

Weighting-Like Diagnostics for Nonlinear Bayesian Hierarchical Models

Ryan Giordano, Alice Cima, Erin Hartman, Jared Murray, Avi Feller

September 2026 Iowa State University



Are US non-voters becoming more Republican?

Blue Rose research says yes:

“Politically disengaged voters have become much more Republican, and because less-engaged voters swung away from [Democrats], an expanded electorate meant a more Republican electorate.”

[Blu24]

(major professional pollsters)

On Data and Democracy says no:

“Claims of a decisive pro-Republican shift among the overall non-voting population are not supported by the most reliable, large-scale post-election data currently available.”

[Bon+25]

(major professional researchers)

Are US non-voters becoming more Republican?

Blue Rose research says yes:

“Politically disengaged voters have become much more Republican, and because less-engaged voters swung away from [Democrats], an expanded electorate meant a more Republican electorate.”

[Blu24]

(major professional pollsters)

On Data and Democracy says no:

“Claims of a decisive pro-Republican shift among the overall non-voting population are not supported by the most reliable, large-scale post-election data currently available.”

[Bon+25]

(major professional researchers)

-
- The problem is very hard (it's difficult to accurately poll non-voters)
 - Different data sources
 - *** **Different statistical methods**
 - Blue Rose uses Bayesian hierarchical modeling (MrP)
 - On Data and Democracy is using calibration weighting (CW)

Are US non-voters becoming more Republican?

Blue Rose research says yes:

“Politically disengaged voters have become much more Republican, and because less-engaged voters swung away from [Democrats], an expanded electorate meant a more Republican electorate.”

[Blu24]

(major professional pollsters)

On Data and Democracy says no:

“Claims of a decisive pro-Republican shift among the overall non-voting population are not supported by the most reliable, large-scale post-election data currently available.”

[Bon+25]

(major professional researchers)

-
- The problem is very hard (it’s difficult to accurately poll non-voters)
 - Different data sources
 - *** **Different statistical methods**
 - Blue Rose uses Bayesian hierarchical modeling (MrP)
 - On Data and Democracy is using calibration weighting (CW)

Our contribution

We define “MrP local equivalent weights” (MrPlew) that:

- Are easily computable from MCMC draws and standard software, and
- Provide MrP versions of key weighting estimator diagnostics.

⇒ **MrPlew provides direct comparisons between MrP and calibration weighting.**

Weighting (linear) estimators are great — they come with easy-to-understand diagnostics.

This talk is about making versions of such diagnostics for **complicated non-linear models**.

Weighting (linear) estimators are great — they come with easy-to-understand diagnostics.

This talk is about making versions of such diagnostics for **complicated non-linear models**.

The key idea is to convert the diagnostic into a local sensitivity analysis of this form:

1. Assume your initial model was accurate
2. Select some perturbation your model should be able to capture
3. Use local sensitivity to detect whether the change is what you expect
4. If the change is not what you expect, either (1) or (2) was wrong

Weighting (linear) estimators are great — they come with easy-to-understand diagnostics.

This talk is about making versions of such diagnostics for **complicated non-linear models**.

The key idea is to convert the diagnostic into a local sensitivity analysis of this form:

1. Assume your initial model was accurate
2. Select some perturbation your model should be able to capture
3. Use local sensitivity to detect whether the change is what you expect
4. If the change is not what you expect, either (1) or (2) was wrong

I'll do this carefully for covariate balance and MCMC.

But many other variants are possible!

- Introduce the statistical problem
 - Contrast calibration weighting and MrP
 - Prior work: Equivalent weights for linear models
 - Equivalent weights for non-linear models
 - Our task: Rigorously justify using locally equivalent weights for diagnostics

- Introduce the statistical problem
 - Contrast calibration weighting and MrP
 - Prior work: Equivalent weights for linear models
 - Equivalent weights for non-linear models
 - Our task: Rigorously justify using locally equivalent weights for diagnostics
- Locally equivalent weights for frequentist variance estimation

- Introduce the statistical problem
 - Contrast calibration weighting and MrP
 - Prior work: Equivalent weights for linear models
 - Equivalent weights for non-linear models
 - Our task: Rigorously justify using locally equivalent weights for diagnostics
- Locally equivalent weights for frequentist variance estimation
- Locally equivalent weights for covariate balance
 - Describe classical covariate balance
 - Introduce a version of covariate balance as sensitivity analysis
 - Examples of real-world results

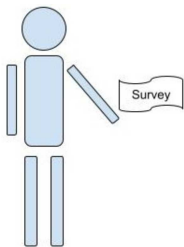
- Introduce the statistical problem
 - Contrast calibration weighting and MrP
 - Prior work: Equivalent weights for linear models
 - Equivalent weights for non-linear models
 - Our task: Rigorously justify using locally equivalent weights for diagnostics
- Locally equivalent weights for frequentist variance estimation
- Locally equivalent weights for covariate balance
 - Describe classical covariate balance
 - Introduce a version of covariate balance as sensitivity analysis
 - Examples of real-world results
- Locally equivalent weights subgroup contribution
- Theoretical support and challenges

The basic problem

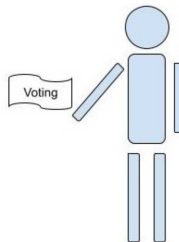
We have a survey population, for whom we observe:

- Covariates $\mathbf{x}$ (e.g. race, gender, zip code, age, education level)
- Responses y (e.g. A binary response to “do you support candidate Z”)

We want the average response in a target population, in which we observe only covariates.



Observe $(\mathbf{x}_i, y_i)$ for $i = 1, \dots, N_S$



Observe $\mathbf{x}_j$ for $j = 1, \dots, N_T$

¹Photo copyright: Mark Taylor / naturepl.com

The basic problem

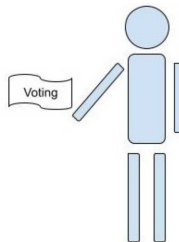
We have a survey population, for whom we observe:

- Covariates $\mathbf{x}$ (e.g. race, gender, zip code, age, education level)
- Responses y (e.g. A binary response to “do you support candidate Z”)

We want the average response in a target population, in which we observe only covariates.



Observe $(\mathbf{x}_i, y_i)$ for $i = 1, \dots, N_S$



Observe $\mathbf{x}_j$ for $j = 1, \dots, N_T$

The problem is that the populations may be very different, maybe leading to bias. ¹

¹Photo copyright: Mark Taylor / naturepl.com

The basic problem

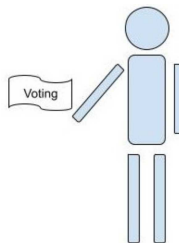
We have a survey population, for whom we observe:

- Covariates $\mathbf{x}$ (e.g. race, gender, zip code, age, education level)
- Responses y (e.g. A binary response to “do you support candidate Z”)

We want the average response in a target population, in which we observe only covariates.



Observe $(\mathbf{x}_i, y_i)$ for $i = 1, \dots, N_S$



Observe $\mathbf{x}_j$ for $j = 1, \dots, N_T$

The problem is that the populations may be very different, maybe leading to bias. ¹

How can we use the covariates to say something about the target responses?

¹Photo copyright: Mark Taylor / naturepl.com

Two approaches

We want $\mu := \frac{1}{N_T} \sum_{j=1}^{N_T} y_j$, but don't observe target y_j . Let $Y_S = \{y_1, \dots, y_{N_S}\}$.

- Assume $p(y|\mathbf{x})$ is the same in both populations,
- But the distribution of $\mathbf{x}$ may be different in the survey and target.

Two approaches

We want $\mu := \frac{1}{N_T} \sum_{j=1}^{N_T} y_j$, but don't observe target y_j . Let $Y_S = \{y_1, \dots, y_{N_S}\}$.

- Assume $p(y|\mathbf{x})$ is the same in both populations,
- But the distribution of $\mathbf{x}$ may be different in the survey and target.

Calibration weighting

- Choose “calibration weights” w_i
using only the regressors $\mathbf{x}$
(e.g. raking weights)

Bayesian hierarchical modeling (MrP)

- Choose $\mathbb{E}[y|\mathbf{x}, \theta] = m(\theta^\top \mathbf{x})$,
choose prior $\mathcal{P}(\theta|\Sigma)\mathcal{P}(\Sigma)$
(e.g. Hierarchical logistic regression)

Two approaches

We want $\mu := \frac{1}{N_T} \sum_{j=1}^{N_T} y_j$, but don't observe target y_j . Let $Y_S = \{y_1, \dots, y_{N_S}\}$.

- Assume $p(y|\mathbf{x})$ is the same in both populations,
- But the distribution of $\mathbf{x}$ may be different in the survey and target.

Calibration weighting

- Choose “calibration weights” w_i
using only the regressors $\mathbf{x}$
(e.g. raking weights)
- Take $\hat{\mu}^{\text{WGT}}(Y_S) = \frac{1}{N_S} \sum_{i=1}^{N_S} w_i y_i$

Bayesian hierarchical modeling (MrP)

- Choose $\mathbb{E}[y|\mathbf{x}, \theta] = m(\theta^\top \mathbf{x})$,
choose prior $\mathcal{P}(\theta|\Sigma)\mathcal{P}(\Sigma)$
(e.g. Hierarchical logistic regression)
- Take $\hat{y}_j = \mathbb{E}_{\mathcal{P}(\theta|\text{Survey data})}[y|\mathbf{x}_j]$ and
 $\hat{\mu}^{\text{MrP}}(Y_S) = \frac{1}{N_T} \sum_{j=1}^{N_T} \hat{y}_j$

Two approaches

We want $\mu := \frac{1}{N_T} \sum_{j=1}^{N_T} y_j$, but don't observe target y_j . Let $Y_S = \{y_1, \dots, y_{N_S}\}$.

- Assume $p(y|\mathbf{x})$ is the same in both populations,
- But the distribution of $\mathbf{x}$ may be different in the survey and target.

Calibration weighting

- ▶ Choose “calibration weights” w_i
using only the regressors $\mathbf{x}$
(e.g. raking weights)
- ▶ Take $\hat{\mu}^{\text{WGT}}(Y_S) = \frac{1}{N_S} \sum_{i=1}^{N_S} w_i y_i$
- ▶ Dependence on y_i is clear

Bayesian hierarchical modeling (MrP)

- ▶ Choose $\mathbb{E}[y|\mathbf{x}, \theta] = m(\theta^\top \mathbf{x})$,
choose prior $\mathcal{P}(\theta|\Sigma)\mathcal{P}(\Sigma)$
(e.g. Hierarchical logistic regression)
- ▶ Take $\hat{y}_j = \mathbb{E}_{\mathcal{P}(\theta|\text{Survey data})}[y|\mathbf{x}_j]$ and
 $\hat{\mu}^{\text{MrP}}(Y_S) = \frac{1}{N_T} \sum_{j=1}^{N_T} \hat{y}_j$
- ▶ Dependence on y_i very complicated
(Typically via MCMC draws from
 $\mathcal{P}(\theta|\text{Survey data})$)

Two approaches

We want $\mu := \frac{1}{N_T} \sum_{j=1}^{N_T} y_j$, but don't observe target y_j . Let $Y_S = \{y_1, \dots, y_{N_S}\}$.

- Assume $p(y|\mathbf{x})$ is the same in both populations,
- But the distribution of $\mathbf{x}$ may be different in the survey and target.

Calibration weighting

- ▶ Choose “calibration weights” w_i using only the regressors $\mathbf{x}$ (e.g. raking weights)
- ▶ Take $\hat{\mu}^{\text{WGT}}(Y_S) = \frac{1}{N_S} \sum_{i=1}^{N_S} w_i y_i$
 - ▶ Dependence on y_i is clear
- ▶ Weights give interpretable diagnostics:
 - Frequentist variability
 - Regressor balance
 - Subgroup contribution

Bayesian hierarchical modeling (MrP)

- ▶ Choose $\mathbb{E}[y|\mathbf{x}, \theta] = m(\theta^\top \mathbf{x})$, choose prior $\mathcal{P}(\theta|\Sigma)\mathcal{P}(\Sigma)$ (e.g. Hierarchical logistic regression)
- ▶ Take $\hat{y}_j = \mathbb{E}_{\mathcal{P}(\theta|\text{Survey data})}[y|\mathbf{x}_j]$ and $\hat{\mu}^{\text{MrP}}(Y_S) = \frac{1}{N_T} \sum_{j=1}^{N_T} \hat{y}_j$
 - ▶ Dependence on y_i very complicated (Typically via MCMC draws from $\mathcal{P}(\theta|\text{Survey data})$)
- ▶ **Black box**

Two approaches

We want $\mu := \frac{1}{N_T} \sum_{j=1}^{N_T} y_j$, but don't observe target y_j . Let $Y_S = \{y_1, \dots, y_{N_S}\}$.

- Assume $p(y|\mathbf{x})$ is the same in both populations,
- But the distribution of $\mathbf{x}$ may be different in the survey and target.

Calibration weighting

- ▶ Choose “calibration weights” w_i using only the regressors $\mathbf{x}$ (e.g. raking weights)
- ▶ Take $\hat{\mu}^{\text{WGT}}(Y_S) = \frac{1}{N_S} \sum_{i=1}^{N_S} w_i y_i$
 - ▶ Dependence on y_i is clear
- ▶ Weights give interpretable diagnostics:
 - Frequentist variability
 - Regressor balance
 - Subgroup contribution

Bayesian hierarchical modeling (MrP)

- ▶ Choose $\mathbb{E}[y|\mathbf{x}, \theta] = m(\theta^\top \mathbf{x})$, choose prior $\mathcal{P}(\theta|\Sigma)\mathcal{P}(\Sigma)$ (e.g. Hierarchical logistic regression)
 - ▶ Take $\hat{y}_j = \mathbb{E}_{\mathcal{P}(\theta|\text{Survey data})}[y|\mathbf{x}_j]$ and $\hat{\mu}^{\text{MrP}}(Y_S) = \frac{1}{N_T} \sum_{j=1}^{N_T} \hat{y}_j$
 - ▶ Dependence on y_i very complicated (Typically via MCMC draws from $\mathcal{P}(\theta|\text{Survey data})$)
 - ▶ **Black box**
- ← Today, we'll open the box and provide MrP analogues of all these diagnostics

Prior work: Equivalent weights for linear models

Calibration weighting estimators:

$$\hat{\mu}^{\text{WGT}}(Y_S) = \frac{1}{N_S} \sum_{i=1}^{N_S} w_i y_i.$$

MrP estimators:

$$\hat{\mu}^{\text{MrP}}(Y_S) = \frac{1}{N_T} \sum_{j=1}^{N_T} \hat{y}_j$$

Prior work: Equivalent weights for linear models

Calibration weighting estimators:

$$\hat{\mu}^{\text{WGT}}(Y_S) = \frac{1}{N_S} \sum_{i=1}^{N_S} w_i y_i.$$

MrP estimators:

$$\hat{\mu}^{\text{MrP}}(Y_S) = \frac{1}{N_T} \sum_{j=1}^{N_T} \hat{y}_j$$

...sometimes coincide!

Gelman [Gel07b] observes that MrP is a weighting estimator when $\hat{y}$ is computed with OLS:

$$\hat{\mu}^{\text{MrP}}(Y_S) = \frac{1}{N_T} \sum_{j=1}^{N_T} \hat{y}_j = \frac{1}{N_T} \sum_{j=1}^{N_T} \underbrace{\mathbf{x}_j^T \hat{\beta}}_{\text{Linear in } Y_S} = \frac{1}{N_S} \sum_{i=1}^{N_S} w_i^{\text{OLS}} y_i$$

Prior work: Equivalent weights for linear models

Calibration weighting estimators:

$$\hat{\mu}^{\text{WGT}}(Y_S) = \frac{1}{N_S} \sum_{i=1}^{N_S} w_i y_i.$$

MrP estimators:

$$\hat{\mu}^{\text{MrP}}(Y_S) = \frac{1}{N_T} \sum_{j=1}^{N_T} \hat{y}_j$$

...sometimes coincide!

Gelman [Gel07b] observes that MrP is a weighting estimator when $\hat{y}$ is computed with OLS:

$$\hat{\mu}^{\text{MrP}}(Y_S) = \frac{1}{N_T} \sum_{j=1}^{N_T} \hat{y}_j = \frac{1}{N_T} \sum_{j=1}^{N_T} \underbrace{\mathbf{x}_j^T \hat{\beta}}_{\text{Linear in } Y_S} = \frac{1}{N_S} \sum_{i=1}^{N_S} w_i^{\text{OLS}} y_i$$

Most existing literature on comparing weighting and MrP focus on such linear models. ²

But what if you use a non-linear link function? Or a hierarchical model?

“It would also be desirable to use nonlinear methods ... but then it would seem difficult to construct even approximately equivalent weights. Weighting and fully nonlinear models would seem to be completely incompatible methods.” — [Gel07a]

²For example, Gelman [Gel07b], B., F., and H. [BFH21], and Chattopadhyay and Zubizarreta [CZ23].

Prior work: Equivalent weights for linear models

Calibration weighting estimators:

$$\hat{\mu}^{\text{WGT}}(Y_S) = \frac{1}{N_S} \sum_{i=1}^{N_S} w_i y_i.$$

MrP estimators:

$$\hat{\mu}^{\text{MrP}}(Y_S) = \frac{1}{N_T} \sum_{j=1}^{N_T} \hat{y}_j$$

...sometimes coincide!

Gelman [Gel07b] observes that MrP is a weighting estimator when $\hat{y}$ is computed with OLS:

$$\hat{\mu}^{\text{MrP}}(Y_S) = \frac{1}{N_T} \sum_{j=1}^{N_T} \hat{y}_j = \frac{1}{N_T} \sum_{j=1}^{N_T} \underbrace{\mathbf{x}_j^T \hat{\beta}}_{\text{Linear in } Y_S} = \frac{1}{N_S} \sum_{i=1}^{N_S} w_i^{\text{OLS}} y_i$$

Useful note. We know the OLS formula, so we can write out w_i^{OLS} by hand.

But we only had access to $\hat{\mu}^{\text{MrP}}(Y_S)$ as a black box, we could compute

$$w_i^{\text{OLS}} = N_S \frac{\partial \hat{\mu}^{\text{MrP}}(Y_S)}{\partial y_i}.$$

Local weights for nonlinear hierarchical logistic regression MrP

- Suppose the model is $m(\mathbf{x}^\top \theta) = \text{Logistic}(\mathbf{x}^\top \theta)$.
- Set a hierarchical prior $\mathcal{P}(\theta|\Sigma)\mathcal{P}(\Sigma)$, use MCMC to draw from $\mathcal{P}(\theta|\text{Survey data})$.
- MrP is $\hat{\mu}^{\text{MrP}}(Y_{\mathcal{S}}) = \frac{1}{N_T} \sum_{j=1}^{N_T} \mathbb{E}_{\mathcal{P}(\theta|\text{Survey data})} \left[m(\mathbf{x}_j^\top \theta) \right]$.

No reason to think $Y_{\mathcal{S}} \mapsto \hat{\mu}^{\text{MrP}}(Y_{\mathcal{S}})$ is even approximately **globally** linear.

Local weights for nonlinear hierarchical logistic regression MrP

- Suppose the model is $m(\mathbf{x}^\top \theta) = \text{Logistic}(\mathbf{x}^\top \theta)$.
- Set a hierarchical prior $\mathcal{P}(\theta|\Sigma)\mathcal{P}(\Sigma)$, use MCMC to draw from $\mathcal{P}(\theta|\text{Survey data})$.
- MrP is $\hat{\mu}^{\text{MrP}}(Y_S) = \frac{1}{N_T} \sum_{j=1}^{N_T} \mathbb{E}_{\mathcal{P}(\theta|\text{Survey data})} \left[m(\mathbf{x}_j^\top \theta) \right]$.

No reason to think $Y_S \mapsto \hat{\mu}^{\text{MrP}}(Y_S)$ is even approximately **globally** linear.

MrP weights for MCMC

$$w_i^{\text{MrP}} := N_S \frac{\partial \hat{\mu}^{\text{MrP}}(Y_S)}{\partial y_i} = N_S \frac{1}{N_T} \sum_{j=1}^{N_T} \underbrace{\text{Cov}_{\mathcal{P}(\theta|\text{Survey data})} \left(m(\mathbf{x}_j^\top \theta), \theta^\top \mathbf{x}_i \right)}_{\text{Can estimate without rerunning MCMC!}}$$

But can still compute $w_i^{\text{MrP}} := N_S \frac{\partial \hat{\mu}^{\text{MrP}}(Y_S)}{\partial y_i}$ using Bayesian sensitivity analysis! [DF86; Gus96; Efr15; GBJ18]

Local weights for nonlinear hierarchical logistic regression MrP

- Suppose the model is $m(\mathbf{x}^\top \theta) = \text{Logistic}(\mathbf{x}^\top \theta)$.
- Set a hierarchical prior $\mathcal{P}(\theta|\Sigma)\mathcal{P}(\Sigma)$, use MCMC to draw from $\mathcal{P}(\theta|\text{Survey data})$.
- MrP is $\hat{\mu}^{\text{MrP}}(Y_S) = \frac{1}{N_T} \sum_{j=1}^{N_T} \mathbb{E}_{\mathcal{P}(\theta|\text{Survey data})} \left[m(\mathbf{x}_j^\top \theta) \right]$.

No reason to think $Y_S \mapsto \hat{\mu}^{\text{MrP}}(Y_S)$ is even approximately **globally** linear.

MrP weights for MCMC

$$w_i^{\text{MrP}} := N_S \frac{\partial \hat{\mu}^{\text{MrP}}(Y_S)}{\partial y_i} = N_S \frac{1}{N_T} \sum_{j=1}^{N_T} \underbrace{\text{Cov}_{\mathcal{P}(\theta|\text{Survey data})} \left(m(\mathbf{x}_j^\top \theta), \theta^\top \mathbf{x}_i \right)}_{\text{Can estimate without rerunning MCMC!}}$$

MrP locally equivalent weights (MrPlew)

For new data $\tilde{Y}_S$, form a **MrP locally equivalent weighting**:

$$\hat{\mu}^{\text{MrP}}(\tilde{Y}_S) \approx \hat{\mu}^{\text{MrP}}(Y_S) + \sum_{i=1}^{N_S} w_i^{\text{MrP}} (\tilde{y}_i - y_i)$$

Local weights for nonlinear hierarchical logistic regression MrP

- Suppose the model is $m(\mathbf{x}^\top \theta) = \text{Logistic}(\mathbf{x}^\top \theta)$.
- Set a hierarchical prior $\mathcal{P}(\theta|\Sigma)\mathcal{P}(\Sigma)$, use MCMC to draw from $\mathcal{P}(\theta|\text{Survey data})$.
- MrP is $\hat{\mu}^{\text{MrP}}(Y_S) = \frac{1}{N_T} \sum_{j=1}^{N_T} \mathbb{E}_{\mathcal{P}(\theta|\text{Survey data})} \left[m(\mathbf{x}_j^\top \theta) \right]$.

No reason to think $Y_S \mapsto \hat{\mu}^{\text{MrP}}(Y_S)$ is even approximately **globally** linear.

MrP weights for MCMC

$$w_i^{\text{MrP}} := N_S \frac{\partial \hat{\mu}^{\text{MrP}}(Y_S)}{\partial y_i} = N_S \frac{1}{N_T} \sum_{j=1}^{N_T} \underbrace{\text{Cov}_{\mathcal{P}(\theta|\text{Survey data})} \left(m(\mathbf{x}_j^\top \theta), \theta^\top \mathbf{x}_i \right)}_{\text{Can estimate without rerunning MCMC!}}$$

MrP locally equivalent weights (MrPlew)

For new data $\tilde{Y}_S$, form a **MrP locally equivalent weighting**:

$$\hat{\mu}^{\text{MrP}}(\tilde{Y}_S) \approx \hat{\mu}^{\text{MrP}}(Y_S) + \sum_{i=1}^{N_S} w_i^{\text{MrP}} (\tilde{y}_i - y_i)$$

Our task is to rigorously show that even such local weights can be meaningfully used diagnostically in the same ways we use global weights.

Real Data: Marital Name Change Survey

Analysis of changing names after marriage².

- **Target population:** ACS survey of US population 2017–2022
- **Survey population:** Marital Name Change Survey (from Twitter)
- **Respose:** Did the female partner keep their name after marriage?
- For regressors, use bins of age, education, state, and decade married.

MrP computed with brms [Bür17]:

```
kept_name ~ (1 | age_group) + (1 | educ_group) + (1 | state_name) + (1 | decade_married)
```

CW used raking on coarsened regressor marginals (survey::calibrate from Lumley [Lum24])

$$N_S = 4,364 \quad N_T = 64,784,329$$

Uncorrected survey mean: $\frac{1}{N_S} \sum_{i=1}^{N_S} y_i = 0.462$

Raking: $\hat{\mu}^{\text{WGT}}(Y_S) = 0.263$

MrP: $\hat{\mu}^{\text{MrP}}(Y_S) = 0.288 \quad (\text{Post. sd} = 0.0165)$

²Based on Alexander [Ale19], Cohen [Coh19], and Ruggles et al. [Rug+24].

Real Data: Marital Name Change Survey

Analysis of changing names after marriage².

- **Target population:** ACS survey of US population 2017–2022
- **Survey population:** Marital Name Change Survey (from Twitter)
- **Respose:** Did the female partner keep their name after marriage?
- For regressors, use bins of age, education, state, and decade married.

MrP computed with brms [Bür17]:

```
kept_name ~ (1 | age_group) + (1 | educ_group) + (1 | state_name) + (1 | decade_married)
```

CW used raking on coarsened regressor marginals (survey::calibrate from Lumley [Lum24])

$$N_S = 4,364 \quad N_T = 64,784,329$$

Uncorrected survey mean: $\frac{1}{N_S} \sum_{i=1}^{N_S} y_i = 0.462$

Raking: $\hat{\mu}^{\text{WGT}}(Y_S) = 0.263$

MrP: $\hat{\mu}^{\text{MrP}}(Y_S) = 0.288 \quad (\text{Post. sd} = 0.0165)$



²Based on Alexander [Ale19], Cohen [Coh19], and Ruggles et al. [Rug+24].

The weights can look very different!

The weights can look very different! Does this mean anything?

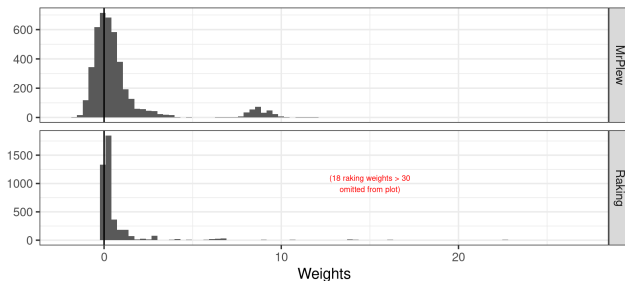


Figure 1: Weight comparison for the Name Change dataset

³See G. and Broderick [GB24]. For weighting variances, see, e.g., Deville, Särndal, and Sautory [DSS93] and Fuller [Ful11].

The weights can look very different!

The weights can look very different! Does this mean anything?

Does the spread relate to frequentist variance?

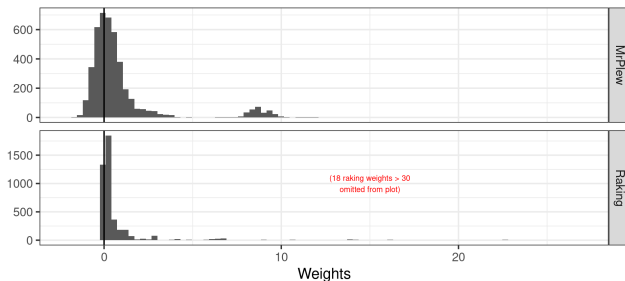


Figure 1: Weight comparison for the Name Change dataset

³See G. and Broderick [GB24]. For weighting variances, see, e.g., Deville, Särndal, and Sautory [DSS93] and Fuller [Ful11].

The weights can look very different!

The weights can look very different! Does this mean anything?

Does the spread relate to frequentist variance?

Theorem [Gio+26]:

Yes: The “spread” relates to frequentist variance just as in weighting estimators, as long as the response y is a sufficient statistic in the model.

This is essentially a corollary of our earlier work on the Bayesian infinitesimal jackknife.³

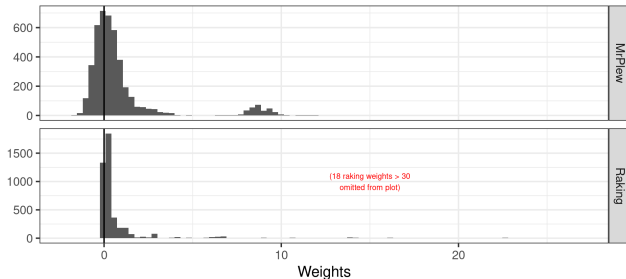


Figure 1: Weight comparison for the Name Change dataset

³See G. and Broderick [GB24]. For weighting variances, see, e.g., Deville, Särndal, and Sautory [DSS93] and Fuller [Ful11].

Standard error estimation experiment

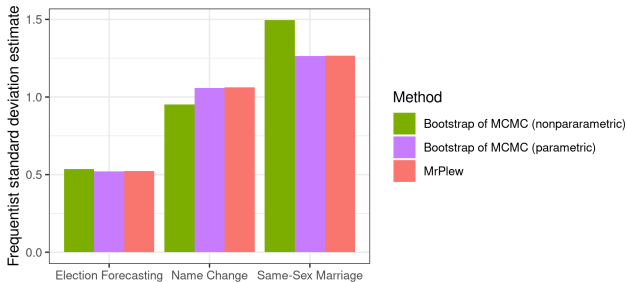


Figure 2: Estimates of the frequentist standard deviation of $\sqrt{N_S} \hat{\mu}^{\text{MrP}}$

Standard error estimation experiment

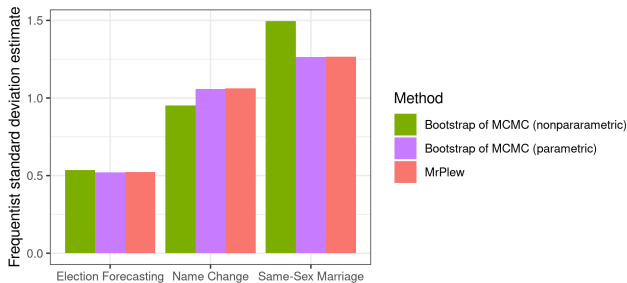


Figure 2: Estimates of the frequentist standard deviation of $\sqrt{N_S} \hat{\mu}^{\text{MrP}}$

Running 100 MCMC bootstraps: $\approx 100 \times 95$ minutes

Computing approximate weights: A few seconds

The weights can look very different!

The weights can look very different! Does this mean anything?

Yes: The “spread” relates to frequentist variance just as in weighting estimators.

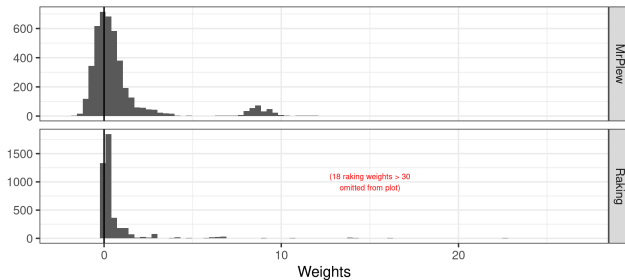


Figure 3: Weight comparison for the Name Change dataset

The weights can look very different!

The weights can look very different! Does this mean anything?

Yes: The “spread” relates to frequentist variance just as in weighting estimators.

What about covariate balance?

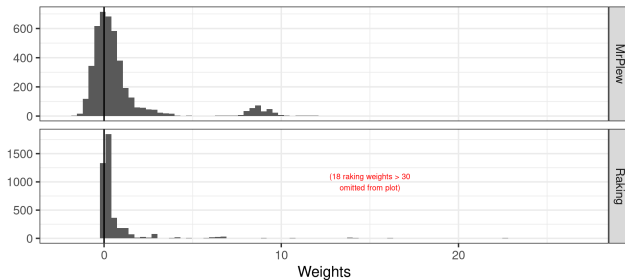


Figure 3: Weight comparison for the Name Change dataset

Introduction to covariate balance: What are we weighting for?⁴

$$\text{Target average response} = \frac{1}{N_T} \sum_{j=1}^{N_T} y_j \approx \frac{1}{N_S} \sum_{i=1}^{N_S} w_i y_i = \text{Weighted survey average response}$$

We can't check this, because we don't observe y_j .

⁴Pun borrowed from Solon, Haider, and Wooldridge [SHW15]

Introduction to covariate balance: What are we weighting for?⁴

$$\text{Target average response} = \frac{1}{N_T} \sum_{j=1}^{N_T} y_j \approx \frac{1}{N_S} \sum_{i=1}^{N_S} w_i y_i = \text{Weighted survey average response}$$

We can't check this, because we don't observe y_j . But we can check whether:

$$\frac{1}{N_T} \sum_{j=1}^{N_T} \mathbf{x}_j \stackrel{\text{check}}{=} \frac{1}{N_S} \sum_{i=1}^{N_S} w_i \mathbf{x}_i$$

Weights that pass this check satisfy “covariate balance” for $\mathbf{x}$.

⁴Pun borrowed from Solon, Haider, and Wooldridge [SHW15]

Introduction to covariate balance: What are we weighting for?⁴

$$\text{Target average response} = \frac{1}{N_T} \sum_{j=1}^{N_T} y_j \approx \frac{1}{N_S} \sum_{i=1}^{N_S} w_i y_i = \text{Weighted survey average response}$$

We can't check this, because we don't observe y_j . But we can check whether:

$$\frac{1}{N_T} \sum_{j=1}^{N_T} \mathbf{x}_j \stackrel{\text{check}}{=} \frac{1}{N_S} \sum_{i=1}^{N_S} w_i \mathbf{x}_i$$

Weights that pass this check satisfy “covariate balance” for $\mathbf{x}$.

You can check covariate balance for any weighting estimator, and any function $f(\mathbf{x})$.

Raking calibration weights aim to exactly balance some set of regressors.

⁴Pun borrowed from Solon, Haider, and Wooldridge [SHW15]

Introduction to covariate balance: What are we weighting for?⁴

$$\text{Target average response} = \frac{1}{N_T} \sum_{j=1}^{N_T} y_j \approx \frac{1}{N_S} \sum_{i=1}^{N_S} w_i y_i = \text{Weighted survey average response}$$

We can't check this, because we don't observe y_j . But we can check whether:

$$\frac{1}{N_T} \sum_{j=1}^{N_T} \mathbf{x}_j \stackrel{\text{check}}{=} \frac{1}{N_S} \sum_{i=1}^{N_S} w_i \mathbf{x}_i$$

Weights that pass this check satisfy “covariate balance” for $\mathbf{x}$.

You can check covariate balance for any weighting estimator, and any function $f(\mathbf{x})$.

Raking calibration weights aim to exactly balance some set of regressors.

Key question

Covariate balance seems inherently linear.

How can there be an analogue of it for non-linear estimators?

⁴Pun borrowed from Solon, Haider, and Wooldridge [SHW15]

Key question

Covariate balance seems inherently linear.

How can there be an analogue of it for non-linear estimators?

Balance checks as local sensitivity

Key question

Covariate balance seems inherently linear.

How can there be an analogue of it for non-linear estimators?

Key idea:

The key idea is to convert the diagnostic into a local sensitivity analysis of this form:

1. Assume your initial model was accurate
2. Select some perturbation your model should be able to capture
3. Use local sensitivity to detect whether the change is what you expect
4. If the change is not what you expect, either (1) or (2) was wrong

Balance checks as local sensitivity

Key question

Covariate balance seems inherently linear.

How can there be an analogue of it for non-linear estimators?

Key idea:

The key idea is to convert the diagnostic into a local sensitivity analysis of this form:

1. Assume your initial model was accurate
2. Select some perturbation your model should be able to capture
3. Use local sensitivity to detect whether the change is what you expect
4. If the change is not what you expect, either (1) or (2) was wrong

Key idea applied to covariate balance

What data perturbation is covariate balance implicitly checking?

One reason to balance $f(\mathbf{x})$ is because we think $\mathbb{E}[y|\mathbf{x}]$ might plausibly vary $\propto f(\mathbf{x})$, and want to check whether our estimator can capture this variability.

For example, linear regression balances regressors, and captures variability in $\mathbb{E}[y|\mathbf{x}]$ that is proportional to the regressors.

Balance checks as local sensitivity

Covariate balance as local sensitivity (informal)

Pick a small $\delta > 0$ and an $f(\cdot)$. Define a new response variable $\tilde{y}$ such that

$$\mathbb{E} [\tilde{y}|\mathbf{x}] = \mathbb{E} [y|\mathbf{x}] + \delta f(\mathbf{x}).$$

We know the change this is supposed to induce in the target population.

Covariate balance checks whether our estimators produce the same change.

Balance checks as local sensitivity

Covariate balance as local sensitivity (formal)

Pick a small $\delta > 0$ and an $f(\cdot)$. Define a new response variable $\tilde{y}$ such that

$$\mathbb{E} [\tilde{y}|\mathbf{x}] = \mathbb{E} [y|\mathbf{x}] + \delta f(\mathbf{x}).$$

We know the expected change this perturbation produces in the target distribution:

$$\mathbb{E} [\mu(\tilde{y}) - \mu(y)|\mathbf{x}] = \frac{1}{N_T} \sum_{j=1}^{N_T} (\mathbb{E} [\tilde{y}|\mathbf{x}_p] - \mathbb{E} [y|\mathbf{x}_p]) = \delta \frac{1}{N_T} \sum_{j=1}^{N_T} f(\mathbf{x}_j)$$

Then, check whether your estimator $\hat{\mu}(\cdot)$ produces the same change for observed $\tilde{Y}_S, Y_S$:

$$\underbrace{\hat{\mu}(\tilde{Y}_S) - \hat{\mu}(Y_S)}_{\text{Replace weighted averages with changes in an estimator}} \stackrel{\text{check}}{\approx} \delta \frac{1}{N_T} \sum_{j=1}^{N_T} f(\mathbf{x}_j).$$

Balance checks as local sensitivity

When $\hat{\mu}(\cdot) = \hat{\mu}^{\text{WGT}}(\cdot)$, we recover the standard covariate balance check.

$$\begin{aligned} \underbrace{\hat{\mu}^{\text{WGT}}(\tilde{Y}_S) - \hat{\mu}^{\text{WGT}}(Y_S)}_{\substack{\text{Replace weighted averages} \\ \text{with changes in an estimator}}} &= \frac{1}{N_S} \sum_{i=1}^{N_S} w_i \tilde{y}_i - \frac{1}{N_S} \sum_{i=1}^{N_S} w_i y_i \\ &= \frac{1}{N_S} \sum_{i=1}^{N_S} w_i (y_i + \delta f(\mathbf{x}_i)) - \frac{1}{N_S} \sum_{i=1}^{N_S} w_i y_i \\ &= \delta \frac{1}{N_S} \sum_{i=1}^{N_S} w_i f(\mathbf{x}_i) \\ &\stackrel{\text{check}}{=} \delta \frac{1}{N_T} \sum_{j=1}^{N_T} f(\mathbf{x}_j). \end{aligned}$$

We will study $\hat{\mu}(\cdot) = \hat{\mu}^{\text{MrP}}(\cdot)$.

Balance checks for MrP

Suppose I have $\tilde{y}$ such that $\mathbb{E} [\tilde{y}|\mathbf{x}] = \mathbb{E} [y|\mathbf{x}] + \delta f(\mathbf{x})$.

Now I need to evaluate $\hat{\mu}^{\text{MrP}}(\tilde{Y}_S) - \hat{\mu}^{\text{MrP}}(Y_S)$.

Balance checks for MrP

Suppose I have $\tilde{y}$ such that $\mathbb{E} [\tilde{y}|\mathbf{x}] = \mathbb{E} [y|\mathbf{x}] + \delta f(\mathbf{x})$.

Now I need to evaluate $\hat{\mu}^{\text{MrP}}(\tilde{Y}_S) - \hat{\mu}^{\text{MrP}}(Y_S)$.

Problem: $\hat{\mu}^{\text{MrP}}(\cdot)$ is computed with MCMC.

- Each MCMC run typically takes hours, and
- MCMC output is noisy, and $\hat{\mu}^{\text{MrP}}(\tilde{Y}_S) - \hat{\mu}^{\text{MrP}}(Y_S)$ may be small.

Balance checks for MrP

Suppose I have $\tilde{y}$ such that $\mathbb{E} [\tilde{y}|\mathbf{x}] = \mathbb{E} [y|\mathbf{x}] + \delta f(\mathbf{x})$.

Now I need to evaluate $\hat{\mu}^{\text{MrP}}(\tilde{Y}_S) - \hat{\mu}^{\text{MrP}}(Y_S)$.

Problem: $\hat{\mu}^{\text{MrP}}(\cdot)$ is computed with MCMC.

- Each MCMC run typically takes hours, and
- MCMC output is noisy, and $\hat{\mu}^{\text{MrP}}(\tilde{Y}_S) - \hat{\mu}^{\text{MrP}}(Y_S)$ may be small.

Solution: Use our local approximation, MrPlew!

Balance informed sensitivity check with MrPlew:

For a wide set of judiciously chosen $f(\cdot)$, check

$$\begin{aligned}\hat{\mu}^{\text{MrP}}(\tilde{Y}_S) - \hat{\mu}^{\text{MrP}}(Y_S) &\approx \frac{1}{N_S} \sum_{i=1}^{N_S} w_i^{\text{MrP}} (\tilde{y}_i - y_i) \\ &\approx \underbrace{\delta \frac{1}{N_S} \sum_{i=1}^{N_S} w_i^{\text{MrP}} f(\mathbf{x}_i)}_{\text{What you actually check}} \stackrel{\text{check}}{\approx} \delta \frac{1}{N_T} \sum_{j=1}^{N_T} f(\mathbf{x}_j).\end{aligned}$$

Covariate balance for primary effects

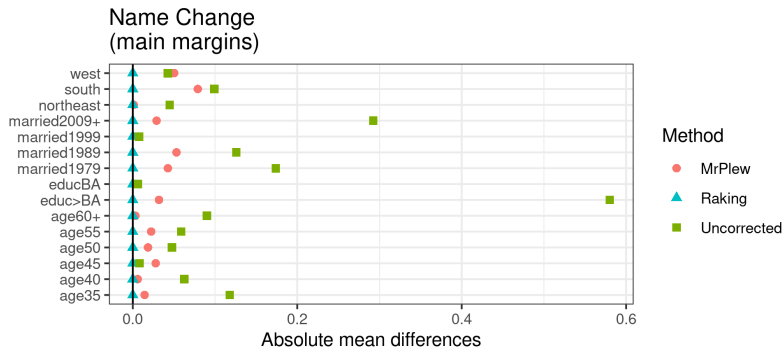


Figure 4: Balance

Covariate balance for interaction effects

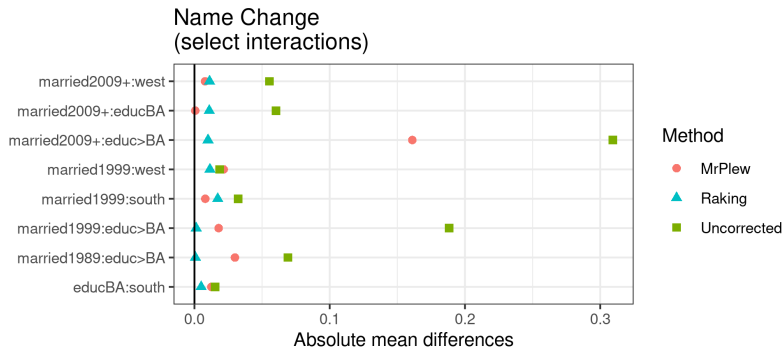


Figure 5: Balance

Subgroup contributions: Same-Sex Marriage Survey

Analysis of US state-level support for same-sex marriage⁵.

- **Survey population:** Five consistently-coded national polls from 2004 ($N_S = 6,341$)
- **Response:** $y = 1$ encodes support for same-sex marriage, $y = 0$ opposition or no opinion.
- **Target population:** US Census 5% PUMS data for CA from 2000

⁵Based on Lax and Phillips [LP09] and Kstellec, Lax, and Phillips [KLP10].

Subgroup contributions: Same-Sex Marriage Survey

Analysis of US state-level support for same-sex marriage⁵.

- **Survey population:** Five consistently-coded national polls from 2004 ($N_S = 6,341$)
- **Response:** $y = 1$ encodes support for same-sex marriage, $y = 0$ opposition or no opinion.
- **Target population:** US Census 5% PUMS data for CA from 2000

Let's analyze this data with two methods: **MrP** and **calibration weighting** (CW).

⁵Based on Lax and Phillips [LP09] and Kestellec, Lax, and Phillips [KLP10].

Subgroup contributions: Same-Sex Marriage Survey

Analysis of US state-level support for same-sex marriage⁵.

- **Survey population:** Five consistently-coded national polls from 2004 ($N_S = 6,341$)
- **Response:** $y = 1$ encodes support for same-sex marriage, $y = 0$ opposition or no opinion.
- **Target population:** US Census 5% PUMS data for CA from 2000

Let's analyze this data with two methods: **MrP** and **calibration weighting** (CW).

MrP computed with `brms` [Bür17]:

$$y \sim \text{logit}\left((1 \mid \text{race.female}) + (1 \mid \text{age.cat}) + (1 \mid \text{edu.cat}) + (1 \mid \text{age.edu.cat}) + \right. \\ \left. (1 \mid \text{state}) + (1 \mid \text{region}) + (1 \mid \text{poll}) + \text{p.relig.full} + \text{p.kerry.full} \right)$$

CW used raking on the following coarsened regressors: (`survey::calibrate` Lumley [Lum24]):

`female, edu.cat, age.cat, race.wbh, race.w.female, age.edu.low`

⁵Based on Lax and Phillips [LP09] and Kstellec, Lax, and Phillips [KLP10].

Subgroup contributions: Same-Sex Marriage Survey

Analysis of US state-level support for same-sex marriage⁵.

- **Survey population:** Five consistently-coded national polls from 2004 ($N_S = 6,341$)
- **Response:** $y = 1$ encodes support for same-sex marriage, $y = 0$ opposition or no opinion.
- **Target population:** US Census 5% PUMS data for CA from 2000

Let's analyze this data with two methods: **MrP** and **calibration weighting** (CW).

MrP computed with `brms` [Bür17]:

$$y \sim \text{logit}\left((1 \mid \text{race.female}) + (1 \mid \text{age.cat}) + (1 \mid \text{edu.cat}) + (1 \mid \text{age.edu.cat}) + \right. \\ \left. (1 \mid \text{state}) + (1 \mid \text{region}) + (1 \mid \text{poll}) + \text{p.relig.full} + \text{p.kerry.full} \right)$$

CW used raking on the following coarsened regressors: (`survey::calibrate` Lumley [Lum24]):

female, edu.cat, age.cat, race.wbh, race.w.female, age.edu.low

Raw survey mean: $\frac{1}{N_S} \sum_{i=1}^{N_S} y_i = 0.333$

Raking: $\hat{\mu}^{\text{WGT}}(Y_S) = 0.345$

MrP: $\hat{\mu}^{\text{MrP}}(Y_S) = 0.457$

⁵Based on Lax and Phillips [LP09] and Kestellec, Lax, and Phillips [KLP10].

Subgroup contributions: Same-Sex Marriage Survey

Analysis of US state-level support for same-sex marriage⁵.

- **Survey population:** Five consistently-coded national polls from 2004 ($N_S = 6,341$)
- **Response:** $y = 1$ encodes support for same-sex marriage, $y = 0$ opposition or no opinion.
- **Target population:** US Census 5% PUMS data for CA from 2000

Let's analyze this data with two methods: **MrP** and **calibration weighting** (CW).

MrP computed with `brms` [Bür17]:

$$y \sim \text{logit}\left((1 \mid \text{race.female}) + (1 \mid \text{age.cat}) + (1 \mid \text{edu.cat}) + (1 \mid \text{age.edu.cat}) + \right. \\ \left. (1 \mid \text{state}) + (1 \mid \text{region}) + (1 \mid \text{poll}) + \text{p.relig.full} + \text{p.kerry.full} \right)$$

CW used raking on the following coarsened regressors: (`survey::calibrate` Lumley [Lum24]):

female, edu.cat, age.cat, race.wbh, race.w.female, age.edu.low

Raw survey mean: $\frac{1}{N_S} \sum_{i=1}^{N_S} y_i = 0.333$

Raking: $\hat{\mu}^{\text{WGT}}(Y_S) = 0.345$

MrP: $\hat{\mu}^{\text{MrP}}(Y_S) = 0.457$

How are MrP and raking using the data differently?

⁵Based on Lax and Phillips [LP09] and Kestellec, Lax, and Phillips [KLP10].

Same-Sex Marriage Survey Weights

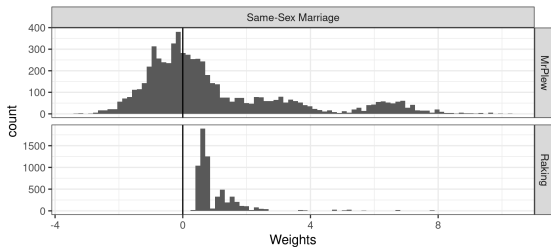


Figure 6: Weight comparison for the Same-sex marriage dataset

Same-Sex Marriage Survey Weights

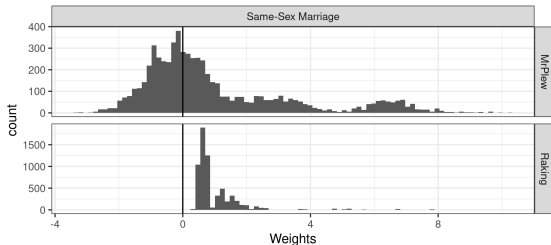


Figure 6: Weight comparison for the Same-sex marriage dataset

What does the difference between these weights mean in terms of

- Frequentist sampling variance (standard errors)?
- Covariate balance / external validity?
- Subgroup contributions?

Same-Sex Marriage subgroup contribution

Note that both methods use national data to infer public opinion in California.

Same-Sex Marriage subgroup contribution

Note that both methods use national data to infer public opinion in California.

Question:

How do responses in Texas affect our estimate of public opinion in CA?



Same-Sex Marriage subgroup contribution

Note that both methods use national data to infer public opinion in California.

More specific question:

Suppose we increased the responses y_i in Texas.

How would our estimate of public opinion in CA change?

Take $\tilde{y}_i = y_i + \delta \mathbb{I}(i \text{ is in Texas})$. Then $\hat{\mu}^{\text{WGT}}(\tilde{y}) - \hat{\mu}^{\text{WGT}}(y) = \delta \underbrace{\sum_{i \text{ in Texas}} w_i}_{\text{Texas's contribution}}.$

Same-Sex Marriage subgroup contribution

Note that both methods use national data to infer public opinion in California.

More specific question:

Suppose we increased the responses y_i in Texas.

How would our estimate of public opinion in CA change?

Take $\tilde{y}_i = y_i + \delta \mathbb{I}(i \text{ is in Texas})$. Then $\hat{\mu}^{\text{MrP}}(\tilde{y}) - \hat{\mu}^{\text{MrP}}(y) \approx \delta \underbrace{\sum_{i \text{ in Texas}} w_i^{\text{MrP}}}_{\text{Texas's contribution}}.$

Same-Sex Marriage subgroup contribution

Note that both methods use national data to infer public opinion in California.

More specific question:

Suppose we increased the responses y_i in Texas.

How would our estimate of public opinion in CA change?

Take $\tilde{y}_i = y_i + \delta \mathbb{I}(i \text{ is in Texas})$. Then $\hat{\mu}^{\text{MrP}}(\tilde{y}) - \hat{\mu}^{\text{MrP}}(y) \approx \delta \underbrace{\sum_{i \text{ in Texas}} w_i^{\text{MrP}}}_{\text{Texas's contribution}}.$

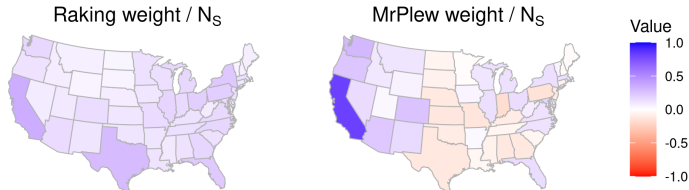


Figure 7: Subgroup contribution for Same-Sex Marriage in California

Same-Sex Marriage subgroup contribution

Note that both methods use national data to infer public opinion in California.

More specific question:

Suppose we increased the responses y_i in Texas.

How would our estimate of public opinion in CA change?

Take $\tilde{y}_i = y_i + \delta \mathbb{I}(i \text{ is in Texas})$. Then $\hat{\mu}^{\text{MrP}}(\tilde{y}) - \hat{\mu}^{\text{MrP}}(y) \approx \delta \underbrace{\sum_{i \text{ in Texas}} w_i^{\text{MrP}}}_{\text{Texas's contribution}}.$

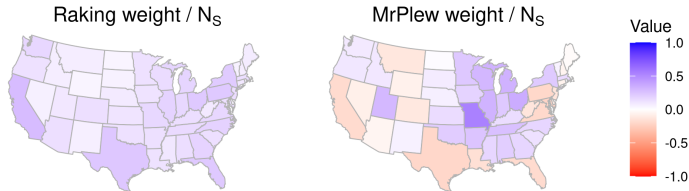


Figure 7: Subgroup contribution for Same-Sex Marriage in Missouri

Generating $\tilde{y}$

- We have defined balance checks in terms of $\tilde{y}$ such that $\mathbb{E} [\tilde{y}|\mathbf{x}] = \mathbb{E} [y|\mathbf{x}] + \delta f(\mathbf{x})$
- We have approximated $\hat{\mu}^{\text{MrP}}(\tilde{Y}_S) - \hat{\mu}^{\text{MrP}}(Y_S)$ for $\tilde{y} \approx y$

How to get such a $\tilde{y}$? **Recall y is binary!**

Generating $\tilde{y}$

- We have defined balance checks in terms of $\tilde{y}$ such that $\mathbb{E} [\tilde{y}|\mathbf{x}] = \mathbb{E} [y|\mathbf{x}] + \delta f(\mathbf{x})$
- We have approximated $\hat{\mu}^{\text{MrP}}(\tilde{Y}_S) - \hat{\mu}^{\text{MrP}}(Y_S)$ for $\tilde{y} \approx y$

How to get such a $\tilde{y}$? **Recall y is binary!** **Two solutions, with their own pros and cons:**

Option 1: Force $\tilde{y}$ to be binary.

Option 2: Allow $\tilde{y}$ to take generic values.

- We have defined balance checks in terms of $\tilde{y}$ such that $\mathbb{E} [\tilde{y}|\mathbf{x}] = \mathbb{E} [y|\mathbf{x}] + \delta f(\mathbf{x})$
- We have approximated $\hat{\mu}^{\text{MrP}}(\tilde{Y}_S) - \hat{\mu}^{\text{MrP}}(Y_S)$ for $\tilde{y} \approx y$

How to get such a $\tilde{y}$? **Recall y is binary!** Two solutions, with their own pros and cons:

Option 1: Force $\tilde{y}$ to be binary.

1. Make *some* guess $\hat{m}(\mathbf{x}) \approx \mathbb{E} [y|\mathbf{x}]$
 - E.g. Posterior mean, or
 - Shrunk posterior mean, or
 - Some values that gives the same posterior
2. Take $u_i \stackrel{iid}{\sim} \text{Unif}(0, 1)$
3. Assume $y_i = \mathbb{I}(u_i \leq \hat{m}(\mathbf{x}_i))$
4. Draw $u_n|y_n$
5. Set $\tilde{y}_i = \mathbb{I}(u_i \leq \hat{m}(\mathbf{x}_i) + \delta \mathbf{x}_i)$

Option 2: Allow $\tilde{y}$ to take generic values.

Generating $\tilde{y}$

- We have defined balance checks in terms of $\tilde{y}$ such that $\mathbb{E} [\tilde{y}|\mathbf{x}] = \mathbb{E} [y|\mathbf{x}] + \delta f(\mathbf{x})$
- We have approximated $\hat{\mu}^{\text{MrP}}(\tilde{Y}_S) - \hat{\mu}^{\text{MrP}}(Y_S)$ for $\tilde{y} \approx y$

How to get such a $\tilde{y}$? **Recall y is binary!** Two solutions, with their own pros and cons:

Option 1: Force $\tilde{y}$ to be binary.

1. Make *some* guess $\hat{m}(\mathbf{x}) \approx \mathbb{E} [y|\mathbf{x}]$
 - E.g. Posterior mean, or
 - Shrunk posterior mean, or
 - Some values that gives the same posterior
2. Take $u_i \stackrel{iid}{\sim} \text{Unif}(0, 1)$
3. Assume $y_i = \mathbb{I}(u_i \leq \hat{m}(\mathbf{x}_i))$
4. Draw $u_n|y_n$
5. Set $\tilde{y}_i = \mathbb{I}(u_i \leq \hat{m}(\mathbf{x}_i) + \delta \mathbf{x}_i)$

Option 2: Allow $\tilde{y}$ to take generic values.

1. Set $\tilde{y}_i = y_i + \delta f(\mathbf{x}_i)$.
2. Then you're done.
3. There is nothing else to do.
4. This space deliberately left blank.

Generating $\tilde{y}$

- We have defined balance checks in terms of $\tilde{y}$ such that $\mathbb{E}[\tilde{y}|\mathbf{x}] = \mathbb{E}[y|\mathbf{x}] + \delta f(\mathbf{x})$
- We have approximated $\hat{\mu}^{\text{MrP}}(\tilde{Y}_{\mathcal{S}}) - \hat{\mu}^{\text{MrP}}(Y_{\mathcal{S}})$ for $\tilde{y} \approx y$

How to get such a $\tilde{y}$? **Recall y is binary!** Two solutions, with their own pros and cons:

Option 1: Force $\tilde{y}$ to be binary.

1. Make *some* guess $\hat{m}(\mathbf{x}) \approx \mathbb{E}[y|\mathbf{x}]$
 - E.g. Posterior mean, or
 - Shrunk posterior mean, or
 - Some values that gives the same posterior
2. Take $u_i \stackrel{iid}{\sim} \text{Unif}(0, 1)$
3. Assume $y_i = \mathbb{I}(u_i \leq \hat{m}(\mathbf{x}_i))$
4. Draw $u_n|y_n$
5. Set $\tilde{y}_i = \mathbb{I}(u_i \leq \hat{m}(\mathbf{x}_i) + \delta \mathbf{x}_i)$

Pros and cons:

- Realistic
- Have to pick $\hat{m}(\mathbf{x})$
- $\tilde{Y}_{\mathcal{S}} - Y_{\mathcal{S}}$ not infinitesimally small
- **Use for checks & experiments**

Option 2: Allow $\tilde{y}$ to take generic values.

1. Set $\tilde{y}_i = y_i + \delta f(\mathbf{x}_i)$.
2. Then you're done.
3. There is nothing else to do.
4. This space deliberately left blank.

Pros and cons:

- Not realistic
- No additional assumptions
- $\tilde{Y}_{\mathcal{S}} - Y_{\mathcal{S}}$ may be infinitesimally small
- **Use for theory**

When is the local approximation accurate?

Theorem [Gio+26]:

Take $\tilde{y}_i = y_i + \delta f(\mathbf{x}_i)$.

We state conditions for Bayesian hierarchical logistic regression under which

$$\left| \hat{\mu}^{\text{MrP}}(\tilde{Y}_S) - \hat{\mu}^{\text{MrP}}(Y_S) - \delta \sum_{i=1}^{N_S} w_i^{\text{MrP}} f(\mathbf{x}_i) \right| = \text{Small}$$

⁶ $\mathcal{F}$ can be any Donsker class of measurable functions with uniformly bounded $\mathbb{E} [\|\mathbf{x}\| f(\mathbf{x})] \leq \frac{2}{\delta}$.

⁷ GB24; KGB25.

When is the local approximation accurate?

Theorem [Gio+26]:

Take $\tilde{y}_i = y_i + \delta f(\mathbf{x}_i)$.

We state conditions for Bayesian hierarchical logistic regression under which

$$\left| \hat{\mu}^{\text{MrP}}(\tilde{Y}_S) - \hat{\mu}^{\text{MrP}}(Y_S) - \delta \sum_{i=1}^{N_S} w_i^{\text{MrP}} f(\mathbf{x}_i) \right| = O(\delta^2)$$

⁶ $\mathcal{F}$ can be any Donsker class of measurable functions with uniformly bounded $\mathbb{E} [\|\mathbf{x}\| f(\mathbf{x})] \leq \frac{2}{\delta}$.

⁷ GB24; KGB25.

When is the local approximation accurate?

Theorem [Gio+26]:

Take $\tilde{y}_i = y_i + \delta f(\mathbf{x}_i)$.

We state conditions for Bayesian hierarchical logistic regression under which

$$\left| \hat{\mu}^{\text{MrP}}(\tilde{Y}_S) - \hat{\mu}^{\text{MrP}}(Y_S) - \delta \sum_{i=1}^{N_S} w_i^{\text{MrP}} f(\mathbf{x}_i) \right| = O(\delta^2) \text{ as } N \rightarrow \infty$$

⁶ $\mathcal{F}$ can be any Donsker class of measurable functions with uniformly bounded $\mathbb{E} [\|\mathbf{x}\| f(\mathbf{x})] \leq \frac{2}{\delta}$.

⁷GB24; KGB25.

When is the local approximation accurate?

Theorem [Gio+26]:

Take $\tilde{y}_i = y_i + \delta f(\mathbf{x}_i)$.

We state conditions for Bayesian hierarchical logistic regression under which

$$\sup_{f \in \mathcal{F}} \left| \hat{\mu}^{\text{MrP}}(\tilde{Y}_S) - \hat{\mu}^{\text{MrP}}(Y_S) - \delta \sum_{i=1}^{N_S} w_i^{\text{MrP}} f(\mathbf{x}_i) \right| = O(\delta^2) \text{ as } N \rightarrow \infty$$

...for a very broad class of $\mathcal{F}$.⁶

Uniformity justifies searching for “imbalanced” f .

⁶ $\mathcal{F}$ can be any Donsker class of measurable functions with uniformly bounded $\mathbb{E} [\|\mathbf{x}\| f(\mathbf{x})] \leq \frac{2}{\delta}$.

⁷ GB24; KGB25.

When is the local approximation accurate?

Theorem [Gio+26]:

Take $\tilde{y}_i = y_i + \delta f(\mathbf{x}_i)$.

We state conditions for Bayesian hierarchical logistic regression under which

$$\sup_{f \in \mathcal{F}} \left| \hat{\mu}^{\text{MrP}}(\tilde{Y}_S) - \hat{\mu}^{\text{MrP}}(Y_S) - \delta \sum_{i=1}^{N_S} w_i^{\text{MrP}} f(\mathbf{x}_i) \right| = O(\delta^2) \text{ as } N \rightarrow \infty$$

...for a very broad class of $\mathcal{F}$.⁶

Uniformity justifies searching for “imbalanced” f .

The uniformity result builds on our earlier work on uniform and finite-sample error bounds for Bernstein–von Mises theorem–like results⁷.

⁶ $\mathcal{F}$ can be any Donsker class of measurable functions with uniformly bounded $\mathbb{E} [\|\mathbf{x}\| f(\mathbf{x})] \leq \frac{2}{\delta}$.

⁷ GB24; KGB25.

Predictions and actual MCMC results

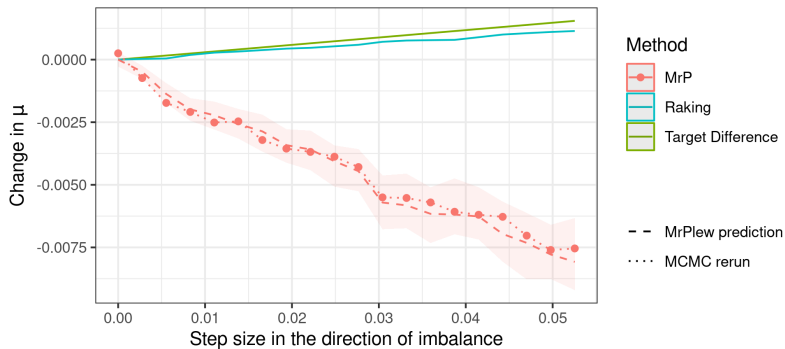


Figure 8: Name Change dataset refit

Predictions and actual MCMC results

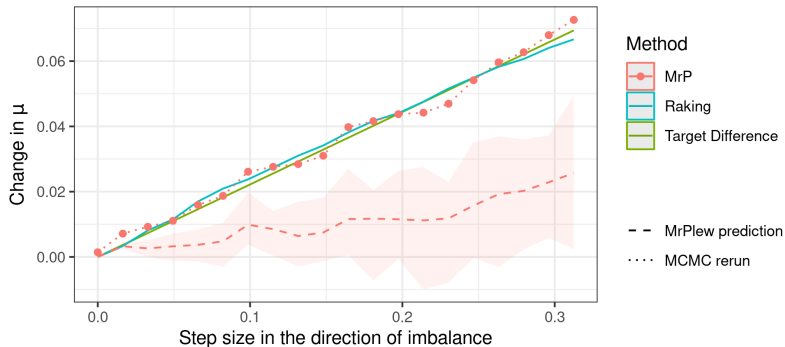


Figure 9: Same-sex marriage dataset refit

More careful investigation shows that the linear approximation is computed correctly, but the posterior is highly curved in this direction, even for small step sizes.

There is real work to do to understand this!

By defining MrP locally equivlanent weights, we can rigorously provide a novel class of model checking metrics for non-linear MrP estimators.

Conclusion

~~By defining MrP locally equivlanent weights, we can rigorously provide a novel class of model checking metrics for non-linear MrP estimators.~~

⚠ That's not the right conclusion!

~~By defining MrP locally equivlanent weights, we can rigorously provide a novel class of model checking metrics for non-linear MrP estimators.~~

⚠️ That's not the right conclusion!

A correctly specified response model can do badly on CW checks, and vice-versa! (See paper.)

This work is best thought of as *model interrogation* for MrP.

Conclusion (let's try that again)

By defining *MrP locally equivlanent weights*, we can rigorously provide analogues of many CW diagnostics for non-linear MrP estimators for model interrogation.

Follow-on work (coming soon...)

- Similar ideas can to lead to actual checks of MrP model misspecification.
- There is a sense in which the MrPlew weights are estimates of the ratio $\mathcal{P}_T(\mathbf{x})/\mathcal{P}_S(\mathbf{x})$ (projected onto a certain subspace)

Thank you for having me!

Conclusion (let's try that again)

By defining *MrP locally equivlanent weights*, we can rigorously provide analogues of many CW diagnostics for non-linear MrP estimators for model interrogation.

Follow-on work (coming soon...)

- Similar ideas can to lead to actual checks of MrP model misspecification.
- There is a sense in which the MrPlew weights are estimates of the ratio $\mathcal{P}_T(\mathbf{x})/\mathcal{P}_S(\mathbf{x})$ (projected onto a certain subspace)

R package available at

<https://github.com/rgiordan/mrplew>

Preprint at

<https://arxiv.org/abs/2606.04250>

Thank you for having me!

- [Ale19] M. Alexander. Analyzing name changes after marriage using a non-representative survey. 2019. URL: <https://www.monicaalexander.com/posts/2019-08-07-mrp/>.
- [BFH21] Eli B., Avi F., and Erin H. Multilevel calibration weighting for survey data. 2021. arXiv: 2102.09052 [stat.ME].
- [Blu24] Blue Rose Research. 2024 Election Retrospective Presentation. <https://data.blueroseresearch.org/2024retro-download>. Accessed on 2024-10-26. 2024.
- [Bon+25] A. Bonica, R. Fordham, J. Grumbach, and E. Tiburcio. Did Non-Voters Really Flip Republican in 2024? The Evidence Says No. <https://data4democracy.substack.com/p/did-non-voters-really-flip-republican>. Apr. 2025.
- [Bür17] Paul-Christian Bürkner. “brms: An R Package for Bayesian Multilevel Models Using Stan”. In: Journal of Statistical Software 80.1 (2017), pp. 1–28. DOI: 10.18637/jss.v080.i01.
- [Coh19] P. Cohen. Marital Name Change Survey. Apr. 2019. DOI: 10.17605/OSF.IO/UZQDN. URL: osf.io/uzqdn.
- [CZ23] A. Chattopadhyay and J. Zubizarreta. “On the implied weights of linear regression for causal inference”. In: Biometrika 110.3 (2023), pp. 615–629.
- [DF86] P. Diaconis and D. Freedman. “On the consistency of Bayes estimates”. In: The Annals of Statistics (1986), pp. 1–26.
- [DSS93] J. Deville, C. Särndal, and O. Sautory. “Generalized raking procedures in survey sampling”. In: Journal of the American statistical Association 88.423 (1993), pp. 1013–1020.
- [Efr15] B. Efron. “Frequentist accuracy of Bayesian estimates”. In: Journal of the Royal Statistical Society Series B: Statistical Methodology 77.3 (2015), pp. 617–646.
- [Ful11] W. Fuller. Sampling statistics. John Wiley & Sons, 2011.
- [G+19] G., W. Stephenson, R. Liu, M. I. Jordan, and T. Broderick. “A swiss army infinitesimal jackknife”. In: The 22nd International Conference on Artificial Intelligence and Statistics. PMLR. 2019, pp. 1139–1147.
- [GB24] G. and T. Broderick. The Bayesian Infinitesimal Jackknife for Variance. 2024. arXiv: 2305.06466 [stat.ME]. URL: <https://arxiv.org/abs/2305.06466>.

- [GBJ18] G., T. Broderick, and M. I. Jordan. “Covariances, robustness and variational bayes”. In: Journal of machine learning research 19.51 (2018).
- [Gel07a] A. Gelman. “Rejoinder: Struggles with survey weighting and regression modelling”. In: Statistical Science 22.2 (2007), pp. 184–188.
- [Gel07b] A. Gelman. “Struggles with survey weighting and regression modeling”. In: (2007).
- [Gio+26] R. Giordano, A. Cima, J. Murray, E. Hartman, and A. Feller. “Locally Equivalent Weights for Multilevel Regression and Poststratification”. In: (2026). arXiv: 2606.04250 [stat.ME]. URL: <https://arxiv.org/abs/2606.04250>.
- [Gus96] P. Gustafson. “Local sensitivity of posterior expectations”. In: The Annals of Statistics 24.1 (1996), pp. 174–195.
- [KGB25] M. Kasprzak, G., and T. Broderick. How good is your Laplace approximation of the Bayesian posterior? Finite-sample computable error bounds for a variety of useful divergences. 2025. arXiv: 2209.14992 [math.ST]. URL: <https://arxiv.org/abs/2209.14992>.
- [KLP10] J. Kastellec, J. Lax, and J. Phillips. “Estimating state public opinion with multi-level regression and poststratification using R”. In: Unpublished manuscript, Princeton University 29.3 (2010).
- [KP12] S. Krantz and H. Parks. The Implicit Function Theorem: History, Theory, and Applications. Springer Science & Business Media, 2012.
- [LP09] J. Lax and J. Phillips. “Gay rights in the states: Public opinion and policy responsiveness”. In: American Political Science Review 103.3 (2009), pp. 367–386.
- [Lum24] T. Lumley. survey: Analysis of complex survey samples. R package version 4.4. 2024.
- [Rug+24] S. Ruggles, S. Flood, M. Sobek, D. Backman, A. Chen, G. Cooper, S. Richards, R. Rodgers, and Megan S. IPUMS USA: Version 15.0 [dataset]. 2024. DOI: 10.18128/D010.V15.0. URL: <https://usa.ipums.org>.
- [SHW15] G. Solon, S. Haider, and J. Wooldridge. “What are we weighting for?” In: Journal of Human resources 50.2 (2015), pp. 301–316.

Extra slides

Frequentist variance estimation

Let $\hat{\text{Var}}(\cdot)$ denote the sample variance.

Calibration weighting standard errors sketch:⁸

If we have $\hat{\mu}^{\text{WGT}}(Y_S) = \frac{1}{N_S} \sum_{i=1}^{N_S} w_i y_i$ and a consistent residual estimate ε_i , then

$$\hat{\text{Var}}(w_i \varepsilon_i) \approx \text{Var} \left(\sqrt{N_S} \hat{\mu}^{\text{WGT}}(Y_S) \right) .$$

⁸E.g., Deville, Särndal, and Sautory [DSS93] and Fuller [Ful11].

Frequentist variance estimation

Let $\hat{\text{Var}}(\cdot)$ denote the sample variance.

Calibration weighting standard errors sketch:⁸

If we have $\hat{\mu}^{\text{WGT}}(Y_S) = \frac{1}{N_S} \sum_{i=1}^{N_S} w_i y_i$ and a consistent residual estimate ε_i , then

$$\hat{\text{Var}}(w_i \varepsilon_i) \approx \text{Var}\left(\sqrt{N_S} \hat{\mu}^{\text{WGT}}(Y_S)\right).$$

MrPlew Standard error consistency theorem sketch (Our contribution):⁹

For Bayesian hierarchical logistic regression, define $\varepsilon_i = y_i - \mathbb{E}_{\mathcal{P}(\theta|\text{Survey data})} [m(\mathbf{x}_i^T \theta)]$.

We state mild conditions under which, as $N_S \rightarrow \infty$, for some μ_∞ and variance V ,

$$\begin{aligned} \sqrt{N_S} (\hat{\mu}^{\text{MrP}}(Y_S) - \mu_\infty) &\rightarrow \mathcal{N}(0, V) \quad \text{and} \\ \hat{\text{Var}}(w_i^{\text{MrP}} \varepsilon_i) &\rightarrow V. \end{aligned}$$

The use of w_i^{MrP} is analogous to the use of w_i for frequentist variance estimation.

⁸E.g., Deville, Särndal, and Sautory [DSS93] and Fuller [Ful11].

⁹This is essentially a corollary of our earlier work on the Bayesian infinitesimal jackknife. (GB24)

Standard error estimation experiment

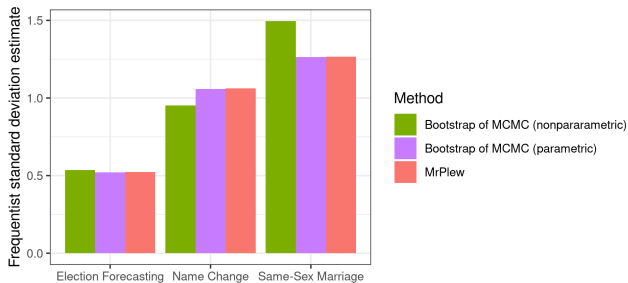


Figure 10: Estimates of the frequentist standard deviation of $\sqrt{N_S} \hat{\mu}^{\text{MrP}}$

Standard error estimation experiment



Figure 10: Estimates of the frequentist standard deviation of $\sqrt{N_S} \hat{\mu}^{\text{MrP}}$

Running fifty MCMC parametric bootstraps: ≈ 95 minutes

Computing approximate weights: 664 seconds

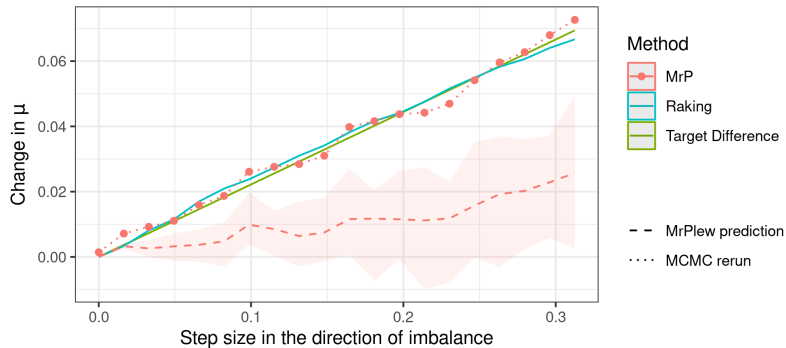


Figure 11: Same-sex marriage dataset refit

Running each MCMC refit: 39 minutes Computing approximate weights: 7 seconds

Regression

Regression

General models

Regression

General models

Consistency /
Unbiased

$$y = \theta^\top \mathbf{x} + \varepsilon$$

$$\tilde{y} = (\theta + \delta)^\top \mathbf{x} + \varepsilon$$

$$\hat{\theta}(\tilde{y}) \stackrel{\text{check}}{=} \hat{\theta}(y) + \delta$$

Some generalized diagnostics

Regression

Consistency /
Unbiased

$$y = \theta^\top \mathbf{x} + \varepsilon$$

$$\tilde{y} = (\theta + \delta)^\top \mathbf{x} + \varepsilon$$

$$\hat{\theta}(\tilde{y}) \stackrel{\text{check}}{=} \hat{\theta}(y) + \delta$$

General models

$$y = f(\mathbf{x}, \varepsilon, \theta)$$

$$\tilde{y} = f(\mathbf{x}, \varepsilon, \theta + \delta)$$

$$\hat{\theta}(\tilde{y}) \stackrel{\text{check}}{=} \hat{\theta}(y) + \delta$$

Some generalized diagnostics

Regression

Consistency /
Unbiased

$$y = \theta^T \mathbf{x} + \varepsilon$$

$$\tilde{y} = (\theta + \delta)^T \mathbf{x} + \varepsilon$$

$$\hat{\theta}(\tilde{y}) \stackrel{\text{check}}{=} \hat{\theta}(y) + \delta$$

General models

$$y = f(\mathbf{x}, \varepsilon, \theta)$$

$$\tilde{y} = f(\mathbf{x}, \varepsilon, \theta + \delta)$$

$$\hat{\theta}(\tilde{y}) \stackrel{\text{check}}{=} \hat{\theta}(y) + \delta$$

Exogenous
residuals

$$y = \theta^T \mathbf{x} + \varepsilon$$

$$\tilde{y} = y + \varepsilon z$$

$$\hat{\theta}(\tilde{y}) \stackrel{\text{check}}{=} \hat{\theta}(y)$$

Some generalized diagnostics

Regression

Consistency /
Unbiased

$$y = \theta^\top \mathbf{x} + \varepsilon$$

$$\tilde{y} = (\theta + \delta)^\top \mathbf{x} + \varepsilon$$

$$\hat{\theta}(\tilde{y}) \stackrel{\text{check}}{=} \hat{\theta}(y) + \delta$$

General models

$$y = f(\mathbf{x}, \varepsilon, \theta)$$

$$\tilde{y} = f(\mathbf{x}, \varepsilon, \theta + \delta)$$

$$\hat{\theta}(\tilde{y}) \stackrel{\text{check}}{=} \hat{\theta}(y) + \delta$$

Exogenous
residuals

$$y = \theta^\top \mathbf{x} + \varepsilon$$

$$\tilde{y} = y + \varepsilon z$$

$$\hat{\theta}(\tilde{y}) \stackrel{\text{check}}{=} \hat{\theta}(y)$$

$$y \sim \mathcal{P}(y|\mathbf{x}) \text{ and } \mathcal{P}(\mathbf{x}) = w$$

$$\tilde{w} = w + \delta z$$

$$\hat{\theta}(\tilde{w}) \stackrel{\text{check}}{=} \hat{\theta}(w)$$

Some generalized diagnostics

	Regression	General models
Consistency / Unbiased	$y = \theta^\top \mathbf{x} + \varepsilon$ $\tilde{y} = (\theta + \delta)^\top \mathbf{x} + \varepsilon$ $\hat{\theta}(\tilde{y}) \stackrel{\text{check}}{=} \hat{\theta}(y) + \delta$	$y = f(\mathbf{x}, \varepsilon, \theta)$ $\tilde{y} = f(\mathbf{x}, \varepsilon, \theta + \delta)$ $\hat{\theta}(\tilde{y}) \stackrel{\text{check}}{=} \hat{\theta}(y) + \delta$
Exogenous residuals	$y = \theta^\top \mathbf{x} + \varepsilon$ $\tilde{y} = y + \varepsilon z$ $\hat{\theta}(\tilde{y}) \stackrel{\text{check}}{=} \hat{\theta}(y)$	$y \sim \mathcal{P}(y \mathbf{x})$ and $\mathcal{P}(\mathbf{x}) = w$ $\tilde{w} = w + \delta z$ $\hat{\theta}(\tilde{w}) \stackrel{\text{check}}{=} \hat{\theta}(w)$
Fisher information	$\mathcal{I} :=$ Fisher information $\Sigma :=$ Score covariance $\mathcal{I}^{-1} \stackrel{\text{check}}{=} \Sigma$	

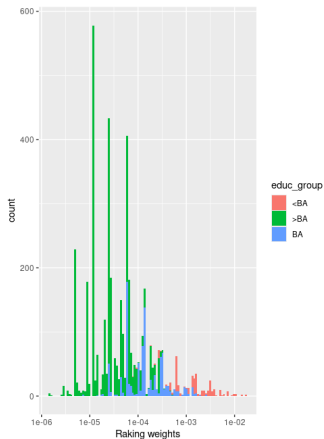
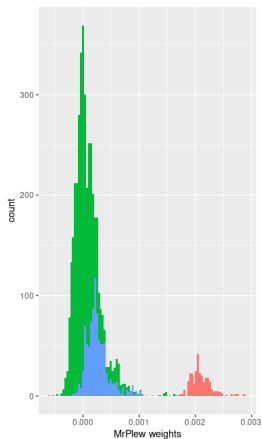
Some generalized diagnostics

	Regression	General models
Consistency / Unbiased	$y = \theta^\top \mathbf{x} + \varepsilon$ $\tilde{y} = (\theta + \delta)^\top \mathbf{x} + \varepsilon$ $\hat{\theta}(\tilde{y}) \stackrel{\text{check}}{=} \hat{\theta}(y) + \delta$	$y = f(\mathbf{x}, \varepsilon, \theta)$ $\tilde{y} = f(\mathbf{x}, \varepsilon, \theta + \delta)$ $\hat{\theta}(\tilde{y}) \stackrel{\text{check}}{=} \hat{\theta}(y) + \delta$
Exogenous residuals	$y = \theta^\top \mathbf{x} + \varepsilon$ $\tilde{y} = y + \varepsilon z$ $\hat{\theta}(\tilde{y}) \stackrel{\text{check}}{=} \hat{\theta}(y)$	$y \sim \mathcal{P}(y \mathbf{x})$ and $\mathcal{P}(\mathbf{x}) = w$ $\tilde{w} = w + \delta z$ $\hat{\theta}(\tilde{w}) \stackrel{\text{check}}{=} \hat{\theta}(w)$
Fisher information	$\mathcal{I} :=$ Fisher information $\Sigma :=$ Score covariance $\mathcal{I}^{-1} \stackrel{\text{check}}{=} \Sigma$	$y \sim \mathcal{P}(y \theta)$ $\tilde{y} \sim$ Importance sample y using $\tilde{w} = \frac{\mathcal{P}(y \hat{\theta} + \delta)}{\mathcal{P}(y \hat{\theta})}$ $\hat{\theta}(\tilde{w}) \stackrel{\text{check}}{=} \hat{\theta}(1) + \delta$

Some generalized diagnostics

	Regression	General models
Consistency / Unbiased	$y = \theta^\top \mathbf{x} + \varepsilon$ $\tilde{y} = (\theta + \delta)^\top \mathbf{x} + \varepsilon$ $\hat{\theta}(\tilde{y}) \stackrel{\text{check}}{=} \hat{\theta}(y) + \delta$	$y = f(\mathbf{x}, \varepsilon, \theta)$ $\tilde{y} = f(\mathbf{x}, \varepsilon, \theta + \delta)$ $\hat{\theta}(\tilde{y}) \stackrel{\text{check}}{=} \hat{\theta}(y) + \delta$
Exogenous residuals	$y = \theta^\top \mathbf{x} + \varepsilon$ $\tilde{y} = y + \varepsilon z$ $\hat{\theta}(\tilde{y}) \stackrel{\text{check}}{=} \hat{\theta}(y)$	$y \sim \mathcal{P}(y \mathbf{x})$ and $\mathcal{P}(\mathbf{x}) = w$ $\tilde{w} = w + \delta z$ $\hat{\theta}(\tilde{w}) \stackrel{\text{check}}{=} \hat{\theta}(w)$
Fisher information	$\mathcal{I} :=$ Fisher information $\Sigma :=$ Score covariance $\mathcal{I}^{-1} \stackrel{\text{check}}{=} \Sigma$	$y \sim \mathcal{P}(y \theta)$ $\tilde{y} \sim$ Importance sample y using $\tilde{w} = \frac{\mathcal{P}(y \hat{\theta} + \delta)}{\mathcal{P}(y \hat{\theta})}$ $\hat{\theta}(\tilde{w}) \stackrel{\text{check}}{=} \hat{\theta}(1) + \delta$

Weights by education for the Name Change analysis



Approximately equivalent weights for (some) logistic regression MrP

- Suppose the model is $m(\mathbf{x}^\top \theta) = \text{Logistic}(\mathbf{x}^\top \theta)$, with MLE $\hat{\theta}$.
- MrP is $\hat{\mu}^{\text{MrP}}(Y_S) = \frac{1}{N_T} \sum_{j=1}^{N_T} m(\mathbf{x}_j^\top \hat{\theta})$.

The map from $Y_S \mapsto m(\mathbf{x}_j^\top \hat{\theta})$ is typically globally nonlinear.

But some sample averages of $m(\mathbf{x}_j^\top \hat{\theta})$ can be approximately linear.

Approximately equivalent weights for (some) logistic regression MrP

- Suppose the model is $m(\mathbf{x}^\top \theta) = \text{Logistic}(\mathbf{x}^\top \theta)$, with MLE $\hat{\theta}$.
- MrP is $\hat{\mu}^{\text{MrP}}(Y_S) = \frac{1}{N_T} \sum_{j=1}^{N_T} m(\mathbf{x}_j^\top \hat{\theta})$.

Example

Suppose $\frac{\mathcal{P}_T(\mathbf{x})}{\mathcal{P}_S(\mathbf{x})} \approx \alpha^\top \mathbf{x}$ for some α . Then MrP is a approximately a weighting estimator.

$$\hat{\mu}^{\text{MrP}}(Y_S) = \frac{1}{N_T} \sum_{j=1}^{N_T} m(\mathbf{x}_j^\top \hat{\theta})$$

Approximately equivalent weights for (some) logistic regression MrP

- Suppose the model is $m(\mathbf{x}^\top \theta) = \text{Logistic}(\mathbf{x}^\top \theta)$, with MLE $\hat{\theta}$.
- MrP is $\hat{\mu}^{\text{MrP}}(Y_S) = \frac{1}{N_T} \sum_{j=1}^{N_T} m(\mathbf{x}_j^\top \hat{\theta})$.

Example

Suppose $\frac{\mathcal{P}_T(\mathbf{x})}{\mathcal{P}_S(\mathbf{x})} \approx \alpha^\top \mathbf{x}$ for some α . Then MrP is a approximately a weighting estimator.

$$\begin{aligned}\hat{\mu}^{\text{MrP}}(Y_S) &= \frac{1}{N_T} \sum_{j=1}^{N_T} m(\mathbf{x}_j^\top \hat{\theta}) \\ &\approx \int m(\mathbf{x}^\top \hat{\theta}) \mathcal{P}_T(\mathbf{x}) d\mathbf{x} \quad (\text{Law of large numbers})\end{aligned}$$

Approximately equivalent weights for (some) logistic regression MrP

- Suppose the model is $m(\mathbf{x}^\top \theta) = \text{Logistic}(\mathbf{x}^\top \theta)$, with MLE $\hat{\theta}$.
- MrP is $\hat{\mu}^{\text{MrP}}(Y_S) = \frac{1}{N_T} \sum_{j=1}^{N_T} m(\mathbf{x}_j^\top \hat{\theta})$.

Example

Suppose $\frac{\mathcal{P}_T(\mathbf{x})}{\mathcal{P}_S(\mathbf{x})} \approx \alpha^\top \mathbf{x}$ for some α . Then MrP is a approximately a weighting estimator.

$$\begin{aligned}\hat{\mu}^{\text{MrP}}(Y_S) &= \frac{1}{N_T} \sum_{j=1}^{N_T} m(\mathbf{x}_j^\top \hat{\theta}) \\ &\approx \int m(\mathbf{x}^\top \hat{\theta}) \mathcal{P}_T(\mathbf{x}) d\mathbf{x} && \text{(Law of large numbers)} \\ &= \int \frac{\mathcal{P}_T(\mathbf{x})}{\mathcal{P}_S(\mathbf{x})} m(\mathbf{x}^\top \hat{\theta}) \mathcal{P}_S(\mathbf{x}) d\mathbf{x} && \text{(Multiply by } \mathcal{P}_S(\mathbf{x})/\mathcal{P}_S(\mathbf{x}) \text{)}\end{aligned}$$

Approximately equivalent weights for (some) logistic regression MrP

- Suppose the model is $m(\mathbf{x}^\top \theta) = \text{Logistic}(\mathbf{x}^\top \theta)$, with MLE $\hat{\theta}$.
- MrP is $\hat{\mu}^{\text{MrP}}(Y_S) = \frac{1}{N_T} \sum_{j=1}^{N_T} m(\mathbf{x}_j^\top \hat{\theta})$.

Example

Suppose $\frac{\mathcal{P}_T(\mathbf{x})}{\mathcal{P}_S(\mathbf{x})} \approx \alpha^\top \mathbf{x}$ for some α . Then MrP is a approximately a weighting estimator.

$$\begin{aligned}\hat{\mu}^{\text{MrP}}(Y_S) &= \frac{1}{N_T} \sum_{j=1}^{N_T} m(\mathbf{x}_j^\top \hat{\theta}) \\ &\approx \int m(\mathbf{x}^\top \hat{\theta}) \mathcal{P}_T(\mathbf{x}) d\mathbf{x} && \text{(Law of large numbers)} \\ &= \int \frac{\mathcal{P}_T(\mathbf{x})}{\mathcal{P}_S(\mathbf{x})} m(\mathbf{x}^\top \hat{\theta}) \mathcal{P}_S(\mathbf{x}) d\mathbf{x} && \text{(Multiply by } \mathcal{P}_S(\mathbf{x})/\mathcal{P}_S(\mathbf{x}) \text{)} \\ &\approx \int (\alpha^\top \mathbf{x}) m(\mathbf{x}^\top \hat{\theta}) \mathcal{P}_S(\mathbf{x}) d\mathbf{x} && \text{(By assumption)}\end{aligned}$$

Approximately equivalent weights for (some) logistic regression MrP

- Suppose the model is $m(\mathbf{x}^\top \theta) = \text{Logistic}(\mathbf{x}^\top \theta)$, with MLE $\hat{\theta}$.
- MrP is $\hat{\mu}^{\text{MrP}}(Y_S) = \frac{1}{N_T} \sum_{j=1}^{N_T} m(\mathbf{x}_j^\top \hat{\theta})$.

Example

Suppose $\frac{\mathcal{P}_T(\mathbf{x})}{\mathcal{P}_S(\mathbf{x})} \approx \alpha^\top \mathbf{x}$ for some α . Then MrP is a approximately a weighting estimator.

$$\begin{aligned}\hat{\mu}^{\text{MrP}}(Y_S) &= \frac{1}{N_T} \sum_{j=1}^{N_T} m(\mathbf{x}_j^\top \hat{\theta}) \\ &\approx \int m(\mathbf{x}^\top \hat{\theta}) \mathcal{P}_T(\mathbf{x}) d\mathbf{x} && \text{(Law of large numbers)} \\ &= \int \frac{\mathcal{P}_T(\mathbf{x})}{\mathcal{P}_S(\mathbf{x})} m(\mathbf{x}^\top \hat{\theta}) \mathcal{P}_S(\mathbf{x}) d\mathbf{x} && \text{(Multiply by } \mathcal{P}_S(\mathbf{x})/\mathcal{P}_S(\mathbf{x}) \text{)} \\ &\approx \int (\alpha^\top \mathbf{x}) m(\mathbf{x}^\top \hat{\theta}) \mathcal{P}_S(\mathbf{x}) d\mathbf{x} && \text{(By assumption)} \\ &\approx \alpha^\top \frac{1}{N_S} \sum_{i=1}^{N_S} \mathbf{x}_i m(\mathbf{x}_i^\top \hat{\theta}) && \text{(Law of large numbers)}\end{aligned}$$

Approximately equivalent weights for (some) logistic regression MrP

- Suppose the model is $m(\mathbf{x}^\top \theta) = \text{Logistic}(\mathbf{x}^\top \theta)$, with MLE $\hat{\theta}$.
- MrP is $\hat{\mu}^{\text{MrP}}(Y_S) = \frac{1}{N_T} \sum_{j=1}^{N_T} m(\mathbf{x}_j^\top \hat{\theta})$.

Example

Suppose $\frac{\mathcal{P}_T(\mathbf{x})}{\mathcal{P}_S(\mathbf{x})} \approx \alpha^\top \mathbf{x}$ for some α . Then MrP is a approximately a weighting estimator.

$$\begin{aligned}\hat{\mu}^{\text{MrP}}(Y_S) &= \frac{1}{N_T} \sum_{j=1}^{N_T} m(\mathbf{x}_j^\top \hat{\theta}) \\ &\approx \int m(\mathbf{x}^\top \hat{\theta}) \mathcal{P}_T(\mathbf{x}) d\mathbf{x} && \text{(Law of large numbers)} \\ &= \int \frac{\mathcal{P}_T(\mathbf{x})}{\mathcal{P}_S(\mathbf{x})} m(\mathbf{x}^\top \hat{\theta}) \mathcal{P}_S(\mathbf{x}) d\mathbf{x} && \text{(Multiply by } \mathcal{P}_S(\mathbf{x})/\mathcal{P}_S(\mathbf{x}) \text{)} \\ &\approx \int (\alpha^\top \mathbf{x}) m(\mathbf{x}^\top \hat{\theta}) \mathcal{P}_S(\mathbf{x}) d\mathbf{x} && \text{(By assumption)} \\ &\approx \alpha^\top \frac{1}{N_S} \sum_{i=1}^{N_S} \mathbf{x}_i m(\mathbf{x}_i^\top \hat{\theta}) && \text{(Law of large numbers)} \\ &= \alpha^\top \frac{1}{N_S} \sum_{i=1}^{N_S} \mathbf{x}_i y_i && \text{(Property of exponential family MLEs)}\end{aligned}$$

Approximately equivalent weights for (some) logistic regression MrP

- Suppose the model is $m(\mathbf{x}^\top \theta) = \text{Logistic}(\mathbf{x}^\top \theta)$, with MLE $\hat{\theta}$.
- MrP is $\hat{\mu}^{\text{MrP}}(Y_S) = \frac{1}{N_T} \sum_{j=1}^{N_T} m(\mathbf{x}_j^\top \hat{\theta})$.

Example

Suppose $\frac{\mathcal{P}_T(\mathbf{x})}{\mathcal{P}_S(\mathbf{x})} \approx \alpha^\top \mathbf{x}$ for some α . Then MrP is a approximately a weighting estimator.

$$\hat{\mu}^{\text{MrP}}(Y_S) = \frac{1}{N_T} \sum_{j=1}^{N_T} m(\mathbf{x}_j^\top \hat{\theta}) = \frac{1}{N_S} \sum_{i=1}^{N_S} \underbrace{w_i^{\text{MrP}}}_{\alpha^\top \mathbf{x}_i} y_i + \text{Small error}$$

Approximately equivalent weights for (some) logistic regression MrP

- Suppose the model is $m(\mathbf{x}^\top \theta) = \text{Logistic}(\mathbf{x}^\top \theta)$, with MLE $\hat{\theta}$.
- MrP is $\hat{\mu}^{\text{MrP}}(Y_S) = \frac{1}{N_T} \sum_{j=1}^{N_T} m(\mathbf{x}_j^\top \hat{\theta})$.

Example

Suppose $\frac{\mathcal{P}_T(\mathbf{x})}{\mathcal{P}_S(\mathbf{x})} \approx \alpha^\top \mathbf{x}$ for some α . Then MrP is a approximately a weighting estimator.

$$\hat{\mu}^{\text{MrP}}(Y_S) = \frac{1}{N_T} \sum_{j=1}^{N_T} m(\mathbf{x}_j^\top \hat{\theta}) = \frac{1}{N_S} \sum_{i=1}^{N_S} \underbrace{w_i^{\text{MrP}}}_{\alpha^\top \mathbf{x}_i} y_i + \text{Small error}$$

But what are the weights? We don't observe $\frac{\mathcal{P}_T(\mathbf{x})}{\mathcal{P}_S(\mathbf{x})}$, so can't estimate α directly.

Approximately equivalent weights for (some) logistic regression MrP

- Suppose the model is $m(\mathbf{x}^\top \theta) = \text{Logistic}(\mathbf{x}^\top \theta)$, with MLE $\hat{\theta}$.
- MrP is $\hat{\mu}^{\text{MrP}}(Y_S) = \frac{1}{N_T} \sum_{j=1}^{N_T} m(\mathbf{x}_j^\top \hat{\theta})$.

Example

Suppose $\frac{\mathcal{P}_T(\mathbf{x})}{\mathcal{P}_S(\mathbf{x})} \approx \alpha^\top \mathbf{x}$ for some α . Then MrP is a approximately a weighting estimator.

$$\hat{\mu}^{\text{MrP}}(Y_S) = \frac{1}{N_T} \sum_{j=1}^{N_T} m(\mathbf{x}_j^\top \hat{\theta}) = \frac{1}{N_S} \sum_{i=1}^{N_S} \underbrace{w_i^{\text{MrP}}}_{\alpha^\top \mathbf{x}_i} y_i + \text{Small error}$$

Key idea (informal)

If $\hat{\mu}^{\text{MrP}}(Y_S)$ is approximately linear, then¹⁰ $w_i^{\text{MrP}} \approx N_S \frac{\partial \hat{\mu}^{\text{MrP}}(Y_S)}{\partial y_i}$.

¹⁰For MLEs, $\frac{\partial \hat{\mu}^{\text{MrP}}(Y_S)}{\partial y_i}$ is given by the implicit function theorem. (KP12; G+19)

Approximately equivalent weights for (some) logistic regression MrP

- Suppose the model is $m(\mathbf{x}^\top \theta) = \text{Logistic}(\mathbf{x}^\top \theta)$, with MLE $\hat{\theta}$.
- MrP is $\hat{\mu}^{\text{MrP}}(Y_S) = \frac{1}{N_T} \sum_{j=1}^{N_T} m(\mathbf{x}_j^\top \hat{\theta})$.

Example

Suppose $\frac{\mathcal{P}_T(\mathbf{x})}{\mathcal{P}_S(\mathbf{x})} \approx \alpha^\top \mathbf{x}$ for some α . Then MrP is a approximately a weighting estimator.

$$\hat{\mu}^{\text{MrP}}(Y_S) = \frac{1}{N_T} \sum_{j=1}^{N_T} m(\mathbf{x}_j^\top \hat{\theta}) = \frac{1}{N_S} \sum_{i=1}^{N_S} \underbrace{w_i^{\text{MrP}}}_{\alpha^\top \mathbf{x}_i} y_i + \text{Small error}$$

Key idea (informal)

If $\hat{\mu}^{\text{MrP}}(Y_S)$ is approximately linear, then¹⁰ $w_i^{\text{MrP}} \approx N_S \frac{\partial \hat{\mu}^{\text{MrP}}(Y_S)}{\partial y_i}$.

Note: The derivatives w_i^{MrP} now have two potentially distinct interpretations:

- **Equivalent weights:** A characterization of $Y_S \mapsto \hat{\mu}^{\text{MrP}}(Y_S)$ for diagnostics
- **Implicit weights:** An estimate of $\mathcal{P}_T(\mathbf{x})/\mathcal{P}_S(\mathbf{x})$

¹⁰For MLEs, $\frac{\partial \hat{\mu}^{\text{MrP}}(Y_S)}{\partial y_i}$ is given by the implicit function theorem. (KP12; G+19)