Finely Stratified Rerandomization Designs

Max Cytrynbaum

Yale University

April 1, 2026

Introduction

Stratification and rerandomization can be used to increase the precision of treatment effect estimation in RCT's.

We show finely stratified rerandomization does **partially linear** regression adjustment "by design."

Nonparametric control over stratification covariates, linear control over rerandomization covariates. Result holds for estimation of generic causal GMM parameters.

Novel rerandomization designs. (1) Designs based on "nonlinear" imbalance criteria. (2) Minimax scheme that minimizes computational cost subject to lower bound on precision.

Inference. Asymptotically exact inference on superpopulation parameters, efficient conservative inference on finite population counterparts.

Related Literature

Stratified Randomization. Bugni et al. (2018), Bai et al. (2022), Bai (2022), Cytrynbaum (2024), Bai et al. (2024).

Rerandomization. Morgan and Rubin (2012), Li et al. (2018), Li and Ding (2020), Wang et al. (2021), Wang and Li (2022), Wang and Li (2024).

Nonlinear/Minimax Designs. Ding and Zhao (2024), Schindl and Branson (2024), Liu et al. (2023).

Finely Stratified Rerandomization

Consider $W_{1:n} = (Y_i(0), Y_i(1), \psi_i, h_i)_{i=1}^n \overset{\text{iid}}{\sim} F$, where $\psi \in \mathbb{R}^{d_\psi}$ are stratification variables and $h \in \mathbb{R}^{d_h}$ are rerandomization variables.

Want to assign proportion $p = a/k$ of n units to treatment $D_i = 1$. Use a rerandomized "matched k-tuples design."

Stratified Rerandomization.

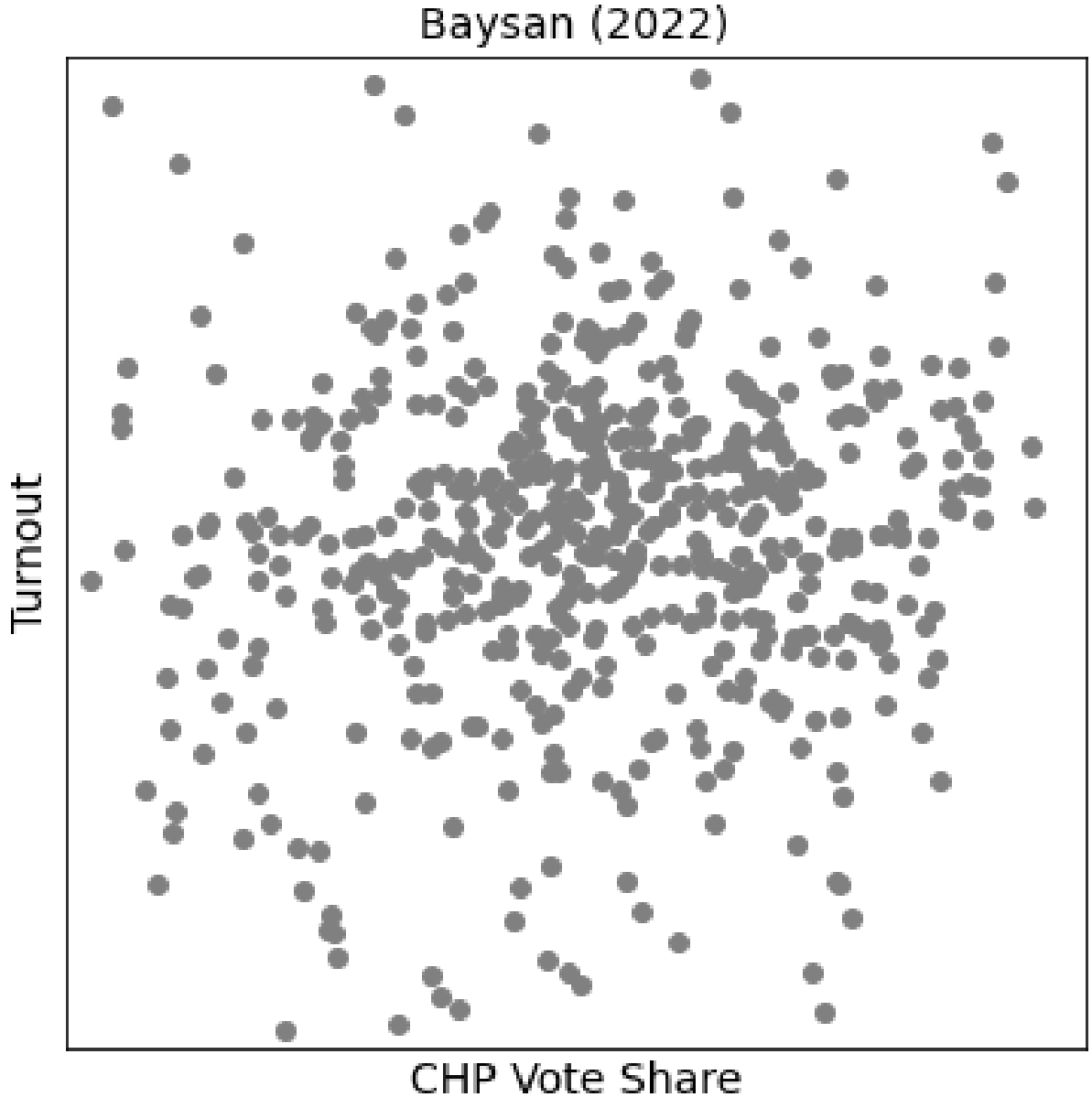
1. Match units into groups $|s| = k$ satisfying

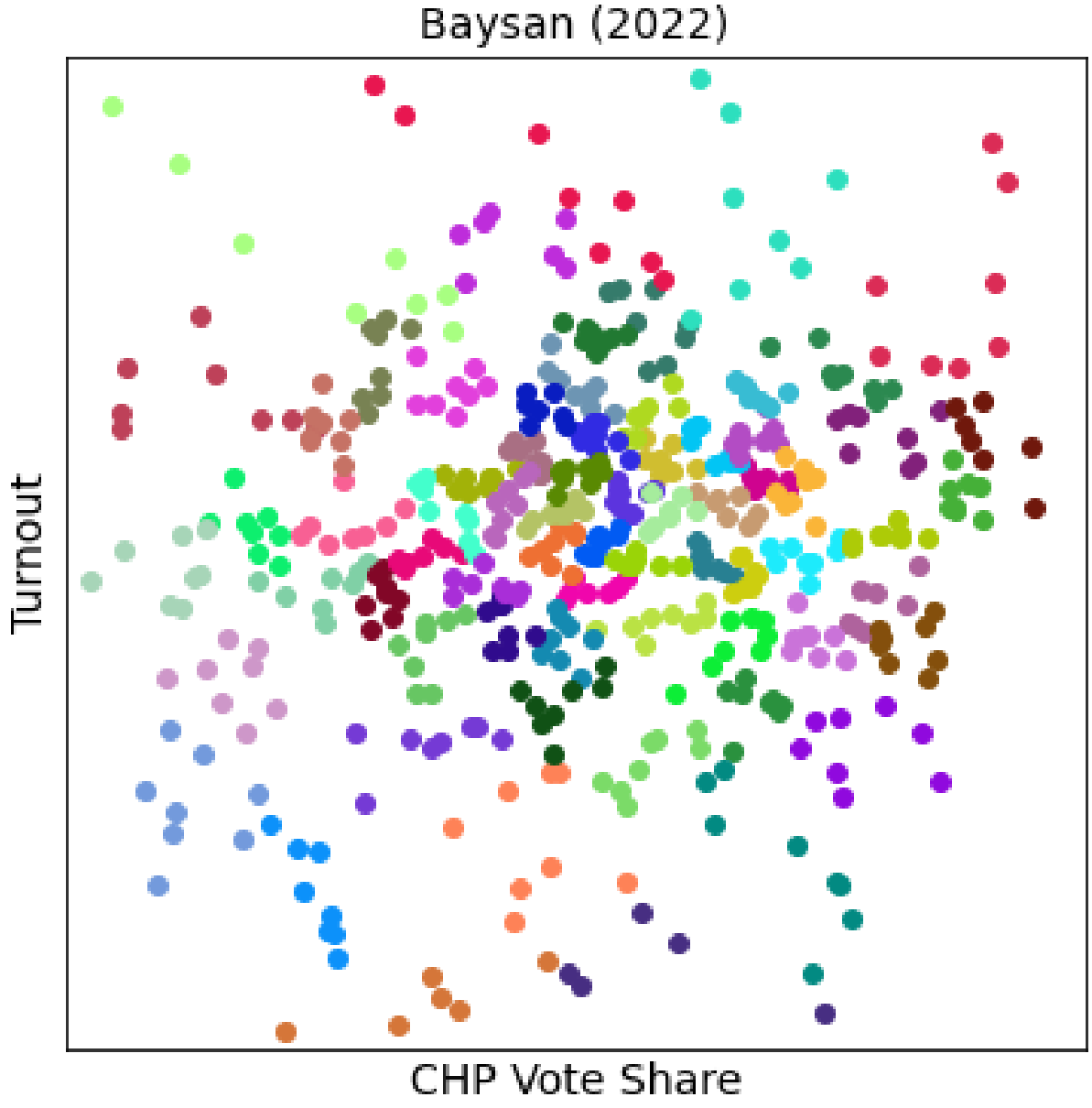
$$n^{-1} \sum_s \sum_{i,j \in s} |\psi_i - \psi_j|_2^2 = o_p(1).$$

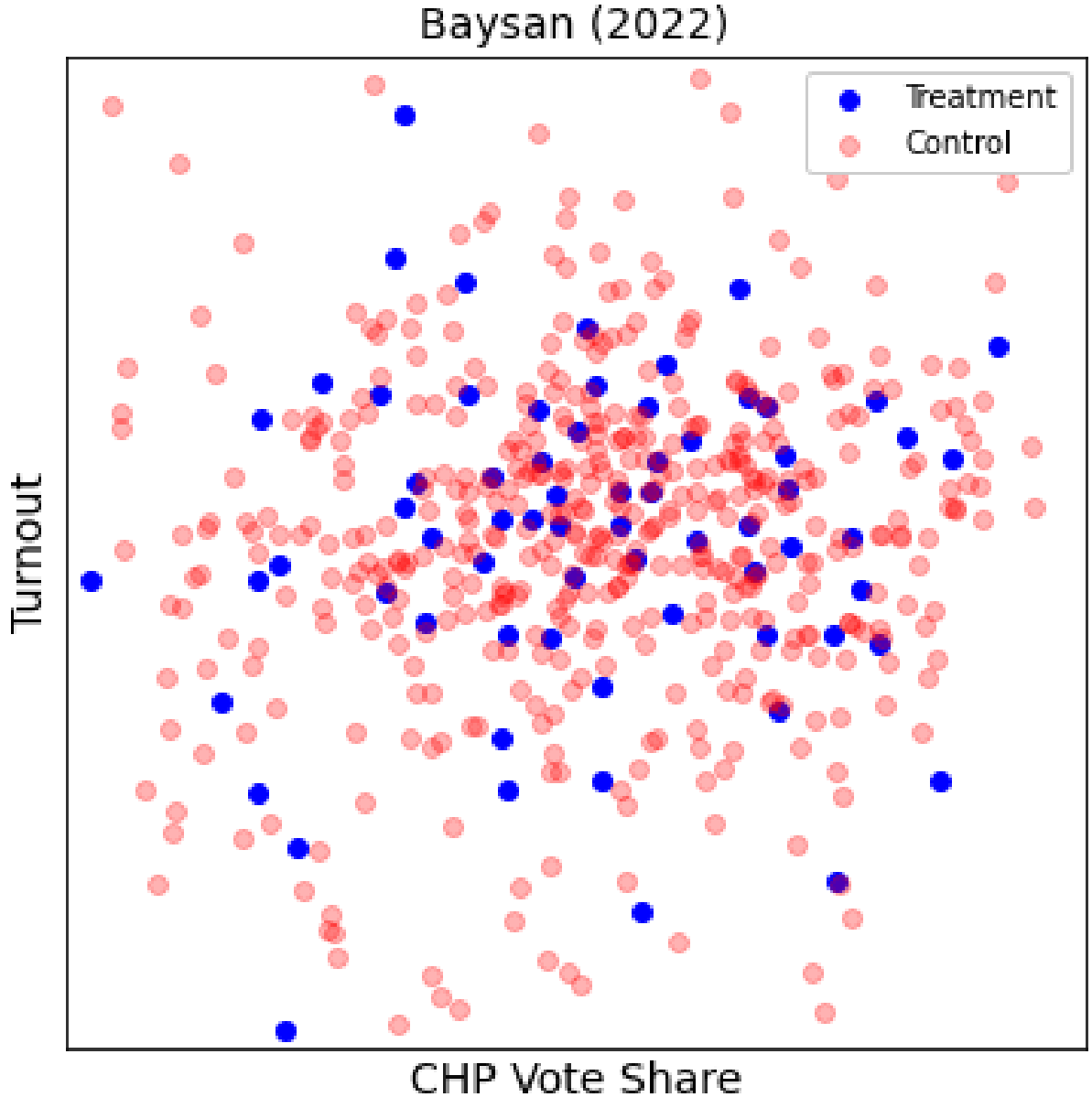
2. Randomize $p = a/k$ units to $D_i = 1$ in each group s.
3. For **acceptance region** $A \subseteq \mathbb{R}^{d_h}$, accept $D_{1:n}$ iff

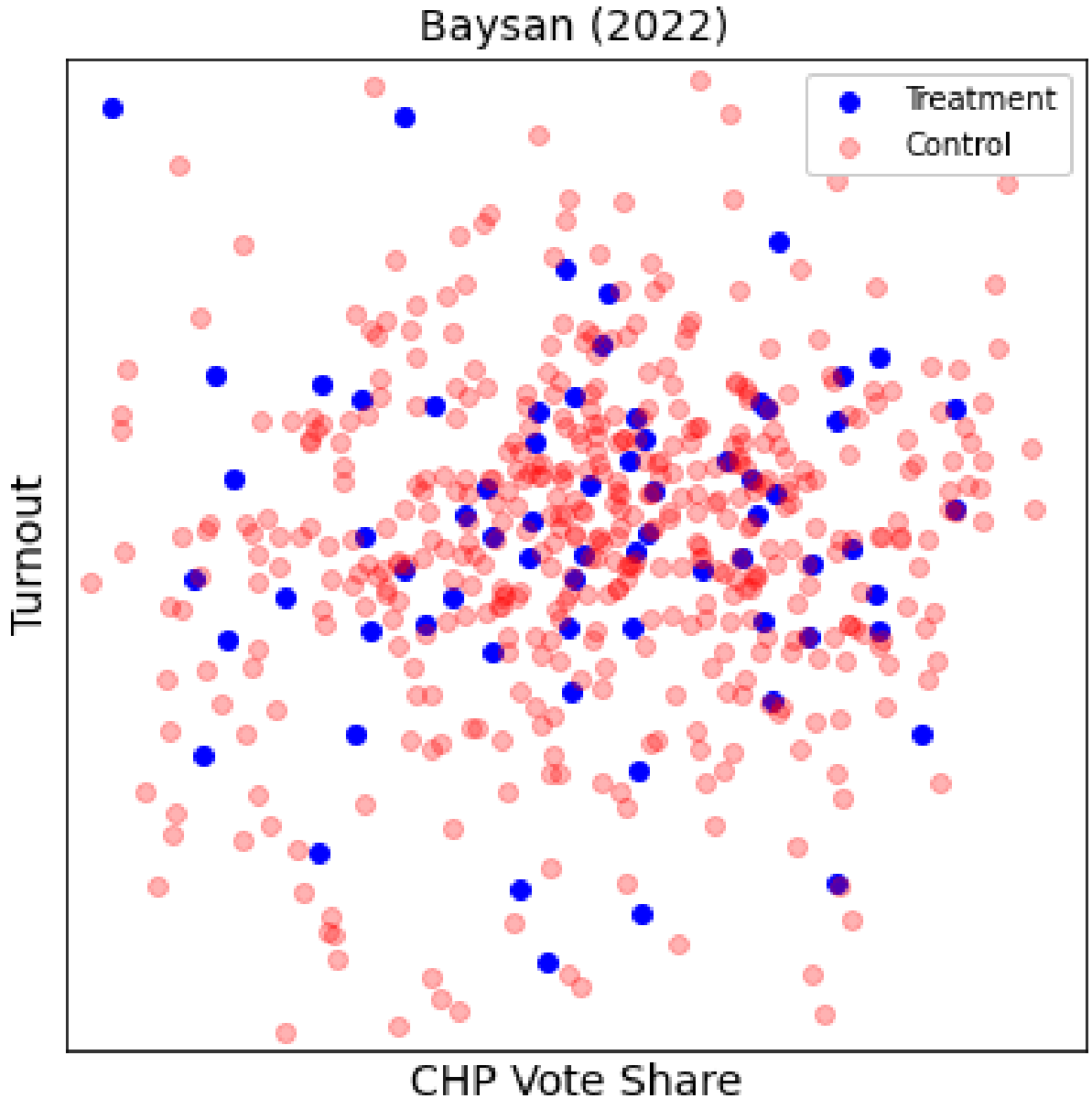
$$\sqrt{n}(\bar{h}_1 - \bar{h}_0) \in A, \quad \bar{h}_d = \frac{\sum_i h_i \mathbb{1}(D_i = d)}{\sum_i \mathbb{1}(D_i = d)}$$

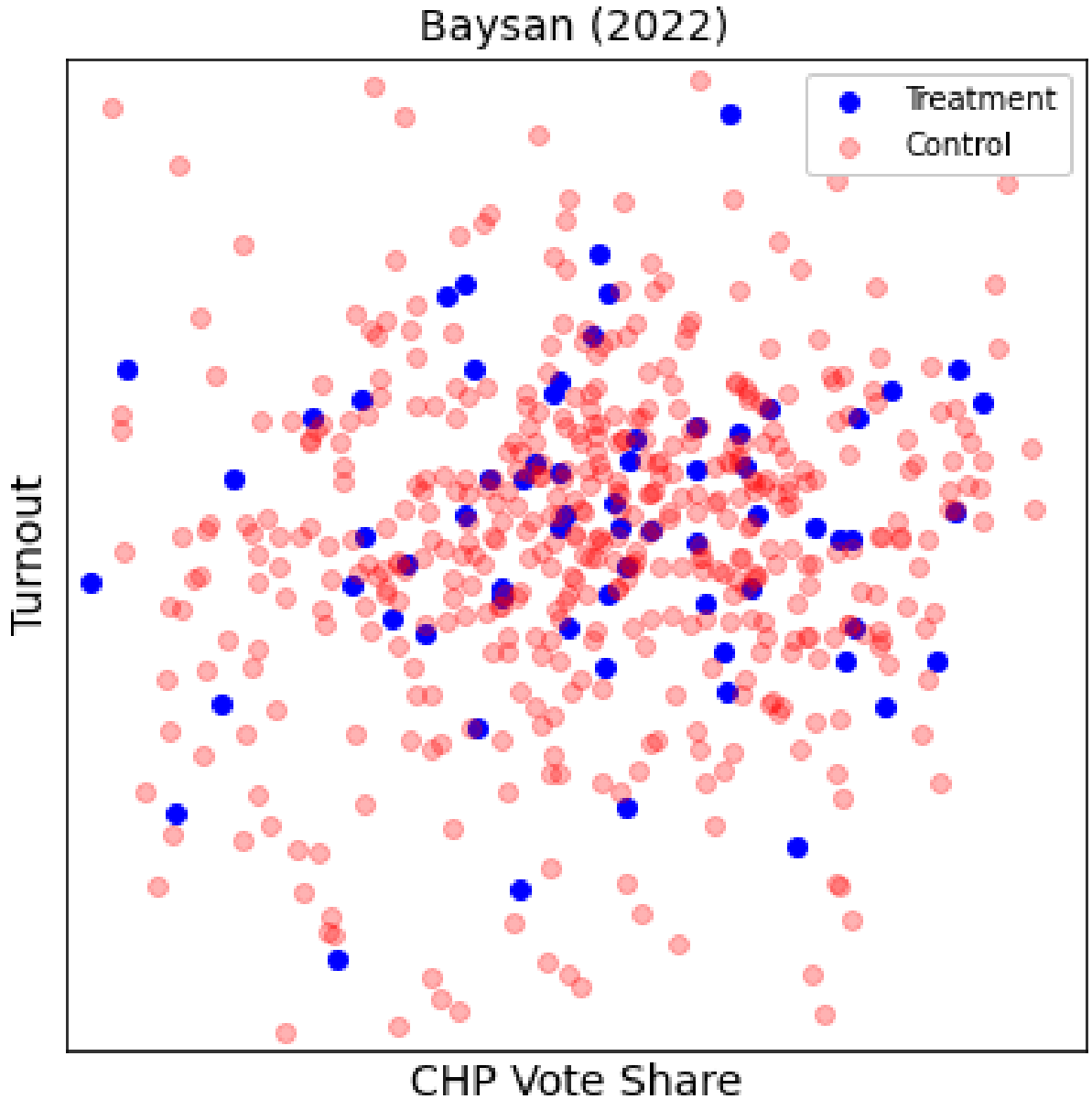
E.g. $A = B(0, \epsilon)$, check $|\sqrt{n}(\bar{h}_1 - \bar{h}_0)| \leq \epsilon$.

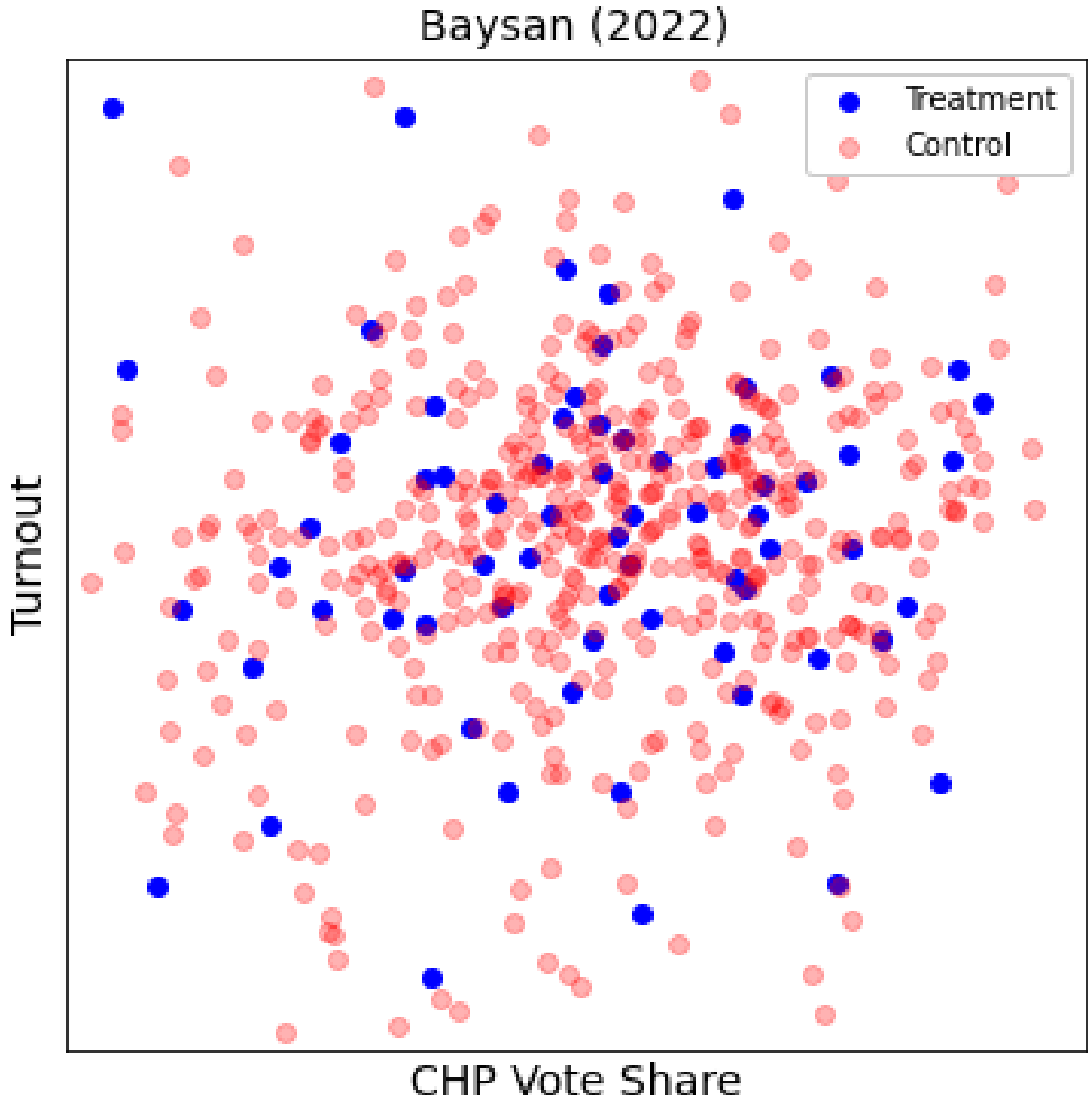

Baysan (2022)
Turnout
CHP Vote Share

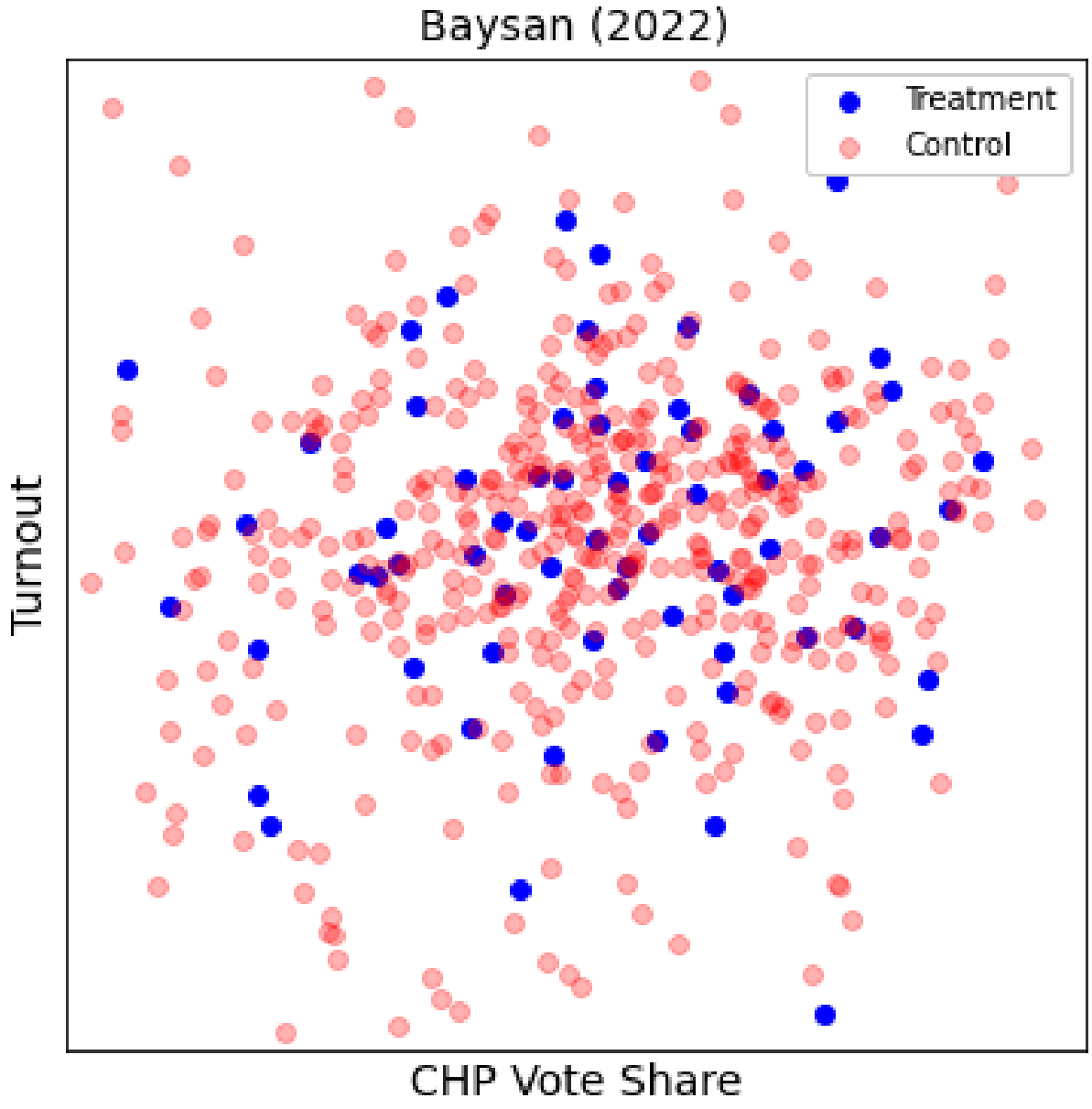

Baysan (2022)
Turnout
CHP Vote Share


Baysan (2022)
Treatment
Control
Turnout
CHP Vote Share


Baysan (2022)
Treatment
Control
Turnout
CHP Vote Share


Baysan (2022)
Treatment
Control
Turnout
CHP Vote Share


Baysan (2022)
Treatment
Control
Turnout
CHP Vote Share


Baysan (2022)
Treatment
Control
Turnout
CHP Vote Share

Motivation - Why Rerandomize?

Stratification. A recent literature in econometrics shows that fine stratification does nonparametric regression adjustment "by design."

Makes unadjusted estimator $Y \sim 1 + D$ **semiparametrically efficient**, achieving Hahn (1998) variance bound V^*_{ATE} for $\text{ATE} = E[Y(1) - Y(0)]$ and observations $(Y(0), Y(1), \psi)$.

However, rate is $n\,\text{Var}(\widehat{\theta}) = V^*_{\text{ATE}} + O(n^{-2/\dim(\psi)+1})$ due to curse of dimensionality in matching.

Goal. Finely stratify to balance small set of important variables ψ nonparametrically. Rerandomize to balance linear functions of remaining variables h.

Treatment Effect Estimation

Estimand. First, consider estimating $\text{SATE} = E_n[Y_i(1) - Y_i(0)]$, the sample average treatment effect using $\widehat{\theta}$ from $Y \sim 1 + D$.

No Stratification. Complete randomization $D_{1:n} \sim \text{C}(p)$ refers to assigning proportion p units to $D_i = 1$ uniformly at random. Obtained in our framework by setting $\psi = 1$.

For $p = a/k$ define **outcome level**:

$$\bar{Y} = (1 - p)Y(1) + pY(0).$$

Warmup. If $D_{1:n} \sim \text{C}(p)$ then $\sqrt{n}(\widehat{\theta} - \text{SATE}) \Rightarrow \mathcal{N}(0, V)$,

$$V = \frac{\text{Var}(\bar{Y})}{\text{Var}(D)}.$$

Treatment Effect Estimation

Theorem. (Pure Stratification) Let $D_{1:n}$ stratified on ψ as above. Then $\sqrt{n}(\widehat{\theta} - \text{SATE}) \Rightarrow \mathcal{N}(0, V)$,

$$V = \frac{E[\text{Var}(\bar{Y}|\psi)]}{\text{Var}(D)} = \min_{t \in L_2(\psi)} \frac{\text{Var}(\bar{Y} - t(\psi))}{\text{Var}(D)}.$$

Fine stratification provides nonparametric control over fluctuations of outcome level $\bar{Y}$ predictable by ψ.

ATE. Similarly, have $\sqrt{n}(\widehat{\theta} - \text{ATE}) \Rightarrow \mathcal{N}(0, V)$ with

$$V = \text{Var}(Y(1) - Y(0)) + \frac{E[\text{Var}(\bar{Y}|\psi)]}{\text{Var}(D)}.$$

First term reflects fluctuations of SATE $-$ ATE. Can reduce this by finely stratified sampling, e.g. see Cytrynbaum (2024).

Stratified Rerandomization

Theorem. Let $D_{1:n}$ assigned by stratified rerandomization as above. Then $\sqrt{n}(\widehat{\theta} - \text{SATE}) | W_{1:n} \Rightarrow \mathcal{N}(0, V^*) + R_A$, indep. RV's

$$V^* = \min_{\gamma \in \mathbb{R}^{d_h}} \frac{E[\text{Var}(\bar{Y} - \gamma' h | \psi)]}{\text{Var}(D)} = \min_{\substack{\gamma \in \mathbb{R}^{d_h} \\ t \in L_2(\psi)}} \frac{\text{Var}(\bar{Y} - \gamma' h - t(\psi))}{\text{Var}(D)}.$$

Stratified rerandomization does partially linear regression adjustment "by design," up to the residual imbalance R_A.

Residual Imbalance. R_A is a truncated Gaussian that arises from slackness in the balance criterion $\sqrt{n}(\bar{h}_1 - \bar{h}_0) \in A$.

For $\gamma_0 = \text{argmin}_{\gamma \in \mathbb{R}^{d_h}} E[\text{Var}(\bar{Y} - \gamma' h | \psi)]$, have $R_A = \gamma_0' Z_{hA}$ with

$$Z_{hA} = Z_h | Z_h \in A, \qquad Z_h \sim \mathcal{N}(0, \Sigma)$$

Residual Imbalance

$$\sqrt{n}(\widehat{\theta} - \text{SATE}) \Rightarrow \mathcal{N}(0, V^*) + \gamma_0' Z_{hA}$$

If A symmetric then $E[Z_{hA}] = 0$, so $\widehat{\theta}$ is asymptotically unbiased. If $p = 1/2$ and A symmetric, then $\widehat{\theta}$ is finite sample unbiased.

For criterion $|\sqrt{n}(\bar{h}_1 - \bar{h}_0)|_2 \leq \epsilon$, have $A = B(0, \epsilon)$, and $Z_{hA} = Z_h | \|Z_h\|_2 \leq \epsilon$. Computationally infeasible for small ϵ.

Choose ϵ for acceptance probability α. $\widehat{\Sigma} \xrightarrow{p} \Sigma$, randomize until

$$\|\widehat{\Sigma}^{-1/2}\sqrt{n}(\bar{h}_1 - \bar{h}_0)\|^2 \leq \epsilon(\alpha) \quad \text{for} \quad P(\chi^2_{\dim(h)} \leq \epsilon(\alpha)) = \alpha.$$

Then number of randomizations $\approx 1/\alpha$.

General Causal Parameters

Superpopulation. We can extend these results to general causal parameters defined by moment equalities,

$$E[g(D, Y, X, \theta_0)] = 0.$$

Finite Population. For experiments in a convenience sample, how do we generate a "SATE-like" version of θ_0?

Let $W = (Y(0), Y(1), X)$. Define finite population version θ_n

$$\frac{1}{n}\sum_{i=1}^{n} E[g(D_i, Y_i, X_i, \theta_n)|W_i] = 0.$$

Example. (ATE) Let $g(D, Y, \theta) = HY - \theta$ for $H_i = \frac{D_i - p}{p - p^2}$. Then we have $\theta_0 = \text{ATE}$ and $\theta_n = \text{SATE}$.

LATE Heterogeneity

Noncompliance. For instrument $Z_i \in \{0, 1\}$, endogenous $D_i(z)$, estimate treatment effect heterogeneity using $f(X, \theta)$.

Define $H_i = \frac{Z_i - p}{p - p^2}$ and score function

$$g = (HY - HD \cdot f(X, \theta))\nabla_\theta f(X, \theta)$$

By LATE reasoning, superpopulation parameter

$$\theta_0 = \underset{\theta}{\text{argmin}}\, E[(Y(1) - Y(0) - f(X, \theta))^2 | D(1) > D(0)].$$

The "SATE-like" finite population version is

$$\theta_n = \underset{\theta}{\text{argmin}}\, E_n[(Y_i(1) - Y_i(0) - f(X_i, \theta))^2 | D_i(1) > D_i(0)]$$

$f(X, \theta) = X'\theta$ gives BLP of LATE. $f(X, \theta) = \theta$ recovers LATE.

General Causal Parameters

Finite Population. $\sqrt{n}(\widehat{\theta}_{GMM} - \theta_n)|W_{1:n} \Rightarrow \mathcal{N}(0, V^*) + R_A$

$$V^* = \min_{\substack{\gamma \in \mathbb{R}^{d_h \times d_\theta} \\ t \in L_2(\psi)^m}} \text{Var}\left(\phi_a - \gamma' h - t(\psi)\right) / \text{Var}(D).$$

Influence function ϕ_a controls variance due to random assignment.

Superpopulation. $\sqrt{n}(\widehat{\theta}_{GMM} - \theta_0) \Rightarrow \mathcal{N}(0, \text{Var}(\phi_s) + V^*) + R_A$ for a "sampling" influence function ϕ_s.

Design. Form of ϕ_a suggests efficient design for estimating θ_0, θ_n.

E.g. for $\theta_0 = \text{argmin}_\theta E[(Y(1) - Y(0) - X'\theta)^2]$, have $\phi_a = E[XX']^{-1}\bar{Y}X$. Choose ψ, h predictive of $\bar{Y}X$.

Nonlinear Rerandomization

Are we missing something by using $\sqrt{n}(\bar{h}_1 - \bar{h}_0) \in A$? Want to rerandomize until distributions $h_i | D_i = 1 \approx h_i | D_i = 0$.

Example. (Density Rerandomization) For parametric density $f(x, \beta)$, estimate within-arm MLE

$$\widehat{\beta}_d \in \operatorname*{argmax}_{\beta} E_n[\log f(X_i, \beta) \,|D_i = d].$$

Rerandomize until $\sqrt{n}(\widehat{\beta}_1 - \widehat{\beta}_0) \approx 0$.

Can generalize to rerandomization based on any nonlinear estimation procedure.

GMM Rerandomization. Stratify on ψ, draw $D_{1:n}$ and estimate

$$E_n[D_i m(X_i, \widehat{\beta}_1)] = 0, \qquad E_n[(1 - D_i) m(X_i, \widehat{\beta}_0)] = 0.$$

Rerandomize until $\sqrt{n}(\widehat{\beta}_1 - \widehat{\beta}_0) \in A$.

GMM Rerandomization

$$m(X, \beta) = X - \beta \qquad \text{(linear)}$$
$$m(X, \beta) = \nabla_\beta \log f(X, \beta) \qquad \text{(density)}$$

Theorem. (GMM Rerandomization) Asymptotically equivalent to linear rerandomization $\sqrt{n}(\bar{h}_1 - \bar{h}_0) \in A$ with

$$h_i = -G^{-1} \cdot m(X_i, \beta^*), \qquad G = E[(\partial/\partial\beta')m(X_i, \beta^*)].$$

Corollary. Density rerandomization in exponential family $f(x, \beta) = \exp(\beta' r(x) - t(\beta))$ is equivalent to $h_i = r(X_i)$.

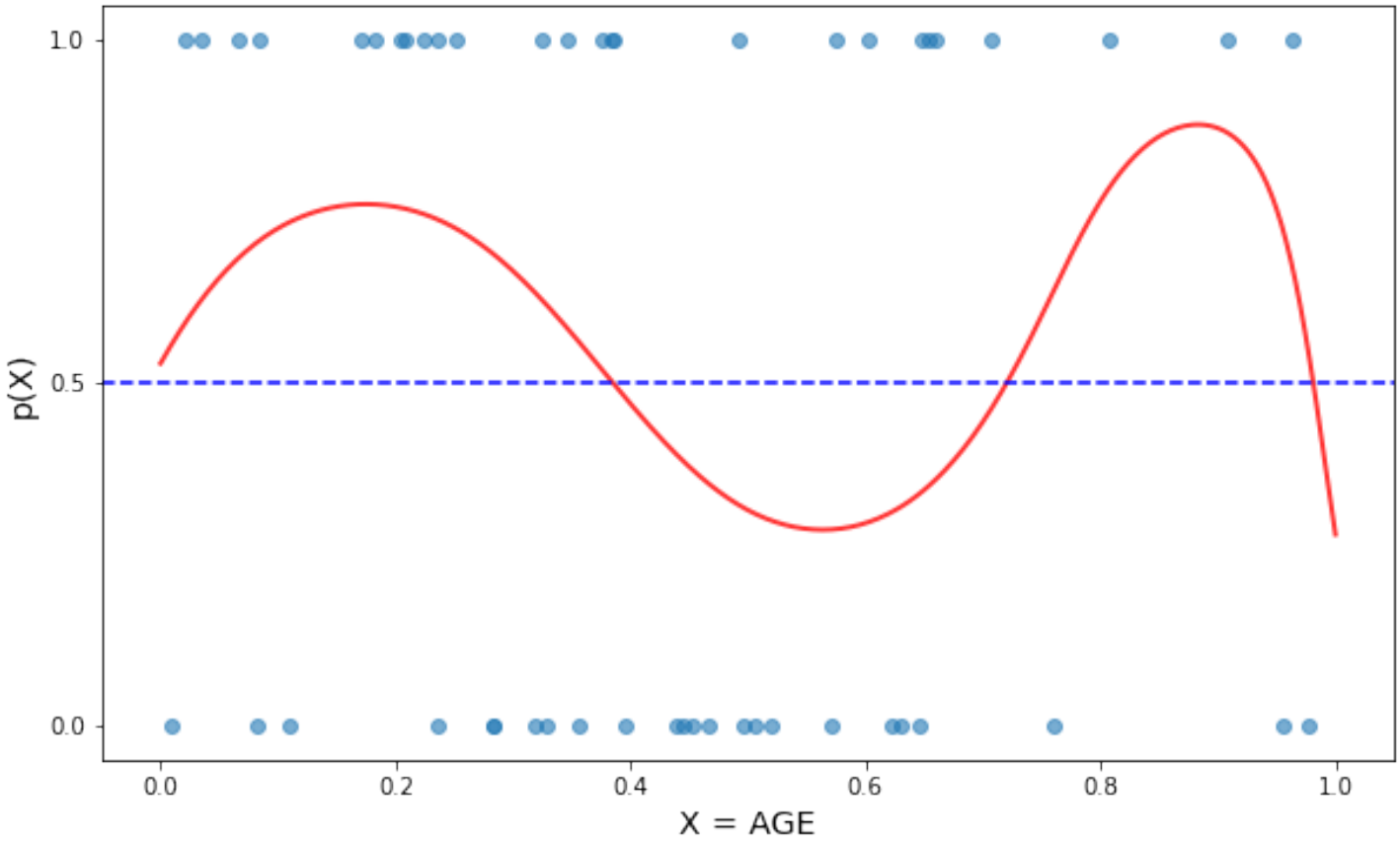
Propensity Rerandomization

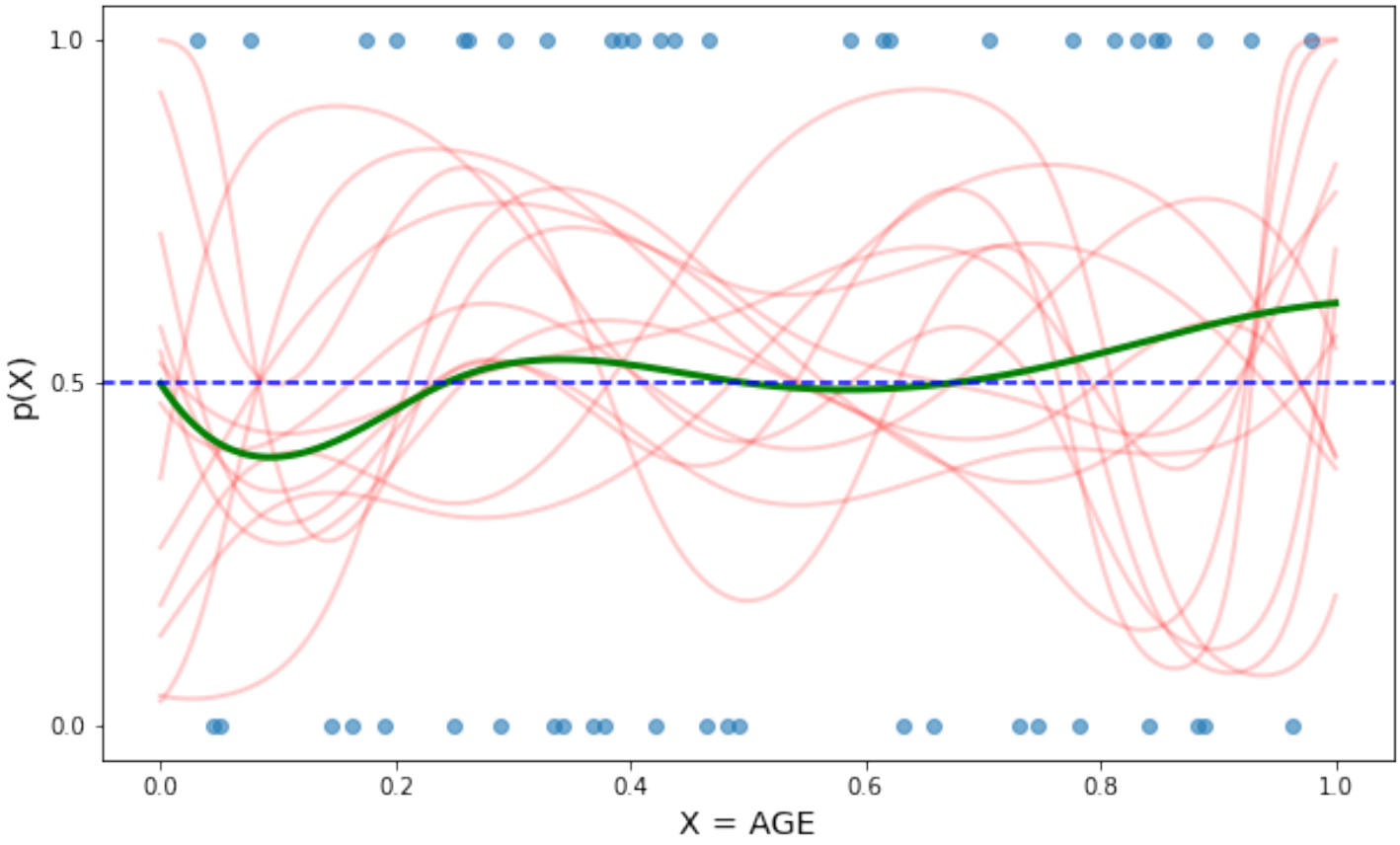
What about using a different distance between distributions?

Classifier Distance. Consider a model $\widehat{p}(X)$ to predict $D_i = 1$ or $D_i = 0$ using covariates X.

If $\widehat{p}(X)$ is unable to predict $D_i = 1$ vs. $D_i = 0$, this is evidence that $X_i|D_i = 1$ and $X_i|D_i = 0$ are close.

Propensity Rerandomization. Rerandomize until X has no predictive power for D according to the propensity $\widehat{p}(X)$.


1.0
0.5
0.0
p(X)
0.0
0.2
0.4
0.6
0.8
1.0
X = AGE


1.0
0.5
0.0
p(X)
0.0
0.2
0.4
0.6
0.8
1.0
X = AGE

Propensity Rerandomization

MLE. For link function $L \in [0,1]$ and $R = (1, X)$, define

$$\widehat{\beta} \in \operatorname*{argmax}_{\beta \in \mathbb{R}^m} E_n[D_i \log L(R_i'\beta) + (1 - D_i) \log(1 - L(R_i'\beta))].$$

Propensity Rerandomization. For $\widehat{p}(X) = L(R_i'\widehat{\beta})$, stratify on ψ and rerandomize until the balance criterion

$$n \cdot E_n[(\widehat{p}(X_i) - p)^2] \leq \epsilon^2.$$

Theorem. Equivalent to rerandomizing until $\sqrt{n}(\bar{h}_1 - \bar{h}_0) \in A$

$$h = \operatorname{Var}(X)^{-1/2} X, \qquad A = B(0, \epsilon / \operatorname{Var}(D)).$$

Result. General rerandomization criteria based on GMM or propensity score estimation equivalent to linear rerandomization with implicit choices of h and A.

Optimizing Acceptance Regions

$$\sqrt{n}(\widehat{\theta} - \text{SATE}) \Rightarrow \mathcal{N}(0, V^*) + \gamma_0' Z_{hA}$$

Optimal Acceptance. How to choose A so $\gamma_0' Z_{hA}$ is small?

Only care about residual imbalances Z_{hA} aligned with γ_0, where $\gamma_0 = \text{argmin}_{\gamma \in \mathbb{R}^{d_h}} E[\text{Var}(\bar{Y} - \gamma' h | \psi)]$.

Oracle Design. Suggests rerandomizing until

$$|\sqrt{n}(\bar{h}_1 - \bar{h}_0)' \gamma_0| \leq \epsilon.$$

Infeasible since γ_0 is unknown at design-time.

Minimax Rerandomization

Minimax. Consider prior belief set $B \subseteq \mathbb{R}^{d_h}$. Rerandomize until:

$$\sup_{\gamma \in B} |\gamma' \sqrt{n}(\bar{h}_1 - \bar{h}_0)| \leq \epsilon.$$

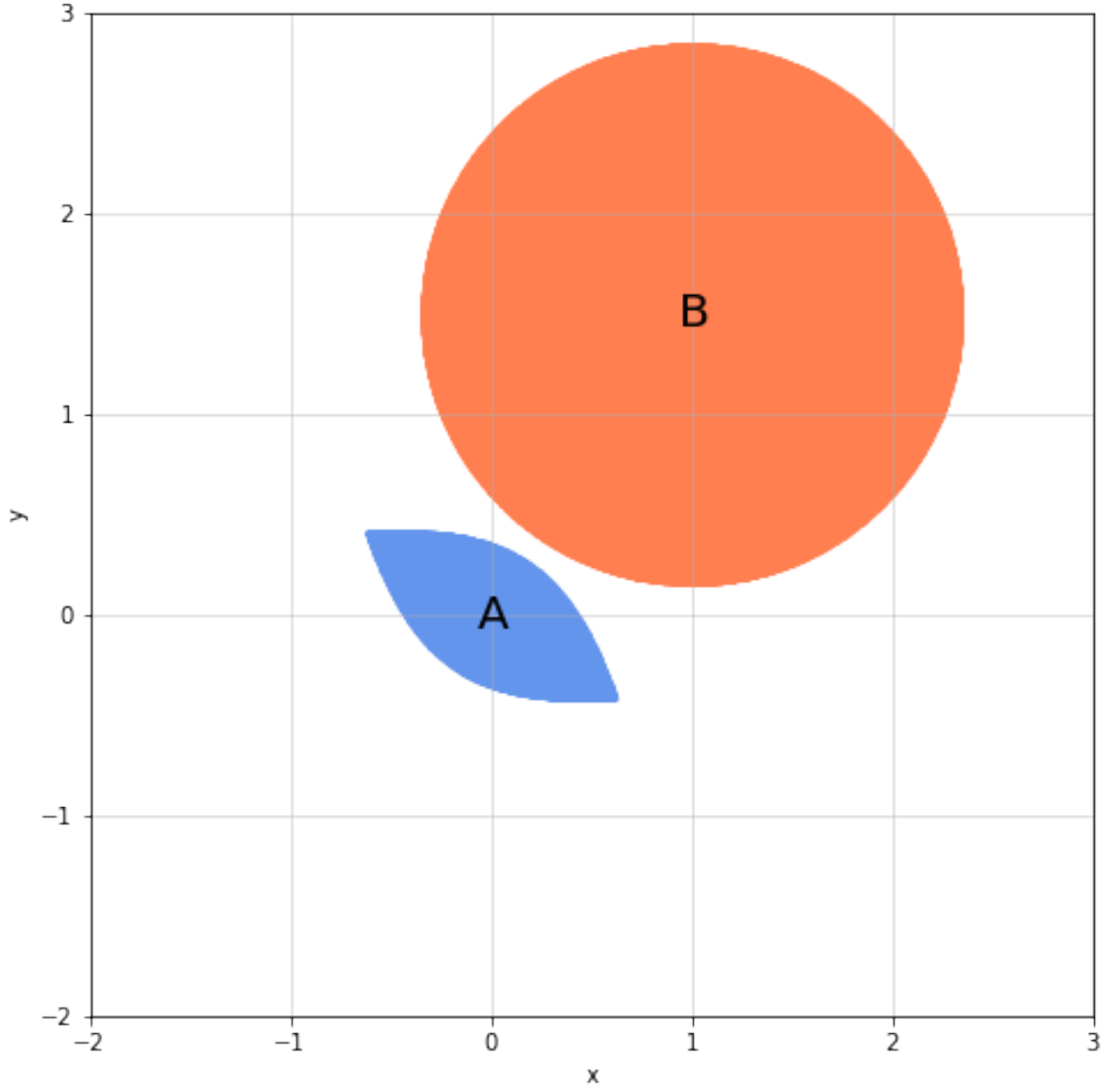
Equivalently, define convex penalty $\text{pen}_B(x) = \sup_{\gamma \in B} |\gamma' x|$ and accept if $\text{pen}_B(\sqrt{n}(\bar{h}_1 - \bar{h}_0)) \leq \epsilon$, so

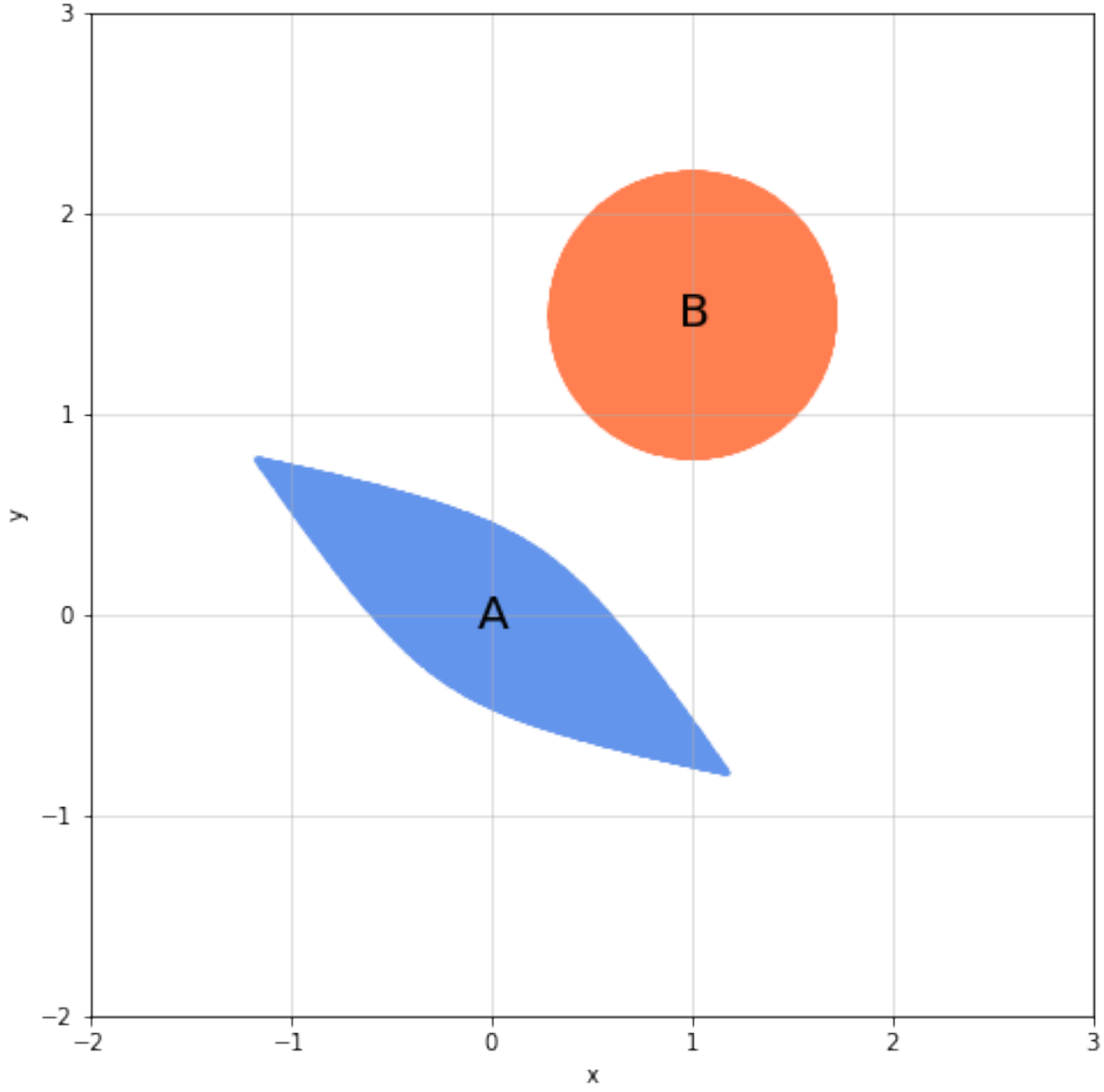
$$A = \{x : \text{pen}_B(x) \leq \epsilon\}.$$

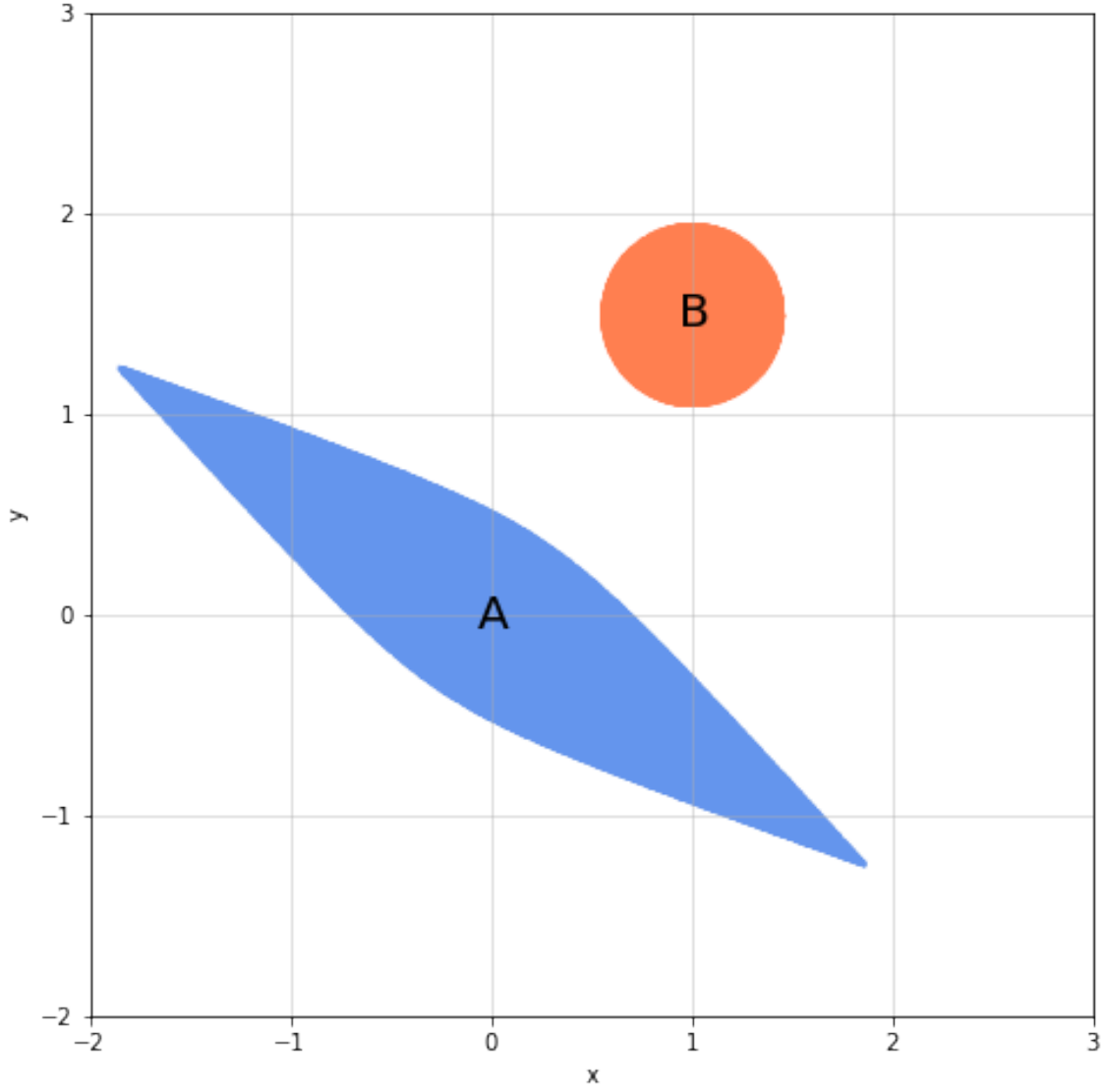
Proposition. (Acceptance Region) A is symmetric and convex. If B is bounded, then A is closed with non-empty interior.

Example. For guess $\gamma_0 \approx \bar{\gamma}$ and uncertainty δ, set $B = B_2(\bar{\gamma}, \delta)$,

$$\text{pen}_B(x) = |x'\bar{\gamma}| + \delta|x|_2, \quad A = \{x : \text{pen}_B(x) \leq \epsilon\}.$$


B
A
3
2
1
0
−1
−2
−2
−1
0
1
2
3
x
y


B
A
3
2
1
0
−1
−2
−2
−1
0
1
2
3
x
y


B
A
3
2
1
0
−1
−2
−2
−1
0
1
2
3
x
y

Minimax Rerandomization

Define family of possible limiting distributions for $\sqrt{n}(\widehat{\theta} - \text{SATE})$ by

$$L_{\gamma A} = \mathcal{N}(0, V(\gamma)) + \gamma' Z_{hA}. \quad \gamma \in \mathbb{R}^{d_h}, A \subseteq \mathbb{R}^{d_h}.$$

Computational Cost. For acceptance probability $P(Z_h \in A)$, expected randomizations until acceptance is $P(Z_h \in A)^{-1}$.

Theorem. The region $A_0 = \{x : \text{pen}_B(x) \leq \epsilon\}$ has

$$A_0 = \underset{A \subseteq \mathbb{R}^{d_h}}{\text{argmax}}\, P(Z_h \in A) \quad \text{s.t.} \quad \sup_{\gamma \in B} |\,\text{bias}(L_{\gamma A} | Z_{hA})| \leq \epsilon.$$

Corollary. If $\gamma_0 \in B$, then $\text{Var}(L_{\gamma_0 A_0}) \leq V^* + \epsilon^2$, for V^* the partially linear variance.

Intuitively, use prior information about outcome model to choose shape of A, form of dimension reduction.

Restoring Normality

We had $\sqrt{n}(\widehat{\theta}_{GMM} - \theta_n)|W_{1:n} \Rightarrow \mathcal{N}(0, V^*) + R_A$.

Adjustment. Consider the adjusted GMM estimator

$$\widehat{\theta}_{adj} = \widehat{\theta}_{GMM} - \widehat{\gamma}'(\bar{h}_1 - \bar{h}_0).$$

Theorem. Suppose $\widehat{\gamma} \xrightarrow{p} \gamma_0 = \text{argmin}_\gamma E[\text{Var}(\phi_a - \gamma' h|\psi)]$. Then for any $A \subseteq \mathbb{R}^{d_h}$, $\sqrt{n}(\widehat{\theta}_{adj} - \theta_n) \Rightarrow \mathcal{N}(0, V^*)$.

Rerandomization and optimal ex-post adjustment are equivalent in first order asymptotics.

For larger dim(h), combining rerandomization and adjustment provides meaningful finite sample efficiency gains, due to a novel double robustness property.

Rerandomization + Adjustment

Consider the case without stratification and $\theta_n = \text{SATE}$, and expand

$$\bar{Y} = c + \gamma_0' h + e \qquad e \perp (1, h).$$

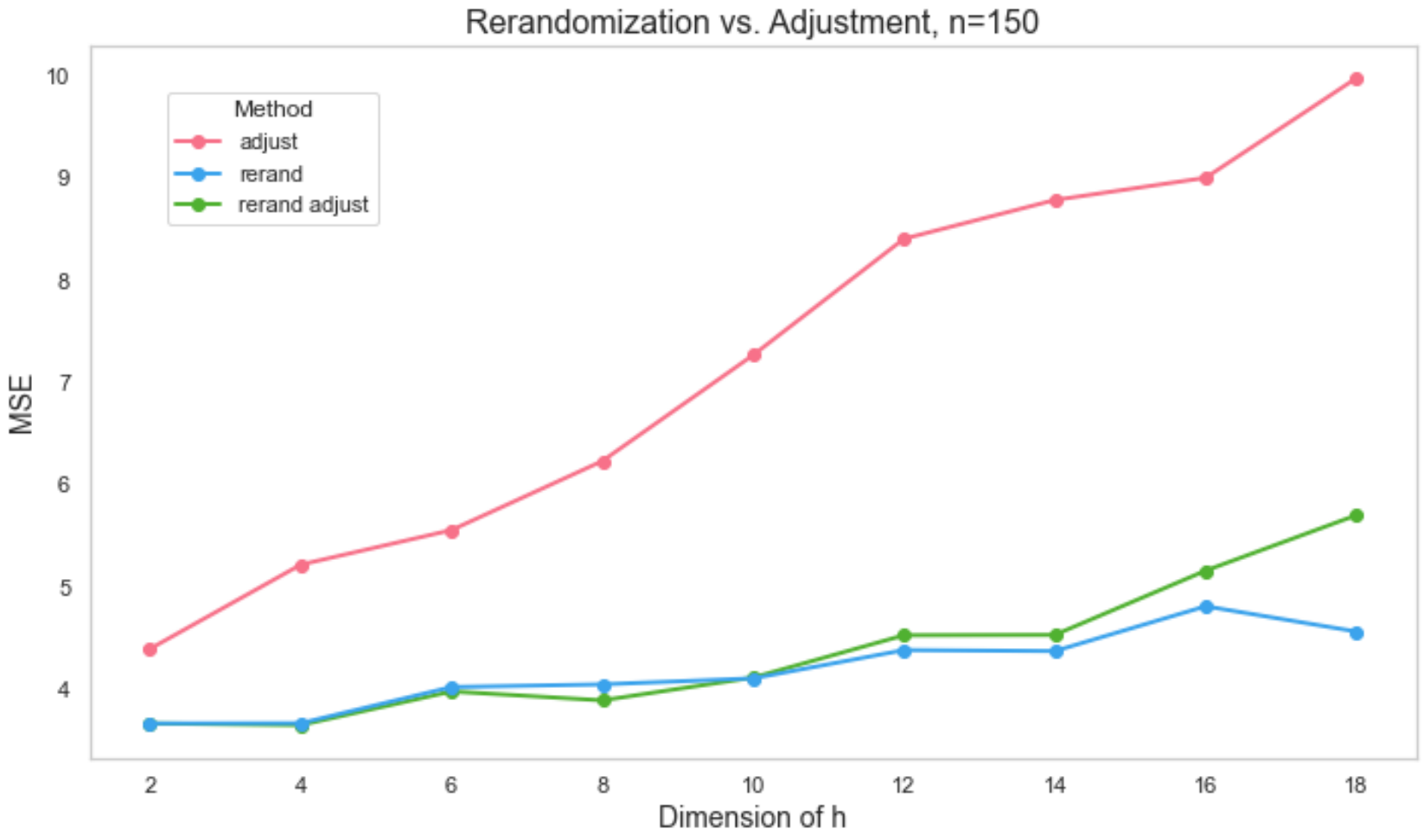
Simple algebra shows that

$$\widehat{\theta} - \text{SATE} = \gamma_0'(\bar{h}_1 - \bar{h}_0) + (\bar{e}_1 - \bar{e}_0)$$

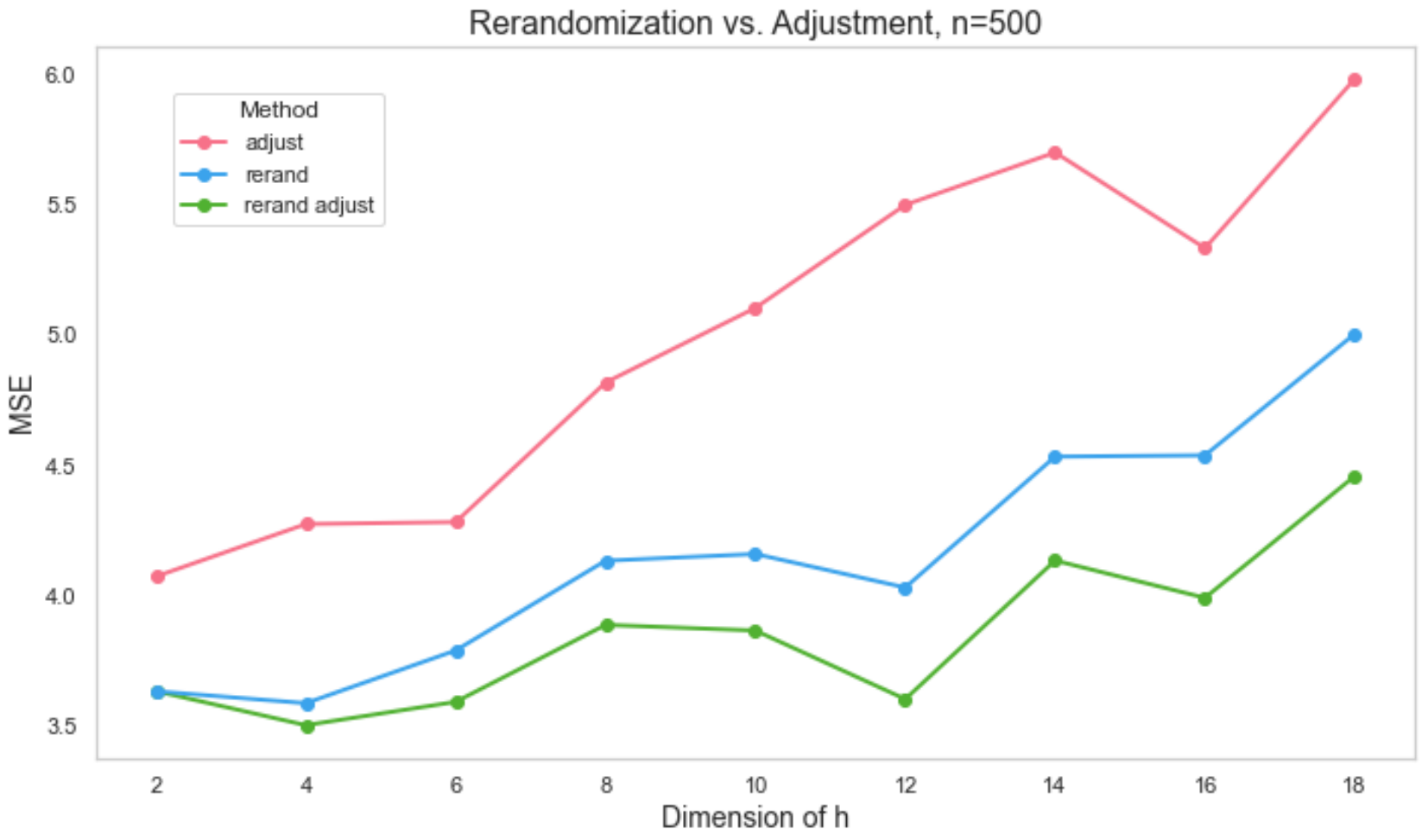
Double Robustness. For $\widehat{\theta}_{adj} = \widehat{\theta} - \widehat{\gamma}'(\bar{h}_1 - \bar{h}_0)$, have

$$\widehat{\theta} - \text{SATE} = (\gamma_0 - \widehat{\gamma})'(\bar{h}_1 - \bar{h}_0) + (\bar{e}_1 - \bar{e}_0)$$

Efficiency


Rerandomization vs. Adjustment, n=150
Method
adjust
rerand
rerand adjust
MSE
10
9
8
7
6
5
4
2
4
6
8
10
12
14
16
18
Dimension of h

Efficiency


Rerandomization vs. Adjustment, n=500
Method
adjust
rerand
rerand adjust
MSE
6.0
5.5
5.0
4.5
4.0
3.5
2
4
6
8
10
12
14
16
18
Dimension of h

Inference

For $\widehat{\theta}_{adj} = \widehat{\theta}_{GMM} - \widehat{\gamma}'(\bar{h}_1 - \bar{h}_0)$, recover $\sqrt{n}(\widehat{\theta}_{adj} - \theta_n) \Rightarrow \mathcal{N}(0, V^*)$.

The variance V^* is not identified. For $\theta_n = \text{SATE}$, $D_{1:n} \sim \text{C}(p)$

$$V^* = \frac{\text{Var}(\bar{Y})}{\text{Var}(D)} \propto \text{Cov}(Y(1), Y(0))$$

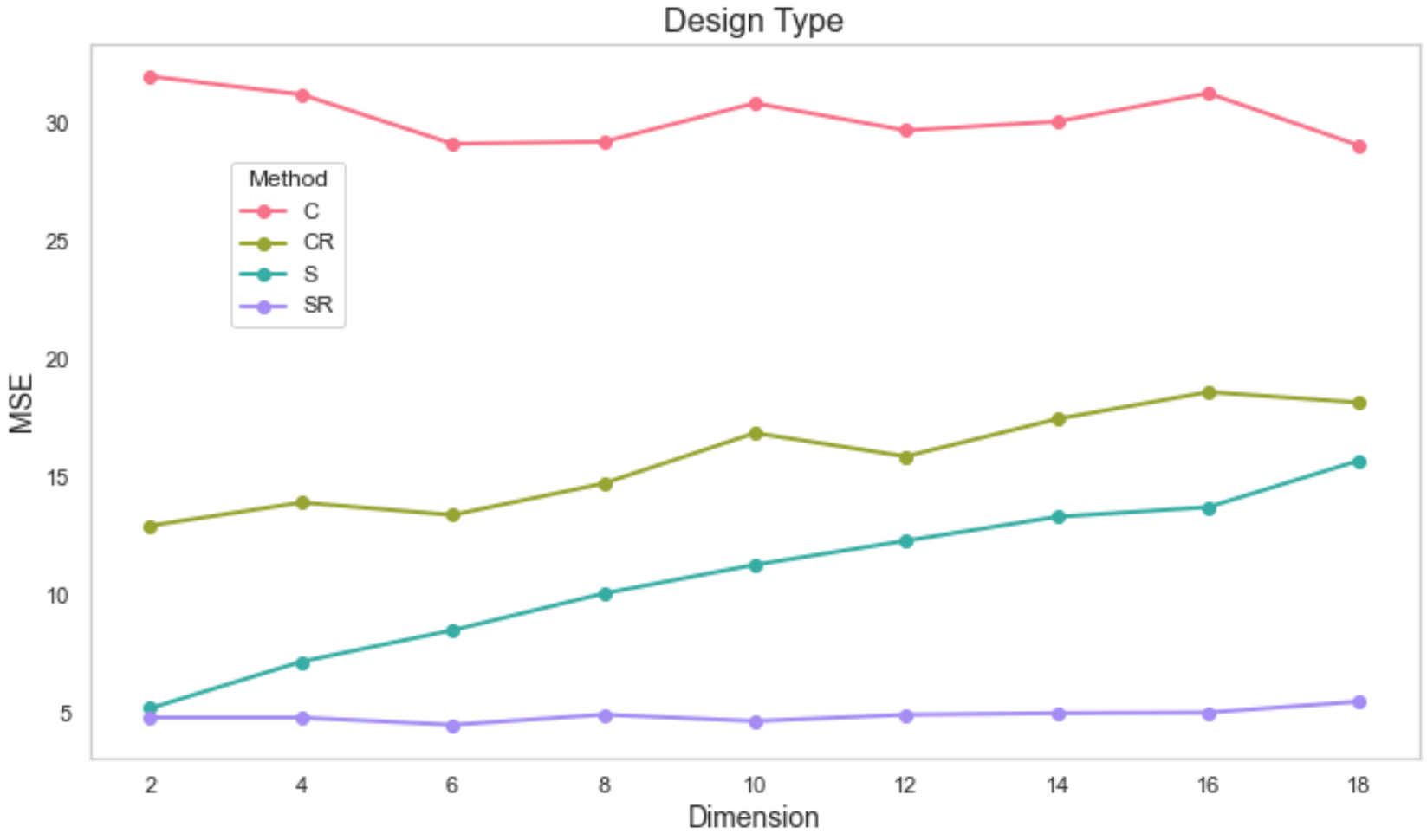
Define $\sigma_d^2 = \text{Var}(Y(d))$. Neyman (1990) showed bounds

$$V^* \leq \frac{\sigma_1^2}{p} + \frac{\sigma_0^2}{1-p} - (\sigma_1 - \sigma_0)^2 \leq \frac{\sigma_1^2}{p} + \frac{\sigma_0^2}{1-p}.$$

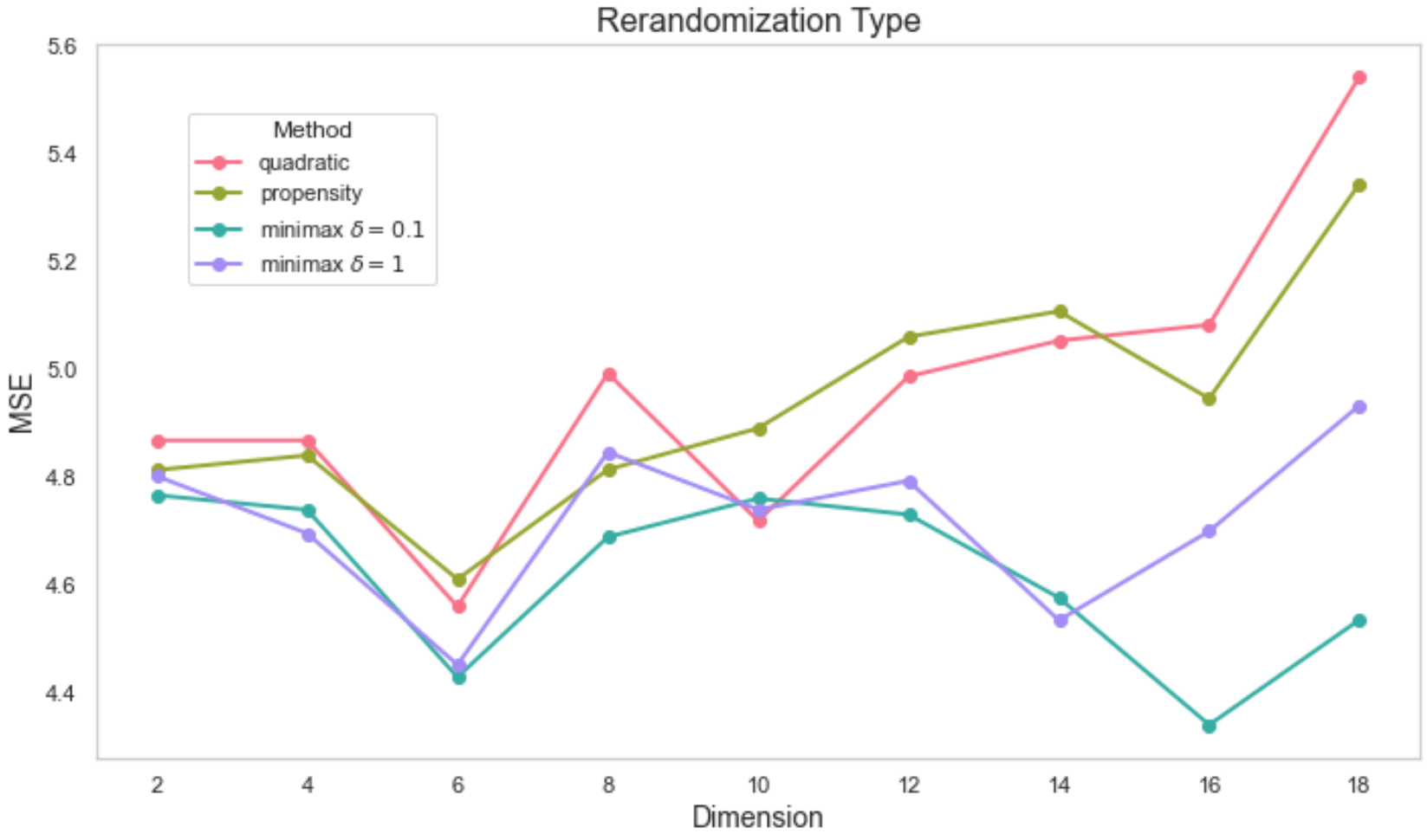
We extend sharper Neyman bound to GMM parameters θ_n, under stratified rerandomization with adjustment, finding $\widehat{V} \xrightarrow{p} \bar{V} \geq V^*$.

Superpopulation. Asymptotically exact inference on θ_0. Both use "collapsed strata" type methods to deal with small strata.

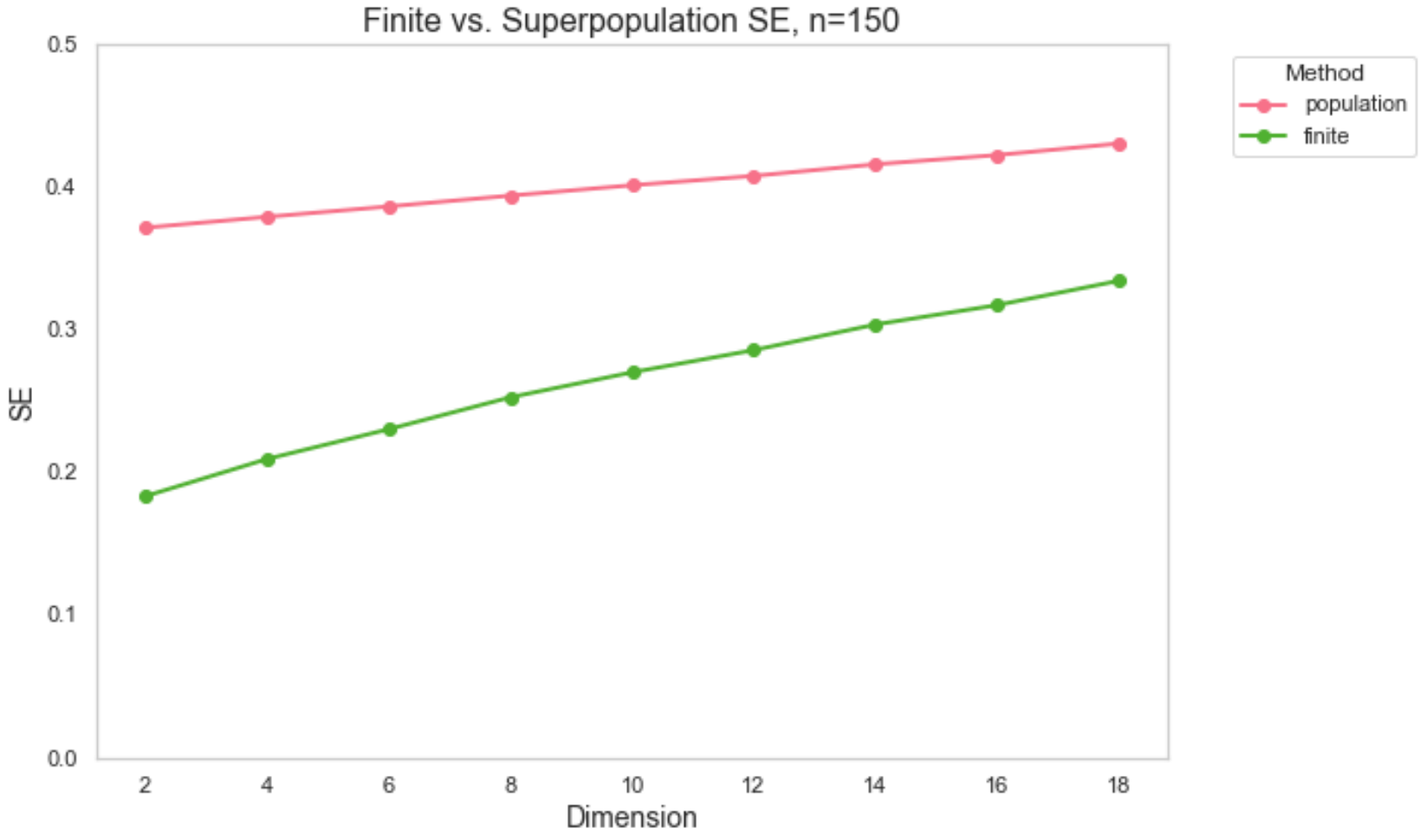
Efficiency


Design Type
Method
C
CR
S
SR
MSE
30
25
20
15
10
5
2
4
6
8
10
12
14
16
18
Dimension

Efficiency


Rerandomization Type
Method
quadratic
propensity
minimax $\delta = 0.1$
minimax $\delta = 1$
MSE
5.6
5.4
5.2
5.0
4.8
4.6
4.4
2
4
6
8
10
12
14
16
18
Dimension

Inference


Finite vs. Superpopulation SE, n=150
SE
0.0
0.1
0.2
0.3
0.4
0.5
2
4
6
8
10
12
14
16
18
Dimension
Method
population
finite

Empirical Application

Angrist et al. (2013), "When Opportunity Knocks, Who Answers?".

Effect of eligibility to receive incentives $Z \in \{0, 1\}$ for good grades on college GPA, graduation, other outcomes.

Treatment $D = 1$ if student actually engaged with program.

Measure covariates like HS GPA, previous college GPA, mother's education, financial need, native language.

Impute missing potential outcomes $\widehat{Y}_i(d)$. Given this full panel $(\widehat{Y}_i(0), \widehat{Y}_i(1))$, simulate various designs.

Estimate LATE and BLP of LATE

$$\min_{\theta} E_n[(Y_i(1) - Y_i(0) - X_i'\theta)^2 | D_i(1) > D_i(0)]$$

Efficiency for BLP of LATE

		MSE		Cover		CI Width	
θ_n (LATE)	Design	$\widehat{\theta}$	$\widehat{\theta}_{adj}$	Pop.	Fin.	Pop.	Fin.
CLATE (GPA)	C	3.06	1.00	0.92	0.98	1.00	1.02
	CR	1.76	0.85	0.95	0.99	0.97	0.97
	S	1.42	0.98	0.94	0.98	0.97	1.01
	SR	1.07	0.89	0.94	0.99	0.97	0.99
	F	0.86	0.92	0.97	0.98	1.10	0.97
	FR	0.79	0.83	0.97	0.99	1.05	0.94
	F+	1.41	1.44	0.96	0.95	1.39	1.32
	FR+	1.32	1.34	0.96	0.97	1.38	1.32

Table: LATE Parameters

Summary

Stratified rerandomization does partially linear regression adjustment "by design."

Results hold for generic causal parameters in a GMM framework.

"Nonlinear" rerandomization designs (e.g. density-based) asymptotically equivalent to linear rerandomization.

Minimax scheme optimizing shape of the acceptance region.

Optimal ex-post linear adjustment restores asymptotic normality, enabling simple and efficient inference methods.

Temporary page!

LaTeX was unable to guess the total number of pages correctly. A
there was some unprocessed data that should have been added t
the final page this extra page has been added to receive it.
If you rerun the document (without altering it) this surplus page
will go away, because LaTeX now knows how many pages to expe
for this document.