Matrix Constraints for Multilevel Observational Studies

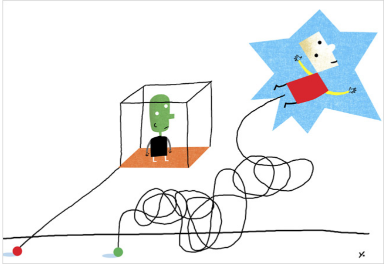
Eli Ben-Michael
(UC Berkeley)

(joint work with Avi Feller)

UAI 2018 Causal Workshop
August 10, 2018

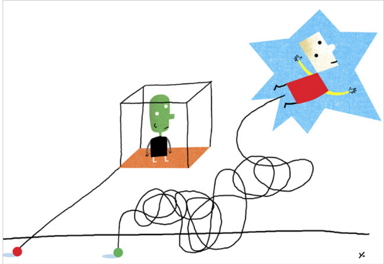
Multi-Site Trials

- National Study of Learning Mindsets [Yeager, 2017]
  - Cheap intervention
  - $>$10,000 students across 76 schools
  - Individual-level features
  - School-level features



Multi-Site Trials

- National Study of Learning Mindsets [Yeager, 2017]
  - Cheap intervention
  - $>$10,000 students across 76 schools
  - Individual-level features
  - School-level features



- Multi-Site Randomized Control Trials
  - Interventions at *many* sites [Raudenbush and Bloom, 2015]
  - Overall and site-specific treatment effects
  - Multilevel structure: individual and site-level features
  - Many analysis approaches including multilevel modeling [Feller and Gelman, 2015]

But what about observational studies?

How do we account for selection bias
and use this multilevel structure to:

But what about observational studies?

How do we account for selection bias
and use this multilevel structure to:

Estimate the overall effect?

But what about observational studies?

How do we account for selection bias
and use this multilevel structure to:

Estimate the overall effect?
Estimate school-level effects?

But what about observational studies?

How do we account for selection bias
and use this multilevel structure to:

Estimate the overall effect?
Estimate school-level effects?

This talk: Extend balancing weights to multilevel settings

But what about observational studies?

How do we account for selection bias
and use this multilevel structure to:

Estimate the overall effect?
Estimate school-level effects?

This talk: Extend balancing weights to multilevel settings

- Trading off global/subgroup balance = partial pooling

But what about observational studies?

How do we account for selection bias
and use this multilevel structure to:

Estimate the overall effect?
Estimate school-level effects?

This talk: Extend balancing weights to multilevel settings

- Trading off global/subgroup balance = partial pooling

Data: Obs. study simulation from real RCT for ACIC workshop

Exact Balancing Weights

Weights that balance covariates

[Hainmueller, 2011; Zubizarreta, 2015]

$$\begin{aligned} \min_{\gamma} \quad & \sum_{T_i=0} \gamma_i \log \gamma_i \\ \text{subject to} \quad & \sum_{T_i=1} X_i = \sum_{T_i=0} \gamma_i X_i \end{aligned}$$

Exact Balancing Weights

Weights that balance covariates

[Hainmueller, 2011; Zubizarreta, 2015]

$$\min_{\gamma} \sum_{T_i=0} \gamma_i \text{log } \gamma_i$$
$$\text{subject to } \sum_{T_i=1} X_i = \sum_{T_i=0} \gamma_i X_i$$

Calibrated propensity score model

[Tan, 2017; Wang and Zubizarreta, 2018]

$$\min_{\alpha,\beta} \mathcal{L}_{\text{cal}}(\alpha, \beta)$$
$$\pi(x) = \text{logit}^{-1}(\alpha + \beta' x)$$

Exact Balancing Weights

Weights that balance covariates

[Hainmueller, 2011; Zubizarreta, 2015]

$$\min_{\gamma} \sum_{T_i=0} \gamma_i \log \gamma_i$$
$$\text{subject to} \sum_{T_i=1} X_i = \sum_{T_i=0} \gamma_i X_i$$

Calibrated propensity score model

[Tan, 2017; Wang and Zubizarreta, 2018]

$$\min_{\alpha,\beta} \mathcal{L}_{\text{cal}}(\alpha, \beta)$$
$$\pi(x) = \text{logit}^{-1}(\alpha + \beta' x)$$

Linking the two: [Zhao and Percival, 2016]

$$\gamma_i = \exp(\alpha + \beta' X_i) = \frac{\text{logit}^{-1}(\alpha + \beta' X_i)}{1 - \text{logit}^{-1}(\alpha + \beta' X_i)}$$

Approximate Balancing Weights

Exact balance is unlikely in high dimensions [Zhao and Percival, 2016; Wang and Zubizarreta, 2018]

Approximate Balancing Weights

Exact balance is unlikely in high dimensions [Zhao and Percival, 2016; Wang and Zubizarreta, 2018]

Approximate balance:

$$\left\| \sum_{T_i=1} X_i - \sum_{T_i=0} \gamma_i X_i \right\|_p$$

Approximate Balancing Weights

Exact balance is unlikely in high dimensions [Zhao and Percival, 2016; Wang and Zubizarreta, 2018]

Approximate balance:

$$\left\| \sum_{T_i=1} X_i - \sum_{T_i=0} \gamma_i X_i \right\|_p$$

Regularized calibrated propensity score:

$$\min_{\alpha,\beta} \; \mathcal{L}_{\mathsf{cal}}(\alpha, \beta) + \|\beta\|_q$$

Why even do this?

Why even do this?

Why not just use outcome modeling?

Why even do this?

Why not just use outcome modeling?

- Flexible outcome modeling is extremely popular [Hahn et al., 2018]

Why even do this?

Why not just use outcome modeling?

- Flexible outcome modeling is extremely popular [Hahn et al., 2018]
- Allowing negative weights = multilevel model regression weights

Why even do this?

Why not just use outcome modeling?

- Flexible outcome modeling is extremely popular [Hahn et al., 2018]
- Allowing negative weights = multilevel model regression weights
- Augment balancing weights with outcome models

Why even do this?

Why not just use outcome modeling?

- Flexible outcome modeling is extremely popular [Hahn et al., 2018]
- Allowing negative weights = multilevel model regression weights
- Augment balancing weights with outcome models

Ok, but why not just use standard IPW?

Why even do this?

Why not just use outcome modeling?

- Flexible outcome modeling is extremely popular [Hahn et al., 2018]
- Allowing negative weights = multilevel model regression weights
- Augment balancing weights with outcome models

Ok, but why not just use standard IPW?

- Many *design based* estimators using IPW

Why even do this?

Why not just use outcome modeling?

- Flexible outcome modeling is extremely popular [Hahn et al., 2018]
- Allowing negative weights = multilevel model regression weights
- Augment balancing weights with outcome models

Ok, but why not just use standard IPW?

- Many *design based* estimators using IPW
- Interest in adapting to multilevel settings [Li et al., 2013]

Why even do this?

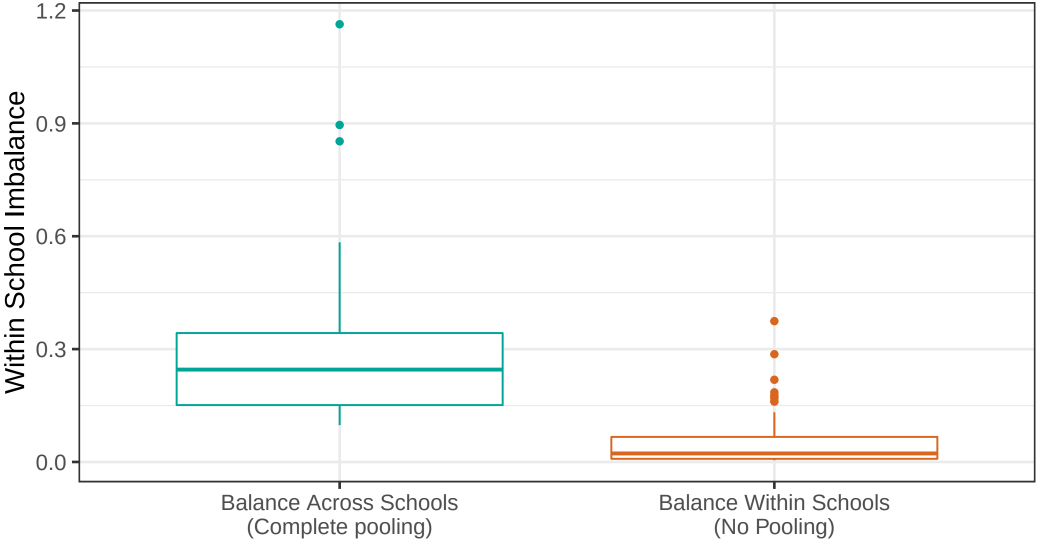
Why not just use outcome modeling?

- Flexible outcome modeling is extremely popular [Hahn et al., 2018]
- Allowing negative weights = multilevel model regression weights
- Augment balancing weights with outcome models

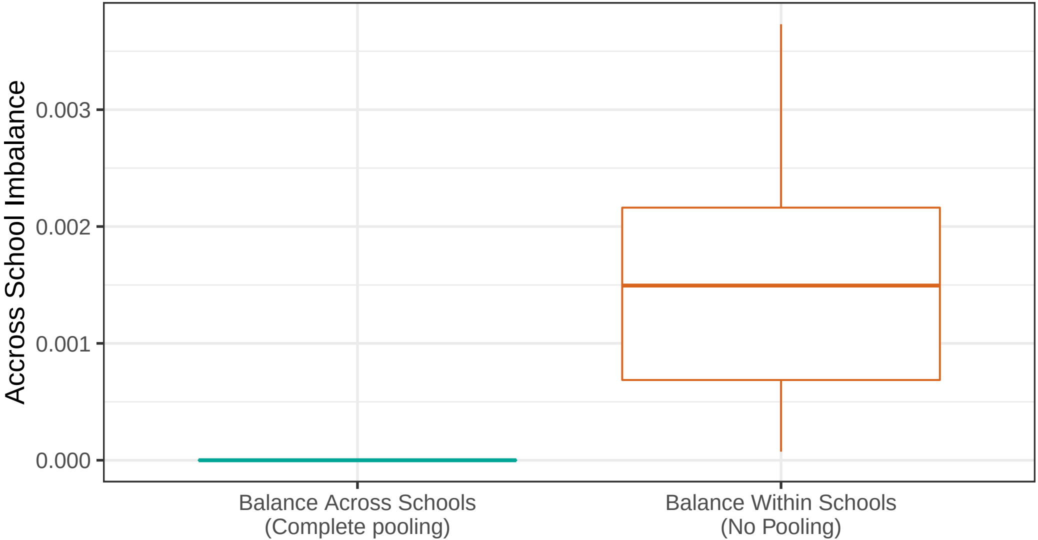
Ok, but why not just use standard IPW?

- Many *design based* estimators using IPW
- Interest in adapting to multilevel settings [Li et al., 2013]
- Poor finite sample performance, especially in high dimensions! [Athey et al., 2018]

So why not just balance within each school?



So why not just balance within each school?



Matrix Balance Constraints

and

Hierarchical Models

What we see and what we assume

For student i in school j observe:

- Student-level covariates $X_i \in \mathbb{R}^d$ and school-level covariates $V_j \in \mathbb{R}^p$
- Treatment status T_i and school indicator Z_i
- Outcome: $Y_i = Y_i(1)T_i + Y_i(0)(1 - T_i)$

What we see and what we assume

For student i in school j observe:

- Student-level covariates $X_i \in \mathbb{R}^d$ and school-level covariates $V_j \in \mathbb{R}^p$
- Treatment status T_i and school indicator Z_i
- Outcome: $Y_i = Y_i(1)T_i + Y_i(0)(1 - T_i)$

Goal: Estimate the ATT

$$\tau = \mathbb{E}[Y(1) - Y(0) \mid T = 1]$$

What we see and what we assume

For student i in school j observe:

- Student-level covariates $X_i \in \mathbb{R}^d$ and school-level covariates $V_j \in \mathbb{R}^p$
- Treatment status T_i and school indicator Z_i
- Outcome: $Y_i = Y_i(1)T_i + Y_i(0)(1 - T_i)$

Goal: Estimate the ATT **and school CATT**

$$\tau = \mathbb{E}[Y(1) - Y(0) \mid T = 1]$$

$$\tau_j = \mathbb{E}[Y(1) - Y(0) \mid T = 1, Z = j]$$

What we see and what we assume

For student i in school j observe:

- Student-level covariates $X_i \in \mathbb{R}^d$ and school-level covariates $V_j \in \mathbb{R}^p$
- Treatment status T_i and school indicator Z_i
- Outcome: $Y_i = Y_i(1)T_i + Y_i(0)(1 - T_i)$

Goal: Estimate the ATT **and school CATT**

$$\tau = \mathbb{E}[Y(1) - Y(0) \mid T = 1]$$

$$\tau_j = \mathbb{E}[Y(1) - Y(0) \mid T = 1, Z = j]$$

Key Identifying assumption: Strong ignorability

$$Y(1), Y(0) \perp T \mid X, V \quad \text{and} \quad \pi(x) < 1$$

What we see and what we assume

For student i in school j observe:

- Student-level covariates $X_i \in \mathbb{R}^d$ and school-level covariates $V_j \in \mathbb{R}^p$
- Treatment status T_i and school indicator Z_i
- Outcome: $Y_i = Y_i(1)T_i + Y_i(0)(1 - T_i)$

Goal: Estimate the ATT **and school CATT**

$$\tau = \mathbb{E}[Y(1) - Y(0) \mid T = 1]$$

$$\tau_j = \mathbb{E}[Y(1) - Y(0) \mid T = 1, Z = j]$$

Key Identifying assumption: Strong ignorability Sorry...

$$Y(1), Y(0) \perp T \mid X, V \quad \text{and} \quad \pi(x) < 1$$

Estimating Effects

Estimate school CATT with matrix weights $\hat{\Gamma} \in \mathbb{R}^{n \times J}$

$$\hat{\tau}_j = \frac{1}{n_1} \sum_{j[i]=j, T_i=1} Y_i - \frac{1}{n_1} \sum_{j[i]=j, T_i=0} \hat{\Gamma}_{ij} Y_i$$

Estimating Effects

Estimate school CATT with matrix weights $\hat{\Gamma} \in \mathbb{R}^{n \times J}$

$$\hat{\tau}_j = \frac{1}{n_1} \sum_{j[i]=j, T_i=1} Y_i - \frac{1}{n_1} \sum_{j[i]=j, T_i=0} \hat{\Gamma}_{ij} Y_i$$

How do we prioritize balance?

Estimating Effects

Estimate school CATT with matrix weights $\hat{\Gamma} \in \mathbb{R}^{n \times J}$

$$\hat{\tau}_j = \frac{1}{n_1} \sum_{j[i]=j, T_i=1} Y_i - \frac{1}{n_1} \sum_{j[i]=j, T_i=0} \hat{\Gamma}_{ij} Y_i$$

How do we prioritize balance?

Global balance only $\leftrightarrow$ Complete pooling

Estimating Effects

Estimate school CATT with matrix weights $\hat{\Gamma} \in \mathbb{R}^{n \times J}$

$$\hat{\tau}_j = \frac{1}{n_1} \sum_{j[i]=j, T_i=1} Y_i - \frac{1}{n_1} \sum_{j[i]=j, T_i=0} \hat{\Gamma}_{ij} Y_i$$

How do we prioritize balance?

Global balance only	$\leftrightarrow$	Complete pooling
Within school balance only	$\leftrightarrow$	No pooling

Estimating Effects

Estimate school CATT with matrix weights $\hat{\Gamma} \in \mathbb{R}^{n \times J}$

$$\hat{\tau}_j = \frac{1}{n_1} \sum_{j[i]=j, T_i=1} Y_i - \frac{1}{n_1} \sum_{j[i]=j, T_i=0} \hat{\Gamma}_{ij} Y_i$$

How do we prioritize balance?

Global balance only	$\leftrightarrow$	Complete pooling
Within school balance only	$\leftrightarrow$	No pooling
Both	$\leftrightarrow$	Partial pooling

Estimating Effects

Estimate school CATT with matrix weights $\hat{\Gamma} \in \mathbb{R}^{n \times J}$

$$\hat{\tau}_j = \frac{1}{n_1} \sum_{j[i]=j, T_i=1} Y_i - \frac{1}{n_1} \sum_{j[i]=j, T_i=0} \hat{\Gamma}_{ij} Y_i$$

How do we prioritize balance?

Global balance only	$\leftrightarrow$	Complete pooling
Within school balance only	$\leftrightarrow$	No pooling
Both	$\leftrightarrow$	Partial pooling

Next: View as hierarchical modeling

The across/within school tradeoff

How do we regularize?
μ_β is a vector $\in \mathbb{R}^d$
β is a **matrix** $\in \mathbb{R}^{d \times J}$

$$\frac{1}{2\sigma_{\mu_\beta}} \|\mu_\beta\|_2^2 + \frac{1}{2\sigma_\beta} \sum_{j=1}^{J} \|\beta_j - \mu_\beta\|_2^2$$

The across/within school tradeoff

How do we regularize?
μ_β is a vector $\in \mathbb{R}^d$
β is a **matrix** $\in \mathbb{R}^{d \times J}$

$$\frac{1}{2\sigma_{\mu_\beta}} \|\mu_\beta\|_2^2 + \frac{1}{2\sigma_\beta} \sum_{j=1}^{J} \|\beta_j - \mu_\beta\|_2^2$$

How do we measure balance?
Global Balance is vector $\in \mathbb{R}^d$
School Balance is **matrix** $\in \mathbb{R}^{d \times J}$

$$\frac{\sigma_{\mu_\beta}}{2} \|\text{Global Balance}\|_2^2 + \frac{\sigma_\beta}{2} \|\text{School Balance}\|_F^2$$

The across/within school tradeoff

How do we regularize?
μ_β is a vector $\in \mathbb{R}^d$
β is a **matrix** $\in \mathbb{R}^{d \times J}$

$$\frac{1}{2\sigma_{\mu_\beta}} \|\mu_\beta\|_2^2 + \frac{1}{2\sigma_\beta} \sum_{j=1}^{J} \|\beta_j - \mu_\beta\|_2^2$$

How do we measure balance?
Global Balance is vector $\in \mathbb{R}^d$
School Balance is **matrix** $\in \mathbb{R}^{d \times J}$

$$\frac{\sigma_{\mu_\beta}}{2} \|\text{Global Balance}\|_2^2 + \frac{\sigma_\beta}{2} \|\text{School Balance}\|_F^2$$

Other examples:

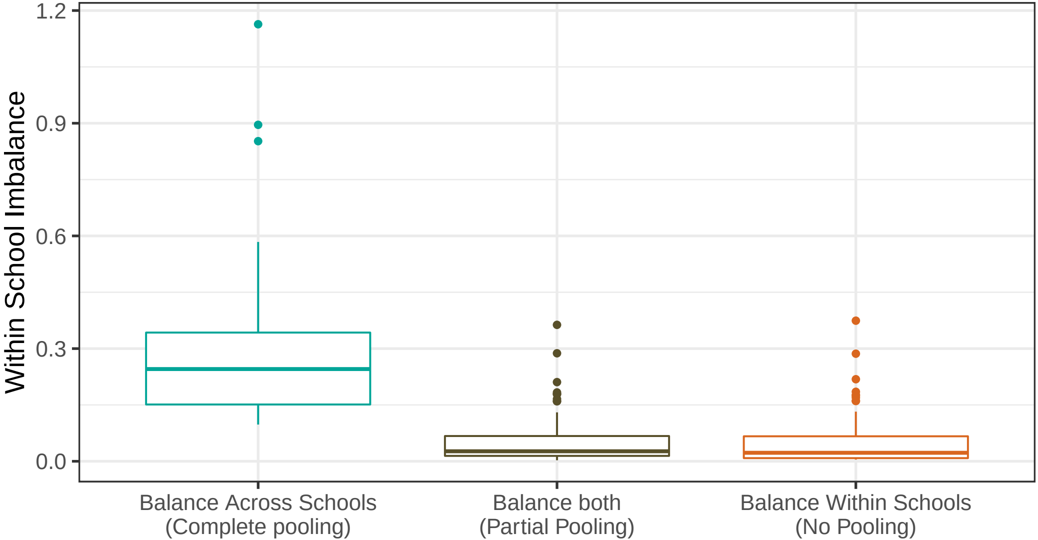
- Sparse deviations from sparse global model
- Low rank deviations from sparse global model

Local vs Global Balance

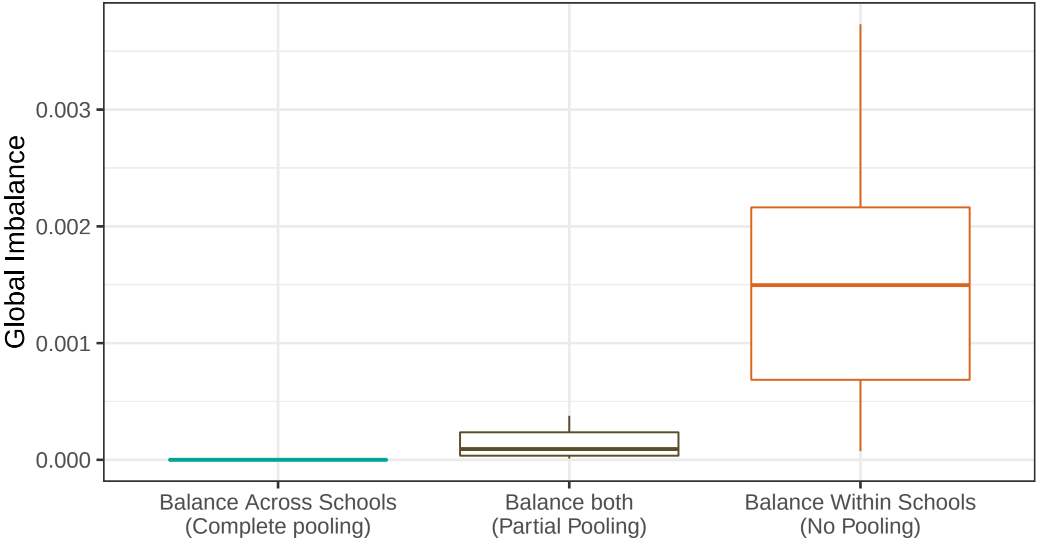
and

Partial Pooling

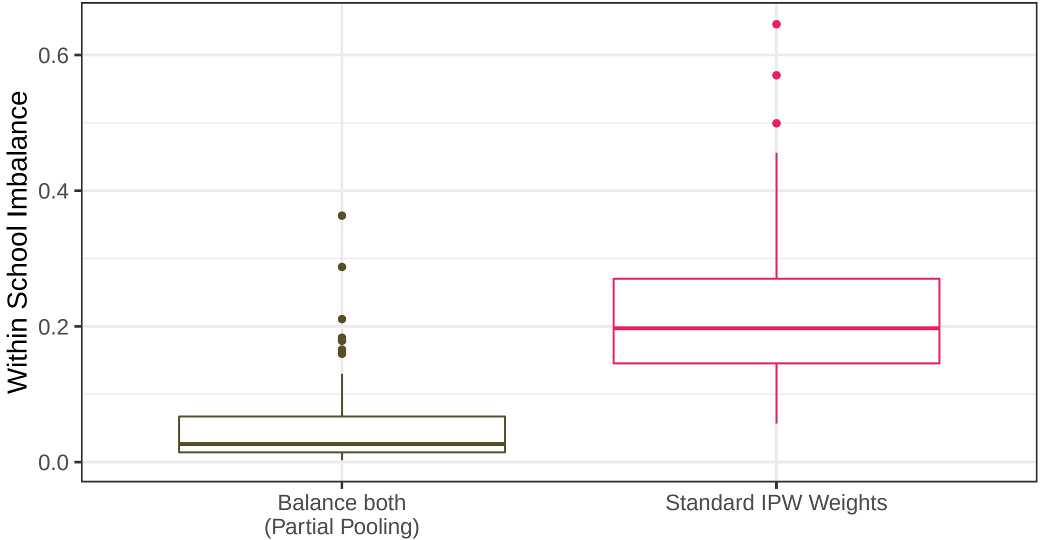
Partial pooling: jack of all trades



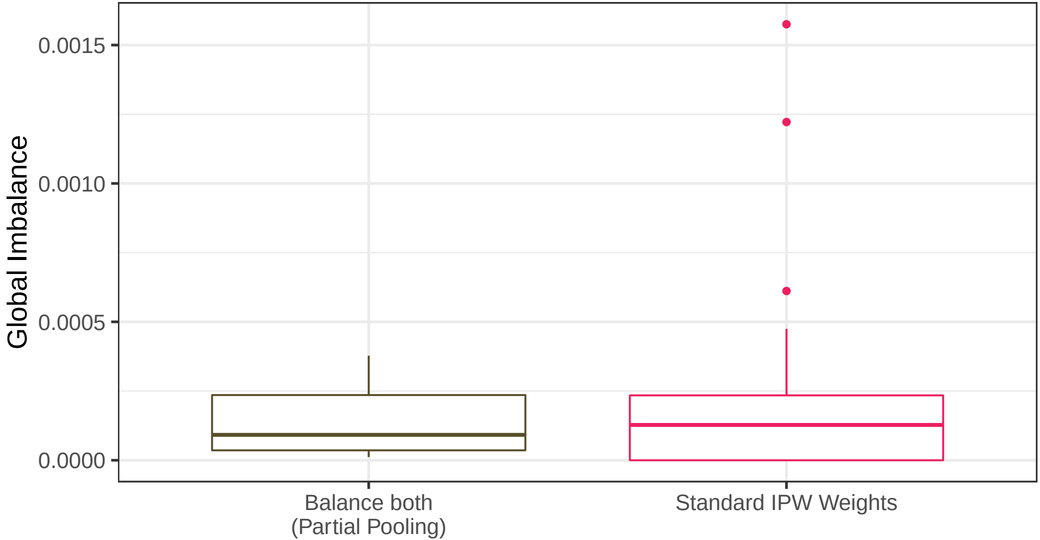
Partial pooling: jack of all trades



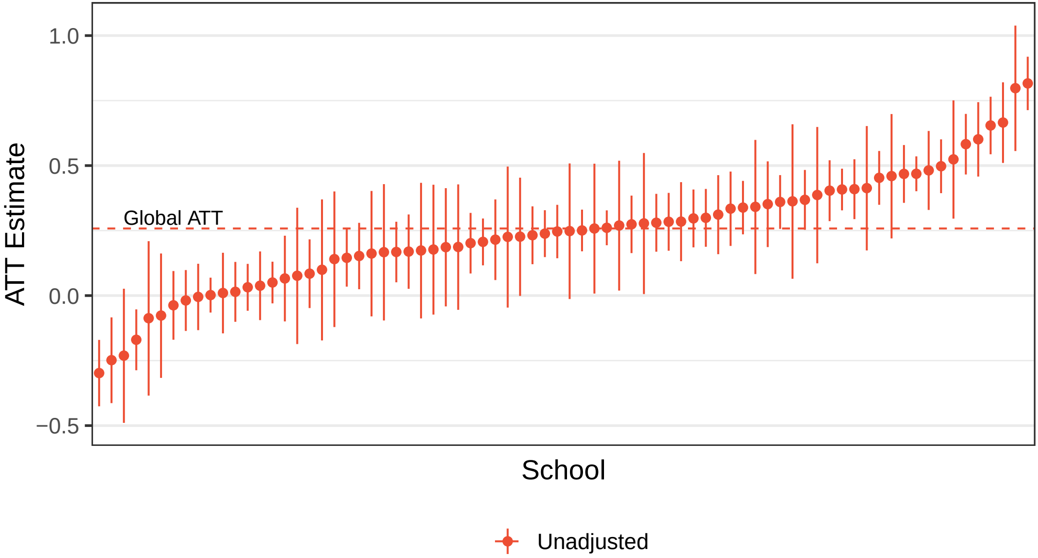
Balancing weights vs standard IPW



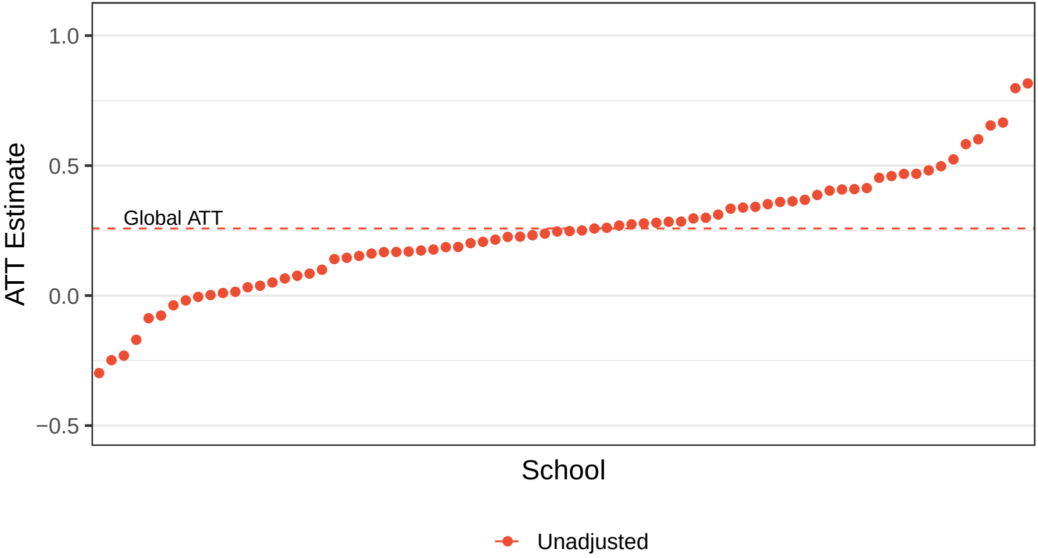
Balancing weights vs standard IPW



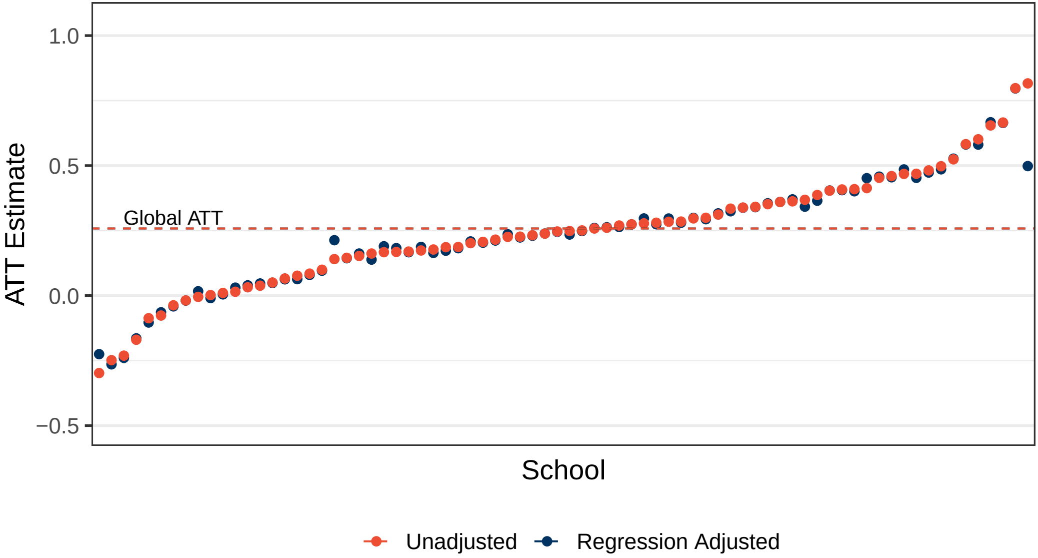
That’s a lot of heterogeneity



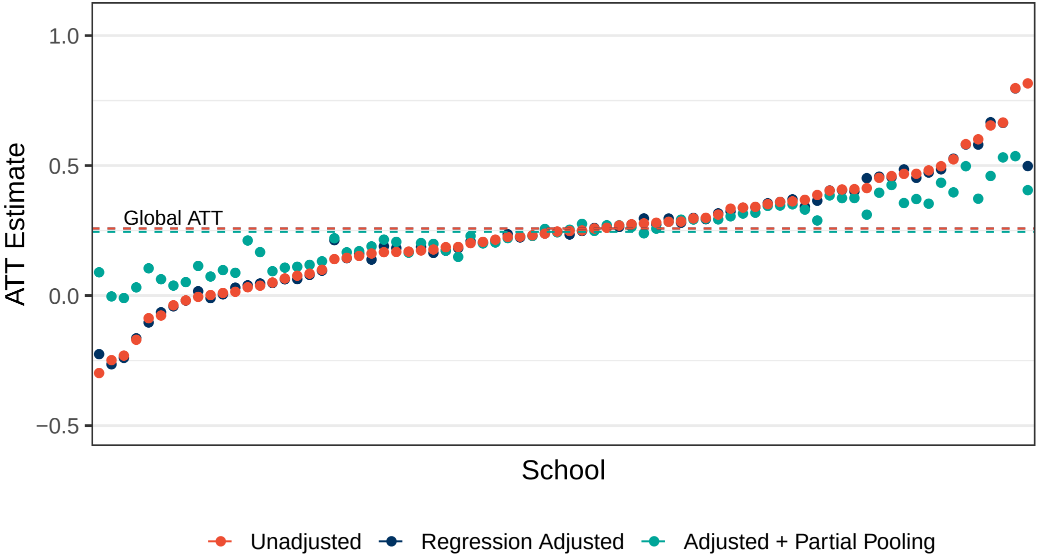
Bias Correction



Bias Correction



Pooling ATT Estimates



Odds and Ends and Future Work

Odds and Ends and Future Work

- Fully incorporate outcome modeling
  - Treatment effect and school-level covariate interactions
  - Augmented IPW [Athey et al., 2018]
  - Design shift and re-weighted risk minimization [Johansson et al., 2018]

Odds and Ends and Future Work

- Fully incorporate outcome modeling
  - Treatment effect and school-level covariate interactions
  - Augmented IPW [Athey et al., 2018]
  - Design shift and re-weighted risk minimization [Johansson et al., 2018]
- Extends to other estimands
  - From ATT to ATE
    - Double the number of weights and measure balance jointly
  - General CATE Estimation
    - Methods extend to general CATE estimation by balancing individuals' features

Odds and Ends and Future Work

- Fully incorporate outcome modeling
  - Treatment effect and school-level covariate interactions
  - Augmented IPW [Athey et al., 2018]
  - Design shift and re-weighted risk minimization [Johansson et al., 2018]
- Extends to other estimands
  - From ATT to ATE
    - Double the number of weights and measure balance jointly
  - General CATE Estimation
    - Methods extend to general CATE estimation by balancing individuals' features
- Future work
  - Inference
  - Double robustness properties
  - Sensitivity Analysis

Thank you!

`ebenmichael@berkeley.edu`

`ebenmichael.github.io`

References

References I

Athey, S., Imbens, G. W., and Wager, S. (2018). Approximate Residual Balancing: De-Biased Inference of Average Treatment Effects in High Dimensions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*.

Feller, A. and Gelman, A. (2015). Hierarchical Models for Causal Effects. *Emerging Trends in the Social and Behavioral Sciences*.

Hahn, P. R., Murray, J., and Carvalho, C. M. (2018). BAYESIAN REGRESSION TREE MODELS FOR CAUSAL INFERENCE: REGULARIZATION, CONFOUNDING, AND HETEROGENEOUS EFFECTS.

Hainmueller, J. (2011). Entropy Balancing for Causal Effects: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies. *Political Analysis*, 20:25–46.

Johansson, F. D., Kallus, N., Shalit, U., and Sontag, D. (2018). Learning Weighted Representations for Generalization Across Designs.

Li, F., Zaslavsky, A. M., and Landrum, M. B. (2013). Propensity score weighting with multilevel data. *Statistics in Medicine*, 32(19):3373–3387.

References II

Raudenbush, S. W. and Bloom, H. S. (2015). Learning About and From Variation in Program Impacts Using Multisite Trials. *MDRC Working Paper*.

Tan, Z. (2017). Regularized calibrated estimation of propensity scores with model misspecification and high-dimensional data.

Wang, Y. and Zubizarreta, J. R. (2018). Minimal Approximately Balancing Weights: Asymptotic Properties and Practical Considerations.

Yeager, D. (2017).

Zhao, Q. and Percival, D. (2016). Entropy Balancing is Doubly Robust. *Journal of Causal Inference*.

Zubizarreta, J. R. (2015). Stable Weights that Balance Covariates for Estimation With Incomplete Outcome Data. *Journal of the American Statistical Association*, 110(511):910–922.

Appendix

Inverse Propensity Score Weights

Good estimate for $\mathbb{E}[Y(1) \mid T = 1]$:

$$\frac{1}{n_1} \sum_{T_i=1} Y_i$$

Inverse Propensity Score Weights

Good estimate for $\mathbb{E}[Y(1) \mid T = 1]$:

$$\frac{1}{n_1} \sum_{T_i=1} Y_i$$

Estimate $\mathbb{E}[Y(0) \mid T = 1]$ with *importance sampling*:

$$\frac{1}{n_1} \sum_{T_i=0} \frac{\pi(X_i, V_{j[i]})}{1 - \pi(X_i, V_{j[i]})} Y_i$$

where

$$\pi(X_i, V_{j[i]}) = \mathbb{E}[T_i \mid X_i, V_{j[i]}]$$

Inverse Propensity Score Weights

Good estimate for $\mathbb{E}[Y(1) \mid T = 1]$:

$$\frac{1}{n_1} \sum_{T_i = 1} Y_i$$

Estimate $\mathbb{E}[Y(0) \mid T = 1]$ with *importance sampling*:*

$$\frac{1}{n_1} \sum_{T_i = 0} \frac{\pi(X_i, V_{j[i]})}{1 - \pi(X_i, V_{j[i]})} Y_i$$

where

$$\pi(X_i, V_{j[i]}) = \mathbb{E}[T_i \mid X_i, V_{j[i]}]$$

* More generally, reweight a loss function

Entropy Balancing = Calibrated Propensity Score Estimation

Primal: Entropy Balancing [Hainmueller, 2011]

$$\begin{aligned} \min_{\gamma} \quad & \sum_{T_i=0} \gamma_i \log \gamma_i \\ \text{subject to} \quad & \sum_{T_i=1} X_i - \sum_{T_i=0} \gamma_i X_i = 0 \end{aligned}$$

Entropy Balancing = Calibrated Propensity Score Estimation

Primal: Entropy Balancing [Hainmueller, 2011]

$$\min_{\gamma} \sum_{T_i=0} \gamma_i \log \gamma_i$$

$$\text{subject to} \sum_{T_i=1} X_i - \sum_{T_i=0} \gamma_i X_i = 0$$

Dual: Calibrated Propensity Score [Zhao and Percival, 2016; Tan, 2017]

$$\min_{\alpha,\beta} \sum_{T_i=0} \exp(\alpha + \beta' X_i) - \sum_{T_i=1} \left[\alpha + \beta' X_i\right]$$

Entropy Balancing = Calibrated Propensity Score Estimation

Primal: Entropy Balancing [Hainmueller, 2011]

$$\min_{\gamma} \sum_{T_i=0} \gamma_i \log \gamma_i$$
$$\text{subject to} \sum_{T_i=1} X_i - \sum_{T_i=0} \gamma_i X_i = 0$$

Dual: Calibrated Propensity Score [Zhao and Percival, 2016; Tan, 2017]

$$\min_{\alpha,\beta} \sum_{T_i=0} \exp(\alpha + \beta' X_i) - \sum_{T_i=1} \left[\alpha + \beta' X_i\right]$$

Linking the two:

$$\gamma_i = \exp(\alpha + \beta' X_i) = \frac{\text{logit}^{-1}(\alpha + \beta' X_i)}{1 - \text{logit}^{-1}(\alpha + \beta' X_i)}$$

The Calibrated Loss Function

Dual: Calibrated Propensity Score [Zhao and Percival, 2016; Tan, 2017]

$$\min_{\alpha,\beta} \sum_{T_i=0} \exp(\alpha + \beta' X_i) - \sum_{T_i=1} \left[\alpha + \beta' X_i\right]$$

The Calibrated Loss Function

Dual: Calibrated Propensity Score [Zhao and Percival, 2016; Tan, 2017]

$$\min_{\alpha,\beta} \sum_{T_i=0} \mathrm{exp}(\alpha + \beta' X_i) - \sum_{T_i=1} \left[\alpha + \beta' X_i\right]$$

Consistent estimator for *logistic propensity score* [Zhao and Percival, 2016]

$$\pi(x) = \mathrm{logit}^{-1}(\alpha + \beta' x)$$

The Calibrated Loss Function

Dual: Calibrated Propensity Score [Zhao and Percival, 2016; Tan, 2017]

$$\min_{\alpha,\beta} \sum_{T_i=0} \exp(\alpha + \beta' X_i) - \sum_{T_i=1} \left[\alpha + \beta' X_i\right]$$

Consistent estimator for *logistic propensity score* [Zhao and Percival, 2016]

$$\pi(x) = \text{logit}^{-1}(\alpha + \beta' x)$$

Loss function *exactly balances* covariates with IPW weights

$$\sum_{T_i=0} \exp(\alpha + \beta' X) X_i - \sum_{T_i=1} X_i = 0$$

The Calibrated Loss Function

Dual: Calibrated Propensity Score [Zhao and Percival, 2016; Tan, 2017]

$$\min_{\alpha,\beta} \sum_{T_i=0} \exp(\alpha + \beta' X_i) - \sum_{T_i=1} \left[\alpha + \beta' X_i\right]$$

Consistent estimator for *logistic propensity score* [Zhao and Percival, 2016]

$$\pi(x) = \text{logit}^{-1}(\alpha + \beta' x)$$

Loss function *exactly balances* covariates with IPW weights

$$\sum_{T_i=0} \exp(\alpha + \beta' X) X_i - \sum_{T_i=1} X_i = 0$$

$$\sum_{T_i=0} \gamma_i X_i - \sum_{T_i=1} X_i = 0$$

Approximate Balance = Regularization

Exact balance is unlikely in high dimensions [Zhao and Percival, 2016; Wang and Zubizarreta, 2018]

So *approximate balance* must suffice

† h^* is the *convex conjugate*

Approximate Balance = Regularization

Exact balance is unlikely in high dimensions [Zhao and Percival, 2016; Wang and Zubizarreta, 2018]

So *approximate balance* must suffice

Primal: Measure balance with a *balance criterion* $h : \mathbb{R}^d \to \mathbb{R}$

$$\min_{\gamma} \quad h\left(\sum_{T_i=1} X_i - \sum_{T_i=0} \gamma_i X_i\right) + \sum_{T_i=0} \gamma_i \log \gamma_i$$

† h^* is the *convex conjugate*

Approximate Balance = Regularization

Exact balance is unlikely in high dimensions [Zhao and Percival, 2016; Wang and Zubizarreta, 2018]

So *approximate balance* must suffice

Primal: Measure balance with a *balance criterion* $h : \mathbb{R}^d \to \mathbb{R}$

$$\min_{\gamma} \quad h\left(\sum_{T_i=1} X_i - \sum_{T_i=0} \gamma_i X_i\right) + \sum_{T_i=0} \gamma_i \log \gamma_i$$

Example: $h(x) = \mathcal{I}(\|x\|_\infty \leq \delta)$ [Zubizarreta, 2015; Athey et al., 2018; Wang and Zubizarreta, 2018]

† h^* is the *convex conjugate*

Approximate Balance = Regularization

Exact balance is unlikely in high dimensions [Zhao and Percival, 2016; Wang and Zubizarreta, 2018]

So *approximate balance* must suffice

Primal: Measure balance with a *balance criterion* $h : \mathbb{R}^d \to \mathbb{R}$

$$\min_{\gamma} \quad h\left(\sum_{T_i=1} X_i - \sum_{T_i=0} \gamma_i X_i\right) + \sum_{T_i=0} \gamma_i \log \gamma_i$$

Example: $h(x) = \mathcal{I}(\|x\|_\infty \leq \delta)$ [Zubizarreta, 2015; Athey et al., 2018; Wang and Zubizarreta, 2018]

Dual: Regularize with $h^* : \mathbb{R}^d \to \mathbb{R}^\dagger$

$$\min_{\alpha,\beta} \sum_{T_i=0} \exp(\alpha + \beta' X_i) - \sum_{T_i=1} \left[\alpha + \beta' X_i\right] + h^*(\beta)$$

$^\dagger h^*$ is the *convex conjugate*

Hierarchical Propensity Score Model

Propensity Score Model

Covariate Balance

Hierarchical Propensity Score Model

Propensity Score Model

Covariate Balance

$$T_i \mid X_i \quad \sim \text{logit}^{-1}(\alpha_{j[i]} + X_i' \beta_{j[i]})$$

Hierarchical Propensity Score Model

Propensity Score Model

$$
\begin{aligned}
T_i \mid X_i &\sim \text{logit}^{-1}(\alpha_{j[i]} + X_i' \beta_{j[i]}) \\
\alpha_j &\overset{\text{iid}}{\sim} N(\mu_\alpha, \sigma_\alpha)
\end{aligned}
$$

Covariate Balance

School Sum-to-n_{1j}

Hierarchical Propensity Score Model

Propensity Score Model

$$\begin{aligned} T_i \mid X_i &\sim \text{logit}^{-1}(\alpha_{j[i]} + X_i'\beta_{j[i]}) \\ \alpha_j &\overset{\text{iid}}{\sim} N(\mu_\alpha, \sigma_\alpha) \\ \mu_\alpha &\overset{\text{iid}}{\sim} N\left(\textstyle\sum_{\ell=1}^{p} V_\ell \gamma_\ell, \sigma_{\mu_\alpha}\right) \end{aligned}$$

Covariate Balance

School Sum-to-n_{1j}

Global Sum-to-n_1

Hierarchical Propensity Score Model

Propensity Score Model

$$
\begin{aligned}
T_i \mid X_i \quad &\sim \text{logit}^{-1}(\alpha_{j[i]} + X_i'\beta_{j[i]}) \\
\alpha_j \quad &\overset{\text{iid}}{\sim} N(\mu_\alpha, \sigma_\alpha) \\
\mu_\alpha \quad &\overset{\text{iid}}{\sim} N\left(\textstyle\sum_{\ell=1}^{p} V_\ell \gamma_\ell, \sigma_{\mu_\alpha}\right) \\
\gamma_\ell \quad &\overset{\text{iid}}{\sim} N(0, \sigma_\gamma)
\end{aligned}
$$

Covariate Balance

School Sum-to-n_{1j}

Global Sum-to-n_1

School-Level Covariate Balance

Hierarchical Propensity Score Model

Propensity Score Model

$$
\begin{aligned}
T_i \mid X_i &\sim \text{logit}^{-1}(\alpha_{j[i]} + X_i' \beta_{j[i]}) \\
\alpha_j &\overset{\text{iid}}{\sim} N(\mu_\alpha, \sigma_\alpha) \\
\mu_\alpha &\overset{\text{iid}}{\sim} N\left(\textstyle\sum_{\ell=1}^{p} V_\ell \gamma_\ell, \sigma_{\mu_\alpha}\right) \\
\gamma_\ell &\overset{\text{iid}}{\sim} N(0, \sigma_\gamma) \\
\beta_{kj} &\overset{\text{iid}}{\sim} N(\mu_{\beta_k}, \sigma_\beta)
\end{aligned}
$$

Covariate Balance

School Sum-to-n_{1j}

Global Sum-to-n_1

School-Level Covariate Balance

School Balance

Hierarchical Propensity Score Model

Propensity Score Model

$$
\begin{aligned}
T_i \mid X_i &\sim \text{logit}^{-1}(\alpha_{j[i]} + X_i' \beta_{j[i]}) \\
\alpha_j &\overset{\text{iid}}{\sim} N(\mu_\alpha, \sigma_\alpha) \\
\mu_\alpha &\overset{\text{iid}}{\sim} N\left(\textstyle\sum_{\ell=1}^{p} V_\ell \gamma_\ell, \sigma_{\mu_\alpha}\right) \\
\gamma_\ell &\overset{\text{iid}}{\sim} N(0, \sigma_\gamma) \\
\beta_{kj} &\overset{\text{iid}}{\sim} N(\mu_{\beta_k}, \sigma_\beta) \\
\mu_{\beta_k} &\overset{\text{iid}}{\sim} N(0, \sigma_{\mu_\beta})
\end{aligned}
$$

Covariate Balance

School Sum-to-n_{1j}

Global Sum-to-n_1

School-Level Covariate Balance

School Balance

Global Balance