Using Multilevel Regression and Poststratification to Estimate Dynamic Public Opinion

Andrew Gelman *
gelman@stat.columbia.edu

Jeffrey Lax
rjl2124@columbia.edu

Justin Phillips
jhp2121@columbia.edu

Jonah Gabry
jgabry@gmail.com

Robert Trangucci
robert.trangucci@gmail.com

September 11, 2016

Abstract

Multilevel Regression and Poststratification (MRP) has emerged as a widely-used technique for estimating subnational preferences from national polls. This technique, however, has a key limitation—existing MRP technology is best utilized for creating static as opposed to dynamic measures of opinion. In this paper, we develop an approach for implementing a "dynamic MRP", doing so in the context of changing public support for same-sex marriage. Using a large dataset of survey respondents, we estimate (in a single model) an annual measure of support for same-sex marriage for each state from 1993 through 2004. To evaluate our estimates we examine their face validity and compare them to estimates produced using the standard MRP approach as well as to the estimates produced by actual state-level polls. We also consider the conditions under which dynamic MRP seems to produce more accurate estimates.

*This project is supported by a grant from the National Science Foundation

1 Introduction

Since its emergence in the 1930s, scientific polling has grown into a large industry, fueled by an insatiable demand for information about the American public. Political scientists now have extensive polling on important policy issues as well as nearly-continuous polling on presidential approval and, during campaign season, we are inundated with trial-heat polls. The Roper Center for Public Opinion Research reports that its archive of opinion surveys has grown to over 18,000 datasets and continues to grow by hundreds more each year. Put simply, researchers are now awash in survey data.

While surveys provide researchers with invaluable data about the public's views, preferences, and beliefs, these data are not without limitations. Key among them, is that most polls are conducted by national survey organizations and are only designed to measure opinion at the national level (e.g., 34% of Americans have a favorable opinion of the Affordable Care Act, 53% support same-sex marriage, and so forth). Many important policies, however, are decided by state governments. For that matter, opinions on contentious national issues are themselves typically translated into national policy based on their potential impact on individual congressional or senatorial races or on the electoral college—i.e., "all politics is local." For these reasons summaries of national opinion provide only limited guidance to most lawmakers. They are also of limited value for those social scientists who want to study the ways in which public opinion varies across the geography of American federalism or those who wish to investigate issues of representation and policy responsiveness.

Relying on subnational surveys is not often a reasonable solution. Despite a rich tradition of state-level polling, finding comparable polls across states is nearly impossible. Similar questions are rarely asked in surveys across all or even most states and when they are, differences in timing, question wording, survey techniques, and response categories make comparisons difficult. An alternative and more practical approach is to use national survey data to simulate subnational opinion. Recently, scholars have revived—or more accurately, reinvented—simulation techniques. The first to truly catch on as a widespread tool, is multilevel regression and poststratification (MRP).

MRP was developed by Gelman and Little (1997) and extended by Park, Gelman, and Bafumi (2004, 2006), Lax and Phillips (2009a), Warshaw and Rodden (2012), and Kastellec, et al. (2014). It uses individual survey responses from national surveys coupled with advances in Bayesian statistics and multilevel modeling to generate opinion estimates by demographic-geographic subgroups, or “types”. The opinion estimates for each demographic-geographic respondent type are then weighted (poststratified) by the percentages of each type in the actual population of each subnational unit of interest. Several research teams have already evaluated and validated MRP (Park, Gelman and Bafumi 2006, Lax and Phillips 2009a, 2013, Warshaw and Rodden 2012, Buttice and Highton 2013). This work suggests that MRP can produce accurate estimates using fairly simple demographic-geographic models of survey response and small amounts of survey data.

However, challenges remain. In particular, the method is not currently well suited for exploring temporal changes in opinion. Existing MRP technology is best utilized for creating static measures of preferences—that is, using national surveys conducted during time t (with t representing a year or set of years) to create a single opinion estimate for each geographic unit. Though such static measures have already proven invaluable, they do not go far enough. Policymakers, the media, and scholars want to understand how and why public opinion is changing. Researchers, across a range of disciplines, need dynamic measures of the public’s preferences in order to better establish causal links between public opinion and outcomes. Political scientists, for example, may want to see whether public policy changes in response to shifting public preferences, while psychologists may want to investigate whether the mental health of lesbian, gay, and bisexual populations improve in places where public where tolerance for homosexuality is rising.

While a few scholars have employed MRP to study opinion change, they have typically created over-time measures of opinion by running separate MRPs on polls from different years or on year subsets and then stringing them together into a time series (see Gelman, Lax and Phillips 2010; Pachecho 2011). Doing so, however, fails to make use of all the available data and employs arbitrary assumptions as to how much change occurs over time. In this paper, we begin to advance MRP such that it can be used to create dynamic measures of public opinion. Advancing MRP in this fashion is not straightforward. To create dynamic measures, MRP needs to allow demographic and geographic effects to vary by year; that is, it must allow for models of much higher complexity wherein many key predictors are interacted with time. In essence, we aim to transform MRP so that it partially pools survey data not just across *space* both also over *time*. Advancing MRP in this fashion will enable researchers to utilize decades of accumulated survey data to create time-varying measures of public opinion across a variety of subnational units and to potentially generate more accurate estimates of public opinion at any given point in time—by incorporating time trends, researchers will be adding information that should result in more accurate estimates than are currently generated using static MRP.

We explore the potential of "dynamic MRP" in the context of changing public support for same-sex marriage. We have assembled a large dataset consisting of all of the respondents to publicly available opinion surveys conducted from 1993 through 2014 that directly ask about support for same-sex marriage. Our dataset consists of 81,127 respondents from 68 separate polls. Using these data we generate, in a single model, an estimate of support for same-sex marriage for each state in each year. This means that we produce 1,100 separate estimates of state opinion. In our analysis we consider alternative specifications of our dynamic model. To evaluate our estimates we examine their face validity and compare them to estimates produced using the standard MRP approach as well as to the estimates produced by actual state-level polls. We also consider the conditions under which dynamic MRP seems to produce more accurate estimates. We vary model complexity, consider the effects of state size, and consider the ways in which the distribution of surveys respondents across time—respondents are quite unevenly distributed across time—affects the accuracy of our estimates.

This exercise, though rather preliminary, will provide some evidence as to whether efforts to develop a fully dynamic MRP are worth continued pursuit. It also sets us on a path towards creating a set of guidelines or "best practices" for properly implementing dynamic MRP. Assuming that this is worthwhile endeavor, we plan to ultimately fully develop and evaluate the tools needed to generate dynamic measures of public opinion and add them to the existing MRP package in R.

Doing so will make the improved MRP toolkit available to a very broad range of social science researchers and poll analysts, not simply those methodologists with the particular skills necessary.

2 Overview of Multilevel Repression and Poststratification

The simulation of subnational public opinion traces back more than forty years (Pool, Abelson, and Popkin 1965). Opinion estimates are created for various geographic units according to the demographic distribution of the population within each. The primary flaw in the older versions of this technique is that respondents were generally modeled as differing in their demographic but not their geographic characteristics, so the prediction for any demographic type was unvaried across the subnational unit of interest (Erikson, Wright, and McIver 1993). In other words, using this method, the opinion of white citizens in New England in 1965 regarding civil rights issues could be thought the same as those held by white citizens in the South. Simulated opinion estimates ignoring geographic variation have been shown inferior in less controversial contexts as well. This tool failed to be widely adopted.

Recently, Park, Gelman, and Bafumi (2006) resuscitated the simulation approach, drawing on pioneering work in poststratification by Rod Little (1991,1993) and small-area estimation by Bob Fay (1979). Unlike early simulation approaches, this reincarnation, which is referred to as multilevel regression and poststratification (MRP), takes geography into account. It is also much more sophisticated than earlier simulation techniques in terms of the way it models individual survey responses and demographic variation.

MRP proceeds in two stages. In the first, a multilevel model of individual survey response is estimated. Instead of relying solely on demographic differences like older incarnations of the method, the geographic location of the respondents is used to estimate geographic effects, which themselves can be modeled using additional predictors such as aggregate demographics of the geographic area. Those residents from a particular area yield information as to how much predictions within that area vary from others after controlling for demographics. All individuals in the survey, no matter their location, yield information about demographic patterns which can be applied to all geographic estimates. These demographic-geographic predictors can interact, as well.

To be specific, MRP uses Bayesian statistics and multilevel modeling (Gelman and Little 1997; Park, Gelman, and Bafumi 2006) to improve upon the estimation of the effects of individual and geographic predictors. For data with hierarchical structure (e.g., individuals within states or congressional districts), multilevel modeling is generally an improvement over classical regression. Rather than using "fixed" ("unmodeled") effects, MRP uses "random" ("modeled") effects, for some predictors. These modeled effects (e.g., state or district effects) are related to each other by their grouping structure and thus are partially pooled towards the group mean, with greater pooling when group-level variance is small and for less-populated groups (this is equivalent to assuming errors are correlated within a grouping structure (Gelman and Hill 2007, 244-65)). The degree of pooling within the grouping emerges from the data endogenously. They can be modeled not only in terms of this "shrinkage" (the assumption that they are drawn from some common distribution) but also by including group-level predictors.

The second step is poststratification: the estimates for each demographic-geographic respondent type are weighted (poststratified) by the percentages of each type in actual populations of the relevant geography, so that we can estimate the percentage of respondents within each who have a particular issue position or preference. Poststratification is done using state or congressional district population frequencies obtained from either the Public Use Micro Data Samples supplied by the Census Bureau (and available going back to early 20th century) or similar data. Compared to previous simulation methods and classical methods, multilevel modeling now makes possible the use of many more respondent types. This too greatly improves accuracy.

Importantly, poststratification corrects for differences between survey samples and the actual population. National surveys, while representative at the national level, are often flawed in terms of representativeness or geographic coverage at the state or congressional district level, due to clustering and other survey techniques utilized by polling firms (Norrander 2007, 154). Indeed, with the increasing popularity of internet survey techniques and cell phones, it is becoming increasingly difficult to even find data with a random sample of the national population, let alone random samples of subsets of interest such as demographic slices or residents of particular states. MRP addresses these concerns. Moreover, it can be difficult to combine different surveys using traditional methods, if they use different question wording, sampling techniques, demographic questions, etc. MRP, on the other hand, can bridge surveys by modeling and accounting for such differences.

The value of using MRP to estimate subnational public opinion has been confirmed in several articles that have appeared in top peer-reviewed journals. Importantly, these articles have been authored by separate research teams. Park, Gelman, and Bafumi (2006) compare MRP estimates of opinion to those of alternative techniques that also model individual survey responses. They show that MRP substantially outperforms the older style simulation approaches as well as simulation approaches that do not partially pool information across respondents. Lax and Phillips (2009a) show that MRP notably outperforms its main competitor (disaggregation), yielding smaller errors, higher correlations, and more reliable estimates. They also establish the face and external validity of MRP estimates by comparing them to actual state polls and election results, demonstrating that a single national poll (approximately 1,400) and a very simple demographic-geographic model can, in some contexts, suffice for MRP to produce highly accurate state-level opinion estimates. In a parallel analysis, Warshaw and Rodden (2012) demonstrate that MRP can produce accurate estimates of opinion by congressional districts (using sample sizes of just 2,500 survey respondents) and state senate districts (with a sample of 5,000). Most recently, scholars have worked to provide sets of guidelines and cautions to MRP users (Buttice and Highton 2013; Lax and Phillips 2013).

Collectively, these efforts demonstrate that MRP can produce accurate estimates of public opinion across a variety of subnational units using fairly simple models of survey preferences and modest sample sizes. Relatively small national samples can be used to produce accurate measures of subnational opinion because the multilevel models used in MRP borrow strength by partially pooling respondent types across space (i.e., the subnational units of interest) to an extent determined (endogenously) by the data. As a result, MRP is “emerging as a widely used gold standard for estimating preferences from national surveys” (Selb and Munzert 2011, 456). Indeed, research employing MRP has already appeared in the top political science journals, and MRP has been employed to study to myriad substantive questions, including the responsiveness of state governments

(Lax and Phillips 2009b, 2012), state supreme court abortion decisions (Caldarone, Canes-Wrone, and Clark 2009), roll call voting in Congress (Kastellec, Lax, and Phillips 2010, Kastellec et al. 2015; Krimmel, Lax, ad Phillips 2016), and the diffusion of public policy (Pacheco 2012). This substantive work has also further developed the method, devising techniques for postratifiying by non-census variables and for estimating uncertainty around MRP estimates and then incorporating this uncertainty into subsequent empirical analyses (Kastellec et al. 2015).

3 Developing a Dynamic MRP

While the methodological and substantive potential of MRP has clearly been established, more work needs to be done before the full potential of national surveys can be unlocked. In particular, existing MRP technology seems to be best utilized for creating static measures of preferences—that is, using national surveys conducted during time t (with t representing a year or set of years) to create a single opinion estimate for each geographic unit of interest. We do know that there are better and worse ways to do MRP (Lax and Phillips 2013). For example, for the most typical use of creating state estimates from national data, it is important to include a state-level variable such as presidential vote or ideology to help smooth across states and enable a proper degree of pooling across states. But how do we explore time trends and combine data across years instead of geographic divisions?

One simple way is just to do MRP for each year, standing alone, to create a state estimate for each year possible. This doesn't make use of the "secret sauce" of MRP for the state problem—whereby partial pooling enables large efficiency gains.So we want to allow for some type of pooling of information over time... but combining data across years raises many concerns. How do we help MRP capture trends across time? How do we enable the "right" degree of pooling across time, without over smoothing and hiding opinion change or opinion differences across states? How do we avoid chasing noise in the guise of opinion swings? Do we just need an over-time state smoother such as presidential vote? Should we model time trends more explicitly? How do we know if we are discovered trends or creating them through modeling?

Some researchers have developed various "patches" for these problems, but these do not necessarily deal with said problems and, perhaps worse, we cannot tell if they do so. For example, consider the creation of over-time measures of opinion by running separate MRPs on polls from different years or on year subsets and then stringing them together into a time series (see Gelman, Lax and Phillips 2010; Pachecho 2011). Doing so, however, fails to make use of all the available data and employs arbitrary assumptions as to how much change occurs over time. Pooling is complete within subset and barred completely across subsets. At least subsetting does some, if oddly constructed, degree of pooling. The most careful of this type of work so far is Pachecho (2011), which created yearly estimates of state-level voter ideology (from 1977 through 2007) by pooling surveys over three or five year time-periods. So, to get an estimate of opinion in year *t* she estimates an MRP using responses from polls conducted in years *t -1*, *t*, and *t + 1*. Doing so assumes away potential short-term changes in opinion. It also means (according to our back-of-the-envelope calculations) that for each year's opinion estimate, 90% of the survey data—just over 292,000 responses—are ignored so that we learn nothing about pooling parameters, etc., from them. This suggests there are gains to be made. We believe that MRP can be improved so that researchers do not need to make these unpalatable decisions.

In particular, we suspect that by incorporating much more data (than static MRP) we should be able to improve opinion estimation for any given year. But we need to know how to model time and opinion change in a flexible-enough way that still yields the desired efficiency gains, without obscuring time trends or creating spurious ones. We need to compare different methods of so doing. For example, Pacheco (2011) presented evidence that three-year windows beat five-year windows for partisanship and ideology data. We seek general evidence across a wider array of possibilities for doing MRP over Time...or MRT.

The models that we employ here are estimated using an R package called RStanarm. RStanarm is a full Bayesian counterpart to lme4 (a more approximate but simpler method, the usual way to implement multilevel model estimation for the purposes of MRP).[1] One advantage of this

[1] It encapsulates several popular model classes, such as linear mixed effect models (estimated using stan-lmer) and generalized linear mixed models (estimated using stan-glmer). These functions accept lme4 model formula syntax but ultimately use Stan to do Bayesian inference on the models. Specifically, Stan uses a flavor of Hamiltonian Monte Carlo, a Markov Chain Monte Carlo algorithm, called the No-U-Turn-Sampler, which adaptively tunes the

Stan approach is that uncertainty around estimates is directly output as part of the estimation process rather than requiring a type of post-estimation bootstrapping from model parameters (as done in the previous MRP paper incorporating uncertainty, Kastellec..). More importantly perhaps, it is a fully Bayesian approach incorporating uncertainty (including uncertainty in the hierarchical variance parameters of the models). While this makes for negligible differences for simple multilevel models and normal MRP, the increased complexity of MRT warrants allowing for as many sources of uncertainty as possible. At this point in time, Stan estimation is feasible for these purposes./footnoteStan is flexible, so it allows you to add terms to your model and estimate the thing in the same framework (i.e. run Stan, extract samples, take means, variances and quantiles, etc), rather than requiring you to change packages and figure out how to shoehorn your model into another estimation procedure.

3.1 Models Estimated

The following is a guide for all the variables used in our models:

---

hyperparameters required by HMC. HMC is powerful because it takes the geometry of the posterior distribution into account when generating proposals for the next step in a Markov Chain. In doing so, the sampler can take sequential steps that are far apart, yielding draws from the posterior that are near independent. This differs from the maximum marginal likelihood (MML) algorithm that undergirds lme4 in several ways. When viewed through the lens of Bayesian inference, lme4 generates maximum a posteriori (MAP) estimates for random effects and hierarchical variance parameters with uniform priors on the half interval (0, Inf). Gelman 2006 writes that this leads the model to overstate the expectation of true variance, which can lead to less-than-optimal shrinkage of random intercepts towards the common mean. Additionally, MML cannot be applied to generalized mixed models in closed form and thus is estimating a normal approximation to the posterior (a Laplace approximation). To summarize, MML's point estimates for hierarchical variance parameters are positively biased, but MML understates posterior uncertainty because it uses a normal approximation to the posterior. Stan doesn't suffer from any of these shortcomings. We can put proper priors on hierarchical variance parameters and then generate random draws from the true posterior. We can diagnose when the sampler encounters problems, but MML cannot offer any theoretical tools to assess the statistical validity of its estimates aside from the convergence of its log-likelihood function.

Table 1: Variables Used in Dynamic MRP Models

Variable	Detail
`favor`	Number of respondents favoring gay marriage in cell i
`oppose`	Number of respondents opposing gay marriage in cell i
`CSL`	Standardized (mean 0 and sd of 0.5) social dimension of the 2-dimensional latent process quantifying the liberalism of states' citizenry from Caughey and Warshaw (2015) Dynamic Responsiveness in the American States 1936-2014. Varies year-by-year and state-by-state. CSL = Citizen Social Liberalism
`year_std`	Year running from 1993 to 2015, standardized
`year_sq_std`	Square of year, running from 1993 to 2015, standardized
`state`	Categorical state variable, 50 levels, unordered
`age`	Categorical age variable, 4 levels
`edu`	Categorical education variable, 4 levels
`sex_race`	Categorical sex and race variable, 8 levels
`year`	Categorical year variable, 19 levels

Baseline MRT model (e.g., Model 1):

$$n_i^{\text{favor}} \sim \text{BinomialLogit} N_i, \alpha_i$$

$$\begin{aligned}\alpha_i = \mu &+ \gamma_{1,\texttt{state}[i]} + \gamma_{2,\texttt{state}}\,\texttt{year\_std} \\ &+ \gamma_{3,\texttt{state}[i]}\,\texttt{year\_sq\_std} + \delta_{\texttt{CSL}}\,\texttt{CSL} \\ &+ \beta_{\text{age}[i]} + \beta_{\text{edu}[i]} + \beta_{\text{sex\_race}[i]} \\ &+ \beta_{\text{year}[i]} + \beta_{\text{year}[i],\text{state}[i]}\end{aligned}$$

$$\begin{pmatrix}\gamma_{1,\texttt{state}} \\ \gamma_{2,\texttt{state}} \\ \gamma_{3,\texttt{state}}\end{pmatrix} \sim \mathcal{N}\begin{pmatrix}\mu_1 \\ \mu_2 \\ \mu_3\end{pmatrix}, \begin{pmatrix}\sigma_1^2 & \rho_{12}\sigma_2\sigma_1 & \rho_{13}\sigma_3\sigma_1 \\ \rho_{21}\sigma_2\sigma_1 & \sigma_2^2 & \rho_{23}\sigma_2\sigma_3 \\ \rho_{31}\sigma_3\sigma_1 & \rho_{32}\sigma_3\sigma_2 & \sigma_3^2\end{pmatrix}$$

The prior specifications follow:

$$\beta_{\text{age}} \sim \mathcal{N}0, \sigma_{\text{age}}$$

$$\beta_{\text{edu}} \sim \mathcal{N}0, \sigma_{\text{edu}}$$

$$\beta_{\text{sex\_race}} \sim \mathcal{N}0, \sigma_{\text{sex\_race}}$$

$$\beta_{\text{year}} \sim \mathcal{N}0, \sigma_{\text{year}}$$

$$\beta_{\text{state,year}} \sim \mathcal{N}0, \sigma_{\text{state,year}}$$

$$\sigma_{\gamma_{i,\texttt{state}}} \sim \mathcal{N}^{+}0, 1, \; i \in \{1, 2, 3\}$$

$$\sigma_{\text{age}}, \, \sigma_{\text{edu}} \sim \mathcal{N}^{+}0, 1$$

$$\sigma_{\text{sex\_race}}, \, \sigma_{\text{year}}, \, \sigma_{\text{state,year}} \sim \mathcal{N}^{+}0, 1$$

`lmer` syntax:

```
cbind(favor, oppose) ~
        CSL +
        (1 | age) +
        (1 | edu) +
        (1 | sex_race) +
        (1 | year) +
        (1 | year:state) +
        (1 + year_std + year_sq_std | state)
```

We will discuss two additional models here (omitting some others for now). Model 2 is identical to Model 1 but does not include year or year-squared trends, though it allows the effect of ideology to vary by state. Skipping Model 3 for now, Model 4 is again similar to Model 1, but allows the "effect" of the state ideology variable to vary over time (say, if ideology maps to opinion well later on but does not do so earlier). Model 5 does not include year and year squared no does it include the measure of state-level ideology. We are also estimating a variety of models that include or do not include time trends and varying sloped in various combinations. Results from these will be added later.

3.2 Our Data

We estimate our dynamic models of MRP using a large database of survey respondents. The database includes respondents from all national publicly available polls from 1993 through 2014 that directly asked about support for same-sex marriage.[2] Respondents were excluded if we do not have data about their demographic characteristics or geographic location (not all polls report respondents state of residence). We located polls using iPoll which is housed at the Roper Center for Public Opinion Research. For each respondent, we code race, age, gender, level of education, and state of residence. In total our dataset includes a total of 81,127 useable respondents from 68 unique polls. Our initial efforts maintain all data, to be trimmed with simulated smaller samples later.

3.3 Partial Preliminary Answers

We are now dealing with the situation where we can feed a large number of national polls into MRP – can MRT make any further improvements in such a situation? Does it do harm?

We begin by considering the simple face validity of our MRT estimates. The top left panel of Figure 1 plots estimated support for same-sex marriage over time for a sample of 12 states. The estimates are generated using Model 4, and the dashed-line shows national opinion. We use Model 4 here because it does slightly better by certain metrics than our other MRT models (more on this later). The 12 states (though not a random sample) contains a mix of large, medium, and small population states as well as states from across the ideological spectrum.

The results (unsurprisingly) are consistent with what one would anticipate. Support for same-sex marriage rises over time, and does so quite dramatically after 2005. The states that are most supporting of same-sex marriage in our sample are those in the northeast (Massachusetts, New Hamsphire, and New York), plus California. The states with the lowest levels of support are located in the south—Tennessee and Alabama. As one can see, the estimates across states clearly

[2]Polls that ask about support for both same-sex marriage and civil unions are excluded as well as those that ask about support for a constitutional amendment banning same-sex marriage.

move together, though there is some independence.

Figure 1 also reports opinion estimates for this sample of states using tradiitonal MRP (i.e., running a separate MRP for each year, only using data from polls conducted during that year) and what we refer to here as MRP-3 (i.e., using MRP on a three year sliding window of survey data: t-1, t, t+1.) The MRP model that we use is as follows:

Baseline MRP model:

$$n_i^{\text{favor}} \sim \text{BinomialLogit} N_i, \alpha_i$$

$$\alpha_i = \mu + \beta_{\text{age}[i]} + \beta_{\text{edu}[i]} + \beta_{\text{sex\_race}[i]} + \beta_{\text{state}[i] + \delta_{\text{CSL}} \text{CSL}}$$

The priors over random intercept terms follow:

$$\beta_{\text{age}} \sim \mathcal{N} 0, \sigma_{\text{age}}$$

$$\beta_{\text{edu}} \sim \mathcal{N} 0, \sigma_{\text{edu}}$$

$$\beta_{\text{sex\_race}} \sim \mathcal{N} 0, \sigma_{\text{sex\_race}}$$

$$\beta_{\text{state}} \sim \mathcal{N} 0, \sigma_{\text{state}}$$

$$\sigma_{\text{age}}, \sigma_{\text{sex\_race}} \sim \mathcal{N}^{+} 0, 1$$

$$\sigma_{\text{edu}}, \sigma_{\text{state}} \sim \mathcal{N}^{+} 0, 1$$

`lmer` syntax:

```
cbind(favor, oppose) ~
    CSL +
        (1 | age) +
        (1 | edu) +
        (1 | sex_race) +
```

```
(1 | state)
```

The model we use for the MRP-3 is the same as our baseline MRP model, with the exception of including more data, from the surrounding year on either side (thus making it a three-year pooling version of normal MRP). This is similar to the approach used by Pacheco (2011),with the inclusion of a state-level predictor (in some results - to be added - we show that leaving that out, as she does, makes for clearly worse estimates).

In Figure 1 we can see that MRT estimates are fairly similar to those generated by the other two approaches. The approach with the greatest amount of smoothing appears to be MRP-3, while MRT and traditional MRP appear to experience similar amounts of smoothing. Of course, we do not know how much smoothing or year-to-year fluctuation in state estimates we should observe since there is no measure of "true" opinion. In future iterations of our efforts we will deal with these issues in part by usuing simulated data. At this point, though, MRT seems to capture a reasonable degree of yearly change.

The next two figures add 90% confidence intervals around our MRT estimates of state-level support for same-sex marriage. Figure 2 plots MRT estimates against traditional MRP estimates, while Figure 3 plots MRT against estimates obtained using MRP-3. Again, estimates across all three methods are highly correlated. When looking at MRP and MRT, it is fairly clear that the confidence interals around MRT tend to be tighter. This suggests that by borrowing strength from the full time series, there are gains from MRT in terms of efficiency. This appears to be particualrly true in the early part of the time series, where survey data is more sparse. When compared to MRP-3, however, the MRT condifence intervals are larger.

These patterns are further illustrated in Figure 4, which reports the average width of 90% confidence intervals for all MRT models (not just model 4), traditional MRP, and MRP-3. The figure also shows results for all states and, separately, for the five largest and five smallest states. The average size of the confidence interval for MRP is approximately 10 percentage points (averaging about 8 points for the largest states and about 11 points for smallest ones). Nearly all of the MRT models produce noticably smaller-sized intervals. (The only excpetion to this patter is Model 5 which does not include year and year squared nor a measure of state level ideology). For example, the opinion estimates from MRT Model 4 have an average confidence interval of approximately 7 percentage points (averaging about 6 points for the largest states and about 8 points for the smallest states).

The figure also makes it clear that MRP-3 tends to produce the smallest confidence intervals. One concern we have about this, however, is that these confidence intervals may be suspiciously small. It is likely that some of this confidence (relative to MRT) comes from ignoring full time series. Not accounting for a time trend in your data can give one confidence intervals that are simply too small.

Figure 5 shows MRT (Model 4) against a number of state polls. For this analysis, we gathered the results of 75 actual state polls conducted 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, at particular snapshots in time. For comparison, Figure 6 shows the same for regular MRP. These state polls are as close as we come to independent measures of true opinion, though they are noisy and scattered, and based on a non-random sample of time-geographic points. Additionally, the polls are themselves estimates and not necessarily truth. Gathering more newly available state polls and later simulating data will make up for this in future revisions.

The MRT estimates of opinion are generally similar to the estimates prodcued by state polls. Often, though not always, the state poll estimates fall within the confidence interval of our MRT estimates. Figure 8 reports the average root-mean-square error between each MRT model, traditional MRP, and MRP-3 and the state polls. This figure gives us an idea as to which apporach is better at predicting the results of actual polls.

Across all states there are only small differences in the overall success of the different approaches. The “winners”, such as they are at this point, are traditional MRP and MRT (model 4). Among the largest states, MRT does best, but among the smallest states, traditional MRP does best. Once concern is that we sinply do not have many polls among small states. In any case, these are the sorts of comparisons we seek and will explore further with large simulated runs.

Finally, one can also ask how much the MRT and MRP-3 estimates differ from regular MRP. This is shown in Figure 9. On average the differences are quite small. MRT-4 is the most similar to traditional MRP, while MRT-5 is the most different.

4 Further Questions/Next Steps

Whereas our previous attempts at handling MRT (not shown) ran into problems of clear over-pooling over time, at this point we have demonstrated that MRT can successfully be implemented. We were able to generate (in a single model) annual opinion estimates for all 50 states over a 22-year period. These estimates are comparable to esitmates obtained through traditional MRP or MRP using a three-year sliding window of survey data (similar to Pacheco 2011). The estimates we produced using MRT are also strongly correlated to actual state-level polls. This is especially true when variables are included to capture time trends.

At this point, however, we cannot say that MRT is clearly advantageous over the alternatives approaches we consider. Though right now we have only consideration a situation in which there are dozens of polls and more than one poll in nearly all years. The next step is do randomly select polls to see how well MRT does when data becomes more scarce and MRP applied to time series starts to break down. We also plan to consider more complicated models. For example, can our predictions become accurate if we allow the democgraphic effects to vary over time etc. A clear next step is to also set up an AR1 framework, which might prove necessary for capturing more complicated time trends.

Our rough research design for that will be to create time series with different degress of heterogeniety in level and trend, assessing whether MRT can recover those degrees, created true by defintion, using simulated poll data samples. Similarly, we will induce different degrees of polarization to be recovered, as well as state-level opinion shocks.

References

Buttice, Matthew K and Benjamin Highton. 2013. "How Does Multilevel Regression and Post-stratification Perform with Conventional National Surveys?" *Political Analysis* 21: 449-67.

Caughey, Devin and Christopher Warshaw. 2015. "Dynamic Estimation of Latent Opinion Using a Hierarchical Group-Level IRT Model," *Political Analysis* 23(2): 191211.

Caldarone, Richard P., Brandice Canes-Wrone, and Tom S. Clark. 2009. "Partisan Labels and Democratic Accountability: An Analysis of State Supreme Court Abortion Decisions." *The Journal of Politics* 71(2):560-73.

Chipman, H. A., George, E. I., and McCulloch, R. E. 2008. "BART: Bayesian additive regression trees. Technical report, Department of Mathematics and Statistics," Acadia University, Nova Scotia, Canada.

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., and Herriot, Roger A. 1979. "Estimates of Income for Small Places: An Application of James-Stein Procedures to Census Data. *Journal of the American Statistical Association* 74: 269-77.

Feller, Avi, Andrew Gelman and Boris Shor. 2012. "Red State/Blue State Divisions in the 2012 Presidential Election." *The Forum* 10(4): 127-31.

Gelman, Andrew. 2004. "Parameterization and Bayesian Modeling." *Journal of the American Statistical Association* 99(466):537-45.

Gelman, Andrew. 2005. "Comment: Anova as a Tool for Structuring and Understanding Hierarchical Models." *Chance* 18(3):33.

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

Gelman, Andrew, Aleks Jakulin, Maria Grazia Pittau, and Yu-Sung Su. 2008. "A Weakly Informative Default Prior Distribution for Logistic and Other Regression Models. Annals Applied Statistic. 2(4): 1360-83.

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

Gelman, Andrew and Yair Ghitza. 2013. "Deep Interactions with MRP: Election Turnout and Voting Patterns Among Small Electoral Subgroups." *American Journal of Political Science* 57(3): 762-76.

Gelman, Andrew, Jeffrey Lax, and Justin Phillips. 2010. "Over Time, a Gay Marriage Groundswell. *The New York Times*, page WK3.

Gelman, Andrew, David Park, Boris Shor, and Jeronimo Cortina. 2009. *Red State, Blue State, Rich State, Poor State: Why Americans Vote the Way They Do*, second edition. Princeton University Press.

Gelman, Andrew, Boris Shor, Joseph Bafumi, and David Park. 2007. "Rich State, Poor State, Red State, Blue State: Whats the Matter with Connecticut?" *Quarterly Journal of Political Science* 2:34567.

Gelman, Andrew, Nate Silver, and Daniel Lee. 2009. "The Senates Health Care Calculations." *New York Times*, 19 November, A35.

Gelman, Andrew, David van Dyk, Zaiying Huang, and John W. Boscardin. 2007. "Transformed and Parameter-Expanded Gibbs Samplers for Multilevel Linear and Generalized Linear Models." *Journal of Computational and Graphical Statistics*.

Kastellec, Jonathan, Jeffrey Lax, and Michael Malecki, and Justin Phillips. 2015. "Distorting the Electoral Connection? Partisan Representation in Supreme Court Confirmation Politics." *The Journal of Politics*.

Kastellec, Jonathan, Jeffrey R. Lax, and Justin H. Phillips. 2010. "Public Opinion and Senate Confirmation of Supreme Court Nominees." *The Journal of Politics* 72(3): 767-84.

Krimmel, Kate, Jeffrey Lax, and Justin Phillips. 2016. "Gay Rights in Congress: Public Opinion and (Mis)Representation." *Public Opinion Quarterly*.

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

Lax, Jeffrey R., and Justin H. Phillips. 2009b. "Gay Rights in the States: Public Opinion and Policy Responsiveness," *American Political Science Review* 103(3): 367385.

Lax, Jeffrey and Justin H. Phillips. 2012. "The Democratic Deficit in the States," *American Journal of Political Science* 56(1), 148166.

Lax, Jeffrey R., and Justin H. Phillips. 2013. "How Should We Estimate Sub-National Opinion Using MRP? Preliminary Findings and Recommendations" Presented at the Annual Meeting of the American Political Science Association, Chicago, IL.

Little, Roderick J. A. 1991. "Inference with survey weights." *Journal of Official Statistics* 7: 405-424.

Little, Roderick J. A. 1993. "Post-stratification: a modelers perspective." *Journal of the American Statistical Association* 88: 1001-12.

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

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

Pacheco, Julianna. 2011. "Using National Surveys to Measure Dynamic U.S. State Public Opinion: A Guideline for Scholars and an Application." *State Politics and Policy Quarterly* 11:41539.

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

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

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

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

Salganik, Matthew, Tian Zheng, and Tyler H. McCormick. 2010. "How Many People Do You Know? Efficiently Estimating Network Size." *Journal of the American Statistical Association* 105(489): 59-70.

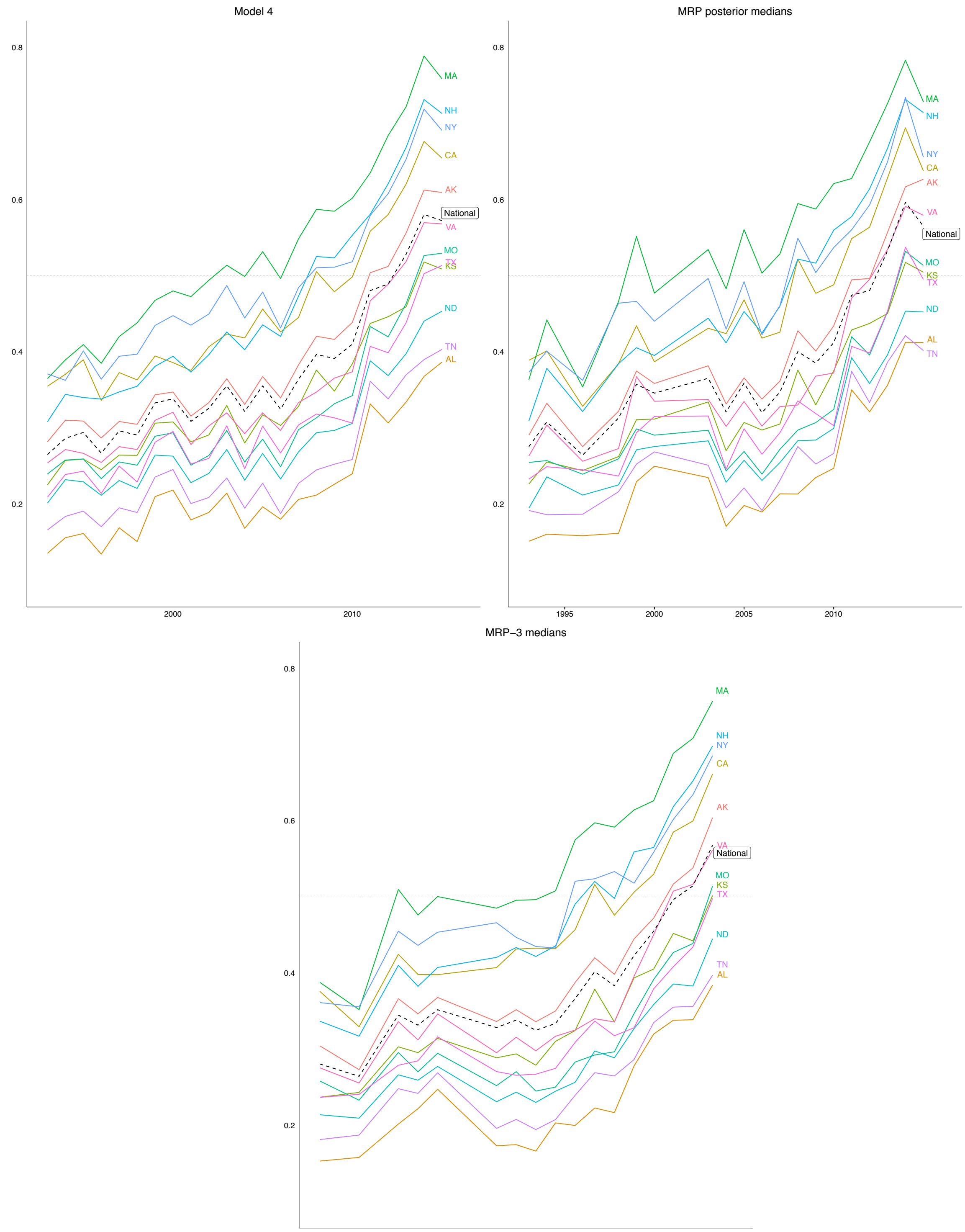
Selb, Peter and Simon Munzert. 2011. "Estimating Constituency Preferences from Sparse Survey Data Using Auxiliary Geographic Information." *Political Analysis* 19(3): 455-70.

Shirley, Kenneth E. and Andrew Gelman. 2014. "Hierarchical Models for Estimating State and Demographic Trends in US Death Penalty Public Opinion." *Journal of the Royal Statistical Society A*

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

Zheng, Tian, Matthew J. Salganik, and Andrew Gelman. 2006 "How Many People Do You Know in Prison?: Using Overdispersion in Count Data to Estimate Social Structure in Networks." *Journal of the American Statistical Association* 101(474): 409-23.

Figure 1: Comparing Estimates for MRT, MRP, & MRP-3

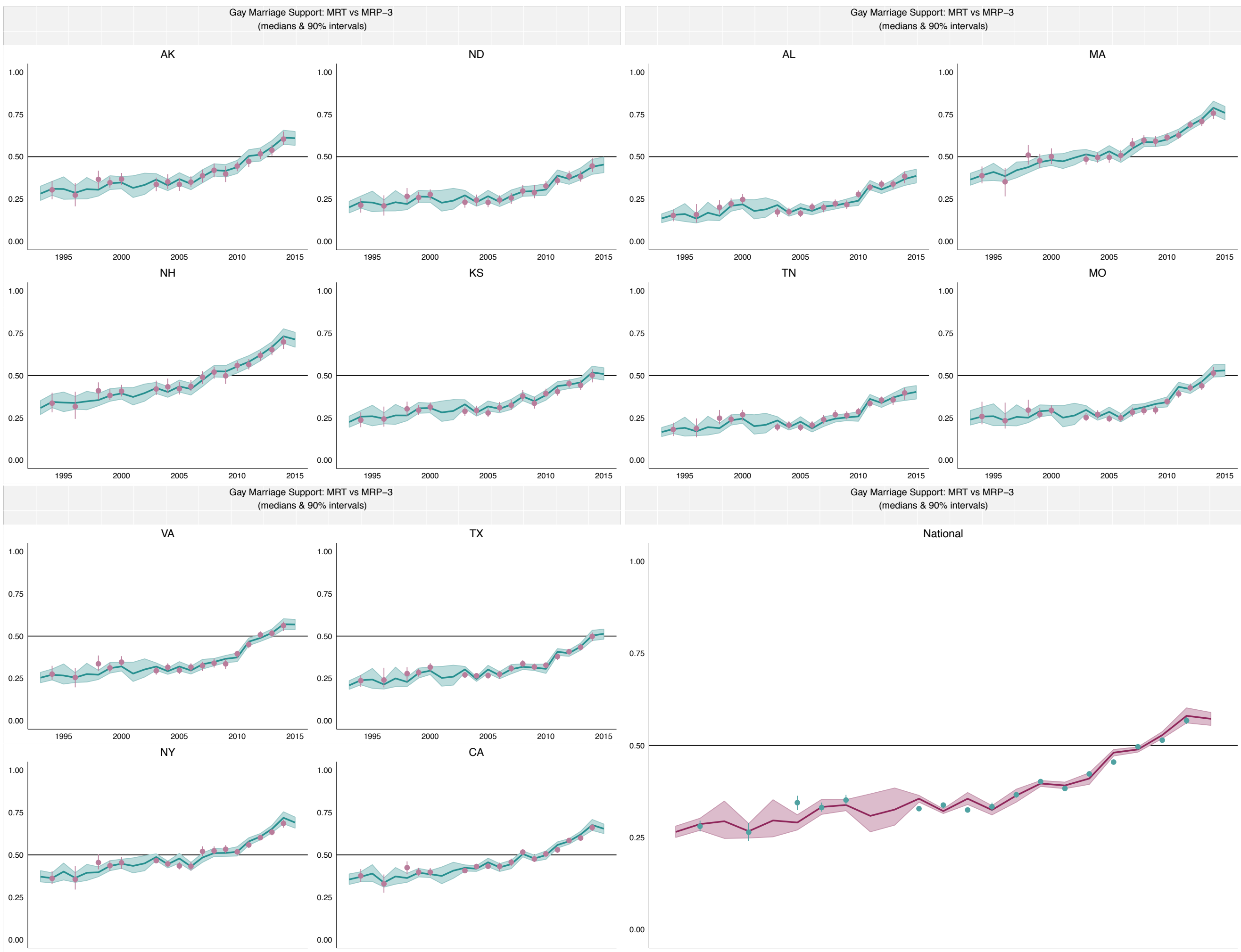


Note: The figure plots estimated support for same-sex marriage over time for a sample of 12 states. The estimates in the top left panel are generated using Model 4; the estimatimes in the top right are generated using traditional MRP; the estimates in the bottom panel are generated using MRP-3. The dashed-lines show national opinion.

Figure 2: Model 4 Estimates Compared to MRP Estimates

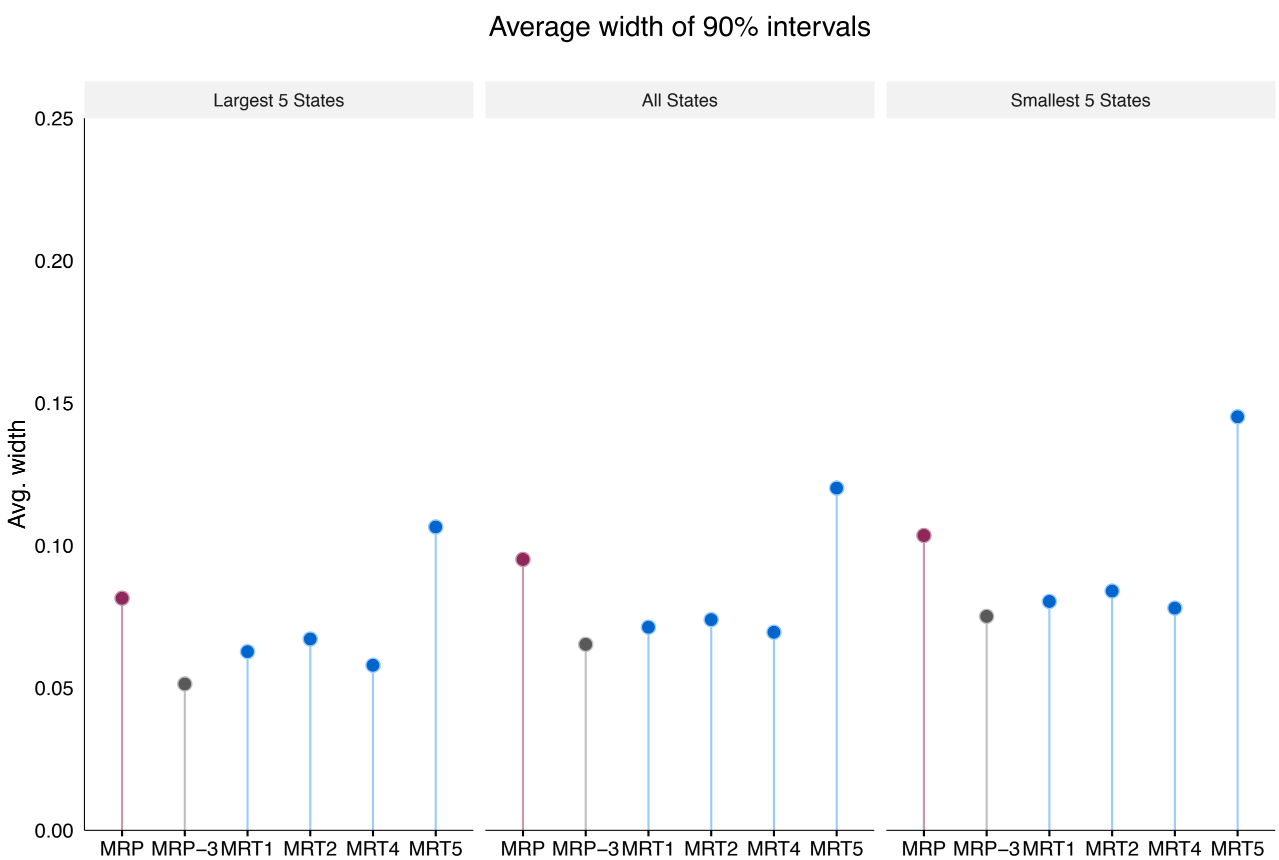
Note: The figure plots state opinion estimates (the solid line) along with their 90% confidence intervals (the shaded ribbon). The dots are state opinion estimates using a traditional MRP model, run for each year separately. The lines around the dots are 90% confidence intervals.

Figure 3: Model 4 Estimates Compared to MRP-3 Estimates



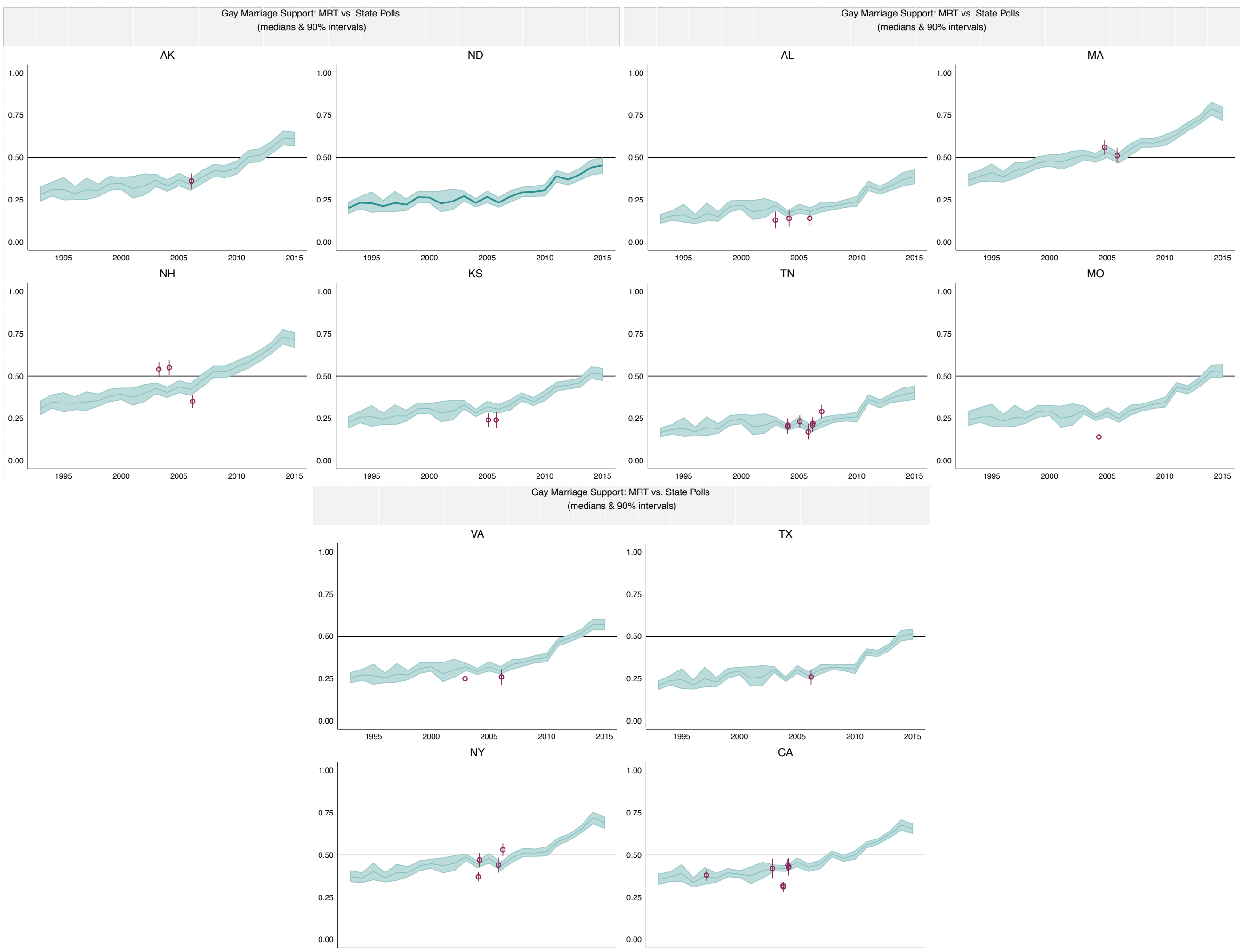
Note: The figure plots state opinion estimates (the solid line) along with their 90% confidence intervals (the shaded ribbon). The dots are state opinion estimates using an MRP-3 model. The lines around the dots are 90% confidence intervals.

Figure 4: Average Width of 90% Confidence Intervals, All Models



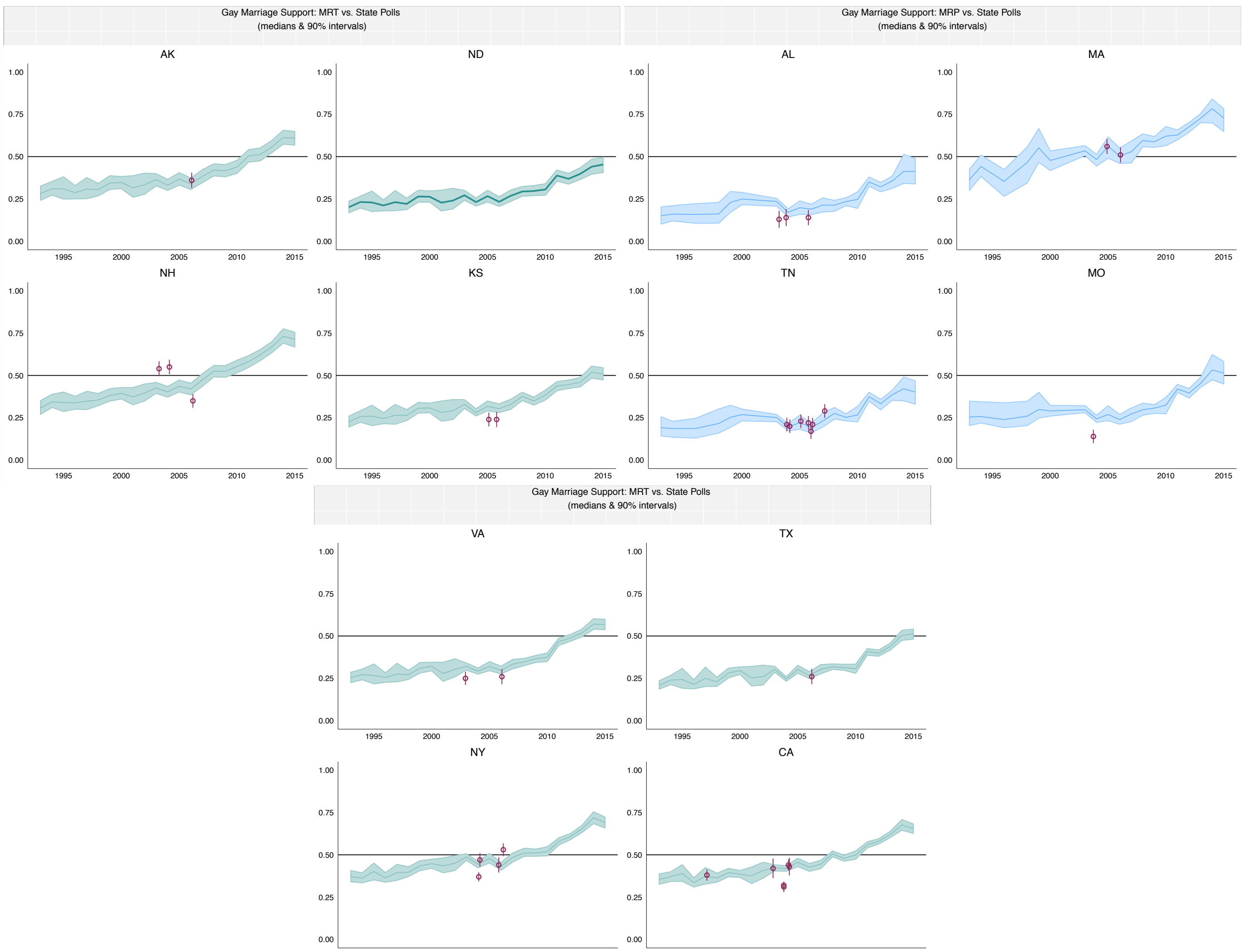
Note: The figure plots the average width of the 90% confidence interval for all MRT models as well the traditional MRP model and MRP-3.

Figure 5: Model 4 Estimates Compared to State Polls



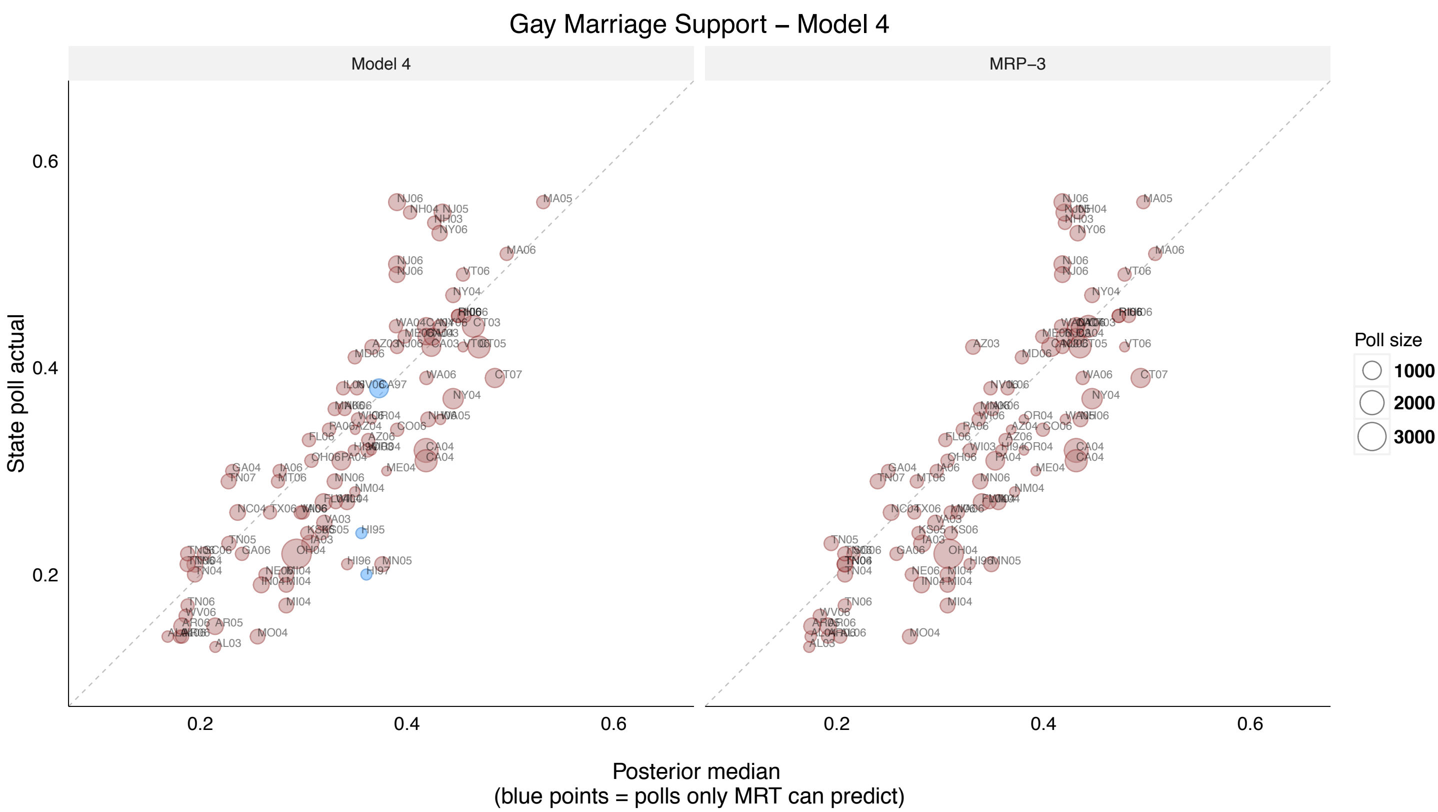
Note: The figure plots state opinion estimates (the solid line) along with their 90% confidence intervals (the shaded ribbon). The dots show the results of actual state polls. The lines around the state polls results show the poll's margin of error.

Figure 6: MRP Estimates Compared to State Polls



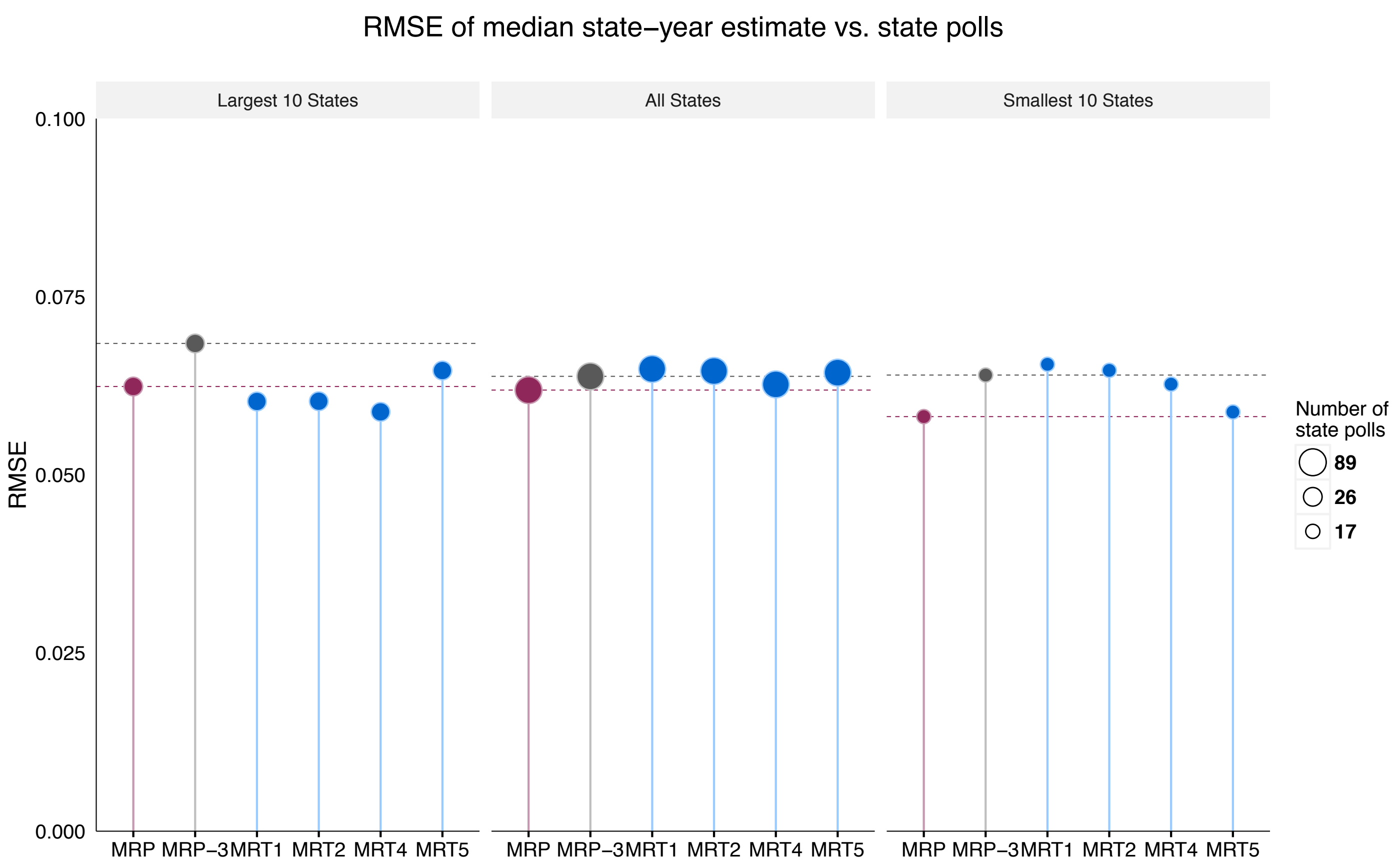
Note: The figure plots state opinion estimates (the solid line) along with their 90% confidence intervals (the shaded ribbon). The dots show the results of actual state polls. The lines around the state polls results show the poll's margin of error.

Figure 7: State Poll Predictions



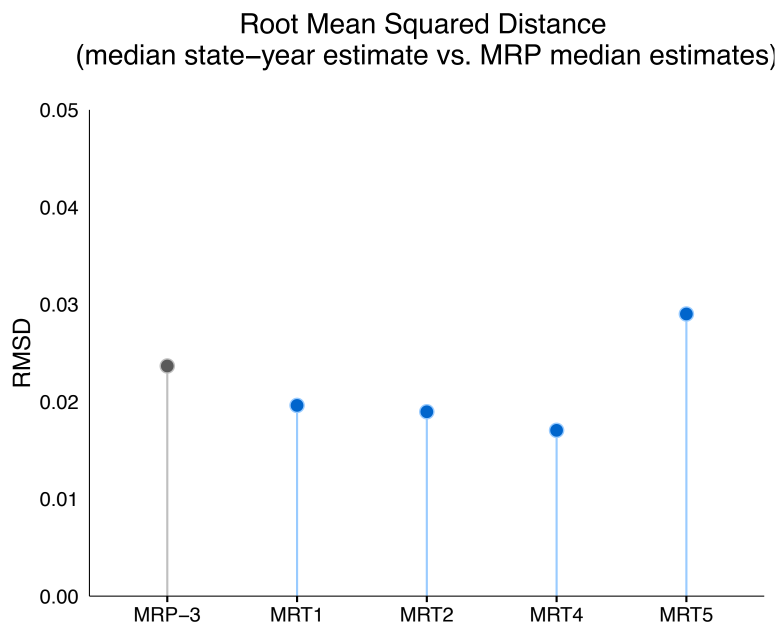
Note: These show the state polls against MRT estimates and MRP-3 estimates.

Figure 8: Average Error: Comparing MRT Estimates to State Polls, All Models



Note: Root-Mean Squared Distance The figure plots the average width of the 90% confidence interval for all dynamic MRP models as well the traditional MRP model and MRP-3. The subsets of largest and smallest states are among those states for which state polls exist.

Figure 9: Comparing MRT and MRP-3 Estimates to Traditional MRP



Note: Root-Mean Squared Distance The figure plots the average width of the 90% confidence interval for all dynamic MRP models as well the traditional MRP model and MRP-3.