Dimension Reduction for Extreme Regression via Contour Projection

Liujun Chen and Jing Zeng

University of Science and Technology of China

Supplementary Material

In the Supplementary Material, we provide additional discussions in Sections S1–S3, including the discussions on the kernel matrix in COPES-DR, technical Assumption (A1), and additional asymptotic results of COPES methods. The additional numerical results are presented in Section S4. In Section S5, we present technical proofs.

S1 Kernel matrix in COPES-DR and its sample estimator

We show that the matrix $\mathbf{M}^h_{\text{eDR}}$ can be re-expressed as $\mathbf{M}^h_{\text{eDR}} = 2\sum_{i=1}^{8} \mathbf{K}_i$, where the expressions of $\mathbf{K}_i$, $i = 1, \ldots, 8$, are

$$\mathbf{K}_1 = \frac{1}{3}\int_0^1 \mathbf{D}_h(u)\mathbf{D}_h^\top(u)\mathrm{d}u, \quad \mathbf{K}_2 = \left(\int_0^1 u\mathbf{D}_h(u)\mathrm{d}u\right)^2, \quad \mathbf{K}_3 = \left(\int_0^1 \mathbf{C}_h(u)\mathbf{C}_h^\top(u)\mathrm{d}u\right)^2,$$

$$\mathbf{K}_4 = \left(\int_0^1 \mathbf{C}_h^\top(u)\mathbf{C}_h(u)\mathrm{d}u\right)\left(\int_0^1 \mathbf{C}_h(u)\mathbf{C}_h^\top(u)\mathrm{d}u\right),$$

$$\mathbf{K}_5 = -\left(\int_0^1 \mathbf{D}_h(u)\mathbf{C}_h(u)\mathrm{d}u\right)\left(\int_0^1 u\mathbf{C}_h^\top(u)\mathrm{d}u\right),$$

$$\mathbf{K}_6 = -\left\{\int_0^1 \mathbf{D}_h(u)\left(\int_0^1 u\mathbf{C}_h(u)\mathrm{d}u\right)\mathbf{C}_h^\top(u)\mathrm{d}u\right\}, \quad \mathbf{K}_7 = \mathbf{K}_5^\top, \quad \mathbf{K}_8 = \mathbf{K}_6^\top.$$

We first decompose $\mathbf{G}_h(u, u^*)$ as follows,

$$\begin{aligned}
\mathbf{G}_h(u,u^*) =& \frac{1}{h^2}\mathbb{E}\left\{(\overrightarrow{\mathbf{X}} - \overrightarrow{\mathbf{X}}^*)(\overrightarrow{\mathbf{X}} - \overrightarrow{\mathbf{X}}^*)^\top I(\widetilde{Y} < F^{-1}(uh))I(\widetilde{Y}^\star < (F^*)^{-1}(u^*h))\right\} \\
&- u^* \cdot \text{median}\{\boldsymbol{\sigma}(\boldsymbol{\Sigma}^{-1/2}\mathbf{T}_h(u)\boldsymbol{\Sigma}^{-1/2})\}\boldsymbol{\Sigma} - u \cdot \text{median}\{\boldsymbol{\sigma}(\boldsymbol{\Sigma}^{-1/2}\mathbf{T}_h(u^*)\boldsymbol{\Sigma}^{-1/2})\}\boldsymbol{\Sigma} \\
=& u^*\frac{1}{h}\mathbb{E}\left\{\overrightarrow{\mathbf{X}}\overrightarrow{\mathbf{X}}^\top I(\widetilde{Y} < F^{-1}(uh))\right\} + u\frac{1}{h}\mathbb{E}\left\{\overrightarrow{\mathbf{X}}^*(\overrightarrow{\mathbf{X}}^*)^\top I(\widetilde{Y}^* < (F^*)^{-1}(u^*h))\right\} \\
&- \frac{1}{h^2}\mathbb{E}\left\{\overrightarrow{\mathbf{X}} I(\widetilde{Y} < F^{-1}(uh))\right\}\mathbb{E}\left\{(\overrightarrow{\mathbf{X}}^*)^\top I(\widetilde{Y}^* < (F^*)^{-1}(u^*h))\right\} \\
&- \frac{1}{h^2}\mathbb{E}\left\{(\overrightarrow{\mathbf{X}}^*) I(\widetilde{Y}^* < (F^*)^{-1}(u^*h))\right\}\mathbb{E}\left\{\overrightarrow{\mathbf{X}}^\top I(\widetilde{Y} < F^{-1}(uh))\right\} \\
&- u^* \cdot \text{median}\{\boldsymbol{\sigma}(\boldsymbol{\Sigma}^{-1/2}\mathbf{T}_h(u)\boldsymbol{\Sigma}^{-1/2})\}\boldsymbol{\Sigma} - u \cdot \text{median}\{\boldsymbol{\sigma}(\boldsymbol{\Sigma}^{-1/2}\mathbf{T}_h(u^*)\boldsymbol{\Sigma}^{-1/2})\}\boldsymbol{\Sigma} \\
=& u^*\mathbf{D}_h(u) + u\mathbf{D}_h(u^*) - \mathbf{C}_h(u)\mathbf{C}_h^\top(u^*) - \mathbf{C}_h(u^*)\mathbf{C}_h^\top(u).
\end{aligned}$$

Then,

$$\begin{aligned}
\mathbf{M}_{\mathrm{DR}}^h =& \int_0^1 \int_0^1 \mathbf{G}_h(u,u^*)\mathbf{G}_h^\top(u,u^*)\mathrm{d}u\mathrm{d}u^* \\
=& \int_0^1 \int_0^1 (u^*)^2 \mathbf{D}_h(u)\mathbf{D}_h^\top(u)\mathrm{d}u\mathrm{d}u^* + \int_0^1 \int_0^1 u^* u \mathbf{D}_h(u)\mathbf{D}_h^\top(u^*)\mathrm{d}u\mathrm{d}u^* \\
&- \int_0^1 \int_0^1 u^* \mathbf{D}_h(u)\mathbf{C}_h(u^*)\mathbf{C}_h^\top(u)\mathrm{d}u\mathrm{d}u^* - \int_0^1 \int_0^1 u^* \mathbf{D}_h(u)\mathbf{C}_h(u)\mathbf{C}_h^\top(u^*)\mathrm{d}u\mathrm{d}u^* \\
&+ \int_0^1 \int_0^1 u^2 \mathbf{D}_h(u^*)\mathbf{D}_h^\top(u^*)\mathrm{d}u\mathrm{d}u^* + \int_0^1 \int_0^1 u^* u \mathbf{D}_h(u^*)\mathbf{D}_h^\top(u)\mathrm{d}u\mathrm{d}u^* \\
&- \int_0^1 \int_0^1 u \mathbf{D}_h(u^*)\mathbf{C}_h(u)\mathbf{C}_h^\top(u^*)\mathrm{d}u\mathrm{d}u^* - \int_0^1 \int_0^1 u \mathbf{D}_h(u^*)\mathbf{C}_h(u^*)\mathbf{C}_h^\top(u)\mathrm{d}u\mathrm{d}u^* \\
&- \int_0^1 \int_0^1 u^* \mathbf{C}_h(u)\mathbf{C}_h^\top(u^*)\mathbf{D}_h^\top(u)\mathrm{d}u\mathrm{d}u^* - \int_0^1 \int_0^1 u \mathbf{C}_h(u)\mathbf{C}_h^\top(u^*)\mathbf{D}_h^\top(u^*)\mathrm{d}u\mathrm{d}u^* \\
&+ \int_0^1 \int_0^1 \mathbf{C}_h(u)\mathbf{C}_h^\top(u^*)\mathbf{C}_h(u^*)\mathbf{C}_h^\top(u)\mathrm{d}u\mathrm{d}u^* + \int_0^1 \int_0^1 \mathbf{C}_h(u)\mathbf{C}_h^\top(u^*)\mathbf{C}_h(u)\mathbf{C}_h^\top(u^*)\mathrm{d}u\mathrm{d}u^* \\
&- \int_0^1 \int_0^1 u \mathbf{C}_h(u^*)\mathbf{C}_h^\top(u)\mathbf{D}_h^\top(u^*)\mathrm{d}u\mathrm{d}u^* - \int_0^1 \int_0^1 u^* \mathbf{C}_h(u^*)\mathbf{C}_h^\top(u)\mathbf{D}_h^\top(u)\mathrm{d}u\mathrm{d}u^* \\
&+ \int_0^1 \int_0^1 \mathbf{C}_h(u^*)\mathbf{C}_h^\top(u)\mathbf{C}_h(u)\mathbf{C}_h^\top(u^*)\mathrm{d}u\mathrm{d}u^* + \int_0^1 \int_0^1 \mathbf{C}_h(u^*)\mathbf{C}_h^\top(u)\mathbf{C}_h(u^*)\mathbf{C}_h^\top(u)\mathrm{d}u\mathrm{d}u^* \\
=& 2\sum_{i=1}^{8} \mathbf{K}_i.
\end{aligned}$$

Note that $\mathbf{K}_3$ follows from the fact that

$$\begin{aligned}
\int_0^1 \int_0^1 \mathbf{C}_h(u^*)\mathbf{C}_h^\top(u)\mathbf{C}_h(u^*)\mathbf{C}_h^\top(u)\mathrm{d}u\mathrm{d}u^* &= \int_0^1 \int_0^1 \mathbf{C}_h(u^*)\mathbf{C}_h^\top(u^*)\mathbf{C}_h(u)\mathbf{C}_h^\top(u)\mathrm{d}u\mathrm{d}u^* \\
&= \left(\int_0^1 \mathbf{C}_h(u)\mathbf{C}_h^\top(u)\mathrm{d}u\right)\left(\int_0^1 \mathbf{C}_h(u^*)\mathbf{C}_h^\top(u^*)\mathrm{d}u^*\right),
\end{aligned}$$

where the first equation holds since $\mathbf{C}_h^\top(u)\mathbf{C}_h(u^*) = \mathbf{C}_h^\top(u^*)\mathbf{C}_h(u)$ always holds.

Consequently, the sample estimator of the kernel matrix is

$$\widehat{\mathbf{M}}_{\mathrm{eDR}}^{k/n} = 2\sum_{i=1}^{8} \widehat{\mathbf{K}}_i^{k/n},$$

where each component $\widehat{\mathbf{K}}_i^{k/n}$, $i = 1, \dots, 8$, is as follows.

$$\widehat{\mathbf{K}}_1 = \frac{1}{3k}\sum_{m=1}^{k} \widehat{\mathbf{D}}_{k/n}\left(\frac{m}{k}\right)\widehat{\mathbf{D}}_{k/n}^{\top}\left(\frac{m}{k}\right), \quad \widehat{\mathbf{K}}_2 = \left(\frac{1}{k}\sum_{m=1}^{k} \frac{m}{k}\widehat{\mathbf{D}}_{k/n}\left(\frac{m}{k}\right)\right)^2$$

$$\widehat{\mathbf{K}}_3 = \left(\frac{1}{k}\sum_{m=1}^{k} \widehat{\mathbf{C}}_{k/n}\left(\frac{m}{k}\right)\widehat{\mathbf{C}}_{k/n}^{\top}\left(\frac{m}{k}\right)\right)^2,$$

$$\widehat{\mathbf{K}}_4 = \left(\frac{1}{k}\sum_{m=1}^{k} \widehat{\mathbf{C}}_{k/n}^{\top}\left(\frac{m}{k}\right)\widehat{\mathbf{C}}_{k/n}\left(\frac{m}{k}\right)\right)\left(\frac{1}{k}\sum_{m=1}^{k} \widehat{\mathbf{C}}_{k/n}\left(\frac{m}{k}\right)\widehat{\mathbf{C}}_{k/n}^{\top}\left(\frac{m}{k}\right)\right),$$

$$\widehat{\mathbf{K}}_5 = -\left(\frac{1}{k}\sum_{m=1}^{k} \widehat{\mathbf{D}}_{k/n}\left(\frac{m}{k}\right)\widehat{\mathbf{C}}_{k/n}\left(\frac{m}{k}\right)\right)\left(\frac{1}{k}\sum_{m=1}^{k} \frac{m}{k}\widehat{\mathbf{C}}_{k/n}^{\top}\left(\frac{m}{k}\right)\right),$$

$$\widehat{\mathbf{K}}_6 = -\left\{\frac{1}{k}\sum_{m=1}^{k} \widehat{\mathbf{D}}_{k/n}\left(\frac{m}{k}\right)\left(\frac{1}{k}\sum_{m=1}^{k} \frac{m}{k}\widehat{\mathbf{C}}_{k/n}\left(\frac{m}{k}\right)\right)\widehat{\mathbf{C}}_{k/n}^{\top}\left(\frac{m}{k}\right)\right\},$$

$$\widehat{\mathbf{K}}_7 = \widehat{\mathbf{K}}_5^{\top}, \quad \widehat{\mathbf{K}}_8 = \widehat{\mathbf{K}}_6^{\top}.$$

S2 Discussion on Assumption (A1)

Aghbalou et al. (2024) assumed that $\lim_{y\to y^+} \mathbb{E}(\mathbf{X}|Y > y)$, which does not hold for many common distributions. We consider a toy example on $\mathbb{E}(X|Y > y)$ where the dimension $p = 1$. Note that $\mathbb{E}(X|Y > y)$ is related to the concept of marginal expected shortfall (Acharya et al. 2017) in systemic risk management.

Assume that the two random variables $X, Y \in \mathbb{R}$ are asymptotically dependent in the sense that for any $(x, y) \in [0, \infty]^2 \setminus \{(\infty, \infty)\}$, the following limit exists and is not identically zero,

$$\lim_{t \to \infty} t \Pr(1 - F_X(X) \leq x/t, 1 - F_Y(Y) \leq y/t),$$

where F_X and F_Y are cumulative distribution functions for X and Y, respectively. Assuming that X has a heavy right-hand tail, it has been established in Cai et al. (2015) that the following asymptotic limit exists,

$$\lim_{y \to y^+} \frac{\mathbb{E}(X \mid Y > y)}{Q(F_Y(y))} \to c,$$

where $c > 0$ is some constant and Q is the quantile function of X such that $Q(\alpha) = \inf\{x \in \mathbb{R} : \Pr(X \leq x) \geq \alpha\}$ for any $0 \leq \alpha \leq 1$. Thus, as y approaches y^+, the conditional expectation $\mathbb{E}(X \mid Y > y) \to \infty$, resulting in the collapse of the convergence condition. Even in the case where X follows the normal distribution, Hua & Joe (2014) showed that with some additional conditions on the tail dependence of Y and X, we still have $\mathbb{E}(X \mid Y > y) \to \infty$ as $y \to y^+$.

Assuming that X is from the univariate EC distribution. When we replace X with the contour-projected predictor $\overrightarrow{X} = \text{sign}(X)\sigma$, where σ^2 is the scatter parameter, the tail moment becomes to

$$\mathbb{E}(\overrightarrow{X} \mid Y > y) = \sigma \mathbb{E}(\text{sign}(X) \mid Y > y) = 2\sigma \Pr\{\text{sign}(X) = 1 \mid Y > y\} - 1.$$

Then, the convergence assumption on $\mathbb{E}(X \mid Y > y)$ reduces to assuming the convergence of the conditional probability $\Pr\{\text{sign}(X) = 1 \mid Y > y\}$ as $y \to y^{+}$, which is fairly mild in most applications.

S3 Additional asymptotic results

We first present the refinement of Theorem 5 under new assumptions. Then, parallel to the development of the asymptotic theory of COPES-DR, we develop the asymptotic theories of COPES-SIR and COPES-SAVE.

Recall that F denotes the cumulative distribution function of $\widetilde{Y} = -Y$. A function $f(y)$ is called *eventually decreasing* if there exists a constant y_0, such that $f(y)$ is decreasing for $y \geq y_0$. We introduce the following two assumptions, adapted from Assumptions (A1) and (A2).

(A1′) Let $\mathbf{a}(y) = \mathbb{E}(\overrightarrow{\mathbf{X}} \mid Y > y) - \boldsymbol{\nu}$ with $\boldsymbol{\nu}$ in Assumption (A1). Assume that $\|\mathbf{a}(y)\|$ is an eventually decreasing function such that $\sqrt{k}\|\mathbf{a}\{-F^{-1}(k/n)\}\| = O(1)$ as $n \to \infty$.

(A2′) Let $\mathbf{b}(y) := \mathbb{E}(\overrightarrow{\mathbf{X}}\overrightarrow{\mathbf{X}}^{\top} | Y > y) - \mathbf{T}$ with $\mathbf{T}$ in Assumption (A2). Assume that $\|\mathbf{b}(y)\|_F$ is an eventually decreasing function such that $\sqrt{k}\|\mathbf{b}\{-F^{-1}(k/n)\}\|_F = O(1)$ as $n \to \infty$.

Corollary S1. *Assume the same assumptions in Theorem 5 and that Assumptions (A1′) & (A2′) hold. Then, as $n \to \infty$, we have (i) $\|\widehat{\mathbf{M}}_{\text{eDR}}^{k/n} -$*

$\mathbf{M}_{\mathrm{eDR}}\|_F = O_P(k^{-1/2})$; *(ii)* $\|\mathbf{P}_{\widehat{\boldsymbol{\beta}}^{k/n}_{\mathrm{eDR}}} - \mathbf{P}_{\mathcal{S}_{Y_\infty|\mathbf{X}}}\|_F = O_P(k^{-1/2})$.

Assumption (A1′) specifically quantifies the convergence rate of $\mathbb{E}(\overrightarrow{\mathbf{X}} \mid Y > y) - \boldsymbol{\nu}$ in Assumption (A1) and imposes restrictions on the choice of k to control the biases of the estimators; namely, k cannot be too large. Similar conditions are often assumed in the literature on extreme value statistics; see, for example, Lalancette et al. (2021) and de Haan & Ferreira (2006). Assumption (A2′) is proposed in the same spirit as Assumption (A1′), specifying the convergence rate of $\mathbb{E}(\overrightarrow{\mathbf{X}}\overrightarrow{\mathbf{X}}^\top|Y > y) - \mathbf{T}$. With these additional convergence rate assumptions, $\widehat{\mathbf{M}}^{k/n}_{\mathrm{eDR}}$ and $\mathrm{span}(\widehat{\boldsymbol{\beta}}^{k/n}_{\mathrm{eDR}})$ exhibit $\sqrt{k}$-consistency in estimating $\mathbf{M}_{\mathrm{eDR}}$ and $\mathcal{S}_{Y_\infty|\mathbf{X}}$.

The following two theorems claim the consistency results for the estimated kernel matrices and working subspaces in COPES-SIR and COPES-SAVE.

Theorem S1. *Assume that* $\mathbf{X}$ *follows the EC distribution and Assumption (A1) & (A3) hold. Moreover, we assume that* $\mathrm{Cov}(\overrightarrow{\mathbf{X}}|Y > y)$ *converges as* $y \to y^+$*. Then, as* $n \to \infty$*, we have (i)* $\|\widehat{\mathbf{M}}^{k/n}_{\mathrm{eSIR}} - \mathbf{M}^{k/n}_{\mathrm{eSIR}}\|_F = O_P(k^{-1/2})$*; (ii)* $\|\widehat{\mathbf{M}}^{k/n}_{\mathrm{eSIR}} - \mathbf{M}_{\mathrm{eSIR}}\|_F = o_P(1)$*; (iii)* $\|\mathbf{P}_{\widehat{\boldsymbol{\beta}}^{k/n}_{\mathrm{eSIR}}} - \mathbf{P}_{\mathcal{S}_{\mathrm{eSIR}}}\|_F = o_P(1)$.

Theorem S2. *Assume that* $\mathbf{X}$ *follows the EC distribution and Assumptions (A2) & (A3) hold. Moreover, we assume that* $\mathrm{Cov}\{\mathrm{vec}(\overrightarrow{\mathbf{X}}\overrightarrow{\mathbf{X}}^\top)|Y > y\}$ *converges as* $y \to y^+$*. Then, as* $n \to \infty$*, we have (i)* $\|\widehat{\mathbf{M}}^{k/n}_{\mathrm{eSAVE}} - \mathbf{M}^{k/n}_{\mathrm{eSAVE}}\|_F =$

$O_P(k^{-1/2})$; *(ii)* $\|\widehat{\mathbf{M}}^{k/n}_{\text{eSAVE}} - \mathbf{M}_{\text{eSAVE}}\|_F = o_P(1)$. *By further assuming that* $\dim(\mathcal{S}_{\text{eSAVE}}) < p/2$, *we have (iii)* $\|\mathbf{P}_{\widehat{\boldsymbol{\beta}}^{k/n}_{\text{eSAVE}}} - \mathbf{P}_{\mathcal{S}_{Y_\infty|\mathbf{X}}}\|_F = o_P(1)$.

With additional convergence Assumptions (A1′) and (A2′), a refined description of the convergence properties for $\mathbf{P}_{\widehat{\boldsymbol{\beta}}^{k/n}_{\text{eSIR}}}$, $\mathbf{P}_{\widehat{\boldsymbol{\beta}}^{k/n}_{\text{eSAVE}}}$, $\text{span}(\widehat{\boldsymbol{\beta}}^{k/n}_{\text{eSIR}})$, and $\text{span}(\widehat{\boldsymbol{\beta}}^{k/n}_{\text{eSAVE}})$ is available, presented in the following two corollaries.

Corollary S2. *Assume the same assumptions in Theorem S1 and that Assumption (A1′) holds. Then, as* $n \to \infty$, *we have (i)* $\|\widehat{\mathbf{M}}^{k/n}_{\text{eSIR}} - \mathbf{M}_{\text{eSIR}}\|_F = O_P(k^{-1/2})$; *(ii)* $\|\mathbf{P}_{\widehat{\boldsymbol{\beta}}^{k/n}_{\text{eSIR}}} - \mathbf{P}_{\mathcal{S}_{\text{eSIR}}}\|_F = O_P(k^{-1/2})$.

Corollary S3. *Assume the same assumptions in Theorem S2 and that Assumption (A2′) holds. Then, as* $n \to \infty$, *we have (i)* $\|\widehat{\mathbf{M}}^{k/n}_{\text{eSAVE}} - \mathbf{M}_{\text{eSAVE}}\|_F = O_P(k^{-1/2})$; *(ii)* $\|\mathbf{P}_{\widehat{\boldsymbol{\beta}}^{k/n}_{\text{eSAVE}}} - \mathbf{P}_{\mathcal{S}_{Y_\infty|\mathbf{X}}}\|_F = O_P(k^{-1/2})$.

S4 Additional numerical results

S4.1 The discussion on numerical performance of COPES

We provide more details about the comparison among the three specific COPES methods. In summary, COPES-DR combines the advantages of COPES-SIR and COPES-SAVE, dominating other competitors in most situations. And in scenarios where $d^* > 1$, COPES-SAVE and COPES-DR are better choices. Specifically, in Model A, where there is a monotone trend in the more contributing component of the model, COPES-SIR exhibits the most favorable estimation performance and COPES-DR perform comparably to COPES-SIR. In comparison, COPES-SAVE performs worse. In Model C, where the monotone trend is absent, COPES-SAVE performs comparably with COPES-SIR, while COPES-DR achieves the best performance. In Models B and D, where $d^* = 2$, the dimension-deficient method COPES-SIR performs the worst. COPES-DR is, once again, the best competitor in either heavy-tailed or light-tailed cases. In fact, under the multivariate normal distribution, similar patterns in the four models have also been observed by their SDR counterparts, SIR, SAVE, and DR, see Li & Wang (2007), for example. Thus, these COPES methods partly inherit the characteristics of their counterparts in SDR.

S4.2 MSE plots under Models A–D

The MSE of the subspace estimation plots under Models A, C, and D are presented in Figures S1–S3.

S4.3 Accuracy of structural dimension determination

In this section, we evaluate the accuracy of the dimension determination based on COPES-SAVE and COPES-DR. The procedure has been de-

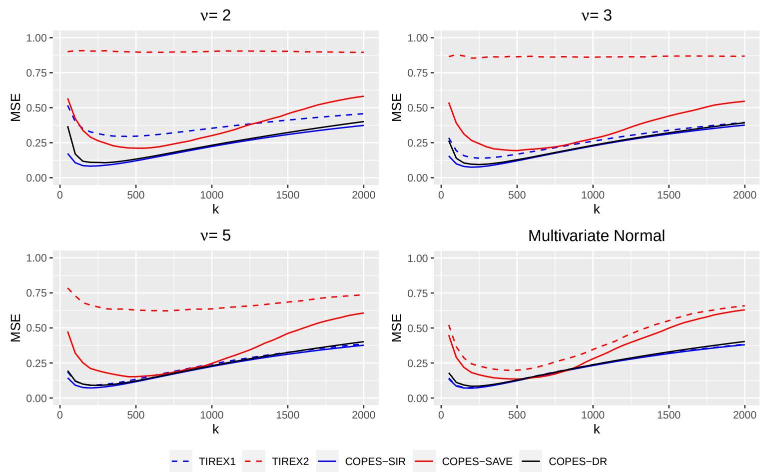

Figure S1: MSEs for different competitors under various ν's under Model A.

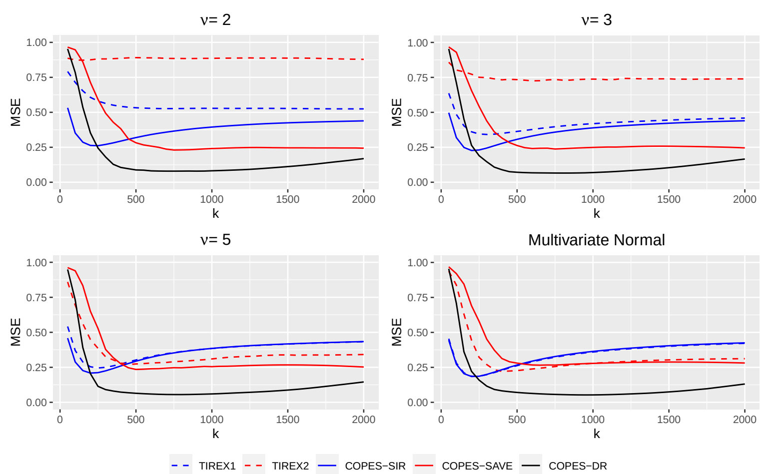

Figure S2: MSEs for different competitors under various ν's under Model C.

tailedly discussed in Section 5.4.

We set the ratio adjustment constant $\varepsilon = 10^{-5}$. In each model, we record the correct dimension selection ratio. We consider three sample sizes, $n = 5000$, 10000, and 20000, and four k's, including $k = 2[n^{0.6}]$, $k = [n^{2/3}]$,

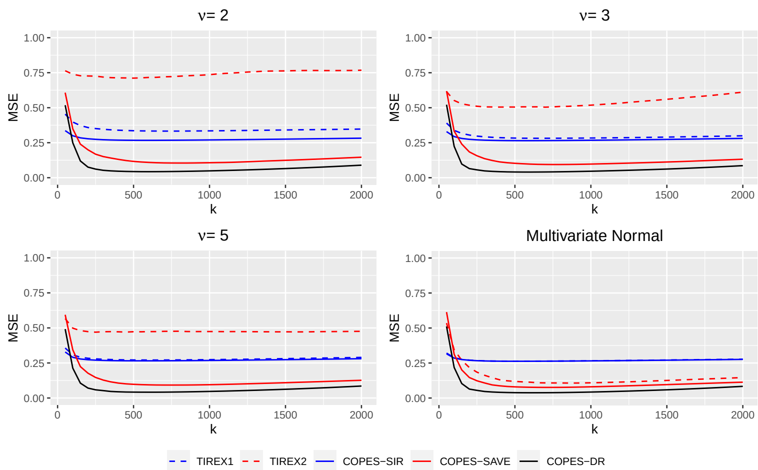

Figure S3: MSEs for different competitors under various ν's under Model D.

$k = [n^{0.7}]$, and $k = [\frac{2}{3}n^{0.75}]$, where $[a]$ denotes the largest integer less or equal to a. In Table S1, we report the dimension determination results of COPES-DR under the scenario where the degree of freedom $\nu = 3$. The results for COPES-SAVE and other ν's yield similar findings and are not reported here.

Generally speaking, when the sample size n is large enough, and k is selected at some reasonable ratio, our procedure achieves accurate determination of the structural dimension. Specifically, for fixed k, the correct selection ratio steadily increases as n increases. For fixed n, the correct selection ratio attains the peak for some properly selected k. Thus, when samples are efficient, our method provides a simple and accurate way of determining the structural dimension. The high accuracy also suggests that

Table S1: Correct selection ratio based on COPES-DR under the scenario where $\nu = 3$.

k	$n = 5000$	$n = 10000$	$n = 20000$	$n = 5000$	$n = 10000$	$n = 20000$
	Model A			Model B		
$2[n^{0.6}]$	0.965	1	1	0.875	0.96	1
$[n^{2/3}]$	0.965	1	1	0.85	0.96	1
$[n^{0.7}]$	0.975	0.995	1	0.9	0.98	1
$[\frac{2}{3}n^{0.75}]$	0.975	0.995	1	0.9	0.98	1
	Model C			Model D		
$2[n^{0.6}]$	0.685	0.905	0.98	0.92	0.99	0.995
$[n^{2/3}]$	0.58	0.88	0.98	0.855	0.985	0.995
$[n^{0.7}]$	0.765	0.96	0.995	0.94	1	1
$[\frac{2}{3}n^{0.75}]$	0.76	0.955	0.995	0.94	1	1

the extreme SAVE subspace $\mathcal{S}_{\text{eSAVE}}$ and extreme DR subspace $\mathcal{S}_{\text{eDR}}$ coincide with the CES $\mathcal{S}_{Y_\infty|\mathbf{X}}$, indicating the exhaustiveness of two extreme subspaces.

S4.4 Simulation results with non-identity scatter matrix

In this subsection, we examine the performance of our methods under Models A–D with $\mathbf{X} = \mathbf{W}/\sqrt{u/\nu}$, where $\mathbf{W} \sim N(\mathbf{0}, \mathbf{\Sigma})$, $u \sim \chi^2_\nu$, and u is independent of $\mathbf{W}$. Here, we consider $\mathbf{\Sigma} = (\sigma_{ij})$ with $\sigma_{ij} = 0.5^{|i-j|}$ for $1 \le i \le j \le p$. All other settings remain consistent with those in Section 6. The MSE of subspace estimation for each method is presented in Figures S4–S7.

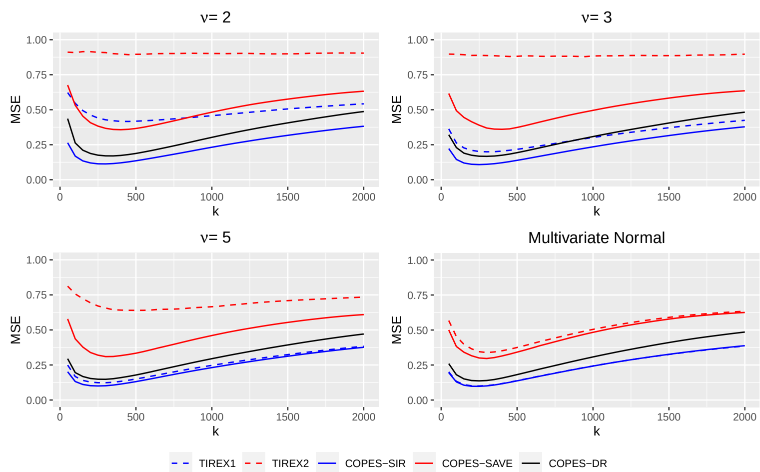

Figure S4: MSEs for different competitors under various ν's under Model A for autoregressive scatter matrix.

The results indicate patterns similar to those observed in Section 6. Specifically, the COPES methods outperform the TIREX methods under the heavy-tailed distribution, while they demonstrate comparable performance under the normal distribution.

S4.5 Simulation results with non-EC distribution

Our proposal assumes the EC distribution for predictors. One intuitive question is how the COPES methods perform when such EC distribution is violated. To this end, we independently generate each covariate X_j, $j = 1, \ldots, p$, from t distribution with degree of freedom ν. We repeat all experiments in Section 6. The results under four models with $\nu = 2, 3, 5$

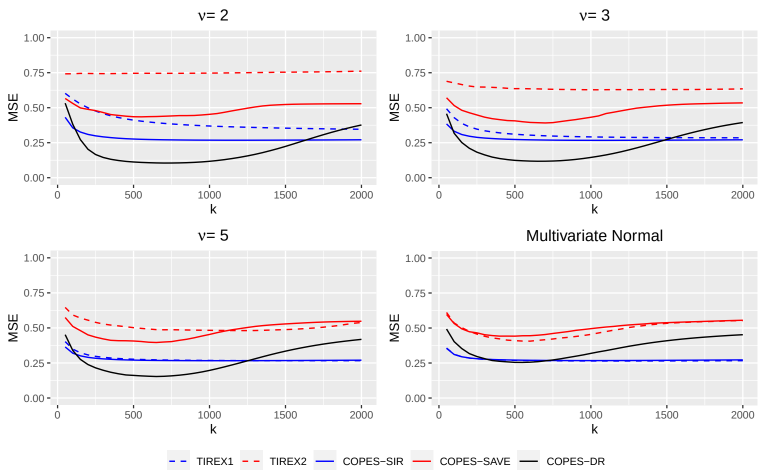

Figure S5: MSEs for different competitors under various ν's under Model B for autoregressive scatter matrix.

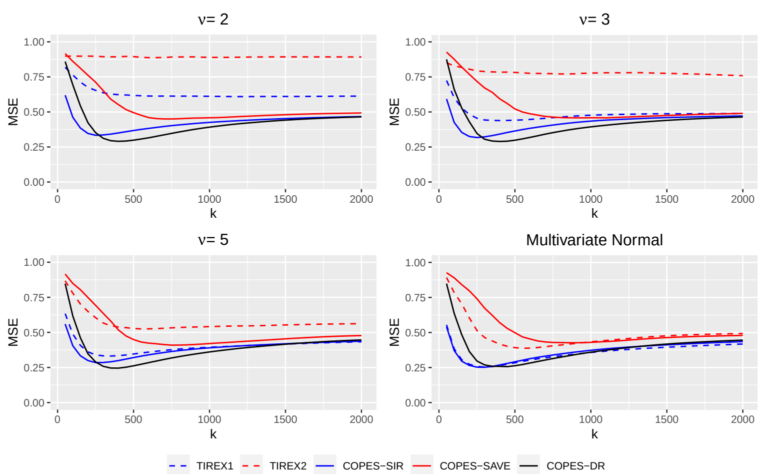

Figure S6: MSEs for different competitors under various ν's under Model C for autoregressive scatter matrix.

are displayed in Figures S8–S10. It can be seen that COPES methods still outperform TIREX1 and TIREX2 even when $\mathbf{X}$ is not from the EC

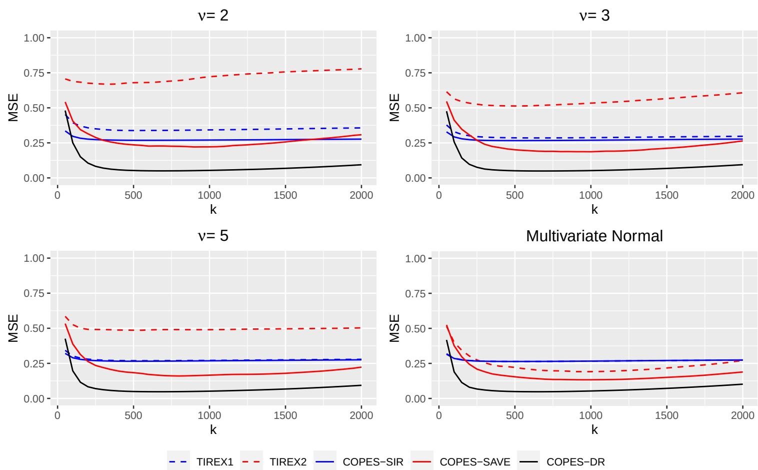

Figure S7: MSEs for different competitors under various ν's under Model D for autoregressive scatter matrix.

distribution, suggesting their insensitivity to the violation of the EC distribution assumption. It is worth noting that COPES-DR performs the best or comparably to the best method throughout all four models.

S4.6 Scatterplot in real data analysis

On Chinese stock dataset, we display the scatterplot of the reduced predictors $\widehat{\boldsymbol{\beta}}_1^\top \mathbf{X}$ and $\widehat{\boldsymbol{\beta}}_2^\top \mathbf{X}$ from COPES-SAVE in Figure S11. The red circles represent the non-tail samples, and the green crosses represent the tail samples.

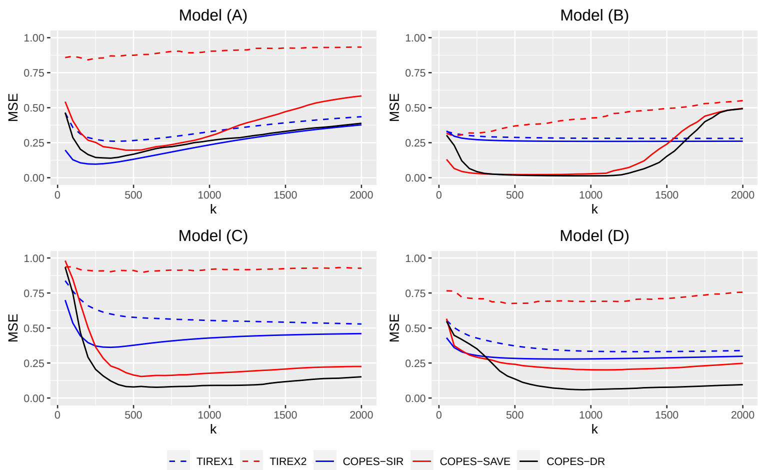

Figure S8: MSE under Models A–D with X_j, $j = 1, \ldots, p$, are independently generated from t_2.

S5 Proofs

S5.1 Proof of Theorem 1

It suffices to show that for any two EDR subspaces, $\mathcal{S}_{\boldsymbol{\alpha}}$ and $\mathcal{S}_{\boldsymbol{\beta}}$, their intersection $\mathcal{S}_{\boldsymbol{\alpha}} \bigcap \mathcal{S}_{\boldsymbol{\beta}} = \mathcal{S}_{\boldsymbol{\delta}}$ is also an extreme SDR subspace. Let $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1, \boldsymbol{\delta})$, $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \boldsymbol{\delta})$, and $\boldsymbol{\eta} = (\boldsymbol{\alpha}_1, \boldsymbol{\beta}_1, \boldsymbol{\delta})$. When $\boldsymbol{\alpha}_1 = \mathbf{0}$ or $\boldsymbol{\beta}_1 = \mathbf{0}$, then $\boldsymbol{\delta} = \boldsymbol{\alpha}$ or $\boldsymbol{\delta} = \boldsymbol{\beta}$, and the result trivially follows. We consider the case where $\boldsymbol{\alpha}_1 \neq \mathbf{0}$ and $\boldsymbol{\beta}_1 \neq \mathbf{0}$. We first introduce the following preliminary lemma.

Lemma S1. *Assume that $\mathcal{S}_{\boldsymbol{\beta}_1}$ is an EDR subspace of Y given $\mathbf{X}$. For any $\mathcal{S}_{\boldsymbol{\beta}_2}$ such that $\mathcal{S}_{\boldsymbol{\beta}_1} \subseteq \mathcal{S}_{\boldsymbol{\beta}_2}$, $\mathcal{S}_{\boldsymbol{\beta}_2}$ is also an EDR subspace of Y given $\mathbf{X}$.*

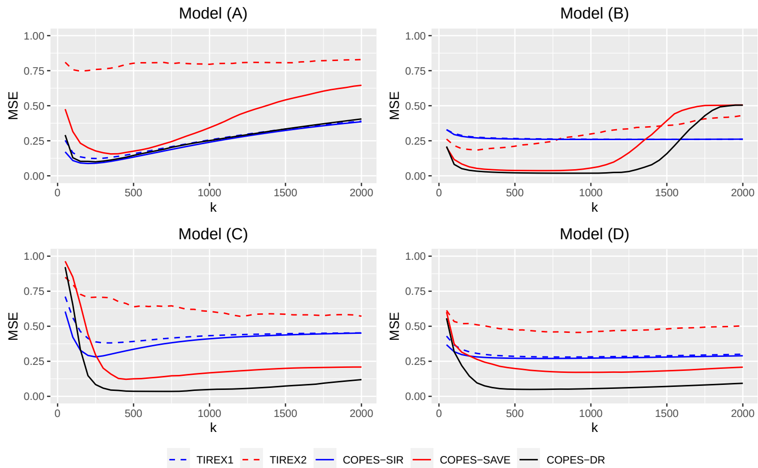

Figure S9: MSE under Models A–D with X_j, $j = 1, \dots, p$, are independently generated from t_3.

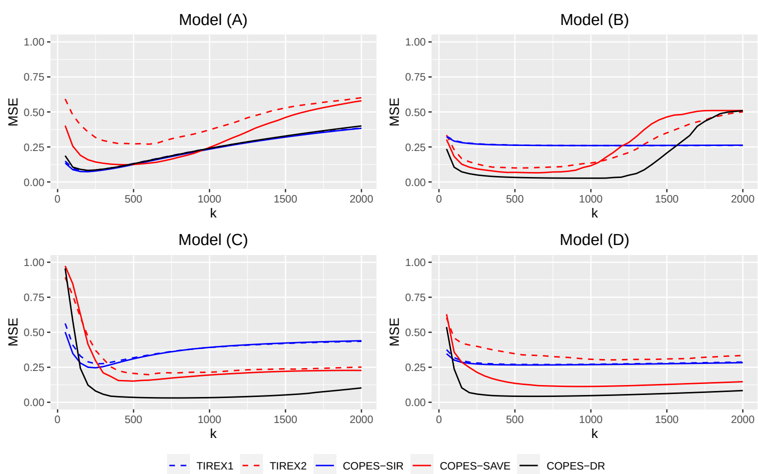

Figure S10: MSE under Models A–D with X_j, $j = 1, \dots, p$, are independently generated from t_5.

Figure S11: Scatter plot of the reduced predictors $\widehat{\boldsymbol{\beta}}_1^\top \mathbf{X}$ and $\widehat{\boldsymbol{\beta}}_2^\top \mathbf{X}$ from COPES-SAVE on Chinese stock dataset. The green crosses correspond to the observations with a response less than y_0, and the red circles correspond to those with a response greater than y_0.

Since $\mathcal{S}_{\boldsymbol{\alpha}}$ and $\mathcal{S}_{\boldsymbol{\beta}}$ are EDR subspaces, then by Lemma S1, $\mathcal{S}_{\boldsymbol{\eta}}$ is also an EDR subspace. Let $\mathbf{W} = \boldsymbol{\eta}^\top \mathbf{X} = (\mathbf{W}_1^\top, \mathbf{W}_2^\top, \mathbf{W}_3^\top)^\top$, where $\mathbf{W}_1 = \boldsymbol{\alpha}_1^\top \mathbf{X}$, $\mathbf{W}_2 = \boldsymbol{\beta}_1^\top \mathbf{X}$, and $\mathbf{W}_3 = \boldsymbol{\delta}^\top \mathbf{X}$. Consider a fixed point $\boldsymbol{x} \in \Omega_{\mathbf{X}}$ and $\boldsymbol{\eta}^\top \boldsymbol{x} = (\boldsymbol{w}_1^\top, \boldsymbol{w}_2^\top, \boldsymbol{w}_3^\top)^\top$. We prove by showing that for any $\varepsilon > 0$, there exists some constant y_0 such that for any $y \geq y_0$, we have

$$\left| \frac{\Pr(Y > y \mid \mathbf{X} = \boldsymbol{x}) - \Pr(Y > y \mid \mathbf{W}_3 = \boldsymbol{w}_3)}{\Pr(Y > y)} \right| \leq \varepsilon, \quad \forall \boldsymbol{x} \in \Omega_{\mathbf{X}}.$$

Let $\Omega_{\mathbf{W}}$ denote the support of $\mathbf{W}$. Also, let $\Omega_{12|3}(\boldsymbol{w}_3)$ denote the support of $(\mathbf{W}_1, \mathbf{W}_2) \mid (\mathbf{W}_3 = \boldsymbol{w}_3)$, which is defined as $\Omega_{12|3}(\boldsymbol{w}_3) = \{(\boldsymbol{w}_1, \boldsymbol{w}_2) : (\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{w}_3) \in \Omega_w\}$. Then, it is also equivalent to show that for any $\varepsilon > 0$, there exists some constant y_0 such that for any $y \geq y_0$, we have

$$\left|\frac{\Pr(Y > y \mid \mathbf{X} = \boldsymbol{x}) - \Pr(Y > y \mid \mathbf{W}_3 = \boldsymbol{w}_3)}{\Pr(Y > y)}\right| \leq \varepsilon, \quad \forall (\boldsymbol{w}_1, \boldsymbol{w}_2) \in \Omega_{12|3}(\boldsymbol{w}_3).$$

We prove by exploiting the fact that any two points in a convex set can be chained by series of linked points. Here, we call that two points $(\boldsymbol{w}_1, \boldsymbol{w}_2)$ and $(\boldsymbol{w}_1^*, \boldsymbol{w}_2^*)$ in $\Omega_{12|3}(\boldsymbol{w}_3)$ are linked if $\boldsymbol{w}_1 = \boldsymbol{w}_1^*$ or $\boldsymbol{w}_2 = \boldsymbol{w}_2^*$. This chaining argument is motivated by the proof of Proposition 6.4 in Cook (1998). However, Cook's proof only involves the conditional distribution. In comparison, we need to deal with the conditional tail probability, used in the definition of EDR subspace, more carefully.

Since $\mathcal{S}_{\boldsymbol{\alpha}}$ is an EDR subspace, for any $\varepsilon > 0$, there exists some constant y_α such that for any $y \geq y_\alpha$, we have

$$\left|\frac{\Pr(Y > y \mid \mathbf{X} = \boldsymbol{x}) - \Pr(Y > y \mid \mathbf{W}_1 = \boldsymbol{w}_1, \mathbf{W}_3 = \boldsymbol{w}_3)}{\Pr(Y > y)}\right| \leq \varepsilon. \tag{S5.1}$$

For any $(\boldsymbol{w}_1^*, \boldsymbol{w}_2^*) \in \Omega_{12|3}(\boldsymbol{w}_3)$, assume $(\boldsymbol{w}_1, \boldsymbol{w}_2^*) \in \Omega_{12|3}(\boldsymbol{w}_3)$, which is linked with both $(\boldsymbol{w}_1^*, \boldsymbol{w}_2^*)$ and $(\boldsymbol{w}_1, \boldsymbol{w}_2)$. Since $(\boldsymbol{w}_1, \boldsymbol{w}_2^*) \in \Omega_{12|3}(\boldsymbol{w}_3)$, then there exists some point $\boldsymbol{x}^* \in \Omega_{\mathbf{X}}$ such that $\boldsymbol{\eta}^\top \boldsymbol{x}^* = (\boldsymbol{w}_1^\top, (\boldsymbol{w}_2^*)^\top, \boldsymbol{w}_3^\top)^\top$, then for $y \geq y_\alpha$,

$$\left|\frac{\Pr(Y > y \mid \mathbf{X} = \boldsymbol{x}^*) - \Pr(Y > y \mid \mathbf{W}_1 = \boldsymbol{w}_1, \mathbf{W}_3 = \boldsymbol{w}_3)}{\Pr(Y > y)}\right| \leq \varepsilon, \tag{S5.2}$$

and there exists some constant y_η such that for any $y \geq y_\eta$,

$$\left|\frac{\Pr(Y > y \mid \mathbf{X} = \boldsymbol{x}^*) - \Pr(Y > y \mid \mathbf{W}_1 = \boldsymbol{w}_1, \mathbf{W}_2 = \boldsymbol{w}_2^*, \mathbf{W}_3 = \boldsymbol{w}_3)}{\Pr(Y > y)}\right| \leq \varepsilon. \tag{S5.3}$$

Combining (S5.2) and (S5.3), for $y \geq \max\{y_\alpha, y_\eta\}$, we have

$$\left|\frac{\Pr(Y > y \mid \mathbf{W}_1 = \boldsymbol{w}_1, \mathbf{W}_3 = \boldsymbol{w}_3) - \Pr(Y > y \mid \mathbf{W}_1 = \boldsymbol{w}_1, \mathbf{W}_2 = \boldsymbol{w}_2^*, \mathbf{W}_3 = \boldsymbol{w}_3)}{\Pr(Y > y)}\right| \leq 2\varepsilon. \tag{S5.4}$$

The subspace $\mathcal{S}_{\boldsymbol{\beta}}$ is also an EDR subspace, there exists some constant y_β such that for any $y \geq y_\beta$,

$$\left|\frac{\Pr(Y > y \mid \mathbf{X} = \boldsymbol{x}^*) - \Pr(Y > y \mid \mathbf{W}_2 = \boldsymbol{w}_2^*, \mathbf{W}_3 = \boldsymbol{w}_3)}{\Pr(Y > y)}\right| \leq \varepsilon. \tag{S5.5}$$

Combining (S5.3) and (S5.5), by taking $y \geq \max\{y_\beta, y_\eta\}$, we have

$$\left|\frac{\Pr(Y > y \mid \mathbf{W}_2 = \boldsymbol{w}_2^*, \mathbf{W}_3 = \boldsymbol{w}_3) - \Pr(Y > y \mid \mathbf{W}_1 = \boldsymbol{w}_1, \mathbf{W}_2 = \boldsymbol{w}_2^*, \mathbf{W}_3 = \boldsymbol{w}_3)}{\Pr(Y > y)}\right| \leq 2\varepsilon. \tag{S5.6}$$

Since $(\boldsymbol{w}_1^*, \boldsymbol{w}_2^*) \in \Omega_{12|3}(\boldsymbol{w}_3)$, then there exists some point $\boldsymbol{x}^{**} \in \Omega_{\mathbf{X}}$ such that $\boldsymbol{\eta}^\top \boldsymbol{x}^{**} = ((\boldsymbol{w}_1^*)^\top, (\boldsymbol{w}_2^*)^\top, \boldsymbol{w}_3^\top)^\top$. Similarly, for $y \geq \max\{y_\beta, y_\eta\}$, we obtain

$$\left|\frac{\Pr(Y > y \mid \mathbf{X} = \boldsymbol{x}^{**}) - \Pr(Y > y \mid \mathbf{W}_1 = \boldsymbol{w}_1^*, \mathbf{W}_2 = \boldsymbol{w}_2^*, \mathbf{W}_3 = \boldsymbol{w}_3)}{\Pr(Y > y)}\right| \leq \varepsilon,$$

and

$$\left|\frac{\Pr(Y > y \mid \mathbf{X} = \boldsymbol{x}^{**}) - \Pr(Y > y \mid \mathbf{W}_2 = \boldsymbol{w}_2^*, \mathbf{W}_3 = \boldsymbol{w}_3)}{\Pr(Y > y)}\right| \leq \varepsilon,$$

which is followed by

$$\left|\frac{\Pr(Y > y \mid \mathbf{W}_2 = \boldsymbol{w}_2^*, \mathbf{W}_3 = \boldsymbol{w}_3) - \Pr(Y > y \mid \mathbf{W}_1 = \boldsymbol{w}_1^*, \mathbf{W}_2 = \boldsymbol{w}_2^*, \mathbf{W}_3 = \boldsymbol{w}_3)}{\Pr(Y > y)}\right| \leq 2\varepsilon. \tag{S5.7}$$

Combining all results in (S5.1)(S5.4)(S5.6)(S5.7), for any $y \geq \max\{y_\alpha, y_\beta, y_\eta\}$, we obtain

$$\left|\frac{\Pr(Y > y \mid \mathbf{X} = \boldsymbol{x}) - \Pr(Y > y \mid \mathbf{W}_1 = \boldsymbol{w}_1^*, \mathbf{W}_2 = \boldsymbol{w}_2^*, \mathbf{W}_3 = \boldsymbol{w}_3)}{\Pr(Y > y)}\right| \leq 7\varepsilon. \tag{S5.8}$$

Since the support $\Omega_{\mathbf{X}}$ is convex, then $\Omega_{12|3}(\boldsymbol{w}_3)$ is also convex. Any two points in $\Omega_{12|3}(\boldsymbol{w}_3)$ are connected by a series of linked points. Therefore, (S5.8) holds for any $(\boldsymbol{w}_1^*, \boldsymbol{w}_2^*) \in \Omega_{12|3}(\boldsymbol{w}_3)$. Then, it follows that

$$\left|\frac{\Pr(Y > y \mid \mathbf{X} = \boldsymbol{x}) - \Pr(Y > y \mid \mathbf{W}_3 = \boldsymbol{w}_3)}{\Pr(Y > y)}\right| \leq 7\varepsilon, \quad \forall(\boldsymbol{w}_1, \boldsymbol{w}_2) \in \Omega_{12|3}(\boldsymbol{w}_3),$$

which completes the proof. □

S5.2 Proof of Lemma 1

For EC distributed $\mathbf{X}$, let $R = \|\mathbf{X} - \boldsymbol{\mu}\|_{\boldsymbol{\Sigma}}$. It is known that $\overrightarrow{\mathbf{X}}$ and R are independent (Johnson 1987). Thus,

$$\begin{aligned}\Pr(Y > y|\overrightarrow{\mathbf{X}}) =& \mathbb{E}\left\{\Pr(Y > y|\overrightarrow{\mathbf{X}}, R)|\overrightarrow{\mathbf{X}}\right\} = \int \Pr(Y > y|\overrightarrow{\mathbf{X}}, r)dF_{R|\overrightarrow{\mathbf{X}}}(r) \\ &= \int \Pr(Y > y|\overrightarrow{\mathbf{X}}, r)dF_R(r) =: \mathbb{E}_R\left\{\Pr(Y > y|\overrightarrow{\mathbf{X}}, R)\right\},\end{aligned}$$

where $\mathbb{E}_R$ denotes the expectation over R. Similarly, we have that,

$$\Pr(Y > y|\boldsymbol{\beta}^\top \overrightarrow{\mathbf{X}}) = \mathbb{E}_R \left\{ \Pr(Y > y|\boldsymbol{\beta}^\top \overrightarrow{\mathbf{X}}, R) \right\}.$$

As a result,

$$\begin{aligned}
&\left| \Pr(Y > y|\overrightarrow{\mathbf{X}}) - \Pr(Y > y|\boldsymbol{\beta}^\top \overrightarrow{\mathbf{X}}) \right| \\
=& \left| \mathbb{E}_R \left\{ \Pr(Y > y|\overrightarrow{\mathbf{X}}, R) - \Pr(Y > y|\boldsymbol{\beta}^\top \overrightarrow{\mathbf{X}}, R) \right\} \right| \\
=& \left| \mathbb{E}_R \left\{ \Pr(Y > y|\mathbf{X}) - \Pr(Y > y|\boldsymbol{\beta}^\top \mathbf{X}, R) \right\} \right| \\
=& \left| \mathbb{E}_R \left\{ \Pr(Y > y|\mathbf{X}) - \Pr(Y > y|\boldsymbol{\beta}^\top \mathbf{X}) + \Pr(Y > y|\boldsymbol{\beta}^\top \mathbf{X}) - \Pr(Y > y|\boldsymbol{\beta}^\top \mathbf{X}, R) \right\} \right| \\
\leq& \mathbb{E}_R \left| \Pr(Y > y|\mathbf{X}) - \Pr(Y > y|\boldsymbol{\beta}^\top \mathbf{X}) \right| + \mathbb{E}_R \left| \Pr(Y > y|\boldsymbol{\beta}^\top \mathbf{X}) - \Pr(Y > y|\boldsymbol{\beta}^\top \mathbf{X}, R) \right| \\
=&: I_1 + I_2.
\end{aligned}$$

Since $\mathcal{S}_{\boldsymbol{\beta}}$ is an EDR subspace, then for any $\varepsilon > 0$, there exists some constant y_0 such that for all $y \geq y_0$,

$$\left| \frac{\Pr(Y > y|\mathbf{X} = \boldsymbol{x}) - \Pr(Y > y|\boldsymbol{\beta}^\top \mathbf{X} = \boldsymbol{\beta}^\top \boldsymbol{x})}{\Pr(Y > y)} \right| \leq \varepsilon, \quad \text{for all } \boldsymbol{x} \in \Omega_{\mathbf{X}}, \tag{S5.9}$$

which leads to

$$I_1 \leq \varepsilon \Pr(Y > y).$$

For I_2, since

$$\Pr(Y > y|\boldsymbol{\beta}^\top \mathbf{X}) = \mathbb{E} \left\{ \Pr(Y > y|\boldsymbol{\beta}^\top \mathbf{X}) | \boldsymbol{\beta}^\top \mathbf{X}, R \right\},$$

and

$$\begin{aligned}\Pr(Y > y|\boldsymbol{\beta}^\top \mathbf{X}, R) =& \mathbb{E}\left\{I(Y > y)|\boldsymbol{\beta}^\top \mathbf{X}, R\right\}\\ =& \mathbb{E}\left[\mathbb{E}\left\{I(Y > y)|\mathbf{X}\right\}|\boldsymbol{\beta}^\top \mathbf{X}, R\right]\\ =& \mathbb{E}\left\{\Pr\left(Y > y|\mathbf{X}\right)|\boldsymbol{\beta}^\top \mathbf{X}, R\right\},\end{aligned}$$

then we have

$$\begin{aligned}I_2 =& \mathbb{E}_R\left|\Pr(Y > y|\boldsymbol{\beta}^\top \mathbf{X}) - \Pr(Y > y|\boldsymbol{\beta}^\top \mathbf{X}, R)\right|\\ =& \mathbb{E}_R\left|\mathbb{E}\left\{\Pr(Y > y|\boldsymbol{\beta}^\top \mathbf{X}) - \Pr(Y > y|\mathbf{X})|\boldsymbol{\beta}^\top \mathbf{X}, R\right\}\right|.\end{aligned}$$

By (S5.9), we have that, for $y \geq y_0$,

$$I_2 \leq \varepsilon \Pr(Y > y).$$

For any $\boldsymbol{x} \in \Omega_{\mathbf{X}}$, let $\overrightarrow{\boldsymbol{x}}$ denote the corresponding contour-projected vector and let $\Omega_{\vec{x}}$ denote the support of $\overrightarrow{\boldsymbol{x}}$. By combining the results for I_1 and I_2, we conclude that, for any $\varepsilon > 0$, there exists some constant y_0 such that for all $y \geq y_0$,

$$\left|\frac{\Pr(Y > y|\overrightarrow{\mathbf{X}} = \overrightarrow{\boldsymbol{x}}) - \Pr(Y > y|\boldsymbol{\beta}^\top \overrightarrow{\mathbf{X}} = \boldsymbol{\beta}^\top \overrightarrow{\boldsymbol{x}})}{\Pr(Y > y)}\right| \leq 2\varepsilon, \quad \text{for all } \overrightarrow{\boldsymbol{x}} \in \Omega_{\vec{x}},$$

and thus the proof is completed. □

S5.3 Proof of Theorem 2

We first introduce the following auxiliary lemma.

Lemma S2. *If $\mathcal{S}_{\boldsymbol{\beta}}$ is an EDR subspace of Y given $\mathbf{X}$, then for any real valued functions g and h, measurable and bounded, we have*

$$\frac{\mathbb{E}\left[h(\boldsymbol{\beta}^{\top}\mathbf{X})I(Y>y)\left\{\mathbb{E}\left(g(\mathbf{X})|Y,\boldsymbol{\beta}^{\top}\mathbf{X}\right)-\mathbb{E}\left(g(\mathbf{X})|\boldsymbol{\beta}^{\top}\mathbf{X}\right)\right\}\right]}{P(Y>y)} \longrightarrow 0, \quad y \to y^{+}. \tag{S5.10}$$

Proof of Lemma S2. Since $\mathrm{span}(\boldsymbol{\beta})$ is an EDR subspace, we have that for any $\varepsilon > 0$, there exists some constant y_0 such that for all $y \geq y_0$,

$$\left|\frac{\Pr(Y>y \mid \mathbf{X}) - \Pr(Y>y \mid \boldsymbol{\beta}^{\top}\mathbf{X})}{\Pr(Y>y)}\right| \leq \varepsilon, \quad \boldsymbol{x} \in \Omega_{\mathbf{X}},$$

By dominance convergency theorem, we have

$$\left|\frac{\Pr(Y>y \mid \mathbf{X}) - \Pr(Y>y \mid \boldsymbol{\beta}^{\top}\mathbf{X})}{\Pr(Y>y)}\right| \longrightarrow 0, \quad y \to y^{+}, \quad \text{in } L_1.$$

Then, by Propositions 3 and 4 in Aghbalou et al. (2024), (S5.10) holds. □

The proof is similar to that of Theorem 1 in Aghbalou et al. (2024). It suffices to show that

$$\mathbf{Q}_{\boldsymbol{\beta}}\boldsymbol{\Sigma}^{-1}\mathbb{E}(\overrightarrow{\mathbf{X}} \mid Y>y) \longrightarrow 0, \quad y \to y^{+}.$$

Let $\pi_y = \Pr(Y>y)$, then

$$\begin{aligned}\mathbf{Q}_{\boldsymbol{\beta}}\boldsymbol{\Sigma}^{-1}\mathbb{E}(\overrightarrow{\mathbf{X}} \mid Y>y) =& \pi_y^{-1}\mathbf{Q}_{\boldsymbol{\beta}}\boldsymbol{\Sigma}^{-1}\mathbb{E}\{\overrightarrow{\mathbf{X}}I(Y>y)\} \\ =& \pi_y^{-1}\mathbf{Q}_{\boldsymbol{\beta}}\boldsymbol{\Sigma}^{-1}\mathbb{E}\{\mathbb{E}(\overrightarrow{\mathbf{X}} \mid \boldsymbol{\beta}^{\top}\overrightarrow{\mathbf{X}}, Y)I(Y>y)\}.\end{aligned}$$

Since $\mathbf{X}$ is EC distributed, so is $\overrightarrow{\mathbf{X}}$. Then $\overrightarrow{\mathbf{X}}$ satisfies the linearity condition and we have that

$$\mathbb{E}(\overrightarrow{\mathbf{X}} \mid \boldsymbol{\beta}^\top \overrightarrow{\mathbf{X}}) = \mathbf{P}^\top_{\boldsymbol{\beta}(p^{-1}\boldsymbol{\Sigma})} \overrightarrow{\mathbf{X}},$$

which follows from Lemma 1.1 in Li (2018) and the fact that $\mathrm{Cov}(\overrightarrow{\mathbf{X}}) = p^{-1}\boldsymbol{\Sigma}$. Therefore,

$$\mathbf{Q}_{\boldsymbol{\beta}} \boldsymbol{\Sigma}^{-1} \mathbb{E}(\overrightarrow{\mathbf{X}} \mid \boldsymbol{\beta}^\top \overrightarrow{\mathbf{X}}) = \mathbf{Q}_{\boldsymbol{\beta}} \boldsymbol{\Sigma}^{-1} \mathbf{P}^\top_{\boldsymbol{\beta}(p^{-1}\boldsymbol{\Sigma})} \overrightarrow{\mathbf{X}} = \mathbf{0}.$$

Then,

$$\mathbf{Q}_{\boldsymbol{\beta}} \boldsymbol{\Sigma}^{-1} \mathbb{E}(\overrightarrow{\mathbf{X}} \mid Y > y) = \pi_y^{-1} \mathbf{Q}_{\boldsymbol{\beta}} \boldsymbol{\Sigma}^{-1} \mathbb{E}\left[\left\{\mathbb{E}(\overrightarrow{\mathbf{X}} \mid \boldsymbol{\beta}^\top \overrightarrow{\mathbf{X}}, Y) - \mathbb{E}(\overrightarrow{\mathbf{X}} \mid \boldsymbol{\beta}^\top \overrightarrow{\mathbf{X}})\right\} I(Y > y)\right].$$

According to Lemma 1, $\mathcal{S}_{\boldsymbol{\beta}}$ is also an EDR subspace of Y given $\overrightarrow{\mathbf{X}}$. We take $h = 1$ and $g(\overrightarrow{\mathbf{X}}) = \overrightarrow{X}_i$, $i = 1, \ldots, p$, in Lemma S2. Since $\overrightarrow{X}_i$ is bounded, then Lemma S2 is satisfied and we obtain that for $i = 1, \ldots, p$,

$$\pi_y^{-1} \mathbb{E}\left[\left\{\mathbb{E}(\overrightarrow{X}_i \mid \boldsymbol{\beta}^\top \overrightarrow{\mathbf{X}}, Y) - \mathbb{E}(\overrightarrow{X}_i \mid \boldsymbol{\beta}^\top \overrightarrow{\mathbf{X}})\right\} I(Y > y)\right] \longrightarrow 0, \quad y \to y^+.$$

As a consequence,

$$\mathbf{Q}_{\boldsymbol{\beta}} \boldsymbol{\Sigma}^{-1} \mathbb{E}(\overrightarrow{\mathbf{X}} \mid Y > y) \longrightarrow 0, \quad y \to y^+,$$

which completes the proof. □

S5.4 Proof of Lemma 2

Since

$$\mathbf{C}_h(u) = u\mathbb{E}\left\{\overrightarrow{\mathbf{X}} \mid \widetilde{Y} < F^{-1}(uh)\right\},$$

then under Assumption (A1), for any $u \in [0,1]$, we have

$$\mathbf{C}_h(u) \longrightarrow u\boldsymbol{\nu}, \quad h \to 0.$$

Since $\|\overrightarrow{\mathbf{X}}\|$ is bounded from above, by dominated convergence theorem, we have

$$\int_0^1 \mathbf{C}_h(u)\mathbf{C}_h^\top(u)\mathrm{d}u \longrightarrow \int_0^1 u^2\boldsymbol{\nu}\boldsymbol{\nu}^\top \mathrm{d}u, \quad h \to 0.$$

Therefore,

$$\mathbf{M}_{\mathrm{eSIR}}^h \longrightarrow \frac{1}{3}\boldsymbol{\nu}\boldsymbol{\nu}^\top := \mathbf{M}_{\mathrm{eSIR}}, \quad h \to 0$$

Then,

$$\boldsymbol{\Sigma}^{-1}\mathrm{span}(\mathbf{M}_{\mathrm{eSIR}}) = \boldsymbol{\Sigma}^{-1}\mathrm{span}(\boldsymbol{\nu}) = \mathcal{S}_{\mathrm{eSIR}},$$

which completes the proof. □

S5.5 Proof of Lemma 3

We define the standardized predictor $\overrightarrow{\mathbf{Z}} = \boldsymbol{\Sigma}^{-1/2}\overrightarrow{\mathbf{X}}$, which is uniformly distributed on the unit sphere $\mathbb{S}^{p-1} = \{\mathbf{u} \in \mathbb{R}^p \mid \|\mathbf{u}\| = 1\}$. Let $\boldsymbol{\alpha} \in \mathbb{R}^{p\times d}$

be the orthonormal basis matrix of subspace $\mathbf{\Sigma}^{1/2}\text{span}(\boldsymbol{\eta})$, then we have

$$\begin{aligned}\text{Cov}(\overrightarrow{\mathbf{X}} \mid \boldsymbol{\eta}^\top \overrightarrow{\mathbf{X}}) &= \mathbf{\Sigma}^{1/2}\text{Cov}(\overrightarrow{\mathbf{Z}} \mid \boldsymbol{\eta}^\top \mathbf{\Sigma}^{1/2}\overrightarrow{\mathbf{Z}})\mathbf{\Sigma}^{1/2} \\ &= \mathbf{\Sigma}^{1/2}\text{Cov}(\overrightarrow{\mathbf{Z}} \mid \boldsymbol{\alpha}^\top \overrightarrow{\mathbf{Z}})\mathbf{\Sigma}^{1/2}. \qquad \text{(S5.11)}\end{aligned}$$

Let $\boldsymbol{\alpha}_0 \in \mathbb{R}^{p\times(p-d)}$ be the orthonormal basis of the orthogonal complement of span($\boldsymbol{\alpha}$) such that $(\boldsymbol{\alpha}, \boldsymbol{\alpha}_0)$ is orthogonal. Following the arguments in the proof of Lemma 3 in Luo et al. (2009), we have that

$$\text{Cov}(\overrightarrow{\mathbf{Z}} \mid \boldsymbol{\alpha}^\top \overrightarrow{\mathbf{Z}}) = \frac{1 - \|\boldsymbol{\alpha}^\top \overrightarrow{\mathbf{Z}}\|^2}{p-d}\mathbf{P}_{\boldsymbol{\alpha}_0}.$$

Since

$$\|\boldsymbol{\alpha}^\top \overrightarrow{\mathbf{Z}}\|^2 = \overrightarrow{\mathbf{Z}}^\top \mathbf{P}_{\boldsymbol{\alpha}} \overrightarrow{\mathbf{Z}} = \overrightarrow{\mathbf{X}}^\top \mathbf{\Sigma}^{-1/2}\mathbf{P}_{\boldsymbol{\alpha}}\mathbf{\Sigma}^{-1/2}\overrightarrow{\mathbf{X}},$$

and

$$\mathbf{P}_{\boldsymbol{\alpha}} = \mathbf{P}_{\mathbf{\Sigma}^{1/2}\boldsymbol{\eta}} = \mathbf{\Sigma}^{1/2}\boldsymbol{\eta}(\boldsymbol{\eta}^\top \mathbf{\Sigma}\boldsymbol{\eta})^{-1}\boldsymbol{\eta}^\top \mathbf{\Sigma}^{1/2}.$$

Then,

$$\|\boldsymbol{\alpha}^\top \overrightarrow{\mathbf{Z}}\|^2 = \overrightarrow{\mathbf{X}}^\top \boldsymbol{\eta}(\boldsymbol{\eta}^\top \mathbf{\Sigma}\boldsymbol{\eta})^{-1}\boldsymbol{\eta}^\top \overrightarrow{\mathbf{X}} = \|\mathbf{P}^\top_{\boldsymbol{\eta}(\mathbf{\Sigma})}\overrightarrow{\mathbf{X}}\|^2_{\mathbf{\Sigma}}.$$

By noting that $\|\overrightarrow{\mathbf{X}}\|^2_{\mathbf{\Sigma}} = 1$, we obtain

$$1 - \|\boldsymbol{\alpha}^\top \overrightarrow{\mathbf{Z}}\|^2 = \|\overrightarrow{\mathbf{X}}\|^2_{\mathbf{\Sigma}} - \|\mathbf{P}^\top_{\boldsymbol{\eta}(\mathbf{\Sigma})}\overrightarrow{\mathbf{X}}\|^2_{\mathbf{\Sigma}} = \|\mathbf{Q}^\top_{\boldsymbol{\eta}(\mathbf{\Sigma})}\overrightarrow{\mathbf{X}}\|^2_{\mathbf{\Sigma}}. \qquad \text{(S5.12)}$$

From (S5.12), it can be seen that $1-\|\boldsymbol{\alpha}^\top \overrightarrow{\mathbf{Z}}\|^2$ is a function of $\boldsymbol{\eta}^\top \overrightarrow{\mathbf{X}}$. There, we denote

$$\zeta(\boldsymbol{\eta}^\top \overrightarrow{\mathbf{X}}) = \frac{1-\|\boldsymbol{\alpha}^\top \overrightarrow{\mathbf{Z}}\|^2}{p-d} = \frac{\|\mathbf{Q}_{\boldsymbol{\eta}(\boldsymbol{\Sigma})}^\top \overrightarrow{\mathbf{X}}\|_{\boldsymbol{\Sigma}}^2}{p-d}$$

and it follows that

$$\mathrm{Cov}(\overrightarrow{\mathbf{Z}} \mid \boldsymbol{\alpha}^\top \overrightarrow{\mathbf{Z}}) = \zeta(\boldsymbol{\eta}^\top \overrightarrow{\mathbf{X}})\mathbf{P}_{\boldsymbol{\alpha}_0}. \tag{S5.13}$$

Combining (S5.11) and (S5.13), we obtain

$$\mathrm{Cov}(\overrightarrow{\mathbf{X}} \mid \boldsymbol{\eta}^\top \overrightarrow{\mathbf{X}}) = \zeta(\boldsymbol{\eta}^\top \overrightarrow{\mathbf{X}})\boldsymbol{\Sigma}^{1/2}\mathbf{P}_{\boldsymbol{\alpha}_0}\boldsymbol{\Sigma}^{1/2} = \zeta(\boldsymbol{\eta}^\top \overrightarrow{\mathbf{X}})\boldsymbol{\Sigma}\mathbf{Q}_{\boldsymbol{\eta}(\boldsymbol{\Sigma})}.$$

□

S5.6 Proof of Theorem 3

We show that

$$\mathbf{Q}_{\boldsymbol{\beta}}\boldsymbol{\Sigma}^{-1}\mathbb{E}\left[\overrightarrow{\mathbf{X}}\overrightarrow{\mathbf{X}}^\top - \zeta(\boldsymbol{\beta}^\top \overrightarrow{\mathbf{X}})\boldsymbol{\Sigma} \mid Y > y\right] \longrightarrow 0, \quad y \to y^+,$$

which is equivalent to

$$\pi_y^{-1}\mathbf{Q}_{\boldsymbol{\beta}}\boldsymbol{\Sigma}^{-1}\mathbb{E}\left(\left[\overrightarrow{\mathbf{X}}\overrightarrow{\mathbf{X}}^\top - \zeta(\boldsymbol{\beta}^\top \overrightarrow{\mathbf{X}})\boldsymbol{\Sigma}\right] I(Y > y)\right) \longrightarrow 0, \quad y \to y^+,$$

where $\pi_y = \Pr(Y > y)$.

We begin with proving

$$\mathbf{Q}_{\boldsymbol{\beta}}\boldsymbol{\Sigma}^{-1}\mathbb{E}\left[\overrightarrow{\mathbf{X}}\overrightarrow{\mathbf{X}}^\top - \zeta(\boldsymbol{\beta}^\top \overrightarrow{\mathbf{X}})\boldsymbol{\Sigma} \mid \boldsymbol{\beta}^\top \overrightarrow{\mathbf{X}}\right] = \mathbf{0}. \tag{S5.14}$$

According to Lemma 3, we have

$$\mathrm{Cov}(\overrightarrow{\mathbf{X}} \mid \boldsymbol{\beta}^\top \overrightarrow{\mathbf{X}}) = \zeta(\boldsymbol{\beta}^\top \overrightarrow{\mathbf{X}}) \boldsymbol{\Sigma} \mathbf{Q}_{\boldsymbol{\beta}(\boldsymbol{\Sigma})}.$$

Then,

$$\begin{aligned}
&\mathbf{Q}_{\boldsymbol{\beta}} \boldsymbol{\Sigma}^{-1} \mathbb{E}\left[\overrightarrow{\mathbf{X}}\overrightarrow{\mathbf{X}}^\top - \zeta(\boldsymbol{\beta}^\top \overrightarrow{\mathbf{X}}) \boldsymbol{\Sigma} \mid \boldsymbol{\beta}^\top \overrightarrow{\mathbf{X}}\right] \\
=&\mathbf{Q}_{\boldsymbol{\beta}} \boldsymbol{\Sigma}^{-1} \left[\mathrm{Cov}(\overrightarrow{\mathbf{X}} \mid \boldsymbol{\beta}^\top \overrightarrow{\mathbf{X}}) + \mathbb{E}(\overrightarrow{\mathbf{X}} \mid \boldsymbol{\beta}^\top \overrightarrow{\mathbf{X}}) \mathbb{E}(\overrightarrow{\mathbf{X}}^\top \mid \boldsymbol{\beta}^\top \overrightarrow{\mathbf{X}}) - \zeta(\boldsymbol{\beta}^\top \overrightarrow{\mathbf{X}}) \boldsymbol{\Sigma}\right] \\
=&\mathbf{Q}_{\boldsymbol{\beta}} \boldsymbol{\Sigma}^{-1} \left\{-\zeta(\boldsymbol{\beta}^\top \overrightarrow{\mathbf{X}}) \boldsymbol{\Sigma} \mathbf{P}_{\boldsymbol{\beta}(\boldsymbol{\Sigma})} + (\mathbf{P}^\top_{\boldsymbol{\beta}(p^{-1}\boldsymbol{\Sigma})} \overrightarrow{\mathbf{X}})(\mathbf{P}^\top_{\boldsymbol{\beta}(p^{-1}\boldsymbol{\Sigma})} \overrightarrow{\mathbf{X}})^\top\right\}, \\
=&-\zeta(\boldsymbol{\beta}^\top \overrightarrow{\mathbf{X}}) \mathbf{Q}_{\boldsymbol{\beta}} \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma} \mathbf{P}_{\boldsymbol{\beta}(\boldsymbol{\Sigma})} + \mathbf{Q}_{\boldsymbol{\beta}} \boldsymbol{\Sigma}^{-1} \mathbf{P}^\top_{\boldsymbol{\beta}(p^{-1}\boldsymbol{\Sigma})} \overrightarrow{\mathbf{X}}\overrightarrow{\mathbf{X}}^\top \mathbf{P}_{\boldsymbol{\beta}(p^{-1}\boldsymbol{\Sigma})} \\
=&-\zeta(\boldsymbol{\beta}^\top \overrightarrow{\mathbf{X}}) \mathbf{Q}_{\boldsymbol{\beta}} \mathbf{P}_{\boldsymbol{\beta}(\boldsymbol{\Sigma})} + \mathbf{Q}_{\boldsymbol{\beta}} \boldsymbol{\Sigma}^{-1} \mathbf{P}^\top_{\boldsymbol{\beta}(\boldsymbol{\Sigma})} \overrightarrow{\mathbf{X}}\overrightarrow{\mathbf{X}}^\top \mathbf{P}_{\boldsymbol{\beta}(\boldsymbol{\Sigma})} \\
=&\mathbf{0},
\end{aligned}$$

where in the second equation, we used the fact that $\mathbb{E}(\overrightarrow{\mathbf{X}} \mid \boldsymbol{\beta}^\top \overrightarrow{\mathbf{X}}) = \mathbf{P}^\top_{\boldsymbol{\beta}(p^{-1}\boldsymbol{\Sigma})} \overrightarrow{\mathbf{X}}$, and in the last equation, we use the fact that

$$\begin{aligned}
\mathbf{Q}_{\boldsymbol{\beta}} \mathbf{P}_{\boldsymbol{\beta}(\boldsymbol{\Sigma})} =&(\mathbf{I} - \boldsymbol{\beta}(\boldsymbol{\beta}^\top \boldsymbol{\beta})^{-1} \boldsymbol{\beta}^\top) \boldsymbol{\beta}(\boldsymbol{\beta}^\top \boldsymbol{\Sigma} \boldsymbol{\beta})^{-1} \boldsymbol{\beta}^\top \boldsymbol{\Sigma} \\
=&\boldsymbol{\beta}(\boldsymbol{\beta}^\top \boldsymbol{\Sigma} \boldsymbol{\beta})^{-1} \boldsymbol{\beta}^\top \boldsymbol{\Sigma} - \boldsymbol{\beta}(\boldsymbol{\beta}^\top \boldsymbol{\Sigma} \boldsymbol{\beta})^{-1} \boldsymbol{\beta}^\top \boldsymbol{\Sigma} \\
=&\mathbf{0},
\end{aligned}$$

and

$$\begin{aligned}
\mathbf{Q}_{\boldsymbol{\beta}} \boldsymbol{\Sigma}^{-1} \mathbf{P}^\top_{\boldsymbol{\beta}(\boldsymbol{\Sigma})} =&(\mathbf{I} - \boldsymbol{\beta}(\boldsymbol{\beta}^\top \boldsymbol{\beta})^{-1} \boldsymbol{\beta}^\top) \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma} \boldsymbol{\beta}(\boldsymbol{\beta}^\top \boldsymbol{\Sigma} \boldsymbol{\beta})^{-1} \boldsymbol{\beta}^\top \\
=&\boldsymbol{\beta}(\boldsymbol{\beta}^\top \boldsymbol{\Sigma} \boldsymbol{\beta})^{-1} \boldsymbol{\beta}^\top - \boldsymbol{\beta}(\boldsymbol{\beta}^\top \boldsymbol{\Sigma} \boldsymbol{\beta})^{-1} \boldsymbol{\beta}^\top \\
=&\mathbf{0}.
\end{aligned}$$

Then, we have

$$
\begin{aligned}
&\pi_y^{-1}\mathbf{Q}_{\boldsymbol\beta}\boldsymbol\Sigma^{-1}\mathbb{E}\left\{\left(\overrightarrow{\mathbf{X}}\overrightarrow{\mathbf{X}}^\top-\zeta(\boldsymbol\beta^\top\overrightarrow{\mathbf{X}})\boldsymbol\Sigma\right)I(Y>y)\right\}\\
=&\pi_y^{-1}\mathbf{Q}_{\boldsymbol\beta}\boldsymbol\Sigma^{-1}\mathbb{E}\left\{\mathbb{E}\left(\overrightarrow{\mathbf{X}}\overrightarrow{\mathbf{X}}^\top-\zeta(\boldsymbol\beta^\top\overrightarrow{\mathbf{X}})\boldsymbol\Sigma\mid\boldsymbol\beta^\top\overrightarrow{\mathbf{X}},Y\right)I(Y>y)\right\}\\
=&\pi_y^{-1}\mathbf{Q}_{\boldsymbol\beta}\boldsymbol\Sigma^{-1}\mathbb{E}\left[\left\{\mathbb{E}\left(\overrightarrow{\mathbf{X}}\overrightarrow{\mathbf{X}}^\top-\zeta(\boldsymbol\beta^\top\overrightarrow{\mathbf{X}})\boldsymbol\Sigma\mid\boldsymbol\beta^\top\overrightarrow{\mathbf{X}},Y\right)-\right.\right.\\
&\left.\left.\mathbb{E}\left(\overrightarrow{\mathbf{X}}\overrightarrow{\mathbf{X}}^\top-\zeta(\boldsymbol\beta^\top\overrightarrow{\mathbf{X}})\boldsymbol\Sigma\mid\boldsymbol\beta^\top\overrightarrow{\mathbf{X}}\right)\right\}I(Y>y)\right]\\
=&\pi_y^{-1}\mathbf{Q}_{\boldsymbol\beta}\boldsymbol\Sigma^{-1}\mathbb{E}\left[\left\{\mathbb{E}\left(\overrightarrow{\mathbf{X}}\overrightarrow{\mathbf{X}}^\top\mid\boldsymbol\beta^\top\overrightarrow{\mathbf{X}},Y\right)-\mathbb{E}\left(\overrightarrow{\mathbf{X}}\overrightarrow{\mathbf{X}}^\top\mid\boldsymbol\beta^\top\overrightarrow{\mathbf{X}}\right)\right\}I(Y>y)\right].
\end{aligned}
$$

where the second equality follows from (S5.14). Since $\mathcal{S}_{\boldsymbol\beta}$ is an EDR subspace of Y given $\overrightarrow{\mathbf{X}}$, according to Lemma S2, by taking $h=1$ and $g(\overrightarrow{\mathbf{X}})=\overrightarrow{X}_i\overrightarrow{X}_j$, for $i,j\in\{1,\dots,p\}$, since $g(\overrightarrow{\mathbf{X}})$ is bounded, we have

$$
\pi_y^{-1}\mathbb{E}\left[\left\{\mathbb{E}\left(\overrightarrow{X}_i\overrightarrow{X}_j\mid\boldsymbol\beta^\top\overrightarrow{\mathbf{X}},Y\right)-\mathbb{E}\left(\overrightarrow{X}_i\overrightarrow{X}_j\mid\boldsymbol\beta^\top\overrightarrow{\mathbf{X}}\right)\right\}I(Y>y)\right]\longrightarrow 0,\quad y\to y^+. \tag{S5.15}
$$

Hence, we have

$$
\pi_y^{-1}\mathbf{Q}_{\boldsymbol\beta}\boldsymbol\Sigma^{-1}\mathbb{E}\left\{\left(\overrightarrow{\mathbf{X}}\overrightarrow{\mathbf{X}}^\top-\zeta(\boldsymbol\beta^\top\overrightarrow{\mathbf{X}})\boldsymbol\Sigma\right)I(Y>y)\right\}\longrightarrow 0,\quad y\to y^+,
$$

which finishes the proof. □

S5.7 Proof of Lemma 4

We equivalently rewrite $\mathcal{S}_{\text{eSAVE}}$ as

$$
\mathcal{S}_{\text{eSAVE}}=\boldsymbol\Sigma^{-1/2}\text{span}(\boldsymbol\Sigma^{-1/2}\mathbf{T}\boldsymbol\Sigma^{-1/2}-\tau_{\boldsymbol\beta}\mathbf{I}).
$$

Since we assume that $\dim(\mathcal{S}_{\text{eSAVE}}) < p/2$, then

$$\#\{i \mid \sigma_i(\boldsymbol{\Sigma}^{-1/2}\mathbf{T}\boldsymbol{\Sigma}^{-1/2}) = \tau_{\boldsymbol{\beta}}\} > p/2,$$

which completes the proof. □

S5.8 Proof of Lemma 5

Since

$$\mathbf{T}_h(u) = u\mathbb{E}\left\{\overrightarrow{\mathbf{X}}\overrightarrow{\mathbf{X}}^\top \mid \widetilde{Y} < F^{-1}(uh)\right\},$$

under Assumption (A2), for any $u \in [0,1]$, we have

$$\mathbf{T}_h(u) \longrightarrow u\mathbf{T}, \quad h \to 0,$$

and

$$\mathbf{D}_h(u) \longrightarrow u\mathbf{T} - u \cdot \text{median}\{\boldsymbol{\sigma}(\boldsymbol{\Sigma}^{-1/2}\mathbf{T}\boldsymbol{\Sigma}^{-1/2})\}\boldsymbol{\Sigma} = u(\mathbf{T} - \tau\boldsymbol{\Sigma}).$$

Since $\|\overrightarrow{\mathbf{X}}\|$ is bounded from above, by dominated convergence theorem, we have

$$\int_0^1 \mathbf{D}_h(u)\mathbf{D}_h^\top(u)\mathrm{d}u \longrightarrow \int_0^1 u^2 \{\mathbf{T} - \tau\boldsymbol{\Sigma}\}\{\mathbf{T} - \tau\boldsymbol{\Sigma}\}^\top \mathrm{d}u, \quad h \to 0.$$

Therefore,

$$\mathbf{M}^h_{\text{eSAVE}} \longrightarrow \frac{1}{3}\{\mathbf{T} - \tau\boldsymbol{\Sigma}\}\{\mathbf{T} - \tau\boldsymbol{\Sigma}\}^\top := \mathbf{M}_{\text{eSAVE}}, \quad h \to 0$$

Since by Lemma 4, by assuming that $\dim(\mathcal{S}_{\text{eSAVE}}) < p/2$, then $\mathcal{S}_{\text{eSAVE}} = \mathbf{\Sigma}^{-1}\text{span}(\mathbf{T} - \tau\mathbf{\Sigma})$, then

$$\mathbf{\Sigma}^{-1}\text{span}(\mathbf{M}_{\text{eSAVE}}) = \mathbf{\Sigma}^{-1}\text{span}(\mathbf{T} - \tau\mathbf{\Sigma}) = \mathcal{S}_{\text{eSAVE}},$$

which completes the proof. □

S5.9 Proof of Lemma 6

We re-express $\mathbb{E}\{(\overrightarrow{\mathbf{X}} - \overrightarrow{\mathbf{X}}^*)(\overrightarrow{\mathbf{X}} - \overrightarrow{\mathbf{X}}^*)^\top \mid Y > y, Y^* > y^*\}$ as

$$\begin{aligned}
&\mathbb{E}\{(\overrightarrow{\mathbf{X}} - \overrightarrow{\mathbf{X}}^*)(\overrightarrow{\mathbf{X}} - \overrightarrow{\mathbf{X}}^*)^\top \mid Y > y, Y^* > y^*\} \\
=&\mathbb{E}(\overrightarrow{\mathbf{X}}\overrightarrow{\mathbf{X}}^\top \mid Y > y, Y^* > y^*) + \mathbb{E}(\overrightarrow{\mathbf{X}}^*(\overrightarrow{\mathbf{X}}^*)^\top \mid Y > y, Y^* > y^*) \\
&- \mathbb{E}(\overrightarrow{\mathbf{X}}(\overrightarrow{\mathbf{X}}^*)^\top \mid Y > y, Y^* > y^*) - \mathbb{E}(\overrightarrow{\mathbf{X}}^*\overrightarrow{\mathbf{X}}^\top \mid Y > y, Y^* > y^*).
\end{aligned}$$

By Lemma 2.1 of Li et al. (2005), since $(\overrightarrow{\mathbf{X}}, Y) \perp\!\!\!\perp (\overrightarrow{\mathbf{X}}^*, Y^*)$, then $\overrightarrow{\mathbf{X}} \perp\!\!\!\perp \overrightarrow{\mathbf{X}}^* \mid (Y, Y^*)$, $\overrightarrow{\mathbf{X}} \perp\!\!\!\perp Y^* \mid Y$ and $\overrightarrow{\mathbf{X}}^* \perp\!\!\!\perp Y \mid Y^*$. Therefore, it follows that

$$\begin{aligned}
&\mathbb{E}\{(\overrightarrow{\mathbf{X}} - \overrightarrow{\mathbf{X}}^*)(\overrightarrow{\mathbf{X}} - \overrightarrow{\mathbf{X}}^*)^\top \mid Y > y, Y^* > y^*\} \\
=&\mathbb{E}(\overrightarrow{\mathbf{X}}\overrightarrow{\mathbf{X}}^\top \mid Y > y) + \mathbb{E}(\overrightarrow{\mathbf{X}}^*(\overrightarrow{\mathbf{X}}^*)^\top \mid Y^* > y^*) - \mathbb{E}(\overrightarrow{\mathbf{X}} \mid Y > y)\mathbb{E}((\overrightarrow{\mathbf{X}}^*)^\top \mid Y^* > y^*) \\
&- \mathbb{E}(\overrightarrow{\mathbf{X}}^* \mid Y^* > y^*)\mathbb{E}(\overrightarrow{\mathbf{X}}^\top \mid Y > y) \\
=&2\mathbb{E}(\overrightarrow{\mathbf{X}}\overrightarrow{\mathbf{X}}^\top \mid Y > y) - 2\mathbb{E}(\overrightarrow{\mathbf{X}} \mid Y > y)\mathbb{E}(\overrightarrow{\mathbf{X}}^\top \mid Y > y).
\end{aligned}$$

Therefore, under Assumptions (A1) and (A2),

$$\mathbf{A} = \lim_{y,y^* \to y^+} \mathbb{E}\{(\overrightarrow{\mathbf{X}} - \overrightarrow{\mathbf{X}}^*)(\overrightarrow{\mathbf{X}} - \overrightarrow{\mathbf{X}}^*)^\top \mid Y > y, Y^* > y^*\} = 2(\mathbf{T} - \boldsymbol{\nu}\boldsymbol{\nu}^\top),$$

which completes the proof. □

S5.10 Proof of Theorem 4

By Lemma 6, $\mathbf{A} = 2(\mathbf{T} - \boldsymbol{\nu}\boldsymbol{\nu}^\top)$. Then

$$\mathcal{S}_{\text{eDR}} = \boldsymbol{\Sigma}^{-1}\text{span}(\mathbf{T} - \tau_{\boldsymbol{\beta}}\boldsymbol{\Sigma} - \boldsymbol{\nu}\boldsymbol{\nu}^\top) \subseteq \mathcal{S}_{\text{eSIR}} \oplus \mathcal{S}_{\text{eSAVE}},$$

where $\mathcal{S} \oplus \mathcal{S}^* := \{\mathbf{u} + \mathbf{v} \mid \forall \mathbf{u} \in \mathcal{S}, \mathbf{v} \in \mathcal{S}^*\}$. By Theorems 2 and 3, we have $\mathcal{S}_{\text{eDR}} \subseteq \text{span}(\boldsymbol{\beta})$. □

S5.11 Proof of Lemma 7

In the proof of Lemma 6, we have shown that $\mathbf{A} = 2(\mathbf{T} - \boldsymbol{\nu}\boldsymbol{\nu}^\top)$. Therefore,

$$\mathcal{S}_{\text{eDR}} = \boldsymbol{\Sigma}^{-1}\text{span}(\mathbf{T} - \boldsymbol{\nu}\boldsymbol{\nu}^\top - \tau_{\boldsymbol{\beta}}\boldsymbol{\Sigma}).$$

Recall that $\mathcal{S}_{\text{eSIR}} = \boldsymbol{\Sigma}^{-1}\text{span}(\boldsymbol{\nu})$ and $\mathcal{S}_{\text{eSAVE}} = \boldsymbol{\Sigma}^{-1}\text{span}(\mathbf{T} - \tau_{\boldsymbol{\beta}}\boldsymbol{\Sigma})$, then it follows that

$$\mathcal{S}_{\text{eSAVE}} \subseteq \mathcal{S}_{\text{eDR}} \oplus \mathcal{S}_{\text{eSIR}},$$

where $\mathcal{S} \oplus \mathcal{S}^* := \{\mathbf{u} + \mathbf{v} \mid \forall \mathbf{u} \in \mathcal{S}, \mathbf{v} \in \mathcal{S}^*\}$, and

$$\dim(\mathcal{S}_{\text{eSAVE}}) \leq \dim(\mathcal{S}_{\text{eDR}}) + \dim(\mathcal{S}_{\text{eSIR}}) \leq \dim(\mathcal{S}_{\text{eDR}}) + 1.$$

By assuming that $\dim(\mathcal{S}_{\text{eDR}}) < p/2 - 1$, we obtain $\dim(\mathcal{S}_{\text{eSAVE}}) < p/2$. Then according to Lemma 4, we have $\tau_{\boldsymbol{\beta}} = \text{median}\{\boldsymbol{\sigma}(\boldsymbol{\Sigma}^{-1/2}\mathbf{T}\boldsymbol{\Sigma}^{-1/2})\}$, which completes the proof. □

S5.12 Proof of Lemma 8

Since

$$\begin{aligned}&\frac{1}{h^2}\mathbb{E}\left\{(\overrightarrow{\mathbf{X}}-\overrightarrow{\mathbf{X}}^*)(\overrightarrow{\mathbf{X}}-\overrightarrow{\mathbf{X}}^*)^\top I(\widetilde{Y}<F^{-1}(uh))I(\widetilde{Y}^\star<F^{-1}(u^*h))\right\}\\=&uu^*\mathbb{E}\left\{(\overrightarrow{\mathbf{X}}-\overrightarrow{\mathbf{X}}^*)(\overrightarrow{\mathbf{X}}-\overrightarrow{\mathbf{X}}^*)^\top \mid \widetilde{Y}<F^{-1}(uh),\widetilde{Y}^\star<F^{-1}(u^*h)\right\},\end{aligned}$$

by Lemma 6, under Assumptions (A1) and (A2), for any $u, u^* \in [0,1]$, we have that as $h \to 0$,

$$\frac{1}{h^2}\mathbb{E}\left\{(\overrightarrow{\mathbf{X}}-\overrightarrow{\mathbf{X}}^*)(\overrightarrow{\mathbf{X}}-\overrightarrow{\mathbf{X}}^*)^\top I(\widetilde{Y}<F^{-1}(uh))I(\widetilde{Y}^\star<(F^*)^{-1}(u^*h))\right\}\longrightarrow uu^*\mathbf{A}.$$

For $\mathbf{T}_h(u)$ and $\mathbf{T}_h(u^*)$, under Assumption (A2), for any $u \in [0,1]$, we have that as $h \to 0$,

$$\mathbf{T}_h(u)\longrightarrow u\mathbf{T},\quad \mathbf{T}_h(u^*)\longrightarrow u^*\mathbf{T},$$

which is followed by

$$\begin{aligned}\mathbf{G}_h(u,u^*)\longrightarrow & uu^*\mathbf{A}-2uu^*\cdot\text{median}\{\boldsymbol{\sigma}(\boldsymbol{\Sigma}^{-1/2}\mathbf{T}\boldsymbol{\Sigma}^{-1/2})\}\boldsymbol{\Sigma}\\ =& uu^*(\mathbf{A}-2\tau\boldsymbol{\Sigma}).\end{aligned}$$

Since $\|\overrightarrow{\mathbf{X}}\|$ is bounded from above, by dominated convergence theorem, we have that as $h \to 0$,

$$\int_0^1\int_0^1\mathbf{G}_h(u,u^*)\mathbf{G}_h^\top(u,u^*)\mathrm{d}u\mathrm{d}u^*\longrightarrow\int_0^1\int_0^1 u^2(u^*)^2(\mathbf{A}-2\tau\boldsymbol{\Sigma})(\mathbf{A}-2\tau\boldsymbol{\Sigma})^\top\mathrm{d}u\mathrm{d}u^*.$$

Therefore,

$$\mathbf{M}_{\text{eDR}}^h\longrightarrow\frac{1}{9}(\mathbf{A}-2\tau\boldsymbol{\Sigma})(\mathbf{A}-2\tau\boldsymbol{\Sigma})^\top:=\mathbf{M}_{\text{eDR}},\quad h\to 0. \tag{S5.16}$$

According to Lemma 7, by assuming that $\dim(\mathcal{S}_{\text{eDR}}) < p/2-1$, then $\mathcal{S}_{\text{eDR}} = \mathbf{\Sigma}^{-1}\text{span}(\mathbf{A} - 2\tau\mathbf{\Sigma})$, then $\mathbf{\Sigma}^{-1}\text{span}(\mathbf{M}_{\text{eDR}}) = \mathbf{\Sigma}^{-1}\text{span}(\mathbf{A} - 2\tau\mathbf{\Sigma}) = \mathcal{S}_{\text{eDR}}$, which completes the proof.

S5.13 Proof of Lemma 9

According to the definition of r_i, we have that, we have that,

$$r_i \leq \frac{\sigma_1(\mathbf{M}_{\text{eDR}}) + \varepsilon}{\sigma_{d^*}(\mathbf{M}_{\text{eDR}}) + \varepsilon}, \quad i = 1, \dots, d^* - 1,$$

and

$$r_i = 1, \quad i = d^* + 1, \dots, p - 1.$$

By choosing $\varepsilon > 0$ such that

$$\frac{\sigma_1(\mathbf{M}_{\text{eDR}}) + \varepsilon}{\sigma_{d^*}(\mathbf{M}_{\text{eDR}}) + \varepsilon} \leq \frac{\sigma_{d^*}(\mathbf{M}_{\text{eDR}})}{\varepsilon} + 1,$$

that is,

$$\varepsilon\{\sigma_1(\mathbf{M}_{\text{eDR}}) - 2\sigma_{d^*}(\mathbf{M}_{\text{eDR}})\} < \sigma_{d^*}^2(\mathbf{M}_{\text{eDR}}),$$

then we have $d^* = \text{argmax}_i \{r_i\}$.

In Theorem 5, we have shown that $\|\widehat{\mathbf{M}}_{\text{eDR}}^{k/n} - \mathbf{M}_{\text{eDR}}\|_F = o_P(1)$ as $n \to \infty$. Then by Weyl's inequality, for $i = 1, \dots, p - 1$, we have

$$|\sigma_i(\widehat{\mathbf{M}}_{\text{eDR}}^{k/n}) - \sigma_i(\mathbf{M}_{\text{eDR}})| = o_P(1), \quad n \to \infty.$$

Then, for $i = 1, \ldots, p-1$,

$$|\widehat{r}_i - r_i| = o_P(1), \quad n \to \infty,$$

which completes the proof.

S5.14 Proofs of Theorems 5, S1, and S2

We need the following two consistency results of the estimated functions $\widehat{\mathbf{C}}_{k/n}(u)$ and $\widehat{\mathbf{T}}_{k/n}(u)$.

Lemma S3. *Under the EC distribution assumption of* $\mathbf{X}$ *and Assumptions (A1) & (A3), as* $n \to \infty$*, we have*

$$\sqrt{k}\left(\widehat{\mathbf{C}}_{k/n}(u) - \mathbf{C}_{k/n}(u)\right) \xrightarrow{d} \mathbf{W}_{\mathbf{C}}(u), \quad u \in [0,1],$$

where $\mathbf{W}_{\mathbf{C}}(u)$ *is a Gaussian process with mean zero and covariance structure*

$$\mathbb{E}\left(\mathbf{W}_{\mathbf{C}}(u_1)\mathbf{W}_{\mathbf{C}}(u_2)\right) = \min(u_1, u_2) \lim_{y \to y^+} \mathrm{Cov}(\overrightarrow{\mathbf{X}}|Y > y), \quad u_1, u_2 \in [0,1].$$

provided that $\lim_{y \to y^+} \mathrm{Cov}(\overrightarrow{\mathbf{X}}|Y > y)$ *exists.*

Lemma S4. *Under the EC distribution assumption of* $\mathbf{X}$ *and Assumptions (A2) & (A3), as* $n \to \infty$*, we have*

$$\sqrt{k}\left(\mathrm{vec}(\widehat{\mathbf{T}}_{k/n}(u)) - \mathrm{vec}(\mathbf{T}_{k/n}(u))\right) \xrightarrow{d} \mathbf{W}_{\mathbf{T}}(u), \quad u \in [0,1],$$

where $\mathbf{W}_{\mathbf{T}}(u)$ *is a Gaussian process with mean zero and covariance structure*

$$\mathbb{E}\left(\mathbf{W}_{\mathbf{T}}(u_1)\mathbf{W}_{\mathbf{T}}(u_2)\right) = \min(u_1, u_2)\left\{\lim_{y \to y^+} \mathrm{Cov}\{\mathrm{vec}(\overrightarrow{\mathbf{X}}\overrightarrow{\mathbf{X}}^T)|Y > y\}\right\}, \quad u_1, u_2 \in [0,1].$$

provided that $\lim_{y \to y^+} \mathrm{Cov}\{\mathrm{vec}(\overrightarrow{\mathbf{X}}\overrightarrow{\mathbf{X}}^T)|Y > y\}$ *exists.*

Proof of Lemmas S3 and S4. We have

$$\begin{aligned}
\widehat{\mathbf{C}}_{k/n}(u) - \mathbf{C}_{k/n}(u) &= \frac{1}{k}\sum_{i=1}^{n} \overrightarrow{\boldsymbol{x}}_i I\left(\widetilde{Y}_i \le \widehat{F}^{-1}(uk/n)\right) - \frac{n}{k}\mathbb{E}\left\{\overrightarrow{\mathbf{X}} I\left(\widetilde{Y} < F^{-1}(uk/n)\right)\right\} \\
&= \frac{1}{k}\sum_{i=1}^{n}\left\{\frac{\mathbf{X}_i - \widehat{\boldsymbol{\mu}}}{\|\mathbf{X}_i - \widehat{\boldsymbol{\mu}}\|_{\widehat{\boldsymbol{\Sigma}}}} - \frac{\mathbf{X}_i - \boldsymbol{\mu}}{\|\mathbf{X}_i - \boldsymbol{\mu}\|_{\widehat{\boldsymbol{\Sigma}}}}\right\} I\left(\widetilde{Y}_i \le \widehat{F}^{-1}(uk/n)\right) \\
&\quad + \frac{1}{k}\sum_{i=1}^{n}\left\{\frac{\mathbf{X}_i - \boldsymbol{\mu}}{\|\mathbf{X}_i - \boldsymbol{\mu}\|_{\widehat{\boldsymbol{\Sigma}}}} - \frac{\mathbf{X}_i - \boldsymbol{\mu}}{\|\mathbf{X}_i - \boldsymbol{\mu}\|_{\boldsymbol{\Sigma}}}\right\} I\left(\widetilde{Y}_i \le \widehat{F}^{-1}(uk/n)\right) \\
&\quad + \left[\frac{1}{k}\sum_{i=1}^{n} \overrightarrow{\mathbf{X}}_i I\left(\widetilde{Y}_i \le \widehat{F}^{-1}(uk/n)\right) - \frac{n}{k}\mathbb{E}\left\{\overrightarrow{\mathbf{X}} I\left(\widetilde{Y} < F^{-1}(uk/n)\right)\right\}\right] \\
&=: \mathbf{I}_1(u) + \mathbf{I}_2(u) + \mathbf{I}_3(u).
\end{aligned}$$

By Theorem 3 in Aghbalou et al. (2024), we have that,

$$\sqrt{k}\mathbf{I}_3(u) \xrightarrow{d} \mathbf{W}_C(u).$$

Thus, it remains to show that, as $n \to \infty$,

$$\|\mathbf{I}_1(u)\| = o_P(k^{-1/2}),$$

$$\|\mathbf{I}_2(u)\| = o_P(k^{-1/2}),$$

uniformly for $u \in [0, 1]$.

For the sake of notation simplicity, we assume without loss of generality that $\boldsymbol{\mu} = \mathbf{0}$ and $\boldsymbol{\Sigma} = I_p$. For $\mathbf{I}_1(u)$, define

$$g_j(t) = \frac{1}{k}\sum_{i=1}^{n}(X_{ij} - t\widehat{\mu}_j)\|\mathbf{X}_i - t\widehat{\boldsymbol{\mu}}\|_{\widehat{\boldsymbol{\Sigma}}}^{-1} I\left(\widetilde{Y}_i \le \widehat{F}^{-1}(uk/n)\right).$$

where $0 \le t \le 1$ and $j = 1, 2, \ldots, p$. By the mean-value theorem, we have that $g_j(1) - g_j(0) = \dot{g}_j(\tilde{t})$ for some $0 \le \tilde{t} \le 1$ and $\dot{g}_j$ is the first-order derivative of g_j:

$$\dot{g}_j(t) = -\widehat{\mu}_j \left\{ \frac{1}{k} \sum_{i=1}^n \|\mathbf{X}_i - t\widehat{\boldsymbol{\mu}}\|_{\widehat{\boldsymbol{\Sigma}}}^{-1} I\left(\widetilde{Y}_i \leq \widehat{F}^{-1}(uk/n)\right) \right\}$$
$$+ \frac{1}{k} \sum_{i=1}^n (X_{ij} - t\widehat{\mu}_j) \|\mathbf{X}_i - t\widehat{\boldsymbol{\mu}}\|_{\widehat{\boldsymbol{\Sigma}}}^{-3} (\mathbf{X}_i - t\widehat{\boldsymbol{\mu}})^T \widehat{\boldsymbol{\Sigma}}^{-1} \widehat{\boldsymbol{\mu}} I\left(\widetilde{Y}_i \leq \widehat{F}^{-1}(uk/n)\right).$$

Let $\widetilde{\boldsymbol{\mu}} = \tilde{t}\widehat{\boldsymbol{\mu}}$. Then, the j-th element of $\mathbf{I}_1(u)$ is

$$\begin{aligned}
\left| I_1^{(j)}(u) \right| &= |g_j(1) - g_j(0)| = \left|\dot{g}_j(\tilde{t})\right| \\
&\leq \frac{1}{k} \sum_{i=1}^n \frac{|\widehat{\mu}_j|}{\|\mathbf{X}_i - \widetilde{\boldsymbol{\mu}}\|_{\widehat{\boldsymbol{\Sigma}}}} I\left(\widetilde{Y}_i \leq \widehat{F}^{-1}(uk/n)\right) \\
&\quad + \frac{1}{k} \sum_{i=1}^n \frac{(X_{ij} - \widetilde{\mu}_j) \left\{ \widehat{\boldsymbol{\Sigma}}^{-1}(\mathbf{X}_i - \widetilde{\boldsymbol{\mu}}) \right\}^T \widehat{\boldsymbol{\mu}}}{\|\mathbf{X}_i - \widetilde{\boldsymbol{\mu}}\|_{\widehat{\boldsymbol{\Sigma}}}^3} I\left(\widetilde{Y}_i \leq \widehat{F}^{-1}(uk/n)\right) \\
&=: I_{1,1}^{(j)}(u) + I_{1,2}^{(j)}(u).
\end{aligned}$$

Let $p = 3$ and $q = 3/2$. The $1/p + 1/q = 1$. By applying the Hölder' inequality, we have that,

$$\begin{aligned}
I_{1,1}^{(j)}(u) &\leq \frac{1}{k} \left\{ \sum_{i=1}^n \frac{|\widehat{\mu}_j|^p}{\|\mathbf{X}_i - \widetilde{\boldsymbol{\mu}}\|_{\widehat{\boldsymbol{\Sigma}}}^p} \right\}^{1/p} \left\{ \sum_{i=1}^n I\left(\widetilde{Y}_i \leq \widehat{F}^{-1}(uk/n)\right) \right\}^{1/q} \\
&\leq \frac{1}{k} \left\{ \sum_{i=1}^n \frac{\|\widehat{\boldsymbol{\mu}}\|^p}{\|\mathbf{X}_i - \widetilde{\boldsymbol{\mu}}\|_{\widehat{\boldsymbol{\Sigma}}}^p} \right\}^{1/p} k^{1/q}.
\end{aligned}$$

Note that,

$$\begin{aligned}
\sum_{i=1}^n \frac{\|\widehat{\boldsymbol{\mu}}\|^p}{\|\mathbf{X}_i - \widetilde{\boldsymbol{\mu}}\|_{\widehat{\boldsymbol{\Sigma}}}^p} &\leq \sum_{i=1}^n \frac{\|\widehat{\boldsymbol{\mu}}\|^p}{\|\widehat{\boldsymbol{\Sigma}}\|^{-p/2} \|\mathbf{X}_i - \widetilde{\boldsymbol{\mu}}\|^p} \\
&= n\|\widehat{\boldsymbol{\mu}}\|^p \times \|\widehat{\boldsymbol{\Sigma}}\|^{p/2} \times \left\{ \frac{1}{n} \sum_{i=1}^n \frac{1}{\|\mathbf{X}_i - \widetilde{\boldsymbol{\mu}}\|^p} \right\}
\end{aligned}$$

By Theorem 4.2 of Tyler (1987), we have $\widehat{\boldsymbol{\mu}} = O_P(n^{-1/2})$ and $\widehat{\boldsymbol{\Sigma}} = O_P(1)$ as $n \to \infty$. Under Assumption (A3), by (A27) of Luo et al. (2009), we have

$\frac{1}{n}\sum_{i=1}^{n} 1/\|\mathbf{X}_i - \widetilde{\boldsymbol{\mu}}\|^p = O_P(1)$. Thus, as $n \to \infty$,

$$I_{1,1}^{(j)}(u) = O_P\left\{\frac{1}{k}\left(nn^{-p/2}\right)^{1/p} k^{1/q}\right\} = O_P(n^{-1/6}k^{-1/3}) = O_P(k^{-1/2}(k/n)^{1/6}) = o_P(k^{-1/2}),$$

uniformly for $u \in [0,1]$. Similarly, we can show that,

$$I_{1,2}^{(j)}(u) = o_P(1),$$

and hence

$$\|\mathbf{I}_1(u)\| = o_P(1),$$

uniformly for $u \in [0,1]$ as $n \to \infty$.

Next, we consider $\mathbf{I}_2(u)$. For the j-th element of $\mathbf{I}_2(u)$,

$$\begin{aligned}
\left|I_2^{(j)}(u)\right| &= \left|\frac{1}{k}\sum_{i=1}^{n}\left\{\frac{X_{ij}}{\|\mathbf{X}_i\|_{\widehat{\boldsymbol{\Sigma}}}} - \frac{X_{ij}}{\|\mathbf{X}_i\|}\right\} I\left(\widetilde{Y}_i \leq \widehat{F}^{-1}(uk/n)\right)\right| \\
&= \left|\frac{1}{k}\sum_{i=1}^{n}\left\{\frac{X_{ij}}{\|\mathbf{X}_i\|_{\widehat{\boldsymbol{\Sigma}}}\|\mathbf{X}_i\|}\cdot\frac{\|\mathbf{X}_i\|^2 - \|\mathbf{X}\|^2_{\widehat{\boldsymbol{\Sigma}}}}{\|\mathbf{X}_i\| + \|\mathbf{X}\|_{\widehat{\boldsymbol{\Sigma}}}}\cdot I\left(\widetilde{Y}_i \leq \widehat{F}^{-1}(uk/n)\right)\right\}\right| \\
&= \left|\frac{1}{k}\sum_{i=1}^{n}\left\{\frac{X_{ij}}{\|\mathbf{X}_i\|_{\widehat{\boldsymbol{\Sigma}}}\|\mathbf{X}_i\|}\cdot\frac{\mathbf{X}_i^T\left(\mathbf{I}_p - \widehat{\boldsymbol{\Sigma}}^{-1}\right)\mathbf{X}_i}{\|\mathbf{X}_i\| + \|\mathbf{X}\|_{\widehat{\boldsymbol{\Sigma}}}}\cdot I\left(\widetilde{Y}_i \leq \widehat{F}^{-1}(uk/n)\right)\right\}\right| \\
&\leq \|\mathbf{I}_p - \widehat{\boldsymbol{\Sigma}}^{-1}\|\left|\frac{1}{k}\sum_{i=1}^{n}\left\{\frac{\|\widehat{\boldsymbol{\Sigma}}\|^{1/2}\|\mathbf{X}_i\|}{\|\mathbf{X}_i\|\|\mathbf{X}_i\|}\cdot\frac{\|\mathbf{X}_i\|^2}{\|\mathbf{X}_i\| + \|\widehat{\boldsymbol{\Sigma}}\|^{-1/2}\|\mathbf{X}_i\|}\cdot I\left(\widetilde{Y}_i \leq \widehat{F}^{-1}(uk/n)\right)\right\}\right| \\
&= \|\mathbf{I}_p - \widehat{\boldsymbol{\Sigma}}^{-1}\|\frac{1}{k}\|\widehat{\boldsymbol{\Sigma}}\|^{1/2}\left\{1 + \|\widehat{\boldsymbol{\Sigma}}\|^{-1/2}\right\}^{-1}\sum_{i=1}^{n} I\left(\widetilde{Y}_i \leq \widehat{F}^{-1}(uk/n)\right) \\
&\leq \frac{1}{k}\|\mathbf{I}_p - \widehat{\boldsymbol{\Sigma}}^{-1}\|\|\widehat{\boldsymbol{\Sigma}}\|^{1/2}\left\{1 + \|\widehat{\boldsymbol{\Sigma}}\|^{-1/2}\right\}^{-1} k \\
&= \|\mathbf{I}_p - \widehat{\boldsymbol{\Sigma}}^{-1}\|\|\widehat{\boldsymbol{\Sigma}}\|^{1/2}\left\{1 + \|\widehat{\boldsymbol{\Sigma}}\|^{-1/2}\right\}^{-1}.
\end{aligned}$$

By Theorem 4.2 of Tyler (1987), we have that, as $n \to \infty$,

$$\|\mathbf{I}_p - \widehat{\boldsymbol{\Sigma}}^{-1}\| = O_P(n^{-1/2}),$$

and

$$\|\widehat{\boldsymbol{\Sigma}}\|^{1/2} = O_P(1), \quad \|\widehat{\boldsymbol{\Sigma}}\|^{-1/2} = O_P(1).$$

Thus, as $n \to \infty$, $\|I_2(u)\| = o_P(k^{-1/2})$ uniformly for $u \in [0,1]$ and the proof is complete.

The proof of Lemma S4 is similar to that of Lemma S3 and is thus omitted. □

We prove Theorem 5 as an example. The proofs of Theorems S1 and S2 are similar to that of Theorem 5 and are thus omitted. By integrating $\widehat{\mathbf{C}}_{k/n}(u)$ and $\widehat{\mathbf{D}}_{k/n}(u)$ on u over $[0,1]$, according to Lemma S3 and S4, we have that,

$$\|\widehat{\mathbf{M}}_{\text{eDR}}^{k/n} - \mathbf{M}_{\text{eDR}}^{k/n}\|_F = O_P(k^{-1/2}), \quad n \to \infty. \tag{S5.17}$$

Note that,

$$\|\widehat{\mathbf{M}}_{\text{eDR}}^{k/n} - \mathbf{M}_{\text{eDR}}\|_F \leq \|\widehat{\mathbf{M}}_{\text{eDR}}^{k/n} - \mathbf{M}_{\text{eDR}}^{k/n}\|_F + \|\mathbf{M}_{\text{eDR}}^{k/n} - \mathbf{M}_{\text{eDR}}\|_F,$$

According to (S5.17) and the result in Lemma 8, we have that,

$$\|\widehat{\mathbf{M}}_{\text{eDR}}^{k/n} - \mathbf{M}_{\text{eDR}}\|_F = o_P(1), \quad n \to \infty.$$

In addition,

$$\|\widehat{\boldsymbol{\Sigma}}^{-1}\widehat{\mathbf{M}}_{\text{eDR}}^{k/n} - \boldsymbol{\Sigma}^{-1}\mathbf{M}_{\text{eDR}}\|_F \leq \|\widehat{\boldsymbol{\Sigma}}^{-1}\|\|\widehat{\mathbf{M}}_{\text{eDR}}^{k/n} - \mathbf{M}_{\text{eDR}}\|_F + \|\mathbf{M}_{\text{eDR}}\|\|\widehat{\boldsymbol{\Sigma}}^{-1} - \mathbf{I}_p\|_F.$$

Since we have

$$\|\widehat{\boldsymbol{\Sigma}}^{-1}\| = O_P(1), \quad \|\mathbf{I}_p - \widehat{\boldsymbol{\Sigma}}^{-1}\|_F = O_P(n^{-1/2}),$$

then

$$\|\widehat{\boldsymbol{\Sigma}}^{-1}\widehat{\mathbf{M}}_{\text{eDR}}^{k/n} - \boldsymbol{\Sigma}^{-1}\mathbf{M}_{\text{eDR}}\|_F = o_P(1).$$

Recall that under the assumption that $\dim(\mathcal{S}_{\text{eDR}}) < p/2 - 1$, we have $\boldsymbol{\Sigma}^{-1}\text{span}(\mathbf{M}_{\text{eDR}}) = \mathcal{S}_{\text{eDR}} = \mathcal{S}_{Y_\infty|\mathbf{X}}$. Since $\widehat{\boldsymbol{\beta}}_{\text{eDR}}^{k/n} = \text{SVD}_{d^*}(\widehat{\boldsymbol{\Sigma}}^{-1}\widehat{\mathbf{M}}_{\text{eDR}}^{k/n})$. Then by Theorem 2 in Yu et al. (2015), we have

$$\|\mathbf{P}_{\widehat{\boldsymbol{\beta}}_{\text{eDR}}^{k/n}} - \mathbf{P}_{\mathcal{S}_{Y_\infty|\mathbf{X}}}\|_F \leq \frac{2\sqrt{2}\|\widehat{\boldsymbol{\Sigma}}^{-1}\widehat{\mathbf{M}}_{\text{eDR}}^{k/n} - \boldsymbol{\Sigma}^{-1}\mathbf{M}_{\text{eDR}}\|_F}{\sigma_{d^*}(\boldsymbol{\Sigma}^{-1}\mathbf{M}_{\text{eDR}})}.$$

Since $\dim(\boldsymbol{\Sigma}^{-1}\mathbf{M}_{\text{eDR}}) = d^*$, then $\sigma_{d^*}(\boldsymbol{\Sigma}^{-1}\mathbf{M}_{\text{eDR}}) > 0$. And we have that

$$\|\mathbf{P}_{\widehat{\boldsymbol{\beta}}_{\text{eDR}}^{k/n}} - \mathbf{P}_{\mathcal{S}_{Y_\infty|\mathbf{X}}}\|_F = o_P(1).$$

□

S5.15 Proofs of Corollaries S1, S2, and S3

We only prove Corollary S1, the proofs for Corollaries S2 and S3 are similar and thus omitted.

In the proof of Theorem 5, we obtain that

$$\|\widehat{\mathbf{M}}_{\text{eDR}}^{k/n} - \mathbf{M}_{\text{eDR}}\|_F \leq \|\widehat{\mathbf{M}}_{\text{eDR}}^{k/n} - \mathbf{M}_{\text{eDR}}^{k/n}\|_F + \|\mathbf{M}_{\text{eDR}}^{k/n} - \mathbf{M}_{\text{eDR}}\|_F$$

and that

$$\|\widehat{\mathbf{M}}_{\text{eDR}}^{k/n} - \mathbf{M}_{\text{eDR}}^{k/n}\|_F = O_P(k^{-1/2}), \quad n \to \infty.$$

It only remains to show that as $n \to \infty$,

$$\|\mathbf{M}_{\text{eDR}}^{k/n} - \mathbf{M}_{\text{eDR}}\|_F = O(k^{-1/2}). \tag{S5.18}$$

Recall that

$$\mathbf{C}_h(k/n) = u\mathbb{E}\left\{\overrightarrow{\mathbf{X}} \mid \widetilde{Y} < F^{-1}(uk/n)\right\}.$$

Under Assumption (A1′), we have that,

$$\|\mathbf{C}_h(u) - u\boldsymbol{\nu}\| = u\|\mathbf{a}(-F^{-1}(uk/n))\| = O(k^{-1/2}),$$

uniformly for all $u \in [0,1]$. Similarly, we can show that

$$\|\mathbf{T}_h(u) - u\mathbf{T}\|_F = O(k^{-1/2}),$$

uniformly for all $u \in [0,1]$. Then (S5.18) holds since $\mathbf{M}_{\text{eDR}}^{k/n}$ are constructed by $\mathbf{C}_h(u)$ and $\mathbf{T}_h(u)$.

S5.16 Proof of Lemma S1

We have

$$\begin{aligned}&\frac{\Pr(Y > y \mid \mathbf{X}) - \Pr(Y > y \mid \boldsymbol{\beta}_2^\top \mathbf{X})}{\Pr(Y > y)}\\ =&\frac{\Pr(Y > y \mid \mathbf{X}) - \Pr(Y > y \mid \boldsymbol{\beta}_1^\top \mathbf{X})}{\Pr(Y > y)} + \frac{\Pr(Y > y \mid \boldsymbol{\beta}_1^\top \mathbf{X}) - \Pr(Y > y \mid \boldsymbol{\beta}_2^\top \mathbf{X})}{\Pr(Y > y)}.\end{aligned} \tag{S5.19}$$

Since $\mathcal{S}_{\boldsymbol{\beta}_1}$ is an EDR subspace of Y given $\mathbf{X}$, then for any $\varepsilon > 0$, there exists some constant y_0 such that for any $y \geq y_0$, we have

$$\left| \frac{\Pr(Y > y \mid \mathbf{X} = \boldsymbol{x}) - \Pr(Y > y \mid \boldsymbol{\beta}_1^\top \mathbf{X} = \boldsymbol{\beta}_1^\top \boldsymbol{x})}{\Pr(Y > y)} \right| \leq \varepsilon, \quad \forall \boldsymbol{x} \in \Omega_{\mathbf{X}}. \tag{S5.20}$$

Meanwhile,

$$\begin{aligned} & \frac{\Pr(Y > y \mid \boldsymbol{\beta}_1^\top \mathbf{X}) - \Pr(Y > y \mid \boldsymbol{\beta}_2^\top \mathbf{X})}{\Pr(Y > y)} \\ =& \frac{\mathbb{E}\left\{\Pr(Y > y \mid \boldsymbol{\beta}_1^\top \mathbf{X}) \mid \boldsymbol{\beta}_2^\top \mathbf{X}\right\} - \mathbb{E}\left\{\Pr(Y > y \mid \mathbf{X}) \mid \boldsymbol{\beta}_2^\top \mathbf{X}\right\}}{\Pr(Y > y)} \\ =& \mathbb{E}\left\{ \frac{\Pr(Y > y \mid \boldsymbol{\beta}_1^\top \mathbf{X}) - \Pr(Y > y \mid \mathbf{X})}{\Pr(Y > y)} \middle| \boldsymbol{\beta}_2^\top \mathbf{X} \right\}. \end{aligned}$$

Then,

$$\left| \frac{\Pr(Y > y \mid \boldsymbol{\beta}_1^\top \mathbf{X}) - \Pr(Y > y \mid \boldsymbol{\beta}_2^\top \mathbf{X})}{\Pr(Y > y)} \right| \leq \mathbb{E}\left\{ \left| \frac{\Pr(Y > y \mid \boldsymbol{\beta}_1^\top \mathbf{X}) - \Pr(Y > y \mid \mathbf{X})}{\Pr(Y > y)} \right| \middle| \boldsymbol{\beta}_2^\top \mathbf{X} \right\}.$$

Combined with (S5.20), we have

$$\left| \frac{\Pr(Y > y \mid \boldsymbol{\beta}_1^\top \mathbf{X} = \boldsymbol{\beta}_1^\top \boldsymbol{x}) - \Pr(Y > y \mid \boldsymbol{\beta}_2^\top \mathbf{X} = \boldsymbol{\beta}_2^\top \boldsymbol{x})}{\Pr(Y > y)} \right| \leq \varepsilon, \quad \forall \boldsymbol{x} \in \Omega_{\mathbf{X}}.$$

Therefore, for any $\varepsilon > 0$, there exists some constant y_0 such that for any $y \geq y_0$, we have

$$\left| \frac{\Pr(Y > y \mid \mathbf{X} = \boldsymbol{x}) - \Pr(Y > y \mid \boldsymbol{\beta}_2^\top \mathbf{X} = \boldsymbol{\beta}_2^\top \boldsymbol{x})}{\Pr(Y > y)} \right| \leq \varepsilon, \quad \forall \boldsymbol{x} \in \Omega_{\mathbf{X}},$$

which finishes the proof.

References

Acharya, V. V., Pedersen, L. H., Philippon, T. & Richardson, M. (2017), 'Measuring systemic risk', *The review of financial studies* **30**(1), 2–47.

Aghbalou, A., Portier, F., Sabourin, A. & Zhou, C. (2024), 'Tail inverse regression: dimension reduction for prediction of extremes', *Bernoulli* **30**(1), 503–533.

Cai, J.-J., Einmahl, J. H., de Haan, L. & Zhou, C. (2015), 'Estimation of the marginal expected shortfall: the mean when a related variable is extreme', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **77**(2), 417–442.

Cook, R. D. (1998), *Regression Graphics: Ideas for Studying Regressions Through Graphics*, Vol. 318, John Wiley & Sons.

de Haan, L. & Ferreira, A. (2006), *Extreme Value Theory: An Introduction*, Vol. 3, Springer.

Hua, L. & Joe, H. (2014), 'Strength of tail dependence based on conditional tail expectation', *Journal of Multivariate Analysis* **123**, 143–159.

Johnson, M. E. (1987), *Multivariate statistical simulation: A guide to selecting and generating continuous multivariate distributions*, Vol. 192, John Wiley & Sons.

Lalancette, M., Engelke, S. & Volgushev, S. (2021), 'Rank-based estimation under asymptotic dependence and independence, with applications to spatial extremes', *The Annals of Statistics* **49**(5), 2552–2576.

Li, B. (2018), *Sufficient dimension reduction: Methods and applications with R*, Chapman and

Hall/CRC.

Li, B. & Wang, S. (2007), 'On directional regression for dimension reduction', *Journal of the American Statistical Association* **102**(479), 997–1008.

Li, B., Zha, H. & Chiaromonte, F. (2005), 'Contour regression: a general approach to dimension reduction', *Annals of Statistics* **33**(4), 1580–1616.

Luo, R., Wang, H. & Tsai, C.-L. (2009), 'Contour projected dimension reduction', *The Annals of Statistics* **37**(6), 3743–3778.

Tyler, D. E. (1987), 'A Distribution-Free M-Estimator of Multivariate Scatter', *The Annals of Statistics* **15**(1), 234–251.

Yu, Y., Wang, T. & Samworth, R. J. (2015), 'A useful variant of the davis–kahan theorem for statisticians', *Biometrika* **102**(2), 315–323.