Inference for Cluster Randomized Experiments with Non-ignorable Cluster Sizes

Federico Bugni
Northwestern

Ivan Canay
Northwestern

Azeem Shaikh
Chicago

Max Tabord-Meehan
Chicago

December 2022

Motivation

Randomized Controlled Trials (RCTs) increasingly used in economics.

Many such RCTs are cluster randomized.

Cluster Randomization

Consider RCT to evaluate educational intervention:

- $Y_{i,g}(0)$: student test score in absence of tutoring program
- $Y_{i,g}(1)$: student test score in presence of tutoring program
- $A_g \in \{0, 1\}$: tutoring program applied at school level
- $Y_{i,g} := Y_{i,g}(1)A_g + Y_{i,g}(0)(1 - A_g)$

Cluster Randomization

Questions to consider:

- What are potential parameters of interest?
  - Schools vary in size. Size may relate to outcomes.
- Might only sample subset of students in each school. Any consequences for estimation/inference?
- Applicability of "standard" approaches to estimation and inference?

Contribution

This paper

- Proposes "super-population" framework where cluster sizes modeled as random and can relate to outcomes
- Distinguishes between two distinct ATE parameters
- Studies estimation and inference under additional complication of two-stage sampling
- Discusses connection to existing finite population results for cluster RCTs

Contribution Part II (Bonus!)

Preview of follow-up paper! (Bai, Liu, Shaikh, Tabord-Meehan)

- Leverages Bugni et al. (2022) framework to study cluster matched-pair designs.
- Formalizes gain in efficiency from matching on cluster size
- Provides asymptotically exact method of inference
- Studies asymptotically-valid and finite-sample robust permutation test

(Some) Related Literature

- **Super-population analyses of unit-level RCTs:** Armstrong (2022), Bai Romano Shaikh (2021), Bai (2022), Bugni Canay Shaikh (2018, 2019), Bugni and Gao (2021), Cytrynbaum (2022), Ma et al. (2020), Negi and Wooldridge (2020), Tabord-Meehan (2021), Zhang and Zheng (2020)

- **Finite-population analyses of cluster RCTs:** Middleton and Aronow (2015), Athey and Imbens (2017), de Chaisemartin and Ramirez-Cuellar (2020), Schochet et al. (2021), Su and Ding (2021)

Outline

- Problem Setup
- Parameters of Interest
- Formal Results
- Simulation Study

Outline

- Problem Setup
- Parameters of Interest
- Formal Results
- Simulation Study
- Bonus Content

Setup of the Problem
Additional Notation

- Z_g observed baseline covariates for cluster g
- N_g size of cluster g
- $S_g \subseteq \{1, 2, \ldots, N_g\}$ sampled observations in cluster g
- $\bar{Y}_g(a) := \frac{1}{|S_g|} \sum_{i \in S_g} Y_{i,g}(a)$

Setup of the Problem

Sampling Framework

- $\{(\bar{Y}_g(1), \bar{Y}_g(0), |S_g|, Z_g, N_g) : 1 \leqslant g \leqslant G\}$ i.i.d
- $E[N_g^2] < \infty$
- $E[Y_{i,g}(a)^2 | N_g, Z_g] \leqslant C$
- $S_g \perp\!\!\!\perp (Y_{i,g}(1), Y_{i,g}(0) : 1 \leqslant i \leqslant N_g) | Z_g, N_g$
- $E[\bar{Y}_g(a)|N_g] = E\left[\frac{1}{N_g}\sum_{1\leqslant i\leqslant N_g} Y_{i,g}(a)|N_g\right]$

Experimental Design

- $\{A_g : 1 \leqslant g \leqslant G\}$ i.i.d with $P(A_g = 1) = \pi$

Parameters of Interest

Two parameters of interest:

$$E\left[\omega_g\left(\frac{1}{N_g}\sum_{1\leqslant i\leqslant N_g}Y_{i,g}(1)-Y_{i,g}(0)\right)\right]$$

with distinct weights $E[\omega_g]=1$.

Parameters of Interest

Equally-weighted ATE

Two parameters of interest:

$$E\left[\omega_g\left(\frac{1}{N_g}\sum_{1\leqslant i\leqslant N_g} Y_{i,g}(1)-Y_{i,g}(0)\right)\right]$$

Setting $\omega_g = 1$ obtains

$$\theta_1 = E\left[\frac{1}{N_g}\sum_{1\leqslant i\leqslant N_g} Y_{i,g}(1)-Y_{i,g}(0)\right]$$

Parameters of Interest

Size-weighted ATE

Two parameters of interest:

$$E\left[\omega_g\left(\frac{1}{N_g}\sum_{1\leqslant i\leqslant N_g} Y_{i,g}(1)-Y_{i,g}(0)\right)\right]$$

Setting $\omega_g = \frac{N_g}{E[N_g]}$ obtains

$$\theta_2 = \frac{E\left[\sum_{1\leqslant i\leqslant N_g} Y_{i,g}(1)-Y_{i,g}(0)\right]}{E[N_g]}$$

Parameters of Interest

Typically, we expect θ_1 and θ_2 to be distinct parameters.

In some cases they are the same, for example:

- If $N_g = k$ for all g
- If $Y_{i,g}(1) - Y_{i,g}(0) = \tau$ for all i, g

Results: Difference-in-Means

Consider

$$\hat{\theta}_G^{\mathrm{alt}} := \frac{\sum_{1 \leqslant g \leqslant G} \sum_{i \in S_g} Y_{i,g} A_g}{\sum_{1 \leqslant g \leqslant G} |S_g| A_g} - \frac{\sum_{1 \leqslant g \leqslant G} \sum_{i \in S_g} Y_{i,g}(1 - A_g)}{\sum_{1 \leqslant g \leqslant G} |S_g|(1 - A_g)} \ .$$

Probability Limit: $\hat{\theta}_G^{\mathrm{alt}}$

$$\hat{\theta}_G^{\mathrm{alt}} \xrightarrow{P} E\left[\frac{1}{E[|S_g|]} \sum_{i \in S_g} Y_{i,g}(1) - Y_{i,g}(0)\right] =: \vartheta$$

Results: Difference-in-Means

$\vartheta = E\left[\frac{1}{E[|S_g|]}\sum_{i\in S_g} Y_{i,g}(1) - Y_{i,g}(0)\right]$ is a sample-weighted ATE:

- Typically distinct from θ_1 and θ_2
- If $|S_g| = k$, then $\vartheta = \theta_1$
- If $|S_g| = \lfloor \gamma N_g \rfloor$ for $\gamma \in (0, 1]$, then $\vartheta \approx \theta_2$

Results: Inference on Equally-weighted ATE

Let

$$\hat{\theta}_{1,G} := \frac{\sum_{1 \leqslant g \leqslant G} \bar{Y}_g A_g}{\sum_{1 \leqslant g \leqslant G} A_g} - \frac{\sum_{1 \leqslant g \leqslant G} \bar{Y}_g (1 - A_g)}{\sum_{1 \leqslant g \leqslant G} (1 - A_g)} \ .$$

Limiting Distribution: $\hat{\theta}_{1,G}$

$$\sqrt{G}(\hat{\theta}_{1,G} - \theta_1) \xrightarrow{d} N(0, \sigma_1^2)$$

where

$$\sigma_1^2 := \frac{1}{\pi} \mathsf{Var}[\bar{Y}_g(1)] + \frac{1}{1-\pi} \mathsf{Var}[\bar{Y}_g(0)]$$

Results: Inference on Equally-weighted ATE

- Equivalent to individual-level analysis on cluster averages
- Estimator $\hat{\sigma}_1^2$ can be obtained as robust variance estimator from regression of $\bar{Y}_g$ on a constant and A_g.

Results: Inference on Size-weighted ATE

Let

$$\hat{\theta}_{2,G} := \frac{\sum_{1 \leqslant g \leqslant G} \bar{Y}_g N_g A_g}{\sum_{1 \leqslant g \leqslant G} N_g A_g} - \frac{\sum_{1 \leqslant g \leqslant G} \bar{Y}_g N_g (1 - A_g)}{\sum_{1 \leqslant g \leqslant G} N_g (1 - A_g)} .$$

Limiting Distribution: $\hat{\theta}_{2,G}$

$$\sqrt{G}(\hat{\theta}_{2,G} - \theta_2) \xrightarrow{d} N(0, \sigma_2^2)$$

where

$$\sigma_2^2 := \frac{1}{E[N_g]^2} \left(\frac{E\left[\left(\frac{N_g}{|S_g|} \right)^2 \left(\sum_{i \in S_g} \epsilon_{i,g}(1) \right)^2 \right]}{\pi} + \frac{E\left[\left(\frac{N_g}{|S_g|} \right)^2 \left(\sum_{i \in S_g} \epsilon_{i,g}(0) \right)^2 \right]}{1 - \pi} \right)$$

with

$$\epsilon_{i,g}(a) = Y_{i,g}(a) - \frac{E[N_g \bar{Y}_g(a)]}{E[N_g]} .$$

Results: Inference on Size-weighted ATE

- $\hat{\theta}_2$ can be obtained from WLS regression of $Y_{i,g}$ on a constant and A_g, with weights $\sqrt{N_g/|S_g|}$.
- Estimator $\hat{\sigma}_2^2$ is then obtained as cluster-robust variance estimator.

Finite Population Variance
(Su and Ding 2021)

Finite population version of σ_2^2 when $S_g = \{1, 2, \ldots, N_g\}$:

$$\sigma^2_{2,G,\text{finpop}} := \left(\frac{G}{N}\right)^2 \left(\frac{1}{G} \sum_{1 \leqslant g \leqslant G} \left[\frac{\left(\sum_{1 \leqslant i \leqslant N_g} \tilde{\epsilon}_{i,g}(1)\right)^2}{\pi} + \frac{\left(\sum_{1 \leqslant i \leqslant N_g} \tilde{\epsilon}_{i,g}(0)\right)^2}{1-\pi} \right] - \frac{1}{G} \sum_{1 \leqslant g \leqslant G} \left[\sum_{1 \leqslant i \leqslant N_g} \left(\tilde{\epsilon}_{i,g}(1) - \tilde{\epsilon}_{i,g}(0)\right) \right]^2 \right),$$

where

$$N := \sum_{1 \leqslant g \leqslant G} N_g$$

$$\tilde{\epsilon}_{i,g}(a) := Y_{i,g}(a) - \frac{1}{N} \sum_{1 \leqslant g \leqslant G} \sum_{1 \leqslant i \leqslant N_g} Y_{i,g}(a) .$$

Finite vs Super Population Variance

$$\sigma^2_{2,G,\text{finpop}} := \left(\frac{G}{N}\right)^2 \left(\frac{1}{G}\sum_{1\leqslant g\leqslant G}\left[\frac{\left(\sum_{1\leqslant i\leqslant N_g}\tilde{\epsilon}_{i,g}(1)\right)^2}{\pi} + \frac{\left(\sum_{1\leqslant i\leqslant N_g}\tilde{\epsilon}_{i,g}(0)\right)^2}{1-\pi}\right] - \frac{1}{G}\sum_{1\leqslant g\leqslant G}\left[\sum_{1\leqslant i\leqslant N_g}\left(\tilde{\epsilon}_{i,g}(1) - \tilde{\epsilon}_{i,g}(0)\right)\right]^2\right)$$

$$\sigma^2_2 := \frac{1}{E[N_g]^2}\left(\frac{E\left[\left(\sum_{1\leqslant i\leqslant N_g}\epsilon_{i,g}(1)\right)^2\right]}{\pi} + \frac{E\left[\left(\sum_{1\leqslant i\leqslant N_g}\epsilon_{i,g}(0)\right)^2\right]}{1-\pi}\right)$$

Finite vs Super Population Variance

$$\sigma^2_{2,G,\text{finpop}} := \left(\frac{G}{N}\right)^2 \left(\frac{1}{G}\sum_{1\leqslant g\leqslant G}\left[\frac{\left(\sum_{1\leqslant i\leqslant N_g}\tilde{\epsilon}_{i,g}(1)\right)^2}{\pi} + \frac{\left(\sum_{1\leqslant i\leqslant N_g}\tilde{\epsilon}_{i,g}(0)\right)^2}{1-\pi}\right] - \frac{1}{G}\sum_{1\leqslant g\leqslant G}\left[\sum_{1\leqslant i\leqslant N_g}(\tilde{\epsilon}_{i,g}(1)-\tilde{\epsilon}_{i,g}(0))\right]^2\right)$$

$$\sigma^2_2 := \frac{1}{E[N_g]^2}\left(\frac{E\left[\left(\sum_{1\leqslant i\leqslant N_g}\epsilon_{i,g}(1)\right)^2\right]}{\pi} + \frac{E\left[\left(\sum_{1\leqslant i\leqslant N_g}\epsilon_{i,g}(0)\right)^2\right]}{1-\pi}\right)$$

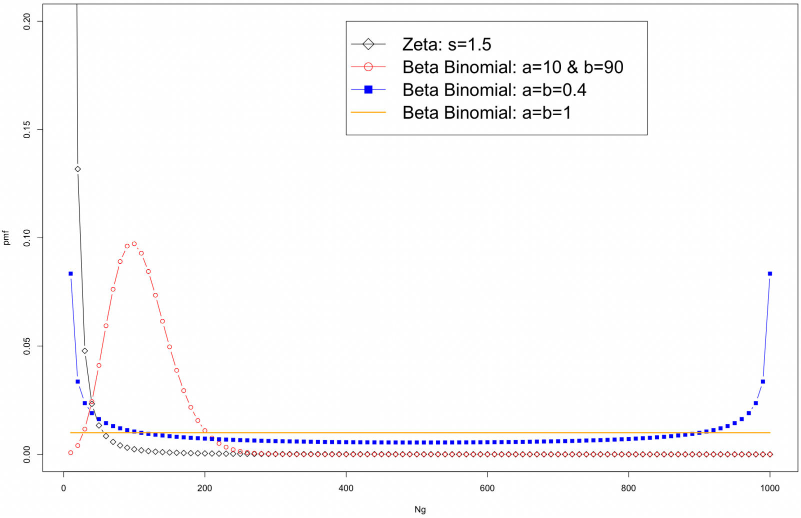
Simulations

DGP

- $Y_{i,g}(a) = \eta_g(a) Z_g + U_{i,g}(a)$
- $Z_g = Z_{g,big} I\{N_g \geqslant E[N_g]\} + Z_{g,small} I\{N_g < E[N_g]\}$
- $N_g = 10(B+1)$ where $B \sim BB(a, b, n_{supp})$ or
  $N_g = 10\zeta$ where $\zeta \sim \text{zeta}(1.5)$
- $|S_g| = N_g$

Simulations

Cluster Distributions



Simulations

Results

Design 2		$G = 100$		$G = 1000$		$G = 5000$	
$\|S_g\|$	N_g	$CS_{1,G}$	$CS_{2,G}$	$CS_{1,G}$	$CS_{2,G}$	$CS_{1,G}$	$CS_{2,G}$
Ng	$BB(1,1)$	0.9492	0.9384	0.9574	0.9532	0.9488	0.9530
	$BB(0.4,0.4)$	0.9486	0.9418	0.9516	0.9482	0.9492	0.9482
	$BB(10,90)$	0.9320	0.9312	0.9018	0.9072	0.9496	0.9492
	zeta(1.5)	0.9258	0.8510	0.8348	0.8918	0.7564	0.8722

Recap

- Proposed framework for cluster RCTs where cluster sizes modeled as random and can affect outcomes.
- Distinguished between two distinct ATE parameters.
- Studied estimation and inference under additional complication of two-stage sampling

Beyond Bernoulli Designs

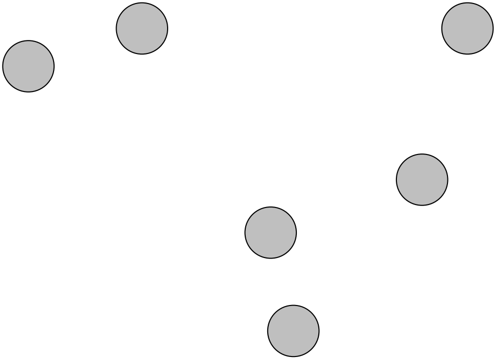
What about "realistic" experimental designs?

Bai, Liu, Shaikh, Tabord-Meehan (2022) study cluster matched-pair designs.

Beyond Bernoulli Designs

What about "realistic" experimental designs?

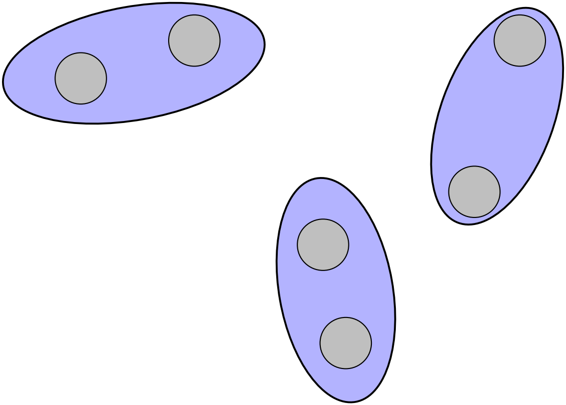
Bai, Liu, Shaikh, Tabord-Meehan (2022) study cluster matched-pair designs.



Beyond Bernoulli Designs

What about "realistic" experimental designs?

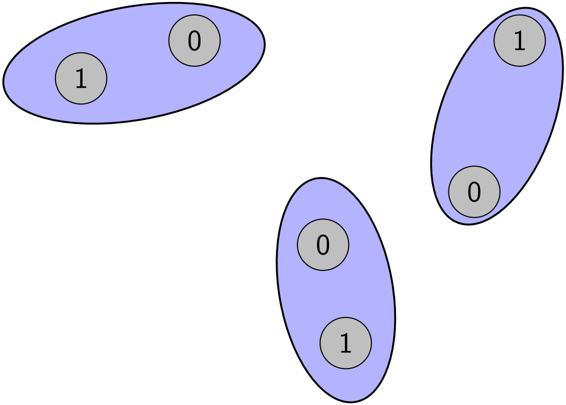
Bai, Liu, Shaikh, Tabord-Meehan (2022) study cluster matched-pair designs.



Beyond Bernoulli Designs

What about "realistic" experimental designs?

Bai, Liu, Shaikh, Tabord-Meehan (2022) study cluster matched-pair designs.



Additional Assumptions

Throughout suppose we have $2G$ clusters.

Matched Pairs

- G pairs represented by $\{\pi(2g-1), \pi(2g)\}$, $g = 1, \ldots, G$, $\pi = \pi_G(Z^{(G)})$ a permutation of $\{1, 2, \ldots, 2G\}$
- Conditional on $Z^{(G)}$, $(A_{\pi(2g-1)}, A_{\pi(2g)})$, $g = 1, \ldots, G$ are i.i.d uniform$\{(0,1),(1,0)\}$
- Pairing satisfies

$$\frac{1}{G}\sum_{g=1}^{G} ||Z_{\pi(2g)} - Z_{\pi(2g-1)}||^r \xrightarrow{P} 0 \ ,$$

for $r \in \{1, 2\}$

Additional Assumptions

Sampling Framework

- $E[\bar{Y}_g^r(a) N_g^\ell | Z_g = z]$, are Lipschitz for $r, \ell \in \{0, 1, 2\}$
- $E[N_g | Z_g] \leqslant C$

Results: Limiting Distribution of $\hat{\theta}_{2,G}$ for MP

Under this design we obtain:

Limiting Distribution: $\hat{\theta}_{2,G}$ for matched-pairs

$$\sqrt{G}(\hat{\theta}_{2,G} - \theta_2) \xrightarrow{d} N(0, \omega^2)$$

as $G \to \infty$, where

$$\omega^2 = E[\tilde{Y}_g^2(1)] + E[\tilde{Y}_g^2(0)] - \frac{1}{2}E[(E[\tilde{Y}_g(1) + \tilde{Y}_g(0)|Z_g])^2] \ ,$$

with

$$\tilde{Y}_g(a) = \frac{N_g}{E[N_g]}\left(\bar{Y}_g(a) - \frac{E[\bar{Y}_g(a)N_g]}{E[N_g]}\right) \ .$$

Results: Limiting Distribution of $\hat{\theta}_{2,G}$ for MP

- Note that $2\omega^2 = \sigma_2^2 - E[(E[\tilde{Y}_g(1) + \tilde{Y}_g(0)|Z_g])^2]$
  - Gain in precision from matched pairs
- We also show that, if matching on cluster size, variance is

$$2\nu^2 = \sigma_2^2 - E[(E[\tilde{Y}_g(1) + \tilde{Y}_g(0)|Z_g, N_g])^2]$$

  - By Jensen's, gain in precision from matching on cluster size

Results: Variance Estimation for ω^2

Note that ω^2 is exactly the asymptotic variance derived in Bai, Romano, Shaikh (2021), with cluster-transformed outcomes $\tilde{Y}_g(a)$.

We use this to construct consistent estimator of ω^2 and ν^2.

Results: Randomization Test

Paper also studies asymptotic validity of pair-permutation test for testing $H_0 : \theta_2 = 0$.

- Displays better size control for small G in simulations
- Crucial to studentize test-statistic using $\hat{v}_G^2$

Test is also finite-sample valid when “sharp”-null holds!

Thank you!

Cluster Size Consequences

Two consequences of our framework:

$$\frac{\sum_{1 \leqslant g \leqslant G} N_g^2}{\sum_{1 \leqslant g \leqslant G} N_g} = O_P(1)$$

$$\frac{\max_{1 \leqslant g \leqslant G} N_g^2}{\sum_{1 \leqslant g \leqslant G} N_g} \xrightarrow{P} 0$$

return

Numerical Example

Two types of classrooms: "Big" ($N_g = 40$) and "small" ($N_g = 10$)

$P(N_g = 40) = P(N_g = 10) = 0.5$

Suppose

$$Y_{i,g}(1) - Y_{i,g}(0) = 1 \text{ if in "big" class}$$

$$Y_{i,g}(1) - Y_{i,g}(0) = -2 \text{ if in "small" class}$$

Then

$$\theta_1 = -\frac{1}{2}$$

$$\theta_2 = \frac{2}{5}\ .$$

return