AJPS AMERICAN JOURNAL of POLITICAL SCIENCE

Extending the Use and Prediction Precision of Subnational Public Opinion Estimation

Lucas Leemann University of Essex
Fabio Wasserfallen University of Salzburg, University of Zurich

Abstract: *The comparative study of subnational units is on the rise. Multilevel regression and poststratification (MrP) has become the standard method for estimating subnational public opinion. Unfortunately, MrP comes with stringent data demands. As a consequence, scholars cannot apply MrP in countries without detailed census data, and when such data are available, the modeling is restricted to a few variables. This article introduces multilevel regression with synthetic poststratification (MrsP), which relaxes the data requirement of MrP to marginal distributions, substantially increases the prediction precision of the method, and extends its use to countries without census data. The findings of Monte Carlo, U.S., and Swiss analyses show that, using the same predictors, MrsP usually performs in standard applications as well as the currently used standard approach, and it is superior when additional predictors are modeled. The better performance and the more straightforward implementation promise that MrsP will further stimulate subnational research.*

Replication Materials: The data, code, and any additional materials required to replicate all analyses in this article are available on the *American Journal of Political Science* Dataverse within the Harvard Dataverse Network, at: https://doi.org/10.7910/DVN/I0VEMG.

The comparative study of subnational units has attracted growing interest in the literature. There are a number of reasons for this: Subnational units are potentially better suited for comparative analysis than countries because they are less heterogeneous, more accurate data are available, country-specific factors are constant, and controlled comparisons allow for the development of interesting identification strategies for causal inference (e.g., Snyder 2001; Tausanovitch and Warshaw 2014; Ziblatt 2008). A critical but challenging element of subnational comparative research is the estimation of public opinion. The recent introduction of multilevel regression with poststratification (MrP) generates reliable public opinion estimates for subnational units. Unfortunately, the use of MrP, as currently applied in the literature, comes with stringent data requirements. We build on the recent methodological advances and develop a new approach that extends the use of MrP beyond a few developed countries and that increases the prediction precision of the method substantially.

Early attempts in the estimation of public opinion disaggregated national surveys into subnational subsamples (Miller and Stokes 1963). One solution for overcoming the small-n problem of the disaggregation approach was to combine multiple surveys with the same questions into one mega-poll (Erikson, Wright, and McIver 1993). Gelman and Little (1997), then, laid the methodological groundwork for MrP, which can be applied by using standard national survey data and is much more precise than disaggregation. More recent articles provide interesting empirical analyses of substantive questions, including further testing and revision of MrP (Lax and Phillips 2009b; Warshaw and Rodden 2012). The quick spread of MrP is quite remarkable—not least when we consider that the method has, to the best of our knowledge, mostly been applied so far in a few countries, such as the United States, the United Kingdom, Germany, and Switzerland.

The narrow spatial scope is mainly because of the stringent data requirement of the current standard

Lucas Leemann is Reader, Department of Government, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, United Kingdom (l.leemann@essex.ac.uk). Fabio Wasserfallen is Assistant Professor, Centre of European Union Studies, University of Salzburg, Mönchsberg 2, 5020 Salzburg, Austria (fabio.wasserfallen@sbg.ac) & Guest Professor, Department of Political Science, University of Zurich, Switzerland.

Many thanks to Noah Buckley, Andrew Gelman, Andy Guess, Jonathan Kastellec, Romain Lachat, Jeff Lax, Tiffany Washburn, and Piero Stanig for helpful comments. We also thank Christopher Warshaw and Jonathan Rodden for sharing their data with us and Werner Seitz and his team at the Swiss Federal Statistics Office for the supply of various data sets. Patrice Siegrist provided excellent research assistance, and the financial support from the Swiss National Science Foundation is gratefully acknowledged (grants #: 100017-1377651 and P1SKP1-148357).

American Journal of Political Science, Vol. 00, No. 0, xxxx 2017, Pp. 1–20

 DOI: 10.1111/ajps.12319

application of the method. The precondition for using MrP is that detailed census data in the form of joint distributions are available for poststratification. Researchers need to know, for example, how many 18–35-year-old women with a university degree live in each subnational unit. This data requirement makes it impossible to apply MrP in countries where such data are not available—whether this is because of data protection laws (e.g., in India) or because the data are not gathered by a single agency (e.g., in Afghanistan). Furthermore, if census data are available, researchers can only use three or four demographic variables that are provided in the restrictive format of joint distributions as individual-level predictors of political preferences. Strong predictors such as party identification and income cannot be modeled.

We develop an alternative application of MrP, which we call multilevel regression with *synthetic* poststratification (MrsP). MrsP relies on marginal distributions. For applying MrsP, researchers only need to know, for example, the shares (marginals) of women, of university graduates, and of 18–35-year-old citizens in each subnational unit. Due to this relaxed data requirement, MrsP can be applied in countries where MrP has not been possible so far (e.g., in India and Afghanistan). A further important advantage of MrsP is that it increases the prediction precision of subnational public opinion estimation for countries with a census (e.g., the United States and Switzerland). We analyze improvements in prediction with U.S. and Swiss data by adding strong predictors—beyond the standard demographic variables—such as income and party identification. The gains in prediction precision are very substantial. The findings of the U.S. application show that MrsP outperforms the standard version of MrP with a 43% reduction of prediction error (measured with the mean squared error), which is even larger than MrP's improvement upon disaggregation (the error decreases by 35%). The Swiss analyses provide even stronger results in favor of MrsP.

This article presents two versions of MrsP: a simpler one that is straightforwardly implemented and a more elaborate one that takes full advantage of the information in the survey data. Although the findings suggest that elaborate MrsP is, generally speaking, the method of choice, researchers can, given some conditions, also apply the simpler version. We provide, based on the presented analyses, premodeling guidance to help researchers specify MrP models. In sum, we believe that MrsP, the approach developed in this article, will further stimulate subnational comparative research relying on public opinion estimates, as it goes beyond the data limitations of the current standard application, extends the use of MrP, and substantially improves the prediction precision of the still young method.

"Classic MrP" and Its Limits

The application of MrP has its origins in the Gelman and Little (1997) study, which combined hierarchical modeling and poststratification. Park, Gelman, and Bafumi (2004) subsequently introduced the method to political science with a remarkable impact on the discipline, as the substantial number of recent studies using the approach demonstrates (Kastellec, Lax, and Phillips 2010; Lax and Phillips 2009a, 2009b, 2012; Pacheco 2012; Tausanovitch and Warshaw 2014; Warshaw and Rodden 2012). In recent years, MrP has been established as the state-of-the-art method for comparative subnational research studying public opinion. Accordingly, Selb and Munzert (2011, 456) conclude that MrP is the "gold standard" in estimating political preferences on the subnational level.

MrP estimates public opinion on the subnational level in four steps. The first is to conduct a survey that identifies personal characteristics and asks a number of political preference questions; second, a hierarchical model is fitted to the data to make predictions for specific voter types (e.g., 18–35-year-old women with a university degree); in the third step, predictions for all predefined voter types are calculated using the estimates of the hierarchical model; and, finally, researchers calculate, based on fine-grained census data, public opinion estimates in the subnational units by weighting the predictions of each voter type according to the number of voters living in the subnational units with the same characteristics.

Let us further illustrate the data requirements and the method. Researchers start with the collection of national survey data of usually somewhere between 500 and 1,500 respondents and model the responses on the support for a specific policy. For illustrative purposes, we explore an MrP response model with two individual-level variables—gender (men/women) and education (no high school, high school, college, postgraduate)—and with random effects for the subnational unit (α_c) and the region (α_r). In addition, we include a number of contextual factors explaining variation between the subnational units ($\mathbf{X}_c$). Accordingly, we write the following hierarchical probit, as a standard response model, to estimate the support for a specific policy:

$$Pr(y_i=1)=\Phi\left(\beta_0+\alpha_{j[i]}^{gender}+\alpha_{m[i]}^{education}+\alpha_c^{subnational\ unit}\right)$$

$$\alpha_j^{gender}\sim N\left(0,\sigma_{gender}^2\right),\quad \text{for } j=1,\ldots,J$$

$$\alpha_m^{education}\sim N\left(0,\sigma_{education}^2\right),\quad \text{for } m=1,\ldots,M$$

Table 1 Census Data Requirement Example of Classic MrP

Gender \ Education	No High School	High School	College	Postgraduate	Total
Men	$\mathbf{N_{11}}$	$\mathbf{N_{12}}$	$\mathbf{N_{13}}$	$\mathbf{N_{14}}$	$N_{1\cdot}$
Women	$\mathbf{N_{21}}$	$\mathbf{N_{22}}$	$\mathbf{N_{23}}$	$\mathbf{N_{24}}$	$N_{2\cdot}$
Total	$N_{\cdot 1}$	$N_{\cdot 2}$	$N_{\cdot 3}$	$N_{\cdot 4}$	N

$$\alpha_c^{subnational\ unit} \sim N\left(\alpha_{r[c]}^{region} + \boldsymbol{\beta}\mathbf{X}_c, \sigma^2_{subnational\ unit}\right), \quad \text{for } c = 1, \ldots, C$$

$$\alpha_r^{region} \sim N\left(0, \sigma^2_{region}\right), \quad \text{for } r = 1, \ldots, R$$

One of the distinguishing features of MrP is the partial pooling and the shrinkage to the mean induced by modeling random effects (Steenbergen and Jones 2002). This partly accounts for the good predictive performance of MrP. In this example, there are eight voter types in each subnational unit (four education categories and gender: $J \times M = 8$). The response model estimates are used to calculate predictions $\hat{\pi}_{cjm}$ for all possible combinations of j and m in each subnational unit c. For poststratification, the researcher needs to know the joint distributions in each subnational unit, that is, the frequency of each voter type ($N_{11}, N_{12}, \ldots, N_{24}$). We call MrP that relies on joint distributions "classic MrP," as it is the current standard method in the literature. Census officials provide, if available, the information on the joint distributions. Table 1 illustrates the data requirement of classic MrP.

Finally, each prediction is weighted by the joint distribution data and the total sum divided by the number of all residents:

$$\hat{\pi}_c = \frac{\sum_j \sum_m \hat{\pi}_{jm\in c} N_{jm\in c}}{N_{n\in c}} = \frac{\sum_j \sum_m \Phi(\hat{\beta}_0 + \hat{\alpha}_m + \hat{\alpha}_j + \hat{\alpha}_c) N_{jm\in c}}{N_{n\in c}}.$$

In this example, we rely on eight voter types. Real-world applications include more individual-level random effects. Lax and Phillips (2009b), for example, work with gender, three races, and four education and age categories (96 types) and thus need the exact number of 18–29-year-old black women with a high school degree in each state, among other data. Such fine-grained information is only available if a detailed census has been carried out. If joint distribution data are available, the specification of the response model is predetermined by the data of the census bureau (not the modeling decisions of researchers) and usually restricted to standard demographic variables. The demanding data requirements of classic MrP are problematic for the following reasons:

- In developing countries, fine-grained census data are not available. In Afghanistan, for example, no census is carried out, whereas in the more developed India, census data on the village level are not available as joint distributions.[1]
- The joint distributions of the census in developed countries do not include variables that are potentially important predictors of political preferences. Party identification, for example, is available in neither the Swiss nor the U.S. census as joint distributions.

Before political scientists started working with MrP, the standard method for deriving preference measures on the subnational level was the disaggregation of national surveys into subnational samples. While this method is free from data availability restrictions, the estimates for small constituencies are unreliable, as they stem from very few observations (Levendusky, Pope, and Jackman 2008). Several studies have shown that MrP estimates are more precise than disaggregation estimates (e.g., Lax and Phillips 2009b).

However, even for countries with good data availability, such as Germany, researchers continue to develop alternative approaches to MrP because of data availability restrictions. Selb and Munzert (2011), for example, refer to MrP as the "gold standard," stating that it cannot be applied in their analysis of German electoral districts because data on joint distributions are not available. They develop, as an alternative, a sophisticated Bayesian hierarchical estimation strategy that exploits auxiliary geographic information. However, Selb and Munzert (2011) argue that their approach is useful if MrP cannot be applied, advising that scholars should exploit information on the constituency populations with MrP, if

[1]These are just two illustrative examples. In principle, MrsP can be applied in all countries, where data on marginal distributions are available.

possible.[2] The contribution of this analysis is exactly that researchers can apply MrP in data-sparse cases, as the approach we develop relies on marginal distributions. Rather than giving up on individual-level predictors, we show how they can be used fruitfully. Thus, we build upon the strength of MrP and make the method more widely applicable.

MrP with Synthetic Joint Distributions (MrsP)

Some of the most sophisticated MrP contributions are also constrained by the data requirements discussed above. Warshaw and Rodden (2012, 208), for example, study district-level public opinion in the United States without using age as a predictor in their response model because the "census factfinder does not include age breakdowns for each race/gender/education subgroup." This example highlights the data availability problem of classic MrP: Whereas U.S. district data on the age structure are available as marginal distributions, data on the exact number of elderly people with a given gender, age, and education are not. More generally, marginal distributions are available for many interesting variables in most countries. In the Afghan case, for example, the Asia Foundation collects data on the ethnic and linguistic structures of subnational populations.

The key deviation of MrsP, the approach developed in this article, is that it relies on synthetic joint distributions that are created with data on the marginal distributions (whereas classic MrP relies on the "true" joint distributions). Our point of departure is that researchers can collect data on the population structures of subnational units as marginal distributions for various variables that are potentially important predictors of political preferences. Instead of relying exclusively on the demographic variables of the census, MrsP allows the modeling of any political, social, economic, or demographic variable asked in the survey. Researchers only need the marginal distributions of these variables in the subnational units, which are more widely available.

Before we further elaborate on MrsP, we would like to discuss how our approach relates to raking, a standard survey research procedure applied for the analysis of nonrepresentative samples. In a nutshell, raking assigns weights based on the marginal distribution of one variable, then calculates—conditional on the derived weights—new weights with the marginal distribution of the second variable, and continues with this iterative proportional fitting until the weighted survey best approximates the distribution of the target population (Deming and Stephan 1940; Fienberg 1970). Appendix E discusses in detail how raking can be integrated in the MrP framework. Like MrsP, raking allows us to model variables for which we only know the marginal distribution. However, both for the prediction of public support and the estimation of uncertainty, MrsP exploits more information from the data compared to the raking approach. Accordingly, it is not surprising that the MrsP predictions in the example presented in Appendix E are more precise. Thus, although the use of raking provides a feasible alternative, we recommend MrsP.

Instead of relying on true joint distributions, which are often unavailable, MrsP calculates synthetic joint distributions with data on the marginal distributions. The synthetic joints can be computed in different ways. Kastellec et al. (2016), for example, extract the data from multiple surveys. For the more standard situation, where just one survey is available, we propose two estimation techniques: a simple and a more elaborate one. In the simple case, the synthetic joint is estimated as the product of the marginal distributions. The accordingly estimated synthetic joint, which we call the *simple* synthetic joint, will be correct when the used variables are independent of one another. If they are not—which is in all likelihood the case—the simple synthetic joint deviates from the true joint distribution. However, even in that case, the example below illustrates that the prediction deviation between MrP and MrsP is only because of the nonconstant marginal effects of the probit link function in the response model. In the following sections, we investigate Monte Carlo simulations and real-world data analyses. Based on the presented findings, we conclude that there are most likely no differences in terms of prediction precision in applied work.

We later propose a more elaborate technique for estimating the synthetic joint distributions that does not make the unrealistic assumption of noncorrelation among individual-level variables, but it is a little more complicated to compute. In the analysis of the real-world application, we will explain the estimation of what we call *adjusted* synthetic joint distributions, which basically take into account information on the correlation among individual-level variables that is available from the survey data. This more elaborate version of MrsP allows for a more nuanced estimation of uncertainty and provides, under some conditions, better point predictions. Based on the presented findings, we will discuss when researchers can use the simple MrsP version and when we would recommend the more elaborate MrsP approach.

[2] Selb and Munzert (2011) recommend applying their method when the number of subnational units under study is exceptionally high compared to the available survey data.

TABLE 2 Example of True and Simple Synthetic Joint Distributions for the Most Extreme Case

v1 \ v2	i=1	i=2	
j=1	60%	0%	60%
j=2	0%	40%	40%
	60%	40%	100%

(a) True Joint Distribution

v1 \ v2	i=1	i=2	
j=1	36%	24%	60%
j=2	24%	16%	40%
	60%	40%	100%

(b) Simple Synthetic Joint Distribution

v1 \ v2	i=1	i=2
j=1	7%	50%
j=2	69%	98%

(c) Predicted Support

Note: An example is shown of a simple synthetic joint distribution for the most extreme case where the two individual-level variables are fully dependent. The predicted support in Table 2(c) is based on a model that includes two random effects (α^{v1} and α^{v2}) with the following estimates: $\hat{\alpha}_1^{v1} = -1$, $\hat{\alpha}_2^{v1} = +1$, $\hat{\alpha}_1^{v2} = -0.5$, and $\hat{\alpha}_2^{v2} = +1$.

But let us first come back to the comparison between classic MrP and MrsP. The following theoretical discussion illustrates the difference between classic MrP and simple MrsP by emphasizing the scenario under which the two estimation procedures are most distinct. Table 4 shows an example of two binary individual-level variables, `v1` and `v2`, that are completely separated and thus totally dependent. The simple MrsP synthetic joint distribution, estimated as the product of the marginal distributions (60% and 40%), deviates quite strongly from the true joint distribution in this most extreme case (compare Table 2a and 2b).

What matters for applied scholars is how much the predictions differ with the use of the simple synthetic joint distribution (MrsP) compared to the true joint distribution (classic MrP). To estimate the difference in predictions, we assume a response model, which is identical in both procedures, with two random effects on the individual level (age and education) and an estimated constant with the value 0. We discuss below how changing the constant affects the findings. As estimates for the random effects, we pick fairly large numbers (from -1 to $+1$) to magnify the difference between classic MrP and MrsP. In applied work, random effects are clearly smaller (see Appendix C). The predicted probability for an individual of a specific cell is estimated using the random effects (e.g., for an individual from the upper left cell: $\hat{p}_{11} = \Phi(\hat{\alpha}_1^{v1} + \hat{\alpha}_1^{v2})$). Table 2(c) reports the predicted support for each cell based on the response model estimates. The support in the subnational unit is estimated by weighting the predictions for each type by its frequency in the population. With the true joint distribution (classic MrP, see Table 2a), the support in the subnational unit, $\hat{p}_{true}$, is estimated as follows:

$$\begin{aligned}\hat{p}_{true} &= 0.6 \cdot \Phi(\hat{\alpha}_1^{v1} + \hat{\alpha}_1^{v2}) + 0.4 \cdot \Phi(\hat{\alpha}_2^{v1} + \hat{\alpha}_2^{v2});\\ &= 0.6 \cdot \underbrace{\Phi(-1.5)}_{0.07} + 0.4 \cdot \underbrace{\Phi(+2)}_{0.98} = 0.431.\end{aligned}$$

Using the simple synthetic joint distribution (MrsP, see Table 2b), the support in the subnational unit, $\hat{p}_{syn}$, is estimated as follows:

$$\begin{aligned}\hat{p}_{syn} = &\underbrace{0.24 \cdot \Phi(\hat{\alpha}_1^{v1} + \hat{\alpha}_2^{v2})}_{\text{should be 0\% of pop}} + \underbrace{0.36 \cdot \Phi(\hat{\alpha}_1^{v1} + \hat{\alpha}_1^{v2})}_{\text{should be 60\% of pop}}\\ &+ \underbrace{0.24 \cdot \Phi(\hat{\alpha}_2^{v1} + \hat{\alpha}_1^{v2})}_{\text{should be 0\% of pop}} + \underbrace{0.16 \cdot \Phi(\hat{\alpha}_2^{v1} + \hat{\alpha}_2^{v1})}_{\text{should be 40\% of pop}};\\ = &\ 0.24 \cdot \underbrace{\Phi(0)}_{0.5} + 0.36 \cdot \underbrace{\Phi(-1.5)}_{0.07} + 0.24 \cdot \underbrace{\Phi(+0.5)}_{0.69}\\ &+ 0.16 \cdot \underbrace{\Phi(+2)}_{0.98};\\ = &\ 0.466.\end{aligned}$$

In this most extreme example, the deviation in predicted support is only 3.5%. The prediction deviation is surprisingly small considering that we choose variables that are perfectly correlated and high values for the estimated random effects. To understand the source of the deviation, it is important to recall that the simple synthetic joint distribution is computed using the correct marginal distributions: It accounts for 60% of the population with the characteristic $\texttt{v1} = 1$, for 60% with $\texttt{v2} = 1$, for 40% with $\texttt{v1} = 2$, and for 40% with $\texttt{v2} = 2$; only the joint distribution values are wrong. However, the prediction deviation of 3.5% is not so much a product of the wrong synthetic joint distribution values but rather a result of the nonconstant marginal effects of the probit link function in the response model. In a probit model, adding

α_1^{v1} to a hypothetical person with $\texttt{v2} = 0$ has a different marginal effect than it has on a person with $\texttt{v2} = 1$.

Let us illustrate that point by looking at how the support in the subnational unit is estimated in both procedures. In this case, all individuals of the first row of the matrix ($\texttt{v1} = 1$) also have the characteristic $\texttt{v2} = 1$ (see Table 2a). For example, all women are also university graduates (and no man has a university degree). Accordingly, women's predicted support of 7% for the policy is weighted by 0.6 (see the first part of the equation using the true joint distribution). In the estimation with the synthetic joint distribution, women's predicted support is overestimated, as the 7% support is only weighted by 0.36 (see the second part of the equation using the synthetic joint distribution), whereas the additional 0.24 of women (with the characteristic $\texttt{v1} = 1$, i.e., no university degree) is multiplied with a higher predicted probability of 50% (see the first part of the equation using the synthetic joint distribution). As there are no women with no university degree in that illustrative example, women's predicted support for the policy is overestimated.

For men ($\texttt{v1} = 2$), the equation using the true joint distribution weights the very high 98% probability of men's support for the policy by 0.40. The prediction using the synthetic joint distribution, however, underestimates men's support for the policy, as it weighs the 98% support only for 0.16 and adds 0.24 with 69% support. Overall, MrsP overestimates women's and underestimates men's support for the policy in this example. In a linear model, the over- and the underestimation in predicted support between men and women would cancel each other out—but not in a probit model: The nonconstant marginal effects in a probit model explain the small prediction deviation between classic MrP and MrsP of 3.5%.[3] Based on this analysis, we derive two important implications. First, the deviation in predictions between classic MrP and MrsP is only because of the nonlinear link function of the probit model. Second, the nonconstant marginal effects of the probit model only cause deviations in predictions to the extent to which the individual-level variables are correlated.

One aspect of the above example is tilted in favor of MrsP: The overall public support is close to 50%. In this area, a linear function provides the best approximation of the logit or probit model, and the first derivative is close to constant. In the subsequent Monte Carlo analyses, we will vary that constant to mimic situations where the average public support ranges from 50% to the highly lopsided 80%. Table 3 summarizes the distinction between classic MrP and MrsP. The most important advantage of MrsP is that it allows the modeling of individual-level variables, of which only the marginal distributions are available, whereas having data on the true joint distribution is a conditio sine qua non for classic MrP. MrsP thus extends the use of MrP to countries without a census and makes the application of MrP more flexible for countries where census data are available.

TABLE 3 Data Requirement and Model Flexibility of Classic MrP and MrsP

	MrsP	Classic MrP
True joint distribution needed	×	√
Marginal distributions are sufficient	√	×
Flexible modeling of response model	√	×

The potential downside of simple MrsP is that the correlation between individual-level variables induces a small deviation in prediction because of the nonconstant marginal effects of the probit model.[4] The discussed illustrative example explored the most extreme (and unrealistic) case with a perfect correlation between the two variables. If they were totally independent—which is also unrealistic—classic MrP and simple MrsP would provide the same results. In general, the deviation between the predictions becomes larger as the correlation increases.[5] The deviation in prediction between the two MrP approaches also increases with higher variances of the random effects. The next sections further analyze whether, and under which circumstances, MrsP performs as well as MrP.

Simple MrsP and Classic MrP with the Same Data

Should we expect that the predictions differ between MrP and MrsP in applied work? To answer that question, we execute Monte Carlo analyses with three manipulated parameters: First, we change the sample sizes from a small

[3]Note that in the linear case, the two equations predicting public opinion are identical. We have explored MrP with a linear model, as in most situations a linear probability model and a binary model will produce similar results (Angrist and Pischke 2008; Beck 2011). However, MrP with a linear model does not perform optimally. The main problem is that it tends to produce estimates that are less than 0 or greater than 1. This is one of the criticisms of using linear models for binary outcome variables (Maddala 1983, 16).

[4]A second limitation is that deep interactions can only be modeled if the two constituting variables are available in the census data (Ghitza and Gelman 2013).

[5]See Appendix A for correlations in Swiss and U.S. data.

sample of 500 respondents to a sample of 1,000 respondents and a large sample of 2,000 respondents. Second, we use four different correlations (ρ) among the individual-level variables (0, 0.2, 0.4, and 0.6). Third, we vary the degree of public support. In the case of no correlation, simple MrsP and classic MrP are equivalent (with perfect independence, the product of the marginals equals the true joint distribution). Correlations of 0.2 and partly also of 0.4 are realistic for standard demographic values, whereas 0.6 is exceptionally high (see Appendix A). As discussed in the previous section, the prediction deviation between the two procedures should increase as the correlation becomes larger.

Our data-generating process (DGP) assumes three variables on the individual level and a variable on the subnational level. We analyze 25 subnational units; 10 of them are large (each covering 7% of the total population) and 15 are small (each covering 3% of the total population). The following equations describe the DGP:

$$
\begin{aligned}
y_i^* &= \beta_0 + \gamma_{1[i]} + \gamma_{2[i]} + \gamma_{3[i]} + \alpha_c^{subnational\ unit} \\
&\quad + \varepsilon_{ic}. \\
Pr(y_i = 1) &= \Phi(y_i^*),\ \varepsilon_{ic} \sim N(0, 1). \\
\bar{\gamma}_{1i} &\sim N(0, 1)\ \&\ \bar{\gamma}_{2i} \sim N(0, 1)\ \&\ \bar{\gamma}_{3i} \sim N(0, 1). \\
\alpha_c^{subnational\ unit} &= \delta_{c1} + \delta_{c2} + \varepsilon_c, \varepsilon_c \sim N(0, 1) \\
&\quad \text{for } c = 1, \ldots, 25. \\
\delta_{c1} &\in \{0, 0.4, 0.85\},\ \delta_{c2} \sim U(-.2, .2).
\end{aligned}
$$

The accordingly simulated data allow us to analyze how the simple version of MrsP performs as public support becomes more lopsided: If $\delta_{c1} = 0$, the average region's public support is 50%; if $\delta_{c1} = 0.4$, the average support is 65%; and if $\delta_{c1} = 0.85$, the average support is 80%. The subnational level variable (δ_{c2}) is evenly spread between ± 0.2. The individual-level variables ($\gamma_1, \gamma_2, \gamma_3$) are based on draws from a multivariate normal distribution and transformed to discrete variables with four categories each:

$$
\gamma_{ki} = \begin{cases} 1 & \text{if } \bar{\gamma}_{ki} < -1, \\ 2 & \text{if } -1 < \bar{\gamma}_{ki} < 0, \\ 3 & \text{if } 0 < \bar{\gamma}_{ki} < 1, \\ 4 & \text{if } \bar{\gamma}_{ki} > 1, \end{cases} \quad \text{and}
$$

$$
\text{Var}(\bar{\boldsymbol{\gamma}}) = \begin{pmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{pmatrix} \quad \text{for } k = 1, 2, 3.
$$

The selected random effects (γ_{ki}) are quite large in size, as they are based on the following normal distributions: ($\gamma_k \sim N(0, 1), \forall\ k$). In real-world examples, the random effects are smaller (the conservative setup of the analysis tends to overestimate the prediction deviation between classic MrP and MrsP).[6] For the Monte Carlo analyses, we create a "true" population of 1 million citizens, draw for every simulation a new sample, and estimate disaggregation, classic MrP, and MrsP predictions.

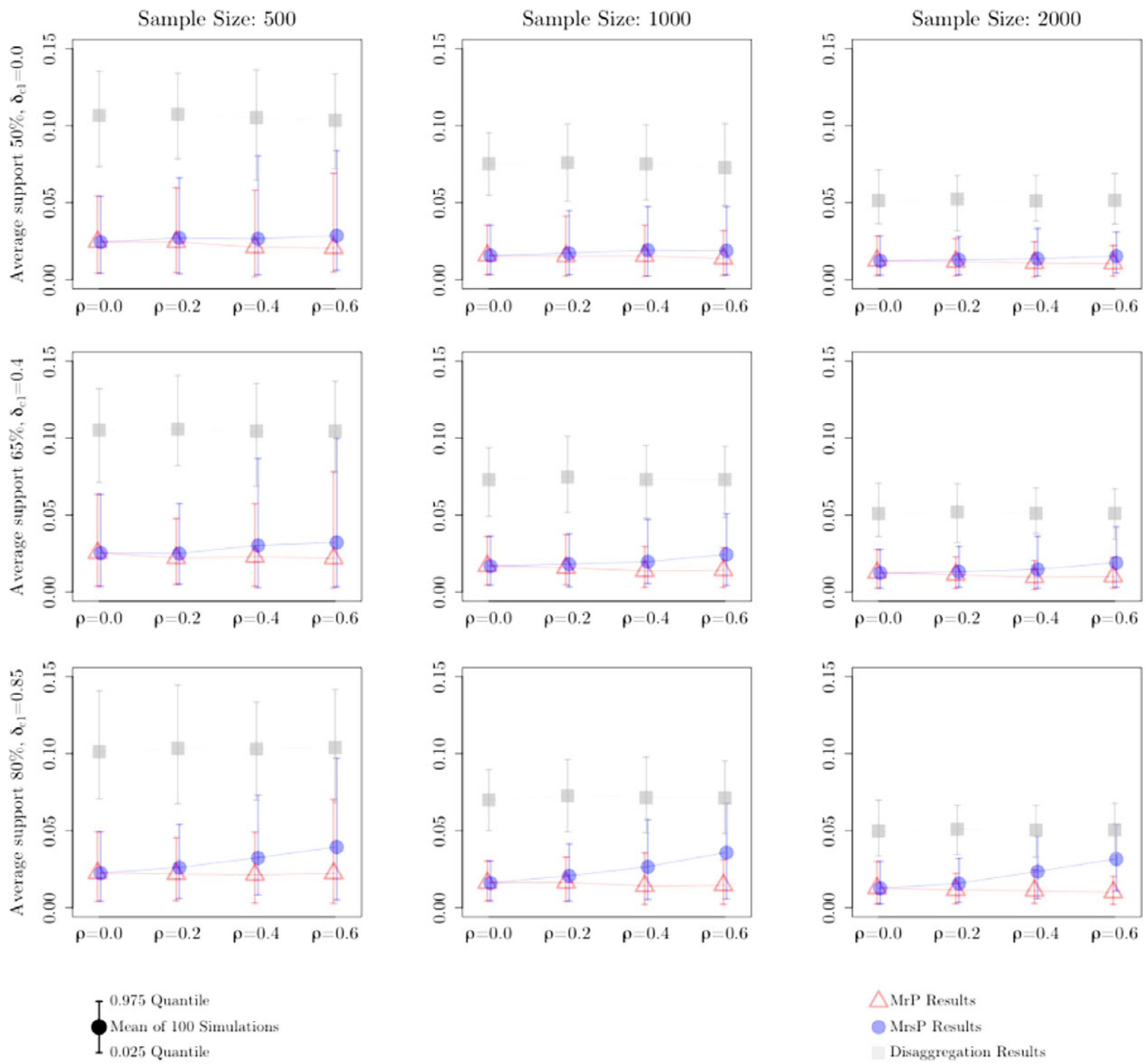
Figure 1 shows the prediction precision for all 36 combinations of the Monte Carlo analyses with three different sample sizes, three different average levels of public support, and four different correlations for disaggregation, classic MrP, and MrsP. Each of the nine plots reports the simulation results for the four different correlations of the individual-level variables (0, 0.2, 0.4, 0.6), with variations in sample sizes (500; 1,000; 2,000) and in average public support (50%, 65%, 80%). Concretely, the dots show the mean absolute error (MAE) for disaggregation, classic MrP, and simple MrsP for 25 subnational unit predictions that were estimated with 100 simulations each. The intervals document the range of the 100 MAEs for the three estimation approaches. Thus, we show the size of the prediction error. Note that a small error in prediction is substantively more important when public support is around 50% than for the case of a very distinct public opinion.

The findings confirm our expectations. First, as other studies have already shown, MrP systematically outperforms disaggregation (e.g., Lax and Phillips 2009b). Second, when the correlation between the individual-level variables is 0, classic MrP and the simple version of MrsP lead to exactly the same predictions. Third, increasing the sample size improves disaggregation but does not change the relative performance of MrP versus MrsP. Fourth, the deviations in prediction between classic MrP and MrsP grow as the correlation between the individual-level variables increases (but note that the predictions between classic MrP and MrsP are only distinguishable if the correlation is at a high level of 0.6). Finally, the relative performance of MrsP declines as the average public support deviates from 50%. While simple MrsP always outperforms disaggregation, its performance is weakest when the individual-level variables are highly correlated ($\rho = 0.6$) and when the average public support is at 80%. Although high correlations of about 0.6 are very unusual in applied work (see Appendix A) and average public support is typically between 20% and 80%, these results show under which conditions applied researchers should be careful when employing the simple version of MrsP.

To further investigate the claim that there are most likely no deviations in prediction between classic MrP

[6]For example, in the MrP analysis of state public opinion by Kastellec, Lax and Philipps (2016), the largest random effect has a variance of 0.3. In this study, the largest random effect has a variance of 0.45 (see Appendix C).

FIGURE 1 Monte Carlo Analyses: Mean Average Prediction Error



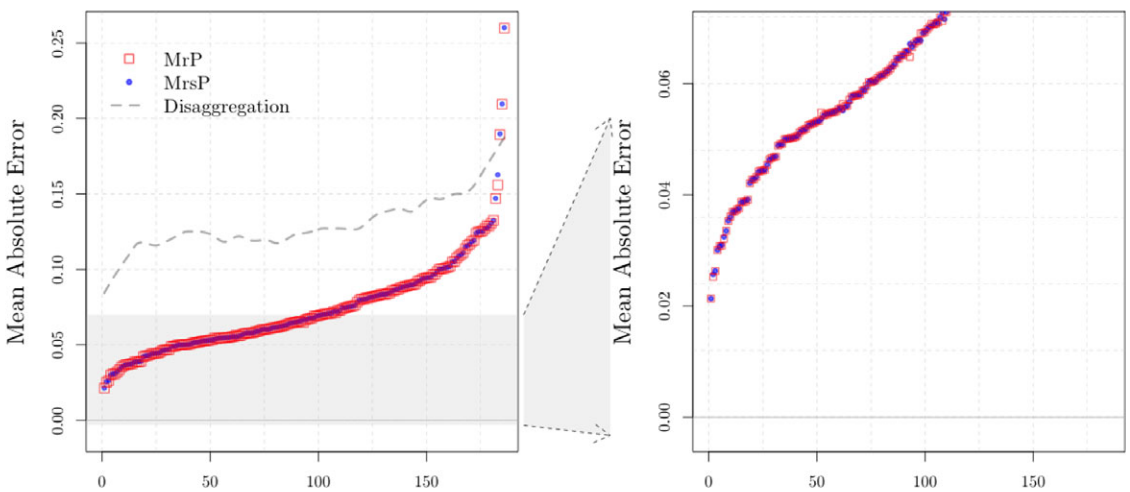
Note: Each plot shows the simulation results for a specific sample size ($N \in \{500, 1000, 2000\}$) with varying levels of average support (50%, 65%, 85%). The y-axis shows the mean absolute error (MAE), and the x-axis shows the different correlations among the individual-level variables.

and MrsP in applied work, we analyze real-world data on 186 direct democratic votes in Switzerland between 1990 and 2010.[7] We estimate the cantonal support with the true joint distributions (classic MrP) and the simple synthetic joint distribution (MrsP) by using data from the national VOX surveys ($n \approx 500 - 1,000$) and compare the predictions to the actual vote outcomes. We rely on a standard response model, including the demographic variables available as joint distributions from the census (gender, education, and age), the shares of German speakers and of Catholics as predictors on the subnational (i.e., cantonal) level, and random effects for regions and cantons. For all 186 votes, we estimate the following response model:

$$Pr(y_i = 1) = \Phi\left(\alpha_0 + \boldsymbol{\beta}\mathbf{X}_c + \alpha^{gender}_{j[i]} + \alpha^{education}_{k[i]} + \alpha^{age}_{m[i]} + \alpha^{canton}_{c[i]} + \alpha^{region}_{r[i]}\right).$$

[7]For the analysis, we use three different data sources: the Federal Statistical Office (BfS) collects vote outcome data for the cantons for all 186 direct democratic votes; the joint distributions for each canton are from the 2000 census; and the survey data are from the VOX research (Kriesi 2005).

FIGURE 2 Prediction Precision of Disaggregation, Classic MrP, and Simple MrsP Estimates with the Same Data for 186 Swiss Votes



Note: MAEs for 186 public votes are displayed. The right plot zooms in on the lower region of the left plot. The gray line reports the MAEs for disaggregation, the red dots show for classic MrP, and the blue dots for simple MrsP.

$$\alpha_j^{gender} \sim N(0, \sigma^2_{gender}), \quad \text{for } j = 1, \ldots, J.$$
$$\alpha_k^{education} \sim N(0, \sigma^2_{education}), \quad \text{for } k = 1, \ldots, K.$$
$$\alpha_m^{age} \sim N(0, \sigma^2_{age}), \quad \text{for } m = 1, \ldots, M.$$
$$\alpha_c^{canton} \sim N(0, \sigma^2_{canton}), \quad \text{for } c = 1, \ldots, C.$$
$$\alpha_r^{region} \sim N(0, \sigma^2_{region}), \quad \text{for } r = 1, \ldots, R.$$

Figure 2 reports the MAEs for disaggregation, classic MrP, and MrsP over all 186 direct democratic votes. The performance similarities of classic MrP and MrsP are striking. The estimates are so close that we can only identify differences once we zoom in (see the right-hand plot). The findings show that there is no difference in prediction precision between the two methods. While the Monte Carlo analyses already suggested that we will not find prediction deviations between classic MrP and MrsP in real-world data applications, the findings of the Swiss analysis support that claim. The results are virtually identical because the factors that theoretically drive the estimates of the two methods apart are not a problem. The random effects in the response model are lower than in the Monte Carlos analyses (the variances of the random effects are less than 0.1), the correlations among the individual-level variables are small (with a maximum of $\rho = -0.2$; see Appendix A), and the public support is on average 49.2%.[8]

[8]The 90% confidence interval is [18.7%, 82.1%].

MrsP with Adjusted Synthetic Joint Distributions

The simple version of MrsP discussed so far was based on the strong and unrealistic assumption that the individual predictors are perfectly uncorrelated with one another. Although the accordingly computed simple synthetic joint distributions deviate quite strongly from the real joint distribution data, the results of the Monte Carlo simulation and the Swiss data analysis have shown how well even this simple version of MrsP performs, and the illustrative theoretical example explained why. However, we can improve upon that by using information on the correlation among the individual-level variables from the survey data for the estimation of what we call *adjusted* synthetic joint distributions. This more elaborate version of MrsP goes beyond the unrealistic independence assumption and takes full advantage of the information provided in the survey.

To illustrate the estimation of adjusted synthetic joint distributions, we assume that the census provides data on the joint distributions for two variables, `gender` (male/female) and `age` (young/elderly), but not for an additional important predictor, `education` (low/middle/high), for which we only have data on the marginal distributions in the subnational units. In that case, we extend the available joint distribution data on `gender` and `age` with information on the marginal distribution of `education` based on the correlations in the survey. The basic point is that the survey data might, for

TABLE 4 Hypothetical Survey Information

(a) Real Survey Information

Gender	Age	N	Education Low	Education Middle	Education High
♂	young	360	120	180	60
♂	elderly	300	80	150	70
♀	young	350	140	150	60
♀	elderly	360	90	180	90
Survey distribution			31.4%	48.2%	20.4%

(b) Corrected Survey Information

Gender	Age	Education Low	Education Middle	Education High
♂	young	96	187	74
♂	elderly	64	156	86
♀	young	111	156	74
♀	elderly	72	187	110
Survey distribution		25%	50%	25%

Note: The top table shows hypothetical survey information, whereas the bottom table shows the corrected version that is necessary to create an optimal synthetic joint distribution.

example, suggest that elderly men tend to have higher educational backgrounds than elderly women. We use this information to estimate the adjusted synthetic joint distributions. If no census data on joint distributions are available for any variables (only marginals), the adjustment relies exclusively on the survey information.

In the illustrative example of Table 4, the distribution of `education` in the population of a given subnational unit is 25% low, 50% middle, and 25% high (note that the marginal distribution of `education` varies across subnational units; we illustrate in the following the adjustment procedure for one subnational unit). In the full survey data, which includes respondents from the entire country, the distribution of `education` will deviate to some extent. Table 4(a) reports the raw survey data from the entire country broken down by `gender`, `age`, and `education`. Every line shows the distribution of `education` for a specific `gender` and `age` combination, and the bottom line reports the overall distribution of `education` in the sample (31.4%/48.2%/20.4%), which differs from the true census information of the illustrative subnational unit (25%/50%/25%).

In a first step, we adjust the survey distribution of `education` to the known marginal distribution of the subnational unit by creating a correction factor for each of the three education categories. Since the survey includes too many less-educated respondents, given the marginal distribution of the subnational unit, we downweight the less-educated respondents by a factor of about 0.796 ($= \frac{0.250}{0.314}$). Table 4(b) shows how the survey information is corrected accordingly for all three `education` categories. Simply put, each line of Table 4(b) represents the best guess, based on the survey information, of how `education` is distributed in the subnational unit with the marginal distribution of 31.4%/48.2%/20.4% for each combination of `gender` and `age`.

In a second step, we compute the adjusted synthetic joint distribution using the census data of the true joint distributions (for `gender` and `age`) and the relative weights of the education distribution for each `gender` and `age` category derived from Table 4(b). In other words, we further break down the `gender/age` joint distributions of the census by extending each cell with the three `education` categories according to the relative shares of the corrected survey data information. For example, if 26% of the population in the illustrative subnational unit are young men, we further split that `gender/age` category using the relative shares of the first line of Table 4(b). Accordingly, 7.0% ($= 0.26 \times \frac{96}{(96+187+74)}$) of the population are young men with low education, 13.6% are young men with middle education, and 5.4% are young men with high education. Following that procedure, we estimate the adjusted synthetic joint distributions for each `gender/age/education` category in each subnational unit. The estimation of adjusted synthetic joint distributions takes full advantage of the survey information, assuming that the correlations among the individual-level variables are the same across subnational units. Thus, MrsP with adjusted synthetic joint distributions is a more fine-grained and elaborate technique than the previously discussed simple version.

In addition, the more elaborate version of MrsP allows for a more precise estimation of uncertainty. The most important source of uncertainty in MrP applications comes from the response model. This uncertainty can be estimated by generating N_{sim} draws of the response model's posterior vector for computing N_{sim} predictions. In MrsP applications, a further source of uncertainty is due to the the lack of precise data on the joint distributions. As discussed above, the derived adjusted synthetic joint distributions are the best guess given what we know from the survey data and all the information available on the joint and marginal distributions. To account for the uncertainty of the adjustment technique, we perform simulations as well. Using again the illustrative example of young men (see Table 4b), we create N_{sim} draws of a sample of 357 young men from a multinomial distribution with relative frequencies of $\frac{96}{357}$, $\frac{187}{357}$, and

$\frac{74}{357}$. We then estimate this form of uncertainty for each `gender/age/education` category in each subnational unit and integrate it with an additional loop in the estimation.[9] The real-world analyses of the next section apply this elaborate MrsP technique.

How MrsP Outperforms Classic MrP

Let us turn now to the key advantage of MrsP—namely, that it allows the modeling of additional powerful individual-level predictors—which promises to improve prediction precision. The specification of the response model on the individual level is in classic MrP predetermined by the census data (and restricted to three or four demographic variables). This limitation constrains even the most sophisticated research in the literature. For example, in the works of Lax and Phillips (2009a) and Warshaw and Rodden (2012), religion and age are important predictors of the political preferences they are investigating. Yet they could not include these variables on the individual level in their studies because of data availability reasons. Following Lax and Phillips (2009a), the standard procedure in the literature is to model variables that are not available in the joint distribution format as predictors on the subnational level of the response model (instead of the individual level). This is a reasonable strategy when these variables better explain variation among the units rather than within them.

In the case of MrsP, however, researchers only need marginal distributions (which is obviously unproblematic for age and religion in subnational units of the United States). Accordingly, the set of variables that can be modeled on the individual level is greatly enhanced with MrsP. Potentially interesting predictors are party identification, income, and employment status—just to name a few. The marginal distributions of these variables are typically available for subnational units. Which of these (or other) variables are potentially powerful predictors depends on the political preferences of interest. The key question is whether we can improve the prediction precision, when interesting predictors of political preferences are modeled as random effects on the individual level (MrsP) as compared to classic MrP, where such variables are included on the subnational level.

We first investigate that question with the Swiss data introduced above. One of the most important recent development in Swiss politics is the rise of the Swiss People's Party (SVP; see Kriesi et al. 2005). Particularly after 2007, when the de facto leader of the SVP was not reelected in the federal government, the party relied strongly on direct democratic campaigns to reinforce the narrative that they are in opposition against the "classe politique." Accordingly, in the legislative period 2007–11, identification with the SVP was a strong predictor of whether voters supported SVP referendums and initiatives, as several exit poll analyses show. We analyze the following four public votes of that legislative period, where the SVP was starkly engaged against the unified coalition of all other relevant Swiss parties and for which VOX survey data are available:

1. Initiative for municipal town hall approval of naturalization decisions.
2. Initiative to limit the government's right to communicate in referendum campaigns.
3. Referendum against an increase of the VAT for disability insurance.
4. Initiative to ban the construction of minarets.

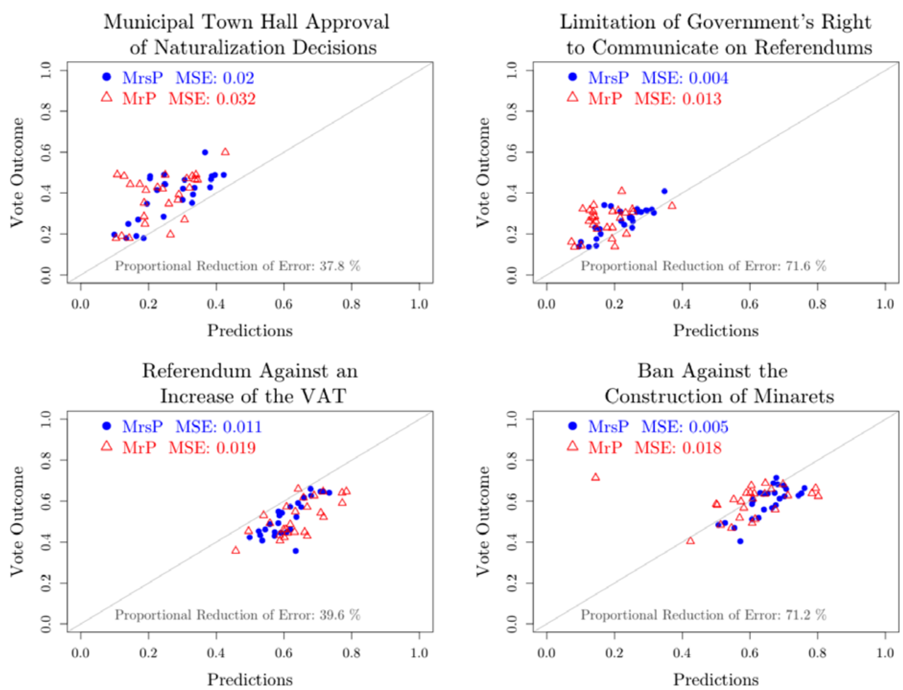
For the estimation of subnational public opinion, we specify the same baseline response model as discussed above with `gender`, `education`, and `age` as individual-level random effects, the shares of German speakers and of Catholics as cantonal variables, and random effects for regions and cantons. The baseline specification is extended for MrsP by adding party identification as an additional random effect on the individual level, whereas party identification is modeled for classic MrP as a subnational (i.e., cantonal) variable (like the shares of German speakers and of Catholics). We predict the cantonal support for the initiatives and referendums with classic MrP and with MrsP using the elaborate technique with adjusted synthetic joint distributions discussed in the previous section. The predictions are compared to the actual results.

Figure 3 plots the MrsP and classic MrP predictions against the true voting outcomes. In all four public votes, MrsP clearly outperforms classic MrP. The improvements in prediction precision are substantial, going up to a 72% reduction of prediction error in the case of the initiative to limit the government's right to communicate in referendum campaigns (i.e., a four times smaller mean squared error [MSE]). The significant improvements show that modeling party identification as a random effect on the individual level for SVP public votes leads to more accurate predictions than introducing that variable on Level 2.[10]

[9] Appendix D provides a detailed description of uncertainty estimation with real-world data.

[10] Appendix B reports a second Swiss example with income as an additional predictor of tax policy preferences. The findings are substantively the same.

FIGURE 3 Public Vote Outcomes and Classic MrP and MrsP Estimates for SVP Initiatives and Referendums with Party Identification as an Additional Predictor



Note: The x-axis reports the estimated share of yes votes, and the y-axis shows the true cantonal vote outcomes. MrsP includes party identification as a random effect on the individual level, whereas classic MrP includes party identification as a cantonal variable on Level 2. The sample sizes (N) vary between 525 and 680.

We additionally investigate a U.S. example, building on the Warshaw and Rodden (2012) analysis of the estimation of public opinion in state legislative districts. To cross-validate the findings, they compare the MrP predictions to the actual voting outcomes for direct democratic votes on same-sex marriage in Arizona, California, Michigan, Ohio, and Wisconsin. The response model includes race, gender, and education as individual-level predictors and, as district-level predictors, the median income and the shares of the urban population, of veterans, and of same-sex couples. The authors explicitly state that age is a critical predictor that they cannot model because the census has no data breakdown for race/gender/education *and* age (Warshaw and Rodden 2012, 208). This is a case where MrsP goes beyond the data limitation of current MrP applications, as the marginal distributions of age are available for U.S. state legislative districts.

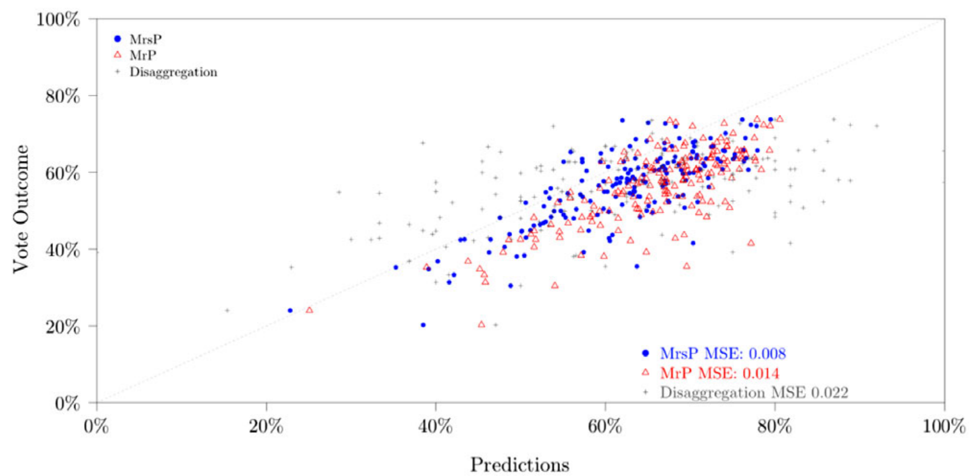
For the analysis, we replicate the Warshaw and Rodden (2012) public opinion estimates[11] before executing the MrsP analysis using adjusted synthetic joint distributions for `race`, `gender`, `education`, and `age`. In the case of MrsP, we model `age`, which cannot be modeled in classic MrP, as an individual-level predictor of same-sex marriage preferences. Figure 4 plots the disaggregation, classic MrP, and MrsP estimates against the true voting outcome. The disaggregation MSE is 0.022, that of classic MrP 0.014, and that of MrsP 0.008. Relying on that measure, MrsP improves upon classic MrP (the error decreases by 43%) even more than classic MrP improves upon disaggregation (the error decreases by 35%).

The U.S. analysis highlights the conditions under which MrsP provides substantially better predictions than classic MrP. The improvement in prediction is because age is an important individual-level predictor of political preferences in same-sex marriage questions: No other estimated random effect variance term is as large in the response model (education has the second largest, with a variance of almost half the size; see Appendix C). This example shows that adding just one powerful individual-level predictor can have an extremely strong effect on the precision of the predictions.

The reported MrsP predictions for the Swiss and U.S. analyses are estimated using adjusted synthetic joint distributions. We have also estimated the models with the

[11]We are indebted to the authors for providing us with detailed replication files and their data set.

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



Note: The x-axis reports the estimated share of yes votes, and the y-axis shows the true vote outcome for state senate districts. MrsP includes age as a additional individual-level random effect, whereas age cannot be modeled in classic MrP.

simple version of MrsP. The predictions of the more elaborate MrsP technique are a little better, but the improvement is quite marginal. In the U.S. case, the mean squared error of MrsP with the simple synthetic joint distribution is 0.0091 (compared to 0.0080 for elaborate MrsP and 0.014 for classic MrP).[12] One reason why elaborate MrsP does not improve more on simple MrsP is because `age` is not strongly correlated with the other individual-level variables.[13] This is consistent with the Monte Carlo analyses, which have shown that simple MrsP is powerful when correlations are low. Besides prediction precision, however, it is important to recall that elaborate MrsP allows for the estimation of uncertainty induced by the adjusted synthetic joint. Appendix D provides a detailed discussion of uncertainty estimation for the U.S. case and shows that the uncertainty induced by the adjusted synthetic joint distribution is very small (2% of the overall uncertainty), which is not surprising, given the large sample size (n = 17,611).

Based on the findings of the Monte Carlo simulations and the Swiss and U.S. analyses, we recommend that applied researchers proceed as follows when considering using MrP:

1. Analyze the survey data and explore which individual-level variables are strong predictors of the studied political preferences.
2. Check which variables, if any, are available as joint distributions from the census.
   - 2.1 Use classic MrP if all important individual-level predictors are available as joint distributions.
   - 2.2 Use MrsP if at least one powerful predictor is not available as joint distributions.
     - 2.2.2 Use elaborate MrsP, particularly when the individual-level predictors are correlated and when the predicted average public support is far from 50%.
     - 2.2.1 If the correlations among the individual-level predictors are at very low levels, simple MrsP is fine for point predictions (note that uncertainty induced by the simple joint distribution cannot be estimated).

Our recommendations are guidelines for helping researchers making sound judgments when applying MrP. All of the three discussed MrP approaches have pros and cons. Simple MrsP is by far the least demanding technique in terms of estimation and data requirements, and the predictions are in many cases much better compared to the classic version. Particularly in developing countries with

[12]In the Swiss example, the precision gains of the elaborate versus simple MrsP are in the same rather marginal magnitude.

[13]The correlations vary from −0.17 (race) to −0.01 (education) and 0.05 (gender). See Appendix A.

sparse data, where classic MrP cannot be applied, simple MrsP might be a good option. However, researchers should pay attention to the limitations of simple MrsP (correlation of individual-level variables, lopsided public support, and inaccurate uncertainty estimation). The presented findings suggest that elaborate MrsP is a very attractive and widely applicable extension of MrP. For some (rare) cases in advanced countries, classic MrP should be as good as (or might even be superior to) elaborate MrsP. This can be the case when the important predictors are available as joint distributions and the survey sample is too small to extract meaningful information for the construction of adjusted synthetic joint distributions.

Conclusion

The comparative study of subnational units has attracted growing interest, and the estimation of reliable public preference measures for subnational units is a critical element in the literature. MrP generates reliable public opinion estimates for subnational units with standard national polling data. The numerous MrP studies published in recent years show that the method stimulates research—for example, on the responsiveness of subnational politicians and administrations to voters' preferences (Lax and Phillips 2012; Leemann and Wasserfallen 2016; Tausanovitch and Warshaw 2014). However, the application of MrP has been restricted to a few countries because of the stringent data requirements of the current standard approach, which requires detailed census data in the form of joint distributions (researchers need to know, for example, how many 18–35-year-old women with a university degree live in each subnational unit).

The presented alternative application of MrP, MrsP, relies on the marginal distribution of individual-level variables (e.g., the shares of women, of university graduates, and of 18–35-year-old citizens in each subnational unit), which extends the use of the method to countries without joint distribution census data. This extension of subnational public opinion estimation with MrP is important for stimulating comparative subnational research in less-developed countries with more restricted data availability. We compared MrsP and the current standard MrP approach theoretically using Monte Carlo analyses and Swiss and U.S. data examples. The findings show that, using the same predictors, MrsP usually performs in standard applications as well as the current standard approach, and that MrsP increases the prediction precision when additional strong predictors beyond the standard demographic variables are added to the response model.

The improvements of MrsP also promise to further stimulate subnational comparative research in developed countries with census data. So far, scholars have relied on rather generic response models with three or four demographic individual-level variables as predictors of various policy preferences. With MrsP, the individual-level predictors can be selected depending on the political preferences of interest. Public opinion polls show, for example, that churchgoing is associated with policy views on abortion, and views on free trade policies are correlated with the trade exposure of an individual's job (Mayda and Rodrik 2005). The presented findings suggest that modeling such strong predictors increases the prediction precision of MrP substantially. MrsP thus takes MrP to new countries and improves the predictive power of the method by allowing more model flexibility. The guidance provided in this article helps scholars to develop MrP applications that take full advantage of the available data for the estimation of subnational policy preferences.

Appendixes

Appendix A: Correlation of Individual-Level Variables

The Monte Carlo analyses suggest that correlations below $\approx |0.4|$ should be unproblematic for MrsP applications. In the U.S. and Swiss data, the correlations are much lower, as Tables A1 and A2 show.

TABLE A1 Average Correlation Matrix over 186 Swiss Exit Polls

	Education	Age	Gender
Education	1.00	−0.12	−0.20
Age	−0.12	1.00	0.02
Gender	−0.20	0.02	1.00

TABLE A2 Correlation Matrix from U.S. Data (Warshaw and Rodden 2012)

	Age	Education	Race	Gender
Age	1.00	−0.01	−0.17	0.05
Education	−0.01	1.00	−0.09	−0.04
Race	−0.17	−0.09	1.00	−0.01
Gender	0.05	−0.04	−0.01	1.00

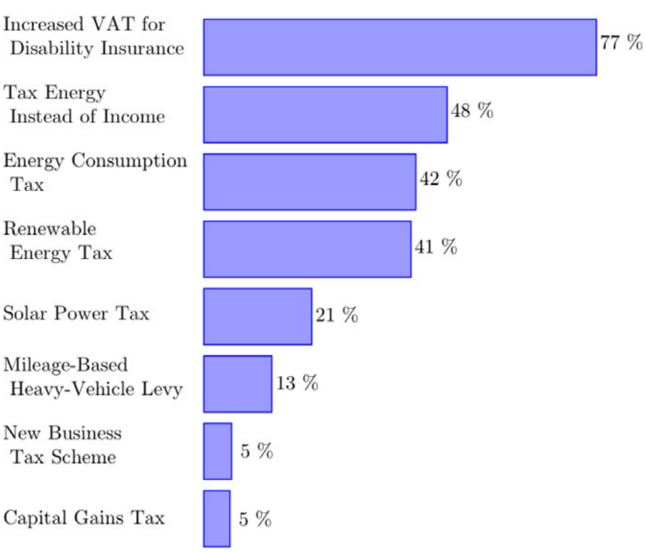
Appendix B: Additional Swiss Analysis

The article shows that MrsP increases the prediction precision compared to classic MrP with the example of introducing party identification as an additional individual-level predictor for SVP initiatives and referendums for the 2007–11 legislative period. Another interesting individual-level predictor of political preferences is income. Earned income has been identified in the literature as an important determinant of tax policy preferences that is typically politicized by the left (Bartels 2008; Corneo and Grüner 2002).

We find eight public votes on taxation with a distinct left-right campaign dynamic in the Swiss survey data (VOX). As in the SVP example, we rely on the baseline specification of the response model. For assessing the gains in prediction precision, we introduce income as an additional random effect for MrsP and compare the MrsP predictions using simple synthetic joint distributions to classic MrP, where we model income as a variable on the subnational level.

The proportional mean squared error reductions reported in Figure A1 corroborate that MrsP outperforms classic MrP when a powerful predictor of the investigated political preferences is introduced on the individual level. The mean squared error reductions between 5% and 77% are again substantial improvements in prediction precision.

FIGURE A1 Reduction in Mean Squared Errors between Classic MrP and MrsP Estimates



Note: Swiss votes on taxation with income as additional predictor are shown; sample sizes vary between 445 and 819.

Appendix C: Replication of Warshaw and Rodden (2012)

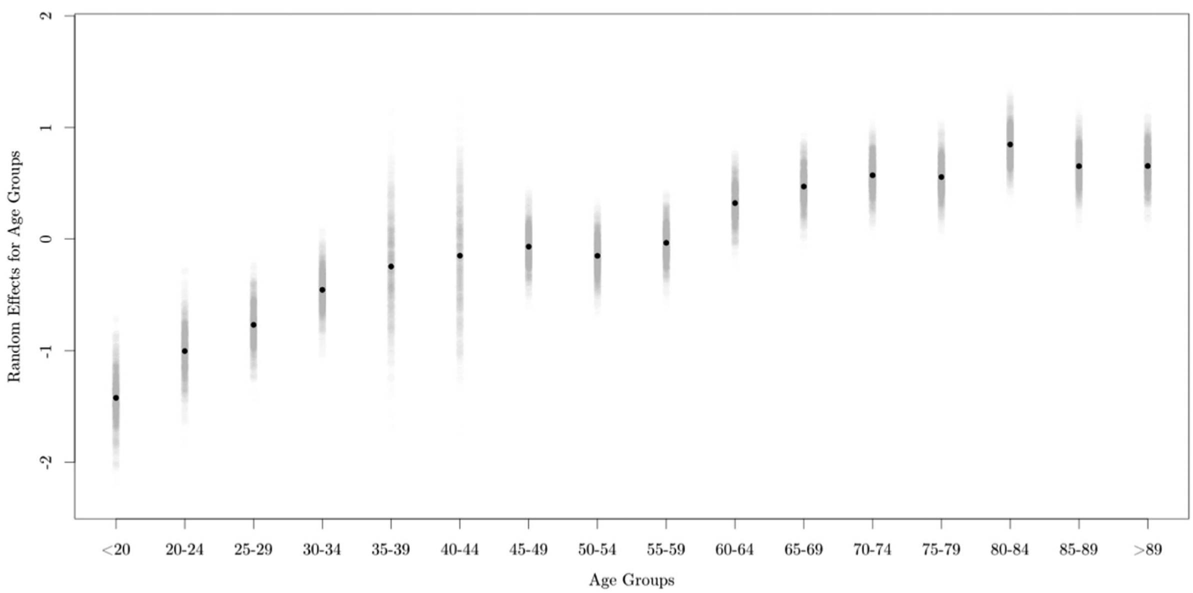
Table A3 presents the response model estimates of the Warshaw and Rodden (2012) replication analysis. The MrsP model includes age as an individual-level predictor. The random effect for age is large (and the other random effects are also larger in the MrsP model), which

TABLE A3 MrP and MrsP Response Model Estimates

	MrsP Model	MrP Model
Gender	-0.49^{***}	-0.43^{***}
	(0.04)	(0.03)
Income (district)	-0.72^{***}	-0.79^{***}
	(0.18)	(0.17)
Urban (district)	-0.65^{***}	-0.60^{***}
	(0.10)	(0.10)
Veteran (district)	-0.59	-0.42
	(0.56)	(0.55)
Religion (state)	1.70^{***}	1.56^{***}
	(0.28)	(0.27)
Union members (state)	-0.91	-0.82
	(0.63)	(0.61)
Same-sex couples (district)	-34.17^{***}	-34.35^{***}
	(3.41)	(3.13)
Constant	2.48^{***}	2.15^{***}
	(0.36)	(0.29)
Variance: `district`	0.08	0.07
Variance: `state`	0.01	0.00
Variance: `age.group`	0.45	not included
Variance: `education.group`	0.24	0.21
Variance: `region`	0.01	0.01
Variance: `race`	0.07	0.04
AIC	20415.85	21103.44
BIC	20524.72	21204.53
Num. obs.	17611	17611
Num. of districts	1779	1779
Num. of states	48	48
Num. groups: age	16	
Num. groups: education	5	5
Num. groups: region	4	4
Num. groups: race	4	4

Note: $^{***}p < .001$, $^{**}p < .01$, $^{*}p < .05$.

FIGURE A2 Random Effects for Age from the MrsP Response Model



Note: Estimates of the random effects and 1,000 simulations drawn from the posterior vector for illustrating estimation uncertainty.

shows, together with the AIC and BIC values, that introducing age increases the predictive power of the model substantially.

To further illustrate how strongly age is correlated with attitudes toward same-sex marriage, Figure A2 plots the 16 different random effects for age from the MrsP response model. The older a respondent is, the larger the estimated random effect ($y = 1$ is equivalent to being opposed to same-sex marriages).

Appendix D: Uncertainty with Adjusted Synthetic Joint Distributions

To further illustrate the estimation of uncertainty for the elaborate version of MrsP, we discuss in more detail the U.S. example. In the U.S. census, age is classified in 15 different groups. To create adjusted synthetic joint distributions, we use the census information on the age structure in each district and the survey data. Table A4 shows the age distribution in the survey (second column) and in district j (third column). The correction factor (fourth column) shows how the data have to be adjusted.

In the adjustment step, we correct the survey data for each district so that the age shares for every ideal type are according to the survey information, while making sure (with the use of the correction factor) that the eventual marginal distribution of age is equal to the true marginal distribution in each district. Table A5 shows the accordingly derived adjusted synthetic joint distribution for district j. Every line represents one of the 40 ideal types (`race`$\times$`gender`$\times$`education`), and each columns refers to a specific age category.

TABLE A4 Generating the Adjusted Synthetic Joint Distribution for District j

Age Group	Survey (All Districts)	Census (for District j)	Correction Factor, c_j
<20	0.06	0.09	1.60
20–24	0.08	0.08	0.97
25–29	0.10	0.07	0.75
30–34	0.09	0.08	0.86
35–39	0.11	0.09	0.78
40–44	0.12	0.10	0.81
45–49	0.11	0.10	0.92
50–54	0.09	0.09	1.02
55–59	0.07	0.08	1.20
60–64	0.06	0.07	1.17
65–69	0.05	0.06	1.12
70–74	0.04	0.04	1.26
75–79	0.03	0.03	1.28
80–84	0.01	0.02	2.38
>84	0.01	0.01	1.29

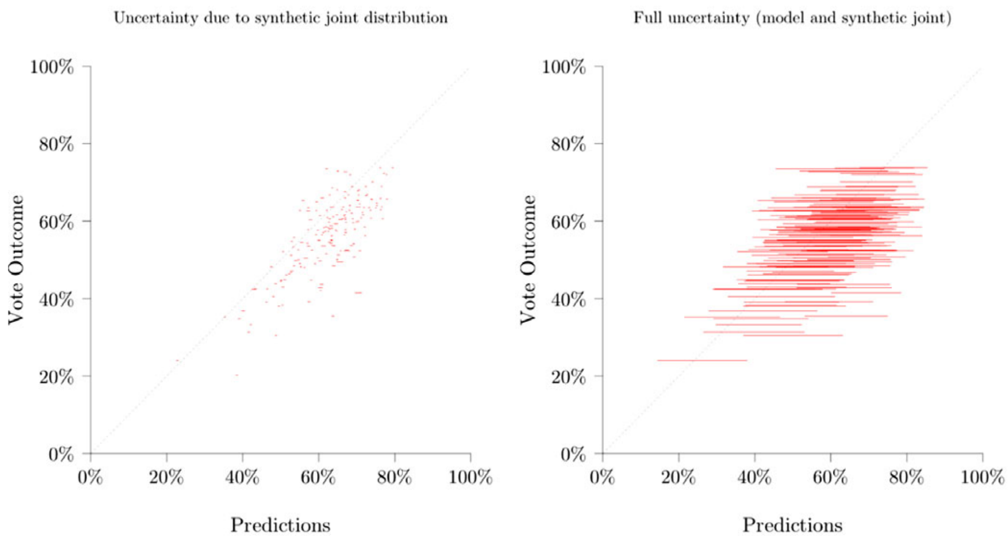
Table A5 Adjusted Synthetic Joint Distribution for District *j*

<20	20–24	25–29	30–34	35–39	40–44	45–49	50–54	55–59	60–64	65–69	70–74	75–79	80–84	>84
4	6	1	4	1	5	5	6	5	4	5	2	3	0	0
13	11	14	16	31	18	21	12	12	10	3	6	1	3	2
4	18	19	7	15	21	20	18	11	6	9	4	2	1	0
0	5	6	14	15	11	14	10	9	4	1	1	1	1	0
0	0	6	2	3	8	4	14	8	4	2	1	0	0	2
10	8	6	9	8	6	4	7	3	7	7	2	3	3	0
18	24	29	23	25	26	26	21	20	13	6	10	5	0	2
2	38	33	29	31	26	39	36	15	7	7	5	3	2	0
0	8	19	11	19	12	28	7	11	3	2	4	1	1	0
0	1	7	14	10	19	18	19	5	5	5	3	1	0	1
11	17	17	27	23	21	14	6	8	2	1	5	3	1	0
19	21	28	22	24	14	15	6	5	5	2	2	0	1	0
10	19	21	21	21	14	19	9	11	3	2	1	2	0	1
0	5	14	19	14	9	13	7	6	3	2	1	0	0	0
0	1	2	7	1	7	6	4	3	2	2	1	0	1	0
6	16	18	28	23	13	11	10	9	6	5	6	2	2	1
14	21	27	33	25	18	9	11	8	6	4	6	1	1	0
11	32	21	24	13	17	17	13	6	7	4	3	3	1	0
0	8	12	14	16	14	9	6	2	1	1	1	2	1	0
0	1	6	12	14	11	5	8	5	1	0	4	0	0	2
28	20	16	22	19	30	27	22	24	29	36	23	22	20	11
62	99	97	139	146	171	195	135	115	108	111	84	51	40	26
20	136	123	130	121	166	188	198	156	122	87	75	48	39	15
0	62	128	178	161	193	168	138	124	89	66	54	40	29	15
0	11	75	116	127	139	147	184	141	122	87	61	57	23	16
27	19	20	20	20	27	25	29	27	43	41	36	55	31	17
50	99	99	111	139	184	200	213	208	197	176	175	125	96	46
32	150	155	190	180	254	306	250	228	168	149	126	85	83	47
0	59	138	185	167	174	212	184	127	92	67	64	35	27	20
1	27	96	128	124	145	178	191	169	90	84	60	32	17	25
7	3	1	2	0	2	3	0	0	3	1	4	0	0	1
6	8	7	7	8	10	7	9	3	6	6	5	2	1	0
5	8	10	11	12	13	10	13	6	2	1	1	3	0	2
0	2	11	21	10	15	7	11	7	5	3	1	0	0	1
0	3	7	15	16	10	17	7	12	6	5	3	1	0	0
2	5	1	3	1	2	1	3	4	4	3	1	2	2	1
5	4	6	10	5	10	6	10	6	8	4	5	4	1	2
3	24	20	10	14	21	8	14	10	5	4	3	3	1	2
0	9	9	13	5	15	19	9	4	5	0	1	2	0	2
0	3	15	13	8	9	4	19	4	4	0	0	0	0	1

To estimate the uncertainty, we take into account that our survey information approximates how age groups are distributed within an ideal type. The distribution of age categories can be described by a multinomial distribution, which depends on the survey sample size (equivalent to the number of trials) and the relative frequencies. Accordingly, we run simulations and draw a number of potential outcomes to capture the uncertainty that is induced by the adjustment step.

Figure A3 reports in the left plot the uncertainty that is induced by the adjusted synthetic joint distribution. The right plot shows the full uncertainty (including the model

FIGURE A3 Sources of Uncertainty with MrsP



Note: The left plot shows the uncertainty (95% confidence intervals) induced by the adjusted synthetic joint distribution, and the right plot shows the full uncertainty (including model uncertainty).

uncertainty). The findings show that the uncertainty induced by the synthetic joint distribution is minimal (only 2% of the total uncertainty). Two important factors drive this finding: first, the large sample size, and, second, the low correlation among the individual-level predictors (no correlation is larger than 0.2; the largest correlation is 0.17 for `race` and `age`). Thus, the drawback of the simple version of MrsP—that we cannot estimate the uncertainty induced by the synthetic joint distribution—is not as consequential when the sample is large and the correlations are low.

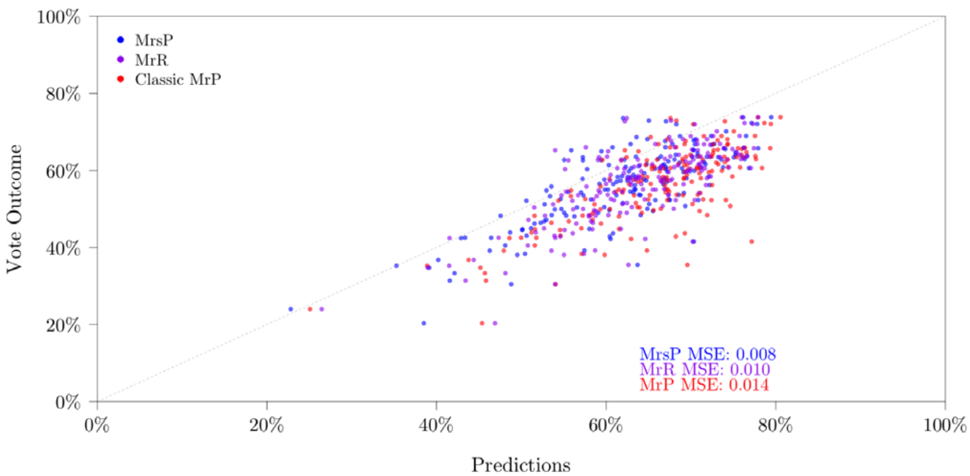
Appendix E: Raking as an Alternative to Synthetic Poststratification

MrsP uses information from marginal distributions for the creation of synthetic joints. Raking, a standard procedure in survey research, offers an alternative for poststratification with marginal distributions. In a nutshell, raking assigns weights based on the marginal distribution of one variable, then calculates—conditional on the derived weights—new weights with the marginal distribution of the second variable, and continues with this iterative proportional fitting until the weighted survey best approximates the distribution of the target population (Deming and Stephan 1940; Fienberg 1970). Like MrsP, raking relies on information from marginal distributions and thus offers a potential alternative for poststratification.

However, poststratification with raking is slightly differently implemented than MrsP. Rather than using the calculated joint distributions from the synthetic joint distributions, multilevel regression with raking (MrR) weighs the predictions for voter types (see Steps 1–3 of MrP discussed in "'Classic MrP' and Its Limits" in the main text) with the weights calculated by iterative proportional fitting. We present below an application of MrR and compare its predictions to classic MrP and MrsP. In this example, MrsP yields more precise predictions than MrR. The reason for this finding is that MrsP with synthetic (adjusted) joint distributions exploits additional information from the survey data. An additional advantage of MrsP is that it allows us to take into account the uncertainty over the true joint distribution for generating the uncertainty measure (this cannot be done with raking). Thus, although MrR certainly is a feasible alternative, we recommend MrsP.

To assess the prediction precision of MrR, we again analyze the Warshaw and Rodden (2012) example discussed in "How MrsP Outperforms Classic MrP" in the main text. In this example, we know the true joint distribution for `race`, `gender`, and `education`, but not for `age`. We want to include `age` because it is expected to be a powerful predictor of an individual's preference

FIGURE A4 Public Vote Outcomes and Classic MrP, MrsP, and MrR Estimates for the Warshaw and Rodden (2012) Analysis on Same-Sex Marriage Referendums in Arizona, California, Michigan, Ohio, and Wisconsin with Age as Additional Predictor



Note: The x-axis reports the estimated share of yes, votes and the y-axis shows the true vote outcome for state senate districts. MrsP and MrR include age as a additional individual-level random effect, whereas age cannot be modeled in classic MrP.

over same-sex marriage. To generate MrR estimates, we take the known joint distribution of `race`, `gender`, and `education` and treat it as if it were a univariate distribution with 40 possible values (all combinations from two `gender` categories, four `race` categories, and five `education` categories). This allows us to keep the known information of the partial joint distribution. We then add the 15 `age` categories, which yields 600 different voter types. Finally, using the predicted probabilities for each of these 600 voter types, we employ raking over two variables: the `race-gender-education` variable (with 40 categories) and the `age` variable (with 15 categories).

The findings presented in Figure A4 show that MrR outperforms classic MrP. This is because MrR takes `age` into account when modeling the individual preferences. Yet MrR has a larger mean squared error (MSE) than MrsP because it does not exploit the survey information about the joint distribution. Please also note that the raking results are sensitive to the sequence of the variables (we use the `rake` command from the `survey` library [Lumley 2004] in R). The MSE is 0.011 when we first specify the `race-gender-education` variable and then the `age` variable. In the reversed iteration, the MSE is 0.010 (we report in Figure A4 the better MrR outcome for this example).

References

Angrist, Joshua D., and Jörn-Steffen Pischke. 2008. *Mostly Harmless Econometrics: An Empiricist's Companion.* Princeton, NJ: Princeton University Press.

Bartels, Larry M. 2008. *Unequal Democracy: The Political Economy of the New Gilded Age.* Princeton, NJ: Princeton University Press.

Beck, Nathaniel. 2011. "Is OLS with a Binary Dependent Variable Really OK? Estimating (Mostly) TSCS Models with Binary Dependent Variables and Fixed Effects." Unpublished manuscript.

Corneo, Giacomo, and Hans Peter Grüner. 2002. "Individual Preferences for Political Redistribution." *Journal of Public Economics* 83(1): 83–107.

Deming, W. Edwards, and Frederick F. Stephan. 1940. "On a Least Squares Adjustment of a Sampled Frequency Table When the Expected Marginal Totals Are Known." *Annals of Mathematical Statistics* 11(4): 427–44.

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.

Fienberg, Stephen E. 1970. "An Iterative Procedure for Estimation in Contingency Tables." *Annals of Mathematical Statistics* 41(3): 907–17.

Gelman, Andrew, and Thomas C. Little. 1997. "Poststratification into Many Categories Using Hierarchical Logistic Regression." *Survey Research* 23(2): 127–35.

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): 762–76.

Kastellec, Jonathan P., Jeffrey R. Lax, Michael Malecki, and Justin H. Phillips. 2016. “Polarizing the Electoral Connection: Partisan Representation in Supreme Court Confirmation Politics.” *Jounal of Politics* 77(3): 787–804.

Kastellec, Jonathan P., Jeffrey R. Lax, and Justin H. Phillips. 2016. “Estimating State Public Opinion with Multi-Level Regression and Poststratification using R.” http://www.princeton.edu/˜jkastell/MRP_primer/mrp_primer.pdf.

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): 767–84.

Kriesi, Hanspeter. 2005. *Direct Democratic Choice.* Lanham, MD: Lexington Books.

Kriesi, Hanspeter, Romain Lachat, Peter Selb, Simon Bornschier, and Marc Helbling. 2005. *Der Aufstieg der SVP. Acht Kantone im Vergleich.* Zürich: Neue Zürcher Zeitung.

Lax, Jeffrey R., and Justin H. Phillips. 2009a. “Gay Rights in the States: Public Opinion and Policy Responsiveness.” *American Political Science Review* 103(3): 367–86.

Lax, Jeffrey R., and Justin H. Phillips. 2009b. “How Should We Estimate Public Opinion in the States?” *American Journal of Political Science* 53(1): 107–21.

Lax, Jeffrey R., and Justin H. Phillips. 2012. “The Democratic Deficit in States.” *American Journal of Political Science* 56(1): 148–66.

Leemann, Lucas, and Fabio Wasserfallen. 2016. “The Democratic Effect of Direct Democracy.” *American Political Science Review* 110(4): 1–13.

Levendusky, Matthew S., Jeremy C. Pope, and Simon D. Jackman. 2008. “Measuring District-Level Partisanship with Implications for the Analysis of U.S. Elections.” *Journal of Politics* 70(3): 736–53.

Lumley, T. 2004. “Survey: Analysis of Complex Survey Samples. R Package Version 3.30.”

Maddala, Gangadharrao S. 1983. *Limited-Dependent and Qualitative Variables in Econometrics.* Vol. 3. Cambridge: Cambridge University Press.

Mayda, Anna Maria, and Dani Rodrik. 2005. “Why are Some People (and Countries) more Protectionist than others?” *European Economic Review* 49: 1393–1430.

Miller, Warren E., and Donald W. Stokes. 1963. “Constituency Influence in Congress.” *American Political Science Review* 57(1): 45–46.

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. “Bayesian Multilevel Estimation with Poststratification: State-Level Estimates from National Polls.” *Political Analysis* 12(4): 375–85.

Selb, Peter, and Simon Munzert. 2011. “Estimating Constituency Preferences from Sparse Survey Data Using Auxiliary Geographic Information.” *Political Analysis* 19(4): 455–70.

Snyder, Richard. 2001. “Scaling Down: The Subnational Comparative Method.” *Studies in Comparative International Development* 36(1): 93–110.

Steenbergen, Marco R., and Bradford S. Jones. 2002. “Modeling Multilevel Data Structures.” *American Journal of Political Science* 46(1): 218–37.

Tausanovitch, Chris, and Christopher Warshaw. 2014. “Representation in Municipal Government.” *American Political Science Review* 108(3): 605–41.

Warshaw, Christopher, and Jontahan Rodden. 2012. “How Should We Measure District-Level Public Opinion on Individual Issues?” *Journal of Politics* 74(1): 203–19.

Ziblatt, Daniel. 2008. “Does Landholding Inequality Block Democratization? A Test of the ‘Bread and Democracy’ Thesis and the Case of Prussia.” *World Politics* 60(4): 610–41.