How Should We Measure District-Level Public Opinion on Individual Issues?

Christopher Warshaw   Stanford University
Jonathan Rodden   Stanford University

Due to insufficient sample sizes in national surveys, strikingly little is known about public opinion at the level of Congressional and state legislative districts in the United States. As a result, there has been virtually no study of whether legislators accurately represent the will of their constituents on individual issues. This article solves this problem by developing a multilevel regression and poststratification (MRP) model that combines survey and census data to estimate public opinion at the district level. We show that MRP estimates are excellent predictors of public opinion and referenda results for both congressional and state senate districts. Moreover, they have less error, higher correlations, and lower variance than either disaggregated survey estimates or presidential vote shares. The MRP approach provides American and Comparative Politics scholars with a valuable new tool to measure issue-specific public opinion at low levels of geographic aggregation.

T he aggregation of citizens' preferences into policy lies at the heart of democracy. In countries like the United States, where legislative representation is based on plurality elections in single-member constituencies, it is important to measure policy preferences at the level of electoral districts in order to make progress in answering even the most basic questions about representation. One of the most central questions in the study of democratic politics is whether the activities of individual legislators reflect the preferences of their constituents on salient issues. In order to approach this question, it is necessary to have a reliable way of characterizing each district's preferences. If a legislator voted in favor of funding for stem cell research, it would be useful to know whether her constituents also favor this policy.

Unfortunately, previous empirical work has been stymied by the fact that the sample size in national surveys is generally too small to make inferences at the district level. Scholars have adopted various techniques to cope with the sparse availability of district-level data. Some have adopted the strategy of employing the district-level presidential vote as a catch-all proxy for district public opinion (Canes-Wrone, Cogan, and Brady 2002). Others have disaggregated national surveys to the district level (Clinton 2006; Miller and Stokes 1963). Still others have employed simulation techniques or Bayesian models based in part on district-level election data (Levendusky, Pope, and Jackman 2008).

The goal of such efforts is often to come up with broad, one-dimensional measures of partisanship or overarching ideology. Of course such encompassing measures can be useful for some questions, but they are poorly suited to the class of questions about representation introduced above, which often require relatively precise measures of preferences on specific issues.

An alternative approach developed by Gelman and Little (1997) and Park, Gelman, and Bafumi (2004) builds upon some of the key strengths of these different techniques. This multilevel regression and poststratification (MRP) approach incorporates demographic and geographic information to improve survey-based estimates of each geographic unit's public opinion on individual issues. First, the model incorporates both demographics and geographic information to partially pool data across districts. Next, predictions are made for each demographic-geographic respondent type. Finally, these predictions are poststratified (weighted) using Census data.

This approach has worked well at the state level, and this article extends it to the district level.[1]

---

[1]An online appendix with supplementary material for this article will be available at http://journals.cambridge.org/jop. Data and supporting materials necessary to reproduce the figures and numerical results will be made available at http://www.stanford.edu/~cwarshaw upon publication.

A lingering question about this approach is whether it is actually an improvement over simpler methods, such as disaggregating national surveys or employing presidential vote shares. In their state-level analysis, Lax and Phillips (2009) answer in the affirmative, showing that MRP significantly outperforms disaggregation for estimating public opinion on same-sex marriage. In this article, we develop a relatively simple MRP model to estimate the public opinion of congressional and state senate districts on six salient issues. Then we use a variety of approaches to evaluate whether the MRP estimates outperform disaggregation and presidential vote shares for estimating district-level public opinion.

First, we combine three large surveys to build a "supersurvey" with more than 100,000 respondents and use a split-sample design to compare disaggregated district-level means to MRP estimates of public opinion. We find that MRP significantly outperforms disaggregation for estimating the public opinion of both congressional and state senate districts on all six issues we examine. Next, we compare the performance of MRP and disaggregation for predicting district-level voting behavior on state ballot initiatives. To our knowledge, this is the first validation of a survey-based measure of district opinion on specific policies against related referenda. Once again, we find that MRP clearly outperforms raw disaggregation. Finally, we compare MRP estimates to presidential vote shares. With some caveats, we find that MRP generally yields higher correlations with "true" district public opinion than presidential vote shares.

A strength of the MRP method is its strong performance in relatively small national survey samples. We find that MRP produces very reliable estimates of congressional districts' public opinion with a national sample of just 2,500 people, and it yields reliable estimates for state senate districts with a national sample of 5,000 people. In general, our results suggest that the value of the MRP approach increases as survey sample sizes get smaller relative to the number of districts. Thus MRP may be especially valuable for students of state politics or for scholars who wish to estimate district-level public opinion in countries like Canada, Australia, France, and the United Kingdom, where postal codes in existing surveys often make it possible to place respondents in districts, survey sample sizes are often not terribly large, and census departments produce detailed district-level demographic reports.

The application of MRP to produce reliable estimates of public opinion at the district level provides new tools for a number of research questions. First, it provides public opinion and political behavior schol-

ars with the opportunity to examine the distribution of opinions across districts on specific issues or bundles of issues rather than attempting to make inference about ideology from noisy and endogenous district-level voting data. Second, it provides U.S. and comparative scholars with the ability to develop stronger tests of representation and democracy by comparing district-level public opinion with the attitudes and behavior of elected representatives. While in this article we focus on individual survey items in order to provide the clearest possible evaluation of the MRP approach, researchers can also combine Item Response Theory (IRT) and MRP models to aggregate information across related survey questions in order to develop more accurate estimates of overall district ideology. This approach would have the added benefit of reducing measurement error associated with individual survey items (Ansolabehere, Rodden, and Snyder 2008), making it possible to generate a high-quality mapping of district-level opinion on multiple issue dimensions. Finally, MRP provides comparative politics scholars with a new tool to estimate district-level public opinion in other countries where survey data is sparser than in the United States.

Estimating District-Level Public Opinion

Disaggregation Overview

The most straightforward approach to estimating district preferences is to use data from a representative survey that asks respondents for their preferences on individual issues (Brace et al. 2002; Erikson, Wright, and McIver 1993). Lax and Phillips (2009) calls this approach "disaggregation." The primary advantage of disaggregation is that scholars can estimate district-level public opinion using only the respondent's survey response and district of residence. Thus, it is very straightforward for applied researchers to quickly build disaggregated estimates for each state and district's public opinion.

District-level disaggregation has a long lineage in political science. In their seminal study of legislative representation, Miller and Stokes (1963) used data from the 1958 American National Election Study to estimate policy preferences at the district level. The problem is that national surveys generally do not have enough respondents to develop efficient estimates of voter's preferences at substate levels. Miller and Stokes (1963) study had an average of just 13 respondents per

district (Erikson 1978). Thus, while their estimates of constituency opinion were unbiased, they had extremely large standard errors.

Several recent studies have turned to large-N surveys, such as National Annenberg Election Survey (NAES), Knowledge Networks, and the Cooperative Congressional Election Survey (CCES) to increase sample sizes. For instance, Clinton (2006) combines data on self-identified preferences from surveys conducted in 1999 and 2000 by Knowledge Networks (KN) and the National Annenberg Election Survey (NAES). The two surveys have over 100,000 combined responses. However, there are significantly fewer responses for many specific issue questions.

MRP Overview

An alternative strategy introduced by Park, Gelman, and Bafumi (2004) and Lax and Phillips (2009) is to estimate district-level public opinion using multilevel regression and poststratification (MRP). This approach employs Bayesian statistics and multilevel modeling to incorporate information about respondents' demographics and geography in order to estimate the public opinion of each geographic subunit (see Gelman and Hill 2007 and Jackman 2009 for more about multilevel modeling). Specifically, each individual's survey responses are modeled as a function of demographic and geographic predictors, partially pooling respondents across districts to an extent determined by the data. The district-level effects are modeled using additional district-, state-, and region-level predictors, such as districts' median-income level and states' religiosity.[2] Thus, all individuals in the survey yield information about demographic and geographic patterns, which can be applied to all district estimates. The final step is poststratification, in which the estimates for each demographic-geographic respondent type are weighted (poststratified) by the percentage of each type in the actual district population.

This approach improves upon the first generation of simulation-based methods, pioneered by Pool and Abelson (1961), in a number of crucial ways. First, the earlier simulation approaches relied exclusively on demographic correlations, such that the prediction

for any demographic type was constant across districts. In the words of Pool and Abelson, "A simulated state therefore consisted of a weighted average of the behaviors of the voter types in that state, the weighting being proportional to the numbers of such persons in that state... We assumed that an 'upper-income Protestant Republican rural white male' was the same in either state" (1961, 175). In contrast, MRP allows researchers to address possible geographic "neighborhood" or "contextual" effects by including a rich array of district-level covariates in the first stage of the model, thus taking into account the fact that people in different locales differ in their opinions even after controlling for individual-level demography. Second, MRP is far more sophisticated in the way it models public opinion, using Bayesian statistics and multilevel modeling to partially pool information about respondents across districts to learn about what drives individual responses (Lax and Phillips 2009).

What Do We Know?

The case for MRP is clear, but the proof is in the performance. Lax and Phillips (2009) show that MRP dramatically outperforms disaggregation at the state level for predicting both public opinion and election outcomes. Compared to baseline opinion measures, it yields smaller errors, higher correlations, and more reliable estimates. Moreover, they show that the performance advantages of MRP are even greater for smaller sample sizes: MRP yields relatively reliable estimates of state public opinion with national samples as small as 1,500.

Lax and Phillips' state-level results suggest that MRP is also likely to produce strong estimates at the district level as well, but a number of questions remain unanswered. First, and most importantly, no previous work has evaluated whether MRP produces more accurate district-level estimates of public opinion than disaggregation or presidential vote shares. There are a number of reasons why MRP might fail to do so. The small district-level sample sizes may stymie MRP analysis by producing too much pooling between districts, and geography may affect district-level estimates in complex ways that are difficult to capture through an MRP model.

Second, many previous MRP models use previous election results to help improve state-level public opinion estimates (Park, Gelman, and Bafumi 2004). However, this approach is less suitable at the district level, where researchers are likely to want to use district public opinion estimates as a right-hand side variable to predict election results. If election results

---

[2]Our multilevel model enables us to partially pool respondents in different geographic areas. We facilitate greater pooling by modeling the differences in public opinion across geography using additional district, state, and region-level predictors. This approach stands in contrast to a typical "fixed effect" model with unmodeled factors for each district (Gelman and Hill 2007, 245). Fixed effects are equivalent to a "no-pooling" model which generates predictors for each group using only the respondents in that group (Gelman and Hill 2007, 255; Jackman 2009, 307).

are used in the estimation process for the public opinion estimates, it makes little sense to use them to predict election results in subsequent work.

Third, if MRP does outperform disaggregation, it is important to examine whether the "performance gap" varies for different sample sizes, geographic levels (e.g., congressional versus state legislative districts), or issues. For instance, it is possible that MRP works for congressional districts but not smaller levels of aggregation such as state legislative districts. Finally, existing research has done little to validate district-level MRP estimates of preferences on specific issues against results of actual district-level votes on those issues. Statewide referenda provide an excellent opportunity.

The MRP Model

Data

In order to evaluate MRP at the district level, we develop a large "super-survey" by combining three large-N surveys of the American public: the 2004 National Annenberg Election Survey, the 2006 Cooperative Congressional Survey, and the 2008 Cooperative Congressional Survey.[3] There are six issue questions with similar question wording on at least two of the three surveys: same-sex marriage, abortion, environmental protection, minimum wage, social security privatization, and federal funding for stem cell research. This yields between 70,000 and 110,000 responses for each question.

We then recode the surveys as necessary to combine them into a single dataset:

- For same-sex marriage, responses are coded 1 for support of an amendment to ban same-sex marriage and 0 otherwise ("no," "don't know," or "refused").
- For abortion, responses are coded 1 if the respondent either believes abortion should never be permitted or permitted only in case of rape, incest, or when the woman's life is in danger, and 0 otherwise.
- For environmental protection, responses are coded 1 if they favor environmental protection over the economy, and 0 otherwise.
- For minimum wage, responses are coded 1 for support of increasing the minimum wage to $7.25, and 0 otherwise.

- For social security, responses are coded 1 for support of privatization and 0 otherwise.
- For stem cell research, responses are coded 1 for support of federal funding for stem cell research, and 0 otherwise.

For each respondent, we have an array of demographic information, including sex (male or female), race (black, Hispanic, white, and other), and one of five education categories (less than a high school education, high school graduate, some college, college graduate, and graduate school).[4] We also have information on each respondent's congressional district, state senate district, state, and region. For each district, we have Census data on the percent that live in an urban area, the median income, the percent of the population that are veterans, and the percent of couples that live with a member of the same sex. For each state, we have the percent of evangelical Protestants and Mormons (Jones et al. 2002).

Modeling Individual Responses

MRP models each individual response as a function of both demographic and geographic predictors. It assumes that the effects within a group of variables are related to each other by their hierarchical or grouping structure. For data with a hierarchical structure (e.g., individuals within districts within states), multilevel modeling is generally an improvement over classical regression. A 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 (Gelman and Hill 2007, 254–58; Lax and Phillips 2009). In contrast, 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 endogenously, "with the relative weights determined by the sample size in the group and the variation within and between groups" (Gelman and Hill 2007, 254).

In our MRP model, we estimate each individual's preferences as a function of his or her demographics, district, and state (for individual i, with indexes $r, g, e, d, p, s,$ and z for race, gender, education category,

---

district, poll-year, state, and region, respectively).[5] This approach allows individual-level demographic factors and geography to contribute to our understanding of district ideology. Moreover, the model incorporates both within and between state geographic variation. We facilitate greater pooling across districts by including in the model several district- and state-level variables that are plausibly correlated with public opinion. For example, we include the percentage of people in each state that are evangelicals or Mormons, and the percentage of people in each district in same-sex couples.

We incorporate this information with the following hierarchical model for respondent's responses:

$$\Pr(y_i = 1) = \text{logit}^{-1}\left(\gamma_0 + \alpha_{r[i]}^{\text{race}} + \alpha_{g[i]}^{\text{gender}} + \alpha_{e[i]}^{edu} + \alpha_{d[i]}^{district} + \alpha_{p[i]}^{year}\right) \quad (1)$$

where

$$\alpha_r^{race} \sim N\left(0, \sigma_{race}^2\right), \text{ for } r = 1, \ldots, 4$$

$$\alpha_g^{gender} \sim N\left(0, \sigma_{gender}^2\right)$$

$$\alpha_e^{edu} \sim N\left(0, \sigma_{edu}^2\right), \text{ for } e = 1, \ldots, 5$$

$$\alpha_p^{year} \sim N\left(0, \sigma_{year}^2\right), \text{ for } p = 1, 2, 3$$

That is, each individual-level variable is modeled as drawn from a normal distribution with mean zero and some estimated variance. Following previous work using MRP, we assume that the effect of demographic factors do not vary geographically. We allow geography to enter into the model by adding a district level to the model and giving each district a separate intercept.[6] However, our model could easily be extended to allow the effect of individual-level demographics to vary across districts or states (see Gelman et al. 2008; Jackson and Carsey 2002). For our models, we tested whether allowing the effects of demographics to vary across states changed our estimates of district preferences, and we found very little effect.[7]

The district effects are modeled as a function of the state into which the district falls, the district's

average income, the percent of the district's residents that live in urban areas, the percentage of the district's residents that are military veterans, and the percentage of couples in each district that are in same-sex couples.[8]

$$\begin{aligned}\alpha_d^{district} \sim N\Big(&\kappa_{s[d]}^{state} + \gamma^{inc.} \bullet income_d \\ &+ \gamma^{urb.} \bullet urban_d + \gamma^{mil.} \bullet military_d \\ &+, \gamma^{samesex} \bullet samesex_d, \sigma_{district}^2\Big), \\ &\text{for } c = (1, \ldots, 436)\end{aligned} \quad (2)$$

The state effects are modeled as a function of the region into which the state falls, the percentage of the state's residents that are union members, and the state's percentage of evangelical or Mormon residents:

$$\begin{aligned}\alpha_s^{state} \sim N\Big(&\alpha_{z[s]}^{region} + \beta^{union} \bullet union_u \\ &+ \beta^{relig} \bullet relig_s, \sigma_{state}^2\Big), \text{ for } s = (1, \ldots, 51)\end{aligned} \quad (3)$$

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

$$\alpha_z^{region} \sim N\left(0, \sigma_{region}^2\right), \text{ for } z = (1, \ldots, 4) \quad (4)$$

We estimate the model using the GLMER function in R (Bates 2005).[9]

Poststratification

For any set of individual demographic and geographic values, cell c, the results above allow us to make a prediction of ideology. Specifically, ϑ_c is a function of the relevant predictors and their estimated coefficients.[10] 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 district populations.

---

[5]Since this article is focused on the general applicability of MRP, we chose to deploy a relatively parsimonious model. However, MRP's performance for any particular issue area would likely be improved by using a model with a stronger theoretical basis for linking specific demographic or geographic characteristics to issue stances.

[6]This intercept is, in turn, modeled based on district and state-level demographic factors.

[7]For example, we tested allowing the effect of race and education to vary across state or regions, and we found very little improvement in model fit.

[8]These data were obtained from Census factfinder.

[9]For simplicity, we employ the mean coefficient estimates yielded by GLMER and ignore the uncertainty in these estimates. In other contexts, however, it may be useful to propagate the measurement error in the model's coefficients into the poststratification step to quantify the uncertainty in the MRP estimates of district public opinion.

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

Previous work using MRP at the state level has used either the 1% or 5% "Public Use Microdata Sample" data from the Census (Lax and Phillips 2009; Park, Gelman, and Bafumi 2004). However, the microdata does not include information about respondents' congressional or state-legislative districts. Fortunately, the census factfinder data includes breakdowns by race, gender, and education in each congressional district for the population 25 and over.[11] We use these data to calculate the necessary population frequencies for our analysis.[12]

For our model of congressional districts, we have 436 districts with 40 demographic types in each, which yields 17,440 possible combinations of demographic and district values. For our model of state senate districts, we have 1,942 districts, which yields 77,680 possible combinations of demographic and district values.[13]

Each cell ϑ_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 district, we calculate the average response, over each cell c in district d:

$$y^{mrp}_{districts} = \frac{\sum_{c \in d} N_c \vartheta_c}{\sum_{c \in d} N_c} \qquad (5)$$

Does MRP Outperform Disaggregation?

In this section, we compare MRP and disaggregation estimates for predicting district-level public opinion. First, we use a split-sample validation approach to compare MRP and disaggregation for same-sex marriage. Focusing on same-sex marriage has a number of advantages. It makes our results directly comparable to

[11]Using the population 25 and over to poststratify our results introduces some error into our analysis. But this error is likely minimal since (1) only about 10% of the voting population is under 25 and (2) in most districts the demographic breakdown of the 25 and under population is similar to the demographic breakdown of the 25 and older population.

[12]Because census factfinder does not include age breakdowns for each race/gender/education subgroup, we are not able to control for respondents' age in our model. However, the omission of predictors for age probably does not significantly affect our results. Previous studies using MRP have found little variation among age groups after controlling for other predictors (Park, Bafumi, and Gelman 2004).

[13]Including the District of Columbia, there are 1,943 state senate districts in the country. However, the Census data on state senate districts is missing a district in West Virginia.

the Lax and Phillips (2009) evaluation of MRP at the state level. In addition, all three of our surveys have almost identical questions on same-sex marriage. This enables us to generate a very large sample that makes district-level disaggregation plausible for both congressional districts and state senate districts. Also, the district estimations may be of substantive interest to scholars and policy makers—particularly at the state-legislative-district level. Moreover, as Lax and Phillips (2009) point out, there is significant variation across districts on same-sex marriage, which avoids biasing results towards MRP.[14] Above all, several states have held statewide votes on same-sex marriage, and the availability of district-level tallies gives us a rare opportunity to undertake a second, and perhaps more convincing validation strategy: we contrast the performance of raw survey means and MRP preference estimates in predicting district-level referendum results. Finally, we replicate our split-sample validation strategy for five additional issues, and our referendum strategy for two additional issues.

Split-Sample Validation Analysis of Same-Sex Marriage Estimates

In order to assess the relative performance of the disaggregation and MRP methods in different sample sizes, we rely upon cross-validation (see Lax and Phillips 2009). We randomly split the data, using roughly three-fourths of the data to define the baseline or "true" district public opinion. We define the baseline data for same-sex marriage as the percentage of people in each district that support a constitutional amendment to ban same-sex marriage.[15] We then use some portion of the remaining data to generate estimates of opinion, using both disaggregation and MRP. We draw these random samples 200 times (both the baseline data and the sample data for comparative estimation) for three or four different-sized samples. For congressional districts, the national sample sizes are 2,500, 5,000, 15,000, and 30,000. For state legislative districts, the national sample sizes are 5,000, 15,000, and 30,000. The sample size in particular districts ranges from 0 to roughly 150. We chose 30,000 as the largest sample size in our validation analysis because most large-N surveys top out at about 30,000 responses. Thus, for the time being, 30,000

[14]The partial pooling employed by MRP may be less reliable when the opinion of voters in one area is not useful for predicting the opinion of voters of the same demographic type in other areas after controlling for intercept shifts due to geographic differences.

[15]This yields approximately 85,000 responses.

responses is likely to be the largest sample size available for most applied research questions.

We follow Erikson, Wright, and McIver (1993) and Lax and Phillips (2009) in using unweighted survey responses for both the baseline data and sample data.[16] Similarly to Lax and Phillips (2009), we measure predictive success (how close each set of estimates is to the measure for the baseline sample) in several ways. In each run of a simulation q, let $y_{q,d}^{base}$ be the opinion percentage in district d in the baseline data (again, measured as the disaggregation method does, totaling up the simple percentage by district), let $y_{q,d}^{dis}$ be the disaggregated percentage in district d on the sampled data, and let $y_{q,d}^{MRP}$ be the estimate in district d using MRP.

For each of the sample sizes, we do the following. We first calculate the errors produced by each simulation. The simplest way to measure these errors is using the absolute differences between the estimates and the baseline measure.

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

This forms two matrices of absolute errors, of size 200 (simulations) x d (districts) each. For district d, we then calculate the mean absolute error for each method across simulations.

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

Next, we calculate the mean absolute error over both districts and simulations, collapsing the means-by-district into a simple number for each sample size and method:

$$\bar{e}_d^{dis} = \frac{\sum_q e_{q,d}^{dis}}{200 * D}, \; \bar{e}_d^{MRP} = \frac{\sum_q e_{q,d}^{MRP}}{200 * D} \quad (8)$$

Second, we take the correlations between each set of estimates and the baseline measure for each sample size. Finally, we ask how often MRP "beats" disaggregation. We calculate this in two ways. First, for each district estimate (i.e., for each district in each run of a simulation), we score whether the MRP estimate or the disaggregation estimate comes closer to the "true" value. Next, we score whether the average absolute error

across districts within a simulation run is smaller for MRP or disaggregation. In other words, would a researcher get less error for that sample using MRP or disaggregation?

The left column of Figure 1 shows the results of our performance measures on same-sex marriage for congressional districts.[17] The top-left panel compares the mean absolute errors for each estimation method for various sample sizes. It indicates that the MRP method's mean absolute error is smaller than disaggregation for all four-sample sizes. Indeed, the MRP method has smaller absolute errors with a national sample of just 2,500 than disaggregation does in the largest sample.

Moreover, MRP outperforms disaggregation in individual districts with both small and large sample sizes. In districts with less than 10 respondents, MRP yields roughly 50% smaller errors than disaggregation. In districts with larger sample sizes, MRP yields smaller improvements, and in districts with very large sample sizes of more than 100 respondents, the performance of MRP and disaggregation converge.[18]

The middle panel on the left shows the correlations between MRP estimates and the baseline measure. Not surprisingly, the disaggregated estimates are only weakly correlated with the baseline for smaller sample sizes. In contrast, MRP is correlated with the baseline at 0.80 or better in every sample size. Finally, the lower-left panel shows how often MRP "beats" disaggregation for estimating the public opinion of congressional districts. For individual district estimates, MRP wins 83% of the comparisons in the smallest sample and 60% in the largest sample. If we use simulated datasets as the unit of comparison, MRP wins 100% of the matchups in each sample size.

The right-hand side of Figure 1 displays broadly similar results for identical analysis at the level of state senate districts. The top-right panel shows that the MRP method's mean absolute error is significantly smaller for all four sample sizes. The smaller size of most state senate districts compared to congressional districts magnifies the advantages of MRP; even with a sample of 30,000 people, MRP has a 40% smaller mean absolute error than disaggregation. The middle panel on the right in Figure 1 shows the correlations between MRP estimates and the

---

[16]This approach biases our findings somewhat against MRP since unweighted data are being used both to define both the baseline "actual public opinion" and the disaggregated analysis of the sample data. In contrast, MRP corrects for lack of weighting through poststratification. Thus, some of the differences between the MRP and baseline results could be due to the lack of survey weighting in the baseline data.

[17]In our MRP model, we find that public opinion on same-sex marriage is correlated with gender, education, the median income in a district, the percentage of couples in a district that are in same-sex couples, and the percentage of people in a state that are evangelical. See the online appendix for full model results of our MRP model of same-sex marriage.

[18]See the online appendix for more information.

FIGURE 1 Cross Validation: Summary Performance Measures for Same–Sex Marriage
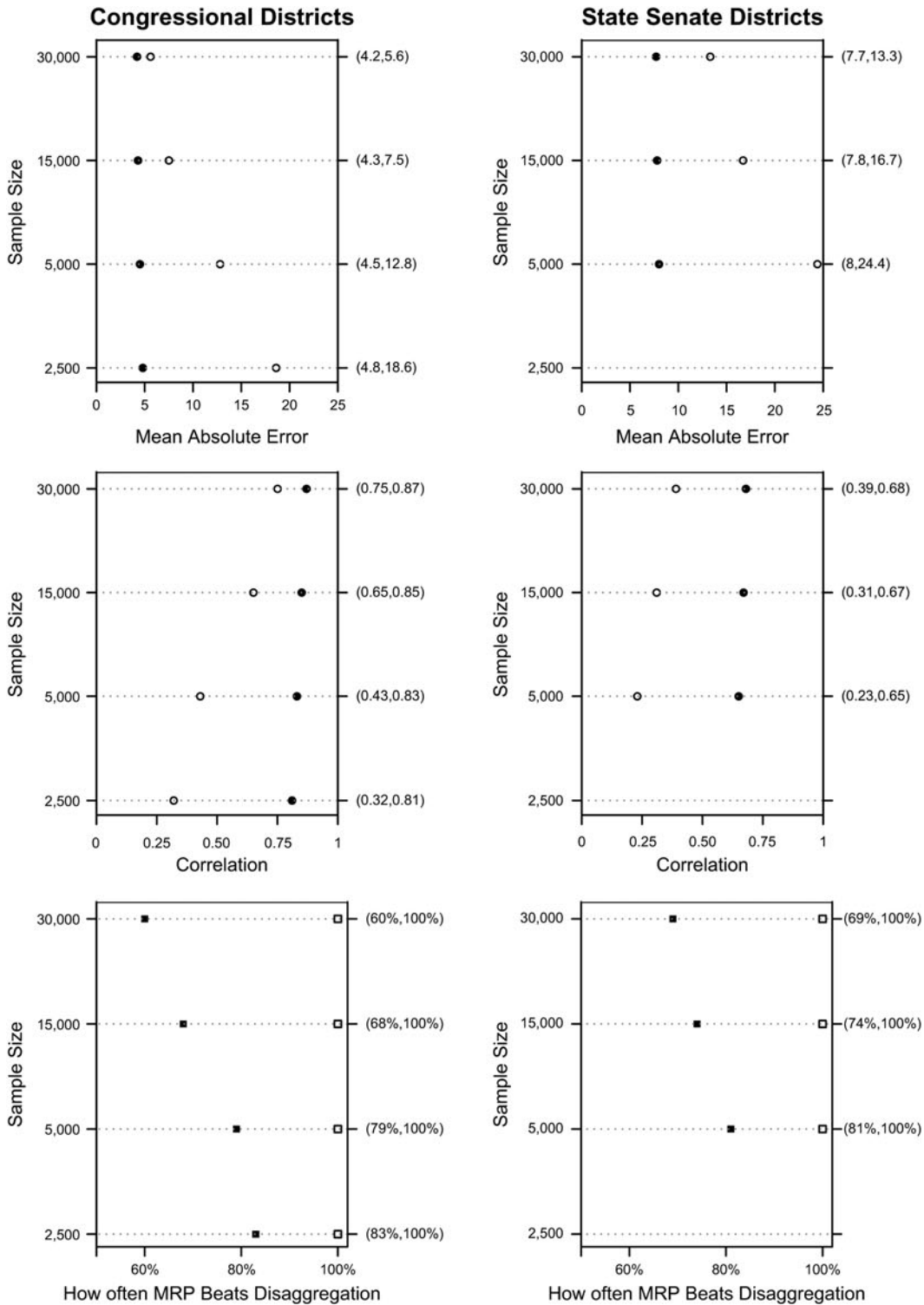
Note: The top panel plots the mean absolute error across districts and simulation runs for MRP (•) and disaggregation (○). The second panel shows the correlation of the MRP and disaggregated estimates with the baseline measures. The bottom panel shows how often the MRP error is smaller than the disaggregation error using (■) each district estimate (across districts and simulation runs) as the unit of analysis and using (□) each simulation run as the unit of analysis (averaging over districts within each simulation run). The national sample sizes are on the left axis, and the values plotted are indicated along the right axis.

baseline measure. MRP has a higher correlation with the baseline than disaggregated estimates at every sample size. Moreover, MRP is correlated with the baseline at 0.65 or better in every sample size. Finally, the lower-right panel shows how often MRP "beats" disaggregation for state senate districts. For individual district estimates, MRP wins 81% of the comparisons in the smallest sample and 69% in the largest sample. If we use simulated datasets as the unit of comparison, MRP wins 100% of the matchups.

External Validation: Same-Sex Marriage Ballot Referendums

We further assess the performance of MRP by comparing the accuracy of MRP and disaggregation for predicting the results of state ballot initiatives on same-sex marriage. We use a question on the 2004 NAES that asks whether respondents support state laws permitting same-sex marriages.[19] We recode the responses as 1 for opposition to state laws permitting same-sex marriage and 0 otherwise. We then compare the district-level results to actual referenda on state constitutional amendments prohibiting same-sex marriage. California, Ohio, and Wisconsin make such data available at the district level, and we were able to build district-level aggregates from the precinct-level results made available by Arizona and Michigan.[20] While these five states represent less state-level variation in size and attitudes than we would like, they are convenient in that all five states have districts with relatively large populations. Thus, any advantage we find for MRP is likely to be magnified in states with smaller districts.

Figure 2 displays scatter plots of referendum results against disaggregated means in the panels on the left and against MRP estimates in the panels on the right. The results are encouraging for scholars who wish to use surveys to gauge district-level preferences of voters. Even the raw means are reasonably correlated with actual district-level votes, but the MRP estimates are clearly a much better predictor of the referendum results than the disaggregated estimates for both congressional and state senate districts. For state senate districts, the MRP estimates show a correlation of 0.76 with the referenda results, while the disaggregated results show a correlation of just 0.51. Similarly, the MRP results have a mean absolute error of just 6.2%, compared to a mean absolute error of 10.0% for the

disaggregated estimates. The results are closer for congressional districts. Here, MRP results have a mean absolute error of 5.0%, while the disaggregated results have a mean absolute error of 6.8%.

Replication for Other Issues

It is possible that same-sex marriage is unique in some way. As a result, we replicated our split-sample validity analysis of same-sex marriage for five other issues: abortion, environmental protection, minimum wage, social security privatization, and stem cell research. Figure 3 shows that the improvements yielded by MRP vary little across issues. For every issue and sample size, the MRP estimates have lower mean absolute error than the disaggregated estimates, and MRP always "beats" disaggregation in terms of total error.

We also examine the relationship between MRP estimates and state referendum results for minimum wage laws (Arizona and Ohio) and stem cell research (California and Michigan).[21] The results, presented in Figure 4, are similar to those for same-sex marriage. For both issues, the MRP estimates are better predictors of referendum results than the disaggregated estimates. The MRP estimates have higher correlations with the referenda results for each issue and state, and generally smaller mean absolute errors.[22]

Does MRP Outperform Presidential Vote Shares?

Even if MRP outperforms disaggregation, presidential vote shares could still be a more reliable correlate of public opinion. Empirical researchers in need of a catchall one-dimensional proxy for district ideology have typically turned to the district-level presidential vote. This strategy has a number of advantages, especially given its potential for meaningful time-series analysis. Yet one of several drawbacks is that presidential voting might be driven by preferences on

---

[19]There were approximately 17,000 respondents on this question.

[20]Most other states make referendum results available at the county level, but the matching of counties to state senate districts is not precise enough for our purposes.

[21]In Arizona and Ohio, we have data on minimum wage referenda that took place in 2006. In California and Michigan, we have referenda data on ballot initiatives to fund stem cell research that took place in 2004 and 2008, respectively. For these issues, we compare the referenda results to the disaggregated and MRP estimates calculated using our 30,000-person sample.

[22]For stem cell referenda, the MRP estimates have mean absolute errors that are 25% smaller than the mean absolute errors for the disaggregated estimates. For minimum wage referenda, MRP and disaggregation yield similar mean absolute errors. The somewhat weaker performance of MRP on the minimum wage referenda may be because these referenda took place in low-turnout off-year elections where the opinions of voters may have differed from those of all adults.

FIGURE 2 Cross Validation of MRP Estimates with Same–Sex Marriage Ref. in Arizona, California, Michigan, Ohio, Wisconsin
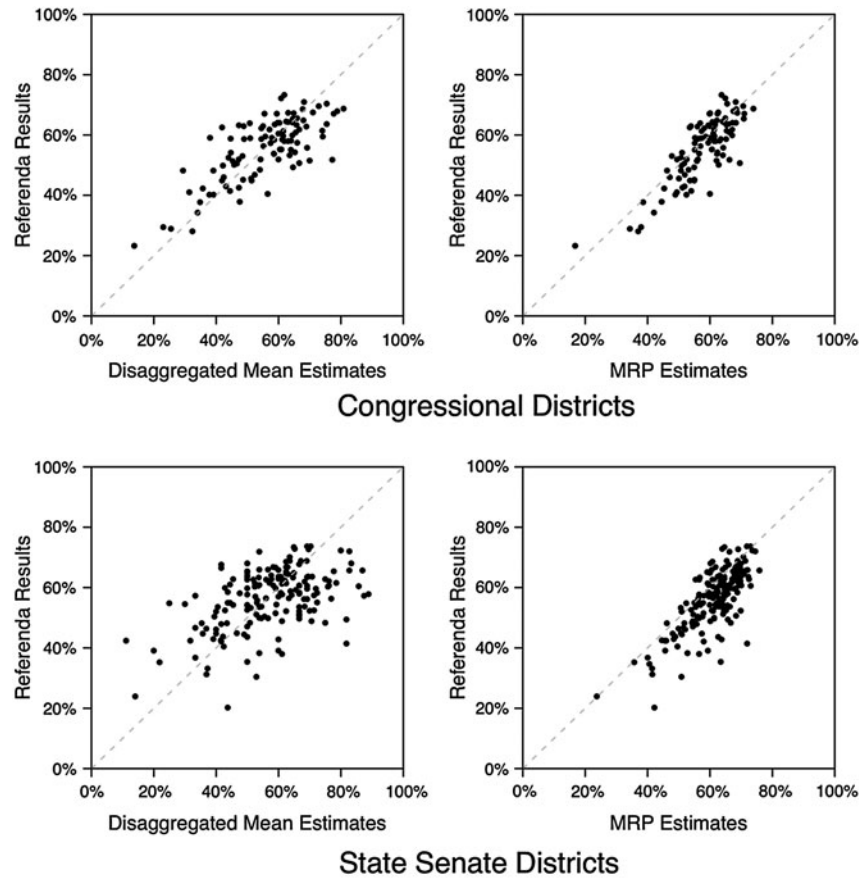
Note: This figure shows that in national samples of 17,000, MRP outperforms disaggregation for predicting state referenda results on same-sex marriage.

multiple issue dimensions, and the salience of these dimensions might vary across districts and over time. Presidential voting might be a useful proxy for district ideology on one dimension and not another.

We are now in a position to evaluate whether MRP estimates of preferences on specific issues display higher correlations with "true" congressional district public opinion on those issues than presidential vote shares, and we can do this for several distinct issues (Figure 5).[23] We find that presidential vote shares generally have a correlation with public opinion between .6 and .7. This is a rather impressive correlation, and it should be somewhat heartening for researchers who wish to continue using presidential vote shares as catchall proxies for district-level ideology, especially those who require time-series variation. Nevertheless, we also find that the MRP estimates generally outperform
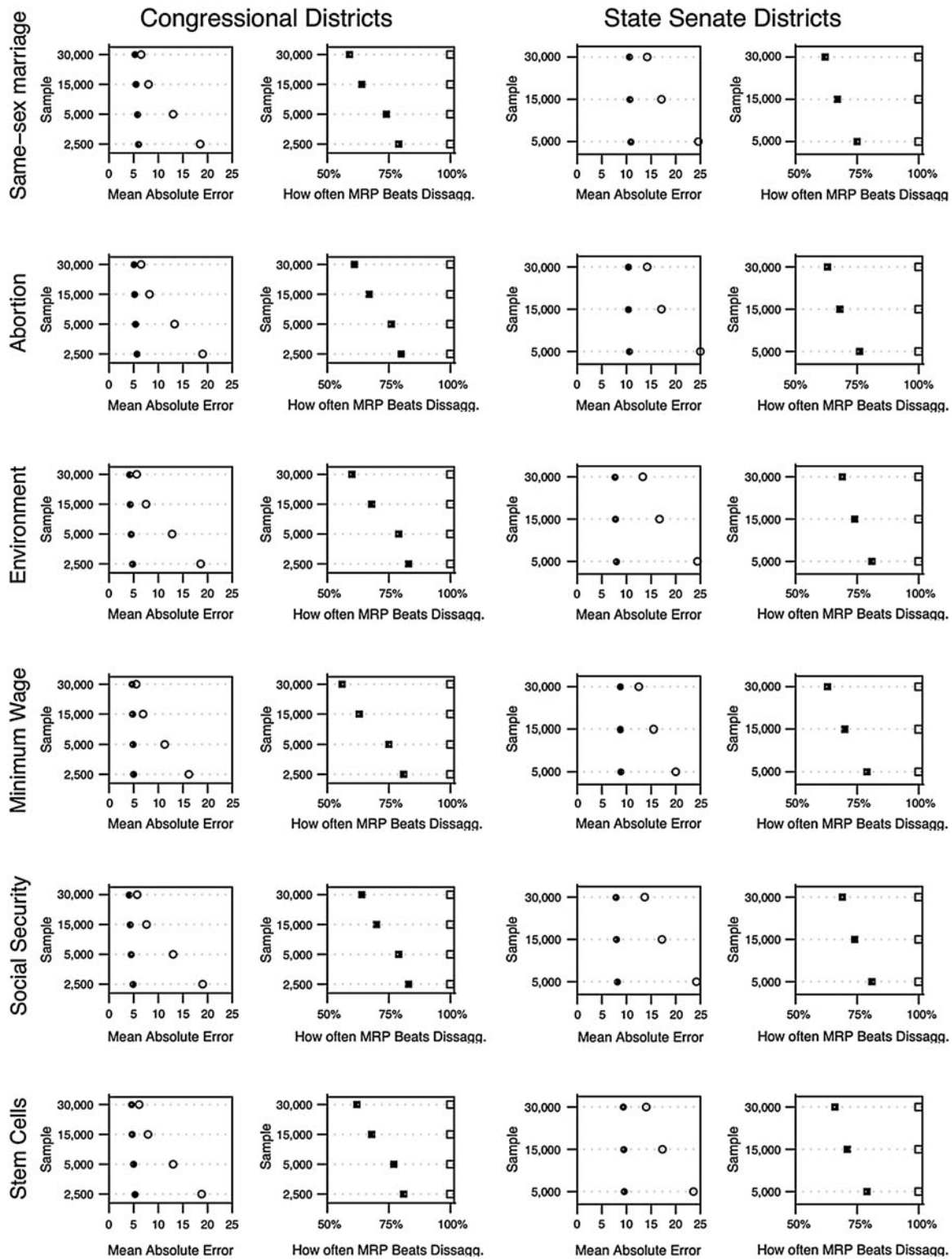
presidential vote shares for estimating public opinion on our selected issues. MRP estimates have substantially higher correlations with "true" public opinion than presidential vote shares at all sample sizes for four of our six issues. On the other two issues (minimum wage and social security), MRP outperforms presidential vote shares in larger samples and presidential vote shares perform better in smaller samples.

## Does Increasing MRP Model Complexity Increase Accuracy?

Our results thus far indicate that MRP generally yields significantly stronger estimates of district public opinion than disaggregation or presidential vote shares. But what causes these gains? The gains could be due to the individual-level demographic predictors in the model. Alternatively, they could be due to the partial pooling of observations across

[23]Unfortunately, data on presidential vote shares are not readily available for state legislative districts. As a result, we focus our analysis on congressional districts.

FIGURE 3 Summary Performance Measures on Six Issues

Note: This figure compares MRP and disaggregation for six issues and four national sample sizes. The left-most column plots the mean absolute error across districts and simulation runs for MRP (•) and disaggregation (◦) for congressional districts. The next column shows how often the MRP error is smaller than the disaggregation error using (■) each district estimate as the unit of analysis and using (□) each simulation run as the unit of analysis (averaging over districts within each simulation run) for congressional districts. The right two columns make the same comparisons for state senate districts.

FIGURE 4   Cross Validation of MRP Estimates with Minimum Wage and Stem Cell Referenda


Congressional Districts

State Senate Districts
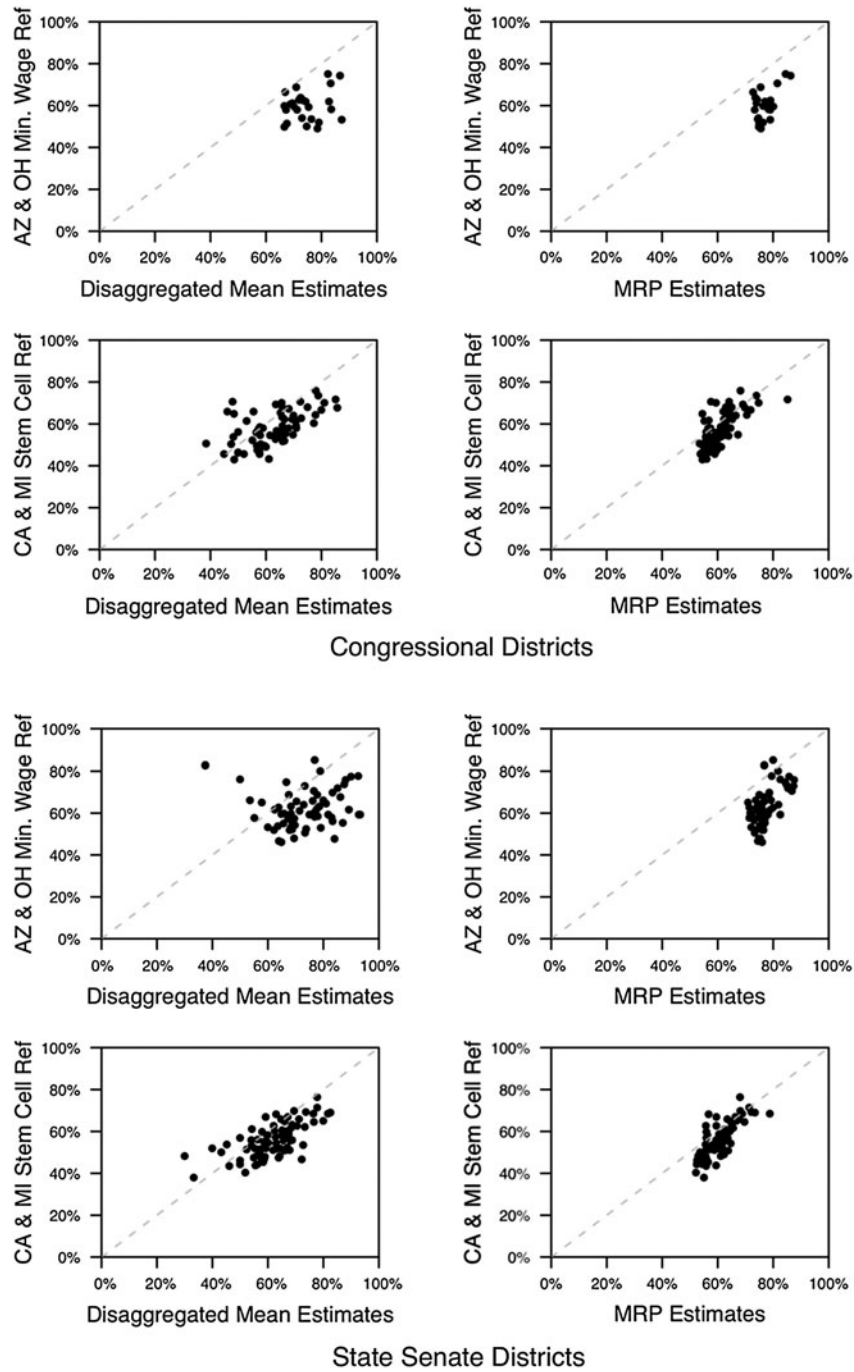
Note: This figure shows that in national samples of 30,000, MRP outperforms disaggregation for predicting state referenda results for minimum wage and stem cell research funding.

districts using the geographic predictors in our multilevel model.[24]

We evaluate four possible MRP models to estimate the public opinion in congressional districts, along with disaggregation, and presidential vote shares (Figure 6). We run 200 simulations, applying each method to a national sample of 5,000 survey respondents. We use the remainder of the sample to measure the baseline public opinion.

FIGURE 5 **Comparing MRP and Presidential Vote Shares: Summary Performance Measures**
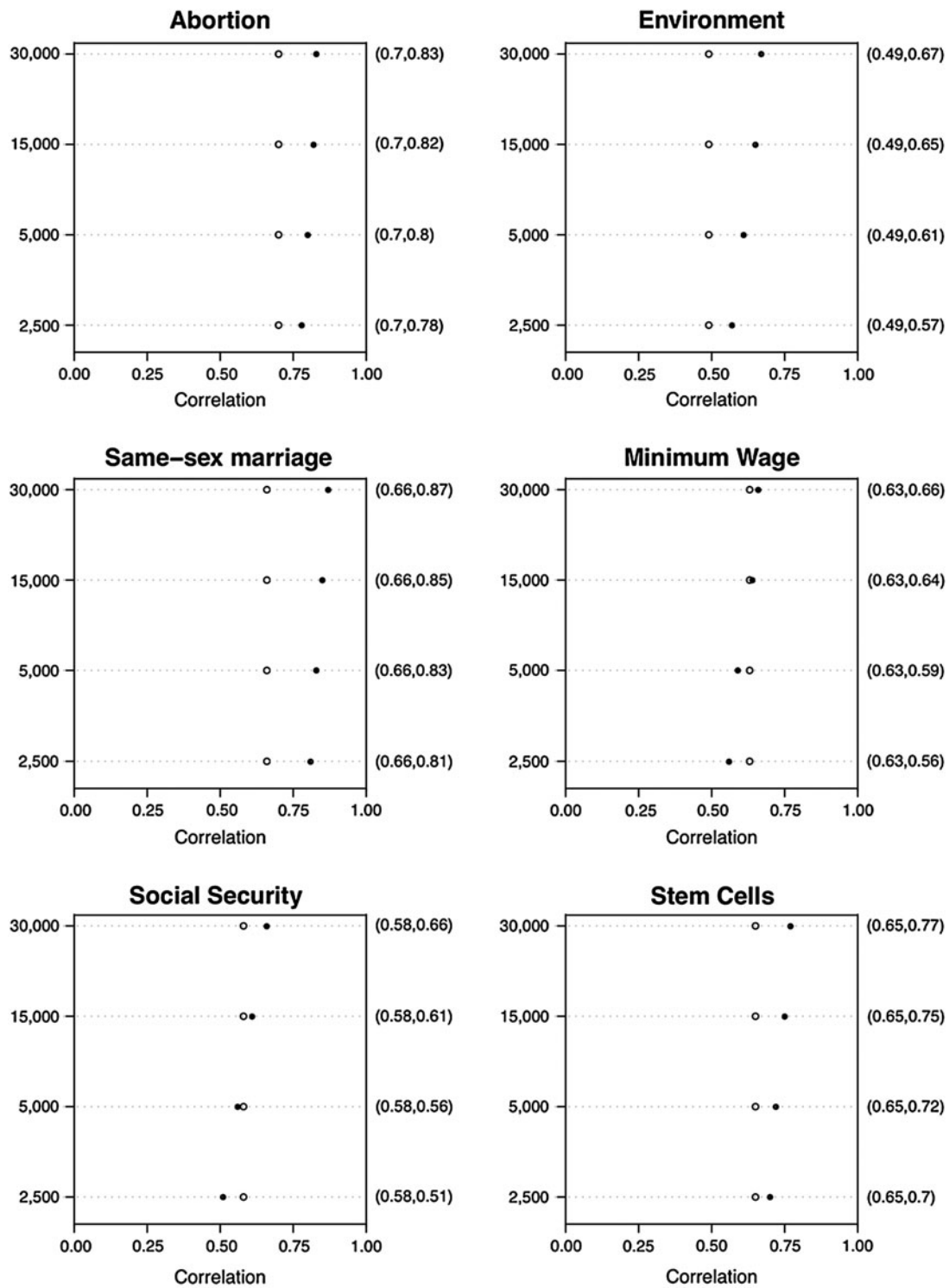
Note: This figure compares MRP estimates and presidential vote shares for six issues and four national sample sizes. For each issue, it plots the correlations of presidential vote shares (○) and MRP (•) with the baseline public opinion in each congressional district.

First, we consider MRP using only individual-level demographic predictors. This model does not include any geographic modeled effects and is similar in spirit to the simulation efforts of the 1960s (e.g., Pool and Abelson 1961) in that the only variation across districts is their demographic composition. We find that a demographics-only model generally outperforms disaggregation, but the improvements are relatively modest.

Second, we consider a model that includes only a simple geographic model, with unmodeled district, state, and region effects. This model allows partial pooling of districts toward the national mean, to an extent determined by the district sample. But the model does not include any district or state-level predictors. We find that the partial pooling in the simple geography model generally yields modest gains compared to the demographics-only model or disaggregation. But the gains are inconsistent across issues.

Third, we evaluate a model that omits individual-level demographic predictors, but includes our full suite of district- and state-level predictors shown in equations (2), (3), and (4) (e.g., the district's median income, percent urban, etc.). This model yields substantial gains on every issue stemming from the inclusion of district and state covariates in our multilevel model.

Finally, we evaluate our full multilevel model that includes both individual-level demographic predictors and multilevel geographic predictors (see equations 1, 2, 3, and 4). We find that the full multilevel model generally increases the correlations with the baseline. However, the gains over the full geography model are relatively modest. Thus, in some contexts in which the researcher has a powerful set of district-level predictors, it may be possible to omit the individual-level demographics entirely. This would reduce the "start-up cost" required to estimate district-level public opinion. Nonetheless, a full MRP model with demographics is appropriate in most applied settings. First, the costs associated with the MRP model are relatively trivial. Second, we find that the full MRP model with demographics yields gains on three of our six issues. Thus, it should increase the accuracy of district-level public opinion estimates for most issues. Third, the demographic variables are often of interest in their own right. For instance, it may be of interest to break down public opinion by race or gender.[25]

Conclusion

This article addresses a crucial question in the study of Congress, state politics, public opinion, and political geography: how should we measure public opinion at the district level? There is no consensus on this important question in the extant literature. Perhaps the most attractive strategy is to obtain a very large sample and take the disaggregated mean or median of the relevant survey response in each district (e.g., Clinton 2006). But this approach falls apart in the smaller datasets that are far more typical in applied research on specific issues. As a result, many applied researchers have simply used district-level presidential votes as a proxy for public opinion. This approach, however, makes it impossible to disaggregate district public opinion into individual issues or issue dimensions or to examine the relationship between district-level preferences and voting behavior.

In this article, we show that MRP yields estimates of issue-specific district public opinion that are consistently superior to disaggregated means or presidential vote shares.[26] Thus, most applied researchers who require an estimate of district-level public opinion on specific issues or bundles of issues should consider employing the MRP approach rather than using disaggregated means or presidential vote shares.

We show that MRP clearly outperforms disaggregation in the estimation of public opinion in congressional districts at even small and moderate sample sizes. At larger sample sizes (30,000+), the difference between MRP and disaggregation is smaller.[27] As a result, given a sufficiently large sample size, some researchers may choose to simply use the disaggregated estimates due to the simplicity and convenience associated with this approach. But even in samples at the limit of most large-scale surveys, MRP consistently outperforms the disaggregation approach and presidential vote shares.

Our results also suggest a number of additional lessons for scholars seeking to use MRP to estimate district-level public opinion in the United States and beyond. First, a potential weakness of MRP is that due to the unavailability of district-level demographics broken down by voters and abstainers, estimates are based on survey responses from all adults rather than voters. Nevertheless, MRP estimates perform well in predicting the opinions of voters as expressed in

---

[25]For example, in the online appendix we show the coefficients for the demographic predictors of public opinion on same-sex marriage.

[26]However, our findings also imply that presidential vote shares may be sufficient for questions requiring a longer time series or where survey data is unavailable.

[27]In national samples significantly larger than 30,000 respondents, disaggregation may sometimes yield better estimates.

FIGURE 6 Correlation by Model Complexity
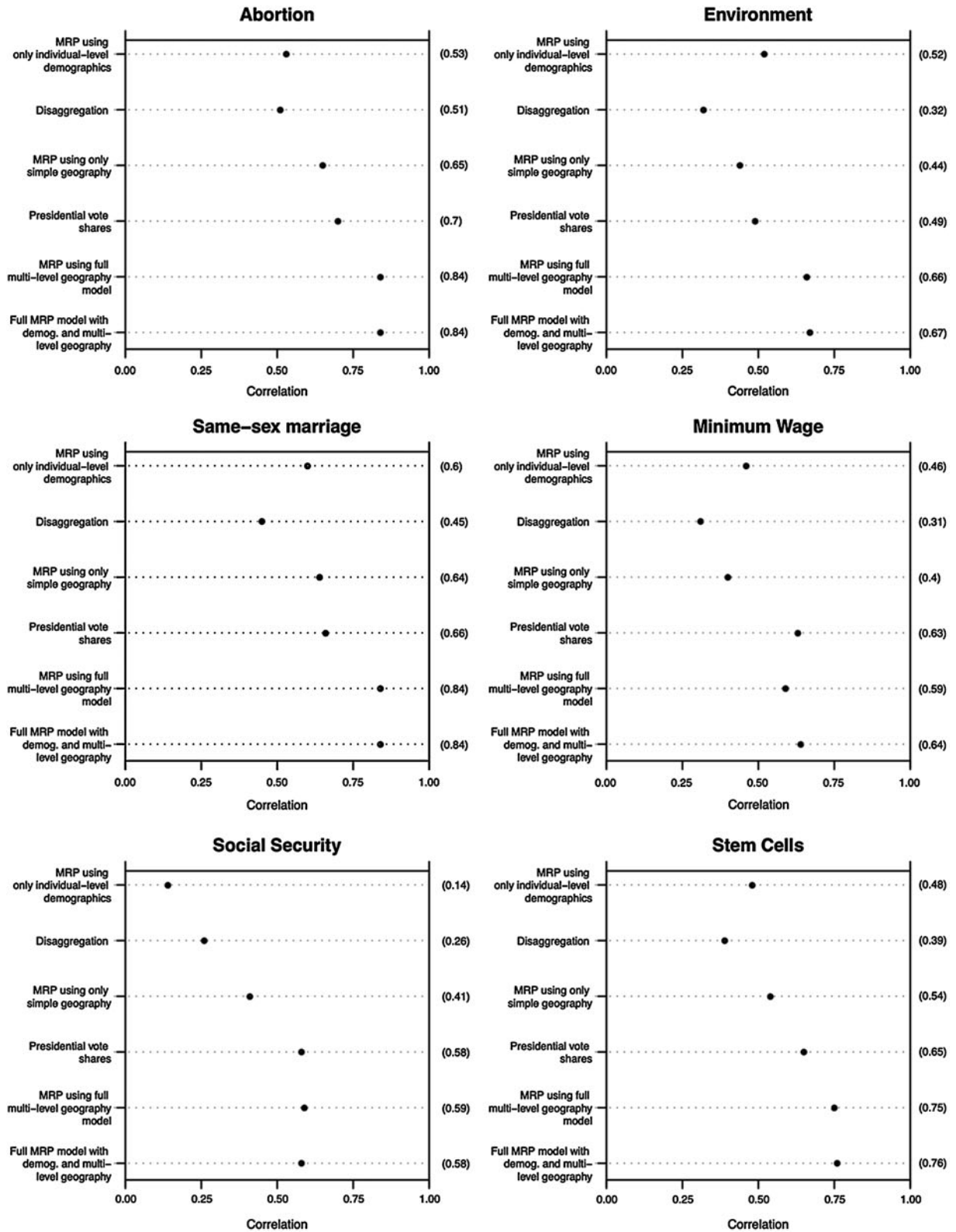
Note: We apply MRP to 5,000 person national samples to estimate congressional district public opinion, using four models of varying complexity. We show the correlation of each set of MRP estimates to the baseline estimate, along with the correlation using disaggregation and presidential vote shares. Values plotted are indicated along the right axis.

referenda (see Figures 2 and 4).[28] Thus, MRP estimates appear to be useful for predicting the public opinion of voters. However, researchers should be mindful of the distinction between voters and all adults and employ caution when deploying MRP, especially if attempting to analyze low-turnout elections where the opinions of voters may differ from those of all adults.

Second, although most existing applications of MRP and related techniques rely on the inclusion of election results in the model (e.g., Park, Gelman, and Bafumi 2004), it is possible to get very reliable estimates of district public opinion without relying on election results. This makes MRP estimates a viable tool for congressional scholars seeking to examine the impact of various district-level issue preferences on contemporaneous elections since the MRP estimates are at least plausibly not endogenous to election outcomes. More broadly, this strategy solves a classic problem in the electoral geography literature: theories often focus on the distribution of issue preferences across districts (Callander 2005; Gudgin and Taylor 1979), but empirical researchers are often forced to examine the distribution of election results instead.

Third, very small national samples (2,500 people) produce reliable estimates for congressional districts and moderate-sized samples (5,000 people) can produce reliable estimates for state legislative districts on many issues. For the first time, this means that congressional and state politics scholars can examine whether legislators are responsive to public opinion on individual issues. Moreover, the distribution of political preferences across districts is an important topic in other countries with winner-take-all districts, such as Australia, Canada, and the United Kingdom. Given the sample sizes of the most commonly used surveys in these countries and the ready availability of district-level census reports, MRP is a promising technique for the production of sensible district-level preference estimates.

Finally, the strength of our model stems partially from strong and predictable relationships between individual- and district-level demographic predictors and public opinion. Many of the covariates in our model are particularly well suited for social issues such as same-sex marriage and abortion. This suggests the importance of optimizing an MRP model for a particular research question. For instance, a researcher seeking district-level public opinion estimates on social security may want to include addi-

tional district-level covariates related to opinion on business and financial issues.

Acknowledgments

References

Ansolabehere, Stephen, Jonathan Rodden, and James M. Snyder, Jr. 2008. "The Strength of Issues: Using Multiple Measures to Gauge Preference Stability, Ideological Constraint, and Issue Voting." *American Political Science Review* 102 (2): 215–32.

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

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

Callander, Steven. 2005. "Electoral Competition in Heterogeneous Districts." *Journal of Political Economy* 113 (5): 1116–45.

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." *Journal of Politics* 68 (2): 397–409.

Erikson, Robert S. 1978. "Constituency Opinion and Congressional Behavior: A Reexamination of the Miller-Stokes Representation Data." *American Journal of Political Science* 22 (3): 511–35.

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

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

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, NJ: Princeton University Press.

Gelman, Andrew, and Thomas C. Little. 1997. "Poststratification into many categories using hierarchical logistic regression." *Survey Methodology* 23 (2): 127–35.

Gudgin, G., and Peter J. Taylor. 1979. Seats, Votes, and the Spatial Organization of Elections. London: Pion.

Jackman, Simon. 2009. Bayesian Analysis for the Social Sciences. West Sussex, UK: John Wiley & Sons Ltd.

Jackson, Robert A., and Thomas Carsey. 2002. "Group Effects on Party Identification and Party Coalitions across the United States." *American Politics Research* 30 (1): 66–92.

Jones, Dale E., Sherri Doty, Clifford Grammich, James E. Horsch, Richard Houseal, Mac Lynn, John P. Marcum, Kenneth M. Sanchagrin and Richard H. Taylor. 2002. Religious Congregations and Membership in the United States 2000: An Enumeration by Region, State and County Based on Data Reported for

[28]See online appendix for more about using MRP to estimate the opinion of voters.

149 Religious Bodies. Nashville, TN: Glenmary Research Center.

Lax, Jeffrey, and Justin Phillips. 2009. "How Should We Estimate Public Opinion in the States?" *American Journal of Political Science* 53 (1): 197–21.

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

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

Park, David, Andrew Gelman, and Joseph Bafumi. 2004. "Bayesian Multilevel Estimation with Poststratification: State-Level Estimates from National Polls." *Political Analysis* 12 (4): 375–85.

Pool, Ithiel De Sola, and Robert Abelson. 1961. "The Simulmatics Project." *Public Opinion Quarterly* 25 (2): 167–83.

Christopher Warshaw is a Ph.D. Candidate in Political Science at Stanford University, Stanford, CA 94305.

Jonathan Rodden is an Associate Professor of Political Science at Stanford University, Stanford, CA 94305.