Double Regression with Post-stratification (DRP)

for high-dimensional survey data

Eli Ben-Michael, Avi Feller, and Erin Hartman

UC Berkeley & UCLA

BigSurv20

November 2020

Even in moderate dimensions, survey adjustment is hard

Even in moderate dimensions, survey adjustment is hard

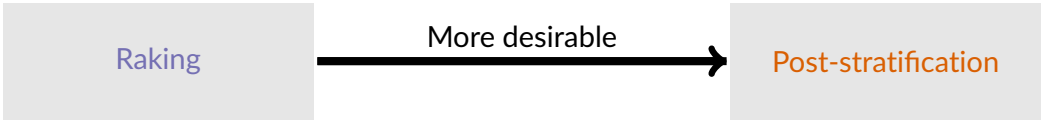
Raking

Even in moderate dimensions, survey adjustment is hard

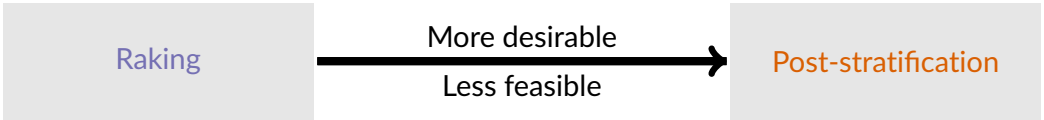
Raking

Post-stratification

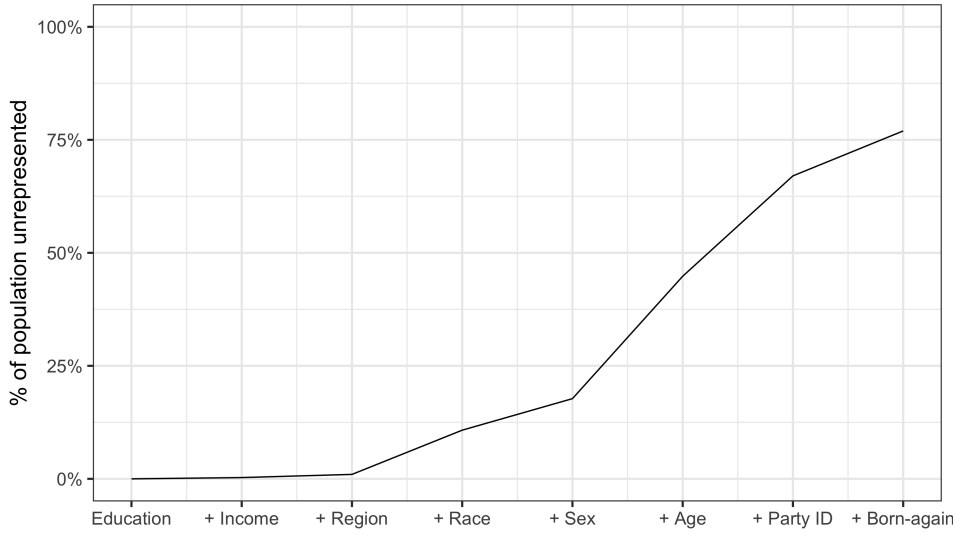
Even in moderate dimensions, survey adjustment is hard



Even in moderate dimensions, survey adjustment is hard



Quickly run into empty cells



How can we account for interactions in a principled way?

Adjusting for interactions in non-response is important

- Example: 2016 presidential election polling [Kennedy et al., 2018]

Ideally we'd post-stratify, but we can't

How can we account for interactions in a principled way?

Adjusting for interactions in non-response is important

- Example: 2016 presidential election polling [Kennedy et al., 2018]

Ideally we'd post-stratify, but we can't

This paper: Approximately post-stratify while at least raking on margins

- Leverage the value of interactions in a parsimonious way
- Dual representation as multilevel model of non-response

Combine with outcome model $\rightarrow$ Double Regression with Post-stratification (DRP)

Approximately post-stratifying
while at least raking

Notation and setup

$i = 1, \ldots, N$ individuals

- Outcome Y_i, Response R_i with prob $P(R_i = 1) = \pi_i$
- d categorical covariates w/levels $J_1, \ldots, J_d$

Notation and setup

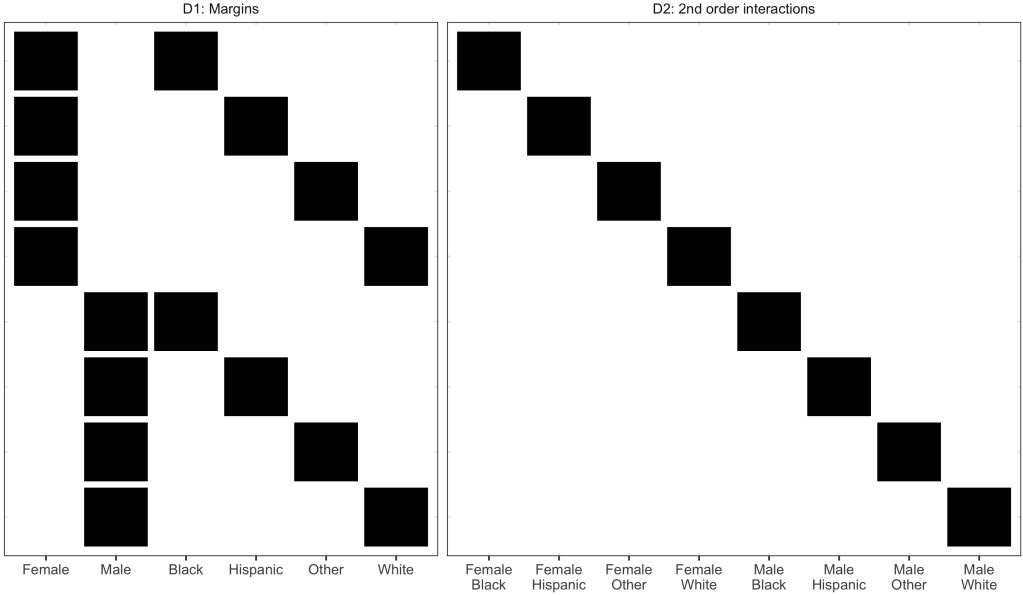
$i = 1, \ldots, N$ individuals

- Outcome Y_i, Response R_i with prob $P(R_i = 1) = \pi_i$
- d categorical covariates w/levels $J_1, \ldots, J_d$

Combine into cells $S_i \in \{1, \ldots, J_1 \times \ldots \times J_d \equiv J\}$

- Overall count vector $N^{\mathcal{P}} \in \mathbb{N}^J$ and response count vector $n^{\mathcal{R}} \in \mathbb{N}^J$

Binary vector of k^{th} order interaction terms for cell s: D_s^k


D1: Margins
Female
Male
Black
Hispanic
Other
White
D2: 2nd order interactions
Female Black
Female Hispanic
Female Other
Female White
Male Black
Male Hispanic
Male Other
Male White

Goal

Impute the average

$$\mu = \frac{1}{N}\sum_{i=1}^{n} Y_i = \frac{1}{N}\sum_{s} N_s^{\mathcal{P}}\mu_s$$

Goal

Impute the average

$$\mu = \frac{1}{N}\sum_{i=1}^{n} Y_i = \frac{1}{N}\sum_{s} N_s^{\mathcal{P}}\mu_s$$

From the respondents

$$\hat{\mu} = \frac{1}{N}\sum_{i=1}^{N} R_i\hat{\gamma}_i Y_i = \frac{1}{N}\sum_{s} n_s^{\mathcal{R}}\hat{\gamma}_s\bar{Y}_s$$

Goal

Impute the average

$$\mu = \frac{1}{N}\sum_{i=1}^{n} Y_i = \frac{1}{N}\sum_{s} N_s^{\mathcal{P}}\mu_s$$

From the respondents

$$\hat{\mu} = \frac{1}{N}\sum_{i=1}^{N} R_i\hat{\gamma}_i Y_i = \frac{1}{N}\sum_{s} n_s^{\mathcal{R}}\hat{\gamma}_s\bar{Y}_s$$

Assume responses are Missing At Random (MAR) so that within each cell

$$\mathbb{E}\left[\bar{Y}_s\right] = \mu_s$$

Choosing weights: Raking and Post-stratification

Raking on margins

Exactly match the counts for margins:

$$\sum_s D_s^1 n_s^{\mathcal{R}} \hat{\gamma}_s = \sum_s D_s^1 N_s^{\mathcal{P}}$$

- Can usually compute if d is moderate
- "Only" accounts for the linear variables

Choosing weights: Raking and Post-stratification

Raking on margins

Exactly match the counts for margins:

$$\sum_s D_s^1 n_s^{\mathcal{R}} \hat{\gamma}_s = \sum_s D_s^1 N_s^{\mathcal{P}}$$

- Can usually compute if d is moderate
- "Only" accounts for the linear variables

Post-stratification

Exactly match the counts within each cell:

$$\hat{\gamma}_s = \frac{N_s^{\mathcal{P}}}{n_s^{\mathcal{R}}}$$

- Impossible to compute with empty cells
- Unbiased when feasible

What do we want from calibration weights?

Is it worth it to try to post-stratify?

What do we want from calibration weights?

Is it worth it to try to post-stratify?

Population regression

$$Y_i = \sum_{k=1}^{K} \beta_k \cdot D_{S_i}^k + e_i$$

What do we want from calibration weights?

Is it worth it to try to post-stratify?

Population regression

$$Y_i = \sum_{k=1}^{K} \beta_k \cdot D^k_{S_i} + e_i$$

Estimation error

$$|\hat{\mu} - \mu|$$

What do we want from calibration weights?

Is it worth it to try to post-stratify?

Population regression

$$Y_i = \sum_{k=1}^{K} \beta_k \cdot D^k_{S_i} + e_i$$

Estimation error

$$|\hat{\mu} - \mu| \leq \frac{1}{N} \sum_{k=1}^{K} \|\beta_k\|_2 \left\| \sum_s D^k_s n^{\mathcal{R}}_s \gamma_s - D^k_s N^{\mathcal{P}}_s \right\|_2$$

What do we want from calibration weights?

Is it worth it to try to post-stratify?

Population regression

$$Y_i = \sum_{k=1}^{K} \beta_k \cdot D^k_{S_i} + e_i$$

Estimation error

$$|\hat{\mu} - \mu| \leq \frac{1}{N} \sum_{k=1}^{K} \|\beta_k\|_2 \left\| \sum_s D^k_s n^{\mathcal{R}}_s \gamma_s - D^k_s N^{\mathcal{P}}_s \right\|_2 + \frac{1}{N} \|\hat{\gamma}\|_2 \|e\|_2$$

What do we want from calibration weights?

Is it worth it to try to post-stratify?

Population regression

$$Y_i = \sum_{k=1}^{K} \beta_k \cdot D^k_{S_i} + e_i$$

Estimation error

$$|\hat{\mu} - \mu| \leq \frac{1}{N} \sum_{k=1}^{K} \|\beta_k\|_2 \left\| \sum_s D^k_s n^{\mathcal{R}}_s \gamma_s - D^k_s N^{\mathcal{P}}_s \right\|_2 + \frac{1}{N} \|\hat{\gamma}\|_2 \|e\|_2$$

If interactions are weak, approximate post-stratification may be enough

- Bias depends on strength of interaction $\times$ imbalance
- Variance depends on sum of the squared weights $\|\hat{\gamma}\|_2^2$

Approximately post-stratify while at least raking on margins

Find weights via convex optimization:

$$\min_{\gamma} \quad \sum_{k=2}^{d} \frac{1}{\lambda_k} \left\| \sum_{s} D_s^k n_s^{\mathcal{R}} \gamma_s - D_s^k N_s^{\mathcal{P}} \right\|_2^2 + \|\gamma\|_2^2$$

Approximately post-stratify while at least raking on margins

Find weights via convex optimization:

$$\min_{\gamma} \quad \sum_{k=2}^{d} \frac{1}{\lambda_k} \left\| \sum_s D_s^k n_s^{\mathcal{R}} \gamma_s - D_s^k N_s^{\mathcal{P}} \right\|_2^2 + \|\gamma\|_2^2$$

$$\text{subject to} \quad \sum_s D_s^1 n_s^{\mathcal{R}} \gamma_s = \sum_s D_s^1 N_s, \quad \gamma_s > 0$$

Approximately post-stratify while at least raking on margins

Find weights via convex optimization:

$$\min_{\gamma} \quad \sum_{k=2}^{d} \frac{1}{\lambda_k} \left\| \sum_s D_s^k n_s^{\mathcal{R}} \gamma_s - D_s^k N_s^{\mathcal{P}} \right\|_2^2 + \|\gamma\|_2^2$$

$$\text{subject to} \quad \sum_s D_s^1 n_s^{\mathcal{R}} \gamma_s = \sum_s D_s^1 N_s, \quad \gamma_s > 0$$

Move smoothly between two extremes

- With $\lambda_k \to \infty$, recover raking
- With $\lambda_k \to 0$, recover post-stratification
- With $\lambda_k = 1$, cell weights are regularized by cell size

Based on calibration weighting and approximate balancing weights

[Deville and Särndal, 1992; Deville et al., 1993; Zubizarreta, 2015; Hirshberg et al., 2019]

Dual view: multilevel regression for response

A **regularized** model for the **inverse** probability of response:

[Zhao and Percival, 2016; Wang and Zubizarreta, 2020; Chattopadhyay et al., 2020]

$$\frac{1}{\pi_i} \sim \left[\theta_1 \cdot D^1_{S_i} + \sum_{k=2}^{K} \theta_k \cdot D^k_{S_i} \right]_+$$

Dual view: multilevel regression for response

A **regularized** model for the **inverse** probability of response:

[Zhao and Percival, 2016; Wang and Zubizarreta, 2020; Chattopadhyay et al., 2020]

$$\frac{1}{\pi_i} \sim \left[\theta_1 \cdot D^1_{S_i} + \sum_{k=2}^{K} \theta_k \cdot D^k_{S_i} \right]_+$$

Regularization makes approximate post-stratification feasible

$$0 \times \|\theta_1\|_2^2 + \sum_{k=2}^{K} \lambda_k \|\theta_k\|_2^2$$

- At least raking $\rightarrow$ no regularization for marginal probabilities [Little and Wu, 1991]
- Approximate post-stratification $\rightarrow$ regularizing interaction terms

Link between calibration weighting and inverse propensity score weighting

Double Regression with Post-Stratification (DRP)

MRP-style approaches

Start with an outcome model (aggregated to cell-level)

$$\hat{\mu}_s = \frac{1}{N_s^P} \sum_{S_i = s} \hat{Y}_i$$

- Multilevel model, high dimensional regression, tree-based methods

MRP-style approaches

Start with an outcome model (aggregated to cell-level)

$$\hat{\mu}_s = \frac{1}{N_s^{\mathcal{P}}} \sum_{S_i = s} \hat{Y}_i$$

- Multilevel model, high dimensional regression, tree-based methods

Post-stratify using predictions instead of outcomes

$$\hat{\mu}^{\text{mrp}} = \frac{1}{N} \sum_s \frac{N_s^{\mathcal{P}}}{n_s^{\mathcal{R}}} n_s^{\mathcal{R}} \hat{\mu}_s = \frac{1}{N} \sum_s N_s^{\mathcal{P}} \hat{\mu}_s$$

DRP: a more surgical approach

Instead of relying on the model everywhere, use it to pick up slack from weighting

DRP: a more surgical approach

Instead of relying on the model everywhere, use it to pick up slack from weighting

$$\hat{\mu}^{\text{drp}} = \hat{\mu}^{\text{weight}}$$

DRP: a more surgical approach

Instead of relying on the model everywhere, use it to pick up slack from weighting

$$\hat{\mu}^{\text{drp}} = \hat{\mu}^{\text{weight}} + \frac{1}{N} \sum_s \hat{\mu}_s \times \underbrace{\left(N_s^{\mathcal{P}} - n_s^{\mathcal{R}} \hat{\gamma}_s\right)}_{\text{imbalance in cell } s}$$

DRP: a more surgical approach

Instead of relying on the model everywhere, use it to pick up slack from weighting

$$\hat{\mu}^{\text{drp}} = \hat{\mu}^{\text{weight}} + \frac{1}{N}\sum_s \hat{\mu}_s \times \underbrace{\left(N_s^{\mathcal{P}} - n_s^{\mathcal{R}}\hat{\gamma}_s\right)}_{\text{imbalance in cell } s}$$

$$= \hat{\mu}^{\text{mrp}}$$

DRP: a more surgical approach

Instead of relying on the model everywhere, use it to pick up slack from weighting

$$\hat{\mu}^{\text{drp}} = \hat{\mu}^{\text{weight}} + \frac{1}{N}\sum_s \hat{\mu}_s \times \underbrace{\left(N_s^{\mathcal{P}} - n_s^{\mathcal{R}}\hat{\gamma}_s\right)}_{\text{imbalance in cell } s}$$

$$= \hat{\mu}^{\text{mrp}} + \frac{1}{N}\sum_s n_s^{\mathcal{R}}\hat{\gamma}_s \times \underbrace{\left(\bar{Y}_s - \hat{\mu}_s\right)}_{\text{error in cell } s}$$

DRP: a more surgical approach

Instead of relying on the model everywhere, use it to pick up slack from weighting

$$\hat{\mu}^{\text{drp}} = \hat{\mu}^{\text{weight}} + \frac{1}{N}\sum_s \hat{\mu}_s \times \underbrace{\left(N_s^{\mathcal{P}} - n_s^{\mathcal{R}}\hat{\gamma}_s\right)}_{\text{imbalance in cell } s}$$

$$= \hat{\mu}^{\text{mrp}} + \frac{1}{N}\sum_s n_s^{\mathcal{R}}\hat{\gamma}_s \times \underbrace{\left(\bar{Y}_s - \hat{\mu}_s\right)}_{\text{error in cell } s}$$

Related to double robust and bias-corrected estimators

[Cassel et al., 1976; Robins et al., 1994; Abadie and Imbens, 2006; Hirshberg and Wager, 2019]

- Relies on outcome model in cells where weighting doesn't get it right
- Relies on weights to adjust cells where model is off
- If post-stratifying, collapses to weighting estimator $\hat{\mu}^{\text{drp}} = \hat{\mu}^{\text{weight}}$

Estimation error

Estimation error depends on how good the weights and the model are together

$$\left|\hat{\mu}^{\mathsf{drp}} - \mu\right| \lesssim \frac{1}{N}\sqrt{\sum_s n_s(\mu_s - \hat{\mu}_s)^2}\sqrt{\sum_s (n_s^{\mathcal{R}}\gamma_s - N_s^{\mathcal{P}})^2} + \frac{1}{N}\|\hat{\gamma}\|_2\|e\|_2$$

Estimation error

Estimation error depends on how good the weights and the model are together

$$\left|\hat{\mu}^{\text{drp}} - \mu\right| \lesssim \frac{1}{N}\sqrt{\sum_s n_s(\mu_s - \hat{\mu}_s)^2}\sqrt{\sum_s (n_s^{\mathcal{R}}\gamma_s - N_s^{\mathcal{P}})^2} + \frac{1}{N}\|\hat{\gamma}\|_2\|e\|_2$$

Special case of linear model:

$$\left|\hat{\mu}^{\text{drp}} - \mu\right| \lesssim \frac{1}{N}\sum_{k=1}^{K}\|\hat{\beta}_k - \beta_k\|_2\left\|\sum_s D_s^k n_s^{\mathcal{R}}\gamma_s - D_s^k N_s^{\mathcal{P}}\right\|_2 + \frac{1}{N}\|\hat{\gamma}\|_2\|e\|_2$$

Example: 2016 presidential election

Pre-election Pew poll of vote intention

- Age, sex, race, region, party ID, education, income, born again Christian

Impute Republican vote share within each state

Ground truth: weighted CCES

Interactions should be important here [Kennedy et al., 2018]

Example: 2016 presidential election

Pre-election Pew poll of vote intention

- Age, sex, race, region, party ID, education, income, born again Christian

Impute Republican vote share within each state

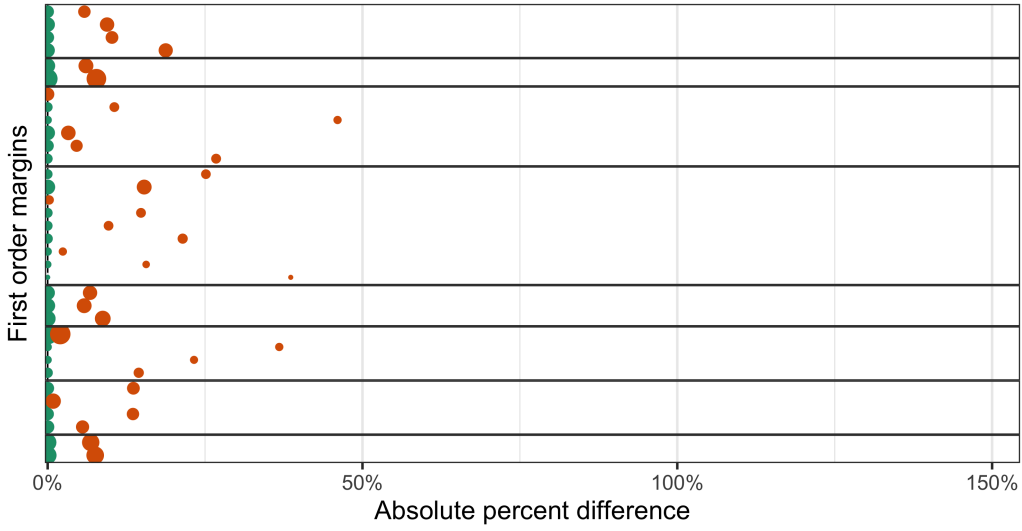
Ground truth: weighted CCES

Interactions should be important here [Kennedy et al., 2018]

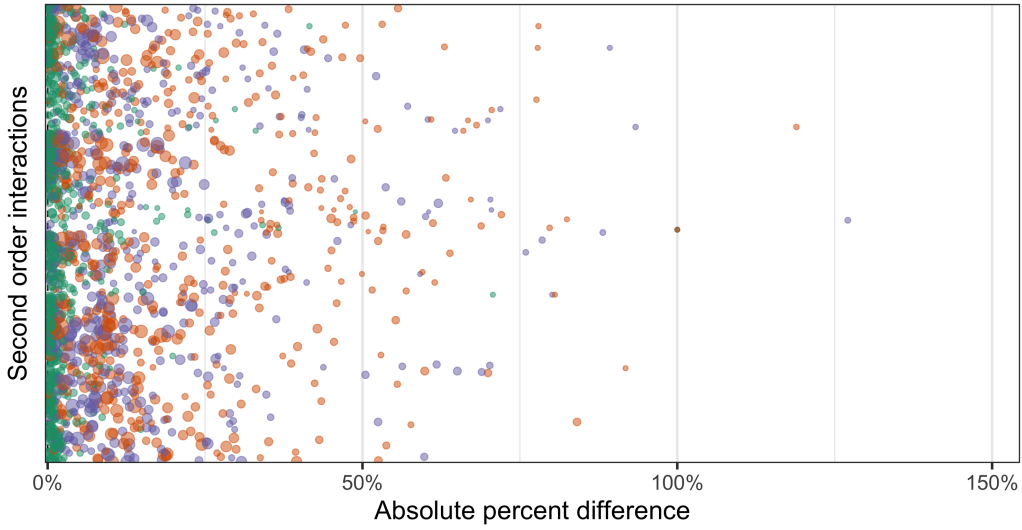
First, look at imbalance with full weighted CCES as target

$$\frac{\left|N_s^{\mathcal{P}} - n_s^{\mathcal{R}}\hat{\gamma}_s\right|}{N_s^{\mathcal{P}}}$$

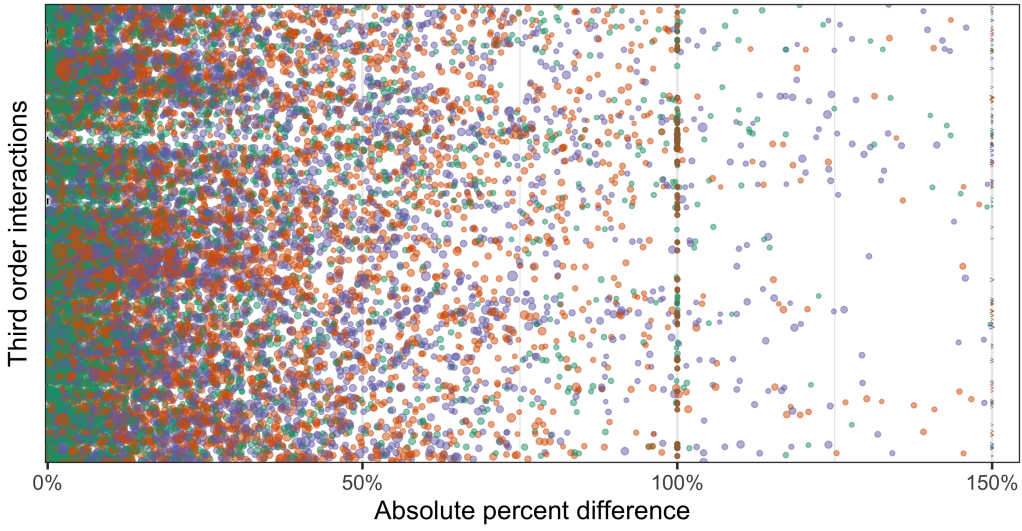
Both multilevel weighting and raking exactly match marginals...



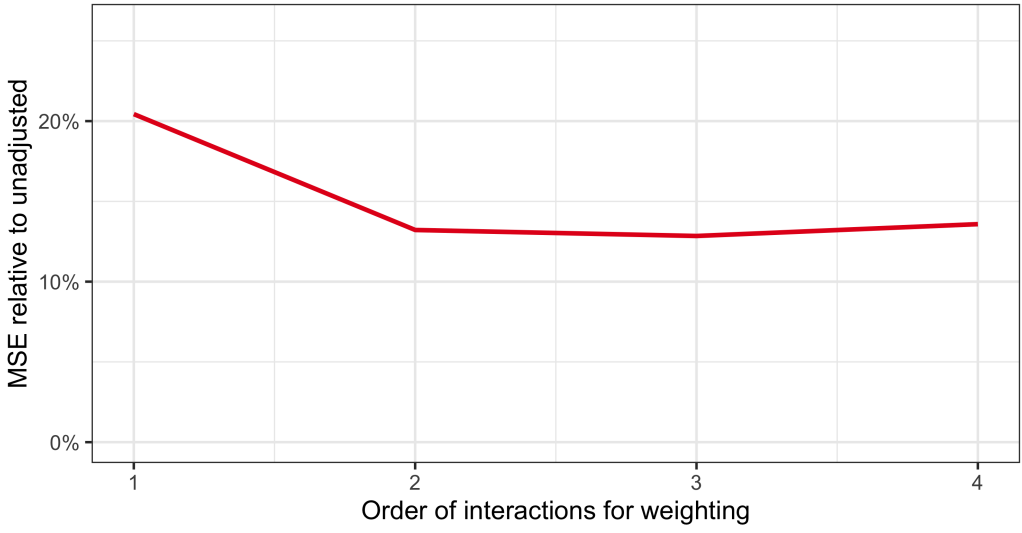
...but raking fails to balance higher order interactions



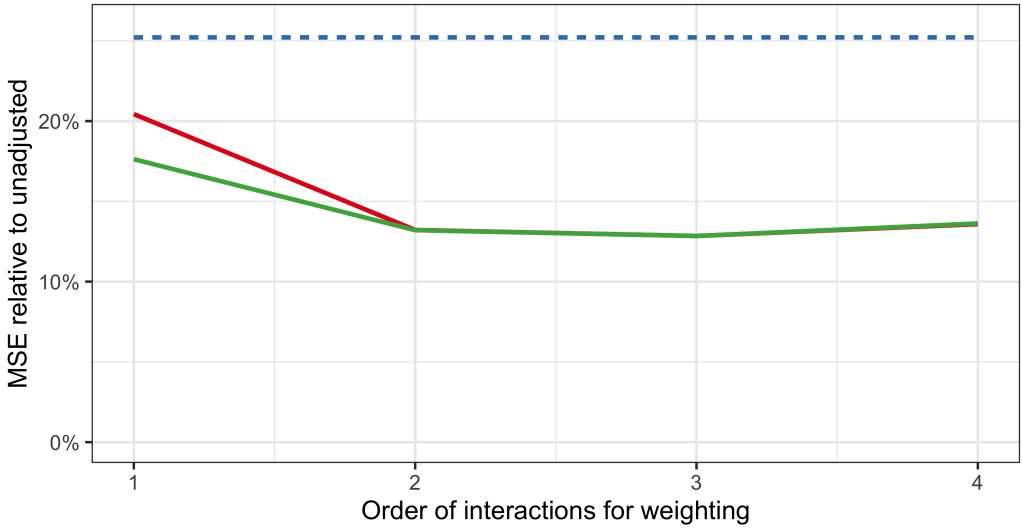
Approximate balance in 3rd order interactions



Weighting on higher order interactions helps



Double regression picks up some slack



Recap

Principled, parsimonious way of leveraging the value of interactions

- Multilevel weighting as middle ground between raking and post-stratification
- Dual view as multilevel model for non-response
- DRP adjusts cells where weighting misses the mark

Recap

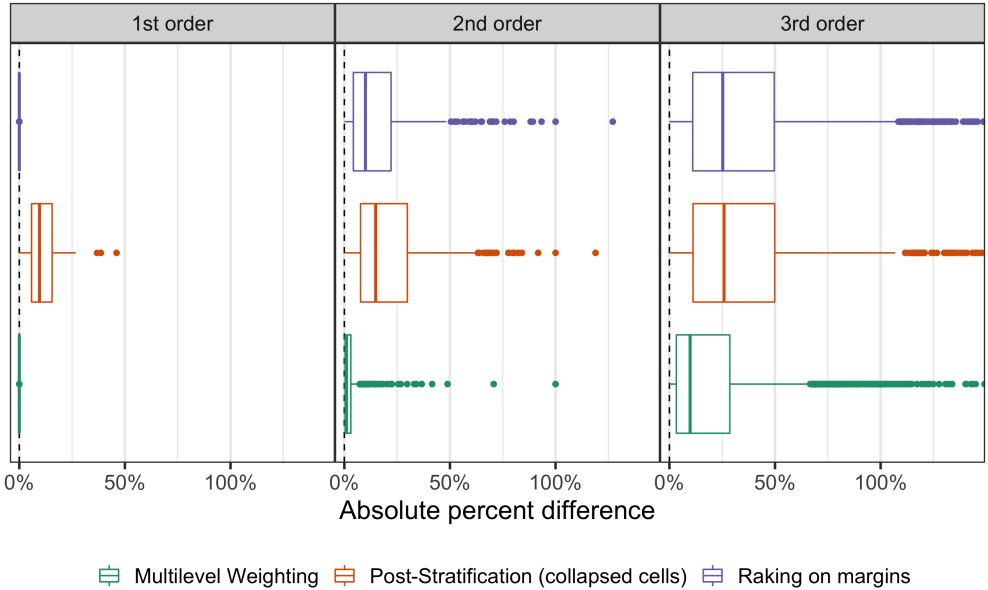
Principled, parsimonious way of leveraging the value of interactions

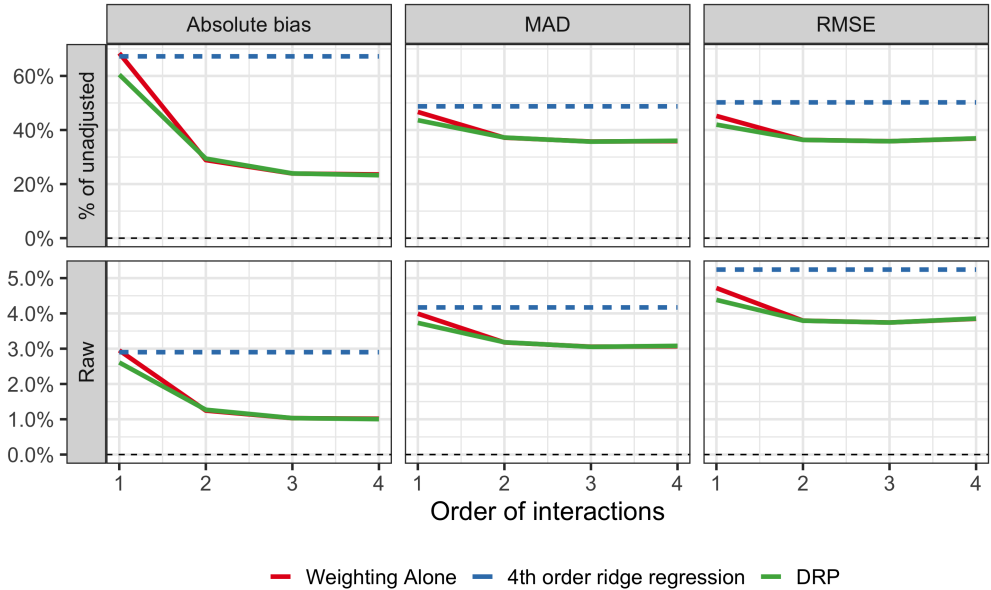
- Multilevel weighting as middle ground between raking and post-stratification
- Dual view as multilevel model for non-response
- DRP adjusts cells where weighting misses the mark

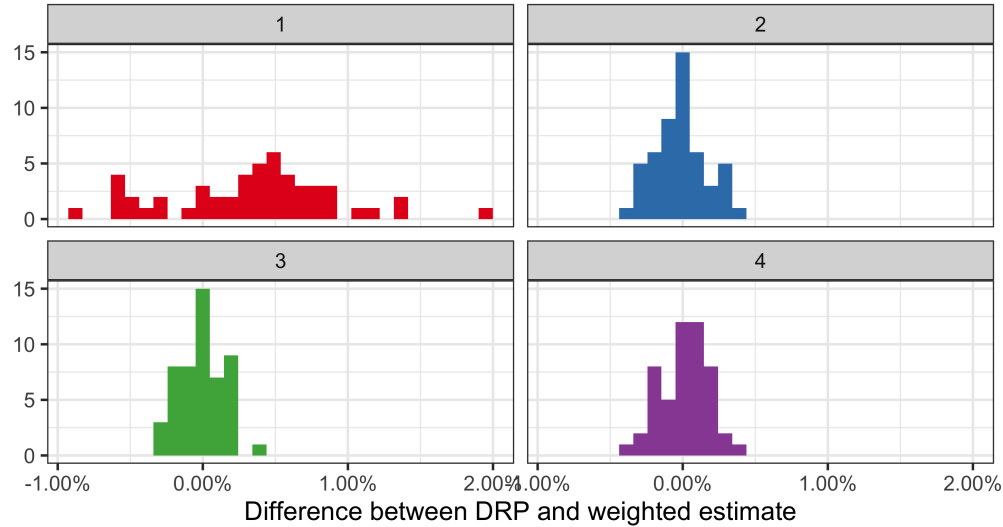
Thank you!

ebenmichael.github.io

Appendix


1st order
2nd order
3rd order
0%
50%
100%
Absolute percent difference
Multilevel Weighting
Post-Stratification (collapsed cells)
Raking on margins


Absolute bias
MAD
RMSE
% of unadjusted
Raw
60%
40%
20%
0%
5.0%
4.0%
3.0%
2.0%
1.0%
0.0%
1
2
3
4
Order of interactions
Weighting Alone
4th order ridge regression
DRP


1
2
3
4
15
10
5
0
-1.00%
0.00%
1.00%
2.00%
Difference between DRP and weighted estimate
Order of interactions
1
2
3
4

References I

Abadie, A. and Imbens, G. W. (2006). Large Sample Properties of Matching Estimators. *Econometrica*, 74(1):235–267.

Cassel, C. M., Sarndal, C.-E., and Wretman, J. H. (1976). Some results on generalized difference estimation and generalized regression estimation for finite populations. *Biometrika*, 63(3):615–620.

Chattopadhyay, A., Christopher H. Hase, and Zubizarreta, J. R. (2020). Balancing Versus Modeling Approaches to Weighting in Practice. *Statistics in Medicine*, in press.

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.

Hirshberg, D. and Wager, S. (2019). Augmented Minimax Linear Estimation.

Hirshberg, D. A., Maleki, A., and Zubizarreta, J. (2019). Minimax Linear Estimation of the Retargeted Mean.

References II

Kennedy, C., Blumenthal, M., Clement, S., Clinton, J. D., Durand, C., Franklin, C., McGeeney, K., Miringoff, L., Olson, K., Rivers, D., Saad, L., Witt, G. E., and Wlezien, C. (2018). An Evaluation of the 2016 Election Polls in the United States. *Public Opinion Quarterly*, 82(1):1–33.

Little, R. J. and Wu, M. M. (1991). Models for contingency tables with known margins when target and sampled populations differ. *Journal of the American Statistical Association*, 86(413):87–95.

Robins, J. M., Rotnitzky, A., Ping Zhao, L., and Ping ZHAO, L. (1994). Estimation of Regression Coefficients When Some Regressors are not Always Observed. *Journal of the American Statistical Association*, 89427:846–866.

Wang, Y. and Zubizarreta, J. R. (2020). Minimal dispersion approximately balancing weights: Asymptotic properties and practical considerations. *Biometrika*, 107(1):93–105.

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.