

Deep Interactions with MRP: Election Turnout and Voting Patterns Among Small Electoral Subgroups

Yair Ghitza Columbia University
Andrew Gelman Columbia University

Using multilevel regression and poststratification (MRP), we estimate voter turnout and vote choice within deeply interacted subgroups: subsets of the population that are defined by multiple demographic and geographic characteristics. This article lays out the models and statistical procedures we use, along with the steps required to fit the model for the 2004 and 2008 presidential elections. Though MRP is an increasingly popular method, we improve upon it in numerous ways: deeper levels of covariate interaction, allowing for nonlinearity and nonmonotonicity, accounting for unequal inclusion probabilities that are conveyed in survey weights, postestimation adjustments to turnout and voting levels, and informative multidimensional graphical displays as a form of model checking. We use a series of examples to demonstrate the flexibility of our method, including an illustration of turnout and vote choice as subgroups become increasingly detailed, and an analysis of both vote choice changes and turnout changes from 2004 to 2008.

The introduction of survey methods into the social sciences in the 1940s preceded an explosion of scholarly work explaining the voting behavior of the American public. This work has touched various behavioral topics: political participation (Brady, Schlozman, and Verba 1995; Downs 1957; Gerber and Green 2000; Hansen and Rosenstone 1993; McDonald and Popkin 2002; Putnam 2001; Rosenstone and Wolfinger 1980; Skocpol 2004), public opinion formation (Achen 1975; Carmines and Stimson 1989; Converse 1964; Page and Shapiro 1992; Zaller 1992), determinants of vote choice (Abrams and Fiorina, 2009; Berelson, Lazarsfeld, and McPhee 1954; Campbell et al. 1964; Fiorina 1981; Key 1966), and the impact of party identification (Bartels 2000; Erikson, MacKuen, and Stimson 2002; Green, Palmquist, and Schickler 2004; Wattenberg 1986), to name a few.

This article presents a set of tools to study these topics and others in greater geographic and demographic detail than has been previously possible. Moving beyond traditional regression and crosstab-based inferences, we use multilevel regression and poststratification (MRP) to derive precise vote choice and turnout estimates for small

subgroups of the population. Lax and Phillips (2009a, 2009b) discuss the benefits of MRP in statistical and substantive terms in the context of state-to-state variation in opinion and policies on gay rights. Following Gelman and Little (1997) and Park, Gelman, and Bafumi (2004), Lax and Phillips estimated public opinion in small subsets of the population but used these only as intermediate quantities to be summed (in the poststratification step) to get averages for each state.

We improve upon this process by deriving estimates for demographic categories within states. As a simple example, imagine we want to break the population into five income categories, four ethnic groups (non-Hispanic white, black, Hispanic, other), and 51 states (including the District of Columbia), and we are interested in estimating the rate of turnout and average vote (Democrat vs. Republican) within each cell. Even this simple example totals $5 \times 4 \times 51 = 1,020$ cells, making the task nontrivial. If we can break the population into mutually exclusive categories, however, we are provided the flexibility of combining them arbitrarily, for example averaging all Hispanic voters together, or all voters in Delaware, or all black low-income voters in Georgia (this would be a

Yair Ghitza is a Ph.D. Candidate, Department of Political Science, Columbia University, 420 West 118th Street, 7th Floor, IAB, New York, NY 10027 (yg2173@columbia.edu). Andrew Gelman is Professor, Department of Statistics and Department of Political Science, Columbia University, 420 West 118th Street, 7th Floor, IAB, New York, NY 10027 (gelman@stat.columbia.edu).

We thank Aleks Jakulin, Daniel Lee, Yu-Sung Su, Michael Malecki, Jeffrey Lax, Justin Phillips, Shigeo Hirano, and Doug Rivers for their many helpful ideas. We also thank the Institute of Education Sciences, National Security Agency, Department of Energy, National Science Foundation, and the Columbia University Applied Statistics Center for partial support of this work.

American Journal of Political Science, Vol. 57, No. 3, July 2013, Pp. 762–776

©2013, Midwest Political Science Association

DOI: 10.1111/ajps.12004

single cell). As we add demographic levels to the analysis, estimating cell values becomes more difficult.

This work is in the forefront of statistical analysis of survey data. *Methodologically*, we improve upon the existing MRP literature in five ways: (1) modeling deeper levels of interaction between geographic and demographic covariates, (2) allowing for the relationship between covariates to be nonlinear and even nonmonotonic, if demanded by the data, (3) accounting for survey weights while maintaining appropriate cell sizes for partial pooling, (4) adjusting turnout and voting levels after estimates are obtained, and (5) introducing a series of informative multidimensional graphical displays as a form of model checking. *Substantively*, we improve upon the work of Bafumi, Gelman, Park, and Shor (2007)—who studied variation among states and regions in the relation of income and voting—by moving from two explanatory factors to four and by modeling turnout and vote choice in an integrated framework.

To demonstrate the flexibility of our method, we will present several examples of analyses that we conducted shortly after the 2008 election. Our goal is not to estimate a single regression coefficient or identify a single effect, as is often the case in social science. Rather, our goal is to paint a broader portrait of the distribution of the electorate. The graphs that we construct along the way clarify our intuitions and help us understand the subgroup-level characteristics of both turnout and vote choice.

Through most of the article, we focus on description and model checking instead of deep causal questions. Given the uncertainties surrounding demographic voting trends and their interaction with state-to-state variation, we feel it is an important contribution to simply put together this information and measure these trends, setting up a firm foundation for future researchers to study fundamental political questions using the best possible survey-based estimates. As such, this work fits with recent literature that devotes considerable effort to deriving better estimates which can be fed into later analyses.¹

The article proceeds as follows. The next section lays out our statistical methods in detail, describing our methods all the way from statistical notation through computational implementation details and graphical model checking. After describing our data sources, we then go through a series of examples, all of which were conducted after the 2008 election as we were analyzing voting and

turnout trends over recent presidential elections. In tandem, the examples illustrate our ability to construct stable and reasonable estimates even for detailed subgroups. We conclude with discussion.

Statistical Methods

Notation

We develop the notation in the context of a general three-way structure:

The population is defined based on three variables, taking on levels $j_1 = 1, \dots, J_1$; $j_2 = 1, \dots, J_2$; $j_3 = 1, \dots, J_3$. For example, in our model of income \times ethnicity \times state, $J = (5, 4, 51)$. Any individual cell in this model can be written as $j = (j_1, j_2, j_3)$. We index the three factors as $k = 1, 2, 3$.

We further suppose that each of the three factors k has L_k group-level predictors and is thus associated with a $J_k \times L_k$ matrix X^k of group-level predictors. The predictors in our example are as follows: for income ($k = 1$), we have a 5×2 matrix X^1 whose two columns correspond to a constant term and a simple index variable that takes on the values 1, 2, 3, 4, 5. For ethnicity ($k = 2$), we only have a constant term, so X^2 is a 4×1 matrix of ones. For state ($k = 3$), we have three predictors in the vote choice model: a constant term, Republican vote share in a past presidential election, and average income in the state (we also will use the classification of states into regions, but we will get back to this later). Thus, X^3 is a 51×3 matrix.

Finally, each of our models has a binary outcome, which could be the decision of whether to vote or for which candidate to vote. In any case, we label the outcome as y and, within any cell j , we label y_j as the number of Yes responses in that cell and n_j as the number of Yes or No responses (excluding no-answers and other responses). Assuming independent sampling, the data model is $y_j \sim \text{Binomial}(n_j, \theta_j)$, where θ_j is what we want to estimate: the proportion of Yes responses in cell j in the population.

Poststratification

For some purposes we are interested in displaying the estimated θ_j 's directly, for example when mapping voter turnout or vote intention by state, with separate maps for each age and income category (see Figure 4). Other times we want to aggregate across categories, for example summing over race to estimate votes by income and state (Figure 2) or averaging nationally or over regions to focus on demographic breakdowns.

¹Examples include better measures of roll-call data (Carroll et al. 2009; Clinton, Jackman, and Rivers 2004), district-level preferences (Jackman, Levendusky, and Pope 2008; Kernell 2009), voter knowledge, the ideological connection between voters, legislators, and candidates (Ansolabehere and Jones 2010; Bafumi and Herron 2010; Jessee 2009), and many others.

In the latter cases, we poststratify—that is, average over groups in proportion to their size in the population. We might be averaging over the voting-age population, or the voting-eligible population, or the population of voters, or even a subset such as the people who voted for John McCain for president. In any case, label N_j as the relevant population in cell j , and suppose we are interested in θ_S : the average of θ_j 's within some set J_S of cells. The poststratified estimate is simply

$$\theta_S = \sum_{j \in J_S} N_j \theta_j / \sum_{j \in J_S} N_j. \quad (1)$$

For example, to prepare the state \times ethnicity estimates in Figure 3, we aggregated the $5 \times 4 \times 51$ cells j into 4×51 sets J_S , with each poststratification (1) having four terms in the numerator and four in the denominator.

When the N_j 's are known, or are treated as known (as in the case of the voting-age population; see the “Data Sources” section), poststratification is easy. When the N_j 's are merely estimated, we continue to apply (1), this time plugging in estimates of the N_j 's obtained from some preliminary analysis. This should be reasonable in our application, although in more complex settings, a fully Bayesian approach might be preferred in order to better propagate the uncertainty in the estimated group sizes (Schutt 2009). For the remainder of this article, we shall treat the N_j 's as known.

Setting Up Multilevel Regression

We fit a model that includes group-level predictors as well as unexplained variation at each of the levels of the factors and their interactions. This resulting nonnested (crossed) multilevel is complicated enough that we build it up in stages.

Classical logistic regression. To start, we fit a simple (nonmultilevel) model on the J cells, with cell-level predictors derived from the group-level predictor matrices X^1, X^2, X^3 . For each cell j (labeled with three indexes as $j = (j_1, j_2, j_3)$), we start with the main effects, which comes to a constant term plus $(L^1 - 1) + (L^2 - 1) + (L^3 - 1)$ predictors (with the -1 terms coming from the duplicates of the constant term from the three design matrices). We then include all the two-way interactions, which give $(L^1 - 1)(L^2 - 1) + (L^1 - 1)(L^3 - 1) + (L^2 - 1)(L^3 - 1)$ additional predictors. In our example, these correspond to different slopes for income among Republican and Democratic states and different slopes for income among rich and poor states. The classical regression is formed by the binomial data model along

with a logistic link, $\theta_j = \text{logit}^{-1}(X_j \beta)$, where X is the combined predictor matrix constructed above.

Multilevel regression with no group-level predictors. If we ignore the group-level predictors, we can form a basic multilevel logistic regression by modeling the outcome for cells j by factors for the components, j_1, j_2, j_3 :

$$\theta_j = \text{logit}^{-1} \left(\alpha^0 + \alpha_{j_1}^1 + \alpha_{j_2}^2 + \alpha_{j_3}^3 + \alpha_{j_1, j_2}^{1,2} + \alpha_{j_1, j_3}^{1,3} + \alpha_{j_2, j_3}^{2,3} + \alpha_{j_1, j_2, j_3}^{1,2,3} \right), \quad (2)$$

where each batch of coefficients has its own scale parameter: for any subset S of $\{1, 2, 3\}$, α^S is an array with $\prod_{s \in S} J_s$ elements, which we give independent prior distributions $\alpha_j^S \sim N(0, (\sigma^S)^2)$. We complete the model with a prior distribution for the group-level variance parameters: $(\sigma^S)^2 \sim \text{inv-}\chi^2(\nu, \sigma_0^2)$, with these last two parameters given weak priors and estimated from data. Because the binomial distribution has an internally specified variance, it is possible to estimate all the variance components, up to and including the three-way interactions.

We can also write (2) in more general notation by summing over subsets S of $\{1, 2, 3\}$:

$$\theta_j = \text{logit}^{-1} \left(\sum_S \alpha_{S(j)}^S \right), \quad (3)$$

where $S(j)$ represents the indexes of j corresponding to the elements of S . In our running example, the subset $S = \{1, 3\}$ represents income \times state interactions, and for this set of terms in the regression, $S(j) = (j_1, j_3)$ indexes the income category and state corresponding to cell j . The terms in the summation within (3) correspond to the eight terms in (2).

Multilevel model with group-level predictors. The next step is to combine the above two models by taking the classical regression and adding the multilevel terms; this is a varying-intercept regression (also called a mixed-effects model):

$$\theta_j = \text{logit}^{-1} \left(X_j \beta + \sum_S \alpha_{S(j)}^S \right), \quad (4)$$

with a uniform prior distribution on the vector beta of “fixed effects” and a hierarchical Gaussian prior distribution for the batches of “random effects” α , as before.

Multilevel model with varying slopes for group-level predictors. The importance of particular demographic factors can vary systematically by state. For example, individual income is more strongly associated with Republican voting in rich states than in poor states. Gelman et al. (2007) fit this pattern using a varying-intercept, varying-slope multilevel logistic regression. Our model is more complicated than theirs, but the same principle

applies: we will fit the data better by allowing regression slopes—not just intercepts—to vary by group.

We implement by allowing each coefficient to vary by all the factors not included in the predictors. In our example, the coefficients for the state-level predictors (Republican presidential vote share and average state income) are allowed to vary by income level and ethnicity, while the coefficient for the continuous income predictor can vary by ethnicity and state. The general form of the model combines (2) and (3) by allowing each batch of coefficients in (3) to vary by group as in (2). To write this more general form requires another stage of notation in which the coefficients for any set of group-level predictors can be labeled β^S and whose components come from a distribution with mean 0 and standard deviation σ^S . This is analogous to the varying intercepts which are labeled α^S , as before.

Adding a multilevel model for one of the group-level factors. The final model includes classical logistic regression as a baseline to shrink to, and then this model's coefficients vary by group (in our example, varying by ethnic group, income category, and state).

Adding region as an additional predictor. We are already using state as one of the groups in the model, but we can add region as an additional predictor to capture effects that can be found for large areas of the country but perhaps do not have enough data to be captured by the state-level groups that are already in the model. We do this by expanding S from the set of subsets of $\{1, 2, 3\}$ to the set of subsets of $\{1, 2, 3, 4\}$, where α_4 will now refer to the region-level varying intercept, and including all relevant interaction terms. Because region is created as a direct mapping of state, there are no interactions between state and region (these interactions would be nonsensical and would add no additional information to the model). The final model, then, is:

$$\theta_j = \text{logit}^{-1} \left(\sum_S X^S \beta_{S(j)}^S + \sum_S \alpha_{S(j)}^S \right), \quad (5)$$

where S is the set of subsets of $\{1, 2, 3, 4\}$, referring to income, ethnicity, state, and region, $\beta_{S(j)}^S$ are varying slopes for group-level predictors, and $\alpha_{S(j)}^S$ are varying intercepts.

Accounting for Survey Weights

Many of our survey data come with weights. When fitting regressions to such data, it is not always necessary to include the weights. Simple unweighted regression is fine—as long as all the variables used in the weighting are included as regression predictors. The population in-

formation encoded in the survey weights enters into the analysis through the poststratification step (see Gelman 2007, for example).

Here, however, we are modeling based on only three factors (ethnicity, income, and state), but the survey adjustments use several other variables, including sex, age, and education. Ultimately we want to fit a complex model including all these predictors, but for now we must accept that our regression does not include all the weighting variables. Thus, our model must account for variation of weights within poststratification cells.

Within each cell j , we make two corrections. First, we replace the raw data estimate with \bar{y}_j^* , the weighted average of the outcome measure, y , within the cell. Second, we adjust the effective sample size for the measurement to account for the increased variance of a weighted average compared to a simple average, using the correction:

$$\text{design.effect}_j = 1 + \left(\frac{\text{sd}(\text{unit weights within cell } j)}{\text{mean}(\text{unit weights within cell } j)} \right)^2 \quad (6)$$

These estimated design effects are noisy, and for any given analysis we average them to get a single design effect for all the cells.

Putting these together, we account for weighting by using the data model $y_j^* \sim \text{Binomial}(n_j^*, \theta_j)$, where $n_j^* = \frac{n_j}{\text{design.effect}_j}$ and $y_j^* = \bar{y}_j^* n_j$. The resulting n_j^* , y_j^* will not in general be integers, but we handle this by simply using the binomial likelihood function with non-integer data, which works fine in practice (and is in fact simply the weighted log-likelihood approach to fitting generalized linear models with weighted data). This falls in the category of quasi-likelihood methods (Wedderburn 1974).

Computation

We ultimately would like to perform fully Bayesian inference for our models, but for now, we have been using the approximate marginal maximum likelihood estimates obtained using the `glmer()` program in R (Bates and Maechler 2009). Such estimates are sometimes justified on their own theoretical and computational grounds (e.g., Skrondal and Rabe-Hesketh 2004), but here we are considering them as approximations to fully Bayesian inference. Recent work on multilevel modeling and poststratification gives us confidence that this approach works well in estimating demographic and state-by-state breakdowns from national surveys (Lax and Phillips 2009a, 2009b).

TABLE 1 Variables in the `lmer()` Model, Along with Analogous Terms from the Statistical Model

<code>lmer()</code> Variable	Description	Type	Number of Groups	Coefficient in Statistical Model
<code>y</code>	Vote choice (1 = McCain, 0 = Obama)	Output variable	—	—
<code>z.incstt</code>	State-level income	Linear predictor	—	Part of $\beta^1, \beta^3, \beta^4$
<code>z.repprv</code>	State-level Republican vote share from previous election	Linear predictor	—	Part of $\beta^1, \beta^3, \beta^4$
<code>z.inc</code>	Income (included as a linear predictor)	Linear predictor/ Varying slope	—	Part of $\beta^2, \beta^3, \beta^4$
<code>inc</code>	Income	Varying intercept	5	α^1
<code>eth</code>	Ethnicity	Varying intercept	4	α^2
<code>stt</code>	State	Varying intercept	51	α^3
<code>reg</code>	Region of the country	Varying intercept	5	α^4
<code>inc.eth</code>	Income \times ethnicity interaction	Varying intercept	$4 \times 5 = 20$	$\alpha^{1,2}$
<code>inc.stt</code>	Income \times state interaction	Varying intercept	$4 \times 51 = 204$	$\alpha^{1,3}$
<code>inc.reg</code>	Income \times region interaction	Varying intercept	$5 \times 5 = 25$	$\alpha^{1,4}$
<code>eth.stt</code>	Ethnicity \times state interaction	Varying intercept	$5 \times 51 = 255$	$\alpha^{2,3}$
<code>eth.reg</code>	Ethnicity \times region interaction	Varying intercept	$4 \times 5 = 20$	$\alpha^{2,4}$

For our running example, the `lmer()` model looks like this:

```
model.fit <- lmer(y ~ z.inc*z.incstt +
  z.inc*z.repprv +
  (1|inc)+(1+z.inc|eth)+(1+z.inc|stt)+
  (1+z.inc|reg)+
  (1|inc.eth)+(1|inc.stt)+(1|inc.reg)+
  (1|eth.stt)+(1|eth.reg),
  family=binomial(link='logit'))
```

Where the vectors here have length $J = J_1 J_2 J_3$ and, for each element j : `y` is a two-column matrix indicating the number of Democratic and Republican voters in cell j (adjusted for varying survey weights). Table 1 describes each of the variables in the computational model and lists the analogous term from the statistical model. `inc`, `eth`, and `stt` are index vectors running from 1–5, 1–4, and 1–51 indexing the grouping factors in the model; `reg` indicates the region of the country, which is mapped directly from `stt`; `z.incstt` and `z.repprv` are state-level predictors of income and previous Republican vote that have been centered and rescaled (hence the “z”) and have been expanded to length J by repeating over the index `stt`; `z.inc` is the 1–5 `inc` index after it has been centered and rescaled (thus including it as a linear predictor as well as nonmonotonically); and `reg.eth`, `reg.inc`, `stt.eth`, `stt.inc`, and `eth.inc` are the interaction terms. Notice that `z.incstt` and `z.repprv` are considered part of β_3 and β_4 because they are included only as group-level predictors for these variables. In contrast, `z.inc` is both a

group-level predictor (for the income cells) and a varying slope (for the other cells), so it is included in all of the β s.

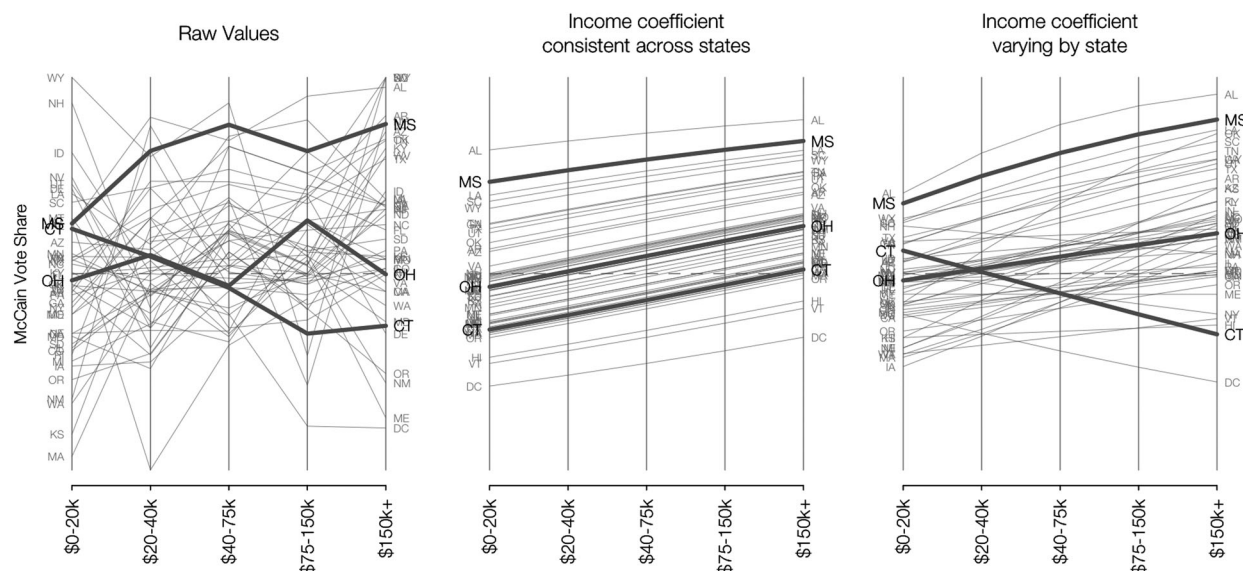
Our model includes income as a continuous variable (through the `z.inc` term) and also as a categorical factor (through the `inc` terms). Including both in the same model allows a nonlinear fit that partially pools toward linearity, thus given both the flexibility of the categorical fit and the statistical efficiency of including a linear term for an approximately linear predictor (see section 12.6 of Gelman and Hill 2007).

In future versions of the model, we can also include a term of the form `(1 | inc.eth.stt)` to allow saturated interactions. In the meantime, we estimate the residual cell-level variance and compare it to the (design-effect-adjusted) binomial variance. If the residual variance is much higher than would be expected from the sampling model, this implies that three-way interactions should be included. In the models fit for the present article, this was not the case. We have developed a freely available R package called `mrp`, which implements all of these steps, including the ability to fit the model with fully saturated interactions.

Graphical Display of Inferences and Model Checking

When social scientists present data, they tend to use tables instead of graphs, despite the ability of graphs to translate

FIGURE 1 The Evolution of a Simple Model of Vote Choice in the 2008 Election for State/Income Subgroups, Non-Hispanic Whites Only



Note: The first panel shows the raw data; the middle panel is a hierarchical model where state coefficients vary, but the (linear) income coefficient is held constant across states; the right panel allows the income coefficient to vary by state. Adding complexity to the model reveals weaknesses in inferences drawn from simpler versions of the model. Three states—Mississippi (the poorest state), Ohio (a middle-income state), and Connecticut (the richest state)—are highlighted to show important trends.

large amounts of data in clearer and less obtrusive ways (Kastellec and Leoni 2007). This is especially true when it comes to the presentation of regression estimates. Dozens of numbers with too many digits are squeezed into a tiny space, the reader drawn to the stars showing significant effects. Substantive effect size is generally interpretable for single coefficients, but what about interactions? Two-way interactions can sometimes be understood but are rarely teased out in any detail, and interpretability of three- and four-way interactions is virtually impossible.

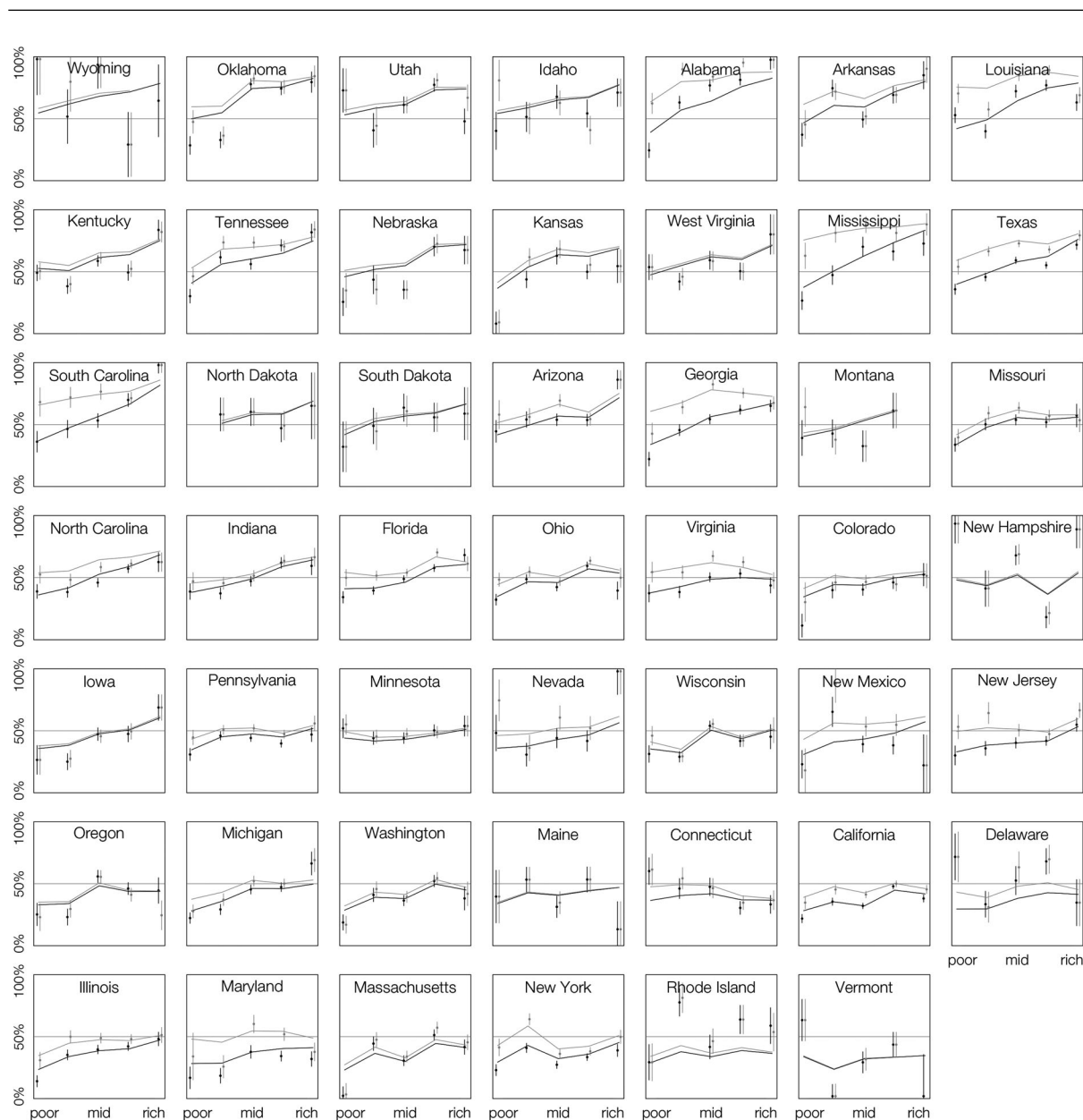
We take the opposite approach here and in our research in general, viewing graphical data visualization as a key step in understanding model fit and building confidence in our inferences. Our approach is to build a full model in stages: build a simple version of the model; graph results to check fit; make the model more complex; graph results to see if and how the model fit changes; continue until reaching the final model. An example of this is Figure 1, which shows the evolution of a simple model of vote choice in the 2008 election for state \times income subgroups. This is a parallel coordinates plot, showing estimates for non-Hispanic whites only. The first panel shows the raw data; the second panel is a simple model where state coefficients are allowed to vary, but the (linear) income coefficient is held constant across states; the last panel allows the income coefficient to vary by state.

Three states—MS, OH, and CT—are highlighted to show important trends.

These simple models and visualizations reveal quite a bit about the data. In the first panel we can see that raw estimates are noisy and insufficient to reveal any clear structure, despite a sample size exceeding 15,000. The second panel indicates high state-level variance and is suggestive that higher income is tied to higher McCain vote. The third panel shows that this simple story is insufficient: there is a wide variance in the income coefficient, with richer voters supporting McCain in Republican-leaning states but supporting Obama in Democratic-leaning states. This distinction is theoretically similar to (and more extreme than) the relationship between income and voting found in Gelman et al. (2008).

We prefer the graphical strategy in general, but even more so in the present project due to the implausibility of checking each parameter estimate for each of our models and the futility in trying to do so, as deep interactions would be uninterpretable in this setting. Instead of trying to interpret regression coefficients one at a time or in conjunction, examining fitted subgroup estimates facilitates a more natural interpretation of the model. In other words, it is easier to notice when subgroup estimates *seem* right, while regression coefficients are more difficult to assess. For example, when we started this analysis, we

FIGURE 2 2008 McCain Share of the Two-Party Vote in Each Income Category within Each State



Note: All voters shown in black and non-Hispanic whites in gray. Dots are weighted averages from pooled June/November Pew surveys; error bars show ± 1 s.e. bounds. Curves are estimated using multilevel models and have a s.e. of about 3% at each point. States are ordered in decreasing order of McCain vote (Alaska, Hawaii, and Washington, DC, excluded).

knew a priori that our estimates for Obama's vote share among African American groups needed to be high, over 90%, but we could not know what regression coefficient was plausible, as the coefficient could change drastically depending on functional form.

Another example of the power of graphical model checking is shown in Figure 2. While iteratively building our model, it became clear that there were problems

in our estimates, especially as they related to ethnicity.² These graphs represent a better way of looking at the data as a whole—they indicate McCain's share of the two-party vote in each income/state cell, estimated for all voters (in

²We posted maps on the Internet, and politically savvy readers noted problems in some of the state/income categories which could be traced back to interactions between ethnicity, income, and state that had not been included in the earlier versions of our model.

black) and non-Hispanic whites (in gray). While some states have similar black and gray trend lines, they diverge tremendously in some cases. Any method for fitting an elaborate model should come with procedures for evaluating it and building confidence. The state-by-state plots in Figure 2 are, we believe, a good start on the way to the general goal of tracing the mapping from data to inference.

Data Sources

For estimates of cell population size, we use the 5% public use microsample (PUMS) of the long-form Census for 2000, which has a sample of size 9,827,156 among voting-age citizens. When broken down into subgroups, this yields, for instance, a weighted estimate of 4,596 white women in Kentucky aged 45–64 with college educations and incomes over \$100,000, or 156 black men in North Carolina aged 30–44 with postgraduate degrees and incomes under \$20,000. Because of the large sample size of the PUMS data and the fact that population-size estimates are not our primary quantities of interest, we treat these PUMS numbers as truth. For 2004 and 2008, we use the American Community Survey (ACS), a large national sample that gives much of the same information as the long-form Census (enough information to construct the same subgroups in each of the years). We use the 2004 ACS ($N = 850,924$ for voting-age citizens) for our 2004 estimates, and we use the pooled 2005–2007 ACS ($N = 6,291,316$ for the same group) for our 2008 estimates. Again, for now we take the weighted ACS values as exact numbers of the voting-age population in each subcategory.³

To construct turnout estimates, we use the Current Population Survey's (CPS) post-election Voting and Registration supplement, conducted every two years in November and generally considered to be the gold standard on voter turnout, especially when it comes to estimating turnout for demographic subgroups. The survey does not ask people how they voted, but it asks whether they voted. We can compare the survey results, nationally and at the state level, with actual number of votes.

³If we reach the stage of being interested in extremely small groups of the population, we can fit an overdispersed Poisson model to capture sampling variability in the context of weighted survey data: in each cell j , let n_j be the number of CPS respondents in the cell and w_j be the average survey weight of the respondents in the cell. The model is $n_j \sim \text{overdispersed-Poisson}(\theta_j/w_j)$, where θ_j is the actual voting-age population in the cell. The simple estimate of θ_j is proportional to $w_j n_j$, but with sparse data a model could be helpful.

We use Michael McDonald's "highest office" vote totals.⁴ The CPS comes close to these numbers. For example, 131,304,731 people voted for president in 2008, representing an estimated 57% of the voting-age population and 62% of the voting-eligible population. The CPS turnout estimate is 68% for voting-age citizens. This estimate is higher than McDonald's estimates, most likely due to vote overreporting bias, but the CPS generally has less overreporting bias than other surveys like the American National Election Studies.

For each election, we use the CPS to estimate the probability that a voting-age citizen will turn out to vote, given his or her demographics and state of residence ($N = 68,643; 79,148; \text{ and } 74,327$ in 2000, 2004, and 2008 after removing missing data). We know the actual number of voters for each state and perform a simple adjustment so that overall turnout matches the state totals, as follows. Let ξ_s indicate the number of voters for each state $s = 1, \dots, 51$, and let S denote the set of cells such that j is in state s . We derive the adjusted turnout estimate θ_j^* for each cell $j \in S$ as follows:

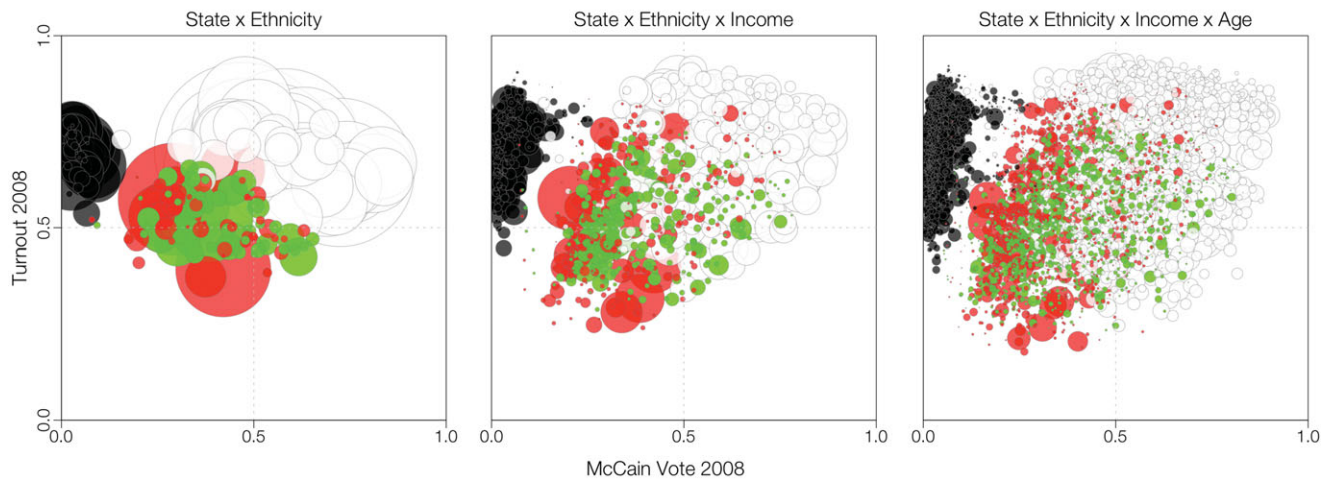
$$\delta_s = \min \left(\text{abs} \left(\xi_s - \sum_{j \in S} (N_j \logit^{-1}(\logit(\theta_j) + \delta)) \right) \right) \quad (7)$$

$$\theta_j^* = \theta_j + \delta_s \quad \forall j \in S, \quad (8)$$

where $\text{abs}()$ is the absolute value function and $\min()$ is a function that finds the δ that minimizes the expression. This process simply applies a constant logistic adjustment δ_s to each cell in state s to make sure that the total number of estimated voters is correct. We assume here that overreporting bias is consistent across cells within state. Because the CPS data have a vote overreport bias, δ_s is usually negative.

We are interested in differences in turnout rate by demographic group, so it is important to consider whether overreport bias is consistent across demographic groups. In the 1980s, scholars investigated demographic correlates of overreporting in the American National Election Studies (ANES). Though evidence is mixed, overall there were no consistently strong relationships between demographics and overreporting, with the exception of African Americans slightly overreporting more than whites (Abramson, Anderson, and Silver 1986; Abramson and Claggett 1984; Katosh and Traugott 1981; Sigelman 1982). Bernstein, Chadha, and Montjoy (2001) revisited the data, arguing that overreporting is more likely among people who are under the most social pressure to vote, such as educated people, partisans, and minorities in minority districts, but Cassel (2003) showed

⁴United States Elections Project, George Mason University.

FIGURE 3 Turnout and Vote Choice for Population Subgroups, Presidential Election 2008

Note: Size = Subgroup population size 2007; Color by ethnicity: White = White, Black = Black, Red = Hispanic, Green = Other. Each bubble represents one demographic subgroup per state, with size and color indicating population size and ethnicity. As additional demographics are added, heterogeneity within subgroups is revealed by the dispersion of the bubbles, while estimates remain reasonable.

that these findings are model-specific and the effect is generally small in typical model specifications. In terms of the CPS data used in this study, McDonald (2007) recently showed that demographic descriptions of the electorate are approximately the same when using the CPS data and voter registration files, in contrast to using exit polls, which show a younger electorate with more minorities. In many ways, then, the CPS data are the “gold standard” of survey data on voter turnout.

To construct vote choice estimates, we use the National Annenberg Election Survey in 2000 and 2004 and Pew Research pre-election polls ($N = 31,719; 43,970$; and $19,170$ in 2000, 2004, and 2008 after removing missing data). These surveys get large samples by aggregating rolling cross-sections and waves conducted over several months. The model is of the same form, and we do a similar adjustment for vote choice as in equations (6) and (7) above, substituting the estimated number of voters $\theta_j^* N_j$ instead of population size N_j . Because vote choice does not suffer overreporting bias, δ_s is usually smaller in magnitude here.

Putting It All Together

Now that our model and graphical approach are fully described, we present examples of the type of analysis that can be done using this method. We fit numerous models using the framework described above, using the following covariates: state, region, ethnicity, income, education, and age. Eventually we want to include additional demo-

graphics such as sex, religion, number of children, and others, and we are currently working on software which will allow model fitting in this higher dimensional space.

Demographic Expansion

Figure 3 shows the distribution of geographic/demographic subgroups in the 2008 presidential election. In each of these graphs, the x - and y -axes show McCain vote and election turnout, respectively. Moving from left to right, we add additional demographic covariates—state and ethnicity are shown on the left, family income is added in the middle,⁵ and age is added on the right.⁶ Bubbles are sized proportionally to population, and colors indicate ethnicity: white = non-Hispanic white, black = African American, red = Hispanic, green = Other, all drawn with transparency to increase visibility.⁷ By the end of the full demographic expansion, $51 \times 4 \times 5 \times 4 = 4,080$ groups are plotted in on the right.

This type of graph helps us confirm top-level trends on race-based voting and turnout, and it builds confidence in our estimates. The left graph shows that, as expected, African Americans in all states voted overwhelmingly for Obama. Hispanic voters and other nonwhites also voted heavily Democratic, while white voters are

⁵ \$0–20k, \$20–40k, \$40–75k, \$75–150k, and >\$150k per year.

⁶ 18–29, 30–44, 45–64, 65+.

⁷ All of these analyses are based on the voting-age citizen population, so turnout numbers are not biased by different levels of citizenship.

spread out and more likely to vote for McCain. In terms of turnout, Hispanics and Others voted less than other groups as a whole. As we add covariates, the bubbles become increasingly dispersed. Although mainstream political commentary tends to think of demographic groups (especially minorities) as homogenous voting blocs, they exhibit substantial heterogeneity. For example, consider African Americans in North Carolina. In a state that went 50–49 for Obama, they comprised roughly 20% of the population, had 72% turnout (similar to the state total of 71%), and voted for Obama 95–5. However, looking more closely we can see that the richest African Americans in North Carolina “only” voted for Obama 86–14 with a turnout of 84%, while the poorest went 97–3 with 53% turnout. These differences (11 points in vote choice and 31 points in turnout) are substantial. As another example, let us compare Hispanics to non-Hispanic whites in New Mexico, another important swing state. As a whole, Hispanic turnout was much lower (53% compared to 74%), but there was basically no difference among the richest and most educated (89% to 93%).

The important takeaways here are that (1) there are substantial and important differences between subgroups, even within demographic categories, and (2) our method captures those differences while keeping estimates stable and reasonable. This is in line with Figure 1: there we showed that raw estimates are too noisy to be interpretable and that increasingly complicated statistical models help reveal trends in the data. Here we show the same thing with more variables included.

Homogeneous or Heterogeneous Vote Swing?

One of the important features of the MRP framework is that we can look at the overall distribution of estimates as well as combine estimates in any way we please. We can, for example, combine estimates to examine shifts in vote choice for demographic/geographic subgroups. It is well known that states, *when measured in the aggregate*, tend to shift uniformly from election to election (Gelman and Lock 2010). This can be referred to as a homogeneous partisan swing—for example, the average change in Republican vote from 1996 to 2000 was 5.6 points with a standard deviation of 3.3 (after removing Washington, DC, as an outlier). That is, after accounting for the national swing toward Bush, most states were within 3–4 points compared to their relative position in 1996. The standard deviations for the 2000–2004 and

2004–2008 swings (excluding DC) are 2.4 and 3.8, showing similar stability.

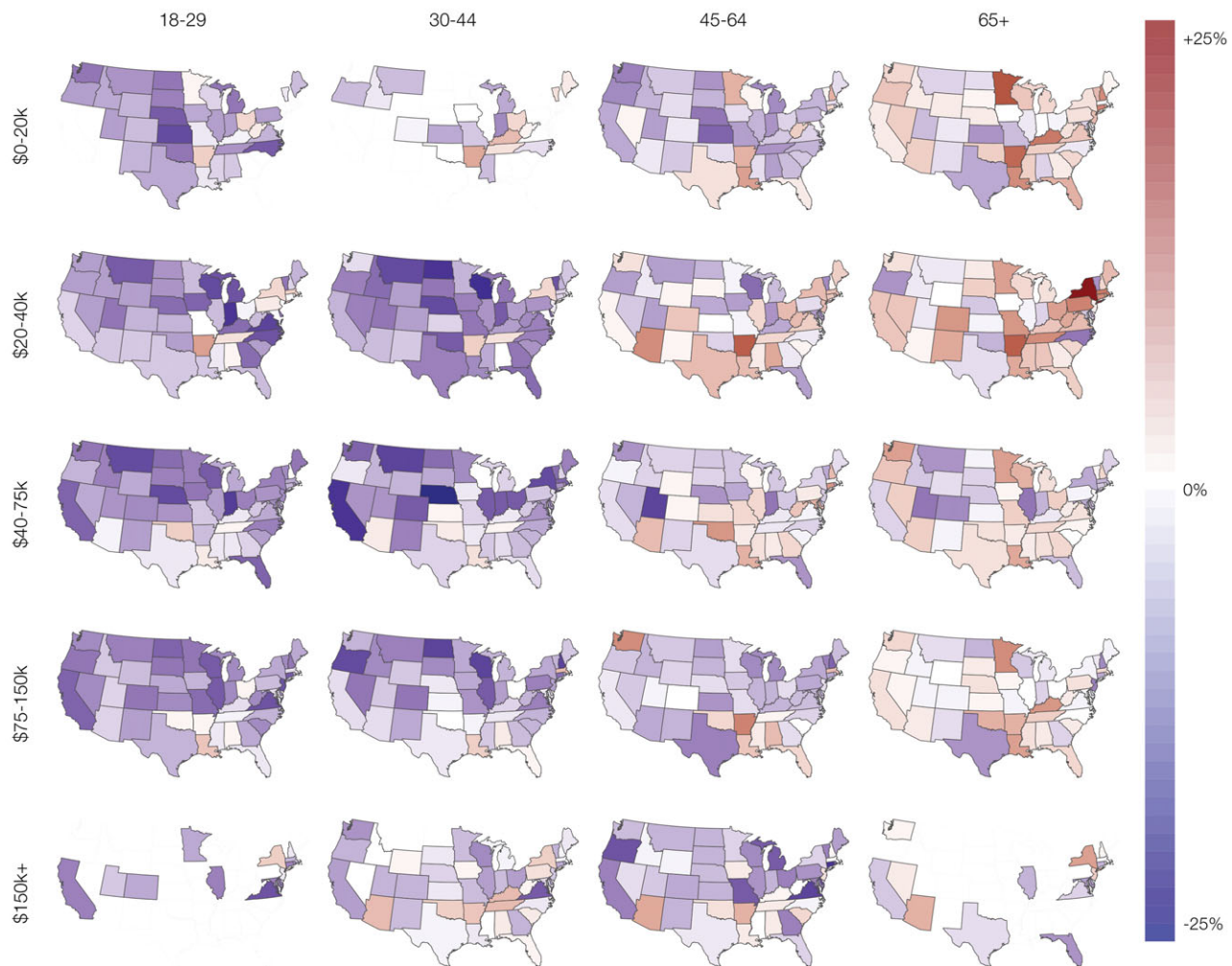
When we break the electorate down by demographics, though, the homogenous swing breaks down. It is easy to show that there was an enormous difference between ethnicities—for example, whites had a 3.3% shift toward Obama and nonwhites had a 7.8% shift—but again a single demographic cut hides much of the variation. Figure 4 displays public opinion change among whites as a series of maps broken down by age and income. Although almost every state moved toward Obama as a whole, this graph clearly shows subgroups that resisted the aggregate electoral forces and moved toward McCain, sometimes by substantial margins. These anti-Obama groups are mostly poorer and older white voters, especially in the South and Appalachia.⁸

Turnout Swing

As mentioned, our framework is not just a way to look at state-level estimates; rather, it allows us to combine subgroups in any way we please. For the following example, we move to national trends by examining change in turnout levels from 2004 to 2008. Turnout went up as a whole this election, but the upward turnout swing was not uniform. One of the main storylines of the 2008 campaign, in fact, was Obama's ability to energize a new group of voters, especially minorities and young people. Figure 5 evaluates that claim by breaking out the turnout swing by age, ethnicity, and income. Each plot here actually shows a number of things. The histograms show the distribution of age/ethnicity group by income—going from left to right in each plot shows low to high income—while the trend line shows the turnout change. The aggregate turnout increase (3.6%) is plotted as the horizontal reference line, with another reference line at 0% change.

The main point that we would like to highlight here is that the turnout swing was primarily driven by African Americans and young minorities. These groups are highlighted with a thick box and lines because they are the only groups with a total turnout change over 5%. Although the popular consensus would imply that young white voters also increased their turnout, that is simply not the case. Poor younger whites indeed turned out at higher rates than before, but this is a small subset of that

⁸Many of these groups also disapproved of Obama's health care reform agenda in 2009. Looking at these data with the benefit of hindsight, it seems that the roots of his political problems with this group were planted during the election.

FIGURE 4 McCain 2008 Minus Bush 2004 among Non-Hispanic Whites

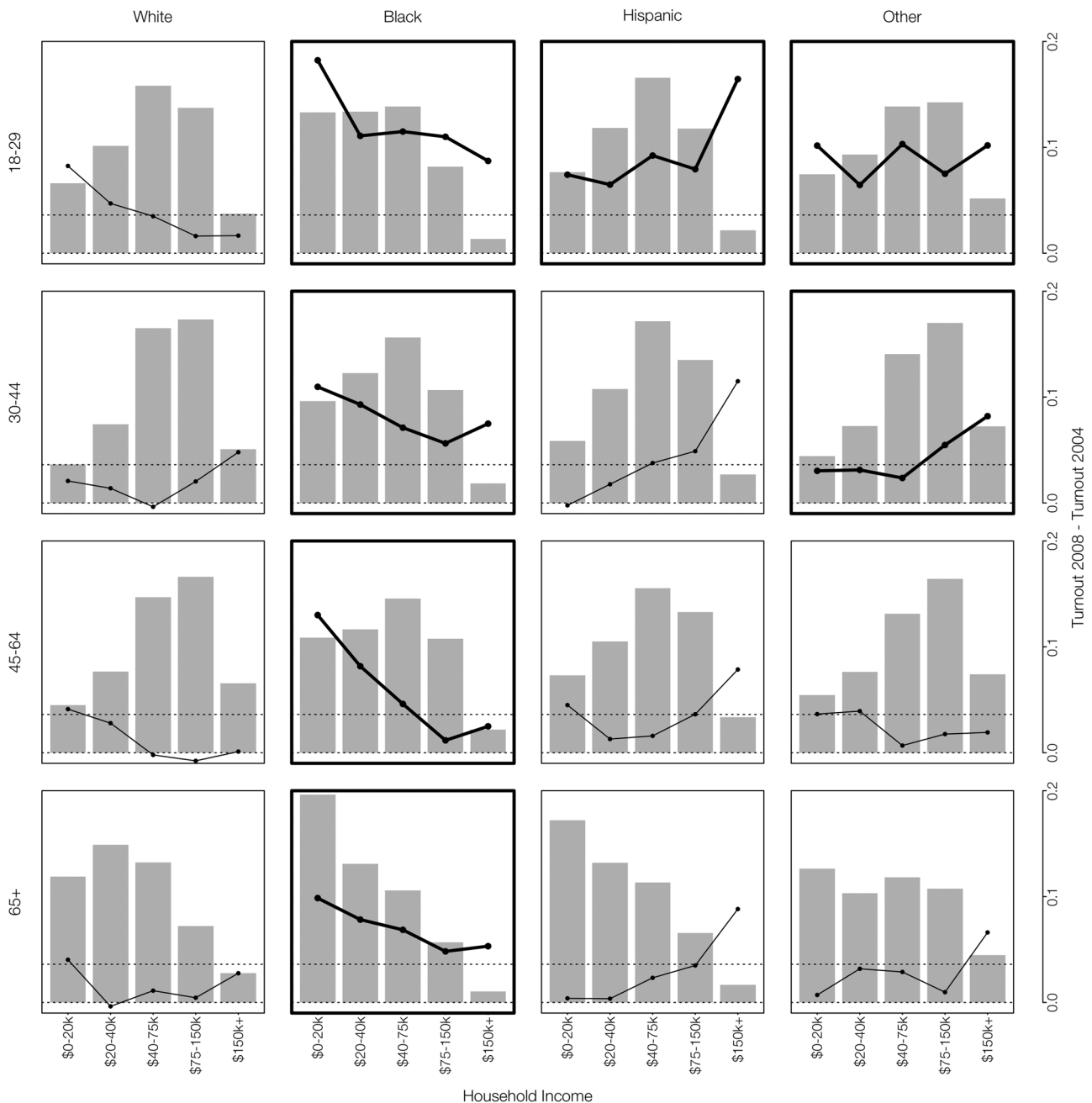
Note: State-by-state shift toward McCain (red) or Obama (blue) among white voters broken down by income and age. Red = McCain better than Bush; Blue = McCain worse than Bush. Only groups with >1% of state voters shown. Although almost every state moved toward Obama in aggregate, there are substantial demographic groups that moved toward McCain all over the map, specifically among older whites.

overall group, as shown in the histograms. The incorrect interpretation that has often been given is driven by improperly combining all young people into a single group. By breaking them out, we see that there is a big difference between white young people and minority young people. Of course, our framework allows us to break this out by state and to show the final turnout levels for each subgroup, as shown in Figure 6. Young white turnout is low, hovering in the 30–40% range for most states.⁹ Hispanic and other ethnicities also remain low in turnout levels.

⁹It only rises to a high level in Minnesota, which is a same-day registration state and has high levels of turnout for all subgroups.

Discussion

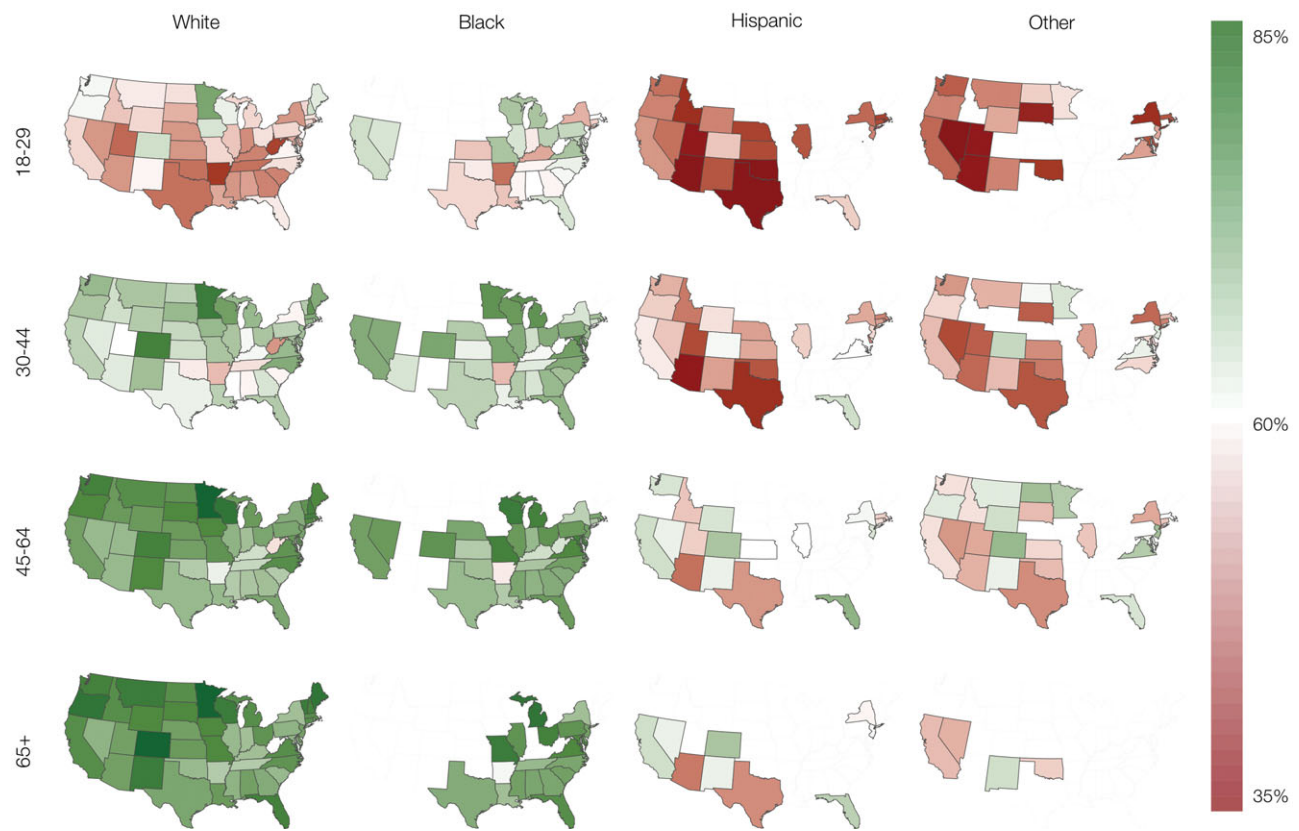
This article has introduced and described our method for producing estimates of turnout and vote choice for deeply interacted subgroups of the population: groups that are defined by multiple demographic and geographic characteristics. Although regression models have been used for decades to infer these estimates, MRP is an improvement over traditional methods for several reasons. Multilevel modeling allows estimates to be partially pooled to take advantage of common characteristics in different parts of the electorate, while poststratification corrects for the underlying distribution of the electorate.

FIGURE 5 Turnout Swing Mainly Isolated to African Americans and Young Minorities

Note: Turnout change shown in line graphs; population distribution shown as bar graphs. Turnout changes in the 2008 election were not consistent across demographic subgroups. African Americans and young minorities increased turnout almost uniformly, but white voters did not. Groups with a total turnout change over 5% are highlighted with a thick box and trend line.

Our method improves upon even the most recent implementations of MRP, though, by modeling deeper levels of interactions and allowing for the relationship between covariates to be nonlinear and even nonmonotonic—in other words, we let the data define the appropriate level

of nonlinearity and interaction between covariates. Our method also respects the design information included in survey weights, and lastly, it makes aggregate adjustments to make sure our final estimates are reasonable. At a substantive level, we have been able to integrate the study of

FIGURE 6 Turnout among Young Whites and Minorities Over-Emphasized

Note: 2008 turnout for ethnicity \times age subgroups; only groups with $>1\%$ of state population shown. This is another way to look at the lower turnout of Hispanics and Others. Despite modest increases, turnout among young white people is still low in comparison to other groups.

vote choice and turnout at a level of specificity that has not been possible before.

We use MRP to make inferences in the presence of sparse data. Given that our model is necessarily imperfect, we can interpret our estimates as smoothed versions of the data. In addition to working on making the model more realistic (for example, by including nonlinearity and interactions), it makes sense at each stage to compare the MRP estimates to corresponding raw-data summaries so that we and other consumers of the analyses can understand the effects of the modeling assumptions on the inferences.

Adding these layers lets the data speak more freely to the final estimates, but it imposes challenges in interpreting the final model. As a result, we recommend a gradual and visual approach to model building: build a simple model, graph inferences, add complexity to that model in the form of additional covariates and interactions, graph, and continue until all appropriate variables are included. The purpose of intermittent graphing is

to ensure that model estimates remain reasonable and that changes induced by additional covariates or interactions are understood. Because there will eventually be too many interactions to be interpreted by simply looking at the coefficients, it is important to graph final estimates as a substitute or in addition to the coefficients alone. The model builder with domain-specific knowledge will more likely find it easier to interpret and understand final estimates—for example, among African Americans, 90% support for Obama is easier to interpret than a coefficient of 2.56, although both may infer the same thing.

We have used the U.S. presidential elections of 2004 and 2008 to illustrate this process and have found a number of nonobvious trends, mainly focusing on the 2008 campaign between Barack Obama and John McCain: (1) Although demographic subgroups are often described as monolithic aggregates—especially when it comes to ethnicity—they are in fact quite diverse when broken down by other characteristics like income and education. (2) States as a whole essentially display a “homogeneous

swing” between elections, but demographic subgroups within those states show more variability. (3) The Obama coalition was weakened by older, white low-income voters who moved away from him in the election, foreshadowing the difficulty he had convincing this group to support his health care initiatives in 2009 and beyond. (4) Despite media reports to the contrary, there was not a substantial turnout swing among young white voters; in fact, most of the increase in turnout came from African Americans and other young minorities. Lastly, (5) despite modest increases, turnout among young white people and among Hispanics is still low in comparison to other demographic groups.

Through this article, our focus has been both introductory and descriptive: introductory because we have provided inferences for a relatively small number of demographic/geographic combinations, and descriptive because we have only briefly touched on substantive topics as illustrations, while intentionally avoiding deep causal questions. Although these methods can certainly be used in conjunction with other tools of causal inference, the purely observational data are inappropriate for that task. We have also ignored issue opinion entirely.¹⁰ Still, given the uncertainties surrounding demographic voting trends and their interaction with state-to-state variation, we feel these methods can be used to derive better survey estimates and set up a firm foundation for future researchers to study fundamental questions using the best possible data.

References

- Abramson, Paul, and William Claggett. 1984. “Race-Related Differences in Self-Reported and Validated Turnout.” *Journal of Politics* 46(3): 719–38.
- Achen, Chris. 1975. “Mass Political Attitudes and the Survey Response.” *American Political Science Review* 69(4): 1218–31.
- Ansolabehere, Stephen, Jonathan Rodden, and James M. Snyder. 2008. “The Strength of Issues: Using Multiple Measures to Gauge Preference Stability, Ideological Constraint, and Issue Voting.” *American Political Science Review* 102(2): 215–32.
- Ansolabehere, Stephen, and Phillip E. Jones. 2010. “Constituents’ Responses to Congressional Roll-Call Voting.” *American Journal of Political Science* 54(3): 583–97.
- Bafumi, Joseph, and Michael C. Herron. 2010. “Leapfrog Representation and Extremism: A Study of American Voters and Their Members in Congress.” *American Political Science Review* 104(3): 519–42.
- Bartels, Larry M. 2000. “Partisanship and Voting Behavior, 1952–1996.” *American Journal of Political Science* 44(1): 35–50.
- Bates, Douglas, and Martin Maechler. 2009. “Package lme4.” lme4.r-forge.r-project.org.
- Berelson, Bernard R., Paul F. Lazarsfeld, and William N. McPhee. 1954. *Voting: A Study of Opinion Formation in a Presidential Campaign*. Chicago: University of Chicago Press.
- Bernstein, Robert, Anita Chadha, and Robert Montjoy. 2001. “Overreporting Voting: Why It Happens and Why It Matters.” *Public Opinion Quarterly* 65(1): 22–44.
- Campbell, Angus, Philip E. Converse, Warren E. Miller, and Donald E. Stokes. 1964. *The American Voter*. New York: Wiley.
- Carmines, Edward G., and James A. Stimson. 1989. *Issue Evolution: Race and the Transformation of American Politics*. Princeton, NJ: Princeton University Press.
- Carroll, Royce, Jeffrey B. Lewis, James Lo, Keith T. Poole, and Howard Rosenthal. 2009. “Measuring Bias and Uncertainty in DW-NOMINATE Ideal Point Estimates via the Parametric Bootstrap.” *Political Analysis* 17(3): 261–75.
- Cassel, Carol A. 2003. “Overreporting and Electoral Participation Research.” *American Politics Research* 31(1): 81–92.
- Clinton, Joshua, Simon Jackman, and Douglas Rivers. 2004. “The Statistical Analysis of Roll Call Data.” *American Political Science Review* 98(2): 355–70.
- Converse, Philip E. 1964. “The Nature of Belief Systems in Mass Publics.” In *Ideology and Discontent*, ed. David E. Apter. New York: Free Press, 206–261.
- Downs, Anthony. 1957. *An Economic Theory of Democracy*. Reading, MA: Addison Wesley.
- Erikson, Robert S., Michael B. MacKuen, and James A. Stimson. 2002. *The Macro Polity*. Cambridge: Cambridge University Press.
- Fiorina, Morris P. 1981. *Retrospective Voting in American National Elections*. New Haven, CT: Yale University Press.
- Fiorina, Morris P., and Samuel J. Abrams. 2009. *Disconnect: The Breakdown of Representation in American Politics*. Norman: University of Oklahoma Press.
- Gelman, Andrew, Boris Shor, Joseph Bafumi, and David K. Park. 2007. “Rich State, Poor State, Red State, Blue State: What’s the Matter with Connecticut?” *Quarterly Journal of Political Science* 2: 345–67.
- Gelman, Andrew, Daniel Lee, and Yair Ghitza. 2010. “Public Opinion on Health Care Reform.” *The Forum* 8: 1–14.
- Gelman, Andrew, David K. Park, Boris Shor, and Joseph Bafumi. 2008. *Red State, Blue State, Rich State, Poor State: Why Americans Vote the Way They Do*. Princeton, NJ: Princeton University Press.
- Gelman, Andrew, and Jennifer Hill. 2007. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge: Cambridge University Press.
- Gelman, Andrew, and Thomas C. Little. 1997. Poststratification into Many Categories Using Hierarchical Logistic Regression.” *Survey Methodology* 23(2): 127–35.
- Gerber, Alan S., and Donald P. Green. 2000. “The Effects of Canvassing, Telephone Calls, and Direct Mail on Voter

¹⁰We describe health care opinions elsewhere; see Gelman, Ghitza, and Lee (2010).

- Turnout: A Field Experiment." *American Political Science Review* 94(3): 653–63.
- Green, Donald P., Bradley Palmquist, and Eric Schickler. 2004. *Partisan Hearts and Minds: Political Parties and the Social Identities of Voters*. New Haven, CT: Yale University Press.
- Jessee, Stephen A. 2009. "Spatial Voting in the 2004 Presidential Election." *American Political Science Review* 103(1): 59–81.
- Kastellec, Jonathan P., and Eduardo L. Leoni. 2007. "Using Graphs Instead of Tables in Political Science." *Perspectives on Politics* 5(4): 755–71.
- Katosh, John, and Michael Traugott. 1981. "The Consequences of Validated and Self-Reported Voting Measures." *Public Opinion Quarterly* 45(4): 519–35.
- Kernell, Georgia. 2009. "Giving Order to Districts: Estimating Voter Distributions with National Election Returns." *Political Analysis* 17(3): 215–35.
- Key, V. O. 1966. *The Responsible Electorate: Rationality in Presidential Voting, 1936–1960*. Boston: Belknap Press.
- Lax, Jeffrey, and Justin Phillips. 2009a. "Gay Rights in the States: Public Opinion and Policy Responsiveness." *American Political Science Review* 103(3): 367–86.
- Lax, Jeffrey, and Justin Phillips. 2009b. "How Should We Estimate Public Opinion in the States?" *American Journal of Political Science* 53(1): 107–21.
- Levendusky, Matthew, Jeremy Pope, and Simon Jackman. 2008. "Measuring District-Level Partisanship with Implications for the Analysis of US Elections." *Journal of Politics* 70(3): 736–53.
- Lock, Kari, and Andrew Gelman. 2010. "Bayesian Combination of State Polls and Election Forecasts." *Political Analysis* 18: 337–48.
- McDonald, Michael P. 2007. "The True Electorate: A Cross-Validation of Voter Registration Files and Election Survey Demographics." *Public Opinion Quarterly* 71(4): 588–602.
- McDonald, Michael P., and Samuel L. Popkin. 2002. "The Myth of the Vanishing Voter." *American Political Science Review* 95(4): 963–74.
- Page, Benjamin I., and Robert Y. Shapiro. 1992. *The Rational Public: Fifty Years of Trends in Americans' Policy Preferences*. Chicago: University of Chicago Press.
- Park, David K., Andrew Gelman, and Joseph Bafumi. 2004. "Bayesian Multilevel Estimation with Poststratification: State-Level Estimates from National Polls." *Political Analysis* 12(4): 375–85.
- Putnam, Robert. 2001. *Bowling Alone: The Collapse and Revival of American Community*. New York: Touchstone Books.
- Rosenstone, Steven J., and John Mark Hansen. 1993. *Mobilization, Participation, and Democracy in America*. London: Longman.
- Schutt, Rachel. 2009. "Topics in Model-Based Population Inference." PhD thesis. Columbia University.
- Sigelman, Lee. 1982. "The Nonvoting Voter in Voting Research." *American Journal of Political Science* 26(1): 47–56.
- Silver, Brian, Barbara Anderson, and Paul Abramson. 1986. "Who Overreports Voting?" *American Political Science Review* 80(2): 613–24.
- Skocpol, Theda. 2004. *Diminished Democracy: From Membership to Management in American Civic Life*. Norman: University of Oklahoma Press.
- Skrondal, Anders, and Sophia Rabe-Hesketh. 2004. *Generalized Latent Variable Modeling: Multilevel, Longitudinal, and Structural Equation Models*. Boca Raton, FL: Chapman & Hall/CRC.
- Verba, Sidney, Kay L. Schlozman, and Henry E. Brady. 1995. *Voice and Equality: Civic Voluntarism in American Politics*. Cambridge, MA: Harvard University Press.
- Wattenberg, Martin P. 1986. *The Decline of American Political Parties, 1952–1984*. Cambridge, MA: Harvard University Press.
- Wedderburn, R.W.M. 1974. "Quasi-Likelihood Functions, Generalized Linear Models, and the Gauss-Newton Method." *Biometrika* 61(3): 439–47.
- Wolfinger, Raymond E., and Steven J. Rosenstone. 1980. *Who Votes?* New Haven, CT: Yale University Press.
- Zaller, John R. 1992. *The Nature and Origins of Mass Opinion*. Cambridge: Cambridge University Press.