Supplementary Materials for "Information projection approach to smoothed propensity score weighting for handling selection bias under missing at random"

Hengfang Wang   Jae Kwang Kim

The supplementary material contains proof of Lemma 1, Theorem 1, Theorem 2 and Corollary 1, regularity conditions of Corollary 2 and derivation details of equation (55) and (56) for Simulation Study Two.

S1 Proof of Lemma 1

We now introduce information projection (I-projection) to derive density ration estimation. Let Π be a non-empty closed, convex set of distributions. The I-projection of $\mathbb{Q}$ onto Π is $\mathbb{P}^\star \in \Pi$ such that

$$D(\mathbb{P}^\star \parallel \mathbb{Q}) = \min_{\mathbb{P}\in\Pi} D(\mathbb{P} \parallel \mathbb{Q}).$$

One important family of distributions is a linear family:

$$\mathcal{L} = \left\{\mathbb{P} : \int B_i(\boldsymbol{x})d\mathbb{P}(\boldsymbol{x}) = \alpha_i, i = 1, \cdots, k\right\} \subset \Pi,$$

where $B_i(\cdot)$'s are Lebesgue integrable functions. Note that the linear family is orthogonal to $B_i(\cdot) - \alpha_i$ for $i = 1, \cdots, k$. Since the function $D(\mathbb{P} \parallel \mathbb{Q})$ is continuous and strictly convex in $\mathbb{P}$, so that $\mathbb{P}^\star$ satisfying

$$D(\mathbb{P}^\star \parallel \mathbb{Q}) = \min_{\mathbb{P}\in\mathcal{L}} D(\mathbb{P} \parallel \mathbb{Q})$$

exists and is unique. Moreover, $\mathbb{P}^\star$, the I-projection of $\mathbb{Q}$ onto $\mathcal{L}$ is of the form

$$\mathbb{P}^\star(x) = \mathbb{Q}(x)\frac{\exp\left\{\sum_{i=1}^K \beta_i B_i(\boldsymbol{x})\right\}}{\mathbb{E}_{\mathbb{Q}}\left[\exp\left\{\sum_{i=1}^K \beta_i B_i(\boldsymbol{x})\right\}\right]}, \tag{S.1}$$

where β_i's are constants in $\mathbb{R}$. The solution $\mathbb{P}^\star(\boldsymbol{x})$ in (S.1) is an exponential tilting of the density $\mathbb{Q}(\boldsymbol{x})$ with the moment restrictions on $B_i(\boldsymbol{x})$, $i = 1, \cdots, k$.

To derive valid DRE, we leverage the I-projection theory with $\mathbb{Q} = \mathbb{P}_1$ and $\mathbb{P} = \mathbb{P}_0$ whose densities are $f_1(\boldsymbol{x})$ and $f_0(\boldsymbol{x})$, respectively. The linear space that we are projecting on is

$$p\int \boldsymbol{b}(x) f_1(\boldsymbol{x}) d\mu + (1-p)\int \boldsymbol{b}(\boldsymbol{x}) f_0(\boldsymbol{x}) d\mu = \mathbb{E}\{\boldsymbol{b}(X)\}. \tag{S.2}$$

By (S.1), the I-projection solution is

$$f_0^\star(\boldsymbol{x}) = f_1(\boldsymbol{x}) \times \frac{\exp\{\boldsymbol{\lambda}_1^{\mathrm{T}}\boldsymbol{b}(\boldsymbol{x})\}}{\mathbb{E}_1\left[\exp\{\boldsymbol{\lambda}_1^{\mathrm{T}}\boldsymbol{b}(\boldsymbol{x})\}\right]},$$

where $\boldsymbol{\lambda}_1$ is chosen to satisfy (S.2).

S2 Proof of Theorem 1

The detailed proof for Theorem 1 is presented as follows.

Proof. By Assumption 1∼ 3 and Corollary II.2 of **?**, we have

$$\widehat{\boldsymbol{\lambda}}_{\boldsymbol{\theta}} - \boldsymbol{\lambda}_{\boldsymbol{\theta}}^\star = o_p(1).$$

Then, by mean value theorem, we have

$$\boldsymbol{U}_{\mathrm{B},N}(\widehat{\boldsymbol{\lambda}}_{\boldsymbol{\theta}}) - \boldsymbol{U}_{\mathrm{B},N}(\boldsymbol{\lambda}^{\star}_{\boldsymbol{\theta}}) = \frac{\partial}{\partial \boldsymbol{\lambda}}\boldsymbol{U}_{\mathrm{B},N}(\widetilde{\boldsymbol{\lambda}}_{\boldsymbol{\theta}}) \left(\widehat{\boldsymbol{\lambda}}_{\boldsymbol{\theta}} - \boldsymbol{\lambda}^{\star}_{\boldsymbol{\theta}}\right), \tag{S.3}$$

where $\widetilde{\boldsymbol{\lambda}}_{\boldsymbol{\theta}}$ is a point between $\boldsymbol{\lambda}^{\star}_{\boldsymbol{\theta}}$ and $\widehat{\boldsymbol{\lambda}}_{\boldsymbol{\theta}}$. Apparently, $\partial \boldsymbol{U}_{\mathrm{B},N}/\partial \boldsymbol{\lambda}$ and $\mathbb{E}[\partial \boldsymbol{U}_{\mathrm{B},N}/\partial \boldsymbol{\lambda}]$ are continuous within the set $\mathcal{G}_1$. Therefore, using the similar technique as above, we can arrive at

$$\frac{\partial}{\partial \boldsymbol{\lambda}}\boldsymbol{U}_{\mathrm{B},N}(\widetilde{\boldsymbol{\lambda}}_{\boldsymbol{\theta}}) = \mathbb{E}\left\{\frac{\partial}{\partial \boldsymbol{\lambda}}\boldsymbol{U}_{\mathrm{B},N}(\boldsymbol{\lambda}^{\star}_{\boldsymbol{\theta}})\right\} + o_p(1). \tag{S.4}$$

Combine (S.3), (S.4) and the fact that $\widehat{\boldsymbol{U}}_2(\widehat{\boldsymbol{\lambda}}) = \mathbf{0}$, we have

$$-\sqrt{N}\boldsymbol{U}_{\mathrm{B},N}(\boldsymbol{\lambda}^{\star}_{\boldsymbol{\theta}}) = \sqrt{N}\mathbb{E}\left\{\frac{\partial}{\partial \boldsymbol{\lambda}}\boldsymbol{U}_{\mathrm{B},N}(\boldsymbol{\lambda}^{\star}_{\boldsymbol{\theta}})\right\}(\widehat{\boldsymbol{\lambda}}_{\boldsymbol{\theta}} - \boldsymbol{\lambda}^{\star}_{\boldsymbol{\theta}}) + o_p\left(\sqrt{N}\left\|\widehat{\boldsymbol{\lambda}}_{\boldsymbol{\theta}} - \boldsymbol{\lambda}^{\star}_{\boldsymbol{\theta}}\right\|\right). \tag{S.5}$$

Then, by Cauchy-Schwarz inequality, we have

$$\begin{aligned}
\sqrt{N}\left\|\widehat{\boldsymbol{\lambda}}_{\boldsymbol{\theta}} - \boldsymbol{\lambda}^{\star}_{\boldsymbol{\theta}}\right\| &\leqslant \left\|\mathbb{E}\left\{\frac{\partial}{\partial \boldsymbol{\lambda}}\boldsymbol{U}_{\mathrm{B},N}(\boldsymbol{\lambda}^{\star}_{\boldsymbol{\theta}})\right\}^{-1}\right\|\left\|\sqrt{N}\mathbb{E}\left\{\frac{\partial}{\partial \boldsymbol{\lambda}}\boldsymbol{U}_{\mathrm{B},N}(\boldsymbol{\lambda}^{\star}_{\boldsymbol{\theta}})\right\}(\widehat{\boldsymbol{\lambda}}_{\boldsymbol{\theta}} - \boldsymbol{\lambda}^{\star}_{\boldsymbol{\theta}})\right\| \\
&= \left\|\mathbb{E}\left\{\frac{\partial}{\partial \boldsymbol{\lambda}}\boldsymbol{U}_{\mathrm{B},N}(\boldsymbol{\lambda}^{\star}_{\boldsymbol{\theta}})\right\}^{-1}\right\|\left\|\sqrt{N}\boldsymbol{U}_{\mathrm{B},N}(\boldsymbol{\lambda}^{\star}_{\boldsymbol{\theta}}) + o_p\left(\sqrt{N}\left\|\widehat{\boldsymbol{\lambda}}_{\boldsymbol{\theta}} - \boldsymbol{\lambda}^{\star}_{\boldsymbol{\theta}}\right\|\right)\right\| \\
&= \mathcal{O}_p(1) + o_p\left(\sqrt{N}\left\|\widehat{\boldsymbol{\lambda}}_{\boldsymbol{\theta}} - \boldsymbol{\lambda}^{\star}_{\boldsymbol{\theta}}\right\|\right),
\end{aligned}$$

which implies the root-n convergence of $\widehat{\boldsymbol{\lambda}}_{\boldsymbol{\theta}}$. Therefore, (S.5) can be written as

$$\widehat{\boldsymbol{\lambda}}_{\boldsymbol{\theta}} - \boldsymbol{\lambda}^{\star}_{\boldsymbol{\theta}} = -\left[\mathbb{E}\left\{\frac{\partial}{\partial \boldsymbol{\lambda}}\boldsymbol{U}_{\mathrm{B},N}(\boldsymbol{\lambda}^{\star}_{\boldsymbol{\theta}})\right\}\right]^{-1}\boldsymbol{U}_{\mathrm{B},N}(\boldsymbol{\lambda}^{\star}_{\boldsymbol{\theta}}) + o_p(N^{-1/2}),$$

where

$$\frac{\partial \boldsymbol{U}_{\mathrm{B},N}(\boldsymbol{\theta}, \boldsymbol{\lambda}_{\theta})}{\partial \boldsymbol{\lambda}} = \frac{1}{N}\sum_{i=1}^{N}\delta_i\{\omega^{\star}(\boldsymbol{x}_i;\boldsymbol{\theta},\boldsymbol{\lambda}_{\theta}) - 1\}\boldsymbol{z}_i(\boldsymbol{\theta})\boldsymbol{z}_i^{\mathrm{T}}(\boldsymbol{\theta}).$$

By Taylor expansion and similar technique we used above, and by mean value theorem, there exists $\bar{\boldsymbol{\lambda}}_{\boldsymbol{\theta}}$ between $\boldsymbol{\lambda}^\star_{\boldsymbol{\theta}}$ and $\widehat{\boldsymbol{\lambda}}_{\boldsymbol{\theta}}$, we have

$$\begin{aligned}
\boldsymbol{U}_{\text{SIPW},N}(\boldsymbol{\theta}) &= \frac{1}{N}\sum_{i=1}^{N}\delta_i\omega^\star(\boldsymbol{x}_i;\boldsymbol{\theta},\widehat{\boldsymbol{\lambda}}_{\boldsymbol{\theta}})\boldsymbol{U}(\boldsymbol{\theta},\boldsymbol{x}_i,\boldsymbol{y}_i) \\
&= \frac{1}{N}\sum_{i=1}^{N}\delta_i\omega^\star(\boldsymbol{x}_i;\boldsymbol{\theta},\boldsymbol{\lambda}^\star_{\boldsymbol{\theta}})\boldsymbol{U}(\boldsymbol{\theta},\boldsymbol{x}_i,\boldsymbol{y}_i) \\
&\qquad + \frac{1}{N}\sum_{i=1}^{N}\delta_i\boldsymbol{U}(\boldsymbol{\theta},\boldsymbol{x}_i,\boldsymbol{y}_i)\left\{\frac{\partial}{\partial\boldsymbol{\lambda}}\omega^\star(\boldsymbol{x}_i;\boldsymbol{\theta},\bar{\boldsymbol{\lambda}}_{\boldsymbol{\theta}})\right\}\left(\widehat{\boldsymbol{\lambda}}_{\boldsymbol{\theta}}-\boldsymbol{\lambda}^\star_{\boldsymbol{\theta}}\right) \\
&= \frac{1}{N}\sum_{i=1}^{N}\left[\boldsymbol{\beta}^\star\boldsymbol{z}_i(\boldsymbol{\theta})+\delta_i\omega^\star(\boldsymbol{x}_i;\boldsymbol{\theta},\boldsymbol{\lambda}^\star_{\boldsymbol{\theta}})\{\boldsymbol{U}(\boldsymbol{\theta};\boldsymbol{x}_i,y_i)-\boldsymbol{\beta}^\star\boldsymbol{z}_i(\boldsymbol{\theta})\}\right]+o_p(N^{-1/2}),
\end{aligned}$$

which completes the proof of Theorem 1.

□

S2.1 Verification of $\boldsymbol{\beta}$

It can be verified that

$$\frac{\partial\boldsymbol{U}_{\text{B},N}(\boldsymbol{\theta},\boldsymbol{\lambda}_{\boldsymbol{\theta}})}{\partial\boldsymbol{\lambda}} = \frac{N_0}{N_1}\frac{1}{N}\sum_{i=1}^{N}\delta_i\{\omega^\star(\boldsymbol{x}_i;\boldsymbol{\theta},\boldsymbol{\lambda}_\theta)-1\}\boldsymbol{z}_i(\boldsymbol{\theta})\boldsymbol{z}_i^{\text{T}}(\boldsymbol{\theta}), \tag{S.6}$$

$$\boldsymbol{U}_{\text{B},N}(\boldsymbol{\theta},\boldsymbol{\lambda}) = \frac{1}{N}\sum_{i=1}^{N}\{\delta_i\omega^\star(\boldsymbol{x}_i;\boldsymbol{\theta},\boldsymbol{\lambda})-1\}\boldsymbol{z}_i(\boldsymbol{\theta}), \tag{S.7}$$

$$\frac{\partial}{\partial\boldsymbol{\lambda}}\omega^\star(\boldsymbol{x}_i;\boldsymbol{\theta},\boldsymbol{\lambda}_{\boldsymbol{\theta}}) = \frac{N_0}{N_1}\{\omega^\star(\boldsymbol{x}_i;\boldsymbol{\theta},\boldsymbol{\lambda}_\theta)-1\}\boldsymbol{z}_i(\boldsymbol{\theta}). \tag{S.8}$$

Combine (S.6), (S.7) and (S.8), we have

$$
\begin{aligned}
&\frac{1}{N}\sum_{i=1}^{N}\delta_i \boldsymbol{U}(\boldsymbol{\theta},\boldsymbol{x}_i,\boldsymbol{y}_i)\left[\left\{\frac{\partial}{\partial\boldsymbol{\lambda}}\omega^\star(\boldsymbol{x}_i;\boldsymbol{\theta},\boldsymbol{\lambda}_{\boldsymbol{\theta}})\right\}\left(\widehat{\boldsymbol{\lambda}}_{\boldsymbol{\theta}}-\boldsymbol{\lambda}^\star_{\boldsymbol{\theta}}\right)\right]\\
&=\left[\frac{1}{N}\sum_{i=1}^{N}\delta_i\{\omega^\star(\boldsymbol{x}_i;\boldsymbol{\theta},\boldsymbol{\lambda})-1\}\boldsymbol{U}(\boldsymbol{\theta},\boldsymbol{x}_i,\boldsymbol{y}_i)\boldsymbol{z}_i^{\mathrm{T}}(\boldsymbol{\theta})\right]\\
&\quad\times\left[\frac{1}{N}\sum_{i=1}^{N}\delta_i\{\omega^\star(\boldsymbol{x}_i;\boldsymbol{\theta},\boldsymbol{\lambda})-1\}\boldsymbol{z}_i(\boldsymbol{\theta})\boldsymbol{z}_i^{\mathrm{T}}(\boldsymbol{\theta})\right]^{-1}\\
&\quad\times\frac{1}{N}\sum_{i=1}^{N}\{1-\delta_i\omega^\star(\boldsymbol{x}_i;\boldsymbol{\theta},\boldsymbol{\lambda})\}\boldsymbol{z}_i(\boldsymbol{\theta}).
\end{aligned}
$$

Note that the first part serves as the estimate of $\boldsymbol{\beta}^\star$.

S3 Proof of Theorem 2

As $\boldsymbol{U}_{\mathrm{B},N}=\mathbf{0}$, we have

$$
\frac{\partial}{\partial\boldsymbol{\theta}}\left\{\boldsymbol{U}_{\mathrm{B},N}(\boldsymbol{\theta},\boldsymbol{\lambda})\right\}=\frac{\partial}{\partial\boldsymbol{\theta}}\left\{\frac{1}{N}\sum_{i=1}^{N}\delta_i\omega^\star(\boldsymbol{x}_i;\boldsymbol{\theta},\boldsymbol{\lambda})\boldsymbol{z}_i(\boldsymbol{\theta})-\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{z}_i(\boldsymbol{\theta})\right\}=\mathbf{0}. \tag{S.9}
$$

It turns out that

$$
\begin{aligned}
&\frac{\partial}{\partial\boldsymbol{\theta}}\left\{\frac{1}{N}\sum_{i=1}^{N}\delta_i\omega^\star(\boldsymbol{x}_i;\boldsymbol{\theta},\boldsymbol{\lambda})\boldsymbol{z}_i(\boldsymbol{\theta})\right\}\\
&=\frac{1}{N}\sum_{i=1}^{N}\delta_i\left[\{\omega^\star(\boldsymbol{x}_i;\boldsymbol{\theta},\boldsymbol{\lambda})-1\}\boldsymbol{z}_i(\boldsymbol{\theta})\left\{\boldsymbol{z}_i^{\mathrm{T}}(\boldsymbol{\theta})\frac{\partial\boldsymbol{\lambda}}{\partial\boldsymbol{\theta}}+\boldsymbol{\lambda}^{\mathrm{T}}\frac{\partial\boldsymbol{z}_i(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right\}+\omega^\star(\boldsymbol{x}_i;\boldsymbol{\theta},\boldsymbol{\lambda})\frac{\partial\boldsymbol{z}_i(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right].
\end{aligned} \tag{S.10}
$$

Combine (S.9) and (S.10), it yields that

$$\frac{\partial \boldsymbol{\lambda}}{\partial \boldsymbol{\theta}} = \left\{ \frac{1}{N}\sum_{i=1}^{N} \delta_i \{\omega^\star(\boldsymbol{x}_i;\boldsymbol{\theta},\boldsymbol{\lambda}) - 1\} \boldsymbol{z}_i(\boldsymbol{\theta})\boldsymbol{z}_i^{\mathrm{T}}(\boldsymbol{\theta}) \right\}^{-1} \times \left[\frac{1}{N}\sum_{i=1}^{N} \left\{1 - \delta_i\omega^\star(\boldsymbol{x}_i;\boldsymbol{\theta},\boldsymbol{\lambda})\right\} \frac{\partial \boldsymbol{z}_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} - \frac{1}{N}\sum_{i=1}^{N} \delta_i\{\omega^\star(\boldsymbol{x}_i;\boldsymbol{\theta},\boldsymbol{\lambda}) - 1\}\boldsymbol{z}_i(\boldsymbol{\theta})\boldsymbol{\lambda}^{\mathrm{T}}\frac{\partial \boldsymbol{z}_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right]. \tag{S.11}$$

By Taylor expansion and mean value theorem, there exists $\widetilde{\boldsymbol{\theta}}$ between $\boldsymbol{\theta}^\star$ and $\widehat{\boldsymbol{\theta}}_{\mathrm{SIPW}}$, such that

$$\begin{aligned}\widehat{\boldsymbol{\theta}}_{\mathrm{SIPW}} - \boldsymbol{\theta}^\star &= -\left[\frac{1}{N}\sum_{i=1}^{N} \frac{\partial}{\partial \boldsymbol{\theta}} \delta_i\omega^\star(\boldsymbol{x}_i;\widetilde{\boldsymbol{\theta}},\widehat{\boldsymbol{\lambda}}_{\widetilde{\boldsymbol{\theta}}})\boldsymbol{U}(\widetilde{\boldsymbol{\theta}};\boldsymbol{x}_i,\boldsymbol{y}_i) \right]^{-1} \\ &\qquad \times \left\{ \frac{1}{\sqrt{N}}\sum_{i=1}^{N} \delta_i\omega^\star(\boldsymbol{x}_i;\boldsymbol{\theta}^\star,\widehat{\boldsymbol{\lambda}}_{\boldsymbol{\theta}^\star})\boldsymbol{U}(\boldsymbol{\theta}^\star,\boldsymbol{x}_i,\boldsymbol{y}_i) \right\} \\ &\overset{(i)}{=} -\mathbb{E}\left[\frac{\partial}{\partial \boldsymbol{\theta}} \delta\omega^\star(\boldsymbol{X};\boldsymbol{\theta}^\star,\boldsymbol{\lambda}^\star)\boldsymbol{U}(\boldsymbol{\theta}^\star;\boldsymbol{X},\mathbf{Y}) \right]^{-1} \times \left\{ \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{d}(\boldsymbol{x}_i, y_i, \delta_i;\boldsymbol{\theta}^\star,\boldsymbol{\lambda}^\star) \right\} + o_p(N^{-1/2}),\end{aligned}$$

where the equality (i) is by Theorem 1. Furthermore, note that

$$\begin{aligned}&\frac{\partial}{\partial \boldsymbol{\theta}}\left\{\delta\omega^\star(\boldsymbol{X};\boldsymbol{\theta}^\star,\boldsymbol{\lambda}^\star)\boldsymbol{U}(\boldsymbol{\theta}^\star;\boldsymbol{X},\mathbf{Y})\right\} \\ &= \delta\frac{\partial}{\partial \boldsymbol{\theta}}\left\{\omega^\star(\boldsymbol{X};\boldsymbol{\theta}^\star,\boldsymbol{\lambda}^\star)\right\}\boldsymbol{U}(\boldsymbol{\theta}^\star;\boldsymbol{X},\mathbf{Y}) + \delta\omega^\star(\boldsymbol{X};\boldsymbol{\theta}^\star,\boldsymbol{\lambda}^\star)\frac{\partial}{\partial \boldsymbol{\theta}}\left\{\boldsymbol{U}(\boldsymbol{\theta}^\star;\boldsymbol{X},\mathbf{Y})\right\},\end{aligned}$$

and

$$\frac{\partial}{\partial \boldsymbol{\theta}}\left\{\omega^\star(\boldsymbol{X};\boldsymbol{\theta},\boldsymbol{\lambda})\right\} = \{\omega^\star(\boldsymbol{X};\boldsymbol{\theta},\boldsymbol{\lambda}) - 1\}\left\{\boldsymbol{Z}^{\mathrm{T}}(\boldsymbol{\theta})\frac{\partial \boldsymbol{\lambda}}{\partial \boldsymbol{\theta}} + \boldsymbol{\lambda}^{\mathrm{T}}\frac{\partial \boldsymbol{Z}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right\},$$

together with (S.11) leads to the form of $\boldsymbol{\tau}$ in Theorem 2, which complete the proof.

S4 Proof of Corollary 1

Note that $\boldsymbol{\beta}^\star \boldsymbol{Z}(\boldsymbol{\theta}^\star) = \sum_{k=0}^{L} b_k(\boldsymbol{X};\boldsymbol{\theta}^\star)\boldsymbol{\beta}_k^\star$. If the outcome model is correctly specified, rearranging the terms in the estimating equation for $\boldsymbol{\beta}^\star$, it turns out that

$$\sum_{i=1}^{N} \delta_i\{\omega^\star(\boldsymbol{x}_i;\boldsymbol{\theta},\boldsymbol{\lambda}_{\boldsymbol{\theta}}^\star)-1\}\boldsymbol{\beta}\boldsymbol{z}_i(\boldsymbol{\theta})\boldsymbol{z}_i^{\mathrm{T}}(\boldsymbol{\theta}) = \sum_{i=1}^{N} \delta_i\{\omega^\star(\boldsymbol{x}_i;\boldsymbol{\theta},\boldsymbol{\lambda}_{\boldsymbol{\theta}}^\star)-1\}\boldsymbol{U}(\boldsymbol{\theta};\boldsymbol{x}_i,y_i)\boldsymbol{z}_i^{\mathrm{T}}(\boldsymbol{\theta})$$

$$\Leftrightarrow \left[\sum_{i=1}^{N} \delta_i\{\omega^\star(\boldsymbol{x}_i;\boldsymbol{\theta},\boldsymbol{\lambda}_{\boldsymbol{\theta}}^\star)-1\}\boldsymbol{z}_i(\boldsymbol{\theta})\boldsymbol{z}_i^{\mathrm{T}}(\boldsymbol{\theta})\right]\boldsymbol{\beta}^{\mathrm{T}} = \sum_{i=1}^{N} \delta_i\{\omega^\star(\boldsymbol{x}_i;\boldsymbol{\theta},\boldsymbol{\lambda}_{\boldsymbol{\theta}}^\star)-1\}\boldsymbol{z}_i(\boldsymbol{\theta})\boldsymbol{U}^{\mathrm{T}}(\boldsymbol{\theta};\boldsymbol{x}_i,y_i). \tag{S.12}$$

The equation (S.12) is the normal equation of a weighted least square estimation of $\boldsymbol{\beta}$. Thus, as long as

$$\mathbb{E}\{\boldsymbol{U}(\boldsymbol{\theta};\boldsymbol{x},Y) \mid \boldsymbol{x}\} \in \mathrm{span}\{1, b_1(\boldsymbol{x};\boldsymbol{\theta}),\ldots,b_L(\boldsymbol{x};\boldsymbol{\theta})\}, \tag{S.13}$$

by the uniqueness solution of weighted least square estimation, we know that $\mathbb{E}\{\boldsymbol{U}(\boldsymbol{\theta};\boldsymbol{X},Y) \mid \boldsymbol{X}=\boldsymbol{x}\} = \boldsymbol{\beta}^\star\boldsymbol{z}(\boldsymbol{\theta})$. It turns out that

$$\begin{aligned}
&\mathbb{V}\left\{\boldsymbol{d}(\boldsymbol{X},Y;\delta;\boldsymbol{\theta}^\star,\boldsymbol{\lambda}^\star)\right\}\\
&= \mathbb{V}\left[\sum_{k=0}^{L} b_k(\boldsymbol{X};\boldsymbol{\theta}^\star)\boldsymbol{\beta}_k^\star + \delta\omega^\star(\boldsymbol{X};\boldsymbol{\theta}^\star,\boldsymbol{\lambda}^\star)\left\{\boldsymbol{U}(\boldsymbol{\theta}^\star;\boldsymbol{X},Y) - \sum_{k=0}^{L} b_k(\boldsymbol{X};\boldsymbol{\theta}^\star)\boldsymbol{\beta}_k^\star\right\}\right]\\
&= \mathbb{V}\left(\mathbb{E}_Y\left[\sum_{k=0}^{L} b_k(\boldsymbol{X};\boldsymbol{\theta}^\star)\boldsymbol{\beta}_k^\star + \delta\omega^\star(\boldsymbol{X};\boldsymbol{\theta}^\star,\boldsymbol{\lambda}^\star)\left\{\boldsymbol{U}(\boldsymbol{\theta}^\star;\boldsymbol{X},Y) - \sum_{k=0}^{L} b_k(\boldsymbol{X};\boldsymbol{\theta}^\star)\boldsymbol{\beta}_k^\star\right\} \mid \boldsymbol{X},\delta\right]\right)+\\
&\quad + \mathbb{E}\left(\mathbb{V}_Y\left[\sum_{k=0}^{L} b_k(\boldsymbol{X};\boldsymbol{\theta}^\star)\boldsymbol{\beta}_k^\star + \delta\omega^\star(\boldsymbol{X};\boldsymbol{\theta}^\star,\boldsymbol{\lambda}^\star)\left\{\boldsymbol{U}(\boldsymbol{\theta}^\star;\boldsymbol{X},Y) - \sum_{k=0}^{L} b_k(\boldsymbol{X};\boldsymbol{\theta}^\star)\boldsymbol{\beta}_k^\star\right\} \mid \boldsymbol{X},\delta\right]\right)\\
&= \mathbb{V}\left\{\boldsymbol{\beta}^\star\boldsymbol{Z}(\boldsymbol{\theta}^\star)\right\} + \mathbb{E}\left[\delta\{\omega^\star(\boldsymbol{X};\boldsymbol{\theta}^\star,\boldsymbol{\lambda}^\star)\}^2\mathbb{V}\{\boldsymbol{U}(\boldsymbol{\theta}^\star;\boldsymbol{X},Y) \mid \boldsymbol{X}\}\right].
\end{aligned}$$

By the argument before Corollary 1, we know that $\boldsymbol{\theta}_0 = \boldsymbol{\theta}^\star$, which completes the proof.

S5 Additional Regularity Conditions for Corollary 2

[**S1**] For any constant M, there exists non-singular matrix $\mathbf{D}$ such that

$$\sup_{|\boldsymbol{\alpha}-\boldsymbol{\alpha}_0|\leqslant MN^{-1/2}} \left| N^{-1/2}\boldsymbol{U}(\boldsymbol{\alpha}) - N^{-1/2}\boldsymbol{U}(\boldsymbol{\alpha}_0) - N^{1/2}\mathbf{D}(\boldsymbol{\alpha}-\boldsymbol{\alpha}_0) \right| = o_p(1).$$

Additionally, $n^{-1/2}\boldsymbol{U}(\boldsymbol{\alpha}_0) \xrightarrow{\mathcal{L}} N(\mathbf{0}, \mathbf{F})$ for a positive definite matrix $\mathbf{F}$.

[**S2**] The tuning parameter λ in (47) satisfies

$$\lambda \to 0, \sqrt{N}\lambda \to \infty$$

as $n \to \infty$.

S6 Basis Function Derivation for Simulation Study Two

Recall that we have

$$f_1(y \mid x; \boldsymbol{\beta}) = \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\left\{ -\frac{(y-\beta_0-\beta_1 x)^2}{2\sigma_0^2} \right\},$$
$$f_2(z, x; \alpha) = f(x) f(z \mid x; \alpha) = \frac{1}{\sqrt{2\pi\sigma_e^2}} \exp\left\{ -\frac{(z-\alpha x)^2}{2\sigma_e^2} \right\}.$$

Further, we have

$$\mathbf{S}_1(\boldsymbol{\beta}; x, y) = (y-\beta_0-\beta_1 x)(1, x)^{\mathrm{T}},$$
$$S_2(\alpha; x, z) = x(z-\alpha x).$$

Next we calculate the quantities involved in basis function.

$$
\begin{aligned}
&\int f_1(y \mid x; \boldsymbol{\beta}) f_2(x, z \mid \alpha) dx \\
&= \int \frac{1}{2\pi\sigma_0\sigma_e} \exp\left[-\frac{1}{2}\left\{\frac{(y-\beta_0-\beta_1 x)^2}{\sigma_0^2} + \frac{(z-\alpha x)^2}{\sigma_e^2}\right\}\right] dx \\
&= \int \frac{1}{2\pi\sigma_0\sigma_e} \exp\left[-\frac{(\beta_1^2\sigma_e^2+\alpha^2\sigma_0^2)x^2 - 2\{\beta_1(y-\beta_0)\sigma_e^2+\alpha z\sigma_0^2\}x + (y-\beta_0)^2\sigma_e^2 + z^2\sigma_0^2}{2\sigma_0^2\sigma_e^2}\right] dx \\
&= \int \frac{1}{2\pi\sigma_0\sigma_e} \exp\left[-\frac{(\beta_1^2\sigma_e^2+\alpha^2\sigma_0^2)x^2 - 2\{\beta_1(y-\beta_0)\sigma_e^2+\alpha z\sigma_0^2\}x + \frac{\{\beta_1(y-\beta_0)\sigma_e^2+\alpha z\sigma_0^2\}^2}{\beta_1^2\sigma_e^2+\alpha^2\sigma_0^2}}{2\sigma_0^2\sigma_e^2}\right. \\
&\qquad\qquad \left. + \frac{\frac{\{\beta_1(y-\beta_0)\sigma_e^2+\alpha z\sigma_0^2\}^2}{\beta_1^2\sigma_e^2+\alpha^2\sigma_0^2} - \{(y-\beta_0)^2\sigma_e^2 + z^2\sigma_0^2\}}{2\sigma_0^2\sigma_e^2}\right] dx \\
&= \sqrt{2\pi\frac{\sigma_0^2\sigma_e^2}{(\beta_1^2\sigma_e^2+\alpha^2\sigma_0^2)}} \times \frac{1}{2\pi\sigma_0\sigma_e} \exp\left[\frac{\frac{\{\beta_1(y-\beta_0)\sigma_e^2+\alpha z\sigma_0^2\}^2}{\beta_1^2\sigma_e^2+\alpha^2\sigma_0^2} - \{(y-\beta_0)^2\sigma_e^2 + z^2\sigma_0^2\}}{2\sigma_0^2\sigma_e^2}\right] \\
&\qquad\qquad \times \underbrace{\int \frac{1}{\sqrt{2\pi\frac{\sigma_0^2\sigma_e^2}{(\beta_1^2\sigma_e^2+\alpha^2\sigma_0^2)}}} \exp\left\{-\frac{\left(x - \frac{\beta_1(y-\beta_0)\sigma_e^2+\alpha z\sigma_0^2}{\beta_1^2\sigma_e^2+\alpha^2\sigma_0^2}\right)^2}{2\frac{\sigma_0^2\sigma_e^2}{\beta_1^2\sigma_e^2+\alpha^2\sigma_0^2}}\right\} dx}_{T_0} \\
&= \frac{1}{\sqrt{2\pi(\beta_1^2\sigma_e^2+\alpha^2\sigma_0^2)}} \exp\left[\frac{\frac{\{\beta_1(y-\beta_0)\sigma_e^2+\alpha z\sigma_0^2\}^2}{\beta_1^2\sigma_e^2+\alpha^2\sigma_0^2} - \{(y-\beta_0)^2\sigma_e^2 + z^2\sigma_0^2\}}{2\sigma_0^2\sigma_e^2}\right] \\
&=: \tau(y, z). \qquad \text{(S.14)}
\end{aligned}
$$

The term T_0 in (S.14) motivates that

$$\begin{aligned}
&\int x f_1(y \mid x; \boldsymbol{\beta}) f_2(x, z \mid \alpha) dx \\
&= \tau(y, z) \frac{\beta_1 (y - \beta_0)\sigma_e^2 + \alpha z \sigma_0^2}{\beta_1^2 \sigma_e^2 + \alpha^2 \sigma_0^2},
\end{aligned}$$

and

$$\begin{aligned}
&\int x^2 f_1 x(y \mid x; \boldsymbol{\beta}) f_2(x, z \mid \alpha) dx \\
&= \tau(y, z) \left[\left\{ \frac{\beta_1 (y - \beta_0)\sigma_e^2 + \alpha z \sigma_0^2}{\beta_1^2 \sigma_e^2 + \alpha^2 \sigma_0^2} \right\}^2 + \frac{\sigma_0^2 \sigma_e^2}{\beta_1^2 \sigma_e^2 + \alpha^2 \sigma_0^2} \right].
\end{aligned}$$

As a result, we have

$$\begin{aligned}
\boldsymbol{b}_1(\theta; y, z) &= \mathbb{E}\{S_1(\boldsymbol{\beta}; X, Y \mid z, y)\} \\
&= \begin{pmatrix} y - \beta_0 - \beta_1 \frac{\beta_1 (y - \beta_0)\sigma_e^2 + \alpha z \sigma_0^2}{\beta_1^2 \sigma_e^2 + \alpha^2 \sigma_0^2} \\ (y - \beta_0) \frac{\beta_1 (y - \beta_0)\sigma_e^2 + \alpha z \sigma_0^2}{\beta_1^2 \sigma_e^2 + \alpha^2 \sigma_0^2} - \beta_1 \left[\left\{ \frac{\beta_1 (y - \beta_0)\sigma_e^2 + \alpha z \sigma_0^2}{\beta_1^2 \sigma_e^2 + \alpha^2 \sigma_0^2} \right\}^2 + \frac{\sigma_0^2 \sigma_e^2}{\beta_1^2 \sigma_e^2 + \alpha^2 \sigma_0^2} \right] \end{pmatrix},
\end{aligned}$$

and

$$\begin{aligned}
b_2(\alpha; x, z) &= \mathbb{E}\left\{ S_2(\alpha; X, Z) \mid y, z \right\} \\
&= z \frac{\beta_1 (y - \beta_0)\sigma_e^2 + \alpha z \sigma_0^2}{\beta_1^2 \sigma_e^2 + \alpha^2 \sigma_0^2} - \alpha \left[\left\{ \frac{\beta_1 (y - \beta_0)\sigma_e^2 + \alpha z \sigma_0^2}{\beta_1^2 \sigma_e^2 + \alpha^2 \sigma_0^2} \right\}^2 + \frac{\sigma_0^2 \sigma_e^2}{\beta_1^2 \sigma_e^2 + \alpha^2 \sigma_0^2} \right].
\end{aligned}$$