mathbf{U}_{l,\mathbf{w}}\right)^{\mathrm{T}}\mathbf{U}_{l,\mathbf{w}}\mathbf{P}_{l,\mathbf{w}}+\mathbf{U}_{l,\mathbf{w}}^{\mathrm{T}}\left(\widehat{\mathbf{U}}_{l,\mathbf{w}}-\mathbf{U}_{l,\mathbf{w}}\right)\mathbf{P}_{l,\mathbf{w}}-\mathbf{U}_{l,\mathbf{w}}^{\mathrm{T}}\mathbf{U}_{l,\mathbf{w}}\left(\widehat{\mathbf{P}}_{l,\mathbf{w}}-\mathbf{P}_{l,\mathbf{w}}\right)}{\mathbf{P}_{l,\mathbf{w}}^{2}}\\
&\times\left(\sum_{l=1}^{H}\frac{\mathbf{U}_{l,\mathbf{w}}\mathbf{U}_{l,\mathbf{w}}^{\mathrm{T}}}{\mathbf{P}_{l,\mathbf{w}}}\right)+\mathbf{o}_{P}\left(\frac{1}{\sqrt{nh}}\right)=\mathbf{B}_{\mathrm{DR}}+\mathbf{C}_{\mathrm{DR}}+\mathbf{o}_{P}\left(\frac{1}{\sqrt{nh}}\right),
\end{aligned}
$$

where

$$\begin{aligned}\mathbf{B}_{\mathrm{DR}} =&2\sum_{l=1}^{H}\Bigg\{\mathbf{B}_{\mathbf{P},\mathbf{w}}\left(\frac{\mathbf{N}_{l,\mathbf{w}}}{\mathbf{P}_{l,\mathbf{w}}}-\mathbf{\Sigma}_{\mathbf{w}}\right)^{2}+2\mathbf{N}_{l,\mathbf{w}}\frac{\mathbf{B}_{\mathbf{N},\mathbf{w}}\mathbf{P}_{l,\mathbf{w}}-\mathbf{N}_{l,\mathbf{w}}\mathbf{B}_{\mathbf{P},\mathbf{w}}}{\mathbf{P}_{l,\mathbf{w}}^{2}}-2\mathbf{N}_{l,\mathbf{w}}\mathbf{B}_{\mathbf{\Sigma},\mathbf{w}}\\&-2\mathbf{\Sigma}_{\mathbf{w}}\frac{\mathbf{B}_{\mathbf{N},\mathbf{w}}\mathbf{P}_{l,\mathbf{w}}-\mathbf{N}_{l,\mathbf{w}}\mathbf{B}_{\mathbf{P},\mathbf{w}}}{\mathbf{P}_{l,\mathbf{w}}}+2\mathbf{P}_{l,\mathbf{w}}\mathbf{\Sigma}_{\mathbf{w}}\mathbf{B}_{\mathbf{\Sigma},\mathbf{w}}\Bigg\}+2\sum_{l=1}^{H}\left(2\frac{\mathbf{U}_{l,\mathbf{w}}\mathbf{U}_{l,\mathbf{w}}^{\mathrm{T}}}{\mathbf{P}_{l,\mathbf{w}}}+\frac{\mathbf{U}_{l,\mathbf{w}}^{\mathrm{T}}\mathbf{U}_{l,\mathbf{w}}}{\mathbf{P}_{l,\mathbf{w}}}\right)\\&\times\sum_{l=1}^{H}\frac{\mathbf{B}_{\mathbf{U},\mathbf{w}}\mathbf{U}_{l,\mathbf{w}}^{\mathrm{T}}\mathbf{P}_{l,\mathbf{w}}+\mathbf{U}_{l,\mathbf{w}}\mathbf{B}_{\mathbf{U},\mathbf{w}}^{\mathrm{T}}\mathbf{P}_{l,\mathbf{w}}-\mathbf{U}_{l,\mathbf{w}}\mathbf{U}_{l,\mathbf{w}}^{\mathrm{T}}\mathbf{B}_{\mathbf{P},\mathbf{w}}}{\mathbf{P}_{l,\mathbf{w}}^{2}}\\&+2\sum_{l=1}^{H}\frac{\mathbf{B}_{\mathbf{U},\mathbf{w}}^{\mathrm{T}}\mathbf{U}_{l,\mathbf{w}}\mathbf{P}_{l,\mathbf{w}}+\mathbf{U}_{l,\mathbf{w}}^{\mathrm{T}}\mathbf{B}_{\mathbf{U},\mathbf{w}}\mathbf{P}_{l,\mathbf{w}}-\mathbf{U}_{l,\mathbf{w}}^{\mathrm{T}}\mathbf{U}_{l,\mathbf{w}}\mathbf{B}_{\mathbf{P},\mathbf{w}}}{\mathbf{P}_{l,\mathbf{w}}^{2}}\left(\sum_{l=1}^{H}\frac{\mathbf{U}_{l,\mathbf{w}}\mathbf{U}_{l,\mathbf{w}}^{\mathrm{T}}}{\mathbf{P}_{l,\mathbf{w}}}\right)\end{aligned} \tag{S4.38}$$

and

$$\begin{aligned}\mathbf{C}_{\mathrm{DR}} =&2\sum_{l=1}^{H}\Bigg\{\mathbf{C}_{\mathbf{P},\mathbf{w}}\left(\frac{\mathbf{N}_{l,\mathbf{w}}}{\mathbf{P}_{l,\mathbf{w}}}-\mathbf{\Sigma}_{\mathbf{w}}\right)^{2}+2\mathbf{N}_{l,\mathbf{w}}\frac{\mathbf{C}_{\mathbf{N},\mathbf{w}}\mathbf{P}_{l,\mathbf{w}}-\mathbf{N}_{l,\mathbf{w}}\mathbf{C}_{\mathbf{P},\mathbf{w}}}{\mathbf{P}_{l,\mathbf{w}}^{2}}-2\mathbf{N}_{l,\mathbf{w}}\mathbf{C}_{\mathbf{\Sigma},\mathbf{w}}\\&-2\mathbf{\Sigma}_{\mathbf{w}}\frac{\mathbf{C}_{\mathbf{N},\mathbf{w}}\mathbf{P}_{l,\mathbf{w}}-\mathbf{N}_{l,\mathbf{w}}\mathbf{C}_{\mathbf{P},\mathbf{w}}}{\mathbf{P}_{l,\mathbf{w}}}+2\mathbf{P}_{l,\mathbf{w}}\mathbf{\Sigma}_{\mathbf{w}}\mathbf{C}_{\mathbf{\Sigma},\mathbf{w}}\Bigg\}+2\sum_{l=1}^{H}\left(2\frac{\mathbf{U}_{l,\mathbf{w}}\mathbf{U}_{l,\mathbf{w}}^{\mathrm{T}}}{\mathbf{P}_{l,\mathbf{w}}}+\frac{\mathbf{U}_{l,\mathbf{w}}^{\mathrm{T}}\mathbf{U}_{l,\mathbf{w}}}{\mathbf{P}_{l,\mathbf{w}}}\right)\\&\times\sum_{l=1}^{H}\frac{\mathbf{C}_{\mathbf{U},\mathbf{w}}\mathbf{U}_{l,\mathbf{w}}^{\mathrm{T}}\mathbf{P}_{l,\mathbf{w}}+\mathbf{U}_{l,\mathbf{w}}\mathbf{C}_{\mathbf{U},\mathbf{w}}^{\mathrm{T}}\mathbf{P}_{l,\mathbf{w}}-\mathbf{U}_{l,\mathbf{w}}\mathbf{U}_{l,\mathbf{w}}^{\mathrm{T}}\mathbf{C}_{\mathbf{P},\mathbf{w}}}{\mathbf{P}_{l,\mathbf{w}}^{2}}\\&+2\sum_{l=1}^{H}\frac{\mathbf{C}_{\mathbf{U},\mathbf{w}}^{\mathrm{T}}\mathbf{U}_{l,\mathbf{w}}\mathbf{P}_{l,\mathbf{w}}+\mathbf{U}_{l,\mathbf{w}}^{\mathrm{T}}\mathbf{C}_{\mathbf{U},\mathbf{w}}\mathbf{P}_{l,\mathbf{w}}-\mathbf{U}_{l,\mathbf{w}}^{\mathrm{T}}\mathbf{U}_{l,\mathbf{w}}\mathbf{C}_{\mathbf{P},\mathbf{w}}}{\mathbf{P}_{l,\mathbf{w}}^{2}}\left(\sum_{l=1}^{H}\frac{\mathbf{U}_{l,\mathbf{w}}\mathbf{U}_{l,\mathbf{w}}^{\mathrm{T}}}{\mathbf{P}_{l,\mathbf{w}}}\right)\end{aligned}$$

Similar to the proof of Theorem 1 and 3, we have

$$\sqrt{nh}\left(\mathbf{vech}\{\widehat{\mathbf{M}}_{\mathrm{DR}}(\mathbf{w})\}-\mathbf{vech}\{\mathbf{M}_{\mathrm{DR}}(\mathbf{w})\}-\mathbf{vech}\{\mathbf{B}_{\mathrm{DR}}(\mathbf{w})\}\right)\xrightarrow{d}\mathbf{N}(0,f^{-1}(\mathbf{w})\omega_0\mathbf{C}^{\mathrm{DR}}(\mathbf{w})), \tag{S4.39}$$

where ω_0 is defined as previously. Hence

$$\mathbf{C}^{\mathrm{DR}}(\mathbf{w})=\mathrm{Cov}[\mathbf{vech}\{\mathbf{C}_{\mathrm{DR}}(\mathbf{w})\}|\mathbf{w}]. \tag{S4.40}$$

Then for singular value decomposition, in partial variable dependent DR, denote $\mathbf{G}(\mathbf{w})=\mathbf{\Sigma}_{\mathbf{w}}^{-1}\mathbf{M}_{\mathrm{DR}}(\mathbf{w})$, then substitute it into (S2.20) and (S2.21),

$$\begin{aligned}\mathbf{B}_k^{\mathrm{DR}}(\mathbf{w})=\sum_{j\neq k}\frac{\boldsymbol{\beta}_j^{\mathrm{DR}}(\mathbf{w})\{\boldsymbol{\beta}_j^{\mathrm{DR}}(\mathbf{w})\}^{\mathrm{T}}}{\{\lambda_j^{\mathrm{DR}}(\mathbf{w})\}^2-\{\lambda_k^{\mathrm{DR}}(\mathbf{w})\}^2}&\Big[\mathbf{M}_{\mathrm{DR}}^{\mathrm{T}}(\mathbf{w})\mathbf{\Sigma}_{\mathbf{w}}^{-1}\mathbf{B}_{\mathbf{\Sigma},\mathbf{w}}\mathbf{\Sigma}_{\mathbf{w}}^{-1}\mathbf{\Sigma}_{\mathbf{w}}^{-1}\mathbf{M}_{\mathrm{DR}}(\mathbf{w})\\&-\mathbf{B}_{\mathrm{DR}}^{\mathrm{T}}(\mathbf{w})\mathbf{\Sigma}_{\mathbf{w}}^{-1}\mathbf{\Sigma}_{\mathbf{w}}^{-1}\mathbf{M}_{\mathrm{DR}}(\mathbf{w})+\{\mathbf{\Sigma}_{\mathbf{w}}^{-1}\mathbf{M}_{\mathrm{DR}}(\mathbf{w})\}^{\mathrm{T}}\mathbf{\Sigma}_{\mathbf{w}}^{-1}\mathbf{B}_{\mathbf{\Sigma},\mathbf{w}}\mathbf{\Sigma}_{\mathbf{w}}^{-1}\mathbf{M}_{\mathrm{DR}}(\mathbf{w})\\&-\{\mathbf{\Sigma}_{\mathbf{w}}^{-1}\mathbf{M}_{\mathrm{DR}}(\mathbf{w})\}^{\mathrm{T}}\mathbf{\Sigma}_{\mathbf{w}}^{-1}\mathbf{B}_{\mathrm{DR}}(\mathbf{w})\Big]\boldsymbol{\beta}_k^{\mathrm{DR}}(\mathbf{w}),\end{aligned} \tag{S4.41}$$

$$\begin{aligned}\mathbf{C}_k^{\mathrm{DR}}(\mathbf{w})=\sum_{j\neq k}\frac{\boldsymbol{\beta}_j^{\mathrm{DR}}(\mathbf{w})\{\boldsymbol{\beta}_j^{\mathrm{DR}}(\mathbf{w})\}^{\mathrm{T}}}{\{\lambda_j^{\mathrm{DR}}(\mathbf{w})\}^2-\{\lambda_k^{\mathrm{DR}}(\mathbf{w})\}^2}&\Big[\mathbf{M}_{\mathrm{DR}}^{\mathrm{T}}(\mathbf{w})\mathbf{\Sigma}_{\mathbf{w}}^{-1}\mathbf{C}_{\mathbf{\Sigma},\mathbf{w}}\mathbf{\Sigma}_{\mathbf{w}}^{-1}\mathbf{\Sigma}_{\mathbf{w}}^{-1}\mathbf{M}_{\mathrm{DR}}(\mathbf{w})\\&-\mathbf{C}_{\mathrm{DR}}(\mathbf{w})^{\mathrm{T}}\mathbf{\Sigma}_{\mathbf{w}}^{-1}\mathbf{\Sigma}_{\mathbf{w}}^{-1}\mathbf{M}_{\mathrm{DR}}(\mathbf{w})+\{\mathbf{\Sigma}_{\mathbf{w}}^{-1}\mathbf{M}_{\mathrm{DR}}(\mathbf{w})\}^{\mathrm{T}}\mathbf{\Sigma}_{\mathbf{w}}^{-1}\mathbf{C}_{\mathbf{\Sigma},\mathbf{w}}\mathbf{\Sigma}_{\mathbf{w}}^{-1}\mathbf{M}_{\mathrm{DR}}(\mathbf{w})\\&-\{\mathbf{\Sigma}_{\mathbf{w}}^{-1}\mathbf{M}_{\mathrm{DR}}(\mathbf{w})\}^{\mathrm{T}}\mathbf{\Sigma}_{\mathbf{w}}^{-1}\mathbf{C}_{\mathrm{DR}}(\mathbf{w})\Big]\boldsymbol{\beta}_k^{\mathrm{DR}}(\mathbf{w}).\end{aligned}$$

Hence,

$$\boldsymbol{\Sigma}_k^{\mathrm{DR}}(\mathbf{w}) = \mathrm{Cov}\{\mathbf{C}_k^{\mathrm{DR}}(\mathbf{w})|\mathbf{w}\} \tag{S4.42}$$

$\square$

S5. The Bandwidth Selection for Variable Dependent Partial SAVE and DR

For variable dependent partial SAVE, since Cook and Yin (2001) has shown that SAVE is closely related to quadratic discriminant analysis, the bandwidth selection can be conducted by first minimizing (S5.43), assuming that $\mathbf{X}|(\widetilde{Y} = l, \mathbf{W} = \mathbf{w}) \sim N(\mathbf{m}_l(\mathbf{w}), \boldsymbol{\Sigma}_{l\mathbf{w}})$, where $\boldsymbol{\Sigma}_{l\mathbf{w}}$ is different in every slice,

$$CV(h_l) = \frac{1}{n_l}\sum_{i=1}^{n_l}\left[\{\mathbf{x}_i - \widehat{\mathbf{m}}_l^{(-i)}(\mathbf{w})\}^{\mathrm{T}}\widehat{\boldsymbol{\Sigma}}_{l\mathbf{w}(-i)}^{-1}(\mathbf{w})\{\mathbf{x}_i - \widehat{\mathbf{m}}_l^{(-i)}(\mathbf{w})\} + \log|\widehat{\boldsymbol{\Sigma}}_{l\mathbf{w}(-i)}(\mathbf{w})|\right]. \tag{S5.43}$$

The optimal bandwidth h_{opt} for variable dependent partial SAVE is then selected by choosing the value of h_l which maximizes (1.11). For variable dependent partial DR, we use the same bandwidth selection procedure as variable dependent partial SAVE since DR synthesizes the dimension reduction methods based on the first two conditional moments.

References

[1] Bickel, P. J., and Freedman, D. A. (1981) Some asymptotic theory for the bootstrap. *The Annals of Statistics*, **9** 1196-1217.

[2] Cook, R., and Hakbae Lee. (1999) Dimension reduction in binary response regression. *Journal of the American Statistical Association*, **94(448)** 1187-1200.

[3] Cook, R. D., and Ni, L. (2005) Sufficient dimension reduction via inverse regression: A minimum discrepancy approach. *Journal of the American Statistical Association*, **100(470)** 410-428.

[4] Cook, R. D., and Weisberg, S. (1991) Comment. *Journal of the American Statistical Association*, **86(414)** 328-332.

[5] Cook, R. D., and Yin, X. (2001) Theory and methods: Special invited paper: Dimension reduction and visualization in discriminant analysis (with discussion). *Australian and New Zealand Journal of Statistics*, **43(2)** 147-199.

[6] Fan, J. (1993) Local linear regression smoothers and their minimax efficiencies. *The Annals of Statistics*, 196-216.

[7] Hardle, W. (1984) Robust regression function estimation. *Journal of Multivariate Analysis*, **14(2)** 169-180.

[8] Li, B., Cook, R. D., and Chiaromonte, F. (2003) Dimension reduction for the conditional mean in regressions with categorical predictors. *The Annals of Statistics*, 1636-1668.

[9] Li, B., and Wang, S. (2007) On directional regression for dimension reduction. *Journal of the American Statistical Association*, **102(479)** 997-1008.

[10] Luo, W., and Li, B. (2016) Combining eigenvalues and variation of eigenvectors for order determination. *Biometrika*, **103(4)** 875-887.

[11] Shao, Y., Cook, R. D., and Weisberg, S. (2009) Partial central subspace and sliced average variance estimation. *Journal of Statistical Planning and Inference*, **139(3)** 952-961.

[12] Yao, Q. and Tong, H. (1996) Asymmetric least squares regression estimation: a nonparametric approach. *Journal of Nonparametric Statistics*, **6(2-3)** 273-292.

[13] Yao, Q., and Tong, H. (1998) Cross-validatory bandwidth selections

for regression estimation based on dependent data. *Journal of Statistical Planning and Inference*, **68(2)** 387-415.

[14] Yin, J., Geng, Z., Li, R., and Wang, H. (2010) Nonparametric covariance model. *Statistica Sinica*, **20** 469.

[15] Yu, Z., Dong, Y. and Huang, M. (2014) General directional regression. *Journal of Multivariate Analysis*, **124** 94-104.

[16] Yu, Z., and Dong, Y. (2016) Model-free coordinate test and variable selection via directional regressionl. *Statistica Sinica*, 1159-1174.

[17] Zhao, L. C., Krishnaiah, P. R., and Bai, Z. D. (1986) On detection of the number of signals in presence of white noise. *Academic Press*, Inc.