How Should We Estimate Sub-National Opinion Using MRP? Preliminary Findings and Recommendations

Jeffrey R. Lax
Department of Political Science
Columbia University
JRL2124@columbia.edu

Justin H. Phillips*
Department of Political Science
Columbia University
JHP2121@columbia.edu

April 10, 2013

Abstract

Over the past few years, multilevel regression and poststratification (MRP) has become an increasingly trusted tool for estimating public opinion in sub-national units from national surveys. Especially given the proliferation of this technique, more evaluation is needed to determine the conditions under which MRP performs best and to establish benchmarks for expectations of performance. Using data from common content of the Cooperative Congressional Election Study, we evaluate the accuracy of MRP across a wide range of survey questions. In doing so, we consider varying degrees of model complexity and identify the measures of model fit and performance that best correlate to the accuracy of MRP estimates. The totality of our results will enable us to develop a set of guidelines for implementing MRP properly as well as a set of diagnostics for identifying instances where MRP is appropriate and instances where its use may be problematic.

*For helpful comments we thank Andrew Gelman. For research assistance we thank Eurry Kim.

1 Introduction

Empirical scholars of representation have long been interested in the relationship between public preferences and government action. In normative accounts of representative democracy there is near universal agreement that some minimal matching of policy (or roll call votes) to public opinion is required. Indeed, the responsiveness of elected officials to mass preferences is one way that political scientists can and do evaluate the quality of a democracy.

Of course, studying the link between public opinion and government action requires accurate measures of constituent preferences. Such measures have been difficult to obtain, particularly for subnational units of interest, including states and congressional districts. National public opinion polls, while commonplace, rarely have sufficiently large samples to draw inferences about subnational units. Additionally, comparable polls across all or even most states or legislative districts are incredibly rare and prohibitively expensive.

To overcome this obstacle, scholars have traditionally relied upon one of two approaches. The first is to employ some proxy for public opinion, such as sociodemographics (Kalt and Zupan 1984, Krehbiel 1993, Levitt 1996) or presidential election returns (Erikson and Wright 1980, Ansolabehere, Snyder and Stewart 2001, Canes-Wrone, Cogan and Brady 2002). These measures, while readily available, have been criticized for their imprecision (Jackson and King 1989, Cohen 2006). The second approach—disaggregation—combine numerous national level surveys (usually over many years) and then computes the mean response by the geographic units of interest (Erikson, Wright, and McIver 1993, Brace et al. 2002, Clinton 2006). Unfortunately, this technique is almost always limited to those survey questions that appear in multiple opinion polls and for which opinion is temporally stable.

More recently, scholars have revived simulation techniques, the most recent iteration of which is multilevel regression and poststratification (MRP). MRP, developed by Gelman and Little (1997) and extended by Park, Gelman, and Bafumi (2004, 2006), uses national surveys and advances in Bayesian statistics and multilevel modeling to generate opinion estimates by demographic-geographic subgroups. MRP uses individual survey responses from national polls and regression analysis to estimate the opinions of thousands of different respondent types. From these estimates, a measure of state or district opinion is created by determining how many of each type live within the geographic unit of interest.

Several research teams have validated MRP, demonstrating that it generally outperforms prior approaches to estimating subnational public opinion (Park, Gelman and Bafumi 2006, Lax and Phillips 2009a, Warshaw and Rodden 2012). This work also suggests that MRP can produce accurate estimates using fairly simple demographic-geographic models of survey response and small amounts of survey data—as little as a single national poll (approximately 1,500 respondents) for state level opinion estimates. As a result, MRP has quickly become an accepted research tool, "emerging as a widely used gold standard for estimating preferences from national surveys" (Selb and Munzert, 2011 p. 456). Research employing MRP has already appeared in the top political science journals and MRP has been employed to study to numerous substantive questions, including the responsiveness of state governments (Lax and Phillips 2009b, 2012), state supreme court abortion decisions (Caldarone, Canes-Wrone, and Clark 2009), roll call voting on U.S. Supreme Court nominations (Kastellec, Lax, and Phillips 2010), and the diffusion of public policy (Pacheco 2012).

However, one might worry that substantive applications of MRP are outpacing our knowledge of the strengths and limitations of the methodology. Systematic evaluations of the predictive accuracy of MRP have been limited largely to presidential voting (where MRP estimates can be compared to actual election returns) and to public support for same sex marriage (where MRP estimates can be compared to state polls and voting on corresponding ballot measures).[1] While MRP has been shown to perform well in these areas, the fairly limited scope of this evaluative work means that several crucial questions remain unanswered. Will MRP perform equally well across a wide range of issues and survey question types? Are there metrics that will allow researchers to identify whether a particular set of estimates are likely to be accurate? What steps might be taken to maximize the performance of MRP? Are there conditions under which MRP should be avoided? The answers to these questions will provide much needed guidance to both users and consumers of MRP.

In this paper, we evaluate the predictive accuracy of MRP using a set of 50 survey questions from the 2010 Cooperative Congressional Election Study (CCES) and the new MRP package in R. For each question we treat the sample of respondents as the population of interest. We then obtain "true" opinion for each state (the mean response among all respondents from that state) and the necessary poststratification data (since we treat the survey respondents as our population it makes sense to use the survey as opposed to the Census to create poststratification weights). We then evaluate the accuracy of MRP by comparing MRP estimates to "true" opinion. By using survey respondents as our population of interest, we overcome two constraints that have limited existing efforts to evaluate MRP. First, it is usually quite difficult to obtain measures of actual state opinion or congressional district opinion (the baseline against which MRP estimates have traditionally been

[1] An exception is Warshaw and Rodden (2012) who also compare MRP estimates of support for minimum wage laws and stem cell research to the results of ballot measures on these topics in a non-random sample of four states.

compared). Our approach makes this much easier, providing us with measures of the true opinion of the population of interest across a very large number of issues. Second, by creating poststratification weights from surveys, we can evaluate MRP models that include individual level predictors that are not available from the Census. This allows us to consider a variety of hitherto untested response models.

In evaluating the predictive accuracy of MRP, we vary the complexity of and variables included in the response model. Doing so not only enables us to speak to the performance of MRP across wide range of policy areas and political attitudes, but also allows us to make recommendations as to the type of response models that ought to be employed and whether significant gains can be realized by tailoring the response model to the specific issue area in question (e.g., should researchers use different models for economic and social issues?). Along the way, we also identify the measures of model fit and performance that best correlate to the accuracy of MRP estimates. The totality of our results will enable us to develop a set of guidelines for implementing MRP properly as well as a set of diagnostics for identifying instances where MRP is appropriate and instances where its use may be problematic.

So far, we have completed a trial run, and we are engaged in producing a far wider assessment of MRP.

2 MRP Overview

MRP allows researchers to simulate subnational public opinion (by states, legislative districts, etc) using national-level survey data. Simulation approaches to opinion estimation have a long history in political science (e.g., Pool, Abelson, and Popkin 1965, and, for critiques, see Weber, et al. 1972, Seidman 1975, and Erikson, Wright, and McIver 1993). MRP however, has important advantages over prior efforts. For example, older applications typically modeled opinion using only demographic variables. In contrast, MRP also includes geographic variables, recognizing that even after controlling for a variety of demographic influences, the state and region of the country in which people live are important predictors of their core political attitudes as well as their opinions on a variety of policy debates (Erikson, Wright, and McIver 1993; Gelman et al 2008). MRP is also far more sophisticated than older approaches in the way that it models individual survey responses, using Bayesian statistics and multilevel modeling. Doing so improves the accuracy of estimates of the effects of individual- and state-level predictors (Gelman and Little 1997). The multilevel model also allows researchers to use many more respondent types than did classical methods.

MRP proceeds in two stages. In the first stage, a multilevel model of individual survey response is estimated, with opinion modeled as a function of a respondent's demographic and geographic characteristics. The multilevel model partially pools respondents across states to an extent determined by the data. Individual responses are explicitly modeled as nested within states, so that all individuals in the survey, no matter their location, yield information about demographic patterns that can be applied to all estimates; state effects capture residual differences. State-level effects can themselves be modeled using additional state-level predictors such as region or aggregate state demographics. The results of this modeling stage are used to generate an estimate of opinion for each demographic-geographic type of voter. "Typical" state-level models estimate the preferences of well over 4,000 demographic-geographic types (Lax and Phillips 2012), while the key work on congressional district modeling estimated over preferences for over 17,000 types (Warshaw and

Rodden 2012), with roughly the same types per geographic sub-unit as other work.

The second step of MRP is poststratification: the opinion estimates for each demographic-geographic respondent type are weighted (poststratified) by the percentages of each type in the actual population of each state. This allows researchers to estimate the percentage of respondents within each state who hold a particular attitude or policy preference. Post-stratification has typically been done using population frequencies obtained from either the Public Use Micro Data Samples supplied by the Census Bureau or similar data.

The potential advantages of MRP are many. First (and most importantly), it should allow researchers to generate accurate opinion estimates by state or legislative district using relatively small amounts of survey data. This is possible because (as we note above) the multilevel model used in stage one "borrows strength" by partially pooling respond types across geographic units. Indeed, by borrowing strength across all observations, each and every stratification cell does not need to be populated with survey respondents. Second, through the process of poststratification, MRP can potentially correct for differences between a survey sample and the actual population. This can help alleviate problems such as survey non-response and concerns over sampling techniques. Finally, MRP can generate opinion estimates for constituencies not included in the surveys employed in the stage-one model (assuming, of course, that constituency-level census is available). This is particularly useful since many surveys intentionally do not sample Alaska and Hawaii and smaller population states, such as New Hampshire, Vermont, and Wyoming, are sometimes unintentionally unsampled.

3 What Do We Know and How Do We Know It?

The handful of studies that have evaluated MRP largely confirm its potential and demonstrate that it generally outperforms its primary alternative—disaggregation. The first work to evaluate MRP is that of Park, Gelman, and Bafumi (2004, 2006), who used MRP to estimate state-level support for President George H.W. Bush during the 1988 and 1992 presidential elections. Their data consisted of responses to CBS News/New York Times national polls conducted the week before each presidential election and they model survey response as a function of a respondent's gender, ethnicity, age, education, and state. With modest sample sizes (2,193 respondents in 1988 and 4,650 respondents in 1992) Park, Gelman, and Bafumi come close to predicting actual state-level election results—their MRP estimates yield a mean absolute error of approximately 4% . They find that the partial pooling that is utilized in MRP produces more accurate estimates of election outcomes than do techniques that employ either full pooling (which is similar to old-style simulation approaches) and techniques that employ no pooling (which is similar to disaggregation).

Lax and Phillips (2009a) explicitly compare MRP estimates of state public opinion to those obtained via disaggregation. They begin by merging a large set of national surveys on support for same-sex marriage, creating a dataset of approximately 28,000 respondents. They then randomly split the data, using half to define "true" state opinion and some portion of the remaining data to generate opinion estimates, either by applying MRP or disaggregation. Using a similar response model as Park, Gelman, and Bafumi, Lax and Phillips find that when compared to baseline measures of true opinion, MRP notably outperforms disaggregation, yielding smaller errors, higher correlations, and more reliable estimates.[2] Importantly, opinion estimates obtained via MRP appear to be quite accurate even when using samples with as few as 1,400 survey respondents. The estimates obtained from these small samples correlate with "true" opinion at 0.74 and possess a mean absolute error of 4.9%. They also show that while the accuracy of MRP improves as sample size increases, such gains are relatively modest.[3] To further validate their findings, Lax and Phillips compare MRP estimates of state-level support for same-sex marriage to actual state polls, finding (once again) that MRP does quite well. Using a single slightly above average-sized national poll, they produce estimates of opinion that correlate with state polls at a very high level 0.73 and have an absolute mean error of 6%.

The most recent evaluation of MRP was conducted by Warshaw and Rodden (2012). Rather than consider the predictive accuracy of MRP at the state level, they evalute it's ability to generate accurate estimates of public opinion by congressional and state senate districts. Warshaw and Rodden begin by combining national surveys to obtain a dataset of 100,000 respondents. Using the same split-sample research design as Lax and Phillips, they compare to opinion estimates obtained via disaggregation and MRP to "true" opinion across six issues—same-sex marriage, abortion, environmental protection, minimum wage, social security privatization, and federal funding for stem cell research.[4] Consistent with

[2]Lax and Phillips do not include an interaction between age and education, but add (as state-level predictors) region and the share of the population that consists of religious conservatives.

[3]For example, increasing the sample size from 1,400 to 14,000 only decreases the mean absolute error from 4.9% to 3.8%.

[4]The survey response model used by Warshaw and Rodden (stage one of MRP) includes the same individual level predictors as used by Lax and Phillips. However, Warshaw and Rodden utilize many more district level predictors—the district's average income, the percent of a the district's residents that live in urban areas, the percentage of the districts residents that are military veterans, and the percentage of couples in each district that are same-sex couples. They also employ different post stratification data, relying on the Census Factfinder as opposed to the 1% or 5% Public Use Microdata Sample.

prior work, they find strong evidence that MRP outperforms disaggregation. For external validation, Warshaw and Rodden examine how well their MRP estimates predict district-level voting on state ballot measures that closely correspond to three of the six issues included in their study. While they can only conduct this analysis for small non-random sample of states, the correlation between MRP estimates and the actual vote is fairly high. Warshaw and Rodden ultimately conclude that MRP generates reliable estimates of congressional district opinion using sample sizes of just 2,500 and yields reliable estimates for state senate districts with a national sample of 5,000.

Overall, existing work presents a favorable evaluation of the potential of MRP, indicating that it can generate accurate measures of public preferences using a modestly-sized national sample of survey respondents and a fairly simple survey response model. This technique, given the large number of national surveys on which it can potentially be used to estimate subnational opinion, may greatly expand the political phenomena than can be systematically studied. Indeed, teams of researchers have already employed MRP to tackle a variety of substantive questions that were, given prior technology, thought to be beyond the bounds of empirical inquiry. However, one might worry that these substantive applications of MRP are outpacing our knowledge of the methodology. Existing evaluative efforts have been limited to a handful of issues and leave unanswered important questions abut the performance of MRP.

4 Further Assessing MRP

In this paper, we consider the predictive accuracy of MRP across a very large set of issues and political attitudes. In doing so, we seek metrics that will allow researchers to identify whether a particular set of estimates are likely to be accurate. We also consider a variety of steps that might be taken to maximize the performance of MRP. We conduct our analysis at the state level, though we see little reason to believe our findings cannot be applied to research that employs MRP to estimate preferences by other subnational geographic units.

4.1 Data

To evaluate MRP we utilize data from the 2010 Cooperative Congressional Election Study (CCES). The 2010 CCES survey contains a large national sample of just under 40,000 respondents, with a large number from each state (ranging from a high of nearly 5,000 respondents from California to a low of 80 from Wyoming). Using this survey we have answers to 50 distinct questions that ask respondents about their political attitudes and issue-specific preferences. We recode each survey question as necessary so that dependent variable (opinion) is measured dichotomously. For each respondent we have a wealth of demographic and geographic information. These data will be used (to varying extents) in our survey response models and to generate our poststratification files (remember, unlike most studies we will not be using the Census data for poststratification, but will be treating CCES respondents as our population).

4.2 Modeling Individual Responses

MRP begins by modeling individual survey responses (opinions) as a function of both demographic and geographic variables. This allows researchers to create predictions for each respondent type. Rather than using "unmodeled" or "fixed" effects, MRP uses "random" or "modeled" effects, at least for some predictors (see Gelman and Hill 2007, 244-8). That is, it assumes that the effects within a group of variables are related to each other by their hierarchical or grouping structure. For data with hierarchical structure (e.g., individuals within states), multilevel modeling is generally an improvement over classical regression—indeed, classical regression is a special case of multilevel models in which the degree to which the data is pooled across subgroups is set to either one extreme or the other (complete pooling or no polling) by arbitrary assumption (see Gelman and Hill 2007, 254-8). The general principal behind this type of modeling is that it is a "compromise between pooled and unpooled estimates, with the relative weights determined by the sample size in the group and the variation within and between groups." A multilevel model pools group-level parameters towards their mean, with greater pooling when group-level variance is small and more smoothing for less-populated groups. The degree of pooling emerges from the data, with similarities and differences across groups estimated endogenously. A additional advantage of this modeling structure is that is allows researchers to estimate preferences by many more demographic-geographic categories, producing more accurate poststratification.

We estimate several alternative stage-one models for each CCES survey question used. However, we begin with what we refer to as the BASELINE model. This baseline is similar to or slightly simpler than MRP models used throughout the literature.

In this model, we treat the probability of a "yes" response for any type of individual as a function of the demographic and geographic characteristics that define those types (each type gets its own cell c, with indexes j, k, l, m, and s for race-gender combination, age category, education category, and state respectively). The demographic categories we employ are as follows: gender (male or female), race (black, Hispanic, white, and other), age (18-29, 30-39, 40-49, 50-59, 60-69, and 70+), and education (less than a high school education, high school graduate, some college, college graduate, post graduate degree).[5]

$$\Pr(y_c = 1) = \text{logit}^{-1}(\beta^0 + \alpha^{gender}_{j[c]} + \alpha^{race}_{k[c]} + \alpha^{age}_{l[c]} + \alpha^{edu}_{m[c]} + \alpha^{state}_{s[c]}) \tag{1}$$

The terms after the intercept are random/modeled effects for the various groups of respondents:

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

$$\alpha^{race}_{k} \sim N(0, \sigma^2_{race}), \text{ for } k = 1, ..., 4 \tag{3}$$

$$\alpha^{age}_{l} \sim N(0, \sigma^2_{age}), \text{ for } l = 1, ..., 6 \tag{4}$$

$$\alpha^{edu}_{m} \sim N(0, \sigma^2_{edu}), \text{ for } m = 1, ..., 5 \tag{5}$$

$$\alpha^{state}_{s} \sim N(0, \sigma^2_{state}), \text{ for } s = 1, ..., 50 \tag{6}$$

We also evaluate several more complicated response models. We increase model complexity through some combination of:

[5]Sometimes the response model does not completely converge or gives a false convergence. Doing any single MRP run, one would extend the number of iterations or simplify the model. Here, we assumed a naive run of MRP in our simulations, leaving in the faulty runs so that our results are a lower bound on MRP accuracy.

1. Adding additional demographic information in the form of my nuanced cell typographies (that is, by splitting our cells into subcells).

$$\alpha_r^{religion} \sim N(0, \sigma^2_{religion}), \text{ for } r = 1, ..., 7 \tag{7}$$

$$\alpha_p^{income} \sim N(0, \sigma^2_{income}), \text{ for } p = 1, ..., 14 \tag{8}$$

The religion categories are: atheist, born-again Protestant, mainline Protestant, Catholic, Jewish, and other. The income categories use the following breakpoints in thousands of dollars: 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 100, 120, and 150.

2. We make further use of the demographic and geographic information in the baseline model by including interactions between existing categories (these do not create any new cell types, but rather allow for more nuanced estimation of probabilities within existing cells). The PAIRED INTERACTION setup includes an interaction between age and eduction as well as an interaction between gender and race.

$$\alpha_{j,k}^{gender,race} \sim N(0, \sigma^2_{gender,race}), \text{ for } j = 1, 2; k = 1, ..., 4 \tag{9}$$

$$\alpha_{l,m}^{age,edu} \sim N(0, \sigma^2_{age,edu}), \text{ for } l = 1, ..., 6; m = 1, ...5 \tag{10}$$

The QUAD INTERACTION setup adds to the above the four-way interaction.

$$\alpha_{j,k,l,m}^{gender,race,age,edu} \sim N(0, \sigma^2_{gender,race,age,edu}), \text{ for } j = 1, 2; k = 1, ..., 4; l = 1, ..., 6; m = 1, ..., 5 \tag{11}$$

Finally, there is the GEOGRAPHIC INTERACTION setup. Here, we interact state and race *or* state with all four demographic descriptors.

$$\alpha_{s,k}^{state,race} \sim N(0, \sigma^2_{state,race}), \text{ for } s = 1, ..., 50; k = 1, ..., 4 \tag{12}$$

$$\text{or } \alpha_{s,j,k,l,m}^{state,gender,race,age,edu} \sim N(0, \sigma^2_{state,gender,race,age,edu}), \tag{13}$$

$$\text{for } s = 1, ..., 50; j = 1, 2; k = 1, ..., 4; l = 1, ..., 6; m = 1, ..., 5$$

3. The GEOGRAPHIC PREDICTOR setup. We can potential improve on the above by adding group-level predictors. Geographic predictors fall into two types. The first adds a hierarchical level to organize the state random effects into regions (that is we add a region random effect).

$$\alpha_q^{region} \sim N(0, \sigma^2_{region}), \text{ for } q = 1, ..., 4 \tag{14}$$

The second brings in additional information, to form a substantive group-level predictor. The numeric value of this predictor is determined by the level of the relevant random effect. For example, using a state level ideology score (as per Erikson, Wright and McIver 1993)[6] does not create any new cells (or types), but rather is a function of the cell as already defined: all cells associated with New York get the New York ideology score. If we were using both region and ideology, the state-level formula would

[6]We tweaked, very slightly, their scores by imputing scores for HI, AK, and NV (they often drop all three of those, the first for lack of the data they use, the last because of their strange result for NV). We impute those values.

be as follows:

$$\alpha_s^{state} \sim N(\alpha_{q[s]}^{region} + \beta^{ideology} \cdot ideology_s, \sigma_{state}^2), \text{ for } s = 1, ..., 50 \tag{15}$$

We use different combinations of region, ideology, presidential vote (in the form of Obama's vote margin over McCain in the 2008 election), percent religious conservative (i.e., Mormon and evangelical Protestant), and a *DPSP* (see note).[7]

4. The Demographic Linear Predictor setup. Because age, education, and income are ordered categories, we create a linearized predictor for each based on the level within the category. For example, in addition to using random effects for the six age categories, we can add an ordinal variable $z.age$ with values ranging from 1 to 6, treated as a continuous predictor for the age random effects. We rescale these predictors by

[7] DPSP stands for demographically purged state predictor. The state-level predictors we use, such as presidential vote or ideology, are usually things we think are correlated with the actual state-level true values—which are, after all, connected strongly to demographics—rather than being directly correlated to the state level random effects, which are the corrections to a purely demographic model. These intercept shifts are to be the corrections to whatever the demographic and other variables would produce. Therefore, it might be odd to use a model for them that assumes that the linear relationship between ultimate state opinion and presidential vote is the same as the linear relationship between state-level corrections and presidential vote. We constructed the DPSP we use herein from the full set of 39 survey sets in Lax and Phillips (2009a). Very simply, this DPSP is the state random effects vector found by excluding all state level predictors and running a somewhat standard model otherwise. This state random effects vector (from a model with 200K observations across many survey questions) is the average desired state level intercept shift across a wide set of policies. We find that DPSP does at least weakly better than other state predictors most of the time and, indeed, when used, reduces the variation in state random effects in MRP applications, showing that more state variation (at the level of corrections to demographic effects) is explained by DPSP than other state-level predictors. DPSP values are included in the MRP package and are shown in the Appendix.

centering to mean zero and dividing by two standard deviations.[8]

$$z.age = rescale(levels(age)) \tag{16}$$

$$z.edu = rescale(levels(edu)) \tag{17}$$

$$z.income = rescale(levels(income)) \tag{18}$$

Then, we substitute for the above, the following models for the specific random effects.

$$\alpha_l^{age} \sim N(\beta^{age} \cdot z.age_l, \sigma^2_{age}), \text{ for } l = 1, ..., 6 \tag{19}$$

$$\alpha_m^{edu} \sim N(\beta^{edu} \cdot z.edu_m, \sigma^2_{edu}), \text{ for } m = 1, ..., 5 \tag{20}$$

$$\alpha_p^{income} \sim N(\beta^{income} \cdot z.incomep, \sigma^2_{income}), \text{ for } p = 1, ..., 14 \tag{21}$$

5. Finally, there are two sub baseline variants that allow us to assess the contributions of even the standard components of the baseline setup. The Race Only variant leaves out age, gender, and education (as well as potential demographics such as income and religion). The No Demographics variants leave out even race. Such variants usually rely on geographic predictors to make up for the loss of demographic information

We invoke the above setups in a variety of different combinations.

[8]One could also use a substantive linear predictor such as mean within each category or some other predictor of the likely effect of the categories, just as one uses substantive predictors for states such as presidential vote and not just more basic predictors such as region.

4.3 Poststratification

For each combination of individual demographic and geographic values that define a cell c, the results from the multilevel model of response are used to make a prediction of public opinion. Specifically, θ_c is the inverse logit given the relevant predictors and their estimated coefficients. The next stage is poststratification, in which our estimates for each respondent demographic-geographic type must be weighted by the percentages of each type in the state population. Again, we assume the state population to be the set of CCES survey respondents from that state.

In the baseline model, we have 50 states with 240 demographic types in each. This yields 12,000 possible combinations of demographic and state values, ranging from "White," "Male," "Age 18-29," "Not high school graduate," in "Alabama," to "Other," "Female," "Age 70+," "Graduate Degree," in "Wyoming." Each cell c is assigned the relevant population frequency N_c. The prediction in each cell, θ_c, needs to be weighted by these population frequencies of that cell. For each state, over each cell c in state s, the predicted affirmative response percentage is this weighted average :

$$y^{\text{MRP}}_{\text{state } s} = \frac{\sum_{c \in s} N_c \theta_c}{\sum_{c \in s} N_c} \tag{22}$$

4.4 Simulations

For each set of runs of our simulation takes a sample of 1,000 responses (from the full set of almost 40,000) on a given question. A set of runs consists of different MRP models applied to the same sample (so that we fix the sample and vary the particular MRP variant applied to it). To do each MRP, we use the newly available MRP package in R (for most current version, use the GITHUB website), which greatly simplifies the multilevel modeling and poststratification steps, in addition to providing a framework for adding benchmarks we will develop from our results therein.[9]

We take 10 samples for each question so that the final number of runs will be Num(questions used) $\times$ 10 $\times$ Num(MRP model variants).[10] For each run we save the vector of MRP estimates for the sample, disaggregated state percentages within sample, MRP and disaggregation for the full CCES (the latter of which defines "true"). We calculate the various metrics we discuss in our results section such as the absolute error between MRP and true and the correlation of the MRP vector to the true vector.

Our plan is to replicate those analyses that show to be promising from our starting set of 2010 CCES questions for a larger set of CCES surveys and others of similar size.

4.5 Results

We use four metrics to measure the success of MRP. All of our current results are summarized in the Appendix tables. The first is the error between the predicted affirmative response percentage (a state's MRP estimate) and the actual affirmative response percentage ("true" state opinion obtained from the full CCES). For robustness, we consider the mean error, median error, and percentage reduction in error across states and simulation runs. Since

[9]Whereas older implementations of MRP ran the response model at the level of the individual, the MRP package reformats the response data into cells (individual types) from the start, where a cell is a complete statement of type. The distinction is innocuous (logit in R simply takes in the number of Yes responses and number of No responses for each cell) but does focus attention properly on the cell level (which is the ultimate level of analysis in the poststratification stage) and simplifies internal MRP processes for standard dichotomous response situations.

[10]Our computers are continuing to process these simulations so not all runs are completed as yet.

these metrics produce very similar results, we focus in the text on the median absolute error across states for a given run (this should reflect the error for an average sized state). To aggregate errors by model variant we take the mean across all runs. The second major metric is the correlation between a set of 50 state MRP estimates and 50 true values. Both of these approaches, which we will refer to as *error* and *correlation*, have been used in previous MRP assessments. These metrics, though similar, are not equivalent.

The third metric that we employ is *congruence*. The substantive literature on government responsiveness increasingly asks whether a policy or roll call vote matches the state or district opinion majority (cf., Lax and Phillips 2009a, 2012; Matsusaka 2010). To make such a determination, scholars need to know the placement of the median constituent (e.g., does she favor or oppose a given policy). Thus, we measure the frequency with which MRP correctly identifies the preferences of this individual. Our fourth metric, *shrinkage*, compares the standard deviation of MRP estimates to the standard deviation of true opinion. This allows us to consider the extent to which MRP reduces cross-state variation in opinion as a result of partial pooling.

We begin by discussing the results of the baseline model. These results demonstrate that even with a small national sample of 1,000 survey respondents and a fairly simple demographic-geographic response model, MRP performs quite well. On average, the mean correlation between true state opinion and our estimates is 0.46, with a quite modest mean error of only 3 percentage points. Furthermore, MRP was able to correctly identify the majority side in 93% of all simulation runs. The low error and high congruence of our MRP estimates is not the result of using survey questions for which there is little cross-state variation in true opinion, though there is a positive correlation between the spread of state opinion and the error of our estimates. This is shown in the graph on the left side of Figure 1 which plots on the x-axis the spread of true state opinion (measured as a standard deviation) and the mean absolute error of our opinion estimates. As one can see, errors tend to be low—just under 4 percentage points—even when the standard deviation of true opinion is high. However, it clear that as the standard deviation of true opinion increases, the accuracy of MRP estimates decline (though only very modestly). Interestingly, the correlation between MRP estimates and true opinion has no clear relationship to the standard deviation of true opinion. This is shown in the graph on the right-hand side of Figure 1.

While MRP produces reasonably accurate estimates of opinion across a range of issues, these estimates do not have as much cross-state variation as true opinion. The standard deviation of the MRP state estimates is 51 percent of the standard deviation of the true state values. This shrinkage can be seen in Figure 2, which plots the standard deviation of true opinion on the x-axis and standard deviation of the MRP estimates on the y-axis. The dark gray line is a lowess curve, showing the relationship between estimated and true opinion; the dashed line is the 45 degree line. The difference between the 45-degree line and the lowess curve is the amount of shrinkage in the MRP estimates. Note that the lowess curve is always below the 45-degree line, indicating that MRP estimates (using the baseline mode) are consistently underestimating the amount of cross-state variation in opinion. As cross state variation in true opinion grows, so the does the extent to which MRP underestimates variation. This finding suggests that a basic MRP model may be over-pooling opinion. We would expect this to go down as our sample size increases and it does. If we estimate the baseline model using our full dataset (approximately 30,000 observations per survey question), the MRP state estimates go from being 51 percent of the standard deviation of the true state values to 79 percent.

Can the results of the baseline model be improved upon? To answer this question, we generate additional opinion estimates using the model variants presented in Section 4.2. Here, we briefly discuss the manner in which these variants affect the accuracy of estimates. We begin by considering the use of additional demographic information in the form of more nuanced cell typographies (that is, by splitting our cells into subcells). Specifically, we add to the baseline model religion and income as predictors, estimating some models with just one of these additional predictors and some models with both. Ultimately, however, adding religion and income (either by themselves or in tandem) results in at best a very slight improvement in the accuracy of MRP. When these predictors are added to the baseline model, religion (on average) reduces mean error by a third of a percentage point, while income (on average) reduces error by only a few hundredths of a point. There is also some improvement in the correlation between MRP estimates and true state opinion when religion is added, but this improvement is also quite modest—0.09 on a scale from 0 to 1. Religion and income make no difference when added to model variants other than the baseline.

Next, we consider interactions to the baseline model. We estimate models that utilize a paired interaction set up (interactions between age and eduction and between gender and race), models that utilize a quad interaction setup (a four-way interaction between race, education, age, and gender), and models that employ the geographic interaction setup (interactions between state and race or state with all four demographic descriptors). On average, we find that the inclusion of some or all of these terms results in no improvements to the accuracy of MRP estimates, even when these terms are added to model variants other than the baseline. This is true holding constant the simulation run as well as the particular survey question asked. To be sure, interactions in some runs and for some questions modestly help the performance MRP, but other times they hurt the accuracy of estimates. The range of gains and losses to median absolute error, when they occur, is only about a third of a percentage point. In subsequent analyses we will seek to identify the conditions under which each occurs. It is important to reiterate, that, that on average, there are no benefits to the use of interactions.

The next approach we consider is the geographic predictor setup, in which we evaluate the benefits of adding a hierarchical level to organize the state random effects into regions (that is we add a region random effect) as well as the benefits of adding a substantive group-level predictor, such as state-level ideology, the share of the population that are religious conservatives, the share of the state electorate who voted for President Obama in the prior presidential election, and our measure DSPS. Our results indicate that there is little to be gained by including region as a random effect. However, utilizing a substantive group level predictor notably enhances the performance of MRP. Figure 3 demonstrates this. The results reported in the figure use the baseline model, but all presidential vote share as a substantive state-level predictor. The improvements of this predictor can be seen by comparing these new results to those reported in Figure 1. The addition of presidential vote reduces the mean error of the MRP estimates. Error in the baseline model averaged 3 percentage points, with a range of approximately 1.5 to 5 percentage points (depending upon the spread of true state opinion). After including presidential vote share, the mean error falls 2.8 percentage points, with a range of 1.5 to 3 points. The correlation between MRP estimates and true opinion also increases from 0.46 to 0.56. Note that even though the average effect of including presidential vote is on average positive, there are some runs in which adding this substantive sate level predictor hurts the accuracy of estimates (though when it hurts, the consequences are quite small).[11] The inclusion of a presidential vote share also reduces the amount of shrinkage in MRP estimates. The standard deviation of the MRP state estimates is now 85 percent of the standard deviation of the true state values. The improvement can be seen quite nicely by comparing Figure 4 with Figure 1.

Each of the four substantive group-level predictors (when added individually to the basic model) improves on average the accuracy and correlation of MRP estimates to true opinion and reduces the amount of shrinkage in the estimates. We find little difference in the degree of improvement across each of the four—in other words, it didn't really matter which of the four we used as long as one was included. Using one state level predictor is on average better than using none (it reduces error by an average of 0.2 percentage points and improves correlation by 0.06). However, using two actually increases error by a small amount relative to using just one substantive state level and also reduces correlation between MRP estimates and true opinion. We can also ask how often across all our simulation runs adding the second state level predictor helps. We find that doing so helps half the time hurts half the time. To be sure the results at this point do not include uncertainty around our estimates—and using multiple state level variables has been shown in our trial runs to increase uncertainty around estimates.

We also consider models that utilize a demographic linear predict setup. In these, we create a linearized predictor for age, education, and income. For example, in addition to using random effects for the six age categories, we can add an ordinal variable *z.age* with

[11] In 57% of the runs the inclusion of presidential vote reduced error. The average gain from using presidential vote is roughly twice the potential loss. This is true for each of the other substantive state-level predictors as well.

values ranging from 1 to 6, treated as a continuous predictor for the age random effects. Doing so does not on average improve the the performance of MRP. Finally, we consider two variants that are less complex than the baseline model. One of these drops all demographic predictors with the exception of race and the second leaves out even race, relying only on state random effects. Unsurprisingly, neither of these models performs nearly as well as the baseline model.

5 Discussion

The results of our analysis demonstrate that MRP can produce reasonably accurate estimates of state-level public opinion using small sample sizes (1,000 survey respondents, even fewer than previously suggested) and fairly simple demographic-geographic response models. The accuracy of MRP estimates that we report here is consistent with what has been found in the existing literature, but across a wider range of survey questions and with particular attention to assessments across MRP variants rather than against other methods. We find that MRP does slightly better when the spread of true state opinion is smaller (while slightly worse for higher spreads, so too would any method be), and we find that MRP has a tendency to shrink cross-state variation in opinion. This is particularly true when sample sizes are small and state level predictors are not used. We also find that tweaks to the baseline model generate modest gains at best and in some cases may actually reduce the performance of MRP.

Our next steps include determining when to expect these tweaks to help or hurt; establishing what average error and correlation to expect based on those factors observable by a researcher without access to “truth”; test how MRP estimates perform when used as predictors of policy choice; test how well MRP performs at estimating demographic group opinion within states; establish diagnostics, benchmarks, and indicators of MRP success; evaluate how our recommendations and findings vary by sample size; and extend our assessment to an even larger pool of questions.

6 Preliminary Recommendations

We currently recommend MRPers follow the pointers and keep in mind the comments below, all of which are based on our runs on 1000 observations at a time, along with previous work on and experience with MRP. We should note that none of our calculations, at this point, take into account the reliability and noise in our measure of “true” opinion—we thus understate MRP performance. Ongoing work will correct for this and extend our assessment significantly.

1. Use a substantive group-level predictor for state. Using more than one is unlikely to be helpful, especially if noisily estimated. The choice is not dispositive though the DPSP variable we recommend is weakly best in our current results.

2. Interactions between individual cell-level predictors are not necessary. Deeper interactions (say, four-way interactions) do nothing for small sample.

3. Adding additional individual types (by religious or income categories) does not improve performance on average.

4. Adding continuous predictors for demographic group-level variables does not improve

performance.

5. Until further diagnostics are provide, and if our recommendations are followed, expect median absolute errors across states to be approximately 2.7 points (and likely in the range 1.4 to 5.0 points) and expect correlation to true state values to be approximately .57. Congruence correct is on average 94% of codings (and those concerned with error in congruence codings can use degree of incongruence instead).

6. Shrinkage of inter-state standard deviations for a sample size of 1000 is approximately .78.

7. Take into account uncertainty around your mrp estimates in substantive work (the package will soon do so and an example of this is shown in Lax, Kastellec, Malecki, and Phillips 2013).

References

Ansolabehere, Stephen, James M. Snyder, Jr., and Charles Stewart III. 2001. "Candidate Positioning in U.S. House Elections." *American Journal of Political Science* 45(1): 136-49.

Bates, Douglas. 2005."Fitting Linear Models in R Using the lme4 Package." *R News* 5(1): 27-30.

Caldarone, Richard P., Brandice Canes-Wrone, and Tom S. Clark. 2009. "Partisan labels and Democratic Accountability: An Analysis of State Supreme Court Abortion Decisions," *The Journal of Politics* 70: 560-73

Canes-Wrone, Brandice, John F. Cogan, and David W. Brady. 2002. "Out of Step, Out of Office: Electoral Accountability and House Members' Voting." *American Political Science Review* 96(1): 127-40.

Clinton, Joshua. 2006. "Representation in Congress: Constituents and Roll Calls in the 106th House." *The Journal of Politics* 68(2): 397-409.

Cohen, Jeffrey E. 2006. "Conclusions: Where Have We Been, Where Should We Go?" In *Public Opinion in State Politics*, ed. Jeffrey E. Cohen. Stanford, Calif.: Stanford University Press.

Erikson, Robert S. 1976. "The Relationship between Public Opinion and State Policy: A New Look Based on Some Forgotten Data." *American Journal of Political Science* 20(1): 25-36.

Erikson, Robert S. and Gerald C. Wright. 1980. "Policy Representation of Constituency Interests." *Political Behavior* 2(1): 91-06.

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. 2007. "Struggles with Survey Weighting and Regression Modeling." *Statistical Science* 22(2): 153-64.

Gelman, Andrew, and Jennifer Hill. 2007. *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 Methodology* 23(2): 127-35.

Gelman, Andrew, David Park, Boris Shor, Joseph Bafumi, and Jeronimo Cortina. 2008. *Red State, Blue State, Rich State, Poor State: Why Americans Vote the Way They Do*. Princeton, N.J.: Princeton University Press.

Jackson, John T. and David C. King. 1989. "Public Goods, Private Interests, and Representation." *American Political Science Review* 83(4): 1,143-64.

Kalt, Joseph P. and Mark A. Zupan. 1984. "Capture and Ideology in the Economic Theory of Politics." *American Economic Review* 74: 279-300.

Kastellec, Jonathan, Jeffrey Lax, and Justin Phillips. 2010. "Public Opinion and Senate Confirmation of Supreme Court Nominees." *The Journal of Politics* 72(3): 76784.

Kastellec, Lax, Malecki, and Phillips. 2013. Working paper - available upon request. "Distorting the Electoral Connection? Partisan Representation in Confirmation Politics."

Krehbiel, Keith. 1993. "Constituency Characteristics and Legislative Preferences." *Public Choice* 76(1): 21-37.

John G. Matsusaka, "Popular Control of Public Policy: A Quantitative Approach," *Quarterly Journal of Political Science* 5 (2010): 133167.

Miller, Warren E. and Donald E. Stokes. 1963. "Constituency Influence in Congress." *American Political Science Review* 57(1): 45-56.

Monroe, Alan D. 1998. "Public Opinion and Policy,1980-1993." *Public Opinion Quarterly* 62: 6-28.

Lax, Jeffrey R., and Justin H. Phillips. 2009a. "How Should We Estimate Public Opinion in the States?" *American Journal of Political Science* 53(1): 10721.

Lax, Jeffrey R., and Justin H. Phillips. 2009b. "Public Opinion and Policy Responsiveness: Gay Rights in the States." *American Political Science Review* 103(3): 36785.

Lax, Jeffrey R., and Justin H. Phillips. 2012. "The Democratic Deficit in the States." *American Journal of Political Science* 56(1): 14866.

Levitt, Steven D. 1996. "How Do Senator's Vote? Disentangling the Role of Voter Preferences, Party Affiliation, and Senator Ideology." *American Economic Review* 86: 425-41.

Kastellec, Jonathan P., Jeffrey R. Lax, and Justin H. Phillips. 2008. "Public Opinion and Senate Confirmation of Supreme Court Nominees." Presented at the Annual Meeting of the American Political Science Association (Boston).

Kastellec, Jonathan P., and Eduardo Leoni. 2007. "Using Graphs Instead of Tables in Political Science." *Perspectives on Politics* 5(4): 755-71.

Norrander, Barbara. 2001. "Measuring State Public Opinion with the Senate National Election Study." *State Politics and Policy Quarterly* 1(1): 111-25.

Norrander, Barbara. 2007. "Choosing Among Indicators of State Public Opinion." *State Politics and Policy Quarterly* 7(2): 111.

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

Page, Benjamin I. 1994. "Democratic Responsiveness? Untangling the Links Between Public Opinion and Policy." *P.S.: Political Science and Politics* 27: 25-9.

Page, Benjamin I., and Robert Y. Shapiro. 1983. "The Effects of Public Opinion on Policy." *The American Political Science Review* 77(1): 175-190.

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

Park, David K., Andrew Gelman, and Joseph Bafumi. 2006. "State Level Opinions from National Surveys: Poststratification using Multilevel Logistic Regression," in *Public Opinion in State Politics* (ed. Jeffrey E. Cohen). Stanford, CA: Stanford University Press.

Pool, Ithiel de Sola, Robert Abelson, and Samuel L. Popkin. 1965. *Candidates, Issues, and Strategies*. Cambridge, MA: M.I.T. Press.

Seidman, David. 1975. "Simulation of Public Opinion: A Caveat." *The Public Opinion Quarterly* 39(3): 331-42.

Selb, Peter and Simon Munzert. 2011. "Estimating Constituency Preferences from Sparse Survey Data Using Auxiliary Geographic Information." *Political Analysis* 19(3): 455-70.

Warshaw, Christopher and Jonathan Rodden. 2012. "How Should We Estimate District-Level Public Opinion on Individual Issues?" *The Journal of Politics* 74(1): 203-19.

Weber, Ronald E., Anne H. Hopkins, Michael L. Mezey, and Frank J. Munger. 1972. "Computer Simulation of State Electorates." *Public Opinion Quarterly* 36(4): 549-565.

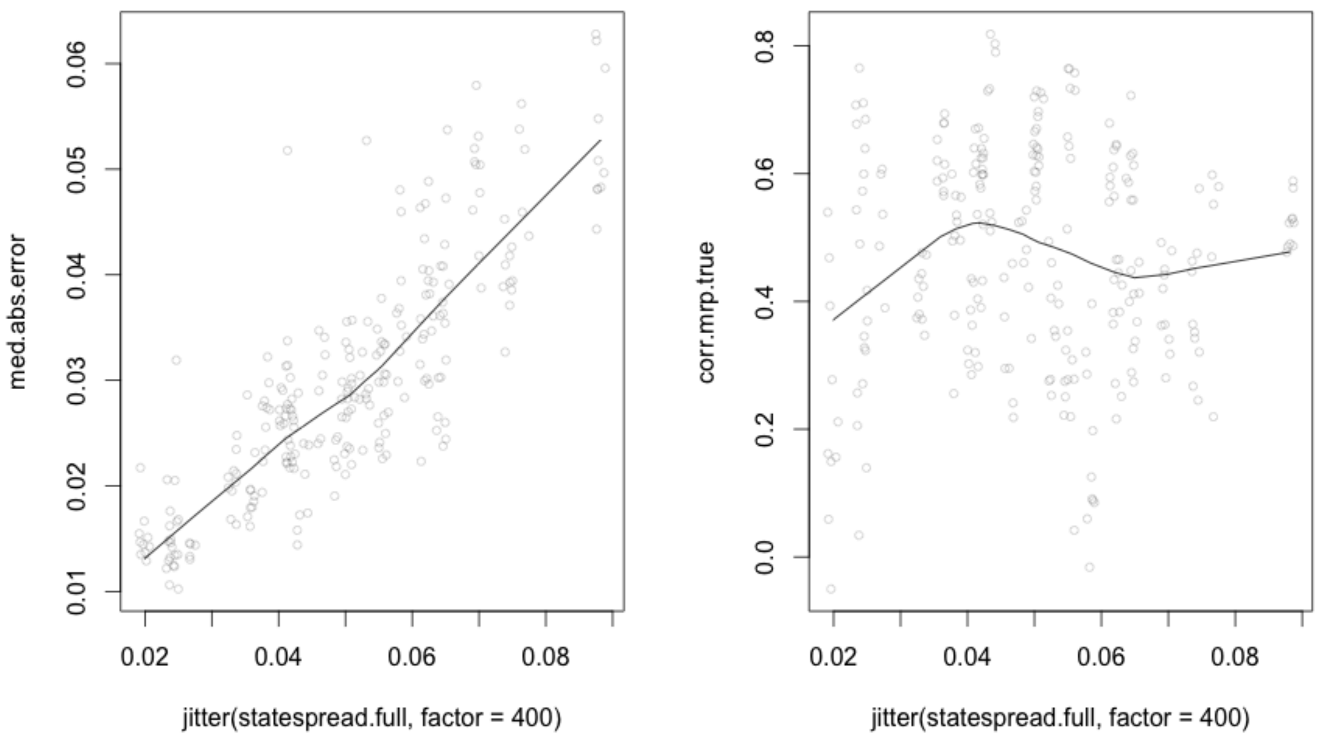

Figure 1: *Baseline Model: Error and Correlation by the Spread of True State Opinion.* This figures shows how the accuracy of MRP varies as the spread of true state opinion (measured as a standard deviation) grows. The x-axis in both graphs is the standard deviation of true state opinion. The y-axis in the graph on the left-hand side of the figure is the mean absolute error of the MRP opinion estimates. The y-axis in the graph on the right-hand side is the correlation between the MRP estimates and true opinion. All of the MRP estimates reported here were obtained using the "baseline" MRP model.

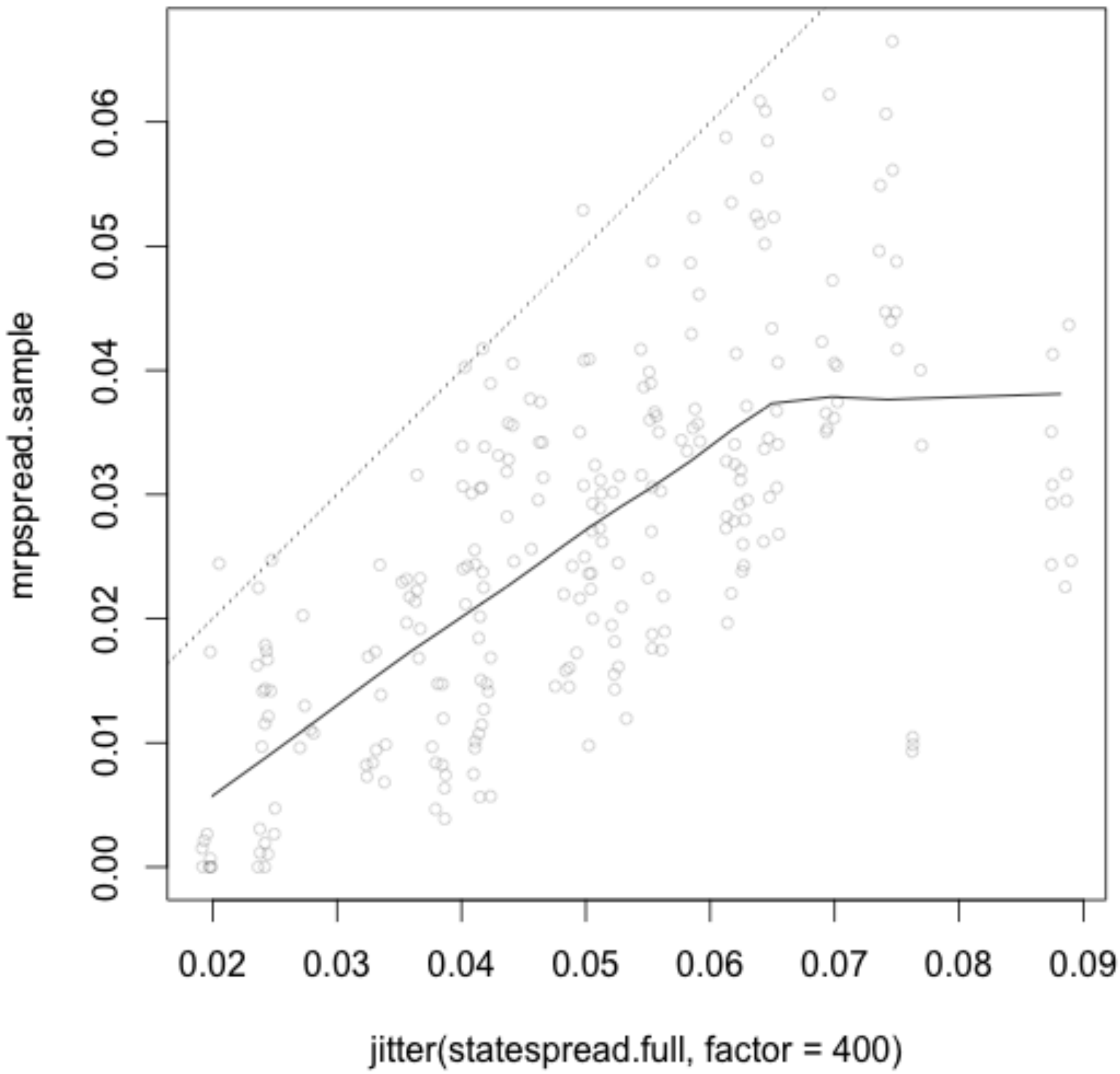

Figure 2: *Baseline Model: Spread of MRP Estimates vs. Spread of True State Opinion.* This figures shows how the spread (standard deviation) of MRP estimates compares to the spread of true state opinion. The x-axis is the standard deviation of true state opinion and the y-axis is the standard deviation of the MRP opinion estimates. All of the MRP estimates reported here were obtained using the "baseline" MRP model.

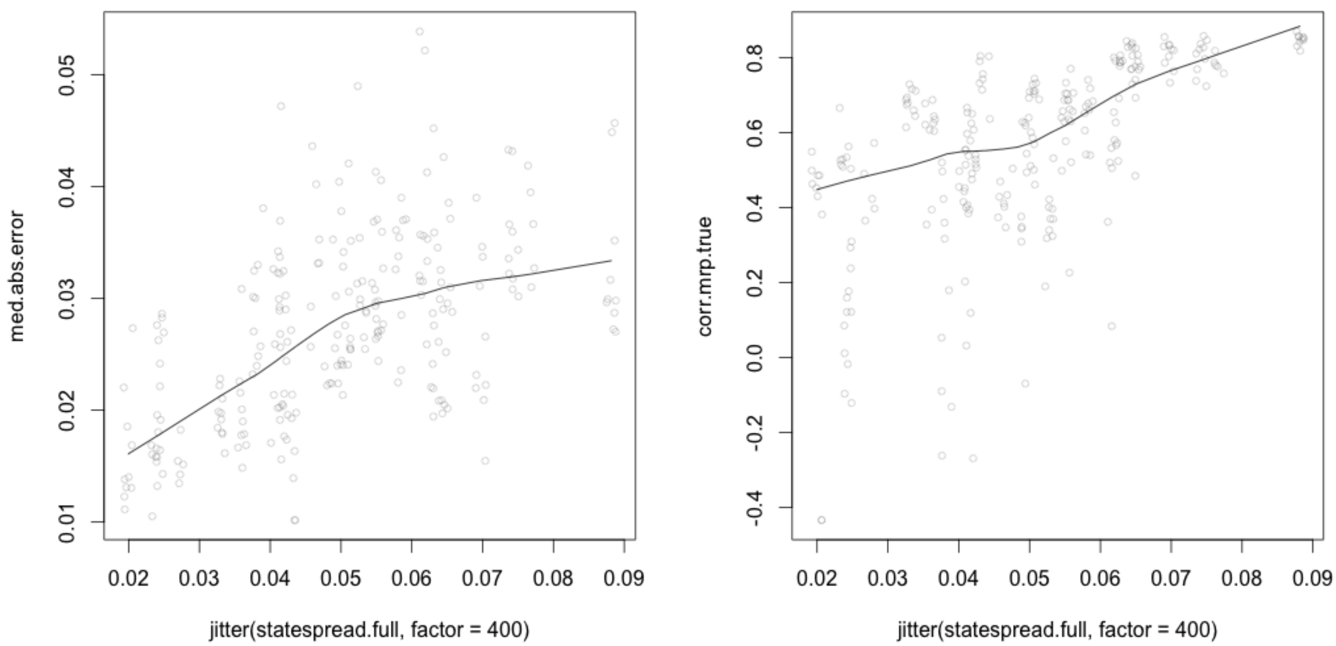

Figure 3: *Geographic Predictor Model: Error and Correlation by the Spread of True State Opinion.* This figures shows how the accuracy of MRP varies as the spread of true state opinion (measured as a standard deviation) grows. The x-axis in both graphs is the standard deviation of true state opinion. The y-axis in the graph on the left-hand side of the figure is the mean absolute error of the MRP opinion estimates. The y-axis in the graph on the right-hand side is the correlation between the MRP estimates and true opinion. All of the MRP estimates reported here were obtained using a Geographic Predictor model. This models is identical to the baseline model, but adds presidential vote share as a substantive state-level predictor.

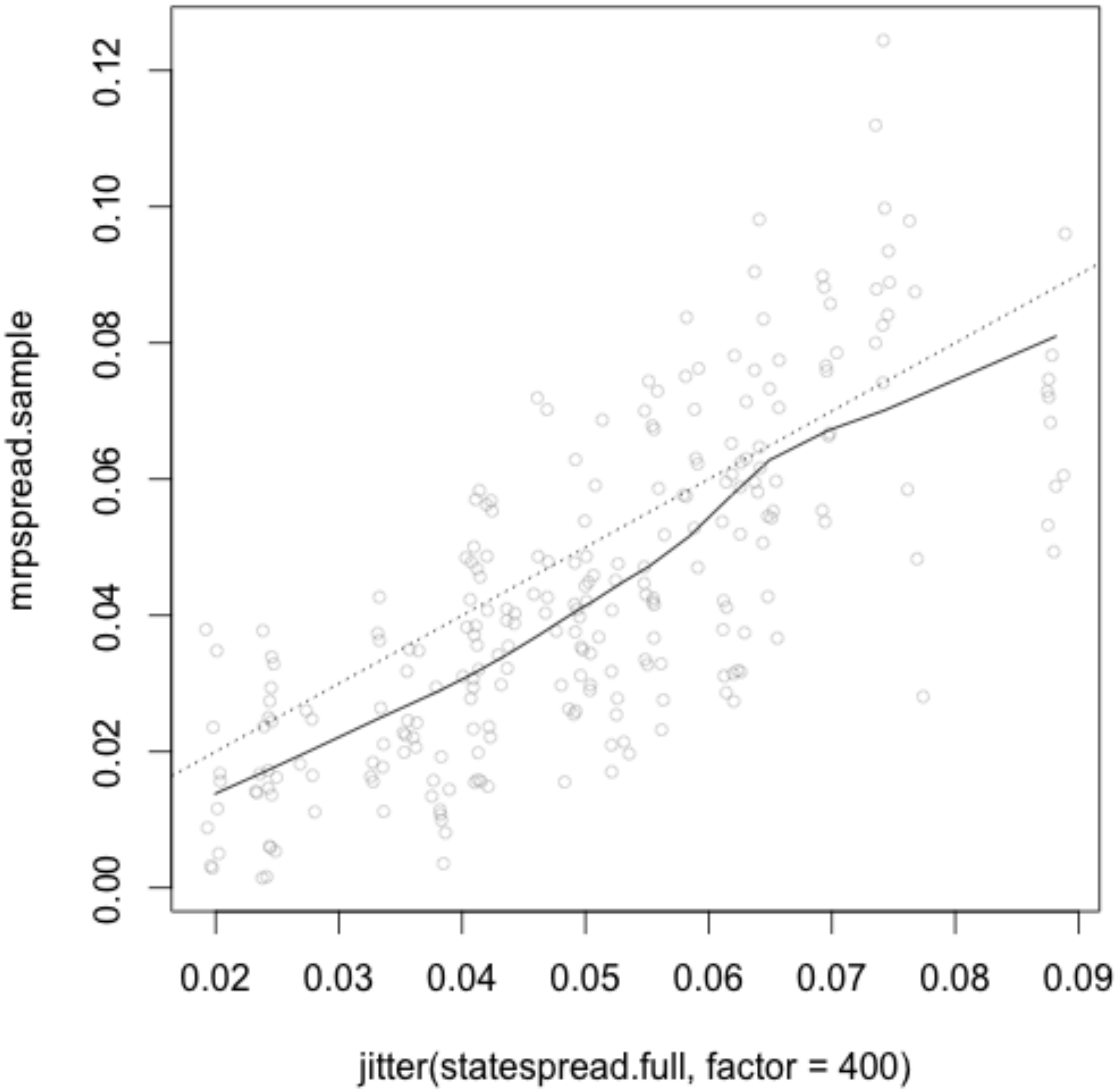

Figure 4: *Geographic Predictor Model: Spread of MRP Estimates vs. Spread of True State Opinion.* This figures shows how the spread (standard deviation) of MRP estimates compares to the spread of true state opinion. The x-axis is the standard deviation of true state opinion and the y-axis is the standard deviation of the MRP opinion estimates. All of the MRP estimates reported here were obtained using a Geographic Predictor model. This models is identical to the baseline model, but adds presidential vote share as a substantive state-level predictor.

7 Appendix

Results for the Baseline Model across Questions

question	corr	error
interest in news and public affairs	0.77	1.87
affirmative action	0.61	2.84
turnout	0.69	2.58
media use - newspaper	0.51	2.25
does r intend to vote in 2010	0.52	1.4
institution approval - congress	0.52	2.3
vote president 2008	0.39	3.97
pol activity - donate money	0.63	2.69
campaign contact	0.6	3.35
party of government knowledge - senate	0.63	1.98
jobs-environment	0.38	2.74
vote 2008	0.64	1.41
racial resentment a	0.26	3.25
stock ownership	0.67	2.99
union member - household	0.62	2.47
phone service	0.37	3.17
gun control	0.52	5.29
health insurance - no	0.41	2.05
roll call votes - recovery and reinvestment	0.37	4.01
tea party favorability	0.16	3.67
home ownership	0.33	2.99
roll call votes - clean energy	0.36	4.05
institution approval - supreme court	0.56	2.71
gay marriage	0.48	5.03
military troops - intervene in genocide or civil war	0.49	2.63
roll call votes - health reform	0.39	4.89
roll call votes - children's health insur	0.35	2.93
military household - family	0.24	1.52

variant	error	corr	congcorrect	shrinkage	residstatevar
basic	3.00	0.46	0.92	0.51	0.18
basic_dpsp	2.72	0.57	0.94	0.77	0.11
basic_dpsp_region	2.73	0.56	0.94	0.78	0.09
basic_dpsp_z	2.71	0.57	0.94	0.76	0.11
basic_ideology	2.70	0.52	0.94	0.77	0.10
basic_ideology_dpsp	2.77	0.53	0.94	0.85	0.09
basic_plus_income	2.96	0.48	0.92	0.53	0.18
basic_plus_income_dpsp	2.68	0.57	0.94	0.77	0.11
basic_plus_income_dpsp_z	2.65	0.59	0.94	0.76	0.11
basic_plus_relig	2.68	0.55	0.93	0.52	0.16
basic_plus_relig_dpsp	2.67	0.57	0.94	0.76	0.11
basic_plus_relig_plus_income	2.64	0.57	0.94	0.53	0.15
basic_prelig	2.80	0.52	0.94	0.78	0.11
basic_pres	2.75	0.56	0.94	0.85	0.10
basic_region	3.00	0.50	0.92	0.56	0.12
just_race	3.19	0.33	0.91	0.46	0.17
just_race_dpsp	2.81	0.51	0.94	0.75	0.11
just_race_ideology	2.82	0.44	0.94	0.74	0.09
just_race_prelig	2.92	0.44	0.93	0.76	0.11
just_race_pres	2.84	0.49	0.94	0.82	0.10
just_race_region	3.13	0.39	0.92	0.52	0.12
nodemog	3.38	0.29	0.91	0.24	0.16
nodemog_dpsp	3.06	0.44	0.93	0.62	0.10
nodemog_dpsp_ideo	3.13	0.41	0.93	0.72	0.08
nodemog_dpsp_pres	3.11	0.46	0.93	0.84	0.08
nodemog_ideo	3.09	0.40	0.93	0.59	0.09
nodemog_ideo_pres	3.13	0.42	0.93	0.81	0.08
nodemog_prelig	3.09	0.40	0.93	0.59	0.09
nodemog_pres	2.98	0.45	0.93	0.70	0.09
paired_interactions	3.00	0.46	0.92	0.51	0.18
paired_interactions_dpsp	2.71	0.57	0.94	0.76	0.11
paired_interactions_dpsp_z	2.71	0.57	0.94	0.76	0.11
paired_interactions_ideology	2.68	0.53	0.94	0.76	0.10
paired_interactions_ideology_dpsp	2.75	0.53	0.94	0.85	0.08
paired_interactions_prelig	2.80	0.52	0.94	0.77	0.12
paired_interactions_pres	2.75	0.56	0.94	0.84	0.10
quad_interactions	2.99	0.46	0.92	0.51	0.18
quad_interactions_dpsp	2.71	0.57	0.94	0.77	0.11
quad_interactions_ideology	2.69	0.53	0.94	0.77	0.10
quad_interactions_ideology_dpsp	2.76	0.53	0.94	0.85	0.08
quad_interactions_prelig	2.79	0.52	0.94	0.78	0.11
quad_interactions_pres	2.74	0.56	0.94	0.84	0.10
st.all_interaction	3.14	0.45	0.92	0.51	0.17
st.all_interaction_dpsp	2.86	0.57	0.94	0.75	0.10
st.r_interaction	3.01	0.46	0.92	0.53	0.14
st.r_interaction_dpsp	2.73	0.57	0.94	0.77	0.09

state	DPSP
AK	0.052
AL	-0.468
AR	-0.349
AZ	0.074
CA	0.257
CO	0.096
CT	0.341
DC	0.358
DE	0.057
FL	0.103
GA	-0.224
HI	0.268
IA	-0.011
ID	-0.037
IL	0.114
IN	-0.134
KS	-0.016
KY	-0.210
LA	-0.220
MA	0.493
MD	0.136
ME	0.228
MI	0.021
MN	0.118
MO	-0.022
MS	-0.474
MT	0.061
NC	-0.178
ND	-0.211
NE	-0.291
NH	0.278
NJ	0.250
NM	0.058
NV	0.028
NY	0.335
OH	-0.040
OK	-0.328
OR	0.183
PA	0.101
RI	0.474
SC	-0.256
SD	-0.243
TN	-0.380
TX	-0.127
UT	-0.516
VA	-0.056
VT	0.367
WA	0.266
WI	0.091
WV	-0.232
WY	-0.185