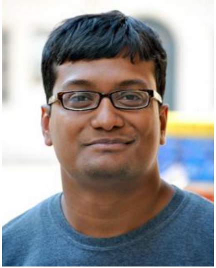
Efficiency and Robustness of Rosenbaum's Rank-based Estimator in Randomized Experiments

Nabarun Deb

University of Chicago Booth School of Business

Statistical learning in causal inference

Joint work with Aditya Ghosh, Bikram Karmakar, Bodhisattva Sen

- Randomized trials – ‘randomization inference’ (Fisher, ’35)
- Mean based estimators (e.g., difference-in-means) yield very wide C.I.’s when potential outcomes are heavy-tailed or have outliers

- Randomized trials – ‘randomization inference’ (Fisher, ’35)
- Mean based estimators (e.g., difference-in-means) yield very wide C.I.’s when potential outcomes are heavy-tailed or have outliers
- Rank-based estimators are generally less sensitive to heavy-tails or extreme observations

- Randomized trials – ‘randomization inference’ (Fisher, ’35)
- Mean based estimators (e.g., difference-in-means) yield very wide C.I.’s when potential outcomes are heavy-tailed or have outliers
- Rank-based estimators are generally less sensitive to heavy-tails or extreme observations
- Rosenbaum (’93) proposed a Hodges-Lehmann type estimator
- Theoretical study of this estimator was missing in the literature.
  Numerical methods do not shed light on efficiency or robustness

Our contributions

A systematic and rigorous study of the asymptotic properties of Rosenbaum's estimator.

Our contributions

A systematic and rigorous study of the asymptotic properties of Rosenbaum's estimator. **Highlights:**

- Robustness against extreme observations
  $\rightarrow$ notion of breakdown point

Our contributions

A systematic and rigorous study of the asymptotic properties of Rosenbaum's estimator. **Highlights:**

- Robustness against extreme observations
  $\rightarrow$ notion of breakdown point
- Establishing its asymptotic distribution
  $\rightarrow$ weaker assumptions, also valid for heavy-tails

Our contributions

A systematic and rigorous study of the asymptotic properties of Rosenbaum's estimator. **Highlights:**

- Robustness against extreme observations
  $\rightarrow$ notion of breakdown point
- Establishing its asymptotic distribution
  $\rightarrow$ weaker assumptions, also valid for heavy-tails
- Consistent estimation of the asymptotic variance
  $\rightarrow$ asymptotically valid C.I.'s in analytic form

Our contributions

A systematic and rigorous study of the asymptotic properties of Rosenbaum's estimator. **Highlights:**

- Robustness against extreme observations
  $\rightarrow$ notion of breakdown point
- Establishing its asymptotic distribution
  $\rightarrow$ weaker assumptions, also valid for heavy-tails
- Consistent estimation of the asymptotic variance
  $\rightarrow$ asymptotically valid C.I.'s in analytic form
- Efficiency relative to the difference-in-means estimator
  $\rightarrow$ an efficiency lower bound

Our contributions

A systematic and rigorous study of the asymptotic properties of Rosenbaum's estimator. **Highlights:**

- Robustness against extreme observations
  $\rightarrow$ notion of breakdown point
- Establishing its asymptotic distribution
  $\rightarrow$ weaker assumptions, also valid for heavy-tails
- Consistent estimation of the asymptotic variance
  $\rightarrow$ asymptotically valid C.I.'s in analytic form
- Efficiency relative to the difference-in-means estimator
  $\rightarrow$ an efficiency lower bound
- Also study OLS adjusted version of Rosenbaum's estimator
  $\rightarrow$ efficiency gain by regression adjustment

The framework and estimation strategy

The framework

- **Randomized trial:** treatment group comprises m subjects chosen uniformly at random from N subjects. Assume $m/N \to \lambda \in (0, 1)$.

The framework

- **Randomized trial:** treatment group comprises m subjects chosen uniformly at random from N subjects. Assume $m/N \to \lambda \in (0, 1)$.
- Z_i = treatment indicator, Y_i = response for i-th subject, given by

$$Y_i = Z_i a_i + (1 - Z_i) b_i,$$

where a_i and b_i are the potential outcomes for the treated and control, respectively. (Neyman, '23; Rubin, '74, '77)

The framework

- **Randomized trial:** treatment group comprises m subjects chosen uniformly at random from N subjects. Assume $m/N \to \lambda \in (0,1)$.
- Z_i = treatment indicator, Y_i = response for i-th subject, given by

$$Y_i = Z_i a_i + (1 - Z_i) b_i,$$

where a_i and b_i are the potential outcomes for the treated and control, respectively. (Neyman, '23; Rubin, '74, '77)

- Constant additive treatment effect model: (Rosenbaum, '93, '02)

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

(analog of location shift model in classical nonparametrics. Non-constant treatment effect case at the end)

The framework

- **Randomized trial:** treatment group comprises m subjects chosen uniformly at random from N subjects. Assume $m/N \to \lambda \in (0, 1)$.
- Z_i = treatment indicator, Y_i = response for i-th subject, given by

$$Y_i = Z_i a_i + (1 - Z_i) b_i,$$

  where a_i and b_i are the potential outcomes for the treated and control, respectively. (Neyman, '23; Rubin, '74, '77)
- Constant additive treatment effect model: (Rosenbaum, '93, '02)

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

- **Randomization inference:** a_i and b_i are fixed, only Z_i is random

Rosenbaum's estimator

- Consider testing $H_0 : \tau = \tau_0 \quad \text{versus} \quad H_1 : \tau \neq \tau_0$.
- Under H_0, $\boldsymbol{Y} - \tau_0 \boldsymbol{Z} = \boldsymbol{b}$, which is non-random $\rightarrow$ we can use any statistic of the form $t(\boldsymbol{Z}, \boldsymbol{Y} - \tau_0 \boldsymbol{Z})$ to draw randomization inference

Rosenbaum's estimator

- Consider testing $H_0 : \tau = \tau_0 \quad \text{versus} \quad H_1 : \tau \neq \tau_0$.
- Under H_0, $\boldsymbol{Y} - \tau_0 \boldsymbol{Z} = \boldsymbol{b}$, which is non-random $\rightarrow$ we can use any statistic of the form $t(\boldsymbol{Z}, \boldsymbol{Y} - \tau_0 \boldsymbol{Z})$ to draw randomization inference
- For a point estimate of τ, Rosenbaum ('02) suggested to invert the above test. Following Hodges and Lehmann ('63), We set

$$\widehat{\tau}^{*} := \sup \{\tau : t(\boldsymbol{Z}, \boldsymbol{Y} - \tau \boldsymbol{Z}) > \mu\}, \ \widehat{\tau}^{**} := \inf \{\tau : t(\boldsymbol{Z}, \boldsymbol{Y} - \tau \boldsymbol{Z}) < \mu\},$$

where $\mu := \mathbb{E}_{\tau_0} t(\boldsymbol{Z}, \boldsymbol{Y} - \tau_0 \boldsymbol{Z})$, and define

$$\widehat{\tau}^{\mathrm{R}} := \frac{\widehat{\tau}^{*} + \widehat{\tau}^{**}}{2}.$$

Note, $\widehat{\tau}^{\mathrm{R}}$ depends on our choice of the test statistic $t(\cdot, \cdot)$.

The Wilcoxon rank-sum statistic

- The Wilcoxon rank-sum (WRS) statistic is defined as

$$t(\boldsymbol{Z}, \boldsymbol{Y} - \tau \boldsymbol{Z}) := \sum_{j: Z_j = 1} q_j^{(\tau)} = \sum_{j=1}^{N} Z_j q_j^{(\tau)},$$

  where

$$q_j^{(\tau)} := \sum_{i=1}^{N} \mathbf{1}_{\{Y_i - \tau Z_i \leq Y_j - \tau Z_j\}}, \ 1 \leq j \leq N.$$

- For rest of the talk, $\widehat{\tau}^{\mathrm{R}}$ will denote Rosenbaum's estimator based on the WRS statistic.

Main theoretical results

Robustness (Breakdown point)

- Indices corresp. to the treatment and control groups are chosen by randomization $\rightarrow$ need an appropriate definition of Breakdown point

Robustness (Breakdown point)

- Indices corresp. to the treatment and control groups are chosen by randomization $\rightarrow$ need an appropriate definition of Breakdown point
- What is the minimum proportion of potential outcomes that, if replaced with arbitrarily extreme values, will cause the estimator to be arbitrarily large, for all configurations of treatment assignments?

Robustness (Breakdown point)

- Indices corresp. to the treatment and control groups are chosen by randomization $\rightarrow$ need an appropriate definition of Breakdown point
- What is the minimum proportion of potential outcomes that, if replaced with arbitrarily extreme values, will cause the estimator to be arbitrarily large, for all configurations of treatment assignments?

Proposition 1 (Asymptotic breakdown point of $\widehat{\tau}^{\mathrm{R}}$)

Let $\widehat{\tau}^{\mathrm{R}}$ be Rosenbaum's estimator based the Wilcoxon rank-sum statistic.

$$\mathrm{ABP}(\widehat{\tau}^{\mathrm{R}}) = \begin{cases} (1-\lambda)/2 & \textit{if } \lambda < 1/3 \\ 1 - \sqrt{2\lambda(1-\lambda)} & \textit{if } 1/3 \leq \lambda \leq 2/3 \\ \lambda/2 & \textit{if } \lambda > 2/3, \end{cases}$$

where λ is the limiting ratio of the treated group to the total sample size.

$\mathrm{ABP}(\widehat{\tau}^{\mathrm{R}}) \geq 1 - 1/\sqrt{2} \approx 0.29$ always; approaches $1/2$ as $\lambda \rightarrow 0$ or 1.

Notations

- We add a subscript N for the subsequent asymptotic results.
- Potential outcomes $\{(a_{N,i}, b_{N,i}) : 1 \le i \le N\}$ are fixed, and $a_{N,i} - b_{N,i} = \tau$ for each i.
- $\boldsymbol{Z}_N$: treatment indicators, $\boldsymbol{Y}_N$: observed responses.
- Wilcoxon Rank-Sum (WRS) statistic for testing $\tau = \tau_0$ is given by

$$t_N \equiv t_N(\boldsymbol{Z}_N, \boldsymbol{Y}_N - \tau_0 \boldsymbol{Z}_N) := \boldsymbol{Z}_N^\top \boldsymbol{q}_N,$$

where

$$q_{N,\,j} = \sum_{i=1}^{N} \mathbf{1}_{\{Y_{N,i} - \tau_0 Z_{N,i} \le Y_{N,j} - \tau_0 Z_{N,j}\}},\ 1 \le j \le N.$$

Asymptotic null distribution of the WRS statistic

Proposition 2 (Asymptotic null distribution of t_N)

Assume that $m/N \to \lambda \in (0,1)$, *and that the ranks* $\{q_{N,j}\}$ *satisfy*

$$\frac{1}{N}\sum_{j=1}^{N}\left(q_{N,j} - \overline{q}_N\right)^2 = \frac{N^2-1}{12} + o(N^2),$$

where $\overline{q}_N := N^{-1}\sum_{j=1}^{N} q_{N,j}$. *Then, under* $\tau = \tau_0$,

$$N^{-3/2}\left(t_N - m\overline{q}_N\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{\lambda(1-\lambda)}{12}\right).$$

- Identical to the result under the infinite population setup.
- Justifies Rosenbaum ('02)'s numerical method to find C.I.'s.
- No moment assumption, and ties are allowed

Strategy: We show that if under $\tau = \tau_N := \tau_0 - hN^{-1/2}$,

$$N^{-3/2}\,(t_N - \mu_N) \xrightarrow{d} \mathcal{N}(-hB, A^2),$$

for every fixed $h \in \mathbb{R}$, where $\mu_N := \mathbb{E}_{\tau_0} t_N$ and $A, B > 0$, then

$$\sqrt{N}\left(\widehat{\tau}^{\mathrm{R}} - \tau_0\right) \xrightarrow{d} \mathcal{N}(0, (A/B)^2).$$

Strategy: We show that if under $\tau = \tau_N := \tau_0 - hN^{-1/2}$,

$$N^{-3/2}\left(t_N - \mu_N\right) \xrightarrow{d} \mathcal{N}(-hB, A^2),$$

for every fixed $h \in \mathbb{R}$, where $\mu_N := \mathbb{E}_{\tau_0} t_N$ and $A, B > 0$, then

$$\sqrt{N}\left(\widehat{\tau}^{\mathrm{R}} - \tau_0\right) \xrightarrow{d} \mathcal{N}(0, (A/B)^2).$$

Thus, it suffices to find the asymptotic distribution of the WRS statistic t_N under the local alternatives $\tau = \tau_N$.

Asymptotic distribution of t_N under $\tau = \tau_N$

Challenges:

- Under $\tau = \tau_N \equiv \tau_0 - hN^{-1/2}$,

$$t_N \overset{d}{=} m + \sum_{j=1}^{N} Z_{N,j} \sum_{i=1, i\neq j}^{N} \mathbf{1}_{\{b_{N,i} - hN^{-1/2} Z_{N,i} \leq b_{N,j} - hN^{-1/2} Z_{N,j}\}}.$$

  linear combination of $Z_{N,i}$'s with random weights depending on $Z_{N,i}$'s themselves. So, classical combinatorial CLTs do not apply.
- Ranks are highly non-linear and not deterministic under $\tau = \tau_N$, so methods similar to Li and Ding ('17) do not apply.
- Le Cam's method using contiguity is also not applicable.

Asymptotic distribution of t_N under $\tau = \tau_N$ (Contd.)

Define

$$I_{h,N}(x) := \begin{cases} \mathbf{1}_{\{0 \le x < hN^{-1/2}\}} & \text{if } h \ge 0, \\ -\mathbf{1}_{\{hN^{-1/2} \le x < 0\}} & \text{if } h < 0. \end{cases} \quad (1)$$

Define

$$I_{h,N}(x) := \begin{cases} \mathbf{1}_{\{0 \le x < hN^{-1/2}\}} & \text{if } h \ge 0, \\ -\mathbf{1}_{\{hN^{-1/2} \le x < 0\}} & \text{if } h < 0. \end{cases} \quad (1)$$

Assumption 1

Assume that, for $I_{h,N}$ as in (1), the following holds:

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

for some $\mathcal{I}_b \in (0, \infty)$, where $b_{N,i}$'s are the potential control outcomes.

e.g., if $b_{N,i}$'s are realizations from a density f_b with $\int_{\mathbb{R}} f_b^2(x)dx < \infty$, then the above holds a.s. with $\mathcal{I}_b = \int_{\mathbb{R}} f_b^2(x)dx$.

Local asymptotic normality of t_N and asymptotic distribution of $\widehat{\tau}^{\mathrm{R}}$

Theorem 1 (Local asymptotic normality of t_N)

Let t_N be the WRS statistic. Suppose that Assumption 1 holds. Fix $h \in \mathbb{R}$ and let $\tau_N = \tau_0 - hN^{-1/2}$. Then, under $\tau = \tau_N$,

$$N^{-3/2}\left(t_N - \frac{m(N+1)}{2}\right) \xrightarrow{d} \mathcal{N}\left(-h\lambda(1-\lambda)\mathcal{I}_b, \frac{\lambda(1-\lambda)}{12}\right).$$

Local asymptotic normality of t_N and asymptotic distribution of $\widehat{\tau}^{\mathrm{R}}$

Theorem 1 (Local asymptotic normality of t_N)

Let t_N be the WRS statistic. Suppose that Assumption 1 holds. Fix $h \in \mathbb{R}$ and let $\tau_N = \tau_0 - hN^{-1/2}$. Then, under $\tau = \tau_N$,

$$N^{-3/2}\left(t_N - \frac{m(N+1)}{2}\right) \xrightarrow{d} \mathcal{N}\left(-h\lambda(1-\lambda)\mathcal{I}_b, \frac{\lambda(1-\lambda)}{12}\right).$$

Theorem 2 (CLT for the estimator $\widehat{\tau}^{\mathrm{R}}$)

Under Assumption 1, it holds that

$$\sqrt{N}\left(\widehat{\tau}^{\mathrm{R}} - \tau_0\right) \xrightarrow{d} \mathcal{N}\left(0, (12\lambda(1-\lambda)\mathcal{I}_b^2)^{-1}\right).$$

$\mathcal{I}_b$ is defined in Assumption 1.

Confidence interval for τ

Theorem 3 (Consistent estimation of $\mathcal{I}_b$)

Let Assumption 1 hold and $m/N \to \lambda \in (0,1)$. Then as $N \to \infty$,

$$\widehat{\mathcal{I}}_N := \left(1 - \frac{m}{N}\right)^{-2} N^{-3/2} \sum_{i \neq j,\ Z_{N,i} = Z_{N,\,j} = 0} \mathbf{1}_{\left\{0 \leq Y_{N,j} - Y_{N,i} < N^{-1/2}\right\}} \xrightarrow{P} \mathcal{I}_b. \tag{2}$$

Theorem 3 (Consistent estimation of $\mathcal{I}_b$)

Let Assumption 1 hold and $m/N \to \lambda \in (0,1)$. Then as $N \to \infty$,

$$\widehat{\mathcal{I}}_N := \left(1 - \frac{m}{N}\right)^{-2} N^{-3/2} \sum_{i \neq j,\ Z_{N,i} = Z_{N,\ j} = 0} \mathbf{1}_{\{0 \leq Y_{N,j} - Y_{N,i} < N^{-1/2}\}} \xrightarrow{P} \mathcal{I}_b. \quad (2)$$

Corollary 4 (Confidence interval for τ)

Under Assumption 1, an approx. $100(1-\alpha)\%$ C.I. for τ is given by

$$\widehat{\tau}^{\mathrm{R}} \pm \frac{z_{\alpha/2}}{\sqrt{N}} \left(12 \frac{m}{N}\left(1 - \frac{m}{N}\right) \widehat{\mathcal{I}}_N^2\right)^{-1/2}$$

where $\widehat{\mathcal{I}}_N$ is as in (2) and z_α is the upper α-th quantile of $\mathcal{N}(0,1)$.

Definition 5 (Asymptotic relative efficiency)

Let $\widehat{\tau}_{N,1}$ and $\widehat{\tau}_{N,2}$ be two asymptotically normal estimators of τ, in the sense that there exist positive sequences $\sigma^2_{N,1}$ and $\sigma^2_{N,2}$ such that

$$\frac{\widehat{\tau}_{N,1} - \tau}{\sigma_{N,1}} \xrightarrow{d} \mathcal{N}(0,1), \quad \text{and} \quad \frac{\widehat{\tau}_{N,2} - \tau}{\sigma_{N,2}} \xrightarrow{d} \mathcal{N}(0,1).$$

Then the asymptotic relative efficiency of $\widehat{\tau}_{N,1}$ with respect to $\widehat{\tau}_{N,2}$ is defined as

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

Efficiency of $\widehat{\tau}^{\mathrm{R}}$ compared to $\widehat{\tau}_{\mathrm{dm}}$

Theorem 6 (Efficiency lower bound, simplified)

Assume that the potential control outcomes $\{b_{N,j} : 1 \leq j \leq N\}$ are i.i.d. samples from a distribution with density f_b satisfying $\int_{\mathbb{R}} f_b^2(x)dx < \infty$. When the density $f_b(\cdot)$ admits a finite variance σ_b^2, the asymptotic efficiency of $\widehat{\tau}^{\mathrm{R}}$ relative to $\widehat{\tau}_{\mathrm{dm}}$ is given by

$$\mathrm{eff}(\widehat{\tau}^{\mathrm{R}}, \widehat{\tau}_{\mathrm{dm}}) = 12\sigma_b^2 \left(\int_{\mathbb{R}} f_b^2(x)dx \right)^2.$$

Further, if $\mathcal{F}$ be the family of all probability densities on $\mathbb{R}$, then

$$\inf_{\mathcal{F}} \mathrm{eff}(\widehat{\tau}^{\mathrm{R}}, \widehat{\tau}_{\mathrm{dm}}) \geq 0.864.$$

A more general version, relaxing the i.i.d. assumption, is available in our paper.

Efficiency of $\widehat{\tau}^{\mathrm{R}}$ compared to $\widehat{\tau}_{\mathrm{dm}}$ (Contd.)

Table 1: The values of $\mathrm{eff}(\widehat{\tau}^{\mathrm{R}}, \widehat{\tau}_{\mathrm{dm}})$ for some common distributions

distribution	density (f_b)	$\mathrm{eff}(\widehat{\tau}^{\mathrm{R}}, \widehat{\tau}_{\mathrm{dm}})$
Normal	$(2\pi)^{-1/2}\exp(-x^2/2)$	$3/\pi \approx 0.955$
Uniform	$\mathbf{1}_{\{0 \le x \le 1\}}$	1
Laplace	$2^{-1}\exp(-\|x\|)$	$3/2$
t_3	$c\left(x^2/3+1\right)^{-2}$	$75/(4\pi^2) \approx 1.9$
Exponential	$\exp(-x)\mathbf{1}_{\{x \ge 0\}}$	3
Pareto(α)	$\alpha x^{-(\alpha+1)}\mathbf{1}_{\{x \ge 1\}}$	$\begin{cases} \frac{\alpha^5}{(\alpha-1)^2(2\alpha+1)^2(\alpha-2)} & \text{if } \alpha > 2 \\ +\infty & \text{if } \alpha \in (0,2] \end{cases}$

Regression adjustment for covariates

Regression adjusted test statistic

Along with the responses Y_j's we also collect data on covariates $\boldsymbol{x}_j \in \mathbb{R}^p$. Let $\boldsymbol{X}_{N\times p}$ be the deterministic matrix of covariates.

- To test $H_0 : \tau = \tau_0$ vs. $H_1 : \tau \neq \tau_0$, first regress $\boldsymbol{Y} - \tau_0 \boldsymbol{Z}$ on $\boldsymbol{X}$

Regression adjusted test statistic

Along with the responses Y_j's we also collect data on covariates $\boldsymbol{x}_j \in \mathbb{R}^p$. Let $\boldsymbol{X}_{N\times p}$ be the deterministic matrix of covariates.

- To test $H_0 : \tau = \tau_0$ vs. $H_1 : \tau \neq \tau_0$, first regress $\boldsymbol{Y} - \tau_0 \boldsymbol{Z}$ on $\boldsymbol{X}$
- Calculate the null residuals: $\boldsymbol{e}_0 := (\boldsymbol{I} - \boldsymbol{P}_X)(\boldsymbol{Y} - \tau_0 \boldsymbol{Z})$.

Regression adjusted test statistic

Along with the responses Y_j's we also collect data on covariates $\boldsymbol{x}_j \in \mathbb{R}^p$. Let $\boldsymbol{X}_{N\times p}$ be the deterministic matrix of covariates.

- To test $H_0 : \tau = \tau_0$ vs. $H_1 : \tau \neq \tau_0$, first regress $\boldsymbol{Y} - \tau_0 \boldsymbol{Z}$ on $\boldsymbol{X}$
- Calculate the null residuals: $\boldsymbol{e}_0 := (\boldsymbol{I} - \boldsymbol{P}_X)(\boldsymbol{Y} - \tau_0 \boldsymbol{Z})$.
- Calculate the WRS statistic based on $\boldsymbol{e}_0$ (instead of $\boldsymbol{Y} - \tau_0 \boldsymbol{Z}$)

$$t_{N,\mathrm{adj}} := t_N(\boldsymbol{Z}, \boldsymbol{e}_0) = \sum_{i=1}^{N} Z_i \sum_{j=1}^{N} \mathbf{1}_{\{e_{0,j} \leq e_{0,i}\}}.$$

Regression adjusted test statistic

Along with the responses Y_j's we also collect data on covariates $\boldsymbol{x}_j \in \mathbb{R}^p$. Let $\boldsymbol{X}_{N\times p}$ be the deterministic matrix of covariates.

- To test $H_0 : \tau = \tau_0$ vs. $H_1 : \tau \neq \tau_0$, first regress $\boldsymbol{Y} - \tau_0 \boldsymbol{Z}$ on $\boldsymbol{X}$
- Calculate the null residuals: $\boldsymbol{e}_0 := (\boldsymbol{I} - \boldsymbol{P}_X)(\boldsymbol{Y} - \tau_0 \boldsymbol{Z})$.
- Calculate the WRS statistic based on $\boldsymbol{e}_0$ (instead of $\boldsymbol{Y} - \tau_0 \boldsymbol{Z}$)

$$t_{N,\mathrm{adj}} := t_N(\boldsymbol{Z}, \boldsymbol{e}_0) = \sum_{i=1}^{N} Z_i \sum_{j=1}^{N} \mathbf{1}_{\{e_{0,j} \leq e_{0,i}\}}.$$

- $t_{N,\mathrm{adj}}$ has the same null distribution as its unadjusted counterpart t_N (since $\boldsymbol{e}_0$ is non-random under H_0, same proof applies).

- Regress $\boldsymbol{Y} - \tau \boldsymbol{Z}$ on $\boldsymbol{X}$ using ordinary least squares. Residuals:

$$\boldsymbol{e}_\tau := (\boldsymbol{I} - \boldsymbol{P}_X)(\boldsymbol{Y} - \tau \boldsymbol{Z})$$

where $\boldsymbol{P}_X$ is the projection matrix onto the column space of $\boldsymbol{X}$.

- As in unadjusted case, set $\widehat{\tau}^{*}_{\mathrm{adj}} := \sup\{\tau : t_N(\boldsymbol{Z}, \boldsymbol{e}_\tau) > \mu\}$, and $\widehat{\tau}^{**}_{\mathrm{adj}} := \inf\{\tau : t_N(\boldsymbol{Z}, \boldsymbol{e}_\tau) < \mu\}$, and define

$$\widehat{\tau}^{\mathrm{R}}_{\mathrm{adj}} := \frac{\widehat{\tau}^{*}_{\mathrm{adj}} + \widehat{\tau}^{**}_{\mathrm{adj}}}{2}.$$

Main theoretical results
(for the regression adjusted estimator)

- We add a subscript N for the subsequent asymptotic results.
- Define

$$\widetilde{b}_{N,j} := b_{N,j} - \boldsymbol{p}_{N,j}^{\top} \boldsymbol{b}_N, \quad 1 \leq j \leq N,$$

where $\boldsymbol{p}_{N,j}$ is the j-th column of the projection matrix $\boldsymbol{P}_{\boldsymbol{X}_N}$ that projects onto the column space of $\boldsymbol{X}_N$.

- We add a subscript N for the subsequent asymptotic results.
- Define

$$\widetilde{b}_{N,j} := b_{N,j} - \boldsymbol{p}_{N,j}^\top \boldsymbol{b}_N, \quad 1 \leq j \leq N,$$

  where $\boldsymbol{p}_{N,j}$ is the j-th column of the projection matrix $\boldsymbol{P}_{\boldsymbol{X}_N}$ that projects onto the column space of $\boldsymbol{X}_N$.
- Under $H_0 : \tau = \tau_0$, the null residuals are given by:

$$(\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{X}_N})(\boldsymbol{Y}_N - \tau_0 \boldsymbol{Z}_N) = (\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{X}_N})\boldsymbol{b}_N = \widetilde{\boldsymbol{b}}_N.$$

- $\widetilde{\boldsymbol{b}}_N$ will play the role of $\boldsymbol{b}_N$ in the unadjusted case.

The following mimics Assumption 1 of the regression unadjusted case.

Assumption 2

Let $\widetilde{b}_{N,j} := b_{N,j} - \boldsymbol{p}_{N,j}^{\top} \boldsymbol{b}_N,\ 1 \leq j \leq N$. *We assume that,*

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

for some fixed $\mathcal{J}_b \in (0, \infty)$.

This holds in probability when $\boldsymbol{b}_N = \boldsymbol{X}_N \boldsymbol{\beta}_N + \boldsymbol{\varepsilon}_N$, where $\varepsilon_{N,1}, \ldots, \varepsilon_{N,N}$ are i.i.d. from $\mathcal{N}(0, \sigma^2)$. In fact, $\mathcal{J}_b = (2\sqrt{\pi}\sigma)^{-1}$ in this case.

Theorem 7 (Local asymptotic normality of $t_{N,\mathrm{adj}}$)

Let $t_{N,\mathrm{adj}}$ be the regression adjusted WRS statistic. Suppose that Assumption 2 holds. Fix $h \in \mathbb{R}$ and let $\tau_N = \tau_0 - hN^{-1/2}$. Then, under $\tau = \tau_N$,

$$N^{-3/2}\left(t_{N,\mathrm{adj}} - \frac{m(N+1)}{2}\right) \xrightarrow{d} \mathcal{N}\left(-h\lambda(1-\lambda)\mathcal{J}_b, \frac{\lambda(1-\lambda)}{12}\right).$$

Asymptotic distribution of $\widehat{\tau}^{\mathrm{R}}_{\mathrm{adj}}$

Theorem 7 (Local asymptotic normality of $t_{N,\mathrm{adj}}$)

Let $t_{N,\mathrm{adj}}$ be the regression adjusted WRS statistic. Suppose that Assumption 2 holds. Fix $h \in \mathbb{R}$ and let $\tau_N = \tau_0 - hN^{-1/2}$. Then, under $\tau = \tau_N$,

$$N^{-3/2}\left(t_{N,\mathrm{adj}} - \frac{m(N+1)}{2}\right) \xrightarrow{d} \mathcal{N}\left(-h\lambda(1-\lambda)\mathcal{J}_b, \frac{\lambda(1-\lambda)}{12}\right).$$

Theorem 8 (CLT for the estimator $\widehat{\tau}^{\mathrm{R}}_{\mathrm{adj}}$)

Under Assumption 2, it holds that

$$\sqrt{N}\left(\widehat{\tau}^{\mathrm{R}}_{\mathrm{adj}} - \tau_0\right) \xrightarrow{d} \mathcal{N}\left(0, (12\lambda(1-\lambda)\mathcal{J}_b^2)^{-1}\right).$$

$\mathcal{J}_b$ is defined in Assumption 2.

Theorem 9 (Efficiency gain by regression adjustment)

Assume that the model $\boldsymbol{b}_N = \boldsymbol{X}_N \boldsymbol{\beta}_N + \boldsymbol{\varepsilon}_N$ *holds, where* $\varepsilon_{N,i}$*'s are i.i.d.* $\mathcal{N}(0, \sigma^2)$*. Then Assumption 2 holds, with* $\mathcal{J}_b = (2\sqrt{\pi}\sigma)^{-1}$*. Further, with* $\boldsymbol{v}_N := \boldsymbol{X}_N \boldsymbol{\beta}_N$*, if* $\lim_{N\to\infty} N^{-2} \sum_{j=1}^{N} \sum_{i=1}^{N} e^{-(v_{N,j}-v_{N,i})^2/4\sigma^2} = \ell$*, then Assumption 1 holds with* $\mathcal{I}_b = \ell \mathcal{J}_b$*, and consequently,*

$$\mathcal{J}_b \geq \mathcal{I}_b, \quad \textit{i.e.,} \quad \mathrm{eff}(\widehat{\tau}^{\mathrm{R}}_{\mathrm{adj}}, \widehat{\tau}^{\mathrm{R}}) \geq 1.$$

Moreover, if $\liminf_{N\to\infty} N^{-1} \sum_{j=1}^{N} (v_{N,j} - \overline{v}_N)^2 > 0$*, then*

$$\mathcal{J}_b > \mathcal{I}_b, \quad \textit{i.e.,} \quad \mathrm{eff}(\widehat{\tau}^{\mathrm{R}}_{\mathrm{adj}}, \widehat{\tau}^{\mathrm{R}}) > 1.$$

What if treatment effect is not constant?

- Confidence intervals with Rosenbaum's estimator are sometimes used in practice without constant treatment assumption. What is the corresponding estimand?

What if treatment effect is not constant?

- Confidence intervals with Rosenbaum's estimator are sometimes used in practice without constant treatment assumption. What is the corresponding estimand?
- Assume no regression adjustment
- Set

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

What if treatment effect is not constant?

- Confidence intervals with Rosenbaum's estimator are sometimes used in practice without constant treatment assumption. What is the corresponding estimand?
- Assume no regression adjustment
- Set

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

Theorem 10 (Efficiency gain by regression adjustment)

Under appropriate assumptions, Rosenbaum's estimator $\hat{\tau}_R$ satisfies

$$\hat{\tau}_R - med_N \xrightarrow{P} 0.$$

- The empirical median should be well separated.
- For $\epsilon > 0$, set

$$\kappa_N^{(1)} := \frac{1}{N(N-1)} \sum_{1 \leq i \neq j \leq N} \mathbf{1}(a_i - b_j \leq \mathrm{med}_N + \epsilon) \geq \frac{1}{2}.$$

Define $\kappa_N^{(2)}$ similarly by replacing $\leq$ with $\geq$ and $\mathrm{med}_N + \epsilon$ with $\mathrm{med}_N - \epsilon$.

On our assumptions

- The empirical median should be well separated.
- For $\epsilon > 0$, set

$$\kappa_N^{(1)} := \frac{1}{N(N-1)} \sum_{1 \le i \ne j \le N} \mathbf{1}(a_i - b_j \le \text{med}_N + \epsilon) \ge \frac{1}{2}.$$

  Define $\kappa_N^{(2)}$ similarly by replacing $\le$ with $\ge$ and $\text{med}_N + \epsilon$ with $\text{med}_N - \epsilon$.
- We need some gap (potentially vanishing) between $\kappa_N^{(1)}$, $\kappa_N^{(2)}$, and $1/2$, i.e.,

$$\sqrt{N}(\kappa_N^{(1)} - 0.5) \to \infty, \qquad \sqrt{N}(0.5 - \kappa_N^{(2)}) \to \infty.$$

On our assumptions

- The empirical median should be well separated.
- For $\epsilon > 0$, set

$$\kappa_N^{(1)} := \frac{1}{N(N-1)} \sum_{1 \le i \neq j \le N} \mathbf{1}(a_i - b_j \le \mathrm{med}_N + \epsilon) \ge \frac{1}{2}.$$

Define $\kappa_N^{(2)}$ similarly by replacing $\le$ with $\ge$ and $\mathrm{med}_N + \epsilon$ with $\mathrm{med}_N - \epsilon$.

- We need some gap (potentially vanishing) between $\kappa_N^{(1)}$, $\kappa_N^{(2)}$, and $1/2$, i.e.,

$$\sqrt{N}(\kappa_N^{(1)} - 0.5) \to \infty, \qquad \sqrt{N}(0.5 - \kappa_N^{(2)}) \to \infty.$$

- Under stronger separability, $\sqrt{N}(\hat{\tau}_R - \mathrm{med}_N) = O_0(1)$.

Applications

- Aim to study the electoral impact of *Progresa*, Mexico's conditional cash transfer program (CCT program) (De La O, 2013; Imai, 2018).

Application 1 : Progresa data

- Aim to study the electoral impact of *Progresa*, Mexico's conditional cash transfer program (CCT program) (De La O, 2013; Imai, 2018).
- Eligible villages were randomly assigned to receive the program either 21 months (treated) or 6 months (control) before the 2000 Mexican presidential election.

Application 1 : Progresa data

- Aim to study the electoral impact of *Progresa*, Mexico's conditional cash transfer program (CCT program) (De La O, 2013; Imai, 2018).
- Eligible villages were randomly assigned to receive the program either 21 months (treated) or 6 months (control) before the 2000 Mexican presidential election.
- 417 observations each representing a precinct, and for each precinct we have its treatment status, outcomes of interest, socioeconomic indicators, and other precinct characteristics.

We use the same regression model as De La O (2013), taking

- outcome variable: *pri2000s* (support rates for the incumbent party (PRI) as shares of the eligible voting population)

Application 1 : Progresa data (Contd.)

We use the same regression model as De La O (2013), taking

- outcome variable: *pri2000s* (support rates for the incumbent party (PRI) as shares of the eligible voting population)
- covariates:
  - average poverty level in a precinct (*avgpoverty*)
  - total precinct population in 1994 (*pobtot1994*)
  - total no. of voters turned out in the previous election (*votos1994*)
  - total no. of votes cast for each of the three main competing parties in the previous election (*pri1994*, *pan1994*, and *prd1994*)
  - include *villages* as factors.

Table 2: Different estimates of the effect of early Progresa on PRI support rates with the corresp. standard errors, 95% approximate C.I.'s and their lengths.

	estimate	std.error	95% C.I.	length
$\widehat{\tau}^{\mathrm{R}}$	1.834	0.446	[0.960, 2.707]	**1.747**
$\widehat{\tau}_{\mathrm{dm}}$	3.622	1.728	[0.235, 7.010]	6.774
$\widehat{\tau}^{\mathrm{R}}_{\mathrm{adj}}$	2.185	0.411	[1.380, 2.989]	**1.610**
$\widehat{\tau}_{\mathrm{adj}}$	3.671	1.510	[0.712, 6.630]	5.917
$\widehat{\tau}_{\mathrm{interact}}$	4.214	1.462	[1.348, 7.079]	5.731

Table 2: Different estimates of the effect of early Progresa on PRI support rates with the corresp. standard errors, 95% approximate C.I.'s and their lengths.

	estimate	std.error	95% C.I.	length
$\widehat{\tau}^{\mathrm{R}}$	1.834	0.446	[0.960, 2.707]	**1.747**
$\widehat{\tau}_{\mathrm{dm}}$	3.622	1.728	[0.235, 7.010]	6.774
$\widehat{\tau}^{\mathrm{R}}_{\mathrm{adj}}$	2.185	0.411	[1.380, 2.989]	**1.610**
$\widehat{\tau}_{\mathrm{adj}}$	3.671	1.510	[0.712, 6.630]	5.917
$\widehat{\tau}_{\mathrm{interact}}$	4.214	1.462	[1.348, 7.079]	5.731

- Std. errors of $\widehat{\tau}^{\mathrm{R}}$ or $\widehat{\tau}^{\mathrm{R}}_{\mathrm{adj}}$ are much less than that of $\widehat{\tau}_{\mathrm{dm}}$, $\widehat{\tau}_{\mathrm{adj}}$, or $\widehat{\tau}_{\mathrm{interact}}$.

Table 2: Different estimates of the effect of early Progresa on PRI support rates with the corresp. standard errors, 95% approximate C.I.'s and their lengths.

	estimate	std.error	95% C.I.	length
$\widehat{\tau}^{\mathrm{R}}$	1.834	0.446	[0.960, 2.707]	**1.747**
$\widehat{\tau}_{\mathrm{dm}}$	3.622	1.728	[0.235, 7.010]	6.774
$\widehat{\tau}^{\mathrm{R}}_{\mathrm{adj}}$	2.185	0.411	[1.380, 2.989]	**1.610**
$\widehat{\tau}_{\mathrm{adj}}$	3.671	1.510	[0.712, 6.630]	5.917
$\widehat{\tau}_{\mathrm{interact}}$	4.214	1.462	[1.348, 7.079]	5.731

- Std. errors of $\widehat{\tau}^{\mathrm{R}}$ or $\widehat{\tau}^{\mathrm{R}}_{\mathrm{adj}}$ are much less than that of $\widehat{\tau}_{\mathrm{dm}}$, $\widehat{\tau}_{\mathrm{adj}}$, or $\widehat{\tau}_{\mathrm{interact}}$.
- Std. error of $\widehat{\tau}^{\mathrm{R}}_{\mathrm{adj}}$ is slightly less than that of $\widehat{\tau}^{\mathrm{R}}$.

Table 2: Different estimates of the effect of early Progresa on PRI support rates with the corresp. standard errors, 95% approximate C.I.'s and their lengths.

	estimate	std.error	95% C.I.	length
$\widehat{\tau}^{\mathrm{R}}$	1.834	0.446	[0.960, 2.707]	**1.747**
$\widehat{\tau}_{\mathrm{dm}}$	3.622	1.728	[0.235, 7.010]	6.774
$\widehat{\tau}^{\mathrm{R}}_{\mathrm{adj}}$	2.185	0.411	[1.380, 2.989]	**1.610**
$\widehat{\tau}_{\mathrm{adj}}$	3.671	1.510	[0.712, 6.630]	5.917
$\widehat{\tau}_{\mathrm{interact}}$	4.214	1.462	[1.348, 7.079]	5.731

- Std. errors of $\widehat{\tau}^{\mathrm{R}}$ or $\widehat{\tau}^{\mathrm{R}}_{\mathrm{adj}}$ are much less than that of $\widehat{\tau}_{\mathrm{dm}}$, $\widehat{\tau}_{\mathrm{adj}}$, or $\widehat{\tau}_{\mathrm{interact}}$.
- Std. error of $\widehat{\tau}^{\mathrm{R}}_{\mathrm{adj}}$ is slightly less than that of $\widehat{\tau}^{\mathrm{R}}$.
- Each of the confidence intervals suggests that the CCT program led to a significant positive increase in support for the incumbent party.

- Property sales for Mecklenburg County, North Carolina (Jan'94 – Dec'04) **source:** replication files of Linden and Rockoff ('08)

Application 2 : House price data

- Property sales for Mecklenburg County, North Carolina (Jan'94 – Dec'04) **source:** replication files of Linden and Rockoff ('08)
- Even after taking logarithm of house prices, the distribution is heavily skewed on the right side.

- Property sales for Mecklenburg County, North Carolina (Jan'94 – Dec'04) **source:** replication files of Linden and Rockoff ('08)
- Even after taking logarithm of house prices, the distribution is heavily skewed on the right side.
- Following Athey et al. (2021), we draw subsamples from the dataset and randomly assign exactly half of each sample to the treatment group and the remaining half to the control group. Thus, we know apriori that the treatment effect is zero.

- We draw subsamples of size $n = 1000$ in each iteration, and take the log of the house prices as the outcome variable.
- Use several features of the houses (e.g., sales year, age of the house, number of bedrooms, etc.) as covariates.
- Model fit is quite satisfactory, with adjusted $R^2 \approx 0.7$.
- The estimates, along with their standard errors and approximate 95% C.I.'s obtained from a single simulation are shown in Table 3.
- Repeating this experiment $B = 1000$ times, we report the coverage and average lengths of the C.I.'s in Table 4.

Table 3: Results from a single simulation from the house price data

	estimate	std.error	95% C.I.	length
$\widehat{\tau}^{\mathrm{R}}$	-0.04	0.03	[-0.09, 0.02]	0.12
$\widehat{\tau}_{\mathrm{dm}}$	-0.05	0.04	[-0.12, 0.03]	0.15
$\widehat{\tau}^{\mathrm{R}}_{\mathrm{adj}}$	-0.01	0.01	[-0.03, 0.02]	0.04
$\widehat{\tau}_{\mathrm{adj}}$	-0.02	0.02	[-0.06, 0.03]	0.08
$\widehat{\tau}_{\mathrm{interact}}$	-0.02	0.02	[-0.06, 0.03]	0.08

Table 4: Coverage and average lengths of the approximate 95% C.I.'s obtained from different estimators by repeated simulations from the house price data

	$\widehat{\tau}^{\mathrm{R}}$	$\widehat{\tau}_{\mathrm{dm}}$	$\widehat{\tau}^{\mathrm{R}}_{\mathrm{adj}}$	$\widehat{\tau}_{\mathrm{adj}}$	$\widehat{\tau}_{\mathrm{interact}}$
coverage	0.944	0.959	0.954	0.952	0.949
avg length	**0.121**	0.143	**0.041**	0.073	0.073

Thank You!

Questions?

Paper: https://arxiv.org/abs/2111.15524

(Major revision at Biometrika)