Paper Review

High-dimensional Covariate Balancing Propensity Score by Ning, Peng, and Imai

Hao Wang

Department of Statistics, Iowa State University

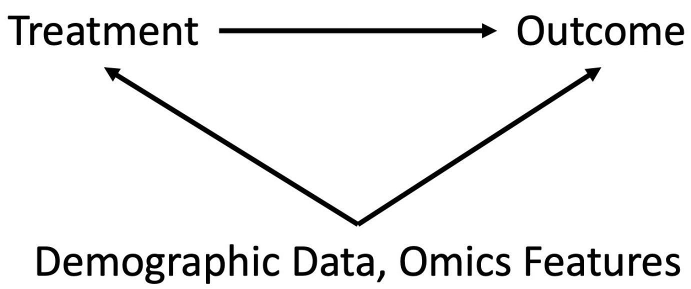
2023-10-10

Motivation

Observation Study

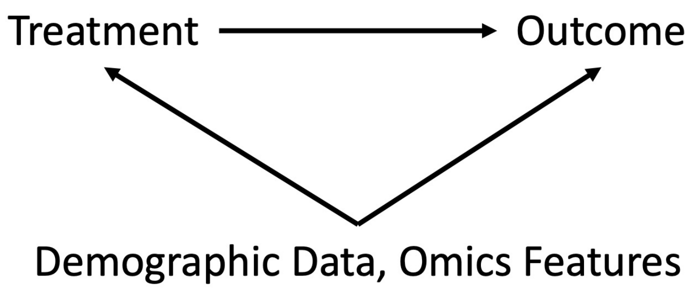
Motivation

Observation Study



Motivation

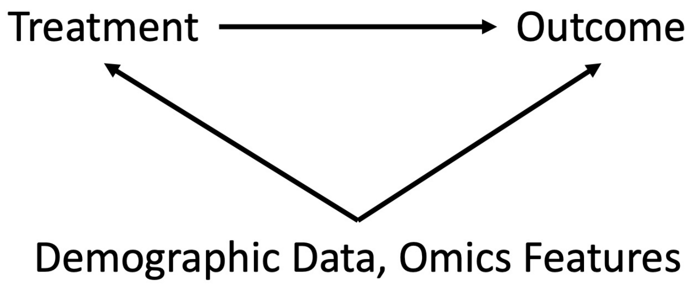
Observation Study



> High-dimensional confounders! $\dim(\mathbf{X}_i) >> N$
>
> How to estimate average treatment effect?
>
> Propensity score? Outcome model?

Motivation

Observation Study

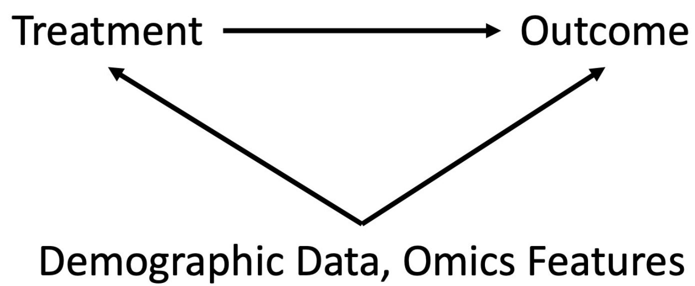


> High-dimensional confounders! $\dim(\mathbf{X}_i) >> N$
>
> How to estimate average treatment effect?
>
> Propensity score? Outcome model?

High-dimensional regression

Motivation

Observation Study



> High-dimensional confounders! $\dim(\mathbf{X}_i) >> N$
>
> How to estimate average treatment effect?
>
> Propensity score? Outcome model?

High-dimensional regression

See Bradic (2019): Rate/sparsity double robust

Background

Covariate Balancing Propensity Score (CBPS)

See Imai (2014): Estimate propensity score so that covariate balance is optimized

Background

Covariate Balancing Propensity Score (CBPS)

See Imai (2014): Estimate propensity score so that covariate balance is optimized

Propensity Score

Conditional probability of treatment assignment

$$\pi(\mathbf{X}_i) = \Pr(T_i = 1|\mathbf{X}_i)$$

Balancing property (without assumption)

$$T_i \perp\!\!\!\perp \mathbf{X}_i|\pi(\mathbf{X}_i)$$

Background

Covariate Balancing Propensity Score (CBPS)

See Imai (2014): Estimate propensity score so that covariate balance is optimized

Propensity Score

Conditional probability of treatment assignment

$$\pi(\mathbf{X}_i) = \Pr(T_i = 1 | \mathbf{X}_i)$$

Balancing property (without assumption)

$$T_i \perp\!\!\!\perp \mathbf{X}_i | \pi(\mathbf{X}_i)$$

Use of Propensity Score for Causal Inference

- Matching
- Subclassification
- Weighting (Horvitz-Thompson)
- Double-robust estimators

Background

Covariate Balancing Propensity Score (CBPS)

See Imai (2014): Estimate propensity score so that covariate balance is optimized

Background

Covariate Balancing Propensity Score (CBPS)

See Imai (2014): Estimate propensity score so that covariate balance is optimized

Propensity Score Tautology

- Propensity score is unknown
- Diagnostics: covariate balance checking
- Misspecification is possible especially for non-binary treatments
- Skewed covariates are common in applied settings
- Propensity score methods can be sensitive to misspecification

Background

Covariate Balancing Propensity Score (CBPS)

See Imai (2014): Estimate propensity score so that covariate balance is optimized

Propensity Score Tautology

- Propensity score is unknown
- Diagnostics: covariate balance checking
- Misspecification is possible especially for non-binary treatments
- Skewed covariates are common in applied settings
- Propensity score methods can be sensitive to misspecification

Covariate balancing condition (C.1)

$$\mathrm{E}[\frac{T_i\tilde{\mathbf{X}}_i}{\pi_\beta(\mathbf{X}_i)} - \frac{(1-T_i)\tilde{\mathbf{X}}_i}{1-\pi_\beta(\mathbf{X}_i)}] = 0,\ \tilde{\mathbf{X}}_i = f(\mathbf{X}_i) \text{ any vector-valued measurable function}$$

Background

Covariate Balancing Propensity Score (CBPS)

See Imai (2014): Estimate propensity score so that covariate balance is optimized

Propensity Score Tautology

- Propensity score is unknown
- Diagnostics: covariate balance checking
- Misspecification is possible especially for non-binary treatments
- Skewed covariates are common in applied settings
- Propensity score methods can be sensitive to misspecification

Covariate balancing condition (C.1)

$$\mathrm{E}[\frac{T_i\tilde{\mathbf{X}}_i}{\pi_\beta(\mathbf{X}_i)} - \frac{(1-T_i)\tilde{\mathbf{X}}_i}{1-\pi_\beta(\mathbf{X}_i)}] = 0,\ \tilde{\mathbf{X}}_i = f(\mathbf{X}_i) \text{ any vector-valued measurable function}$$

Score condition (C.2)

$$\mathrm{E}[\frac{T_i\pi'_\beta(\mathbf{X}_i)}{\pi_\beta(\mathbf{X}_i)} - \frac{(1-T_i)\pi'_\beta(\mathbf{X}_i)}{1-\pi_\beta(\mathbf{X}_i)}] = 0,\ (f(\mathbf{X}_i) = \pi'(\mathbf{X}_i))$$

Background

Covariate Balancing Propensity Score (CBPS)

- Just-identified CBPS: covariate balancing conditions alone
- Over-identified CBPS: combine them with score conditions

Background

Covariate Balancing Propensity Score (CBPS)

- Just-identified CBPS: covariate balancing conditions alone
- Over-identified CBPS: combine them with score conditions

MLE

$$\hat{\beta}_{MLE} = \text{argmax}_{\beta \in \Theta} \sum_{i=1}^{N} T_i \log(\pi_\beta(\mathbf{X}_i)) + (1 - T_i)\log(1 - \pi_\beta(\mathbf{X}_i))$$

$$\frac{1}{N}\sum_{i=1}^{N} s_\beta(T_i, \mathbf{X}_i) = 0, \;\; s_\beta(T_i, \mathbf{X}_i) = \frac{T_i \pi'_\beta(\mathbf{X}_i)}{\pi_\beta(\mathbf{X}_i)} - \frac{(1 - T_i)\pi'_\beta(\mathbf{X}_i)}{1 - \pi_\beta(\mathbf{X}_i)}$$

Background

Covariate Balancing Propensity Score (CBPS)

- Just-identified CBPS: covariate balancing conditions alone
- Over-identified CBPS: combine them with score conditions

Background

Covariate Balancing Propensity Score (CBPS)

- Just-identified CBPS: covariate balancing conditions alone
- Over-identified CBPS: combine them with score conditions

GMM

$$\hat{\boldsymbol{\beta}}_{GMM} = \operatorname{argmin} \bar{g}_{\beta}(\mathbf{T}, \mathbf{X})' \boldsymbol{\Sigma}_{\beta}(\mathbf{T}, \mathbf{X})^{-1} g_{\beta}(\mathbf{T}, \mathbf{X}), \text{ where } g_{\beta}(\mathbf{T}, \mathbf{X}) = \frac{1}{N} \sum_{i=1}^{N} \underbrace{\begin{bmatrix} \text{score condition} \\ \text{balancing condition} \end{bmatrix}}_{g_{\beta}(T_i, \mathbf{X}_i)}$$

Background

Covariate Balancing Propensity Score (CBPS)

- Just-identified CBPS: covariate balancing conditions alone
- Over-identified CBPS: combine them with score conditions

GMM

$$\hat{\boldsymbol{\beta}}_{GMM} = \text{argmin } \bar{g}_\beta(\mathbf{T},\mathbf{X})'\boldsymbol{\Sigma}_\beta(\mathbf{T},\mathbf{X})^{-1}g_\beta(\mathbf{T},\mathbf{X}), \text{ where } g_\beta(\mathbf{T},\mathbf{X}) = \frac{1}{N}\sum_{i=1}^{N}\underbrace{\begin{bmatrix}\text{score condition}\\ \text{balancing condition}\end{bmatrix}}_{g_\beta(T_i,\mathbf{X}_i)}$$

The balancing condition: $\frac{1}{N}\sum_{i=1}^{N} w_\beta(T_i,\mathbf{X}_i)\tilde{\mathbf{X}}_i$

- $w_\beta(T_i,\mathbf{X}_i) = \frac{T_i-\pi_\beta(\mathbf{X}_i)}{\pi_\beta(\mathbf{X}_i)[1-\pi_\beta(\mathbf{X}_i)]}$, the sample analogue of C.1
- $w_\beta(T_i,\mathbf{X}_i) = \frac{N}{N_1}\frac{T_i-\pi_\beta(\mathbf{X}_i)}{1-\pi_\beta(\mathbf{X}_i)}$, the sample analogue of C.2

Background

Covariate Balancing Propensity Score (CBPS)

Background

Covariate Balancing Propensity Score (CBPS)

Pros

- Single model determines the treatment assignment mechanism and the covariate balancing weights
- Optimize the covariate balance while modeling the treatment assignment
- Estimate propensity score without consulting the outcome data
- Improve WLS and DR estimator (reduce s.e.)
- **Robust to model misspecification**

Background

Covariate Balancing Propensity Score (CBPS)

Pros

- Single model determines the treatment assignment mechanism and the covariate balancing weights
- Optimize the covariate balance while modeling the treatment assignment
- Estimate propensity score without consulting the outcome data
- Improve WLS and DR estimator (reduce s.e.)
- **Robust to model misspecification**

A simple and yet powerful method to improve the parameteric estimation of the propensity score

Background

Covariate Balancing Propensity Score (CBPS)

Pros

- Single model determines the treatment assignment mechanism and the covariate balancing weights
- Optimize the covariate balance while modeling the treatment assignment
- Estimate propensity score without consulting the outcome data
- Improve WLS and DR estimator (reduce s.e.)
- **Robust to model misspecification**

A simple and yet powerful method to improve the parametric estimation of the propensity score

Cons

- Can **introduce bias** to HT, IPW estimators, when both models are correctly specified, or PS model is correct

Background

Covariate Balancing Propensity Score (CBPS)

Pros

- Single model determines the treatment assignment mechanism and the covariate balancing weights
- Optimize the covariate balance while modeling the treatment assignment
- Estimate propensity score without consulting the outcome data
- Improve WLS and DR estimator (reduce s.e.)
- **Robust to model misspecification**

A simple and yet powerful method to improve the parameteric estimation of the propensity score

Cons

- Can **introduce bias** to HT, IPW estimators, when both models are correctly specified, or PS model is correct

Empirical Study

The CBPS method can be applied to HT, IPW, WLS, and DR estimator of ATE

High-dimensional CBPS

HD-CBPS Methodology

HD-CBPS Methodology

Set-up

- Data: $T_i, Y_i, \mathbf{X}_i$, $\dim(\mathbf{X}_i) = d$, $d \gg N$
- Potential outcomes: $Y_i(1), Y_i(0)$
- SUTVA: $Y_i = T_i Y_i(1) + (1 - T_i) Y_i(0)$
- Positivity: $0 < \Pr(T_i = 1 | \mathbf{X}_i) < 1$
- Ignorability: $\{Y_i(0), Y_i(1) \perp\!\!\!\perp T_i | \mathbf{X}_i\}$
- ATE: $\mu^* = \mathrm{E}[Y_i(1) - Y_i(0)]$
- Focus on: $\mu_1^* = \mathrm{E}[Y_i(1)]$

HD-CBPS Methodology

Set-up

- Data: $T_i, Y_i, \mathbf{X}_i$, $\dim(\mathbf{X}_i) = d$, $d \gg N$
- Potential outcomes: $Y_i(1), Y_i(0)$
- SUTVA: $Y_i = T_i Y_i(1) + (1 - T_i) Y_i(0)$
- Positivity: $0 < \Pr(T_i = 1|\mathbf{X}_i) < 1$
- Ignorability: $\{Y_i(0), Y_i(1) \perp\!\!\!\perp T_i|\mathbf{X}_i\}$
- ATE: $\mu^* = \mathrm{E}[Y_i(1) - Y_i(0)]$
- Focus on: $\mu_1^* = \mathrm{E}[Y_i(1)]$

Two correctly specified high-dimensional working models

- $\Pr(T_i = 1|\mathbf{X}_i) = \pi(\mathbf{X}_i' \boldsymbol{\beta}^* + u_i)$
  - where u_i is the approximation error
- $\mathrm{E}[Y_i(1)|\mathbf{X}_i] = K_1(\mathbf{X}_i) + r_i$, $K_1(\mathbf{X}_i) = \boldsymbol{\alpha}^{*\prime} \mathbf{X}_i$
  - where r_i is the approximation error

HD-CBPS Methodology

HD-CBPS Methodology

Recall

Covariate balancing properties

Let $\hat{\pi} = \hat{\pi}(\mathbf{X}'\hat{\boldsymbol{\beta}})$ be an estimator of the propensity score

$$\sum_{i=1}^{N}(\frac{T_i}{\hat{\pi}} - 1)\mathbf{X}_i = \mathbf{0}, \text{ strong condition}$$

e.g., balancing the mean or every component of $\mathbf{X}_i$ (Imai, 2014; Fan et al., 2016)

$$\sum_{i=1}^{N}(\frac{T_i}{\hat{\pi}} - 1)\boldsymbol{\alpha}^{*\prime}\mathbf{X}_i = 0, \text{ weak condition}$$

HD-CBPS Methodology

Recall

Covariate balancing properties

Let $\hat{\pi} = \hat{\pi}(\mathbf{X}'\hat{\boldsymbol{\beta}})$ be an estimator of the propensity score

$$\sum_{i=1}^{N}(\frac{T_i}{\hat{\pi}} - 1)\mathbf{X}_i = \mathbf{0}, \text{ strong condition}$$

e.g., balancing the mean or every component of $\mathbf{X}_i$ (Imai, 2014; Fan et al., 2016)

$$\sum_{i=1}^{N}(\frac{T_i}{\hat{\pi}} - 1)\boldsymbol{\alpha}^{*\prime}\mathbf{X}_i = 0, \text{ weak condition}$$

Challenge

In high-dimensional setting, $\hat{\boldsymbol{\beta}}$ that satisfies strong condition is not unique ?

HD-CBPS Methodology

HD-CBPS Methodology

Main idea

Update estimated propensity score by balancing a **subset** of covariates that are predictive to the outcome

HD-CBPS Methodology

Main idea

Update estimated propensity score by balancing a **subset** of covariates that are predictive to the outcome

Step 1 (Generalized Quasilikelihood)

$$Q_N(\boldsymbol{\beta}) = \frac{1}{N}\sum_{i=1}^{N}\int_0^{\boldsymbol{\beta}'\mathbf{X}_i}[\frac{T_i}{\pi(u)} - 1]w_1(u)du$$

$$\hat{\boldsymbol{\beta}} = \text{argmin}_{\beta}\{-Q_N(\boldsymbol{\beta}) + \lambda_{\beta}||\boldsymbol{\beta}||_1\}$$

Remark: $\frac{\partial Q_N(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \frac{1}{N}\sum_{i=1}^{N}[\frac{T_i}{\pi(\boldsymbol{\beta}'\mathbf{X}_i)} - 1]w_1(\boldsymbol{\beta}'\mathbf{X}_i)\mathbf{X}_i$

HD-CBPS Methodology

Main idea

Update estimated propensity score by balancing a **subset** of covariates that are predictive to the outcome

Step 1 (Generalized Quasilikelihood)

$$Q_N(\boldsymbol{\beta}) = \frac{1}{N}\sum_{i=1}^{N}\int_0^{\boldsymbol{\beta}'\mathbf{X}_i}[\frac{T_i}{\pi(u)} - 1]w_1(u)du$$

$$\hat{\boldsymbol{\beta}} = \text{argmin}_\beta\{-Q_N(\boldsymbol{\beta}) + \lambda_\beta||\boldsymbol{\beta}||_1\}$$

Remark: $\frac{\partial Q_N(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \frac{1}{N}\sum_{i=1}^{N}[\frac{T_i}{\pi(\boldsymbol{\beta}'\mathbf{X}_i)} - 1]w_1(\boldsymbol{\beta}'\mathbf{X}_i)\mathbf{X}_i$

Step 2

$$L_N(\boldsymbol{\alpha}) = \frac{1}{N}\sum_{i=1}^{N}T_i w_2(\hat{\boldsymbol{\beta}}'\mathbf{X}_i)(Y_i - \boldsymbol{\alpha}'\mathbf{X}_i)^2$$

$$\tilde{\boldsymbol{\alpha}} = \text{argmin}_\alpha\{L_N(\boldsymbol{\alpha}) + \lambda_\alpha||\boldsymbol{\alpha}||_1\}$$

HD-CBPS Methodology

Main idea

HD-CBPS Methodology

Main idea

Step 3 (Update! Remove bias)

Active set: $\tilde{S} = \{j : \tilde{\alpha}_j \neq 0\}$; and $\mathbf{X}_{\tilde{S}}$

$$g_N(\boldsymbol{\gamma}) = \frac{1}{N}\sum_{i=1}^{N}\left\{\frac{T_i}{\pi(\boldsymbol{\gamma}'\mathbf{X}_{i\tilde{S}} + \hat{\boldsymbol{\beta}}'_{\tilde{S}^c}\mathbf{X}_{i\tilde{S}^c})} - 1\right\}\mathbf{X}_{i\tilde{S}}$$

$$\tilde{\gamma} = \text{argmin}_{\gamma\in\mathbb{R}^{|\tilde{S}|}} ||g_N(\boldsymbol{\gamma})||_2^2$$

HD-CBPS Methodology

Main idea

Step 3 (Update! Remove bias)

Active set: $\tilde{S} = \{j : \tilde{\alpha}_j \neq 0\}$; and $\mathbf{X}_{\tilde{S}}$

$$g_N(\boldsymbol{\gamma}) = \frac{1}{N}\sum_{i=1}^{N}\left\{\frac{T_i}{\pi(\boldsymbol{\gamma}'\mathbf{X}_{i\tilde{S}} + \hat{\boldsymbol{\beta}}'_{\tilde{S}^c}\mathbf{X}_{i\tilde{S}^c})} - 1\right\}\mathbf{X}_{i\tilde{S}}$$

$$\tilde{\gamma} = \operatorname{argmin}_{\gamma\in\mathbb{R}^{|\tilde{S}|}} ||g_N(\boldsymbol{\gamma})||_2^2$$

Step 4

Let $\tilde{\boldsymbol{\beta}} = (\tilde{\boldsymbol{\gamma}}', \hat{\boldsymbol{\beta}}'_{\tilde{S}^c})'$, and $\tilde{\pi}_i = \pi(\tilde{\boldsymbol{\beta}}'\mathbf{X}_i)$

- Remark: $\sum_{i=1}^{N}\left(\frac{T_i}{\tilde{\pi}_i} - 1\right)\mathbf{X}_i'\boldsymbol{\alpha}^* \approx \sum_{i=1}^{N}\left(\frac{T_i}{\tilde{\pi}_i} - 1\right)\mathbf{X}_i'\tilde{\boldsymbol{\alpha}} = \sum_{i=1}^{N}\left(\frac{T_i}{\tilde{\pi}_i} - 1\right)\mathbf{X}'_{i\tilde{S}}\tilde{\boldsymbol{\alpha}}_{\tilde{S}} = 0$

Horvitz–Thompson estimator:

$$\hat{\mu}_1 = \frac{1}{N}\sum_{i=1}^{N}\frac{T_iY_i}{\tilde{\pi}_i}$$

Theoretical Properties

Assumptions

A.1 (Unconfoundedness) $\{Y_i(0), Y_i(1) \perp\!\!\!\perp T_i | \mathbf{X}_i\}$.

A.2 (Overlap) $\pi_i^* \geq c_0$, for some constant c_0.

A.3 (Sub-Guassian condition) both $\epsilon_1 = Y(1) - \boldsymbol{\alpha}^{*\prime}\mathbf{X}$ and X_j satisfy $||\epsilon_1||_{\psi_2} \leq c_\epsilon$, $||X_j||_{\psi_2} \leq c_X$.

A.4 (Sparsity) $\frac{(s_1 \vee s_2)\log(d \vee N)}{\sqrt{N}} = o(1)$, $s_1 = ||\boldsymbol{\beta}^*||_1$, $s_2 = ||\boldsymbol{\alpha}^*||_1$; $\sum_{i=1}^N u_i^2 = O(s_1)$, $\sum_{i=1}^N r_i^2 = O(s_2)$, and $\sum_{i=1}^N u_i r_i = o(\sqrt{N})$.

A.5 (Eigenvalue condition) $\boldsymbol{\Sigma} = \mathrm{E}(\mathbf{X}\mathbf{X}')$, then $C \leq \lambda_{min}(\Sigma_{SS}) \leq \lambda_{max}(\Sigma_{SS}) \leq 1/C$ for some constant C and any $S \subset \{1, 2, \ldots, d\}$ with $|S| \leq (s_1 \vee s_2)\log(N)$.

A.6 (Propensity score and weight functions)

- i) PS model $\pi(u)$ is bounded away from 0, and $\pi'(u)$ is locally Lipschitz
- ii) weight function $w_1(u)$ is bounded away from 0 and $w_1'(u)$ is locally Lipschitz
- iii) weight function $w_2(u)$ is bounded away from 0 and $w_2'(u)$ is continuous.

Theoretical Properties

Under correct model specification

Theorem 1 (Asymp. Normality & Semi-Effi)

When $\Pr(T_i = 1|\mathbf{X}_i) = \pi(\mathbf{X}_i'\boldsymbol{\beta}^* + u_i)$ and $\mathrm{E}[Y_i(1)|\mathbf{X}_i] = K_1(\mathbf{X}_i) + r_i$ are correctly specified, and A.1-A.6 hold, $\hat{\mu}_1$ with ANY weight functions satisfies

Theoretical Properties

Under correct model specification

Theorem 1 (Asymp. Normality & Semi-Effi)

When $\Pr(T_i = 1|\mathbf{X}_i) = \pi(\mathbf{X}_i'\boldsymbol{\beta}^* + u_i)$ and $\mathrm{E}[Y_i(1)|\mathbf{X}_i] = K_1(\mathbf{X}_i) + r_i$ are correctly specified, and A.1-A.6 hold, $\hat{\mu}_1$ with ANY weight functions satisfies

$$\hat{\mu}_1 - \mu_1^* = \frac{1}{N}\sum_{i=1}^{N}\left[\frac{T_i}{\pi_i^*}\{Y_i(1) - \boldsymbol{\alpha}^{*\prime}\mathbf{X}_i\} + \boldsymbol{\alpha}^{*\prime}\mathbf{X}_i - \mu_1^*\right] + O_p\left\{\frac{(s_1 \vee s_2)\log(d \vee N)}{N}\right\}$$

as $d, N \to \infty$.

Theoretical Properties

Under correct model specification

Theorem 1 (Asymp. Normality & Semi-Effi)

When $\Pr(T_i = 1|\mathbf{X}_i) = \pi(\mathbf{X}_i'\boldsymbol{\beta}^* + u_i)$ and $\mathrm{E}[Y_i(1)|\mathbf{X}_i] = K_1(\mathbf{X}_i) + r_i$ are correctly specified, and A.1-A.6 hold, $\hat{\mu}_1$ with ANY weight functions satisfies

$$\hat{\mu}_1 - \mu_1^* = \frac{1}{N}\sum_{i=1}^{N}\Big[\frac{T_i}{\pi_i^*}\{Y_i(1) - \boldsymbol{\alpha}^{*\prime}\mathbf{X}_i\} + \boldsymbol{\alpha}^{*\prime}\mathbf{X}_i - \mu_1^*\Big] + O_p\Big\{\frac{(s_1 \vee s_2)\log(d \vee N)}{N}\Big\}$$

as $d, N \to \infty$.

Let V be the semiparametric asymptotic variance bound, i.e.

$$V = \mathrm{E}\Big[\frac{1}{\pi^*}\mathrm{E}(\epsilon_1^2|\mathbf{X}) + (\boldsymbol{\alpha}^{*\prime}\mathbf{X} - \mu_1^*)^2\Big].$$

Assume $\mathrm{E}(\epsilon_1^2|\mathbf{X}) \geq c$ for some $c > 0$ and $\mathrm{E}(\boldsymbol{\alpha}^{*\prime}\mathbf{X})^4 = O(s_2^2)$. Then

$$\sqrt{N}(\hat{\mu}_1 - \mu_1^*)/V^{1/2} \xrightarrow{d} N(0, 1).$$

Theoretical Properties

Under correct model specification

Theoretical Properties

Under correct model specification

True variance

$$V = \mathrm{E}\Big[\frac{1}{\pi^*}\mathrm{E}(\epsilon_1^2|\mathbf{X}) + (\boldsymbol{\alpha}^{*\prime}\mathbf{X} - \mu_1^*)^2\Big]$$

Estimated variance ?

$$\hat{V} = \frac{1}{N}\sum_{i=1}^{N}\Big\{\frac{T_i}{\tilde{\pi}_i^2}(Y_i - \tilde{\boldsymbol{\alpha}}'\mathbf{X}_i)^2 + (\tilde{\boldsymbol{\alpha}}'\mathbf{X}_i - \hat{\mu}_1)^2\Big\}$$

Theoretical Properties

Under correct model specification

True variance

$$V = \mathrm{E}\Big[\frac{1}{\pi^*}\mathrm{E}(\epsilon_1^2|\mathbf{X}) + (\boldsymbol{\alpha}^{*\prime}\mathbf{X} - \mu_1^*)^2\Big]$$

Estimated variance ?

$$\hat{V} = \frac{1}{N}\sum_{i=1}^{N}\Big\{\frac{T_i}{\tilde{\pi}_i^2}(Y_i - \tilde{\boldsymbol{\alpha}}'\mathbf{X}_i)^2 + (\tilde{\boldsymbol{\alpha}}'\mathbf{X}_i - \hat{\mu}_1)^2\Big\}$$

Honest CI

When all assumptions in Theorem 1 hold, then

$$|\hat{V} - V| = O_p\Big[(s_1 \vee s_2)\Big\{\frac{\log(d \vee N)}{N}\Big\}^{1/2}\Big].$$

Given $0 < \eta \leq 1$, $\mathcal{I} = \big(\hat{\mu}_1 - z_{1-\eta/2}(\hat{V}/N)^{1/2}, \hat{\mu}_1 + z_{1-\eta/2}(\hat{V}/N)^{1/2}\big)$

Theoretical Properties

Under misspecified PS model (Robustness)

$\pi^* = \Pr(T = 1|\mathbf{X})$ does not belong to the assumed parametric class $\{\pi(\mathbf{X}'\boldsymbol{\beta})\}$

Theoretical Properties

Under misspecified PS model (Robustness)

$\pi^* = \Pr(T = 1|\mathbf{X})$ does not belong to the assumed parametric class $\{\pi(\mathbf{X}'\boldsymbol{\beta})\}$

New step 1

$$\boldsymbol{\beta}^o = \text{argmax E}\Big[\int_0^{\mathbf{X}_i'\boldsymbol{\beta}} \{\frac{T_i}{\pi(u)} - 1\} w_1(u) du\Big],\ \boldsymbol{\beta}^o \text{ is sparse}$$

Theoretical Properties

Under misspecified PS model (Robustness)

$\pi^* = \Pr(T=1|\mathbf{X})$ does not belong to the assumed parametric class $\{\pi(\mathbf{X}'\boldsymbol{\beta})\}$

New step 1

$$\boldsymbol{\beta}^o = \text{argmax } \mathrm{E}\Big[\int_0^{\mathbf{X}_i'\boldsymbol{\beta}} \{\frac{T_i}{\pi(u)} - 1\} w_1(u) du\Big], \ \boldsymbol{\beta}^o \text{ is sparse}$$

Remark

- When PS model is correct, $\boldsymbol{\beta}^o$ reduces to $\boldsymbol{\beta}^*$
- The estimand, $\boldsymbol{\beta}^o$ implicitly depends on the choice of $w_1(u)$
- By choosing $w_1(u) = 1$, $\boldsymbol{\beta}^o$ satisfies

$$\mathrm{E}\big[\{\frac{T_i}{\pi(\mathbf{X}_i'\boldsymbol{\beta}^o)} - 1\}\mathbf{X}_i\big] = 0$$

- $\boldsymbol{\beta}^o$: an estimand by solving the strong covariate balancing equation
- $\pi^o = \pi(\mathbf{X}_i'\boldsymbol{\beta}^o)$

Theoretical Properties

Under misspecified PS model (Robustness)

Theoretical Properties

Under misspecified PS model (Robustness)

$$\hat{\mu}_1 - \mu_1^* = \frac{1}{N}\sum_{i=1}^{N}\Big[\frac{T_i}{\pi_i^o}\{Y_i(1) - \boldsymbol{\alpha}^{*\prime}\mathbf{X}_i\} + \boldsymbol{\alpha}^{*\prime}\mathbf{X}_i - \mu_1^*\Big] + I_1 + I_2$$

Theoretical Properties

Under misspecified PS model (Robustness)

$$\hat{\mu}_1 - \mu_1^* = \frac{1}{N}\sum_{i=1}^{N}\Big[\frac{T_i}{\pi_i^o}\{Y_i(1) - \boldsymbol{\alpha}^{*\prime}\mathbf{X}_i\} + \boldsymbol{\alpha}^{*\prime}\mathbf{X}_i - \mu_1^*\Big] + I_1 + I_2$$

$$I_1 = \frac{1}{N}\sum_{i=1}^{N}\Big(\frac{T_i}{\tilde{\pi}_i} - \frac{T_i}{\pi_i^o}\Big)\{Y_i(1) - \mathbf{X}_i'\boldsymbol{\alpha}^*\}$$

$$I_2 = \frac{1}{N}\sum_{i=1}^{N}\Big(\frac{T_i}{\tilde{\pi}_i} - 1\Big)\mathbf{X}_i'\boldsymbol{\alpha}^*$$

Theoretical Properties

Under misspecified PS model (Robustness)

$$\hat{\mu}_1 - \mu_1^* = \frac{1}{N}\sum_{i=1}^{N}\Big[\frac{T_i}{\pi_i^o}\{Y_i(1) - \boldsymbol{\alpha}^{*\prime}\mathbf{X}_i\} + \boldsymbol{\alpha}^{*\prime}\mathbf{X}_i - \mu_1^*\Big] + I_1 + I_2$$

$$I_1 = \frac{1}{N}\sum_{i=1}^{N}\Big(\frac{T_i}{\tilde{\pi}_i} - \frac{T_i}{\pi_i^o}\Big)\{Y_i(1) - \mathbf{X}_i'\boldsymbol{\alpha}^*\}$$

$$I_2 = \frac{1}{N}\sum_{i=1}^{N}\Big(\frac{T_i}{\tilde{\pi}_i} - 1\Big)\mathbf{X}_i'\boldsymbol{\alpha}^*$$

Proposition 1 (Only outcome model is correctly specified)

......,

$$\hat{\mu}_1 - \mu_1^* = \frac{1}{N}\sum_{i=1}^{N}\Big[\frac{T_i}{\pi_i^o}\{Y_i(1) - \boldsymbol{\alpha}^{*\prime}\mathbf{X}_i\} + \boldsymbol{\alpha}^{*\prime}\mathbf{X}_i - \mu_1^*\Big] + O_p\Big\{\frac{(s_1 \vee s_2)\log(d \vee N)}{N}\Big\}$$

$$V_{mis-ps} = \mathrm{E}\Big[\frac{\pi^*}{(\pi^o)^2}\mathrm{E}(\epsilon_1^2|\mathbf{X}) + (\boldsymbol{\alpha}^{*\prime}\mathbf{X} - \mu_1^*)^2\Big],\ \sqrt{N}(\hat{\mu}_1 - \mu_1^*)/V_{mis-ps}^{1/2} \xrightarrow{d} N(0,1)$$

Remark: under misspecified PS model, $\hat{V}$ is still consistent for V_{mis-ps}, see the proof of Corollary 1.

Theoretical Properties

Under misspecified outcome model (Robustness)

$\mathrm{E}[Y_i(1)|\mathbf{X}_i]$ is non-linear in $\mathbf{X}_i$

Theoretical Properties

Under misspecified outcome model (Robustness)

$\mathrm{E}[Y_i(1)|\mathbf{X}_i]$ is non-linear in $\mathbf{X}_i$

New step 2

$$\boldsymbol{\alpha}^o = \operatorname{argmin} \mathrm{E}\{T_i w_2(\mathbf{X}_i'\boldsymbol{\beta})(Y_i - \mathbf{X}_i'\boldsymbol{\alpha})^2\}$$

Theoretical Properties

Under misspecified outcome model (Robustness)

$\mathrm{E}[Y_i(1)|\mathbf{X}_i]$ is non-linear in $\mathbf{X}_i$

New step 2

$$\boldsymbol{\alpha}^o = \operatorname{argmin} \mathrm{E}\{T_i w_2(\mathbf{X}_i'\boldsymbol{\beta})(Y_i - \mathbf{X}_i'\boldsymbol{\alpha})^2\}$$

Proposition 2 (Only PS model is correctly specified)

......,

$$\hat{\mu}_1 - \mu_1^* = \frac{1}{N}\sum_{i=1}^{N}\left[\frac{T_i}{\pi_i^*}\{Y_i(1) - \boldsymbol{\alpha}^{o\prime}\mathbf{X}_i\} + \boldsymbol{\alpha}^{o\prime}\mathbf{X}_i - \mu_1^*\right] + O_p\left\{\frac{(s_1 \vee s_2)\log(d \vee N)}{N}\right\}$$

$$V_{mis-o} = \mathrm{E}\left[\frac{1}{\pi^*}\mathrm{E}(\epsilon^{o2}|\mathbf{X}) + (\boldsymbol{\alpha}^{o\prime}\mathbf{X} - \mu_1^*)^2\right],\ \sqrt{N}(\hat{\mu}_1 - \mu_1^*)/V_{mis-o}^{1/2} \xrightarrow{d} N(0,1)$$

Remark 1: $w_2(u)$ can be $\frac{\pi'(u)}{\pi^2(u)}$

Remark 2: under misspecified outcome model, $\hat{V}$ is still consistent for V_{mis-ps}

Theoretical Properties

Double Robust

> Proposition 1 and Proposition 2 together imply $\hat{\mu}_1$ is root-n consistent and asymptotically normal provided either the PS-model or O-model is correctly specified. And $\hat{V}$ is always consistent, the same confidence interval is valid so long as one of the two models is correctly specified.
>
> Recommendation: $w_1(u) = 1$ and $w_2(u) = \pi'(u)/\pi^2(u)$.

Remark: this method can also be used for generalized linear models, along with great theoretical properties. See chapter 4.

Simulation Study

Simulation

Simulation

Set-up

$$\mathbf{X}_i \sim N(0, \Sigma),\ \Sigma_{jk} = \rho^{|j-k|},\ \rho = .2,\ j, k \in \{1, 2, \ldots, d\}$$

$$\pi(\mathbf{X}_i) = 1 - \frac{1}{1 + \exp(-X_{i1} + \frac{1}{2}X_{i2} - \frac{1}{4}X_{i3} - \frac{1}{10}X_{i4} - \frac{1}{10}X_{i5} + \frac{1}{10}X_{i6})}$$

$$Y_i(1) = 2 + .137(X_{i5} + X_{i6} + X_{i7} + X_{i8}) + \epsilon_{1i}$$

$$Y_i(0) = 1 + .291(X_{i5} + X_{i6} + X_{i7} + X_{i8} + X_{i9} + X_{i10}) + \epsilon_{0i}$$

$$\begin{aligned} X_{mis} =& \{\exp(X_1/2),\ X_2/\{1 + \exp(X_1)\} + 10,\ (X_1X_3/25 + 0.6)^3,\ (X_2 + X_4 + 20)^2, \\ & X_6,\ \exp(X_6 + X_7),\ X_9^2,\ X_7^3 - 20,\ X_9, \ldots,\ X_d\} \end{aligned}$$

Simulation

Set-up

$$\mathbf{X}_i \sim N(0, \Sigma),\ \Sigma_{jk} = \rho^{|j-k|},\ \rho = .2,\ j,k \in \{1,2,\ldots,d\}$$

$$\pi(\mathbf{X}_i) = 1 - \frac{1}{1 + \exp(-X_{i1} + \frac{1}{2}X_{i2} - \frac{1}{4}X_{i3} - \frac{1}{10}X_{i4} - \frac{1}{10}X_{i5} + \frac{1}{10}X_{i6})}$$

$$Y_i(1) = 2 + .137(X_{i5} + X_{i6} + X_{i7} + X_{i8}) + \epsilon_{1i}$$

$$Y_i(0) = 1 + .291(X_{i5} + X_{i6} + X_{i7} + X_{i8} + X_{i9} + X_{i10}) + \epsilon_{0i}$$

$$\begin{aligned} X_{mis} =& \{\exp(X_1/2),\ X_2/\{1+\exp(X_1)\} + 10,\ (X_1X_3/25 + 0.6)^3,\ (X_2 + X_4 + 20)^2, \\ & X_6,\ \exp(X_6 + X_7),\ X_9^2,\ X_7^3 - 20,\ X_9, \ldots,\ X_d\} \end{aligned}$$

Candidate methods

- HD-CBPS: the proposed method
- PB: approximate residual balancing
- AIPW: regularized augmented inverse probability weighting estimator
- D-SELECT: double selection

Simulation

Set-up

$$\mathbf{X}_i \sim N(0, \Sigma),\ \Sigma_{jk} = \rho^{|j-k|},\ \rho = .2,\ j,k \in \{1,2,\dots,d\}$$

$$\pi(\mathbf{X}_i) = 1 - \frac{1}{1 + \exp(-X_{i1} + \frac{1}{2}X_{i2} - \frac{1}{4}X_{i3} - \frac{1}{10}X_{i4} - \frac{1}{10}X_{i5} + \frac{1}{10}X_{i6})}$$

$$Y_i(1) = 2 + .137(X_{i5} + X_{i6} + X_{i7} + X_{i8}) + \epsilon_{1i}$$

$$Y_i(0) = 1 + .291(X_{i5} + X_{i6} + X_{i7} + X_{i8} + X_{i9} + X_{i10}) + \epsilon_{0i}$$

$$\begin{aligned} X_{mis} =& \{\exp(X_1/2),\ X_2/\{1+\exp(X_1)\} + 10,\ (X_1X_3/25 + 0.6)^3,\ (X_2 + X_4 + 20)^2, \\ & X_6,\ \exp(X_6 + X_7),\ X_9^2,\ X_7^3 - 20,\ X_9, \dots,\ X_d\} \end{aligned}$$

Candidate methods

- HD-CBPS: the proposed method
- PB: approximate residual balancing
- AIPW: regularized augmented inverse probability weighting estimator
- D-SELECT: double selection

Evaluation metrics

- Std.err: standard error
- RMSE: standardized root-mean-squared-error
- Coverage: coverage probability of 95% confidence intervals

Simulation

Both models correct

Simulation

Both models correct

N=500, d=1000

```
##             HD-CBPS      RB    AIPW D-SELECT
## Bias       -0.0026 -0.0017 -0.0498  -0.0910
## Std err     0.0936  0.1074  0.0926   0.0979
## RMSE        0.0936  0.0174  0.1052   0.1337
## Coverage    0.9650  0.9300  0.9150   0.8900
## CI length   0.3867  0.4231  0.3775   0.4294
```

Simulation

Both models correct

N=500, d=1000

```
##             HD-CBPS      RB    AIPW D-SELECT
## Bias        -0.0026 -0.0017 -0.0498  -0.0910
## Std err      0.0936  0.1074  0.0926   0.0979
## RMSE         0.0936  0.0174  0.1052   0.1337
## Coverage     0.9650  0.9300  0.9150   0.8900
## CI length    0.3867  0.4231  0.3775   0.4294
```

N=500, d=2000

```
##             HD-CBPS      RB    AIPW D-SELECT
## Bias        -0.0595 -0.0580 -0.1200  -0.0397
## Std err      0.1061  0.1155  0.1011   0.1279
## RMSE         0.1216  0.1292  0.1569   0.1279
## Coverage     0.9100  0.9100  0.8550   0.9450
## CI length    0.3862  0.4359  0.3731   0.5034
```

Simulation

Propensity score model is misspecified

Simulation

Propensity score model is misspecified

N=500, d=1000

```
##             HD-CBPS      RB    AIPW D-SELECT
## Bias       -0.0120 -0.0303 -0.1078  -0.0782
## Std err     0.0984  0.1153  0.0963   0.1034
## RMSE        0.0991  0.1193  0.1446   0.1296
## Coverage    0.9650  0.9450  0.8150   0.9050
## CI length   0.3864  0.4431  0.3732   0.4227
```

Simulation

Propensity score model is misspecified

N=500, d=1000

```
##             HD-CBPS      RB    AIPW D-SELECT
## Bias        -0.0120 -0.0303 -0.1078  -0.0782
## Std err      0.0984  0.1153  0.0963   0.1034
## RMSE         0.0991  0.1193  0.1446   0.1296
## Coverage     0.9650  0.9450  0.8150   0.9050
## CI length    0.3864  0.4431  0.3732   0.4227
```

N=500, d=2000

```
##             HD-CBPS      RB    AIPW D-SELECT
## Bias        -0.0446 -0.0685 -0.1234  -0.0357
## Std err      0.0924  0.1041  0.0921   0.1214
## RMSE         0.1025  0.1246  0.1540   0.1265
## Coverage     0.9300  0.9100  0.7400   0.9400
## CI length    0.3839  0.4382  0.3702   0.5023
```

Simulation

Outcome model is misspecified

Simulation

Outcome model is misspecified

N=500, d=1000

```
##             HD-CBPS      RB    AIPW D-SELECT
## Bias        -0.0034 -0.0321 -0.0562  -0.0991
## Std err      0.0917  0.0982  0.0914   0.1023
## RMSE         0.0917  0.1033  0.1072   0.1424
## Coverage     0.9600  0.9600  0.9050   0.8450
## CI length    0.3874  0.4292  0.3815   0.4327
```

Simulation

Outcome model is misspecified

N=500, d=1000

```
##             HD-CBPS      RB    AIPW D-SELECT
## Bias        -0.0034 -0.0321 -0.0562  -0.0991
## Std err      0.0917  0.0982  0.0914   0.1023
## RMSE         0.0917  0.1033  0.1072   0.1424
## Coverage     0.9600  0.9600  0.9050   0.8450
## CI length    0.3874  0.4292  0.3815   0.4327
```

N=500, d=2000

```
##             HD-CBPS      RB    AIPW D-SELECT
## Bias        -0.0317 -0.0572 -0.1215  -0.0443
## Std err      0.0944  0.0992  0.0921   0.1026
## RMSE         0.0995  0.1145  0.1525   0.1118
## Coverage     0.9500  0.9550  0.7700   0.9450
## CI length    0.3890  0.4403  0.3728   0.4261
```

Simulation

Both are misspecified

Simulation

Both are misspecified

N=500, d=1000

```
##             HD-CBPS      RB    AIPW D-SELECT
## Bias       -0.0547 -0.1201 -0.1873  -0.1005
## Std err     0.1106  0.1038  0.0903   0.0950
## RMSE        0.1234  0.1588  0.2079   0.1383
## Coverage    0.8900  0.8150  0.7750   0.8750
## CI length   0.3994  0.4586  0.3790   0.4333
```

Simulation

Both are misspecified

N=500, d=1000

```
##             HD-CBPS      RB    AIPW D-SELECT
## Bias       -0.0547 -0.1201 -0.1873  -0.1005
## Std err     0.1106  0.1038  0.0903   0.0950
## RMSE        0.1234  0.1588  0.2079   0.1383
## Coverage    0.8900  0.8150  0.7750   0.8750
## CI length   0.3994  0.4586  0.3790   0.4333
```

N=500, d=2000

```
##             HD-CBPS      RB    AIPW D-SELECT
## Bias       -0.0243 -0.0599 -0.1393  -0.0518
## Std err     0.0969  0.1060  0.0921   0.0965
## RMSE        0.0999  0.1218  0.1670   0.1095
## Coverage    0.9400  0.9400  0.7200   0.9500
## CI length   0.3948  0.4545  0.3781   0.4334
```

Summary

Summary

HD-CBPS

- Method: straightforward
- Appealling theoretical results
- Good empirical performance
    - CBPS tends to outperform the AIPW in low-dimensional settings
    - Conclusion appears to hold in high-dimensional settings as well

Summary

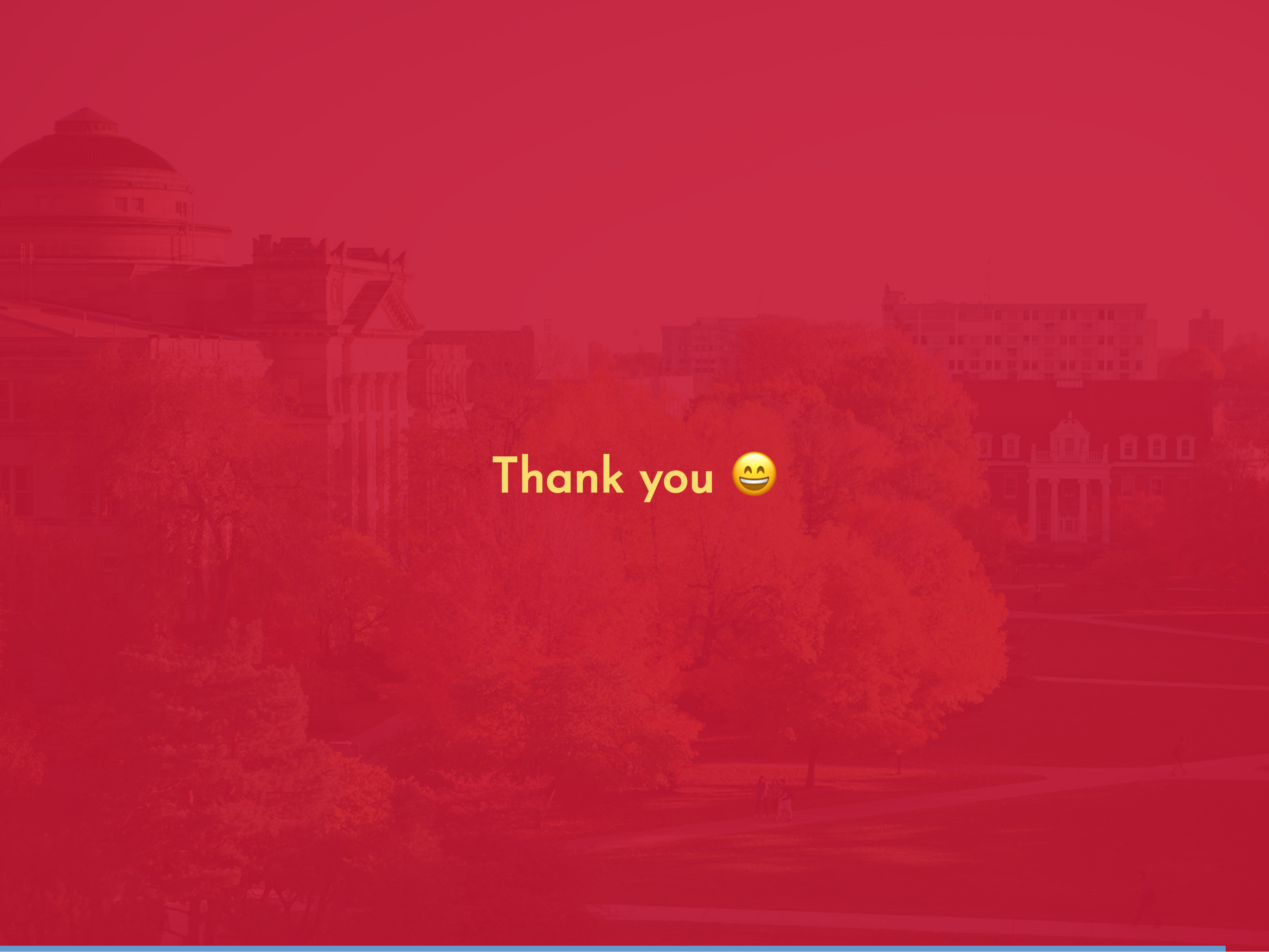
HD-CBPS

- Method: straightforward
- Appealling theoretical results
- Good empirical performance
  - CBPS tends to outperform the AIPW in low-dimensional settings
  - Conclusion appears to hold in high-dimensional settings as well

Potential

- Applied on different high-dimensional dataset
- Inference based on cross-validated estimators
  - Sample-splitting?
- Dense theory
  - Relax some conditions?

Questions


Thank you 😄

Reference

Bradic, J., Wager, S., & Zhu, Y. (2019). Sparsity double robust inference of average treatment effects. arXiv preprint arXiv:1905.00744.

Imai, K., & Ratkovic, M. (2014). Covariate balancing propensity score. Journal of the Royal Statistical Society Series B: Statistical Methodology, 76(1), 243-263.

Fan, J., Imai, K., Liu, H., Ning, Y., & Yang, X. (2016). Improving covariate balancing propensity score: A doubly robust and efficient approach. URL: https://imai. fas. harvard. edu/research/CBPStheory. html.

Ning, Y., Sida, P., & Imai, K. (2020). Robust estimation of causal effects via a high-dimensional covariate balancing propensity score. Biometrika, 107(3), 533-554.