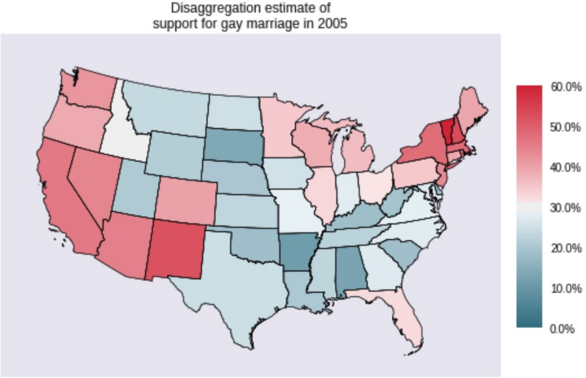
Multilevel Regression and Poststratification

Lucas Leemann
University of Zürich

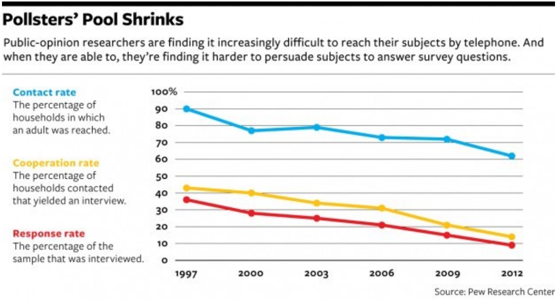
July 14, 2021

Motivation 1: Measuring public opinion in sub-national polls



Source: https://austinrochford.com/posts/2017-07-09-mrpymc3.html

Motivation 2: Non-Representative Samples and Low Participation



Motivation 1: Measuring public opinion in sub-national polls

- Disaggregation: Take mean of all respondent in a given district j
  - Regular poll: For some units very few observations....
  - Mega-Poll: Aggregate many polls with the same question and look at the average per district (Erikson et al. 1993)
- Post-Stratification (or raking): Use post-stratification to come up with weights for each person to mimic characteristics of state j
  - Fine if Unit 1 and Unit 2 only differ because in Unit 1 has a different structure
  - Fails to acknowledge local idiosyncrasies

Motivation 2: Non-Representative Samples and Low Participation Samples

- Random sampling assumes that participation is orthogonal to variable of interest.
- Many polls are not based anymore on random samples. There is variation: YouGov (good) vs. GOP poll on Trump

  see e.g. https://news.vice.com/en_us/article/d34wda/the-seriously-frugged-up-practice-of-using-fake-polls-in-politics
- May help but will not always help.
- Lack of *ex ante* indicator – we are left with *ex post* indicators.

Motivation(s)

- (1) Classic academic interest:
  We have a national poll but would like to exploit the information for sub-national measures.
- (2) Survey research faces challenges - sometimes MrP can help.

Partial Pooling

(the secret ingredient)

Example: Partial Pooling

Hypothetical example:

Imagine that you observe from every canton (j) a low number of people (i) and you want to estimate average attitudes based on those few observations per canton. Every person will tell you where they would locate themselves on a left-right axis from 0 to 9 (Y_{ij}).

- Consider political left-right self-placement in Switzerland.
- We have 26 cantons: $j = 1, 2, \cdots, 26$.
- In each canton we have a sample of n_j voters.
- We would like an estimate for $E(y_j)$, canton-specific value of y.

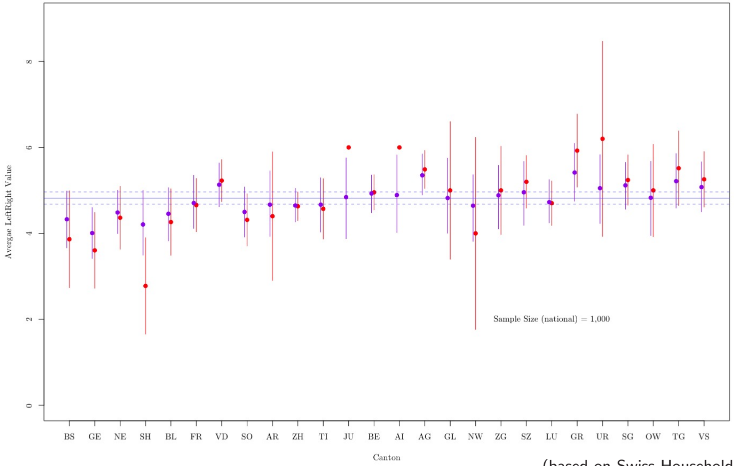
Partial Pooling 1

We have, in principle, two possible estimates we can use:

- Estimate for the entire country: $\bar{y} = \frac{\sum_j \sum_i y_{ij}}{N}$
  - Will have low variance...
  - ...but cannot distinguish between cantons.
- Estimate for each canton: $\bar{y}_j = \frac{\sum_{i \in j} y_{ij}}{N_j}$
  - Will yield an estimate for each canton...
  - ...but with high variance since some cantons contribute very few observations.

$\rightarrow$ It would be nice if we could exploit the information from other cantons as well but still produce different estimates for each canton.

Random effects enable partial pooling



(based on Swiss Household Panel 1999)

Partial Pooling 2

Grand mean (GM):

You disregard the structure of the data:
$y_{ij} = \beta_0 + \varepsilon_{ij}$
You only estimate one mean for all units.

Fixed effect (FE):

You add a dummy variable for each unit j leading to a model like this:
$y_{ij} = \beta_0 + \beta_1 \cdot d_1 + \beta_2 \cdot d_2 + ... + \beta_{j-1} \cdot d_{j-1} + \varepsilon_{ij}$ That is the same as estimating a separate $\hat{y}_j$ for each unit j.

Random Effect (RE)

A compromise between the *FE* and the *grand mean* - it's like magic:
$y_{ij} = \beta_0 + \alpha_j + \varepsilon_{ij}$ where we define that $\alpha_j \sim N(0, \sigma^2_\alpha)$

Partial Pooling 3

Random effects:

- Something between overall mean (GM) and unit specific mean (FE)
    - If there are few observation in a unit it should be closer to the grand mean
    - If our unit estimate is noisy it should be closer to the grand mean

Approximation from Gelman and Hill (2007: 253):

$$\bar{\alpha}_j^{multilevel} \approx \frac{\frac{n_j}{\sigma_j^2}\bar{y}_j + \frac{1}{\sigma_\alpha^2}\bar{y}_{all}}{\frac{n_j}{\sigma_j^2} + \frac{1}{\sigma_\alpha^2}}$$

$n_j \rightarrow$ Number of observations in unit j

$\bar{y}_j \rightarrow$ Average value in unit j

$\bar{y}_{all} \rightarrow$ Average value over all observations

$\sigma_j^2 \rightarrow$ Variance within unit j

$\sigma_\alpha^2 \rightarrow$ Variance among the unit averages $\bar{y}_j$

Multilevel Regression with Post-Stratification
(MrP)

MrP

- Gelman and Little (1997)
  - We use a model to make predictions for ideal types and use *post-stratification*
  - We use a MLM and have a RE for locality (keeping local idiosyncrasies)
- MRP outperforms alternatives like disaggregation (Lax and Phillips, 2009; Warshaw and Rodden, 2012)
- "...emerging as a widely used gold standard for estimating constituency preferences from national surveys." (Selb and Munzert, 2011: 456)

→ Much more efficient use of the data.
→ Allows to include level 2 predictors.

MRP Example (simplified) in 4 steps

1. Survey with N respondents
2. $Pr(y_i = 1) = \Phi\left(\beta_0 + \alpha^{gender}_{j[i]} + \alpha^{educ}_{m[i]} + \alpha^{constituency}_{c[i]}\right)$, whereas c is for constituency, j for sex, and m for education groups
3. For all ideal voter types (men/women and low/high education) we predict the support of a policy

$$\hat{\pi}_{jm\in c} = \Phi\left(\hat{\beta}_0 + \hat{\alpha}^{gender}_{j} + \hat{\alpha}^{educ}_{m} + \hat{\alpha}^{constituency}_{c}\right)$$

4. Weigh each prediction by the relative share of such voters in a constituency

$$\hat{\pi}_c = \frac{\sum_j \sum_m N_{jm\in c} \cdot \hat{\pi}_{jm\in c}}{N_c}$$

MRP: The second step

$$Pr(y_i = 1) \quad = \quad \Phi\left(\beta_0 + \alpha^{gender}_{j[i]} + \alpha^{education}_{k[i]} + \alpha^{age}_{m[i]} + \alpha^{district}_{n[i]}\right)$$

$$\alpha^{region}_{o} \sim \quad N(0, \sigma^2_{region}),\ \text{for } o = 1, ..., 7$$

$$\alpha^{district}_{n} \sim \quad N(\alpha^{region}_{o[n]} + \boldsymbol{\beta}\mathrm{X}_n, \sigma^2_{district}),\ \text{for } n = 1, ..., N$$

$$\alpha^{gender}_{j} \sim \quad N(0, \sigma^2_{gender}),\ \text{for } j = 1, 2$$

$$\alpha^{education}_{k} \sim \quad N(0, \sigma^2_{education}),\ \text{for } k = 1, ..., 6$$

$$\alpha^{age}_{m} \sim \quad N(0, \sigma^2_{age}),\ \text{for } m = 1, ..., 4$$

$\boldsymbol{\beta}\mathrm{X}_n$: Level-2 variables, explaining differences among districts (presidential vote share, german-speakers).

β_0 is the "grand mean".

MRP: The third step

We analyze the variance among individuals and districts...

The estimate for the average support of a 20-35 year old female university graduate in a specific unit is influenced by

- all people 20-35 years old in the survey,
- all woman,
- all university graduates,
- all people from that unit,
- and all people from the same region in which that unit lies

Partial pooling of MLM facilitates to retrieving more precise estimates – here, we create for each district 48 ideal types and their average support ($2 \times 6 \times 4 = 48$).

MRP: The forth step

We now weigh each of the 48 ideal types by their relative share in the population....

.... what do I need to know about the people in a specific unit?

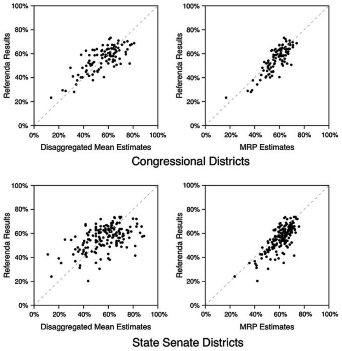
Differences in district estimates will hence come from:

- Different population structure
- Different responses ($\alpha_n^{district}$)
- Different level 2 variables

How Good is MrP?

Motivation 1: Small Area Estimation

FIGURE 2 Cross Validation of MRP Estimates with Same–Sex Marriage Ref. in Arizona, California, Michigan, Ohio, Wisconsin



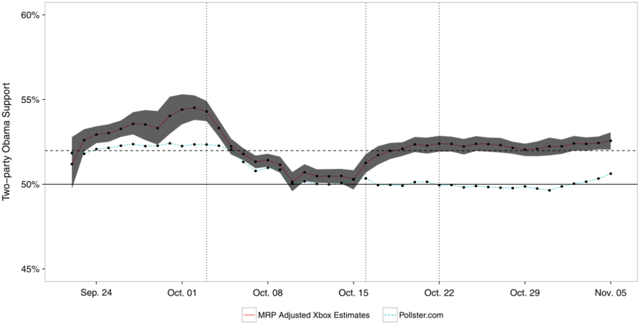
Note: This figure shows that in national samples of 17,000, MRP outperforms disaggregation for predicting state referenda results on same-sex marriage.

Warshaw and Rodden (2012: 212)

Motivation 2: Non-Probability Samples

XBox players in the US, Presidential election 2012 (Obama vs Romney)

W. Wang et al. / International Journal of Forecasting 31 (2015) 980–991



Wang et al. (2014) →Link to paper

"We conclude by arguing that non-representative polling shows promise not only for election forecasting, but also for measuring public opinion on a broad range of social, economic and cultural issues"

Extensions

(Meeting the entire family, i.e. MrsP and autoMrP)

1) Level 1 Improvement

Leemann, Lucas and Fabio Wasserfallen. 2017. “Extending the Use and Prediction Precision of Subnational Public Opinion Estimation” *American Journal of Political Science* 61(4): 1003-1021.

Level 1 Problem

Restrictive requirement of MrP...

- One needs very fine-grained information for the post-stratification step
  (e.g. # of white men with high school degree between 30-44 years old)
  $\rightarrow$ requires census
- One can only use individual information (demographic predictors) which is part of the census data

Stylized example: We need to know exactly how many highly educated men we have in constituency c to compute $\hat{\pi}_c$

	♂	♀	
low education	40%	20%	60%
high education	20%	20%	40%
	60%	40%	100%

Level 1 Problem

- The non-constant first derivative of the link function implies that we need the joint distribution (j: sex, m: educ)

$$\frac{\sum_j \sum_m F\left(\hat{\beta}_0 + \hat{\alpha}_m + \hat{\alpha}_j + \hat{\alpha}_c\right) \cdot N_{jm \in c}}{N_{n \in c}} \stackrel{?}{=} F\left(\frac{\sum_j \sum_m (\hat{\beta}_0 + \hat{\alpha}_m + \hat{\alpha}_j + \hat{\alpha}_c) \cdot N_{jm \in c}}{N_c}\right)$$

- If $F()$ is identity fct equality holds!
- If $F()$ is logit fct equality DOES NOT hold!
- The effect of adding $\hat{\alpha}^{sex}_{j=♂}$ to a *low education* person is different than when adding it to a *high education* person
  (non constant marginal effect in logit model)

University of Zurich[UZH]

Level 1 Problem

Very stringent data requirements:

One needs to know for each sub-national unit the exact number of people who correspond to an ideal type.

- MrP is mostly used in US and sometimes in developed countries (Switzerland, Germany, UK)
- MrP is used with *suboptimal* response models, Warshaw and Rodden (2012): "Because `census factfinder` does not include age breakdowns for each race/gender/education subgroup, we are not able to control for respondents' age in our model." (p.208)

Alternative: A linear link function for response model (MPSA 2013).

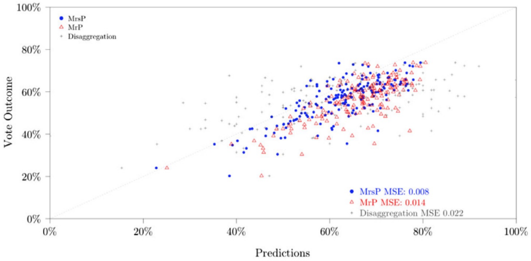
Level 1 Solution: MrsP (MrP's Better Half)

- Multilevel regression with *synthetic* post-stratification allows to include them.
  - Simple MrsP: Assume that new variable is uncorrelated with any other variable.
  - Elaborate MrsP: Use Survey data to estimate correlation structure.

Level 1 Solution: MrsP (MrP's Better Half)

Example Elaborate MrsP

FIGURE 4 Public Vote Outcomes and Disaggregation, Classic MrP, and MrsP Estimates for the Warshaw and Rodden (2012) Analysis on Same-Sex Marriage Referendums in Arizona, California, Michigan, Ohio, and Wisconsin



2) Level 2 Improvements

Broniecki, Philipp, Lucas Leemann, and Reto Wueest. 2021. “Improved Multilevel Regression with Post-Stratification Through Machine Learning (autoMrP)” *Journal of Politics* forthcoming.

https://lucasleemann.files.wordpress.com/2020/07/automrp-r2pa.pdf
https://cran.r-project.org/web/packages/autoMrP/index.html

Context-Level Variables

$$
\begin{aligned}
Pr(y_i = 1) \quad &= \quad \Phi\left(\beta_0 + \alpha^{gender}_{j[i]} + \alpha^{education}_{k[i]} + \alpha^{age}_{m[i]} + \alpha^{district}_{n[i]}\right) \\
\alpha^{region}_{o} \sim \quad & N(0, \sigma^2_{region}),\ \text{for } o = 1, ..., 7 \\
\alpha^{district}_{n} \sim \quad & N(\alpha^{region}_{o[n]} + \boldsymbol{\beta}\mathrm{X}_n, \sigma^2_{district}),\ \text{for } n = 1, ..., N \\
\alpha^{gender}_{j} \sim \quad & N(0, \sigma^2_{gender}),\ \text{for } j = 1, 2 \\
\alpha^{education}_{k} \sim \quad & N(0, \sigma^2_{education}),\ \text{for } k = 1, ..., 6 \\
\alpha^{age}_{m} \sim \quad & N(0, \sigma^2_{age}),\ \text{for } m = 1, ..., 4
\end{aligned}
$$

$\boldsymbol{\beta}\mathrm{X}_n$: Level-2 variables, explaining differences among districts (presidential vote share, german-speakers).

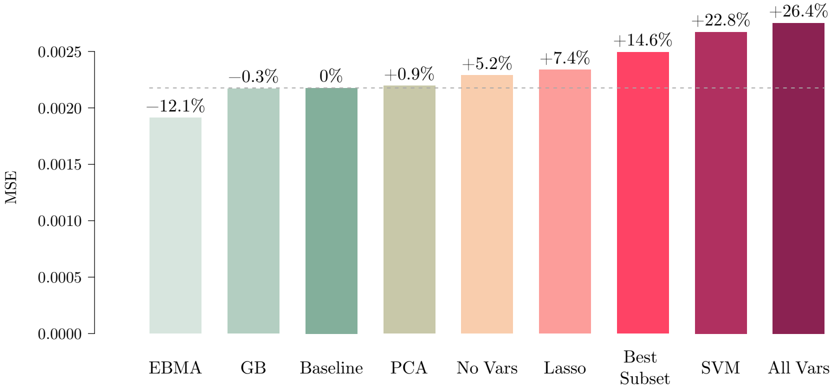
The Standard for Selection of Contextual Information

- Level 2 features are important (Warshaw and Rodden, 2012)
- Inclusion of plausible candidates but neither explicitly justified nor systematically chosen
- See Park et al. (2006); Enns and Koch (2013); Lax and Phillips (2009); Warshaw and Rodden (2012); Ghitza and Gelman (2013); Tausanovitch and Warshaw (2013); Eggers and Lauderdale (2016)
- Systematic Selection?
    - Maximize model fit?
    - Select variables that seem to correlate with the DV?
- Select no context variables: Underfitting
- Select too many context variables: Overfitting

autoMrP

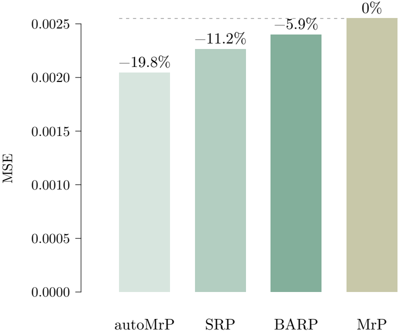
- Five simple classifiers (best subset, Lasso, PCA, GB, SVM)
- Additional classifiers can be added
- Systematic & flexible combination (via EBMA)
- $\rightarrow$ automatic MrP allowing for systematic model specification

Performance of Classifiers and Baselines



Note: $N = 1,500$.

autoMrP vs Alternatives



Note: Average MSE of state-level predictions over 89 survey items. *MrP* is the standard MrP model with all context-level variables. Percentage numbers: Comparison to standard MrP model.

Conclusion

Summary

- MrP as a model-based alternative to raking or post-stratification.
- MrP allowing to generate good estimates for small areas.
- Cost (1) : Data requirement and complexity.
- Cost (2) : Not observation-specific but outcome-specific.
- A question that will not go away: How can we handle non probability samples?
- Practical input: Some examples in the lab