Propensity score weighting under lack of positivity

Yi Liu

CIMA Group Meeting

NC STATE UNIVERSITY

October 20, 2023

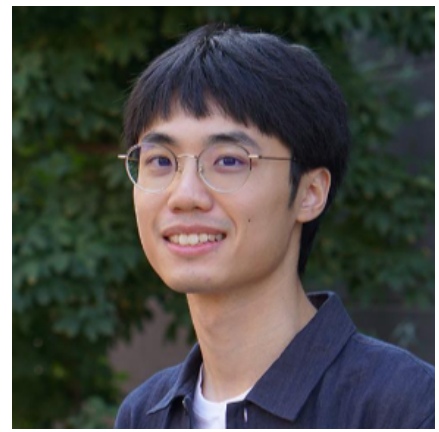
Overview

Related papers [1–4]:

- **Liu Y**, Li H, Zhou Y, and Matsouaka RA (2023+). Average treatment effect on the treated, under lack of positivity. Under review at *SMMR*.
- Matsouaka RA, **Liu Y**, and Zhou Y (2023+). Overlap, matching, or entropy weights: what are we weighting for? In revision at *Comm in Stat – S&C*.
- Matsouaka RA, and Zhou Y (2023+). A framework for causal inference in the presence of extreme inverse probability weights: the role of overlap weights. In revision at *Biomatrical Journal*.
- Zhou Y, Matsouaka RA, and Thomas L. Propensity score weighting under limited overlap and model misspecification (2020). *SMMR*.

Acknowledgement





Roland Matsouaka, PhD (Duke Biostat & Bioinfo, DCRI)
Huiyue Li (Baim Institute for Clinical Research, Boston)
Yunji Zhou (Univ. Washington Biostat)

Overview

- Propensity score methods have been ubiquitous in observational studies for identifying causal effects. At the heart of the propensity score methods is the **positivity** assumption.
- Each participant should have certain (non-zero) probabilities to receive either treatment or control, given their baseline covariates.

Overview

Set-up

- Observational data: X (baseline covariate vector), observed outcome Y and binary treatment $Z \in \{0, 1\}$.
- Potential outcomes $Y(0)$ and $Y(1)$, and the observed outcome from the data has a consistency with potential outcomes: $Y = ZY(1) + (1 - Z)Y(0)$.
- In this talk, we also focus on the following 2 assumptions hold:
  - Stable unit treatment values assumption (SUTVA): there is only one version of the treatment, and the potential outcome $Y(z), z = 0, 1$, of an individual does not depend on nor impact another's received treatment.
  - Unconfoundedness: $(Y(0), Y(1)) \perp\!\!\!\perp Z|X$. Furthermore, we have
- **Positivity/Overlap**: the propensity score $e(X) = P(Z = 1|X)$ must satisfy $0 < e(X) < 1$ with probability 1.
- Or, **strict positivity/overlap**[1]: $c_1 \leq e(X) \leq c_2$ for some $0 < c_1 < c_2 < 1$ with probability 1.

[1] Hirano, K. *et al.* Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica* **71,** 1161–1189 (2003).

Overview

Positivity vs. Strict positivity[2]

- Strict positivity implies positivity.
- Positivity is sufficient for non-parametric identification for ATE.
- But the existence of regular semiparametric estimators of ATE requires strict positivity.
- Strict positivity is assumed by a large body of literature.
- Not easy to relax strict positivity , which may involve non-standard asymptotic analyses.

[2] D'Amour, A. *et al.* Overlap in observational studies with high-dimensional covariates. *Journal of Econometrics* **221,** 644–654 (2021).

Overview

A quick view of the remaining slides

- ATE-type inference (30%)
  - Trimming, smooth trimming, and truncation methods
  - The weighted ATE (WATE) framework
  - What are we weighting for?
- ATT-type inference (70%; our original and most recent work)
  - Extended the WATE framework to ATT-type estimation and inference – proposed weighted ATT (WATT)
  - Overlap weighted ATT (OWATT) as a special member
  - Simulation and data example
- Discussion and future research

Positivity violation

The average treatment effect (ATE): $\tau = \mathbb{E}\{Y(1) - Y(0)\} = \mathbb{E}\{\tau(X)\}$ where $\tau(X) = \mathbb{E}\{Y(1) - Y(0)|X\}$. The ATE under assumptions aforementioned can also be written as

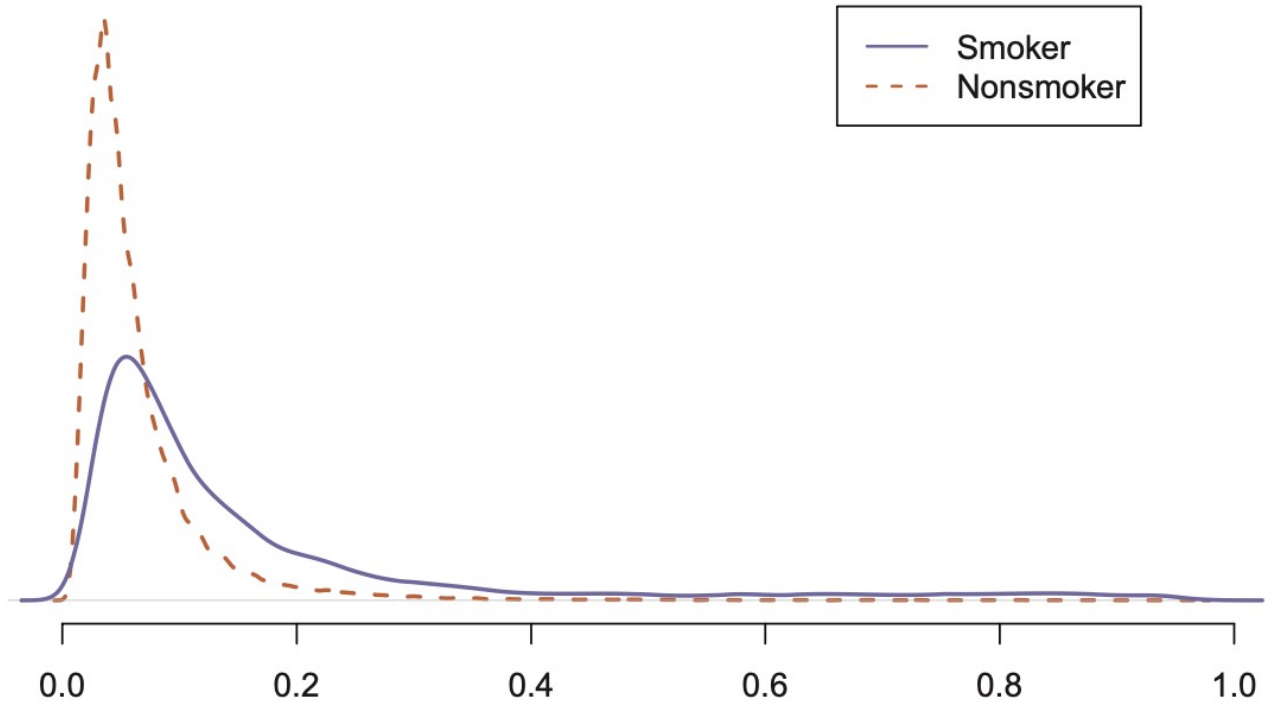
$$\tau = \mathbb{E}\left\{\frac{ZY}{e(X)} - \frac{(1-Z)Y}{1-e(X)}\right\}.$$

The weights $\left(\frac{1}{e(X)}, \frac{1}{1-e(X)}\right)$—inverse probability weights (IPW), generate a pseudo-population by re-weighting the contribution of each participant. The weights reflect that $e(X)$ (in treated group) and $1 - e(X)$ (in control group) shouldn't be nor close 0 or there can be extremely large weights.

- Extreme weights affect the finite-sample performance and hurt the asymptotic normality.
- $e(X)$ (of treated) close to 1 and $1 - e(X)$ (of control) close to 1 incur poor overlap of propensity score distributions of two treatment groups.

Positivity violation - Example

Extreme weights: NC birth weights data[3]

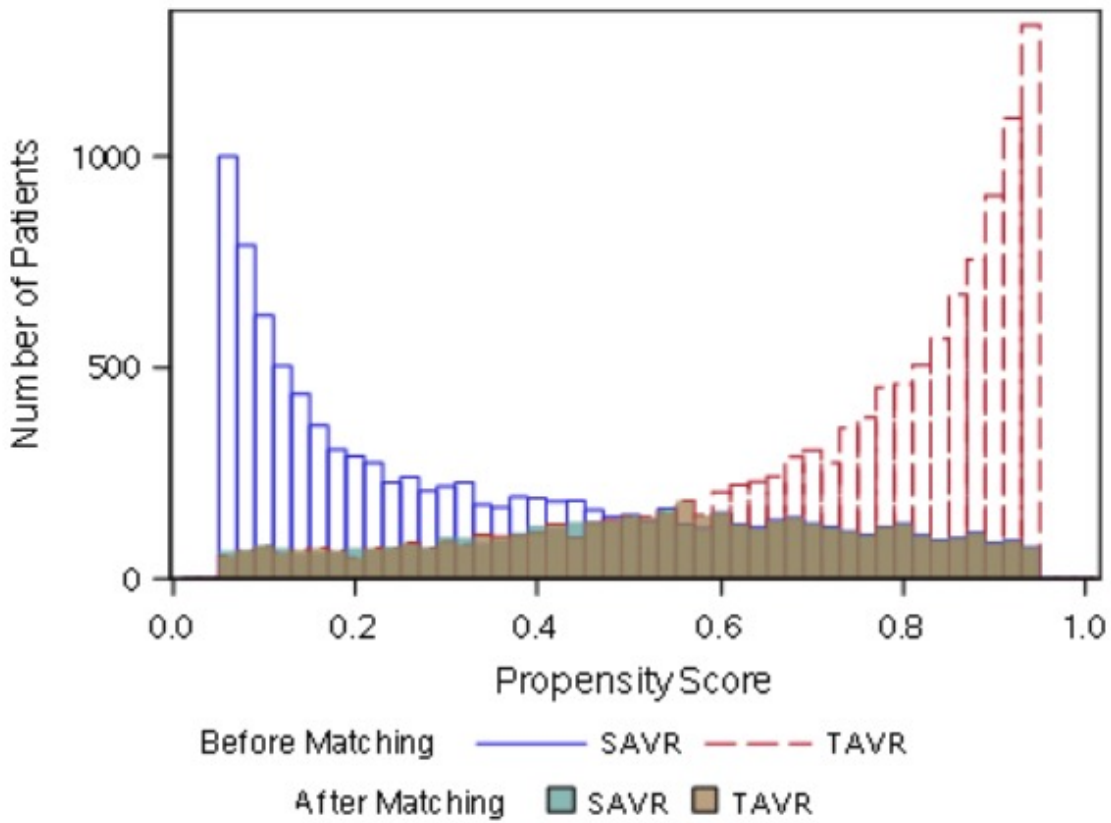


	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Smoker	0.0044	0.06	0.10	0.17	0.18	0.98
Nonsmoker	0.0014	0.03	0.05	0.07	0.08	0.97

[3] Zhou, Y. *et al.* Propensity score weighting under limited overlap and model misspecification. *Statistical Methods in Medical Research* **29,** 3721–3756 (2020).

Positivity violation - Example

Poor overlap



Types of positivity violation

The literature defines two types of violation of positivity.[4]

- Random violation (i.e., by chance): due to small sample sizes, model misspecifications, etc.
  - Reparameterization and direct optimization can offer better ATE estimation via IPW.
- Structural violation: expected due to the inherent characteristics of the target population.
  - ATE via IPW is technically not identifiable.

[4] Petersen, M. L. *et al.* Diagnosing and responding to violations in the positivity assumption. *Statistical methods in medical research* **21,** 31–54 (2012).

Trimming or truncating extreme weights

Two common practices for excluding/capping extreme weights:

- Trimming: exclude participants with estimated $e(X)$ outside a range $[c_1, c_2]$, where $0 < c_1 < c_2 < 1$.
  - Symmetric trimming[5]: $c_1 = \alpha$, $c_2 = 1 - \alpha$ for some $\alpha \in (0, 0.5)$.
    - A rule of thumb: $\alpha = 0.1$.
  - Asymmetric trimming[6]:
    1. Exclude participants with $e(X)$ outside the common $e(X)$ range formed by the treated and control.
    2. Among all participants, also exclude those whose $e(X)$ is below the qth quantile of the treated, and those whose $e(X)$ is above the $(1 - q)$th quantile of the control.
- Truncation: a weight capping, i.e., assign c_1 as the new propensity score to those $e(X) < c_1$ and c_2 to those $e(X) > c_2$.
- In our discussion later, we focus on the symmetric case of both trimming and truncation, i.e., $c_1 = \alpha$, $c_2 = 1 - \alpha$ for illustration and simplicity.

[5] Crump, R. K. *et al.* Dealing with limited overlap in estimation of average treatment effects. *Biometrika* **96,** 187–199 (2009), Crump, R. *et al. Moving the goalposts: Addressing limited overlap in the estimation of average treatment effects by changing the estimand.* 2006.

[6] Stürmer, T. *et al.* Treatment effects in the presence of unmeasured confounding: dealing with observations in the tails of the propensity score distribution—a simulation study. *American journal of epidemiology* **172,** 843–854 (2010).

Moving the goalpost

- There is a (even asymptotically) non-negligible bias for ATE estimation when using trimming or truncation, thus it is detrimental to statistical inference.
- In fact, they moved the goalpost of inference. For example, the (symmetric) trimming targets $\mathcal{O}(X) = \{X : \alpha < e(X) < 1 - \alpha\}$.
- Some practical considerations on trimming and truncation:
  - Ad-hoc selection on threshold α.
  - The effect estimate is often sensitive to change of the threshold.

Moving the goalpost: weighted ATE (WATE)

- The IPW $\left(\frac{1}{e(X)}, \frac{1}{1-e(X)}\right)$ is a special case of a general class of balancing weights:[7]

$$\left(\frac{h(X)}{e(X)}, \frac{h(X)}{1-e(X)}\right),$$

  where $h(\cdot)$ is called a tilting/selection function.
- Trimming: $h(X) = \mathbb{1}\{\alpha < e(X) < 1-\alpha\}$.
- Truncation: $h(X) = \mathbb{1}\{\alpha < e(X) < 1-\alpha\} + \alpha^{-1} e(X)\mathbb{1}\{e(X) \leq \alpha\} + (1-\alpha)^{-1}\{1-e(X)\}\mathbb{1}\{e(X) \geq 1-\alpha\}$.
- Weights for identifying ATT is by $h(X) = e(X)$, i.e., $\left(1, \frac{e(X)}{1-e(X)}\right)$.
- General target estimand by balancing weights: weighted average treatment effect (WATE), defined by

$$\tau_{wate} = \frac{\mathbb{E}\{h(X)\tau(X)\}}{\mathbb{E}\{h(X)\}}, \text{ where } \tau(X) = \mathbb{E}\{Y(1) - Y(0)|X\}.$$

  Clearly when $h(X) \propto 1$, it is just ATE.

[7] Li, F. *et al.* Balancing covariates via propensity score weighting. *Journal of the American Statistical Association* **113,** 390–400 (2018).

Smooth weights for trimming

Yang and Ding[8] proposed a smooth version of trimming. The smooth trimming replace the $h(X)$ of trimming, i.e., $\mathbb{1}\{\alpha < e(X) < 1-\alpha\}$ by

$$h_\varepsilon(X) = \Phi_\varepsilon(e(X)-\alpha)\Phi_\varepsilon(1-\alpha-e(X)),$$

where $\Phi_\varepsilon(\cdot)$ is the CDF of $\mathcal{N}(0,\varepsilon^2)$ distribution.

- $h_\varepsilon(X) \to \mathbb{1}\{\alpha < e(X) < 1-\alpha\}$ as $\varepsilon \to 0$. The smooth trimming still targets $\mathcal{O}(X) = \mathbb{1}\{X : \alpha < e(X) < 1-\alpha\}$ when ε is small.
- Bias-variance trade-off: smoothing reduces the variance.

[8] Yang, S & Ding, P. Asymptotic inference of causal effects with observational studies trimmed by the estimated propensity scores. *Biometrika* **105,** 487–493 (2018).

Overlap weights

Li et al.[9] proposed the overlap weights (OW) who has the tilting function $h(X) = e(X)\{1 - e(X)\}$. The target estimand is called the average treatment effect on the overlap (ATO).

- The balancing weights by OW are $\left(\frac{h(X)}{e(X)}, \frac{h(X)}{1-e(X)}\right) = (1 - e(X), e(X))$.
- Intuition on OW: the $h(X)$ achieves the maximal at $e(X) = 0.5$ and is decreasing at both sides of $e(X) = 0.5$, thus it targets/highlights a the "clinical equipose" population, mimicking the trait of a randomized clinical trial where the propensity score of each participant is 0.5.
- Some concerns on OW and ATO: OW shifts the target of inference and in practice can be based only on estimated propensity score.
  - Zhou et al.[10] and Li et al.[11] further provided some numerical analyses and analytic proofs.

---

[9] Li, F. *et al.* Balancing covariates via propensity score weighting. *Journal of the American Statistical Association* **113,** 390–400 (2018).

[10] Zhou, Y. *et al.* Propensity score weighting under limited overlap and model misspecification. *Statistical Methods in Medical Research* **29,** 3721–3756 (2020).

[11] Li, F. *et al.* Addressing extreme propensity scores via the overlap weights. *American journal of epidemiology* **188,** 250–257 (2018).

WATE summary

Tilting functions of trimming, smooth trimming and overlap weights:[12]

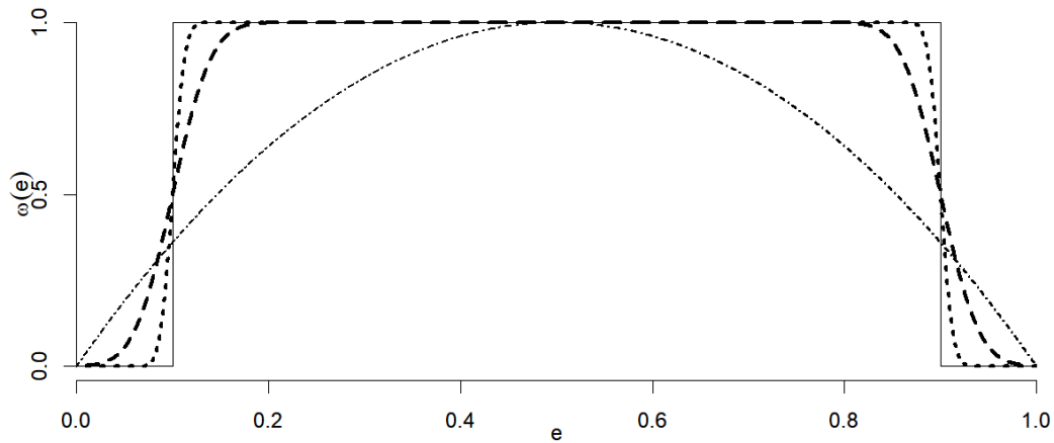


Fig. 1. Weight functions: the solid line is $1\{0.1 \leq e \leq 0.9\}$; the thick dot and dash lines are the smooth weight function $\omega_\epsilon(e)$ with $\epsilon = 0.0001$ and $\epsilon = 0.001$, respectively; the thin dash line is the overlap weight function $\omega(e) \propto e(1-e)$ in Li et al. (2016) up to some constant.

[12] Yang, S & Ding, P. Asymptotic inference of causal effects with observational studies trimmed by the estimated propensity scores. *Biometrika* **105,** 487–493 (2018).

What are we weighting for?

- Matsouaka et al. also assessed the impact of $p = P(Z = 1)$, i.e., the proportion of treated, on different WATE estimators via an extensive simulation study.[13]
  - What impacts your p? Is it just by chance by your sample (e.g., education research interventions tend to enroll a smaller number of participants than can be found in the general population), or it is a characteristic of the target population (e.g., exposure to toxic chemicals).
  - ATE = pATT + $(1 - p)$ATC
  - ATO has the opposite result, i.e., when p is large, it is closer to ATC (simulation results only).
- Practice suggestion: Beware of what you ultimately get by using a specific weighting method. ATE may not lead you where you expected; ATO takes you to the overlap land.

[13] Matsouaka, R. A. *et al.* Overlap, matching, or entropy weights: what are we weighting for? *arXiv preprint arXiv:2210.12968* (2022).

Positivity issue in ATT estimation

- The average treatment effect on the treated (ATT) is defined by $\tau_{att} = \mathbb{E}\{Y(1) - Y(0)|Z = 1\} = \mathbb{E}\{Y|Z = 1\} - \mathbb{E}\{Y(0)|Z = 1\}$.
- It can be shown that

$$\tau_{att} = \frac{\mathbb{E}(ZY)}{\mathbb{E}(Z)} - \frac{\mathbb{E}\{w_0(X)(1-Z)Y\}}{\mathbb{E}\{w_0(X)(1-Z)\}},$$

  where $w_0(X) = \dfrac{e(X)}{1-e(X)}$.

- Extreme weights occur when $e(X) \approx 1$ in control participants.
- A weighting estimator for ATT:

$$\widehat{\tau}_{att} = \frac{\sum_{i=1}^N Z_i Y_i}{\sum_{i=1}^N Z_i} - \frac{\sum_{i=1}^N (1-Z_i)\widehat{w}_0(X_i) Y_i}{\sum_{i=1}^N (1-Z_i)\widehat{w}_0(X_i)}.$$

Positivity issue in ATT estimation

- The key differences compared to ATE identification: (1) We do not have weights on treated participants; (2) We only miss $Y(0)$ for treated group.
- Positivity assumptions for identifying ATT using weighting:
  - (a) $P(Z = 1) > 0$. We need a fraction of the population to receive treatment.
  - (b) $e(X) < 1$ with probability 1 on control participants.
- More insights can be found from Abadie et al. and Heckman et al. for these assumptions.[14]

[14] Abadie, A. & Imbens, G. W. *Matching on the estimated propensity score.* Tech. rep. (National Bureau of Economic Research, 2009), Heckman, J. J. *et al.* Matching as an econometric evaluation estimator. *The review of economic studies* **65,** 261–294 (1998).

Moving the goalpost: weighted ATT (WATT)

The WATT is defined by:

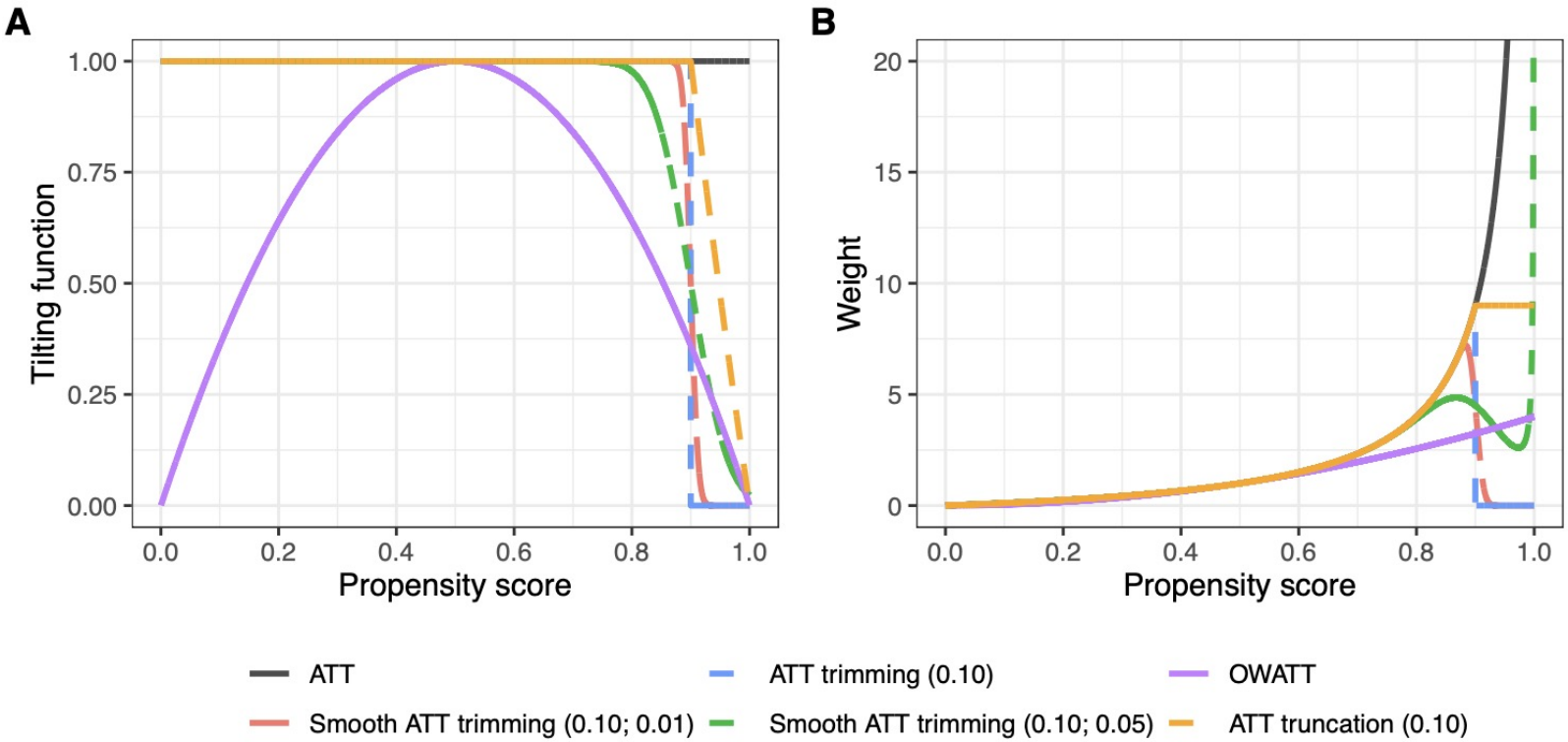
$$\tau^h_{watt} = \frac{\mathbb{E}(ZY)}{\mathbb{E}(Z)} - \frac{\mathbb{E}\{\omega_{0h}(X)(1-Z)Y\}}{\mathbb{E}\{\omega_{0h}(X)(1-Z)\}}, \text{ with } \omega_{0h}(x) = w_0(x)h(x) = \frac{e(x)h(x)}{1-e(x)}.$$

- $h(x)$ is the tilting function again. It generalizes the weights on control and thus generalizes the estimand.
- A weighting estimator for ATT:

$$\widehat{\tau}^h_{watt} = \frac{\sum_{i=1}^N Z_i Y_i}{\sum_{i=1}^N Z_i} - \frac{\sum_{i=1}^N (1-Z_i)\widehat{\omega}_{0h}(X_i)Y_i}{\sum_{i=1}^N (1-Z_i)\widehat{\omega}_{0h}(X_i)}.$$

Overlap weighted ATT (OWATT)

Figures of $h(X)$ (A) and the resulting weight on control (B):



Panel A: tilting functions $h(x)$; Panel B: PS weights $\omega_{0h}(x) = \dfrac{e(x)h(x)}{1-e(x)}$.
For trimming or truncation, $\alpha = 0.1$; also for smooth trimming $\varepsilon = 0.01$ (red line) and 0.05 (green line).
For OWATT, we use $h(x) = 4e(x)\{1-e(x)\}$ (purple line) for illustration and comparison purposes.

Overlap weighted ATT (OWATT)

Under some regularity conditions, the estimator $\widehat{\tau}^h_{watt}$ is regular and asymptotic linear (RAL), if the propensity score is specified by some generalized linear model $e(X) = e(X;\beta)$. Furthermore,

$$\sqrt{N}(\widehat{\tau}^h_{watt} - \tau^h_{watt}) \to_d \mathcal{N}(0, \sigma^2 + b_1'\mathcal{I}(\beta^*)^{-1}b_1 - b_2'\mathcal{I}(\beta^*)^{-1}b_2),$$

where $\sigma^2 = \sum_{z=0}^{1} \mathbb{E}\left\{\eta_z(X)\{\mu\{z, e(X)\}^2 + \sigma^2\{z, e(X)\} + \sigma^2(z, X)\}\right\}$ with

$$\eta_1(X) = \frac{e(X)}{\mathbb{E}\{e(X)\}^2}, \quad \eta_0(X) = \frac{\omega_{0h}(X)^2\{1 - e(X)\}}{\mathbb{E}\{e(X)h(X)\}^2},$$

$$\mu\{z, e(X)\} = \mathbb{E}\{Y|e(X), Z = z\},$$

$$\sigma^2\{z, e(X)\} = \text{var}\{Y|e(X), Z = z\},$$

$$\sigma^2(z, X) = \text{var}\{Y|X, Z = z\}, \quad \text{for } z = 0, 1,$$

where $\mathcal{I}(\beta^*)$ is the Fisher's information matrix of β, with β^* the truth of β, and

$$b_1' = \mathbb{E}\left\{\frac{\partial}{\partial\beta'}\left[\frac{e(X'\beta^*)}{\mathbb{E}\{e(X'\beta^*)\}}\right]\mu(1, X) - \frac{\partial}{\partial\beta'}\left[\frac{e(X'\beta^*)h(X'\beta^*)}{\mathbb{E}\{e(X'\beta^*)h(X'\beta^*)\}}\right]\mu(0, X)\right\},$$

$$b_2' = \mathbb{E}\left\{\left[\frac{\mathbb{E}\{X\mu(1, X)|e(X)\}}{\mathbb{E}\{e(X)\}} + \frac{\omega_{0h}(X)\mathbb{E}\{X\mu(0, X)|e(X)\}}{\mathbb{E}\{e(X)h(X)\}}\right]f(X)\right\}.$$

Overlap weighted ATT (OWATT)

Remarks from the previous theorem:

- The asymptotic linearity allows the use of bootstrap for variance estimation.
- In the asymptotic variance term, $\eta_0(X) = \frac{\omega_{0h}(X)^2\{1-e(X)\}}{\mathbb{E}\{e(X)h(X)\}^2}$. Thus,
  - when $h(x) \propto 1$, $\eta_0(X) \propto e(x)^2/\{1-e(x)\}$, which can still be extreme.
  - when $h(x) \propto e(x)\{1-e(x)\}$ (overlap tilting function), $\eta_0(x) \propto e(x)^4\{1-e(x)\}$.

Overlap weighted ATT (OWATT)

When the propensity score model is possibly misspecified and equal to $\widetilde{e}(X)$, the asymptotic bias of estimating τ^h_{watt} by $\widehat{\tau}^h_{watt}$ is

$$\text{ABias}(\widehat{\tau}^h_{watt}) = \frac{\mathbb{E}\{\omega_{0h}(X)\{1-e(X)\}m_0(X)\}}{\mathbb{E}\{\omega_{0h}(X)\{1-e(X)\}\}} - \frac{\mathbb{E}\{\widetilde{\omega}_{0h}(X)\{1-e(X)\}m_0(X)\}}{\mathbb{E}\{\widetilde{\omega}_{0h}(X)\{1-e(X)\}\}},$$

where $m_0(X) = \mathbb{E}\{Y(0)|X\}$. Clearly, whether we choose $h(x) = 1$ vs. $h(x) = e(x)\{1-e(x)\}$, the asymptotic bias is, respectively,

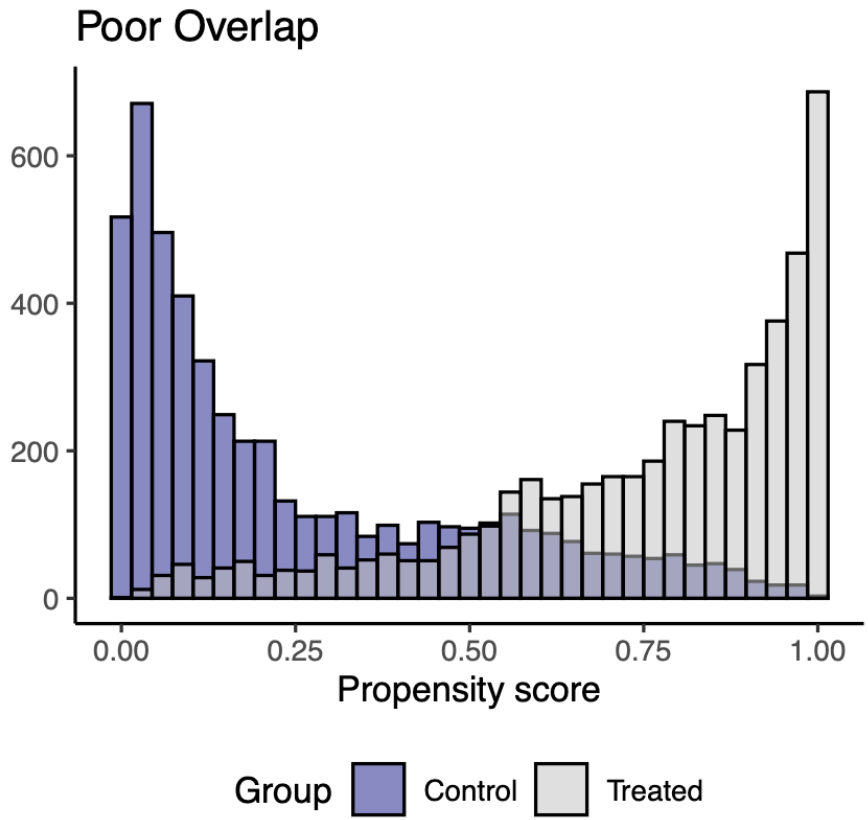
$$\frac{\mathbb{E}\{e(X)m_0(X)\}}{\mathbb{E}\{e(X)\}} - \frac{\mathbb{E}\left\{\dfrac{\widetilde{e}(X)}{1-\widetilde{e}(X)}\{1-e(X)\}m_0(X)\right\}}{\mathbb{E}\left\{\dfrac{\widetilde{e}(X)}{1-\widetilde{e}(X)}\{1-e(X)\}\right\}}$$

$$\text{or} \quad \frac{\mathbb{E}\{e(X)^2\{1-e(X)\}m_0(X)\}}{\mathbb{E}\{e(X)^2\{1-e(X)\}\}} - \frac{\mathbb{E}\left\{\widetilde{e}(X)^2\{1-e(X)\}m_0(X)\right\}}{\mathbb{E}\left\{\widetilde{e}(X)^2\{1-e(X)\}\right\}}.$$

This shows that the latter bias is less sensitive and less extreme to large values of $\widetilde{e}(x)$, i.e., when $\widetilde{e}(x) \to 1$.

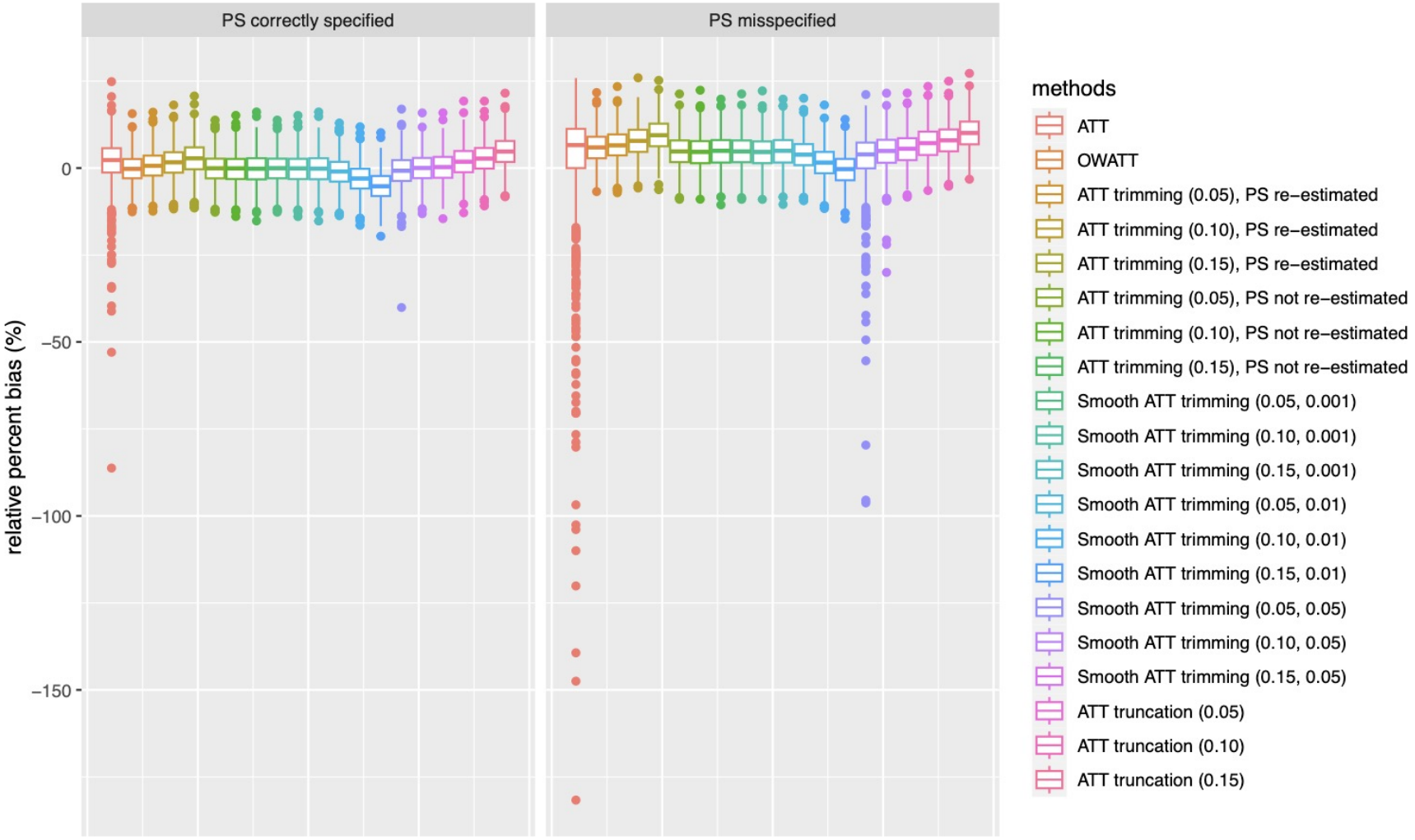
Simulation study

We conducted a simulation study with propensity score model such that the overlap is as follows. There are certain extreme weights as well by this model.



The results are assessed by relative bias, RMSE and coverage probability using bootstrap variance estimation. We also compared the cases where the propensity score model is correctly specified and misspecified.

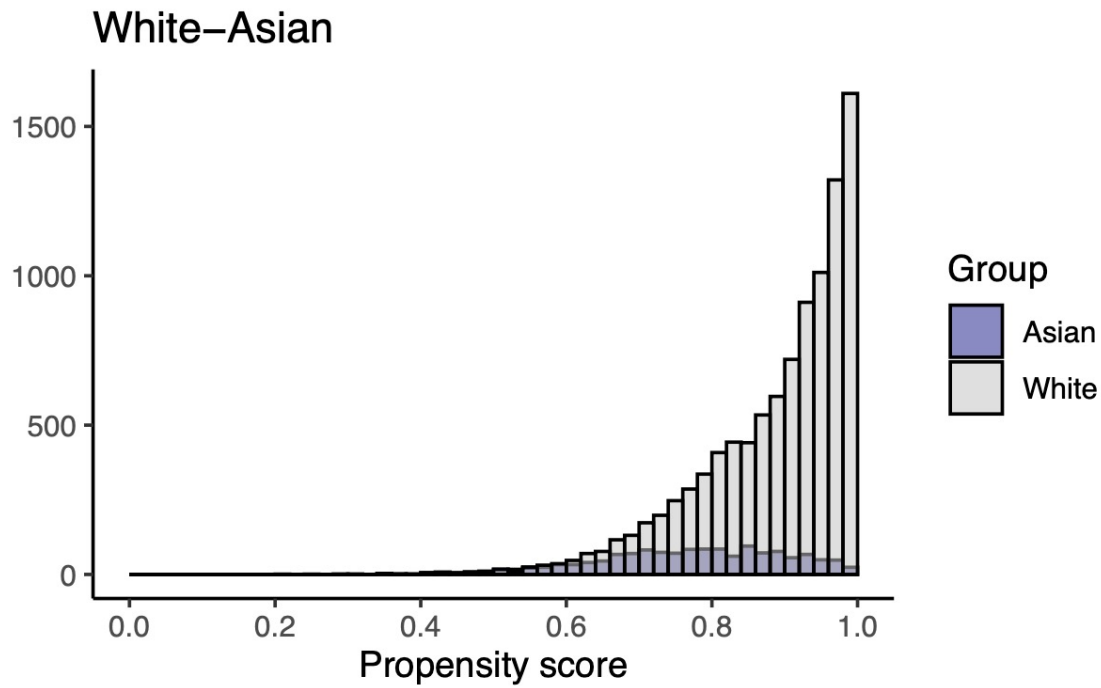
Simulation study



For smooth ATT trimming methods, the parameter in the bracket is (α, ε), i.e., trimming threshold and standard error of the normal cdf in the tilting function, respectively.

Racial disparities in health care expenditure

- Data from the Medical Expenditure Panel Survey (MEPS): `https://www.meps.ahrq.gov/mepsweb/`
- We include 11276 individuals, with 9830 (87.18%) non-Hispanic White as treated and 1446 (12.82%) Asian as control. We included 31 covariates, and considered the health care expenditure as the outcome of interest.



Racial disparities in health care expenditure

Method	Point estimate	Standard error	p-value
ATT	2399.32	787.37	0.002
OWATT	**2511.91**	**255.20**	< 0.001
ATT trimming ($\alpha = 0.05$), PS re-estimated	2363.09	403.42	< 0.001
ATT trimming ($\alpha = 0.10$), PS re-estimated	2666.13	356.62	< 0.001
ATT trimming ($\alpha = 0.15$), PS re-estimated	3054.09	352.98	< 0.001
ATT trimming ($\alpha = 0.05$), PS not re-estimated	2487.25	352.16	< 0.001
ATT trimming ($\alpha = 0.10$), PS not re-estimated	2928.39	286.52	< 0.001
ATT trimming ($\alpha = 0.15$), PS not re-estimated	3286.90	270.04	< 0.001
Smooth ATT trimming ($\alpha = 0.05, \varepsilon = 0.001$)	2488.98	348.88	< 0.001
Smooth ATT trimming ($\alpha = 0.10, \varepsilon = 0.001$)	2926.52	285.92	< 0.001
Smooth ATT trimming ($\alpha = 0.15, \varepsilon = 0.001$)	3291.05	268.68	< 0.001
Smooth ATT trimming ($\alpha = 0.05, \varepsilon = 0.01$)	2419.59	327.68	< 0.001
Smooth ATT trimming ($\alpha = 0.10, \varepsilon = 0.01$)	2881.88	277.57	< 0.001
Smooth ATT trimming ($\alpha = 0.15, \varepsilon = 0.01$)	3229.41	259.47	< 0.001
Smooth ATT trimming ($\alpha = 0.05, \varepsilon = 0.05$)	2337.55	373.65	< 0.001
Smooth ATT trimming ($\alpha = 0.10, \varepsilon = 0.05$)	2638.19	250.78	< 0.001
Smooth ATT trimming ($\alpha = 0.15, \varepsilon = 0.05$)	3014.23	232.06	< 0.001
ATT truncation ($\alpha = 0.05$)	1945.35	385.00	< 0.001
ATT truncation ($\alpha = 0.10$)	2211.56	307.63	< 0.001
ATT truncation ($\alpha = 0.15$)	2419.23	271.39	< 0.001

Racial disparities in health care expenditure

- OWATT estimator has the smallest standard error, except when compared to smooth trimming for $(\alpha, \varepsilon) = (0.15, 0.05)$ (both parameters are large).
- Trimming, smooth trimming and truncation: the point estimates increase as the threshold α increases.
- ATT estimator has obviously the largest standard error.

Discussion

Summary

- Our proposal on weighted ATT under lack of positivity.
- OWATT has some practical advantages:
  - No selection on any threshold parameters like trimming.
  - Statistically sound and efficient under lack of positivity.
- The methodology can easily be extended to ATC.

Future research on WATT

- Semiparametric efficiency estimation via augmentation, other robust estimator, . . .
- Empirical sandwich variance estimation
- Multi-valued treatment extension
- User-friendly R package development, similar to the `PSweight` and `PSW` packages.

Discussion

Other related work

- Data-driven based trimming/truncation for targeting ATE or ATT under lack of positivity[15]
  - They use trimming to defend poor overlap and violation of positivity, and use some bias-correction procedures after trimming. They showed their methods are asymptotically unbiased and normal. The choice of trimming threshold is also data-driven, avoiding ad-hoc selections.
  - Practical limitation: too technical and no available software developed.
  - Only reliable when the lack of positivity is by happenstance. When there is structural violation of positivity, we cannot guarantee that these methods truly estimate the ATE or ATT.

Other questions

- Can we develop methods to distinguish random and structural violations of positivity?
- Sandwich variance is invalid for trimming due to the non-smooth indicator function. Can we use smoothing and what ε to be chosen?

[15] Ma, X. & Wang, J. Robust inference using inverse probability weighting. *Journal of the American Statistical Association* **115,** 1851–1860 (2020), Chaudhuri, S. & Hill, J. B. *Heavy tail robust estimation and inference for average treatment effects.* Tech. rep. (Working paper, 2014), Sasaki, Y. & Ura, T. Estimation and inference for moments of ratios with robustness against large trimming bias. *Econometric Theory* **38,** 66–112 (2022).

References I

1. Liu, Y., Li, H., Zhou, Y. & Matsouaka, R. Average treatment effect on the treated, under lack of positivity. *arXiv preprint arXiv:2309.01334* (2023).
2. Matsouaka, R. A., Liu, Y. & Zhou, Y. Overlap, matching, or entropy weights: what are we weighting for? *arXiv preprint arXiv:2210.12968* (2022).
3. Matsouaka, R. A. & Zhou, Y. A framework for causal inference in the presence of extreme inverse probability weights: the role of overlap weights. *arXiv preprint arXiv:2011.01388* (2020).
4. Zhou, Y., Matsouaka, R. A. & Thomas, L. Propensity score weighting under limited overlap and model misspecification. *Statistical Methods in Medical Research* **29,** 3721–3756 (2020).
5. Hirano, K., Imbens, G. W. & Ridder, G. Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica* **71,** 1161–1189 (2003).
6. D'Amour, A., Ding, P., Feller, A., Lei, L. & Sekhon, J. Overlap in observational studies with high-dimensional covariates. *Journal of Econometrics* **221,** 644–654 (2021).
7. Petersen, M. L., Porter, K. E., Gruber, S., Wang, Y. & van der Laan, M. J. Diagnosing and responding to violations in the positivity assumption. *Statistical methods in medical research* **21,** 31–54 (2012).

References II

8. Crump, R. K., Hotz, V. J., Imbens, G. W. & Mitnik, O. A. Dealing with limited overlap in estimation of average treatment effects. *Biometrika* **96,** 187–199 (2009).

9. Crump, R., Hotz, V. J., Imbens, G. & Mitnik, O. *Moving the goalposts: Addressing limited overlap in the estimation of average treatment effects by changing the estimand.* 2006.

10. Stürmer, T., Rothman, K. J., Avorn, J. & Glynn, R. J. Treatment effects in the presence of unmeasured confounding: dealing with observations in the tails of the propensity score distribution—a simulation study. *American journal of epidemiology* **172,** 843–854 (2010).

11. Li, F., Morgan, K. L. & Zaslavsky, A. M. Balancing covariates via propensity score weighting. *Journal of the American Statistical Association* **113,** 390–400 (2018).

12. Yang, S & Ding, P. Asymptotic inference of causal effects with observational studies trimmed by the estimated propensity scores. *Biometrika* **105,** 487–493 (2018).

13. Li, F., Thomas, L. E. & Li, F. Addressing extreme propensity scores via the overlap weights. *American journal of epidemiology* **188,** 250–257 (2018).

14. Abadie, A. & Imbens, G. W. *Matching on the estimated propensity score.* Tech. rep. (National Bureau of Economic Research, 2009).

References III

15. Heckman, J. J., Ichimura, H. & Todd, P. Matching as an econometric evaluation estimator. *The review of economic studies* **65,** 261–294 (1998).

16. Ma, X. & Wang, J. Robust inference using inverse probability weighting. *Journal of the American Statistical Association* **115,** 1851–1860 (2020).

17. Chaudhuri, S. & Hill, J. B. *Heavy tail robust estimation and inference for average treatment effects.* Tech. rep. (Working paper, 2014).

18. Sasaki, Y. & Ura, T. Estimation and inference for moments of ratios with robustness against large trimming bias. *Econometric Theory* **38,** 66–112 (2022).

Thank you!

Q & A

Email: yliu297@ncsu.edu

Find me on Twitter: @YiLiu1998