Optimal Covariate Balancing Conditions in Propensity Score Estimation

Jianqing Fan, Kosuke Imai, Inbeom Lee, Han Liu, Yang Ning & Xiaolin Yang

April 18, 2023

Outline

Content

Framework

- observe $(T_i, Y_i, \mathbf{X}_i)$: binary treatment, outcome, and covariates
- common assumptions
  1. SUTVA
  $$Y_i = T_i Y_i(1) + (1 - T_i) Y_i(0),$$
  2. no unmeasured confounders
  $$\{Y^*(1), Y^*(0)\} \perp T_i | \mathbf{X}_i,$$
  3. positivity
  $$0 < \mathbf{P}(T_i = 1 | \mathbf{X}_i) < 1,$$
- parameter of interest: ATE
$$\mu = \mathbf{E}(Y_i(1) - Y_i(0))$$

Framework

- Propensity Score
$$\pi(\mathbf{X}_i) = \mathbf{P}(T_i = 1|\mathbf{X}_i),$$
where a popular parametric choice is $\pi_{\boldsymbol{\beta}} = exp(\mathbf{X}_i^{\intercal}\beta)/\{1 + exp(\mathbf{X}_i^{\intercal}\beta)\}$.
- conditional mean functions
$$\mathbf{E}(Y_i(0)|\mathbf{X}_i) = K(\mathbf{X}_i),$$
and
$$\mathbf{E}(Y_i(1)|\mathbf{X}_i) = K(\mathbf{X}_i) + L(\mathbf{X}_i).$$

Research Puzzles

thinking about ...

- Inverse Propensity Weight Estimator
- Doubly Robust Estimator
- AIPW
- Calibration (Covariate Balancing)
- Projection
- Bias Correction
- ...

Content

IPTW: starting point

- IPTW estimator of ATE:

$$\hat{\mu}_{\boldsymbol{\beta}} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{T_i Y_i}{\pi_{\hat{\boldsymbol{\beta}}}} - \frac{(1-T_i)Y_i}{1-\pi_{\hat{\boldsymbol{\beta}}}}\right)$$

- IPTW estimator does not attain the semiparametric efficiency bound with known propensity score.
- can be sensitive to the misspecification of the propensity score model and the outcome model.

Calibration: a general form

Estimating β with calibration constraints and GMM:

- Define the m-dimensional estimating equation as

$$\bar{g}_{\beta} = \frac{1}{n}\sum_{i=1}^{n} g_{\beta}(T_i, \mathbf{X}_i) = 0,$$

where $g_{\beta}(T_i, \mathbf{X}_i) = \left(\frac{T_i}{\pi_{\beta}} - \frac{(1-T_i)}{1-\pi_{\beta}}\right) f(\mathbf{X}_i)$.

- Solving GMM ($m¿q$) and obtain

$$\hat{\beta} = argmin \quad \bar{\mathbf{g}}_{\beta}(\mathbf{T}, \mathbf{X})^{\intercal}\hat{\mathbf{W}}\bar{\mathbf{g}}_{\beta}(\mathbf{T}, \mathbf{X}).$$

Calibration: choice of covariate balancing function

Remarks for $f(\cdot) : \mathbb{R}^d \mapsto \mathbb{R}^m$:

- Imai and Ratkociv (2014) point out that the common practice of fitting a logistic model is equivalent to balancing the score function

$$f(\mathbf{X}_i) = \pi'_{\boldsymbol{\beta}}(\mathbf{X}_i) = \partial \pi_{\boldsymbol{\beta}}(\mathbf{X}_i)/\partial \boldsymbol{\beta};$$

- Another common choice $f(\mathbf{X}_i) = \mathbf{X}_i$, which balances the first moment between the treatment and control groups, significantly reduces the bias of the estimated ATE;
- Some higher moments and/or interactions include $f(\mathbf{X}_i) = (\mathbf{X}_i, \mathbf{X}_i^2)$.

Content

Bias Elimination: investigation on local misspecification

Assume the true propensity score $\pi(\mathbf{X}_i)$ is related to the working model $\pi_{\boldsymbol{\beta}}(\mathbf{X}_i)$ through the exponential tilt for some $\boldsymbol{\beta}^*$,

$$\pi\left(\mathbf{X}_i\right) = \pi_{\boldsymbol{\beta}^*}\left(\mathbf{X}_i\right) \exp\left(\xi u\left(\mathbf{X}_i; \boldsymbol{\beta}^*\right)\right),$$

where $u\left(\mathbf{X}_i; \boldsymbol{\beta}^*\right)$ is a function determining the direction of misspecification and $\xi \in \mathbb{R}$ represents the magnitude of misspecification.
We assume $\xi = o(1)$ as $n \to \infty$ so that the true propensity score $\pi(\mathbf{X}_i)$ is in a local neighborhood of the working model $\pi_{\boldsymbol{\beta}^*}$.

Bias Elimination: investigation on local misspecification

Under Assumption B.1, the asymptotic bias of $\hat{\mu}_{\boldsymbol{\beta}}$ is given by

$$B = \left\{ \mathbb{E}\left[\frac{u(\mathbf{X}_i; \boldsymbol{\beta}^*) \{K(\mathbf{X}_i) + L(\mathbf{X}_i)(1 - \pi_{\beta^*}(\mathbf{X}_i))\}}{1 - \pi_{\beta^*}(\mathbf{X}_i)} \right] \right. \\ \left. + \boldsymbol{H}_y^* \left(\boldsymbol{H}_{\mathrm{f}}^{*\top} \mathbf{W}^* \boldsymbol{H}_{\mathrm{f}}^* \right)^{-1} \boldsymbol{H}_{\mathrm{f}}^{*\top} \mathbf{W}^* \mathbb{E}\left(\frac{u(\mathbf{X}_i; \boldsymbol{\beta}^*) \mathbf{f}(\mathbf{X}_i)}{1 - \pi_{\beta^*}(\mathbf{X}_i)} \right) \right\} \tag{1}$$

The detailed expression of the theory is defined in Theorem 2.1 (p99).

- Our goal: minimize $|B|$.

Bias Elimination: theoretical optimum

Corollary 2.1

Suppose $m = q$, and $f(\mathbf{X}_i)$ satisfies

$$\boldsymbol{\alpha}^{\intercal}\mathbf{f}\left(\mathbf{X}_i\right) = \pi_{\beta^*}\left(\mathbf{X}_i\right)\mathbb{E}\left(Y_i(0) \mid \mathbf{X}_i\right) + \left(1 - \pi_{\beta^*}\left(\mathbf{X}_i\right)\right)\mathbb{E}\left(Y_i(1) \mid \mathbf{X}_i\right). \quad (2)$$

Then, under the conditions in Theorem 2.1, the asymptotic bias is eliminated, i.e., $B = 0$.

Corollary 2.2

Under the same conditions in Corollary 2.1, the asymptotic variance of $\hat{\mu}_{\boldsymbol{\beta}}$ is minimized by any $f(\mathbf{X}_i)$ which satisfies condition (2). In this case, $\hat{\mu}_{\boldsymbol{\beta}}$ attains the semiparametric asymptotic variance bound and

$$V_{\mathrm{opt}} = \mathbb{E}\left[\frac{\operatorname{var}\left(Y_i(1) \mid X_i\right)}{\pi\left(X_i\right)} + \frac{\operatorname{var}\left(Y_i(0) \mid X_i\right)}{1 - \pi\left(X_i\right)} + \left\{L\left(X_i\right) - \mu\right\}^2\right].$$

Bias Elimination: theoretical optimum

Remark:

- (2) is referred to as the optimality condition, which implies that the weighted average of the potential outcome lies in the space spanned by $f(\mathbf{X}_i)$.
- However, some initial estimator for $\boldsymbol{\beta}^*$ is required and the empirical performance of the plug-in estimator is often unstable due to the estimation error of $\boldsymbol{\beta}^*$.
- An alternative way to construct an optimal estimator is proposed, which does not require any initial estimator.

Bias Elimination: one-step further

Note that the optimality condition for calibration function $f(\mathbf{X}_i)$ is equivalent to

$$\begin{aligned}\alpha^{\intercal} g_{\beta^*}\left(T_i, X_i\right) &= \left(\frac{T_i}{\pi_{\beta^*(X_i)}} - \frac{1-T_i}{1-\pi_{\beta^*(X_i)}}\right) \\ &\quad \times \left[\pi_{\beta^*}\left(X_i\right) K\left(X_i\right) + \left(1-\pi_{\beta^*}\left(X_i\right)\right)\left(K\left(X_i\right)+L\left(X_i\right)\right)\right] \\ &= \left(\frac{T_i}{\pi_{\beta^*(X_i)}} - \frac{1-T_i}{1-\pi_{\beta^*(X_i)}}\right) K\left(X_i\right) \\ &\quad + \left(\frac{T_i}{\pi_{\beta^*(X_i)}} - 1\right) L\left(X_i\right).\end{aligned} \tag{3}$$

Following this observation, the calibration estimating functions could then be constructed separately by

$$\bar{g}_\beta(T, X) = \begin{pmatrix} \bar{g}_{1\beta}(T, X) \\ \bar{g}_{2\beta}(T, X) \end{pmatrix}$$

Bias Elimination: one-step further

Balancing $K(\mathbf{X}_i)$ and $L(\mathbf{X}_i)$ separately with $h_1(\cdot) : \mathbb{R}^d \mapsto \mathbb{R}^{m_1}$, $h_2(\cdot) : \mathbb{R}^d \mapsto \mathbb{R}^{m_2}$:

- $g_{1\beta}\left(T_i, \mathbf{X}_i\right) = \left(\frac{T_i}{\pi_\beta(\mathbf{X}_i)} - \frac{1-T_i}{1-\pi_\beta(\mathbf{X}_i)}\right) \boldsymbol{h}_1\left(\mathbf{X}_i\right)$,
- $g_{2\beta}\left(T_i, \mathbf{X}_i\right) = \left(\frac{T_i}{\pi_\beta(\mathbf{X}_i)} - 1\right) \boldsymbol{h}_2\left(\mathbf{X}_i\right)$

As long as $K(\cdot)$ and $L(\cdot)$ lie in the linear space spanned by $\mathbf{h}_1(\cdot)$ and $\mathbf{h}_2(\cdot)$ respectively, we no longer require an initial estimate of $\boldsymbol{\beta}$ or the conditional mean models.

Optimal Covariate Balancing: interpretation

- $g_{1\beta}\left(T_i, \mathbf{X}_i\right)$ is the same as the existing covariate balancing moment functions, which balances the covariate $\mathbf{h}_1(\mathbf{X}_i)$ between the treatment and control groups.
- $g_{2\beta}\left(T_i, \mathbf{X}_i\right)$ matches the weighted covariates $\mathbf{h}_2(\mathbf{X}_i)$ in the treatment group and unweighted covariates $\mathbf{h}_2((\mathbf{X}_i)$ in the control group:

$$\sum_{T_i=1} \frac{1-\pi_\beta\left(X_i\right)}{\pi_\beta\left(\mathbf{X}_i\right)} \boldsymbol{h}_2\left(\mathbf{X}_i\right) = \sum_{T_i=0} \boldsymbol{h}_2\left(\mathbf{X}_i\right)$$

Optimal Covariate Balancing: double robustness

Theorem 3.1

Under Assumption 3.1, the proposed oCBPS-based IPTW estimator $\hat{\mu}_{\hat{\beta}}$ is doubly robust.
That is, $\hat{\mu}_{\hat{\beta}} \to \mu$ if as least one of the following two conditions holds:

1. (propensity score model) $\mathbb{P}(T_i = 1|\mathbf{X}_i) = \pi_{\beta^0}(\mathbf{X}_i)$;
2. (outcome model) $\mathbf{h}_1(\cdot)$, $\mathbf{h}_2(\cdot)$, and $\mathbf{W}^*$ in Assumption 3.1 satisfies:. $\exists \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2$ such that $K(\mathbf{X}_i) = \boldsymbol{\alpha}_1 \mathbf{M}_1 \mathbf{h}_1(\mathbf{X}_i)$ and $L(\mathbf{X}_i) = \boldsymbol{\alpha}_2 \mathbf{M}_2 \mathbf{h}_2(\mathbf{X}_i)$, where $\mathbf{M}_1, \mathbf{M}_2$ are partitions of $\mathbf{G}^{*\intercal} \mathbf{W}^* = (\mathbf{M}_1, \mathbf{M}_2)$.

Optimal Covariate Balancing: theoretical remarks

- If both condition 1 and 2 holds, the asymptotic variance

$$V = \Sigma_\mu - (\boldsymbol{\alpha}_1^\intercal \mathbf{M}_1, \boldsymbol{\alpha}_2^\intercal \mathbf{M}_2)\, \mathbf{G}^* \left(\mathbf{G}^{*\intercal} \boldsymbol{\Omega}^{-1} \mathbf{G}^*\right)^{-1} \mathbf{G}^{*\intercal} \begin{pmatrix} \mathbf{M}_1^\intercal \boldsymbol{\alpha}_1 \\ \mathbf{M}_2^\intercal \boldsymbol{\alpha}_2 \end{pmatrix}$$

  where Σ_μ is the variance of the IPW estimator with true propensity score, and the second term is the effect of estimating β via covariate balance conditions.
- We can make the proposed estimator more robust by incorporating more basis functions into $\mathbf{h}_1(\cdot)$ and $\mathbf{h}_2(\cdot)$, such that the optimality condition is more likely to hold. However, doing so may inflate the variance of the proposed estimator(Corollary 3.1).
- $\hat{\mu}_{\hat{\beta}}$ is semiparametrically efficient if $\mathbf{G}^*$ is a square matrix $(m = q)$ and invertible,

$$V_{\text{opt}} = \Sigma_\mu - (\boldsymbol{\alpha}_1^\intercal \mathbf{M}_1, \boldsymbol{\alpha}_2^\intercal \mathbf{M}_2)\, \boldsymbol{\Omega} \begin{pmatrix} \mathbf{M}_1^\intercal \boldsymbol{\alpha}_1 \\ \mathbf{M}_2^\intercal \boldsymbol{\alpha}_2 \end{pmatrix}$$

Content

Compare with AIPW

- Recall that

$$\begin{aligned}\hat{\mu}^{\mathrm{AIPW}}_{\beta,\alpha,\gamma} =& \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{T_i Y_i}{\pi_\beta(\mathbf{X}_i)} - \frac{(1-T_i)\,Y_i}{1-\pi_\beta(\mathbf{X}_i)}\right.\\ &\left. - (T_i - \pi_\beta(\mathbf{X}_i))\left(\frac{K(\mathbf{X}_i,\alpha)+L(\mathbf{X}_i,\gamma)}{\pi_\beta(\mathbf{X}_i)} + \frac{K(\mathbf{X}_i,\alpha)}{1-\pi_\beta(\mathbf{X}_i)}\right)\right\}.\end{aligned}$$

  It's interesting to note that the proposed estimator could be rewritten as an AIPW estimator if $K(\mathbf{X}_i,\alpha) = \boldsymbol{\alpha}^\intercal \mathbf{h}_1(\mathbf{X}_i)$ and $K(\mathbf{X}_i,\gamma) = \boldsymbol{\gamma}^\intercal \mathbf{h}_2(\mathbf{X}_i)$ $(m = q)$.
- Both the AIPW estimator and the proposed method are doubly robust and locally efficient, but the proposed method converges at a faster rate when both propensity score and outcome models are *locally* misspecified.

Compare with AIPW

Locally Misspecification ($\xi, \delta = o(1)$):

- $\pi(\mathbf{X}_i) = \pi_{\boldsymbol{\beta}^*}(\mathbf{X}_i) \exp(\xi u(\mathbf{X}_i; \boldsymbol{\beta}^*))$,
- $K(\mathbf{X}_i, \alpha) = \boldsymbol{\alpha}^\intercal \mathbf{h}_1(\mathbf{X}_i) + \delta r_1(\mathbf{X}_i)$, $L(\mathbf{X}_i, \alpha) = \boldsymbol{\gamma}^\intercal \mathbf{h}_2(\mathbf{X}_i) + \delta r_2(\mathbf{X}_i)$

Corresponding estimator:

- the proposed oCBPS estimator:

$$\hat{\mu}_{\widehat{\beta}} - \mu = \phi^{eff}(\mathbf{X}_i) + O_p\left(\xi^2\delta + \delta n^{-1/2} + \xi n^{-1/2}\right);$$

- the AIPW estimator:

$$\hat{\mu}^{\mathrm{AIPW}}_{\tilde{\beta},\tilde{\alpha},\tilde{\gamma}} - \mu = \phi^{eff}(\mathbf{X}_i) + O_p\left(\xi\delta + \delta n^{-1/2} + \xi n^{-1/2}\right)$$

where the leading terms

$$\phi^{eff}(\mathbf{X}_i) = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{T_i}{\pi(\mathbf{X}_i)}\{Y_i(1) - K(\mathbf{X}_i) - L(\mathbf{X}_i)\} - \frac{1-T_i}{1-\pi(\mathbf{X}_i)}\{Y_i(0) - K(\mathbf{X}_i)\} + L(\mathbf{X}_i) - \mu\right]$$

in the asymptotic expansions are identical and are known as the efficient influence function.

Nonparametric oCBPS

- It's a natural idea to extend $\mathbf{h}_1(\cdot)$, $\mathbf{h}_2(\cdot)$ to a large number of basis functions.
- The parametric assumption for the propensity score model $\mathbb{P}(T_i = 1|\mathbf{X}_i) = \pi_{\beta^0}(\mathbf{X}_i)$ is too restrictive (which requires $m = q$ to achieve optimal variance).
- To release the assumption, suppose $\mathbb{P}(T_i = 1|\mathbf{X}_i) = J(\psi^*(\mathbf{X}_i))$, where $J(\cdot)$ is a known link function and ψ^* is an unknown smooth function. Then we could approximate $\psi^*(\mathbf{x})$ by $\boldsymbol{\beta}^{*\top}B(\boldsymbol{x})$ for some $\boldsymbol{\beta}^* \in \mathbb{R}^\kappa$ and allow κ to grow with n.

Content

Data Generation

The following setting is used to estimate the propensity score:

- Potential Outcome

$$Y_i(1) = 200 + 27.4X_{i1} + 13.7X_{i2} + 13.7X_{i3} + 13.7X_{i4} + \varepsilon_i,$$
$$Y_i(0) = 200 + 13.7X_{i2} + 13.7X_{i3} + 13.7X_{i4} + \varepsilon_i;$$

- Propensity Score

$$\mathbb{P}(T_i = 1 \mid X_i = x_i) = \frac{\exp(-\beta_1 x_{i1} + 0.5x_{i2} - 0.25x_{i3} - 0.1x_{i4})}{1 + \exp(-\beta_1 x_{i1} + 0.5x_{i2} - 0.25x_{i3} - 0.1x_{i4})},$$

  where β_1 varies from 0 to 1.

- Calibration function: $\mathbf{h}_1(\boldsymbol{x}_i) = (1, x_{i2}, x_{i3}, x_{i4})$ and $\mathbf{h}_2(\boldsymbol{x}_i) = x_{i1}$;
- Covariate: $X_{i1} \sim \mathcal{N}(3, 2)$, $X_{i2}, X_{i3}, X_{i4} \sim \mathcal{N}(0, 1)$.

Each set of results is based on 500 Monte Carlo simulations.

Data Generation

To evaluate the method performance under misspecification, the actual data maybe generated differently.

- local misspecification of Propensity Score: set $\xi = n^{-1/2}$ and $u(\mathbf{X}_i, \boldsymbol{\beta}) = X_{i2}^2$;
- arbitrary misspecification of Propensity Score:

$$\mathbb{P}\left(T_i = 1 \mid X_i = x_i\right) = \frac{\exp\left(-\beta_1 x_{i1}^* + 0.5x_{i2}^* - 0.25x_{i3}^* - 0.1x_{i4}^*\right)}{1 + \exp\left(-\beta_1 x_{i1} + 0.5x_{i2}^* - 0.25x_{i3}^* - 0.1x_{i4}^*\right)},$$

  with $x_{i1}^* = exp(x_{i1}/3)$, $x_{i2}^* = x_{i2}/\{1 + exp(x_{i1})\} + 10$, $x_{i3}^* = x_{i1}x_{i3}/25 + 0.6$, $x_{i4}^* = xi1 + x_{i4} + 20$;
- Potential Outcome

$$Y_i(1) = 200 + 27.4X_{i1}^2 + 13.7X_{i2}^2 + 13.7X_{i3}^2 + 13.7X_{i4}^2 + \varepsilon_i,$$
$$Y_i(0) = 200 + 13.7X_{i2}^2 + 13.7X_{i3}^2 + 13.7X_{i4}^2 + \varepsilon_i;$$

.

Result: both correctly specified

Table 1. The bias, standard deviation, root mean squared error (RMSE), and the coverage probability of the constructed 95% C.I. of the IPTW estimator with known propensity score (True), the IPTW estimator when the propensity score is fitted using the maximum likelihood (GLM), the IPTW estimator when the propensity score is fitted using the generalized additive model (GAM), the targeted maximum likelihood estimator (DR), the standard CBPS estimator balancing the first moment (CBPS), and the proposed optimal CBPS estimator (oCBPS) under the scenario that both the outcome model and the propensity score model are correctly specified.

		$n = 300$				$n = 1000$			
	β_1	0	0.33	0.67	1	0	0.33	0.67	1
Bias	True	−0.43	−0.01	1.15	−5.19	0.00	0.09	−2.43	9.99
	GLM	−0.18	−0.86	0.15	−4.32	−0.04	0.02	0.32	11.15
	GAM	−0.74	−4.60	−15.55	−35.38	−0.19	−1.16	−2.85	−6.86
	DR	0.08	−1.04	−3.41	−8.32	0.18	−0.56	−2.14	−4.50
	CBPS	−0.05	−0.09	0.54	−0.27	0.04	0.04	0.20	0.45
	oCBPS	−0.04	0.03	0.07	0.06	0.04	0.06	0.16	0.08
Std Dev	True	29.52	39.46	72.56	138.33	15.73	22.36	38.18	88.33
	GLM	4.45	12.31	63.35	144.25	2.21	5.49	22.93	114.45
	GAM	4.31	14.91	43.08	100.16	2.06	5.22	21.27	51.96
	DR	2.39	2.57	4.25	8.06	1.20	1.29	1.76	3.32
	CBPS	2.39	2.35	2.66	15.94	1.24	1.26	1.27	1.45
	oCBPS	2.26	2.16	2.27	2.39	1.20	1.20	1.18	1.22
RMSE	True	29.52	39.46	72.57	138.43	15.73	22.36	38.26	88.89
	GLM	4.46	12.34	63.35	144.32	2.21	5.49	22.93	114.99
	GAM	4.37	15.60	45.81	106.23	2.07	5.35	21.46	52.41
	DR	2.39	2.77	5.45	11.58	1.21	1.41	2.77	5.59
	CBPS	2.39	2.35	2.72	15.94	1.24	1.26	1.29	1.52
	oCBPS	2.26	2.16	2.27	2.39	1.20	1.20	1.19	1.23
Coverage Probability (of the 95% C.I.)	True	0.936	0.938	0.922	0.948	0.962	0.942	0.926	0.948
	GLM	0.946	0.946	0.946	0.946	0.944	0.954	0.954	0.958
	GAM	0.704	0.310	0.090	0.028	0.754	0.382	0.108	0.048
	DR	0.928	0.876	0.576	0.278	0.960	0.906	0.562	0.268
	CBPS	0.944	0.944	0.944	0.944	0.960	0.958	0.958	0.968
	oCBPS	0.950	0.964	0.962	0.982	0.956	0.954	0.962	0.966

NOTE: We vary the value of β_1 in the data-generating model (5.1).

Result: misspecified propensity score

Table 2. Correct outcome model with a misspecified propensity score model.

		$n = 300$				$n = 1000$			
	β_1	0	0.33	0.67	1	0	0.33	0.67	1
Bias	True	0.00	2.13	0.08	4.79	−1.28	−0.36	1.83	3.62
	GLM	0.41	−6.67	−18.84	−32.15	0.19	−6.33	−19.21	−32.96
	GAM	15.61	3.11	−7.16	−20.76	4.07	0.28	−4.98	−14.11
	DR	−0.29	−0.68	−1.89	−3.60	−0.21	−0.39	−1.23	−2.75
	CBPS	0.84	−0.05	−2.06	−2.44	0.06	−0.79	−2.74	−3.28
	oCBPS	−0.20	−0.02	−0.13	0.07	−0.04	0.03	0.01	−0.05
Std Dev	True	45.43	36.03	39.77	77.26	26.32	19.36	39.15	88.45
	GLM	11.23	12.66	15.73	26.82	2.17	5.32	8.61	10.92
	GAM	19.91	9.40	8.81	16.18	4.29	2.87	4.14	8.52
	DR	3.35	2.57	2.52	3.16	1.42	1.27	1.28	1.57
	CBPS	3.21	2.74	3.18	3.61	1.25	1.41	1.74	2.04
	oCBPS	2.26	2.30	2.28	2.34	1.24	1.26	1.24	1.29
RMSE	True	45.43	36.10	39.77	77.40	26.36	19.36	39.20	88.52
	GLM	11.24	14.31	24.55	41.86	2.18	8.27	21.05	34.72
	GAM	25.30	9.90	11.35	26.32	5.91	2.89	6.47	16.48
	DR	3.37	2.65	3.15	4.79	1.44	1.33	1.78	3.16
	CBPS	3.32	2.74	3.79	4.36	1.26	1.62	3.24	3.86
	oCBPS	2.27	2.30	2.29	2.34	1.24	1.26	1.24	1.29
Coverage probability (of the 95% C.I.)	True	0.952	0.936	0.964	0.972	0.946	0.950	0.960	0.988
	GLM	0.964	0.898	0.740	0.834	0.948	0.714	0.300	0.346
	GAM	0.236	0.434	0.286	0.066	0.356	0.648	0.178	0.042
	DR	0.882	0.904	0.822	0.596	0.908	0.938	0.788	0.392
	CBPS	0.956	0.978	0.924	0.914	0.944	0.928	0.742	0.654
	oCBPS	0.946	0.944	0.952	0.944	0.950	0.950	0.954	0.954

Result: locally misspecified propensity score

Table 3. Correctly specified outcome with a locally misspecified propensity score model.

		$n = 300$				$n = 1000$			
	β_1	0	0.33	0.67	1	0	0.33	0.67	1
Bias	True	−1.96	0.69	0.80	4.87	0.04	0.87	−0.42	3.07
	GLM	−16.73	8.43	5.85	19.96	8.55	0.84	4.65	21.07
	GAM	−8.19	7.68	−4.35	−10.79	4.62	−0.25	−0.63	2.95
	DR	0.43	0.34	−0.83	−3.67	0.38	0.08	−1.39	−3.50
	CBPS	−0.76	−2.15	0.56	1.34	−1.92	−0.34	0.22	0.37
	oCBPS	−0.41	0.05	0.10	0.06	−0.05	0.02	−0.01	−0.02
Std Dev	True	41.03	33.16	41.86	82.09	20.65	18.39	28.44	59.63
	GLM	67.79	9.55	23.67	72.99	9.43	3.23	13.86	81.20
	GAM	46.08	8.92	21.56	52.34	11.06	2.91	11.78	52.31
	DR	3.10	2.51	2.87	5.74	1.37	1.29	1.59	2.60
	CBPS	3.26	2.56	2.44	2.77	1.58	1.28	1.33	1.43
	oCBPS	2.47	2.24	2.25	2.26	1.29	1.22	1.26	1.29
RMSE	True	41.07	33.17	41.87	82.24	20.65	18.41	28.44	59.70
	GLM	69.82	12.74	24.39	75.67	12.73	3.34	14.62	83.89
	GAM	46.80	11.77	21.99	53.44	11.98	2.92	11.80	52.39
	DR	3.13	2.53	2.99	6.81	1.42	1.29	2.11	4.36
	CBPS	3.35	3.34	2.51	3.07	2.49	1.32	1.34	1.48
	oCBPS	2.50	2.24	2.26	2.27	1.29	1.22	1.26	1.29
Coverage probability (of the 95% C.I.)	True	0.962	0.948	0.962	0.938	0.934	0.946	0.954	0.942
	GLM	0.804	0.788	0.888	0.916	0.652	0.936	0.918	0.910
	GAM	0.132	0.294	0.238	0.076	0.154	0.612	0.144	0.052
	DR	0.856	0.922	0.866	0.556	0.916	0.936	0.736	0.332
	CBPS	0.912	0.914	0.926	0.958	0.752	0.954	0.954	0.952
	oCBPS	0.916	0.946	0.936	0.954	0.950	0.948	0.958	0.954

Result: misspecified outcome model

Table 4. Misspecified outcome model with correct propensity score model.

		$n = 300$				$n = 1000$			
	β_1	0	0.13	0.27	0.4	0	0.13	0.27	0.4
Bias	True	−4.37	−0.03	−4.24	1.51	0.80	−1.00	2.31	2.67
	GLM	0.38	−0.64	−2.67	−1.33	0.11	−0.44	0.05	0.75
	GAM	−2.03	−5.49	−10.43	−13.66	−0.65	−1.72	−1.95	−3.04
	DR	−2.77	−5.06	−9.92	−14.36	−2.98	−4.98	−7.43	−10.11
	CBPS	0.07	−0.69	−2.59	−3.94	0.05	−0.55	−0.71	−1.63
	oCBPS	−0.56	−0.97	−3.05	−4.37	−0.03	−0.68	−0.84	−1.70
Std Dev	True	49.87	58.75	74.32	100.35	27.61	33.62	44.75	53.58
	GLM	18.12	24.87	34.83	56.17	9.68	12.37	18.45	31.16
	GAM	17.59	23.19	34.72	49.87	9.07	11.36	16.85	26.50
	DR	14.02	14.65	15.58	16.65	7.95	8.26	8.21	8.40
	CBPS	15.51	17.60	18.83	20.66	8.74	9.47	10.64	12.05
	oCBPS	14.74	16.15	17.13	18.55	8.44	9.03	9.68	10.87
RMSE	True	50.06	58.75	74.45	100.36	27.62	33.64	44.81	53.60
	GLM	18.13	24.88	34.93	56.18	9.68	12.37	18.45	31.17
	GAM	17.71	23.83	36.25	51.71	9.09	11.49	16.96	26.67
	DR	14.29	15.50	18.47	21.99	8.49	9.65	11.07	13.15
	CBPS	15.51	17.62	19.01	21.03	8.74	9.49	10.66	12.16
	oCBPS	14.75	16.18	17.40	19.06	8.44	9.06	9.72	11.00
Coverage probability (of the 95% C.I.)	True	0.948	0.954	0.946	0.920	0.938	0.950	0.910	0.922
	GLM	0.896	0.852	0.870	0.868	0.908	0.862	0.816	0.802
	GAM	0.912	0.832	0.676	0.476	0.932	0.846	0.690	0.516
	DR	0.930	0.910	0.838	0.716	0.924	0.874	0.794	0.688
	CBPS	0.920	0.870	0.790	0.676	0.914	0.862	0.776	0.668
	oCBPS	0.950	0.930	0.908	0.904	0.954	0.920	0.902	0.862

Result: both misspecified

Table 5. Misspecified outcome with misspecified propensity score models.

		$n = 300$				$n = 1000$			
	β_1	0	0.13	0.27	0.4	0	0.13	0.27	0.4
Bias	True	0.54	−1.74	1.71	−3.56	−2.66	−2.52	−2.06	−0.36
	GLM	2.94	−1.70	−8.47	−20.25	−0.18	−2.07	−8.89	−18.79
	GAM	20.74	12.05	3.42	−8.06	4.95	2.35	−1.03	−5.01
	DR	9.16	6.66	4.52	0.46	6.55	4.91	2.80	0.36
	CBPS	9.57	4.10	0.37	−7.62	0.46	−0.81	−4.94	−11.18
	oCBPS	2.51	−0.24	−1.62	−4.82	0.04	−0.61	−2.29	−4.54
Std Dev	True	59.12	55.64	54.16	58.35	34.79	31.31	28.41	31.62
	GLM	25.00	19.44	22.49	26.01	9.67	9.79	11.17	12.44
	GAM	30.85	23.01	19.46	21.72	10.23	9.53	9.19	9.25
	DR	15.18	15.14	13.71	13.60	7.86	7.85	7.69	7.70
	CBPS	26.74	18.65	19.74	18.92	9.16	9.11	9.36	9.63
	oCBPS	16.28	15.38	15.08	14.42	8.93	8.60	8.32	8.27
RMSE	True	59.12	55.66	54.19	58.45	34.89	31.42	28.48	31.62
	GLM	25.18	19.51	24.03	32.96	9.67	10.00	14.28	22.53
	GAM	37.17	25.97	19.76	23.17	11.37	9.81	9.25	10.52
	DR	17.73	16.54	14.43	13.60	10.24	9.26	8.19	7.71
	CBPS	28.40	19.10	19.75	20.40	9.17	9.15	10.59	14.76
	oCBPS	16.47	15.38	15.17	15.20	8.93	8.62	8.63	9.43
Coverage probability (of the 95% C.I.)	True	0.952	0.940	0.936	0.952	0.936	0.940	0.952	0.916
	GLM	0.854	0.902	0.866	0.788	0.890	0.878	0.772	0.540
	GAM	0.714	0.810	0.860	0.832	0.868	0.902	0.916	0.834
	DR	0.878	0.920	0.934	0.946	0.876	0.906	0.936	0.946
	CBPS	0.866	0.892	0.890	0.866	0.894	0.888	0.852	0.670
	oCBPS	0.940	0.964	0.926	0.934	0.944	0.942	0.926	0.894

Reference

- Jianqing Fan, Kosuke Imai, Inbeom Lee, Han Liu, Yang Ning & Xiaolin Yang (2023) Optimal Covariate Balancing Conditions in Propensity Score Estimation, Journal of Business & Economic Statistics, 41:1, 97-110, DOI: 10.1080/07350015.2021.2002159