Analytical and graphical methods for understanding hierarchical models with an application to estimating state trends in death penalty public opinion [1]

Kenneth E. Shirley[2] and Andrew Gelman[3]

July 26, 2011

Abstract

One of the longest running questions that has been regularly included in Gallup's national public opinion poll is "Do you favor or oppose the death penalty for persons convicted of murder?" Because the death penalty is governed by state laws rather than federal laws, it is of special interest to know how public opinion varies by state, and how it has changed over time within each state. In this paper we combine dozens of national polls taken over a fifty-year span and fit a Bayesian multilevel logistic regression model

[1] We thank the National Science Foundation, Institute of Education Sciences, Department of Energy, and National Security Agency for partial support of this work.

[2] AT&T Labs Research, Florham Park, NJ. kshirley@research.att.com, http://www.research.att.com/~kshirley/. A portion of this work was done while the author was a postdoc at Columbia University.

[3] Department of Statistics and Department of Political Science, Columbia University, New York. gelman@stat.columbia.edu, http://www.stat.columbia.edu/~gelman/

to individual response data to estimate changes in state-level public opinion over time. Such a long span of polls has not been analyzed this way before, partly because doing so requires a suitable model for the overall national time trend of death penalty public opinion, which is challenging to formulate.

In the context of the death penalty example, we develop here a suite of methods, largely graphical, for manipulating and understanding a fitted hierarchical model. In the death penalty problem we resolve the issue of modeling the national trend of support by using redundant parametrization and a structured prior distribution for the yearly effects. The resulting model can be fit using standard MCMC techniques, but the output of the model-fitting process is difficult to analyze immediately, as it is for many large hierarchical Bayesian models. The fitted model analyses we discuss in this paper include computing finite population contrasts and average predictive comparisons, and plotting posterior intervals of within-group standard deviations to compare different sources of variation within the data. We discuss inferences about the changing nature of death penalty support across time, states, and demographic groups that could not be made without using a variety of advanced tools for model understanding.

Keywords: Hierarchical Bayes, MCMC, Political Science, Gallup Survey, GSS, Multilevel Model

1 Introduction

Statisticians and applied researchers are increasingly turning to hierarchical models to handle problems with multiple sources of variation. Often our ability to fit such models has surpassed our techniques for understanding them. In the present article, we develop here a suite of methods, largely graphical, for manipulating and understanding a fitted hierarchical model, taking advantage of redundant-parameter models that were originally developed for computational purposes. We develop these ideas in a study of state-level trends in public opinion about the death penalty.

Capital punishment is perennially popular in the United States but is only legal in about two-thirds of the states (and is implemented rarely in many of these). To better understand the relationship between public opinion and policy, it would be desirable to know the support for the death penalty in each of the fifty states, and how this support has changed over time. To estimate state-level effects over time, it is necessary to control for the effects of national swings in public opinion as well as demographic variables, both of which are known to be large. As Figure 1 illustrates, national support for the death penalty has fluctuated substantially during the past fifty years, beginning with low support in the 1960s, increasing support throughout the 1970s (when capital punishment was ruled unconstitutional by the Supreme Court, followed by new rules under which the death penalty reappeared, one state at a time) high suppport in the 1980s (during which time a national concern about crime made the death penalty a prominent political issue), and finally, decreased support during the past decade (when five states either explicitly illegalized or indirectly suspended the death penalty in part due to the exoneration of numerous death row inmates due to DNA evidence).[4] Additionally, studies of individual surveys have repeatedly shown strong

[4]The issue even reached national politics, as in the 1988 presidential debate when death-penalty opponent

relationships between death penalty support and demographic variables such as sex, race, age, education, income, and religion, to name just a few. To effectively model state trends, national trends, demographic effects, and interactions between these sets of effects, we must gather a large amount of data and use a complex model.

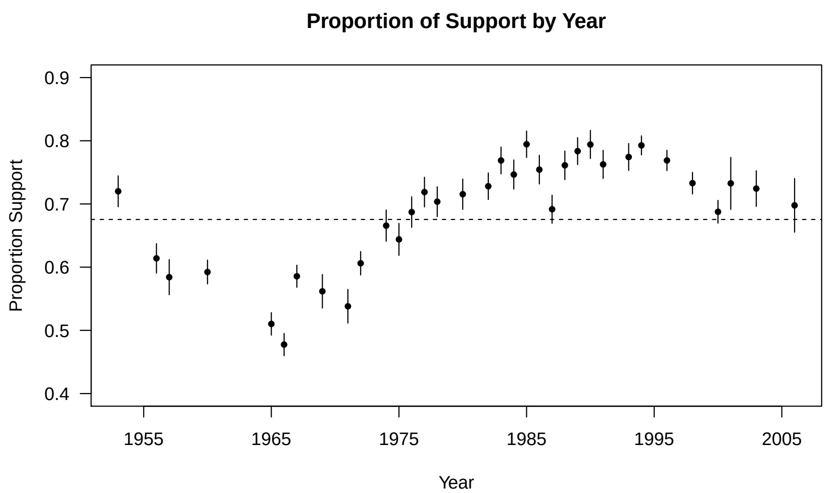

Figure 1: The proportion of respondents who support the death penalty by year, with bars representing 95% confidence intervals. The dashed horizontal line represents the overall proportion of death penalty support across all years, which is 67.5%.

Despite the extensive literature on the time series of aggregate death penalty support in the U.S. over the past fifty years, and also on individual-level predictors of death penalty support at specific time points, there has been very little rigorous, simultaneous analysis of both. One of the challenges in modeling demographic, time series, and state-specific effects simultaneously is that even after combining multiple surveys, the data are still too

Michael Dukakis was asked if his views would changed if his wife were raped and murdered.

sparse to estimate high-level interactions without some form of regularization. Furthermore, the national yearly fluctuations in death penalty support don't follow a straightforward pattern that can be embedded in a standard linear model. If such a model was fit to the differences between state-by-state levels of support and their average across states, this would be equivalent to treating the national average time series as if it were a known quantity, and the result would be the underestimation of the uncertainty of modeled quantities. Lax and Phillips (2009b) and Franklin (2001) discuss these problems in the context of public opinion polls and voting behavior.

Our approach to the problem is to fit a Bayesian multilevel logistic regression model to the data, where the model is overparametrized to allow for the simultaneous estimation of the national fluctuations in death penalty support as well as state-by-state deviations from it. The Bayesian approach handles the regularization required for sparse data via prior distributions for groups of parameters. The overparametrization of the model is a consequence of modeling yearly effects using an AR(1) model, and also modeling state trends as linear deviations from the national average. We follow Gelfand and Sahu (1999) and use weakly informative priors for the unidentified parameters in order to speed convergence of the contrasts of exchangeable parameters within groups, rather than impose constraints to make the model identifiable.

The model can be fit using standard MCMC techniques, but assessing convergence, checking the fit of the model, and interpreting the parameter estimates requires a variety of non-standard tools for understanding complex hierarchical models. We highlight our use of these tools throughout our discussion of the fit of the model and the resulting conclusions about death penalty public opinion. First, when we analyze posterior samples, we compute finite population contrasts to compare different units within groups, and we focus our convergence assessments and posterior summaries on these quantities rather than the "raw" parameters from the model. Second, we include graphical summaries of the fitted model that visualize the variation among respondents within different groups, in order to discover which categorical predictors explain the most variation in the outcome. The large number of categorical predictors and resulting interaction terms in our model make the visual summary of variability absolutely necessary for model understanding. Last, we compute average predictive comparisons to provide an additional high-level summary of which predictors explain the most variation in death penalty support. This is how we compare, on an equal scale, how much death penalty support has changed as a function of time, state residency, and demographic variables – a novel comparison that requires the combination of pooling many surveys over time, fitting a complex model, and summarizing the model fit in a succinct way.

We find that public support for the death penalty is highly associated with certain demographic variables, such as sex, race, and education, which is consistent with previous research. Our model, however, provides novel estimates of how these effects have changed over time, and how they vary across states, especially with regard to opinion as a function of race. We also find that support for the death penalty has changed significantly within certain states over time compared to the national average, holding constant the effects of demographic variables. In particular, we find that before the 1970s, capital punishment was more popular in the North than the South, a surprise given the current pattern in which the vast majority of executions are carried out in southern states. Additionally, we find that some of the variation among state trends can be explained by state-level variables including the legality of the death penalty and shifting partisan support over time, and the rest of the variation among states is explained by state-specific effects that we estimate with our model.

The models being developed and evaluated here are relevant not just for death sentencing and criminal justice but also more generally for studying the interactions between state-level opinion and policies, as discussed in literature from Erikson (1976) to Lax and Phillips (2009a). What is important is that the model allows for interactions between demographic, geographic, and time patterns. We fit the model using the open-source Bayesian program JAGS (Plummer, 2003). We make no apologies for using off-the-shelf software here; it is a strength of our methodology that it can work with relatively sophisticated models without intense programming effort. The key step is conceptual—modeling interactions at sufficient depth to allow for full poststratification, and summarizing the model using analytical and graphical tools that allow us to make the desired inferences.

The rest of the paper is organized as follows: Section 2 reviews previous work on death penalty opinions, Section 3 contains a description of the data to which we fit the model, Section 4 contains a detailed description of the model, Section 5 contains a description of the model parameter estimates and other analysis of the fitted model, Section 6 discusses the goodness of the fit of the model via posterior predictive checks and residual plots, Section 7 discusses the results of out-of-sample predictions made by the model and similar competing models, and last, Section 8 contains a discussion of the results.

2 Public opinion on the death penalty

Opinion on capital publishment has received a large amount of attention in the political science literature for a variety of reasons. First, the death penalty has consistently been an issue of national interest since the earliest national opinion polls were conducted by Gallup in the mid 1930's; thus, there exists a large amount of historical data concerning death penalty public opinion. Second, there have been multiple Supreme Court decisions concerning death penalty laws that cite changing public opinion as a factor in determining whether aspects of the death penalty constitute "cruel and unusual punishment." In *Weems v. United States* in 1910 and later in *Trop v. Dulles* in 1958, Supreme Court decisions specifically pointed out that the definition of "cruel and unusual" can change over time according to societal standards (Vidmar and Ellsworth, 1974). Much later, in 2002, Supreme Court Justice John Paul Stevens specifically mentioned that public opinion polls provided insight concerning the public's feelings toward the death penalty for mentally retarded prisoners (Hanley, 2008). Although individuals rarely exert direct control over death penalty laws through state-level referendums, these Supreme Court precedents show that indirectly, death penalty public opinion affects public policy. A number of articles have explored this opinion-policy relationship (Erikson, 1976; Norrander, 2000).

The basic relationships between demographic variables and death penalty support have been well understood for decades, but higher-level interaction effects have been less studied. During the period in the 1970s when the Supreme Court was shaping modern death penalty policy, Vidmar and Ellsworth (1974) reviewed then-current public opinion of the death penalty based on a 1972 Gallup poll and previous work by Erskine (1970) tabulating poll results from the late 1960s. They found that higher support for death penalty was associated with respondents who were male, white, old, and less-educated. Analyses in the 1990s found similar results (Fox et al., 1991; Ellsworth and Gross, 1994). The question of whether these effects have changed over time has not been answered rigorously. Baumgartner et al. (2008) suggest that associations between demographic variables and death penalty support are mostly fixed over time, but they do not fit a quantitative model to back up this claim. Hanley (2008) considers changes in support as a function of sex and race over time, and claims that in the 1990s, support among all race $\times$ sex subgroups was higher than it was in the 1970s, but does not consider the relative magnitudes of these differences (in Section 5 we show that while absolute levels of support may have been higher for all groups in the 1990s, relative levels of support for blacks were decreasing, and at slightly different rates for males and females). Hanley (2008) also points out that the relationship between age and support varied during the period from 1970-2000. Last, they found that the negative correlation between education level and support is not monotonic; the two least supportive educational groups are those without a high school degree, and those with a graduate degree of some kind (these two groups occupy opposite ends of the education scale). Our results in Section 5 are consistent with this finding.

The most recent and thorough time series model of death penalty public opinion is found in Baumgartner et al. (2008). They focused on modeling long-term trends in death penalty public opinion as a function of changing media coverage, using an index of support for death penalty based on a weighted average of yearly changes in support of death penalty from 292 statewide and national surveys between 1953 and 2006 that asked about the death penalty using different question wordings. They found a relationship between changes in public opinion and the tone of media coverage and levels of crime. They don't, however, model individual responses simultaneously as a function of time and demographic variables.

3 The data

We put together data on public opinion of the death penalty in two stages. We started with the 21 GSS polls given between 1974 and 2000, all of which asked the question: "Do you favor or oppose the death penalty for persons convicted of murder?" Second, we included data from Gallup polls taken before 1974 and after 2000 that asked a similar question: "Are you in favor of the death penalty for persons convicted of murder?" We searched for these Gallup polls in the archive of the Roper Center for Public Opinion Research, and to our knowledge, we included every Gallup poll in the Roper archive that (1) asked the question of interest, and (2) was given before 1974 or after 2000, with the exception of the Gallup polls given in 1936 and 1937, which we excluded because they didn't include detailed information about the education level of each respondent. The source of the poll is not a significant factor in the level of support expressed by respondents: Baumgartner et al (2008) found that the correlation between support levels from the two polling sources, GSS and Gallup, is 0.90 for the years in which the both asked the question of interest, and Schuman and Presser (1981) show that having formal balance in a survey question (explicitly suggesting a positive *and* negative answer in the question) rarely affects the outcome. In all, we modeled data from 34 polls, all of which were taken in distinct years between 1953 and 2006, where the maximum number of years between consecutive polls was 5 years (between the 1960 and 1965 Gallup polls). The number of respondents per poll ranged from 445 to 3085, and the total number of responses was $N = 58,253$.

Between 1953 and 2006, the proportion of poll respondents supporting the death penalty in a given year fluctuated between 47% and 79%, excluding those who had no opinion (the proportion of respondents with no opinion rarely exceeded 10%). Figure 1 contains a plot of death penalty support by year during this time. In each of the 34 polls we recorded the state of residence of each respondent, with the District of Columbia considered as the 51st state. The number of respondents per state per year was highly imbalanced, ranging from an average of 0.5 respondents per year in Hawaii (17 responses among the 34 surveys) to 170 respondents per year in California. We also classified states into 4 regions—Northeast, South, Midwest, and West—according to the U.S. Census state classification (with the District of Columbia included in the South).

The other information we used for our models was demographic information about each respondent, consisting of the following four variables:

1. Sex (male or female)
2. Race (black or non-black)
3. Age (a categorical variable with 4 levels, where 1 = 18–29, 2 = 30–44, 3 = 45–64, and 4 = 65 or older).
4. Education (measured as the highest degree achieved by the respondent, a categorical variable with 5 levels, where 1 = Less than high school, 2 = High school, 3 = Some college or trade school, 4 = College graduate, and 5 = Graduate degree). From this point forward, we call this variable "Degree."

Figure 2 displays some summary statistics of the distributions of each of these four variables, and includes the sample percentage of respondents in each main demographic category who supported the death penalty.

Last, we considered three state-level variables as potential explanatory variables in our model. The first two state-level variables are related to the state's partisan political support, as measured by their support for Republican vs. Democratic presidential candidates throughout the years. Specifically, for each presidential election year from 1952–2004, we recorded the Republican share of the vote in each state, discarding third party votes, and then we fit a linear regression model separately for each state to these percentages using time as the sole predictor. We recorded the estimated intercept and slope of this regression model for each state to create two state-level variables per state, which we call "Republican Share Intercept" and "Republican Share Slope."

Third, for each year from 1953 to 2006, we recorded the percentage of years in this 54-year time span that the death penalty was legal in each state. There is only a moderate amount

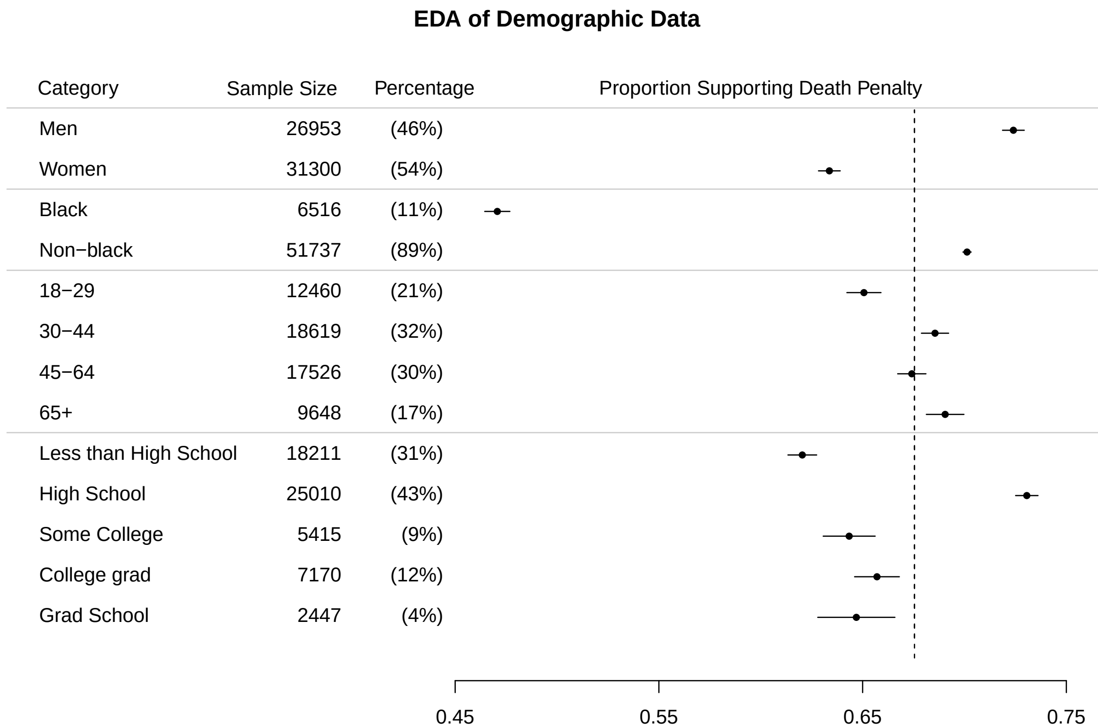

Figure 2: Exploratory data analysis of the demographic variables. The points are the proportions of respondents in each group who supported the death penalty (pooled across all years and states), and the horizontal lines are 95% confidence intervals for these proportions. Of the four demographic variables recorded, the largest difference in death penalty support exists between blacks and non-blacks. Age and education, which are ordinal variables, have non-monotonic relationships with death penalty support. The dashed vertical line is the overall proportion of death penalty support across all subgroups, which is about 67.5%.

of variation in this variable across states. The death penalty was legal in 40 states from 1953-1972 (and in Oregon from 1953-1963), and then a supreme court decision, *Furman v. Georgia*, effectively rendered every state's death penalty statute illegal in June, 1972. About 30 states rewrote their death penalty statute during the next five years, and seven other states rewrote their statutes at some point during the following 20 years.[5] In 1976, another supreme court decision, *Gregg v. Georgia*, overturned the 1972 ruling and made the death penalty legal again.[6] For our definition of this variable, which we call "Legality," if a state rewrote their death penalty statute between 1972 and 1976, we include these years as additional years in which the death penalty was legal for that state, even though technically every state was waiting for the court system to rule on the new version of the law during this time, and no executions were attempted. We code the variable this way to measure the degree to which each state supported the death penalty legislatively over this period; it is a slightly less precise variable than the strict percentage of years in which the death penalty was legal, but it increases the variation of this state-level variable, and it may be associated with public opinion. Across all states, the variable Legality has a maximum of 1 (Florida rewrote their death penalty law in 1972, the same year as *Furman v. Georgia*), and its minimum is 0 (10 states have never had a death penalty statute during this time span). The mean and standard deviation of Legality are 0.70 and 0.39, respectively.

In summary, there are six categorical variables per data point: year, state, sex, race, age, and education. Age and education are actually ordinal, but we include them as unordered categories in our model. Of the $54 \times 51 \times 2 \times 2 \times 4 \times 5 = 220,320$ distinct categories, we have

[5] Iowa, West Virginia, and the District of Columbia have never rewritten their death penalty statutes since 1972.

[6] In both the 1972 decision and the 1976 decision, the supreme court ruled on a specific death penalty case, and their ruling set a precedent that was applicable to death penalty laws nation-wide.

at least one observation from only 24,103 of them. This motivates the use of a multilevel Bayesian model.

4 The model

We fit a multilevel model to the the data with the main goal of understanding changes in public opinion of the death penalty for different states and demographic groups across time. We began by fitting simple models to the data, and we gradually increased the complexity of the models to include more explanatory variables at each step. In this section we present the most complicated model that fit the data well (as measured by a combination of criteria including out-of-sample prediction error, posterior predictive checks, prior subject area knowledge, and interpretability of results), and later in Section 7 we describe the gains we achieved by using this model over some simpler ones.

The first level of the model states that

$$\begin{aligned} p(Y_i = 1) &= \text{logit}^{-1}\Big(\alpha^{\text{state-year}}_{(s,t)[i]} + \alpha^{\text{degree-state}}_{(d,s)[i]} + \alpha^{\text{age-state}}_{(a,s)[i]} + \delta^{\text{age-state}}_{(a,s)[i]} X^{\text{year}}_i \\ &+ \beta^{\text{black-state}}_{s[i]} X^{\text{black}}_i + \beta^{\text{female-state}}_{s[i]} X^{\text{female}}_i + \beta^{\text{black-female-state}}_{s[i]} X^{\text{black}}_i X^{\text{female}}_i \\ &+ \delta^{\text{black-state}}_{s[i]} X^{\text{black}}_i X^{\text{year}}_i + \delta^{\text{female-state}}_{s[i]} X^{\text{female}}_i X^{\text{year}}_i \\ &+ \delta^{\text{black-female-state}}_{s[i]} X^{\text{black}}_i X^{\text{female}}_i X^{\text{year}}_i\Big), \end{aligned} \qquad (1)$$

for individual responses $i = 1, ..., 58253$, states $s = 1, ..., 51$, years $t = 1, ..., 54$, degrees $d = 1, ..., 5$, and ages $a = 1, ..., 4$. X^{year}_i is the year of response i measured as a continuous variable, scaled to have a mean of zero and a standard deviation of 1 ($X^{\text{year}}_i = 0$ corresponds to a response given in the mean survey year, 1980). X^{black}_i and X^{female}_i are likewise scaled to each have mean 0 and standard deviation 1 (which means that a black woman has the value $(X^{\text{black}}, X^{\text{female}}) = (2.82, 0.93)$, and a white man has the value $(X^{\text{black}}, X^{\text{female}}) =$

$(-0.35, -1.08)$). The subscript notation $s[i]$ denotes the state of residence, $s = 1, ..., 51$, for individual i.

The main feature of the priors for the parameters in Equation 1 is that every group of state-level parameters is normally distributed around a regional mean. Additionally, the state-year interaction effects have a structured model of their own. There is a lot of repetition in the setup of the priors, so here we write down some of them in full, and later we explain how the rest of the priors are analogous to these.

$$\alpha^{\text{state-year}}_{(s,t)} \sim \mathrm{N}(\alpha^{\text{year}}_t + \alpha^{\text{state}}_s + \delta^{\text{state}}_s X^{\text{year}}_t, \sigma^2_{\text{state-year}}), \tag{2}$$

$$\alpha^{\text{year}}_t \sim \mathrm{N}(\mu + \mu_\delta X^{\text{year}}_t + \phi(\alpha^{\text{year}}_{t-1} - \mu - \mu_\delta X^{\text{year}}_{t-1}), \sigma^2_{\text{year}}), \tag{3}$$

$$\alpha^{\text{year}}_1 \sim \mathrm{N}(\mu + \mu_\delta X^{\text{year}}_1, \sigma^2_{\text{year}}/(1 - \phi^2)), \tag{4}$$

$$\mu \sim N(0, 5^2), \tag{5}$$

$$\mu_\delta \sim N(0, 5^2), \tag{6}$$

$$\phi \sim U(-1, 1), \tag{7}$$

$$\alpha^{\text{state}}_s \sim \mathrm{N}(\alpha^{\text{region}}_{r[s]} + \boldsymbol{\beta X}^{\text{state}}_s, \sigma^2_{\text{state}_{r[s]}}), \tag{8}$$

$$\alpha^{\text{region}}_r \sim \mathrm{N}(0, \sigma^2_{\text{region}}), \tag{9}$$

$$\beta_j \sim \mathrm{N}(0, 5^2) \qquad \text{for } j = 1, 2 \tag{10}$$

$$\delta^{\text{state}}_s \sim \mathrm{N}(\delta^{\text{region}}_{r[s]} + \boldsymbol{\gamma Z}^{\text{state}}_s, \tau^2_{\text{state}_{r[s]}}), \tag{11}$$

$$\delta^{\text{region}}_r \sim \mathrm{N}(0, \tau^2_{\text{region}}), \tag{12}$$

$$\gamma_j \sim \mathrm{N}(0, 5^2) \qquad \text{for } j = 1, 2 \tag{13}$$

$$\alpha^{\text{age-state}}_{(a,s)} \sim \mathrm{N}(\alpha^{\text{age-region}}_{(a,r[s])}, \sigma^2_{\text{age-state}_{(a,r[s])}}), \tag{14}$$

$$\alpha^{\text{age-region}}_{(r,s)} \sim \mathrm{N}(\alpha^{\text{age}}_a, \sigma^2_{\text{age-region}_a}), \tag{15}$$

$$\alpha^{\text{age}}_a \sim \mathrm{N}(0, \sigma^2_{\text{age}}), \tag{16}$$

$$\beta_s^{\text{black-state}} \sim \mathrm{N}(\beta_{r[s]}^{\text{black-region}}, \sigma^2_{\text{black-state}_{r[s]}}), \tag{17}$$

$$\beta_r^{\text{black-region}} \sim \mathrm{N}(\beta^{\text{black}}, \sigma^2_{\text{black-region}}), \tag{18}$$

$$\beta^{\text{black}} \sim N(0, 5^2). \tag{19}$$

We also specify that the prior distribution for every standard deviation parameter is a half-t distribution with scale 5, and 3 degrees of freedom. There are 110 such parameters in the model. $\boldsymbol{X}^{\text{state}}$ and $\boldsymbol{Z}^{\text{state}}$ denote the (51 x 2) matrices of state-level covariates that affect the state intercepts and slopes, respectively, where $\boldsymbol{X}_s^{\text{state}} =$ (Republican Share Intercept, Legality)$_s$, and $\boldsymbol{Z}_s^{\text{state}} = (\text{Republican Share Slope}, \text{Legality})_s$ for states $s = 1, ..., 51$.

The parameters left out of equations 2 through 19 have priors that are identical in structure to some of those in equations 2 through 19. First, the prior distributions for the degree-state intercepts, $\alpha_{(d,s)}^{\text{degree-state}}$, are exactly the same as the priors for $\alpha_{(a,s)}^{\text{age-state}}$, where "degree" replaces "age" in every specification, and there are five levels of degree effects (rather than four for the age effects). Next, the prior distributions for the age-state slopes, $\delta_{(a,s)}^{\text{age-state}}$, are exactly the same as the priors for the age-state intercepts, except that slope parameters, δ, replace the intercept parameters, α, and the standard deviation parameters are denoted by τ rather than σ. Last, there are five additional sets of race-sex effects whose priors are not specified in equations 2 through 19. Each set has a prior distribution identical in structure to the prior distribution for β_s^{black}, where intercepts, β, and slopes, δ, for individual states are normally distributed around regional means, which are, in turn, normally distributed around the grand mean, which is given a weakly informative $N(0, 5^2)$ prior. The supplemental files contain a graphical illustration of the full model in the form of a directed acyclic graph (DAG).

We opt to model the national average yearly effects, α_t^{year}, as an AR(1) process with a linear trend, where the differences between individual states and this national average yearly pattern are modeled as linear (on the logistic scale). We assume the AR(1) process is stationary so that we can estimate the overall mean across years (rather than conditioning on the first year, or anchoring the mean to a given year as we would have to do if the process weren't stationary). We chose not to expand the AR(1) model to individual states because we felt the assumption that each state was individually stationary might not be realistic. We discuss the goodness-of-fit of the linearity assumption for states in Section 6.

We fit the model using JAGS (Plummer, 2003), which implements a mix of Gibbs sampling, slice sampling, and Metropolis jumping, and we performed all pre- and post-processing in R. Before we ran the MCMC, we computed the binomial count of those who supported the death penalty for each observed state-year-demographic 6-way combination (there were 24,103 unique state-year-demographic combinations with at least one observation in the data), so as to save time in the model fitting by modeling the sufficient statistics rather than each individual data point.

5 Results

5.1 MCMC Convergence and Post-processing

We ran the MCMC algorithm on three separate chains for 10,000 iterations each, and we saved every fifth iteration among the last 5,000 to form a posterior sample of size 1,000 for each of the three chains. (We thinned the output only for convenience of manipulating a smaller amount of posterior output in R).

One of the main challenges to understanding the raw output from the MCMC algorithm is that the model we fit is overparametrized. This overparametrization comes in two different varieties, and in each case we disentangle non-identifiable parameters using parameter adjustments in the post-processing stage of model-fitting. The first case is simple: We adjust our MCMC output to reflect finite population contrasts, because for many of our groups of parameters (such as groups of states, age levels, and education levels), the members of the group constitute the entire population, rather than a sample from a larger population. Doing this allows us to make more precise inferences about differences between observed units in a given group (Gelman and Hill, 2007). So, for example, when a group of parameters is distributed around an unknown mean, such as the set of race intercepts, $\beta_s^{\text{black-state}}$ and their mean, β^{black}, we compute an adjusted version of each parameter:

$$\beta_s^{\text{black-state}\prime} = \beta_s^{\text{black-state}} - \beta_{\cdot}^{\text{black-state}}, \tag{20}$$

$$\beta^{\text{black}} = \beta_{\cdot}^{\text{black-state}}. \tag{21}$$

The final set of these adjustments is made slightly more complicated by the nesting of state effects within regions, but the basic principle remains the same.

The second form of overparametrization is slightly more complicated to disentangle, and is related to the model for the time trend of death penalty support. One of the key components of our model that allows us to simultaneously estimate national fluctuations in opinion as well as state-by-state trends is the AR(1) model we use for the yearly effects, which treats the years as different levels of a categorical variable. We also modeled state effects as linear time trends with a mean linear trend, μ_δ. Such an overparametrized model is used because it preserves the uncertainty in the overall year-to-year fluctuations in death penalty support while simultaneously modeling state deviations from this pattern in a parsimonious way. The result is a lack of identifiability among the yearly effects and the mean slope. In order to capture the true mean slope, we compute the adjusted version of μ_δ by summing over every component of the model that allows for a linear trend in death penalty support:

$$\mu'_\delta = \frac{\sum_{t=1}^{T} X_t^{\text{year}}(\alpha_{.t}^{\text{state-year}} - \alpha_{..}^{\text{state-year}})}{\sum_{t=1}^{T}(X_t^{\text{year}})^2} + \delta_{..}^{\text{age-state}}. \tag{22}$$

The first term on the right side of Equation 22 is the estimated slope of death penalty support embedded within the state-year effects, and the second term is the mean slope across all age-state combinations. Similarly, to capture more precise estimates of the yearly effects, subtracting out the mean slope, we compute the adjusted yearly effects:

$$\alpha_t^{\text{year}'} = \alpha_{.t}^{\text{state-year}} - \alpha_{..}^{\text{state-year}} - (\mu'_\delta - \delta_{..}^{\text{age-state}})X_t^{\text{year}}. \tag{23}$$

These adjustments result in more precise estimates of the quantities of interest, which are sets of centered parameters whose mean is zero, and the accompanying means themselves, which are adjusted to be identifiable. The full set of adjustments that we make is detailed in the supplementary materials. From this point forward, when we refer to a parameter, we are actually referring to its "adjusted" value as written with a prime symbol ($'$) in the supplement.

We computed the potential scale reduction factor (Gelman and Rubin, 1992) and the effective sample size for each adjusted parameter (there were 4099 of them). The range of the potential scale reduction factors was (1.00, 1.03), indicating that the chains mixed well on all measured dimensions. The effective sample sizes of these adjusted parameters ranged from about 200 to 3000 (where an effective sample size of 3000 means that the autocorrelation of the three chains for a given parameter was virtually zero).

5.2 National average trends and yearly effects

The quantity $\text{logit}^{-1}(\mu + \mu_\delta X_t^{\text{year}} + \alpha_t^{\text{year}})$ is the mean support for the death penalty in year t for a respondent from an average state with average demographic variables. By "average" state and demographic variables, we mean that the respondent's state of residence, age, and education level are unknown (which is equivalent to the average of the group, because these group effects have been scaled to have a mean of zero), and their values of X^{black} and X^{female} are zero.

The mean proportion of support for the death penalty for the average respondent in this time period is estimated to be about 66.8% (calculated as the posterior mean of $\text{logit}^{-1}(\mu)$), and the mean linear change in support per year is estimated to be about 0.50%. This means that the linear component of support for the death penalty for the average respondent increased by about 1% every 2 years during the years 1953-2006, or by a total of about 27% during this time span. This trend is modeled as being linear on the logistic scale ($\bar{\mu_\delta} = 0.37$), which means that the actual *proportion* of people supporting the death penalty is not technically modeled as linear. The difference, however, is slight: on the probability scale, the estimated curve has a slope of about 0.58% at the beginning of the time span (1953-54), and a slope of about 0.39% at the end of the time span (2005-06).

The yearly effects, which are modeled as deviations from the linear time trend, follow an AR(1) model (on the logistic scale) where $\hat{\phi} = 0.92$ and the standard error of ϕ is about 0.06. The posterior distribution of ϕ is skewed to the left since it is bounded on the right by 1. The year-to-year standard deviation, σ_{year}, is about 0.17 (+/- 0.03) on the logistic scale. On the probability scale, this means that the estimated standard deviation of the change in one year's proportion of support, given the previous year's proportion, is about 3-4%. The marginal standard deviation of the estimated yearly support for an average respondent is about 11.1% (with a mean, as we said earlier, of about 66.8%).

Figure 3 shows posterior means and intervals for the proportion of death penalty support in each year for an average respondent, including the posterior means of the intercept and slope of the linear trend. Each yearly estimate is essentially a weighted average of the observed proportion of support in that year, the linear trend across all years, and the proportions in neighboring years.

5.3 State and regional trends

There was substantial variation in the trends among states. Figure 4 shows the estimated differences between each state's level of support (for the average respondent) and the national average on the probability scale, where the states are grouped by region. The different patterns of variation between the four regions are clear—the western states are the most variable in their intercepts (with Utah respondents showing high support and Hawaii respondents showing low support), and the northern states are the most variable in their slopes (with Massachusetts and Maine exhibiting low and high slopes, respectively). The average support among western and northern states decreased over time relative to the national average. The southern and midwestern states are somewhat less variable in their slopes and intercepts, and both of these groups of states gradually increased their support over the time span, relative to the average. The differences between states and the national average are modeled as linear on the logistic scale, but again, when transformed to the probability scale, they are not linear, especially when the proportions approach the upper boundary of one.

Our model explains this variation among state trends using three types of variables: State-level variables, regional effects, and state-specific variation. Figure 5 summarizes the amount of variation explained by each of these three sets of variables using point estimates and intervals of the group-level standard deviation estimates. We describe these effects in more detail in the following two subsections.

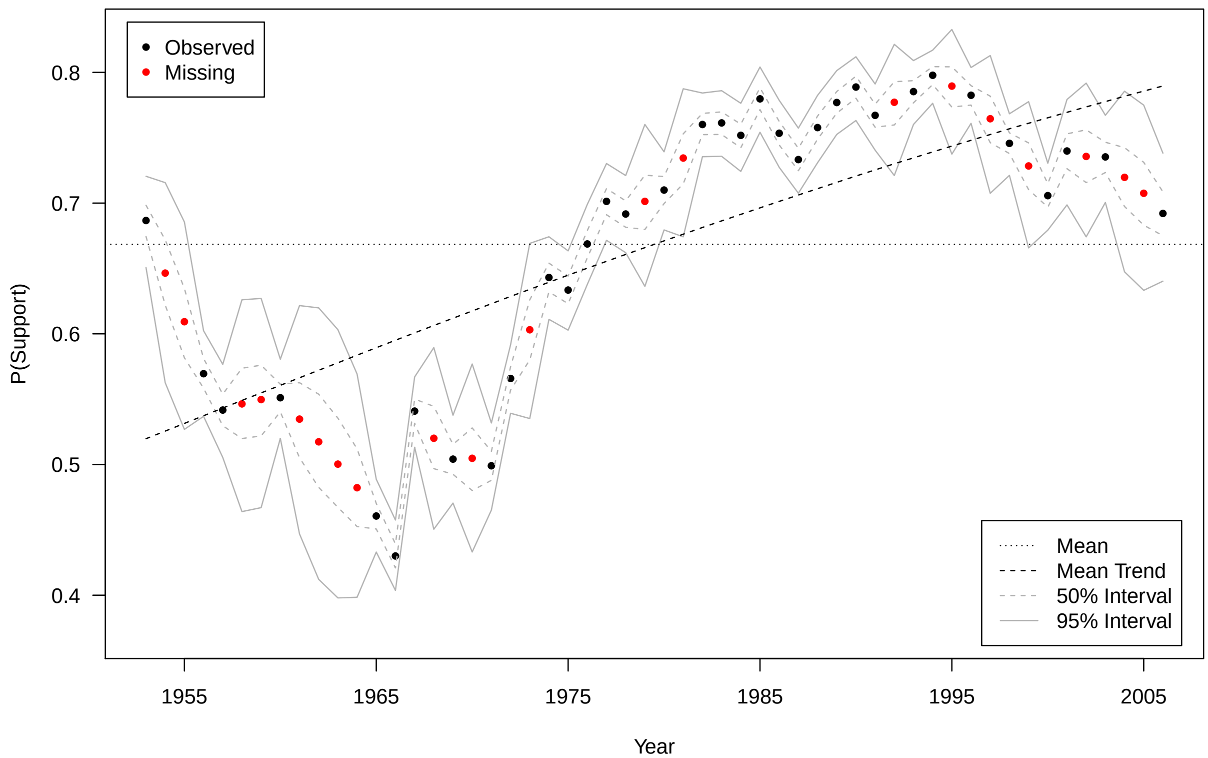

Figure 3: National average trend and yearly effects. The black and red points are posterior means of the proportion of support for an average respondent in each year from 1953-2006. The black points denote the years in which we have survey data, and the red points denote years in which there was no survey data. As expected, the intervals are wider for years of missing data, and are especially wide when there is a multi-year gap between consecutive surveys. When we extrapolate far into the future, we find that the estimates gradually shrink toward the average trend (dashed line), as expected, and they lie exactly on this curve after about 20 years.

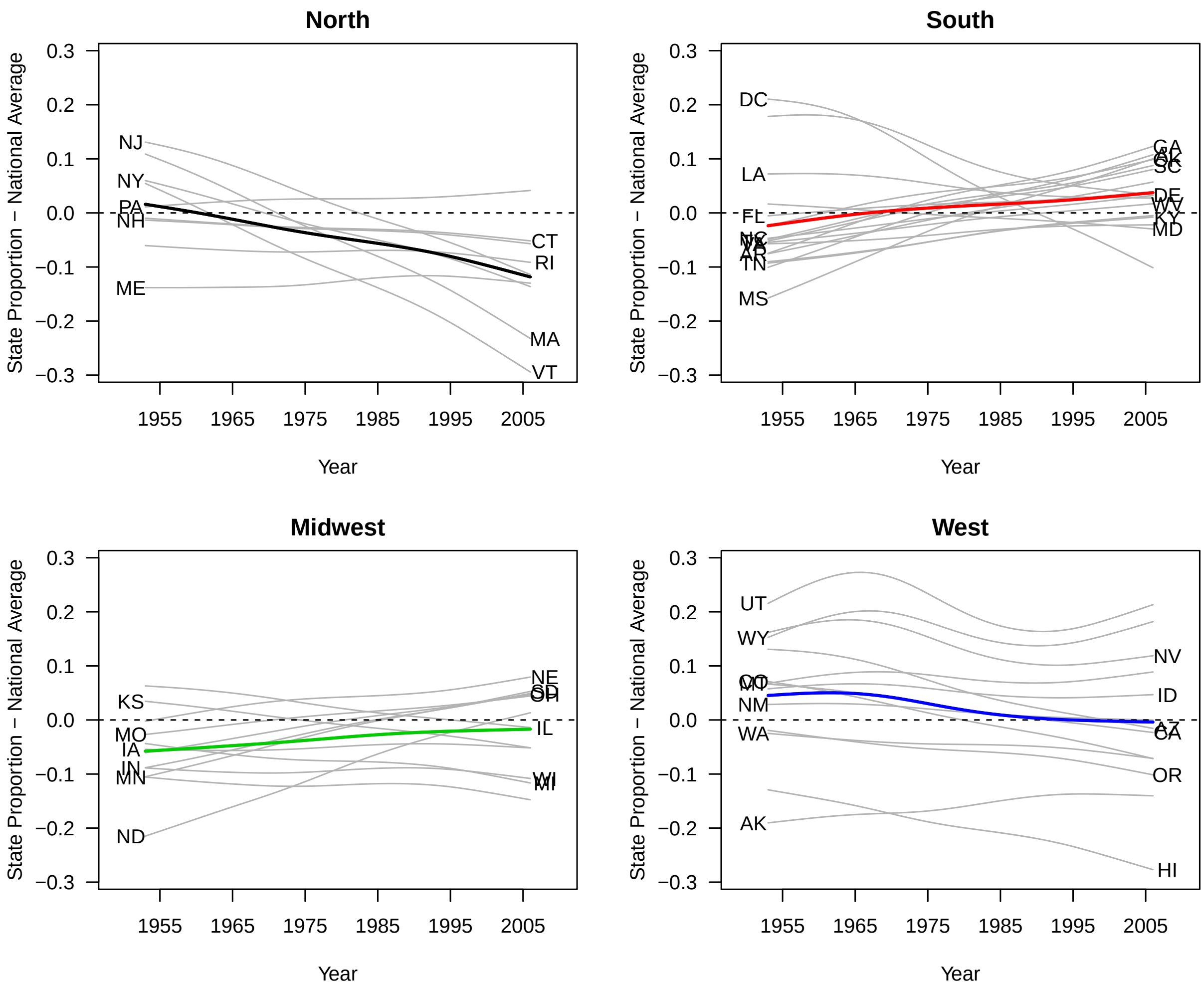

Figure 4: Posterior means of the differences in the proportion of support between states and the national average for a respondent of "average" demographics (see Section 5.2), grouped by region, where the colored curves are regional means. The curves are not linear because of the transformation from the logistic scale to the probability scale. All 4 plots are on the same scale to highlight the differences in variation among states between the four regions. The state intercepts vary the most among western states, while the state slopes vary the most among northern states. The state labels here aren't all legible; for clearer state summaries, see Figure 6.

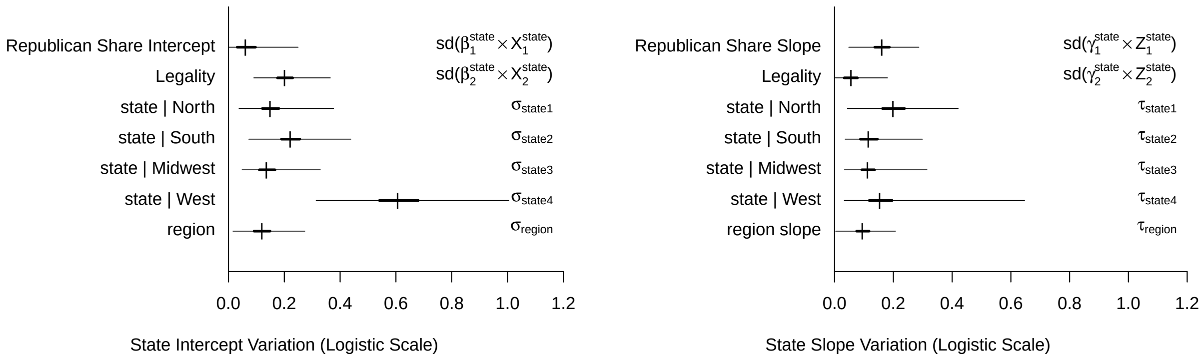

Figure 5: The vertical hash marks are point estimates of group-level finite-population standard deviation parameters, and the thick and thin horizontal lines represent 50% and 95% intervals for these parameters. In the left plot, it is clear that the intercepts of the western state-specific effects have the largest variability, and "Legality" also explains a substantial proportion of variation in the state-level intercepts. In the right plot, the same is true of the slopes of the northern states, and of the state-level variable "Republican Share Slope."

5.3.1 Effects of state-level variables

The estimated effects of the state-level variables "Republican Share Intercept" and "Legality" on the intercepts of the state trends ($\boldsymbol{\beta}$ from Equation 8) are -0.06 and 0.20, with standard errors of 0.06 and 0.04, respectively. The first estimate shows that higher average levels of support for Republican presidential candidates during this time span are associated with lower levels of death penalty support, but the interval estimate of this effect contains zero, so it is not statistically significant. On the other hand, the estimated effect of legality of the death penalty on state intercepts is large and statistically significant, indicating that states where the death penalty was legal during a longer proportion of the time span 1953-2006 also show higher average levels of support for the death penalty.

A different story emerges when we look at the effects of state-level variables on the *rate of change* of death penalty support over time (the parameters $\boldsymbol{\gamma}$ from Equation 11). In this case, the estimated effect of the slope of Republican presidential support over this time span on the slope of state-level death penalty support is 0.16 with and sd of 0.03. That is, states that increased their support for Republican presidential candidates over this time span also increased their death penalty support during this time, and the association is large and statistically significant. The converse is also true: states that decreased their support for Republican presidential candidates during this time period also decreased their level of support for the death penalty. The legality of the death penalty has a small, positive, statistically insignificant association with the slope of state-level support for the death penalty (the estimated effect is 0.05 with an sd of 0.03). In other words, support for the death penalty increased slightly more during this time period among states in which it was legal compared to those in which is was illegal.

From Figure 5, you can see that these state-level variables explain a substantial amount of the variation in state trends—comparable to the amount of variation explained by the state and region-level errors. The associations between death penalty support and other state-level variables, such as executions, crime rates, and other state-level public policies, could be estimated this way, too.

5.3.2 Regional effects and state-level errors

The state trends also depend on regional effects, and state-specific coefficients that are centered at zero within each region. The posterior means of both sets of these effects are plotted in Figure 6. The state-specific varying intercepts and slopes estimate the predictive effects of residing in a given state that are not already explained by whatever state-level variables are included in the model—in our case, each state's Republican voting trend and proportion of years of legality.

Recall that the amount of variability among the state-specific intercepts and slopes is allowed to be different for each of the four regions; in fact, the data justify these separate estimates of variability. The posterior means of the standard deviations of the state-specific varying intercepts and slopes are, in order of the regions (North, South, Midwest, West), $\hat{\boldsymbol{\sigma}}_{\text{state}} = (0.15, 0.22, 0.14, 0.61)$, and $\hat{\boldsymbol{\tau}}_{\text{state}} = (0.20, 0.12, 0.12, 0.16)$ (these point estimates are plotted in Figure 5). That is, the intercepts of the western states are highly variable, as are the slopes of the northern states. The variabilities of the state-specific intercepts and slopes in the south and midwest are less, but still non-zero. We already saw this in Figure 4, but that figure displayed state trends that included the effects of state-level variables. Here we are describing the residual trends in each state attributed to unmeasured state-level variables. The regional means of the state-specific slopes and intercepts are $\hat{\boldsymbol{\alpha}}^{\text{region}} = (-0.10, -0.03, -0.05, 0.16)$, and $\hat{\boldsymbol{\delta}}^{\text{region}} = (-0.02, -0.01, 0.11, -0.11)$. In other words, these are the estimated intercepts and slopes of death penalty support relative to the national average for a respondent of average demographics from a random state in one of these regions. The average of the state-specific varying slopes for northern states is almost zero, whereas most northern states have strong decreasing trends of support compared to the national average as pictured in Figure 4. This is because most of the decline in support for the death penalty among northern states is explained by their decreased support for Republican presidential candidates during this time span. This isn't a *causal* effect, but it is merely an association between Republican vote share and death penalty support that helps us measure the effects of unmeasured state-level variables more efficiently.

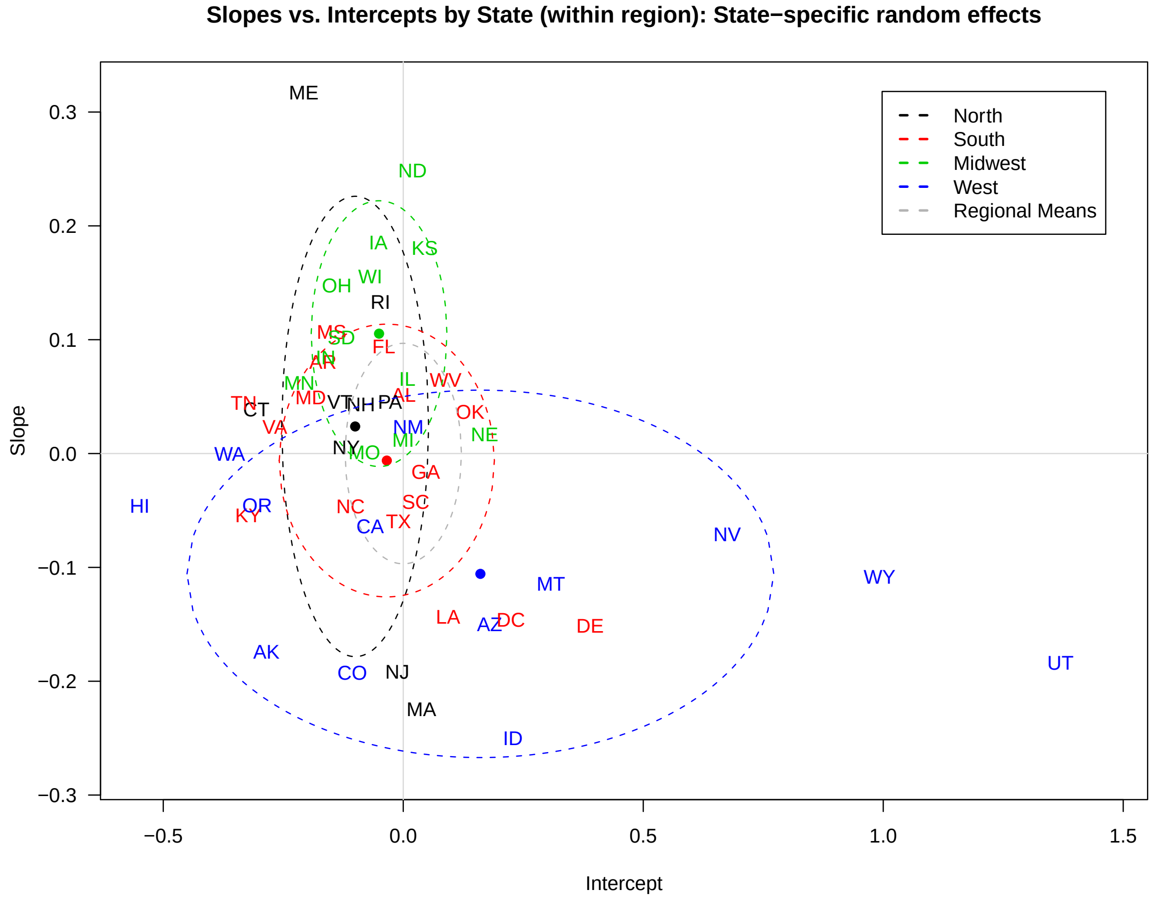

Figure 6: Posterior means of state-specific slopes vs. intercepts for all 51 states. The dotted ellipses are drawn so that the lengths of the horizontal and vertical axes are equal to the posterior means of the regional standard deviations of the intercepts and slopes, respectively, and the gray dotted ellipse in the center is drawn with (horizontal axis, vertical axis) = (0.12,0.08), the posterior means of σ_{region} and τ_{region}, respectively. States are colored by region as usual. These effects in this figure do not include the effects of state-level variables on state slopes and intercepts.

We'll make a few more comments here regarding state-level trends, as seen in Figures 5 and 6:

- Among the 9 northern states, Maine and Rhode Island have positive state-specific slopes, and relatively flat slopes when the effects of state-level variables are included. The rest of the northern states have relatively flat or negative state-specific slopes, and negative slopes when the effects of state-level variables are included. The three states with the fastest-decreasing support for the death penalty over this time period are Massachusetts, New Jersey, and Vermont. Most of Vermont's decrease in death penalty support is predictable from the model given Vermont's declining support for Republican presidential candidates over this time period.

- The slopes of the western states vary little. Their intercepts are highly variable, though, with Hawaii and Alaska showing low support for the death penalty, and Utah and Wyoming showing high support. Most of the lower levels of support in Hawaii and Alaska can be explained in the model from the fact that the death penalty has never been legal in those states, whereas for the rest of the western states, it has been legal during most of the time span in question.

- Although the District of Columbia and Delaware are classified by the U.S. Census Bureau as southern, their slopes are much more similar to the northern states' slopes. Their slopes are negative, indicating declining support for the death penalty relative to the national average.

Last, we visualize the intercepts and slopes for each state relative to the national average using colored maps in Figure 7. Some regional correlations are visible in the maps, but overall the maps make it clear that the variation between the states is greater than that between the regions.

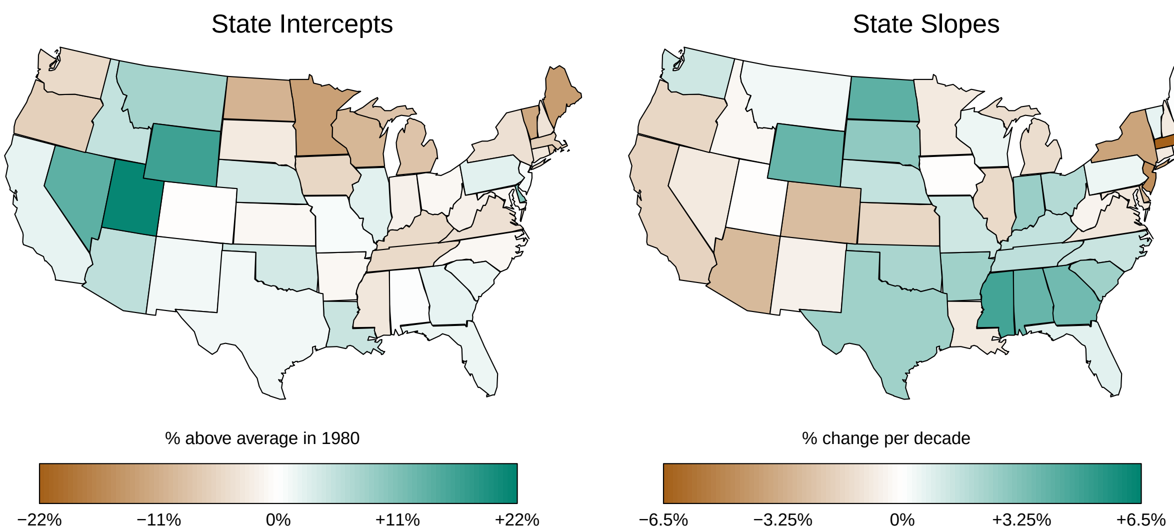

Figure 7: The slopes and intercepts of the state trends in death penalty support relative to the national average (67%) are visualized here on maps of the United States using a color scale. Time was scaled to have a mean of zero, which means that the intercepts are estimated levels of support for the mean year of the sample, which is 1980.

5.3.3 State-year interaction effects

Our model also includes state-year interaction effects to account for additional variation on the state-year level. For example, when a highly-publicized crime occurs, or when a murder trial receives a large amount of media attention, it is plausible that death penalty support in that state at that time could experience a relatively sudden increase or decrease that would not be captured by the state trends that are included in our model. The posterior mean of the standard deviation of the state-year interaction effects, $\sigma_{\text{state-year}}$, was 0.27 with a standard error of about 0.02; in other words, the state-year interaction effects explain a substantial amount of variation in the response, and the estimate of their variability is precise. The precision of the estimate of $\sigma_{\text{state-year}}$ is a result of the large number of state-year interactions that are contained in the model (there are 51×54 of them)—it is fairly easy to estimate the variability of such a large set of parameters compared to, say, the variability of the slopes of states within a region, where there are only about 10 parameters in the group. A potentially useful extension to the model that we didn't explore for this paper would be to allow the variance of the state-year interactions to vary by state, to see if certain states show more volatility from year to year in their state-year effects. If this were true, it might suggest that the death penalty is a more salient issue in some states than others, and is therefore more prone to public opinion swings on a short time scale.

A few examples of individual state-year effects are visualized in Figure 13 (in Section 6). They are re-centered (see the supplement on parameter adjustments) so that the mean of the state-year effects is zero for each year and state. Their general characteristic is that they are a weighted average of the level of support in a given state and year and the mean level of support for that state according to its state-level trend. The amount of shrinkage in the estimate (from the observed level of support in a given year and state toward the state-level trend) depends on the sample size for that given year and state, where larger samples shrink less.

5.4 Demographic effects and trends

Individual demographic variables also explain a substantial amount of the variation in death penalty support during the time span 1953-2006. Recall from Section 4 that we model the effects of race, sex and their two-way interaction as a linear trend on the logistic scale, and we allow these trends to vary by state (for a total of $2 \times 2 \times 51$ intercepts, and the same number of slopes). We also model the effects of education (measured by the highest degree earned by the respondent) on the intercept of the logit of the probability of support, and we allow these effects to vary for each state (for a total of 5×51 degree-state effects). Last, we model the effects of age on death penalty support as a linear trend, where the linear trends vary by state, resulting in 4×51 age-state lines that are estimated.

5.4.1 Trends related to race and sex

Figure 8 shows the estimated difference between the level of support among individuals of each of the four race-sex combinations that we consider (black/non-black $\times$ female/male) compared to their weighted average (which is weighted by the frequencies of each of these 4 categories in the sample, pooled across all years, as described in Section 5.2).

Black males have shown the sharpest decline in support over this time period, with an average decrease in support of about 1% every two years compared to the national average, starting with 6% lower estimated support in 1953, declining to estimated levels of support that were about 37% below the national average in 2006. Black females have also decreased their support over time, at almost the same rate, but not quite as steeply; their estimated support decreased from about 16% below the national average to about 38% below the national average. Averaged across these years, black females showed the least support over time for the death penalty of the four race-sex combinations that we consider here. Non-black males showed the highest average levels of support over this time period, increasing their estimated support from about 7% above the national average to about 10% above the national average. Last, non-black females began this time period with estimated support

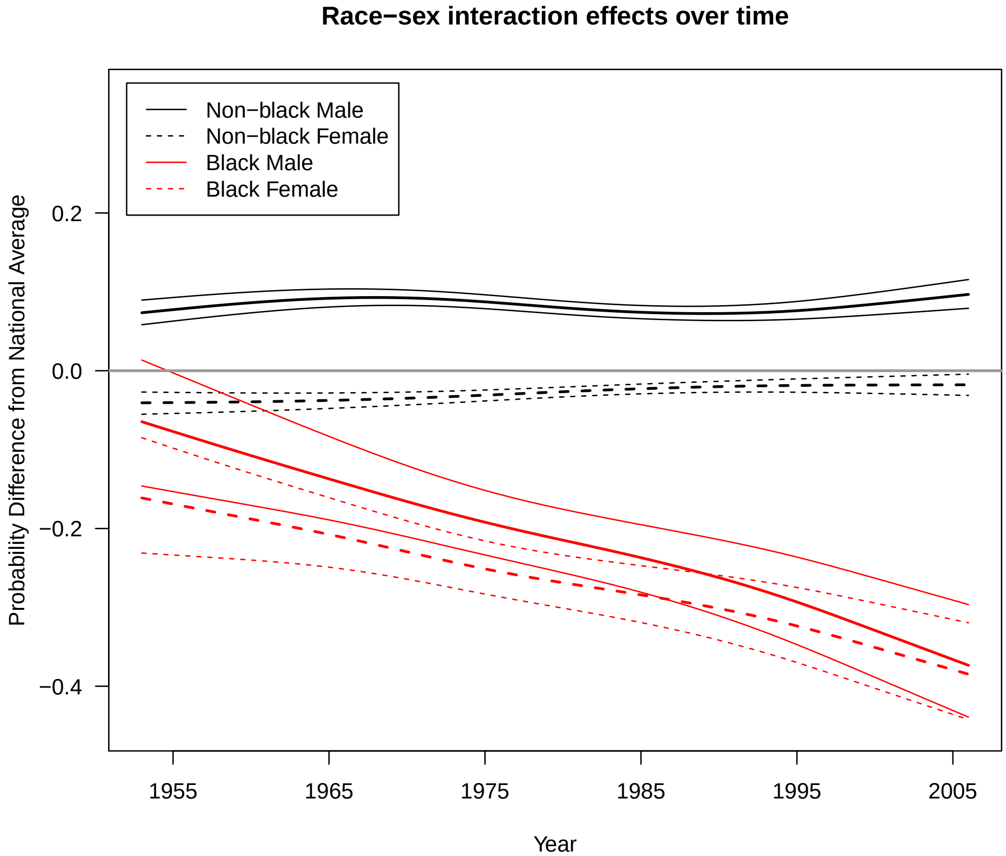

Figure 8: This plot shows the estimated differences between support levels over time for each of the four race-sex demographic groups compared to their average (for an unknown state, and average levels of age and education). The three lines associated with each group represent the mean and the boundaries of the 95% interval for the difference in support for that group and their average over time. Residuals are plotted in Figure 9.

about 4% below average, and increased their relative support by a total of about 2% over the time period.

These trends among race-sex groups were previously confounded by the nationwide fluctuations in death penalty support over time. Hanley (2008) points out that support was higher in the 1990s than in the 1970s among many subgroups of the population, including blacks, despite high-profile court cases in the late 1980's that found evidence for bias against blacks in death sentencing (which presumably would result in lower support among blacks). To investigate such a claim, the relevant analysis is to compare support among blacks to support among non-blacks, controlling for nationwide trends and fluctuations over time. Our model provides this comparison and shows that support among blacks declined on average during the 54-year stretch from 1953-2006 relative to non-blacks, as pictured in Figure 8. The residual plots in Figure 9 generally show no pattern among the residuals over time, indicating that a linear model for support among race-sex groups over time is a good fit to the data.

Further, Hanley (2008) claim that "since 1990 black men have been more supportive of the death penalty than black women, after a period of fluctuation in the late 1970s and 1980s." Our analysis suggests that any such fluctuation was due to confounding variables or randomness, and that (1), on average, black men have always supported the death penalty at higher levels than black women, and (2) this gap is actually decreasing over time, so that in the near future it is possible that black men will show lower levels of support for the death penalty than black women.

These trends among race-sex groups are allowed to vary by state, where the variability of the state-specific trends for each race-sex subgroup is estimated separately for each of the four U.S. regions. In general, there was more variability in race-sex trends between states within

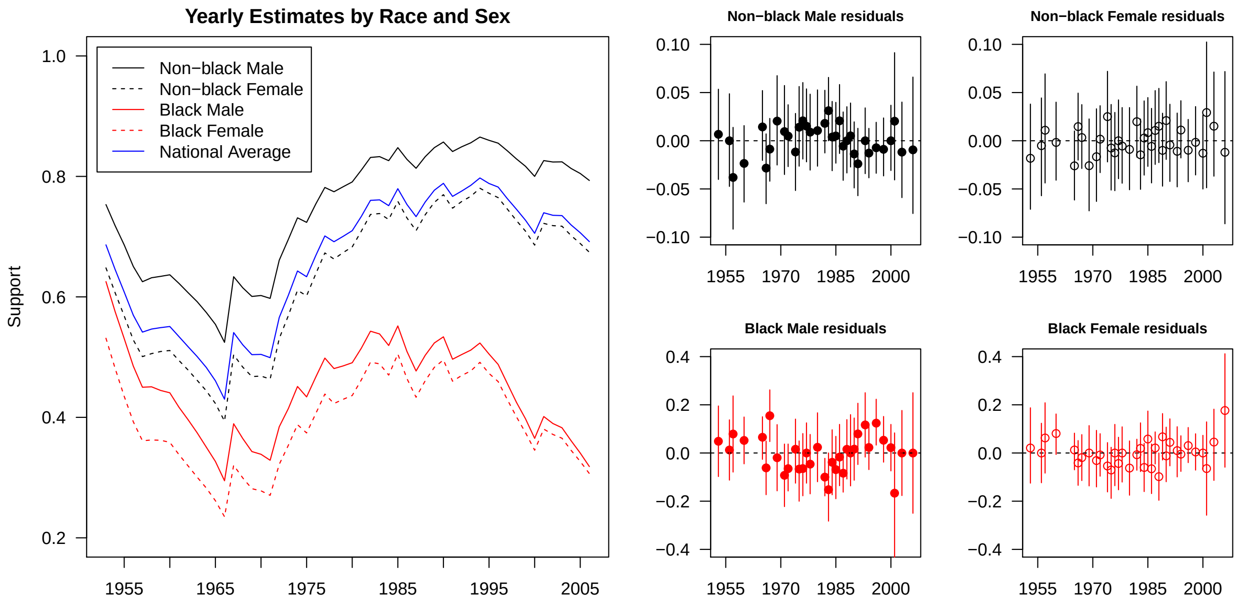

Figure 9: On the left we plot the estimated levels of support for the death penalty for each of the four race-sex groups; this plot is similar to Figure 8, except the yearly pattern has not been subtracted. On the right, we plot residuals of support (observed - fitted) for each race-sex group over time, conditional on the observed states or residence, ages, and educational levels of each group. The vertical lines are 95% posterior intervals for the estimated differences. The y-axes of the residual plots are different for blacks and non-blacks, to better visualize the residuals.

the same region than there was between the regions themselves. There was a particularly large amount of variation between the intercepts of racial trends among northern states ($\sigma_{\text{black-state}_1} \approx 0.11$ +/- 0.06), and also between the slopes of racial trends among southern states ($\tau_{\text{black-state}_2} \approx 0.14$ +/- 0.02). We won't summarize each group of state-specific trends here, but a full set of parameters estimates is available in the supplementary material.

To check the fit of the model with regard to individual states (see Figure 9 for residual plots across all states), we plotted the observed difference between a given state's support over time among a particular race-sex group and the national average support over time for that race-sex group. Figure 10 shows this comparison for Maryland, the state with the fastest increasing support among black females of all the southern states (the region where there was a lot of variation among state slopes). The raw data show an increasing trend over time, just as the model fit suggests. Figure 10 also shows the shrinkage of the estimated slopes by comparing them to "naive" slopes estimated from the raw differences in percentages between each state's support among a given race-sex group and the national average. The multilevel Bayesian model generally shrinks the estimated slopes toward zero, as it does for Maryland. Here, Maryland is used just as an example – this type of residual plot can and should be used to check fitted trends for all states.

5.4.2 Trends related to age

We model the effect of age on death penalty support as a linear trend over time (on the logistic scale) for each state and age category. Figure 11 shows the estimated differences between each age-state trend and the mean trend for a given state. The blue lines in each of the four plots shows the average trend for each age category (averaged across the 51 states). Here we see that 18-29 year-olds supported the death penalty the least on average (about 3%

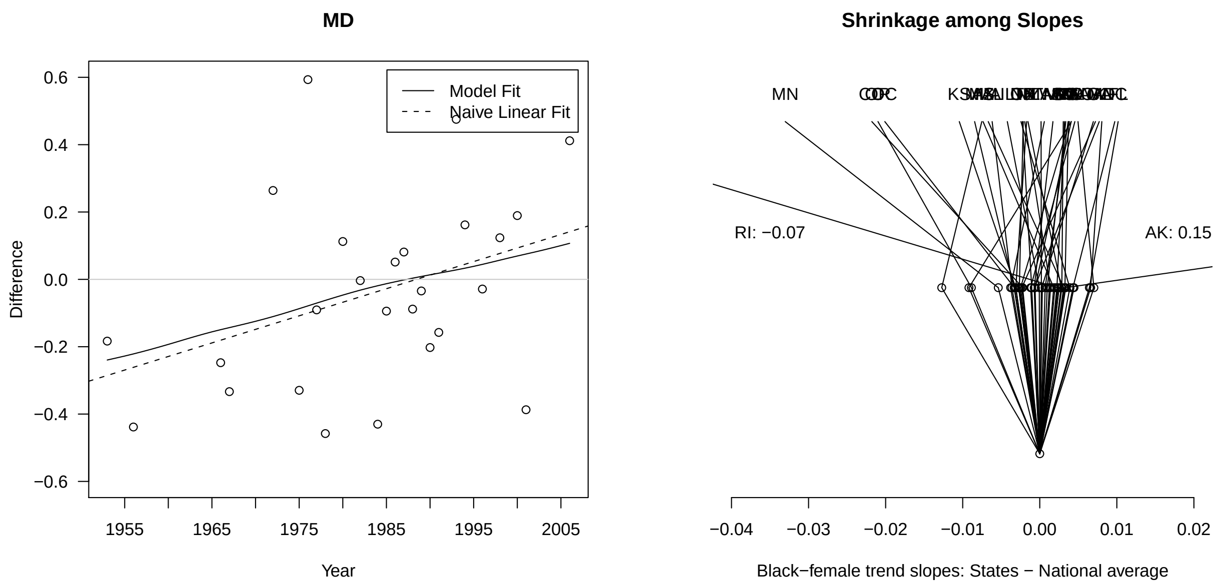

Figure 10: The left plot contains the observed difference between support among black women in Maryland over time and the support of black women across the nation over time. The dotted line is least-squares regression line through the differences in percentages, and the dark line is the difference estimated by the model (for a person of average age and education). The slope and intercept of the model fit are both shrunk slightly toward zero. The right plot shows the shrinkage of estimated slopes like those pictured in the left plot. The top set of points in the plot is the set of 51 estimated slopes from fitting a linear model to the difference in observed percentages of support among black women between a given state the and national average (equivalent to the dotted line in the left plot); the middle set of points are the estimated slopes for each state from the multilevel Bayesian model, and the bottom point is the grand mean.

less than average), and their support has not changed over time. 30-44 year-olds have shown the most support for the death penalty on average (about 1.8% above average), and their support has remained steadily above average for the whole time period. There are trends visible in support among the other two age groups. 45-64 year-olds have shown increasing support for the death penalty over time, increasing their support by about 6% on average over the 54-year time span. Respondents over 65 years old decreased their support by about 5% over the 54-year time span. In general, there is more variation among state slopes for the two elder age categories, as shown by the variation among the gray lines in the bottom two plots of Figure 11.

One thing that Figure 11 does not show is a breakdown of the variation in age trends by region. The largest amount of variation in a group of intercepts is among western states in the 30-44 year-old age category, where $\hat{\sigma}_{\text{age-state}_{(2,4)}} = 0.26$, and the estimated differences between a state's mean (1980) support for this age group and that state's overall mean support ranges from -4% (Idaho) to +6% (Hawaii). There is also a substantial amount of variation among the slopes of 45-64 year-olds in the midwestern states, for example, where $\hat{\tau}_{\text{age-state}_{(3,3)}} = 0.24$, and among the slopes of 65+ year-olds in the western states, where $\hat{\tau}_{\text{age-state}_{(4,4)}} = 0.30$. We don't discuss specific hypotheses regarding trends among age-state cohorts in this paper, but if further investigation were to be done, we suggest plotting the raw data vs. the model fit in a single figure as a tool for further understanding, similar to how Figure 10 displayed the trend in support among one particular race-sex group (black females) in a given state (Maryland).

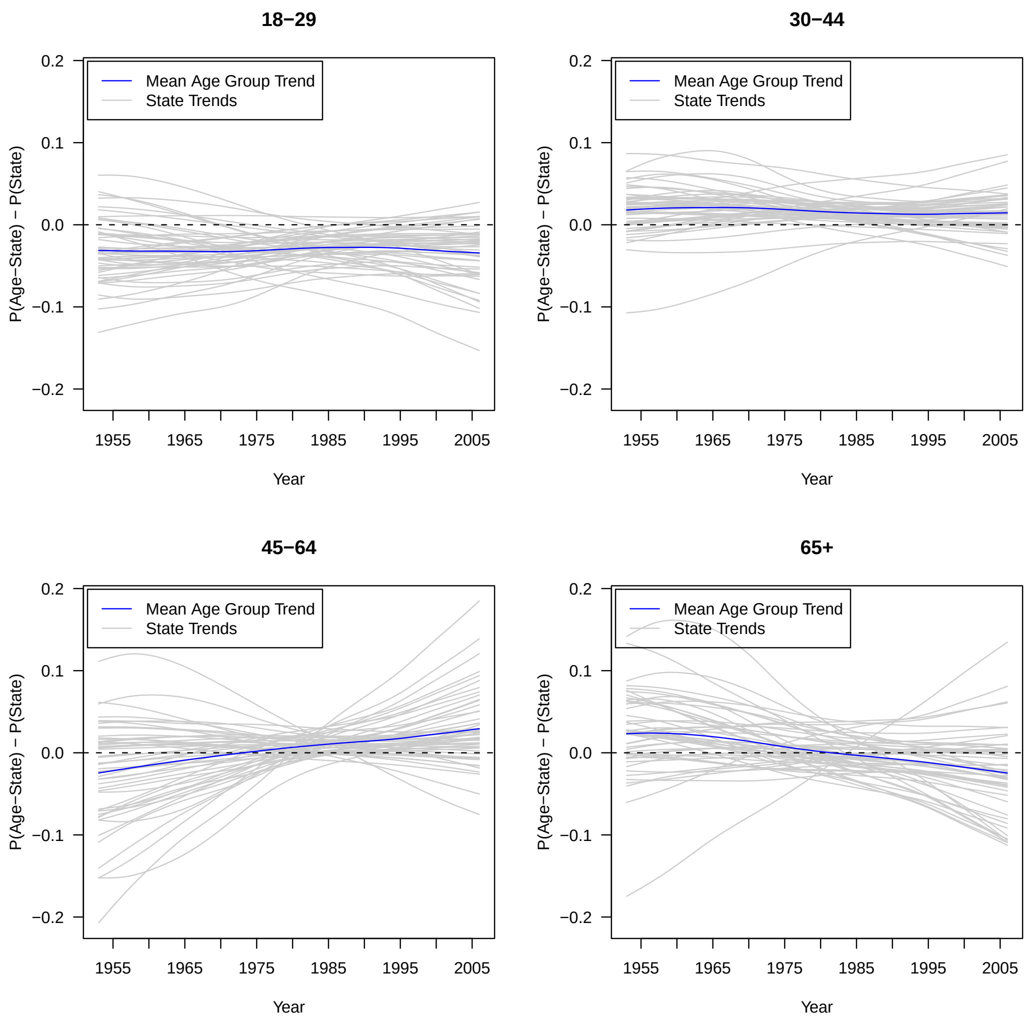

Figure 11: The gray lines show the differences between individual age-state trends and the average trend for each state, and the blue lines show the average difference across all states for each age category.

5.4.3 Coefficients of education

The education level of a respondent is measured by the highest degree they earned, and we model the effects of degree on death penalty support separately for each state. We don't, however, model these effects as a trend. Degree level explains a substantial amount of variation in death penalty support, and its interaction with state of residence is also a strong predictor.

Figure 12 illustrates the differences among states for each degree category with colored maps. The five maps that correspond to the different degree levels illustrate the state-to-state variation in public opinion within each degree category:

- Death penalty support between states varies the most for respondents with less than a high school education, compared to the other educational categories. On average, respondents with this degree level support the death penalty at a level equal to the national average, but in some states (Vermont, New York, and Iowa, for example) respondents in this degree group support the death penalty about 12% more than average, and in other states (Idaho, North Carolina, and Montana), about 12% fewer respondents support the death penalty than the national average. Nevada is the most extreme state—respondents there with less than a high school degree support the death penalty 22% less than the national average. The standard deviation of the differences (by state) for this degree category is about 9%.

- Respondents with a high school degree support the death penalty about 7% more than average—the highest level of support across the degree categories.

- Respondents with a graduate degree support the death penalty at much lower levels than average (about 12% lower than average); they are the degree group that differs from average the most.

- The standard deviations of the differences across states for the highest four categories of degree level are between 3% and 5%.

- We don't find evidence for any time trend for degree categories; residual plots illustrating this are in Section 6.

6 Goodness of fit checks

We include in our analysis a number of plots that display raw data and various estimates to give a sense of the importance we attach to model understanding. With these complex models, it is not enough to simply present results; the serious researcher must also be able to visualize the path from data to inferences. We have already included plots of model estimates vs. observed data in Figures 9 and 10, showing the fit of the model with respect to trends among racial and gender-based groups. In this section, we display graphical checks of the fit of the model with respect to state trends, state-year interaction effects, and trends among educational groups.

Figure 13 illustrates the fit of the model with respect to the difference between the national average level of death penalty support and support in two specific states, Massachusetts and Ohio, over time. For both states, the assumption of a linear trend explaining the differences between state-level opinion and national opinion seems sound. There may be some dependence from year to year for Ohio (especially in the late 1970's, when a group of about 4-5 years were all different from the national average by more than the linear model suggests), but on the whole, the independence assumption for the state-year interaction effects looks realistic. In years where more data is observed, the estimated probabilities do

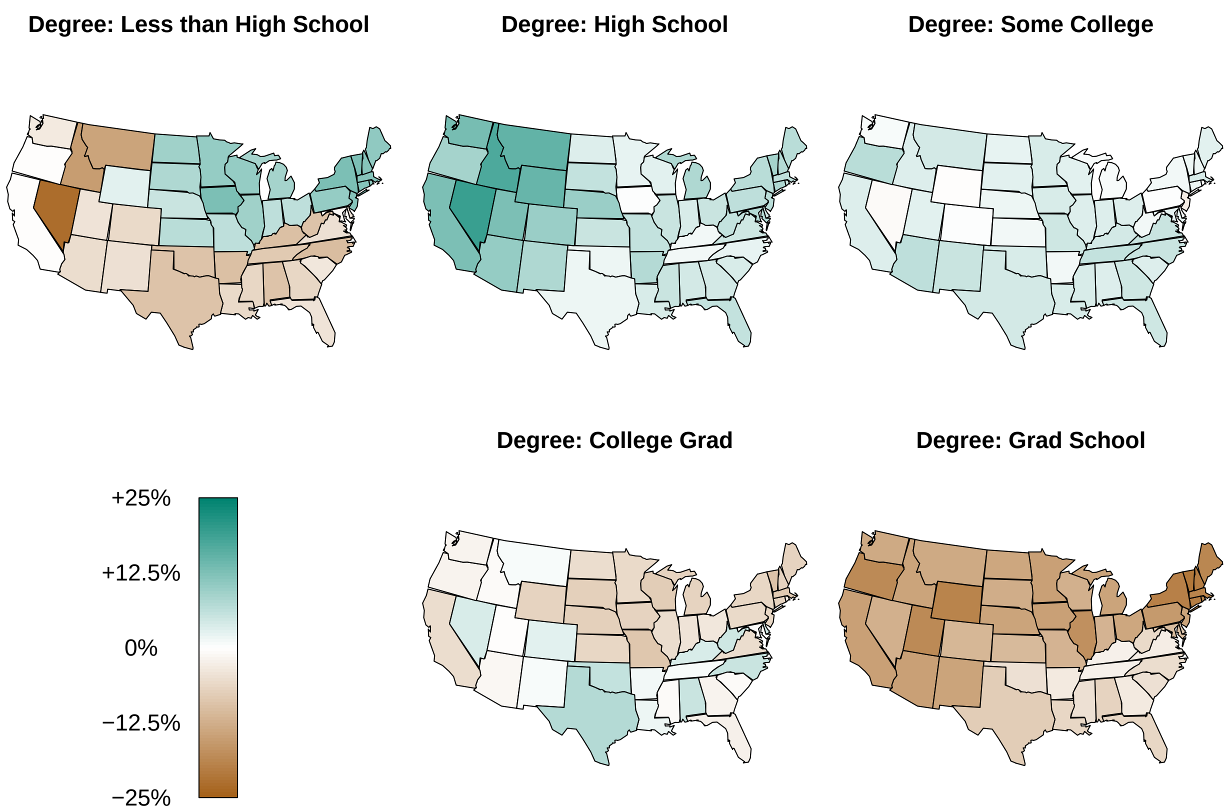

Figure 12: Each state in each map is shaded from brown to blue, where brown indicates a level of support 25% below the national average, and blue indicates a level of support 25% above the national average. White shading indicates support equal to the national average.

not shrink as far toward the state trend line (1965, for example). In years where there is little data, the estimated probabilities are pooled almost all the way to the state trend line (Massachusetts in 2001, for example). Last, when there is no data from a given year, the estimated percentages for a given state lie exactly on the state trend line.

We considered extending the model to allow for the effects of degree to vary across time.

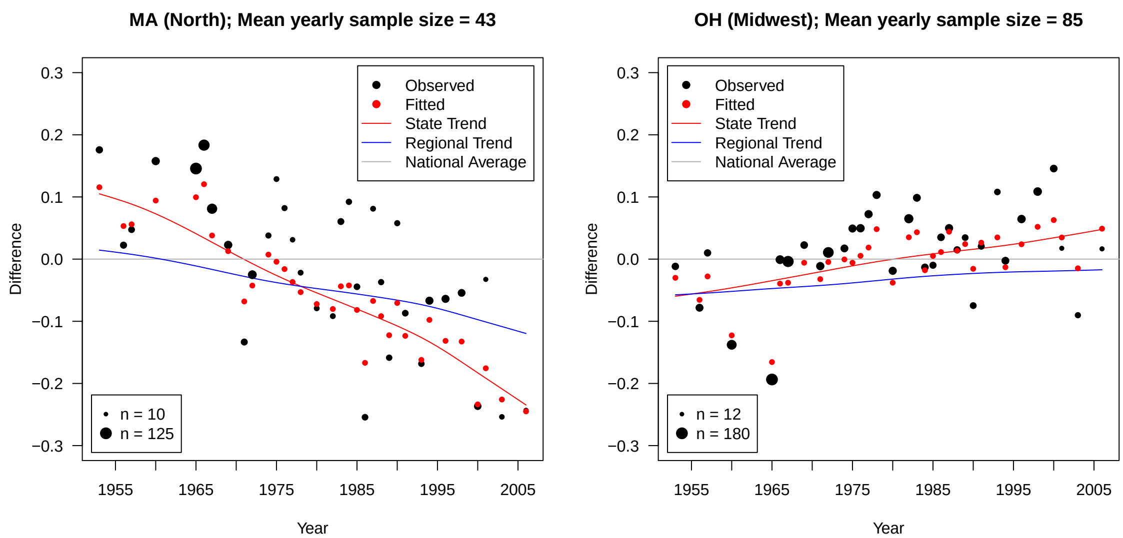

Figure 13: Differences between individual states and the national average over time on the probability scale, with estimates from the fitted model plotted with observed data. The model-based estimates are of the estimated support for an individual of "average" demographics.

To investigate whether this would be likely to improve the fit of the model, we performed a posterior predictive check in which we simulated data for each respondent, and compared the predicted levels of support for each degree level over time, holding the other variables constant at their observed levels. The results are displayed in the five residuals plots in Figure 14. There don't appear to be any patterns across time among the residuals in any of the degree-level groups that would indicate that our model is missing an important time trend. There appear to be influential points at the extremes of the x-axis for the high school degree category and the college degree category, but in each case, the 95% intervals for these residuals still contain zero.

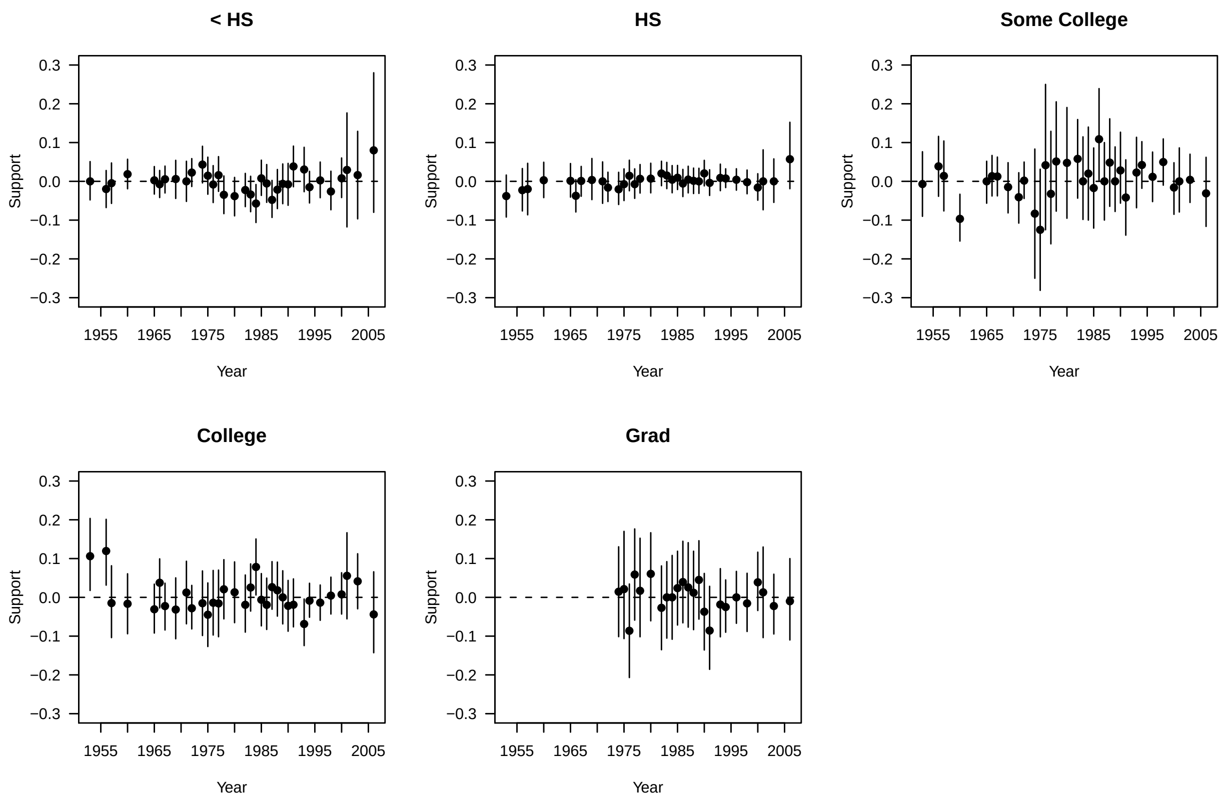

Figure 14: Residual plots of estimated support over time for each education level. The vertical lines are 95% intervals of the estimated differences between the observed levels of support for each degree category in a given year, and the estimated support for that group and year. None of the plots show a pattern across time among the residuals.

7 Out-of-sample predictions

To assess the predictive accuracy of our model and compare it to competing models, we randomly divided our data set into a training set and a test set, containing 80% and 20% of the data, respectively. We then re-fit the model to the training data and made out-of-sample predictions on the test data.

We started with a series of relatively simple models estimated via maximum likelihood using the `glm()` function in R. We fit these models mostly as a form of EDA, so that we could see which interaction terms were likely to improve the predictive accuracy of the larger, multilevel Bayesian models that we would fit later. The simplest model we fit was a model with main effects for each of the six variables (year, state, sex, race, age, and education), where we treated each variable as categorical. We then fit a variety of models (30 in total) that contained two-way interactions and occasionally a three-way interaction. For each model, we recorded three statistics: (1) the average training deviance, (2) the average test deviance, and (3) the number of parameters in the model. Figure 15 displays a summary of these model fits, where the x-axis contains the number of parameters in each model, and the y-axis contains the average deviance for training and test data points, where deviance is defined as -2 times the log-likelihood. Of this set of models, the one that had the lowest average deviance on the test set contained the main effects and eight sets of two-way interaction terms (and zero sets of 3-way interaction terms), for a total of 489 parameters, and an average deviance on the test set of 1.168.

Using the `glm()` function in R to estimate MLE's for the parameters presented a few limitations. First, the program crashed when we tried to include state-year interaction terms, because there were so many of them (thus, none of the 30 models we attempted included these interactions). Second, maximum likelihood estimates do not help for cells in which

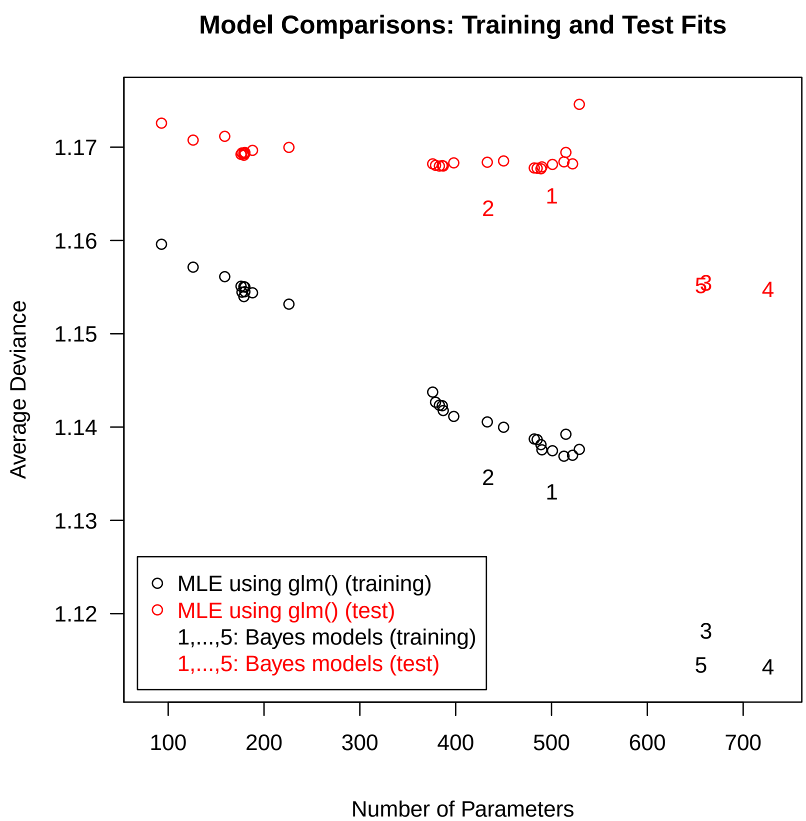

Figure 15: Each point represents a model, and the plot contains the average deviance for each model on the training (black) and test (red) data sets, plotted against the number of parameters in the model. The points denoted by small circles are models fit using maximum likelihood estimation in R (using the glm() function). The five numbers in the plot correspond to multilevel Bayesian models, where the number of parameters plotted (along the x-axis) is the estimate of the "effective" number of parameters, pD (see Spiegelhalter et al. (2002)) for that model.

there is no training data; for example, we didn't get estimates of yearly effects for 20 of the years in the 54-year span of the data set because there were no surveys given in these years. Third, it isn't clear exactly how to incorporate prior information or state-level variables. Last, these models are prone to overfitting when the number of parameters becomes large relative to the amount of data that is observed; some regularization is necessary.

We fit five multilevel Bayesian models to the data, where the fourth model, which we call the "main" model, is the one described in Equation 1 and pictured by the DAG in the supplemental materials. The first two Bayesian models are much simpler than the main model, but generally more complicated than the models fit using `glm()`. The first Bayesian model contained main effects for the four demographic variables and state-year interaction terms, where the state-year interaction terms were centered at the sum of their respective year and state main effects, and each group of main effect parameters had a prior mean of zero and a half-t prior distribution on its standard deviation (the same as described in Section 4). This model, labeled "1" in Figure 15, performed substantially better than the 30 frequentist models. The effective number of parameters, p_D, as estimated by the difference between the posterior mean of the log-likelihood and the log-likelihood evaluated at the posterior mean of the parameters in the likelihood equation (Spiegelhalter et al., 2002), was about 500, which is close to the number of parameters in the best-predicting non-Bayesian model.

The second Bayesian model contained state-year interaction terms, but this time with the exact same prior structure as the main model—that is, this model included state trends, state-level variables, and an AR(1) prior distribution (with a linear component) for the yearly effects. The rest of the demographic main effects were given the same structure and priors as in the first Bayesian model—means of zero, and half-t prior distributions on their group-level standard deviations. This model had fewer effective parameters than the first one (about 434 compared to 500) because of the partial pooling induced by the structured prior on the state-year interaction terms, and the inclusion of state-level variables (which help to induce even more pooling on the state-specific slopes and intercepts, decreasing the estimate of pD). This model fit the training data worse than the first Bayesian model, but made better predictions on the test set.

Ther third and fifth Bayesian models were similar to the main model, with only minor modifications. The third Bayesian model omitted the age-state slopes, and was otherwise identical to the main model. This model performed slightly worse on the test set than the main model, and substantially worse on the training set. The fifth Bayesian model included all of the parameters in the main model, and also included four additional sets of two-way interactions: (sex, age), (sex, degree), (race, age), and (race, degree). The inclusion of each of these two-way interactions improved the predictive performance of the models fit using `glm()`, which is why we also included them in a Bayesian model. The result of their inclusion in a Bayesian model, though, was a poorer fit than the main model on both the training and test data sets. Ultimately, the main model (the fourth Bayesian model) performed the best on the training set, with an average deviance of about 1.155.

8 Discussion

We fit a multilevel Bayesian model to 58,253 individual responses to the question "Are you in favor of the death penalty for persons convicted of murder?" using data from a 54-year time span and including demographic and state-level variables. The use of a structured prior distribution on the yearly effects allowed us to simultaneously estimate their variation, while also estimating various main effects, trends, and interaction effects that shed new light on certain relationships between demographic variables and death penalty support. We found that blacks have decreased their support over time dramatically compared to the general population, with the support among black men decreasing slightly faster, on average, than among black women. This pattern was previously difficult to identify because of yearly fluctuations in overall death penalty support. We also found substantial variation in trends of support between states, and especially between states within certain classes of education level, where the least-educated respondents showed the most state-to-state variation in support. We also found that death penalty support has grown faster in states where support for Republican presidential candidates has grown during the past 50 years, and average support over time has been higher in states where the death penalty has been legal for a larger proportion of the past 50 years.

One useful visual summary of the model we fit is given in Figure 16. It displays average predictive comparisons (Gelman and Hill, 2007) for each of the six main effects. These are estimates of the amount of variation, on the percentage scale, that is the result of differences in each of the six main variables. To compute the average predictive comparison interval for the variable Year, for example, we did the following for each posterior sample: For year $t = 1, ..., T$ plug in year t for the survey year for each of the 58,253 respondents in the data set, and compute the average level of estimated death penalty support across all respondents. This results in T numbers, each of which is the average of 58,253 estimated proportions; next, compute the standard deviation of these T average proportions. This is the estimated amount of variation due to the variable Year, averaged across all respondents, holding the other variables constant at their observed values for each individual data point, for this iteration of the MCMC algorithm. We did this for all 3,000 posterior samples, for all six main variables, and plotted the resulting 95% intervals in Figure 16. Our conclusion is that race explains the most variation in death penalty support, and it accounts for a difference of about 25% in support for the death penalty. The next two variables that explain the most variation are state and year, and each account for about a 12% difference in support levels. Then, sex, education, and age account for about 10%, 7.5%, and 3% of variation in death penalty support. In the case of all these variables, the 95% intervals have a width of only about 2%. The resulting comparisons are easy to interpret. For example, you would expect a larger difference in the probability of death penalty support by picking two individuals from different years than you would from by picking two individuals of different sexes (or different degrees earned, or age groups, for that matter).

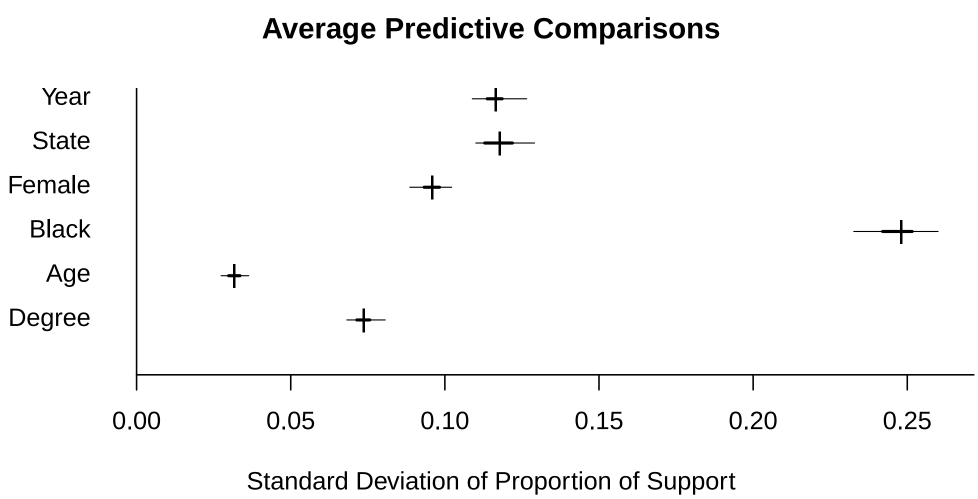

Figure 16: Average predictive comparisons for the main effects of the six variables: year, state, sex, race, age, and education. These are estimates of the amount of variation in death penalty support, on the percentage scale, attributable to each of these variables, holding the other variables constant at their observed levels.

Future work on modeling death penalty public opinion could go in a number of directions. One could include additional survey data and, if necessary, model differences in data collected by different organizations. One could also attempt to gather more demographic variables for each respondent, such as political party affiliation, income, or religion, to better understand the factors that predict attitudes on this issue. It could also make sense to consider nonparametric regression methods such as BART (Chipman et al., 2010), perhaps in combination with existing regression models to pull out any structure in the data beyond what is captured by our logistic regression.

To return to the political questions that motivated this work: public opinion is, presumably, both a cause and a consequence of policies on capital punishment. In order to study these connections, researchers need measures of state-level opinion. In a study of state-level attitudes on gay rights, Lax and Phillips (2009a) showed a level of responsiveness to opinion that was surprising given some of the earlier literature on state politics. The present research goes further by modeling trends at the state level using sparse data.

9 Supplementary Materials

1. DAG.pdf: A directed acylclic graph (DAG) of the full model

2. Adjusted-parameters.pdf: A description of the calculations to derive the adjusted parameters.

3. estimates.RData: Interval estimates of the level of support for the death penalty for each possible combination of (state, year, race, sex, degree, age).

4. varnames.RData: Variable names that correspond to the dimensions of the array in the file estimates.RData.

References

Baumgartner, F. R., De Boef, S. L., and Boydstun, A. E. (2008), *The Decline of the Death Penalty and the Discovery of Innocence*, Cambridge University Press.

Chipman, H. A., George, E. I., and McCulloch, R. E. (2010), "BART: Bayesian Additive Regression Trees," *Annals of Applied Statistics*, 4, 266–298.

Ellsworth, P. C. and Gross, S. R. (1994), "Hardening of the attitudes: Americans' views on the death penalty," *Journal of Social Issues*, 50, 19–52.

Erikson, R. (1976), "The relationship between public opinion and state policy: A new look based on some forgotten data," *American Journal of Political Science*, 20, 25–36.

Erskine, H. (1970), "The polls: Capital Punishment," *Public Opinion Quarterly*, 34, 290–307.

Fox, J. A., Radelet, M. L., and Bonsteel, J. L. (1991), "Death penalty opinion in the post-Furman years," *New York University Review of Law and Social Change*, 28, 499–528.

Franklin, C. (2001), "Pre-Election Polls in Nation and State: A Dynamic Bayesian Hierarchical Model," Tech. rep., University of Wisconsin.

Gelfand, A. E. and Sahu, S. K. (1999), "Identifiability, improper priors, and Gibbs sampling for generalized linear models," *Journal of the American Statistical Association*, 94, 247–253.

Gelman, A. and Hill, J. (2007), *Data Analysis Using Regression and Multilevel/Hierarchical Models*, Cambridge University Press.

Gelman, A. and Rubin, D. (1992), "Inference from Iterative Simulation Using Multiple Sequences," *Statistical Science*, 7, 457–472.

Hanley, J. (2008), *Public Opinion and Constitutional Controversy*, Oxford University Press, chap. The Death Penalty, pp. 108–138.

Lax, J. R. and Phillips, J. H. (2009a), "Gay rights in the states: Public opinion and policy responsiveness," *American Political Science Review*, 103, 367–386.

— (2009b), "How should we estimate public opinion in the states?" *American Journal of Political Science*, 53, 107–121.

Norrander, B. (2000), "The multi-layered impact of public opinion on capital punishment implementation in the American states," *Political Research Quarterly*, 53, 771–793.

Plummer, M. (2003), "JAGS: A Program for Analysis of Bayesian Graphical Models Using Gibbs Sampling," in *Proceedings of the 3rd International Workshop on Distributed Statistical Computing.*

Schuman, H. and Presser, S. (1981), *Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording, and Context*, Sage Publications, Inc.

Spiegelhalter, D. J., Best, N. G., P., C. B., and Van der Linde, A. (2002), "Bayesian measures of model complexity and fit (with discussion)," *Journal of the Royal Statistical Society, Series B*, 64, 583–616.

Vidmar, N. and Ellsworth, P. (1974), "Public opinion and the death penalty," *Stanford Law Review*, 26, 1245–1270.