Supplement to "Semiparametric efficient estimators in heteroscedastic error models"

Mijeong Kim

Department of Statistics, Ewha Womans University, Seoul 03760, Republic of Korea

m.kim@ewha.ac.kr

and

Yanyuan Ma

Department of Statistics, Penn State University, University Park, PA 16802, USA

yzm63@psu.edu

September 27, 2017

S.1 Proof of Proposition 2

Similar to the proof of Proposition 1, before starting the formal proof, we note that the construction of t ensures the property

$$E(\epsilon t) = E(\epsilon^3) - E(\epsilon^3)E(\epsilon^2) - E(\epsilon) = E(\epsilon^3) - E(\epsilon^3) = 0,$$

which is useful in the later development.

We derive nuisance tangent space Λ and its orthogonal complement $\Lambda^{\perp}$ for this model and then we find the semiparametric efficient estimator by projecting score function onto $\Lambda^{\perp}$.

We will construct $\Lambda = \Lambda_x \oplus \Lambda_\epsilon$, where Λ_x is a subspace with functions of $\mathbf{X}$ and Λ_ϵ is a subspace with functions of ϵ.

Since the predicting variable $\mathbf{X}$ does not have any constraint and all functions are defined in Hilbert space, we have the following property for Λ_x.

$$\Lambda_x \quad = \quad \{\mathbf{a}(\mathbf{x}) : E\{\mathbf{a}(\mathbf{X})\} = \mathbf{0}\}.$$

We now investigate Λ_ϵ. The nuisance tangent space Λ and its orthogonal complement $\Lambda^{\perp}$ can be derived based on the relations

$$\int f_\epsilon(\epsilon) d\epsilon = 1, \quad \int \epsilon f_\epsilon(\epsilon) d\epsilon = 0, \quad \int \epsilon^2 f_\epsilon(\epsilon) d\epsilon = 1.$$

An arbitrary function $\mathbf{b}(\epsilon)$ in Λ_ϵ satisfies $E\{\mathbf{b}(\epsilon)\} = \mathbf{0}$ in Hilbert space. Following Tsiatis (2006, Section 4.5), the second constraint allows $E\{\epsilon\mathbf{b}(\epsilon)\} = \mathbf{0}$. Similarly, we obtain $E\{\epsilon^2\mathbf{b}(\epsilon)\} = \mathbf{0}$ from the third constraint. Thus,

$$\begin{aligned} \Lambda_\epsilon &= \{\mathbf{b}(\epsilon) : E\{\mathbf{b}(\epsilon)\} = E\{\epsilon\mathbf{b}(\epsilon)\} = E\{\epsilon^2\mathbf{b}(\epsilon)\} = \mathbf{0}\} \\ &= \{\mathbf{b}(\epsilon) : E\{\mathbf{b}(\epsilon)\} = E\{\epsilon\mathbf{b}(\epsilon)\} = E\{t\mathbf{b}(\epsilon)\} = \mathbf{0}\}, \end{aligned}$$

where $t = \epsilon^2 - E(\epsilon^3)\epsilon - 1$. In the above,

$$E\{t\mathbf{b}(\epsilon)\} = E\{\epsilon^2\mathbf{b}(\epsilon)\} - E(\epsilon^3)E\{\epsilon\mathbf{b}(\epsilon)\} - E\{\mathbf{b}(\epsilon)\} = E\{\epsilon^2\mathbf{b}(\epsilon)\} = \mathbf{0}.$$

Thus, we have

$$\Lambda = \Lambda_x \oplus \Lambda_\epsilon = \{\mathbf{a}(\mathbf{x}) + \mathbf{b}(\epsilon) : E\{\mathbf{a}(\mathbf{X})\} = \mathbf{0}, E\{\mathbf{b}(\epsilon)\} = \mathbf{0}, E\{\epsilon\mathbf{b}(\epsilon)\} = \mathbf{0}, E\{t\mathbf{b}(\epsilon)\} = \mathbf{0}\}.$$

Note that Λ_x and Λ_ϵ are orthogonal because we assume $\mathbf{x}$ and ϵ are independent.

We now prove that

$$\Lambda^\perp = \{\mathbf{g}(\mathbf{x}, \epsilon) : E\{\mathbf{g}(\mathbf{X}, \epsilon)|\mathbf{X}\} = \mathbf{0}, E\{\mathbf{g}(\mathbf{X}, \epsilon)|\epsilon\} = \mathbf{c}_1\epsilon + \mathbf{c}_2 t : \mathbf{c}_1, \mathbf{c}_2 \in \mathbb{R}^{k+l}\}.$$

Let $\mathbf{K} = \{\mathbf{g}(\mathbf{x}, \epsilon) : E\{\mathbf{g}(\mathbf{X}, \epsilon)|\mathbf{X}\} = \mathbf{0}, E\{\mathbf{g}(\mathbf{X}, \epsilon)|\epsilon\} = \mathbf{c}_1\epsilon + \mathbf{c}_2 t, \mathbf{c}_1, \mathbf{c}_2 \in \mathbb{R}^{k+l}\}$. We will show $\mathbf{K} \subset \Lambda^\perp$ and $\Lambda^\perp \subset \mathbf{K}$.

For any $\mathbf{a}(\mathbf{x}) + \mathbf{b}(\epsilon) \in \Lambda$ and $\mathbf{g}(\mathbf{x}, \epsilon) \in \mathbf{K}$, we have

$$\begin{aligned} E\{\mathbf{g}(\mathbf{X}, \epsilon)^{\mathrm{T}}\mathbf{a}(\mathbf{X})\} &= E\left[E\{\mathbf{g}(\mathbf{X}, \epsilon)|\mathbf{X}\}^{\mathrm{T}}\mathbf{a}(\mathbf{X})\right] = 0, \\ E\{\mathbf{g}(\mathbf{X}, \epsilon)^{\mathrm{T}}\mathbf{b}(\epsilon)\} &= E\left[E\{\mathbf{g}(\mathbf{X}, \epsilon)|\epsilon\}^{\mathrm{T}}\mathbf{b}(\epsilon)\right] = E\left\{(\mathbf{c}_1\epsilon + \mathbf{c}_2 t)^{\mathrm{T}}\mathbf{b}(\epsilon)\right\} = 0. \end{aligned}$$

Hence, we obtain

$$E[\mathbf{g}(\mathbf{X}, \epsilon)^{\mathrm{T}}\{\mathbf{a}(\mathbf{X}) + \mathbf{b}(\epsilon)\}] = 0.$$

Thus, $\mathbf{K} \subset \Lambda^\perp$.

Next, we need to show $\Lambda^\perp \subset \mathbf{K}$. We assume $\mathbf{h}(\mathbf{x}, \epsilon) \in \Lambda^\perp$. We can decompose $\mathbf{h}(\mathbf{x}, \epsilon)$ as $\mathbf{h}(\mathbf{X}, \epsilon) = \mathbf{h}_1(\mathbf{X}, \epsilon) + \mathbf{r}(\mathbf{X}, \epsilon)$. Here, $\mathbf{h}_1(\mathbf{X}, \epsilon)$ and $\mathbf{r}(\mathbf{X}, \epsilon)$ are given below.

$$\begin{aligned} \mathbf{h}_1(\mathbf{X}, \epsilon) &= \mathbf{h}(\mathbf{X}, \epsilon) - E\{\mathbf{h}(\mathbf{X}, \epsilon)|\mathbf{X}\} - E\{\mathbf{h}(\mathbf{X}, \epsilon)|\epsilon\} + E\{\epsilon\mathbf{h}(\mathbf{X}, \epsilon)\}\epsilon + \frac{E\{t\mathbf{h}(\mathbf{X}, \epsilon)\}}{E(t^2)}t, \\ \mathbf{r}(\mathbf{X}, \epsilon) &= E\{\mathbf{h}(\mathbf{X}, \epsilon)|\mathbf{X}\} + E\{\mathbf{h}(\mathbf{X}, \epsilon)|\epsilon\} - E\{\epsilon\mathbf{h}(\mathbf{X}, \epsilon)\}\epsilon - \frac{E\{t\mathbf{h}(\mathbf{X}, \epsilon)\}}{E(t^2)}t. \end{aligned}$$

Also, we can decompose $\mathbf{r}(\mathbf{x}, \epsilon)$ as $\mathbf{r}(\mathbf{x}, \epsilon) = \mathbf{r}_1(\mathbf{x}) + \mathbf{r}_2(\epsilon)$, where

$$\begin{aligned}
\mathbf{r}_1(\mathbf{X}) &= E\{\mathbf{h}(\mathbf{X}, \epsilon)|\mathbf{X}\}, \\
\mathbf{r}_2(\epsilon) &= E\{\mathbf{h}(\mathbf{X}, \epsilon)|\epsilon\} - E\{\epsilon\mathbf{h}(\mathbf{X}, \epsilon)\}\epsilon - \frac{E\{t\mathbf{h}(\mathbf{X}, \epsilon)\}}{E(t^2)}t.
\end{aligned}$$

Then, the followings satisfy.

$$\begin{aligned}
E\{\mathbf{r}_1(\mathbf{X})\} &= E\left[E\{\mathbf{h}(\mathbf{X}, \epsilon)|\mathbf{X}\}\right] = E\{\mathbf{h}(\mathbf{X}, \epsilon)\} = \mathbf{0}, \\
E\{\mathbf{r}_2(\epsilon)\} &= E\left[E\{\mathbf{h}(\mathbf{X}, \epsilon)|\epsilon\} - E\{\epsilon\mathbf{h}(\mathbf{X}, \epsilon)\}\epsilon - \frac{E\{t\mathbf{h}(\mathbf{X}, \epsilon)\}}{E(t^2)}t\right] = \mathbf{0} \\
&= E\{\mathbf{h}(\mathbf{x}, \epsilon)\} + E\{\epsilon\mathbf{h}(\mathbf{X}, \epsilon)\}E(\epsilon) - \frac{E\{t\mathbf{h}(\mathbf{X}, \epsilon)\}}{E(t^2)}E(t) = \mathbf{0}, \\
E\{\epsilon\mathbf{r}_2(\epsilon)\} &= E\left\{E\{\mathbf{h}(\mathbf{X}, \epsilon)|\epsilon\}\epsilon - E\{\epsilon\mathbf{h}(\mathbf{X}, \epsilon)\}\epsilon^2 - \frac{E\{t\mathbf{h}(\mathbf{X}, \epsilon)\}}{E(t^2)}\epsilon t\right\} \\
&= E(\epsilon\mathbf{h}) - E(\epsilon\mathbf{h}) \cdot 1 - \frac{E\{t\mathbf{h}(\mathbf{X}, \epsilon)\}}{E(t^2)} \cdot 0 = \mathbf{0}, \\
E\{t\mathbf{r}_2(\epsilon)\} &= E\left\{E\{t\mathbf{h}(\mathbf{X}, \epsilon)|\epsilon\} - E\{\epsilon\mathbf{h}(\mathbf{X}, \epsilon)\}\epsilon t - \frac{E\{t\mathbf{h}(\mathbf{X}, \epsilon)\}}{E(t^2)}t^2\right\} \\
&= E\{t\mathbf{h}(\mathbf{X}, \epsilon)\} - E(\epsilon\mathbf{h}) \cdot 0 - \frac{E\{t\mathbf{h}(\mathbf{X}, \epsilon)\}}{E(t^2)}E(t^2) = \mathbf{0}.
\end{aligned}$$

Thus, $\mathbf{r}_1(\mathbf{x}) \in \Lambda_x$ and $\mathbf{r}_2(\mathbf{x}) \in \Lambda_\epsilon$. It implies that $\mathbf{r}(\mathbf{X}, \epsilon) \in \Lambda$. For $\mathbf{h}_1(\mathbf{x}, \epsilon) = \mathbf{h}(\mathbf{x}, \epsilon) - \mathbf{r}(\mathbf{x}, \epsilon)$, we can easily calculate $E\{\mathbf{h}_1(\mathbf{X}, \epsilon)|\mathbf{X}\}$ and $E\{\mathbf{h}_1(\mathbf{X}, \epsilon)|\epsilon\}$.

$$\begin{aligned}
E\{\mathbf{h}_1(\mathbf{X}, \epsilon)|\mathbf{X}\} &= E\{\mathbf{h}(\mathbf{X}, \epsilon)|\mathbf{X}\} - E\{\mathbf{h}(\mathbf{X}, \epsilon)|\mathbf{X}\} - E[E\{\mathbf{h}(\mathbf{X}, \epsilon)|\epsilon\}] + E\{\epsilon\mathbf{h}(\mathbf{X}, \epsilon)\}E(\epsilon) \\
&\quad + \frac{E\{t\mathbf{h}(\mathbf{X}, \epsilon)\}}{E(t^2)}E(t) = \mathbf{0}, \\
E\{\mathbf{h}_1(\mathbf{X}, \epsilon)|\epsilon\} &= E\{\mathbf{h}(\mathbf{X}, \epsilon)|\epsilon\} - E[E\{\mathbf{h}(\mathbf{X}, \epsilon)|\mathbf{X}\}] - E\{\mathbf{h}(\mathbf{X}, \epsilon)|\epsilon\} + E\{\epsilon\mathbf{h}(\mathbf{X}, \epsilon)\}\epsilon \\
&\quad + \frac{E\{t\mathbf{h}(\mathbf{X}, \epsilon)\}}{E(t^2)}t \\
&= E\{\epsilon\mathbf{h}(\mathbf{X}, \epsilon)\}\epsilon + \frac{E\{t\mathbf{h}(\mathbf{X}, \epsilon)\}}{E(t^2)}t.
\end{aligned}$$

From the above results, we obtain $\mathbf{h}_1(\mathbf{x}, \epsilon) \in \mathbf{K} \subset \Lambda^\perp$. Since $\mathbf{h}(\mathbf{x}, \epsilon) \in \Lambda^\perp$, $\mathbf{r}(\mathbf{x}, \epsilon) = \mathbf{h}(\mathbf{X}, \epsilon) - \mathbf{h}_1(\mathbf{x}, \epsilon) \in \Lambda^\perp$. The fact $\mathbf{r}(\mathbf{x}, \epsilon) \in \Lambda^\perp$ and $\mathbf{r}(\mathbf{x}, \epsilon) \in \Lambda$ implies $\mathbf{r}(\mathbf{x}, \epsilon) = \mathbf{0}$. We have $\mathbf{h}(\mathbf{x}, \epsilon) = \mathbf{h}_1(\mathbf{x}, \epsilon) \in \mathbf{K}$ for arbitrary $\mathbf{h}(\mathbf{x}, \epsilon) \in \Lambda^\perp$. Thus $\Lambda^\perp \subset \mathbf{K}$. □

S.2 Proof of Theorem 2

The joint probability distribution of $\mathbf{X}$ and Y can be written as

$$f_{\mathbf{X},Y}(\mathbf{x},y) = f_{\mathbf{X}}(\mathbf{x}) f_\epsilon \left\{ \frac{y - m(\mathbf{x},\boldsymbol{\alpha})}{e^{\sigma(\mathbf{x},\boldsymbol{\beta})}} \right\} \frac{1}{e^{\sigma(\mathbf{x},\boldsymbol{\beta})}} = f_{\mathbf{X}}(\mathbf{x}) f_\epsilon(\epsilon) e^{-\sigma(\mathbf{X},\boldsymbol{\beta})}.$$

Score functions of the parameters $\boldsymbol{\theta} = (\boldsymbol{\alpha}^{\mathrm{T}}, \boldsymbol{\beta}^{\mathrm{T}})^{\mathrm{T}}$ are given by

$$\begin{aligned}
\boldsymbol{S}_{\boldsymbol{\alpha}} &= \frac{\partial \log f_{\mathbf{X},Y}(\mathbf{x},y)}{\partial \boldsymbol{\alpha}} = -\frac{f'_\epsilon(\epsilon)}{f_\epsilon(\epsilon)} e^{-\sigma(\mathbf{X},\boldsymbol{\beta})} \mathbf{m}'_{\boldsymbol{\alpha}}(\mathbf{X},\boldsymbol{\alpha}), \\
\boldsymbol{S}_{\boldsymbol{\beta}} &= \frac{\partial \log f_{\mathbf{X},Y}(\mathbf{x},y)}{\partial \boldsymbol{\beta}} = -\frac{f'_\epsilon(\epsilon)}{f_\epsilon(\epsilon)} \epsilon \boldsymbol{\sigma}'_{\boldsymbol{\beta}}(\mathbf{x},\boldsymbol{\beta}) - \boldsymbol{\sigma}'_{\boldsymbol{\beta}}(\mathbf{x},\boldsymbol{\beta}).
\end{aligned}$$

We claim $\boldsymbol{S}_{\mathrm{eff}} = (\boldsymbol{S}^{\mathrm{T}}_{\mathrm{eff},\boldsymbol{\alpha}}, \boldsymbol{S}^{\mathrm{T}}_{\mathrm{eff},\boldsymbol{\beta}})^{\mathrm{T}}$, where

$$\begin{aligned}
\boldsymbol{S}_{\mathrm{eff},\boldsymbol{\alpha}} &= -\frac{f'_\epsilon(\epsilon)}{f_\epsilon(\epsilon)} \left[e^{-\sigma(\mathbf{X},\boldsymbol{\beta})} \mathbf{m}'_{\boldsymbol{\alpha}}(\mathbf{X},\boldsymbol{\alpha}) - E\{e^{-\sigma(\mathbf{X},\boldsymbol{\beta})} \mathbf{m}'_{\boldsymbol{\alpha}}(\mathbf{X},\boldsymbol{\alpha})\} \right] \\
&\quad + \left\{ \epsilon - \frac{E(\epsilon^3)}{E(t^2)} t \right\} E\{e^{-\boldsymbol{\alpha}(\mathbf{X},\boldsymbol{\beta})} \mathbf{m}'_{\boldsymbol{\alpha}}(\mathbf{X},\boldsymbol{\alpha})\}, \\
\boldsymbol{S}_{\mathrm{eff},\boldsymbol{\beta}} &= -\left\{ \frac{f'_\epsilon(\epsilon)}{f_\epsilon(\epsilon)} \epsilon + 1 \right\} \left[\boldsymbol{\sigma}'_{\boldsymbol{\beta}}(\mathbf{X},\boldsymbol{\beta}) - E\{\boldsymbol{\sigma}'_{\boldsymbol{\beta}}(\mathbf{X},\boldsymbol{\beta})\} \right] + \frac{2t}{E(t^2)} E\{\boldsymbol{\sigma}'_{\boldsymbol{\beta}}(\mathbf{X},\boldsymbol{\beta})\}.
\end{aligned}$$

To prove the above claim, we first verified that $\boldsymbol{S}_{\boldsymbol{\theta}} - \boldsymbol{S}_{\mathrm{eff}} \in \Lambda$.

First, $\boldsymbol{S}_{\boldsymbol{\alpha}} - \boldsymbol{S}_{\mathrm{eff},\boldsymbol{\alpha}} = E\{e^{-\boldsymbol{\alpha}(\mathbf{X},\boldsymbol{\beta})} \mathbf{m}'_{\boldsymbol{\alpha}}(\mathbf{X},\boldsymbol{\alpha})\} \left\{ -\frac{f'_\epsilon(\epsilon)}{f_\epsilon(\epsilon)} - \epsilon + \frac{E(\epsilon^3)}{E(t^2)} t \right\}$ is a pure function of ϵ.

$$\begin{aligned}
E(\boldsymbol{S}_{\boldsymbol{\alpha}} - \boldsymbol{S}_{\mathrm{eff},\boldsymbol{\alpha}}) &= E\{e^{-\boldsymbol{\alpha}(\mathbf{X},\boldsymbol{\beta})} \mathbf{m}'_{\boldsymbol{\alpha}}(\mathbf{X},\boldsymbol{\alpha})\} E\left\{ -\frac{f'_\epsilon(\epsilon)}{f_\epsilon(\epsilon)} - \epsilon + \frac{E(\epsilon^3)}{E(t^2)} t \right\} = \mathbf{0}, \\
E\{\epsilon(\boldsymbol{S}_{\boldsymbol{\alpha}} - \boldsymbol{S}_{\mathrm{eff},\boldsymbol{\alpha}})\} &= E\{e^{-\boldsymbol{\alpha}(\mathbf{X},\boldsymbol{\beta})} \mathbf{m}'_{\boldsymbol{\alpha}}(\mathbf{X},\boldsymbol{\alpha})\} E\left\{ -\frac{f'_\epsilon(\epsilon)}{f_\epsilon(\epsilon)} \epsilon - \epsilon^2 + \frac{E(\epsilon^3)}{E(t^2)} \epsilon t \right\} = \mathbf{0}, \\
E\{t(\boldsymbol{S}_{\boldsymbol{\alpha}} - \boldsymbol{S}_{\mathrm{eff},\boldsymbol{\alpha}})\} &= E\{e^{-\boldsymbol{\alpha}(\mathbf{X},\boldsymbol{\beta})} \mathbf{m}'_{\boldsymbol{\alpha}}(\mathbf{X},\boldsymbol{\alpha})\} E\left\{ -\frac{f'_\epsilon(\epsilon)}{f_\epsilon(\epsilon)} t - \epsilon t + \frac{E(\epsilon^3)}{E(t^2)} t^2 \right\} = \mathbf{0}.
\end{aligned}$$

Additionally, $\boldsymbol{S}_{\boldsymbol{\beta}} - \boldsymbol{S}_{\mathrm{eff},\boldsymbol{\beta}} = E\{\boldsymbol{\sigma}'_{\boldsymbol{\beta}}(\mathbf{X},\boldsymbol{\beta})\} \left\{ -\frac{f'_\epsilon(\epsilon)}{f_\epsilon(\epsilon)} \epsilon - 1 - \frac{2t}{E(t^2)} \right\}$ is also a pure function of ϵ.

$$\begin{aligned}
E(\boldsymbol{S}_{\boldsymbol{\beta}} - \boldsymbol{S}_{\mathrm{eff},\boldsymbol{\beta}}) &= E\{\boldsymbol{\sigma}'_{\boldsymbol{\beta}}(\mathbf{X},\boldsymbol{\beta})\} E\left\{ -\frac{f'_\epsilon(\epsilon)}{f_\epsilon(\epsilon)} \epsilon - 1 - \frac{2t}{E(t^2)} \right\} = \mathbf{0}, \\
E\{\epsilon(\boldsymbol{S}_{\boldsymbol{\beta}} - \boldsymbol{S}_{\mathrm{eff},\boldsymbol{\beta}})\} &= E\{\boldsymbol{\sigma}'_{\boldsymbol{\beta}}(\mathbf{X},\boldsymbol{\beta})\} E\left\{ -\frac{f'_\epsilon(\epsilon)}{f_\epsilon(\epsilon)} \epsilon^2 - \epsilon - \frac{2\epsilon t}{E(t^2)} \right\} = \mathbf{0}, \\
E\{t(\boldsymbol{S}_{\boldsymbol{\beta}} - \boldsymbol{S}_{\mathrm{eff},\boldsymbol{\beta}})\} &= E\{\boldsymbol{\sigma}'_{\boldsymbol{\beta}}(\mathbf{X},\boldsymbol{\beta})\} E\left\{ -\frac{f'_\epsilon(\epsilon)}{f_\epsilon(\epsilon)} \epsilon t - t - \frac{2t^2}{E(t^2)} \right\} = \mathbf{0}.
\end{aligned}$$

The above calculation justifies that $\boldsymbol{S}_{\boldsymbol{\theta}} - \boldsymbol{S}_{\text{eff}} \in \Lambda$.

$$
\begin{aligned}
E(\boldsymbol{S}_{\text{eff},\boldsymbol{\alpha}}|\mathbf{X}) &= -E\left\{\frac{f_\epsilon'(\epsilon)}{f_\epsilon(\epsilon)}\right\}\left[e^{-\sigma(\mathbf{X},\boldsymbol{\beta})}\mathbf{m}_{\boldsymbol{\alpha}}'(\mathbf{X},\boldsymbol{\alpha}) - E\{e^{-\sigma(\mathbf{X},\boldsymbol{\beta})}\mathbf{m}_{\boldsymbol{\alpha}}'(\mathbf{X},\boldsymbol{\alpha})\}\right] \\
&\quad + E\{e^{-\boldsymbol{\alpha}(\mathbf{X},\boldsymbol{\beta})}\mathbf{m}_{\boldsymbol{\alpha}}'(\mathbf{X},\boldsymbol{\alpha})\}E\left\{\epsilon - \frac{E(\epsilon^3)}{E(t^2)}t\right\} = \mathbf{0}, \\
E(\boldsymbol{S}_{\text{eff},\boldsymbol{\alpha}}|\epsilon) &= -\frac{f_\epsilon'(\epsilon)}{f_\epsilon(\epsilon)}E\left[e^{-\sigma(\mathbf{X},\boldsymbol{\beta})}\mathbf{m}_{\boldsymbol{\alpha}}'(\mathbf{X},\boldsymbol{\alpha}) - E\{e^{-\sigma(\mathbf{X},\boldsymbol{\beta})}\mathbf{m}_{\boldsymbol{\alpha}}'(\mathbf{X},\boldsymbol{\alpha})\}\right] \\
&\quad + E\{e^{-\boldsymbol{\alpha}(\mathbf{X},\boldsymbol{\beta})}\mathbf{m}_{\boldsymbol{\alpha}}'(\mathbf{X},\boldsymbol{\alpha})\}\left\{\epsilon - \frac{E(\epsilon^3)}{E(t^2)}t\right\} \\
&= E\{e^{-\boldsymbol{\alpha}(\mathbf{X},\boldsymbol{\beta})}\mathbf{m}_{\boldsymbol{\alpha}}'(\mathbf{X},\boldsymbol{\alpha})\}\epsilon - \frac{E\{e^{-\boldsymbol{\alpha}(\mathbf{X},\boldsymbol{\beta})}\mathbf{m}_{\boldsymbol{\alpha}}'(\mathbf{X},\boldsymbol{\alpha})\}E(\epsilon^3)}{E(t^2)}t.
\end{aligned}
$$

Note that the fact $E\{f_\epsilon'(\epsilon)/f_\epsilon(\epsilon)\} = 0$ and $E(\epsilon) = E(t) = 0$ is used in the above calculation.

In the above derivation, $E(\boldsymbol{S}_{\text{eff},\boldsymbol{\alpha}}|\mathbf{X}) = \mathbf{0}$ and $E(\boldsymbol{S}_{\text{eff},\boldsymbol{\alpha}}|\epsilon)$ has a form of $\mathbf{c}_1\epsilon + \mathbf{c}_2 t$, where $\mathbf{c}_1$ and $\mathbf{c}_2$ are constant vectors.

$$
\begin{aligned}
E(\boldsymbol{S}_{\text{eff},\boldsymbol{\beta}}|\mathbf{X}) &= E\left\{-\frac{f_\epsilon'(\epsilon)}{f_\epsilon(\epsilon)}\epsilon - 1\right\}\left[\boldsymbol{\sigma}_{\boldsymbol{\beta}}'(\mathbf{X},\boldsymbol{\beta}) - E\{\boldsymbol{\sigma}_{\boldsymbol{\beta}}'(\mathbf{X},\boldsymbol{\beta})\}\right] + E\{\boldsymbol{\sigma}_{\boldsymbol{\beta}}'(\mathbf{X},\boldsymbol{\beta})\}E\left\{\frac{2t}{E(t^2)}\right\} = \mathbf{0}, \\
E(\boldsymbol{S}_{\text{eff},\boldsymbol{\beta}}|\epsilon) &= \left\{-\frac{f_\epsilon'(\epsilon)}{f_\epsilon(\epsilon)}\epsilon - 1\right\}E\left[\boldsymbol{\sigma}_{\boldsymbol{\beta}}'(\mathbf{X},\boldsymbol{\beta}) - E\{\boldsymbol{\sigma}_{\boldsymbol{\beta}}'(\mathbf{X},\boldsymbol{\beta})\}\right] + E\{\boldsymbol{\sigma}_{\boldsymbol{\beta}}'(\mathbf{X},\boldsymbol{\beta})\}\frac{2t}{E(t^2)} \\
&= \frac{2E\{\boldsymbol{\sigma}_{\boldsymbol{\beta}}'(\mathbf{X},\boldsymbol{\beta})\}}{E(t^2)}t.
\end{aligned}
$$

The fact $E\{-f_\epsilon'(\epsilon)\epsilon/f_\epsilon(\epsilon)\} = 1$ simplifies the above equation.

Thus, for the $\boldsymbol{S}_{\text{eff},\boldsymbol{\beta}}$, we also have $E(\boldsymbol{S}_{\text{eff},\boldsymbol{\beta}}|\mathbf{X}) = \mathbf{0}$ and $E(\boldsymbol{S}_{\text{eff},\boldsymbol{\beta}}|\epsilon)$ is of the form a constant vector times t. Therefore, $\boldsymbol{S}_{\text{eff}} \in \Lambda^{\perp}$.

Now we find the optimal efficiency matrix $E(\boldsymbol{S}_{\text{eff}}\boldsymbol{S}_{\text{eff}}^{\text{T}})$. We can consider $E(\boldsymbol{S}_{\text{eff},\boldsymbol{\alpha}}\boldsymbol{S}_{\text{eff},\boldsymbol{\alpha}}^{\text{T}})$, $E(\boldsymbol{S}_{\text{eff},\boldsymbol{\alpha}}\boldsymbol{S}_{\text{eff},\boldsymbol{\beta}}^{\text{T}})$, $E(\boldsymbol{S}_{\text{eff},\boldsymbol{\beta}}\boldsymbol{S}_{\text{eff},\boldsymbol{\beta}}^{\text{T}})$ separately.

1. $E(\boldsymbol{S}_{\text{eff},\boldsymbol{\alpha}}\boldsymbol{S}_{\text{eff},\boldsymbol{\alpha}}^{\text{T}})$

$$
\begin{aligned}
&\boldsymbol{S}_{\text{eff},\boldsymbol{\alpha}}\boldsymbol{S}_{\text{eff},\boldsymbol{\alpha}}^{\text{T}} \\
= &\left\{\frac{f_\epsilon'(\epsilon)}{f_\epsilon(\epsilon)}\right\}^2\left[e^{-\sigma(\mathbf{X},\boldsymbol{\beta})}\mathbf{m}_{\boldsymbol{\alpha}}'(\mathbf{X},\boldsymbol{\alpha}) - E\{e^{-\sigma(\mathbf{X},\boldsymbol{\beta})}\mathbf{m}_{\boldsymbol{\alpha}}'(\mathbf{X},\boldsymbol{\alpha})\}\right] \\
&\times\left[e^{-\sigma(\mathbf{X},\boldsymbol{\beta})}\mathbf{m}_{\boldsymbol{\alpha}}'(\mathbf{X},\boldsymbol{\alpha}) - E\{e^{-\sigma(\mathbf{X},\boldsymbol{\beta})}\mathbf{m}_{\boldsymbol{\alpha}}'(\mathbf{X},\boldsymbol{\alpha})\}\right]^{\text{T}} \\
&+\left\{\epsilon - \frac{E(\epsilon^3)}{E(t^2)}t\right\}^2 E\{e^{-\sigma(\mathbf{X},\boldsymbol{\beta})}\mathbf{m}_{\boldsymbol{\alpha}}'(\mathbf{X},\boldsymbol{\alpha})\}E\{e^{-\sigma(\mathbf{X},\boldsymbol{\beta})}\mathbf{m}_{\boldsymbol{\alpha}}'^{\text{T}}(\mathbf{X},\boldsymbol{\alpha})\} \\
&-\frac{f_\epsilon'(\epsilon)}{f_\epsilon(\epsilon)}\left\{\epsilon - \frac{E(\epsilon^3)}{E(t^2)}t\right\}E\{e^{-\sigma(\mathbf{X},\boldsymbol{\beta})}\mathbf{m}_{\boldsymbol{\alpha}}'(\mathbf{X},\boldsymbol{\alpha})\}\left[e^{-\sigma(\mathbf{X},\boldsymbol{\beta})}\mathbf{m}_{\boldsymbol{\alpha}}'(\mathbf{X},\boldsymbol{\alpha}) - E\{e^{-\sigma(\mathbf{X},\boldsymbol{\beta})}\mathbf{m}_{\boldsymbol{\alpha}}'(\mathbf{X},\boldsymbol{\alpha})\}\right]^{\text{T}} \\
&-\frac{f_\epsilon'(\epsilon)}{f_\epsilon(\epsilon)}\left\{\epsilon - \frac{E(\epsilon^3)}{E(t^2)}t\right\}\left[e^{-\sigma(\mathbf{X},\boldsymbol{\beta})}\mathbf{m}_{\boldsymbol{\alpha}}'(\mathbf{X},\boldsymbol{\alpha}) + E\{e^{-\sigma(\mathbf{X},\boldsymbol{\beta})}\mathbf{m}_{\boldsymbol{\alpha}}'(\mathbf{X},\boldsymbol{\alpha})\}\right]E\{e^{-\sigma(\mathbf{X},\boldsymbol{\beta})}\mathbf{m}_{\boldsymbol{\alpha}}'^{\text{T}}(\mathbf{X},\boldsymbol{\alpha})\}.
\end{aligned}
$$

Note that because $\mathbf{X}$ and ϵ are independent, we have $E\{\mathbf{g}_1(\mathbf{X})g_2(\epsilon)\} = E\{\mathbf{g}_1(\mathbf{X})\}E\{g_2(\epsilon)\}$. Hence, we have

$$
\begin{aligned}
& E(\boldsymbol{S}_{\mathrm{eff},\boldsymbol{\alpha}}\boldsymbol{S}_{\mathrm{eff},\boldsymbol{\alpha}}^{\mathrm{T}}) \\
= \; & E\left\{\frac{f_\epsilon'(\epsilon)^2}{f_\epsilon(\epsilon)^2}\right\} \\
& \times \left\{E\{e^{-2\sigma(\mathbf{X},\boldsymbol{\beta})}\mathbf{m}_{\boldsymbol{\alpha}}'(\mathbf{X},\boldsymbol{\alpha})\mathbf{m}_{\boldsymbol{\alpha}}'^{\mathrm{T}}(\mathbf{X},\boldsymbol{\alpha})\} - E\{e^{-\sigma(\mathbf{X},\boldsymbol{\beta})}\mathbf{m}_{\boldsymbol{\alpha}}'(\mathbf{X},\boldsymbol{\alpha})\}E\{e^{-\sigma(\mathbf{X},\boldsymbol{\beta})}\mathbf{m}_{\boldsymbol{\alpha}}'^{\mathrm{T}}\}\right\} \\
& + \left[1 + \frac{\{E(\epsilon^3)\}^2}{E(t^2)}\right] E\{e^{-\sigma(\mathbf{X},\boldsymbol{\beta})}\mathbf{m}_{\boldsymbol{\alpha}}'(\mathbf{X},\boldsymbol{\alpha})\}E\{e^{-\sigma(\mathbf{X},\boldsymbol{\beta})}\mathbf{m}_{\boldsymbol{\alpha}}'^{\mathrm{T}}(\mathbf{X},\boldsymbol{\alpha})\}.
\end{aligned}
$$

2. $E(\boldsymbol{S}_{\mathrm{eff},\boldsymbol{\alpha}}\boldsymbol{S}_{\mathrm{eff},\boldsymbol{\beta}}^{\mathrm{T}})$

$$
\begin{aligned}
& \boldsymbol{S}_{\mathrm{eff},\boldsymbol{\alpha}}\boldsymbol{S}_{\mathrm{eff},\boldsymbol{\beta}}^{\mathrm{T}} \\
= \; & \frac{f_\epsilon'(\epsilon)}{f_\epsilon(\epsilon)}\left\{\frac{f_\epsilon'(\epsilon)}{f_\epsilon(\epsilon)}\epsilon + 1\right\}\left[e^{-\sigma(\mathbf{X},\boldsymbol{\beta})}\mathbf{m}_{\boldsymbol{\alpha}}'(\mathbf{X},\boldsymbol{\alpha}) - E\{e^{-\sigma(\mathbf{X},\boldsymbol{\beta})}\mathbf{m}_{\boldsymbol{\alpha}}'(\mathbf{X},\boldsymbol{\alpha})\}\right] \\
& \times \left[\boldsymbol{\sigma}_{\boldsymbol{\beta}}'(\mathbf{X},\boldsymbol{\beta}) - E\{\boldsymbol{\sigma}_{\boldsymbol{\beta}}'(\mathbf{X},\boldsymbol{\beta})\}\right]^{\mathrm{T}} \\
& - \frac{f_\epsilon'(\epsilon)}{f_\epsilon(\epsilon)}\frac{2t}{E(t^2)}\left[e^{-\sigma(\mathbf{X},\boldsymbol{\beta})}\mathbf{m}_{\boldsymbol{\alpha}}'(\mathbf{X},\boldsymbol{\alpha}) - E\{e^{-\sigma(\mathbf{X},\boldsymbol{\beta})}\mathbf{m}_{\boldsymbol{\alpha}}'(\mathbf{X},\boldsymbol{\alpha})\}\right] E\{\boldsymbol{\sigma}_{\boldsymbol{\beta}}'(\mathbf{X},\boldsymbol{\beta})^{\mathrm{T}}\} \\
& - \left\{\epsilon - \frac{E(\epsilon^3)t}{E(t^2)}\right\}\left\{\frac{f_\epsilon'(\epsilon)}{f_\epsilon(\epsilon)}\epsilon + 1\right\} E\{e^{-\sigma(\mathbf{X},\boldsymbol{\beta})}\mathbf{m}_{\boldsymbol{\alpha}}'(\mathbf{X},\boldsymbol{\alpha})\}\left[\boldsymbol{\sigma}_{\boldsymbol{\beta}}'(\mathbf{X},\boldsymbol{\beta}) - E\{\boldsymbol{\sigma}_{\boldsymbol{\beta}}'(\mathbf{X},\boldsymbol{\beta})\}\right]^{\mathrm{T}} \\
& + \left\{\epsilon - \frac{E(\epsilon^3)t}{E(t^2)}\right\}\frac{2t}{E(t^2)} E\{e^{-\sigma(\mathbf{X},\boldsymbol{\beta})}\mathbf{m}_{\boldsymbol{\alpha}}'(\mathbf{X},\boldsymbol{\alpha})\}E\{\boldsymbol{\sigma}_{\boldsymbol{\beta}}'(\mathbf{X},\boldsymbol{\beta})^{\mathrm{T}}\}.
\end{aligned}
$$

Now we apply the independence assumption of $\mathbf{X}$ and ϵ, then we calculate

$$
\begin{aligned}
E\left[\frac{f_\epsilon'(\epsilon)}{f_\epsilon(\epsilon)}\left\{\frac{f_\epsilon'(\epsilon)}{f_\epsilon(\epsilon)}\epsilon + 1\right\}\right] &= E\left\{\frac{f_\epsilon'(\epsilon)^2}{f_\epsilon(\epsilon)^2}\epsilon\right\}, \\
E\left[\left\{\epsilon - \frac{E(\epsilon^3)t}{E(t^2)}\right\}\frac{2t}{E(t^2)}\right] &= -\frac{2E(\epsilon^3)}{E(t^2)}.
\end{aligned}
$$

This leads to

$$
\begin{aligned}
& E(\boldsymbol{S}_{\mathrm{eff},\boldsymbol{\alpha}}^{\mathrm{T}}\boldsymbol{S}_{\mathrm{eff},\boldsymbol{\beta}}^{\mathrm{T}}) \\
= \; & E\left\{\frac{f_\epsilon'(\epsilon)^2}{f_\epsilon(\epsilon)^2}\epsilon\right\} \\
& \times \left\{E\{e^{-\sigma(\mathbf{X},\boldsymbol{\beta})}\mathbf{m}_{\boldsymbol{\alpha}}'(\mathbf{X},\boldsymbol{\alpha})\boldsymbol{\sigma}_{\boldsymbol{\beta}}'^{\mathrm{T}}(\mathbf{X},\boldsymbol{\beta})\} - E\{e^{-\sigma(\mathbf{X},\boldsymbol{\beta})}\mathbf{m}_{\boldsymbol{\alpha}}'(\mathbf{X},\boldsymbol{\alpha})\}E\{\boldsymbol{\sigma}_{\boldsymbol{\beta}}'(\mathbf{X},\boldsymbol{\beta})^{\mathrm{T}}\}\right\} \\
& - \frac{2E(\epsilon^3)}{E(t^2)} E\{e^{-\sigma(\mathbf{X},\boldsymbol{\beta})}\mathbf{m}_{\boldsymbol{\alpha}}'(\mathbf{X},\boldsymbol{\alpha})\}E\{\boldsymbol{\sigma}_{\boldsymbol{\beta}}'(\mathbf{X},\boldsymbol{\beta})^{\mathrm{T}}\}.
\end{aligned}
$$

3. $E(\boldsymbol{S}_{\mathrm{eff},\boldsymbol{\beta}}\boldsymbol{S}_{\mathrm{eff},\boldsymbol{\beta}}^{\mathrm{T}})$

$$
\begin{aligned}
& \boldsymbol{S}_{\mathrm{eff},\boldsymbol{\beta}}\boldsymbol{S}_{\mathrm{eff},\boldsymbol{\beta}}^{\mathrm{T}} \\
= & \left\{\frac{f_\epsilon'(\epsilon)}{f_\epsilon(\epsilon)}\epsilon+1\right\}^2 \left[\boldsymbol{\sigma}_{\boldsymbol{\beta}}'(\mathbf{X},\boldsymbol{\beta})-E\{\boldsymbol{\sigma}_{\boldsymbol{\beta}}'(\mathbf{X},\boldsymbol{\beta})\}\right]\left[\boldsymbol{\sigma}_{\boldsymbol{\beta}}'(\mathbf{X},\boldsymbol{\beta})-E\{\boldsymbol{\sigma}_{\boldsymbol{\beta}}'(\mathbf{X},\boldsymbol{\beta})\}\right]^{\mathrm{T}} \\
& +\left\{\frac{2t}{E(t^2)}\right\}^2 E\{\boldsymbol{\sigma}_{\boldsymbol{\beta}}'(\mathbf{X},\boldsymbol{\beta})\}E\{\boldsymbol{\sigma}_{\boldsymbol{\beta}}'(\mathbf{X},\boldsymbol{\beta})^{\mathrm{T}}\} \\
& -\left\{\frac{f_\epsilon'(\epsilon)}{f_\epsilon(\epsilon)}\epsilon+1\right\}\frac{2t}{E(t^2)}\left[\boldsymbol{\sigma}_{\boldsymbol{\beta}}'(\mathbf{X},\boldsymbol{\beta})-E\{\boldsymbol{\sigma}_{\boldsymbol{\beta}}'(\mathbf{X},\boldsymbol{\beta})\}\right]E\{\boldsymbol{\sigma}_{\boldsymbol{\beta}}'(\mathbf{X},\boldsymbol{\beta})^{\mathrm{T}}\} \\
& -\left\{\frac{f_\epsilon'(\epsilon)}{f_\epsilon(\epsilon)}\epsilon+1\right\}\frac{2t}{E(t^2)}E\{\boldsymbol{\sigma}_{\boldsymbol{\beta}}'(\mathbf{X},\boldsymbol{\beta})\}\left[\boldsymbol{\sigma}_{\boldsymbol{\beta}}'(\mathbf{X},\boldsymbol{\beta})-E\{\boldsymbol{\sigma}_{\boldsymbol{\beta}}'(\mathbf{X},\boldsymbol{\beta})\}\right]^{\mathrm{T}}.
\end{aligned}
$$

Taking expectations, we have

$$
\begin{aligned}
& E(\boldsymbol{S}_{\mathrm{eff},\boldsymbol{\beta}}\boldsymbol{S}_{\mathrm{eff},\boldsymbol{\beta}}^{\mathrm{T}}) \\
= & E\left\{\left(\frac{f_\epsilon'}{f_\epsilon}\epsilon+1\right)^2\right\}\left[E\{\boldsymbol{\sigma}_{\boldsymbol{\beta}}'(\mathbf{X},\boldsymbol{\beta})\boldsymbol{\sigma}_{\boldsymbol{\beta}}'^{\mathrm{T}}(\mathbf{X},\boldsymbol{\beta})\}-E\{\boldsymbol{\sigma}_{\boldsymbol{\beta}}'(\mathbf{X},\boldsymbol{\beta})\}E\{\boldsymbol{\sigma}_{\boldsymbol{\beta}}'(\mathbf{X},\boldsymbol{\beta})^{\mathrm{T}}\}\right] \\
& +\frac{4}{E(t^2)}E\{\boldsymbol{\sigma}_{\boldsymbol{\beta}}'(\mathbf{X},\boldsymbol{\beta})\}E\{\boldsymbol{\sigma}_{\boldsymbol{\beta}}'(\mathbf{X},\boldsymbol{\beta})^{\mathrm{T}}\} \\
= & E\left\{\frac{f_\epsilon'(\epsilon)^2}{f_\epsilon(\epsilon)^2}\epsilon^2-1\right\}\left[E\{\boldsymbol{\sigma}_{\boldsymbol{\beta}}'(\mathbf{X},\boldsymbol{\beta})\boldsymbol{\sigma}_{\boldsymbol{\beta}}'^{\mathrm{T}}(\mathbf{X},\boldsymbol{\beta})\}-E\{\boldsymbol{\sigma}_{\boldsymbol{\beta}}'(\mathbf{X},\boldsymbol{\beta})\}E\{\boldsymbol{\sigma}_{\boldsymbol{\beta}}'(\mathbf{X},\boldsymbol{\beta})^{\mathrm{T}}\}\right] \\
& +\frac{4}{E(t^2)}E\{\boldsymbol{\sigma}_{\boldsymbol{\beta}}'(\mathbf{X},\boldsymbol{\beta})\}E\{\boldsymbol{\sigma}_{\boldsymbol{\beta}}'(\mathbf{X},\boldsymbol{\beta})^{\mathrm{T}}\}.
\end{aligned}
$$

□

S.3 Proof of Proposition 3

Similar to the proof of Proposition 2, before starting the formal proof, we note that

$$E(\epsilon t|\mathbf{X}) = E(\epsilon^3|\mathbf{X}) - E(\epsilon|\mathbf{X}) = 0,$$

because $f_{\epsilon|\mathbf{X}}(\epsilon|\mathbf{X})$ is a symmetric function of ϵ at any $\mathbf{X}$. This property will be used throughout the proof.

We can construct a nuisance tangent space which satisfies the model assumptions. According to Tsiatis (2006, Section 4.5), we derive the nuisance tangent space Λ_1 with respect to $f_{\mathbf{X}}(\mathbf{x})$.

$$\Lambda_1 = \{\mathbf{a}(\mathbf{x}) : E\{\mathbf{a}(\mathbf{X})\} = \mathbf{0}\}.$$

Similarly, we have the following conditions for a nuisance tangent space Λ_2 which is associated with $f_{\epsilon|\mathbf{X}}(\epsilon, \mathbf{x})$. For $\mathbf{b}(\mathbf{x}, \epsilon) \in \Lambda_2$,

$$E\{\mathbf{b}(\mathbf{X}, \epsilon)|\mathbf{X}\} = E\{\epsilon\mathbf{b}(\mathbf{X}, \epsilon)|\mathbf{X}\} = E\{\epsilon^2\mathbf{b}(\mathbf{X}, \epsilon)|\mathbf{X}\} = \mathbf{0}.$$

Suppose that $f_{\epsilon|\mathbf{X}}(\epsilon, \mathbf{x}, \boldsymbol{\gamma})$ is a parametric submodel of $f_{\epsilon|\mathbf{X}}(\epsilon, \mathbf{x})$. The nuisance tangent space Λ_2 is spanned by $\boldsymbol{S}_{\boldsymbol{\gamma}} = \frac{\partial \log f_{\epsilon|\mathbf{X}}(\epsilon, \mathbf{x}, \boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}}|_{\boldsymbol{\gamma}_0}$, where $f_{\epsilon|\mathbf{X}}(\epsilon, \mathbf{x}, \boldsymbol{\gamma}_0)$ is the true density $f_{\epsilon|\mathbf{X}}(\epsilon, \mathbf{x})$. $\boldsymbol{S}_{\boldsymbol{\gamma}}$ refers to a nuisance score vector. From the symmetry assumption of $f_{\epsilon|\mathbf{X}}(\epsilon, \mathbf{x})$, we have

$$\frac{\partial \log f_{\epsilon|\mathbf{X}}(\epsilon, \mathbf{x}, \boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}} = \frac{\partial \log f_{\epsilon|\mathbf{X}}(-\epsilon, \mathbf{x}, \boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}}.$$

This indicates a nuisance score vector $\boldsymbol{S}_{\boldsymbol{\gamma}}$ is also a symmetric function of ϵ for fixed $\mathbf{x}$. Hence, we have

$$\begin{aligned}
\Lambda_2 &= \{\mathbf{b}(\mathbf{x}, \epsilon) : \mathbf{b}(\mathbf{x}, \epsilon) = \mathbf{b}(\mathbf{x}, -\epsilon), E\{\mathbf{b}(\mathbf{X}, \epsilon)|\mathbf{X}\} = E\{\epsilon\mathbf{b}(\mathbf{X}, \epsilon)|\mathbf{X}\} = E\{\epsilon^2\mathbf{b}(\mathbf{X}, \epsilon)|\mathbf{X}\} = \mathbf{0}\} \\
&= \{\mathbf{b}(\mathbf{x}, \epsilon) : \mathbf{b}(\mathbf{x}, \epsilon) = \mathbf{b}(\mathbf{x}, -\epsilon), E\{\mathbf{b}(\mathbf{X}, \epsilon)|\mathbf{X}\} = E\{\epsilon^2\mathbf{b}(\mathbf{X}, \epsilon)|\mathbf{X}\} = \mathbf{0}\} \\
&= \{\mathbf{b}(\mathbf{x}, \epsilon) : \mathbf{b}(\mathbf{x}, \epsilon) = \mathbf{b}(\mathbf{x}, -\epsilon), E\{\mathbf{b}(\mathbf{X}, \epsilon)|\mathbf{X}\} = E\{t\mathbf{b}(\mathbf{X}, \epsilon)|\mathbf{X}\} = \mathbf{0}\},
\end{aligned}$$

where $t = \epsilon^2 - 1$.

The above equality holds because $\epsilon^{2k-1}\mathbf{b}(\mathbf{x}, \epsilon)$ is an odd function of ϵ at any fixed $\mathbf{x}$ for any integer k. In the above,

$$\begin{aligned}
E\{t\mathbf{b}(\mathbf{X}, \epsilon)|\mathbf{X}\} &= E\{\epsilon^2\mathbf{b}(\mathbf{X}, \epsilon)|\mathbf{X}\} - E\{\mathbf{b}(\mathbf{X}, \epsilon)|\mathbf{X}\} \\
&= E\{\epsilon^2\mathbf{b}(\mathbf{X}, \epsilon)|\mathbf{X}\} = \mathbf{0}.
\end{aligned}$$

Note that $E(\epsilon t|\mathbf{X}) = E(\epsilon^3|\mathbf{X}) - E(\epsilon|\mathbf{X}) = 0$, because ϵ is symmetrically distributed given $\mathbf{X} = \mathbf{x}$.

From the above two subspace Λ_1 and Λ_2, we have

$$\begin{aligned}
\Lambda &= \Lambda_1 \bigoplus \Lambda_2 \\
&= \{\mathbf{h}_1(\mathbf{x}) + \mathbf{h}_2(\mathbf{x}, \epsilon) : E\{\mathbf{h}_1(\mathbf{X})\} = E(\mathbf{h}_2|\mathbf{X}) = E\{\epsilon^2\mathbf{h}_2(\mathbf{X}, \epsilon)|\mathbf{X}\} = \mathbf{0}, \mathbf{h}_2(\mathbf{x}, \epsilon) = \mathbf{h}_2(\mathbf{x}, -\epsilon)\} \\
&= \{\mathbf{h}(\mathbf{x}, \epsilon) : E\{\mathbf{h}(\mathbf{x}, \epsilon)\} = E\{t\mathbf{h}(\mathbf{X}, \epsilon)|\mathbf{X}\} = \mathbf{0}, \mathbf{h}(\mathbf{x}, \epsilon) = \mathbf{h}(\mathbf{x}, -\epsilon)\}.
\end{aligned}$$

First, we find $\Lambda_1^{\perp}$. We claim $\Lambda_1^{\perp} = \{\mathbf{g}(\mathbf{x}, \epsilon) : E\{\mathbf{g}(\mathbf{X}, \epsilon)|\mathbf{X}\} = \mathbf{0}\}$.

Assume $\mathbf{K}_1 = \{\mathbf{g}(\mathbf{x}, \epsilon) : E\{\mathbf{g}(\mathbf{X}, \epsilon)|\mathbf{X}\} = \mathbf{0}\}$. For any $\mathbf{h}_1(\mathbf{x}) \in \Lambda_1$ and $\mathbf{g}(\mathbf{x}, \epsilon) \in \mathbf{K}_1$,

$$E\{\mathbf{h}_1(\mathbf{X})^{\mathrm{T}}\mathbf{g}(\mathbf{X}, \epsilon)\} = E[E\{\mathbf{h}_1(\mathbf{X})^{\mathrm{T}}\mathbf{g}(\mathbf{X}, \epsilon)|\mathbf{X}\}] = E[\mathbf{h}_1(\mathbf{X})^{\mathrm{T}}E\{\mathbf{g}(\mathbf{X}, \epsilon)|\mathbf{X}\}] = 0.$$

Therefore, $\mathbf{K}_1 \subset \Lambda_1^{\perp}$.

We need to show $\Lambda_1^{\perp} \subset \mathbf{K}_1$. Assume $\mathbf{g}(\mathbf{x}, \epsilon) \in \Lambda_1^{\perp}$. We can decompose

$$\mathbf{g}(\mathbf{X}, \epsilon) = \mathbf{g}(\mathbf{X}, \epsilon) - E\{\mathbf{g}(\mathbf{X}, \epsilon)|\mathbf{X}\} + E\{\mathbf{g}(\mathbf{X}, \epsilon)|\mathbf{X}\}.$$

Since $\mathbf{g}(\mathbf{x}, \epsilon)$ is included in the Hilbert space $\mathcal{H}$,

$$E\left[E\{\mathbf{g}(\mathbf{X}, \epsilon)|\mathbf{X}\}\right] = E\{\mathbf{g}(\mathbf{X}, \epsilon)\} = \mathbf{0}.$$

From the above relationship, $E\{\mathbf{g}(\mathbf{X}, \epsilon)|\mathbf{X}\} \in \Lambda_1$. Consider $\mathbf{g}(\mathbf{X}, \epsilon) - E\{\mathbf{g}(\mathbf{X}, \epsilon)|\mathbf{X}\}$.

$$E\left[\mathbf{g}(\mathbf{X}, \epsilon) - E\{\mathbf{g}(\mathbf{X}, \epsilon)|\mathbf{X}\}|\mathbf{X}\right] = E\{\mathbf{g}(\mathbf{X}, \epsilon)|\mathbf{X}\} - E\{\mathbf{g}(\mathbf{X}, \epsilon)|\mathbf{X}\} = \mathbf{0}.$$

This indicates $\mathbf{g}(\mathbf{X}, \epsilon) - E\{\mathbf{g}(\mathbf{X}, \epsilon)|\mathbf{X}\} \in \mathbf{K}_1 \subset \Lambda_1^{\perp}$. Since $\mathbf{g}(\mathbf{x}, \epsilon) \in \Lambda_1^{\perp}$, it is naturally obtained that $E\{\mathbf{g}(\mathbf{X}, \epsilon)|\mathbf{X}\} \in \Lambda_1^{\perp}$. Simultaneously, $E\{\mathbf{g}(\mathbf{X}, \epsilon)|\mathbf{X}\} \in \Lambda_1$ and $E\{\mathbf{g}(\mathbf{X}, \epsilon)|\mathbf{X}\} \in \Lambda_1^{\perp}$. Consequently, we have $E\{\mathbf{g}(\mathbf{X}, \epsilon)|\mathbf{X}\} = \mathbf{0}$. This results in $\mathbf{g}(\mathbf{x}, \epsilon) \in \mathbf{K}_1$ and $\Lambda_1^{\perp} \subset \mathbf{K}_1$.

Next, we derive $\Lambda_2^{\perp}$. We claim $\Lambda_2^{\perp} = \{\mathbf{a}(\mathbf{x}, \epsilon) + \mathbf{b}(\mathbf{x})t + \mathbf{c}(\mathbf{x}) : \mathbf{a}(\mathbf{x}, \epsilon) + \mathbf{a}(\mathbf{x}, -\epsilon) = \mathbf{0}\}$. Let $\mathbf{K}_2 = \{\mathbf{a}(\mathbf{x}, \epsilon) + \mathbf{b}(\mathbf{x})t + \mathbf{c}(\mathbf{x}) : \mathbf{a}(\mathbf{x}, \epsilon) + \mathbf{a}(\mathbf{x}, -\epsilon) = \mathbf{0}\}$. We will show $\mathbf{K}_2 = \Lambda_2^{\perp}$. For an arbitrary $\mathbf{h}(\mathbf{x}, \epsilon) \in \Lambda_2$ and an arbitrary $\mathbf{g}(\mathbf{x}, \epsilon) = \mathbf{a}(\mathbf{x}, \epsilon) + \mathbf{b}(\mathbf{x})t + \mathbf{c}(\mathbf{x}) \in \mathbf{K}_2$, we have

$$\begin{aligned}
& E\{\mathbf{h}(\mathbf{X}, \epsilon)^{\mathrm{T}}\mathbf{g}(\mathbf{X}, \epsilon)\} \\
= \;& E\left[\mathbf{h}(\mathbf{X}, \epsilon)^{\mathrm{T}}\{\mathbf{a}(\mathbf{X}, \epsilon) + \mathbf{b}(\mathbf{X})t + \mathbf{c}(\mathbf{X})\}\right] \\
= \;& E\{\mathbf{h}(\mathbf{X}, \epsilon)^{\mathrm{T}}\mathbf{a}(\mathbf{X}, \epsilon)\} + E\left[E\{\mathbf{h}(\mathbf{X}, \epsilon)t|\mathbf{X}\}^{\mathrm{T}}\mathbf{b}(\mathbf{X})\right] + E\left[E\{\mathbf{h}(\mathbf{X}, \epsilon)|\mathbf{X}\}^{\mathrm{T}}\mathbf{c}(\mathbf{X})\right]. \quad \text{(S.1)}
\end{aligned}$$

Consider the first term in (S.1). Because $\mathbf{h}(\mathbf{x}, \epsilon)$ is an even function of ϵ at any fixed $\mathbf{x}$, $\mathbf{a}(\mathbf{x}, \epsilon)$ is an odd function of ϵ at any fixed $\mathbf{x}$ and $f_{\epsilon|\mathbf{X}}(\epsilon, \mathbf{x})$ is an even function of ϵ, we have

$$E\{\mathbf{h}(\mathbf{X}, \epsilon)^{\mathrm{T}}\mathbf{a}(\mathbf{X}, \epsilon)\} = E\left[E\{\mathbf{h}(\mathbf{X}, \epsilon)^{\mathrm{T}}\mathbf{a}(\mathbf{X}, \epsilon)|\mathbf{X}\}\right] = 0.$$

The second term and the third term in (S.1) are zero because $E\{\mathbf{h}(\mathbf{X}, \epsilon)t|\mathbf{X}\} = E\{\mathbf{h}(\mathbf{X}, \epsilon)|\mathbf{X}\} = \mathbf{0}$ if $\mathbf{h}(\mathbf{x}, \epsilon) \in \Lambda_2$. Thus, we conclude that $\mathbf{K}_2 \subset \Lambda_2^{\perp}$. We will prove that $\Lambda_2^{\perp} \subset \mathbf{K}_2$.

Assume $\mathbf{g}(\mathbf{x},\epsilon) \in \Lambda_2^{\perp}$. We can decompose $\mathbf{g}(\mathbf{x},\epsilon)$ as

$$\mathbf{g}(\mathbf{x},\epsilon) = \boldsymbol{\gamma}_1(\mathbf{x},\epsilon) + \boldsymbol{\gamma}_2(\mathbf{x},\epsilon),$$

where

$$\begin{aligned}\boldsymbol{\gamma}_1(\mathbf{x},\epsilon) &= \frac{\mathbf{g}(\mathbf{x},\epsilon)+\mathbf{g}(\mathbf{x},-\epsilon)}{2} - \frac{E(t\mathbf{g}|\mathbf{X})}{E(t^2|\mathbf{X})}t - E\{\mathbf{g}(\mathbf{X},\epsilon)|\mathbf{X}\},\\ \boldsymbol{\gamma}_2(\mathbf{x},\epsilon) &= \frac{\mathbf{g}(\mathbf{x},\epsilon)-\mathbf{g}(\mathbf{x},-\epsilon)}{2} + \frac{E(t\mathbf{g}|\mathbf{X})}{E(t^2|\mathbf{X})}t + E\{\mathbf{g}(\mathbf{X},\epsilon)|\mathbf{X}\}.\end{aligned}$$

We let

$$\mathbf{a}(\mathbf{x},\epsilon) = \frac{\mathbf{g}(\mathbf{x},\epsilon)-\mathbf{g}(\mathbf{x},-\epsilon)}{2}.$$

Then $\mathbf{a}(\mathbf{x},\epsilon)$ is an odd function of ϵ. Therefore, $\boldsymbol{\gamma}_2(\mathbf{x},\epsilon) \in \mathbf{K}_2 \subset \Lambda_2^{\perp}$. Since both $\mathbf{g}(\mathbf{x},\epsilon) \in \Lambda_2^{\perp}$ and $\boldsymbol{\gamma}_2(\mathbf{x},\epsilon) \in \Lambda_2^{\perp}$, we obtain $\boldsymbol{\gamma}_1(\mathbf{x},\epsilon) \in \Lambda_2^{\perp}$. On the other hand,

$$\begin{aligned}\boldsymbol{\gamma}_1(\mathbf{x},\epsilon) &= \frac{\mathbf{g}(\mathbf{x},\epsilon)+\mathbf{g}(\mathbf{x},-\epsilon)}{2} - \frac{E(t\mathbf{g}|\mathbf{X})}{E(t^2|\mathbf{X})}t - E\{\mathbf{g}(\mathbf{x},\epsilon)|\mathbf{X}\} = \boldsymbol{\gamma}_1(\mathbf{x},-\epsilon),\\ E\{\boldsymbol{\gamma}_1(\mathbf{X},\epsilon)|\mathbf{X}\} &= \frac{E\{\mathbf{g}(\mathbf{X},\epsilon)|\mathbf{X}\}+E\{\mathbf{g}(\mathbf{X},-\epsilon)|\mathbf{X}\}}{2} - \frac{E(t\mathbf{g}|\mathbf{X})}{E(t^2|\mathbf{X})}E(t|\mathbf{X}) - E\{\mathbf{g}(\mathbf{X},\epsilon)|\mathbf{X}\} = \mathbf{0}.\end{aligned}$$

The last equality hold because $E\{\mathbf{g}(\mathbf{x},\epsilon)|\mathbf{X}\} = E\{\mathbf{g}(\mathbf{x},-\epsilon)|\mathbf{X}\}$, specifically,

$$\begin{aligned}E\{\mathbf{g}(\mathbf{X},\epsilon)|\mathbf{X}\} &= \int_{-\infty}^{\infty} \mathbf{g}(\mathbf{x},\epsilon) f_{\epsilon|\mathbf{X}}(\epsilon,\mathbf{x})d\epsilon\\ &= \int_{\infty}^{-\infty} \mathbf{g}(\mathbf{x},-\epsilon) f_{\epsilon|\mathbf{X}}(-\epsilon,\mathbf{x})d(-\epsilon)\\ &= \int_{-\infty}^{\infty} \mathbf{g}(\mathbf{x},-\epsilon) f_{\epsilon|\mathbf{X}}(-\epsilon,\mathbf{x})d\epsilon\\ &= \int_{-\infty}^{\infty} \mathbf{g}(\mathbf{x},-\epsilon) f_{\epsilon|\mathbf{X}}(\epsilon,\mathbf{x})d\epsilon = E\{\mathbf{g}(\mathbf{X},-\epsilon)|\mathbf{X}\}.\end{aligned}$$

In addition,

$$\begin{aligned}& E\{t\boldsymbol{\gamma}_1(\mathbf{X},\epsilon)|\mathbf{X}\}\\ =\ & \frac{E\{t\mathbf{g}(\mathbf{X},\epsilon)|\mathbf{X}\}+E\{t\mathbf{g}(\mathbf{X},-\epsilon)|\mathbf{X}\}}{2} - \frac{E(t\mathbf{g}|\mathbf{X})}{E(t^2|\mathbf{X})}E(t^2|\mathbf{X}) - E\{\mathbf{g}(\mathbf{X},\epsilon)|\mathbf{X}\}E(t|\mathbf{X}) = \mathbf{0}.\end{aligned}$$

Similarly, the above equality holds because $E\{t\mathbf{g}(\mathbf{x},\epsilon)|\mathbf{X}\} = E\{t\mathbf{g}(\mathbf{x},-\epsilon)|\mathbf{X}\}$. It is calculated in the following way.

$$
\begin{aligned}
E\{t\mathbf{g}(\mathbf{X},\epsilon)|\mathbf{X}\} &= \int_{-\infty}^{\infty}(\epsilon^2-1)\mathbf{g}(\mathbf{x},\epsilon)f_{\epsilon|\mathbf{X}}(\epsilon,\mathbf{x})d\epsilon \\
&= \int_{\infty}^{-\infty}(\epsilon^2-1)\mathbf{g}(\mathbf{x},-\epsilon)f_{\epsilon|\mathbf{X}}(-\epsilon,\mathbf{x})d(-\epsilon) \\
&= \int_{-\infty}^{\infty}(\epsilon^2-1)\mathbf{g}(\mathbf{x},-\epsilon)f_{\epsilon|\mathbf{X}}(-\epsilon,\mathbf{x})d\epsilon \\
&= \int_{-\infty}^{\infty}(\epsilon^2-1)\mathbf{g}(\mathbf{x},-\epsilon)f_{\epsilon|\mathbf{X}}(\epsilon,\mathbf{x})d\epsilon = E\{t\mathbf{g}(\mathbf{X},-\epsilon)|\mathbf{X}\}.
\end{aligned}
$$

Thus, $\boldsymbol{\gamma}_1(\mathbf{x},\epsilon)\in\Lambda_2$. Since $\boldsymbol{\gamma}_1(\mathbf{x},\epsilon)\in\Lambda_2$ and $\boldsymbol{\gamma}_1(\mathbf{x},\epsilon)\in\Lambda_2^{\perp}$, we have $\boldsymbol{\gamma}_1(\mathbf{x},\epsilon)=\mathbf{0}$. Consequently, we have $\mathbf{g}=\boldsymbol{\gamma}_2\in\mathbf{K}_2$. It proves that $\Lambda_2^{\perp}\subset\mathbf{K}_2$.

From the fact $\Lambda^{\perp}=\Lambda_1^{\perp}\cap\Lambda_2^{\perp}$, we obtain

$$
\Lambda^{\perp}=\{\mathbf{a}(\mathbf{x},\epsilon)+\mathbf{b}(\mathbf{x})t,\mathbf{a}(\mathbf{x},\epsilon)+\mathbf{a}(\mathbf{x},-\epsilon)=\mathbf{0}\}.
$$

□

S.4 Proof of Theorem 3

The joint probability distribution of $\mathbf{X}$ and Y is given by

$$
f_{\mathbf{X},Y}(\mathbf{x},y)=f_{\mathbf{X}}(\mathbf{x})e^{-\sigma(\mathbf{X},\boldsymbol{\beta})}f_{\epsilon|\mathbf{X}}(\epsilon,\mathbf{x}),
$$

where $\epsilon=\dfrac{y-m(\mathbf{x},\boldsymbol{\alpha})}{e^{\sigma(\mathbf{x},\boldsymbol{\beta})}}$.

We obtain the score functions of $\boldsymbol{\theta}=(\boldsymbol{\alpha}^{\mathrm{T}},\boldsymbol{\beta}^{\mathrm{T}})^{\mathrm{T}}$ as

$$
\begin{aligned}
\boldsymbol{S}_{\boldsymbol{\alpha}} &= \frac{\partial\log f_{\mathbf{X},Y}(\mathbf{x},y)}{\partial\boldsymbol{\alpha}} = -\frac{\partial f_{\epsilon|\mathbf{X}}(\epsilon,\mathbf{x})/\partial\epsilon}{f_{\epsilon|\mathbf{X}}(\epsilon,\mathbf{x})}e^{-\sigma(\mathbf{X},\boldsymbol{\beta})}\mathbf{m}'_{\boldsymbol{\alpha}}(\mathbf{X},\boldsymbol{\alpha}), \\
\boldsymbol{S}_{\boldsymbol{\beta}} &= \frac{\partial\log f_{\mathbf{X},Y}(\mathbf{x},y)}{\partial\boldsymbol{\beta}} = -\frac{\partial f_{\epsilon|\mathbf{X}}(\epsilon,\mathbf{x})/\partial\epsilon}{f_{\epsilon|\mathbf{X}}(\epsilon,\mathbf{x})}\epsilon\boldsymbol{\sigma}'_{\boldsymbol{\beta}}(\mathbf{x},\boldsymbol{\beta})-\boldsymbol{\sigma}'_{\boldsymbol{\beta}}(\mathbf{x},\boldsymbol{\beta}).
\end{aligned}
$$

Here $\boldsymbol{S}_{\boldsymbol{\alpha}}$ is an odd function of ϵ, $\boldsymbol{S}_{\boldsymbol{\beta}}$ is an even function of ϵ. We now show that the efficient scores are

$$
\begin{aligned}
\boldsymbol{S}_{\mathrm{eff},\boldsymbol{\alpha}} &= -\frac{\partial f_{\epsilon|\mathbf{X}}(\epsilon,\mathbf{X})/\partial\epsilon}{f_{\epsilon|\mathbf{X}}(\epsilon,\mathbf{X})}e^{-\sigma(\mathbf{X},\boldsymbol{\beta})}\mathbf{m}'_{\boldsymbol{\alpha}}(\mathbf{X},\boldsymbol{\alpha}), \\
\boldsymbol{S}_{\mathrm{eff},\boldsymbol{\beta}} &= \frac{2t}{E(t^2|\mathbf{X})}\boldsymbol{\sigma}'_{\boldsymbol{\beta}}(\mathbf{X},\boldsymbol{\beta}).
\end{aligned}
$$

Here, $\boldsymbol{S}_{\text{eff},\boldsymbol{\alpha}}$ is an odd function of ϵ for a fixed $\mathbf{x}$ and $\boldsymbol{S}_{\text{eff},\boldsymbol{\beta}}$ has a form of $\mathbf{b}(\mathbf{x})t$. Thus, $\boldsymbol{S}_{\text{eff}} \in \Lambda^{\perp}$.

Now we prove that $\boldsymbol{S}_{\boldsymbol{\theta}} - \boldsymbol{S}_{\text{eff}} \in \Lambda$. $\boldsymbol{S}_{\boldsymbol{\theta}} - \boldsymbol{S}_{\text{eff}}$ is given by

$$\begin{aligned}
\boldsymbol{S}_{\boldsymbol{\alpha}} - \boldsymbol{S}_{\text{eff},\boldsymbol{\alpha}} &= \mathbf{0}, \\
\boldsymbol{S}_{\boldsymbol{\beta}} - \boldsymbol{S}_{\text{eff},\boldsymbol{\beta}} &= \left\{ -\frac{\partial f_{\epsilon|\mathbf{X}}(\epsilon, \mathbf{X})/\partial \epsilon}{f_{\epsilon|\mathbf{X}}(\epsilon, \mathbf{X})} \epsilon - 1 - \frac{2t}{E(t^2|\mathbf{X})} \right\} \boldsymbol{\sigma}'_{\boldsymbol{\beta}}(\mathbf{X}, \boldsymbol{\beta}).
\end{aligned}$$

Obviously, $\boldsymbol{S}_{\boldsymbol{\alpha}} - \boldsymbol{S}_{\text{eff},\boldsymbol{\alpha}} \in \Lambda$. For $\boldsymbol{S}_{\boldsymbol{\beta}} - \boldsymbol{S}_{\text{eff},\boldsymbol{\beta}}$, we need to show it is an even function of ϵ and

$$E(\boldsymbol{S}_{\boldsymbol{\beta}} - \boldsymbol{S}_{\text{eff},\boldsymbol{\beta}}) = \mathbf{0}, \quad E\{t(\boldsymbol{S}_{\boldsymbol{\beta}} - \boldsymbol{S}_{\text{eff},\boldsymbol{\beta}})|\mathbf{X}\} = \mathbf{0}.$$

For fixed $\mathbf{x}$, we can easily verify that $-\frac{\partial f_{\epsilon|\mathbf{X}}(\epsilon, \mathbf{X})/\partial \epsilon}{f_{\epsilon|\mathbf{X}}(\epsilon, \mathbf{X})} \epsilon$ and t are even function of ϵ. It follows that $\boldsymbol{S}_{\boldsymbol{\beta}} - \boldsymbol{S}_{\text{eff},\boldsymbol{\beta}}$ is even function of ϵ. Also, we obtain

$$\begin{aligned}
E(\boldsymbol{S}_{\boldsymbol{\beta}} - \boldsymbol{S}_{\text{eff},\boldsymbol{\beta}}|\mathbf{X}) &= E\left\{ -\frac{\partial f_{\epsilon|\mathbf{X}}(\epsilon, \mathbf{X})/\partial \epsilon}{f_{\epsilon|\mathbf{X}}(\epsilon, \mathbf{X})} \epsilon - 1 - \frac{2t}{E(t^2|\mathbf{X})} |\mathbf{X} \right\} \boldsymbol{\sigma}'_{\boldsymbol{\beta}}(\mathbf{X}, \boldsymbol{\beta}) = \mathbf{0}, \\
E\{t(\boldsymbol{S}_{\boldsymbol{\beta}} - \boldsymbol{S}_{\text{eff},\boldsymbol{\beta}})|\mathbf{X}\} &= E\left\{ -\frac{\partial f_{\epsilon|\mathbf{X}}(\epsilon, \mathbf{X})/\partial \epsilon}{f_{\epsilon|\mathbf{X}}(\epsilon, \mathbf{X})} \epsilon t - t - \frac{2t^2}{E(t^2|\mathbf{X})} |\mathbf{X} \right\} \boldsymbol{\sigma}'_{\boldsymbol{\beta}}(\mathbf{X}, \boldsymbol{\beta}) = \mathbf{0}.
\end{aligned}$$

Thus, $\boldsymbol{S}_{\boldsymbol{\beta}} - \boldsymbol{S}_{\text{eff},\boldsymbol{\beta}} \in \Lambda$.

Now we find the optimal efficiency matrix.

$$E(\boldsymbol{S}_{\text{eff}} \boldsymbol{S}_{\text{eff}}^{\text{T}}) = E \left\{ \begin{array}{cc} \boldsymbol{S}_{\text{eff},\boldsymbol{\alpha}} \boldsymbol{S}_{\text{eff},\boldsymbol{\alpha}}^{\text{T}} & \boldsymbol{S}_{\text{eff},\boldsymbol{\alpha}} \boldsymbol{S}_{\text{eff},\boldsymbol{\beta}}^{\text{T}} \\ \boldsymbol{S}_{\text{eff},\boldsymbol{\beta}} \boldsymbol{S}_{\text{eff},\boldsymbol{\alpha}}^{\text{T}} & \boldsymbol{S}_{\text{eff},\boldsymbol{\beta}} \boldsymbol{S}_{\text{eff},\boldsymbol{\beta}}^{\text{T}} \end{array} \right\},$$

where

$$\begin{aligned}
E(\boldsymbol{S}_{\text{eff},\boldsymbol{\alpha}} \boldsymbol{S}_{\text{eff},\boldsymbol{\alpha}}^{\text{T}}) &= E\left[\left\{ \frac{\partial f_{\epsilon|\mathbf{X}}(\epsilon, \mathbf{X})/\partial \epsilon}{f_{\epsilon|\mathbf{X}}(\epsilon, \mathbf{X})} \right\}^2 e^{-2\sigma(\mathbf{X}, \boldsymbol{\beta})} \mathbf{m}'_{\boldsymbol{\alpha}}(\mathbf{X}, \boldsymbol{\alpha}) \mathbf{m}'^{\text{T}}_{\boldsymbol{\alpha}}(\mathbf{X}, \boldsymbol{\alpha}) \right], \\
E(\boldsymbol{S}_{\text{eff},\boldsymbol{\alpha}} \boldsymbol{S}_{\text{eff},\boldsymbol{\beta}}^{\text{T}}) &= E\left\{ -\frac{\partial f_{\epsilon|\mathbf{X}}(\epsilon, \mathbf{X})/\partial \epsilon}{f_{\epsilon|\mathbf{X}}(\epsilon, \mathbf{X})} \frac{2te^{-\sigma(\mathbf{X}, \boldsymbol{\beta})}}{E(t^2|\mathbf{X})} \mathbf{m}'_{\boldsymbol{\alpha}}(\mathbf{X}, \boldsymbol{\alpha}) \boldsymbol{\sigma}'^{\text{T}}_{\boldsymbol{\beta}}(\mathbf{X}, \boldsymbol{\beta}) \right\}, \\
E(\boldsymbol{S}_{\text{eff},\boldsymbol{\beta}} \boldsymbol{S}_{\text{eff},\boldsymbol{\beta}}^{\text{T}}) &= E\left[\frac{4t^2}{\{E(t^2|\mathbf{X})\}^2} \boldsymbol{\sigma}'_{\boldsymbol{\beta}}(\mathbf{X}, \boldsymbol{\beta}) \boldsymbol{\sigma}'^{\text{T}}_{\boldsymbol{\beta}}(\mathbf{X}, \boldsymbol{\beta}) \right].
\end{aligned}$$

Since $E(\boldsymbol{S}_{\text{eff}} \boldsymbol{S}_{\text{eff}}^{\text{T}}) = E\{E(\boldsymbol{S}_{\text{eff}} \boldsymbol{S}_{\text{eff}}{}^{\text{T}}|\mathbf{X})\}$, $E(\boldsymbol{S}_{\text{eff},\boldsymbol{\alpha}} \boldsymbol{S}_{\text{eff},\boldsymbol{\beta}}^{\text{T}})$ is simplified as

$$\begin{aligned}
& E\left\{ -\frac{\partial f_{\epsilon|\mathbf{X}}(\epsilon, \mathbf{X})/\partial \epsilon}{f_{\epsilon|\mathbf{X}}(\epsilon, \mathbf{X})} \frac{2te^{-\sigma(\mathbf{X}, \boldsymbol{\beta})}}{E(t^2|\mathbf{X})} \mathbf{m}'_{\boldsymbol{\alpha}}(\mathbf{X}, \boldsymbol{\alpha}) \boldsymbol{\sigma}'^{\text{T}}_{\boldsymbol{\beta}}(\mathbf{X}, \boldsymbol{\beta}) \right\} \\
=\ & E\left\{ E\left(-\frac{\partial f_{\epsilon|\mathbf{X}}(\epsilon, \mathbf{X})/\partial \epsilon}{f_{\epsilon|\mathbf{X}}(\epsilon, \mathbf{X})} t|\mathbf{X} \right) \frac{2e^{-\sigma(\mathbf{X}, \boldsymbol{\beta})} \mathbf{m}'_{\boldsymbol{\alpha}}(\mathbf{X}, \boldsymbol{\alpha}) \boldsymbol{\sigma}'^{\text{T}}_{\boldsymbol{\beta}}(\mathbf{X}, \boldsymbol{\beta})}{E(t^2|\mathbf{X})} \right\} \\
=\ & E\left\{ 0 \cdot \frac{2e^{-\sigma(\mathbf{X}, \boldsymbol{\beta})} \mathbf{m}'_{\boldsymbol{\alpha}}(\mathbf{X}, \boldsymbol{\alpha}) \boldsymbol{\sigma}'^{\text{T}}_{\boldsymbol{\beta}}(\mathbf{X}, \boldsymbol{\beta})}{E(t^2|\mathbf{X})} \right\} = \mathbf{0}.
\end{aligned}$$

□

S.5 Proof of Proposition 4

Similar as before, we note that

$$E(\epsilon t) = E(\epsilon^3) - E(\epsilon) = 0,$$

because $f_\epsilon(\epsilon)$ is a symmetric function of ϵ.

We construct $\Lambda = \Lambda_{\mathbf{x}} \oplus \Lambda_\epsilon$, where $\Lambda_{\mathbf{x}}$ is a subspace with functions of $\mathbf{X}$ and Λ_ϵ is a subspace with functions of ϵ. Since $\mathbf{X}$ does not have any constraint and all functions are defined in Hilbert space, we obtain the following subspace $\Lambda_{\mathbf{x}}$ in $\mathcal{H}$.

$$\Lambda_{\mathbf{x}} = \{\mathbf{a}(\mathbf{x}) : E\{\mathbf{a}(\mathbf{X})\} = \mathbf{0}\}.$$

We also have the following conditions for a nuisance tangent space Λ_ϵ which is associated with $f_\epsilon(\epsilon)$. For $\mathbf{b}(\epsilon) \in \Lambda_\epsilon$,

$$E\{\mathbf{b}(\epsilon)\} = E\{\epsilon\mathbf{b}(\epsilon)\} = E\{\epsilon^2\mathbf{b}(\epsilon)\} = \mathbf{0}.$$

Suppose that $f_\epsilon(\epsilon, \boldsymbol{\gamma})$ is a parametric submodel of $f_\epsilon(\epsilon)$. The nuisance tangent space Λ_ϵ is spanned by $\boldsymbol{S}_\gamma = \frac{\partial \log f_\epsilon(\epsilon,\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}}|_{\boldsymbol{\gamma}_0}$, where $f_\epsilon(\epsilon, \boldsymbol{\gamma}_0)$ is the true density $f_\epsilon(\epsilon)$. $\boldsymbol{S}_\gamma$ refers to a nuisance score vector. From the symmetry assumption of $f_\epsilon(\epsilon)$, we have

$$\frac{\partial \log f_\epsilon(\epsilon, \boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}} = \frac{\partial \log f_\epsilon(-\epsilon, \boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}}.$$

This indicates a nuisance score vector $\boldsymbol{S}_{\boldsymbol{\gamma}}$ is also a symmetric function of ϵ. Hence, we have

$$\begin{aligned}
\Lambda_\epsilon &= \{\mathbf{b}(\epsilon) : \mathbf{b}(\epsilon) = \mathbf{b}(-\epsilon), E\{\mathbf{b}(\epsilon)\} = E\{\epsilon\mathbf{b}(\epsilon)\} = E\{\epsilon^2\mathbf{b}(\epsilon)\} = \mathbf{0}\} \\
&= \{\mathbf{b}(\epsilon) : \mathbf{b}(\epsilon) = \mathbf{b}(-\epsilon), E\{\mathbf{b}(\epsilon)\} = E\{\epsilon^2\mathbf{b}(\epsilon)\} = \mathbf{0}\} \\
&= \{\mathbf{b}(\epsilon) : \mathbf{b}(\epsilon) = \mathbf{b}(-\epsilon). E\{\mathbf{b}(\epsilon)\} = E(t\mathbf{b}) = \mathbf{0}\},
\end{aligned}$$

where $t = \epsilon^2 - 1$. The above equality holds because $\epsilon^{2k-1}\mathbf{b}(\epsilon)$ is an odd function of ϵ for any integer k.

In the above,

$$E\{t\mathbf{b}(\epsilon)\} = E\{\epsilon^2\mathbf{b}(\epsilon)\} - E\{\mathbf{b}(\epsilon)\} = E\{\epsilon^2\mathbf{b}(\epsilon)\} = \mathbf{0}.$$

Note that $E(\epsilon t) = E(\epsilon^3) - E(\epsilon) = 0$, because ϵ is symmetrically distributed.

Thus, we have

$$\Lambda = \Lambda_{\mathbf{x}} \oplus \Lambda_{\epsilon} = \{\mathbf{a}(\mathbf{x}) + \mathbf{b}(\epsilon) : E\{\mathbf{a}(\mathbf{x})\} = \mathbf{0}, E\{\mathbf{b}(\epsilon)\} = \mathbf{0}, E\{t\mathbf{b}(\epsilon)\} = \mathbf{0}, \mathbf{b}(\epsilon) = \mathbf{b}(-\epsilon)\}.$$

Note that $\Lambda_{\mathbf{x}}$ and Λ_{ϵ} are orthogonal because we assume $\mathbf{x}$ and ϵ are independent.

We claim $\Lambda^{\perp} = \{\mathbf{g}(\mathbf{x}, \epsilon) : E\{\mathbf{g}(\mathbf{X}, \epsilon)|\mathbf{X}\} = \mathbf{0}, E\{\mathbf{g}(\mathbf{X}, \epsilon)|\epsilon\} = \mathbf{a}(\epsilon) + \mathbf{b}t, \mathbf{a}(\epsilon) + \mathbf{a}(-\epsilon) = \mathbf{0}, \mathbf{b} \in \mathbb{R}^{k+l}\}$. The orthogonal tangent space $\Lambda^{\perp}$ can be obtained by constructing $\Lambda^{\perp} = \Lambda_{\mathbf{x}}^{\perp} \cap \Lambda_{\epsilon}^{\perp}$.

First, We claim $\Lambda_{\mathbf{x}}^{\perp} = \{\mathbf{g}(\mathbf{x}, \epsilon) : E\{\mathbf{g}(\mathbf{X}, \epsilon)|\mathbf{X}\} = \mathbf{0}\}$. Assume $\mathbf{K}_1 = \{\mathbf{g}(\mathbf{x}, \epsilon) : E\{\mathbf{g}(\mathbf{X}, \epsilon)|\mathbf{X}\} = \mathbf{0}\}$. For any $\mathbf{h}(\mathbf{x}) \in \Lambda_{\mathbf{x}}$ and $\mathbf{g}(\mathbf{x}, \epsilon) \in \mathbf{K}_1$,

$$E\{\mathbf{h}(\mathbf{X})^{\mathrm{T}}\mathbf{g}(\mathbf{X}, \epsilon)\} = E[E\{\mathbf{h}(\mathbf{X})^{\mathrm{T}}\mathbf{g}(\mathbf{X}, \epsilon)|\mathbf{X}\}] = E[\mathbf{h}(\mathbf{X})^{\mathrm{T}}E\{\mathbf{g}(\mathbf{X}, \epsilon)|\mathbf{X}\}] = 0.$$

Therefore, $\mathbf{K}_1 \subset \Lambda_{\mathbf{x}}^{\perp}$.
We need to show $\Lambda_{\mathbf{x}}^{\perp} \subset \mathbf{K}_1$. Assume $\mathbf{g}(\mathbf{x}, \epsilon) \in \Lambda_{\mathbf{x}}^{\perp}$. We can decompose

$$\mathbf{g}(\mathbf{x}, \epsilon) = \mathbf{g}(\mathbf{x}, \epsilon) - E\{\mathbf{g}(\mathbf{X}, \epsilon)|\mathbf{X}\} + E\{\mathbf{g}(\mathbf{X}, \epsilon)|\mathbf{X}\}.$$

Since $\mathbf{g}(\mathbf{x}, \epsilon)$ is in Hilbert space $\mathcal{H}$,

$$E\{\mathbf{g}(\mathbf{X}, \epsilon)\} = E\left[E\{\mathbf{g}(\mathbf{X}, \epsilon)|\mathbf{X}\}\right] = \mathbf{0}.$$

It also indicates that $E\{\mathbf{g}(\mathbf{X}, \epsilon)|\mathbf{X}\} \in \Lambda_{\mathbf{x}}$. Now consider $\mathbf{g}(\mathbf{x}, \epsilon) - E\{\mathbf{g}(\mathbf{X}, \epsilon)|\mathbf{X}\}$.

$$E\left[\mathbf{g}(\mathbf{X}, \epsilon) - E\{\mathbf{g}(\mathbf{X}, \epsilon)|\mathbf{X}\}|\mathbf{X}\right] = E\{\mathbf{g}(\mathbf{X}, \epsilon)|\mathbf{X}\} - E\{\mathbf{g}(\mathbf{X}, \epsilon)|\mathbf{X}\} = \mathbf{0}.$$

This indicates $\mathbf{g}(\mathbf{x}, \epsilon) - E\{\mathbf{g}(\mathbf{X}, \epsilon)|\mathbf{X}\} \in \mathbf{K}_1 \subset \Lambda_{\mathbf{x}}^{\perp}$. Since we assume that $\mathbf{g}(\mathbf{x}, \epsilon) \in \Lambda_{\mathbf{x}}^{\perp}$, it is naturally obtained that $E\{\mathbf{g}(\mathbf{X}, \epsilon)|\mathbf{X}\} \in \Lambda_{\mathbf{x}}^{\perp}$. Simultaneously, $E\{\mathbf{g}(\mathbf{X}, \epsilon)|\mathbf{X}\} \in \Lambda_{\mathbf{x}}$ and $E\{\mathbf{g}(\mathbf{X}, \epsilon)|\mathbf{X}\} \in \Lambda_{\mathbf{x}}^{\perp}$. It results in $E\{\mathbf{g}(\mathbf{X}, \epsilon)|\mathbf{X}\} = \mathbf{0}$. Thus, $\mathbf{g}(\mathbf{x}, \epsilon) \in \mathbf{K}_1$ for an arbitrary $\mathbf{g}(\mathbf{x}, \epsilon) \in \Lambda_{\mathbf{x}}^{\perp}$. Consequently, we have $\Lambda_{\mathbf{x}}^{\perp} \subset \mathbf{K}_1$.

Let $\mathbf{K}_2 = \{\mathbf{g}(\mathbf{x}, \epsilon) : E\{\mathbf{g}(\mathbf{X}, \epsilon)|\epsilon\} = \mathbf{a}(\epsilon) + \mathbf{b}t, \mathbf{a}(\epsilon) + \mathbf{a}(-\epsilon) = \mathbf{0}, \mathbf{b} \in \mathbb{R}^{k+l}\}$.

We will show $\mathbf{K}_2 = \Lambda_{\epsilon}^{\perp}$. For an arbitrary $\mathbf{h}(\epsilon) \in \Lambda_{\epsilon}$ and an arbitrary $\mathbf{g}(\mathbf{x}, \epsilon) \in \mathbf{K}_2$, we have

$$\begin{aligned} E\{\mathbf{h}(\epsilon)^{\mathrm{T}}\mathbf{g}(\mathbf{X}, \epsilon)\} &= E\left\{\mathbf{h}(\epsilon)^{\mathrm{T}}E\{\mathbf{g}(\mathbf{X}, \epsilon)|\epsilon\}\right\} = E\left[\mathbf{h}(\epsilon)^{\mathrm{T}}\{\mathbf{a}(\epsilon) + \mathbf{b}t\}\right] \\ &= E\{\mathbf{h}(\epsilon)^{\mathrm{T}}\mathbf{a}(\epsilon)\} + E\{\mathbf{h}(\epsilon)t\}^{\mathrm{T}}\mathbf{b} = 0. \end{aligned}$$

In the above, because $\mathbf{h}(\epsilon)$ is an even function, $\mathbf{a}(\epsilon)$ is an odd function of ϵ and $f_\epsilon(\epsilon)$ is an even function, then we have $E\{\mathbf{h}(\epsilon)^{\mathrm{T}}\mathbf{a}(\epsilon)\} = 0$. And $E\{\mathbf{h}(\epsilon)t\}^{\mathrm{T}}\mathbf{b} = 0$ since $E\{\mathbf{h}(\epsilon)t\} = \mathbf{0}$ for $\mathbf{h}(\mathbf{x}, \epsilon) \in \Lambda_\epsilon$. Thus, we conclude that $\mathbf{K}_2 \subset \Lambda_\epsilon^\perp$.

Now we will show that $\Lambda_\epsilon^\perp \subset \mathbf{K}_2$. For an arbitrary $\mathbf{g}(\mathbf{x}, \epsilon) \in \Lambda_\epsilon^\perp$, we can decompose $\mathbf{g}(\mathbf{x}, \epsilon)$ as

$$\mathbf{g}(\mathbf{x}, \epsilon) = \boldsymbol{\gamma}_1(\epsilon) + \boldsymbol{\gamma}_2(\mathbf{x}, \epsilon),$$

where

$$\begin{aligned}
\boldsymbol{\gamma}_1(\epsilon) &= \frac{E\{\mathbf{g}(\mathbf{x}, \epsilon)|\epsilon\} + E\{\mathbf{g}(\mathbf{x}, -\epsilon)|\epsilon\}}{2} - \frac{E\{t\mathbf{g}(\mathbf{x}, \epsilon) + t\mathbf{g}(\mathbf{x}, -\epsilon)\}}{2E(t^2)}t, \\
\boldsymbol{\gamma}_2(\mathbf{x}, \epsilon) &= \mathbf{g}(\mathbf{x}, \epsilon) - \frac{E\{\mathbf{g}(\mathbf{x}, \epsilon)|\epsilon\} + E\{\mathbf{g}(\mathbf{x}, -\epsilon)|\epsilon\}}{2} + \frac{E\{t\mathbf{g}(\mathbf{x}, \epsilon) + t\mathbf{g}(\mathbf{x}, -\epsilon)\}}{2E(t^2)}t.
\end{aligned}$$

Consider $\boldsymbol{\gamma}_1(\epsilon)$.

$$\begin{aligned}
\boldsymbol{\gamma}_1(\epsilon) &= \frac{E\{\mathbf{g}(\mathbf{x}, \epsilon)|\epsilon\} + E\{\mathbf{g}(\mathbf{x}, -\epsilon)|\epsilon\}}{2} - \frac{E\{t\mathbf{g}(\mathbf{x}, \epsilon) + t\mathbf{g}(\mathbf{x}, -\epsilon)\}}{2E(t^2)}t = \boldsymbol{\gamma}_1(-\epsilon), \\
E\{\boldsymbol{\gamma}_1(\epsilon)\} &= \frac{E\{\mathbf{g}(\mathbf{x}, \epsilon)\} + E\{\mathbf{g}(\mathbf{x}, -\epsilon)\}}{2} - \frac{E\{t\mathbf{g}(\mathbf{x}, \epsilon) + t\mathbf{g}(\mathbf{x}, -\epsilon)\}}{2E(t^2)}E(t) = \mathbf{0}, \\
E\{t\boldsymbol{\gamma}_1(\epsilon)\} &= \frac{E[E\{t\mathbf{g}(\mathbf{x}, \epsilon)|\epsilon\}] + E[E\{t\mathbf{g}(\mathbf{x}, -\epsilon)|\epsilon\}]}{2} - \frac{E\{t\mathbf{g}(\mathbf{x}, \epsilon) + t\mathbf{g}(\mathbf{x}, -\epsilon)\}}{2E(t^2)}E(t^2) \\
&= \frac{E\{t\mathbf{g}(\mathbf{x}, \epsilon) + t\mathbf{g}(\mathbf{x}, -\epsilon)\}}{2E(t^2)} - \frac{E\{t\mathbf{g}(\mathbf{x}, \epsilon) + t\mathbf{g}(\mathbf{x}, -\epsilon)\}}{2E(t^2)} = \mathbf{0}.
\end{aligned}$$

Hence, we have $\boldsymbol{\gamma}_1(\epsilon) \in \Lambda_\epsilon$.

In $\boldsymbol{\gamma}_2(\mathbf{x}, \epsilon)$, we let

$$\mathbf{g}_1(\mathbf{x}, \epsilon) = \mathbf{g}(\mathbf{x}, \epsilon) - \frac{E\{\mathbf{g}(\mathbf{x}, \epsilon)|\epsilon\} + E\{\mathbf{g}(\mathbf{x}, -\epsilon)|\epsilon\}}{2}.$$

Then we have

$$\begin{aligned}
E\{\mathbf{g}_1(\mathbf{x}, \epsilon)|\epsilon\} &= E\{\mathbf{g}(\mathbf{x}, \epsilon)|\epsilon\} - \frac{E\{\mathbf{g}(\mathbf{x}, \epsilon)|\epsilon\} + E\{\mathbf{g}(\mathbf{x}, -\epsilon)|\epsilon\}}{2} \\
&= \frac{E\{\mathbf{g}(\mathbf{x}, \epsilon)|\epsilon\} - E\{\mathbf{g}(\mathbf{x}, -\epsilon)|\epsilon\}}{2} = -E\{\mathbf{g}_1(\mathbf{x}, -\epsilon)|\epsilon\}.
\end{aligned}$$

It implies that $E\{\mathbf{g}_1(\mathbf{x}, \epsilon)|\epsilon\} + E\{\mathbf{g}_1(\mathbf{x}, -\epsilon)|\epsilon\} = \mathbf{0}$. We calculate $E(\boldsymbol{\gamma}(\mathbf{x}, \epsilon)|\epsilon)$ as

$$E(\boldsymbol{\gamma}_2(\mathbf{x}, \epsilon)|\epsilon) = E\{\mathbf{g}_1(\mathbf{x}, \epsilon)|\epsilon\} + \frac{E\{t\mathbf{g}(\mathbf{x}, \epsilon) + t\mathbf{g}(\mathbf{x}, -\epsilon)\}}{2E(t^2)}t.$$

Then we have that $\boldsymbol{\gamma}_2(\mathbf{x},\epsilon) \in \mathbf{K}_2 \subset \Lambda_\epsilon^\perp$. Since both $\mathbf{g}(\mathbf{x},\epsilon) \in \Lambda_\epsilon^\perp$ and $\boldsymbol{\gamma}_2(\mathbf{x},\epsilon) \in \Lambda_\epsilon^\perp$, we obtain $\boldsymbol{\gamma}_1(\epsilon) = \mathbf{g}(\mathbf{x},\epsilon) - \boldsymbol{\gamma}_2(\mathbf{x},\epsilon) \in \Lambda_\epsilon^\perp$. Note that $\boldsymbol{\gamma}_1(\mathbf{x},\epsilon) \in \Lambda_\epsilon$ and $\boldsymbol{\gamma}_1(\mathbf{x},\epsilon) \in \Lambda_\epsilon^\perp$. Thus, we have $\boldsymbol{\gamma}_1(\mathbf{x},\epsilon) = \mathbf{0}$. Consequently, we have $\mathbf{g}(\mathbf{x},\epsilon) = \boldsymbol{\gamma}_2(\mathbf{x},\epsilon) \in \mathbf{K}_2$ for an arbitrary $\mathbf{g} \in \Lambda_\epsilon^\perp$. It proves that $\Lambda_\epsilon^\perp \subset \mathbf{K}_2$.

We conclude that $\Lambda_{\mathbf{x}}^\perp = \{\mathbf{g}(\mathbf{x},\epsilon) : E\{\mathbf{g}(\mathbf{X},\epsilon)|\mathbf{X}\} = \mathbf{0}\}$ and $\Lambda_\epsilon^\perp = \{\mathbf{g}(\mathbf{x},\epsilon) : E\{\mathbf{g}(\mathbf{X},\epsilon)|\epsilon\} = \mathbf{a}(\epsilon) + \mathbf{b}t, \mathbf{a}(\epsilon) + \mathbf{a}(-\epsilon) = \mathbf{0}, \mathbf{b} \in \mathbb{R}^{k+l}\}$. Further, we have

$$\Lambda^\perp = \{\mathbf{g}(\mathbf{x},\epsilon) : E\{\mathbf{g}(\mathbf{X},\epsilon)|\mathbf{X}\} = \mathbf{0}, E\{\mathbf{g}(\mathbf{X},\epsilon)|\epsilon\} = \mathbf{a}(\epsilon) + \mathbf{b}t, \mathbf{a}(\epsilon) + \mathbf{a}(-\epsilon) = \mathbf{0}, \mathbf{b} \in \mathbb{R}^{k+l}\}.$$

□

S.6 Proof of Theorem 4

We have the joint probability distribution function as

$$f_{\mathbf{X},Y}(\mathbf{x},y) = f_{\mathbf{X}}(\mathbf{x}) f_\epsilon\left\{\frac{y - m(\mathbf{x},\boldsymbol{\alpha})}{e^{\sigma(\mathbf{x},\boldsymbol{\beta})}}\right\} \frac{1}{e^{\sigma(\mathbf{x},\boldsymbol{\beta})}} = f_{\mathbf{X}}(\mathbf{x}) f_\epsilon(\epsilon) e^{-\sigma(\mathbf{X},\boldsymbol{\beta})},$$

where $\epsilon = \dfrac{y - m(\mathbf{x},\boldsymbol{\alpha})}{e^{\sigma(\mathbf{x},\boldsymbol{\beta})}}$. This model assumes that $E(\epsilon) = 0$ and $E(\epsilon^2) = 1$. From the fact $f_\epsilon(\epsilon) = f_\epsilon(-\epsilon)$, we also have $E(\epsilon^3) = 0$.

We have score functions of $\boldsymbol{\theta} = (\boldsymbol{\alpha}^{\mathrm{T}}, \boldsymbol{\beta}^{\mathrm{T}})^{\mathrm{T}}$ as

$$\begin{aligned}
\boldsymbol{S}_{\boldsymbol{\alpha}} &= \frac{\partial \log f_{\mathbf{X},Y}(\mathbf{x},y)}{\partial \boldsymbol{\alpha}} = -\frac{\partial f_\epsilon(\epsilon)/\partial\epsilon}{f_\epsilon(\epsilon)} e^{-\sigma(\mathbf{X},\boldsymbol{\beta})} \mathbf{m}'_{\boldsymbol{\alpha}}(\mathbf{X},\boldsymbol{\alpha}), \\
\boldsymbol{S}_{\boldsymbol{\beta}} &= \frac{\partial \log f_{\mathbf{X},Y}(\mathbf{x},y)}{\partial \boldsymbol{\beta}} = -\frac{\partial f_\epsilon(\epsilon)/\partial\epsilon}{f_\epsilon(\epsilon)} \epsilon \boldsymbol{\sigma}'_{\boldsymbol{\beta}}(\mathbf{x},\boldsymbol{\beta}) - \boldsymbol{\sigma}'_{\boldsymbol{\beta}}(\mathbf{x},\boldsymbol{\beta}).
\end{aligned}$$

Note that $\boldsymbol{S}_{\boldsymbol{\alpha}}$ is an odd function of ϵ, $\boldsymbol{S}_{\boldsymbol{\beta}}$ is an even function of ϵ for fixed $\mathbf{x}$. We claim $\boldsymbol{S}_{\mathrm{eff}} = (\boldsymbol{S}_{\mathrm{eff},\boldsymbol{\alpha}}^{\mathrm{T}}, \boldsymbol{S}_{\mathrm{eff},\boldsymbol{\beta}}^{\mathrm{T}})^{\mathrm{T}}$, where

$$\begin{aligned}
\boldsymbol{S}_{\mathrm{eff},\boldsymbol{\alpha}} &= -\frac{f'_\epsilon(\epsilon)}{f_\epsilon(\epsilon)} e^{-\sigma(\mathbf{X},\boldsymbol{\beta})} \mathbf{m}'_{\boldsymbol{\alpha}}(\mathbf{X},\boldsymbol{\alpha}), \\
\boldsymbol{S}_{\mathrm{eff},\boldsymbol{\beta}} &= \left\{-\frac{f'_\epsilon(\epsilon)}{f_\epsilon(\epsilon)}\epsilon - 1\right\} \left[\boldsymbol{\sigma}'_{\boldsymbol{\beta}}(\mathbf{X},\boldsymbol{\beta}) - E\{\boldsymbol{\sigma}'_{\boldsymbol{\beta}}(\mathbf{X},\boldsymbol{\beta})\}\right] + \frac{2t}{E(t^2)} E\{\boldsymbol{\sigma}'_{\boldsymbol{\beta}}(\mathbf{X},\boldsymbol{\beta})\}.
\end{aligned}$$

We calculate conditional expectations of $\boldsymbol{S}_{\mathrm{eff}}$ in the following.

$$\begin{aligned}
E(\boldsymbol{S}_{\mathrm{eff},\boldsymbol{\alpha}}|\mathbf{X}) &= -E\left\{\frac{f'_\epsilon(\epsilon)}{f_\epsilon(\epsilon)}\right\} e^{-\sigma(\mathbf{X},\boldsymbol{\beta})} \mathbf{m}'_{\boldsymbol{\alpha}}(\mathbf{X},\boldsymbol{\alpha}) = 0 \cdot e^{-\sigma(\mathbf{X},\boldsymbol{\beta})} \mathbf{m}'_{\boldsymbol{\alpha}}(\mathbf{X},\boldsymbol{\alpha})\} = \mathbf{0}, \\
E(\boldsymbol{S}_{\mathrm{eff},\boldsymbol{\alpha}}|\epsilon) &= -\frac{f'_\epsilon(\epsilon)}{f_\epsilon(\epsilon)} E\{e^{-\sigma(\mathbf{X},\boldsymbol{\beta})} \mathbf{m}'_{\boldsymbol{\alpha}}(\mathbf{X},\boldsymbol{\alpha})\}.
\end{aligned}$$

Since $E(\boldsymbol{S}_{\text{eff},\boldsymbol{\alpha}}|\mathbf{X}) = \mathbf{0}$ and $E(\boldsymbol{S}_{\text{eff},\boldsymbol{\alpha}}|\epsilon)$ is an odd function, $\boldsymbol{S}_{\text{eff},\boldsymbol{\alpha}} \in \Lambda_\epsilon^\perp$.

$$\begin{aligned}
E(\boldsymbol{S}_{\text{eff},\boldsymbol{\beta}}|\mathbf{X}) &= E\left(-\frac{f'}{f_\epsilon}\epsilon - 1\right)[\boldsymbol{\sigma}'_{\boldsymbol{\beta}}(\mathbf{X},\boldsymbol{\beta}) - E\{\boldsymbol{\sigma}'_{\boldsymbol{\beta}}(\mathbf{X},\boldsymbol{\beta})\}] + \frac{2E(t)}{E(t^2)}E\{\boldsymbol{\sigma}'_{\boldsymbol{\beta}}(\mathbf{X},\boldsymbol{\beta})\} = \mathbf{0}, \\
E(\boldsymbol{S}_{\text{eff},\boldsymbol{\beta}}|\epsilon) &= \left(-\frac{f'}{f_\epsilon}\epsilon - 1\right)[E\{\boldsymbol{\sigma}'_{\boldsymbol{\beta}}(\mathbf{X},\boldsymbol{\beta})\} - E\{\boldsymbol{\sigma}'_{\boldsymbol{\beta}}(\mathbf{X},\boldsymbol{\beta})\}] + \frac{2t}{E(t^2)}E\{\boldsymbol{\sigma}'_{\boldsymbol{\beta}}(\mathbf{X},\boldsymbol{\beta})\} \\
&= \frac{2t}{E(t^2)}E\{\boldsymbol{\sigma}'_{\boldsymbol{\beta}}(\mathbf{X},\boldsymbol{\beta})\}.
\end{aligned}$$

In the above, $E(\boldsymbol{S}_{\text{eff},\boldsymbol{\beta}}|\mathbf{X}) = \mathbf{0}$ and $E(\boldsymbol{S}_{\text{eff},\boldsymbol{\beta}}|\epsilon)$ is a constant vector times t. Therefore, we have $\boldsymbol{S}_{\text{eff},\boldsymbol{\beta}} \in \Lambda_\epsilon^\perp$.

To prove the claim, we also need to verify that $\boldsymbol{S}_{\boldsymbol{\theta}} - \boldsymbol{S}_{\text{eff}} \in \Lambda$. Since $\boldsymbol{S}_{\boldsymbol{\alpha}} = \boldsymbol{S}_{\text{eff},\boldsymbol{\alpha}}$, obviously we have $\boldsymbol{S}_{\boldsymbol{\alpha}} - \boldsymbol{S}_{\text{eff},\boldsymbol{\alpha}} = \mathbf{0} \in \Lambda$. $\boldsymbol{S}_{\boldsymbol{\beta}} - \boldsymbol{S}_{\text{eff},\boldsymbol{\beta}}$ is given by

$$\boldsymbol{S}_{\boldsymbol{\beta}} - \boldsymbol{S}_{\text{eff},\boldsymbol{\beta}} = E\{\boldsymbol{\sigma}'_{\boldsymbol{\beta}}(\mathbf{X},\boldsymbol{\beta})\}\left\{-\frac{f'_\epsilon(\epsilon)}{f_\epsilon(\epsilon)}\epsilon - 1 - \frac{2t}{E(t^2)}\right\}.$$

$\boldsymbol{S}_{\boldsymbol{\beta}} - \boldsymbol{S}_{\text{eff},\boldsymbol{\beta}}$ is a pure function of ϵ. Also, $\boldsymbol{S}_{\boldsymbol{\beta}} - \boldsymbol{S}_{\text{eff},\boldsymbol{\beta}}$ is an even function of ϵ.

$$\begin{aligned}
E(\boldsymbol{S}_{\boldsymbol{\beta}} - \boldsymbol{S}_{\text{eff},\boldsymbol{\beta}}) &= E\{\boldsymbol{\sigma}'_{\boldsymbol{\beta}}(\mathbf{X},\boldsymbol{\beta})\}E\left\{-\frac{f'_\epsilon(\epsilon)}{f_\epsilon(\epsilon)}\epsilon - 1 - \frac{2t}{E(t^2)}\right\} = \mathbf{0}, \\
E\{t(\boldsymbol{S}_{\boldsymbol{\beta}} - \boldsymbol{S}_{\text{eff},\boldsymbol{\beta}})\} &= E\{\boldsymbol{\sigma}'_{\boldsymbol{\beta}}(\mathbf{X},\boldsymbol{\beta})\}E\left\{-\frac{f'_\epsilon(\epsilon)}{f_\epsilon(\epsilon)}\epsilon t - t - \frac{2t^2}{E(t^2)}\right\} = \mathbf{0}.
\end{aligned}$$

The above calculations justify that $\boldsymbol{S}_{\boldsymbol{\beta}} - \boldsymbol{S}_{\text{eff},\boldsymbol{\beta}} \in \Lambda_\epsilon$. Thus, we obtain $\boldsymbol{S}_{\boldsymbol{\theta}} - \boldsymbol{S}_{\text{eff}} \in \Lambda$.

In order to find optimal efficiency matrix, we calculate $E(\boldsymbol{S}_{\text{eff},\boldsymbol{\alpha}}\boldsymbol{S}_{\text{eff},\boldsymbol{\alpha}}^{\text{T}})$, $E(\boldsymbol{S}_{\text{eff},\boldsymbol{\alpha}}\boldsymbol{S}_{\text{eff},\boldsymbol{\beta}}^{\text{T}})$ and $E(\boldsymbol{S}_{\text{eff},\boldsymbol{\beta}}\boldsymbol{S}_{\text{eff},\boldsymbol{\beta}}^{\text{T}})$.

1. $E(\boldsymbol{S}_{\text{eff},\boldsymbol{\alpha}}\boldsymbol{S}_{\text{eff},\boldsymbol{\alpha}}^{\text{T}})$

$$\begin{aligned}
\boldsymbol{S}_{\text{eff},\boldsymbol{\alpha}}\boldsymbol{S}_{\text{eff},\boldsymbol{\alpha}}^{\text{T}} &= \left\{\frac{f'_\epsilon(\epsilon)^2}{f_\epsilon(\epsilon)^2}\right\} e^{-2\sigma(\mathbf{X},\boldsymbol{\beta})}\mathbf{m}'_{\boldsymbol{\alpha}}(\mathbf{X},\boldsymbol{\alpha})\mathbf{m}'_{\boldsymbol{\alpha}}{}^{\text{T}}(\mathbf{X},\boldsymbol{\alpha}), \\
E(\boldsymbol{S}_{\text{eff},\boldsymbol{\alpha}}\boldsymbol{S}_{\text{eff},\boldsymbol{\alpha}}^{\text{T}}) &= E\left\{\frac{f'_\epsilon(\epsilon)^2}{f_\epsilon(\epsilon)^2}\right\} E\{e^{-2\sigma(\mathbf{X},\boldsymbol{\beta})}\mathbf{m}'_{\boldsymbol{\alpha}}(\mathbf{X},\boldsymbol{\alpha})\mathbf{m}'_{\boldsymbol{\alpha}}{}^{\text{T}}(\mathbf{X},\boldsymbol{\alpha})\}.
\end{aligned}$$

2. $E(\boldsymbol{S}_{\text{eff},\boldsymbol{\alpha}}\boldsymbol{S}_{\text{eff},\boldsymbol{\beta}}^{\text{T}})$

$$\begin{aligned}
\boldsymbol{S}_{\text{eff},\boldsymbol{\alpha}}\boldsymbol{S}_{\text{eff},\boldsymbol{\beta}}^{\text{T}} &= \left[\left\{\frac{f_\epsilon'(\epsilon)}{f_\epsilon(\epsilon)}\right\}^2\epsilon + \frac{f_\epsilon'(\epsilon)}{f_\epsilon(\epsilon)}\right] e^{-\sigma(\mathbf{X},\boldsymbol{\beta})}\mathbf{m}_{\boldsymbol{\alpha}}'(\mathbf{X},\boldsymbol{\alpha}) \\
&\quad \times[\boldsymbol{\sigma}_{\boldsymbol{\beta}}'(\mathbf{X},\boldsymbol{\beta}) - E\{\boldsymbol{\sigma}_{\boldsymbol{\beta}}'(\mathbf{X},\boldsymbol{\beta})\}]^{\text{T}} \\
&\quad + \left\{-\frac{f_\epsilon'(\epsilon)}{f_\epsilon(\epsilon)}\right\}\frac{2t}{E(t^2)}e^{-\sigma(\mathbf{X},\boldsymbol{\beta})}\mathbf{m}_{\boldsymbol{\alpha}}'(\mathbf{X},\boldsymbol{\alpha})E\{\boldsymbol{\sigma}_{\boldsymbol{\beta}}'(\mathbf{X},\boldsymbol{\beta})\}^{\text{T}}, \\
E(\boldsymbol{S}_{\text{eff},\boldsymbol{\alpha}}\boldsymbol{S}_{\text{eff},\boldsymbol{\beta}}^{\text{T}}) &= E\left[\left\{\frac{f_\epsilon'(\epsilon)}{f_\epsilon(\epsilon)}\right\}^2\epsilon\right] \\
&\quad \times E\left(e^{-\sigma(\mathbf{X},\boldsymbol{\beta})}\mathbf{m}_{\boldsymbol{\alpha}}'(\mathbf{X},\boldsymbol{\alpha})[\boldsymbol{\sigma}_{\boldsymbol{\beta}}'(\mathbf{X},\beta) - E\{\boldsymbol{\sigma}_{\boldsymbol{\beta}}'(\mathbf{X},\boldsymbol{\beta})\}]^{\text{T}}\right) = \mathbf{0}.
\end{aligned}$$

3. $E(\boldsymbol{S}_{\text{eff},\boldsymbol{\beta}}\boldsymbol{S}_{\text{eff},\boldsymbol{\beta}}^{\text{T}})$

$$\begin{aligned}
\boldsymbol{S}_{\text{eff},\boldsymbol{\beta}} &= \left\{-\frac{f_\epsilon'(\epsilon)}{f_\epsilon(\epsilon)}\epsilon - 1\right\}\left[\boldsymbol{\sigma}_{\boldsymbol{\beta}}'(\mathbf{X},\boldsymbol{\beta}) - E\{\boldsymbol{\sigma}_{\boldsymbol{\beta}}'(\mathbf{X},\boldsymbol{\beta})\}\right] + \frac{2t}{E(t^2)}E\{\boldsymbol{\sigma}_{\boldsymbol{\beta}}'(\mathbf{X},\boldsymbol{\beta})\}, \\
\boldsymbol{S}_{\text{eff},\boldsymbol{\beta}}\boldsymbol{S}_{\text{eff},\boldsymbol{\beta}}^{\text{T}} &= \left(\frac{f_\epsilon'(\epsilon)^2}{f_\epsilon(\epsilon)^2}\epsilon^2 + 2\frac{f_\epsilon'(\epsilon)}{f_\epsilon(\epsilon)}\epsilon + 1\right)[\boldsymbol{\sigma}_{\boldsymbol{\beta}}'(\mathbf{X},\beta) - E\{\boldsymbol{\sigma}_{\boldsymbol{\beta}}'(\mathbf{X},\boldsymbol{\beta})\}] \\
&\quad \times[\boldsymbol{\sigma}_{\boldsymbol{\beta}}'(\mathbf{X},\beta) - E\{\boldsymbol{\sigma}_{\boldsymbol{\beta}}'(\mathbf{X},\boldsymbol{\beta})\}]^{\text{T}} \\
&\quad + \left\{-\frac{f_\epsilon'(\epsilon)}{f_\epsilon(\epsilon)}\epsilon - 1\right\}\frac{2t}{E(t^2)}[\boldsymbol{\sigma}_{\boldsymbol{\beta}}'(\mathbf{X},\beta) - E\{\boldsymbol{\sigma}_{\boldsymbol{\beta}}'(\mathbf{X},\boldsymbol{\beta})\}]E\{\boldsymbol{\sigma}_{\boldsymbol{\beta}}'(\mathbf{X},\boldsymbol{\beta})\}^{\text{T}} \\
&\quad + \left\{-\frac{f_\epsilon'(\epsilon)}{f_\epsilon(\epsilon)}\epsilon - 1\right\}\frac{2t}{E(t^2)}E\{\boldsymbol{\sigma}_{\boldsymbol{\beta}}'(\mathbf{X},\boldsymbol{\beta})\}[\boldsymbol{\sigma}_{\boldsymbol{\beta}}'(\mathbf{X},\beta) - E\{\boldsymbol{\sigma}_{\boldsymbol{\beta}}'(\mathbf{X},\boldsymbol{\beta})\}]^{\text{T}} \\
&\quad + \frac{4t^2}{\{E(t^2)\}^2}E\{\boldsymbol{\sigma}_{\boldsymbol{\beta}}'(\mathbf{X},\boldsymbol{\beta})\}E\{\boldsymbol{\sigma}_{\boldsymbol{\beta}}'(\mathbf{X},\boldsymbol{\beta})\}^{\text{T}}.
\end{aligned}$$

Taking expectation for the above, we have

$$
\begin{aligned}
& E(\boldsymbol{S}_{\text{eff},\boldsymbol\beta}\boldsymbol{S}_{\text{eff},\boldsymbol\beta}^{\rm T}) \\
= & E\left\{\frac{f_\epsilon'(\epsilon)^2}{f_\epsilon(\epsilon)^2}\epsilon^2+2\frac{f_\epsilon'(\epsilon)}{f_\epsilon(\epsilon)}\epsilon+1\right\} \\
& \times E\left([\boldsymbol\sigma'_{\boldsymbol\beta}(\mathbf{X},\beta)-E\{\boldsymbol\sigma'_{\boldsymbol\beta}(\mathbf{X},\boldsymbol\beta)\}][\boldsymbol\sigma'_{\boldsymbol\beta}(\mathbf{X},\beta)-E\{\boldsymbol\sigma'_{\boldsymbol\beta}(\mathbf{X},\boldsymbol\beta)\}]^{\rm T}\right) \\
& +E\left(-\frac{f_\epsilon'(\epsilon)}{f_\epsilon(\epsilon)}\epsilon t-t\right)\frac{2}{E(t^2)}E[\boldsymbol\sigma'_{\boldsymbol\beta}(\mathbf{X},\beta)-E\{\boldsymbol\sigma'_{\boldsymbol\beta}(\mathbf{X},\boldsymbol\beta)\}]E\{\boldsymbol\sigma'_{\boldsymbol\beta}(\mathbf{X},\boldsymbol\beta)\}^{\rm T} \\
& +E\left(-\frac{f_\epsilon'(\epsilon)}{f_\epsilon(\epsilon)}\epsilon t-t\right)\frac{2}{E(t^2)}E\{\boldsymbol\sigma'_{\boldsymbol\beta}(\mathbf{X},\boldsymbol\beta)\}E[\boldsymbol\sigma'_{\boldsymbol\beta}(\mathbf{X},\beta)-E\{\boldsymbol\sigma'_{\boldsymbol\beta}(\mathbf{X},\boldsymbol\beta)\}]^{\rm T} \\
& +\frac{4E(t^2)}{\{E(t^2)\}^2}E\{\boldsymbol\sigma'_{\boldsymbol\beta}(\mathbf{X},\boldsymbol\beta)\}E\{\boldsymbol\sigma'_{\boldsymbol\beta}(\mathbf{X},\boldsymbol\beta)\}^{\rm T} \\
= & E\left\{\frac{f_\epsilon'(\epsilon)^2}{f_\epsilon(\epsilon)^2}\epsilon^2-1\right\}E\{\boldsymbol\sigma'_{\boldsymbol\beta}(\mathbf{X},\boldsymbol\beta){\boldsymbol\sigma'_{\boldsymbol\beta}}^{\rm T}(\mathbf{X},\boldsymbol\beta)\} \\
& +\left[-E\left\{\frac{f_\epsilon'(\epsilon)^2}{f_\epsilon(\epsilon)^2}\epsilon^2\right\}+1+\frac{4}{E(t^2)}\right]E\{\boldsymbol\sigma'_{\boldsymbol\beta}(\mathbf{X},\boldsymbol\beta)\}E\{\boldsymbol\sigma'_{\boldsymbol\beta}(\mathbf{X},\boldsymbol\beta)^{\rm T}\}.
\end{aligned}
$$

□

S.7 Proof of Theorem 5

Under the assumption of Case 2, we have

$$
\mathbf{M}_2-\mathbf{M}_1=\begin{bmatrix}\mathbf{M}_{1,1} & \mathbf{M}_{1,2}\\ \mathbf{M}_{1,2}^{\rm T} & \mathbf{M}_{2,2}\end{bmatrix},
$$

where

$$
\begin{aligned}
\mathbf{M}_{1,1} = & E\left[\frac{f_\epsilon'(\epsilon)^2}{f_\epsilon(\epsilon)^2}-1-\frac{\{E(\epsilon^3)\}^2}{E(t^2)}\right] \\
& \times\left[E\left\{e^{-2\sigma(\mathbf{X},\boldsymbol\beta)}\mathbf{m}'_{\boldsymbol\alpha}(\mathbf{X},\boldsymbol\alpha){\mathbf{m}'}^{\rm T}_{\boldsymbol\alpha}(\mathbf{X},\boldsymbol\alpha)\right\}-E\left\{e^{-\sigma(\mathbf{X},\boldsymbol\beta)}\mathbf{m}'_{\boldsymbol\alpha}(\mathbf{X},\boldsymbol\alpha)\right\}E\left\{e^{-\sigma}(\mathbf{X},\boldsymbol\beta){\mathbf{m}'_{\boldsymbol\alpha}}^{\rm T}(\mathbf{X},\boldsymbol\alpha)\right\}\right], \\
\mathbf{M}_{1,2} = & E\left\{\frac{f_\epsilon'(\epsilon)^2}{f_\epsilon(\epsilon)^2}\epsilon+\frac{2E(\epsilon^3)}{E(t^2)}\right\} \\
& \times\left[E\left\{e^{-\sigma(\mathbf{X},\boldsymbol\beta)}\mathbf{m}'_{\boldsymbol\alpha}(\mathbf{X},\boldsymbol\alpha){\boldsymbol\sigma'}^{\rm T}_{\boldsymbol\beta}(\mathbf{X},\boldsymbol\alpha)\right\}-E\left\{e^{-\sigma(\mathbf{X},\boldsymbol\beta)}\mathbf{m}'_{\boldsymbol\alpha}(\mathbf{X},\boldsymbol\alpha)\right\}E\left\{{\boldsymbol\sigma'}^{\rm T}_{\boldsymbol\beta}(\mathbf{X},\boldsymbol\beta)\right\}\right], \\
\mathbf{M}_{2,2} = & E\left\{\frac{f_\epsilon'(\epsilon)^2}{f_\epsilon(\epsilon)^2}\epsilon^2-1-\frac{4}{E(t^2)}\right\}\left[E\left\{\boldsymbol\sigma'_{\boldsymbol\beta}(\mathbf{X},\boldsymbol\beta){\boldsymbol\sigma'}^{\rm T}_{\boldsymbol\beta}(\mathbf{X},\boldsymbol\beta)\right\}-E\left\{\boldsymbol\sigma'{}_{\boldsymbol\beta}(\mathbf{X},\boldsymbol\beta)\right\}E\left\{{\boldsymbol\sigma'}^{\rm T}_{\boldsymbol\beta}(\mathbf{X},\boldsymbol\beta)\right\}\right].
\end{aligned}
$$

From

$$\mathbf{u} = \left[\begin{array}{c} \left\{\epsilon - \frac{E(\epsilon^3)}{E(t^2)}t + \frac{f'_\epsilon(\epsilon)}{f_\epsilon(\epsilon)}\right\} \left[e^{-\sigma(\mathbf{X},\boldsymbol{\beta})}\mathbf{m}'_{\boldsymbol{\alpha}}(\mathbf{X},\boldsymbol{\alpha}) - E\{e^{-\sigma(\mathbf{X},\boldsymbol{\beta})}\mathbf{m}'_{\boldsymbol{\alpha}}(\mathbf{X},\boldsymbol{\alpha})\}\right] \\ \left\{\frac{f'_\epsilon(\epsilon)}{f_\epsilon(\epsilon)}\epsilon + 1 + \frac{2t}{E(t^2)}\right\} \left[\boldsymbol{\sigma}'_{\boldsymbol{\beta}}(\mathbf{X},\boldsymbol{\beta}) - E\{\boldsymbol{\sigma}'_{\boldsymbol{\beta}}(\mathbf{X},\boldsymbol{\beta})\}\right] \end{array} \right],$$

we have

$$\mathbf{u}\mathbf{u}^{\mathrm{T}} = \begin{pmatrix} \mathbf{u}\mathbf{u}^{\mathrm{T}}_{1,1} & \mathbf{u}\mathbf{u}^{\mathrm{T}}_{1,2} \\ \mathbf{u}\mathbf{u}^{\mathrm{T}}_{2,1} & \mathbf{u}\mathbf{u}^{\mathrm{T}}_{2,2} \end{pmatrix},$$

where

$$\begin{aligned}
\mathbf{u}\mathbf{u}^{\mathrm{T}}_{1,1} &= \left[\epsilon^2 + \frac{\{E(\epsilon^3)\}^2}{\{E(t^2)\}^2}t^2 + \frac{f'_\epsilon(\epsilon)^2}{f_\epsilon(\epsilon)^2} - \frac{2E(\epsilon^3)}{E(t^2)}\epsilon t + 2\frac{f'_\epsilon(\epsilon)}{f_\epsilon(\epsilon)}\epsilon - 2\frac{E(\epsilon^3)}{E(t^2)}\frac{f'_\epsilon(\epsilon)}{f_\epsilon(\epsilon)}t\right] \\
&\quad \times \Big[e^{-2\sigma(\mathbf{X},\boldsymbol{\beta})}\mathbf{m}'_{\boldsymbol{\alpha}}(\mathbf{X},\boldsymbol{\alpha})\mathbf{m}'^{\mathrm{T}}_{\boldsymbol{\alpha}}(\mathbf{X},\boldsymbol{\alpha}) - e^{-\sigma(\mathbf{X},\boldsymbol{\beta})}\mathbf{m}'_{\boldsymbol{\alpha}}(\mathbf{X},\boldsymbol{\alpha})E\left\{e^{-\sigma}(\mathbf{X},\boldsymbol{\beta})\mathbf{m}'_{\boldsymbol{\alpha}}{}^{\mathrm{T}}(\mathbf{X},\boldsymbol{\alpha})\right\} \\
&\quad - E\left\{e^{-\sigma(\mathbf{X},\boldsymbol{\beta})}\mathbf{m}'_{\boldsymbol{\alpha}}(\mathbf{X},\boldsymbol{\alpha})\right\}e^{-\sigma}(\mathbf{X},\boldsymbol{\beta})\mathbf{m}'_{\boldsymbol{\alpha}}{}^{\mathrm{T}}(\mathbf{X},\boldsymbol{\alpha}) \\
&\quad + E\left\{e^{-\sigma(\mathbf{X},\boldsymbol{\beta})}\mathbf{m}'_{\boldsymbol{\alpha}}(\mathbf{X},\boldsymbol{\alpha})\right\}E\left\{e^{-\sigma}(\mathbf{X},\boldsymbol{\beta})\mathbf{m}'_{\boldsymbol{\alpha}}{}^{\mathrm{T}}(\mathbf{X},\boldsymbol{\alpha})\right\}\Big], \\
\mathbf{u}\mathbf{u}^{\mathrm{T}}_{1,2} &= \Big[\frac{f'_\epsilon(\epsilon)}{f_\epsilon(\epsilon)}\epsilon^2 + \epsilon + \frac{2\epsilon t}{E(t^2)} - \frac{E(\epsilon^3)}{E(t^2)}\frac{f'_\epsilon(\epsilon)}{f_\epsilon(\epsilon)}\epsilon t - \frac{E(\epsilon^3)}{E(t^2)}t - \frac{2t^2E(\epsilon^3)}{\{E(t^2)\}^2} + \frac{f'_\epsilon(\epsilon)^2}{f_\epsilon(\epsilon)^2}\epsilon + \frac{f'_\epsilon(\epsilon)}{f_\epsilon(\epsilon)} \\
&\quad + \frac{2t}{E(t^2)}\frac{f'_\epsilon(\epsilon)}{f_\epsilon(\epsilon)}\Big] \\
&\quad \times \Big[e^{-\sigma(\mathbf{X},\boldsymbol{\beta})}\mathbf{m}'_{\boldsymbol{\alpha}}(\mathbf{X},\boldsymbol{\alpha})\boldsymbol{\sigma}'_{\boldsymbol{\beta}}{}^{\mathrm{T}}(\mathbf{X},\boldsymbol{\beta}) - e^{-\sigma(\mathbf{X},\boldsymbol{\beta})}\mathbf{m}'_{\boldsymbol{\alpha}}(\mathbf{X},\boldsymbol{\alpha})E\left\{\boldsymbol{\sigma}'_{\boldsymbol{\beta}}(\mathbf{X},\boldsymbol{\beta})\right\}^{\mathrm{T}} \\
&\quad - E\left\{e^{-\sigma}(\mathbf{X},\boldsymbol{\beta})\mathbf{m}'_{\boldsymbol{\alpha}}(\mathbf{X},\boldsymbol{\alpha})\right\}\boldsymbol{\sigma}'_{\boldsymbol{\beta}}{}^{\mathrm{T}}(\mathbf{X},\boldsymbol{\beta}) + E\left\{e^{-\sigma}(\mathbf{X},\boldsymbol{\beta})\mathbf{m}'_{\boldsymbol{\alpha}}(\mathbf{X},\boldsymbol{\alpha})\right\}E\left\{\boldsymbol{\sigma}'_{\boldsymbol{\beta}}(\mathbf{X},\boldsymbol{\beta})\right\}^{\mathrm{T}}\Big], \\
\mathbf{u}\mathbf{u}^{\mathrm{T}}_{2,1} &= \left(\mathbf{u}\mathbf{u}^{\mathrm{T}}_{1,2}\right)^{\mathrm{T}}, \\
\mathbf{u}\mathbf{u}^{\mathrm{T}}_{2,2} &= \left[\frac{f'_\epsilon(\epsilon)^2}{f_\epsilon(\epsilon)^2}\epsilon^2 + 1 + \frac{4t^2}{\{E(t^2)\}^2} + 2\frac{f'_\epsilon(\epsilon)}{f_\epsilon(\epsilon)}\epsilon + \frac{4\epsilon t}{E(t^2)}\frac{f'_\epsilon(\epsilon)}{f_\epsilon(\epsilon)} + \frac{4t}{E(t^2)}\right] \\
&\quad \times \Big[\boldsymbol{\sigma}'_{\boldsymbol{\beta}}(\mathbf{X},\boldsymbol{\beta})\boldsymbol{\sigma}'^{\mathrm{T}}_{\boldsymbol{\beta}}(\mathbf{X},\boldsymbol{\beta}) - \boldsymbol{\sigma}'_{\boldsymbol{\beta}}(\mathbf{X},\boldsymbol{\beta})E\left\{\boldsymbol{\sigma}'_{\boldsymbol{\beta}}(\mathbf{X},\boldsymbol{\beta})\right\}^{\mathrm{T}} - E\left\{\boldsymbol{\sigma}'_{\boldsymbol{\beta}}(\mathbf{X},\boldsymbol{\beta})\right\}\boldsymbol{\sigma}'^{\mathrm{T}}_{\boldsymbol{\beta}}(\mathbf{X},\boldsymbol{\beta}) \\
&\quad + E\left\{\boldsymbol{\sigma}'_{\boldsymbol{\beta}}(\mathbf{X},\boldsymbol{\beta})\right\}E\left\{\boldsymbol{\sigma}'_{\boldsymbol{\beta}}(\mathbf{X},\boldsymbol{\beta})\right\}^{\mathrm{T}}\Big].
\end{aligned}$$

It can be easily verified that $\mathbf{M}_2 - \mathbf{M}_1 = E(\mathbf{u}\mathbf{u}^{\mathrm{T}})$. Hence, $\mathbf{M}_2 - \mathbf{M}_1$ is nonnegative definite. □

S.8 Proof of Theorem 6

Under the assumption of Case 3, we have

$$\mathbf{M}_3 - \mathbf{M}_1 = \begin{bmatrix} \mathbf{M}_{1,1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix},$$

where

$$\mathbf{M}_{1,1} \quad = \quad E\left(\left[\left\{\frac{\partial f_{\epsilon|\mathbf{X}}(\epsilon,\mathbf{X})/\partial\epsilon}{f_{\epsilon|\mathbf{X}}(\epsilon,\mathbf{X})}\right\}^2 - 1\right] e^{-2\sigma(\mathbf{X},\boldsymbol{\beta})}\mathbf{m}'_{\boldsymbol{\alpha}}(\mathbf{X},\boldsymbol{\alpha})\mathbf{m}'^{\mathrm{T}}_{\boldsymbol{\alpha}}(\mathbf{X},\boldsymbol{\alpha})\right).$$

From

$$\mathbf{u} = \begin{bmatrix} \left\{\frac{\partial f_{\epsilon|\mathbf{X}}(\epsilon,\mathbf{X})/\partial\epsilon}{f_{\epsilon|\mathbf{X}}(\epsilon,\mathbf{X})} + \epsilon\right\} e^{-\sigma}(\mathbf{X},\boldsymbol{\beta})\mathbf{m}'_{\boldsymbol{\alpha}}(\mathbf{X},\boldsymbol{\alpha}) \\ \mathbf{0} \end{bmatrix},$$

we have

$$\mathbf{u}\mathbf{u}^{\mathrm{T}} = \begin{pmatrix} \mathbf{u}\mathbf{u}^{\mathrm{T}}_{1,1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix},$$

where

$$\mathbf{u}\mathbf{u}^{\mathrm{T}}_{1,1} = \left[\frac{\{\partial f_{\epsilon|\mathbf{X}}(\epsilon,\mathbf{X})/\partial\epsilon\}^2}{\{f_{\epsilon|\mathbf{X}}(\epsilon,\mathbf{X})\}^2} + 2\frac{\partial f_{\epsilon|\mathbf{X}}(\epsilon,\mathbf{X})/\partial\epsilon}{f_{\epsilon|\mathbf{X}}(\epsilon,\mathbf{X})}\epsilon + \epsilon^2\right] e^{-2\sigma(\mathbf{X},\boldsymbol{\beta})}\mathbf{m}'_{\boldsymbol{\alpha}}(\mathbf{X},\boldsymbol{\alpha})\mathbf{m}'^{\mathrm{T}}_{\boldsymbol{\alpha}}(\mathbf{X},\boldsymbol{\alpha}).$$

It is easy to verify that $\mathbf{M}_3 - \mathbf{M}_1 = E(\mathbf{u}\mathbf{u}^{\mathrm{T}})$. Thus, $\mathbf{M}_3 - \mathbf{M}_1$ is nonnegative definite. □

S.9 Proof of Theorem 7

Under the assumption of Case 4, we have

$$\mathbf{M}_4 - \mathbf{M}_2 = \begin{bmatrix} \mathbf{M}_{1,1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix},$$

where

$$\mathbf{M}_{1,1} = \left[E\left\{\frac{f'_\epsilon(\epsilon)^2}{f_\epsilon(\epsilon)^2}\right\} - 1\right] E\{e^{-\sigma(\mathbf{X},\boldsymbol{\beta})}\mathbf{m}'_{\boldsymbol{\alpha}}(\mathbf{X},\boldsymbol{\alpha})\}E\{e^{-\sigma(\mathbf{X},\boldsymbol{\beta})}\mathbf{m}'^{\mathrm{T}}_{\boldsymbol{\alpha}}(\mathbf{X},\boldsymbol{\alpha})\}.$$

From

$$\mathbf{u} = \begin{bmatrix} \left\{\frac{f'_\epsilon(\epsilon)}{f_\epsilon(\epsilon)} + \epsilon\right\} E\{e^{-\sigma(\mathbf{X},\boldsymbol{\beta})}\mathbf{m}'_{\boldsymbol{\alpha}}(\mathbf{X},\boldsymbol{\alpha})\} \\ \mathbf{0} \end{bmatrix},$$

we have

$$\mathbf{u}\mathbf{u}^{\mathrm{T}} = \begin{pmatrix} \mathbf{u}\mathbf{u}^{\mathrm{T}}_{1,1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix},$$

where

$$\mathbf{u}\mathbf{u}^{\mathrm{T}}_{1,1} = \left\{ \frac{f'_\epsilon(\epsilon)^2}{f_\epsilon(\epsilon)^2} + 2\frac{f'_\epsilon(\epsilon)}{f_\epsilon(\epsilon)}\epsilon + \epsilon^2 \right\} E\{e^{-\sigma(\mathbf{X},\boldsymbol{\beta})}\mathbf{m}'_{\boldsymbol{\alpha}}(\mathbf{X},\boldsymbol{\alpha})\} E\{e^{-\sigma(\mathbf{X},\boldsymbol{\beta})}\mathbf{m}'_{\boldsymbol{\alpha}}{}^{\mathrm{T}}(\mathbf{X},\boldsymbol{\alpha})\}.$$

It is easily obtained that $\mathbf{M}_4 - \mathbf{M}_2 = E(\mathbf{u}\mathbf{u}^{\mathrm{T}})$. Thus, $\mathbf{M}_4 - \mathbf{M}_2$ is nonnegative definite. □

S.10 Proof of Theorem 8

Under the assumption of Case 4, we have

$$\mathbf{M}_4 - \mathbf{M}_3 = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{M}_{2,2} \end{bmatrix},$$

where

$$\mathbf{M}_{2,2} = \left[E\left\{ \frac{f'_\epsilon(\epsilon)^2}{f_\epsilon(\epsilon)^2}\epsilon^2 \right\} - 1 - \frac{4}{E(t^2)} \right] \left[E\{\boldsymbol{\sigma}'_{\boldsymbol{\beta}}(\mathbf{X},\boldsymbol{\beta})\boldsymbol{\sigma}'_{\boldsymbol{\beta}}{}^{\mathrm{T}}(\mathbf{X},\boldsymbol{\beta})\} - E\{\boldsymbol{\sigma}'_{\boldsymbol{\beta}}(\mathbf{X},\boldsymbol{\beta})\} E\{\boldsymbol{\sigma}'_{\boldsymbol{\beta}}(\mathbf{X},\boldsymbol{\beta})^{\mathrm{T}}\} \right].$$

From

$$\mathbf{u} = \begin{bmatrix} \mathbf{0} \\ \left\{ \frac{f'_\epsilon(\epsilon)}{f_\epsilon(\epsilon)}\epsilon + 1 + \frac{2t}{E(t^2)} \right\} \left[\boldsymbol{\sigma}'_{\boldsymbol{\beta}}(\mathbf{X},\boldsymbol{\beta}) - E\{\boldsymbol{\sigma}'_{\boldsymbol{\beta}}(\mathbf{X},\boldsymbol{\beta})\} \right] \end{bmatrix},$$

we have

$$\mathbf{u}\mathbf{u}^{\mathrm{T}} = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{u}\mathbf{u}^{\mathrm{T}}_{2,2} \end{pmatrix},$$

where

$$\begin{aligned} \mathbf{u}\mathbf{u}^{\mathrm{T}}_{2,2} &= \left[\frac{f'_\epsilon(\epsilon)^2}{f_\epsilon(\epsilon)^2}\epsilon^2 + 1 + \frac{4t^2}{\{E(t^2)\}^2} + 2\frac{f'_\epsilon(\epsilon)}{f_\epsilon(\epsilon)}\epsilon + \frac{4\epsilon t}{E(t^2)}\frac{f'_\epsilon(\epsilon)}{f_\epsilon(\epsilon)} + \frac{4t}{E(t^2)} \right] \\ &\times \Big[\boldsymbol{\sigma}'_{\boldsymbol{\beta}}(\mathbf{X},\boldsymbol{\beta})\boldsymbol{\sigma}'^{\mathrm{T}}_{\boldsymbol{\beta}}(\mathbf{X},\boldsymbol{\beta}) - \boldsymbol{\sigma}'_{\boldsymbol{\beta}}(\mathbf{X},\boldsymbol{\beta}) E\left\{\boldsymbol{\sigma}'_{\boldsymbol{\beta}}(\mathbf{X},\boldsymbol{\beta})\right\}^{\mathrm{T}} - E\left\{\boldsymbol{\sigma}'_{\boldsymbol{\beta}}(\mathbf{X},\boldsymbol{\beta})\right\} \boldsymbol{\sigma}'^{\mathrm{T}}_{\boldsymbol{\beta}}(\mathbf{X},\boldsymbol{\beta}) \\ &+ E\left\{\boldsymbol{\sigma}'_{\boldsymbol{\beta}}(\mathbf{X},\boldsymbol{\beta})\right\} E\left\{\boldsymbol{\sigma}'_{\boldsymbol{\beta}}(\mathbf{X},\boldsymbol{\beta})\right\}^{\mathrm{T}} \Big]. \end{aligned}$$

It is easy to verify that $\mathbf{M}_4 - \mathbf{M}_3 = E(\mathbf{u}\mathbf{u}^{\mathrm{T}})$. Hence, $\mathbf{M}_4 - \mathbf{M}_3$ is nonnegative definite. □

S.11 Simulation Details

We describe in detail how to obtain the efficient estimators in the simulation studies in Section 4.1.2.

In each simulation, we generated 1000 data sets with sample size $n = 500$. For each data set, we obtain the semiparametric efficient estimator $\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}})^{\mathrm{T}}$ through solving $\sum_{i=1}^{n} \boldsymbol{S}_{\text{eff}}(\mathbf{x}_i, y_i, \boldsymbol{\theta}) = \mathbf{0}$. We then calculate $\boldsymbol{S}_{\text{eff}}(\mathbf{x}_i, y_i, \widehat{\boldsymbol{\theta}})$ and obtain the estimated variance for $\widehat{\boldsymbol{\theta}}$ through $\left[\sum_{i=1}^{n}\{\boldsymbol{S}_{\text{eff}}(\mathbf{x}_i, y_i, \widehat{\boldsymbol{\theta}})\boldsymbol{S}_{\text{eff},i}^{\mathrm{T}}(\mathbf{x}_i, y_i, \widehat{\boldsymbol{\theta}})\}\right]^{-1}$. We reported the simulation results in Tables 1 to 4, where the notations have the following meaning.

- Estimator: median of 1000 estimators
- Bias: Estimator - true $\boldsymbol{\theta}$
- Bias(%): (Bias/true $\boldsymbol{\theta}$) $\times 100$
- Var: variance of 1000 estimators
- Var1: median of 1000 estimated variances
- 95% cov: (the number of 95% confidence intervals which include true $\boldsymbol{\theta}/1000) \times 100$ (95% confidence interval is obtained as $(\widehat{\boldsymbol{\theta}} - 1.96\ \text{s.e}(\widehat{\boldsymbol{\theta}}),\ \widehat{\boldsymbol{\theta}} + 1.96\ \text{s.e}(\widehat{\boldsymbol{\theta}}))$, where s.e$(\widehat{\boldsymbol{\theta}})$ is the square root of the estimated variance described above.)
- 95% CI: (25th value, 975th value) of 1000 sorted estimators

Now we explain how to construct the efficient score $\boldsymbol{S}_{\text{eff}}(\mathbf{X}, Y)$. The efficient score functions are given in Theorems 1 to 4 in Section 2 for the four different methods. For all Cases, the efficient score functions include the terms $\mathbf{m}'_{\boldsymbol{\alpha}}(\mathbf{X}, \boldsymbol{\alpha})$ and $\boldsymbol{\sigma}'_{\boldsymbol{\beta}}(\mathbf{X}, \boldsymbol{\beta})$ which depends on the assumed model. The mean and log-standard deviation functions considered in the simulation are given in (7) and (8) respectively. From the m and σ models considered in the simulation study, we get

$$\begin{aligned}
\mathbf{m}'_{\boldsymbol{\alpha}}(\mathbf{X},\boldsymbol{\alpha}) &= \begin{bmatrix} \exp(\alpha_1 X_1 + \alpha_2 X_2) \\ X_1 \exp(\alpha_1 X_1 + \alpha_2 X_2) \\ X_2 \exp(\alpha_1 X_1 + \alpha_2 X_2) \end{bmatrix}, \\
\boldsymbol{\sigma}'_{\boldsymbol{\beta}}(\mathbf{X},\boldsymbol{\beta}) &= \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}.
\end{aligned}$$

We then plug the above expressions into the efficient score function $\boldsymbol{S}_{\text{eff}}(\mathbf{X}, Y)$ in each case as described below.

S.11.1 Case 1 method

The semiparametric efficient estimator is obtained thorough solving the efficient score function (3). The efficient score function (3) contains terms $E(\epsilon^3|\mathbf{X})$ and $E(t^2|\mathbf{X})$, which need to be calculated based on the distribution of ϵ conditional $\mathbf{X}$.

Since Case 1 method is based on the possibly asymmetric $f_{\epsilon|\mathbf{X}}(\epsilon, \mathbf{X})$ where ϵ and $\mathbf{X}$ are not independent, we choose such an model $f_{\epsilon|\mathbf{X}}(\epsilon, \mathbf{X})$ for estimation, regardless of data generation scheme. Among many possible distributions, we chose a chisquared distribution with degree of freedom $p(\mathbf{X})$ which is given in (9), even for the simulation where the ϵ is not generated from the density function (9). Note that ϵ was not generated from the density (9) in Simulation 2, 3, and 4. Then we calculate $E(\epsilon^3|\mathbf{X})$, $E(\epsilon^4|\mathbf{X})$ and $E(t^2|\mathbf{X})$ as the following.

$$\begin{aligned}
E(\epsilon^3|\mathbf{X}) &= \frac{2\sqrt{2}}{\sqrt{p(\mathbf{X})}}, \\
E(\epsilon^4|\mathbf{X}) &= 3 + \frac{12}{p(\mathbf{X})}, \\
E(t^2|\mathbf{X}) &= E(\epsilon^4|\mathbf{X}) - \left\{E(\epsilon^3|\mathbf{X})\right\}^2 - 1 = 2 + \frac{4}{p(\mathbf{X})}.
\end{aligned}$$

Because each simulation generates different data according to its assumption, we use different degree of freedom $p(\mathbf{X})$ for estimation as follows.

1. Simulation 1, 2 and 4

   We used the chisquared distribution of ϵ with degree of freedom $p(\mathbf{X}) = (X_1+X_2)+0.5$.

The calculated $E(\epsilon^3|\mathbf{X})$ and $E(t^2|\mathbf{X})$ according $p(\mathbf{X}) = (X_1 + X_2) + 0.5$ are plugged into the efficient score function.

2. Simulation 3

   We used the chisquared distribution of ϵ with degree of freedom $p(\mathbf{X}) = 10(X_1+X_2+1)$. The calculated $E(\epsilon^3|\mathbf{X})$ and $E(t^2|\mathbf{X})$ according $p(\mathbf{X}) = 10(X_1 + X_2 + 1)$ are plugged into the efficient score function.

S.11.2 Case 2 method

The efficient score function of Case 2 method is given in (4). To solve the efficient score function (4), we should calculate terms $E(\epsilon^3)$, $E(t^2)$ and $f'_\epsilon(\epsilon)/f_\epsilon(\epsilon)$ in (4). Case 2 method is based on possibly asymmetric $f_\epsilon(\epsilon)$. For such an asymmetric $f_\epsilon(\epsilon)$, we used a chisquared distribution with degree of freedom q which is given in (10) for estimation, even for the simulation where the ϵ is not generated from the chisquared distribution (10). Note that ϵ was not generated from (10) in Simulation 1, 3, and 4. According the standardized $\chi^2(q)$, we calculate the following quantities.

$$
\begin{aligned}
E(\epsilon^3) &= \frac{2\sqrt{2}}{\sqrt{q}}, \\
E(\epsilon^4) &= 3 + \frac{12}{q}, \\
E(t^2) &= E(\epsilon^4) - \left\{E(\epsilon^3)\right\}^2 - 1 = 2 + \frac{4}{q}, \\
\frac{f'_\epsilon(\epsilon)}{f_\epsilon(\epsilon)} &= \frac{\sqrt{2q}(q/2-1)}{\sqrt{2q}\epsilon + q} - \frac{\sqrt{2q}}{2}.
\end{aligned}
$$

Because each simulation generates different data according to its assumption, we use different degrees of freedom q for estimation as the following.

- Simulation 1 and 2

  We used the chisquared distribution ϵ with degree of freedom as $q = 13$. We plug $E(\epsilon^3) = 0.7845$, $E(t^2) = 2.3077$ and $f'_\epsilon(\epsilon)/f_\epsilon(\epsilon) = 11\sqrt{26}/(2\sqrt{26}\epsilon + 26) - \sqrt{26}/2$ into the efficient score function.

- Simulation 3

  We used the chisquared distribution ϵ with degree of freedom as $q = 21$. We plug

$E(\epsilon^3) = 0.6172$, $E(t^2) = 2.1905$ and $f'_\epsilon(\epsilon)/f_\epsilon(\epsilon) = 19\sqrt{42}/(2\sqrt{42}\epsilon + 42) - \sqrt{42}/2$ into the efficient score function.

- Simulation 4

  We used the chisquared distribution ϵ with degree of freedom as $q = 180$. We plug $E(\epsilon^3) = 0.2108$, $E(t^2) = 2.0222$ and $f'_\epsilon(\epsilon)/f_\epsilon(\epsilon) = 89\sqrt{10}/(\sqrt{10}\epsilon + 30) - 3\sqrt{10}$ into the efficient score function.

S.11.3 Case 3 Method

The efficient score function of Case 3 method is given in (5). To solve the efficient score function (5), we should calculate $E(t^2|\mathbf{X})$ and $f'_{\epsilon|\mathbf{X}}(\epsilon, \mathbf{X})/f_{\epsilon|\mathbf{X}}(\epsilon, \mathbf{X})$ in (5). Case 3 method is based on the symmetric $f_{\epsilon|\mathbf{X}}(\epsilon, \mathbf{X})$ conditional on $\mathbf{X}$. For estimation, we chose a generalized normal distribution with scale parameter $s = \sqrt{\Gamma(1/u(\mathbf{X}))/\Gamma(3/u(\mathbf{X}))}$ and shape parameter $k = u(\mathbf{X})$ which is given by

$$f_{\epsilon|\mathbf{X}}(\epsilon, \mathbf{X}) = \frac{u(\mathbf{X})}{2\Gamma(1/u(\mathbf{X}))\sqrt{\Gamma(1/u(\mathbf{X}))/\Gamma(3/u(\mathbf{X}))}} \exp\left[-\left\{\frac{|\epsilon|}{\sqrt{\Gamma(1/u(\mathbf{X}))/\Gamma(3/u(\mathbf{X}))}}\right\}^{u(\mathbf{X})}\right].$$

From the above distribution, we obtain

$$E(\epsilon^4|\mathbf{X}) = \frac{\Gamma(5/u(\mathbf{X}))\Gamma(1/u(\mathbf{X}))}{\Gamma(3/u(\mathbf{X}))^2}, \tag{S.2}$$

$$E(t^2|\mathbf{X}) = E(\epsilon^4|\mathbf{X}) - 1, \tag{S.3}$$

$$\frac{\partial f_{\epsilon|\mathbf{X}}(\epsilon, \mathbf{X})/\partial\epsilon}{f_{\epsilon|\mathbf{X}}(\epsilon, \mathbf{X})} = (-1)^{-\text{sgn}(\epsilon)}|\epsilon|^{u(\mathbf{X})-1}u(\mathbf{X})\left\{\frac{\Gamma(3/u(\mathbf{X}))}{\Gamma(1/u(\mathbf{X}))}\right\}^{u(\mathbf{X})/2}, \tag{S.4}$$

where $\text{sgn}(\epsilon)$ is a signum function of ϵ. Note that $u(\mathbf{X})$ determines both the scale parameter and the shape parameter of a generalized normal distribution. Also, note that ϵ was not generated from the generalized normal distribution in Simulation 1, 2, and 4.

Because each simulation generates different data according to its assumption, we use different parameter for estimation of each simulation as follows.

- Simulation 1, 2 and 4

  We choose $u(\mathbf{X}) = 0.06(X_1 + X_2) + 1.5$ which varies in [1.5,3]. According to $u(\mathbf{X})$, the calculated $E(t^2|\mathbf{X})$ and $f'_{\epsilon|\mathbf{X}}(\epsilon, \mathbf{X})/f_{\epsilon|\mathbf{X}}(\epsilon, \mathbf{X})$ are plugged into the efficient score function.

- Simulation 3

  We choose $u(\mathbf{X}) = 20I(\mathbf{X} \in A) + 1.7I(\mathbf{X} \in A^c)$ where $A = \{(0,5) \times (0,7.5), (5,10) \times (7.5,15)\}$ and $I(\cdot)$ is an indicator function. Using (S.2), (S.3) and (S.4), we can calculate $E(t^2|\mathbf{X})$ and $f'_{\epsilon|\mathbf{X}}(\epsilon,\mathbf{X})/f_{\epsilon|\mathbf{X}}(\epsilon,\mathbf{X})$ conditional on $\mathbf{X}$ for the above distributions and plug them into the efficient score functions.

S.11.4 Case 4 Method

The efficient score function of Case 3 method is given in (6). To construct the efficient score function, we should calculate $E(t^2)$ and $f'_\epsilon(\epsilon)/f_\epsilon(\epsilon)$.

- Simulation 1, 2 and 4

  We use Logistic$(0, \sqrt{3}/\pi)$ distribution for symmetric $f_\epsilon(\epsilon)$. Then we obtained

$$E(\epsilon^4) = 4.2, \quad \frac{f'_\epsilon(\epsilon)}{f_\epsilon(\epsilon)} = -\frac{\pi}{\sqrt{3}} + \frac{2\exp\left(-\epsilon\pi/\sqrt{3}\right)}{\sqrt{3}\{1+\exp(-\epsilon\pi/\sqrt{3})\}/\pi}.$$

  We plug the above terms in the efficient score function.

- Simulation 3

  For Simulation 3, we choose a GN$(0, s, k)$ for a symmetric $f_\epsilon(\epsilon)$, where $s = \sqrt{\Gamma(1/k)/\Gamma(3/k)}$ and $k = 5.4$. We calculate

$$\begin{aligned} E(\epsilon^4) &= \frac{\Gamma(5/k)\Gamma(1/k)}{\Gamma(3/k)^2}, \\ E(t^2) &= E(\epsilon^4) - 1, \\ \frac{f'_\epsilon(\epsilon)}{f_\epsilon(\epsilon)} &= (-1)^{-\text{sgn}(\epsilon)}|\epsilon|^{k-1}k\left\{\frac{\Gamma(3/k)}{\Gamma(1/k)}\right\}^{k/2}. \end{aligned}$$

  The above terms are plugged in the efficient score function.

References

Tsiatis, A. (2006), *Semiparametric Theory and Missing Data*, New York: Springer.