How Should We Estimate Public Opinion in The States?

Jeffrey R. Lax Columbia University
Justin H. Phillips Columbia University

We compare two approaches for estimating state-level public opinion: disaggregation by state of national surveys and a simulation approach using multilevel modeling of individual opinion and poststratification by population share. We present the first systematic assessment of the predictive accuracy of each and give practical advice about when and how each method should be used. To do so, we use an original data set of over 100 surveys on gay rights issues as well as 1988 presidential election data. Under optimal conditions, both methods work well, but multilevel modeling performs better generally. Compared to baseline opinion measures, it yields smaller errors, higher correlations, and more reliable estimates. Multilevel modeling is clearly superior when samples are smaller—indeed, one can accurately estimate state opinion using only a single large national survey. This greatly expands the scope of issues for which researchers can study subnational opinion directly or as an influence on policymaking.

Democratic theory suggests that the varying attitudes and policy preference of citizens across states should play a large role in shaping both electoral outcomes and policymaking. Accurate measurements of state-level opinion are therefore needed to study a wide range of related political issues, issues at the heart of political science such as representation and policy responsiveness.

Unfortunately, measuring state opinion is not easy. Despite the proliferation of public opinion polls, state-level surveys are still quite rare. Finding comparable surveys across all (or even many) states is nearly impossible. And, while most national poll data include the home state of the respondents, there are almost always too few respondents within each state to be considered an adequate sample.

In response to these problems, scholars have devised sophisticated techniques for coping with sparse data, techniques which allow them to use national surveys to generate estimates of state-level opinion. The two main methods are *disaggregation* and *simulation*. However, each method raises some concerns—and important questions remain as to which method should be used, when, and how.

The currently dominant method is disaggregation, developed and popularized by Erikson, Wright, and McIver (1993). This method pools large numbers of national surveys and then disaggregates the data so as to calculate opinion percentages by state. Erikson, Wright, and McIver's work grew, in part, out of a critique of earlier methods that simulated state-level public opinion using only demographic data.[1] Erikson, Wright, and McIver showed that states vary *even after* controlling for demographics and that the difference between state effects is often the same magnitude as the effect of shifting demographic categories (1993, 53). In short, we should not ignore geography.

Disaggregation is easily implemented, in that it skips any analysis of demographic correlations. It does, however, have its own drawbacks. Typically, surveys over many

[1]This work, dating at least as far back as Pool, Abelson, and Popkin (1965), estimated state opinion using demographic correlations estimated at the national level and then weighted the predictions by demographic type given each state's demographic composition. Differences between states were *only* incorporated in terms of demographics, so that two demographically identical states would have identical predictions.

years, say 10 to 25, must be pooled to guarantee sufficient samples within states (e.g., Brace et al. 2002; Gibson 1992; Norrander 2001).[2] This blocks any inquiry into temporal dynamics. And, if such dynamics exist, they would call into question how well the estimates reflect current opinion. There are also sampling issues, such as clustering, that undermine sample randomness within states.

As an alternative, recent work by Park, Gelman, and Bafumi (2006) presents a new version of simulating state opinion, based on multilevel regression and poststratification (MRP).[3] This has the potential to combine the best features of both disaggregation and simulation techniques. It revives the old simulation method, incorporating demographic information to improve state estimation, while allowing for nondemographic differences between states. That is, opinion is modeled as a function of both demographics *and* state-specific effects. The estimation of these effects is improved by using a multilevel model and partially pooling data across states (to an extent warranted by the data). Predictions are made for each demographic-geographic respondent type, and these predictions are then poststratified (weighted) by population data. The drawback here is the need for detailed demographic data on respondents and states, along with greater methodological complexity.[4]

Is it worth it? Are the estimates from MRP as good as those from disaggregation? Under what conditions, if any, can they match or improve upon the estimates from disaggregation, and by how much? Which method should scholars adopt?

This study presents the first systematic comparison between the predictive accuracy of disaggregation and that of MRP. We explore sample size, model complexity, and the balance between demographic and geographic predictors. We use our findings to address questions crucial to the applied researcher: How many national surveys does one need to do MRP? How complicated or accurate

a demographic typology is necessary? How important is the incorporation of demographic versus geographic predictors?

We attack these questions as follows. As we explain in the next section, we start with a large set of national surveys, a random sample of which is used to calculate a baseline measure of "true" opinion. We then use samples of the remaining respondents to assess how well each method does in matching the baseline measure.[5] The third section shows that, with very large samples, both methods work well, but multilevel modeling performs better generally. MRP yields smaller errors, higher correlations, and more reliable estimates—even though we use disaggregation (on the baseline sample) to establish "true" opinion. MRP is clearly superior when samples are smaller and even works well on samples the size of a single large national poll. The fourth section considers varying individual response models and how large a role demographic and geographic predictors play in successful state estimates.

In the fifth section, we further explore the possibility of using single national polls to generate MRP estimates. We first establish the face validity of the estimates and then check external validity by using MRP estimates to predict actual state polls, which serve as a second measure of "true" state opinion. We find that estimates from a single national poll correlate strongly to the actual state polls.

To confirm that our findings are not artifacts of the particular surveys used, we replicate the findings above using survey responses on other gay rights issues and survey data from the 1988 presidential election (the sixth section). Results are highly similar.

We then conclude, offering advice as to when and how to use each method, and drawing out the implications of our findings for the study of subnational opinion and policy responsiveness. Our results provide new and useful guidance to scholars in assessing the trade-offs between estimation methods and determining whether MRP is worth the implementation costs. Most importantly, we show that (1) MRP should be employed when data samples are small to medium size; (2) for very large samples, the gains from MRP are less likely to be worth its implementation costs; (3) relatively simple demographic typologies can suffice for MRP, but additional demographic information improves estimation; and (4) MRP can be used successfully even on small samples, such as individual national polls.

---

[2]For any national survey sample size (say, 1,000 respondents), approximately eight such national surveys must be pooled to obtain the same targeted number of respondents in California (the largest state), given its population share of 12.5%; South Carolina (the median state) requires 70 surveys; Wyoming (the smallest state) requires 571 surveys.

[3]For substantive applications, see Lax and Phillips (2008) and Kastellec, Lax, and Phillips (2008).

[4]The MRP method is more complicated than simple small area estimation that does not poststratify (such as in the large applied literature on small area estimation in public health surveys; see, e.g., Fay and Herriot 1979), but less complicated than various techniques used in truly thorny problems such as census adjustment. Another approach might be to combine survey weighting with multilevel modeling; while this is not currently possible, it might offer some advantages in the future (see Gelman 2007).

[5]This approach is similar to cross-validation of reliability across samples.

Estimating Opinion

Disaggregation Overview

The most commonly used method for estimating state-level opinion is disaggregation. The main advantage relative to MRP is its simplicity. After combining a set of national polls, one calculates the opinion percentages disaggregated by state. The only necessary data are the respondent's answer and state of residence. No further statistical analysis is necessary.

There are potential problems, however. The principle disadvantage, as noted above, is that it requires a large number of national surveys to create a sufficient sample size within each state (see, e.g., Brace et al. 2002; Gibson 1992; Miller and Stokes 1963; Norrander 2001). And smaller states (e.g., Rhode Island) or those seldom surveyed (e.g., Alaska and Hawaii) must sometimes be dropped entirely.

Where many contemporaneous surveys are available, it may not be particularly problematic to combine them. Usually, however, one must collect surveys over a long time window to achieve sufficient state sample sizes. (For example, Erikson, Wright, and McIver 1993 combine 12 years and Brace et al. 2002 combine 25 years.) Survey pooling would then be most appropriate where opinion is stable. If opinion is not stable over time, then this method will be less accurate as to opinion at any particular point in time. Furthermore, disaggregation obscures any such dynamics over time within states. For those survey questions that are asked less frequently, or for newer issues, it simply may not be possible to collect a sufficient number of compatible surveys.

Additionally, national surveys, while representative at that level, are often flawed in terms of representativeness or geographic coverage at the state level, due to clustering and other survey techniques utilized by polling firms (Norrander 2007, 154).

MRP Overview

One alternative estimation strategy is the simulation of state opinion using national surveys, a method which has a long history (e.g., Pool, Abelson, and Popkin 1965; and, for critiques, see Erikson, Wright, and McIver 1993; Seidman 1975; and Weber et al. 1972). The current implementation of such simulation has certain advantages over earlier efforts. For example, some older applications used only demographic correlations. That is, respondents were generally modeled as differing in their demographic but not their geographic characteristics, so the prediction for any demographic type was unvaried by state. In contrast,

MRP takes into account geography as well, incorporating the criticism that people differ in their opinions even after controlling for the standard demographic typologies. In short, place matters and the updated simulation method allows it to.

MRP is also far more sophisticated in the way it models individual survey responses, using Bayesian statistics and multilevel modeling (Gelman and Little 1997; Park, Gelman, and Bafumi 2006). It improves upon the estimation of the effects of individual- and state-level predictors by employing recent advances in multilevel modeling, a generalization of linear and generalized linear modeling, in which relationships between grouped variables are themselves modeled and estimated. This partially pools information about respondents across states to learn about what drives individual responses.[6] Whereas the disaggregation method copes with insufficient samples within states by combining many surveys, MRP compensates for small within-state samples by using demographic and geographic correlations.

Specifically, individual survey responses are modeled as a function of demographic and geographic predictors, partially pooling respondents across states to an extent determined by the data. (We elaborate on this shortly.) Unlike the earlier simulation method, the location of the respondents is used to estimate state-level effects on responses. These state-level effects can be modeled using additional state-level predictors such as region or state-level (aggregate) demographics (e.g., those not available at the individual level). In this way, all individuals in the survey, no matter their location, yield information about demographic patterns which can be applied to all state estimates, and those residents from a particular state or region yield further information as to how much predictions within that state or region vary from others after controlling for demographics. The final step is poststratification, in which the estimates for each demographic-geographic respondent type are weighted (poststratified) by the percentages of each type in the actual state populations.

The multilevel model allows us to use many more respondent types than would classical methods. This improves accuracy by incorporating more detailed population information. Earlier simulation methods, rather than using poststratification by full respondent type, would poststratify on the margins ("raking," e.g., Deville, Sarndal, and Sautory 1993). Another advantage of MRP is that poststratification can correct for clustering and other statistical issues that may bias estimates obtained via survey

---

[6]Disaggregation does not pool information across states (only across surveys within states).

pooling. That is, poststratification can correct for differences between samples and population.[7] A final benefit of MRP is that modeling individual responses is itself substantively interesting, in that one can study the relationship between demographics and opinion and inquire as to what drives differences between states—demographic composition or residual cultural differences.[8]

Obviously, this method and similar methods are statistically more complex, as compared to disaggregation. For some scholars, these methods will require learning new statistical techniques[9] and obtaining additional data. One needs demographic information on individual survey respondents, along with census data to poststratify the demographic-geographic types. That is, consider poststratification by sex, race, and education in, say, Nevada. MRP requires knowing not just the percentage of women and the percentage of Hispanics and the percentage of college graduates, but rather the share of Nevada's population that consists of female Hispanic college graduates. The problem is that not all cross-tabulations are available, particularly for smaller geographic units (say, congressional districts). This could limit the number of subtypes, though we show below that simpler typologies can suffice.[10] Of course, some of the start-up costs—in particular, learning the new method and setting up the census cross-tabulations—need only be paid once.

Data

To evaluate the two methods, we first use a set of 26 national polls from 1996 through 2005 that ask respondents about their support for same-sex marriage. The polls are random national samples conducted by Gallup, Pew, ABC News, CBS News, AP, Kaiser, and Newsweek (the list of specific polls is available upon request). We then recode

[7]NES and other studies are generally not set up to sample within states representatively. In-person surveys tend to have this problem, although telephone surveys are usually adequate by state, unless clustering is used. In terms of survey nonresponse, if we get a biased sample of respondent types, poststratification will correct for it (whereas disaggregation will not, of course); if we get a biased sample within a respondent type, that will affect both sets of estimates, but the MRP estimates might suffer less due to partial pooling.

[8]Also, opinion dynamics are not "squashed" by the MRP method, as they are in the disaggregation method. More than that, one can actually model opinion dynamics, by controlling for poll differences in the response model, or by running a model for each poll.

[9]Gelman and Hill (2007) provide code for various packages.

[10]One approach is to try the analysis using various available combinations of cross-tabulated frequencies and averaging over the estimates produced (see the analysis for school districts in Berkman and Plutzer 2005).

as necessary to combine these polls into a single internally consistent data set.[11] For each respondent, we have sex (male or female), race (black, Hispanic, or white and other), one of four age categories (18–29, 30–44, 45–64, and 65+), and one of four education categories (less than a high school education, high school graduate, some college, and college graduate). Race and gender are combined to form six possible categories (from male-white to female-Hispanic). Finally, each respondent's state and region is indicated (Washington, DC, is included as a "state" and its own region, along with Northeast, Midwest, South, and West). For each state, we have the percent of evangelical Protestants and Mormons (American Religion Data Archive 1990).

Responses are coded 1 for support of same-sex marriage and 0 otherwise ("no," "don't know," or "refused"). This captures positive support among all respondents, not simply those expressing an opinion. (Coding refusals as missing does not change our results. There are slight variations across polls in question wording and ordering, though each polling firm tends to use the same wording over time.)

While many survey questions could yield useful data for assessing the relative merits of disaggregation and MRP, same-sex marriage has certain advantages. First, the state estimations are themselves of substantive interest to scholars, policy makers, and pundits alike, and this is a policy that is in large part set at the state level. There is also substantial opinion variation across states, which avoids biasing results towards MRP, which partially pools across states (the greater the opinion differences between residents of different states, the less useful, say, Ohio respondents are for understanding Texas respondents). Next, there is a sufficient number of national polls concerning same-sex marriage so as to make disaggregation plausible and so that survey size issues can be studied. Finally, there are also enough state polls to enable meaningful comparisons to the estimates using MRP.

Modeling Individual Responses

MRP begins by modeling individual responses, so as to create predictions for each respondent type. We use a multilevel logistic regression model, estimated using the LMER function ("linear mixed effects in R"; Bates 2005).[12] Rather than using "unmodeled" or "fixed"

[11]To the best of our knowledge, we included *all* available surveys from reputable sources that have the necessary demographic and geographic information.

[12]We used R version 2.6.2 and lme4 version 0.99875-9. Code available upon request.

effects, the model uses "random" or "modeled" effects, at least for some predictors (see Gelman and Hill 2007, 244–48). That is, we assume that the effects within a grouping of variables are related to each other by their hierarchical or grouping structure. For example, we model the effects of the four educational levels as drawn from some common distribution. The state effects are drawn from a common distribution, controlling for percent Evangelical/Mormon and region, and these regional effects are in turn drawn from their own common distribution.

For data with hierarchical structure (e.g., individuals within states within regions), 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 pooling) by arbitrary assumption (see Gelman and Hill 2007, 254–58).[13] The general principle 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.[14] The degree of pooling emerges from the data, with similarities and differences across groups estimated endogenously.

This modeling structure also lets us break down our respondents into tighter demographic categories, for more accurate poststratification. For example, we include interaction effects between demographic predictors and can separate Hispanic respondents from white respondents. Also, in a multilevel model, we can include indicators for all groups without needing to omit one as a baseline (because of the prior distribution for the coefficients, the matrix is invertible), so that many results are easier to interpret (Gelman and Hill 2007, 275, 393). We do find significant differences between racial/ethnic groups.

While there is more than one way to express a multilevel model (see Gelman and Hill 2007, 262), the following is the most intuitive.[15] We model each individual's response as a function of his or her demographics and state (for individual i, with indexes j, k, l, m, s, and p for race-gender combination, age category, education category, region, state, and poll year, respectively):

$$\Pr(y_i = 1) = \text{logit}^{-1}\big(\beta^0 + \alpha^{race,gender}_{j[i]} + \alpha^{age}_{k[i]} + \alpha^{edu}_{l[i]} + \alpha^{state}_{s[i]} + \alpha^{year}_{p[i]}\big) \quad (1)$$

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

$$\alpha^{race,gender}_{j} \sim N\big(0, \sigma^2_{race,gender}\big), \text{ for } j = 1, \ldots, 6$$
$$\alpha^{age}_{k} \sim N\big(0, \sigma^2_{age}\big), \text{ for } k = 1, \ldots, 4$$
$$\alpha^{edu}_{l} \sim N\big(0, \sigma^2_{edu}\big), \text{ for } l = 1, \ldots, 4$$
$$\alpha^{year}_{p} \sim N\big(0, \sigma^2_{year}\big), \text{ for } p = 1, \ldots, 7 \quad (2)$$

That is, each is modeled as drawn from a normal distribution with mean zero and some estimated variance. The state effects[16] are in turn modeled as a function of the region into which the state falls and the state's percentage of evangelical or Mormon residents:[17]

$$\alpha^{state}_{s} \sim N\big(\alpha^{region}_{m[s]} + \beta^{relig} \cdot relig_s, \sigma^2_{state}\big),$$
$$\text{for } s = 1, \ldots, 49 \quad (3)$$

The region variable is, in turn, another modeled effect:

$$\alpha^{region}_{m} \sim N\big(0, \sigma^2_{region}\big), \text{ for } m = 1, \ldots, 5 \quad (4)$$

We use standard demographic indicators: race, gender, age, and education have all been shown to be important predictors of social attitudes, in particular towards gays and lesbians (e.g., Cook 1999; Haider-Markel and Meier 1996). We have kept the model relatively simple, to show that even such a sparse specification can do quite well in terms of precision at the state level, as compared to disaggregation. Using a simple model of opinion response should bias findings against the multilevel model's success. Our findings are robust to variations in this specification (such as running race and gender as unmodeled fixed effects or adding interaction terms between age and education), and our state-level predictions are robust even when using simpler respondent typologies. While one

---

[13]Park, Gelman, and Bafumi (2006) compare MRP to these two extremes. Partial pooling across states did better than running a separate model for each state's respondents (no pooling across states) and better than pooling all respondents across states (so that only demographic information was used to model individual response before poststratification).

[14]There is a lengthy theoretical literature in statistics showing that multilevel models reduce mean squared errors when the number of groups is three or more (e.g., Efron and Morris 1975; James and Stein 1960).

[15]It can also be expressed as a classical regression with correlated errors.

[16]We have to drop Hawaii and Alaska in the disaggregation, though we could generate predictions for those states using MRP by setting each state's coefficient to its regional mean or that of a similar state.

[17]Group-level predictors such as these can be directly of interest but also reduce any unexplained group-level variation, meaning more precise estimation (Gelman and Hill 2007, 271). One could of course include other state-level predictors.

might think to include religion at the individual level, rather than include it only as a state-level indicator, that datum is less commonly available in surveys and is not available in census data.

Poststratification

For any set of individual demographic and geographic values, cell *c*, the results above allow us to make a prediction of same-sex marriage support. Specifically, θ_c is the inverse logit given the relevant predictors and their estimated coefficients.[18] 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 actual state populations.

We calculate the necessary population frequencies using the "1-Percent Public Use Microdata Sample" from the 2000 census, which gives us the necessary demographic information for 1% of each state's voting-age population. After dropping Alaska and Hawaii, which are almost never polled, and including Washington, DC, as a "state," we have 49 states with 96 demographic types in each. This yields 4,704 possible combinations of demographic and state values, ranging from "White," "Male," "Age 18–29," "Not high school graduate," in "Alabama," to "Hispanic," "Female," "Age 65+," "College degree or more," in "Wyoming." Each cell *c* is assigned the relevant population frequency N_c. For example, for the cells mentioned above the frequencies are 581 (1.7% of Alabama's total population) and 0, respectively.

The prediction in each cell, θ_c, needs to be weighted by these population frequencies of that cell. For each state, we calculate the average response, over each cell *c* in state *s*:

$$y_{\text{state}s}^{\text{MRP}} = \frac{\sum_{c \in s} N_c \theta_c}{\sum_{c \in s} N_c} \qquad (5)$$

This yields our prediction of the affirmative response rate in state *s*.

Comparing Methods Using National Polls

Data and Methods

To assess the relative performance of the disaggregation and MRP methods in different sample sizes, we

rely upon cross validation.[19] We randomly split the data, using half to define the baseline or "true" state opinion. In the baseline data, we disaggregate the sample and measure each state's actual percentage of pro-gay-marriage support within the sample. That is, we treat disaggregation of the baseline sample as the prediction goal.

We then use some portion of the remaining data to generate estimates of opinion, once employing disaggregation and a second time using MRP. We draw such random samples 200 times (both the baseline data and the data for comparative estimation) for four different size samples (800 simulation runs in all). The approximate sample sizes are 14,000 for the baseline sample; 1,400 for the 5% sample; 2,800 for the 10% sample; 7,000 for the 25% sample; and 14,000 for the 50% sample (that is, all data not already in the baseline sample).[20] These run from the size of perhaps a single good-sized national poll to a sample 10 times this size.

By using the disaggregation method to calculate our standard for the target level of state opinion, we set the baseline in favor of disaggregation and potentially bias findings against MRP, thus taking a conservative position in favor of the status quo. We follow Erikson, Wright, and McIver (1993) and Brace et al. (2002) in using unweighted survey responses, for both the baseline data and the sample data. To the extent that poststratification corrects for any lack of weighting, this also biases our findings against MRP—because the unweighted data is being used both to define the baseline and in the disaggregation on the sampled data. (This all, of course, means that where MRP and disaggregation differ, even if MRP has the larger "error," it could actually be closer to true state opinion.)
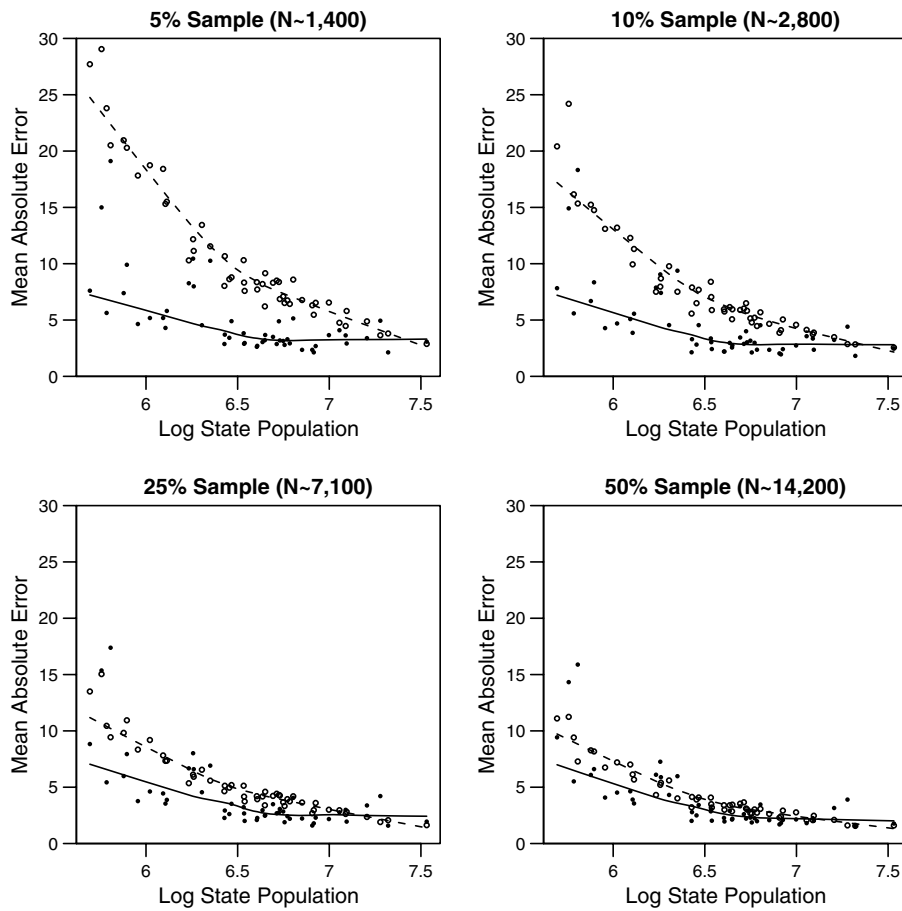
Results

We now measure predictive success—how close each set of estimates is to the measure for the baseline sample—in various ways, discussed in more detail below. In each run of a simulation *q*, let $y_{q,s}^{base}$ be the opinion percentage in state *s* in the baseline data (again, measured as the

---

[18]Since we allow different poll-year intercepts when estimating the individual's response, we must choose a specific year coefficient when generating these predicted values using the inverse logit. We simply use the average value of the coefficients, which is zero by assumption.

[19]We also calculated the reliability and stability of the estimates, using standard split-sample techniques. We follow Erikson, Wright, and McIver (1993, 22) and Brace et al. (2002, 187) in using the Spearman-Brown prophesy formula on the split-halves correlation (for reliability, splitting the poll data into two random halves; for stability, splitting them into early and late sets with roughly equal population). Disaggregation estimates had reliability and stability coefficients of .91 and .90, respectively; MRP had reliability and stability coefficients of .99 and .99, respectively.

[20]The number of observations in the estimation samples varies slightly.

FIGURE 1 Cross Validation—Mean Errors by State and Estimation

Each panel shows the results for a particular sample size. We show the mean error by state against the log of state population, using MRP (●) and disaggregation (○). Lowess curves for each are shown (solid and dashed, respectively).

disaggregation method does, totaling up the simple percentage by state), let $y_{q,s}^{dis}$ be the disaggregated percentage in state s on the sampled data, and let $y_{q,s}^{MRP}$ be the estimate in state s using MRP.[21]

For each of the four sample sizes, we do the following. We first calculate the errors produced by each method in each state in each simulation, the simplest measure being the absolute difference between the estimates and the baseline measure:

$$e_{q,s}^{dis} = \left| y_{q,s}^{dis} - y_{q,s}^{base} \right|, \quad e_{q,s}^{MRP} = \left| y_{q,s}^{MRP} - y_{q,s}^{base} \right| \quad (6)$$

[21]Occasionally, in the smaller samples, the model does not converge in a particular run and so we drop those observations for both methods. This does not affect any results. Were this to happen when running a single model, one would rerun it, changing LMER settings or simplifying the model.

This forms two matrices of absolute errors, of size 200 (simulations) × 49 (states) each. For state s, we then calculate the mean absolute error for each method across simulations (49 x 2 mean errors):[22]

$$\bar{e}_s^{dis} = \frac{\sum_q e_{q,s}^{dis}}{200}, \quad \bar{e}_s^{MRP} = \frac{\sum_q e_{q,s}^{MRP}}{200} \quad (7)$$

The four panels in Figure 1 show the results for the four sample sets, plotting the mean error for each state against the log of state population. The solid dots show MRP's errors, while the open circles show the mean errors for disaggregation. We add locally weighted regression (lowess) curves for each.

Figure 1 reveals three patterns of interest. First, within each panel, as expected, errors are smaller in larger states.

[22]Focusing on median errors yields equivalent results.

FIGURE 2 Cross Validation—Summary Performance Measures
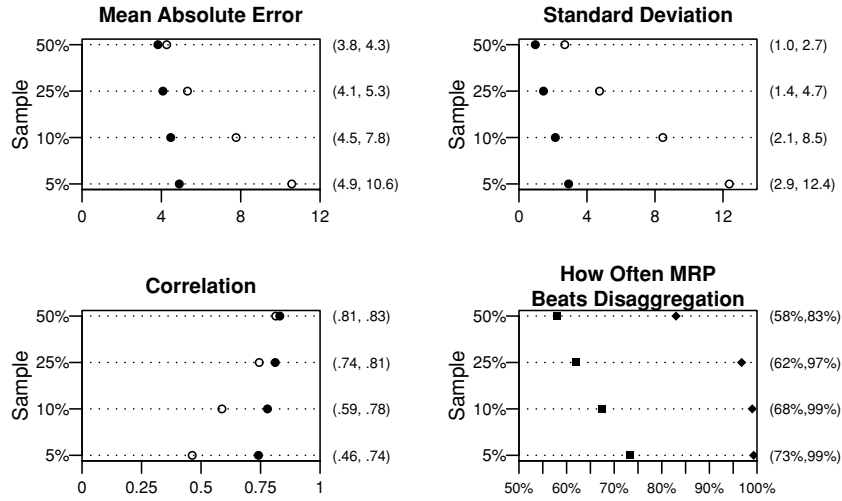
The top-left panel plots the mean absolute error across states and simulation runs for MRP (●) and disaggregation (○). The top-right panel plots, for each method, the average (over states) of the standard deviation of state estimates across simulation runs. The bottom-left panel shows the correlation of each set of estimates to the baseline measures. The bottom-right panel shows how often the MRP error is smaller than the disaggregation error using (■) each state estimate (across states and simulation runs) as the unit of analysis and using (◆) each simulation run as the unit of analysis (averaging over states within each simulation run). Values plotted are indicated along the right axis.

However, disaggregation's errors vary more with state size, drastically so for the smaller samples (the top panels). Second, again within each panel, the MRP estimate beats disaggregation on average and for almost every state in every panel. The differences between the two methods for the 50% sample are smaller, suggesting that it matters less which method is used. But the differences increase significantly as we move back across the panels to the 5% sample. Finally, whereas the mean errors for disaggregation increase significantly as sample size decreases (the curves are both higher and steeper), the mean errors for the MRP estimates hardly vary across panels. That is, using MRP on a sample the size of a single reasonably large national survey is very nearly as successful as using the MRP method on a far larger sample: throwing away roughly 12,600 random observations does little damage on average. Indeed, MRP on the 5% samples is nearly as accurate as using disaggregation on the 50% samples, so, to put it another way, it is like getting 12,000 or more observations free.

We next construct various summary measures of performance, shown in Figure 2.[23] First, we calculate the mean absolute error over both states and simulations,

collapsing the means-by-state above into a single number for each sample size and method:

$$\bar{e}^{dis} = \frac{\sum_{q,s} e^{dis}_{q,s}}{200 \cdot 49}, \quad \bar{e}^{MRP} = \frac{\sum_{q,s} e^{MRP}_{q,s}}{200 \cdot 49} \quad (8)$$

Figure 2's top-left panel shows these means, with solid circles for MRP and open circles for disaggregation. Note, as suggested by our discussion of Figure 1, that the MRP method's mean absolute error is smaller no matter what the sample size; that the mean absolute error varies little for MRP (ranging from 4 to 5), but greatly for disaggregation (ranging from 4 to 11); and that MRP using the 5% samples is nearly as accurate as disaggregation over the largest sample.

We next ask how much the estimates for a state vary depending on the particular sample used. For each method, we calculate the standard deviation in the estimates for each state across the simulations. We then take the mean across states. The top-right panel shows these mean standard deviations for each method. Note that the mean standard deviation is always smaller for MRP, approximately one-fourth to one-third the size of the disaggregation method. The variation in the disaggregation estimates is also far more sensitive to sample size than that for MRP. Moving from the largest sample to the smallest triples the mean standard deviation for MRP and more

[23]We follow the advice of Kastellec and Leoni (2007) in presenting results graphically.

than quadruples the mean standard deviation for disaggregation. Overall, the results achieved by poststratifying are far more stable across samples than those achieved by using raw survey data.

The bottom-left panel shows the correlations between estimates and the baseline measure. Again, disaggregation is much more sensitive to sampling size, only achieving a strong correlation for the larger samples. On the other hand, MRP is correlated at .75 or better even in the smallest samples. A level of correlation this high is surely sufficient for using state estimates directly or as independent variables for most research purposes.

As a final summary measure, we ask how often MRP "beats" disaggregation. We calculated this in two ways. First, for each state estimate (that is, for each state in each run of a simulation), we score whether the MRP estimate or the disaggregation estimate comes closer to the "true" value. Next, we scored whether the average absolute error across states within a simulation run was smaller for MRP or disaggregation—if a researcher used this one simulated data set, would she or he be better off with MRP or disaggregation? The final panel in Figure 2 plots the percentage of estimates for which MRP improves on disaggregation, comparing estimate-by-estimate and simulation-by-simulation. For individual state estimates, in the smallest samples, MRP "wins" 73% of the error comparisons. Even in the large 50% sample, MRP yields smaller errors 57% of the time. If we look at which did better with simulated data sets as the unit of comparison, MRP "won" 99% of matchups in the smaller samples and still "won" 83% in the largest samples.

Overall, the two methods mostly converge in the largest samples, but otherwise MRP yields clear advantages in reducing error size, in reducing the variance of predictions, and in improving correlation to the "true" state opinion levels. Even in the largest samples, MRP estimates are more stable and errors are smaller a clear majority of the time. MRP does better at predicting the results of disaggregation in the baseline sample than disaggregation itself.

How Good a Response Model Do You Need?

The results above show that there are clear gains from MRP over simple disaggregation, particularly in smaller samples. To what can we attribute such gains? They could be due to partially pooling observations across states, given the use of a multilevel model. That is, if estimates from less populated states are pooled somewhat towards the national mean, that alone might produce better

estimates due to smoothing. Or the gains from MRP could be due to use of demographic or geographic predictors which allow for a more accurate model of individual responses. That is, the demographic information or the geographic information or their interaction could be responsible for the bulk of the gains. We next seek to apportion these gains across these potential contributors. This also allows us to explore another question: how complicated a multilevel model is needed to achieve a reasonable correlation to "true" state opinion?

We consider four possible MRP models, along with disaggregation (i.e., raw state estimates). The first possibility is MRP including only demographic predictors. No state-level modeled effects are used, so that states are allowed to vary only in demographic composition, and we do not include the state-level religion variable. This resembles the older-style simulation that ignored nondemographic differences across states.

Second, we consider a model that uses only geographic predictors, in the form of state and region modeled effects (again excluding the state-level religion variable). This is similar to disaggregation, except that we partially pool states towards the national mean, to an extent determined by the size of the state sample (e.g., if we only have two respondents in Wyoming, its estimate is strongly smoothed towards the national mean). This will determine to what extent simple partial pooling yields the gains we found, as opposed to full modeling and poststratification by demographic type.

Third, we use a simplified version of our full multilevel model, including the state-level model (including region), but only including race, and not the remaining demographic predictors (such as education or religion). (We choose race as it is a demographic predictor that varies greatly from state to state.) This can show how complex the demographic partitioning need be to achieve reliable estimates. Finally, we use our full model as above.

We run 200 simulations, applying each method to 10% of the data, and using the remaining 90% to define the baseline measure of opinion, again using disaggregation. Figure 3 shows the correlation of the estimates from each method to baseline opinion. The demographics-only model does quite poorly, a finding strongly in accord with that of Erikson, Wright, and McIver (1993). Geography clearly matters; the demographic predictors used do not come close to capturing all of the variation across states.

As compared to raw state totals, the partial pooling in the geography-only model yields modest gains, with the correlation increasing by .03. In comparison, the correlation increases by a further .11 when even just "race" is included in the model. Using the full multilevel model increases the correlation by .06 to the full .82. If given

FIGURE 3 Correlation by Model Complexity

Correlation by Model Complexity
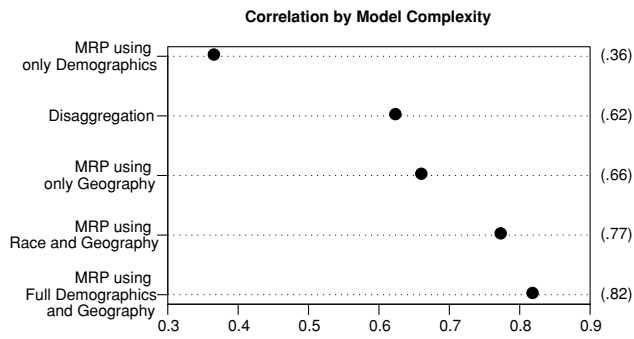
Using 90% of the data disaggregated by state as the baseline, we apply MRP to the remaining 10% to estimate state opinion, using four models of varying complexity. We show the correlation of each set of MRP estimates to the baseline estimate, along with the correlation using disaggregation on the 10% sample. The models are ordered by increasing correlation. Values plotted are indicated along the right axis.

the choice between demographics and geographics, the latter seems more important to include—but the benefits of using both are clear in these results. Even a simple demographic model, in combination with state modeled effects, can do quite well in measuring opinion. That said, the addition of further demographic information does yield increasingly more accurate predictions.

In short, the bulk of gains from MRP over disaggregation are achieved *not* from simple pooling of states towards the national mean, but due to the more accurate modeling of responses by demographic-geographic type. We, of course, recognize that this is not the last word on the subject but rather that others might wish to explore how much further one can push the envelope and to assess how large the gains are from MRP in varying contexts (we begin this process in the replication section below).

The size of the gains will depend on the quality of the demographic model for the issue at hand. Certainly, the researcher should make sure to include any demographic variables thought to be relevant in the policy area being studied. If there were no or very weak demographic correlations, one would still get the benefits of partial pooling the state-level effects, which will improve upon raw disaggregation to an extent based on sample size and heterogeneity.[24] This alone can lead to strong results as shown by Figure 3. Any demographic correlations one

[24]Partial pooling of the state effects towards the national mean should only be a problem if, after controlling for region and any demographic effects, the state effects are bimodal, etc. If this is controlled for, by including further state-level predictors or groupings, then this problem is alleviated. One should be able to anticipate such issues.

can find using one's subject-area expertise will improve estimation from there, and it seems unlikely that there will be many policy areas without *any* demographic correlations. One should seek a "good" demographic (and geographic) model of the individual response, but Figure 3 also shows we do not need a perfect one and that even a limited set of demographic correlates can strongly improve estimation.

Further interaction terms could also be included, and, at the state level, we could possibly improve precision by including additional state-level information (e.g., median income). If we expect demographic effects to differ across states or regions (say, if the "effect" of being black on opinion may differ between the South and elsewhere), they can be allowed to do so in the multilevel model (whereas for now we have assumed independence between demographic effects and geographic effects). While complicating the model somewhat and possibly requiring larger samples, this could improve predictive success further. This allows the researcher to capture differences in politics across geographic areas, even though we did not find this necessary here. (They could also be allowed to vary over time.) A more complete model would allow the researcher to better study substantive demographic effects, of course, even if omitting them would not unduly affect estimates of aggregate state opinion.

Estimation Using a Single National Poll

The previous section strongly suggested the plausibility of using MRP from a single national survey. In this section, we explore this possibility further. How well can estimates produced from a single national poll do? Are they sensible results? And how good are the estimates in predicting support for same-sex marriage as measured by actual state-level polls?

Face Validity

We first present some representative estimates. We use the methods above on four Pew polls from 2004. Figure 4 presents our estimates, by state, of support for same-sex marriage using each of these four polls. For reference, the raw state-by-state disaggregated support levels are shown in the left panel. In the right panel, the four estimated support levels are shown for each state. The states are listed from top to bottom in descending order of their mean estimated support for same-sex marriage. While there is some variation in state-level estimates across polls,

## FIGURE 4 Representative Estimates of State Opinion
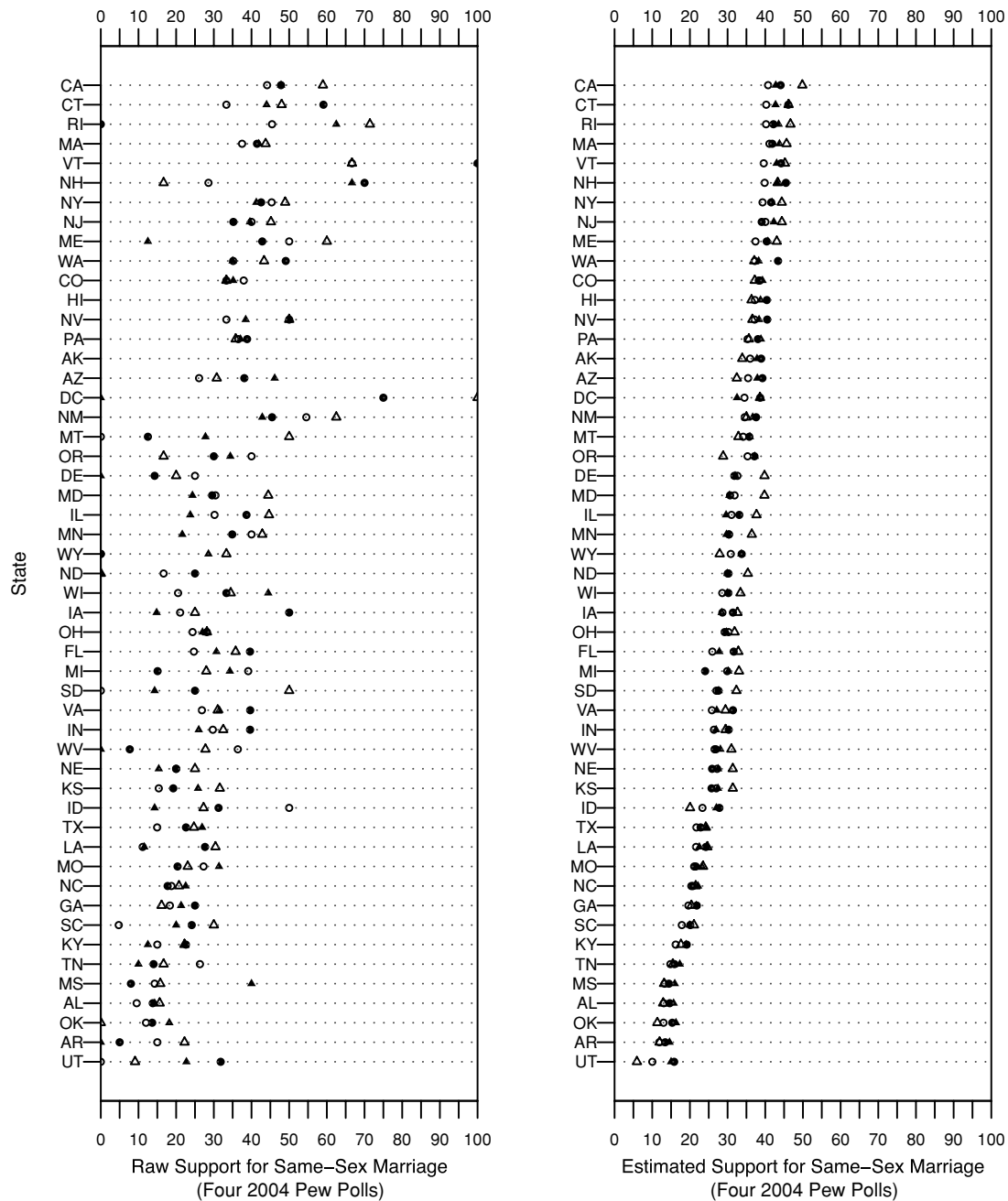
The left-hand panel shows the raw levels of support across four 2004 Pew polls, disaggregated by state. The right-hand panel shows the MRP estimates from each of these four polls. States are ordered by mean estimated support for same-sex marriage. The polls were taken in February (○), March (△), July (●), and December (▲).

which is natural given that the polls span nearly a calendar year and show variation at the national level, the results are generally quite consistent.

Additionally, the results have a great deal of face validity. The states near the top, not surprisingly, are California, Connecticut, Rhode Island, and Massachusetts. In fact, of

the top 10 states, all but two have some form of recognition of same-sex relationships. The states that are least supportive of same-sex marriage are the southern states plus Utah. Of the bottom 10 states, all prohibit same-sex marriage and, with the exception of North Carolina, each does so at the constitutional level.

FIGURE 5 Predicting State Polls
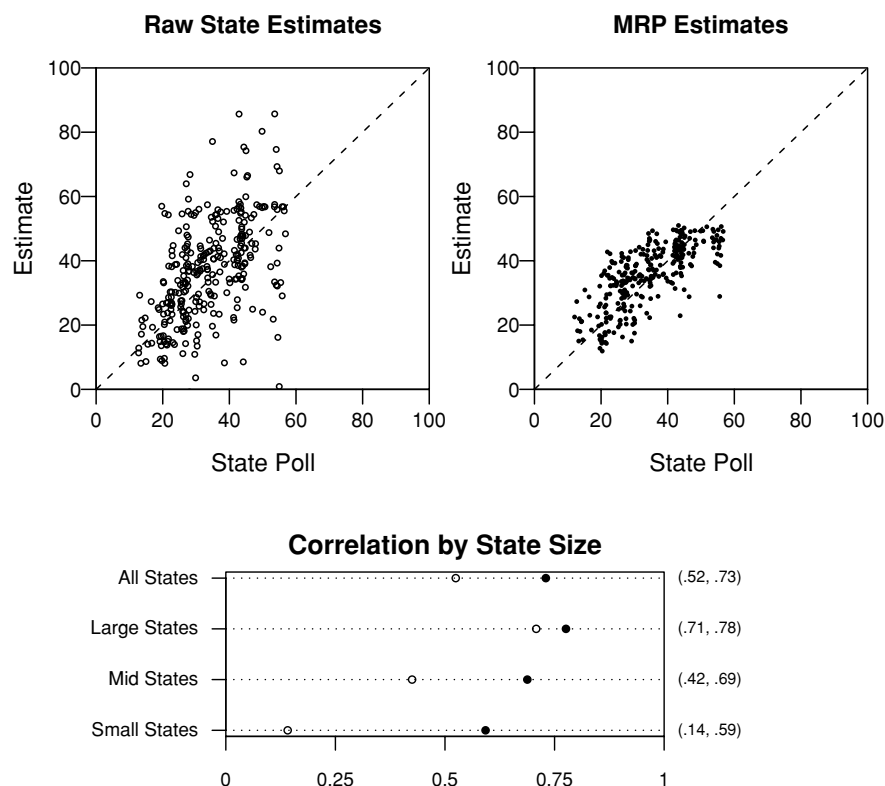
We match state polls to estimates in those states using individual national polls. The top-left panel shows the state polls against the raw estimates from disaggregating the national polls. The top-right panel shows the MRP estimates against the actual state polls. In each, the dotted line represents a perfect correlation. Finally, in the bottom panel, we show the correlation between actual state polls and MRP using individual national polls (●) and the correlation between the state polls and raw state support (○). The correlations for all states are shown on the right, whereas on the left the states are split into thirds by population. Values plotted are indicated along the right axis.

External Validity: Predicting State Polls

In addition to the set of national surveys used above, we have gathered the results of 75 actual state-level surveys on support for same-sex marriage. These were located using news archives and interest group websites. Note that we have multiple polls from some states and no polls for other states, so that 37 states are covered in all. As can be seen, state-level surveys are relatively sporadic, even on a politically relevant issue such as same-sex marriage. We do not, however, detect any troubling biases in terms of which states were polled.

For each state poll, we took each national poll in the same year of sufficient size (roughly 1,000 or more) and used it to generate a prediction using MRP.[25] We

also calculated the disaggregated state-by-state percentages within each national poll. We then compared our prediction to the state poll.

The top-right panel in Figure 5 plots the actual state polls against MRP's predictions, with the raw state percentages in the top-left panel (points are jittered slightly for clarity). We have added a reference line for perfect correlations. Our predictions are tightly clustered around this line, indicating a strong correlation between these estimates of true state opinion. The mean error was only 6% and the MRP estimates fall within each state survey's margin of error 44% of the time.[26]

---

[25]We combined the Gallup polls from March 18 and April 29, 2005, to reach this threshold. Since we use no 2006 national polls, we estimate the 2006 state polls with the 2005 national polls.

[26]We can also see how much MRP improves upon raw state data, at the level of a single national poll, given that the multilevel model incorporates demographic and geographic correlations. The mean error in the raw data, for example, was 11%, and the raw percentages only fall within the state margins of error 30% of the time. The MRP error was smaller than or equal to the raw estimate 73% of the time. The median reduction of error was 36%.

The final panel in Figure 5 shows the correlations between the estimates and the state polls, for all states, and broken down into thirds by population. The solid circles show the correlation for the MRP estimates, with the open circles showing the correlation for the raw state percentage. Overall, the MRP estimates correlate at .73 to the actual state polls, whereas the raw state data only correlate at .52.

As is expected, correlations are higher for larger states, for which the national sample offers a reasonably sized sample by state. Note that the MRP estimates improve upon the raw state estimates and that the difference is larger for the smaller states. The contribution made by the demographic information, along with the multilevel model's pooling of information across states, is quite clear. MRP on even a single, large national survey correlates very strongly with actual state poll results. Correlations this high are likely to be sufficient for using these state estimates as independent variables in models of policy-making and the like, and for drawing conclusions about opinion itself.[27]

Replications

One might be concerned that there is something idiosyncratic about the same-sex marriage data that led to an advantage for the MRP estimates, despite the theoretical arguments as to why these estimates should work well. To allay these concerns, we replicated our simulations for other survey questions. First, we used other survey data on gay rights issues, including civil unions, sodomy laws, employment protection, and adoption rights. The results were quite similar. The magnitude of the gains varied by question and sample size, but the errors in the MRP estimates were smaller on average across sample sizes, the errors were smaller in a large majority of the estimate-by-estimate comparisons across simulations, the standard deviation of estimates was smaller, and the correlations to the baseline samples were higher (results available upon request).

Next, we turned to data completely distinct from gay rights issues, poll data on the 1988 presidential election. The data are those analyzed in Park, Gelman, and Bafumi (2006), from seven preelection national tracking polls

conducted by CBS/NYT in the nine days preceding the election, for a total sample size of 11,566. The dependent variable is support for Bush (over Dukakis), coding "leaners" as supporters. The multilevel model remains the same with one exception. Given that the data did not identify Hispanic respondents, we use only four race-gender categories (male or female by white or black). We used census data from 1990. We ran simulations for random sample sizes of 10% ($N \sim 1,150$), 25% ($N \sim 2,900$), and 50% ($N \sim 5,800$), again keeping a random 50% sample to define the baseline by disaggregation. (Figure 3 suggests this will not unduly hurt the MRP estimates.)

Results were again similar to those for same-sex marriage. The following results parallel those in Figure 1. Comparing estimates to those from disaggregation on the baseline sample, the MRP and disaggregation errors were 8.0 and 13.4, respectively (for the 10% samples), 7.4 and 9.4 (for the 25% samples), and 6.9 and 7.5 (for the 50% samples).[28] Again, MRP's mean errors were consistently smaller, the standard deviations of estimates within states were consistently smaller, the correlations were consistently higher, and MRP "won" the majority of estimate-by-estimate matchups. If we compare by simulation, focusing on which method had the lower mean absolute error within each simulation run, MRP "won" 100%, 95%, and 80% of such runs across the three sample sizes. Even in the largest of these, the researcher would have been better off using MRP.

We next assessed external validity by comparing the MRP and disaggregation estimates from our simulations to the *actual* Bush vote shares by state, with MRP again outperforming disaggregation and doing well in absolute terms. The pairs of mean absolute errors for the three sample sizes were (5.0, 12.8) at 10%, (4.3, 8.6) at 25%, and (3.9, 6.5) at 50%. The correlations were (.52, .37), (.63, .50), and (.72, .64). The MRP errors were smaller in 73%, 70%, and 68% of estimate-by-estimate comparisons across sample sizes. And, if we measure the mean absolute error for the estimates from each simulated data set, MRP had the smaller mean absolute error in 100% of them. If we consider a simulation run to be a single study, a researcher would always have been better off using MRP.

Finally, we created estimates from the entire set of surveys to compare to the actual vote shares. The mean absolute error from MRP was 3.5 as compared to 5.2 from disaggregation, a 32% reduction in mean error even in this large set of data. The correlations were .78 and .74, respectively (a 6% gain for MRP). MRP yields smaller errors in 34 of 48 states. Thus, even with the full sample, MRP improves estimation and predictive success.

---

[27]Furthermore, these are only the correlations to another noisy measure of actual opinion, suggesting that the correlation to the true opinion level is likely to be even higher. That is, sampling error in the actual state polls alone will reduce correlation, even if our estimates were perfectly correlated with actual opinion. And some state polls are of registered voters whereas others are of all voting-age residents. These are all sources of error such that we might be understating the correlation to true state opinion.

[28]Given tiny sample size, we omit DC from these error calculations.

These results strongly suggest that the gains from MRP were not unique to the same-sex marriage data or other gay rights data. Again, the MRP estimates from samples even as small as the size of a single national poll do quite well—indeed, they are competitive with those from disaggregation on a sample many times larger.

Conclusions

This article addresses a crucial question in the public opinion literature: Which of the available statistical techniques should be employed to estimate state-level public opinion? In particular, we compare the pooling of national surveys as pioneered by Erikson, Wright, and McIver (1993) to the modified simulation approach more recently developed by Park, Gelman, and Bafumi (2006). Thirty years ago, Erikson correctly described the simulation approach as a "second-best substitute for the real thing" (1976, 25). We no longer believe this to be the case.

Three principal findings emerge from our analysis. First, the results show that when working with smaller numbers of survey respondents MRP clearly outperforms disaggregation. It consistently produces estimates of public opinion that are both more accurate and more robust (these benefits are shown to be due to the joint contribution of demographics and geography). This is especially true when predicting opinion in small- and medium-sized states. Second, while disaggregation and MRP perform similarly when using large samples (14,000 or so survey respondents), both producing reasonably accurate estimates of public opinion, MRP estimates still tend to beat disaggregation estimates and are more reliable.

Finally, and perhaps most importantly, MRP does well in an absolute sense regardless of sample size. This approach, if implemented using a single, large national survey, produces estimates of state-level public opinion that are virtually as accurate as those it generates using 10 or more surveys. (Note that these gains from MRP exist even when we use disaggregation itself to set the baseline measure of opinion.)

In other words, a sample of approximately 1,400 respondents or more can produce respectable estimates of opinion, such that the correlation to actual state opinion should be sufficiently high. This can save researchers time, money, and effort. If a higher correlation is needed, if the researcher desires even more accurate measures of opinion, additional data should be gathered for the multilevel model and poststratification. Additional data will also allow the researcher to study more nuanced demographic effects and more complicated interactions between demographic and geographic predictors. The gains from MRP

may vary by context and given how well the response model used captures opinion.

A number of recommendations concerning the use of MRP are suggested by our findings. Most obviously, MRP ought to be utilized when the number of national-level surveys on a given topic is limited. MRP is also appropriate when researchers suspect temporal instability in public opinion. One can accurately estimate current public opinion by simply disregarding older surveys and instead using the most recent one or two. Or, one can use year effects or temporal interaction terms. Researchers can deal similarly with question wording differences, by controlling for individual polls. (These solutions are not available to those scholars employing disaggregation.)

If a very large number of national surveys is available and opinion is stable over time, either technique can be employed, since each will produce similar and fairly accurate results. Even if errors are slightly larger, disaggregation may be preferable to some researchers, however, due to its ease of implementation. Because this approach generates estimates of state-level public opinion without poststratification, researchers do not need to collect population frequency data from the U.S. Census Bureau. In it simplest form, disaggregation requires few specialized methodological skills and can be calculated in any statistical package. In comparison, the MRP approach requires knowledge of hierarchical modeling and statistical packages that can readily implement such techniques.

On the other hand, many researchers may prefer MRP despite its added costs. First, under normal conditions, MRP should at worst converge to disaggregation, even if it does not improve upon it. Second, we have shown that less complicated typologies can be used without significant loss of predictive accuracy. Third, MRP produces much more information that may be of interest. In particular, it provides insights about the determinants of public opinion and the degree to which state variation is based on demographic characteristics versus residual cultural differences. Finally, it also allows for an expanded N—for example, Alaska and Hawaii are usually not surveyed in national polls and therefore opinion in these states cannot be measured using disaggregation. MRP allows estimates to be created for these states, as well as for less populated states more generally.

Overall, the results above have significant implications for the study of public opinion at the state level. Our finding that MRP performs equally as well with small and large samples of survey respondents suggests that MRP can greatly expand the number of issues for which scholars can estimate state opinion and the nuance with which they can do so. Thus far, researchers have had to limit themselves to those questions which

have been asked in dozens of compatible national surveys. These tend to be questions asking voters about their general attitudes or ideology as opposed to their opinions on specific policy issues. As a result, state-level opinion research has focused almost exclusively on general attitudes. Using the MRP approach, scholars should now be able to measure opinion across a large set of specific policy concerns. This will greatly enhance research into the responsiveness of state governments. Additionally, since MRP can effectively be used with relatively little data and simple demographic typologies, it can also be applied to studies of public opinion over smaller time periods or in smaller geographic units, such as congressional districts or school districts, for which detailed demographic data are limited, or for other subsets of the population.

References

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

Berkman, Michael B., and Eric Plutzer. 2005. *Ten Thousand Democracies: Politics and Public Opinion in America's School Districts.* Washington, DC: Georgetown University Press.

Brace, Paul, Kellie Sims-Butler, Kevin Arceneaux, and Martin Johnson. 2002. "Public Opinion in the American States: New Perspectives Using National Survey Data." *American Journal of Political Science* 46(1): 173–89.

Cook, Timothy E. 1999. "The Empirical Study of Lesbian, Gay, and Bisexual Politics: Assessing the First Wave of Research." *American Political Science Review* 93(3): 679–92.

Deville, J., C. Sarndal, and O. Sautory. 1993. "Generalizing Raking Procedures in Survey Sampling." *Journal of the American Statistical Association* 88(421): 335–54.

Efron, Bradley, and Carl Morris. 1975. "Data Analysis Using Stein's Estimator and Its Generalizations." *Journal of the American Statistical Association* 70(350): 311–19.

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

Fay, Robert E., III, and Roger A. Herriot. 1997. "Estimates of Income for Small Places: An Application of James-Stein Procedures to Census Data." *Journal of the American Statistical Association* 74(6): 269–77.

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.

Gibson, James L. 1992. "The Policy Consequences of Political Intolerance: Cultural Conformity and Political Freedom." *American Political Science Review* 86(2): 338–56.

Haider-Markel, Donald P., and Kenneth J. Meier. 1996. "The Politics of Gay and Lesbian Rights: Expanding the Scope of the Conflict." *The Journal of Politics* 58(2): 332–49.

James, W., and C. Stein. 1960. "Estimation with Quadratic Loss." In *Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability 1*, ed. J. Neyman. Berkeley: University of California Press, 361–80.

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.

Lax, Jeffrey R., and Justin H. Phillips. 2008. "Gay Rights in the States: Public Opinion and Policy Responsiveness." Presented at the annual State Politics and Policy Conference, Philadelphia.

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

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.

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, 209–28.

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

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

Weber, Ronald E., Anne H. Hopkins, Michael L. Mezey, and Frank J. Munger. 1972. "Computer Simulation of State Electorates." *Public Opinion Quarterly* 36(4): 549–65.