Think Globally, Balance Locally:

Multi-Level Balancing Weights for

Multi-Site Observational Studies

Eli Ben-Michael
(UC Berkeley)

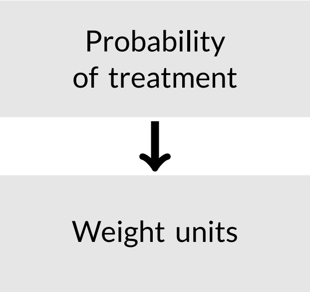
(joint work with Avi Feller)

SREE 2019
March 7, 2019

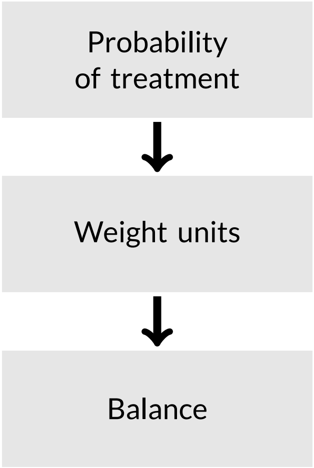
Typical Inverse Propensity Weighting (IPW) workflow

Probability
of treatment

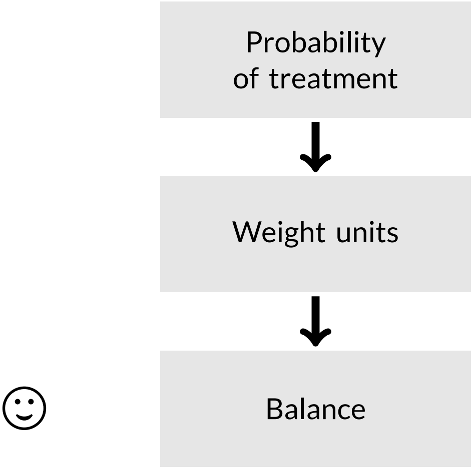
Typical Inverse Propensity Weighting (IPW) workflow



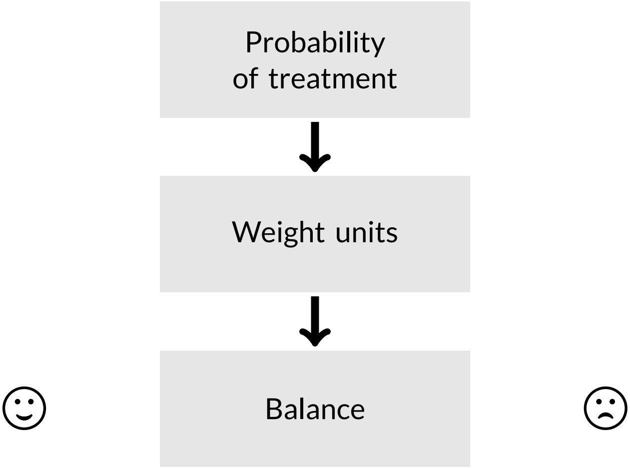
Typical Inverse Propensity Weighting (IPW) workflow



Typical Inverse Propensity Weighting (IPW) workflow



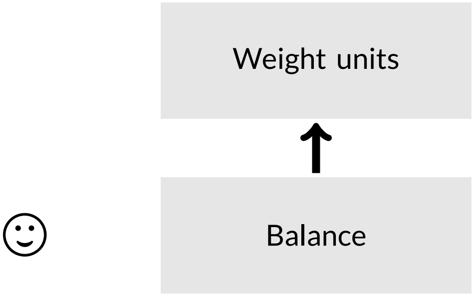
Typical Inverse Propensity Weighting (IPW) workflow



Balancing weights workflow

☺ Balance

Balancing weights workflow



Balancing weights workflow

Probability of treatment

↑

Weight units

↑

☺ Balance

Multi-level balancing weights

This talk: Extend balancing weights to multi-level setting

Multi-level balancing weights

This talk: Extend balancing weights to multi-level setting

Multilevel Matching [Zubizarreta and Keele, 2017; Pimentel et al., 2018]	—	Hierarchical IPW [Arpino and Mealli, 2011; Li et al., 2013]

Multi-level balancing weights

This talk: Extend balancing weights to multi-level setting

Multilevel Matching

[Zubizarreta and Keele, 2017; Pimentel et al., 2018]

Hierarchical IPW

[Arpino and Mealli, 2011; Li et al., 2013]

Multilevel Weighting

Multi-level selection mechanisms

Cluster-level treatment assignment

[Zubizarreta and Keele, 2017; Pimentel et al., 2018]

Multi-level selection mechanisms

Cluster-level treatment assignment

[Zubizarreta and Keele, 2017; Pimentel et al., 2018]

Multi-site studies

- **Ex**: Obs. study simulation from real RCT for ACIC workshop

  Based off of National Study of Learning Mindsets [Yeager, 2017]

The balance objective depends on the estimand

Two balancing goals:

Overall effect	$\longleftrightarrow$	Balance globally across all schools
School-level effects	$\longleftrightarrow$	Balance locally within each school

The balance objective depends on the estimand

Two balancing goals:

Overall effect	$\longleftrightarrow$	Balance globally across all schools
School-level effects	$\longleftrightarrow$	Balance locally within each school

Really want to do both:

The balance objective depends on the estimand

Two balancing goals:

Overall effect $\longleftrightarrow$ Balance globally across all schools

School-level effects $\longleftrightarrow$ Balance locally within each school

Really want to do both:

Balancing within school and globally

$\updownarrow$

Partial pooling in multilevel IPW

What we see

For student i in school $j[i]$ observe:

- Student-level covariates $X_i \in \mathbb{R}^d$ and school-level covariates $V_{j[i]} \in \mathbb{R}^p$
- Treatment status T_i, School indicator S_i
- Outcome: $Y_i = Y_i(1)T_i + Y_i(0)(1 - T_i)$

What we see

For student i in school $j[i]$ observe:

- Student-level covariates $X_i \in \mathbb{R}^d$ and school-level covariates $V_{j[i]} \in \mathbb{R}^p$
- Treatment status T_i, School indicator S_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

For student i in school $j[i]$ observe:

- Student-level covariates $X_i \in \mathbb{R}^d$ and school-level covariates $V_{j[i]} \in \mathbb{R}^p$
- Treatment status T_i, School indicator S_i
- Outcome: $Y_i = Y_i(1)T_i + Y_i(0)(1 - T_i)$

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

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

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

What we assume

Key Identifying assumption: Strong ignorability

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

What we assume

Key Identifying assumption: Strong ignorability

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

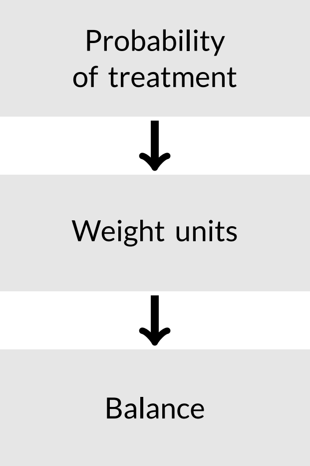
Weighting control units by the odds of treatment:

$$\mathbb{E}[Y(0) \mid T = 1] = \mathbb{E}\left[\frac{\pi(X, V)}{1 - \pi(X, V)} Y \mid T = 0\right]$$

Two views of the P-score imply different estimation methods

Conditional probability of treatment

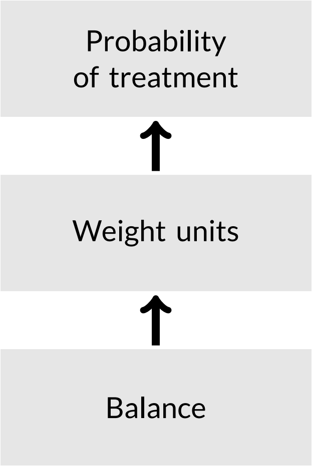
- Estimate $\hat{\pi}(X_i, V_i)$ with MLE, then estimate weights $\hat{\gamma}_i = \frac{\hat{\pi}(X_i, V_{j[i]})}{1-\hat{\pi}(X_i, V_{j[i]})}$
- Indirectly balances covariates
- Poor finite sample performance, especially with many covariates
  [Athey et al., 2018]



Two views of the P-score imply different estimation methods

Balancing score [Rosenbaum and Rubin, 1983]

- Find weights $\hat{\gamma}$ that directly balance covariates
- Indirectly estimates the P-score
- Old history as raking and calibration in survey sampling with non-response
  [Deming and Stephan, 1940; Deville et al., 1993]



Balancing globally estimates the overall propensity score

Balancing weights:

[Hainmueller, 2011; Zubizarreta, 2015]

Global Balance $= 0$

Balancing globally estimates the overall propensity score

Balancing weights:

[Hainmueller, 2011; Zubizarreta, 2015]

Global Balance $= 0$

Calibrated propensity score:

[Tan, 2017; Wang and Zubizarreta, 2018; Zhao, 2017]

$$\pi(X_i, V_{j[i]}) \quad = \text{logit}^{-1}(\alpha + {\mu_\beta}' X_i + \eta' V_{j[i]})$$

Balancing globally estimates the overall propensity score

Balancing weights:

[Hainmueller, 2011; Zubizarreta, 2015]

Calibrated propensity score:

[Tan, 2017; Wang and Zubizarreta, 2018; Zhao, 2017]

$$\text{Global Balance} = 0$$

$$\pi(X_i, V_{j[i]}) \quad = \text{logit}^{-1}(\alpha + {\mu_\beta}' X_i + \eta' V_{j[i]})$$

Linking the two: [Zhao and Percival, 2016]

$$\hat{\gamma}_i = \exp(\hat{\alpha} + {\hat{\mu}_\beta}' X_i + \hat{\eta}' V_{j[i]}) = \frac{\hat{\pi}(X_i, V_{j[i]})}{1 - \hat{\pi}(X_i, V_{j[i]})}$$

Balancing globally estimates the overall propensity score

Balancing weights:
[Hainmueller, 2011; Zubizarreta, 2015]

Calibrated propensity score: Complete Pooling
[Tan, 2017; Wang and Zubizarreta, 2018; Zhao, 2017]

Global Balance $= 0$

$$\pi(X_i, V_{j[i]}) \quad = \text{logit}^{-1}(\alpha + {\mu_\beta}' X_i + \eta' V_{j[i]})$$

Linking the two: [Zhao and Percival, 2016]

$$\hat{\gamma}_i = \exp(\hat{\alpha} + {\hat{\mu}_\beta}' X_i + \hat{\eta}' V_{j[i]}) = \frac{\hat{\pi}(X_i, V_{j[i]})}{1 - \hat{\pi}(X_i, V_{j[i]})}$$

Balancing each school fits separate propensity scores

With multiple sites, can restrict to within school analysis, follow same procedure

But now exact balance is unlikely [Zhao and Percival, 2016; Wang and Zubizarreta, 2018]

Balancing each school fits separate propensity scores

With multiple sites, can restrict to within school analysis, follow same procedure

But now exact balance is unlikely [Zhao and Percival, 2016; Wang and Zubizarreta, 2018]

Measuring balance:

$$\frac{\sigma_\beta^2}{2J} \sum_{j=1}^{J} \|\text{School Balance}_j\|_2^2$$

Balancing each school fits separate propensity scores

With multiple sites, can restrict to within school analysis, follow same procedure

But now exact balance is unlikely [Zhao and Percival, 2016; Wang and Zubizarreta, 2018]

Measuring balance:

$$\frac{\sigma_\beta^2}{2J}\sum_{j=1}^{J} \|\text{School Balance}_j\|_2^2$$

Propensity score:

$$\pi(X_i, V_{j[i]}) = \text{logit}^{-1}(\alpha_j + \beta_j{}' X_i)$$

$$\beta_j \sim N(0, \sigma_\beta^2)$$

Balancing each school fits separate propensity scores

With multiple sites, can restrict to within school analysis, follow same procedure

But now exact balance is unlikely [Zhao and Percival, 2016; Wang and Zubizarreta, 2018]

Measuring balance:

$$\frac{\sigma_\beta^2}{2J}\sum_{j=1}^{J}\|\text{School Balance}_j\|_2^2$$

Propensity score:*

$$\begin{aligned}\pi(X_i, V_{j[i]}) &= \text{logit}^{-1}(\alpha_j + \beta_j{}' X_i)\\ \beta_j &\sim N(0, \sigma_\beta^2)\end{aligned}$$

*Technically regularization

Balancing each school fits separate propensity scores

With multiple sites, can restrict to within school analysis, follow same procedure

But now exact balance is unlikely [Zhao and Percival, 2016; Wang and Zubizarreta, 2018]

Measuring balance:

$$\frac{\sigma_\beta^2}{2J}\sum_{j=1}^{J}\|\text{School Balance}_j\|_2^2$$

Propensity score:* No Pooling

$$\begin{aligned}\pi(X_i, V_{j[i]}) &= \text{logit}^{-1}(\alpha_j + \beta_j{}' X_i)\\ \beta_j &\sim N(0, \sigma_\beta^2)\end{aligned}$$

*Technically regularization

Balancing each school fits separate propensity scores

With multiple sites, can restrict to within school analysis, follow same procedure

But now exact balance is unlikely [Zhao and Percival, 2016; Wang and Zubizarreta, 2018]

Measuring balance:

$$\frac{\sigma_\beta^2}{2J}\sum_{j=1}^{J}\|\text{School Balance}_j\|_2^2$$

Propensity score:* No Pooling

$$\pi(X_i, V_{j[i]}) = \text{logit}^{-1}(\alpha_j + \beta_j{}' X_i)$$

$$\beta_j \sim N(0, \sigma_\beta^2)$$

Within school balance $\not\Rightarrow$ global balance because it is only *approximate*

*Technically regularization

Balancing both across and within schools is partial pooling

Measuring balance:

$$\text{Global Balance} = 0$$

$$+$$

$$\frac{\sigma_\beta^2}{2J} \sum_{j=1}^{J} \|\text{School Balance}_j\|_2^2$$

Balancing both across and within schools is partial pooling

Measuring balance:

$$\text{Global Balance} = 0$$

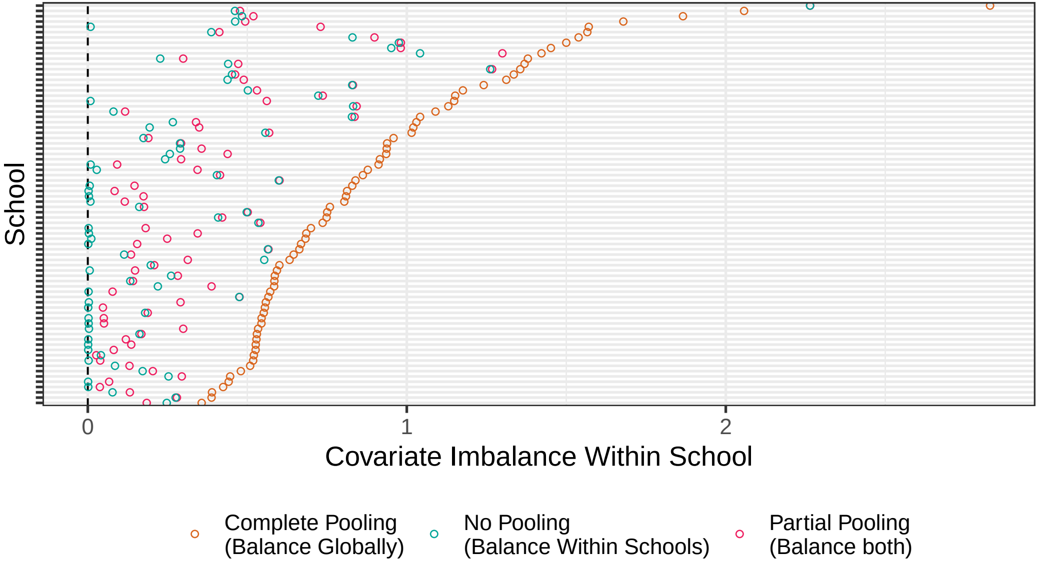
+

$$\frac{\sigma_\beta^2}{2J}\sum_{j=1}^{J}\|\text{School Balance}_j\|_2^2$$

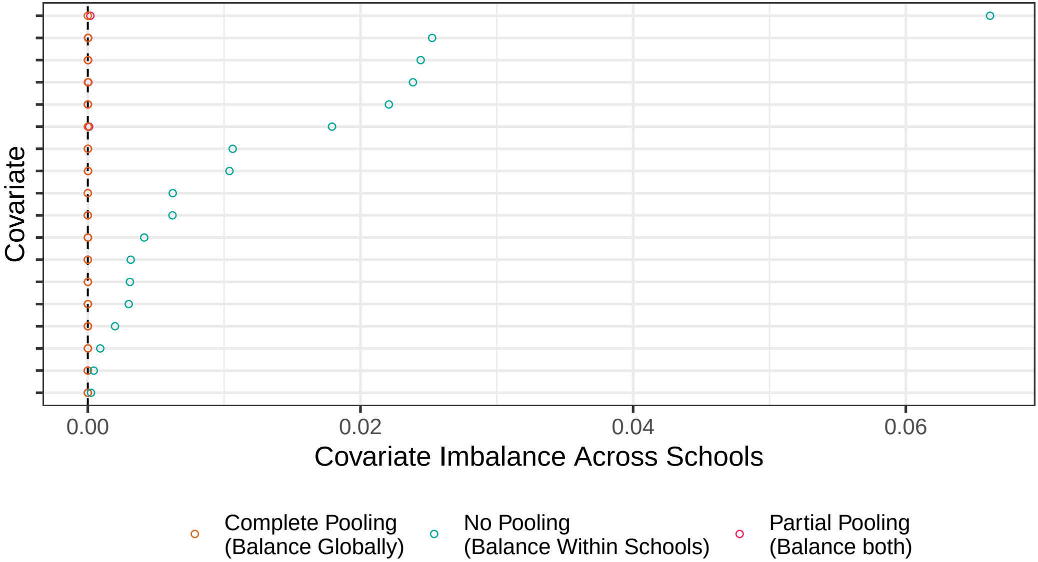
Hierarchical propensity score: (FIRC)

$$\begin{aligned}\pi(X_i, V_{j[i]}) &= \text{logit}^{-1}(\alpha_j + \beta_j{}' X_i)\\ \beta_j &\sim N\left(\mu_\beta + \eta' V_j, \sigma_\beta^2\right)\end{aligned}$$

Partial pooling achieves good balance within school



Partial pooling achieves nearly perfect balance across schools



Summary

In multi-site observational studies we want to balance:

- Globally across schools
- Locally within schools

But just balancing one does not balance the other

Solution: balance both

- Implicitly fits a multi-level propensity score model (fixed intercept random coefficients)

Summary

In multi-site observational studies we want to balance:

- Globally across schools
- Locally within schools

But just balancing one does not balance the other

Solution: balance both

- Implicitly fits a multi-level propensity score model (fixed intercept random coefficients)

Thanks!

`ebenmichael@berkeley.edu`

`ebenmichael.github.io`

References

References I

Arpino, B. and Mealli, F. (2011). The specification of the propensity score in multilevel observational studies. *Computational Statistics and Data Analysis*, 55(4):1770–1780.

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)*.

Deming, W. E. and Stephan, F. F. (1940). On a Least Squares Adjustment of a Sampled Frequency Table When the Expected Marginal Totals are Known. *The Annals of Mathematical Statistics*, 11(4):427–444.

Deville, J. C. and Särndal, C. E. (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association*, 87(418):376–382.

Deville, J. C., Särndal, C. E., and Sautory, O. (1993). Generalized raking procedures in survey sampling. *Journal of the American Statistical Association*, 88(423):1013–1020.

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

References II

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

Pimentel, S. D., Page, L. C., Lenard, M., and Keele, L. (2018). Optimal multilevel matching using network flows: An application to a summer reading intervention. *Annals of Applied Statistics*.

Rosenbaum, P. R. and Rubin, D. B. (1983). The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika*, 70(1):41–55.

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. (2017). Covariate Balancing Propensity Score by Tailored Loss Functions.

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

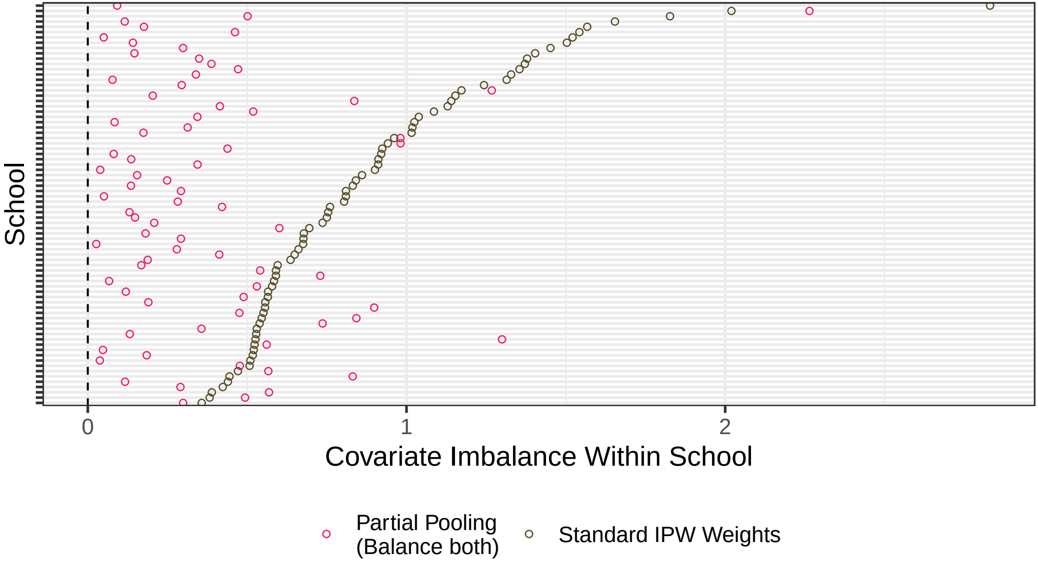
References III

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.

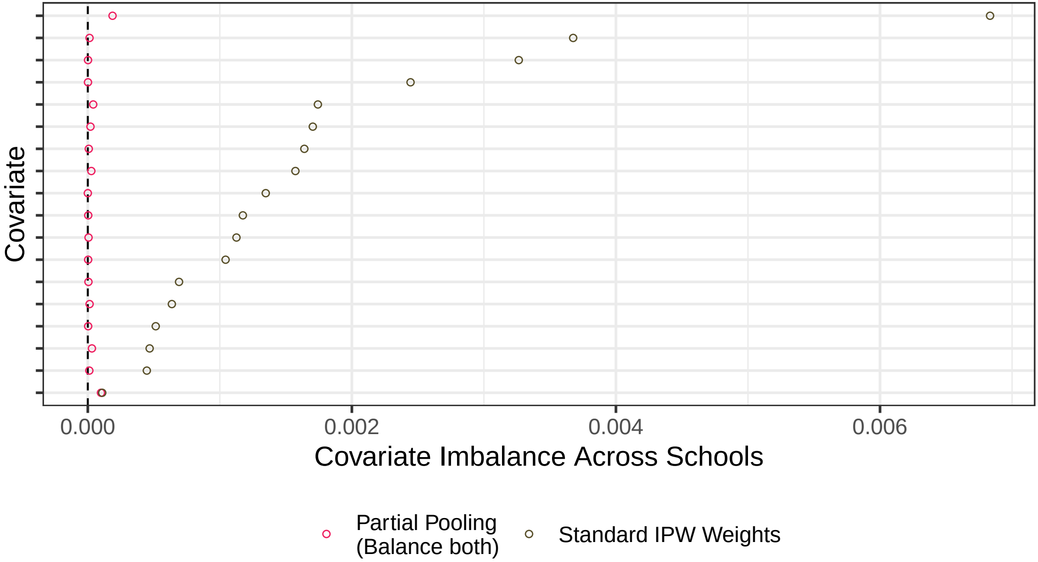
Zubizarreta, J. R. and Keele, L. (2017). Optimal Multilevel Matching in Clustered Observational Studies: A Case Study of the Effectiveness of Private Schools Under a Large-Scale Voucher System. *Journal of the American Statistical Association*, 112(518):547–560.

Appendix

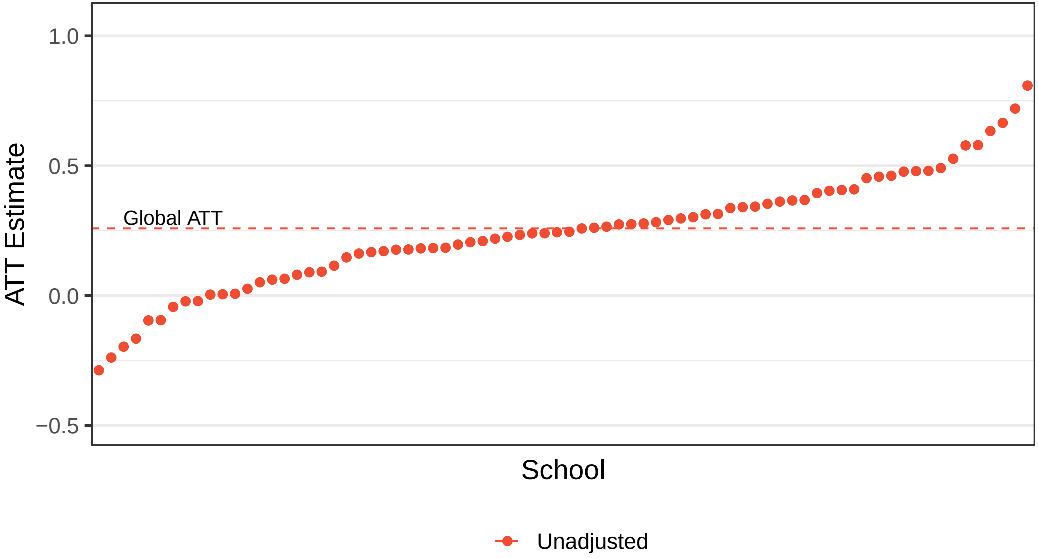
Standard multilevel IPW achieves worse balance within school



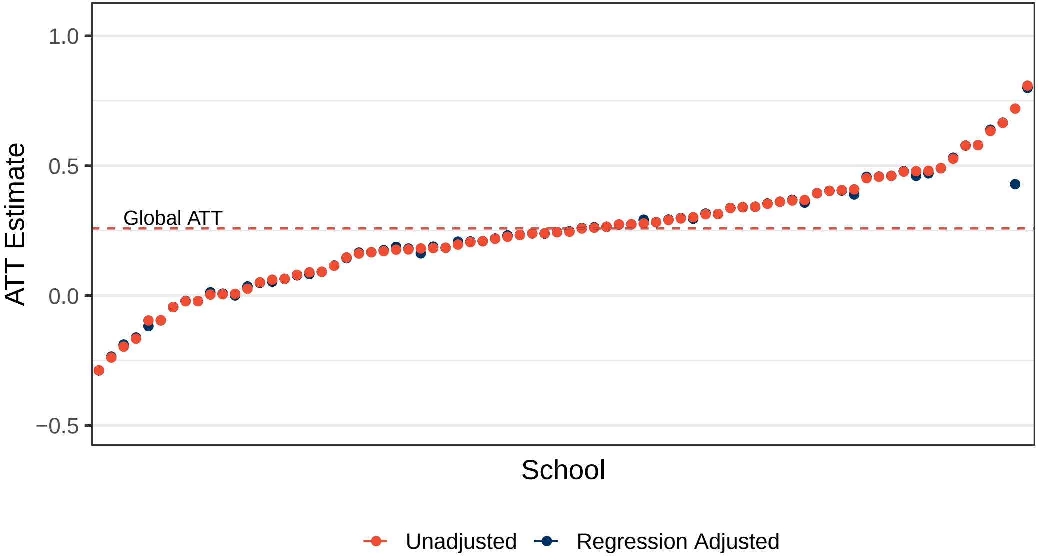
Standard multilevel IPW achieves worse balance globally



Estimating overall and school-specific effects



Bias Correction



Pooling ATT Estimates