Measuring Attitudes – Multilevel Modeling with Post-Stratification (MrP)*

Lucas Leemann and Fabio Wasserfallen

Abstract

Multilevel regression with post-stratification, short MrP, has quickly become a standard model to measure public opinion across e.g. geographical units. This chapter shows how it outperforms other alternatives and illustrates MrP with concrete examples. Readers are walked step-by-step through the construction of a MrP model – code and data are available for replication. We also discuss and illustrate the relevance of context-level factors and how to generate uncertainty estimates of our measurements. The chapter ends with a discussion of extensions and ongoing research projects that promise to further improve MrP.

1 Introduction

Reliable information on public opinion is for the empirical analysis of many political science questions a *conditio sine qua non.* Not only for the whole sub-field of political behaviour, but also for researchers interested in, for example, the congruence of public opinion with the legislator or government action (and how that is shaped by institutions). The basic methodological challenge for research on public opinion is to make valid inferences from a collected survey sample to the larger (underlying) population. Also, there may be several (sub-)populations of interests if we want to compare, for example, public opinion across the U.S. states with national polling

*A complete replication file with an illustrative example can be found here: https://github.com/lleemann/MrP_chapter. Draft of a Chapter Contribution to the *Sage Handbook of Research Methods in Political Science and International Relations* edited by Luigi Curini and Robert J. Franzese Jr.

data that only includes a small number of respondents for certain states. We will present and discuss in this chapter multilevel regression and post-stratification, so-called MrP, which has made a seminal contribution to the estimation of public opinion for sub-national units, and also advances public opinion research in several other respects.

Often, we collect national polls, but may be interested in the estimation (and comparison) of public opinion on lower constituency levels (such as states or congressional districts in the U.S., Bundesländer or Wahlkreise in Germany, and cantons and municipalities in Switzerland). The problem is that the sub-samples of respondents in the poll for some units—particularly the smaller ones—are too small. Small samples will increase the estimates mean squared error - these estimates remain unbiased but their variance is large in small samples. In addition (and related to) this small-n problem, samples are often not representative for the larger population. The usual fixes for these caveats have been that researchers pool data from several surveys (to increase the sample) and that they use raking or post-stratification for calibration, which, in effect, induces the structure of the underlying population on the sample (Miller and Stokes, 1963; Erikson, Wright, and McIver, 1993; Zhang, 2000). The more recent literature has established MrP as superior solution to these methodological problems with several studies testing, validating, end extending the method (Gelman and Little, 1997; Lax and Phillips, 2009*b*; Warshaw and Rodden, 2012; Leemann and Wasserfallen, 2017).

MrP is powerful because it combines the multilevel modeling of the survey data with post-stratification, using information on the population structures for which we aim to derive estimates. In short, researchers using MrP a) organize the data in fine-grained ideal types, b) make predictions for each ideal type using (national) survey data, and c) use post-stratification for the estimation of public opinion of the (sub-national) constituencies of interest. This approach provides more precise estimates in case researchers face the small-n problem discussed above and/or work with skewed samples. Equally (or even more) important is that MrP provides a flexible and general framework that has opened avenues for constant improvements, fruitful combinations with other methods, and sophisticated applications that are nicely tailored to specific data and research design challenges.

In the next section we present the basics of MrP and discuss technical issues, methodological limitations and extensions of the method. In essence, MrP is a stepwise framework that is conducive to flexible adaptations. Many different substantive political science puzzles have been empirically addressed with different versions of MrP. The contributions to various political science literatures in that respect are quite impressive, considering that MrP is still a young method. Among others, MrP has contributed to the literatures on Political Behavior, American Politics, Federalism, Comparative Politics, and Institutions. More specifically, several MrP studies generated important insights to the study of partisanship, ideology, the responsiveness of institutions to public opinion, congruence of elite and voter preferences, and polarization (e.g., Lax and Phillips, 2012; Tausanovitch and Warshaw, 2013; Leemann and Wasserfallen, 2016; McCarty et al., 2018). The impressive track-record of MrP and its broad scope of substantive applications (also beyond political science in sociology, for example) may be considered to be the greatest contribution of the still rather young method. In that sense, MrP is much more than the successor of raking or post-stratification—with substantial potential for future research.

As far as methodological advances are concerned, Ghitza and Gelman (2013) elaborated a method for the analysis of deeply interacted subgroups, Leemann and Wasserfallen (2017) developed a MrP approach that relaxes the data requirement and increases the prediction precision, while Kousser, Phillips, and Shor (2016) and Tausanovitch and Warshaw (2013) combine MrP with joint scaling and IRT. These are just a few examples of some recent methodological improvements that are, in all these cases, sophisticated answers to specific research design challenges. In addition to making academic contributions, MrP is also widely applied in the real world. This comes as no surprise, given that political actors are eager to get precise estimates of public opinion in election and referendum campaigns (Hanretty, Lauderdale, and Vivyan, 2016*a*). In additon, MrP's logic of ideal types lends itself straightforwardly to the analysis of public opinion in specific sub-groups, which is of importance for political actors that want to develop campaigning efforts targeting specific sub-groups of the population. A further attractive feature of MrP is that it allows to leverage large datasets, also from samples with unequal participation (Wang et al., 2015; Downes et al., 2018). In the digital age, such data become the norm.

We illustrate in this chapter the basics of MrP with a running example, before we discuss several technical issues, extensions, and advances of the method. Moreover, we provide a broad overview of MrP, as it is applied in several (sub-)fields of the political science literature, by also discussing some (selected) substantive insights that we could gain thanks to several sophisticated studies. The broad spectrum of MrP applications is quite impressive and shows that MrP is, next to its methodological appeal, an approach that has great potential to further contribute important insights to various puzzles of political science. Finally, we also elaborate on challenges and limitations of MrP and speculate on future developments.

2 Multilevel Regression with Post-Stratification (MrP)

Raking and post-stratification are two ways to generate weights for each individual observation. The idea of both methods is to rely on some stratifying variables (e.g., education, gender, age) to generate weights such that the weighted sample has the same marginal or joint distribution (with respect to these variables) as the target population.

Let's take a step back and look at the estimates of the outcome variable rather than how to generate weights for the observation. In both, raking and post-stratification, we generate weights for each individual observation. Another way to think about this is that we organize the data in ideal types. If we were to post-stratify by education (six levels), gender (two levels), and age (four categories). We would think of the population as consisting of 48 different ideal types. Each individual in the sample can be assigned to a specific ideal type that is defined by the individual's education, gender, and age. To generate weights, we can look at how large the share of respondents is that belong to a specific ideal type and how large that share should be in the target population. The weight is chosen such that the weighted share in the sample is equal to the actual share in the target population.

For raking or post-stratification, one just takes the mean of the outcome variable among all people belonging to the same ideal type. That estimate is then the best guess for the outcome variable for people of that ideal type. If the outcome variable is the support for a specific politician, the estimate will be between 0 and 1 and be interpreted as the share of people of that ideal type that support a specific politician.

What multilevel regression with post-stratification (MrP) does is different in the way we determine the estimate of the outcome variable for a specific ideal type. Rather than using this estimate for the ideal type's average support, MrP relies on a model to estimate the support among all survey respondents (Gelman and Little, 1997; Park, Gelman, and Bafumi, 2004*a*). Based on this model, we can generate different predictions for different ideal types. This is the main technical difference to simpler methods, such as raking or post-stratification. As we show later, relying on a model rather than just a simple mean opens up many doors for analysis.

2.1 An Example: Voting on a Minaret Ban in Switzerland

An example will help to illustrate MrP and its extensions discussed later. The example here is an (in)famous vote in Switzerland in 2009 when a slight majority of the Swiss population supported an initiative to ban the future construction of minarets (cite some work on the minaret ban). Surveys in advance of the vote indicated that it would be rejected and to the surprise of many it was accepted. The example here will rely on a dataset from a post-vote CATI carried out by universities in cooperation with a private company.

We start the illustration with a fairly simple MrP analysis with only include individual-level variables (education, gender, age). We thus think of the sample and also the population of consisting of 48 different ideal types.[1] At the core of MrP is a multilevel (or hierarchical) response model with the variable of interest as the outcome (in this example a yes or no vote for the minaret ban). It can be modeled as a probit or a logit - we opt for the probit. Education, gender, and age are added to the model as random effects such that each education group has its own draw from a common normal distribution as well as the groups for gender and age. In addition, we add a random effect for the subnational units (here cantons) to the model. If attitudes vary regionally (beyond variation due to different demographics), we account for this in the subnational (here cantonal) random effects. There are 26 cantons in Switzerland and this leads to 1248 (26×48) different ideal types.

[1]Gender has two categories, education has six categories ((mandatory schooling or no response, apprenticeship, university-entrance diploma (Matura) and teachers college, additional job training, advanced training, university degree including universities. of applied sciences), and age has four categories (18-34, 35-49, 50-64, 75–).

$$
\begin{aligned}
Pr(y_i = 1) &= \Phi\left(\beta_0 + \alpha^{education}_{k[i]} + \alpha^{gender}_{j[i]} + \alpha^{age}_{l[i]} + \alpha^{canton}_{n[i]}\right) \qquad (1) \\
\alpha^{education}_{k} &\sim N(0, \sigma^2_{education}), \text{ for } k = 1, ..., 6 \\
\alpha^{gender}_{j} &\sim N(0, \sigma^2_{gender}), \text{ for } j = 1, 2 \\
\alpha^{age}_{m} &\sim N(0, \sigma^2_{age}), \text{ for } m = 1, ..., 4 \\
\alpha^{canton}_{n} &\sim N(0, \sigma^2_{canton}), \text{ for } n = 1, ..., 26
\end{aligned}
$$

This model yields estimates for the parameter $\hat{\beta}_0$ as well as for the realizations of the random effects (e.g., for the first education group it is $\hat{\alpha}^{education}_{k=1}$). This allows us to create a prediction for any of the 48 ideal types by adding the respective random effect realization of that ideal type. For any ideal type, we can add the constant to the four chosen random effect realizations and have an ideal type's score on the latent variable. After transforming that - via the cumulative standard normal distribution - to a probability, we derive an estimate of the share of a specific ideal type that is expected to support the minaret ban. This is markedly different from what one does when relying on raking or post-stratification, where we would have just taken the average response (among all respondents of that ideal type). In this example, there are actually less observations than ideal types, but since the predictions are model-based this does not pose a problem.

The second step of MrP is post-stratification. We generate an average support for each ideal type and then weigh this by the relative share of a type in the target population (N_{ng} denotes the the number of people of type g living in subnational unit n). This is only possible with precise information on the structure of the target population. Often used variables such as age, education, gender, and (in the US) race are known due to the census. Here, we need to know exactly how many people of a specific age group, education, and gender live in a specific subnational unit. This allows us to determine the average support in a subnational unit by weighing the support per ideal type (denoted as g) by the share of that type in the unit:

$$
\hat{\pi}_n = \frac{\sum_{g \in n}^{G} \hat{\pi}_{ng} N_{ng}}{N_{n\cdot}} \qquad (2)
$$

The running example here is the vote on the minaret ban in Switzerland in 2009. The raw data of the exit poll suggests that a majority of 51.5% of the electorate was against the ban. In other surveys leading up to the vote, a even larger majority seemed to oppose the ban. It was to the great surprise of many observers when on the vote's day the official results were announced: 57.5% of voters supported the ban. This triggered a larger debate about the values of public opinion polling in Switzerland. Would MrP have helped here? Yes, the simple MrP model described above provides an estimate of 58.0%, which is very close to the true value.

2.2 Using MrP to Generate Subnational Preference Estimates

As mentioned above, MrP has received a lot of scholarly attention mostly due its ability to generate credible estimates of subnational preferences (Lax and Phillips, 2009*a*,*b*). So far, we have only used MrP to correct for the unequal make-up of the sample's structure (non-response bias). In a next step, we will use additional context-level variables to generate subnational estimates. We can add context-level explanatory variables by redefining the distribution of the cantonal random effect to be $\alpha_n^{canton} \sim N(\boldsymbol{\beta}\mathbf{X}_n, \sigma^2_{canton})$ whereas $\mathbf{X}$ is a matrix with a leading column of 1's. Context-level variables were not included in the original MrP paper (Gelman and Little, 1997), but from Park, Gelman, and Bafumi (2004*a*) onwards many published MrP models include context-level variables.

Warshaw and Rodden (2012) show – looking at US data– that using context-level variables improves the district-level estimates. Since the estimated support for an ideal type is based on a model prediction rather than a simple sample average, one can easily include additional information into the model. With US data, the most common variable used is the presidential vote share in the preceding election, but other frequently used variables are median income, share of veterans, or share of evangelical Protestants and Mormons.

In the example here, we can also add a context-level variable. We are looking for a variable that only varies across cantons and that is likely correlated with the collective part of the voting decision. Variables picking up on political culture or more structural variables such as unemployment or income levels could be used as well. But since Swiss voters vote frequently on issues we can actually use a past vote on a related issue. In 2008 an initiative wanted to change the constitution – to de facto reverse a ruling by the highest court – to make it easier for municipalities to vote on naturalization cases. The issue at hand and the likely motivating factors as well as the partisan vote recommendations were similar as in the minaret ban vote. We include the share of people voting in favor of that initiative (which was eventually rejected) to the response model.

Figure 1: Cantonal Estimates and True Approval Rate

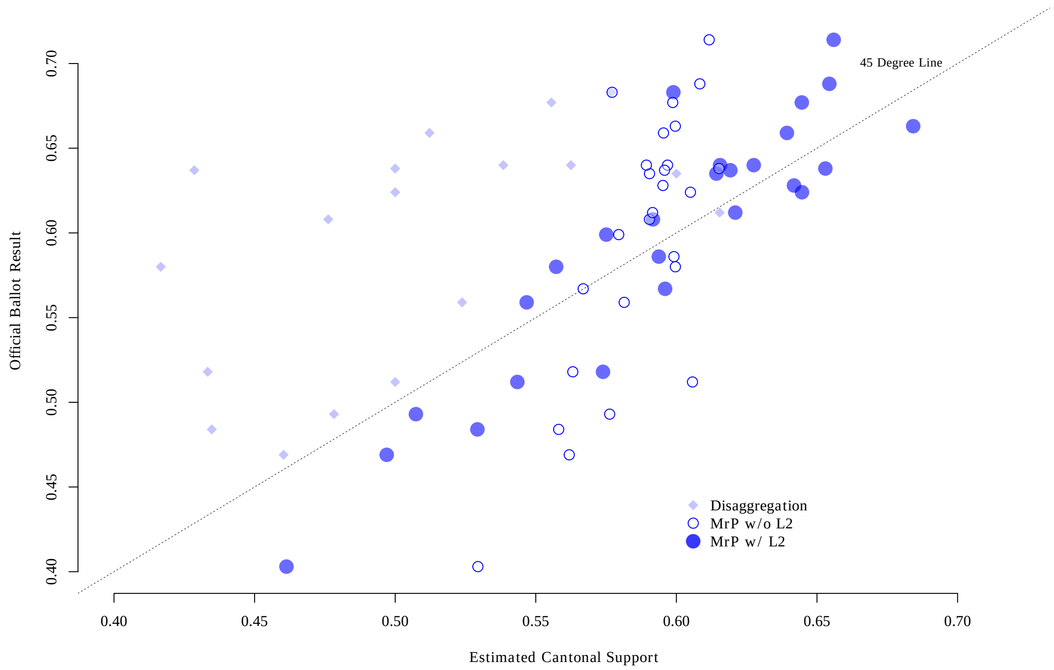


Figure 1 shows the estimates based on the raw data and two different MrP models. The estimates based on the raw data take the average response per canton. The MrP without any context-level variable is what was defined in Equation 1, and the third model includes the vote share in the earlier vote. The disaggregation estimates do not perform well due to the low sample size per canton. This is in line with prior research (Lax and Phillips, 2009*b*; Warshaw and Rodden, 2012; Leemann and Wasserfallen, 2017). In addition, we see that the MrP estimates based on a model with no context-level variable yields estimates biased towards the sample mean (here 48.5%). We discuss this in subsection 2.3 in more detail. Finally, adding a context-level predictor increases prediction accuracy and the estimates based on MrP with a context-level variable are neatly clustered around the 45 degree line.

2.3 Some Technical Remarks

Why is it that MrP performs so much better than raw data in the estimation of sub-national public opinion? Part of the answer is that the model is estimated on all observations in the sample, while the disaggregation results only rely on responses from a specific canton. The key is in the multilevel model used here that allows for partial pooling. When estimating the realizations of the random effects, for example, for the four age groups, all observations are used. When then generating predictions (e.g., for a young well-educated woman in the canton of Zurich) the model also benefits from men in other cantons in different education groups.

Partial pooling comes into the model by incorporating random effects. Random effects can be thought of as a weighted average between a global estimate and a local estimate, whereas the specific weights are based on the entire variation and the local variation (see p. 253-254 in Gelman and Hill (2007*a*) for a more thorough discussion). Hence, the multilevel model allows for a more efficient use of the data.

A consequence of partial pooling is that the cantonal random effects can be biased towards 0 and this leads in the predictions to a bias towards the sample mean. The sample mean within a unit is less influential on that unit's random effect realization if there are few observations within the unit (Gelman and Hill, 2007*a*, p. 253). As a consequence, the partial pooling will be stronger for smaller units with fewer respondents (Buttice and Highton, 2013, p. 5). This can be nicely seen in Figure 1 where the estimates of MrP without a context-level variable are too steep and show little variation across cantons.

Once we add context-level variables the subnational estimates vary much more since they enter the model as a regular variable and we estimate a fixed coefficient.[2] Almost all MrP applications now rely on a multilevel model with context-level variables and it is good practice to include them.

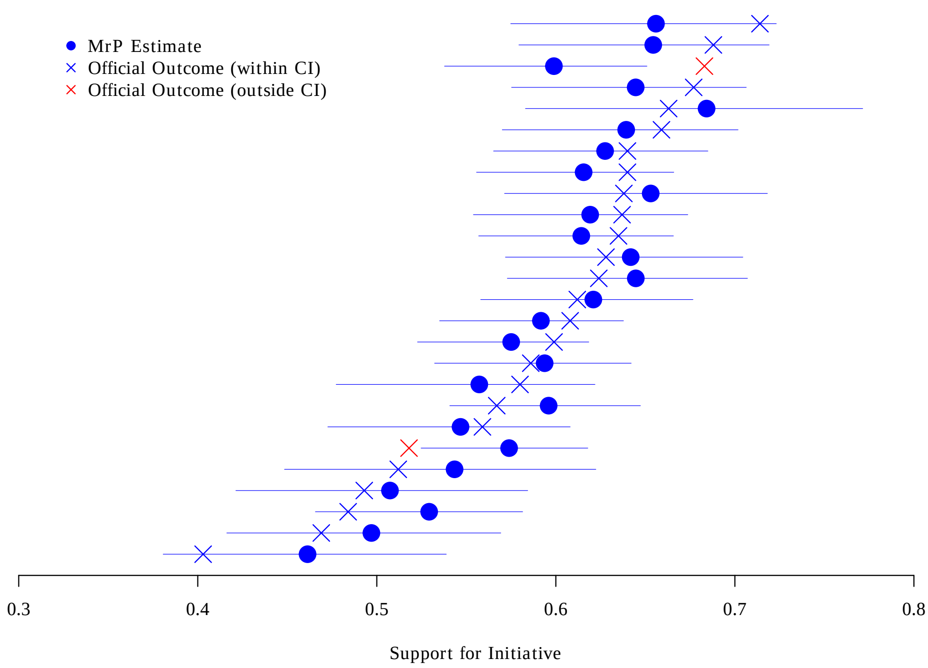
Finally, it is also worth considering that MrP is partly based on post-stratification. And post-stratification will perform best when the there is little variation within a cell (here: specific combination of gender, age, education, and canton) but large variation across cells. That is the

[2]Fixed refers here to a coefficient that is estimated as a specific value unlike random effects where we estimate the variance part. This is not what is often referred to as with-in estimator. See p. 245-246 in Gelman and Hill (2007*a*) for a more detialed discussion.

case for any procedure that relies on strata, cells, or ideal types. MrP will perform better the more homogenous preferences are within an ideal type and the more variation there is among different ideal types.

One issue we have not yet touched upon is uncertainty. We presented the estimates as point estimates and did not discuss any strategy to generate the appropriate confidence intervals. Since the estimates are based on model predictions, we can easily generate the uncertainty estimates by relying on simulation (Herron, 1999; Gelman and Hill, 2007*b*). This approach works whether one relies on `stan` or `lme4` to estimate the response model. We can illustrate how certain or uncertain we are by displaying the 95% confidence interval for the minaret vote.

Figure 2: Cantonal Estimates and Uncertainty



Most confidence intervals have a width of about 12 percentage points and 24 of the 26 cantons had an actual result that was in line with the MrP estimate, i.e. the official outcome was within the confidence interval.

This section described the standard MrP model and used an example to illustrate that MrP can help with two different - but frequent - challenges encountered when working with survey data.

On the one hand, MrP can deal well with data sets that suffer from non-response bias as it is found in many standard surveys used by academics. On the other hand, it enables researchers to generate subnational estimates that are by far superior to disaggregation.

Finally, a last point that has not yet received a lot of attention is how to select context-level variables. In the running example here it is easy to come up with a related but older ballot vote and to use those results as the context-level variable. But what do we do if we want to estimate support or opposition to a new EU treaty in a country that does not hold regular ballot votes on issues? It is worth stressing that the selection of context-level variables is important for two reasons. First, those are the only variables that enter the model as 'fixed effect' (in the hierarchical meaning and not meant to describe the within-estimator). Since MrP engages in prediction we should worry about overfitting (James et al., 2013). Second, it has been shown that using optimal context-level variables can greatly increase the prediction precision of MrP (Warshaw and Rodden, 2012). While authors discuss the variables included in the model, there is often no systematic justification provided for the specific choice.[3] We return to this issue below.

2.4 Methodological Limitations and Extensions of MrP

MrP has proven to be a valuable methods in the political science tool box, but it also comes with limitations and downsides. For one, MrP may work well to estimate district-level preferences when we have at least some responses in every district. However, MrP will not work to one's satisfaction at increasing degrees of granularity – at some point, the identified small areas are too small and one will have hardly any respondents in the sample.

But beside this general limitation there are also more specific points. Buttice and Highton (2013) published a paper that shows that the performance of MrP can vary and that there is no guarantee that MrP estimates are precise. They argue that MrP performs well for cultural political attitudes in the US but that this does not extend to all areas. The strength of this paper is that they work through various aspects that are important.

One relevant aspect to assess whether MrP can do well or not is the intra-class correlation

[3] There is one exception - Leemann and Wasserfallen (2016) touch on this issue in the appendix and show that they try to pick context-level variables based on AIC and BIC measures of the estimated models.

(ICC). The ICC accounts for how strongly opinions vary across states and how strongly they vary within states. Buttice and Highton (2013) argue that MrP performs strongest when almost all variance is on the context-level (for them the state level). In addition, they also underline the value of having context-level variables in the multilevel model.

Another point made by Buttice and Highton (2013) is that researchers cannot rely on a canonical response model when analyzing different issues. More practically expressed, the hierarchical response model should be differently specified when researchers analyze public opinion on cultural issues (e.g., same sex-marriage) or economic questions (e.g., tax levels). This echos the point made by Warshaw and Rodden (2012), who advice to include variables that are strong predictors of variation across units in the response model. How to chose optimal context-level variables remains a difficult task. As we discuss in Section 4, machine learning applications of MrP promise improvements in that respect.

Some of the discussed shortcomings have already been addressed in the literature. First, the problem of *too many* sub-national units was early on encountered by researchers working on German elections where there are 299 districts. This is the problem that Selb and Munzert (2011) tackled - they want to estimate constituency-level estimates for 299 German *Wahlkreise*. Technically, nothing is wrong and one can estimate an MrP model. Realizations of the context-level random effect are 0 for those units that have no respondent in the sample. But this approach might be inefficient as it does not exploit all information we have. While we may not have any respondents from a specific unit, we might have respondents from neighboring units and could exploit that.

Another problem in such a situation is that we may not have socio-economic data broken down by unit as electoral districts need not to perfectly overlap with administrative boundaries which are usually the standard units of census data. Hence, it is not immediately clear how one would cary out the post-stratification step. The major contribution in Selb and Munzert (2011) is to formulate a model where the unit-level random effect is spatially autocorrelated – this then allows to exploit that we know which respondents live closer or further away from a specific district. They show how information can be added in such a model and also how this can be used within an MrP model (given that one does have administrative data on the electoral districts to be able to post-stratify). They replicate the empirical example of Park, Gelman, and Bafumi (2004*b*) which only has 50 units and show that by adding a spatially autocorrelated random effect the estimates do not improve by very much. With a well-specified MrP model there is not that much unexplained variation that is spatially correlated and can be absorbed. But that does just show that it is not always necessary to rely on their extension - nevertheless, when facing too many districts Selb and Munzert (2011) have shown a reliable and feasible procedure to produce estimates.

A second extensions relates to a researcher's ability to build a strong response model. While Buttice and Highton (2013) and Warshaw and Rodden (2012) advocate improving model quality by selecting better context-level variables there is also another complementary strategy. It is also possible to add additional individual-level variables that will greatly improve the estimates. Variables such as party affiliation and income might be powerful predictors but are not part of the census. If a variable is not part of the census, we cannot get the joint distribution and hence it is not possible to post-stratify by such additional variables. This is unfortunate since one usually relies on less strong models due to the data constraints following in the post-stratification step. This was the motivating problem that Leemann and Wasserfallen (2017) address.

The first part of their argument is that the need for the full joint distribution stems from the binary part of the response model. When generating predicted probabilities the impact of the realization of e.g. the gender random effect will vary depending on where an ideal type's score on the latent variable is. Simply put, if the response model was a linearly additive model (as e.g. OLS) there would be no problem. But since we rely on binary models, we are confined the using a joint distribution. The second part is that they show how one can generate a synthetic joint distribution whereby one assumes independence between gender/education.age and e.g. party ID (simple synthetic) or learns about the correlational structure from the survey data itself (adjusted synthetic). Following their approach allows to rely on powerful predictors on the individual-level as long one can find marginal distributions for those variables at the sub-national level. This leads to multilevel regression with synthetic post-stratification (MrsP). For the authors, MrsP is MrP's better half. They show in simulations and replications that MrsP can outperform MrP and reduce the estimation error by as much as MrP does over disaggregation.

3 The Use of MrP in the Literature

The chapter so far presented the underlying methodology of MrP by illustrating the method with a running example and discussed several methodological limitations and extensions of MrP. What has become clear out of this discussion is that MrP provides a more precise and sophisticated method for subnational unit estimation as compared to disaggregation or raking. In addition, the discussion has shown that MrP builds on a distinct methodological structure, which provides the ground for ongoing and future innovations, adaptations, as well as combinations of MrP with other methods. Accordingly, MrP is not simply the more precise successor of disaggregation or raking. Rather, the advances of MrP allow for the study of a new set of substantively interesting political science research questions.

Noteworthy is that the development of the method is not (exclusively) out-sourced to specialized method journals, but has also been advanced by substantive political science research. This combination of methodological development and substantive advances of various sub-field literatures is one of the most exciting features of the recent MrP literature. In the following, we discuss the state-of-the-art literature by focusing on three themes: testing and validation in different contexts, combinations MrP with other methods, and the broad substantive applications of MrP.

Multiple articles using MrP provide a rigorous methodological testing of the method. The previous section already discussed important methodological limitations and extensions of MrP in that respect. Given that MrP is still a relatively young method, the analyses of both the estimation precision and the scope conditions under which it performs well are critical for its establishment as "gold standard" (Selb and Munzert, 2011, 456). An important factor for the quick establishment of MrP is that several substantive articles using MrP test and validate the precision of MrP in various different settings and contexts, before they apply the method. Situated in the scholarship on American Politics, Enns and Koch (2013), for example, estimate state public opinion on the level of U.S. states and validate their application of MrP with specialized polls that were conducted for a subset of six states.

Going one level deeper, Warshaw and Rodden (2012) made a seminal contribution in popularizing MrP by showing how the method can be applied for the estimation of public opinion on

Congressional and State legislative districts in the United States. Among others, they validate the method by comparing MrP estimates with results of same-sex marriage referendums in Arizona, California, Michigan, Ohio, and Wisconsin. Also other cross-validation methods have been applied, such as randomly splitting data from the whole sample (using only part of the sample for predictions). However, the validation with voting data is particularly stringent, as it uses the real world as a benchmark (like the example illustrated in the previous section does). Leemann and Wasserfallen (2016) make a similar test of their Swiss MrP models by comparing cantonal MrP estimates with results of 186 direct democratic votes, whereas Tausanovitch and Warshaw (2013) use, as real-world comparison, 2008 presidential vote shares.

Besides showing that MrP outperforms other approaches, several of these studies also advance MrP in some respect, often by combining MrP with other methods. For example, Enns and Koch (2013) estimate state-level policy preferences from 1956 to 2010, extending the use of MrP to the analysis of long time series. Also, Pacheco (2011, 2012) estimates time-series data for her analysis of change in U.S. state public opinion, using imputation techniques to improve the estimates. A classic question of MrP studies is the study of congruence between public opinion and the views of their political representatives—and between public opinion and policy outcomes. The MrP literature has been careful in making sure that the opinions of voters and representatives is measured on the same scale. To that end, Kousser, Phillips, and Shor (2016) combine MRP with joint scaling, Tausanovitch and Warshaw (2013) scale survey items with IRT, and Leemann and Wasserfallen (2016) field their own survey, accounting for the joint scaling in the design of the questionnaire. The availability of precise time-series public opinion data and consistent estimates of elite and voter opinions on the same issues produced important empirical insights for the literatures on public opinion, representation, and democratic performance.

The previous section already discussed the pertinent challenges of the specification of the hierarchical response model as well as the methodological contributions by Selb and Munzert (2011) and Leemann and Wasserfallen (2017). More generally speaking, the literature on the limitations and extensions of MrP highlight that researchers ought to be familiar with the basics of MrP. They should develop an application of the method that is carefully tailored to their research design by taking into account all available data sources for their units of analysis. In that respect, also the work by Ghitza and Gelman (2013) makes an important contribution by reducing model-dependency with the analysis of deeply interacted subgroups, which are subsets of the population that are defined by multiple demographic and geographic characteristics.

This line of research is of interest for political scientists that are interested in very specific subgroups of the population. Also, for political actors that are seeking information on turnout and vote choice—preferably on a very specific level in terms of voter characteristics that allow for targeted campaigning efforts. Accordingly, it does not come as a surprise that MrP has become a standard tool in the field of campaigning and forecasting. A further powerful feature is that MrP can leverage large data sources that nowadays can be collected through online surveys. We just name a few examples here (out of a long list): Hanretty, Lauderdale, and Vivyan (2016*a*) present a forecast model for the 2015 British general election, and, among many others, CBS (in the U.S.), YouGov (in the UK), and LeeWas (in Switzerland) conduct regular surveys using MrP. A further, rather general, attractive feature of MrP is that the it can be applied to the analysis of large datasets collected online, including from samples with unequal participation (Wang et al., 2015; Downes et al., 2018).

In the digital age, large datasets that are not random samples of the target population become the norm. MrP has been successfully used in the analysis of such non-random samples, even for the case of Xbox players that were used for predictions of the 2012 presidential election in the US (Wang et al., 2015).[4] While MrP has received some attention for its ability to work seemingly well with non-probability polls, there is no secret magic entailed in the use of a regression model and a post-stratification procedure. When non-probability samples can yield valuable insights and when not will likely be a question that will attract much more research, particularly from polling companies that have an economic interests in further developing such approaches. For now, we refer to Ansolabehere and Rivers (2013) for a useful overview and discussion of the necessary assumptions.

Coming back to the use of MrP in the political science literature, MrP has been applied to a broad spectrum of substantive research. Among others, MrP has advanced literatures on partisanship, ideology, the responsiveness of institutions to public opinion, congruence of elite and voter preferences, and polarization. For example, Kastellec, Lax, and Phillips (2010) analyze

[4] See also the work of Hanretty, Lauderdale, and Vivyan (2016*a*,*b*) on UK YouGov samples for various elections and votes.

how public opinion matters for senate vote confirming supreme court nominees, McCarty et al. (2018) analyze how polarization differs between voters and legislators, while a series of studies investigates the congruence between public opinion, representatives, and policy outcomes — typically as a function of different institutional and electoral arrangements (Kousser, Phillips, and Shor, 2016; Lax and Phillips, 2012; Tausanovitch and Warshaw, 2013; Leemann and Wasserfallen, 2016). MrP has, for a still rather young method, an impressive track-record in providing substantive empirical insights to various (sub-)disciplines of political science. This may be the greatest success (and promise for the future) of MrP.

Beyond political science, MrP has also been used, for example, in sociology. Claassen and Traunmüller (2017) use MrP as a tool for the analysis of hard to measure populations (not as a method for estimating public opinion). More specifically, they estimate the demographic structures of Hindus, Moslems, and Jews in Great Britain using survey data. Since there is good census data for these religious minorities in Great Britain, they can validate the accuracy of their estimates. Thus, methodological advances and extensions of MrP (that are motivated by substantive research) are not restricted to the political science literature.

4 Conclusion

MrP has in a rather short time established itself as the standard method for the estimation of public opinion for sub-national units (and other subgroups of populations and samples). MrP provides a distinct and flexible framework in three steps with the specification of the hierarchical response model, the prediction for ideal types, and the post-stratification step. This general methodological structure allows for tailored applications, combinations with other methods, and innovative extensions. In this chapter, we presented all steps of MrP with a running example, before discussing in more detail technical issues, methodological limitations, and extensions.

We have also discussed multiple advances of MrP and expect much more to come. As discussed, the selection of optimal context-level variables for the response model is critical, yet not straightforward. With the increasing use of machine learning approaches in political science (see e.g. Montgomery and Olivella, 2018), there are promising ideas and projects that are likely to provide further useful solutions for the challenge of selecting the response model in

MrP applications. Machine learning is an umbrella term for a lot of methodologies that improve model-specification and prediction accuracy with a disciplined approach. Several working papers combine classifiers with post-stratification and aim to bring the promise of statistical learning to MrP (Goplerud et al., 2018; Ornstein, 2017; Broniecki, Leemann, and Wüest, 2018). The combination of MrP with other methods has already shown to be very productive and semi-automated procedures will likely help further improve MrP for certain applications.

However, most important for the continuing success story of MrP will be that methodological innovations are not and end in itself. Rather, future advances of MrP have to provide solutions for technical problems that are motivated by substantive research puzzles and are powerful to solve them. To the extent to which this will continue to shape the development and applications of MrP, the importance of MrP for the political science literature will continue to grow. As the discussion of the broad substantive applications of MrP in the literature has shown, the track-record of MrP in that respect is already quite impressive.

References

Ansolabehere, Stephen, and Douglas Rivers. 2013. "Cooperative survey research." *Annual Review of Political Science* 16: 307–329.

Broniecki, Philipp, Lucas Leemann, and Reto Wüest. 2018. "Improved Multilevel Regression with Post-Stratification Through Optimal Model Selection." *Working Paper* .

Buttice, Matthew K, and Benjamin Highton. 2013. "How does multilevel regression and post-stratification perform with conventional national surveys?" *Political Analysis* 21(4): 449–467.

Claassen, Christopher, and Richard Traunmüller. 2017. "Improving and Validating Survey Estimates of Religious Demography Using Bayesian Multilevel Models and Poststratification." *Sociological Methods & Research* p. 0049124118769086.

Downes, Marnie, Lyle C Gurrin, Dallas R English, Jane Pirkis, Dianne Currier, Matthew J Spittal, and John B Carlin. 2018. "Multilevel Regression and Poststratification: A Modelling Approach to Estimating Population Quantities From Highly Selected Survey Samples." *American journal of epidemiology* p. kwy070.

Enns, Peter K, and Julianna Koch. 2013. "Public Opinion in the US States: 1956 to 2010." *State Politics & Policy Quarterly* 13(3): 349–372.

Erikson, Robert S., Gerald C. Wright, and John P. McIver. 1993. *Statehouse Democracy: Public Opinion and Policy in the American States.* Cambridge: Cambridge University Press.

Gelman, Andrew, and Jennifer Hill. 2007*a*. *Data Analysis Using Regression and Multilevel/Hierarchical Models.* Cambridge: Cambridge University Press.

Gelman, Andrew, and Jennifer Hill. 2007*b*. *Data Analysis Using Regression and Multilevel/Hierarchical Models.* Cambridge: Cambridge University Press.

Gelman, Andrew, and Thomas C. Little. 1997. "Poststratification into many categories using hierarchical logistic regression." *Survey Research* 23: 127–135.

Ghitza, Yair, and Andrew Gelman. 2013. "Deep interactions with MRP: Election turnout and voting patterns among small electoral subgroups." *American Journal of Political Science* 57(3): 762–776.

Goplerud, Max, Shiro Kuriwaki, Marc Ratkovic, and Dustin Tingley. 2018. "Sparse Multilevel Regression (and Poststratification (sMRP))." *Working Paper* Harvard University.

Hanretty, Chris, Ben Lauderdale, and Nick Vivyan. 2016*a*. "Combining national and constituency polling for forecasting." *Electoral Studies* 41: 239–243.

Hanretty, Chris, Benjamin E Lauderdale, and Nick Vivyan. 2016*b*. "Comparing strategies for estimating constituency opinion from national survey samples." *Political Science Research and Methods* pp. 1–21.

Herron, Michael C. 1999. "Postestimation uncertainty in limited dependent variable models." *Political Analysis* 8(1): 83–98.

James, Gareth, Daniela Witten, Trevor Hastie, and Robert Tibshirani. 2013. *An introduction to statistical learning.* Vol. 112 Springer.

Kastellec, Jonathan P., Jeffrey R. Lax, and Justin H. Phillips. 2010. "Public opinion and Senate confirmation of Supreme Court nominees." *Journal of Politics* 72(3): 767–784.

Kousser, Thad, Justin Phillips, and Boris Shor. 2016. "Reform and representation: a new method applied to recent electoral changes." *Political Science Research and Methods* pp. 1–19.

Lax, Jeffrey R., and Justin H. Phillips. 2009*a*. "Gay Rights in the States: Public Opinion and Policy Responsiveness." *American Political Science Review* 103(3): 367–386.

Lax, Jeffrey R., and Justin H. Phillips. 2009*b*. "How Should We Estimate Public Opinion in the States?" *American Journal of Political Science* 53(1): 107–121.

Lax, Jeffrey R., and Justin H. Phillips. 2012. "The Democratic Deficit in States." *American Journal of Political Science* 56(1): 148–166.

Leemann, Lucas, and Fabio Wasserfallen. 2016. "The Democratic Effect of Direct Democracy." *American Political Science Review* 110(4): 750–762.

Leemann, Lucas, and Fabio Wasserfallen. 2017. "Extending the Use and Prediction Precision of Subnational Public Opinion Estimation." *American Journal of Political Science* 61(4): 1003–1022.

McCarty, Nolan, Jonathan Rodden, Boris Shor, Chris Tausanovitch, and Christopher Warshaw. 2018. "Geography, uncertainty, and polarization." *Political Science Research and Methods* pp. 1–20.

Miller, Warren E., and Donald W. Stokes. 1963. "Constituency Influence in Congress." *American Political Science Review* 57: 45–46.

Montgomery, Jacob M, and Santiago Olivella. 2018. "Tree-Based Models for Political Science Data." *American Journal of Political Science* 62(3): 729–744.

Ornstein, Joseph T. 2017. "Machine Learning and Poststratification.".

Pacheco, Julianna. 2011. "Using National Surveys to Measure Dynamic U.S. State Public Opinion: A Guideline for Scholars and an Application." *State Politics & Policy Quarterly* 11(4): 415–439.

Pacheco, Julianna. 2012. "The Social Contagion Model: Exploring the Role of Public Opinion on the Diffusion of Antismoking Legislation across the American States." *Journal of Politics* 74(1): 187–202.

Park, David K., Andrew Gelman, and Joseph Bafumi. 2004*a*. "Bayesian Multilevel Estimation with Poststratification: State-Level Estimates from National Polls." *Political Analysis* 12: 375–385.

Park, David K, Andrew Gelman, and Joseph Bafumi. 2004*b*. "Bayesian multilevel estimation with poststratification: state-level estimates from national polls." *Political Analysis* 12(4): 375–385.

Selb, Peter, and Simon Munzert. 2011. "Estimating constituency preferences from sparse survey data using auxiliary geographic information." *Political Analysis* 19(4): 455–470.

Tausanovitch, Chris, and Christopher Warshaw. 2013. "Measuring constituent policy preferences in congress, state legislatures, and cities." *The Journal of Politics* 75(2): 330–342.

Wang, Wei, David Rothschild, Sharad Goel, and Andrew Gelman. 2015. "Forecasting elections with non-representative polls." *International Journal of Forecasting* 31(3): 980–991.

Warshaw, Christopher, and Jontahan Rodden. 2012. "How Should We Measure District-Level Public Opinion on Individual Issues?" *Journal of Politics* 74(1): 203–219.

Zhang, Li-Chun. 2000. "Post-stratification and calibration—a synthesis." *The American Statistician* 54(3): 178–184.