How Bayesian analysis cracked the red-state, blue state problem*

Andrew Gelman

29 May 2012

1. Introduction

Income and economic redistribution are central to electoral politics. In the United States as in other countries, political and economic divisions cut along geographic and demographic lines. Richer *people* are more likely to vote for Republican candidates while poorer voters lean Democratic; this is consistent with the positions of the two parties on economic issues. At the same time, richer *states* on the coasts are bastions of the Democrats, while most of the generally lower-income areas in the middle of the country strongly support Republicans. This geographic pattern is consistent with the sense of a culture war between richer, more socially liberal cosmopolitans and middle-class proponents of traditional American values.

Thus, a statistical pattern of voting patterns at the individual and group level is central to political debates about economic and social polarization. During a research project lasting several years, we resolved the statistical questions by fitting a series of multilevel models to study the differences in voting between rich and poor voters, and rich and poor states. We were using national survey data with relatively small samples in some states, ethnic groups, and income categories; this motivated the use of Bayesian inference to partially pool between fitted models and local data.

Previous, non-Bayesian analyses of income and voting had failed to connect individual and state-level patterns. Typical analyses would be either at the individual or the aggregate levels but not both. In the studies that did model voting based on individual and goegraphic characteristics, the focus was on estimating some particular regression coefficient (or, more generally, on identification of some average causal effect). Classical statistics tends to focus on estimation or testing for a single parameter or low-dimensional vector, whereas Bayesian methods work particularly well when the goal is inference about a large number of uncertain quantities (in this case, coefficients within each of the fifty states).

Now that our analysis has been done done, we believe it could be replicated using non-Bayesian methods. However, one can also view our fitting of a series of models as a form of exploratory data analysis. It is only through active engagement with the data that we got a sense what to look for. Thus, the flexibile generality of the Bayesian approach facilitated our substantive research breakthrough here.

This is the opposite of the paradigm common in classical theoretical statistics, of laser-like focus on identification of a single effect and a concern with frequency properties of a prechosen statistical procedure.

From a statistical perspective, we are confident that our Bayesian procedure has worked well because our inferences make sense, are consistent with the data, and have performed well in external validation (in that we developed our models to fit to the 2000 election and then they successfully worked for 2004 and 2008).

The real world impact of this work is twofold. First, we have established that income is more strongly predictive of Republican voting in poor states than in rich states, and that this difference has arisen in the past two decades. Second, political scientists and journalists now have a clearer view of the relation between social, economic, and political polarization. The political differences

*For *Statistical Science*. We thank Yair Ghitza for helpful comments, Christian Robert and Kerrie Mengersen for organizing this special issue, and the National Science Foundation and Institute of Education Sciences for partial support of this research.

between "red America" and "blue America" are concentrated among the upper half of the income distribution. By allowing us to model a pattern of income and voting that varies across states, Bayesian analysis allowed us to get a grip on this important political trend.

2. Background

For the past fifteen years or so, Americans have been divided poliitcally into "red states" (the conservative, Republican-leaning areas in the south and middle of the country) and "blue states" (the more urban areas in the northeast and west coast, whose residents consistently vote for Democrats). Here is the red-state, blue-state paradox: since the 1990s, the poorer states have vote for conservative Republicans while rich states favor liberal Democrats. This has surprised political observers, given that Republicans are traditionally the rich with the Democrats representing the poor. And, indeed, Republican candidates do about 20 percentage points better among rich voters than among poor voters, a gap that has persisted for decades.

The red-state blue-state pattern became widely apparent in the aftermath of the disputed 2000 presidential election, when television viewers became all too familiar with the iconic electoral map: blue states on the coasts and upper midwest voting for Al Gore, red states in the American heartland supporting George Bush, and Florida colored blank awaiting the decision of the courts.

The result has confused political observers on both sides of the political spectrum. On the right came a much-discussed magazine article by David Brooks, comparing Montgomery County, Maryland, the liberal, upper-middle-class suburb where he and his friends live, to rural, conservative Franklin County, Pennsylvania, a short drive away but distant in attitudes and values, with "no Starbucks, no Pottery Barn, no Borders or Barnes & Noble," plenty of churches but not so many Thai restaurants, "a lot fewer sun-dried-tomato concoctions on restaurant menus and a lot more meatloaf platters." On the left, Thomas Frank's bestselling *What's the Matter with Kansas* (2004) was widely interpreted to answer the question of why low-income Americans vote Republican: "For more than thirty-five years, American politics has followed a populist pattern . . . the average American, humble, long- suffering, working hard, and paying his taxes; and the liberal elite, the know-it-alls of Manhattan and Malibu, sipping their lattes as they lord it over the peasantry with their fancy college degrees and their friends in the judiciary."

Here is a summary from Gelman (2011):

> Republicans, who traditionally represented Americas elites, had dominated in lower-income areas in the South and Midwest and in unassuming suburbs, rather than in Americas glittering centers of power. What could explain this turnaround? The most direct story—hinted at by Brooks in his articles and books on Americas new, cosmopolitan, liberal upper class—is that the parties simply switched, with the new-look Democrats representing hedge-fund billionaires, college professors, and other urban liberals, and Republicans getting the votes of middle-class middle Americans. This story of partisan reversal has received some attention from pundits. For example, TV talk show host Tucker Carlson said, "Okay, but heres the fact that nobody ever, ever mentions—Democrats win rich people. Over $100,000 in income, you are likely more than not to vote for Democrats. People never point that out. Rich people vote liberal." And Michael Barone, the editor of the Almanac of American Politics, wrote that the Democratic Party "does not run very well among the common people." But Tucker Carlson and Michael Barone were both wrong . . . obviously wrong, from the standpoint of any political scientist who knows opinion polls. Republican candidates consistently do best among upper-income voters and worst at the low end. In the country as a whole and

separately among Whites, Blacks, Hispanics, and others, richer Americans are more likely to vote Republican. . . .

Misconceptions about income and voting are all over the place in the serious popular press. For example, James Ledbetter in *Slate* claimed that "Americas rich now tilt politically left in their opinions." In the *London Review of Books*, political theorist David Runciman wrote, "It is striking that the people who most dislike the whole idea of healthcare reform—the ones who think it is socialist, godless, a step on the road to a police state—are often the ones it seems designed to help. . . . Right-wing politics has become a vehicle for channeling this popular anger against intellectual snobs. The result is that many of Americas poorest citizens have a deep emotional attachment to a party that serves the interests of its richest." No, no, and no. An analysis of opinion polls [Gelman, Lee, and Ghitza, 2010] finds, unsurprisingly, but in contradiction to the above claims, that older and high-income voters are the groups that most strongly oppose health care reform.

It has been difficult for poitical journalists to accept that richer voters prefer Republicans while richer states lean Democratic. At first this may appear to be a simple example of Simpson's paradox: the correlation of income with Republican voting voting is negative at the aggregate level and positive at the individual level.

But from a political perspective, the story is more complicated. The United States has a federal system of government, with some policies determined nationally and others at the state level. When considering campaigns, however, voters' opinions are paramount, and here it is a central fact of American politics that richer voters lean Republican.

3. Statistical model and Bayesian inference: overview

It turns out that the statistical story is more complicated too. Red state and blue states do not only differ in their political complexions; in addition, the relation between income and voting varies systematically by state. In richer, liberal states such as New York and California, there is essentially no correlation between income and voting—rich and poor vote the same way—while in conservative states such as Texas, the rich are much more Republcian than the poor. Politcal divisions by social class look different in red and blue America.

This key statistical part of our analysis is the estimation of the relation between income and voting (later including religious attendance and ethnicity as additional explanatory factors) separately in each state. This is difficult because even a large national survey will not have a huge sample size in all fifty state—and recall that we are not merely estimating an average in each state but we are attempting to estimate a regression or even a nonlinear functional relationship. Political scientists armed with conventional statistical tools sometimes try to get around this sample size problem by pooling data from multiple years—but this would not work here because we are interested in changes over time.

The Bayesian resolution was a multilevel model allowing different patterns of income and voting in different states. The model was built on a hierarchical logistic regression but included error terms at every level so that the ultimate fit was nonparametric. Because of the complexity of our model, it was necessary to check its fit by comparing data to posterior simulations. Classical approaches—even classical multilevel models—would not fully express the uncertainty in the fit. In contrast, our Bayesian approach not only allowed us to fit the data; it also provided a structure for us to consider a series of dfferent models to explore the data.

In general, estimating state-level patterns from national polls requires two tasks: *survey weighting* or adjustment for known differences between sample and population (for example, surveys tend to overrepresent women, whites, and older Americans, while underrepresenting young male ethnic minorities), and *small-area estimation* or regularized estimates for subsets where raw-data averages would be too noisy.

In order to estimate the pattern of income and voting within each state, we used the strategy of *multilevel regression and poststratification* (MRP), a general approach to survey inference for subsets of the population that has two steps:

1. Use a multilevel model to estimate the distribution of the outcome of interest (in this case, vote preference, among those people who plan to vote in the presidential election) given demographic and geographic predictors which divide the population into categories. Here we start with 250 cells (5 income categories within each of 50 states); a later model considers four ethnicity categories as well, and more generally the analysis could categorize people by sex, age, income, marital status, religion, religious attendance, and so on, easily leading to more cells than survey respondents. It is the job of the Bayesian model to come up with a reasonable inference for the joint distribution of the Republican vote share within whatever categories are included.

2. Poststratify to sum the inferences across cells. For example, the estimated percentage of support for Obama among Hispanics in the midwest is simply the weighted average of his estimated support within each of the relevant poststratification cells (in the ethnicity/income/state model, this would be one cell for each income category within each midwestern state). The weights in this weighted average are simply the number of voters in each cell, which we can get from the U.S. Census. (To obtain voter weights is itself a two-stage process in which we first take the number of adult Americans in each cell, then multiply within each cell by the proportion of adults who voted, as estimated from a multilevel logistic regression fit to a Census post-election survey that asks about voting behavior; again, see Ghitza and Gelman, 2012, for details.)

(In sampling jargon, *strata* are defined based on the design of the survey—a stratified design has separate sampling within each stratum—whereas *post-strata* are chosen based on the analysis. This is why our method is MRP and not MRS.)

It is clear how MRP fits in with Bayesian statistics: the number of observations per cell is small, so our problem is one of small-area estimation (Fay and Herriot, 1979), hence it makes sense to partially pool inferences, averaging local data and a larger fitted regression model. Bayesian inference is a well-recognized tool for combining local information with predictions from a stochastic model (Clayton and Kaldor, 1987).

But it may be less obvious how our method connects with the vast literature on survey weighting, a field traditionally draws a strong distinction between "model-based" procedures such as Bayesian or even likelihood methods that posit a probability model for the data and "design-based" inference which leave data unmodeled and apply a probability distirbution only to the sampling process. The connection was made clear by Little (1991, 1993), who showed how model-based inference fits in a larger design-based framework (or, conversely, how design-based inferences are possible within a larger probability model). Little's key insight is centered on the *poststratification identity*:

$$\theta = \frac{\sum_j N_j \theta_j}{\sum_j N_j},$$

where θ is some aggregate quantity of interest (for example, the estimated support for Obama among Hispanics in the midwest), j's are the cells within this aggregate, N_j is the population size of each cell (in our case, obtained from the Census), and θ_j is the (unknown) population quantity with the cell.

As noted, the above equation is an identitiy, that is, a tautology. Its connection to statistical inference comes in the inferences for the θ_j's. Assuming simple random sampling within cells (the implied basis for classical survey weighting), one can estimate the θ_j's through simple raw cell means (statistically inefficient if sample sizes are less than huge) or more effectively via regression modeling which quickly leads to Bayes if the number of cells is large and the model is realistically complex. The inforomation that would go into classical survey weights instead enters our MRP calculations through the population sizes N_j. This is important: you can't get something for nothing, and the price we pay for our poststratified lunch is the array of population numbers N_j.

MRP combines long-existing ideas in sample surveys but has become recently popular as a way to learn about state-level opinions from national polls (Gelman and Little, 1997, Lax and Phillips, 2008, 2009), perhaps as a result of increasing ease of handling large datasets as well as improvements in off-the-shelf hierarchical modeling tools. In many political science applications, state averages are of primary interest, and we estimaie opinion in within-state slices (for example, white women aged 30–44 in Missouri) only because we feel we need to, in order to adjust for differential nonresponse. We fit the multilevel model to get reasonable inferences within all these cells but then immediately poststratify to get state-level estimates. All these steps are needed—a simple Bayesian analysis of state-level data would fail to adjust for known demographic differences between sample and population. This is the sense that MRP forms a bridge between Bayesian inference (so flexible and powerful for estimating large numbers of parameters and making large numbers of uncertain predictions at once) and classical survey adjustment (given that real surveys can be clearly nonrepresentative of the population). This latter step is crucial in many applications in which data are combined from many disparate surveys.

In the Red State Blue State project, MRP plays a slightly different role. Here we actually are interested in categories within a state (initially, the five income categories; later, voters cross-classified by income, education, and religious attendance). The poststratification is less important here (although it does come up: after we sum our inferences over cells within each state, we adjust our predictions of state-level averages to line up with actual recorded vote totals, a completely reasonable step given this additional information separate from the survey data). What is relevant for the present discussion is that our method harnesses the power of Bayes within a framework that accounts for concerns specific to survey sampling.

4. Statistical model and Bayesian inference: details

We fit our models separately to pre-election poll data from 2000, 2004, and 2008, with about 20–40,000 respondents in each year. This sample size is large enough for us to estimate variation among states but not so large that we could just estimate each state's pattern using its own data alone.

For the purposes of learning about opinion from a sample, the multilevel model is a way to obtain estimates for mutually exclusive slices of the population (and implicltly corresponds to the assumption that the respondents being analyzed are a simple random sample within each cell). From the perspective of statistical inference, however, our model is simply a hierarchical regression with discrete predictors. Thus, if we want to perform inference for 4 ethnicities $\times$ 5 income categories $\times$ 50 states, we just need to include predictors for ethnicities, income levels, and states (along with various interactions), and perform inferences for the vector of regression coefficients, and inferences for the 1000 cells just pop out as predictions from the fitted regression model.

The most basic form of the model is a varying-intercept logistic regression of survey responses:

$$\Pr(y_i = 1) = \text{logit}^{-1}(\alpha_{j[i]} + X_i\beta),$$

where:

- $y_i = 1$ if respondent i intends to vote for the Republican candidate for president or 0 if he or she supports the Democrat (with those expressing no opinion excluded from the analysis),
- $\alpha_{j[i]}$ is a varying intercept for the state $j[i]$ where the respondent lives (that is, $j[i]$ is an index taking on a value between 1 and 50),
- X_i is a vector of demographic predictors (indicators for state, age, ethnicity, education, and some of their interactions, and also income, discretized on a scale of $-2, -1, 0, 1, 2$), and β is a vector of estimated coefficients.

The intercepts α_j are themselves modeled by a regression:

$$\alpha_j \sim \text{N}(W_j\gamma, \sigma_\alpha^2),$$

where:

- W_j is a vector of state-level predictors (including average income of the residents of the state, Republican vote share in the previous presidential election, and indicators for region of the country),
- γ is a vector of state-level coefficients, and
- σ_α is the standard deviation of the unexplained state-level variance.

We completed the Bayesian model by assigning to the otherwise-unmodeled parameters $\beta, \gamma, \sigma_\alpha$ a uniform prior distributions: in retrospect not the best choice (we do in fact have prior information on these quantities, starting with results from the model fit to earlier elections) but enough to give us reasonable results. As this work goes forward we plan to think harder about hyperprior distributions and additional levels of the hierarchy such as building in time-series models.

The varying-intercept model above fails because it assumes a constant relation between income and voting across states. Acutally, the data show that income is much more highly correlated with Republican voting in some states than others. We fit this pattern using a model in which the intercept and the coefficient for individual income varies by state. The two varying coefficients within each state are then themselves are modeled given state-level predictors and with a 2×2 covariance matrix for the state-level errors. (We coded income as -2 to 2 rather than 1–5 so that the joint distribution of intercept and slope would be easier to model, following standard practice in regressions with interactions.) This new model fit reasonably well but we further elaborated it by adding varying coefficients for each income category, thus allowing a nonlinear relation (on the logistic scale) of income and vote preference that could vary by state and ethnicity. Income is included in this regression in three ways at once, but because of the hierarchical Bayesian model there is no multicolliearity problem.

Other versions of the model include additional individual-level predictors such as age, education, and religious attendance. For some polls that are "self-weighting" or approximately so—this refers to surveys where adjustments are made within the sampling process to minimize demographic differences between sample and population—we also sometimes fit models with *fewer* individual predictors. Ideally it makes sense to include important predictors such as sex, age, and ethnicity to improve statistical efficiency, even when they are not needed to correct for sampling biases, but for simplicity in computation and analysis we have fit models including only income as a respondent-level variable.

The different pieces of the Bayesian predictive model for vote preferences connect in different ways to our statistical and substantive goals. Adjustments for sex, age, ethnicity, and education correspond to survey weightng for these variables to correct for important known differences between sample and population. Including individual income as a predictor serves the goal of comparing the votes of rich and poor within states, while including state income as a group-level predictor allows us to compare rich and poor states. Finally, the varying-intercept model for state with its error term allows unexplained variation among states, which is crucial because we know that states vary in many other ways beyond that predicted by state income levels

The poststratification step points to a difference between our Bayesian solution and traditional statistical analyses. Even our basic model had many parameters but none of them mapped directly to our summaries of interest. To obtain the relation between income and voting within a state, we did not look at the coefficient for the income predictor. Rather, we used our model to estimate opinion in each poststratification cell and then summed up to infer about each income category within each state. Similarly, we compare rich and poor states not by focusing on the coefficient of state income in the group-level regression but by using MRP to estimate the slope in each of the 50 states and then plotting the estimates vs. state income. The individual and state-level income coefficients are relevant to the model, but our ultimate inferences are constructed from pieced-together predictions. This sort of simulation-based inference may seem awkward to classically-trained statisticians but its flexibility makes it ideal for problems in political science where we are interested in studying variation rather than in estimating some sort of universal constant such as the speed of light. In addition, simulation-based estimates can be directly and easily expressed on the probability scale; there is no need to try to interpret log-odds or logistic regression coefficients.

5. Gains from Bayes

Income and voting had been studied by political scientists for decades, but it was only through Bayesian methods that we were able to discover the different patterns of income and voting in rich and poor states, an important and exciting pattern that had never been noticed before. (At a technical level, our approach also accounted for the design of the survey data by adjusting for demographic factors that were used in survey weighting.)

Often the key to a statistical method is not what it does with the data but, rather, what data it allows one to use. MRP combines design-based and model-based inference and can handle data from multiple surveys as well as census totals on demographics. As always, Bayesian inference works well with models with large numbers of parameters, allowing adjustment for many factors, which is another way of including more information in the inferential procedure.

That said, we believe that an analysis just as good as ours could be constructed entirely using non-Bayesian methods. It would require a lot of extra work (for us) but it should be possible. In fact, many of the patterns we discovered (most notably, that income predicts Republican voting better in rich states than in poor states, and that religious attendance predicts Republican voting better among rich than poor voters) appear directly in the raw data—if you know to look for them. In that sense, multilevel Bayesian modeling (adapted to the sample survey context using poststratification) can be considered as an elaborate form of exploratory data analysis, giving us the chance to see patterns of complex interactions that are in the data but would not appear in simple regression models.

The key pieces in the Bayesian inference were: (a) weighted averages for small-area estimation; (b) poststratification, which detached the modeling stage of the analysis from the inferences for quantities of interest; (c) state-level predictors, which gave us reasonable estimates even for small states; (d) individual-level income included as a continuous and discrete variable at the same time, allowing a nonparametric form for the income-voting relation but partially pooling to linearity; (e) and flexibility in modeling, letting us see the data and examine the problem from many different angles without the burden of requiring a fully-specified model. In standard statistical thoery—Bayesian or otherwise—a model is either already built or is one of some discrete class of candidate models. In this sort of applied exploration, however, the model is always evolving, and it is helpful to have a statistical and computational framework in which we can explore different possibilities.

6. Moving forward

Our book that built upon the analysis described above has changed how journalists and political professionals think about the social and political bases of support for America's two major political parties. On a more methodological level, MRP is being used in a variety of settings to understand local attitudes and to integrate demographic and geographic modeling in social science. Many statistical challenges remain, most notably how to build and compute models with many predictive factors (age, ethnicity, education, family structure, . . .) and correspondingly huge numbers of interactions, how to visualize such model fits, and how to poststratify on characteristics such as religious attendance that are not known in the population. More generally, our increasing ability to fit large statistical models puts more of a burden on checking and understanding these models. Given that a mere two-way model of income and state turned out to be complicated enough to require a multi-year research project, we anticipate new challenges in digesting larger models that allow more accurate inferences from sample to population.

7. References

Clayton, D. G., and Kaldor, J. M. (1987). Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics* **43**, 671–682.

Fay, R. E., and Herriot, R. A. (1979). Estimates of income for small places: an application of James-Stein procedures to census data. *Journal of the American Statistical Association* **74**, 269–277.

Gelman, A. (2011). Economic divisions and political polarization in red and blue America. *Pathways* (Summer), 3–6.

Gelman, A., Lee, D., and Ghitza, Y. (2010). Public opinion on health care reform. *The Forum* **8** (1), article 8.

Gelman, A., and Little, T. C. (1997). Poststratification into many categories using hierarchical logistic regression. *Survey Methodology* **23**, 127–135.

Gelman, A., Park, D., Shor, B., and Cortina, J. (2009). *Red State, Blue State, Rich State, Poor State: Why Americans Vote the Way They Do*, second edition. Princeton University Press.

Gelman, A., Shor, B., Bafumi, J., and Park, D. (2008). Rich state, poor state, red state, blue state: What's the matter with Connecticut? *Quarterly Journal of Political Science* **2**, 345–367.

Ghitza, Y., and Gelman, A. (2011). Deep interactions with MRP: Presidential turnout and voting patterns among small electoral subgroups. Technical report, Department of Political Science, Columbia University.

Lax, J. and Phillips, J. (2008). How should we estimate public opinion in the states? *American Journal of Political Science* **53**.

Lax, J. and Phillips, J. (2009). Gay rights in the states: public opinion and policy responsiveness. *American Political Science Review*.

Lindley, D. V., and Smith, A. F. M. (1972). Bayes estimates for the linear model. *Journal of the Royal Statistical Society B* **34**, 1–41.

Little, R. J. A. (1991). Inference with survey weights. *Journal of Official Statistics* **7**, 405-424.

Little, R. J. A. (1993). Post-stratification: A modelers perspective. *Journal of the American Statistical Association* **88**, 1001-1012.

Louis, T. A. (1984). Estimating a population of parameter values using Bayes and empirical Bayes methods. *Journal of the American Statistical Association* **78**, 393–398.

Snijders, T. A. B., and Bosker, R. J. (1999). *Multilevel Analysis*. London: Sage.