Robustness and Efficiency of Rosenbaum's Rank-based Estimator in Randomized Trials: A Design-based Perspective

(Forthcoming in *Biometrika*)

Aditya Ghosh

Statistics Ph.D. student, Stanford University

Joint work with Nabarun Deb (University of Chicago), Bikram Karmakar (University of Wisconsin-Madison) & Bodhi Sen (Columbia University)

Motivation

- Heavy-tailed outcomes are common in modern randomized experiments

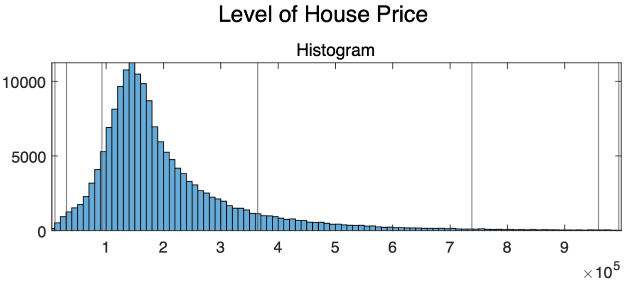


Figure 1: An example of heavy-tailed outcomes from Athey et al. (2023)

Motivation

- Heavy-tailed outcomes are common in modern randomized experiments
- Mean-based estimators of causal effects may behave poorly if outcomes have a *heavy tail* or contain *outliers*
  - Hence confidence intervals can be too wide to be useful!

Motivation

- Heavy-tailed outcomes are common in modern randomized experiments
- Mean-based estimators of causal effects may behave poorly if outcomes have a *heavy tail* or contain *outliers*
- One solution: Use the *ranks* of the outcomes instead their values
  - Popular in the Nonparametric Statistics literature
  - How to employ ranks in this Causal context?

Motivation

- Heavy-tailed outcomes are common in modern randomized experiments
- Mean-based estimators of causal effects may behave poorly if outcomes have a *heavy tail* or contain *outliers*
- One solution: Use the *ranks* of the outcomes instead their values
- Rosenbaum (1993)[1] proposed a Hodges-Lehmann type estimator that inverts a rank-based test.

[1] Rosenbaum, P. R. (1993). Hodges-Lehmann point estimates of treatment effect in observational studies. *JASA* 88, 1250–1253.

Motivation

- Heavy-tailed outcomes are common in modern randomized experiments
- Mean-based estimators of causal effects may behave poorly if outcomes have a *heavy tail* or contain *outliers*
- One solution: Use the *ranks* of the outcomes instead their values
- Rosenbaum (1993)[1] proposed a Hodges-Lehmann type estimator that inverts a rank-based test.

What are the robustness and efficiency properties of Rosenbaum's rank-based estimator under the *design-based* framework?

[1] Rosenbaum, P. R. (1993). Hodges-Lehmann point estimates of treatment effect in observational studies. *JASA* 88, 1250–1253.

Design-based framework

Randomized trial: m subjects chosen uniformly at random from N subjects.

- $Z_i =$ treatment indicator, $Y_i =$ outcome for i-th subject, given by

$$Y_i = Z_i a_i + (1 - Z_i) b_i = \begin{cases} a_i & \textit{if} \quad \textit{treated}, \\ b_i & \textit{if} \quad \textit{control}. \end{cases}$$

Design-based framework

Randomized trial: m subjects chosen uniformly at random from N subjects.

- $Z_i =$ treatment indicator, $Y_i =$ outcome for i-th subject, given by

$$Y_i = Z_i a_i + (1 - Z_i) b_i = \begin{cases} a_i & \textit{if} \quad \textit{treated}, \\ b_i & \textit{if} \quad \textit{control}. \end{cases}$$

- **Randomization/Design-based/finite-population inference:**

The potential outcomes a_i and b_i are fixed, treatment indicators Z_i are the only source of randomness.

$+$ Cleaner: Validity by design rather than modeling assumptions

$-$ Proofs are often harder than in model-based (i.i.d.) framework

Rosenbaum's estimator

- Constant additive treatment effect model: (Rosenbaum, 1993, 2002)

$$a_i - b_i = \tau \quad \text{for each} \quad 1 \leq i \leq N.$$

  - Analog of the location shift model from Nonparametric Statistics.
  - Generalizes Fisher's sharp null (and often a convenient starting point).
  - We will relax this at the end of the talk.

Rosenbaum's estimator

- Constant additive treatment effect model: (Rosenbaum, 1993, 2002)

$$a_i - b_i = \tau \quad \text{for each} \quad 1 \leq i \leq N.$$

- Testing $H_0 : \tau = \tau_0$ vs $H_1 : \tau \neq \tau_0$.
  - Under H_0,

$$Y_i - \tau_0 Z_i = a_i Z_i + b_i(1 - Z_i) - (a_i - b_i)Z_i = b_i \quad \text{(non-random)}$$

  - Use any test statistic $t(\mathbf{Z}, \mathbf{Y} - \tau_0 \mathbf{Z})$ to draw randomization inference

- Constant additive treatment effect model: (Rosenbaum, 1993, 2002)

$$a_i - b_i = \tau \quad \text{for each} \quad 1 \leq i \leq N.$$

- Testing $H_0 : \tau = \tau_0$ vs $H_1 : \tau \neq \tau_0$.

- We focus on the Wilcoxon rank-sum (WRS) test statistic:

$$t(\mathbf{Z}, \mathbf{Y} - \tau_0 \mathbf{Z}) = \sum_{Z_i = 1} \text{rank of } (Y_i - \tau_0 Z_i) \text{ in the treated} + \text{control pool.}$$

Rosenbaum's estimator

- Constant additive treatment effect model: (Rosenbaum, 1993, 2002)

$$a_i - b_i = \tau \quad \text{for each} \quad 1 \leq i \leq N.$$

- Testing $H_0 : \tau = \tau_0$ vs $H_1 : \tau \neq \tau_0$.

- We focus on the Wilcoxon rank-sum (WRS) test statistic:

$$t(\mathbf{Z}, \mathbf{Y} - \tau_0 \mathbf{Z}) = \sum_{Z_i = 1} \text{rank of } (Y_i - \tau_0 Z_i) \text{ in the treated} + \text{control pool.}$$

- Rosenbaum's estimator $\widehat{\tau}_{\mathrm{R}}$ is obtained by inverting the above test.

"Solve for τ_0 such that $t(\mathbf{Z}, \mathbf{Y} - \tau_0 \mathbf{Z}) = \mu$ where $\mu := \mathbb{E}_{\tau_0} t(\mathbf{Z}, \mathbf{Y} - \tau_0 \mathbf{Z})$"

$$\widehat{\tau}_{\mathrm{R}} := \frac{1}{2} \left(\sup \{\tau_0 : t(\mathbf{Z}, \mathbf{Y} - \tau_0 \mathbf{Z}) > \mu\} + \inf \{\tau_0 : t(\mathbf{Z}, \mathbf{Y} - \tau_0 \mathbf{Z}) < \mu\} \right).$$

Rosenbaum's estimator (contd.)

- Rosenbaum's rank-based estimator inverts the Wilcoxon rank-sum test

$\mathbf{Y} - \tau_0\mathbf{Z}$ → Compute WRS test statistic → Invert the test → $\widehat{\tau}_{\mathrm{R}}$

- Easy to get permutation confidence intervals → inference ✔

Rosenbaum's estimator (contd.)

- Rosenbaum's rank-based estimator inverts the Wilcoxon rank-sum test

$\mathbf{Y} - \tau_0 \mathbf{Z}$ → Compute WRS test statistic → Invert the test → $\widehat{\tau}_{\mathrm{R}}$

- Easy to get permutation confidence intervals $\rightarrow$ inference ✓

> What are the robustness and efficiency properties of this estimator?

Robustness against contamination/outliers

- In Robust Statistics, a natural way to quantify the robustness of an estimator is via its (asymptotic) breakdown point:

> What minimum proportion of data, when replaced with extreme values, makes the estimator arbitrarily large?

- To our knowledge, such definitions are scarce in Causal Inference.

Robustness against contamination/outliers

- In Robust Statistics, a natural way to quantify the robustness of an estimator is via its (asymptotic) breakdown point:

> What minimum proportion of data, when replaced with extreme values, makes the estimator arbitrarily large?

- Under the design-based framework for randomized trials, we ask:

> What minimum proportion of potential outcomes, when replaced with arbitrarily extreme values, will cause the estimator to be arbitrarily large (in absolute value), for *some* treatment assignment?

Robustness against contamination/outliers

- In Robust Statistics, a natural way to quantify the robustness of an estimator is via its (asymptotic) breakdown point:

> What minimum proportion of data, when replaced with extreme values, makes the estimator arbitrarily large?

- Under the design-based framework for randomized trials, we ask:

> What minimum proportion of potential outcomes, when replaced with arbitrarily extreme values, will cause the estimator to be arbitrarily large (in absolute value), for *some* treatment assignment?

- Asymptotic breakdown point (ABP) is the limit of this proportion as the sample size grows to infinity.

Weighted average quantile (WAQ) estimators

[2] Athey, S., Bickel, P. J.,Chen, A., Imbens, G. W. & Pollmann, M. (2023). Semi-parametric estimation of treatment effects in randomised experiments. *JRSSB* 85, 1615–1638.

Weighted average quantile (WAQ) estimators

- Define weighted average quantile (WAQ) estimators as

$$\widehat{\tau}_{waq}(\nu) = \sum_{i=1}^{m} \nu\left(\left[\frac{i-1}{m}, \frac{i}{m}\right]\right) \boxed{Y_{(i)}^1} - \sum_{i=1}^{N-m} \nu\left(\left[\frac{i-1}{N-m}, \frac{i}{N-m}\right]\right) \boxed{Y_{(i)}^0}$$

  where ν is any signed measure with $\nu([0,1]) = 1$, and $Y_{(i)}^1$ (resp. $Y_{(i)}^0$) are the order statistics of the treated (resp. control) group.
- **Examples:**
  - Difference-in-means (when ν is uniform),
  - Difference-in-medians (when ν is point mass at $1/2$),
  - α-trimmed diff-in-means, α-Winsorized diff-in-means,
  - Novel estimators proposed by Athey et al. (2023)[2] (semiparametically efficient under certain classes)

[2] Athey, S., Bickel, P. J.,Chen, A., Imbens, G. W. & Pollmann, M. (2023). Semi-parametric estimation of treatment effects in randomised experiments. *JRSSB* 85, 1615–1638.

Robustness of Rosenbaum's estimator

Assume that $m/N \to \lambda \in (0,1)$ where m is the size of the treated group. Then,

$$\sup_{\nu} \mathrm{ABP}(\widehat{\tau}_{waq}(\nu)) \leq \frac{1}{2}\lambda(1-\lambda) = \mathrm{ABP}(\widehat{\tau}_{\mathrm{R}}).$$

In words, the asymptotic breakdown point (ABP) of Rosenbaum's estimator $\widehat{\tau}_{\mathrm{R}}$ uniformly dominates the ABP of *all* weighted average quantile estimators!

Robustness of Rosenbaum's estimator

Assume that $m/N \to \lambda \in (0,1)$ where m is the size of the treated group. Then,

$$\sup_{\nu} \mathsf{ABP}(\widehat{\tau}_{waq}(\nu)) \leq \frac{1}{2}\lambda(1-\lambda) = \mathsf{ABP}(\widehat{\tau}_{\mathsf{R}}).$$

In words, the asymptotic breakdown point (ABP) of Rosenbaum's estimator $\widehat{\tau}_{\mathsf{R}}$ uniformly dominates the ABP of *all* weighted average quantile estimators!

- Moreover, the ABP of semiparametrically efficient WAQ estimators proposed by Athey et al. (2023) can often be zero, highlighting the trade-off between robustness and efficiency.

Robustness of Rosenbaum's estimator

Assume that $m/N \to \lambda \in (0,1)$ where m is the size of the treated group. Then,

$$\sup_{\nu} \mathrm{ABP}(\widehat{\tau}_{waq}(\nu)) \leq \frac{1}{2}\lambda(1-\lambda) = \mathrm{ABP}(\widehat{\tau}_{\mathrm{R}}).$$

In words, the asymptotic breakdown point (ABP) of Rosenbaum's estimator $\widehat{\tau}_{\mathrm{R}}$ uniformly dominates the ABP of *all* weighted average quantile estimators!

- Moreover, the ABP of semiparametrically efficient WAQ estimators proposed by Athey et al. (2023) can often be zero, highlighting the trade-off between robustness and efficiency.
- The inequality is tight: Difference-in-medians has the same breakdown point, thus requiring a comparison of asymptotic relative efficiency.

Asymptotic distribution

- Permutation CIs do not shed light on (relative) efficiency; which motivates us to derive the asymptotic distribution.

[3] Li, X. & Ding, P. (2017). General forms of finite population central limit theorems with applications to causal inference. *JASA* 112, 1759–1769.

Asymptotic distribution

- Permutation CIs do not shed light on (relative) efficiency; which motivates us to derive the asymptotic distribution.
- Li & Ding (2017)[3]: versatile toolkit for deriving finite-population CLTs.
  - This yields null distribution of our test statistic.

[3] Li, X. & Ding, P. (2017). General forms of finite population central limit theorems with applications to causal inference. *JASA* 112, 1759–1769.

Asymptotic distribution

- Permutation CIs do not shed light on (relative) efficiency; which motivates us to derive the asymptotic distribution.
- Li & Ding (2017)[3]: versatile toolkit for deriving finite-population CLTs.
  - This yields null distribution of our test statistic.
  - However, Rosenbaum's point estimator (which inverts the above test) is highly non-linear and their techniques do not apply.

[3] Li, X. & Ding, P. (2017). General forms of finite population central limit theorems with applications to causal inference. *JASA* 112, 1759–1769.

Asymptotic distribution

- Permutation CIs do not shed light on (relative) efficiency; which motivates us to derive the asymptotic distribution.
- Li & Ding (2017)[3]: versatile toolkit for deriving finite-population CLTs.
  - This yields null distribution of our test statistic.
  - However, Rosenbaum's point estimator (which inverts the above test) is highly non-linear and their techniques do not apply.
- We derive the asymptotic distribution of Rosenbaum's estimator by studying the test statistic under local alternatives.

[3] Li, X. & Ding, P. (2017). General forms of finite population central limit theorems with applications to causal inference. *JASA* 112, 1759–1769.

Fighting test inversion with local alternatives

We now index vectors with a subscript N (sample size). Following is a strategy due to Hodges & Lehmann (1963), adapted to our context.

Fighting test inversion with local alternatives

We now index vectors with a subscript N (sample size). Following is a strategy due to Hodges & Lehmann (1963), adapted to our context.

- WRS test statistic $t_N = t(\mathbf{Z}_N, \mathbf{Y}_N - \tau_0 \mathbf{Z}_N)$ for $H_0 : \tau = \tau_0$.
- Local alternatives: $\tau_N = \tau_0 - h/\sqrt{N}$ where $h \in \mathbb{R}$ is fixed.

Fighting test inversion with local alternatives

We now index vectors with a subscript N (sample size). Following is a strategy due to Hodges & Lehmann (1963), adapted to our context.

- WRS test statistic $t_N = t(\mathbf{Z}_N, \mathbf{Y}_N - \tau_0 \mathbf{Z}_N)$ for $H_0 : \tau = \tau_0$.
- Local alternatives: $\tau_N = \tau_0 - h/\sqrt{N}$ where $h \in \mathbb{R}$ is fixed.
- Suppose we are able to show that, under $\tau = \tau_N$,

$$N^{-3/2} \left(t_N - \mathbb{E}_{\tau_0}[t_N] \right) \xrightarrow{d} \mathcal{N}(-hB, A^2).$$

Fighting test inversion with local alternatives

We now index vectors with a subscript N (sample size). Following is a strategy due to Hodges & Lehmann (1963), adapted to our context.

- WRS test statistic $t_N = t(\mathbf{Z}_N, \mathbf{Y}_N - \tau_0 \mathbf{Z}_N)$ for $H_0 : \tau = \tau_0$.
- Local alternatives: $\tau_N = \tau_0 - h/\sqrt{N}$ where $h \in \mathbb{R}$ is fixed.
- Suppose we are able to show that, under $\tau = \tau_N$,

$$N^{-3/2} \left(t_N - \mathbb{E}_{\tau_0}[t_N]\right) \xrightarrow{d} \mathcal{N}(-hB, A^2).$$

- Then,

$$\sqrt{N} \left(\widehat{\tau}_{\mathrm{R}} - \tau\right) \xrightarrow{d} \mathcal{N}(0, A^2/B^2).$$

The scaling $N^{-3/2}$ is specific to our test statistic, but the above recipe applies to many test statistics.

Local Asymptotic Normality

- Le Cam's results do not apply in our fixed design setting.
- We handle this using decompositions (into a null component and a random bias component)
- Under the local alternative $\tau = \tau_0 - hN^{-1/2}$,

$$\text{Test statistic} = \sum_{Z_i=1} \sum_{j=1}^{N} \mathbf{1}\left\{ \boxed{Y_{N,j} - \tau_0 Z_{N,j} \leq Y_{N,i} - \tau_0 Z_{N,i}} \right\}$$

$$= \sum_{Z_i=1} \sum_{j=1}^{N} \mathbf{1}\left\{ \boxed{b_{N,j} - b_{N,i} \leq \frac{h}{\sqrt{N}}(Z_{N,j} - Z_{N,i})} \right\}$$

- We impose conditions on how many of the pairs $b_{N,j} - b_{N,i}$ belong to intervals like $\left[0, \frac{h}{\sqrt{N}}\right)$ or $\left[-\frac{h}{\sqrt{N}}, 0\right)$.

Local Asymptotic Normality

Define $I_{h,N}(x) = \mathbf{1}\left\{0 \leq x < \frac{h}{\sqrt{N}}\right\}$ for $h > 0$, and $-\mathbf{1}\left\{\frac{h}{\sqrt{N}} \leq x < 0\right\}$ for $h < 0$.

Local Asymptotic Normality

Define $I_{h,N}(x) = \mathbf{1}\left\{0 \leq x < \frac{h}{\sqrt{N}}\right\}$ for $h > 0$, and $-\mathbf{1}\left\{\frac{h}{\sqrt{N}} \leq x < 0\right\}$ for $h < 0$.

We assume that the control potential outcomes $\{b_{N,i}\}_{1 \leq i \leq N}$ satisfy

$$\lim_{N \to \infty} N^{-3/2} \sum_{i=1}^{N} \sum_{j=1}^{N} I_{h,N}(b_{N,j} - b_{N,i}) = h\mathcal{I}_b$$

for every $h \in \mathbb{R}$, where $\mathcal{I}_b \in (0, \infty)$ is a fixed constant.

The proportion of the pairwise differences $b_{N,j} - b_{N,i}$ falling into small intervals (shrinking at the rate of $N^{-1/2}$) scales with the lengths of those intervals.

Local Asymptotic Normality

Define $I_{h,N}(x) = \mathbf{1}\left\{0 \leq x < \frac{h}{\sqrt{N}}\right\}$ for $h > 0$, and $-\mathbf{1}\left\{\frac{h}{\sqrt{N}} \leq x < 0\right\}$ for $h < 0$.

We assume that the control potential outcomes $\{b_{N,i}\}_{1 \leq i \leq N}$ satisfy

$$\lim_{N\to\infty} N^{-3/2} \sum_{i=1}^{N} \sum_{j=1}^{N} I_{h,N}(b_{N,j} - b_{N,i}) = h\mathcal{I}_b$$

for every $h \in \mathbb{R}$, where $\mathcal{I}_b \in (0, \infty)$ is a fixed constant.

The proportion of the pairwise differences $b_{N,j} - b_{N,i}$ falling into small intervals (shrinking at the rate of $N^{-1/2}$) scales with the lengths of those intervals.

Under the sequence $\tau_N = \tau_0 - hN^{-1/2}$, the WRS test statistic t_N satisfies

$$N^{-3/2}\left(t_N - \mathbb{E}_{\tau_0}[t_N]\right) \xrightarrow{d} \mathcal{N}\left(-h\lambda(1-\lambda)\mathcal{I}_b, \frac{\lambda(1-\lambda)}{12}\right).$$

Asymptotic distribution of Rosenbaum's estimator

Assume that

(A1) Treatment is randomly assigned to m units with $m/N \to \lambda \in (0,1)$.

(A2) Constant additive treatment effect: $a_i - b_i = \tau$ for each $1 \le i \le N$.

(A3) The control potential outcomes $\{b_{N,i}\}_{1 \le i \le N}$ satisfy

$$\lim_{N\to\infty} N^{-3/2} \sum_{i=1}^{N} \sum_{j=1}^{N} I_{h,N}(b_{N,j} - b_{N,i}) = h\mathcal{I}_b$$

for every $h \in \mathbb{R}$, where $\mathcal{I}_b \in (0,\infty)$ is a fixed constant.

Then,

$$\sqrt{N}\left(\widehat{\tau}_{\mathrm{R}} - \tau\right) \xrightarrow{d} \mathcal{N}\left(0, \left(12\lambda(1-\lambda)\mathcal{I}_b^2\right)^{-1}\right).$$

A more interpretable version

We can replace the regularity assumption (A3) with the following.

(A3') The empirical distribution F_N of the control potential outcomes $\{b_{N,i}\}$ converges weakly to a distribution with a square-integrable density f, and satisfies a smoothness condition at the $N^{-1/2}$ scale[4].

[4] For every $h \in \mathbb{R}$, $\sup_{x \in \mathbb{R}} |\sqrt{N}(F_N(x + h/\sqrt{N}) - F_N(x)) - h\, f(x)| = o(1)$ as $N \to \infty$

A more interpretable version

We can replace the regularity assumption (A3) with the following.

(A3') The empirical distribution F_N of the control potential outcomes $\{b_{N,i}\}$ converges weakly to a distribution with a square-integrable density f, and satisfies a smoothness condition at the $N^{-1/2}$ scale[4].

- Control potential outcomes behave, in the limit, as i.i.d. from density f.
- Moreover,

$$\mathcal{I}_b = \int_{\mathbb{R}} f^2(x)\, dx.$$

[4] For every $h \in \mathbb{R}$, $\sup_{x\in\mathbb{R}} |\sqrt{N}(F_N(x + h/\sqrt{N}) - F_N(x)) - h\, f(x)| = o(1)$ as $N \to \infty$

A more interpretable version

We can replace the regularity assumption (A3) with the following.

(A3') The empirical distribution F_N of the control potential outcomes $\{b_{N,i}\}$ converges weakly to a distribution with a square-integrable density f, and satisfies a smoothness condition at the $N^{-1/2}$ scale[4].

Under (A1), (A2) and (A3'),

$$\sqrt{N}\,(\widehat{\tau}_{\mathrm{R}} - \tau) \xrightarrow{d} \mathcal{N}\left(0, \left(12\lambda(1-\lambda)\left(\int_{\mathbb{R}} f^2(x)\,dx\right)^2\right)^{-1}\right).$$

[4]For every $h \in \mathbb{R}$, $\sup_{x\in\mathbb{R}} |\sqrt{N}(F_N(x + h/\sqrt{N}) - F_N(x)) - h\,f(x)| = o(1)$ as $N \to \infty$

Efficiency comparisons

- When $(\widehat{\tau}_{N,k} - \tau)/\sigma_{N,k} \xrightarrow{d} \mathcal{N}(0,1)$ $(k = 1, 2)$, we can define asymptotic relative efficiency as

$$\mathrm{eff}(\widehat{\tau}_{N,1}, \widehat{\tau}_{N,2}) = \lim_{N\to\infty} \sigma^2_{N,2}/\sigma^2_{N,1}.$$

$\mathrm{eff}(\widehat{\tau}_{N,1}, \widehat{\tau}_{N,2}) > 1 \implies \widehat{\tau}_{N,1}$ is more efficient

Efficiency comparisons

- $\text{eff}(\widehat{\tau}_{N,1}, \widehat{\tau}_{N,2}) > 1 \implies \widehat{\tau}_{N,1}$ is more efficient
- Under previous assumptions + $N^{-1}\sum_{i=1}^{N}(b_{N,i} - \bar{b}_N)^2 \to \sigma^2$, and $N^{-1}\max_{i \le N}(b_{N,i} - \bar{b}_N)^2 \to 0$,

$$\text{eff}\left(\widehat{\tau}_{\mathrm{R}}, \widehat{\tau}_{\mathrm{dm}}\right) = 12\sigma^2 \left(\int_{\mathbb{R}} f^2(x)dx\right)^2 \ge 0.864 \text{ (worst-case lower bound)}$$

  - Coincides with a famous lower bound on asymptotic relative efficiency of the Wilcoxon rank-sum test and the t-test (Chernoff & Savage)

Efficiency comparisons

- $\text{eff}(\widehat{\tau}_{N,1}, \widehat{\tau}_{N,2}) > 1 \implies \widehat{\tau}_{N,1}$ is more efficient
- Under previous assumptions + $N^{-1}\sum_{i=1}^{N}(b_{N,i} - \bar{b}_N)^2 \to \sigma^2$, and $N^{-1}\max_{i\le N}(b_{N,i} - \bar{b}_N)^2 \to 0$,

$$\boxed{\text{eff}\left(\widehat{\tau}_{\mathrm{R}}, \widehat{\tau}_{\mathrm{dm}}\right) = 12\sigma^2\left(\int_{\mathbb{R}} f^2(x)dx\right)^2 \ge 0.864 \text{ (worst-case lower bound)}}$$

f	Normal	Cauchy	Student's t_3	Exponential	Pareto(α)
$\text{eff}\left(\widehat{\tau}_{\mathrm{R}}, \widehat{\tau}_{\mathrm{dm}}\right)$	0.95	∞	1.9	3.0	$\ge 3, \nearrow \infty$ as $\alpha \to 2$
$\text{eff}\left(\widehat{\tau}_{\mathrm{R}}, \widehat{\tau}_{\mathrm{med}}\right)$	1.5	0.75	1.2	3.0	$\ge 3, \nearrow 3.8$ as $\alpha \to 2$

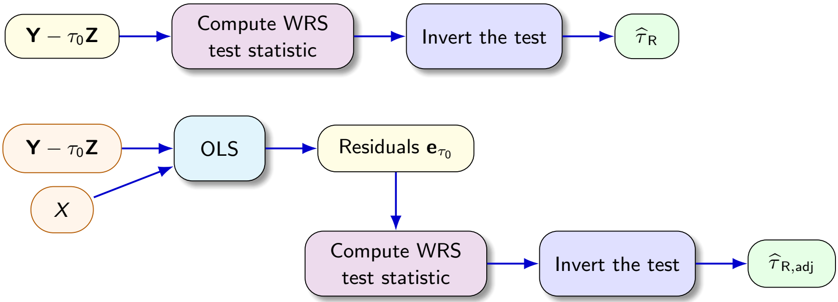
Table 1: Asymptotic efficiency of Rosenbaum's estimator ($\widehat{\tau}_{\mathrm{R}}$) relative to the difference-in-means ($\widehat{\tau}_{\mathrm{dm}}$) and the difference-in-medians ($\widehat{\tau}_{\mathrm{med}}$) estimators.

Regression adjustment

- It is common to collect additional (pre-treatment) covariates $\mathbf{X}_{N\times p}$ (assume non-random, and contains $\mathbf{1}$ (intercept)).

Regression adjustment

- It is common to collect additional (pre-treatment) covariates $\mathbf{X}_{N\times p}$ (assume non-random, and contains $\mathbf{1}$ (intercept)).
- One advantage of Rosenbaum's approach is that it is relatively simple to incorporate these covariates!



Asymptotics of the regression adjusted estimator

- Analyzing the regression adjusted estimator $\widehat{\tau}_{\mathrm{R,adj}}$ naturally presents new challenges compared to $\widehat{\tau}_{\mathrm{R}}$.
  - Dealing with 'global dependence' in the indicators that define the ranks

$$\text{Earlier: } t(\boldsymbol{Z}, \mathbf{Y} - \tau_0 \mathbf{Z}) = \sum_{Z_i=1} \sum_{j=1}^{N} \mathbf{1}\left\{ \boxed{Y_j - \tau_0 Z_j \leq Y_i - \tau_0 Z_i} \right\}$$

$$\text{Now: } t(\boldsymbol{Z}, \boldsymbol{e}_{\tau_0}) = \sum_{Z_i=1} \sum_{j=1}^{N} \mathbf{1}\left\{ \boxed{e_{\tau_0,j} \leq e_{\tau_0,i}} \right\}, \quad \boldsymbol{e}_{\tau_0} = (\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{X}})(\boldsymbol{Y} - \tau_0 \boldsymbol{Z})$$

Asymptotics of the regression adjusted estimator

- Analyzing the regression adjusted estimator $\widehat{\tau}_{\mathsf{R,adj}}$ naturally presents new challenges compared to $\widehat{\tau}_{\mathsf{R}}$.
- Luckily, our proof techniques could be adapted to follow same roadmap: Derive the local asymptotic normality of the test statistic and thus find the asymptotic distribution of $\widehat{\tau}_{\mathsf{R,adj}}$.

Asymptotics of the regression adjusted estimator

- Analyzing the regression adjusted estimator $\widehat{\tau}_{\mathrm{R,adj}}$ naturally presents new challenges compared to $\widehat{\tau}_{\mathrm{R}}$.
- Luckily, our proof techniques could be adapted to follow same roadmap: Derive the local asymptotic normality of the test statistic and thus find the asymptotic distribution of $\widehat{\tau}_{\mathrm{R,adj}}$.
- Similar to the unadjusted case, we derive efficiency of $\widehat{\tau}_{\mathrm{R,adj}}$ relative to Lin's estimator $\widehat{\tau}_{\mathrm{Lin}} \leftarrow \mathrm{lm}(Y \sim Z + X + Z : X)$.

Asymptotics of the regression adjusted estimator

- Analyzing the regression adjusted estimator $\widehat{\tau}_{\mathrm{R,adj}}$ naturally presents new challenges compared to $\widehat{\tau}_{\mathrm{R}}$.
- Luckily, our proof techniques could be adapted to follow same roadmap: Derive the local asymptotic normality of the test statistic and thus find the asymptotic distribution of $\widehat{\tau}_{\mathrm{R,adj}}$.
- Similar to the unadjusted case, we derive efficiency of $\widehat{\tau}_{\mathrm{R,adj}}$ relative to Lin's estimator $\widehat{\tau}_{\mathrm{Lin}} \leftarrow \mathrm{lm}(Y \sim Z + X + Z : X)$.
- We also show that regression adjustment generally improves precision ($\widehat{\tau}_{\mathrm{R,adj}}$ is better than $\widehat{\tau}_{\mathrm{R}}$).

Local asymptotic normality

Define OLS-adjusted control potential outcomes as $\widetilde{b}_{N,i} = b_{N,i} - \mathbf{p}_{N,i}^{\top}\mathbf{b}_N$ where $\mathbf{p}_{N,i}$ is the i-th column of the projection matrix that projects onto range$(\mathbf{X}_N)$.

Define $I_{h,N}(x) = \mathbf{1}\left\{0 \leq x < \frac{h}{\sqrt{N}}\right\}$ for $h > 0$, and $-\mathbf{1}\left\{\frac{h}{\sqrt{N}} \leq x < 0\right\}$ for $h < 0$.

Local asymptotic normality

Define OLS-adjusted control potential outcomes as $\widetilde{b}_{N,i} = b_{N,i} - \mathbf{p}_{N,i}^{\top}\mathbf{b}_N$ where $\mathbf{p}_{N,i}$ is the i-th column of the projection matrix that projects onto range$(\mathbf{X}_N)$.

Define $I_{h,N}(x) = \mathbf{1}\left\{0 \le x < \frac{h}{\sqrt{N}}\right\}$ for $h > 0$, and $-\mathbf{1}\left\{\frac{h}{\sqrt{N}} \le x < 0\right\}$ for $h < 0$.

Assume that the OLS-adjusted control potential outcomes $\{\widetilde{b}_{N,i}\}$ satisfy

$$\lim_{N\to\infty} N^{-3/2} \sum_{i=1}^{N} \sum_{j=1}^{N} I_{h,N}(\widetilde{b}_{N,j} - \widetilde{b}_{N,i}) = h\,\mathcal{J}_b$$

for every $h \in \mathbb{R}$, where $\mathcal{J}_b \in (0,\infty)$ is a fixed constant.

The proportion of the pairwise differences $\widetilde{b}_{N,j} - \widetilde{b}_{N,i}$ falling into small intervals (shrinking at the rate of $N^{-1/2}$) scales with the lengths of those intervals.

Local asymptotic normality

Define OLS-adjusted control potential outcomes as $\widetilde{b}_{N,i} = b_{N,i} - \mathbf{p}_{N,i}^\top \mathbf{b}_N$ where $\mathbf{p}_{N,i}$ is the i-th column of the projection matrix that projects onto range($\mathbf{X}_N$).

Define $I_{h,N}(x) = \mathbf{1}\left\{0 \leq x < \frac{h}{\sqrt{N}}\right\}$ for $h > 0$, and $-\mathbf{1}\left\{\frac{h}{\sqrt{N}} \leq x < 0\right\}$ for $h < 0$.

Assume that the OLS-adjusted control potential outcomes $\{\widetilde{b}_{N,i}\}$ satisfy

$$\lim_{N\to\infty} N^{-3/2} \sum_{i=1}^{N} \sum_{j=1}^{N} I_{h,N}(\widetilde{b}_{N,j} - \widetilde{b}_{N,i}) = h\,\mathcal{J}_b$$

for every $h \in \mathbb{R}$, where $\mathcal{J}_b \in (0, \infty)$ is a fixed constant.

Under $\tau_N = \tau_0 - hN^{-1/2}$, the regression-adjusted test statistic $t_{N,\mathrm{adj}}$ satisfies

$$N^{-3/2}\left(t_{N,\mathrm{adj}} - \mathbb{E}_{\tau_0}[t_N]\right) \xrightarrow{d} \mathcal{N}\left(-h\lambda(1-\lambda)\mathcal{J}_b, \frac{\lambda(1-\lambda)}{12}\right).$$

Assume that

(A1) Treatment is randomly assigned to m units with $m/N \to \lambda \in (0,1)$.

(A2) Constant additive treatment effect: $a_i - b_i = \tau$ for each $1 \le i \le N$.

(A4) The OLS-adjusted control potential outcomes $\{\widetilde{b}_{N,i}\}_{1 \le i \le N}$ satisfy

$$\lim_{N\to\infty} N^{-3/2} \sum_{i=1}^{N} \sum_{j=1}^{N} I_{h,N}(\widetilde{b}_{N,j} - \widetilde{b}_{N,i}) = h\,\mathcal{J}_b$$

for every $h \in \mathbb{R}$, where $\mathcal{J}_b \in (0,\infty)$ is a fixed constant.

Then,

$$\sqrt{N}\,(\widehat{\tau}_{\mathrm{R,adj}} - \tau) \xrightarrow{d} \mathcal{N}\left(0, \left(12\lambda(1-\lambda)\mathcal{J}_b^2\right)^{-1}\right).$$

A more interpretable version

Replace the regularity assumption (A4) with the following.

(A4') The empirical distribution G_N of the OLS-adjusted control P.O. $\{\widetilde{b}_{N,i}\}$ converges weakly to a distribution G with a square-integrable density g, and satisfies a smoothness condition at the $N^{-1/2}$ scale[5].

The residuals $\widetilde{b}_{N,i}$ behave, in the limit, as i.i.d. from a density g and

$$\mathcal{J}_b = \int_{\mathbb{R}} g^2(x)\, dx.$$

[5] For every $h \in \mathbb{R}$, $\sup_{x\in\mathbb{R}} |\sqrt{N}(G_N(x + h/\sqrt{N}) - G_N(x)) - h\, g(x)| = o(1)$ as $N \to \infty$.

A more interpretable version

Replace the regularity assumption (A4) with the following.

(A4') The empirical distribution G_N of the OLS-adjusted control P.O. $\{\widetilde{b}_{N,i}\}$ converges weakly to a distribution G with a square-integrable density g, and satisfies a smoothness condition at the $N^{-1/2}$ scale[5].

- The residuals $\widetilde{b}_{N,i}$ behave, in the limit, as i.i.d. from a density g and

$$\mathcal{J}_b = \int_{\mathbb{R}} g^2(x)\, dx.$$

- Indeed, if $\mathbf{Y}_N = \tau \mathbf{Z}_N + \mathbf{X}_N \boldsymbol{\beta}_N + \boldsymbol{\varepsilon}_N$ where $\varepsilon_{N,i}$ are i.i.d. mean zero with square-integrable density g, then the above holds.

[5] For every $h \in \mathbb{R}$, $\sup_{x\in\mathbb{R}} |\sqrt{N}(G_N(x + h/\sqrt{N}) - G_N(x)) - h\, g(x)| = o(1)$ as $N \to \infty$.

Efficiency comparison

Under (A1), (A2), (A4'),

$$\sqrt{N}\left(\widehat{\tau}_{\mathrm{R,adj}} - \tau\right) \xrightarrow{d} \mathcal{N}\left(0, \left(12\lambda(1-\lambda)\left(\int_{\mathbb{R}} g^2(x)\,dx\right)^2\right)^{-1}\right).$$

Under additional moment assumptions as in Lin (2013),

$$\mathrm{eff}\left(\widehat{\tau}_{\mathrm{R,adj}}, \widehat{\tau}_{\mathrm{Lin}}\right) = 12\sigma^2\left(\int_{\mathbb{R}} g^2(x)dx\right)^2 \geq 0.864 \text{ (worst-case lower bound)}$$

g	Normal	Cauchy	Student's t_3	Exponential	Pareto(α)
$\mathrm{eff}\left(\widehat{\tau}_{\mathrm{R,adj}}, \widehat{\tau}_{\mathrm{Lin}}\right)$	0.95	∞	1.9	3.0	$\nearrow \infty$ as $\alpha \to 2$

Table 2: Asymptotic efficiency of Rosenbaum's $\widehat{\tau}_{\mathrm{R,adj}}$ relative to Lin's estimator ($\widehat{\tau}_{\mathrm{Lin}}$).

Efficiency gain by regression adjustment

We impose an asymptotic independence-like condition on the residuals $\widetilde{b}_{N,i}$ and predictions $(b_{N,i} - \widetilde{b}_{N,i})$

$$\sup_{x,y} \left| \frac{1}{N} \sum_{i=1}^{N} \mathbf{1}\left\{ \widetilde{b}_{N,i} \leq x, b_{N,i} - \widetilde{b}_{N,i} \leq y \right\} - G_N(x) \frac{1}{N} \sum_{i=1}^{N} \mathbf{1}\left\{ b_{N,i} - \widetilde{b}_{N,i} \leq y \right\} \right| = o(1).$$

✔ This holds, in particular, when $\mathbf{Y}_N = \tau \mathbf{Z}_N + \mathbf{X}_N \boldsymbol{\beta}_N + \boldsymbol{\varepsilon}_N$ where $\varepsilon_{N,i}$ are i.i.d. mean zero with square-integrable density g, independent of $\mathbf{Z}_N$.

Efficiency gain by regression adjustment

We impose an asymptotic independence-like condition on the residuals $\widetilde{b}_{N,i}$ and predictions $(b_{N,i} - \widetilde{b}_{N,i})$

Rosenbaum's regression-adjusted estimator is *at least as efficient as* its unadjusted counterpart:

$$\mathrm{eff}\left(\widehat{\tau}_{\mathrm{R,adj}}, \widehat{\tau}_{\mathrm{R}}\right) \geq 1.$$

- Our assumptions allow heavy-tailed noise as well as heavy-tailed covariates.

Interestingly, the proof uses Parseval–Plancherel theorem from Fourier analysis!

Heterogeneous treatment effects

What happens to Rosenbaum's estimator $\widehat{\tau}_{\mathrm{R}}$ if the constant treatment effect assumption does not hold?

Heterogeneous treatment effects

What happens to Rosenbaum's estimator $\widehat{\tau}_{\mathrm{R}}$ if the constant treatment effect assumption does not hold?

- We can define Rosenbaum's estimator in a way that does not assume constant treatment effect.
- Robustness to contamination (the breakdown point result) still holds.

Heterogeneous treatment effects

What happens to Rosenbaum's estimator $\widehat{\tau}_{\mathrm{R}}$ if the constant treatment effect assumption does not hold?

- ✓ We can define Rosenbaum's estimator in a way that does not assume constant treatment effect.
- ✓ Robustness to contamination (the breakdown point result) still holds.
- ⚠ The results on asymptotic distribution of the estimator no longer hold.

Which estimand does Rosenbaum's estimator target when the constant treatment effect assumption does not hold?

Heterogeneous treatment effects (contd.)

- Under constant treatment effect, Rosenbaum's estimator can be written as

$$\widehat{\tau}_{\mathrm{R}} := \operatorname{median}\{Y_i - Y_j : Z_i = 1, Z_j = 0\}.$$

[6] Lei, L. (2024). Causal interpretation of regressions with ranks. arXiv preprint 2406.05548.

Heterogeneous treatment effects (contd.)

- Under constant treatment effect, Rosenbaum's estimator can be written as

$$\widehat{\tau}_{\mathrm{R}} := \operatorname{median}\{Y_i - Y_j : Z_i = 1, Z_j = 0\}.$$

- Taking the above as definition, we show that $\widehat{\tau}_{\mathrm{R}} - \mathrm{med}_N \to_p 0$, where

$$\mathrm{med}_N = \operatorname{median}\{a_i - b_j : 1 \le i \ne j \le N\}.$$

[6] Lei, L. (2024). Causal interpretation of regressions with ranks. arXiv preprint 2406.05548.

Heterogeneous treatment effects (contd.)

- Under constant treatment effect, Rosenbaum's estimator can be written as

$$\widehat{\tau}_{\mathrm{R}} := \text{median}\left\{ Y_i - Y_j \,:\, Z_i = 1, Z_j = 0 \right\}.$$

- Taking the above as definition, we show that $\widehat{\tau}_{\mathrm{R}} - \text{med}_N \to_p 0$, where

$$\text{med}_N = \text{median}\left\{ a_i - b_j \,:\, 1 \leq i \neq j \leq N \right\}.$$

- Being the median of pairwise differences between treatment and control outcomes, the causal estimand med_N is more robust than the ATE.

[6] Lei, L. (2024). Causal interpretation of regressions with ranks. arXiv preprint 2406.05548.

Heterogeneous treatment effects (contd.)

- Under constant treatment effect, Rosenbaum's estimator can be written as

$$\widehat{\tau}_{\mathrm{R}} := \operatorname{median}\{\, Y_i - Y_j \,:\, Z_i = 1, Z_j = 0 \}\,.$$

- Taking the above as definition, we show that $\widehat{\tau}_{\mathrm{R}} - \mathrm{med}_N \to_p 0$, where

$$\mathrm{med}_N = \operatorname{median}\{a_i - b_j \,:\, 1 \leq i \neq j \leq N\}\,.$$

- Being the median of pairwise differences between treatment and control outcomes, the causal estimand med_N is more robust than the ATE.
- Lei (2024)[6] studies the same estimand under i.i.d. framework.

[6] Lei, L. (2024). Causal interpretation of regressions with ranks. arXiv preprint 2406.05548.

Heterogeneous treatment effects (contd.)

- Under constant treatment effect, Rosenbaum's estimator can be written as

$$\widehat{\tau}_{\mathrm{R}} := \operatorname{median}\{Y_i - Y_j : Z_i = 1, Z_j = 0\}.$$

- Taking the above as definition, we show that $\widehat{\tau}_{\mathrm{R}} - \mathrm{med}_N \to_p 0$, where

$$\mathrm{med}_N = \operatorname{median}\{a_i - b_j : 1 \leq i \neq j \leq N\}.$$

- Being the median of pairwise differences between treatment and control outcomes, the causal estimand med_N is more robust than the ATE.
- Lei (2024)[6] studies the same estimand under i.i.d. framework.
- While we show consistency, finite-sample limiting distribution is left for future work.

[6] Lei, L. (2024). Causal interpretation of regressions with ranks. arXiv preprint 2406.05548.

Summary

- We derive finite-population asymptotics of Rosenbaum's rank-based estimator in randomized trials, with or without covariate adjustment.

Summary

- We derive finite-population asymptotics of Rosenbaum's rank-based estimator in randomized trials, with or without covariate adjustment.
- Our results highlight important trade-offs between robustness & efficiency.

Summary

- We derive finite-population asymptotics of Rosenbaum's rank-based estimator in randomized trials, with or without covariate adjustment.
- Our results highlight important trade-offs between robustness & efficiency.
  - Rosenbaum's estimator has a breakdown point at least as high as any weighted quantile estimator.

Summary

- We derive finite-population asymptotics of Rosenbaum's rank-based estimator in randomized trials, with or without covariate adjustment.
- Our results highlight important trade-offs between robustness & efficiency.
  - Rosenbaum's estimator has a breakdown point at least as high as any weighted quantile estimator.
  - We formally establish efficiency gain over $\widehat{\tau}_{\mathrm{dm}}$ or $\widehat{\tau}_{\mathrm{Lin}}$ when outcomes have heavy-tails or contain outliers.

Summary

- We derive finite-population asymptotics of Rosenbaum's rank-based estimator in randomized trials, with or without covariate adjustment.
- Our results highlight important trade-offs between robustness & efficiency.
  - Rosenbaum's estimator has a breakdown point at least as high as any weighted quantile estimator.
  - We formally establish efficiency gain over $\widehat{\tau}_{\mathsf{dm}}$ or $\widehat{\tau}_{\mathsf{Lin}}$ when outcomes have heavy-tails or contain outliers.
  - Also show efficiency gain from regression-adjustment for Rosenbaum's estimator under suitable conditions.

Summary

- We derive finite-population asymptotics of Rosenbaum's rank-based estimator in randomized trials, with or without covariate adjustment.
- Our results highlight important trade-offs between robustness & efficiency.
  - Rosenbaum's estimator has a breakdown point at least as high as any weighted quantile estimator.
  - We formally establish efficiency gain over $\widehat{\tau}_{\mathrm{dm}}$ or $\widehat{\tau}_{\mathrm{Lin}}$ when outcomes have heavy-tails or contain outliers.
  - Also show efficiency gain from regression-adjustment for Rosenbaum's estimator under suitable conditions.
- Our proof techniques can be of independent interest for analyzing other rank-based estimators under the design-based framework.

Paper: arXiv:2111.15524 (*Biometrika*, forthcoming)

Backup slides

Implication of breakdown point for confidence intervals

One practical implication of our notion of breakdown point is that, any confidence interval based on an estimator with zero ABP must either

- fail to maintain uniform validity under small contamination, or
- have infinite expected length, even under arbitrarily small contamination rates.

In contrast, any estimator with strictly positive ABP can be used to construct CIs with bounded expected length that maintain uniform validity under contamination, provided the contamination level is below the ABP.

Table 3: Empirical coverage and average length of 95% CIs for simulation Setting 1: Generate x_i i.i.d. from Unif$(-4, 4)$, and set $a_i = 3x_i + \varepsilon_i$ and $b_i = a_i - 2$. Noise ϵ_i's are i.i.d. from: (a) Gaussian, (b) Cauchy, and (c) Student's t_3.

Estimator	(a) Gaussian errors		(b) Cauchy errors		(c) t_3 errors	
	coverage	length	coverage	length	coverage	length
Difference-in-Means ($\widehat{\tau}_{\mathrm{dm}}$)	94.8%	1.75	94.2%	24.26	94.8%	1.79
Difference-in-Medians ($\widehat{\tau}_{\mathrm{med}}$)	93.1%	2.90	94.9%	3.07	94.7%	2.91
0.1-trimmed Diff-in-Means	94.1%	2.03	95.2%	2.19	95.2%	2.04
0.1-Winsorized Diff-in-Means	94.2%	1.82	95.6%	2.01	94.9%	1.84
$\widehat{\tau}_{\mathrm{eif}}$ (Athey et al., 2023)	96.3%	1.30	96.2%	1.74	96.7%	1.39
$\widehat{\tau}_{\mathrm{waq}}$ (Athey et al., 2023)	95.7%	1.32	94.5%	2.50	96.2%	1.44
Rosenbaum's estimator ($\widehat{\tau}_{\mathrm{R}}$)	94.9%	1.84	95.6%	2.14	95.4%	1.88
OLS adjusted ($\widehat{\tau}_{\mathrm{adj}}$)	99.0%	0.35	94.2%	24.30	96.5%	0.49
Lin's estimator ($\widehat{\tau}_{\mathrm{Lin}}$)	99.0%	0.35	94.2%	24.30	96.5%	0.49
Rosenbaum's adjusted ($\widehat{\tau}_{\mathrm{R,adj}}$)	99.4%	0.37	97.2%	1.34	98.7%	0.43

Simulation results (contd.)

Table 4: Empirical coverage and average length of 95% CIs for simulation Setting 2: Same as Setting 1, except here we contaminate 5% of the potential outcomes with an arbitrary large value $M = 500$.

Estimator	(a) Gaussian errors		(b) Cauchy errors		(c) t_3 errors	
	coverage	length	coverage	length	coverage	length
Difference-in-Means ($\widehat{\tau}_{\mathrm{dm}}$)	100.0%	27.52	100.0%	43.71	100.0%	27.51
Difference-in-Medians ($\widehat{\tau}_{\mathrm{med}}$)	94.5%	3.04	94.7%	3.22	94.9%	3.07
0.1-trimmed Diff-in-Means	95.4%	2.15	96.7%	2.35	96.9%	2.16
0.1-Winsorized Diff-in-Means	96.1%	1.95	97.9%	2.32	97.4%	1.98
$\widehat{\tau}_{\mathrm{eif}}$ (Athey et al., 2023)	92.9%	1.63	90.2%	9.16	92.1%	1.68
$\widehat{\tau}_{\mathrm{waq}}$ (Athey et al., 2023)	100.0%	59.17	94.1%	659.9	100.0%	56.97
Rosenbaum's estimator ($\widehat{\tau}_{\mathrm{R}}$)	95.5%	1.97	96.4%	2.29	96.0%	2.01
OLS adjusted ($\widehat{\tau}_{\mathrm{adj}}$)	100.0%	27.77	99.9%	43.85	100.0%	27.75
Lin's estimator ($\widehat{\tau}_{\mathrm{Lin}}$)	100.0%	27.77	99.9%	43.85	100.0%	27.75
Rosenbaum's adjusted ($\widehat{\tau}_{\mathrm{R,adj}}$)	97.3%	0.95	97.4%	1.88	97.8%	1.01

Simulation results (contd.)

Table 5: Empirical coverage and average length of 95% CIs for simulation Setting 3 (model misspecification): Generate u_i i.i.d. from Unif$(-4, 4)$, and set $x_i = e^{u_i}$, $a_i = \frac{1}{4}(x_i + \sqrt{x_i}) + \varepsilon_i$ and $b_i = a_i - 2$.

Estimator	(a) Gaussian errors		(b) Cauchy errors		(c) t_3 errors	
	coverage	length	coverage	length	coverage	length
Difference-in-Means ($\widehat{\tau}_{\mathrm{dm}}$)	96.3%	0.91	94.2%	24.05	95.9%	0.98
Difference-in-Medians ($\widehat{\tau}_{\mathrm{med}}$)	98.0%	0.68	96.1%	0.84	96.5%	0.71
0.1-trimmed Diff-in-Means	95.6%	0.83	95.7%	1.14	96.4%	0.88
0.1-Winsorized Diff-in-Means	95.1%	1.00	95.3%	1.33	95.8%	1.02
$\widehat{\tau}_{\mathrm{eif}}$ (Athey et al., 2023)	99.6%	0.52	98.1%	0.76	98.8%	0.60
$\widehat{\tau}_{\mathrm{waq}}$ (Athey et al., 2023)	99.4%	0.52	96.1%	0.86	97.8%	0.61
Rosenbaum's estimator ($\widehat{\tau}_{\mathrm{R}}$)	98.6%	0.62	97.8%	0.91	98.5%	0.69
OLS adjusted ($\widehat{\tau}_{\mathrm{adj}}$)	99.2%	0.35	94.7%	24.24	96.6%	0.49
Lin's estimator ($\widehat{\tau}_{\mathrm{Lin}}$)	99.2%	0.35	94.7%	24.23	96.6%	0.49
Rosenbaum's adjusted ($\widehat{\tau}_{\mathrm{R,adj}}$)	99.2%	0.37	97.7%	0.74	98.5%	0.43

Data analysis

Table 6: Different estimates of the effect of early Progresa on PRI support rates with the corresponding standard errors, 95% permutation CIs and their lengths.

Estimator	Estimate	Std. Error	95% CI	CI Length
Difference-in-Means ($\widehat{\tau}_{\mathrm{dm}}$)	3.62	1.92	[−0.14, 7.39]	7.53
Difference-in-Medians ($\widehat{\tau}_{\mathrm{med}}$)	0.69	1.56	[−2.37, 3.75]	6.12
0.1-trimmed Difference-in-Means	2.00	1.68	[−1.29, 5.28]	6.57
0.1-Winsorized Difference-in-Means	2.59	1.72	[−0.78, 5.96]	6.74
$\widehat{\tau}_{\mathrm{eif}}$ (Athey et al., 2023)	1.95	1.72	[−1.42, 5.33]	6.75
$\widehat{\tau}_{\mathrm{waq}}$ (Athey et al., 2023)	1.31	1.71	[−2.04, 4.65]	6.69
Rosenbaum's estimator ($\widehat{\tau}_{\mathrm{R}}$)	1.83	1.67	[−1.43, 5.10]	6.53
OLS adjusted $\widehat{\tau}_{\mathrm{adj}}$	3.67	1.70	[0.34, 7.00]	6.67
$\widehat{\tau}_{\mathrm{Lin}}$ (Lin, 2013)	4.21	1.99	[0.32, 8.11]	7.78
Rosenbaum's adjusted ($\widehat{\tau}_{\mathrm{R,adj}}$)	2.19	1.38	[−0.53, 4.90]	5.43