A METHOD OF LOCAL INFLUENCE ANALYSIS IN SUFFICIENT DIMENSION REDUCTION

Fei Chen[1], Lei Shi[1], Lin Zhu[1,2] and Lixing Zhu[3*]

[1] *Yunnan University of Finance and Economics,*

[2] *Tiantai County Branch of the People's Bank of China,*

[3] *Hong Kong Baptist University*

Supplementary Material

S1 Proof of the equality (3.3)

Under assumption 2, it can be easily shown that $\hat{\mathbf{B}}^\top\mathbf{Z}\mathbf{Z}^T\hat{\mathbf{B}}$ is invertible and for any given $\mathbf{h}$, the matrix $\hat{\mathbf{B}}^\top\mathbf{Z}\mathbf{Z}^T\hat{\mathbf{B}}$ also has full rank in a neighborhood of $t = 0$. Then we have $D(\boldsymbol{\omega}) = d(\hat{\mathbf{B}}(\boldsymbol{\omega}))$. Firstly, we expand $d(\mathbf{A})$ at $\mathbf{A} = \hat{\mathbf{B}}$:

$$\begin{aligned} d(\mathbf{A}) &= d(\hat{\mathbf{B}}) + \left\{ \frac{\partial d(\mathbf{A})}{\partial \mathrm{vec}(\mathbf{A})}\bigg|_{\mathbf{A}=\hat{\mathbf{B}}} \right\}^{\mathrm{T}} \mathrm{vec}(\mathbf{A}-\hat{\mathbf{B}}) + \frac{1}{2}\mathrm{vec}(\mathbf{A}-\hat{\mathbf{B}})^{\mathrm{T}} \\ &\quad \frac{\partial^2 d(\mathbf{A})}{\partial \mathrm{vec}(\mathbf{A})\partial \mathrm{vec}(\mathbf{A})^{\mathrm{T}}}\bigg|_{\mathbf{A}=\hat{\mathbf{B}}} \mathrm{vec}(\mathbf{A}-\hat{\mathbf{B}}) + o[\mathrm{tr}\{(\mathbf{A}-\hat{\mathbf{B}})^{\mathrm{T}}(\mathbf{A}-\hat{\mathbf{B}})\}]. \end{aligned}$$

From $d(\hat{\mathbf{B}}) = 0$ and $0 \le d(\mathbf{A}) \le 1$, we know $\hat{\mathbf{B}}$ minimizes $d(\mathbf{A})$ and

$\partial d(\mathbf{A})/\partial \mathrm{vec}(\mathbf{A})|_{\mathbf{A}=\hat{\mathbf{B}}} = 0$. Hence,

$$\begin{aligned} d(\mathbf{A}) &= \frac{1}{2}\mathrm{vec}(\mathbf{A}-\hat{\mathbf{B}})^{\mathrm{T}} \left.\frac{\partial^2 d(\mathbf{A})}{\partial \mathrm{vec}(\mathbf{A})\partial \mathrm{vec}(\mathbf{A})^{\mathrm{T}}}\right|_{\mathbf{A}=\hat{\mathbf{B}}} \mathrm{vec}(\mathbf{A}-\hat{\mathbf{B}}) \\ &\quad + o[\mathrm{tr}\{(\mathbf{A}-\hat{\mathbf{B}})^{\mathrm{T}}(\mathbf{A}-\hat{\mathbf{B}})\}]. \end{aligned} \tag{S1.1}$$

As $D(\boldsymbol{\omega}_{(0)}+t\mathbf{h}) = d\{\hat{\mathbf{B}}(\boldsymbol{\omega}_{(0)}+t\mathbf{h})\}$,substituting $\mathbf{A}$ in (S1.1) by $\hat{\mathbf{B}}(\boldsymbol{\omega}_{(0)}+t\mathbf{h})$ under assumption 2 gives

$$\begin{aligned} D(\boldsymbol{\omega}_{(0)}+t\mathbf{h}) &= \frac{1}{2}\mathrm{vec}\{t\mathbf{F}_{\mathbf{B},\mathbf{h}}+o(t)\}^{\mathrm{T}} \left.\frac{\partial^2 d(\mathbf{A})}{\partial \mathrm{vec}(\mathbf{A})\partial \mathrm{vec}(\mathbf{A})^{\mathrm{T}}}\right|_{\mathbf{A}=\hat{\mathbf{B}}} \mathrm{vec}\{t\mathbf{F}_{\mathbf{B},\mathbf{h}} \\ &\quad + o(t)\} + o(\mathrm{tr}[\{t\mathbf{F}_{\mathbf{B},\mathbf{h}}+o(t))^{\mathrm{T}}(t\mathbf{F}_{\mathbf{B},\mathbf{h}}+o(t)\}]). \end{aligned}$$

Then the equality (3.3) holds obviously and its proof has been finished.

S2 Proof of Lemma 1

Let $\hat{\mathbf{V}} = \mathbf{Z}^{\mathrm{T}}\hat{\mathbf{B}}$, $\hat{\mathbf{H}} = \mathbf{P}_{\mathbf{Z}^{\mathrm{T}}\hat{\mathbf{B}}}$, $\tilde{\mathbf{V}} = \mathbf{Z}^{\mathrm{T}}\mathbf{A}$ and $\tilde{\mathbf{H}} = \tilde{\mathbf{V}}(\tilde{\mathbf{V}}^{\mathrm{T}}\tilde{\mathbf{V}})^{-1}\tilde{\mathbf{V}}^{\mathrm{T}}$. To simplify the proof, we employ some matrix differentiation techniques. Now we define some notations. For any vectors $\boldsymbol{\xi}_1$ and $\boldsymbol{\xi}_2$, let $\partial\boldsymbol{\xi}_1/\partial\boldsymbol{\xi}_2$ denote the matrix with its (i,j)th element $\partial\xi_{1,j}/\partial\xi_{2,i}$ where $\xi_{k,j}$ denotes the jth element of ξ_k. For any matrix $\boldsymbol{\Xi} = (\boldsymbol{\xi}_1,\ldots,\boldsymbol{\xi}_m)^{\mathrm{T}}$ with $\boldsymbol{\xi}_i$ the ith row vector of $\boldsymbol{\Xi}$, let $\mathrm{vec}_{\mathrm{R}}(\boldsymbol{\Xi}) = (\boldsymbol{\xi}_1^{\mathrm{T}},\boldsymbol{\xi}_2^{\mathrm{T}},\ldots\boldsymbol{\xi}_m^{\mathrm{T}})^{\mathrm{T}}$. For any matrices $\boldsymbol{\Xi}$ and $\boldsymbol{\Gamma}$, let $\partial\boldsymbol{\Xi}/\partial\boldsymbol{\Gamma} = \partial\mathrm{vec}_{\mathrm{R}}(\boldsymbol{\Xi})/\partial\mathrm{vec}_{\mathrm{R}}(\boldsymbol{\Gamma})$.

(i) We first calculate $\partial d(\mathbf{A})/\partial\mathbf{A}$. Note that $d(\mathbf{A}) = 1-\hat{K}^{-1}\mathrm{tr}(\hat{\mathbf{H}}\tilde{\mathbf{H}})$.

Then we have

$$\frac{\partial d(\mathbf{A})}{\partial \mathbf{A}} = -\hat{K}^{-1}\frac{\partial \tilde{\mathbf{H}}}{\partial \mathbf{A}}\mathrm{vec}_{\mathrm{R}}(\hat{\mathbf{H}}) = -\hat{K}^{-1}\frac{\partial \tilde{\mathbf{V}}}{\partial \mathbf{A}}\frac{\partial \tilde{\mathbf{H}}}{\partial \tilde{\mathbf{V}}}\mathrm{vec}_{\mathrm{R}}(\hat{\mathbf{H}}), \tag{S2.1}$$

where $\partial \tilde{\mathbf{V}}/\partial \mathbf{A} = \mathbf{Z} \otimes \mathbf{I}_{\hat{K}}$. The expressions of $\partial \tilde{\mathbf{H}}/\partial \mathbf{A}$ can be derived as follows. Firstly,

$$\begin{aligned}
\frac{\partial \tilde{\mathbf{H}}}{\partial \tilde{\mathbf{V}}} &= \frac{\partial \tilde{\mathbf{V}}}{\partial \tilde{\mathbf{V}}}[\mathbf{I}_n \otimes \{(\tilde{\mathbf{V}}^{\mathrm{T}}\tilde{\mathbf{V}})^{-1}\tilde{\mathbf{V}}^{\mathrm{T}}\}] + \frac{\partial (\tilde{\mathbf{V}}^{\mathrm{T}}\tilde{\mathbf{V}})^{-1}\tilde{\mathbf{V}}^{\mathrm{T}}}{\partial \tilde{\mathbf{V}}}(\tilde{\mathbf{V}}^{\mathrm{T}} \otimes \mathbf{I}_n) \\
&= \mathbf{I}_n \otimes \{(\tilde{\mathbf{V}}^{\mathrm{T}}\tilde{\mathbf{V}})^{-1}\tilde{\mathbf{V}}^{\mathrm{T}}\} + \frac{\partial (\tilde{\mathbf{V}}^{\mathrm{T}}\tilde{\mathbf{V}})^{-1}}{\partial \tilde{\mathbf{V}}}(\tilde{\mathbf{V}}^{\mathrm{T}} \otimes \tilde{\mathbf{V}}^{\mathrm{T}}) \\
&\quad + \frac{\partial \tilde{\mathbf{V}}^{\mathrm{T}}}{\partial \tilde{\mathbf{V}}}[\{(\tilde{\mathbf{V}}^{\mathrm{T}}\tilde{\mathbf{V}})^{-1}\tilde{\mathbf{V}}^{\mathrm{T}}\} \otimes \mathbf{I}_n].
\end{aligned} \tag{S2.2}$$

Moreover,

$$\frac{\partial (\tilde{\mathbf{V}}^{\mathrm{T}}\tilde{\mathbf{V}})^{-1}}{\partial \tilde{\mathbf{V}}} = -\frac{\partial \tilde{\mathbf{V}}^{\mathrm{T}}\tilde{\mathbf{V}}}{\partial \tilde{\mathbf{V}}}\{(\tilde{\mathbf{V}}^{\mathrm{T}}\tilde{\mathbf{V}})^{-1} \otimes (\tilde{\mathbf{V}}^{\mathrm{T}}\tilde{\mathbf{V}})^{-1}\},$$

and then we have

$$\begin{aligned}
\frac{\partial (\tilde{\mathbf{V}}^{\mathrm{T}}\tilde{\mathbf{V}})^{-1}}{\partial \tilde{\mathbf{V}}} &= -\frac{\partial \tilde{\mathbf{V}}^{\mathrm{T}}}{\partial \tilde{\mathbf{V}}}[(\tilde{\mathbf{V}}^{\mathrm{T}}\tilde{\mathbf{V}})^{-1} \otimes \{\tilde{\mathbf{V}}(\tilde{\mathbf{V}}^{\mathrm{T}}\tilde{\mathbf{V}})^{-1}\}] - \{\tilde{\mathbf{V}}(\tilde{\mathbf{V}}^{\mathrm{T}}\tilde{\mathbf{V}})^{-1}\} \\
&\quad \otimes (\tilde{\mathbf{V}}^{\mathrm{T}}\tilde{\mathbf{V}})^{-1}.
\end{aligned} \tag{S2.3}$$

Combining (S2.2) and (S2.3), we have

$$\begin{aligned}
\frac{\partial \tilde{\mathbf{H}}}{\partial \mathbf{A}} &= (\mathbf{Z} \otimes \mathbf{I}_{\hat{K}})\Big(\mathbf{I}_n \otimes \{(\tilde{\mathbf{V}}^{\mathrm{T}}\tilde{\mathbf{V}})^{-1}\tilde{\mathbf{V}}^{\mathrm{T}}\} - \tilde{\mathbf{H}} \otimes \{(\tilde{\mathbf{V}}^{\mathrm{T}}\tilde{\mathbf{V}})^{-1}\tilde{\mathbf{V}}^{\mathrm{T}}\} \\
&\quad - \frac{\partial \tilde{\mathbf{V}}^{\mathrm{T}}}{\partial \tilde{\mathbf{V}}}[\{(\tilde{\mathbf{V}}^{\mathrm{T}}\tilde{\mathbf{V}})^{-1}\tilde{\mathbf{V}}^{\mathrm{T}}\} \otimes \tilde{\mathbf{H}}] \\
&\quad + \frac{\partial \tilde{\mathbf{V}}^{\mathrm{T}}}{\partial \tilde{\mathbf{V}}}\,[\{(\tilde{\mathbf{V}}^{\mathrm{T}}\tilde{\mathbf{V}})^{-1}\tilde{\mathbf{V}}^{\mathrm{T}}\} \otimes \mathbf{I}_n]\Big).
\end{aligned} \tag{S2.4}$$

Note that

$$(\boldsymbol{\Xi} \otimes \boldsymbol{\Gamma}^{\mathrm{T}})\mathrm{vec}_{\mathrm{R}}(\boldsymbol{\Psi}) = \mathrm{vec}_{\mathrm{R}}(\boldsymbol{\Xi}\boldsymbol{\Psi}\boldsymbol{\Gamma})$$

for any matrices $\boldsymbol{\Xi}$, $\boldsymbol{\Gamma}$ and $\boldsymbol{\Psi}$ of appropriate dimensions. Hence, combining (S2.1) and (S2.4) gives

$$\begin{aligned}\frac{\partial d(\mathbf{A})}{\partial \mathbf{A}} &= -\hat{K}^{-1}(\mathbf{Z} \otimes \mathbf{I}_{\hat{K}}) \Big[\mathrm{vec}_{\mathrm{R}} \Big\{\hat{\mathbf{H}}\tilde{\mathbf{V}}(\tilde{\mathbf{V}}^{\mathrm{T}}\tilde{\mathbf{V}})^{-1} - \tilde{\mathbf{H}}\hat{\mathbf{H}}\tilde{\mathbf{V}}(\tilde{\mathbf{V}}^{\mathrm{T}}\tilde{\mathbf{V}})^{-1}\Big\} \\ &\quad - \frac{\partial \tilde{\mathbf{V}}^{\mathrm{T}}}{\partial \tilde{\mathbf{V}}}\mathrm{vec}_{\mathrm{R}} \Big\{(\tilde{\mathbf{V}}^{\mathrm{T}}\tilde{\mathbf{V}})^{-1}\tilde{\mathbf{V}}^{\mathrm{T}}\hat{\mathbf{H}}\tilde{\mathbf{H}} - (\tilde{\mathbf{V}}^{\mathrm{T}}\tilde{\mathbf{V}})^{-1}\tilde{\mathbf{V}}^{\mathrm{T}}\hat{\mathbf{H}}\Big\}\Big]. \qquad \text{(S2.5)}\end{aligned}$$

It can be shown that

$$\frac{\partial \boldsymbol{\Xi}^{\mathrm{T}}}{\partial \boldsymbol{\Xi}}\mathrm{vec}_{\mathrm{R}}(\boldsymbol{\Psi}) = \mathrm{vec}_{\mathrm{R}}(\boldsymbol{\Psi}^{\mathrm{T}})$$

for any matrices $\boldsymbol{\Xi}_{k \times l}$ and $\boldsymbol{\Psi}_{l \times k}$, and so (S2.5) can be written as

$$\begin{aligned}\frac{\partial d(\mathbf{A})}{\partial \mathbf{A}} &= -2\hat{K}^{-1}(\mathbf{Z} \otimes \mathbf{I}_{\hat{K}})\mathrm{vec}_{\mathrm{R}} \Big\{\hat{\mathbf{H}}\tilde{\mathbf{V}}(\tilde{\mathbf{V}}^{\mathrm{T}}\tilde{\mathbf{V}})^{-1} - \tilde{\mathbf{H}}\hat{\mathbf{H}}\tilde{\mathbf{V}}(\tilde{\mathbf{V}}^{\mathrm{T}}\tilde{\mathbf{V}})^{-1}\Big\} \\ &= -2\hat{K}^{-1}\mathrm{vec}_{\mathrm{R}} \Big\{\mathbf{Z}(\mathbf{I} - \tilde{\mathbf{H}})\hat{\mathbf{H}}\tilde{\mathbf{V}}(\tilde{\mathbf{V}}^{\mathrm{T}}\tilde{\mathbf{V}})^{-1}\Big\}. \qquad \text{(S2.6)}\end{aligned}$$

(ii) Next we calculate $\partial^2 d(\mathbf{A})/\partial \mathrm{vec}_{\mathrm{R}}(\mathbf{A})\partial \mathrm{vec}_{\mathrm{R}}(\mathbf{A})^{\mathrm{T}}|_{\mathbf{A}=\hat{\mathbf{B}}}$.Let $\tilde{\mathbf{S}} = \tilde{\mathbf{S}}_1\tilde{\mathbf{S}}_2$ where $\tilde{\mathbf{S}}_1 = \mathbf{Z}(\mathbf{I} - \tilde{\mathbf{H}})\hat{\mathbf{H}}$ and $\tilde{\mathbf{S}}_2 = \tilde{\mathbf{V}}(\tilde{\mathbf{V}}^{\mathrm{T}}\tilde{\mathbf{V}})^{-1}$. Then it holds that

$$\frac{\partial^2 d(\mathbf{A})}{\partial \mathrm{vec}_{\mathrm{R}}(\mathbf{A})\partial \mathrm{vec}_{\mathrm{R}}(\mathbf{A})^{\mathrm{T}}} = -2\hat{K}^{-1}\left\{\frac{\partial \tilde{\mathbf{S}}_1}{\partial \mathbf{A}}(\mathbf{I}_p \otimes \tilde{\mathbf{S}}_2) + \frac{\partial \tilde{\mathbf{S}}_2}{\partial \mathbf{A}}(\tilde{\mathbf{S}}_1^{\mathrm{T}} \otimes \mathbf{I}_{\hat{K}})\right\}. \qquad \text{(S2.7)}$$

As $\hat{\mathbf{H}}$ is idempotent, we have $\tilde{\mathbf{S}}_1|_{\mathbf{A}=\hat{\mathbf{B}}} = \mathbf{Z}(\mathbf{I} - \hat{\mathbf{H}})\hat{\mathbf{H}} = 0$. Therefore,

$$\left.\frac{\partial^2 d(\mathbf{A})}{\partial \mathrm{vec}_{\mathrm{R}}(\mathbf{A})\partial \mathrm{vec}_{\mathrm{R}}(\mathbf{A})^{\mathrm{T}}}\right|_{\mathbf{A}=\hat{\mathbf{B}}} = -2\hat{K}^{-1}\left.\frac{\partial \tilde{\mathbf{S}}_1}{\partial \mathbf{A}}\right|_{\mathbf{A}=\hat{\mathbf{B}}}(\mathbf{I}_p \otimes \tilde{\mathbf{S}}_2|_{\mathbf{A}=\hat{\mathbf{B}}}). \qquad \text{(S2.8)}$$

Combining (S2.4), (S2.8) and the equation

$$\frac{\partial \tilde{\mathbf{S}}_1}{\partial \mathbf{A}} = -\frac{\partial \tilde{\mathbf{H}}}{\partial \mathbf{A}}(\mathbf{Z}^{\mathrm{T}} \otimes \hat{\mathbf{H}})$$

gives

$$\left.\frac{\partial^2 d(\mathbf{A})}{\partial \mathrm{vec}_{\mathrm{R}}(\mathbf{A}) \partial \mathrm{vec}_{\mathrm{R}}(\mathbf{A})^{\mathrm{T}}}\right|_{\mathbf{A}=\hat{\mathbf{B}}} = 2\hat{K}^{-1}\{\mathbf{Z}(\mathbf{I}-\hat{\mathbf{H}})\mathbf{Z}^{\mathrm{T}}\} \otimes (\hat{\mathbf{V}}^{\mathrm{T}}\hat{\mathbf{V}})^{-1}. \quad \text{(S2.9)}$$

Then the equality in Lemma 1 can be easily shown from (S2.9).

The proof of Lemma 1 has been finished.

S3 Proof of Theorem 1

The proof of Theorem 1 is straightforward and omitted.

S4 Proof of Theorem 2

Let $\mathbf{Z}^*$ be a $p \times n$ matrix with the ith column $\mathbf{z}_i^* = \mathbf{x}_i^* - \bar{\mathbf{x}}^*$ where $\bar{\mathbf{x}}^*$ denotes the sample mean of $\mathbf{x}^*$. According to the definition of space displacement,

$$D^*(\boldsymbol{\omega}) \;=\; 1 - \frac{1}{\hat{K}^*}\mathrm{tr}\{\mathbf{P}_{\mathbf{Z}^{*\mathrm{T}}\hat{\mathcal{B}}^*}\mathbf{P}_{\mathbf{Z}^{*\mathrm{T}}\hat{\mathcal{B}}^*(\boldsymbol{\omega})}\}, \quad \text{(S4.1)}$$

where $\hat{\mathcal{B}}^* = \mathcal{M}(\hat{\mathbf{B}}^*)$, $\hat{\mathcal{B}}^*(\boldsymbol{\omega}) = \mathcal{M}\{\hat{\mathbf{B}}^*(\boldsymbol{\omega})\}$, $\hat{\mathbf{B}}^* = (\hat{\mathbf{b}}_1^*, \ldots, \hat{\mathbf{b}}_{\hat{K}^*}^*)$ and $\hat{\mathbf{B}}^*(\boldsymbol{\omega}) = (\hat{\mathbf{b}}_1^*(\boldsymbol{\omega}), \ldots, \hat{\mathbf{b}}_{\hat{K}^*}^*(\boldsymbol{\omega}))$ in which $\hat{\mathbf{b}}_1^*, \ldots, \hat{\mathbf{b}}_{\hat{K}^*}^*$ and $\hat{\mathbf{b}}_1^*(\boldsymbol{\omega}), \ldots, \hat{\mathbf{b}}_{\hat{K}^*}^*(\boldsymbol{\omega})$ denote, respectively, the estimates and perturbed estimates of dimension reduction directions for $\mathbf{x}^*$ obtained by sliced inverse regression. By the similar reasoning as in Li (2000), we show that $\hat{\mathbf{B}}^*(\boldsymbol{\omega}) = (\mathbf{A}^{\mathrm{T}})^{-1}\hat{\mathbf{B}}(\boldsymbol{\omega})$ as follows. Under the multiplicative scheme, the samples of $\mathbf{x}$ is perturbed to $\mathbf{x}_i(\boldsymbol{\omega}) = \omega_i \mathbf{x}_i (i = 1, \ldots, n)$, and the samples of $\mathbf{x}^*$, the transformed variate vector, are perturbed to $\mathbf{x}_i^*(\boldsymbol{\omega}) = \omega_i \mathbf{A}\mathbf{x}_i$ $(i = 1, \ldots, n)$. Let $\mathbf{X}(\boldsymbol{\omega}) = (\mathbf{x}_1(\boldsymbol{\omega}), \ldots, \mathbf{x}_n(\boldsymbol{\omega}))$, $\mathbf{X}^*(\boldsymbol{\omega}) = (\mathbf{x}_1^*(\boldsymbol{\omega}), \ldots, \mathbf{x}_n^*(\boldsymbol{\omega}))$ and $\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^*(\boldsymbol{\omega})$ denote the perturbed estimate of the covariance matrix of $\mathbf{x}^*$. Note that $\mathbf{X}^*(\boldsymbol{\omega}) = \mathbf{A}\mathbf{X}(\boldsymbol{\omega})$. Then

$$\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^*(\boldsymbol{\omega}) \;=\; n^{-1}\mathbf{X}^*(\boldsymbol{\omega})(\mathbf{I} - n^{-1}\mathbf{1}_a\mathbf{1}_a^{\mathrm{T}})\mathbf{X}^*(\boldsymbol{\omega})^{\mathrm{T}} = \mathbf{A}\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}(\boldsymbol{\omega})\mathbf{A}^{\mathrm{T}}. \quad \text{(S4.2)}$$

Let $\bar{\mathbf{x}}_l(\boldsymbol{\omega}) = \sum_{i \in \mathcal{I}_l} \mathbf{x}_i(\boldsymbol{\omega})/n_l$, $\bar{\mathbf{x}}(\boldsymbol{\omega}) = \sum_{i=1}^{n} \mathbf{x}_i(\boldsymbol{\omega})/n$, $\bar{\mathbf{x}}_l^*(\boldsymbol{\omega}) = \sum_{i \in \mathcal{I}_l} \mathbf{x}_i^*(\boldsymbol{\omega})/n_l$ and $\bar{\mathbf{x}}^*(\boldsymbol{\omega}) = \sum_{i=1}^{n} \mathbf{x}_i^*(\boldsymbol{\omega})/n$. Then $\bar{\mathbf{x}}^*(\boldsymbol{\omega}) = \mathbf{A}\bar{\mathbf{x}}(\boldsymbol{\omega})$ and $\bar{\mathbf{x}}_l^*(\boldsymbol{\omega}) = \mathbf{A}\bar{\mathbf{x}}_l(\boldsymbol{\omega})$. Hence

$$\begin{aligned} \hat{\boldsymbol{\Sigma}}_{\eta}^*(\boldsymbol{\omega}) &= n^{-1}\sum_{l=1}^{\tau} n_l\{\bar{\mathbf{x}}_l^*(\boldsymbol{\omega}) - \bar{\mathbf{x}}^*(\boldsymbol{\omega})\}\{\bar{\mathbf{x}}_l^*(\boldsymbol{\omega}) - \bar{\mathbf{x}}^*(\boldsymbol{\omega})\}^{\mathrm{T}} \\ &= \mathbf{A}\hat{\boldsymbol{\Sigma}}_{\eta}(\boldsymbol{\omega})\mathbf{A}^{\mathrm{T}}. \end{aligned} \quad \text{(S4.3)}$$

Hence, $\hat{\boldsymbol{\Sigma}}_{\eta}^*(\boldsymbol{\omega})\mathbf{b} = \lambda\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^*(\boldsymbol{\omega})\mathbf{b}$ is equivalent to $\hat{\boldsymbol{\Sigma}}_{\eta}(\boldsymbol{\omega})(\mathbf{A}^{\mathrm{T}}\mathbf{b}) = \lambda\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}(\boldsymbol{\omega})(\mathbf{A}^{\mathrm{T}}\mathbf{b})$. Moreover, the equalities $\mathbf{b}^{\mathrm{T}}\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^*(\boldsymbol{\omega})\mathbf{b} = 1$ and $\mathbf{a}^{\mathrm{T}}\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^*(\boldsymbol{\omega})\mathbf{b} = 0$ are, respectively, equivalent to the equalities

$$(\mathbf{A}^{\mathrm{T}}\mathbf{b})^{\mathrm{T}}\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}(\boldsymbol{\omega})(\mathbf{A}^{\mathrm{T}}\mathbf{b}) = 1, \quad (\mathbf{A}^{\mathrm{T}}\mathbf{a})^{\mathrm{T}}\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}(\boldsymbol{\omega})(\mathbf{A}^{\mathrm{T}}\mathbf{b}) = 0.$$

That means $\hat{K}^* = \hat{K}$ and $(\mathbf{A}^{\mathrm{T}})^{-1}\hat{\mathbf{b}}_1(\boldsymbol{\omega}), \ldots, (\mathbf{A}^{\mathrm{T}})^{-1}\hat{\mathbf{b}}_{\hat{K}}(\boldsymbol{\omega})$ are the perturbed estimates of dimension reduction directions for $\mathbf{x}^*$ in sliced inverse regression. Then we have shown that $\hat{\mathbf{B}}^*(\boldsymbol{\omega}) = (\mathbf{A}^{\mathrm{T}})^{-1}\hat{\mathbf{B}}(\boldsymbol{\omega})$, which also indicates $\hat{\mathbf{B}}^* = (\mathbf{A}^{\mathrm{T}})^{-1}\hat{\mathbf{B}}$. In addition, $\mathbf{Z}^* = \mathbf{A}\mathbf{Z}$. Therefore, it holds that $\mathbf{Z}^{*\mathrm{T}}\hat{\mathbf{B}}^* = \mathbf{Z}^{\mathrm{T}}\hat{\mathbf{B}}$ and $\mathbf{Z}^{*\mathrm{T}}\hat{\mathbf{B}}^*(\boldsymbol{\omega}) = \mathbf{Z}^{\mathrm{T}}\hat{\mathbf{B}}(\boldsymbol{\omega})$ which indicate $D^*(\boldsymbol{\omega}) = D(\boldsymbol{\omega})$.

The proof of Theorem 2 is finished.

S5 Proof of Lemma 2

1. We prove that $\hat{\lambda}(t)$ and $\hat{\mathbf{b}}(t)$ are differentiable at $t = 0$ and continuous in a real neighborhood of $t = 0$.

Note that $\hat{\lambda}(t)$ and $\hat{\mathbf{b}}(t)$ satisfy

$$\hat{\boldsymbol{\Sigma}}_\eta(\boldsymbol{\omega}_{(0)}+t\mathbf{h})\hat{\mathbf{b}}(t) = \hat{\lambda}(t)\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}(\boldsymbol{\omega}_{(0)}+t\mathbf{h})\hat{\mathbf{b}}(t) \quad \text{and} \quad \hat{\mathbf{b}}(t)^\top\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}(\boldsymbol{\omega}_{(0)}+t\mathbf{h})\hat{\mathbf{b}}(t) = 1.$$

Let $\tilde{\boldsymbol{\Sigma}}(t) = \hat{\boldsymbol{\Sigma}}_{\mathbf{x}}(\boldsymbol{\omega}_{(0)}+t\mathbf{h})^{-1/2}\hat{\boldsymbol{\Sigma}}_\eta(\boldsymbol{\omega}_{(0)}+t\mathbf{h})\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}(\boldsymbol{\omega}_{(0)}+t\mathbf{h})^{-1/2}$, where $\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}(\boldsymbol{\omega}_{(0)}+t\mathbf{h})^{-1/2}$ is well defined since $\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}(\boldsymbol{\omega}_{(0)} + t\mathbf{h})$ will be proved to be positively definite in a real neighborhood of $t = 0$. We have

$$\tilde{\boldsymbol{\Sigma}}(t)\tilde{\mathbf{b}}(t) = \hat{\lambda}(t)\tilde{\mathbf{b}}(t),$$

where $\tilde{\mathbf{b}}(t) = \hat{\boldsymbol{\Sigma}}_{\mathbf{x}}(\boldsymbol{\omega}_{(0)} + t\mathbf{h})^{1/2}\hat{\mathbf{b}}(t)$ and $\tilde{\mathbf{b}}(t)^\top\tilde{\mathbf{b}}(t) = 1$. That is,

$$\hat{\mathbf{b}}(t) = \hat{\boldsymbol{\Sigma}}_{\mathbf{x}}(\boldsymbol{\omega}_{(0)} + t\mathbf{h})^{-1/2}\tilde{\mathbf{b}}(t)$$

and $\hat{\lambda}(t)$ and $\tilde{\mathbf{b}}(t)$ are, respectively, the eigenvalue and corresponding standardized eigenvector of $\tilde{\boldsymbol{\Sigma}}(t)$.

Firstly, we prove that $\hat{\mathbf{\Sigma}}_{\mathbf{x}}(\boldsymbol{\omega}_{(0)} + t\mathbf{h})$ is positively definite and $\hat{\mathbf{\Sigma}}_{\mathbf{x}}(\boldsymbol{\omega}_{(0)} + t\mathbf{h})^{-1/2}$ is differentiable in a real neighborhood of $t = 0$. Some related concepts and theorems can be found in the section 1 of the chapter II in Kato (2013). Note that $\hat{\mathbf{\Sigma}}_{\mathbf{x}}(\boldsymbol{\omega}_{(0)} + t\mathbf{h})$ is well defined and holomorphic in a complex neighborhood of $t = 0$, say $\mathcal{C}_0$. Then the number of eigenvalues of $\hat{\mathbf{\Sigma}}_{\mathbf{x}}(\boldsymbol{\omega}_{(0)} + t\mathbf{h})$ is a constant s independent of t, with the exception of some special values of t, and there are only a finite number of such exceptional points t in each compact subset of $\mathcal{C}_0$. Since $\hat{\mathbf{\Sigma}}_{\mathbf{x}}(\boldsymbol{\omega}_{(0)} + t\mathbf{h})$ is diagonalizable when t is a real number, a real number in the neighborhood will not be an exceptional point. Hence, there exists a complex simple subdomain of $\mathcal{C}_0$ containing a real neighborhood of $t = 0$ but no exceptional point. Let $\mathcal{C}$ denote this subdomain. Let $\lambda_{\mathbf{x},1}(t) > \cdots > \lambda_{\mathbf{x},m}(t)$ and $\mathbf{P}_{\mathbf{x},1}(t), \ldots, \mathbf{P}_{\mathbf{x},m}(t)$ be all the eigenvalues of $\hat{\mathbf{\Sigma}}_{\mathbf{x}}(\boldsymbol{\omega}_{(0)} + t\mathbf{h})$ and eigenprojection matrices for these eigenvalues. Then $\lambda_{\mathbf{x},1}(t), \cdots, \lambda_{\mathbf{x},m}(t)$ and $\mathbf{P}_{\mathbf{x},1}(t), \ldots, \mathbf{P}_{\mathbf{x},m}(t)$ are all holomorphic in $\mathcal{C}$. On the other hand, for t in the above real neighborhood of $t = 0$ contained in $\mathcal{C}$, which is denoted by $\mathcal{C}_1$, $\hat{\mathbf{\Sigma}}_{\mathbf{x}}(\boldsymbol{\omega}_{(0)} + t\mathbf{h})$ is symmetric and hence diagonalizable. Note that $\hat{\mathbf{\Sigma}}_{\mathbf{x}} = \hat{\mathbf{\Sigma}}_{\mathbf{x}}(\boldsymbol{\omega}_{(0)} + t\mathbf{h})|_{t=0}$ is assumed to be positively definite indicating that $\lambda_{\mathbf{x},1}(0), \cdots, \lambda_{\mathbf{x},m}(0)$ are all positive. Then there exists a real neighborhood of $t = 0$, which is contained by $\mathcal{C}_1$ and denoted by $\mathcal{C}_2$, such that for t in $\mathcal{C}_2$, $\lambda_{\mathbf{x},1}(t), \cdots, \lambda_{\mathbf{x},m}(t)$ are all positive, which means $\hat{\mathbf{\Sigma}}_{\mathbf{x}}(\boldsymbol{\omega}_{(0)} + t\mathbf{h})$ is positively definite for t in $\mathcal{C}_2$. Since $\mathcal{C}_2$ is contained by $\mathcal{C}$, we have $\lambda_{\mathbf{x},1}(t), \cdots, \lambda_{\mathbf{x},m}(t)$ and $\mathbf{P}_{\mathbf{x},1}(t), \ldots, \mathbf{P}_{\mathbf{x},m}(t)$ are all differentiable in $\mathcal{C}_2$. Furthermore, for t in $\mathcal{C}_2$, $\hat{\mathbf{\Sigma}}_{\mathbf{x}}(\boldsymbol{\omega}_{(0)} + t\mathbf{h})^{-1/2}$ is diagonalizable and

$$\hat{\mathbf{\Sigma}}_{\mathbf{x}}(\boldsymbol{\omega}_{(0)} + t\mathbf{h})^{-1/2} = \sum_{j=1}^{m} \lambda_{\mathbf{x},1}(t)^{-1/2}\mathbf{P}_{\mathbf{x},j}(t).$$

Then $\hat{\mathbf{\Sigma}}_{\mathbf{x}}(\boldsymbol{\omega}_{(0)} + t\mathbf{h})^{-1/2}$ is differentiable in $\mathcal{C}_2$, which is a real neighborhood of $t = 0$. That also means, for t in $\mathcal{C}_2$, $\hat{\mathbf{\Sigma}}_{\mathbf{x}}(\boldsymbol{\omega}_{(0)} + t\mathbf{h})^{-1/2}$ can be written as

$$\hat{\mathbf{\Sigma}}_{\mathbf{x}}(\boldsymbol{\omega}_{(0)} + t\mathbf{h})^{-1/2} = \hat{\mathbf{\Sigma}}_{\mathbf{x}}^{-1/2} + t\hat{\mathbf{\Sigma}}_{\mathbf{x},2} + o(t), \tag{S5.1}$$

for some matrix $\hat{\mathbf{\Sigma}}_{\mathbf{x},2}$.

As both $\hat{\mathbf{\Sigma}}_{\mathbf{x}}(\boldsymbol{\omega}_{(0)} + t\mathbf{h})^{-1/2}$ and $\hat{\mathbf{\Sigma}}_{\eta}(\boldsymbol{\omega}_{(0)} + t\mathbf{h})$ are differentiable in a real neighborhood of $t = 0$, we have $\tilde{\mathbf{\Sigma}}(t)$ is differentiable in this real neighborhood. Note that $\hat{\lambda}$ is a simple eigenvalue. According to the section 5.1 of the chapter II in Kato (2013), the dimension of the total eigenspace for the $\hat{\lambda}$-group equals the dimension of eigen-subspace of $\hat{\lambda}$, which is one, and $\hat{\lambda}(t)$ and its eigenprojection matrix are continuous in a real neighborhood of $t = 0$, which also means that $\tilde{\mathbf{b}}(t)$ can be chosen to be continuous from the two standardized eigenvectors in the one-dimensional eigenspace of $\hat{\lambda}(t)$ in this neighborhood. Then $\hat{\mathbf{b}}(t)$ is continuous in the real neighborhood of $t = 0$ where both $\hat{\mathbf{\Sigma}}_{\mathbf{x}}(\boldsymbol{\omega}_{(0)} + t\mathbf{h})^{-1/2}$ and $\tilde{\mathbf{b}}(t)$ are continuous. Moreover, according to the theorem 5.4 in the section 5.4 of the chapter II in Kato (2013), $\hat{\lambda}(t)$ and $\mathbf{P}_\lambda(t)$ are both differentiable at $t=0$, where $\mathbf{P}_\lambda(t)$ is the eigen-projection matrix for $\hat{\lambda}(t)$, and furthermore, since the dimension of the total eigenspace for the $\hat{\lambda}$-group equals one as discussed above, the standardized eigenvector $\tilde{\mathbf{b}}(t)$ for $\hat{\lambda}(t)$ is differentiable at $t=0$, which means $\tilde{\mathbf{b}}(t)$ can be chosen in such a way:

$$\tilde{\mathbf{b}}(t) = \tilde{\mathbf{b}} + t\mathbf{v} + o(t). \tag{S5.2}$$

Note that $\tilde{\mathbf{b}}(t)$ can also be chosen as $-\tilde{\mathbf{b}} - t\mathbf{v} + o(t)$, but this does not change $\text{QC}_{\mathbf{h}}$ (see Theorem 1). Combining (S5.1) and (S5.2) gives that $\hat{\mathbf{b}}(t)$ is differentiable at $t=0$ in the real space.

The above proof needs a comment that, the theories for the operators given by Kato (2013), which are mentioned above, hold for the matrices because the matrices can be regarded as operators.

2. We prove (6.2) and (6.3).

(i) Note that $\hat{\boldsymbol{\Sigma}}_\eta(\boldsymbol{\omega}_{(0)} + t\mathbf{h})\hat{\mathbf{b}}(t) = \hat{\lambda}(t)\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}(\boldsymbol{\omega}_{(0)} + t\mathbf{h})\hat{\mathbf{b}}(t)$. Substituting $\hat{\lambda}(t) = \hat{\lambda} + t\hat{\lambda}_{*,1} + o(t)$, $\hat{\mathbf{b}}(t) = \hat{\mathbf{b}} + t\mathbf{f} + o(t)$,

$$\hat{\boldsymbol{\Sigma}}_\eta(\boldsymbol{\omega}_{(0)} + t\mathbf{h}) = \hat{\boldsymbol{\Sigma}}_\eta + t\hat{\boldsymbol{\Sigma}}_{\eta,1} + o(t), \text{ and } \hat{\boldsymbol{\Sigma}}_{\mathbf{x}}(\boldsymbol{\omega}_{(0)} + t\mathbf{h}) = \hat{\boldsymbol{\Sigma}}_{\mathbf{x}} + t\hat{\boldsymbol{\Sigma}}_{\mathbf{x},1} + o(t)$$

into this equality and equating coefficients of t in it gives

$$(\hat{\boldsymbol{\Sigma}}_\eta - \hat{\lambda}\hat{\boldsymbol{\Sigma}}_{\mathbf{x}})\mathbf{f} = \hat{\lambda}_{*,1}\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}\hat{\mathbf{b}} + \hat{\lambda}\hat{\boldsymbol{\Sigma}}_{\mathbf{x},1}\hat{\mathbf{b}} - \hat{\boldsymbol{\Sigma}}_{\eta,1}\hat{\mathbf{b}}. \tag{S5.3}$$

Note that $\hat{\boldsymbol{\Sigma}}_\eta \hat{\mathbf{b}} = \hat{\lambda}\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}\hat{\mathbf{b}}$ and $\hat{\mathbf{b}}^{\mathrm{T}}\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}\hat{\mathbf{b}} = 1$. Then premultiplying (S5.3) by $\hat{\mathbf{b}}^{\mathrm{T}}$ gives (6.2).

(ii) Next we prove (6.3). Firstly, we rewrite (S5.3) as

$$\begin{aligned}(\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{-1/2}\hat{\boldsymbol{\Sigma}}_\eta\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{-1/2} - \hat{\lambda}\mathbf{I})(\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{1/2}\mathbf{f}) &= \hat{\lambda}_{*,1}\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{1/2}\hat{\mathbf{b}} + \hat{\lambda}\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{-1/2}\hat{\boldsymbol{\Sigma}}_{\mathbf{x},1}\hat{\mathbf{b}} \\ &\quad -\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{-1/2}\hat{\boldsymbol{\Sigma}}_{\eta,1}\hat{\mathbf{b}}. \qquad \text{(S5.4)}\end{aligned}$$

We try to solve (S5.4) though $\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{-1/2}\hat{\boldsymbol{\Sigma}}_\eta\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{-1/2} - \hat{\lambda}\mathbf{I}$ is not invertible. Substituting $\hat{\mathbf{b}}(t) = \hat{\mathbf{b}} + t\mathbf{f} + o(t)$ and $\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}(\boldsymbol{\omega}_{(0)} + t\mathbf{h}) = \hat{\boldsymbol{\Sigma}}_{\mathbf{x}} + t\hat{\boldsymbol{\Sigma}}_{\mathbf{x},1} + o(t)$ into

$$\hat{\mathbf{b}}(t)^{\mathrm{T}}\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}(\boldsymbol{\omega}_{(0)} + t\mathbf{h})\hat{\mathbf{b}}(t) = 1$$

and equating coefficients of t gives the equality

$$2\hat{\mathbf{b}}^{\mathrm{T}}\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}\mathbf{f} + \hat{\mathbf{b}}^{\mathrm{T}}\hat{\boldsymbol{\Sigma}}_{\mathbf{x},1}\hat{\mathbf{b}} = 0,$$

which means

$$\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{1/2}\hat{\mathbf{b}} \perp 2\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{1/2}\mathbf{f} + \hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{-1/2}\hat{\boldsymbol{\Sigma}}_{\mathbf{x},1}\hat{\mathbf{b}}.$$

Moreover, zero is a simple eigenvalue of $\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{-1/2}\hat{\boldsymbol{\Sigma}}_\eta\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{-1/2} - \hat{\lambda}\mathbf{I}$ with $\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{1/2}\hat{\mathbf{b}}$ associated standardized eigenvector, which means the space $\mathcal{M}(\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{-1/2}\hat{\boldsymbol{\Sigma}}_\eta\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{-1/2} - \hat{\lambda}\mathbf{I})$ is the orthogonal complement of $\mathcal{M}(\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{1/2}\hat{\mathbf{b}})$. Hence,

$$2\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{1/2}\mathbf{f} + \hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{-1/2}\hat{\boldsymbol{\Sigma}}_{\mathbf{x},1}\hat{\mathbf{b}} \in \mathcal{M}(\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{-1/2}\hat{\boldsymbol{\Sigma}}_\eta\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{-1/2} - \hat{\lambda}\mathbf{I}).$$

Let $\mathbf{P}_\lambda$ be the orthogonal projection matrix of the space $\mathcal{M}(\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{-1/2}\hat{\boldsymbol{\Sigma}}_\eta\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{-1/2} -$

$\hat{\lambda}\mathbf{I})$. Then

$$\mathbf{P}_\lambda = (\hat{\mathbf{\Sigma}}_{\mathbf{x}}^{-1/2}\hat{\mathbf{\Sigma}}_\eta\hat{\mathbf{\Sigma}}_{\mathbf{x}}^{-1/2} - \hat{\lambda}\mathbf{I})^+(\hat{\mathbf{\Sigma}}_{\mathbf{x}}^{-1/2}\hat{\mathbf{\Sigma}}_\eta\hat{\mathbf{\Sigma}}_{\mathbf{x}}^{-1/2} - \hat{\lambda}\mathbf{I})$$

and

$$\mathbf{P}_\lambda(2\hat{\mathbf{\Sigma}}_{\mathbf{x}}^{1/2}\mathbf{f} + \hat{\mathbf{\Sigma}}_{\mathbf{x}}^{-1/2}\hat{\mathbf{\Sigma}}_{\mathbf{x},1}\hat{\mathbf{b}}) = 2\hat{\mathbf{\Sigma}}_{\mathbf{x}}^{1/2}\mathbf{f} + \hat{\mathbf{\Sigma}}_{\mathbf{x}}^{-1/2}\hat{\mathbf{\Sigma}}_{\mathbf{x},1}\hat{\mathbf{b}}. \tag{S5.5}$$

On the other hand, zero is also a simple eigenvalue of $(\hat{\mathbf{\Sigma}}_{\mathbf{x}}^{-1/2}\hat{\mathbf{\Sigma}}_\eta\hat{\mathbf{\Sigma}}_{\mathbf{x}}^{-1/2} - \hat{\lambda}\mathbf{I})^+$ with $\hat{\mathbf{\Sigma}}_{\mathbf{x}}^{1/2}\hat{\mathbf{b}}$ associated standardized eigenvector, which means

$$(\hat{\mathbf{\Sigma}}_{\mathbf{x}}^{-1/2}\hat{\mathbf{\Sigma}}_\eta\hat{\mathbf{\Sigma}}_{\mathbf{x}}^{-1/2} - \hat{\lambda}\mathbf{I})^+(\hat{\mathbf{\Sigma}}_{\mathbf{x}}^{1/2}\hat{\mathbf{b}}) = 0. \tag{S5.6}$$

From (S5.5) and (S5.6), premultiplying (S5.4) by $(\hat{\mathbf{\Sigma}}_{\mathbf{x}}^{-1/2}\hat{\mathbf{\Sigma}}_\eta\hat{\mathbf{\Sigma}}_{\mathbf{x}}^{-1/2} - \hat{\lambda}\mathbf{I})^+$ gives (6.3).

The proof of Lemma 2 has been finished.

S6 Simplifying the expression of f in Lemma 2

Now we simplify the expression of $\mathbf{f}$ in Lemma 2. Let $\hat{\lambda}_2, \ldots, \hat{\lambda}_p$ be the other eigenvalues of $\hat{\mathbf{\Sigma}}_{\mathbf{x}}^{-1/2}\hat{\mathbf{\Sigma}}_\eta\hat{\mathbf{\Sigma}}_{\mathbf{x}}^{-1/2}$. Since $\hat{\mathbf{\Sigma}}_{\mathbf{x}}^{-1/2}\hat{\mathbf{\Sigma}}_\eta\hat{\mathbf{\Sigma}}_{\mathbf{x}}^{-1/2}$ is symmetric, there exists an orthogonal matrix $\mathbf{P}_{\eta,\mathbf{x}}$ such that

$$\hat{\mathbf{\Sigma}}_{\mathbf{x}}^{-1/2}\hat{\mathbf{\Sigma}}_\eta\hat{\mathbf{\Sigma}}_{\mathbf{x}}^{-1/2} = \mathbf{P}_{\eta,\mathbf{x}}\mathrm{diag}(\hat{\lambda}, \hat{\lambda}_2, \ldots, \hat{\lambda}_p)\mathbf{P}_{\eta,\mathbf{x}}^\top.$$

Then it holds that

$$(\hat{\mathbf{\Sigma}}_{\mathbf{x}}^{-1/2}\hat{\mathbf{\Sigma}}_\eta\hat{\mathbf{\Sigma}}_{\mathbf{x}}^{-1/2} - \hat{\lambda}\mathbf{I})^+ = \mathbf{P}_{\eta,\mathbf{x}}\mathrm{diag}\{0, (\hat{\lambda}_2 - \hat{\lambda})^{-1}, \ldots, (\hat{\lambda}_p - \hat{\lambda})^{-1}\}\mathbf{P}_{\eta,\mathbf{x}}^\top,$$

$$\begin{aligned}
&\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{-1/2}(\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{-1/2}\hat{\boldsymbol{\Sigma}}_{\eta}\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{-1/2}-\hat{\lambda}\mathbf{I})^{+}\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{-1/2}\hat{\boldsymbol{\Sigma}}_{\eta}\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{-1}\hat{\boldsymbol{\Sigma}}_{\mathbf{x},1}\hat{\mathbf{b}}\\
=\;&\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{-1/2}\mathbf{P}_{\eta,\mathbf{x}}\mathrm{diag}\{0,\hat{\lambda}_2(\hat{\lambda}_2-\hat{\lambda})^{-1},\ldots,\hat{\lambda}_p(\hat{\lambda}_p-\hat{\lambda})^{-1}\}\mathbf{P}_{\eta,\mathbf{x}}^{\top}\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{-1/2}\hat{\boldsymbol{\Sigma}}_{\mathbf{x},1}\hat{\mathbf{b}},
\end{aligned}$$

$$\begin{aligned}
&\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{-1/2}(\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{-1/2}\hat{\boldsymbol{\Sigma}}_{\eta}\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{-1/2}-\hat{\lambda}\mathbf{I})^{+}\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{-1/2}(\hat{\lambda}\hat{\boldsymbol{\Sigma}}_{\mathbf{x},1}\hat{\mathbf{b}})\\
=\;&\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{-1/2}\mathbf{P}_{\eta,\mathbf{x}}\mathrm{diag}\{0,\hat{\lambda}(\hat{\lambda}_2-\hat{\lambda})^{-1},\ldots,\hat{\lambda}(\hat{\lambda}_p-\hat{\lambda})^{-1}\}\mathbf{P}_{\eta,\mathbf{x}}^{\top}\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{-1/2}\hat{\boldsymbol{\Sigma}}_{\mathbf{x},1}\hat{\mathbf{b}},
\end{aligned}$$

$$\begin{aligned}
&\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{-1/2}(\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{-1/2}\hat{\boldsymbol{\Sigma}}_{\eta}\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{-1/2}-\hat{\lambda}\mathbf{I})^{+}\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{-1/2}(\hat{\boldsymbol{\Sigma}}_{\eta}\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{-1}\hat{\boldsymbol{\Sigma}}_{\mathbf{x},1}\hat{\mathbf{b}}+\hat{\lambda}\hat{\boldsymbol{\Sigma}}_{\mathbf{x},1}\hat{\mathbf{b}})\\
&-\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{-1}\hat{\boldsymbol{\Sigma}}_{\mathbf{x},1}\hat{\mathbf{b}}\\
=\;&\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{-1/2}\mathbf{P}_{\eta,\mathbf{x}}\mathrm{diag}\{-1,2\hat{\lambda}(\hat{\lambda}_2-\hat{\lambda})^{-1},\ldots,2\hat{\lambda}(\hat{\lambda}_p-\hat{\lambda})^{-1}\}\mathbf{P}_{\eta,\mathbf{x}}^{\top}\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{-1/2}\hat{\boldsymbol{\Sigma}}_{\mathbf{x},1}\hat{\mathbf{b}}\\
=\;&\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{-1/2}\mathbf{P}_{\eta,\mathbf{x}}\mathrm{diag}(-1,0,\ldots,0)\mathbf{P}_{\eta,\mathbf{x}}^{\top}\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{-1/2}\hat{\boldsymbol{\Sigma}}_{\mathbf{x},1}\hat{\mathbf{b}}+\\
&\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{-1/2}\mathbf{P}_{\eta,\mathbf{x}}\mathrm{diag}\{0,2\hat{\lambda}(\hat{\lambda}_2-\hat{\lambda})^{-1},\ldots,2\hat{\lambda}(\hat{\lambda}_p-\hat{\lambda})^{-1}\}\mathbf{P}_{\eta,\mathbf{x}}^{\top}\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{-1/2}\hat{\boldsymbol{\Sigma}}_{\mathbf{x},1}\hat{\mathbf{b}}\\
=\;&-\hat{\mathbf{b}}\hat{\mathbf{b}}^{\top}\hat{\boldsymbol{\Sigma}}_{\mathbf{x},1}\hat{\mathbf{b}}+\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{-1/2}(\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{-1/2}\hat{\boldsymbol{\Sigma}}_{\eta}\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{-1/2}-\hat{\lambda}\mathbf{I})^{+}\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{-1/2}(2\hat{\lambda}\hat{\boldsymbol{\Sigma}}_{\mathbf{x},1}\hat{\mathbf{b}}).
\end{aligned}$$

Hence, we have

$$\mathbf{f}=-\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{-1/2}(\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{-1/2}\hat{\boldsymbol{\Sigma}}_{\eta}\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{-1/2}-\hat{\lambda}\mathbf{I})^{+}\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{-1/2}(\hat{\boldsymbol{\Sigma}}_{\eta,1}-\hat{\lambda}\hat{\boldsymbol{\Sigma}}_{\mathbf{x},1})\hat{\mathbf{b}}-\frac{1}{2}\hat{\mathbf{b}}\hat{\mathbf{b}}^{\top}\hat{\boldsymbol{\Sigma}}_{\mathbf{x},1}\hat{\mathbf{b}}.$$

Moreover, it can be shown that

$$\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{-1/2}(\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{-1/2}\hat{\boldsymbol{\Sigma}}_{\eta}\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{-1/2}-\hat{\lambda}\mathbf{I})^{+}\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{-1/2}=\mathbf{P}_{\Sigma,b}(\hat{\boldsymbol{\Sigma}}_{\eta}-\hat{\lambda}\hat{\boldsymbol{\Sigma}}_{\mathbf{x}})^{+}\mathbf{P}_{\Sigma,b}^{\top},$$

where $\mathbf{P}_{\Sigma,b}=\mathbf{I}-\hat{\mathbf{b}}\hat{\mathbf{b}}^{\top}\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}$ is the projection matrix along $\mathcal{M}(\hat{\mathbf{b}})$ to the orthogonal complement of $\mathcal{M}(\boldsymbol{\Sigma}_{\mathbf{x}}\hat{\mathbf{b}})$. Then we have

$$\mathbf{f}=-\mathbf{P}_{\Sigma,b}(\hat{\boldsymbol{\Sigma}}_{\eta}-\hat{\lambda}\hat{\boldsymbol{\Sigma}}_{\mathbf{x}})^{+}\mathbf{P}_{\Sigma,b}^{\top}(\hat{\boldsymbol{\Sigma}}_{\eta,1}-\hat{\lambda}\hat{\boldsymbol{\Sigma}}_{\mathbf{x},1})\hat{\mathbf{b}}-\frac{1}{2}(\hat{\mathbf{b}}^{\top}\hat{\mathbf{b}})\mathbf{P}_{\mathbf{b}}\hat{\boldsymbol{\Sigma}}_{\mathbf{x},1}\hat{\mathbf{b}},$$

where $\mathbf{P_b}$ denotes the orthogonal projection matrix on $\mathcal{M}(\hat{\mathbf{b}})$.

S7 Lemma 3 and its proof

Lemma 3. *Under the sliced inverse regression and scheme (5.1) with* $\boldsymbol{\omega} = \boldsymbol{\omega}_{(0)} + t\mathbf{h}$, *the matrices* $\hat{\boldsymbol{\Sigma}}_{\mathbf{x},1}$ *and* $\hat{\boldsymbol{\Sigma}}_{\eta,1}$ *defined in lemma 2 satisfy*

$$\begin{aligned}
\hat{\boldsymbol{\Sigma}}_{\mathbf{x},1}\boldsymbol{\zeta} &= \frac{1}{n}\mathbf{X}\mathrm{diag}(\mathbf{Z}^{\mathrm{T}}\boldsymbol{\zeta})\mathbf{h} + \frac{1}{n}\mathbf{Z}\mathrm{diag}(\mathbf{X}^{\mathrm{T}}\boldsymbol{\zeta})\mathbf{h}, \\
\hat{\boldsymbol{\Sigma}}_{\eta,1}\boldsymbol{\zeta} &= \frac{1}{n}\mathbf{X}\mathrm{diag}(\mathbf{Z}_{\eta}^{\mathrm{T}}\boldsymbol{\zeta})\mathbf{h} + \frac{1}{n}\mathbf{Z}_{\eta}\mathrm{diag}(\mathbf{X}^{\mathrm{T}}\boldsymbol{\zeta})\mathbf{h},
\end{aligned}$$

for any vector $\boldsymbol{\zeta}$, *where* $\mathbf{Z}_\eta$ *is a* $p \times n$ *matrix with the* i*th column* $\bar{\mathbf{z}}_l = \bar{\mathbf{x}}_l - \bar{\mathbf{x}}$ *for* $i \in \mathcal{I}_l$.

Proof: (i) We first prove the expression of $\hat{\boldsymbol{\Sigma}}_{\mathbf{x},1}\boldsymbol{\zeta}$. Let x_{ij} denote the jth sample of X_i, that is, the ith element of $\mathbf{x}_j$. Under (5.1), x_{ij} is perturbed to $x_{ij}(\boldsymbol{\omega}) = \omega_j x_{ij}$ ($j = 1, \ldots, n$, $i = 1, \ldots, p$). Then $\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}(\boldsymbol{\omega}) = (\hat{\sigma}_{x,ij}(\boldsymbol{\omega}))_{p\times p}$, where

$$\begin{aligned}
\hat{\sigma}_{x,ij}(\boldsymbol{\omega}) &= n^{-1}\sum_{k=1}^{n}\{x_{ik}(\boldsymbol{\omega}) - \bar{x}_{i\cdot}(\boldsymbol{\omega})\}\{x_{jk}(\boldsymbol{\omega}) - \bar{x}_{j\cdot}(\boldsymbol{\omega})\} \\
&= n^{-1}\sum_{k=1}^{n}(x_{ik}\omega_k - n^{-1}\sum_{u=1}^{n}x_{iu}\omega_u)(x_{jk}\omega_k - n^{-1}\sum_{v=1}^{n}x_{jv}\omega_v)
\end{aligned}$$

with $\bar{x}_{i\cdot}(\boldsymbol{\omega}) = n^{-1}\sum\limits_{k=1}^{n}x_{ik}(\boldsymbol{\omega})$. As $\omega = \boldsymbol{\omega}_{(0)} + t\mathbf{h}$ with $\boldsymbol{\omega}_{(0)} = (1, \ldots, 1)^{\mathrm{T}}$, we have $\partial\omega_k/\partial t = h_k$ and

$$\left.\frac{\partial\hat{\sigma}_{x,ij}(\omega)}{\partial t}\right|_{t=0} = n^{-1}\sum_{k=1}^{n}\{(x_{ik}h_k - n^{-1}\sum_{u=1}^{n}x_{iu}h_u)(x_{jk}-\bar{x}_{j\cdot}) + (x_{ik}-\bar{x}_{i\cdot})(x_{jk}h_k - n^{-1}\sum_{v=1}^{n}x_{jv}h_v)\},$$

where $\bar{x}_{j\cdot}$ is the sample mean of X_j, that is, the jth element of $\bar{\mathbf{x}}$, and h_k is the kth element of the vector $\mathbf{h}$. Let $(\hat{\Sigma}_{x,1}\zeta)_i$ denote the ith element of $\hat{\boldsymbol{\Sigma}}_{\mathbf{x},1}\boldsymbol{\zeta}$. As $\hat{\Sigma}_{\mathbf{x},1} = (\partial\hat{\sigma}_{x,ij}(\omega)/\partial t|_{t=0})_{p\times p}$, it is obvious that

$$(\hat{\Sigma}_{x,1}\zeta)_i = n^{-1}\sum_{k=1}^{n}\{(x_{ik}h_k - n^{-1}\sum_{u=1}^{n}x_{iu}h_u)\sum_{j=1}^{p}\zeta_j(x_{jk}-\bar{x}_{j\cdot})\} + n^{-1}\sum_{k=1}^{n}\{(x_{ik}-\bar{x}_{i\cdot})\sum_{j=1}^{p}\zeta_j(x_{jk}h_k - n^{-1}\sum_{v=1}^{n}x_{jv}h_v)\},$$

where ζ_j denotes the jth element of $\boldsymbol{\zeta}$. Note that

$$\sum_{k=1}^{n}[(\sum_{u=1}^{n}x_{iu}h_u)\sum_{j=1}^{p}\{\zeta_j(x_{jk}-\bar{x}_{j\cdot})\}] = (\sum_{u=1}^{n}x_{iu}h_u)\sum_{j=1}^{p}\{\zeta_j\sum_{k=1}^{n}(x_{jk}-\bar{x}_{j\cdot})\} = 0$$

and

$$\sum_{k=1}^{n}[(x_{ik}-\bar{x}_{i\cdot})\sum_{j=1}^{p}\{\zeta_j\sum_{v=1}^{n}(x_{jv}h_v)\}] = 0.$$

Hence, it holds that

$$\begin{aligned}(\hat{\Sigma}_{x,1}\zeta)_i &= n^{-1}\sum_{k=1}^{n}\{(x_{ik}h_k)\sum_{j=1}^{p}\zeta_j(x_{jk}-\bar{x}_{j\cdot})\} \\ &\quad + n^{-1}\sum_{k=1}^{n}\{(x_{ik}-\bar{x}_{i\cdot})\sum_{j=1}^{p}\zeta_j(x_{jk}h_k)\} \\ &= n^{-1}[(\boldsymbol{\zeta}^{\mathrm{T}}\mathbf{Z})\{\mathrm{diag}(\mathbf{x}^{(i)})\mathbf{h}\} + \mathbf{z}^{(i)\mathrm{T}}\mathrm{diag}(\mathbf{X}^{\mathrm{T}}\boldsymbol{\zeta})\mathbf{h}] \\ &= n^{-1}\{\mathbf{x}^{(i)\mathrm{T}}\mathrm{diag}(\mathbf{Z}^{\mathrm{T}}\boldsymbol{\zeta})\mathbf{h} + \mathbf{z}^{(i)\mathrm{T}}\mathrm{diag}(\mathbf{X}^{\mathrm{T}}\boldsymbol{\zeta})\mathbf{h}\},\end{aligned}$$

where $\mathbf{x}^{(i)}$ and $\mathbf{z}^{(i)}$ are, respectively, the original and centralized sample vectors of X_i. As $\mathbf{x}^{(i)}$ and $\mathbf{z}^{(i)}$ are the transposes of the ith rows of $\mathbf{X}$ and $\mathbf{Z}$, respectively, the expression of $\hat{\boldsymbol{\Sigma}}_{\mathbf{x},1}\boldsymbol{\zeta}$ given by Lemma 3 has been proved.

(ii) Next we prove the expression of $\hat{\boldsymbol{\Sigma}}_{\eta,1}\boldsymbol{\zeta}$. It is obvious that $\hat{\boldsymbol{\Sigma}}_{\eta}(\boldsymbol{\omega}) = (\hat{\sigma}_{\eta,ij}(\boldsymbol{\omega}))_{p\times p}$, where

$$\hat{\sigma}_{\eta,ij}(\boldsymbol{\omega}) = n^{-1}\sum_{l=1}^{\tau} n_l\{n_l^{-1}\sum_{k\in\mathcal{I}_l} x_{ik}\omega_k - \bar{x}_{i\cdot}(\boldsymbol{\omega})\}\{n_l^{-1}\sum_{k\in\mathcal{I}_l} x_{jk}\omega_k - \bar{x}_{j\cdot}(\boldsymbol{\omega})\}.$$

Moreover, it can be shown that

$$\begin{aligned}\left.\frac{\partial\hat{\sigma}_{\eta,ij}(\boldsymbol{\omega})}{\partial t}\right|_{t=0} \;=\;& n^{-1}\sum_{l=1}^{\tau} n_l\{(n_l^{-1}\sum_{k\in\mathcal{I}_l} x_{ik}h_k - n^{-1}\sum_{u=1}^{n} x_{iu}h_u)(\bar{x}_{lj} - \bar{x}_{j\cdot})\\ &+(\bar{x}_{li} - \bar{x}_{i\cdot})(n_l^{-1}\sum_{k\in\mathcal{I}_l} x_{jk}h_k - n^{-1}\sum_{v=1}^{n} x_{jv}h_v)\},\end{aligned}$$

where $\bar{x}_{lj}$ denotes the jth element of the vector $\bar{\mathbf{x}}_l$, the lth slice mean of $\mathbf{x}$. Let $(\hat{\Sigma}_{\eta,1}\zeta)_i$ denote the ith element of $\hat{\boldsymbol{\Sigma}}_{\eta,1}\boldsymbol{\zeta}$. As $\hat{\boldsymbol{\Sigma}}_{\eta,1} = (\partial\hat{\sigma}_{\eta,ij}(\boldsymbol{\omega})/\partial t|_{t=0})_{p\times p}$, it is obvious that

$$\begin{aligned}(\hat{\Sigma}_{\eta,1}\zeta)_i \;=\;& n^{-1}\sum_{l=1}^{\tau}[n_l(n_l^{-1}\sum_{k\in\mathcal{I}_l} x_{ik}h_k - n^{-1}\sum_{u=1}^{n} x_{iu}h_u)\sum_{j=1}^{p}\{\zeta_j(\bar{x}_{lj} - \bar{x}_{j\cdot})\}]\\ &+n^{-1}\sum_{l=1}^{\tau}[n_l(\bar{x}_{li} - \bar{x}_{i\cdot})\sum_{j=1}^{p}\{\zeta_j(n_l^{-1}\sum_{k\in\mathcal{I}_l} x_{jk}h_k - n^{-1}\sum_{v=1}^{n} x_{jv}h_v)\}].\end{aligned}$$

As $\sum_{l=1}^{\tau}\{n_l(\bar{x}_{lj} - \bar{x}_{j\cdot})\} = \sum_{l=1}^{\tau}\sum_{k\in\mathcal{I}_l} x_{jk} - \sum_{k=1}^{n} x_{jk} = 0$, it holds that

$$\sum_{l=1}^{\tau}[n_l(n^{-1}\sum_{u=1}^{n} x_{iu}h_u)\sum_{j=1}^{p}\{\zeta_j(\bar{x}_{lj} - \bar{x}_{j\cdot})\}] = 0$$

and

$$\sum_{l=1}^{\tau}[n_l(\bar{x}_{li}-\bar{x}_{i\cdot})\sum_{j=1}^{p}\{\zeta_j(n^{-1}\sum_{v=1}^{n}x_{jv}h_v)\}]=0.$$

Hence,

$$\begin{aligned}(\hat{\Sigma}_{\eta,1}\zeta)_i &= n^{-1}\sum_{l=1}^{\tau}[(\sum_{k\in\mathcal{I}_l}x_{ik}h_k)\sum_{j=1}^{p}\{\zeta_j(\bar{x}_{lj}-\bar{x}_{j\cdot})\}] \\ &\quad +n^{-1}\sum_{l=1}^{\tau}[(\bar{x}_{li}-\bar{x}_{i\cdot})\sum_{j=1}^{p}\{\zeta_j(\sum_{k\in\mathcal{I}_l}x_{jk}h_k)\}].\end{aligned}$$

Let Ξ_1 and Ξ_2 denote, respectively, the first and second terms in the above equation. Then we have $\Xi_1 = n^{-1}\sum_{l=1}^{\tau}\sum_{k\in\mathcal{I}_l}(x_{ik}h_k\bar{\mathbf{z}}_l^{\mathrm{T}}\boldsymbol{\zeta})$ and $\Xi_2 = n^{-1}\sum_{l=1}^{\tau}\sum_{k\in\mathcal{I}_l}(\bar{z}_{li}h_k\mathbf{x}_k^{\mathrm{T}}\boldsymbol{\zeta})$, where $\bar{\mathbf{z}}_l = \bar{\mathbf{x}}_l - \bar{\mathbf{x}}$ and $\bar{z}_{li}$ denotes the ith element of $\bar{\mathbf{z}}_l$. The kth elements of the vectors $\mathrm{diag}(\mathbf{Z}_\eta^{\mathrm{T}}\boldsymbol{\zeta})\mathbf{h}$ and $\mathbf{z}_\eta^{(i)}$ are, respectively, $h_k\bar{z}_l^{\mathrm{T}}\zeta$ and $\bar{z}_{li}$ for $k\in\mathcal{I}_l$, where $\mathbf{z}_\eta^{(i)}$ denotes the ith column of $\mathbf{Z}_\eta^{\mathrm{T}}$, and the kth element of the vector $\mathrm{diag}(\mathbf{X}^{\mathrm{T}}\boldsymbol{\zeta})\mathbf{h}$ is $h_kx_k^{\mathrm{T}}\zeta$. Then we have $\Xi_1 = n^{-1}\mathbf{x}^{(i)\mathrm{T}}\mathrm{diag}(\mathbf{Z}_\eta^{\mathrm{T}}\boldsymbol{\zeta})\mathbf{h}$ and $\Xi_2 = n^{-1}\mathbf{z}_\eta^{(i)\mathrm{T}}\mathrm{diag}(\mathbf{X}^{\mathrm{T}}\boldsymbol{\zeta})\mathbf{h}$ indicating the expression of $\hat{\boldsymbol{\Sigma}}_{\eta,1}\boldsymbol{\zeta}$ given by Lemma 3.

The proof of Lemma 3 is finished.

S8 Proof of Theorem 3

Note that $\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}(\boldsymbol{\omega}) = (\hat{\sigma}_{x,ij}(\boldsymbol{\omega}))_{p\times p}$, where

$$\hat{\sigma}_{x,ij}(\boldsymbol{\omega}) = n^{-1}\sum_{k=1}^{n}(x_{ik}\omega_k - n^{-1}\sum_{u=1}^{n}x_{iu}\omega_u)(x_{jk}\omega_k - n^{-1}\sum_{v=1}^{n}x_{jv}\omega_v)$$

and $\hat{\mathbf{\Sigma}}_\eta(\boldsymbol{\omega}) = (\hat{\sigma}_{\eta,ij}(\boldsymbol{\omega}))_{p\times p}$, where

$$\hat{\sigma}_{\eta,ij}(\boldsymbol{\omega}) = n^{-1}\sum_{l=1}^{\tau} n_l\{n_l^{-1}\sum_{k\in\mathcal{I}_l} x_{ik}\omega_k - \bar{x}_{i\cdot}(\boldsymbol{\omega})\}\{n_l^{-1}\sum_{k\in\mathcal{I}_l} x_{jk}\omega_k - \bar{x}_{j\cdot}(\boldsymbol{\omega})\}$$

with $\bar{x}_{i\cdot}(\boldsymbol{\omega}) = n^{-1}\sum_{k=1}^{n} x_{ik}(\boldsymbol{\omega})$ and $x_{ij}(\boldsymbol{\omega}) = \omega_j x_{ij}$. As $\boldsymbol{\omega} = \boldsymbol{\omega}_{(0)} + t\mathbf{h}$, we have $\omega_i = 1 + th_i$, where h_i is the ith element of $\mathbf{h}$. Hence, both $\hat{\sigma}_{x,ij}(\boldsymbol{\omega})$ and $\hat{\sigma}_{\eta,ij}(\boldsymbol{\omega})$ are quadratic functions of t. It is well known that all the polynomial functions are holomorphic in the complex plane. Moreover, $\hat{\mathbf{\Sigma}}_{\mathbf{x}}(\boldsymbol{\omega})$ and $\hat{\mathbf{\Sigma}}_\eta(\boldsymbol{\omega})$ are both obviously symmetric, and $\hat{\mathbf{\Sigma}}_\eta$ and $\hat{\mathbf{\Sigma}}_{\mathbf{x}}$ are, respectively, symmetric and positively definite matrices. In addition, it is assume that $\mathrm{rk}(\mathbf{Z}) = p$ and the eigenvalues $\hat{\lambda}_1, \ldots, \hat{\lambda}_{\hat{K}}$ of $\hat{\mathbf{\Sigma}}_\eta$ with respect to $\hat{\mathbf{\Sigma}}_{\mathbf{x}}$ are all simple. In summary, all the conditions demanded by Lemma 2 are satisfied. Then according to Lemma 2, Assumption 2 holds, and combining Lemmas 2 and 3 gives that $\mathrm{vec}(\mathbf{F}_{\mathbf{B},\mathbf{h}}) = (\mathbf{\Delta}_{\mathbf{B},1}^{\mathrm{T}}, \ldots, \mathbf{\Delta}_{\mathbf{B},\hat{K}}^{\mathrm{T}})^{\mathrm{T}}\mathbf{h}$. From Lemma 1, substituting this equality into the expression of $\mathrm{QC}_{\mathbf{h}}$ gives $\mathrm{QC}_{\mathbf{h}} = \mathbf{h}^{\mathrm{T}}\ddot{\mathbf{D}}_{\boldsymbol{\omega}_{(0)}}\mathbf{h}$.

The proof of Theorem 3 is completed.

S9 About the condition of simple eigenvalues

In Theorem 3, there is a condition that the eigenvalues $\hat{\lambda}_1, \ldots, \hat{\lambda}_{\hat{K}}$ of $\hat{\mathbf{\Sigma}}_\eta$ with respect to $\hat{\mathbf{\Sigma}}_{\mathbf{x}}$ are all simple. Here, we make a comment about it, using the case of sliced inverse regression as an example. That $\hat{\lambda}$ is a simple eigenvalue of $\hat{\boldsymbol{\Sigma}}_\eta$ with respect to $\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}$ is equivalent to that $\hat{\lambda}$ is a simple eigenvalue of $\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{-1/2}\hat{\boldsymbol{\Sigma}}_\eta\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{-1/2}$. We show that, in sliced inverse regression, it is not a usual case that a nonzero eigenvalue of $\boldsymbol{\Sigma}_{\mathbf{x}}^{-1/2}\boldsymbol{\Sigma}_\eta\boldsymbol{\Sigma}_{\mathbf{x}}^{-1/2}$ is not simple. In the sliced inverse regression, $\boldsymbol{\Sigma}_\eta = \mathrm{cov}(\mathrm{E}(\mathbf{x} \mid Y))$, and without loss of generality, we assume $\boldsymbol{\Sigma}_{\mathbf{x}} = \mathbf{I}$. Now suppose a nonzero eigenvalue, say λ, of $\boldsymbol{\Sigma}_\eta$ is not simple. Then the dimension of eigen-subspace for this eigenvalue, denoted by $\mathcal{B}_\lambda$, will be at least two. On the other hand, for any standardized vector $\boldsymbol{\beta}$ in $\mathcal{B}_\lambda$, it always holds that

$$\mathrm{var}\{\mathrm{E}(\boldsymbol{\beta}^\top\mathbf{x} \mid Y)\} = \boldsymbol{\beta}^\top\boldsymbol{\Sigma}_\eta\boldsymbol{\beta} = \lambda,$$

$$\mathrm{E}\{\mathrm{var}(\boldsymbol{\beta}^\top\mathbf{x} \mid Y)\} = \mathrm{var}(\boldsymbol{\beta}^\top\mathbf{x}) - \mathrm{var}\{\mathrm{E}(\boldsymbol{\beta}^\top\mathbf{x} \mid Y)\} = 1 - \lambda.$$

That means for any standardized vector $\boldsymbol{\beta}$ in $\mathcal{B}_\lambda$, $\mathrm{var}\{\mathrm{E}(\boldsymbol{\beta}^\top\mathbf{x} \mid Y)\}$ and $\mathrm{E}\{\mathrm{var}(\boldsymbol{\beta}^\top\mathbf{x} \mid Y)\}$ are both constant independent of $\boldsymbol{\beta}$. This is not a usual case. To illustrate this, we further give an simple example as follows.

Consider a model $Y = g(\boldsymbol{\beta}_1^\top\mathbf{x}, \boldsymbol{\beta}_2^\top\mathbf{x})$, where $\mathbf{x} \sim N(\mathbf{0}, \mathbf{I})$, $\boldsymbol{\beta} = (1, 0, \ldots, 0)^\top$, $\boldsymbol{\beta}_2 = (0, 1, 0, \ldots, 0)^\top$, and $g(\cdot, \cdot)$ is given as follows:

$g(x_1, x_2) = 1$, for $x_1 \in (-\infty, a_1)$ and $x_2 \in (-\infty, a_2)$;

$g(x_1, x_2) = 2$, for $x_1 \in (-\infty, a_1)$ and $x_2 \in [a_2, +\infty)$;

$g(x_1, x_2) = 3$, for $x_1 \in [a_1, +\infty)$ and $x_2 \in (-\infty, a_2)$;

$g(x_1, x_2) = 4$, for $x_1 \in [a_1, +\infty)$ and $x_2 \in [a_2, +\infty)$.

Then it can be shown that $\mathrm{cov}(\mathrm{E}(\mathbf{x} \mid Y))$ is a diagonal matrix with the first two diagonal elements, respectively,

$$p(a_1)^2/\{\Phi(a_1)\Phi(-a_1)\} \quad \text{and} \quad p(a_2)^2/\{\Phi(a_2)\Phi(-a_2)\},$$

and the other diagonal elements all zero, where $p(x)$ and $\Phi(x)$ denote the density function and distributional function of $N(0,1)$. Hence, in this example, unless $a_1 = a_2$ or $a_1 = -a_2$, the nonzero eigenvalues are always simple.

Moreover, for $\mathbf{\Sigma}_{\mathbf{x}}^{-1/2}\mathbf{\Sigma}_{\eta}\mathbf{\Sigma}_{\mathbf{x}}^{-1/2}$, when $\mathbf{\Sigma}_{\eta}$ and $\mathbf{\Sigma}_{\mathbf{x}}$ are substituted by their estimates, the chance of nonzero eigenvalues which are not simple may be even smaller, due to the errors of estimates.

S10 About re-weighting-case perturbation scheme

As commented in Remark 1 in the main content, we can also consider the re-weighting-case perturbation scheme. Let F and F_n denote the cumulative distribution function and empirical distribution function of $(\mathbf{x}^{\mathrm{T}}, Y)^{\mathrm{T}}$, respectively, and let $\mathcal{T}$ be the functional that satisfies $\mathcal{T}(F) = \mathbf{B}$ and $\mathcal{T}(F_n) = \hat{\mathbf{B}}$. We perturb F_n to $F_{n,\boldsymbol{\omega}} = \sum_{i=1}^{n} \omega_i \delta_{(\mathbf{x}_i^T, y_i)^{\mathrm{T}}}$ with $\sum_{i=1}^{n} \omega_i = 1$, and then $\hat{\mathbf{B}}(\boldsymbol{\omega}) = \mathcal{T}(F_{n,\boldsymbol{\omega}})$, where $\delta_{\mathbf{a}}$ denotes the distribution with probability massed at $\mathbf{a}$.

We now give some properties in this section. The columns of $\hat{\mathbf{B}}(\boldsymbol{\omega})$,

$\hat{\mathbf{b}}_1(\boldsymbol{\omega}),\ldots,\hat{\mathbf{b}}_{\hat{K}}(\boldsymbol{\omega})$, are the standardized orthogonal eigenvectors of $\hat{\boldsymbol{\Sigma}}_\eta(\boldsymbol{\omega})$ with respect to $\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}(\boldsymbol{\omega})$ associated with the largest eigenvalues $\lambda_1(\boldsymbol{\omega}) \geq \ldots \geq \lambda_{\hat{K}}(\boldsymbol{\omega})$, where

$$\begin{aligned} \hat{\boldsymbol{\Sigma}}_\eta(\boldsymbol{\omega}) &= \sum_{l=1}^{\tau}(\sum_{i\in\mathcal{I}_l}\omega_i)\bar{\mathbf{x}}_l(\boldsymbol{\omega})\bar{\mathbf{x}}_l(\boldsymbol{\omega})^{\mathrm{T}} - \bar{\mathbf{x}}(\boldsymbol{\omega})\bar{\mathbf{x}}(\boldsymbol{\omega})^{\mathrm{T}}, \\ \hat{\boldsymbol{\Sigma}}_{\mathbf{x}}(\boldsymbol{\omega}) &= \sum_{i=1}^{n}\omega_i\mathbf{x}_i\mathbf{x}_i^{\mathrm{T}} - \bar{\mathbf{x}}(\boldsymbol{\omega})\bar{\mathbf{x}}(\boldsymbol{\omega})^{\mathrm{T}}, \end{aligned} \tag{S10.1}$$

in which $\bar{\mathbf{x}}(\boldsymbol{\omega}) = \sum_{i=1}^n \omega_i\mathbf{x}_i, \bar{\mathbf{x}}_l(\boldsymbol{\omega}) = \sum_{i\in\mathcal{I}_l}\{(\omega_i/\sum_{j\in\mathcal{I}_l}\omega_j)\mathbf{x}_i\}$, and $\sum_{i=1}^n \omega_i = 1$. Under this scheme, $\boldsymbol{\omega}_{(0)} = (1/n,\ldots,1/n)^{\mathrm{T}}$. We have the following invariance property.

Theorem S1. *Let $\mathbf{x}_1^*,\ldots,\mathbf{x}_n^*$ denote the sample of $\mathbf{x}^*$ under the invertible affine transformation $\mathbf{x}^* = \mathbf{A}\mathbf{x}$ and $D^*(\boldsymbol{\omega})$ denote the space displacement function under the model where Y is regressed on $\mathbf{x}^*$. Then under the re-weighting-case perturbation scheme, it holds for sliced inverse regression that $D^*(\boldsymbol{\omega}) = D(\boldsymbol{\omega})$, which means the space displacement function, the quasi-curvature and the influential direction are all invariant under this transformation.*

Proof:

Let $\hat{\boldsymbol{\Sigma}}^*_{\mathbf{x}}(\boldsymbol{\omega})$ and $\hat{\boldsymbol{\Sigma}}^*_\eta(\boldsymbol{\omega})$, respectively, denote the matrices $\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}(\boldsymbol{\omega})$ and $\hat{\boldsymbol{\Sigma}}_\eta(\boldsymbol{\omega})$ defined by (S10.1) with $\mathbf{x}_1,\ldots,\mathbf{x}_n$ replaced by $\mathbf{x}_1^*,\ldots,\mathbf{x}_n^*$. It can be easily shown that $\hat{\boldsymbol{\Sigma}}^*_\eta(\boldsymbol{\omega}) = \mathbf{A}\hat{\boldsymbol{\Sigma}}_\eta(\boldsymbol{\omega})\mathbf{A}^{\mathrm{T}}$ and $\hat{\boldsymbol{\Sigma}}^*_{\mathbf{x}}(\boldsymbol{\omega}) = \mathbf{A}\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}(\boldsymbol{\omega})\mathbf{A}^{\mathrm{T}}$.

Then by similar reasoning as that in the proof of Theorem 2, we can get $D^*(\boldsymbol{\omega}) = D(\boldsymbol{\omega})$. The proof of Theorem S1 is completed.

To obtain the expression of the quasi-curvature under the re-weighting-case perturbation scheme, we need the following lemma.

Lemma S1. Under (S10.1) with $\boldsymbol{\omega} = \boldsymbol{\omega}_{(0)} + t\mathbf{h}$, the matrices $\hat{\boldsymbol{\Sigma}}_{\mathbf{x},1}$ and $\hat{\boldsymbol{\Sigma}}_{\eta,1}$ defined in Lemma 2 satisfy

$$\hat{\boldsymbol{\Sigma}}_{\mathbf{x},1}\boldsymbol{\zeta} = \mathbf{X}\mathrm{diag}(\mathbf{Z}^{\mathrm{T}}\boldsymbol{\zeta})\mathbf{h} + \mathbf{Z}\mathrm{diag}(\mathbf{X}^{\mathrm{T}}\boldsymbol{\zeta})\mathbf{h} - \mathbf{X}\mathrm{diag}(\mathbf{X}^{\mathrm{T}}\boldsymbol{\zeta})\mathbf{h}, \quad \text{(S10.2)}$$

$$\hat{\boldsymbol{\Sigma}}_{\eta,1}\boldsymbol{\zeta} = \mathbf{X}\mathrm{diag}(\mathbf{Z}_\eta^{\mathrm{T}}\boldsymbol{\zeta})\mathbf{h} + \mathbf{Z}_\eta\mathrm{diag}(\mathbf{X}^{\mathrm{T}}\boldsymbol{\zeta})\mathbf{h} - \mathbf{X}_\eta\mathrm{diag}(\mathbf{X}_\eta^{\mathrm{T}}\boldsymbol{\zeta})\mathbf{h}, \quad \text{(S10.3)}$$

for any vector $\boldsymbol{\zeta}$, where $\mathbf{X}_\eta$ is a $p \times n$ matrix with the ith column $\bar{\mathbf{x}}_l$ for $i \in \mathcal{I}_l$.

Proof:

(i) We first prove the expression of $\hat{\boldsymbol{\Sigma}}_{\mathbf{x},1}\boldsymbol{\zeta}$. We still let $\hat{\sigma}_{x,ij}(\boldsymbol{\omega})$ be the (i,j)th element of $\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}(\boldsymbol{\omega})$. Then

$$\left.\frac{\partial \hat{\sigma}_{x,ij}(\boldsymbol{\omega}_{(0)} + t\mathbf{h})}{\partial t}\right|_{t=0} = \sum_{k=1}^n h_k x_{ik} x_{jk} - \bar{x}_{j\cdot} \sum_{k=1}^n h_k x_{ik} - \bar{x}_{i\cdot} \sum_{k=1}^n h_k x_{jk}.$$

Let $(\hat{\Sigma}_{x,1}\zeta)_i$ denote the ith element of $\hat{\boldsymbol{\Sigma}}_{\mathbf{x},1}\boldsymbol{\zeta}$. As $\hat{\boldsymbol{\Sigma}}_{\mathbf{x},1} = (\partial \hat{\sigma}_{x,ij}(\omega)/\partial t|_{t=0})_{p\times p}$, it holds that

$$
\begin{aligned}
(\hat{\Sigma}_{x,1}\zeta)_i &= \sum_{j=1}^{p}(\sum_{k=1}^{n} h_k x_{ik}x_{jk} - \bar{x}_{j\cdot}\sum_{k=1}^{n} h_k x_{ik} - \bar{x}_{i\cdot}\sum_{k=1}^{n} h_k x_{jk})\zeta_j \\
&= \sum_{k=1}^{n} h_k x_{ik}\sum_{j=1}^{p}(x_{jk}-\bar{x}_{j\cdot})\zeta_j + \sum_{k=1}^{n} h_k(x_{ik}-\bar{x}_{i\cdot})\sum_{j=1}^{p} x_{jk}\zeta_j \\
&\quad -\sum_{k=1}^{n} h_k x_{ik}\sum_{j=1}^{p} x_{jk}\zeta_j \\
&= \mathbf{x}^{(i)\mathrm{T}}\mathrm{diag}(\mathbf{Z}^{\mathrm{T}}\boldsymbol{\zeta})\mathbf{h} + \mathbf{z}^{(i)\mathrm{T}}\mathrm{diag}(\mathbf{X}^{\mathrm{T}}\boldsymbol{\zeta})\mathbf{h} - \mathbf{x}^{(i)\mathrm{T}}\mathrm{diag}(\mathbf{X}^{T}\boldsymbol{\zeta})\mathbf{h}.
\end{aligned}
$$

Then (S10.2) has been proved.

(ii) Next we prove (S10.3). We still let $\hat{\sigma}_{\eta,ij}(\omega)$ be the (i,j)th element of $\hat{\mathbf{\Sigma}}_\eta(\boldsymbol{\omega})$. Then

$$
\begin{aligned}
\left.\frac{\partial\hat{\sigma}_{\eta,ij}(\boldsymbol{\omega}_{(0)}+t\mathbf{h})}{\partial t}\right|_{t=0} &= \sum_{l=1}^{\tau}\{-(\sum_{k\in\mathcal{I}_l} h_k)\bar{x}_{li}\bar{x}_{lj} + (\sum_{k\in\mathcal{I}_l} h_k x_{ik})\bar{x}_{lj} \\
&\quad +\bar{x}_{li}(\sum_{k\in\mathcal{I}_l} h_k x_{jk})\} - (\sum_{k=1}^{n} h_k x_{ik})\bar{x}_{j\cdot} \\
&\quad -\bar{x}_{i\cdot}(\sum_{k=1}^{n} h_k x_{jk}).
\end{aligned}
$$

Let $(\hat{\Sigma}_{\eta,1}\zeta)_i$ denote the ith element of $\hat{\mathbf{\Sigma}}_{\eta,1}\boldsymbol{\zeta}$. As $\hat{\mathbf{\Sigma}}_{\eta,1} = (\partial\hat{\sigma}_{\eta,ij}(\omega)/\partial t|_{t=0})_{p\times p}$, it holds that

$$
\begin{aligned}
(\hat{\Sigma}_{\eta,1}\zeta)_i &= \sum_{j=1}^{p}\zeta_j[\sum_{l=1}^{\tau}\{-(\sum_{k\in\mathcal{I}_l}h_k)\bar{x}_{li}\cdot\bar{x}_{lj}+(\sum_{k\in\mathcal{I}_l}h_kx_{ik})\bar{x}_{lj}+\bar{x}_{li}(\sum_{k\in\mathcal{I}_l}h_kx_{jk})\} \\
&\quad -(\sum_{k=1}^{n}h_kx_{ik})\bar{x}_{j\cdot}-\bar{x}_{i\cdot}(\sum_{k=1}^{n}h_kx_{jk})] \\
&= \sum_{l=1}^{\tau}\sum_{k\in\mathcal{I}_l}\sum_{j=1}^{p}(-h_k\bar{x}_{li}\bar{x}_{lj}\zeta_j+h_kx_{ik}\bar{x}_{lj}\zeta_j+h_kx_{jk}\bar{x}_{li}\zeta_j) \\
&\quad -\sum_{k=1}^{n}\sum_{j=1}^{p}(h_kx_{ik}\bar{x}_{j\cdot}\zeta_j+h_kx_{jk}\bar{x}_{i\cdot}\zeta_j) \\
&= \sum_{l=1}^{\tau}\sum_{k\in\mathcal{I}_l}\{-h_k\bar{x}_{li}\sum_{j=1}^{p}\bar{x}_{lj}\zeta_j+h_kx_{ik}\sum_{j=1}^{p}(\bar{x}_{lj}-\bar{x}_{j\cdot})\zeta_j \\
&\quad +h_k(\bar{x}_{li}-\bar{x}_{i\cdot})\sum_{j=1}^{p}x_{jk}\zeta_j\}
\end{aligned}
$$

Let $\mathbf{x}_\eta^{(i)}$ denote the ith column of $\mathbf{X}_\eta^{\mathrm{T}}$. From the above equality, we have

$$
(\hat{\Sigma}_{\eta,1}\zeta)_i = -\mathbf{x}_\eta^{(i)\mathrm{T}}\mathrm{diag}(\mathbf{X}_\eta^{\mathrm{T}}\boldsymbol{\zeta})\mathbf{h}+\mathbf{x}^{(i)\mathrm{T}}\mathrm{diag}(\mathbf{Z}_\eta^{\mathrm{T}}\boldsymbol{\zeta})\mathbf{h}+\mathbf{z}_\eta^{(i)\mathrm{T}}\mathrm{diag}(\mathbf{X}^{\mathrm{T}}\boldsymbol{\zeta})\mathbf{h}.
$$

Then the equality (S10.3) has been proved. The proof of Lemma S1 is concluded.

If the eigenvalues $\hat{\lambda}_1,\ldots,\hat{\lambda}_{\hat{K}}$ of $\hat{\boldsymbol{\Sigma}}_\eta$ with respect to $\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}$ are all simple, Lemma 2 and Lemma S1 indicate that, under the re-weighting-case scheme, $\mathrm{vec}(\mathbf{F}_{\mathbf{B},\mathbf{h}}) = (\boldsymbol{\Delta}_{\mathbf{B},1}^{(R)\mathrm{T}},\ldots,\boldsymbol{\Delta}_{\mathbf{B},\hat{K}}^{(R)\mathrm{T}})^{\mathrm{T}}\mathbf{h}$, where

$$
\begin{aligned}
\boldsymbol{\Delta}_{\mathbf{B},i}^{(R)} &= \frac{1}{2}(\boldsymbol{\Sigma}_{\eta,\mathbf{x},i}\hat{\boldsymbol{\Sigma}}_\eta\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{-1}+\hat{\lambda}_i\boldsymbol{\Sigma}_{\eta,\mathbf{x},i}-\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{-1})\{\mathbf{X}\mathrm{diag}(\mathbf{Z}^{\mathrm{T}}\hat{\mathbf{b}}_i)+\mathbf{Z}\mathrm{diag}(\mathbf{X}^{\mathrm{T}}\hat{\mathbf{b}}_i) \\
&\quad -\mathbf{X}\mathrm{diag}(\mathbf{X}^{\mathrm{T}}\hat{\mathbf{b}}_i)\}-\boldsymbol{\Sigma}_{\eta,\mathbf{x},i}\{\mathbf{X}\mathrm{diag}(\mathbf{Z}_\eta^{\mathrm{T}}\hat{\mathbf{b}}_i)+\mathbf{Z}_\eta\mathrm{diag}(\mathbf{X}^{\mathrm{T}}\hat{\mathbf{b}}_i) \\
&\quad -\mathbf{X}_\eta\mathrm{diag}(\mathbf{X}_\eta^{\mathrm{T}}\hat{\mathbf{b}}_i)\} \quad (i=1,\ldots,\hat{K}).
\end{aligned}
$$

Combining Lemma 1, Lemma 2 and Lemma S1 gives the following theorem.

Theorem S2. *Under the re-weighting-case perturbation scheme, the quasi-curvature of lifted line along* $\mathbf{h}$ *at* $\boldsymbol{\omega}_{(0)}$ *can be expressed as* $\mathrm{QC}_{\mathbf{h}} = \mathbf{h}^{\mathrm{T}}\ddot{\mathbf{D}}^{(R)}_{\boldsymbol{\omega}_{(0)}}\mathbf{h}$*, where* $\ddot{\mathbf{D}}^{(R)}_{\boldsymbol{\omega}_{(0)}}$ *denotes*

$$\ddot{\mathbf{D}}^{(R)}_{\boldsymbol{\omega}_{(0)}} = \frac{2}{n\hat{K}}\sum_{k=1}^{\hat{K}}\boldsymbol{\Delta}^{(R)\mathrm{T}}_{\mathbf{B},k}\{\mathbf{Z}(\mathbf{I}-\mathbf{P}_{\mathbf{Z}^{\mathrm{T}}\hat{\mathbf{B}}})\mathbf{Z}^{\mathrm{T}}\}\boldsymbol{\Delta}^{(R)}_{\mathbf{B},k}.$$

We now study the influential direction under the re-weighting-case perturbation scheme. Since there is a constraint $\sum_{i=1}^{n}\omega_i = 1$ under the re-weighting-case scheme, we slightly modify the definition of influential direction and aggregate contribution vector under this scheme. As we set $\boldsymbol{\omega} = \boldsymbol{\omega}_{(0)} + t\mathbf{h}$, the constraint $\sum_{i=1}^{n}\omega_i = 1$ is equivalent with $\mathbf{1}^T\mathbf{h} = 0$, where $\mathbf{1} = (1,\ldots,1)^{\mathrm{T}}$. Inspired by Shi and Huang (2011), the influential direction under the re-weighting-case scheme is naturally defined as

$$\mathbf{h}_{\max} = \arg\max_{\|\mathbf{h}\|=1,\mathbf{1}^T\mathbf{h}=0}\mathrm{QC}_{\mathbf{h}}.$$

Moreover, the aggregate contribution vector under this scheme is defined as

$$\mathbf{M}_0 = \sum_{i=1}^{v}\lambda_i^{(R)}\boldsymbol{\vartheta}_i^{(s)},$$

where $\boldsymbol{\vartheta}_i^{(s)} = (\vartheta_{i1}^2,\ldots,\vartheta_{in}^2)^{\mathrm{T}}$, ϑ_{ij} is the jth entry of $\boldsymbol{\vartheta}_i$, and $(\lambda_1^{(R)},\boldsymbol{\vartheta}_1)$ and

$(\lambda_k^{(R)}, \boldsymbol{\vartheta}_k)$ denote the solutions of

$$\max_{\|\mathbf{h}\|=1, \mathbf{1}^{\mathrm{T}}\mathbf{h}=0} \mathbf{h}^{\mathrm{T}} \ddot{\mathbf{D}}_{\boldsymbol{\omega}_{(0)}}^{(R)} \mathbf{h} \tag{S10.4}$$

and

$$\max_{\|\mathbf{h}\|=1, \mathbf{1}^{\mathrm{T}}\mathbf{h}=0, \mathbf{h}^{\mathrm{T}}\boldsymbol{\vartheta}_j, j=1,\ldots,k-1} \mathbf{h}^{\mathrm{T}} \ddot{\mathbf{D}}_{\boldsymbol{\omega}_{(0)}}^{(R)} \mathbf{h}, \tag{S10.5}$$

respectively, and v is the number of non-zero $\lambda_k^{(R)}$s.

Let $\mathbf{P}_{\mathbf{1}^{\perp}} = \mathbf{I}_n - \mathbf{1}\mathbf{1}^{\mathrm{T}}/n$. Then $\mathbf{P}_{\mathbf{1}^{\perp}}$ is the projection matrix onto the orthogonal complement of $\mathcal{M}(\mathbf{1})$ and $\mathbf{1}^{\mathrm{T}}\mathbf{h} = 0$ is equivalent with $\mathbf{h} \in \mathcal{M}(\mathbf{P}_{\mathbf{1}^{\perp}})$. Then (S10.4) and (S10.5) are equivalent with

$$\max_{\|\mathbf{h}\|=1, \mathbf{h}\in\mathcal{M}(\mathbf{P}_{\mathbf{1}^{\perp}})} \mathbf{h}^{\mathrm{T}} \mathbf{P}_{\mathbf{1}^{\perp}} \ddot{\mathbf{D}}_{\boldsymbol{\omega}_{(0)}}^{(R)} \mathbf{P}_{\mathbf{1}^{\perp}} \mathbf{h}$$

and

$$\max_{\|\mathbf{h}\|=1, \mathbf{h}\in\mathcal{M}(\mathbf{P}_{\mathbf{1}^{\perp}}), \mathbf{h}^{\mathrm{T}}\boldsymbol{\vartheta}_j, j=1,\ldots,k-1} \mathbf{h}^{\mathrm{T}} \mathbf{P}_{\mathbf{1}^{\perp}} \ddot{\mathbf{D}}_{\boldsymbol{\omega}_{(0)}}^{(R)} \mathbf{P}_{\mathbf{1}^{\perp}} \mathbf{h},$$

respectively, because $\mathbf{P}_{\mathbf{1}^{\perp}}\mathbf{h} = \mathbf{h}$ when $\mathbf{h} \in \mathcal{M}(\mathbf{P}_{\mathbf{1}^{\perp}})$. Note that all the eigenvectors of $\mathbf{P}_{\mathbf{1}^{\perp}} \ddot{\mathbf{D}}_{\boldsymbol{\omega}_{(0)}}^{(R)} \mathbf{P}_{\mathbf{1}^{\perp}}$ associated with non-zero eigenvalues fall into $\mathcal{M}(\mathbf{P}_{\mathbf{1}^{\perp}})$. Then $\lambda_1^{(R)}, \ldots, \lambda_g^{(R)}$ are the non-zero eigenvalues of $\mathbf{P}_{\mathbf{1}^{\perp}} \ddot{\mathbf{D}}_{\boldsymbol{\omega}_{(0)}}^{(R)} \mathbf{P}_{\mathbf{1}^{\perp}}$ and $\boldsymbol{\vartheta}_1, \ldots, \boldsymbol{\vartheta}_g$ are the orthonormal eigenvectors associated with these non-zero eigenvalues, where g denotes $\mathrm{rk}(\mathbf{P}_{\mathbf{1}^{\perp}} \ddot{\mathbf{D}}_{\boldsymbol{\omega}_{(0)}}^{(R)} \mathbf{P}_{\mathbf{1}^{\perp}})$. Moreover, as $\ddot{\mathbf{D}}_{\boldsymbol{\omega}_{(0)}}^{(R)}$ is non-negative and

$$\max_{\|\mathbf{h}\|=1, \mathbf{h}\in\mathcal{M}(\mathbf{P}_{\mathbf{1}^{\perp}}), \mathbf{h}^{\mathrm{T}}\boldsymbol{\vartheta}_j, j=1,\ldots,g} \mathbf{h}^{\mathrm{T}} \mathbf{P}_{\mathbf{1}^{\perp}} \ddot{\mathbf{D}}_{\boldsymbol{\omega}_{(0)}}^{(R)} \mathbf{P}_{\mathbf{1}^{\perp}} \mathbf{h}$$

$$\leq \max_{\|\mathbf{h}\|=1,\mathbf{h}^{\mathrm{T}}\boldsymbol{\vartheta}_j, j=1,\ldots,g} \mathbf{h}^{\mathrm{T}}\mathbf{P}_{\mathbf{1}^{\perp}}\ddot{\mathbf{D}}^{(R)}_{\boldsymbol{\omega}_{(0)}}\mathbf{P}_{\mathbf{1}^{\perp}}\mathbf{h} = 0,$$

it holds that

$$\max_{\|\mathbf{h}\|=1,\mathbf{h}\in\mathcal{M}(\mathbf{P}_{\mathbf{1}^{\perp}}),\mathbf{h}^{\mathrm{T}}\boldsymbol{\vartheta}_j, j=1,\ldots,g} \mathbf{h}^{\mathrm{T}}\mathbf{P}_{\mathbf{1}^{\perp}}\ddot{\mathbf{D}}^{(R)}_{\boldsymbol{\omega}_{(0)}}\mathbf{P}_{\mathbf{1}^{\perp}}\mathbf{h} = 0,$$

which means $v = g$. That is $v = \mathrm{rk}(\mathbf{P}_{\mathbf{1}^{\perp}}\ddot{\mathbf{D}}^{(R)}_{\boldsymbol{\omega}_{(0)}}\mathbf{P}_{\mathbf{1}^{\perp}})$. Then we have $\mathbf{P}_{\mathbf{1}^{\perp}}\ddot{\mathbf{D}}^{(R)}_{\boldsymbol{\omega}_{(0)}}\mathbf{P}_{\mathbf{1}^{\perp}} = \sum_{i=1}^{v}\lambda_i^{(R)}\boldsymbol{\vartheta}_i\boldsymbol{\vartheta}_i^{\mathrm{T}}$, which means $\mathbf{M}_0$ defined above is the vector that consists of the diagonal elements of $\mathbf{P}_{\mathbf{1}^{\perp}}\ddot{\mathbf{D}}^{(R)}_{\boldsymbol{\omega}_{(0)}}\mathbf{P}_{\mathbf{1}^{\perp}}$. To sum up, we give the following proposition.

Proposition S1. *Under the re-weighting-case perturbation scheme, the influential direction is the eigenvector of* $\mathbf{P}_{\mathbf{1}^{\perp}}\ddot{\mathbf{D}}^{(R)}_{\boldsymbol{\omega}_{(0)}}\mathbf{P}_{\mathbf{1}^{\perp}}$ *associated with its largest eigenvalue and the aggregate contribution vector* $\mathbf{M}_0$ *is the vector that consists of the diagonal elements of* $\mathbf{P}_{\mathbf{1}^{\perp}}\ddot{\mathbf{D}}^{(R)}_{\boldsymbol{\omega}_{(0)}}\mathbf{P}_{\mathbf{1}^{\perp}}$.

Because $\sum_{i=1}^{n}\omega_i = 1$, the influential direction is the eigenvector of $\mathbf{P}_{\mathbf{1}^{\perp}}\ddot{\mathbf{D}}^{(R)}_{\boldsymbol{\omega}_{(0)}}\mathbf{P}_{\mathbf{1}^{\perp}}$ associated with its largest eigenvalue (Shi and Huang, 2011), and the aggregate contribution vector $\mathbf{M}_0$ is the vector that consists of the diagonal elements of $\mathbf{P}_{\mathbf{1}^{\perp}}\ddot{\mathbf{D}}^{(R)}_{\boldsymbol{\omega}_{(0)}}\mathbf{P}_{\mathbf{1}^{\perp}}$, where $\mathbf{P}_{\mathbf{1}^{\perp}} = \mathbf{I}_n - \mathbf{1}\mathbf{1}^{\mathrm{T}}/n$ and $\mathbf{1} = (1,\ldots,1)^{\mathrm{T}}$. Let $\mathbf{u}_i$ be a vector with the ith entry 1 and the other entries 0. It holds that

$$D\left(\boldsymbol{\omega}_{(0)} - t\frac{\mathbf{P}_{\mathbf{1}^{\perp}}\mathbf{u}_i}{\|\mathbf{P}_{\mathbf{1}^{\perp}}\mathbf{u}_i\|}\right) = \frac{n}{n-1}M_{0,i}t^2 + o(t^2), \tag{S10.6}$$

where $M_{0,i}$ is the ith element of $\mathbf{M}_0$. On the other hand, for $t = 1/\{n(n -$

$1)\}^{1/2}$,

$$D\left(\boldsymbol{\omega}_{(0)} - t\frac{\mathbf{P}_{\mathbf{1}^\perp}\mathbf{u}_i}{\|\mathbf{P}_{\mathbf{1}^\perp}\mathbf{u}_i\|}\right) = \frac{n}{n-1}\text{SIFC}(i)t^2, \tag{S10.7}$$

where SIFC(i) is the influence measure of the ith observation given by the case-deletion method proposed by Prendergast and Smith (2010). In this sense, the quasi-curvature method under the re-weighting-case scheme is similar to the case-deletion method.

Now we prove the equalities (S10.6) and (S10.7). These two equalities indicate that the quasi-curvature method under re-weighting-case scheme is similar to the case-deletion method proposed by Prendergast and Smith (2010) in a sense. First, from the definition of the quasi-curvature, it holds under the re-weighting-case scheme that,

$$D(\boldsymbol{\omega}_{(0)} + t\mathbf{h}) = t^2\mathbf{h}^T\ddot{\mathbf{D}}^{(R)}_{\boldsymbol{\omega}_{(0)}}\mathbf{h} + o(t^2). \tag{S10.8}$$

Since $\mathbf{M}_0$ is the vector that consists of the diagonal elements of $\mathbf{P}_{\mathbf{1}^\perp}\ddot{\mathbf{D}}^{(R)}_{\boldsymbol{\omega}_{(0)}}\mathbf{P}_{\mathbf{1}^\perp}$ and obviously $\|\mathbf{P}_{\mathbf{1}^\perp}\mathbf{u}_i\|^2 = (n-1)/n$, the equality (S10.6) is proved by taking $\mathbf{h} = -\mathbf{P}_{\mathbf{1}^\perp}\mathbf{u}_i/\|\mathbf{P}_{\mathbf{1}^\perp}\mathbf{u}_i\|$ in (S10.8), where $\mathbf{u}_i$ is a vector with the ith entry 1 and the other entries 0. The following is the proof of (S10.7). Let $t_0 = 1/\{n(n-1)\}^{1/2}$, it holds that $\boldsymbol{\omega}_{(0)} - t_0\mathbf{P}_{\mathbf{1}^\perp}\mathbf{u}_i/\|\mathbf{P}_{\mathbf{1}^\perp}\mathbf{u}_i\|$ is a vector with the ith entry 0 and the other entries $1/(n-1)$. Let $\boldsymbol{\omega}_{(i)}$ denote

$\boldsymbol{\omega}_{(0)} - t_0\mathbf{P}_{\mathbf{1}^\perp}\mathbf{u}_i/\|\mathbf{P}_{\mathbf{1}^\perp}\mathbf{u}_i\|$. Then it holds that

$$F_{n,\boldsymbol{\omega}_{(i)}} = \frac{1}{n-1}\sum_{j=1,\ldots,n,j\neq i}\delta_{(\mathbf{x}_j^T,y_j)^{\mathrm{T}}}.$$

Let $\hat{\mathbf{B}}_{(i)}$ denote the estimate of $\mathbf{B}$ using sliced inverse regression with the ith observation deleted. Then

$$\hat{\mathbf{B}}(\boldsymbol{\omega}_{(i)}) = \mathcal{T}(F_{n,\boldsymbol{\omega}_{(i)}}) = \hat{\mathbf{B}}_{(i)}. \tag{S10.9}$$

Because $\mathrm{SIFC}(i) = (n-1)^2\{1 - \mathrm{tr}(\mathbf{P}_{\mathbf{Z}^{\mathrm{T}}\hat{\mathbf{B}}}\mathbf{P}_{\mathbf{Z}^{\mathrm{T}}\hat{\mathbf{B}}_{(i)}})/\hat{K}\}$ according to Prendergast and Smith (2010) and $t_0^2 = 1/\{n(n-1)\}$, combining the definition of $D(\boldsymbol{\omega})$ and the equality (S10.9) completes the proof of (S10.7). The proof is completed.

S11 Local influence of dMAVE

S11.1 A brief review of MAVE based on conditional density function

MAVE based on conditional density function (dMAVE) was proposed by Xia (2007). Its idea is based on the fact that the column space of the gradient $\partial m_b(\mathbf{x}, Y)/\partial\mathbf{x}$ is the subset of the central subspace, where $m_b(x,y) = \mathrm{E}\{H_b(Y-y)|\mathbf{x}=x\}$, $H(v)$ is a symmetric density function, and $H_b(v) = b^{-1}H(v/b)$ with $b>0$ a bandwidth. Suppose the structural dimension K is known. By minimizing $G(\mathbf{a},\mathbf{d},\mathbf{B})$, which is defined as

$$n^{-3}\sum_{k=1}^{n}\sum_{j=1}^{n}\hat{\rho}_{jk}\sum_{i=1}^{n}\{H_{b,i}(y_k)-a_{jk}-\mathbf{d}_{jk}\mathbf{B}^{\top}\mathbf{x}_{ij}\}^2K_h(\mathbf{B}^{\top}\mathbf{x}_{ij}) \qquad \text{(S11.1)}$$

with $\mathbf{a}$ and $\mathbf{d}$ being vectors containing, respectively, a_{jk}s and $\mathbf{d}_{jk}$s for $j,k=1,\ldots,n$, with respect to $a_{jk},\mathbf{d}_{jk},j=1,\ldots,n$ and $\mathbf{B}:\mathbf{B}^{\top}\mathbf{B}=\mathbf{I}_K$, dMAVE method obtains the central subspace estimate $\hat{\mathcal{B}}=\mathcal{M}(\hat{\mathbf{B}})$, where $\hat{\mathbf{B}},\hat{a}_{jk},\hat{\mathbf{d}}_{jk},j=1,\ldots,n$ is the minimizer, $K_h(\mathbf{u})=h^{-d}K(\mathbf{u}/h)$ with d the dimension of $\mathbf{u}$, $K(\mathbf{v})=K_0(\mathbf{v}^{\top}\mathbf{v})$, $K_0(v^2)$ is a univariate symmetric density function, $h>0$ is a bandwidth, $H_{b,i}(y)=H_b(y_i-y)$, $\mathbf{x}_{ij}=\mathbf{x}_i-\mathbf{x}_j$, $\hat{\rho}_{jk}=\rho(\hat{f}_{\mathbf{B},h}(\mathbf{x}_j))\rho(\hat{f}_{Y,b}(y_k))$, $\hat{f}_{Y,b}(y)=n^{-1}\sum_{i=1}^{n}H_{b,i}(y)$, $\hat{f}_{\mathbf{B},h}(x)=n^{-1}\sum_{i=1}^{n}K_h(\mathbf{B}^{\top}(\mathbf{x}_i-x))$, $\rho(\cdot)$ is a bounded function with bounded second order derivatives such that $\rho(v)>$ if $v>v_0$; $\rho(v)=0$ if $v\le v_0$ for some small $v_0>0$.

Xia (2007) proposed the following algorithm to implement the estimation.

Step 0. Let $\mathbf{B}_{(1)}$ be an initial estimator of the central subspace directions. Set $s=1$.

Step 1. Let $\mathbf{B}=\mathbf{B}_{(s)}$, calculate the solutions of $(a_{jk},\mathbf{d}_{jk})$,$j,k=1,\ldots,n$, to the minimization problem in (S11.1):

$$
\begin{aligned}
(a_{jk}^{(s)}, \mathbf{d}_{jk}^{(s)\top})^\top &= \left\{\sum_{i=1}^{n} K_{h_s}(\mathbf{B}_{(s)}^\top \mathbf{x}_{ij})(1, \mathbf{x}_{ij}^\top \mathbf{B}_{(s)})^\top (1, \mathbf{x}_{ij}^\top \mathbf{B}_{(s)})\right\}^{-1} \\
&\quad \times \sum_{i=1}^{n} K_{h_s}(\mathbf{B}_{(s)}^\top \mathbf{x}_{ij})(1, \mathbf{x}_{ij}^\top \mathbf{B}_{(s)})^\top H_{b_s,i}(y_k), \qquad \text{(S11.2)}
\end{aligned}
$$

where h_s and b_s are two bandwidths.

Step 2. Let $\rho_{jk}^{(s)} = \rho(\hat{f}_{\mathbf{B}_{(s)},h_s}(\mathbf{x}_j))\rho(\hat{f}_{Y,b_s}(y_k))$. Fixing $a_{jk} = a_{jk}^{(s)}$ and $\mathbf{d}_{jk} = \mathbf{d}_{jk}^{(s)}$, calculate the solution of $\mathbf{B}$ or vec($\mathbf{B}$) to (S11.1):

$$
\begin{aligned}
\mathbf{b}^{(s+1)} &= \left\{\sum_{k,j,i=1}^{n} \rho_{jk}^{(s)} K_{h_s}(\mathbf{B}_{(s)}^\top \mathbf{x}_{ij})\mathbf{x}_{ijk}^{(s)}(\mathbf{x}_{ijk}^{(s)})^\top\right\}^{-1} \\
&\quad \times \sum_{k,j,i=1}^{n} \rho_{jk}^{(s)} K_{h_s}(\mathbf{B}_{(s)}^\top \mathbf{x}_{ij})\mathbf{x}_{ijk}^{(s)}\{H_{b_s,i}(y_k) - a_{jk}^{(s)}\},
\end{aligned}
$$

where $\mathbf{x}_{ijk}^{(s)} = \mathbf{d}_{jk}^{(s)} \otimes \mathbf{x}_{ij}$.

Step 3. Calculate $\mathbf{\Lambda}_{(s+1)} = \{\mathcal{V}(\mathbf{b}^{(s+1)})\}^\top\{\mathcal{V}(\mathbf{b}^{(s+1)})\}$ and $\mathbf{B}_{(s+1)} = \mathcal{V}(\mathbf{b}^{(s+1)}) \times \mathbf{\Lambda}_{(s+1)}^{-1/2}$, where $\mathcal{V}((\mathbf{v}_1^\top, \ldots, \mathbf{v}_q^\top)^\top) = (\mathbf{v}_1, \ldots, \mathbf{v}_q)$. Set $s := s+1$ and go to step 1.

Step 4. Repeat Steps 1-3 until convergence. The final value of $\mathbf{B}_{(s)}$ can be taken as $\hat{\mathbf{B}}$.

In the above algorithm, $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$ and $\mathbf{y} = (y_1, \ldots, y_n)^\top$ are assumed to be standardized. Now we let $\mathbf{X}$ and $\mathbf{y}$ denote the original data and $\tilde{\mathbf{X}} = (\tilde{\mathbf{x}}_1, \ldots, \tilde{\mathbf{x}}_n)$ and $\tilde{\mathbf{y}} = (\tilde{y}_1, \ldots, \tilde{y}_n)^\top$ be the standardized data. Then the central subspace estimate associated with the original data should be

$\hat{\mathcal{B}} = \mathcal{M}(\tilde{\mathbf{B}})$, where $\tilde{\mathbf{B}} = \hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{-1/2}\hat{\mathbf{B}}$ and $\hat{\mathbf{B}}$ is the final value of $\mathbf{B}_{(s)}$ in the above algorithm with $\mathbf{X}$ and $\mathbf{y}$ replaced by $\tilde{\mathbf{X}}$ and $\tilde{\mathbf{y}}$, respectively.

S11.2 A local influence analysis of dMAVE

Now we try to assess local influence of the original observations $(\mathbf{x}_i^\top, y_i)^\top, i = 1, \ldots, n$ on $\hat{\mathcal{B}}$ under the above dMAVE algorithm. Inspired by Zhu and Lee (2001), we assess the local influence of the original observations on $\hat{\mathcal{B}} = \mathcal{M}(\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{-1/2}\mathbf{B}_{(s+1)})$ in the final iteration of the dMAVE algorithm. Firstly, we set the following perturbation scheme.

$$\mathbf{X}(\boldsymbol{\omega}) = \mathbf{X}\mathrm{diag}(\boldsymbol{\omega}) \quad \text{and} \quad \mathbf{y}(\boldsymbol{\omega}) = \mathrm{diag}(\boldsymbol{\omega})\mathbf{y}. \tag{S11.3}$$

Under (S11.3), $\mathbf{x}_i(\boldsymbol{\omega}) = \mathbf{x}_i\omega_i$ and $y_i(\boldsymbol{\omega}) = y_i\omega_i$. In the final iteration under the perturbation (S11.3), $\mathbf{b}^{(s+1)}$ is perturbed to

$$\begin{aligned}
& \mathbf{b}^{(s+1)}(\boldsymbol{\omega}) \\
= & \left\{ \sum_{k,j,i=1}^{n} \tilde{\rho}_{jk}^{(s)}(\boldsymbol{\omega}) K_{h_s}(\mathbf{B}_{(s)}^\top \tilde{\mathbf{x}}_{ij}(\boldsymbol{\omega})) \tilde{\mathbf{x}}_{ijk}^{(s)}(\boldsymbol{\omega}) (\tilde{\mathbf{x}}_{ijk}^{(s)}(\boldsymbol{\omega}))^\top \right\}^{-1} \\
& \times \sum_{k,j,i=1}^{n} \tilde{\rho}_{jk}^{(s)}(\boldsymbol{\omega}) K_{h_s}(\mathbf{B}_{(s)}^\top \tilde{\mathbf{x}}_{ij}(\boldsymbol{\omega})) \tilde{\mathbf{x}}_{ijk}^{(s)}(\boldsymbol{\omega}) \{H_{b_s}(\tilde{y}_i(\boldsymbol{\omega}) - \tilde{y}_k(\boldsymbol{\omega})) - \tilde{a}_{jk}^{(s)}(\boldsymbol{\omega})\},
\end{aligned}$$

where

$$\tilde{\mathbf{x}}_{ij}(\boldsymbol{\omega}) = \tilde{\mathbf{x}}_i(\boldsymbol{\omega}) - \tilde{\mathbf{x}}_j(\boldsymbol{\omega}), \tilde{\mathbf{x}}_i(\boldsymbol{\omega}) = \hat{\boldsymbol{\Sigma}}_{\mathbf{x}}(\boldsymbol{\omega})^{-1/2}(\mathbf{x}_i(\boldsymbol{\omega}) - \bar{\mathbf{x}}(\boldsymbol{\omega})),$$

$$\bar{\mathbf{x}}(\boldsymbol{\omega}) = n^{-1}\sum_{i=1}^{n}\mathbf{x}_i(\boldsymbol{\omega}), \hat{\boldsymbol{\Sigma}}_{\mathbf{x}}(\boldsymbol{\omega}) = n^{-1}\sum_{i=1}^{n}(\mathbf{x}_i(\boldsymbol{\omega}) - \bar{\mathbf{x}}(\boldsymbol{\omega}))(\mathbf{x}_i(\boldsymbol{\omega}) - \bar{\mathbf{x}}(\boldsymbol{\omega}))^\top,$$

$\tilde{\mathbf{x}}_{ijk}^{(s)}(\boldsymbol{\omega}) = \tilde{\mathbf{d}}_{jk}^{(s)}(\boldsymbol{\omega}) \otimes \tilde{\mathbf{x}}_{ij}(\boldsymbol{\omega}), \tilde{\rho}_{jk}^{(s)}(\boldsymbol{\omega}) = \rho(\hat{f}^{(\boldsymbol{\omega})}_{\mathbf{B}_{(s)},h_s}(\tilde{\mathbf{x}}_j(\boldsymbol{\omega})))\rho(\hat{f}^{(\boldsymbol{\omega})}_{Y,b_s}(\tilde{y}_k(\boldsymbol{\omega}))),$

$\hat{f}^{(\boldsymbol{\omega})}_{Y,b_s}(y) = n^{-1}\sum_{i=1}^{n} H_{b_s}(\tilde{y}_i(\boldsymbol{\omega}) - y), \hat{f}^{(\boldsymbol{\omega})}_{\mathbf{B}_s,h_s}(x) = n^{-1}\sum_{i=1}^{n} K_{h_s}(\mathbf{B}_{(s)}^{\top}(\tilde{\mathbf{x}}_i(\boldsymbol{\omega}) - x)),$

$\tilde{y}_i(\boldsymbol{\omega}) = (y_i(\boldsymbol{\omega}) - \bar{y}(\boldsymbol{\omega}))/\sqrt{S_Y(\boldsymbol{\omega})}, \bar{y}(\boldsymbol{\omega}) = n^{-1}\sum_{i=1}^{n} y_i(\boldsymbol{\omega}),$

$S_Y(\boldsymbol{\omega}) = n^{-1}\sum_{i=1}^{n}(y_i(\boldsymbol{\omega}) - \bar{y}(\boldsymbol{\omega}))^2,$

and

$$\begin{aligned}
& (\tilde{a}_{jk}^{(s)}(\boldsymbol{\omega}), \tilde{\mathbf{d}}_{jk}^{(s)}(\boldsymbol{\omega})^{\top})^{\top} \\
= & \left\{\sum_{i=1}^{n} K_{h_s}(\mathbf{B}_{(s)}^{\top}\tilde{\mathbf{x}}_{ij}(\boldsymbol{\omega}))(1, \tilde{\mathbf{x}}_{ij}(\boldsymbol{\omega})^{\top}\mathbf{B}_{(s)})^{\top}(1, \tilde{\mathbf{x}}_{ij}(\boldsymbol{\omega})^{\top}\mathbf{B}_{(s)})\right\}^{-1} \\
& \times \sum_{i=1}^{n} K_{h_s}(\mathbf{B}_{(s)}^{\top}\tilde{\mathbf{x}}_{ij}(\boldsymbol{\omega}))(1, \tilde{\mathbf{x}}_{ij}(\boldsymbol{\omega})^{\top}\mathbf{B}_{(s)})^{\top} H_{b_s}(\tilde{y}_i(\boldsymbol{\omega}) - \tilde{y}_k(\boldsymbol{\omega})).
\end{aligned}$$

Then the central subspace estimate is perturbed to

$$\hat{\mathcal{B}}(\boldsymbol{\omega}) = \mathcal{M}\{\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}(\boldsymbol{\omega})^{-1/2}\mathbf{B}_{(s+1)}(\boldsymbol{\omega})\},$$

where

$$\mathbf{B}_{(s+1)}(\boldsymbol{\omega}) = \mathcal{V}(\mathbf{b}^{(s+1)}(\boldsymbol{\omega}))\boldsymbol{\Lambda}_{(s+1)}(\boldsymbol{\omega})^{-1/2},$$

in which $\boldsymbol{\Lambda}_{(s+1)}(\boldsymbol{\omega}) = \{\mathcal{V}(\mathbf{b}^{(s+1)}(\boldsymbol{\omega}))\}^{\top}\mathcal{V}(\mathbf{b}^{(s+1)}(\boldsymbol{\omega}))$. Let

$$\tilde{\mathbf{B}}_{(s+1)}(\boldsymbol{\omega}) = \hat{\boldsymbol{\Sigma}}_{\mathbf{x}}(\boldsymbol{\omega})^{-1/2}\mathcal{V}(\mathbf{b}^{(s+1)}(\boldsymbol{\omega})) \quad \text{and} \quad \tilde{\mathbf{B}}_{(s+1)} = \hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{-1/2}\mathcal{V}(\mathbf{b}^{(s+1)}),$$

In the expression of $D(\boldsymbol{\omega})$, taking $\hat{\mathbf{B}}$ and $\hat{\mathbf{B}}(\boldsymbol{\omega})$ to be $\tilde{\mathbf{B}}_{(s+1)}$ and $\tilde{\mathbf{B}}_{(s+1)}(\boldsymbol{\omega})$, respectively, we can employ the proposed methodologies of quasi-curvature.

Here we omit $\mathbf{\Lambda}_{(s+1)}(\boldsymbol{\omega})^{-1/2}$ and $\mathbf{\Lambda}_{(s+1)}^{-1/2}$ in the expression of $\tilde{\mathbf{B}}_{(s+1)}(\boldsymbol{\omega})$ and $\tilde{\mathbf{B}}_{(s+1)}$ because from theorem 1, this omission brings no change of $D(\boldsymbol{\omega})$. Let

$$\tilde{\mathbf{b}}^{(s+1)}(\boldsymbol{\omega}) = (\mathbf{I}_K \otimes \mathbf{\Sigma}_{\mathbf{x}}(\boldsymbol{\omega})^{-1/2})\mathbf{b}^{(s+1)}(\boldsymbol{\omega}).$$

Then we have

$$\mathrm{QC}_{\mathbf{h}} = \mathrm{vec}(\mathbf{F}_{\mathbf{B},\mathbf{h}})^\top \left.\frac{\partial^2 d(\mathbf{A})}{\partial \mathrm{vec}(\mathbf{A}) \partial \mathrm{vec}(\mathbf{A})^{\mathrm{T}}}\right|_{\mathbf{A}=\tilde{\mathbf{B}}_{(s+1)}} \mathrm{vec}(\mathbf{F}_{\mathbf{B},\mathbf{h}}),$$

where

$$\mathrm{vec}(\mathbf{F}_{\mathbf{B},\mathbf{h}}) = \left(\left.\frac{\partial \tilde{\mathbf{b}}^{(s+1)}(\boldsymbol{\omega})}{\partial \boldsymbol{\omega}}\right|_{\boldsymbol{\omega}_{(0)}} \right)^\top \mathbf{h}.$$

Hence, to obtain $\mathbf{h}_{\max}$, we only need to calculate $\partial \tilde{\mathbf{b}}^{(s+1)}(\boldsymbol{\omega})/\partial \boldsymbol{\omega}\Big|_{\boldsymbol{\omega}_{(0)}}$. Let

$$\mathbf{C}_1(\boldsymbol{\omega}) = \sum_{k,j,i=1}^{n} \tilde{\rho}_{jk}^{(s)}(\boldsymbol{\omega}) K_{h_s}(\mathbf{B}_{(s)}^\top \tilde{\mathbf{x}}_{ij}(\boldsymbol{\omega})) \tilde{\mathbf{x}}_{ijk}^{(s)}(\boldsymbol{\omega}) (\tilde{\mathbf{x}}_{ijk}^{(s)}(\boldsymbol{\omega}))^\top,$$

$$\mathbf{C}_2(\boldsymbol{\omega}) = \sum_{k,j,i=1}^{n} \tilde{\rho}_{jk}^{(s)}(\boldsymbol{\omega}) K_{h_s}(\mathbf{B}_{(s)}^\top \tilde{\mathbf{x}}_{ij}(\boldsymbol{\omega})) \tilde{\mathbf{x}}_{ijk}^{(s)}(\boldsymbol{\omega}) \{H_{b_s}(\tilde{y}_i(\boldsymbol{\omega}) - \tilde{y}_k(\boldsymbol{\omega})) - \tilde{a}_{jk}^{(s)}(\boldsymbol{\omega})\},$$

$$\mathbf{D}_1(\boldsymbol{\omega}) = \sum_{i=1}^{n} K_{h_s}(\mathbf{B}_{(s)}^\top \tilde{\mathbf{x}}_{ij}(\boldsymbol{\omega}))(1, \tilde{\mathbf{x}}_{ij}(\boldsymbol{\omega})^\top \mathbf{B}_{(s)})^\top (1, \tilde{\mathbf{x}}_{ij}(\boldsymbol{\omega})^\top \mathbf{B}_{(s)}),$$

$$\mathbf{D}_2(\boldsymbol{\omega}) = \sum_{i=1}^{n} K_{h_s}(\mathbf{B}_{(s)}^\top \tilde{\mathbf{x}}_{ij}(\boldsymbol{\omega}))(1, \tilde{\mathbf{x}}_{ij}(\boldsymbol{\omega})^\top \mathbf{B}_{(s)})^\top H_{b_s}(\tilde{y}_i(\boldsymbol{\omega}) - \tilde{y}_k(\boldsymbol{\omega})),$$

$\mathbf{x}_{(i)}$ be the ith row of $\mathbf{X}$, $w \cdot \mathbf{V}$ or $\mathbf{V} \cdot w$ be the scalar multiplication of the scalar w and matrix $\mathbf{V}$, and $\mathbf{E}_{n,i}$ be a $n \times 1$ vector with the ith element one and the others zeros. The calculation of $\partial \tilde{\mathbf{b}}^{(s+1)}(\boldsymbol{\omega})/\partial \boldsymbol{\omega}\Big|_{\boldsymbol{\omega}_{(0)}}$ is given by the following expressions:

$$\frac{\partial\tilde{\mathbf{b}}^{(s+1)}(\boldsymbol{\omega})}{\partial\boldsymbol{\omega}} = \frac{\partial\mathbf{b}^{(s+1)}(\boldsymbol{\omega})}{\partial\boldsymbol{\omega}}(\mathbf{I}_K\otimes\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}(\boldsymbol{\omega})^{-1/2})+\frac{\partial\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}(\boldsymbol{\omega})^{-1/2}}{\partial\boldsymbol{\omega}}(\mathcal{V}(\mathbf{b}^{(s+1)}(\boldsymbol{\omega}))\otimes\mathbf{I}_p);$$

$$\frac{\partial\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}(\boldsymbol{\omega})^{-1/2}}{\partial\boldsymbol{\omega}} = -\frac{\partial\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}(\boldsymbol{\omega})^{1/2}}{\partial\boldsymbol{\omega}}(\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}(\boldsymbol{\omega})^{-1/2}\otimes\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}(\boldsymbol{\omega})^{-1/2});$$

$$\frac{\partial\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}(\boldsymbol{\omega})^{1/2}}{\partial\boldsymbol{\omega}} = \frac{\partial\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}(\boldsymbol{\omega})}{\partial\boldsymbol{\omega}}\{\mathbf{I}_p\otimes\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}(\boldsymbol{\omega})^{1/2}+\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}(\boldsymbol{\omega})^{1/2}\otimes\mathbf{I}_p\}^{-1};$$

$$\begin{aligned}\frac{\partial\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}(\boldsymbol{\omega})}{\partial\boldsymbol{\omega}} &= n^{-1}(\operatorname{diag}(\mathbf{x}_{(1)}),\ldots,\operatorname{diag}(\mathbf{x}_{(p)}))\{\mathbf{I}_p\otimes(\mathbf{P}_{\mathbf{1}^\perp}\mathbf{X}(\boldsymbol{\omega})^\top)\}\\ &\quad+n^{-1}\operatorname{diag}(\mathbf{x}_1^\top,\ldots,\mathbf{x}_n^\top)\{(\mathbf{P}_{\mathbf{1}^\perp}\mathbf{X}(\boldsymbol{\omega})^\top)\otimes\mathbf{I}_p\};\end{aligned}$$

$$\frac{\partial\mathbf{b}^{(s+1)}(\boldsymbol{\omega})}{\partial\boldsymbol{\omega}} = -\frac{\partial\mathbf{C}_1(\boldsymbol{\omega})}{\partial\boldsymbol{\omega}}\{\mathbf{C}_1(\boldsymbol{\omega})^{-1}\otimes\mathbf{b}^{(s+1)}(\boldsymbol{\omega})\}+\frac{\partial\mathbf{C}_2(\boldsymbol{\omega})}{\partial\boldsymbol{\omega}}\mathbf{C}_1(\boldsymbol{\omega})^{-1};$$

$$\begin{aligned}\frac{\partial\mathbf{C}_1(\boldsymbol{\omega})}{\partial\boldsymbol{\omega}} &= \sum_{k,j,i=1}^{n}\Bigg[\frac{\partial\tilde{\rho}_{jk}^{(s)}(\boldsymbol{\omega})K_{h_s}(\mathbf{B}_{(s)}^\top\tilde{\mathbf{x}}_{ij}(\boldsymbol{\omega}))}{\partial\boldsymbol{\omega}}\{\tilde{\mathbf{x}}_{ijk}^{(s)}(\boldsymbol{\omega})^\top\otimes\tilde{\mathbf{x}}_{ijk}^{(s)}(\boldsymbol{\omega})^\top\}\\ &\quad+\frac{\partial\tilde{\mathbf{x}}_{ijk}^{(s)}(\boldsymbol{\omega})\tilde{\mathbf{x}}_{ijk}^{(s)}(\boldsymbol{\omega})^\top}{\partial\boldsymbol{\omega}}\cdot\{\tilde{\rho}_{jk}^{(s)}(\boldsymbol{\omega})K_{h_s}(\mathbf{B}_{(s)}^\top\tilde{\mathbf{x}}_{ij}(\boldsymbol{\omega}))\}\Bigg];\end{aligned}$$

$$\frac{\partial\tilde{\mathbf{x}}_{ijk}^{(s)}(\boldsymbol{\omega})\tilde{\mathbf{x}}_{ijk}^{(s)}(\boldsymbol{\omega})^\top}{\partial\boldsymbol{\omega}} = \frac{\partial\tilde{\mathbf{x}}_{ijk}^{(s)}(\boldsymbol{\omega})}{\partial\boldsymbol{\omega}}\{\mathbf{I}_{p\cdot K}\otimes\tilde{\mathbf{x}}_{ijk}^{(s)}(\boldsymbol{\omega})^\top+\tilde{\mathbf{x}}_{ijk}^{(s)}(\boldsymbol{\omega})^\top\otimes\mathbf{I}_{p\cdot K}\};$$

$$\frac{\partial\tilde{\mathbf{x}}_{ijk}^{(s)}(\boldsymbol{\omega})}{\partial\boldsymbol{\omega}} = \frac{\partial\tilde{\mathbf{d}}_{jk}^{(s)}(\boldsymbol{\omega})}{\partial\boldsymbol{\omega}}\{\mathbf{I}_K\otimes\tilde{\mathbf{x}}_{ij}(\boldsymbol{\omega})^\top\}+\frac{\partial\tilde{\mathbf{x}}_{ij}(\boldsymbol{\omega})}{\partial\boldsymbol{\omega}}\{\tilde{\mathbf{d}}_{jk}^{(s)}(\boldsymbol{\omega})^\top\otimes\mathbf{I}_K\};$$

$$\begin{aligned}\frac{\partial\tilde{\rho}_{jk}^{(s)}(\boldsymbol{\omega})K_{h_s}(\mathbf{B}_{(s)}^\top\tilde{\mathbf{x}}_{ij}(\boldsymbol{\omega}))}{\partial\boldsymbol{\omega}} &= \frac{\partial\tilde{\rho}_{jk}^{(s)}(\boldsymbol{\omega})}{\partial\boldsymbol{\omega}}K_{h_s}(\mathbf{B}_{(s)}^\top\tilde{\mathbf{x}}_{ij}(\boldsymbol{\omega}))\\ &\quad+\frac{\partial K_{h_s}(\mathbf{B}_{(s)}^\top\tilde{\mathbf{x}}_{ij}(\boldsymbol{\omega}))}{\partial\boldsymbol{\omega}}\tilde{\rho}_{jk}^{(s)}(\boldsymbol{\omega});\end{aligned}$$

$$\frac{\partial\mathbf{C}_2(\boldsymbol{\omega})}{\partial\boldsymbol{\omega}} = \sum_{k,j,i=1}^{n}(\mathbf{C}_{2,ijk}^{(1)}+\mathbf{C}_{2,ijk}^{(2)}+\mathbf{C}_{2,ijk}^{(3)}+\mathbf{C}_{2,ijk}^{(4)}),$$

$$\mathbf{C}_{2,ijk}^{(1)} = \frac{\partial \tilde{\mathbf{x}}_{ijk}^{(s)}(\boldsymbol{\omega})}{\partial \boldsymbol{\omega}} \cdot [\tilde{\rho}_{jk}^{(s)}(\boldsymbol{\omega}) K_{h_s}(\mathbf{B}_{(s)}^\top \tilde{\mathbf{x}}_{ij}(\boldsymbol{\omega}))\{H_{b_s}(\tilde{y}_i(\boldsymbol{\omega}) - \tilde{y}_k(\boldsymbol{\omega})) - \tilde{a}_{jk}^{(s)}(\boldsymbol{\omega})\}],$$

$$\mathbf{C}_{2,ijk}^{(2)} = \frac{\partial \tilde{\rho}_{jk}^{(s)}(\boldsymbol{\omega})}{\partial \boldsymbol{\omega}} K_{h_s}(\mathbf{B}_{(s)}^\top \tilde{\mathbf{x}}_{ij}(\boldsymbol{\omega}))\{H_{b_s}(\tilde{y}_i(\boldsymbol{\omega}) - \tilde{y}_k(\boldsymbol{\omega})) - \tilde{a}_{jk}^{(s)}(\boldsymbol{\omega})\}\tilde{\mathbf{x}}_{ijk}^{(s)}(\boldsymbol{\omega})^\top,$$

$$\mathbf{C}_{2,ijk}^{(3)} = \frac{\partial K_{h_s}(\mathbf{B}_{(s)}^\top \tilde{\mathbf{x}}_{ij}(\boldsymbol{\omega}))}{\partial \boldsymbol{\omega}} \tilde{\rho}_{jk}^{(s)}(\boldsymbol{\omega})\{H_{b_s}(\tilde{y}_i(\boldsymbol{\omega}) - \tilde{y}_k(\boldsymbol{\omega})) - \tilde{a}_{jk}^{(s)}(\boldsymbol{\omega})\}\tilde{\mathbf{x}}_{ijk}^{(s)}(\boldsymbol{\omega})^\top,$$

$$\begin{aligned}\mathbf{C}_{2,ijk}^{(4)} &= \left\{\frac{\partial H_{b_s}(\tilde{y}_i(\boldsymbol{\omega}) - \tilde{y}_k(\boldsymbol{\omega}))}{\partial \boldsymbol{\omega}} - \frac{\partial \tilde{a}_{jk}^{(s)}(\boldsymbol{\omega})}{\partial \boldsymbol{\omega}}\right\} \\ &\quad \times \tilde{\rho}_{jk}^{(s)}(\boldsymbol{\omega}) K_{h_s}(\mathbf{B}_{(s)}^\top \tilde{\mathbf{x}}_{ij}(\boldsymbol{\omega}))\tilde{\mathbf{x}}_{ijk}^{(s)}(\boldsymbol{\omega})^\top;\end{aligned}$$

$$\begin{aligned}\frac{\partial(\tilde{a}_{jk}^{(s)}(\boldsymbol{\omega}), \tilde{\mathbf{d}}_{jk}^{(s)}(\boldsymbol{\omega})^\top)^\top}{\partial \boldsymbol{\omega}} &= -\frac{\partial \mathbf{D}_1(\boldsymbol{\omega})}{\partial \boldsymbol{\omega}}\{\mathbf{D}_1(\boldsymbol{\omega})^{-1} \otimes (\tilde{a}_{jk}^{(s)}(\boldsymbol{\omega}), \tilde{\mathbf{d}}_{jk}^{(s)}(\boldsymbol{\omega})^\top)^\top\} \\ &\quad + \frac{\partial \mathbf{D}_2(\boldsymbol{\omega})}{\partial \boldsymbol{\omega}} \mathbf{D}_1(\boldsymbol{\omega})^{-1};\end{aligned}$$

$$\begin{aligned}\frac{\partial \mathbf{D}_1(\boldsymbol{\omega})}{\partial \boldsymbol{\omega}} &= \sum_{i=1}^n \left[\frac{\partial K_{h_s}(\mathbf{B}_{(s)}^\top \tilde{\mathbf{x}}_{ij}(\boldsymbol{\omega}))}{\partial \boldsymbol{\omega}}\{(1, \tilde{\mathbf{x}}_{ij}(\boldsymbol{\omega})^\top \mathbf{B}_{(s)}) \otimes (1, \tilde{\mathbf{x}}_{ij}(\boldsymbol{\omega})^\top \mathbf{B}_{(s)})\}\right. \\ &\quad \left. + \frac{\partial (1, \tilde{\mathbf{x}}_{ij}(\boldsymbol{\omega})^\top \mathbf{B}_{(s)})^\top (1, \tilde{\mathbf{x}}_{ij}(\boldsymbol{\omega})^\top \mathbf{B}_{(s)})}{\partial \boldsymbol{\omega}} \cdot K_{h_s}(\mathbf{B}_{(s)}^\top \tilde{\mathbf{x}}_{ij}(\boldsymbol{\omega}))\right];\end{aligned}$$

$$\begin{aligned}&\frac{\partial (1, \tilde{\mathbf{x}}_{ij}(\boldsymbol{\omega})^\top \mathbf{B}_{(s)})^\top (1, \tilde{\mathbf{x}}_{ij}(\boldsymbol{\omega})^\top \mathbf{B}_{(s)})}{\partial \boldsymbol{\omega}} = \\ &\quad \left(\mathbf{0}, \frac{\partial \tilde{\mathbf{x}}_{ij}(\boldsymbol{\omega})}{\partial \boldsymbol{\omega}} \mathbf{B}_{(s)}\right)\{\mathbf{I}_{K+1} \otimes (1, \tilde{\mathbf{x}}_{ij}(\boldsymbol{\omega})^\top \mathbf{B}_{(s)}) + (1, \tilde{\mathbf{x}}_{ij}(\boldsymbol{\omega})^\top \mathbf{B}_{(s)}) \otimes \mathbf{I}_{K+1}\};\end{aligned}$$

$$\begin{aligned}\frac{\partial \mathbf{D}_2(\boldsymbol{\omega})}{\partial \boldsymbol{\omega}} &= \sum_{i=1}^n \left[\left(\mathbf{0}, \frac{\partial \tilde{\mathbf{x}}_{ij}(\boldsymbol{\omega})}{\partial \boldsymbol{\omega}} \mathbf{B}_{(s)}\right) \cdot \{K_{h_s}(\mathbf{B}_{(s)}^\top \tilde{\mathbf{x}}_{ij}(\boldsymbol{\omega})) H_{b_s}(\tilde{y}_i(\boldsymbol{\omega}) - \tilde{y}_k(\boldsymbol{\omega}))\}\right. \\ &\quad + \frac{\partial K_{h_s}(\mathbf{B}_{(s)}^\top \tilde{\mathbf{x}}_{ij}(\boldsymbol{\omega}))}{\partial \boldsymbol{\omega}} H_{b_s}(\tilde{y}_i(\boldsymbol{\omega}) - \tilde{y}_k(\boldsymbol{\omega}))(1, \tilde{\mathbf{x}}_{ij}(\boldsymbol{\omega})^\top \mathbf{B}_{(s)}) \\ &\quad \left. + \frac{\partial H_{b_s}(\tilde{y}_i(\boldsymbol{\omega}) - \tilde{y}_k(\boldsymbol{\omega}))}{\partial \boldsymbol{\omega}} K_{h_s}(\mathbf{B}_{(s)}^\top \tilde{\mathbf{x}}_{ij}(\boldsymbol{\omega}))(1, \tilde{\mathbf{x}}_{ij}(\boldsymbol{\omega})^\top \mathbf{B}_{(s)})\right];\end{aligned}$$

$$\begin{aligned}\frac{\partial \tilde{\rho}_{jk}^{(s)}(\boldsymbol{\omega})}{\partial \boldsymbol{\omega}} &= \left\{ n^{-1} \sum_{i=1}^{n} \frac{\partial K_{h_s}(\mathbf{B}_{(s)}^{\top} \tilde{\mathbf{x}}_{ij}(\boldsymbol{\omega}))}{\partial \boldsymbol{\omega}} \right\} \left. \frac{\partial \rho(v)}{\partial v} \right|_{v=\hat{f}^{(\boldsymbol{\omega})}_{\mathbf{B}_{(s)},h_s}(\tilde{\mathbf{x}}_j(\boldsymbol{\omega}))} \\ &\quad \times \rho(\hat{f}^{(\boldsymbol{\omega})}_{Y,b_s}(\tilde{y}_k(\boldsymbol{\omega}))) + \left\{ n^{-1} \sum_{i=1}^{n} \frac{\partial H_{b_s}(\tilde{y}_i(\boldsymbol{\omega}) - \tilde{y}_k(\boldsymbol{\omega}))}{\partial \boldsymbol{\omega}} \right\} \\ &\quad \times \left. \frac{\partial \rho(v)}{\partial v} \right|_{v=\hat{f}^{(\boldsymbol{\omega})}_{Y,b_s}(\tilde{y}_k(\boldsymbol{\omega}))} \rho(\hat{f}^{(\boldsymbol{\omega})}_{\mathbf{B}_{(s)},h_s}(\tilde{\mathbf{x}}_j(\boldsymbol{\omega})));\end{aligned}$$

$$\frac{\partial K_{h_s}(\mathbf{B}_{(s)}^{\top} \tilde{\mathbf{x}}_{ij}(\boldsymbol{\omega}))}{\partial \boldsymbol{\omega}} = \frac{\partial \tilde{\mathbf{x}}_{ij}(\boldsymbol{\omega})}{\partial \boldsymbol{\omega}} \mathbf{B}_{(s)} \left. \frac{\partial K_{h_s}(u)}{\partial u} \right|_{u=\mathbf{B}_{(s)}^{\top}\tilde{\mathbf{x}}_{ij}(\boldsymbol{\omega})};$$

$$\frac{\partial H_{b_s}(\tilde{y}_i(\boldsymbol{\omega}) - \tilde{y}_k(\boldsymbol{\omega}))}{\partial \boldsymbol{\omega}} = \left\{ \frac{\tilde{y}_i(\boldsymbol{\omega})}{\partial \boldsymbol{\omega}} - \frac{\tilde{y}_k(\boldsymbol{\omega})}{\partial \boldsymbol{\omega}} \right\} \left. \frac{\partial H_{b_s}(u)}{\partial u} \right|_{u=\tilde{y}_i(\boldsymbol{\omega})-\tilde{y}_k(\boldsymbol{\omega})};$$

$$\frac{\partial \tilde{\mathbf{x}}_{ij}(\boldsymbol{\omega})}{\boldsymbol{\omega}} = \frac{\partial \tilde{\mathbf{x}}_{i}(\boldsymbol{\omega})}{\boldsymbol{\omega}} - \frac{\partial \tilde{\mathbf{x}}_{j}(\boldsymbol{\omega})}{\boldsymbol{\omega}};$$

$$\frac{\partial \tilde{\mathbf{x}}_{i}(\boldsymbol{\omega})}{\partial \boldsymbol{\omega}} = \frac{\partial \hat{\boldsymbol{\Sigma}}_{\mathbf{x}}(\boldsymbol{\omega})^{-1/2}}{\partial \boldsymbol{\omega}} \{\mathbf{I}_p \otimes (\mathbf{x}_i \omega_i - \bar{\mathbf{x}}(\boldsymbol{\omega}))\} + (\mathbf{E}_{n,i}\mathbf{x}_i^{\top} - n^{-1}\mathbf{X}^{\top}) \hat{\boldsymbol{\Sigma}}_{\mathbf{x}}(\boldsymbol{\omega})^{-1/2};$$

$$\frac{\partial \tilde{y}_i(\boldsymbol{\omega})}{\partial \boldsymbol{\omega}} = \frac{\partial S_Y(\boldsymbol{\omega})^{-1/2}}{\partial \boldsymbol{\omega}} \{y_i \omega_i - \bar{y}(\boldsymbol{\omega})\} + (\mathbf{E}_{n,i} y_i - n^{-1}\mathbf{y}) S_Y(\boldsymbol{\omega})^{-1/2};$$

$$\frac{\partial S_Y(\boldsymbol{\omega})^{-1/2}}{\partial \boldsymbol{\omega}} = -\frac{1}{2} \frac{\partial S_Y(\boldsymbol{\omega})}{\partial \boldsymbol{\omega}} S_Y(\boldsymbol{\omega})^{-3/2};$$

$$\frac{\partial S_Y(\boldsymbol{\omega})}{\partial \boldsymbol{\omega}} = 2n^{-1}\mathrm{diag}(\mathbf{y}) \mathbf{P}_{\mathbf{1}^{\perp}} \mathbf{y}(\boldsymbol{\omega}).$$

Remark S1 For dMAVE, we conduct the influence analysis directly for the algorithm minimizing $G(\mathbf{a}, \mathbf{d}, \mathbf{B})$, because the central subspace estimate depends on the algorithm including some important respects such as the selection of the bandwidths which are changed for each iteration. Theoretically, we have another option of the influence analysis method for dMAVE which starts from the objective function. Here we just present the basic idea. Since the central subspace estimate is obtained by minimizing

$$\tilde{G}(\mathbf{a}, \mathbf{d}, \mathbf{B}, \mathbf{\Lambda}) = G(\mathbf{a}, \mathbf{d}, \mathbf{B}) - (1/2)\mathrm{tr}\{\mathbf{\Lambda}(\mathbf{B}^\top \mathbf{B} - \mathbf{I}_K)\}$$

where $\mathbf{\Lambda}$ is a Lagrange multiplier, we can perturb $\tilde{G}(\mathbf{a}, \mathbf{d}, \mathbf{B}, \mathbf{\Lambda})$ to $\tilde{G}(\mathbf{a}, \mathbf{d}, \mathbf{B}, \mathbf{\Lambda}|\boldsymbol{\omega})$ in some scheme such as weighting scheme. Then the profile of $\tilde{G}(\mathbf{a}, \mathbf{d}, \mathbf{B}, \mathbf{\Lambda}|\boldsymbol{\omega})$ for $\mathbf{B}$ can be obtained as

$$G^*(\mathbf{B}, \boldsymbol{\omega}) = \tilde{G}(\mathbf{a}(\mathbf{B}|\boldsymbol{\omega}), \mathbf{d}(\mathbf{B}|\boldsymbol{\omega}), \mathbf{B}, \mathbf{\Lambda}(\mathbf{B}|\boldsymbol{\omega})|\boldsymbol{\omega}).$$

As $\hat{\mathbf{B}}(\boldsymbol{\omega})$, the minimizer of $\tilde{G}(\mathbf{a}, \mathbf{d}, \mathbf{B}, \mathbf{\Lambda}|\boldsymbol{\omega})$, is also the minimizer of $G^*(\mathbf{B}, \boldsymbol{\omega})$, we have

$$\left.\frac{\partial G^*(\mathbf{B}, \boldsymbol{\omega})}{\partial \mathbf{B}}\right|_{\mathbf{B}=\hat{\mathbf{B}}(\boldsymbol{\omega})} = 0.$$

Differentiating both sides of the above equation with respect to $\boldsymbol{\omega}$, we can construct an equation by chain rule, and solving this equation will give $\partial\hat{\mathbf{B}}(\boldsymbol{\omega})/\partial\boldsymbol{\omega}$. The remaining steps are just similar to those of the given method. This method does not depend on the specific algorithm of minimization except that we need the algorithm to obtain $\hat{\mathbf{B}}$, which is used to substitute $\mathbf{B}$ in the expression of $\partial\hat{\mathbf{B}}(\boldsymbol{\omega})/\partial\boldsymbol{\omega}$, and we need to select the bandwidths. In addition, the equation based on chain rule involves two second-order differentiation matrices, and that, combined with the summations $\sum_{j,k,i=1}^{n}(\cdot)$, may bring very heavy computational burden.

S12 Local influence of cumulative mean estimation

The cumulative mean estimation (CUME) was proposed by Zhu et al. (2010). This method is based on the fact that $\mathcal{M}(\mathbf{M})$ is a subset of $\boldsymbol{\Sigma}_{\mathbf{x}}\mathcal{B}$, where $\mathbf{M}$ is the CUME matrix defined as $\mathbf{M} = \mathrm{E}\{\mathbf{m}(\tilde{Y})\mathbf{m}(\tilde{Y})^\top W(\tilde{Y})\}$, in which $\tilde{Y}$ denotes an independent copy of Y, $\mathbf{m}(\tilde{y}) = \mathrm{E}\{\mathbf{x}\mathbf{1}(Y \leq \tilde{y})\}$ with $\mathbf{1}(Y \leq \tilde{y})$ being a indicator function, and $W(\cdot)$ is a nonnegative weight function. Let $\mathbf{m}_n(\tilde{y}) = n^{-1}\sum_{i=1}^{n}(\mathbf{x}_i - \bar{\mathbf{x}})\mathbf{1}(y_i \leq \tilde{y})$ and

$$\mathbf{M}_n = n^{-1}\sum_{i=1}^{n}\{\mathbf{m}_n(y_i)\mathbf{m}_n(y_i)^\top W(y_i)\}.$$

By assuming a known K, the K eigenvectors of $\mathbf{M}_n$ with respect to $\boldsymbol{\Sigma}_{\mathbf{x}}$ associated with the largest eigenvalues are used as an estimate of the basis of $\mathcal{B}$. We consider the scenarios where Y is continuously distributed. In that case, $y_1, \dots, y_n$ are different from each other in probability one and a perturbation small enough to y_i will not change the estimate of $\mathcal{B}$. Hence, we still use the perturbation scheme (5.1), that is,

$$\mathbf{X}(\boldsymbol{\omega}) = \mathbf{X}\mathrm{diag}(\boldsymbol{\omega}).$$

Under this perturbation scheme, we try to obtain $\mathrm{QC}_{\mathbf{h}} = \mathbf{h}^{\mathrm{T}}\ddot{\mathbf{D}}^{(C)}_{\boldsymbol{\omega}_{(0)}}\mathbf{h}$ for CUME, which is a quadratic form of $\mathbf{h}$ with $\ddot{\mathbf{D}}^{(C)}_{\boldsymbol{\omega}_{(0)}}$ to be calculated. It can be shown that, for any $p \times 1$ vector $\boldsymbol{\zeta}$,

$$\begin{aligned}
& \left.\frac{\partial\{\mathbf{M}_n(\boldsymbol{\omega}_{(0)} + t\mathbf{h})\}}{\partial\{t\}}\right|_{t=0} \boldsymbol{\zeta} \\
= & \left(n^{-1}\sum_{i=1}^{n}[n^{-1}\mathbf{X}\tilde{\mathbf{D}}_{\mathbf{1}}(y_i) \cdot \{\mathbf{m}_n(y_i)^\top \boldsymbol{\zeta} W(y_i)\}\right. \\
& \left. + n^{-1}\mathbf{m}_n(y_i)\boldsymbol{\zeta}^\top \mathbf{X}\tilde{\mathbf{D}}_{\mathbf{1}}(y_i) \cdot W(y_i)] + n^{-1}\mathbf{M}_w(\boldsymbol{\zeta})\right)\mathbf{h}, \quad \text{(S12.1)}
\end{aligned}$$

where $\partial\{\mathbf{A}(t)\}/\partial\{t\}$ denotes the matrix with its (i,j)th element being the derivative of the (i,j)th element of $\mathbf{A}(t)$ with respect to t,

$$\tilde{\mathbf{D}}_{\mathbf{1}}(y_i) = \mathbf{D}_{\mathbf{1}}(y_i) - \{n^{-1}\mathbf{1}_n^\top \mathbf{D}_{\mathbf{1}}(y_i)\mathbf{1}_n\} \cdot \mathbf{I},$$

$$\mathbf{D}_{\mathbf{1}}(y_i) = \mathrm{diag}\{\mathbf{1}(y_1 \leq y_i), \ldots, \mathbf{1}(y_n \leq y_i)\},$$

and $\mathbf{M}_w(\boldsymbol{\zeta})$ denotes a matrix with the ith column being

$$\mathbf{m}_n(y_i)\mathbf{m}_n(y_i)^\top \boldsymbol{\zeta} \cdot \left\{ y_i \left.\frac{\mathrm{d}W(u)}{\mathrm{d}u}\right|_{u=y_i} \right\}.$$

To save space, we omit the proof of (S12.1) since it is somewhat direct. Combining lemma 1, lemma 2, the expression of $\hat{\boldsymbol{\Sigma}}_{\mathbf{x},1}\boldsymbol{\zeta}$ in lemma 3 and (S12.1), the matrix $\ddot{\mathbf{D}}^{(C)}_{\boldsymbol{\omega}_{(0)}}$ is given for CUME.

S13 Simulation studies

In each of the following models, $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are independently generated from $N(0, \boldsymbol{\Sigma}_{\mathbf{x}})$ with $\boldsymbol{\Sigma}_{\mathbf{x}} = \mathrm{diag}(\sigma^2_{x,1}, \ldots, \sigma^2_{x,p})$. The errors $\epsilon_1, \ldots, \epsilon_n$ are independently generated from $N(0, \sigma_\epsilon^2)$. We always first generate all the observations from the model under study, and then we change several observations into artificial outliers by resetting some of their entries and keeping all the other entries unchanged. We first consider

$$y_i = (x_{i1} - 3x_{i2})^3 + \epsilon_i \quad (i = 1, \ldots, 100), \tag{S13.1}$$

where x_{ij} is the jth element of $\mathbf{x}_i$. Under (S13.1), we let $p = 20$, $\sigma_\epsilon = 1$, $\sigma_{x,1} = 2$, $\sigma_{x,2} = 0.7$, and $\sigma_{x,i} = 1$ for $i = 3, \ldots, p$, and consider two settings of artificial outliers. In setting (S-I), we reset $x_{10,2} = 5\sigma_{x,2}$, and in setting (S-II), we reset $x_{10,1} = 5\sigma_{x,1}$. Moreover, we also consider

$$y_i = \frac{x_{i1}}{(x_{i2} + 1.5)^2 + 0.5} + \epsilon_i \quad (i = 1, \ldots, 200), \tag{S13.2}$$

where the function expression was proposed by Li (1991). Under (S13.2), $p = 10$, $\sigma_\epsilon = 0.2$, and $\sigma_{x,i} = 1$ for $i = 1, \ldots, p$. In this model, we first consider two settings, called (S-III) and (S-IV). Under (S-III), let $x_{10,1}{=}x_{10,2}{=}x_{10,3}{=}5$. Under (S-IV), let $x_{10,1}{=}\ x_{10,2}{=}x_{10,3}{=}5$, $x_{11,1}{=}x_{11,2}{=}x_{11,3}{=}4.9$, $x_{12,1} = x_{12,2} = x_{12,3} = 4.8$, $y_{11}{=}1.1y_{10}$, and $y_{12}{=}1.2y_{10}$. The three artificial outliers are set to be close to each other under (S-IV). That is to check whether the quasi-curvature method can overcome the difficulties brought by masking effect, since the local influence methods are supposed to have some advantage over the case-deletion methods in the scenarios where several outliers are close to each other. Moreover, we assign specific numbers to the artificial outliers, instead of generating them randomly, to make them outlying stably in all the replications.

Under each of these four settings, we assess the influence of observations on $\hat{\mathcal{B}}$ given by sliced inverse regression. For the slicing strategy, we obtain $[n/v_s]$ slices with each of the first $[n/v_s]-1$ slices containing v_s observations and the last slice containing the remaining observations, where $[\xi]$ denotes the integer closest to ξ. For comparison, three methods are used, including our quasi-curvature approach, which is denoted by QC, and two sample influence functions, which were proposed by Prendergast (2006, 2007) and Prendergast and Smith (2010) and denoted by SIFB and SIFC, respectively. The latter two are both case-deletion methods, and we denote the influence measures that they provide for the ith observation by SIFB(i) and SIFC(i). For both of them, the slices are always kept unchanged after the deletion of each observation. For the quasi-curvature method, the influential direction $\mathbf{h}_{\max}$ under the perturbation scheme (5.1) is used with $|h_{\max,i}|$ to be the influence measure of the ith observation.

We conduct 200 replications. The estimate $\hat{K}$ is obtained through sequential tests with the test level being α_T/p in each step. For now, we take $\alpha_T = 0.05$. Table S1 presents the numbers of replications in which

Table S1: Numbers of replications with artificial outliers identified

setting	(S-I)	(S-II)	(S-III)	(S-IV)				(S-IV)			
v_s	25	25	25	20				15			
outlier	10th	10th	10th	10th	11th	12th	all	10th	11th	12th	all
SIFB	137	164	198	112	115	117	84	126	119	123	91
SIFC	199	200	200	183	181	162	152	162	156	146	123
QC	198	197	200	197	193	187	184	193	187	185	173

Note: The columns 'all' include the numbers of replications where all the 10th, 11th and 12th observations are identified as influential.

the artificial outliers are identified as influential among the 200 total replications. The results show that the quasi-curvature method under scheme (5.1) has stable performance under all the settings. Prendergast's method using SIFC(i) performs well under settings (I), (II) and (III), but its detection power sharply decreases under setting (IV), which may be due to the masking effect. We have also obtained the index plots of influence measures for observations given by the quasi-curvature method under scheme (5.1) in a replication under settings (S-I)–(S-IV). They are presented in Figure S1 and Figure S2, respectively. In all these plots, the artificial outliers all stand out, as expected.

We also consider a scenario where the elements of $\mathbf{x}$ are correlated.

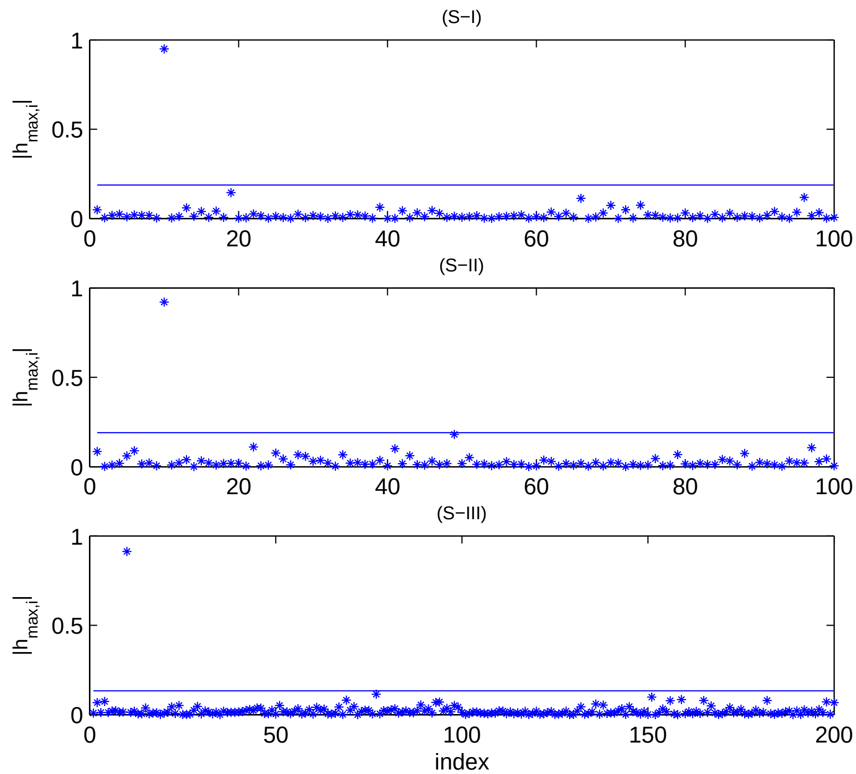

Figure S1: Index plots of influence measures for observations given by quasi-curvature method under scheme (5) with bench-marks $\bar{M} + 1.645 s_M$ (solid line) for one replication in the simulation under the settings (S-I)–(S-III).

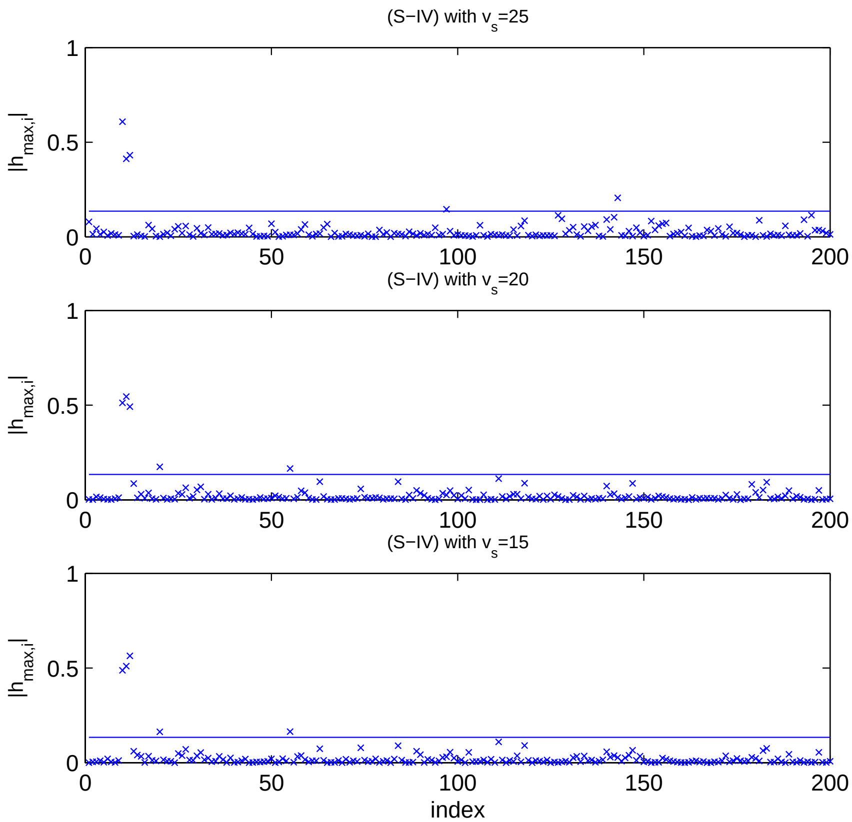

Figure S2: Index plots of influence measures for observations given by quasi-curvature method under scheme (5) with bench-marks $\bar{M} + 1.645 s_M$ (solid line) for one replication in the simulation under the setting (S-IV).

The model (S13.2) is still considered with $\text{cov}(\mathbf{x}) = \mathbf{I}_p + \boldsymbol{\xi}\boldsymbol{\xi}^\top$ and the other settings not changed, where $\boldsymbol{\xi} = (1,-1,1,-1,1,-1,1,-1,1,-1)^\top$. This covariance matrix means some predictors are positively correlated and some are negatively correlated. Firstly, we set artificial outliers with $x_{10,1}$=$x_{10,2}$=$x_{10,3}$=$5\times\sqrt{2}$, $x_{11,1}$=$x_{11,2}$=$x_{11,3}$=$4.9\times\sqrt{2}$, $x_{12,1} = x_{12,2} = x_{12,3} = 4.8\times\sqrt{2}$, y_{11}=$1.1y_{10}$, and y_{12}=$1.2y_{10}$. Let $\boldsymbol{\pi} = (\pi_0, \pi_1, \pi_2, \pi_a)$, where π_0, π_1 and π_2 denote, respectively, the numbers of replications in which the 10th, 11th and 12th observations are identified as influential among all the 200 replications, and π_a denotes the number of replications in which the 10th, 11th and 12th observations are all identified as influential. It turns out under scheme (5.1) $\boldsymbol{\pi}$ for QC, SIFC, and SIFB is, respectively, $(185, 180, 179, 165)$, $(172, 163, 153, 134)$ and $(119, 130, 128, 95)$. When we keep the model unchanged, but set only one artificial outlier with $x_{10,1} = x_{10,2} = x_{10,3} = 5\times\sqrt{2}$, π_0 for QC, SIFC, and SIFB is, respectively, $200, 198, 199$.

We now show that several outliers may result in a sharp decrease in the accuracies of $\hat{K}$ and $\hat{\mathcal{B}}$ and that data trimming is helpful in reducing this decrease. Our data trimming strategy is to conduct sliced inverse regression after deleting the influential observations detected by the quasi-curvature method under (5.1). The benchmark is still taken to be

$\bar{M} + 1.645 s_M$. Four types of settings (S-V)–(S-VIII) are considered under model (S13.2) with $v_s = 20$, where (S-V) means $x_{10,1}$=$x_{10,2}$=$x_{10,3}$=5, $x_{11,1}$=$x_{11,2}$=$x_{11,3}$=4.8 and y_{11}=1.2y_{10}; (S-VI) means for $i = 10, 12$, $x_{i,1}$= $x_{i,2}$=$x_{i,3}$=5, $x_{i+1,1}$=$x_{i+1,2}$=$x_{i+1,3}$=4.8 and y_{i+1}=1.2y_i; (S-VII) means for $i = 10, 12, 14, 16, 18$, $x_{i,1}$= $x_{i,2}$=2.5, $x_{i+1,1}$=$x_{i+1,2}$=2.3 and y_{i+1}=1.2y_i; and (S-VIII) means no artificial outliers. We still perform 200 replications. For the estimate of K, the performances of the sequential tests and Bayesian information criterion are both investigated, which depend on the test level and C_n, respectively. For the former, we take the test level to be α_T/p in each step. The accuracy of $\hat{K}$ is described by the percentage of $\hat{K} = K$ in 200 replications. We present the scatter plots of the accuracies of $\hat{K}$ versus α_T and C_n in Figure S3 and Figure S4, respectively, where the accuracies of $\hat{K}$ with and without data trimming can be compared. These figures show that data trimming makes $\hat{K}$ considerably more robust with respect to the values of C_n and α_T under (S-V)–(S-VII), and under (S-VIII), the loss caused by data trimming, if any, is very slight. The performances of $\hat{\mathcal{B}}$ with and without data trimming are compared in Table S2, which shows that the data trimming provides a substantial improvement in the robustness of $\hat{\mathcal{B}}$.

For the re-weighting-case scheme, equalities (S10.6) and (S10.7) can be

Table S2: Accuracy of $\hat{\mathcal{B}}$ and $\hat{K}$ by sliced inverse regression with and without data trimming

setting	data	corr. mean	corr. std	% of the corr. < 0.6	< 0.7	< 0.8	< 0.85	% of $\hat{K} = K$ by test	by BIC
(S-V)	original	0.870	0.060	0	2.5	9	30.5	82.5	86.5
(S-V)	trimmed	0.9360	0.028	0	0	0	0.5	94.5	98.5
(S-VI)	original	0.816	0.069	0.5	7.5	36	63.5	71.5	70.5
(S-VI)	trimmed	0.924	0.048	0	1	2	6	89.5	96
(S-VII)	original	0.864	0.070	0.5	3.5	15	32.5	58	62
(S-VII)	trimmed	0.904	0.050	0	1	3.5	12	87	88.5
(S-VIII)	original	0.937	0.026	0	0	0	0	95	98.5
(S-VIII)	trimmed	0.937	0.026	0	0	0	0.5	97	99.5

Note: The corr.=$\mathrm{tr}[\hat{\mathbf{B}}(\hat{\mathbf{B}}^{\mathrm{T}}\hat{\mathbf{B}})^{-1}\hat{\mathbf{B}}^{\mathrm{T}}\mathbf{B}(\mathbf{B}^{\mathrm{T}}\mathbf{B})^{-1}\mathbf{B}^{\mathrm{T}}]/K$ since $\mathrm{cov}(\mathbf{X}) = \mathbf{I}$ in the simulation. The percentages are calculated among the 200 replications. For the estimate of K, we take $\alpha_T = 0.05$ and $C_n = 0.3$. 'Std' means standard deviation.

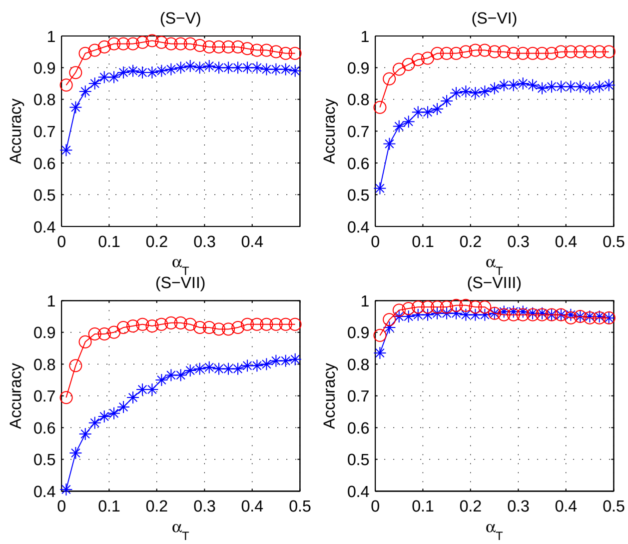

Figure S3: Plots of accuracy of $\hat{K}$ versus α_T in the sequential tests with (circles) and without (stars) data trimming.

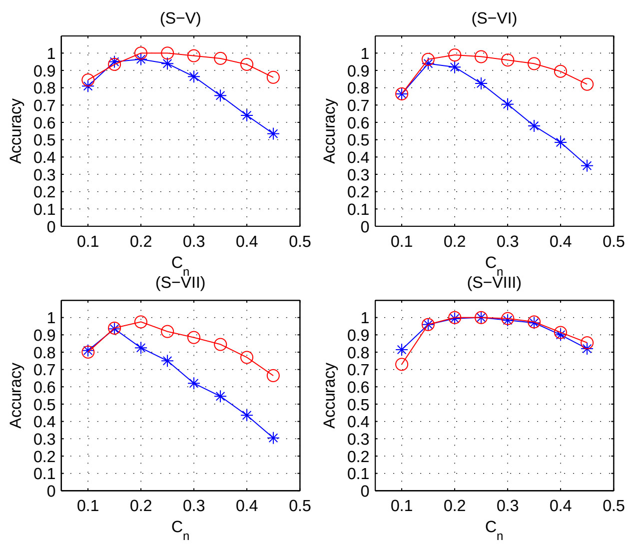

Figure S4: Plots of accuracy of $\hat{K}$ versus C_n using Bayesian information criterion with (circles) and without (stars) data trimming.

illustrated by the following simulation. In the replications under setting (S-IV) with $v_s = 20$, the mean and standard deviation of cosines of the angles between the aggregate contribution vectors ($\mathbf{M}_0$s) based on quasi-curvature under the re-weighting-case scheme and $(\textsc{sifc}(1), \ldots, \textsc{sifc}(n))^{\mathrm{T}}$s are 0.9877 and 0.0085, respectively, whereas the quasi-curvature method under scheme (5.1) using $\mathbf{h}_{\max}$ performs quite differently. Moreover, under the re-weighting-case scheme, the methods using $\mathbf{M}_0$s and $\mathbf{h}_{\max}$ identify all three artificial outliers in only 123 and 80 replications, respectively.

The invariance property can also be illustrated by numerical studies. For example, under (S-IV), we make the transformation $\mathbf{x}_i^* = \mathbf{A}\mathbf{x}_i, i = 1, \ldots, n$ and obtain the influential direction based on $(y_1, \mathbf{x}_1^{*\mathrm{T}}), \ldots, (y_n, \mathbf{x}_n^{*\mathrm{T}})$ under scheme (5.1) with its absolute value vector denoted by $(|h^*_{\max,1}|, \ldots, |h^*_{\max,n}|)^{\mathrm{T}}$, where $\mathbf{A} \hat{=} (a_{ij})$ is a 10×10 matrix with the diagonal elements being 1,3,2,5,4,4,5,2,3,1, $a_{i,i+1} = 1$ for $i = 1, \ldots, 9$ and all the other elements being zeros. The mean of $\sum_{i=1}^n ||h_{\max,i}| - |h^*_{\max,i}||/n$ among the 200 replications is $3.6727 \times e^{-10}$, which appears to be extremely small and is only caused by calculation errors.

Now we conduct a simulation study for the local influence analysis of dMAVE. Inspired by Xia (2007), consider the model

$$Y = \operatorname{sign}(2\mathbf{x}^\top \boldsymbol{\beta}_1 + 0.1\epsilon_1) \log(|2\mathbf{x}^\top \boldsymbol{\beta}_2 + 4 + 0.1\epsilon|),$$

where sign$(\cdot)$ is the sign function. The predictor vector $\mathbf{x} \sim N(\mathbf{0}, \mathbf{I}_p)$ with $p = 10$ and the random errors $\epsilon_1 \sim N(0,1)$ and $\epsilon_2 \sim N(0,1)$ are independent. For $\boldsymbol{\beta}_1$, the first four elements are all 0.5 and the others are zero. For $\boldsymbol{\beta}_2$, the first four elements are $0.5, -0.5, 0.5, -0.5$, respectively, and all the others are zero. One hundred of replications are conducted. In each replication, one hundred of observations are produced from the above model with three artificial outliers. For the artificial outliers, we set $\mathbf{x}_2 = \mathbf{x}_1 + \mathbf{1} * 0.2$, $\mathbf{x}_3 = \mathbf{x}_1 - \mathbf{1} * 0.2$, and $y_i = \text{sign}(2\mathbf{x}_i^\top \boldsymbol{\beta}_1 + 0.1\epsilon_1) \log(|2\mathbf{x}_i^\top \boldsymbol{\beta}_2 + 4 + 0.1\epsilon|) + 10\sigma_Y$, where σ_y is the standard deviation of Y. Here we artificially set the value of $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$ to make the outliers close to each other. We compare dMAVE and SIR in this simulation. Figure S5 presents the plots of $(i, \bar{h}_{a,i})$, where $\bar{h}_{a,i}$ denotes the mean of the absolute values for the ith elements of $\mathbf{h}_{\max}$s among all the replications. For dMAVE, the plot (a) shows that the artificial outliers produce much stronger influence on the central subspace estimate in the sense of average performance. However, in the plot (b), the artificial outliers seems not very influential. It seems that, in this simulation, SIR is more robust than dMAVE. Let $\rho(\hat{\mathcal{B}}, \mathcal{B}) = \text{tr}(\mathbf{P}_{\hat{\mathcal{B}}} \mathbf{P}_{\mathcal{B}})/K$, where $\mathbf{P}_{\mathcal{A}}$ denotes the projection matrix on the subspace $\mathcal{A}$. Then $\rho(\hat{\mathcal{B}}, \mathcal{B})$ can describe the accuracy of $\hat{\mathcal{B}}$ as the estimate of $\mathcal{B}$. For SIR, the mean and standard deviation of $\rho(\hat{\mathcal{B}}, \mathcal{B})$s among the replications are, respectively, 0.908 and 0.055, and under the data set without artificial outliers, they are 0.927 and 0.043. However, for dMAVE, the mean and standard deviation of $\rho(\hat{\mathcal{B}}, \mathcal{B})$s are, respectively, 0.877 and 0.100, while under uncontaminated data set, they are 0.954 and 0.027. The estimating accuracy of dMAVE decreases much sharper than SIR when they suffer the same artificial outliers. That means, in this simulation, the results of the influence analyses for dMAVE and SIR coincide with their performance on the estimating accuracy. In a real data analysis, where the estimating accuracy is generally unknown, the influence analysis may give some useful information for the selection of the central subspace estimate method. Normally, a method with no or less extremely large value of influence measures for the observations may be preferred.

A simulation study is conducted for CUME. We consider the model (S13.2) and the setting (S-III). Under this setting, we compare CUME and SIR. Figure S6 presents the plots of $(i, \bar{h}_{a,i})$ for CUME (a) and SIR (b). For both CUME and SIR, the plots show that the artificial outlier produces much stronger influence on $\hat{\mathcal{B}}$ than the other observations in the sense of average performance, and the difference between CUME and SIR is not significant in this figure. Now we check the decrease of $\rho(\hat{\mathcal{B}}, \mathcal{B})$ produced by the artificial outlier for SIR and CUME, respectively. For SIR, the mean and standard deviation of $\rho(\hat{\mathcal{B}}, \mathcal{B})$s among the replications are, respectively,

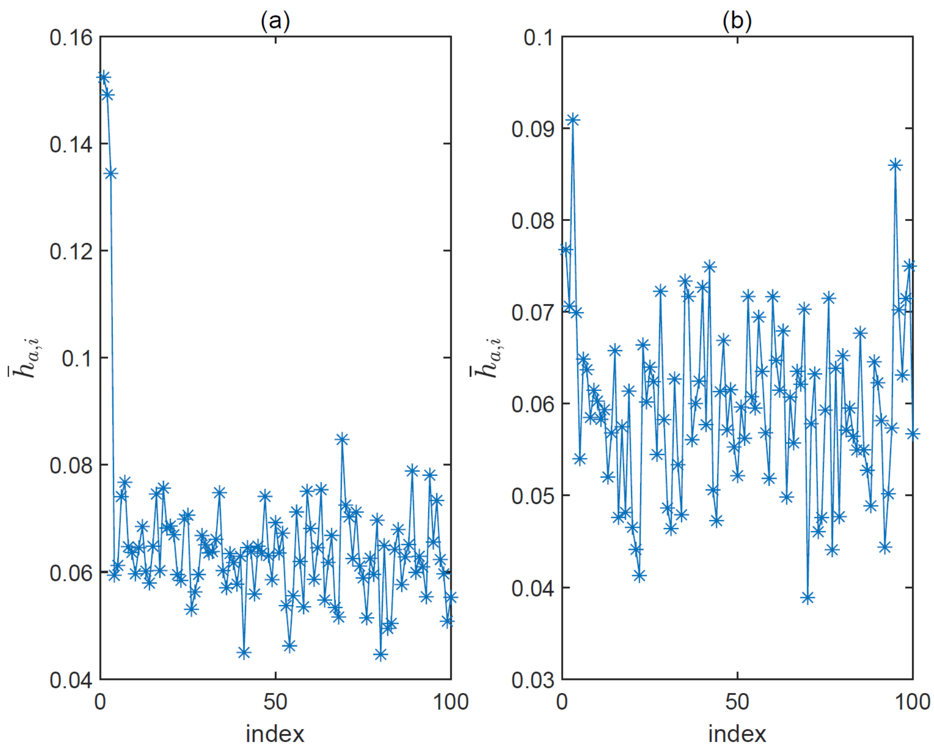

Figure S5: Index plots of mean influence measures for observations given by quasi-curvature method using dMAVE (a) and SIR (b).

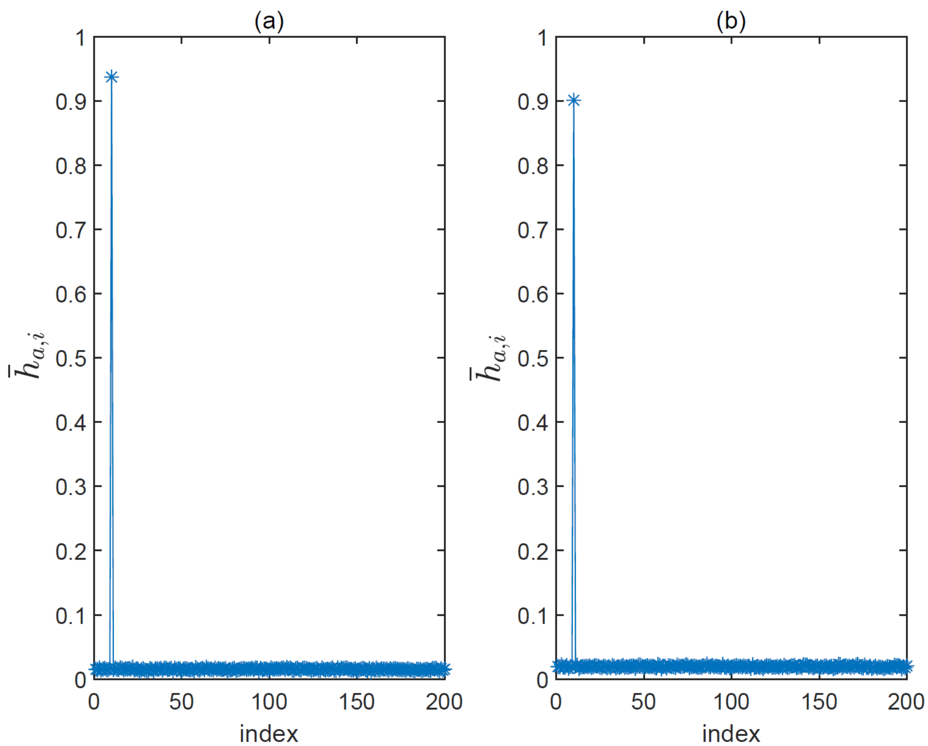

Figure S6: Index plots of mean influence measures for observations given by quasi-curvature method using CUME (a) and SIR (b).

0.902 and 0.039, and under the data set without the artificial outlier, they are 0.933 and 0.028. Similarly, for CUME, the mean and standard deviation of $\rho(\hat{\mathcal{B}}, \mathcal{B})$s are, respectively, 0.907 and 0.037, while under uncontaminated data set, they are 0.936 and 0.028. The artificial outlier leads to similar decreases of the estimating accuracy for CUME and SIR. That means, in this simulation, the results of the influence analyses for CUME and SIR coincide with their performance on the estimating accuracy.

References

Kato, T. (2013). *Perturbation theory for linear operators.* Springer Science & Business Media.

Li, K. C. (1991). Sliced inverse regression for dimension reduction (with discussion). *Journal of the American Statistical Association* **86**, 312–342.

Li, K. C. (2000). *High dimensional data analysis via the SIR/PHD approach.* unpublished manuscript, available at http://www.stat.ucla.edu/~kcli/sir-PHD.pdf.

Prendergast, L. A. (2006). Detecting influential observations in sliced inverse regression analysis. *Australian and New Zealand Journal of Statistics* **48**, 285–304.

Prendergast, L. A. (2007). Implications of influence function analysis for sliced inverse regression and sliced average variance estimation. *Biometrika* **94**, 585–601.

Prendergast, L. A. and J. A. Smith (2010). Influence functions for dimension reduction methods: an example influence study of principal Hessian direction analysis. *Scandinavian Journal* *of Statistics* **37**, 588–611.

Shi, L. and M. Huang (2011). Stepwise local influence analysis. *Computational Statistics and Data Analysis* **55**, 973–982.

Xia, Y. (2007). A constructive approach to the estimation of dimension reduction directions. *The Annals of Statistics* **35**, 2654–2690.

Zhu, H. T. and S. Y. Lee (2001). Local influence for incomplete data models. *Journal of the Royal Statistical Society, Series B* **63**, 111–126.

Zhu, L. P., L. X. Zhu, and Z. H. Feng (2010). Dimension reduction in regressions through cumulative slicing estimation. *Journal of the American Statistical Association* **105**, 1455–1466.

Fei Chen
School of Statistics and Mathematics,
Yunnan University of Finance and Economics,
237 Longquan Road, 650221, Kunming, China
E-mail: chen_fei_77@hotmail.com

Lei Shi
School of Statistics and Mathematics,
Yunnan University of Finance and Economics,
237 Longquan Road, 650221, Kunming, China
E-mail: shi_lei65@hotmail.com

Lin Zhu
Tiantai County Branch of the People's Bank of China
No. 1 West Renmin Road, Tiantai County, Taizhou, Zhejiang, China
E-mail: 783268587@qq.com

Lixing Zhu (corresponding author)
Department of Mathematics,
Hong Kong Baptist University,
Kowloon, Hong Kong, China
E-mail: lzhu@hkbu.edu.hk
Phone: (852) 3411-7016
Fax: (852) 3411-5811