Revisiting Randomization with the Cube Method

38th meeting of the European Economic Association 2024

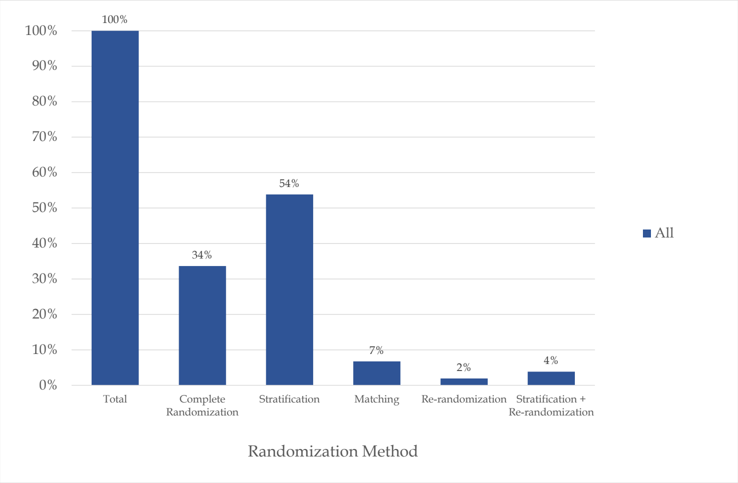
Laurent Davezies[1] Guillaume Hollard[2] Pedro Vergara Merino[1]

[1]CREST - ENSAE
[2]CREST - CNRS & Ecole Polytechnique

August 28, 2024

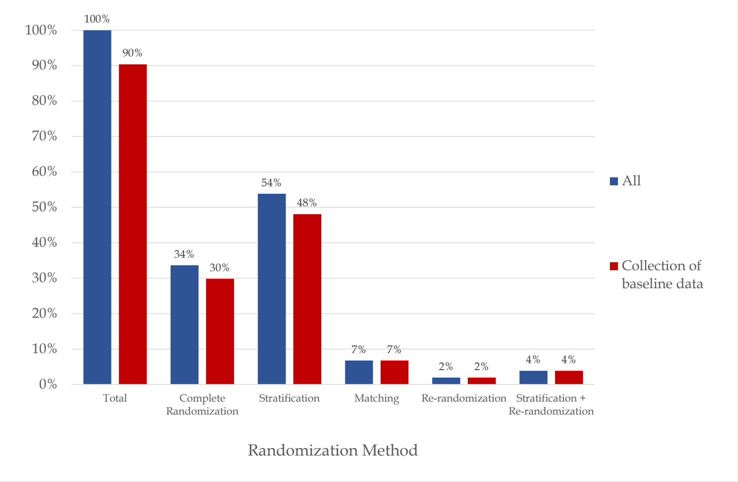
Golden Rule for Randomizing?

Figure 1: Distribution of randomization methods for 104 RCTs in top-5 journals (2019-2023)



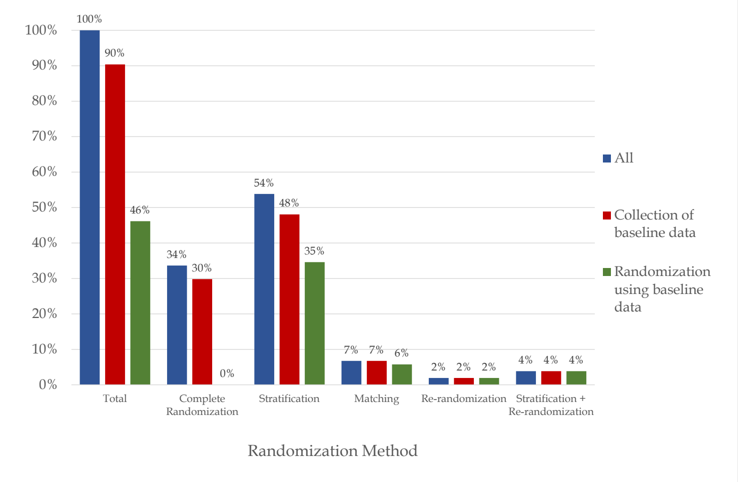
Golden Rule for Randomizing?

Figure 1: Distribution of randomization methods for 104 RCTs in top-5 journals (2019-2023)



Golden Rule for Randomizing?

Figure 1: Distribution of randomization methods for 104 RCTs in top-5 journals (2019-2023)



Why Balancing?

However, using available information (e.g., stratifying) allows to:

1. Improve precision of treatment effect estimates
2. Improve balance between the treatment and control groups (ex-post randomization checks)

Evidence of p-hacking and/or publication bias on ex-post balancing tests:

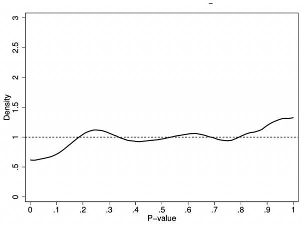


Figure 2: Distribution of 2,981 *p*-values of balancing checks in ‘pure’ RCTs
Source: Snyder and Zhuo (2018)

Research Questions

Is it possible to improve balancing using pre-randomization information?

Research Questions

Is it possible to improve balancing using pre-randomization information?

$\hookrightarrow$ Introduce the cube method to the RCT framework

Research Questions

Is it possible to improve balancing using pre-randomization information?

$\hookrightarrow$ Introduce the cube method to the RCT framework

How does the cube method compare to other randomization techniques?

Research Questions

Is it possible to improve balancing using pre-randomization information?

↪ Introduce the cube method to the RCT framework

How does the cube method compare to other randomization techniques?

↪ The cube method outperforms existing randomization methods on *many* dimensions!

Literature Review

1. The cube method (Chauvet and Tillé, 2006; Deville and Tillé, 2004, 2005; Tillé, 2011, 2022; Tillé and Favre, 2004, 2005).

   $\hookrightarrow$ Extend the scope of the cube method, a sampling algorithm, and formalize precision gains

2. Covariate-adaptive randomization and its benefits

   - *Stratification* (Bugni, Canay, and Shaikh, 2018; R. A. Fisher, 1926; S. R. A. Fisher, 1935)
   - *Matched pairs and local stratification* (Bai, 2022; Bai, Romano, and Shaikh, 2022; Cytrynbaum, 2023; Greevy et al., 2004; Higgins, Sävje, and Sekhon, 2016; Imai, King, and Nall, 2009)
   - *Re-randomization* (Li and Ding, 2017; Li, Ding, and Rubin, 2018; Morgan and Rubin, 2012)
   - *Gram-Schmidt Walk Design* (Harshaw et al., 2023)

   $\hookrightarrow$ Compare the performance of such methods as the number of balanced covariates increases

   $\hookrightarrow$ Introduce the cube method and inference methods to achieve greater precision gains

3. Practical implications for randomistas (Athey and Imbens, 2017; Bai, Shaikh, and Tabord-Meehan, 2024; Bruhn and McKenzie, 2009)

   $\hookrightarrow$ Discuss benefits arising from the cube method

Roadmap

1. Setup
2. The Cube Method
3. Results
4. Simulations
5. Empirical Application
6. Practical Implications

Data Generating Process

We consider the Neyman-Rubin causal framework, where nature generates for individual $i \in \{1, \ldots, n\}$

- $Y_i(1)$ the potential outcome when treated
- $Y_i(0)$ the potential outcome when untreated
- X_i a vector of p baseline characteristics

Assumption 1 (iid-ness+2nd moment).

$(Y_i(0), Y_i(1), X_i)$ are iid across i and $\mathbb{E}\left(Y(0)^2 + Y(1)^2 + ||X||^2\right) < \infty$

The empiricist only observes $(X_1, \ldots, X_n)$ before the experiment.

Assignment Design

Empiricists want to allocate treatment to n units according to a design Π.

D_i takes value 1 if i is treated, and 0 if untreated.

Empiricists choose Π, a probability distribution for $(D_i)_{i=1,\dots,n}|(X_i)_{i=1,\dots,n}$.

Assumption 2 (Restriction on the design Π).

- $(D_i)_{i=1,\dots,n} \perp\!\!\!\perp (Y_i(0), Y_i(1))_{i=1,\dots,n}|(X_i)_{i=1,\dots,n}$
- $\mathbb{P}_\Pi(D_i = 1|X_1, ..., X_n) = p(X_i) \in [c, 1-c], \forall i \in \{1, \dots, n\}$ with p a function chosen by the empiricist and for $c \in (0, 1/2)$

We note $\pi_i := p(X_i)$

After the experiment, she observes $Y_i = D_i Y_i(1) + (1 - D_i) Y_i(0)$.

Estimands and Estimators

We focus here in estimating the population average treatment effect (PATE):

$$\theta_0^* = \mathbb{E}\left[Y_i(1) - Y_i(0)\right].$$

via the Horvitz-Thompson estimator

$$\widehat{\theta}_{HT} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{Y_i D_i}{\pi_i} - \frac{Y_i(1-D_i)}{1-\pi_i}\right).$$

In the paper, we also show the results for the sample average treatment effect (SATE) and the Hájek estimator.

Perfect Balance

Definition.

A design Π is perfectly-balanced over $X = (X_1, ..., X_p)'$ if for $(D_i)_{i=1,...,n}$ sampled in Π we always have for any $j = 1, ..., p$:

- Balance in the treatment group:

$$\frac{1}{n}\sum_{i=1}^{n}\frac{X_{ji}D_i}{\pi_i} = \frac{1}{n}\sum_{i=1}^{n}X_{ji}$$

- Balance in the control group:

$$\frac{1}{n}\sum_{i=1}^{n}\frac{X_{ji}(1-D_i)}{1-\pi_i} = \frac{1}{n}\sum_{i=1}^{n}X_{ji}$$

Balancing Approximations

Remark.

If $(\pi_i)_{i=1,\dots,n}$ are heterogeneous, the set of constraints is defined as:

$$\frac{1}{n}\sum_{i=1}^{n}\frac{Z_{1i}D_i}{\pi_i} = \frac{1}{n}\sum_{i=1}^{n} Z_{1i}, \text{ with } Z_{1i} = \left(1, \frac{\pi_i}{1-\pi_i}, \pi_i, X_i', \frac{X_i'\pi_i}{1-\pi_i}\right)'.$$

Perfectly balanced designs are not always attainable!

For instance, if n is odd and $\pi_i = \frac{1}{2}$, then $\sum_{i=1}^{n}\pi_i = \frac{n}{2}$ is a non-integer, so

$$\sum_{i=1}^{n} D_i \neq \sum_{i=1}^{n}\pi_i$$

However, it is sufficient to have asymptotic balance:

$$\boxed{\frac{1}{n}\sum_{i=1}^{n}\frac{Z_{1i}D_i}{\pi_i} = \frac{1}{n}\sum_{i=1}^{n} Z_{1i} + o_p\left(\frac{1}{\sqrt{n}}\right)}$$

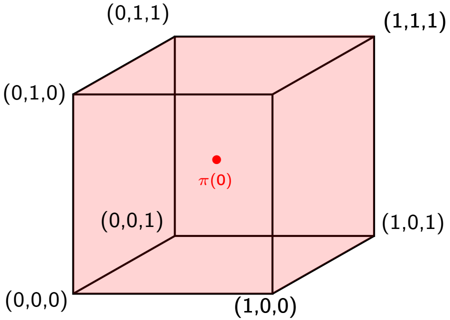
Optimal balancing

Roadmap

1. Setup
2. The Cube Method
3. Results
4. Simulations
5. Empirical Application
6. Practical Implications

Random Assignment as a Geometrical Problem

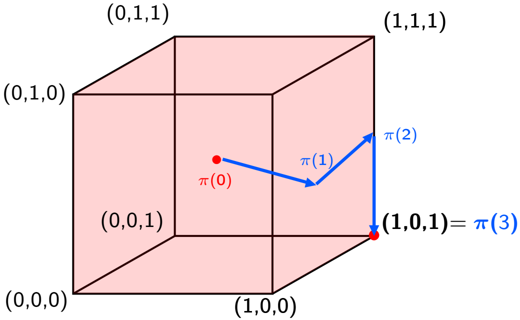
Treatment assignment can be seen as a random walk in a n-cube $\{0,1\}^n$. For $n = 3$:



$$\pi(0) \equiv (\pi_i)_{i=1,\dots,n}$$

Random Assignment as a Geometrical Problem

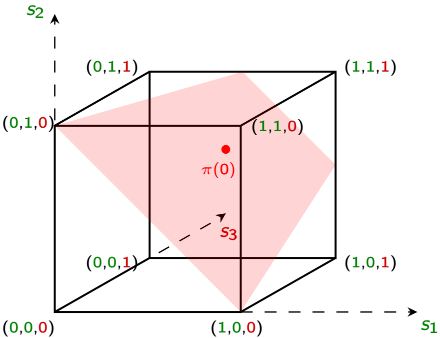
Treatment assignment can be seen as a random walk in a n-cube $\{0,1\}^n$. For $n = 3$:



$$\mathbb{E}[\pi(t+1)|\pi(t)] = \pi(t)$$

Random Assignment with Constraints

Imagine we want to balance the amount of savings. Let $X_1 = X_2 = 1000$ and $X_3 = -500$. We set $\pi_i = 2/3$ (i.e., two treated units on average).

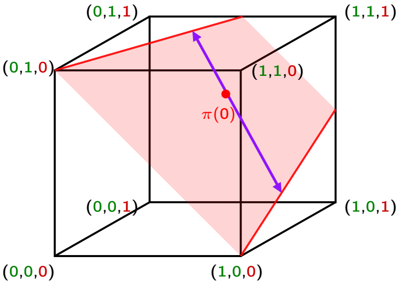


$$\text{Red Area} = \left\{ s \in [0,1]^3 \quad \middle| \quad s_1 + s_2 - \frac{1}{2} s_3 = 1 \right\}$$

Example balance constraints

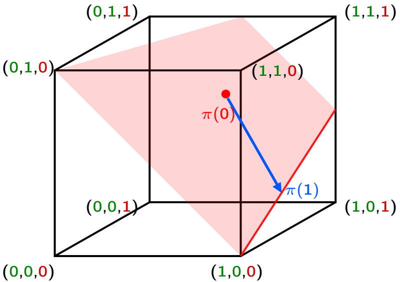
Random Assignment with Constraints

Imagine we want to balance the amount of savings. Let $X_1 = X_2 = 1000$ and $X_3 = -500$. We set $\pi_i = 2/3$ (i.e., two treated units on average).



Random Assignment with Constraints

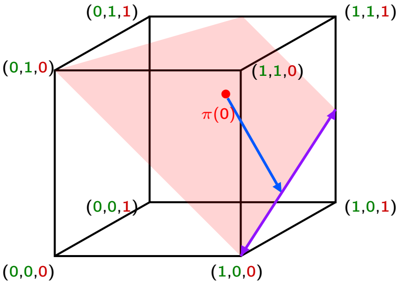
Imagine we want to balance the amount of savings. Let $X_1 = X_2 = 1000$ and $X_3 = -500$. We set $\pi_i = 2/3$ (i.e., two treated units on average).



$$\mathbb{E}[\pi(1)|\pi(0)] = \pi(0)$$

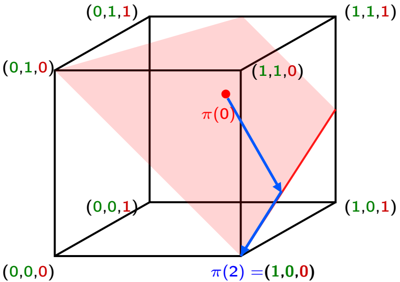
Random Assignment with Constraints

Imagine we want to balance the amount of savings. Let $X_1 = X_2 = 1000$ and $X_3 = -500$. We set $\pi_i = 2/3$ (i.e., two treated units on average).



Random Assignment with Constraints

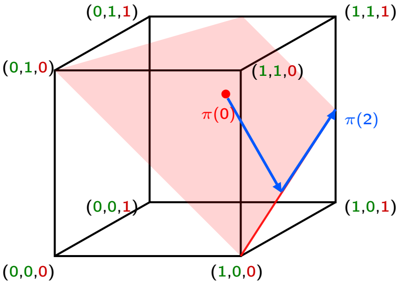
Imagine we want to balance the amount of savings. Let $X_1 = X_2 = 1000$ and $X_3 = -500$. We set $\pi_i = 2/3$ (i.e., two treated units on average).



$$\mathbb{E}[\pi(2)|\pi(1)] = \pi(1)$$

Random Assignment with Constraints

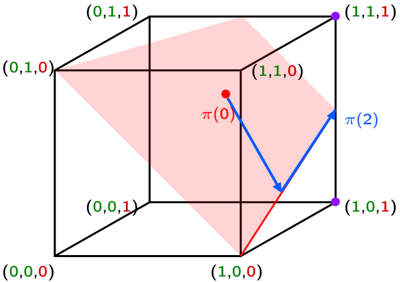
Imagine we want to balance the amount of savings. Let $X_1 = X_2 = 1000$ and $X_3 = -500$. We set $\pi_i = 2/3$ (i.e., two treated units on average).



$$\mathbb{E}[\pi(2)|\pi(1)] = \pi(1)$$

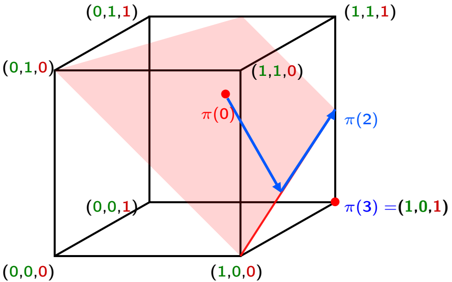
Random Assignment with Constraints

Imagine we want to balance the amount of savings. Let $X_1 = X_2 = 1000$ and $X_3 = -500$. We set $\pi_i = 2/3$ (i.e., two treated units on average).



Random Assignment with Constraints

Imagine we want to balance the amount of savings. Let $X_1 = X_2 = 1000$ and $X_3 = -500$. We set $\pi_i = 2/3$ (i.e., two treated units on average).



$$\mathbb{E}[\pi(3)|\pi(2)] = \pi(2)$$

Roadmap

Asymptotically-balancing Design

Proposition 1 (Balancing approximations with the cube method).

Let

$$\Delta^{\Pi}_{j,n} = \frac{1}{n}\sum_{i=1}^{n} \frac{X_{ji}D_i}{\pi_i} - \frac{X_{ji}(1-D_i)}{1-\pi_i}$$

If Assumptions 1 and 2 hold, then

$$\Delta^{Cube}_{j,n} = o_p\left(\frac{q}{\sqrt{n}}\right),$$

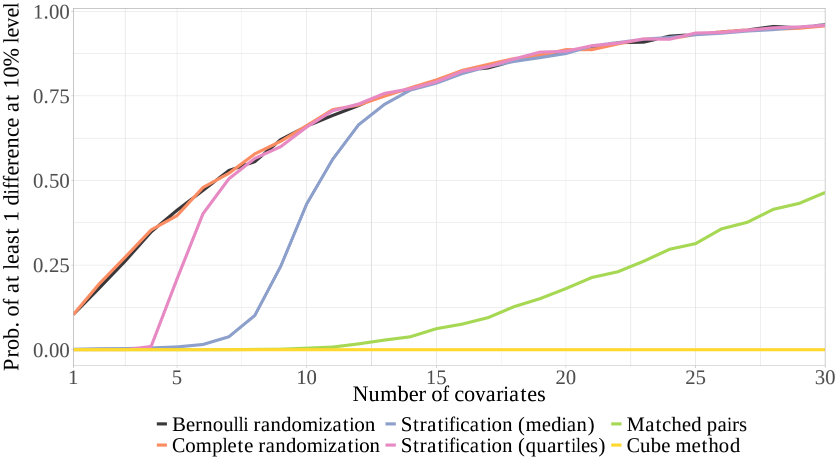
where q is the number of constraints.

In practice, balance tests become irrelevant for variables balanced with the cube method.

Full Proposition

Simulations with X_{1j} Independently Uniform

Figure 3: Impact of additional covariates on balance tests



Curse of Dimensionality and Balance Quality

Assumption 3.

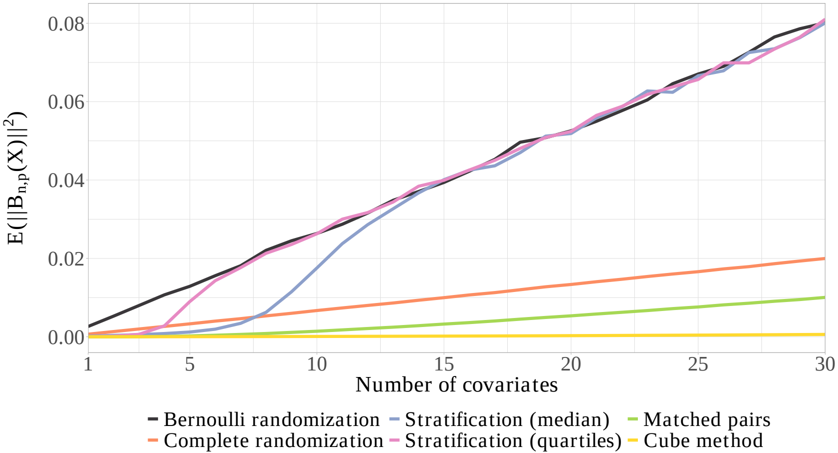
$\pi_i = 1/2$, n is a positive even number, and X_i are some $i.i.d$ random vectors of dimension p such that X_1 admits a density f_X with respect to the Lebesgue measure on $[0,1]^p$ and there exists some positive constants $\underline{C}$ and $\overline{C}$ (independent of p) such that for any $x \in [0,1]^p$, $\underline{C} < f_X(x) < \overline{C}$.

Measure of imbalance:

$$||B_{n,p}(X)||^2 = \sum_{j=1}^{p} \left(\frac{2}{n} \sum_{i=1}^{n} X_{ji} D_i - X_{ji}(1 - D_i) \right)^2$$

Simulations with X_{1j} Independently Uniform

Figure 4: Impact of additional covariates on balance quality



Comparison with Other Methods

These results hold in general under Assumption 3.

Expected imbalances $\mathbb{E}\left[||B_{n,p}(X)||^2\right]$ grow asymptotically at a rate:

- p^2/n^2 under the cube method Proposition
- p/n under SoA designs (stratification, matched pairs, ...) Proposition

This difference emerges from two perspectives on balancing:

- Moment approach: Trying to balance selected moments of X between treatment and control
- Distribution approach Trying to balance the joint density of X between treatment and control

Conjecture and Assumptions

Assumption 4 (Linearity).

For $d = 0, 1$,

$$Y_i(d) = Z'_{di}\beta_d + \varepsilon_i(d)$$

with $\mathbb{E}[\varepsilon_i(d)|Z_{di}] = 0$

Conjecture (Poisson approximation).

For any $k \in \mathbb{N}^*$ we have with probability one:

$$\lim_{n\to\infty} \sup_{i_1,\ldots,i_k} \left| \mathbb{E}\left(\prod_{j=1}^{k} \left(D_{i_j} - \pi_{i_j}\right) \middle| X_1, ..., X_n \right) \right| = 0$$

Asymptotic Normality

Proposition 3 (Asymptotic normality).

Let Assumptions 1, 2 and 3, and Conjecture 1 hold, the cube method yields for any $\pi_i \in [c, 1-c]$, $c > 0$,

$$\sqrt{n}\left(\widehat{\theta} - \theta_0^*\right) \xrightarrow{d} \mathcal{N}\left(0, V_0^*\right).$$

$$\text{for } V_0^* = \mathbb{V}(Z_1'\beta_1 - Z_0'\beta_0) + \mathbb{E}\left[\frac{\varepsilon_i(1)^2}{\pi_i}\right] + \mathbb{E}\left[\frac{\varepsilon_i(0)^2}{1-\pi_i}\right]$$

V_0^* is equal to the semiparametric efficiency bound from Hahn (1998).

Sketch of the proof

Asymptotic-Based Inference

Using the expression for V_0^* we can perform inference in the following steps:

1. Regress Y on Z_0 for the control group. Store coefficients $\hat{\beta}_0$ and residuals $\hat{\varepsilon}(0)$.
2. Regress Y on Z_1 for the treatment group. Store coefficients $\hat{\beta}_1$ and residuals $\hat{\varepsilon}(1)$.
3. Compute

$$\hat{V} = \frac{1}{n}\left[\hat{\mathbb{V}}((Z_1'\hat{\beta}_1 - Z_0'\hat{\beta}_0) + \frac{1}{n}\sum_{i=1}^{n}\frac{\hat{\varepsilon}_i(1)^2 D_i}{\pi_i^2} + \frac{1}{n}\sum_{i=1}^{n}\frac{\hat{\varepsilon}_i(0)^2(1-D_i)}{(1-\pi_i)^2}\right]$$

4. Compute $(1-\alpha)$-confidence intervals based on $\hat{\theta} \pm \Phi^{-1}\left(1-\frac{\alpha}{2}\right)\sqrt{(\hat{V})}$.

Randomization-based inference

Roadmap

Data Generating Process

We consider a simple DGP with non-linearities and a null ATE.
For $k \in \{1, \dots, K\}$, we simulate

- $Y_{ik}(0) = 1 + (X_{ik} - 1/2)'\beta_0 + \varepsilon_{ik}(0)$
- $Y_{ik}(1) = 1 + (X_{ik} - 1/2)'\beta_1 + (X_{ik} - 1/2)'A(X_{ik} - 1/2) + \varepsilon_{ik}(1)$

with

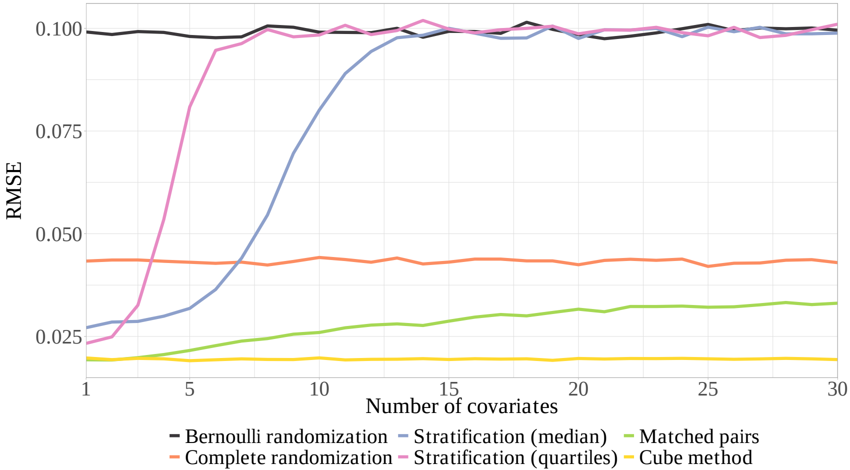
- $X_{jik} \sim 2 \times (\text{Beta}(2,2) - 1/2)$
- $\varepsilon_{ik}(d) \sim 0.1 \times \mathcal{N}(0,1)$
- $A = (1/20) \times (\mathbb{1}\mathbb{1}' - \text{diag}(1))$

and

$$\beta_0 = (1, 0, \dots, 0) \quad \text{and } \beta_1 = 2\beta_0$$

Design Effect on RMSE

Figure 5: Impact of additional covariates on sd $\left(\widehat{\theta}_n^{\Pi}\right)$

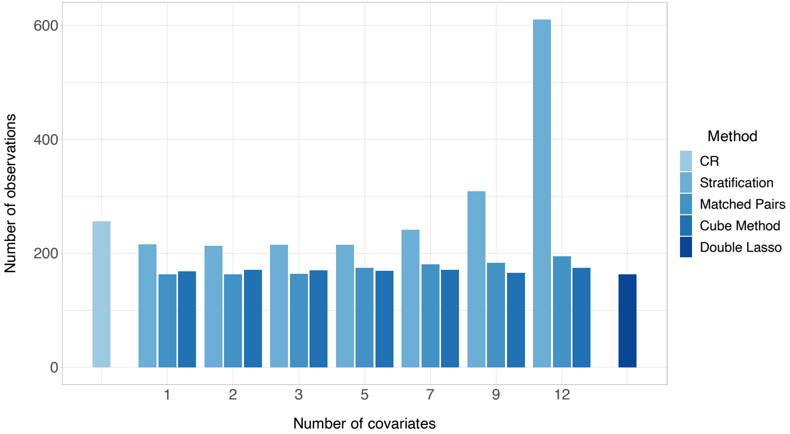


Roadmap

Effective Sample Size

We use Gerber et al. (2020) who investigate the effect of polls on beliefs and voting behavior.

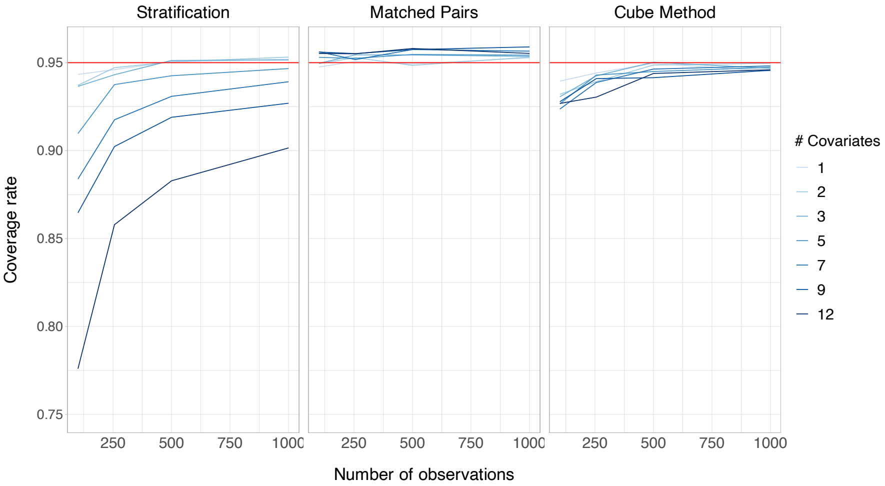
Figure 6: # of observations giving the same precision as in CR



DGP

Coverage Rate

Figure 7: Coverage rate of 95% confidence intervals



Roadmap

Practical Implications

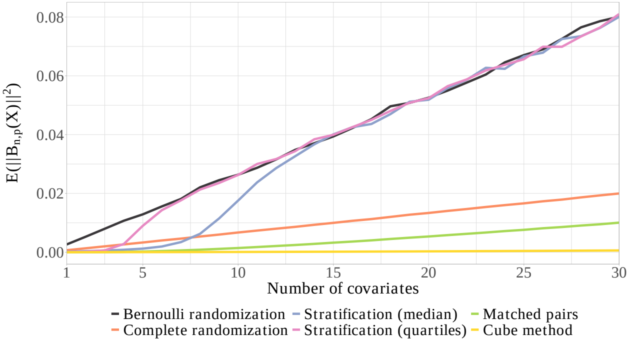
Apart from being able to balance more covariates, the cube method provides benefits across many other dimensions:

	Distribution Approach			Moment Approach		Either Approach
	Stratified	**Matched Pairs**	**Local Rand.**	**GS Walk**	**Cube Method**	**Re-randomization**
Curse of dim.	✗✗	✗✗	✗✗	✗	✗	✗✗ or ✗
C^0 **covariates**	✗	✓	✓	✓	✓	✓
Inference	✓	✓	✓	✓	✓	✗ or ✓
Prob. $\neq 1/2$	✓	✗	✓	✓	✓	✓
Het. prob.	✓	✗	✓	✓	✓	✓
No tuning par.	✓	✓	✓	✗	✓	✗

Conclusion

Whereas there is a large consensus about the importance of collecting baseline information, how this information is used varies a lot across experiments.

This paper presents a randomization design that allows to further exploit this information for precision gains and avoiding publication bias. By comparing with other methods, we introduce new nontrivial questions concerning randomization.



Link to the Paper

Thank you!

pedro.vergaramerino@ensae.fr

References I

Athey, S. and G. W. Imbens (Jan. 2017). "Chapter 3 - The Econometrics of Randomized Experimentsa". en. *Handbook of Economic Field Experiments*. Ed. by Abhijit Vinayak Banerjee and Esther Duflo. Vol. 1. Handbook of Field Experiments. North-Holland, pp. 73–140. doi: 10.1016/bs.hefe.2016.10.003.

Bai, Yuehao (Dec. 2022). "Optimality of Matched-Pair Designs in Randomized Controlled Trials". en. *American Economic Review* 112.12, pp. 3911–3940. issn: 0002-8282. doi: 10.1257/aer.20201856.

Bai, Yuehao, Joseph P. Romano, and Azeem M. Shaikh (Oct. 2022). "Inference in Experiments With Matched Pairs". *Journal of the American Statistical Association* 117.540. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/01621459.2021.1883437, pp. 1726–1737. issn: 0162-1459. doi: 10.1080/01621459.2021.1883437.

Bai, Yuehao, Azeem M. Shaikh, and Max Tabord-Meehan (May 2024). *A Primer on the Analysis of Randomized Experiments and a Survey of some Recent Advances*. arXiv:2405.03910 [econ, stat]. doi: 10.48550/arXiv.2405.03910.

References II

Bruhn, Miriam and David McKenzie (Oct. 2009). “In Pursuit of Balance: Randomization in Practice in Development Field Experiments”. en. *American Economic Journal: Applied Economics* 1.4, pp. 200–232. issn: 1945-7782. doi: 10.1257/app.1.4.200.

Bugni, Federico A., Ivan A. Canay, and Azeem M. Shaikh (Oct. 2018). “Inference Under Covariate-Adaptive Randomization”. *Journal of the American Statistical Association* 113.524. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/01621459.2017.1375934, pp. 1784–1796. issn: 0162-1459. doi: 10.1080/01621459.2017.1375934.

Chauvet, Guillaume and Yves Tillé (Mar. 2006). “A fast algorithm for balanced sampling”. en. *Computational Statistics* 21.1, pp. 53–62. issn: 1613-9658. doi: 10.1007/s00180-006-0250-2.

Chen, Jiahua and J. N. K. Rao (2007). “Asymptotic Normality Under Two-Phase Sampling Designs”. *Statistica Sinica* 17.3. Publisher: Institute of Statistical Science, Academia Sinica, pp. 1047–1064. issn: 1017-0405.

Cytrynbaum, Max (Aug. 2023). *Optimal Stratification of Survey Experiments*. arXiv:2111.08157 [econ, math, stat]. doi: 10.48550/arXiv.2111.08157.

References III

Deville, Jean-Claude and Yves Tillé (Dec. 2004). "Efficient balanced sampling: The cube method". *Biometrika* 91.4, pp. 893–912. issn: 0006-3444. doi: `10.1093/biomet/91.4.893`.

— (Feb. 2005). "Variance approximation under balanced sampling". en. *Journal of Statistical Planning and Inference* 128.2, pp. 569–591. issn: 03783758. doi: `10.1016/j.jspi.2003.11.011`.

Fisher, R. A. (1926). "The arrangement of field experiments". en. *Journal of the Ministry of Agriculture* 33. Publisher: Ministry of Agriculture and Fisheries, pp. 503–515. issn: 0368-3087. doi: `10.23637/rothamsted.8v61q`.

Fisher, Sir Ronald Aylmer (1935). *The Design of Experiments*. en. Oliver and Boyd.

Gerber, Alan et al. (July 2020). "One in a Million: Field Experiments on Perceived Closeness of the Election and Voter Turnout". en. *American Economic Journal: Applied Economics* 12.3, pp. 287–325. issn: 1945-7782. doi: `10.1257/app.20180574`.

Greevy, Robert et al. (Apr. 2004). "Optimal multivariate matching before randomization". eng. *Biostatistics (Oxford, England)* 5.2, pp. 263–275. issn: 1465-4644. doi: `10.1093/biostatistics/5.2.263`.

References IV

Hahn, Jinyong (Mar. 1998). "On the Role of the Propensity Score in Efficient Semiparametric Estimation of Average Treatment Effects". en. *Econometrica* 66.2, p. 315. issn: 00129682. doi: 10.2307/2998560.

Harshaw, Christopher et al. (June 2023). *Balancing Covariates in Randomized Experiments with the Gram-Schmidt Walk Design*. arXiv:1911.03071 [cs, math, stat]. doi: 10.48550/arXiv.1911.03071.

Higgins, Michael J., Fredrik Sävje, and Jasjeet S. Sekhon (July 2016). "Improving massive experiments with threshold blocking". *Proceedings of the National Academy of Sciences* 113.27. Publisher: Proceedings of the National Academy of Sciences, pp. 7369–7376. doi: 10.1073/pnas.1510504113.

Imai, Kosuke, Gary King, and Clayton Nall (Feb. 2009). "The Essential Role of Pair Matching in Cluster-Randomized Experiments, with Application to the Mexican Universal Health Insurance Evaluation". en. *Statistical Science* 24.1. issn: 0883-4237. doi: 10.1214/08-STS274.

References V

Li, Xinran and Peng Ding (Oct. 2017). "General Forms of Finite Population Central Limit Theorems with Applications to Causal Inference". en. *Journal of the American Statistical Association* 112.520, pp. 1759–1769. issn: 0162-1459, 1537-274X. doi: 10.1080/01621459.2017.1295865.

Li, Xinran, Peng Ding, and Donald B. Rubin (Sept. 2018). "Asymptotic theory of rerandomization in treatment–control experiments". *Proceedings of the National Academy of Sciences* 115.37. Publisher: Proceedings of the National Academy of Sciences, pp. 9157–9162. doi: 10.1073/pnas.1808191115.

Morgan, Kari Lock and Donald B. Rubin (Apr. 2012). "Rerandomization to improve covariate balance in experiments". en. *The Annals of Statistics* 40.2. arXiv:1207.5625 [math, stat]. issn: 0090-5364. doi: 10.1214/12-AOS1008.

Snyder, Christopher and Ran Zhuo (Sept. 2018). *Sniff Tests as a Screen in the Publication Process: Throwing out the Wheat with the Chaff*. en. Tech. rep. w25058. Cambridge, MA: National Bureau of Economic Research, w25058. doi: 10.3386/w25058.

References VI

Takacs, Lajos (1991). "A Moment Convergence Theorem". *The American Mathematical Monthly* 98.8. Publisher: Mathematical Association of America, pp. 742–746. issn: 0002-9890. doi: 10.2307/2324428.

Tillé, Yves (2011). "Ten years of balanced sampling with the cube method: An appraisal". en. *Survey Methodology* 37.2, pp. 215–226.

— (2022). "Some Solutions Inspired by Survey Sampling Theory to Build Effective Clinical Trials". en. *International Statistical Review* 90.3. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/insr.12498, pp. 481–498. issn: 1751-5823. doi: 10.1111/insr.12498.

Tillé, Yves and Anne-Catherine Favre (2004). "Coordination, Combination and Extension of Balanced Samples". *Biometrika* 91.4. Publisher: [Oxford University Press, Biometrika Trust], pp. 913–927. issn: 0006-3444.

— (Aug. 2005). "Optimal allocation in balanced sampling". en. *Statistics & Probability Letters* 74.1, pp. 31–37. issn: 0167-7152. doi: 10.1016/j.spl.2005.04.027.

Optimal Balancing

Let $m_0(X) = \mathbb{E}(Y(0)|X_1, ..., X_p)$ and $m_1(X) = \mathbb{E}(Y(0)|X_1, ..., X_p)$ and consider estimating the SATE or PATE.

- If m_0 and m_1 were known, a random assignment balancing $m_j(X)$ for $j = 0, 1$ will:
  - eliminate bias created by any unlucky imbalances
  - minimize the variance of the estimator
  - m_0 and m_1 are the "optimal moments" to balance on
- But m_0 and m_1 are unknown.
- Balancing some known moment functions $(f_k)_{k=1,\dots,K}$, $\mathbb{E}(f_k(X)|D = 1) = \mathbb{E}(f_k(X)|D = 0)$ such that

$$\min_{b_j} \mathbb{E}\left(\left(m_j(X) - \sum_{k=1}^{K} b_{jk} f_k(X)\right)^2\right)$$

for $j = 0, 1$ are "small" will mimic the balancing on m_0, m_1.

Optimal Balancing

Balancing on $m_0(X)$ and $m_1(X)$

- eliminates bias created by any imbalances in the sense that

$$
\begin{aligned}
\mathbb{E}\left(\widehat{\theta}_{HT} \middle| (D_i, X_i)_{i=1,\dots,n}\right) &= \frac{1}{n}\sum_{i=1}^{n} \frac{m_1(X_i)D_i}{\pi_i} - \frac{m_0(X_i)(1-D_i)}{1-\pi_i} \\
\text{(balancing} \rightarrow\text{)} \qquad &= \frac{1}{n}\sum_{i=1}^{n} m_1(X_i) - m_0(X_i) \\
&= \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} Y_i(1) - Y_i(0) \middle| (X_{i'})_{i'=1,\dots,n}\right)
\end{aligned}
$$

Optimal Balancing

Balancing on $m_0(X)$ and $m_1(X)$

- minimizes the variance of the estimator in the sense that if $Pr(D_i = 1|X_1, ..., X_n) = \pi_i$ and $(Y_i(0), Y_i(1))_{i=1,...,n} \perp\!\!\!\perp (D_i)_{i=1,...,n}|(X_i)_{i=1,...,n}$ we have:

$$\mathbb{V}\left(\widehat{\theta}_{HT}|(D_i, X_i)_{i=1,...,n}\right) = \frac{1}{n^2}\sum_{i=1}^{n} \frac{\mathbb{V}(Y_i(1)|X_i)D_i}{\pi_i^2} + \frac{\mathbb{V}(Y_i(0)|X_i)(1-D_i)}{(1-\pi_i)^2}$$

  and next by variance decomposition

$$\mathbb{V}\left(\widehat{\theta}_{HT}|(X_i)_{i=1,...,n}\right) \geq \frac{1}{n^2}\sum_{i=1}^{n} \frac{\mathbb{V}(Y_i(1)|X_i)}{\pi_i} + \frac{\mathbb{V}(Y_i(0)|X_i)}{1-\pi_i}$$

  with equality if and only if $\mathbb{E}\left(\widehat{\theta}_{HT}|(D_i, X_i)_{i=1,...,n}\right)$ does not depend on $(D_i)_{i=1,...,n}$ but only on $(X_i)_{i=1,...,n}$ which is ensured by balancing on m_0 and m_1.

Back

Balancing Constraints: an Example

We want to balance the number of female participants in the treatment and control groups.

Let $X_i = \mathbb{1}\{i \text{ is female}\}$. Individuals 1 and 2 are women and individual 3 is a man, so $(X_1, X_2, X_3) = (1, 1, 0)$.

Every unit has the same probability $\pi_i = 1/2$ of being treated.

Let $s = (s_1, s_2, s_3) \in [0,1]^3$ be any point in the unit cube. Then, the set of points satisfying the balancing constraints are:

$$\left\{ s \in [0,1]^3 \;\middle|\; \frac{1}{3}\sum_{i=1}^{3} \frac{X_i s_i}{\pi_i} = \frac{1}{3}\sum_{i=1}^{3} X_i \right\}$$

i.e.

$$\left\{ s \in [0,1]^3 \;\middle|\; s_1 + s_2 = 1 \right\}.$$

Balancing Constraints: Graphical Representation

Graphically, if we balance female units, we have

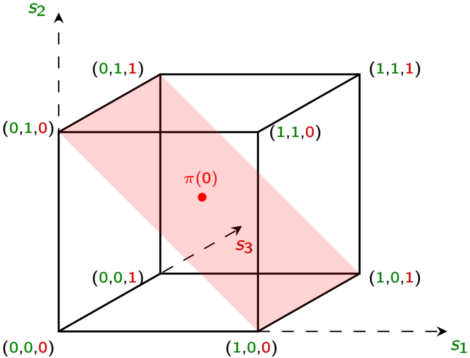


Figure 8: Balancing constraint with $\pi_i = 1/2$ and $X_1 = X_2 = 1 - X_3 = 1$

$$\text{Red Area} = \left\{ s \in [0,1]^3 \quad \middle| \quad s_1 + s_2 = 1 \right\}$$

Imperfect Balance : Graphical Representation

However, if we balance male units, we have $X_i = \mathbb{1}\{i \text{ is male}\}$ and

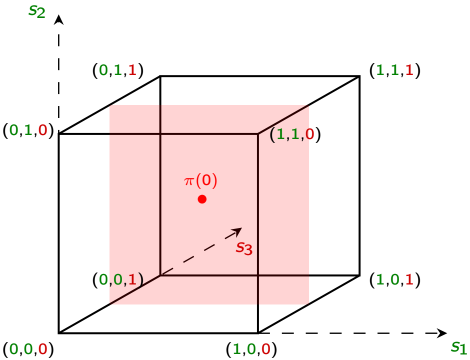
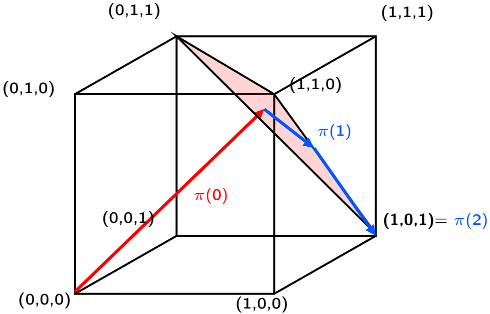


Figure 9: Balancing constraint with $\pi_i = 1/2$ and $1 - X_1 = 1 - X_2 = X_3 = 1$

$$\text{Red Area} = \left\{ s \in [0,1]^3 \quad \middle| \quad s_3 = \frac{1}{2} \right\}$$

Random Assignment with Fixed Group Size



$$\text{Red Area} = \left\{ s \in [0,1]^3 \quad \middle| \quad s_1 + s_2 + s_3 = 2 \right\}$$

Back

Proposition 1 (Balancing approximations with the cube method).

Let

$$\Delta^{\Pi}_{j,n} = \frac{1}{n}\sum_{i=1}^{n} \frac{X_{ji}D_i}{\pi_i} - \frac{X_{ji}(1-D_i)}{1-\pi_i}$$

If Assumptions 1 and 2 hold, then

$$\Delta^{Cube}_{j,n} = o_p\left(\frac{q}{\sqrt{n}}\right).$$

Moreover,

- if $\mathbb{E}\left[|X_{j1}|^r\right] < \infty$ for $r \geq 2$, then $\Delta^{Cube}_{j,n} = o_p\left(\frac{q}{n^{1-1/r}}\right)$
- if X_{j1} is sub-Gaussian, then $\Delta^{Cube}_{j,n} = O_p\left(\frac{q\sqrt{\ln(n)}}{n}\right)$
- if X_{j1} has a bounded support, then $\left|\Delta^{Cube}_{j,n}\right| < \frac{Kq}{cn}$ for K such that $|X_{j1}| < K$.

Back

Proposition 2.a (Imbalance under the cube method).

Suppose Assumption 3 holds.
Under the cube method using linear programming with positive-definite matrix M for the landing phase, we have

$$\mathbb{E}\left[||B_{n,p}(X)||^2\right] \leq 4\frac{(p+1)^2}{n^2}\frac{\lambda_{max}(M)}{\lambda_{min}(M)}$$

for $\lambda_{max}(M)$ and $\lambda_{min}(M)$ the largest and the smallest eigenvalues of M.

Back

Proposition 2.b (Imbalance under other designs).

Suppose Assumption 3 holds.

1. Under Bernoulli randomization: $\frac{4\underline{C}}{3}\frac{p}{n} \leq \mathbb{E}\left[||B_{n,p}(X)||^2\right] = \frac{4}{n}\sum_{k=1}^{p}\mathbb{E}\left[X_{k1}^2\right] \leq \frac{4\overline{C}}{3}\frac{p}{n}$
2. Under complete randomization: $\frac{\underline{C}}{3}\frac{p}{n} \leq \mathbb{E}\left[||B_{n,p}(X)||^2\right] = \frac{4}{n}\sum_{k=1}^{p}\mathbb{V}\left(X_{k1}^2\right) \leq \frac{\overline{C}}{3}\frac{p}{n}$
3. Under stratification with ℓ-quantiles:
   3.1 if $n\ell^{-p} \to \infty$: $||B_{n,p}(X)||^2 = B_1^2 + o_p\left(\frac{p}{n}\right)$, with $\frac{\underline{C}}{6\ell^2\overline{C}}\frac{p}{n}(1-o(1)) \leq \mathbb{E}\left[B_1^2\right] \leq \frac{4}{n}\sum_{k=1}^{p}\mathbb{V}(X_{k1})$
   3.2 if f $n\ell^{-p} \to 0$: $||B_{n,p}(X)||^2 = B_2^2 + o_p\left(\frac{p}{n}\right)$, with $\mathbb{E}\left[B_2^2\right] = \frac{4}{n}\sum_{k=1}^{p}\mathbb{E}\left[X_{k1}^2\right]$
4. Under matched-pairs desing: $\frac{p}{n}\left(\frac{1}{3} - \sqrt{\frac{2\ln(n-1)+4\ln\overline{C}}{p}}\right) \leq \mathbb{E}\left[||B_{n,p}(X)||^2\right] \leq \frac{4}{n}\sum_{k=1}^{p}\mathbb{V}(X_{1k})$

Back

Randomization-Based Inference

We can also perform randomization-based inference by running the cube method B times and computing $\widehat{\theta}_b$, for $b = \{1, \ldots, n\}$. Then, we define

$$\phi_n^{rand} = \mathbb{1}\left\{\left|\widehat{\theta}\right| > c_n(1-\alpha)\right\}$$

with

$$c_n(1-\alpha) = \inf\left\{t \in \mathbb{R} : \frac{1}{B}\sum_{b=1}^{B} \mathbb{1}\{\left|\widehat{\theta}_b\right| \leq t\} \geq 1-\alpha\right\}.$$

Proposition 4.

Under Assumptions 1 and 2, and the null hypothesis $H_0 : (Y_i(1), X_i) \stackrel{d}{=} (Y_i(0), X_i)$,

$$\mathbb{E}\left[\phi_n^{rand}\right] \leq \alpha.$$

Back

Sketch of Proof for Proposition 2

1. Assumptions 1, 2 and 3, and Conjecture 1 ensure that, conditional on $(X_i)_{i\geq 1}$, by the moment convergence theorem in Takacs (1991),

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} f_i + g_i D_i \xrightarrow{d} \mathcal{N}\left(0, \sigma_1^2\right)$$

   for any functions f and g such that for $f_i = f(\varepsilon_i(1), \varepsilon_i(0), X_i)$ and $g_i = g(\varepsilon_i(1), \varepsilon_i(0), X_i)$ we have $\mathbb{E}(f_i^2 + g_i^2) < \infty$ and $\mathbb{E}(f_i|X_i) + \mathbb{E}(g_i|X_i) = 0$.

2. Let us consider a function h, such that for $h_i = h(X_i)$ we have, $\frac{1}{\sqrt{n}}\sum_{i=1}^{n} h_i \xrightarrow{d} \mathcal{N}\left(0, \sigma_2^2\right)$. Then, by Theorem 2 in Chen and Rao (2007),

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} h_i + f_i + g_i D_i \xrightarrow{d} \mathcal{N}\left(0, \sigma_1^2 + \sigma_2^2\right)$$

3. Both $\sqrt{n}(\widehat{\theta}_{HT} - \theta_0)$ and $\sqrt{n}(\widehat{\theta}_{HT} - \theta_0^*)$ can be decomposed as $\frac{1}{\sqrt{n}}\sum_{i=1}^{n} h_i + f_i + g_i D_i$.

Back

Methodology

We use Gerber et al. (2020) who investigate the effect of polls on beliefs and voting behavior.

We create a superpopulation from the observed data:

- LASSO of Y on all X (+ interactions and squared values), separately for treated and control units. Gives models $f_1(.)$ and $f_0(.)$ and estimators for $\sigma_1^2 = \text{Var}(Y - f_1(X)|D = 1)$ and $\sigma_0^2 = \text{Var}(Y - f_0(X)|D = 0)$.
- Draw $(X)_{i=1,\ldots,5e4}$ with replacement from the observed data.
- Impute $Y_i(1) = f_1(X_i) + \varepsilon_i(1)$ and $Y_i(0) = f_0(X_i) + \varepsilon_i(0)$, with $(\varepsilon_1, \varepsilon_0) \sim \mathcal{N}(0, \Sigma)$ and $\Sigma = \begin{pmatrix} \widehat{\sigma}_1^2 & 0.5\widehat{\sigma}_1\widehat{\sigma}_0 \\ 0.5\widehat{\sigma}_1\widehat{\sigma}_0 & \widehat{\sigma}_0^2 \end{pmatrix}$.

We draw n observations without replacement and estimate treatment effects $10,000$ times under different allocation designs.

Back