# 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 $\delta$ 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$\acute{e}$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,

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

□

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

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

□

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

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

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

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

and

$$\mathbf{R}_3(\mathbf{w}) = \frac{1}{nf(\mathbf{w})}\left\{\left|\sum_{i=1}^{n} K\left(\frac{\mathbf{w}_i - \mathbf{w}}{h}\right)(\mathbf{x}_i - \mathbf{m}(\mathbf{w}_i))\right|\right\} + \mathbf{o}(h^2),$$

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

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

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

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

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

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

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

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

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

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

$$= \{\widehat{\lambda}_k(\mathbf{w}) - \lambda_k(\mathbf{w})\}\lambda_k(\mathbf{w}) + \lambda_k\{\widehat{\lambda}_k(\mathbf{w}) - \lambda_k(\mathbf{w})\} + \mathbf{o}_p(\frac{1}{\sqrt{nh}}),$$

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 $(\boldsymbol{\beta}_1(\mathbf{w}), \ldots, \boldsymbol{\beta}_p(\mathbf{w}))$ is a basis of $\mathbb{R}^p$, then there exists $c_{kj}^*$ for $j = 1, \ldots, p$, 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

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

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 $\mathbf{G}(\mathbf{w})\mathbf{G}(\mathbf{w})^{T} \in \mathbb{R}^{p\times p}$ 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* $\mathbf{G}(\mathbf{w})\mathbf{G}(\mathbf{w})^{\tau} \in \mathbb{R}^{p \times p}$ *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* $i \neq j$*,*

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

*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

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

and

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

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

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

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 $i \neq j$,

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

□

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), $\text{pr}(A_1) = 1$. 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 $\text{pr}(A_2) = 1$. Hence $\text{pr}(A_1 \cap A_2) = 1$. 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 $\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\mathbf{M}_{\mathrm{SAVE}}(\mathbf{w}) \subseteq \mathcal{S}_{Y_{\mathbf{w}}|\mathbf{X}_{\mathbf{w}}}$ 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

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

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 $\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\mathbf{M}_{\mathrm{SAVE}}(\mathbf{w})$. 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

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

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

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

thus

$$\widehat{\mathbf{M}}_{\text{SAVE}}(\mathbf{w}) - \mathbf{M}_{\text{SAVE}}(\mathbf{w}) = \sum_{l=1}^{H} \left(\widehat{\mathbf{P}}_{l,\mathbf{w}}\widehat{\mathbf{E}}^2_{l,\mathbf{w}} - \mathbf{P}_{l,\mathbf{w}}\mathbf{E}^2_{l,\mathbf{w}}\right)$$

$$= \sum_{l=1}^{H} \left\{ \left(\widehat{\mathbf{P}}_{l,\mathbf{w}} - \mathbf{P}_{l,\mathbf{w}}\right) \widehat{\mathbf{E}}^2_{l,\mathbf{w}} + \mathbf{P}_{l,\mathbf{w}} \left(\widehat{\mathbf{E}}^2_{l,\mathbf{w}} - \mathbf{E}^2_{l,\mathbf{w}}\right) \right\}$$

$$= \sum_{l=1}^{H} \left(\widehat{\mathbf{P}}_{l,\mathbf{w}} - \mathbf{P}_{l,\mathbf{w}}\right) \mathbf{E}^2_{l,\mathbf{w}} + \sum_{l=1}^{H} \mathbf{P}_{l,\mathbf{w}} \left(\widehat{\mathbf{E}}_{l,\mathbf{w}} - \mathbf{E}_{l,\mathbf{w}}\right) \times \left(\widehat{\mathbf{E}}_{l,\mathbf{w}} + \mathbf{E}_{l,\mathbf{w}}\right) + \mathbf{o}_P\left(\frac{1}{\sqrt{nh}}\right)$$

$$= \sum_{l=1}^{H} \left(\widehat{\mathbf{P}}_{l,\mathbf{w}} - \mathbf{P}_{l,\mathbf{w}}\right) \mathbf{E}^2_{l,\mathbf{w}} + \sum_{l=1}^{H} 2\mathbf{P}_{l,\mathbf{w}} \Bigg\{ \left(\widehat{\boldsymbol{\Sigma}}_{\mathbf{w}} - \boldsymbol{\Sigma}_{\mathbf{w}}\right) \boldsymbol{\Sigma}_{\mathbf{w}}$$

$$+ \frac{\mathbf{N}_{l,\mathbf{w}}\left(\widehat{\mathbf{P}}_{l,\mathbf{w}} - \mathbf{P}_{l,\mathbf{w}}\right) - \left(\widehat{\mathbf{N}}_{l,\mathbf{w}} - \mathbf{N}_{l,\mathbf{w}}\right)\mathbf{P}_{l,\mathbf{w}}}{\mathbf{P}^2_{l,\mathbf{w}}} \boldsymbol{\Sigma}_{\mathbf{w}}$$

$$+ \frac{\left(\widehat{\mathbf{U}}_{l,\mathbf{w}} - \mathbf{U}_{l,\mathbf{w}}\right)\mathbf{U}^{\mathrm{T}}_{l,\mathbf{w}}\mathbf{P}_{l,\mathbf{w}} + \mathbf{U}_{l,\mathbf{w}}\left(\widehat{\mathbf{U}}_{l,\mathbf{w}} - \mathbf{U}_{l,\mathbf{w}}\right)^{\mathrm{T}}\mathbf{P}_{l,\mathbf{w}} - 2\mathbf{U}_{l,\mathbf{w}}\mathbf{U}^{\mathrm{T}}_{l,\mathbf{w}}\left(\widehat{\mathbf{P}}_{l,\mathbf{w}} - \mathbf{P}_{l,\mathbf{w}}\right)}{\mathbf{P}^3_{l,\mathbf{w}}} \boldsymbol{\Sigma}_{\mathbf{w}}$$

$$- \left(\widehat{\boldsymbol{\Sigma}}_{\mathbf{w}} - \boldsymbol{\Sigma}_{\mathbf{w}}\right)\frac{\mathbf{N}_{l,\mathbf{w}}}{\mathbf{P}_{l,\mathbf{w}}} - \frac{\mathbf{N}_{l,\mathbf{w}}\left(\widehat{\mathbf{P}}_{l,\mathbf{w}} - \mathbf{P}_{l,\mathbf{w}}\right) - \left(\widehat{\mathbf{N}}_{l,\mathbf{w}} - \mathbf{N}_{l,\mathbf{w}}\right)\mathbf{P}_{l,\mathbf{w}}}{\mathbf{P}^2_{l,\mathbf{w}}} \frac{\mathbf{N}_{l,\mathbf{w}}}{\mathbf{P}_{l,\mathbf{w}}}$$

$$- \frac{\left(\widehat{\mathbf{U}}_{l,\mathbf{w}} - \mathbf{U}_{l,\mathbf{w}}\right)\mathbf{U}^{\mathrm{T}}_{l,\mathbf{w}}\mathbf{P}_{l,\mathbf{w}} + \mathbf{U}_{l,\mathbf{w}}\left(\widehat{\mathbf{U}}_{l,\mathbf{w}} - \mathbf{U}_{l,\mathbf{w}}\right)^{\mathrm{T}}\mathbf{P}_{l,\mathbf{w}} - 2\mathbf{U}_{l,\mathbf{w}}\mathbf{U}^{\mathrm{T}}_{l,\mathbf{w}}\left(\widehat{\mathbf{P}}_{l,\mathbf{w}} - \mathbf{P}_{l,\mathbf{w}}\right)}{\mathbf{P}^3_{l,\mathbf{w}}} \frac{\mathbf{N}_{l,\mathbf{w}}}{\mathbf{P}_{l,\mathbf{w}}}$$

$$+ \left(\widehat{\boldsymbol{\Sigma}}_{\mathbf{w}} - \boldsymbol{\Sigma}_{\mathbf{w}}\right)\frac{\mathbf{U}_{l,\mathbf{w}}\mathbf{U}^{\mathrm{T}}_{l,\mathbf{w}}}{\mathbf{P}^2_{l,\mathbf{w}}} + \frac{\mathbf{N}_{l,\mathbf{w}}\left(\widehat{\mathbf{P}}_{l,\mathbf{w}} - \mathbf{P}_{l,\mathbf{w}}\right) - \left(\widehat{\mathbf{N}}_{l,\mathbf{w}} - \mathbf{N}_{l,\mathbf{w}}\right)\mathbf{P}_{l,\mathbf{w}}}{\mathbf{P}^2_{l,\mathbf{w}}} \frac{\mathbf{U}_{l,\mathbf{w}}\mathbf{U}^{\mathrm{T}}_{l,\mathbf{w}}}{\mathbf{P}^2_{l,\mathbf{w}}}$$

$$+ \frac{\left(\widehat{\mathbf{U}}_{l,\mathbf{w}} - \mathbf{U}_{l,\mathbf{w}}\right)\mathbf{U}^{\mathrm{T}}_{l,\mathbf{w}}\mathbf{P}_{l,\mathbf{w}} + \mathbf{U}_{l,\mathbf{w}}\left(\widehat{\mathbf{U}}_{l,\mathbf{w}} - \mathbf{U}_{l,\mathbf{w}}\right)^{\mathrm{T}}\mathbf{P}_{l,\mathbf{w}} - 2\mathbf{U}_{l,\mathbf{w}}\mathbf{U}^{\mathrm{T}}_{l,\mathbf{w}}\left(\widehat{\mathbf{P}}_{l,\mathbf{w}} - \mathbf{P}_{l,\mathbf{w}}\right)}{\mathbf{P}^3_{l,\mathbf{w}}} \frac{\mathbf{U}_{l,\mathbf{w}}\mathbf{U}^{\mathrm{T}}_{l,\mathbf{w}}}{\mathbf{P}^2_{l,\mathbf{w}}} \Bigg\}$$

$$+ \mathbf{o}_P\left(\frac{1}{\sqrt{nh}}\right) = \mathbf{B}_{\text{SAVE}} + \mathbf{C}_{\text{SAVE}} + \mathbf{o}_P\left(\frac{1}{\sqrt{nh}}\right),$$

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

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

where $\omega_0$ is defined as follows. Hence

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

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

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

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

Hence,

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

□

## 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,

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

Where $\mathrm{E}_{Y,\breve{Y}}$ 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

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

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}}} - \mathbf{\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{\mathbf{\Sigma}}_{\mathbf{w}} + \frac{\mathbf{N}_{l,\mathbf{w}}}{\mathbf{P}_{l,\mathbf{w}}} - \mathbf{\Sigma}_{\mathbf{w}}\right) \\
&\times \left(\frac{\widehat{\mathbf{N}}_{l,\mathbf{w}}}{\widehat{\mathbf{P}}_{l,\mathbf{w}}} - \widehat{\mathbf{\Sigma}}_{\mathbf{w}} - \frac{\mathbf{N}_{l,\mathbf{w}}}{\mathbf{P}_{l,\mathbf{w}}} + \mathbf{\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}}} - \mathbf{\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{\mathbf{\Sigma}}_{\mathbf{w}} - \mathbf{\Sigma}_{\mathbf{w}}\right) \\
&- 2\mathbf{P}_{l,\mathbf{w}}\mathbf{\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}}\mathbf{\Sigma}_{\mathbf{w}}\left(\widehat{\mathbf{\Sigma}}_{\mathbf{w}} - \mathbf{\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}}} - \mathbf{\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{\mathbf{\Sigma}}_{\mathbf{w}} - \mathbf{\Sigma}_{\mathbf{w}}\right) - 2\mathbf{\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}}\mathbf{\Sigma}_{\mathbf{w}}\left(\widehat{\mathbf{\Sigma}}_{\mathbf{w}} - \mathbf{\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 $\omega_0$ 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,

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

□

## 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.