ON SUPERVISED REDUCTION AND ITS DUAL

Peirong Xu and Tao Wang

Shanghai Normal University and Shanghai Jiao Tong University

Supplementary Materials

THE SUPPLEMENTARY FILE CONTAINS THE PROOFS.

PROOF OF PROPOSITION 1. By Proposition 11.1 of Cook (1998),

$$\mathcal{S}_{\mathrm{E}(\boldsymbol{X}|Y)} = \mathrm{span}[\mathrm{Var}\{\mathrm{E}(\boldsymbol{X} \mid Y)\}], \tag{S0.1}$$

the subspace spanned by the columns of $\mathrm{Var}\{\mathrm{E}(\boldsymbol{X} \mid Y)\}$. This, together with condition (C1), implies that for any $\boldsymbol{v} \in \mathcal{S}_{Y|\boldsymbol{X}}$,

$$\boldsymbol{v} - \{\mathrm{Var}(\boldsymbol{X})\}^{-1}\mathrm{Var}\{\mathrm{E}(\boldsymbol{X} \mid Y)\}\boldsymbol{v} \in \mathcal{S}_{Y|\boldsymbol{X}}.$$

By condition (C2) and the law of total covariance,

$$\begin{aligned}\mathrm{Var}(\boldsymbol{X} \mid Y)\boldsymbol{v} &= [\mathrm{Var}(\boldsymbol{X}) - \mathrm{Var}\{\mathrm{E}(\boldsymbol{X} \mid Y)\}]\boldsymbol{v} \\ &= \mathrm{Var}(\boldsymbol{X})[\boldsymbol{v} - \{\mathrm{Var}(\boldsymbol{X})\}^{-1}\mathrm{Var}\{\mathrm{E}(\boldsymbol{X} \mid Y)\}\boldsymbol{v}].\end{aligned}$$

Consequently,

$$\mathrm{Var}(\boldsymbol{X} \mid Y)\mathcal{S}_{Y|\boldsymbol{X}} \subseteq \mathrm{Var}(\boldsymbol{X})\mathcal{S}_{Y|\boldsymbol{X}}. \tag{S0.2}$$

Since $\text{Var}(\boldsymbol{X} \mid Y)$ is positive definite,

$$\text{Var}(\boldsymbol{X})\boldsymbol{v} = \text{Var}(\boldsymbol{X} \mid Y)\boldsymbol{v}^*,$$

where $\boldsymbol{v}^* = \{\text{Var}(\boldsymbol{X} \mid Y)\}^{-1}\text{Var}(\boldsymbol{X})\boldsymbol{v}$. Let $\text{Var}\{\text{E}(\boldsymbol{X} \mid Y)\} = \mathbf{H}\boldsymbol{\Lambda}\mathbf{H}^\top$ be the eigen-decomposition of $\text{Var}\{\text{E}(\boldsymbol{X} \mid Y)\}$. By the matrix inversion lemma,

$$\begin{aligned} & \{\text{Var}(\boldsymbol{X} \mid Y)\}^{-1} \\ = \; & [\text{Var}(\boldsymbol{X}) - \text{Var}\{\text{E}(\boldsymbol{X} \mid Y)\}]^{-1} \\ = \; & \{\text{Var}(\boldsymbol{X})\}^{-1} + \{\text{Var}(\boldsymbol{X})\}^{-1}\mathbf{H}[\boldsymbol{\Lambda}^{-1} - \mathbf{H}^\top\{\text{Var}(\boldsymbol{X})\}^{-1}\mathbf{H}]^{-1}\mathbf{H}^\top\{\text{Var}(\boldsymbol{X})\}^{-1}. \end{aligned}$$

Together with (S0.1) and condition (C1), this implies that $\boldsymbol{v}^* \in \mathcal{S}_{Y|\boldsymbol{X}}$, and hence

$$\text{Var}(\boldsymbol{X})\mathcal{S}_{Y|\boldsymbol{X}} \subseteq \text{Var}(\boldsymbol{X} \mid Y)\mathcal{S}_{Y|\boldsymbol{X}}. \tag{S0.3}$$

Combining (S0.2) and (S0.3), the proof is complete.

Lemma 1. *Assume the conditions of Theorem 1. Then,* $\hat{\boldsymbol{\Delta}}^{-1}$ *is a* $\sqrt{n}$ *consistent estimator of* $\boldsymbol{\Delta}^{-1}$*, and* $\hat{\boldsymbol{\beta}}$ *is a* $\sqrt{n}$ *consistent estimator of* $\boldsymbol{\beta}$ *up to a rotation.*

Proof of Lemma 1. Under the stated assumptions,

$$\begin{aligned}\frac{1}{n}\mathbf{X}^\top\mathbf{X} &= \mathrm{Var}(\boldsymbol{X}) + O_P\left(\frac{1}{\sqrt{n}}\right),\\ \frac{1}{n}\mathbf{F}^\top\mathbf{F} &= \mathrm{Var}(\boldsymbol{f}_Y) + O_P\left(\frac{1}{\sqrt{n}}\right),\\ \frac{1}{n}\mathbf{F}^\top\mathbf{X} &= \mathrm{Cov}(\boldsymbol{f}_Y, \boldsymbol{X}) + O_P\left(\frac{1}{\sqrt{n}}\right).\end{aligned}$$

Hence,

$$\begin{aligned}\hat{\boldsymbol{\Delta}} &= \frac{\mathbf{X}^\top\mathbf{X}}{n} - \frac{\mathbf{X}^\top\mathbf{F}}{n}\left(\frac{\mathbf{F}^\top\mathbf{F}}{n}\right)^{-1}\frac{\mathbf{F}^\top\mathbf{X}}{n}\\ &= \mathrm{Var}(\boldsymbol{X}) - \mathrm{Cov}(\boldsymbol{X}, \boldsymbol{f}_Y)\{\mathrm{Var}(\boldsymbol{f}_Y)\}^{-1}\mathrm{Cov}(\boldsymbol{f}_Y, \boldsymbol{X}) + O_P\left(\frac{1}{\sqrt{n}}\right).\end{aligned}$$

Note that

$$\mathrm{Var}(\boldsymbol{X}) = \boldsymbol{\Gamma}\boldsymbol{\beta}\mathrm{Var}(\boldsymbol{f}_Y)\boldsymbol{\beta}^\top\boldsymbol{\Gamma}^\top + \boldsymbol{\Delta}, \tag{S0.4}$$

and

$$\mathrm{Cov}(\boldsymbol{f}_Y, \boldsymbol{X}) = \mathrm{Var}(\boldsymbol{f}_Y)\boldsymbol{\beta}^\top\boldsymbol{\Gamma}^\top. \tag{S0.5}$$

We have

$$\begin{aligned}\hat{\boldsymbol{\Delta}} &= \boldsymbol{\Gamma}\boldsymbol{\beta}\mathrm{Var}(\boldsymbol{f}_Y)\boldsymbol{\beta}^\top\boldsymbol{\Gamma}^\top + \boldsymbol{\Delta} - \boldsymbol{\Gamma}\boldsymbol{\beta}\mathrm{Var}(\boldsymbol{f}_Y)\boldsymbol{\beta}^\top\boldsymbol{\Gamma}^\top + O_P\left(\frac{1}{\sqrt{n}}\right)\\ &= \boldsymbol{\Delta} + O_P\left(\frac{1}{\sqrt{n}}\right),\end{aligned}$$

and hence

$$\hat{\boldsymbol{\Delta}}^{-1} = \boldsymbol{\Delta}^{-1} + O_P\left(\frac{1}{\sqrt{n}}\right).$$

Similarly,

$$
\begin{aligned}
& (\mathbf{F}^\top\mathbf{F})^{-1/2}\mathbf{F}^\top\mathbf{X}\hat{\boldsymbol{\Delta}}^{-1}\mathbf{X}^\top\mathbf{F}(\mathbf{F}^\top\mathbf{F})^{-1/2} \\
= \;& \{\mathrm{Var}(\boldsymbol{f}_Y)\}^{-1/2}\mathrm{Cov}(\boldsymbol{f}_Y, \boldsymbol{X})\boldsymbol{\Delta}^{-1}\mathrm{Cov}(\boldsymbol{X}, \boldsymbol{f}_Y)\{\mathrm{Var}(\boldsymbol{f}_Y)\}^{-1/2} + O_P\left(\frac{1}{\sqrt{n}}\right) \\
= \;& \{\mathrm{Var}(\boldsymbol{f}_Y)\}^{1/2}\boldsymbol{\beta}^\top\boldsymbol{\Gamma}^\top\boldsymbol{\Delta}^{-1}\boldsymbol{\Gamma}\boldsymbol{\beta}\{\mathrm{Var}(\boldsymbol{f}_Y)\}^{1/2} + O_P\left(\frac{1}{\sqrt{n}}\right) \\
= \;& \{\mathrm{Var}(\boldsymbol{f}_Y)\}^{1/2}\boldsymbol{\beta}^\top\boldsymbol{\beta}\{\mathrm{Var}(\boldsymbol{f}_Y)\}^{1/2} + O_P\left(\frac{1}{\sqrt{n}}\right),
\end{aligned}
$$

where the last equality follows because $\boldsymbol{\Gamma}^\top\boldsymbol{\Delta}^{-1}\boldsymbol{\Gamma} = \mathbf{I}_d$. This implies that

$$
\hat{\boldsymbol{\beta}}^\top\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}^\top\boldsymbol{\beta} + O_P\left(\frac{1}{\sqrt{n}}\right).
$$

The proof is complete.

Proof of Theorem 1. Note that

$$
\frac{1}{n}\hat{\mathbf{V}}^\top\hat{\boldsymbol{s}} = \frac{1}{n}\sum_{i=1}^{n}\hat{\boldsymbol{v}}_{y_i}\hat{s}_i = \hat{\boldsymbol{\beta}}\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{f}_{y_i}\hat{s}_i\right)
$$

and

$$
\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{f}_{y_i}\hat{s}_i &= \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{f}_{y_i}(\|\hat{\boldsymbol{v}}_{y_i}\|_2^2 - \|\hat{\boldsymbol{\Delta}}^{-1/2}(\boldsymbol{x}_{y^*} - \boldsymbol{x}_{y_i})\|_2^2) \\
&= \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{f}_{y_i}\|\hat{\boldsymbol{v}}_{y_i}\|_2^2 - \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{f}_{y_i}\|\hat{\boldsymbol{\Delta}}^{-1/2}(\boldsymbol{x}_{y^*} - \boldsymbol{x}_{y_i})\|_2^2 \\
&= T_1 - T_2.
\end{aligned}
$$

Consider the first term. By Lemma 1,

$$
T_1 = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{f}_{y_i}\boldsymbol{f}_{y_i}^\top\hat{\boldsymbol{\beta}}^\top\hat{\boldsymbol{\beta}}\boldsymbol{f}_{y_i} = \mathrm{E}(\boldsymbol{f}_Y\boldsymbol{f}_Y^\top\boldsymbol{\beta}^\top\boldsymbol{\beta}\boldsymbol{f}_Y) + O_P\left(\frac{1}{\sqrt{n}}\right). \qquad \text{(S0.6)}
$$

Consider the second term. We have

$$
\begin{aligned}
\|\hat{\boldsymbol{\Delta}}^{-1/2}(\boldsymbol{x}_{y^*} - \boldsymbol{x}_{y_i})\|_2^2 &= \|\hat{\boldsymbol{\Delta}}^{-1/2}(\boldsymbol{\Gamma}\boldsymbol{v}_{y^*} + \boldsymbol{\epsilon}_{y^*} - \boldsymbol{\Gamma}\boldsymbol{v}_{y_i} - \boldsymbol{\epsilon}_{y_i})\|_2^2 \\
&= (\boldsymbol{v}_{y^*} - \boldsymbol{v}_{y_i})^\top \boldsymbol{\Gamma}^\top \hat{\boldsymbol{\Delta}}^{-1} \boldsymbol{\Gamma}(\boldsymbol{v}_{y^*} - \boldsymbol{v}_{y_i}) \\
&\quad + 2(\boldsymbol{v}_{y^*} - \boldsymbol{v}_{y_i})^\top \boldsymbol{\Gamma}^\top \hat{\boldsymbol{\Delta}}^{-1}(\boldsymbol{\epsilon}_{y^*} - \boldsymbol{\epsilon}_{y_i}) \\
&\quad + (\boldsymbol{\epsilon}_{y^*} - \boldsymbol{\epsilon}_{y_i})^\top \hat{\boldsymbol{\Delta}}^{-1}(\boldsymbol{\epsilon}_{y^*} - \boldsymbol{\epsilon}_{y_i}),
\end{aligned}
$$

and hence

$$
\begin{aligned}
T_2 &= \frac{1}{n}\sum_{i=1}^n \boldsymbol{f}_{y_i}(\boldsymbol{v}_{y^*} - \boldsymbol{v}_{y_i})^\top \boldsymbol{\Gamma}^\top \hat{\boldsymbol{\Delta}}^{-1} \boldsymbol{\Gamma}(\boldsymbol{v}_{y^*} - \boldsymbol{v}_{y_i}) \\
&\quad + \frac{2}{n}\sum_{i=1}^n \boldsymbol{f}_{y_i}(\boldsymbol{v}_{y^*} - \boldsymbol{v}_{y_i})^\top \boldsymbol{\Gamma}^\top \hat{\boldsymbol{\Delta}}^{-1}(\boldsymbol{\epsilon}_{y^*} - \boldsymbol{\epsilon}_{y_i}) \\
&\quad + \frac{1}{n}\sum_{i=1}^n \boldsymbol{f}_{y_i}(\boldsymbol{\epsilon}_{y^*} - \boldsymbol{\epsilon}_{y_i})^\top \hat{\boldsymbol{\Delta}}^{-1}(\boldsymbol{\epsilon}_{y^*} - \boldsymbol{\epsilon}_{y_i}) \\
&= T_{21} + T_{22} + T_{23}.
\end{aligned}
$$

By Lemma 1,

$$
\begin{aligned}
T_{21} &= \frac{1}{n}\sum_{i=1}^n \boldsymbol{f}_{y_i}(\boldsymbol{f}_{y^*} - \boldsymbol{f}_{y_i})^\top \boldsymbol{\beta}^\top \boldsymbol{\Gamma}^\top \hat{\boldsymbol{\Delta}}^{-1} \boldsymbol{\Gamma}\boldsymbol{\beta}(\boldsymbol{f}_{y^*} - \boldsymbol{f}_{y_i}) \\
&= -2\mathrm{Var}(\boldsymbol{f}_Y)\boldsymbol{\beta}^\top \boldsymbol{\beta}\boldsymbol{f}_{y^*} + \mathrm{E}(\boldsymbol{f}_Y \boldsymbol{f}_Y^\top \boldsymbol{\beta}^\top \boldsymbol{\beta}\boldsymbol{f}_Y) + O_P\left(\frac{1}{\sqrt{n}}\right).
\end{aligned}
\tag{S0.7}
$$

Similarly,

$$
\begin{aligned}
T_{22} &= \frac{2}{n}\sum_{i=1}^n \boldsymbol{f}_{y_i}(\boldsymbol{f}_{y^*} - \boldsymbol{f}_{y_i})^\top \boldsymbol{\beta}^\top \boldsymbol{\Gamma}^\top \hat{\boldsymbol{\Delta}}^{-1}(\boldsymbol{\epsilon}_{y^*} - \boldsymbol{\epsilon}_{y_i}) \\
&= -2\mathrm{Var}(\boldsymbol{f}_Y)\boldsymbol{\beta}^\top \boldsymbol{\Gamma}^\top \boldsymbol{\Delta}^{-1}\boldsymbol{\epsilon}_{y^*} + O_P\left(\frac{1}{\sqrt{n}}\right)
\end{aligned}
\tag{S0.8}
$$

and

$$T_{23} = O_P\left(\frac{1}{\sqrt{n}}\right). \tag{S0.9}$$

From (S0.6)-(S0.9), we have

$$\frac{1}{n}\hat{\mathbf{V}}^\top \hat{\boldsymbol{s}} = 2\mathbf{R}\boldsymbol{\beta}\mathrm{Var}(\boldsymbol{f}_Y)\boldsymbol{\beta}^\top\boldsymbol{\beta}\boldsymbol{f}_{y^*} + 2\mathbf{R}\boldsymbol{\beta}\mathrm{Var}(\boldsymbol{f}_Y)\boldsymbol{\beta}^\top\boldsymbol{\Gamma}^\top\boldsymbol{\Delta}^{-1}\boldsymbol{\epsilon}_{y^*} + O_P\left(\frac{1}{\sqrt{n}}\right) \tag{S0.10}$$

for some $d \times d$ rotation matrix $\mathbf{R}$. Note that $\hat{\mathbf{V}}^\top = \hat{\boldsymbol{\beta}}\mathbf{F}^\top$. By Lemma 1,

$$\frac{1}{n}\hat{\mathbf{V}}^\top\hat{\mathbf{V}} = \hat{\boldsymbol{\beta}}\left(\frac{1}{n}\mathbf{F}^\top\mathbf{F}\right)\hat{\boldsymbol{\beta}}^\top = \mathbf{R}\boldsymbol{\beta}\mathrm{Var}(\boldsymbol{f}_Y)\boldsymbol{\beta}^\top\mathbf{R}^\top + O_P\left(\frac{1}{\sqrt{n}}\right). \tag{S0.11}$$

Combining (S0.10) and (S0.11),

$$\hat{\boldsymbol{v}}_{y^*} = \mathbf{R}\boldsymbol{\beta}\boldsymbol{f}_{y^*} + \mathbf{R}\boldsymbol{\Gamma}^\top\boldsymbol{\Delta}^{-1}\boldsymbol{\epsilon}_{y^*} + O_P\left(\frac{1}{\sqrt{n}}\right).$$

The proof is complete.

Proof of Corollary 1. By Theorem 1, there exists a rotation matrix $\mathbf{R}$, such that

$$\hat{\boldsymbol{v}}_{y^*} = \mathbf{R}\boldsymbol{v}_{y^*} + \mathbf{R}\boldsymbol{\Gamma}^\top\boldsymbol{\Delta}^{-1}\boldsymbol{\epsilon}_{y^*} + O_P\left(\frac{1}{\sqrt{n}}\right).$$

Let $\tilde{\boldsymbol{v}}_{Y^*} = \mathbf{R}\boldsymbol{v}_{Y^*} + \mathbf{R}\boldsymbol{\Gamma}^\top\boldsymbol{\Delta}^{-1}\boldsymbol{\epsilon}_{Y^*}$. Then, by the independence of Y^* and $\boldsymbol{\epsilon}_{Y^*}$,

$$\mathrm{Var}(\tilde{\boldsymbol{v}}_{Y^*}) = \mathbf{R}\{\mathrm{Var}(\boldsymbol{v}_{Y^*}) + \mathbf{I}_d\}\mathbf{R}^\top$$

and

$$\mathrm{Cov}(\tilde{\boldsymbol{v}}_{Y^*}, \boldsymbol{v}_{Y^*}) = \mathbf{R}\mathrm{Var}(\boldsymbol{v}_{Y^*}).$$

It follows that

$$\begin{aligned}\rho^2(\tilde{\boldsymbol{v}}_{Y^*}, \boldsymbol{v}_{Y^*}) &= \frac{1}{d}\,\mathrm{trace}[\mathbf{R}\mathrm{Var}(\boldsymbol{v}_{Y^*})\{\mathrm{Var}(\boldsymbol{v}_{Y^*})\}^{-1}\mathrm{Var}(\boldsymbol{v}_{Y^*})\mathbf{R}^\top\mathbf{R}\{\mathrm{Var}(\boldsymbol{v}_{Y^*}) + \mathbf{I}_d\}^{-1}\mathbf{R}^\top] \\ &= \frac{1}{d}\,\mathrm{trace}[\mathrm{Var}(\boldsymbol{v}_{Y^*})\{\mathrm{Var}(\boldsymbol{v}_{Y^*}) + \mathbf{I}_d\}^{-1}].\end{aligned}$$

The proof is complete.

Lemma 2. *Assume the conditions of Theorem 2. Then,* $\hat{\boldsymbol{\Delta}}^{-1}$ *is a* $\sqrt{n}$ *consistent estimator of* $\boldsymbol{\Omega}^{-1}$*, and* $\hat{\boldsymbol{\beta}}$ *is a* $\sqrt{n}$ *consistent estimator of* $\boldsymbol{\Phi}$ *up to a rotation.*

Proof of Lemma 2. We mimic the proof of Lemma 1. Under the stated conditions,

$$\begin{aligned}\hat{\boldsymbol{\Delta}} &= \mathrm{Var}(\boldsymbol{X}) - \boldsymbol{\Gamma}\mathrm{Cov}(\boldsymbol{v}_Y, \boldsymbol{f}_Y)\{\mathrm{Var}(\boldsymbol{f}_Y)\}^{-1}\mathrm{Cov}(\boldsymbol{f}_Y, \boldsymbol{v}_Y)\boldsymbol{\Gamma}^\top + O_P\left(\frac{1}{\sqrt{n}}\right) \\ &= \boldsymbol{\Omega} + O_P\left(\frac{1}{\sqrt{n}}\right).\end{aligned}$$

It is easy to verify that $\boldsymbol{\Omega}$ is positive definite. Hence

$$\hat{\boldsymbol{\Delta}}^{-1} = \boldsymbol{\Omega}^{-1} + O_P\left(\frac{1}{\sqrt{n}}\right).$$

Together with the constraint $\boldsymbol{\Gamma}^\top\boldsymbol{\Omega}^{-1}\boldsymbol{\Gamma} = \mathbf{I}_d$ (which reduces to $\boldsymbol{\Gamma}^\top\boldsymbol{\Delta}^{-1}\boldsymbol{\Gamma} = \mathbf{I}_d$, if $\boldsymbol{v}_y$ is correctly specified as $\boldsymbol{v}_y = \boldsymbol{\beta}\boldsymbol{f}_y$), this implies that

$$\begin{aligned}&(\mathbf{F}^\top\mathbf{F})^{-1/2}\mathbf{F}^\top\mathbf{X}\hat{\boldsymbol{\Delta}}^{-1}\mathbf{X}^\top\mathbf{F}(\mathbf{F}^\top\mathbf{F})^{-1/2} \\ = \;& \{\mathrm{Var}(\boldsymbol{f}_Y)\}^{-1/2}\mathrm{Cov}(\boldsymbol{f}_Y, \boldsymbol{v}_Y)\boldsymbol{\Gamma}^\top\boldsymbol{\Omega}^{-1}\boldsymbol{\Gamma}\mathrm{Cov}(\boldsymbol{v}_Y, \boldsymbol{f}_Y)\{\mathrm{Var}(\boldsymbol{f}_Y)\}^{-1/2} + O_P\left(\frac{1}{\sqrt{n}}\right) \\ = \;& \{\mathrm{Var}(\boldsymbol{f}_Y)\}^{-1/2}\mathrm{Cov}(\boldsymbol{f}_Y, \boldsymbol{v}_Y)\mathrm{Cov}(\boldsymbol{v}_Y, \boldsymbol{f}_Y)\{\mathrm{Var}(\boldsymbol{f}_Y)\}^{-1/2} + O_P\left(\frac{1}{\sqrt{n}}\right).\end{aligned}$$

Consequently,

$$\hat{\boldsymbol{\beta}}^\top \hat{\boldsymbol{\beta}} = \{\mathrm{Var}(\boldsymbol{f}_Y)\}^{-1}\mathrm{Cov}(\boldsymbol{f}_Y, \boldsymbol{v}_Y)\mathrm{Cov}(\boldsymbol{v}_Y, \boldsymbol{f}_Y)\{\mathrm{Var}(\boldsymbol{f}_Y)\}^{-1} + O_P\left(\frac{1}{\sqrt{n}}\right).$$

The proof is complete.

Proof of Theorem 2. Recall that

$$\frac{1}{n}\hat{\mathbf{V}}^\top \hat{\boldsymbol{s}} = \hat{\boldsymbol{\beta}}\left(\frac{1}{n}\sum_{i=1}^n \boldsymbol{f}_{y_i}\hat{s}_i\right)$$

and

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^n \boldsymbol{f}_{y_i}\hat{s}_i &= \frac{1}{n}\sum_{i=1}^n \boldsymbol{f}_{y_i}\boldsymbol{f}_{y_i}^\top \hat{\boldsymbol{\beta}}^\top \hat{\boldsymbol{\beta}}\boldsymbol{f}_{y_i} \\
&\quad -\frac{1}{n}\sum_{i=1}^n \boldsymbol{f}_{y_i}(\boldsymbol{v}_{y^*} - \boldsymbol{v}_{y_i})^\top \boldsymbol{\Gamma}^\top \hat{\boldsymbol{\Delta}}^{-1}\boldsymbol{\Gamma}(\boldsymbol{v}_{y^*} - \boldsymbol{v}_{y_i}) \\
&\quad -\frac{2}{n}\sum_{i=1}^n \boldsymbol{f}_{y_i}(\boldsymbol{v}_{y^*} - \boldsymbol{v}_{y_i})^\top \boldsymbol{\Gamma}^\top \hat{\boldsymbol{\Delta}}^{-1}(\boldsymbol{\epsilon}_{y^*} - \boldsymbol{\epsilon}_{y_i}) \\
&\quad -\frac{1}{n}\sum_{i=1}^n \boldsymbol{f}_{y_i}(\boldsymbol{\epsilon}_{y^*} - \boldsymbol{\epsilon}_{y_i})^\top \hat{\boldsymbol{\Delta}}^{-1}(\boldsymbol{\epsilon}_{y^*} - \boldsymbol{\epsilon}_{y_i}) \\
&= T_1 - (T_{21} + T_{22} + T_{23}).
\end{aligned}$$

By Lemma 2,

$$T_1 = \frac{1}{n}\sum_{i=1}^n \boldsymbol{f}_{y_i}\boldsymbol{f}_{y_i}^\top \hat{\boldsymbol{\beta}}^\top \hat{\boldsymbol{\beta}}\boldsymbol{f}_{y_i} = \mathrm{E}(\boldsymbol{f}_Y\boldsymbol{f}_Y^\top \boldsymbol{\Phi}^\top \boldsymbol{\Phi}\boldsymbol{f}_Y) + O_P\left(\frac{1}{\sqrt{n}}\right), \tag{S0.12}$$

$$T_{21} = -2\mathrm{Cov}(\boldsymbol{f}_Y, \boldsymbol{v}_Y)\boldsymbol{v}_{y^*} + \mathrm{E}(\boldsymbol{f}_Y\boldsymbol{v}_Y^\top \boldsymbol{v}_Y) + O_P\left(\frac{1}{\sqrt{n}}\right), \tag{S0.13}$$

$$T_{22} = -2\mathrm{Cov}(\boldsymbol{f}_Y, \boldsymbol{v}_Y)\boldsymbol{\Gamma}^\top \boldsymbol{\Omega}^{-1}\boldsymbol{\epsilon}_{y^*} + O_P\left(\frac{1}{\sqrt{n}}\right), \tag{S0.14}$$

$$T_{23} = O_P\left(\frac{1}{\sqrt{n}}\right). \tag{S0.15}$$

From (S0.12)-(S0.15), we have

$$\begin{aligned}\frac{1}{n}\hat{\mathbf{V}}^\top\hat{\boldsymbol{s}} &= \mathbf{R\Phi}\mathrm{E}(\boldsymbol{f}_Y\boldsymbol{f}_Y^\top\boldsymbol{\Phi}^\top\boldsymbol{\Phi}\boldsymbol{f}_Y) - \mathbf{R\Phi}\mathrm{E}(\boldsymbol{f}_Y\boldsymbol{v}_Y^\top\boldsymbol{v}_Y) \\ &\quad +2\mathbf{R\Phi}\mathrm{Cov}(\boldsymbol{f}_Y,\boldsymbol{v}_Y)\boldsymbol{v}_{y^*} + 2\mathbf{R\Phi}\mathrm{Cov}(\boldsymbol{f}_Y,\boldsymbol{v}_Y)\boldsymbol{\Gamma}^\top\boldsymbol{\Omega}^{-1}\boldsymbol{\epsilon}_{y^*} + O_P\left(\frac{1}{\sqrt{n}}\right), \quad \text{(S0.16)}\end{aligned}$$

for some $d\times d$ rotation matrix $\mathbf{R}$. By Lemma 2,

$$\frac{1}{n}\hat{\mathbf{V}}^\top\hat{\mathbf{V}} = \hat{\boldsymbol{\beta}}\left(\frac{1}{n}\mathbf{F}^\top\mathbf{F}\right)\hat{\boldsymbol{\beta}}^\top = \mathbf{R\Phi}\mathrm{Var}(\boldsymbol{f}_Y)\boldsymbol{\Phi}^\top\mathbf{R}^\top + O_P\left(\frac{1}{\sqrt{n}}\right). \quad \text{(S0.17)}$$

Combining (S0.16) and (S0.17),

$$\hat{\boldsymbol{v}}_{y^*} = \mathbf{R}\boldsymbol{c} + \mathbf{R}\mathbf{A}\boldsymbol{v}_{y^*} + \mathbf{RA}\boldsymbol{\Gamma}^\top\boldsymbol{\Omega}^{-1}\boldsymbol{\epsilon}_{y^*} + O_P\left(\frac{1}{\sqrt{n}}\right).$$

The proof is complete.

Proof of Theorem 3. For the moment we assume the conditions of Theorem 1. By Lemma 1,

$$\hat{\boldsymbol{\Delta}}^{-1}\hat{\boldsymbol{\Gamma}} = \boldsymbol{\Delta}^{-1}\mathrm{Cov}(\boldsymbol{X},\boldsymbol{f}_Y)\boldsymbol{\beta}^\top\{\boldsymbol{\beta}\mathrm{Var}(\boldsymbol{f}_Y)\boldsymbol{\beta}^\top\}^{-1} + O_P\left(\frac{1}{\sqrt{n}}\right).$$

This, together with (S0.5), implies that

$$\begin{aligned}\hat{\boldsymbol{\Delta}}^{-1}\hat{\boldsymbol{\Gamma}} &= \boldsymbol{\Delta}^{-1}\boldsymbol{\Gamma}\boldsymbol{\beta}\mathrm{Var}(\boldsymbol{f}_Y)\boldsymbol{\beta}^\top\{\boldsymbol{\beta}\mathrm{Var}(\boldsymbol{f}_Y)\boldsymbol{\beta}^\top\}^{-1} + O_P\left(\frac{1}{\sqrt{n}}\right) \\ &= \boldsymbol{\Delta}^{-1}\boldsymbol{\Gamma} + O_P\left(\frac{1}{\sqrt{n}}\right).\end{aligned}$$

Therefore, $\mathrm{span}(\hat{\boldsymbol{\Delta}}^{-1}\hat{\boldsymbol{\Gamma}})$ is a $\sqrt{n}$ consistent estimate of $\mathcal{S}_{Y|\boldsymbol{X}}$.

We now give the proof under the conditions of Theorem 2. By Lemma 2,

$$\hat{\boldsymbol{\Delta}}^{-1}\hat{\boldsymbol{\Gamma}} = \boldsymbol{\Omega}^{-1}\boldsymbol{\Gamma}\mathrm{Cov}(\boldsymbol{v}_Y,\boldsymbol{f}_Y)\boldsymbol{\Phi}^\top\{\boldsymbol{\Phi}\mathrm{Var}(\boldsymbol{f}_Y)\boldsymbol{\Phi}^\top\}^{-1} + O_P\left(\frac{1}{\sqrt{n}}\right).$$

Consequently, $\text{span}(\hat{\mathbf{\Delta}}^{-1}\hat{\mathbf{\Gamma}})$ is a $\sqrt{n}$ consistent estimate of $\text{span}(\mathbf{\Omega}^{-1}\mathbf{\Gamma})$.

We first show that

$$\text{span}(\mathbf{\Omega}^{-1}\mathbf{\Gamma}) = \text{span}[\{\text{Var}(\boldsymbol{X})\}^{-1}\mathbf{\Gamma}].$$

Let $\mathbf{C} = \text{Cov}(\boldsymbol{v}_Y, \boldsymbol{f}_Y), \mathbf{\Sigma}_{\boldsymbol{v}} = \text{Var}(\boldsymbol{v}_Y), \mathbf{\Sigma}_{\boldsymbol{f}} = \text{Var}(\boldsymbol{f}_Y)$, and $\mathbf{\Sigma}_{\boldsymbol{X}} = \text{Var}(\boldsymbol{X})$. By the Woodbury matrix identity,

$$\mathbf{\Omega}^{-1} = \mathbf{\Sigma}_{\boldsymbol{X}}^{-1} + \mathbf{\Sigma}_{\boldsymbol{X}}^{-1}\mathbf{\Gamma}\mathbf{C}(\mathbf{\Sigma}_{\boldsymbol{f}} - \mathbf{C}^\top\mathbf{\Gamma}^\top\mathbf{\Sigma}_{\boldsymbol{X}}^{-1}\mathbf{\Gamma}\mathbf{C})^{-1}\mathbf{C}^\top\mathbf{\Gamma}^\top\mathbf{\Sigma}_{\boldsymbol{X}}^{-1}.$$

We can then write

$$\begin{aligned}
\mathbf{\Omega}^{-1}\mathbf{\Gamma} &= \mathbf{\Sigma}_{\boldsymbol{X}}^{-1}\mathbf{\Gamma} + \mathbf{\Sigma}_{\boldsymbol{X}}^{-1}\mathbf{\Gamma}\mathbf{C}(\mathbf{\Sigma}_{\boldsymbol{f}} - \mathbf{C}^\top\mathbf{\Gamma}^\top\mathbf{\Sigma}_{\boldsymbol{X}}^{-1}\mathbf{\Gamma}\mathbf{C})^{-1}\mathbf{C}^\top\mathbf{\Gamma}^\top\mathbf{\Sigma}_{\boldsymbol{X}}^{-1}\mathbf{\Gamma} \\
&= \mathbf{\Sigma}_{\boldsymbol{X}}^{-1}\mathbf{\Gamma}\mathbf{H},
\end{aligned}$$

where $\mathbf{H} = \mathbf{I}_d + \mathbf{C}[\mathbf{\Sigma}_{\boldsymbol{f}} - \mathbf{C}^\top\mathbf{\Gamma}^\top\mathbf{\Sigma}_{\boldsymbol{X}}^{-1}\mathbf{\Gamma}\mathbf{C}]^{-1}\mathbf{C}^\top\mathbf{\Gamma}^\top\mathbf{\Sigma}_{\boldsymbol{X}}^{-1}\mathbf{\Gamma}$. This implies that $\mathbf{H}$ is non-singular, and that $\mathbf{\Omega}^{-1}\mathbf{\Gamma}$ and $\mathbf{\Sigma}_{\boldsymbol{X}}^{-1}\mathbf{\Gamma}$ have the same column subspace.

It remains to show that

$$\text{span}[\{\text{Var}(\boldsymbol{X})\}^{-1}\mathbf{\Gamma}] = \text{span}(\mathbf{\Delta}^{-1}\mathbf{\Gamma}).$$

Notice that $\text{Var}(\boldsymbol{X}) = \mathbf{\Gamma}\mathbf{\Sigma}_{\boldsymbol{v}}\mathbf{\Gamma}^\top + \mathbf{\Delta}$. Let $\mathbf{A} = (\mathbf{\Gamma}^\top\mathbf{\Delta}^{-1}\mathbf{\Gamma})^{-1}$. By the Woodbury matrix identity,

$$\begin{aligned}
\{\text{Var}(\boldsymbol{X})\}^{-1}\mathbf{\Gamma} &= \mathbf{\Delta}^{-1}\mathbf{\Gamma} - \mathbf{\Delta}^{-1}\mathbf{\Gamma}(\mathbf{\Sigma}_{\boldsymbol{v}}^{-1} + \mathbf{\Gamma}^\top\mathbf{\Delta}^{-1}\mathbf{\Gamma})^{-1}\mathbf{\Gamma}^\top\mathbf{\Delta}^{-1}\mathbf{\Gamma} \\
&= \mathbf{\Delta}^{-1}\mathbf{\Gamma} - \mathbf{\Delta}^{-1}\mathbf{\Gamma}\{\mathbf{A} - \mathbf{A}(\mathbf{\Sigma}_{\boldsymbol{v}} + \mathbf{A})^{-1}\mathbf{A}\}\mathbf{\Gamma}^\top\mathbf{\Delta}^{-1}\mathbf{\Gamma} \\
&= \mathbf{\Delta}^{-1}\mathbf{\Gamma}\mathbf{A}(\mathbf{\Sigma}_{\boldsymbol{v}} + \mathbf{A})^{-1}.
\end{aligned}$$

Because $\mathbf{A}(\mathbf{\Sigma}_v + \mathbf{A})^{-1}$ is non-singular, the proof is complete.

PROOF OF THEOREM 4. Let $\hat{\mathbf{\Delta}}^{-1/2}\mathbf{X}^\top\mathbf{F}(\mathbf{F}^\top\mathbf{F})^{-1/2} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^\top$ denote the singular value decomposition of $\hat{\mathbf{\Delta}}^{-1/2}\mathbf{X}^\top\mathbf{F}(\mathbf{F}^\top\mathbf{F})^{-1/2}$; that is, $\mathbf{U} = (\mathbf{U}_1, \ldots, \mathbf{U}_r)$ is $p \times r$ with orthonormal columns, $\mathbf{V} = (\mathbf{V}_1, \ldots, \mathbf{V}_r)$ is $r \times r$ orthogonal, and $\mathbf{\Lambda}$ is an $r \times r$ diagonal matrix with diagonal entries $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r \geq 0$. Let $\mathbf{\Psi} = (\mathbf{U}_1, \ldots, \mathbf{U}_d)$. Then, span($\mathbf{\Psi}$) is the subspace spanned by the first d eigenvectors of $\hat{\mathbf{\Delta}}^{-1/2}\hat{\mathbf{\Sigma}}\hat{\mathbf{\Delta}}^{-1/2}$.

Let $\mathbf{\Phi} = (\lambda_1\mathbf{V}_1, \ldots, \lambda_d\mathbf{V}_d)^\top$. By definition, $\hat{\boldsymbol{\beta}}(\mathbf{F}^\top\mathbf{F})^{1/2} = \mathbf{\Phi}$. Hence

$$\begin{aligned}
\text{span}(\hat{\mathbf{\Delta}}^{-1/2}\hat{\mathbf{\Gamma}}) &= \text{span}(\hat{\mathbf{\Delta}}^{-1/2}\mathbf{X}^\top\mathbf{F}\hat{\boldsymbol{\beta}}^\top) \\
&= \text{span}\{\hat{\mathbf{\Delta}}^{-1/2}\mathbf{X}^\top\mathbf{F}(\mathbf{F}^\top\mathbf{F})^{-1/2}\mathbf{\Phi}^\top\} \\
&= \text{span}(\mathbf{\Psi}).
\end{aligned}$$

This proves the first part. The second part follows from Corollary 3.4 of Cook and Forzani (2008). The proof is complete.

PROOF OF THEOREM 5. Let $\mathbf{A} = \{\text{Var}(\boldsymbol{f}_Y)\}^{-1/2}\text{Cov}(\boldsymbol{f}_Y, \boldsymbol{v}_Y)$. Under the stated conditions, $\mathbf{A}$ has full column rank, d_0. Furthermore, by Lemma 1 and Lemma 2,

$$(\mathbf{F}^\top\mathbf{F})^{-1/2}\mathbf{F}^\top\mathbf{X}\hat{\mathbf{\Delta}}^{-1}\mathbf{X}^\top\mathbf{F}(\mathbf{F}^\top\mathbf{F})^{-1/2} = \mathbf{A}\mathbf{A}^\top + O_P\left(\frac{1}{\sqrt{n}}\right).$$

The rest of the proof can be found in Zhu et al. (2012). The proof is complete.

The predictive equivalence of SRIR and PFC. In the following we show that if the sample multiple correlation coefficient is used for measuring predictive performance, then SRIR and PFC are equivalent.

Let $\hat{\mathbf{V}}^*_{SRIR}$ and $\hat{\mathbf{V}}^*_{PFC}$ be predicted coordinates of m new observations $\{\boldsymbol{x}_{y_1^*}, \ldots, \boldsymbol{x}_{y_m^*}\}$ by SRIR and PFC. Let $\hat{\boldsymbol{\Sigma}}_{SRIR}$, $\hat{\boldsymbol{\Sigma}}_{PFC}$, and $\hat{\boldsymbol{\Sigma}}_{SRIR,PFC}$ be the sample covariance matrix of $\hat{\mathbf{V}}^*_{SRIR}$, the sample covariance matrix of $\hat{\mathbf{V}}^*_{PFC}$, and the sample covariance matrix between $\hat{\mathbf{V}}^*_{SRIR}$ and $\hat{\mathbf{V}}^*_{PFC}$, respectively. By definition, the squared sample multiple correlation coefficient

$$MCC^2(\hat{\mathbf{V}}^*_{SRIR}, \hat{\mathbf{V}}^*_{PFC}) = \frac{1}{d}\,\text{trace}(\hat{\boldsymbol{\Sigma}}_{SRIR,PFC}\hat{\boldsymbol{\Sigma}}_{PFC}^{-1}\hat{\boldsymbol{\Sigma}}_{SRIR,PFC}^{\top}\hat{\boldsymbol{\Sigma}}_{SRIR}^{-1}).$$

Without loss of generality, assume that $\hat{\mathbf{V}}^*_{SRIR}$ and $\hat{\mathbf{V}}^*_{PFC}$ are centered. Then it is easy to check that

$$MCC^2(\hat{\mathbf{V}}^*_{SRIR}, \hat{\mathbf{V}}^*_{PFC}) = \frac{1}{d}\,\text{trace}(\mathbf{P}_{SRIR}\mathbf{P}_{PFC}),$$

where $\mathbf{P}_{SRIR}$ and $\mathbf{P}_{PFC}$ are projection matrices onto the column spaces of $\hat{\mathbf{V}}^*_{SRIR}$ and $\hat{\mathbf{V}}^*_{PFC}$, respectively. It remains to prove that $\mathbf{P}_{SRIR} = \mathbf{P}_{PFC}$.

Write $\hat{\mathbf{V}}^*_{SRIR} = (\boldsymbol{u}_1^*, \ldots, \boldsymbol{u}_m^*)^\top$ and $\hat{\mathbf{V}}^*_{PFC} = (\boldsymbol{w}_1^*, \ldots, \boldsymbol{w}_m^*)^\top$. By (4.6), up to a common constant vector, $\boldsymbol{u}_j^* = (\hat{\mathbf{V}}^\top\hat{\mathbf{V}})^{-1}\hat{\mathbf{V}}^\top\mathbf{X}\hat{\boldsymbol{\Delta}}^{-1}\boldsymbol{x}_{y_j^*}$. On the other hand, $\boldsymbol{w}_j^* = \boldsymbol{\Psi}^\top\hat{\boldsymbol{\Delta}}^{-1/2}\boldsymbol{x}_{y_j^*}$. By the definition of $\hat{\boldsymbol{\Gamma}}$, $\hat{\boldsymbol{\Delta}}^{-1}\mathbf{X}^\top\hat{\mathbf{V}}(\hat{\mathbf{V}}^\top\hat{\mathbf{V}})^{-1} = \hat{\boldsymbol{\Delta}}^{-1}\hat{\boldsymbol{\Gamma}}$. From the proof of Theorem 4, it follows that $\text{span}\{\hat{\boldsymbol{\Delta}}^{-1}\mathbf{X}^\top\hat{\mathbf{V}}(\hat{\mathbf{V}}^\top\hat{\mathbf{V}})^{-1}\} = \text{span}\{\hat{\boldsymbol{\Delta}}^{-1/2}\boldsymbol{\Psi}\}$. The proof is complete.

Bibliography

Cook, R. D. (1998). *Regression Graphics: Ideas for Studying Regressions Through Graphics*. Wiley, New York.

Cook, R. D. and L. Forzani (2008). Principal fitted components for dimension reduction in regression. *Statistical Science 23*(4), 485–501.

Zhu, L., L. Zhu, and Z. Feng (2012). Dimension reduction in regressions through cumulative slicing estimation. *Journal of the American Statistical Association 105*(492), 1455–1466.