Semi-parametric estimation of potential outcome distributions and general causal estimands by borrowing information from both treatments and controls

Manli Cheng[1,2] and Yukun Liu[2]

[1] *University of Waterloo and* [2] *East China Normal University*

Supplementary Material

This supplementary material consists of ten sections. Section S1 presents technical preliminaries. Sections S2- S6 contain proofs of Theorems 1 and 2, Corollary 1, and Theorems 3 and 4, respectively. Section S7 proposes a testing procedure to evaluate the adequacy of the proposed semiparametric likelihood ratio model (SPLRM). Section S8 contains additional simulation results and also investigates the finite-sample agreement between the bootstrap variance estimates and the corresponding theoretical variances. Section S9 presents additional analysis results about the LLvsPSID dataset. Finally, Section S10 presents additional simulation results for the estimation of conditional average treatment effects (CATE).

S1 Preparation

Before presenting the proofs of the theoretical results stated in the main paper, we first introduce some model-related notations and preliminary results to facilitate the subsequent analysis.

Denote $Z = (Y, X, D)$, recall that the likelihood function of F and β based on the observations $\{Z_i, 1 \leqslant i \leqslant n\}$ is

$$L_n(F, \beta) = \prod_{i=1}^{n} \left\{ \frac{e^{(X_i^\top \beta_1)Y_i} dF(Y_i)}{\int e^{(X_i^\top \beta_1)t} dF(t)} \right\}^{D_i} \left\{ \frac{e^{(X_i^\top \beta_0)Y_i} dF(Y_i)}{\int e^{(X_i^\top \beta_0)t} dF(t)} \right\}^{1-D_i}. \tag{S1.1}$$

Let $\mathbb{P}_n$ be the empirical measure induced by the data-set $\{Z_i : 1 \leqslant i \leqslant n\}$, and denote P_0 the probability measure of Z. Let Q_n be the empirical measure induced by the data-set $\{(X_i, D_i) : 1 \leqslant i \leqslant n\}$, and denote Q_0 the true probability measure of (X, D). In general, for a function g of Z and an estimator $\hat{\xi}$, we use $\mathbb{P}_n g(Z; \hat{\xi})$ to denote $\mathbb{P}_n g(Z; \xi)$ evaluated at $\xi = \hat{\xi}$. We write $P_0 g(Z; \hat{\xi})$ likewise. Let $\mathbb{G}_n = \sqrt{n}(\mathbb{P}_n - P_0)$. Throughout the supplementary material, we use the notations $F_{Y(k)|X}(\cdot)$ and $F_{(k)|X}(\cdot)$ interchangeably, as well as $F_{Y(k)}(\cdot)$ and $F_{(k)}(\cdot)$, without further distinction.

Denote the log-likelihood function based on an ideal observation $Z = (Y, X, D)$ by

$$\begin{aligned} \ell(F, \beta) \quad = \quad & D \cdot \left[Y \cdot \beta_1^\top X + \ln dF(Y) - \ln \int e^{y \cdot \beta_1^\top X} dF(y) \right] \\ & + (1 - D) \left[Y \cdot \beta_0^\top X + \ln dF(Y) - \ln \int e^{y \cdot \beta_0^\top X} dF(y) \right]. \end{aligned} \tag{S1.2}$$

For fixed (F, β), define a path, through F indexed by ε, as $dF_\varepsilon = (1 + \varepsilon h_1) dF$, where h_1 is an arbitrary nonnegative bounded function. The derivative of the log-likelihood with respect to ε along the path evaluated $\varepsilon = 0$ yields a score function in the direction h_1, i.e.

$$\begin{aligned}\ell_{1,F,\beta}[h_1](Z) &= (1-D)\cdot\left\{h_1(Y)-\frac{\int\exp(y\cdot\beta_0^\top X)h_1(y)dF(y)}{\int\exp(y\cdot\beta_0^\top X)dF(y)}\right\}\\ &\quad +D\cdot\left\{h_1(Y)-\frac{\int\exp(y\cdot\beta_1^\top X)h_1(y)dF(y)}{\int\exp(y\cdot\beta_1^\top X)dF(y)}\right\}. \qquad \text{(S1.3)}\end{aligned}$$

Similarly, define a path through β, index by ε, as $\beta_\varepsilon=\beta+\varepsilon h_2$, where $h_2=(h_{21}^\top,h_{22}^\top)^\top$ and h_2 is an arbitrary vector in $\mathbb{R}^d$. Differentiation of the log-likelihood with respect to ε along this path and evaluation at $\varepsilon=0$ yield the score function for β in the direction h_2:

$$\begin{aligned}h_2^\top\ell_{2,F,\beta}(Z) &= h_2^\top\cdot\begin{pmatrix}(1-D)\left(Y-\frac{\int y\exp(y\cdot\beta_0^\top X)dF(y)}{\int\exp(y\cdot\beta_0^\top X)dF(y)}\right)\\ D\left(Y-\frac{\int y\exp(y\cdot\beta_1^\top X)dF(y)}{\int\exp(y\cdot\beta_1^\top X)dF(y)}\right)\end{pmatrix}\\ &= h_{21}^\top X\left\{(1-D)\cdot\left(Y-\frac{\int y\exp(y\cdot\beta_0^\top X)dF(y)}{\int\exp(y\cdot\beta_0^\top X)dF(y)}\right)\right\}\\ &\quad +h_{22}^\top X\left\{D\cdot\left(Y-\frac{\int y\exp(y\cdot\beta_1^\top X)dF(y)}{\int\exp(y\cdot\beta_1^\top X)dF(y)}\right)\right\}. \qquad \text{(S1.4)}\end{aligned}$$

Along any direction $h=(h_1(\cdot),h_2)$, define the score of $\ell(F,\beta)$ with respect to F and β like

$$S_{F,\beta}[h](Z)=\ell_{1,F,\beta}[h_1](Z)+h_2^\top\ell_{2,F,\beta}(Z). \qquad \text{(S1.5)}$$

Because $(\hat{F},\hat{\beta})$ is the maximizer of the empirical log-likelihood $\ell_n(F,\beta)=\sum_{i=1}^n \text{loglik}(Z_i;F,\beta)$, where

$$\text{loglik}(Z_i;F,\beta) = D_i\cdot\left[Y_i\cdot\beta_1^\top X_i+\ln dF(Y_i)-\ln\int e^{y\cdot\beta_1^\top X_i}dF(y)\right]$$

$$+ (1 - D_i)\left[Y \cdot \beta_0^\top X_i + \ln dF(Y_i) - \ln \int e^{y \cdot \beta_0^\top X_i} dF(y)\right]\right\}.$$

It then follows that

$$0 = \mathbb{P}_n\{S_{\hat{F},\hat{\beta}}[h](Z)\} = P_0\{S_{F,\beta}[h](Z)\}.$$

Let $A_{kg}(u; F) = \int y^k g(y) \exp(u \cdot y) dF(y)$ and we write $A_{k1}(u; F)$ when $g(y) \equiv 1$, where $k = 0$ or 1 and g is some function. Since the boundness assumption of the support X and Y, we can easily get that $A_{kg}(\beta_k^\top X; F)$ is continuous, even uniformly continuous, with respect to β_k. Denote the information operator by $\sigma(h) = (\sigma_1(h), \sigma_2(h))$, where $\sigma_1(h)$ is a bounded function of bounded variation and $\sigma_2(h)$ is a $(2d)$-dimensional vector. The operator σ under our SPLRM in the main paper is defined through

$$P_0\left\{\ell_{1,F^*,\beta^*}[h_1](Z) + h_2^\top \ell_{2,F^*,\beta^*}(Z)\right\}^2 = \langle \sigma[h], h\rangle = \int \sigma_1(h)(y) h_1(y) dF^* + h_2^\top \sigma_2(h), \forall h, \tag{S1.6}$$

where

$$\begin{aligned}
\sigma_1(h)(y) \quad = \quad & Q_0\left\{(1 - D)\left[h_1(y)\frac{e^{\beta_0^{*\top} X \cdot y}}{A_{01}(\beta_0^{*\top} X; F^*)} - \frac{A_{0h_1}(\beta_0^{*\top} X; F^*) e^{\beta_0^{*\top} X \cdot y}}{A_{01}^2(\beta_0^{*\top} X; F^*)}\right.\right. \\
& \left.+ \; y h_{21}^\top X \frac{e^{\beta_0^{*\top} X \cdot y}}{A_{01}(\beta_0^\top X; F^*)} - h_{21}^\top X \frac{e^{\beta_0^\top X \cdot y} A_{11}(\beta_0^\top X; F^*)}{A_{01}^2(\beta_0^{*\top} X; F^*)}\right]
\end{aligned}$$

$$+D\left[h_1(y)\frac{e^{\beta_1^{*\top}X\cdot y}}{A_{01}(\beta_1^{*\top}X;F^*)}-\frac{A_{0h_1}(\beta_1^{*\top}X;F^*)e^{\beta_1^{*\top}X\cdot y}}{A_{01}^2(\beta_1^{*\top}X;F^*)}\right.$$
$$\left.+\ yh_{22}^\top X\frac{e^{\beta_1^{*\top}X\cdot y}}{A_{01}(\beta_1^{*\top}X;F^*)}-h_{22}^\top X\frac{e^{\beta_1^{*\top}X\cdot y}A_{11}(\beta_1^{*\top}X;F^*)}{A_{01}^2(\beta_1^{*\top}X;F^*)}\right]\Bigg\}$$

and

$$\begin{aligned}
&\sigma_2(h)\\
&= Q_0\begin{pmatrix} (1-D)\left[\left(\frac{A_{21}(\beta_0^{*\top}X;F^*)}{A_{01}(\beta_0^{*\top}X;F^*)}-\frac{A_{11}^2(\beta_0^{*\top}X;F^*)}{A_{01}^2(\beta_0^{*\top}X;F^*)}\right)XX^\top h_{21}\right]\\ D\left[\left(\frac{A_{21}(\beta_1^{*\top}X;F^*)}{A_{01}(\beta_1^{*\top}X;F^*)}-\frac{A_{11}^2(\beta_1^{*\top}X;F^*)}{A_{01}^2(\beta_1^{*\top}X;F^*)}\right)XX^\top h_{22}\right]\end{pmatrix}\\
&\quad +Q_0\begin{pmatrix} (1-D)\left[\left(\frac{A_{1h_1}(\beta_0^{*\top}X;F^*)}{A_{01}(\beta_0^{*\top}X;F^*)}-\frac{A_{11}(\beta_0^{*\top}X;F^*)A_{0h_1}(\beta_0^{*\top}X;F^*)}{A_{01}^2(\beta_0^{*\top}X;F^*)}\right)X\right]\\ D\left[\left(\frac{A_{1h_1}(\beta_1^{*\top}X;F^*)}{A_{01}(\beta_1^{*\top}X;F^*)}-\frac{A_{11}(\beta_1^{*\top}X;F^*)A_{0h_1}(\beta_1^{*\top}X;F^*)}{A_{01}^2(\beta_1^{*\top}X;F^*)}\right)X\right]\end{pmatrix}.
\end{aligned}$$

The information operator $\sigma : BV_M \times R^{2d} \to BV_M \times R^{2d}$ is onto and continuously invertible following from a lemma stated by Murphy et al. (1997) and Luo and Tsai (2012). This conclusion is also valid under our model settings. Denote its inverse by $\tilde{\sigma} : BV_M \times R^{2d} \to BV_M \times R^{2d}$ with the first component a function in BV and the second component a d-dimensional vector.

Similar to the conclusion in the proof of Theorem 3 in Luo and Tsai (2012), we can also prove the linear approximation of $\hat{F}$ and $\hat{\beta}$, that is, for any h, setting $h_2^* = \tilde{\sigma}_2(h)$ and $h_1^* = \tilde{\sigma}_1(h)$, we have

$$\begin{aligned}
\sqrt{n}\{\int h_1 d(\hat{F}-F^*) + h_2^\top(\hat{\beta}-\beta^*)\} &= \mathbb{G}_n\{\ell_{1,F^*,\beta^*}(Z)[h_1^*] + (h_2^*)^\top \ell_{2,F^*,\beta^*}(Z)\} + o_p(1)\\
&= \mathbb{G}_n\{S_{F^*,\beta^*}[\tilde{\sigma}(h)](Z)\} + o_p(1), \qquad \text{(S1.7)}
\end{aligned}$$

where $\tilde{\sigma}[h](\cdot) = (\tilde{\sigma}_1[h](\cdot), \tilde{\sigma}_2[h])$ is the primary image of h under the operator σ.

S2 Proof of Theorem 1 in main paper

In order to complete the proof of Theorem 1, we first introduce the following important lemmas, which will simplify our proof procedure. For convenience of notation, we rewrite the support set of Y as $\mathcal{Y} = [l, u]$ and define $\mathcal{Y}\mathcal{X} = \{(y, x) : y \in \mathcal{Y}, x \in \mathcal{X}\}$. Let $\ell^\infty(\mathcal{Y}\mathcal{X})$ be the set of all bounded and measurable mappings $\mathcal{Y}\mathcal{X} \to \mathbb{R}$. We consider $\mathcal{Y}\mathcal{X}$ as a subset of $\bar{\mathbb{R}}^{1+d}$, with relative topology and let ρ denote a standard metric on $\bar{\mathbb{R}}^{1+d}$, where $\bar{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$.

Lemma 1 (Hadamard differentiability of potential operator)**.** *For any functions* g_1 *and* g_2*, Define their semi-metric as* $\delta(\mathrm{g}_1, \mathrm{g}_2) = \left[P_0(\mathrm{g}_1 - \mathrm{g}_2)^2\right]^{1/2}$*. Let* $\mathcal{G}_k$ *be the class of bounded functions that contains* $\left\{F_{Y(k)|X}(y|\cdot) : y \in \mathcal{Y}\right\}$ *as well as the indicators of all the rectangles in* $\bar{\mathbb{R}}^p$*, such that* $\mathcal{G}_k$ *is totally bounded under* δ*,* $k \in \{0, 1\}$*. Let* $\mathbb{D}_\phi$ *be the product of the space of measureable functions* $\Gamma : \mathcal{Y}\mathcal{X} \mapsto [0, 1]$ *defined by* $(y, x) \mapsto \Gamma(y, x)$ *and the bounded maps* $\Pi : \mathcal{G}_k \mapsto \mathbb{R}$ *defined by* $f \mapsto \int f d\Pi$*, where* Π *is restricted to be a probability measure on* $\mathcal{X}$*. Consider the map* $\phi : \mathbb{D}_\phi \subset \ell^\infty(\mathcal{Y}\mathcal{X}) \times \ell^\infty(\mathcal{G}_k) \mapsto \ell^\infty(\mathcal{Y})$*, defined by*

$$(\Gamma, \Pi) \mapsto \phi(\Gamma, \Pi) = \int \Gamma(\cdot, x) d\Pi(x).$$

Then for $k = 0$ *and* 1*, the map* ϕ *is well defined. Moreover, the map* ϕ *is Hadamard-*

differentiable at $(\Gamma, \Pi) = (F_{Y(k)|X}, F_X)$ *with the derivativ map* $(\gamma, \pi) \mapsto \phi'_{F_{Y(k)|X}, F_X}(\gamma, \pi)$ *mapping* $\ell^\infty(\mathcal{YX}) \times \ell^\infty(\mathcal{G}_k)$ *to* $\ell^\infty(\mathcal{Y})$ *defined by* $\phi'_{F_{Y(k)|X}, F_X}(\gamma, \pi)(y) = \int \gamma(y, x) dF_X(x) + \pi(F_{Y(k)|X}(y|\cdot))$*, which is continuous on* $\ell^\infty(\mathcal{YX}) \times \ell^\infty(\mathcal{G}_k)$*.*

Lemma 1 is similar to Lemma D.1 in Chernozhukov et al. (2013), so we omit its proof here. Next, we will provide a formal proof of Theorem 1 in main paper.

Proof. It is easy to find that for any g $\in \mathcal{G}_k$ in Lemma 1, the functional central limit theorem implies $\mathbb{G}_n(\mathrm{g}) \xrightarrow{d} \mathbb{G}(\mathrm{g})$ under our SPLRM model and Assumption 2 in the main paper, where $\mathbb{G}$ is a zero-mean tight Gaussian process. Recall the definition of $F_{Y(k)|X}(y|x)$, a function of $F(y)$ and β, so our plg-in estimate $\hat{F}_{Y(k)|X}(y|x)$ is a function of $\hat{F}$ and $\hat{\beta}$. Applying the delta method along the direction $\hat{h}_k = (\hat{F} - F^*, \hat{\beta}_k - \beta_k^*)$ to linearly approximate $\hat{F}_{Y(k)|X}(y|x)$ for $k = 0$ or 1 and some (y, x), we get

$$\hat{F}_{(k)}(y|x) \quad = \quad F_{(k)}(y|x) + \int S_F(u; y, x, F^*, \beta^*, k) d(\hat{F}(u) - F^*(u)) + S_{\beta_k}^\top(x, F^*, \beta^*, k)(\hat{\beta}_k - \beta_k^*),$$

where

$$\begin{aligned} S_F(u; y, x, F^*, \beta^*, k) &= \frac{e^{\beta_k^{*\top} x \cdot u}}{A_{01}(\beta_k^{*\top} x; F^*)} \cdot \left\{ I_{(-\infty, y]}(u) - \frac{A_{0I_y}(\beta_k^{*\top} x; F^*)}{A_{01}(\beta_k^{*\top} x; F^*)} \right\}, \\ S_{\beta_k}(x, F^*, \beta^*, k) &= x \cdot \left[\frac{A_{0I_y}(\beta_k^{*\top} x; F^*)}{A_{01}(\beta_k^{*\top} x; F^*)} - \frac{A_{0I_y}(\beta_k^{*\top} x; F^*) A_{11}(\beta_k^{*\top} x; F^*)}{A_{01}^2(\beta_k^{*\top} x; F^*)} \right] \end{aligned}$$

and $I_y = I_{(-\infty, y]}$. We denote $S_\beta^\top(x, F, \beta, k) = (k S_{\beta_k}^\top(x, F, \beta, k), (1-k) S_{\beta_k}^\top(x, F, \beta, k))$

, then $\hat{F}_{(k)}(y|x)$ can be re-writen as

$$\hat{F}_{(k)}(y|x) \quad = \quad F_{(k)}(y|x) + \int S_F(u; y, x, F^*, \beta^*, k) d(\hat{F}(u) - F^*(u)) + S_\beta^\top(x, F^*, \beta^*, k)(\hat{\beta} - \beta^*).$$

Let $S_{F,\beta}^{y,x,k} = (S_F(u; y, x, F, \beta, k), S_\beta^\top(x, F, \beta, k))^\top$, the conclusion in (S1.7) gives

$$\begin{aligned}\hat{Z}_{nk}(y, x) &= \sqrt{n}(\hat{F}_{Y(k)|X}(y|x) - F_{Y(k)|X}(y|x)) \\ &= \mathbb{G}_n \left\{ \ell_{1,F^*,\beta^*}(Z)[\tilde{\sigma}_1[S_{F^*,\beta^*}^{y,x,k}]] + \tilde{\sigma}_2[S_{F^*,\beta^*}^{y,x,k}]^\top \ell_{2,F^*,\beta^*}(Z) \right\} + o_p(1) \\ &\xrightarrow{d} \mathbb{G} \left\{ \ell_{1,F^*,\beta^*}(Z)[\tilde{\sigma}_1[S_{F^*,\beta^*}^{y,x,k}]] + \tilde{\sigma}_2[S_{F^*,\beta^*}^{y,x,k}]^\top \ell_{2,F^*,\beta^*}(Z) \right\} := Z_k(y, x),\end{aligned} \tag{S2.8}$$

where $Z_k(y, x)$ is a zero-mean tight Gaussian process with a.s. uniformly continuous paths with respect to the semi-metric ρ by the functional central limit theorem.

Recall the definition of ϕ in Lemmma 1, we re-write $F_{Y(k)}(y) = \phi(F_{Y(k)|X}, F_X)$ and $\hat{F}_{Y(k)}(y) = \phi(\hat{F}_{Y(k)|X}, \hat{F}_X)$. Throughout the article, we estimate the marginal distribution F_X of the covariable X using the empirical distribution function,that is $\hat{F}_X(x) = \sum_{i=1}^n I(x_i \leqslant x)/n = \sum_{k=0}^1 \sum_{i=1}^{n_k} I(x_{ki} \leqslant x)$. Lemmma 1 established that map ϕ (potential operator) is Hadamard differentiable at $(F_{Y(k)|X}, F_X)$. By the Functional Delta Method, it follows that, for $\gamma(y, x) = \hat{F}_{Y(k)|X} - F_{Y(k)|X}$ and $\pi(x) = \hat{F}_X - F_X$,

$$\begin{aligned}&\sqrt{n}(\hat{F}_{Y(k)}(y) - F_{Y(k)}(y)) \\ = \quad &\sqrt{n}(\phi(\hat{F}_{Y(k)|X}, \hat{F}_X) - \phi(F_{Y(k)|X}, F_X)) \\ = \quad &\sqrt{n}\phi'_{F_{Y(k)|X}, F_X}(\gamma(y, x), \pi(x)) + \sqrt{n} o_P(||\gamma(y, x)||_\infty + ||\pi(x)||_\infty),\end{aligned}$$

$$
\begin{aligned}
&= \sqrt{n}\left\{\int \hat{F}_{Y(k)|X} - F_{Y(k)|X}dF_X + \pi(F_{Y(k)|X})\right\} + o_p(1) \\
&= \int \sqrt{n}(\hat{F}_{Y(k)|X} - F_{Y(k)|X})dF_X + \int F_{Y(k)|X}(y|x)\sqrt{n}d(\hat{F}_X - F_X) + o_p(1) \\
&= \int \sqrt{n}(\hat{F}_{Y(k)|X} - F_{Y(k)|X})dF_X + \mathbb{G}_n\left\{F_{Y(k)|X}\right\} + o_p(1) \\
&\xrightarrow{d} \int Z_k(y,x)dF_X + \mathbb{G}(F_{Y(k)|X}) := Z_k(y),
\end{aligned}
$$

where the third equation holds because we have proved that $\sqrt{n}(\hat{F}_{Y(k)|X} - F_{Y(k)|X})$ and $\sqrt{n}(\hat{F}_X - F_X)$ respectively converge weakly to tight Gaussian processes with zero mean and a.s uniformly continuous sample path. Therefore, the limiting process $Z_k(y) \in \ell^\infty(\mathcal{Y})$ is also a zero-mean tight Gaussian process with a.s. uniformly continuous sample path. □

S3 Proof of Theorem 2 in main paper

Proof. Since the boundness of the support X and Y, we can easily get that $A_{01}(\beta_k^\top X; F)$ is continuous, even uniformly continuous, with respect to β_k. For fixed β_k, considering two functions that are very close together, F_1 and F_2, in $\mathcal{F}$. When the distribution function F is a step function, let's assume that F only has a jump at some y_0 for simplicity. Assuming $dF_2(y_0) = dF_1(y_0) + \Delta$, where Δ arbitrarily small. Then, the difference $\Delta A_{01}(\beta_k^\top X; \cdot) = \exp(\beta_k^T X \cdot y_0)\Delta$ is arbitrarily small. If F is a continuous distribution function, $|\Delta A_{01}(\beta_k^\top X; \cdot)| \leqslant \int(\beta_k^T X \cdot y)|d(F_2(y) - F_1(y))| \leqslant M \int |d(F_2(y) - F_1(y))|$, where $\int |d(F_2(y) - F_1(y))|$ represents the variation of the function $F_2(y) - F_1(y)$. When the distance between F_1 and F_2 approaches 0 , the variation of $F_2(y) - F_1(y)$ will also approach 0. $A(\beta_k^\top X)$ actually is identical to $A_{01}(\beta_k^\top X; F)$. Therefore, we have proven the continuity of $A(\beta_k^\top X)$ with respect to β_k and F, regardless of whether F is a continuous distribution function or a step function. $\int e^{\beta_k^\top X y} dF(y)$ is the moment generating function corresponding to the distribution function F, and it is strictly away from 0. Therefore, $\hat{\tau}(x)$ is continuous with respect to F and β. Combining the conclusions in Theorem 1, the Continuous Mapping Theorem directly implies the uniform consistency of $\hat{\tau}(x)$.

Below, we give a proof of the asymptotic normality of $\hat{\tau}(x)$ for fixed x. Applying the delta method along the direction $\hat{h}^\top = (\hat{F} - F^*, \hat{\beta}^\top - \beta^{*\top})$ to $\hat{\tau}_1(x)$, we have

$$\begin{aligned}
\sqrt{n}(\hat{\tau}_1(x) - \tau_1(x)) &= \sqrt{n} K_1(\hat{h}; x) + o_p(\sqrt{n}\hat{h}) \\
&= \sqrt{n} \left\{ \int K_{11}(y, x) d(\hat{F}(y) - F^*(y)) + K_{12}(x)^\top (\hat{\beta}_1 - \beta_1^*) \right\} + o_p(1),
\end{aligned}$$

where $K_1(\cdot; x)$, a bounded linear operator, is the Fréchet derivative of $\hat{\tau}_1(x)$ with respect to (F, β_1) at (F^*, β_1^*).Recall the definition of $\tau_1(x)$ and $A_{kg}(u; F)$, it can be seen that

$$K_{11}(y, x) = \left[\frac{y}{A_{01}(\beta_1^{*\top} x; F^*)} - \frac{A_{11}(\beta_1^{*\top} x; F^*)}{A_{01}^2(\beta_1^{*\top} x; F^*)} \right] \cdot e^{\beta_1^{*\top} x y}$$

is a bounded function of bounded variation,

$$K_{12}(x) = \left[\frac{A_{21}(\beta_1^{*\top} x; F^*)}{A_{01}(\beta_1^{*\top} x; F^*)} - \frac{A_{11}^2(\beta_1^{*\top} x; F^*)}{A_{01}^2(\beta_1^{*\top} x; F^*)} \right] \cdot x$$

is a $d-$variate bounded vector function of bounded variation.

Similarly, we also apply delta method along the direction $\hat{h}$ to $\hat{\tau}_0(x)$, we get

$$\begin{aligned}\sqrt{n}(\hat{\tau}_0(x)-\tau_0(x)) &= \sqrt{n}K_0(\hat{h};x)+o_p(\sqrt{n}\hat{h})\\ &= \sqrt{n}\left\{\int K_{01}(y,x)d(\hat{F}(y)-F^*(y))+K_{02}(x)^\top(\hat{\beta}_0-\beta_0^*)\right\}+o_p(1),\end{aligned}$$

where $K_{01}(y,x)$ and $K_{02}(x)$ are similar to $K_{11}(y,x)$ and $K_{12}(x)$, except that they are obtained by replacing β_1^* with β_0^* in $K_{11}(y,x)$ and $K_{12}(x)$ respectively.

Denote $K^1(y,x) = K_{11}(y,x) - K_{01}(y,x)$, $K^2(x) = (K_{02}(x)^\top, K_{12}(x)^\top)^\top$ and $K^{12}(y,x) = (K^1(y,x), K^2(x)^\top)^\top$, we express the linear approximation of $\hat{\tau}(x)$ at (F^*,β^*) along the direction $\hat{h}$ as

$$\hat{\tau}(x)=\tau(x)+\int K^1(y,x)d(\hat{F}(y)-F^*(y))+K^2(x)^\top(\hat{\beta}-\beta^*)+o_p(\hat{h}).$$

Therefore, using the result in (S1.7), we further have

$$\begin{aligned}\sqrt{n}(\hat{\tau}(x)-\tau(x)) &= \sqrt{n}\left\{\int K^1(y,x)d(\hat{F}(y)-F^*(y))+K^2(x)^\top(\hat{\beta}-\beta^*)\right\}+o_p(\sqrt{n}\hat{h})\\ &= \mathbb{G}_n\{\ell_{1,F*,\beta*}(Z)[\tilde{\sigma}_1[K^{12}(Y,x)]]+\tilde{\sigma}_2[K^{12}(Y,x)]^\top\ell_{2,F*,\beta*}(Z)\}+o_p(1)\\ &= \mathbb{G}_n\{S_{F*,\beta*}[\tilde{\sigma}[K^{12}(Y,x)]](Z)\}+o_p(1), \qquad \text{(S3.9)}\end{aligned}$$

which converges to $N(0,\sigma^2_{cate})$ in distribution with

$$\sigma^2_{cate} = \mathbb{V}\mathrm{ar}(\ell_{1,F*,\beta*}(Z)[\tilde{\sigma}_1[K^{12}(Y,x)]]+\tilde{\sigma}_2[K^{12}(Y,x)]^\top\ell_{2,F*,\beta*}(Z))$$

$$= P_0\{S_{F*,\beta*}[\tilde{\sigma}[K^{12}(Y,x)]](Z)\}^{\otimes 2}$$

$$= \langle K^{12}(Y,x), \tilde{\sigma}[K^{12}(Y,x)]\rangle. \tag{S3.10}$$

□

S4 Proof of Corollary 1 in the main paper

Proof. It is evident that τ_{ate}, τ_{att}, and τ_{atc} are functions of $\tau(x)$ (conditional expectations of $\tau(x)$ under different populations).

We first prove the asymptotic property of $\hat{\tau}_{ate}$. To achieve the asymptotic normality of $\hat{\tau}_{ate}$, we consider a class of functions

$$\mathcal{T} = \left\{\tau(x;F,\beta) = \frac{A_{11}(\beta_1^\top x;F)}{A_{01}(\beta_1^\top x;F)} - \frac{A_{11}(\beta_0^\top x;F)}{A_{01}(\beta_0^\top x;F)} : F \in \mathcal{F}, \beta = (\beta_0^\top, \beta_1^\top)^\top, \beta \in \mathcal{B}\right\}, \tag{S4.11}$$

where $\mathcal{F}$ and $\mathcal{B}$ defined in Assumption 2 are both compact. The functions in $\mathcal{T}$ are clearly uniformly bounded, therefore $\mathcal{T}$ forms a P-Donsker class. From the results in (S3.9), we consequently have

$$\begin{aligned}
&\sqrt{n}(\hat{\tau}_{ate} - \tau_{ate}) \\
= \;& \sqrt{n}(\mathbb{P}_n(\tau(X;\hat{F},\hat{\beta})) - P_0(\tau(X;F^*,\beta^*))) \\
= \;& \mathbb{G}_n\{\tau(X;F^*,\beta^*)\} + \mathbb{G}_n\left\{\tau(X;\hat{F},\hat{\beta}) - \tau(X;F^*,\beta^*)\right\} + \sqrt{n}P_0(\tau(X;\hat{F},\hat{\beta}) - \tau(X;F^*,\beta^*)) \\
= \;& \mathbb{G}_n\{\tau(X;F^*,\beta^*)\} + \mathbb{G}_n\left\{\mathbb{E}_{X_*}(\ell_{1,F*,\beta*}(Z)[\tilde{\sigma}_1[K^{12}(Y,X_*)]] + \tilde{\sigma}_2[K^{12}(Y,X_*)]^\top \ell_{2,F*,\beta*}(Z))\right\} + o_P(1)
\end{aligned}$$

$$= \mathbb{G}_n \{\tau(X; F^*, \beta^*)\} + \mathbb{G}_n \left\{\mathbb{E}_{X_*}(S_{F^*,\beta^*}[\tilde{\sigma}[K^{12}(Y, X_*)]](Z))\right\} + o_P(1), \quad \text{(S4.12)}$$

where $\mathbb{G}_n \left\{\tau(X; \hat{F}, \hat{\beta}) - \tau(X; F^*, \beta^*)\right\} = o_p(1)$ is obtained by asymptotic equicontinuity.

After simple algebraic calculation, we find that $\tau(X; F^*, \beta^*)$ and $S_{F,\beta}[h](Z)$ are not correlated for any direction h. Therefore, we can derive from the equation above that $\sqrt{n}(\hat{\tau}_{ate} - \tau_{ate}) \xrightarrow{d} N(0, \sigma^2_{ate})$, where

$$\begin{aligned} \sigma^2_{ate} &= \mathbb{V}\text{ar}(\tau(X; F^*, \beta^*) + \mathbb{E}_{X_*}(S_{F^*,\beta^*}[\tilde{\sigma}[K^{12}(Y, X_*)]](Z))) \quad \text{(S4.13)} \\ &= \mathbb{E}_{X_*}\langle K^{12}(Y, X_*), \tilde{\sigma}[K^{12}(Y, X_*)]\rangle + P_0\{\mathbb{E}_{X_*}(S_{F^*,\beta^*}[\tilde{\sigma}[K^{12}(Y, X_*)]](Z))\}^{\otimes 2}. \end{aligned}$$

Let $\mathbb{P}_{n1}$ be the empirical measure induced by the data-set $\{X_i : D_i = 1, 1 \leqslant i \leqslant n\}$, and denote P_{01} as the conditional probability measure of X given $D = 1$. Similar notations for control group $D = 0$ are $\mathbb{P}_{n0}$ and P_{00}. Accordingly we write $\mathbb{G}_{n0} = \sqrt{n_0}(\mathbb{P}_{n0} - P_{00})$ and $\mathbb{G}_{n1} = \sqrt{n_0}(\mathbb{P}_{n1} - P_{01})$. And let $\mathbb{E}^0_X$ and $\mathbb{E}^1_X$ denote the conditional expectation of X given $D = 0$ and $D = 1$ respectively. Similar to (S4.12), we can get

$$\begin{aligned} &\sqrt{n}(\hat{\tau}_{att} - \tau_{att}) \\ = \;& \sqrt{n}(\mathbb{P}_{n1}(\tau(X; \hat{F}, \hat{\beta})) - P_{01}(\tau(X; F^*, \beta^*))) \\ = \;& \sqrt{\frac{n}{n_1}} \cdot \mathbb{G}_{n1} \{\tau(X; F^*, \beta^*)\} + \sqrt{\frac{n}{n_1}} \cdot \mathbb{G}_{n1} \left\{\tau(X; \hat{F}, \hat{\beta}) - \tau(X; F^*, \beta^*)\right\} \\ & + \sqrt{n} P_{01}(\tau(X; \hat{F}, \hat{\beta}) - \tau(X; F^*, \beta^*)) \end{aligned}$$

$$= \sqrt{\frac{n}{n_1}} \cdot \mathbb{G}_{n1}\left\{\tau(X)\right\} + \mathbb{G}_n\left\{\mathbb{E}^1_{X_*}(\ell_{1,F*,\beta*}(Z)[\tilde{\sigma}_1[K^{12}(Y,X_*)]] + \tilde{\sigma}_2[K^{12}(Y,X_*)]^\top \ell_{2,F*,\beta*}(Z))\right\} + o_P(1)$$

$$= \sqrt{\frac{n}{n_1}} \cdot \mathbb{G}_{n1}\left\{\tau(X)\right\} + \mathbb{G}_n\left\{\mathbb{E}^1_{X_*}(S_{F*,\beta*}[\tilde{\sigma}[K^{12}(Y,X_*)]](Z))\right\} + o_P(1), \quad \text{(S4.14)}$$

which implies $\sqrt{n}(\hat{\tau}_{att} - \tau_{att}) \xrightarrow{d} N(0, \sigma^2_{att})$ with

$$\sigma^2_{att} = \rho_1 \mathbb{E}^1_{X_*}\langle K^{12}(Y,X_*), \tilde{\sigma}[K^{12}(Y,X_*)]\rangle + P_0\{\mathbb{E}^1_{X_*}(S_{F*,\beta*}[\tilde{\sigma}[K^{12}(Y,X_*)]](Z))\}^{\otimes 2}. \quad \text{(S4.15)}$$

Similarly, we can directly obtain the asymptotic distribution of $\hat{\tau}_{atc}$, $\sqrt{n}(\hat{\tau}_{atc} - \tau_{atc}) \xrightarrow{d} N(0, \sigma^2_{atc})$, where

$$\sigma^2_{atc} = \rho_0 \mathbb{E}^0_{X_*}\langle K^{12}(Y,X_*), \tilde{\sigma}[K^{12}(Y,X_*)]\rangle + P_0\{\mathbb{E}^0_{X_*}(S_{F*,\beta*}[\tilde{\sigma}[K^{12}(Y,X_*)]](Z))\}^{\otimes 2}. \quad \text{(S4.16)}$$

□

S5 Proof of Theorem 3 in the main paper

The following Lemma 2 establishes the limiting property of a quantile process, and the Lemma is actually Theorem 1 of Doss and Gill (1992). With the help of this Lemma and Lemma 1, we will easily complete the proof of the asymptotic property of Theorem 2 in the main paper.

Lemma 2 (Limiting process of quantile process, Doss and Gill, 1992)**.** *Let* ξ *be* *a function defined on a bounded interval* $[l, u]$ *with a derivative* ξ' *that is positive and continuous. Let* ξ_n *be nondecreasing right-continuous process on* $[l, u]$ *satisfying* $\xi_n(l) = \xi(l)$ *a.s. and* $\sqrt{n}(\xi_n - \xi) \xrightarrow{d} W$ *in* $D[l, u]$*, where* W *has a.s. continuous sample paths, then for every* $\varepsilon \geqslant 0$*,* $\sqrt{n}(\xi_n^{-1} - \xi^{-1}) \xrightarrow{d} -W(\xi^{-1})/\xi'(\xi^{-1})$ *in* $D[\xi(l), \xi(u) - \varepsilon]$.

In the following, we give the complete proof of Theorem 3, which is a simple application of Lemma 2.

Proof. It is obvious that the estimates of $\hat{F}_{Y(k)|X}$ and $\hat{F}_{Y(k)}$ we construct are a non-decreasing function of y under our SPLRM model. Under Assumption 4, the potential distribution of $Y(k), (k = 0, 1)$,has continuous density function $f_{Y(k)}(y)$. Then the conclusion in lemma 2 implies

$$\begin{aligned} \sqrt{n}\left\{\hat{Q}_{Y(k)}(q) - Q_{Y(k)}(q)\right\} &= \sqrt{n}(\hat{F}^{-1}_{Y(k)} - F^{-1}_{Y(k)})(q) \\ &\xrightarrow{d} -\frac{Z_k(Q_{Y(k)}(q))}{f_{Y(k)}(Q_{Y(k)}(q))} := V_k(q), \end{aligned} \tag{S5.17}$$

where $V_k(q)$ is a stochastic process indexed by $q \in (0, 1)$ and it is also zero-mean tight Gaussian process. Then by the extended continuous mapping theorem, we can easily get that

$$\begin{aligned} \sqrt{n}(\hat{\tau}_{qte}(q) - \tau_{qte}(q)) &= \sqrt{n}(\hat{Q}_{Y(1)}(q) - Q_{Y(1)}(q)) - \sqrt{n}(\hat{Q}_{Y(0)}(q) - Q_{Y(0)}(q)) \\ &\xrightarrow{d} \frac{Z_0(Q_{Y(0)}(q))}{f_{Y(0)}(Q_{Y(0)}(q))} - \frac{Z_1(Q_{Y(1)}(q))}{f_{Y(1)}(Q_{Y(1)}(q))}. \end{aligned} \tag{S5.18}$$

Up to now, we have completed the entire proof of Theorem 3. □

S6 Proof of Theorem 4 in main paper

Under the null hypothesis, suppose the true value of β is $\beta^* = (\beta_1^{*\top}, \beta_1^{*\top})^\top$, where $\beta_1^* \in R^d$. The empirical log-likelihood function based on our observations is

$$\ell_n(F, \beta) = n\mathbb{P}_n \ell(Z; F, \beta).$$

The likelihood ratio test statistic is

$$R = 2\{\sup_{F,\beta} \ell_n(F, \beta) - \sup_{F,\beta_1} \ell_n(F, (\beta_1, \beta_1))\} = R_1 - R_2,$$

where

$$\begin{aligned} R_1 &= 2\{\sup_{F,\beta} \ell_n(F, \beta) - \sup_{F} \ell_n(F, (\beta_1^*, \beta_1^*)\} = 2\{\ell_n(\hat{F}, \hat{\beta}) - \sup_{F} \ell_n(F, (\beta_1^*, \beta_1^*)\} \\ R_2 &= 2\{\sup_{F,\beta_1} \ell_n(F, (\beta_1, \beta_1)) - \sup_{F} \ell_n(F, (\beta_1^*, \beta_1^*)\} = 2\{\ell_n(\tilde{F}, (\tilde{\beta}_1, \tilde{\beta}_1)) - \sup_{F} \ell_n(F, (\beta_1^*, \beta_1^*)\}. \end{aligned}$$

S6.1 Approximation of R_1

Luo and Tsai (2012) has proved that $\hat{\beta}$ is an asymptotic semi-parametric efficient estimate of β, and its corresponding efficient score function is

$$\ell_{\text{eff}}(Z) = \ell_{2,F,\beta}(Z) - \ell_{1,F,\beta}[g^*](Z),$$

where $g^* = -\tilde{\sigma}_1 \left({0 \atop E_{2d\times 2d}} \right) \Sigma$ is the least favorable direction and $\Sigma = \tilde{\sigma}_2 \left({0 \atop E_{2d\times 2d}} \right)$ is the asymptotic efficient variance. Then, it follows that

$$R_1 = \{\mathbb{G}_n \ell_{\text{eff}}(Z)\}^\top \Sigma^{-1} \{\mathbb{G}_n \ell_{\text{eff}}(Z)\} + o_p(1).$$

S6.2 Approximation of R_2

We will look for similar approximations to R_2. First, we calculate the efficient score for β when the null hypothesis H_0 holds.

Under the null hypothesis, the conditional log-likelihood function based on an ideal observation $Z = (Y, X, D)$ is

$$\bar{\ell}(Z; F, \beta_1) = \ell(Z; F, (\beta_1, \beta_1)) = Y \cdot \beta_1^\top X + \ln dF(Y) - \ln \int e^{y \cdot \beta_1^\top X} dF(y). \tag{S6.19}$$

The score function with respect to F along the direction h_1 is

$$\bar{\ell}_{1,F,\beta_1}[h_1](Z) = h_1(Y) - \frac{\int \exp(y \cdot \beta_1^\top X) h_1(y) dF(y)}{\int \exp(y \cdot \beta_1^\top X) dF(y)} = \ell_{1,F,(\beta_1,\beta_1)}[h_1](Z)$$

The score function for β_1 in the direction $h_{2\dagger}$ is

$$h_{2\dagger}^\top \bar{\ell}_{2,F,\beta_1}(Z) = h_{2\dagger}^\top X \left(Y - \frac{\int y \exp(y \cdot \beta_1^\top X) dF(y)}{\int \exp(y \cdot \beta_1^\top X) dF(y)} \right) = h_{2\dagger}^\top E \ell_{2,F,(\beta_1,\beta_1)}(Z),$$

where $E = (E_{2d\times 2d}, E_{2d\times 2d})$ and $E_{2d\times 2d}$ is $(2d)$-dimeansional identity matrix. Then along any direction $h_\dagger = (h_1(\cdot), h_{2\dagger})$, the score of $\bar{\ell}(F, \beta_1)$ with respect to F and β_1 is

$$\begin{aligned} S_{F,\beta_1}[h_\dagger](Z) &= \bar{\ell}_{1,F,\beta_1}[h_1](Z) + h_{2\dagger}^\top \bar{\ell}_{2,F,\beta_1}(Z) \\ &= \ell_{1,F,(\beta_1,\beta_1)}[h_1](Z) + h_{2\dagger}^\top E\ell_{2,F,(\beta_1,\beta_1)}(Z). \end{aligned}$$

Because $(\tilde{F}, \tilde{\beta})$ is the maximizer of $\bar{\ell}_n(F, \beta_1) = n\mathbb{P}_n\bar{\ell}(Z; F, \beta_1)$ we have

$$0 = \mathbb{P}_n\{S_{\tilde{F},\tilde{\beta}_1}[h_\dagger](Z)\}. \tag{S6.20}$$

Recall that

$$\text{pr}(Y = y|X = x, D = 1) = \exp\{\bar{\ell}(z; F, \beta_1)\}.$$

Then

$$1 = \int \exp\{\bar{\ell}(z; F_\varepsilon, \beta_{1\varepsilon})\}dQ(z)$$

for each $|\varepsilon|$ small enough. Taking the derivative with respect to ε on both sides of the above equation gives

$$\begin{aligned} 0 &= \int \{\nabla_\varepsilon \bar{\ell}(z; F_\varepsilon, \beta_{1\varepsilon})\} \exp\{\bar{\ell}(z; F_\varepsilon, \beta_{1\varepsilon})\}dQ(z) \\ 0 &= \int [\{\nabla_{\varepsilon\varepsilon} \bar{\ell}(z; F_\varepsilon, \beta_{1\varepsilon})\} + \{\nabla_\varepsilon \bar{\ell}(z; F_\varepsilon, \beta_{1\varepsilon})\}^2] \exp\{\bar{\ell}(z; F_\varepsilon, \beta_{1\varepsilon})\}dQ(z) \end{aligned}$$

Setting $\varepsilon = 0$ in the first equation, we have

$$0 = P_0\{S_{F*,\beta_1^*}[h_\dagger](Z)\} = P_0\{\ell_{1,F*,\beta*}[h_1](Z) + h_{2\dagger}^\top E\ell_{2,F*,\beta_1^*}(Z)\}. \quad \text{(S6.21)}$$

Setting $\varepsilon = 0$ in the second equation leads to

$$0 = -\langle \sigma_\dagger[h_\dagger], h_\dagger\rangle + \langle S_{F*,\beta_1^*}[h_\dagger](Z), S_{F*,\beta_1^*}[h_\dagger](Z)\rangle$$

or equivalently

$$\langle \sigma_\dagger[h_\dagger], h_\dagger\rangle = \langle S_{F*,\beta_1^*}[h_\dagger](Z), S_{F*,\beta_1^*}[h_\dagger](Z)\rangle, \quad \text{(S6.22)}$$

where the inner product operator is defined as $\langle g_1, g_2\rangle = P(g_1 g_2)$ for any g_1 and g_2 and the operator $\sigma_\dagger$ under our model (1) in the main paper is defined similarly through

$$P_0\left\{\bar{\ell}_{1,F*,\beta_1^*}[h_1](Z) + h_{2\dagger}^\top \bar{\ell}_{2,F*,\beta_1^*}(Z)\right\}^2 = \int \sigma_1[h_\dagger](y)h_1(y)dF^*(y) + h_{2\dagger}^\top \sigma_{2\dagger}[h_\dagger], \qquad \forall h,$$

where

$$\begin{aligned}\sigma_1(h_\dagger)(y) \;&=\; P_0\left\{h_1(y)\frac{e^{\beta_1^{*\top}X\cdot y}}{A_{01}(\beta_1^{*\top}X;F^*)} - \frac{A_{0h_1}(\beta_1^{*\top}X;F^*)e^{\beta_1^{*\top}X\cdot y}}{A_{01}^2(\beta_1^{*\top}X;F^*)}\right.\\ &\left.\qquad + yh_{21}^\top X\frac{e^{\beta_1^{*\top}X\cdot y}}{A_{01}(\beta_1^{*\top}X;F^*)} - h_{21}^\top X\frac{e^{\beta_1^{*\top}X\cdot y}A_{11}(\beta_1^{*\top}X;F)}{A_{01}^2(\beta_1^{*\top}X;F^*)}\right\}\end{aligned}$$

and

$$\begin{aligned}\sigma_{2\dagger}(h_\dagger) &= P_0\left[\left(\frac{A_{21}(\beta_1^{*\top}X;F^*)}{A_{01}(\beta_1^{*\top}X;F^*)}-\frac{A_{11}^2(\beta_1^{*\top}X;F^*)}{A_{01}^2(\beta_1^{*\top}X;F^*)}\right)XX^\top h_{2\dagger}\right]\\ &\quad +P_0\left[\left(\frac{A_{1h_1}(\beta_1^{*\top}X;F^*)}{A_{01}(\beta_1^{*\top}X;F^*)}-\frac{A_{11}(\beta_1^{*\top}X;F^*)A_{0h_1}(\beta_1^{*\top}X;F^*)}{A_{01}^2(\beta_1^{*\top}X;F^*)}\right)X\right]\end{aligned}$$

In a similar computational procedure to Luo and Tsai (2012), we can define the least favorable direction for estimation $\beta = (\beta_1^\top, \beta_1^\top)^\top$ is $g_\dagger^* = -\tilde{\sigma}_{1\dagger}\left(\begin{smallmatrix}0\\ E_{d\times d}\end{smallmatrix}\right)\Sigma_\dagger^{-1}$, where $\Sigma_\dagger = \tilde{\sigma}_{2\dagger}\left(\begin{smallmatrix}0\\ E_{d\times d}\end{smallmatrix}\right)$ is the asymptotic variance. Then we can approximate R_2 using the following formula

$$R_2 = \{\mathbb{G}_n\ell_{\text{eff}\dagger}(Z)\}^\top\Sigma_\dagger^{-1}\{\mathbb{G}_n\ell_{\text{eff}\dagger}(Z)\} + o_p(1),$$

where

$$\ell_{\text{eff}\dagger}(Z) = \bar{\ell}_{2,F^*,\beta_1^*}(Z) - \bar{\ell}_{1,F^*,\beta_1^*}(Z)[g_\dagger^*].$$

Based on the above results, we have

$$\begin{aligned}R &= R_1 - R_2\\ &= \{\mathbb{G}_n\ell_{\text{eff}}(Z)\}^\top\Sigma^{-1}\{\mathbb{G}_n\ell_{\text{eff}}(Z)\} - \{\mathbb{G}_n\ell_{\text{eff}\dagger}(Z)\}^\top\Sigma_\dagger^{-1}\{\mathbb{G}_n\ell_{\text{eff}\dagger}(Z)\} + o_p(1).\end{aligned}$$

S6.3 Asymptotic Chi-square property

In order to find the relationship between the efficient score $\ell_{\text{eff}}(Z)$ and $\ell_{\text{eff}\dagger}(Z)$, we need redifine the scores of $\ell(F,\beta)$ with respect to F and β using the operator concept. The score operator can be expected to take the form

$$A[h](Z) = A_1[h_1](Z) + \boldsymbol{h}_2^\top A_2(Z),$$

where $A_1[\cdot] = \ell_{1,F*,\beta*}[\cdot]$ is the score operator for the nuisance parameter F and $A_2 = \ell_{2,F*,\beta*}$. The adjoint operator A_1^* is a linear map characterized by the requirement $\langle A_1[h], g\rangle_{\dot{\mathbb{H}}_F} = \langle h, A_1^*[g]\rangle_{P_{F,\beta}}$ for every $h \in L_2(P_{F*,\beta*})$ and $g \in \dot{\mathbb{H}}_F$, where $\dot{\mathbb{H}}_F$ is nuisance tangent space. Here, the brackets denote the inner products in the Hilbert space $L_2(P_{F,\beta})$ and $\dot{\mathbb{H}}_F$. Denote $A = (A_1, A_2)$ and A^*A is called the information operator, which is actually is σ defined in section S1 of this proof. σ is onto and continuously invertible, so is A^*A. Under the notation of the operator, the least favorable direction is $g^* = -A_2^*A_1(A_1^*A_1)^{-1}$ and asymptotic efficient variance can be rewritten as $\Sigma = A_2^*A_2 - A_2^*A_1(A_1^*A_1)^{-1}A_1^*A_2$.

When the null hypothesis holds, the score operator can be re-expressed as

$$A_\dagger[h](Z) = A_{1\dagger}[h_1](Z) + h_{2\dagger}^\top A_{2\dagger}(Z),$$

where $A_{1\dagger}[\cdot] = \bar{\ell}_{1,F*,\beta_1^*}[\cdot] = \ell_{1,F*,\beta*}[\cdot] = A_1[\cdot]$ is the score operator for the nuisance parameter F and $A_{2\dagger}(Z) = \bar{\ell}_{2,F*,\beta_1^*}(Z) = E\ell_{2,F*,\beta_1^*}(Z) = EA_2(Z)$. Denote $A_\dagger = (A_1, A_{2\dagger})$ and $A_\dagger^* A_\dagger$ is the information operator, which is actually $\sigma_\dagger$. Similarly, the least favorable direction under the null hypothesis is re-expressed as $g_\dagger^* = A_{2\dagger}^* A_1 (A_1^* A_1)^{-1}$ and directly $\Sigma_\dagger = A_{2\dagger}^* A_{2\dagger} - A_{2\dagger}^* A_1 (A_1^* A_1)^{-1} A_1^* A_{2\dagger} = E\Sigma E^\top$.

Recall that

$$
\begin{aligned}
\ell_{\text{eff}\dagger}(Z) &= \bar{\ell}_{2,F*,\beta_1^*}(Z) - \bar{\ell}_{1,F*,\beta_1^*}[g_\dagger^*](Z) \\
\ell_{\text{eff}}(Z) &= \ell_{2,F*,\beta_1^*}(Z) - \ell_{1,F,\beta*}[g^*](Z)
\end{aligned}
$$

and

$$
\bar{\ell}_{2,F*,\beta_1^*}(Z) = E\ell_{2,F*,\beta_1^*}(Z), \quad \bar{\ell}_{1,F*,\beta_1^*}[\cdot](Z) = \ell_{1,F,\beta*}[\cdot](Z),
$$

where $E = (E_{d\times d}, E_{d\times d})$. In addition, as $A_{2\dagger} = EA_2$, we have

$$
g_\dagger^* = Eg^*
$$

and therefore

$$
\bar{\ell}_{1,F*,\beta_1^*}(Z)[g_\dagger^*] = E\ell_{1,F*,\beta*}(Z)[g^*]
$$

Consequently,

$$
\ell_{\text{eff}\dagger}(Z) = E\ell_{\text{eff}}(Z).
$$

In addition, we have shown that

$$\Sigma_\dagger = E\Sigma E^\top.$$

Let $U_n = \Sigma^{-1/2}\mathbb{G}_n \ell_{\text{eff}}(Z)$. We have shown that $U_n \xrightarrow{d} N(0, E_d)$. Now we have

$$\begin{aligned} R &= R_1 - R_2 \\ &= U_n^\top U_n - U_n^\top \Sigma^{1/2}\Sigma_\dagger^{-1}\Sigma^{1/2}U_n + o_p(1) \\ &= U_n^\top (E_d - \Sigma^{1/2}E^\top (E\Sigma E^\top)^{-1}E\Sigma^{1/2})U_n + o_p(1) \\ &= U_n^\top \Omega U_n + o_p(1) \end{aligned}$$

where

$$\Omega = E_d - \Sigma^{1/2}E^\top (E\Sigma E^\top)^{-1}E\Sigma^{1/2}.$$

It can be seen that

$$\begin{aligned} \Omega^2 &= E_d - 2\Sigma^{1/2}E^\top (E\Sigma E^\top)^{-1}E\Sigma^{1/2} \\ &\quad + \Sigma^{1/2}E^\top (E\Sigma E^\top)^{-1}E\Sigma^{1/2} \times \Sigma^{1/2}E^\top (E\Sigma E^\top)^{-1}E\Sigma^{1/2} \\ &= E_d - \Sigma^{-1/2}E^\top (E\Sigma E^\top)^{-1}E\Sigma^{-1/2} \\ &= \Omega. \end{aligned}$$

Thus Ω is an idempotent matrix with rank d, therefore

$$R = U_n^\top \Omega U_n + o_p(1) \xrightarrow{d} \chi_d^2.$$

S7 Evaluate the adequacy of the SPLRM model

This section develops a practical diagnostic procedure for evaluating the adequacy of the SPLRM. We formulate the associated joint testing problem and present a simple bootstrap-based implementation. We will establish the entire diagnostic process in three steps.

Step I. Testing the single SPLRM structure.

The testing problem (i) in Remark 1 reduces to the following question. Given a sample of observations $\{(X_i, Y_i)\}_{i=1}^n$ from one treatment group, we assess whether their conditional distribution admits the single SPLRM representation

$$dF(y \mid x; F, \beta) = \frac{\exp\{y \cdot x^\top \beta\}\, dF(y)}{\int \exp\{t \cdot x^\top \beta\}\, dF(t)}, \tag{S7.23}$$

where $\beta \in \mathbb{R}^d$ is an unknown finite-dimensional parameter and F denotes some unspecified baseline distribution.

we address this issue by a score test. To capture possible departure of the true data generating process from the SPLRM, we relax the fixed β to be a random vector $\beta + e$, where e follows a multivariate normal distribution with mean zero and variance-covariance matrix Σ. For simplicity, we further assume that $\Sigma = \tau I$, a diagonal matrix. Then the problem of testing whether the SPLRM fits the data is equivalent to testing whether $\tau = 0$ or not. Thus, testing model adequacy reduces to

$$H_0 : \tau = 0 \quad \text{versus} \quad H_1 : \tau > 0.$$

We follow the score test procedure in Liu et al. (2021) to check the goodness of model (S7.23). Let $\phi(u)$ denote a multivariate standard normal distribution. Under this assumption, the likelihood of (τ, β, F) is

$$L_n(F, \beta, \tau) = \prod_{i=1}^{n} \int \frac{\exp(Y_i \cdot X_i^\top \beta + Y_i \cdot X_i^\top u\sqrt{\tau}) dF(Y_i)}{\int \exp(t \cdot X_i^\top \beta + t \cdot X_i^\top u\sqrt{\tau}) dF(t)} \phi(u) du. \tag{S7.24}$$

The log-likelihood is

$$\ell_n(F, \beta, \tau) = \sum_{i=1}^{n} \ln \{m_i(F, \beta, \tau)\}, \tag{S7.25}$$

where

$$m_i(F, \beta, \tau) = \int \frac{\exp(Y_i \cdot X_i^\top \beta + Y_i \cdot X_i^\top u\sqrt{\tau}) dF(Y_i)}{\int \exp(t \cdot X_i^\top \beta + t \cdot X_i^\top u\sqrt{\tau}) dF(t)} \phi(u) du$$

Note that

$$\begin{aligned} \nabla_\tau m_i(F, \beta, \tau) \quad = \quad & \int \frac{\exp(Y_i \cdot X_i^\top \beta + Y_i \cdot X_i^\top u\sqrt{\tau}) dF(Y_i)}{\int \exp(t \cdot X_i^\top \beta + t \cdot X_i^\top u\sqrt{\tau}) dF(t)} Y_i \cdot X_i^\top u \frac{1}{2\sqrt{\tau}} \phi(u) du \\ & - \int \frac{\exp(Y_i \cdot X_i^\top \beta + Y_i \cdot X_i^\top u\sqrt{\tau}) dF(Y_i)}{\{\int \exp(t \cdot X_i^\top \beta + t \cdot X_i^\top u\sqrt{\tau}) dF(t)\}^2} \end{aligned}$$

$$
\begin{aligned}
& \times \int \exp(s \cdot X_i^\top \beta + s \cdot X_i^\top u\sqrt{\tau}) s \cdot X_i^\top u \frac{1}{2\sqrt{\tau}} dF(s)\phi(u)du \\
= \; & \frac{1}{2\sqrt{\tau}} q_i(F, \beta, \tau),
\end{aligned}
$$

where

$$
\begin{aligned}
q_i(F, \beta, \tau) \;=\; & \int \frac{\exp(Y_i \cdot X_i^\top \beta + Y_i \cdot X_i^\top u\sqrt{\tau}) dF(Y_i)}{\int \exp(t \cdot X_i^\top \beta + t \cdot X_i^\top u\sqrt{\tau}) dF(t)} Y_i \cdot X_i^\top u \phi(u) du \\
& - \int \frac{\exp(Y_i \cdot X_i^\top \beta + Y_i \cdot X_i^\top u\sqrt{\tau}) dF(Y_i)}{\{\int \exp(t \cdot X_i^\top \beta + t \cdot X_i^\top u\sqrt{\tau}) dF(t)\}^2} \\
& \times \int \exp(s \cdot X_i^\top \beta + s \cdot X_i^\top u\sqrt{\tau}) s \cdot X_i^\top u dF(s) \phi(u) du
\end{aligned}
$$

The score function is

$$
\begin{aligned}
\nabla_\tau \ell_n(F, \beta, \tau) \;&=\; \sum_{i=1}^n \frac{1}{m_i(F, \beta, \tau)} \frac{1}{2\sqrt{\tau}} q_i(F, \beta, \tau) \\
&=\; \frac{1}{2\sqrt{\tau}} \sum_{i=1}^n \frac{q_i(F, \beta, \tau)}{m_i(F, \beta, \tau)}
\end{aligned}
$$

To calculate the limit of the above score as $\tau \to 0$, we use the L'Hopital's rule:

$$
\begin{aligned}
\lim_{\tau \to 0} \nabla_\tau \ell_n(F, \beta, \tau) \;&=\; \frac{1}{1/\sqrt{\tau}} \nabla_\tau \sum_{i=1}^n \frac{q_i(F, \beta, \tau)}{m_i(F, \beta, \tau)} \Big|_{\tau=0} \\
&=\; \frac{1}{1/\sqrt{\tau}} \sum_{i=1}^n \left\{ \frac{\nabla_\tau q_i(F, \beta, \tau)}{m_i(F, \beta, \tau)} - \frac{q_i(F, \beta, \tau) \nabla_\tau m_i(F, \beta, \tau)}{(m_i(F, \beta, \tau))^2} \right\} \Big|_{\tau=0} \\
&=\; \frac{1}{2} \sum_{i=1}^n \left\{ \frac{2\sqrt{\tau} \nabla_\tau q_i(F, \beta, \tau)}{m_i(F, \beta, \tau)} - \frac{q_i(F, \beta, \tau) \times q_i(F, \beta, \tau)}{(m_i(F, \beta, \tau))^2} \right\} \Big|_{\tau=0}
\end{aligned}
$$

Note that

$$m_i(F,\beta,0) = \frac{\exp(Y_i \cdot X_i^\top \beta) dF(Y_i)}{\int \exp(t \cdot X_i^\top \beta) dF(t)}$$

and

$$\begin{aligned} q_i(F,\beta,0) &= \int \frac{\exp(Y_i \cdot X_i^\top \beta) dF(Y_i)}{\int \exp(t \cdot X_i^\top \beta) dF(t)} Y_i \cdot X_i^\top u \phi(u) du \\ &\quad - \int \frac{\exp(Y_i \cdot X_i^\top \beta) dF(Y_i)}{\{\int \exp(t \cdot X_i^\top \beta) dF(t)\}^2} \times \int \exp(s \cdot X_i^\top \beta) s \cdot X_i^\top u dF(s) \phi(u) du \\ &= 0 \end{aligned}$$

Therefore

$$\lim_{\tau \to 0} \nabla_\tau \ell_n(F,\beta,\tau) = \frac{1}{2} \sum_{i=1}^{n} \frac{\lim_{\tau \to 0} 2\sqrt{\tau} \nabla_\tau q_i(F,\beta,\tau)}{m_i(F,\beta,0)}$$

We need to calculate $\nabla_\tau q_i(F,\beta,\tau)$. It can be seen that

$$\begin{aligned} \nabla_\tau q_i(F,\beta,\tau) &= \int \frac{\exp(Y_i \cdot X_i^\top \beta + Y_i \cdot X_i^\top u\sqrt{\tau}) dF(Y_i)}{\int \exp(t \cdot X_i^\top \beta + t \cdot X_i^\top u\sqrt{\tau}) dF(t)} (Y_i \cdot X_i^\top u)^2 \frac{1}{2\sqrt{\tau}} \phi(u) du \\ &\quad - \int \frac{\exp(Y_i \cdot X_i^\top \beta + Y_i \cdot X_i^\top u\sqrt{\tau}) dF(Y_i)}{\{\int \exp(t \cdot X_i^\top \beta + t \cdot X_i^\top u\sqrt{\tau}) dF(t)\}^2} \\ &\quad \times \int \exp(v \cdot X_i^\top \beta + v \cdot X_i^\top u\sqrt{\tau}) v \cdot \frac{1}{2\sqrt{\tau}} dF(v) \times Y_i \cdot (X_i^\top u)^2 \phi(u) du \\ &\quad - \int \frac{\exp(Y_i \cdot X_i^\top \beta + Y_i \cdot X_i^\top u\sqrt{\tau}) dF(Y_i)}{\{\int \exp(t \cdot X_i^\top \beta + t \cdot X_i^\top u\sqrt{\tau}) dF(t)\}^2} \cdot \frac{Y_i \cdot X_i^\top u}{2\sqrt{\tau}} \\ &\quad \times \int \exp(s \cdot X_i^\top \beta + s \cdot X_i^\top u\sqrt{\tau}) s \cdot X_i^\top u dF(s) \phi(u) du \\ &\quad - \int \frac{\exp(Y_i \cdot X_i^\top \beta + Y_i \cdot X_i^\top u\sqrt{\tau}) dF(Y_i)}{\{\int \exp(t \cdot X_i^\top \beta + t \cdot X_i^\top u\sqrt{\tau}) dF(t)\}^2} \end{aligned}$$

$$
\begin{aligned}
&\times \int \exp(s \cdot X_i^\top \beta + s \cdot X_i^\top u\sqrt{\tau}) s^2 dF(s) \cdot \frac{(X_i^\top u)^2}{2\sqrt{\tau}} \phi(u) du \\
&+2 \int \frac{\exp(Y_i \cdot X_i^\top \beta + Y_i \cdot X_i^\top u\sqrt{\tau}) dF(Y_i)}{\{\int \exp(t \cdot X_i^\top \beta + t \cdot X_i^\top u\sqrt{\tau}) dF(t)\}^2} \\
&\times \int \exp(s \cdot X_i^\top \beta + s \cdot X_i^\top u\sqrt{\tau}) s \cdot X_i^\top u dF(s) \\
&\times \int \exp(t \cdot X_i^\top \beta + t \cdot X_i^\top u\sqrt{\tau}) \frac{t \cdot X_i^\top u}{2\sqrt{\tau}} dF(t) \times \phi(u) du
\end{aligned}
$$

Therefore

$$
\begin{aligned}
2\sqrt{\tau} \times \nabla_\tau q_i(F,\beta,\tau) \quad = \quad & \int \frac{\exp(Y_i \cdot X_i^\top \beta + Y_i \cdot X_i^\top u\sqrt{\tau}) dF(Y_i)}{\int \exp(t \cdot X_i^\top \beta + t \cdot X_i^\top u\sqrt{\tau}) dF(t)} (Y_i \cdot X_i^\top u)^2 \phi(u) du \\
& - \int \frac{\exp(Y_i \cdot X_i^\top \beta + Y_i \cdot X_i^\top u\sqrt{\tau}) dF(Y_i)}{\{\int \exp(t \cdot X_i^\top \beta + t \cdot X_i^\top u\sqrt{\tau}) dF(t)\}^2} \\
& \times \int \exp(v \cdot X_i^\top \beta + v \cdot X_i^\top u\sqrt{\tau}) v dF(v) \times Y_i \cdot (X_i^\top u)^2 \phi(u) du \\
& - \int \frac{\exp(Y_i \cdot X_i^\top \beta + Y_i \cdot X_i^\top u\sqrt{\tau}) dF(Y_i)}{\{\int \exp(t \cdot X_i^\top \beta + t \cdot X_i^\top u\sqrt{\tau}) dF(t)\}^2} \cdot Y_i \cdot X_i^\top u \\
& \times \int \exp(s \cdot X_i^\top \beta + s \cdot X_i^\top u\sqrt{\tau}) s \cdot X_i^\top u dF(s) \phi(u) du \\
& - \int \frac{\exp(Y_i \cdot X_i^\top \beta + Y_i \cdot X_i^\top u\sqrt{\tau}) dF(Y_i)}{\{\int \exp(t \cdot X_i^\top \beta + t \cdot X_i^\top u\sqrt{\tau}) dF(t)\}^2} \\
& \times \int \exp(s \cdot X_i^\top \beta + s \cdot X_i^\top u\sqrt{\tau}) s^2 dF(s) \cdot (X_i^\top u)^2 \phi(u) du \\
& +2 \int \frac{\exp(Y_i \cdot X_i^\top \beta + Y_i \cdot X_i^\top u\sqrt{\tau}) dF(Y_i)}{\{\int \exp(t \cdot X_i^\top \beta + t \cdot X_i^\top u\sqrt{\tau}) dF(t)\}^2} \\
& \times \int \exp(s \cdot X_i^\top \beta + s \cdot X_i^\top u\sqrt{\tau}) s \cdot X_i^\top u dF(s) \\
& \times \int \exp(t \cdot X_i^\top \beta + t \cdot X_i^\top u\sqrt{\tau}) t \cdot X_i^\top u dF(t) \times \phi(u) du
\end{aligned}
$$

Setting $\tau = 0$ on the right hand side of the above equation gives

$$
\begin{aligned}
\Delta_i(F,\beta) &= \int \frac{\exp(Y_i \cdot X_i^\top \beta) dF(Y_i)}{\int \exp(t \cdot X_i^\top \beta) dF(t)} (Y_i \cdot X_i^\top u)^2 \phi(u) du \\
&\quad - \int \frac{\exp(Y_i \cdot X_i^\top \beta) dF(Y_i)}{\{\int \exp(t \cdot X_i^\top \beta) dF(t)\}^2} \times \int \exp(v \cdot X_i^\top \beta) v dF(v) \times Y_i \cdot (X_i^\top u)^2 \phi(u) du \\
&\quad - \int \frac{\exp(Y_i \cdot X_i^\top \beta) dF(Y_i)}{\{\int \exp(t \cdot X_i^\top \beta) dF(t)\}^2} \cdot Y_i \cdot X_i^\top u \times \int \exp(s \cdot X_i^\top \beta) s \cdot X_i^\top u dF(s) \phi(u) du \\
&\quad - \int \frac{\exp(Y_i \cdot X_i^\top \beta) dF(Y_i)}{\{\int \exp(t \cdot X_i^\top \beta) dF(t)\}^2} \times \int \exp(s \cdot X_i^\top \beta) s^2 dF(s) \cdot (X_i^\top u)^2 \phi(u) du \\
&\quad + 2 \int \frac{\exp(Y_i \cdot X_i^\top \beta) dF(Y_i)}{\{\int \exp(t \cdot X_i^\top \beta) dF(t)\}^2} \times \{\int \exp(s \cdot X_i^\top \beta) s dF(s)\}^2 \times (X_i^\top u)^2 \phi(u) du \\
&= \frac{\exp(Y_i \cdot X_i^\top \beta) dF(Y_i)}{\int \exp(t \cdot X_i^\top \beta) dF(t)} Y_i^2 \cdot X_i^\top X_i \\
&\quad - 2 \frac{\exp(Y_i \cdot X_i^\top \beta) dF(Y_i)}{\{\int \exp(t \cdot X_i^\top \beta) dF(t)\}^2} \times \int \exp(v \cdot X_i^\top \beta) v dF(v) \times Y_i \cdot X_i^\top X_i \\
&\quad - \frac{\exp(Y_i \cdot X_i^\top \beta) dF(Y_i)}{\{\int \exp(t \cdot X_i^\top \beta) dF(t)\}^2} \times \int \exp(s \cdot X_i^\top \beta) s^2 dF(s) \cdot X_i^\top X_i \\
&\quad + 2 \frac{\exp(Y_i \cdot X_i^\top \beta) dF(Y_i)}{\{\int \exp(t \cdot X_i^\top \beta) dF(t)\}^2} \times \{\int \exp(s \cdot X_i^\top \beta) s dF(s)\}^2 \times X_i^\top X_i
\end{aligned}
$$

Therefore

$$
\begin{aligned}
\lim_{\tau \to 0} \nabla_\tau \ell_n(F,\beta,\tau) &= \frac{1}{2} \sum_{i=1}^n \frac{\lim_{\tau \to 0} 2\sqrt{\tau} \nabla_\tau q_i(F,\beta,\tau)}{m_i(F,\beta,\tau)} \\
&= \frac{1}{2} \sum_{i=1}^n \frac{\Delta_i}{m_i(F,\beta,0)} \\
&= \frac{1}{2} \sum_{i=1}^n \Bigg[Y_i^2 \cdot X_i^\top X_i \\
&\quad - \frac{2 Y_i \cdot X_i^\top X_i}{\int \exp(t \cdot X_i^\top \beta) dF(t)} \times \int \exp(v \cdot X_i^\top \beta) v dF(v) \\
&\quad - \frac{X_i^\top X_i}{\int \exp(t \cdot X_i^\top \beta) dF(t)} \times \int \exp(s \cdot X_i^\top \beta) s^2 dF(s)
\end{aligned}
$$

$$+\frac{2X_i^\top X_i}{\int \exp(t\cdot X_i^\top\beta)dF(t)}\times\{\int \exp(s\cdot X_i^\top\beta)sdF(s)\}^2\Big]$$

The above limit corresponds to the score function with respect to τ evaluated at $\tau=0$. We therefore define it as

$$S_n(F,\beta)=\frac{1}{2}\sum_{i=1}^n\frac{\Delta_i(F,\beta)}{m_i(F,\beta,0)}.$$

Since (F,β) is unknown, implementing the score-based goodness-of-fit test requires consistent estimation under H_0. For the single SPLRM (S7.23), Cheng et al. (2025) reformulate the estimation problem as two biased-sample empirical likelihood problems, whose solutions admit closed forms. We adopt their iterative procedure to obtain $(\hat{F},\hat{\beta})$ under H_0. Under the proposed algorithm in Cheng et al. (2025),

$$\hat{F}(y)=\sum_{i=1}^n\hat{p}_iI(Y_i\leqslant y),\qquad \hat{G}(x)=\sum_{i=1}^n\hat{q}_iI(X_i\leqslant x),$$

where $\hat{p}_i=\hat{p}_i(\hat{\beta},\hat{G}),\hat{q}_i=\hat{q}_i(\hat{\beta},\hat{F})$. Here G is an auxiliary distribution function introduced for reparametrization.

Specifically,

$$p_i(\beta,G)=\frac{1/\Delta_2(Y_i,\beta,G)}{\sum_{j=1}^n 1/\Delta_2(Y_j,\beta,G)},\qquad q_i(\beta,F)=\frac{1/\Delta_1(X_i,\beta,F)}{\sum_{j=1}^n 1/\Delta_1(X_j,\beta,F)},$$

where

$$\Delta_1(x,\beta,F) = \int e^{(\beta^\top x)y}\, dF(y), \qquad \Delta_2(y,\beta,G) = \int e^{(\beta^\top x)y}\, dG(x).$$

The estimated conditional distribution in (S7.23) is then

$$F(y|x;\hat F,\hat\beta) = \sum_{i=1}^{n} w_i(x;\hat\beta,\hat F,\hat G) I(Y_i \leqslant y), \tag{S7.26}$$

where

$$w_i(x;\hat\beta,\hat F,\hat G) = \hat p_i(\hat\beta,\hat G)\frac{\exp(x^\top\hat\beta\, Y_i)}{\hat A(x^\top\hat\beta;\hat F,\hat G)}, \tag{S7.27}$$

and $\hat A(x^\top\hat\beta;\hat F,\hat G) = \sum_{i=1}^{n} \hat p_i(\hat\beta,\hat G)\exp(x^\top\hat\beta\, Y_i)$.

The test statistic is constructed as

$$T_n = S_n(\hat F,\hat\beta).$$

To approximate the null distribution, we employ a bootstrap procedure under H_0:

1. Generate $X_1^*,\dots,X_n^*$ from the empirical distribution function of $\{X_i\}_{i=1}^n$;

2. For each $j = 1,\dots,n$, generate Y_j^* from $\{Y_i\}_{i=1}^n$ according to the discrete distribution (S7.26). Specifically, conditional on X_j^*, sample Y_j^* from $\{Y_i\}_{i=1}^n$ with probabilities $\{w_1(X_j^*;\hat\beta,\hat F,\hat G),\dots,w_n(X_j^*;\hat\beta,\hat F,\hat G)\}$.

3. Based on the bootstrap sample $\big\{(X_j^*,Y_j^*)\big\}_{j=1}^n$, re-estimate $(\hat F^*,\hat\beta^*)$ using the

iterative procedure of Cheng et al. (2025), and compute the bootstrap statistic $T_n^* = S_n^*(\hat{F}^*, \hat{\beta}^*)$.

Repeating this procedure B times, let $c^*_{1-\alpha}$ denote the $(1-\alpha)$ quantile of $\left\{\max\{T_n^{*(b)}, 0\}\right\}_{b=1}^{B}$. We reject H_0 if

$$T_n > c^*_{1-\alpha}.$$

Step II. Testing equality of baseline distributions.

We now consider testing whether the two groups share a common baseline distribution.

Suppose we have observations $\{(X_i, Y_i, D_i)\}_{i=1}^n$, for the general SPLRM,

$$dF_{(k)}(y|x) = \frac{\exp\{y \cdot x^\top \beta_k\}\, dF_k(y)}{\int \exp\{t \cdot x^\top \beta_k\}\, dF_k(t)}, \qquad k \in \{0, 1\}, \tag{S7.28}$$

we consider the hypothesis test

$$H_0' : F_0 = F_1 \equiv F \quad \text{versus} \quad H_1' : F_0 \neq F_1.$$

Specifically, we construct a profile likelihood ratio statistic under the SPLRM formulation and approximate its null distribution via bootstrap. Let $\tilde{\ell}_n(F_0, F_1, \beta_0, \beta_1)$ denote the corresponding profile log-likelihood, which has the same form as (2.3) in the main paper, namely,

$$\tilde{\ell}_n(F_0, F_1, \beta_0, \beta_1) = \sum_{k=0}^{1} \left[\sum_{i=1}^{n_k} Y_{ki} \cdot X_{ki}^\top \beta_k + \log(p_{ki}) - \log\left\{ \sum_{s=0}^{1} \sum_{r=1}^{n_s} p_{sr} \cdot \exp(Y_{sr} \cdot X_{ki}^\top \beta_k) \right\} \right].$$

The likelihood ratio statistic is defined as

$$\widetilde{R}_n = 2\{\sup_{F_0,F_1,\beta_0,\beta_1} \tilde{\ell}_n(F_0, F_1, \beta_0, \beta_1) - \sup_{F,\beta_0,\beta_1} \tilde{\ell}_n(F, F, \beta_0, \beta_1)\}$$

Under the unrestricted model, we estimate (F_k, β_k) separately for $k = 0, 1$ by applying the iterative procedure of Cheng et al. (2025) to each group-specific sample. This yields the maximizer $\sup_{F_0,F_1,\beta_0,\beta_1} \tilde{\ell}_n(F_0, F_1, \beta_0, \beta_1)$, with corresponding estimates $(\breve{F}_0, \breve{F}_1, \breve{\beta}_0, \breve{\beta}_1)$. Under H_0', we impose the constraint $F_0 = F_1 \equiv F$ and estimate (F, β_0, β_1) using the iterative procedure described in the main paper. This gives $\sup_{F,\beta_0,\beta_1} \tilde{\ell}_n(F, F, \beta_0, \beta_1)$, with estimates $(\hat{F}, \hat{F}, \hat{\beta}_0, \hat{\beta}_1)$. Recall that $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1)$.

Under H_0', the estimated conditional distribution has the same form as in Equation (2.6) of the main paper:

$$\hat{F}_{(k)}(y \mid x) = \sum_{i=1}^{n} w_{i,k}(x; \hat{F}, \hat{\beta}) I(Y_i \leqslant y), \tag{S7.29}$$

where

$$w_{i,k}(x; \hat{F}, \hat{\beta}) = \hat{p}_i(\hat{\beta}) \frac{\exp(x^\top \hat{\beta}_k Y_i)}{\hat{A}(x^\top \hat{\beta}_k; \hat{F})}, \tag{S7.30}$$

and $\hat{p}_i(\hat{\beta})$ denotes the limiting iterative value of p_{uv} in (2.5) with (β, F) replaced by $(\hat{\beta}, \hat{F})$.

Because the asymptotic distribution of $\widetilde{R}_n$ is nonstandard, we approximate its null distribution via bootstrap. Under the unconfoundedness assumption, the propensity score $\pi(x) = P(D = 1 \mid X = x)$ can be consistently estimated. Let

$\hat{\pi}(x)$ denote such an estimator.

1. Generate $X_1^*, \dots, X_n^*$ from the empirical distribution function of $\{X_i\}_{i=1}^n$;

2. For each $j = 1, \dots, n$, generate D_j^* from a Bernoulli$\{\hat{\pi}(X_j^*)\}$ distribution.

3. For each $j = 1, \dots, n$ and $k \in \{0, 1\}$, generate potention outcome $Y_j^*(k)$ from $\{Y_i\}_{i=1}^n$ according to the discrete distribution (S7.29). Specifically, conditional on X_j^*, sample $Y_j^*(k)$ from $\{Y_i\}_{i=1}^n$ with probabilities $\{w_{1,k}(X_j^*; \hat{F}, \hat{\beta}), \dots, w_{n,k}(X_j^*; \hat{F}, \hat{\beta})\}$. Set $Y_i^* = D_i^* Y_i^*(1) + (1 - D_i^*) Y_i^*(0)$.

4. Based on the bootstrap sample $\{(X_i^*, Y_i^*, D_i^*)\}_{i=1}^n$, compute the bootstrap statistic

$$\widetilde{R}_n^* = 2\{\sup_{F_0, F_1, \beta_0, \beta_1} \tilde{\ell}_n^*(F_0, F_1, \beta_0, \beta_1) - \sup_{F, F, \beta_0, \beta_1} \tilde{\ell}_n^*(F, F, \beta_0, \beta_1)\}.$$

Repeating the above procedure B times, let $q_{1-\alpha}^*$ denote the $(1 - \alpha)$ quantile of $\{\widetilde{R}_n^{*(b)}\}_{b=1}^B$. We reject H_0' if

$$\widetilde{R}_n > q_{1-\alpha}^*.$$

Step III. Joint diagnostic via Bonferroni adjustment.

The overall diagnostic for SPLRM requires simultaneously testing the two hypotheses corresponding to (i) and (ii). To control the family-wise error rate at level α, we apply a Bonferroni correction. Specifically, each test is conducted at significance level $\alpha/2$. The SPLRM assumption is rejected if either of the two null hypotheses is rejected under this adjusted level.

S8 Additional simulation results

Tables S1-S4 present additional simulation results when data were generated from Examples 1 and 2 in the main paper.

Table S1: Simulated Biases, Variances, and MSEs under data generated from Example 1 with low-dimensional ($d = 5$) covariates of the **ps-N** type. All numbers reported have been multiplied by 100.

	τ_{ate}	τ_{att}	τ_{atc}	$\tau_{0.1}$	$\tau_{0.5}$	$\tau_{0.9}$	τ_{ate}	τ_{att}	τ_{atc}	$\tau_{0.1}$	$\tau_{0.5}$	$\tau_{0.9}$
	HOM:Bias						**HET:Bias**					
SPLRM	-0.43	-0.33	-0.49	-0.67	-0.44	0.15	-0.62	-0.2	-0.84	-0.44	-0.41	-0.29
GLM	-2.97	21.48	-17.35	17.8	-3.44	-19.4	-5.98	21.58	-21.6	18.16	-6.39	-26.77
POW	3.9	11.75	-8.65	9.72	13.06	-15.86	3.15	11.86	-10.68	10.27	14.24	-21.07
GRF	-32.99	0.3	-44.86	-17.95	-32.84	-47.26	-40.74	0.42	-56.09	-20.77	-40.21	-60.33
GBM	-21.19	13.45	-39.22	-1.9	-20.89	-39.58	-28	13.36	-49	-3.57	-27.26	-52.02
IDEAL	-0.29	-0.28	-0.29	-0.16	-0.17	-0.75	-0.26	-0.18	-0.3	0.12	-0.1	-1.09
T-PAR	-0.4	0.09	-0.64	0.42	-0.34	-1.03	-0.37	0.19	-0.65	0.68	-0.33	-1.3
F-PAR	-38.23	-23.35	-45.46	-13.11	-39.27	-34.34	-45.81	-23.25	-56.79	-30.14	-46.94	-78.43
POOL	-0.55	-0.2	-0.73	-0.48	-0.49	-0.43	-0.52	-0.1	-0.74	-0.12	-0.48	-0.8
AIPW-1	-0.47	0.04	-0.73	0.33	-0.79	-1.3	-0.45	0.15	-0.76	0.83	-0.94	-1.65
AIPW-2	-26.9	0.85	-40.41	-11.45	-27.27	-41.67	-33.65	0.95	-50.5	-13.31	-33.7	-53.91
	HOM:Variance						**HET:Variance**					
SPLRM	1.37	0.9	1.83	0.65	1.4	2.81	1.38	0.92	1.85	0.93	1.71	4.36
GLM	1.9	3.17	2.53	7.42	2.85	8.76	2.06	3.14	2.92	7.86	3.18	10.98
POW	16.11	10.1	20.32	27.23	35.52	35.52	19.96	10.12	25.34	32.45	43.21	42.88
GRF	1.68	2.09	2.18	5.09	2.93	6.24	1.86	2.24	2.45	5.64	3.31	7.21
GBM	2.3	3.58	2.42	7.4	3.91	7.06	2.5	3.59	2.82	7.8	4.22	7.88
IDEAL	0.41	1.27	0.62	1.74	0.76	1.73	0.42	1.28	0.64	1.81	0.82	1.84
T-PAR	1.38	1.27	1.8	1.72	1.38	3.08	1.39	1.28	1.82	1.78	1.39	3.34
F-PAR	1.63	1.98	1.75	357.3	1.61	687.79	1.84	2.16	1.98	116.58	1.82	513.56
POOL	1.27	0.96	1.75	1.01	1.27	2.93	1.29	0.98	1.77	1.1	1.29	3.2
AIPW-1	1.47	1.15	2.02	3.85	2.35	5.13	1.48	1.16	2.04	4.2	2.58	5.45
AIPW-2	1.64	2.09	1.95	4.57	2.52	5.19	1.8	2.22	2.18	5	2.8	5.99
	HOM:MSE						**HET:MSE**					
SPLRM	1.38	0.9	1.84	0.65	1.4	2.81	1.38	0.92	1.85	0.93	1.71	4.36
GLM	1.98	7.78	5.54	10.59	2.96	12.53	2.41	7.8	7.59	11.16	3.59	18.14
POW	16.26	11.48	21.07	28.17	37.23	38.03	20.06	11.53	26.49	33.5	45.24	47.32
GRF	12.57	2.09	22.31	8.31	13.72	28.58	18.45	2.24	33.91	9.95	19.48	43.61
GBM	6.8	5.39	17.8	7.44	8.27	22.73	10.34	5.38	26.82	7.93	11.65	34.94
IDEAL	0.41	1.28	0.62	1.74	0.76	1.74	0.42	1.28	0.64	1.81	0.82	1.85
T-PAR	1.38	1.27	1.8	1.73	1.38	3.09	1.39	1.28	1.82	1.79	1.39	3.36
F-PAR	16.25	7.44	22.42	359.02	17.03	699.59	22.83	7.56	34.23	125.67	23.86	575.08
POOL	1.28	0.97	1.76	1.01	1.28	2.93	1.29	0.98	1.77	1.1	1.29	3.21
AIPW-1	1.47	1.15	2.03	3.85	2.35	5.14	1.48	1.16	2.05	4.2	2.59	5.48
AIPW-2	8.87	2.1	18.28	5.88	9.96	22.55	13.12	2.23	27.68	6.77	14.15	35.05

Table S2: Simulated Biases, Variances, and MSEs under data generated from Example 1 with large-dimensional ($d = 10$) covariates of the **ps-L** type. All numbers reported have been multiplied by 100.

	τ_{ate}	τ_{att}	τ_{atc}	$\tau_{0.1}$	$\tau_{0.5}$	$\tau_{0.9}$	τ_{ate}	τ_{att}	τ_{atc}	$\tau_{0.1}$	$\tau_{0.5}$	$\tau_{0.9}$
	HOM:Bias						**HET:Bias**					
SPLRM	-0.5	0.07	-0.64	-3.65	-0.66	3.01	-0.9	0.23	-1.18	-2.44	-0.25	1.51
GLM	-11.47	-3.87	-12.24	10.3	-6.67	-40	-14.24	-3.82	-15.35	11.55	-8.52	-48.57
POW	-18.19	-8.1	-17.2	19.74	-12.5	-64.69	-22.4	-8.05	-21.67	24.92	-15.65	-78.57
GRF	-25.9	-21.31	-23.49	-20.7	-24.7	-30.95	-31.04	-21.27	-29.35	-24.24	-29.17	-38.47
GBM	-23.37	-14.99	-25.15	-13.13	-21.92	-32.58	-28.56	-14.95	-31.46	-15.05	-27.04	-40.14
IDEAL	-0.28	-0.48	-0.24	-0.1	-0.52	-0.16	-0.21	-0.42	-0.18	0.54	-0.37	-0.43
T-PAR	-0.61	0	-0.77	-1.97	-0.82	1.11	-0.54	0.05	-0.7	-1.09	-0.75	0.63
F-PAR	1.77	-0.3	2.24	-9.14	2.33	59.45	2.37	-0.24	2.96	-13.7	-1.04	2.98
POOL	-0.7	-0.22	-0.83	-4.07	-0.91	3.05	-0.62	-0.16	-0.76	-2.88	-0.83	2.26
AIPW-1	0.15	-0.09	0.16	-0.15	-0.27	-0.11	0.18	-0.1	0.2	-0.28	-0.09	-0.75
AIPW-2	3.51	-0.89	4.52	-9.54	1.07	19.36	4.46	-0.82	5.68	-9.71	1.2	21.95
	HOM:Variance						**HET:Variance**					
SPLRM	5.23	2.05	7.41	2.86	5.33	9.51	5.18	2.17	7.31	3.44	6.22	11.9
GLM	30.38	14.77	40.38	54	64.28	58.39	39.71	14.95	53.12	67.74	85.18	77.82
POW	42.19	24.48	56.47	67.61	88.39	73.99	52.24	24.66	70.43	84.74	110.87	95
GRF	9.5	5.69	14.07	23.99	16.11	35.78	12.66	6.46	18.78	30.56	21.56	48.62
GBM	10.01	9.23	13.37	27.52	17.38	33.91	12.85	9.63	17.28	34.17	22.43	42.74
IDEAL	0.4	2.29	0.5	2.01	0.92	2.01	0.42	2.39	0.53	2.24	1.04	2.22
T-PAR	5.48	2.25	7.66	5.12	5.48	9.31	5.52	2.34	7.71	5.19	5.52	9.68
F-PAR	48.51	9.2	72.15	139.48	145.64	2074.24	63.66	10.04	94.95	48.93	49.58	620.94
POOL	5.49	1.97	7.69	4.66	5.49	9	5.53	2.06	7.74	4.76	5.53	9.41
AIPW-1	6.54	2.62	9.43	16.11	10.5	21.99	6.65	2.7	9.56	18.24	10.83	24.34
AIPW-2	47.76	9.33	71.42	54.17	43.32	103.92	63.16	10.09	94.71	69.05	57.22	140.69
	HOM:MSE						**HET:MSE**					
SPLRM	5.23	2.05	7.41	2.99	5.33	9.6	5.19	2.17	7.32	3.5	6.22	11.92
GLM	31.69	14.92	41.88	55.06	64.72	74.4	41.74	15.1	55.47	69.07	85.91	101.42
POW	45.5	25.14	59.43	71.51	89.95	115.84	57.26	25.31	75.12	90.95	113.32	156.73
GRF	16.21	10.23	19.58	28.28	22.21	45.36	22.29	10.98	27.39	36.44	30.06	63.42
GBM	15.47	11.48	19.69	29.25	22.19	44.52	21	11.86	27.17	36.43	29.74	58.85
IDEAL	0.4	2.29	0.5	2.01	0.92	2.01	0.42	2.39	0.53	2.24	1.04	2.22
T-PAR	5.48	2.25	7.67	5.16	5.48	9.32	5.52	2.34	7.72	5.2	5.52	9.69
F-PAR	48.54	9.2	72.2	140.32	145.69	2109.59	63.71	10.05	95.04	50.81	49.59	621.03
POOL	5.5	1.97	7.7	4.82	5.5	9.09	5.53	2.06	7.75	4.84	5.54	9.46
AIPW-1	6.54	2.62	9.43	16.11	10.5	21.99	6.65	2.7	9.56	18.25	10.83	24.34
AIPW-2	47.88	9.34	71.62	55.08	43.33	107.67	63.36	10.1	95.03	70	57.24	145.51

Table S3: Simulated Biases, Variances, and MSEs under data generated from Example 1 with large-dimensional ($d = 10$) covariates of the **ps-N** type. All numbers reported have been multiplied by 100.

	τ_{ate}	τ_{att}	τ_{atc}	$\tau_{0.1}$	$\tau_{0.5}$	$\tau_{0.9}$	τ_{ate}	τ_{att}	τ_{atc}	$\tau_{0.1}$	$\tau_{0.5}$	$\tau_{0.9}$
	HOM:Bias						**HET:Bias**					
SPLRM	0.04	-0.6	0.34	-2.33	-0.47	2.98	-0.24	-0.58	-0.1	-2.38	-0.3	2.55
GLM	-3.01	19.95	-16.66	10.99	-2.94	-14.09	-6.03	19.83	-20.86	9.9	-6.21	-19.98
POW	-1.13	9.03	-14.14	5.91	8.81	-23.19	-2.99	8.91	-17.7	6.67	9.19	-32.2
GRF	-35.46	-7.98	-45.03	-25.2	-35.17	-45.67	-43.3	-8.09	-56.34	-29.97	-43.25	-56.65
GBM	-24.99	5.9	-39.86	-11.67	-24.84	-37.64	-31.96	5.68	-49.87	-15.4	-31.74	-48.68
IDEAL	-0.2	-0.77	0.09	-1.78	-0.16	1.22	-0.17	-0.89	0.17	-2.14	-0.24	0.97
T-PAR	-0.07	-0.91	0.33	-2.61	-0.42	2.64	-0.04	-1.03	0.42	-2.8	-0.67	2.35
F-PAR	-38.78	-24.78	-45.58	-29.14	-39.98	-36.67	-46.46	-24.89	-56.95	-34.04	-53.82	-92.93
POOL	0.06	-0.69	0.41	-3.02	-0.29	3.31	0.09	-0.8	0.5	-3.01	-0.54	2.83
AIPW-1	0.08	-0.61	0.41	-1.59	-0.46	2.12	0.12	-0.71	0.5	-2.25	-0.6	1.9
AIPW-2	-28.47	-4.23	-40.24	-17.6	-28.73	-38.82	-35.29	-4.33	-50.33	-21.65	-35.58	-49.25
	HOM:Variance						**HET:Variance**					
SPLRM	1.46	1.06	1.97	0.88	1.6	2.95	1.49	1.14	1.99	1.92	2.16	5.55
GLM	2.72	3.85	4.11	10.19	4.61	13.85	3.27	3.87	5.3	11.31	5.58	18.46
POW	30.66	19.38	35.31	56.54	62.28	51.87	39.21	19.42	46.95	70.42	79.26	62.77
GRF	2.49	2.73	3.27	7.5	4.08	7.76	2.97	3.14	3.95	8.84	4.88	9.6
GBM	2.86	4.1	3.55	9.9	4.66	9.07	3.53	4.59	4.5	11.51	5.73	11
IDEAL	0.42	1.21	0.63	1.94	1.03	2.09	0.45	1.28	0.68	2.15	1.13	2.27
T-PAR	1.48	1.26	1.97	1.94	1.48	3.37	1.5	1.32	1.98	2.06	1.5	3.54
F-PAR	2.46	2.75	2.69	4.38	2.43	280.12	2.99	3.19	3.3	4.86	172.64	915.58
POOL	1.38	1.02	1.93	1.17	1.38	3.21	1.41	1.08	1.94	1.3	1.41	3.41
AIPW-1	1.62	1.24	2.26	4.76	2.96	5.96	1.65	1.31	2.27	5.29	3.25	6.59
AIPW-2	2.48	2.82	3.06	6.65	3.62	6.68	2.97	3.22	3.71	7.78	4.38	8.32
	HOM:MSE						**HET:MSE**					
SPLRM	1.46	1.07	1.97	0.93	1.6	3.04	1.49	1.14	1.99	1.97	2.16	5.62
GLM	2.81	7.82	6.88	11.39	4.7	15.84	3.64	7.8	9.65	12.29	5.96	22.45
POW	30.68	20.19	37.31	56.89	63.06	57.25	39.3	20.21	50.08	70.87	80.11	73.14
GRF	15.06	3.37	23.55	13.85	16.45	28.61	21.72	3.8	35.69	17.82	23.59	41.69
GBM	9.1	4.45	19.44	11.26	10.82	23.24	13.75	4.91	29.37	13.89	15.8	34.7
IDEAL	0.42	1.22	0.63	1.97	1.03	2.1	0.45	1.29	0.68	2.19	1.13	2.28
T-PAR	1.48	1.27	1.97	2.01	1.48	3.44	1.5	1.33	1.99	2.14	1.51	3.6
F-PAR	17.5	8.89	23.46	12.88	18.41	293.56	24.57	9.39	35.73	16.45	201.61	1001.95
POOL	1.38	1.03	1.93	1.26	1.38	3.32	1.41	1.09	1.94	1.39	1.41	3.49
AIPW-1	1.62	1.24	2.26	4.78	2.96	6.01	1.65	1.31	2.27	5.34	3.25	6.63
AIPW-2	10.59	3	19.26	9.75	11.87	21.75	15.42	3.41	29.04	12.47	17.04	32.57

Table S4: Simulated Biases, Variances, and MSEs under data generated from Example 2 with large-dimensional ($d = 10$) covariates. All numbers reported have been multiplied by 100.

	τ_{ate}	τ_{att}	τ_{atc}	$\tau_{0.1}$	$\tau_{0.5}$	$\tau_{0.9}$	τ_{ate}	τ_{att}	τ_{atc}	$\tau_{0.1}$	$\tau_{0.5}$	$\tau_{0.9}$
	ps-L:Bias						**ps-N:Bias**					
SPLRM	8.69	-7.56	12.39	0.02	12.53	30.29	-2.3	-5.7	-0.59	-3.69	2.73	14.59
GLM	14.93	-3.84	17.28	22.71	19.82	3.12	11.15	-0.37	15.92	4.31	13.18	15.6
POW	23.9	-1.3	25.11	33.22	33.53	-3.42	10.34	0.12	17.61	12.1	16.3	1.26
GRF	31.49	-1.48	36.17	20.02	35.03	40.67	28.35	-0.91	42.28	13.86	32.13	37.45
GBM	29.8	-1.51	36.62	19.54	35.53	37.19	24.25	-0.86	35.98	12.87	28.38	29.6
IDEAL	-0.02	-1.07	0.01	-0.38	0.31	-0.81	-0.85	-1.27	-0.61	-0.57	-0.97	-1.64
T-PAR	0.42	-0.95	0.53	0.72	1.14	-0.88	-0.25	-0.65	-0.03	-0.05	0.18	-1.15
F-PAR	93.13	27.58	108.74	19.3	123.62	125.14	47.4	28.72	56.5	5.24	79.7	46.39
POOL	49.14	28.93	53.79	12.91	87.61	30.62	33.6	28.32	36.2	4.84	72.48	9.5
AIPW-1	-1.04	-1.41	-1.19	2.83	1.68	-4.71	-0.71	-0.86	-0.6	0.2	-0.53	-0.48
AIPW-2	48.9	4.31	59.45	19.6	45.34	84.64	23.19	1.36	33.82	9.57	26.7	33.11
	ps-L:Variance						**ps-N:Variance**					
SPLRM	6.08	5.29	8	0.82	9.62	9.18	1.98	2.49	2.27	0.6	3.94	4.29
GLM	21.05	17.19	26.98	7.73	41.86	92.12	3.2	4.27	4.18	1.41	6.73	15.44
POW	39.54	31.54	51.45	21.87	74.13	144.61	22.81	23.51	22.45	11.41	48.09	88.08
GRF	9.19	6.32	12.97	4.31	15.01	55.56	2.47	2.98	3.08	1.15	4.85	14.26
GBM	9.37	9.05	12.08	4.4	15.82	63.52	2.66	3.89	3.35	1.29	5.56	14.74
IDEAL	0.91	5.23	1.07	0.53	1.95	4.22	0.88	2.98	1.2	0.48	1.99	4.11
T-PAR	1.74	3.34	1.74	0.37	1.76	7.06	1.09	1.72	1.04	0.16	0.89	5.43
F-PAR	19.97	10.39	29.1	2.01	15.74	76.41	7	6.96	9.21	0.75	5.15	32.88
POOL	4.37	0.84	7.69	1.88	4.07	20.68	1.03	0.53	1.57	0.4	1	4.91
AIPW-1	6	5.38	7.91	1.69	10.02	49.47	1.85	2.61	2.1	0.77	3.83	11.38
AIPW-2	15.89	7.15	23.38	3.79	15.74	83.27	3.21	3.12	4.93	1.09	4.92	14.58
	ps-L:MSE						**ps-N:MSE**					
SPLRM	6.83	5.87	9.53	0.82	11.19	18.35	2.03	2.82	2.28	0.73	4.01	6.41
GLM	23.28	17.33	29.96	12.88	45.78	92.22	4.44	4.28	6.72	1.6	8.47	17.87
POW	45.25	31.56	57.75	32.9	85.38	144.73	23.88	23.51	25.56	12.88	50.75	88.1
GRF	19.11	6.34	26.05	8.32	27.28	72.1	10.51	2.99	20.95	3.07	15.17	28.29
GBM	18.25	9.07	25.49	8.22	28.45	77.36	8.54	3.9	16.3	2.95	13.61	23.49
IDEAL	0.91	5.24	1.07	0.54	1.95	4.23	0.89	3	1.2	0.49	2	4.13
T-PAR	1.74	3.35	1.75	0.37	1.78	7.07	1.09	1.73	1.04	0.16	0.89	5.45
F-PAR	106.69	18	147.35	5.74	168.56	233	29.47	15.21	41.14	1.03	68.68	54.4
POOL	28.52	9.21	36.62	3.55	80.82	30.06	12.33	8.55	14.68	0.64	53.54	5.82
AIPW-1	6.01	5.4	7.92	1.77	10.05	49.69	1.85	2.62	2.1	0.77	3.83	11.38
AIPW-2	39.8	7.33	58.72	7.63	36.3	154.92	8.58	3.14	16.37	2.01	12.05	25.54

S8.1 On asymptotic variances and their bootstrap estimates

In the main paper, we have derived the asymptotic variance of the proposed estimators for various estimands. In the real data analysis of the main paper, we report bootstrap-based estimates of the asymptotic variances. We may wonder whether the bootstrap-based estimates are close to the theoretical variances.

To address this concern, we consider a representative simulation setting: Example 1 under the low-dimensional setting ($d = 5$), the propensity score satisfying the **ps-L** model, and $\beta_0 = \beta_1 = (0.4, \ldots, 0.4)$. See Section 4.2 in the main paper for more information on this setting. To calculate the theoretical (asymptotic) variance or approximate it as closely as possible, we set sample size n to be 5000 or 10000. For each n, we generate 1000 samples of size n, calculate the proposed point estimate based on each sample, and calculate the sample variance, say S, of the resulting 1000 point estimates. Then the quantity $n \times S$ is an approximate of the asymptotic variance of this point estimator. The first panel of Table Table S5 gives such results for all the causal parameter of interest. We observe that, as the sample size increases from $n = 5000$ to $n = 10000$, the estimated variances for all causal parameters remain essentially unchanged. This stability suggests that the values reported in the first panel can be regarded as reliable approximations to the theoretical asymptotic variances.

The second panel of Table S5 presents the bootstrap variance estimates for various causal parameters based on $B = 300$ bootstrap samples of size $n \in \{500, 1000, 2000\}$.

We see that the bootstrap variance estimates are very close to the theoretical counterparts when the causal parameters are τ_{ate}, τ_{att}, τ_{atc} and $\tau_{0.5}$. The bootstrap variance estimates for $\tau_{0.1}$ and $\tau_{0.9}$ are a bit larger than the theoretical variances when the sample size is as small as 500. Their gaps decrease as the sample size increases to 1000 and 2000. In practice, using a slightly inflated variance estimate is more acceptable than using an underestimated one, because it produces a conservative confidence intervala result that is typically deemed acceptable.

Table S5: Table of theoretical (asymptotic) and bootstrap variance estimates under Example 1 ($d = 5$, **ps-L**)

n	τ_{ate}	τ_{att}	τ_{atc}	$\tau_{0.1}$	$\tau_{0.5}$	$\tau_{0.9}$
Theoretical Variance						
5000	20.935	7.409	29.971	8.299	21.003	40.100
10000	20.926	7.425	30.047	8.212	20.801	39.605
Variance estimation based on bootstrap						
500	23.543	8.145	33.824	12.734	25.237	51.766
1000	22.441	7.901	32.305	10.809	23.420	46.287
2000	21.606	7.824	31.050	9.677	22.337	43.145

S9 Additional results for LLvsPSID dataset

Estimates based on IPW and AIPW are highly sensitive to propensity score estimation and may become unstable in the presence of extreme weights. To assess this sensitivity in the LLvsPSID data, we additionally implemented truncated IPW and AIPW estimators; the corresponding results are reported in Tables S6, and S7.

Relative to the untruncated estimators, truncation generally brings the IPW and AIPW estimates closer to those obtained from SPLRM and REG. The effec-

Table S6: Truncated estimates and 95% bootstrap confidence intervals ([Lower, Upper]) for ATE, ATT, and ATC on the earnings of participants aged below 30 years based on the LLvsPSID data. Cut-off point for propensity scores: 0.02.

		GLM-trunc	POW-trunc	GRF-trunc	GBM-trunc	AIPW-trunc
ATE	EST	-9290.48	-8143.09	-11182.87	-10466.90	-9914.25
	Lower	-12807.09	-12610.63	-14230.23	-13681.51	-13172.60
	Upper	-5773.87	-3675.54	-8135.52	-7252.30	-6655.90
ATT	EST	90.79	-859.52	-3672.18	-1109.07	-2899.99
	Lower	-1880.17	-10021.38	-5581.97	-3358.96	-4619.60
	Upper	2061.74	8302.34	-1762.40	1140.82	-1180.39
ATC	EST	-12445.57	-13093.24	-12566.16	-12849.14	-11616.54
	Lower	-16664.14	-16496.55	-16889.23	-17071.06	-15554.84
	Upper	-8227.00	-9689.94	-8243.09	-8627.22	-7678.24

t is particularly pronounced for the POW-based estimators, in which truncation substantially alters both ATE and QTE estimates, even reversing their signs. For example, QTE50, QTE75, and QTE90 change from non-significant pre-truncation values of 3131.09, 5224.29, and 7999.13 to significant negative values of -8919.46, -13921.67, and -17803.98. For these quantiles, the truncated AIPW estimates also move markedly closer to SPLRM. These findings suggest that IPW and AIPW estimates are strongly influenced by extreme propensity scores. In contrast, SPLRM provides a model-flexible alternative that avoids reliance on inverse probability weighting and yields empirically more reliable estimates in the presence of extreme propensity scores.

Table S7: Truncated estimates and 95% bootstrap confidence intervals ([Lower, Upper]) for QTEs on the earnings of participants aged below 30 years based on the LLvsPSID data. Cut-off point for propensity scores: 0.02.

		GLM-trunc	POW-trunc	GRF-trunc	GBM-trunc	AIPW-trunc
	EST	0.00	0.00	-218.70	0.00	-128.21
QTE10	Lower	-1013.12	-974.65	-1605.81	-1200.99	-997.34
	Upper	1013.12	974.65	1168.41	1200.99	740.91
	EST	-5484.68	-3151.43	-8766.79	-6719.89	-8432.41
QTE25	Lower	-8871.68	-7944.01	-10403.43	-9593.19	-9967.01
	Upper	-2097.68	1641.16	-7130.15	-3846.58	-6897.80
	EST	-10744.56	-8919.46	-14174.78	-12358.20	-13675.30
QTE50	Lower	-14930.00	-14054.00	-17382.50	-16060.32	-16601.17
	Upper	-6559.11	-3784.91	-10967.07	-8656.07	-10749.43
	EST	-13986.37	-13921.67	-15724.50	-14574.74	-14870.22
QTE75	Lower	-17324.71	-18112.40	-19405.01	-17231.87	-17787.45
	Upper	-10648.04	-9730.95	-12043.99	-11917.62	-11952.99
	EST	-19281.70	-17803.98	-19622.80	-19919.27	-13918.65
QTE90	Lower	-34113.33	-29698.79	-32777.39	-35555.01	-26490.32
	Upper	-4450.08	-5909.17	-6468.22	-4283.53	-1346.97

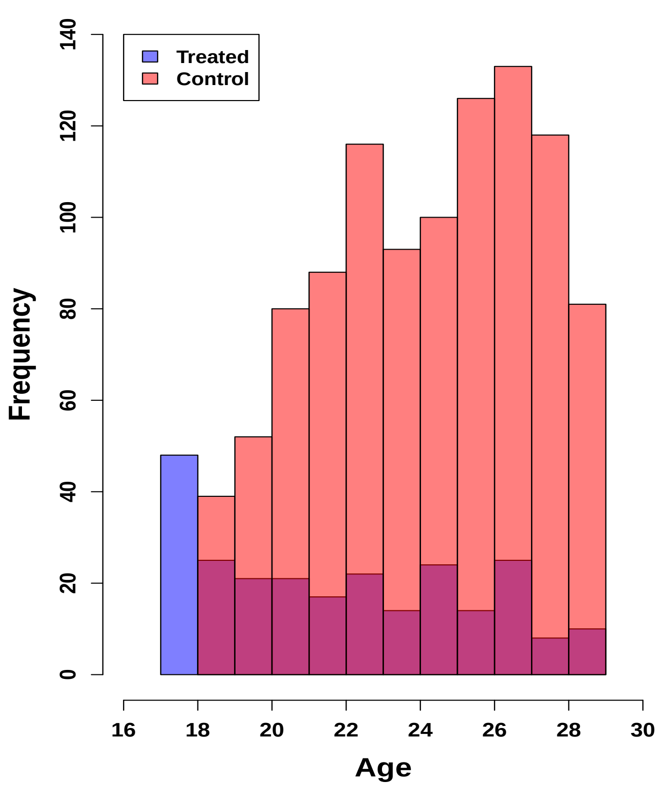

Figure S1: Age distribution of treated and control groups for individuals under age 30 in LLvsPSID Data.

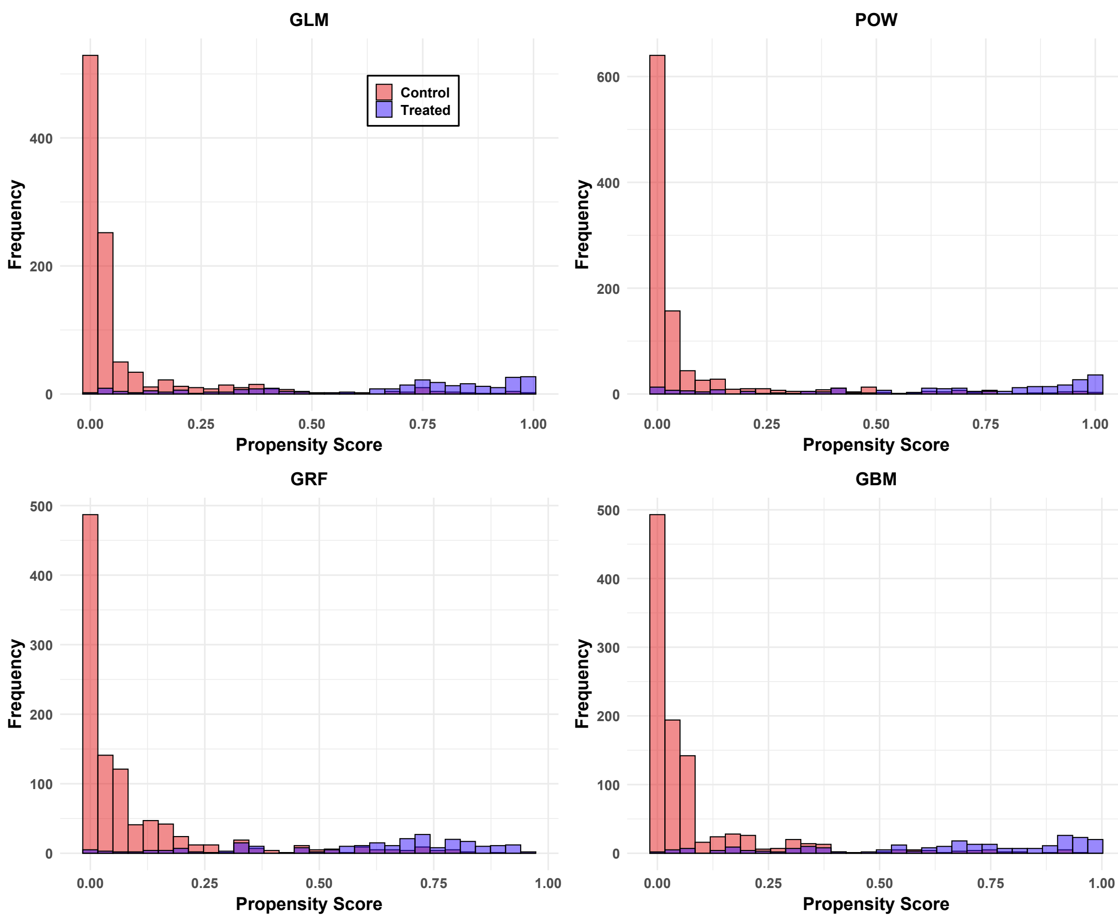

Figure S2: Estimated propensity score distributions using different methods for Individuals under Age 30 in LLvsPSID data.

S10 Additional simulation for conditional average treatment effect

Under the same simulation settings as in Section 4 of the main paper, we further evaluate the estimation of the conditional average treatment effect (CATE), denoted by $\tau(X)$. The IPW method is excluded from this comparison due to its well-documented instability. As shown in the main simulation results, IPW offers no clear advantage in terms of bias or MSE for estimating ATE and ATT under these settings, and performs similarly for CATE estimation. Since ATE and ATT are averages of CATE over specific subpopulations, we focus instead on three more widely used methods for CATE estimation:

(1) **GRF**: the causal forest method proposed by Wager and Athey (2018), implemented in the `R` package `grf` (Tibshirani et al., 2018).

(2) **X-learner**: a meta-learning algorithm for CATE estimation proposed by Künzel et al. (2019). This method is particularly efficient when the treatment and control group sizes are highly imbalanced, making it well-suited for our simulation scenarios. We use BART as the base learner due to the relatively small sample size.

(3) **BART**: The Bayesian additive regression trees method, widely used for causal inference, which also provides credible intervals for CATE estimation (Dorie and Hill, 2017; Dorie et al., 2019).

We assess the performance of the CATE estimators using the expected mean squared error (MSE) for estimating $\tau(X)$ at a randomly drawn test point X, and the empirical coverage probability with a nominal coverage level of 0.95. The simulation settings mirror those in Section 4 of the main paper, except for the **HET** case in Example 1, where the coefficients differ across treatment groups. For each experiment, we generate a training sample of $n = 500$ units to train the CATE estimators and assess their performance on a test set of 10,000 units with known true CATE values. Each experiment is repeated 50 times, and the reported results are averaged over these replications. This evaluation procedure follows the approach used by Künzel et al. (2019) and Wager and Athey (2018) for calculating the MSE of CATE estimators.

Figure S3 reports the MSE performance of six CATE estimators. In the two leftmost columns of subplots, all methods achieve low MSEs across both dimensions, except for the F-PAR method. The performance of SPLRM, T-PAR, and F-PAR aligns with their corresponding ATE results reported in Tables 1–3 of the main paper. Specifically, under the "**HOM+ps-L**" setting, the MSEs of F-PAR for ATE estimation are 0.348 and 0.485, and under "**HOM+ps-N**" they are 0.163 and 0.175–substantially higher than those of SPLRM and T-PAR. Among the three machine learning-based estimators, GRF consistently achieves lower MSEs compared to BART and the X-learner.

The two rightmost columns of subplots in Figure S3 highlight the clear superiority of our proposed SPLRM method, which consistently achieves the smallest MSE across all scenarios. The MSEs obtained by the BART, X-learner, and GRF methods are comparable, with the X-learner performing slightly better overall. This may be attributed to the X-learner's ability to handle data imbalance effectively, as noted by Künzel et al. (2019). In contrast, the T-PAR method, even under perfect model specification, exhibits consistently larger MSEs than the other methods due to its sensitivity to data imbalance.

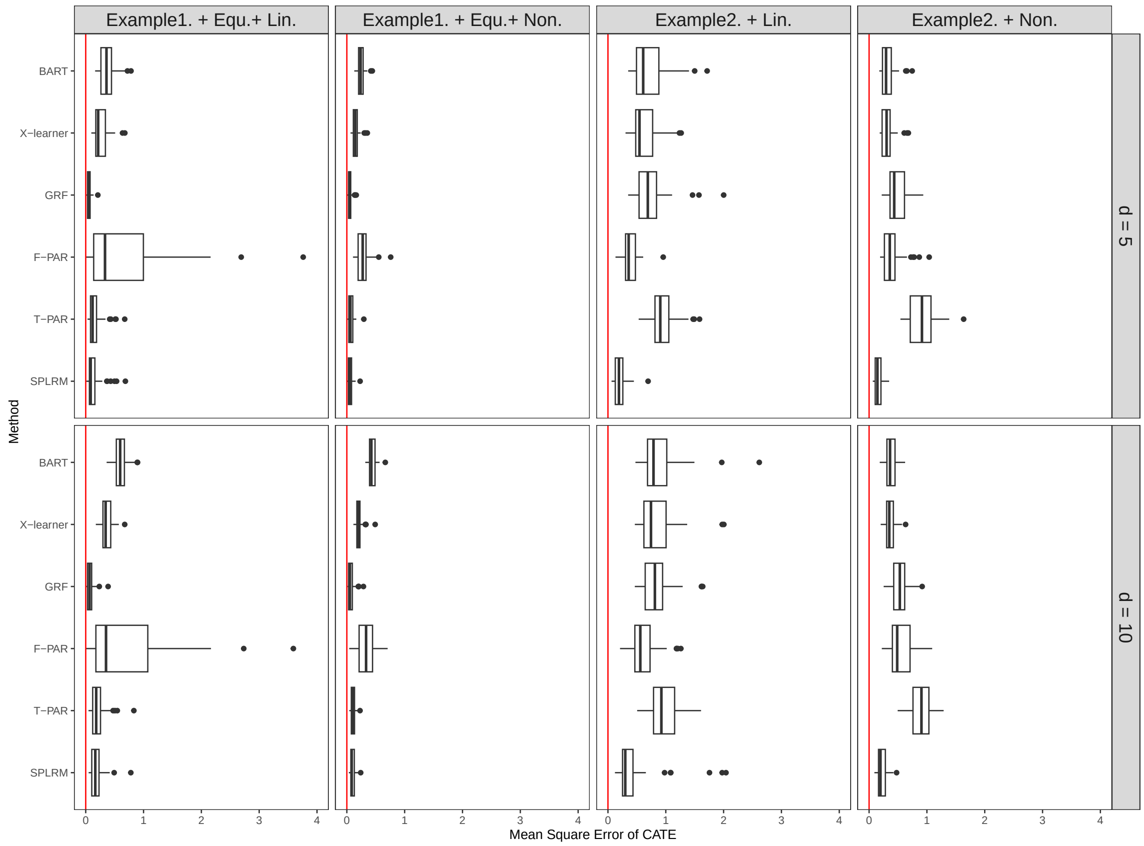

Figure S3: The boxplot of mean square error (MSE) of CATE. Each panel corresponds to one of the four scenarios, except for the case where the $\beta_0 \neq \beta_1$ in Example 1. The abbreviations "Equ.", "Lin.", and "Non." stand for "Equal", "Linear", and "Nonlinear", respectively, and correspond to the **HOM**, **ps-L**, and **ps-N** scenarios introduced in the main paper.

Figure S4 presents the empirical coverage results for CATE, and Figure S5 shows the corresponding interval lengths. We use a nonparametric bootstrap method with $B = 50$ resamples to estimate the variance of CATE for the SPLRM, T-PAR, and F-PAR methods. Our proposed SPLRM method achieves satisfactory coverage across all scenarios while maintaining relatively short interval lengths. However, the interval length of SPLRM increases slightly as the covariate dimension d increases from 5 to

10. BART and X-learner generally provide higher coverage probabilities, but their intervals tend to be conservative. GRF achieves better coverage and shorter intervals than BART and X-learner in Example 1; however, it fails to maintain adequate coverage in Example 2. For the T-PAR method, the intervals provide reasonable coverage with short lengths in the first two scenarios, but severe under-coverage is observed in the last two scenarios, again highlighting its inability to handle data imbalance effectively.

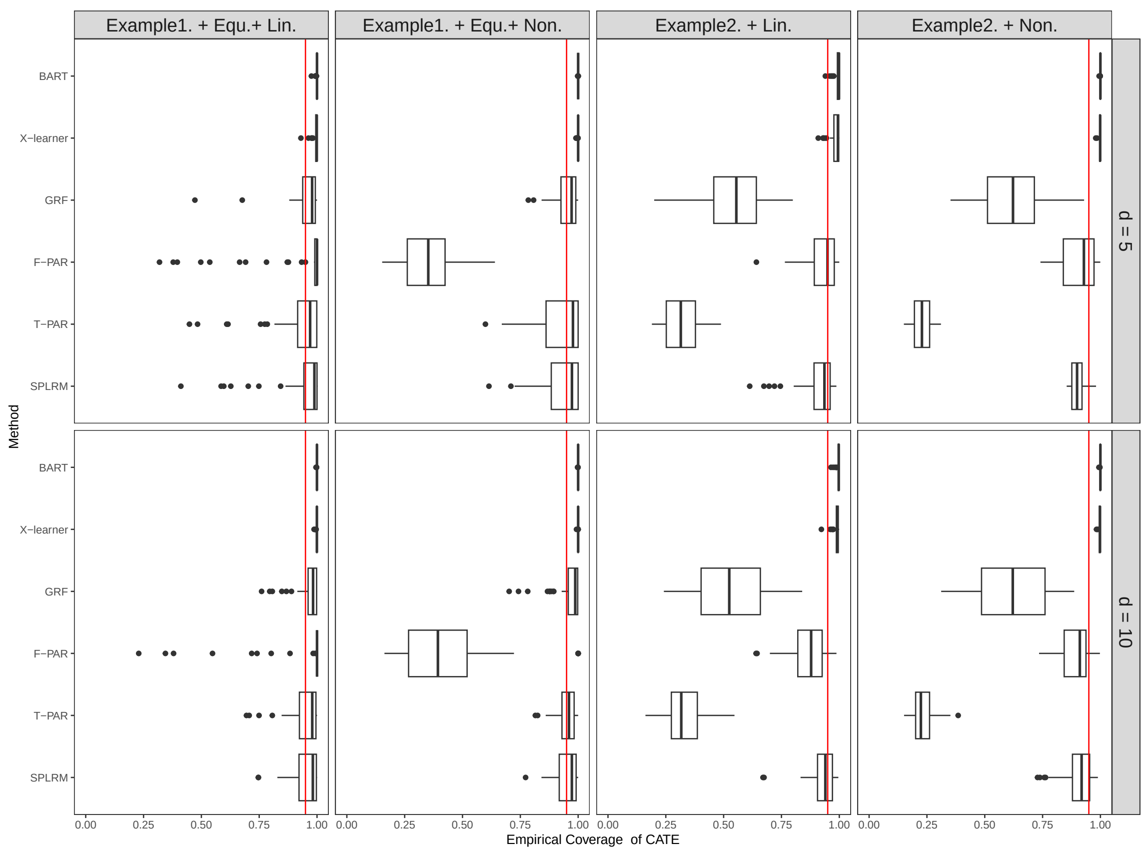

Figure S4: Empirical 95% coverage of CATE. The solid red line is the 95% coverage level value.

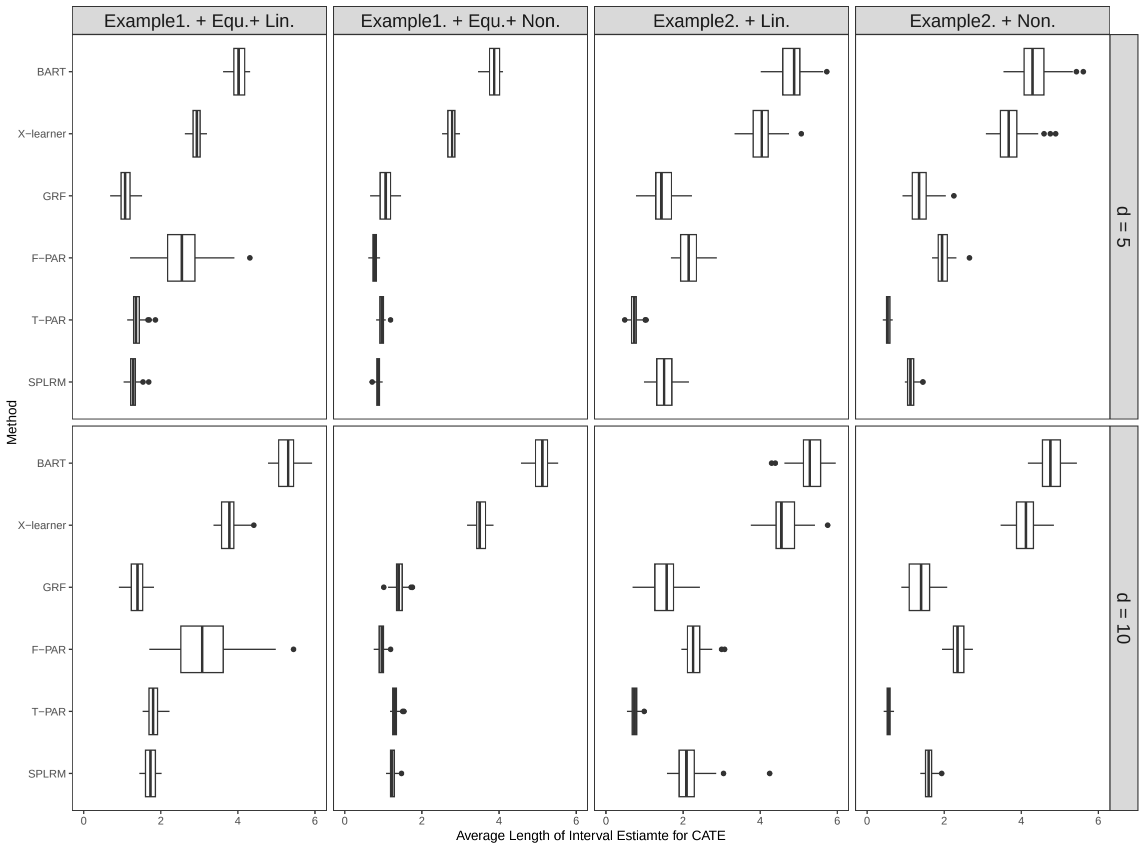

Figure S5: Average length of interval estimate for CATE, which corresponds to the results in Figure S4.

Bibliography

Cheng, M., Y. Liu, and J. Qin (2025). A novel semiparametric approach to non-ignorable missing data by catching covariate marginal information. *Scandinavian Journal of Statistics 52*(2), 691–709.

Chernozhukov, V., I. Fernández-Val, and B. Melly (2013). Inference on counterfactual distributions. *Econometrica 81*(6), 2205–2268.

Dorie, V. and J. Hill (2017). aciccomp2016: Atlantic causal inference conference competition 2016 simulation. *R package Version 0.1-0 4*, 19.

Dorie, V., J. Hill, U. Shalit, M. Scott, and D. Cervone (2019). Automated versus do-it-yourself methods for causal inference: Lessons learned from a data analysis competition. *Statistical Science 34*, 43–68.

Doss, H. and R. D. Gill (1992). An elementary approach to weak convergence for quantile processes, with applications to censored survival data. *Journal of the American Statistical Association 87*(419), 869–877.

Künzel, S. R., J. S. Sekhon, P. J. Bickel, and B. Yu (2019). Metalearners for estimating heterogeneous treatment effects using machine learning. *Proceedings of the National Academy of Sciences of the United States of America 116*(10), 4156–4165.

Liu, Y., P. Li, L. Song, K. Yu, and J. Qin (2021). Retrospective versus prospective score tests for genetic association with case-control data. *Biometrics 77*(1), 102–112.

Luo, X. and W. Y. Tsai (2012). A proportional likelihood ratio model. *Biometrika 99*(1), 211–222.

Murphy, S., A. Rossini, and A. W. van der Vaart (1997). Maximum likelihood estimation in the proportional odds model. *Journal of the American Statistical Association 92*(439), 968–976.

Tibshirani, J., S. Athey, R. Friedberg, V. Hadad, D. Hirshberg, L. Miner, E. Sverdrup, S. Wager, M. Wright, and M. J. Tibshirani (2018). Package grf. *Comprehensive R Archive Network*.

Wager, S. and S. Athey (2018). Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association 113*(523), 1228–1242.